DEVELOPMENT OF A MOLECULAR-BASED RISK ASSESSMENT ASSAY FOR HUMAN-PATHOGENIC *CAMPYLOBACTER JEJUNI*

CODY BUCHANAN
Bachelor of Science, University of Victoria, 2009

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Biological Sciences
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

DEVELOPMENT OF A MOLECULAR-BASED RISK ASSESSMENT ASSAY FOR HUMAN-PATHOGENIC *CAMPYLOBACTER JEJUNI*

CODY BUCHANAN

Date of Defence: November 28, 2018

Dr. James E. Thomas Co-Supervisor	Professor Emeritus	Ph.D.
Dr. Eduardo N. Taboada Co-Supervisor	Adjunct Professor	Ph.D.
Dr. Victor P.J. Gannon Thesis Examination Committee Member	Adjunct Professor	Ph.D.
Dr. G. Douglas Inglis Thesis Examination Committee Member	Adjunct Professor	Ph.D.
Dr. Dan L. Johnson Thesis Examination Committee Member	Professor	Ph.D.
Dr. Igor Kovalchuk Chair, Thesis Examination Committee	Professor	Ph.D.

Abstract

Despite the fact that *C. jejuni* is a leading cause of bacterial enteritis in Canada, there is no diagnostic assay available to identify high-risk strains. This represents a significant challenge towards the prevention and control of campylobacteriosis. Additionally, molecular epidemiological studies have shown that not all *C. jejuni* strains or lineages appear to pose an equal risk to human health, and the factors underlying this subtype-dependent pathogenesis are not understood. A Genome-Wide Association Study was conducted to identify genetic markers associated with *C. jejuni* strains linked to human illness. These markers were implemented in a molecular-based risk assessment assay, and used to screen isolates collected as part of a national microbial survey of Canadian poultry products. This study suggests that strains with a diverse metabolic toolkit may pose an elevated risk by virtue of their ability to overcome ecological barriers that might otherwise mitigate exposure to humans.

Acknowledgements

Let me start by saying that this thesis does not represent the typical M.Sc experience, and for this reason, there are many people to thank. First and foremost, my supervisors, Drs. Eduardo Taboada and Jim Thomas. Thank you for the opportunity, guidance and seemingly unlimited patience. It's not often you have to pull your supervisor out of retirement to defend your thesis. By rights, you could have kicked me to the curb a very long time ago. Thank you for hanging in there and for letting me finish my degree. Secondly, thank you to the members of my committee: Dr. Gannon, Dr. Inglis and Dr. Johnson. Again, your patience and support over the course of this degree was most generous. Thank you to Dr. Oliver Lung, who took me on as his technician many moons ago, but always gave me the time to work on my thesis when necessary. And finally, thank you to the University of Lethbridge for not instituting a time limit to complete a degree until *after* I started my program.

The other most important thing you should know about this thesis is how it changed my life. The people I met, and the opportunities it has afforded me, are everything that I am today. To my dear friends and family away from family, thank you all for your friendship, support, guidance, contributions and everything else along the way. You know who you are. I will treasure my memories and look forward to making new ones! Finally, my parents; jeez, they had to put up with a lot of BS from me. Ahh, who am I kidding? They still will, but at least they won't have to listen to me complain about my thesis anymore. Love you!

Contents

List of tables

List of figures

List of abbreviations

VBNC	viable but non-culturable
GBS	Guillain-Barré Syndrome
MFS	Miller Fisher Syndrome
IBS	Irritable Bowel Syndrome
MEE	Multi-enzyme Electrophoresis
SNP	Single Nucleotide Polymorphism
PFGE	Pulsed-Field Gel Electrophoresis
SVR	Short Variable Repeat
MLST	Multi-locus Sequence Typing
ST	Sequence Type
CC	Clonal Complex
mCGH	microarray-based Comparative Genomic Hybridisation
CGF	Comparative Genomic Fingerprinting
C3GFdb	Canadian *Campylobacter* Comparative Genomic Fingerprinting Database
NGS	Next Generation Sequencing
WGS	Whole Genome Sequencing
cgMLST	core genome Multi-locus Sequence Typing
wgMLST	whole genome Multi-locus Sequence Typing
Mb	Megabase
CJIE	*C. jejuni* Integrated Element
CPS	Capsular polysaccharides
LOS	Lipooligosaccharides
Cia	*Campylobacter* invasion antigen
CT	*V. cholerae* Toxin
LT	Heat-labile Toxin
CDT	Cytolethal Distending Toxin
GWAS	Genome-Wide Association Study
NCA	Non-Clinically-Associated
UN	Undefined
CA	Clinically-Associated
CJRA	*C. jejuni* Risk Assessment
CFIA	Canadian Food Inspection Agency
MBS	Molecular Baseline Study
SRA	Short Read Archive
NCBI	National Centre for Biotechnology Information
ORF	Open Reading Frame
LG	Linkage Group
MRA	Molecular-based Risk Assessment
MSA	Multiple Sequence Alignment
MST	Minimum Spanning Tree

Chapter 1

1 Literature review

1.1 The genus *Campylobacter:* then and now

The earliest description of *Campylobacter* may have been penned by Theodor Escherich in his1886 post-doctoral thesis entitled "The Intestinal Bacteria of the Infant and Their Regulation to the Physiology of Digestion" (Shulman et al., 2007) where he described vibrio-like organisms in the stools of infants with diarrhea. However, it would be another 20 years before it was finally isolated by researchers from the United Kingdom. The British government commissioned a study in 1905 to investigate the cause of a significant number of epizootic abortions in cattle and sheep. The investigation fell to John McFadyean, the Dean and Professor of Pathology and Bacteriology at the Royal Veterinary College in London, and his colleague, Stewart Stockman, the Chief Veterinary Officer to the Board of Agriculture and Fisheries (Skirrow, 2006). It was early February in 1906 when McFadyean and Stockman isolated a pure culture of an unknown, "comma" shaped organism from the uterine mucous of a pregnant sheep. They described an organism that was spiral shaped, motile, and grew after 1-4 days at temperatures between 35-37°C. Growth was absent or poor on the surface of agar under aerobic conditions, but was favoured in a rarefied atmosphere. McFadyean and Stockman also described the appearance of coccal "granular" forms in culture and pathogenic material, which today, along with their other observations, are well-recognised features of the *Campylobacter* species (Blaser, 1997; McFadyean & Stockman, 1913; Ng et al., 1985).

Several years later in the United States, Theobald Smith and Marian Taylor isolated an organism from aborting cattle that was identical to the organism isolated in the United Kingdom. They called it *Vibrio fetus* (Smith & Taylor, 1919) and the disease became known as "vibrionic" abortion, which was of major economic concern at the time (Skirrow, 2006). Over the next several decades, researchers identified several other "vibrio" species including *V. jejuni* (Jones et al., 1931), *V. sputorum* subsp. *sputorum* (Prévot, 1940), *V. coli* (Doyle, 1944), *V. sputorum* subsp. *bubulus* (Florent, 1953), and *V. fetus* subsp. *venerealis* and *V. fetus* subsp. *intestinalis* (Florent, 1959). However, in 1963, French researchers, Véron and Chatelain, determined that *Vibrio fetus* required reclassification as its own genus (Sebald & Veron, 1963). Their experiments assessing the phenotypic characteristics and the G+C content of the DNA showed that *Vibrio fetus* was very different from *Vibrio cholerae* and related *Vibrio* species. In fact, the only characteristic they had in common was the "comma" shape of their cells, and thus, the genus *Campylobacter,* meaning "twisted bacteria", was proposed (Veron & Chatelain, 1973). The first documented cases where *Campylobacter* was linked to diarrhea in humans was in 1957 (King, 1957); however, it wasn't until the 1970's when new techniques were developed for the isolation of *Campylobacter* from stool that led to its widespread recognition as a human pathogen causing gastrointestinal illness (Blaser et al., 1979).

Belonging to the epsilon class of proteobacteria and the order *Campylobacteriales*, the genus *Campylobacter* is now known to comprise 26 species, of which 22 have been associated with human illness (Fitzgerald, 2015). *Campylobacter* species are gram-negative, non-spore forming bacilli that typically possess a spiral, or curved shape between 0.2 to 0.8 µm wide and 0.5 to 5 µm long. They are motile via a single,

unsheathed uni- or bipolar flagellum, which propels the organism with a characteristic corkscrew-like motion (Blaser, 1997).

In the laboratory, campylobacters are seen as fragile: sensitive to aerobic, acidic and highly osmotic environments, desiccation and temperature extremes. They are microaeophilic and grow best in atmospheres comprising 3-15% oxygen and 2-10% carbon dioxide. Campylobacters live in the digestive tracts of many different wild and domesticated animal species, which act as reservoirs for human disease (Blaser, 1997). They are found regularly in animals used for food production including cattle, sheep, pigs, and poultry, have been isolated from household pets including cats, dogs, and birds, and despite their fastidious nature in the laboratory, campylobacters are isolated readily outside the host from environmental waters and soils (Blaser, 1997).

Change in the morphological structure of campylobacters from spiral to coccoid in shape has been observed in old cultures and under unfavourable conditions. This "viable but non culturable" (VBNC) state is thought to enhance survivability outside of the host where growth conditions are poor (Cappelier & Federighi, 1998; Thomas, C. et al., 2002). Recent work has demonstrated that such cells can still express virulence genes and adhere to epithelial cells (Chaisowwong et al., 2012), substantiating concerns that VBNC cells, which are not always detectable by culture, may remain infectious and pose a threat to human health. Campylobacters have also been shown to form mono-culture biofilms or integrate into pre-existing biofilms comprising other bacteria, which renders them much more resistant to environmental extremes as well as disinfectants, and may be a contributing factor in their persistence and spread (Magajna & Schraft, 2015).

A thermotolerant subset of *Campylobacter* species that are capable of growth between 30°C to 42°C includes *C. lari, C. upseliensis, C. hyointestinalis, C. lanienae, C.*

sputorum, C. jejuni and C. coli. Campylobacter jejuni is delineated into two subspecies: *jejuni* and *doylei.* The pathogenic role of *C. jejuni* subspecies *doylei* is unknown, and is infrequently isolated. It is primarily isolated from human clinical samples associated with patients suffering from bacteremia, especially infants. It has yet to be isolated from an animal host, grows poorly at 42°C, and does not reduce nitrate (Parker et al., 2007). Thus, hereafter, *C. jejuni* will refer to *C. jejuni* subspecies *jejuni* exclusively. The subset of thermotolerant species are most commonly associated with human infection, but an overwhelming majority of cases (>95%) are attributed to *C. jejuni* (80-94%) and *C. coli* (2-15%) based on data from developed countries that use selective media culture for identification (Baylis et al., 2000; Blaser, 1997; Park, 2002). However, because of technical limitations of the current culture and phenotypic methods commonly used for detection, isolation, and typing of *Campylobacter,* the aforementioned figures likely result in under reporting of non-*jejuni/coli* species in clinical specimens, and therefore, the pathogenic potential of other *Campylobacter* species may be underestimated (Fratamico et al., 2008; Moore & Goldsmith, 2001).

1.2 Campylobacteriosis: A human disease

1.2.1 Overview of the disease

Infection with *Campylobacter* can result in a variety of diseases including acute enteritis, as well as extraintestinal infections such as bacteremia, abscess, meningitis, and other more severe post-infection complications including reactive arthritis, Guillain-Barré Syndrome (GBS), Miller Fisher Syndrome (MFS) and Irritable Bowel Syndrome (IBS) (Allos & Blaser, 1995; Blaser, 1997; Bremell et al., 1991; Garcia-Rodríguez et al., 2006;

Lecuit et al., 2004; Pope et al., 2007). In fact, case studies have demonstrated that *C. jejuni* is the leading bacterial antecedent of GBS having been found in 25-50% of patients, and there is evidence showing that some strains of *C. jejuni* produce ganglioside mimics found in neural tissues that are recognised by antibodies in the sera of afflicted patients (Willison et al., 2016). Typically, campylobacteriosis is hallmarked by severe abdominal pain, bloody diarrhoea, and sometimes fever, nausea and vomiting (Blaser, 1997). The symptoms of the disease manifest themselves 12-24 hours after infection, and are most severe for 24-48 hours, while dissipating over the course of a week. The disease is typically self-limiting lasting 5-7 days, though a small percentage of patients relapse after their initial illness. Appreciable numbers of bacteria can persist in faeces for 2-3 weeks after the initial infection (Blaser, 1997), but human to human transmission is rare (Musher & Musher, 2004). Long-term carriage of *Campylobacter* is often only observed in patients with immune deficiencies such as hypogammaglobulinemia and AIDS. Extraintestinal infections are more commonly observed in individuals who are immunocompromised, elderly or pregnant, and typically present as bacteremia with or without diarrheal illness. In most cases, the disease is self-managed by the patient with water and electrolytes. However, in cases where the patient is immunocompromised, or symptoms are especially severe, antibiotics such as erythromycin, tetracyclines or fluoroquinolones may be administered (Kirkpatrick & Tribble, 2011).

Investigations involving the experimental infection of human volunteers with *Campylobacter* have shown that the infectious dose can be as low as 500 cells (Black et al., 1988). These data are consistent with the observation that most recorded *Campylobacter* outbreaks result from the consumption of contaminated water or unpasteurised milk products since these matrices can act to lower the infectious dose

through buffering mechanisms, and rapid wash-through of stomach contents (Blaser et al., 1980; Murphy et al., 2006). However, the clinical consequences of an infection can vary in part based on the virulence of the infecting strain, the infectious dose, and the susceptibility of the patient (Black et al., 1988).

1.3 Prevalence and ecology

1.3.1 Overall prevalence

Campylobacter is considered to be the leading cause of bacterial foodborne gastroenteritis in the world (Batz et al., 2012; Galanis, 2007; WHO, 2013) and a significant cause of child morbidity and mortality in developing nations (Lanata et al., 2013). It is estimated to be responsible for as much as 14% of all diarrheal disease, translating to more than 400 million cases of campylobacteriosis annually (Duong & Konkel, 2009). In countries with active surveillance for this pathogen, the incidence of campylobacteriosis has risen annually (Park, 2002). The organism is ubiquitous in nature having been isolated from a broad range of wild animals and birds, domesticated animals including food animals and companion animals, as well as from environmental sources including waters and soils (Blaser, 1997).

1.3.2 Human incidence

In Canada, *Campylobacter* is the most common cause of bacterial gastroenteritis with annual incidence rates approaching 33 cases per 100,000 individuals (PHAC, 2015a). However, statistical models that account for unreported and undiagnosed cases suggest this rate could be as high as 447 cases per 100,000 individuals (Thomas, M. et al.,

2013). The highest rates are observed in children under 5 years of age and in young adults, particularly males, between the ages of 20-29 (PHAC, 2009). *Campylobacter* is also the most common cause of "Traveler's Diarrhea" among Canadians travelling to developing nations (PHAC, 2015b). These trends are typical of what has been reported in other industrialised nations (Blaser, 1997). The economic burden of campylobacteriosis in Canada has not been calculated, however, acute gastrointestinal illness is estimated to cost $115 per capita (Majowicz et al., 2006). In the United States, it is estimated that a case of campylobacteriosis costs approximately $2,283 totalling between $903 million to $4.9 billion annually, and was second only to norovirus for quality-adjusted life years lost (Batz et al., 2012; Hoffmann et al., 2015).

In developing nations, *Campylobacter* is endemic, but disease is generally confined to young children. It is thought that frequent and multiple exposures to a wide variety of strains early in life (i.e. via contaminated drinking water and close contact with farm animals) may result in protective immunity against subsequent infections and/or disease (Blaser, 1997; Calva et al., 1988; Rao et al., 2001), and could also reflect why asymptomatic *Campylobacter* infections are common in these regions (Havelaar et al., 2009). However, there is emerging evidence suggesting that both symptomatic and asymptomatic *Campylobacter* infections in children can result in growth faltering, which has been associated with a number of long-term negative effects including poorer cognitive development, lower adult work capacity and income, and poorer pregnancy outcomes (Lee et al., 2013).

1.3.3 Prevalence in agriculture animals

1.3.3.1 Poultry

Campylobacter is a commensal organism in poultry (Dhillon et al., 2006). It can colonise to high concentrations in the lower gastrointestinal tract, and can be found naturally within the tissues of lymphoid organs, liver and gallbladder (Cox et al., 2006). The most recent national poultry baseline survey in Canada found that 24% of flocks were colonised with *Campylobacter,* and a similar survey in the European Union found that 71% of flocks were colonised with *Campylobacter* (CFIA, 2016; EFSA, 2010). Once *Campylobacter* has been detected in a flock, it is recovered in very high numbers from most of the remaining birds within a few days, as well as from the surrounding environment (Shreeve et al., 2000). Colonised flocks typically remain positive until slaughter, where high levels of *Campylobacter* in the intestinal content can contaminate meat during processing (Oosterom et al., 1983). Retail poultry products are often *Campylobacter*-positive in Canada and in other industrialised nations, with prevalence ranging from 30% - 93% and is considered to be the leading source of exposure to human campylobacteriosis (CFIA, 2016; EFSA, 2010; PHAC, 2015b; Pointon et al., 2008).

1.3.3.2 Cattle

Cattle have been shown to be reservoirs for a number of thermophilic *Campylobacter* species including *C. jejuni, C. coli, C. hyointestinalis, C. fetus* and *C. lanienae* (Busato et al., 1999; Inglis et al., 2003; Stanley & Jones, 2003). Numerous investigations probing the prevalence of campylobacters in cattle populations have yielded estimated carriage rates ranging between 0.8% - 100% , depending on factors

such as age, cattle type, herd size, season, and whether the herd is on pasture, or in a feedlot (Garcia et al., 1985; Giacoboni et al., 2013; Hoar et al., 2001; Humphrey & Beckett, 1987; Inglis et al., 2004; Minihan et al., 2004; Nielsen, 2002; Stanley et al., 1998a; Stanley & Jones, 2003; Wesley et al., 2000). Studies have shown that a high percentage of cattle in feedlots chronically shed high numbers of campylobacter cells in their faeces (Besser, T. et al., 2005; Minihan et al., 2004). Campylobacters excreted in cattle faeces sent to compost can persist for up to 10 months (Inglis et al., 2010), which is in contrast to the common belief that campylobacters fare poorly in solid manure once excreted (Gilpin et al., 2009; Hoar et al., 1999; Nicholson et al., 2005; Sinton et al., 2007; Xu et al., 2009), and could be a persistent source of contamination on the farm and surrounding environment. While the prevalence of *Campylobacter* on cattle carcasses can be as high as 89% at time of slaughter, with the exception of offal, the prevalence of *Campylobacter* in retail products is negligible and not generally considered to be a risk for human campylobacteriosis (Bolton et al., 1985; Little & de Louvois, 1998; Stanley & Jones, 2003). In Canada, the prevalence of campylobacters in ground beef at retail is less than 1%, and data are no longer available after 2010 (PHAC, 2014; PHAC, 2015b).

1.3.3.3 Swine

Swine have been established as a reservoir for *Campylobacter* species including *jejuni* and *coli,* although, in contrast with poultry and cattle, *C. coli* is more prevalent than *C. jejuni* (Horrocks et al., 2009). Between 2011 and 2012 in Canada, *Campylobacter* was recovered from 85% of pooled-manure samples on swine farms, with 95% of isolates speciated as *C. coli* (PHAC, 2015a). Other studies in Canada have indicated that *Campylobacter* is highly prevalent on swine farms where between 40-100% tested

positive for the organism, with carriage rates in samples (e.g. faecal samples or caecal contents) between 30-100% and *C. coli* was the dominant species isolated (Farzan et al., 2010; Guevremont et al., 2004; Munroe et al., 1983; Varela et al., 2007). The Canadian data presented here are similar to those from other industrialised nations (Alter et al., 2005; Thakur et al., 2010). Quantification of caecal content at the abattoir indicated that pigs continue to harbour high levels of *Campylobacter* until slaughter with an average of 10^4 CFU/g faeces (Weijtens et al., 1997), and thus, can be a source of contamination during processing of pork products. However, while the prevalence of campylobacters on swine carcasses during various stages of processing ranged between 9-76%, it was not isolated after the chilling stage (Bracewell et al., 1985; Oosterom et al., 1985; Pearce et al., 2003). In Canada campylobacters are rarely isolated from pork products at retail, and are not generally considered to be a risk for human campylobacteriosis (PHAC, 2015a).

1.3.4 Prevalence in companion animals

It is well-established that dogs are carriers of many different thermophilic *Campylobacter* species including predominantly *C. upsaliensis,* but also, *C. jejuni, C. coli, C. helveticus, C. lari* and others (Engvall et al., 2003; Hald et al., 2004; Koene et al., 2004; Rossi et al., 2008; Tsai et al., 2007). Dogs can naturally carry multiple *Campylobacter* species simultaneously; however, the species richness can be elevated in diarrheic animals (Chaban et al., 2010), and illness may be associated with *C. jejuni,* especially in young and immunocompromised dogs (Burnens et al., 1992; Fox et al., 1983). In Canada, studies investigating the prevalence of *Campylobacter* species in pet dogs have demonstrated carriage rates between 22-58% in healthy dogs and as high as 97% in diarrheic dogs (Chaban et al., 2010; Leonard et al., 2011), which are similar to

data from other countries, where prevalence rates ranged between 2-100% (Hald et al., 2004; Leahy et al., 2016; Mohan et al., 2017; Tsai et al., 2007).

Cats are also known carriers of several thermophilic *Campylobacter* species including predominantly *C. helveticus,* but also *C. jejuni, C. coli and C. upsaliensis,* and can also be diversely colonised with multiple species (Rossi et al., 2008; Shen et al., 2001; Wieland et al., 2005). While data from Canada is unavailable, the prevalence of campylobacters in cats reported by other industrialised nations ranged between 28-75% (Acke et al., 2009; Shen et al., 2001; Wieland et al., 2005), and with the exception of one study, which found *C. jejuni* to be more prevalent in diarrheic and young cats (Acke et al., 2009), there is no apparent association between a particular species of *Campylobacter* and gastroenteritis (Burnens et al., 1992).

Although the prevalence of *C. jejuni* and *C. coli* are relatively low in dogs and cats, other more prevalent species, especially *C. upsaliensis,* have been associated with disease in humans (Logan et al., 2000). And because of the exceptionally close contact shared between humans and companion animals (including their faeces), pets are considered to be a risk factor for contracting campylobacteriosis (Horrocks et al., 2009).

1.3.5 Prevalence in wild birds

Thermophilic campylobacters, especially *C. jejuni* and to a lesser extent, *C. coli* and *C. lari,* are considered to be widely distributed amongst wild bird species, with prevalence ranging from 0-90% (Colles et al., 2008; Hald et al., 2016; Keller et al., 2011; Van Dyke et al., 2010). A 2017 study from South Korea assessed the prevalence of campylobacters in 71 different species and demonstrated an overall isolation rate of 15.3%, although the prevalence varied between different avian families (Kwon et al.,

2017). In Canada, studies investigating the prevalence of campylobacters in wild birds have focused largely on waterfowl and seabird species with isolation rates ranging between 0-29% (Dobbin et al., 2005; Hoar et al., 2007; Inglis et al., 2007; Quessy & Messier, 1992; Van Dyke et al., 2010). The variability of *Campylobacter* prevalence observed among different species or families of birds is thought to be partially linked to foraging strategies; for example, gulls and crows found foraging in landfills, and species that forage off the ground within close proximity to agricultural operations, were found to have a higher prevalence of *Campylobacter* (Hald et al., 2016; Waldenstrom et al., 2002). Other factors such as social behaviours (e.g. communal roosting) and habitat-use have been shown to contribute to both intra-and inter-species spread of campylobacters (Taff et al., 2016). Wild birds, therefore, may act as a reservoir and conduit for the transmission of campylobacters amongst themselves, other wild and domesticated animals, as well as humans, via contamination of the environment, animal feed supplies and human food sources (Weis et al., 2016). For example, in the United States, a recent outbreak of campylobacteriosis in humans was clinically linked to raw peas contaminated with wild bird faeces (Gardner et al., 2011; Kwan et al., 2014).

1.3.6 Prevalence in wild animals

A variety of thermophilic *Campylobacter* species have been isolated from various terrestrial and aquatic mammals including hedgehogs, various mouse, vole, rat and shrew species, moles, squirrels, hares, badgers, skunks, racoons, foxes, wild boars, deer, horses, buffalo, non-human primates, and seals, as well as from reptiles including a tortoise and various species of lizards with isolation rates ranging between 0-67% (Baily et al., 2015; Benejat et al., 2014; Gilbert et al., 2014; Jokinen et al., 2011; Kakoyiannis et al., 1988;

Koga et al., 2015; Mutschall et al., 2015; Navarro-Gonzalez et al., 2014; Petersen et al., 2001; Rainwater et al., 2017; Viswanathan et al., 2017; Whiley et al., 2016). Although it is generally accepted that campylobacters isolated from wildlife are commensal in nature, some species have been associated with disease in specific hosts (Baily et al., 2015; Benejat et al., 2014; Gilbert et al., 2014). The role of wildlife as possible reservoirs and vectors of campylobacters is not well understood; several recent molecular-based epidemiological investigations have indicated that wildlife tend to carry different *Campylobacter* species as compared to wild birds, domestic animals and humans, and if species overlap was observed, isolates were genetically distinct (Gilbert et al., 2014; Navarro-Gonzalez et al., 2014; Petersen et al., 2001; Viswanathan et al., 2017). However, a strain of *C. jejuni* isolated from raccoons at a wildlife rehabilitation centre was implicated as the causative agent in an outbreak of human campylobacteriosis, and another study investigating the prevalence of *C. jejuni* in wild grey seal pups determined that isolates from these animals were genetically similar to human clinical isolates (Baily et al., 2015; Saunders et al., 2017). Thus, while campylobacters have been isolated from a broad range of wild animals, their importance in the epidemiology of human campylobacteriosis may be restricted to species that regularly interface with agricultural operations and/or human population centres, which can result in their exposure to and spread of campylobacters atypical of the wild environment, but relevant to human health.

1.3.7 Prevalence in environmental waters

Thermophilic campylobacters have been isolated from all types of environmental water sources, including surface waters (e.g. ponds, lakes and rivers), groundwater, wastewater effluent and seawater, and have been implicated as the cause of waterborne

outbreaks in Canada and around the world (Clark et al., 2003; Khan et al., 2014; Pitkänen, 2013; Schuster et al., 2005). Sources of contamination include run-off from farms, abattoir effluent and slurries, human sewage, and faecal contamination from domestic animals and wildlife (Khan et al., 2014). Watersheds serve many needs, including drinking, irrigation and recreation and support both livestock and wildlife. Surface waters, particularly those near major agricultural operations, have been well-studied in Canada, where thermophilic campylobacters, including *C. jejuni* and *C. coli,* are frequently detected with prevalences from 0-100% depending on the season, proximity to and type of agricultural activities (Huang et al., 2015; Jokinen et al., 2010; Jokinen et al., 2011; Khan et al., 2014). Furthermore, studies have shown isolates from water are genetically similar to those from domestic and wild animals that inhabit the same geographical region (Jokinen et al., 2011). These data suggest that water is an important source of *Campylobacter* and an efficient vehicle for transmission between wildlife, domestic animals, and humans.

As described earlier, campylobacters ha -culturable" state, or integrate into biofilms when conditions are not favorable (Cappelier & Federighi, 1998). Additionally, there is evidence that microbial eukaryotes may act as a nonvertebrate reservoir of campylobacters in the environment (Kaakoush et al., 2015). These mechanisms may act to enhance the persistence and transmissibility of campylobacters in environmental waters.

1.4 Epidemiology of *Campylobacter* infections

1.4.1 The endemic nature of *Campylobacter*

The vast majority of cases of campylobacteriosis, greater than 95%, are said to occur as sporadic, or endemic infections (Blaser, 1997). In Canada, between 2000 and 2004, 99.8% of all confirmed cases were classified as sporadic, and only 177 cases were attributed to outbreaks (PHAC, 2009). Epidemiological investigations using case control studies have indicated that contact with, or consumption of improperly cooked poultry products account for 50-70% of all endemic cases (Blaser, 1997; Friedman et al., 2004). The consumption of unpasteurised milk and milk products (Hopkins et al., 1984), consumption of untreated, contaminated water (Fullerton et al., 2007), contact with livestock and pets (Friedman et al., 2004; Potter et al., 2003), occupational exposure in abattoirs (Cawthraw et al., 2000; Eberhart-Phillips et al., 1997), contact with infected individuals, and travel in developing countries (Blaser, 1997) are other dominant routes of infection.

1.4.2 *Campylobacter* outbreaks

Despite its relatively low infectious dose and status as an enteric pathogen, outbreaks of campylobacteriosis are rare. It is possible that the fragile nature of *Campylobacter,* such as sensitivity to atmospheric oxygen, heat, ultraviolet radiation, desiccation, and that it does not readily multiply at room temperature (Baylis et al., 2000; Blaser et al., 1980) precludes it from large-scale outbreaks related to solid foods (Franco, 1988). This is consistent with the fact that most documented outbreaks are a result of the consumption of unpasteurised milk and untreated and/or contaminated water, although

outbreaks due to the consumption of contaminated foods, especially poultry, have been reported (Blaser, 1997). More recently, several *Campylobacter* outbreaks have been attributed to contact with animals, or their faeces, including an outbreak caused by faecal contamination of peas by wild birds, contact with raccoons at a wildlife rehabilitation centre, and contact with puppies at a chain of pet stores (CDC, 2018; Gardner et al., 2011; Kwan et al., 2014; Saunders et al., 2017). In Canada, of 138 documented outbreaks between 1978 and 2005, a majority were associated with poultry and dairy products, although *Campylobacter* has been associated with serious waterborne outbreaks including in Walkerton, Ontario in 2000 (Clark et al., 2003).

In addition to the considerable underreporting of *Campylobacter,* the lack of, or inconsistent application of surveillance programs and subtyping methodologies may limit the detection of outbreaks, especially if they are small or widely dispersed. Furthermore, the ubiquitous nature of *Campylobacter*, its low infectious dose, the ease at which cross-contamination may occur, the possibility of contamination by multiple subtypes, and the relatively long incubation period may act to obfuscate the identification of a discrete source of contamination that results in a cluster of related illnesses (Ethelberg et al., 2004; PHAC, 2009; The *Campylobacter* Sentinel Surveillance Scheme Collaborators, 2003).

1.4.3. Seasonal variation of *Campylobacter* infections

The number of *Campylobacter* infections, particularly in countries with temperate climates, including Canada, follows a well-defined seasonal pattern that is characterised by a dramatic increase in the spring, followed by a peak in the summer, and a gradual decline in the fall and winter months (Kapperud & Aasen, 1992; Kovats et al., 2005; Miller et al., 2004; Nylen et al., 2002; PHAC, 2009; Skirrow, 1987). Some countries

observed a peak, or multiple peaks in the summer months, which are believed to be a result of travel related infections (Ekdahl & Andersson, 2004; Kovats et al., 2005).

Given the complexity of the epidemiology of *Campylobacter*, it is uncertain what drives this peculiar phenomenon. However, it is possible that the seasonal changes in environmental conditions may alter the dynamics of transmission of *Campylobacter* by changing the behaviour of wild animals and migrating birds, known reservoirs of the pathogen, and the recreational activities of humans, resulting in increased exposure and subsequent illness (Lacey, 1993; Olson et al., 2008; Southern et al., 1990). The warmer weather also represents a marked increase in agricultural activity of both production animals and produce. Interestingly, the carriage rates are highest in the summer for poultry tested at various stages of production including on the farm, at slaughter, and at retail, with a similar seasonal peak in the percentage of infected broiler flocks occurring just prior to that of human infections (Meldrum et al., 2005; Patrick et al., 2004; Wilson, I., 2002). Other studies showed that seasonal peaks in isolates from lambs, cattle and sewage occurred one to two months prior to the peak of human infections, which along with poultry, may be associated with the seasonal cycle of campylobacteriosis in humans (Kovats et al., 2005; Stanley et al., 1998a; Stanley et al., 1998b; Wallace et al., 1997).

1.4.4 Transmission

Elucidating the transmission pathways of *Campylobacter* is critical towards developing mitigation strategies to reduce human campylobacteriosis. However, this is proving a daunting task given the widespread prevalence amongst wild and domesticated animal species and in the environment (Blaser, 1997). Generally speaking, transmission occurs via the consumption of contaminated food, water and contact with infected or

reservoir hosts and their faeces (Humphrey et al., 2007). With recent advancements in molecular typing strategies for *Campylobacter*, it has become apparent that the transmission of pathogen to human can vary based on regional factors, and that there are other important points of *Campylobacter* contamination outside the "farm to fork" continuum (Hannon et al., 2009; Taboada et al., 2015). For example, recently described outbreaks associated with campylobacters from wildlife, wild birds and pets highlight potentially underrepresented sources of human infection, and necessitates further surveillance of species that regularly interface with humans, and of environmental sources at sites where this occurs (CDC, 2018; Kwan et al., 2014; Saunders et al., 2017).

While poultry has long been considered a predominant source of human campylobacteriosis (Blaser, 1997), recent molecular epidemiological studies in Canada have indicated there is a marked difference in the attribution of human *Campylobacter* cases in rural versus urban environments, where human isolates are more closely related to those from cattle versus those from poultry, respectively (Hannon et al., 2009; Taboada et al., 2015). Moreover, subtyping of isolates collected as part of a recent Canadian initiative to establish a national baseline estimate of the prevalence of campylobacters in broiler chickens and chicken products (CFIA, 2016) showed that while poultry remains a significant reservoir for *Campylobacter* in Canada, subtypes exclusive to chicken account for only a small percentage of clinical cases, and a majority of clinical cases arise from subtypes of mixed-host origin (presented in Chapter 3).

Thus, in order for future mitigation strategies to be effective, it is important to consider regional influences on *Campylobacter* ecology, and to continue to invest in broader molecular epidemiological studies and surveillance programs that not only investigate the ecology of these clinically important mixed-host subtypes observed in the

agriculture domain, but to also establish the transmission dynamics between animals, humans and the environment outside of the "farm-to-fork" continuum.

1.5 Subtyping of *Campylobacter jejuni*

1.5.1 A rationale for subtyping

The fundamental purpose of subtyping is to identify distinct lineages or sub-lineages that comprise the population structure of a bacterium by assaying phenotypic or genotypic markers. However, the true utility of subtyping is as a framework for which metadata, such as important phenotypic traits or epidemiological features, can be overlaid to identify subtypes of interest. Subtyping is an indispensable component of modern day epidemiology; it is used for surveillance, outbreak detection and mitigation, and as part of a more holistic approach to understanding the ecological epidemiology of *C. jejuni,* which can be used to develop more efficacious and targeted mitigation strategies (Ahmed et al., 2012; Taboada et al., 2013).

1.5.2 Phenotypic subtyping methods

The first subtyping methods developed for *C. jejuni* were designed to differentiate strains based on phenotypic – physical and/or biochemical – characteristics (Taboada et al., 2013). Many phenotypic subtyping schemes were developed including biotyping (Bolton et al., 1984; Lior, 1984; Skirrow & Benjamin, 1980), serotyping (Lior et al., 1982; Penner & Hennessy, 1980), phage typing (Grajewski et al., 1985), resistotyping (Ribeiro et al., 1996), and multilocus enzyme electrophoresis (MEE) (Aeschbacher & Piffaretti, 1989; Moore et al., 2002). However, with the exception of the serotyping

schemes, the widespread adoption of these methods was limited; technical aspects between independently developed biotyping and phage typing schemes, for example, lacked consistency across the scientific community, which hindered inter-laboratory comparison of subtyping data. Serotyping and phage typing were labour intensive, technically demanding and time consuming, while the maintenance and quality control of antisera and phage panels were expensive (Wassenaar & Newell, 2000). Additionally, serotyping and phage typing were prone to non-typeable strains and to problems with cross-reactivity, and although these methods had higher discriminatory power than biotyping and resistotyping, none had the discriminatory power to consistently reconcile subtyping data with epidemiological data (Moore & Madden, 2003; On et al., 2008; Sails et al., 2003). While MEE appeared to be highly discriminatory relative to other phenotypic methods (Moore et al., 2002), advances in molecular biological techniques led to the concurrent development and popularisation of genotypic methods.

1.5.3 Genotypic subtyping methods

Genotypic subtyping methods differentiate bacterial strains by interrogating the organism's DNA for certain genomic features such as different restriction sites, the presence or absence of genes or other genetic markers, and single nucleotide polymorphisms (SNPs) (Taboada et al., 2013). In addition to superior discriminatory power, standardisation, reproducibility and typeability, genotypic subtyping methods can be used to infer genomic relatedness amongst isolates (Eberle & Kiess, 2012). While a wide variety of methods have been used to subtype *C. jejuni*, (reviewed by Eberle & Kiess, 2012), only a few have been widely used including pulsed field gel electrophoreses

(PFGE), *flaA* Short Variable Region (*flaA*-SVR), multi-locus sequence typing (MLST), as well as several more recently developed methods based on whole genome analysis.

1.5.3.1 Pulsed-Field Gel Electrophoresis (PFGE)

PFGE is a subtyping method that uses one or more rare-cutting restriction enzymes to digest intact genomic DNA into a small number of fragments, which are subsequently electrophoresed in an agarose gel to separate them based on size. The resulting banding pattern is a strain's subtype. For *C. jejuni,* the *Sma*I and *Kpn*I restriction enzymes are most frequently used. Although banding patterns can be compared directly on gel, softwares have been developed to facilitate the inter-laboratory comparison of large numbers of isolates electronically, and to correct for minor variations in PFGE patterns that can occur between gels and laboratories (On et al., 2008).

While PFGE is regarded as being highly discriminatory, it has failed, in some instances, to differentiate epidemiologically-defined outbreak strains from sporadic strains in highly clonal lineages of *C. jejuni* (Champion et al., 2002). Moreover, genomic instability, through processes such as genetic recombination, has been shown to alter PFGE profiles in closely related strains (Hanninen et al., 1999; Wassenaar et al., 1998), and although standardised methods for *Campylobacter* have been developed by PulseNet, PFGE is considered to be too costly and time consuming for surveillance of sporadic cases of campylobacteriosis (Taboada et al., 2013).

1.5.3.2 Multi-Locus Sequence Typing (MLST)

MLST is a typing strategy that relies upon PCR amplification and subsequent sequencing of fragments from seven relatively stable housekeeping genes (*asp, glnA,*

gltA, glyA, pgm, uncA, and *tkt).* The sequenced PCR products are assigned an allele number based on an identical match to an allele in the global PubMLST database (https://pubmlst.org/campylobacter/), and the combination of the seven allele numbers is assigned a sequence type (ST) according to the database. Sequence Types that share four or more of the same alleles are said to be part of the same clonal complex (CC) or lineage (Dingle et al., 2001; Suerbaum et al., 2001).

Because of its increased discriminatory power, reproducibility, and relative ease to generate, interpret and share data, MLST has become one of the most extensively used subtyping methods for *C. jejuni*, and has led to important insights into its epidemiology, population structure and evolution (Taboada et al., 2013). Notably, MLST analysis of a large number of isolates has shown that *C. jejuni* is very genetically diverse, weakly clonal in population structure, and both intra- and inter-species recombination is common (Dingle et al., 2001). Moreover, MLST analysis of isolates from human, animal and environmental sources has shown that some lineages are widely distributed in both their geography and ecology, while other lineages appear to be more restricted (Dingle et al., 2001; French et al., 2005; Manning et al., 2003), which has been useful for source attribution to clinical cases (Sheppard et al., 2009). However, if an outbreak with a common ST occurs, MLST may not have sufficient resolution to differentiate outbreak strains from epidemiologically-unrelated strains, and a secondary typing scheme may be required to provide additional discrimination (Sails et al., 2003).

1.5.3.3 Microarray-based comparative genomic hybridisation (mCGH)

Completion of the genomic sequence of *C. jejuni* NCTC 11168 in 2000 paved the way for comparative genomics analyses by providing the nucleotide gene sequences

necessary to generate whole-genome DNA microarrays (Parkhill et al., 2000). Microarray-based Comparative Genomic Hybridisation (mCGH) is a technique that compares the presence or absence of thousands of genes in a single experiment based on the co-hybridisation of differentially-labelled tester and reference strains (Taboada et al., 2013).

After a series of comparative genomics analyses came available, it became clear that *C. jejuni* exhibited extensive genomic diversity, including differential carriage of accessory genes not conserved across the entire species. This led to the suggestion its evolution is driven by recombination, with the exchange of genetic material influencing its ecology and pathogenesis (Champion et al., 2005; Dorrell et al., 2001; Leonard et al., 2003; Pearson et al., 2003; Taboada et al., 2004). In comparison to other subtyping methods, mCGH was found to be highly discriminatory, capable of differentiating between strains with identical genotypes consistent with differences in their epidemiology (Carrillo et al., 2012; Gripp et al., 2011; Taboada et al., 2008). Although technical limitations (i.e. low throughout and high cost) precluded the deployment of mCGH as a tool for molecular subtyping, it led to the development of comparative genomics-informed typing methods based on assaying differences in gene content (Taboada et al., 2013).

1.5.3.4 Comparative Genomic Fingerprinting (CGF)

Comparative Genomic Fingerprinting (CGF) is a subtyping method that uses multiplex PCR to assess the conservation status of 40 genes from the accessory genome of *C. jejuni* to reproduce phylogenetic relationships inferred by mCGH analysis.

Subtyping data are generated through the binarisation of gene conservation profiles, which can then be compared to infer phylogenetic relationships (Taboada et al., 2012).

Results from a panel comprising more than 400 isolates showed that CGF is highly concordant with MLST, but with higher discriminatory power (Taboada et al., 2012). In comparison to MLST, fla-SVR and porA sequence typing, CGF was best suited for the identification of case clusters (i.e. outbreaks) in a routine surveillance setting, and was the only method capable of discriminating outbreak isolates from similar, but epidemiologically unrelated isolates (Clark et al., 2012). In addition to its enhanced discriminatory power, CGF is rapid, portable, inexpensive, amenable to high throughput workflows, and as such, has been deployed in Canada as the primary subtyping method for *C. jejuni.*

The Canadian *Campylobacter* Comparative Genomic Fingerprinting Database (C3GFdb) contains subtyping data for 24,142 *Campylobacter* isolates representing 4,882 unique subtypes isolated from human (n = 4,697), animal (n = 14,750) and environmental (n = 4,457) sources from across Canada. The database, in addition to storing subtyping data, also contains basic epidemiological metadata for each isolate including host source, date and location. This allows for contextualisation of a given subtype into the broader epidemiology of *C. jejuni* in Canada, and provides a snapshot of the characteristics associated with it. Together, the subtyping and epidemiological data provide an analytical framework to establish linkages between genotypes and phenotypes of public health concern. With the enhanced discriminatory power of CGF, the identification of more precise linkages consistent with local epidemiological trends may be useful for the development and implementation of cost-effective, targeted mitigation strategies to minimise human exposure to *C. jejuni* (Taboada et al., 2015).

1.5.3.5 Whole genome typing methods

With the advent of rapid and cost-effective next generation sequencing (NGS) technologies, whole genome sequencing (WGS) is increasingly being deployed for outbreak analysis and surveillance of bacterial pathogens including *Campylobacter* (Llarena et al., 2017). WGS is advantageous because it yields unparalleled genomic resolution (i.e. the complete genetic sequence), is theoretically backwards compatible with most molecular subtyping techniques, and simultaneously provides a wealth of additional information such as the presence of antimicrobial resistance genes and virulence factors (Besser, J. et al., 2016; Carleton & Gerner-Smidt, 2016).

Predominant analytical approaches for bacterial comparative genomics using WGS data include whole-genome SNP-based comparisons (either reference based or non-referenced based), and gene-by gene comparisons such as core genome MLST (cgMLST) and whole-genome MLST (wgMLST) (Quainoo et al., 2017). The whole-genome SNP-based approach is highly discriminatory and has proven to be useful for outbreak analysis; however, a major drawback is the requirement of a highly-related reference genome to generate SNP-calls, which essentially limits the scope of the analysis to a particular dataset. On the contrary, for wgMLST, a strain is characterised on a gene-by-gene basis against a database comprising all known genes or loci (i.e. the pan-genome) for a bacterial species or genus, which means strains can be compared across datasets and laboratories. Additionally, the scheme can be extended as novel genes are discovered, and the analysis yields information on a strain's gene content, which can be used as the basis for preliminary genotypic and phenotypic interpretations.

Although WGS is replacing conventional subtyping tools for outbreak analysis and routine surveillance of pathogens such as *Listeria* and *Salmonella,* existing knowledge

gaps pertaining to the baseline levels of diversity within and between *C. jejuni* lineages, and how the organism evolves during colonisation and infection need to be addressed to establish robust criteria for case cluster definition during outbreak analyses (Llarena et al., 2017).

1.6 Genomics of *Campylobacter jejuni*

1.6.1 Genomic features of *C. jejuni*

At time of writing, there were 151 completed *C. jejuni* genomes available on Genbank (https://www.ncbi.nlm.nih.gov/genome/genomes/149), with thousands of additional draft genome assemblies available in public repositories including Genbank and PubMLST (http://pubmlst.org/). The overall genome size of *C. jejuni* is relatively small, with an average length of 1.68 megabases (Mb), relatively low G+C content of 30.4%, and is among the densest bacterial genomes with more than 90% of the genome coding for an average number of 1,633 proteins (Parkhill et al., 2000; Wu et al., 2016). Although the genomic structure of *C. jejuni* is largely syntenic, four genomic islands, termed *C. jejuni*-Integrated Elements (CJIEs), originally identified in the *C. jejuni* strain RM1221, have been shown to be broadly distributed across the species (Fouts et al., 2005; Parker et al., 2006). CJIE1 has been characterised as a Mu-like phage, while CJIEs 2 and 4 contain phage-related genes including novel endonucleases that disrupt the natural transformation of infected *C. jejuni* strains, which may help explain why some genotypes are relatively stable versus others, and CJIE3 appears to be an integrated plasmid (Fouts et al., 2005; Gaasbeek et al., 2009; Gaasbeek et al., 2010). Other

biological effects these genomic islands may mediate in *C. jejuni* have not been determined (Clark, 2011).

The *C. jejuni* genome is hallmarked by the presence of hypervariable sequences comprising homopolymeric tracts found mostly in regions encoding proteins involved in the biosynthesis or modification of cell-surface structures including the capsule, lipooligosaccharide and flagellum. Slipped-strand mispairing in these regions during replication can result in variation in the length of the homopolymer tract leading to phase variation during translation, and may play a role in immune avoidance, virulence, host adaptation or rapid adaptation to environmental changes (Jerome et al., 2011; Parkhill et al., 2000; Thomas, D. et al., 2014; Young et al., 2007). The frequency of variation within these sequences is very high, which has led to the suggestion that *C. jejuni* may behave similarly to a quasispecies where a population may be made up of many genotypes with many phenotypes rather than a clonal isolate (Jerome et al., 2011; Parkhill et al., 2000).

Following sequencing of the genome from the *C. jejuni* strain NCTC 11168, CGH-based studies demonstrated that the species exhibits extensive genomic diversity (Dorrell et al., 2001; Leonard et al., 2011; Pearson et al., 2003; Taboada et al., 2004) confirming earlier observations from MLST-based analyses suggesting that *C. jejuni* constitutes a weakly clonal and genetically diverse species (Dingle et al., 2001). A meta-analysis of the available CGH data revealed that approximately 79% of the *C. jejuni* genome is composed of core genes. Core genes are present in all strains, and are typically involved in vital processes that are essential for the organism's survival such as energy metabolism, cell division, and the synthesis of RNA and DNA (Dorrell et al., 2001; Pearson et al., 2003; Taboada et al., 2004). Interestingly, many genes that are associated with virulence also appear to be conserved (Dorrell et al., 2001). The remaining genes,

termed accessory genes, are generally confined to 16 discrete hypervariable, or plasticity regions dispersed throughout the *C. jejuni* genome and have been shown to have differential carriage across the species (Dorrell et al., 2001; Leonard et al., 2003; Pearson et al., 2003; Taboada et al., 2004). These genes encode proteins associated with the biosynthesis of cell surface structures, iron acquisition, respiration, DNA restriction/modification, sialyation, as well as those with no known function (Dorrell et al., 2001; Leonard et al., 2003; Pearson et al., 2003; Taboada et al., 2004). Additionally, *C. jejuni* is naturally competent (Wang & Taylor, 1990), which has led to the suggestion that recombination is the primary driver in its evolution. The ability to generate variability in the accessory genome may represent an additional mechanism through which *C. jejuni* can enhance its ecological fitness towards an array of different environmental conditions and hosts, and may play a role in its pathogenesis (Pearson et al., 2003; Taboada et al., 2004; Wilson, D. et al., 2009).

1.6.2 Plasmids

Carriage of plasmids occurs in between 19 and 53% of *C. jejuni* isolates, where a majority are classified as antibiotic resistance plasmids that are transmissible between *Campylobacter* species, but not *Escherichia coli* (Bacon et al., 2000). Several plasmids have been sequenced including a representative from a novel subgroup of conjugative plasmids called *mob* plasmids, a tetracycline-resistance encoding plasmid called pTet, and a virulence plasmid called pVir that encodes a Type IV Secretion System (Bacon et al., 2002; Batchelor et al., 2004; Schmidt-Ott et al., 2005).

The pVir plasmid, originally isolated from the highly pathogenic *C. jejuni* strain 81-176, encodes a Type IV Secretion System similar to one in the *cag* pathogenicity island

from *Helicobacter pylori*. Mutation of some of the plasmid-encoded genes resulted in reduced adherence to and invasion of INT407 enteric epithelial cells *in vitro*. However, when the plasmid was transferred to *C. jejuni* NCTC 11168, no difference in invasiveness was observed (Bacon et al., 2000). The role of pVir in *Campylobacter* pathogenesis is unclear; in one study from Canada, pVir was significantly associated with bloody diarrhea, while another study from Holland showed no correlation between the two (Louwen et al., 2006; Tracz et al., 2005). Additionally, pVir is found in only a small subset of human clinical isolates, which implies it is not essential for *C. jejuni* pathogenesis (Louwen et al., 2006; Schmidt-Ott et al., 2005).

1.7 Virulence factors of *C. jejuni*

C. jejuni is somewhat unique amongst enteric pathogens, as it does not contain any pathogenicity islands, and encodes relatively few classical virulence factors, most of which are conserved (Dasti et al., 2009; Dorrell et al., 2001). The molecular mechanisms underpinning the pathogenesis of *C. jejuni* have not been completely elucidated, however virulence factors associated with motility, evasion of the host immune response, host cell adherence, host cell invasion, protein secretion and alteration of host cell signalling pathways have been described (Flanagan et al., 2009; Guerry, 2007; Konkel et al., 1999; Konkel et al., 2001).

1.7.1 Flagella, flagellar-mediated protein secretion and chemotaxis

Flagella and flagellar-motility have long been considered essential for the pathogenesis of *C. jejuni* where early work demonstrated that mutants with defective

flagella were deficient for host colonisation and cellular internalisation (Morooka et al., 1985; Wassenaar et al., 1991; Yao et al., 1994).

Genes involved in the biogenesis of *C. jejuni* flagella are tightly regulated by the FlgS/R two-component system, in addition to three sigma factors (σ^{28}, σ^{54}, σ^{50}), which act in concert to regulate a large number of flagellar-structural and related genes involved in motility, protein secretion and invasion (Wösten et al., 2004). As with other bacteria, the flagellar filament is composed of multimers of the protein flagellin, which is highly immunogenic (Wassenaar & Blaser, 1999). However, in *C. jejuni,* flagellin is encoded by two differentially expressed genes (*flaA* and *flaB*) that are subject to variation by mechanisms including recombination and horizontal gene transfer, which can result in new epitopes and contribute to immune avoidance (Wassenaar et al., 1995).

In addition to motility, it is now understood that the flagellum of *C. jejuni* also functions as a Type III Secretion System, which secretes a number of proteins directly into the host cell, including the *Campylobacter* invasion antigens (Cia proteins), that help coordinate its internalisation and aid in its intracellular survival (Neal-McKinney & Konkel, 2012). *C. jejuni* flagella may also have a role in the formation of microcolonies and biofilms, which could promote its survival in the environment (Guerry, 2007; Haddock et al., 2010).

Chemotaxis is important for the pathogenesis of *C. jejuni,* where non-chemotactic mutants failed to colonise the intestine *in vivo* (Takata et al., 1992). The genome of *C. jejuni* possesses orthologues for a number of chemotaxis genes including *cheA, cheW, cheV, cheY, cheR* and *cheB* (Dasti et al., 2009; Parkhill et al., 2000). The protein CheY is the response regulator responsible for transmitting sensory signals from chemoreceptors to the flagellar motor, and plays a key role in *C. jejuni* pathogenesis by directing the

bacterium towards the mucous lining in the intestine. Mutants lacking, or overexpressing CheY, showed decreased virulence *in vivo;* it was postulated this could be due to the inability to properly chemotact, or in the case of CheY overexpression, affect subsequent interactions with intestinal epithelial cells (Yao et al., 1997).

1.7.2 Capsule and lipooligosaccharide

C. jejuni produces cell surface structures including capsular polysaccharides (CPSs), which are organised into a membrane-linked, extracellular capsule surrounding the bacterial cell and lipooligosaccharides (LOS) (Gourley & Konkel, 2016; Roberts, 1996). The capsule is thought to provide protection against desiccation, and from innate and acquired immune responses (Keo et al., 2011; Roberts, 1996; Young et al., 2007). *C. jejuni* with deficient capsule has shown decreased adherence and invasion *in vitro* and decreased fitness *in vivo* in chicken and ferrets (Bacon et al., 2001; Louwen et al., 2006). After flagella, LOS play a major role in cellular invasion, where variability in their structure can modulate the invasiveness of *C. jejuni* (Guerry et al., 2002; Louwen et al., 2008). Structural similarity between *C. jejuni* LOS and human neuronal gangliosides can trigger the onset of Guillain-Barré Syndrome, a serious autoimmune disorder that causes paralysis (Allos, 1997; Yuki, 1997). More recent work has demonstrated that *C. jejuni* LOS are involved in immune cell recognition, and provide protection from bacteriophages and antimicrobial peptides (Cullen et al., 2013; de Zoete et al., 2010; Louwen et al., 2013).

Both the CPS and LOS genes are located within hypervariable regions of the *C. jejuni* genome and are subject to recombination (Dorrell et al., 2001). In addition, some genes contain homopolymeric nucleotide tracts and are thus subject to phase variation

(Parkhill et al., 2000). This genetic variation results in variability of the CPS and LOS structures, which may be a strategy for avoidance of the host immune response, and can influence aspects of a strain's pathogenicity and disease outcome (Bacon et al., 2001; Guerry et al., 2002).

1.7.3 Adhesion and invasion

Adherence to intestinal epithelial cells is considered to be fundamental to *C. jejuni* pathogenesis, and while the genome of *C. jejuni* does not appear to encode any adherence organelles such as pili, several adhesins including CadF, FlpA, CapA, PorA, PEB1 and JlpA have been identified (Larson et al., 2008; Young et al., 2007). Mutants for CadF, FlpA, CapA, PEB1 and JlpA displayed reduced adherence and invasion *in vitro* and reduced or completely abrogated colonisation *in vivo* (Ashgar et al., 2007; Flanagan et al., 2009; Jin et al., 2001; Konkel et al., 2010; Monteville & Konkel, 2002; Monteville et al., 2003; Pei et al., 1998; Ziprin et al., 1999). With the exception of *capA,* all adhesin-encoding genes appear to be conserved in *C. jejuni* (Flanagan et al., 2009).

Invasion is mediated through the binding of CadF and FlpA to the host cell extracellular protein fibronectin. This interaction, in conjunction with the flagellar-secreted *Campylobacter* invasion antigens (Cia), trigger a signal cascade that ultimately co-opts the host cell's cytoskeleton to promote internalisation of *C. jejuni* and its subsequent intracellular survival (Eucker & Konkel, 2012; Monteville et al., 2003; Neal-McKinney & Konkel, 2012). Of the approximately 18 Cia proteins predicted to exist, only four have been characterised including CiaB, CiaC, CiaD and CiaI, and while their exact functions have not been determined, both *in vitro* and *in vivo* studies suggest they

are required for maximal invasion and intracellular survival (Buelow et al., 2011; Christensen et al., 2009; Konkel et al., 1999; Larson et al., 2008; Samuelson et al., 2013).

1.7.4 Toxins

Although early evidence suggested that *C. jejuni* expresses enterotoxins similar to the *V. cholerae* toxin (CT) and the *E. coli* heat-labile toxin (LT), subsequent studies yielded contradictory results, and as of yet, the genetic basis for their presence has not been verified (Dasti et al., 2009; Wassenaar, 1997). To date, the only toxin identified in the *C. jejuni* genome is the Cytolethal Distending Toxin (CDT), which causes G_2/M cell cycle arrest and eventual cell death via apoptosis (Parkhill et al., 2000; Whitehouse et al., 1998). CDT is encoded by three genes including *cdtA, cdtB,* and *cdtC*, which when translated, form a tripartite complex, classified as an AB_2 toxin, where the CdtB subunit is the catalytically active "A" subunit, and the CdtA and CdtC subunits form the binding "B" subunit (Lara-Tejero & Galán, 2001). The CdtB subunit alone is responsible for its toxic effects, which when singularly microinjected or transfected into host cells, induced the same effects as the complete toxin (Lara-Tejero & Galán, 2000). CdtA and CdtC mediate the toxin's binding to the host cell and internalisation of the CdtB subunit, which is then localised to the nucleus where it causes damage to the DNA in a fashion similar to proteins of the DNase-1 family (Elwell & Dreyfus, 2000; Lara-Tejero & Galán, 2000; Nishikubo et al., 2003).

Although CDT appears to be conserved across the species, its role and importance in the pathogenesis of *C. jejuni* are not clear (Pickett & Whitehouse, 1999). For example, a study by Mortensen and colleagues demonstrated that the disease outcome of humans infected by strains with anomalous *cdtABC* loci that did not express CDT was not

distinguishable from those with functional CDT (Mortensen et al., 2011). However, it has been determined that the presence of CDT elicits production of the proinflammatory cytokine interleukin-8, which results in a localised inflammatory response at the site of infection, and contributes to the symptoms of campylobacteriosis (Hickey et al., 2000).

1.8 Overview of the thesis

Based on molecular epidemiological studies, it appears that not all *C. jejuni* strains or genetic lineages pose an equal risk to human health. For example, the use of MLST has identified host restricted clonal complexes, and in Canada, the use of CGF has demonstrated the presence of subtypes that are often associated with human illness and those that are not (Buchanan et al., 2017; Dingle et al., 2001; French et al., 2005; Manning et al., 2003). *C. jejuni* is somewhat unique amongst enteric pathogens; it does not possess pathogenicity islands, and it expresses relatively few classical virulence factors (Dasti et al., 2009). Although a number of genes involved in various aspects of its pathogenesis have been described, they appear to be conserved across the species, and therefore, the genetic basis for this observed subtype-dependent pathogenesis is not well understood (Dorrell et al., 2001; Young et al., 2007). From a public health perspective, this is problematic; unlike other pathogens, such as *E. coli* for which well-characterised virulence factors (i.e. shiga toxins), or well-established epidemiological linkages between certain subtypes (i.e. O157:H7) and human disease are used for risk assessment, no such tools are available for *C. jejuni.* Given its status as a leading cause of bacterial gastroenteritis and its widespread prevalence in nature, the inability to characterise and track dangerous *C. jejuni* populations is a major knowledge gap in our understanding of

the epidemiology of campylobacteriosis, and a detriment to overall *Campylobacter*-associated disease management in Canada.

With the advent of inexpensive and accurate whole genome sequencing, Genome-Wide Association Studies (GWAS) are increasingly being implemented to study bacterial genomics as tools for the identification of genetic markers associated with a phenotype of interest. A GWAS represents a "top-down" approach that compares all genomic content between test and control groups to identify genetic factors that are statistically associated with a phenotype of interest. This is in contrast to a "bottom-up" approach where certain genetic factors are manipulated to test their effect on the phenotype. The utility of the GWAS approach lies in the ability to rapidly test large numbers of genetic factors, and potentially identify novel associations between genetic factors and a phenotype under study (Read & Massey, 2014). GWAS have been utilised to identify mutations and other polymorphisms associated with antibiotic resistance in *Mycobacterium tuberculosis* (Farhat et al., 2013), *Staphylococcus aureaus* (Alam et al., 2014), and *Streptococcus pneumonia* (Chewapreecha et al., 2014). In *Campylobacter,* GWAS have been used to identify genetic factors related to Guillain-Barré Syndrome (Taboada et al., 2007), host adaptation in *C. jejuni* and *C. coli* (Sheppard et al., 2013), and markers associated with the survival of *C. jejuni* in the poultry production chain (Yahara et al., 2017).

In this thesis, isolates from the C3GFdb were used to perform a GWAS aimed at identifying genes preferentially found amongst *C. jejuni* isolates from lineages associated with human disease. In addition with the goal of using them as diagnostic markers in a molecular-based risk assessment assay for the rapid detection of human-pathogenic *C. jejuni* isolates, identification of such genes could lead to insights in the molecular mechanisms behind the ecology and pathogenesis of this organism.

1.8.1 Research Objectives

1.8.1.1 Chapter 2: A Genome-wide association study to identify diagnostic markers for human-pathogenic *Campylobacter jejuni* strains

1) A dataset composed of 166 *C. jejuni* isolates representing 34 of the 100 most prevalent CGF subtypes circulating in Canada were selected for WGS. These subtypes represent approximately 31% of all isolates in the C3GFdb, and over 55% of isolates from the 100 most prevalent subtypes. Isolates were selected from Non-Clinically-Associated (NCA; ≤5% human clinical isola , Undefined (UN; 5–25% human clinical isolates), and Clinically-Associated (CA; ≥25% human clinical isolates) CGF subtypes.

2) The WGS data were processed and a pan-genome (n=3,358) for the dataset was established. The distribution of the accessory genome content (n=1,981) was determined and used in a GWAS to identify genes (n=28) statistically-significantly associated with CGF subtypes comprising isolates that cause disease in humans.

3) The availability of thousands of additional draft genome assemblies (n=3,902) from our laboratory and other public repositories presented an opportunity to conduct an *in silico* validation of the putative diagnostic markers against a much larger, and more geographically and genetically diverse dataset. In total, 25 markers were considered robust enough for testing in the wetlab.

1.8.1.2 Chapter 3: Laboratory validation and implementation of a molecular-based risk assessment assay for the identification of human-pathogenic *Campylobacter jejuni* strains

4) PCR primers were designed for each putative diagnostic marker (n=28), and organised into multiplex PCRs (n=6) for testing in the wetlab. The multiplex PCRs were optimised, then validated against a panel of 362 strains, including 221 strains that were previously sequenced. In order to streamline the assay, a subset comprising the most robust markers (n=11) were re-organised into two new multiplex PCRs and deployed as the *Campylobacter jejuni* Risk Assessment (CJRA) assay.

5) The CJRA Assay was screened against additional *C. jejuni* strains (n=591) collected as part of the Canadian Food Inspection Agency's (CFIA) National Microbial Baseline Study in Broiler Chicken (CFIA, 2016).

1.9 References

Acke, E., McGill, K., Golden, O., Jones, B., Fanning, S., and Whyte, P. (2009). Prevalence of thermophilic *Campylobacter* species in household cats and dogs in Ireland. *The Veterinary Record* 164, 44-47.

Aeschbacher, M., and Piffaretti, J. (1989). Population genetics of human and animal enteric *Campylobacter* strains. *Infection and Immunity* 57, 1432-1437.

Ahmed, M., Dunn, L., and Ivanova, E. (2012). Evaluation of current molecular approaches for genotyping of *Campylobacter jejuni* strains. *Foodborne Pathogens and Disease* 9, 375-385.

Alam, M., Petit, R., Crispell, E., Thornton, T., Conneely, K., Jiang, Y., Satola, S., and Read, T. (2014). Dissecting vancomycin-intermediate resistance in *Staphylococcus aureus* using genome-wide association. *Genome Biology and Evolution* 6, 1174-1185.

Allos, B. (1997). Association between *Campylobacter* infection and Guillain-Barré Syndrome. *The Journal of Infectious Diseases* 176, S128.

Allos, B., and Blaser, M. (1995). *Campylobacter jejuni* and the expanding spectrum of related infections. *Clinical Infectious Diseases* 20, 1092-1099.

Alter, T., Gaull, F., Kasimir, S., Gürtler, M., Mielke, H., Linnebur, M., and Fehlhaber, K. (2005). Prevalences and transmission routes of *Campylobacter* spp. strains within multiple pig farms. *Veterinary Microbiology* 108, 251-261.

Ashgar, S., Oldfield, N., Wooldridge, K., Jones, M., Irving, G., Turner, D., and Ala'Aldeen, D. (2007). CapA, an autotransporter protein of *Campylobacter jejuni*, mediates association with human epithelial cells and colonization of the chicken gut. *Journal of Bacteriology* 189, 1856-1865.

Bacon, D., Alm, R., Burr, D., Hu, L., Kopecko, D., Ewing, C., Trust, T., and Guerry, P. (2000). Involvement of a plasmid in virulence of *Campylobacter jejuni* 81-176. *Infection and Immunity* 68, 4384-4390.

Bacon, D., Alm, R., Hu, L., Hickey, T., Ewing, C., Batchelor, R., Trust, T., and Guerry, P. (2002). DNA sequence and mutational analyses of the pVir plasmid of *Campylobacter jejuni* 81-176. *Infection and Immunity* 70, 6242-6250.

Bacon, D., Szymanski, C., Burr, D., Silver, R., Alm, R., and Guerry, P. (2001). A phase-variable capsule is involved in virulence of *Campylobacter jejuni* 81-176. *Molecular Microbiology* 40, 769-777.

Baily, J., Méric, G., Bayliss, S., Foster, G., Moss, S., Watson, E., Pascoe, B., Mikhail, J., Pizzi, R., Goldstone, R., Smith, D., Willoughby, K., Hall, A., Sheppard, S., and Dagleish, M. (2015). Evidence of land-sea transfer of the zoonotic pathogen *Campylobacter* to a wildlife marine sentinel species. *Molecular Ecology* 24, 208-221.

Batchelor, R., Pearson, B., Friis, L., Guerry, P., and Wells, J. (2004). Nucleotide sequences and comparison of two large conjugative plasmids from different *Campylobacter* species. *Microbiology* 150, 3507-3517.

Batz, M., Hoffmann, S., and Morris, J. (2012). Ranking the disease burden of 14 pathogens in food sources in the United States using attribution data from outbreak investigations and expert elicitation. *Journal of Food Protection* 75, 1278-1291.

Baylis, C., MacPhee, S., Martin, K., Humphrey, T., and Betts, R. (2000). Comparison of three enrichment media for the isolation of *Campylobacter* spp. from foods. *Journal of Applied Microbiology* 89, 884-891.

Benejat, L., Gravet, A., Sifré, E., Ben Amor, S., Quintard, B., Mégraud, F., and Lehours, P. (2014). Characterization of a *Campylobacter fetus*-like strain isolated from the faeces of a sick leopard tortoise (*Stigmochelys pardalis*) using matrix-assisted laser desorption/ionization time of flight as an alternative to bacterial 16S rDNA phylogeny. *Letters in Applied Microbiology* 58, 338-343.

Besser, J., Carleton, H., Goering, R., & Gerner-Schmidt, P. (2016). The impact of molecular diagnostics on surveillance of foodborne infections. In D. Persing, F. Tenover, T. Randall, M. Leven, M. Miller, F. Nolte, Y. Tang & A. van Belkum (Eds.), *Molecular Microbiology: Diagnostic Principles and Practice* (Third ed., pp. 235-244). Washington, DC: ASM Press.

Besser, T., LeJeune, J., Rice, D., Berg, J., Stilborn, R., Kaya, K., Bae, W., and Hancock, D. (2005). Increasing prevalence of *Campylobacter jejuni* in feedlot cattle through the feeding period. *Applied and Environmental Microbiology* 71, 5752-5758.

Black, R., Levine, M., Clements, M., Hughes, T., and Blaser, M. (1988). Experimental *Campylobacter jejuni* infection in humans. *The Journal of Infectious Diseases* 157, 472-479.

Blaser, M. (1997). Epidemiologic and clinical features of *Campylobacter jejuni* infections. *The Journal of Infectious Diseases* 176, S105.

Blaser, M., Berkowitz, I., LaForce, F., Cravens, J., Reller, L., and Wang, W. (1979). *Campylobacter* enteritis: clinical and epidemiologic features. *Annals of Internal Medicine* 91, 179-185.

Blaser, M., Hardesty, H., Powers, B., and Wang, W. (1980). Survival of *Campylobacter fetus* subsp. *jejuni* in biological milieus. *Journal of Clinical Microbiology* 11, 309-313.

Bolton, F., Dawkins, H., and Hutchinson, D. (1985). Biotypes and serotypes of thermophilic campylobacters isolated from cattle, sheep and pig offal and other red meats. *Epidemiology and Infection* 95, 1-6.

Bolton, F., Hutchinson, D., and Coates, D. (1984). Blood-free selective medium for isolation of *Campylobacter jejuni* from feces. *Journal of Clinical Microbiology* 19, 169-171.

Bracewell, A., Reagan, J., Carpenter, J., and Blankenship, L. (1985). Incidence of *Campylobacter jejuni*/*coli* on pork carcasses in the northeast Georgia area. *Journal of Food Protection* 48, 808-810.

Bremell, T., Bjelle, A., and Svedhem, A. (1991). Rheumatic symptoms following an outbreak of *Campylobacter* enteritis: a five year follow up. *Annals of the Rheumatic Diseases* 50, 934-938.

Buchanan, C., Webb, A., Mutschall, S., Kruczkiewicz, P., Barker, D., Hetman, B., Gannon, V., Abbott, D., Thomas, J., Inglis, G., and Taboada, E. (2017). A genome-wide association study to identify diagnostic markers for human pathogenic *Campylobacter jejuni* strains. *Frontiers in Microbiology* 8, 1-9.

Buelow, D., Christensen, J., Neal-McKinney, J., and Konkel, M. (2011). *Campylobacter jejuni* survival within human epithelial cells is enhanced by the secreted protein CiaI. *Molecular Microbiology* 80, 1296-1312.

Burnens, A., Angéloz-Wick, B., and Nicolet, J. (1992). Comparison of *campylobacter* carriage rates in diarrheic and healthy pet animals. *Journal of Veterinary Medicine, Series B* 39, 175-180.

Busato, A., Hofer, D., Lentze, T., Gaillard, C., and Burnens, A. (1999). Prevalence and infection risks of zoonotic enteropathogenic bacteria in Swiss cow-calf farms. *Veterinary Microbiology* 69, 251-263.

Calva, J., Ruiz-Palacios, G., Lopez-Vidal, A., Ramos, A., and Bojalil, R. (1988). Cohort study of intestinal infection with *Campylobacter* in Mexican children. *Lancet* 1, 503-506.

Cappelier, J., and Federighi, M. (1998). Demonstration of viable but non-culturable state for *Campylobacter jejuni*. *Revue De Medecine Veterinaire* 149, 319-326.

Carleton, H., and Gerner-Smidt, P. (2016). Whole-genome sequencing is taking over foodborne disease surveillance. *Microbe* 11, 311-317.

Carrillo, C., Kruczkiewicz, P., Mutschall, S., Tudor, A., Clark, C., and Taboada, E. (2012). A framework for assessing the concordance of molecular typing methods and the true strain phylogeny of *Campylobacter jejuni* and *C. coli* using draft genome sequence data. *Frontiers in Cellular and Infection Microbiology* 2, 1-12.

Cawthraw, S., Lind, L., Kaijser, B., and Newell, D. (2000). Antibodies, directed towards *Campylobacter jejuni* antigens, in sera from poultry abattoir workers. *Clinical & Experimental Immunology* 122, 55-60.

CDC. (2018). Multistate outbreak of multidrug-resistant *Campylobacter* infections linked to contact with pet store puppies. Retrieved from https://www.cdc.gov/campylobacter/outbreaks/puppies-9-17/index.html

CFIA. (2016). *National microbiological baseline study in broiler chicken December 2012 - December 2013*. Canada: Retrieved from: http://www.inspection.gc.ca/food/chemical-residues-microbiology/food-safety-testing-bulletins/2016-08-17/december-2012-december-2013/eng/1471358115567/1471358175297?chap=5.

Chaban, B., Ngeleka, M., and Hill, J. (2010). Detection and quantification of 14 *Campylobacter* species in pet dogs reveals an increase in species richness in feces of diarrheic animals. *BMC Microbiology* 10, 1-73.

Chaisowwong, W., Kusumoto, A., Hashimoto, M., Harada, T., Maklon, K., and Kawamoto, K. (2012). Physiological characterization of *Campylobacter jejuni* under cold stresses conditions: its potential for public threat. *Journal of Veterinary Medical Science* 74, 43-50.

Champion, O., Best, E., and Frost, J. (2002). Comparison of pulsed-field gel electrophoresis and amplified fragment length polymorphism techniques for investigating outbreaks of enteritis due to campylobacters. *Journal of Clinical Microbiology* 40, 2263-2265.

Champion, O., Gaunt, M., Gundogdu, O., Elmi, A., Witney, A., Hinds, J., Dorrell, N., Wren, B., and Falkow, S. (2005). Comparative phylogenomics of the food-borne pathogen *Campylobacter jejuni* reveals genetic markers predictive of infection source. *Proceedings of the National Academy of Sciences of the United States of America* 102, 16043-16048.

Chewapreecha, C., Marttinen, P., Croucher, N., Salter, S., Harris, S., Mather, A., Hanage, W., Goldblatt, D., Nosten, F., Turner, C., Turner, P., Bentley, S., and Parkhill, J. (2014). Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes. *PLOS Genetics* 10, 1-11.

Christensen, J., Pacheco, S., and Konkel, M. (2009). Identification of a *Campylobacter jejuni*-secreted protein required for maximal invasion of host cells. *Molecular Microbiology* 73, 650-662.

Clark, C. (2011). Sequencing of CJIE1 prophages from *Campylobacter jejuni* isolates reveals the presence of inserted and (or) deleted genes. *Canadian Journal of Microbiology* 57, 795-809.

Clark, C., Price, L., Ahmed, R., Woodward, D., Melito, P., Rodgers, F., Jamieson, F., Ciebin, B., Li, A., and Ellis, A. (2003). Characterization of waterborne outbreak-associated *Campylobacter jejuni*, Walkerton, Ontario. *Emerging Infectious Diseases* 9, 1232-1241.

Clark, C., Taboada, E., Grant, C., Blakeston, C., Pollari, F., Marshall, B., Rahn, K., MacKinnon, J., Daignault, D., Pillai, D., and Ng, L. K. (2012). Comparison of molecular typing methods useful for detecting clusters of *Campylobacter jejuni* and *C. coli* isolates through routine surveillance. *Journal of Clinical Microbiology* 50, 798-809.

Colles, F., Jones, T., McCarthy, N., Sheppard, S., Cody, A., Dingle, K., Dawkins, M., and Maiden, M. (2008). *Campylobacter* infection of broiler chickens in a free-range environment. *Environmental Microbiology* 10, 2042-2050.

Cox, N., Richardson, L., Buhr, R., Fedorka-Cray, P., Bailey, J., Wilson, J., and Hiett, K. (2006). Natural presence of *Campylobacter* spp. in various internal organs of commercial broiler breeder hens. *Avian Diseases* 50, 450-453.

Cullen, T., O'Brien, J., Hendrixson, D., Giles, D., Hobb, R., Thompson, S., Brodbelt, J., and Trent, M. (2013). EptC of *Campylobacter jejuni* mediates phenotypes involved in host interactions and virulence. *Infection and Immunity* 81, 430-440.

Dasti, J., Tareen, A., Lugert, R., Zautner, A., and Gross, U. (2009). *Campylobacter jejuni*: A brief overview on pathogenicity-associated factors and disease-mediating mechanisms. *International Journal of Medical Microbiology* 300, 205-211.

de Zoete, M., Keestra, A., Roszczenko, P., and van Putten, J. (2010). Activation of human and chicken toll-like receptors by *Campylobacter* spp. *Infection and Immunity* 78, 1229-1238.

Dhillon, A., Shivaprasa, H., Schaberg, D., Wier, F., Weber, S., and Bandli, D. (2006). *Campylobacter jejuni* infection in broiler chickens. *Avian Diseases* 50, 55-58.

Dingle, K., Colles, F., Wareing, D., Ure, R., Fox, A., Bolton, F., Bootsma, H., Willems, R., Urwin, R., and Maiden, M. (2001). Multilocus sequence typing system for *Campylobacter jejuni*. *Journal of Clinical Microbiology* 39, 14-23.

Dobbin, G., Hariharan, H., Daoust, P., Hariharan, S., Heaney, S., Coles, M., Price, L., and Muckle, C. (2005). Bacterial flora of free-living Double-Crested Cormorant (*Phalacrocorax auritus*) chicks on Prince Edward Island, Canada, with reference to enteric bacteria and antibiotic resistance. *Comparative Immunology, Microbiology & Infectious Diseases* 28, 71-82.

Dorrell, N., Mangan, J., Laing, K., Hinds, J., Linton, D., Al-Ghusein, H., Barrell, B., Parkhill, J., Stoker, N., Karlyshev, A., Butcher, P., and Wren, B. (2001). Whole genome

comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Research* 11, 1706-1715.

Doyle, L. (1944). A *Vibrio* associated with swine dysentery. *American Journal of Veterinary Research* 5, 3-5.

Duong, T., and Konkel, M. (2009). Comparative studies of *Campylobacter jejuni* genomic diversity reveal the importance of core and dispensable genes in the biology of this enigmatic food-borne pathogen. *Current Opinion in Biotechnology* 20, 158-165.

Eberhart-Phillips, J., Walker, N., Garrett, N., Bell, D., Sinclair, D., Rainger, W., and Bates, M. (1997). Campylobacteriosis in New Zealand: results of a case-control study. *Journal of Epidemiology and Community Health* 51, 686-691.

Eberle, K., and Kiess, A. (2012). Phenotypic and genotypic methods for typing *Campylobacter jejuni* and *Campylobacter coli* in poultry. *Poultry Science* 91, 255-264.

EFSA. (2010). Analysis of the baseline survey on the prevalence of *Campylobacter* in broiler batches and of *Campylobacter* and *Salmonella* on broiler carcasses in the EU, 2008. *EFSA Journal* 8, 1503-1603.

Ekdahl, K., and Andersson, Y. (2004). Regional risks and seasonality in travel-associated campylobacteriosis. *BMC Infectious Diseases* 4, 1-7.

Elwell, C., and Dreyfus, L. (2000). DNase I homologous residues in CdtB are critical for Cytolethal Distending Toxin-mediated cell cycle arrest. *Molecular Microbiology* 37, 952-963.

Engvall, E., Brändström, B., Andersson, L., Båverud, V., Trowald-Wigh, G., and Englund, L. (2003). Isolation and identification of thermophilic *Campylobacter* species in faecal samples from Swedish dogs. *Scandinavian Journal of Infectious Diseases* 35, 713-718.

Ethelberg, S., Olsen, K., Gerner-Smidt, P., and Molbak, K. (2004). Household outbreaks among culture-confirmed cases of bacterial gastrointestinal disease. *American Journal of Epidemiology* 159, 406-412.

Eucker, T., and Konkel, M. (2012). The cooperative action of bacterial fibronectin-binding proteins and secreted proteins promote maximal *Campylobacter jejuni* invasion of host cells by stimulating membrane ruffling. *Cellular Microbiology* 14, 226-238.

Farhat, M., Shapiro, B., Kieser, K., Sultana, R., Jacobson, K., Victor, T., Warren, R., Streicher, E., Calver, A., Sloutsky, A., Kaur, D., Posey, J., Plikaytis, B., Oggioni, M., Gardy, J., Johnston, J., Rodrigues, M., Tang, P., Kato-Maeda, M., Borowsky, M., Muddukrishna, B., Kreiswirth, B., Kurepina, N., Galagan, J., Gagneux, S., Birren, B., Rubin, E., Lander, E., Sabeti, P., and Murray, M. (2013). Genomic analysis identifies

targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nature Genetics* 45, 1183-1189.

Farzan, A., Friendship, R., Cook, A., and Pollari, F. (2010). Occurrence of *Salmonella, Campylobacter, Yersinia enterocolitica, Escherichia coli* O157 and *Listeria monocytogenes* in swine. *Zoonoses and Public Health* 57, 388-396.

Fitzgerald, C. (2015). *Campylobacter*. *Clinics in Laboratory Medicine* 35, 289-298.

Flanagan, R., Neal-McKinney, J., Dhillon, A., Miller, W., and Konkel, M. (2009). Examination of *Campylobacter jejuni* putative adhesins leads to the identification of a new protein, designated FlpA, required for chicken colonization. *Infection and Immunity* 77, 2399-2407.

Florent, A. (1953). Isolation of a saprophytic *Vibrio* from bull sperm and vagina of cow, *Vibrio bulbulus*. *Comptes Rendus Des Seances De La Societe De Biologie Et De Ses Filiales* 147, 2066-2069.

Florent, A. (1959). Les deux vibrioses génitales: la vibriose due à *V. fetus venerialis* et la vibriose d'origine intestinale due à *V. fetus intestinalis*. *Meded . Veeartsenijsch. Rijksuniv. Gent, 3*, 1-60.

Fouts, D., Mongodin, E., Mandrell, R., Miller, W., Rasko, D., Ravel, J., Brinkac, L., DeBoy, R., Parker, C., Daugherty, S., Dodson, R., Durkin, A., Madupu, R., Sullivan, S., Shetty, J., Ayodeji, M., Shvartsbeyn, A., Schatz, M., Badger, J., Fraser, C., and Nelson, K. (2005). Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. *PLOS Biology* 3, 72-85.

Fox, J., Moore, R., and Ackerman, J. (1983). *Campylobacter jejuni*-associated diarrhea in dogs. *Journal of the American Veterinary Medical Association* 183, 1430-1433.

Franco, D. (1988). *Campylobacter* species: considerations for controlling a foodborne pathogen. *Journal of Food Protection* 51, 145-153.

Fratamico, P., Kawasaki, S., & Kawamoto, S. (2008). United States Patent No. 20080305484. *Genetic methods for speciating Campylobacter*. United States of America.

French, N., Barrigas, M., Brown, P., Ribiero, P., Williams, N., Leatherbarrow, H., Birtles, R., Bolton, E., Fearnhead, P., and Fox, A. (2005). Spatial epidemiology and natural population structure of *Campylobacter jejuni* colonizing a farmland ecosystem. *Environmental Microbiology* 7, 1116-1126.

Friedman, C., Hoekstra, R., Samuel, M., Marcus, R., Bender, J., Shiferaw, B., Reddy, S., Ahuja, S., Helfrick, D., Hardnett, F., Carter, M., Anderson, B., and Tauxe, R. (2004). Risk factors for sporadic *Campylobacter* infection in the United States: a case-control study in FoodNet sites. *Clinical Infectious Diseases* 38, S296.

Fullerton, K., Ingram, L., Jones, T., Anderson, B., McCarthy, P., Hurd, S., Shiferaw, B., Vugia, D., Haubert, N., Hayes, T., Wedel, S., Scallan, E., Henao, O., and Angulo, F. (2007). Sporadic *Campylobacter* infection in infants: a population-based surveillance case-control study. *The Pediatric Infectious Disease Journal* 26, 19-24.

Gaasbeek, E., Wagenaar, J., Guilhabert, M., van Putten, J., Parker, C., and van der Wal, F. (2010). Nucleases encoded by the integrated elements CJIE2 and CJIE4 inhibit natural transformation of *Campylobacter jejuni*. *Journal of Bacteriology* 192, 936-941.

Gaasbeek, E., Wagenaar, J., Guilhabert, M., Wösten, M., van Putten, J., van der Graaf-van Bloois, L, Parker, C., and van der Wal, F. (2009). A DNase encoded by integrated element CJIE1 inhibits natural transformation of *Campylobacter jejuni*. *Journal of Bacteriology* 191, 2296-2306.

Galanis, E. (2007). *Campylobacter* and bacterial gastroenteritis. *Canadian Medical Association Journal* 177, 570-571.

Garcia, M., Lior, H., Stewart, R., Ruckerbauer, G., Trudel, J., and Skljarevski, A. (1985). Isolation, characterization, and serotyping of *Campylobacter jejuni* and *Campylobacter coli* from slaughter cattle. *Applied and Environmental Microbiology* 49, 667-672.

Garcia-Rodríguez, L., Ruigómez, A., and Panés, J. (2006). Acute gastroenteritis is followed by an increased risk of Inflammatory Bowel Disease. *Gastroenterology* 130, 1588-1594.

Gardner, T., Fitzgerald, C., Xavier, C., Klein, R., Pruckler, J., Stroika, S., and McLaughlin, J. (2011). Outbreak of campylobacteriosis associated with consumption of raw peas. *Clinical Infectious Diseases* 53, 26-32.

Giacoboni, G., Itoh, K., Hirayama, K., Takahashi, E., and Mitsuoka, T. (2013). Comparison of fecal *Campylobacter* in calves and cattle of different ages and areas in Japan. *Journal of Veterinary Medical Science* 55, 555-559.

Gilbert, M., Kik, M., Timmerman, A., Severs, T., Kusters, J., Duim, B., and Wagenaar, J. (2014). Occurrence, diversity, and host association of intestinal *Campylobacter*, *Arcobacter*, and *Helicobacter* in reptiles. *PLOS One* 9, 1-8.

Gilpin, B., Robson, B., Scholes, P., Nourozi, F., and Sinton, L. (2009). Survival of *Campylobacter* spp. in bovine faeces on pasture. *Letters in Applied Microbiology* 48, 162-166.

Gourley, C., & Konkel, M. (2016). *Campylobacter jejuni*: molecular mechanisms and animal models of colonization and disease. In S. Singh (Ed.), *Human Emerging and Re-emerging Infections: Bacterial & Mycotic Infections, Volume II* (First ed., pp. 535-553). Hoboken, NJ: John Wiley & Sons, Inc.

Grajewski, B., Kusek, J., and Gelfand, H. (1985). Development of a bacteriophage typing system for *Campylobacter jejuni* and *Campylobacter coli*. *Journal of Clinical Microbiology* 22, 13-18.

Gripp, E., Hlahla, D., Didelot, X., Kops, F., Maurischat, S., Tedin, K., Alter, T., Ellerbroek, L., Schreiber, K., Schomburg, D., Janssen, T., Bartholomäus, P., Hofreuter, D., Woltemate, S., Uhr, M., Brenneke, B., Grüning, P., Gerlach, G., Wieler, L., Suerbaum, S., and Josenhans, C. (2011). Closely related *Campylobacter jejuni* strains from different sources reveal a generalist rather than a specialist lifestyle. *BMC Genomics* 12, 584-605.

Guerry, P. (2007). *Campylobacter* flagella: not just for motility. *Trends in Microbiology* 15, 456-461.

Guerry, P., Szymanski, C., Prendergast, M., Hickey, T., Ewing, C., Pattarini, D., and Moran, A. (2002). Phase variation of *Campylobacter jejuni* 81-176 lipooligosaccharide affects ganglioside mimicry and invasiveness in vitro. *Infection and Immunity* 70, 787-793.

Guevremont, E., Higgins, R., and Quessy, S. (2004). Characterization of *Campylobacter* isolates recovered from clinically healthy pigs and from sporadic cases of campylobacteriosis in humans. *Journal of Food Protection* 67, 228-234.

Haddock, G., Mullin, M., MacCallum, A., Sherry, A., Tetley, L., Watson, E., Dagleish, M., Smith, D., and Everest, P. (2010). *Campylobacter jejuni* 81-176 forms distinct microcolonies on in vitro-infected human small intestinal tissue prior to biofilm formation. *Microbiology* 156, 3079-3084.

Hald, B., Pedersen, K., Waino, M., Jorgensen, J., and Madsen, M. (2004). Longitudinal study of the excretion patterns of thermophilic *Campylobacter* spp. in young pet dogs in Denmark. *Journal of Clinical Microbiology* 42, 2003-2012.

Hald, B., Skov, M., Nielsen, E., Rahbek, C., Madsen, J., Waino, M., Chriel, M., Nordentoft, S., Baggesen, D., and Madsen, M. (2016). *Campylobacter jejuni* and *Campylobacter coli* in wild birds on Danish livestock farms. *Acta Veterinaria Scandinavica* 58, 11.

Hanninen, M., Hakkinen, M., and Rautelin, H. (1999). Stability of related human and chicken *Campylobacter jejuni* genotypes after passage through chick intestine studied by pulsed-field gel electrophoresis. *Applied and Environmental Microbiology* 65, 2272-2275.

Hannon, S., Taboada, E., Russell, M., Allan, B., Waldner, C., Wilson, H., Potter, A., Babiuk, L., and Townsend, H. (2009). Genomics-based molecular epidemiology of *Campylobacter jejuni* isolates from feedlot cattle and from people in Alberta, Canada. *Journal of Clinical Microbiology* 47, 410-420.

Havelaar, A., van Pelt, W., Ang, C., Wagenaar, J., van Putten, J., Gross, U., and Newell, D. (2009). Immunity to *Campylobacter*: its role in risk assessment and epidemiology. *Critical Reviews in Microbiology* 35, 1-22.

Hickey, T., McVeigh, A., Scott, D., Michielutti, R., Bixby, A., Carroll, S., Bourgeois, A., and Guerry, P. (2000). *Campylobacter jejuni* Cytolethal Distending Toxin mediates release of Interleukin-8 from intestinal epithelial cells. *Infection and Immunity* 68, 6535-6541.

Hoar, B., Atwill, E., Elmi, C., and Farver, T. (2001). An examination of risk factors associated with beef cattle shedding pathogens of potential zoonotic concern. *Epidemiology and Infection* 127, 147-155.

Hoar, B., Atwill, E., Elmi, C., Utterback, W., and Edmondson, A. (1999). Comparison of fecal samples collected per rectum and off the ground for estimation of environmental contamination attributable to beef cattle. *American Journal of Veterinary Research* 60, 1352-1356.

Hoar, B., Whiteside, D., Ward, L., Inglis, G., and Morck, D. (2007). Evaluation of the enteric microflora of captive Whooping Cranes (*Grus americana*) and Sandhill Cranes (*Grus canadensis*). *Zoo Biology* 26, 141-153.

Hopkins, R., Olmsted, R., and Istre, G. (1984). Endemic *Campylobacter jejuni* infection in Colorado: identified risk factors. *American Journal of Public Health* 74, 249-250.

Horrocks, S., Anderson, R., Nisbet, D., and Ricke, S. (2009). Incidence and ecology of *Campylobacter jejuni* and *coli* in animals. *Anaerobe* 15, 18-25.

Huang, H., Brooks, B., Lowman, R., and Carrillo, C. (2015). *Campylobacter* species in animal, food, and environmental sources, and relevant testing programs in Canada. *Canadian Journal of Microbiology* 61, 701-721.

Humphrey, T., and Beckett, P. (1987). *Campylobacter jejuni* in dairy cows and raw milk. *Epidemiology and Infection* 98, 263-269.

Humphrey, T., O'Brien, S., and Madsen, M. (2007). Campylobacters as zoonotic pathogens: a food production perspective. *International Journal of Food Microbiology* 117, 237-257.

Inglis, G., Hoar, B., Whiteside, D., and Morck, D. (2007). *Campylobacter canadensis* sp. nov., from captive Whooping Cranes in Canada. *International Journal of Systematic and Evolutionary Microbiology* 57, 2636-2644.

Inglis, G., Kalischuk, L., and Busz, H. (2003). A survey of *Campylobacter* species shed in faeces of beef cattle using Polymerase Chain Reaction. *Canadian Journal of Microbiology* 49, 655-661.

Inglis, G., Kalischuk, L., and Busz, H. (2004). Chronic shedding of *Campylobacter* species in beef cattle. *Journal of Applied Microbiology* 97, 410-420.

Inglis, G., McAllister, T., Larney, F., and Topp, E. (2010). Prolonged survival of *Campylobacter* species in bovine manure compost. *Applied and Environmental Microbiology* 76, 1110-1119.

Jerome, J., Bell, J., Plovanich-Jones, A., Barrick, J., Brown, C., and Mansfield, L. (2011). Standing genetic variation in contingency loci drives the rapid adaptation of *Campylobacter jejuni* to a novel host. *PLOS One* 6, 1-11.

Jin, S., Joe, A., Lynett, J., and Hani, E. (2001). JlpA, a novel surface-exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial cells. *Molecular Microbiology* 39, 1225-1236.

Jokinen, C., Edge, T., Ho, S., Koning, W., Laing, C., Mauro, W., Medeiros, D., Miller, J., Robertson, W., Taboada, E., Thomas, J., Topp, E., Ziebell, K., and Gannon, V. (2011). Molecular subtypes of *Campylobacter* spp., *Salmonella enterica*, and *Escherichia coli* O157:H7 isolated from faecal and surface water samples in the Oldman River watershed, Alberta, Canada. *Water Research* 45, 1247-1257.

Jokinen, C., Schreier, H., Mauro, W., Taboada, E., Isaac-Renton, J., Topp, E., Edge, T., Thomas, J., and Gannon, V. (2010). The occurrence and sources of *Campylobacter* spp., *Salmonella enterica* and *Escherichia coli* O157:H7 in the Salmon River, British Columbia, Canada. *Journal of Water and Health* 8, 374-386.

Jones, F., Orcutt, M., and Little, R. (1931). Vibrios (*Vibrio jejuni*, n. sp.) associated with intestinal disorders of cows and calves. *The Journal of Experimental Medicine* 53, 853-863.

Kaakoush, N., Castaño-Rodríguez, N., Mitchell, H., and Man, S. (2015). Global epidemiology of *Campylobacter* infection. *Clinical Microbiology Reviews* 28, 687-720.

Kakoyiannis, C., Winter, P., and Marshall, R. (1988). The relationship between intestinal *Campylobacter* species isolated from animals and humans as determined by BRENDA. *Epidemiology and Infection* 100, 379-387.

Kapperud, G., and Aasen, S. (1992). Descriptive epidemiology of infections due to thermotolerant *Campylobacter* spp. in Norway, 1979–1988. *Journal of Pathology, Microbiology and Immunology* 100, 883-890.

Keller, J., Shriver, W., Waldenström, J., Griekspoor, P., and Olsen, B. (2011). Prevalence of *Campylobacter* in wild birds of the mid-Atlantic region, USA. *Journal of Wildlife Diseases* 47, 750-754.

Keo, T., Collins, J., Kunwar, P., Blaser, M., and Iovine, N. (2011). *Campylobacter* capsule and lipooligosaccharide confer resistance to serum and cationic antimicrobials. *Virulence* 2, 30-40.

Khan, I., Gannon, V., Jokinen, C., Kent, R., Koning, W., Lapen, D., Medeiros, D., Miller, J., Neumann, N., Phillips, R., Schreier, H., Topp, E., van Bochove, E., Wilkes, G., and Edge, T. (2014). A national investigation of the prevalence and diversity of thermophilic *Campylobacter* species in agricultural watersheds in Canada. *Water Research* 61, 243-252.

King, E. (1957). Human infections with *Vibrio fetus* and a closely related *Vibrio*. *The Journal of Infectious Diseases* 101, 119-128.

Kirkpatrick, B., and Tribble, D. (2011). Update on human *Campylobacter jejuni* infections. *Current Opinion in Gastroenterology* 27, 1-7.

Koene, M., Houwers, D., Dijkstra, J., Duim, B., and Wagenaar, J. (2004). Simultaneous presence of multiple *Campylobacter* species in dogs. *Journal of Clinical Microbiology* 42, 819-821.

Koga, T., Aoki, W., Mizuno, T., Wakazono, K., Ohno, J., Nakai, T., Nomiya, T., Fujii, M., Fusegawa, K., Kinoshita, K., Hamada, T., and Ikeda, Y. (2015). Antimicrobial resistance in *Campylobacter coli* and *Campylobacter jejuni* in Cynomolgus monkeys (*Macaca fascicularis*) and eradication regimens. *Journal of Microbiology, Immunology and Infection* 50, 75-82.

Konkel, M., Kim, B., Rivera-Amill, V., and Garvis, S. (1999). Bacterial secreted proteins are required for the internalization of *Campylobacter jejuni* into cultured mammalian cells. *Molecular Microbiology* 32, 691-701.

Konkel, M., Larson, C., and Flanagan, R. (2010). *Campylobacter jejuni* FlpA binds fibronectin and is required for maximal host cell adherence. *Journal of Bacteriology* 192, 68-76.

Konkel, M., Monteville, M., Rivera-Amill, V., and Joens, L. (2001). The pathogenesis of *Campylobacter jejuni*-mediated enteritis. *Current Issues in Intestinal Microbiology* 2, 55-71.

Kovats, R., Edwards, S., Charron, D., Cowden, J., D'Souza, R., Ebi, K., Gauci, C., Gerner-Smidt, P., Hajat, S., Hales, S., Hernández Pezzi, G., Kriz, B., Kutsar, K., McKeown, P., Mellou, K., Menne, B., O'Brien, S., van Pelt, W., and Schmid, H. (2005). Climate variability and *Campylobacter* infection: an international study. *International Journal of Biometeorology* 49, 207-214.

Kwan, P., Xavier, C., Santovenia, M., Pruckler, J., Stroika, S., Joyce, K., Gardner, T., Fields, P., McLaughlin, J., Tauxe, R., and Fitzgerald, C. (2014). Multilocus sequence

typing confirms wild birds as the source of a *Campylobacter* outbreak associated with the consumption of raw peas. *Applied and Environmental Microbiology* 80, 4540-4546.

Kwon, Y., Oh, J., Jeong, O., Moon, O., Kang, M., Jung, B., An, B., Youn, S., Kim, H., Jang, I., and Lee, H. (2017). Prevalence of *Campylobacter* species in wild birds of South Korea. *Avian Pathology* 46, 474-480.

Lacey, R. (1993). Food-borne bacterial infections. *Human Nutrition and Parasitic Infection* 107, 75-93.

Lanata, C., Fischer-Walker, C., Olascoaga, A., Torres, C., and Aryee, M. (2013). Global causes of diarrheal disease mortality in children *PLOS One* 8, 1-11.

Lara-Tejero, M., and Galán, J. (2000). A bacterial toxin that controls cell cycle progression as a Deoxyribonuclease I-like protein. *Science* 290, 354-357.

Lara-Tejero, M., and Galán, J. (2001). CdtA, CdtB, and CdtC form a tripartite complex that is required for Cytolethal Distending Toxin activity. *Infection and Immunity* 69, 4358-4365.

Larson, C., Christensen, J., Pacheco, S., Minnich, S., & Konkel, M. (2008). *Campylobacter jejuni* secretes proteins via the flagellar Type III Secretion System that contribute to host cell invasion and gastroenteritis. In I. Nachamkin, C. Szymanski & M. Blaser (Eds.), *Campylobacter* (Third ed., pp. 315-332). Washington, DC: ASM Press.

Leahy, A., Cummings, K., Rodriguez-Rivera, L., Rankin, S., and Hamer, S. (2016). Evaluation of faecal *Salmonella* shedding among dogs at seven animal shelters across Texas. *Zoonoses and Public Health* 63, 515-521.

Lecuit, M., Abachin, E., Martin, A., Poyart, C., Pochart, P., Suarez, F., Bengoufa, D., Feuillard, J., Lavergne, A., Gordon, J., Berche, P., Guillevin, L., and Lortholary, O. (2004). Immunoproliferative small intestinal disease associated with *Campylobacter jejuni*. *The New England Journal of Medicine* 350, 239-248.

Leonard, E., Pearl, D., Janecko, N., Weese, J., Reid-Smith, R., Peregrine, A., and Finley, R. (2011). Factors related to *Campylobacter* spp. carriage in client-owned dogs visiting veterinary clinics in a region of Ontario, Canada. *Epidemiology and Infection* 139, 1531-1541.

Leonard, E., Takata, T., Blaser, M., Falkow, S., Tompkins, L., and Gaynor, E. (2003). Use of an open-reading frame-specific *Campylobacter jejuni* DNA microarray as a new genotyping tool for studying epidemiologically related isolates. *The Journal of Infectious Diseases* 187, 691-694.

Lior, H. (1984). New, extended biotyping scheme for *Campylobacter jejuni, Campylobacter coli,* and *"Campylobacter laridis"*. *Journal of Clinical Microbiology* 20, 636-640.

Lior, H., Woodward, D., Edgar, J., Laroche, L., and Gill, P. (1982). Serotyping of *Campylobacter jejuni* by slide agglutination based on heat-labile antigenic factors. *Journal of Clinical Microbiology* 15, 761-768.

Little, C., and de Louvois, J. (1998). The microbiological examination of butchery products and butchers' premises in the United Kingdom. *Journal of Applied Microbiology* 85, 177-186.

Llarena, A., Taboada, E., and Rossi, M. (2017). Whole-genome sequencing in epidemiology of *Campylobacter jejuni* infections. *Journal of Clinical Microbiology* 55, 1269-1275.

Logan, J., Burnens, A., Linton, D., Lawson, A., and Stanley, J. (2000). *Campylobacter lanienae* sp. nov., a new species isolated from workers in an abattoir. *International Journal of Systematic and Evolutionary Microbiology* 50, 865-872.

Louwen, R., Heikema, A., van Belkum, A., Ott, A., Gilbert, M., Ang, W., Endtz, H., Bergman, M., and Nieuwenhuis, E. (2008). The sialylated lipooligosaccharide outer core in *Campylobacter jejuni* is an important determinant for epithelial cell invasion. *Infection and Immunity* 76, 4431-4438.

Louwen, R., Horst-Kreft, D., de Boer, A., van der Graaf, L., de Knegt, G., Hamersma, M., Heikema, A., Timms, A., Jacobs, B., Wagenaar, J., Endtz, H., van der Oost, J., Wells, J., Nieuwenhuis, E., van Vliet, A., Willemsen, P., van Baarlen, P., and van Belkum, A. (2013). A novel link between *Campylobacter jejuni* bacteriophage defence, virulence and Guillain-Barré Syndrome. *European Journal of Clinical Microbiology and Infectious Diseases* 32, 207-226.

Louwen, R., van Belkum, A., Wagenaar, J., Doorduyn, Y., Achterberg, R., and Endtz, H. (2006). Lack of association between the presence of the pVir plasmid and bloody diarrhea in *Campylobacter jejuni* enteritis. *Journal of Clinical Microbiology* 44, 1867-1868.

Magajna, B., and Schraft, H. (2015). *Campylobacter jejuni* biofilm cells become viable but non-culturable (VBNC) in low nutrient conditions at 4°C more quickly than their planktonic counterparts. *Food Control* 50, 45-50.

Majowicz, S., McNab, W., Sockett, P., Henson, S., Dore, K., Edge, V., Buffett, M., Fazil, A., Read, S., McEwen, S., Stacey, D., and Wilson, J. (2006). Burden and cost of gastroenteritis in a Canadian community. *Journal of Food Protection* 69, 651-659.

Manning, G., Dowson, C., Bagnall, M., Ahmed, I., West, M., and Newell, D. (2003). Multilocus sequence typing for comparison of veterinary and human isolates of *Campylobacter jejuni*. *Applied and Environmental Microbiology* 69, 6370-6379.

McFadyean, J., & Stockman, S. (1913). *Report of the departmental committee appointed by the Board of Agriculture and Fisheries to inquire into epizootic abortion.* London:

Meldrum, R., Griffiths, J., Smith, R., and Evans, M. (2005). The seasonality of human *Campylobacter* infection and *Campylobacter* isolates from fresh, retail chicken in Wales. *Epidemiology and Infection* 133, 49-52.

Miller, G., Dunn, G., Smith-Palmer, A., Ogden, I., and Strachan, N. (2004). Human campylobacteriosis in Scotland: seasonality, regional trends and bursts of infection. *Epidemiology and Infection* 132, 585-593.

Minihan, D., Whyte, P., O'Mahony, M., Fanning, S., McGill, K., and Collins, J. (2004). *Campylobacter spp.* in Irish feedlot cattle: a longitudinal study involving pre-harvest and harvest phases of the food chain. *Journal of Veterinary Medicine, Series B* 51, 28-33.

Mohan, V., Stevenson, M., Marshall, J., and French, N. (2017). Characterisation by multilocus sequence and porA and flaA typing of *Campylobacter jejuni* isolated from samples of dog faeces collected in one city in New Zealand. *New Zealand Veterinary Journal* 65, 209-213.

Monteville, M., and Konkel, M. (2002). Fibronectin-facilitated invasion of T84 eukaryotic cells by *Campylobacter jejuni* occurs preferentially at the basolateral cell surface. *Infection and Immunity* 70, 6665-6671.

Monteville, M., Yoon, J., and Konkel, M. (2003). Maximal adherence and invasion of INT 407 cells by *Campylobacter jejuni* requires the CadF outer-membrane protein and microfilament reorganization. *Microbiology* 149, 153-165.

Moore, J., Garcia, M., and Madden, R. (2002). Subspecies characterization of porcine *Campylobacter coli* and *Campylobacter jejuni* by multilocus enzyme electrophoresis typing. *Veterinary Research Communications* 26, 1-9.

Moore, J., and Goldsmith, C. (2001). Phenotyping of *Campylobacter jejuni* and *Campylobacter coli* by a quantitative antibiogram [MIC] typing scheme using euclidean distances [QATED]. *BMC Microbiology* 1, 1-4.

Moore, J., and Madden, R. (2003). Comparison of eight phenotypic methods for subspecies characterization of thermophilic *Campylobacter* spp. isolated from pig liver. *Journal of Food Protection* 66, 1079-1084.

Morooka, T., Umeda, A., and Amako, K. (1985). Motility as an intestinal colonization factor for *Campylobacter jejuni*. *Journal of General Microbiology* 131, 1973-1980.

Mortensen, N., Schiellerup, P., Boisen, N., Klein, B., Locht, H., Abuoun, M., Newell, D., and Krogfelt, K. (2011). The role of *Campylobacter jejuni* Cytolethal Distending Toxin in gastroenteritis: toxin detection, antibody production, and clinical outcome. *Apmis* 119, 626-634.

Munroe, D., Prescott, J., and Penner, J. (1983). *Campylobacter jejuni* and *Campylobacter coli* serotypes isolated from chickens, cattle, and pigs. *Journal of Clinical Microbiology* 18, 877-881.

Murphy, C., Carroll, C., and Jordan, K. (2006). Environmental survival mechanisms of the foodborne pathogen *Campylobacter jejuni*. *Journal of Applied Microbiology* 100, 623-632.

Musher, D., and Musher, B. (2004). Contagious acute gastrointestinal infections. *The New England Journal of Medicine* 351, 2417-2427.

Mutschall,S., B. Hetman, C. Jardine and E. Taboada. 2015. *Do raccoons play a significant role in the transmission dynamics of Campylobacter jejuni at the interface of wild animal populations and livestock?* Presented at the *18th International Workshop on Campylobacter, Helicobacter and Related Organisms,* New Zealand. 87.

Navarro-Gonzalez, N., Ugarte-Ruiz, M., Porrero, M., Zamora, L., Mentaberre, G., Serrano, E., Mateos, A., Lavín, S., and Domínguez, L. (2014). *Campylobacter* shared between free-ranging cattle and sympatric wild ungulates in a natural environment (NE spain). *EcoHealth* 11, 333-342.

Neal-McKinney, J., and Konkel, M. (2012). The *Campylobacter jejuni* CiaC virulence protein is secreted from the flagellum and delivered to the cytosol of host cells. *Frontiers in Cellular and Infection Microbiology* 2, 1-15.

Ng, L., Sherburne, R., Taylor, T., and Stiles, M. (1985). Morphological forms and viability of *Campylobacter* species studied by electron microscopy. *Journal of Bacteriology* 164, 338-343.

Nicholson, F., Groves, S., and Chambers, B. (2005). Pathogen survival during livestock manure storage and following land application. *Bioresource Technology* 96, 135-143.

Nielsen, E. (2002). Occurrence and strain diversity of thermophilic campylobacters in cattle of different age groups in dairy herds. *Letters in Applied Microbiology* 35, 85-89.

Nishikubo, S., Ohara, M., Ueno, Y., Ikura, M., Kurihara, H., Komatsuzawa, H., Oswald, E., and Sugai, M. (2003). An N-terminal segment of the active component of the bacterial genotoxin Cytolethal Distending Toxin B (CDTB) directs CDTB into the nucleus. *Journal of Biological Chemistry* 278, 50671-50681.

Nylen, G., Dunstan, F., Palmer, S., Andersson, Y., Bager, F., Cowden, J., Feierl, G., Galloway, Y., Kapperud, G., Megraud, F., Molbak, K., Petersen, L., and Ruutu, P. (2002). The seasonal distribution of *Campylobacter* infection in nine European countries and New Zealand. *Epidemiology and Infection* 128, 383-390.

Olson, C., Ethelberg, S., van Pelt, W., & Tauxe, R. (2008). Epidemiology of *Campylobacter jejuni* infections in industrialized nations. In I. Nachamkin, C. Szymanski & M. Blaser (Eds.), *Campylobacter* (Third ed., pp. 163-189). Washington, DC: ASM Press.

On, S., McCarthy, N., Miller, W., & Gilpin, B. (2008). Molecular epidemiology of *Campylobacter* species. In I. Nachamkin, C. Szymanski & M. Blaser (Eds.), *Campylobacter* (Third ed., pp. 191-212). Washington, DC: ASM Press.

Oosterom, J., Dekker, R., de Wilde, G., van Kempen-de Troye, F., and Engels, G. (1985). Prevalence of *Campylobacter jejuni* and *Salmonella* during pig slaughtering. *The Veterinary Quarterly* 7, 31-34.

Oosterom, J., Notermans, S., Karman, H., and Engels, G. (1983). Origin and prevalence of *Campylobacter jejuni* in poultry processing. *Journal of Food Protection* 46, 339-344.

Park, S. (2002). The physiology of *Campylobacter* species and its relevance to their role as foodborne pathogens. *International Journal of Food Microbiology* 74, 177-188.

Parker, C., Miller, W., Horn, S., and Lastovica, A. (2007). Common genomic features of *Campylobacter jejuni* subsp. *doylei* strains distinguish them from *C. jejuni* subsp. *jejuni*. *BMC Microbiology* 7, 1-9.

Parker, C., Quiñones, B., Miller, W., Horn, S., and Mandrell, R. (2006). Comparative genomic analysis of *Campylobacter jejuni* strains reveals diversity due to genomic elements similar to those present in *C. jejuni* strain RM1221. *Journal of Clinical Microbiology* 44, 4125-4135.

Parkhill, J., Wren, B., Mungall, K., Ketley, J., Churcher, C., Basham, D., Chillingworth, T., Davies, R., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A., Moule, S., Pallen, M., Penn, C., Quail, M., Rajandream, M., Rutherford, K., van Vilet, A., Whitehead, S., and Barrell, B. (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* 403, 665-668.

Patrick, M., Christiansen, L., Wainø, M., Ethelberg, S., Madsen, H., and Wegener, H. (2004). Effects of climate on incidence of *Campylobacter* spp. in humans and prevalence in broiler flocks in Denmark. *Applied and Environmental Microbiology* 70, 7474-7480.

Pearce, R., Wallace, F., Call, J., Dudley, R., Oser, A., Yoder, L., Sheridan, J., and Luchansky, J. (2003). Prevalence of *Campylobacter* within a swine slaughter and processing facility. *Journal of Food Protection* 66, 1550-1556.

Pearson, B., Pin, C., Wright, J., I'Anson, K., Humphrey, T., and Wells, J. (2003). Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. *FEBS Letters* 554, 224-230.

Pei, Z., Burucoa, C., Grignon, B., Baqar, S., Huang, X., Kopecko, D., Bourgeois, A., Fauchere, J., and Blaser, M. (1998). Mutation in the Peb1A locus of *Campylobacter jejuni* reduces interactions with epithelial cells and intestinal colonization of mice. *Infection and Immunity* 66, 938-943.

Penner, J., and Hennessy, J. (1980). Passive hemagglutination technique for serotyping *Campylobacter fetus* subsp. *jejuni* on the basis of soluble heat-stable antigens. *Journal of Clinical Microbiology* 12, 732-737.

Petersen, L., Nielsen, E., Engberg, J., On, S., and Dietz, H. (2001). Comparison of genotypes and serotypes of *Campylobacter jejuni* isolated from Danish wild mammals and birds and from broiler flocks and humans. *Applied and Environmental Microbiology* 67, 3115-3121.

PHAC. (2009). Canadian Integrated Surveillance Report: *Salmonella, Campylobacter,* verotoxigenic *E. coli* and *Shigella,* from 2000 to 2004. *Canada Communicable Disease Report* 35S3, 1-50.

PHAC. (2015a). *FoodNet Canada Biennial Report 2011–2012.* Canada: Retrieved from: https://www.canada.ca/en/public-health/services/surveillance/foodnet-canada/publications.html.

PHAC. (2015b). *FoodNet Canada Short Report 2014.* Canada:

Pickett, C., and Whitehouse, C. (1999). The Cytolethal Distending Toxin family. *Trends in Microbiology* 7, 292-297.

Pitkänen, T. (2013). Review of *Campylobacter* spp. in drinking and environmental waters. *Journal of Microbiological Methods* 95, 39-47.

Pointon, A., Sexton, M., Dowsett, P., Saputra, T., Kiermeier, A., Lorimer, M., Holds, G., Arnold, G., Davos, D., Combs, B., Fabiansson, S., Raven, G., McKenzie, H., Chapman, A., and Sumner, J. (2008). A baseline survey of the microbiological quality of chicken portions and carcasses at retail in two Australian states (2005 to 2006). *Journal of Food Protection* 71, 1123-1134.

Pope, J., Krizova, A., Garg, A., Thiessen-Philbrook, H., and Ouimet, J. (2007). *Campylobacter* Reactive Arthritis: a systematic review. *Seminars in Arthritis and Rheumatism* 37, 48-55.

Potter, R., Kaneene, J., and Hall, W. (2003). Risk factors for sporadic *Campylobacter jejuni* infections in rural Michigan: a prospective case-control study. *American Journal of Public Health* 93, 2118-2123.

Prévot, A. (1940). Études de systématique bactérienne. V. essai de classification des vibrions anaérobies. *Annales De L'Institut Pasteur* 61, 117-125.

Quainoo, S., Coolen, J., van Hijum, S., Huynen, M., Melchers, W., van Schaik, W., and Wertheim, H. (2017). Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. *Clinical Microbiology Reviews* 30, 1015-1063.

Quessy, S., and Messier, S. (1992). Prevalence of *Salmonella* spp., *Campylobacter* spp. and *Listeria* spp. in Ring-Billed Gulls (*Larus delawarensis*). *Journal of Wildlife Diseases* 28, 526-531.

Rainwater, K., Marchese, K., Slavinski, S., Humberg, L., Dubovi, E., Jarvis, J., McAloose, D., and Calle, P. (2017). Health survey of free-ranging raccoons *(Procyon lotor)* in Central Park, New York, New York, USA: implications for human and domestic animal health. *Journal of Wildlife Diseases* 53, 272-284.

Rao, M., Naficy, A., Savarino, S., Abu-Elyazeed, R., Wierzba, T., Peruski, L., Abdel-Messih, I., Frenck, R., and Clemens, J. (2001). Pathogenicity and convalescent excretion of *Campylobacter* in rural Egyptian children. *American Journal of Epidemiology* 154, 166-173.

Read, T., and Massey, R. (2014). Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology. *Genome Medicine* 6, 1-11.

Ribeiro, C., Thomas, M., Kembrey, D., Magee, J., and North, Z. (1996). Resistotyping of campylobacters: fulfilling a need. *Epidemiology and Infection* 116, 169-175.

Roberts, I. (1996). The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annual Review of Microbiology* 50, 285-315.

Rossi, M., Hänninen, M., Revez, J., Hannula, M., and Zanoni, R. (2008). Occurrence and species level diagnostics of *Campylobacter* spp., enteric *Helicobacter* spp. and *Anaerobiospirillum* spp. in healthy and diarrheic dogs and cats. *Veterinary Microbiology* 129, 304-314.

Sails, A., Swaminathan, B., and Fields, P. (2003). Utility of multilocus sequence typing as an epidemiological tool for investigation of outbreaks of gastroenteritis caused by *Campylobacter jejuni*. *Journal of Clinical Microbiology* 41, 4733-4739.

Samuelson, D., Eucker, T., Bell, J., Dybas, L., Mansfield, L., and Konkel, M. (2013). The *Campylobacter jejuni* CiaD effector protein activates MAP kinase signaling pathways and is required for the development of disease. *Cell Communication and Signalling* 11, 79-94.

Saunders, S., Smith, K., Schott, R., Dobbins, G., and Scheftel, J. (2017). Outbreak of campylobacteriosis associated with raccoon contact at a wildlife rehabilitation centre, Minnesota, 2013. *Zoonoses and Public Health* 64, 222-227.

Schmidt-Ott, R., Pohl, S., Burghard, S., Weig, M., and Groß, U. (2005). Identification and characterization of a major subgroup of conjugative *Campylobacter jejuni* plasmids. *Journal of Infection* 50, 12-21.

Schuster, C., Ellis, A., Robertson, W., Charron, D., Aramini, J., Marshall, B., and Medeiros, D. (2005). Infectious disease outbreaks related to drinking water in Canada, 1974-2001. *Canadian Journal of Public Health* 96, 254-258.

Sebald, M., and Veron, M. (1963). Teneur en bases de l'ADN et classification des vibrions. *Annales De L'Institut Pasteur* 105, 897-910.

Shen, Z., Feng, Y., Dewhirst, F., and Fox, J. (2001). Coinfection of enteric *Helicobacter* spp. and *Campylobacter* spp. in cats. *Journal of Clinical Microbiology* 39, 2166-2172.

Sheppard, S., Dallas, J., MacRae, M., McCarthy, N., Sproston, E., Gormley, F., Strachan, N., Ogden, I., Maiden, M., and Forbes, K. (2009). *Campylobacter* genotypes from food animals, environmental sources and clinical disease in Scotland 2005/6. *International Journal of Food Microbiology* 134, 96-103.

Sheppard, S., Didelot, X., Meric, G., Torralbo, A., Jolley, K., Kelly, D., Bentley, S., Maiden, M., Parkhill, J., and Falush, D. (2013). Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. *Proceedings of the National Academy of Sciences of the United States* 110, 11923-11927.

Shreeve, J., Toszeghy, M., Pattison, M., and Newell, D. (2000). Sequential spread of *Campylobacter* infection in a multipen broiler house. *Avian Diseases* 44, 983-988.

Shulman, S., Friedman, H., and Sims, R. (2007). Theodor Escherich: The first pediatric infectious diseases physician? *Clinical Infectious Diseases* 45, 1025-1029.

Sinton, L., Braithwaite, R., Hall, C., and Mackenzie, M. (2007). Survival of indicator and pathogenic bacteria in bovine feces on pasture. *Applied and Environmental Microbiology* 73, 7917-7925.

Skirrow, M. (1987). A demographic survey of *Campylobacter, Salmonella* and *Shigella* infections in England: a public health laboratory service survey. *Epidemiology and Infection* 99, 647-657.

Skirrow, M. (2006). John McFadyean and the centenary of the first isolation of *Campylobacter* species. *Clinical Infectious Diseases* 43, 1213-1217.

Skirrow, M., and Benjamin, J. (1980). Differentiation of enteropathogenic *Campylobacter*. *Journal of Clinical Pathology* 33, 1122-1122.

Smith, T., and Taylor, M. (1919). Some morphological and biological characters of the spirilla (*vibrio fetus*, n. sp.) associated with disease of the fetal membranes in cattle. *The Journal of Experimental Medicine* 30, 299-311.

Southern, J., Smith, R., and Palmer, S. (1990). Bird attack on milk bottles: possible mode of transmission of *Campylobacter jejuni* to man. *Lancet* 336, 1425-1427.

Stanley, K., and Jones, K. (2003). Cattle and sheep farms as reservoirs of *Campylobacter*. *Journal of Applied Microbiology* 94, 104-113.

Stanley, K., Wallace, J., Currie, J., Diggle, P., and Jones, K. (1998a). The seasonal variation of thermophilic campylobacters in beef cattle, dairy cattle and calves. *Journal of Applied Microbiology* 85, 472-480.

Stanley, K., Wallace, J., Currie, J., Diggle, P., and Jones, K. (1998b). Seasonal variation of thermophilic campylobacters in lambs at slaughter. *Journal of Applied Microbiology* 84, 1111-1116.

Suerbaum, S., Lohrengel., M., Sonnevend, A., Ruberg, F., and Kist, M. (2001). Allelic diversity and recombination in *Campylobacter jejuni*. *Journal of Bacteriology* 183, 2553-2559.

Taboada, E., Acedillo, R., Carrillo, C., Findlay, W., Medeiros, D., Mykytczuk, O., Roberts, M., Valencia, C., Farber, J., and Nash, J. (2004). Large-scale comparative genomics meta-analysis of *campylobacter jejuni* isolates reveals low level of genome plasticity. *Journal of Clinical Microbiology* 42, 4566-4576.

Taboada, E., Clark, C., Sproston, E., and Carrillo, C. (2013). Current methods for molecular typing of *Campylobacter* species. *Journal of Microbiological Methods* 95, 24-31.

Taboada, E., Mackinnon, J., Luebbert, C., Gannon, V., Nash, J., and Rahn, K. (2008). Comparative genomic assessment of multi-locus sequence typing: rapid accumulation of genomic heterogeneity among clonal isolates of *Campylobacter jejuni*. *BMC Evolutionary Biology* 8, 229-241.

Taboada,E., S. Mutschall, B. Hetman, V. Boras, V. Suttorp, P. Hodgkinson and G. Inglis. 2015. *Molecular epidemiology of C. jejuni and C. coli circulating in a model ecosystem with intense agricultural activity and high rates of campylobacteriosis.* Presented at the

18th International Workshop on Campylobacter, Helicobacter and Related Organisms, New Zealand. 34.

Taboada, E., Ross, S., Mutschall, S., MacKinnon, J., Roberts, M., Buchanan, C., Kruczkiewicz, P., Jokinen, C., Thomas, J., Nash, J., Gannon, V., Marshall, B., Pollari, F., and Clark, C. (2012). Development and validation of a comparative genomic fingerprinting method for high-resolution genotyping of *Campylobacter jejuni*. *Journal of Clinical Microbiology* 50, 788-797.

Taboada, E., van Belkum, A., Yuki, N., Acedillo, R., Godschalk, P., Koga, M., Endtz, H., Gilbert, M., and Nash, J. (2007). Comparative genomic analysis of *Campylobacter jejuni* associated with Guillain-Barré and Miller Fisher Syndromes: neuropathogenic and enteritis-associated isolates can share high levels of genomic similarity. *BMC Genomics* 8, 1-11.

Taff, C., Weis, A., Wheeler, S., Hinton, M., Weimer, B., Barker, C., Jones, M., Logsdon, R., Smith, W., Boyce, W., and Townsend, A. (2016). Influence of host ecology and behavior on *Campylobacter jejuni* prevalence and environmental contamination risk in a synanthropic wild bird species. *Applied and Environmental Microbiology* 82, 4811-4820.

Takata, T., Fujimoto, S., and Amako, K. (1992). Isolation of nonchemotactic mutants of *Campylobacter jejuni* and their colonization of the mouse intestinal tract. *Infection and Immunity* 60, 3596-3600.

Thakur, S., Zhao, S., McDermott, P., Harbottle, H., Abbott, J., English, L., Gebreyes, W., and White, D. (2010). Antimicrobial resistance, virulence, and genotypic profile comparison of *Campylobacter jejuni* and *Campylobacter coli* isolated from humans and retail meats. *Foodborne Pathogens and Disease* 7, 1-10.

The Campylobacter Sentinel Surveillance Scheme Collaborators. (2003). Point source outbreaks of *Campylobacter jejuni* infection - are they more common than we think and what might cause them? *Epidemiology and Infection* 130, 367-375.

Thomas, C., Hill, D., and Mabey, M. (2002). Culturability, injury and morphological dynamics of thermophilic *Campylobacter* spp. within a laboratory-based aquatic model system. *Journal of Applied Microbiology* 92, 433-442.

Thomas, M., Murray, R., Flockhart, L., Pintar, K., Pollari, F., Fazil, A., Nesbitt, A., and Marshall, B. (2013). Estimates of the burden of foodborne illness in Canada for 30 specified pathogens and unspecified agents, circa 2006. *Foodborne Pathogens and Disease* 10, 639-648.

Tracz, D., Keelan, M., Ahmed-Bentley, J., Gibreel, A., Kowalewska-Grochowska, K., and Taylor, D. (2005). pVir and bloody diarrhea in *Campylobacter jejuni* enteritis. *Emerging Infectious Diseases* 11, 838-843.

Tsai, H., Huang, H., Lin, C., Lien, Y., and Chou, C. (2007). Salmonellae and campylobacters in household and stray dogs in northern Taiwan. *Veterinary Research Communications* 31, 931-939.

Van Dyke, M., Morton, V., McLellan, N., and Huck, P. (2010). The occurrence of *Campylobacter* in river water and waterfowl within a watershed in southern Ontario, Canada. *Journal of Applied Microbiology* 109, 1053-1066.

Varela, N., Friendship, R., and Dewey, C. (2007). Prevalence of *Campylobacter* spp isolated from grower-finisher pigs in Ontario. *The Canadian Veterinary Journal* 48, 515-517.

Veron, M., and Chatelain, R. (1973). Taxonomic study of the genus *Campylobacter* Sebald and Veron and designation of the neotype strain for the type species, *Campylobacter fetus* (Smith and Taylor) Sebald and Veron. *International Journal of Systematic Bacteriology* 23, 122-134.

Viswanathan, M., Pearl, D., Taboada, E., Parmley, E., Mutschall, S., and Jardine, C. (2017). Cluster analysis of *Campylobacter jejuni* genotypes isolated from small and medium-sized mammalian wildlife and bovine livestock from Ontario farms. *Zoonoses and Public Health* 64, 185-193.

Waldenstrom, J., Broman, T., Carlsson, I., Hasselquist, D., Achterberg, R., Wagenaar, J., and Olsen, B. (2002). Prevalence of *Campylobacter jejuni*, *Campylobacter lari*, and *Campylobacter coli* in different ecological guilds and taxa of migrating birds. *Applied and Environmental Microbiology* 68, 5911-5917.

Wallace, J., Stanley, K., Currie, J., Diggle, P., and Jones, K. (1997). Seasonality of thermophilic *Campylobacter* populations in chickens. *Journal of Applied Microbiology* 82, 219-224.

Wang, Y., and Taylor, D. (1990). Natural transformation in *Campylobacter* species. *Journal of Bacteriology* 172, 949-955.

Wassenaar, T. (1997). Toxin production by *Campylobacter* spp. *Clinical Microbiology Reviews* 10, 466-476.

Wassenaar, T., and Blaser, M. (1999). Pathophysiology of *Campylobacter jejuni* infections of humans. *Microbes and Infection* 1, 1023-1033.

Wassenaar, T., Bleumink-Pluym, N., and Zeijst, B. (1991). Inactivation of *Campylobacter jejuni* flagellin genes by homologous recombination demonstrates that flaA but not flaB is required for invasion. *The EMBO Journal* 10, 2055-2061.

Wassenaar, T., Fry, B., and van der Zeijst, B. (1995). Variation of the flagellin gene locus of *Campylobacter jejuni* by recombination and horizontal gene transfer. *Microbiology* 141, 95-101.

Wassenaar, T., Geilhausen, B., and Newell, D. (1998). Evidence of genomic instability in *Campylobacter jejuni* isolated from poultry. *Applied and Environmental Microbiology* 64, 1816-1821.

Wassenaar, T., and Newell, D. (2000). Genotyping of *Campylobacter* spp. *Applied and Environmental Microbiology* 66, 1-9.

Weijtens, M., van der Plas, J., Bijker, P., Urlings, H., Koster, D., van Logtestijn, J., and Huis in't Veld, J. (1997). The transmission of *Campylobacter* in piggeries; an epidemiological study. *Journal of Applied Microbiology* 83, 693-698.

Weis, A., Storey, D., Taff, C., Townsend, A., Huang, B., Kong, N., Clothier, K., Spinner, A., Byrne, B., and Weimer, B. (2016). Genomic comparison of *Campylobacter* spp. and their potential for zoonotic transmission between birds, primates, and livestock. *Applied and Environmental Microbiology* 82, 7165-7175.

Wesley, I., Wells, S., Harmon, K., and Green, A. (2000). Fecal shedding of *Campylobacter* and *Arcobacter* spp. in dairy cattle. *Applied and Environmental Microbiology* 66, 1994-2000.

Whiley, H., McLean, R., and Ross, K. (2016). Detection of *Campylobacter jejuni* in lizard faeces from central Australia using quantitative PCR. *Pathogens* 6, 1-6.

Whitehouse, C., Balbo, P., Pesci, E., Cottle, D., Mirabito, P., and Pickett, C. (1998). *Campylobacter jejuni* Cytolethal Distending Toxin causes a G2-phase cell cycle block. *Infection and Immunity* 66, 1934-1940.

WHO. (2013). *The global view of campylobacteriosis.*World Health Organization Retrieved from: http://www.who.int/foodsafety/publications/campylobacteriosis/en/.

Wieland, B., Regula, G., Danuser, J., Wittwer, W., Burnens, A., Wassenaar, T., and Stärk, K. (2005). *Campylobacter* spp. in dogs and cats in Switzerland: risk factor analysis and molecular characterization with AFLP. *Journal of Veterinary Medicine* 52, 183-189.

Willison, H., Jacobs, B., and van Doorn, P. (2016). Guillain-Barré Syndrome. *Lancet* 388, 717-727.

Wilson, D., Gabriel, E., Leatherbarrow, A., Cheesbrough, J., Gee, S., Bolton, E., Fox, A., Hart, C., Diggle, P., and Fearnhead, P. (2009). Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. *Molecular Biology and Evolution* 26, 385-397.

Wilson, I. (2002). *Salmonella* and *Campylobacter* contamination of raw retail chickens from different producers: a six year survey. *Epidemiology and Infection* 129, 635-645.

Wösten, M., Wagenaar, J., and van Putten, J. (2004). The FlgS/FlgR two-component signal transduction system regulates the fla regulon in *Campylobacter jejuni*. *Journal of Biological Chemistry* 279, 16214-16222.

Wu, Z., Mou, K., Sahin, O., and Zhang, Q. (2016). Genomic insights into *Campylobacter jejuni* virulence and population genetics. *Infectious Diseases and Translational Medicine* 2, 109-119.

Xu, W., Reuter, T., Inglis, G., Larney, F., Alexander, T., Guan, J., Stanford, K., Xu, Y., and McAllister, T. (2009). A biosecure composting system for disposal of cattle carcasses and manure following infectious disease outbreak. *Journal of Environmental Quality* 38, 437-450.

Yahara, K., Méric, G., Taylor, A., de Vries, S., Murray, S., Pascoe, B., Mageiros, L., Torralbo, A., Vidal, A., Ridley, A., Komukai, S., Wimalarathna, H., Cody, A., Colles, F., McCarthy, N., Harris, D., Bray, J., Jolley, K., Maiden, M., Bentley, S., Parkhill, J., Bayliss, C., Grant, A., Maskell, D., Didelot, X., Kelly, D., and Sheppard, S. (2017). Genome-wide association of functional traits linked with *Campylobacter jejuni* survival from farm to fork. *Environmental Microbiology* 19, 361-380.

Yao, R., Burr, D., Doig, P., Trust, T., Niu, H., and Guerry, P. (1994). Isolation of motile and non-motile insertional mutants of *Campylobacter jejuni*: the role of motility in adherence and invasion of eukaryotic cells. *Molecular Microbiology* 14, 883-893.

Yao, R., Burr, D., and Guerry, P. (1997). CheY-mediated modulation of *Campylobacter jejuni* virulence. *Molecular Microbiology* 23, 1021-1031.

Young, K., DiRita, V., and Davis, L. (2007). *Campylobacter jejuni*: molecular biology and pathogenesis. *Nature Reviews Microbiology* 5, 665-679.

Yuki, N. (1997). Molecular mimicry between gangliosides and lipopolysaccharides of *Campylobacter jejuni* isolated from patients with Guillain-Barré Syndrome and Miller Fisher Syndrome. *The Journal of Infectious Diseases* 176, S153.

Ziprin, R., Young, C., Stanker, L., Hume, M., and Konkel, M. (1999). The absence of cecal colonization of chicks by a mutant of *Campylobacter jejuni* not expressing bacterial fibronectin-binding protein. *Avian Diseases* 43, 586-589.

Chapter 2

2 A Genome-wide association study to identify diagnostic markers for human-pathogenic *Campylobacter jejuni* strains

2.1 Abstract

Campylobacter jejuni is a leading human enteric pathogen worldwide and despite an improved understanding of its biology, ecology, and epidemiology, limited tools exist for identifying strains that are likely to cause disease. In the current study, we used subtyping data in a database representing over 24,000 isolates collected through various surveillance projects in Canada to identify 166 representative genomes from prevalent *C. jejuni* subtypes for whole genome sequencing. The sequence data was used in a Genome-Wide Association Study (GWAS) aimed at identifying accessory gene markers associated with clinically-related *C. jejuni* subtypes. Prospective markers (n=28) were then validated against a large number (n=3,902) of clinically-associated and non-clinically-associated genomes from a variety of sources. A total of 25 genes, including six sets of genetically linked genes, were identified as robust putative diagnostic markers for clinically-related *C. jejuni* subtypes. Although some of the genes identified in this study have been previously shown to play a role in important processes such as iron acquisition and vitamin B_5 biosynthesis, others have unknown function or are unique to the current study and warrant further investigation. As few as four of these markers could be used in combination to detect up to 90% of clinically-associated isolates in the validation dataset, and such markers could form the basis for a screening assay to rapidly identify strains that pose an increased risk to public health. The results of the current study are consistent

with the notion that specific groups of *C. jejuni* strains of interest are defined by the presence of specific accessory genes.

2.2 Introduction

Although molecular epidemiological evidence suggests that not all *C. jejuni* strains or genetic lineages pose an equal risk to human health, our current understanding of *C. jejuni* subtype-dependent pathogenesis is incomplete (French et al., 2005; Sheppard et al., 2012). In contrast to other enteric pathogens, *C. jejuni* does not possess a number of the classical virulence factors (e.g. Type III or Type IV Secretion Systems, enterotoxins) found in other pathogens (Havelaar et al., 2009). Previous studies have identified genetic determinants that are important for *C. jejuni* pathogenicity (Dasti et al., 2009), but they are generally conserved across the species. Therefore, these factors have little predictive power for the identification of isolates with a higher propensity to cause disease in humans.

With the advent of inexpensive and high-throughput whole genome sequencing, Genome-Wide Association Studies (GWAS) are increasingly being applied to bacterial genomics as tools for the identification of genetic markers associated with a phenotype or trait of interest (Read & Massey, 2014). For example, GWAS have been utilised to identify mutations and other polymorphisms associated with antibiotic resistance in *Mycobacterium tuberculosis* (Farhat et al., 2013), *Staphylococcus aureus* (Alam et al., 2014) and *Streptococcus pneumoniae* (Chewapreecha et al., 2014). In *Campylobacter*, GWAS have been used to identify genetic factors related to the Guillain-Barré Syndrome (Taboada et al., 2007), host adaptation in *C. jejuni* and *Campylobacter coli* (Sheppard et

al., 2013), and more recently, used to identify markers associated with the survival of *C. jejuni* in the poultry production chain (Yahara et al., 2017).

In this study, we have used isolates from the Canadian *Campylobacter* Comparative Genomic Fingerprinting Database (C3GFdb) to perform a GWAS aimed at identifying genetic determinants preferentially found among *C. jejuni* lineages associated with human disease. The goal of the current study was to identify accessory genes with a statistically significant difference in carriage rates in two *C. jejuni* cohorts that differ in terms of their association with human campylobacteriosis. These genes could be used as diagnostic markers for molecular-based risk assessment and the rapid detection of *C. jejuni* isolates that pose the greatest risk to human health.

2.3 Materials and methods

2.3.1 Strain selection

A total of 166 *C. jejuni* isolates representing 34 of the 100 most prevalent CGF subtypes circulating in Canada were selected from the C3GFdb for whole genome sequencing (Supplementary Table 2.1). The selected isolates and their respective subtypes represented approximately 31% (7,407/24,142) of all isolates in the database and over 55% (7,407/13,367) of the isolates from the 100 most prevalent CGF subtypes (Figure 2.1). They have been observed in multiple provinces, sources and hosts, and over multiple years, suggesting that they are endemic and in wide circulation. The dataset selected for WGS was composed of 72 isolates from animals or retail meat, 54 isolates from environmental sources, and 40 isolates from human clinical cases (Table 2.1).

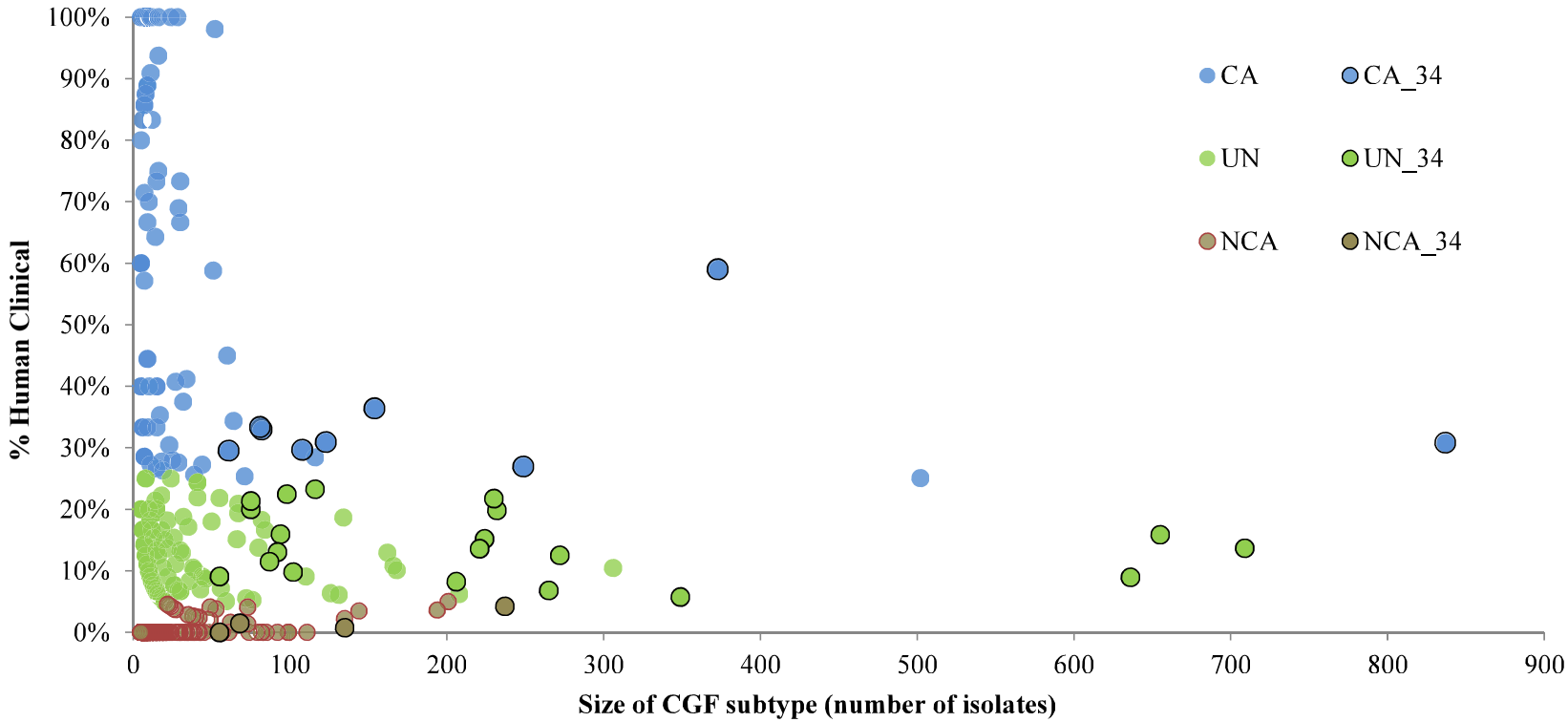

Figure 2.1: Identification of CGF subtypes for GWAS analysis of Clinically-Associated (CA) vs. Non-Clinically-Associated (NCA) *C. jejuni* subtypes. The C3GFdb was used to identify 166 *C. jejuni* isolates for whole genome sequencing from 34 highly prevalent CGF subtypes (black outline) that together account for nearly 31% of all isolates in the database, and over 55% of all isolates from the 100 most prevalent CGF subtypes circulating in Canada. These subtypes exhibit differences in their association with human campylobacteriosis, and sequence data from representative isolates was used in a genome-wide association study aimed at identifying accessory genes associated with clinically relevant *C. jejuni* subtypes.

Table 2.1: Epidemiological characteristics of 34 CGF subtypes targeted for whole-genome sequencing based on the Canadian *Campylobacter* Comparative Genomic Fingerprinting Database (C3GFdb).

				Proportion of isolates in subtype (%)[4]			
CGF Subtype	Cohort[1]	Cluster Size[2]	Cluster Rank[3]	H	A	E	U
0169.001.002	CA	837	1	30.8%	62.7%	6.3%	0.1%
0695.006.001	UN	709	2	13.7%	80.4%	5.9%	0.0%
0083.001.002	UN	655	3	15.9%	83.2%	0.8%	0.2%
0926.002.001	UN	636	4	9.0%	74.2%	16.8%	0.0%
0044.003.001	CA	373	6	59.0%	40.5%	0.5%	0.0%
0957.001.001	UN	349	7	5.7%	69.6%	24.6%	0.0%
0853.011.001	UN	272	9	12.5%	87.1%	0.4%	0.0%
0882.005.001	UN	265	10	6.8%	81.1%	9.4%	2.6%
0982.001.002	CA	249	11	26.9%	68.3%	4.8%	0.0%
0811.009.002	NCA	237	12	4.2%	43.9%	51.9%	0.0%
0735.005.001	UN	232	13	19.8%	66.8%	13.4%	0.0%
0253.004.001	UN	230	14	21.7%	75.2%	3.0%	0.0%
0960.007.001	UN	224	15	15.2%	76.8%	5.4%	2.7%
0731.001.005	UN	221	16	13.6%	81.9%	4.5%	0.0%
0923.002.001	UN	206	18	8.3%	61.7%	30.1%	0.0%
0269.004.001	CA	154	24	36.4%	63.6%	0.0%	0.0%
0811.008.001	NCA	135	26.5	0.7%	45.9%	53.3%	0.0%
0173.004.001	CA	123	31	30.9%	57.7%	11.4%	0.0%
0173.002.004	UN	116	32.5	23.3%	76.7%	0.0%	0.0%
0933.004.002	CA	108	36	29.6%	65.7%	4.6%	0.0%
0893.001.001	UN	102	37	9.8%	82.4%	7.8%	0.0%
0933.008.001	UN	98	40	22.4%	75.5%	2.0%	0.0%
0949.001.002	UN	94	41	16.0%	72.3%	11.7%	0.0%
0960.003.002	UN	92	42.5	13.0%	67.4%	19.6%	0.0%
0904.002.002	UN	87	44	11.5%	74.7%	12.6%	1.1%
0103.001.002	CA	82	48.5	32.9%	67.1%	0.0%	0.0%
0077.001.003	CA	81	51	33.3%	66.7%	0.0%	0.0%
0238.007.002	UN	75	55.5	20.0%	80.0%	0.0%	0.0%
0260.007.001	UN	75	55.5	21.3%	78.7%	0.0%	0.0%
0844.001.001	NCA	68	63	1.5%	23.5%	75.0%	0.0%
0253.001.002	CA	61	69.5	29.5%	68.9%	0.0%	1.6%
0535.001.003	UN	55	76.5	9.1%	36.4%	54.5%	0.0%
0817.003.001	NCA	55	76.5	0.0%	20.0%	80.0%	0.0%
0083.007.001	CA	51	81	58.8%	41.2%	0.0%	0.0%

[1] Cohorts: Clinically Associated (CA); Non-Clinically Associated (NCA); Undefined (UN). [2] Number of isolates observed with the CGF subtype in the C3GFdb. [3] Rank of CGF subtype (based cluster size) in the C3GFdb. [4] Proportion of isolates in the subtype from Human (H), Animal (A), Environmental (E), and Unknown (U) sources.

2.3.2 Genome sequencing, assembly and annotation

Sequencing was conducted at Canada's Michael Smith Genome Sciences Centre, Vancouver, Canada using the Illumina HiSeq 2000 platform. Whole genome sequence data for this study has been deposited in the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) under the BioProject PRJNA368735. Draft *de novo* genome assembly of paired-end reads was performed using SPAdes v.2.4.0 (Bankevich et al., 2012) with pre-assembly BayesHammer read correction, default k-mer size testing options, and post-assembly Burrows Wheeler Aligner mismatch correction. Contigs with low coverage or shorter than 500 bp were removed from all subsequent analyses. Genome assembly quality was assessed using QUAST v.2.1 (Gurevich et al., 2013). Prediction of Open Reading Frames (ORFs) and annotation was performed using the PROKKA pipeline v.1.5.2 (Seemann, 2014) using a custom database of non-redundant gene sequences representing five complete and well-annotated *C. jejuni* reference genomes available from NCBI (Supplementary Table 2.2).

2.3.3 Definition of a *C. jejuni* reference pan-genome for the dataset

Predicted ORFs were queried using a reciprocal best hit approach (Moreno-Hagelsieb & Latimer, 2008; Ward & Moreno-Hagelsieb, 2014) with BLAST v 2.2.29 (Camacho et al., 2009) in order to define a reference pan-genome, the non-redundant set of genes for a set of genome sequences (Méric et al., 2014). Paired BLAST queries were treated as *orthologous* if they shared ≥80% sequence identity and ≥50% alignment coverage and a single exemplar was included in the pan-genome. The pan-genome defined using this process was used in the subsequent GWAS.

2.3.4 Genome-wide association study

Carriage across the dataset of all genes representing the pan-genome was assessed by BLAST analysis. The nucleotide sequence of each gene was queried against the 166 draft genome assemblies using Blastn. Genes were considered to be *present* if a hit representing ≥80% sequence identity over ≥50% of the length of the query gene was found and considered *absent* otherwise. In order to facilitate statistical comparison, subtypes were defined as either Non-Clinically Associated (NCA; ≤5% human clinical isolates), Undefined (UN; 5-25% human clinical isolates), or Clinically Associated (CA; ≥25% human clinical isola). The statistical significance of each gene ($p < 0.05$) was defined based on its carriage rate in the CA and NCA cohorts and was computed using Fisher's Exact test statistic in GenomeFisher (https://bitbucket.org/peterk87/genomefisher/wiki/Home); p-values were adjusted for multiple testing using the method of Holm (Aickin & Gensler, 1996; Gaetano, 2013; Holm, 1979). Statistically significant genes were subjected to further analysis and validation as outlined below.

2.3.5 *In silico* validation of putative diagnostic marker genes.

In order to select markers with the highest potential for downstream assay development, candidate genes identified by the GWAS analysis were filtered in a stepwise process according to the following conditions: 1) complete absence in the NCA cohort and presence in ≥50% of CA genomes; 2) high sequence identity (>99%) and complete, or near complete, conservation of sequence length (>90%) in the corresponding orthologous gene, when present, among a set of reference genomes (Table 2.2); and 3) statistical significance ($p < 0.05$) when the NCA cohort was compared to a combined

CA+UN cohort, in which the UN (i.e. undefined) genomes were treated as CA and pooled with the CA genomes. Genes that passed all criteria were selected for an *in silico* validation using a larger set of genome sequences. This validation dataset was created by combining genomes sequenced in house as part of current or previous studies (n=325) and additional genomes acquired from public repositories (n=3,955). Publicly available genomes were restricted to those with available epidemiological data (e.g. sample source, country of origin, etc.). To facilitate assignment into NCA, UN, and CA cohorts, *in silico* CGF was performed on these genomes using MIST (Kruczkiewicz et al., 2013), with a concordance between CGF profiles predicted *in silico* and those generated in the laboratory estimated to be 96.8% on a subset of 325 isolates (12,583/13,000 matching loci; data not shown); only genomes from CGF subtypes previously observed in the C3GFdb were retained in the validation set (n=3,902). Each genome was designated to its respective cohort based on the corresponding epidemiological data of the *in silico* CGF subtype. Finally, the putative diagnostic genes identified by the GWAS using the original set of 166 genomes were tested for statistically significance with the expanded cohorts. The combinatorial effect of different subsets of markers was also assessed to determine if a reduced number of markers could be applied to detect clinically-related *C. jejuni* subtypes without a subsequent loss of sensitivity.

Table 2.2: Significant genes observed after GWAS analysis of genome sequences from representative Clinically-Associated (CA) and Non-Clinically-Associated (NCA) *C. jejuni* subtypes.

	p-value[1]					
Marker	Raw	Holm-corrected[2]	11168 Ortholog	Gene Name	Function	Linkage Group
11168_00051	4.29E-10	8.39E-07	*Cj0055c*		hypothetical protein	LG1
11168_00052	5.28E-10	1.03E-06	*Cj0056c*		hypothetical protein	
11168_00169	3.36E-11	6.61E-08	*Cj0177*		putative iron transport protein	LG2
11168_00170	3.36E-11	6.61E-08	*Cj0178*		putative TonB-denpendent outer membrane receptor	
11168_00171	3.36E-11	6.60E-08	*Cj0179*	*exbB1*	biopolymer transport protein	
11168_00172	3.36E-11	6.60E-08	*Cj0180*	*exbD1*	biopolymer transport protein	
11168_00173	3.36E-11	6.60E-08	*Cj0181*	*tonB1*	TonB transport protein	
11168_00230	6.12E-19	1.21E-15	*Cj0246c*		putative MCP-domain signal transduction protein	
11168_00243	6.48E-34	1.28E-30	*Cj0259*	*pyrC*	putative dihydroorotase	LG3
11168_00244	3.10E-27	6.14E-24	*Cj0260c*		small hydrophobic protein	
11168_00248	6.57E-25	1.30E-21	*Cj0264c*		putative molybdopterin containing oxidoreductase	LG4
11168_00249	6.57E-25	1.30E-21	*Cj0265c*		putative cytochrome C-type haem-binding periplasmic protein	
11168_00277	1.30E-17	2.57E-14	*Cj0295*		putative acetyltransferease	LG5
11168_00278	1.35E-18	2.66E-15	*Cj0296c*	*panD*	aspartate 1-decarboxylase precursor	
11168_00279	1.35E-18	2.66E-15	*Cj0297c*	*panC*	pantoate-beta-alanine ligase	
11168_00280	1.35E-18	2.66E-15	*Cj0298c*	*panB*	3-methyl-2-oxobutanoate hydroxymethyltransferase	
11168_00281	1.09E-16	2.15E-13	*Cj0299*		putative periplasmic beta-lactamase	
11168_00703	6.98E-24	1.38E-20	*Cj0731*		putative ABC transport system permease	
11168_00718	3.36E-11	6.59E-08	*Cj0753c*	*tonB3*	TonB transport protein	LG6
11168_00719	3.36E-11	6.59E-08	*Cj0755*	*cfrA*	ferric enterobactin uptake receptor	
11168_01072	4.90E-11	9.59E-08	*Cj1122c*		putative integral membrane protein.	
11168_01201	6.12E-19	1.21E-15	*Cj1255*		putative isomerase	
11168_01309	5.30E-15	1.04E-11	*Cj1365c*		putative secreted serine protease	
11168_01519	4.29E-10	8.39E-07	*Cj1585c*		putative oxidoreductase	
11168_01610	4.29E-10	8.38E-07	*Cj1679*		hypothetical protein	
06_2866_00597	6.89E-28	1.36E-24			di-/tripeptide transporter	
06_7515_00723	4.19E-16	8.24E-13			prophage Lp2 protein 6	
07_0675_00227	2.62E-11	5.15E-08		*tetO*	elongation factor G	

[1] p-value based on 2-tailed Fisher's Exact Test. [2] p-values were adjusted using the Holm-correction (Holm, 1979)

2.4 Results and discussion

2.4.1 Genome sequencing, assembly and annotation.

The quality of the *de novo* assembly of the 166 genomes selected as representatives of 34 highly prevalent CGF subtypes in Canada was assessed using QUAST (Gurevich et al., 2013). The average number of reads produced for each genome was 4,161,271 (±1,223,304), for an average coverage of 253x (±74.7x). Individual genome assemblies had an average of 67 (±27) contigs and an N75 of 34,631 bp (± 13,815 bp). All genome assemblies had additional parameters in range with what has typically been observed for *C. jejuni.* The average assembly length (1,660,986 ±51,283.5 bp), predicted ORFs (1,719±71), and %G+C (30.4±0.13%) were typical of *C. jejuni* genome assemblies available in the public domain. Annotation of the 166 draft genomes from this study using the PROKKA pipeline (Seemann, 2014) resulted in the identification of 291,502 ORFs. The genome of strain NCTC 11168, which has been completely sequenced (Parkhill et al., 2000), was included in the analysis as a control to assess the completeness of the ORF prediction and annotation process. The original annotation of NCTC 11168 predicted 1,654 ORFs, while a subsequent re-annotation predicted 1,643 ORFs (Gundogdu et al., 2007); in our analysis, the PROKKA pipeline predicted 1,659 ORFs. This small discrepancy is related to the advanced curation used in the re-annotation of NCTC 11168, which resulted in the merging and removal of coding sequences belonging to pseudogenes and phase variable genes. The pan-genome established using this dataset consisted of 3,358 unique ORFs, of which 1,377 were present in all genomes (i.e. core genes) and 1,981 were present in a varying number of genomes (i.e. accessory genes).

2.4.2 Genome-wide association study

Of the 166 C. *jejuni* isolates selected for this study, 35 were assigned to the NCA cohort and represented four different CGF subtypes, 80 were assigned to the UN cohort and represented 20 CGF subtypes, and 51 were assigned to the CA cohort and represented ten CGF subtypes (Table 2.1). A GWAS was performed in order to identify accessory genes with a biased distribution in CA and NCA cohorts. Although in principle GWAS can be used to identify genetic variation ranging from SNPs to indels involving multiple genes, we chose to focus on accessory genes, as they have excellent potential for the development of rapid, robust, and inexpensive PCR-based diagnostic assays for screening of large numbers of strains. At the same time, it is important to note that other forms of genetic variation may represent valuable targets for tracking strains of interest. Recently, Clark et al. showed that large-scale chromosomal inversion could be used to distinguish a subset of outbreak-associated isolates from epidemiologically unrelated co-circulating isolates (Clark et al., 2016).

In total, 595 genes showed statistically significant differences in carriage between NCA and CA cohorts ($p < 0.05$) (Figure 2.2). Of these, 71 genes were completely absent from the NCA cohort but were present in at least ≥50% of isolates in the CA cohort (Condition 1), and 63 of these genes also maintained high sequence identity (>99%) and near complete sequence coverage (>90%) compared to their respective reference genes (Condition 2). Of these, 28 continued to exhibit robust statistical significance when the NCA cohort was compared to a pooled cohort comprising all UN and CA genomes (Condition 3). These included six sets of genes that appear to be found in linkage groups (LGs) (Table 2.2), with members of each linkage group possessing similar rates of carriage in the dataset. Since linked genes, which are located adjacently on the

chromosome, tend to be functionally related and are typically transmitted as a functional unit (Muley & Acharya, 2013), it is likely that their identification in this study was not due to spurious statistical signal.

Among the linkage groups observed in the GWAS were two sets of genes responsible for encoding iron acquisition systems. We observed that genes encoding both the *TonB1*-mediated *Cj0178* (LG2; *Cj0177 – Cj0181*) and the *TonB3*-mediated *CfrA* (LG6; *Cj0753c*/*Cj0755*) iron acquisition systems were significantly associated with *C. jejuni* isolates from clinically-related CGF subtypes. As is the case in most pathogens, iron acquisition is considered to be a virulence determinant in *C. jejuni* and has been linked to successful colonisation *in vivo* (Kim et al., 2003; Naikare et al., 2006; Palyada et al., 2004). *CfrA* has been shown to be capable of transporting a wide variety of structurally different siderophores, which may contribute to the ability of isolates with these genes to colonise a wide variety of hosts/niches (Naikare et al., 2013).

Another linkage group associated with CA and UN subtypes was composed of genes that encode the pantothenate (vitamin B_5) biosynthesis pathway and β-lactam antibiotic resistance. LG5 encompasses a total of five genes, including a putative acetyltransferase (*Cj0295*), the *panBCD* operon (*Cj0296c – Cj0298c*), which encodes for the pantothenate (vitamin B_5) biosynthesis pathway, as well as the gene bla_{OXA-61} (*Cj0299*), w -lactam antibiotics.

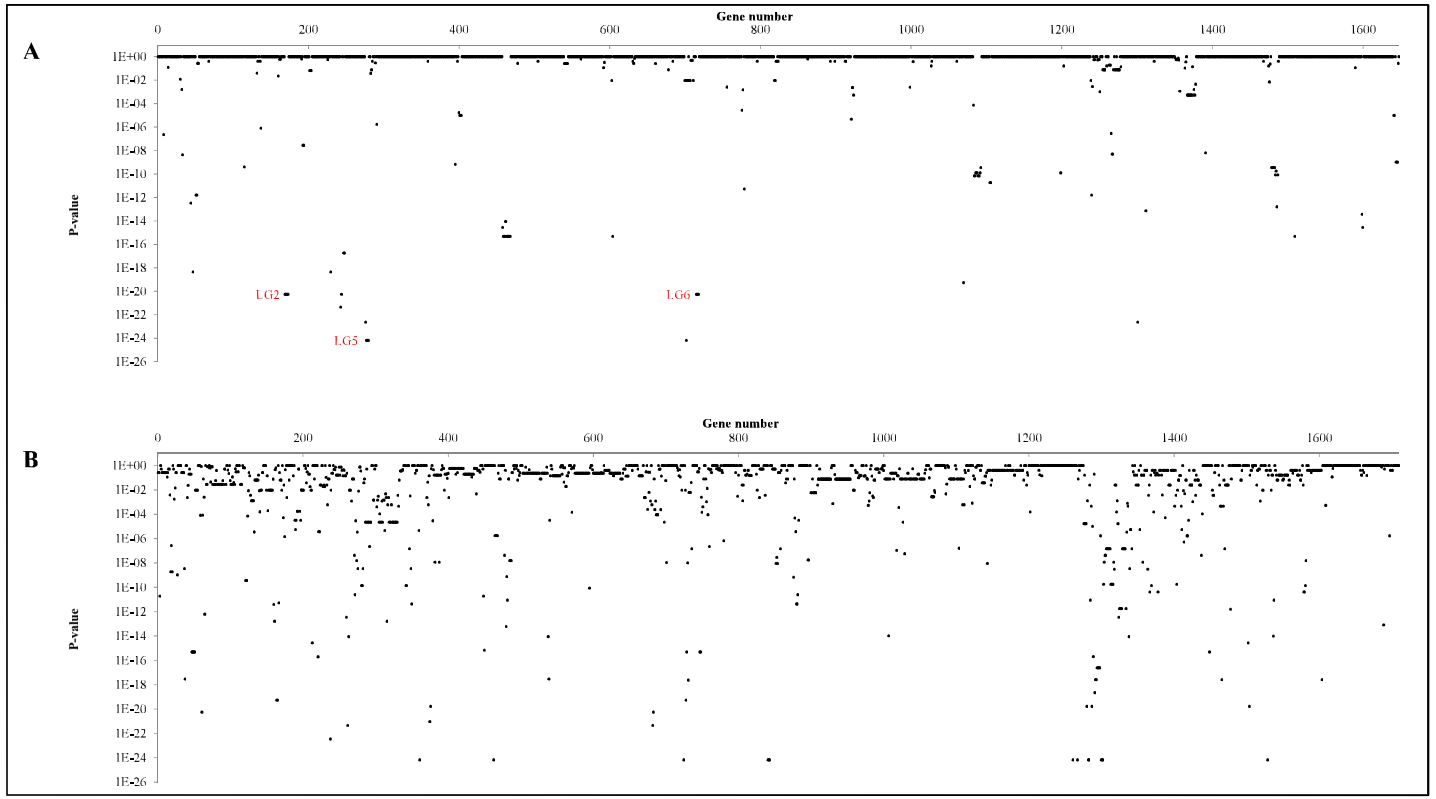

Figure 2.2: GWAS-based identification of genes with significant differences in carriage between Non-Clinically-Associated (NCA) and Clinically-Associated (CA) cohorts. The distribution of p-values observed for 3,358 genes in the *C. jejuni* pan-genome computed for this study after Genome Fisher analysis of gene carriage data for CA and NCA cohorts. (A) Genes from NCTC 11168 genome strains (n=1,648). (B) Genes from all other genomes (n=1,710).

These genes were recently implicated in host adaptation in *C. jejuni* and *C. coli,* where they were found to be more strongly associated with cattle-specific lineages relative to chicken-specific lineages, possibly as a result of selective pressures created by contemporary and geographically-dependent agricultural practices (Sheppard et al., 2013). Although it is generally recognised that chickens are a primary source of human exposure leading to infection, we observed strong statistical signal among CA subtypes for genes previously identified as cattle-associated (Sheppard et al., 2013). Sheppard *et al.* suggested that maintenance of these genes in chickens, albeit at a reduced rate, may facilitate rapid host switching as part of a host-generalist strategy. Moreover, we have observed that a majority of the most prevalent clinically-related CGF subtypes, many of which are represented in our GWAS dataset, are associated with both cattle and chickens. This is consistent with the possible role of cattle as an important reservoir for strains that go on to contaminate the chicken production system, ultimately leading to human cases of campylobacteriosis. As this manuscript was being readied for publication, GWAS was used to identify several loci that could be used as "host-segregating" epidemiological markers for source attribution (Thépault et al., 2017). Interestingly, one of the loci (*Cj0260c*) was also identified in our analysis. Thus, while our data suggests that presence of this gene is strongly associated with human clinical isolates, data from the study by Thépault et al. further suggests the allelic information appears to highly predictive of host source.

2.4.3 *In silico* validation of putative diagnostic marker genes

Population structure has been identified as a potential confounding factor in GWAS analyses, in that statistically significant associations may ultimately be due to

oversampling of certain subpopulations rather than with the phenotypic trait under investigation (Read & Massey, 2014). Since the focus of the current study was the examination of prevalent *C. jejuni* subtypes in Canada in the context of population structure, it was necessary to exclude the possibility that the markers we identified represent a biased distribution resulting from oversampling within certain lineages in the population. The large-scale marker validation that we performed using available WGS data included a dataset composed of genomes largely from the United Kingdom (3,871/4,280; 90%) and Canada (327/4,280; 8%), and an overwhelming majority of isolates were recovered from human clinical sources (3,559/4,280; 83%), while those from animal (626/4,280; 15%) and environmental (95/4,280; 2%) sources made up the remainder. A total of 539 CGF subtypes were identified by *in silico* CGF, however, 279 subtypes were novel and had not been previously observed in the C3GFdb and were omitted from the analysis since their epidemiological characteristics could not be determined. Of the remaining 260 CGF subtypes, 38 CGF subtypes (160 genomes) were identified as NCA, nine CGF subtypes (99 genomes) were identified as UN, and 213 CGF subtypes (3,742 isolates) were identified as CA. Despite the influx of genetically and geographically diverse isolates introduced as part of the expanded dataset, a majority (n=25) of the markers in the original GWAS analysis continued to show statistical significance with CA subtypes; on average these markers were present in 73% of CA isolates compared to only 36% of NCA isolates (Figure 2.3). Moreover, results of our combinatorial marker analysis show that as few as four markers could be used in combination to detect up to 90% of CA isolates in the validation dataset, with a modest carriage rate of 21% among NCA isolates. These findings suggest that the robust signal detected in the original GWAS analysis stems from genes that appear to have diagnostic

value for the identification of *C. jejuni* subtypes with an increased association with campylobacteriosis.

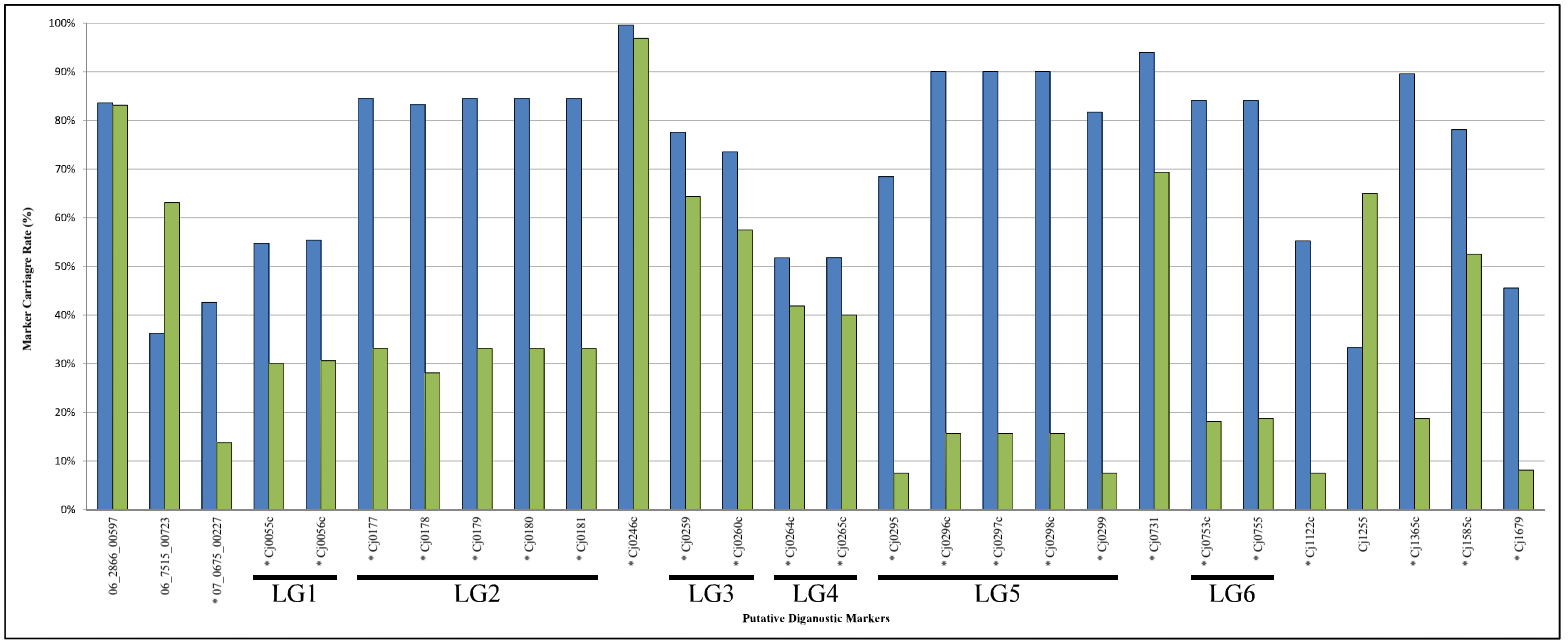

Figure 2.3: *In silico* validation of putative diagnostic marker genes against expanded CA and NCA cohorts. The putative diagnostic genes identified by the GWAS using the original set of 166 genomes were tested for statistical significance with expanded CA (blue bars; n=3,742) and NCA (green bars; n=160) cohorts comprising additional genomes sequenced in house and from public repositories. Despite the influx of genetically and geographically diverse isolates introduced as part of the expanded dataset, a majority (n=25) of the markers continued to show statistically significant signal with CA subtypes. This suggests that the robust signal detected in the original GWAS analysis stems from genes that appear to have diagnostic value for the identification of *C. jejuni* subtypes with an increased association with campylobacteriosis. * denotes genes that showed statistically significant signal with CA subtypes.

2.5 Conclusion

A major challenge in the prevention and control of campylobacteriosis is our current inability to identify strains of *C. jejuni* that pose greatest risk to human health. Addressing this issue would pave the way to better tracking of high-risk strains, leading to a better understanding of their distribution in the food chain and providing critical information towards the development of targeted mitigation strategies to reduce human exposure.

The goal of this study was to identify markers associated with *C. jejuni* lineages known to cause disease in humans and that have a high prevalence in Canada. The genomes of 166 isolates representing 34 highly prevalent *C. jejuni* subtypes were sequenced and a GWAS was performed to identify 28 genes significantly associated with highly-prevalent and clinically-related *C. jejuni* subtypes. While some putative gene markers identified as part of this study have previously been associated with important aspects of *C. jejuni* biology including iron acquisition and vitamin B_5 biosynthesis, others represent putative proteins associated with catalysis and transport, which may play roles in processes important for infection and warrant further investigation.

Although these genes were identified within a dataset of Canadian origin, 25 of them continued to display strong statistical significance when validated against a more genetically and geographically diverse dataset. This suggests that they may represent robust markers for clinically associated *C. jejuni* subtypes, paving the way for future development of molecular assays for rapid identification of *C. jejuni* strains that pose an increased risk to human health.

2.6 References

Aickin, M., and Gensler, H. (1996). Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm Methods. *American Journal of Public Health* 86, 726-728.

Alam, M., Petit, R., Crispell, E., Thornton, T., Conneely, K., Jiang, Y., Satola, S., and Read, T. (2014). Dissecting vancomycin-intermediate resistance in *Staphylococcus aureus* using genome-wide association. *Genome Biology and Evolution* 6, 1174-1185.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A., Dvorkin, M., Kulikov, A., Lesin, V., Nikolenko, S., Pham, S., Prjibelski, A., Pyshkin, A., Sirotkin, A., Vyahhi, N., Tesler, G., Alekseyev, M., and Pevzner, P. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19, 455-477.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 1-9.

Chewapreecha, C., Marttinen, P., Croucher, N., Salter, S., Harris, S., Mather, A., Hanage, W., Goldblatt, D., Nosten, F., Turner, C., Turner, P., Bentley, S., and Parkhill, J. (2014). Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes. *PLOS Genetics* 10, 1-11.

Clark, C., Berry, C., Walker, M., Petkau, A., Barker, D., Guan, C., Reimer, A., and Taboada, E. (2016). Genomic insights from whole genome sequencing of four clonal outbreak *Campylobacter jejuni* assessed within the global *C. jejuni* population. *BMC Genomics* 17, 1-16.

Dasti, J., Tareen, A., Lugert, R., Zautner, A., and Gross, U. (2009). *Campylobacter jejuni* : a brief overview on pathogenicity-associated factors and disease-mediating mechanisms. *International Journal of Medical Microbiology* 300, 205-211.

Farhat, M., Shapiro, B., Kieser, K., Sultana, R., Jacobson, K., Victor, T., Warren, R., Streicher, E., Calver, A., Sloutsky, A., Kaur, D., Posey, J., Plikaytis, B., Oggioni, M., Gardy, J., Johnston, J., Rodrigues, M., Tang, P., Kato-Maeda, M., Borowsky, M., Muddukrishna, B., Kreiswirth, B., Kurepina, N., Galagan, J., Gagneux, S., Birren, B., Rubin, E., Lander, E., Sabeti, P., and Murray, M. (2013). Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nature Genetics* 45, 1183-1189.

French, N., Barrigas, M., Brown, P., Ribiero, P., Williams, N., Leatherbarrow, H., Birtles, R., Bolton, E., Fearnhead, P., and Fox, A. (2005). Spatial epidemiology and natural population structure of *Campylobacter jejuni* colonizing a farmland ecosystem. *Environmental Microbiology* 7, 1116-1126.

Gaetano, J. (2013). *Holm-Bonferroni sequential correction: an EXCEL calculator.* v. 1.2. Retrieved from: https://www.researchgate.net/publication/242331583_Holm-Bonferroni_Sequential_Correction_An_EXCEL_Calculator_-_Ver_12.

Gundogdu, O., Bentley, S. D., Holden, M. T., Parkhill, J., Dorrell, N., and Wren, B. W. (2007). Re-annotation and re-analysis of the *Campylobacter jejuni* NCTC11168 genome sequence. *BMC Genomics* 8, 162.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072-1075.

Havelaar, A., van Pelt, W., Ang, C., Wagenaar, J., van Putten, J., Gross, U., and Newell, D. (2009). Immunity to *Campylobacter*: its role in risk assessment and epidemiology. *Critical Reviews in Microbiology* 35, 1-22.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70.

Kim, E., Sabra, W., and Zeng, A. (2003). Iron deficiency leads to inhibition of oxygen transfer and enhanced formation of virulence factors in cultures of *Pseudomonas aeruginosa* PAO1. *Microbiology* 149, 2627-2634.

Kruczkiewicz,P., S. Mutschall, D. Barker, J. Thomas, G. Van Domselaar, V. Gannon, C. Carrillo and E. Taboada. 2013. *MIST: a tool for rapid in silico generation of molecular data from bacterial genome sequences.* Presented at the *Biostec 2013,* 316-323.

Méric, G., Yahara, K., Mageiros, L., Pascoe, B., Maiden, M., Jolley, K., and Sheppard, S. (2014). A reference pan-genome approach to comparative bacterial genomics: identification of novel epidemiological markers in pathogenic *Campylobacter*. *PLOS One* 9, 1-12.

Moreno-Hagelsieb, G., and Latimer, K. (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* 24, 319-324.

Muley, V., and Acharya, V. (2013). Chromosomal proximity of genes as an indicator of functional linkage. *Springer Briefs in Systems Biology* , 1-60.

Naikare, H., Butcher, J., Flint, A., Xu, J., Raymond, K., and Stintzi, A. (2013). *Campylobacter jejuni* ferric-enterobactin receptor CfrA is TonB3 dependent and mediates iron acquisition from structurally different catechol siderophores. *Metallomics* 5, 988-996.

Naikare, H., Palyada, K., Panciera, R., Marlow, D., and Stintzi, A. (2006). Major role for FeoB in *Campylobacter jejuni* ferrous iron acquisition, gut colonization, and intracellular survival. *Infection and Immunity* 74, 5433-5444.

Palyada, K., Threadgill, D., and Stintzi, A. (2004). Iron acquisition and regulation in *Campylobacter jejuni*. *Journal of Bacteriology* 186, 4714-4729.

Parkhill, J., Wren, B., Mungall, K., Ketley, J., Churcher, C., Basham, D., Chillingworth, T., Davies, R., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A., Moule, S., Pallen, M., Penn, C., Quail, M., Rajandream, M., Rutherford, K., van Vilet, A., Whitehead, S., and Barrell, B. (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* 403, 665-668.

Read, T., and Massey, R. (2014). Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology. *Genome Medicine* 6, 1-11.

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

Sheppard, S., Didelot, X., Meric, G., Torralbo, A., Jolley, K., Kelly, D., Bentley, S., Maiden, M., Parkhill, J., and Falush, D. (2013). Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. *Proceedings of the National Academy of Sciences of the United States* 110, 11923-11927.

Sheppard, S., Jolley, K. A., and Maiden, M. C. J. (2012). A gene-by-gene approach to bacterial population genomics: whole genome MLST of *Campylobacter*. *Genes* 3, 261-277.

Taboada, E., van Belkum, A., Yuki, N., Acedillo, R., Godschalk, P., Koga, M., Endtz, H., Gilbert, M., and Nash, J. (2007). Comparative genomic analysis of *Campylobacter jejuni* associated with Guillain-Barré and Miller Fisher syndromes: neuropathogenic and enteritis-associated isolates can share high levels of genomic similarity. *BMC Genomics* 8, 1-11.

Thépault, A., Méric, G., Rivoal, K., Pascoe, B., Mageiros, L., Touzain, F., Rose, V., Béven, V., Chemaly, M., and Sheppard, S. (2017). Genome-wide identification of host-segregating epidemiological markers for source attribution in *Campylobacter jejuni*. *Applied and Environmental Microbiology* 83, 1-13.

Ward, N., and Moreno-Hagelsieb, G. (2014). Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do we miss? *PLOS One* 9, 1-6.

Yahara, K., Méric, G., Taylor, A., de Vries, S., Murray, S., Pascoe, B., Mageiros, L., Torralbo, A., Vidal, A., Ridley, A., Komukai, S., Wimalarathna, H., Cody, A., Colles, F., McCarthy, N., Harris, D., Bray, J., Jolley, K., Maiden, M., Bentley, S., Parkhill, J., Bayliss, C., Grant, A., Maskell, D., Didelot, X., Kelly, D., and Sheppard, S. (2017). Genome-wide association of functional traits linked with *Campylobacter jejuni* survival from farm to fork. *Environmental Microbiology* 19, 361-380.

2.7 Supplementary information

Table S2.1: Characteristics and NCBI accession numbers for strains characterised by whole genome sequencing.

Strain	CGF Subtype	Clinical Relevance	Host	Isolation Year	Isolation Location	Accession Number
C3GFDB009514	0083.007.001	CA	Human	2007	Ontario	SAMN06270746
C3GFDB009574	0083.007.001	CA	Chicken	2007	Ontario	SAMN06270747
C3GFDB008268	0103.001.002	CA	Cow	2010	Ontario	SAMN06270748
C3GFDB016493	0253.001.002	CA	Human	2005	Alberta	SAMN06270749
C3GFDB016524	0077.001.003	CA	Human	2005	Alberta	SAMN06270750
C3GFDB016539	0077.001.003	CA	Human	2005	Alberta	SAMN06270751
C3GFDB016546	0253.001.002	CA	Human	2005	Alberta	SAMN06270752
C3GFDB016564	0933.004.002	CA	Human	2005	Alberta	SAMN06270753
C3GFDB016571	0044.003.001	CA	Human	2005	Alberta	SAMN06270754
C3GFDB016572	0169.001.002	CA	Human	2005	Alberta	SAMN06270755
C3GFDB016592	0169.001.002	CA	Human	2005	Alberta	SAMN06270756
C3GFDB016601	0169.001.002	CA	Human	2005	Alberta	SAMN06270757
C3GFDB016634	0169.001.002	CA	Human	2005	Alberta	SAMN06270758
C3GFDB017536	0269.004.001	CA	Human	2004	Alberta	SAMN06270759
C3GFDB017537	0269.004.001	CA	Human	2004	Alberta	SAMN06270760
C3GFDB017540	0044.003.001	CA	Human	2004	Alberta	SAMN06270761
C3GFDB017542	0173.004.001	CA	Human	2004	Alberta	SAMN06270762
C3GFDB017610	0269.004.001	CA	Human	2005	Alberta	SAMN06270763
C3GFDB017620	0044.003.001	CA	Human	2005	Alberta	SAMN06270764
C3GFDB017638	0169.001.002	CA	Human	2005	Alberta	SAMN06270765
C3GFDB017646	0933.004.002	CA	Human	2005	Alberta	SAMN06270766
C3GFDB014399	0173.004.001	CA	Cow	2004	Alberta	SAMN06270767
C3GFDB014501	0044.003.001	CA	Cow	2004	Alberta	SAMN06270768
C3GFDB014638	0269.004.001	CA	Cow	2004	Alberta	SAMN06270769
C3GFDB014681	0269.004.001	CA	Cow	2005	Alberta	SAMN06270770
C3GFDB014710	0044.003.001	CA	Cow	2005	Alberta	SAMN06270771
C3GFDB014716	0169.001.002	CA	Cow	2005	Alberta	SAMN06270772
C3GFDB014747	0933.004.002	CA	Sheep	2005	Alberta	SAMN06270773
C3GFDB014751	0044.003.001	CA	Cow	2005	Alberta	SAMN06270774
C3GFDB014892	0169.001.002	CA	Chicken	2006	Alberta	SAMN06270775
C3GFDB014897	0169.001.002	CA	Chicken	2006	Alberta	SAMN06270776
C3GFDB014986	0269.004.001	CA	Cow	2011	Alberta	SAMN06270777
C3GFDB015015	0269.004.001	CA	Cow	2011	Alberta	SAMN06270778
C3GFDB015022	0253.001.002	CA	Cow	2011	Alberta	SAMN06270779
C3GFDB015057	0253.001.002	CA	Cow	2011	Alberta	SAMN06270780
C3GFDB015090	0253.001.002	CA	Cow	2011	Alberta	SAMN06270781
C3GFDB015145	0269.004.001	CA	Cow	2011	Alberta	SAMN06270782
C3GFDB015154	0269.004.001	CA	Cow	2011	Alberta	SAMN06270783
C3GFDB015165	0269.004.001	CA	Cow	2011	Alberta	SAMN06270784
C3GFDB015192	0269.004.001	CA	Cow	2011	Alberta	SAMN06270785
C3GFDB015200	0269.004.001	CA	Cow	2011	Alberta	SAMN06270786
C3GFDB015228	0169.001.002	CA	Cow	2012	Alberta	SAMN06270787
C3GFDB015257	0169.001.002	CA	Cow	2012	Alberta	SAMN06270788
C3GFDB015303	0169.001.002	CA	Cow	2012	Alberta	SAMN06270789
C3GFDB015390	0173.004.001	CA	Cow	2012	Alberta	SAMN06270790
C3GFDB016536	0982.001.002	CA	Human	2005	Alberta	SAMN06270791
C3GFDB016537	0982.001.002	CA	Human	2005	Alberta	SAMN06270792
C3GFDB014656	0982.001.002	CA	Cow	2005	Alberta	SAMN06270793
C3GFDB014661	0982.001.002	CA	Cow	2005	Alberta	SAMN06270794
C3GFDB015708	0982.001.002	CA	Water	2005	Alberta	SAMN06270795
C3GFDB016010	0982.001.002	CA	Water	2010	Alberta	SAMN06270796
C3GFDB014551	0817.003.001	NCA	Goose	2004	Alberta	SAMN06270797
C3GFDB014554	0817.003.001	NCA	Goose	2004	Alberta	SAMN06270798
C3GFDB014559	0817.003.001	NCA	Goose	2004	Alberta	SAMN06270799
C3GFDB015571	0811.009.002	NCA	Water	2004	Alberta	SAMN06270800

Strain	CGF Subtype	Clinical Relevance	Host	Isolation Year	Isolation Location	Accession Number
C3GFDB015649	0817.003.001	NCA	Water	2004	Alberta	SAMN06270801
C3GFDB015666	0811.009.002	NCA	Water	2004	Alberta	SAMN06270802
C3GFDB015668	0817.003.001	NCA	Water	2004	Alberta	SAMN06270803
C3GFDB014630	0811.009.002	NCA	Goose	2004	Alberta	SAMN06270804
C3GFDB016172	0811.009.002	NCA	Water	2006	British Columbia	SAMN06270805
C3GFDB015754	0811.009.002	NCA	Water	2006	Alberta	SAMN06270806
C3GFDB016266	0811.009.002	NCA	Water	2006	British Columbia	SAMN06270807
C3GFDB014923	0811.008.001	NCA	Goose	2006	Alberta	SAMN06270808
C3GFDB014927	0811.008.001	NCA	Duck	2006	Alberta	SAMN06270809
C3GFDB015823	0811.008.001	NCA	Water	2006	Alberta	SAMN06270810
C3GFDB019796	0844.001.001	NCA	Water	2006	Ontario	SAMN06270811
C3GFDB019832	0811.009.002	NCA	Water	2006	Ontario	SAMN06270812
C3GFDB015843	0817.003.001	NCA	Water	2006	Alberta	SAMN06270813
C3GFDB015846	0811.009.002	NCA	Water	2006	Alberta	SAMN06270814
C3GFDB016311	0844.001.001	NCA	Water	2006	British Columbia	SAMN06270815
C3GFDB019847	0811.009.002	NCA	Water	2006	Ontario	SAMN06270816
C3GFDB015878	0817.003.001	NCA	Water	2006	Alberta	SAMN06270817
C3GFDB016325	0811.009.002	NCA	Water	2006	British Columbia	SAMN06270818
C3GFDB015958	0811.008.001	NCA	Water	2007	Alberta	SAMN06270819
C3GFDB016401	0844.001.001	NCA	Water	2007	British Columbia	SAMN06270820
C3GFDB016403	0844.001.001	NCA	Water	2007	British Columbia	SAMN06270821
C3GFDB015968	0844.001.001	NCA	Water	2007	Alberta	SAMN06270822
C3GFDB015981	0811.008.001	NCA	Water	2007	Alberta	SAMN06270823
C3GFDB015988	0817.003.001	NCA	Water	2007	Alberta	SAMN06270824
C3GFDB016437	0817.003.001	NCA	Water	2007	British Columbia	SAMN06270825
C3GFDB019988	0811.008.001	NCA	Water	2008	Ontario	SAMN06270826
C3GFDB020240	0811.008.001	NCA	Water	2009	Ontario	SAMN06270827
C3GFDB016025	0811.009.002	NCA	Water	2010	Alberta	SAMN06270828
C3GFDB020625	0811.009.002	NCA	Water	2010	Ontario	SAMN06270829
C3GFDB021309	0811.009.002	NCA	Water	2010	Quebec	SAMN06270830
C3GFDB020827	0811.008.001	NCA	Water	2011	Ontario	SAMN06270831
C3GFDB009352	0926.002.001	UN	Chicken	2005	Ontario	SAMN06270832
C3GFDB009426	0926.002.001	UN	Human	2006	Ontario	SAMN06270833
C3GFDB009442	0173.002.004	UN	Human	2006	Ontario	SAMN06270834
C3GFDB009472	0735.005.001	UN	Human	2006	Ontario	SAMN06270835
C3GFDB009494	0853.011.001	UN	Human	2007	Ontario	SAMN06270836
C3GFDB008237	0960.003.002	UN	Cow	2010	Ontario	SAMN06270837
C3GFDB008715	0893.001.001	UN	Chicken	2011	Ontario	SAMN06270838
C3GFDB008759	0173.002.004	UN	Chicken	2011	Ontario	SAMN06270839
C3GFDB008776	0923.002.001	UN	Chicken	2011	Ontario	SAMN06270840
C3GFDB008798	0882.005.001	UN	Turkey	2011	Ontario	SAMN06270841
C3GFDB008833	0882.005.001	UN	Chicken	2011	British Columbia	SAMN06270842
C3GFDB008868	0893.001.001	UN	Chicken	2011	British Columbia	SAMN06270843
C3GFDB016476	0695.006.001	UN	Human	2005	Alberta	SAMN06270844
C3GFDB016480	0253.004.001	UN	Human	2005	Alberta	SAMN06270845
C3GFDB016487	0695.006.001	UN	Human	2005	Alberta	SAMN06270846
C3GFDB016492	0933.008.001	UN	Human	2005	Alberta	SAMN06270847
C3GFDB016521	0695.006.001	UN	Human	2005	Alberta	SAMN06270848
C3GFDB016545	0083.001.002	UN	Human	2005	Alberta	SAMN06270849
C3GFDB016553	0083.001.002	UN	Human	2005	Alberta	SAMN06270850
C3GFDB016561	0083.001.002	UN	Human	2005	Alberta	SAMN06270851
C3GFDB016566	0960.007.001	UN	Human	2005	Alberta	SAMN06270852
C3GFDB016580	0926.002.001	UN	Human	2005	Alberta	SAMN06270853
C3GFDB016583	0253.004.001	UN	Human	2005	Alberta	SAMN06270854
C3GFDB016602	0933.008.001	UN	Human	2005	Alberta	SAMN06270855

Strain	CGF Subtype	Clinical Relevance	Host	Isolation Year	Isolation Location	Accession Number
C3GFDB016630	0083.001.002	UN	Human	2005	Alberta	SAMN06270856
C3GFDB017566	0695.006.001	UN	Human	2004	Alberta	SAMN06270857
C3GFDB017618	0260.007.001	UN	Human	2005	Alberta	SAMN06270858
C3GFDB015479	0731.001.005	UN	Water	2004	Alberta	SAMN06270859
C3GFDB014430	0949.001.002	UN	Cow	2004	Alberta	SAMN06270860
C3GFDB014444	0949.001.002	UN	Cow	2004	Alberta	SAMN06270861
C3GFDB015494	0695.006.001	UN	Water	2004	Alberta	SAMN06270862
C3GFDB014498	0960.007.001	UN	Cow	2004	Alberta	SAMN06270863
C3GFDB014510	0695.006.001	UN	Cow	2004	Alberta	SAMN06270864
C3GFDB014519	0238.007.002	UN	Sheep	2004	Alberta	SAMN06270865
C3GFDB015508	0949.001.002	UN	Sewage	2004	Alberta	SAMN06270866
C3GFDB014531	0731.001.005	UN	Cat	2004	Alberta	SAMN06270867
C3GFDB015612	0904.002.002	UN	Water	2004	Alberta	SAMN06270868
C3GFDB015628	0933.008.001	UN	Sewage	2004	Alberta	SAMN06270869
C3GFDB014616	0083.001.002	UN	Chicken	2004	Alberta	SAMN06270870
C3GFDB014624	0083.001.002	UN	Chicken	2004	Alberta	SAMN06270871
C3GFDB014647	0731.001.005	UN	Cow	2005	Alberta	SAMN06270872
C3GFDB014687	0695.006.001	UN	Cow	2005	Alberta	SAMN06270873
C3GFDB014743	0957.001.001	UN	Sheep	2005	Alberta	SAMN06270874
C3GFDB014764	0957.001.001	UN	Cow	2005	Alberta	SAMN06270875
C3GFDB014772	0957.001.001	UN	Sheep	2005	Alberta	SAMN06270876
C3GFDB018263	0535.001.003	UN	Water	2006	New Brunswick	SAMN06270877
C3GFDB019677	0535.001.003	UN	Water	2006	Ontario	SAMN06270878
C3GFDB014863	0695.006.001	UN	Cow	2006	Alberta	SAMN06270879
C3GFDB015779	0960.007.001	UN	Water	2006	Alberta	SAMN06270880
C3GFDB014881	0926.002.001	UN	Goose	2006	Alberta	SAMN06270881
C3GFDB014887	0926.002.001	UN	Dog	2006	Alberta	SAMN06270882
C3GFDB014933	0926.002.001	UN	Duck	2006	Alberta	SAMN06270883
C3GFDB014939	0926.002.001	UN	Chicken	2006	Alberta	SAMN06270884
C3GFDB018269	0535.001.003	UN	Water	2006	New Brunswick	SAMN06270885
C3GFDB019812	0957.001.001	UN	Water	2006	Ontario	SAMN06270886
C3GFDB015839	0904.002.002	UN	Water	2006	Alberta	SAMN06270887
C3GFDB019857	0957.001.001	UN	Water	2006	Ontario	SAMN06270888
C3GFDB015915	0960.007.001	UN	Water	2007	Alberta	SAMN06270889
C3GFDB019903	0735.005.001	UN	Water	2007	Ontario	SAMN06270890
C3GFDB014971	0949.001.002	UN	Cow	2009	Alberta	SAMN06270891
C3GFDB014974	0949.001.002	UN	Cow	2009	Alberta	SAMN06270892
C3GFDB020075	0731.001.005	UN	Water	2009	Ontario	SAMN06270893
C3GFDB020274	0957.001.001	UN	Water	2009	Ontario	SAMN06270894
C3GFDB020424	0535.001.003	UN	Water	2010	Ontario	SAMN06270895
C3GFDB021276	0535.001.003	UN	Water	2010	Quebec	SAMN06270896
C3GFDB020474	0957.001.001	UN	Water	2010	Ontario	SAMN06270897
C3GFDB016023	0535.001.003	UN	Water	2010	Alberta	SAMN06270898
C3GFDB020618	0957.001.001	UN	Water	2010	Ontario	SAMN06270899
C3GFDB015045	0695.006.001	UN	Cow	2011	Alberta	SAMN06270900
C3GFDB015120	0695.006.001	UN	Cow	2011	Alberta	SAMN06270901
C3GFDB020782	0957.001.001	UN	Water	2011	Ontario	SAMN06270902
C3GFDB018644	0904.002.002	UN	Raccoon	2011	Ontario	SAMN06270903
C3GFDB018673	0904.002.002	UN	Raccoon	2011	Ontario	SAMN06270904
C3GFDB018718	0904.002.002	UN	Raccoon	2011	Ontario	SAMN06270905
C3GFDB018774	0535.001.003	UN	Raccoon	2011	Ontario	SAMN06270906
C3GFDB020936	0957.001.001	UN	Water	2011	Ontario	SAMN06270907
C3GFDB015346	0695.006.001	UN	Cow	2012	Alberta	SAMN06270908
C3GFDB015373	0695.006.001	UN	Cow	2012	Alberta	SAMN06270909
C3GFDB015427	0695.006.001	UN	Horse	2012	Alberta	SAMN06270910
C3GFDB015431	0695.006.001	UN	Horse	2012	Alberta	SAMN06270911

Table S2.2: Well-characterised genomes used to generate the reference pan-genome.

Accession Number	Species	Strain	Reference
AL111168.1	*C. jejuni* jejuni	NCTC 11168	Parkhill et al., 2000
NC_003912.7	*C. jejuni* jejuni	RM1221	Fouts et al., 2005
NC_008787.1	*C. jejuni* jejuni	81-176	Direct Submission
NC_009707.1	*C. jejuni* doylei	269.97	Direct Submission
NC_009839.1	*C. jejuni* jejuni	81116	Direct Submission

Chapter 3

3 Laboratory validation and implementation of a molecular-based risk assessment assay for the identification of human-pathogenic *Campylobacter jejuni*

3.1 Abstract

A major challenge in the prevention and control of campylobacteriosis is our inability to identify strains that pose the greatest risk to human health. In a previous study, we conducted a Genome-Wide Association Study (GWAS) to identify genetic markers associated with *C. jejuni* strains linked to human illness that could be used in a molecular-based risk assessment assay. Here, these markers were assembled into six multiplex PCR assays for wetlab validation using an expanded panel of strains (n=362) representing Non-Clinically-Associated (NCA) and Clinically-Associated (CA) subtypes circulating in Canada. These data were used to select a subset of the best-performing markers (n=11), which were re-organised into two multiplex PCRs comprising the finalised *Campylobacter jejuni* Risk Assessment (CJRA) Assay. The CJRA Assay was then deployed to screen field isolates (n=601) collected as part of the Canadian Food Inspection Agency's (CFIA) Microbial Baseline Survey in poultry. Overall, the assay detected 99.7% (579/581) of strains classified as being clinically associated, and 64.7% (247/382) of strains classified as being non-clinically associated. The detection of human-pathogenic strains belonging to subtypes not previously associated with human illness (i.e. those that are novel or emerging), highlights the utility of the risk assessment assay to identify high-risk strains in the absence of other supporting metadata used to infer risk.

3.2 Introduction

A major challenge in the prevention and control of campylobacteriosis is the inability to identify pathogenic *C. jejuni* strains, especially in light of molecular epidemiological evidence that suggests not all *C. jejuni* strains or genetic lineages pose an equal risk to human health (Buchanan et al., 2017; French et al., 2005; Sheppard et al., 2012). Unlike other enteric pathogens, for which there are well-described features (i.e. virulence factors or specific serotypes) linked to human disease that are used as markers in risk assessment assays, no such tool exists for *C. jejuni.* Addressing this issue would result in better tracking of high-risk strains, leading to a better understanding of their distribution in the food chain, and provide critical data for the development of targeted mitigation strategies to reduce human exposure to *C. jejuni.*

The previous chapter described a Genome-Wide Association Study (GWAS) designed to identify genetic markers that could be used in a molecular-based risk assessment (MRA) assay for *C. jejuni.* The gene content between strains from prominent non-pathogenic and pathogenic subtypes was compared, which resulted in the identification of 28 putative diagnostic markers that were statistically significantly linked to strains from subtypes associated with human illness. These markers were subjected to an *in silico* validation using a much larger and epidemiologically diverse dataset composed of all available WGS data that possessed sufficient epidemiological metadata to determine a strain's pathogenic status. Based on this analysis, 25 markers continued to exhibit robust diagnostic behaviour for the identification of pathogenic *C. jejuni* strains.

In the current study, the markers described above were implemented as a wetlab assay and subjected to additional validation before being used to screen a large panel of

isola llected as part of the Canadian Food Inspection Agency's (CFIA) Microbial Baseline Study (MBS) in broiler chicken (CFIA, 2016).

3.3 Materials and methods

3.3.1 Putative diagnostic markers

A detailed description of methods used to identify the putative diagnostic markers is described in Chapter 2 (Buchanan et al., 2017). Briefly, 166 strains representing 34 highly prevalent subtypes circulating in Canada were selected for whole genome sequencing. The non-redundant pan-genome was defined for the dataset, and a GWAS was conducted to compare the distribution of gene content between cohorts of strains from Non-Clinically-Associated (NCA; <5% human clinical isolates) and Undefined/Clinically-Associated (UN/CA; ≥5% human clinical isolates) subtypes to identify genes that were statistically significantly associated with the latter. A total of 28 putative diagnostic markers were identified and validated *in silico* (Table 3.1).

Table 3.1: Putative diagnostic markers identified in Buchanan et al., 2017 after GWAS analysis of genome sequences from representative Clinically-Associated (CA) and Non-Clinically-Associated (NCA) *C. jejuni* subtypes.

Marker	11168 Ortholog	Gene Name	Function	Linkage Group
11168_00051†*	*Cj0055c*		hypothetical protein	LG1
11168_00052†	*Cj0056c*		hypothetical protein	
11168_00169†	*Cj0177*		putative iron transport protein	LG2
11168_00170†	*Cj0178*		putative TonB-denpendent outer membrane receptor	
11168_00171†	*Cj0179*	*exbB1*	biopolymer transport protein	
11168_00172†	*Cj0180*	*exbD1*	biopolymer transport protein	
11168_00173†	*Cj0181*	*tonB1*	TonB transport protein	
11168_00230†	*Cj0246c*		putative MCP-domain signal transduction protein	
11168_00243	*Cj0259*	*pyrC*	putative dihydroorotase	LG3
11168_00244†*	*Cj0260c*		small hydrophobic protein	
11168_00248†*	*Cj0264c*		putative molybdopterin containing oxidoreductase	LG4
11168_00249†	*Cj0265c*		putative cytochrome C-type haem-binding periplasmic protein	
11168_00277†*	*Cj0295*		putative acetyltransferease	LG5
11168_00278†*	*Cj0296c*	*panD*	aspartate 1-decarboxylase precursor	
11168_00279	*Cj0297c*	*panC*	pantoate-beta-alanine ligase	
11168_00280†	*Cj0298c*	*panB*	3-methyl-2-oxobutanoate hydroxymethyltransferase	
11168_00281†	*Cj0299*		putative periplasmic beta-lactamase	
11168_00703†	*Cj0731*		putative ABC transport system permease	
11168_00718†*	*Cj0753c*	*tonB3*	TonB transport protein	LG6
11168_00719†	*Cj0755*	*cfrA*	ferric enterobactin uptake receptor	
11168_01072†*	*Cj1122c*		putative integral membrane protein.	
11168_01201†	*Cj1255*		putative isomerase	
11168_01309†*	*Cj1365c*		putative secreted serine protease	
11168_01519†*	*Cj1585c*		putative oxidoreductase	
11168_01610†*	*Cj1679*		hypothetical protein	
06_2866_00597†			di-/tripeptide transporter	
06_7515_00723†			prophage Lp2 protein 6	
07_0675_00227†*		*tetO*	elongation factor G	

† denotes markers screened using PCR

* denotes markers selected for the final *Campylobacter jejuni* Risk Assessment Assay

3.3.2 Primer and multiplex PCR assay design

Orthologous sequences from the draft genome assemblies and NCTC 11168 for each putative diagnostic marker were identified by homology searching with BLAST v 2.2.29 (Camacho et al., 2009). Paired BLAST queries were treated as orthologous if they shared ≥80% sequence identity and ≥50% alignment coverage. A multiple-sequence alignment (MSA) for each set of orthologues was generated using MAFFT v. 7.245 (Katoh & Standley, 2013) and a consensus sequence was generated from each MSA using BioEdit v. 7.2.5 (Hall, 1998). The consensus sequences representing the putative diagnostic markers were split into six groups such that each group was composed of a range of short to long gene sequences to promote primer design that would yield amplicons of staggered size to facilitate scoring on agarose gels. The PrimerPlex Module in AlleleID v 7.83 (Premier Biosoft International, Palo Alto, CA, USA) was used to design primers suitable for multiplexing with the following guidelines: a length of 25-30 bp, a melting temperature between 57-61°C and with minimal self- and hetero-dimerisation (i.e. Delta G ≥ -8.0 kcal/mole). Amplicons within a multiplex had to fall between 100 bp – 1000 bp in length with a minimum size difference of at least 30 bp. Specificity of the primers was assessed using NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the *C. jejuni* taxid (197).

3.3.3 Multiplex PCR optimisation

Each of the six multiplex PCR assays (MP1-MP6; Table 3.2) was optimised for primer concentration, magnesium concentration, annealing temperature, cycle number and extension time (data not shown). A panel of 11 strains with corresponding whole genome sequence data was used to validate the multiplex PCRs during the optimisation

process. The optimised multiplex PCR mixtures consisted of 1 µL of nucleic acid, 0.2 µL of Taq Polymerase (MP Biomedicals, Solon, OH, USA), either 1.5 mM $MgCl_2$ for MP5 or 3.0 mM $MgCl_2$ for all other multiplexes (MP Biomedicals), 0.2 mM of each dNTP (Invitrogen, Burlington, ON, CAN), 0.2 µM of each primer (IDT DNA Technologies, Coralville, IA, USA) in 10 x reaction buffer (MP Biomedicals) with PCR-grade water to a final volume of 25 µL. PCR was carried out with a pre-denaturation stage of 94°C for 5 min followed by 30 cycles of 94°C for 30 s, 65°C for 30 s, 72°C for 1 min, with a final extension step of 72°C for 5 min on the Mastercycler Pro thermal cycler (Eppendorf, Mississauga, ON, CAN). PCR amplified material was assayed on the QIAxcel Capillary Gel Electrophoresis System (QiAgen, Mississauga, ON, CAN) with the DNA Screening Kit (QiAgen) for visualisation of the amplification products; markers were scored as present if a PCR product of the correct size was observed.

Table 3.2: The multiplexed PCR primers used to screen strains for the putative diagnostic markers. MP1-F and MP2-F represent the finalised multiplex PCR assays for the *Campylobacter jejuni* Risk Assessment (CJRA) Assay.

		Primer Sequence (5' - 3')		
Multiplex	Marker	Forward	Reverse	Amplicon Size
MP1	11168_00169	GTTAGCAAGCACAGAGCAAAATCACTT	CAATACATCAGCCATACGCCTGTCT	373
	11168_00170	GGGTGGCAAGATAGGTAGTAGTATGGT	AAGCAGGTCTGTGTCCTTTGATATTCC	623
	11168_00172	ACAAAGAAGAGGAGTTAAGTGAGATCAA	ACTTGTGATAAGAGGAGTAACAGCCATA	101
	11168_00718	AACCCACACAAACTATACAAGAGCCTTC	AAGAACCTCCCCACTCATACGCATT	314
	11168_01201	CGGAGAACCTACAACAGAACAAAAACAA	TCCACCTAGTCCGTAATTATCCGTATCG	139
MP2	11168_00051	AGGTGCTTTGCCTTATCGTGGAAAA	ACCCATAGCCATTTCGTATTCTTTCACA	400
	11168_00244	TTGGACGCTTGATTCTTGATGCAATT	ACTGACTATCAATGACAATAAACCAGCT	119
	11168_00248	TCCTTTACAGCACTATACCGCCGATA	TCCACTTGATTTCCAGCCATAACTTCC	241
	11168_00281	ACTTTCTCTCCCGCTTCCACTTTTAA	TTGATACCAATTCTTCTTGCCACTTCTT	209
MP3	11168_00173	CAAGGAGCGGAATTTACTTCCATTATGA	ATCTTTATTACTTTGTAGCGGAGCACTG	282
	11168_00249	TACCGCCTATACCGATGAAGTTGTATC	AATCACAGAAGGTGCTTTAGGATTTACA	166
	11168_00277	GCAGAAATTGCTTGATTTGCTTTTTACT	TTTTGTCCCTCTTTGATAACCTTGTTTT	198
	11168_01519	CTTACGAGAGATTAAACGCCTTGAAAGC	ACGCCCTAGTTCTTGGTATAATGCTAAG	356
	06_7515_00723	TGAAACAATAGCTGAACGCAAAACACTT	TGGCACTGACACTAGCAAGAATTGATTT	803
MP4	11168_00171	ATCAAAATGCTGATGAGGTGGTAAAAGG	TGCTGTGATTTGCTTTGAGATGGTATTT	305
	11168_00230	GTCTAGGTGAAATGGAAGGAGCTATCAA	TCAAAATCTTCGCTGACATTGCTTCTTT	205
	11168_00703	GCCTGATGATCTTACGCTTAAATGGTAT	GTATCCAAGCTGTGCCTCCTATATTGTC	275
	11168_01309	GGAACTTCTATGGCAGCACCTATGG	TCCTTGCACGGTTAAATCATTCAAATCT	870
	06_2866_00597	GTGGATGGCTTGCGGATAATTACTTAG	AGTGTATTGTTGCTCATATACTGACCAA	710
MP5	11168_00278	AGCGTAACAGAAGCTAGGCTTGATTAT	GCCATTATTGACATTGACTACTTGGACT	117
	11168_00719	CCTTATAGAGATGTTGCAGAGGCTATCG	CCTTGAACTCTTTGTCCACTTCCATTTG	506
	11168_01072	GGGGTGGATTGTCATTAATTTACTAACC	ACAGAAGAAACTGATACTTGCAAACTCC	276
	11168_01610	ACTACAAGAAGCCAATCAACACCATG	AAATTATCATCCACTCCCATACGCATCA	183
MP6	11168_00052	TGAAGTTATAGCCTTTGCCAAACCTGAT	CGCACATTCCACCAAATCGAACAAG	119
	11168_00280	GAAAGCCACGCAAATGGGGTAAAAG	CCTGAACCTATGCCGATTGTTGGAA	302
	07_0675_00227	ACAATGACGATATGGAACAGTGGGATG	CAGATTGACCTTCAGGCGTTGATGA	240
MP1-F	11168_00051	AGGTGCTTTGCCTTATCGTGGAAAA	ACCCATAGCCATTTCGTATTCTTTCACA	400
	11168_00248	TCCTTTACAGCACTATACCGCCGATA	TCCACTTGATTTCCAGCCATAACTTCC	241
	11168_00278	AGCGTAACAGAAGCTAGGCTTGATTAT	GCCATTATTGACATTGACTACTTGGACT	117
	11168_01072	GGGGTGGATTGTCATTAATTTACTAACC	ACAGAAGAAACTGATACTTGCAAACTCC	276
	11168_01519	CTTACGAGAGATTAAACGCCTTGAAAGC	ACGCCCTAGTTCTTGGTATAATGCTAAG	356
	11168_01610	ACTACAAGAAGCCAATCAACACCATG	AAATTATCATCCACTCCCATACGCATCA	183
MP2-F	11168_00244	TTGGACGCTTGATTCTTGATGCAATT	ACTGACTATCAATGACAATAAACCAGCT	119
	11168_00277	GCAGAAATTGCTTGATTTGCTTTTTACT	TTTTGTCCCTCTTTGATAACCTTGTTTT	198
	11168_00718	AACCCACACAAACTATACAAGAGCCTTC	AAGAACCTCCCCACTCATACGCATT	314
	11168_01309	GGAACTTCTATGGCAGCACCTATGG	TCCTTGCACGGTTAAATCATTCAAATCT	870
	07_0675_00227	ACAATGACGATATGGAACAGTGGGATG	CAGATTGACCTTCAGGCGTTGATGA	240

3.3.4 Validation

The optimised multiplex PCR assays were screened against a total of 362 *C. jejuni* strains, which included a validation panel of 221 strains previously characterised by whole genome sequencing. The dataset was composed of 188 isolates from 34 CGF subtypes classified as UN/CA, and 174 isolates from 31 CGF subtypes classified as NCA. The statistical significance of each marker ($p < 0.05$) was defined based on its carriage rate in the UN/CA and NCA cohorts and computed using Fisher's Exact test statistic in GenomeFisher (https://bitbucket.org/peterk87/genomefisher/wiki/Home); p-values were adjusted for multiple testing using the method of Holm (Aickin & Gensler, 1996; Gaetano, 2013; Holm, 1979). The concordance between the *in silico* gene presence/absence data from the original GWAS and experimental PCR data from the sequenced strains, as well as the overall distribution of markers between the UN/CA and NCA cohorts, were used to select a subset of robust markers to facilitate high throughput screening. A total of 11 markers were reorganised into the finalised *C. jejuni* Risk Assessment (CJRA) assay comprising two multiplex PCRs (MP1-F and MP2-F, Table 3.2), and optimised as described above. The final PCR mixtures consisted of 1 µL of nucleic acid, 0.2 µL of Taq Polymerase, 3.0 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 µM of each primer in 10 x reaction buffer with PCR-grade water to a final volume of 25 µL. The PCR cycling conditions were unchanged.

3.3.5 Deployment of the *C. jejuni* Risk Assessment (CJRA) Assay

An additional 591 *C. jejuni* isolates collected as part of the CFIA's National MBS in broiler chicken (CFIA, 2016) were screened using the CJRA Assay. CGF subtyping (Clark et al., 2012; Taboada et al., 2012) was subsequently carried out by the National

Microbiology Laboratory's Campylobacter Genomics Laboratory to classify these strains as NCA or UN/CA based on the percentage of clinical isolates present in their respective CGF subtypes as described above.

3.3.6 Generation of a minimum spanning tree

A Minimum Spanning Tree (MST) was generated and visualised using PHYLOViZ 2.0 (Nascimento et al., 2017) based on all unique *C. jejuni* CGF subtypes (n=2,610) characterised in the Canadian Campylobacter CGF Database (C3GFdb). Subtyping data were input in a format compatible with the Multi-Locus Sequence Type (MLST) module in PHYLOViZ, and the goeBURST Full MST algorithm was used to construct the tree. A companion file based on all *C. jejuni* strains (n=17,279) characterised in the C3GFdb was used to overlay strain metadata onto the MST to visualise features of the population structure of *C. jejuni* in Canada.

3.4 Results and discussion

3.4.1 Validation of the putative diagnostic markers

Although the putative diagnostic markers were subjected to a rigorous *in silico* validation that clearly demonstrated a biased distribution amongst strains from clinically associated subtypes, a large majority of the publicly available *C. jejuni* genomes used in the study were of human clinical origin, and there was a paucity of genomic data derived from strains of subtypes rarely associated with human illness (Buchanan et al., 2017). Therefore, in addition to assessing the concordance between the *in silico* and

experimental PCR results, this validation was critical towards testing the markers against additional strains from Non-Clinically-Associated subtypes.

Twenty-six of the 28 markers identified in the GWAS were organised into six multiplex PCR assays. Two markers (Cj0259 and Cj0297c) were dropped due to poor amplification results, however, since both were members of different linkage groups comprising genes with similar carriage rates, their signal was considered redundant, and their exclusion from the assay was deemed acceptable. The six multiplex PCR assays were screened against a total of 362 *C. jejuni* strains including 188 strains from 34 CGF subtypes classified as UN/CA (≥5% human clinical), and 174 strains from 31 CGF subtypes classified as NCA (<5% human clinical). Among the validation dataset were 221 strains with whole genome sequence data. The overall concordance between the *in silico* and experimental PCR results was 98.8% (5678/5746 matching loci), and the number of discordant results for individual markers ranged from 0 to 13 discrepancies. The statistical significance of each marker based on its distribution between both cohorts was assessed, and all 26 markers exhibited statistically significant signal with clinically associated subtypes, with p-values ranging from 2.8×10^{-2} to 7.3×10^{-63} (Table 3.3). Overall, these results are similar to those reported in Buchanan et al., 2017, except three markers (06_2866_00597, 06_7515_00723 and Cj1255), which were not statistically associated with UN/CA subtypes in the *in silico* study, were found to be significant in this validation. On average, the markers were present in 76% of strains from UN/CA subtypes compared to only 26% of strains from NCA subtypes (Figure 3.1), while the average number of markers present in strains from UN/CA subtypes was 20, and seven in strains from NCA subtypes. Overall, the assay detected 100% of strains from UN/CA

subtypes, and 65% of strains from NCA subtypes (a strain was considered to be detected if at least one marker was present).

Table 3.3: The statistical significance of each putative diagnostic marker after screening against the validation dataset of *C. jejuni* strains.

	p-value[1]	
Marker	Raw	Holm-corrected[2]
11168_00051	6.54E-30	9.16E-29
11168_00052	1.95E-29	2.53E-28
11168_00169	3.84E-04	1.54E-03
11168_00170	4.64E-08	3.68E-07
11168_00171	1.14E-04	5.69E-04
11168_00172	3.84E-04	1.54E-03
11168_00173	7.44E-05	4.46E-04
11168_00230	1.30E-09	1.17E-08
11168_00244	4.24E-62	1.02E-60
11168_00248	2.80E-64	7.27E-63
11168_00249	8.06E-56	1.77E-54
11168_00277	8.37E-62	1.93E-60
11168_00278	6.32E-55	1.33E-53
11168_00280	2.88E-53	5.46E-52
11168_00281	1.79E-62	4.48E-61
11168_00703	5.21E-14	5.21E-13
11168_00718	1.02E-48	1.84E-47
11168_00719	1.84E-35	2.76E-34
11168_01072	1.30E-53	2.60E-52
11168_01201	2.87E-02	2.87E-02
11168_01309	3.21E-47	5.46E-46
11168_01519	2.24E-19	2.47E-18
11168_01610	1.61E-44	2.58E-43
06_2866_00597	4.60E-08	3.68E-07
06_7515_00723	4.22E-04	1.54E-03
07_0675_00227	1.80E-26	2.16E-25

[1] p-value based on 2-tailed Fisher's Exact Test

[2] p-values were adjusted using the Holm-correction (Holm, 1979)

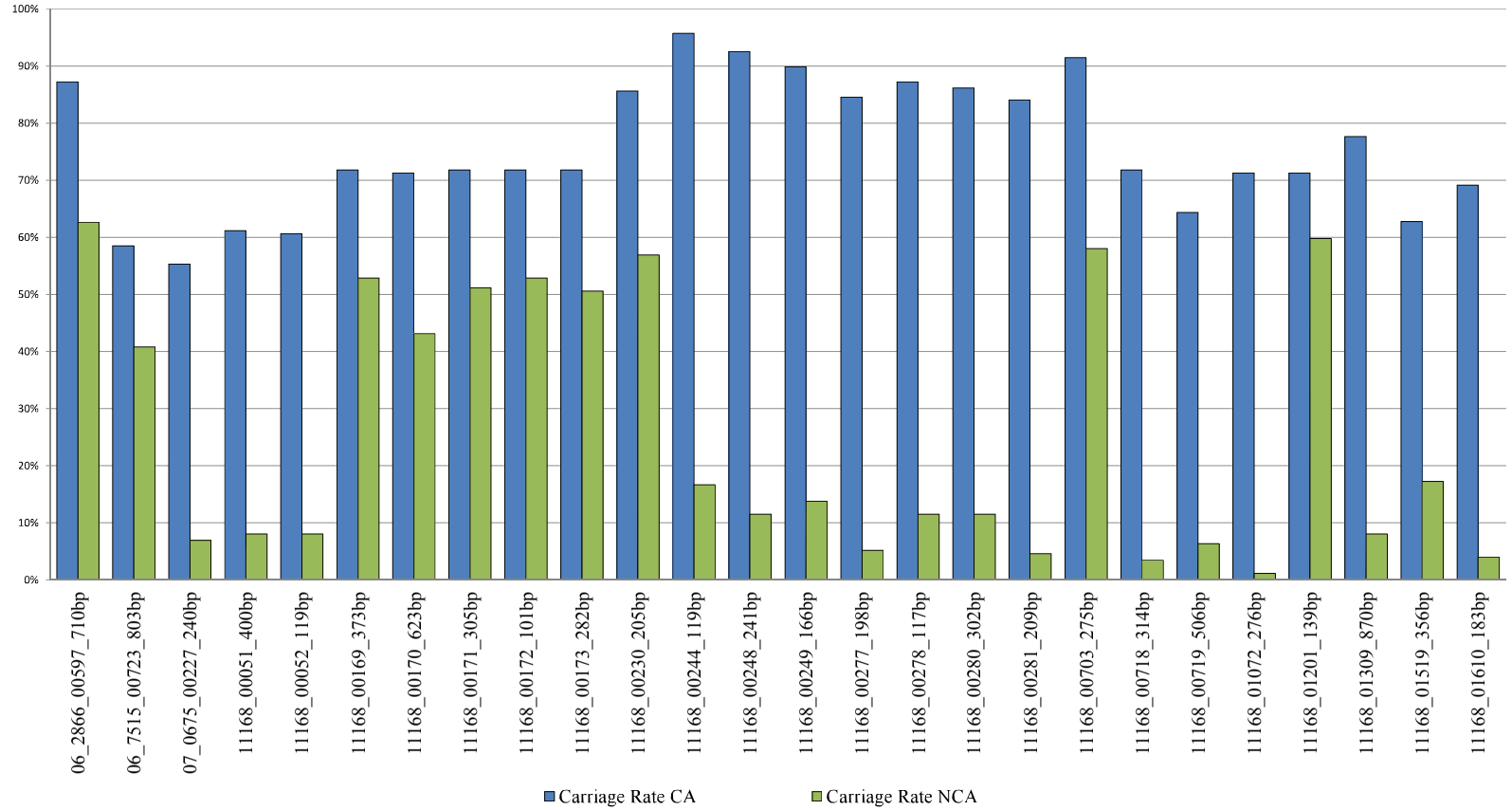

Figure 3.1: The distribution of the wetlab PCR results for 26 putative diagnostic markers identified in the GWAS conducted by Buchanan et al., 2017 against strains from UN/CA subtypes (blue bars; n = 188) and NCA subtypes (green bars; n = 174). All markers were statistically-significantly associated with strains from CA subtypes (Holm-corrected p-values ranging from 2.8×10^{-2} to 7.3×10^{-63}). On average, the markers were present in 76% of strains from UN/CA subtypes compared to only 26% of strains from NCA subtypes. A subset of robust markers was selected for an abbreviated, more user-friendly assay, which excluded any marker present in more than 20% of strains from NCA subtypes, and included one or two markers representative of each linkage group.

Of the 174 strains from NCA subtypes, no markers were present in 61 strains, while various numbers of markers were observed in the remainder, which could be indicative of some level of risk despite membership in CGF subtypes not presently associated with human campylobacteriosis. A minimum spanning tree based on data from the C3GFdb was used to examine NCA subtypes with positive markers to better contextualise their underlying epidemiological properties based on the characteristics of related subtypes. For example, strains from the NCA subtype 0170_003_002 were positive for all 26 markers, which would suggest an elevated risk for human pathogenesis. In fact, this subtype is highly related to one of the most prevalent CA subtypes in Canada. Of the 24 NCA subtypes encompassing strains positive for one or more putative diagnostic marker, 14 were exclusive to, or predominantly carried by raccoons, which more recently, have been shown to carry strains capable of causing disease in humans (Saunders et al., 2017). Raccoons occupy a unique habitat at the interface of farm and/or urban environments, which can bring them into close proximity with livestock, pets and people (Rainwater et al., 2017). While investigation of similar subtypes did not yield many instances of strains from human clinical cases, overlap with strains from poultry, cattle and environmental waters was observed, which underscores a potentially underappreciated facet of campylobacteriosis epidemiology involving wild animals, and their ability to transmit a diverse range of *C. jejuni* lineages between the wild, farm and urban environments (Whiley et al., 2013). Although these subtypes have not been previously associated with human campylobacteriosis, it should be noted that this may not necessarily be a reflection of pathogenicity, but a potential knowledge gap resulting from the incomplete surveillance of *C. jejuni* including from human clinical cases (Figure 3.2).

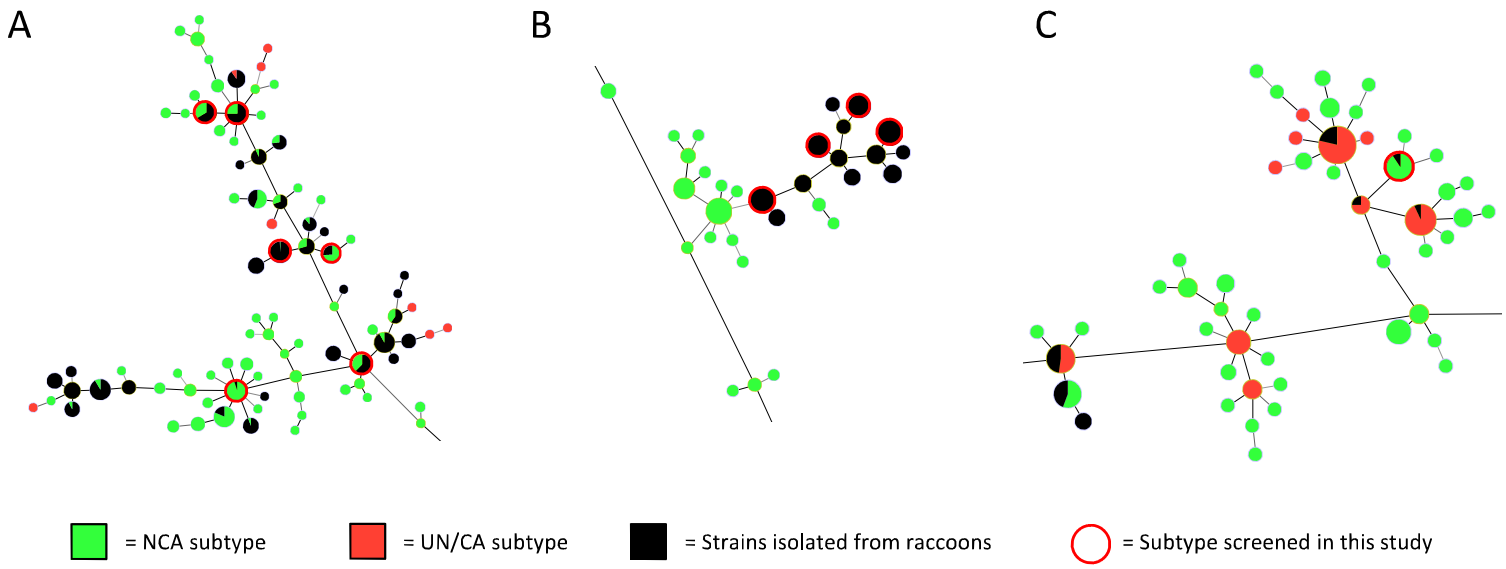

Figure 3.2: A minimum spanning tree (MST) depicting CGF subtypes where the risk level is not accurately reflected based on the current understanding of the population structure of *C. jejuni* in Canada. Green represents NCA subtypes, while red represents UN/CA subtypes. Strains isolated from raccoons are black. A red outline represents subtypes screened in this study. Panels A and B depict examples of NCA subtypes that contain strains with one or more putative diagnostic markers, which suggests an elevated level of risk. However, based on the current understanding of the population structure of *C. jejuni,* these subtypes would appear to be benign due to the absence of human clinical isolates. In Panel C, we see examples of the opposite: NCA subtypes comprising strains with elevated risk that are more closely related to pathogenic subtypes. Also note, many subtypes are composed of strains from raccoons, other animal and environmental sources (i.e. mixed colouration), which demonstrates the potential for transmission of strains from the wild to farm and urban environments.

A subset of robust markers was selected for an abbreviated, more user-friendly assay, which excluded any marker present in more than 20% of strains from NCA subtypes, and included one or two markers representative of each linkage group. A total of 11 markers were selected, which could still detect all strains from the UN/CA subtypes. These markers were re-organised into two multiplex PCR assays and re-optimised before implementation in the laboratory as the *C. jejuni* Risk Assessment (CJRA) assay (Table 3.2).

3.4.2 Deployment of the *C. jejuni* Risk Assessment (CJRA) Assay

The CFIA's Microbial Baseline Study is the first ever Canadian study examining *Campylobacter* in broiler chicken, and thus, provided an attractive opportunity to further assess the CJRA Assay for several reasons. *C. jejuni* is highly prevalent in raw poultry meat and poultry by-products, accounting for a significant proportion (65-69%) of human campylobacteriosis cases in Canada (Ravel et al., 2017; Suzuki & Yamamoto, 2009; Williams & Oyarzabal, 2012), and the scope of this survey is unparalleled, targeting poultry products at the farm, abattoir and retail sectors across Canada accounting for 92.9% of total chicken production and retail sampling designed to reflect the purchasing habits of a majority of the Canadian population (i.e. 62%) (CFIA, 2016). To that end, an additional 601 *C. jejuni* isolates were screened using the CJRA Assay, bringing the total number of isolates tested to 963. The final dataset encompassed 581 *C. jejuni* strains from 76 UN/CA subtypes and 382 strains from 112 NCA subtypes. Overall, the markers were present, on average, in 68% of strains from UN/CA subtypes compared to only 24% of strains from NCA subtypes (Figure 3.3), while the average number of markers present in strains from UN/CA subtypes was eight, and three in strains from NCA subtypes.

Overall, the assay detected 99.7% (579/581) of strains from UN/CA subtypes, and 65% (247/382) of strains from NCA subtypes. Amongst the strains from the MBS, the markers were present, on average, in 65% of strains from UN/CA subtypes, and 37% of strains from NCA subtypes, while the average numbers of markers present in strains from UN/CA subtypes was seven, and four in strains from NCA subtypes. The assay detected 99.5% (389/391) of strains from the UN/CA subtypes, and 90.0% (180/200) of strains from NCA subtypes.

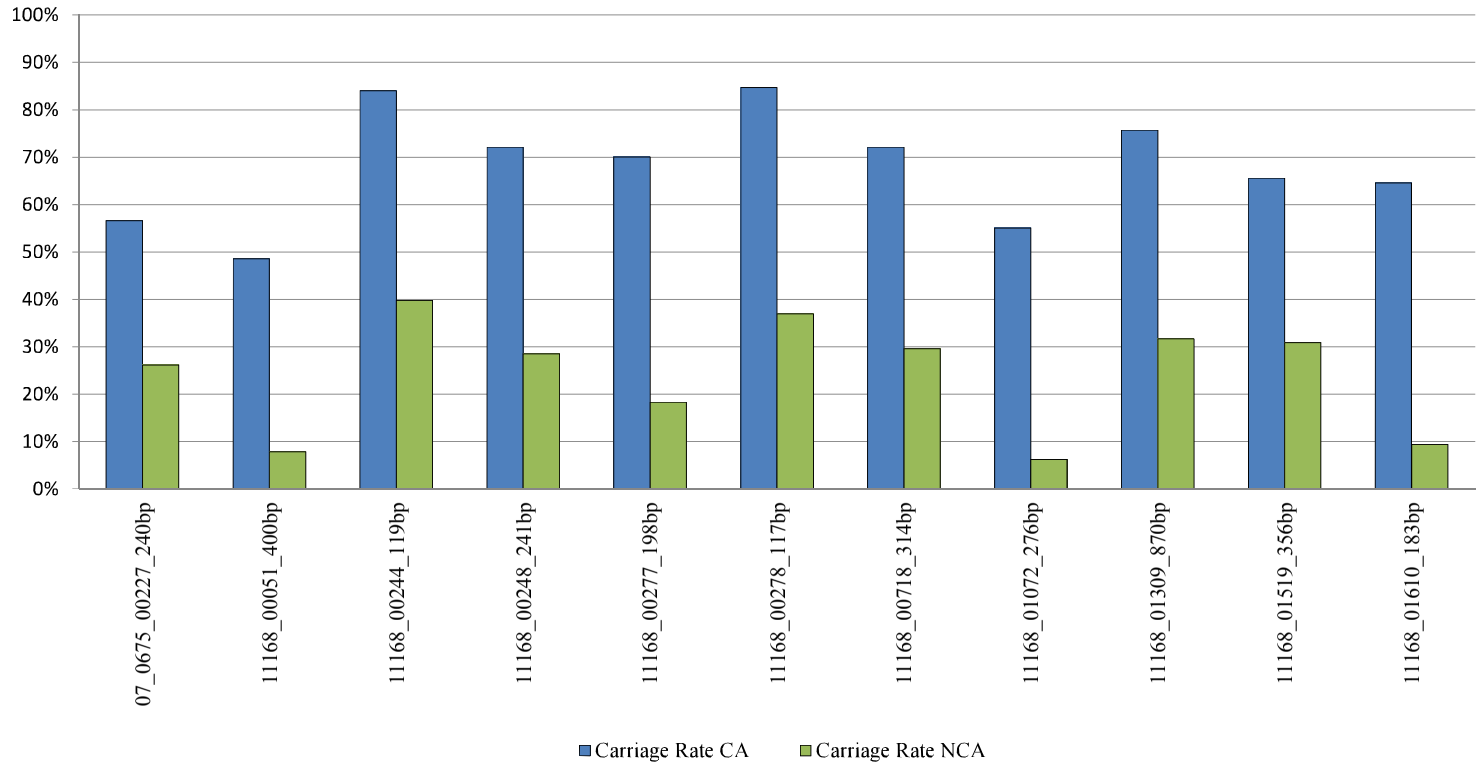

Figure 3.3: The distribution of the wetlab PCR results for 11 diagnostic markers representing the finalised CJRA Assay against strains from UN/CA subtypes (blue bars; n = 581) and NCA subtypes (green bars; n = 382). All markers were statistically-significantly associated with strains from CA subtypes (Holm-corrected p-values ranging from 4.8×10^{-21} to 3.8×10^{-70}). Overall, the markers were present, on average, in 68% of strains from UN/CA subtypes compared to only 24% of strains from NCA subtypes.

The National Microbiology Laboratory's *Campylobacter* Genomics Laboratory completed CGF subtyping on all *C. jejuni* isolates collected during the MBS, including those assessed in this study using the CJRA Assay. In total, 352 subtypes were characterised in the survey, and *C. jejuni* strains from 166 of those subtypes, representing 85% (2,045/2,401) of the collection, were assessed with the CJRA Assay. Despite the breadth of the C3GFdb, which includes more than 26,000 *Campylobacter* strains representing over 4,800 unique subtypes, over a third of the subtypes characterised in the MBS were novel (n=137). This is a testament to the ability of *C. jejuni* to generate genetic diversity, and a stark reminder of the complexity of its population structure. Overall, we observed a relationship suggesting an increasing proportion of clinical isolates in CGF subtypes and the number of positive CJRA markers in strains they include, which could eventually be used to develop a risk prediction scheme (i.e. low, medium, high risk) for a strain based on the number of markers present. However, there were a number of strains from CGF subtypes classified as NCA that were positive for at least one CJRA marker (Figure 3.4). For example, among the MBS isolates tested with the CJRA Assay, 37 strains were from 31 novel CGF subtypes classified as NCA based on the absence of any clinical association; more than half of these were positive for five or more CJRA markers suggesting a greater level of risk than the NCA classification would imply. Novel subtypes, in addition to rare subtypes, or those biased by oversampling are problematic because they lack robust epidemiological signal to confidently infer pathogenic risk. This problem is further exacerbated by underreporting at the clinical level, or the lack of clinical surveillance (Allos & Blaser, 1995), which limits our ability to project these important metadata over the population structure of *C. jejuni*. In fact, this is a possible explanation for the higher than expected detection of

high-risk strains from subtypes classified as NCA in the MBS. Additionally, we observed that strains with identical CGF subtypes also shared an identical or similar distribution of risk markers (Figure 3.5). Although more validation is required, these data suggest that it might be possible to assign risk to the subtype itself, similar to how other human pathogens are classified (e.g. the "seropathotyping" scheme for shigatoxigenic *E. coli* (Karmali et al., 2003)). This would allow the CJRA Assay to be used to assign risk to emerging and poorly-defined subtypes in the absence of other metadata used to predict pathogenic risk, and allow public health officials to expedite a response when a high-risk strain is detected.

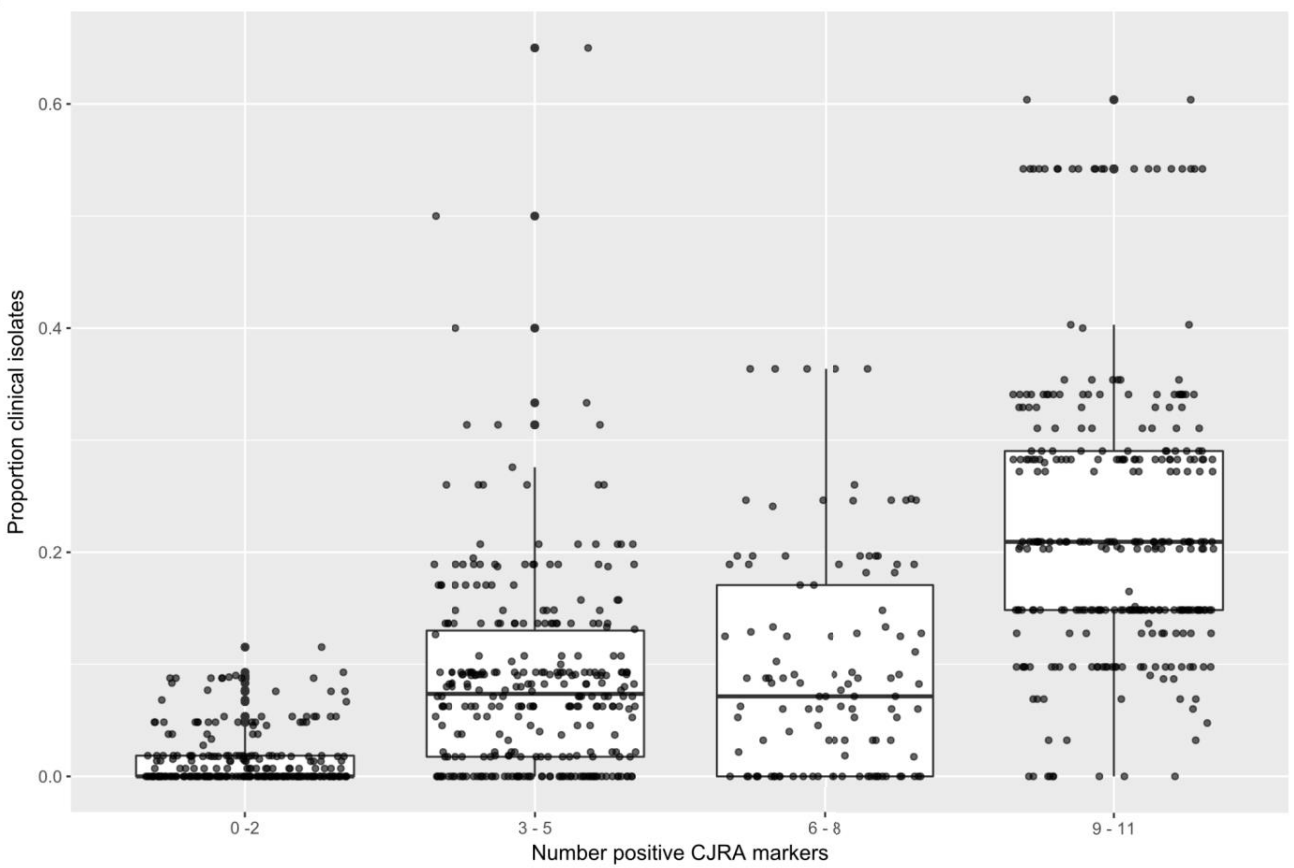

Figure 3.4: Strains screened using the finalised CJRA Assay were categorised into bins based on the number of CJRA markers present (i.e. 0-2, 3-5, 6-8, 9-11 markers), then plotted against the proportion of clinical isolates from the CGF subtypes they were derived from. Overall, we observed a trend towards an increasing proportion of clinical isolates with the number of positive CJRA markers. However, we also observed a number of strains that were positive for one or more CJRA markers from CGF subtypes classified as NCA, which would suggest they possess a greater level of risk than is implied based on their current designation. These data suggest that such subtypes lack robust epidemiological signal to confidently infer their pathogenic risk based on the current understanding of the population structure of *C. jejuni* in Canada, and highlights the utility of the CJRA Assay to assign risk to emerging and poorly-defined subtypes in the absence of other data.

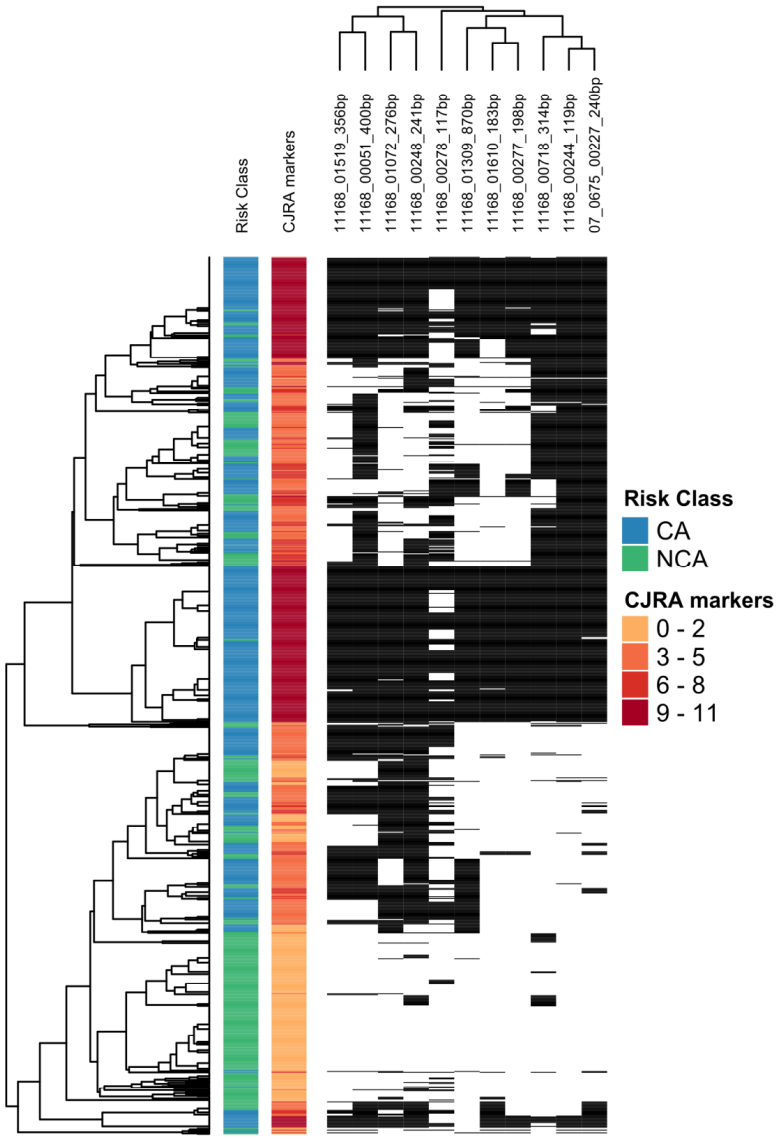

Figure 3.5: Strains screened using the finalised CJRA Assay were clustered based on their CGF subtype. Strains from subtypes classified as NCA are coloured green, and those classified as CA are coloured blue. The number of CJRA markers present were binned (i.e. 0-2, 3-5, 6-8, 9-11) and colour-coded from light orange to dark red. The binarised CJRA Assay risk profile was added for each strain. As was observed in Figure 4, there is a possible correlation between the risk classification (i.e. NCA and CA) and the number of CJRA markers present, which could eventually be used to develop a risk prediction scheme (i.e. low, medium, and high-risk) based on the number of markers present. Strains with identical CGF subtypes generally possessed an identical or similar distribution of risk markers. Although further validation is required, these data suggest it might be possible to assign a risk level to the subtype itself, similar to how other human pathogens are classified (e.g. the "seropathotyping" scheme for shigatoxigenic *E. coli* (Karmali et al., 2003)).

3.5 Conclusion

The CFIA's MBS confirmed that poultry continues to be an important reservoir for *C. jejuni* on farm (average prevalence of 24.1%), at slaughter (average prevalence of 33%) and at retail (average prevalence of 41.8%) (CFIA, 2016). Subsequent characterisation of isolates using CGF revealed a wide variety of lineages circulating within poultry flocks, including a large number of novel subtypes where the pathogenic potential could not be established based on the current understanding of the population structure of *C. jejuni* and a paucity of human surveillance data in Canada. Based on the *in silico* validation described in Chapter 2 (Buchanan et al., 2017), and the expanded validation described here, the putative diagnostic markers identified in the GWAS are demonstrably associated with clinically associated CGF subtypes. Therefore, it is reasonable to assume that the detection of high-risk strains amongst subtypes classified as NCA in the MBS is legitimate and demonstrates the utility of the CJRA Assay to infer risk in cases where the underlying epidemiological data is lacking. It is possible the CJRA Assay could enhance the current framework used to study the epidemiology of campylobacteriosis retrospectively by assessing strains from known subtypes that lack robust epidemiological signal, and proactively by assessing strains from novel and emerging subtypes. This would enable public health officials to better identify and track high-risk *C. jejuni* strains circulating within the environment and food chain, and prioritise cases based on risk level as is done with other bacterial pathogens. These data can also be used to help implement targeted mitigation strategies to limit human exposure to *C. jejuni.*

3.6 References

Aickin, M., and Gensler, H. (1996). Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm Methods. *American Journal of Public Health* 86, 726-728.

Allos, B., and Blaser, M. (1995). *Campylobacter jejuni* and the expanding spectrum of related infections. *Clinical Infectious Diseases* 20, 1092-1099.

Buchanan, C., Webb, A., Mutschall, S., Kruczkiewicz, P., Barker, D., Hetman, B., Gannon, V., Abbott, D., Thomas, J., Inglis, G., and Taboada, E. (2017). A genome-wide association study to identify diagnostic markers for human pathogenic *Campylobacter jejuni* strains. *Frontiers in Microbiology* 8, 1-9.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 1-9.

CFIA. (2016). *National microbiological baseline study in broiler chicken December 2012 - December 2013.* Canada: Retrieved from: http://www.inspection.gc.ca/food/chemical-residues-microbiology/food-safety-testing-bulletins/2016-08-17/december-2012-december-2013/eng/1471358115567/1471358175297?chap=5.

Clark, C., Taboada, E., Grant, C., Blakeston, C., Pollari, F., Marshall, B., Rahn, K., MacKinnon, J., Daignault, D., Pillai, D., and Ng, L. K. (2012). Comparison of molecular typing methods useful for detecting clusters of *Campylobacter jejuni* and *C. coli* isolates through routine surveillance. *Journal of Clinical Microbiology* 50, 798-809.

French, N., Barrigas, M., Brown, P., Ribiero, P., Williams, N., Leatherbarrow, H., Birtles, R., Bolton, E., Fearnhead, P., and Fox, A. (2005). Spatial epidemiology and natural population structure of *Campylobacter jejuni* colonizing a farmland ecosystem. *Environmental Microbiology* 7, 1116-1126.

Gaetano, J. (2013). *Holm-Bonferroni sequential correction: an EXCEL calculator.* v. 1.2. Retrieved from: https://www.researchgate.net/publication/242331583_Holm-Bonferroni_Sequential_Correction_An_EXCEL_Calculator_-_Ver_12.

Hall, T. (1998). *BioEdit. Biological sequence alignment editor for Windows.* v. 7.2.5. Retrieved from: http://www.mbio.ncsu.edu/BioEdit/bioedit.html.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70.

Karmali, M.A., Mascarenhas, M., Shen, S., Ziebell, K., Johnson, S., Reid-Smith, R., Isaac-Renton, J., Clark, C., Rahn, K., and Kaper, J. B. (2003). Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli*

seropathotypes that are linked to epidemic and/or serious disease. *Journal of Clinical Microbiology* 41, 4930-4940.

Katoh, K., and Standley, D. (2013). MAFFT multiple sequence alignment software Version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30, 772-780.

Nascimento, M., Sousa, A., Ramirez, M., Francisco, A., Carriço, J., and Vaz, C. (2017). PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. *Bioinformatics (Oxford, England)* 33, 128-129.

Rainwater, K., Marchese, K., Slavinski, S., Humberg, L., Dubovi, E., Jarvis, J., McAloose, D., and Calle, P. (2017). Health survey of free-ranging raccoons *(Procyon lotor)* in Central Park, New York, New York, USA: implications for human and domestic animal health. *Journal of Wildlife Diseases* 53, 272-284.

Ravel, A., Hurst, M., Petrica., N., David, N., Mutschall, S., Pintar, K., Taboada, E., and Pollari, F. (2017). Source attribution of human campylobacteriosis at the point of exposure by combining comparative exposure assessment and subtype comparison based on comparative genomic fingerprinting. *PLOS One* 12, 1-21.

Saunders, S., Smith, K., Schott, R., Dobbins, G., and Scheftel, J. (2017). Outbreak of campylobacteriosis associated with raccoon contact at a wildlife rehabilitation centre, Minnesota, 2013. *Zoonoses and Public Health* 64, 222-227.

Sheppard, S., Jolley, K. A., and Maiden, M. C. J. (2012). A gene-by-gene approach to bacterial population genomics: whole genome MLST of *Campylobacter*. *Genes* 3, 261-277.

Suzuki, H., and Yamamoto, S. (2009). *Campylobacter* contamination in retail poultry meats and by-products in the world: a literature survey. *Journal of Veterinary Medical Science* 71, 255-261.

Taboada, E., Ross, S., Mutschall, S., MacKinnon, J., Roberts, M., Buchanan, C., Kruczkiewicz, P., Jokinen, C., Thomas, J., Nash, J., Gannon, V., Marshall, B., Pollari, F., and Clark, C. (2012). Development and validation of a comparative genomic fingerprinting method for high-resolution genotyping of *Campylobacter jejuni*. *Journal of Clinical Microbiology* 50, 788-797.

Whiley, H., van den Akker, B., Giglio, S., and Bentham, R. (2013). The role of environmental reservoirs in human campylobacteriosis. *International Journal of Environmental Research and Public Health* 10, 5886-5907.

Williams, A., and Oyarzabal, O. (2012). Prevalence of *Campylobacter* spp. in skinless, boneless retail broiler meat from 2005 through 2011 in Alabama, USA. *BMC Microbiology* 12, 1-7.

Chapter 4

4 Conclusions and future directions

4.1 Conclusions

Despite the fact that *C. jejuni* is a leading cause of bacterial enteritis in Canada and abroad, there is currently no diagnostic assay available to identify strains that present an elevated risk for human health (Blaser, 1997; Thomas, M. et al., 2013). Molecular epidemiological data suggest that not all *C. jejuni* subtypes pose an equal risk to human health, however, it is unclear what drives this subtype-specific pathogenesis (Buchanan et al., 2017; French et al., 2005; Sheppard et al., 2012). Therefore, the first component of this thesis was to identify genetic markers associated with *C. jejuni* strains known to cause disease in humans. This was accomplished by conducting a Genome-Wide Association Study (GWAS) comparing the gene content between strains from prominent non-pathogenic and pathogenic CGF subtypes circulating in Canada. The analysis yielded 28 putative markers that were statistically significantly linked to *C. jejuni* strains from subtypes associated with human illness, which included genes involved in iron acquisition and vitamin B5 biosynthesis, in addition to those encoding putative proteins associated with catalysis and transport. Some of the genes identified have been shown experimentally to be important for colonisation, survivability and host transmissibility (Naikare et al., 2013; Palyada et al., 2004; Sheppard et al., 2013), while the role in *C. jejuni* pathogenesis (if any) of the other markers is as of yet undefined. Broadly, these data suggest that the hallmark of high-risk strains might be their ability to overcome ecological barriers enroute to the human host, namely, the ability to withstand adverse

environmental conditions, and the ability to efficiently transition between hosts (Dearlove et al., 2016; Gripp et al., 2011; Oh et al., 2018). The markers were subjected to an *in silico* validation using a much larger and epidemiologically diverse dataset comprising all available WGS data that possessed sufficient epidemiological metadata to determine a strain's pathogenic status. Based on this analysis, 25 markers continued to exhibit robust diagnostic behaviour for the identification of pathogenic *C. jejuni* strains.

The second component of this thesis was to develop a PCR-based diagnostic assay using the genetic markers identified in the GWAS. Primers for all 28 markers were designed, of which 26 were implemented into six multiplex PCR assays and screened against a panel of *C. jejuni* strains from both non-clinically and clinically-associated lineages. Although, all markers exhibited robust diagnostic behaviour for detection of pathogenic *C. jejuni* strains, a subset comprising the best-performing markers was re-organised into two multiplex PCR assays to facilitate higher throughput and faster turnaround time. The finalised *Campylobacter jejuni* Risk Assessment (CJRA) Assay comprising two multiplex PCR assays representing 11 markers, was deployed against a pan-Canadian set of *C. jejuni* isolates collected as part of the CFIA's Microbial Baseline Study in broiler chicken (CFIA, 2016). In combination with CGF subtyping, the analysis revealed several salient features: 1) *C. jejuni* is highly prevalent in Canadian chickens and poultry products; 2) there are many novel subtypes circulating in the Canadian poultry chain comprising strains that possess an elevated risk for human illness; 3) the CJRA Assay was able to identify potentially high-risk strains from novel subtypes in the absence of other epidemiological data; 4) when strains are clustered based on CGF subtype, identical strains also share an identical or similar distribution of risk markers; and 5) there was a positive trend between the number of positive markers and the

proportion of human clinical isolates within a CGF subtype. Therefore, not only will the CJRA Assay be useful for identifying high-risk *C. jejuni* strains, but with further validation, it could be used to establish "risk levels" for a given CGF subtype, similar to the seropathotyping scheme devised for shigatoxigenic *E. coli*. In the seropathotyping scheme, serotypes are ranked for risk based on their occurrence in human disease, severity of disease and association with outbreaks (Karmali et al., 2003). Such a model for *C. jejuni* would allow public health officials to make more informed decisions on how to deploy resources to mitigate campylobacteriosis in Canada.

4.2 Future directions

4.2.1 Are all *C. jejuni* pathogenic?

C. jejuni is unique amongst human bacterial enteric pathogens as it does not possess many classical virulence factors, and those it does have, are broadly conserved across the species (Dorrell et al., 2001). A possible implication of this is that all *C. jejuni* are inherently pathogenic in a naïve human host. However, molecular epidemiological data, including MLST and CGF, have shown that not all lineages are equally implicated in disease in humans (Buchanan et al., 2017; French et al., 2005; Sheppard et al., 2012). The GWAS conducted in this work revealed a number of genetic markers that were statistically associated with strains from subtypes known to cause disease in humans. Amongst these markers were genes encoding multiple iron acquisition systems, which have been well-studied in *C. jejuni* and shown to be important for colonisation in model animals, and may play role in survivability and host transmission (Naikare et al., 2013; Palyada et al., 2004; Zeng et al., 2013). Additionally, genes encoding the vitamin B5

biosynthesis pathway were identified in this study, which more recently, have been shown to be important for rapid transmission between hosts (Sheppard et al., 2013). While the analysis revealed a number of other genes related to transport and catabolism, in addition to those with no known function, their roles, if any, in the pathogenesis of *C. jejuni* remain undefined, and represent an avenue for future research. Overall, these results suggest the pathogenicity of *C. jejuni* may not be linked to some as of yet identified molecular machinery, but rather, to its ability to overcome ecological barriers or bottlenecks enroute to the human host, and thus, the best markers for high-risk *C. jejuni* might be genes that enhance its ecological robustness.

As stated above, while some of the markers represent well-studied genes in *C. jejuni,* the analysis identified other genes encoding proteins with roles that are less well defined, and represent an unknown quantity in terms of their role in *C. jejuni* pathogenesis. Thus, future work could involve elucidating the functions of the proteins to determine if and how they might enhance *C. jejuni* pathogenicity. CRISPR/Cas9 technology has greatly simplified our ability to precisely modify an organism's genome. Therefore, this system can be used to generate knockout mutants for each gene to assess subsequent effects on the phenotype. Broad-spectrum phenotypic arrays that assess bacterial metabolism and chemical sensitivities can be used to rapidly compare the mutant and wildtype strains, which may shed light on the function of the deleted protein. Although the use of animal models to study *C. jejuni* has been restricted largely due to its inability to cause disease in most common research animals, recent work has established several mouse models that develop acute enterocolitis in response to *C. jejuni.* These models could provide an avenue to evaluate the potential role of these genes in the colonisation and infection by *C. jejuni in vivo* (Stahl et al., 2017).

4.2.2 Increasing the resolution: whole genome SNP analysis

When the study was conceptualised, an early decision was made to conduct the GWAS at the gene level for two reasons: 1) the hypothesis was based on the premise that differences in pathogenicity would be linked to genes from the accessory genome; and 2) to facilitate downstream implementation of a presence/absence PCR-based diagnostic assay. Thus, by design, only the accessory gene content was interrogated. Therefore, future work could be centred upon repeating the GWAS at the nucleotide level, which would have the added advantage of interrogating the whole genome including both the core and accessory gene content, as well as non-coding and regulatory regions of the genome. Furthermore, with the imminent integration of whole genome sequencing in routine diagnostics, important SNPs and other polymorphisms can be easily detected using bioinformatic tools, making their use as diagnostic markers more palatable. GWAS have been used to identify SNPs and other mutations associated with antibiotic resistance in *Mycobacterium tuberculosis* (Farhat et al., 2013), *Staphylococcus aureus* (Alam et al., 2014) and *Streptococcus pneumoniae* (Chewapreecha et al., 2014), as well as a recent study investigating variant alleles in *C. jejuni* and *C. coli* associated with survival during poultry processing (Yahara et al., 2017).

4.2.3 Developing a risk framework for *C. jejuni*

In general, strains that were identical by CGF, also shared an identical or similar distribution of risk markers. Therefore, if risk profiles are found to be stable within a CGF subtype, the risk of any given strain could be inferred based on its subtype so long as that subtype has been previously characterised. This would negate having to run the risk assessment assay, resulting in an expedited risk analysis.

While the markers identified in this study have not, as of yet, been linked to virulence, if the appropriate metadata comes available, it could be beneficial to determine if the presence of certain markers can be correlated to certain epidemiological features (e.g. associated with outbreaks), or more severe disease outcomes (e.g. diarrhea, hospitalisation, etc.). Additionally, we observed a trend where strains with more risk markers came from CGF subtypes comprising a higher proportion of human clinical strains. These features could be used to develop a risk framework for *C. jejuni* whereby a strain is assigned a "risk classification" (e.g. low, medium, high) based on criteria such as its CGF subtype, the presence of certain markers and the number of markers present. An analogous system exists for shigatoxigenic *E. coli,* where serotypes are grouped into "seropathotypes", and ranked for risk (from A-E) based on their epidemiological characteristics including their occurrence in human disease, the severity of disease, and whether they are associated with outbreaks (Karmali et al., 2003).

In order to develop such a framework for *C. jejuni,* further validation is required. More strains from well-characterised NCA and CA CGF subtypes need to be screened to ensure 1) that the markers continue to show a biased distribution amongst strains associated with human disease, 2) that the intra-subtype distribution of risk markers is stable, and thus, a reliable proxy for risk for a given subtype, and 3) whether risk can be inferred from one subtype to a similar one. However, it is important to note that a significant roadblock to implementing such a framework for *C. jejuni* in Canada is the paucity of human surveillance data. These data are necessary for determining which subtypes are causing disease in humans, how frequently and to what extent (i.e. severity of disease, manifestation of post-infection sequelae), and are needed in combination with data from the CJRA Assay to help inform the development of "risk classifications"

similar to the "seropathotypes" for shigatoxigenic *E. coli*. In the absence of domestic data, a possible work around is to leverage publicly-available WGS data from partner countries for further validation of the CJRA Assay and to help develop a risk framework for *C. jejuni.*

4.2.4 Towards implementation of the CJRA Assay: the poultry production continuum as a model

As described in the CFIA's Microbial Baseline Survey, *C. jejuni* was isolated at all stages of the poultry production continuum including the farm, abattoir and retail levels (CFIA, 2016). In this study, CGF subtyping in conjunction with the *C. jejuni* Risk Assessment Assay demonstrated that there is a wide variety of lineages circulating within Canadian poultry flocks, including many comprising strains that pose an elevated risk to human health. Because poultry is considered to be the most important route of infection for *C. jejuni*, intervention at key points of the poultry production chain could be the most effective strategy for mitigating human campylobacteriosis in Canada. As such, the poultry production chain can be used as a framework for the implementation of an enhanced molecular-based surveillance system that is capable of characterising which subtypes are in circulation along with their associated risk at each stage of production. These data can be used to develop targeted mitigation strategies designed to minimise both human campylobacteriosis, as well as food waste.

Implementation of on-farm sampling could be used to identify potential sources of contamination leading to flock infection, which could be used to inform the implementation of farm-specific and potentially industry-wide intervention strategies. Additionally, on-farm testing can be used to develop a gating system where chicken

carrying high-risk strains are diverted from general processing at centrally-located abattoirs minimising the potential for cross-contamination during slaughter and processing. Chicken flagged as carrying high-risk strains could then be processed separately and used for specialised products (i.e. cooked and frozen chicken nuggets), which should eliminate the bacterium, and reduce the risk of campylobacteriosis. Sampling at the abattoir along each stage of the slaughtering process could be used to determine if and where contamination is occurring; these data could be used to inform the implementation of new intervention strategies, and determine whether currently-used intervention strategies are efficacious. Testing products prior to their shipment to retail outlets could provide a final gating measure to prevent products contaminated with high-risk strains from being sold.

Recognising that this strategy would be costly to implement, and therefore, unlikely to garner support from industry partners and diagnostic laboratories over the long-term, the above strategy might be best implemented as a pilot study used to inform the development of a more sustainable, culture-free assay where raw samples (e.g. faeces, tissues, rinses, etc.) are tested directly. The pilot study would therefore be critical for determining which risk marker(s) are the most diagnostic for human pathogenic *C. jejuni,* as well as which points along the poultry production chain should be tested. Technologies such as real-time PCR and loop-mediated isothermal amplification are sensitive, and both rapid and inexpensive compared to isolation and subsequent downstream characterisation (i.e. subtyping and risk assessment) of *C. jejuni.* However, there are several limitations that need to be considered: 1) since both real-time PCR and loop-mediated isothermal amplification are less-amenable to multiplexing, it is necessary to determine which risk markers represent the best candidates to carry forward; 2) raw

samples will comprise nucleic acids from any number of microorganisms including the host, and therefore a suitable species-specific marker gene unique to *C. jejuni* would need to be incorporated into the assay to bolster a positive result; 3) the culture-free assay would need to be validated against current isolation/characterisation techniques to ensure that a positive result (i.e. detection of the species-specific *C. jejuni* marker and risk marker(s)) is due to the presence of *C. jejuni* with the selected risk marker, and not a result of non-specific amplification, cross-reactivity, or the combination of signal from multiple microorganisms.

4.3 References

Alam, M., Petit, R., Crispell, E., Thornton, T., Conneely, K., Jiang, Y., Satola, S., and Read, T. (2014). Dissecting vancomycin-intermediate resistance in *Staphylococcus aureus* using genome-wide association. *Genome Biology and Evolution* 6, 1174-1185.

Blaser, M. (1997). Epidemiologic and clinical features of *Campylobacter jejuni* infections. *The Journal of Infectious Diseases* 176, S105.

Buchanan, C., Webb, A., Mutschall, S., Kruczkiewicz, P., Barker, D., Hetman, B., Gannon, V., Abbott, D., Thomas, J., Inglis, G., and Taboada, E. (2017). A genome-wide association study to identify diagnostic markers for human pathogenic *Campylobacter jejuni* strains. *Frontiers in Microbiology* 8, 1-9.

CFIA. (2016). *National microbiological baseline study in broiler chicken December 2012 - December 2013.* Canada: Retrieved from: http://www.inspection.gc.ca/food/chemical-residues-microbiology/food-safety-testing-bulletins/2016-08-17/december-2012-december-2013/eng/1471358115567/1471358175297?chap=5.

Chewapreecha, C., Marttinen, P., Croucher, N., Salter, S., Harris, S., Mather, A., Hanage, W., Goldblatt, D., Nosten, F., Turner, C., Turner, P., Bentley, S., and Parkhill, J. (2014). Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes. *PLOS Genetics* 10, 1-11.

Dearlove, B., Cody, A., Pascoe, B., Méric, G., Wilson, D., and Sheppard, S. (2016). Rapid host switching in generalist *Campylobacter* strains erodes the signal for tracing human infections. *International Society for Microbial Ecology* 10, 721-729.

Dorrell, N., Mangan, J., Laing, K., Hinds, J., Linton, D., Al-Ghusein, H., Barrell, B., Parkhill, J., Stoker, N., Karlyshev, A., Butcher, P., and Wren, B. (2001). Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Research* 11, 1706-1715.

Farhat, M., Shapiro, B., Kieser, K., Sultana, R., Jacobson, K., Victor, T., Warren, R., Streicher, E., Calver, A., Sloutsky, A., Kaur, D., Posey, J., Plikaytis, B., Oggioni, M., Gardy, J., Johnston, J., Rodrigues, M., Tang, P., Kato-Maeda, M., Borowsky, M., Muddukrishna, B., Kreiswirth, B., Kurepina, N., Galagan, J., Gagneux, S., Birren, B., Rubin, E., Lander, E., Sabeti, P., and Murray, M. (2013). Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nature Genetics* 45, 1183-1189.

French, N., Barrigas, M., Brown, P., Ribiero, P., Williams, N., Leatherbarrow, H., Birtles, R., Bolton, E., Fearnhead, P., and Fox, A. (2005). Spatial epidemiology and natural

population structure of *Campylobacter jejuni* colonizing a farmland ecosystem. *Environmental Microbiology* 7, 1116-1126.

Gripp, E., Hlahla, D., Didelot, X., Kops, F., Maurischat, S., Tedin, K., Alter, T., Ellerbroek, L., Schreiber, K., Schomburg, D., Janssen, T., Bartholomäus, P., Hofreuter, D., Woltemate, S., Uhr, M., Brenneke, B., Grüning, P., Gerlach, G., Wieler, L., Suerbaum, S., and Josenhans, C. (2011). Closely related *Campylobacter jejuni* strains from different sources reveal a generalist rather than a specialist lifestyle. *BMC Genomics* 12, 584-605.

Karmali, M.A., Mascarenhas, M., Shen, S., Ziebell, K., Johnson, S., Reid-Smith, R., Isaac-Renton, J., Clark, C., Rahn, K., and Kaper, J. B. (2003). Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. *Journal of Clinical Microbiology* 41, 4930-4940.

Naikare, H., Butcher, J., Flint, A., Xu, J., Raymond, K., and Stintzi, A. (2013). *Campylobacter jejuni* ferric-enterobactin receptor CfrA is TonB3 dependent and mediates iron acquisition from structurally different catechol siderophores. *Metallomics* 5, 988-996.

Oh, E., Chui, L., Bae, J., Li, V., Ma, A., Mutschall, S., Taboada, E., McMullen, L., and Jeon, B. (2018). Frequent implication of multistress-tolerant *Campylobacter jejuni* in human infections. *Emerging Infectious Diseases* 24, 1037-1044.

Palyada, K., Threadgill, D., and Stintzi, A. (2004). Iron acquisition and regulation in *Campylobacter jejuni*. *Journal of Bacteriology* 186, 4714-4729.

Sheppard, S., Didelot, X., Meric, G., Torralbo, A., Jolley, K., Kelly, D., Bentley, S., Maiden, M., Parkhill, J., and Falush, D. (2013). Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. *Proceedings of the National Academy of Sciences of the United States* 110, 11923-11927.

Sheppard, S., Jolley, K. A., and Maiden, M. C. J. (2012). A gene-by-gene approach to bacterial population genomics: whole genome MLST of *Campylobacter*. *Genes* 3, 261-277.

Stahl, M., Graef, F., & Vallance, B. (2017). Mouse models for *Campylobacter jejuni* colonization and infection. In J. Butcher, & A. Stintzi (Eds.), *Campylobacter jejuni Methods and Protocols* (pp. 171-188). New York, NY: Springer New York.

Thomas, M., Murray, R., Flockhart, L., Pintar, K., Pollari, F., Fazil, A., Nesbitt, A., and Marshall, B. (2013). Estimates of the burden of foodborne illness in Canada for 30 specified pathogens and unspecified agents, circa 2006. *Foodborne Pathogens and Disease* 10, 639-648.

Yahara, K., Méric, G., Taylor, A., de Vries, S., Murray, S., Pascoe, B., Mageiros, L., Torralbo, A., Vidal, A., Ridley, A., Komukai, S., Wimalarathna, H., Cody, A., Colles, F., McCarthy, N., Harris, D., Bray, J., Jolley, K., Maiden, M., Bentley, S., Parkhill, J., Bayliss, C., Grant, A., Maskell, D., Didelot, X., Kelly, D., and Sheppard, S. (2017). Genome-wide association of functional traits linked with *Campylobacter jejuni* survival from farm to fork. *Environmental Microbiology* 19, 361-380.

Zeng, X., Xu, F., and Lin, J. (2013). Specific TonB-ExbB-ExbD energy transduction systems required for ferric enterobactin acquisition in *Campylobacter*. *FEMS Microbiology Letters* 347, 83-91.