Predicted Sporulation Capacities and Molecular Characterization of 2006/2007 Manitoba TcdA$^+$/TcdB$^+$ *Clostridium difficile* Isolates

Vanessa M. Wolf

Supervisor: Dr. M. Graham (National Microbiology Laboratory)

A thesis submitted in partial fulfillment of the Honours Thesis Course (BIOL-4111/6)

Department of Biology

The University of Winnipeg

2010

Abstract

Clostridium difficile, the major causative agent of antibiotic-associated diarrhoea and pseudomembranous colitis, has recently emerged as an important source of healthcare-associated infections. The Gram-positive, spore-forming microbe is especially problematic because spores are resistant to standard hospital cleaning methods. Moreover, the protective gastrointestinal flora is compromised in persons subjected to antibiotic therapy, making them particularly susceptible to colonization by *C. difficile*. The RNA polymerase sigma factor-encoding gene *sigK* is essential for spore coat formation. Regulated excision of a prophage-like element (designated $skin^{Cd}$) from *sigK* is crucial for achieving proper activation of SigK and thus, impacts sporulation efficiency as it is required for sporulation. Polymerase chain reaction (PCR) was used to amplify intact *sigK*, circular fragment (CF; excised $skin^{Cd}$), and 3′ and 5′ *sigK::*$skin^{Cd}$ genetic fusions for DNA templates from 302 Manitoba strains. Based on interpretation of the PCR results, it has been possible to predict each strain's putative sporulation capacity. Follow-up studies are now being conducted to experimentally confirm these molecular-based sporulation capacity predictions. Additional molecular typing was performed for more detailed strain characterization and assessment of factors associated with sporulation capacity, such as sequence type and hospital source. The capacity to predict hypersporulating strains would be highly advantageous in healthcare environments.

Acknowledgements

I would like to thank my supervisor Dr. Morag Graham at the National Microbiology Laboratory (Genomics Core department) for guidance and insight through the course of this year; her unbelievable patience and kindness were invaluable and truly made the project a great success. To our collaborator Dr. Michelle Alfa and her "team" at the St. Boniface Research Institute for providing the *Clostridium difficile* strains along with performing follow-up *in vitro* spore counts. To Dr. Michael Mulvey, Jennifer Campbell and Tim Du at the National Microbiology Laboratory (Antimicrobial Resistance and Nosocomial Infections department) for providing initial study data on the provided strains along with teaching me how to use BioNumerics: Applied Maths. I would also like to thank Dr. Murray Wiegand, Dr. Anne Adkins and Ms. Dawn Rittberg at the University of Winnipeg Biology department for their feedback and allowing me to have this opportunity.

Table of Contents

Title page……..... i

Abstract……... ii

Acknowledgements……..... iii

Table of Contents……..... iv

List of Figures……..... v

List of Tables……... vi

List of Abbreviations……... vii

List of Appendices…….. viii

Introduction……. 1

- Predicting Sporulation Capacities via PCR Amplification……..... 4
- Toxin Regulators……... 5
- Multi-Locus Sequence Typing…….. 8

Methods……... 11

- Predicting Sporulation Capacities via PCR Amplification……... 12
- Multi-Locus Sequence Typing…….. 13
- Toxin Regulators……... 15

Results……..... 17

- Predicting Sporulation Capacities via PCR Amplification……... 20
- Multi-Locus Sequence Typing…….. 24
- Toxin Regulators……... 30

Discussion…… 34

- Predicting Sporulation Capacities via PCR Amplification……... 34
- Multi-Locus Sequence Typing…….. 37
- Toxin Regulators……... 38

Conclusions……. 41

References……... 42

Appendix……. 44

List of Figures

Figure 1. Regulated excision of *skin*^{Cd} from *sigK* is crucial for sporulation in *C. difficile*……... 5

Figure 2. Two large toxins, TcdA and TcdB, are encoded on the pathogenicity locus (PaLoc), which is comprised of five genes…….. 7

Figure 3. The binary toxin, CDT, is encoded on CdtLoc that is comprised of three genes……. 8

Figure 4. Condensed dendrogram displaying PFGE-*SmaI* restriction digestion pattern similarity…… 19

Figure 5. PCR strategy used to determine predicted sporulation capacities……. 21

Figure 6. Visualization of fragments amplified via PCR for predicting a strain's capacity for sporulation……. 23

Figure 7. Frequency of predicted sporulation capacities taking MLST sequence type into consideration……..... 23

Figure 8. Distribution of MLST sequence types across the 306 strains analyzed……. 26

Figure 9. Dendrogram displaying MLST sequence type similarity…….. 28

Figure 10. Dendrogram displaying allelic variation within *C. difficile*'s *tcdC* toxin regulator and their associated relatedness……. 32

List of Tables

Table 1. PCR amplification primers used in testing the sporulation capacities of *C. difficile* isolates…….. 13

Table 2. Predicted *C. difficile* sporulation capacity scenarios based on PCR amplification results…… 13

Table 3. Primers used for multi-locus sequence typing (MLST) of *C. difficile*. 15

Table 4. Primers used for molecular characterization of *C. difficile* toxin regulators……. 16

Table 5. Clinical strain distribution across eight Manitoba hospitals and one laboratory; seven control strains also were provided, totalling 309 *C. difficile* strains analyzed……... 17

Table 6. Summary of *C. difficile* predicted sporulation capacities and distribution of each scenario across hospital source……... 22

Table 7. Allelic variation in the seven *C. difficile* housekeeping genes analyzed via MLST……. 25

Table 8. Allelic variation in the *C. difficile* toxin regulator *tcdC* for 301 strains…….. 31

List of Abbreviations

ADP: Adenosine diphosphate

ARNI: Antimicrobial Resistance and Nosocomial Infections (department at the National Microbiology Laboratory)

bp: base pairs

CDI: *Clostridium difficile* infection

CDT: *Clostridium difficile* toxin

CdtLoc: *Clostridium difficile* toxin locus

CF: Circular fragment (of $skin^{Cd}$)

DNA: Deoxyribonucleic acid

kb: kilobase

MLST: Multi-locus sequence typing

NAP: North American Pulsed-field (type)

NCBI: National Center for Biotechnology Information

PaLoc: Pathogenicity locus

PCR: Polymerase chain reaction

PFGE: Pulsed-field gel electrophoresis

ppm: parts per million

RAPD: Random amplified polymorphic DNA

ROI: Region of interest

skin*^{Cd}: *sigK intervening sequence of *Clostridium difficile*

ST: Sequence type (of MLST)

TAE: Tris-acetate-ethylenediaminetetraacetic acid (EDTA)

TcdA: Toxin A

TcdB: Toxin B

TRs: Tandem repeats

List of Appendices

Appendix 1. Nucleotide sequences of novel *aroE* alleles defined herein for *C. difficile*……... 44

Appendix 2. Nucleotide sequences of novel *ddl* alleles defined herein for *C. difficile*……... 44

Appendix 3. Nucleotide sequence of novel *dutA* allele defined herein for *C. difficile*……... 44

Appendix 4. Nucleotide sequences of novel *sodA* alleles defined herein for *C. difficile*……... 45

Appendix 5. Novel MLST sequence types defined herein for *C. difficile*… 46

Appendix 6. Distribution of MLST sequence types used herein for 306 strains of *C. difficile*…… 47

Appendix 7. Nucleotide sequences of novel *tcdC* alleles defined herein for *C. difficile*……... 48

Appendix 8. Summary table (spanning seven pages) of PFGE-*SmaI* pattern, NAP type, predicted sporulation capacity, MLST sequence type and *tcdC* allele results for each of the 309 strains analyzed in this study… 50

Introduction

The anaerobic, Gram-positive, spore-forming bacillus *Clostridium difficile* is found naturally within the colonic microflora of mammals (Cookson 2007; Dawson *et al.*, 2009). *C. difficile* has recently emerged as an important source of healthcare-associated infections as it is the major causative agent of antibiotic-associated diarrhoea (15 to 25%) and the more severe condition, pseudomembranous colitis (95%; Haraldsen and Sonenshein 2003; Lemee *et al*., 2004; Bouvet and Popoff 2008; Rupnik *et al*., 2009; Zaiß *et al*., 2009). *C. difficile* infection (CDI) is a toxin-mediated gastrointestinal disease. Symptoms can range from asymptomatic colonization, mild diarrhoea, to more severe disease syndromes with abdominal pain, fever and leukocytosis (Cookson 2007; Dawson *et al.*, 2009; Rupnik *et al*., 2009; Zaiß *et al*., 2009). Pseudomembranous colitis is characterized by pseudomembranes, which are visible patches of red and white blood cells, fibrin and bacteria, that form on the inner lining of the inflamed colon or rectum (Cookson 2007; Rupnik *et al*., 2009). Other severe conditions such as bowel perforation, toxic megacolon, sepsis, shock, and even death may result subsequent to CDI (Cookson 2007; Dawson *et al.*, 2009; Rupnik *et al*., 2009; Zaiß *et al*., 2009).

Risks for CDI include exposure to antibiotics, a spore-contaminated environment, inadequate hand hygiene, increased subject age, inadequate sanitation and cleaning of healthcare premises and overcrowded hospitals (Cookson 2007; Rupnik *et al*., 2009). Up to 50% of elderly patients in long-term care or nursing home facilities have been shown to carry and shed toxigenic *C. difficile* strains (Rupnik *et al*., 2009). In North America, hospital-acquired (nosocomial) CDI infections have a higher incidence than all other bacterial gastrointestinal pathogens combined (Alfa *et al*., 2008) and in the healthcare

environment, a universal association between antibiotic therapy and the occurrence of *C. difficile* has been observed. Antibiotic therapy has been shown to disrupt otherwise protective normal gut microflora (Dawson *et al*., 2009) enabling pathogenic and even multi-drug resistant microorganisms to colonize. Hence, persons with compromised colonic flora are susceptible to *C. difficile* colonization and often subsequently progress to infection (Haraldsen and Sonenshein 2003; Lemee *et al*., 2004; Cookson 2007; Dawson *et al.*, 2009; Rupnik *et al*., 2009). As a result, prompt restoration of normal gut microflora is crucial in preventing CDI after one undergoes antibiotic therapy (Rupnik *et al*., 2009).

Although *C. difficile* is a leading cause of healthcare-associated infections, CDIs are now also arising in the community and in animals used as food sources (Rupnik *et al*., 2009). Consequently, CDI in younger patients with no prior hospital or antibiotic contact, are occurring (Rupnik *et al*., 2009).

Early and aggressive intervention is paramount to CDI recovery (Cookson 2007). The current antibiotics used to treat CDI are metronidazole and vancomycin (Rupnik *et al*., 2009). However, diarrhoea often resumes within days (to weeks) after treatment has ceased, a condition referred to as recurrent CDI (Dawson *et al.*, 2009; Rupnik *et al*., 2009). Twenty to fifty percent of such recurrences are caused by reinfection with a "new" strain of *C. difficile* rather than a relapse of the original infection (Dawson *et al.*, 2009; Rupnik *et al*., 2009). Thus, patients often receive smaller doses of vancomycin every second or third day in the hope of minimizing *C. difficile* regrowth while the normal microflora recovers (Rupnik *et al*., 2009). Compromised microflora may also be replenished through fecal transplant delivery from healthy volunteers via nasogastric tube

or enema (Rupnik *et al*., 2009). If these methods are ineffective, surgical removal of the severely damaged colon can be the only remaining life-saving measure (Rupnik *et al*., 2009).

Two main forms of CDI control measures are enforced in healthcare systems: i) those that attempt to prevent *C. difficile* spores from reaching patients, and ii) those that reduce the risk of CDI should the organism be ingested (Rupnik *et al*., 2009). Spore ingestion occurs after a contaminated surface is touched (such as bed linens, pull strings, bedpans, toilets) followed by hand-to-mouth transfer. Whenever an individual is diagnosed with CDI, isolation precautions are immediately undertaken to protect others. Healthcare workers are notified to wear effective barrier protection such as gowns and gloves (Cookson 2007; Alfa *et al*., 2008; Rupnik *et al*., 2009). Housekeeping cleaning protocols are also enhanced, attempting to reduce the environmental load of *C. difficile* and transmission between patient rooms and thus, nosocomial infection risk (Alfa *et al*., 2008). Unfortunately, alcohol rubs and gels are not effective in removing *C. difficile* spores (Rupnik *et al*., 2009). Moreover, when Alfa *et al*. (2008) tested the cleaning efficiency of hospitals, neither physical cleaning action nor cleaning agent used were found effective in killing or removing *C. difficile* spores from toilet surfaces. Rupnik *et al*. (2009) have also demonstrated that traditional antimicrobial detergents and quaternary ammonium cleaning agents are not sporicidal, and may actually enhance *C. difficile* sporulation (Rupnik *et al*., 2009). Spore contamination and CDI rates can be significantly reduced via use of a sporicidal hypochlorite solution (1,000 to 5,000 ppm chlorine bleach; Alfa *et al*., 2008; Rupnik *et al*., 2009). However, due to the corrosivity, pungent odour, and cutaneous and respiratory irritation to those using such a high

strength of hypochlorite, such drastic cleaning measures are usually reserved for serious CDI outbreaks or epidemic CDI (Rupnik *et al*., 2009).

Predicting Sporulation Capacities via PCR Amplification

C. difficile is found either as vegetative cells or dormant spores (Paredes *et al*., 2005). Vegetative cells are metabolically active and are killed when exposed to oxygen, whereas spores are neither metabolically active nor are they killed by oxygen. *C. difficile* spores are formed as a means for the bacterium to survive hostile or unfavourable conditions (Paredes *et al*., 2005). Spores are highly resistant to dessication, chemicals and extreme temperatures, and have been detected intact on floor surfaces after five months (Alfa *et al*., 2008; Rupnik *et al*., 2009). Temperatures and disinfectant treatments used in typical hospital laundering cycles are not adequate to destroy *C. difficile* spores. Hence, cross-contamination of bed linens can occur, thereby facilitating *C. difficile* spore spreading (Dawson *et al*., 2009; Rupnik *et al*., 2009). During and following CDI treatment, patients may shed up to 10^4 spores per gram of feces (Alfa *et al*., 2008). Consequently, spores frequently contaminate and persist in the environment for long periods around CDI patients (Paredes *et al*., 2005; Rupnik *et al*., 2009).

In order for mature *C. difficile* spores to form, an intact *sigK* gene is essential. The translated product (SigK) functions as an RNA polymerase sigma factor, σ^K, that regulates the transcription of late expression spore-coat proteins that are critical for spore survival (Haraldsen and Sonenshein 2003). Surprisingly, Haraldsen and Sonenshein (2003) found that a 14.6 kb prophage-like element was inserted in *sigK*, thus interrupting the locus between the amino- and carboxy-terminal encoding portions. Importantly,

regulated excision of this prophage-like element, termed *skin*^{Cd} for *sigK* intervening sequence of *C. difficile*, into a circular (extrachromosomal) fragment appears to be critical for achieving proper temporal activation of σ^K (Figure 1; Haraldsen and Sonenshein 2003). As a result, both the presence of *skin*^{Cd} followed by its excision is required for proper late spore coat production (Haraldsen and Sonenshein 2003). Absence of *skin*^{Cd} correlates with a sporulation defect (Haraldsen and Sonenshein 2003).

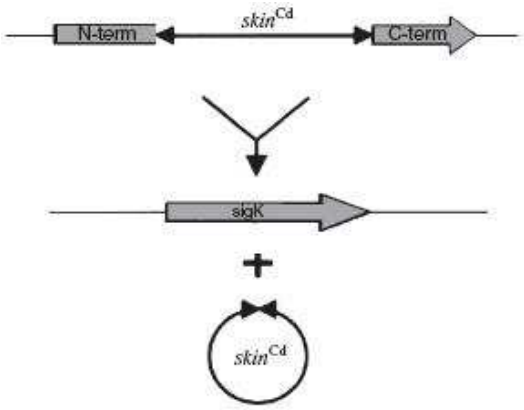

Figure 1. Regulated excision of *skin*^{Cd} from *sigK* is crucial for sporulation in *C. difficile*. At the top, shaded rectangle and arrow represent *sigK* disrupted by *skin*^{Cd} (solid line bounded by arrowheads). With appropriate regulatory signals, *skin*^{Cd} excises into a circular fragment to produce intact *sigK*. Adapted from Haraldsen and Sonenshein 2003.

Following disruption of the protective gut microflora, ingested *C. difficile* spores germinate to become actively growing vegetative cells, which then colonize the gastrointestinal tract and produce toxins (Dawson *et al*., 2009). Spore germination is likely triggered by the stomach's low pH or other host environmental cues (Rupnik *et al*., 2009).

Toxin Regulators

Biochemical and molecular studies have shown that the main virulence factors of *C. difficile* are the large clostridial toxins, designated Toxin A (TcdA) and Toxin B

(TcdB), which together are responsible for the principal clinical symptoms of CDI (namely secretory diarrhoea and inflammation of the colonic mucosa; Paredes *et al*., 2005; Bouvet and Popoff 2008; Dawson *et al.*, 2009; Kuijper *et al*., 2009; Rupnik *et al*., 2009). TcdA and TcdB are single-chain proteins with three main functional domains: an amino-terminal binding domain with characteristic repeats, a carboxy-terminal catalytic domain and a putative translocation domain (Rupnik *et al*., 2009). TcdA and TcdB are enterotoxins (toxins specific for cells of the intestinal mucosa) which have the ability to glycosylate small GTPases of the Rho and Ras families in the host cell (Rupnik *et al*., 2009). This causes disruption of the host actin cytoskeleton and tight junctions, resulting in decreased transepithelial resistance, fluid accumulation and destruction of the intestinal epithelium (Dawson *et al.*, 2009). Inflammatory mediators are also released from damaged intestinal epithelial cells, as well as recruited mast cells and macrophages, promoting an influx of inflammatory cells which adds to fluid secretion and intestinal inflammation (Rupnik *et al*., 2009).

Through experiments on human colonic tissue, Riegler *et al*. (1995) demonstrated that TcdB was more potent than TcdA in causing mucosal necrosis and decreasing barrier function. Reasoning came from the fact that TcdA binds more effectively on the apical side of the host cell and TcdB binds to an unknown receptor on the basolateral side of the host cell (Rupnik *et al*., 2009). Both toxins are taken up via endocytosis. Due to a lowering of the endosome's pH, TcdA becomes activated, cleaved from the holotoxin and is translocated into the cytosol (Dawson *et al.*, 2009; Rupnik *et al*., 2009). However, it is eukaryotic signals that induce autocatalytic cleavage of TcdB into two domains: the carboxy-terminal delivery domain that aids in the translocation of the amino-terminal

domain to the cytosol, whereupon its glucosyltransferase acts against GTPases (Dawson *et al.*, 2009).

TcdA and TcdB are produced during late exponential and stationary phases of vegetative growth due to strain and environmental factors such as limiting nutrient levels (glucose, amino acids, biotin), temperature and the presence of sub-inhibitory levels of antibiotics (Paredes *et al*., 2005; Rupnik *et al*., 2009). Together, TcdA and TcdB are encoded with *tcdR* (formerly known as *tcdD*), *tcdC* and *tcdE* in a well-defined genetic element, called the pathogenicity locus (PaLoc; Figure 2; Rupnik *et al*., 2009). *tcdR* encodes a sigma factor involved in positive transcriptional regulation of *tcdA* and *tcdB*, whereas *tcdC* encodes a negative regulator that destabilizes the TcdR-holoenzyme (Bouvet and Popoff 2008; Dawson *et al.*, 2009). The DNA sequence of the PaLoc is variable, and strains with changes in this region are defined as different toxinotypes (Rupnik *et al*., 2009).

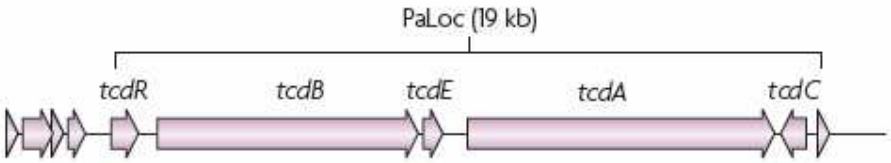

Figure 2. Two large toxins, TcdA and TcdB, are encoded on the pathogenicity locus (PaLoc), which is comprised of five genes. *tcdR* and *tcdC* are the positive and negative transcriptional regulators, respectively, of *tcdA* and *tcdB*. Adapted from Rupnik *et al*., 2009.

A third toxin, CDT, is produced by some *C. difficile* strains – usually those with sequence variant but functional *tcdA* and *tcdB* genes (Dawson *et al.*, 2009; Rupnik *et al*., 2009). CDT is composed of two proteins, CdtA and CdtB, which act together upon host cells (Bouvet and Popoff 2008; Rupnik *et al*., 2009). CdtB binds to host cells and translocates the catalytic component, CdtA, into the cytosol where it ADP-ribosylates

actin molecules and causes deleterious cytoskeletal rearrangements (Rupnik *et al*., 2009). The exact role of CDT in CDI is unknown; however, this toxin is speculated to cause secretion of fluid from colonic cells even though it does not directly cause cell death, thereby enhancing pathogenesis and disease otherwise caused by TcdA and TcdB (Cookson 2007; Dawson *et al.*, 2009; Kuijper *et al*., 2009). Genes encoding CdtA and CdtB are located outside the PaLoc in a so-called binary toxin locus, designated CdtLoc (Figure 3; Kuijper *et al*., 2009; Rupnik *et al*., 2009). Production of CDT is tightly and positively regulated by the *cdtR* gene, also located within CdtLoc (Bouvet and Popoff 2008; Dawson *et al.*, 2009). Strains that do not express CDT have a conserved 68 bp sequence replacing the CdtLoc (Dawson *et al.*, 2009). The production of CDT along with TcdA and TcdB cannot fully explain *C. difficile* pathogenesis. Thus, other, yet unidentified, factors must be important in the transmissibility, survival, colonization, and virulence of *C. difficile* (Dawson *et al.*, 2009).

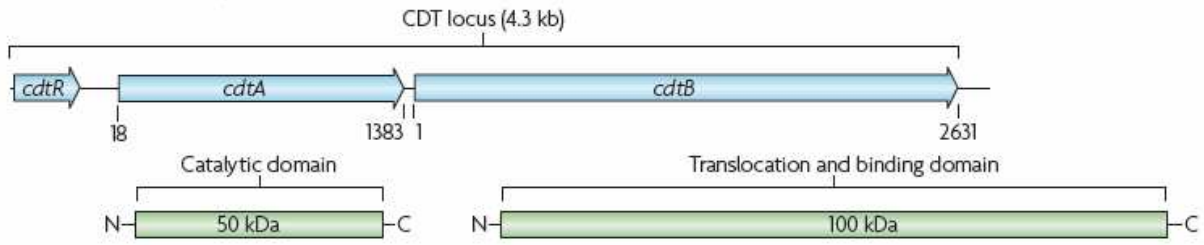

Figure 3. The binary toxin, CDT, is encoded on CdtLoc that is comprised of three genes. The binary toxin is composed of two unlinked proteins, CdtB and CdtA. CdtB has a binding function and CdtA is the enzymatic component. Adapted from Rupnik *et al*., 2009.

Multi-Locus Sequence Typing (MLST)

Improved surveillance and strain typing can help detect and control *C. difficile* (Cookson 2007). Typing methods have been classified into two categories: phenotypic or genotypic (Kuijper *et al*., 2009). Phenotypic methods focus on the products of gene

expression whereas genotypic methods analyze the genetic profile of strains and therefore, have advantages over phenotypic methods in terms of providing greater levels of typeability and molecular markers (Kuijper *et al*., 2009). A typing method with high discriminatory power, typeability, and reproducibility is required for studying the epidemiology of CDI (Kuijper *et al*., 2009). Deoxyribonucleic acid (DNA) sequence typing procedures overcome these limitations since sequence data is easily exchangeable and stored in databases accessible via the Internet (Kuijper *et al*., 2009). Molecular typing methods that have been used to type *C. difficile* strains include, but are not limited to, pulsed-field gel electrophoresis (PFGE; giving rise to pulsotypes), polymerase chain reaction (PCR) ribotyping, random amplified polymorphic DNA (RAPD) analysis, toxinotyping (based on sequence data of TcdA and TcdB), and multi-locus sequence typing (MLST; Lemee *et al*., 2004; Dawson *et al.*, 2009). These typing methods have been used to confirm single or multiple hospital-acquired outbreaks of CDI (Dawson *et al.*, 2009).

Traditional MLST consists of PCR amplification of housekeeping genes followed by DNA sequence variation analysis for studying genetic relationships and population structures (Kuijper *et al*., 2009). MLST characterizes multi-locus genotypes of bacterial isolates by using 400- to 500-bp intragenic sequences of several (generally seven) housekeeping genes (Lemee *et al*., 2004). However, for *C. difficile*, MLST including virulence-associated genes has been shown to have higher discriminatory power than housekeeping genes alone (Lemée *et al*., 2005). The major advantages of MLST over other molecular typing methods are: i) ability to produce high throughput comparable and unambiguous sequence data that can be stored in centralized databases and, ii) allows for

the molecular study of strain evolution as MLST describes neutral (ie. non-coding) sequence variations affecting several housekeeping genes (Lemee *et al*., 2004; Bouvet and Popoff 2008; Dawson *et al.*, 2009; Kuijper *et al*., 2009).

Transmission of *C. difficile* is mediated via spore shedding. Inadequate cleaning procedures in the healthcare environment have allowed *C. difficile* spores to contaminate and persist in the environment. As a result, *C. difficile* has emerged as an important cause of healthcare-associated infections that can become chronic, debilitating and fatal. Therefore, a fast and efficient way to detect the element responsible for influencing sporulation, $skin^{Cd}$, is important for risk assessment (i.e. sporulation capacity) of clinical strains and needs to be validated. A PCR test has been designed that confirms the presence or absence of $skin^{Cd}$ and predicts a strain's ability to sporulate; these will be assessed for congruency with *in vitro* spore count data when available. MLST on seven housekeeping genes – *aroE* (shikimate dehydrogenase), *ddl* (D-alanine:D-alanine ligase), *dutA* (dUTP pyrophosphatase), *gmk* (guanylate kinase), *recA* (recombinase), *sodA* (superoxide dismutase), and *tpi* (triosephosphate isomerase) – and three toxin regulatory genes (*tcdC*, *tcdR*, and *cdtR*) was performed to molecularly characterize the strains via assignment of allelic profiles and sequence types. Strain relationships were examined via sequence variability and associations between sequence type, hospital source and sporulation capacity predictions were evaluated.

Methods

The extracted DNA from 447 *C. difficile* strains was provided by Dr. Michelle Alfa at St. Boniface Research Institute in Winnipeg, Manitoba, Canada. These province-wide samples were obtained from fecal specimens taken from patients diagnosed with *C. difficile*-associated diarrhoea between 2006 and 2007. All strains provided tested positive for TcdA and TcdB toxins (hence, $TcdA^+/TcdB^+$) by PCR and phenotypic tests. These strains have previously been supplied to the Antimicrobial Resistance and Nosocomial Infections (ARNI) department at the National Microbiology Laboratory (Winnipeg, Manitoba, Canada) where PFGE-*SmaI* patterns and North American pulsed-field (NAP) types were assigned, when appropriate. Dr. Michelle Alfa also provided seven *C. difficile* strains as "controls" for this study. However, owing to the double-blind nature of this study, strain information was not provided.

A subset of 302 clinical Manitoba *C. difficile* strains and the seven controls were selected for further analyses. These strains were selected from a dendrogram constructed based on PFGE-*SmaI* patterns assigned by ARNI. All strains at dendrogram terminal branches were represented in the subset and several samples were analyzed for larger strain clusters. The 302 Manitoba strains analyzed included 122 distinct PFGE-*SmaI* patterns, and included NAP types 1 through 12, except 5 and 8. Primers were designed for each gene or region of interest (ROI) based upon the complete genome sequence for *C. difficile* 630 retrieved from the National Center for Biotechnology Information (NCBI; accession number NC_009089; Hächler *et al*., 1987) or from previously published *C. difficile* studies. Initial sample concentrations were determined via a Quant-iT™

PicoGreen® dsDNA Assay Kit (Invitrogen) to ensure PCR reactions were not going to fail due to the lack of input DNA template.

Predicting Sporulation Capacities via PCR Amplification

Primers used to amplify various regions of the *sigK* gene and flanking sequences from *C. difficile* are listed in Table 1. Amplification was performed using the manufacturer's protocol as outlined by the *Taq* DNA Polymerase, recombinant kit (Invitrogen). Thermocycling consisted of an initial denaturation at 95ºC for 1 minute, followed by 40 cycles comprised of denaturation (95ºC, 1 min), annealing (50ºC, 1 min), and extension (72ºC, 1.5 min), followed by a final extension at 72ºC for 10 minutes. PCR products were visualized on a 1% ultrapure agarose gel (in a 1X tris-acetate-EDTA buffering system; TAE) containing 0.2 µg/mL ethidium bromide and resultant images captured using a gel documentation device (Quantity One® 1-D Analysis Software v.4.6.3; Bio-Rad). Strains were scored for the presence or absence of *sigK* (primers OJH39 and OJH42; 828 bp); circular fragment (CF; primers 43new and OJH44; 1079 bp); 3′ (primers OJH42 and OJH44; 478 bp) and 5′ (primers OJH39 and 43new; 1428 bp) *sigK::skin*^{Cd} fusions. Each isolate was then classified according to scenario (A, B, C or D), and sporulation capacity predictions were made based on the PCR product results (Table 2). To confirm that appropriate regions were amplified, amplicons from a subset of 14 strains were chosen for purification with Agencourt AMPure (Beckman Coulter Genomics) followed by DNA sequencing using the same primers as were used for the PCR reactions. Throughout the study, PCR amplicon templates were sequenced by Sanger's primer extension methodology from appropriate primers using BigDye™

Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems®). Capillary-sequencing was performed on 3730 XL genetic analyzers (Applied Biosystems®) equipped with Sequence Analysis v5.2 software (Applied Biosystems®). Subsequent data analysis was performed using Lasergene v8.1 SeqMan software (DNASTAR®).

Table 1. PCR amplification primers used in testing the sporulation capacities of *C. difficile* isolates.

Primer Name	5′ to 3′ Sequence
OJH39	CCGAATTCAAGCTTCCCACTCTTTATGTCATATTCC
OJH42	CCCGAGCTCCTAATGATGTACGATATATCC
43new	ATCAATGTATTGGCATTGATTGTG
OJH44	CCATGATTCAGATTCCCTTGG

Primers OJH39, OJH42 and OJH44 are the same as those used by Haraldsen and Sonenshein 2003.

Table 2. Predicted[a] *C. difficile* sporulation capacity scenarios based on PCR amplification results.

sigK	CF	3′ Fusion	5′ Fusion
+	-	-	-
Scenario A: *skin*^{Cd} not present in *sigK*; **won't** sporulate			
-	-	+	+
Scenario B: *skin*^{Cd} present but not excised; thus **poor** sporulator			
+	+	+	+
Scenario C: *skin*^{Cd} present and partially excised from *sigK*; thus **good** sporulator			
+	+	-	-
Scenario D: *skin*^{Cd} once present but now completely excised from *sigK*; thus **strong** sporulator			

[a] Sporulation capacities are predicted only, as it is a double-blind study and actual spore counts are currently being performed at the St. Boniface Research Institute in Winnipeg, Manitoba, Canada.

Multi-Locus Sequence Typing (MLST)

Internal fragments of seven housekeeping genes were selected for molecular characterization of *C. difficile*: *aroE* (shikimate dehydrogenase; 408 bp), *ddl* (D-alanine:D-alanine ligase; 420 bp), *dutA* (dUTP pyrophosphatase; 309 bp), *gmk* (guanylate kinase; 279 bp), *recA* (recombinase; 378 bp), *sodA* (superoxide dismutase; 408 bp), and *tpi* (triosephosphate isomerase; 393 bp). Primers used to amplify each loci in *C. difficile*

are listed in Table 3. Amplification was performed using the protocol outlined by the Platinum® *Taq* DNA Polymerase High Fidelity kit from Invitrogen. Thermocycling consisted of an initial denaturation at 94ºC for 2 minutes; followed by 40 cycles comprised of denaturation (94ºC, 30 sec), annealing (54ºC, 1 min) and extension (68ºC, 1 min), with a final round of denaturation (94ºC, 1.5 min), annealing (54ºC, 1.5 min), and extension (68ºC, 5 min). A subset of PCR products were visualized on a 1% ultrapure agarose gel (in a 1X TAE buffering system) containing 0.2 µg/mL ethidium bromide and resultant images captured using a gel documentation device (Quantity One® 1-D Analysis Software v.4.6.3; Bio-Rad) to ensure appropriate fragment sizes were amplified. All PCR products were purified with Agencourt AMPure (Beckman Coulter Genomics) followed by DNA sequence analysis as described above. Primers for DNA sequencing were the same as those used for PCR reactions. Sequences were subsequently analyzed in BioNumerics 6.0.1 (Applied Maths) wherein allelic profiles (based on the score for each of the seven alleles analyzed) were created for each isolate and sequence types were assigned. Allelic profiles and sequence types assigned were based on the *C. difficile* MLST scheme set forth by the Pasteur website (http://www.pasteur.fr/recherche/genopole/PF8/mlst/Cdifficile.html) – now designated "old" scheme and maintained only prior to October 2009.

Table 3. Primers used for multi-locus sequence typing (MLST) of *C. difficile*.

Primer Name	5′ to 3′ Sequence
aroE-F	CTAGTAGGTGAAAAACTCTCTCA
aroE-R	ACTGGTGTAGCATTTAATATTATATC
ddl-F	CATAAACTTGTTCATTCAGAAGG
ddl-R	CTATGAGAAGTAAAGCCAGGAAT
dutA-F	CCTAATTTTGCTCACAAAGGT
dutA-R	AAATCCAGTTGAGCCAAACC
gmk-F	TCAGGTGCAGGAAAAGGTAC
gmk-R	TCTGTTTCTGTACCTCTTCCAAC
recA-F	CCAGATACAGGTGAACAGGC
recA-R	TTTAACATTTTCTCTTCCTTGTCC
sodA-F	TATSCWTATGATGCWYTWGARCC
sodA-R	TARTAAGCATGYTCCCAAACATC
tpi-F	GCAGGAAACTGGAAAATGCATAA
tpi-R	CAGATTGGCTCATATGCAACAAC

All primers listed and their sequences are from the Pasteur website for the old *C. difficile* MLST scheme (http://www.pasteur.fr/recherche/genopole/PF8/mlst/Cdifficile.html).

Toxin Regulators

Based on prior publications, three toxin regulators were selected for molecular characterization of *C. difficile*: *tcdC* (primers C1 and C2; 600 bp), *tcdR* (primers PaL11 and PaL12; 385 bp) and *cdtR* (P1605F and P1510R; 720 bp). Primers used to amplify each toxin regulator in *C. difficile* are listed in Table 4. Amplification was performed as per the Platinum® *Taq* DNA Polymerase High Fidelity kit protocol (Invitrogen). Thermocycling consisted of an initial denaturation at 94ºC for 5 minutes, followed by 40 cycles comprised of denaturation (94ºC, 1 min), annealing (50ºC, 1 min) and extension (68ºC, 1 min), with a final extension of 68ºC for 10 minutes. A subset of PCR products were visualized on a 1% ultrapure agarose gel (in a 1X TAE buffering system) containing 0.2 µg/mL ethidium bromide and resultant images captured using a gel documentation device (Quantity One® 1-D Analysis Software v.4.6.3; Bio-Rad) to ensure appropriate fragment sizes were amplified. All PCR amplicons were purified with Agencourt AMPure (Beckman Coulter Genomics) followed by DNA sequencing and data analysis

as described above. Primers for DNA sequencing were the same as those used for PCR reactions.

Table 4. Primers used for molecular characterization of *C. difficile* toxin regulators.

Primer Name	5′ to 3′ Sequence
C1[a]	TTAATTAATTTTCTCTACAGCTATCC
C2[a]	TCTAATAAAAGGGAGATTGTATTATG
PaL11[b]	CTCAGTAGATGATTTGCAAGAA
PaL12[b]	TTTTAAATGCTCTATTTTTAGCC
P1605F[c]	AGCATAAATATACCTTAATTCTAACTATC
P1510R[c]	TCTTGAGACATCTCCTTTTTCT

[a] C1 and C2 primers are the same as those used by Spigaglia and Mastrantonio 2002.
[b] PaL11 and PaL12 primers are the same as those used by Spigaglia and Mastrantonio 2004.
[c] P1605F and P1510R primers are the same as those used by Bouvet and Popoff 2008.

Results

Of 302 province-wide *C. difficile* clinical samples analyzed, St. Boniface General Hospital had greatest representation with 73 samples (Table 5). Westman Regional Hospital (in Brandon) was represented by 49 samples, Grace Hospital by 38 and Cadham Provincial Lab by 32. Misericordia Hospital and Seven Oaks Hospital had the fewest samples, with 11 and 15, respectively. The remaining three hospitals, Concordia, Victoria and Health Sciences Centre were represented by 28, 30 and 26 samples, respectively. In addition to the 302 clinical isolates analyzed, seven *C. difficile* "control" strains also were analyzed.

Table 5. Clinical strain distribution across eight Manitoba hospitals and one laboratory; seven control strains also were provided, totalling 309 *C. difficile* strains analyzed.

Hospital Source	All 454 Samples Provided	309 Sample Subset	Percent Represented from Hospital Source	Subset Percent Represented from Total Examined
St. Boniface General Hospital	73	73	100.0	23.6
Grace Hospital (Deer Lodge)	38	38	100.0	12.3
Concordia Hospital	28	28	100.0	9.1
Victoria Hospital	30	30	100.0	9.7
Misericordia Hospital	11	11	100.0	3.6
Health Sciences Centre	54	26	48.2	8.4
Seven Oaks Hospital	32	15	46.9	4.9
Westman Regional Hospital	100	49	49.0	15.9
Cadham Provincial Lab	81	32	39.5	10.4
Controls	7	7	100.0	2.3
Total/Average	454	309	68.1	10.0 ± 6.3

Prior PFGE analyses were performed on the majority of the 302 selected province-wide *C. difficile* clinical strains by the ARNI department at the National Microbiology Laboratory. Only three clinical strains (SB321, C361 and V019) and the seven controls were not previously analyzed (the latter will be analyzed in the near future). A dendrogram (Figure 4) was constructed to visualize PFGE-*SmaI* restriction

digestion patterns across strains. Three major clusters (A, B and C) formed, sharing 53% overall pattern similarity; all except four strains were found within these three clusters. Clusters A and B shared 54% overall pattern similarity, and all strains contained within also were negative by PCR for the *cdtB* locus (data provided by ARNI). Cluster A strains (all NAP untypable strains) revealed 22 PFGE-*SmaI* patterns sharing 58% overall pattern similarity. A single strain (CPL56) grouped alone within cluster A sharing only 58% similarity with the other cluster A strains. Two additional internal subclusters within cluster A were detected exhibiting 60% similarity; one (A.1) exhibited 66% similarity (11 strains) and the other (A.2) exhibited 72% similarity (78 strains). Cluster B strains also shared 60% overall pattern similarity. Two subclusters also were observed within cluster B: cluster B.1 exhibited 64% similarity (NAP3, NAP6 and NAP10 strains; totalling 58 strains) whereas subcluster B.2 revealed 66% similarity (NAP2 and NAP12 strains; totalling 146 strains). Containing six internal subclusters, cluster C shared 56% overall similarity. Subcluster C.1 was comprised solely of NAP4 strains with 78% similarity (32 strains); subcluster C.2 was comprised solely of NAP untypable strains with 70% overall similarity (12 strains); subcluster C.3 exhibited 70% similarity (comprised of NAP1, NAP11 and NAP untypable strains; totalling 97 strains), and subcluster C.4 with 64% similarity (comprised of NAP7 and NAP untypable strains; totalling 13 strains). The remaining four strains (CPL36, W76, H13 and SB201) clustered outside of the three major groupings (A, B and C). SB201 was the most divergent, exhibiting 68% similarity to the remaining three strains, which shared 79% similarity.

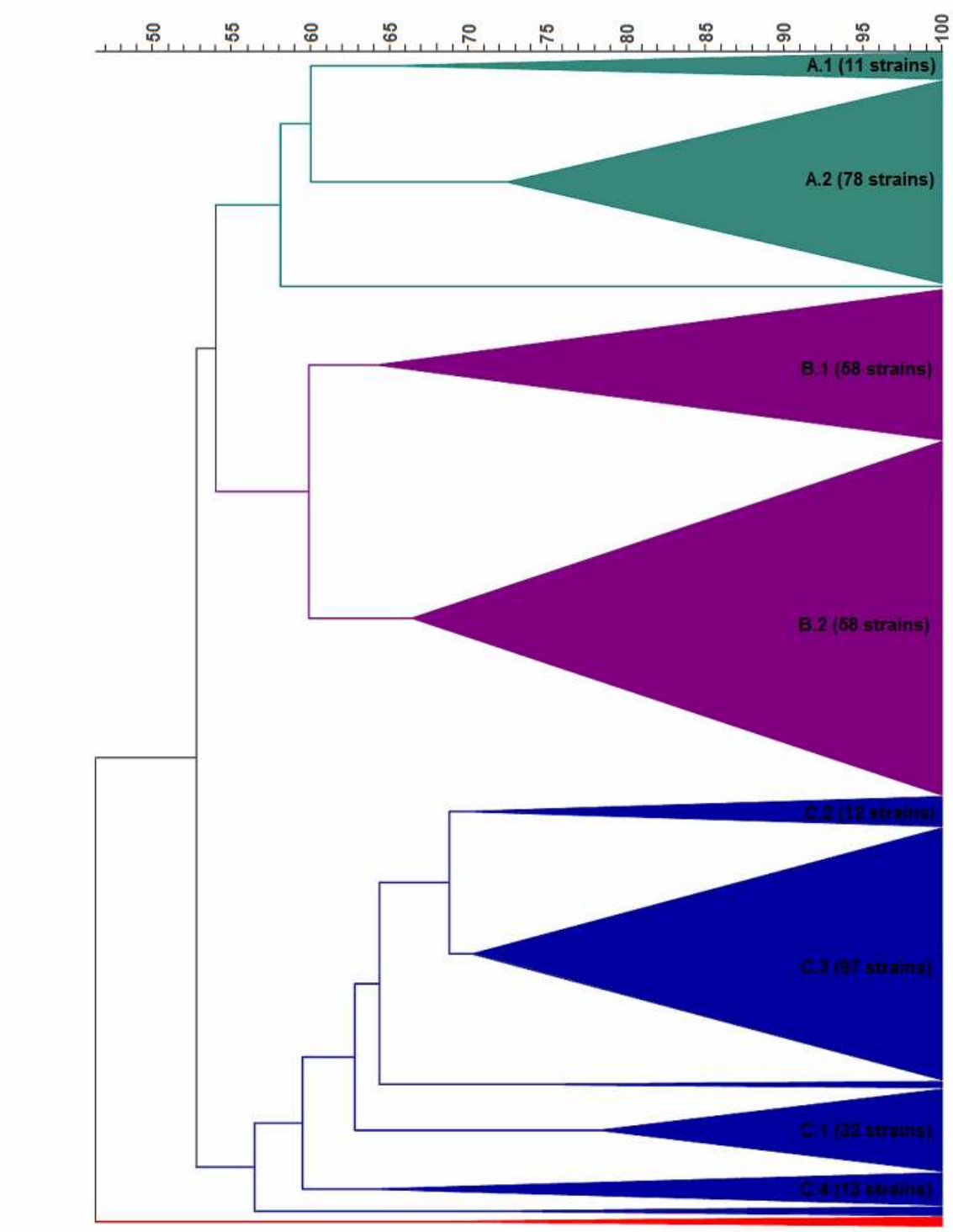

Figure 4. Condensed dendrogram displaying PFGE-*SmaI* restriction digestion pattern similarity. The red cluster at the bottom represents the four divergent strains (CPL36, W76, H13 and SB201) that exhibited 46% similarity to the three major groupings (A, B and C). Dendrogram was generated in BioNumerics and modified in Microsoft Paint for visual clarity.

Predicting Sporulation Capacities via PCR Amplification

Using DNA extracted from *C. difficile* spore preparations (but originally derived from fecal specimens and "control" *C. difficile* strains) as input template in polymerase chain reaction assays, predictions were made on the molecular capacity of an isolate to sporulate. PCR with primer sets [OJH39 + 43new] and [OJH42 + OJH44] yielded two unique amplification products containing part of both *sigK* and $skin^{Cd}$ sequences, designated 5′Fus (1428 bp) and 3′Fus (478 bp) respectively, if present (i.e. the primer sets flank the *sigK::*$skin^{Cd}$ fusions; see Figure 5). A single amplicon product, *sigK* (828 bp), was obtained if the primer pair [OJH39 + OJH42] annealed and an undisrupted form of *sigK* was present on the chromosome. If $skin^{Cd}$ was efficiently excised from *sigK* and subsequently circularized, the primer pair [43new + OJH44] (now facing each other) yielded an amplicon, designated CF (1079 bp). To ensure PCR reactions were amplifying the expected products, *sigK*, CF, 3′Fus and 5′Fus amplicons of 14 strains (selected at random) were sequenced. The resultant sequencing data confirmed that the PCR reactions were specific to the predicted loci (data not shown). They also confirmed that intact *sigK* is created following the excision of $skin^{Cd}$, and that an in-frame twelve base pair repeat is generated at the site of $skin^{Cd}$ insertion and maintained following its excision (data not shown), which occurs commonly with prophage insertions.

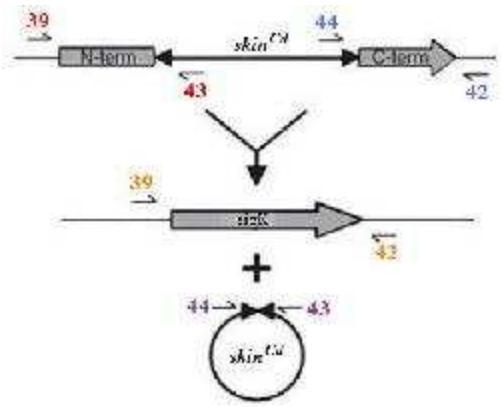

Figure 5. PCR strategy used to determine predicted sporulation capacities. Primers [OJH39 + 43new] and [OJH42 + OJH44] amplify 5′Fus and 3′Fus products, respectively; [OJH39 + OJH42] amplify intact *sigK*; [43new + OJH44] amplify circularized $skin^{Cd}$ as only in this configuration are primers facing each other.

Based on molecular analysis, the putative sporulation capacity was predicted for 308 strains; data summarized in Table 6. One strain (W36) was eliminated from this portion of the study, as it did not yield conclusive results. Four predicted sporulation capacity scenarios were detected. Scenario A was designated when a *sigK* fragment was amplified but not fragments CF, 3′Fus, or 5′Fus (Figure 6A). Hence, $skin^{Cd}$ was not detectable within *sigK* and undetectable as a circular fragment. Scenario A strains (115 clinical strains and one control) were predicted to be "incapable" of sporulation. Scenario B was designated when $skin^{Cd}$ was still found interrupting *sigK* in the spore DNA preparations and thus, was not efficiently excised. Scenario B strains (one clinical strain and one control), had both 3′Fus and 5′Fus fragments amplified but not intact *sigK* nor CF (Figure 6B). These scenario B strains were designated as "poor" sporulators, as they can only initiate sporulation whenever $skin^{Cd}$ is finally excised, which appears to occur with lower efficiency. Scenario C was designated when all four fragments (*sigK*, CF, 3′Fus and 5′Fus) were amplifiable (Figure 6C), indicating that $skin^{Cd}$ was both

present within *sigK* but also partially excised in circular form. Scenario C strains (180 clinical strains and five controls) were designated as "good" sporulators. Finally, scenario D was designated when *sigK* and CF amplicons were detectable, but not 3′Fus and 5′Fus fragments (Figure 6D). Scenario D strains (5 clinical strains) were designated as "strong" sporulators, since these PCR results implied that $skin^{Cd}$ (once present and interrupting *sigK*) now was completely excised from *sigK* and present in its circular form.

Table 6. Summary of *C. difficile* predicted sporulation capacities and distribution of each scenario across hospital source.

Scenario A: Won't Sporulate (N = 116)	
Hospital Source	Number of Isolates
St. Boniface General Hospital	31
Grace Hospital (Deer Lodge)	22
Concordia Hospital	19
Victoria Hospital	6
Misericordia Hospital	2
Health Sciences Centre	8
Seven Oaks Hospital	3
Westman Regional Hospital	14
Cadham Provincial Lab	10
Controls	1
Scenario B: Poor Sporulator (N = 2)	
Hospital Source	Number of Isolates
Concordia Hospital	1
Controls	1
Scenario C: Good Sporulator (N = 185)	
Hospital Source	Number of Isolates
St. Boniface General Hospital	40
Grace Hospital (Deer Lodge)	15
Concordia Hospital	8
Victoria Hospital	24
Misericordia Hospital	9
Health Sciences Centre	18
Seven Oaks Hospital	12
Westman Regional Hospital	33
Cadham Provincial Lab	21
Controls	5
Scenario D: Strong Sporulator (N = 5)	
Hospital Source	Number of Isolates
St. Boniface General Hospital	2
Grace Hospital (Deer Lodge)	1
Westman Regional Hospital	1
Cadham Provincial Lab	1

Sample W36 was eliminated from predicted sporulation capacities study as no conclusive results could be obtained.

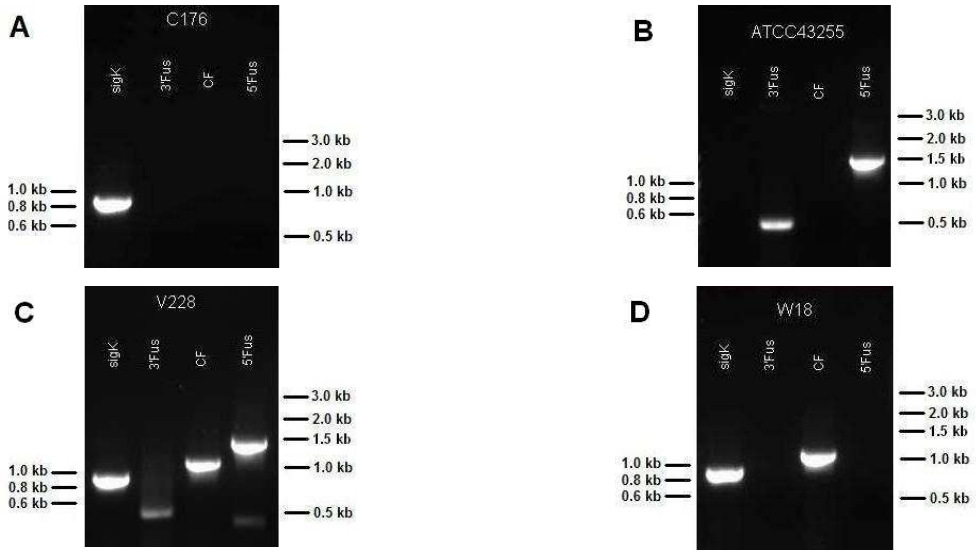

Figure 6. Visualization of fragments amplified via PCR for predicting a strain's capacity for sporulation. (A) Scenario A: won't sporulate as *sigK* is intact and $skin^{Cd}$ is undetectable as a circular fragment (CF); (B) Scenario B: poor sporulator as $skin^{Cd}$ is present within *sigK* but not excised; (C) Scenario C: good sporulator as $skin^{Cd}$ is found interrupted *sigK* and as a CF; (D) Scenario D: strong sporulator as *sigK* is intact and only the CF form of $skin^{Cd}$ is detectable.

When the frequency of predicted sporulation capacity was graphed taking MLST sequence type into consideration, the majority (84) of strains deemed unable to sporulate were revealed to be ST3 (Figure 7). Conversely, a large proportion of ST36 (53) and ST44 (22) strains were classified as good sporulators (Figure 7).

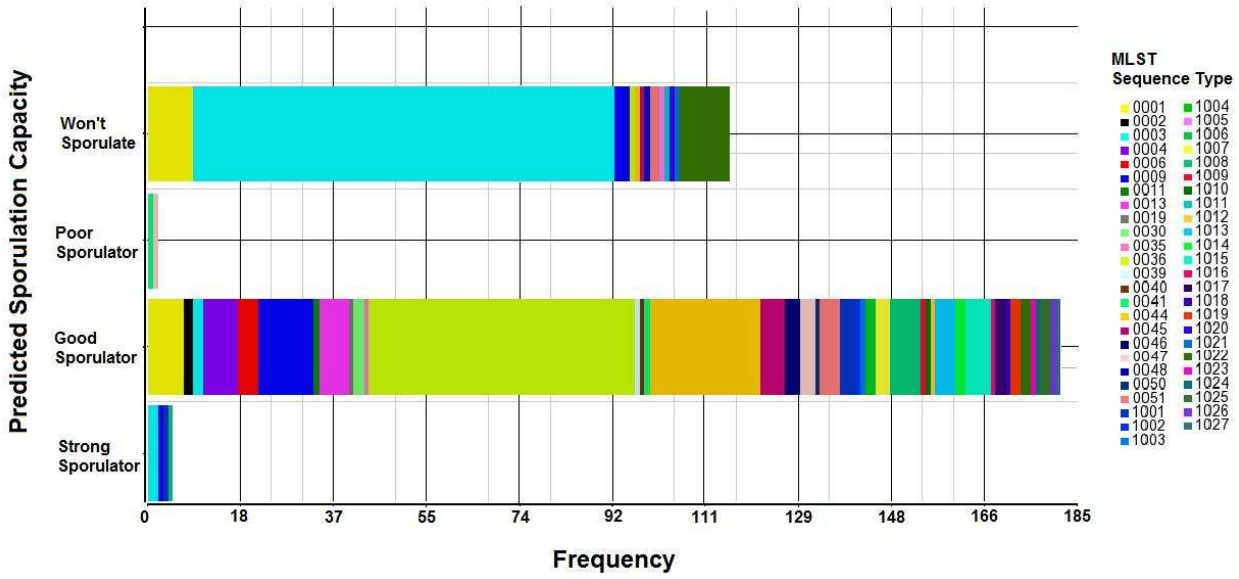

Figure 7. Frequency of predicted sporulation capacities taking MLST sequence type into consideration.

Interestingly, two isolates analyzed (ATCC43598 and SB201) had exceptionally large (about 5.5 kb) CF and 5′Fus PCR amplicons (compared to the expected sizes of 1079 bp and 1428 bp, respectively). Whether the insert and its excised form carry foreign DNA is purely conjecture at this time. Additional primers have been designed for sequencing by primer walking across these large fragments; but conclusive sequencing has yet not been obtained, owing to repetitive features within the inserted DNA. Primers annealed in multiple locations, resulting in messy (and un-interpretable) sequence data. Moreover, although interesting, analyses of these larger amplicons are beyond the scope and timeframe of this project. It is notable that SB201 was initially "flagged" as one of four outliers in the initial PFGE-*SmaI* pattern dendrogram (Figure 4). As mentioned previously, prior studies (PCR assays and PFGE analyses) on ATCC43598 were not completed by the ARNI department at the National Microbiology Laboratory.

Multi-Locus Sequence Typing (MLST)

The seven housekeeping gene loci amplified and sequenced for this study are listed in Table 7. The fragments analyzed within each locus were trimmed according to the guidelines for the "old" *C. difficile* MLST scheme on the Pasteur website. Three strains (CPL103, M178 and SB419) were eliminated from the MLST portion of this study owing to repeated sequencing failures. The number of individual alleles called for each of the seven loci ranged from six for *ddl*, *dutA*, *gmk*, *recA* and *tpi* to eleven for *sodA* and thirteen for *aroE*. For *ddl*, one of these six alleles was a null allele (designated allele type '999', denoting that no *ddl* gene is present); several amplification attempts remained unsuccessful and ultimately 999 was deemed an accurate determination since this allele

type is seen (albeit infrequently) by other groups. Four loci (*aroE*, *ddl*, *dutA* and *sodA*) revealed novel alleles. Three alleles identified for *aroE* were novel: designated new1 (four strains), new2 (one strain), and new3 (one strain). Two novel alleles were defined for *ddl*: new1 (16 strains) and new2 (one strain). Twelve strains had the novel *dutA* allele, designated new1 herein. Three novel alleles also were identified for *sodA*: new1 (one strain), new2 (five strains) and new3 (one strain). Data reporting allelic variations for the seven housekeeping genes are summarized in Table 7.

Table 7. Allelic variation in the seven *C. difficile* housekeeping genes analyzed via MLST.

Loci	Size (bp) of Fragment Analyzed	Number of Alleles Identified
aroE	408	13[a]
ddl	420	6[b]
dutA	309	6[c]
gmk	279	6
recA	378	6
sodA	408	11[d]
tpi	393	6

[a] three novel alleles were defined for *aroE*
[b] two novel alleles were defined for *ddl*
[c] one novel allele was defined for *dutA*
[d] three novel alleles were defined for *sodA*

For *aroE*, allele 2 was the most common allele type with a presence in 165 strains and allele 13 was in 53 strains, the second most common *aroE* allele. For *ddl*, allele 2 was present in 146 strains (the most common *ddl* allele) and allele 4 was present in 114 strains. For *dutA*, 233 strains had allele 2 whereas 47 strains had allele 5. Allele 1 was the most common *recA* allele (220 strains) whereas allele 2 was present in 71 strains. For *sodA*, allele 5 was present in 97 strains, followed by allele 1 in 92 strains and then allele 2 in 88 strains. For *tpi*, 229 strains contained allele 1 and 67 strains contained allele 2.

Allele identification for the sum of seven *C. difficile* housekeeping genes gave rise to 49 distinct allele profiles (denoted as multi-locus sequence types; STs). The

majority (31) of sequence types appeared only once or twice. Three sequence types accounted for over 54% of the total strains analyzed. ST3 represented 29.1% (89), ST36 represented 17.6% (54), and ST44 represented 7.5% (23) of the total strains analyzed. It was not surprising that ST3 was the most common sequence type revealed, since its MLST allelic profile (2,2,2,1,1,5,1) is comprised of the most common alleles for each locus. Data reporting the distribution of sequence types across the 306 strains analyzed are summarized in Figure 8.

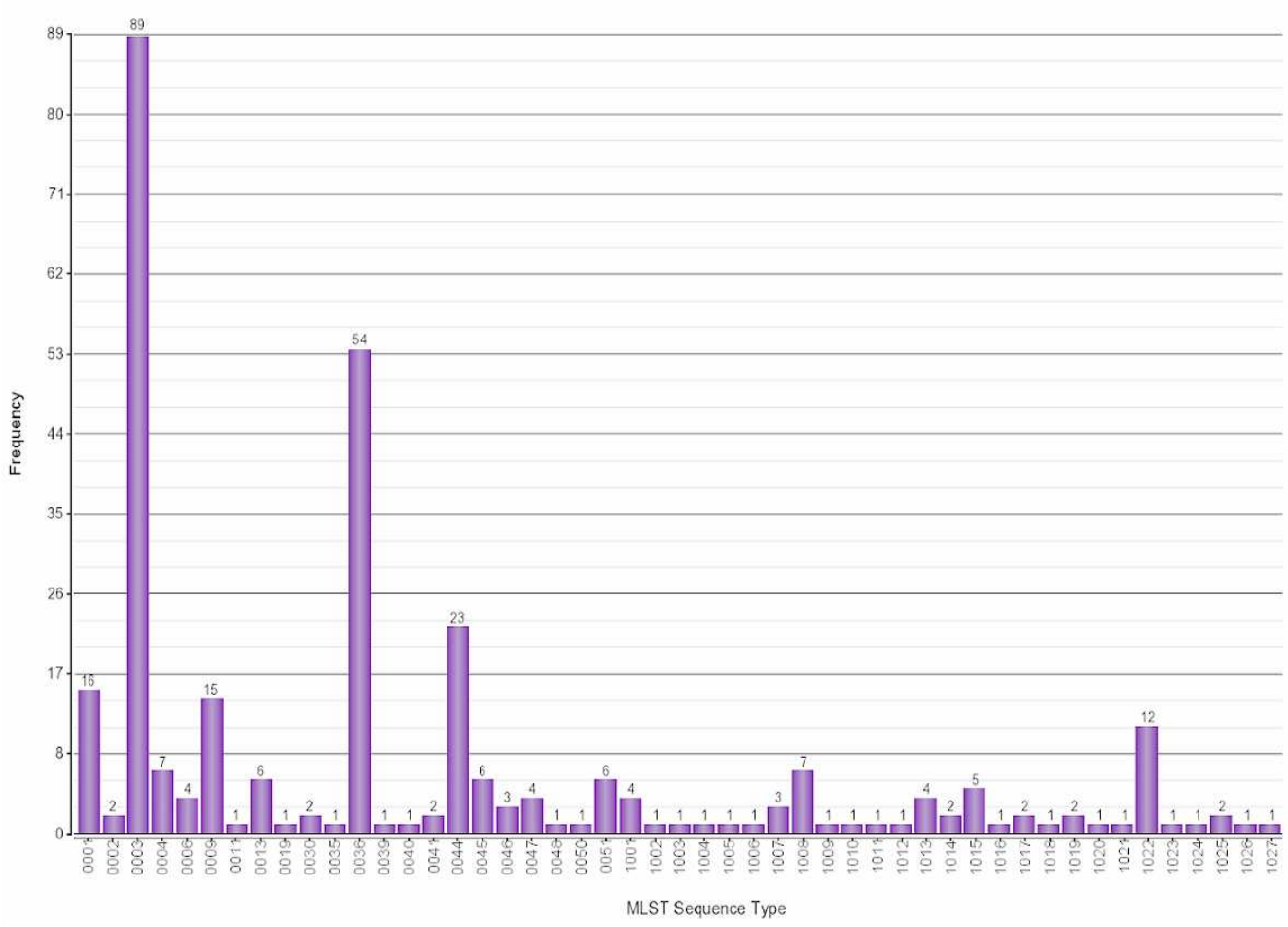

Figure 8. Distribution of MLST sequence types across the 306 strains analyzed.

A dendrogram displaying similarity between the defined *C. difficile* MLST sequence types (Figure 9) revealed a highly divergent strain cluster, comprised of ST2, ST8, ST26 and ST29 and the novel sequence types 1005, 1017, 1020 and 1021. The next most divergent group of strains included ST16, ST56, ST57 and ST1026. These twelve sequence types in these two divergent clusters represented allelic profiles that differed by three or more alleles when compared to the sequence types sharing higher degree of similarity (uppermost in dendrogram).

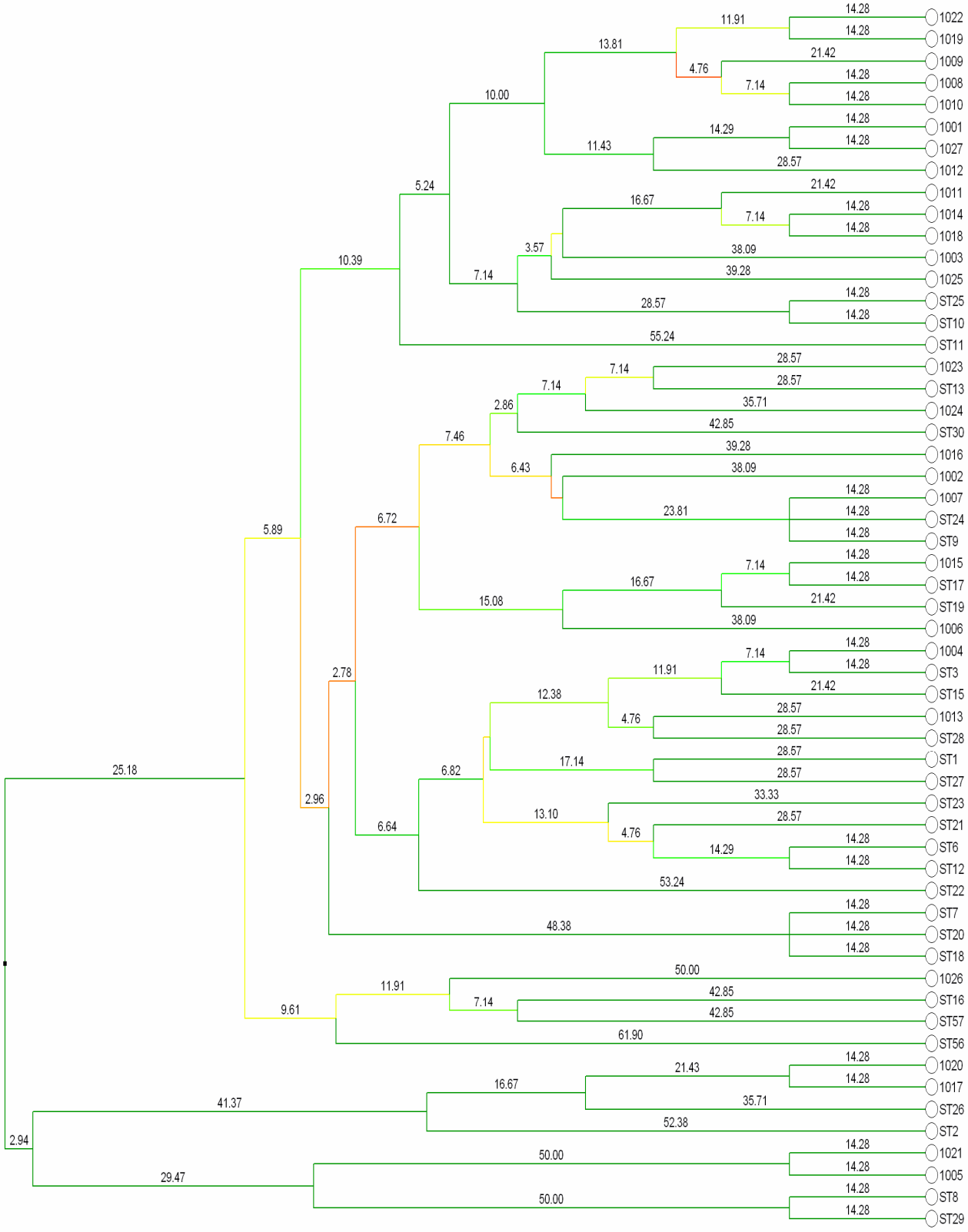

Figure 9. Dendrogram displaying MLST sequence type similarity. Dendrogram was generated in BioNumerics.

A few connections were made when comparing MLST sequence type to NAP type and predicted sporulation capacity. All strains designated as ST1 were NAP4 (representing 16 of the possible 18 NAP4's) and have been predicted to either be "incapable" of sporulation or conversely, to be "good" sporulators. In total, 89 strains were designated as ST3. The majority (86) of ST3 strains were NAP2 (with PFGE-*SmaI* pattern 0003); one strain's NAP type was not determined, another was NAP6 and the other was NAP untypable. Of ST3s, the vast majority (84) also were predicted to be "incapable" of sporulating, two to be "good" sporulators, two to be "strong" sporulators, and one strain's (W36) capacity for sporulation was undeterminable owing to inconsistent amplification results. The seven strains designated as ST4 all were predicted to be "good" sporulators (four were NAP12 and three were NAP untypable). Of 54 strains designated as ST36, 53 were predicted to be "good" sporulators whereas the remaining strain was predicted to be "incapable" of sporulation. Most of these ST36 strains were NAP untypable however, two were not determined and one was NAP2. All 23 strains designated as ST44 were predicted to be "good" sporulators, with one exception (predicted to be "incapable" of sporulation). Seventeen of these ST44 strains were NAP1 (24 NAP1 strains in the 299 samples analyzed), four were NAP untypable and two were not determined. Of twelve strains assigned the novel sequence type ST1022, ten were predicted to be "incapable" of sporulating and two were predicted to be "good" sporulators. Four of these strains were NAP1 and the remainder were untypable. ST13 and the novel sequence type ST1013 stood out as both were only comprised of strains with molecular predictions as "good" sporulators. The six ST13 strains were all of the NAP10s in the study and three of the four ST1013's were NAP11 (all NAP11s in

the study); the other strain was NAP untypable. Therefore, all NAP10s were designated as ST13 and all NAP11s were designated as ST1013. In addition, the four NAP12s studied were designated as ST4.

Toxin Regulators

This study intended to analyze three *C. difficile* toxin regulators, *tcdC*, *tcdR* and *cdtR*. Sequencing was completed for all three toxin regulators; however, analyses were only completed for *tcdC* sequencing data owing to time constraints. Hence, allelic profiles based on all three toxin regulators (*tcdC*, *tcdR* and *cdtR*) could not be determined to date, but will be completed in the near future.

Eight strains (SB373, SB400, C333, W7, W36, W44, ATCC700057 and 765) were eliminated from the *tcdC* study owing to repeated sequencing failures or incomplete sequencing data. Twenty-six different *tcdC* alleles were identified in 301 *C. difficile* strains; ten novel alleles were revealed. The most common *tcdC* alleles were sc3 (89 strains) and sc9 (89 strains). The novel *tcdC* allele, designated as New4, was the third most common (27 strains). Data reporting *tcdC* allelic variations are summarized in Table 8. Thorough analyses of deletions, mutations, and predicted protein sequences are yet to be completed for the *tcdC* alleles. Deletions and mutations within the nucleotide sequence introduce reading frame shifts that may result in premature stop codons and therefore, truncated TcdC proteins. Deletions of various sizes were detected in the 301 *C. difficile* strains analyzed and often occurred between positions 277 to 388 of the *tcdC* gene; an area characterized by tandem repeats (TRs). Mutations also were observed to be

more frequent in this region of DNA. Recombination events between repeat elements are suspected for giving rise to the multiple deletions and mutations observed.

Table 8. Allelic variation in the *C. difficile* toxin regulator *tcdC* for 301 strains[a].

Allele[b]	Size (bp) of Fragment Analyzed	Number of Strains with Allele
A1[c]	636	4
mb1[d]	657	11
mb4[d]	657	5
mb6[d]	657	1
mb8[d]	657	4
sc1[c]	656	21
sc2[c]	657	1
sc3[c]	675	89
sc4[c]	657	1
sc5[c]	675	6
sc7[c]	675	3
sc9[c]	675	89
sc13[c]	675	14
sc14[c]	675	1
sc15[c]	675	3
sc16[c]	639	2
New1	675	1
New2	675	1
New3	675	2
New4	675	27
New5	657	1
New6	657	1
New7	657	2
New8	657	4
New9	639	5
New10	657	2

[a] Eight *C. difficile* strains were eliminated from the *tcdC* study owing to repeated sequencing failures or incomplete sequencing data.
[b] Ten novel alleles were identified for *tcdC* and were designated as New1 to New10.
[c] All "sc-" *tcdC* alleles and allele A1 were first define by Curry *et al*., 2007.
[d] All "mb-" *tcdC* alleles were defined by Rupnik and Kocuvan on NCBI under accession numbers FJ409548 through FJ409555 (unpublished data).

Figure 10 displays a dendrogram revealing sequence similarity of the 26 *tcdC* alleles observed. Coloured allele designations correlated with deletions detected by ARNI when prior *tcdC* PCR assays were completed (teal for deletions under 18-bp; purple for 18-bp deletions; orange for deletions greater than 18-bp). Although alleles mb8, sc1, sc2 and New7 (all containing an 18-bp deletion) clustered together and alleles

sc16 and New9 (greater than 18-bp deletion) clustered together, clustered grouping based solely on deletion type was not generally observed. Allele A1 appeared to have the greatest divergence owing to multiple sequence deletions. However, all allele sequences were very similar as is reflected in the right-sidedness of the dendrogram; sequences vary by one base per 100 bases within deletion classifications. These results, clustering based on *tcdC* DNA sequences, do not reveal whether deletions occurred prior to mutations or vice versa. Therefore, the *tcdC* locus on *C. difficile*'s chromosome appears to continuously evolve via random mutation and deletion events.

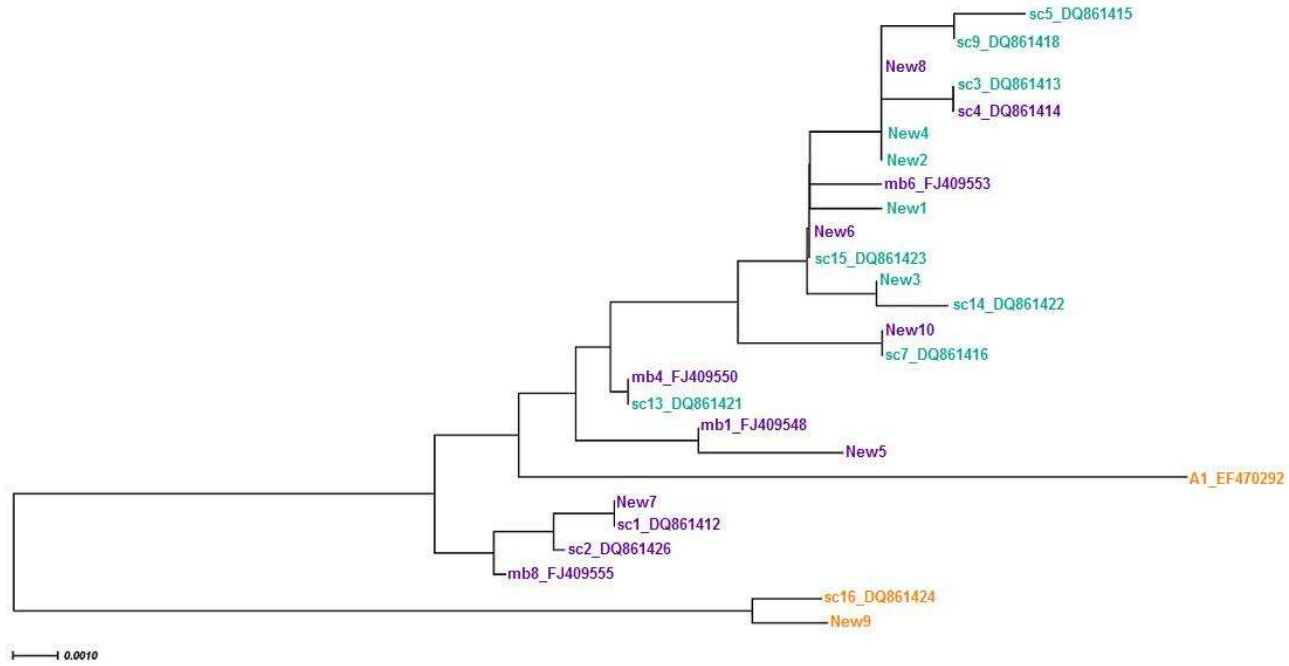

Figure 10. Dendrogram displaying allelic variation within *C. difficile*'s *tcdC* toxin regulator and their associated relatedness. Dendrogram was generated in MEGA and modified for visual clarification with Dendroscope.

When taking NAP type, predicted sporulation capacities, MLST sequence type and *tcdC* allele into consideration, a few connections were evident. All strains designated as *tcdC* allele mb1 were MLST ST1022 (eleven of the twelve possible ST1022 strains; the twelfth strain, C333, was eliminated from *tcdC* analyses due to short sequencing products). Seven of these strains were NAP untypable, the remaining four were NAP1;

nine strains were predicted to be “incapable” of sporulating, two to be “good” sporulators. Twenty-seven strains were designated with the novel *tcdC* allele New4; 24 strains were predicted to be “good” sporulators and one strain each for the predicted sporulation capacities of “incapable”, “poor” and “strong”. Within these 27 strains, MLST sequence types (11 observed) and NAP types (NAP2, NAP3, NAP4, NAP6, NAP untypable and not determined) were inconsistent. *tcdC* allele sc1 was observed in 21 strains; 19 were designated to be NAP1 (the remaining two were not determined). Predicted sporulation capacity was “good” for 19 strains and “incapable” for the other two strains; 18 strains were ST44, two were ST1019 and one was ST1011. The majority (87) of the 89 strains designated to have *tcdC* allele sc3 were also designated to be ST3 (in 306 strains assigned sequence types, 89 strains were ST3; ST3 strains SB373 and W36 were eliminated from the *tcdC* portion of this study). The other two *tcdC* allele sc3 strains (H03 and ATCC43594) were designated to be ST9. Strains were predicted to be “incapable” of sporulation (85 strains) or to be “good” (two strains) or “strong” (two strains) sporulators. All strains designated to have sc5 *tcdC* allele also were NAP untypable and designated as ST51 (all ST51 strains in the study); four strains were predicted to be “good” sporulators and the other two were predicted to be “incapable” of sporulating. Allele sc9 was observed in 89 strains; within this population, 15 MLST sequence types were defined (53 strains were ST36), with a variety of NAP types. The majority (79) of sc9 strains were predicted to be “good” sporulators and the remaining ten strains were predicted to be “incapable” of sporulation.

Discussion

As of thesis submission, the study blind had not yet been broken; hence thorough epidemiological linkage analyses and strain characterization could not be completed.

Predicting Sporulation Capacities via PCR Amplification

Healthcare facilities bring together a variety of CDI risks including, individuals exposed to antibiotics, a spore contaminated environment, inadequate sanitation and cleaning procedures and a constant flux of patients (Cookson 2007; Rupnik *et al*., 2009). All risks can give rise to the rapid spread of *C. difficile* as a healthcare-associated infection. Spore contamination and healthcare cleaning procedures are of great interest because a means that diminishes these risks can greatly reduce transmission. To date, standard healthcare cleaning protocols do not incorporate a method adequate for removing *C. difficile* spores from the environment (Fawley *et al*., 2007). Only under certain circumstances, such as *C. difficile* outbreaks, are cleaning procedures enhanced to eliminate spores from the environment. Thus, the ability to predict hypersporulating strains would be highly advantageous for proactive protection of patients, staff, and visitors in healthcare facilities.

Prokaryotic organisms sometimes use chromosomal rearrangements for regulating important developmental processes. Haraldsen and Sonenshein (2003) found a 14.6 kb prophage-like element, $skin^{Cd}$, interrupted the *sigK* gene of some *C. difficile* strains and that regulated excision of $skin^{Cd}$ is critical for achieving proper temporal activation of σ^K, the translation product of *sigK* that is essential for late sporulation gene expression. In agreement with Haraldsen and Sonenshein (2003), not all strains have disruption of *sigK*

and the absence of *skin*^{Cd} correlates with a sporulation defect (i.e. sporulation will not occur). If present, *skin*^{Cd} excision from the chromosome occurs during sporulation, generating a circular molecule and uninterrupted *sigK*. Thus, a strain's efficiency for excising *skin*^{Cd} reflects its propensity for sporulation – strains that excise *skin*^{Cd} rapidly have greater sporulation capacities. Haraldsen and Sonenshein (2003) and this study have demonstrated that predictions for each strain's putative sporulation capacity are possible through careful molecular analysis of the genetic markers *sigK* and *skin*^{Cd}. In the analysis of province-wide *C. difficile* strains, four predicted sporulation capacity scenarios arose, reflecting varying sporulation ability and thus, risk for contaminating the environment. Removal of spores from the environment would reduce transmission between patients, staff, and visitors. Based on these study findings, it would be prudent to upgrade cleaning protocols when high levels of strain sporulation are predicted. Therefore, a cleaning protocol altered to the infecting *C. difficile* strain's sporulation capacity shows greater promise for reducing spore contamination than generalized cleaning protocols currently used throughout hospitals.

The majority of MLST allele ST3 strains were predicted to be "incapable" of sporulation whereas, the majority of both ST36 and ST44 were predicted to be "good" sporulators. Based on these results, it is hypothesized that ST3 strains arose in community-acquired infections owing to ingestion of live vegetative *C. difficile* microorganisms contained within contaminated food products. Conversely, infections arising from ST36 and ST44 strains could be hospital- or health facility-acquired when individuals in close contact with *C. difficile* spores present in the environment

subsequently ingest them and develop CDI. Forthcoming epidemiological data analysis will be undertaken to investigate these two hypotheses.

Molecular sporulation capacity predictions are currently being validated by our collaborators (blinded) at St. Boniface Research Institute using a subset of strains from the sequenced data set. If molecular based predictions and *in vitro* spore counts are found to be congruent, a multiplex PCR assay will next be optimized using the same four primer pairs as was used in this study (this is possible as differences in fragment sizes are large enough for efficient separation in an agarose gel), thus reducing four PCR reactions into one. These multiplex reactions will be validated with newly-prepared spore DNAs from an expanded panel of *C. difficile* strains. It is hoped that such a validated multiplexed assay might be used in clinical microbiology laboratories to assess the risk of sporulation for inpatients diagnosed with CDI.

If the forthcoming *in vitro* spore counts fail to validate the molecular sporulation capacity predictions, there are several plausible explanations for why such a high number of provincial strains exhibited the "incapable" of sporulation scenario. i) Spore preparations for DNA isolation are difficult to monitor, inconsistent, and may contain variable amounts of vegetative cells; improved methods for spore preparations are currently being investigated at St. Boniface Research Institute to limit this possibility. ii) Susceptible individuals may be succumbing to CDI caused by indigenous *C. difficile* (from within strains), since many of us are colonized by *C. difficile* soon after birth, rather than external sources; or iii) all primers used to amplify CF, 3′Fus, or 5′Fus simultaneously have sequence divergence in some strains, resulting in false negative PCR results. $skin^{Cd}$ insert sequence analysis in strains deemed "incapable" of sporulating

would be worthwhile to assess whether sequence divergence has occurred, but is beyond the scope of this project.

Multi-Locus Sequence Typing (MLST)

MLST is based on allelic differences in the nucleotide sequences of several housekeeping genes among bacterial strains and thus, distinguishes strains based on observed variations in the nucleotide sequences of several loci rather than the degree of sequence variation in any single gene or locus (Bouvet and Popoff 2008). Within this study, 299 *C. difficile* strains isolated from patients within Manitoba hospitals were analyzed along with seven control *C. difficile* strains. Allelic variation within seven housekeeping genes (*aroE*, *ddl*, *dutA*, *gmk*, *sodA*, *recA* and *tpi*) gave rise to 49 different MLST sequence types (STs), 27 of which were novel. As mentioned previously, ST3 strains are hypothesized to be community-acquired through the ingestion of live vegetative *C. difficile* microorganisms whereas, ST36 and ST44 strains may give rise to hospital-acquired infection owing to their ability to be "good" sporulators.

According to Lemee *et al*. (2004), ST2, ST7, and ST11 strains should not have been observed in this study as they reported ST7 and ST11 strains as nontoxigenic ($TcdA^-/TcdB^-$) and ST2 strains as $TcdA^-/TcdB^+$ variant strains. All strains studied herein were designated as $TcdA^+/TcdB^+$ by former PCR assays (performed elsewhere); yet, ST2 was observed twice and ST11 was observed once. Confirmation that no data translation error has occurred is advised. Similar to Lemee *et al*. (2004), some MB sequence types were revealed to be very divergent in cluster analysis; ST2, ST8, ST26 and ST29, along with the novel sequence types ST1005, ST1017, ST1020 and ST1021.

After this study began, the MLST typing scheme for *C. difficile* was changed on the Pasteur website; this includes the replacement of the *ddl* loci with *groEL* and alteration of primer sets for *gmk* and *sodA*. Soon after, Griffiths *et al*. (2009) proposed yet another new seven-loci MLST scheme – replacing four loci (*aroE*, *ddl*, *dutA* and *gmk*) with *adk*, *atpA*, *dxr* and *glyA* loci along with new primer sets for the remaining three loci under study (*recA*, *sodA* and *tpi*). These recent changes indicate that *C. difficile* MLST is currently in a state of flux. Analyses that will clearly correlate sequence types across the now three *C. difficile* MLST schemes will be highly appreciated by future *C. difficile* MLST researchers.

Toxin Regulators

MLST that takes virulence-associated genes into consideration is believed to have a higher discriminatory power than traditional MLST in which only several housekeeping genes are analyzed (Bouvet and Popoff 2008). The aim of this study was to analyze variability within toxin regulator genes *tcdC*, *tcdR* and *cdtR* in *C. difficile* strains from Manitoba while assessing the relatedness of the isolates, and to also determine whether connections exist between traditional MLST based on housekeeping genes and sequence types that include virulence-associated genes. Due to time constraints, initial sequence analyses have only been completed on *tcdC*, the negative regulator of the *tcdA* and *tcdB* toxin genes. Genetic variation amongst *tcdC* alleles of 301 Manitoba *C. difficile* strains gave rise to 26 *tcdC* alleles, 11 of which are novel.

Tandem repeats between positions 277 to 388 of the *tcdC* gene, giving rise to eight-amino acid repeats, have been observed in this and other studies (Curry *et al*., 2007;

Bouvet and Popoff 2008; Murray *et al*., 2009). These tandem repeats are known to be responsible for *recA*-independent intragenic homologous recombination which subsequently results in deletion or duplication of repeats by slipped-strand mispairing (Spigaglia and Mastrantonio 2002; Bouvet and Popoff 2008). These deletions within *tcdC*, primarily a one base-pair deletion at position 117, are believed to be the cause of hypertoxic epidemic strains (Bouvet and Popoff 2008; Dupuy *et al*., 2008); thus, *tcdC* nucleotide sequence is of great interest. The one base-pair deletion at position 117 causes a premature stop codon, resulting in a truncated TcdC protein (Bouvet and Popoff 2008; Dupuy *et al*., 2008). When full-length TcdC is not present to negatively regulate the production of toxins TcdA and TcdB, increased toxin production and strain hypertoxicity result. High levels of toxin expression have been observed in strains with a one-base pair deletion at position 117 (Spigaglia and Mastrantonio 2002). However, Murray *et al*. (2009) did not have data that supported a direct role for alterations in the *tcdC* gene as a predictor of hyperproduction of toxins TcdA and TcdB, at least in NAP1-related strains. Thorough analyses of *tcdC* deletions and mutations and their resultant effect on the TcdC translated product are yet to be completed for these Manitoba strains. However, shifts in reading frames and premature stop codons giving rise to truncated TcdC proteins are suspected in some strains as other studies have described them (Curry *et al*., 2007; Bouvet and Popoff 2008; Dupuy *et al*., 2008). It would be interesting to investigate whether the *tcdC* sequence variants described herein functionally affect TcdC and thereby, production of toxins TcdA and TcdB. Completion of *tcdR* and *cdtR* sequence analyses will be undertaken for thorough strain characterization and epidemiological linkage analyses.

Only after the study blind is removed, can thorough epidemiological analyses be completed – including linking sequence types with community versus hospital acquisition and whether a specific sequence type was predominant in a particular outbreak or multiple outbreaks within a given hospital.

Conclusions

1. It was possible to predict each *C. difficile* strain's putative sporulation capacity based on PCR results of *sigK::skin*^{Cd} genetic markers.
    a. Follow-up studies are required to confirm the sporulation capacity predictions.
2. We suspect infections due to MLST designated ST3 *C. difficile* strains are community-acquired infections whereas, infections owing to ST36 and ST44 strains may be hospital-acquired.
3. MLST allows for discrimination of *C. difficile* strains; however, one reliable scheme must be adopted.
4. The toxin regulator *tcdC* gene nucleotide sequence is highly variable and this variability may influence the toxigenicity of *C. difficile*.

References

Alfa, M.J., Dueck, C., Olson, N., DeGagne, P., Papetti, S., Wald, A., Lo, E., and G. Harding. 2008. UV-visible marker confirms that environmental persistence of *Clostridium difficile* spores in toilets of patients with *C. difficile*-associated diarrhea is associated with lack of compliance with cleaning protocol.e. BMC Infectious Diseases, 8: 64-70.

Bouvet, P.J.M. and M.R. Popoff. 2008. Genetic relatedness of *Clostridium difficile* isolates from various origins determined by triple-locus sequence analysis based on toxin regulatory genes *tcdC*, *tcdR*, and *cdtR*. Journal of Clinical Microbiology, 46(11): 3703-3713.

B. Cookson. 2007. Hypervirulent strains of *Clostridium difficile*. Postgraduate Medical Journal, 83: 291-295.

Curry, S.R., Marsh, J.W., Muto, C.A., O'Leary, M.M., Pasculle, A.W., and L.H. Harrison. 2007. *tcdC* genotypes associated with severed TcdC truncation in an epidemic clone and other strains of *Clostridium difficile*. Journal of Clinical Microbiology, 45(1): 215-221.

Dawson, L.F., Valiente, E., and B.W. Wren. 2009. *Clostridium difficile* – A continually evolving and problematic pathogen. Infection, Genetics and Evolution, 9(6): 1410-1417.

Dupuy, B., Govind, R., Antunes, A., and S. Matamouros. 2008. *Clostridium difficile* toxin synthesis is negatively regulated by TcdC. Journal of Medical Microbiology, 57: 685-689.

Fawley, W.N., Underwood, S., Freeman, J., Baines, S.D., Saxton, K., Stephenson, K., Owens, R.C., and M.H. Wilcox. 2007. Efficacy of hospital cleaning agents and germicides against epidemic *Clostridium difficile* strains. Infection Control and Hospital Epidemiology, 28(8): 920-925.

Griffiths, D., Fawley, W., Kachrimanidou, M., Bowden, R., Crook, D.W., Fung, R., Golubchik, T., Harding, R.M., Jeffery, K.J.M., Jolly, K.A., Kirton, R., Peto, T.E., Rees, G., Stoesser, N., Vaughan, A., Walker, A.S., Young, B.C., Wilcox, M., and K.E. Dingle. 2009. Multilocus sequence typing of *Clostridium difficile*. Journal of Clinical Microbiology, doi:10.1128/JCM.01796-09.

Hächler, H., Berger-Bächi, B., and F.H. Kayser. 1987. Genetic characterization of a *Clostridium difficile* erythromycin-clindamycin resistance determinant that is transferable to *Staphylococcus aureus*. Antimicrobial Agents and Chemotherapy, 31(7): 1039-1045.

Haraldsen, J.D. and A.L. Sonenshein. 2003. Efficient sporulation in *Clostridium difficile* requires disruption of the σ^K gene. Molecular Microbiology, 48(3): 811-821.

Kuijper, E.J., van den Berg, R.J., and J.S. Brazier. 2009. Comparison of molecular typing methods applied to *Clostridium difficile*. Chapter 13, p. 159-171. In D.A. Caugant (ed.). Molecular Epidemiology of Microorganisms, Methods in Molecular Biology: (Volume 551) Humana Press, a part of Springer Science.

Lemee, L., Dhalluin, A., Pestel-Caron, M., Lemeland, J.F., and J.L. Pons. 2004. Multilocus sequence typing analysis of human and animal *Clostridium difficile* isolates of various toxigenic types. Journal of Clinical Microbiology, 42(6): 2609-2617.

Lemée, L., Bourgeois, I., Ruffin, E., Collignon, A., Lemeland, J.F., and J.L. Pons. 2005. Multilocus sequence analysis and comparative evolution of virulence-associated genes and housekeeping genes of *Clostridium difficile*. Microbiology, 151: 3171-3180.

Murray, R., Boyd, D., Levett, P.N., Mulvey, M., and M.J. Alfa. 2009. Truncation in the *tcdC* region of the *Clostridium difficile* PathLoc of clinical isolates does not predict increased biological activity of Toxin B or Toxin A. BMC Infectious Diseases, 9: 103-113.

Paredes, C.J., Alsaker, K.V., and E.T. Papoutsakis. 2005. A comparative genomic view of clostridial sporulation and physiology. Nature Reviews Microbiology, 3(12): 969-978.

Riegler, M., Sedivy, R., Pothoulakis, C., Hamilton, G., Zacherl, J., Bischof, G., Cosentini, E., Feil, W., Schiessel, R, LaMont, T, and E. Wenzl. 1995. *Clostridium difficile* toxin B is more potent than toxin A in damaging human colonic epithelium in vitro. The Journal of Clinical Investigation, 95(5): 2004-2011.

Rupnik, M., Wilcox, M.H., and D.N. Gerding. 2009. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. Nature Reviews Microbiology, 7: 526-536.

Spigaglia, P. and P. Mastrantonio. 2002. Molecular analysis of the pathogenecity locus and polymorphism in the putative negative regulator of toxin production (TcdC) among *Clostridium difficile* clinical isolates. Journal of Clinical Microbiology, 40(9): 3470-3475.

Spigaglia, P. and P. Mastrantonio. 2004. Comparative analysis of *Clostridium difficile* clinical isolates belonging to different genetic lineages and time periods. Journal of Medical Microbiology, 53: 1129-1136.

Zaiß, N.H., Rupnik, M., Kuijper, E.J., Harmanus, C., Michielsen, D., Janssens, K., and U. Nübel. 2009. Typing *Clostridium difficile* strains based on tandem repeat sequences. BMC Microbiology, 9: 6-16.

Appendix

Appendix 1. Nucleotide sequences of novel *aroE* alleles defined herein for *C. difficile*.

aroE-New1: atccacaagagagtattcgagatattgaatattgaaagtgcttataaaaactttgaaatttcaaaagaggatatatcta
agctagacggtgctataaaattattaggaatacaaggggtaaatgttactgtaccatataaggagcgtattatgaaatatcttgattttata
tcaccagaggcaaaacgtataggagcagttaatacaattttactcagagaaaatatgttatatggttacaatacagattactttggtcttga
tagtatgtttaaaatggctaatattgatgttcaaggaaaagtagcagttatattaggaactggaggtgcttctaaagctgctcttacatatttt
attgattctggtatagaaaagctttatgtttcaactagaaaaaaggatgat

aroE-New2: atccacaagagagtattcgagatattgaatattgaaagtgcttataaaaactttgaaatttcaaaagaagatatatcta
agctagacagtgctataaaattattaggaatacaaggagtaaatgttactgtaccatataaggagcgtattatgaaatatcttgattttata
tcaccagaggcaaaacgtataggagcagttaatacaattttactcagagaaaatatgttatatggttacaatacagattactttggtcttga
tagtatgtttaaaatggctaatattgatgttcaaggaaaagtagcagttatattaggaactggaggtgcttctaaagctgctcttacatatttt
attgattctggtatagaaaagctttatgtttcaactagaaaaaaggatgat

aroE-New3: atccacaagagaatattcgagatattgaatattgaaagtgcttataaaaactttgaaattccaaaagaagatatatcta
agctagacggtgctataaaattattaggaatacaaggagtaaatgttactgtaccatataaggagcgtattatgaaatatcttgattttata
tcaccagaggcaaaacgtatagaagcagttaatacaattttactcagagaaaatatgttatatggttacaatacagattactttggtcttga
tagtatgtttaaaatggctaatattgatgttcaaggaaaagtagcagttatattaggaactggaggtgcttctaaagctgctcttacatatttt
attgattctggtatagaaaagctttatgtttcaactagaaaaaaggatgat

Appendix 2. Nucleotide sequences of novel *ddl* alleles defined herein for *C. difficile*.

ddl-New1: tcagataaatttaatgtggctgaactgtcaaaagacttaacttatccattatttataaaacctgtgcgtgcaggttcatcattt
ggtattagtaaaatttataaaaaagaagagttacaagtagcagtagacttggcttttgaacatgataatgaagtaataattgaggaaaatatt
gatggttttgaggtagggtgtgcagttttaggaaacgatgacttacttgttggaagagtagatgaaatagaattgtcaaaaggatttttgat
tatacagaaaaatatacattgaaaagttcaaaaattcatatgcgagctagaattgatttggattgtgagaaaagaatacaagagacagcaa
aaattatttatagaactcttggatgttctggatttgctagagtagatatgtttttaact

ddl-New2: tcagataaatttaatgtggctgaactgtcaaaagacttaacttatccattatttataaaacctgtgcgtgcaggttcatcattt
ggtattagtaaaatttataaaaaagaagagttacaagtagtagtagacttggcttttgaacatgataatgaagtaataattgaggaaaatatt
gatggttttgaggtagggtgtgcagttttaggaaacgatgacttacttgttggaagagtagatgaaatagaattgtcaaaaggatttttgat
tatacagaaaaatatacattgaaaagttcaaaaattcatatgccagctagaattgatttggattgtgagaaaagaatacaagagacagcaa
aaattatttatagaactcttggatgttctggatttgctagagtagatatgtttttaact

Appendix 3. Nucleotide sequence of novel *dutA* allele defined herein for *C. difficile*.

dutA-New1: gttgtaatacctcctagagagactaaacttatcaaaactgggatatgtatagagcttcctactatgacagaagcgcag
gtaagacctagaagtgggcttgccttaaagcattctgttactgtattaaatacaccaggaactattgacgaaggatacagaggagagctt
aaaattatattaataaatcatgggaaaaatgattttaaggttgaaaagcatatgaaaattgctcaaatgatagtaaaacctatctatgatatta
acatagaagaagtaaaagaattaagtgactccgaaagaggaaaaggt

Appendix 4. Nucleotide sequences of novel *sodA* alleles defined herein for *C. difficile*.

sodA-New1: catcatgataagcattatcaagcttatgttgataaattaaatgctgctcttgaaaaatatcctgagctttataattattcttt
atgtgaattattgcaaaatttagattctttacctaaagatattgctacaactgtaagaaataatgcaggtggagcttataatcataaattctttt
ttgatataatgacgccagaaaaaaccataccttctgaatctttaaaagaagctattgatagagactttggttcttttgaaaaatttaagcaag
aattccaaaaatctgctttagatgtctttggttctggttgggcttggcttgtagctactaaagatgggaaattatctattatgactactccaaat
caggatagccctgtaagtaaaaacctaactcctataatagga

sodA-New2: catcatgataagcattatcaagcttatgttgataaattaaatgctgctcttgaaaaatatcctgagctttataattattcttt
atgtgaattattgcaaaatttagattctttacctaaagatattgctacaactgtaagaaataatgcaggtggggcttataatcataaattctttt
ttgatataatgacaccagaaaaaaccataccttctgaatctttaaaagaagctattgatagagactttggttcttttgaaaaatttaagcaag
agttccaaaaatctgctttagatgtctttggttctggttgggcttggcttgtagctactaaagatgggaaattatctattatgactactccaaat
caggatagccctgtaagtaaaaacctaactcctataatagga

sodA-New3: catcatgataagcattatcaagcttatgttgataaattaaatgctgctcttgaaaaatatcctgagctttataattattcttt
atgtgaattattgcaaaatttagattctttacctaaagatattgctacaactgtaagaaataatgcaggtggagcttataatcataaattctttt
ttgatataatgacaccagaaaaaaccataccttctgaatctttaaaagaagctattgatagagactttggttcttttgaaaaatttaagcaag
agttccaaaaatctgctttagatgtctttggttctggttgggcttggcttatagctactaaagatggcaaattatctattatgactactccaaat
caggatagccctgtaagtaaaaacctaactcctataatagga

Appendix 5. Novel MLST sequence types defined herein for *C. difficile*.

aroE	*ddl*	*dutA*	*gmk*	*recA*	*sodA*	*tpi*	Sequence Type
1	2	5	1	1	New2	1	1001
6	4	2	2	1	New3	1	1002
13	4	2	1	1	New1	1	1003
6	2	2	1	1	5	1	1004
7	2	4	6	7	9	5	1005
5	1	2	1	6	1	2	1006
5	4	2	2	11	1	1	1007
13	2	5	1	1	2	1	1008
13	1	5	1	1	2	1	1009
13	2	5	1	1	1	1	1010
11	4	5	1	1	2	1	1011
New2	2	5	1	1	5	1	1012
New1	2	2	1	1	1	1	1013
11	4	2	1	1	2	1	1014
5	2	2	2	1	1	2	1015
5	999	2	2	1	5	1	1016
4	999	3	7	3	5	8	1017
11	New2	2	1	1	2	1	1018
13	New1	5	1	1	2	1	1019
4	New1	3	7	3	5	8	1020
7	New1	4	6	7	9	5	1021
13	New1	New1	1	1	2	1	1022
1	1	2	2	1	1	1	1023
2	1	2	2	2	1	1	1024
3	4	2	9	1	2	1	1025
5	1	1	2	1	5	2	1026
New3	2	5	1	1	New2	1	1027

Appendix 6. Distribution of MLST sequence types used herein for 306 strains of *C. difficile*.

Sequence Type	Number of Isolates	Hospital Distribution
1	16	St. B (5), Conc. (2), Vic. (2), Mis. (1), HSC (1), SO (1), Westman (2), Cadham (2)
2	2	St. B (1), Controls (1)
3	89	St. B (32), Grace (21), Conc. (15), HSC (5), SO (3), Westman (8), Cadham (5)
4	7	St. B (4), Westman (2), Cadham (1)
6	4	St. B (1), Grace (1), Vic. (1), Westman (1)
9	15	St. B (3), Grace (3), Vic. (1), HSC (2), SO (1), Westman (3), Cadham (1), Controls (1)
11	1	Controls (1)
13	6	Grace (1), Vic. (1), Mis. (1), Westman (2), Cadham (1)
19	1	Grace (1)
30	2	HSC (1), Westman (1)
35	1	Grace (1)
36	54	St. B (14), Grace (2), Conc. (2), Vic. (15), Mis. (3), HSC (6), SO (6), Westman (4), Cadham (1), Controls (1)
39	1	Conc. (1)
40	1	Grace (1)
41	2	Conc. (1), Vic. (1)
44	23	St. B (6), Grace (2), Conc. (2), Vic. (2), Mis. (1), SO (1), Westman (5), Cadham (2), Controls (2)
45	6	St. B (1), Conc. (2), Westman (2), Cadham (1)
46	3	Vic. (1), HSC (1), SO (1)
47	4	Conc. (1), Westman (1), Cadham (1), Controls (1)
48	1	Westman (1)
50	1	Westman (1)
51	6	Grace (2), Westman (3), Cadham (1)
1001	4	Westman (2), Cadham (2)
1002	1	HSC (1)
1003	1	Mis. (1)
1004	1	Cadham (1)
1005	1	St. B (1)
1006	1	HSC (1)
1007	3	HSC (2), Westman (1)
1008	7	St. B (2), HSC (1), Westman (3), Cadham (1)
1009	1	Cadham (1)
1010	1	Cadham (1)
1011	1	Cadham (1)
1012	1	Westman (1)
1013	4	Grace (1), Vic. (2), SO (1)
1014	2	Grace (1),Westman (1)
1015	5	St. B (1), Grace (1), Westman (2), Cadham (1)
1016	1	Cadham (1)
1017	2	HSC (1), Westman (1)
1018	1	HSC (1)
1019	2	HSC (1), Westman (1)
1020	1	Cadham (1)
1021	1	Cadham (1)
1022	12	Conc. (1), Vic. (4), Mis. (1), HSC (1), Westman (2), Cadham (3)
1023	1	Conc. (1)
1024	1	HSC (1)
1025	2	Mis. (2)
1026	1	SO (1)
1027	1	St. B (1)

Abbreviations: St.B (St. Boniface), Conc. (Concordia), Vic. (Victoria), Mis. (Misericordia), HSC (Health Sciences Centre), SO (Seven Oaks), Westman (Westman Regional), Cadham (Cadham Provincial Lab).

Appendix 7. Nucleotide sequences of novel *tcdC* alleles defined herein for *C. difficile*.

New1: aaaaaaatgagggtaacgaatttagtaatgaaggaaaaggaagctctaagaaaataattaaattctttaagagcacaaa
tggtattgctctactggcatttatttgggcgtgttttttggcaatatatcctcaccagcttgttctgaagaccatgaggaggtcatttct
aatcaaacatcagttatagattctcaaaaaacagaaatagaaactttaaatagcaaattgtctgatgctgaaccatggttcaaaatga
aagacgacgaaaagaaagctattgaagctgaaaatcaacgtaaagctgaagaagctaaaaaagctgaagaagctaaaaaggc
tgaagaacaacgtaaaaaagaagaagaggagaagaaaggatatgatactggtattacttatgaccaattagctagaacacctgat
gattataagtacaaaaaggtaaaatttgaaggtaaggttattcaagttattgaagatggtgatgaggtgcaaataagattagctgtgt
ctggaaattatgataaggttgtactatgtagttataaaaaatcaataactccttcaagagtattagaggatgattacataactataaga
ggtataagtgctggaactataacttatgaatcaactatgggtggaaatataactataccaggg

New2: aaaaaaatgagggtaacgaatttagtaatgaaggaaaaggaagctctaagaaaataattaaattctttaagagcacaaa
gggtattgctctactggcatttatttttggcgtgttttttggcaatatatcctcaccagcttgttctgaagaccatgaggaggtcatttct
aaccaaacatcagttatagattctcaaaaaacagaaatagaaactttaaatagcaaattgtctgatgctgaaccatggttcaaaatg
aaagacgacgaaaagaaagctattgaagctgaaaatcaacgtaaagctgaagaagctaaaaaagctgaagaagctaaaaagg
ctgaagaacaacgtaaaaaagaagaagaggagaagaaaggatatgatactggtattacttatgaccaattagctagaacacctg
atgattataagtacaaaaaggtaaaatttgaaggtaaggttattcaagttattgaagatggtgatgaggtgcaaataagattagctgt
gtctggaaattatgataaggttgtactatgtagttataaaaaatcaataactccttcaagagtattagaggatgattacataactataa
gaggtataagtgctggaactataacttatgaatcaactatgggtggaaatataactataccaggg

New3: aaaaaaatgagggtaacgaatttagtaatgaaggaaaaggaagctctaagaaaataattaaattctttaagagcacaaa
gggtattgctctactggcatttatttgggcgtgttttttggcaatatatcctcaccagcttgttctgaagaacatgaggaggtcatttct
aatcaaacatcagttatagattctcaaaaaacagaaatagaaactttaaatagcaaattgtctgatgctgaaccatggttcaaaatga
aagacgacgaaaagaaagctattgaagctgaaaatcaacgtaaagctgaagaagctaaaaaagctgaagaagctaaaaaggc
tgaagaacaacgtaaaaaagaagaagaggagaagaaaggatatgatactggtattacttatgaccaattagctagaacacctgat
gattataagtacaaaaaggtaaaatttgaaggtaaggttattcaagttattgaagatggtgatgaggtgcaaataagattagctgtgt
ctggaaattatgataaggttgtactatgtagttataaaaaatcaataactccttcaagagtattagaggatgattacataactataaga
ggtataagtgctggaactataacttatgaatcaactatgggtggaaatataactataccaggg

New4: aaaaaaatgagggtaacgaatttagtaatgaaggaaaaggaagctctaagaaaataattaaattctttaagagcacaaa
gggtattgctctactggcatttatttaggcgtgttttttggcaatatatcctcaccagcttgttctgaagaccatgaggaggtcatttct
aaccaaacatcagttatagattctcaaaaaacagaaatagaaactttaaatagcaaattgtctgatgctgaaccatggttcaaaatg
aaagacgacgaaaagaaagctattgaagctgaaaatcaacgtaaagctgaagaagctaaaaaagctgaagaagctaaaaaagg
ctgaagaacaacgcaaaaaagaagaagaggagaagaaaggatatgatactggtattacttatgaccaattagctagaacacctg
atgattataagtacaaaaaggtaaaatttgaaggtaaggttattcaagttattgaagatggtgatgaggtgcaaataagattagctgt
gtctggaaattatgataaggttgtactatgtagttataaaaaatcaataactccttcaagagtattagaggatgattacataactataa
gaggtataagtgctggaactataacttatgaatcaactatgggtggaaatataactataccaggg

New5: aaaaaaatgagggtaacgaatttagtaatgaaggaaaaggaagctctaagaaaataattaaattctttaagagcacaaa
gggtattgctctactggcatttatttgggcgtgttttttggcaatatatcctcaccagcttgttctgaagaccatgaggaggtcatttct
aatcaaacatcagttatagattctcaaaaaacagaaatagaaactttaaatagcaaattgtctgatgctgaaccatggttcaaaatga
aagacgacgaaaagaaagctattgaagctgaaaatcaacgtaaagctgaagaagctaaaaaggctgaagaacaacgtaaaaa
agaagaagaagaaaagaaaggatatgatactggtattacttatgaccaattagctagaacacctgatgattataagtacaaaaag
gtaaaatttgaaggtaaggttattcaagttattgaagatggtgatgaggtgcaaataagattagctgtgtctggaaattatgataagg
tcgtactatgtagttataaaaaatcaataactccttcaagagtgttagaggatgattacataactataagaggtataagtgctggaact
ataacttatgaatcagctatgggtggaaaaataaccataccaggg

New6: aaaaaaatgagggtaacgaatttagtaatgaaggaaaaggaagctctaagaaaataattaaattctttaagagcacaaa
gggtattgctctactggcatttattttgggcgtgttttttggcaatatatcctcaccagcttgttctgaagaccatgaggaggtcatttct
aatcaaacatcagttatagattctcaaaaaacagaaatagaaactttaaatagcaaattgtctgatgctgaaccatggttcaaaatga
aagacgacgaaaagaaagctattgaagctgaaaatcaacgtaaagctgaagaagctaaaaaggctgaagaacaacgtaaaaa
agaagaagaggagaagaaaggatatgatactggtattacttatgaccaattagctagaacacctgatgattataagtacaaaaag
gtaaaatttgaaggtaaggttattcaagttattgaagatggtgatgaggtgcaaataagattagctgtgtctggaaattatgataagg
ttgtactatgtagttataaaaaatcaataactccttcaagagtattagaggatgattacataactataagaggtataagtgctggaact
ataacttatgaatcaactatgggtggaaatataactataccaggg

New7: aaaaaaatgagggtaacgaatttagtaatgaaagaaaaggaagctctaagaaaataattaaattctttaagagcacaaa
ggatattgctctactggcatttatttggtgtgttttttggcaatatatcctcaccagcttgttctgaagaccatgaggaggtcatttctaa
tcaaacatcagttatagattctcaaaaaactagaaatagaaactttaaatagcaaattgtctgatgctgaaccatggttcaaaatgaa
agacgacgaaaagaaagctattgaagctgaaaatcaacgtaaagctgaagaagctaaaaaggctgaagaacaacgtaaaaaa
gaagaagaagagaagaaaggatatgatactggtattacttatgaccaattagctagaacacctgatgattataagtacaaaaaggt
aaaatttgaaggtaaggttattcaagttattgaagatggtgatgaggtgcaaataagattagctgtgtctggaaattatgataaggtc
gtactatgtagttataaaaaatcaataactccttcaagagtgttagaggatgattacataactataagaggtataagtgctggaactat
aacttatgaatcaactatgggtggaaaaataaccataccaggg

New8: aaaaaaatgagggtaacgaatttagtaatgaaggaaaaggaagctctaagaaaataattaaattctttaagagcacaaa
gggtattgctctactggcatttatttaggcgtgttttttggcaatatatcctcaccagcttgttctgaagaccatgaggaggtcatttct
aaccaaacatcagttatagattctcaaaaaacagaaatagaaactttaaatagcaaattgtctgatgctgaaccatggttcaaaatg
aaagacgacgaaaagaaagctattgaagctgaaaatcaacgtaaagctgaagaagctaaaaaggctgaagaacaacgcaaaa
aagaagaagaggagaagaaaggatatgatactggtattacttatgaccaattagctagaacacctgatgattataagtacaaaaa
ggtaaaatttgaaggtaaggttattcaagttattgaagatggtgatgaggtgcaaataagattagctgtgtctggaaattatgataag
gttgtactatgtagttataaaaaatcaataactccttcaagagtattagaggatgattacataactataagaggtataagtgctggaac
tataacttatgaatcaactatgggtggaaatataactataccaggg

New9: aaaaaaatgagggtaacgaatttagtaatgaaggaaaaggaagctctaagaaaataattaaattctttaagagcacaaa
ggatattgctctactggcatttatttttggtgtgttttttggcaatatatcctcaccagcttgttctgaagaccatgaggaggtcatttcta
atcaaacataagttatagattctcaaaaaacagaaatagaaactttaaatagcaaattgtctgatgctgaaccatggttcaaaatgaa
agatgacgaaaagaaagctattgaagctgaagaagctaagaaagctgaagaacaacgtaaaaaagaagaagaagagaagaa
aggatatgatactggtattacttatgaccaattagctagaacatctgatgattataagtacaaaaaggtaaaatttgaaggtaaggtt
attcaagttattgaagatggtgatgaggtgcaaataagattagctgtgtctggaaattatgatgaggtcgtactatgtagttataaaaa
atcaataactccttcaagagtgttagaggatgattacataactataagaggtataagtgctggaactataacttatgaatcagctatg
ggtggaaaaataactataccaggg

New10: aaaaaaatgagggtaacgaatttagtaatgaaggaaaaggaagctctaagaaaataattaaattctttaagagcaca
aagggtattgctctactggtatttatttttaggcgtgttttttggcaatatatcctcaccagcttgttctgaagaccatgagaaggtcattt
ctaatcaaacatcagttatagattctcaaaaaacagaaatagaaactttaaatagcaaattgtctgatgctgaaccatggttcaaaat
gaaagacgacgaaaagaaagctattgaagctgaaaatcaacgtaaagctgaagaagctaaaaaggctgaagaacaacgtaaa
aaagaagaagaagagaagaaaggatatgatactggtattacttatgaccaattagctagaacacctgatgattataagtacaaaaa
ggtaaaatttgaaggtaaggttattcaagttattgaagatggtgatgaggtgcaaataagattagctgtgtctggaaattatgataag
gttgtactatgtagttataaaaaatcaataactccttcaagagtattagaggatgattacataactataagaggtataagtgctggaac
tataacttatgaatcaactatgggtggaaatataactataccaggg

Appendix 8. Summary table (spanning seven pages) of PFGE-*SmaI* pattern, NAP type, predicted sporulation capacity, MLST sequence type and *tcdC* allele results for each of the 309 strains analyzed in this study.

Strain[a]	PFGE-*SmaI* Pattern[b]	NAP Type[b,c]	Predicted Sporulation Capacity	MLST Sequence Type	*tcdC* Allele
SB04	0233		Good sporulator	36	sc9
SB08	0003	NAP2	Won't sporulate	3	sc3
SB10	0003	NAP2	Won't sporulate	3	sc3
SB13	0251	NAP1	Good sporulator	44	sc1
SB16	0142		Good sporulator	36	sc9
SB26	0042		Good sporulator	1015	sc9
SB34	0033	NAP4	Good sporulator	1	sc9
SB35	0122		Good sporulator	36	sc9
SB40	0003	NAP2	Won't sporulate	3	sc3
SB43	0122		Good sporulator	36	sc9
SB46	0122		Good sporulator	36	sc9
SB48	0012		Good sporulator	45	sc13
SB53	0003	NAP2	Won't sporulate	3	sc3
SB65	0003	NAP2	Won't sporulate	3	sc3
SB66	0012		Good sporulator	44	sc13
SB89	0018	NAP1	Good sporulator	44	sc1
SB118	0033	NAP4	Good sporulator	1	sc9
SB128	0006	NAP1	Good sporulator	44	sc1
SB129	0012		Good sporulator	44	sc13
SB146	0012		Good sporulator	1008	sc13
SB171	0274		Good sporulator	6	sc9
SB186	0003	NAP2	Won't sporulate	3	sc3
SB200	0003	NAP2	Won't sporulate	3	sc3
SB201	0349		Good sporulator	2	sc7
SB246	0139		Good sporulator	36	sc9
SB255	0003	NAP2	Won't sporulate	3	sc3
SB279	0139		Good sporulator	36	sc9
SB284	0038	NAP6	Good sporulator	9	sc9
SB295	0003	NAP2	Won't sporulate	3	sc3
SB321	N.D.	N.D.	Won't sporulate	3	sc3
SB330	0012		Good sporulator	1008	sc13
SB331	0026	NAP4	Won't sporulate	1	sc9
SB368	0122		Good sporulator	36	sc9
SB369	0145		Good sporulator	1027	sc2
SB370	0003	NAP2	Strong sporulator	3	sc3
SB372	0331		Good sporulator	36	sc9
SB373	0003	NAP2	Won't sporulate	3	FAIL
SB374	0066	NAP12	Good sporulator	4	New8
SB376	0003	NAP2	Won't sporulate	3	sc3
SB378	0356	NAP2	Good sporulator	3	sc3
SB381	0251	NAP1	Good sporulator	44	sc1
SB382	0038	NAP6	Good sporulator	9	sc9
SB383	0003	NAP2	Won't sporulate	3	sc3
SB387	0139		Good sporulator	36	sc9

SB391	0003	NAP2	Won't sporulate	3	sc3
SB392	0331		Good sporulator	36	sc9
SB393	0003	NAP2	Won't sporulate	3	sc3
SB394	0003	NAP2	Won't sporulate	3	sc3
SB396	0331		Good sporulator	36	sc9
SB398	0003	NAP2	Won't sporulate	3	sc3
SB400	0366		Good sporulator	4	New11
SB401	0003	NAP2	Won't sporulate	3	sc3
SB403	0003	NAP2	Won't sporulate	3	sc3
SB404	0233		Good sporulator	36	sc9
SB405	0331		Good sporulator	36	sc9
SB414	0003	NAP2	Won't sporulate	3	sc3
SB419	0123		Good sporulator	FAIL	New5
SB420	0003	NAP2	Won't sporulate	3	sc3
SB427	0080	NAP7	Won't sporulate	1005	A1
SB430	0003	NAP2	Won't sporulate	3	sc3
SB433	0003	NAP2	Won't sporulate	3	sc3
SB440	0066	NAP12	Good sporulator	4	New8
SB443	0003	NAP2	Won't sporulate	3	sc3
SB447	0003	NAP2	Won't sporulate	3	sc3
SB448	0003	NAP2	Won't sporulate	3	sc3
SB451	0003	NAP2	Strong sporulator	3	sc3
SB454	0003	NAP2	Won't sporulate	3	sc3
SB457	0020	NAP6	Good sporulator	9	New4
SB461	0003	NAP2	Won't sporulate	3	sc3
SB462	0066	NAP12	Good sporulator	4	New8
SB465	0003	NAP2	Won't sporulate	3	sc3
SB466	0033	NAP4	Good sporulator	1	New4
SB468	0033	NAP4	Good sporulator	1	New4
DL01	0003	NAP2	Won't sporulate	3	sc3
G03	0194	NAP6	Strong sporulator	9	New4
G020	0003	NAP2	Won't sporulate	3	sc3
G026	0003	NAP2	Won't sporulate	3	sc3
G028	0003	NAP2	Won't sporulate	3	sc3
G31	0003	NAP2	Won't sporulate	3	sc3
G034	0313		Good sporulator	1015	New4
G035	0003	NAP2	Won't sporulate	3	sc3
G36	0003	NAP2	Won't sporulate	3	sc3
G037	0003	NAP2	Won't sporulate	3	sc3
G42	0088		Good sporulator	1014	New9
G045	0003	NAP2	Won't sporulate	3	sc3
G046	0003	NAP2	Won't sporulate	3	sc3
G123	0251	NAP1	Won't sporulate	44	sc1
G137	0003	NAP2	Won't sporulate	3	sc3
G155	0122		Good sporulator	36	sc9
G156	0274		Good sporulator	6	sc9
G217	0003	NAP2	Won't sporulate	3	sc3
G236	0347		Good sporulator	40	sc15
G260	0337		Good sporulator	36	sc9
G261	0041		Good sporulator	51	sc5

G302	0049	NAP11	Good sporulator	1013	sc14
G314	0019	NAP10	Good sporulator	13	sc9
G356	0207	NAP1	Good sporulator	44	sc1
G360	0115		Good sporulator	35	sc9
G367	0003	NAP2	Won't sporulate	3	sc3
G397	0024	NAP6	Good sporulator	9	sc9
G406	0003	NAP2	Won't sporulate	3	sc3
G415	0003	NAP2	Won't sporulate	3	sc3
G421	0116		Good sporulator	51	sc5
G422	0003	NAP2	Won't sporulate	3	sc3
G426	0003	NAP2	Won't sporulate	3	sc3
G429	0003	NAP2	Won't sporulate	3	sc3
G434	0003	NAP2	Won't sporulate	3	sc3
G445	0003	NAP2	Won't sporulate	3	sc3
G452	0003	NAP2	Won't sporulate	3	sc3
G458	0360		Good sporulator	19	sc15
G471	0024	NAP6	Good sporulator	9	sc9
A362	0003	NAP2	Won't sporulate	3	sc3
C06	0009		Good sporulator	47	New4
C008	0072	NAP2	Won't sporulate	3	sc3
C013	0368	NAP2	Won't sporulate	3	sc3
C24	0054		Good sporulator	36	sc9
C039	0003	NAP2	Won't sporulate	3	sc3
C042	0033	NAP4	Won't sporulate	1	sc9
C43	0122		Won't sporulate	3	sc3
C47	0003	NAP2	Won't sporulate	3	sc3
C050	0003	NAP2	Won't sporulate	3	sc3
C67	0003	NAP2	Won't sporulate	3	sc3
C68	0003	NAP2	Won't sporulate	3	sc3
C72	0061		Good sporulator	45	sc13
C91	0018	NAP1	Good sporulator	44	sc1
C122	0003	NAP2	Won't sporulate	3	sc3
C176	0012		Won't sporulate	45	sc13
C189	0003	NAP2	Won't sporulate	3	sc3
C283	0033	NAP4	Won't sporulate	1	sc9
C292	0251	NAP1	Good sporulator	44	sc1
C333	0342		Won't sporulate	1022	FAIL
C361	N.D.	N.D.	Poor Sporulator	41	A1
C379	0359	NAP4	Good sporulator	39	sc15
C386	0337		Good sporulator	36	sc9
C413	0003	NAP2	Won't sporulate	3	sc3
C424	0003	NAP2	Good sporulator	1023	New4
C431	0003	NAP2	Won't sporulate	3	sc3
C437	0003	NAP2	Won't sporulate	3	sc3
C439	0003	NAP2	Won't sporulate	3	sc3
V02	0233		Good sporulator	36	sc9
V05	0233		Good sporulator	36	sc9
V007	0122		Won't sporulate	36	sc9
V009	0373		Won't sporulate	1022	mb1
V18	0122		Good sporulator	36	sc9

V019	N.D.	N.D.	Good sporulator	36	sc9
V021	0122		Good sporulator	36	sc9
V032	0175	NAP11	Good sporulator	1013	New3
V33	0331		Good sporulator	36	sc9
V44	0001	NAP1	Good sporulator	44	sc1
V048	0014	NAP1	Won't sporulate	1022	mb1
V054	0122		Good sporulator	36	sc9
V055	0328		Good sporulator	46	New9
V119	0233		Good sporulator	36	sc9
V190	0108		Good sporulator	36	sc9
V214	0019	NAP10	Good sporulator	13	sc9
V227	0340		Good sporulator	1013	New6
V228	0012		Good sporulator	44	sc13
V245	0153	NAP7	Good sporulator	41	A1
V305	0214	NAP4	Good sporulator	1	New4
V311	0122		Good sporulator	36	sc9
V366	0122		Good sporulator	36	sc9
V371	0122		Good sporulator	36	sc9
V402	0122		Good sporulator	36	sc9
V407	0202		Won't sporulate	1022	mb1
V432	0249	NAP1	Won't sporulate	1022	mb1
V449	0020	NAP6	Good sporulator	9	New4
V450	0033	NAP4	Won't sporulate	1	sc9
V453	0122		Good sporulator	36	sc9
V475	0274		Good sporulator	6	New4
M033	0361		Good sporulator	36	sc9
M58	0037		Good sporulator	1025	sc9
M109	0018	NAP1	Good sporulator	44	sc1
M113	0037		Good sporulator	1025	sc9
M163	0328		Won't sporulate	1022	mb1
M178	0088		Good sporulator	FAIL	sc16
M238	0327		Good sporulator	1003	New9
M256	0214	NAP4	Won't sporulate	1	sc9
M357	0019	NAP10	Good sporulator	13	sc9
M418	0358		Good sporulator	36	sc9
M467	0361		Good sporulator	36	sc9
H01	0139		Good sporulator	36	sc9
H02	0122		Good sporulator	36	sc9
H03	0225		Won't sporulate	9	sc3
H04	0024	NAP6	Good sporulator	1007	sc9
H05	0194	NAP6	Good sporulator	9	sc9
H06	0024	NAP6	Good sporulator	1007	sc9
H07	0003	NAP2	Won't sporulate	3	sc3
H08	0003	NAP2	Won't sporulate	3	sc3
H10	0014	NAP1	Good sporulator	1019	sc1
H13	0326		Good sporulator	1017	sc7
H14	0091		Good sporulator	1002	sc9
H24	0332	NAP2	Won't sporulate	3	sc3
H26	0330		Good sporulator	36	sc9
H31	0334		Good sporulator	30	sc9

H36	0003	NAP2	Won't sporulate	3	sc3
H45	0033	NAP4	Won't sporulate	1	sc9
H46	0111	NAP1	Good sporulator	46	New9
H49	0338		Good sporulator	1018	New9
H52	0100		Good sporulator	36	sc9
H55	0336		Good sporulator	1006	sc9
H59	0012		Good sporulator	1008	mb4
H65	0214	NAP4	Good sporulator	1024	sc9
H66	0202		Won't sporulate	1022	mb1
H84	0361		Good sporulator	36	sc9
H85	0361		Good sporulator	36	sc9
H95	0003	NAP2	Won't sporulate	3	sc3
SO12	0135	NAP4	Good sporulator	1	New4
SO16	0332	NAP2	Good sporulator	36	sc9
SO18	0333	NAP2	Won't sporulate	3	sc3
SO22	0024	NAP6	Good sporulator	9	New4
SO29	0003	NAP2	Won't sporulate	3	sc3
SO30	0018	NAP1	Good sporulator	44	sc1
SO35	0139		Good sporulator	36	sc9
SO39	0233		Good sporulator	36	sc9
SO48	0341	NAP11	Good sporulator	1013	mb6
SO57	0348		Good sporulator	1026	sc9
SO62	0139		Good sporulator	36	sc9
SO79	0203		Good sporulator	46	sc16
SO81	0361		Good sporulator	36	sc9
SO82	0003	NAP2	Won't sporulate	3	sc3
SO90	0361		Good sporulator	36	sc9
W1	0016		Good sporulator	1019	sc1
W2	0003	NAP2	Won't sporulate	3	sc3
W3	0329		Good sporulator	1001	mb8
W5	0198		Won't sporulate	48	New4
W6	0332	NAP2	Won't sporulate	3	sc3
W7	0208	NAP12	Good sporulator	4	New11
W8	0167		Good sporulator	44	sc13
W12	0054		Good sporulator	36	sc9
W13	0335		Won't sporulate	1022	mb1
W17	0003	NAP2	Won't sporulate	3	sc3
W18	0012		Strong sporulator	1008	mb4
W20	0110		Good sporulator	36	sc9
W23	0251	NAP1	Good sporulator	44	sc1
W28	0047		Good sporulator	45	sc13
W29	0343	NAP6	Good sporulator	9	New4
W30	0090		Good sporulator	50	New2
W31	0038	NAP6	Good sporulator	9	New4
W32	0332	NAP2	Won't sporulate	3	sc3
W33	0346	NAP10	Good sporulator	13	sc9
W34	0012		Good sporulator	45	sc13
W36	332	NAP2	FAIL	3	FAIL
W44	0042		Good sporulator	1015	FAIL
W47	0077		Won't sporulate	51	sc5

W50	0342		Won't sporulate	1022	mb1
W55	0033	NAP4	Won't sporulate	1	sc9
W60	0067		Good sporulator	4	New4
W61	0077		Won't sporulate	51	sc5
W63	0003	NAP2	Won't sporulate	3	sc3
W64	0006	NAP1	Good sporulator	44	sc1
W65	0024	NAP6	Good sporulator	1007	New4
W71	0050	NAP10	Good sporulator	13	sc9
W76	0363		Good sporulator	1017	New10
W97	0148		Good sporulator	51	sc5
W111	0029		Won't sporulate	9	sc4
W122	0251	NAP1	Good sporulator	44	sc1
W125	0371		Good sporulator	36	sc9
W127	0313		Good sporulator	1015	New4
W128	0006	NAP1	Good sporulator	44	sc1
W131	0054		Good sporulator	36	sc9
W132	0085	NAP4	Good sporulator	1	New4
W133	0011	NAP3	Good sporulator	6	New4
W134	0376	NAP2	Won't sporulate	3	sc3
W135	0012		Good sporulator	1008	mb4
W145	0375		Good sporulator	1014	sc13
W148	0003	NAP2	Won't sporulate	3	sc3
W151	0012		Good sporulator	1008	mb4
W162	0140		Good sporulator	47	New8
W174	0079		Good sporulator	1001	New7
W176	0374		Good sporulator	1012	mb8
CPL1	0350	NAP1	Good sporulator	44	New7
CPL4	0352		Good sporulator	1016	New1
CPL9	0050	NAP10	Good sporulator	13	sc9
CPL15	0024	NAP6	Won't sporulate	3	sc3
CPL19	0045		Good sporulator	1001	mb8
CPL21	0012		Good sporulator	1009	sc13
CPL31	0003	NAP2	Won't sporulate	3	sc3
CPL34	0014	NAP1	Good sporulator	1022	mb1
CPL36	0362	NAP9	Won't sporulate	1020	New10
CPL38	0345		Strong sporulator	1001	mb8
CPL43	0012		Good sporulator	45	sc13
CPL46	0148		Good sporulator	51	sc5
CPL47	0249	NAP1	Won't sporulate	1022	mb1
CPL48	0364		Good sporulator	30	New4
CPL50	0053	NAP2	Good sporulator	3	sc3
CPL51	0365	NAP1	Good sporulator	44	sc1
CPL56	0367		Good sporulator	4	New4
CPL60	0313		Good sporulator	1015	New4
CPL66	0369	NAP6	Good sporulator	9	New4
CPL70	0003	NAP2	Won't sporulate	3	sc3
CPL72	0153	NAP7	Won't sporulate	1021	A1
CPL78	0013		Good sporulator	1008	mb4
CPL80	0033	NAP4	Won't sporulate	1	sc9
CPL83	0140		Good sporulator	47	sc9

CPL88	0100		Good sporulator	36	sc9
CPL94	0063		Good sporulator	1022	mb1
CPL96	0254		Good sporulator	1010	New3
CPL99	0370		Good sporulator	1004	New4
CPL101	0023	NAP4	Won't sporulate	1	sc9
CPL103	0140		Good sporulator	FAIL	New4
CPL104	0003	NAP2	Won't sporulate	3	sc3
CPL105	0300	NAP1	Won't sporulate	1011	sc1
ATCC700057	N.D.	N.D.	Good sporulator	11	FAIL
ATCC43594	N.D.	N.D.	Won't sporulate	9	sc3
ATCC43255	N.D.	N.D.	Poor Sporulator	47	New4
ATCC43598	N.D.	N.D.	Good sporulator	2	sc7
57A	N.D.	N.D.	Good sporulator	44	sc1
83	N.D.	N.D.	Good sporulator	44	sc1
765	N.D.	N.D.	Good sporulator	36	FAIL

[a] Strain's are named according to their hospital source: "SB-" for St. Boniface, "G-" for Grace, "C-" for Concordia, "V-" for Victoria, "M-" for Misericordia, "H-" for Health Sciences Centre, "SO-" for Seven Oaks, "W-" for Westman, "CPL-" for Cadham Provincial Lab, and "ATCC-" strains with 57A, 83, and 765 were the control strains.

[b] N.D. denotes not determined.

[c] Empty cells in the NAP type column designate strains that were NAP untypable.