Emergence and Molecular Characterization of NAP4 *Clostridium difficile* Healthcare-Associated Infections in Canada

By

Tyler Kosowan

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Medical Microbiology

University of Manitoba

Winnipeg

Abstract

Clostridium difficile is an important nosocomial pathogen and the leading cause of antibiotic-associated diarrhea. In the early 2000's, major outbreaks of *C. difficile* occurred in the Canada and the United States, both caused by the same strain, the North American Pulsed-Field type 1 (NAP1). Recently a steady rise in prevalence of a NAP4 strain was observed. This study utilized phenotypic and genotypic methods, including whole genome sequencing, to fully characterize and better understand the emergence of NAP4 in Canada.

Through WGS it was revealed that an emerging NAP4 PFGE type (PFGE type 0033) has acquired a novel phage (phiCD505), which harbors transcriptional regulators and is inserted in close proximity to a GMP synthase gene. The role of these phage encoded transcriptional regulators on *C. difficile* virulence factors and/or disruption of normal expression of GMP synthase remain to be elucidated, but might be contributing to the emergence of this PFGE type.

Acknowledgments

I would like to express my gratitude to my supervisors, Dr. Michael Mulvey and Dr. George Golding, for accepting me into the Antimicrobial Resistance and Nosocomial Infections (ARNI) lab so that I may acquire my Master's degree. Their attitude towards scientific exploration encouraged me to work harder and to aspire for further discovery. I would also like to thank my committee members, Dr. George Zhanel, and Dr. Teresa de Kievit, for their guidance and professional expertise on this project. I would like to express thanks to the members of the ARNI lab for their support and the enjoyable work environment they've created. As well, my thanks to the bioinformatics group for their assistance, especially Adrian Zetner for his expertise and patience. Finally, I thank my parents for helping me get through the hard times during my project.

Table of Contents

List of Abbreviations

BAK	Brucella Agar with hemin and vitamin K
BHI	Brain Heart Infusion broth
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CA-CDI	Community-associated *Clostridium difficile* Infection
CDC	Center for Disease Control
CDI	*Clostridium difficile* Infection
CDMN	*Clostridium difficile* Moxalactam Norfloxacin selective supplement
CLB	Cell Lysis Buffer
CLSI	Clinical and Laboratory Standards Institute
CNISP	Canadian Nosocomial Infection Surveillance Program
CRE	Carbapenem-resistant *Enterobacteriaceae*
DARPIC	Death Attribution Rules for Patients Infected by *Clostridium difficile*
DNA	Deoxyribonucleic acid
ECDC	European Centre for Disease Prevention and Control
EDTA	Ethylenediaminetetraacetic acid
EF-TU	Elongation Factor TU
eMLVA	Extended Multilocus Variable Number Tandem Repeat Analysis
EPA	Environmental Protection Agency
FAB	Fastidious Anaerobic Broth
FAM	Carboxyfluorescein primer tag
FMT	Fecal Microbiota Transplant

FQR	Fluoroquinolone Resistant Group
GMP	Guanosine Monophosphate
HA-CDI	Healthcare-associated *Clostridium difficile* Infection
ICE	Integrative and Conjugative Element
ICU	Intensive Care Unit
kb	Kilobase
MIC	Minimum Inhibitory Concentration
MLST	Multilocus Sequence Typing
MLVA	Multilocus Variable Number Tandem Repeat Analysis
MPC	MasterPure-Complete
MRSA	Methicillin-resistant *Staphylococcus aureus*
NAP	North American Pulsed-field type
NML	National Microbiology Laboratory
ns	No Standard
OD	Optical Density
PaLoc	Pathogenicity Locus
PBS	Phosphate-buffered Saline
PCR	Polymerase Chain Reaction
PFGE	Pulsed-field Gel Electrophoresis
PHAC	Public Health Agency of Canada
PHAST	PHAge Search Tool
PPE	Personal Protective Equipment
PPI	Proton Pump Inhibitor

REA Restriction Endonuclease Assay

RFLP Restriction Fragment Length Polymorphism

RNA Ribonucleic acid

rpm Rotations per minute

SDS Sodium Dodecyl Sulfate

SMRT Single Molecule Real Time

SNP Single Nucleotide Polymorphism

SNV Single Nucleotide Variant

ST Sequence type

STRD Summed Tandem Repeat Difference

TBE Tris-Borate-EDTA

TE Tris-EDTA

TEM Transmission Electron Microscopy

TSA Trypticase Soy Agar

UDP-MurNac Undecaprenyl-N-acetylmuramic acid

UDP-PP-GlcNac Undecaprenyl-pyrophosphate-N-acetylglucosamine

UPGMA Unweighted Pair Group Method with Arithmetic Mean

UV Ultraviolet

V/cm Volts per centimeter

VNTR Variable Number Tandem Repeat

WGS Whole genome sequencing

WTA Cell Wall Teichoic Acid

1. Introduction

Clostridium difficile is a Gram-positive, obligate anaerobic bacillus that is acquired from the environment or through the fecal-oral route [114]. Once the bacterium colonizes the human gut it may produce toxins A and B which are a major cause of subsequent symptoms related to a *C. difficile* infection (CDI) [38]. Symptoms of CDI can range in severity depending on the host as well as the strain causing infection. A mild case of CDI commonly presents as diarrhea. More severe cases of CDI can involve pseudomembranous colitis, abdominal distension, or end organ failure [114, 138].

C. difficile is a pathogen commonly associated with healthcare environments. A link between the use of antimicrobials and the potential development of CDI has been well documented. This is known to be due to the disruptive effect antimicrobials have on the natural microflora of the gut. The intestinal dysbiosis that takes place facilitates the colonization of the gut by *C. difficile* and its potential progression to CDI. Diarrheal symptoms that arise from this chain of events are known as antibiotic-associated diarrhea [24]. In 2001 CDI was shown to cause between 42% and 63% of all cases of antibiotic-associated diarrhea. This rate was somewhat variable between different hospitals and antibiotic regimen [148].

Recent outbreaks of *C. difficile* in hospital settings have been related to the emergence of a hypervirulent *C. difficile* strain. This strain, designated North American Pulsed-field Type 1 (NAP1), has presented high recurrence rates and significant mortality compared to other NAP strains in North America and Europe [71]. The first recorded NAP1 outbreak in Canada was in the province of Quebec in 2002 [103]. Soon after national surveillance of *C. difficile* was established through the Canadian Nosocomial Infection Surveillance Program (CNISP), which has revealed NAP1 as the most prevalent strain type of *C. difficile* across Canada. (unpublished

data from CNISP). Over time, NAP4 has become the second most common NAP type in Canada, but not much research has been done on this NAP type to date. With changes in *C. difficile* epidemiology being observed in Canada, and arguably other countries around the world, the need for continued surveillance and study is warranted.

1.1. Physical Characteristics

Clostridium difficile is defined in Bergey's Manual of Systemic Bacteriology [114] as a spore-forming, Gram-positive, obligate anaerobic bacillus. The bacteria can present a motile phenotype by the use of peritrichous flagella, but motility is not always observed. The spores *C. difficile* forms are oval in shape and form at a subterminal position in the cell. Optimal growth temperature is 37^0C although growth is still observed at temperatures as low as 25^0C or as high as 45^0C. Colonies formed on solid, blood agar plates are 2–5mm in circumference, circular, and can be flat or slightly convex. The colonies may form rhizoidal extensions, are opaque, gray to white in colour, and have a matt or glossy surface. All *C. difficile* colonies grown on blood agar plates fluoresce pale green under long-wave ultraviolet (UV) light. Common sources of *C. difficile* are, soil, sand, and marine sediment; feces from animals such as dogs, birds, and camels; the human genital tract, human feces; and most notably the hospital environment [114].

1.2. History

Clostridium difficile was first isolated in 1935. The study in which it was found involved characterizing the microbial changes of the intestinal flora in new-born infants over ten days by analysing infant stool samples [51]. *C. difficile* was examined using microscopy to observe cell morphology after a Gram stain as well as a Bailey's stain. The bacterium was originally named *Bacillus difficilus* because of how difficult it was to isolate and study [51].

Later, in 1978, it was discovered that antibiotic-associated pseudomembranous colitis was caused by toxins being produced by *C. difficile* through studies of cytotoxicity in tissue culture and enterocolitis in the hamster model [8].

In the year 2000 the first major outbreak of NAP1 *C. difficile* infection in North America was observed at the University of Pittsburgh Medical Center. In late 2002 NAP1 presented the first major outbreak in Canada in the province of Quebec [103]. This led to the creation of a mandatory, province-wide *Clostridium difficile* infection (CDI) surveillance program in 2004 [19]. In Europe, the first recorded outbreak of *C. difficile* was detected in 2003 at the Stoke Mandeville Hospital in the United Kingdom [70].

1.3. Burden on Healthcare

The common association of *C. difficile* with hospital environments is due to the fact that the bacterium is able to flourish in the human gut after the host's natural gut microflora is disrupted by antibiotics. The Center for Disease Control (CDC) has recently surmised that *C. difficile* causes around 250,000 infections, 14,000 deaths, and led to 1 billion dollars in excess medical costs annually. From this, the CDC has flagged *C. difficile* as an urgent-level threat to public health, the highest hazard level they present, alongside carbapenem-resistant *Enterobacteriaceae* (CRE) and drug-resistant *Neisseria gonorrhoeae* [138]. A study by Lessa et al. (2015) found that in the US *C. difficile* caused about 453,000 infections in the study year and was implicated as the cause of death in around 15,000 cases [81]. The cost for acute-care facilities alone has also been updated to an estimated 4.8 billion dollars annually [29].

1.4. Prevention and Management

The Public Health Agency of Canada (PHAC) has produced a guide for handling the presence or threat of CDI in acute care facilities [109]. It outlines that patients with suspected or confirmed cases of CDI must be placed in a single room with a dedicated toilet or commode. Proper hand hygiene techniques should be frequently performed after contact with the patient or their room. Personal protective equipment (PPE) such as gloves and gowns should be worn by all staff and visitors entering the patient's room. Disinfecting reusable patient-care equipment should be done with a chlorine-containing cleaning agent that is comprised of at least 1,000 parts per million [109]. Other attempts to reduce *C. difficile* contamination of hospital environments that have been shown to decrease the incidence of CDI include a hypochlorite-based solution and an ammonium compound cleaning agent [93, 143]. The use of hydrogen peroxide in a vapor form has also been investigated for use of decontaminating hospital rooms with promising results [10]. The CDC recommends a disinfectant registered by the Environmental Protection Agency (EPA) that has a *C. difficile*-sporicidal label claim. Evidence from a study by Rutala et al. (2008) supports the use of chlorine-containing cleaning agents with a minimum of 5,000 p.p.m. of available chlorine with an exposure time of at least 10 minutes [120].

The PHAC guidelines strongly support the use of antibiotic stewardship programs [109]. Antibiotic stewardship is a program with two main goals. The first is to tailor the antibiotic regimen to the patient to ensure a positive outcome from treating a bacterial infection. The second is to educate physicians and patients to minimize the redundant use of certain antibiotics and introduce more effective treatment plans. This goal is, in part, meant to decrease or slow the appearance of antibiotic resistant bacteria in both healthcare and community settings.

1.5. Risk of Infection

There are many risks associated with a person becoming infected with *C. difficile*. The link between taking antimicrobials and developing CDI [24] presents a risk to many people in hospital settings. For instance, it has previously been reported that the risk of a patient acquiring CDI when taking antimicrobials is 7-10 fold greater than a person who has taken no antimicrobials [55].

Risks have been related to certain outcomes of CDI as well as the number and types of comorbidities in the patient. Several outcomes of CDI have shown an association with certain risk factors described in a review of *C. difficile* related studies from 1978 to 2013 by Chakra et al. [15]. The outcomes with patient-related risk factors described in the review are complicated CDI, 30-day all-cause mortality, and CDI recurrence. Complicated CDI is defined in the review as a case involving any complication, shock, admission to an intensive care unit (ICU), fulminant colitis, and/or death with a composite outcome. Patient risk factors for complicated CDI include age over 65 [15, 102], renal failure [15], leucocytosis [15], and underlying co-morbidities such as infection with methicillin-resistant *Staphylococcus aureus* (MRSA) or cancer [28]. 30-day all-cause mortality is recorded if the patient dies up to 30 days after diagnosis with CDI, but *C. difficile* was not the direct cause of death. Risk factors associated with 30-day all-cause mortality were associated with underlying co-morbidities, leucocytosis, older age, infection with ribotype 027, decreased serum albumin, and increased levels of serum creatinine and/or urea [15]. Recurrence has been defined as the re-emergence of symptoms of CDI at least 10 days after resolution of the primary infection [6]. The common risks for CDI recurrence include the prescription of non-CDI antibiotics after the diagnosis of CDI, older age, the use of proton pump inhibitors (PPIs), and strain type [15].

1.6. Symptoms

The severity of CDI cases in patients has been shown to vary. A mild case of CDI is defined by a case of mild diarrhea [131]. A moderate case of CDI involves diarrhea and extra signs or symptoms that are not associated with a severe case of CDI. A severe case of CDI involves a patient's white blood cell count ≥ 15,000 cells/μl [18, 100, 131]. As well, severe CDI is further described by a wide range of possible signs or symptoms including ileus or megacolon [18], end organ failure (e.g., renal failure), hypotension without vasodepressors [131], or presence of pseudomembranous colitis [100].

Some individuals have been found to be asymptomatic carriers of *C. difficile*. The age of these individuals ranges from children under 1 year old to elderly adults. The percentage of neonates asymptomatically colonized with *C. difficile* has been shown to be between 40% and 60%. The rate of asymptomatic colonization of healthy individuals over the age of 1, however, drops to around 2% to 3% [7]. It should be noted that the presentation of symptoms associated with CDI in neonates is considered rare [153]. The reason for the lack of symptoms is not currently understood. It was hypothesized the reason for this is related to a lack of expression of a receptor the *C. difficile* toxin binds in neonates. A preliminary study by Eglow et al. (1992) demonstrated the effect of purified *C. difficile* Toxin A on the intestinal brush border on rabbit ileum. Toxin A showed stronger binding affinity and caused more damage to the ileal loop of 90-day old rabbits compared to that of 5-day old rabbits [31].

1.7. Recurrence

Once the initial episode of CDI has been cleared a secondary episode may occur which is known as a recurrence of CDI in the patient. The infection is considered a recurrence if CDI re-emerges at least 10 days after the patient has resolved the initial case of CDI [6]. In a study by Barbut et al. (2000) they observed the time between the first episode and the recurrence of the infection to be between 10 and 211 days [6]. In 2010 a study by Lowy et al. found the rate of the first episode of recurrence in a group of patients with CDI given a placebo treatment was 25%. The chance of a subsequent recurrence was seen to rise in accordance with the number of previous recurrences a patient had experienced [89].

Recurrence of CDI can be categorized into two types, relapse or reinfection. A relapse is considered a secondary infection by the same strain of *C. difficile* that caused the primary infection. A reinfection is an infection of a different strain of *C. difficile,* but can be considered a recurrence according to current clinical definitions [6]. Reasons for relapse or reinfection have been associated with prolonged disturbance of the natural gut microflora [92], or an inability of the patient to mount an immune response against *C. difficile* or its toxins [74]. It has been hypothesized by Tedesco et al. that the persistence of spores in the gut may also be related to recurrence of CDI, but this has not been fully investigated [132].

The differentiation between relapse and reinfection is important as it can theoretically help us better treat a recurrent infection. In the Barbut study previously mentioned, they found CDI recurrence was due to reinfection 48% of the time [6]. Another study about CDI recurrence by Wilcox et al. showed a similar conclusion finding cases of recurrence to be reinfections in 53% of the patients studied [145].

1.8. Case Definition of CDI

Presently, Community-associated CDI (CA-CDI) is differentiated from Hospital-associated CDI (HA-CDI) by a basic set of guidelines (Figure 1) [25]. There are two terms that help fully describe a case of CDI, onset and association.

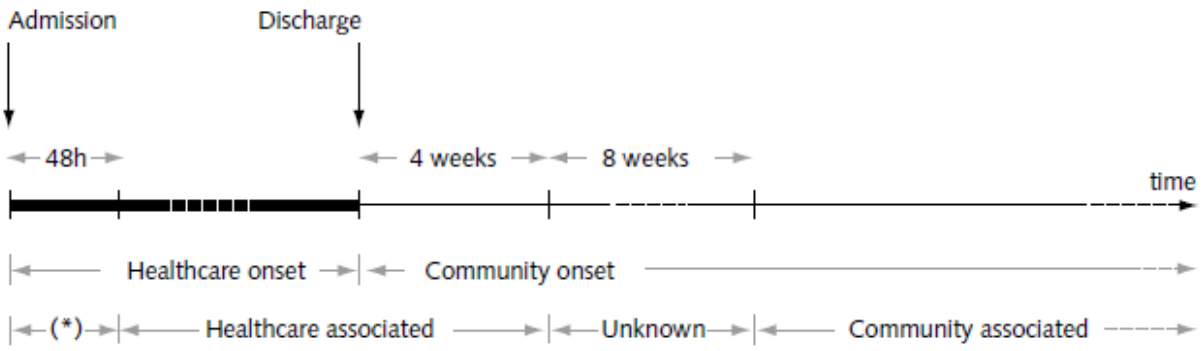

Admission
Discharge
48h
4 weeks
8 weeks
time
Healthcare onset
Community onset
(*)
Healthcare associated
Unknown
Community associated

Figure 1. A differentiation between *Clostridium difficile* infection onset and association within a healthcare or a community setting [25].

The (*) denotes a period of time where CDI could be healthcare or community associated depending on the patient's history.

Onset is used to describe where the symptoms of infection first appeared. It is a lot simpler than association, but still helps to describe the infection. CDI is considered healthcare-onset if the symptoms began in a healthcare setting. If the patient starts experiencing symptoms in the community then the case of CDI is considered community-onset [67].

Association is used to determine in what setting the infection originated from. The true origin of where the patient acquired CDI is rarely known. Elderly individuals have shown carriage of *C. difficile*, some of whom may then develop CDI at a later stage. This can confound analysis of infection association making it difficult to understand. It is preferable to use the term community-associated CDI to encompass all of the cases of symptomatic CDI that occur in patients in the community setting [146]. A case of CDI is defined as healthcare-associated if symptoms occurred less than 4 weeks after discharge from, or more than 48 hours after admission to, a healthcare facility. The other definition of association is community-associated CDI. This definition is used if symptoms of CDI occurred before hospital admission to, or more than 12 weeks after discharge from, a healthcare facility. As well, CDI is typically defined as community-associated if onset of patient symptoms occurs within 48 hours of admission to a hospital, but only if symptom onset was more than 12 weeks after the last discharge from a hospital [67]. It should be noted that these guidelines are not standardized. Another source defines community-associated CDI if onset occurred within 72 hours of admission, not 48 hours [131]. Association of a case of CDI to a setting is considered an unknown association if symptoms occurred between 4 and 12 weeks after hospital discharge. In this time period it is difficult to judge if the infection was from a healthcare or community setting [67].

1.9. Treatment

Treatment must take into account the patient's age and serum creatinine levels [18] as well as serum albumin levels [131] as these can factor into potential risks outside of the presented symptoms of the patient. For example, an elderly patient who presents with a mild case of CDI may be at risk for a more severe type of CDI as their immune system is not functioning at the same effectiveness as a younger individual.

The initial consideration for the treatment of CDI involves taking the patient off the antibiotic regimen that is leading to the proliferation of *C. difficile* [18, 100]. This step involves allowing the patient's natural gut microflora to return to normal proportions to out-compete the *C. difficile* while providing the patient with rehydration to counteract the fluid loss from diarrhea. This method may also decrease the risk of recurrent CDI [18]. A study that looked at the effect of ciprofloxacin on the distal gut microflora in healthy adults found that the recovery time of the microflora, after antibiotic exposure, took one week to two months depending on the patient [26]. With some cases of CDI it may be feasible to take the patient off the antibiotic as the major part of therapy. However, considering patient condition and CDI severity this method may not be best. The above study only described data from healthy individuals. If the immune system is under stress from a microbial infection the time for return of the natural microflora to normal levels may take even longer. This could lead to the patient experiencing the symptoms of CDI for a longer amount of time or an increased hospital stay which can put a stress on the patient as well as the hospital's resources. The patient should also be kept off antiperistaltic agents if possible [18, 100, 131] to help avoid the formation of toxic megacolon [18].

Since the late 1980's metronidazole and vancomycin have been the main antimicrobials used to treat CDI [18]. Metronidazole is in a class of antimicrobials called the nitromidazoles. It

is activated when its nitro-group is reduced by bacterial ferredoxin-based enzymes/hydrogenases within the bacterial cell. This active form, which contains a reactive radical species, is able to destroy the prokaryotic DNA. Vancomycin is in a class of antimicrobials called the glycopeptides. It acts to block cell wall synthesis of bacterial cells by binding the D-ala-D-ala amino acid sequence of pentapeptide precursor molecules. This stops the precursor from binding other precursors to form a net of mature peptidoglycan around the bacterial cell [2]. A study from 2007 compared the effectiveness of both metronidazole and vancomycin against all severities of CDI. They found that against mild CDI clinical cure rates were 90% and 98%, respectively. However, against severe cases of CDI the clinical cure rates were 76% and 97%, respectively [150]. Both antimicrobials were shown to be comparative against mild CDI, but metronidazole is typically chosen in hospital settings due to lower cost and to avoid the risk of selecting for vancomycin-resistant enterococci.

The first line therapy for a mild to moderate case of CDI referenced in both American and Canadian CDI treatment guidelines is a course of metronidazole that lasts for 10-14 days [18, 100, 131]. The first line therapy for a severe case of CDI referenced in both American and Canadian CDI treatment guidelines is a course of vancomycin that lasts for 10-14 days [18, 100, 131].

Recurrent CDI is treated differently depending on the status of the infection. That is, the first recurrent infection is typically treated with the same antimicrobial regimen as the initial case of CDI unless the symptoms relate to a severe infection. If the first recurrence presents as severe then the patient should be put on a course of vancomycin [18, 131]. The second recurrence of CDI calls for a pulse regimen of vancomycin [18, 131] whereby a short, high-dose burst of the antimicrobial is followed by an extended dose-free period [59]. This method has been

demonstrated with metronidazole and was shown to be bactericidal against strains of *Bacteroides* spp. that were either sensitive or resistant to metronidazole [59]. If there is a third recurrence of CDI in the patient a pulse regimen of vancomycin should be administered again. At this point, a fecal microbiota transplant (FMT) from a healthy donor may be considered to help out-compete the *C. difficile* colonizing the patient's gut [100, 131]. The use of probiotics to help negate CDI recurrence has not shown enough compelling evidence of effectiveness to permit use with all cases of CDI. However, it can still be considered at an individual patient level if the condition of the patient warrants it [18, 100, 131].

A few alternative options for treating CDI with antimicrobials have been tested. These include bacitracin, fusidic acid, nitazoxanide, rifampin, rifaximin, teicoplanin, and fidaxomicin. Of these, only teicoplanin and fidaxomicin appeared to have promising results [99]. The antimicrobial fidaxomicin has been further compared to vancomycin in terms of effectiveness and feasibility for treatment of CDI [100, 131]. Fidaxomicin is in a class of antimicrobials called the macrolides. Its mechanism of action involves inactivating the σ-subunit halting the creation of a functional prokaryotic RNA polymerase [75]. An initial, but limited study of fidaxomicin found that the antimicrobial demonstrated non-inferiority to vancomycin and in two trials demonstrated 16.9% and 19.6% reduction in risk of CDI recurrence compared to vancomycin. This is contrasted by the fact that the current 10-day fidaxomicin treatment course in an American guideline for CDI treatment was quoted at $2,800. This is compared to a 10-day course of vancomycin which costs $1270 under the same guidelines. As well, it should be noted that a *C. difficile* strain has already been found with an elevated minimum inhibitory concentration (MIC) to fidaxomicin [75].

Prospective new antimicrobials for treating CDI include LFF571 and teixobactin. LFF571 is a semi-synthetic broad-spectrum thiopeptide antimicrobial that acts to inhibit the function of elongation factor TU (EF-TU) to hinder bacterial protein synthesis [136]. Teixobactin is a cell wall synthesis inhibitor that targets highly conserved sugar and phosphate moieties of cell wall precursor pentapeptides exposed on the surface of Gram-positive bacteria. The two main target molecules containing these moieties are undecaprenyl-N-acetylmuramic acid-pentapeptide (UDP-MurNac-pentapeptide; a target of vancomycin as well) and undecaprenyl-pyrophosphate-N-acetylglucosamine (UDP-PP-GlcNac). Binding of these molecules leads to a disruption of both call wall synthesis and cell wall teichoic acid (WTA) synthesis. Interference with WTA synthesis is not lethal; however some mature WTA molecules anchor autolysins. These autolysins are released when teixobactin binds WTA precursor molecules which go on to uncontrollably hydrolyse peptidoglycan further disrupting the cell wall [87].

Treatment options for CDI that do not involve antimicrobials are also being investigated. A preliminary study by Gerding et al. hypothesized that giving doses of nontixogenic *C. difficile* spores orally helped patients avoid recurrence after resolution of CDI. As discussed earlier, the gut microflora does not return to normal immediately after antibiotic treatment. Nontoxigenic spores were given to the patients immediately after the first episode of CDI had resolved. Taken early enough, these spores are theorized to occupy the same environmental niches in the gut as toxigenic *C. difficile,* although this remains to be tested [43]. This idea is supported by findings from another study that focused on asymptomatic individuals colonized by *C. difficile*. These individuals were found to have a lower rate of CDI compared to those not colonized by *C. difficile* at 1.0% and 4.5% respectively after a two week follow-up by the hospital [126]. The

spore treatment creates another barrier for toxigenic *C. difficile* strains to colonize the gut and cause subsequent disease. The Gerding study analysed the results of the phase two trials for this treatment and revealed promising results. Recurrence was found to be significantly lower in the patients who received the nontoxigenic *C. difficile* spores (11%) compared to the placebo control group (30%). The nontoxigenic strain successfully colonized 69% of the patients given the spore treatment. If this treatment is to become a viable option it may require a nontoxigenic strain that is more effective at colonizing the human gut. As well, this study only involved 157 patients and the researchers could not rule out the possibility that toxigenic strains could pass on toxin genes to their nontoxigenic treatment strain [43].

1.10. Pathogenesis Mechanisms

There are 3 major virulence traits that factor into the pathogenicity of *Clostridium difficile*. These include the collection of toxin genes that are part of the pathogenicity locus (PaLoc), the ability of *C. difficile* to sporulate, and the binary toxin.

1.10.1. Pathogenicity Locus

Strains of *C. difficile* that are able to produce toxins have what is known as the PaLoc in their genomes. PaLoc contains all of the genes needed to produce and regulate the *C. difficile* toxins A and B, which are encoded by *tcdA* and *tcdB*, respectively (Figure 2) [140]. These two toxins act to modify the Ras superfamily of GTPases in the host cell by an irreversible glycosylation event to inactivate the GTPases. This leads to the disruption of downstream signalling of host cell processes. The inactivation of GTPases by TcdA disrupts host cell control over actions such as apoptosis, actin condensation, and transcriptional activation [140]. TcdB on the other hand appears to lead to the disruption of tight junctions between host cells as well as signalling cascades related to apoptosis [140]. This disturbance of the natural state of the host

cell leads to host immune responses and fluid accumulation around the area of infected cells. Other genes in the PaLoc are regulatory in nature. TcdC is a negative regulator of toxin production. Three types of in-frame base pair (bp) deletions have been described in the *tcdC* gene that disrupts its natural function. These deletions can be 18 bp, 39 bp, and 54 bp in length [105]. As well, two kinds of mutations have also been identified in *C. difficile*. These include a 1 bp truncating deletion present in the gene at position 117 and a nucleotide transition at position 184 from a cytosine to a threonine [105]. Additionally, both a deletion and a mutation have been described in the *tcdC* gene of individual *C. difficile* genomes. In contrast to this, *tcdD* encodes a positive regulator of toxin production. The true effect of *tcd*E has not yet been elucidated, but it exhibits homology to a gene encoding a holin protein, which is thought to aid in toxin release from the *C. difficile* cell by permeabilization of the bacterial cell wall [140].

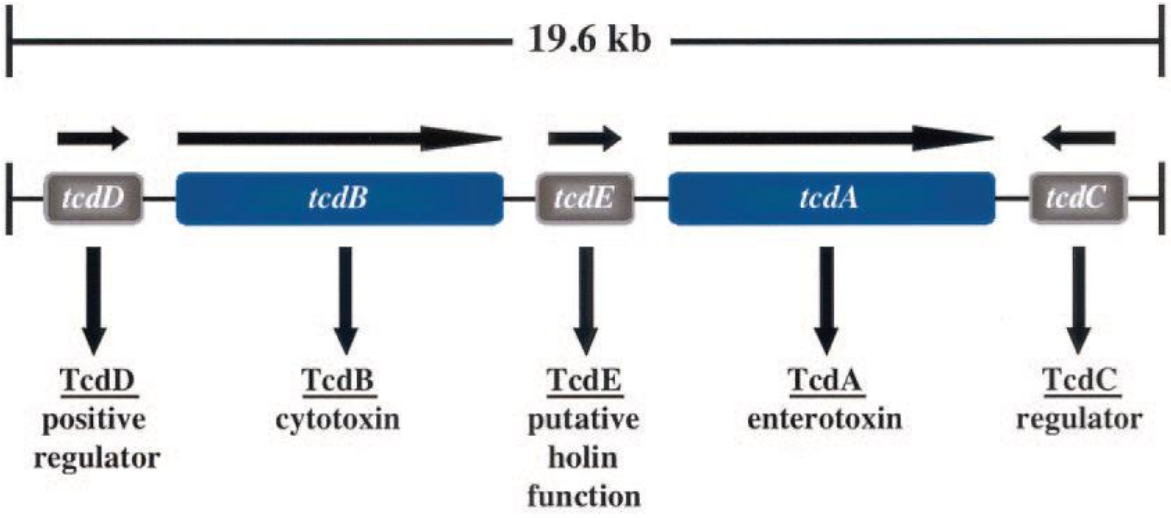

19.6 kb
tcdD
tcdB
tcdE
tcdA
tcdC
TcdD
positive
regulator
TcdB
cytotoxin
TcdE
putative
holin
function
TcdA
enterotoxin
TcdC
regulator

Figure 2. The organization of genes in the Pathogenicity Locus (PaLoc) in *Clostridium difficile* [140].

1.10.2. Binary Toxin

Binary toxin, produced by the genes *cdtA* and *cdtB* [44], is a relatively uncharacterized toxin compared to the two main *C. difficile* toxins in the pathogenicity locus. Originally described in 1988 by Popoff et al. [108], these genes and their toxin have only recently begun to be characterized. Research of binary toxin described the toxin inducing microtubule reformation inside intestinal epithelial cells to create long, surface protrusions. These protrusions appeared to wrap and embed bacterial cells on the outside of the host cell, increasing *C. difficile* intestinal adherence [123]. The prevalence of binary toxin in *C. difficile* isolates has been found at approximately 22% [30]. The effect of a *C. difficile* isolate being able to produce binary toxin has been shown to increase case fatality rates compared to isolates that only produced Toxin A and Toxin B [4]. Only recently have *C. difficile* isolates been found that were positive for the binary toxin genes, but negative for *tcdA* and *tcdB* activity [129]. Despite lacking *tcdA* and *tcdB*, these isolates remained virulent as demonstrated in a hamster model [4].

1.10.3. Sporulation

The ability of *C. difficile* to transition from a vegetative state to a spore is a major contributor to the survivability and transmission of the organism. The spore is able to resist cell damage from bleach-free disinfectants, antibiotics, and host immune cells [101]. For one, this allows *C. difficile* to persist in an aerobic environment outside of the host where a vegetative form would otherwise die. This ability also allows for the bacterium to survive in the colonic tract of the host in a form that is difficult to destroy. Spore formation is thought to be a major reason for the recurrence of CDI. A cross section of *C. difficile* spores helps us understand the major spore layers (Figure 3). Closest to the inside is the cortex which is a modified peptidoglycan layer. The next layer is the spore coat which doesn't infer any antibiotic resistant capabilities, but has been shown in *Bacillus subtilis* to aid in resistance to commonly used decontaminants. Finally, the outer-most layer of the spore is the exosporium, which in *C difficile* is not as well understood with regards to its fragility. What the exosporium does confer is the ability of the otherwise dormant spore to interact with its environment with respect to receptor binding or adhesion [101].

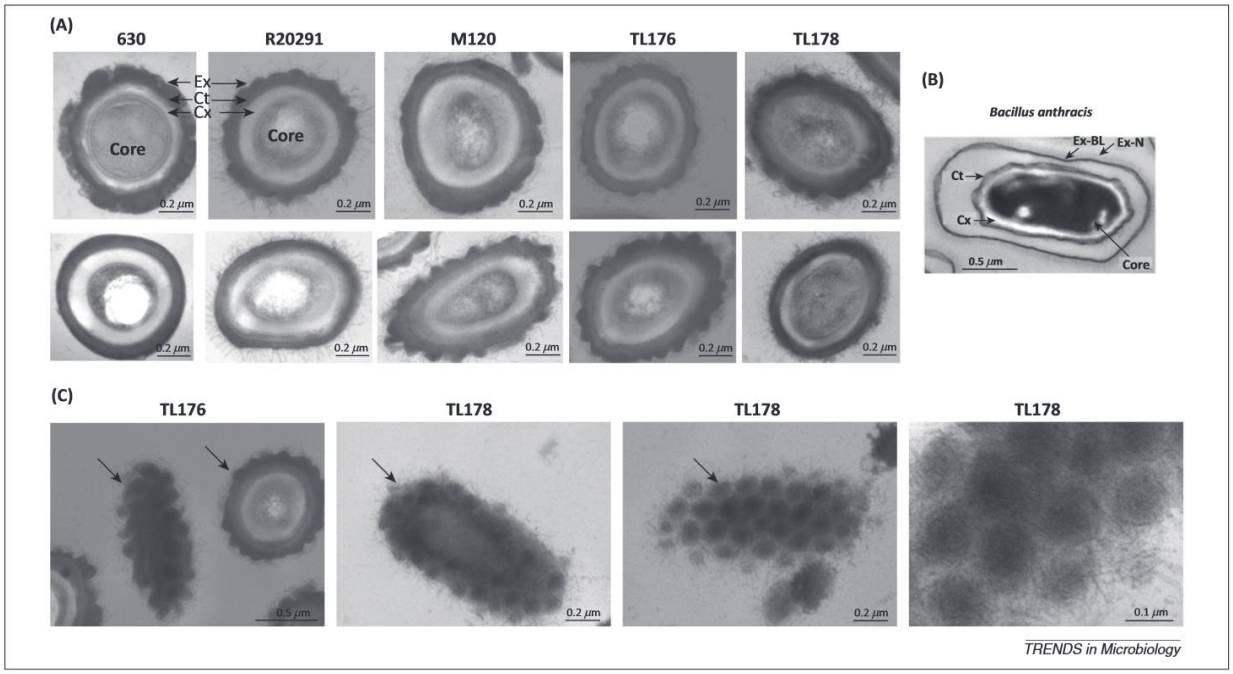

(A)
630
R20291
M120
TL176
TL178
Ex
Ct
Cx
Core
Core
0.2 μm
(B)
Bacillus anthracis
Ex-BL
Ex-N
Ct
Cx
Core
0.5 μm
(C)
TL176
TL178
TL178
TL178
0.5 μm
0.2 μm
0.2 μm
0.1 μm
TRENDS in Microbiology

Figure 3. Ultrastructure of *Clostridium difficile* spores [101]. (A) Thin sections of various *C. difficile* strains in spore form. Upper and lower panels show transmission electron microscopy (TEM) images of the five strains. In the upper panels, spores with a typical exosporium-like structure with high electron-dense material are shown. The lower panels show spores with less electron-dense exosporium-like ultrastructures. Abbreviations: Ex, exosporium; Ct, coat; Cx, cortex. (B) The ultrastructure of a *Bacillus anthracis* spore is shown for comparison. Abbreviations: Ex-BL, exosporium basal layer; Ex-N, exosporium hair-like nap; Ct, coat; Cx, cortex. (C) The ultrastructure of the surface of spores from two *C. difficile* strains is shown to highlight the well-organized surface structure comprised of unknown exosporium proteins.

1.10.4. Antibiotic Resistance

Antimicrobial resistance to clindamycin, rifampin, moxifloxacin, and metronidazole in *C. difficile* has been well described in a recent study by Tenover et al. (2012) in which 316 North American *C. difficile* isolates were analysed. The observed resistance rates of the tested antimicrobials were 41.5%, 7.9%, 38%, and 0%, respectively [135]. It was noted in another study that reduced susceptibility to vancomycin or metronidazole is rarely observed in *C. difficile* [57]. However, resistance to metronidazole was recently reported, in which single nucleotide polymorphisms (SNPs) were found in several genes involved in key metabolic pathways [90]. For example, a SNP was found in the *fur* gene which has been associated with metronidazole resistance in *Helicobacter pylori* leading to altered binding to superoxide dismutase. This would decrease oxidative stress within the cell [90].

1.11. Typing Methods

1.11.1. Toxinotyping

Toxinotyping is a way of differentiating between *C. difficile* isolates by utilizing both polymerase chain reaction (PCR) amplifications and restriction fragment length polymorphism (RFLP) digestions. The entire pathogenicity locus of a *C. difficile* genome is amplified using 10 different primer pairs. Typically, only one primer pair for each of the toxin genes, *tcdA* and *tcdB*, is required for use in toxinotyping. The amplification targets are the A3 fragment of *tcdA* and the B1 fragment of the *tcdB* gene. Both fragments are around 3.1 kb in length. The amplified fragments are then digested with restriction enzymes and ran on a 1% agarose gel. Banding patterns are assigned a number based on which known patterns they match. Of the two amplified fragments, A3 is digested with restriction enzyme *Eco*RI whereas B1 is digested using the restriction enzymes *Acc*I and *Hinc*II. Fragment A3 has fourteen known banding patterns.

Fragment B1 has 7 known banding patterns. When both numbers are known for each gene fragment they are compared to a database of known combinations of A3/B1 numbers to find the isolate's toxinotype [118]. There are currently 31 known *C. difficile* toxinotypes [119]. In a case where the A3 and B1 banding patterns are the same, other fragments of the toxin genes must be amplified and digested to find the isolate's true toxinotype.

1.11.2. Restriction Endonuclease Assay

A restriction endonuclease assay (REA) involves cutting the *C. difficile* genome with a restriction enzyme and comparing the banding patterns of isolates. REA uses *Hin*dIII as its restriction enzyme. This method was first used in a hospital in The Netherlands in 1986 to track the spread of *C. difficile*. This method involves overnight runs on a large, expensive electrophoresis machine. With proper standard ladders loaded on the gel, banding patterns can be compared without too much difficulty [58]. It should be noted that banding patterns produced from using REA are not as easy to differentiate as those produced by pulsed-field gel electrophoresis [61].

1.11.3. Multilocus Variable Number Tandem Repeat Analysis

Coding and non-coding segments of a genome may contain regions of repeating genetic sequence. When the size of these repeating nucleotide patterns vary between isolates of the same species these regions are called variable number tandem repeats (VNTRs). Multilocus variable number tandem repeat analysis (MLVA) is a banding method that focuses on amplifying several of these VNTRs in a genome at one time. MLVA is a PCR-based method [121]. Different isolates of the same species can have varying numbers of a repeat sequence at a locus [34]. Initially, MLVA of *C. difficile* involved two different schemes, one from 2005 and one from 2006. Both schemes involved seven target loci. After the PCR, amplicons of all loci are either

run on a gel or through a capillary gel electrophoresis machine. Each band is given a number based on the size of the band equating to the amount of times a sequence repeats in a locus. The number of each locus is concatenated into a string of seven numbers for each isolate. This string is called the summed tandem repeat difference (STRD). Clonal clustering is based on a STRD value ≤ 2 whereas genetically related clustering has a STRD value ≤ 10. The lower amount of differences between MLVA number strings, the lower the STRD, and the closer the isolates are related. Over the years several loci have been replaced by more optimal loci or PCR conditions have been adjusted. The current MLVA method for *C. difficile* was established in 2011 and has been used to assess several outbreaks in Europe [69]. *C. difficile* MLVA has an optional method which is known as extended multilocus variable number tandem repeat analysis (eMLVA) which involves 15 target loci. Four multiplex PCR assays are done to cover all 15 targets. Multiplex PCR uses several primer pairs in an assay tube at once each amplifying a different region of the genome to save time and resources [63]. MLVA has a high discriminatory power for good phylogenetic comparisons. Also, with current methods MLVA output is in a digital format facilitating data exchange between laboratories. However, MLVA is expensive because it requires a lot of resources and can take several days for isolate processing [58]. In terms of eMLVA, the discriminatory power is even higher, but more resources are required because there are more targets to amplify [63].

1.11.4. Multilocus Sequence Typing

Multilocus sequence typing (MLST) is a sequence-based method that uses PCR to amplify and sequence fragments of housekeeping genes. The first time this method was used was in 1998 on *Neisseria meningitidis* [121] and it was subsequently adapted for *C. difficile* in 2004 [78]. In the current scheme for *C. difficile* MLST there are seven target housekeeping genes. The

targets are *adk, atpA, dxr, glyA, recA, soda,* and *tpi,* which are an adjustment from the initial scheme targets outlined in 2004. The sequence of each allele is assigned a number based on which known allele sequence it matches. The seven allele numbers for each isolate are combined to generate a multilocus allelic profile which translates to a sequence type [49]. On the current *C. difficile* MLST database there are 347 unique MLST sequence types listed [64]. MLST has good discriminatory power for *C. difficile* isolates. As well, since it produces electronic sequence files results are easily transferrable between laboratories. The largest downside to doing MLST is the cost as one isolate requires seven primer pairs. Also, not every lab contains a sequencer so outsourcing the sequencing step of MLST will increase the time and cost associated with this method [58].

1.11.5. Ribotyping

Ribotyping is a banding-based method for typing whereby conserved intergenic spacer regions between the 16S and 23S ribosomal RNA genes of the genome are amplified by PCR. The method was originally described by Gürtler in 1993, but has since been refined [58]. The copy number of this combination of a 16S gene, intergenic spacer, and a 23S gene differ between *C. difficile* isolates lending to the variation seen by ribotyping. It was estimated that there are between 1 and 15 copies of this intergenic spacer region that can be found in a *C. difficile* isolate [58]. The theorized mechanisms resulting in different isolate copy numbers include slipped-strand mispairing as well as inter- and intrachromosomal homologous recombination events [72].

Originally, amplicons were run on a 1.5% agarose gel and the banding patterns of each isolate were used to determine their respective ribotypes. A more refined approach was later implemented by running the amplicons on a capillary gel electrophoresis machine. For this approach, ribotyping primers have fluorescent tags attached to them. These tags allow the PCR

amplicons to be run through a genetic analyser that has capillary tubes loaded with POP4 polymer. The genetic analyser detects the amplicons by the fluorescent tag attached to them allowing for computer software to create an electronic copy of the ribotype banding pattern for each isolate. This capillary-based approach increased the rate of ribotype determination and improved inter-laboratory reproducibility of ribotyping results [60].

Ribotyping is the preferred typing method for *C. difficile* in Europe and Australia and has been recently employed in the United States [58]. A paper from 2008 by Killgor et al. described two primer pairs used for ribotyping. One set, designed by Stubbs et al., was used in the United Kingdom whereas another set, designed by Bidet et al. was employed in America [68]. Currently over 400 unique PCR ribotype patterns are known [69]. Ribotyping has its own strengths and weaknesses, each of which should be described in terms of the method employed to obtain the final banding pattern. In terms of the agarose gel-based method, there is a reference library available for comparative purposes. However, results are not easily interchangeable between laboratories. The capillary gel-based method is interchangeable between labs, is easily reproducible, and is quite accurate. Recently, a standardized reference library has been made available for electronic ribotyping endeavors [33].

1.11.6. Pulsed-field Gel Electrophoresis

Pulsed-field gel electrophoresis (PFGE) is a restriction enzyme-based method analysing genome fragment banding patterns. It was one of the first *C. difficile* typing methods [58] and is currently the most common method used in North America [71, 72]. PFGE involves a whole-genome macro-restriction digest with an enzyme that cuts infrequently in the genome. For *C. difficile* the restriction enzyme of choice is typically *Sma*I [69, 127] although *Sac*II and *Ksp*I have been described for use in PFGE for *C. difficile* as well [61]. The restriction digests are run

on an agarose gel with an electric field switching between three different directions. One direction is through the center axis of the gel, but the other two fields are angled 60 degrees away on either side of the central axis. Only one of the three fields is on at one time. The constant changing of direction facilitates migration of the long DNA fragments through the agarose gel matrix. The time it takes to change direction of the electric field (pulse time) increases over the length of the run to aid in better separation of fragment sizes especially large bands. The banding patterns observed for each isolate are compared to a standard set [69]. PFGE banding patterns that have ≥80% similarity are grouped into North American Pulsed-field types (NAP types) [68]. A NAP type can also be referred to as a pulsotype, but that particular nomenclature is less common [61]. There are currently 12 known NAP types for *C. difficile*. There are several downsides to choosing PFGE as a typing method. It typically involves overnight runs on the electrophoresis machine [61], which are expensive and require a lot of space. PFGE is limited to running a number of samples based on how many wells are made in the gel.

1.11.7. Whole Genome Sequencing

1.11.7.1. Sequencing Platforms

The first common method for sequencing bacterial genomes was established by Sanger et al. in 1977. The method now known as Sanger sequencing utilizes dideoxy terminator nucleotides. These custom nucleotides prematurely stop amplification of a certain proportion of DNA strands in an amplification reaction. The output of this method requires the interpretation of amplicon banding patterns on a gel to read the sequence base-by-base [122]. There are currently several modern sequencing platforms in use today. Two popular platforms include MiSeq (Illumina, San Diego, CA) and PacBio (Pacific Biosciences Menlo Park, CA).

MiSeq involves a synthesis-focused sequencing approach. Once the extracted DNA is sheared into fragments adapters are ligated to the ends. This sequencing system clonally amplifies DNA templates immobilized on an acrylamide coating on the surface of a glass flowcell. The adapters ligated to the ends of the sheared DNA fragments are used for this immobilization. The process of creating sequence-ready DNA requires fluorescently labeled reversible-terminator nucleotides. When the amplified DNA fragments are run through the sequencer the fluorescent terminators attached to the DNA strand are exposed to a detector. The detector signals the sequencer that a certain nucleotide has passed. With the size of DNA fragments increasing one base at a time the machine can determine the sequence of the DNA fragments on each flowcell [112].

The PacBio platform uses what is known as single molecule real time (SMRT) sequencing. With SMRT, DNA polymerase molecules are attached to the bottom of 50 nm-wide wells. The polymerases amplify single-stranded DNA molecules. Each polymerase carries out DNA synthesis of the complementary strand in the presence of fluorescently labeled γ-phosphate nucleotides. The width of the well is such that light cannot propagate through it, but energy can penetrate a short distance to excite the fluorophores attached to the added nucleotides that are close to the polymerase at the bottom of the well. As each base is incorporated, a distinctive pulse of fluorescence is detected in real time by a detector in the PacBio sequencer [112].

1.11.7.2. Comparing Sequencing Platforms

Comparison between these platforms has been previously described [112] and an evaluation of the two platforms is provided in Table 1.

Table 1. Technical specifications of next generation sequencing platforms

Platform	**MiSeq**	**PacBio**
Instrument Cost	$128K	$695K
Sequence Yield per Run	1.5 - 2.0 Gb	1.0 Gb
Sequencing Cost per Gb	$502	$2,000
Run Time	27 Hours	2 Hours
Reported Accuracy	> Q30	< Q10
Observed Error Rate	0.80%	12.86%
Read Length	up to 150 bases	Average 1500 bases
Insert Size	up to 700 bases	up to 10 Kb
Starting DNA Amount	50-1,000 ng	1 μg

Of note, MiSeq has the lowest cost per gigabase of sequence data, but in some cases due the lower yield per run, a top-up run may be required to obtain sufficient data. PacBio has the lower run time, but this doesn't include time taken to prepare the DNA for sequencing which can take longer depending on the library kit used. Sequence read accuracy is measured as quality scores. Q10 relates to 1 miscalled base every 10 bases, Q20 is 1 miscalled base every 100 bases, and Q30 is 1 miscalled base every 1,000 bases. MiSeq has the best average quality of sequence reads [88, 112]. There are methods in place to improve the poor quality of PacBio long reads. An example of this includes using high quality short reads to map to the long reads. Computer software can determine the proper sequence using the short reads in regions of poor quality sequence data. This does, however, increase the time required to complete PacBio sequencing [3]. MiSeq has the lowest observed error rate of the two platforms. Read length describes how far a DNA polymerase can synthesize a complementary strand of DNA before releasing the template DNA. PacBio has been shown to achieve the longer read length of the two platforms. Insert size describes the length of the sheared template DNA for the type of sequencing being done. This can be related to the insert size of a vector used for adapter addition or the read length capacity of a DNA polymerase. PacBio has the longest read length with an average of 10Kb per genomic fragment. Finally, starting DNA required is much lower when using the MiSeq platform. Costs associated with DNA extraction increase with the amount of DNA required by the sequencing platform. Having a low amount of starting DNA can be seen as lower cost and time to acquire the final sequence data [88, 112].

Whole genome sequencing (WGS) has finally reached a point where accuracy and reliability of the method have improved to be considered for typing large numbers of isolates in a short time. This sequencing-based method aims to assess the entire core genome of a species

looking for single nucleotide polymorphisms (SNPs) to help differentiate isolates. The core genome is a collection of the genes shared by all organisms of that species. Finding SNPs helps to evaluate population structures and epidemic lineages to track not only strains involved in infection, but also virulence traits such as antibiotic resistance features. Whole genome sequencing is conceptually the best genotypic typing method in that all of the results obtained from most typing methods are provided by this one method. In order to determine if an isolate is part of an outbreak the goal is to compare its core genome to well characterized genomes. To do so, the amount of SNPs that can occur, yet be associated with those from an outbreak, must be established [69]. One source found the mutation rate of *C. difficile* to be at 1-2 SNPs per genome per year at a 95% confidence interval [53]. Another found the rate of mutation to be at 2.3 SNPs per genome per year [32]. This typing method is very accurate and has the highest discriminatory power among all modern typing methods. Furthermore, the data is stored electronically allowing for easy transfer [58]. The weakness of whole genome sequencing is the associated costs. Depending on the sequencing platform chosen results could take days to weeks to acquire. Sequencing also requires bioinformatics expertise to properly process the data [58]. However, computational pipelines are being established to ease the bioinformatics requirement of this method [69].

Whole genome sequencing has proven to be a promising tool as shown in the study by He et al. where 151 *C. difficile* genomes from 1985 to 2010 were analysed. This study was done to determine the cause of the major outbreaks of NAP1 in the early 2000's and its subsequent global spread. Researchers found two distinct lineages of *C. difficile* that had individually acquired the same SNP in a DNA gyrase subunit gene, *gyrA*. This mutation led to both lineages gaining resistance to fluoroquinolones [53]. The two lineages were known as fluoroquinolone

resistant group 1 (FQR1) and group 2 (FQR2) (Figure 4). A Bayesian comparison showed the most recent common ancestor of FQR1 and FQR2 appeared around 1993 and 1994, respectively. The earliest isolate of FQR1 was found in 2001 from the first major outbreak in Pittsburgh, Pennsylvania representing the hypervirulent NAP1 strain. One isolate that is thought to be the first isolate from FQR2 appeared in Montreal in 2003, around the time the first major Canadian outbreak occurred. It has been hypothesized that the hypervirulent nature of NAP1 in North America was due to its virulence traits. It is known that NAP1 isolates have a deletion in the *tcdC* gene that allows for overexpression of toxin. However, with this study the observation of the PaLoc in FQR1 and FQR2 showed no change before or after the North American outbreaks. It must be noted that fluoroquniolones were one of the most commonly prescribed antibiotic classes during the late 1990's and the early 2000's. This represents an ideal time for bacteria such as *C. difficile* to acquire and maintain fluoroquinolone resistance mechanisms. This could help explain the emergence of the FQR lineages as they appeared to flourish and spread around the same time [53].

At this time, whole genome sequencing may not be feasible for many healthcare facilities or surveillance groups. However, the potential benefits are enticing organizations to consider this superior method for *C. difficile* typing.

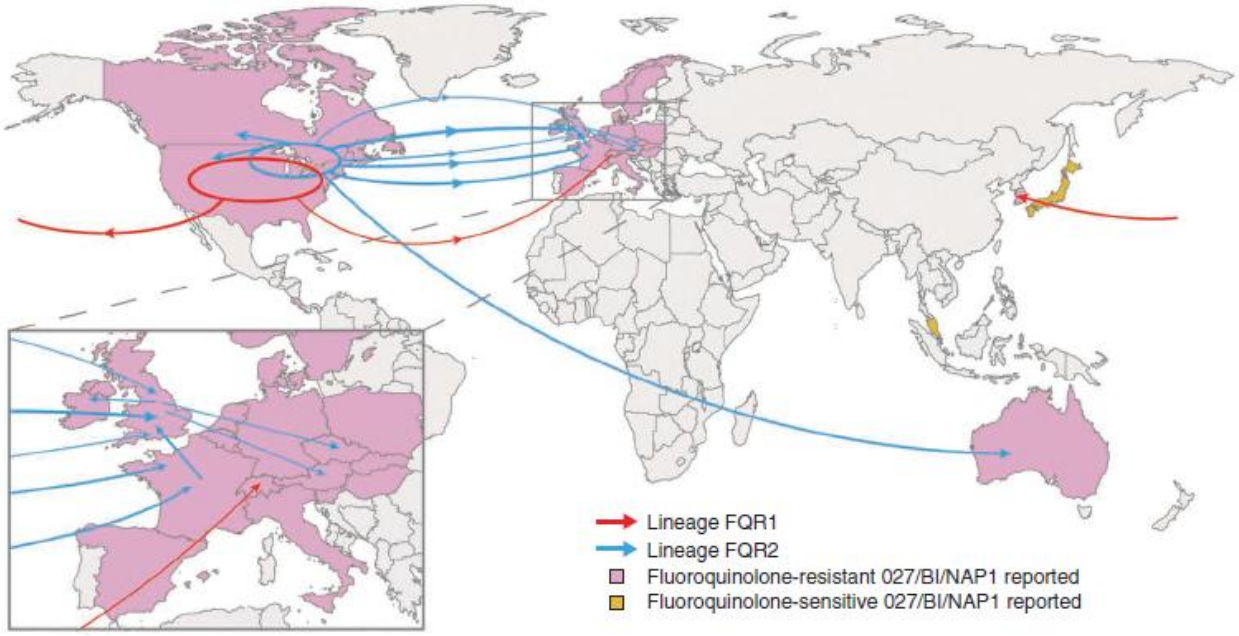

Lineage FQR1
Lineage FQR2
Fluoroquinolone-resistant 027/BI/NAP1 reported
Fluoroquinolone-sensitive 027/BI/NAP1 reported

Figure 4. The hypothesized spread of *Clostridium difficile* lineages FQR1 and FQR2 across the globe based whole genome sequencing data from years 1985 to 2010 [53].

2. Thesis Rationale and Objectives

Between 2004 and 2011 NAP4 has become one of the most prevalent *C. difficile* strains in the country. Only a limited amount of data is available from published research describing NAP4. Most *C. difficile* studies to date have focused on NAP1 due to its previously described hypervirulent nature [104]. Therefore, to explain the increase in prevalence of NAP4 throughout Canada our limited understanding of this strain needs to be expanded upon through phenotypic and genomic analysis.

The goals for this project were 5-fold: To (1) analyse data acquired from toxin PCR, PFGE, antimicrobial susceptibility testing, and ribotyping on all of our NAP4 isolates allowing full characterization of these isolates; (2) carry out growth and sporulation assays on isolates of all NAP types to see how NAP4 compares; (3) analyse patient data of NAP4 isolates to determine the range of epidemiological characteristics that exist within this group and compare them to the well characterized NAP1 type; (4) evaluate the phylogenetic arrangement of select NAP4 isolates based on their core genome content to understand the evolution and dissemination of this emerging strain; and (5) scan all of our genomes for unique nucleotide sequences that may potentially contribute to the successful emergence of this strain in Canada.

3. Hypothesis

The *Clostridium difficile* epidemic type NAP4 is increasing in incidence as a hospital-acquired infection by acquiring genomic changes over time that increase the fitness and virulence of this strain type.

4. Methods

4.1. *Clostridium difficile* Isolate Surveillance

Three hundred and thirty eight NAP4 stool samples were received from over 50 acute-care facilities across Canada through the Canadian Nosocomial Infection Surveillance Program (CNISP). Samples were collected between 2004 and 2011 with no samples from 2006. Sample collection occurred between March 1st and April 30th for patients ≥18 yrs of age and year round for patients <18 years of age. Patient information was collected from a chart review consisting of questions regarding patient age, gender, date of hospital admission, date of symptom onset, and severity markers such as ICU admission, colectomy, and 30 day outcomes.

4.2. Patient Eligibility

Patients eligible for inclusion in the CNISP CDI study had to meet several requirements. The patient had to be one year of age or older and they must have been admitted to an acute-care ward within the hospital. They also must not have been dedicated to long-term care wards, transitional care wards, or alternate level of care wards. Patients who were discharged before positive culture of *C. difficile* was determined were included in the study. Patients who returned to the ER or outpatient clinic within 4 weeks of previous discharge with a new case of CDI, and were not readmitted, were not included in the study.

4.3. Surveillance Case Definition

CDI was identified in a patient if any one of the following cases were recorded. If the patient had diarrhea or fever, abdominal pain or ileus, along with a laboratory-confirmed positive toxin assay for *C. difficile*. The patient had CDI recorded if they had a histological or pathological diagnosis of CDI, or a diagnosis of pseudomembranes from a sigmoidoscopy. Finally, a diagnosis of toxic megacolon was also considered an identification of CDI in a patient.

The definition of diarrhea can vary. It may involve 3 or more unformed stools in a 24 hour period if considered unusual for the patient, 6 or more watery stools from the patient in a 36 hour period, or 8 or more unformed stools over a 48 hour period. Diarrheal stool consistency was identified if the stool matched a Type 7 on the Bristol Stool Scale. All isolates in this study were from healthcare-associated cases of CDI. Healthcare-associated CDI was defined as CDI in a patient who was hospitalized and discharged within the previous 4 weeks or a patient who had symptoms of CDI at least 48 hours after admission. This definition was changed in 2009 to require symptoms to appear at least 72 hours post-admission to fall under the definition of healthcare-associated CDI. Only primary cases of CDI were included in this study. A primary case of CDI was defined as the first episode the patient had ever experienced or a new episode of CDI that occurred at least 8 weeks after the first positive toxin assay. All cases of death within 30 days after the first positive CDI culture were evaluated using the Death Attribution Rules for Patients Infected by *C. difficile* (DARPIC) algorithm.

4.4. Recovery and Storage of *Clostridium difficile*

4.4.1. Alcohol shock

One ml of each unformed stool sample received was transferred to a separate, sterile 15 ml centrifuge tube. Any leftover stool not used for the following alcohol shock was stored at -80°C. Formed samples were diluted with 1 ml of Phosphate-buffered saline (PBS; Thermo Fisher Scientific Inc., Wilmington, DE) pH 7.2 before further processing. All samples were emulsified by vortexing. One ml of 95% ethanol was then added to each stool suspension, vortexed, and left stationary at room temperature for 35 minutes. Shocked samples were then centrifuged at 2675 xg for 5 minutes and supernatants were discarded. Using a sterile swab, shocked stool sediment was cultured onto CDMN plates (*Clostridium difficile* Moxalactam Norfloxacin selective

supplement; Oxoid, Nepean, ON), streaking for isolation. The swab was then placed in a tube of FAB media (Fastidious Anaerobic Broth; Lab M Limited, Heywood, UK) to use as a backup if there was no growth on the CDMN plate. All volumes of broth media used in this study were pre-reduced for at least 24 hours in an anaerobic atmosphere before being used to culture *C. difficile*. Both CDMN and FAB inoculations were incubated at 37°C for 48 hours in anaerobic conditions. The remaining sediment was stored at -80°C. The anaerobic chamber used in this study was a Type A Vinyl Anaerobic Chamber (Coy Laboratory Inc., Glass Lake, MI). The accessories used with the anaerobic chamber included a forced air incubator (Coy Laboratory Inc.) and palladium catalyst (Coy Laboratory Inc.); mixed gas 5% H_2, 10% CO_2, 85% N_2; Linde Canada Ltd., Winnipeg, MB); compressed nitrogen gas (Linde Canada Ltd.); and T.H.E. Desiccant beads (VWR, Mississauga, Ontario). The incubator inside the anaerobic chamber was used to grow all *C. difficile* cultures at 37°C. If there is less than 0.5 ml of stool sample received, then the sample was alcohol shocked and sediment was inoculated into FAB media. FAB cultures were incubated at 37°C for up to 7 days in anaerobic conditions until growth was observed. If growth appeared, a sterile swab was used to collect the broth culture and streak onto a CDMN plate for isolation.

4.4.2. CDMN Growth Examination

If typical *C. difficile* colony morphology was observed on the CDMN plate, one of those colonies was streaked onto a BAK plate (Brucella Agar with hemin, laked sheep's blood, and vitamin K; BD Bacto™, Mississauga, ON) and incubated at 37°C for 48 hours in anaerobic conditions. If no growth of *C. difficile* colonies was observed, the swab from the FAB culture was used to streak onto a new CDMN plate to attempt isolation again or the primary CDMN plate was left to incubate for up to 7 days.

4.4.3. BAK Growth Examination

After incubation, *C. difficile* colonies on BAK plates were examined for purity. Colonies on the BAK plate were exposed to long-wave UV light using a UVGL-25 Compact UV Lamp (UVP, Upland, CA). *C. difficile* colonies fluoresce a chartreuse (yellow-green) colour under UV exposure and were further assessed using the Microgen Latex Agglutination Kit (Microgen, London, UK). If the colonies on the plate were determined to be pure, 1 *C. difficile* colony from the plate was streaked onto a new BAK plate and incubated as previously described. After incubation, 5 to 10 colonies from the secondary BAK plate were transferred to a Microbank™ vial (Pro-Lab Diagnostics, Richmond Hill, ON) using a sterile loop. The vial was capped and mixed by shaking. The vial was left at room temperature for 5 to 20 minutes to allow the bacteria to bind to the porous beads. Afterwards, the cryopreservative was aspirated off the beads by pipetting and discarded. All inoculated cryo-vials were stored at -80°C. One colony from the primary BAK plate was also inoculated into 4 ml of BHI broth (Brain Heart Infusion; BD Bacto™) and incubated at 37°C for 6 hours in anaerobic conditions. These broth cultures were used for further characterization by PFGE and toxin PCR tests.

4.5. Analysis of Patient data

Patient data from 338 NAP4 *C. difficile* isolates and 1,387 NAP1 isolates collected by CNISP between 2004 and 2011 were compared using a chi-squared test or a Fisher's Exact test [96] when appropriate. Patient data from NAP4 isolates was also used to compare isolates with PFGE type 0023 or 0033.

4.6. Growth Assays

C. difficile isolates from cryo-bead stocks were streaked for isolation on BHI plates and incubated anaerobically at 37°C for 48 hours. Two to three colonies from each plate were transferred to a new BHI plate, streaking for isolation. These new plates were incubated anaerobically at 37°C for 48 hours. Two to three colonies from each plate were then inoculated into 5 ml of pre-reduced BHI broth using a sterile loop and incubated anaerobically at 37°C overnight. The turbidity of each broth culture was then normalized to a #3 McFarland turbidity standard (BioMérieux, Marcy-l'Étoile, France), measuring 10^8 cfu/ml, using pre-reduced BHI broth. A 100-µl volume of this standardized suspension was then transferred to 10 ml of pre-reduced BHI broth for a 1:100 dilution. Absorbance readings (OD600nm) were measured using 400 µl of each broth culture at each time point using an Eppendorf Biophotometer (Eppendorf, Hamburg, Germany). The first assay monitored the growth of *C. difficile* isolates representing all 12 known NAP types. Growth was measured every hour for 12 hours and at 24 hours. Each isolate was analysed in triplicate. Next, several isolates with NAP1, NAP9, NAP4 PFGE type 0023, or NAP4 PFGE type 0033 were monitored for growth every 2 hours for 12 hours and at 24 hours. Each isolate was analyzed in duplicate. Both assays included two technical replicates of each isolate. The slope of the line between two time points was used to represent the growth rate of each isolate. These slopes were compared using a two-tailed student's t-test assuming unequal variance with an alpha of 0.05.

4.7. Sporulation Assays

C. difficile isolates from cryo-bead stocks were streaked for isolation on TSA + 5% sheep's blood plates (Trypticase soy agar; BD Bacto™) and incubated anaerobically at 37°C for 48 hours. Two to three colonies from each plate were transferred to 5 ml of pre-reduced BHI broth using a sterile loop and incubated anaerobically at 37°C overnight. Next, the turbidity of each broth tube was diluted with pre-reduced BHI broth to match a #3 McFarland turbidity standard measuring 10^8 cfu/ml. A 10-µl volume of this standardized suspension was further diluted in 10 ml of pre-reduced BHI broth. A 100-µl aliquot of this new suspension was then diluted in 10 ml of pre-reduced BHI broth for a final dilution of approximately 10^3 cfu/ml. These broth cultures were incubated anaerobically at 37°C for 48 hours. After incubation the broth cultures were vortexed and 500 µl of each culture was aliquoted into a sterile 1.5 ml tube for alcohol shock. A separate 500-µl aliquot was transferred to a second sterile 1.5 ml tube to be used as a control. Alcohol shock was carried out by adding 500 µl of 95% ethanol to each sample. The tubes were placed in a Shaker 20 tube shaker (Labnet, Edison, NJ) at room temperature for 40 minutes. Afterwards, the tubes were centrifuged at 16,000 xg for 2 minutes and the supernatants discarded. The cell pellets were resuspended in 500 µl of pre-reduced BHI broth. Ten-fold serial dilutions of each tube were made to 10^{-5} using pre-reduced BHI broth. A 100-µl volume of each dilution was plated in triplicate on BHI + 0.5% Taurocholate plates and incubated anaerobically at 37°C for 48 hours. After incubation, colonies were counted manually. Colony counts were averaged across all three plates for each dilution. The 1 ml control culture aliquots were centrifuged at 16,000 xg for 2 minutes. Supernatants were removed and the cell pellets were resuspended in 500 µl of pre-reduced BHI broth. Ten-fold serial dilutions of each tube were made to 10^{-5} using pre-reduced BHI broth. A 100-µl aliquot of each dilution was

plated in triplicate on BHI + 0.5% Taurocholate plates and incubated anaerobically at 37°C for 48 hours. After incubation, colonies were counted manually. Colony counts were averaged across all three plates for each dilution. Colony counts between 30 and 300 were averaged (correcting for dilutions) for both shocked and control tubes for each isolate separately. The average colony counts from the shocked tube were divided by the colony counts from the control tube for each isolate to represent that isolate's sporulation ratio. Sporulation ratios were compared using a two-tailed student's t-test assuming unequal variance with an alpha of 0.05.

4.8. Ribotyping

4.8.1. DNA Extraction

C. difficile isolates from cryo-bead stocks were streaked for isolation on TSA + 5% sheep's blood plates and incubated anaerobically at 37°C for 48 hours. Two to three colonies from each plate were subcultured into a separate, sterile 15 ml tube containing 5 ml of pre-reduced BHI broth. Tubes were incubated anaerobically at 37°C overnight; afterwhich 500 µl of each culture was transferred to a sterile 1.5 ml tube. The tubes were centrifuged at 16,000 xg for 2 minutes and the supernatants were discarded. The cell pellet was resuspended in 100 µl of Instagene Matrix (BioRad, Hercules, CA) and mixed by vortexing. Samples were heated in a heatblock at 99°C for 10 minutes to lyse the cells. Sample tubes were then centrifuged at 16,000 xg for 2 minutes and supernatants aspirated into a new, sterile 1.5 ml tube. DNA concentration of the supernatants was quantified using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Waltham, Massachusetts). Each DNA extraction was adjusted to 100 ng/µl using molecular-grade water as a diluent.

4.8.2. PCR Preparation and Amplification

We used Bidet primers for our ribotyping tests (Table 2) [68]. Both primers were made by Applied Biosystems (Foster City, CA).

Table 2. Ribotyping primers

Name	Sequence (5'-3')	Size (bp)
16S-FAM* forward primer	GTG CGG CTG GAT CAC CTC CT	20
23S reverse primer	CCC TGC ACC CTT AAT AAC TTG ACC	24

***** - 5' carboxyfluorescein (5-FAM) labelled primer

PCR master mix was made according to Table 3.

Table 3. *Clostridium difficile* ribotyping PCR mastermix contents

Component	Volume (µL)
HotStar Taq Master Mix (Qiagen)	12.5
16S-FAM forward primer (10 µM)	0.25
23S reverse primer (10 µM)	0.25
H_2O (autoclaved)	10.0
DNA*	2.00

*- DNA added at a later step

PCR master mix was kept on ice while dispensing 23 µl of master mix into each well of a sterile 96-well plate. To each well containing master mix, 2 µl of DNA lysate was added for each sample for a final reaction volume of 25 µl. MicroAmp Optical Adhesive Film (Applied Biosystems) was applied to the top of the wells to seal the plate. Well contents were mixed by vortexing the plate. Using a plate rotor, the plate was centrifuged at 1,238 xg for 2 minutes to remove air bubbles and collect the reaction mixture at the bottom of each well. The positive control was DNA from *C. difficile* isolate N10-2127. The negative control was sterile 18 MΩ ddH_2O. The PCR amplification was performed using a Verti 96-well Thermal Cycler (Applied Biosystems) with the appropriate cycling program (Table 4). Samples were analyzed immediately or transferred to a 4°C fridge for storage.

Table 4. *Clostridium difficile* ribotyping PCR assay cycling conditions

Cycles	Temperature (°C)	Time	Description
1	95	15 min	Initial denaturation
24	95	60 sec	Denaturation
	57	60 sec	Annealing
	72	60 sec	Extension
1	72	30 min	Final extension
1	4	∞	Storage

4.8.3. Capillary Gel Electrophoresis

A working solution for fragment analysis was made consisting of 8.5 µl of Hi-Di Formamide (Applied Biosystems) and 0.5 µl of GeneScan 1200 LIZ size standard (Applied Biosystems) for each sample. The mixture was vortexed and 9 µl aliquots were dispensed into each well of a sterile 96-well plate. A 1-µl aliquot of each amplimer from the PCR plate was added to the wells for a final volume of 10 µl. MicroAmp Optical Adhesive Film (Applied Biosystems) was applied to the top of the wells to seal the plate. Well contents were mixed by vortexing the plate. Using a plate rotor, the plate was centrifuged at 1,238 xg for 2 minutes to remove air bubbles and collect the reaction mixture at the bottom of each well. Samples were denatured at 95°C for 2 minutes in a Verti 96-well Thermal Cycler (Applied Biosystems) and snap-cooled on ice for 5 minutes. The plate was, again, centrifuged at 1,238 xg for 2 minutes. The plate was loaded into a 3130XL Genetic Analyzer (Applied Biosystems), configured for fragment analysis, that contained validated sequence hardware consisting of an ABI 3130xl/POP7/36 cm array (Applied Biosystems). Loading of the plate was done as per the instructions for the sequencer guidelines and plate layout was configured as required by the ABI software. Sample injection was set at 5kV for 5 seconds and run time was set for 103 minutes with a 6.5kV separation. The sequencer produced a *.fsa file for each sample.

4.8.4. Capillary Gel Electrophoresis Fragment Processing

All fsa files were processed using ABI GeneMapper software (Applied Biosystems) that was configured with the 68 fragment sizes of the GeneScan 1200 LIZ size standard, 20 bp through 1200 bp. The settings selected for the processing specified the GeneScan 1200 LIZ as the size standard, a 10% maximum peak threshold, and a sample peak size on the 5-FAM, blue channel. Any files that failed due to poor trace quality or ones that fell off scale were marked to

re-run the sample. The remaining peak files were exported as a combined file containing all trace data that passed analysis for each sample.

4.8.5. Capillary Gel Electrophoresis Fragment Analysis

The combined peak file was imported into BioNumerics software v5.10 (Applied Maths, Austin, TX) with the default settings for adding new fingerprint types. Only the 5-FAM blue-dye peaks were chosen for import for each sample. A synthetic gel image was created for each sample through the import process. The banding patterns on these images were compared to our standard ribotype banding patterns provided by the Leeds Reference Laboratory that were already imported into BioNumerics. Samples with banding patterns that matched to a known standard were recorded as having the same ribotype as the standard. Samples whose banding pattern did not match to a standard, and were confirmed by repeat testing, were designated as ns (no standard). Samples recorded with ns banding patterns were added to the standard list to keep track of all putative patterns for consideration in future ribotyping runs.

4.9. Toxin PCR Assays

4.9.1. PCR Mix Preparation

A 400-µl volume of BHI broth culture from the end of the isolation procedure was transferred to a sterile 1.5 ml tube, centrifuged at 16,000 xg for 2 minutes, and the supernatant was removed. The cell pellet was resuspended in 100 µl of Instagene Matrix (BioRad) and mixed by vortexing. Samples were heated in a heatblock at 99°C for 10 minutes to lyse the cells. Sample tubes were then centrifuged at 16,000 xg for 2 minutes and supernatants dispensed into a fresh, sterile 1.5 ml tube. DNA concentration was quantified using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific). Each DNA extraction was adjusted to a concentration of 100 ng/µl using molecular-grade water as a diluent.

4.9.2. PCR Reaction Preparation

Two multiplex PCR assays were used for this protocol (Table 5). This assay was run as per the instructions in the standard operating procedures at the National Microbiology Laboratory (*C. difficile* Multiplex PCRs I and II (*tcdA, tcdB, tcdC, tpi, cdtB*); Version 1.0). All primers were made in-house by the DNA Core facility at the National Microbiology Laboratory (NML).

Table 5. Toxin PCR primers

Name	Sequence (5'-3')	T_m (˚C)	Size (bp)	Gene	Description
cdtB-F1	TGGACAGGAAGAATAATTCCTTC	68.2	23	*cdtB*	binary toxin subunit B
cdtB-R1	TGCAACTAACGGATCTCTTGC	68.9	21		
tcdA-F	AGATTCCTATATTTACATGACAATAT	65.0	26	*tcdA*	toxin A
A3B	ACCATCAATCTCGAAAAGTCCAC	70.0	26		
tcdB-3	AATGCATTTTTGATAAACACATTG	63.6	24	*tcdB*	toxin B
tcdB-4	AAGTTTCTAACATCATTTCCAC	63.9	22		
Cdtpi-F	AAAGAAGCTACTAAGGGTACAAA	66.4	23	*tpi*	triose phosphate isomerasae
Cdtpi-R	CATAATATTGGGTCTATTCCTAC	66.4	23		
PaL15	TCTCTACAGCTATCCCTGGT3	68.2	20	*tcdC*	negative regulator of PaLoc
PaL16	AAAAATGAGGGTAACGAATTT	61.1	21		

Multiplex mastermixes were prepared according to Table 6.

Table 6. *Clostridium difficile* toxin PCR multiplex mastermix contents

Multiplex I		Multiplex II	
Component	**Volume (µL)**	**Component**	**Volume (µL)**
2X Master Mix (Qiagen)	12.5	2X Master Mix	12.5
cdtB F1/R1 (10 µM)	0.75	*tpi* (10 µM)	0.50
tcdA F/A3B (10 µM)	0.75	*Pal* 15/16 (10 µM)	0.50
tcdB 3+4 (10 µM)	0.50	-	-
H_2O (autoclaved)	8.50	H_2O (autoclaved)	9.50
DNA*	2.00	DNA*	2.00

*- DNA added at a later step

Keeping PCR master mix on ice, a 23-µl volume was added to each well of a sterile 96-well plate. A 2-µl aliquot of DNA lysate was then added for a final reaction volume of 25 µl.

MicroAmp Optical Adhesive Film was applied to the top of the wells to seal the plate. Well contents were mixed by vortexing the plate. Using a plate rotor, the plate was centrifuged at 1,238 xg for 2 minutes to remove air bubbles and collect the reaction mixture at the bottom of each well. Four controls were included on each toxin PCR assay plate (Table 7).

Table 7. Toxin gene profiles of positive control strains used in the toxin PCR assay

Control	**Description**
11ACD0028 (NAP7)	39 bp *tcdC* deletion (637 bp product) Toxin A positive (*tcdA*+) Toxin B positive (*tcdB*+) Binary toxin positive (*cdtB*+) Triose phosphate isomerase positive (*tpi*+)
11ACD0075 (NAP1)	*tcdC* deletion of 19 bp (657 bp product) Binary toxin positive (*cdtB*+) Toxin A positive (*tcdA*+) Toxin B positive (*tcdB*+) Triose phosphate isomerase positive (*tpi*+)
N07-01533 (toxin A-/B+)	No *tcdC* deletion (676 bp product) Toxin A positive (*tcdA*-) Toxin B positive (*tcdB*+) Binary toxin negative (*cdtB*-) Triose phosphate isomerase positive (*tpi*+)
ATCC 43255	No *tcdC* deletion (676 bp product) Toxin A positive (*tcdA*+) Toxin B positive (*tcdB*+) Binary toxin negative (*cdtB*-) Triose phosphate isomerase positive (*tpi*+)
Sterile ddH_2O as negative control	Negative control

Figure 5 displays the toxin PCR banding patterns of these four control strains.

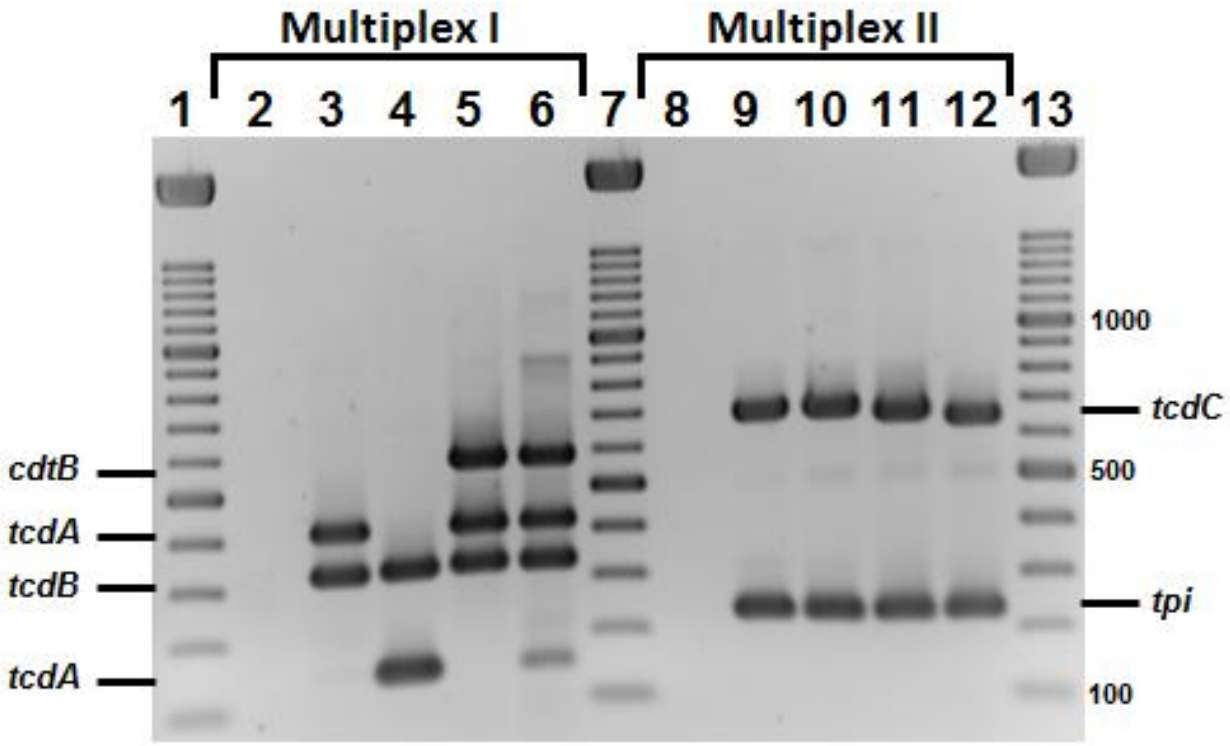

Multiplex I
Multiplex II
1 2 3 4 5 6 7 8 9 10 11 12 13
cdtB
tcdA
tcdB
tcdA
1000
tcdC
500
tpi
100

Figure 5. Toxin PCR banding profiles of the four *C. difficile* control strains by gel electrophoresis. Lanes 1, 7 and 13 are 100 bp ladders; Multiplex I: lane 2 is the negative control (water blank), lane 3 is the ATCC 43255 positive control (*cdtB*$^{-}$, *tcdA*$^{+}$, *tcdB*$^{+}$), lane 4 is the N07-1533 (*cdtB*$^{-}$, *tcdA*$^{-}$, *tcdB*$^{+}$) positive control, lane 5 is the 11ACD0075 positive control (*cdtB*$^{+}$, *tcdA*$^{+}$, *tcdB*$^{+}$), and lane 6 is the 11ACD0028 positive control (*cdtB*$^{+}$, *tcdA*$^{+}$, *tcdB*$^{+}$); Multiplex II: lane 8 is the negative control (water blank), lane 9 is the ATCC 43255 positive control (*tcdC*$^{+}$, *tpi*$^{+}$), lane 10 is the N07-1533 positive control (*tcdC*$^{+}$, *tpi*$^{+}$), lane 11 is the 11ACD0075 positive control (*tcdC*$^{\Delta 18\text{-}1}$, *tpi*$^{+}$), and lane 12 is the 11ACD0028 positive control (*tcdC*$^{\Delta 39}$, *tpi*$^{+}$).

4.9.3. PCR Amplification

The PCR amplification was performed using a Verti 96-well Thermal Cycler (Applied Biosystems) with the appropriate cycling program (Table 8). Samples were analyzed immediately or transferred to a 4°C fridge for storage.

Table 8. *Clostridium difficile* toxin PCR assay cycling conditions

Cycles	Temperature (°C)	Time	Description
1	95	15 min	Initial denaturation
30	94	30 sec	Denaturation
	57	90 sec	Annealing
	72	60 sec	Extension
1	72	7 min	Final extension
1	4	∞	Storage

The PCR amplicons were analyzed by submarine gel electrophoresis. 1% Agarose Gels were made by melting together 1 g of agarose (Sigma Aldrich, St. Louis, MO) with 100 ml of 1x TBE (10x stock; Sigma Aldrich). The melted agarose was slowly poured into the casting tray and allowed to cool at room temperature for 30 minutes. The gel was then loaded into the electrophoresis chamber containing 1x TBE (10x stock; Sigma Aldrich) up to the fill line. Gels were run for 60 minutes at 120 volts.

4.9.4. Staining and Analysis of Agarose Gel

Gels were submerged in a 0.5 μg/ml ethidium bromide staining solution (10 mg/ml stock; Sigma Aldrich) and placed on an Adjustable Tilt Rocker (Mandel Scientific, Guelph, ON) for at least 30 minutes. Stained gels were visualized using an AlphaImager HP (Alpha Innotech, San Leandro, CA) and photographs printed as physical documentation of the PCR assay.

4.10. Pulsed-field Gel Electrophoresis

4.10.1. Agarose Plug Preparation

This method was run as described in the standard operating procedures at the National Microbiology Laboratory (Pulsed Field Gel Electrophoresis of *Clostridium difficile*; Version 1.0). Two bead baths (Lab Armor, Cornelius, OR) were pre-heated to 55 ± 2°C and 37 ± 2°C, respectively. A 0.25 g aliquot of SeaKem Gold agarose (Lonza, Basel, Switzerland) was combined with 2.5 ml Sodium dodecyl sulfate (SDS, 10% stock; Sigma Aldrich) and 22.5 ml of 1x Tris-EDTA buffer (Sigma Aldrich) and placed in the 55°C water bath for at least 10 minutes. The control isolate for this protocol was *Salmonella enterica* serovar Branderup H9812. This isolate was grown up in 5 ml of BHI broth that was incubated at 37°C for 24 hours. The turbidity of each BHI broth culture to be tested was normalized to 0.33 using pre-reduced BHI broth as the diluent. Turbidity measurements were found using a Dade Microscan Turbidity Meter (Dade Behring, Inc., West Sacramento, CA). A turbidity of 0.33 equates to the turbidity of a 3.0 McFarland standard. A 400-µl volume of the normalized cultures was transferred to a sterile 1.5 ml tube, centrifuged at 21,130 xg for 1 minute, and the supernatants discarded. Each cell pellet was resuspended in 150 µl of Cell Lysis Buffer (CLB). CLB stock was made with 2.4 ml Tris-HCl pH 8.0 (1M stock; Sigma Aldrich), 58.80 ml NaCl (5M stock; Thermo Fisher Scientific Inc.), 80 ml ethlylenediaminetetraacetic acid disodium salt dehydrate pH 8.0 (EDTA, 0.5M stock; Sigma Aldrich), 8 ml sodium deoxycholate (10% stock; Sigma Aldrich), 10 ml N-lauroylsarcosine sodium salt (20% stock; Sigma Aldrich), 20 ml Brij 58 (10% stock; Sigma Aldrich), and 158 ml 18 MΩ ddH_2O. To each cell suspension, 150 µl of the melted SeaKem Gold agarose mixture was added and mixed gently by pipetting. The entire volume of each

agarose-cell suspension was transferred evenly into two wells of a disposable plug mold (BioRad). The plugs were allowed to solidify for 10 minutes at room temperature.

4.10.2 Cell Lysis of Agarose Plugs

For each sample from the plug preparation step, 500 µl of Cell Lysis Buffer was combined with 1 µl of RNAse (20 µg/ml stock; Sigma Aldrich), 50 µl of lysozyme (2.0 mg/ml stock; Sigma Aldrich), and 1.25 µl of mutanolysin (12.5 u/ml stock; Sigma Aldrich) in a sterile 1.5 ml tube, creating the lysis tube. The pair of solidified plugs of each sample were placed into a lysis tube and incubated at 37 ± 2°C overnight in a bead bath. Next, the lysis buffer solution was aspirated off of the plugs and discarded. To each tube, 495 µl of Proteinase K buffer and 5 µl of 20 mg/ml Proteinase K (50 mg/ml stock; Roche, Basel, Switzerland) were added. Proteinase K buffer stock was made from 399 ml ethlylenediaminetetraacetic acid disodium salt dehydrate pH 8.0 (EDTA, 0.5M stock; Sigma Aldrich) mixed with 4 g N-lauroylsarcosine sodium salt (Sigma Aldrich). Plug tubes were then transferred to the 55 ± 2°C bead bath to incubate for 2 hours.

4.10.3 Plug Washing

A stock of 1x TE buffer was made from 5 ml of Tris-HCl pH 8.0 (1M stock; Sigma Aldrich) and 0.1 ml of EDTA (Sigma Aldrich) per tube. The plugs were rinsed three times with 1.5 ml of room temperature 1x TE buffer. Next, the tubes were washed with 1.5 ml of 1x TE buffer in a Max Q 7000 rotator (Thermo Scientific) at intervals of 5, 10, 15, and 20 minutes.

4.10.4 Restriction Digest of DNA in Agarose Plugs

One plug from each sample was cut into thirds, using a sterile scalpel, on a PFGE casting plate that was pre-washed with 70% ethanol. One third of each sample and control was transferred to a separate, sterile 0.6 ml tube. Sample plug slices were equilibrated by adding 150 µl of 1x buffer A (10x stock; Roche). The positive control slice was equilibrated using 150 µl of 1x buffer H (10x stock; Roche). Equilibrations were left for 10 to 15 minutes at room temperature. The buffers were removed from the tubes by pipetting. A digestion solution was

made with 100 µl of 1x buffer A and 1 µl of *Sma*I enzyme (40U/ µl stock; Roche) per sample. A digestion solution for the control was made with 100 µl of 1x buffer H and 1 µl of *Xba*I enzyme (40U/ µl stock; Roche). Each digestion solution was mixed well by vortexing and 100 µl was added to the respective plug. The *Sma*I sample tubes were digested at 25°C for 2 hours in an Echo Therm Chilling Incubator (Torrey Pines Scientific, Carlsbad, CA). The *Xba*I control tube was digested at 37 ± 2°C in a bead bath for 2 hours.

4.10.5. PFGE Electrophoretic Run Conditions

With the black gel frame assembled into a CHEF DRIII chamber (BioRad) the chamber was filled with 100 ml of 10x Tris-Borate-EDTA (TBE, Sigma Aldrich) and 1.0 ml of 100mM Thiourea (Sigma Aldrich). The cooling module was set to 14°C. The buffer was allowed to cycle for at least 30 minutes.

4.10.6. Casting Agarose Gels and Loading Digested Plugs into Wells

1% Agarose gels were made by melting together 1.5 g of Pulsed Field Certified Agarose (BioRad) with 150 ml of 0.5x TBE (10x stock; Sigma Aldrich). With the teeth of the 15-well comb touching the bottom of the gel casting tray digested plug slices were loaded onto the bottom of the comb teeth. Gel slices were allowed to dry for 5 minutes. The positive control slices were placed on lanes 1, 5, 10, and 15. The 150 ml of melted agarose was slowly poured into the casting tray and allowed to cool for 30 to 45 minutes. The gel was then removed from the tray and loaded into the CHEF chamber to begin the gel electrophoresis. The details of the run conditions are shown in Table 9.

Table 9. Pulsed-field gel electrophoresis run conditions

CHEF DRIII:		**CHEF Mapper:**	
Initial A Switch time	1.0s	Initial A Switch time	1.0s
Final A Switch Time	40.0s	Final A Switch Time	40.0s
Voltage	6V/cm	Voltage	6V/cm
Included angle	120	Included angle	120
Start Ratio	1	Start Ratio	1
Run Time	*22hrs.	Run Time	*22hrs.

*- gels were run until positive controls were within one centimeter from the bottom of the gel.

4.10.7. Gel Staining

Gels were submerged in a 0.5 µg/ml ethidium bromide staining solution (10 mg/ml stock; Sigma Aldrich) to stain on an Adjustable Tilt Rocker (Mandel Scientific) for at least 30 minutes. Stained gels were then destained in a container filled with 1.0 L of 18 MΩ ddH_2O for 30 minutes to 3 hours. Gels were visualized in an AlphaImager HP (Alpha Innotech) and photographs printed as physical documentation of the PFGE run.

4.10.8 BioNumerics Analysis

PFGE banding patterns were analyzed using BioNumerics v5.10 (Applied Maths). Comparison settings included a 1.0% band tolerance and 1.0% optimization. Dendrogram cluster analysis was calculated using an unweighted pair group method with arithmetic mean (UPGMA) along with the Dice coefficient. Isolates were categorized into groups based on the number and size of PFGE bands which was compared to standard PFGE patterns in our database.

4.11. Selection Criteria for Isolate Sequencing

Of the 338 NAP4 isolates, 131 were selected for sequencing. Isolates were chosen to the best of our ability to represent an even distribution of NAP4 PFGE types 0023 and 0033 across the three regions of Canada (west, central, and eastern provinces) and all study years (2004-2011). The western region of Canada included the provinces of British Columbia, Alberta, Saskatchewan, and Manitoba. Central Canada included Ontario and Québec. The eastern region

of Canada encompassed the provinces of New Brunswick, Nova Scotia, Prince Edward Island, and Newfoundland.

4.12. MiSeq Sequencing

4.12.1. DNA Extraction

DNA extraction was carried out using a modified protocol from the MasterPureTM Complete DNA and RNA Purification Kit (Epicentre®, Madison, WI) to obtain purity and high yields. Each *C. difficile* isolate, from cryo-bead stock, was streaked for isolation on TSA + 5% sheep's blood agar and incubated at 37°C for 48 hours. One colony from the plate was transferred to 5ml of pre-reduced BHI broth and was allowed to grow overnight at 37°C in anaerobic conditions. One ml of broth culture was then dispensed into a sterile 1.5 mL tube and centrifuged at 21,000 xg for 2 minutes. The supernatant was aspirated off of the cell pellet. The pellet was resuspended in 150 µL of lysis solution [50 µL of 20 mg/ml lysozyme (Sigma-Aldrich, Oakville, ON), 10 µL of 5KU/ml mutanolysin (Sigma-Aldrich), 1 µL of 10 µg/ml RNAse A (Sigma-Aldrich), and 89 µL of 1x TE buffer (Sigma-Aldrich)]. The suspension was incubated overnight at 37°C in a Thermomixer R (Fisher Scientific Inc.) with gentle vortexing. 150 µL of Epicentre T&C Lysis Solution (Epicentre®) was then added to the tubes along with 10 µL of 20 mg/ml Proteinase K (Sigma-Aldrich) and vortexed. The tubes were then placed in a 65°C bead bath, vortexing every 5 minutes, until the contents became clear (10-15 minutes). The tubes were chilled on ice for 5 minutes. A 175-µL volume of Epicentre MPC Protein Precipitation Reagent (Epicentre®) was then added to the tubes and vortexed to mix. Precipitated cell debris was pelleted by centrifugation at 16,000 xg at 4°C for 10 minutes. The supernatant was transferred to a sterile 1.5 mL tube. To the aspirated supernatant, 2 µL of 10 µg/ml RNAse A (Sigma-Aldrich) was added. The tube was placed in a 37°C bead bath for 30 to 60 minutes.

4.12.2. Nucleic acid precipitation

A 500-µL volume of isopropanol (Sigma-Aldrich) was added to each tube and mixed by inversion 30 to 40 times. The tubes were then centrifuged at 16,000 xg for 10 minutes at 4°C. The isopropanol was removed by decanting and 1 ml of 70% ethanol (Sigma-Aldrich) was added to the tube. The tube was centrifuged at 16,000 xg for 10 minutes. The ethanol was removed by aspiration with a pipette. This ethanol wash was repeated twice more. DNA pellets were then allowed to dry at room temperature until the pellets became clear. The DNA pellets were resuspended in 35 µL of 1x Tris-HCl pH 8.0 (1M stock; Sigma Aldrich) overnight at room temperature. Prior to placing purified DNA at -40°C for storage, a Qubit® Fluorometer 2.0 (Invitrogen, Carlsbad, CA) was used to measure the quantity and quality of each sample as per the instructions in the User Manual. DNA extractions were submitted to the DNA Core facility at the National Microbiology Laboratory for genomic sequencing. Library construction was done using a Nextera XT Library Kit (Illumina). Genome sequencing was carried out with a MiSeq Sequencer (Illumina).

4.13. Pacific Biosciences Sequencing

4.13.1. DNA Extraction

DNA extraction was carried out using a modified protocol from the Metagenomic DNA Isolation Kit for Water (Epicentre®) to obtain purity and high yields. *C. difficile* isolates, from cryo-bead stock, were streaked for isolation on TSA + 5% sheep's blood plates and incubated at 37°C for 48 hours. One colony from the plate was transferred to 5 ml of pre-reduced BHI broth and was incubated overnight at 37°C under anaerobic conditions. Afterwards, 500 µL of broth culture was transferred to a sterile 1.5 mL tube and centrifuged at 14,000 xg for 2 minutes. The supernatant was discarded and the cell pellets were resuspended in 300 µL of 1x TE buffer. 2 µL

of Ready-Lyse Lysozyme Solution (Epicentre®) and 2 µL of 10 µg/ml RNAse A (20 µg/ml stock; Sigma Aldrich) were added. The mixtures were gently vortexed and incubated at 37°C in a Thermomixer R (Fisher Scientific Inc.). 300 µL of 2x Meta-Lysis Solution (Epicentre®) and 3 µL of 20 mg/ml Proteinase K (50 mg/ml stock; Roche) were then added. The tubes were mixed by vortexing and incubated in an Isotemp 205 bead bath (Fisher Scientific Inc.) set to 65°C for 15 minutes. Tubes were cooled at room temperature for 5 minutes and then chilled on ice for an additional 5 minutes. 350 µL of MPC Protein Precipitation Reagent (Epicentre®) was added. Tubes were then mixed by vortexing for 10 seconds each. Cell debris was pelleted by centrifuging each tube at 14,000 xg for 10 minutes at 4°C. Supernatants were each transferred by pipetting to a sterile 1.5 ml tube. To the supernatants, 4 µL of RNAse A (20 µg/ml stock; Sigma Aldrich) was added and tubes were incubated in a 37°C bead bath for 5 minutes.

4.13.2. Nucleic acid precipitation

To precipitate the DNA, 500 µL of room temperature isopropanol (Sigma-Aldrich) was added to all DNA extractions. All tubes were mixed by inversion 30 to 40 times. Tubes were then centrifuged at 15,000 xg for 30 minutes at 4°C. Supernatants were then decanted and the pellets washed with 1 ml of 70% ethanol (Sigma-Aldrich). Wash tubes were centrifuged at 15,000 xg for 10 minutes at 4°C. The supernatants were decanted and the pellets were air dried for 20 minutes or until they became clear. The DNA pellets were then resuspended in 50 µL of Tris-HCl pH 8.0 (1M stock; Sigma Aldrich) overnight at room temperature. Génome Québec requires a minimum of 10 µg of DNA in volumes not exceeding 100 µl. To achieve this, three DNA extractions from each *C. difficile* isolate were combined into a single, sterile 1.5 ml tube and subject to DNA precipitation. Prior to placing purified DNA at -40°C for storage, a Qubit® Fluorometer 2.0 (Invitrogen) was used to measure the quantity and quality of each

sample according to the instructions in the User Manual. Also, as per the Génome Québec Pacific Biosciences submission user guide, the OD^{260}/OD^{280} ratio of each DNA extraction was checked to make sure the measurement was between 1.8 and 2.0. This was done using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific). Two DNA extractions were submitted to Génome Québec for sequencing using the Pacific Biosciences platform. One extraction was from a PFGE type 0023 isolate and the other extraction was from a PFGE type 0033 isolate. Library construction utilized a SMRT Template Prep Kit 1.0 (Pacific Biosciences) and DNA/Polymerase Binding Kit P4 (Pacific Biosciences). Genome sequencing was run on a PacBio RS II (Pacific Biosciences) sequencer using SMRT Cell 8-Pac v3 (Pacific Biosciences). Quality of the sequencing was confirmed by analyzing the two sequence files using Freebayes version 0.9.8 [40] to compare the base calls of Pacific Biosciences to that of MiSeq for same genome. This tool revealed the position and characteristics of single-nucleotide polymorphisms (SNPs) throughout the genome. All positions with SNP coverage greater than 50% of all base calls at a position in the genome exhibiting higher than 10x coverage at that position were investigated. The location of these SNPs was visualized in Tablet version 1.14.10.21 [97] using the pileup of MiSeq sequence against the Pacific Biosciences sequence from the same isolate. These SNPs of interest were all found in repeat regions or homopolyer regions and deemed not impactful. One end of the linearized Pacific Biosciences sequence file was confirmed to be a duplication of the beginning of the genome by PCR amplification. The primers used for this confirmation are described in Table 10. This duplicated sequence was removed from the sequence file in a text editor.

Table 10. Primers used to confirm Pacific Biosciences sequence duplication

Name	Sequence (5'-3')	Size (bp)
Pac1R	CAATTAACGGGACATGGAGG	20
Pac2F	ACGGCTCTATAGTTGCAGCT	20
Pac3F	GCTACTACTGATGGAGTCTG	20

Mastermixes were prepared as described in Table 11.

Table 11. PCR mastermix contents for confirmation of Pacific Biosciences sequence duplication

Reaction #1		Reaction #2	
Component	**Volume (µL)**	**Component**	**Volume (µL)**
2X HotStart Taq Mastermix (Qiagen)	25	2X HotStart Taq Mastermix (Qiagen)	25
Pac2F (10 µM)	1.25	Pac3F (10 µM)	1.25
Pac1R (10 µM)	1.25	Pac1R (10 µM)	1.25
H_2O (autoclaved)	20.5	H_2O (autoclaved)	20.5
DNA*	2.00	DNA*	2.00

The PCR amplifications were performed using a Verti 96-well Thermal Cycler (Applied Biosystems) with the appropriate cycling program (Table 12). Samples were analyzed immediately or transferred to a 4°C fridge for storage.

Table 12. PCR assay cycling conditions for confirmation of Pacific Biosciences sequence duplication

Cycles	Temperature (°C)	Time	Description
1	95	15 min	Initial denaturation
35	95	30 sec	Denaturation
	51	40 sec	Annealing
	72	70 sec	Extension
1	72	5 min	Final extension
1	4	∞	Storage

4.14. Constructing the Core SNP Phylogeny

At this point further isolate selection was implemented. Of the 131 genomes sequenced, ones that did not meet the pre-determined cut-off value for inclusion in the core SNP analysis were not included in further genomic investigations. This cut-off required at least 80% of the nucleotides in each genome being mapped to be found in the core sequence defined by the reference genome. As well, genomes were removed due to poor sequence quality or coverage issues that were related to a coverage bias associated with Nextera XT Sample Preparation Kits for library construction for MiSeq sequencing [137]. After this secondary selection, 85 *C. difficile* genomes remained for further analysis. Genomic sequence data from 48 PFGE type 0033 *C. difficile* isolates and 37 PFGE type 0023 *C. difficile* isolates were processed by the Core SNP Genomics Pipeline [106]. This pipeline was an in-house workflow created by the Bioinformatics group at the National Microbiology Laboratory. The pipeline first mapped the raw sequence data of each isolate to a reference using SMALT version 0.7.6 [107]. The reference sequence was the genome of a PFGE type 0023 *C. difficile* isolate that was closed by Pacific Biosciences sequencing included in this study. The pipeline then used FreeBayes version 0.9.8 [40] to detect single-nucleotide variations (SNVs) in the mapped sequences. SAMtools version 1.0 [83, 84] was used to determine depth of coverage over the entirety of every mapped genome. Next, SAMtools was used to align SNVs to a pseudo-alignment consensus sequence made from the most common base call at every position in the genome. All SNVs at an invalid position were removed. Invalid positions consisted of regions not considered part of the core genome. These regions were found in the reference genome using PHAST [152] for phages, IslandViewer [27] for genomic islands, and nucmer version 3.1 [73] for repeat regions. Finally, the phylogenetic

tree comparing the core genome of all isolates was created using PhyML [50] and analyzed using FigTree [79].

4.15. eBurst Analysis

The relationship between the SNPs described in the SNP matrix from the Core SNP Genomics Pipeline run mentioned previously and data for all 85 isolates were analyzed using the global optimal eBURST (goeburst) algorithm. The product of this analysis was a minimum spanning tree. The goeBURST algorithm was included as part of the Phyloviz software version 1.1 [117].

4.16. Antibiotic Resistance Evaluations

Antibiogram data of all 85 isolates were analyzed. Isolates collected between 2007 and 2011 had their antimicrobial susceptibilities determined including clindamycin, metronidazole, moxifloxacin, rifampin, tigecycline, and vancomycin using Etest strips. MIC breakpoints for each antimicrobial tested were defined by the Clinical and Laboratory Standards Institute (CLSI). Antimicrobial susceptibility data for rifampin and tigecycline was not collected for isolates from the 2004/2005 surveillance year. Every isolate's antimicrobial susceptibility profiles were translated into a data matrix to be combined with the core SNP phylogenetic tree. A score of 0.0 was for no recorded data, 0.5 for sensitivity to an antimicrobial, 1.0 for intermediate resistance, and 1.5 for resistance. This data matrix was used along with the SNP matrix created from the Core SNP Genomics Pipeline to create a heatmap of antimicrobial resistance profiles in RStudio version 0.99.467 [37]. The commands to construct the heatmap are described in Section 1 of the Appendix. Clindamycin and moxifloxacin were the only antimicrobials which showed variation of resistance profiles among all 85 *C. difficile* isolates. The presence of antimicrobial resistance genes related to these antimicrobials was investigated

further. First, MiSeq sequence data from all isolates was assembled using SPAdes assembly software version 3.6.2 [5]. Assembled draft genomes were subsequently annotated using Prokka version 1.10 [22, 125]. All annotation files were searched for the presence of the rRNA adenine N-6-methyltransferase ErmB (clindamycin resistance) and DNA gyrases GyrA and GyrB (moxifloxacin resistance). The Megalign tool version 11.2.1 [13], part of the DNAStar software, was used to find amino acid substitutions among protein annotations containing these genes. ResFinder version 2.1 [149] was also used to search all draft genomes for the presence of antimicrobial resistance genes.

4.17. Nucleic Acid and Protein Profile Characterization

Lengths of unique nucleic acid sequence in all draft genomes were investigated using the default settings of Neptune version 1.2.2 [91]. Draft genomes from both PFGE type 0033 isolates and PFGE type 0023 isolates were used in the exclusion group and inclusion group in separate runs of the software.

4.18. PFGE Pattern Differentiation

To determine the difference between the PFGE banding patterns of PFGE type 0023 and PFGE type 0033 an in silico digestion algorithm built-in to the Seqbuilder tool version 11.2.9 [13], part of the DNAStar software, was used. The PFGE digestion was carried out with *Sma*I on the closed reference genome from the Core SNP Genomics Pipeline run mentioned previously (PFGE type 0023). Sequence fragments smaller than 10 kb were discarded. In silico digestion fragment sizes were compared to the approximate band sizes of the PFGE pattern of isolates with PFGE type 0023 and 0033 in our BioNumerics database. The in silico PFGE digestion requires closed genomes. No closed genomes with PFGE type 0033 were available to us. To complete the comparison of the digestion patterns, a draft genome from a PFGE type 0033 isolate was aligned

to the PFGE type 0023 reference genome using Mauve version 2.3.1 [23]. The locations of the *Sma*I cut sites that created the major fragments in the in silico digestion were overlaid onto the aligned PFGE type 0033 genome to determine what genetic content created the PFGE type 0033 band shift.

4.18.1. Phage Analysis

All 85 *C. difficile* genomes were run with PHAST [152] to find putative phages. Only high-scoring phage hits, defined as intact phages by PHAST, were considered for further processing. The genomic sequences of intact phages found by PHAST were excised on a Linux command line as per the instructions outlined in Section 2 of the Appendix. All extracted intact phage sequences were annotated using Prokka version 1.10 [22, 125]. The orthologs in these phage annotations were found using OrthoMCL clustering software version 2.0.9 [16, 17, 85]. The presence or absence of the orthologs found in each phage from the OrthoMCL output was translated into a binary table using a python script [41]. RStudio was used to transform this binary table into an ORF tree that differentiates phage genomes based on the presence or absence of orthologous groups of proteins found by OrthoMCL. The instructions for creating this tree are described in Section 3 of the Appendix.

4.18.2. Intact Phage Location in *Clostridium difficile* Genomes

The instructions from the end of the in silico phage extraction guide, described in Section 2 of the Appendix, was used to extract intact phages plus 1,000 bp of flanking *C. difficile* genomic sequence. Instructions from the guide were also used to remove the phage sequence from between these flanking sequences. This left 1,000 bp of flanking *C. difficile* genomic sequences which were concatenated together. The resulting 2,000 bp sequences around each phage were used as queries for a blastX search [45] to determine if genes were disrupted by the

phage insertion. The contig(s) from each genome in this database search that had a high-scoring hit to the representative flanking sequence were aligned in Mauve. This Mauve alignment was used as a visual confirmation of the blastN results as to the presence or absence of a phage insertion.

5. Results

5.1. Molecular Epidemiology of *Clostridium difficile* in Canada

Since 2004 there have been significant changes in the molecular epidemiology of *C. difficile* in Canada (Figure 6). Overall, NAP1 has been the most prevalent strain type, reaching a maximum in 2008, but has been declining ever since. Between 2004-2007 NAP2 was the second most prevalent strain type, but has since decreased substantially, representing only 4.2% of all *C. difficile* in 2011. From 2007-2011 NAP4 has been emerging and beginning in 2010 has become the 2nd most prevalent strain type in Canada.

Compared to the relatively well-studied NAP1 strain, little is known about NAP4 in Canada. This study sets out to fully characterize all NAP4 isolates in Canada through molecular, phenotypic, and epidemiological comparisons with other *C. difficile* strain types, in particular the hypervirulent NAP1, in order to gain a better understanding of this emerging strain type in Canada.

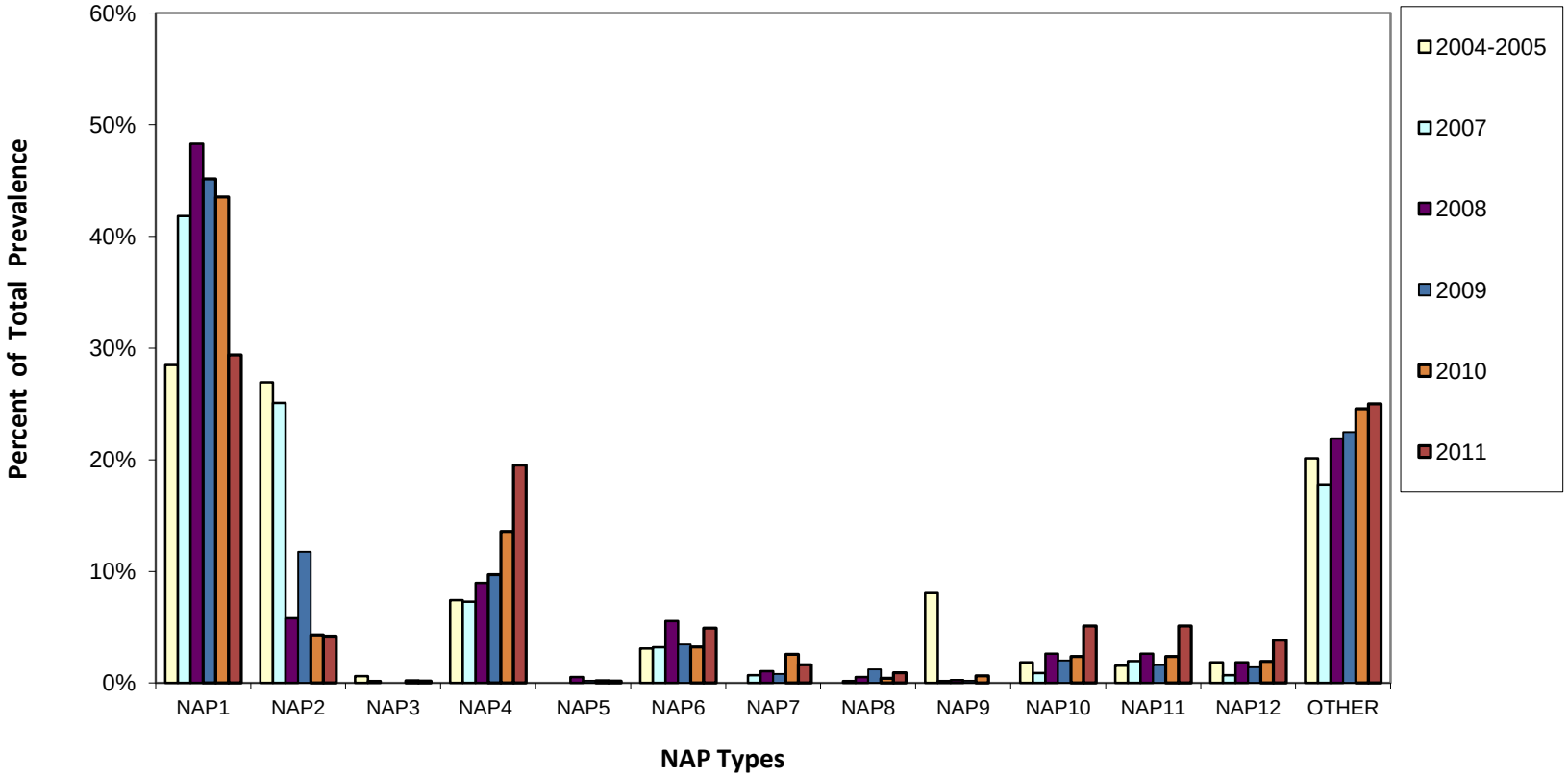

60%
50%
40%
30%
20%
10%
0%
Percent of Total Prevalence
NAP1
NAP2
NAP3
NAP4
NAP5
NAP6
NAP7
NAP8
NAP9
NAP10
NAP11
NAP12
OTHER
NAP Types
2004-2005
2007
2008
2009
2010
2011

Figure 6. Prevalence of NAP Types in Canada from 2004 to 2011.

(unpublished data from the Canadian Nosocomial Infection Surveillance Program)

5.2. Growth Assay

In order to measure any potential differences in fitness, 24-hour growth curve assays were performed on 19 *C. difficile* isolates (Figure 7). This pool of isolates consisted of one or two representatives of each of the 12 *C. difficile* NAP types with the main focus being comparison of the growth rate of NAP4 to that of the other NAP types. Entry into log-phase appeared to occur around the 3-4 hour time points. Mid-log phase was achieved approximately 5-6 hours post-inoculation. Stationary phase growth was reached at approximately 8-9 hours post-inoculation. Between 4 and 8 hours the growth rate of NAP4 was not found to be significantly greater than any of the other NAP types, as determined by comparing the slopes on the growth curves. It should be noted that NAP9 experienced a 1 hour lag to reach exponential phase in both runs. This may be due to the effect of cryo-storage on this specific isolate.

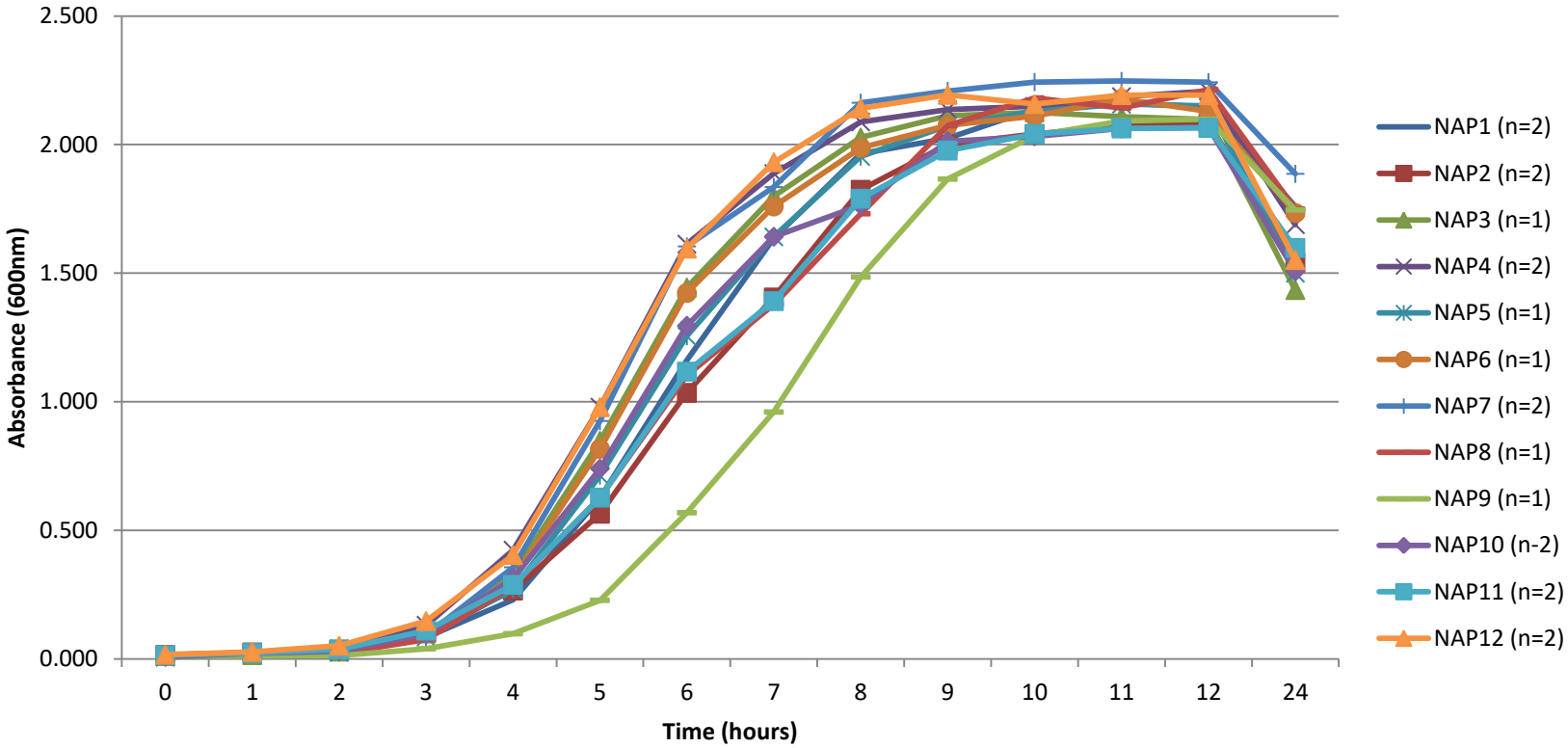

2.500
2.000
1.500
1.000
0.500
0.000
Absorbance (600nm)
0
1
2
3
4
5
6
7
8
9
10
11
12
24
Time (hours)
NAP1 (n=2)
NAP2 (n=2)
NAP3 (n=1)
NAP4 (n=2)
NAP5 (n=1)
NAP6 (n=1)
NAP7 (n=2)
NAP8 (n=1)
NAP9 (n=1)
NAP10 (n-2)
NAP11 (n=2)
NAP12 (n=2)

Figure 7. Growth curves generated from the average absorbance readings of isolates representing all currently known NAP types of *Clostridium difficile*. All isolates were tested in duplicate with each separate run including a set of technical replicates of each isolate. Growth was recorded by measuring the absorbance (OD_{600}) of the liquid cultures over time.

5.3. Sporulation Assay

To evaluate the sporulation ratio of NAP4 isolates compared to that of isolates from other NAP types sporulation assays were performed on 19 *C. difficile* isolates. The average sporulation ratio among all NAP types was 46% with a standard deviation of 26%. Ratios that fell one standard deviation outside of the average ratio were considered outliers and were repeated for confirmation. The lowest sporulation ratio observed was 6.1%, from a NAP2 isolate, and the highest was at 93.6%, from a NAP3 isolate (Figure 8). No significant differences in sporulation ratios were observed between NAP4 and the other NAP types in this experiment. Several NAP groups demonstrated a large variation in sporulation ratios which might simply be due to the limited amount of isolates in those groups. A larger isolate pool may reduce this variation.

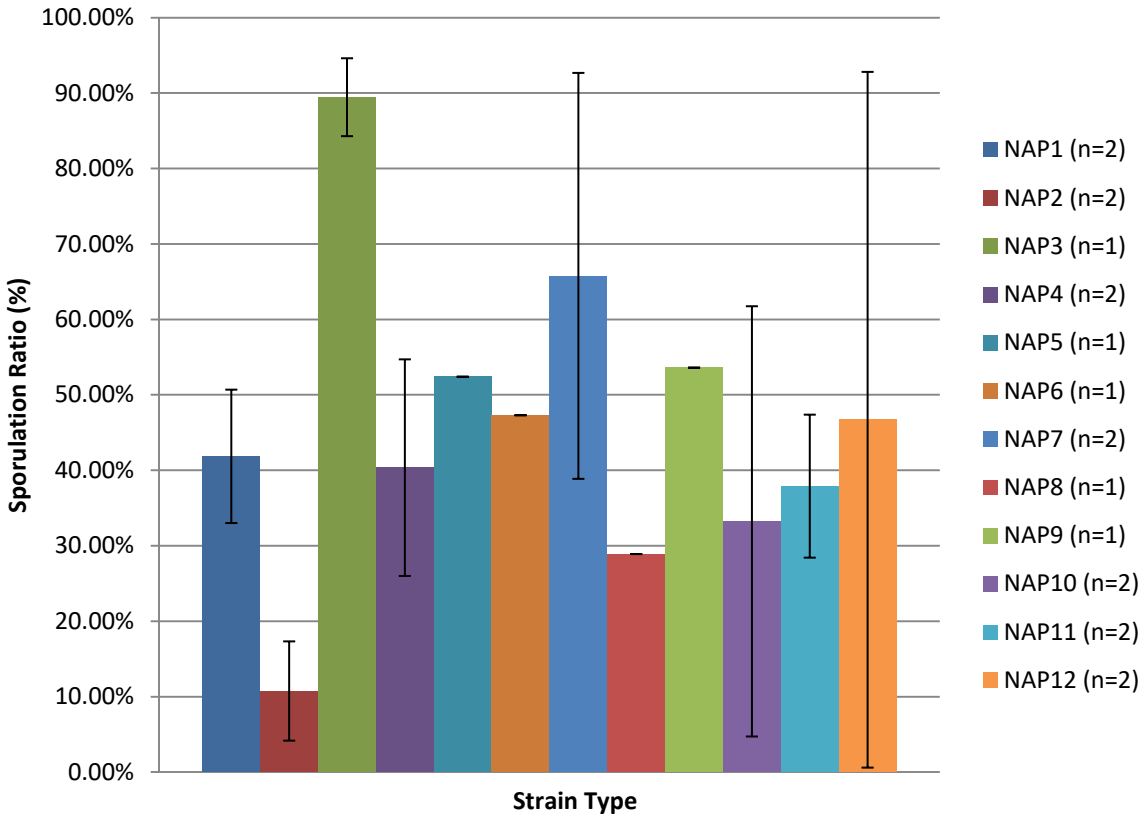

100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
Sporulation Ratio (%)
Strain Type
NAP1 (n=2)
NAP2 (n=2)
NAP3 (n=1)
NAP4 (n=2)
NAP5 (n=1)
NAP6 (n=1)
NAP7 (n=2)
NAP8 (n=1)
NAP9 (n=1)
NAP10 (n=2)
NAP11 (n=2)
NAP12 (n=2)

Figure 8. Average sporulation ratios of *Clostridium difficile* isolates representing each NAP type. Sporulation ratios are shown with standard deviation error bars depicting ratio variation between isolates within each group.

5.4. Epidemiology

Patient data was collected between 2004 and 2011 for 1,387 NAP1 and 338 NAP4 cases of CDI by members of CNISP as previously defined. From 338 cases of CDI with NAP4 only relatively a few would be indicated as severe as <1% of patients required ICU admission, colectomy, or died as a direct result of infection. For patients with NAP4 infections, gender was evenly distributed, the average age was 53 years, and cases were distributed across Canada. 37% of NAP4 isolates were resistant to clindamycin and only 6.5% were resistant to moxifloxacin. No resistance was observed to metronidazole, vancomycin or tigecycline. In comparison to NAP4, a significantly higher proportion of NAP1 cases resulted in colectomy and ICU admission as well as both 30 day all-cause mortality and attributable mortality (Table 13). Infections of either male or female patients were independent of NAP1 or NAP4 isolates. Interestingly, NAP4 isolates were shown to infect a significantly larger proportion of patients between ages 1 to 18 years of age as well as 19 to 64 years of age than NAP1 isolates. However, NAP1 isolates were shown to infect a significantly larger proportion of patients of age 65 years old or older than NAP4 isolates. NAP1 isolates were found to be collected in significantly higher proportions than NAP4 isolates in central and western Canada. Finally, we found NAP1 to have a significantly higher proportion of isolates with resistance to clindamycin, rifampin, and moxifloxacin than that of NAP4 isolates.

Table 13. Summary of patient data from NAP1 and NAP4 isolates collected between 2004 and 2011

	NAP1	NAP4	
Total Cases	1387	338	P-value
Morbidity & Mortality			
Cases Requiring Colectomy	24 (1.7%)	1 (0.3%)	0.04
Cases Requiring ICU Admission	46 (3.3%)	2 (0.6%)	<0.01
30-day All-Cause Mortality	235 (16.9%)	27 (7.9%)	<0.01
30-day Attributable Mortality	113 (8.1%)	3 (0.9%)	<0.01
Gender*			
Male	664 (47.8%)	166 (49%)	0.92
Female	672 (48.4%)	166 (49%)	0.92
Age**			
Average Age	71	53	<0.01
≤ 18	38 (2.7%)	78 (23.0%)	<0.01
19-64	332 (23.9%)	105 (31.0%)	0.01
65+	970 (69.9%)	149 (44.0%)	<0.01
Location***			
West	548 (27.2%)	260 (12.9%)	<0.01
Central	1375 (44.3%)	312 (10.0%)	<0.01
East	58 (16.5%)	50 (14.2%)	0.40
Antimicrobial Resistance****			
Metronidazole			
Resistant Isolates	0 (0.0%)	0 (0.0%)	1.00
MIC50	1.00	0.25	
MIC90	2.00	0.50	
MIC Range	0.047 - 8.00	0.023 - 2.00	
Clindamycin			
Resistant Isolates	671 (48.4%)	125 (37.0%)	<0.01
MIC50	4.00	4.00	
MIC90	16.00	8.00	
MIC Range	0.047 - >256	0.016 - >256	
Vancomycin			
Resistant Isolates	0 (0.0%)	0 (0.0%)	1.00
MIC50	0.75	0.75	
MIC90	2.00	1.00	
MIC Range	0.094 - 4.00	0.38 - 3.00	
Rifampin			
Resistant Isolates	35 (2.5%)	1 (0.3%)	<0.01
MIC50	0.002	<0.002	
MIC90	0.003	0.003	
MIC Range	0.002 - >32	<0.002 - >32	
Moxifloxacin			

Resistant Isolates	1234 (88.9%)	22 (6.5%)	<0.01
MIC50	>32	1.50	
MIC90	>32	2.00	
MIC Range	0.50 - >32	<0.002 - >32	
Tigecycline			
Resistant Isolates	0 (0.0%)	0 (0.0%)	1.00
MIC50	0.094	0.125	
MIC90	0.125	0.190	
MIC Range	0.023 - 0.38	0.047 - 0.38	

***-** 51 NAP1 isolates and 6 NAP4 isolates did not have patient gender recorded

** - 47 NAP1 isolates and 6 NAP4 isolates did not have patient age recorded

***- isolate counts represent proportion of all *C. difficile* isolates collected in that region

****- all MIC50, MIC90, and MIC range units are measured in μg/ml

5.5. Toxin PCR

For NAP4, all 338 isolates tested positive for *tcdA* and *tcdB*, whereas only 1 isolate was found to carry the binary toxin *cdtB* (Table 14). 3 NAP4 isolates were also found to carry a deletion in the *tcdC* regulatory gene. Toxin PCR for these isolates was not repeated. Patients infected with these 4 NAP4 isolates all survived their respective infections. In comparison to NAP4, NAP1 was found to have a significantly higher proportion of isolates with a *tcdC* deletions as well as *cdtB*.

Table 14. Toxin PCR results of NAP1 and NAP4 isolates

		NAP1	NAP4
Total Cases		**1387**	**338**
tcdA			
	Positive	1387 (100.0%)	338(100.0%)
	Negative	0 (0%)	0 (0.0%)
tcdB			
	Positive	1387 (100.0%)	338 (100.0%)
	Negative	0 (0.0%)	0 (0.0%)
tcdC			
	Positive	27 (1.9%)	335 (99.1%)
	Positive with Deletion	1360 (98.1%)	3 (0.9%)
cdtB			
	Positive	1379 (99.4%)	1 (0.3%)
	Negative	8 (0.6%)	337 (99.7%)
tpi			
	Positive	1387 (100.0%)	338 (100.0%)
	Negative	0 (0.0%)	0 (0.0%)

5.6. PFGE 0023 vs 0033

5.6.1. Pulsed-field Gel Electrophoresis

All *C. difficile* strains were characterized by PFGE. From the 338 NAP4 isolates analyzed, 28 distinct PFGE types were observed (Figure 9). Two clades of PFGE patterns are distinguished, clade A (n=146) and clade B (n= 192). The PFGE pattern with the highest

representation from clades A and B are PFGE type 0033 (n=98, 67.1%) and PFGE type 0023 (n=88, 46.8%), respectively.

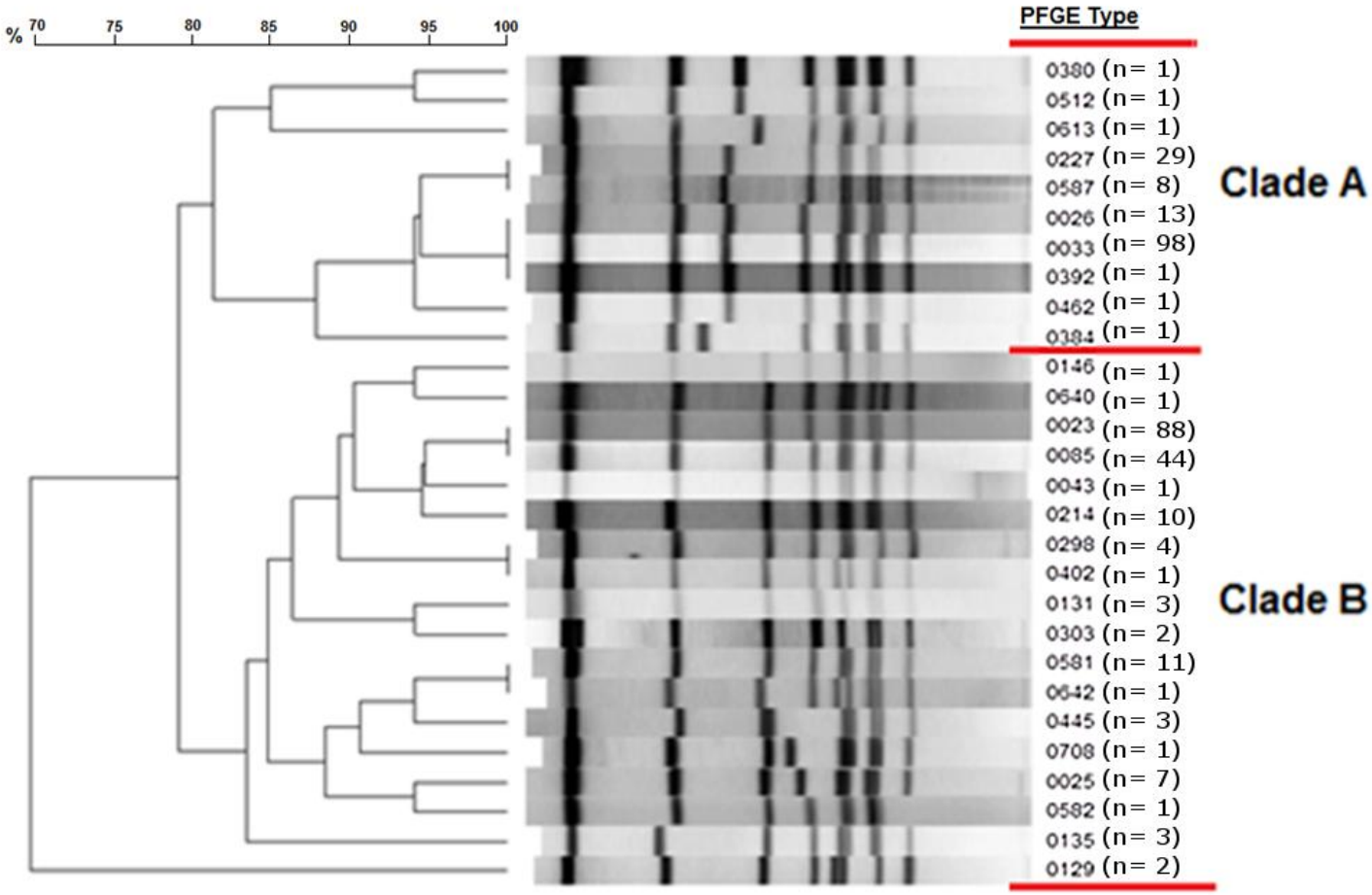
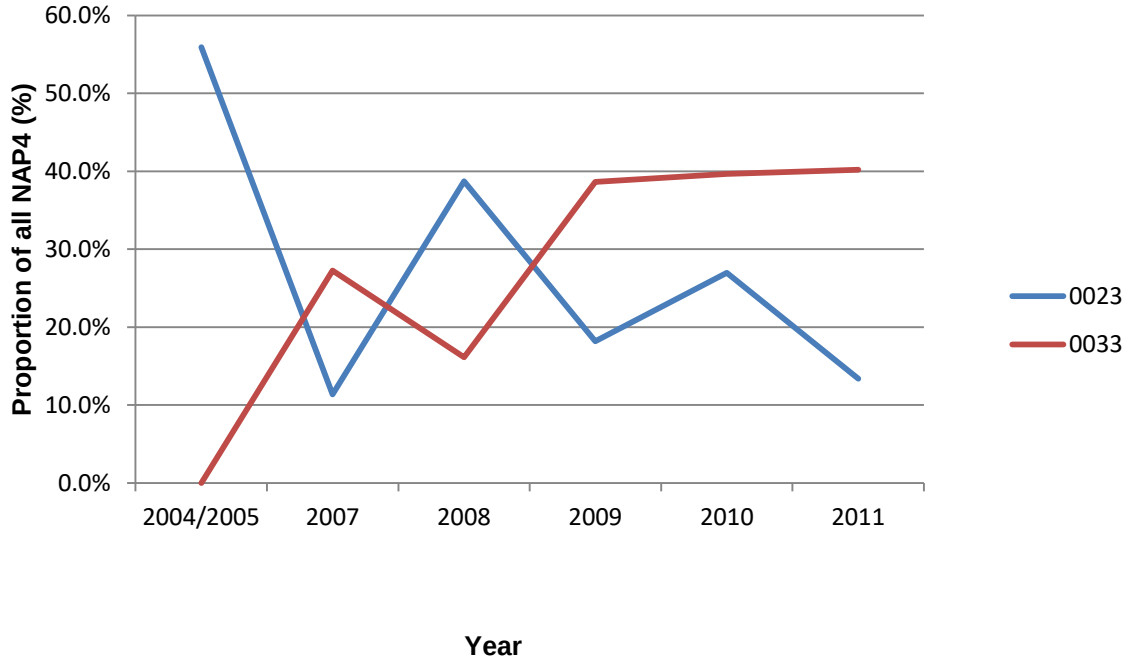

%
70
75
80
85
90
95
100
PFGE Type
0380 (n= 1)
0512 (n= 1)
0513 (n= 1)
0227 (n= 29)
0587 (n= 8)
0026 (n= 13)
0033 (n= 98)
0392 (n= 1)
0462 (n= 1)
0384 (n= 1)
Clade A
0146 (n= 1)
0640 (n= 1)
0023 (n= 88)
0085 (n= 44)
0043 (n= 1)
0214 (n= 10)
0298 (n= 4)
0402 (n= 1)
0131 (n= 3)
0303 (n= 2)
0581 (n= 11)
0642 (n= 1)
0445 (n= 3)
0708 (n= 1)
0025 (n= 7)
0582 (n= 1)
0135 (n= 3)
0129 (n= 2)
Clade B

Figure 9. A dendrogram comparing the PFGE banding patterns of the 28 unique PFGE types found in our NAP4 isolates. Percent similarity between the branches is shown at the top of the figure. PFGE types are displayed to the right of each banding pattern.

The prevalence of these two PFGE types has significantly changed over time (Figure 10). PFGE type 0023 was the most prevalent type of NAP4 in 2004/2005 representing 56% of all NAP4 *C. difficile* isolates collected in that that period. The prevalence of this PFGE type showed a sharp decline in 2007 and represented only 13% of all NAP4 types by 2011. In contrast to this, PFGE type 0033 was not observed in the 2004/2005 study year. By 2009, PFGE type 0033 had steadily become the most prevalent PFGE type among NAP4 *C. difficile* isolates collected during this study period.

a)


60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Proportion of all NAP4 (%)
2004/2005
2007
2008
2009
2010
2011
Year
0023
0033

b)

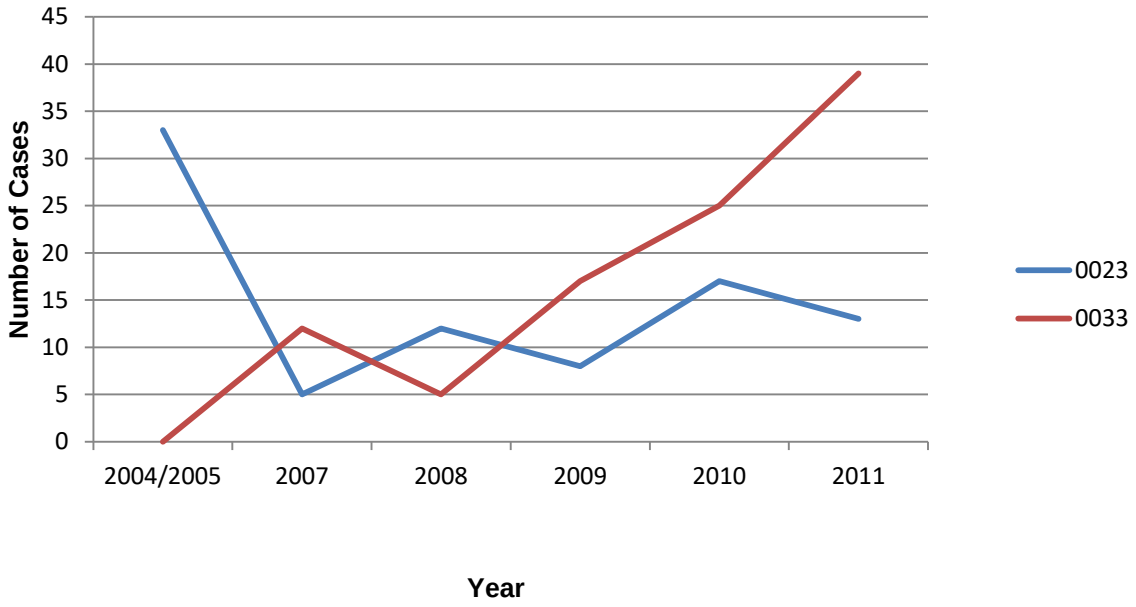

45
40
35
30
25
20
15
10
5
0
Number of Cases
2004/2005
2007
2008
2009
2010
2011
Year
0023
0033

Figure 10. Prevalence of NAP4 PFGE type 0023 and 0033 in Canada between 2004 and 2011. a) Total proportion of all NAP4 isolates collected that were PFGE types 0023 or 0033 collected over the study years. b) Total amount of NAP4 isolates collected that were either PFGE type 0023 or 0033 over the study years.

5.6.2. Ribotyping

All 338 NAP4 *C. difficile* isolates were typed by ribotyping. 25 unique ribotype banding patterns were observed. From this set, 6 ribotype patterns did not match a known standard and were designated a No Standard (ns) ribotype. All ns ribotypes were observed once and were confirmed through repeat testing.

Ribotype 020 was the most common ribotype accounting for 42% of all NAP4 isolates (Table 15). The second most common ribotype observed, ribotype 014, made up 29% of all NAP4 isolates. Other common ribotypes included ribotype 076 and 207 which accounted for 6% and 5% of all NAP4 isolates, respectively. All other ribotypes observed made up less than 3% of all NAP4 isolates tested.

Table 15. Proportions of ribotypes observed in NAP4 isolates

	All Isolates	PFGE Type 0023	PFGE Type 0033
Total Isolates	**338**	**85**	**98**
Ribotype			
011	4 (1.1%)	1 (1.1%)	0 (0.0%)
014	99 (29.2%)	15 (17.6%)	19 (19.3%)
020	142 (42.0%)	53 (62.3%)	52 (53.0%)
064	2 (0.5%)	1 (1.1%)	0 (0.0%)
076	20 (5.9%)	4 (4.7%)	9 (9.1%)
077	5 (1.4%)	0 (0.0%)	2 (2.0%)
106	1 (0.3%)	1 (1.1%)	0 (0.0%)
154	1 (0.3%)	0 (0.0%)	0 (0.0%)
207	17 (5.0%)	1 (1.1%)	0 (0.0%)
216	1 (0.3%)	0 (0.0%)	0 (0.0%)
221	8 (2.3%)	1 (1.1%)	4 (4.0%)
296	1 (0.3%)	0 (0.0%)	0 (0.0%)
325	5 (1.4%)	0 (0.0%)	1 (1.0%)
341	2 (0.5%)	0 (0.0%)	2 (2.0%)
354	2 (0.5%)	0 (0.0%)	1 (1.0%)
503	1 (0.3%)	0 (0.0%)	1 (1.0%)
511	9 (2.6%)	3 (3.5%)	3 (3.0%)
530	1 (0.3%)	0 (0.0%)	0 (0.0%)
629	7 (2.0%)	4 (4.7%)	1 (1.0%)
ns11	1 (0.3%)	1 (1.1%)	0 (0.0%)
ns12	1 (0.3%)	0 (0.0%)	1 (1.0%)
ns21	1 (0.3%)	0 (0.0%)	0 (0.0%)
ns23	1 (0.3%)	0 (0.0%)	0 (0.0%)
ns56	1 (0.3%)	0 (0.0%)	1 (1.0%)
ns70	1 (0.3%)	0 (0.0%)	1 (1.0%)

A dendrogram comparing the ribotype banding patterns of all 338 isolates (Figure 11) revealed two different ribotype 014 patterns which were confirmed to be the same ribotype (pers. commun. Dr. Warren Fawley at the Leeds Institute of Biomedical & Clinical Sciences). Ribotypes did not correlate with PFGE types and there was no significant difference in the proportion of ribotypes 014 or 020 found among the PFGE type 0023 or 0033 isolates.

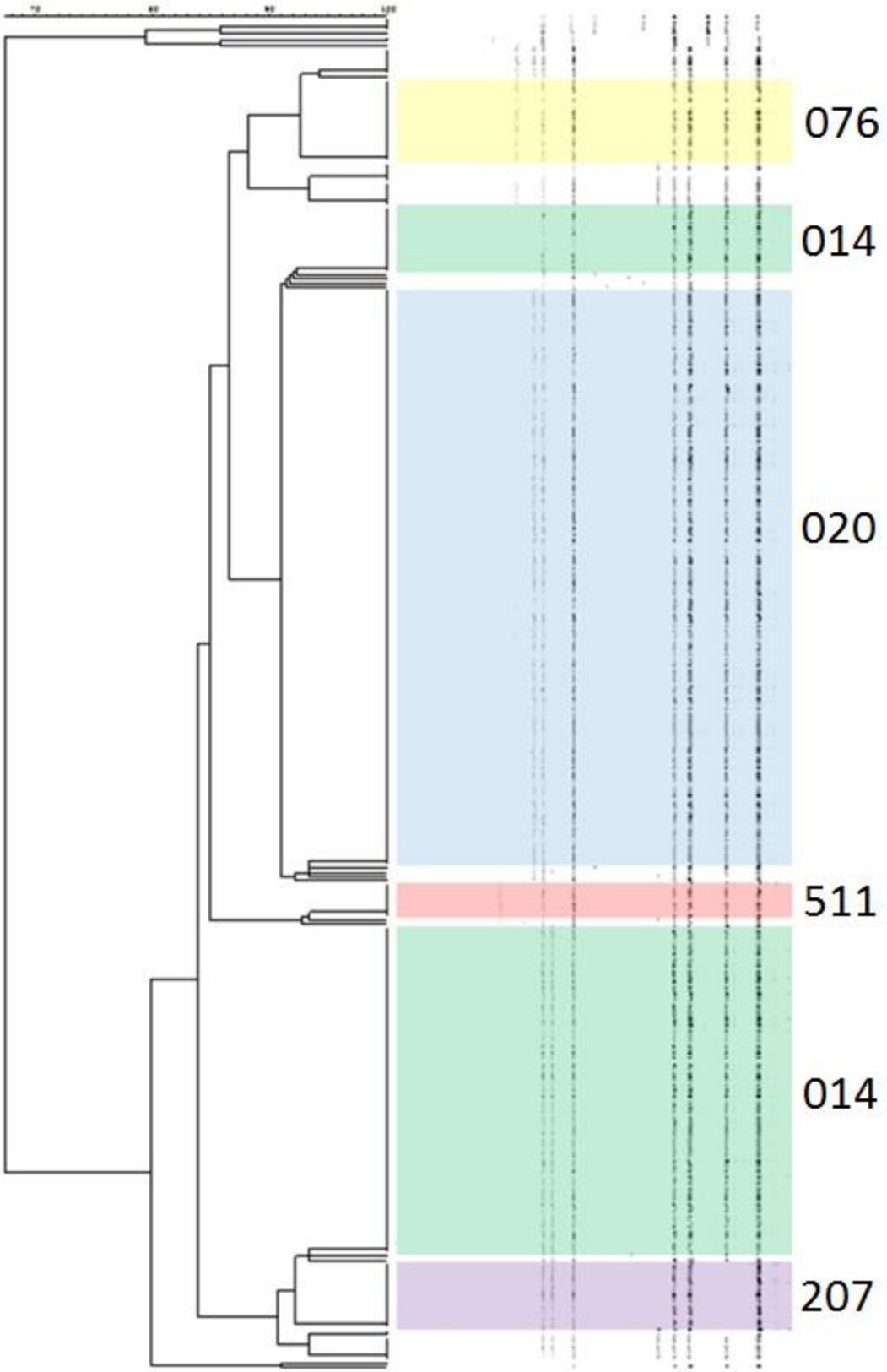

076
014
020
511
014
207

Figure 11. A dendrogram of the ribotyping banding patterns from all isolates in this study. Major clades are highlighted and their respective ribotype is shown on the right side of the clade.

5.6.3. Epidemiological Comparison of PFGE Type 0023 and 0033

Focusing in on the two most predominant NAP4 PFGE types, 0023 and 0033, further analysis was performed in order to determine any additional epidemiological, phenotypic, or molecular differences that could be contributing to the emergence of PFGE type 0033 isolates. Patient data of CDI cases involving NAP4 isolates with PFGE types 0023 or 0033 were compared (Table 16). No statistically significant differences were observed between these two PFGE types with respect to any of the recorded epidemiological data analyzed.

Table 16. Summary of patient data from NAP4 isolates with PFGE type 0023 or 0033 collected between 2004 and 2011

	All Isolates	PFGE Type 0023	PFGE Type 0033	
Total Cases	**338**	**88**	**98**	**P-value**
Morbidity & Mortality				
Cases Requiring Colectomy	1 (0.3%)	0 (0.0%)	1 (1.0%)	1.00
Cases Requiring ICU Admission	2 (0.6%)	1 (1.1%)	0 (0.0%)	0.47
30-day All-Cause Mortality	27 (7.9%)	9 (10.2%)	8 (8.1%)	0.63
30-day Attributable Mortality	3 (0.9%)	1 (1.1%)	2 (2.0%)	1.00
Gender*				
Male	166 (49%)	37 (42.0%)	49 (50.0%)	0.38
Female	166 (49%)	47 (53.4%)	48 (49.0%)	0.38
Age**				
Average Age	53	51	53	0.39
≤ 18	78 (23.5%)	20 (22.7%)	24 (24.4%)	0.81
19-64	105 (31.6%)	32 (36.3%)	27 (27.5%)	0.17
65+	149 (44.9%)	33 (37.5%)	45 (45.9%)	0.27
Location				
West	131 (38.7%)	32 (36.5%)	41 (41.8%)	0.45
Central	181 (53.5%)	49 (55.7%)	44 (44.8%)	0.14
East	26 (7.7%)	7 (7.9%)	13 (13.2%)	0.24
Antimicrobial Resistance****				
Metronidazole				
Resistant Isolates	0 (0.0%)	0 (0.0%)	0 (0.0%)	1.00
MIC50	0.25	0.38	0.25	
MIC90	0.50	0.50	0.50	
MIC Range	0.023 - 2.00	0.094 - 2.00	0.064 - 0.75	
Clindamycin				
Resistant Isolates	125 (37.0%)	36 (40.9%)	29 (30.0%)	0.11
MIC50	4.00	4.00	3.00	
MIC90	8.00	8.00	6.00	
MIC Range	0.016 - >256	0.75 - 256	0.094 - 12.0	
Vancomycin				
Resistant Isolates	0 (0.0%)	0 (0.0%)	0 (0.0%)	1.00
MIC50	0.75	0.75	0.75	
MIC90	1.00	1.00	0.75	
MIC Range	0.38 - 3.00	0.50 - 2.00	0.38 - 1.50	
Rifampin				
Resistant Isolates	1 (0.3%)	1 (1.1%)	0 (0.0%)	0.47
MIC50	<0.002	0.002	<0.002	
MIC90	0.003	0.003	0.002	
MIC Range	<0.002 - >32	<0.002 - >32	<0.002 – 0.003	
Moxifloxacin				

Resistant Isolates	22 (6.5%)	7 (8.0%)	6 (6.1%)	0.62
MIC50	1.50	1.50	1.50	
MIC90	2.00	2.00	2.00	
MIC Range	0.19 - >32	0.75 - >32	1.0 - >32	
Tigecycline				
Resistant Isolates	0 (0.0%)	0 (0.0%)	0 (0.0%)	1.00
MIC50	0.125	0.094	0.125	
MIC90	0.190	0.190	0.190	
MIC Range	0.047 - 0.38	0.047 - 0.25	0.047 - 0.250	

***** - 6 NAP4 isolates did not have patient gender recorded

****** - 6 NAP4 isolates did not have patient age recorded

****- all MIC50, MIC90, and MIC range units are measured in μg/ml

5.6.4. Toxin PCR

PCR analysis of PFGE types 0023 and 0033 was undertaken to look for the presence of toxin-encoding genes (Table 17). There were no significant differences found between the two PFGE types regarding proportion of isolates with a certain toxin PCR profile.

Table 17. Toxin PCR results of NAP4 isolates with PFGE type 0023 or 0033

	All Isolates	PFGE Type 0023	PFGE Type 0033
Total Cases	**338**	**88**	**98**
tcdA			
Positive	338 (100.0%)	88 (100.0%)	98 (100.0%)
Negative	0 (0.0%)	0 (0.0%)	0 (0.0%)
tcdB			
Positive	338 (100.0%)	88 (100.0%)	98 (100.0%)
Negative	0 (0.0%)	0 (0.0%)	0 (0.0%)
tcdC			
Positive	335 (99.1%)	86 (97.7%)	98 (100.0%)
Positive with Deletion	3 (0.9%)	2 (2.3%)	0 (0.0%)
cdtB			
Positive	1 (0.3%)	0 (0.0%)	0 (0.0%)
Negative	337 (99.7%)	88 (100.0%)	98 (100.0%)
tpi			
Positive	338 (100.0%)	88 (100.0%)	98 (100.0%)
Negative	0 (0.0%)	0 (0.0%)	0 (0.0%)

5.6.5. Growth Assay

24-hour growth curve assays were also performed on 20 additional *C. difficile* isolates (Figure 12). This pool of isolates contained representatives of PFGE type 0023, PFGE type 0033, NAP1, and NAP9. The NAP1 isolates were included as controls representing the most prevalent NAP-type in all years monitored during this study. The NAP9 isolates were included as controls representing a NAP-type that had one of the lowest prevalence levels during all study years (Figure 6).

Isolates appeared to enter the exponential phase between 2 to 4 hours. Mid-log phase was achieved approximately 5-6 hours post-inoculation into nutrient broth. Stationary phase growth was reached at approximately 8 hours post-inoculation. PFGE type 0023 isolates tested appeared to enter log phase earlier than the other *C. difficile* isolates. Upon further analysis, PFGE type 0023 demonstrated a significantly faster entry into exponential phase than NAP4 PFGE type 0033, NAP1, and NAP9 isolates with a p-value of 0.03, 0.01, and 0.02, respectively.
When focus was shifted to the growth rates during exponential phase NAP4 PFGE type 0033 isolates demonstrated a significantly greater growth rate between 4 and 8 hours than NAP4 PFGE type 0023 isolates with a p-value of <0.01.

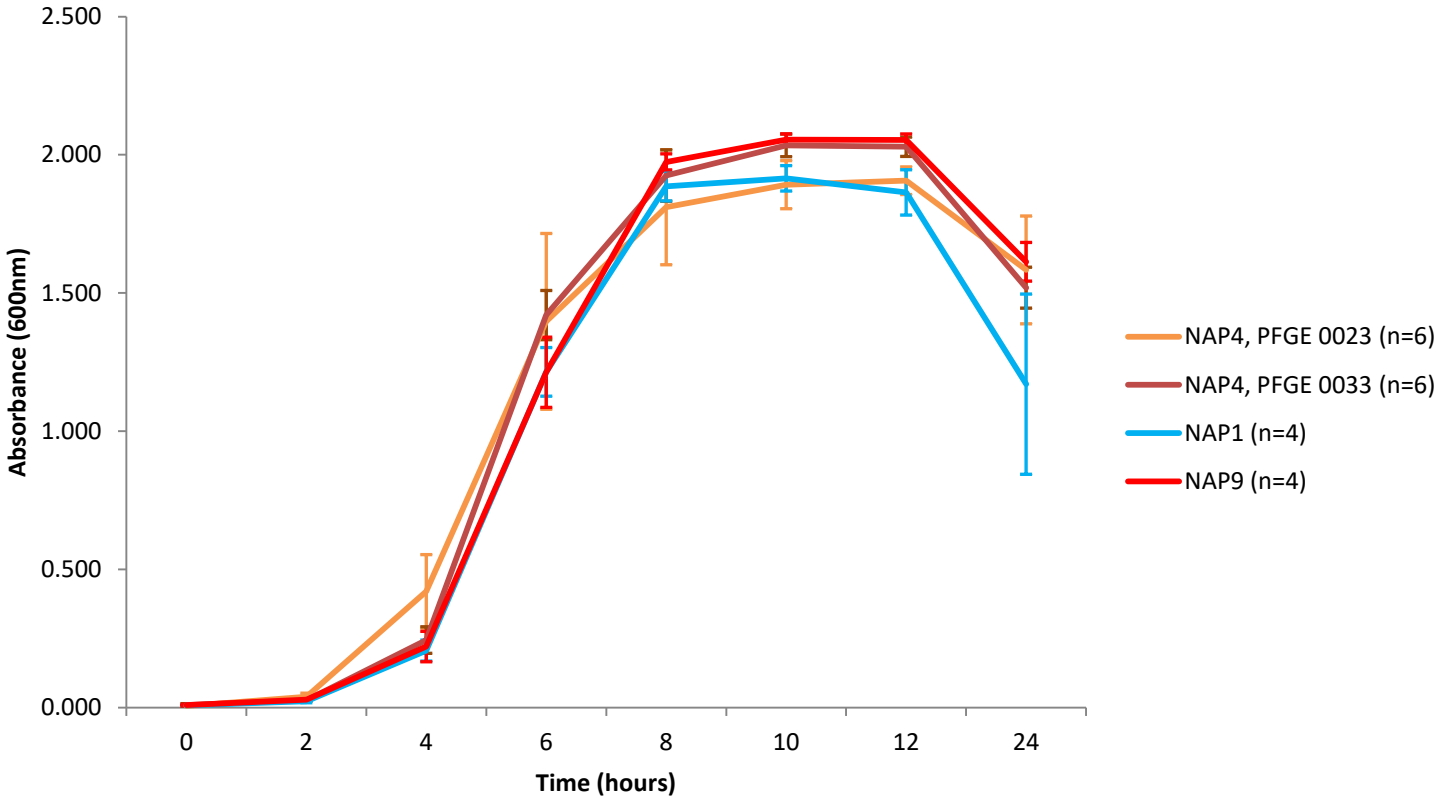

2.500
2.000
1.500
1.000
0.500
0.000
Absorbance (600nm)
0
2
4
6
8
10
12
24
Time (hours)
NAP4, PFGE 0023 (n=6)
NAP4, PFGE 0033 (n=6)
NAP1 (n=4)
NAP9 (n=4)

Figure 12. Growth curves generated from the absorbance readings of several NAP1, NAP4, and NAP9 *Clostridium difficile* isolates. The NAP4 isolates were stratified into groups based on their PFGE type, 0023 or 0033. All isolates were tested in triplicate. Growth curves are shown with standard deviation error bars depicting broth culture absorbance (OD_{600}) over time.

5.6.6. Sporulation Assay

The sporulation assay was also performed on 20 additional *C. difficile* isolates. This particular assay was performed to help reveal any differences in the sporulation ratio of two NAP4 PFGE types, 0023 and 0033. A difference in sporulation ratios may have contributed to the differences in their national prevalence rates. The average sporulation ratio of the isolates tested was found to be 44% with a standard deviation of 23%. Ratios that fell one standard deviation outside of the average ratio were considered outliers and were repeated for confirmation. Ratios ranged from 5.5% to 96% (Figure 13). No significant differences in sporulation ratios were observed between any groups of isolates in this experiment, but PFGE type 0033 had a noticeably larger sporulation ratio (51.4% ratio; 19.1% standard deviation) than PFGE type 0023 (31.2% ratio; 25.5% standard deviation) with a p-value of 0.09.

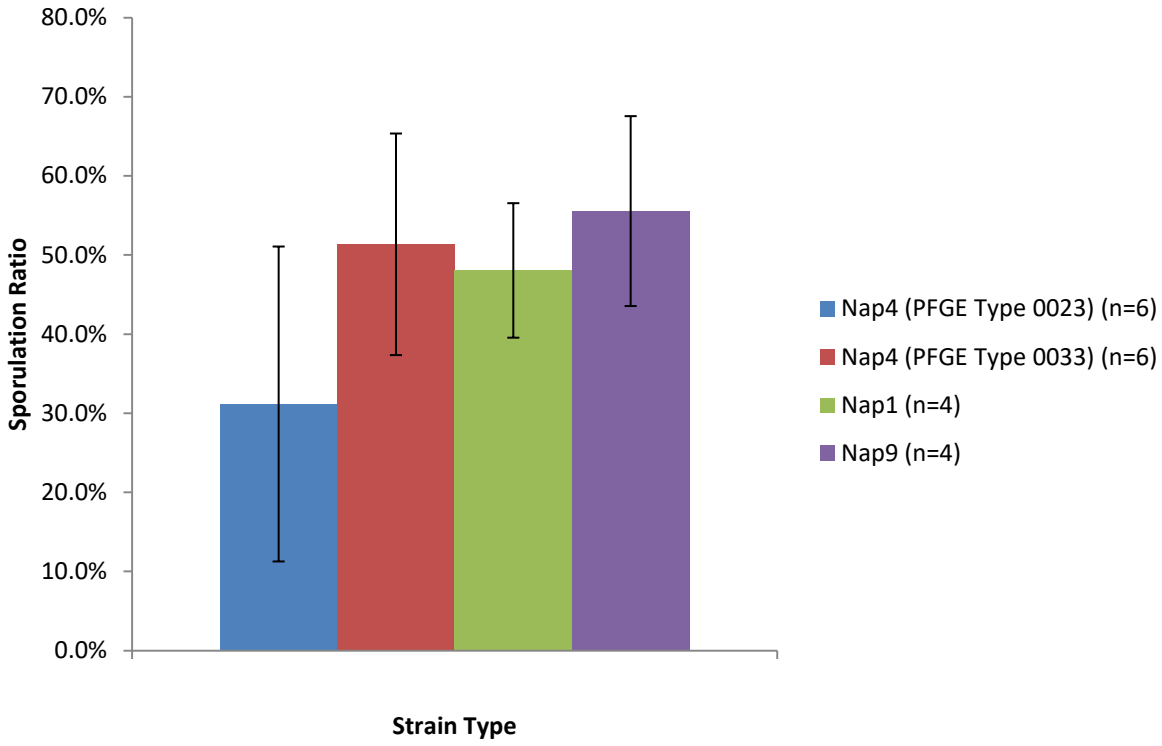

80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Sporulation Ratio
Strain Type
Nap4 (PFGE Type 0023) (n=6)
Nap4 (PFGE Type 0033) (n=6)
Nap1 (n=4)
Nap9 (n=4)

Figure 13. Average sporulation ratios calculated from running the sporulation assay on *Clostridium difficile* isolates of various types; NAP4 PFGE type 0023, NAP4 PFGE type 0033, NAP1, or NAP9. Sporulation ratios are shown with standard deviation error bars depicting ratio variation between isolates within each group.

5.7. Genomic Analysis

We constructed a phylogenetic tree of the core genome made from 28 *C. difficile* genomes representing all known NAP types to better understand the exclusive and inclusive SNP differences of NAP types (Figure 14). We observed 130,007 unique SNP positions in the core genome. Up to 388 SNPs were observed within the NAP1 cluster containing 8 isolates. It should be noted that the SNP differences among all 10 NAP1 isolates was as high as 7,730 SNPs. The NAP4 cluster showed a difference of 411 SNPs between the two isolates. The NAP4 and NAP1 genomes differed by roughly 19,700 SNPs. Although the inclusive SNP differences among the cluster containing both NAP7 and NAP8 isolates was only 36 SNPs, the NAP7/8 cluster differed from all other NAP types by approximately 87,500 SNPs.

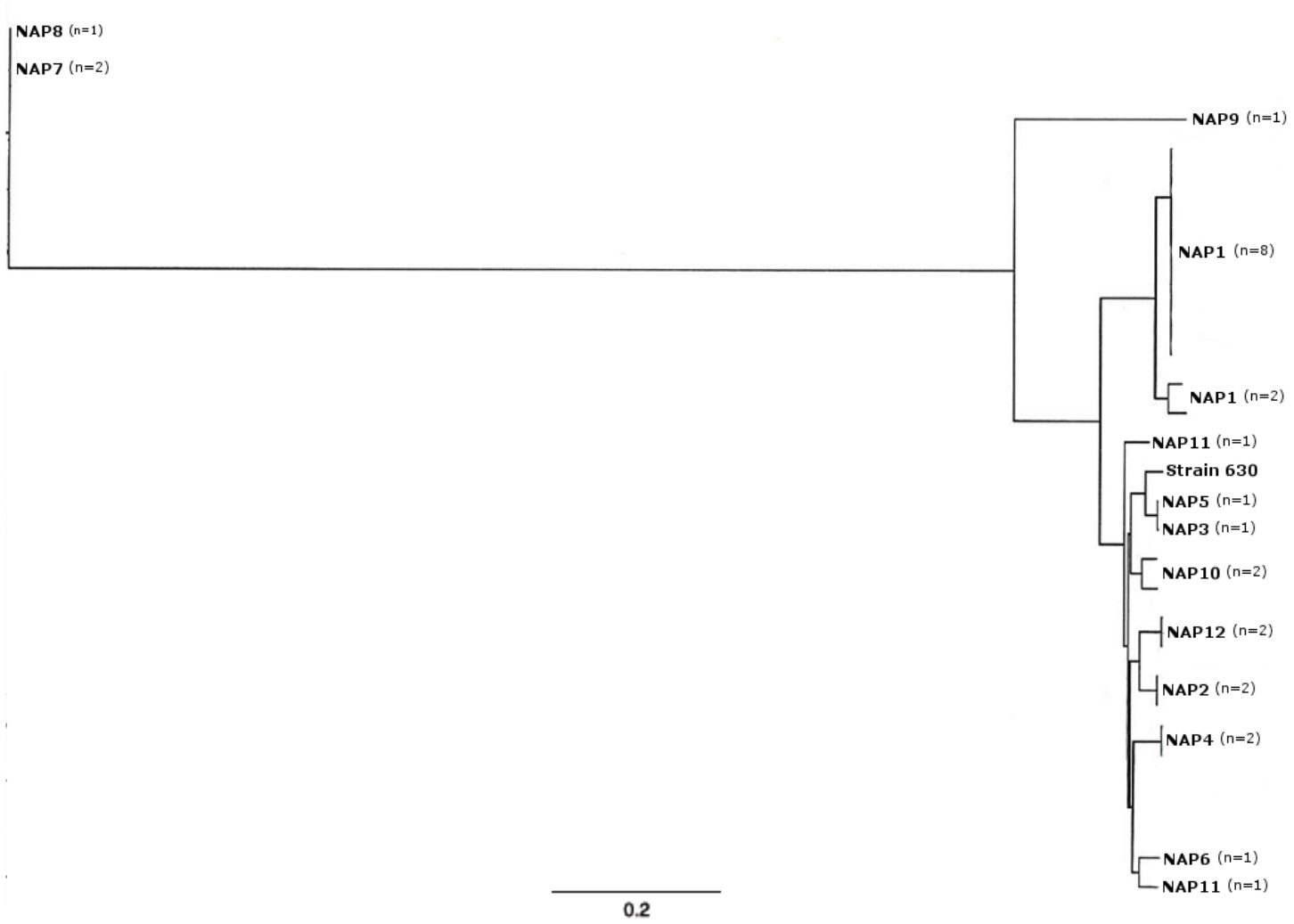

NAP8 (n=1)
NAP7 (n=2)
NAP9 (n=1)
NAP1 (n=8)
NAP1 (n=2)
NAP11 (n=1)
Strain 630
NAP5 (n=1)
NAP3 (n=1)
NAP10 (n=2)
NAP12 (n=2)
NAP2 (n=2)
NAP4 (n=2)
NAP6 (n=1)
NAP11 (n=1)
0.2

Figure 14. Core SNP phylogenetic tree of 28 *Clostridium difficile* genomes with the well-characterized Strain 630 *C. difficile* genome used as the reference. Clusters have NAP types defined and the amount of genomes in those clusters is shown to the right of the cluster definitions.

5.7.1. Pacific Biosciences Sequencing

With the lack of a closed NAP4 *C. difficile* genome available we submitted DNA from two NAP4 isolates to Génome Québec for sequencing using the Pacific Biosciences platform as previously described. Only the PFGE type 0023 strain (08ACD0030) was closed. It had 61 times overall sequence coverage and was received as one, complete contig. The other strain, 21A09CD0060, had 65 times overall coverage, but was received as 161 separate contigs.

5.7.2. MiSeq Sequencing

Of the 85 NAP4 genomes included for study, the average genome read coverage was 125 times (range 60 to 263). The average number of reads was 2,020,816. Assembled genomes consisted of an average of 64 contigs (range 23 to 237). All 85 genomes were used to run the core SNP genomics pipeline as previously described. SNPs were observed at a total of 8,353 different positions in the core genome amongst the genomes being mapped to the reference. The largest amount of SNPs differentiating the core content of two genomes was 4,615 SNPs. According to OrthoMCL, the core genome was found to contain 2,569 genes. The gene counts for all isolates are shown in Table 18.

Table 18. Total amount of genes found amongst 85 NAP4 genomes

	All Isolates	PFGE Type 0023	PFGE Type 0033
Total Genomes	**85**	**37**	**48**
Core Gene Counts			
Average	3,706	3,699	3,710
Minimum*	3,041	3,526	3,041
Maximum*	3,839	3,808	3,839

The phylogenetic tree generated by the core SNP pipeline contained four clusters consisting of more than 3 isolates, labeled A-D. Cluster D contained the majority of isolates (n=61) that were found to differ by as many as 71 SNPs. The average SNP difference between two isolates in Cluster D was 29 SNPs. Isolates in the tree were labeled based on PFGE type,

ribotype, and MLST sequence type in Figures 15, 16, and 17, respectively. All isolates in Cluster A were found to be ribotype 014 while Cluster B contained isolates with only ribotype 020. All isolates in Cluster A and B had MLST sequence type 2 (ST-2) whereas all isolates in Cluster C were sequence type 110 (ST-110). No other clustering based on isolate labels was observed.

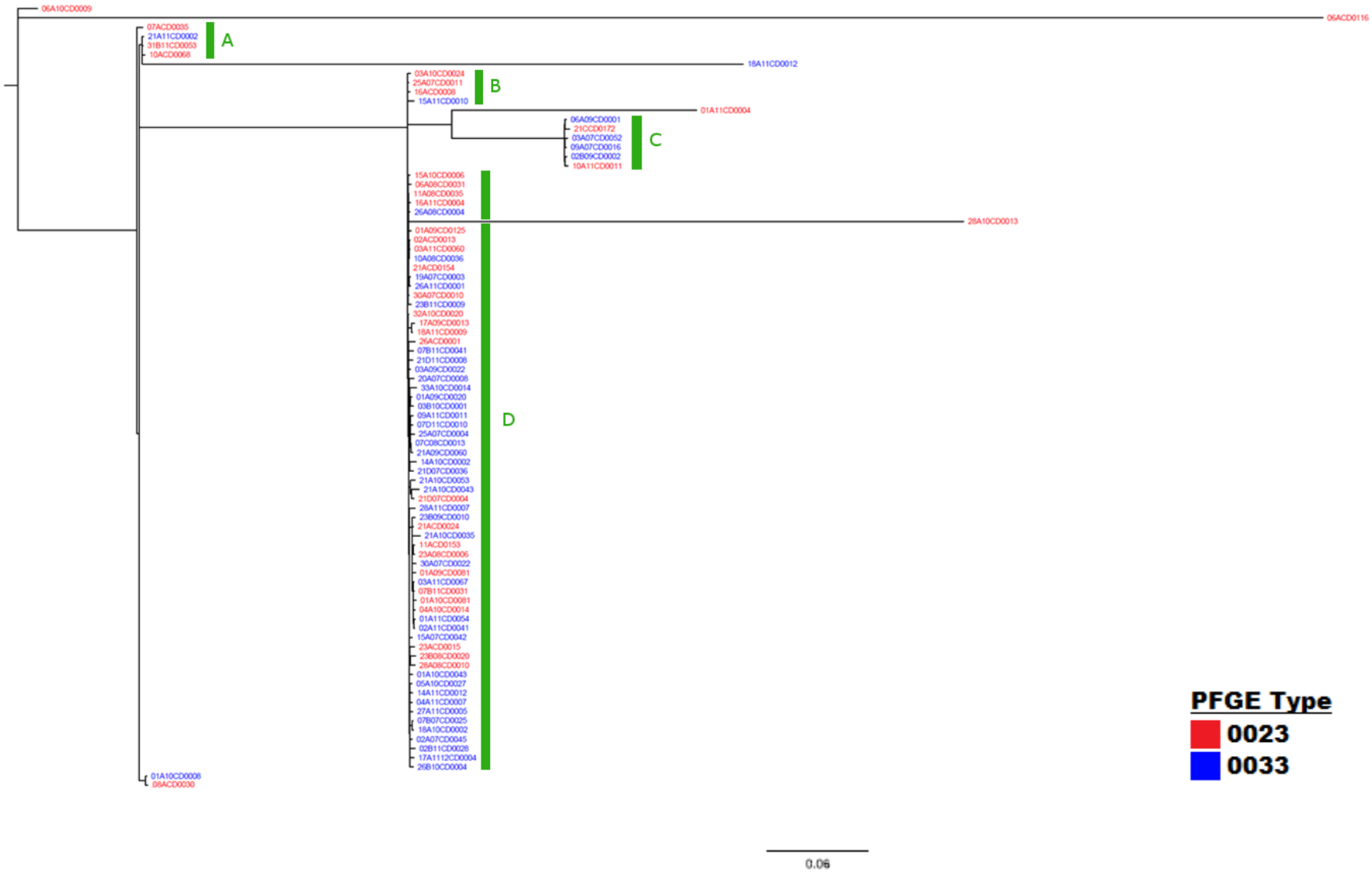

06A10CD0009
06ACD0116
07ACD0035
21A11CD0002
31B11CD0053
10ACD0065
A
18A11CD0012
03A10CD0024
25A07CD0011
16ACD0008
15A11CD0010
B
01A11CD0004
06A09CD0001
21CCD0172
03A07CD0052
09A07CD0016
02B09CD0002
10A11CD0011
C
15A10CD0006
06A08CD0031
11A08CD0035
16A11CD0004
26A08CD0004
28A10CD0013
01A09CD0125
02ACD0013
03A11CD0060
10A08CD0036
21ACD0154
19A07CD0003
26A11CD0001
30A07CD0010
23B11CD0009
32A10CD0020
17A09CD0013
18A11CD0009
26ACD0001
07B11CD0041
21D11CD0008
03A09CD0022
20A07CD0008
33A10CD0014
01A09CD0020
03B10CD0001
09A11CD0011
07D11CD0010
25A07CD0004
07C08CD0013
21A09CD0060
14A10CD0002
21D07CD0036
21A10CD0053
21A10CD0043
21D07CD0004
28A11CD0007
23B09CD0010
21ACD0024
21A10CD0035
11ACD0153
23A08CD0006
30A07CD0022
01A09CD0081
03A11CD0067
07B11CD0031
01A10CD0081
04A10CD0014
01A11CD0054
02A11CD0041
15A07CD0042
23ACD0015
23B08CD0020
28A08CD0010
01A10CD0043
05A10CD0027
14A11CD0012
04A11CD0007
27A11CD0005
07B07CD0025
18A10CD0002
02A07CD0045
02B11CD0028
17A112CD0004
26B10CD0004
D
01A10CD0008
08ACD0030
PFGE Type
0023
0033
0.06

Figure 15. Core SNP phylogenetic tree of 84 NAP4 *Clostridium difficile* genomes labeled by PFGE type. Isolates were mapped against a closed NAP4 *C. difficile* reference genome (08ACD0030). Genomes are colour-coded based on their PFGE type. The 4 major clusters (green bars) are denoted as A, B, C, and D.

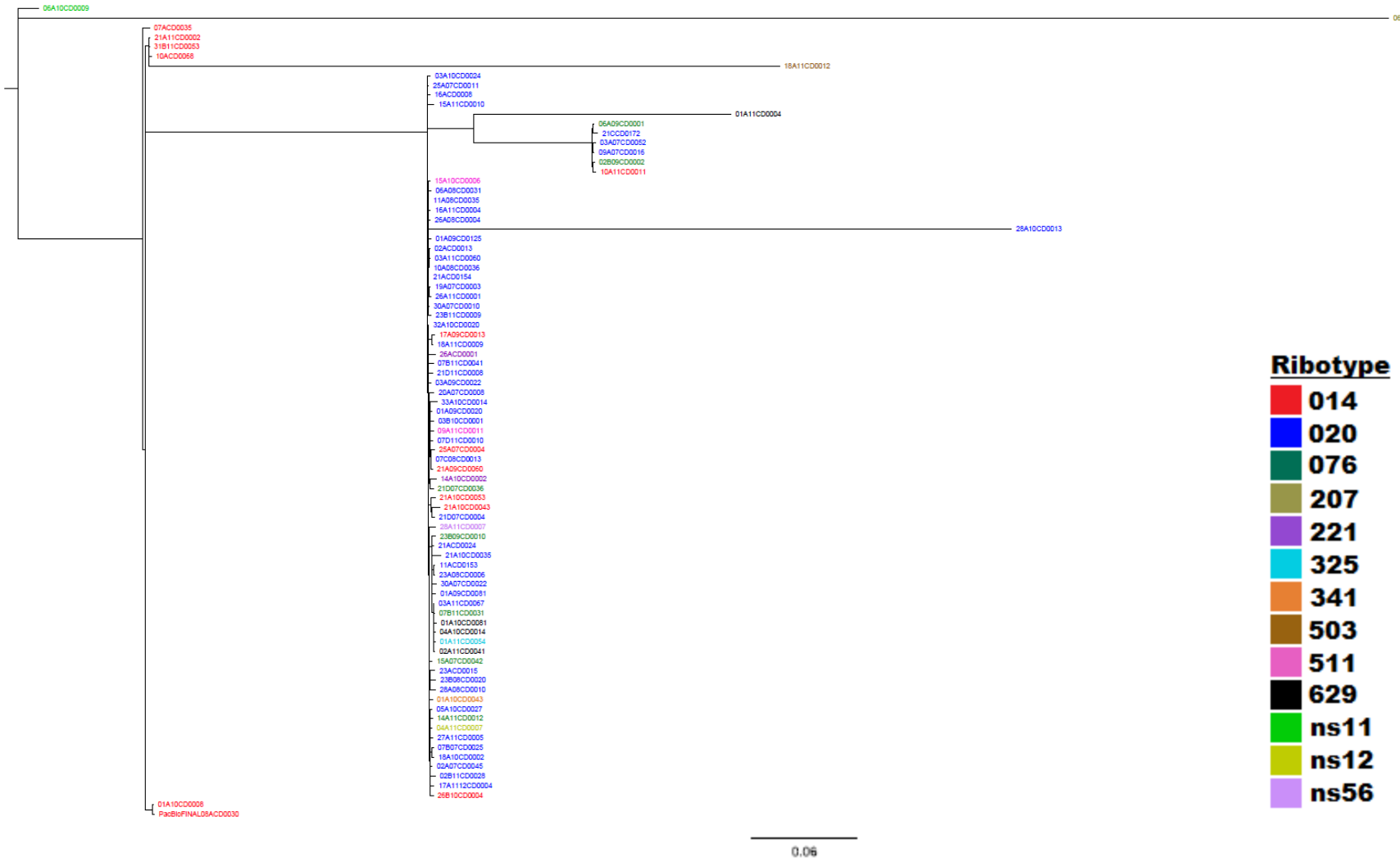

06A10CD0009
06ACD0116
07ACD0036
21A11CD0002
31B11CD0053
10ACD0068
18A11CD0012
03A10CD0024
25A07CD0011
16ACD0006
15A11CD0010
01A11CD0004
06A09CD0001
21CCD0172
03A07CD0052
09A07CD0016
02B09CD0002
10A11CD0011
15A10CD0006
06A08CD0031
11A08CD0035
16A11CD0004
26A08CD0004
28A10CD0013
01A09CD0125
02ACD0013
03A11CD0060
10A06CD0036
21ACD0154
19A07CD0003
26A11CD0001
30A07CD0010
23B11CD0009
32A10CD0020
17A09CD0013
18A11CD0009
26ACD0001
07B11CD0041
21D11CD0008
03A09CD0022
29A07CD0008
33A10CD0014
01A09CD0020
03B10CD0001
09A11CD0011
07D11CD0010
25A07CD0004
07C08CD0013
21A09CD0060
14A10CD0002
21D07CD0036
21A10CD0053
21A10CD0043
21D07CD0004
28A11CD0007
23B09CD0010
21ACD0024
21A10CD0035
11ACD0153
23A09CD0006
30A07CD0022
01A09CD0081
03A11CD0067
07B11CD0031
01A10CD0081
04A10CD0014
01A11CD0054
02A11CD0041
15A07CD0042
23ACD0015
23B08CD0020
28A08CD0010
01A10CD0043
05A10CD0027
14A11CD0012
04A11CD0007
27A11CD0005
07B07CD0025
18A10CD0002
02A07CD0045
02B11CD0028
17A1112CD0004
26B10CD0004
01A10CD0008
PacBioFINAL08ACD0030
0.06
Ribotype
014
020
076
207
221
325
341
503
511
629
ns11
ns12
ns56

Figure 16. Core SNP phylogenetic tree of 84 NAP4 *Clostridium difficile* genomes labeled by ribotype. Isolates were mapped against a closed NAP4 *C. difficile* reference genome (08ACD0030). Genomes are colour-coded based on their ribotype.

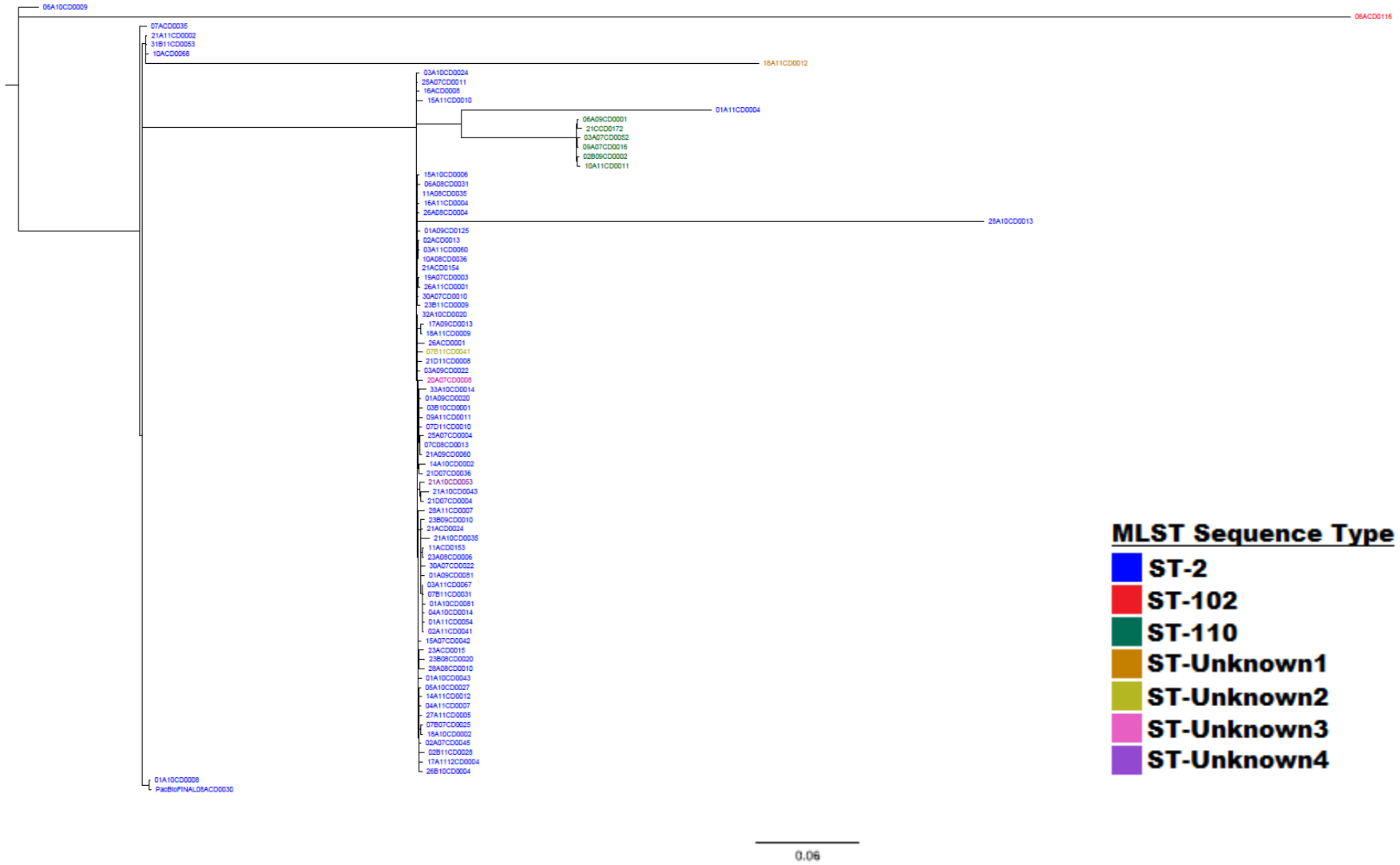

06A10CD0009
06ACD0116
07ACD0035
21A11CD0002
31B11CD0053
10ACD0068
18A11CD0012
03A10CD0024
25A07CD0011
16ACD0008
15A11CD0010
01A11CD0004
06A09CD0001
21CCD0172
03A07CD0052
09A07CD0016
02B09CD0002
10A11CD0011
15A10CD0006
06A08CD0031
11A08CD0035
16A11CD0004
25A08CD0004
28A10CD0013
01A09CD0125
02ACD0013
03A11CD0060
10A08CD0036
21ACD0154
19A07CD0003
26A11CD0001
30A07CD0010
23B11CD0009
32A10CD0020
17A09CD0013
18A11CD0009
26ACD0001
07B11CD0041
21D11CD0008
03A09CD0022
20A07CD0008
33A10CD0014
01A09CD0020
03B10CD0001
09A11CD0011
07D11CD0010
25A07CD0004
07C08CD0013
21A09CD0060
14A10CD0002
21D07CD0036
21A10CD0053
21A10CD0043
21D07CD0004
28A11CD0007
23B09CD0010
21ACD0024
21A10CD0035
11ACD0153
23A08CD0006
30A07CD0022
01A09CD0081
03A11CD0057
07B11CD0031
01A10CD0081
04A10CD0014
01A11CD0054
02A11CD0041
15A07CD0042
23ACD0015
23B08CD0020
28A08CD0010
01A10CD0043
05A10CD0027
14A11CD0012
04A11CD0007
27A11CD0005
07B07CD0025
18A10CD0002
02A07CD0045
02B11CD0028
17A11CD0004
26B10CD0004
01A10CD0008
PacBioFINAL08ACD0030
MLST Sequence Type
ST-2
ST-102
ST-110
ST-Unknown1
ST-Unknown2
ST-Unknown3
ST-Unknown4
0.06

Figure 17. Core SNP phylogenetic tree of 84 NAP4 *Clostridium difficile* genomes labeled by MLST sequence type. Isolates were mapped against a closed NAP4 *C. difficile* reference genome (08ACD0030). Genomes are colour-coded based on their MLST sequence type.

In silico MLST revealed 3 known sequence types, ST-2, ST-102, and ST-110 as well as 4 unknown sequence types (Table 19).

Table 19. In silico MLST sequence types found amongst 85 NAP4 genomes

	Allele Pattern							
Sequence Type	***adk***	***atpA***	***dxr***	***glyA***	***recA***	***sodA***	***tpi***	**Number Found**
ST-2	1	1	2	1	5	3	1	74
ST-102	1	1	2	1	5	5	1	1
ST-110	1	1	2	1	13	3	1	6
ST-Unknown1	1	5	2	1	5	3	1	1
ST-Unknown2	1	1	2	1	5	46	1	1
ST-Unknown3	1	1	22	1	5	37	1	1
ST-Unknown4	1	1	2	1	5	18	1	1

PFGE typing was repeated on 5 isolates that were considered outliers in the tree and confirmed to match their original PFGE type. As well, the ribotype and MLST sequence type of these isolates was also considered (Table 20). Three outliers had uncommon ribotypes (207, 503, and ns11) in that those ribotypes were only observed once among the 85 isolates in the core SNP tree. For MLST, one outlier was ST-102 and the other had an unknown sequence type (ST-Unknown1).

Table 20. Core SNP tree outlier analysis using common typing methods.

Outlier ID	PFGE Type	Ribotype	MLST Sequence Type
01A11CD0004	0023	020	ST-2
06A10CD0009	0023	ns11	ST-2
06ACD0116	0023	207	ST-102
18A11CD0012	0033	503	ST-Unknown1
28A10CD0013	0023	020	ST-2

5.8. eBurst Analysis

Cluster analysis was further carried out using minimum spanning trees. Minimum spanning trees were generated using the eBurst algorithm. The SNP matrix, created by the core SNP genomic pipeline, was used to generate node connections and calculate branch lengths. Figures 18 through 20 show minimum spanning trees created using the same SNP matrix that were labeled with various patient data to search for clustering of NAP4 isolates amongst different age groups (Figure 18), patient outcome (Figure 19), and regional distribution (Figure 20). From this analysis no obvious clustering of isolates were observed.

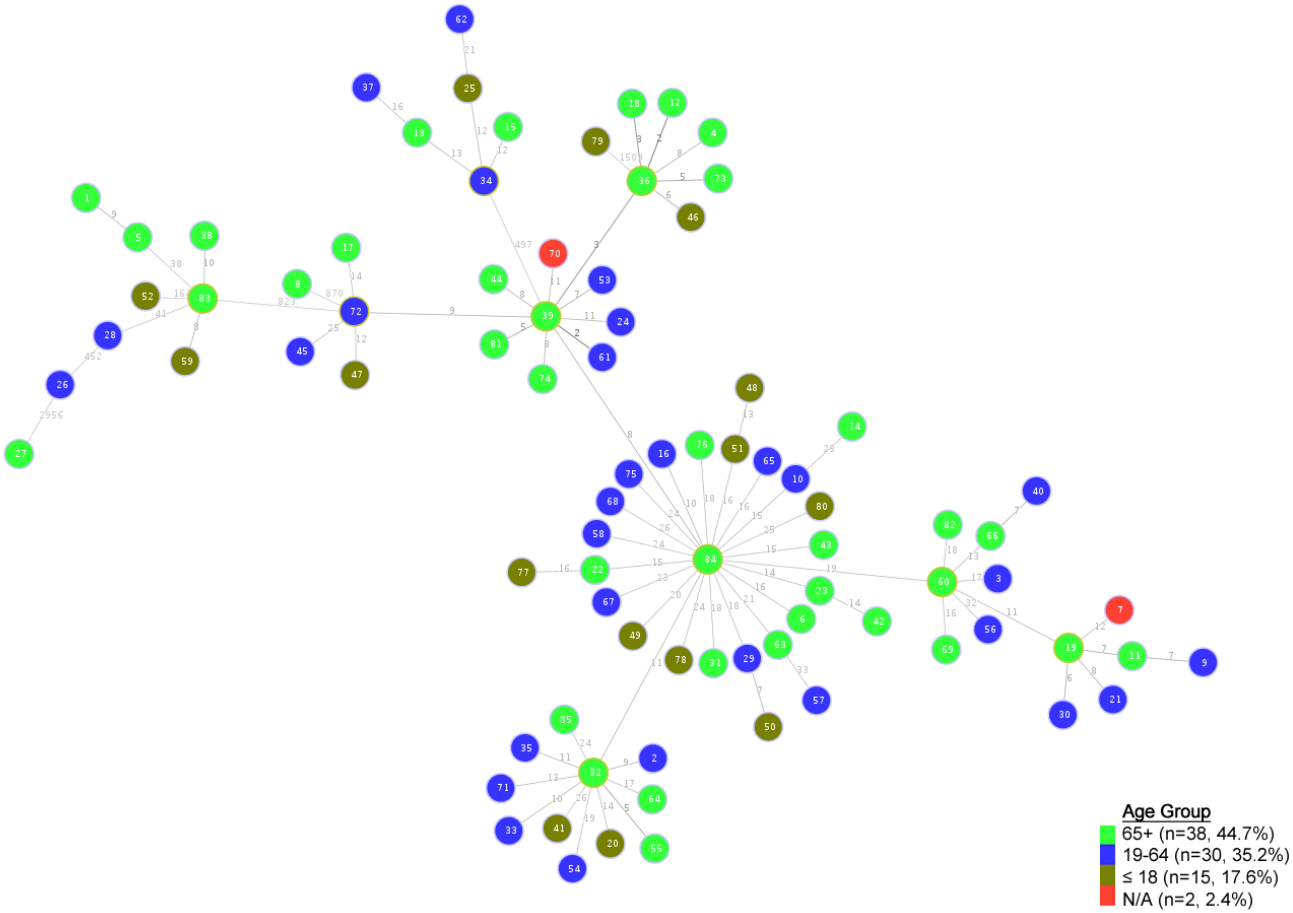

Age Group
65+ (n=38, 44.7%)
19-64 (n=30, 35.2%)
≤ 18 (n=15, 17.6%)
N/A (n=2, 2.4%)

Figure 18. Minimum spanning tree generated from the core SNP matrix of 85 NAP4 isolates with nodes labeled based on the age groups of patients infected by those isolates. Numbered nodes represent individual NAP4 isolates. The age of the patient was not recorded for two isolates (age group N/A). The amount and proportion of isolates associated with a patient age group are shown in the colour-coded legend. Numbers within each node specify a particular isolate. The number on the line connecting two nodes describes the amount of core SNP differences between the genomes of those two isolates.

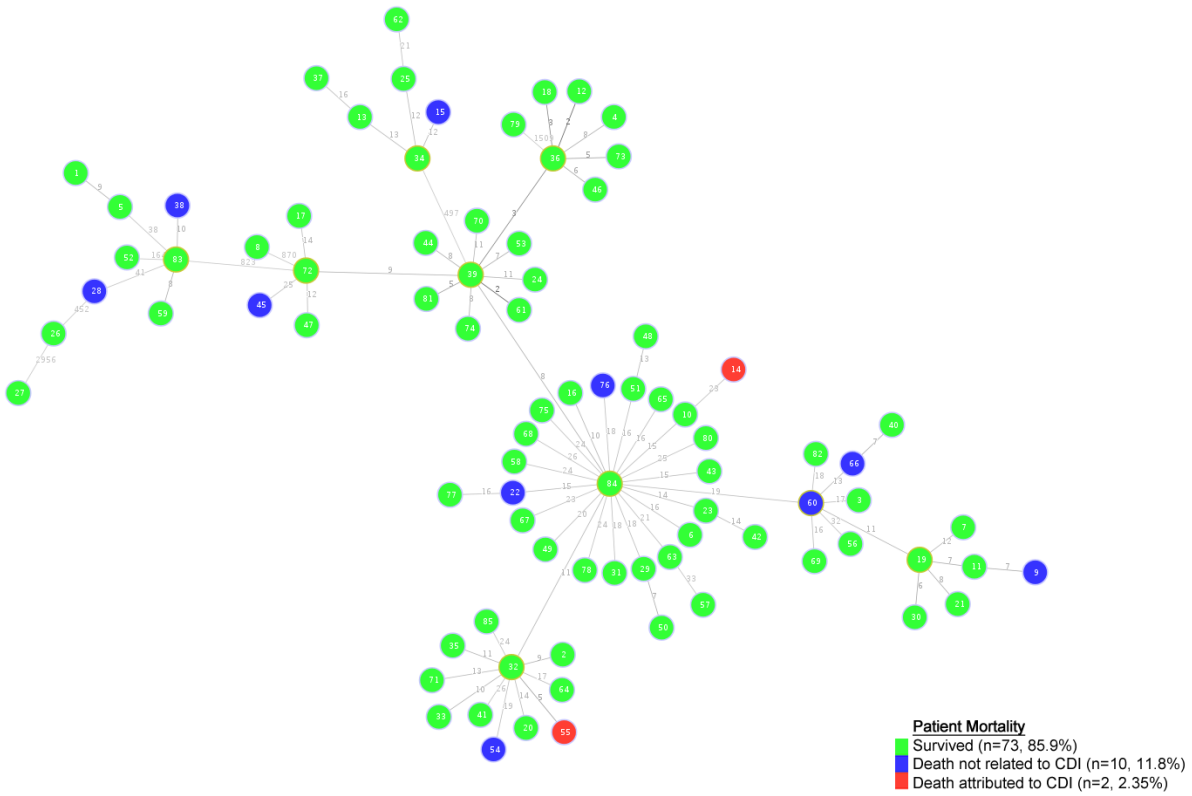

Patient Mortality
Survived (n=73, 85.9%)
Death not related to CDI (n=10, 11.8%)
Death attributed to CDI (n=2, 2.35%)

Figure 19. Minimum spanning tree generated from the core SNP matrix of 85 NAP4 isolates with nodes labeled based on the relation of the isolate to the patient's death, if applicable. Numbered nodes represent individual NAP4 isolates. Green nodes represent isolates collected from patients who were alive at the time of hospital discharge. Blue nodes represent isolates collected from patients who died from non-CDI causes. Red nodes represent isolates that were related to the death of the patient. The amount and proportion of isolates is described in the colour-coded legend. Numbers within each node specify a particular isolate. The number on the line connecting two nodes describes the amount of core SNP differences between the genomes of those two isolates.

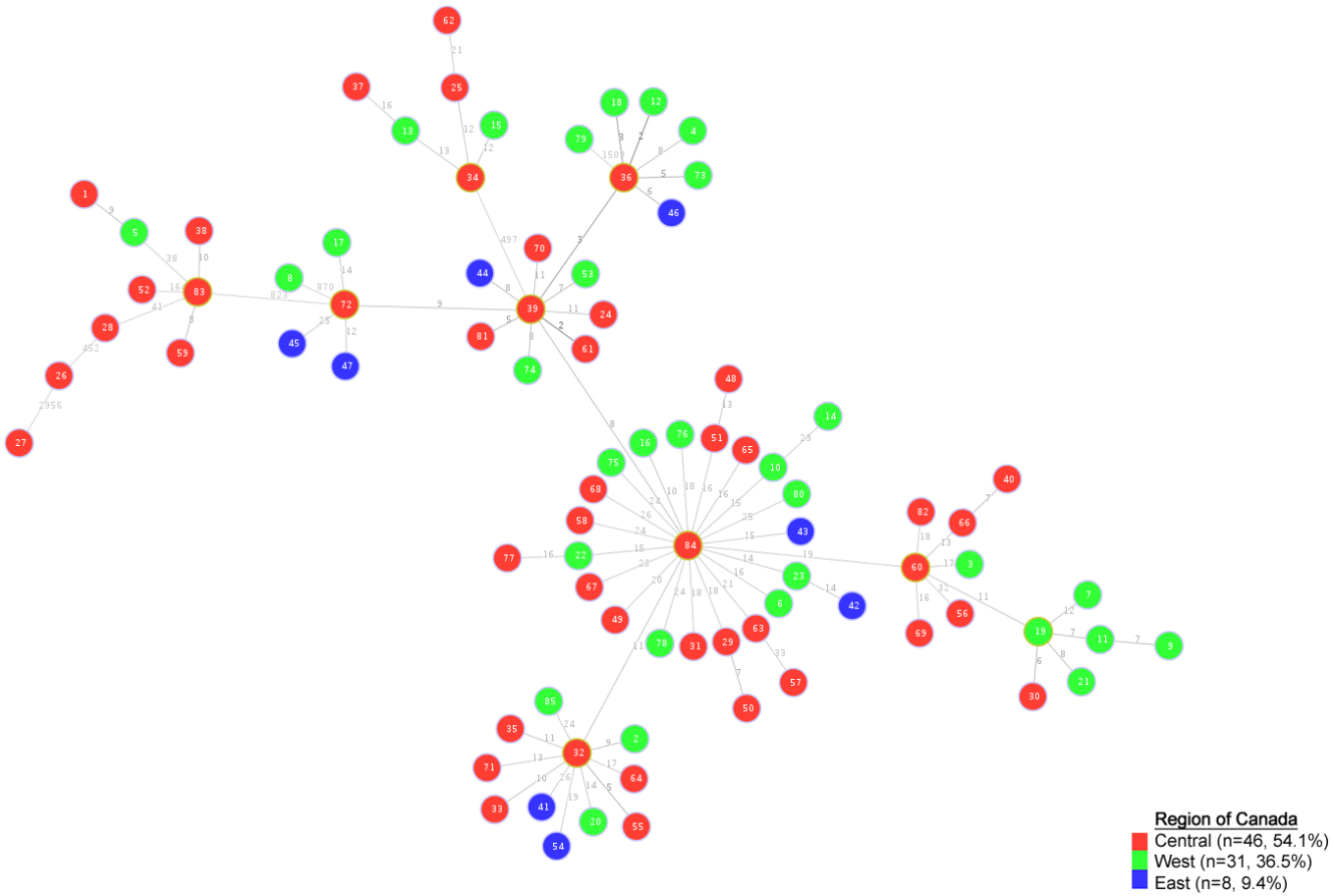

Region of Canada
Central (n=46, 54.1%)
West (n=31, 36.5%)
East (n=8, 9.4%)

Figure 20. Minimum spanning tree generated from the core SNP matrix of 85 NAP4 isolates with nodes labeled based on the region of Canada the isolates were collected from. Numbered nodes represent individual NAP4 isolates. The amount and proportion of isolates collected from each region are described in the colour-coded legend. Numbers within each node specify a particular isolate. The number on the line connecting two nodes describes the amount of core SNP differences between the genomes of those two isolates.

5.9. Antimicrobial Susceptibility Testing

Among the 85 isolates in the core SNP tree, intermediate and resistant phenotypes were only observed for clindamycin and moxifloxacin (Figure 21).

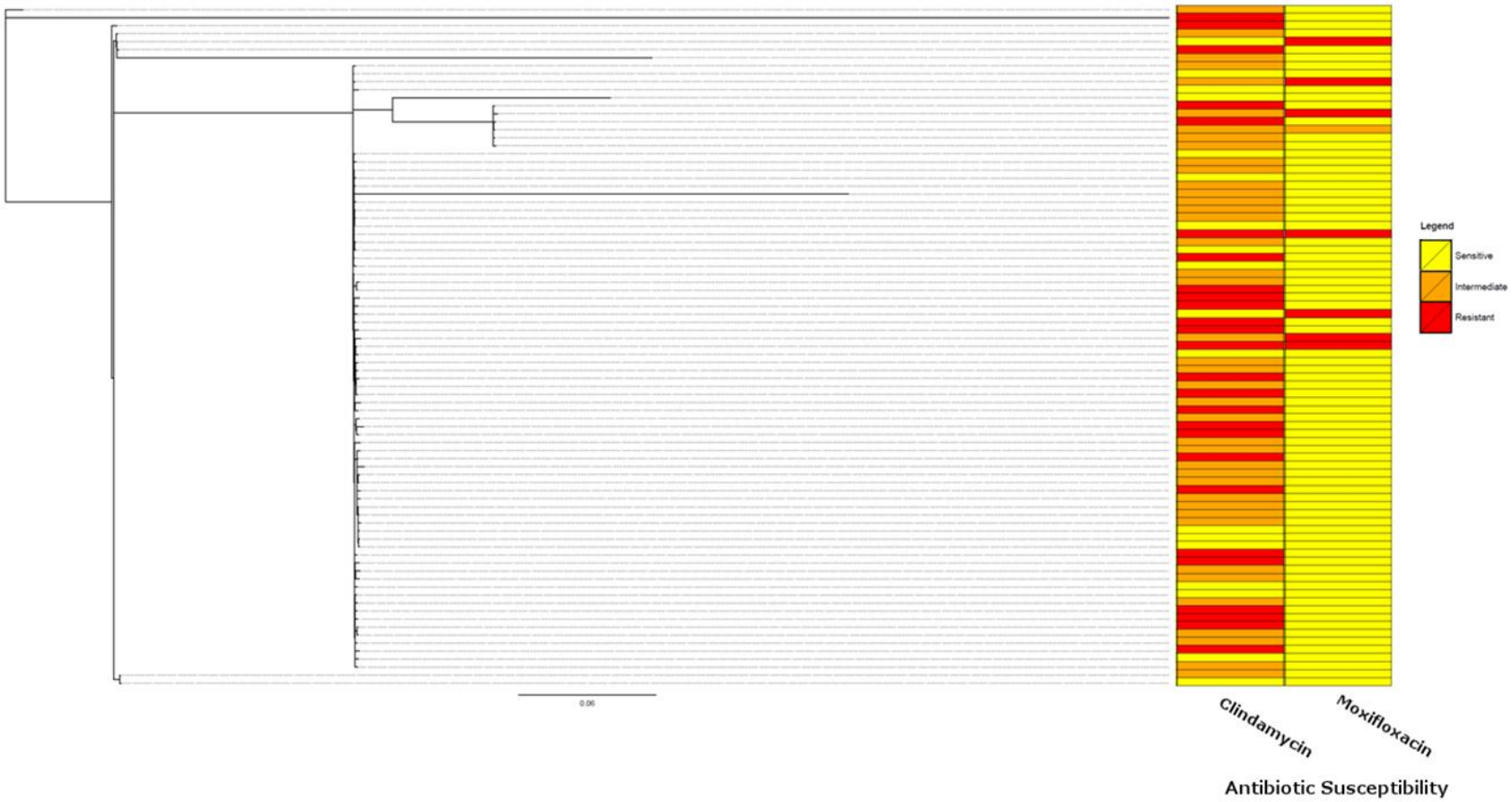

Legend
Sensitive
Intermediate
Resistant
0.06
Clindamycin
Moxifloxacin
Antibiotic Susceptibility

Figure 21. Phylogenetic tree of 85 NAP4 genomes with a heatmap of the antimicrobial susceptibility profiles of each isolate to clindamycin and moxifloxacin are shown to the right of the tree. Yellow cells represent a susceptible antimicrobial phenotype. Orange cells represent an intermediate resistance antimicrobial phenotype. Red cells represent a resistant antimicrobial phenotype.

In the case of clindamycin resistance the erythromycin ribosomal methylase gene B (*ermB*) was not identified in any of the resistant NAP4 isolates. For moxifloxacin resistance, the presence of mutations in the DNA gyrase genes *gyrA* and *gyrB* was found by manual alignment of assembled genomic data (Table 21).

Table 21. Correlation of mutations found in DNA gyrase genes to moxifloxacin resistance profiles and PFGE type of NAP4 isolates

Isolate	PFGE Type (*SmaI*)	Moxifloxacin MIC (µg/ml)	*gyrA* Mutation	*gyrB* Mutation
01A09CD0020	0033	>32	Thr82-Ile	none
33A10CD0014	0033	>32	Thr82-Ile	Asp426-Asn
21D11CD0008	0033	8	none	none
16ACD0008	0023	8	none	Asp426-Asn
21ACD0154	0023	8	Ala118-Thr	none
21CCD0172	0023	32	Thr82-Ile	none
31B11CD0053	0023	>32	Thr82-Ile	none

Within NAP4, MICs of ≥ 32 µg/ml appeared to be associated with a Thr82-Ile mutation in *gyrA*. It is interesting to note that one of the isolates with a moxifloxacin-resistant phenotype did not have a mutation in either of the two DNA gyrase genes. MIC testing of this isolate to moxifloxacin was not repeated due to time constraints.

To account for the absence of *ermB* and the 1 moxifloxacin resistant isolate with no mutations in *gyrA* or *gyrB* ResFinder was used to search all draft genomes for the presence of any additional antimicrobial resistance genes that could possibly be contributing to resistance. However, no resistance genes from the ResFinder database were found in any of these isolates.

5.10. Nucleic Acid and Protein Profile Characterization

5.10.1 Neptune Genomic Characterization

The 85 *C. difficile* genomes were processed by Neptune to find nucleotide sequences that are unique to either PFGE type 0023 genomes or PFGE type 0033 genomes. No sequences were found to be unique to PFGE type 0023 genomes. Two nucleotide sequences were found to be unique to PFGE type 0033 genomes spanning 1,177 bp and 840 bp herein referred to as NepSeq1 and NepSeq2, respectively. These sequences are detailed in Section 4 of the Appendix. A blastn search using NepSeq1 as a query revealed a 100% match to Clostridium phage phiCD505 (accession number: NC_028764). This matching sequence contained two genes for hypothetical proteins 131 bp and 401 bp in length, respectively. A blastp search of the 131 bp gene product did not reveal any further description in any other organism. However, a blastp search of the 401 bp gene product had a 98% match to a previously described transcriptional regulator from *C. difficile*. A blastn search using NepSeq2 as a query revealed a 100% match to the same Clostridium phage phiCD505 as NepSeq1. This matching segment contained 2 genes spanning 240 bp and 389 bp long, respectively. The 240 bp gene putatively encoded a transcriptional regulator related to the BlaI/MecI/CopY family. The 389 bp gene encoded a hypothetical protein. A blastp search of the hypothetical protein from NepSeq2 was found to have been further described as a penicillinase repressor family protein from *PeptoClostridium difficile* (accession number: WP_021376798) with a 100% match. The protein sequences found within the top blastn matches of NepSeq1 and NepSeq2 are shown in Section 5 of the Appendix. A summary of all 76 open reading frames in phiCD505 is described in section 6 of the Appendix [98].

5.11. PFGE Pattern Differentiation

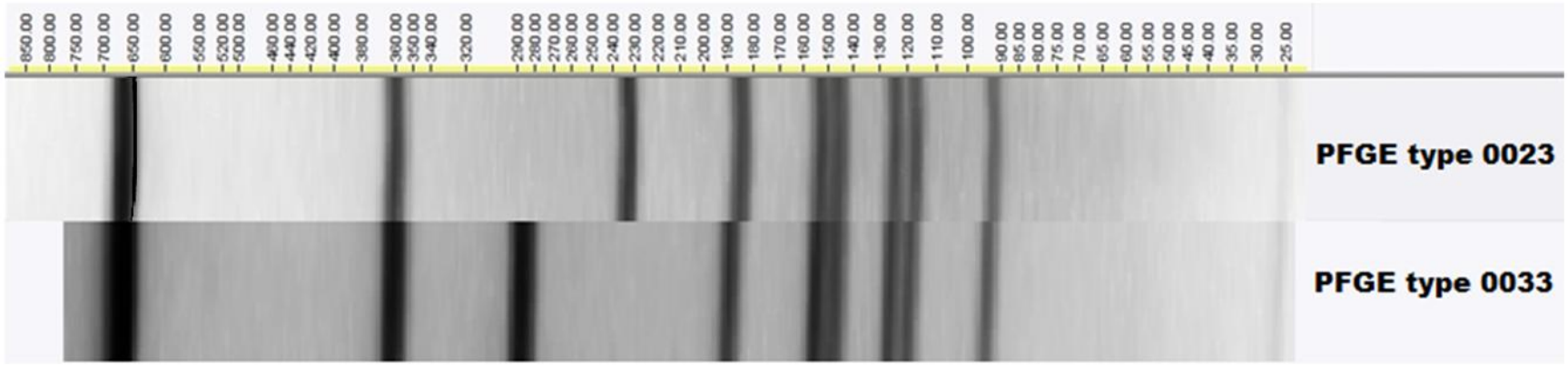
To investigate the one-band shift difference between PFGE types 0033 and 0023 an in silico PFGE digestion of the genomes was performed. This method required closed genomes of which we only had one NAP4 available to us. The band sizes called by the in silico digestion were compared to the band sizes called by the in vitro method (Table 22).

Table 22. Comparison of an in silico *Sma*I digestion of a PFGE type 0023 genome with in vitro *Sma*I digestions of PFGE type 0023 and 0033 genomes

0023 in silico band sizes (bp)	0023 in vitro band sizes (bp)	0033 in vitro band sizes (bp)
1992985		
692836	644960	676690
357361	350980	362690
233812	228150	289530
178303	174820	190470
151836	144240	153660
146241	140570	148150
124606	11590	126940
119057	11270	121450
97983	88840	94050
22698		
14232		

To compare the two PFGE types, the digestion pattern of the PFGE type 0023 was overlaid onto a Mauve alignment of two genomes with either PFGE type (Figure 22). The band that is 233,812 bp in length from the in silico digestion is the band that shifts in size. It is clear that there is extra nucleotide sequence in this region in the PFGE type 0033 genome. The extra nucleotide sequence in this region is 49,313 bp in length.

a)


PFGE type 0023
PFGE type 0033

b)

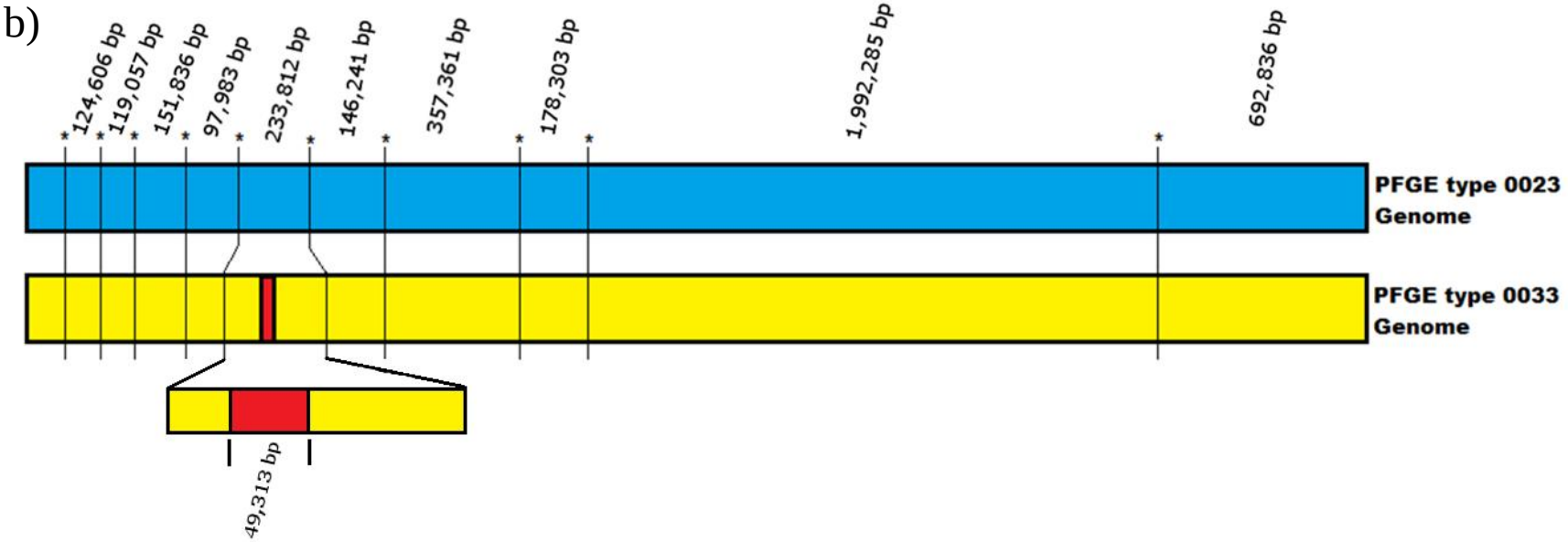

124,606 bp
119,057 bp
151,836 bp
97,983 bp
233,812 bp
146,241 bp
357,361 bp
178,303 bp
1,992,285 bp
692,836 bp
PFGE type 0023 Genome
PFGE type 0033 Genome
49,313 bp

Figure 22. A depiction of in vitro PFGE versus in silico PFGE explaining the band shift between PFGE type 0023 and 0033. a) PFGE fingerprints of PFGE type 0023 and 0033 *Clostridium difficile* genomes digested with *Sma*I restriction enzyme. An electronic sizing ladder is shown above the gel images. Sizes on the ladder are measured in kilobases. b) A representation of an alignment of a PFGE type 0023 genome and a PFGE type 0033 genome. In silico digestion positions in the PFGE type 0023 genome are shown as starred lines. The location of these digestion positions is extended to the PFGE type 0033 genome where a phage-like element (red box) was found increasing the size of the band in that region of the genome.

5.11.1. Phage Analysis

The extra genetic content in the genome of a representative PFGE type 0033 *C. difficile* isolate that was causing the band shift was used as a query for a blastn search. All blastn searches from this region in this isolate matched to the clostridium phage phiCD505, with a query coverage of 100%. This phage was found inserted in close proximity to a guanosine monophosphate (GMP) synthase gene. PHAST identified 17 intact phages at this insertion site, of which 16 were also identified as phiCD505. The other intact phage at this insertion site differed from phiCD505 by 59 orthologues and was most closely related to the clostridium phage phiMMP03. Due to contig breaks, PHAST was unable to determine complete intact phages at this insertion site for the remaining 31 PFGE type 0033 isolates. Blast hits and MAUVE alignments to the non-host chromosome fragments found on contigs within the GMP synthase region confirmed known phage sequence in all PFGE type 0033 isolates and not in any of the 0023 isolates. However, these sequences were often too short to be able to determine the true identity of the phage in the PFGE type 0033 isolates.

Further investigation and characterization of clostridium phages in our isolates was done to determine if other phages could be possibly contributing to the difference in prevalence of PFGE type 0023 and 0033. Only phage genomes that were given a high confidence score by PHAST (defined as “intact phage”) were selected for further study. In total, 123 intact phages were found among the 85 *C. difficile* isolates screened. Individual *C. difficile* genomes contained between zero and five intact phages. PHAST identified 78 intact phages from the 47 PFGE type 0033 isolates whereas only 45 intact phages were found from the 38 PFGE type 0023 isolates. The size range of intact phages was between 26,325 bp to 154,738 bp.

5.11.2. Phage Ortholog Tree

All extracted intact phages were annotated with Prokka. These annotations were used with OrthoMCL to make a large binary table of the orthologous genes each phage contained. This binary table was used to create a phylogenetic tree similar to the SNP matrix used to create the core SNP tree (Figure 23). Among the 123 intact phage genomes and 34 previously described clostridium phages with available sequence data, two clusters were identified. Orthocluster1 contained 16 phages. Using a blastn search, phage genomes from orthocluster1 all matched a known clostridium phage, phiCD505, with 100% query coverage. All phages from this cluster were from PFGE type 0033 genomes as described above. Orthocluster2 contained 22 phages, which did not match to any previously described phages. Phages in this cluster were distributed amongst both PFGE type 0023 and 0033 genomes. The average amount of genes amongst all 157 phage genomes in the ortholog tree was 79. The highest amount of genes found in an intact phage was 154 whereas the lowest amount was 29.

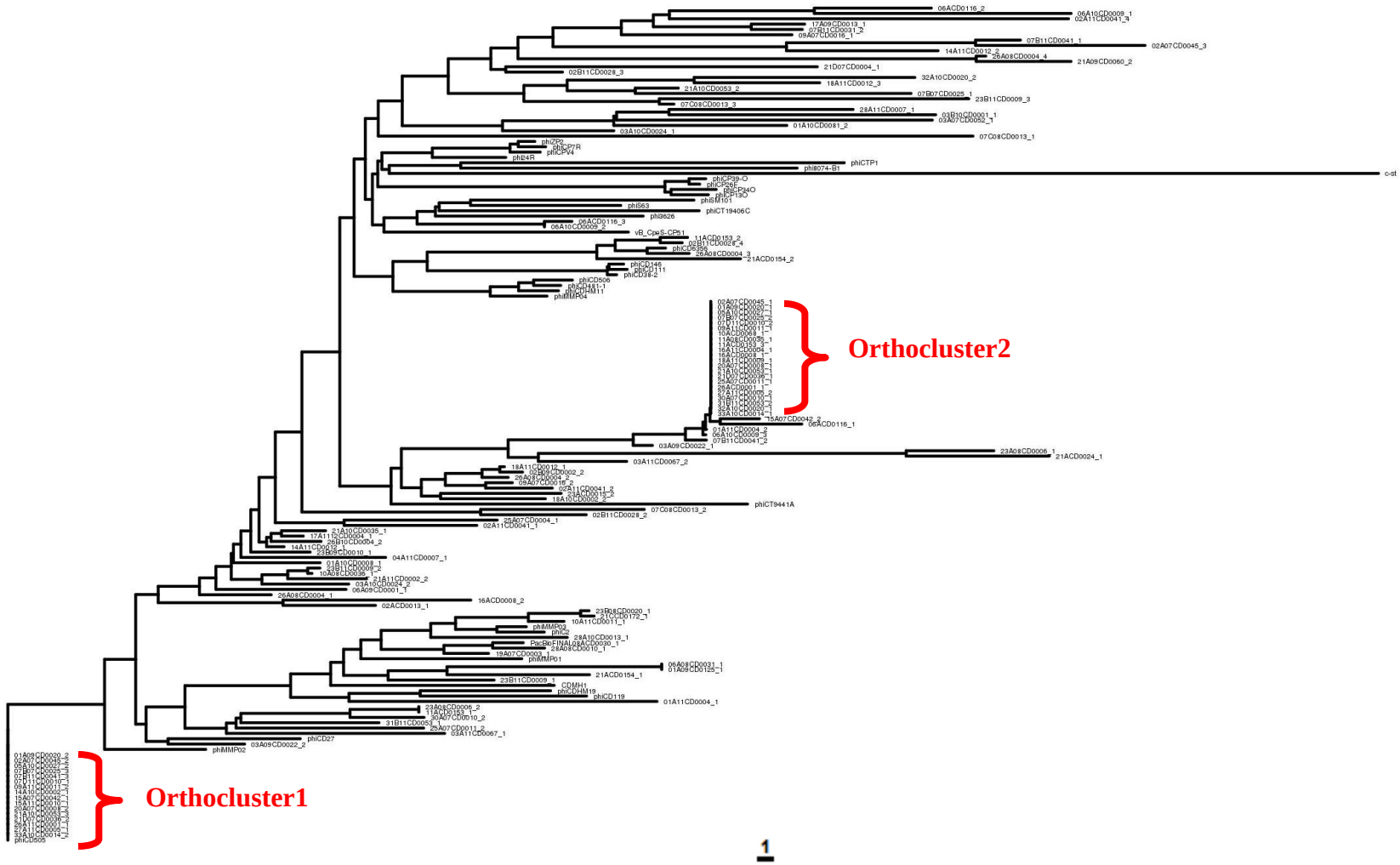

Orthocluster2
Orthocluster1
1

Figure 23. A phylogenetic tree made from the orthologous gene profiles of 123 intact phage genomes among our 85 NAP4 genomes and 34 previously sequenced clostridium phage genomes.

6. Discussion

The pathogen *C. difficile* has become a major cause of nosocomial infection since its correlation with antibiotic-associated pseudomembranous colitis was established in 1978 [8]. A variety of clinical symptoms may be observed, from mild diarrhea to pseudomembranous colitis and possibly death [131]. Major risks of contracting CDI and being afflicted with more severe symptoms have been related to patient age over 65 [15] and the presence of underlying co-morbidities such as cancer [28]. The association of *C. difficile* with hospitals is due to the common use of antimicrobials in this setting which disrupt the natural microflora in the gut [24]. This dysbiosis not only favors *C. difficile* colonization, but also increases the chance of CDI recurrence if populations of the patient's natural gut microflora cannot be restored to normal levels in time [89]. Aside from the antimicrobial agents commonly used for CDI treatment such as metronidazole, vancomycin, and fidaxomicin, research has been performed on alternative methods, including fecal microbiota transplants [131] and phage therapy [95, 115].

A recent report by the CDC ranked *C. difficile* as an urgent-level threat alongside carbapenem-resistant *Enterobacteriaceae* (CRE). Surveillance programs in Canada, the United States, and Germany have recorded a significant increase in incidence rates of CDI that ranged from 2-6 fold in the years 2004 to 2005, 1996 to 2003, and 2002 to 2006, respectively. [12, 48, 94]. In recent years, *C. difficile* has demonstrated its harsh impact on both the health and economy of various nations. In 2012, an analysis of the impact *C. difficile* had on the Canadian healthcare system revealed an estimate of 37,900 cases of CDI, of which 27% of these experienced a recurrence. In this study year alone, *C. difficile* had cost the Canadian healthcare system $281 million, 92% of that being in-hospital expenses [82]. Comparatively, a study of the burden caused by *C. difficile* on the American healthcare system found 453,000 infections in

2013 alone [81]. The annual estimated economic burden in the USA was recently updated to $4.8 billion [29]. The burden of *C. difficile* in acute-care facilities is, in part, due to the increased length of hospital stay, which in one Canadian hospital was found to be upwards of 26 days [36].

6.1. *C. difficile* Epidemiology in Canada

A Canada-wide study from 2005 to 2008 found the top three most prevalent NAP types to be NAP1, NAP2, and NAP4, respectively [11]. In concordance with this study (Figure 6), they observed a decline in NAP2 from 28% prevalence to 6%, whereas NAP4 increased from 5% prevalence to 9% in the same time period [11]. The increased prevalence of NAP4 we observed may be related to community-acquired *C. difficile* infections. There is not a lot of data available regarding the true proportions of community-acquired *C. difficile* infections since it is much more difficult to record these infections outside of strictly regulated hospital environments. Individuals infected with community-acquired strains may become infected through environmental or food sources, but some spores may lie dormant in their gut. The spores may germinate later when the individual is admitted to a hospital and put on a regimen of antimicrobials. NAP4 may be increasing in prevalence in community-acquired sources and we are only beginning to see that increase from the previously infected individuals being admitted to hospitals. The decrease we observed in prevalence of NAP1 may be related to the recent decrease in ciprofloxacin usage in Canada which declined by 21% between 2009 and 2013 (unpublished data). Fluoroquinolone-resistant *C. difficile* was related to major outbreaks in both Canada and the US [53]. With decreased usage of ciprofloxacin (a fluoroquinolone), there would be less selection pressure for fluoroquinolone-resistant strains which would create more opportunities for non-resistant strains to cause infection.

We observed the two most prevalent PFGE types within NAP4, types 0023 and 0033, to have largely contrasting prevalence trends over the study years. From 2004 to 2011, PFGE type 0023 dropped in prevalence from 56% to 13% of all NAP4 isolates collected. In the same time span, PFGE type 0033 increased from 0% to 40% of all NAP4 isolates collected. PFGE type 0033 has maintained this high degree of prevalence since 2009.

When studies analyzed CDI in pediatric populations, NAP4 was found as one of the most prevalent strain types. For example, a nation-wide American study focusing on CDI in pediatric populations found NAP4 to be the third most prevalent NAP type at 11%, behind NAP1 and NAP11 [142]. Conversely, Canada-wide studies have found NAP4 to be the most prevalent NAP type in pediatric populations at 25% prevalence [76, 124]. In the same patient population, NAP4 prevalence has been found as high as 35% in a smaller Canadian study [77]. These reports coincide with this study that shows a much larger proportion of NAP4 isolates (23%) from CDI cases in patients ≤ 18 years of age when compared to NAP1 (2.7%) (Table 13). Differences in age however were not identified when comparing NAP4 PFGE types 0023 and 0033.

6.2. PFGE and Ribotyping

Considering the fact that North America and Europe use different typing methods for identifying *C. difficile* with PFGE NAP typing and ribotyping, respectively, the true global prevalence of NAP4 isolates is unknown. However, correlation of ribotypes with NAP types has been investigated previously in a study by Tenover et al. (2011). Of 350 isolates typed by PFGE and ribotyping, they were able to relate common *C. difficile* ribotypes, including 014, 017, 027, and 078, with specific NAP types [133]. However 30% of the isolates from the Tenover study did not match known ribotyping nomenclature defined by the European Centre for Disease Prevention and Control (ECDC). Approximately 2% of our isolates did not match known

ribotypes in our database. With no proper way to define new or unknown ribotypes it is apparent that global standardized ribotyping nomenclature for *C. difficile* is warranted.

Previously described ribotypes of NAP4 include 014, 020, and 077 [42, 62, 80, 86, 116, 130, 133] and were all observed in our isolate collection. The 23 other ribotypes observed have not been described in the literature, to the best of our knowledge, as being related to a NAP type. Of the previously described NAP4 ribotypes, 014 and 020 were found to be the second most common types of *C. difficile* in the United States behind ribotype 027, which is the predominant ribotype for NAP1. In Europe, ribotypes 014 and 020 have surpassed ribotype 027 to become the most common type of *C. difficile* in community-acquired cases, specifically [42]. From this data it is apparent that NAP4 ribotypes 014 and 020 are emerging as a predominant strain type of *C. difficile* globally. Of the commonly represented ribotypes, 014 and 020, a significantly higher proportion of these ribotypes in either PFGE type 0023 or 0033 was not observed (Table 15). Therefore, neither ribotype appears to be associated with or related to the change in prevalence of these two PFGE types throughout the study years.

6.3. Virulence of NAP4

Severity of CDI can be assessed based on many aspects, including ICU admission, colectomy, or if the infection led to the death of the patient. From this study, 0.3% of patients infected with NAP4 required colectomy (Table 13), which is lower than NAP1 (1.7%). By comparison, Kasper et al. recorded 75 patients from five tertiary-care facilities in the United States that required colectomy due to the CDI from of 8,569 patients with CDI recorded on their charts (0.9%) [66]. As well, a 2008 European study by Bauer et al. found three cases of colectomy due to CDI from 460 patients with recorded CDI (0.7%) from many hospitals across

34 European countries [9]. It should be noted that both the Kasper and the Bauer studies did not specify NAP types for their isolates.

In terms of ICU admission, our data suggests that NAP4 also has an overall lower rate of cases requiring colectomy compared to NAP1 isolates. Only 0.6% of patients with NAP4 required admission to an ICU in comparison to 3.3% of NAP1 cases (Table 13), both of which are lower than the previously reported rate of 7% for *C. difficile* [9].

We determined the 30 day all-cause mortality rate of NAP4 to be 7.7% which is significantly lower than that of NAP1 at 16.9%. This rate is comparable to those previously reported by Bauer et al. (2011), which found the all-cause mortality of CDI in their study to be 7%. Likewise, a study by Hota et al. (2012) that analyzed patient data involving CDI from three Canadian hospitals between 2007 and 2008, reported that the all-cause mortality was 8.8% [56]. It should be noted that the Hota study did not specify NAP types of their isolates. Attributable mortality defines a more specific scenario whereby a case of CDI is determined to be the direct cause of death of the patient. The Bauer and Hota studies observed 2% and 1.4% of their respective CDI cases to be directly attributed to the death of a patient [9, 56] in comparison to 0.9% of NAP4 cases in this study. Quite a significant difference is further noted in our findings when comparing NAP1 cases to NAP4, in which 8.1% of the patients with NAP1 died (Table 13). The severity of CDI may be multi-variable, but the dissimilarity of our data in relation to NAP1 and previous research findings suggest that NAP4 might not be as virulent as other *C. difficile* strain types. Due to the overall limited number of severe cases of NAP4 CDI we could not determine statistically whether one PFGE type, 0023 or 0033, was more virulent than the other.

The toxin PCR profiles of our NAP4 isolates were consistent with profiles previously described for this strain with detectable *tcdA, tcdB,* wild-type *tcdC,* and an absence of *cdtB* [20, 116]. Exceptions being 3 NAP4 isolates with a *tcdC* deletion and 1 that was positive for *cdtB.* In comparison, 98.1% of NAP1 isolates contained a deletion in *tcdC,* which was only 0.9% in NAP4 (Table 14). In NAP1, deletions in the *tcdC* gene were shown to lead to overproduction of toxins A and B upwards of 23 times that of toxin production in isolates without a *tcdC* deletion [71]. As well, binary toxin, produced by the gene *cdtB,* is also associated with increased fatality in *C. difficile* [4]. Although binary toxin is considered to be present in only 22% of all *C. difficile* isolates [30], we found that 99.4% of NAP1 isolates contained this gene in comparison to only 0.3% in NAP4 (Table 14). From this comparison it seems reasonable to speculate that the differences in severity of NAP1 and NAP4 cases are related to these differences in toxin profiles.

6.4. NAP4 Phenotypic Analysis

With the increasing prevalence of NAP4 in Canada, this strain type was investigated further by measuring growth and sporulation and comparing it to isolates of other known NAP-types and between NAP4 PFGE types 0023 and 0033.

6.4.1. Growth Assay

Growth assays performed on all NAP types produced similar growth curves to those previously observed [90, 139] and were not significantly different from each other. For all strain types, the entry into exponential phase was at approximately 3 hours and the entry into stationary phase was achieved around 8 to 9 hours post-inoculation. With the lack of colony counts, growth rates were further analyzed during exponential phase by comparing the slope of the line between 4 and 8 hours. From this analysis, NAP4 did not have a significantly greater growth rate than any of the other NAP types tested. However, the growth rate of PFGE type 0033 isolates was found

to be significantly greater than PFGE type 0023 isolates (Figure 12), which could be one of the factors contributing to its increased prevalence. For instance, a faster growth rate during exponential phase could aid *C. difficile* by allowing it to achieve a fulminant infection of the host sooner. A larger experiment with more isolates and more replicates is needed to confirm this hypothesis.

6.4.2. Sporulation Assay

Few reports describing the sporulation ratio of *C. difficile* have been generated. The average sporulation ratio of NAP4 isolates in this study was 42% and was not significantly greater than any other NAP type tested (Figure 8). As well, our data was similar to previously reported ratios of 42.2% and 87.2% [90]. There was, however, an observable difference between PFGE types 0023 and 0033 whereby PFGE type 0033 appeared to have a greater sporulation rate than PFGE type 0023. However, this observation was not statistically significant (Figure 13). As with the growth assays, additional isolates and further testing is required. The importance of sample size is important because of the previously held notion that NAP1 is associated with higher sporulation ratios. This trait is believed to be one of the major contributors to the dissemination and success of the NAP1 strain type [144]. A study by Burns et al. (2011) using a much larger collection of isolates however, found that not all NAP1 isolates had a greater sporulation ratio relative to other *C. difficile* isolates. With sporulation related to virulence, the study warns against assuming the findings for one or a select few isolates of *C. difficile* to encapsulate all isolates of that strain type [14].

6.5. Antimicrobial Susceptibility Analysis

There is a clear relationship between the use of antimicrobials and CDI whereby the antimicrobials are thought to deplete the host's natural gut microflora which predisposes the gut to colonization by *C. difficile*. The presence of *C. difficile* that are resistant to certain antimicrobials is becoming a concern as resistance can potentially lead to an increase in CDI and decrease favorable patient outcomes, as was the case for the rapid dissemination of fluoroquinolone-resistant NAP1 isolates [53].

The first antimicrobial we investigated for resistance mechanisms amongst our NAP4 isolates was clindamycin. A study by Tang-Feldman et al. related the mechanism of clindamycin resistance in *C. difficile* to the presence of the *ermB* gene, which is a 23S rRNA methylase. We searched for the presence of *ermB* in the genomes of all 85 sequenced isolates using a blast search with a wild-type sequence of the gene and found none. Further searches using ResFinder failed to detect additional resistance genes in any of the NAP4 isolates. In comparison to the study by Tang-Feldam, where the MICs were >256 µg/ml, the MICs of our NAP4 isolates were low (2 to 8 µg/ml). Therefore, the resistance profile of the NAP4 isolates may simply be related to other, less specific mechanisms of resistance to clindamycin. This has been previously reported by Ackermann et al. who suggested that resistance to clindamycin in *ermB* negative *C. difficile* isolates could be the result of yet-to-be-determined mutations in target sequences in the 23S rRNA or that it could be mediated by an unknown efflux system [1].

Research by Walkty et al. has helped further the understanding of moxifloxacin resistance mechanisms in *C. difficile*. These researchers analyzed *C. difficile* isolates for mutations in the genes *gyrA* and *gyrB*. These genes encode subunits of a DNA gyrase that is normally inhibited by fluoroquinolones including moxifloxacin [151]. In the Walkty paper, they

describe several separate mutations that alone or in combination appear to lead to different levels of moxifloxacin resistance in *C. difficile* (Table 23) [141].

Table 23. Mutations in *gyrA* and *gyrB* that are related to fluoroquinolone resistance [141]

Moxifloxacin MIC (µg/ml)	***gyrA* Mutation**	***gyrB* Mutation**
8	Asp71-Val	-
8	-	Asp426-Asn
8	-	Glu466-Val
16 - 32	Thr82-Ile	-
≥64	Thr82-Ile	Ser336-Ala Asp426-Val
≥64	Thr82-Ile	Asp426-Asn
≥64	Thr82-Ile	Leu444-Phe
≥64	Thr82-Ile Asp71-Glu	-
≥64	Thr82-Ile Pro116-Ala	-
≥64	Thr82-Ile Ala118-Ser	-

One of the resistant isolates (21ACD0154) with a MIC of 8 µg/ml had a single *gyrA* mutation (Ala118-Thr), which has not been previously related to moxifloxacin resistance in *C. difficile*. From this study and Waltky et al. it appears that the Thr82-Ile is essential for high level resistance. Finally, one of our isolates was resistant to moxifloxacin, with a MIC of 8 µg/ml, had no *gyrA* or *gyrB* mutations. We believe that this isolate most likely has another mechanism of underlying antimicrobial resistance that has yet to be revealed.

Of these clindamycin and moxifloxacin resistant isolates, an even amount of both NAP4 PFGE type 0023 and 0033 isolates with similar MICs leads us to believe that the difference between these two types is not related to antimicrobial resistance.

6.6. Core SNP Phylogeny

Of all 12 known NAP types, 130,007 unique SNP positions were present in the core genome. A previous genomic comparison of 14 *C. difficile* core genomes done by Forgetta et al. found a total of 127,442 SNPs among them. The study consisted of 8 NAP1 genomes, 3 NAP7/8 genomes, 1 NAP2, strain 630, and 1 reference strain from ATCC. The NAP2, strain 630, and ATCC genomes were placed into a remainder group (R-group). Similar to this study, the NAP7/8 group had the largest number of SNPs (89,954) separating it from the other groups. Group inclusive SNP differences were observed in the NAP7/8 group and NAP1 group at 1,851 and 62 SNPs, respectively [35]. In this study, group inclusive SNP differences amongst the 85 NAP4 genomes was 8, 353 SNPs. This high variation is likely attributed to the 7 year timeline, geographic distribution across Canada, and the much greater number of isolates included in this study. To determine the true SNP variation of the core genome of each NAP type would require many isolates of specific NAP types to be analysed at one time.

The 2,569 core genes we found in the genomes of our 85 NAP4 isolates is lower than previous reports by Forgetta et al. (2011), describing 3,063 core genes [35], and Stabler et al. (2009), describing 3,247 core genes [128]. The Forgetta study included only 14 genomes whereas the Stabler study included only 3 isolates. We would expect a collection of genomes from one single NAP type to have a core genome with a larger number of genes compared to one constructed from collections of genomes with a variety of NAP types. The opposite is seen here, whereby the Forgetta and Stabler papers describe more core genes from collections of genomes of different NAP types compared to our collection of only NAP4 genomes. The case again may simply be that by including fewer genomes of each NAP type the gene counts found by the Forgetta and Stabler studies do not represent the true core gene counts of *C. difficile*.

6.6.1. Outlier Analysis

The 5 outliers in our core SNP tree were confirmed to consist of four PFGE type 0023 isolates and one PFGE type 0033 isolate. The ribotypes of three outliers were rare considering that those ribotypes only appeared once out of the 338 NAP4 isolates typed with ribotyping. Those ribotypes were 207, 503, and ns11. Similarly, the in silico MLST revealed two outliers that had a sequence type unique to all 338 NAP4 isolates in our study, of which both were 1 allele different from the predominant ST-2. Only two outliers had both a unique ribotype and a unique MLST type.

6.7. Accessory Genome Analysis

Further variation amongst NAP4 genomes may be present in the accessory genome, which is not represented in the core genome phylogenetic tree, such as phages or mobile genetic elements. In a study by Garneau et al. they found most of the diversity between their 81 NAP1 genomes was accounted for by the presence of various prophage elements found by PHAST [39]. The same might be the case for our PFGE type 0023 and 0033 isolates as a plethora of different phages were identified. .

6.8. Differentiation of PFGE Banding Patterns

The difference in the NAP4 PFGE banding patterns of PFGE type 0023 and 0033 revealed the band shift was due to the insertion of a phage element measuring 49,313 bp in length. The known phage genome it matched to, phiCD505, measures 49,316 bp in length and was confirmed as only being present in the PFGE type 0033 isolates. Several genes found within this phage, not seen in any PFGE type 0023 isolates, included hypothetical and putative transcriptional regulators. The transcriptional regulators could have many possible downstream effects on cell activity and chromosomal gene expression. As previously reported, a putative

gene repressor in a phage genome was shown to downregulate the expression of *C. difficile tcdA* and *tcdR* genes of the pathogenicity locus [46].

6.8.1. Phage Insert Location Investigation

A great deal of research can be done to investigate the genetic content of a phage genome and there is a myriad of ways the genes of a lysogenic phage can affect a host cell, both positively and negatively. A lysogenic phage can affect a host by inserting into or nearby a gene in the host genome. This has been described in *C. difficile* by observing the preferred integration sites of the clostridium phage phiCD27. This phage appears to integrate into one of two sites in the *C. difficile* genome. The first site is the open reading frame of a putative ATPase gene the product of which is used by an ABC transporter. The second site is in the open reading frame of another putative ATPase gene which affects a flagellar protein export apparatus [147]. The disruption of such genes is expected to have deleterious downstream effects on many cellular functions.

There appears to be a phage inserted nearby the GMP synthase gene in all PFGE type 0033 genomes, but none of the PFGE type 0023 genomes. The effect of overexpression and inhibited expression of this gene in *C. difficile* has been described in a study by Purcell et al. (2012). Here they describe the inverse relationship between the amounts of GMP synthase product, c-di-GMP, and cell motility. Increased expression of GMP synthase, producing more c-di-GMP than normal, promotes cell clumping which may have an effect on the ability of *C. difficile* to form biofilms. Conversely, a decrease in the expression of GMP synthase led to an increase in motile phenotypes [111]. With an effect on either biofilm formation or motility, a change to GMP synthase expression could affect virulence of *C. difficile*. In our case, this would be expected to impact specifically PFGE type 0033 isolates. It is unknown how the insertion

affects the GMP synthase expression in our PFGE type 0033 isolates. To test this, the levels of c-di-GMP would first need to be measured in PFGE type 0033 isolates and compared to levels in isolates without a phage near the GMP synthase gene, such as PFGE type 0023.

These phages were found in *C. difficile* PFGE type 0033 isolates collected as early as 2007. The fact that no PFGE type 0033 *C. difficile* isolates from the 2004/2005 study year were observed makes it difficult to determine when this phage insertion might have happened. If it is a relatively recent event, and either regulatory genes on the phage or disruption of GMP synthase increases the fitness of NAP4, then that might explain the recent increase in prevalence of PFGE type 0033 across Canada. A more in depth analysis on the evolutionary history of this PFGE type in particular would help to confirm this.

6.8.2. Phage Protein Ortholog Phylogeny

To date, 34 known clostridium phages have been documented. The sizes of these phage genomes range from 17,972 bp to 185,683. It should be noted that the second largest clostridium phage genome is 59,199 bp long. We decided to characterize the phage elements in our 85 genomes by their proteomic relatedness to each other, to known clostridium phage genomes. Of the 123 phages defined as intact by PHAST most did not cluster based on the presence of orthologous genes within the phage genomes.

We only observed two orthoclusters. One orthocluster, labelled orthocluster1, contained 16 phages from our 85 *C. difficile* isolates that clustered with a known clostridium phage, phiCD505. The 16 phages from this orthocluster were from 16 different *C. difficile* isolates, all PFGE type 0033. The second orthocluster, labelled orthocluster2, contained 22 phages. This orthocluster did not group with a known phage. The phages in this orthocluster were also from individual *C. difficile* isolates, but the isolates were a mix of PFGE type 0023 and 0033. A high

amount of variation between the other 85 phages in the ortholog tree was obvious due to the lack of clustering amongst themselves or with the known phages. A similar case was observed in a study by Hargreaves et al. who used orthoclustering to group new clostridium phages with known ones. This variability amongst clostridium phages may be due to the high degree of recombination between phages. Clostridium phage genomes have been previously described as containing multiple gene cassettes. That is, certain genes have been observed in multiple phage genomes, but not in the same order, orientation, or combination as that seen in other phage genomes [52]. This could help explain the lack of orthoclustering among most of the intact phages found amongst our 85 NAP4 genomes.

The presence of these different phages found amongst all of our NAP4 isolates could add to the genetic variation between those isolates and possibly be contributing to the evolution and dissemination of NAP4 across Canada.

7. Limitations

This study only began to analyze phage elements. A more in-depth characterization of the different types of clostridium phages could yet reveal more diversity between the two PFGE types. Along with the theme of the accessory genome, we did not look into the presence of mobile elements or integrative and conjugative elements (ICEs) in our genomes.

8. Future work

For future work, it is suggested that the effect of the phage insertion close to the GMP synthase gene should be determined. Expression of this gene with the phage insertion present should be quantified. If the expression levels differ from the wild type, motility and biofilm assays present a simple continuation to correlate the findings to those of Purcell et al. (2012) [111].

The effect of the putative transcriptional regulators unique to PFGE type 0033 should also be investigated. Such work could include gene knockout experiments of the putative regulatory genes. Measuring expression of toxin-related genes *tcdA* and *tcdR* in PFGE type 0033 isolates in the knockout experiment would be a suitable start to the investigation. Results could be compared to that of Govind et al. (2009) [46].

9. Conclusions

Ever since the major outbreaks in the early 2000's, *C. difficile* has steadily become classified as one of the largest threats to public health in Canada and the US. Most of the focus has been on the NAP1 type of *C. difficile* that caused the original outbreaks. A steady increase in the prevalence of the relatively under-studied NAP4 type throughout Canada, however, has not been investigated. Understanding the capabilities of this strain of *C. difficile* would help us prepare for and manage potential future outbreaks as well as broaden our knowledge of this rising public health threat.

To date, aside from prevalence data, studies describing NAP4 isolates have not gone further than antibiotic susceptibility profiling [20, 21, 47, 65, 76, 113, 141] and ribotyping [42, 62, 80, 86, 116, 130, 133, 134]. Recent research of *C. difficile* utilizing genomic analysis has portrayed the genome to be promiscuous. Even the core genome of isolates with the same

ribotype has shown variation with evidence of horizontal gene transfer, homologous recombination, and single nucleotide polymorphisms [54] including entire phage islands [54, 128]. In this study differentiating the two most common PFGE types of NAP4 in Canada, increased growth capability amongst the PFGE type 0033 isolates could be contributing to the emergence of this PFGE type overtime. Through genomic analysis several the major differences were observed between the two PFGE types. The insertion of a phage, phiCD505, near a GMP synthase gene and the presence of two putative transcriptional regulators along with a hypothetical protein were unique to PFGE type 0033 isolates. The effect these differences have on the changing prevalence rates of PFGE types 0033 and 0023 of NAP4 remain to be determined.

10. References

1. **Ackerman, G., Degner, A., Cohen, S. et al.** 2003. Prevalence and association of macrolide–lincosamide–streptogramin B (MLSB) resistance with resistance to moxifloxacin in *Clostridium difficile*, *Journal of Antimicrobial Chemotherapy*, 51(3): 599-603

2. **Anderson, R., Groundwater, P., Todd, A., et al.** 2012. Chapter 2.3 and Chapter 5.2. Pg. 85-96 and 305-316 in: Anderson, R., Groundwater, P., Todd, A., and Worsley, A. (eds.) *Antibacterial Agents: Chemistry, Mode of Action, Mechanisms of Resistance and Clinical Applications*. West Sussex, United Kingdom: John Wiley & Sons, Ltd.

3. **Au, K., Underwood, J., Lee, L. et al.** 2012. Improving PacBio Long Read Accuracy by Short Read Alignment. *PLoS One*, 7: 1-8

4. **Bacci, S., Mølbak, K., Kjeldsen, M. et al.** 2011. Binary toxin and death after *Clostridium difficile* infection, *Emerging Infectious Diseases*, 17(6): 976-982

5. **Bankevich, A., Nurk, S., Antipov, D., et al.** 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing, *Journal of Computational Biology*, 19(5): 455-477

6. **Barbut, F., Richard, A., Hamadi, K. et al.** 2000. Epidemiology of Recurrences or Reinfections of *Clostridium difficile*-Associated Diarrhea, *Journal of Clinical Microbiology*, 38(6): 2386-2388

7. **Bartlett, J.** 2008. Historical Perspectives on Studies of *Clostridium difficile* and *C. difficile* Infection, *Clinical Infectious Diseases*, 46, S4-S11

8. **Bartlett, J., Chang, T., Gurwith, M. et al.** 1978. Antibiotic-Associated Pseudomembranous Colitis Due to Toxin-Producing Clostridia, *New England Journal of Medicine*, 298: 531-534

9. **Bauer, M., Notermans, D., Benthem, B. et al.** 2011. *Clostridium difficile* infection in Europe: a hospital-based survey, *The Lancet*, 377(9759): 63-73

10. **Boyce J., Havill N., Otter J., et al.** 2008. Impact of hydrogen peroxide vapor room decontamination on *Clostridium difficile* environmental contamination and transmission in a healthcare setting, *Infection Control and Hospital Epidemiology*, 29(8):723-729

11. **Boyd, D., Miller, M., Gravel, D. et al.** 2009. Dynamic changes in molecular epidemiology of *Clostridium difficile* from inpatients at Canadian hospitals, 2005 to 2008, *Abstracts from the 26th International Congress of Chemotherapy and Infection*, 34(2): S47

12. **Burckhardt, F., Friedrich, A., Beier, D. et al.** 2008. *Clostridium difficile* Surveillance Trends, Saxony, Germany; *Emerging Infectious Diseases*, 14(4): 691-692

13. **Burland, T.** 2000. DNASTAR's Lasergene sequence analysis software, *Methods in Molecular Biology*, 132: 71-91

14. **Burns, D., Heeg, D., Cartman, S., et al.** 2011. Reconsidering the Sporulation Characteristics of Hypervirulent *Clostridium difficile* BI/NAP1/027, *PLoS ONE*, 6(9): e24894

15. **Chakra, C., Pepin, J., Sirard, S., et al.** 2013. Risk Factors for Recurrence, Complications and Mortality in *Clostridium difficile* Infection: A Systematic Review, *PLoS ONE*, 9(6): e98400

16. **Chen, F., Mackey, A., Stoeckert, M., et al.** 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups, *Nucleic Acids Research*, 34: D363-368

17. **Chen, F., Mackey, A., Vermunt, J. et al.** 2007. Assessing Performance of Orthology Detection Strategies Applied to Eukaryotic Genomes, *PLoS ONE*, 2(4): e383

18. **Cohen, S. Gerding, D., Johnson, S., et al.** 2010. Clinical Practice Guidelines for *Clostridium difficile* Infection in Adults: 2010 Update by the Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA), *Infection Control and Hospital Epidemiology*, 31(5): 431-455

19. **Comité de surveillance provinciale des infections nosocomiales (SPIN-CD).** 2015. Context. Pg. 1 in: Longtin, Y., Tremblay, C., Galameau, L., Baeulieu, F., Bolduc, D., Frenette, C., Garenc, C., Lévesque, S., Lalancette, C., Vachon, J., Loo, V., and Trudeau, M. (eds.), *Protocole de la Surveillance des diarrhées associées au Clostridium difficile dans les centres hospitaliers du Québec*, Institut national de santé publique du Québec (INSPQ), https://www.inspq.qc.ca/sites/default/files/documents/infectionsnosocomiales/2016-02-01_protocole_dacd.pdf, 2015-last update

20. **Costa, C.** 2014. *Clostridium difficile* : incidência da infecção e caracterização das cepas isoladas de pacientes com diarreia internados em um hospital oncológico de Fortaleza, Ceará, Doctoral dissertation, Universidade Federal do Ceará

21. **Costa, C., Quesada-Gómez, C., Mano de Carvalho, C. et al.** 2014. Community-acquired diarrhea associated with *Clostridium difficile* in an HIV-positive cancer patient: first case report in Latin America, *International Journal of Infectious Diseases*, 26: 138-139

22. **Cuccuru, G., Orsini, M., Pinna, A. et al.** 2014. Orione, a web-based framework for NGS analysis in microbiology, *Bioinformatics*, 30(13): 1928-1929

23. **Darling, A., Mau, B., Blattner, F., et al.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements, *Genome Research*, 14(7): 1394-1403

24. **De La Cochetière, M., Durand, T., Lalande, V. et al.** 2008. Effect of Antibiotic Therapy on Human Fecal Microbiota and the Relation to the Development of *Clostridium difficile*, Microbial Ecology, 56: 395-402

25. **Department of Health and Health Protection Agency**. 2008. *Clostridium difficile* infection: How to deal with the problem. Pg. 86-90 in: *Healthcare Associated Infection and Antimicrobial Resistance*

26. **Dethlefsen, l. and Relman, D.** 2011. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation, *Proceedings of the National Academy of Sciences in the United States of America*, 108(1): 4554-4561

27. **Dhillon, B., Laird, M., Shay, J., et al.** 2015. IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis, *Nucleic Acids Research*, 43(W1): W104-W108

28. **Dial, S., Delaney, C., Schneider, V., et al.** 2006. Proton pump inhibitor use and risk of community-acquired *Clostridium difficile*-associated disease defined by prescription for oral vancomycin therapy, *Canadian Medical Association Journal*, 175(7): 745-748

29. **Dubberke, E. and Olsen, M.** 2012. Burden of *Clostridium difficile* on the Healthcare System, *Clinical Infectious Diseases*, 55(2): S88-S92

30. **Eckert, C., Coignard, B., Hebert, M. et al.** 2013. Clinical and microbiological features of *Clostridium difficile* infections in France: The ICD-RAISIN 2009 national survey, *Médecine et Maladies Infectieuses*, 43(2): 67-74

31. **Eglow, R., Pothoulakis, C., Itzkowitz, S.**, et al. 1992. Diminished *Clostridium difficile* Toxin A Sensitivity in Newborn Rabbit Ileum Is Associated with Decreased Toxin A Receptor, *The Journal of*

Clinical Investigation, 90(3): 822-829

32. **Eyre, D., Golubchik, T., Gordon, N. et al.** 2012. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open*, 2: 1-9

33. **Fawley, W., Knetsch, C., MacCannell, D. et al.** 2015. Development and Validation of an Internationally-Standardized, High-Resolution Capillary Gel-Based Electrophoresis PCR-Ribotyping Protocol for *Clostridium difficile*, *PLoS One*, 10(2): e0118150

34. **Food Safety News**, 2009, Genetic Testing. Available: http://www.foodsafetynews.com/2009/08/genetic-testing-1/#.VONf9EC8DTp 2015-last update.

35. **Forgetta, V., Oughton, M., Marquis, P. et al.** 2011. Fourteen-Genome Comparison Identifies DNA Markers for Severe-Disease-Associated Strains of *Clostridium difficile*, *Journal of Clinical Microbiology*, 49(6): 2230-2238

36. **Forster, A., Taljaard, M., Oake, N. et al.** 2011. The effect of hospital-acquired infection with *Clostridium difficile* on length of stay in hospital, *Canadian Medical Association Journal*, doi: 10.1503/cmaj.110543

37. **Francisco, A., Vaz, C., Monteiro, P. et al.** 2012. PHYLOViZ: Phylogenetic inference and data visualization for sequence based typing methods, *BMC Bioinformatics*, 13(1): 87

38. **Fry, D.** 2013. Surgical Infections, *JP Medical Ltd.*, 83 Victoria Street, London, UK, (11): 123

39. **Garneau, J., Monot, M., Valiquette, L., et al.** 2016. Prophage elements as a significant source of genetic diversity among epidemic R027 clinical isolates of *Clostridium difficile*, *University of Sherbrooke*, presented at the Canadian Society for Microbiologists annual conference, 2016

40. **Garrison, E. and Marth, G.** 2012. Haplotype-based variant detection from short-read sequencing, *arXiv*:1207.3907

41. Gene accumulation curves in R, https://web.archive.org/web/20120725061000/http://smokeandumami.com/2010/01/21/gene-accumulation-curves-in-r/, 2010-last update

42. **Gerding, D. and Lessa, F.** 2015. The Epidemiology of *Clostridium difficile* Infection Inside and Outside Health Care Institutions, pg. 42 in: Wilcox, M. (ed.), *Clostridium difficile Infection, An Issue of Infectious Disease Clinics of North America*. Elsevier Health Sciences

43. **Gerding, D., Meyer, T., Lee, C., et al.** 2015. Administration of Spores of Nontoxigenic *Clostridium difficile* Strain M3 for Prevention of Recurrent C difficile Infection A Randomized Clinical Trial, The Journal of the American Medical Association, 313(17): 1719-1727

44. **Geric, B., Rupnik, M., Gerding, D. et al.** 2004. Distribution of *Clostridium difficile* variant toxinotypes and strains with binary toxin genes among clinical isolates in an American hospital, *Journal of Medical Microbiology*, 53: 887-894

45. **Gish, W. and States, D.** 1993. Identification of protein coding regions by database similarity search, *Nature Genetics*, 3:266-27

46. **Govind, R., Vediyappan, G., Rolfe, R. et al.** 2009. Bacteriophage-Mediated Toxin Gene Regulation in *Clostridium difficile*, *Journal of Virology*, 83(23): 12037-12045

47. **Gravel, D., and Miller, M.** 2007. *Clostridium difficile* associated diarrhea in acute-care hospitals participating in CNISP: November 1, 2004 to April 30, 2005; Montreal PQ: Canadian Nosocomial

Infection Surveillance Program (CNISP. PCSIN)

48. **Gravel, D., Miller, M., Simor, A. et al.** 2009. Health care-associated *Clostridium difficile* infection in adults admitted to acute care hospitals in Canada: a Canadian Nosocomial Infection Surveillance Program Study, *Clinical Infectious Diseases*, 48(5): 568-576

49. **Griffiths, D., Fawley, W., Kachrimanidou, M. et al.** 2010. Multilocus Sequence Typing of *Clostridium difficile*. *Journal of Clinical Microbiology*, 48: 770-778

50. **Guindon, S. and Gascuel, O.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, *Systematic Biology*, 52(5):696-704

51. **Hall, I. and O'Toole, E.** 1935. Intestinal flora in new-born infants: with a description of a new pathogenic anaerobe, Bacillus difficilis, *American Journal of Diseases in Children*, 49(2): 390-402

52. **Hargreaves, K. and Clokie, M.** 2014. *Clostridium difficile* phages: still difficult?, *Frontiers in Microbiology*, 5:184

53. **He, M., Miyajima, F., Roberts, P. et al.** 2012. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nature Genetics*, 45: 109-114

54. **He, M., Sebaihia, M., Lawley, T. et al.** 2010. Evolutionary dynamics of *Clostridium difficile* over short and long time scales, *PNAS*, 107(16): 7527-7532

55. **Hensgens, M., Goorhuis, A., Dekkers, O. et al.** 2012. Time interval of increased risk for *Clostridium difficile* infection after exposure to antibiotics, *The Journal of Antimicrobial Chemotherapy*, 67(3):742-748

56. **Hota, S., Achonu, C., Crowcroft, N. et al.** 2012. Determining Mortality Rates Attributable to *Clostridium difficile* Infection, *Emerging Infectious Diseases*, 18(2): 305-307

57. **Huang, H., Weintraub, A., Fang, H., et al.** 2009. Antimicrobial resistance in *Clostridium difficile*, *International Journal of Antimicrobial Agents*, 34: 516-522

58. **Huber, A., Foster, N., Riley, T. et al.** 2013. Challenges for Standardization of *Clostridium difficile* Typing Methods, *Journal of Clinical Microbiology*, 51: 2810-2814

59. **Ibrahim, K., Gunderson, B., Hermeson, E., et al.** 2004. Pharmacodynamics of Pulse Dosing versus Standard Dosing: In Vitro Metronidazole Activity against Bacteroides fragilis and Bacteroides thetaiotaomicron, *Antimicrobial Agents and Chemotherapy*, 48(11): 4195-4199

60. **Indra, A., Huhulescu, S., Schneeweis, M. et al.** 2008. Characterization of *Clostridium difficile* isolates using capillary gel electrophoresis-based PCR ribotyping. *Journal of Medical Microbiology*, 57: 1377-1382

61. **Janezic, S., and Rupnik, M.** 2010. Molecular Typing Methods for *Clostridium difficile*: Pulsed-Field Gel Electrophoresis and PCR Ribotyping. Pg. 55-66 in: Mullany, P. & Roberts, A. (eds.), *Clostridium difficile Methods and Protocols*. Hertfordshire, UK: Humana Press.

62. **Jassem, A., Prystajecky, N., Marra, F. et al.** 2016. Characterization of *Clostridium difficile* Strains in British Columbia, Canada: A Shift from NAP1 Majority (2008) to Novel Strain Types (2013) in One Region, *Canadian Journal of Infectious Diseases and Medical Microbiology*, Article ID 8207418

63. **Jensen, M., Engberga, J., Larsson, J. et al.** 2015. Novel multiplex format of an extended multilocus variable-number-tandem-repeat analysis of *Clostridium difficile* correlates with tandem repeat sequence typing. *Journal of Microbiological Methods*, 110: 98-101

64. **Jolley, K.**, 2015, *Clostridium difficile* Multi Locus Sequence Typing. Available: http://pubmlst.org/cdifficile/ 2015-last update.

65. **Karlowsky, J., Zhanel, G., Hammond, G. et al.** 2012. Multidrug-resistant North American pulsotype 2 *Clostridium difficile* was the predominant toxigenic hospital-acquired strain in the province of Manitoba, Canada, in 2006–2007, *Journal of Medical Microbiology*, 61: 693-700

66. **Kasper, A., Nyazee, H., Yokoe, D. et al.** 2012. A Multicenter Study of *Clostridium difficile* Infection-Related Colectomy, 2000-2006; *Infection Control and Hospital Epidemiology*, 33(5): 470-476

67. **Khanna, S., Pardi, D., Aronson, S. et al.** 2012. The Epidemiology of Community-Acquired *Clostridium difficile* Infection: A Population Based Study, *The American Journal of Gastroneterology*, 107: 89-95

68. **Killgore, G., Thompson, A., Johnson, S. et al.** 2008. Comparison of Seven Techniques for Typing International Epidemic Strains of *Clostridium difficile*: Restriction Endonuclease Analysis, Pulsed-Field Gel Electrophoresis, PCR-Ribotyping, Multilocus Sequence Typing, Multilocus Variable-Number Tandem-Repeat Analysis, Amplified Fragment Length Polymorphism, and Surface Layer Protein A Gene Sequence Typing, *The Journal of Clinical Microbiology*, 46: 431-437

69. **Knetsch, W., Lawley, T., Hensgens, M. et al.** 2013. Current application and future perspectives of molecular typing methods to study *Clostridium difficile* infections, *Eurosurveillance*, 18(7): 1-11

70. **Kuijper, E., Coignard, B., T*ü*ll, P. et al.** 2006. Emergence of *Clostridium difficile*-associated disease in North America and Europe, Clinical Microbiology and Infection, 12(6): 2-18

71. **Kuijper, E., van Dissel, J., and Wilcox, M.** 2007. *Clostridium difficile*: changing epidemiology and new treatment options, *Current Opinion in Infectious Diseases*, 20(4): 376-383

72. **Kurka, H., Ehrenreich, A., Ludwig, W. et al.** 2014. Sequence Similarity of *Clostridium difficile* Strains by Analysis of Conserved Genes and Genome Content Is Reflected by Their Ribotype Affiliation. *PLoS One*, 9: 1-11

73. **Kurtz, S., Phillippy, A., Delcher, A., et al.** 2004. Versatile and open software for comparing large genomes, *Genome Biology*, 5:R12

74. **Kyne L., Warny M., Qamar A. et al.** 2001. Association between antibody response to toxin A and protection against recurrent *Clostridium difficile* diarrhoea, *Lancet*, 357:189-193

75. **Lancaster, J. and Matthews, S.** 2012. Fidaxomicin: The Newest Addition to the Armamentarium Against *Clostridium difficile* Infections, *Clinical Therapeutics*, 34(1): 1-13

76. **Le Saux, N., Gravel, D., Mulvey, M. et al.** 2015. Healthcare-Associated *Clostridium difficile* Infections and Strain Diversity in Pediatric Hospitals in the Canadian Nosocomial Infection Surveillance Program, 2007–2011, *Journal of the Pediatric Infectious Diseases Society*, 4(4): e151-e154

77. **Le Saux, N., Gravel, D., Mulvey, M. et al.** 2016. Pediatric *Clostridium difficile* infection: 6-year active surveillance in a defined patient population, *Infection Control & Hospital Epidemiology*, 35(7): 904-906

78. **Lemee, L., Dhalluin, A., Pestel-Caron, M. et al.** 2004. Multilocus Sequence Typing Analysis of Human and Animal *Clostridium difficile* Isolates of Various Toxigenic Types. *Journal of Clinical Microbiology*, 42: 2609-2617

79. **Lemey, P., Rambaut, A., Drummond, A. et al.** 2009. Bayesian Phylogeography Finds Its Roots, *PLoS Computational Biology*, 5(9): e1000520

80. **Lessa, F., Mu, Y., Bamberg, W. et al.** 2015. Burden of *Clostridium difficile* Infection in the United States, *New England Journal of Medicine*, 372: 825-834

81. **Lessa, F., Yi, M., and Bamberg, W.** 2015. Burden of *Clostridium difficile* Infection in the United States, The *New England Journal of Medicine*, 372(9): 825-834

82. **Levy, A., Szabo, S., Lozano-Ortega, G. et al.** 2015. Incidence and Costs of *Clostridium difficile* Infections in Canada, *Open Forum Infectious Diseases*, 2(3): ofv076

83. **Li, H.** 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, *Bioinformatics*, 27(21): 2987-2993

84. **Li, H., Handsaker, B., Wysoker, A., et al.** 2009. The Sequence alignment/map (SAM) format and SAMtools, *Bioinformatics*, 25(16): 2078-2079

85. **Li, L., Stoeckert, C., and Roos, D.** 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes, *Genome Research*, 13: 2178-2189

86. **Limbago, B., Long, C., Thompson, A. et al.** 2009. *Clostridium difficile* Strains from Community-Associated Infections, *Journal of Clinical Microbiology*, 47(9): 3004-3007

87. **Ling, L., Schneider, T., Peoples, A., et al.** 2015. A new antibiotic kills pathogens without detectable resistance, *Nature*, 517: 455-459

88. **Loman, N., Misra, R., Dallman, T. et al.** 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, 30: 434-441

89. **Lowy, I., Molrine, D., Leav, B. et al.** 2010. Treatment with Monoclonal Antibodies against *Clostridium difficile* Toxins, *The New England Journal of Medicine*, 362(3): 197-205

90. **Lynch, T., Chong, P., Zhang, J. et al.** 2013. Characterization of a Stable, Metronidazole-Resistant *Clostridium difficile* Clinical Isolate, *PLOS One*, 8(1): e53757

91. **Marinier, E. Zaheer, R., Berry, C. et al.** 2015. Neptune: A Tool for Rapid Microbial Genomic Signature Discovery, *bioRxiv*

92. **Maroo, S. and Lamont, J.** 2006. Recurrent *Clostridium difficile*. *Gastroenterology*, 130:1311-1316

93. **Mayfield J., Leet T., Miller J., et al.** 2000. Environmental control to reduce transmission of *Clostridium difficile*, *Clinical Infectious Diseases*, 31(4): 995-1000

94. **McDonald, L., Owings, M., and Jernigan D.** 2006. *Clostridium difficile* infection in patients discharged from US short-stay hospitals, 1996-2003; *Emerging Infectious Diseases*, 12(3): 409-415

95. **Meader, E., Mayer, M., Steverding, D. et al.** 2013. Evaluation of bacteriophage therapy to control *Clostridium difficile* and toxin production in an in vitro human colon model system, *Anaeroobe*, 22: 25-30

96. **Microsoft Research**. 2016. Fisher's Exact Test Calculator for 2x2 Contingency Tables, http://research.microsoft.com/en-us/um/redmond/projects/MSCompBio/FisherExactTest/, 2016-last update

97. **Milne, I., Stephen, G., Bayer, M., et al.** 2013. Using Tablet for visual exploration of second-generation sequencing data, *Briefings in Bioinformatics*, 14(2): 193-202

98. **Monot, M.** 2014. Whole Genome Sequence and Molecular Characterization of Siphoviridae / Myoviridae Phage Infecting *Clostridium difficile*, Submitted to the EMBL/GenBank/DDBJ databases

99. **Nelson, R., Kelsey, P., Leeman, H., et al.** 2011. Antibiotic treatment for *Clostridium difficile*-associated diarrhea in adults, The Cochrane Database of Systematic Reviews, 9: CD004610

100. **Ontario Agency for Health Protection and Promotion, Provincial Infectious Diseases Advisory Committee**. 2013. Annex C – Testing, Surveillance and Management of *Clostridium difficile*. Annexed to: Routine Practices and Additional Precautions in All Health Care Settings. Toronto, ON: Queen's Printer for Ontario.

101. **Paredes,-Sabja, D., Shen, A., and Sorg, J.** 2014. *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins; *Trends in Microbiology*, 22(7): 406-416

102. **Pépin, J., Saheb, N., Coulombe, M. et al.** 2005. Emergence of Fluoroquinolones as the Predominant Risk Factor for *Clostridium difficile*–Associated Diarrhea: A Cohort Study during an Epidemic in Quebec, *Clinical Infectious Diseases*, 41(9): 1254-1260

103. **Pépin, J., Valiquette, L., Alary, M. et al.** 2004. *Clostridium difficile*-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity, *Canadian Medical Association Journal*, 171(5): 466-472

104. **Pépin, J., Valiquette, L., and Cossette, B.** 2005. Mortality attributable to nosocomial *Clostridium difficile*–associated disease during an epidemic caused by a hypervirulent strain in Quebec, *Canadian Medical Association Journal*, 173(9): 1037-1042

105. **Persson, S., Jensen, J., and Olsen, K.** 2011. Multiplex PCR Method for Detection of *Clostridium difficile tcdA, tcdB, cdtA*, and *cdtB* and Internal In-Frame Deletion of *tcd*C, *Journal of Clinical Microbiology*, 49(12): 4299-4300

106. **Petkau, A., Van Domselaar, G., Mabon, P., et al.** 2015. Pipeline for identifying core SNPs and building a phylogenetic tree, Available: https://github.com/apetkau/core-phylogenomics.git

107. **Ponstingl, H.** 2011. SMALT manual, Available: http://www.sanger.ac.uk/science/tools/smalt-0

108. **Popoff, M., Rubin, E., Gill, D.et al.** 1988. Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain, *Infection and Immunity*, 56:2299-2306

109. **Public Health Agency of Canada**, 2013, Infection Prevention and Control Guidance for Management in Acute Care Settings. http://www.phac-aspc.gc.ca/nois-sinp/guide/c-dif-acs-esa/index-eng.php, 2013-last update

110. **Public Health Agency of Canada**, 2013, The Chief Public Health Officer's Report on the State of Public Health in Canada, 2013 Infectious Disease—The Never-ending Threat. http://www.phac-aspc.gc.ca/cphorsphc-respcacsp/2013/infections-eng.php, 2013-last update.

111. **Purcell, E., McKee, R., McBride, S. et al.** 2012. Cyclic Diguanylate Inversely Regulates Motility and Aggregation in *Clostridium difficile*, *Journal of Bacteriology*, 194(13): 3307-3316

112. **Quail, M., Smith, M., Coupland, P. et al.** 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13: 1-13

113. **Quesada-Gómez, C., López-Ureñaa, D., Acuña-Amador, L. et al.** 2015. Emergence of an Outbreak-Associated *Clostridium difficile* Variant with Increased Virulence, *Journal of Clinical Microbiology*, 53(4): 1216-1226

114. **Rainey, F., Hollen, B., and Small, A.** 2009. Genus I. Clostridium. Pg. 738,772-773 in: De Vos, P., Garrity, G., Jones, D., Kreig, N., Ludwig, W., Rainey, F., Schleifer, K., and Whitman, W. (eds.) *Bacteriology Bergey's Manual of Systemic, Second Ed., Vol. 3*, Springer Dordrecht Heidelberg London

New York

115. **Ramesh, V., Fralick, J., and Rolfe, R.** 1999. Prevention of *Clostridium difficile*-induced ileocecitis with Bacteriophage, *Anaerobe*, 5: 69-78

116. **Rodriguez-Palacios, A., Reid-Smith, R., Staempfli, H. et al.** 2009. Possible Seasonality of *Clostridium difficile* in Retail Meat, Canada; *Emerging Infectious Diseases*, 15(5): 802-805

117. **RStudio Team.** 2015. RStudio: Integrated Development for R, *RStudio, Inc.*, Boston, MA, Available: http://www.rstudio.com

118. **Rupnik, M.** 2010. *Clostridium difficile* Toxinotyping. Pg. 67-75 in: Mullany, P. & Roberts, A. (eds.) *Clostridium difficile Methods and Protocols*. Hertfordshire, UK: Humana Press.

119. **Rupnik, M.** 2011. *Clostridium difficile* Toxinotypes. Available: http://www.mf.uni-mb.si/tox/ 2011-last update.

120. **Rutala, W., Weber, D., and the Healthcare Infection Control Practices Advisory Committee (HICPAC)**. 2008. Guideline for Disinfection and Sterilization in Healthcare Facilities, 2008, *The Center for Disease Control and Prevention*, http://www.cdc.gov/hicpac/pdf/guidelines/Disinfection_Nov_2008.pdf, 2008-last update

121. **Sabat, A., Budimir, A., Nashev, D., et al.** 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Eurosurveillance*, 18(3): 1-15

122. **Sanger, F., Nicklen, S. & Coulson, R.** 1977. DNA Sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Science*, 74: 5463-5467

123. **Schwan, C., Stecher, B., Tzivelekidis, T. et al.** 2009. *Clostridium difficile* toxin *CDT* induces formation of microtubule-based protrusions and increases adherence of bacteria, *PLoS Pathogens*, 5(10): e1000626

124. **Schwartz, K., Darwish, I., Richardson, S. et al.** 2014. Severe clinical outcome is uncommon in *Clostridium difficile* infection in children: a retrospective cohort study, *BMC Pediatrics*, 14:28

125. **Seemann, T.** 2014. Prokka: rapid prokaryotic genome annotation, *Bioinformatics*, 30(14): 2068-2069

126. **Shim, J., Johnson, S., Samore, M. et al.** 1998. Primary symptomless colonisation by *Clostridium difficile* and decreased risk of subsequent diarrhoea, *Lancet*, 351: 633-636

127. **Singelton, P.** 2000. Nucleic-Acid-Based Typing. Pg. 181 in: *DNA Methods in Clinical Microbiology*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

128. **Stabler, R., He, M., Dawson, L. et al.** 2009. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium, *Genome Biology*, 10:R102

129. **Stare, G. and Rupnik, M.** 2010. *Clostridium difficile* toxinotype XI (A-B-) exhibits unique arrangement of PaLoc and its upstream region, *Anaerobe*, 16(4): 393-395

130. **Sugeng, C.** 2012. Determining the Growth Limiting Conditions and Prevalence of *Clostridium difficile* in Foods, Master's thesis, The University of Ottawa

131. **Surawicz, C., Brandt, L., Binion, D., et al.** 2013. Guidelines for Diagnosis, Treatment, and Prevention of *Clostridium difficile* Infections, *The American Journal of Gastroenterology*, 108: 478-498

132. **Tedesco, F., Gordon, D., and Fortson, W.** 1985. Approach to patients with multiple relapses of antibiotic-associated pseudomembranous colitis, American Journal of Gastroenterology, 80:867-868

133. **Tenover, F., Åkerlund, T., Gerding, D. et al.** 2011. Comparison of Strain Typing Results for *Clostridium difficile* Isolates from North America, *Journal of Clinical Microbiology*, 49(5): 1831-1837

134. **Tenover, F., Novak-Weekley, S., Woods, C. et al.** 2010. Impact of Strain Type on Detection of Toxigenic *Clostridium difficile*: Comparison of Molecular Diagnostic and Enzyme Immunoassay Approaches, *Journal of Clinical Microbiology*, 48(10): 3719-3724

135. **Tenover, F., Tickler, I., and Persing, D.** 2012. Antimicrobial-Resistant Strains of *Clostridium difficile* from North America, *Antimicrobial Agents and Chemotherapy*, 56(6): 292-2932

136. **Trzasko, A., Leeds, J., Praestgaard, J., et al.** 2012. Efficacy of LFF571 in a Hamster Model of *Clostridium difficile* Infection, *Antimicrobial Agents and Chemotherapy*, 56(8): 4459-4462

137. **Tyler, A., Christianson, S., Knox, N. et al.** 2016. Comparison of Sample Preparation Methods Used for the Next-Generation Sequencing of *Mycobacterium tuberculosis*, *PLoS One*, 11(2): e0148676

138. **US Department of Health and Human Services.** 2013. Antibiotic Resistance Threats in the United States, 2013, *Centers for Disease Control and Prevention*, pg. 7, 11, 13, 17, 21, 25, 26, 50-56

139. **Vohra, P. and Poxton, I.** 2011. Comparison of toxin and spore production in clinically relevant strains of *Clostridium difficile*, *Microbiology*, 157: 1343-1353

140. **Voth, D. and Ballard, D.** 2005. *Clostridium difficile* Toxins: Mechanism of Action and Role in Disease, *Clinical Microbiology Reviews*, 18(2): 247-263

141. **Walkty, A., Boyd, D., Gravel, D. et al.** 2009. Molecular characterization of moxifloxacin resistance from Canadian *Clostridium difficile* clinical isolates, *Diagnostic Microbiology and Infectious Diseases*, 66(4): 419-424

142. **Wendt, J., Cohen, J., Mu, Y. et al.** 2014. *Clostridium difficile* Infection Among Children across Diverse US Geographic Locations, *Pediatrics*, 133(4): 651-658

143. **Wilcox M., Fawley W., Wigglesworth N., et al.** 2003. Comparison of the effect of detergent versus hypochlorite cleaning on environmental contamination and incidence of *Clostridium difficile* infection, *The Journal of Hospital Infection*, 54(2): 109-114

144. **Wilcox, M. and Fawley, W.** 2000. Hospital disinfectants and spore formation by *Clostridium difficile*, *Lancet*, 356(9238): 1324

145. **Wilcox, M., Fawley, W., Settle, C. et al.** 1998. Recurrence of symptoms in *Clostridium difficile* infection – relapse or reinfection?, *Journal of Hospital Infection*, 38: 93-100

146. **Wilcox, M., Mooney, L., Bendall, R. et al.** 2008. A case-control study of community-associated *Clostridium difficile* infection. *The Journal of Antimicrobial Chemotherapy*, 62: 388-396

147. **Williams, R., Meader, E., Mayer, M. et al.** 2013. Determination of the attP and attB sites of phage ϕCD27 from *Clostridium difficile* NCTC 12727, *Journal of Medical Microbiology*, 62: 1439-1443

148. **Wiström, J., Norrbyb, S., Myhre, E. et al.** 2001. Frequency of antibiotic-associated diarrhoea in 2462 antibiotic-treated hospitalized patients: a prospective study, *The Journal of Antimicrobial Chemotherapy*, 47(1): 43-50

149. **Zankari, E., Hasman, H., Cosentino, S.,et al.** 2012. Identification of acquired antimicrobial resistance genes, Journal of Antimicrobial Chemotherapy, 67(11):2640-2644

150. **Zar, F., Bakkanagari, S., Moorthi, M., et al.** 2007. A Comparison of Vancomycin and Metronidazole for the Treatment of *Clostridium difficile*–Associated Diarrhea, Stratified by Disease Severity; *Clinical Infectious Disease*, 45(3): 302-307

151. **Zhanel, G., Walkty, A., Vercaigne, L. et al.** 1999. The new fluoroquinolones: a critical review, *The Canadian Journal of Infectious Diseases*, 10(3): 207-238

152. **Zhou, Y., Liang, Y., Lynch, K., et al.** 2011. PHAST: A Fast Phage Search Tool, *Nucleic Acids Research*, 39(2): W347-W352

153. **Zilberberg, M., Tillotson, G., and McDonald, L.** 2010. *Clostridium difficile* Infections among Hospitalized Children, United States, 1997–2006, *Emerging Infectious Diseases*, 16(4): 604-609

11. Appendix

Section 1. RStudio commands for creating a heatmap.

```
#Import Packages
library("reshape2")
library("ggplot2")
library("dichromat")
library("ape")
library("dplyr")
library("heatmap.plus")

#Importing data
snp.matrix <- read.csv("snp_matrix.csv")
ab.matrix <- read.csv("antimicrobial_resistance_matrix.csv")

ab.matrix.long <- arrange(melt(ab.matrix), strain)
ab.matrix.long$strain <- as.character(ab.matrix.long$strain)
ab.matrix.long$strain <- gsub(pattern = " ", replacement = "", ab.matrix.long$strain)
ab.matrix.long$strain <- ordered(ab.matrix.long$strain,
levels=c(as.character(snp.matrix$strain)))
arrange(ab.matrix.long, strain)

# Replace numeric values with text
ab.matrix.long$value <- as.character(ab.matrix.long$value)
ab.matrix.long$value[ab.matrix.long$value==0]  <- "No Data"
ab.matrix.long$value[ab.matrix.long$value==0.5]  <- "Sensitive"
ab.matrix.long$value[ab.matrix.long$value==1.0]  <- "Intermediate"
ab.matrix.long$value[ab.matrix.long$value==1.5]  <- "Resistant"

# Convert values into ordered factors
ab.matrix.long$value <- as.factor(ab.matrix.long$value)
ab.matrix.long$value <- ordered(ab.matrix.long$value, levels = c("No Data", "Sensitive",
"Intermediate", "Resistant")) # Assign order of factors

# Make base plot
ab.plot <- ggplot(ab.matrix.long, aes(variable, strain, fill=value))

base_size <- 9

# Add aesthetics to plot
ab.plot +
  geom_tile(aes(fill=value, width=0.8, height=1), colour = "black")  +
  scale_fill_manual(values = c("white", "yellow", "orange", "red"))  +
  labs(y='Isolates',x='Antibiotics')  +
  theme_grey(base_size = base_size)  +
  labs(x = "", y = "")  +
```

```
  scale_x_discrete(expand = c(0, 0))  +
  scale_y_discrete(expand = c(0, 0))  +
  theme(legend.position = "left", axis.ticks = element_blank(), axis.text.x = element_text(size =
base_size *0.8, angle = 330, hjust = 0, colour = "grey50"))

# Save the image in project directory
ggsave (ab.plot, filename = "difficile_ABr_heatmap.png", height = 12, width = 16)
```

Section 2. Instructions for the in silico extraction of PHAST-defined phage sequence from draft genomes using the Linux command line.

Running and Collecting PHAST Output

```
#Open a Screen to watch Phast links for results files
screen -S new_screen

#Navigate to where you want Phast output to go to
cd path/to/phast/output/folder

#Make folders you will need later
mkdir phage_extracts all_bed_files watch_output assembled_fasta
```

#Place the assembled fasta files you are going to give Phast into the assembled_fasta folder. There is no command here as everyone probably has their fasta files in different directories. If you want to use the command line I would suggest using the mv (Move) command. Otherwise, move them manually using copy-paste in your project folder outside of the command line.

```
#Go into the assembled_fasta folder
cd assembled_fasta

#create a folder for later
mkdir Gapless_fasta

#Go into phage_extracts. You will need folders in here as well
cd ../phage_extracts

#Create folders you will need later
mkdir extracted_phages FINAL_separated_files

#Go to watch_output folder where phast output will be collected
cd ../watch_output
```

For each assembled fasta file:

#Upload the fasta file on the Phast website (http://phast.wishartlab.com/) in the third upload option at the bottom of the webpage.
#If your genome is a draft (multiple contigs denoting gaps in the sequence) remember to check off the box under the Browse button to concatenate the contigs together before uploading. Otherwise Phast will run on only the first 10 contigs in the fasta file separately.
#Once uploaded the website will estimate how long your genome will be in the queue for. This page includes a link you need to copy into the next code in the command line.

#Create a watch command to wait for Phast output to put into a folder for that specific genome
watch -n 28800 "nml_download_phast_links.pl -i '[link]' -d [strain name]"

#If you have another fasta file to run you need to make a new screen window to run another watch command in (I haven't found a better solution). While on the commmand line press:
'Ctrl + a' then press 'c'

#This is the end of the section regarding an individual fasta file. From here you can go back to the Phast website to upload another assembled draft genome and run another watch command

#You can parse through the screen windows on the command line to check on each of your watch commands using:
'Ctrl + a' and then press 'n' to go to the next window. You can tell by the screen number changing at the top of the command line window "[screen 1: bash]" moving to "[screen 2: bash]" for example.

#When you're done with the watch commands it is good etiquette to close your screens. Start with:
'Ctrl + a' then press 'd'

#Permanently delete the screen you created earlier (screen created was named new_screen)
screen -X -S new_screen quit

#You want to get into the phast output directory with folders phage_extracts, all_bed_files, watch_output, and assembled_fasta
cd path/to/phast/folders

Finding Phage Location in Assembled Genomes

#Make list of isolate names from watch_output
ls watch_output > list.txt

#Extract phage information from Phast summary file output (for your future reference)
for i in `cat list.txt`; do sed -rne "s/\s+\S+\s+\S+\s+(\S+)\s+\S+\s+([[:digit:]]+)-([[:digit:]]+).*/$i\t\2\t\3\t\1/p" watch_output/$i/NC_000000_summary_link_1 >> phast_summary_locations.txt; done

#Remove lines of phages you don't want. In this case, to ignore both questionable and incomplete phages use:
grep -v 'questionable\|incomplete' phast_summary_locations.txt > phast_summary_locations_2.txt

#Remove phage status (intact) from phast_summary_location_2.txt for future steps
sed -r 's/intact\S+|questionable\S+|incomplete\S+//g' phast_summary_locations_2.txt > phast_summary_locations_3.txt

#Subtract 1 from start position of phage for future bedtools use (starts counting at 0 instead of 1)
cat phast_summary_locations_3.txt | awk '{print $1"\t"$2-1"\t"$3}' > phast_summary_locations_4.txt

#You can keep the first 3 phast_summary_locations files (_1.txt, _2.txt, _3.txt) for your own reference, but phast_summary_locations_4.txt is the only file of the 4 we are going to use later on

#Next, we need to change the assembled fasta files to the form Phast was using them in
cd assembled_fasta

Changing Fasta Files to PHAST Format

#Rmv Headers (except the first one) and gaps in the assembled fasta files you gave Phast
for i in *.fasta; do union -sequence $i -outseq Gapless_fasta/"${i%%.*}".fasta; done

Navigate into gapless_fasta folder
cd Gapless_fasta

#Make list of fasta file names
for i in *.fasta; do basename $i .fasta >> list.txt; done

#Simplify headers for a later step
for i in `cat list.txt`; do sed -i "1 s/^.*$/>$i/g" $i.fasta; done

#Rmv Newlines
for i in `cat list.txt`; do awk '/^>/{print s? s"\n"$0:$0;s="";next}{s=s sprintf("%s",$0)}END{if(s)print s}' $i.fasta > "${i%%.*}".final; done

Extracting Phage Sequence Found by PHAST

#Create Bed file for all isolates (positions to extract from fasta)
for i in `cat list.txt`; do grep -i $i ../../phast_summary_locations_4.txt > ../../all_bed_files/"${i%%.*}".bed; done

#Delete .fasta files (set-up for proper rename of .final)
rm *.fasta

```
#Rename .final to .fasta
for i in *.final; do mv "$i" "${i/.final}".fasta; done

#Extract PHAST phage between basepair positions
for i in `cat list.txt`; do bedtools getfasta -fi $i.fasta -bed ../../all_bed_files/$i.bed -fo
../../phage_extracts/extracted_phages/"${i%%.*}"_phast_phages.fasta; done

Naviagate to extracted files
cd ../../phage_extracts/extracted_phages

#Make new list for next step
for i in *.fasta; do basename $i .fasta >> new_list.txt; done

#Place every found phage into its own fasta file
for i in `cat new_list.txt`; do sed -n '1,2p' $i.fasta >
../FINAL_separated_files/"${i%%.*}"_1.fasta; done

#Repeat above script as many times as there are phages found in your isolates (example below
for up to 5 phages in any genome)
for i in `cat new_list.txt`; do sed -n '1,2p' $i.fasta >
../FINAL_separated_files/"${i%%.*}"_1.fasta; done
for i in `cat new_list.txt`; do sed -n '3,4p' $i.fasta >
../FINAL_separated_files/"${i%%.*}"_2.fasta; done
for i in `cat new_list.txt`; do sed -n '5,6p' $i.fasta >
../FINAL_separated_files/"${i%%.*}"_3.fasta; done
for i in `cat new_list.txt`; do sed -n '7,8p' $i.fasta >
../FINAL_separated_files/"${i%%.*}"_4.fasta; done
for i in `cat new_list.txt`; do sed -n '9,10p' $i.fasta >
../FINAL_separated_files/"${i%%.*}"_5.fasta; done

#Go into separated_files folder
cd ../FINAL_separated_files

#Delete all empty files
find -size  0 -print0 |xargs -0 rm

#Wrap sequences after 60 bases (matching regular fasta format)
for i in *.fasta; do fold -w 60 $i > $i.fixed; done

#Remove phast_phages_#.fasta files
rm *.fasta

#Shorten file names
for i in *.fixed; do mv "$i" "${i/_phast_phages}"; done
```

#Finish shortening file names
for i in *.fixed; do mv "$i" "${i/.fasta.fixed}".fasta; done

#The DNA sequence of the phages found by PHAST in each assembled draft genome has been isolated and placed into its own fasta file. For example, 3 phages found means that isolate will have 3 fasta files in the FINAL_separated_files folder ([isolate name]_1.fasta [isolate name]_2.fasta [isolate name]_3.fasta). All of these will be found in the "FINAL_separated_files" folder.
#It should be noted that PHAST concatenated the contigs of your draft genomes (and so did you just now). The true position of the phages in the genome is different than the positions specified by Phast in the summary file. This is due to the absence of the gaps of unknown length between the contigs.

#To see if the phage inserted into a gene or area of interest, try this:

#The positions in the Phast output summary file (#-#) need to be scaled to capture more than the phage. I did this manually in the phast_summary_locations_4.txt file. I chose to extend the positions by 1000bp on either side of the phage (#[-1000]-#[+1000]; ex. 3333-4444 convert to 2333-5444) in a text editor. I'm sure there's an easier way to do this with a script.

#Process the fasta files containing your genome sequence as before using the new phast_summary_locations_4.txt file, but don't do the "#Wrap sequences after 60 bases" step near the end.

#Make a list of file names for the upcoming bash script
for i in *.fasta; do basename $i .fasta >> list.txt; done

#In the FINAL_separated_files folder cut out the phage in each fasta file which will leave you with the bacteria's DNA sequence on either side of the phage.
for i in `cat list.txt`; do sed -e '2,$s/^\(.\{1000\}\).*\(.\{1000\}\)$/\1\2/' $i.fasta > "${i%%.*}".final; done

#These .final files contain the DNA sequence without the phage sequence. If the phage inserted into a gene, a blastx search should show the gene repaired. The phage may have also inserted into a promoter region so keep an eye out for that in the blastx as well.

Section 3. RStudio commands for creating a phylogenetic tree from a binary table.

#User library
library("ape")
library("BiocInstaller")
library("ggplot2")
library("ggtree")

#System library

```
library("datasets")
library("graphics")
library("grDevices")
library("methods")
library("stats")
library("utils")

snp <- rbind(binary_table)
stree <- nj(dist.gene(snp))
nbtree <- ggtree(stree, ladderize = TRUE, layout = 'rectangular',branch.length='none', size=0.3) +
geom_tiplab(size=0.6, color="black") + geom_treescale(width=1, y=-2)
plot(nbtree)
ggsave("rectangle_tree.jpg", plot = nbtree)
```

Section 4. Nucleotide sequences in PFGE type 0033 genomes not found in PFGE type 0023 genomes.

Sequence ID	Nucleotide Sequence (5'-3')	Length (bp)
NepSeq1	AGTAAGAAAAGCACTTTCAAAGCACTATAAAAATAGTGCTGAAATAGATAAACAAATAAAAGTAATAA GGAGGAAAGATAGATTCTAGAGGGAAATCATTATCATTTTTAAGAACGTATGTTCGGTTGGTGGGATAG AAAAATACTTTACAGGGGATGGTTTATTTGAATGAGAATGAAAATATGAGTTTATATGTCAGTAAGTTA AAAGAAGTATTGAAAAATAATTCAGAAGAGTATAAAAAAATAAGAGCTAAAATAAATGAAATTTATAA TCTAAATGAAAAGAAAGAATAAAAGGGCTTTAGCTCTTTTATTCTTTCTTAGTATTAACTGCTTTATAAT AAGCATTATCTATCATTTTTTTTATTGCATTTTTATCATCTGAATCTAAAGATGCTAATCTATTTATTAGT TCGTTAACTTCATTATCTTCTTTTGAGTTAACTTTTTTTTAATAATTCATCATCTTGTTGAGTGTTAGGTTGT ATTTTACCTAATTTAGTTAATAGCTCCTTTAAAGTTAAATTTAATCCAAAACTCATTTTTTCAAGCATATC TAAAGTAGGTTCAACAGCTTTACCATTCCTTGGGTCTCTGTTTTTTTCTATCTTATCTATATATGTATGGC TCACACCACAAAGATTAGAAAATTCTCTAAGAGATAAGTTATGTTCACTTCTATATTCTTTAATAATTTC TGCTAATGTTTGCATAATTTCACACCTCATCTAAGTAATAGTTGACTTAATTATATCAGAAAAATGTAAA CTATTGTTAACAAAAATAAAAATTTAGTAAAAAATAATTGACAAGAGTAAAAAAAAGACGTATACTAT GATTAACAAAAAGGTTATAAATAAAGTTAGGAGGGATTAATATAAAAAATAATTTACAATATATTAGA AAAGAATCTAGAATATCACAAAAAAAATTCGCTGATAAGATAGGTATATCAAGACCGTATTTATCTAGA ATAGAAAACGGTAAGGTTAATCCGAGTTTAGAAATAGCGCATAAAATCTCGCTCGAAACAGGGAAAAC AATAAATGAAATTTTTTTTGATTTTACTGTAAATCATAGTAAACTATAAGGTTTTAATTGTACCTTGAAA ACTAAATATAAAATATTTTAAAGGAGCAAGTATATGAAAAATAAAGAAAAAGAAATAATCATTAG	1,177
NepSeq2	CTTTTAAATAATCTTTTTCTCTTACAGCGATAGTATAATATGTACACTTGCTTGTTTTTTGAGAATACAAA TAGTTTTTCTTTGCAAGCTTAGACAATATTGTTAATGTCGTATTTTTATTCCAGTCATAAATTTGTTTCAT TGCTTCCACAACTTCTCCTGATGCTACCTTATTTCCATCTAATCCCCAAATAAACTTCATTACTTTTAACT CAGAATCTGGCAATTTCTTGAATCCCATAAGCTACACCCCCTTTTTGTAAATTTATGTATGTTTATGTATA TTTAAATATTATAACAAATTTATTTTAAAAATAATTAATATATATATAAATTATTCTTTTCCAATTTTTTC CATATTTTCTATATATTTTTTTATTTGTTCTTCTGTCATCTTTTCTTTTTTATGCATAGTGGTAAACATTCG TATAAAAAAACTATCATATTTAGTTACCTTTCTATCTTCTTTTATTGCATTTAAATATTCTTTTTTAGTTAC TAGCACTTCATAACTAAACTGAAAACCTATCCTCTTCTTTTTTAATGTTTTCATTTTTACTAATCTTTTTAA TAGAATTTCTGTAGTACTTTTCCTCCACTTATATTTTTCTTTCATTGCTTCTATTATTTCTTTTTTAGATAA AATAGAATTTTTCCTCCACAAGTATTCCATCAACATCAATTCTCCCCTGCGTAGTTTATTAACTACTTCTT GTTGTTTCATTTTATGTCACTCCCTCAAATTGTTCTAATCTAAATCAATAATAAAACCTTCTGATTTCAAT TTATACAGGTAAGTTTTACCAAATATTCTACGCATTTAGTCTGTTTTCTTAGAAG	840

Section 5. Protein sequences described in the top blastn matches of NepSeq1 and NepSeq2.

Sequence ID	Top blastn Match	Protein found in Top blastn Match	Protein Sequence	Length (aa)
NepSeq1	Clostridium phage phiCD505, complete genome	Hypothetical Protein	MNENENMSLYVSKLKEVLKNNSEEYK KIRAKINEIYNLNEKKE	43
		Hypothetical Protein	MQTLAEIIKEYRSEHNLSLREFSNLCGV SHTYIDKIEKNRDPRNGKAVEPTLDMLE KMSFGLNLTLKELLTKLGKIQPNTQQD DELLKKVNSKEDNEVNELINRLASLDSD DKNAIKKMIDNAYYKAVNTKKE	133
NepSeq2	Clostridium phage phiCD505, complete genome	Hypothetical Protein	MKQQEVVNKLRRGELMLMEYLWRKN SILSKKEIIEAMKEKYKWRKSTTEILLKR LVKMKTLKKKRIGFQFSYEVLVTKKEY LNAIKEDRKVTKYDSFFIRMFTTMHKK EKMTEEQIKKYIENMEKIGKE	129
		Transcriptional regulator, BlaI/MecI/CopY family	MGFKKLPDSELKVMKFIWGLDGNKVA SGEVVEAMKQIYDWNKNTTLTILSKLA KKNYLYSQKTSKCTYYTIAVREKDYLK VETKKFFSFFHNNSLKSFFTALNDEENL SDDKLDMLEEWVKNWEEDE	127

Section 6. All proteins produced by clostridium phage phiCD505 [151].

Entry name	Protein names	Organism	Length (amino acids)
A0A0A8WI83_9CAUD	Holin	Clostridium phage phiCD505	85
A0A0A8WI91_9CAUD	ICEBs1 excisionase	Clostridium phage phiCD505	69
A0A0A8WJ04_9CAUD	N-acetylmuramoyl-L-alanine amidase	Clostridium phage phiCD505	270
A0A0A8WJ11_9CAUD	Putative antirepressor	Clostridium phage phiCD505	212
A0A0A8WJ22_9CAUD	Putative antirepressor	Clostridium phage phiCD505	282
A0A0A8WI87_9CAUD	Putative cI repressor	Clostridium phage phiCD505	61
A0A0A8WIA1_9CAUD	Putative endodeoxyribonuclease	Clostridium phage phiCD505	135
A0A0A8WIZ2_9CAUD	Putative head morphogenesis protein	Clostridium phage phiCD505	337
A0A0A8WF08_9CAUD	Putative integrase	Clostridium phage phiCD505	348
A0A0A8WI78_9CAUD	Putative peptidoglycan-binding LysM	Clostridium phage phiCD505	223
A0A0A8WFL3_9CAUD	Putative phage DNA modification methylase	Clostridium phage phiCD505	357
A0A0A8WF23_9CAUD	Putative phage essential recombination function protein	Clostridium phage phiCD505	204
A0A0A8WFK7_9CAUD	Putative phage protein	Clostridium phage phiCD505	118
A0A0A8WJQ5_9CAUD	Putative phage protein	Clostridium phage phiCD505	73
A0A0A8WI97_9CAUD	Putative phage protein	Clostridium phage phiCD505	38
A0A0A8WJQ3_9CAUD	Putative phage protein	Clostridium phage phiCD505	49
A0A0A8WFK1_9CAUD	Putative phage protein	Clostridium phage phiCD505	80
A0A0A8WI94_9CAUD	Putative phage protein	Clostridium phage phiCD505	286
A0A0A8WJN8_9CAUD	Putative phage tail tape measure protein	Clostridium phage phiCD505	1,442
A0A0A8WJN3_9CAUD	Putative phage-related terminase small subunit skin element	Clostridium phage phiCD505	233
A0A0A8WI73_9CAUD	Putative portal protein	Clostridium phage phiCD505	500
A0A0A8WIZ9_9CAUD	Putative tail fiber protein	Clostridium phage phiCD505	368
A0A0A8WEW2_9CAUD	Putative terminase B	Clostridium phage phiCD505	472
A0A0A8WJP7_9CAUD	Transcriptional regulator, BlaI/MecI/CopY family	Clostridium phage phiCD505	127
A0A0A8WF20_9CAUD	Transcriptional regulator, Phage-type	Clostridium phage phiCD505	65
A0A0A8WJ06_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	47
A0A0A8WJ20_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	123
A0A0A8WJP3_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	521

A0A0A8WI80_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	209
A0A0A8WFJ6_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	56
A0A0A8WI89_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	56
A0A0A8WJN5_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	127
A0A0A8WEX9_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	264
A0A0A8WJN6_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	273
A0A0A8WFF6_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	136
A0A0A8WIZ7_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	320
A0A0A8WJ12_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	70
A0A0A8WI74_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	112
A0A0A8WF29_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	149
A0A0A8WEX5_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	67
A0A0A8WJ18_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	58
A0A0A8WI84_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	62
A0A0A8WJ08_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	43
A0A0A8WEZ2_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	60
A0A0A8WJP0_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	152
A0A0A8WJQ7_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	113
A0A0A8WJP8_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	45
A0A0A8WJQ8_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	225
A0A0A8WJQ0_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	77
A0A0A8WI81_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	411
A0A0A8WFG0_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	157
A0A0A8WFF1_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	444
A0A0A8WF14_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	103
A0A0A8WF38_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	37
A0A0A8WJ15_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	83
A0A0A8WJP5_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	188
A0A0A8WFH1_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	212
A0A0A8WFM4_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	162

A0A0A8WIZ3_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	125
A0A0A8WJ01_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	99
A0A0A8WF33_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	334
A0A0A8WIZ6_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	140
A0A0A8WI77_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	473
A0A0A8WEZ7_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	76
A0A0A8WFL9_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	46
A0A0A8WFI6_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	129
A0A0A8WF04_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	50
A0A0A8WEW9_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	345
A0A0A8WFH6_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	45
A0A0A8WJP1_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	70
A0A0A8WFJ2_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	133
A0A0A8WEY6_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	378
A0A0A8WJQ2_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	169
A0A0A8WI99_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	109
A0A0A8WFG6_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	121
A0A0A8WFI1_9CAUD	Uncharacterized protein	Clostridium phage phiCD505	163