THE STUDY OF THE BACTERIAL PATHOGEN *CLOSTRIDIOIDES DIFFICILE*

FROM GENETIC TOOLS TO *IN VITRO* AND *IN VIVO* PHYSIOLOGY

A Dissertation

by

KATHLEEN NICOLE MCALLISTER

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Joseph A. Sorg
Committee Members,	Jennifer K. Herman
	Deborah A. Siegele
	James L. Smith
Head of Department,	Thomas D. McKnight

December 2019

Major Subject: Microbiology

ABSTRACT

C. difficile is a Gram-positive, anaerobic gut pathogen which infects hundreds of thousands of individuals each year and is a significant concern as a nosocomial and community-acquired pathogen. Genetic tools are important when analyzing the physiology of such organisms so that the underlying physiology / pathogenesis of the organisms can be studied. We find that Stickland metabolism, a metabolic process that is used by only a small fraction of the microbiota, is important for *C. difficile* physiology. We first used TargeTron mutagenesis to investigate the role of selenoproteins in *C. difficile* Stickland metabolism. In this study, we found that a TargeTron insertion into *selD*, encoding the selenophosphate synthetase that is essential for the specific incorporation of selenium into selenoproteins, results in a significant growth defect and a global loss of selenium incorporation. However, because of potential polar effects of the TargeTron insertion as well as other drawbacks of other available genetic tools, we developed a CRISPR-Cas9 mutagenesis system for *C. difficile*. We then built upon our initial characterization of the CRISPR-Cas9-generated *selD* mutant by creating a CRISPR-Cas9-mediated restoration of the *selD* gene at the native locus and used these strains to analyze the importance of selenium-containing proteins on *C. difficile* physiology. Our findings support the hypothesis that selenium-containing proteins are important for several aspects of *C. difficile* physiology (*e.g.*, vegetative growth, spore formation, and outgrowth post-germination). Using RNAseq, we identified multiple candidate genes that likely aid the cell in overcoming the global loss of selenoproteins to

grow in medium which is favorable for using Stickland metabolism. Lastly, we analyzed samples from hospitalized patients for their bile acid content and abundances in order to study the effects of antibiotic treatment, diarrheal symptoms, *C. difficile* infection, and recurrence of infection. We found that groups of bile acids are associated with all health statuses of patients. These studies collectively give insight into the importance of *C. difficile* physiology and the development of genetic tools in an attempt to analyze pathways that will stop the *C. difficile* life cycle so that targeted therapeutics can be developed.

DEDICATION

I dedicate this to my husband, Jeremy. I would not be where I am today without him.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Sorg, for his patience and guidance to better myself as a scientist. I would also like to thank my committee members, Dr. Herman, Dr. Siegele, and Dr. Smith, for their scientific advice and for volunteering their time to serve on my committee.

I would like to sincerely thank my lab members for their friendship, scientific advice, and overall help (especially during animal experiments). I would also like to thank my colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Last and most important, I would like to thank my parents, Susan and Steven Strunk, and my husband, Jeremy McAllister, for their support and constant words of encouragement.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Joseph A. Sorg, Professor Deborah A. Siegele, and Professor James L. Smith of the Department of Biology and Professor Jennifer K. Herman of the Department of Biochemistry and Biophysics.

In Section 2, Dr. Laurent Bouillaut from Tufts University School of Medicine, Dr. William T. Self from the University of Central Florida, and Jennifer N. Kahn from Texas A&M University performed experiments and contributed significantly to this work for which they were authors on this publication. In Section 4, this work was performed in collaboration with Drs. Abraham L. Sonenshein and Yoav Golan from Tufts University School of Medicine. All other work conducted for the dissertation was completed by the student independently.

Funding Sources

This work was funded by the National Institute of Allergy and Infectious Disease (NIAID) under grant numbers 5R01AI116895 and 1U01AI124290 awarded to Joseph A. Sorg. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of Allergy and Infectious Disease or the National Institutes of Health.

NOMENCLATURE

A	adenine
acetyl-CoA	acetyl coenzyme A
aTet	anhydrous tetracycline
ATP	adenine triphosphate
BHIS	brain heart infusion medium supplemented with yeast extract
bp	base pair
C	cytosine
CA	cholic acid
Cas	CRISPR-associated
Cas9n	“nickase” Cas9
CDCA	chenodeoxycholic acid
CDI	*C. difficile* infection
CDMM	*C. difficile* defined minimal medium
cDNA	complementary DNA
CDT	*C. difficile* binary toxin
CFU	colony forming unit
CO	codon optimized
CO_2	carbon dioxide
CRISPR	clustered regularly interspaced short palindromic repeats
CRISPRi	CRISPR interference

crRNA	CRISPR RNA
DCA	deoxycholic acid
dCas9	catalytically “dead” Cas9
dH_2O	distilled water
DHHW	aspartic acid-histidine-histidine-tryptophan residues
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
EIA	enzyme immunoassay
ELSD	evaporative light scattering detector
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
FOA	5-fluoroorotic acid
G	guanine
gBlock	double-stranded DNA fragment from Integraded DNA Technologies
GCA	glycocholic acid
GCDCA	glycochenodeoxycholic acid
gRNA	guide RNA
GTPase	guanosine triphosphate (GTP) hydrolase
HA	homology arm
HDCA	hyodeoxycholic acid
HNH	histidine-asparagine-histidine residues
HPLC	high-performance liquid chromatography

ILV	isoleucine, leucine, or valine
kb	kilobase
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Luria-Bertani medium
LCA	lithocholic acid
mRNA	messenger RNA
NAD^+	oxidized nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NHEJ	non-homologous end joining
nt	nucleotide
OD_{600}	optical density at a wavelength of 600 nm
PAM	protospacer adjacent motif
PBS	phosphate buffered saline
PCR	polymerase chain reaction
qPCR	quantitative PCR
RAM	retrotransposable activated marker
RNA	ribonucleic acid
RNAP	RNA polymerase
RNA-seq	RNA sequencing
rRNA	ribosomal RNA
RT-qPCR	quantitative reverse transcription – PCR
SCLE	spore cortex lytic enzyme

SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sec	selenocysteine
SECIS	selenocysteine insertion sequence
sgRNA	single guide RNA
ssDNA	single-stranded DNA
T	thymine
TA	taurocholic acid
TCA	tricarboxylic acid cycle
TCDCA	taurochenodeoxycholic acid
tracrRNA	trans-activating crRNA
tRNA	transfer RNA
TY	tryptone yeast extract medium
TYG	TY medium supplemented with glucose

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION TO *C. DIFFICILE* AND CRISPR-CAS9 GENETIC TOOLS

1.1. *C. difficile* epidemiology

Clostridioides difficile, recently reclassified from and more commonly known as *Clostridium difficile* (1, 2), is a Gram-positive, anaerobic, spore-forming bacterium that has become the most frequent cause of antibiotic-associated diarrhea (3, 4). The bacterium, originally named *Bacillus difficilis*, was first isolated in 1935 from newborn infants' microflora (5). *C. difficile* toxins were later found to be the cause of antibiotic-associated pseudomembranous colitis (6). Now, *C. difficile* is a significant concern as a widely-distributed nosocomial pathogen and more frequently being found as community-acquired (4, 7). Nearly 15% of all hospitalized patients who are treated with antibiotics develop an antibiotic-associated diarrhea (8). Of those 15%, ~20% - 30% are caused by *C. difficile* (3). The frequency and severity of *C. difficile* infection (CDI) demands for continued development of potential therapeutics.

The most common risk factors for CDI include elderly age, immunocompromised status, hospitalization, and broad spectrum antibiotic treatment (e.g., clindamycin, fluoroquinolones and cephalosporins) (9, 10). Antibiotics alter the diversity of the colonic microflora that normally provides "colonization resistance" to pathogens such as *C. difficile* and this perturbation potentiates CDI (11-13). CDI symptoms range from mild diarrhea to pseudomembranous colitis and toxic megacolon (7). Vancomycin and metronidazole, both broad-spectrum antibiotics, are most frequently used to treat CDI (though metronidazole is no longer recommended to treat *C.*

difficile infection) (4, 14). Fidaxomicin is commonly used in treatment of recurrent infections due to its narrow spectrum and low rate of recurrence (15, 16). Due to the continued disruption of the colonic microbiota by the antibiotics used to treat CDI, between 15% to 35% of patients who are cleared of the primary symptoms of disease go on to recur with CDI, and these recurrences can lead to a more severe and prolonged infections (17, 18). In 2011 in the United States, approximately 500,000 cases of *C. difficile* infections were identified and nearly 29,000 of those resulted in death (19). Due to the rise of hypervirulent / epidemic strains and significant amounts of recurring infections, the Centers for Disease Control and Prevention has labeled *C. difficile* as an "urgent threat" to the US healthcare system (20).

Clinical manifestations of CDI are due, in part, from the production of toxins by *C. difficile* vegetative cells. The two toxins, TcdA (Toxin A; an enterotoxin (21)) and TcdB (Toxin B; a cytotoxin (21)), are essential for virulence (22, 23) and are encoded by the *tcdA* and *tcdB* genes, respectively. These genes are part of a large pathogenicity locus which encodes three other genes whose products are essential for toxin gene regulation (an alternative RNA-polymerase sigma factor, TcdR, a holin-like protein, TcdE, and an anti-sigma factor, TcdC) (21). A third toxin, CDT or binary toxin, is also produced by some *C. difficile* strains and is encoded by two genes: *cdtA* and *cdtB*. The role of CDT in pathogenesis is unclear (4). Toxin A and B both catalyze the glucosylation of Rho-family GTPases, inactivating them (24, 25). This inactivation leads to cell death and loss of barrier function in the host colonic epithelium (26).

1.2. *C. difficile* physiology: spores and germination

The mature spore is composed of a desiccated core surrounded by an inner membrane. Surrounding the inner membrane is a thin germ cell wall, a thick cortex peptidoglycan, an outer membrane and layers of coat proteins. The core contains the genomic DNA, RNA, and many enzymes needed for a vegetative cell outgrowth from the spore (27). To help protect the DNA from environmental insult (e.g., UV radiation, heat, and household chemicals), DNA is saturated with α / β type small acid soluble proteins (SASP) (28). Moreover, much of the water within the core has been replaced with pyridine-2,6-dicarboxylic acid (dipicolinic acid), which exists, primarily, as a 1:1 chelate of calcium (CaDPA) and accounts for as much as 10% of the spore's dry weight (27, 28). The inner spore membrane has a low permeability for small molecules, including water, and this further protects the core from damaging molecules (29, 30). The germ cell wall peptidoglycan surrounds the inner membrane and becomes the cell wall of the outgrowing cell (27, 31). The spore cortex is composed of specialized peptidoglycan containing muramic-δ-lactam residues which are recognized by spore cortex lytic enzymes (SCLEs) to degrade cortex during spore germination (27). The metabolically dormant spore is resistant to heat, common disinfectants, and antibiotics, which allows for the spore to survive on environmental surfaces and in a host during an antibiotic treatment (32, 33).

The *C. difficile* pathogenesis begins when the spores are either ingested by an individual from the environment or are already present within the host. Spores then germinate in response to small molecule germinants that trigger the start of the

transformation from a spore into a vegetative cell (34, 35). Vegetative *C. difficile* cells produce toxins and cause damage to the epithelium of the colon (21). The vegetative cells eventually sporulate (the signals that induce *C. difficile* spore formation are unclear) and escape into the environment to either be taken up by other hosts or remain in the same host to begin the disease process anew (36). If the life cycle were to be disrupted before toxin production could occur, CDI would be prevented. With this in mind, spore germination could also be an attractive target for potential therapeutics.

Previous work has demonstrated that taurocholic acid (TA), a bile acid, enhances recovery of *C. difficile* spores on agar medium (37, 38). Primary bile acids, cholic acid and chenodeoxycholic acid (CDCA), are produced in the liver and conjugated with either glycine or taurine (cholic acid + glycine = glycocholic acid, *etc.*) (Figure 1.1). The difference between these two primary bile acids is the presence of a 12α-hydroxyl group in cholic acid and its absence in CDCA. These bile acids are secreted from the liver and travel to the gallbladder where they are stored and released, most commonly when a meal has been ingested. Bile acids are released into the small intestines where the molecules aid in fat absorption. Although most bile acids are reabsorbed, approximately 5% - 10% (400-800 mg) of the total bile acids per day enter the large intestine where the resident microflora converts the molecules into secondary bile acids through deconjugation and 7α-dehydroxylation. Deoxycholic acid and lithocholic acid are both secondary bile acids which are derived from cholic acid and chenodeoxycholic acid, respectively (39). Bile acids belonging to the cholic acid family, in conjunction with glycine, are the most efficient germinants for *C. difficile* spores (35, 40, 41). Conversely,

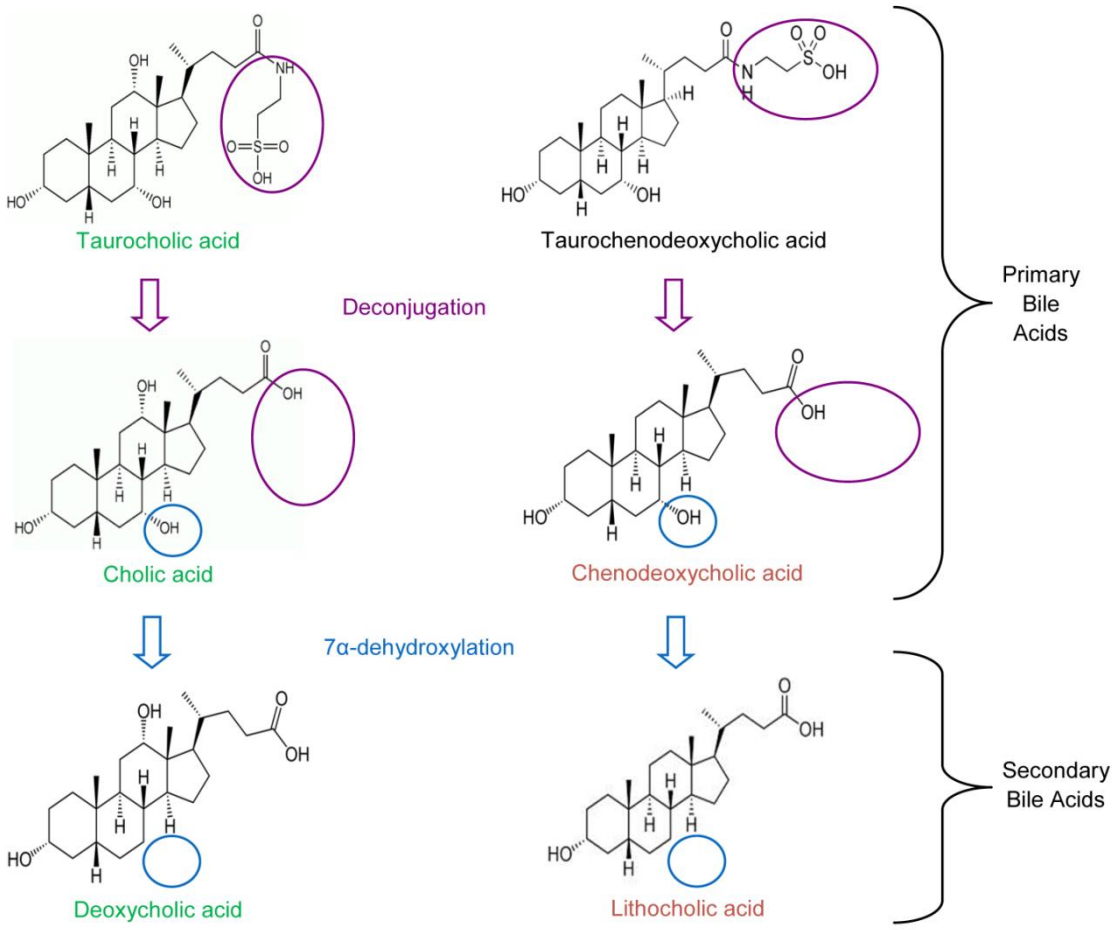


Figure 1.1 Bile acid families.

Primary bile acids, taurocholic acid and taurochenodeoxycholic acid, are modified by bacteria in the GI tract by removing the conjugated amino acid (deconjugation shown in purple). Primary bile acids can also be conjugated by glycine. Subsequently, cholic acid and chenodeoxycholic acid are further modified by other bacteria by removal of the 7α-hydroxyl group (shown in blue) to form secondary bile acids, deoxycholic acid and lithocholic acid, respectively. Cholic acid, a pro-germinating bile acid (shown in green), contains a 12α-hydroxyl group whereas, chenodeoxycholic acid, a growth-inhibitory and anti-germinating bile acid, does not (shown in red).

bile acids of the CDCA family are anti-germinants that competitively inhibit cholic acid-mediated germination (42, 43). Bile acids also play a role in inhibiting vegetative cell growth where deoxycholic acid and CDCA are two bile acids with this ability (40, 42). It is interesting that these two bile acids have an inhibitory role on vegetative cells since they belong to separate primary bile acid groups. Regardless, bile acids play an important role in *C. difficile* physiology, specifically on spore germination (16).

1.3. *C. difficile* physiology: growth

The metabolic and catabolic processes that occur during *C. difficile* growth have been studied in recent years and have suggested a large importance for carbohydrate and amino acid metabolism for survival of this bacterium. Pyruvate and acetyl-CoA are important metabolites that tie together multiple metabolic reactions. Pyruvate is formed by traditional glycolysis and gluconeogenesis pathways (44). Acetyl-CoA is not produced in traditional manners but is produced through the Wood-Ljundahl-pathway, also known as the reductive acetyl-CoA pathway, where two molecules of CO_2 act as terminal electron acceptors and are reduced to acetate (45). *C. difficile* also has an incomplete tricarboxylic acid cycle (TCA) which is proposed to be largely important for generating metabolically important intermediates rather than ATP generation and NADH regeneration (46, 47). These main metabolites are key for central carbon metabolism and contribute to the balance of NADH / NAD^+.

Fermentation pathways in *C. difficile*, specifically of amino acids, are possibly the major form of energy source *in vivo* (44, 48-50). Pyruvate is also used in fermentation pathways either by Stickland metabolism to produce propionate or by

electron bifurcation for butyrate production (44). The most well-known and prevalent pathway is Stickland metabolism where amino acids act as the sole carbon and nitrogen source (51-54). In these reactions, the oxidation and reduction of amino acids are coupled to NADH / NAD^+ generation, respectively (55). Many amino acids can be used in the oxidative branch of Stickland metabolism where isoleucine, leucine or alanine are the most common and efficient electron donors and are oxidatively decarboxylated or deaminated to produce the key products of NADH, different alcohol forms and ATP (53, 56, 57). Other amino acids produce other products, which are used in other pathways (44). In the reductive branch, proline or glycine act as electron acceptors to be reduced to 5-aminovalerate or acetate, respectively, and regenerate NAD^+ in the process. Proline and glycine are reduced by their respective reductases, proline reductase (PrdB) and glycine reductase (GrdA and GrdB) (54-57). These reductases are selenoproteins, meaning they contain selenocysteines in their primary amino acid sequences (55, 58).

Selenium is incorporated into proteins, and this occurs throughout all forms of life (59). Selenocysteine is made and incorporated into proteins through a selenocysteine synthesis pathway. SelD, a selenophosphate synthetase, is the first enzyme in this pathway and generates selenophosphate from selenide and inorganic phosphate. The selenophosphate is then incorporated into a serine-charged tRNA by selenocysteine synthase, SelA. Selenocysteines (Sec) are then incorporated during translation into elongating proteins, such as the proline and glycine reductases, by SelB and this is dependent on a selenocysteine insertion sequence (SECIS) comprised of a stem-loop to halt translation which is just downstream from an in-frame alternate codon, UGA. After

incorporation of the Sec, translation continues (59). It is unclear what the role(s) of selenoproteins are in *C. difficile* but they are hypothesized to be involved in growth due to the presence of selenoproteins in Stickland metabolism.

The importance of amino acids and their fermentation in *C. difficile* has been suggested by many researchers in studying gene transcripts and metabolites *in vivo* (11, 48-50, 60). Multiple studies have found high concentrations of proline or hydroxyproline in an antibiotic-treated or dysbiotic gut, potentially making the environment favorable for *C. difficile* colonization and growth (49, 61). Moreover, the levels of these amino acids as well as others decreased during *C. difficile* infection while products of Stickland metabolism like 5-aminovalerate increased (49, 50). Amino acids are not the only source of energy used *in vivo*. *C. difficile* utilizes carbohydrates, sugar alcohols or carboxylic acids as well, and this depends heavily on the available metabolites present at that time (49).

Until recently, the processes involved in *C. difficile* growth had been understudied (44). With the advent of more genetic tools, the study of *C. difficile* physiology has increased dramatically starting around the mid-2000s. Despite these advances, our understanding of *C. difficile* physiology has still lagged behind that of other bacteria due to the difficult manipulation of this organism (62).

1.4. Current state of genetic tools in clostridia

Clostridia are comprised of obligate anaerobic, endospore-forming firmicutes, and the best studied act as either important producers of industrially-relevant components (*e.g.*, *Clostridium acetobutylicum* or *C. cellulolyticum*) or pathogens (*e.g.*,

C. botulinum or *Clostridioides difficile*) (63, 64). Genetic tool development in clostridia increased over the last few years and continues to expand (65). Counter-selection tools for homologous recombination or allelic exchange is common across many bacterial organisms, including clostridia (66). Such counter-selection markers include I-*Sce*I (67), *upp* (68), *mazF* (69), *codA* (70), *galK* (71), *pyrE* (72) and *pyrF* (73). Though common, each tool has limitations, which have been discussed in length (62, 74-76). More recently, a single-stranded DNA annealing protein was used to engineer a recombineering system for *C. acetobutylicum* (77). This protein, RecT, is an ortholog from *C. perfringens* and is used by phage (*i.e.*, Rac or lambda prophages) to help increase the efficiency of homologous recombination of single-stranded DNA into bacterial genomes. Through the help of RecT, a short oligonucleotide is introduced into the target site (77, 78). This genetic tool shows the potential for exploitation of bacteriophage-derived single-stranded DNA binding proteins for homologous recombination (77).

In the mid-2000s, TargeTron technology (ClosTron) was repurposed for use in clostridia (79, 80). This development led to a surge in our understanding of the physiology and molecular processes that occur in clostridia, especially in *C. difficile*. The TargeTron system relies on the re-targeting of mobilizable group II introns to create insertion mutations into the desired target site. The inclusion of retrotransposable activated markers (RAMs) permits the activation of an antibiotic resistance gene upon splicing from the intron RNA, thereby providing antibiotic resistance upon insertion into the genome (79, 80). Despite the popularity of this tool, it has major drawbacks when

compared to clean deletions of target genes. The TargeTron target sites can be difficult to identify in coding regions that are less than 400 bp. Because of this, not every gene can be targeted for mutation (69, 74). Moreover, due to the nature of the mutation, the insertion results in polarity on downstream genes (74). Also targeting at some sites results in aberrant insertions at other sites in the genome. Lastly, the antibiotic resistance marker has led to perceived hypervirulence of mutant strains in the hamster model of *C. difficile* infection (81, 82).

Because of the above mentioned issues with current genetic systems available for modification of clostridia, there is a large need for fast and efficient genetic tools. With the advent of the CRISPR-Cas systems for genome editing, there has been a surge in the application of this tool in many organisms. The ease of use and pliability of these systems makes them an attractive target to be developed for use in clostridia.

1.5. CRISPR

The clustered regularly interspaced short palindromic repeats (CRISPR), along with their CRISPR-associated (Cas) proteins, have been developed into one of the most promising and successful genome editing tools to date (83). Although the significance was unknown at the time, the CRISPR system was first discovered in 1987 (84). CRISPR loci have been found in ~47% of bacteria, ~87% of archaea and have been suggested to have spread through horizontal transfer among prokaryotes as a basis of adaptive immunity. CRISPR arrays within bacteria and archaea are divided into three major types and sixteen different subtypes based on the *cas* genes present within the organisms (85).

CRISPR systems function as defense mechanisms to evade attacks from phage and other mobile genetic elements by incorporating a small fragment of the invader's DNA into the host's genome. When an invading organism inserts its DNA or RNA into the bacterial cell, the native Cas proteins help adapt to the infection by facilitating the incorporation of ~30 bp fragments of this invading DNA (spacers) into the CRISPR locus encoded on the host's chromosome. Within the CRISPR locus is a series of spacers, which have identity to prior adaptation events, flanked by repeat sequences of ~30 bp in length, forming the CRISPR array. The acquisition of spacers occurs in an ordered fashion where each spacer is incorporated at the beginning of the array behind the leader repeat sequence, resulting in a timeline of the previous infections (86).

The CRISPR array is constitutively transcribed into pre-crRNAs. These pre-crRNAs are processed and cleaved into individual crRNAs which are loaded onto a Cas endonuclease (87). In Type II CRISPR systems, a trans-activating crRNA (tracrRNA) is encoded upstream of the CRISPR locus and is essential for crRNA maturation by ribonuclease III and Cas9 (88). The tracrRNA and crRNA form a duplex that is loaded onto the Cas9 endonuclease (87). The most well-known and characterized Cas endonuclease is Cas9 from *Streptococcus pyogenes* (83). Once the Cas proteins are loaded with duplexed tracrRNA:crRNAs, the Cas-RNA complex will find complementary sequences within invading nucleic acid (89). The ability of the Cas protein to discriminate this foreign DNA from the complementary sequence encoded by the CRISPR array of the host is through the use of a protospacer adjacent motif (PAM). The PAM sequence is two to five nucleotides in length and found at either the 5' end or

3’ end of the protospacer sequence in the invader’s genome; the location and sequence of the PAM sequence recognized on the invading DNA is unique for each type of Cas endonuclease (90, 91). Upon recognition of the complementary sequence by the CRISPR RNA and the PAM by the Cas endonuclease, the Cas endonuclease introduces a double-stranded DNA break into the invading DNA or a single-stranded RNA break into the invading RNA (90-92).

Soon after the discovery of the function of the Cas9 endonuclease, the first CRISPR genome editing systems were developed in human cells, zebrafish embryos, and bacteria (*S. pneumoniae* and *Escherichia coli*) (93-100). In bacteria, a Cas endonuclease, a single guide RNA (sgRNA), and a donor region for homologous recombination are required for *in vivo* editing (Figure 1.2A) (100). The Cas endonuclease cleaves double-stranded DNA. In order for the endonuclease to cleave at a specific location, a guide RNA is used to direct the endonuclease to the intended target site (101). To simplify the system, the gRNA has been engineered to be produced as a single RNA molecule by fusing the crRNA and the tracrRNA. The result is a sgRNA which loads onto the endonuclease and directs it to the target site by binding to the DNA (91). *In silico* identification of sgRNAs are chosen where a PAM site is directly up- or down-stream from the target sequence (90). The sgRNA brings the Cas endonuclease to the intended target and endonuclease recognizes the PAM sequence and cleaves the DNA (90-92). To date, the most commonly used Cas endonuclease for genome editing in bacteria is Cas9 from *S. pyogenes*; the Cpf1 endonuclease from *Acidaminococcus sp.* is being developed as an alternative to Cas9 (83, 102, 103). Because many bacteria either

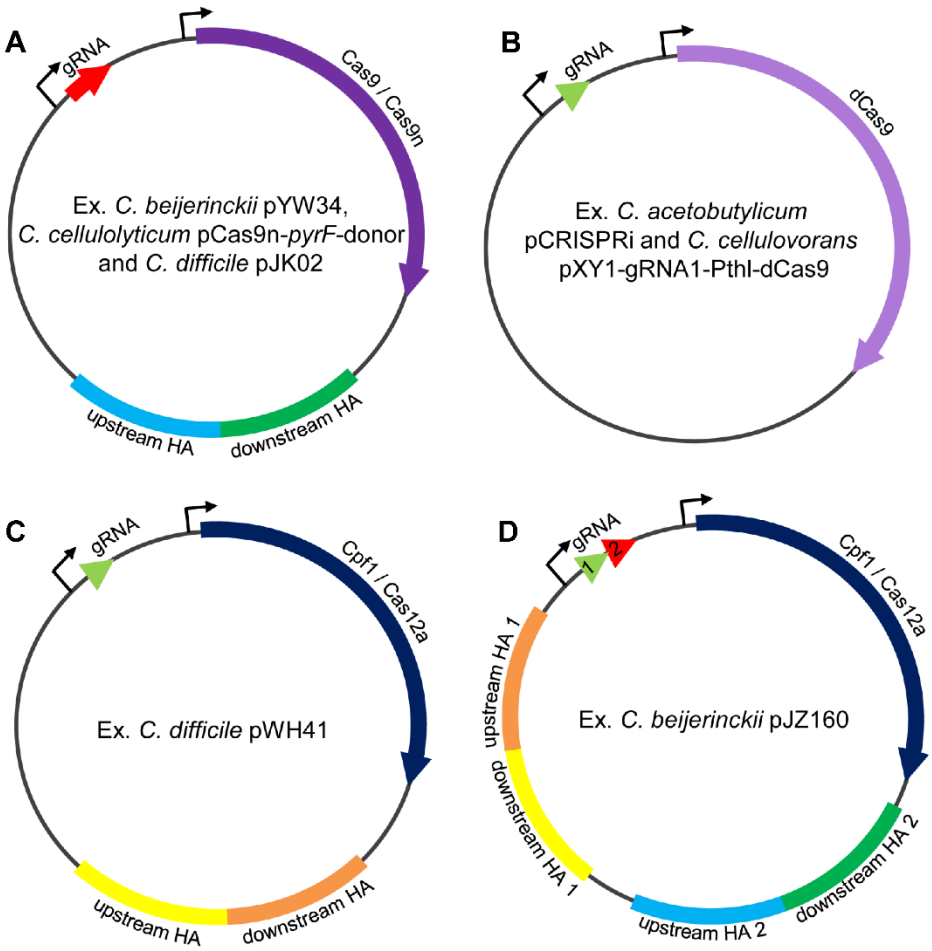
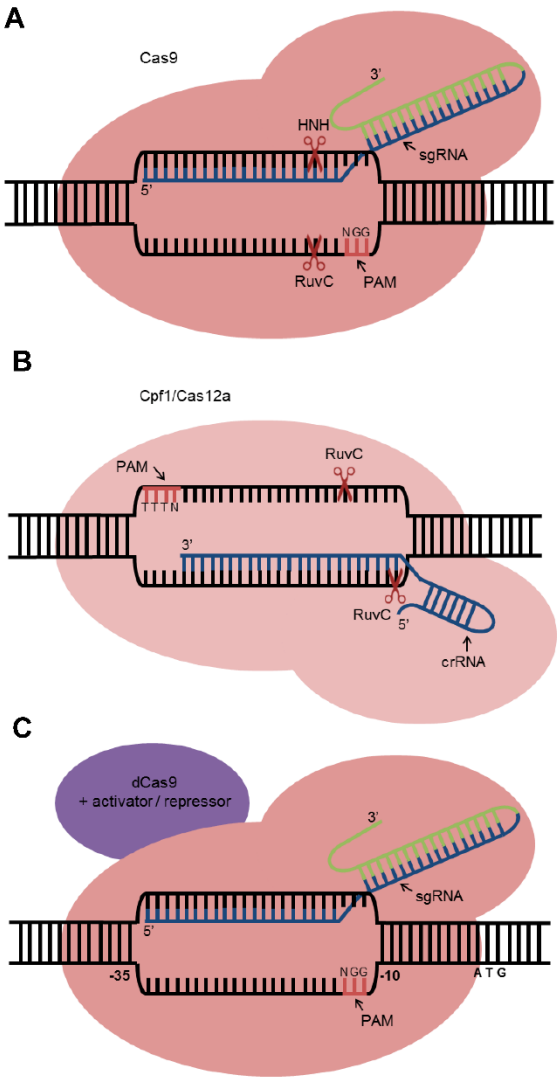


Figure 1.2 Comparison of CRISPR system plasmids.
Representative sample mutagenesis plasmids for CRISPR editing systems are shown. Each contains the three regions: a homology region (denoted by homology arms (HA)), an endonuclease (Cas9 or Cpf1), and a gRNA. Plasmid for editing using (A) Cas9 or Cas9n, (B) dCas9 with no necessary homology region, (C) Cpf1 / Cas12a with one target and (D) Cpf1 / Cas12a with two targets are shown. Examples of CRISPR-Cas mutagenesis plasmids are listed inside each plasmid.

do not have or have inefficient non-homologous end joining (NHEJ) repair systems, repair of double-stranded DNA breaks through homologous recombination is preferred (104). To facilitate this repair, a donor region that encodes the desired change in the genome is included in these systems (101). In prior work, Jiang *et. al.* (100) demonstrated that homologous recombination post-Cas9 cleavage of the DNA resulted in only a small increase in editing events. Their work suggests that Cas9 cleavage of the bacterial DNA counter-selects the population of cells that has not recombined with the donor region; recombination of the donor region with the chromosome removes the sgRNA site from the chromosome. Here, we review the developed CRISPR systems for genetic modification in clostridia, which greatly expanded in 2015 and continues to expand today.

1.6. Wild-type Cas9

In 2012, the function of the type II Cas9 endonuclease was characterized and would kick off the CRISPR-Cas9 genome editing revolution (91, 93, 94). The Cas9 enzyme relies on the base-pairing between the tracrRNA and the crRNA to cleave the complementary DNA adjacent to the PAM sequence (5' – NGG – 3' in *S. pyogenes*) located at the 3' end of the target sequence (90-92). Cas9 has two nuclease domains (HNH and RuvC-like) that are essential for the nuclease to induce a double-stranded DNA break (Figure 1.3A). The HNH domain, a ββα-metal fold which contains the active site, cleaves the DNA strand complementary to the targeting sequence located on the crRNA and the RuvC-like domain cleaves the opposite strand of DNA and shares an RNase H fold similar to the RuvC Holliday junction resolvase (91, 105, 106). Mutation



Mutant Alleles

Mutation	Catalytic activity	Name
D10A	ΔRuvC	Cas9n, "nickase"
D10A / H840A	Dead	dCas9 / CRISPRi

Figure 1.3 Review of CRISPR-Cas genetic modification systems.
Shown are graphical representations of each CRISPR-Cas system discussed in this review. Each contains the different endonucleases: (A) Cas9, (B) Cpf1/Cas12a, and (C) dCas9 with an activator / repressor, as well as the sgRNA or crRNA, PAM site and sequence, cleavage locations and their respective cleavage domains for each. A table is included in (A) to describe the different mutant alleles of Cas9. (C) also shows the promoter region area and the start codon for reference.

of either domain results in a Cas9 nuclease that only cleaves a single strand of DNA (Figure 1.3A). For example, introducing an aspartic acid to alanine mutation at the 10th amino acid (D10A) inactivates the RuvC-like domain (Cas9n) and results in the enzyme that only makes single-stranded DNA breaks ("nickase") (91). In addition, mutation of both nuclease domains, D10A and H840A, results in a catalytically dead nuclease, also known as dCas9, which will be discussed later (95, 107).

In some organisms, including clostridia, the expression of the Cas9 endonuclease has a high cost on the bacterial cell in terms of toxicity (74, 75, 108-111). Despite this cost, CRISPR-Cas9 systems have been successfully applied in *C. acetobutylicum* (75, 76, 112), *C. autoethanogenum* (113), *C. beijerinckii* (74, 75, 111, 114), *C. cellulolyticum* (109, 115), *C. difficile* (62, 116), *C. ljungdahlii* (117), *C. pasteurianum* (110, 112), and *C. saccharoperbutylacetonicum* (118) (Table 1.1) by tightly regulating the expression of the *cas9* gene. Specifically, the implementation of inducible promoters (*e.g.*, tetracycline-, xylose-, or lactose-inducible promoters) has helped to circumvent the issue of Cas9 toxicity (62, 76, 111, 116, 118). Except for *C. cellulolyticum, C. cellulovorans*, and *C. tyrobutyricum,* all of these clostridia use CRISPR-Cas9 systems that encode wild-type Cas9 and have mutation efficiencies exceeding ~50% (Table 1.1).

Since the regulation of Cas9 is an important factor for achieving high editing efficiencies, Cañadas *et al*. developed a "RiboCas" system which is a universal CRISPR-Cas9 editing tool for clostridium. Here they use a theophylline riboswitch to control translation of Cas9 by utilizing the ferredoxin (*fdx*) promoter and fusing the riboswitch with a synthetic Shine-Dalgarno sequence to the core region (-35 and -10). Using this

system, the authors were able to achieve editing efficiencies of 100% by targeting *spoIIE* in *C. pasteurianum*, *C. difficile*, *C. botulinum*, and *C. sporogenes*, demonstrating its versatility (119).

Table 1.1 CRISPR/Cas systems and their properties in clostridia.

Species	Endonuclease used				Homology arm lengths (bp)	endonuclease promoter	sgRNA promoter	Types of mutations	Editing efficiencies or Repression (%)	Reference
	Cas9	dCas9	Cpf1	endogenous						
C. acetobutylicum	S				500 to 1000	*thl*	sRNA sCbei_5830	replacements	100	(112)
	*CO				500 to 1000	*tet* (inducible)	*thl*	NM, small deletions, small replacements	100	(76)
	N				500 to 1000	*ptb* and *thl*	pJ23119	small deletions	7 to 100	(75)
		S			–	*thl*	sRNA sCbei_5830	knockdown	20 to 90	(112)
		S			–	*ptb*	pJ23119	knockdown	45	(75)
C. autoethanogenum	*CO				1000	IPL12	native Wood-Ljungdahl cluster	deletions	50 to 83	(113)
C. beijerinckii	S				1000	*spoIIE*	*thl*	small deletions	47 to 100	(74)
	S				1000	*spoIIE* and *lac* (inducible)	sRNA sCbei_5830	small deletions, small insertions, NM	0 to >99	(111)
	N				500 to 1000	*thl*	pJ23119	small deletions	19 to 100	(75)
	N				–	*thl*	pJ23119	random NM	20 to 94	(114)
		S			–	*thl*	sRNA sCbei_5830	knockdown	65 to 95	(120)
		S			–	*ptb*	pJ23119	knockdown	84	(75)
			A		500	*lac* (inducible)	sRNA sCbei_5830	small deletions	100	(121)
C. botulinum	S				~ 600 - 700	*fdx* + riboswitch-E	*araE*	large deletions	100	(119)

S – wild-type *S. pyogenes*, N – *S. pyogenes* Cas9 nickase, T – truncated / nickase *S. pyogenes*, CO - codon optimized *S. pyogenes*, A – *Acidaminococcus sp.*, P – *C. pasteurianum*, Ty – *C. tyrobutyricum*,
* - endonuclease was included on a separate plasmid from the other components (sgRNA and HAs), NR - not reported, – - not present, NM - nucleotide mutation, small deletion/insertion - smaller than 2 KB, large deletion/insertion - larger than 2 KB

Table 1.1 Continued.

Species	Endonuclease used				Homology arm lengths (bp)	endonuclease promoter	sgRNA promoter	Types of mutations	Editing efficiencies or Repression (%)	Reference
	Cas9	dCas9	Cpf1	endogenous						
C. cellulolyticum	N,CO				100 to 1000	*fdx*	P4 (synthetic)	small deletions, small insertions	<95 to 100	(109)
	N,CO				NR	*fdx*	P4 (synthetic)	small insertion	NR	(115)
C. cellulovorans		S			–	*thl*	pJ23119	knockdown	77 to 95	(122)
C. difficile	CO				1000	*tet* (inducible)	*gdh*	small deletions	20 to 50	(62)
	S				500 to 1000	*lac* (inducible)	sRNA sCbei_5830	small deletions and insertions	80 to 100	(116)
	T				1000	*thl*	*tcdB*	small and large deletions	89 to 96	(123)
	S				~ 600 - 700	*fdx* + riboswitch-E	*araE*	large deletions	100	(119)
		CO			–	*xyl* (inducible)	*gdh*	knockdown	~90	(124)
			A		500	*lac* (inducible)	sRNA sCbei_5830	small and large deletions, double deletions	25 to 100	(125)
C. ljungdahlii	S				1000	*thl*	*araE*	small and large deletions	50 to 100	(117)
		S			–	*lac* (inducible) and *tet* (inducible, from *C. acetobutylicum*)	P4 (synthetic)	knockdown	97 to 99	(126)
C. pasteurianum	S				1000	*thl*	sRNA sCbei_5830	small deletions	100	(112)
	S				1000	*thl*	sRNA sCbei_5830	small deletion	100	(110)
	S				~ 600 - 700	*fdx* + riboswitch-E	*araE*	large deletions	100	(119)
		S			–	*thl*	sRNA sCbei_5830	knockdown	85	(112)

S – wild-type *S. pyogenes*, N – *S. pyogenes* Cas9 nickase, T – truncated / nickase *S. pyogenes*, CO - codon optimized *S. pyogenes*, A – *Acidaminococcus sp.*, P – *C. pasteurianum*, Ty – *C. tyrobutyricum*,
* - endonuclease was included on a separate plasmid from the other components (sgRNA and HAs), NR - not reported, – - not present, NM - nucleotide mutation, small deletion/insertion - smaller than 2 KB, large deletion/insertion - larger than 2 KB

Table 1.1 Continued.

Species	Endonuclease used				Homology arm lengths (bp)	endonuclease promoter	sgRNA promoter	Types of mutations	Editing efficiencies or Repression (%)	Reference
	Cas9	dCas9	Cpf1	endogenous						
C. pasteurianum				P	1000	native CRISPR leader	–	small deletions	100	(110)
C. saccharoperbutyl-acetonicum	S				1000	*lac* (inducible)	sRNA sCbei_5830 and pJ23119	deletions	6 to 75	(118)
C. sporogenes	S				~ 600 - 700	*fdx* + riboswitch-E	*araE*	large deletions	100	(119)
C. tyrobutyricum				Ty	1000	*lac* (inducible)	–	small deletions	6.7 to 100	(127)

S – wild-type *S. pyogenes*, N – *S. pyogenes* Cas9 nickase, T – truncated / nickase *S. pyogenes*, CO - codon optimized *S. pyogenes*, A – *Acidaminococcus sp.*, P – *C. pasteurianum*, Ty – *C. tyrobutyricum*,
* - endonuclease was included on a separate plasmid from the other components (sgRNA and HAs), NR - not reported, – - not present, NM - nucleotide mutation, small deletion/insertion - smaller than 2 KB, large deletion/insertion - larger than 2 KB

A few clostridial CRISPR systems have used the Cas9n allele (75, 109, 114, 115). The major advantage to using this allele of Cas9 is in bacteria which have poor double-stranded DNA break repair systems or in bacteria where expressing both wild-type Cas9 and the gRNA is lethal (109, 128). By using the Cas9n allele, the organism can overcome the toxic effects induced by wild-type Cas9. In *C. cellulolyticum*, Cas9n was required to merely obtain transformants of the mutagenesis plasmid. In this study, however, the endonuclease was under the control of a constitutive promoter (109). It is unclear if the use of an inducible system in *C. cellulolyticum* would allow for tighter regulation of the wild-type Cas9 endonuclease and overcome the limitation of using the wild-type *cas9* allele. Interestingly, other clostridia that have developed Cas9n as a tool also have tools that use wild-type Cas9 (*i.e.*, *C. acetobutylicum* (75, 76, 112) and *C.*

beijerinckii (74, 75, 111, 114)). It is unclear if it is necessary to use the Cas9n allele in these systems; the editing efficiencies are high for both Cas9 alleles.

The most recent use of the Cas9n allele in *C. beijerinckii* is different than what is used in any other clostridia (114). Here, the nickase is fused to both a cytidine deaminase and a uracil DNA glycosylase inhibitor. In this unique system, the complex is guided to the target site by the sgRNA, a nick is made in the DNA, specific base-pair substitutions (C / G to T / A) are introduced, and the site is repaired. These changes result in possible missense mutations or null mutations in the targeted gene / coding region. The efficiencies reported here for base editing are comparable to those seen for gene editing using Cas9. While this method avoids the use of a donor region on the plasmid (decreasing the size of the plasmid as well as obviating the need of the bacteria to repair the lesion), it does not introduce specific mutations within the chromosome (114).

In another different use of a "nickase" Cas9, Nigel Minton's laboratory regenerated the *C. difficile* CD630Δ*erm* and CD630Δ*erm*Δ*pyrE* strains using a novel *C. difficile* CRISPR-Cas9 gene editing system where Cas9 naturally mutated to a truncated form. In this truncated Cas9, a frameshift resulted in a premature stop codon with a start codon and ribosomal binding site located downstream. In this second portion of Cas9, the RuvC domain and should therefore act as a "nickase". They were able to achieve high editing efficiencies in an isogenic background of 89 – 96% (123).

One aspect of the CRISPR-Cas9 system is the composition of the target sequence and the design of the sgRNA that recognizes this sequence (*e.g.*, G+C content and specific nucleotide positions within the intrinsic sequence). Oddly, none of the

developed clostridial CRISPR-Cas9 systems - have discussed this aspect of the system. The authors report differences in editing efficiencies when targeting different genes, but there are no direct discussions about the sgRNAs used or if this impacted the efficiencies of gene editing. In all of these systems, one sgRNA is used to make one mutation. Whether or not the authors tested other potential target sites within the mutated gene is unclear. The inclusion of multiple sgRNAs within the same plasmid to make multiple mutations at one time has not been accomplished in clostridia using the wild-type Cas9 allele. To accomplish this, each sgRNA would require its own promoter to drive its expression, or different spacers could be used if the tracrRNA is expressed separately. Unfortunately, this would result in an increase to the size of the mutagenesis plasmid which may result in issues when working with such a large plasmid. Smaller promoter regions such as the *C. beijerinckii* sCbei_5830 small RNA promoter (110-112, 120) or the *C. cellulolyticum* P4 synthetic promoter (109, 115, 126) used in some clostridia CRISPR systems could be functional in other *Clostridium* species. Both promoters, sCbei_5830 small RNA promoter and P4 promoter, are small (42 bp and 36 bp, respectively) and are functional in several clostridia (Table 1.1) (109-112, 115, 116, 118, 120, 121, 125, 126). This is promising for those clostridia which do not have many genetic tools and for which strong promoters are unknown.

The use of a donor region for homologous recombination is often necessary in bacteria to achieve high editing efficiencies (100). A double-stranded chromosomal break is highly toxic to cells and recombination of the homology region with the chromosome protects against Cas9 targeting by removing the targeting site (109, 129).

Nearly 28% of bacteria have the NHEJ component, Ku (130). NHEJ components are either not present in some bacteria or their expression is low. This poses a problem for genome editing when trying to repair a Cas9-induced double-stranded DNA break. Without added expression of NHEJ components, such as on a multi-copy plasmid, repairing the lesion using this method is not practical (131). Therefore, homologous recombination-based protection is much more efficient, and likely more plausible. The most commonly used homology arm length in clostridia CRISPR-Cas systems is 1 kilobase each with 500 base pairs each being the second most common(62, 74-76, 109-113, 116-118). The size of the homology arms may differ for each bacteria or CRISPR-Cas9 system, but the lowest reported homology region that achieves reasonable editing efficiency in clostridia is 100 bp (109).

1.7. dCas9/CRISPRi

CRISPR-Cas9-mediated genome editing creates a marker-less edit within the target DNA (*e.g.*, bacterial chromosome or plasmid) (83). Not all genes can be targeted as some genes are essential for the survival of the bacteria. For these reasons, a genetic system which does not make chromosomal deletions and only regulates the expression of genes is necessary (132). CRISPRi, or CRISPR interference, has been developed where both catalytic sites of Cas9 have been mutated: D10A (located in the RuvC-like domain) and H840A (located in the HNH domain) (Figure 1.3A) (91, 107). These mutations render Cas9 catalytically dead. Here, Cas9 and the sgRNA function to bind to and block a region of DNA from transcriptional activity. First developed and implemented in *E. coli* (107), the authors suggest that the Cas9:sgRNA complex collides

with the elongating RNA polymerase (RNAP). They found that targeting the template strand permits RNAP to read through and not come in contact with the Cas9:sgRNA complex. Thus, targeting the non-template strand of DNA has been reported to yield higher repression than targeting the template strand in bacteria (107). A common use of this tool is to target the promoter region of a gene or operon to downregulate expression upon induction of the CRISPRi system (107, 133). A positive attribute of this system is that the plasmid size for the CRISPRi system is smaller than that of the original genome editing CRISPR-Cas systems in that no donor region for DNA repair is necessary (Figure 1.2B). More importantly the functionality or characterization of essential genes can be explored using the CRISPRi system (132).

A few *Clostridium* species have developed CRISPRi systems: *C. acetobutylicum* (75, 112), *C. beijerinckii* (75, 120), *C. cellulovorans* (122), *C. difficile* (124), *C. ljungdahlii* (126), and *C. pasteurianum* (112) (Table 1.1). In all cases, the successful repression of genes in these organisms ranges from 20% (*e.g.*, plasmid-encoded *afp* in *C. acetobutylicum* (112)) to 99% (*e.g.*, *pta* in *C. ljungdahlii* (126)). These values depended on the gene being targeted and the target sequence used for the sgRNA. The target sequence chosen for the implementation of CRISPRi is more important here due to the function of repression; an inefficient target sequence may not hold dCas9 on the DNA as tightly or consistently compared to a more efficient target sequence. Li *et al*. targeted the *spo0A* gene in both *C. acetobutylicum* and *C. beijerinckii*, using the same system, and obtained different repression percentages, 45% and 84%, respectively (75). The two sgRNA sequences are only 35% identical. Because the two sequences are different, and

in different organisms, direct comparisons between the two studies cannot be made. Similarly, the location of targeting has an effect on efficiency of repression. Target sites that are further away from the transcription start site have lower efficiencies than those closer to the transcription start site. Moreover, the use of multiple sgRNAs increases the efficiency of repression, as long as the target sequences do not overlap. Finally, truncated sgRNA target sequences, those with less than 12 bp, do not result in repression of the target gene; full length sgRNAs, those of 20 bp in length, are preferred for efficient repression (107). One issue with the CRISPRi genetic tool is that of polarity on downstream genes. The effects of knocking down one gene in an operon will likely have an effect on the downstream genes.

It has previously been demonstrated that a system can be made where dCas9 can be coupled to an activator or repressor to regulate transcription of a specific gene (Figure 1.3C) (133, 134). While it has been demonstrated that CRISPRi works effectively in some clostridial species, it would be interesting to apply a system which uses a transcriptional activator coupled to dCas9 to enhance transcription of specific genes. For example, fine tuning of biofuel production and increasing the use of clostridia to generate valuable end products.

1.8. Cpf1/Cas12a

Cpf1 (Cas12a) is a type V CRISPR system effector protein which has been studied in *Francisella novicida, Acidaminococcus sp*. (AsCpf1), *Moraxella bovoculi*, and *Lachnospiraceae bacterium* (108). The Cpf1 endonuclease specifically recognizes T-rich PAMs instead of G-rich PAMs, as in the case for Cas9 (106). In another

divergence from how the Type II CRISPR system works, Cpf1 is itself responsible for the maturation of pre-crRNA; no tracrRNA is needed. AsCpf1 is guided to its target by the mature crRNA to recognize the PAM sequence 5' – TTTN-N23 – 3' (*i.e.*, 5' – TTTN – 3' followed by a 23 bp protospacer). AsCpf1 then cleaves the double-stranded DNA resulting in a staggered, 5-nt 5' overhang that is 18-23 bp downstream from the PAM site (103). Unlike Cas9, AsCpf1 only uses one, RuvC-like, domain to digest both DNA strands rather than one domain for each strand of DNA (Figure 1.3B) (103, 135). Even though this endonuclease is not as well-studied/characterized as the Cas9 endonuclease, Cpf1 has been used for genome engineering in *Escherichia coli*, *Yersinia pestis*, *Mycobacterium smegmatis*, and *Corynebacterium glutamicum* (136, 137) as well as in *C. beijerinckii* and *C. difficile* (Table 1.1) (121, 125).

One advantage to using Cpf1 over Cas9 is that the T-rich PAM sequence used by Cpf1 are expected to be more probable in AT-rich organisms, such as clostridia, as well as promoter regions, which are commonly AT-rich (106). It has been suggested that Cpf1 is better suited for bacteria, such as *C. difficile*, which have low DNA conjugation efficiencies. This suggestion is based on the described lower toxicity of the Cpf1 endonuclease (125). These toxicity claims were based upon a study from *Corynebacterium glutamicum* where the authors could not obtain transconjugants when introducing their CRISPR-Cas9 plasmids; however, they were able to obtain several transformants from the same system using Cpf1 instead of Cas9 (137). A separate study found that Cas9 is toxic when compared to Cpf1 in *Synechococcus* 2973 (138). However, because the sample size is low, the lower toxicity of Cpf1 compared to Cas9

may be organism-dependent. Finally, Cpf1 was suggested to have lower off-target effects than Cas9 (125, 139). However, these studies were conducted in human cells and were not tested in bacteria (139, 140). In *C. difficile*, Cas9 has no known or detectable off-target effects as of the date of this review (62).

Another proposed advantage to using Cpf1 is that since a tracrRNA is not needed, the cost for constructing and using plasmids would be cheaper due to the shorter gRNA (Figure 1.2C). For Cpf1, a gRNA consisting of only 42 nucleotides would need to be synthesized compared to the >100 nucleotide sgRNA (crRNA and tracrRNA) needed for other type II endonucleases, such as *S. pyogenes* Cas9 (106, 138). Moreover, the smaller size of the gRNA could lead to the use of multiple gRNAs on a single plasmid to simultaneously make multiple mutations in a single application of this system (Figure 1.2D) (121, 125, 141).

CRISPR-Cpf1 systems have been used in two clostridia, *C. beijerinckii* and *C. difficile*, and the efficiencies of mutagenesis of these systems range from 25% (*i.e.*, *C. difficile ermB1 / ermB2*) to 100% (*i.e.*, *C. beijerinckii spo0A*, *C. beijerinckii pta* and *C. difficile fur*) (121, 125). These values are similar to those obtained from using wild-type Cas9 (Table 1.1). More genes would have to be targeted to determine if there are significant differences between using these two endonucleases in their respective CRISPR systems. Using *C. difficile* as an example, the promoters used in Wang *et al*. (Cas9) and Hong *et al*. (Cpf1) for the endonuclease and the sgRNA are the same. Even though the efficiencies for the CRISPR-Cas9 system are higher than that of Cpf1, few genes were targeted and the same genes were not targeted between the studies (Table

1.1) (116, 125). As for *C. beijerinckii*, the same mutation efficiencies were reported for two genes, *spo0A* and *pta*, using either Cas9 or Cpf1 as the endonuclease. In this study, the same gRNA was not used in each system (121). Based on these few studies, there is low evidence of either endonuclease, Cas9 or Cpf1, as being superior to the other in clostridia.

1.9. Endogenous CRISPR systems

In order to overcome the toxic effects of *S. pyogenes* Cas9 or *Acidaminococcus sp.* Cpf1 endonucleases, some genome editing systems were designed that rely on the endogenous CRISPR array for editing (142-144). 74% of clostridial species harbor CRISPR-Cas loci, including many of the well-studied clostridia (110, 127, 145-155). For instance, *C. pasteurianum* and *C. tyrobutyricum* have developed CRISPR-Cas genome editing systems based on their respective endogenous systems (Figure 1.4, Table 1.1) (110, 127).

As a general starting place to develop endogenous CRISPR systems as genetic tools, Pyne *et al*. analyzed the Cas proteins within the CRISPR array in *C. pasteurianum* and determined the PAM sequences by analyzing spacer sequences within the CRISPR array to find common nucleotide sequences among them. The authors then developed a single-plasmid system which mimics the native CRISPR system by including a synthetic CRISPR array and crRNAs with spacers corresponding to the target region which was to be mutated. The authors also predicted PAM sequences for three other clostridia: *C. autoethanogenum, C. tetani,* and *C. thermocellum* (110). Zhang *et al*. also developed an endogenous CRISPR-Cas system but for *C. tyrobutyricum*. The authors used similar

approaches described above to analyze the CRISPR arrays and determined the PAM sequence for *C. tyrobutyricum*. In addition, the authors multiplexed the gRNAs and made simultaneous genomic deletions using one CRISPR-Cas plasmid construct (127).

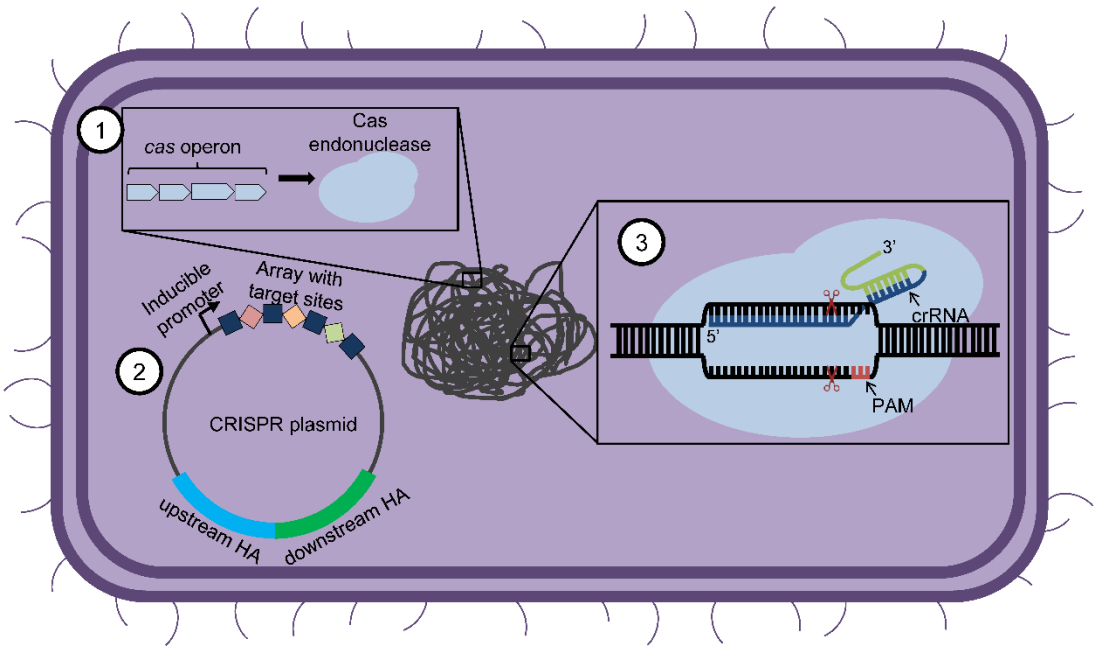


Figure 1.4 Endogenous CRISPR genome editing in clostridia.
Shown is a graphical representation of CRISPR-Cas genome editing in a clostridia vegetative cell. (1) The endogenous CRISPR region contains a *cas* operon in which a Cas endonuclease is encoded and subsequently generated. (2) Meanwhile, a plasmid containing a synthesized CRISPR array under the control of an inducible promoter and a donor region for homologous recombination containing an upstream homology arm (HA) and a downstream HA. When induced, the plasmid transcribes the crRNAs to be used by the endogenous system to generate individual crRNAs. (3) The endogenous Cas endonuclease complexes with a crRNA to target and cleave the DNA. This is then repaired by the donor region located on the CRISPR plasmid.

Both of the tools using endogenous CRISPR-Cas systems differ from those using Cas9 or Cpf1 in that the endogenous systems rely on the endonucleases encoded on the genome to form mature crRNAs as well as make the double-stranded DNA breaks within the genome; no endonucleases are encoded within the plasmid. Genome editing

control lies within the design of the CRISPR array containing the pre-crRNAs as well as the homology arms to be used as donor regions for homologous recombination. The CRISPR array would contain multiple pre-crRNAs, all under the control of a single promoter. Once transcribed, the endogenous system processes the pre-crRNAs into mature crRNAs that are then loaded onto the endogenous Cas endonucleases. This unique system potentially allows for multiple targets to be engineered using a single plasmid system. This avoids the use of multiple different promoters for each gRNA (110, 127, 143).

The development of a CRISPR-Cas gene editing tool using an endogenous system requires the presence of a known CRISPR system within that organism. Due to a large number of bacteria encoding CRISPR-Cas systems, there is great potential for the development of this tool in non-model organisms (85). For those organisms which have previously identified or well-studied endogenous CRISPR systems, the development for genetic modification is streamlined. For those that do not have characterized systems, there is work to be done prior to developing the genetic tool. The type / class of CRISPR array needs to be determined and this will help determine how the system will function. Another vital part of developing an endogenous CRISPR system is identifying the PAM recognition sequence for the encoded endonuclease. Once the CRISPR array and PAM are identified, the endogenous CRISPR system could be exploited for genome editing (110, 127). Unfortunately, different strains of an organism could encode slightly different components to their CRISPR systems and, therefore, may have different requirements for editing (*e.g.*, PAM recognition). This could be seen as a drawback for

exploiting the endogenous CRISPR locus, but the benefits could outweigh the long development time to establish other genetic systems. Because the majority of clostridia have endogenous CRISPR systems, these could be exploited for use as genetic tools (110, 156). In particular, those bacteria which have little to no genetic tools available are good candidates for using this type of genetic engineering since it relies on existing components of the genome, provided the bacterium can easily be transformed or have an effective conjugation system.

1.10. Conclusions

While much has been elucidated about *C. difficile* physiology, in particular growth and spore germination, there is much to be learned. In order to study these processes in *C. difficile*, efficient, fast, and user-friendly genetic systems are needed. Here, we developed a *C. difficile* CRISPR-Cas9 genetic tool to be used by the *C. difficile* community which is more effective and time-efficient than other genome editing tools developed to date. We then applied this system to study the global role of selenoproteins in different aspects of the *C. difficile* life cycle. Since metabolites have an important impact on the different stages of the *C. difficile* life cycle, we analyzed the role of bile acids in human patient samples with different health statuses relating to antibiotic treatment, CDI and recurrence of infection.

2. USING CRISPR-CAS9-MEDIATED GENOME EDITING TO GENERATE *C. DIFFICILE* MUTANTS DEFECTIVE IN SELENOPROTEINS SYNTHESIS[*]

In Section 2, we used TargeTron to investigate the role of selenoproteins in *C. difficile* Stickland metabolism and found that a TargeTron insertion into *selD*, encoding the selenophosphate synthetase that is essential for the specific incorporation of selenium into selenoproteins, results in a significant growth defect and a global loss of selenium incorporation. However, because of potential polar effects of the TargeTron insertion, we developed a CRISPR-Cas9 mutagenesis system for *C. difficile*. This system rapidly and efficiently introduces site-specific mutations into the *C. difficile* genome (20% - 50% mutation frequency). The *selD* CRISPR deletion mutant had a growth defect in protein-rich medium and mimicked the phenotype of a generated TargeTron *selD* mutation. Our findings suggest that Stickland metabolism could be a target for future antibiotic therapies and that the CRISPR-Cas9 system can introduce rapid and efficient modifications into the *C. difficile* genome.

2.1. Introduction

Clostridioides difficile, more commonly known as *Clostridium difficile* (1, 2), is a Gram-positive, anaerobic, spore-forming bacterium that is the major cause of antibiotic-associated diarrhea (3, 4). Most-commonly, patients undergoing antibiotic

[*] Reprinted with permission from "Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis" by McAllister, K.M. et al., 2017, *Scientific Reports*, Vol 7, 14672, Copyright 2017 by Springer Nature.

treatment are at high risk for *C. difficile* infection (CDI) due to the disruption of the normal colonic microbiota by broad-spectrum antibiotics (157). Spores, the metabolically-dormant form of *C. difficile* that can survive passage between hosts in the aerobic environment, are ingested by a host and germinate in response to small-molecule germinants (*i.e.*, cholic acid, and its derivatives, and an amino acid, such as glycine) into a vegetative, toxin-producing cell (34, 40, 158). These toxins, TcdA and TcdB, damage epithelial cells which results in the symptoms of CDI (7, 159, 160).

Much of our understanding of *C. difficile* physiology has come in the last few years and coincided with the development of genetic tools for this organism (70, 79, 161-164). The tools available to genetically manipulate *C. difficile* include: i) single-crossover integration of segregationally unstable plasmids(162, 163); ii) mobile, group II introns (TargeTron / ClosTron technology) (79); iii) Allelic-Coupled Exchange using either the *codA* or *pyrE* systems (70, 161); and iv) Mariner transposition (164, 165). To date, the most widely used system is the TargeTron (or ClosTron) system which relies on the re-targeting of mobilizable group II introns and the use of retrotransposable activated markers (RAM) (80). Though RAM markers allow for the easy identification of potential mutants, unfortunately, this system only creates insertion mutations resulting in potential polar effects on downstream genes. In addition, the inserted antibiotic resistance marker has led to perceived hypervirulence of mutant strains in the hamster model of CDI (81, 82).

Segregationally unstable plasmids can be used to create single insertions into the *C. difficile* genome (162, 163) or can be used as allelic exchange plasmids using *codA*

counter selection or 5-fluoroorotic acid (FOA) in a pre-generated *pyrE* mutant (70, 161). Due to the segregationally unstable nature of the plasmids, daughter cells may not receive a copy of the plasmid and are killed by the antibiotic in the surrounding medium. Strains with single-integration events of the plasmid, due to homologous recombination, grow more rapidly. These 'large' colonies are then spread on nutrient-poor medium supplemented with 5-fluorocytosine (for *codA*-based plasmids) or FOA for the *pyrE* allelic exchange system (70, 161). Due to nutrient-poor media being a requirement for counter-selecting the integrated plasmids, mutations that generate slow-growth phenotypes or loss of metabolic pathways may result in difficulties in growth on such medium. Moreover, the *pyrE* system requires the correction of the pre-generated *pyrE* mutation before progressing experimentally – increasing the effort and time required to generate mutant strains (161). On the other hand, advantages to this system are the ability to create single nucleotide mutations and clean deletions and the ability to integrate a single-copy chromosomal restoration in strains.

In a previous study, the TargeTron system was used to create insertions in genes (*i.e.*, *prdR*, *prdB* and *grdA*) whose products are involved in Stickland metabolism (55). Stickland reactions are a primary source of energy for a small group of anaerobic bacteria grown that typically use amino acids as their sole carbon and nitrogen sources (166). During these reactions, one amino acid, such as alanine or leucine, is oxidatively decarboxylated or deaminated (166). Subsequently, in the reductive branch, D-proline or glycine acts as electron acceptors. In the case of D-proline the amino acid ring is reduced and converted to δ-amino valeric acid (58). In the case of glycine reductase,

two selenoproteins are involved (GrdB and GrdA) and the glycine is deaminated in a process that results in acetyl-phosphate (and thus ATP) production. PrdB, GrdA and GrdB are the selenium-containing subunits of the respective reductases and their expression is regulated by PrdR (55). These enzymes are important for growth in protein-rich medium supplemented with proline or glycine indicating that selenium-containing enzymes are important for *C. difficile* physiology (55).

Selenophosphate synthetase (SelD, CDR20291_2388) has been shown to be necessary for the activation of selenium for specific incorporation into biological macromolecules in several bacterial model systems (59). Selenium incorporation in *C. difficile* is likely to be dependent on a selenophosphate synthetase (SelD) which generates selenophosphate from selenide and inorganic phosphate that is incorporated into a serine-charged tRNA by selenocysteine synthase (SelA, CDR20291_2387) (167). The selenocysteine is then incorporated into proteins such as PrdB and GrdA during translation with the aid of a selenocysteine-specific elongation factor, SelB (CDR20291_2386) (167). SelD, SelA and SelB are encoded in a single genetic locus and likely are part of the same transcriptional unit. Selenium is used for other enzymatic processes as well and requires SelD to generate selenophosphate for these processes (59). In order to test the importance of selenium-containing factors on *C. difficile* growth, we engineered a mutation in *selD* using the established TargeTron gene knock out system and found that *selD* is important for *C. difficile* growth and incorporation of selenium into proteins. To avoid any potential polar effects of the TargeTron system, we developed a CRISPR-Cas9 mutagenesis system and used this system to engineer a *selD*

in-frame deletion mutation. Our results highlight the importance of selenoproteins in *C. difficile* physiology and suggest that these proteins could be used as targets for future antibiotic therapies. Moreover, this newly developed *C. difficile* genetic system can be used to rapidly and efficiently introduce mutations into the *C. difficile* genome.

2.2. Materials and methods

2.2.1. *Bacterial strains and growth conditions*

C. difficile strains (Table 5.1) were routinely grown in an anaerobic atmosphere (10% H_2, 5% CO_2, 85% N_2) at 37°C in brain heart infusion medium supplemented with 5 g / L yeast extract and 0.1% L-cysteine (BHIS), as described previously (15, 34, 168, 169) or TY medium (3% tryptone, 2% yeast extract) (170). For conjugation experiments, cells were plated on BHI agar medium, as described previously for *E. coli*-based conjugations (171), or on TY medium for *Bacillus subtilis*-based conjugations. For selenium incorporation (see below), strains were grown in TY (1% tryptone, 0.5% yeast extract) (58). Where indicated, growth was supplemented with taurocholate (TA; 0.1% w / v), thiamphenicol (10 µg / mL), kanamycin (50 µg / mL), D-cycloserine (250 µg / mL), erythromycin (5 µg / mL) and/or glucose (1% w / v) as needed. Induction of the CRISPR-Cas9 system was performed in TY medium (170) supplemented with thiamphenicol (10 µg / mL) and anhydrous tetracycline (aTet; 100 ng / mL). A defined minimal medium for *C. difficile* growth (CDMM), described previously (70, 172), was used for selection of *pyrE* mutants by supplementation with 5-fluoroorotic acid (FOA; 2 mg / mL) and uracil (5 µg / mL). *E. coli* strains (Table 5.1) were routinely grown at 37°C in LB medium. Strains were supplemented with chloramphenicol (20 µg / mL),

kanamycin (50 µg / mL), and/or ampicillin (100 µg / mL) as needed. *B. subtilis* BS49 was routinely grown at 37°C in BHIS broth or on LB agar plates. Strains were supplemented with chloramphenicol (2.5 µg / mL) and/or tetracycline (5 µg / mL).

2.2.2. *Plasmid construction and molecular biology*

The JIR8094 *selD* TargeTron insertion was created in several steps. First, plasmid pBL38 was constructed by retargeting of the group II intron from pCE240 (173), using the primers oLB70, oLB71, oLB72 and EBS-Universal, as outlined in the TargeTron user manual (Sigma-Aldrich), followed by initial cloning of the retargeted fragment in pCE240 digested with *BsrG*I and *Hind*III. The retargeted group II intron from pBL38 was then extracted by digestion with *Sfo*I and *Sph*I and cloned between the *Sph*I and *SnaB*I sites of pMC123 (174), resulting in pBL54. Conjugation experiments between *C. difficile* and *E. coli* were carried out as described previously (175). *C. difficile* transconjugants were selected on BHIS plates supplemented with D-cycloserine, kanamycin, and thiamphenicol and potential TargeTron mutants were identified by plating on erythromycin. Erythromycin-resistant colonies were screened for the insertion of the intron into *C. difficile selD* by PCR using primers specific for full-length *C. difficile selD* (oLB76 and oLB77). A positive clone, strain LB-CD7, was identified. To identify if the TargeTron system integrated at sites other than the *selD* gene, we sequenced the *C. difficile* LB-CD7 strain and the parent JIR8094 strain. No additional TargeTron insertions were found except the one in *selD*.

To construct the CRISPR-*cas9* plasmid, the constitutively-expressed *cwp2* promoter was chosen to drive expression of a sgRNA targeted to *spoVAC*.

Oligonucleotides were designed and the fragments generated were stitched together by PCR SOEing with Phusion DNA polymerase (due to the nature of the A : T-rich sequence, the entire fragment could not be synthesized directly and thus had to be stitched together by PCR) (all oligonucleotide sequences can be found in Table 5.2). The first round of amplification was done with primer sets gRNA_1_for and gRNA_2_rev, gRNA_3_for and gRNA_4_rev, and gRNA_5_for and gRNA_6_rev. The resulting fragments, gRNA_1_2 and gRNA_3_4, were used along with primers gRNA_1_for and gRNA _4_rev in a PCR to yield fragment gRNA_1_4. Fragment gRNA_5_6 was expanded in a PCR using primers gRNA_5_for and gRNA_7_rev to yield fragment gRNA_5_7. The complete sgRNA was made by PCR sowing of fragments gRNA_1_4 and gRNA_5_7 using primers 5'gRNA and 3'gRNA. This DNA fragment was introduced into pJS116 (all plasmid descriptions can be found in Table 5.1) at the *Pme*I restriction site using Gibson Assembly (176) and transformed into *E. coli* DH5α to generate pKM22. The *tetR* repressor gene along with the P_{tet} promoter for conditional expression of *cas9* was amplified by PCR from pRPF215 (165) (Table 5.1) using primers 5'MTL_tetRprom and 3'tetR_Cas9. *cas9* from *Streptococcus pyogenes* was codon optimized for expression in *C. difficile* by Thermo Fisher (Thermo Fisher Scientific, Waltham, MA). Codon-optimized *cas9* was amplified using primers 5'tetR_CO_cas9 and 3'MTL_CO_cas9 from pMK-RQ-Bs-cas9. To introduce a D10A mutation into *cas9*, we used primer set 5'tetR_CO_cas9_D10A and 3'MTL_CO_cas9 for PCR amplification. The P_{tet} promoter and the wild-type *cas9* and $cas9_{D10A}$ alleles were

introduced into pKM22 at the *Hind*III restriction site using Gibson Assembly to generate pKM46 and pKM48, respectively.

To facilitate future crRNA changes and increase the expression of the sgRNA, the stronger, constitutively-expressed *gdh* promoter was used to replace the *cwp2* promoter. Also, *Kpn*I and *Mlu*I restriction sites were added at the 5' and 3' ends of the sgRNA, respectively. 5'gdh and 3'gdh_gRNA were used to amplify the *gdh* promoter from *C. difficile* R20291. The sgRNA was amplified from pKM22 using primers 5'gRNA_gdh and 3'gRNA 2. The resulting two fragments were both inserted into the *Pme*I and *BsrG*I sites using Gibson Assembly and transformed into *E. coli* DH5α to generate pKM54 and pKM55.

In order to easily select for mutants, we designed a plasmid to target the *pyrE* gene. The donor region for homology directed repair was made such that 1-kb upstream and 1-kb downstream were stitched together to generate a clean deletion of *pyrE*. 1-kb upstream and 1-kb downstream of *pyrE* were separately amplified by PCR from *C. difficile* R20291 using primers 5'pyrE_UP and 3'pyrE_UP and 5'pyrE_DOWN and 3'pyrE_DOWN, respectively. The two resulting fragments were inserted into the *Not*I and *Xho*I restriction sites in pKM54 and pKM55 background by Gibson Assembly and transformed into *E. coli* DH5α to generate pKM64 and pKM65, respectively. A gBlock (Integrated DNA Technologies, Coralville, IA), pyrE_gRNA_gBlock, was designed which contained the sgRNA DNA sequence between, and including, the *Kpn*I restriction site and the *Mlu*I site. This DNA fragment was introduced between the *Kpn*I and *Mlu*I restriction sites of pKM64 and pKM65 to generate pKM71 and pKM72, respectively.

To introduce the plasmids into *C. difficile* by conjugation with *E. coli*, the *B. subtilis Tn*916 *oriT* was replaced with the *E. coli traJ* gene. *traJ* was amplified from pMTL84151 by PCR using primers 5'traJ and 3'traJ. The resulting fragment was introduced into pKM71 and pKM72 using the *Apa*I restriction site by Gibson Assembly and transformed into *E. coli* DH5α to generate pJK02 (accession number MF782679) and pKM93 respectively.

The subsequent CRISPR-Cas9 plasmid targeting *selD* was made using pJK02. The homology regions for *selD* targeting were amplified by primer sets 5'MTL_selD_UP and 3'MTL_selD_UP and 5'MTL_selD_DN and 3'MTL_selD_DN. The resulting fragments were cloned by Gibson assembly into pJK02 at the *Not*I and *Xho*I restriction sites and transformed into *E. coli* DH5α resulting in pJS170. The gBlock for *selD* targeting sgRNA, CRISPR_selD_183, was introduced by ligation into the *KpnI* and *MluI* sites and transformed into *E. coli* DH5α resulting in pJS187. The *selD* targeting plasmid was modified by replacing *traJ* with oriT *tn916* for *B. subtilis* conjugation by amplification from pJS116 using primers 5'Tn916ori and 3'Tn916ori. The resulting fragment was introduced into pJS187 by Gibson assembly at the *Apa*I site and transformed into *E. coli* DH5α resulting in pJS194.

To make a plasmid which would complement the *selD* mutation, *selD* along with 500 bp upstream to include the native promoter was amplified by primer sets 5'selD_comp and 3'selD_comp. The resulting fragment was cloned by Gibson assembly into pJS116 at the *Not*I and *Xho*I restriction sites and transformed into E. coli

DH5α resulting in pKM142. The sequences of all plasmids were verified by DNA sequencing.

2.2.3. *Conjugation for CRISPR-Cas9 and complementation plasmid insertion*

All complete CRISPR-Cas9 and complement plasmids were transformed into either *E. coli* HB101 pRK24 or *B. subtilis* BS49 to be used as donor for conjugation with *C. difficile*. For *E. coli* conjugations, the strains were grown overnight at 37°C in LB supplemented with ampicillin and chloramphenicol. *C. difficile* R20291 was grown anaerobically in TY medium overnight. Five hundred microliters of *C. difficile* overnight culture / mating was heated to either 52°C for 5 minutes for an 8 hour conjugation or 50°C for 15 minutes for a 24 hour conjugation, as described previously (171). *C. difficile* cultures were removed from the heat block and let cool to 37°C for 2 minutes. Meanwhile, 1 mL of *E. coli* HB101 pRK24 containing the CRISPR-Cas9 plasmid cultures were pelleted at 4,000 x g for 2 minutes and the supernatant was removed. The *E. coli* pellets were transferred to the anaerobic chamber and gently suspended in the heat shocked *C. difficile* sample. The resulting mix was plated onto pre-reduced BHI agar plates by spotting ten, 20 µL drops of culture. After either 8 or 24 hours, the growth was harvested by collecting in 1 mL pre-reduced TY broth. One hundred microliters of the resuspended growth was plated onto multiple BHIS agar plates supplemented with thiamphenicol, kanamycin, and D-cycloserine. Growth was monitored for 2 to 3 days. Individual colonies were restreaked for isolation and tested for insertion of plasmid by PCR amplification of the *catP* gene with primers 5’ catP 3 and 3’ catP 2.

For *B. subtilis* conjugation, *C. difficile* R20291 was grown anaerobically in BHIS broth overnight. The *C. difficile* overnight culture was diluted in fresh pre-reduced BHIS broth and grown anaerobically for 4 hours. Meanwhile, *B. subtilis* BS49 was grown aerobically at 37°C in BHIS broth supplemented with tetracycline and chloramphenicol for 4 hours. One hundred microliters of each culture was plated on TY agar medium. After 24 hours, the growth was harvested by suspending in 2 mL pre-reduced BHIS broth. One hundred microliters of the resuspended growth was spread onto several BHIS agar plates supplemented with thiamphenicol, kanamycin, and D-cycloserine. *C. difficile* transconjugants were screened for the presence of Tn*916* using tetracycline resistance, as described previously. Thiamphenicol-resistant, tetracycline-sensitive transconjugants were selected and used for further experiments.

2.2.4. *Radiolabeling studies with ^{75}Selenium*

Selenium is taken up with high affinity and specifically incorporated into macromolecules through exposure of cells to ^{75}Se in the form of selenite (59). For these studies, a 1:100 dilution of overnight cultures were added to 10 mL TY medium supplemented with approximately 10 µCi ^{75}Se in the form of selenite (corresponding to 50 nM cold). After 24 hours growth, nine milliliter cultures were grown overnight in 12 x 75 mm capped culture tubes in an atmosphere of 95% nitrogen and 5% hydrogen. Cells were harvested by centrifugation (5,000 x g for 5 minutes) and resuspended in a small amount (typically 0.2 mL) of lysis buffer (50 mM Tris, pH 8.0, 0.1 mM benzamidine, 0.5 mM EDTA). Cells were lysed by sonication (model 100 Fisher Scientific) for short 10 second bursts until lysis was seen. The crude cell lysates were

further clarified by centrifugation (12,500 x g) for 10 minutes at 4°C. Protein concentrations were determined by Bradford assay (177) using albumin to generate a standard curve. Radiolabeled selenoproteins were separated by SDS-PAGE (15% resolving gel) and, after the gels were dried, selenoproteins were identified by phosphorimager analysis (Molecular Dynamics phosphorimager).

2.2.5. *Induction of the CRISPR-Cas9 system and isolating mutants*

C. difficile R20291 strains containing the *pyrE*-targeting plasmids were grown overnight in TY medium supplemented with thiamphenicol. In the morning, 250 µL of an overnight culture was diluted into 4.75 mL of fresh TY medium supplemented with thiamphenicol and aTet (100 ng / mL) and grown for 6 hours. Subsequently, cultures serially diluted and spread on CDMM medium supplemented with FOA and uracil. Colonies were isolated; DNA was extracted and tested for the desired mutation by PCR amplification of the target gene, 5'pyrE 2 and 3'pyrE 2 (Table 5.2). Once an isolate was confirmed, it was passaged ~ 3 times in BHIS liquid medium in order to lose the CRISPR-Cas9 plasmid. After pick-and-patch on BHIS agar with and without thiamphenicol, loss of plasmid was confirmed by PCR amplification of the *catP* gene using primer set 5' catP 3 and 3' catP 2. *C. difficile* R20291 strains containing the *selD* targeting plasmid was induced for 24 hours. Then ~10 µL of culture was spread onto BHIS medium. Colonies were tested as described above using primer sets 5'selD and 3'selD. Confirmed isolates were passaged on BHIS agar once in order to lose the CRISPR-Cas9 plasmid due to the slow growth of the mutant. Loss of plasmid was

confirmed by PCR amplification of the *catP* gene using primer set 5' catP 3 and 3' catP 2.

2.2.6. *RT-PCR*

RNA was extracted from wild-type *C. difficile* R20291 pMTL84151 (empty vector) and *C. difficile* R20291 pJK02 induced for 30 minutes with or without aTet using a FastRNA Blue Kit (MP Biomedical). DNA contamination was eliminated by using a TURBO DNA-*free* Kit (Thermo Scientific) according to the standard protocol. cDNA was made using the SuperScript III First-Strand Synthesis System (Thermo Scientific) according to the protocol, including controls for each sample without the presence of reverse transcriptase. To determine if *cas9* and the gRNA were being transcribed, the 5' end of *cas9*, sgRNA, and *catP* were amplified from isolated cDNA in a PCR using primers sets 5'COcas9_RT and 3'COcas9_RT, 5'gRNA_RT and 3'gRNA_RT, and 5'catP_RT and 3'catP_RT and Taq DNA polymerase.

2.2.7. *Illumina sequencing*

High-quality, high-molecular weight genomic DNA from *C. difficile* R20291 (WT), aTet-induced *C. difficile* R20291 pMTL84151 (empty vector), two isolates of *C. difficile* KNM5, *C. difficile* LB-CD7, *C. difficile* KNM6 pJS116, and *C. difficile* KNM6 pKM142 was extracted as described previously (175, 178). The genomic DNA was submitted to Tufts University School of Medicine Genomics Core facility for Paired-End 50 Illumina re-sequencing as described previously (34). Alignment and analysis of the sequences was performed using DNASTAR Lasergene program MegAlign Pro 14.

2.2.8. *Determining mutation efficiencies*

C. difficile R20291 strains containing the *pyrE* targeting CRISPR-Cas9 plasmids were induced as described above. Induced cultures were serially diluted and 100 µL was spread on CDMM supplemented with uracil and CDMM supplemented with FOA and uracil. After 4 days, colony forming units (CFUs) were counted for each dilution on each media and the total CFU/mL of the mutants (those on CDMM-FOA and uracil) and the total cell count (CDMM-uracil) were calculated.

C. difficile R20291 strains containing the *selD* targeting CRISPR-Cas9 plasmid was induced as described above. A loop containing ~10 µL of culture was spread onto rich BHIS agar medium. Individual colonies were isolated; DNA was extracted, and tested for the desired mutation by PCR amplification of the target genes using primers 5'selD and 3'selD (Table 5.2). These oligonucleotides only amplify DNA from the chromosome regardless of whether the CRISPR-Cas9 plasmid is present or not.

2.2.9. *Statistical analysis*

Data points represent the mean from two or three independent experiments and, where indicated, error bars represent one standard deviation from the mean.

2.2.10. *Availability of materials and data*

The datasets generated during and/or analyzed during the current study are available in the NCBI SRA repository. The first accession number SRP115702 includes the following sequences and accession numbers: R20291 (SRX3104072), R20291 foaR (SRX3104073), and two KNM5 isolates (SRX3104071 and SRX3104074). The second accession number SRP119051 includes the following sequences and accession numbers:

LB-CD7 (SRX3236353), KNM6 pJS116 (SRX3236354), and KNM6 pKM142 (SRX3236352). Finally, the pJK02 plasmid generated in this study is freely available to the scientific community.

2.3. Results

2.3.1. *Generation of a TargeTron mutation in* C. difficile *JIR8094* selD

In order to investigate the role of selenoproteins on *C. difficile* physiology, we generated a TargeTron insertion into the *selD* gene of *C. difficile* strain JIR8094. *C. difficile selD* is the first gene in the operon and is upstream of the genes encoding a selenocysteine synthase (*selA*) and a selenocysteine-specific elongation factor (*selB*) (Figure 2.1A). The parental strain, JIR8094, and the *selD* mutant, LB-CD7, grew to nearly equal levels in rich BHIS medium (Figure 2.1B). We next tested the growth of the two strains in protein-rich conditions; tryptone is a rich source of amino nitrogen (179) and we reasoned that this medium may favor Stickland metabolism. In either TY or TYG medium, growth of strain LB-CD7 was significantly decreased compared to that of its parent strain (Figure 2.1C).

2.3.2. C. difficile selD::ermB *does not incorporate selenium into proteins*

The growth defect observed for strain LB-CD7 (*selD::ermB*) suggests that selenoproteins are important for growth. To confirm that the *selD* mutation led to a loss of selenium incorporation during protein synthesis, we measured the incorporation of radioactively labeled selenium in TY medium (Figure 2.2). A 1:100 dilution of overnight cultures were added to 10 mL TY medium supplemented with approximately 10 µCi of ^{75}Se in the form of selenite (50 nM cold). After 24 hours cultures were

harvested by centrifugation, lysed by brief sonication and clarified cell extracts were analyzed by SDS-PAGE.

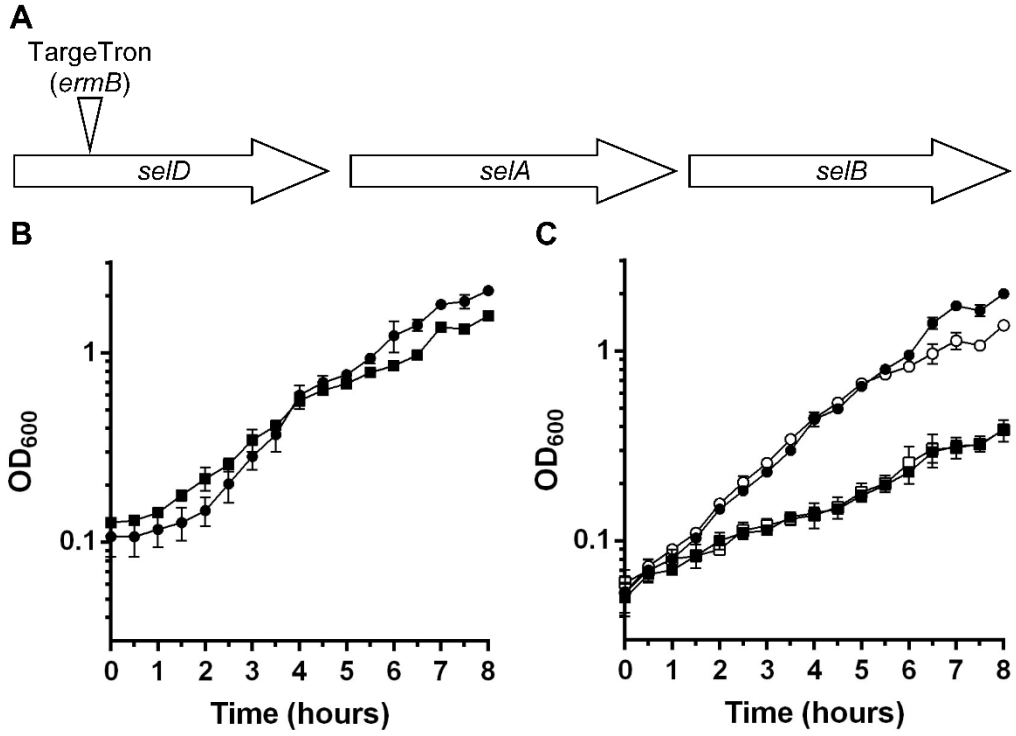


Figure 2.1 *C. difficile selD::ermB* has a defect in growth compared to the WT strain.

(A) Genetic organization of the *selD* locus: *selD* (CDR20291_2388), *selA* (CDR20291_2387) and *selB* (CDR20291_2386). The location of the TargeTron insertion into *selD* is illustrated. (B) *C. difficile* JIR8094 (●) and *C. difficile* LB-CD7 (*selD::ermB*) (■) were grown in BHIS medium and growth was monitored over time. (C) *C. difficile* JIR8094 was grown in TYG (●) or TY (○) medium and *C. difficile* LB-CD7 was grown in TYG (■) or TY (□) medium and growth was monitored over time. In both graphs, data points represent the average from three independent experiments and error bars represent the standard deviation of the mean.

When separated by SDS-PAGE and analyzed using autoradiography, we detected three distinct bands in the parental strain JIR8094 (Figure 2.2). These likely correspond to GrdB (largest), PrdB and GrdA (smallest band) based on a previous study (58). In both *prdB* and *prdR* TargeTron mutants, the PrdB band was lost; PrdR is required for *prdB* expression (55) (Figure 2.2). Similarly, in the *grdA* TargeTron mutant, both GrdA and GrdB proteins are not present; due to the insertion of the group II intron into *grdA* there were polar effects on *grdB*. Significantly, in two separate *selD::ermB* isolates (LB-CD7), we observed a lack of radioactive signal, suggesting that this mutation prevents the incorporation of selenium into all three of these established proteins thereby limiting optimal growth (Figure 2.2) (58).

2.3.3. *Generation of a CRISPR-Cas9 mutagenesis system for use in* C. difficile

Because the generated TargeTron insertion into *C. difficile selD* is likely polar on the downstream genes (*selA* and *selB*), we sought to generate a non-polar deletion in *selD* to confirm its growth disadvantage in protein-rich medium. However, the current state of *C. difficile* genetics requires that strains be isolated under nutrient-poor conditions. Because the growth of the *C. difficile selD* deletion may behave similarly to the TargeTron insertion, we sought to develop a mutagenesis technique that permits the isolation of mutants under nutrient-rich conditions. The CRISPR-Cas9 system was an attractive target. To design a Cas9-producing plasmid (Figure 2.3), we placed a wild-type *cas9* gene, which was codon-optimized for expression in *C. difficile*, under the conditional expression of the tetracycline-inducible *tetR* promoter (165). Also, based on the finding that non-homologous end-joining in *C. cellulolyticum* is inefficient (109), we

engineered *cas9*$_{D10A}$ to determine if the native double-stranded break repair system in *C. difficile* is also inefficient; the Cas9$_{D10A}$ protein functions as a 'nickase' and does not introduce double-stranded breaks into the targeted DNA (180).

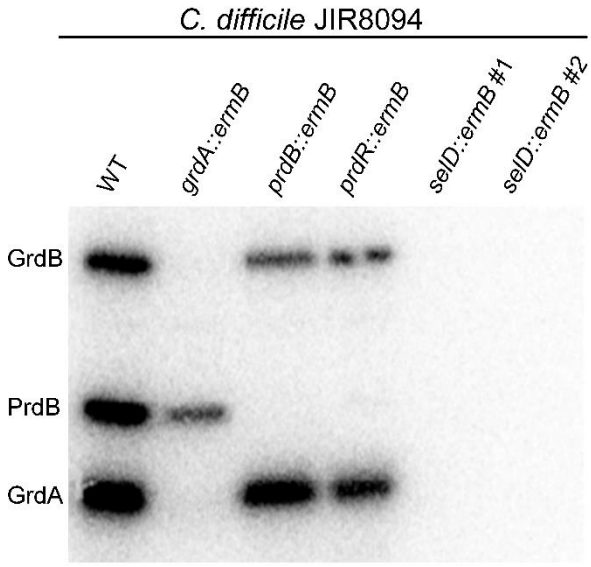


Figure 2.2 *C. difficile selD::ermB* does not incorporate selenium into selenoproteins. *C. difficile* strains were grown in TY medium and 10 µCi of ^{75}Se was added to the culture medium during growth. After 24 hours of growth, cultures were harvested, cells were lysed and samples from clarified lysates were separated by SDS-PAGE (15%). Radioactively-labeled protein was detected using phosphorimage analysis. GrdB, PrdB and GrdA are labeled based on previously published data (58).

The expression of the single guide RNA (sgRNA) that directs Cas9 to the intended target site was placed under the control of the native glutamate dehydrogenase (*gdh*) promoter (181, 182). This promoter is constitutively expressed and allows for sufficient levels of the sgRNA to be present within the cell for Cas9 to act upon. We engineered the sgRNA to be produced as a single RNA molecule by fusing the crRNA

and tracrRNA, as previously described (83). Potential crRNA target sites were determined using an algorithm provided by the CRISPRscan.org website (183) and sites were chosen within the first 200 bp of the gene.

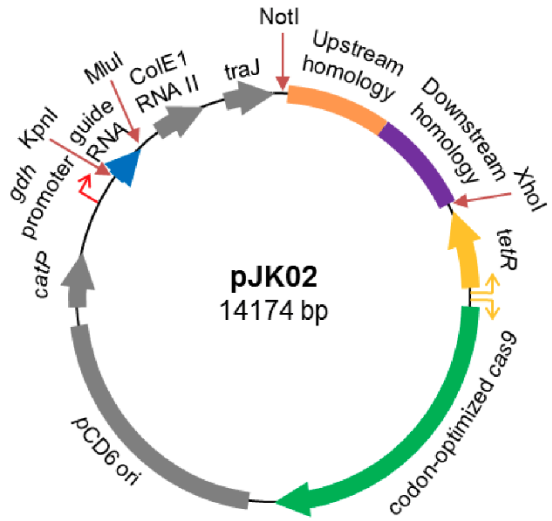


Figure 2.3 *C. difficile* CRISPR-Cas9 plasmid map.
The pMTL84151 backbone, depicted in gray, consists of the pCD6 *C. difficile* (Gram-positive) replicon *oriV*, *orfB* and *repA*, the thiamphenicol resistance marker *catP*, the Gram-negative replicon *colE1*, and *traJ* for conjugal transfer from *E. coli*. The modifications to the backbone, depicted in various colors, consists of the insertions of the targeting region for homologous recombination, sgRNA under the expression of the *gdh* promoter, and the *tetR* promoter that regulates the expression of the *S. pyogenes cas9* gene that was codon optimized for expression in *C. difficile*.

The final portion of the *C. difficile* CRISPR-Cas9 plasmid is the donor region which is necessary to insert a desired mutation. This 2-kb region (one kb upstream of the targeted DNA and one kb downstream of the targeted DNA) surrounds the region to be deleted. The function of this donor region is to provide a template for the native DNA repair system to correct the Cas9-mediated double-stranded DNA (dsDNA) or single-stranded DNA (ssDNA) break.

2.3.4. *Isolation of a CRISPR-Cas9 mediated* pyrE *mutation*

To test the efficiency of the system in *C. difficile*, the first gene targeted for deletion was *pyrE*. The *pyrE* mutant strain is a uracil auxotroph and resistant to FOA-mediated toxicity. Therefore, mutations can easily be selected by incorporating FOA into growth medium. To engineer the deletion, we cloned 1-kb upstream of the *pyrE* start codon and 1-kb downstream of the stop codon in the mutagenesis plasmid (Figure 2.4A). Next, a site near the 5' end of the *pyrE* sequence was chosen for the crRNA. The resulting plasmid, pJK02, was introduced into *C. difficile* R20291 by conjugation from *E. coli*. The *tetR* promoter system was induced by aTet and mutants were isolated by selecting for those that were resistant to FOA. Colonies containing the mutation were confirmed by PCR amplification of the *pyrE* gene and surrounding DNA (Figure 2.4B). A mutation in *pyrE* results in a 585-bp deletion within the 1.59 kb *pyrE* coding sequence (Figure 2.4A and 2.4B). Subsequently, the phenotype of the *pyrE* mutant strain, *C. difficile* KNM5, was confirmed by plating on defined medium or medium supplemented with uracil (Figure 2.4C and 2.4D). Only the wild-type strain was able to grow on medium without uracil supplementation. Surprisingly, we observed a *pyrE* deletion in the uninduced R20291 (pJK02) strain, which was to be used here as a control (Figure 2.4B). This strain was also a uracil auxotroph (Figure 2.4C and 2.4D). Based on results from a previous study (184), we hypothesized that the *tetR* promoter has leaky expression and could lead to a small amount of transcription of *cas9* in the absence of aTet. To confirm this, we extracted RNA from an uninduced culture and amplified a portion of *cas9* (Figure 2.4E). As expected, in the absence of induction, we observed

cas9 transcript without amplification of contaminating DNA (Figure 5.1). These results suggest that the *tetR* promoter is not tightly regulated and uncontrolled expression of *cas9* led to a deletion of *pyrE* in the uninduced strain.

Incorporation of FOA in the growth medium permits the selection against *pyrE*-containing strains, but not all mutations lend themselves to a selection process. Thus, we determined the efficiency of the CRISPR-Cas9-mediated *pyrE* mutations. *C. difficile* R20291 containing the CRISPR-Cas9 plasmid (pJK02) was induced and the resulting cells were plated on defined medium alone or medium supplemented with FOA. The number of colonies were enumerated and the efficiency of mutagenesis was calculated (Table 2.1). Surprisingly, wild-type Cas9 yielded a mutation frequency of approximately 50%. Moreover, this mutation frequency was dependent on the ability of Cas9 to introduce a dsDNA break at the target site. That is, the $Cas9_{D10A}$ protein yielded a much lower mutation frequency (~2×10^{-4}, or 1 mutation in every 5,000 cells). This mutation frequency is above the empty vector control (~4×10^{-7}), suggesting that *C. difficile* R20291 can use homology-directed repair to replace ssDNA breaks, but not efficiently. Importantly, the high mutation frequency of the wild-type Cas9 plasmid was not merely due to homologous recombination of the plasmid with the genome. If this were the case, the $cas9_{D10A}$ plasmid should have yielded a similar frequency. These results suggest that the wild-type *cas9*-containing plasmid is best suited for introducing mutations into the *C. difficile* genome.

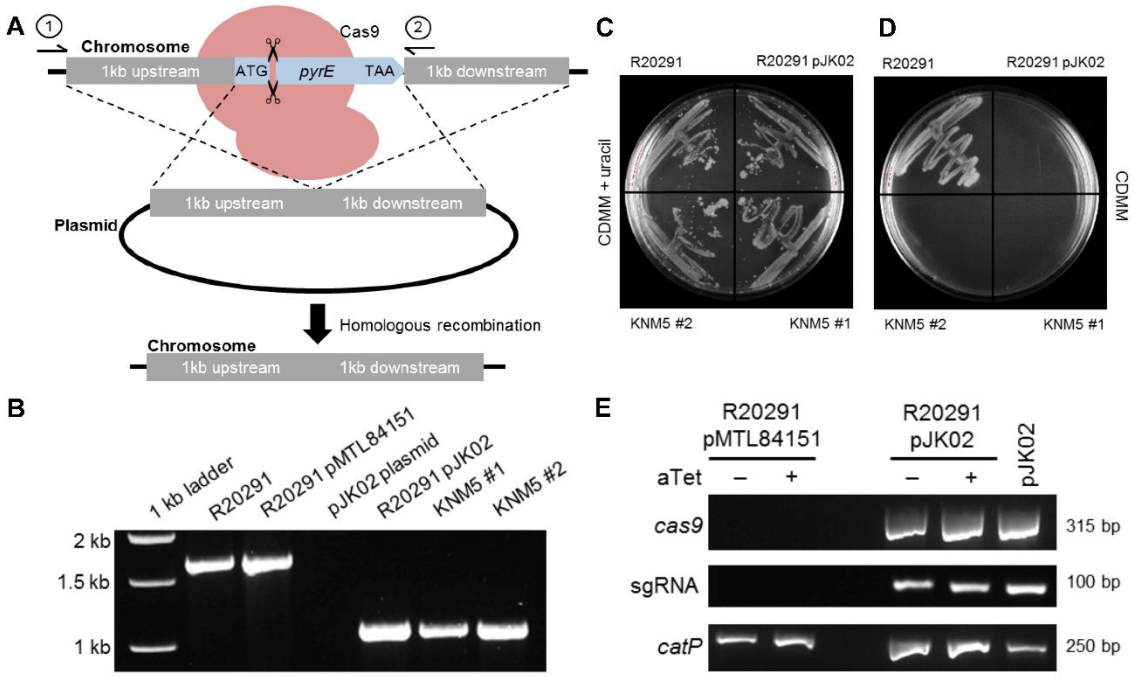


Figure 2.4 Isolating Cas9-mediated *C. difficile pyrE* mutants.
(A) A deletion of the chromosomally-encoded *pyrE* gene was made by homologous recombination from a donor region located on the CRISPR-Cas9 plasmid during repair of a Cas9-mediated double-stranded DNA break. The location of the crRNA target region in *pyrE* is indicated by the cut DNA. Amplification of wildtype *pyrE* using primers 1 and 2 results in a 1.59 kb band on an agarose gel. (B) DNA was isolated from potential mutants. The region surrounding the *pyrE* gene was amplified from the chromosome, and the resulting DNA was separated on an agarose gel. A clean deletion of *pyrE* is indicated by a faster-migrating DNA band. (C-D) *C. difficile* R20291, *C. difficile* R20291 pJK02, *C. difficile* KNM5 isolate 1 and *C. difficile* KNM5 isolate 2 were streaked onto either (C) CDMM supplemented with uracil or (D) CDMM alone, and incubated anaerobically for 4 days. (E) RT-PCR showing a comparison of *C. difficile* R20291 pMTL8151 and *C. difficile* R20291 pJK02 induced without aTet and those induced in the presence of aTet to turn on the expression of *cas9*. Also tested was the sgRNA and *catP*, as a positive control. The white dividing bar between R20291 pMTL84151 and R20291 pJK02 samples indicates an empty lane between samples. The white dividing bar between amplified genes indicates different gels.

Table 2.1 Efficiencies for CRISPR-Cas9-mediated *pyrE* mutations in *C. difficile* R20291.

Strain	Total cell count (CFU/mL)	Mutant cell count (CFU/mL)	Calculated efficiencies	Average
R20291 pMTL84151 (empty vector)	3.5×10^8	1.0×10^2	2.86×10^{-7}	$4.42 \times 10^{-7} \pm 2.46 \times 10^{-7}$
	1.9×10^8	1.5×10^2	7.89×10^{-7}	
	2.0×10^8	5.0×10^1	2.50×10^{-7}	
R20291 pKM93 (*cas9*$_{D10A}$)	1.9×10^7	3.0×10^3	1.58×10^{-4}	$2.04 \times 10^{-4} \pm 5.27 \times 10^{-5}$
	9.0×10^7	2.5×10^4	2.78×10^{-4}	
	8.5×10^7	1.5×10^4	1.76×10^{-4}	
R20291 pJK02 (WT *cas9*)	8.1×10^6	4.0×10^6	0.494	0.495 ± 0.049
	8.1×10^6	4.5×10^6	0.556	
	9.9×10^6	4.3×10^6	0.434	

2.3.5. *No off-target effects in* C. difficile *KNM5*

To understand if the generated *pyrE* deletion mutants had other mutations in the genome, Illumina genome re-sequencing was performed for *C. difficile* KNM5 (Δ*pyrE*) and *C. difficile* R20291 (pMTL84151) (empty vector) after induction for the CRISPR-Cas9 system to account for any mutations that may have occurred due to the induction procedure (though *C. difficile* R20291 (pMTL84151) does not contain a region that could be used for homology directed repair, FOA-resistant colonies were observed on FOA-containing medium). As shown in Figure 2.5, >300 reads were observed for every position across the genome (only the region surrounding *pyrE* is shown). However, at the *pyrE* gene, the reads drop drastically to ~20-30 reads, dropping further to undetectable levels until the end of the stop codon. The reads increase again after the stop codon; there were no other mutations in the KNM5 strains. Interestingly, the *C.*

difficile R20291 (pMTL84151) FOA-resistant strain had a single-nucleotide polymorphism (SNP) in *pyrE* which resulted in a premature stop codon. These results suggest that there were no off-target effects generated by the CRISPR-Cas9 system and that it could be applied to other genes.

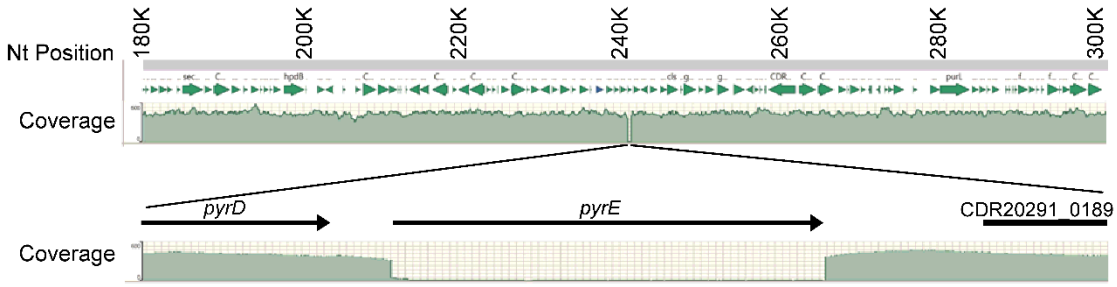


Figure 2.5 Monitoring off-target effects of the CRISPR-Cas9 system. Coverage of the sequencing reads from *C. difficile* KNM5 (Δ*pyrE*) in relation to their position on the *C. difficile* R20291 chromosome. The gap in the coverage is expanded along with annotations for the region.

2.3.6. *Deletion of* selD *using the CRISPR-Cas9 system*

Because the TargeTron mutation results in the insertion of the group II intron into the *selD* gene, there are likely polar effects on the downstream *selA* and *selB* sequences. To generate a deletion in *selD*, we targeted the sgRNA to the *selD* gene and cloned the upstream and downstream regions for use in homology-directed repair. This plasmid was introduced into *C. difficile* R20291, the resulting strain was induced and cells were plated directly on BHIS medium (rich medium). Colonies that grew were tested directly for the desired mutation by PCR. Though the efficiency of mutation was lower compared to *pyrE*, we observed a frequency of ~1 *selD* deletion in every 5 colonies tested (~20% mutation frequency) (Table 2.2).

Table 2.2 Efficiencies for CRISPR-Cas9-mediated *selD* deletion in *C. difficile* R20291.

Strain	Total colonies tested	PCR positive mutants	Calculated efficiencies	Average
R20291 pJS194 (*selD* target)	16	2	0.125	0.194 ± 0.161
	24	1	0.042	
	24	10	0.417	

We then tested the growth phenotype of *C. difficile* KNM6 (Δ*selD*) and compared it to the growth observed for the wild-type R20291 strain. In the same experiment, we tested whether there were no polar effects on downstream genes in the operon, *selA* and *selB*, from this clean deletion of *selD*. To do this, we introduced *selD* with its native promoter on a multi-copy plasmid, *C. difficile* KNM6 pKM142. We included empty vector, pJS116, in wild-type R20291 as well as the mutant KNM6 strains. When grown in rich, BHIS medium (Figure 2.6A), no difference is seen between the three strains, R20291 pJS116, KNM6 pJS116, and KNM6 pKM142. When grown in medium that may favor Stickland metabolism (TY or TYG medium), the wild-type strain with empty vector, R20291 pJS116, grew well while the mutant strain with empty vector, KNM6 pJS116, had a reduction in growth in comparison (Figure 2.6B). When this strain was complemented by expressing *selD*, not the entire locus, KNM6 pKM142 grew to wild-type levels (Figure 2.6B). This confirms there were no polar effects on the downstream genes, *selA* and *selB*, due to the deletion of *selD*.

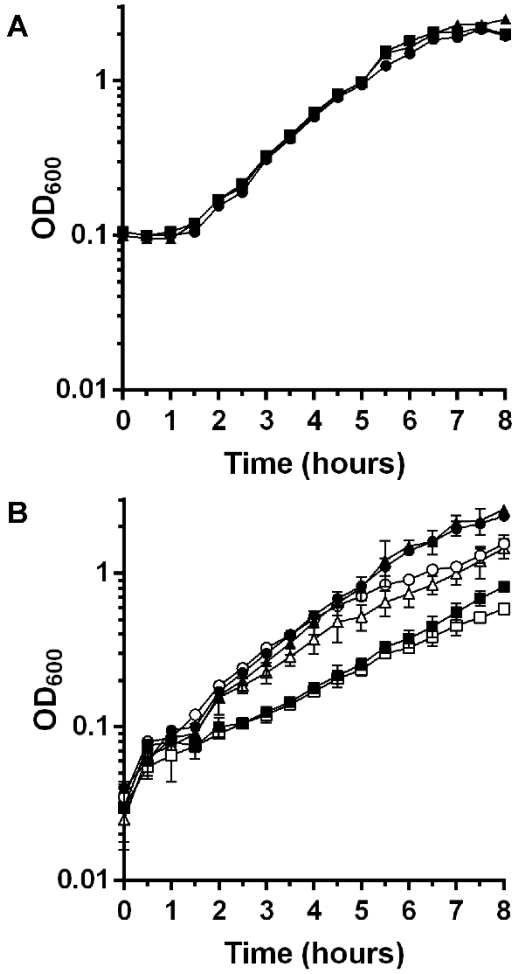


Figure 2.6 *C. difficile* Δ*selD* has a moderate growth defect compared to the WT strain.

(A) *C. difficile* R20291 pJS116 (empty vector) (●), *C. difficile* KNM6 (Δ*selD*) pJS116 (empty vector) (■) and *C. difficile* KNM6 pKM142 (*selD* complement) (▲) were grown in BHIS medium and growth was monitored over time. (B) *C. difficile* R20291 pJS116, KNM6 pJS116 and KNM6 pKM142 were grown in TYG (closed shapes, ●) or TY (open shapes, ○) medium and growth was monitored over time. Data points represent the average from two independent experiments and error bars represent the standard deviation of the mean.

Taken together, our results suggest that selenium incorporation into proteins is important for *C. difficile* growth and that CRISPR-Cas9 gene editing can be used to rapidly and efficiently introduce mutations with no polar effects into the *C. difficile* genome.

2.4. Discussion

CRISPRs were originally discovered in *Escherichia coli* and later in archaea and other bacteria, including *C. difficile* (84, 152, 185). The development of utilizing CRISPRs and Cas proteins has led to functional gene editing tools that are widely used. For gene editing in bacteria, the components necessary for this system include: a Cas protein, a guide RNA, and a region of donor DNA to make the desired mutation (83, 101). A short sequence proximal to the target sequence, which helps the CRISPR-*cas9* system to distinguish between self and non-self-sequences, is called the protospacer adjacent motif (PAM) sequence (90). *S. pyogenes* Cas9 recognizes the 5'–NGG–3' PAM sequence (90, 91).

C. difficile encodes a native CRISPR-Cas system and belongs to the class I-B subtype (152). The *S. pyogenes* Cas9, used in this study, belongs to the class II group. The classes are defined by their mechanisms and also the composition of Cas proteins (83); thus, *C. difficile* has different Cas proteins than that of *S. pyogenes*. The CRISPR-Cas system of *C. difficile* is predicted to recognize a PAM sequence of 5'–CCW–3', where "W" indicates either an adenine (A) or thymine (T) (152). Due to the differences in PAM recognition sequences of these two CRISPR-Cas systems, we do not predict that the *C. difficile* CRISPR-Cas locus will interfere with this genetic tool.

We successfully developed the first application of a CRISPR-Cas9 system for genetic modification in *C. difficile*. Due to the problems inherent to each genetic system used in *C. difficile*, we wanted to create a plasmid containing the system which was simple and easy to modify for future use by others in the field. Towards this goal, the fragments to generate the homology region can be cloned in one step using Gibson Assembly; a gBlock containing the entirety of the sgRNA is also cloned into the plasmid using Gibson Assembly, each at unique restriction sites in the plasmid. Thus, a new mutagenesis plasmid can be generated in two consecutive cloning steps.

We used the *pyrE* gene as a starting point to optimize the system. By doing so, we were able to determine that *C. difficile* has a very poor ssDNA break repair system, as evident by the low efficiency of the $Cas9_{D10A}$-mediated *pyrE* deletion. However, $Cas9_{D10A}$ had a greater frequency of FOA-resistant colonies than did *C. difficile* R20291 pMTL84151 (empty vector), suggesting that *C. difficile* can use homology-directed repair to correct ssDNA breaks, but not at an efficiency which could allow for the isolation of mutants without selective pressure. Thus, the $cas9_{D10A}$ allele is not a viable option for this CRISPR-based genetic system.

Moving forward with the wild-type Cas9, a concern was whether the mutations were due simply to homologous recombination between the chromosome and the donor DNA on the plasmid or required the repair of the CRISPR-Cas9-mediated dsDNA break. The efficiency of FOA-resistant cells (*pyrE* mutants) was greater for strains containing wild-type Cas9 than for strains that contained the $cas9_{D10A}$ allele. Thus, without the aid of the wild-type Cas9 nuclease, *C. difficile* cannot introduce the desired mutation by

homologous recombination with a high enough efficiency to allow for isolation of a mutant without a selection.

Selenophosphate synthetase is an enzyme that uses ATP, water, inorganic phosphate and hydrogen selenide to generate selenophosphate (167, 186). Selenophosphate is used as a donor to generate selenocysteine-charged tRNAs by attaching selenide to serine-charged tRNAs, leading to incorporation of selenocysteine into selenoproteins (*e.g.*, PrdB or GrdA) (167, 187). Previously, a *prdB* mutant was shown to have a decreased growth rate compared to the parent strain (55). We had hypothesized that, because SelD is required for generating the precursor to selenocysteine-charged tRNAs, the CRISPR *selD* mutant would have a greater effect on growth rate than a *prdB* mutation due to the global reduction in selenoproteins. Indeed both the TargeTron mutant and the CRISPR-generated *selD* deletion had reduced growth in protein-rich medium (medium where Stickland metabolism is important for growth). We also show there were no polar effects on the downstream *selA* and *selB* genes from the clean deletion of *selD*. In the future, this mutant will help in studying the global effect of the disruption of selenium incorporation into proteins in *C. difficile*, not just growth and metabolism.

The mutation efficiencies for *pyrE* and *selD* were similar and both were well within testable limits. The crRNA chosen for *pyrE* began at the 36th nucleotide of the 585-bp *pyrE* gene and had a score in CRISPRscan.org of 30, which is low compared to the highest score of 62 for a crRNA for this gene. The crRNA chosen for *selD* had a score in CRISPRscan.org of 67, the highest listed, and started at the 183rd nucleotide of

the 951-bp *selD* gene. From these values for the respective genes, there appears to be no pattern for how efficient the CRISPR-Cas9 system in *C. difficile* can make the mutation. The rules for choosing the optimal crRNA that will yield the highest efficiency for generating a mutation in *C. difficile* is still under investigation.

In summary, we have developed a functional CRISPR-Cas9 system for use in *C. difficile*. Because other systems rely on the integration of segregationally-unstable plasmids into the genome, an event that can take several passages and the eventual regeneration of a chromosomal deletion in *pyrE*, this CRISPR-based plasmid has the potential to rapidly generate mutations within the *C. difficile* genome. With future adjustments to this system, a larger range of mutations, insertions and even point mutations, could possibly be made in *C. difficile* which has been difficult or even impossible in the past.

3. THE SELENOPHOSPHATE SYNTHETASE, *SELD*, IS IMPORTANT FOR *CLOSTRIDIOIDES DIFFICILE* PHYSIOLOGY

In Section 3, we analyze the importance of selenium-containing proteins on *C. difficile* physiology. We found that multiple aspects of *C. difficile* physiology were affected (*i.e.*, growth, sporulation and outgrowth of a vegetative cell post-spore germination). Using RNAseq, we compared the global gene expression of wildtype *C. difficile* cells to the *selD* mutant and identified multiple candidate genes which likely aid the cell in overcoming the global loss of selenoproteins.

3.1. Introduction

Clostridioides (*Clostridium*) *difficile* is a major concern as a nosocomial gut pathogen (7). This pathogen has become the most common cause of health care-associated infections in United States hospitals (3, 4). In 2011 in the United States, approximately 500,000 cases of *C. difficile* infections were identified and nearly 29,000 of those resulted in death (19). In addition to the burden this pathogen places on the patient, this pathogen also places a $4.8 billion burden to the United States health care system in treatment associated health-care costs (188). Due to the stresses this pathogen places on the patient and health care system, and it being a multi-drug resistant organism, the Centers for Disease Control and Prevention labeled bacterium as an urgent threat (20).

Key to battling this aggressive pathogen is understanding the basic processes *C. difficile* uses to complete its life cycle. While our understanding of *C. difficile* physiology has increased dramatically in the last decade, mostly in toxin production / regulation, sporulation and germination, our understanding of metabolic processes is lacking (22, 23, 34, 36, 162, 165, 189). A recent review by Neumann-Schaal *et al*. discussed the known metabolic processes involved in energy generation in *C. difficile* (44). *C. difficile* has multiple metabolic pathways that overlap to ensure generation of key metabolites. Carbon metabolism includes breaking down sugars such as glucose and mannitol to generate pyruvate and acetyl-CoA for glycolysis and the tricarboxylic acid cycle, although the latter is incomplete. Pyruvate is a key metabolite in several carbon metabolism and fermentation pathways (44). For example, it can be used by pyruvate formate-lyase to generate the CO_2 that is used by the Wood-Ljungdahl pathway to fix CO_2 (45, 190), it can be degraded to acetyl-CoA to generate butyrate, and it is also used in fermentation pathways to generate propionate via the reductive branch of Stickland metabolism (44). While many of the metabolic processes in *C. difficile* have been elucidated, how these processes interact when others are impaired has yet to be studied (44). This is especially important if these metabolic pathways are to be targets for antimicrobial development.

Stickland metabolism is a primary source of energy for a small group of anaerobic bacteria that use amino acids as their sole carbon and nitrogen source (*e.g.*, *C. difficile*, *Clostridium sporogenes*, and *Clostridium sticklandii*). The main function of this metabolic pathway is to generate NAD^+ and ATP for the cell (51-54). In the oxidative

branch, an amino acid, most frequently isoleucine, leucine or valine (ILV), is decarboxylated or deaminated and NAD^+ is converted to NADH in this process (52, 53, 55). In the reductive branch, D-proline or glycine are reduced by proline reductase (PrdB) and glycine reductase (GrdA), respectively, to regenerate NAD^+ (53, 54, 56, 57).

Both proline and glycine reductases are selenoproteins (55, 58). Selenoproteins are made through the incorporation of selenium, as selenocysteine, during protein synthesis. Selenocysteine is made through a synthesis pathway where inorganic phosphate is reacted with hydrogen selenide by the selenophosphate synthetase, SelD, to generate selenophosphate. Through the use of a selenocysteinyl-tRNA (Sec) synthase, SelA, selenophosphate is incorporated into serine-charged tRNAs to generate selenocysteine. The selenocysteine-specific elongation factor, SelB, recognizes an in-frame stop codon followed by a selenocysteine insertion sequence (SECIS), and results in a halt in translation to allow for the incorporation of the selenocysteine into the protein (59).

Recently, Stickland metabolism was hypothesized to be a significant contributor to *C. difficile* growth during *C. difficile* infection in a murine model. Proline and hydroxyproline were found to be the among the highest abundant molecules at the start of infection, and 5-aminovalerate, a product of the proline reductase in Stickland metabolism, is an abundant molecule towards the end of infection (49, 50). These results suggest Stickland metabolism is used by *C. difficile*, or by other organisms that are present during *C. difficile* infection, *in vivo*.

We hypothesized that if we were to eliminate the global production of selenoproteins, the two Stickland reductases, PrdB and GrdA, would not be active and resulting strain would be incapable of Stickland metabolism. Previously, we generated a CRISPR-Cas9 genome editing tool for use in *C. difficile*. In that work, we created a *C. difficile* Δ*selD* strain and analyzed the growth phenotype of this strain compared to the wild-type parent and the mutant complemented with a wildtype *selD* allele expressed from a multicopy plasmid. We showed that *C. difficile* R20291 Δ*selD* (KNM6) had no growth defect in rich BHIS medium but had a growth defect in a peptide rich medium where *C. difficile* is expected to rely on Stickland metabolism for growth (62). Here, we build upon this work by using the CRISPR-Cas9 system to 'knock in' the *selD* gene to its native locus. Using this newly *selD*-restored strain, we sought to further characterize the importance of selenoproteins on *C. difficile* physiology. In this study we found that SelD has a significant role in growth of vegetative cells, spore formation, and outgrowth of a vegetative cell post-germination. Our findings suggest that selenoprotein synthesis could be an attractive target for non-antibiotic based therapeutic development.

3.2. Materials and methods

3.2.1. *Bacterial strains and growth conditions*

C. difficile strains were routinely grown in an anaerobic atmosphere (1.7% - 4% H_2, 5% CO_2, 85% N_2) at 37 °C in brain heart infusion supplemented with 5 g / L yeast extract and 0.1% L-cysteine (BHIS), as described previously (15, 34, 168, 169) or TY medium (3% tryptone, 2% yeast extract) (170). Anaerobic gas mix introduced into the chamber was 10% H_2, 5% CO_2 and 85% N_2. Atmospheric hydrogen levels were

determined using a COY Anaerobic Monitor (CAM-12). For conjugation experiments, cells were plated onto TY medium for *Bacillus subtilis*-based conjugations. Where indicated, growth was supplemented with taurocholate (TA; 0.1% w / v), thiamphenicol (10 µg / mL), D-cycloserine (250 µg / mL), xylose (1% w / v) and / or glucose (1% w / v) as needed. Induction of the CRISPR-Cas9 system was performed on TY agar plates supplemented with thiamphenicol (10 µg / mL) and xylose (1% w / v). *E. coli* strains were routinely grown at 37 °C in LB medium. Strains were supplemented with chloramphenicol (20 µg / mL) as needed. *B. subtilis* BS49 was routinely grown at 37 °C in LB broth or on LB agar plates. Strains were supplemented with chloramphenicol (2.5 µg / mL) and / or tetracycline (5 µg / mL). All strains used are listed in Table 5.4.

3.2.2. *Plasmid construction and molecular biology*

To construct the CRISPR-Cas9 *selD* restoration plasmid, the previously published CRISPR-Cas9 *pyrE* targeting plasmid, pJK02 (62) (all plasmids are listed in Table 5.4), was modified by replacing *traJ* with oriT *tn*916 for *B. subtilis* conjugation by amplification from pJS116 using primers 5'Tn916ori and 3'Tn916ori (all oligonucleotides used in this section are listed in Table 5.5). The resulting fragments was introduced into pJK02 by Gibson assembly at the *Apa*I site and transformed into *E. coli* DH5α to generate pKM126. Next, the donor region to be used for homology directed repair was PCR amplified from *C. difficile* R20291 genomic DNA using primers 5'selD_comp and 3'selD_comp 2 where this fragment contains a 500 bp upstream homology arm and a 500 bp downstream homology arm surrounding the *selD* gene. The resulting fragment was cloned by Gibson assembly into pKM126 at the *Not*I and *Xho*I

restriction sites and transformed into *E. coli* DH5α to generate pKM181. Lastly, the gBlock (Integrated DNA Technologies, Coralville, IA) for *selD* targeting sgRNA, CRISPR_selD_comp2, was introduced by Gibson assembly into the *Kpn*I and *Mlu*I restriction sites and transformed into *E. coli* DH5α resulting in pKM183. To improve efficiency of the CRISPR-Cas9 editing system and provide more control of Cas9 expression, the tetracycline-inducible system was replaced with the xylose inducible system (124). To do this, the xylose inducible promoter was PCR amplified from pIA33 using primers 5'selDcomp_HR_xylR 2 and 3'cas9_Pxyl 2 and inserted by Gibson assembly into pKM183 at the *Xho*I and *Pac*I restriction sites and transformed into *E. coli* DH5α to generate pKM194.

The xylose inducible promoter was PCR amplified from pIA33 using primers 5'pyrE_HR_xylR 2 and 3'cas9_Pxyl 2 and inserted by Gibson assembly into pKM126 at the *Xho*I and *Pac*I restriction sites and transformed into *E. coli* DH5α to generate pKM197. To construct the CRISPR-Cas9 *spo0A* deletion mutant, the gBlock for *spo0A* targeting sgRNA, CRISPR_spo0A_2, was introduced by Gibson assembly into the *Kpn*I and *Mlu*I restriction sites and transformed into *E. coli* DH5α resulting in pKM213. The homology arms to be used for homology directed repair was PCR amplified from *C. difficile* R20291 genomic DNA using primers 5'spo0A_UP and 3'spo0A_UP for the 500 bp upstream arm and primers 5'spo0A_DN and 3'spo0A_DN for the 500 bp downstream arm. The resulting fragments were then cloned by Gibson assembly at the *Not*I and *Xho*I restriction sites and transformed into *E. coli* DH5α resulting in pKM215.

3.2.3. *Conjugation for CRISPR-Cas9 plasmid insertion*

CRISPR-Cas9 plasmid pKM194 was transformed into *B. subtilis* BS49 and used as a conjugal donor with *C. difficile* KNM6. Likewise, the pKM215 plasmid was transformed into *B. subtilis* BS49 to be used as a donor for conjugation with *C. difficile* R20291. *C. difficile* R20291 or KNM6 was grown anaerobically in BHIS broth overnight. These cultures were then diluted in fresh pre-reduced BHIS broth and grown anaerobically for 4 hours. Meanwhile, *B. subtilis* BS49 containing the CRISPR plasmids were grown aerobically at 37 °C in LB broth supplemented with tetracycline and chloramphenicol for 3 hours. One hundred microliters of each culture was pated on TY agar medium. After 24 hours, the growth was harvested by suspending in 2 mL pre-reduced BHIS broth. A loopful of this suspended growth was spread onto several BHIS agar plates supplemented with thiamphenicol, kanamycin, and D-cycloserine. *C. difficile* transconjugants were screened for the presence of Tn916 using tetracycline resistance, as described previously. Thiamphenicol-resistant, tetracycline-sensitive transconjugants were selected and used for further experiments.

3.2.4. *Induction of the CRISPR-Cas9 system and isolating mutants*

C. difficile KNM6 pKM194 was streaked onto TY agar medium supplemented with thiamphenicol and xylose for *cas9* induction. Growth was then passaged a second time on the same medium. After restreaking colonies onto BHIS medium supplemented with xylose to isolate colonies, DNA was extracted and tested for insertion of the wildtype *selD* allele by PCR amplification of the *selD* region using primers 5’selD and

3’selD. From this, one mixed colony out of 12 samples was isolated. This mixed colony was then passaged on TY agar medium supplemented with thiamphenicol and xylose. Colonies were isolated on BHIS agar medium supplemented with xylose, DNA was extracted, and isolates were tested by PCR amplification again as above. From this, 14 of 15 colonies were insertions. Confirmed restored strains were passaged 3 times in BHIS liquid medium in order to cure the CRISPR-Cas9 plasmid. Loss of plasmid was confirmed by PCR amplification of a portion of *cas9* using primer set 5’tetR_CO_Cas9 and 3’COcas9 (975) and the gRNA using primer set 5’gdh and 3’gRNA 2.

The *C. difficile* R20291 strain containing the *spo0A*-targeting plasmid, pKM215, was streaked onto TY agar medium supplemented with thiamphenicol and xylose for induction. This was then passaged a second time on the same medium. After restreaking colonies onto BHIS supplemented with xylose to isolate colonies, DNA was extracted and tested for the deletion by PCR amplification of the *spo0A* region using primers 5’spo0A_del and 3’spo0A_del. All tested colonies (36 total) were mutants. The plasmid was cured as described above and the loss of plasmid was confirmed by PCR amplification of a portion of *cas9* using primer set 5’tetR_CO_Cas9 and 3’COcas9 (975) and the gRNA using primer set 5’gdh and 3’gRNA 2.

3.2.5. *Sporulation and heat resistance assay*

To determine differences in sporulation efficiencies between the *C. difficile* R20291 (wild-type), KNM6 (Δ*selD*), KNM9 ((Δ*selD*::*selD*$^{+}$) and KNM10 (Δ*spo0A*) strains, sporulation and heat resistance was determined as described previously (191). Briefly, *C. difficile* R20291, KNM6, KNM9, and KNM10 strains were spread onto BHIS

agar medium supplemented with taurocholate. From this, colonies were restreaked onto either BHIS or TY agar media, making a lawn on the plate. After 48 hours of growth, half of the plate was harvested and suspended in 600 µL of pre-reduced PBS. Then, 300 µL of the sample was transferred to a separate tube and incubated in a heat block for 40 minutes at 65 °C. During incubation, the tube was inverted every 10 minutes to ensure even heating. Both the untreated and the heat-treated samples were serially diluted in PBS and plated onto BHIS agar medium supplemented with taurocholate. CFUs were counted 24 hours after plating. Heat resistance was calculated by dividing the CFUs for the heat-treated sample by the CFUs for the untreated sample for each replicate. Efficiencies for each replicate were calculated by comparison of the heat resistance of the experiment strain compared to the reference strain (wild-type, R20291), and the average and standard deviation was calculated for each strain. This assay does not account for the number of vegetative cells that are viable on the agar plate and the differences between strains/replicates.

3.2.6. *Spore purification*

Spores were purified from *C. difficile* R20291, KNM6, and KNM9 strains as previously described (41, 168, 192, 193). Briefly, spores were streaked onto BHIS agar medium (20 – 30 plates) and allowed to sporulate for 5 to 6 days before scraping each into microcentrifuge tubes containing 1 mL of sterile dH_2O and kept at 4 °C overnight. The spores and debris mixture was washed five times in sterile dH_2O by centrifuging for 1 minute at 14,000 x g per wash. Spores were combined into 2 mL aliquots in sterile dH_2O and layered on top of 8 mL of 50% sucrose and centrifuged at 4,000 x g for 20

minutes. The supernatant containing vegetative cells and cell debris was discarded and the spore pellet was resuspended in sterile dH_2O and washed five more times as before. The spores were stored at 4 °C until use.

3.2.7. *Germination assay*

Purified *C. difficile* R20291, KNM6, and KNM9 spores were first heat activated at 65 °C for 30 minutes and then placed on ice until use. To compare germination between the three strains, the OD_{600} was measured over time in different media. Germination was carried out in clear Falcon 96-well plates at 37 °C in a final volume of 100 µL and final concentrations of 10 mM taurocholate and 1X BHIS or TY. Spores were added to a final OD_{600} of 0.5 and germination was analyzed for 1 hour using a plate reader (Spectramax M3 Plate Reader, Molecular Devices, Sunnyvale, CA).

3.2.8. *Outgrowth assays*

Purified *C. difficile* R20291, KNM6, and KNM9 spores were heat activated at 65 °C for 30 minutes and then placed on ice until use. Spore samples were washed one time with dH_2O and the spores were added to a final optical density (OD_{600}) of 0.5 in 8 mL of BHIS medium supplemented with taurocholate (10 mM final concentration). The spores were incubated in the germination solution for 10 minutes at 37 °C and then immediately placed on ice. In all subsequent steps, the germinated spores were kept on ice or at 4 °C. The germinated spores were washed once with BHIS or TY medium, to remove germinant, and the supernatant was removed. The germinated spores were passed into the anaerobic chamber without cycling the airlock (the airlock was flooded with gas to reduce the amount of oxygen that would be transferred into the chamber) and

resuspended each in 20 mL of pre-reduced BHIS or TY medium. Outgrowth was monitored by measuring OD_{600} over time.

3.2.9. *RNA processing*

C. difficile R20291 (wild-type) and KNM6 (Δ*selD*) were grown to an OD_{600} of 0.6 in TYG medium at low (1.7%) and high (4%) hydrogen levels. At that time, RNA was extracted as described using the FastRNA Blue Kit (MP Biologicals). DNA was depleted using the TURBO DNA-free kit (Invitrogen), repeating the steps in the kit three times to achieve depletion. rRNA was then depleted using the Ribo-Zero rRNA Removal kit for bacteria (Illumina). Enrichment of mRNA and generation of cDNA libraries were completed using the TruSeq Stranded mRNA library prep kit (Illumina).

3.2.10. *RNA-seq*

cDNA libraries were submitted for Illumina high output single-end 50 sequencing at the Tufts University Genomics Core. Reads were assembled using the DNASTAR SeqMan NGen 15 program using a combined assembly noting replicates. Raw expression data between wild-type and mutant strains was then normalized to the *rpoB* gene transcript and then quantified Using DNASTAR ArrayStar 15. Assemblies were normalized by assigning Reads Per Kilobase of template per Million mapped reads (RPKM), experimental values were capped at a minimum of 1, and all genes were normalized by calibration to the median expression value of *rpoB*. Fold-change was determined four ways: wild-type at 4% hydrogen to wild-type at 1.7% hydrogen, wild-type at 1.7% to wild-type at 4%, wild-type at 4% to mutant at 4%, and wild-type at 1.7% to mutant at 1.7% (Table 5.3). To rule out the effects of hydrogen abundance on gene

expression, expression values with ≥ 2-fold change in response to the amount of hydrogen were excluded when comparing wild-type to mutant comparisons. From this, any gene expression that was ≥ 2-fold up- or down-regulated was considered significant. As a final step, we excluded any genes which had little coverage across the gene or genes with very low reads, visualized using DNASTAR GenVision.

3.2.11. *Quantitative RT-PCR*

C. difficile R20291 (wild-type) and KNM6 (Δ*selD*) were grown, RNA was extracted, and DNA was depleted as described above. 50 ng of total RNA was used in cDNA synthesis using the SuperScript III First-Strand Synthesis System (Thermo Scientific) according to the protocol, including controls for each sample without the presence of reverse transcriptase. Primers for quantitative reverse transcription-PCR (RT-qPCR) were designed using the Primer Express 3.0 software (Applied Biosystems) and efficiencies were validated prior to use. cDNA samples were then used as templates for qPCR reactions to amplify *rpoB*, *tcdB*, *CDR20291_0963*, *CDR20291_0962*, *prdB*, *grdB*, *grdA*, and *mtlF* (primers are listed in Table 5.5) using PowerUp SYBR Green Master Mix (Applied Biosystems) and a QuantStudio 6 Flex Real-Time PCR machine (Applied Biosystems). Reactions were performed in a final volume of 10 µL including 1 µL undiluted cDNA sample and 500 nM of each primer. For both wild-type and mutant samples, at each hydrogen level, reactions were run in technical triplicate for each biological triplicate. Outliers of technical replicate samples were omitted from analysis. Results were calculated using the comparative cycle threshold method (194), in which the amount of target mRNA was normalized to that of an internal control (*rpoB*).

3.2.12. *Virulence studies*

All animal studies were performed with prior approval from the Texas A&M University Institutional Animal Care and Use Committee. Female Syrian golden hamsters (60 - 100 g) were housed and tested for *C. difficile* susceptibility as previously described (34, 195). To induce susceptibility to *C. difficile* infection, hamsters were gavaged with 30 mg / kg clindamycin 5 days prior to inoculating with 5,000 *C. difficile* spores of the indicated strain. 5 hamsters were given vehicle only (dH_2O) as a control. Hamsters showing disease symptoms (wet tail, poor fur coat and lethargy) were euthanized by CO_2 asphyxia followed by thoracotomy as a secondary means of death.

3.2.13. *Ethics statement*

All animal procedures were performed with prior approval from the Texas A&M Institutional Animal Care and Use Committee under the approved Animal Use Protocol number 2017–0102. Animals showing signs of disease were euthanized by CO_2 asphyxia followed by thoracotomy as a secondary means of death, in accordance with Panel on Euthanasia of the American Veterinary Medical Association. Texas A&M University's approval of Animal Use Protocols is based upon the United States Government's Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research and Training and complies with all applicable portions of the Animal Welfare Act, the Public Health Service Policy for the Humane Care and Use of Laboratory Animals, and all other federal, state, and local laws which impact the care and use of animals.

3.2.14. *Statistical analysis*

C. difficile R20291 pJS116 (wild-type, empty vector), *C. difficile* KNM6 pJS116 (Δ*selD*, empty vector), and *C. difficile* KNM9 (Δ*selD*::*selD*$^{+}$, p*selD*) strains were grown in biological duplicate. *C. difficile* R20291, KNM6, and KNM9 strains were grown in biological triplicate. *C. difficile* R20291, KNM6, KNM9, and KNM10 strains were allowed to sporulate in biological triplicate. *C. difficile* R20291, KNM6, and KNM9 spores were purified in biological triplicate where each replicate was grown / sporulated and purified separately to be used for germination and outgrowth assays. RNA was extracted from *C. difficile* R20291 and KNM6 strains for RNA-seq in biological duplicate for each strain. RNA was extracted from *C. difficile* R20291 and KNM6 strains for RT-qPCR in biological triplicate and each was run in technical triplicate as well. In each experiment, data represents averages of each of the indicated replicates and error bars represent the standard deviation from the mean. Statistical significance for heat resistance efficiency was determined using a Two way ANOVA with Tukey's multiple comparisons test (* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$).

3.3. Results

3.3.1. *Complementation from a plasmid results in growth differences at different hydrogen levels*

When originally characterizing the *C. difficile* KNM6 (Δ*selD*) mutant strain, we found that the complementing plasmid would not restore the mutant phenotype back to wildtype growth, when grown in either TY or TYG medium (Figure 3.1A). However, as published in the prior work, we were able to complement the phenotype and this

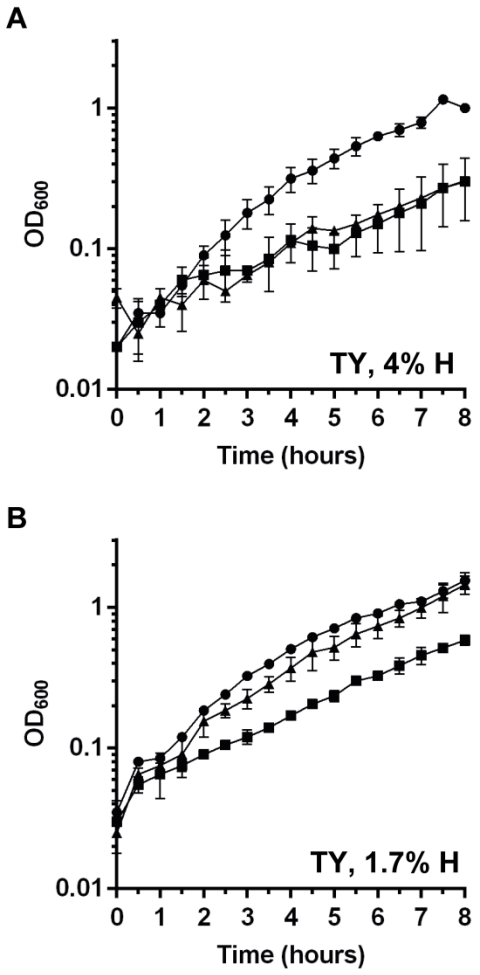


Figure 3.1 Growth curves with complementing plasmid.
C. difficile R20291 pJS116 (wild-type, empty vector) (●), *C. difficile* KNM6 pJS116 (Δ*selD*, empty vector) (■), and *C. difficile* KNM9 (Δ*selD*::*selD*$^{+}$, p*selD*) (▲) were grown in TY medium at (A) 4% hydrogen and (B) 1.7% hydrogen and growth was monitored over an 8 hour period. Data points represent the average from two independent experiments and error bars represent the standard deviation from the mean.

complementation was dependent on the abundance of hydrogen gas in the anaerobic chamber, monitored using a COY Anaerobic Monitor (CAM-12). A 4% hydrogen level did not permit the complementing plasmid to restore the growth phenotype to wildtype levels (Figure 3.1A). However, when the hydrogen amount was lowered to ~1.7% we observed complementation in TY medium (Figure 3.1B). This observation was

interesting and suggests that hydrogen abundance in the anaerobic chamber influences *C. difficile* physiology or the ability of *selD*, when expressed from a multicopy plasmid, to function within *C. difficile*. However, to fully characterize the *C. difficile* Δ*selD* mutant, we wanted to avoid this problem altogether by restoring the CRISPR-Cas9-mediated deletion with *selD*+ at its native locus.

3.3.2. *Restoration of the* selD *mutation with 'knock in'* selD *allele at its native locus using CRISPR-Cas9 genome editing*

The prior iteration of the CRISPR-Cas9 genome editing tool was limited to generating deletion mutations. By modifying the existing CRISPR-Cas9 plasmid, we have improved the functionality of this system to include making insertions into the genome. Recently, Muh *et al*. developed a *C. difficile* CRISPRi tool which encoded *dcas9* controlled by the conditional expression of the xylose promoter (124). We replaced our previous tetracycline-inducible promoter system, which was shown to have uncontrolled expression of Cas9, with the xylose-inducible promoter. We replaced the homology region and gRNA target sequences in our new CRISPR-Cas9 gene editing plasmid to target 26 bp away from the deletion site so as to reintroduce *selD* at its native locus (Figure 3.2A). *C. difficile* Δ*selD* pKM194 was passaged on xylose-containing agar medium. Using this strategy, we generated a *C. difficile selD*+ restored strain (KNM9; Δ*selD*::*selD*$^{+}$) (Figure 3.2B).

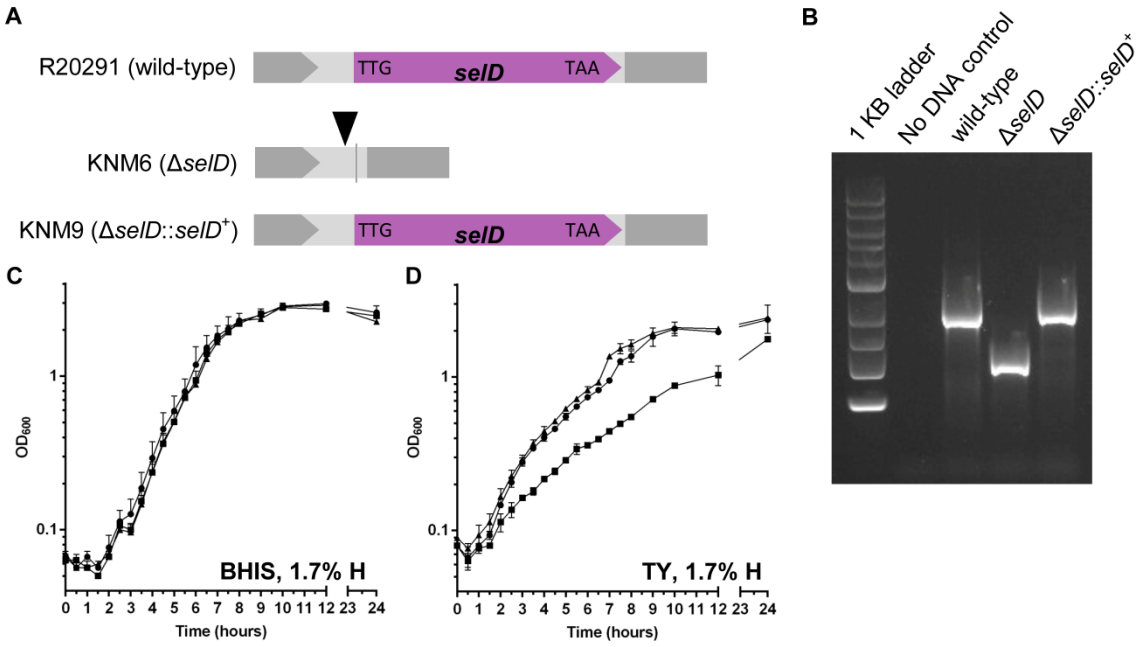


Figure 3.2 CRISPR-Cas9-based restoration of the wildtype *selD* allele to the *C. difficile selD* mutant restores the growth defect.

(A) Graphical representations of the three strains used in this study. Shown as a triangle is the target site of the gRNA and the line indicates the site of the deletion. (B) DNA was isolated from *C. difficile* R20291 (wild-type), KNM6 (Δ*selD*), and KNM9 ((Δ*selD*::*selD*$^+$). The region surrounding the *selD* gene was amplified from the chromosome, and the resulting DNA was separated on an agarose gel. A clean deletion of *selD* is indicated by a faster-migrating DNA band while wild-type and the insertion mutation (restoration) is indicated by a slower-migrating DNA band. (C-D) *C. difficile* R20291 (wild-type) (●), *C. difficile* KNM6 (Δ*selD*) (■), and *C. difficile* KNM9 (Δ*selD*::*selD*$^+$) (▲) were grown in (C) BHIS medium and (D) TY medium at 1.7% hydrogen and growth was monitored over a 24 hour period. Data points represent the average from three independent experiments and error bars represent the standard deviation from the mean.

3.3.3. C. difficile *Δ*selD *strain has a growth defect in peptide rich medium*

We repeated the growth experiments to test whether the restored strain had the wild-type phenotype. When the *C. difficile* wild-type (R20291), Δ*selD* (KNM6), and Δ*selD*::*selD*$^+$ (KNM9) strains were grown in rich BHIS medium, we again saw no difference between the growth of these three strains over a time period of 24 hours

(Figure 3.2C). When these three strains were grown in TY (Figure 3.2D) or TYG (Figure 5.2), we observed a growth defect of the Δ*selD* strain when compared to that of wild-type or Δ*selD*::*selD*$^{+}$ (Figures 3.2D and 5.2). Again, there was no observable difference whether glucose was supplemented in the medium or not on the growth of these strains. For most of the subsequent experiments, TY medium will be used. These results are consistent with our previous findings along with the additional finding that the Δ*selD* mutant strain grows to similar levels as the wild-type and restored strains after 24 hours of growth. We will note that these and all subsequent experiments were performed at ~1.7% hydrogen and will discuss our rationale for this later in the manuscript.

3.3.4. selD *is important for sporulation*

To understand how the absence of selenoproteins impacted *C. difficile* spore formation, we determined the efficiency of sporulation in *C. difficile* R20291, KNM6 (Δ*selD*) and KNM9 (Δ*selD*::*selD*$^{+}$) strains. As a negative control, we generated a mutation in *C. difficile* R20291 *spo0A* using CRISPR-Cas9 (KNM10 strain) (Figure 5.3). In order to test sporulation, we performed a heat resistance assay since fully formed endospores are heat resistant (191). After performing this assay and calculating heat resistance efficiency, we found that regardless of media type, rich BHIS or peptide-rich TY, there was a significant sporulation defect in the KNM6 (Δ*selD*) strain compared to the wildtype R20291 strain ($p \leq 0.05$ in BHIS, $p \leq 0.01$ in TY) (Figure 3.3). This reduced propensity to form spores in the KNM6 (Δ*selD*) strain was restored to wild-type levels in the KNM9 (Δ*selD*::*selD*$^{+}$) strain ($p \leq 0.05$ in BHIS, $p \leq 0.01$ in TY). Our results

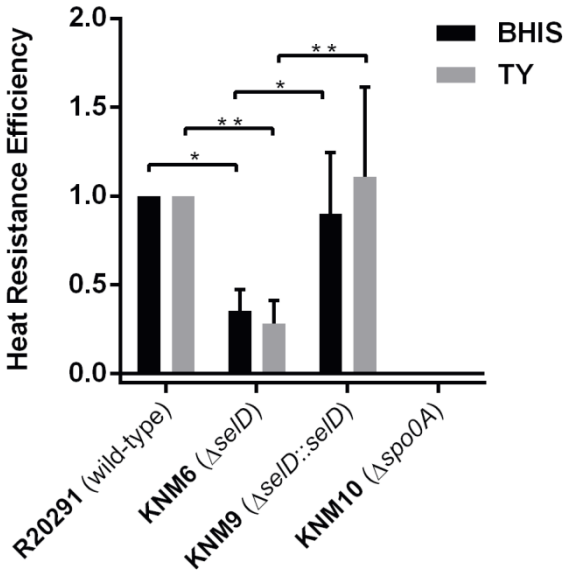


Figure 3.3 *selD* plays a significant role in sporulation of *C. difficile* spores in rich and peptide-based media.

C. difficile R20291 (wild-type), *C. difficile* KNM6 (Δ*selD*), *C. difficile* KNM9 (Δ*selD*::*selD*$^{+}$), and *C. difficile* KNM10 (Δ*spo0A*) strains were allowed to sporulate for 48 hours. Cells were harvested and separated into heat-treated and untreated aliquots. The heat-treated samples were heated at 65 °C for 40 minutes. Both heat-treated and untreated samples were serially diluted in PBS and plated onto BHIS-TA. CFUs were enumerated and efficiencies were calculated. The data represents the averages from three biological replicates and error bars represent the standard error of the mean. Statistical significance was determined using a Two way ANOVA with Tukey's multiple comparisons test (* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$), and the *C. difficile* KNM10 (Δ*spo0A*) strain was excluded from statistical analysis.

indicate that *selD* is necessary for maximal spore formation in the C. difficile R20291 strain.

3.3.5. *Selenoprotein synthesis has no effect on* C. difficile *spore germination*

We then wanted to determine if the Δ*selD* strain had any defect in germination since we saw a defect in both growth and sporulation in peptide-rich medium as well as a sporulation defect in rich BHIS medium. *C. difficile* cells were grown and allowed to sporulate on BHIS medium to minimize affects from both growth and sporulation

defects. The spores were purified and analyzed for germination using the optical density assay. When *C. difficile* wild-type (R20291), Δ*selD* (KNM6), and Δ*selD*::*selD*$^{+}$ (KNM9) spores were suspended in rich BHIS medium supplemented with 10 mM taurocholate (TA), rapid germination occurred (Figure 3.4A). Since the mutant phenotype is most apparent in TY medium, we next germinated spores in TY medium supplemented with 10 mM TA and measured the drop in optical density. Similar to what we observed in rich medium, the spores rapidly germinated in TY medium (Figure 3.4B). These results suggest that the selenophosphate synthetase, *selD*, plays no significant role in the early events during spore germination.

3.3.6. *Deletion of* selD *results in spores with a defect during the outgrowth to a vegetative cell*

Because we observed a defect in growth and sporulation of the *C. difficile* KNM6 (Δ*selD*) strain, we hypothesized that the strain would have a deficiency in outgrowth from spores. Purified spores from wildtype *C. difficile* R20291, KNM6 (Δ*selD*), and KNM9 (Δ*selD*::*selD*$^{+}$) were germinated for 10 minutes at 37 °C in rich BHIS medium to ensure that the germination conditions would not impact outgrowth of a vegetative cell from the germinated spore. After washing the germinated spores in either BHIS or TY medium, the germinated spores were resuspended in the anaerobic chamber in pre-reduced BHIS or TY medium. We then monitored the optical densities over a 24 hour period (Figure 3.5). When cells were allowed to outgrow in rich BHIS medium, a slightly longer lag was observed for the *C. difficile* KNM6 (Δ*selD*) strain of

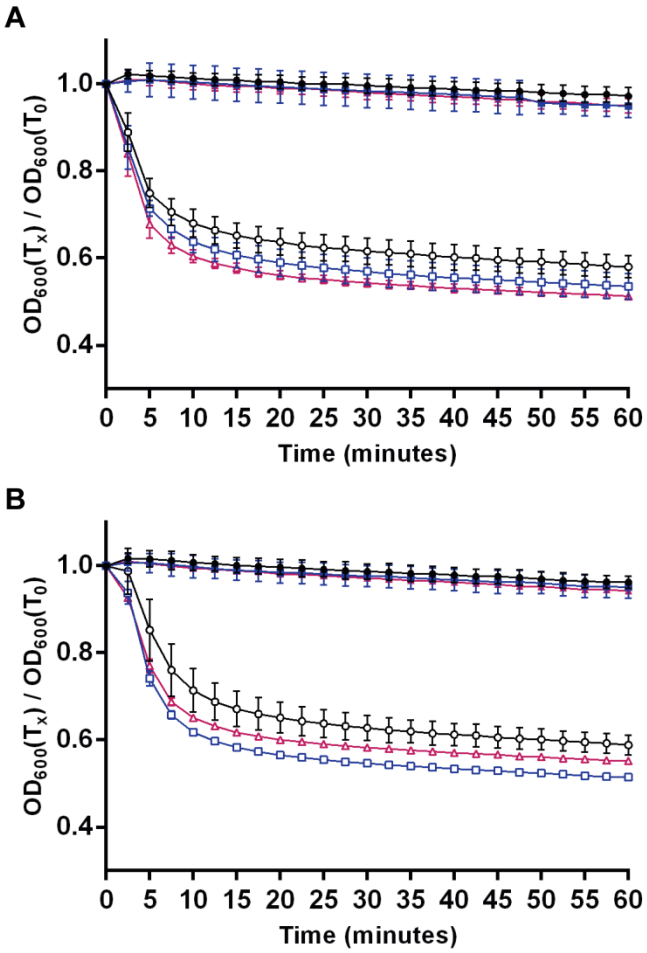


Figure 3.4 Selenoproteins do not have an effect on germination of *C. difficile* R20291 spores.

Purified *C. difficile* R20291 (wild-type) (●/○), *C. difficile* KNM6 (Δ*selD*) (■/□), and *C. difficile* KNM9 (Δ*selD*::*selD*$^+$) (▲/△) spores were suspended in (A) BHIS or (B) TY medium supplemented with 10 mM taurocholate (open symbols) or without taurocholate (closed symbols). The change in OD_{600} during germination was measured over time at 37 °C. The data represents the average of three biological replicates and error bars represent the standard deviation from the mean.

approximately thirty minutes, when compared to the wild-type and restored strains (Figure 3.5A). On the other hand, when cells were allowed to outgrow in peptide-rich, TY medium, *C. difficile* KNM6 (Δ*selD*) had an outgrowth defect, compared to the wild-type and restored strains (Figure 3.5B). Importantly, there was no delay in the initiation of outgrowth. Rather, there was a defect in growth in this medium.

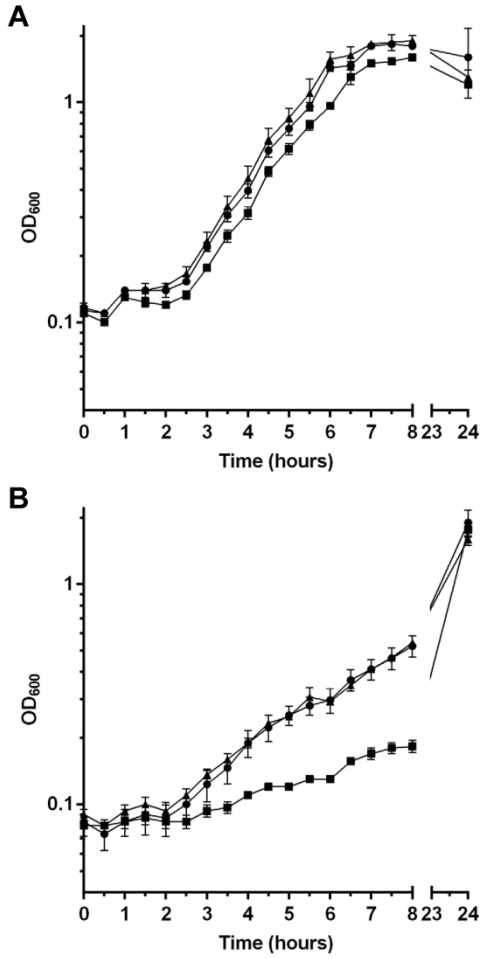


Figure 3.5 *selD* plays a role in outgrowth of *C. difficile* R20291 spores. Purified *C. difficile* R20291 (wild-type) (●), *C. difficile* KNM6 (Δ*selD*) (■), and *C. difficile* KNM9 (Δ*selD*::*selD*$^{+}$) (▲) spores were germinated and then resuspended in (A) BHIS or (B) TY medium, and the OD_{600} was measured over an 8 hour period and then again at 24 hours. Data points represent the average from three independent experiments and error bars represent the standard deviation from the mean.

3.3.7. *RNA-seq comparison of wild-type versus Δ*selD *strains*

If Stickland metabolism is a primary source of energy for *C. difficile*, we wondered why the growth defect in peptide-rich medium was not more severe compared to wild-type and restored strains. We hypothesized that the *C. difficile* KNM6 (Δ*selD*) strain was able to compensate for the loss of global selenoprotein synthesis by up- or down-regulating other metabolic pathways. To determine what pathways were being differentially expressed, we performed RNA-seq on the wild-type *C. difficile* R20201 and *C. difficile* KNM6 (Δ*selD*) strains during exponential growth in TY medium supplemented with glucose. At the time of performing these experiments, we were not aware of whether glucose would have an effect on gene expression and, therefore, decided to include this carbohydrate in the medium. We performed these experiments at both ~1.7% and 4% hydrogen levels to determine if the abundance of atmospheric hydrogen impacts *C. difficile* physiology.

First, we noticed that when comparing expression of genes in wild-type cells grown at low and high hydrogen levels, the cells had significant down-regulation of ribosomal RNA protein in the high hydrogen condition. This suggests that the cells are under stress and downregulating the synthesis of ribosomes under these conditions (196). Due to this finding, we chose to exclude highly up- or down-regulated gene expression due to hydrogen levels when comparing wild-type and mutant expression levels. After this exclusion, we obtained a narrower list for comparison of wild-type and Δ*selD* strains at low hydrogen levels and further determined their known or hypothesized function and what pathway that gene belongs (197, 198) (Table 3.1). As

expected, many of the differentially expressed genes have functions that are involved in metabolism; 36.35% of functions had genes which were up-regulated, and 22.24% of functions had genes which were down-regulated (Figure 3.6). In both cases, there were large numbers of genes of unknown function. Other than metabolism and unknown function, the largest group of up-regulated genes in the Δ*selD* strain were characterized as transferases; although, many of these genes also fit into another KEGG pathway, *e.g.* metabolism. Another group of genes whose expression was increased were those of transporters and those involved in secretion, potentially increasing the cell's import and export systems to compensate for this growth impairment. The largest group of genes which were down-regulated, besides metabolism or unknown functions, belonged to secretion systems or hydrolases.

On first glance, *CDR20291_0963* and *CDR20291_0962* are the two most highly up-regulated genes in the Δ*selD* strain. *CDR20291_0963* is uncharacterized *in C. difficile* but has high similarity to the AlgI protein, an alginate O-acetyltransferase, characterized in *Pseudomonas aeruginosa*. The other gene *CDR20291_0962* is an uncharacterized gene of no known function. When ran through NCBI BLAST and the Conserved Domain Database (199, 200), this protein sequence has a DHHW domain which is common among bacteria. Interestingly, both of these genes have recently been characterized to contain a riboswitch upstream, potentially explaining the expression profile (during growth for RNA extraction, the riboswitch may have flipped to the 'on' orientation) (201).

Table 3.1 Differentially expressed genes in the *C. difficile* Δ*selD* strain compared to the wild-type strain.

Name	Fold-change	Annotation/KEGG function	KEGG pathway
Up-regulated			
CDR20291_0963	70.544	alginate O-acetyl transferase complex protein	Unclassified Metabolism
CDR20291_0962	44.982	putative uncharacterized protein	Unknown
CDR20291_0440	17.707	cell surface protein (putative hemagglutinin / adhesin)	Unknown
CDR20291_1747	12.712	putative conjugative transposon regulatory protein	Unknown
cysM	8.129	putative O-acetylserine sulfhydrylase	Energy and Amino Acid Metabolism
cysA	5.460	serine O-acetyltransferase	Energy and Amino Acid Metabolism
mtlF	4.901	PTS system, mannitol-specific IIA component	Carbohydrate Metabolism, Transferases
mtlD	4.753	mannitol-1-phosphate 5-dehydrogenase	Carbohydrate Metabolism, Oxidoreductases
CDR20291_1260	4.374	putative membrane protein	Unknown
mtlR	3.886	mannitol operon transcriptional antiterminator	Transcription factors
CDR20291_2627	3.031	cytosine permease	Transporters
mtlA	2.984	PTS system, mannitol-specific IIBC component	Carbohydrate Metabolism, Transferases
CDR20291_2626	2.978	putative carbon-nitrogen hydrolase	Unknown
CDR20291_0474	2.623	putative exported protein	Unknown
ribH	2.430	6,7-dimethyl-8-ribityllumazine synthase (riboflavin synthase beta chain)	Metabolism of Cofactors and Vitamins, Transferases
CDR20291_1593	2.319	putative arsenical pump membrane protein	Unknown
ribD	2.232	diaminohydroxyphosphoribosylaminopyrimidine deaminase / 5-amino-6- (5-phosphoribosylamino) uracil reductase (riboflavin biosynthesis protein)	Metabolism of Cofactors and Vitamins, Cell Community, Oxidoreductases, Hydrolases
fdhA	2.185	L-seryl-tRNA (Ser) seleniumtransferase	Metabolism of Other Amino Acids, Translation, Transferases
modA	2.176	molybdate transport system substrate-binding protein	Membrane Transport, Transporters
CDR20291_0182	2.171	putative membrane-associated metalloprotease	Unknown
CDR20291_2397	2.167	TetR-family transcriptional regulator	Unknown
CDR20291_0146	2.153	putative riboflavin transporter	Unknown
CDR20291_0178	2.152	cyclopropane-fatty-acyl-phospholipid synthase	Unclassified Metabolism, Transferases

Table 3.1 Continued.

Name	Fold-change	Annotation/KEGG function	KEGG pathway
Up-regulated			
CDR20291_2191	2.145	putative pilin protein, general secretion pathway protein G	Membrane Transport, Secretion System
CDR20291_1446	2.124	prophage antirepressor-related protein	Unknown
CDR20291_2012	2.100	ABC-2 type transport system ATP-binding protein	Transporters
CDR20291_1405	2.089	putative polysaccharide deacetylase	Unknown
ribB	2.083	riboflavin synthase	Metabolism of Cofactors and Vitamins, Transferases
CDR20291_0763	2.073	aconitate hydratase	Carbohydrate Metabolism, Lyases
glyA	2.069	glycine hydroxymethyltransferase	Carbohydrate, Energy, Amino Acid, Other Amino Acid, and Cofactors and Vitamin Metabolism, Transferases
iorB	2.027	indolepyruvate ferredoxin oxidoreductase, beta subunit	Unclassified Metabolism, Oxidoreductases
CDR20291_1816	2.007	putative lipoprotein	Unknown
CDR20291_0765	2.001	MarR-family transcriptional regulator	Unknown
Down-regulated			
nfo	2.042	deoxyribonuclease IV	Replication and Repair, Hydrolases
CDR20291_1819	2.090	putative phage-related deoxycytidylate deaminase (putative late competence protein)	Nucleotide Metabolism, Secretion Systems, Hydrolases
CDR20291_0177	2.112	putative oxidoreductase, NAD / FAD binding subunit	Unknown
CDR20291_3153	2.114	type IV pilus assembly protein	Secretion Systems
CDR20291_1319	2.257	putative phage shock protein	Unknown
CDR20291_1685	2.274	putative exonuclease	Unknown
CDR20291_0175	2.333	anaerobic carbon-monoxide dehydrogenase catalytic subunit	Energy Metabolism, Xenobiotics biodegradation and Metabolism, Oxidoreductases
CDR20291_2875	2.424	conserved hypothetical protein	Unknown
CDR20291_2931	2.599	putative amino acid permease	Unknown
CDR20291_2077	3.005	putative sodium:dicarboxylate symporter	Unknown
CDR20291_2932	3.055	putative glutamine amidotransferase	Peptidases
selD	49.582	selenide, water dikinase	Metabolism of Other Amino Acids, Transferases

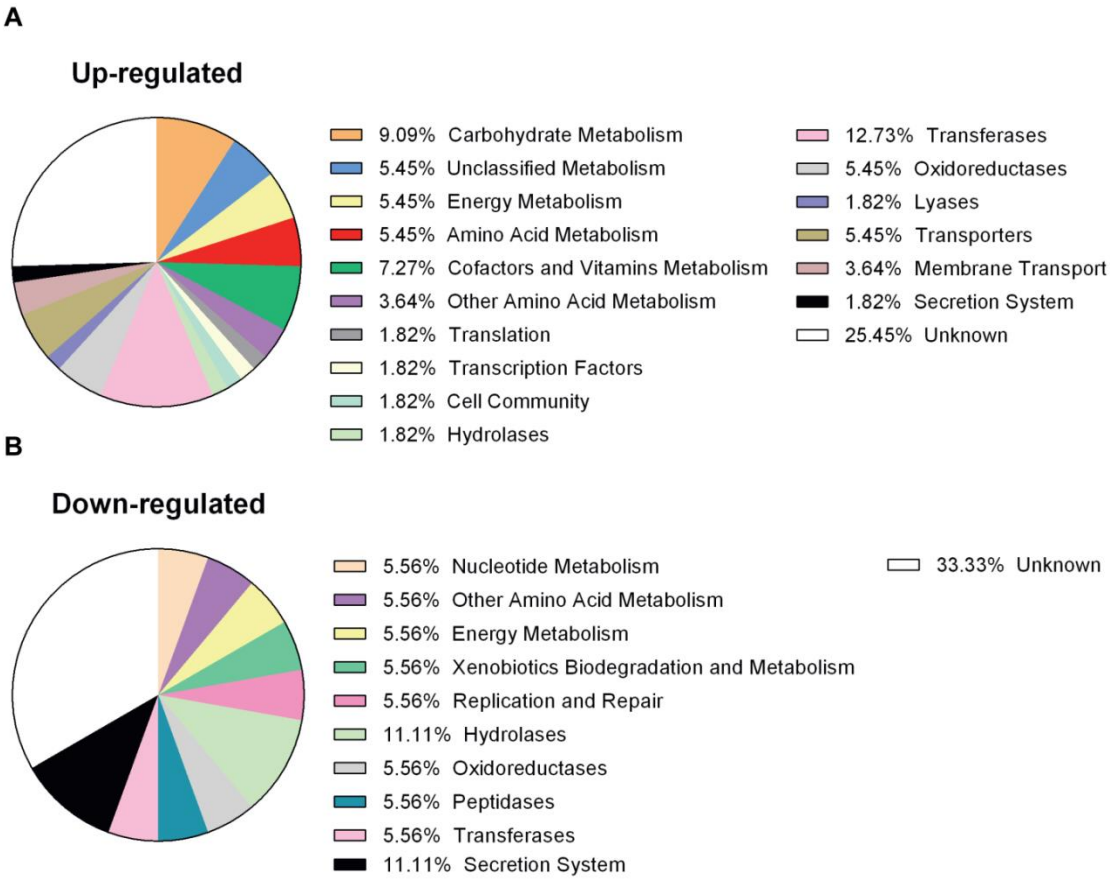


Figure 3.6 Distribution of functions of up- and down-regulated genes from RNA-seq.

Pie charts of the KEGG functional pathways of genes which were (A) up-regulated and (B) down-regulated from RNA-seq analysis. Percentages of each function are shown in the legends.

Every gene in the *mtlARFD* operon had highly upregulated expression in the Δ*selD* strain. The expression levels varied between almost 3- to 5-fold up-regulation compared to wild-type. We hypothesize that mannitol metabolism may play a larger role in a strain lacking selenoproteins. In this pathway, mannitol is taken into the cell and broken down to mannitol-1-phosphate by MtlF and MtlA. Then, MtlD uses NAD^+ to convert mannitol-1-phosphate into NADH and fructose-6-phosphate which can be used

in glycolysis (202-205). It is possible that the reaction could run in the reverse to generate NAD^+ instead of consuming it, but this does not appear to be the case since the other genes in the *mtl* operon, besides *mtlD*, are also upregulated.

Besides mannitol catabolism, cysteine and methionine metabolism are increased by the up-regulation of *cysM* and *cysA*. It is interesting that these genes utilize H_2S produced by *C. difficile* to generate L-cysteine and acetate (206, 207). Acetate is also a final product in the reduction of glycine by Stickland metabolism (53). It is possible that the cell is generating acetate through this method in order to utilize the molecule in other metabolic pathways.

Another metabolic process that had multiple genes up-regulated was riboflavin synthesis: *ribD*, a deaminase and reductase; *ribH*, a lumazine synthase; and *ribB*, a riboflavin synthase. Not all of the genes involved in riboflavin synthesis were up-regulated. Riboflavin is an essential cofactor which can be broken down into flavins (flavin mononucleotide, FMNs) and flavin adenine dinucleotide (FAD) to be used by the cell (208). A gene proposed to be a riboflavin transporter, *CDR20291_0146*, was also up-regulated from our RNA-seq analysis. This essential cofactor is likely being used in multiple processes and its synthesis or uptake is being up-regulated to feed these metabolic pathways.

Multiple genes involved in transport or utilization of amino acids were down-regulated, possibly due to decreased dependence on Stickland metabolism. *CDR20291_2931*, a putative amino acid permease, and *CDR20291_2932*, a putative glutamine amidotransferase which can remove the amido group from glutamine and, in a

second reaction, transfer the resulting ammonia group to a specific substrate, both are proposed to be involved in amino acid usage (209, 210). As well, a putative sodium:dicarboxylate symporter, *CDR20291_2077*, could also be involved in uptake of some amino acids into the cell. The down-regulation of these three genes indicates that the cell is moving its resources away from the degradation of amino acids as a source of energy.

3.3.8. *Validation of RNA-seq by quantitative RT-PCR*

We then chose some of those genes which we have highlighted as well as others of interest to validate the RNA-seq results by RT-qPCR. *C. difficile* R20291 and *C. difficile* KNM6 strains were grown as previously described for RNA-seq. RNA was extracted, DNA was depleted, and cDNA libraries were generated. Using the housekeeping gene, *rpoB*, as an internal normalization control, we determined the fold change of transcripts in the KNM6 (Δ*selD*) strain compared to the R20291 (wild-type) strain at both 4% and 1.7% hydrogen levels. The target genes *CDR20291_0962* and *CDR20291_0963* had transcript levels which were interesting at growth in 1.7% hydrogen percentage in the mutant compared to wild-type since it appeared that, in one biological replicate, the riboswitch was likely turned on while other biological replicates had fold change values close to one and were likely turned in the off position (Figure 3.7A and 3.7B). One target gene, *mtlF*, had slightly increased fold change in transcript levels at both 4% and 1.7% by RT-qPCR and the trend correlated to the results found from RNA-seq (Figure 3.7C). It is likely that mannitol utilization is one large factor which is helping the strain grow in absence of selenoproteins.

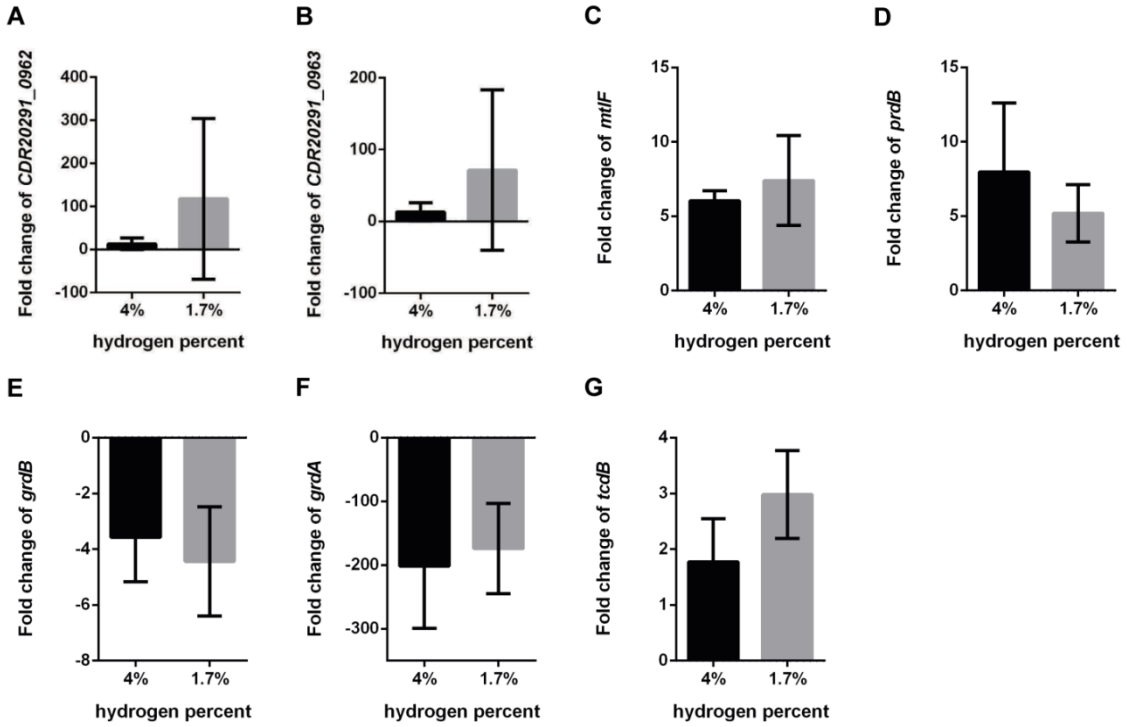


Figure 3.7 Fold change of transcript levels in *C. difficile* KNM6 (Δ*selD*) compared to *C. difficile* R20291.

RNA was extracted from *C. difficile* R20291 and KNM6 (Δ*selD*) strains grown to an OD_{600} of 0.6 in TYG medium, DNA was depleted, and cDNA was synthesized. Quantitative reverse transcription-PCR was performed to determine the fold change of transcript levels in the KNM6 (Δ*selD*) strain compared to wild-type R20291 strain at each 4% and 1.7% hydrogen levels. Fold change for the following genes are shown: (A) *CDR20291_0962*, (B) *CDR20291_0963*, (C) *mtlF*, (D) *prdB*, (E) *grdB*, (F) *grdA*, and (G) *tcdB*. Data represents the average from three independent experiments run in technical triplicate and error bars represent the standard deviation from the mean. Technical replicate outliers were excluded.

Since we hypothesized Stickland metabolism to be decreased in the mutant strain, we chose to analyze the proline reductase, *prdB*, as well as the glycine reductases, *grdB* and *grdA*. The fold change in expression of *prdB* was higher than what was observed in the RNA-seq analysis (Figure 3.7D). In RT-qPCR, the transcription of *grdB* was down-regulated in the mutant compared to wild-type (3.6- and 4.4-fold down-

regulated in 4% and 1.7% hydrogen, respectively) (Figure 3.7E). On the other hand, the transcription of *grdB* was slightly up-regulated in RNA-seq (1.3- and 2-fold up-regulated in 4% and 1.7% hydrogen, respectively). In addition, the down-regulation of *grdA* was more severe in RT-qPCR than seen in RNA-seq (Figure 3.7F). It appears that the *grd* operon was more down-regulated in this extraction of RNA than that for RNA-seq or the coverage was poor when analyzing the RNA-seq data. Since the levels of proline and glycine have been shown to have an effect on toxin production (55), we wanted to confirm the results from RNA-seq through RT-qPCR and found similar levels of *tcdB* expression at 4% hydrogen (no change), but levels were increased to 3-fold up-regulated in 1.7% hydrogen compared to no change from RNA-seq (Figure 3.7G).

3.3.9. *Virulence*

Finally, we wanted to test whether a *selD* mutant had any affect in an animal model of infection due to the growth, sporulation, and outgrowth defects we observed *in vitro*. Syrian hamsters, a model of acute *C. difficile* infection, were used to analyze the *in vivo* importance of *selD*. The clindamycin-treated hamsters were orally gavaged with 5,000 *C. difficile* spores derived from wild-type (R20291), Δ*selD* (KNM6), or Δ*selD*::*selD*$^{+}$ (KNM9) strains and monitored for signs of CDI. Our *C. difficile* R20291 strain does not seem to be as virulent as most other strains used in this model, which cause wild-type hamsters to succumb to disease within four days with 5 times less spores (1,000 spores). Regardless, the differences of our three strains in this infection were not significantly different (Figure 3.8). We are currently creating

these mutants in a strain shown to cause significant and severe infection in the Syrian hamster model of infection, *C. difficile* UK1.

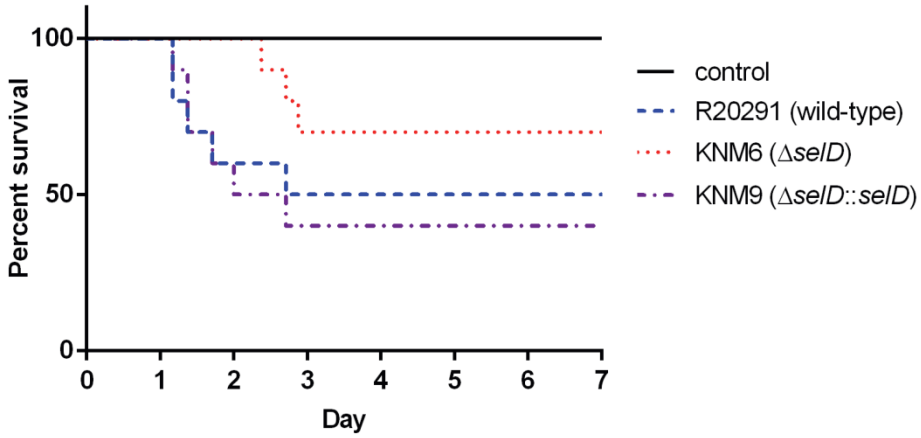


Figure 3.8 *In vivo* model of infection using Syrian hamsters.
Kaplan-Meier survival curve of clindamycin-treated Syrian hamsters inoculated with 5,000 spores of *C. difficile* R20291 (wild-type), *C. difficile* KNM6 (Δ*selD*), and *C. difficile* KNM9 (Δ*selD*::*selD*$^{+}$) strains. Control hamsters were antibiotic-treated but were given vehicle (dH_2O) instead of spores.

3.4. Discussion

The use of CRISPR-Cas9 gene editing tools for the manipulation of bacteria to understand key physiology has been invaluable to science as a whole. The use of our recently developed *C. difficile* CRISPR-Cas9 gene editing system led us to make a mutation in *selD*, which codes for the first enzyme in the synthesis of selenocysteines. We took this tool further to increase the usefulness and range of mutations by replacing the tetracycline-inducible promoter, that regulated the expression of *cas9*, with the recently developed xylose-inducible promoter by Muh *et al* (124). From this we were able to make the first *C. difficile* CRISPR-Cas9 insertion (at the time of this publication) by restoring *selD* at its native locus. This restored strain is genetically identical to that of

wild-type *C. difficile* R20291. In future uses, we plan to leave a "scar" to differentiate the restored strain from wild-type (*e.g.*, addition of a second stop codon at the end of the gene).

Selenium is a naturally occurring and essential trace element for all organisms. It is a necessary cofactor for certain enzymes (*e.g.*, thioredoxin reductase) and is also incorporated into proteins through a selenocysteine synthesis pathway. This element is commonly found in high protein foods such as meats and dairy products and is consumed in small amounts (211). It is listed as an antioxidant as well as known to have antibacterial and antifungal properties (211, 212). To this effect, sodium selenite has been shown to enhance the efficacy of antibiotics against *Staphylococcus aureus* and, more recently, *C. difficile* by increasing the sensitivity of the bacterium to the antibiotic (212, 213). The need for studying the uses and incorporation of selenium in bacteria such as the gut pathogen *C. difficile* is important for developing new therapeutics.

In this study, we demonstrate the importance of selenocysteine synthesis for *C. difficile* to survive and, potentially, to disseminate. In addition to our previous finding that SelD is important for growth of *C. difficile* vegetative cells, here we find that *selD* is required for maximum spore formation and outgrowth of vegetative cells from the spore. These new findings correlate with a recent article by Pellissery *et al*. where they found that sub-minimal inhibitory concentration levels of sodium selenite inhibited spore outgrowth of two hypervirulent *C. difficile* strains. Here they discuss that the concentration of sodium selenite used is much larger (0.14 mM) compared to the minimal requirement for *C. difficile* growth (50 nM) and this is likely toxic to the cell

(213). In the future we would like to test whether the KNM6 (Δ*selD*) strain is resistant to the toxic effects of sodium selenite. One possibility for the physiological defects we observed in our study is that the selenium in the medium is not being processed and incorporated into proteins through the selenocysteine synthesis pathway. Either this is toxic to the cell or the lack of incorporation prevents the production of essential or important proteins for growth, spore formation, and / or outgrowth. We suggest the latter hypothesis due to the growth defect in peptide-rich medium, which should nudge the cells to use Stickland metabolism.

SelD was found to have no significant role in the early events of *C. difficile* spore germination. This finding was not surprising since germination is an enzymatic process and would likely have no difference in different media. The spores were generated on BHIS agar medium and would have no germination defect phenotype. In the future, spores purified after sporulating on TY agar medium should be tested for their ability to germinate.

From our RNA-seq experiment, we found a few metabolic pathways which were up-regulated in the mutant KNM6 (Δ*selD*) strain compared to the wild-type R20291 strain. One pathway which has been of interest is the mannitol utilization pathway where the entire operon was upregulated, *mtlARFD*. Interestingly, this data correlates with an *in vivo* analysis of gene expression and metabolites present after antibiotic-treatment where genes for mannitol utilization, along with other ABC sugar transporters and sugar alcohol catabolism, were some of the highest expressed and mannitol was one of the most abundant metabolites after treatment with cefoperazone (49). Current work is

ongoing to make a deletion mutant of this operon as well as a double mutant with the KNM6 (Δ*selD*) strain to study *C. difficile* physiology in these strains, with an emphasis on growth, sporulation, and outgrowth. Along these lines, we would like to test whether the addition of mannitol restores or enhances growth of *C. difficile* strains, in particular the KNM6 (Δ*selD*) strain. In addition, we would like to test the roles of *cysA* and *cysM* (cysteine synthesis) as well as *ribD*, *ribB*, and *ribH* (riboflavin synthesis) in contributing to the growth of the KNM6 (Δ*selD*) strain.

In summary, we have demonstrated the importance of selenoproteins in *C. difficile* growth, spore formation, and outgrowth of vegetative cells from the spore. As well, we performed RNA-seq which has led to the finding of multiple pathways being upregulated when selenocysteine synthesis is abolished. There is much to be studied about the global loss of selenoproteins in *C. difficile* given the phenotypes we see here in this study. This work is the start of a new project which will help in understanding the impact of selenoproteins on growth and metabolism in *C. difficile*.

4. DETERMINING THE CONTRIBUTION OF FECAL BILE ACIDS TO THE INFECTION OF *CLOSTRIDIOIDES DIFFICILE* IN HUMANS

In Section 4, we analyze the bile acid profiles of hospitalized patients' fecal samples in order to study the roles of antibiotic usage, diarrheal symptoms, *C. difficile* infection, and recurring *C. difficile* infection. This study gives a clinical insight into the importance of bile acids on human health and infection. This study was performed in collaboration with Drs. Abraham Sonenshein and Yoav Golan at Tufts University Medical School.

4.1. Introduction

Clostridioides difficile is the major cause of antibiotic-associated diarrhea and a major threat to hospitalized patients with disrupted colonic microflora (3, 4). Importantly, community acquired *C. difficile* infection (CDI) is on the rise and is associated with increased virulence and susceptibility (4). In most strains of *C. difficile*, the vegetative cells secrete two toxins, TcdA (Toxin A) and TcdB (Toxin B) (21). In some strains, a third toxin, a ribosyl-transferase (CDT; binary toxin) is produced (4). The two primary toxins damage the colonic epithelial cells by targeting the Rho family of small GTPases resulting in the reorganization of the actin cytoskeleton. This damage causes apoptosis and shedding of the epithelial cells which inhibits reabsorption of water and causes diarrhea, the main and most well-known symptom of CDI (24-26).

There are several testing methods for the detection of *C. difficile* in patients. An enzyme immunoassay (EIA) and polymerase chain reaction (PCR) are the two most

common testing strategies. One type of EIA detects *C. difficile* glutamate dehydrogenase (GDH) and has specificity as low as 60% but a high sensitivity (85-100%). Even though this test is quick and inexpensive, it does not indicate whether the strain is toxigenic. Another EIA is directed against toxins A and B (Tox A/B). This detection method is quick but less sensitive (63-94%) than other available tests, has a specificity of 75-100%, and is used by more than 90% of clinicians. However, this test is being phased out due to the more precise nucleic acid tests. PCR amplification of *C. difficile tcdA* and *tcdB* are becoming more commonplace in the healthcare setting due to the rapid testing time and accurate results with a specificity of 99% and a sensitivity of 100%. It should be noted that asymptomatic / transiently-colonized patients also test positive and clinical manifestations should be warranted before performing a PCR test (antibiotic treatment of these individuals can lead to active *C. difficile* infection) (214).

C. difficile, a strict anaerobe, persists in the environment as a dormant endospore and this form is resistant to many environmental stressors such as UV radiation, heat, and household cleaning agents (28, 215). *C. difficile* spores germinate in response to certain host-derived bile acids and certain amino acid co-germinants (34, 35, 40, 41). The human liver synthesizes two primary bile acids, cholic acid and chenodeoxycholic acid, which are then conjugated with either taurine or glycine (*e.g.*, cholic acid + taurine yields taurocholic acid) and diet regulates the ratio of taurine-conjugated to glycine-conjugated bile acids in humans. These bile acids are stored in the gallbladder until secreted into the upper small intestines when a meal is ingested. During digestion, most of the bile acids are reabsorbed and recycled back to the liver to be used again in another

round of digestion. Those that escape enterohepatic recirculation enter the large intestine where they are modified by the resident microflora. Many bacteria in the large intestine express on their cell surfaces, bile salt hydrolases that deconjugate the amino acids from the base bile acid structure (*e.g.*, taurine is removed from taurocholic acid to yield cholic acid and taurine; Figure 1.1). Once deconjugated, a small number of bacteria (e.g., *Bacteroides thetaiotaomicron*, *Clostridium scindens*, *Clostridium sordellii*, and *Clostridium hiranonis*) take up the deconjugated bile acid and remove the 7α-hydroxyl group resulting in secondary bile acids. In this process, cholic acid is converted deoxycholic acid and chenodeoxycholic acid is converted to lithocholic acid (39, 216).

The role of bile acids in *C. difficile* physiology is that of activators and inhibitors of germination of spores and / or growth of vegetative cells. The primary bile acid cholic acid and its derivatives act as activators of *C. difficile* spore germination while the primary bile acid chenodeoxycholic acid and its derivative lithocholic acid are inhibitors of *C. difficile* spore germination (Figure 1.1) (38, 40, 42, 43, 217). Interestingly, the secondary bile acid, deoxycholic acid, is both an activator of spore germination and an inhibitor of vegetative cell growth (40, 218). Another growth inhibitory bile acid is chenodeoxycholic acid, which is also an inhibitor of spore germination (40, 42). Despite no direct *in vivo* evidence, a hypothesis that has reached dogma-like status in the field is that in a healthy individual, *C. difficile* spores germinate in response to cholic acid or its derivatives in the small intestine. When these germinated spores reach the large intestine, deoxycholic acid inhibits growth and prevents colonization. In a dysbiotic

colon, the bacteria that produce the growth-inhibitory secondary bile acids are not present, resulting in a *C. difficile* susceptible environment (219).

In a double-blinded study we quantified the abundance of bile acids in fecal samples from hospitalized patients from Tufts Medical Center in Boston, MA. After quantification and normalization, we generated bile acid profiles for each patient and compared these profiles between different health statuses: non-diarrheal healthy individuals, those with diarrhea but not associated with CDI, individuals with diarrhea caused by CDI confirmed by either EIA or PCR, non-antibiotic treated patients, antibiotic treated patients, non-recurrent CDI patients, and those with recurrent CDI.

4.2. Materials and methods

4.2.1. *Stool sample collection*

Stool samples were collected from hospitalized patients at Tufts Medical Center, Boston, MA, and stored at -80 ºC. The patients chosen for sample collection were treated or not treated with antibiotics, were diagnosed as negative or positive for *C. difficile* infection (CDI), and, if positive, were separated according to whether they did or did not have recurrent CDI. Samples were assigned random six-digit numbers (*e.g.*, 111-222) unrelated to the identity of the patient; the researchers who carried out the experiments had no access to the names of the patients. This method of sample collection was approved by the Tufts Health Sciences Institutional Review Board and Texas A&M Institutional Review Board #IRB2013-3030D.

4.2.2. *Lyophilization for bile acid analysis*

The caps of tubes holding frozen stool sample aliquots were pierced with a 15-gauge needle to provide a single vent. Batches of 15 tubes were then placed in a lyophilization flask that had been pre-chilled to -80 ^{o}C; the flask was then moved directly to a Vertex lyophilization unit equipped with a TriVac pump. Lyophilization was at a final pressure of approximately 100 millitorr and was usually completed within 16-24 hours. Tubes were weighed for calculation of final dry weight, sealed with Parafilm and stored at -80 ^{o}C. Multiple tubes derived from each stool sample were lyophilized in order to achieve a total dry weight of at least 200 mg. The lyophilized samples were transferred to Texas A&M University for bile acid analysis.

4.2.3. *Bile acid extraction*

At least 200 mg of lyophilized stool was incubated for 15 minutes at 100 °C in 7 mL of 100% ethanol in a 15-mL Falcon screw-cap tube. This solution was centrifuged for 10 min at 3,000 x g at room temperature, and the resulting supernatant was transferred to a clean tube. The pellet was again extracted, and the resulting supernatant was combined with the first. Fifty-microliters of 10 mM hyodeoxycholic acid was added to each supernatant as an internal standard, and extraction control, and total volume of the supernatant was then dried in a glass tube at 65 °C under a stream of air. Once dry, 800 µL of 100% methanol was added and the tube was incubated at 65 °C overnight to solubilize the extract. The sample was centrifuged at 14,000 x g for 1 min to remove any insoluble material, and the supernatant fluid was transferred to a glass HPLC vial.

4.2.4. *Bile acid separation and quantification*

The extracted bile acids and pure bile acid standards (used to generate a standard curve, see below) were separated by reverse phase HPLC using a Shimadzu Prominence HPLC system. One-hundred microliter samples were injected and separated on a Syncronis C18 column using a separation buffer consisting of 53% methanol, 24% acetonitrile, 23% water, and 30 mM ammonium acetate, pH 5.6 (pH was modified by addition of glacial acetic acid). The bile acids were detected using a Sedere Sedex model 80 LT-ELSD (low temperature – evaporative light scattering detector) with an air pressure of 50 psi at a temperature of 94 °C.

Samples of pure bile acids taurocholic acid (TA), glycocholic acid (GCA), taurochenodeoxycholic acid (TCDCA), glycochenodeoxycholic acid (GCDCA), hyodeoxycholic acid (HDCA), cholic acid (CA), chenodeoxycholic acid (CDCA), deoxycholic acid (DCA), and lithocholic acid (LCA) (listed in order of elution time) were separated by HPLC to generate a standard curve. The amounts tested were 10, 25, 50, and 100 nmol of each bile acid, calculated from injections of 10, 25, 50, and 100 µL of a 1 mM standard mix. Individual bile acid standards were run separately if the peaks overlapped. The area under each peak was calculated and plotted against the concentration of bile acid and a trend line was generated for each. The R^2 value from each standard curve was > 0.9.

In each of the HPLC-separated fecal extracts, each peak was manually identified and the area under the curve determined using the Shimadzu LC solution software. Based on the elution times of each bile acid from the standard curves, each peak was

assigned to the bile acid that eluted at the same time. The amount of bile acid (in nmol) was calculated using the generated standard curve for each bile acid and normalized to the amount of the HDCA internal standard present in the sample. The amount of each bile acid was divided by the weight of the sample's lyophilized feces to calculate the amount of each bile acid per mg of feces.

4.2.5. *Statistical analysis*

Data represents the average and standard deviation from the mean of all samples in the given group.

4.3. Results

4.3.1. *Patient and data collection in double-blinded study*

To determine if patient bile acid profiles can be used as a marker for determining *C. difficile* susceptibility, fecal samples from patients at the Tufts Medical Center in Boston, MA were collected and analyzed in a double-blinded fashion. Samples were lyophilized and then analyzed for bile acid content and profiles by bile acid extraction and separation methods. After analysis, groups into which the de-identified patient samples fell were revealed. The samples were categorized into one or more of eight different patient groups (Table 4.1). In Study 1, bile acid profiles were analyzed for patients who either exhibited diarrheal symptoms or not; those that had diarrhea were then separated into three categories: diarrhea associated with CDI confirmed by EIA, diarrhea associated with CDI confirmed by PCR, and diarrhea not associated with CDI. Study 2 analyzes patient groups with or without the treatment of antibiotics. All patients in this study were not associated with CDI. Lastly, Study 3 analyzes bile acid profiles of

individuals with non-recurrent CDI versus those with recurrent CDI. In total, 218 samples were used for bile acid analysis where 137 of these samples were unique; meaning, 81 samples belonged to two patient groups.

The detected bile acids in each patient sample were quantified, normalized to an internal standard, and then calculated in reference to the amount of fecal sample used in the extraction. The average total bile acid concentration of individuals in each patient group varied without any significant difference (Table 4.2).

Table 4.1 Distribution of patient samples.

Study	Patient groups	Total number of samples per group	Samples used for bile acid analysis
1	A. Non-diarrheal	30	29
	B. Diarrheal, CDI-negative	46	45
	C. Diarrheal, CDI-positive by EIA	36	28
	D. Diarrheal, CDI-positive by PCR	17	13
2	A. Not treated with antibiotics (CDI-negative)	30	29
	B. Treated with antibiotics (CDI-negative)	30	27
3	A. Non-recurrent CDI	19	19
	B. Recurrent CDI	39	28

Samples were excluded from analysis due to low sample yield.

Table 4.2 Average bile acid concentration in stool samples.

Study	Patient Group	Average total bile acid concentration (nmol/mg feces)
1	1A. Non-diarrheal	13.35 ± 13.52
	1B. Diarrheal, CDI-negative	14.03 ± 22.67
	1C. Diarrheal, CDI-positive (EIA)	5.74 ± 7.55
	1D. Diarrheal, CDI-positive (PCR)	17.43 ± 37.97
2	2A. Non-antibiotic treated	11.49 ± 10.78
	2B. Antibiotic-treated	16.71 ± 26.97
3	3A. Non-recurrent CDI	13.50 ± 32.06
	3B. Recurrent CDI	10.13 ± 15.61

4.3.2. *Relationship between diarrheal disease and abundance of specific bile acids*

In order to determine the impact of diarrheal disease on bile acid profiles of fecal samples, individuals with no diarrhea, those with diarrhea not associated with CDI, and those with diarrhea associated with CDI confirmed either by EIA or PCR were analyzed for their bile acid content and abundances. Those with no diarrhea generally had higher quantities of secondary bile acids and deconjugated bile acids out of the total bile acid amounts compared to those with diarrheal symptoms (Table 4.3 and Figure 4.1A and 4.1B). Also, there were low amounts of TA in each patient sample (germination promoting bile acid), and most individuals contained some level of CDCA and LCA instead (germination inhibiting bile acids) (Figure 4.1C). These results correlate with the hypothesis that in healthy individuals (*i.e.*, those without diarrhea or disruption to the colonic microflora) higher numbers of bile acids that inhibit *C. difficile* growth are present in stool.

Table 4.3 Relative abundance of bile acids in stool samples of CDI-negative and CDI-positive patients.

Patient Group	Relative abundance of bile acid types				
	Conjugated/ Total	Primary/ Total	Secondary/T otal	TA/ Total	(CDCA+LCA)/ Total
A. Non-diarrheal	10.31% ± 22.29	27.10% ± 37.01	72.90% ± 37.01	2.23% ± 6.59	36.10% ± 16.79
B. Diarrheal, CDI-negative	29.80% ± 40.55	56.71% ± 45.58	34.41% ± 43.35	13.55% ± 26.93	22.54% ± 27.80
C. Diarrheal, CDI-positive (EIA)	21.26% ± 29.47	57.38% ± 44.41	31.91% ± 41.23	16.10% ± 29.04	19.32% ± 27.16
D. Diarrheal, CDI-positive (PCR)	29.12% ± 33.28	68.14% ± 44.85	8.78% ± 25.33	1.84% ± 4.45	8.62% ± 17.05

Regardless of whether the patients' diarrheal symptoms were associated with CDI (CDI-negative or CDI-positive), those patients with diarrheal symptoms had higher amounts of conjugated and / or primary bile acids when compared to those without diarrheal symptoms (Table 4.3 and Figures 4.1D, 4.1E, 4.1G, and 4.1H). Conjugated bile acids were higher in those with diarrhea compared to those without diarrhea, conjugated bile acids were not a majority of the total bile acids present in the individual diarrheal positive samples. Of these, general primary bile acids were more abundant in these patients than just those which were conjugated. There were no significant correlations in individual diarrheal patient samples comparing TA abundances to CDCA and LCA abundances (Figures 4.1F and 4.1I). Of note, patients who were confirmed CDI-positive by PCR had lower amounts of CDCA and LCA compared to all other groups in the study and had more similar abundances of taurocholate to non-diarrheal patients (Table 4.3).

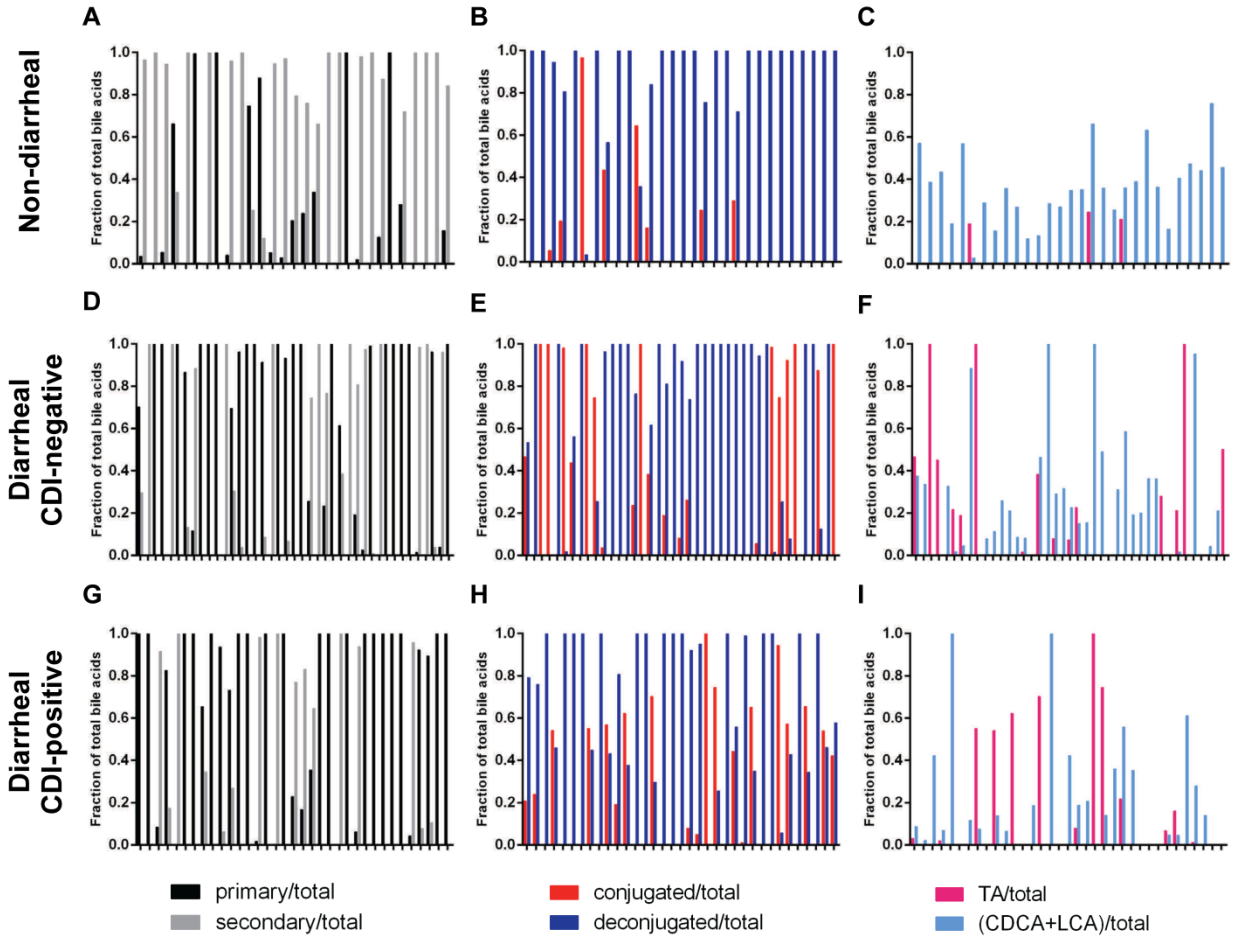


Figure 4.1 Bile acid abundances per individual sample in patients with different diarrheal symptoms.

In each group (non-diarrheal, diarrheal CDI-negative, and diarrheal CDI-positive) bile acid abundances were calculated and normalized. The ratio of individual samples of primary vs secondary (A, D, G), conjugated vs deconjugated (B, E, H), and TA abundance vs CDCA + LCA (C, F, I) abundances were analyzed for their fraction out of total bile acids for each group. The colors of the bile acid category analyzed are indicated in the legends.

Many in the field hypothesize that *C. difficile*-infected individuals would have higher abundances of the bile acids that favor germination or growth of vegetative *C. difficile* cells (11). Looking at the differences in abundances of bile acids in patients with different states of diarrheal disease, it cannot be concluded that individuals with CDI

have higher abundances of TA compared to CDCA and LCA or conjugated bile acids when looking at the group as a whole. In some individuals, there are high abundances of TA or conjugated bile acids, but this is not represented across all individuals (Figures 4.1H and 4.1I). These findings indicate that individuals with diarrhea have a greater abundance of primary bile acids than secondary bile acids, as shown in other studies.

4.3.3. *Impact of antibiotic treatment, independent of CDI, on the abundance of specific bile acids*

Next, antibiotic-treated and non-antibiotic-treated patients were analyzed for their bile acid profiles. In both cases, patients who had diarrheal symptoms were included in these groups, but the diarrhea was not associated with CDI (CDI-negative). It is well established that broad-spectrum antibiotic treatment significantly changes the ecology of the gut flora (9, 60, 61, 220). Here, we wanted to see whether antibiotic treatment affected the abundances and types of bile acids in stool.

Non-antibiotic treated patient samples had very high amounts of deconjugated bile acids and secondary bile acids (Table 4.4). Most individuals in this group had nearly all secondary bile acids or deconjugated bile acids (Figures 4.2 A and 4.2B). Generally, non-antibiotic treated patient samples contained greater abundances of CDCA and LCA compared to TA (Table 4.4 and Figure 4.2C). In contrast, patients who underwent antibiotic treatment had samples which contained a majority of primary bile acids compared to secondary bile acids (Table 4.4 and Figure 4.2D). These data correlate with previously published data suggesting that antibiotic treatment removes the bacteria which convert primary bile acids to secondary bile acid (60, 220). Antibiotic treated

Table 4.4 Relative abundance of bile acids in stool samples of antibiotic-treated and non-treated patients.

Patient Group	Relative abundance of bile acid types				
	Conjugated/Total	Primary/ Total	Secondary/ Total	TA/ Total	(CDCA+LCA)/ Total
A. Non-antibiotic treated	7.96% ± 20.78	21.78% ± 37.53	74.77% ± 39.89	2.44% ± 8.23	35.63% ± 17.49
B. Antibiotic-treated	41.76% ± 42.62	73.34% ± 39.23	19.26% ± 33.74	11.83% ± 22.30	11.51% ± 24.48

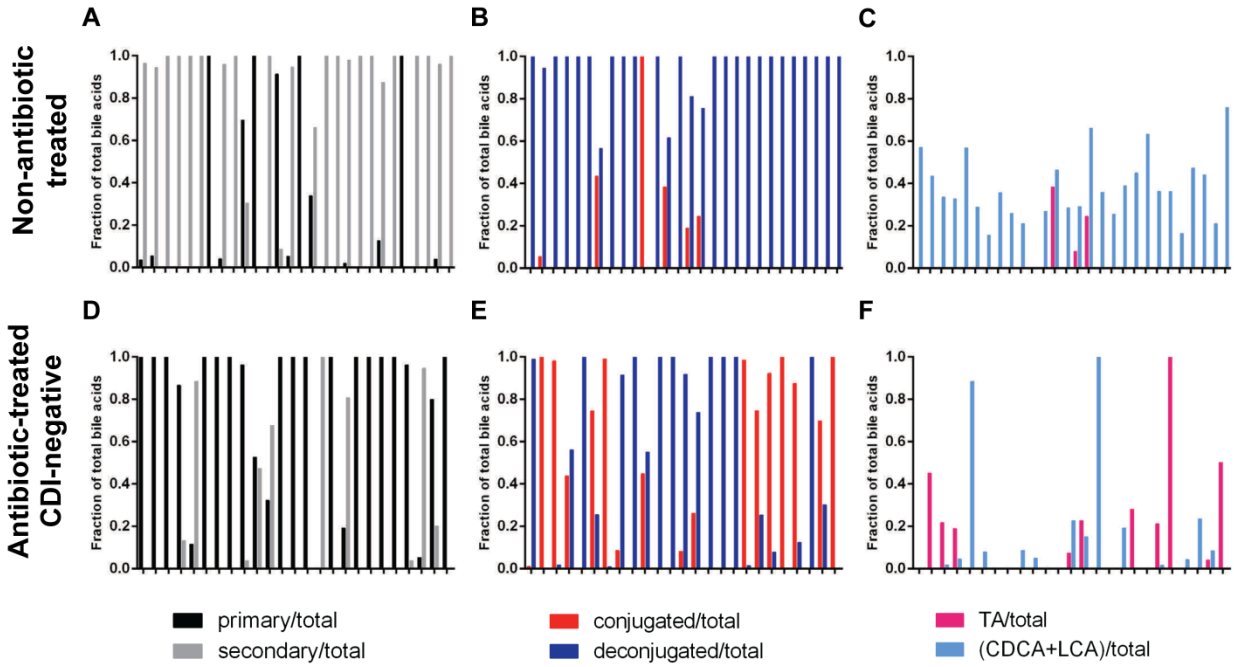


Figure 4.2 Bile acid abundances per individual sample in patients with or without antibiotic treatment.
In each group (non-diarrheal, diarrheal CDI-negative, and diarrheal CDI-positive) bile acid abundances were calculated and normalized. The ratio of individual samples of primary vs secondary (A and D), conjugated vs deconjugated (B and E), and TA abundance vs CDCA + LCA (C and F) abundances were analyzed for their fraction out of total bile acids for each group. The colors of the bile acid category analyzed are indicated in the legends.

samples contained higher abundances of conjugated bile acids compared to non-treated patients (Table 4.4 and Figure 4.2B and 4.2E). Lastly, there were slightly higher abundances of TA and lower abundances of CDCA and LCA in these samples compared to non-antibiotic treated samples (Table 4.4 and Figure 4.2C and 4.2F).

4.3.4. *Relationship between recurrent and non-recurrent CDI and abundance of specific bile acids*

In order to assess whether bile acids play a role in affecting recurrence of CDI, fecal samples of patients with recurrent or non-recurrent CDI were analyzed for bile acid content and abundances. As in antibiotic-treated and diarrheal symptom patient samples, the abundances of primary bile acids in both recurrent and non-recurrent groups are high and almost identical, and primary bile acids make up the majority of the bile acids present in these samples (Table 4.5 and Figures 4.3A and 4.3D). There were no significant differences in regards to primary or secondary bile acid abundances between recurrent and non-recurrent groups.

Table 4.5 Relative abundance of bile acids in stool samples of patients with recurrent vs. non-recurrent CDI.

Patient Group	Relative abundance of bile acid types				
	Conjugated/ Total	Primary/ Total	Secondary/ Total	TA/ Total	(CDCA+LCA)/ Total
A. Non-recurrent CDI	28.78% ± 31.04	60.66% ± 44.85	18.29% ± 33.46	8.35% ± 18.07	12.77% ± 18.45
B. Recurrent CDI	16.67% ± 23.95	60.55% ± 41.78	32.31% ± 39.30	6.48% ± 16.02	22.49% ± 27.53

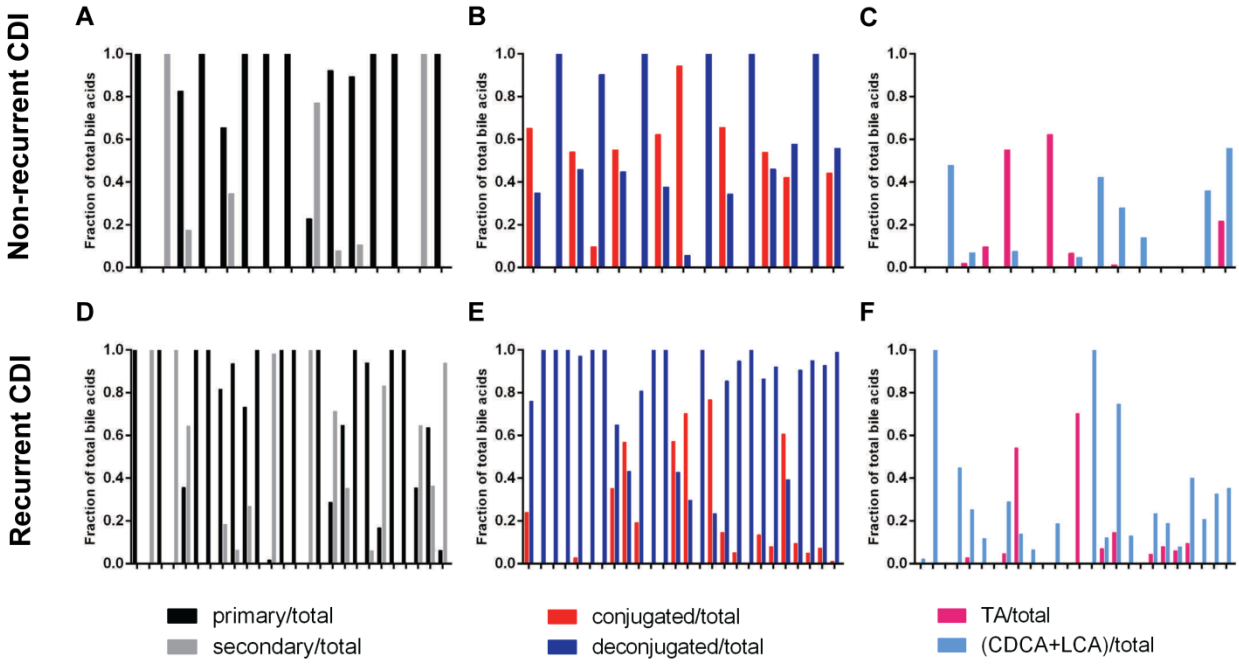


Figure 4.3 Bile acid abundances per individual sample in patients with non-recurrent or recurrent CDI.

In each group (non-diarrheal, diarrheal CDI-negative, and diarrheal CDI-positive) bile acid abundances were calculated and normalized. The ratio of individual samples of primary vs secondary (A and D), conjugated vs deconjugated (B and E), and TA abundance vs CDCA + LCA (C and F) abundances were analyzed for their fraction out of total bile acids for each group. The colors of the bile acid category analyzed are indicated in the legends.

The abundances of conjugated bile acids in both non-recurrent and recurrent groups are similar to those of patients with diarrheal symptoms or are antibiotic treated (Table 4.5). No significant differences were seen as well. When comparing individual samples, there are slightly more recurrent CDI individuals who have almost all deconjugated bile acids, and no conjugated bile acids, than compared to the non-recurrent CDI group (11 out of 26 individuals in the recurrent group have all deconjugated bile acids compared to 5 out of 15 in the non-recurrent group) (Figures 4.3B and 4.3E). Otherwise, there are no significant differences. There are marginally

more abundances of CDCA and LCA in recurrent CDI patient samples compared to non-recurrent samples, but this change is not significant. There were also similar levels of TA in both groups (Table 4.5 and Figures 4.3C and 4.3F).

4.4. Discussion

With the increase in antibiotic usage and increased virulence and dissemination of *C. difficile* toxigenic strains, it is imperative for alternative therapeutics to be developed to combat CDI (4). Because bile acids play not only a critical role in spore germination but also an important role in inhibiting vegetative growth, studying the impact of bile acids on human health *in vivo* is necessary (40). This study has given valuable insight into the bile acid profiles of hospitalized patients with varying levels of disease in relation to CDI.

The results obtained here have given indications of what the role of specific bile acids are in different disease states. Patients without diarrheal symptoms and not subjected to antibiotic treatment can convert their primary bile acids to deconjugated, secondary bile acids which are antagonistic to *C. difficile* vegetative growth. These patients act as healthy individuals, at least in respect to their colonic health. If *C. difficile* spores were able to germinate, they would not be able to survive due to the inhibitory growth role of the bile acids present as well as not be able to colonize the healthy or normal colon which is colonized with the resident microflora. On the other hand, patients with diarrheal symptoms and / or treated with antibiotics maintain conjugated and 7α-hydroxylated bile acids which support CDI and fail to prevent infection. Other studies have also found that an increase in primary bile acids is associated with

individuals with CDI (11, 12). Another point to add from this study is that antibiotic treatment increases the abundance of primary bile acids (some of which retain conjugation, including TA). It is likely the disruption in the gut microbiota which is causing the higher abundance of primary bile acids which is a favorable state for *C. difficile* spore germination and growth (11, 12). This work further supports the idea that specific bile acids can be associated with specific disease states relating to a health individual or one with diarrhea and / or on antibiotics, which a good majority is CDI positive.

A question which has been difficult to study is if there are differences in bile acid abundances in individuals with recurrent or non-recurrent CDI and if they play any role in recurrence. Patients with non-recurrent and recurrent CDI have similar relative levels of conjugated and primary (7α-hydroxylated) bile acids. There were no significant changes which can be attributed to recurrence of CDI found in this study.

While this study provides valuable information of *in vivo* bile acid profiles of hospitalized patients, the study does have some limitations. Many samples were collected and analyzed, but the sample size was still small. Some individuals had bile acid levels which were undetectable by our methods. A larger sample population size is needed to more accurately assess the bile acid profiles of each group. As well, not all samples had enough material for the 200 mg weight needed for maximal bile acid extraction and larger quantities of the samples were needed. The only information released was the health status of the patients. Other biological variable such as sex, age, and diet were not considered in this study. All patients were hospitalized and presumably

given the same diet, but this was not controlled. Overall, the study needs to be expanded and given more controlled parameters for sample collection in order to more accurately represent these different health status groups.

Analyzing the bile acid profiles of these patients is not the only aspect of the double-blinded study. Collaborators are working to analyze these fecal samples for microbiome analyses as well as measuring levels of specific bacteria, *i.e.*, *C. difficile*, *C. leptum*, *C. scindens*, and *Bacteroides thetaiotaomicron*, by qPCR amplifying for 16s rRNA or the *baiCD* gene from *C. scindens*, a gene whose product is involved in 7α-dehydroxylation of bile acids (216). The combination of all of this data will give further insight into the role of bile acids as well as the role of the resident microflora on diarrheal symptoms, antibiotic usage, and *C. difficile* infection. Regardless, the results obtained in this study not only help to support some previously established hypotheses but also give a well characterized view of bile acids in human samples, which has not been done before.

5. CONCLUSIONS

Clostridioides difficile is the leading cause of antibiotic-associated diarrhea and a significant health threat to individuals taking antibiotics (4). Understanding the underlying mechanisms and physiology of this pathogen both *in vitro* and *in vivo* will aid in developing antimicrobial therapies. In these sections, I studied the current genetic tools available for *C. difficile* and discussed the need for a more efficient and precise genome editing system, such as the CRISPR-Cas9 system. We were the first to develop and adapt this system for use in *C. difficile* using wild-type Cas9 under the control of a tetracycline inducible promoter. We later modified the system to control the expression of c*as9*, using the recently-published xylose inducible promoter, in order to create the first *C. difficile* insertion by CRISPR-Cas9 genome editing (124). Our *C. difficile* CRISPR-Cas9 genome editing tool is freely available and has become popular in the *C. difficile* community. This system has already been used by another in the field (221) as well as acted as a reference for developing a CRISPRi tool for *C. difficile* (124).

While this tool has proven its usefulness, there are still aspects of the system which are unknown and which need to be improved. The main issue with the *C. difficile* CRISPR-Cas9 gene editing system is determining what makes a "good" target sequence to be targeted by the gRNA. Here we describe "good" as efficient and making mutations within the first few passages of induction. The current issue is that researchers have to make multiple iterations of their deletion / insertion plasmid each with a different 20 nt target sequence within or surrounding the gene or fragment to be deleted / inserted. It is

currently a guessing game of which target sequence will result in a mutation. Ongoing work in the lab has included performing *in vitro* cleavage assays with Cas9 and different target sequences located on a gRNA to analyze the locations of cleavages within the *C. difficile* genome. These data need to be confirmed *in vivo*. Moreover, a larger experiment is ongoing in which we analyze the *in vivo* incorporation of new spacers into a CRISPR-Cas array located on a plasmid in the *C. difficile* R20291 strain. This experiment is based on previous experiments performed in *Staphylococcus aureus* that analyzed spacer acquisition from the *S. aureus* genome (222, 223). From these results, we hope to be able to determine where in the *C. difficile* genome cleavage and/or spacer acquisition is preferred and hope to conclude from these results what makes an efficient target sequence.

We utilized our system to develop a series of *C. difficile* R20291 mutant strains in which we made a clean deletion of *selD*, a selenophosphate synthetase important in selenoprotein synthesis, as well as a restoration strain of *selD* at its native locus. The resulting, *selD*-restored, strain was genetically identical to the wild-type *C. difficile* R20291 strain. We characterized the physiology of the *C. difficile* Δ*selD* strain by analyzing the different processes of the *C. difficile* life cycle and found defects in growth, spore formation, and outgrowth of a vegetative cell from the spore. We then identified potential genes or operons which are highly expressed during exponential growth of the Δ*selD* strain through RNA-seq and validation through qPCR. We found multiple pathways which were upregulated in the *selD* mutant strain. For example, the entire *mtl* operon (mannitol catabolism), *cysA* and *cysM* (cysteine synthesis), and *ribD*,

ribB, and *ribH* (riboflavin synthesis) were upregulated. The first is interesting due to the correlation with *in vivo* expression and metabolite data post-cefoperazone treatment (49). Possibly, the most interesting is the expression for mannitol utilization (*mtlARFD*) because the entire operon was highly up-regulated. Future work should be done to make mutations in these genes in combination with the *selD* mutant. It has been mentioned that mannitol utilization is repressed by the presence of glucose (202). An experiment looking at the expression of the *mtl* operon when grown in the presence or absence of glucose would be interesting to determine if the absence of glucose increases expression of these genes. However, because the expression for mannitol utilization is increased in the *selD* mutant, our data suggests that selenoproteins may directly or indirectly repress the utilization of mannitol in a wild-type cell. We are interested to know whether mannitol plays a significant role for *C. difficile* selenoprotein synthesis as well as why this metabolite is increased *in vivo* after treatment of mice with cefoperazone.

Though these studies gave a thorough characterization of *selD*, there is still much to be learned. RNA-seq needs to be performed again but rather than collecting RNA during exponential growth, RNA can also be collected during stationary phase to give a better understanding of the role of *selD* in this growth phase. This could lead to an explanation for the sporulation defect we observed. We hypothesize that different transcripts will be expressed during stationary phase compared to exponential growth in the mutant compared to wild-type. We would like to see if there is a significant pathway, or multiple pathways, being upregulated at this time. This could also give some insight into why the *selD* mutants have difficulty sporulating.

In a study by Jackson-Rosario *et al*. (2009), the authors found that auranofin, a gold-containing compound which has been used to treat rheumatoid arthritis, inhibited the growth of *C. difficile* vegetative cells (224). This drug has a proposed mechanism of action of binding to selenide and forming a stable complex that inhibits the uptake of selenium into the cell. The addition of auranofin to growth medium blocked the uptake of selenium into *C. difficile*, and the authors saw an accumulation of selenium in the medium. If additional selenite or L-selenocysteine was added to the growth medium, the inhibition of auranofin was decreased or competitively inhibited. The authors concluded by suggesting that selenium metabolism is a good target for antimicrobial development due to its target of only those organisms which require selenoproteins for growth, but more work needs to be done to determine the effect on the host, which possibly also contains selenoproteins (224). This work is particularly interesting to us due to the correlations we saw when analyzing our *C. difficile* Δ*selD* strain. In future studies, we would like to test the effects of auranofin on the *C. difficile* R20291 wild-type strain as well as the *C. difficile* Δ*selD* strain particularly in growth but, more importantly, what happens when auranofin is included in the growth medium with our *selD* mutant. Another experiment would be to add in auranofin during exponential growth to the *C. difficile* R20291 wild-type and Δ*selD* strains and analyze gene expression through RNA-seq. Due to the prevalence of Stickland metabolism on *C. difficile* growth *in vivo* and that this relies on two selenoproteins, we are eager to understand more in depth the effect of auranofin on selenoprotein synthesis (48-50).

The metabolites present during infection of *C. difficile* in a host have an important impact on growth and survival of this bacterium. Other metabolites that impact another part of the *C. difficile* life cycle are bile acids during germination. We analyzed the presence and abundance of specific bile acids in hospitalized patients' fecal samples. These patients from Tufts Medical Center were part of a double-blinded study to determine the differences in bile acid profiles of patients with different health statuses. These health statuses were broken down based on antibiotic treatment, diarrheal symptoms, whether the patients were CDI positive (EIA or PCR), and recurring CDI. We found that healthy individuals not on antibiotics and without diarrheal symptoms tended to have larger abundances of deconjugated and secondary bile acids. This data strongly correlates with other hypotheses that healthy individuals have this bile acid profile and this helps to protect against CDI, which is largely accepted in the field as the mechanism by which the microbiome protects against CDI (though all studies on this are entirely correlative in nature) (12, 220). Conversely individuals who had diarrheal symptoms or were on antibiotic treatment generally had larger abundances of conjugated and primary bile acids. These bile acids are associated with favorable conditions for *C. difficile* to germinate along with the disrupted microbiome due to the antibiotic usage. Lastly, there was no difference in bile acids profiles of patients with recurrent CDI versus a single occurrence of CDI. This suggests that the composition and abundance of bile acids do not contribute to recurrence of infection. These results help further our understanding of bile acids content and abundance *in vivo*.

Further studies include analyzing the microbiomes of each of these patients as well as testing for specific bacteria which are known to carry out 7α-dehydroxylation, as previously discussed. This study will need to be expanded to determine significance. More data needs to be collected on the patient to go along with the data from analysis. Information needed includes age, sex, what antibiotics were taken if applicable, how many recurrences if applicable, and diet to see if these factors correspond to specific bile acid profiles and could be analyzed rather than assumed the individual was an outlier. These controls would help narrow the focus of the study.

Overall, the studies outlined in each section help contribute to our understanding of *C. difficile* physiology as well as expand on current genetic tools available for this pathogen. While there is much to learn about these topics, my studies provide a solid basis for future directions that, we hope, will ultimately aid in the development of non-antibiotic therapies for the treatment of CDI.

REFERENCES

1. **Oren A, Garrity GM.** 2016. List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol **66:**1-3.

2. **Lawson PA, Citron DM, Tyrrell KL, Finegold SM.** 2016. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prevot 1938. Anaerobe **40:**95-99.

3. **Rupnik M, Wilcox MH, Gerding DN.** 2009. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. Nat Rev Microbiol **7:**526-536.

4. **Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ.** 2016. *Clostridium difficile* infection. Nature Reviews Disease Primers **2:**16020.

5. **Hall IC, O'Toole E.** 1935. Intenstinal flora in new-born infants with a description of a new pathogenic anaerobe, *Bacillus difficilis*. American Journal of Diseases in Children **49:**390-402.

6. **Bartlett JG, Chang TW, Gurwith M, Gorbach SL, Onderdonk AB.** 1978. Antibiotic-associated pseudomembranous colitis due to toxin-producing clostridia. N Engl J Med **298:**531-534.

7. **Poutanen SM, Simor AE.** 2004. *Clostridium difficile*-associated diarrhea in adults. CMAJ **171:**51-58.

8. **McFarland LV.** 1998. Epidemiology, risk factors and treatments for antibiotic-associated diarrhea. Dig Dis **16:**292-307.

9. **Bignardi GE.** 1998. Risk factors for *Clostridium difficile* infection. J Hosp Infect **40:**1-15.

10. **Pepin J, Saheb N, Coulombe MA, Alary ME, Corriveau MP, Authier S, Leblanc M, Rivard G, Bettez M, Primeau V, Nguyen M, Jacob CE, Lanthier L.** 2005. Emergence of fluoroquinolones as the predominant risk factor for *Clostridium difficile*-associated diarrhea: a cohort study during an epidemic in Quebec. Clin Infect Dis **41:**1254-1260.

11. **Theriot CM, Koenigsknecht MJ, Carlson PE, Jr., Hatton GE, Nelson AM, Li B, Huffnagle GB, J ZL, Young VB.** 2014. Antibiotic-induced shifts in the

mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. Nat Commun **5:**3114.

12. **Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A, Littmann E, van den Brink MR, Jenq RR, Taur Y, Sander C, Cross JR, Toussaint NC, Xavier JB, Pamer EG.** 2015. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. Nature **517:**205-208.

13. **Wilson KH, Perini F.** 1988. Role of competition for nutrients in suppression of *Clostridium difficile* by the colonic microflora. Infect Immun **56:**2610-2614.

14. **Zar FA, Bakkanagari SR, Moorthi KM, Davis MB.** 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.

15. **Allen CA, Babakhani F, Sears P, Nguyen L, Sorg JA.** 2013. Both fidaxomicin and vancomycin inhibit outgrowth of *Clostridium difficile* spores. Antimicrob Agents Chemother **57:**664-667.

16. **Zhu D, Sorg JA, Sun X.** 2018. *Clostridioides difficile* biology: sporulation, germination, and corresponding therapies for *C. difficile* infection. Front Cell Infect Microbiol **8:**29.

17. **Maroo S, Lamont JT.** 2006. Recurrent *Clostridium difficile*. Gastroenterology **130:**1311-1316.

18. **Leffler DA, Lamont JT.** 2015. *Clostridium difficile* infection. N Engl J Med **372:**1539-1548.

19. **Kelly CP, LaMont JT.** 2008. *Clostridium difficile*--more difficult than ever. N Engl J Med **359:**1932-1940.

20. **(US) C.** 2013. Antibiotic resistant threats in the United States. US Department of Health and Human Services.

21. **Voth DE, Ballard JD.** 2005. *Clostridium difficile* toxins: mechanism of action and role in disease. Clin Microbiol Rev **18:**247-263.

22. **Lyras D, O'Connor JR, Howarth PM, Sambol SP, Carter GP, Phumoonna T, Poon R, Adams V, Vedantam G, Johnson S, Gerding DN, Rood JI.** 2009. Toxin B is essential for virulence of *Clostridium difficile*. Nature **458:**1176-1179.

23. **Kuehne SA, Cartman ST, Heap JT, Kelly ML, Cockayne A, Minton NP.** 2010. The role of toxin A and toxin B in *Clostridium difficile* infection. Nature **467:**711-713.

24. **Qa'Dan M, Ramsey M, Daniel J, Spyres LM, Safiejko-Mroczka B, Ortiz-Leduc W, Ballard JD.** 2002. *Clostridium difficile* toxin B activates dual caspase-dependent and caspase-independent apoptosis in intoxicated cells. Cell Microbiol **4:**425-434.

25. **Just I, Selzer J, Wilm M, von Eichel-Streiber C, Mann M, Aktories K.** 1995. Glucosylation of Rho proteins by *Clostridium difficile* toxin B. Nature **375:**500-503.

26. **Pruitt RN, Lacy DB.** 2012. Toward a structural understanding of *Clostridium difficile* toxins A and B. Front Cell Infect Microbiol **2:**28.

27. **Setlow P.** 2014. Germination of spores of *Bacillus* species: what we know and do not know. J Bacteriol **196:**1297-1305.

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

29. **Cowan AE, Olivastro EM, Koppel DE, Loshon CA, Setlow B, Setlow P.** 2004. Lipids in the inner membrane of dormant spores of *Bacillus* species are largely immobile. Proc Natl Acad Sci U S A **101:**7733-7738.

30. **Moir A, Corfe BM, Behravan J.** 2002. Spore germination. Cell Mol Life Sci **59:**403-409.

31. **Warth AD, Strominger JL.** 1969. Structure of the peptidoglycan of bacterial spores: occurrence of the lactam of muramic acid. Proc Natl Acad Sci U S A **64:**528-535.

32. **Ali S, Moore G, Wilson AP.** 2011. Spread and persistence of *Clostridium difficile* spores during and after cleaning with sporicidal disinfectants. J Hosp Infect **79:**97-98.

33. **Baines SD, O'Connor R, Saxton K, Freeman J, Wilcox MH.** 2009. Activity of vancomycin against epidemic *Clostridium difficile* strains in a human gut model. J Antimicrob Chemother **63:**520-525.

34. **Francis MB, Allen CA, Shrestha R, Sorg JA.** 2013. Bile acid recognition by the *Clostridium difficile* germinant receptor, CspC, is important for establishing infection. PLoS Pathog **9:**e1003356.

35. **Shrestha R, Cochran AM, Sorg JA.** 2019. The requirement for co-germinants during *Clostridium difficile* spore germination is influenced by mutations in *yabG* and *cspA*. PLoS Pathog **15:**e1007681.

36. **Deakin LJ, Clare S, Fagan RP, Dawson LF, Pickard DJ, West MR, Wren BW, Fairweather NF, Dougan G, Lawley TD.** 2012. The *Clostridium difficile spo0A* gene is a persistence and transmission factor. Infect Immun **80:**2704-2711.

37. **Bliss DZ, Johnson S, Clabots CR, Savik K, Gerding DN.** 1997. Comparison of cycloserine-cefoxitin-fructose agar (CCFA) and taurocholate-CCFA for recovery of *Clostridium difficile* during surveillance of hospitalized patients. Diagn Microbiol Infect Dis **29:**1-4.

38. **Wilson KH, Kennedy MJ, Fekety FR.** 1982. Use of sodium taurocholate to enhance spore recovery on a medium selective for *Clostridium difficile*. J Clin Microbiol **15:**443-446.

39. **Ridlon JM, Kang DJ, Hylemon PB.** 2006. Bile salt biotransformations by human intestinal bacteria. J Lipid Res **47:**241-259.

40. **Sorg JA, Sonenshein AL.** 2008. Bile salts and glycine as cogerminants for *Clostridium difficile* spores. J Bacteriol **190:**2505-2512.

41. **Shrestha R, Sorg JA.** 2018. Hierarchical recognition of amino acid co-germinants during *Clostridioides difficile* spore germination. Anaerobe **49:**41-47.

42. **Sorg JA, Sonenshein AL.** 2009. Chenodeoxycholate is an inhibitor of *Clostridium difficile* spore germination. J Bacteriol **191:**1115-1117.

43. **Sorg JA, Sonenshein AL.** 2010. Inhibiting the initiation of *Clostridium difficile* spore germination using analogs of chenodeoxycholic acid, a bile acid. J Bacteriol **192:**4983-4990.

44. **Neumann-Schaal M, Jahn D, Schmidt-Hohagen K.** 2019. Metabolism the Difficile Way: The Key to the Success of the Pathogen *Clostridioides difficile*. Front Microbiol **10:**219.

45. **Ragsdale SW.** 1997. The eastern and western branches of the Wood/Ljungdahl pathway: how the east and west were won. Biofactors **6:**3-11.

46. **Dannheim H, Will SE, Schomburg D, Neumann-Schaal M.** 2017. *Clostridioides difficile* 630Δerm *in silico* and *in vivo* - quantitative growth and extensive polysaccharide secretion. FEBS Open Bio **7:**602-615.

47. **Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeno-Tarraga AM, Wang H, Holden MT, Wright A, Churcher C, Quail MA, Baker S, Bason N, Brooks K, Chillingworth T, Cronin A, Davis P, Dowd L, Fraser A, Feltwell T, Hance Z, Holroyd S, Jagels K, Moule S, Mungall K, Price C, Rabbinowitsch E, Sharp S, Simmonds M, Stevens K, Unwin L, Whithead S, Dupuy B, Dougan G, Barrell B, Parkhill J.** 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet **38:**779-786.

48. **Janoir C, Deneve C, Bouttier S, Barbut F, Hoys S, Caleechum L, Chapeton-Montes D, Pereira FC, Henriques AO, Collignon A, Monot M, Dupuy B.** 2013. Adaptive strategies and pathogenesis of *Clostridium difficile* from *in vivo* transcriptomics. Infect Immun **81:**3757-3769.

49. **Jenior ML, Leslie JL, Young VB, Schloss PD.** 2017. *Clostridium difficile* colonizes alternative nutrient niches during infection across distinct murine gut microbiomes. mSystems **2**.

50. **Jenior ML, Leslie JL, Young VB, Schloss PD.** 2018. *Clostridium difficile* alters the structure and metabolism of distinct cecal microbiomes during initial infection to promote sustained colonization. mSphere **3**.

51. **Nisman B, Raynaud M, Cohen GN.** 1948. Extension of the Stickland reaction to several bacterial species. Arch Biochem **16:**473.

52. **Stickland LH.** 1934. Studies in the metabolism of the strict anaerobes (genus *Clostridium*): The chemical reactions by which *Cl. sporogenes* obtains its energy. Biochem J **28:**1746-1759.

53. **Stickland LH.** 1935. Studies in the metabolism of the strict anaerobes (genus *Clostridium*): The oxidation of alanine by *Cl. sporogenes*. IV. The reduction of glycine by *Cl. sporogenes*. Biochem J **29:**889-898.

54. **Stickland LH.** 1935. Studies in the metabolism of the strict anaerobes (Genus *Clostridium*): The reduction of proline by *Cl. sporogenes*. Biochem J **29:**288-290.

55. **Bouillaut L, Self WT, Sonenshein AL.** 2013. Proline-dependent regulation of *Clostridium difficile* Stickland metabolism. J Bacteriol **195:**844-854.

56. **Stadtman TC.** 1956. Studies on the enzymic reduction of amino acids: a proline reductase of an amino acid-fermenting *Clostridium*, strain HF. Biochem J **62:**614-621.

57. **Stadtman TC, Elliott P.** 1957. Studies on the enzymic reduction of amino acids. II. Purification and properties of D-proline reductase and a proline racemase from *Clostridium sticklandii*. J Biol Chem **228:**983-997.

58. **Jackson S, Calos M, Myers A, Self WT.** 2006. Analysis of proline reduction in the nosocomial pathogen *Clostridium difficile*. J Bacteriol **188:**8487-8495.

59. **Self WT.** 2010. Specific and Nonspecific Incorporation of Selenium into Macromolecules. Comprehensive Natural Products II Chemistry and Biology **5:**121-148.

60. **Theriot CM, Young VB.** 2014. Microbial and metabolic interactions between the gastrointestinal tract and *Clostridium difficile* infection. Gut Microbes **5:**86-95.

61. **Battaglioli EJ, Hale VL, Chen J, Jeraldo P, Ruiz-Mojica C, Schmidt BA, Rekdal VM, Till LM, Huq L, Smits SA, Moor WJ, Jones-Hall Y, Smyrk T, Khanna S, Pardi DS, Grover M, Patel R, Chia N, Nelson H, Sonnenburg JL, Farrugia G, Kashyap PC.** 2018. *Clostridioides difficile* uses amino acids associated with gut microbial dysbiosis in a subset of patients with diarrhea. Sci Transl Med **10**.

62. **McAllister KN, Bouillaut L, Kahn JN, Self WT, Sorg JA.** 2017. Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. Sci Rep **7:**14672.

63. **Finegold SM, Song Y, Liu C.** 2002. Taxonomy--General comments and update on taxonomy of Clostridia and Anaerobic cocci. Anaerobe **8:**283-285.

64. **Papoutsakis ET.** 2008. Engineering solventogenic clostridia. Curr Opin Biotechnol **19:**420-429.

65. **Joseph RC, Kim NM, Sandoval NR.** 2018. Recent developments of the synthetic biology toolkit for *Clostridium*. Front Microbiol **9:**154.

66. **Reyrat JM, Pelicic V, Gicquel B, Rappuoli R.** 1998. Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. Infect Immun **66:**4011-4017.

67. **Zhang N, Shao L, Jiang Y, Gu Y, Li Q, Liu J, Jiang W, Yang S.** 2015. I-SceI-mediated scarless gene modification via allelic exchange in *Clostridium*. J Microbiol Methods **108:**49-60.

68. **Hillmann F, Fischer RJ, Saint-Prix F, Girbal L, Bahl H.** 2008. PerR acts as a switch for oxygen tolerance in the strict anaerobe *Clostridium acetobutylicum*. Mol Microbiol **68:**848-860.

69. **Al-Hinai MA, Fast AG, Papoutsakis ET.** 2012. Novel system for efficient isolation of *Clostridium* double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration. Appl Environ Microbiol **78:**8112-8121.

70. **Cartman ST, Kelly ML, Heeg D, Heap JT, Minton NP.** 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.

71. **Nariya H, Miyata S, Suzuki M, Tamai E, Okabe A.** 2011. Development and application of a method for counterselectable in-frame deletion in *Clostridium perfringens*. Appl Environ Microbiol **77:**1375-1382.

72. **Heap JT, Ehsaan M, Cooksley CM, Ng YK, Cartman ST, Winzer K, Minton NP.** 2012. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. Nucleic Acids Res **40:**e59.

73. **Tripathi SA, Olson DG, Argyros DA, Miller BB, Barrett TF, Murphy DM, McCool JD, Warner AK, Rajgarhia VB, Lynd LR, Hogsett DA, Caiazza NC.** 2010. Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a *pta* mutant. Appl Environ Microbiol **76:**6591-6599.

74. **Wang Y, Zhang ZT, Seo SO, Choi K, Lu T, Jin YS, Blaschek HP.** 2015. Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. J Biotechnol **200:**1-5.

75. **Li Q, Chen J, Minton NP, Zhang Y, Wen Z, Liu J, Yang H, Zeng Z, Ren X, Yang J, Gu Y, Jiang W, Jiang Y, Yang S.** 2016. CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. Biotechnol J **11:**961-972.

76. **Wasels F, Jean-Marie J, Collas F, Lopez-Contreras AM, Lopes Ferreira N.** 2017. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. J Microbiol Methods **140:**5-11.

77. **Dong H, Tao W, Gong F, Li Y, Zhang Y.** 2014. A functional *recT* gene for recombineering of *Clostridium*. J Biotechnol **173:**65-67.

78. **Kusano K, Takahashi NK, Yoshikura H, Kobayashi I.** 1994. Involvement of RecE exonuclease and RecT annealing protein in DNA double-strand break repair by homologous recombination. Gene **138:**17-25.

79. **Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP.** 2007. The ClosTron: a universal gene knock-out system for the genus *Clostridium*. J Microbiol Methods **70:**452-464.

80. **Heap JT, Cartman ST, Kuehne SA, Cooksley C, Minton NP.** 2010. ClosTron-targeted mutagenesis. Methods Mol Biol **646:**165-182.

81. **Kelly ML, Ng YK, Cartman ST, Collery MM, Cockayne A, Minton NP.** 2016. Improving the reproducibility of the NAP1/B1/027 epidemic strain R20291 in the hamster model of infection. Anaerobe **39:**51-53.

82. **Bakker D, Buckley AM, de Jong A, van Winden VJ, Verhoeks JP, Kuipers OP, Douce GR, Kuijper EJ, Smits WK, Corver J.** 2014. The HtrA-like protease CD3284 modulates virulence of *Clostridium difficile*. Infect Immun **82:**4222-4232.

83. **Doudna JA, Charpentier E.** 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science **346:**1258096.

84. **Ishino Y, Shinagawa H, Makino K, Amemura M, 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.

85. **Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJ, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV.** 2015. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol **13:**722-736.

86. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709-1712.

87. **van der Oost J, Westra ER, Jackson RN, Wiedenheft B.** 2014. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Microbiol **12:**479-492.

88. **Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E.** 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature **471:**602-607.

89. **Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA.** 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature **507:**62-67.

90. **Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C.** 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology **155:**733-740.

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

92. **Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S.** 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature **468:**67-71.

93. **Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V.** 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. Nucleic Acids Res **39:**9275-9282.

94. **Gasiunas G, Barrangou R, Horvath P, 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.

95. **Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J.** 2013. RNA-programmed genome editing in human cells. Elife **2:**e00471.

96. **Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F.** 2013. Multiplex genome engineering using CRISPR/Cas systems. Science **339:**819-823.

97. **Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM.** 2013. RNA-guided human genome engineering via Cas9. Science **339:**823-826.

98. **Cho SW, Kim S, Kim JM, Kim JS.** 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol **31:**230-232.

99. **Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK.** 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol **31:**227-229.

100. **Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA.** 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol **31:**233-239.

101. **Sternberg SH, Doudna JA.** 2015. Expanding the biologist's toolkit with CRISPR-Cas9. Mol Cell **58:**568-574.

102. **Wu WY, Lebbink JHG, Kanaar R, Geijsen N, van der Oost J.** 2018. Genome editing by natural and engineered CRISPR-associated nucleases. Nat Chem Biol **14:**642-651.

103. **Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV, Zhang F.** 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell **163:**759-771.

104. **Shuman S, Glickman MS.** 2007. Bacterial DNA repair by non-homologous end joining. Nat Rev Microbiol **5:**852-861.

105. **Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O.** 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell **156:**935-949.

106. **Swarts DC, Jinek M.** 2018. Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. Wiley Interdiscip Rev RNA doi:10.1002/wrna.1481**:**e1481.

107. **Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA.** 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell **152:**1173-1183.

108. **Bayat H, Modarressi MH, Rahimpour A.** 2018. The conspicuity of CRISPR-Cpf1 system as a significant breakthrough in genome editing. Curr Microbiol **75:**107-115.

109. **Xu T, Li Y, Shi Z, Hemme CL, Li Y, Zhu Y, Van Nostrand JD, He Z, Zhou J.** 2015. Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. Appl Environ Microbiol **81:**4423-4431.

110. **Pyne ME, Bruder MR, Moo-Young M, Chung DA, Chou CP.** 2016. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. Sci Rep **6:**25666.

111. **Wang Y, Zhang ZT, Seo SO, Lynn P, Lu T, Jin YS, Blaschek HP.** 2016. Bacterial genome editing with CRISPR-Cas9: deletion, integration, single

nucleotide modification, and desirable "clean" mutant selection in *Clostridium beijerinckii* as an example. ACS Synth Biol **5:**721-732.

112. **Bruder MR, Pyne ME, Moo-Young M, Chung DA, Chou CP.** 2016. Extending CRISPR-Cas9 technology from genome editing to transcriptional engineering in the genus *Clostridium*. Appl Environ Microbiol **82:**6109-6119.

113. **Nagaraju S, Davies NK, Walker DJ, Kopke M, Simpson SD.** 2016. Genome editing of *Clostridium autoethanogenum* using CRISPR/Cas9. Biotechnol Biofuels **9:**219.

114. **Li Q, Seys FM, Minton NP, Yang J, Jiang Y, Jiang W, Yang S.** 2019. CRISPR-Cas9(D10A) nickase-assisted base editing in solvent producer *Clostridium beijerinckii*. Biotechnol Bioeng doi:10.1002/bit.26949.

115. **Xu T, Li Y, He Z, Van Nostrand JD, Zhou J.** 2017. Cas9 nickase-assisted RNA repression enables stable and efficient manipulation of essential metabolic genes in *Clostridium cellulolyticum*. Front Microbiol **8:**1744.

116. **Wang S, Hong W, Dong S, Zhang ZT, Zhang J, Wang L, Wang Y.** 2018. Genome engineering of *Clostridium difficile* using the CRISPR-Cas9 system. Clin Microbiol Infect doi:10.1016/j.cmi.2018.03.026.

117. **Huang H, Chai C, Li N, Rowe P, Minton NP, Yang S, Jiang W, Gu Y.** 2016. CRISPR/Cas9-based efficient genome editing in *Clostridium ljungdahlii*, an autotrophic gas-fermenting bacterium. ACS Synth Biol **5:**1355-1361.

118. **Wang S, Dong S, Wang P, Tao Y, Wang Y.** 2017. Genome editing in *Clostridium saccharoperbutylacetonicum* N1-4 with the CRISPR-Cas9 system. Appl Environ Microbiol **83**.

119. **Canadas IC, Groothuis D, Zygouropoulou M, Rodrigues R, Minton NP.** 2019. RiboCas: a universal CRISPR-based editing tool for *Clostridium*. ACS Synth Biol doi:10.1021/acssynbio.9b00075.

120. **Wang Y, Zhang ZT, Seo SO, Lynn P, Lu T, Jin YS, Blaschek HP.** 2016. Gene transcription repression in *Clostridium beijerinckii* using CRISPR-dCas9. Biotechnol Bioeng **113:**2739-2743.

121. **Zhang J, Hong W, Zong W, Wang P, Wang Y.** 2018. Markerless genome editing in *Clostridium beijerinckii* using the CRISPR-Cpf1 system. J Biotechnol **284:**27-30.

122. **Wen Z, Minton NP, Zhang Y, Li Q, Liu J, Jiang Y, Yang S.** 2017. Enhanced solvent production by metabolic engineering of a twin-clostridial consortium. Metab Eng **39:**38-48.

123. **Ingle P, Groothuis D, Rowe P, Huang H, Cockayne A, Kuehne SA, Jiang W, Gu Y, Humphreys CM, Minton NP.** 2019. Generation of a fully erythromycin-sensitive strain of *Clostridioides difficile* using a novel CRISPR-Cas9 genome editing system. Sci Rep **9:**8123.

124. **Muh U, Pannullo AG, Weiss DS, Ellermeier CD.** 2019. A xylose-inducible expression system and a CRISPRi-plasmid for targeted knock-down of gene expression in *Clostridioides difficile*. J Bacteriol doi:10.1128/JB.00711-18.

125. **Hong W, Zhang J, Cui G, Wang L, Wang Y.** 2018. Multiplexed CRISPR-Cpf1-mediated genome editing in *Clostridium difficile* toward the understanding of pathogenesis of *C. difficile* infection. ACS Synth Biol **7:**1588-1600.

126. **Woolston BM, Emerson DF, Currie DH, Stephanopoulos G.** 2018. Rediverting carbon flux in *Clostridium ljungdahlii* using CRISPR interference (CRISPRi). Metab Eng **48:**243-253.

127. **Zhang J, Zong W, Hong W, Zhang ZT, Wang Y.** 2018. Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. Metab Eng **47:**49-59.

128. **Metzger MJ, McConnell-Smith A, Stoddard BL, Miller AD.** 2011. Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. Nucleic Acids Res **39:**926-935.

129. **Barrangou R, Marraffini LA.** 2014. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. Mol Cell **54:**234-244.

130. **Bowater R, Doherty AJ.** 2006. Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. PLoS Genet **2:**e8.

131. **Su T, Liu F, Gu P, Jin H, Chang Y, Wang Q, Liang Q, Qi Q.** 2016. A CRISPR-Cas9 assisted non-homologous end-joining strategy for one-step engineering of bacterial genome. Sci Rep **6:**37895.

132. **Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, Hawkins JS, Lu CHS, Koo BM, Marta E, Shiver AL, Whitehead EH, Weissman JS, Brown ED, Qi LS, Huang KC, Gross CA.** 2016. A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. Cell **165:**1493-1506.

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

134. **Roman E, Coman I, Prieto D, Alonso-Monge R, Pla J.** 2019. Implementation of a CRISPR-based system for gene regulation in *Candida albicans*. mSphere **4**.

135. **Yamano T, Nishimasu H, Zetsche B, Hirano H, Slaymaker IM, Li Y, Fedorova I, Nakane T, Makarova KS, Koonin EV, Ishitani R, Zhang F, Nureki O.** 2016. Crystal structure of Cpf1 in complex with guide RNA and target DNA. Cell **165:**949-962.

136. **Yan MY, Yan HQ, Ren GX, Zhao JP, Guo XP, Sun YC.** 2017. CRISPR-Cas12a-assisted recombineering in bacteria. Appl Environ Microbiol **83**.

137. **Jiang Y, Qian F, Yang J, Liu Y, Dong F, Xu C, Sun B, Chen B, Xu X, Li Y, Wang R, Yang S.** 2017. CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*. Nat Commun **8:**15179.

138. **Ungerer J, Pakrasi HB.** 2016. Cpf1 is a versatile tool for CRISPR genome editing across diverse species of Cyanobacteria. Sci Rep **6:**39681.

139. **Kim D, Kim J, Hur JK, Been KW, Yoon SH, Kim JS.** 2016. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat Biotechnol **34:**863-868.

140. **Kleinstiver BP, Tsai SQ, Prew MS, Nguyen NT, Welch MM, Lopez JM, McCaw ZR, Aryee MJ, Joung JK.** 2016. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nat Biotechnol **34:**869-874.

141. **Zetsche B, Heidenreich M, Mohanraju P, Fedorova I, Kneppers J, DeGennaro EM, Winblad N, Choudhury SR, Abudayyeh OO, Gootenberg JS, Wu WY, Scott DA, Severinov K, van der Oost J, Zhang F.** 2017. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat Biotechnol **35:**31-34.

142. **Li Y, Pan S, Zhang Y, Ren M, Feng M, Peng N, Chen L, Liang YX, She Q.** 2016. Harnessing Type I and Type III CRISPR-Cas systems for genome editing. Nucleic Acids Res **44:**e34.

143. **Luo ML, Mullis AS, Leenay RT, Beisel CL.** 2015. Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. Nucleic Acids Res **43:**674-681.

144. **Vercoe RB, Chang JT, Dy RL, Taylor C, Gristwood T, Clulow JS, Richter C, Przybilski R, Pitman AR, Fineran PC.** 2013. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. PLoS Genet **9:**e1003454.

145. **Peng L, Pei J, Pang H, Guo Y, Lin L, Huang R.** 2014. Whole genome sequencing reveals a novel CRISPR system in industrial *Clostridium acetobutylicum*. J Ind Microbiol Biotechnol **41:**1677-1685.

146. **Brown SD, Nagaraju S, Utturkar S, De Tissera S, Segovia S, Mitchell W, Land ML, Dassanayake A, Kopke M.** 2014. Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant Clostridia. Biotechnol Biofuels **7:**40.

147. **Tomazetto G, Hahnke S, Koeck DE, Wibberg D, Maus I, Puhler A, Klocke M, Schluter A.** 2016. Complete genome analysis of *Clostridium bornimense* strain M2/40(T): A new acidogenic *Clostridium* species isolated from a mesophilic two-phase laboratory-scale biogas reactor. J Biotechnol **232:**38-49.

148. **Negahdaripour M, Nezafat N, Hajighahramani N, Rahmatabadi SS, Ghasemi Y.** 2017. Investigating CRISPR-Cas systems in *Clostridium botulinum* via bioinformatics tools. Infect Genet Evol **54:**355-373.

149. **Chylinski K, Makarova KS, Charpentier E, Koonin EV.** 2014. Classification and evolution of type II CRISPR-Cas systems. Nucleic Acids Res **42:**6091-6105.

150. **Rychener L, InAlbon S, Djordjevic SP, Chowdhury PR, Ziech RE, de Vargas AC, Frey J, Falquet L.** 2017. *Clostridium chauvoei*, an evolutionary dead-end pathogen. Front Microbiol **8:**1054.

151. **Thomas P, Semmler T, Eichhorn I, Lubke-Becker A, Werckenthin C, Abdel-Glil MY, Wieler LH, Neubauer H, Seyboldt C.** 2017. First report of two complete *Clostridium chauvoei* genome sequences and detailed *in silico* genome analysis. Infect Genet Evol **54:**287-298.

152. **Boudry P, Semenova E, Monot M, Datsenko KA, Lopatina A, Sekulovic O, Ospina-Bedoya M, Fortier LC, Severinov K, Dupuy B, Soutourina O.** 2015. Function of the CRISPR-Cas system of the human pathogen *Clostridium difficile*. MBio **6**.

153. **Cohen JE, Wang R, Shen RF, Wu WW, Keller JE.** 2017. Comparative pathogenomics of *Clostridium tetani*. PLoS One **12:**e0182909.

154. **Richter H, Zoephel J, Schermuly J, Maticzka D, Backofen R, Randau L.** 2012. Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. Nucleic Acids Res **40:**9887-9896.

155. **Lee J, Jang YS, Han MJ, Kim JY, Lee SY.** 2016. Deciphering *Clostridium tyrobutyricum* metabolism based on the whole-genome sequence and proteome analyses. MBio **7**.

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

157. **Theriot CM, Young VB.** 2015. Interactions between the gastrointestinal microbiome and *Clostridium difficile*. Annu Rev Microbiol **69:**445-461.

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

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

160. **Aktories K, Schwan C, Jank T.** 2017. *Clostridium difficile* toxin biology. Annu Rev Microbiol doi:10.1146/annurev-micro-090816-093458.

161. **Ng YK, Ehsaan M, Philip S, Collery MM, Janoir C, Collignon A, Cartman ST, Minton NP.** 2013. Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using *pyrE* alleles. PLoS One **8:**e56051.

162. **Dineen SS, Villapakkam AC, Nordman JT, Sonenshein AL.** 2007. Repression of *Clostridium difficile* toxin gene expression by CodY. Mol Microbiol **66:**206-219.

163. **O'Connor JR, Lyras D, Farrow KA, Adams V, Powell DR, Hinds J, Cheung JK, Rood JI.** 2006. Construction and analysis of chromosomal *Clostridium difficile* mutants. Mol Microbiol **61:**1335-1351.

164. **Cartman ST, Minton NP.** 2010. A mariner-based transposon system for *in vivo* random mutagenesis of *Clostridium difficile*. Appl Environ Microbiol **76:**1103-1109.

165. **Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, Fairweather NF, Fagan RP.** 2015. High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*. MBio **6:**e02383.

166. **Bouillaut L, Dubois T, Sonenshein AL, Dupuy B.** 2015. Integration of metabolism and virulence in *Clostridium difficile*. Res Microbiol **166:**375-383.

167. **Srivastava M, Mallard C, Barke T, Hancock LE, Self WT.** 2011. A selenium-dependent xanthine dehydrogenase triggers biofilm proliferation in *Enterococcus faecalis* through oxidant production. J Bacteriol **193:**1643-1652.

168. **Francis MB, Allen CA, Sorg JA.** 2015. Spore cortex hydrolysis precedes DPA release during *Clostridium difficile* spore germination. J Bacteriol doi:10.1128/JB.02575-14.

169. **Sorg JA, Dineen SS.** 2009. Laboratory maintenance of *Clostridium difficile*. Curr Protoc Microbiol **Chapter 9:**Unit9A 1.

170. **Dupuy B, Sonenshein AL.** 1998. Regulated transcription of *Clostridium difficile* toxin genes. Mol Microbiol **27:**107-120.

171. **Kirk JA, Fagan RP.** 2016. Heat shock increases conjugation efficiency in *Clostridium difficile*. Anaerobe **42:**1-5.

172. **Karlsson S, Burman LG, Akerlund T.** 1999. Suppression of toxin production in *Clostridium difficile* VPI 10463 by amino acids. Microbiology **145 (Pt 7):**1683-1693.

173. **Ho TD, Ellermeier CD.** 2011. PrsW is required for colonization, resistance to antimicrobial peptides, and expression of extracytoplasmic function sigma factors in *Clostridium difficile*. Infect Immun **79:**3229-3238.

174. **McBride SM, Sonenshein AL.** 2011. Identification of a genetic locus responsible for antimicrobial peptide resistance in *Clostridium difficile*. Infect Immun **79:**167-176.

175. **Bouillaut L, McBride SM, Sorg JA.** 2011. Genetic manipulation of *Clostridium difficile*. Curr Protoc Microbiol **Chapter 9:**Unit 9A 2.

176. **Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO.** 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods **6:**343-345.

177. **Bradford MM.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem **72:**248-254.

178. **Wren BW, Tabaqchali S.** 1987. Restriction endonuclease DNA analysis of *Clostridium difficile*. J Clin Microbiol **25:**2402-2404.

179. **Power DAZ, M.J.** 2003. Difco & BBL Manual: Manual of Microbiological Culture Media, *on* Beckton, Dickson and Company. http://galachem.ru/upload/iblock/c79/difcobblmanual_2nded.pdf. Accessed

180. **Shen B, Zhang W, Zhang J, Zhou J, Wang J, Chen L, Wang L, Hodgkins A, Iyer V, Huang X, Skarnes WC.** 2014. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nat Methods **11:**399-402.

181. **Mani N, Dupuy B.** 2001. Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. Proc Natl Acad Sci U S A **98:**5844-5849.

182. **Mani N, Dupuy B, Sonenshein AL.** 2006. Isolation of RNA polymerase from *Clostridium difficile* and characterization of glutamate dehydrogenase and rRNA gene promoters *in vitro* and *in vivo*. J Bacteriol **188:**96-102.

183. **Moreno-Mateos MA, Vejnar CE, Beaudoin JD, Fernandez JP, Mis EK, Khokha MK, Giraldez AJ.** 2015. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting *in vivo*. Nat Methods **12:**982-988.

184. **Oliveira Paiva AM, Friggen AH, Hossein-Javaheri S, Smits WK.** 2016. The signal sequence of the abundant extracellular metalloprotease PPEP-1 can be used to secrete synthetic reporter proteins in *Clostridium difficile*. ACS Synth Biol **5:**1376-1382.

185. **Mojica FJ, Diez-Villasenor C, Soria E, Juez G.** 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol Microbiol **36:**244-246.

186. **Kim IY, Veres Z, Stadtman TC.** 1992. *Escherichia coli* mutant SELD enzymes. The cysteine 17 residue is essential for selenophosphate formation from ATP and selenide. J Biol Chem **267:**19650-19654.

187. **Forchhammer K, Bock A.** 1991. Selenocysteine synthase from *Escherichia coli*. Analysis of the reaction sequence. J Biol Chem **266:**6324-6328.

188. **Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, Farley MM, Holzbauer SM, Meek JI, Phipps EC, Wilson LE, Winston LG, Cohen JA, Limbago BM, Fridkin SK, Gerding DN, McDonald LC.** 2015. Burden of *Clostridium difficile* infection in the United States. N Engl J Med **372:**825-834.

189. **Fimlaid KA, Bond JP, Schutz KC, Putnam EE, Leung JM, Lawley TD, Shen A.** 2013. Global analysis of the sporulation pathway of *Clostridium difficile*. PLoS Genet **9:**e1003660.

190. **Pinske C, Sawers RG.** 2016. Anaerobic formate and hydrogen metabolism. EcoSal Plus **7**.

191. **Shen A, Fimlaid KA, Pishdadian K.** 2016. Inducing and quantifying *Clostridium difficile* spore formation. Methods Mol Biol **1476:**129-142.

192. **Bhattacharjee D, McAllister KN, Sorg JA.** 2016. Germinants and their receptors in clostridia. J Bacteriol doi:10.1128/JB.00405-16.

193. **Francis MB, Sorg JA.** 2016. Detecting cortex fragments during bacterial spore germination. J Vis Exp doi:10.3791/54146.

194. **Schmittgen TD, Livak KJ.** 2008. Analyzing real-time PCR data by the comparative Ct method. Nat Protoc **3:**1101-1108.

195. **Sambol SP, Tang JK, Merrigan MM, Johnson S, Gerding DN.** 2001. Infection of hamsters with epidemiologically important strains of *Clostridium difficile*. J Infect Dis **183:**1760-1766.

196. **Hansen MC, Nielsen AK, Molin S, Hammer K, Kilstrup M.** 2001. Changes in rRNA levels during stress invalidates results from mRNA blotting: fluorescence *in situ* rRNA hybridization permits renormalization for estimation of cellular mRNA levels. J Bacteriol **183:**4747-4751.

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

198. **Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M.** 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res **44:**D457-462.

199. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

200. **Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH.** 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res **45:**D200-D203.

201. **Sekulovic O, Mathias Garrett E, Bourgeois J, Tamayo R, Shen A, Camilli A.** 2018. Genome-wide detection of conservative site-specific recombination in bacteria. PLoS Genet **14:**e1007332.

202. **Behrens S, Mitchell W, Bahl H.** 2001. Molecular analysis of the mannitol operon of *Clostridium acetobutylicum* encoding a phosphotransferase system and a putative PTS-modulated regulator. Microbiology **147:**75-86.

203. **Watanabe S, Hamano M, Kakeshita H, Bunai K, Tojo S, Yamaguchi H, Fujita Y, Wong SL, Yamane K.** 2003. Mannitol-1-phosphate dehydrogenase (MtlD) is required for mannitol and glucitol assimilation in *Bacillus subtilis*: possible cooperation of *mtl* and *gut* operons. J Bacteriol **185:**4816-4824.

204. **Zhang M, Gu L, Cheng C, Ma J, Xin F, Liu J, Wu H, Jiang M.** 2018. Recent advances in microbial production of mannitol: utilization of low-cost substrates, strain development and regulation strategies. World J Microbiol Biotechnol **34:**41.

205. **Byer T, Wang J, Zhang MG, Vather N, Blachman A, Visser B, Liu JM.** 2017. MtlR negatively regulates mannitol utilization by *Vibrio cholerae*. Microbiology doi:10.1099/mic.0.000559.

206. **Pye VE, Tingey AP, Robson RL, Moody PC.** 2004. The structure and mechanism of serine acetyltransferase from *Escherichia coli*. J Biol Chem **279:**40729-40736.

207. **Rabeh WM, Cook PF.** 2004. Structure and mechanism of O-acetylserine sulfhydrylase. J Biol Chem **279:**26803-26806.

208. **Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS.** 2002. Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. Nucleic Acids Res **30:**3141-3151.

209. **Mantsala P, Zalkin H.** 1984. Glutamine amidotransferase function. Replacement of the active-site cysteine in glutamine phosphoribosylpyrophosphate amidotransferase by site-directed mutagenesis. J Biol Chem **259:**14230-14236.

210. **Massiere F, Badet-Denisot MA.** 1998. The mechanism of glutamine-dependent amidotransferases. Cell Mol Life Sci **54:**205-222.

211. **Kieliszek M.** 2019. Selenium - fascinating microelement, properties and sources in food. Molecules **24**.

212. **Estevam EC, Witek K, Faulstich L, Nasim MJ, Latacz G, Dominguez-Alvarez E, Kiec-Kononowicz K, Demasi M, Handzlik J, Jacob C.** 2015. Aspects of a distinct cytotoxicity of selenium salts and organic selenides in living cells with possible implications for drug design. Molecules **20:**13894-13912.

213. **Pellissery AJ, Vinayamohan PG, Yin HB, Mooyottu S, Venkitanarayanan K.** 2019. *In vitro* efficacy of sodium selenite in reducing toxin production, spore outgrowth and antibiotic resistance in hypervirulent *Clostridium difficile*. J Med Microbiol doi:10.1099/jmm.0.001008.

214. **Stanley JD, Bartlett JG, Dart BWt, Ashcraft JH.** 2013. *Clostridium difficile* infection. Curr Probl Surg **50:**302-337.

215. **Setlow P.** 2014. Spore resistance properties. Microbiol Spectr **2:**001-003.

216. **Ridlon JM, Harris SC, Bhowmik S, Kang DJ, Hylemon PB.** 2016. Consequences of bile salt biotransformations by intestinal bacteria. Gut Microbes **7:**22-39.

217. **Wilson KH.** 1983. Efficiency of various bile salt preparations for stimulation of *Clostridium difficile* spore germination. J Clin Microbiol **18:**1017-1019.

218. **Francis MB, Allen CA, Sorg JA.** 2013. Muricholic acids inhibit *Clostridium difficile* spore germination and growth. PLoS One **8:**e73653.

219. **Savidge T, Sorg JA.** 2019. Role of bile in infectious disease: the gall of 7alpha-dehydroxylating gut bacteria. Cell Chem Biol **26:**1-3.

220. **Theriot CM, Bowman AA, Young VB.** 2016. Antibiotic-induced alterations of the gut microbiota alter secondary bile acid production and allow for *Clostridium difficile* spore germination and outgrowth in the large intestine. mSphere **1**.

221. **Edwards AN, Anjuwon-Foster BR, McBride SM.** 2019. RstA is a major regulator of *Clostridioides difficile* toxin production and motility. MBio **10**.

222. **Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, Bikard D, Marraffini LA.** 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature **519:**199-202.

223. **Modell JW, Jiang W, Marraffini LA.** 2017. CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. Nature **544:**101-104.

224. **Jackson-Rosario S, Cowart D, Myers A, Tarrien R, Levine RL, Scott RA, Self WT.** 2009. Auranofin disrupts selenium metabolism in *Clostridium difficile* by forming a stable Au-Se adduct. J Biol Inorg Chem **14:**507-519.

225. **Hanahan D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol **166:**557-580.

226. **Haraldsen JD, Sonenshein AL.** 2003. Efficient sporulation in *Clostridium difficile* requires disruption of the σK gene. Mol Microbiol **48:**811-821.

227. **Stabler RA, He M, Dawson L, Martin M, Valiente E, Corton C, Lawley TD, Sebaihia M, Quail MA, Rose G, Gerding DN, Gibert M, Popoff MR, Parkhill J, Dougan G, Wren BW.** 2009. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol **10:**R102.

228. **Heap JT, Pennington OJ, Cartman ST, Minton NP.** 2009. A modular system for *Clostridium* shuttle plasmids. J Microbiol Methods **78:**79-85.

APPENDIX A

SUPPLEMENTAL FIGURES

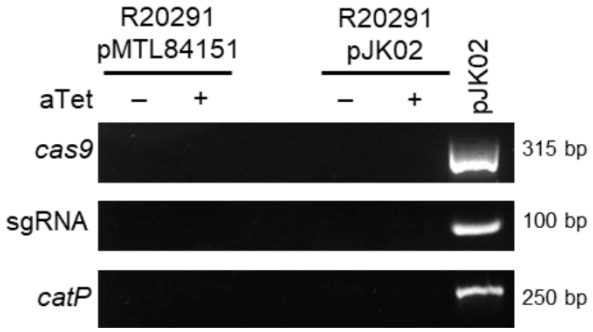


Figure 5.1 No DNA contamination in RNA samples of *C. difficile* R20291 vectors. RT-PCR without reverse transcriptase showing the comparison of *C. difficile* R20291 pMTL8151 and *C. difficile* R20291 pJK02 induced without aTet and those induced in the presence of aTet to turn on the expression of *cas9*. Also tested was the sgRNA and *catP*, as positive controls. The full, uncropped gels corresponding to this data can be found below.

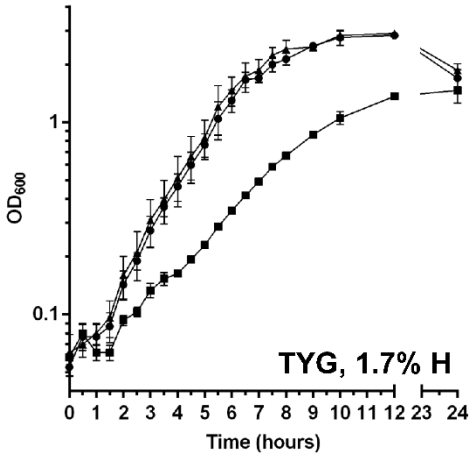


Figure 5.2 Growth of *C. difficile* strains in TY medium supplemented with glucose.
C. difficile R20291 (wild-type) (●), *C. difficile* KNM6 (Δ*selD*) (■), and *C. difficile* KNM9 (Δ*selD*::*selD*$^{+}$) (▲) were grown in TY medium supplemented with glucose and growth was monitored over a 24 hour period. Data points represent the average from three independent experiments and error bars represent the standard deviation from the mean.

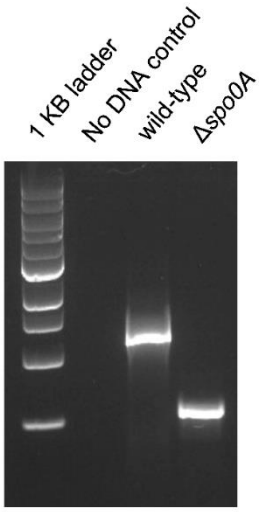


Figure 5.3 Deletion of *spo0A*.
DNA was isolated from *C. difficile* R20291 (wild-type) and *C. difficile* KNM10 (Δ*spo0A*). The region surrounding the *spo0A* gene was amplified from the chromosome, and the resulting DNA was separated on an agarose gel. A clean deletion of *spo0A* is indicated by a faster-migrating DNA band while wild-type is indicated by a slower-migrating DNA band.

APPENDIX B

SUPPLEMENTAL TABLES

Table 5.1 Strains and plasmids used in Section 2.

Strain	**Description/Phenotype**	**Source or Reference**
E. coli DH5α	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–	(225)
E. coli HB101 pRK24	*lavYI galK2 xyl-6 mtl-I repsL20* carrying pRK24	B. Dupuy
B. subtilis BS49	Tn*916* donor strain, Tet^R	(226)
C. difficile JIR8094	*erm*-sensitive derivative of 630	(163)
C. difficile LB-CD7	*selD* TargeTron mutant	This Study
C. difficile LB-CD12	*grdA* TargeTron mutant	(55)
C. difficile LB-CD4	*prdB* TargeTron mutant	(55)
C. difficile LB-CD8	*prdR* TargeTron mutant	(55)
C. difficile R20291	Wild type, ribotype 027	(227)
C. difficile KNM5	*pyrE* targeted CRISPR-*cas9* mutant, uracil auxotroph	This study
C. difficile KNM6	*selD* targeted CRISPR-*cas9* mutant	This study
Plasmids		
pCE240	Derivative of pJIR750ai (Sigma-Aldrich)	(173)
pBL38	*selD* TargeTron in pCE240	This study
pBL54	pMC123-containing group II intron from pBL38	This study
pJK02	*traJ* in pKM71	This study
pJS116	*B. subtilis-C. difficile* shuttle vector (pCD6 ColE1 *Tn916 oriT* Cm^R)	(34)

Table 5.1 Continued.

Plasmid	Description/Phenotype	Source or Reference
pJS170	*selD* homology region in pJK02	This study
pJS187	*selD*-targeted sgRNA in pJS170	This study
pJS194	*tn916* oriT in pJS187	This study
pKM22	*spoVAC*- targeted sgRNA in pJS116	This study
pKM46	codon-optimized wildtype *cas9* in pKM22	This study
pKM48	codon-optimized $cas9_{D10A}$ in pKM22	This study
pKM54	*gdh* promoter with sgRNA in pKM46	This study
pKM55	*gdh* promoter with sgRNA in pKM48	This study
pKM64	*pyrE* homology region in pKM54	This study
pKM65	*pyrE* homology region in pKM55	This study
pKM71	*pyrE*-targeted sgRNA in pKM64	This study
pKM72	*pyrE*-targeted sgRNA in pKM65	This study
pKM93	*traJ* in pKM72	This study
pKM142	*selD* with 500 bp upstream in pJS116	This study
pMC123	*C. difficile* shuttle vector	(174)
pMK-RQ-Bs-cas9	codon-optimized *cas9* for *C. difficile* in pMK-RQ-Bs	Invitrogen, This study
pMTL84151	*E. coli-C. difficile* shuttle vector (pCD6 ColE1 *traJ* Cm^R)	(228)
pRPF215	tetracycline-inducible P_{tet} promoter	R. Fagan (165)

Table 5.2 Oligonucleotides used in Section 2.

Oligonucleotide	Sequence
gRNA_1_for	AAATACGGTGTTTTTTGTTACCCTAAGTTTCATAAA AATAAGAAGCCTGCAAATGCAGGCTTCTTATTTTT ATGGTTTAAACCCGCATTATTAAA
gRNA_2_rev	CTTTTTCTTATATTTTATTTTTTTAATATTTATGTAT ACAAAAAAGTAACAAAATTGTAATTTTTTTTAATA ATGCGGGTTTAAACCATAAAAAT
gRNA_3_for	AATATTAAAAAAATAAAATATAAGAAAAAGTTTA TATCTTTTGGTTAATTATTACAATAAGTCTCATTTA TTGAAATAATATCAAATATATATTA
gRNA_4_rev	ATTTCTAGCTCTAAAACAATAAAGAAAAAATCTAT TTTTATTATTTTTTCCTTATTTACCAATTATAATATA TATTTGATATTATTTCAATAAAT
gRNA_5_for	GATTTTTTCTTTATTGTTTTAGAGCTAGAAATAGCA AGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTT
gRNA_6_rev	TAAAAATAGTTGCAGAGCTTTGTACACTAGTCAGA CATCATGCTGATCTAGATTTCTCCATAGAAAAAAA GCACCGACTCGGTGCCACTTTTTCA
gRNA_7_rev	ATTTTGGTCATGAGATTATCAAAAAGGAGTTTAAT AAAAAATAAAAATAAGCTCTGCAACCATCTAAAA ATAGTTGCAGAGCTTTGTACACTAGT
5’gRNA	AAATACGGTGTTTTTTGTTACCCTAAGTTTCATAAA AATAAGAAGCCTGCAAATGCAGGC
3’gRNA	TTGGTCATGAGATTATCAAAAAGGAGTTTAATAAA AAATAAAAATAAGCTCTGCAACCA
5’MTL_tetRprom	CCATGGAGATCTCGAGGCCTGCAGACATGCGCGAT CGCAGACCCACTTTCACATTTAAG
3’tetR_Cas9	ATCTAAGCCTATTGAGTATTTCTTATCCATTTAATT AAAAAAACCTCCTAGTATTATTGAGC
5’tetR_CO_cas9	CTGAGCTCAATAATACTAGGAGGTTTTTTTAATTA AATGGATAAAAAATATAGTATAGGATTAGATATA GGAAC
3’MTL_CO_cas9	TCACGACGTTGTAAAACGACGGCCAGTGCCAAGCT TTTAATCACCACCTAATTGAGATAAATCTAT
5’tetR_CO_cas9_D10A	CTGAGCTCAATAATACTAGGAGGTTTTTTTAATTA AATGGATAAAAAATATAGTATAGGATTAGCTATAG GAACAAATAG
5’gdh	TGCAGGCTTCTTATTTTTATGGTTTAAACGGTTTTA GCTGGGATATCG
3’gdh_gRNA	AAAACAATAAAGAAAAAATCTATTTGGTACCTTTA CAGTTTAATTATAGCACACTTTATT

Table 5.2 Continued.

Oligonucleotide	Sequence
5'gRNA_gdh	AGTGTGCTATAATTAAACTGTAAAGGTACCAAATA GATTTTTTCTTTATTGTTTTAG
3'gRNA 2	CATCTAAAAATAGTTGCAGAGCTTACGCGTCTAGT CAGACATCATGCTGA
5'pyrE_UP	TTATCAGGAAACAGCTATGACCGCGGCCGCGACGT GATTTTTAATGGGTA
3'pyrE_UP	TAAGTAACACTATAAATAATTAAGTTTTTAATTATT TTTCCTCCATGTTAATG
5'pyrE_DOWN	ATAACATTAACATGGAGGAAAAATAATTAAAAAC TTAATTATTTATAGTGTTACTTAAAA
3'pyrE_DOWN	GTGGGTCTGCGATCGCGCATGTCTGCAGGCCTCGA GAAGCATTGATGTTCTTCCT
pyrE_gRNA_gBlock	GGTACCGAAAAGTGATGCATTGTTGGGTTTTAGAG CTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTT ATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT TTTTCTATGGAGAAATCTAGATCAGCATGATGTCT GACTAGACGCGT
5'traJ	GCGAGGAAGCGGAAGAGCGCCCAATACGCAGGGC CCCCTGCTTCGGGGTCA
3'traJ	AATTTATCTACAATTTTTTTATCCTGCAGGGGGCCC GATCGGTCTTGCCTTG
5'MTL_selD_UP	TTATCAGGAAACAGCTATGACCGCGGCCGCATTAA TAAAAGTGATAAATTTCTTTTCAT
3'MTL_selD_UP	TCAAAACAATCACTCTTTCTCTATAATATTTTGAGA TAAAACTGATGCCA
5'selD_MTL_DN	CCAGAGGTTCTGGCATCAGTTTTATCTCAAAATAT TATAGAGAAAGAGTGATTGTTTT
3'selD_MTL_DN	CTGCGATCGCGCATGTCTGCAGGCCTCGAGTCTAG TAAGTTGATTTTTCTTCATTT
CRISPR_selD_183	GGTACCTGCTATGGGAGGCAAACCTTGTTTTAGAG CTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTT ATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTT TTTTCTATGGAGAAATCTAGATCAGCATGATGTCT GACTAGACGCGT
5’Tn916ori	AAGCGGAAGAGCGCCCAATACGCAGGGCCCTAAC ATCTTCTATTTTTCCCA
3’Tn916ori	TATCTACAATTTTTTTATCCTGCAGGGGGCCCCTAA AGGGAATGTAGATAAATTATTAG
5'selD_comp	CAATTTTTTTATCAGGAAACAGCTATGACCGCGGC CGCACCTAAAATAGGTGAAGCAAC

Table 5.2 Continued.

Oligonucleotide	Sequence
3'selD_comp	GCCAGTGCCAAGCTTGCATGTCTGCAGGCCTCGAG TTATAAAACTGTAATATATTTTTCC
5' tcdB	TTACATTTTGTTTGGATTGGAGGTC
3' tcdB	AGCAGCTAAATTCCACCTTTCTACC
5' catP 3	ATGGTATTTGAAAAAATTGATAAAAATAG
3' catP 2	TTAACTATTTATCAATTCCTGCAATTCG
5'pyrE 2	GTCCAGTGTTCTGGGGAG
3'pyrE 2	AAAATTTACATTTTTTAAGTAACACTATAAATAAT TAAGTTTTTA
5’selD	GAGCTTCCTAAAAATGAAGTAAATATCAATAAACA G
3’selD	TTTTGCTCAAAACAATCACTCTTTCTCTATAATATT
5’COcas9_RT	GTGTAGGATGGGCAGTAATAAC
3’COcas9_RT	CCAAATATTGGATGTCTTTCGTG
5'gRNA_RT	GTTTTAGAGCTAGAAATAGCAAG
3'gRNA_RT	CTGATCTAGATTTCTCCATAG
5'catP_RT	CTTTGCAAGTGTACCTTGTACC
3'catP_RT	GTCAGACTTACACTCAGTCC
oLB70	AAAAGCTTTTGCAACCCACGTCGATCGTGAAGAGC CTAAATATGTGCGCCCAGATAGGGTG
oLB71	CAGATTGTACAAATGTGGTGATAACAGATAAGTCA AATATGGTAACTTACCTTTCTTTGT
oLB72	CGCAAGTTTCTAATTTCGGTTGGCTCTCGATAGAG GAAAGTGTCT
EBS-universal	CGAAATTA GAAACTTGCGTTCAGTAAAC
oLB76	ATGACAACAGCAGGTGGTTGAGCCGC
oLB77	ATTTGTGCCATTTCTATTGCACC

Table 5.3 Complete list of gene expression fold-change from RNA-seq of wild-type and *selD* mutant strains.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
prdC	11.773 up	CDR20291_0963	70.544 up	CDR20291_r14	8.105 up	CDR20291_0284	7.541 up
CDR20291_3096	9.976 up	CDR20291_0962	44.982 up	CDR20291_r26	6.972 up	CDR20291_1726	4.566 up
CDR20291_3100	7.748 up	CDR20291_0440	17.707 up	CDR20291_r03	6.568 up	CDR20291_1445	3.950 up
CDR20291_3098	7.618 up	CDR20291_1747	12.712 up	CDR20291_r19	6.392 up	CDR20291_2807	3.777 up
CDR20291_3099	7.616 up	cysM	8.129 up	CDR20291_r08	6.320 up	CDR20291_2122	3.365 up
prdF	6.729 up	prdC	6.994 up	CDR20291_r21	6.108 up	CDR20291_1639	3.349 up
cysM	6.102 up	CDR20291_3096	5.791 up	CDR20291_r25	6.011 up	CDR20291_0739	3.289 up
CDR20291_0963	6.083 up	cysA	5.460 up	CDR20291_r11	5.061 up	CDR20291_2801	3.220 up
prdA	5.521 up	mtlF	4.901 up	nanA	4.724 up	CDR20291_3408	3.178 up
CDR20291_0962	4.683 up	mtlD	4.753 up	csrA	4.600 up	CDR20291_0020	3.133 up
prdB	4.477 up	CDR20291_r03	4.495 up	fliS1	4.501 up	CDR20291_2490	2.960 up
CDR20291_3102	4.454 up	CDR20291_r08	4.421 up	flgL	4.261 up	CDR20291_1823	2.928 up
cysA	4.213 up	CDR20291_r25	4.386 up	CDR20291_0271	4.194 up	CDR20291_1744	2.879 up
mtlR	3.412 up	CDR20291_r21	4.380 up	fliS2	4.157 up	CDR20291_t31	2.878 up
CDR20291_2627	3.260 up	CDR20291_1260	4.374 up	CDR20291_2138	3.800 up	CDR20291_2121	2.869 up
mtlA	3.154 up	CDR20291_r19	4.358 up	fliD	3.780 up	CDR20291_t30	2.856 up
CDR20291_2626	3.153 up	CDR20291_r26	4.213 up	CDR20291_2136	3.593 up	sspB	2.836 up
CDR20291_1689	2.974 up	CDR20291_r14	3.913 up	CDR20291_0234	3.570 up	CDR20291_0537	2.826 up
CDR20291_1260	2.950 up	mtlR	3.886 up	nanE	3.482 up	CDR20291_t32	2.822 up
serA	2.881 up	CDR20291_r11	3.462 up	flgK	3.477 up	CDR20291_3098	2.760 up
CDR20291_0979	2.795 up	CDR20291_1746	3.419 up	CDR20291_0243	3.461 up	CDR20291_1246	2.748 up
mtlD	2.774 up	CDR20291_2627	3.031 up	flgE	3.395 up	CDR20291_2182	2.742 up
CDR20291_0847	2.714 up	mtlA	2.984 up	CDR20291_2979	3.388 up	CDR20291_t29	2.732 up
mtlF	2.709 up	CDR20291_2626	2.978 up	CDR20291_0746	3.376 up	CDR20291_1241	2.715 up
CDR20291_0849	2.689 up	CDR20291_1783	2.810 up	grdE	3.351 up	prdF	2.682 up
CDR20291_2251	2.650 up	CDR20291_0474	2.623 up	CDR20291_0259	3.346 up	CDR20291_3099	2.678 up
CDR20291_0714	2.543 up	CDR20291_3098	2.600 up	flgG	3.333 up	CDR20291_t02	2.653 up

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
CDR20291_2081	2.513 up	CDR20291_3100	2.505 up	CDR20291_0239	3.277 up	fdxA	2.604 up
CDR20291_2080	2.505 up	ribH	2.430 up	flhF	3.275 up	CDR20291_3096	2.575 up
CDR20291_2078	2.484 up	prdF	2.377 up	CDR20291_0241	3.271 up	CDR20291_1760	2.559 up
CDR20291_2082	2.380 up	prdA	2.368 up	CDR20291_0242	3.253 up	CDR20291_t40	2.554 up
CDR20291_2875	2.340 up	CDR20291_1593	2.319 up	flhB	3.250 up	CDR20291_t56	2.544 up
CDR20291_2083	2.288 up	CDR20291_0512	2.286 up	flgM	3.236 up	CDR20291_1706	2.538 up
CDR20291_0636	2.224 up	CDR20291_t16	2.261 up	CDR20291_r27	3.230 up	CDR20291_1759	2.535 up
CDR20291_2079	2.198 up	CDR20291_1111	2.258 up	CDR20291_r09	3.220 up	CDR20291_t38	2.535 up
CDR20291_0764	2.192 up	CDR20291_t53	2.249 up	flgD	3.172 up	CDR20291_1204	2.531 up
CDR20291_0191	2.192 up	CDR20291_3099	2.248 up	flhA	3.169 up	CDR20291_0879	2.517 up
nrdE	2.179 up	ribD	2.232 up	CDR20291_r17	3.168 up	CDR20291_2517	2.514 up
fdhA	2.131 up	cat1	2.217 up	CDR20291_r01	3.149 up	CDR20291_0340	2.511 up
CDR20291_1931	2.115 up	aksA	2.200 up	fliP	3.144 up	CDR20291_1949	2.489 up
CDR20291_0707	2.113 up	CDR20291_0777	2.193 up	flgB	3.140 up	CDR20291_1780	2.479 up
gapN	2.090 up	fdhA	2.185 up	CDR20291_r06	3.140 up	CDR20291_0191	2.446 up
glyA	2.061 up	modA	2.176 up	fleN	3.133 up	CDR20291_0214	2.443 up
CDR20291_2164	2.041 up	CDR20291_0182	2.171 up	flgC	3.128 up	CDR20291_1753	2.440 up
sucD	2.032 up	CDR20291_2440	2.169 up	CDR20291_r12	3.125 up	CDR20291_t07	2.424 up
CDR20291_0763	2.018 up	CDR20291_2397	2.167 up	CDR20291_r16	3.124 up	CDR20291_t24	2.413 up
CDR20291_1624	2.014 up	CDR20291_0146	2.153 up	fliA	3.121 up	CDR20291_2272	2.394 up
CDR20291_2868	2.006 up	CDR20291_0178	2.152 up	CDR20291_r15	3.109 up	CDR20291_1368	2.383 up
CDR20291_2963	2.000 down	CDR20291_2191	2.145 up	CDR20291_3387	3.099 up	CDR20291_1813	2.380 up
srlB	2.007 down	CDR20291_1230	2.124 up	CDR20291_r04	3.075 up	CDR20291_2350	2.366 up
CDR20291_3463	2.008 down	CDR20291_1446	2.124 up	flgG	3.060 up	prdB	2.360 up
pduQ	2.009 down	CDR20291_0283	2.118 up	motB	3.050 up	CDR20291_t58	2.358 up
CDR20291_2932	2.011 down	CDR20291_0191	2.108 up	CDR20291_3282	3.014 up	prdC	2.352 up
CDR20291_0549	2.012 down	CDR20291_3471	2.107 up	CDR20291_3267	2.981 up	CDR20291_1752	2.347 up
fhuB	2.014 down	CDR20291_2012	2.100 up	fliK	2.978 up	CDR20291_1750	2.335 up

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
rbsR	2.014 down	CDR20291_1929	2.092 up	CDR20291_r18	2.958 up	CDR20291_1223	2.332 up
clpB	2.017 down	CDR20291_1405	2.089 up	CDR20291_0263	2.929 up	CDR20291_3012	2.322 up
CDR20291_2850	2.020 down	ribB	2.083 up	fliI	2.909 up	CDR20291_0347	2.321 up
CDR20291_t08	2.022 down	CDR20291_0763	2.073 up	CDR20291_r20	2.904 up	CDR20291_2509	2.317 up
CDR20291_2450	2.023 down	glyA	2.069 up	CDR20291_r13	2.901 up	CDR20291_2907	2.313 up
gatA	2.024 down	CDR20291_0947	2.068 up	CDR20291_0231	2.894 up	CDR20291_3100	2.308 up
CDR20291_0392	2.037 down	CDR20291_0418	2.064 up	fliF	2.883 up	CDR20291_0775	2.302 up
CDR20291_1437	2.043 down	CDR20291_1476	2.054 up	CDR20291_r10	2.880 up	CDR20291_3189	2.301 up
CDR20291_1460	2.044 down	CDR20291_1075	2.048 up	CDR20291_r05	2.876 up	CDR20291_t08	2.300 up
nagA	2.045 down	CDR20291_t39	2.037 up	CDR20291_r22	2.875 up	CDR20291_1779	2.293 up
oraE	2.049 down	CDR20291_3474	2.035 up	CDR20291_1230	2.869 up	CDR20291_3480	2.292 up
CDR20291_1681	2.050 down	CDR20291_0778	2.032 up	fliH	2.865 up	CDR20291_1749	2.281 up
CDR20291_1934	2.050 down	CDR20291_1233	2.028 up	CDR20291_r23	2.860 up	CDR20291_2651	2.270 up
CDR20291_0782	2.054 down	iorB	2.027 up	fliC	2.852 up	CDR20291_1592	2.267 up
CDR20291_1790	2.055 down	asnB	2.012 up	fliM	2.844 up	CDR20291_1764	2.264 up
bclA3	2.060 down	CDR20291_3109	2.010 up	CDR20291_r07	2.832 up	CDR20291_t33	2.261 up
CDR20291_0594	2.065 down	CDR20291_1816	2.007 up	CDR20291_r02	2.825 up	CDR20291_2273	2.256 up
CDR20291_0781	2.070 down	CDR20291_1745	2.002 up	CDR20291_r24	2.823 up	CDR20291_1778	2.256 up
leuB	2.071 down	CDR20291_0765	2.001 up	CDR20291_3080	2.815 up	CDR20291_1771	2.249 up
CDR20291_0597	2.078 down	CDR20291_3034	2.035 down	fliQ	2.808 up	CDR20291_3279	2.241 up
CDR20291_0572	2.085 down	nfo	2.042 down	motA	2.805 up	CDR20291_2617	2.231 up
CDR20291_2300	2.091 down	sigE	2.071 down	grdC	2.775 up	CDR20291_2775	2.228 up
CDR20291_2289	2.091 down	CDR20291_1819	2.090 down	CDR20291_0135	2.765 up	CDR20291_3113	2.223 up
CDR20291_0733	2.094 down	CDR20291_1478	2.111 down	CDR20291_0226	2.718 up	CDR20291_0738	2.213 up
CDR20291_1187	2.098 down	CDR20291_0177	2.112 down	ribH	2.713 up	CDR20291_0960	2.206 up
CDR20291_t15	2.099 down	CDR20291_3153	2.114 down	nagA	2.697 up	CDR20291_2011	2.205 up
CDR20291_0924	2.100 down	CDR20291_2514	2.116 down	fliJ	2.684 up	CDR20291_0283	2.203 up
CDR20291_1782	2.101 down	CDR20291_0340	2.191 down	CDR20291_0245	2.669 up	CDR20291_t51	2.197 up

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
CDR20291_2174	2.103 down	CDR20291_1432	2.208 down	CDR20291_2049	2.647 up	CDR20291_1775	2.192 up
CDR20291_t40	2.104 down	CDR20291_1319	2.257 down	fliN	2.644 up	acpP	2.185 up
CDR20291_0319	2.105 down	CDR20291_1685	2.274 down	fliL	2.626 up	CDR20291_1308	2.177 up
tagT	2.106 down	CDR20291_0176	2.293 down	CDR20291_1911	2.620 up	CDR20291_3108	2.176 up
CDR20291_0807	2.107 down	CDR20291_0175	2.333 down	CDR20291_0225	2.617 up	hpt	2.172 up
CDR20291_3387	2.109 down	CDR20291_2875	2.424 down	CDR20291_2050	2.616 up	CDR20291_3094	2.165 up
CDR20291_2595	2.115 down	CDR20291_1441	2.452 down	gapA	2.605 up	CDR20291_0511	2.160 up
CDR20291_t65	2.117 down	CDR20291_2931	2.599 down	fliE	2.594 up	CDR20291_3188	2.152 up
thiG	2.120 down	sspB	2.627 down	CDR20291_1453	2.574 up	CDR20291_3110	2.148 up
CDR20291_1428	2.127 down	CDR20291_3515	2.709 down	CDR20291_2048	2.524 up	CDR20291_1772	2.146 up
CDR20291_1431	2.135 down	CDR20291_2137	2.781 down	CDR20291_3404	2.504 up	CDR20291_1006	2.141 up
CDR20291_3413	2.136 down	CDR20291_0907	2.922 down	opuCA	2.477 up	CDR20291_1514	2.139 up
CDR20291_3287	2.145 down	CDR20291_2077	3.005 down	CDR20291_0228	2.468 up	CDR20291_t43	2.134 up
abgT	2.150 down	CDR20291_2932	3.055 down	CDR20291_2323	2.467 up	CDR20291_3187A	2.133 up
srlE	2.163 down	CDR20291_2801	3.128 down	CDR20291_0865	2.457 up	CDR20291_2941	2.131 up
CDR20291_2919	2.172 down	nanE	3.292 down	CDR20291_0224	2.427 up	CDR20291_1953	2.123 up
CDR20291_1911	2.178 down	nanA	4.753 down	CDR20291_3236	2.393 up	CDR20291_t06	2.117 up
CDR20291_1191	2.186 down	CDR20291_2138	5.302 down	CDR20291_0227	2.381 up	CDR20291_1658	2.113 up
CDR20291_3504	2.187 down	selD	49.582 down	CDR20291_2051	2.378 up	CDR20291_0920	2.113 up
cspA	2.190 down			CDR20291_3239	2.356 up	CDR20291_t59	2.106 up
phnI	2.197 down			CDR20291_2381	2.347 up	phnA	2.105 up
CDR20291_0030	2.200 down			CDR20291_1791	2.343 up	CDR20291_1917	2.101 up
CDR20291_3511	2.201 down			CDR20291_2516	2.325 up	CDR20291_1748	2.100 up
sodA	2.218 down			CDR20291_0276	2.299 up	CDR20291_0928	2.097 up
CDR20291_0292	2.219 down			CDR20291_t53	2.267 up	prdA	2.095 up
CDR20291_1545	2.228 down			ribA	2.259 up	CDR20291_1924	2.085 up
CDR20291_1389	2.239 down			CDR20291_0247	2.257 up	CDR20291_t15	2.069 up
CDR20291_2281	2.242 down			CDR20291_1130	2.257 up	CDR20291_1624	2.067 up

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
cspC	2.249 down			CDR20291_0600	2.250 up	CDR20291_1952	2.065 up
CDR20291_0606	2.254 down			CDR20291_3362	2.231 up	CDR20291_3191	2.063 up
eutM	2.259 down			CDR20291_2075	2.231 up	glvR	2.059 up
CDR20291_0790	2.259 down			clpB	2.223 up	CDR20291_1527	2.058 up
CDR20291_t59	2.266 down			CDR20291_0319	2.217 up	CDR20291_2286	2.057 up
CDR20291_3282	2.266 down			vanG	2.201 up	CDR20291_0587	2.053 up
CDR20291_1791	2.269 down			dacF	2.193 up	CDR20291_1751	2.053 up
CDR20291_2891	2.270 down			CDR20291_0244	2.172 up	CDR20291_3064	2.049 up
CDR20291_2818	2.288 down			CDR20291_2492	2.171 up	CDR20291_2961	2.040 up
CDR20291_1459	2.295 down			CDR20291_0179	2.164 up	CDR20291_2592	2.034 up
CDR20291_0841	2.299 down			CDR20291_0137	2.152 up	CDR20291_1271	2.029 up
CDR20291_1461	2.302 down			CDR20291_0176	2.131 up	CDR20291_t03	2.029 up
CDR20291_1472	2.306 down			CDR20291_3301	2.123 up	CDR20291_2396	2.029 up
CDR20291_0426	2.313 down			grdG	2.111 up	rpsT	2.028 up
araD	2.316 down			CDR20291_2978	2.111 up	CDR20291_2194	2.026 up
CDR20291_3474	2.327 down			CDR20291_3240	2.094 up	CDR20291_0847	2.026 up
cspB	2.332 down			CDR20291_2732	2.081 up	CDR20291_1622	2.020 up
gapA	2.337 down			CDR20291_2937	2.080 up	CDR20291_2453	2.015 up
CDR20291_0876	2.344 down			gutA	2.063 up	CDR20291_0713	2.008 up
CDR20291_1026	2.354 down			CDR20291_2200	2.058 up	CDR20291_2195	2.000 up
CDR20291_1905	2.361 down			CDR20291_2938	2.047 up	bioY	2.009 down
CDR20291_2514	2.366 down			srlE'	2.041 up	CDR20291_0548	2.019 down
CDR20291_2318	2.371 down			fliN	2.041 up	fliN	2.041 down
tagK	2.374 down			CDR20291_0548	2.019 up	srlE'	2.041 down
comE	2.375 down			bioY	2.009 up	CDR20291_2938	2.047 down
CDR20291_3473	2.385 down			CDR20291_2195	2.000 down	CDR20291_2200	2.058 down
ydiB	2.397 down			CDR20291_0713	2.008 down	gutA	2.063 down
CDR20291_2709	2.397 down			CDR20291_2453	2.015 down	CDR20291_2937	2.080 down

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
asrA	2.400 down			CDR20291_1622	2.020 down	CDR20291_2732	2.081 down
phnG	2.400 down			CDR20291_0847	2.026 down	CDR20291_3240	2.094 down
CDR20291_1382	2.401 down			CDR20291_2194	2.026 down	CDR20291_2978	2.111 down
CDR20291_2753	2.401 down			rpsT	2.028 down	grdG	2.111 down
CDR20291_2978	2.405 down			CDR20291_2396	2.029 down	CDR20291_3301	2.123 down
CDR20291_0039	2.413 down			CDR20291_t03	2.029 down	CDR20291_0176	2.131 down
CDR20291_1962	2.414 down			CDR20291_1271	2.029 down	CDR20291_0137	2.152 down
proS	2.426 down			CDR20291_2592	2.034 down	CDR20291_0179	2.164 down
CDR20291_t19	2.436 down			CDR20291_2961	2.040 down	CDR20291_2492	2.171 down
CDR20291_1863	2.441 down			CDR20291_3064	2.049 down	CDR20291_0244	2.172 down
CDR20291_1334	2.443 down			CDR20291_1751	2.053 down	dacF	2.193 down
CDR20291_2883	2.450 down			CDR20291_0587	2.053 down	vanG	2.201 down
CDR20291_0163	2.452 down			CDR20291_2286	2.057 down	CDR20291_0319	2.217 down
CDR20291_2545	2.464 down			CDR20291_1527	2.058 down	clpB	2.223 down
srlE'	2.468 down			glvR	2.059 down	CDR20291_2075	2.231 down
CDR20291_0385	2.470 down			CDR20291_3191	2.063 down	CDR20291_3362	2.231 down
CDR20291_1948	2.474 down			CDR20291_1952	2.065 down	CDR20291_0600	2.250 down
CDR20291_3184	2.482 down			CDR20291_1624	2.067 down	CDR20291_1130	2.257 down
CDR20291_1203	2.487 down			CDR20291_t15	2.069 down	CDR20291_0247	2.257 down
leuC	2.493 down			CDR20291_1924	2.085 down	ribA	2.259 down
CDR20291_0728	2.495 down			prdA	2.095 down	CDR20291_t53	2.267 down
CDR20291_2940	2.498 down			CDR20291_0928	2.097 down	CDR20291_0276	2.299 down
CDR20291_0922	2.505 down			CDR20291_1748	2.100 down	CDR20291_2516	2.325 down
CDR20291_0289	2.505 down			CDR20291_1917	2.101 down	CDR20291_1791	2.343 down
CDR20291_2452	2.520 down			phnA	2.105 down	CDR20291_2381	2.347 down
feoA1	2.521 down			CDR20291_t59	2.106 down	CDR20291_3239	2.356 down
CDR20291_t45	2.525 down			CDR20291_0920	2.113 down	CDR20291_2051	2.378 down
CDR20291_2999	2.531 down			CDR20291_1658	2.113 down	CDR20291_0227	2.381 down

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
CDR20291_2204	2.531 down			CDR20291_t06	2.117 down	CDR20291_3236	2.393 down
CDR20291_0386	2.536 down			CDR20291_1953	2.123 down	CDR20291_0224	2.427 down
CDR20291_0287	2.542 down			CDR20291_2941	2.131 down	CDR20291_0865	2.457 down
CDR20291_2186	2.545 down			CDR20291_3187A	2.133 down	CDR20291_2323	2.467 down
CDR20291_3454	2.545 down			CDR20291_t43	2.134 down	CDR20291_0228	2.468 down
CDR20291_2105	2.561 down			CDR20291_1514	2.139 down	opuCA	2.477 down
CDR20291_1838	2.561 down			CDR20291_1006	2.141 down	CDR20291_3404	2.504 down
leuD	2.561 down			CDR20291_1772	2.146 down	CDR20291_2048	2.524 down
CDR20291_0288	2.577 down			CDR20291_3110	2.148 down	CDR20291_1453	2.574 down
spoIIIAC	2.579 down			CDR20291_3188	2.152 down	fliE	2.594 down
CDR20291_2480	2.593 down			CDR20291_0511	2.160 down	gapA	2.605 down
CDR20291_2575	2.601 down			CDR20291_3094	2.165 down	CDR20291_2050	2.616 down
CDR20291_2597	2.601 down			hpt	2.172 down	CDR20291_0225	2.617 down
CDR20291_t41	2.601 down			CDR20291_3108	2.176 down	CDR20291_1911	2.620 down
CDR20291_2193	2.601 down			CDR20291_1308	2.177 down	fliL	2.626 down
CDR20291_2307	2.601 down			acpP	2.185 down	fliN	2.644 down
CDR20291_2190	2.601 down			CDR20291_1775	2.192 down	CDR20291_2049	2.647 down
CDR20291_2189	2.601 down			CDR20291_t51	2.197 down	CDR20291_0245	2.669 down
asrB	2.601 down			CDR20291_0283	2.203 down	fliJ	2.684 down
CDR20291_2180	2.601 down			CDR20291_2011	2.205 down	nagA	2.697 down
CDR20291_2181	2.601 down			CDR20291_0960	2.206 down	ribH	2.713 down
CDR20291_2306	2.601 down			CDR20291_0738	2.213 down	CDR20291_0226	2.718 down
srlE	2.601 down			CDR20291_3113	2.223 down	CDR20291_0135	2.765 down
CDR20291_2574	2.601 down			CDR20291_2775	2.228 down	grdC	2.775 down
srlA	2.601 down			CDR20291_2617	2.231 down	motA	2.805 down
thlA2	2.601 down			CDR20291_3279	2.241 down	fliQ	2.808 down
CDR20291_2315	2.601 down			CDR20291_1771	2.249 down	CDR20291_3080	2.815 down
CDR20291_2316	2.601 down			CDR20291_1778	2.256 down	CDR20291_r24	2.823 down

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
ctfA	2.601 down			CDR20291_2273	2.256 down	CDR20291_r02	2.825 down
CDR20291_2317	2.601 down			CDR20291_t33	2.261 down	CDR20291_r07	2.832 down
CDR20291_2321	2.601 down			CDR20291_1764	2.264 down	fliM	2.844 down
CDR20291_2335	2.601 down			CDR20291_1592	2.267 down	fliC	2.852 down
ctfB	2.601 down			CDR20291_2651	2.270 down	CDR20291_r23	2.860 down
gpr	2.601 down			CDR20291_1749	2.281 down	fliH	2.865 down
CDR20291_0021	2.601 down			CDR20291_3480	2.292 down	CDR20291_1230	2.869 down
CDR20291_0034	2.601 down			CDR20291_1779	2.293 down	CDR20291_r22	2.875 down
CDR20291_0033	2.601 down			CDR20291_t08	2.300 down	CDR20291_r05	2.876 down
asrC	2.601 down			CDR20291_3189	2.301 down	CDR20291_r10	2.880 down
CDR20291_3277	2.601 down			CDR20291_0775	2.302 down	fliF	2.883 down
CDR20291_1606	2.601 down			CDR20291_3100	2.308 down	CDR20291_0231	2.894 down
CDR20291_0950	2.601 down			CDR20291_2907	2.313 down	CDR20291_r13	2.901 down
CDR20291_2920	2.601 down			CDR20291_2509	2.317 down	CDR20291_r20	2.904 down
CDR20291_2108	2.601 down			CDR20291_0347	2.321 down	fliI	2.909 down
gatY	2.601 down			CDR20291_3012	2.322 down	CDR20291_0263	2.929 down
gltC	2.601 down			CDR20291_1223	2.332 down	CDR20291_r18	2.958 down
glvC	2.601 down			CDR20291_1750	2.335 down	fliK	2.978 down
CDR20291_2893	2.601 down			CDR20291_1752	2.347 down	CDR20291_3267	2.981 down
CDR20291_2882	2.601 down			prdC	2.352 down	CDR20291_3282	3.014 down
CDR20291_1738	2.601 down			CDR20291_t58	2.358 down	motB	3.050 down
CDR20291_1742	2.601 down			prdB	2.360 down	flgG	3.060 down
CDR20291_1745	2.601 down			CDR20291_2350	2.366 down	CDR20291_r04	3.075 down
CDR20291_1754	2.601 down			CDR20291_1813	2.380 down	CDR20291_3387	3.099 down
rbsA	2.601 down			CDR20291_1368	2.383 down	CDR20291_r15	3.109 down
CDR20291_1755	2.601 down			CDR20291_2272	2.394 down	fliA	3.121 down
CDR20291_1756	2.601 down			CDR20291_t24	2.413 down	CDR20291_r16	3.124 down
CDR20291_1757	2.601 down			CDR20291_t07	2.424 down	CDR20291_r12	3.125 down

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
CDR20291_1761	2.601 down			CDR20291_1753	2.440 down	flgC	3.128 down
CDR20291_1762	2.601 down			CDR20291_0214	2.443 down	fleN	3.133 down
CDR20291_1763	2.601 down			CDR20291_0191	2.446 down	CDR20291_r06	3.140 down
CDR20291_1765	2.601 down			CDR20291_1780	2.479 down	flgB	3.140 down
CDR20291_1767	2.601 down			CDR20291_1949	2.489 down	fliP	3.144 down
CDR20291_1768	2.601 down			CDR20291_0340	2.511 down	CDR20291_r01	3.149 down
CDR20291_1769	2.601 down			CDR20291_2517	2.514 down	CDR20291_r17	3.168 down
d4	2.601 down			CDR20291_0879	2.517 down	flhA	3.169 down
spoIIIAA	2.601 down			CDR20291_1204	2.531 down	flgD	3.172 down
CDR20291_1025	2.601 down			CDR20291_t38	2.535 down	CDR20291_r09	3.220 down
CDR20291_0571	2.601 down			CDR20291_1759	2.535 down	CDR20291_r27	3.230 down
CDR20291_3123	2.601 down			CDR20291_1706	2.538 down	flgM	3.236 down
CDR20291_1784	2.601 down			CDR20291_t56	2.544 down	flhB	3.250 down
CDR20291_1786	2.601 down			CDR20291_t40	2.554 down	CDR20291_0242	3.253 down
CDR20291_2921	2.601 down			CDR20291_1760	2.559 down	CDR20291_0241	3.271 down
CDR20291_2922	2.601 down			CDR20291_3096	2.575 down	flhF	3.275 down
CDR20291_0387	2.601 down			fdxA	2.604 down	CDR20291_0239	3.277 down
spoIIIAB	2.601 down			CDR20291_t02	2.653 down	flgG	3.333 down
CDR20291_3061	2.601 down			CDR20291_3099	2.678 down	CDR20291_0259	3.346 down
CDR20291_3060	2.601 down			prdF	2.682 down	grdE	3.351 down
CDR20291_1099	2.601 down			CDR20291_1241	2.715 down	CDR20291_0746	3.376 down
CDR20291_0544	2.601 down			CDR20291_t29	2.732 down	CDR20291_2979	3.388 down
CDR20291_1235	2.601 down			CDR20291_2182	2.742 down	flgE	3.395 down
CDR20291_1236	2.601 down			CDR20291_1246	2.748 down	CDR20291_0243	3.461 down
CDR20291_0536	2.601 down			CDR20291_3098	2.760 down	flgK	3.477 down
CDR20291_1279	2.601 down			CDR20291_t32	2.822 down	nanE	3.482 down
phnH	2.601 down			CDR20291_0537	2.826 down	CDR20291_0234	3.570 down
tcdD	2.601 down			sspB	2.836 down	CDR20291_2136	3.593 down

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
CDR20291_2977	2.601 down			CDR20291_t30	2.856 down	fliD	3.780 down
CDR20291_3116	2.601 down			CDR20291_2121	2.869 down	CDR20291_2138	3.800 down
spoIIIAE	2.601 down			CDR20291_t31	2.878 down	fliS2	4.157 down
sleB	2.601 down			CDR20291_1744	2.879 down	CDR20291_0271	4.194 down
oraS	2.601 down			CDR20291_1823	2.928 down	flgL	4.261 down
sleC	2.601 down			CDR20291_2490	2.960 down	fliS1	4.501 down
d2	2.601 down			CDR20291_0020	3.133 down	csrA	4.600 down
d3	2.601 down			CDR20291_3408	3.178 down	nanA	4.724 down
hom2	2.601 down			CDR20291_2801	3.220 down	CDR20291_r11	5.061 down
CDR20291_0433	2.601 down			CDR20291_0739	3.289 down	CDR20291_r25	6.011 down
CDR20291_2953	2.601 down			CDR20291_1639	3.349 down	CDR20291_r21	6.108 down
CDR20291_1506	2.601 down			CDR20291_2122	3.365 down	CDR20291_r08	6.320 down
CDR20291_2939	2.601 down			CDR20291_2807	3.777 down	CDR20291_r19	6.392 down
CDR20291_0409	2.601 down			CDR20291_1445	3.950 down	CDR20291_r03	6.568 down
CDR20291_1546	2.601 down			CDR20291_1726	4.566 down	CDR20291_r26	6.972 down
CDR20291_0398	2.601 down			CDR20291_0284	7.541 down	CDR20291_r14	8.105 down
CDR20291_0391	2.601 down						
CDR20291_0390	2.601 down						
CDR20291_1787	2.601 down						
CDR20291_1776	2.601 down						
zupT	2.601 down						
CDR20291_0732	2.601 down						
CDR20291_1928	2.601 down						
CDR20291_0832	2.601 down						
CDR20291_0921	2.601 down						
CDR20291_0206	2.601 down						
CDR20291_3464	2.601 down						
CDR20291_0791	2.601 down						

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
CDR20291_0789	2.601 down						
CDR20291_3465	2.601 down						
CDR20291_3466	2.601 down						
CDR20291_3467	2.601 down						
argB	2.601 down						
argJ	2.601 down						
argC	2.601 down						
CDR20291_0868	2.601 down						
CDR20291_2729	2.601 down						
CDR20291_2727	2.601 down						
garR	2.601 down						
CDR20291_0158	2.601 down						
CDR20291_2698	2.601 down						
CDR20291_2066	2.601 down						
crt1	2.601 down						
CDR20291_2106	2.601 down						
CDR20291_0729	2.601 down						
CDR20291_0723	2.601 down						
CDR20291_2107	2.601 down						
CDR20291_0130	2.601 down						
leuA	2.601 down						
CDR20291_0207	2.601 down						
CDR20291_0210	2.601 down						
CDR20291_2786	2.601 down						
CDR20291_1837	2.601 down						
eutC	2.601 down						
eutA	2.601 down						
CDR20291_1827	2.601 down						
CDR20291_1826	2.601 down						
CDR20291_3453	2.601 down						

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
CDR20291_0875	2.601 down						
CDR20291_1842	2.601 down						
CDR20291_3194	2.601 down						
CDR20291_1811	2.601 down						
CDR20291_1796	2.601 down						
CDR20291_1793	2.601 down						
CDR20291_0884	2.601 down						
CDR20291_1792	2.601 down						
eutT	2.601 down						
eutB	2.601 down						
CDR20291_1137	2.601 down						
CDR20291_1848	2.601 down						
ispD	2.601 down						
CDR20291_3459	2.601 down						
CDR20291_1851	2.601 down						
CDR20291_1913	2.601 down						
CDR20291_1844	2.601 down						
CDR20291_t21	2.611 down						
sspB	2.629 down						
CDR20291_2491	2.640 down						
CDR20291_1935	2.677 down						
CDR20291_2515	2.681 down						
CDR20291_2492	2.688 down						
CDR20291_1452	2.706 down						
CDR20291_t64	2.740 down						
CDR20291_3003	2.763 down						
CDR20291_3002	2.854 down						
CDR20291_0865	2.856 down						
CDR20291_2516	2.863 down						

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
CDR20291_t62	2.913 down						
CDR20291_1558	2.926 down						
CDR20291_2642	2.944 down						
CDR20291_1458	2.969 down						
CDR20291_1514	3.066 down						
CDR20291_1554	3.079 down						
CDR20291_2598	3.110 down						
CDR20291_t20	3.116 down						
spoIIID	3.248 down						
CDR20291_t49	3.251 down						
gutA	3.259 down						
CDR20291_t22	3.312 down						
CDR20291_1453	3.406 down						
CDR20291_1326	3.410 down						
trxB3	3.410 down						
CDR20291_0550	3.442 down						
CDR20291_t42	3.476 down						
CDR20291_t57	3.605 down						
grdX	3.646 down						
CDR20291_2979	3.682 down						
grdD	4.267 down						
CDR20291_2077	4.276 down						
CDR20291_t16	4.334 down						
trxA2	5.383 down						
CDR20291_2137	5.434 down						
grdA	5.710 down						
CDR20291_2136	6.295 down						
CDR20291_t39	7.956 down						
nanE	10.629 down						
grdE	11.248 down						
nanA	11.281 down						

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
grdC	11.337 down						
CDR20291_2138	12.087 down						
CDR20291_r10	13.978 down						
CDR20291_r24	14.209 down						
CDR20291_r02	14.220 down						
CDR20291_r22	14.286 down						
CDR20291_r07	14.424 down						
CDR20291_r23	14.881 down						
CDR20291_r13	14.949 down						
CDR20291_r20	15.184 down						
CDR20291_r18	15.198 down						
CDR20291_t53	24.184 down						
CDR20291_r04	42.022 down						
CDR20291_r15	43.193 down						
CDR20291_r16	43.444 down						
CDR20291_r06	44.122 down						
CDR20291_r12	44.317 down						
CDR20291_r01	44.485 down						
CDR20291_r27	44.945 down						
CDR20291_r09	45.596 down						
CDR20291_r17	46.213 down						
CDR20291_r03	70.307 down						
CDR20291_r08	92.221 down						
selD	97.359 down						
CDR20291_r21	108.714 down						
CDR20291_r25	116.858 down						
CDR20291_r19	118.412 down						
CDR20291_r26	135.282 down						

Table 5.3 Continued.

Name	Fold change from control 4H R20291 to experiment 4H KNM6	Name	Fold change from control 1.7H R20291 to experiment 1.7H KNM6	Name	Fold change from control 1.7H R20291 to experiment 4H R20291	Name	Fold change from control 4H R20291 to experiment 1.7H R20291
CDR20291_r11	149.688 down						
CDR20291_r14	153.222 down						

Table 5.4 Strains and plasmids used in Section 3.

Strain	Description/Phenotype	Source or Reference
E. coli DH5α	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lac*ZΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–	Hanahan 1983
E. coli HB101 pRK24	*lavYI galK2 xyl-6 mtl-I repsL20* carrying pRK24	B. Dupuy
B. subtilis BS49	Tn*916* donor strain, Tet^R	Haraldsen 2003
C. difficile R20291	Wild type, ribotype 027	Stabler 2009
C. difficile KNM6	*selD* targeted CRISPR-*cas9* mutant	McAllister 2017
C. difficile KNM9	KNM6 strain CRISPR-*cas9* restoration of *selD* at its native locus, Δ*selD*::*selD*$^+$	This study
C. difficile KNM10	*spo0A* targeted CRISPR-*cas9* mutant	This study
Plasmids		
pIA33	P_{xyl}::*dCas9-opt* P_{gdh}::*sgRNA-rfp catP*	Muh 2019
pJK02	P_{tet}-*Cas9-opt* P_{gdh}-*sgRNA colE1 pCD6 traJ catP*	McAllister 2017
pJS116	*B. subtilis-C. difficile* shuttle vector (pCD6 ColE1 *Tn916 oriT* Cm^R)	Francis 2013
pKM126	*tn916* oriT in pJK02	McAllister 2017, This study
pKM142	*selD* with 500 bp upstream in pJS116	McAllister 2017
pKM181	*selD* complementing homology region in pKM126	This study
pKM183	sgRNA for targeting region surrounding *selD* deletion in pKM181	This study
pKM194	xylose-inducible P_{xyl} promoter in pKM183	This study
pKM197	xylose-inducible P_{xyl} promoter in pKM126	This study
pKM213	*spo0A*-targeting sgRNA in pKM197	This study
pKM215	*spo0A* deletion homology region in pKM213	This study

Table 5.5 Oligonucleotides used in Section 3.

Oligonucleotide	Sequence	Reference
5'Tn916ori	AAGCGGAAGAGCGCCCAATACGCAGGGCCCT AACATCTTCTATTTTTCCCA	McAllister 2017
3'Tn916ori	TATCTACAATTTTTTTATCCTGCAGGGGGCCC CTAAAGGGAATGTAGATAAATTATTAG	McAllister 2017
5'selD_comp	CAATTTTTTTATCAGGAAACAGCTATGACCGC GGCCGCACCTAAAATAGGTGAAGCAAC	This study
3'selD_comp 2	GGTCTTAAGCGATCGCGCATGTCTGCAGGCCT CGAGCGCTGCATTATTATTTACAA	This study
CRISPR_selD_comp2	GTGTGCTATAATTAAACTGTAAAACGCGTAG CCGCTAAAATAGGGCCAGGTTTTAGAGCTAG AAATAGCAAGTTAAAATAAGGCTAGTCCGTT ATCAACTTGAAAAAGTGGCACCGAGTCGGTG CTTTTTTTCTATGGAGAAATCTAGATCAGCAT GATGTCTGACTAGACGCGTAAGCTCTGCAAC TATTTTTAGAT	This study
5'selDcomp_HR_xylR 2	GTACTAGTTGTAAATAATAATGCAGCGCTCG AGCTAGCATAAAAATAAGAAGCCT	This study
3'cas9_Pxyl 2	TAATCCTATACTATATTTTTTATCCATTTAATT AACTCTCCTCTTTACCCTCCTT	This study
5'pyrE_HR_xylR 2	CATTCAAAAGAAGGAAGAACATCAATGCTTC TCGAGCTAGCATAAAAATAAGAAGCCT	This study
CRISPR_spo0A_2	GTGTGCTATAATTAAACTGTAAAACGCGTGA CATGCAATAGAGGTTGCAGTTTTAGAGCTAG AAATAGCAAGTTAAAATAAGGCTAGTCCGTT ATCAACTTGAAAAAGTGGCACCGAGTCGGTG CTTTTTTTCTATGGAGAAATCTAGATCAGCAT GATGTCTGACTAGACGCGTAAGCTCTGCAAC TATTTTTAGAT	This study
5' spo0A_UP	TTTTTTTATCAGGAAACAGCTATGACCGCGGC CGCCAGAAAACCATAATAAAGAGTTTAA	This study
3' spo0A_UP	TGTCTTGTCCTGTTGAATGTCTTCCTTCTGCTA AAAAACATCTTCTTATTACAGAAAACT	This study
5' spo0A_DN	GATGTTTTTTAGCAGAAGGAAGACATTCAAC AGGACAAGACATAAAAAGTAAGGC	This study
3' spo0A_DN	AATGCAGGCTTCTTATTTTTATGCTAGCTCGA GGATTTATAACTGCTATTTCCCC	This study
5'selD	GAGCTTCCTAAAAATGAAGTAAATATCAATA AACAG	McAllister 2017
3'selD	TTTTGCTCAAAACAATCACTCTTTCTCTATAA TATT	McAllister 2017

Table 5.5 Continued.

Oligonucleotide	Sequence	Reference
5'spo0A_del	CAAATAATTCAGAGCTAGGTATAAGTGGTAATAT	This study
3'spo0A_del	CAATGCCTTAATTAAAAAGCCTTACTTTTTATGTCTTG	This study
5' tcdB	TTACATTTTGTTTGGATTGGAGGTC	McAllister 2017
3' tcdB	AGCAGCTAAATTCCACCTTTCTACC	McAllister 2017
5' catP 3	ATGGTATTTGAAAAAATTGATAAAAATAG	McAllister 2017
3' catP 2	TTAACTATTTATCAATTCCTGCAATTCG	McAllister 2017
5'tetR_CO_cas9	CTGAGCTCAATAATACTAGGAGGTTTTTTTAATTAAATGGATAAAAAATATAGTATAGGATTAGATATAGGAAC	McAllister 2017
3'COcas9 (975)	GTAATGTTAAATCTTGATGATGTTCATC	This study
5'gdh	TGCAGGCTTCTTATTTTTATGGTTTAAACGGTTTTAGCTGGGATATCG	McAllister 2017
3'gRNA 2	CATCTAAAAATAGTTGCAGAGCTTACGCGTCTAGTCAGACATCATGCTGA	McAllister 2017
5'rpoB_qPCR	GAGTGTAAAGAGAGAGATGC	Fimlaid 2013
3'rpoB_qPCR	CTTCCGCATAGTAAACACC	Fimlaid 2013
5' tcdB_qPCR	GGCAAATGTAAGATTTCGTGTTCA	Edwards 2016
3' tcdB_qPCR	TCGACTACAGTATTCTCTGAC	Edwards 2016
5'0963_qPCR	CAGACTGTTGCAGATAGCATTGAGTA	This study
3'0963_qPCR	CAACAACAAATCTGTTTACACCTTGA	This study
5'0962_qPCR	TCAGGCTCCTACAACACTTTTATTTG	This study
3'0962_qPCR	TCTGCATTACTTTCCTCGATTATCTC	This study
5'prdB_qPCR	GGAAGAGGGAGTAGACGGTGTAGTT	This study
3'prdB_qPCR	ACGATCACGGCAGTTCTATGG	This study
5'grdB_qPCR	TATAGCAGGAGTTATGGATTTAACAGAAGAG	This study
3'grdB_qPCR	CTAAATTTGCATACACTGGGTCATATC	This study
5'grdA_qPCR	TTTCGCTGGACCACTTGCT	This study
3'grdA_qPCR	TGGTTCCTCAACAACGTGGTAA	This study
5'mtlF_qPCR	CATATATGGGAATGGGAGTTGCTAT	This study
3'mtlF_qPCR	TTTCTCCATCAAAATCTATACCATTAGG	This study