IDENTIFICATION AND CHARACTERIZATION OF *CLOSTRIDIUM SORDELLII* TOXIN GENE REGULATOR

by

APOORVA REDDY SIRIGI REDDY

B.Tech, Anna University, 2011

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Division of Biology
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2014

Approved by:

Major Professor
Dr. Revathi Govind

Copyright

Abstract

Toxigenic *Clostridium sordellii* causes uncommon but highly lethal infections in humans and animals. Recently, an increased incidence of *C. sordellii* infections has been reported in women undergoing obstetric interventions. Pathogenic strains of *C. sordellii* produce numerous virulence factors, including sordellilysin, phospholipase, neuraminidase, and two large clostridial glucosylating toxins, TcsL and TcsH. Recent studies have demonstrated that TcsL toxin is an essential virulence factor for the pathogenicity of *C. sordellii*. In this study, we identified and characterized TcsR as the toxin gene (*tcsL*) regulator in *C. sordellii*. High-throughput sequencing of two *C. sordellii* strains revealed that *tcsR* lies within a genomic region that encodes TcsL, TcsH, and TcsE, a putative holin. By using ClosTron technology, we inactivated the *tcsR* gene in strain ATCC 9714. Toxin production and *tcsL* transcription were decreased in the *tcsR* mutant strain. However, the complemented *tcsR* mutant produced large amounts of toxins, similar to the parental strain. Expression of the *Clostridium difficile* toxin gene regulator *tcdR* also restored toxin production to the *C. sordellii tcsR* mutant, showing that these sigma factors are functionally interchangeable.

Table of Contents

List of Figures

List of Tables

Acknowledgements

On the very outset, I bow in reverence to the almighty but for whose blessings I would not have come this far in life.

I would like to express my sincere gratitude to my advisor Dr. Revathi Govind for trusting in my abilities and giving me an opportunity to work as a graduate student under her guidance.

I extend my thanks to my committee members Dr. Govindsamy Vediyappan and Dr. Lynn Hancock for their kind assitance and constant guidance.

I would like to thank all the members of Dr. Govind's lab for their friendship and support. Brintha Parasumanna Girinathan, Ryan Zapotocny, Sterlin Braun, Rebekah nichols and Trisha Rettig have made working in the lab a fun experience.

I am especially grateful to my family, my parents (Mr. Sirigi Reddy Veera Reddy , Mrs.Sirigi Reddy Aruna) and my brother (Sirigi Reddy Avinaash Reddy) for all their love, encouragement and constant support.

I would like to thank all my friends from here and India for making my stay here feel at home, interesting and crazy at times.

Chapter 1 - Introduction

Anaerobic, Gram-positive bacteria belonging to the phylum of Firmicutes define the genus *Clostridium*. They are rod shaped bacteria and are capable of forming spores. There are about 100 different species of non-pathogenic and pathogenic *Clostridium.* The majority of the human intestinal microbiota belongs to the phylum Firmicutes, where clostridia constitute a significant proportion (22, 24). *Clostridium botulinum, Clostridium tetani, Clostridium perfringens, Clostridium difficile, Clostridium septicum,* and *Clostridium sordellii* are the important clostridial species that cause human disease. *C. botulinum* and *C. tetani* cause paralysis by producing neurotoxins. *C. perfringens* infections result in gangrene and enteritis. *C. difficile,* a nosocomial pathogen causes diarrhea and pseudomembranous colitis. *C. septicum,* a resident of the human micro flora causes myonecrosis. *C. sordellii* , an emerging pathogen is known to causes myonecrosis, sepsis and shock (8).

C. sordellii is an anaerobic, spore forming bacteria having peritrichous flagella. In 1922, *C. sordellii* was first isolated by an Argentinean microbiologist Alfredo Sordelli and based on the morphology and tissue edema characteristics of the bacteria; it was named as *Bacillus oedematis sporogenes.* In 1929, based on its similarities with *Clostridium oedematoides,* it was given the name *Clostridium sordellii*. It is commonly found in soil and intestines of animals, including 0.5% in humans. Most of the strains are non-pathogenic, but virulent strains can cause lethal infections. In animals, especially sheep and cattle they cause enteritis and enterotoxaemia. In humans, they cause myonecrosis and gangrene (1). Women in post-partem or having undergone drug induced abortions are more susceptible to *C. sordellii* infections. Mortality rate associated

with *C. sordellii* infections is 69%; out of which 63% of these cases are of women with a mean age of 33.6 years. One in two hundred deaths in women of reproductive age were caused by the toxic shock syndrome caused by *C. sordellii* (1). *C. sordellii* associated systemic infections are rare, but cases have been reported of *C. sordelli* infections with pericarditis (9). Several incidences of *C. sordellii* infections have been reported in intravenous heroin users (7).

Pathogenesis

Importance as animal pathogen

C. sordellii has been associated with animal infections in horses, sheep, cattle, guinea pigs, lions, quails, etc (4, 10, 11, 15, 18). The kind of infection varies from animal to animal and the pathogenesis of these infections is not known.

Sudden deaths in sheep by *C. sordellii* have been reported in Britain where 37 sheep died between the years 1993 and 1995. It was found that *C. sordellii* affect four to ten week old lambs (15). Similar deaths in sheep have been reported before where *C. sordellii* was isolated but no proper investigations were made. In USA, *C. sordellii* is identified as a cattle pathogen. Cases in sheep have been recorded in the year 1962 and 1976 where it was isolated from the liver, spleen, lower intestine and lungs (15).

C. sordellii infections were reported in wild animals in captivity as well. Literature describes few cases of clostridial disease in felines (11). The first report of death in felines by *C. sordellii* was recorded in the year 2003, where five lions after the onset of the disease symptoms died within

24 to 36 hours. *C. sordellii* was isolated from all five lions and based on histological findings, the death was caused by acute enteritis (inflammation of the intestines), myositis (inflammation of the muslces) and cellulitis (inflammation of connective tissue). Two types of clostridial myositis are identified – a non-gangrenous form which is characterized by cellulitis rather than myositis and a gangrenous form characterized by gas production anddisintegration of muscles (11). Several perinatal mortality cases in animals have been caused by omphalitis and septicemia. Eight cases of perinatal mortality in foals by *C. sordellii* was reported so far. The bacterial pathogen was isolated from the internal umbilical remnant and peritoneal fluid and it was found to have caused peritonitis and septicemia (18). The first *C. sordelli* infection in bear was reported in Spain (4). *C. sordellii* was found in the cultures from the liver, muscle and intestine of a brown bear in Spain. It was a fatal case of gangrenous myositis and septicemia (4).

C. sordelli associated ulcerative enteritis was reported in avian species as well (10). Two such outbreaks in quail farms were reported, where fifty quails died over a 3-week period and sixteen died in the other. On further studying these infected quails they found severe lesions in the liver and intestine that resembled ulcerative enteritis. *C. sordelli* cultures were identified on extensive anaerobic culturing of the intestine and liver from the quails. Several cases of *C. sordellii* associated disease were reported in different kinds of birds, and led to the suggestion of *C. sordelli* being a normal intestinal flora in avian species (10).

Impact on human health

C. sordellii infections have found to be highly lethal and cause toxic shock syndrome similar to that caused by *Clostridium novyi.* It was mostly found to affect women of reproductive age (2). The clinical symptoms caused by *C. sordellii* infection in women include abdominal pain,

hypotension, tachycardia, third-space fluid accumulations, and marked leukemoid response. The mortality associated with these gynecologic infections was found to be greater than ninety-five percent and those of injection drug users and soft tissue infections were found to be fifty percent approximately (26). This disease got widespread publicity in 2003, when four fatal cases of *C. sordellii* infections in women were investigated (26). The progress of *C. sordellii* infections was so rapid that death precedes diagnosis. In a review by Aldape in the year 2006 (1) it was found that out of 45 reported cases of *C.sordellii* infections, 8 were in women with normal child birth, 5 in women with medically induced abortions and 2 in women with spontaneous abortion. The leukemoid response was an indicator of the fatal outcome of the patients, a white blood count of 18,000 cells/mm^3 was found in patients who survived and 75,000 cells/mm^3 in patients that did not survive. This leukemoid response was thought to be caused by the exotoxin neuraminidase produced by *C. sordellii* (6). Eight cases of *C. sordellii* infections after medical abortions using drugs mifepristone and misoprostol were reported between the years 2000 and 2010. The investigations on these cases led to changing the regimen of medical abortions, the way the drugs were administered and also administering antibiotics during these treatments (17). Fatal cases of *C. sordellii* infections with abortions associated with Misoprostol lead to the investigation of the association of this drug with the infection. Misoprostol, an analog of prostaglandin E_1 functions like prostaglandin E_2 and acts as an immune suppressor. Studies on the effect of misoprostol of *C. sordellii* infections in rats showed that, intrauterine delivery of misoprostol worsened the mortality of *C. sordellii* infection but not that of *Lactobacillus crispatus,* a vaginal commensal. It also reduced the production of TNF-α in the uterus during infection. The *in vitro* studies showed that misoprostol suppressed TNF-α production, impaired the leukocyte phagocytosis of *C. sordellii* and also inhibited the expression of the uterine epithelial cell β – defensin. These

studies suggest that misoprostol further enhances the chances of *C. sordellii* infections and can explain the post abortion sepsis, leading to death. But the effect of misoprostol in colonization and initiation of infection is not known (3). Several theories were put forward regarding *C. sordellii* infections and their association with patients who have undergone medical abortions and also on prevention of this disease (25). The different theories included possibilities of contamination of the drugs mifepristone and misoprostol that were used, mutations in *C. sordellii* that made it more pathogenic now than before, use of misoprostol vaginally allows existing vaginal *C. sordellii* to cause illness, self-insertion of misoprostol vaginally leading to contamination, and using antibiotics during abortions to prevent such infections. All such assumptions were questioned and the exact reason for *C. sordellii* infections and their association with women post abortions are still not known (25).

C. sordellii is known to mostly affect women of reproductive age, but several cases of *C. sordellii* associated bacteremia have also been reported. In the year 2009, a 59-year-old man with rectum carcinoma and liver metastases was diagnosed with *C. sordellii* associated bacteremia. Out of the 12 cases reported till the year 2009, the mortality was 70% and the patients died within a few hours or days after the infection (16). Cases have been reported of fatal *C. sordellii* associated soft tissue necrosis, a 4-year-old boy after surgery of transverse fracture of the arm started showing symptoms of *C. sordellii* infection and died in four days. A similar death occurred in a 21 year old woman, four days post-delivery after a third degree vaginal laceration during delivery (1). Omphalitis (inflammation of the umbilical cord stump) in newborns were reported to be associated with *C. sordellii* (18). On one occasion a patient with fatal *C. sordellii* infection was mistakenly treated for *C. difficile* infection based on the stool assay. This reiterates

the fact that toxins of *C. sordellii* and *C. difficile* are similar antigenically and that diagnosis for *C. sordellii* infections should be further confirmed with PCR and other techniques (14).

Major Virulence Factors

Virulence factors are those that play a role in the pathogenesis of the organism. Toxins and surface molecules are the two major classes of virulence factors. Compared to other bacterial genus, clostridia produce more toxins. In clostridia different kinds of toxins, extracellular proteins, spreading factors and proteolytic enzymes were found to contribute to their virulence (20). In *C. sordellii* , virulent strains produce up to seven exotoxins, out of which the major virulence factors are the Lethal toxin (LT) and Hemorrhagic toxin (HT). These toxins are classified under the family of Large Clostridial Toxins (LCT). Other members of the family includes toxins A and B produced by *C. difficile* and the alpha-toxin of *C. novyi* (13)

The LT and HT of *C. sordellii* were found to be similar to the cytotoxin and enterotoxin of *C. difficile* respectively. 76% similarity has been shown between the *C. sordellii* TcsL and *C. difficile* toxin B.

The LCT are generally 250 to 300 kDa in size and are the largest bacterial protein toxins known. They typically carry an enzymatic domain in its N-terminus region and a receptor-binding domain in its C-terminus. A hydrophobic region spans the center of the protein and plays a role in translocating proteins across the membrane. All large clostridial cytotoxins are glycosyltransferases. Conserved amino acids were recognized that are responsible for enzyme activity in the LCTs. A DXD motif located in the center of the enzyme domain is the best-characterized conserved region. The role of the DXD motif in the enzyme activity is not defined yet but has been found in other glycosyltransferases as well (13).The bacteria release toxins in

the environment that then enter into the eukaryotic cells by various cellular uptake mechanisms of the cell. The toxins bind to the cell surface receptors and enter into the cell by endocytosis as a receptor-toxin complex and are then translocated to the cytosol by a change in pH in the endosome. These toxins on entering into the target cell, glycosylate specific threonine residues and modify the Rho- and Ras- subfamily of proteins. It was found that the modification in the GTPases occurs at the Threonine 35 or Threonine 37 residue. The Rho and Ras family of proteins are active when bound to GTP and inactive when bound to GDP, and the Guanine nucleotide exchange factors help in the release of GDP and in binding of GTP. The clostridial toxins glycosylate the GTPases and inactive them. (5).

LT was found to be more toxic than HT. The role of the HT in virulence has not been studied yet. The other virulence factors and exotoxins produced by *C. sordellii* are sordellilysin, cholesterol dependent cytolysin, phospholipase C, hemolysin, neuraminidase, DNase, collagenase, and lysolecithinase. (7, 21).

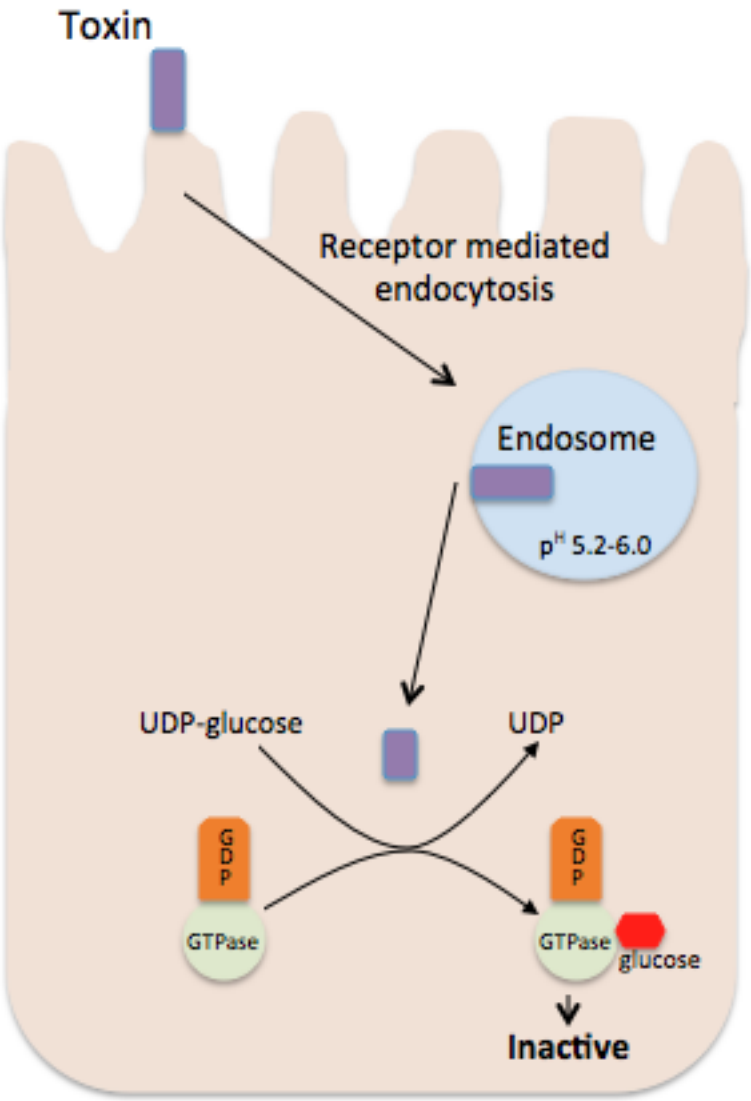


Figure 1.1 Mechanism of action of *C. difficile* toxins .

The toxins enter the cell by receptor mediated endocytosis and in the endosome the active enyme domain of the toxins are released into the cell. These glyxosyltransferases inactivate the Rho,Rac pathways by transfering glucose using UDP-glucose as substrate. These modifications result in transcirptional inactivation, actin condensation and apoptosis (Adapted from (23))

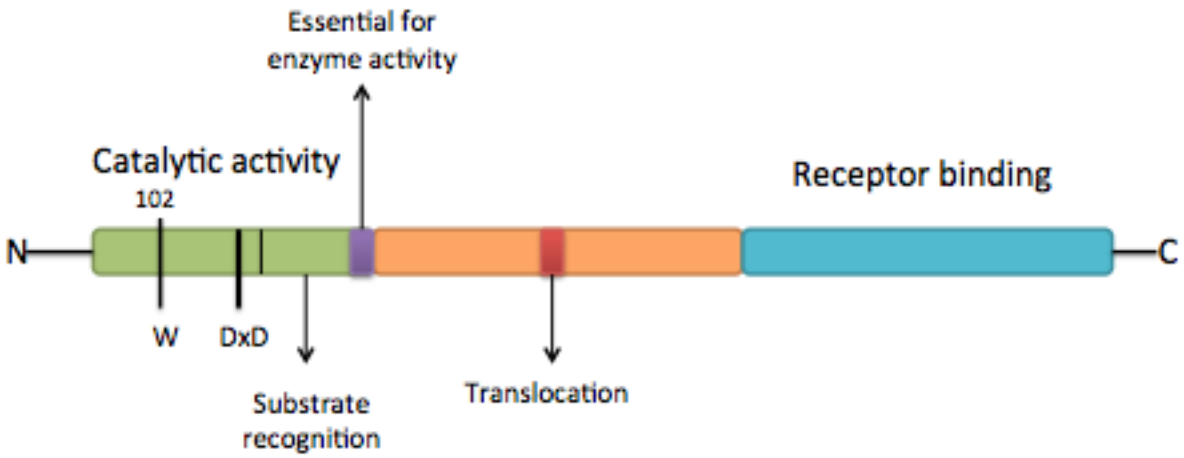


Figure 1.2 Functional domain of large clostridial toxins.

The toxins are divided into three functional domains. The N-terminus region carries the enzymatic domain , the C-terminus the receptor binding domain and a hydrophobic region in the center of the protein used in translocating them across cell membranes (Adapted from (13))

Regulation of toxin genes

The major checkpoint for regulating gene expression is transcription initiation. The core RNA polymerase has 5 subunits and on binding of the sixth subunit, the sigma factor makes the RNA polymerase complete and is called the holoenzyme. Recognition of promoter in DNA and transcription initiation depends on the sigma factor component of RNA polymerase holoenzyme. In bacteria, the house keeping sigma factor, sigma 70 directs the RNA polymerase towards the promoters of the house keeping genes. In addition, other sigma factors called alternate sigma factors that are specialized to recognize distinct class of promoters to drive transcription of discrete sets of genes whose products are needed for adaptive response. The role of alternative sigma factors in various cellular functions like spore formation, stress have been studied in *E. coli* and *B. subtilis* and are used as model organisms (19). The sigma factor makes contact with the core RNA polymerase to form open promoter complex during transcription initiation and is released during the step of elongation which in turn are recycled to bind to RNA polymerase to initiate another round of transcription. The σ^{70} are the major family of sigma factors that are further divided into four groups based on the four conserved domains. Group I comprises of the housekeeping sigma factors, Group II those related to stress, Group III are distantly related to group I and are those of the developmental checkpoints and heat shock and Group IV is that of the extra cytoplasmic subfamily. The next class of sigma factors is the σ^{54} family, which is related to nitrogen assimilation, utilization of carbon sources. The major difference between the σ^{70} and the σ^{54} family of sigma factors is that, the σ^{54} family of sigma factors do not easily form the open promoter complex like the σ^{70} family, they need help from bacterial enhancer binding proteins, which use ATP to achieve the required conformation (19).

In *E.coli,* studies have proven that the alternative sigma factors are found at a concentration lower than that of the house keeping sigma factors. Even under conditions where the alternative sigma factors are the most active, the cellular concentrations of the house keeping sigma factors is always higher than that of the alternative sigma factors. It has been found that the levels of the house keeping sigma factors and the core RNA polymerase is found to be constant at all growth conditions and housekeeping sigma factors exceed the core RNA polymerase concentrations by three fold. (19).

TcdR, the alternative sigma factor in *C. difficile* was the first alternative sigma factor that was shown to regulate the toxin gene expression (12). The biochemical characterization of TcdR showed it to be similar to group IV, the Extra Cytoplasmic Function (ECF) sigma factors. Several experiments were then done to confirm its ability to act as a sigma factor. It was found that TcdR could bind to RNA polymerase core enzyme to initiate transcription from the toxin gene promoters. Further studies found that TcdR like sigma factors in other *Clostridium* species like BotR in *C. botulinum,* TetR in *C. tetani* and UviA in *C. perfringens*. It was found that all these sigma factors were auto regulators and were turned on during nutrient limitation, DNA damage and when temperatures were not optimal. Phylogenetic studies were done between the four sigma factors – TcdR, BotR, TetR and UviA. The promoter regions of the genes targeted by all the four sigma factors were found to be nearly identical in the -35 region with a TTTACA conserved motif and were found to be different in the -10 region. The *clostridium* sigma factors differ in their structure and function and hence assigned their own group – Group V in the σ^{70} family as they are distantly related to the σ^{70} family. It was found that the sigma factors BotR

and TetR can be interchanged and can still regulate the expression of the other sigma factor dependent genes. Similarly UviA and TcdR were found to be interchangeable. Similarly, UviA and TcdR were fully inter-changeable. Neither BotR nor TetR could substitute for UviA or TcdR, however, and neither UviA nor TcdR could direct transcription of the natural targets of BotR or TetR. This property or function of interchangeability was found to be due to the conserved -35 regions of the target promoters and the restrictions of interchangeability were found to be due to the variations in the -10 target site sequences (12).

References

1. Aldape, M. J., A. E. Bryant, and D. L. Stevens. 2006. *Clostridium sordellii* infection: epidemiology, clinical findings, and current perspectives on diagnosis and treatment. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 43:1436-1446.

2. Aronoff, D. M. 2013. *Clostridium novyi*, *sordellii*, and *tetani*: mechanisms of disease. Anaerobe 24:98-101.

3. Aronoff, D. M., Y. Hao, J. Chung, N. Coleman, C. Lewis, C. M. Peres, C. H. Serezani, G. H. Chen, N. Flamand, T. G. Brock, and M. Peters-Golden. 2008. Misoprostol impairs female reproductive tract innate immunity against *Clostridium sordellii*. Journal of immunology 180:8222-8230.

4. Balseiro, A., A. Oleaga, L. Polledo, G. Aduriz, R. Atxaerandio, N. Kortabarria, and J. F. Marin. 2013. *Clostridium sordellii* in a brown bear (Ursus arctos) from Spain. Journal of wildlife diseases 49:1047-1051.

5. Belyi, Y., and K. Aktories. 2010. Bacterial toxin and effector glycosyltransferases. Biochimica et biophysica acta 1800:134-143.

6. Bryant, A. E., and D. L. Stevens. 2010. Clostridial myonecrosis: new insights in pathogenesis and management. Current infectious disease reports 12:383-391.

7. Carter, G. P., M. M. Awad, Y. Hao, T. Thelen, I. L. Bergin, P. M. Howarth, T. Seemann, J. I. Rood, D. M. Aronoff, and D. Lyras. 2011. TcsL is an essential virulence factor in *Clostridium sordellii* ATCC 9714. Infection and immunity 79:1025-1032.

8. Carter, G. P., J. K. Cheung, S. Larcombe, and D. Lyras. 2014. Regulation of toxin production in the pathogenic clostridia. Molecular microbiology 91:221-231.

9. Chaudhry, R., N. Verma, T. Bahadur, P. Chaudhary, P. Sharma, and N. Sharma. 2011. *Clostridium sordellii* as a cause of constrictive pericarditis with pyopericardium and tamponade. Journal of clinical microbiology 49:3700-3702.

10. Crespo, R., M. Franca, and H. L. Shivaprasad. 2013. Ulcerative enteritis-like disease associated with *Clostridium sordellii* in quail. Avian diseases 57:698-702.

11. de la Fe, C., J. M. Rodriguez, G. A. Ramirez, J. Hervas, J. Gil, and J. B. Poveda. 2006. Sudden death associated with *Clostridium sordellii* in captive lions (Panthera leo). Veterinary pathology 43:370-374.

12. Dupuy, B., and S. Matamouros. 2006. Regulation of toxin and bacteriocin synthesis in *Clostridium* species by a new subgroup of RNA polymerase sigma-factors. Research in microbiology 157:201-205.

13. Dürre, P. 2004. Handbook on clostridia. CRC Press.

14. Elsayed, S., and K. Zhang. 2006. Positive *Clostridium difficile* stool assay in a patient with fatal *C. sordellii* infection. The New England journal of medicine 355:1284-1285.

15. Lewis, C. J., and R. D. Naylor. 1998. Sudden death in sheep associated with *Clostridium sordellii*. The Veterinary record 142:417-421.

16. Matten, J., V. Buechner, and R. Schwarz. 2009. A rare case of *Clostridium sordellii* bacteremia in an immunocompromised patient. Infection 37:368-369.

17. Meites, E., S. Zane, C. Gould, and C. s. Investigators. 2010. Fatal *Clostridium sordellii* infections after medical abortions. The New England journal of medicine 363:1382-1383.

18. Ortega, J., B. Daft, R. A. Assis, H. Kinde, L. Anthenill, J. Odani, and F. A. Uzal. 2007. Infection of internal umbilical remnant in foals by *Clostridium sordellii*. Veterinary pathology 44:269-275.

19. Osterberg, S., T. del Peso-Santos, and V. Shingler. 2011. Regulation of alternative sigma factor use. Annual review of microbiology 65:37-55.

20. Rood, J. I., B. A. McClane, J. G. Songer, and R. W. Titball. 1997. The clostridia: molecular biology and pathogenesis. Academic Press.

21. Stevens, D. L., M. J. Aldape, and A. E. Bryant. 2012. Life-threatening clostridial infections. Anaerobe 18:254-259.

22. Tuovinen, E., J. Keto, J. Nikkila, J. Matto, and K. Lahteenmaki. 2013. Cytokine response of human mononuclear cells induced by intestinal *Clostridium* species. Anaerobe 19:70-76.

23. Voth, D. E., and J. D. Ballard. 2005. *Clostridium difficile* toxins: mechanism of action and role in disease. Clinical microbiology reviews 18:247-263.

24. Wells, C. L., and T. D. Wilkins. 1996. Clostridia: Sporeforming Anaerobic Bacilli. *In* S. Baron (ed.), Medical Microbiology, 4th ed, Galveston (TX).

25. Winikoff, B. 2006. *Clostridium sordellii* infection in medical abortion. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 43:1447-1448.

26. Zane, S., and J. Guarner. 2011. Gynecologic clostridial toxic shock in women of reproductive age. Current infectious disease reports 13:561-570.

Chapter 2 - Identification and Characterization of *Clostridium sordellii* Toxin Gene Regulator

Introduction

Clostridium sordellii is an anaerobic, gram-positive spore-forming bacterium and a common inhabitant of soil and animal gastrointestinal tract. Virulent strains of *C. sordellii* are recognized as the causative agent of broad spectrum of diseases including myonecrosis, uterine infections, and sepsis in humans. *C. sordellii* is also known to cause lethal infections in several animal species including sheep, foals and lambs (7,25,26,35) Recently, fatal cases of *C. sordellii* endometritis following medical abortions with mifepristone-misoprostol combinations have been reported(15). It's been suggested that Mifepristone/misoprostol may facilitate colonization of *C. sordellii* in uterine tissue, trigger toxin expression, and induce hypotension and systemic shock by deregulating the host's immune response (3).

Pathogenic *C. sordellii* strains produce up to seven identified exotoxins (39). Of these, the two major toxins, lethal toxin (TcsL) and the hemorrhagic toxin (TcsH) are regarded as major virulence factors (31,36). The lethal toxin produced by *C. sordellii* was shown to invoke enteritis in animals and was proved to be essential for the virulence of *C. sordellii*(6,36). TcsH and TcsL are members of the large clostridial cytotoxin (LCC) family with a predicted molecular weight of 300kDa, and 250 kDa respectively (31,36). The *C. sordellii* toxins were reported to be similar to *C. diffcile* toxins A and B in terms of biological activity and antigenicity (30). To date

only TcsL encoding gene has been sequenced and it was found to be 76% identical to *C. difficile* toxin B (*tcdB*).

In this study we have sequenced two *C. sordellii* strains ATCC 9714 and VPI 9048 by high throughput sequencing techniques and have identified many open reading frames (ORFs) surrounding *tcsL* gene. Consistent with the previous reports, in ATCC 9714 strain only lethal toxin TcsL encoding gene is present and the VPI 9048 strain carry both TcsL and TcsH encoding genes. Near the toxin genes, we identified a small ORF with similarities to sigma factors including TcdR, the positive regulator of *C. difficile* toxin genes. We named the ORF *tcsR* and have tested its role in toxin gene regulation in *C. sordellii*. Here we show that *C. sordellii tcsR* mutant is defective in toxin production due to the lower transcription of *tcsL*. Further we complemented the mutant with a functional *tcsR* and showed that it can reactivate the toxin production in the mutant. *C. sordellii tcsR* mutant could also be complemented with *C. difficile* toxin gene regulator TcdR, showing that these sigma factor are closely related to each other. This is the first report on *C. sordellii* toxin gene regulator and on the toxin locus region in *C. sordellii* genome.

Table 2.1 Bacterial strains and plasmids used in this study

Strain or plasmid	Characteristic(s) and/or origin	Source or reference
***C. sordelii* strains**		
ATCC 9714	$TcsL^+$ $TcsH^-$	American Type Culture Collection (20)
VPI 9048	$TcsL^+$ $TcsH^+$	Tec Lab (VA) (18)
***E. coli* strains**		
DH5α	λ^- ϕ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)*U169 recA1 endA1 hsdR17*(r_K^- m_K^-) *supE44 thi-1 gyrA relA1*	New England BioLabs, MA
S17 - 1	Favors conjugation	42
Plasmids		
pMTL007C-E5	ClosTron plasmid	23
pTUM007::Cdi-*tcsR*-342	pMTL007C-E5 carrying *tcsR* -specific intron	This study
pRPF185	pMTL960, Cm^r, Tm^r,$gusA^+$, inducible tetracycline (Tet) promoter	14
pRGL153A	pRPF185 with a promoter less *gusA* gene	This study
pRGL100	Tet promoter in pRPF185 replaced with *tcsL* promoter	This study

pRGL161	Tet promoter in pRPF185 replaced with *tcsR* promoter	This study
pRGL162	Tet promoter in pRPF185 replaced with *tcsH* promoter	This study
pRGL163	Tet promoter in pRPF185 replaced with *tcsE* promoter	This study
pRGL154	pRPF185 without a *gusA* gene	This study
pRGL145-1	pRGL154 with wild-type *tcsR* under inducible Tet promoter	This study
pRGL144-1	pRPF154 carrying the wild-type *tcdR* under inducible Tet promoter	This study

Materials and Methods

Bacterial strains and growth conditions

C. sordellii strains VPI9048 (TechLab, VA, USA)(18), ATCC9714 (19) (Table 1), and the *tcsR* mutant were grown anaerobically (10% H_2, 10% CO_2 and 80% N_2) in TY broth or TY agar. *E. coli* strain SL-17 used for conjugation was cultured aerobically in LB medium. When necessary, *E. coli* cultures were supplemented with chloramphenicol or ampicillin, at 30 μg ml^{-1} and 100 μg ml^{-1}, respectively. All routine plasmid constructions were carried out using standard procedures.

High throughput genome sequencing, assembly and annotation

The genomes of *C. sordellii* ATCC 9714 and VPI 9048 were sequenced using the Roche 454 Life Sciences Genome Sequencer FLX (Roche 454 Life Science, Branford, CT, USA) and with Illumina Genome Analyzer (Illumina, San Diego, CA, USA) following the manufacturer's instructions. For Roche 454 sequencer, shotgun library was prepared with 5 μg genomic DNA using the "Standard DNA Library Preparation Kit" (04852265001, Roche). Nebulized, purified, and adaptors attached single strand DNA fragments were clonally amplified using the "Emulsion PCR Kit I" (04852290001, Roche). Sequencing on the GS FLX was performed using the "Standard LR70 Sequencing Kit" (04932315001, Roche). Images were processed using the "Genome Sequencer FLX Data Processing Pipeline 1.1.02.15" and sequences generated were assembled using Newbler assembler (Roche). For illumina sequencing, the DNA template Library was prepared using the Illumina "Genomic DNA Sample Prep Kit" (Illumina). Briefly, 5 μg of genomic DNA was broken into fragments of approximately 100 bp by nebulization. After end repairing and adaptor ligation, the samples were gel-purified to recover fragments of 150–

250 bp, which were PCR amplified for 15 cycles. The DNA template Library was then used for flow-cell preparation using the "Standard Cluster Generation Kit" (Illumina). Sequencing on the Illumina Genome Analyzer was performed using "Genomic DNA sequencing primer V2" for 36 cycles. At the end of the run, images were processed using the "Solexa Data Analysis Pipeline 0.2.2.6". Reads from Roche 454 and Illumina were mapped to contigs using SOAP package at default parameters. Assembled contigs were submitted to Institute for Gennomic Sciences at University of Maryland for annotation Service, where it was run through the prokaryotic annotation pipeline. Along with gene finding, Glimmer, Blast-extend-repraze (BER) searches, HMM searches, TMHMM searches, SignalP predictions, and automatic annotations from Auto-Annotate are included in the annotation pipeline.

The nucleotide sequences and the corresponding automated annotations for the first versions of the genomes of *C. sordellii* str. ATCC 9714 and VPI9048 were submitted to GenBank, with accession numbers APWR00000000 and AQGJ00000000, respectively.

Construction of *tcsR* mutant

A *tcsR* mutant was generated in *C. sordellii* ATCC9714 by the insertion of a bacterial group II intron using the ClosTron gene knockout system as described by Heap et al. (23). The insertion site in antisense orientation between nucleotides 234-235 of the *tcsR* ORF was selected to design the retargeting intron. The intron was designed using Perutka algorithm, a web based design tool available at Clostron.com site. It was then synthesized and cloned in plasmid pMTL007-E5C. The resulting plasmid, pTUM007: *Cdi-tcsR*-234a, was transferred to *C. sordellii* ATCC9714 strain by conjugation as described previously (23). Thiamphenicol-resistant transconjugants were

resuspended in 200 µl of TY broth and plated on TY agar plates containing erythromycin (5µg ml^{-1}) to select potential Ll.ltrB insertions. The putative *tcsR* mutants were then screened by PCR using *tcsR*-specific primers (ORG94, ORG95) in combination with the EBS-U universal primer (Supplementary Table 1).

Southern Blot analysis

Southern blot analysis was performed as described previously (to verify a specific single integration of the group II intron into the genome (17). Ten µg of genomic DNA was digested with EcoRV enzyme and separated on a 0.8% agarose gel by electrophoresis. DNA was transferred onto IMMOBILON-NY+ nylon membrane (Millipore, Bedford, MA) by the capillary transfer method. Prehybridization of the filter was done for 2 hours at 60°C in 5x SCC, 5x Denhart and 100 mg/ml of salmon sperm DNA. Probes specific for the group II intron *ermB* gene and *tcsR* genes were radiolabeled (^{32}P dATP) using High Prime kit (Roche) and hybridized overnight in 10 ml fresh pre-hybridization buffer at 60°C. The hybridized membrane was washed twice for 30 minutes in 2x SCC, 0.5% SDS and 30 minutes and in 1X SSC, 0.5% SDS and analyzed using phosphorimage screen and a Typhoon 9410 scanner (GE healthcare).

Toxin assay

C. sordellii ATCC 9714 and its *tcsR* mutant cultures were grown in TY for 10 hours and the bacterial cells were collected after centrifugation. Cell pellets were re-suspended in 10 mM Tris buffer (pH 8.0), containing a protease inhibitor cocktail (Roche, Mannheim, Germany). The cytosolic contents were obtained by sonicating the cells followed by a brief centrifugation to

remove the unbroken cells and cell debris. Total protein concentration was determined using the Bio-Rad protein assay reagent. Equal amounts of cytosolic proteins (50µgs) were assayed for their relative toxin contents using the *C. difficile* Premier Toxin A&B ELISA kit from Meridian Diagnostics Inc., [Cincinnati, Ohio]. This ELISA kit is known to recognize with *C. sordellii* toxins as well (32).

RNA extraction and quantitative RT- PCR

Total RNA was extracted from *C. sordellii* cultures grown for 10 h in TY media following protocol described previously (6, 23). After treating the total RNA with DNAse (Turbo-Ambion), reverse transcription was performed using AMV Reverse Transcriptase (Promega) using random hexamer oligonucleotides primers with1 µg template RNA. The cDNA samples were then stored at -20°C until needed. Primers specific for *tcsL, tcsE* and *tcsR* (Supplementary Table 1) were designed using Primer 3 software (Geneious Software). Quantitative RT-PCR was performed using iQPCR real-time PCR instrument (BioRad). Reactions were carried out using SYBR Green Master Mix (BioRad) with 20 ng of cDNA as template. Samples were normalized using *C. sordellii* 16sRNA.

Construction of reporter plasmids and Beta Glucuronidase assay

Approximately, six hundred base pairs of the upstream DNA regions of *tcsL*, *tcsH*, *tcsR* or *tcsE* genes, along with their potential ribosomal binding sites (RBS) were PCR amplified using specific primers with KpnI and SacI recognition sequences (Supplementary table1) using ATCC 9714 chromosomal DNA as a template. Plasmid pRPF185 carries a *gusA* gene encoding beta

glucuronidase under the tetracycline inducible (tet) promoter (14). Using KpnI and SacI digestion, we removed the tet promoter and replaced it with either *tcsL*, *tcsR*, *tcsH* or *tcsE* upstream regions to create plasmids pRGL100, pRGL161, pRGL162 and pRGL163 respectively (Table 1). To create a plasmid (pRGL153A) with promoter less gusA we removed the tet promoter from plasmid pRPF185 with KpnI, SacI digestion and was self-ligated by creating blunt ends. Plasmids pRGL100, pRGL161, pRGL162, pRGL163 and pRGL153A (control) were introduced into ATCC9714 and its *tcsR* mutant through conjugation as described in the earlier section. The transconjugants were then grown in TY medium in the presence of thioamphenicol (15ugs/ml) for overnight. Overnight cultures were used as inoculum at 1:100 dilution to start a new culture. Bacterial cultures were harvested at 10 hours of growth and the amount of beta glucuronidase activity was assessed as described earlier (29) with minor modifications. Briefly, the cells were washed and suspended in 0.8 ml of Z buffer (60 mM $Na_2HPO_4 \cdot 7H_2O$ [pH 7.0], 40 mM $NaH_2PO_4 \cdot H_2O$, 10 mM KCl, 1 mM $MgSO_4$ $7H_2O$, 50 mM 2-mercaptoethanol) and were sonicated. The enzyme reaction was started by the addition of 0.16 ml of 6 mM p-nitrophenyl beta-D-glucuronide (Sigma) to the broken cells and stopped by the addition of 0.4 ml of 1.0 M $NaCO_3$. Beta glucuronidase activity units were calculated as described earlier (11, 29).

Complementation of *C. sordellii tcsR* mutant

The *tcsR* ORF along with its RBS was 215 PCR amplified from ATCC9714 chromosomal DNA using primers ORG203 and ORG204 (Supplementary Table 1). Similarly the *tcdR* ORF with its RBS was amplified from JIR8094 chromosomal DNA using primers ORG208 and ORG209. The resulting PCR products digested with SacI and BamHI were eventually cloned into vector pRFP185 (14) under a tetracycline inducible promoter to create plasmids pRGL145-1(with *tcsR*)

and pRGL144-1 (with *tcdR*), which were then introduced into *tcsR* mutant strain. Transconjugants carrying either pRGL145-1, pRGL144-1 or the vector pRGL154 (pRPF185 without gusA) alone were grown overnight in TY medium supplemented with thiamphenicol. Fresh 10 ml cultures were initiated using 0.1 ml of overnight cultures and were grown for 4 hours in TY medium up to 0.5 OD_{600nm} with thiamphenicol before the induction with ATc (Anhydrotetracycline) at the concentration of 50ngs/ml. Cultures were harvested 4 hrs after induction and cytosolic proteins were extracted for the detection of toxins using ELISA.

Table 2.2 Salient features of *C. sordellii* draft genomes and other *Clostridium* spp. strain genomes

	C. sordellii **(ATCC 9714) ***	***C. sordellii*** **(VPI 9048) ***	***C. difficile*** **(630) †**	***C. difficile*** **(R20291)†**	***C. difficile*** **(196)†**	***C. perfringens*** **(Strain 13) †**	***C. botulinum*** **(ATCC 3502)†**
Size (Mbp)	~3.03	~3.32	4.29	4.19	4.11	3. 03	3.88
G+C %	27.4%	27.3%	29.06%	28.8	28.6	28.57	28.24
Proteins coding genes	3271#	3985#	3798	3757	3454	2723	3590
RNA							
tRNA genes	66[#]	36[#]	87	82	82	96	80
rRNA 23S	12[#]	6[#]	11	10	10	10	9
rRNA 16S	12[#]	6[#]	11	9	10	10	9
rRNA 5S	12[#]	6[#]	10	8	9	10	9

[a] Draft genome.
[b] Genome information available as of May 2013.
[c] Predicted value; the number may change in the future.
[d] Only ORFs with more than 50 amino acid residues are included in this value

Results

Sequencing and de novo assembly of *C. sordellii* genomes:

Genome sequences for two *C. sordellii* strains were generated via 454 and Illumina sequencing technologies. Strain ATCC9714 is known to produce only TcsL, whereas strain VPI9048 produces both TcsL and TcsH toxins. The Roche 454 GS FLX system was used to generate sequences of strain ATCC 9714, which were assembled using Newbler assembler (Roche). A total of 637,164 reads with an average length of 387 bases was obtained and was assembled into 164 contigs with an average contig size of 21,629 bps. Strain ATCC 9714 was resequenced using Illumina technology along with strain VPI 9048. Totals of 3.563 and 3.616 million reads of 35 bases in length were obtained for ATCC 9714 and VPI 9048, respectively. Sequences from both 454 and Illumina were assembled using the SOAP package. This resulted in 104 and 166 contigs for strains ATCC 9714 and VPI9048, respectively. Gaps in genome coverage were not filled in with manual sequencing due to resource constraints. The overall characteristics of the draft *C.sordellii* genomes are summarized in Table 2.

Features of *C. sordellii* toxin gene locus and its similarity to the *C. difficile* PaLoc:

Bacterial virulence-associated genes are often found in mobile genetic elements. In Clostridia, the tetanus toxin gene of *Clostridium tetani* is encoded by a plasmid while the genes for botulinum toxins in *Clostridium botulinum* strains are within bacteriophage genomes (5, 13). The gene for *Clostridium novyi* alpha toxin, which shows high homology to TcsL of *C. sordellii* and TcdA of *C. difficile*, is also carried by a phage (12). However, the PaLoc in *C. difficile* is distinct and is not associated with an actively mobile genetic element (4). The *C. difficile* PaLoc includes

five genes, *tcdR*, *tcdB*, *tcdE*, *tcdA* and *tcdC*, and is found in the same locus in all *C. difficile* toxigenic strains (Figure 1C). Regions adjacent to the *C. difficile* PaLoc are not similar to any known transposon, plasmid or phage-like element (4). However, the base composition of the PaLoc differs from that of the genome as a whole, suggesting that it was acquired by horizontal transfer (4).

In *C. sordellii*, the lethal toxin-encoding *tcsL* gene is 76% similar to *tcdB* and the hemorrhagic toxin-encoding gene *tcsH* is 78% similar to *tcdA*. We identified the *tcsL* gene within the 57,746-bp contig #88 in strain ATCC9714 and within the 77,359-bp contig #152 of the VPI 9048 strain. Most of the genes surrounding the toxin-encoding genes are conserved in strains VPI9048 and ATCC9714 (Figure 1A and 1B). Immediately downstream of the *tcsL* gene is *tcsE*, a gene that encodes a holin-like protein that is homologous to the *tcdE* gene of *C. difficile*. TcdE is essential for the efficient secretion of toxins by *C. difficile* (16). In strain VPI 9048, *tcsH*, the hemorrhagic toxin-encoding gene, lies downstream of *tcsE*, but in strain ATCC 9714 only a truncated *tcsH* gene is present. Immediately upstream of *tcsH* is *tcsR*, a gene that is homologous to several sigma factor-encoding genes, including *tcdR* of *C. difficile*. Organization of the toxin genes in *C. sordellii* differs from that of the *C. difficile* pathogenicity locus. In *C. difficile*, the toxin genes *tcdA* and *tcdB* are transcribed in the same direction, but in *C. sordellii tcsL* and *tcsH* are transcribed in opposite directions. In *C. difficile*, *tcdR* is upstream of the *tcdB* gene and is transcribed in the same direction as the toxin genes. In *C. sordellii* the *tcsR* gene is upstream of the *tcsH* gene and is transcribed in the same direction as the *tcsH* gene. In *C. difficile*, *tcdC*, a gene downstream of *tcdA* codes for an anti-sigma factor, which affects toxin gene transcription by regulating TcdR activity. We were unable to identify any *tcdC* homologue near the toxin

locus of *C. sordellii* or at any other location in the incompletely sequenced genomes. This suggests that the toxin genes in *C. sordellii* may be regulated differently from *C. difficile*.

Unlike the case for the *C. difficile* PaLoc, the genes adjacent to the toxin genes in *C. sordellii* show several hallmarks of a mobile genetic element. Specifically, the toxin locus of *C. sordellii* shows signatures of integrative and conjugative elements (ICEs) (Figure 1A, 1B; Supplementary tables 2 and 3). ICEs are self-transmissible, mobile, genetic elements that encode the machinery for conjugation as well as for the regulatory systems to control their excision from the chromosome and their conjugative transfer. Unlike conjugative plasmids, ICEs do not replicate autonomously; instead, they integrate into the host chromosome. Predicted coding sequences in the toxin locus show homology to conjugal transfer proteins, plasmid replication proteins, transposases, recombinases and resolvases (Supplementary tables 2 and 3). The proteins encoded by genes VPI 9048-*H476_0274* and ATCC 9714- *H477_0270* that lie upstream of *tcsR* show homology to various plasmid replication proteins (Supplementary tables 2 and 3). The genes (VPI9048-*H476_0297*, VPI9048-*H476_0302*, ATCC 9714- *H477_0297* and ATCC9714-*H477_0298*) appear to encode type IV secretory system conjugative DNA transfer family protein is also part of toxin locus. In addition, a type IV secretion-system coupling DNA-binding domain protein could be identified near the toxin-encoding genes. Moreover, a TraB homologue (VPI9048-*H476_0319*) that plays a role in DNA transfer in other bacteria is also present in the toxin loci, indicating the possibility of conjugal transfer of the genetic element. The presence of transposase-encoding genes (VPI 9048- *H476_0321*, VPI9048-*H476_0286*) along with conjugative elements suggests that the toxin loci in *C. sordellii* may be part of an integrative-conjugative element. All ICEs encode an integrase, which enables their integration into the host chromosome by site-specific recombination. In the *C. sordellii* toxin loci, we couldn't identify

any genes likely to encode an integrase, but did find a *recA*-type gene that might encode a protein involved in homologous recombination (Supplementary Tables 2 and 3). This is an unusual signature for an ICE and requires further functional characterization. Besides the main pathogenicity factors TcsH and TcsL, the toxin locus in *C. sordellii* codes for several proteins that may be involved in virulence associated processes during infection. The genes *H476_0289* in VPI9048 and *H477_0286* in ATCC9714 code for a possible collagen-binding protein with predicted CNA peptide repeats in the C-terminal region. In Staphylococcus aureus the CNA repeat protein mediates bacterial adherence to collagenous tissues, such as cartilage, a process that is important in the pathogenesis of septic arthritis caused by Staphylococci (37). The strain VPI 9048 toxin locus also includes a gene that codes for a probable GNAT family acetyl transferase (*H476_0290*). Such proteins include aminoglycoside acetyl transferases that confer resistance to the antibiotics kanamycin and gentamycin. The gene VPI9048- *H476_0305* appears to encode a cell wall protein with a SPR domain and may be involved in the invasion of host cells.

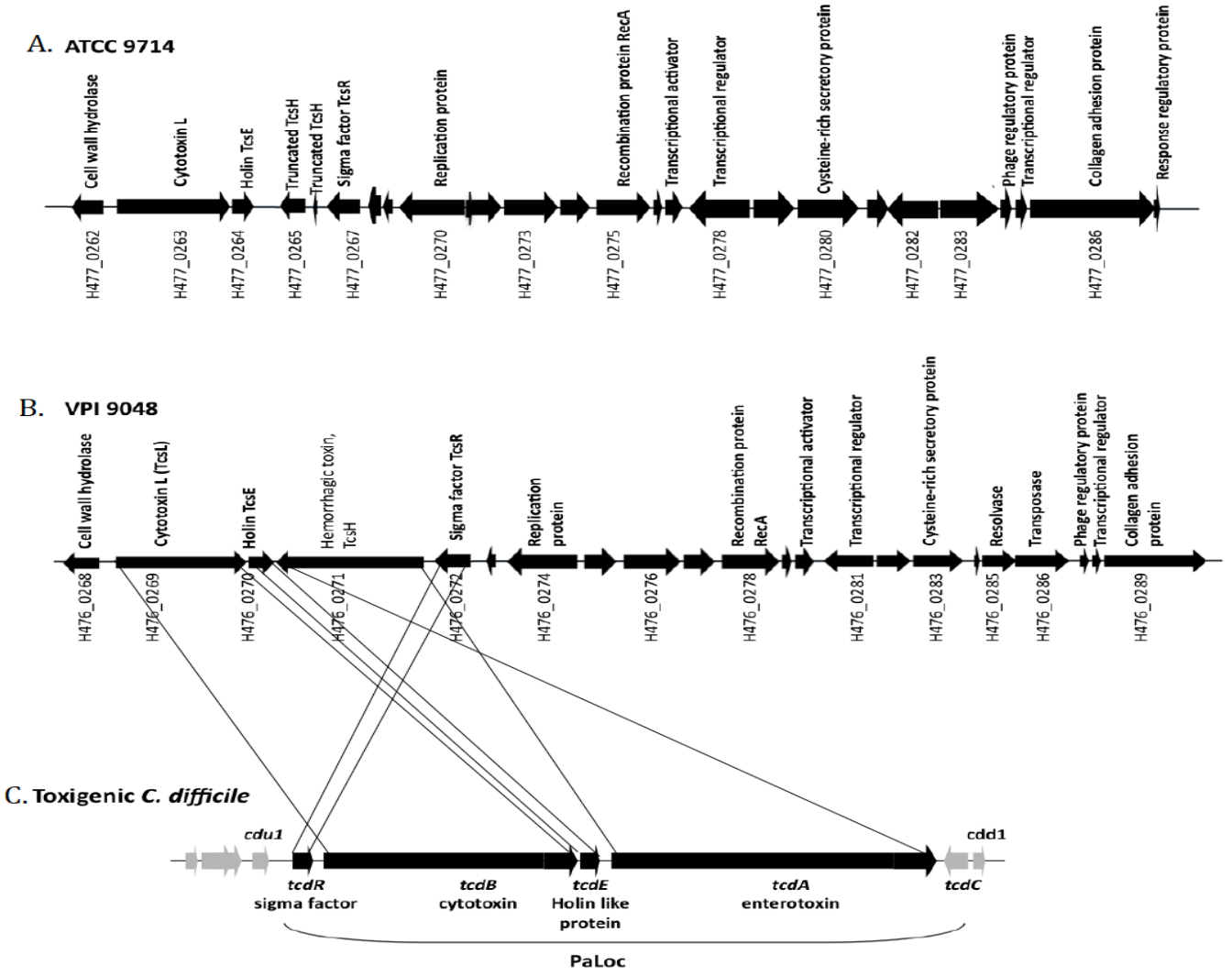

A. ATCC 9714
Cell wall hydrolase
Cytotoxin L
Holin TcsE
Truncated TcsH
Truncated TcsH
Sigma factor TcsR
Replication protein
Recombination protein RecA
Transcriptional activator
Transcriptional regulator
Cysteine-rich secretory protein
Phage regulatory protein
Transcriptional regulator
Collagen adhesion protein
Response regulatory protein
H477_0262
H477_0263
H477_0264
H477_0265
H477_0267
H477_0270
H477_0273
H477_0275
H477_0278
H477_0280
H477_0282
H477_0283
H477_0286
B. VPI 9048
Cell wall hydrolase
Cytotoxin L (TcsL)
Holin TcsE
Hemorrhagic toxin, TcsH
Sigma factor TcsR
Replication protein
Recombination protein RecA
Transcriptional activator
Transcriptional regulator
Cysteine-rich secretory protein
Resolvase
Transposase
Phage regulatory protein
Transcriptional regulator
Collagen adhesion protein
H476_0268
H476_0269
H476_0270
H476_0271
H476_0272
H476_0274
H476_0276
H476_0278
H476_0281
H476_0283
H476_0285
H476_0286
H476_0289
C. Toxigenic C. difficile
cdu1
cdd1
tcdR
sigma factor
tcdB
cytotoxin
tcdE
Holin like protein
tcdA
enterotoxin
tcdC
PaLoc

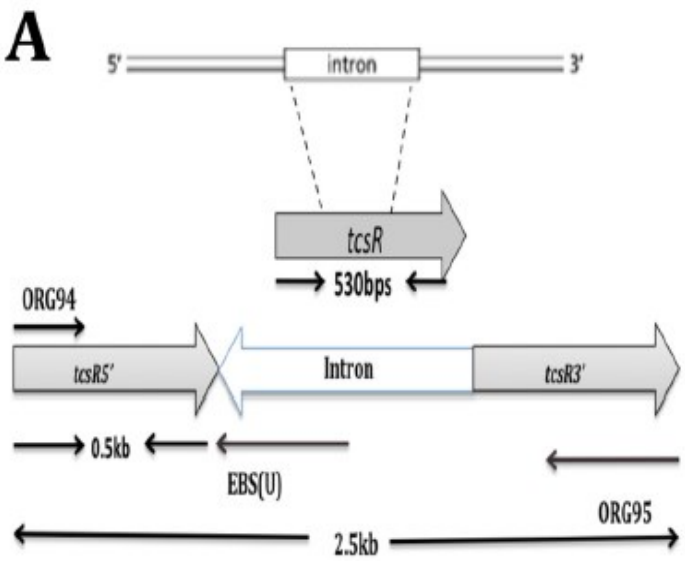
Figure 2.1 Pathogenecity locus in *C. sordellii* and *C. difficile*

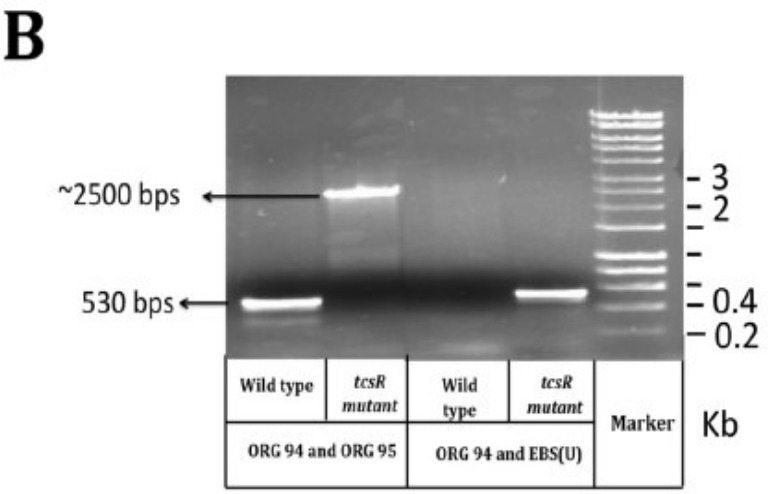
(A and B) Genetic organization of genes near *tcsL* and *tcsH* in *C. sordellii* strains ATCC 9714(A) and VPI 9048 (B). (c) Genetic organization of genes in the *C. difficile* PaLoc. The ORFs are indicated by arrowheads pointing in the direction of transcription. Black lines drawn between the genes of *C. sordellii* (VPI 9048) and the *C. difficile* PaLoc represent regions with sequence similarity between the two bacteria.

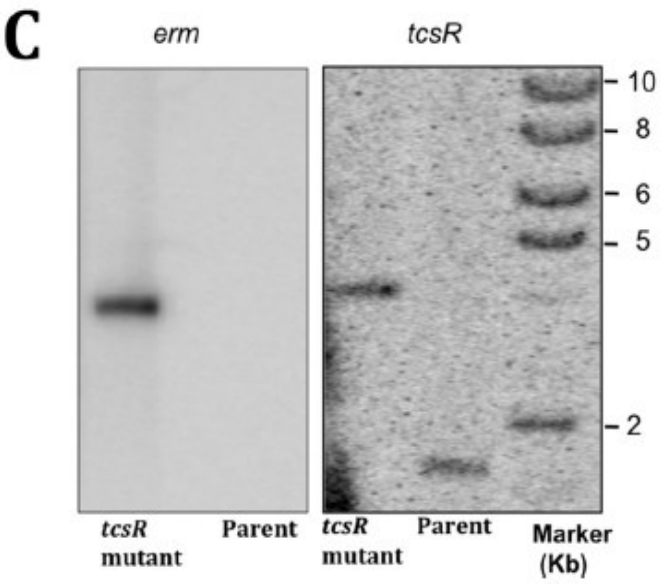
Mutation in *tcsR* affects cytotoxin production in *C. sordellii*

We initiated functional characterization of the toxin loci genes by characterizing the *tcsR* gene that is present downstream of *tcsL*. TcsR appears to be a 174-residue protein that shows homology to *C. difficile* TcdR, an alternative sigma factor that drives transcription of the toxin genes *tcdA* and *tcdB* (11, 28). To determine whether TcsR is necessary for high-level expression of *C. sordellii tcsL*, a *tcsR* mutant was created in the ATCC9714 strain using the ClosTron technique (23). Intron insertion sites in *tcsR* gene were identified using the Perutka algorithm available at ClosTron.com. The group II intron in plasmid pMTL007C-E5 was retargeted to integrate within the *tcsR* coding sequence at position 342 bp on the DNA sense strand (Figure 2A). The *tcsR*-retargeted plasmid pTUM007: Cdi-tcsR-342 was introduced into *C. sordellii* ATCC9714 by conjugation with E. coli and thioamphenicol-resistant transconjugants were selected. We could not introduce plasmids into VPI 9048 even with repeated attempts. Hence our *tcsR* characterization study was limited to the ATCC9714 strain. To confirm the successful inactivation of *tcsR* gene, PCR was performed using gene-specific primers (ORG94 and ORG95) and an intron-specific primer (EBS Universal primer). When tcsR-specific primers were used, amplified bands of 2.5 Kb and 0.5 Kb were obtained from the *tcsR* mutant and parent strains, respectively (Figure 2B). The presence of the 2.5 Kb amplification product indicates the presence of a 2.0 Kb intron within the *tcsR* gene. PCR was also performed using the intron specific primer EBS-universal and the *tcsR*-specific primer ORG94. A PCR product of 0.5 Kb was observed only in the *tcsR* mutant (Figure 2B). Further, Southern blot hybridizations were performed to confirm single 335 integration site of the group II intron within the *tcsR* gene in the mutant strain chromosome. Chromosomal DNA from strain ATCC9714 and its *tcsR* mutant strain were digested with EcoRV and subjected to Southern blot hybridization using ^{32}-P labeled

tcsR and *ermB* probes. As expected, the *tcsR* probe hybridized with both the mutant and the parent strains. In the parent strain a band at 1.9 Kb could be observed and in mutant the probe hybridized with a band of 4 Kb, consistent with the insertion of the intron into the *tcsR* gene. The intron-specific *ermB* probe hybridized only with the *tcsR* mutant strain in the same 4 Kb band, further confirming the presence of the intron within the *tcsR* gene (Figure 2C). Growth curves of parent and the *tcsR* mutant strains in TY medium over a 24-hour period were essentially identical (Figure 2D). To see if TcsR plays a role in toxin production, toxin ELISAs were performed with equal amounts of cytosolic proteins (50 ug/ well) from the 10 hour-old parent and *tcsR* mutant strains. Absorbance recorded at 405 nm represents the toxin titer. The absorbance for mutant strain samples was approximately four- fold lower than for the parent strain. This result suggests that TcsR is required for maximal toxin production in *C. sordellii* (Figure 3A).


A
5'
intron
3'
tcsR
530bps
ORG94
tcsR5'
Intron
tcsR3'
0.5kb
EBS(U)
ORG95
2.5kb


B
~2500 bps
530 bps
3
2
0.4
0.2
Wild type
tcsR mutant
Wild type
tcsR mutant
Marker
Kb
ORG 94 and ORG 95
ORG 94 and EBS(U)


C
erm
tcsR
10
8
6
5
2
tcsR mutant
Parent
tcsR mutant
Parent
Marker (Kb)

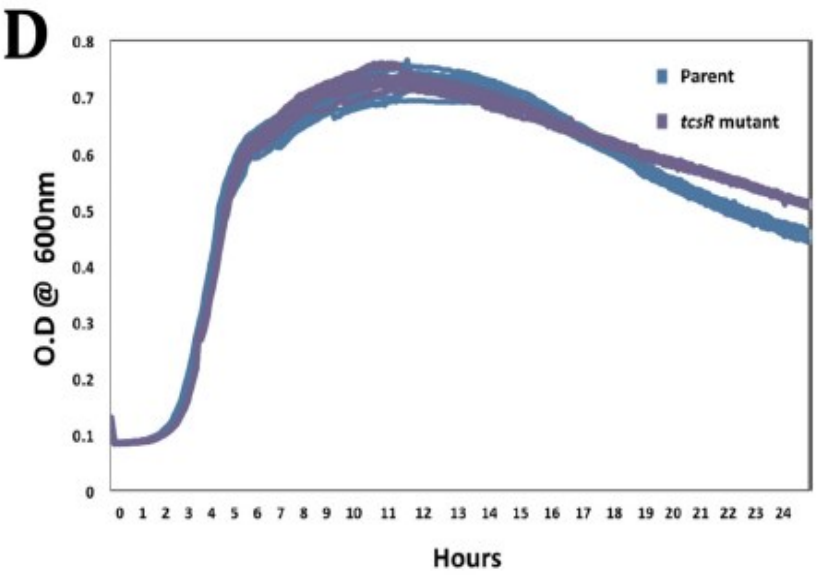

D
O.D @ 600nm
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Parent
tcsR mutant
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Hours

Figure 2.2 Construction and characterization of the *tcsR* mutant in *C. sordellii* ATCC 9714. (A) Schematic representation of ClosTron (group II intron)-mediated disruption of the *tcsR* gene in *C. sordellii*. (B) PCR verification of the intron insertion, conducted with gene-specific primers ORG94 and ORG95 or the intron-specific primer EBS universal [EBS(U)] with ORG94. (C) Southern blot analysis of genomic DNA from *C. sordellii* ATCC 9714 and *tcsR* mutant strains with *erm* (intron-specific) and *tcsR* probes. Chromosomal DNA was digested with EcoRV. (D) Growth curves of parent ATCC 9714 and the *tcsR* mutant in TY medium.

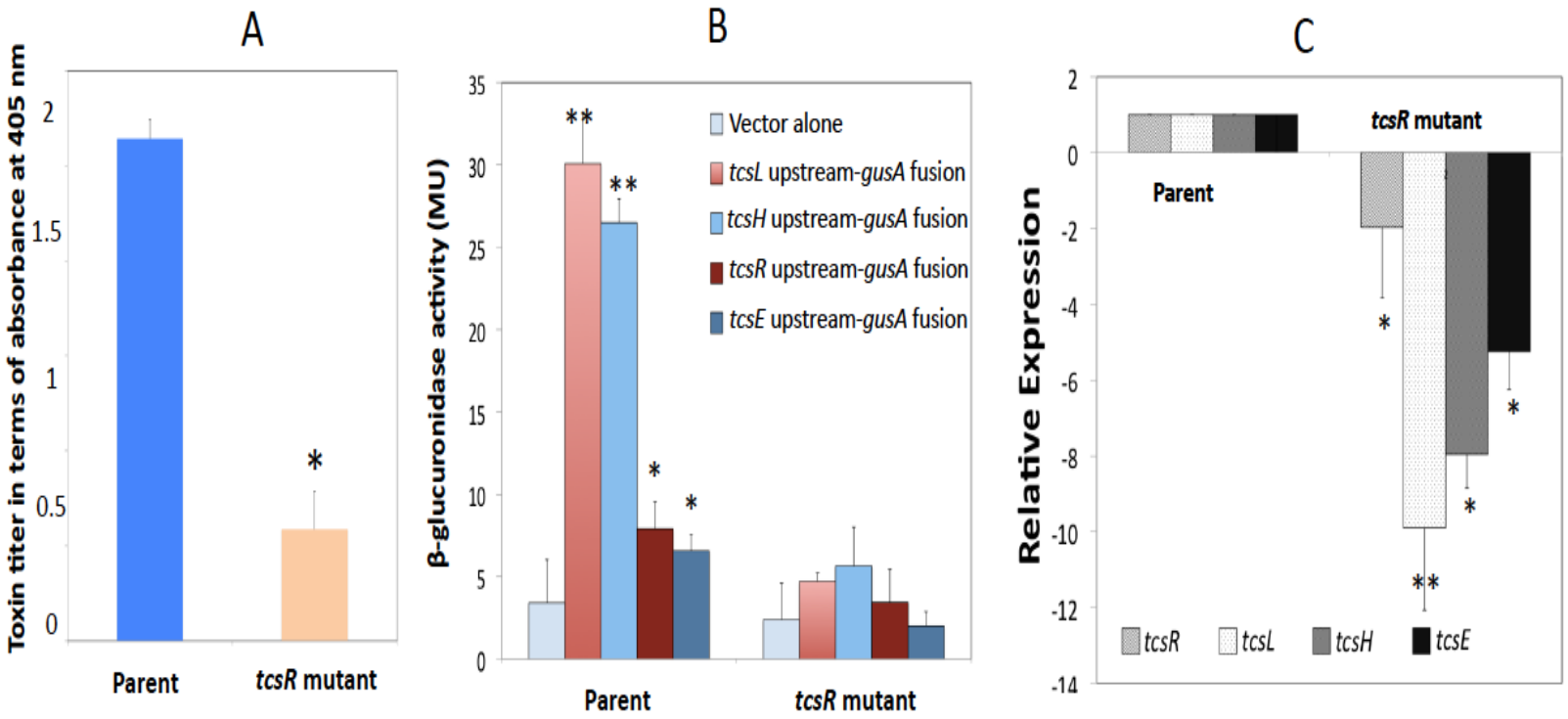

A
Toxin titer in terms of absorbance at 405 nm
2
1.5
1
0.5
0
*
Parent
tcsR mutant
B
β-glucuronidase activity (MU)
35
30
25
20
15
10
5
0
**
**
*
*
Vector alone
tcsL upstream-gusA fusion
tcsH upstream-gusA fusion
tcsR upstream-gusA fusion
tcsE upstream-gusA fusion
Parent
tcsR mutant
C
Relative Expression
2
0
-2
-4
-6
-8
-10
-12
-14
Parent
tcsR mutant
*
**
*
*
tcsR
tcsL
tcsH
tcsE

Figure 2.3 TcsR mediates the transcription of toxin genes

(A) Quantification of toxins in parent ATCC 9714 and *tcsR* mutant strains. Toxin titers in cytoplasmic proteins harvested from parent and *tcsR* mutant were determined by ELISA, and the signal from the test was recorded as the absorbance at 405 nm. The data shown are means ± standard errors of three replicate samples. Student's *t* test was used for statistical analysis. *, $P < 0.05$. (B) Expression of β-glucuronidase in parent ATCC 9714 and *tcsR* mutant strains carrying plasmids with *gusA* as the reporter gene fused to the promoters of *tcsL*, *tcsH*, *tcsR*, and *tcsE*. Strains carrying a promoterless *gusA* plasmid (pRGL153A) were used as control. Data represent the means ± standard errors of the means (SEM) ($n = 3$). (C) Comparison of transcript levels of *tcsL*, *tcsH*, *tcsR*, and *tcsE* in parent and *tcsR* mutant strains based on QRT-PCR. Data represent the mean fold change in expression ± SEM ($n = 3$) compared to the parent ATCC 9714 strain. Ten-hour-old bacterial cultures were used in all the experiments presented.

TcsR affects *tcsL* transcription in *C. sordellii*

To verify that TcsR regulates toxin gene transcription in *C. sordellii*, a *tcsL* promoter-gusA fusion was constructed. A 600-bp region upstream of the tcsL gene was PCR-amplified and cloned in the vector pRGL153A to create plasmid pRGL100, which was then introduced into the parent strain ATCC9714 and its *tcsR* mutant strain by transconjugation. Similarly, constructs with *tcsH*, *tcsR* and *tcsE* promoter-*gusA* fusion constructs were also made and introduced into parent and *tcsR* mutant strains. Strains carrying promoter-*gusA* fusions (pRGL100, pRGL161, pRGL162, pRGL163) or vector alone (pRGL153A) were grown in TY medium with thioamphenicol and the beta-glucuronidase assay was performed using samples collected after 10 hour of growth (late exponential phase). A six-fold higher level of β- glucuronidase activity was recorded in the parent strain than in the *tcsR* mutant strain (Figure 3B). Similarly, a *tcsH-gusA* fusion was expressed at a 5-fold higher level in the parent strain than in the tcsR mutant. Approximately, two-fold higher expression of *tcsR-gusA* and *tcsE-gusA* fusions was recorded in parent strain than in the *tcsR* mutant. These results provided evidence that TcsR positively influence the transcription of *tcsL*, *tcsH*, *tcsR* and *tcsE* genes in *C. sordellii*.

QRT-PCR was also performed with the RNA extracted from 10hour cultures of the parent and *tcsR* mutant strains. In Q-RT PCR, the transcript levels for *tcsL* and *tcsH* were ten-fold and eight fold lower, respectively, in *tcsR* mutant compared to the parent strain (Figure 3C). We also compared the transcript levels of *tcsR* and *tcsE* in the *tcsR* mutant vs. the parent strain. *C. difficile* TcdR positively regulates its own production (28, 29). Six-fold and two-fold decreases in transcript levels of *tcsE* and *tcsR* were recorded in the *tcsR* mutant when compared to parent strain (Figure 3C). These results suggest that TcsR may activate its own transcription and of the

tcsE gene in *C. sordellii*.

Complementation of the *tcsR* mutant

To further confirm that disruption of the *tcsR* gene causes underexpression of the toxin genes, the *tcsR* mutant was complemented with the wild-type *tcsR* gene from ATCC9714. The *tcsR* gene was cloned under the control of a tetracycline-inducible promoter in the vector pRGL154 and the resulting construct, pRGL145-1, was then introduced into the *tcsR* mutant by conjugation. The *tcsR* mutant with pRGL154 alone served as a control in these experiments. Bacterial strains were grown in TY medium with thioamphenicol and at OD_{600nm} = 0.5, ATc was added to 50 ng/ml to induce the expression of TcsR. Three hour post-induction, bacterial cultures were harvested and equal amounts of cytosolic proteins (50 ug/well) from each strains were used for toxin ELISAs. The complemented *tcsR* mutant strain had a toxin titer nearly 20-fold higher than that of the *tcsR* mutant with the vector alone (Figure 4A).

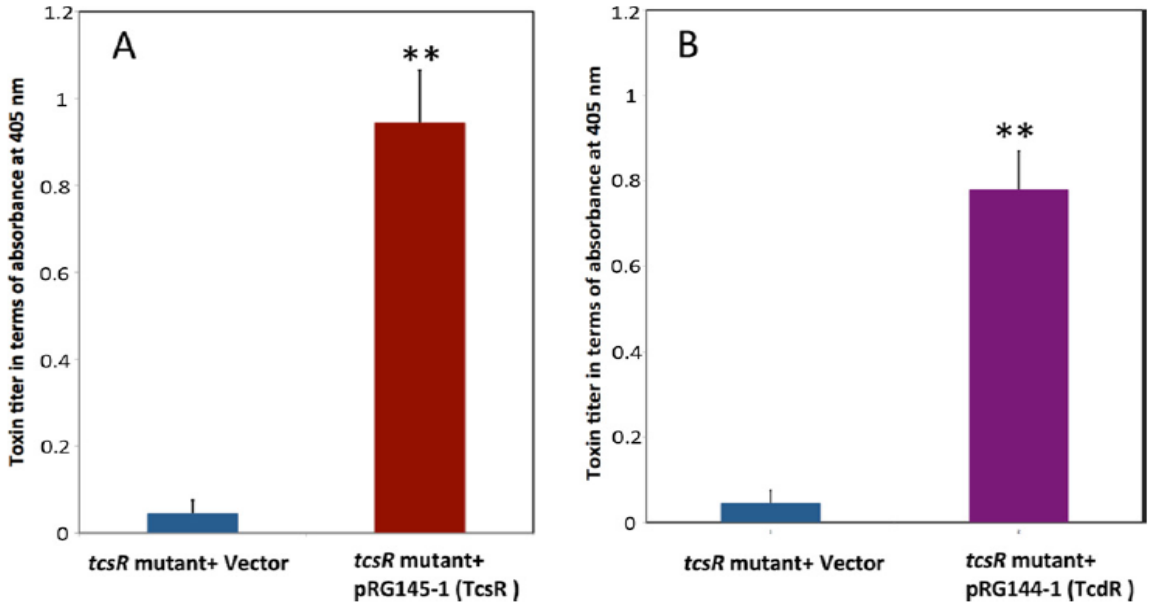


Figure 2.4 Complementation of *tcsR* mutant

Complementation of the *tcsR* mutant with *C. sordellii tcsR* (A) or *C. difficile tcdR* (B). The *tcsR* or the *tcdR* genes were cloned under a tetracycline-inducible promoter. The resulting plasmid constructs and the vector alone were introduced into the *tcsR* mutant for complementation. Bacterial cultures at an OD_{600} of 0.5 were induced for 4 h, and the toxins in the cytoplasm were quantified by ELISA. The signal from the test was recorded as the absorbance at 405 nm. The data shown are means ± standard errors of the means of three replicate samples. Student's *t* test was used for statistical analysis. **, $P \leq 0.01$.

C. difficile tcdR can complement the *C. sordellii tcsR* mutant

The Clostridial sigma factors TcdR, TetR, BotR and UviA are similar to the ECF sigma factor family (group 4 of the σ_{70}-family) but they differ enough in structure and function that they have been assigned to their own group (group 5) (10). Since they belong to a similar group, These sigma factors are interchangeable in terms of activation of transcription by RNA polymerase core enzyme in vitro and are partially interchangeable in vivo (9, 10). To determine whether TcdR is interchangeable with TcsR, we complemented the *C. sordellii* tcsR mutant with *C. difficile* tcdR. The tcdR gene was cloned under the control of the tetracycline-inducible promoter in pRGL154 and the resulting plasmid, pRGL144-1, was introduced into the *tcsR* mutant by conjugation. Cytosolic proteins (50 μg) collected from cultures that had been induced for three hours with ATc were tested for their toxin content using ELISA. The *C. sordellii tcsR* mutant complemented with *tcdR* produced nearly 16-fold more toxin than the control (*tcsR* mutant with vector alone) (Figure 4B). This result shows that TcdR can function in *C. sordellii* to drive transcription of the *tcsL* toxin gene, implying that TcsR is also a sigma factor. Promoters recognized by TcdR and its most closely related sigma factor, UviA, are thought to have a conserved "TTTACA" hexanucleotide motif in the -35 region and the sequence CTC/TTTT in the -10 region (10). The amino acid sequences of TscR and TcdR showed high conservation in the putative region 4.2, which interacts with the -35 sequence (Figure 5B). Moreover, the regions upstream of the *tcsL*, *tcsH*, *tcsE* and *tcsR* genes contain the highly conserved "TTTACA" sequence and less well-conserved potential -10 sequences (Figure 5A). Our complementation studies along with the sequence analysis suggests strongly that TcsR is a new member of the group 5 sigma factors, all of which to date have been discovered in Clostridium sp

A


TcsR regulated genes
-35
-10
+1
tcsH upstream ttttag TTTACA ttttgcttaaaatttat TTTTTA atataggg gtag...110bp...ATG
tcsL upstream agtatg TTTACA tattcatataaaaataa TTCTTA ctttatgtaagg...132bp...ATG
tcsE(upstream 1) tattat TTTACA gatgaatatatagcagc AACTAG ctcattgactat...269bp...ATG
tcsE(upstream 2) tagtag TTTACA tccttatatctatatga ATTCTA taataatataat....31bp...ATG
tcsR upstream tataat TTTACA tatgattataactctat TTTTAT aaagaaatttgt...109bp...ATG
TcsR consensus TTTACA TTTTTA
TcdR regulated genes
tcdA upstream gatatg TTTACA aattactatcagacaat CTCCTT atctaatagaag....56bp...ATG
tcdB upstream acaaag TTTACA tatttatttcagacaac GTCTTT attattcaatcg...162bp...ATG
TcdR consensus TTTACA CTCTTT

B

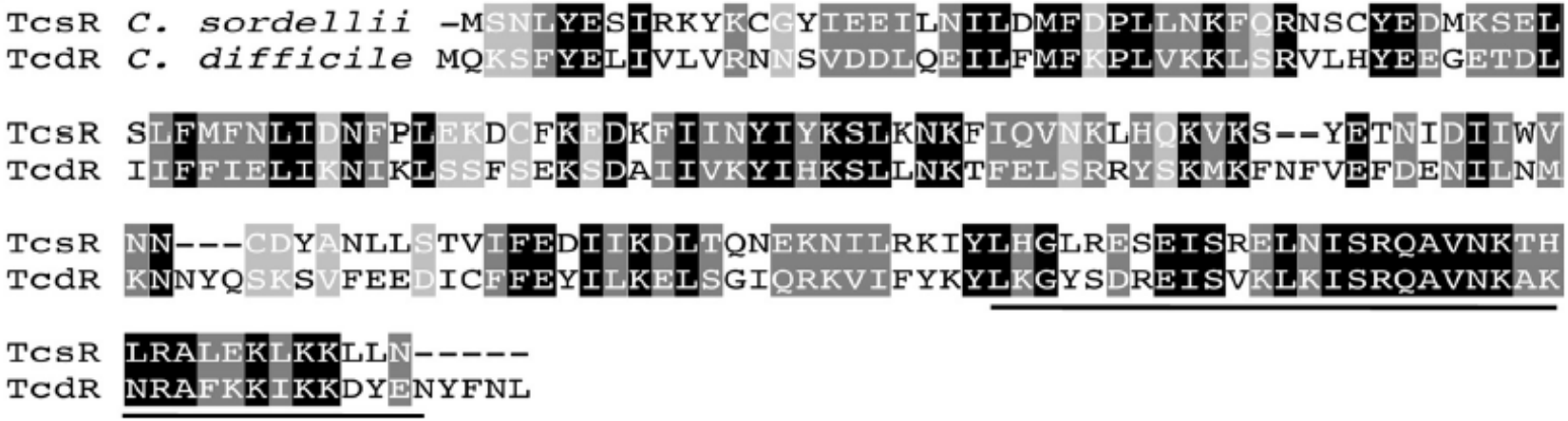

TcsR C. sordellii -MSNLYESIRKYKCGYIEEILNILDMFDPLLNKFQRNSCYEDMKSEL
TcdR C. difficile MQKSFYELIVLVRNNSVDDLQEILFMFKPLVKKLSRVLHYEEGETDL
TcsR SLFMFNLIDNFPLEKDCFKEDKFIINYIYKSLKNKFIQVNKLHQKVKS--YETNIDIIWV
TcdR IIFFIELIKNIKLSSFSEKSDAIIVKYIHKSLLNKTFELSRRYSKMKFNFVEFDENILNM
TcsR NN---CDYANLLSTVIFEDIIKDLTQNEKNILRKIYLHGLRESEISRELNISRQAVNKTH
TcdR KNNYQSKSVFEEDICFFEYILKELSGIQRKVIFYKYLKGYSDREISVKLKISRQAVNKAK
TcsR LRALEKLKKLLN-----
TcdR NRAFKKIKKDYENYFNL

Figure 2.5 Sequence alignment

(A) Sequence alignment of the predicted promoter regions of the cytotoxin and hemorrhagic toxic genes (*tcsL* and *tcsH*), the regulator *tcsR*, and two predicted promoters of the *tcsE* gene. TcdR-regulated *tcdA*, *tcdB*, and *tcdR* promoters are also presented. Predicted −35 and −10 regions and the +1 site are boxed. The distance of the ATG codon from the +1 site is indicated. The promoter consensus sequences recognized by TcsR and TcdR sigma factor are shown at the bottom. (B) Alignment of TcsR from *C. sordellii* with TcdR from *C. difficile*. Shaded in black are identical residues, shaded in gray are similar residues, and the dashes represent gaps in the alignment. Underlined are the region 4.2 sequences of the group 5 sigma factors that are predicted to interact with the highly conserved −35 region.

Discussion

Clostridium sordellii is known to cause lethal infections in animals and humans, in women undergoing medically induced abortions with mifepristone/misoprostol (15, 33). Incidences of *C. sordellii* infections in intravenous heroin users have also been reported (24). Although *C. sordellii* infections are relatively rare in humans, the high mortality rate (approximately 70%) associated with these infections makes *C. sordellii* one of the important pathogens requiring more attention (1).

In this study, we present the draft genome sequences of *C. sordellii* strains, ATCC9714 and VPI9048. High throughput sequencing and subsequent assembly generated 104 and 166 contigs of ATCC9714 and VPI9048 genomes respectively. Gaps in genome coverage were not filled in with manual sequencing due to resource constraints. This approach is consistent with de novo sequencing and the publication of other pathogen genomes, given that the length of the draft genomes were consistent with other sequenced Clostridial genomes (Table 2) and that the two strains whose genome sequences reported here are vastly similar. Gaps are typically caused by large (greater than the library "insert" size) fragments, which tend to be rRNA operons, large mobile elements or duplicated regions and will likely do not materially detract from the quality of the data analysis presented here. The nearly complete genome sequences of the two *C. sordellii* strains that we report here will help in the further characterization of *C. sordellii* pathogenesis. While functional analysis is ultimately required to elucidate the roles of individual genes in pathogenesis, sequence information can assist greatly in this effort.

Similar to those of many related Clostridial pathogens, the *C. sordellii* genome is highly A+T-rich (G+C content = 27%). Manatee annotation analysis of the *C. sordellii* genomes showed that only 1% of the genome represents mobile genetic elements. This feature of the *C. sordellii* genome is in drastic contrast to its close relative *C. difficile*, where nearly 11% of the genome consists of mobile genetic elements (38). The mobile genetic elements in the ATCC 9714 and VPI9048 genomes are primarily cryptic phages. In many *C. difficile* strains and in *Bacillus subtilis*, a skin element (sigK intervening sequence) is inserted within the gene *sigK* that codes for a sporulation-specific sigma factor (22, 41). Skin is required for efficient sporulation and its excision occurs at the onset of sporulation in *C. difficile* (22). Skin is absent in *C. sordellii*; the SigK-encoding ORF (VPI 9048 -*H476_1977*; ATCC 9714- *H477_0922*) is intact without any insertion elements. *C. difficile*, in the presence of glucose and other metabolizable sugars down regulates toxin gene transcription (11, 29). Carbon catabolite respressor, CcpA is responsible for this carbon catabolite repression (CCR) of toxin genes in *C. difficile* (2). *C. sordellii* carries a CcpA encoding gene (VPI 9048- *H476_1939*; ATCC 9714-*H477_3163*), which is 80% identical to the *C. difficile* CcpA. Preliminary data from our lab show that toxin transcription in *C. sordellii* is also under CCR (data not shown) and suggests a possible role for CcpA-mediated transcriptional repression of toxin genes. CodY is another global regulator that controls gene expression in response to nutrient availability (40). CodY is widely present in many Gram-positive bacterial pathogens, including *C. difficile* (8), *Staphyloccocus aureus* (34) and *Listeria monocytogenes* (27). A *C. difficile codY* mutant expresses high levels of toxins during growth in rich medium and CodY was found to repress the expression of *tcdR*, the alternate sigma factor that is specific for the toxin genes (8). *C. sordellii* also encodes a CodY homolog (VPI9048-*H476_0550*; ATCC9714- *H477_0862*), which is 78% identical to *C. difficile* CodY and may play

a role in virulence gene expression.

For many years, the lack of molecular and genetic tools to manipulate *C. sordellii* has made it difficult to study the importance of potential virulence factors. Recently, the cytotoxin TcsL encoding gene in *C. sordellii* was inactivated using Targetron technology (21), which demonstrates the feasibility of genetic manipulations in *C. sordellii*. Using similar technology, in this work we have inactivated the putative sigma factor-encoding gene, *tcsR*, and have shown that TcsR is needed for the transcription of the toxin-encoding genes *tcsL* and *tcsH*. Our promoter reporter fusion studies and the quantitative real-time PCR analysis provided evidence that TcsR is required for the transcription of genes *tcsL*, *tcsH*, *tcsR* and *tcsE*. Genetic experiments in this study showed that TcsR and TcdR are interchangeable to regulate toxin gene transcription on *C. sordellii* and suggests that TcsR to be the new member in the group 5 of Clostridial sigma factors. Further, in vitro transcription experiment that test the ability of purified TcsR and RNA polymerase core enzyme to initiate transcription from the *tcsL* or *tcsH* promoters are needed to conclusively prove TcsR as a sigma factor. Such experiments are currently under progress in our lab.

In conclusion, we have sequenced and presented genomes of two *C. sordellii* strains. Sequencing information revealed the toxin genes *tcsL* and *tcsH* to be part of a region that shows signatures of an integrative conjugative element. Mutational and computational analysis revealed TcsR to belong to the group 5 sigma factors. Availability of genome sequence from these two different *C. sordellii* strains now facilitates the identification of more virulence-associated factors in this pathogen. More functional studies on these putative virulence factors in *C. sordellii* may well

help us to determine their contribution to bacterial pathogenesis.

References

1. Aldape, M. J., A. E. Bryant, and D. L. Stevens. 2006. *Clostridium sordellii* infection: epidemiology, clinical findings, and current perspectives on diagnosis and treatment. Clin Infect Dis 43:1436-1446.

2. Antunes, A., I. Martin-Verstraete, and B. Dupuy. 2011. CcpA-mediated repression of *Clostridium difficile* toxin gene expression. Mol Microbiol 79:882-899.

3. Aronoff, D. M., Y. Hao, J. Chung, N. Coleman, C. Lewis, C. M. Peres, C. H. Serezani, G. H.Chen, N. Flamand, T. G. Brock, and M. Peters-Golden. 2008. Misoprostol impairs female reproductive tract innate immunity against *Clostridium sordellii*. J Immunol 180:8222-8230.

4. Braun, V., T. Hundsberger, P. Leukel, M. Sauerborn, and C. von Eichel-Streiber. 1996. Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. Gene 181:29-38.

5. Bruggemann, H. 2005. Genomics of clostridial pathogens: implication of extrachromosomal elements in pathogenicity. Curr Opin Microbiol 8:601-605.

6. Carter, G. P., M. M. Awad, Y. Hao, T. Thelen, I. L. Bergin, P. M. Howarth, T. Seemann, J.I. Rood, D. M. Aronoff, and D. Lyras. 2011. TcsL is an essential virulence factor in *Clostridium sordellii* ATCC 9714. Infect Immun 79:1025-1032.

7. Clark, S. 2003. Sudden death in periparturient sheep associated with *Clostridium sordellii*. Vet Rec 153:340.

8. Dineen, S. S., A. C. Villapakkam, J. T. Nordman, and A. L. Sonenshein. 2007. Repression of *Clostridium difficile* toxin gene expression by CodY. Mol Microbiol 66:206-219.

9. Dupuy, B., N. Mani, S. Katayama, and A. L. Sonenshein. 2005. Transcription activation of a

UV-inducible *Clostridium perfringens* bacteriocin gene by a novel sigma factor. Mol Microbiol 55:1196-1206.

10. Dupuy, B., and S. Matamouros. 2006. Regulation of toxin and bacteriocin synthesis in Clostridium species by a new subgroup of RNA polymerase sigma-factors. Res Microbiol 157:201-205.

11. Dupuy, B., and A. L. Sonenshein. 1998. Regulated transcription of *Clostridium difficile* toxin genes. Mol Microbiol 27:107-120.

12. Eklund, M. W., F. T. Poysky, M. E. Peterson, and J. A. Meyers. 1976. Relationship of bacteriophages to alpha toxin production in *Clostridium novyi* types A and B. Infect Immun 14:793-803.

13. Eklund, M. W., F. T. Poysky, S. M. Reed, and C. A. Smith. 1971. Bacteriophage and the toxigenicity of *Clostridium botulinum* type C. Science 172:480-482.

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

15. Fischer, M., J. Bhatnagar, J. Guarner, S. Reagan, J. K. Hacker, S. H. Van Meter, V. Poukens, D. B. Whiteman, A. Iton, M. Cheung, D. E. Dassey, W. J. Shieh, and S. R. Zaki. 2005. Fatal toxic shock syndrome associated with *Clostridium sordellii* after medical abortion. N Engl J Med 353:2352-2360.

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

17. Govind, R., J. A. Fralick, and R. D. Rolfe. 2006. Genomic organization and molecular characterization of *Clostridium difficile* bacteriophage PhiCD119. J Bacteriol 188:2568-2577.

18. Green, G. A., V. Schue, R. Girardot, and H. Monteil. 1996. Characterisation of an

enterotoxin-negative, cytotoxin-positive strain of *Clostridium sordellii*. J Med Microbiol 44:60-64.

19. Hall, I. C., and J. P. Scott. . 1927. *Bacillus sordellii*, a cause of malignant edema in man. . J. Infect. Dis. 41:329-335. .

20. Hall IC, S. J. 1927. *Bacillus sordellii*, a cause of malignant edema in man. . Journal of Infectious Diseases 41:329-335.

21. Hao, Y., T. Senn, J. S. Opp, V. B. Young, T. Thiele, G. Srinivas, S. K. Huang, and D. M. Aronoff. 2010. Lethal toxin is a critical determinant of rapid mortality in rodent models of *Clostridium sordellii* endometritis. Anaerobe 16:155-160.

22. Haraldsen, J. D., and A. L. Sonenshein. 2003. Efficient sporulation in *Clostridium difficile* requires disruption of the sigmaK gene. Mol Microbiol 48:811-821.

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

24. Kimura, A. C., J. I. Higa, R. M. Levin, G. Simpson, Y. Vargas, and D. J. Vugia. 2004. Outbreak of necrotizing fasciitis due to *Clostridium sordellii* among black-tar heroin users. Clin Infect Dis 38:e87-91.

25. Lewis, C. J., and R. Naylor. 1996. Sudden death in lambs associated with *Clostridium sordellii* infection. Vet Rec 138:262.

26. Lewis, C. J., and R. D. Naylor. 1998. Sudden death in sheep associated with *Clostridium sordellii*. Vet Rec 142:417-421.

27. Lobel, L., N. Sigal, I. Borovok, E. Ruppin, and A. A. Herskovits. 2012. Integrative genomic analysis identifies isoleucine and CodY as regulators of *Listeria monocytogenes* virulence. PLoS

Genet 8:e1002887.

28. Mani, N., and B. Dupuy. 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.

29. Mani, N., D. Lyras, L. Barroso, P. Howarth, T. Wilkins, J. I. Rood, A. L. Sonenshein, and B. Dupuy. 2002. Environmental response and autoregulation of *Clostridium difficile* TxeR, a sigma factor for toxin gene expression. J Bacteriol 184:5971-5978.

30. Martinez, R. D., and T. D. Wilkins. 1992. Comparison of *Clostridium sordellii* toxins HT and LT with toxins A and B of *C. difficile*. J Med Microbiol 36:30-36.

31. Martinez, R. D., and T. D. Wilkins. 1988. Purification and characterization of *Clostridium sordellii* hemorrhagic toxin and cross-reactivity with *Clostridium difficile* toxin A (enterotoxin). Infect Immun 56:1215-1221.

32. Meridian Diagnositics. http://www.meridianbioscience.com/diagnostic-products/cdifficile/premier/premier-toxins-a-and-b.aspx. [Online.]

33. Miech, R. P. 2005. Pathophysiology of mifepristone-induced septic shock due to *Clostridium sordellii*. Ann Pharmacother 39:1483-1488.

34. Montgomery, C. P., S. Boyle-Vavra, A. Roux, K. Ebine, A. L. Sonenshein, and R. S. Daum. 2012. CodY deletion enhances in vivo virulence of community-associated methicillin-resistant *Staphylococcus aureus* clone USA300. Infect Immun 80:2382-2389.

35. Ortega, J., B. Daft, R. A. Assis, H. Kinde, L. Anthenill, J. Odani, and F. A. Uzal. 2007. Infection of internal umbilical remnant in foals by *Clostridium sordellii*. Vet Pathol 44:269-275.

36. Popoff, M. R. 1987. Purification and characterization of *Clostridium sordellii* lethal toxin and cross-reactivity with *Clostridium difficile* cytotoxin. Infect Immun 55:35-43.

37. Rich, R. L., B. Demeler, K. Ashby, C. C. Deivanayagam, J. W. Petrich, J. M. Patti, S. V.

Narayana, and M. Hook. 1998. Domain structure of the *Staphylococcus aureus* collagen adhesin. Biochemistry 37:15423-15433.

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

39. Smith, L. 1975. *Clostridium sordellii*, 2nd ed. Thomas Publishing, Springfield, IL.

40. Sonenshein, A. L. 2005. CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria. Curr Opin Microbiol 8:203-207.

41. Stragier, P., B. Kunkel, L. Kroos, and R. Losick. 1989. Chromosomal rearrangement generating a composite gene for a developmental transcription factor. Science 243:507-512.

42. Teng, F., B. E. Murray, and G. M. Weinstock. 1998. Conjugal transfer of plasmid DNA from *Escherichia coli* to enterococci: a method to make insertion mutations. Plasmid 39:182-186

Chapter 3 - Conclusion and Future studies

C. sordellii, an emerging pathogen is known to cause myonecrosis, sepsis and shock in humans and enteritis, omphalitis and myositis in animals. Increasing reports of *C. sordellii* infection in women of reproductive age are reported. The major virulence factors in *C. sordellii* are the hemorrhagic toxin and lethal toxin. These toxins belong to the large clostridial cytotoxin family and are glycosyltransferases. They act by inactivating the Ras and Rac proteins and increase the cell permeability resulting in fluid infiltration and inflammation. On sequencing the genome, similarities were found with other clostridium species. The PaLoc in *C. sordellii* was partially characterized and was found to be very similar to the PaLoc in *C. difficile.* The *C. sordellii* hemorrhagic and lethal toxin were found to be similar to the *C. difficile* toxin A and toxin B respectively. A sigma factor like protein encoding gene similar to *tcdR* was present in *C. sordellii.* This study shows that TcsR, is a sigma factor like TcdR and regulates the toxin gene production in *C. sordellii.* It was also shown to regulate other genes in the PaLoc as well. The *C. difficile* sigma factor TcdR was found to be auto regulatory and the regulation of TcsR in *C. sordellii* is yet to be studied. Preliminary data from our lab shows that TcdR plays a role in regulating sporulation in *C. difficile.* Similarly, studies have to be done to see if TcsR plays a role in regulating sporulation and other genes in *C. sordellii.* With the help of the genome sequences available, other virulence associated factors can be identified and studied.

Appendix A - Supplemental tables

Table A.1 Oligonucleotides used in the study:

Oligo Name	Sequence (5' to 3')	Description
ORG 210	GGTACCGCCTTTAAAATCCATTTATTTGGATTCTAAGGG	*tcsR* upstream (Forward with KpnI)
ORG 211	GAGCTC TACTCATATTATGCACTCCTTTATGG	*tcsR* upstream (Reverse with SacI)
ORG 208	GAGCTCATATAAGAGAGGATGATTTTATGC	*tcdR* ORF (Forward with SacI)
ORG 209	GGATCCTTAATGATGATGATGATGATGCAAGTTAAAATAATTTTC	*tcdR*-ORF (Reverse with BamHI)
ORG 203	GAGCTCACCATAAAGGAGTGCATAATTATGA	*tcsR* ORF (Forward with SacI)
ORG 204	GGATCCTTAATGATGATGATGATGATGATTTATTAATTTTTTTAG	tcsR ORF (Reverse with BamHI)
ORG 194	GGAATGGTACCTCACTTCCTCCATGACCTAA	*tcsL* upstream (Forward with KpnI)
ORG 195	GGAATGAGCTCTAAAATTCCCCCTTTATATAACAT	*tcsL* upstream (Reverse with SacI)
ORG 197	ACGGCTCAACCGTAGTAAGC	16srRNA (Q-RT PCR-Forward)
ORG 198	TCTACGCATTTCACCGCTAC	16srRNA (Q-RT PCR- Reverse)

ORG 192	GGGATGTTTGAGGGCAATAA	*tcsE* (Q-PCR RT- Forward)
ORG 193	TTCCTCCATTGATTCCAAAACT	*tcsE* (Q-PCR RT- Reverse)
ORG 190	CATGGACTAAGGGAATCTGAAA	*tcsR* (Q-PCR RT- Forward)
ORG 191	TTTTCCAAAGCTCTTAAATGTGTTT	*tcsR* (Q-PCR RT- Reverse)
ORG 186	CAGCAGCTTCAACTGCAATC	*tcsL* (Q-PCR RT-Forward)
ORG 187	CTGAAATTCCTGCCAAAGGA	*tcsL* (Q- PCR RT-Reverse)
ORG 310	TTTAGCTGCAGCATCTGATTT	*tcsH* (Q-PCR RT- Forward)
ORG 311	TAAATCTGGTTGTATCCCTGGC	*tcsH* (Q- PCR RT – Reverse)
ORG 17	GGTCGGTACCATGGACCCAAGAGATGCTGGTGCTTCT	*erm B* (Forward)
ORG 18	GCTAGAGCTCGAACGCGTGCGACTCATAGAATTATTTCCT	*erm B* (Reverse)
ORG 292	GGTACCTTAAATAAATTTCAAAGAAATAGTTGC	*tcsH* upstream ((Forward with KpnI)
ORG 293	GAGCTCAAGACCCTCCTAATATTATTATTTTGGGC	*tcsH* upstream (Reverse with SacI)
ORG 294	GGTACCGGTATAATGAAAACAGGGCTTATCATTTG	*tcsE* upstream (Forward with KpnI)
ORG 295	GAGCTCAATGCCTCCTAGTTTTGTAATTAAAATAAT	*tcsE* upstream (Reverse with SacI)

Table A.2 ATCC toxin locus

Locus tag	Length (aa)	Predicted product	Hit description	E-value	% identity	Conserved domains
H477_0243	149	leucine Rich repeat family protein	S4 domain protein [*Clostridium perfringens* D str. JGS1721 ZP_02954940.1	93-27	49% 71/145	No putative conserved domains
H477_0244	126	leucine Rich repeat family protein	uncharacterized protein LOC100215918 [*Hydra magnipapillata* XP_002169743.2	4.9	38% 29/77	LRR_4
H477_0245	126	Hypothetical protein	uncharacterized protein LOC100215918 [*Hydra magnipapillata*] XP_002169743.2	4.9	38% 29/77	No putative conserved domains
H477_0246	206	RNA polymerase sigma factor, sigma-70 family	putative RNA polymerase sigma factor SigV	33-24	33% 61/187	Sigma70_r2, r4 superfamily

			[*Clostridium sp*. D5] >gb|EGB91713.1 ZP_08131089.1			
H477_0247	216	EAL domain protein	cyclic diguanylate phosphodiesterase (EAL) domain protein [*Clostridium sp*. HGF2] >gb|EFR38126.1 ZP_07832144.1	2e-18	27% 60/221	EAL domain, bacterial signallling protein
H477_0248	252	RNA polymerase sigma factor, sigma-70 family	RNA polymerase sigma-70 factor [*Clostridium perfringens* D str. JGS1721]ZP_02953213.1	2e-22	38% 65/173	Sigma70_r2 superfamily, RNA polymerase sigma factor, sigma 70 region 4
H477_0249	59	Hypothetical protein	polysaccharide deacetylase family protein [*Pseudomonas fluorescens* Pf0-1] >gb|ABA72018.1 YP_346007.1	1.6	38% 18/47	No putative conserved domains
H477_0250	174	Hypothetical protein	sigma-70 family RNA	3e-07	28%	No putative conserved domain

			polymerase sigma factor [*Stomatobaculum longum*] >gb|EHO17516.1ZP_09521451.1		32/115	
H477_0251	158	bacterial Ig-like domain family protein"	bacterial group 3 Ig-like protein [*Clostridium celatum* DSM 1785] ZP_19299437.1	8e-08	56% 31/55	Big_3 Bacterial Ig like domain group III
H477_0252	316	actinobacterial surface-anchored domain protein	bacterial group 3 Ig-like protein [*Clostridium celatum* DSM 1785] >gb|EKY22587.1|ZP_19299437.1	2e-69	48% 153/325	Big_3 Bacterial Ig like domain group III
H477_0253	296	bacterial Ig-like domain family protein	chitinase B [*Clostridium paraputrificum*]BAA23796.1	7e-16	46% 65/141	No putative conserved domain
H477_0254	218	Hypothetical protein	hypothetical protein HMPREF9466_00181 [*Fusobacterium necrophorum* subsp. funduliforme 1_1_36S]	0.41	32% 33/102	No putative conserved domain

			ZP_09526148.1			
H477_0255	57	transcriptional regulator family protein	Rrf2 family transcriptional regulator [*Clostridium difficile* 630] >ref|ZP_05272825.1 YP_001089272.1	0.030	45% 21/47	RrF2 superfamily transcriptional regulator
H477_0256	93	Hypothetical protein	hypothetical protein phiCD27_gp59 [*Clostridium* phage phiCD27] YP_002290935.1	0.63	64% 16/25	No putative conserved domain
H477_0257	65	antidote-toxin recognition MazE family protein	hypothetical protein CdifQCD_20326 [*Clostridium difficile* QCD-37x79] ZP_05399442.1	6e-11	42% 25/60	No putative conserved domain
H477_0258	46	Hypothetical protein	unnamed protein product [*Babesia* microti strain RI CCF73377.1	2.2	43% 19/44	No putative conserved domain
H477_0259	278	parB-like nuclease domain protein	ParB protein [*Clostridium perfringens* C str. JGS1495]	2e-105	69% 164/237	ParBc superfamily

			>gb|EDS79302.1 ZP_02865546.1			
H477_0260	69	soj domain protein	Soj protein [*Clostridium perfringens* str. 13] >dbj|BAB62438.1 NP_149993.1	4e-15	54% 37/68	Soj, mind_arch, flhG domain
H477_0261	146	cobQ/CobB/MinD/ParA nucleotide binding domain	chromosome partitioning related protein [*Clostridium perfringens*] >ref|ZP_16429000.1 YP_209675.1	5e-73	76% 105/139	Plasmid partitioning protein, ParA, Fer4_NifH superfamily
H477_0262	220	cell wall hydrolase/autolysin	sporulation-specific N-acetylmuramoyl-L-alanine amidase (Cell wall hydrolase) [*Clostridium sp*. D5] ZP_08128126.1	3e-25	37% 74/200	N-Acetylmuramoyl-L-alanine amidase I , II
H477_0263	2364	cytotoxin L (TcsL)	*C. sordellii*-cytotoxin L CAA57959.1	0.0	99% 2358/ 2364	N terminal Glycosyltransferase domain

H477_0264	141	Holin (TcsE)	TcdE protein [*Clostridium difficile*] CAC79642.1	2e-66	73% 103/141	Phage Holin
H477_0265	60	Haemorrhagic toxin (TcsH (Truncated)	truncated toxin A [*Clostridium difficile*] AAC08437.1	3e-08	93% 26/28	TcdA_TcdB
H477_0266	142	Haemorrhagic toxin (TcsH (Truncated)	truncated toxin A [*Clostridium difficile*] AAC08437.1	1e-66	77% 110/142	Glycosyltransferase domain
H477_0267 (*tcsR*)	170	Sigma factor (TcsR)	TcdR [*Clostridium difficile*] ABI93806.1	2e-15	35% 60/170	Sigma 70- region 4
H477_0268	87	Hypothetical protein	conserved hypothetical protein [*Beggiatoa sp.* PS] >gb|EDN70470.1| ZP_01999532.1	0.36	34%	No putative conserved domain

H477_0269	60	Hypothetical protein	hypothetical protein LsueK3_08561 [*Lactobacillus suebicus* KCTC 3549] ZP_09451275.1	2.1	43% 16/37	No putative conserved domain
H477_0270	407	Initiator family Replication protein	RepB protein [*Clostridium difficile* 002-P50-2011] ZP_17071833.1	7e-56	36% 122/340	Rep_3 superfamily, Replication protein
H477_0271	63	Hypothetical protein	PAB-like protein [*Saccharomyces uvarum*] CAC87105.1	5.3	46% 13/28	No putative conserved domain
H477_0272	126	Conserved hypothetical protein	conserved hypothetical protein [*Clostridium difficile* NAP08] ZP_06892370.1	0.003	25% 25/110	DNA binding helix turn helix motif, intron encoded nuclease repeat motif NUMOD1 domain protein
H477_0273	370	Hypothetical protein	hypothetical protein CBCST_p6CBCSt0005	2e-09	48% 32/66	No putative conserved domain

			[*Clostridium botulinum*] ZP_11680199.1			
H477_0274	135	Hypothetical protein	waterborne settlement pheromone [*Amphibalanus amphitrite*] BAM34601.1	9.3	40% 17/42	DNA binding helix turn helix motif, intron encoded nuclease repeat motif, group I intron endonuclease
H477_0275	369	Recombination protein	recombinase A [*Clostridium difficile* 630] >ref|ZP_05271345.1 YP_001087824.1	0.0	80% 264/329	Recombination protein
H477_0276	48	Hypothetical protein	conserved hypothetical protein [*Clostridium difficile* NAP08] >ref|ZP_06903143.1 ZP_06892599.1	1.9	44% 16/36	No putative conserved domain
H477_0277	100	Hypothetical protein	adenylosuccinate synthetase [*Lactobacillus mali* KCTC 3596 = DSM 20444] ZP_09448373.1	1.2	38% 18/47	No putative conserved domain
H477_0278	325	Transcriptional regulator	Dna binding helix-turn-helix	2e-09	49%	DNA binding helix turn helix

			protein [*Clostridium difficile* 050-P50-2011] ZP_17073318.1		31/63	protein, transcription repressor and regulator protein
H477_0279	162	Hypothetical protein	Conserved hypothetical protein [*Clostridium perfringens* B str ATCC 3626 ZP_02636631.1]	1e-05	29% 32/110	No putative conserved domain
H477_0280	341	Cysteine rich secretory family Extracellular protein	SCP-like protein [*Clostridium difficile* 002-P50-2011] >ref|ZP_17073080.1| ZP_17069969.1	3e-06	28% 61/218	SCP-like Extracellular protein domain, Cysteine-rich secretory protein
H477_0281	79	Hypothetical protein	hypothetical protein CJD_A0353 [*Clostridium perfringens* D str. JGS1721] ZP_02954525.1	0.039	40% 20/50	Glycoside hydrolase GH99_GH71_like superfamily
H477_0282	177	Hypothetical protein	conserved hypothetical protein	Ie-10	37%	No putative conserved domain

			[*Clostridium perfringens* E str. JGS1987 ZP_02633035.1		46/125	
H477_0283	144	Hypothetical protein	uncharacterized protein LOC100816858 [Glycine max] XP_003529550.1	4.2	29% 24/84	No putative conserved domain
H477_0284	75	Regulatory protein	putative phage regulatory protein [*Clostridium difficile* ATCC 43255] ZP_05349371.1	4e-04	38% 24/64	Helix Turn Helix DNA binding protein, HTH_XRE superfamily, HipB
H477_0285	76	Regulatory protein	transcriptional Regulator, XRE family with Cupin sensor domain [*Clostridium perfringens* D str. JGS1721] ZP_02952728.1	2.3	31% 21/68	No putative conserved domain
H477_0286	1354	Collagen adhesion protein	collagen adhesin [*Clostridium perfringens* str. 13] >dbj|BAB62495.1 NP_150050.1	0	48% 496/1039	Cna peptidase protein

H477_0287	107	Hypothetical protein	deoxyguanosinetriphosphate triphosphohydrolase-related protein [*Streptococcus mutans* NV1996] >gb|EMC16034.1| ZP_21855083.1	9.6	31% 17/55	No putative conserved domain
H477_0288	120	Hypothetical protein	hypothetical protein AC3_A0045 [*Clostridium perfringens* E str. JGS1987] >gb|EDT14347.1 ZP_02632952.1	2e-06	34% 30/89	No putative conserved domain
H477_0289	84	Hypothetical protein	hypothetical protein Cspa_c57090 [*Clostridium saccharoperbutylacetonicum* N1-4(HMT)] >gb|AGF59434.1	93-06	35% 29/83	No putative conserved domain
H477_0290	213	Dna Topoisomerase family protein	putative DNA topoisomerase [*Clostridium difficile* QCD-97b34]ZP_05385472.1	6e-48	45% 96/213	Putative nucleotide, metal binding protein, TOPRIM_TopoIA_TopoIIITOP1Ac

						superfamily, topoisomerase primase, topoisomerase I, III
H477_0291	276	Dna Topoisomerase family protein	DNA topoisomerase type IA [*Clostridium difficile* 630] >emb|CAJ69289.1 YP_001088916.1	2e-29	33% 89/266	TOP1Ac, DNA topoisomerase 1, III
H477_0292	132	DNA topoisomerase	DNA topoisomerase [*Clostridium difficile* QCD-23m63] ZP_05400436.1	4e-17	40% 46/114	PRK08173, DNA topoisomerase III
H477_0293	161	Hypothetical protein	conserved hypothetical protein [*Clostridium perfringens* D str. JGS1721] >gb|EDT70393.1 ZP_02954621.1	4e-08	26% 37/144	No putative conserved domain
H477_0294	54	Hypothetical protein	hypothetical protein EFOG_03015 [*Enterococcus faecalis* X98] >gb|EEU94826.1	0.047	48% 15/31	No putative conserved domain

			ZP_05600032.1			
H477_0295	76	Putative membrane protein	conserved hypothetical protein [*Clostridium perfringens* D str. JGS1721] >gb|EDT72340.1 ZP_02952703.1	2E-12	50% 38/76	No putative conserved domain
H477_0296	94	Hypothetical protein	sulfate transporter [*Maribacter sp*. HTCC2170] >gb|EAR02051.1 YP_003861354.1	0.35	31% 28/89	No putative conserved domain
H477_0297	339	type IV secretory system Conjugative DNA	type IV secretory pathway, VirD4 component [*Clostridium perfringens* D str. JGS1721] >gb|EDT72265.1 ZP_02952628.1	2e-113	51% 192/377	RecA-like_NTPases superfamily Type IV secretory pathway, VirD4 components [Intracellular trafficking and secretion]
H477_0298	210	type IV secretory system Conjugative	Type IV secretory pathway, VirD4 component [*Clostridium perfringens* D str.	1e-94	66% 139/210	TraG-D_C superfamily, VirD4, type IV conjugal transfer coupling protein

		DNA	JGS1721] >gb|EDT72265.1 ZP_02952628.1			
H477_0299	45	No homology found	No homology found	--	--	No putative conserved domain
H477_0300	163	Hypothetical protein	putative membrane protein [*Clostridium perfringens* C str. JGS1495] >gb|EDS79403.1 ZP_02865646.1	4e-30	41% 60/148	No putative conserved domain
H477_0301	264	Hypothetical protein	putative membrane protein [*Clostridium perfringens* B str. ATCC 3626] >gb|EDT23194.1 ZP_02636580.1	6e-35	35% 110/312	No putative conserved domain
H477_0302	80	Hypothetical protein	hypothetical protein IGO_05453, partial [*Bacillus cereus* HuB5-5] >gb|EJQ78697.1 ZP_17525376.1	0.008	32% 31/96	No putative conserved domain

H477_0303	97	Hypothetical protein	conserved hypothetical protein [*Clostridium perfringens* B str. ATCC 3626] >gb|EDT23793.1| ZP_02635907.1	7e-17	46% 41/89	No putative conserved domain
H477_0304	47	Hypothetical protein	hypothetical protein E5Q_00477 [*Mixia osmundae* IAM 14324] GAA93831.1	0.084	47% 18/38	No putative conserved domain
H477_0305	141	Hypothetical protein	conserved hypothetical protein [*Clostridium perfringens* E str. JGS1987] >gb|EDT14223.1 ZP_02633032.1	4e-29	42% 60/143	No putative conserved domain
H477_0306	51	Hypothetical protein	conserved hypothetical protein [*Clostridium perfringens* D str. JGS1721] >gb|EDT72342.1 ZP_02952705.1	0.84	56% 14/25	No putative conserved domain

H477_0307	328	Hypothetical protein	conserved hypothetical protein [*Clostridium perfringens* B str. ATCC 3626] >gb|EDT23144.1 ZP_02636571.1	7e-100	51% 161/313	ATP-binding cassette domain of multidrug resistance protein-like transporters. AAA domain
H477_0308	151	zonular occludens toxin family protein	conserved hypothetical protein [*Clostridium perfringens* C str. JGS1495] >ref|ZP_16922172.1 ZP_02865588.1	9e-57	63% 91/148	Zot superfamily, AAA like domain, conjugal transfer ATP binding protein
H477_0309	221	Hypothetical protein	Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5 [*Chelonia mydas* EMP28438.1	2.0	30% 35/116	No putative conserved domain

Table A.3 VPI toxin locus

Locus tag	Length (aa)	Predicted product	Hit description	E-value	% identity	Conserved domains
H476_0259	497	Bacterial Ig-Like domain family protein	*Enterococcus faecalis* TX0470, collagen binding protein; |ZP_07761016.1|	3e-08	24% 107/449	No putative conserved domain
H476_0260	137	Transcriptional regulator	BadM/Rrf2 family transcriptional regulator [*Moorella thermoacetica* ATCC 39073 YP_430929.1	8e-18	31% 42/134	Rrf2 transcriptional regulator
H476_0261	120	Hypothetical protein	hypothetical protein CdifQCD-6_20123 ZP_05332106.1	0.031	31% 31/101	GIY-YIG_C term superfamily
H476_0262	42	Transposase IS200 like family protein	transposase [*Clostridium sp.* M62/1] ZP_06346833.2	4e-10	59% 24/41	Y1_Tnp superfamily
H476_0263	362	Transposase, IS605 OrfB family	transposase [*Clostridium perfringens* C str. JGS1495] ZP_02865789.1	0.0	73% 267/364	Putative transposase dna binding domain

H476_0264	65	Antidote-toxin recognition MazE family protein	putative PemI [*Clostridium difficile* ATCC 43255] ZP_05350752.1	9e-06	37% 21/57	No putative conserved domain
H476_0265	445	Partitioning protein, RepB family	partition protein [*Clostridium perfringens* E str. JGS1987] ZP_02630615.1	8e-115	62% 191/310	ParBc nuclease domain
H476_0266	250	Cell division protein	Sporulation initiation inhibitor protein soj [*Clostridium perfringens* C str. JGS1495] ZP_02865591.1	2e-125	70% 174/250	cobQ/CobB/MinD/ParA nucleotide binding domain ParA Fer4_NifH superfamily
H476_0267	44	Hypothetical protein	Hypothetical protein CAEBREN_04895 [*Caenorhabditis brenneri*] EGT33401.1	6.7	51% 19/37	No putative conserved domain
H476_0268	220	Cell wall hydrolase/autolysin	*Clostridium cellulovorans* 743B - cell wall hydrolase YP_003842332.1	2e-28	36% 81/226	N-Acetylmuramoyl-L-alanine amidase

H476_0269	2364	cytotoxin L (TcsL)	*C. sordellii*-cytotoxin L CAA57959.1	0.0	96% 2272/2364	N terminal Glycosyltransferase domain. Toxin B
H476_0270	141	Holin (TcsE)	TcdE protein [*Clostridium difficile]* CAC79642.1	1e-67	74% 103/141	Toxin secretion Phage lysis Holin family
H476_0271	2597	Hemorrhagic toxin, TcsH	*C. difficile* Toxin A (TcdA) QCD-66c26] ZP_05270743.1	0.0	78% 2055/2618	Glycosyltransferase domain
H476_0272	173	Sigma factor (TcsR)	TcdR [*Clostridium] difficile*] ABI93806.1	1e-14	34% 60/174	RNA polymerase Sigma 70 family- region 4
H476_0273	60	hypothetical protein	Hypothetical -*Lactobacillus suebicus* KCTC 3549 ZP_09451275.1	2.1	43% 16/37	No putative conserved domain
H476_0274	407	Replication protein	RepB protein [*Clostridium difficile* 002-P50-2011] ZP_17071833.1	9e-56	36% 122/340	Initiator Replication family protein
H476_0275	128	Hypothetical protein	conserved hypothetical protein	0.005	26%	Helix turn helix motif

			[*Clostridium difficile* NAP08] ZP_06892370.1		29/112	NUMOD1 domain protein
H476_0276	368	Hypothetical protein	hypothetical protein CBCST_p6CBCSt0005 [*Clostridium botulinum*] ZP_11680199.1	2e-09	48% 32/66	No putative conserved domain
H476_0277	135	Predicted protein	kelch-like protein 17-like [*Saccoglossus kowalevskii*] XP_002740860.1	8.1	36% 27/76	NUMOD1 domain protein, GIY-YIG_C term superfamily
H476_0278	368	Recombination protein	recombinase A [*Clostridium difficile* 630] >ref|ZP_05271345.1 YP_001087824.1	0.0	80% 268/336	Protein RecA like NTPases superfamily
H476_0279	48	Hypothetical protein	hypothetical protein CD196_1775 [*Clostridium difficile* CD196] >ref|YP_003218309.1 YP_003214800.1	0.64	42% 16/38	No putative conserved domain
H476_0280	101	Hypothetical protein	Transcriptional activator, putative [*Giardia lamblia* P15]	1.7	24% 25/106	No putative conserved domain

			EFO63198.1			
H476_0281	306	Transcriptional regulator	SOS-response transcriptional Lex A repressor [*Clostridium perfringens* D str. JGS1721] ZP_02952675.1	2e-22	31%	DNA binding helix turn helix protein, transcription repressor and regulator protein
H476_0282	162	Hypothetical protein	conserved hypothetical protein [*Clostridium perfringens* C str. JGS1495]ZP_02863539.1	3e-05	28% 31/110	No putative conserved domain
H476_0283	331	Extracellular protein	Putative calcium chelating exported protein [*Clostridium difficile* QCD-23m63] ZP_05402067.1	1.2	26% 71/275	Extracellular protein Cysteine rich secretory family protein
H476_0284	44	Hypothetical protein	hypothetical protein CLOHIR_00995 [*Clostridium hiranonis* DSM 13275] ZP_03293048.1	3.5	38% 15/40	No putative conserved domain
H476_0285	188	Resolvase	resolvase domain-containing protein [*Clostridium difficile*	2e-93	90% 132/146	Resolvase N terminal domain protein

			QCD-76w55] ZP_05358063.1			
H476_0286	387	Transposase	IS605 family transposase OrfB [*Clostridium difficile* BI1] YP_006197023.1	0.0	87% 335/387	Putative Transposase family protein
H476_0287	75	Regulatory protein	putative phage regulator [*Clostridium difficile* ATCC 43255] ZP_05349371.1	4e-04	38% 24/64	Helix turn helix proteins
H476_0288	76	Regulatory proteins	Transcriptional regulator [*Clostridium cellulolyticum*] YP_002507585.1	1.2	29% 16/56	Hypothetical protein
H476_0289	1357	Collagen adhesion protein	collagen adhesin [*Clostridium perfringens* str. 13] NP_150050.1	0.0	42% 553/1323	CnaB type peptidase domain protein
H476_0290	171	Hypothetical protein	conserved hypothetical protein [*Clostridium perfringens* E str. JGS1987]ZP_02633035.1	9e-06	36% 48/133	No putative conserved domain
H476_0291	160	Hypothetical protein	putative rhizopine uptake ABC	9.7	33%	No putative conserved

			transporter periplasmic solute-binding protein [*Clostridium butyricum* 5521] ZP_02948719.1		33/100	domain
H476_0292	216	Hypothetical protein	beta-lactamase [*Bacillus cereus* VD200]ZP_17634301.1	7.1	25% 32/126	No putative conserved domain
H476_0293	700	Topoisomerase	DNA topoisomerase [*Clostridium difficile* QCD-76w55]ZP_05358121.1	2e-105	34% 247/731	TOPRIM_TopoIA_TopoIII superfamily,TOP1Ac superfamily, DNA topoisomerase family protein
H476_0294	167	Hypothetical protein	hypothetical protein CLL_0015 [*Clostridium botulinum* B str. Eklund 17B] YP_001893638.1	6e-12	36% 43/119	No putative conserved domain
H476_0295	54	Hypothetical protein	hypothetical protein EFOG_03015 [*Enterococcus faecalis* X98] ZP_05600032.1	0.059	48% 15/31	No putative conserved domain
H476_0296	76	Putative membrane protein	conserved hypothetical protein [*Clostridium perfringens* D str. JGS1721]	2e-12	50% 38/76	No putative conserved domain

			>gb|EDT72340.1|ZP_02952703.1			
H476_0297	925	Type IV secretory system conjugative DNA	type IV secretory pathway, VirD4 component [*Clostridium perfringens* D str. JGS1721] ZP_02952628.1	0	47% 393/835	TraG-D_C superfamily, VirD4
H476_0298	682	Putative membrane protein	putative membrane protein [*Clostridium perfringens* C str. JGS1495] >gb|EDS79403.1 ZP_02865646.1	2e-90	36% 218/606	No putative conserved domain
H476_0299	97	Hypothetical protein	conserved hypothetical protein [*Clostridium perfringens* B str. ATCC 3626] >gb|EDT23793.1| ZP_02635907.1	1e-16	46% 41/89	No putative conserved domain
H476_0300	201	Hypothetical protein	hypothetical protein HA1_15800 [*Clostridium perfringens* F262] >gb|EIA15630.1 ZP_16922174.1	6e-38	44% 88/201	No putative conserved domain
H476_0301	59	Hypothetical protein	hypothetical protein CdifA_02598 [*Clostridium difficile* ATCC	1e-04	58% 22/38	No putative conserved domain

			43255] >ref|ZP_17069585.1 ZP_05349623.1			
H476_0302	635	Type IV secretion system coupling dna binding	type IV secretion system VirB4 component [*Bacillus thuringiensis* MC28] >gb|AFU17455.1 YP_006815694.1	5e-141	45% 223/497	Conjugal transfer ATP binding cassette domain of MRP, iron sulphur cluster transporters
H476_0303	54	Hypothetical protein	hypothetical protein CdifA_02598 [*Clostridium difficile* ATCC 43255] >ref|ZP_17069585.1 ZP_05349623.1	0.019	58% 19/33	No putative conserved domain
H476_0304	225	Hypothetical protein	endonuclease/exonuclease/phospha tase [*Ignavibacterium album* JCM 16511] >gb|AFH49241.1 YP_005846509.1	8.7	25% 38/151	No putative conserved domain
H476_0305	392	Cell wall binding protein	cell wall-binding protein [*Clostridium perfringens* str. 13] >dbj|BAB62482.1 NP_150037.1	6e-149	58% 225/387	Lysozyme like superfamily, nlpC/P60 family protein

H476_0306	125	SS Dna binding protein	single-strand binding protein [*Clostridium difficile* QCD-63q42] ZP_05332107.1	1e-15	38% 41/107	Single stranded DNA binding family protein
H476_0307	300	Hypothetical protein	hypothetical protein PBCN14 [*Clostridium perfringens*] >dbj|BAD90616.1 YP_209672.1	1e-22	30% 85/279	No putative conserved domain
H476_0308	77	ribbon-helix-helix , copG family protein	Fe-S oxidoreductase [*Streptococcus thermophilus* CNCM I-1630] >gb|EHE87156.1 ZP_12667090.1	3.3	39% 21/54	No putative conserved domain
H476_0309	89	ribbon-helix-helix , copG family protein	toxin-antitoxin system, antitoxin component, ribbon-helix-helix domain protein [*Enterococcus faecium* R497] ZP_18306856.1	5e-06	38% 28/74	No putative conserved domain
H476_0310	142	Hypothetical protein	peptidoglycan GlcNAc deacetylase [*Streptococcus pneumoniae* TIGR4] ZP_22472958.1	7.2	27% 29/106	No putative conserved domain

H476_0311	386	Hypothetical protein	putative conjugative relaxase [*Lactococcus garvieae*] >emb|CCF71034.1| YP_005352364.1	2e-63	32% 133/420	TraI, conjugal transfer relaxase
H476_0312	80	Hypothetical protein	Cell division cycle protein 123-like [*Cricetulus griseus*]EGV95431.1	.90	33% 27/78	No putative conserved domain
H476_0313	607	Hypothetical protein	LtrC-like protein [*Bacillus thuringiensis* MC28] >gb|AFU17432.1 YP_006815671.1	2e-102	50%	Domain of undefined function (DUF)955 superfamily
H476_0314	60	Hypothetical protein	conserved hypothetical protein [*Flavobacteria bacterium* MS024-2A] >gb|EEG41291.1 ZP_03702776.1	4.3	32% 15/47	No putative conserved domain
H476_0315	95	Hypothetical protein	hypothetical protein Clole_3081 [*Clostridium lentocellum* DSM 5427] >gb|ADZ84777.1	2e-15	44% 44/100	No putative conserved domain

			YP_004309975.1			
H476_0316	259	Hypothetical protein	hypothetical protein Clocel_0429 [*Clostridium cellulovorans* 743B] >ref|ZP_07630939.1 YP_003841970.1	3e-13	50% 37/74	No putative conserved domain
H476_0317	54	Hypothetical protein	cytochrome P450 monooxigenase GliC2 [*Arthroderma otae* CBS 113480] >gb|EEQ35852.1 XP_002842840.1	4.5	47% 15/32	No putative conserved domain
H476_0318	187	resolvase	resolvase domain-containing protein [*Clostridium difficile* QCD-66c26] >ref|ZP_05320915.1 ZP_05273846.1	4e-90	72% 128/179	Ser_Recombinase superfamily, Transposase, resolvase N terminal domain, SR_Res Inv
H476_0319	187	traB family protein	traB family protein [*Vibrio cholerae* HE-40] >ref|ZP_17738822.1 ZP_17735271.1	1e-04	25% 38/152	No putative conserved domain
H476_0320	109	ycxB like family	primase-like protein	3.1	29%	No putative conserved

		protein	[*Staphylococcus aureus*] ADN53668.1		23/80	domain
H476_0321	418	transposase,	transposase, IS605 OrfB family [*Clostridium difficile* 70-100-2010] >gb|EHJ40079.1 ZP_17077481.1	6e-116	45% 189/419	IS605 OrfB family
H476_0322	119	Hypothetical protein	permease [*Lactobacillus reuteri* CF48-3A] >ref|YP_004649442.1ZP_03975395.1	8.6	32% 13/61	No putative conserved domain
H476_0323	204	Hypothetical protein	hypothetical protein HMPREF9469_00031 [*Clostridium citroniae* WAL-17108] >gb|EHF01154.1| ZP_09056994.1	1e-34	38% 78/203	DUF3267 superfamily
H476_0324	77	Hypothetical protein	hypothetical protein CdifQCD_20341 [*Clostridium difficile* QCD-37x79]	4e-13	65% 31/48	No putative conserved domain

			ZP_05399445.1			
H476_0325	46	Peptidase M28 family protein	putative aminopeptidase [*Clostridium sp.* JC122] ZP_10929384.1	2e-05	65% 24/37	Zinc peptidase like superfamily M28
H476_0326	42	Peptidase M28 family protein	putative aminopeptidase [*Clostridium sp.* JC122]ZP_10929384.1	.001	54% 19/35	Zinc peptidase like superfamily M28