Development of Comparative Genomic Fingerprinting for Molecular Epidemiological Studies of *Campylobacter jejuni*

Susan Ross
Bachelor of Science, University of Lethbridge, 2006

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirement for the Degree

MASTER OF SCIENCE

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

© Susan Ross, 2009

Thesis Abstract

This thesis reports the development of Comparative Genomic Fingerprinting (CGF), a rapid genotyping method for *Campylobacter jejuni* that assesses the conservation status of 20 genes previously described as having high intraspecies variability based on comparative genomics studies. This novel method for genotyping *C. jejuni*, CGF was validated two-fold. First, by comparison to *fla*A restriction fragment length polymorphism analysis, and second a subset of isolates was validated using two higher resolution CGF assays assessing 35 and 119 genes. CGF was then tested in a molecular epidemiological study of *C. jejuni* isolated from environmental, animal and human clinical samples from southern Alberta. Reservoirs of infection, subtypes associated with higher incidence of human infection, and the persistence of prevalent subtypes in animal/environmental reservoirs were identified. This thesis demonstrates that CGF analysis is robust and can be used to rapidly assess genetic similarity of *C. jejuni* isolates and to detect epidemiologically relevant clonal groups.

Acknowledgements

This thesis would not have been possible without the support of my supervisors James Thomas and Victor Gannon. Their help has been invaluable, and I am grateful for the encouragement, guidance and expertise provided. When I was presented with a problem, they were able to suggest alternatives, and provide motivation. I am indebted that they brought me in under their wings, and provided me with the skills (and reference letters) needed to be successful.

I would like to thank Eduardo Taboada for guidance in project direction, and thesis writing. His knowledge of computer programming and genotyping guided me to develop my abilities as a scientist. Throughout my program, his support has provided the confidence to apply for conferences, and to reach for achievement.

It is an honour for me to thank my thesis examination committee members, Dan Johnson and Brent Selinger. Their knowledge and expertise pushed me to deepen my understanding of the subject. I appreciate the effort put forth by Douglas Inglis and thank for providing an external opinion during my graduate thesis defence. Additionally, I thank Elizabeth Schultz for fulfilling chairing the thesis examination committee.

I must thank Chris Armstrong-Ester. His encouragement from after-class conversations inspired me to see graduate training in Public Health. I would also like to thank Wayne Lippa and Helena Danyk for teaching me how to teach. They both ignited a fire that I never knew existed and I admire their patience and passion.

I would to thank several individuals who were able to make this research possible. Joanne MacKinnon, William Mauro, Steven Mutschall, and Catherine Yoshida were instrumental to my project for their efforts in genotyping. Jim Miller, Bruce Beasly, Ed Topp, Thomas Edge, Will Tobinson, Judy Issac Rention all contributed to sample collection. I am grateful to the many individuals in the Gannon laboratory for obtaining many of the *C. jejuni* isolates in this study, Unfortunately, you are too numerous to list by name. I would like to thank William Yates and Pam Gale for use of research facilities at the Animal Diseases Research Institute. I must also thank Betty Golsteyn-Thomas and the women of the Bacteriology department for assistance and accommodation during the final months of my graduate studies.

I would like to thank the Public Health Agency of Canada, the Canadian Food Inspection Agency, the University of Lethbridge, the Canadian Society of Microbiologists, the University of Montreal, The Alberta Ingenuity Fund and the School of Graduate Studies for financial support either directly or via scholarships.

Lastly, I owe my deepest gratitude to my friends and family for late night coffees, computer assistance, photo shoots and midnight brain-storming sessions. Additionally, they provided balance during precarious periods and occasionally a roof over my head, both of which will not be forgotten. Without them, this goal would have been unachievable.

Table of Contents

Thesis Abstract iii
Acknowledgements iv
Table of Contents vi
List of Tables viii
List of Abbreviations xi
1. Literature Review 1
1.1 Epidemiology of *Campylobacter* 1
1.1.1 Overview of *Campylobacter* and *C. jejuni* 1
1.1.2 *Campylobacter* as a human disease 2
1.1.3 Overview of *Campylobacter* illness 4
1.1.3.1 Complications resulting from *Campylobacter* infections 7
1.1.4 *Campylobacter* Outbreaks 9
1.1.5 Transmission and Prevalence of *C. jejuni* 10
1.1.5.1 *Campylobacter* bacteria in poultry and other avian sources 13
1.1.5.2 *Campylobacter* and swine 16
1.1.5.3 Prevalence in cattle 18
1.1.5.4 Prevalence in other animals 19
1.2 Molecular Genetics, Typing Methods, and *Campylobacter* 20
1.2.1 Introduction to molecular sub-typing 20
1.2.2 Phenotypic sub-typing methods 21
1.2.3 Molecular-based Methods of subtyping strains beyond the Genus-Species level and application to *Campylobacter jejuni* 21
1.2.3.1 An introduction to molecular methods for subtyping *Campylobacter* 22
1.2.3.1.1 Pulsed-field gel electrophoresis 23
1.2.3.1.2 Ribotyping 24
1.2.3.1.3 Random amplified polymorphic DNA 25
1.2.3.1.4 Analysis of the *fla*A allele (also known as *fla* typing) 26
1.2.3.1.5 Multi-Locus Sequence Typing 27
1.2.3.2 Current challenges in subtyping of *Campylobacter* 29
1.2.4 Development of next-generation molecular typing methods 31
1.2.4.1 Evolution of bacterial genomes 31
1.2.4.2 Genomics of *C. jejuni* 32
1.2.5 Comparative genomics of *C. jejuni* 33
1.2.6 Comparative genomics in understanding pathogenicity 33
1.2.7 Comparative genomic fingerprinting 36
1.3 Overview of the thesis 37
1.3.1 Research Objectives 37
2. Materials and Methods 40
2.1 Isolation of *Campylobacter* spp. from water 40
2.2 Isolation of *Campylobacter* spp from animal fecal samples 41
2.3 Isolation of *Campylobacter* spp. from clinical isolates 42
2.4 Extraction of *Campylobacter* DNA 43
2.5 Confirmation and species identification of *C. jejuni* and *C. coli* 44
2.6 Sample Subsets / Strain composition 46

2.7 Subtyping of strains using *fla*A-RFLP: 47
2.8 Subtyping using Comparative Genome Fingerprinting with 20 genes (CGF-20) .. 48
2.9 Subtyping using Comparative Genome Fingerprinting with 35 genes (CGF-35) .. 54
2.10 Subtyping using Comparative Genome Fingerprinting with 119 genes (CGF-119) 61
2.11 Molecular epidemiological analysis 65
2.12 Analysis of fingerprints using Bionumerics 66
2.13 Statistical analysis of clustering results 66
3. Development and validation of a comparative genomic fingerprint for *Campylobacter jejuni* 68
3.1 Introduction 68
3.2 Results 69
3.2.1 Analysis and validation of the CGF-20 assay 69
3.2.2 Development of a high-throughput CGF assay 72
3.2.2.1 Analysis of reproducibility: A comparison of agarose gel electrophoresis vs. capillary gel electrophoresis 73
3.2.2.2 Analysis of assay reproducibility using PCR replicates 76
3.2.3 Development of High Resolution CGF Assays: CGF-35 and CGF-119 76
3.2.3.1 Validation of CGF-20 clusters using CGF-35 and CGF-119 79
3.4 Discussion 79
4. Molecular epidemiology of *Campylobacter jejuni* in southern Alberta 86
4.1 Introduction 86
4.2 Results 87
4.2.1 Evaluation of samples from southern Alberta 87
4.2.2 Epidemiological analysis of *C. jejuni* isolates using a CGF-20 assay 89
4.2.3 Epidemiological analysis of *C. jejuni* isolates using a *fla*A-RFLP assay 94
4.2.4 Comparison of CGF-20 and *fla*A-RFLP typing results 99
4.2.4.1 Statistical comparison of CGF-20 and *fla*A-RFLP 99
4.2.5 Assessment of the ability of CGF-20 and *fla*A-RFLP assays to predict overall genetic similarity among *C. jejuni* isolates 103
4.2.6 Epidemiological comparison of the CGF-20 assay and *fla*A-RFLP assay 106
4.2.7 Examination of highly prevalent CGF-20 genotypes among the southern Alberta data set 114
4.3 Discussion 120
5. Review of Thesis 124
5.1 Thesis Conclusions 130
References 135
Appendix A: Table of 641 isolates that compose the southern Alberta *C. jejuni* data set 150
Appendix B: Detailed Clustering of the southern Alberta Data Set using Bionumerics v5.0 (Applied Maths, Austin, TX) 169

List of Tables

Table 1: Primer sequences used to confirm *Campylobacter* spp. 45
Table 2: Primer sequences for PCR sets 1-7 for CGF of *Campylobacter jejuni*. 51
Table 3: Olgionucleotide probe sequences for CGF-119 61
Table 4: Average intra-cluster and inter-cluster genetic similarities identified for CGF-20 clusters based on CGF-35 and CGF-119 data.. 81
Table 5: Isolation rates of *C. jejuni* from animal fecal samples 88
Table 6: Distribution of *fla*A-RFLP types among CGF-20 clusters composed of a single *fla*A-RFLP type.. 100
Table 7: *fla*A-RFLP composition of multi-member CGF-20 clusters 102
Table 8: Average intra-cluster and inter-cluster genetic similarities for multi-member CGF-20 clusters 104
Table 9: Average intra-cluster and inter-cluster genetic similarities for multi-member *fla*A-RFLP clusters 105

List of Figures

Figure 1: The route of infection of *C. jejuni* 5
Figure 2: *Campylobacter* in the environment and routes of transmission 11
Figure 3: Development of the CGF-20 Assay 49
Figure 4: Location of hyper-variable regions and the CGF targets within the *C. jejuni* genome 53
Figure 5: Visualization of CGF-20 targets using agarose gel electrophoresis........ 55
Figure 6: Visual representation of the CGF-35 images using the QIAxcel high-throughput capillary gel electrophoresis system. 57
Figure 7: Dendrogram of the CGF-20 data from the southern Alberta data set 71
Figure 8: A comparison of CGF-20 amplification products analyzed by capillary gel electrophoresis using the QIAxcel instrument with CGF-20 amplifications analyzed using agarose gel electrophoresis 74
Figure 9: Distribution of mismatches obtained for CGF-20 fingerprints when the amplicons were measured by capillary gel electrophoresis using a QIAxcel instrument. 75
Figure 10: Reproducibility of CGF-20 fingerprints by PCR replication. CGF-20 fingerprints were compared from 2 sets of PCR to establish the reproducibility of method 77
Figure 11: Distribution of mismatches obtained for CGF-20 fingerprints between replications of PCR. 78
Figure 12: Detailed genomic analysis of select CGF-20 clusters using CGF-35 and CGF-119 assays 80
Figure 13: Analysis of CGF-20 cluster composition by source type. 90
Figure 14: *C. jejuni* isolates from animal sources with matching CGF-20 fingerprints to human clinical isolates 92
Figure 15: Proportion of clinical isolates by CGF-20 cluster represented as a percentage of all clinical isolates in the dataset (n=260) 93
Figure 16: Analysis of *flaA* RFLP cluster composition by source type. Strains from the three source types (Animal fecal, water or clinical sample) were sub-divided into 182 *flaA*-RFLP clusters (including singletons) 95
Figure 17: *C. jejuni* isolates from animal sources with matching *fla*A-RFLP fingerprints to human clinical isolates 97
Figure 18: Clinical isolates arranged by *fla*A-RFLP type 98
Figure 19: Distribution of intra-cluster CGF-119 similarities for all pair-wise comparisons between the 74 isolates selected for CGF-119 analysis........ 107
Figure 20: Distribution of inter-cluster CGF-119 similarities for all pair-wise comparisons between the 74 isolates selected for CGF-119 analysis........ 108
Figure 21: CGF-20 relationships and fingerprints of the 6 isolates in *fla*A-RFLP cluster #15........ 110
Figure 22: Evidence of lateral transfer, or *flaA allelic transfer*, is observed when strains with very different CGF profiles share the same *fla*A-RFLP allele 111
Figure 23: Evidence of *flaA allelic replacement* or exchange of *fla*A alleles between highly related *C. jejuni* isolates........ 113
Figure 24: Proportion of clinical isolates by CGF-20 cluster represented as a percentage of all clinical isolates in the dataset (n=260) 115

Figure 25: Detailed genomic analysis of suspected endemic clones in southern Alberta 118
Figure 26: Temporal distribution of the six CGF-20 clusters containing a high proportion of clinical isolates in the dataset 119

List of Abbreviations

ADRI – Animal Diseases Research Institute

AFLP – Amplified Fragment Length Polymorphism

bp – Base Pair

CFIA – Canadian Food Inspection Agency

CGF – Comparative Genomic Fingerprint

CI – Confidence Interval

CLS – Clinical Laboratory Services

dNTP – Deoxynucleotide

DNA – deoxyribose nucleic acid

GI – Gastrointestinal

GBS – Guillain-Barré syndrome

*fla*A – Flagellin A gene

HGT – Horizontal Gene Transfer

ID – Index of Diversity

ORF – Open Reading Frame

PBS – Phosphate Buffered Saline

PCR – Polymerase Chain Reaction

PFGE – Pulsed-field Gel Electrophoresis

PHAC – Public Health Agency of Canada

PI-IBS – Post Infection Irritable Bowel Syndrome

PoP – Percent of Positive

LFZ – Laboratory for Foodborne Zoonosis

LGT – Lateral Gene Transfer

mCCDA – modified Charcoal Cefoperzone deoxycholate agar

MFS – Miller Fisher Syndrome

MLST – Multilocus Sequence Typing

NaPoi – Sodium phosophate dibasic

NCBI – National Center for Biotechnology Information

RAPD – Random Amplified Polymorphic DNA

ReA – Reactive Arthritis

RFLP – Restriction Fragment Length Polymorphism

SDS – Sodium Dodecyl Sulfate

SSC – Saline-Sodium Citrate

SNP – Single Nucleotide Polymorphism

SVR – Sequencing of the Variable Region

TD – Traveler's Diarrhea

UPGMA – Unweighted Pair Group Method with Arithmetic Mean

VAP – Variably Absent/Present

Chapter 1

1. Literature Review

1.1 Epidemiology of *Campylobacter*

Campylobacter organisms have long been known as a veterinary pathogen attributed to infectious abortion in cattle and sheep under the name *Vibro fetus* since 1909, but only in the recently decades become important as a cause of human illness (Blaser & Reller, 1981). To understand why *Campylobacter* may be seen as an emerging human pathogen (Altekruse, Stern, Fields, & Swerdlow, 1999) although studied since for nearly a century, first *Campylobacter* must be understood.

1.1.1 Overview of *Campylobacter* and *C. jejuni*

Campylobacter is a Gram-negative group of bacteria that is nearly ubiquitous in both animal, in particular mammalian, and water reservoirs (Koga *et al.,* 2005; Mangen *et al.,* 2005; Papworth and Olson, 2002). The bacterium has been isolated from water and animal sources including pigs, chicken, and cattle. *Campylobacter* is sensitive to many conditions including oxygen levels and prefers to grow under microaerobic conditions with low oxygen (3-15%), and high carbon dioxide levels (2-10%) (Snelling, Matsuda, Moore, & Dooley, 2005). These special growth requirements make it difficult to isolate some strains of *Campylobacter* in a laboratory setting.

Members of the genus *Campylobacter* that are most commonly associated with human infection, also known as campylobacteriosis, are referred to as "thermophilic

campylobacteria". This classification is determined by growth at 42°C and 37°C, but not below 30°C. Members of this group include *Campylobacter jejuni*, *Campylobacter coli*, *Campylobacter lari* and *Campylobacter upsaliensis*. *Campylobacter jejuni* is the predominant cause of gastrointestinal disease of the genus *Campylobacter* and is isolated in 80% or more of clinical human cases (Skirrow, 1997; Thomas, Gibson, & Mabey, 1999). The second most common is *C. coli*. Together *C. coli* and *C. jejuni* compose 95% of clinical isolates (Thomas *et al.*, 1999). It has been suggested that the high frequency of these two species may be a reflection of the isolation methods resulting in an underestimation of other species that might be the cause of clinical illness (Bolton, Hutchison, & Parker, 1987). However, the consensus remains that *C. jejuni* is the most common *Campylobacter* species associated with human enteritis (Allos, 2001; Walker *et al.*, 1986).

1.1.2 *Campylobacter* as a human disease

Campylobacter has long been associated with Traveler's Diarrhea (TD) and incidence of enteric disease in developing countries. Up to 55% of travelers to developing countries will acquire TD (Yates & Stetz, 2006) and 45% of troops deployed in the middle-east have reported diarrhea resulting in decreased job performance for a median of three days (Sanders, Putnam, Riddle, & Tribble, 2004). Travelers have a high incidence of diarrheal disease (1 to 15 episodes per person per year) while residing in developing countries (Payment, 2001). Children in developing countries exhibit a peak incidence of diarrheal disease at 24 months of age before the incidence declines in adulthood (Payment, 2001), possibly due to the acquisition of protective immunity from

constant exposure (Thomas *et al.*, 1999). These high childhood disease rates are thought to be one reason for the high child mortality rates in certain developing countries (Payment, 2001). In many locations, such as Thailand, *Campylobacter* infections are the leading cause of TD (Yates & Stetz, 2006). *C. jejuni* is responsible for nearly all cases of human campylobacteriosis acquired overseas (Schrotz-King, Prokhorova, Nielsen, Crawford, & Morsczeck, 2007).

Infection by *Campylobacter* also is a leading cause of illness in many developed countries. Each year in the US, 2 to 2.5 million cases of gastroenteritis are attributed to *Campylobacter* (Gadewar & Fasano, 2005; Olsen *et al.*, 2001; Poly & Guerry, 2008). If each case requires 2-3 sick days to be taken, 6 million work days could be lost annually. In the Netherlands, the cost associated with campylobacteriosis, including both health care costs and lost productivity, is an estimated 21 million Euros (Mangen, Havelaar, Bernsen, Koningsveld, & Wit, 2005). In Canada, there were 12,000 reported cases of campylobacteriosis during 2000. However, because of high rates of under reporting, it is thought that as many as 300,000 – 600,000 cases actually occurred in this year (Thomas *et al.*, 2006). Even the cost of diagnosing watery or bloody diarrhea caused by *Campylobacter* can be high. In the US, the estimated cost of a positive test result for *Campylobacter* is can be as much as $900 to $1000 (Gadewar & Fasano, 2005). Although in developed countries, diarrheal disease is not traditionally considered to be life-threatening, mortalities are also observed, primarily in infants, the elderly and previously ill patients (Altekruse, Stern, Fields, & Swerdlow, 1999). In the United States *Campylobacter* is estimated to be responsible for 124 deaths annually, 99 of which are food-borne acquired (Mead *et al.*, 2000). This accounts for 5.5% of bacterial food-borne

deaths, second only to *Listeria monocytogenes,* and non-typhoidal *Salmonella* (Mead *et al.*, 2000). In developed countries young children (less than three years of age) and infants are also at risk of acquiring campylobacteriosis, in particular those in contact with puppies and pet chickens (Tenkate & Stafford, 2001) although contact with adult dogs does not appear to be a risk factor for disease in this age group (Kapperud, Skjerve, Bean, Ostroff, & Lassen, 1992).

1.1.3 Overview of *Campylobacter* illness

Campylobacter infection begins with consumption of the bacteria through ingestion of contaminated food or water (Figure 1). The infectious dose for *Campylobacter* is low and is estimated to be 500-800 cells (Skirrow, 1997; Young, Davis, & DiRita, 2007). A correlation between ingested dose and infection rate exists where approximately 10% of individuals who ingest 800 or fewer cells will develop clinical symptoms of the disease (Thomas *et al.*, 1999). Volunteer studies have shown that consumption of even 500 cells can cause an infection (Robinson, 1981). The normal incubation period prior to illness associated with *C. jejuni* averages 2 to 3 days (usually 24-72 hours past ingestion of cells), but could be as short as one day or as long as a week. During this period the bacterium must survive passage through the acidic environment of the stomach (Blaser & Engberg, 2008). Chemotactic motility towards amino acids which are found in high concentrations in mucus coated regions of the gastrointestinal tract (Young *et al.*, 2007) may help direct the bacteria to the small intestine. Once in the small intestine the bacteria move through the mucus layer, aided by their spiral rod shape and motility, to colonize the jejunum and ileum (Blaser & Engberg, 2008). During

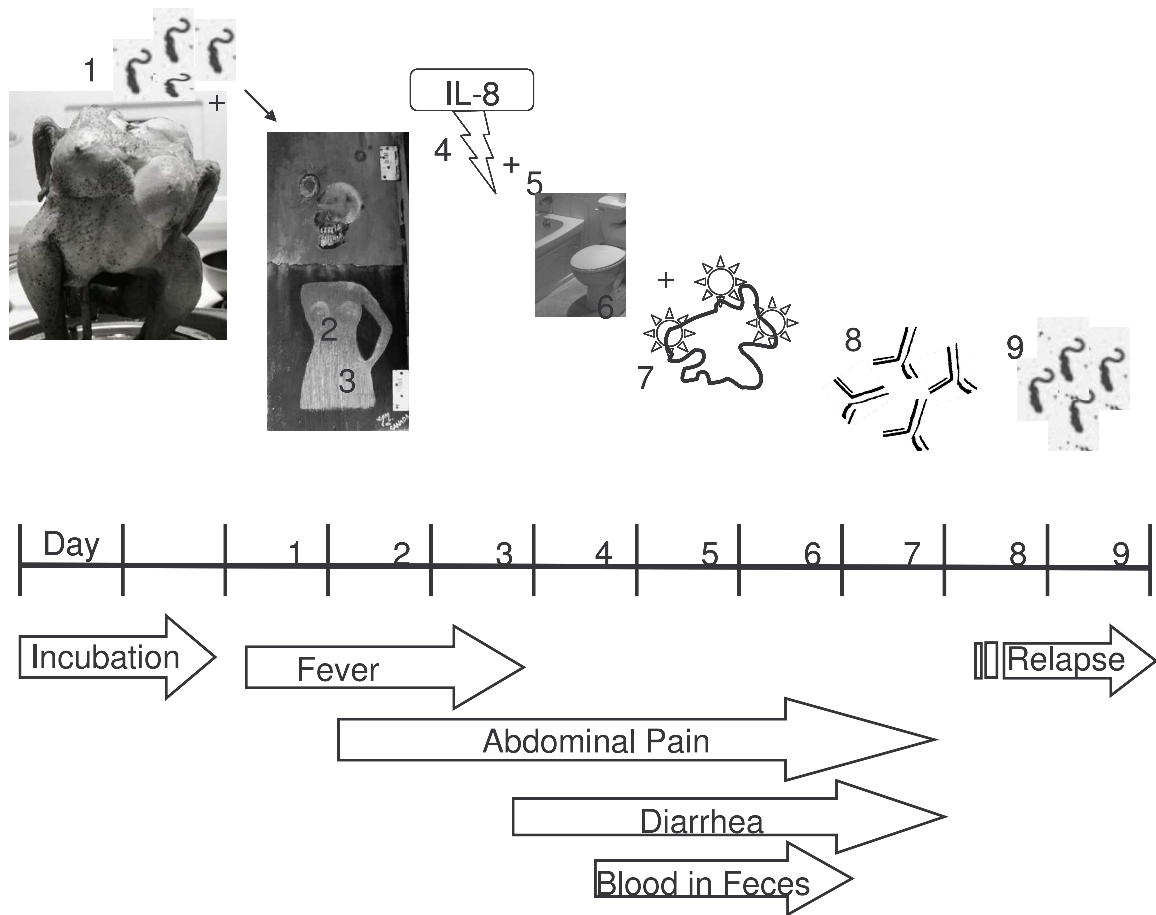


Figure 1: The route of infection of *C. jejuni.* 1) *C. jejuni* must enter the body, primarily by consumption of contaminated foods (for example, from a contaminated turkey) or by other means of ingestion. The normal incubation period before clinical symptoms appear is between 1 to 7 days, averaging between 2 or 3 days. 2) The bacterium must survive the acidic environment of the stomach. 3) After passing through the mucus layer coating the intestine, *C. jejuni* are able to colonize cells lining the jejunum and ileum of the small intestine. The spiral shape and motility of the bacterium aid in colonization. 4) The host begins a cell mediated immune response by production of interleukin-8 which begins a signaling cascade leading to production of other cytokines. 5) The first response by the host is fever, which is quickly followed by abdominal pain. 6) This stage of infection can be followed by acute and/or bloody diarrhea, especially in severe cases where the bacteria has been able to progress to the colon and rectum. 7) By the time diarrhea is observed, the host immune response is activated; macrophages, neutrophils and other cytokine growth factors can be found at the site of infection. 8) Approximately 5 days after onset of fever, antibodies can be found in the serum. These antibodies have little impact on the current infection but can offer some short-term protection against additional infections of a similar strain of *Campylobacter*. 9) In patients that have not been treated with antibiotics, *C. jejuni* can be shed in the feces for several weeks following infection. Relapses can occur, resulting in abdominal pain and diarrhea. Figure adapted from Skirrow (1997); Blaser and Engberg (2008). Images courtesy of Michael Maguire and Susan Ross.

colonization, the host's immune system begins to respond by starting a cascade of molecular events beginning with production of interleukin-8, a cytokine which helps stimulate a cell mediated immune response (Blaser & Engberg, 2008). At this point the patient will begin to feel ill and will display physical signs of illness including development of a high fever (Skirrow, 1997), which may be accompanied by fear and anxiety (Goehler, Lyte, & Gaykema, 2007). These symptoms are followed by moderate to severe abdominal pain which typically leads to acute and in some cases bloody diarrhea (Skirrow, 1997).

At this point the patient's immune response is fully activated; macrophages, neutrophils and other cytokine growth factors can be found at the site of infection. Other symptoms of *C. jejuni* infection may include anorexia, headache, nausea, chills and sweats, myalgia, vomiting and gross blood in the stools (Peterson, 1994). In some cases, the symptoms may resemble a dysentery-like syndrome which mimics inflammatory bowel disease (Walker *et al.*, 1986). Presence of blood in the stool usually signals a severe case of infection where the bacteria have been able to progress to the colon and rectum (Blaser & Engberg, 2008). Still, a majority of infections are self-limiting and can be treated by monitoring hydration and maintaining electrolyte balance; the ability to self treat most infections is likely responsible for the high number of unreported cases. Total duration of an infection is about 7 days where diarrheal symptoms are experienced and for most individuals, severe symptoms last about 3-5 days (Skirrow, 1997). By the end of an infection, IgA, IgM and IgG antibodies can be found in the serum of infected individuals. The ability of patients previously infected with *Campylobacter* to produce these antibodies may aid in protection against a second infection by the same or similar

strains of *C. jejuni* (Skirrow, 1997). This type of protection has been suggested to occur in young captive Macaque monkeys; however, these young monkeys might also possess maternal antibodies capable of preventing clinical disease (Dassanayake *et al.*, 2005).

In severe cases of *Campylobacter* infection, antibiotic treatment with macrolides and fluoroquinolones are often prescribed by heath care professionals (Keller & Perreten, 2006), although antibiotic therapy needs to be started as soon as symptoms appear in order to produce a dramatic reduction in the bacteria causing the disease to occur. Therapy is most important for individuals that are immune compromised; *e.g.*, the elderly and the very young. Except in patients that have been treated with antibiotics, *C. jejuni* will continue to be shed in the feces for several weeks following an infection (Kapperud *et al.*, 1992; Skirrow, 1997). *Campylobacter* cells shed in feces can contribute to human-to-human transmission via a fecal-oral route. Presence of the bacteria in some patients for long periods following an infection can lead to a relapse accompanied by both abdominal pain and diarrhea. Occasionally, a relapse also can be triggered in patients who consume a large quantity of food very quickly after symptoms of an infection subside (Blaser & Engberg, 2008).

1.1.3.1 Complications resulting from *Campylobacter* infections

While the majority of *Campylobacter* infections are considered to be self-limiting, and the host's immune system is able to clear the bacteria from the body without the need for medical intervention, serious, life-long complications can occur. Irritable bowel syndrome is associated with gastroenteritis, in particular TD (Janda & Abbott, 2006). Acute myocarditis, which may lead to heart failure, also can develop after infection by *C.*

jejuni (Mera, Lopez, & Serralta, 2007). Although the mechanism for causing damage to the myoctyes is not known, it has been hypothesized to occur either by a direct bacterial / host interaction or as a consequence of circulating toxins (Mera *et al.*, 2007).

Reiter's syndrome, also known as Reactive Arthritis (ReA), is trigged by gastrointestinal or genitourinary infections, and campylobacteriosis is no exception. ReA occurs in about 7% of people after infection with *Campylobacter* (Yates & Stetz, 2006), and often is seen in patients who have had TD. Onset occurs about 1- 4 weeks after the triggering infection. The symptoms of ReA have been shown to be extremely long lasting; *e.g.*, for up to 5 years in the majority of patients infected with *Salmonella enterica* serovar Typhimurium-triggered ReA (Yates & Stetz, 2006). This suggests that *Campylobacter*, the most common cause of TD, may be responsible for a significant proportion of ReA.

Campylobacter has been identified as a risk factor for post infection irritable bowel syndrome (PI-IBS) and inflammatory bowel disease. Up to 25% of patients who were diagnosed with *Campylobacter* enteritis will develop PI-IBS as gut changes persist long after infection (Spiller *et el.*, 2000).

Guillan-Barré Syndrome (GBS) and its variant, Miller Fisher Syndrome (MFS), both result from auto-immune responses which cause damage to nervous tissue, usually resulting in paralysis. Extreme cases of paralysis can compromise the function of respiratory muscles resulting in death (Sahin, Kobalka, & Zhang, 2003). The most commonly identified infectious agent associated with GBS is *Campylobacter jejuni* (Godschalk *et al.*, 2006); up to 30% of GBS cases and 20% of MFS cases are preceded by *C. jejuni* infection (Jacobs *et al.*, 1998; Koga *et al.*, 2005; Rees, Soudain, Gregson, &

Hughes, 1995). An estimated one in one thousand cases of gastroenteritis will develop some form of GBS (Allos, 2001; Poly & Guerry, 2008). It is believed that molecular mimicry between lipo-oligosaccharides found on the cell surface of *Campylobacter* and gangliosides found in human nervous tissue may lead to a cross-reactive immune response resulting in a damaged nervous system (Godschalk *et al.*, 2006; Poly & Guerry, 2008). Genes including *cgt*A, *neu*A1, *orf*11, Cj1421c and cj1428c/*fcl* all appear to contribute to the pathogenesis in cases of GBS and MFS as they often appear conserved in strains obtained from clinical cases which involve neuropathic complications (Taboada *et al.*, 2007). Differences in host background may also play a significant role in the expression of either GBS or MFS following *Campylobacter* infection, since many of these genes are also shared by strains isolated from cases of enteritis without complications (Taboada *et al.*, 2007).

1.1.4 *Campylobacter* Outbreaks

Despite a low infectious dose and its status as a common enteric pathogen, cases of *Campylobacter* infection are thought to be sporadic in nature (Pebody, Ryan, & Wall, 1997), only resulting in rare outbreaks of disease conditions which can be traced to discrete sources of an individual bacterial contaminant. This may be due to the sensitivity of *Campylobacter* to desiccation, heat, ultra-violet radiation and salt (Wagenaar, Mevius, & Havelaar, 2006), all of which work to limit concentrations of bacteria in food and water sources that can be distributed to the public. Identified risk factors for sporadic infection include consumption of contaminated produce such as poultry, eating at barbeques, and contact with animals, in particular dogs (Kapperud *et al.*, 1992).

In one outbreak of campylobacteriosis, individuals from New Zealand became ill when they drank spring water that appeared to be contaminated by runoff from adjacent farmland grazed by cattle and sheep following torrential rains (Centers for Disease Control and Prevention, 1991). Forty-four individuals developed symptoms of the disease, but after introduction of water chlorination no further cases were detected (Stehr-Green, Nicholls, McEwan, Payne, & Mitchell, 1991). In another outbreak, a water holding tank at a boys' school appeared to be contaminated by fecal material from birds (Palmer *et al.*, 1983; Walker *et al.*, 1986). Contamination of food by food handlers also has given rise to outbreaks, but these cases have been extremely rare (Olsen *et al.*, 2001).

In Canada our most notable outbreak occurred in 2000, when *Escherichia coli* O157:H7 and *C. jejuni* entered the public water supply for the community of Walkerton, Ontario, population 4800 (Hrudey, Payment, Huck, Gillham, & Hrudey, 2003). The pathogens likely entered a well used to supply the town with water through farm run-off which occurred immediately following a period of heavy rains. Many of the individuals who consumed the water became ill, resulting in approximately 2300 cases of infection, 67 hospitalizations, 27 cases of haemolytic uraemic syndrome (a potentially fatal kidney ailment) and seven deaths (Auld, MacIver, & Klaassen, 2004; Clark *et al.*, 2003).

1.1.5 Transmission and Prevalence of *C. jejuni*

Transmission of *Campylobacter* is multi-faceted with many potential routes (Figure 2). Poultry and bovine milk are well-established sources of human infectious *C. jejuni* cells, and provide a route for foodborne infection. Water also is thought to be an important source for the acquisition of *Campylobacter*, and may occasionally act as a

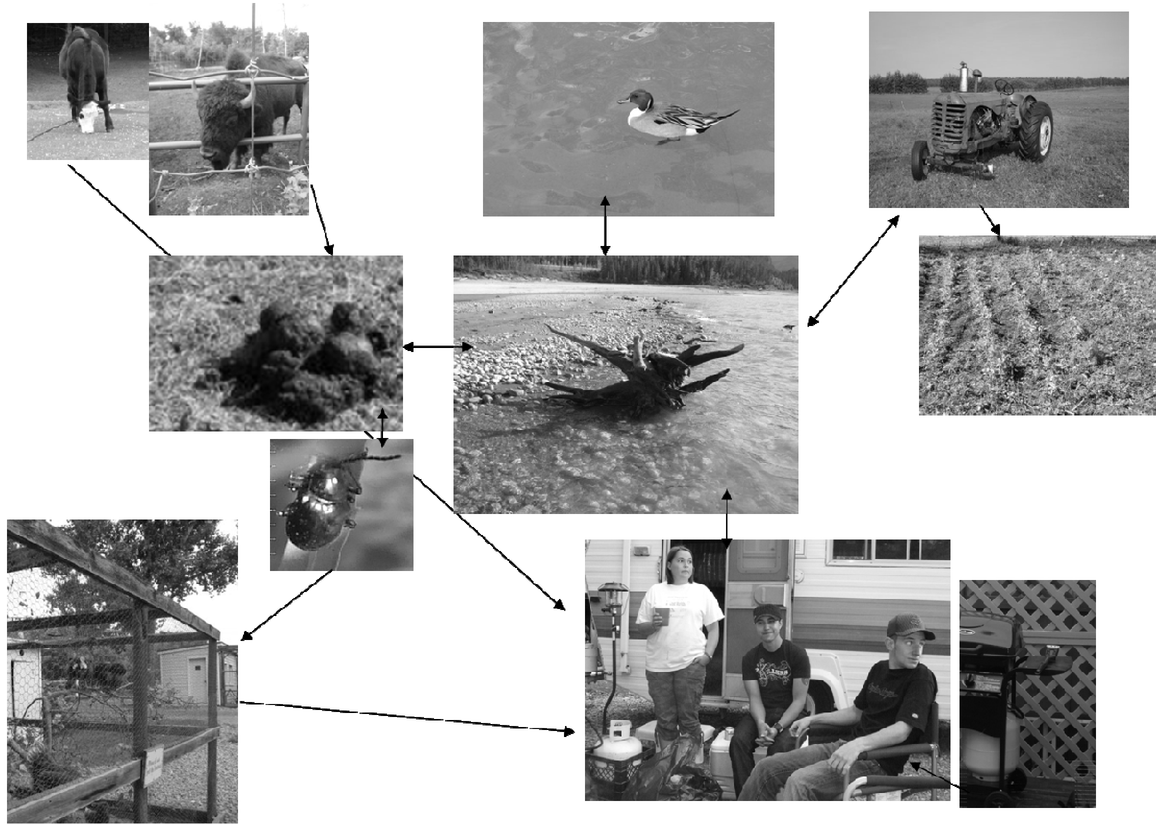


Figure 2: *Campylobacter* in the environment and routes of transmission. Dairy cattle have been associated with *Campylobacter* contamination and unpasteurized milk has been identified as a source of these infectious bacteria in humans (Levesque, Frost, Arbeit, & Michaud, 2008; Robinson, 1981). Livestock can act as carriers of *Campylobacter* and are known to shed it in feces (Açik & Çetinkaya, 2006; Inglis, Kalischuk, & Busz, 2004). Fecal material may enter water directly as run-off or may contaminate insects that spread the bacteria within the environment, possibly even to exposed food. Contaminated poultry, especially chickens can spread *Campylobacter* to humans as food. Migratory birds, such as geese can be contaminated by bacteria found in the water and then, can increase bacterial load in the environment via their own excretions (Aydin, Atabay, & Akan, 2001). Some birds may carry the bacteria into new geographical regions. Agriculture may contribute to the bacterial burden either through irrigation with contaminated water sources, or via use of improperly treated manure which can contaminate run-off. Consumption of raw or under-cooked contaminated meat products peaks during the barbeque season leading to an increase in human infections (Evans, Ribeiro, & Salmon, 2003; Kapperud *et al.*, 1992). Consumption of contaminated surface and well water (Stehr-Green *et al.*, 1991) also can result in campylobacteriosis. In addition, humans may impact the environment through improper hygienic practices and improper treatment of sewage. Images courtesy of Melissa Leece and Susan Ross. Note: importance of sources not presented in scale.

vector. Migratory birds may act as vectors, transporting bacteria to new environments. Insects such as flies may also act as vectors carrying bacterial cells to new environments such as chicken coops or directly on to food.

Use of contaminated water and manure in agriculture can increase the environmental bacterial load, leading to contamination of produce such as fruits and vegetables that generally are not cooked prior to human consumption. Farm animals and house hold pets, as well as large game animals found in the wild may become contaminated from naturally occurring bacteria in the environment and then act as carriers for *Campylobacter*. These animals can shed the bacteria in feces that can subsequently flow back into environmental reservoirs.

Water is seen as a potential source for the transmission of *Campylobacter* because of its relatively low infectious dose. Fecal contamination of water most likely occurs as a result of a combination of problems, including improper wastewater discharge, use of improperly treated waste or run-off from agricultural activities, and by direct contamination of the environment by wild animals, including migratory birds (Gannon *et al.*, 2005; Hyland *et al.*, 2003; Johnson *et al.*, 2003; Waldenstrom *et al.*, 2002). There is also a seasonal cycle related to numbers of *Campylobacter* found in aquatic systems; *i.e.*, bacterial contamination often rises to a peak in summer months after which levels tend to decline into the winter months (Thomas *et al.*, 1999). Because of the cost and difficulty associated with isolation methods required for *Campylobacter* detection from clinical samples, it has been suggested that *Campylobacter* may be a contributor to many more waterborne cases of infection where the disease agent has not be identified (Thomas *et al.*, 1999).

Mode of consumption is a key factor in the spread of *Campylobacter* to humans. Approximately 80% of *Campylobacter* related illnesses appear to be transferred through food (Logue, Sherwood, Elijah, Olah, & Dockter, 2003). *C. jejuni* is the most common *Campylobacter* isolated from clinical samples, with most of the remaining isolates being *C. coli*. Campylobacteriosis is usually attributed to improperly handled or undercooked poultry or livestock meat, unpasteurized milk, and contaminated water. An emerging source of foodborne illness is raw produce, as *Campylobacter* isolation rates range from 68%-97% from produce, with the majority of isolates being *C. jejuni* (Chai *et al.*, 2007).

C. jejuni is able to survive for 2 to 5 weeks in bovine milk stored at 4°*C* (Walker *et al.*, 1986) indicating the importance for use of pasteurization to process raw milk. Use of sterilization by hydrostatic pressure is already being used to preserve foods such as fruit juices (Martinez-Rodriguez & Mackey, 2005) and bacterial phage inactivation of meat contaminants prior to consumer purchase is being developed (Bigwood, Hudson, Billington, Carey-Smith, & Heinemann, 2008; Wagenaar, Van Bergen, Mueller, Wassenaar, & Carlton, 2005). In Iceland, freezing of chicken prior to consumer availability has limited case numbers of human infection (Reiersen *et al.*, 2002; Stern *et al.*, 2003).

1.1.5.1 *Campylobacter* bacteria in poultry and other avian sources

Campylobacter often is associated with poultry and consumption of contaminated poultry is deemed the main source of food borne illness, contributing to 20-40% of all cases of campylobacteriosis (Wagenaar *et al.*, 2006). Attempts to eliminate *Campylobacter* contamination from chickens combined with an extensive educational

campaign was tested in Iceland after identification of an epidemic of campylobacteriosis (Reiersen *et al.*, 2002; Stern *et al.*, 2003). As expected, implementation of a surveillance program aimed at eliminating *Campylobacter*, resulted in a drop in the incidence of campylobacteriosis to levels seen in other European countries (Reiersen *et al.*, 2002). Freezing of chickens prior to distribution also was implemented during peak periods of campylobacteriosis infection in Iceland; in combination with an educational program and enhanced on-farm biological security measures, these approaches appeared effective in reducing the number of domestic cases (Stern *et al.*, 2003). During a four-week period where the sale of chicken meat was prohibited in Belgium, the incidence of campylobacteriosis dropped by 40% from expected values, but returned to normal after the ban was lifted (Vellinga & Van Loock, 2002). A decrease in incidence of campylobacteriosis also was observed in the Netherlands after a reduction in poultry consumption (Van Pelt, Wannet, van de Giessen, Mevius, & van Duynhoven, 2004; Wagenaar *et al.*, 2006).

Isolation rates of *Campylobacter* species depends on the manner in which chickens are raised (*e.g.*, use of intensive or free-range rearing facilities) and the environment used for transport prior to arrival at the slaughter house. Prevalence of contamination is lowest in free-range chickens with approximately one-third of chickens carrying *C. jejuni* (Avrain *et al.*, 2003). A study in Quebec, Canada, found that chickens and turkeys raised on farms without rodent control, or with the flocks of birds located less than two hundred meters from a manure heap, were four to five times more likely to be colonized (Arsenault, Letellier, Quessy, Normand, & Boulianne, 2007). Most birds in an

infected flock will carry the organisms until slaughter, allowing for transmission of the bacteria to other members of the flock (Avrain *et al.*, 2003).

It is currently unknown how poultry destined for food production are contaminated by *C. jejuni*. Vertical transmission to chicken eggs can occur in a primary fashion though the reproductive tract of the hen, by the negative pressure produced by egg cooling and contracting which draws the bacteria in though the eggshell, or alternatively through environmental contamination of chicks and adults by fecal shedding (Sahin *et al.*, 2003), although this idea is controversial. *Campylobacter* is unlikely to penetrate the eggshell or remain viable within the open environment; a previous study done at a group of broiler farms has shown that *Campylobacter* cannot be isolated from eggs or chicks less than 2 days old, even though most hens were shedding the bacteria (Sahin *et al.*, 2003). There has been little evidence to support the suggestion that there is linear transmission from hen to chick of *Campylobacter*, a possible hypothesis considering the almost ubiquitous colonization of *Campylobacter* bacteria in chickens (Callicott *et al.*, 2006).

Campylobacter colonizes the gut of avian species, and likely contaminates the skin of birds during slaughter and processing. In chickens, the cecum can be colonized by up to 10^{10} organisms per gram of cecal content without causing symptoms of disease, and is found on as many as 90% of raw chicken carcasses (Evans, 1997; Wassenaar, Geilhausen, & Newell, 1998). Colonization is not limited to the gut, and contamination of other organs such as the liver may occur by migration of the bacteria from the intestine through tubules such as the bile duct or into intestinal tissues through breaks in epithelial layers of the intestine (Whyte, Hudson, & Graham, 2006). Risk of *Campylobacter*

infection in chickens increases with age of the chicken before slaughter (Arsenault *et al.*, 2007; Barrios *et al.*, 2006); these infections may occur independent of *Campylobacter* contamination acquired at the slaughter house. During transport to slaughter, shipping crates used to house the birds are not always thoroughly cleaned, leaving fecal contamination or debris from the feet and feathers (Hansson, Ederoth, Andersson, Vagsholm, & Olsson Engvall, 2005). A majority of crates (57%) that are washed and santized continue to test positive for the presence of thermophilic *Campylobacter*. This persistence may result in transmission of the bacteria to humans by consumption of contaminated meat when poor slaughter-house conditions and/or handling of the chicken meat has taken place.

As well as being prevalent in chickens, *Campylobacter* can be isolated from a large portion of turkey carcasses prior to slaughter (Logue *et al.*, 2003). In addition, *C. jejuni* has been isolated from migratory birds such as geese (Aydin *et al.*, 2001). Mean prevalence of *Campylobacter* in migratory birds tested in Sweden was 21.6%, with the prevalence varying from 0% to 100% depending on the species examined (Waldenstrom *et al.*, 2002). Raptors, opportunistic feeders and ground-foraging birds were found more likely to be contaminated with *Campylobacter* (Waldenstrom *et al.*, 2002).

1.1.5.2 *Campylobacter* and swine

Transmission of *Campylobacter* from pig to human is thought to occur mostly by contaminated pork products and has been associated with the consumption of products that may have been contaminated at slaughter. Two sources of contamination appear likely as a majority of bacteria are found on the skin surface and within the GI tract.

Additionally, animal organs may also contain *Campylobacter*s including liver (Alexandrina & Botus, 2008). It is unclear if these organs are infected during the life time of the animal, or contamination from the skin surface or GI tract during slaughter as many *Campylobacter* species are isolated from fecal material.

It is believed that *C. coli* is a commensal bacteria in the intestinal tract of pigs that can be pathogenic to humans (Alter *et al.*, 2005). Although piglets are not born colonized with *Campylobacter*, acquisition is rapid. The majority of piglets are colonized with *Campylobacter* within hours after birth. The primary source of contamination is thought to be maternal in origin where transmission of *Campylobacter* is from mother to piglet (Alter *et al.*, 2005). The majority of farm raised pigs test positive for *C. coli* in any age group. For piglets, a dramatic increase in number and strain diversity is found in herds both before and after transportation to a nursery. Because these piglets are no longer colonized with the maternal strain alternative sources of contamination must exist (Alter *et al.*, 2005). This increase in acquired strain diversity may be related to age, as it has been suggested that after 3 weeks of age, new infections can increase by 50% per sampling period up until approximately 12 weeks of age (Alter *et al.*, 2005). Typically, by the age of fattening, the Campylobacter genotype pool present in pigs has stabilized, and tends to remain constant until slaughter (Alter *et al.*, 2005).

Another study reported that 100% of organic piglets between 8 and 13 weeks of age excreted *Campylobacter* (Jensen, Dalsgarrd, Baggesen, & Nielson, 2006). However, there were temporal differences in the percentage of pigs that were *C. jejuni* positive; *i.e.*, contamination rates ranged from 0%-78.6% in samples taken from April to September (Jensen *et al.*, 2006). A high incidence in the number of contaminated young organic pigs

also was observed compared with conventionally raised pigs (Alter *et al.*, 2005). *C. jejuni* also persisted longer in the organic pigs. This may have been due to greater environmental exposure and the challenge of maintaining hygienic practices within an outdoor paddock. As well as the potential of spread between pigs, additional contamination can result from environmental sources, which may introduce new strains into swine herds (Jensen *et al.*, 2006).

Different porcine strains of *Campylobacter* have been shown to undergo genetic recombination. These recombinants were able to infect other species such as chickens (Ziprin, Hume, Young, & Harvey, 2002). Thus a pig-to-chicken species transfer represents a potential route for inter-species transfer of the bacteria. Such low host specificity would allow for potential contamination and spread of *Campylobacter* through environmental contact.

1.1.5.3 Prevalence in cattle

The possibility that cattle represent a significant reservoir for *Campylobacter* has been raised by work showing that adult cattle showed a consistent presence of *Campylobacter*; *i.e.*, 89.6% of the samples contained *Campylobacter*, and 66.1% shed the bacteria over all sampling times (Inglis *et al.*, 2004). Excretion of *Campylobacter* has also been shown to increase in cattle after transport, and when the animal may be stress, such as before slaughter (Humphrey, O'Brien, & Madsen, 2007). Albeit some studies show the majority of cattle shed *C. lanienae*, *C. jejuni* has been attributed to contamination of milk (Levesque *et al.*, 2008; Robinson, 1981). This requires the application of meat control later in the food processing system, including improved

hygiene during milking and slaughter, and the pasteurization of milk (Humphrey *et al.*, 2007). Although associated with *Campylobacter,* cattle has largely been ignored as carrying potential human pathogens and little is known about the consequences of contact with these strains, although many have been isolated from healthy abattoir workers in Switzerland (Inglis *et al.*, 2004). High prevalence of *Campylobacter* shedding in cattle and cattle-associated strains found in humans suggests that *Campylobacter* also may be spread from cattle to humans, and may be able to cause disease.

1.1.5.4 Prevalence in other animals

Campylobacter has been isolated from a large number of sheep (49.5%) that did not display clinical symptoms of a disease; isolation rates for *Campylobacter* were highest within the intestine suggesting that sheep may cause significant contamination to the environment, and are potential sources for introduction of the bacteria into the human food chain (Açik & Çetinkaya, 2006). Lambs can be colonized by the bacteria; this occurs most often where continuous breeding is being used, which allows fecal-oral spread of the bacteria to the young of various ages after accumulation of pathogens in the breeding/lambing pen (Andrés *et al.*, 2007). This type of evidence suggests that *Campylobacter* strains that are able to colonize ovines could be more widely distributed in the environment than previously thought (Diker, Esendal, & Akan, 2000).

Campylobacter has been found in farm and large game animals (Gill, 2007; Humphrey & Jorgensen, 2006). In addition, *Campylobacter* is often isolated from flies and household pets including cats and dogs (Chai *et al.*, 2007; Evans *et al.*, 2003; Gannon

et al., 2004; Humphrey *et al.*, 2007; Johnson *et al.*, 2003; Koene, Houwers, Dijikstra, Duim, & Wagenaar, 2008).

C. jejuni has been isolated from a large number of zoological sources including primates, felids, red panda, hoofed animals, birds and reptile suggesting the ubiquitous nature of *Campylobacter* (Misawa *et al.*, 2000). Transmission is suggested to occur by host-to-host, and fecal-to-oral contact or by ingestion of contaminated of water (Misawa *et al.*, 2000). Many aspects of the epidemiology are not well understood and bacterial source tracking, identification of bacterial reservoirs, and identification of potential routes for human infection with *C. jejuni* and other *Campylobacter* spp. will require a significant effort to produce a complete picture of the molecular epidemiology of *Campylobacter*.

1.2 Molecular Genetics, Typing Methods, and *Campylobacter*

1.2.1 Introduction to molecular sub-typing

Sub-species typing methods play an important role in the classification of strains, surveillance to identify trends in pathogen transmission which may incite disease, outbreak detection, and monitoring of prevention/control efforts. *Campylobacter* is no exception and significant efforts have been targeted at the development of sub-typing methods for use in molecular epidemiology.

1.2.2 Phenotypic sub-typing methods

Although most *Campylobacter* isolation protocols require biochemical screening to confirm *Campylobacter* spp., species subtyping methods based on phenotypic traits are few. Among the more prominent and accepted phenotypic typing methods are the serotyping methods based on the Penner scheme of heat-stable antigens, and the Lior scheme of heat-labile antigens. The Penner serogrouping scheme recognizes 65 serogroups, originally believed to be heat-stable lipopolysaccharide O antigens (Penner & Hennessy, 1980), but which have subsequently been found to be capsular polysaccharide antigens (Karlyshev, Linton, Gregson, Lastovica, & Wren, 2000). The Lior serotyping scheme is based on heat-labile surface and flagellar protein antigens (Lior, Woodward, Edgar, Laroche, & Gill, 1982) which rarely produce the cross-reactivity seen with the Penner sero-grouping scheme, thus reducing the difficulty of interpretation (Patton, M., Barrett, & Morris, 1985). Both of these methods have been used to track epidemiological trends and study species diversity (Skirrow, 1997; Woodward & Rodgers, 2002), and have provided an effective and practical early approach to the identification of pathogenic strains (Patton *et al.*, 1991) before the development of DNA-based typing methods. For additional typing resolution, both schemes can be combined (Patton *et al.*, 1985), and when used together, have been successful in establishing epidemiological links after outbreaks have occurred (Patton *et al.*, 1991; Woodward & Rodgers, 2002).

1.2.3 Molecular-based Methods of subtyping strains beyond the Genus-Species level and application to *Campylobacter jejuni*

1.2.3.1 An introduction to molecular methods for subtyping *Campylobacter*

Typing of *Campylobacter* is a dynamic field with older methods continually being advanced and new methodologies constantly being developed. These methods vary in through-put, cost, resolution, and ease of use/analysis. The optimal method depends on the purpose of the study; *i.e.*, whether or not a test is being used for diagnostic confirmation, short-term epidemiological study, long- term epidemiological study, or determination of genotype/phenotype relationships.

The general trend in the bacterial typing community has been to move away from phenotypic approaches towards DNA-based genotypic subtyping methods. Some problems associated with phenotypic typing methods include cost, large numbers of untypeable strains and instability of phenotypic markers under certain environmental and culture conditions (Farber, 1996). DNA-based typing methods are preferred as all bacterial isolates should give consistent results regardless of culture conditions (Farber, 1996).

In general, molecular typing methods require a combination of two elements: genetic loci that bear regions of high conservation within a specific group and that can be targeted in all isolates (*i.e.*, several methods require a PCR amplification step, and this ensures that the genetic marker(s) can be amplified in all isolates) and loci that also encompass highly variable regions that will form the basis for differentiation among isolates. Methods of genotypic subtyping for *Campylobacter* include, but are not limited to, pulsed-field gel electrophoresis, ribotyping, random amplified polymorphic DNA analysis, *fla*A short variable region sequencing or restriction fragment length polymorphism, and multi-locus sequence typing (described in the sections below).

1.2.3.1.1 Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis (PFGE) is based on analysis of macro-restriction profiles obtained by digestion of genomic DNA with a rare-cutting endonuclease to produce distinct banding patterns of large fragments on gels that are resolved using a specialized electrophoresis unit capable of generating electrical fields that alternate in direction (Finney, 1993).

PFGE has been used successfully in the identification of outbreaks associated with many bacterial pathogens (Fitzgerald *et al.*, 2001; Olsen *et al.*, 2001). This method was used to discriminate between bacterial isolates obtained from a school outbreak of campylobacteriosis and those from sporadic community cases that occurred during the same time period. In combination with available epidemiological data, PFGE was able to provide the genetic evidence needed to confirm that strains of the pathogen isolated from the outbreak were related (Olsen *et al.*, 2001). PFGE was also determined to be critical to understanding an outbreak that occurred in Kansas; it was shown to provide adequate discrimination between unrelated isolates of *Campylobacter*, greater than that of serotyping and *flaA*-SVR typing (Fitzgerald *et al.*, 2001). However, differences in insertion sites of bacteriophages by acquisition, and loss or transposition can result in PFGE profiles which obscure the true relationship among strains (Barton, Ng, Tyler, & Clark, 2007).

One of the major challenges of PFGE has been the standardization of analytical protocols to enable inter-laboratory comparisons. Analysis of banding patterns typically requires the use of specialized software capable of comparing the macro-restriction

fingerprints. Online comparison between laboratories has been made possible with the development of the PulseNet system (Swaminathan, Barrett, Hunter, & Tauxe, 2001). This is one of the few database systems that provides both epidemiological data and typing data over widespread geographical regions (Maiden, 2006). Methods and protocols have been standardized, and the network has been successful at detecting disease outbreaks, particularly using PFGE (Swaminathan *et al.*, 2001). However, this network has only been shown to be effective in bacterial populations where the PFGE patterns are relatively stable (Maiden, 2006), which appears to have been a problem in the case of *C. jejuni* (Barton *et al.*, 2007; Clark *et al.*, 2007)

1.2.3.1.2 Ribotyping

The ribosomal genes for 5S, 16S and 23S rRNAs are located in several locations in the *C. jejuni* genome. These genes are highly conserved, but there can be significant variation in the surrounding regions. They thus represent suitable targets for use in subtyping.

The traditional method of ribotyping begins with enzymatic digestion of genomic DNA with *Pst*I, *Hae*III, *Hind*III and *Pvu*I in a single, double or triple digest followed by Southern blot hybridization with a probe specific derived from RNA genes (Cox, Stern, Hiett, & Berrang, 2002; Fayos, Owen, Desai, & Hernandez, 1992; Fitzgerald, Owen, & Stanley, 1996; Gibson, Fitzgerald, & Owen, 1995; Manfreda, De Cesare, Bondioli, & Franchini, 2003; Waldenstrom, On, Ottvall, Hasselquist, & Olsen, 2007; Ziprin, Sheffield, Hume, Drinnon, & Harvey, 2003). Digested DNA is run on an agarose gel, then hybridized to the rDNA probes by Southern blot hybridization followed by

comparison of banding fragments. This method is effective for determining the species of *Campylobacter* (Denes, Lutze-Wallace, Cormier, & Manzano, 1995), but tends to lack the resolution needed to define specific subtypes of the organism (Kiehlbauch, Plikaytis, Swaminathan, Cameron, & Wachsmuth, 1991). The lack of consistency within a laboratory and difficulty in reproducing methods between laboratories, and conflicting evidence of distinguishing *C. coli* from *C. jejuni* suggests that this method is unsuitable for discrimination and subtyping of *C. jejuni.* Automated systems have been designed for ribotyping but they are costly and of low throughput, factors which further limit the use of this typing method (Wassenaar & Newell, 2000).

1.2.3.1.3 Random amplified polymorphic DNA

Random amplified polymorphic DNA (RAPD) is based on PCR amplification of genomic DNA under low stringency conditions (*i.e.*, low annealing temperatures) using arbitrarily assigned 10-mer primers. Strain RAPD banding patterns are then compared after gel electrophoresis (Açik & Çetinkaya, 2006; Ertas, Cetinkaya, Muz, & Ongor, 2004; Fujimoto, Allos, Misawa, Patton, & Blaser, 1997; Ono, Kurazono, Niwa, & Itoh, 2003; Payne, Lee, Dreesen, & Barnhart, 1999).

The primers usually allow for some level of mismatch between primer and target, reducing the need to target identical, highly conserved DNA sequences. It is this allowance for mismatch that sometimes results in a combination of weak and faint bands, a common complication of the technique, which can lead to difficulties in reproducing results from run to run and between laboratories (Wassenaar & Newell, 2000). RAPD is less discriminatory than PFGE and ribotyping (Eriksson *et al.*, 2005) and a high

percentage of strains must be categorized as untypeable. These disadvantages negate the low cost, simplicity, and throughput of the method.

A similar genotyping method is amplified fragment length polymorphism (AFLP) which is based on the selective amplification of small restriction fragments (50 to 500 bp) of chromosomal DNA to represent the entire genome (Duim, Wassenaar, Rigter, &Wagenaar, 1999). AFLP has been successfully applied to the differentiation of *Campylobacter* (Duim, Wassenaar, Rigter, &Wagenaar, 1999; Duim *et al.,* 2001). Although this method is highly discriminatory, additional genotyping methods are required for optimal typing of *Campylobacter jejuni* (Boer *et al.,* 2000) requiring additional expertise and expense for effective genotyping.

1.2.3.1.4 Analysis of the *fla*A allele (also known as *fla* typing)

C. jejuni contains two flagellin genes, *fla*A and *fla*B. The genes consist of both conserved and variable sequence regions. These genes are also found in other pathogens and may be valuable for typing of closely related pathogens (Wassenaar & Newell, 2000). Generally the first of the two genes, *fla*A, has been targeted for *C. jejuni fla* typing.

There are several competing protocols used in *fla* typing, which differ by DNA preparation method, primer design and PCR target location, PCR temperatures and time, restriction enzymes used, and the nomenclature used to describe the typing results (Wassenaar & Newell, 2000). As of 2000, seven different *fla* typing procedures were available, however, the basis of procedure remains the same; *i.e.*, extracted DNA is amplified by primers within (or flanking) the *fla*A allele, so that the PCR amplicon

(typically 0.7kb to 1.7kb) can be purified and digested with restriction enzymes, followed by gel electrophoresis. Differences in *fla*A alleles between strains result in restriction fragment length polymorphisms (RFLP) and, when fragments are visualized using gel electrophoresis, RFLP profiles can be compared and used to assess their similarity. Restriction enzymes that have been used in the literature include *Alu*I, *Dde*I, *Hin*fI, *Eco*RI and *Pst*I (Duim, Wassenaar, Rigter, & Wagenaar, 1999; Wassenaar & Newell, 2000).

A more detailed analysis of the *fla*A gene, *flaA*-short variable region (SVR) typing, can be obtained by sequence comparison of the *fla*A SVR, a segment of 321 bp corresponding to positions 283 to 603 of the *fla*A gene that is highly polymorphic (Dingle *et al.*, 2001). The *fla*A gene is amplified by PCR, purified and sequenced. The sequences of the alleles from different strains can be compared and used to measure the relatedness of genes (and by extension strains) using standard phylogenetic analysis. However, *fla* typing relies on a single locus and is ideally suited for clonal bacteria, but may be problematic when applied to non-clonal or weakly clonal species such as *C. jejuni* (Terletski, Brenner, & Schwarz, 2004) since the *fla*A allele may not necessarily be related to gene content at other areas of the genome. Over-reliance on this locus for the comparative analysis of strains can result in a skewed picture of the "true", whole genome, genetic relationship between strains.

1.2.3.1.5 Multi-Locus Sequence Typing

Multi-Locus Sequence Typing (MLST) is the leading molecular typing method for *Campylobacter* at this time. MLST was proposed in 1998 as a universal and definitive method for characterizing bacterial populations (Dingle *et al.*, 2001; Maiden, 2006) and

has been used as an established technique for characterization of *C. jejuni* for several years (Dingle *et al.*, 2001).

The typing scheme is based on PCR amplification of a 350-600 bp region of seven highly conserved house keeping genes (*asp*A, *gln*A, *glt*A, *gyl*A, *pgm*, *tkt* and *unc*A), the sequencing of the fragments, followed by comparison of their nucleotide sequences using standard phylogenetic analysis (Maiden, 2006). A database of published MLST sequence types is available at http://pubmlst.org/campylobacter/ . As typing data are nucleotide based, MLST can be easily compared between laboratories regardless of the methods used to generate the allelic sequence. Also, because MLST data can be exchanged freely over the internet, this allows for widespread comparison of MLST sequence types and can be used for population studies and evolutionary biology (Maiden, 2006).

Outbreaks have been identified or confirmed using MLST, thus proving that it has practical use in some epidemiological applications (Sails, Swaminathan, & Fields, 2003). Some alleles have been identified that appear to be species-specific (Wagenaar *et al.*, 2006) which suggests that an application for source tracking may also be possible. MLST has been shown to be better correlated with epidemiological information than *fla*A-SVR (Clark *et al.*, 2005) and its resolution is high allowing subtypes of *C. jejuni* to be differentiated (Parker, Miller, Horn, & Lastovica, 2007).

Although more loci are evaluated than with other typing methods, recent work has shown that the seven loci used may not be sufficient to provide an accurate picture of gene content in all areas of the *C. jejuni* genome (Taboada *et al.*, 2008). Additionally, MLST is unable to distinguish closely related strains in short-term outbreak

investigations and additional methods, such as *fla* typing, may be required in order to obtain sufficient resolution (Sails *et al.*, 2003) increasing the cost and time to acquire results adequate for detailed molecular epidemiological analysis.

1.2.3.2 Current challenges in subtyping of *Campylobacter*

The general consensus is that phenotypic methods of subtyping *Campylobacter* lack adequate discrimination and reproducibility (Steele *et al.*, 1998). Serotyping and biotyping only target specific and potentially transient phenotypic properties of an isolate. This is important for clinical purposes when a diagnosis is required for immediate treatment, but does not address the underlying genetic composition. Additionally, these methods require maintenance of a panel of costly antisera, and may give ambiguous results due to non-typable strains and cross-reactivity between antigens (Fitzgerald *et al.*, 2001). These challenges have led to a shift towards molecular and genotypic subtyping methods.

Although band-based genotyping methods (*i.e.*, methods that rely on the comparison of electrophoretic fingerprints) such as PFGE, RFLP, RAPDs have proved to be useful in some contexts, analysis of fingerprint banding patterns are limited in their application to population studies and evolutionary studies as one of the prime analysis assumptions is that the multiple bands are unrelated (*i.e.,* bands between different fingerprints that do not align within reason are not the same) which may not account for genetic rearrangement such as SNPs, or insertion/deletion events (Maiden, 2006). Additionally such methods can be difficult to reproduce accurately for each sample, such that large-scale application and inter-laboratory data comparisons require specialized and

expensive software for band analysis and fingerprint comparison (Maiden, 2006). MLST can remedy problems associated with reproducibility, data portability (Levesque *et al.*, 2008) but remains expensive and slow, even with the increasing speed and reduced cost of nucleotide sequencing (Maiden, 2006). Although MLST alone can generate relevant data for deciphering epidemiological relationships among isolates, it usually requires a secondary method, such as *fla*A-SVR, for optimal discrimination (Clark *et al.*, 2005), adding to the cost and reducing throughput of the method.

Among the major challenges with current genotyping methods for *C. jejuni* is that they do not adequately account for processes that result in genome plasticity. *Campylobacter* is naturally competent and can acquire genetic material "laterally" through uptake of DNA from the environment (Wang & Taylor, 1990). This process can contribute clinically relevant genetic material such as virulence genes or antimicrobial resistance genes to strains. Chromosomal rearrangements are also common in *Campylobacter*, which has limited the usefulness of PFGE as a typing method (Sails *et al.*, 2003). Most molecular typing methods are designed for use with clonal bacteria, however, their use with non-clonal or weakly clonal species such as *C. jejuni* is problematic (Terletski *et al.*, 2004). This genomic plasticity causes problems for methods that rely on the evaluation of one or a few loci, for example MLST and *flaA* typing, since whole-genome comparative genomic analysis has revealed that information on a small number of genetic loci does not necessarily allow for inferences to be made on the gene content at other areas of the genome (Taboada *et al.*, 2008). For example, strains that are similar at specific typing loci may still harbor significant genomic differences

that could account for differences in pathogenicity. Conversely, strains that differ at the typing loci may have a similar complement of virulence factors.

1.2.4 Development of next-generation molecular typing methods

1.2.4.1 Evolution of bacterial genomes

A significant portion of genetic change in prokaryotes occurs via horizontal gene transfer (HGT). HGT, which may also be referred to as lateral gene transfer, leads to genetic exchange between members of the same species or between different bacterial species through DNA uptake and subsequent incorporation of the genetic material by a method such as homologous recombination (Hacker & Carniel, 2001). When HGT involves DNA encoding an entire operon, novel functions and metabolic abilities can be acquired in a single event (Lawrence, 1999). Such events can lead to an increase in metabolic diversity of recipient bacteria, or confer the ability to explore a greater variety of ecological niches (Marri, Hao, & Golding, 2007). Perhaps the most interesting aspect of HGT, at least from a clinical and epidemiological point of view, is the acquisition of virulence factors or antibiotic resistance genes as in the step-wise acquisition of virulence factors such as gene encoding shiga toxins in *Escherichia coli* (Boerlin, 1999).

When HGT is combined with frequent point mutations, high recombination rates, high rearrangement rates, and gene silencing or gene loss, the speed at which bacterial genomes become diversified is rapidly increased. Because of this high rate of genome evolution, new challenges are presented in the analysis and interpretation of molecular typing and phylogenetic data due to the rapid erosion of genetic similarities or rapid

increase of differences between clonally related strains, reducing the extent to which such relationships can be firmly established with current technologies. In particular, because of HGT, not all genes in an organism's genome will share the same evolutionary history, which presents a challenge in phylogenetic analysis (Olsen & Woese, 1997).

1.2.4.2 Genomics of *C. jejuni*

The first *C. jejuni* strain had its genome sequence published in 2000. *C. jejuni* NCTC 11168, a human clinical isolate, has a circular genome of 1,641,481 bp that is predicted to encode 1,654 proteins and 54 stable RNA species (Parkhill *et al.*, 2000). In contrast to most other bacteria, there appear to be no phage-associated sequences, and few insertion sequences or repeat sequences (Parkhill *et al.*, 2000).

The genomes of three additional *C. jejuni* strains have been completely sequenced since the publication of the NCTC 11168 genome sequence; *i.e.*, RM1221, CG8468 and 81-176 (Fouts *et al.*, 2005; Hofreuter *et al.*, 2006; Poly *et al.*, 2007). Strain RM1221, a strain isolated from a chicken carcass, has a slightly larger gene content than NCTC 11168 (1,884 ORFs vs. 1,634 ORFs) and both strains have similar genome organization, although the RM1221 genome is interrupted by inserted prophage/genomic islands (Fouts *et al.*, 2005). Strains CG8486 (Poly *et al.*, 2007) and 81-176 (Hofreuter *et al.*, 2006) were isolated from diarrheal patients and their genome sequences reveal that they are similar in size to NCTC 11168. The genomes have a similar organization to the other two genomes, with the bulk of genomic variability residing within hypervariable carbohydrate biosynthetic loci and a limited number of variable genes mapping outside these. To date,

an additional eight strains have whole-genome shotgun sequencing data available at NCBI, with more expected to come in the very near future.

1.2.5 Comparative genomics of *C. jejuni*

Comparative genomics, the analysis of and comparison of two or more genomes, has not only served to uncover the large amount of within-species genomic diversity and the rapid pace of evolution of bacterial genomes, but it has also served to underscore some of the new challenges in bacterial genotyping and phylogenetic analysis.

C. jejuni has approximately 1300 core genes, house-keeping genes that encode functions required for survival, as determined by comparative genomic sequencing (Fouts *et al.*, 2005) and by microarray-based comparative genomic hybridization analysis (Dorrell *et al.*, 2001; Pearson *et al.*, 2003; Taboada *et al.*, 2004). This accounts for about 60-70% of the entire *C. jejuni* genome with the remaining genes forming part of an "accessory" pool of genes that are variably present or absent in different strains. As a result, genome content varies between strains of *C. jejuni*, with a select number of genetic regions in the *C. jejuni* genome being highly plastic (Pearson *et al.*, 2003; Taboada *et al.*, 2004), leading to the suggestion that recombinogenic processes such as HGT play a dominant role in *C. jejuni* genome evolution.

1.2.6 Comparative genomics in understanding pathogenicity

It has been suggested that genomic differences exist between pathogenic and non-pathogenic *Campylobacter* spp. (Humphrey *et al.*, 2007). Comparative genomics can yield practical data for assessing genetic differences among strains that could account for

differences in their pathogenicity. For example, a large number of strains can be collected that are isolated from clinical samples (*i.e.*, pathogenic strains) and compared at the whole genome level to strains isolated from food or environmental sources to identify genes associated with pathogenic strains and which may help contribute to that organism's ability to cause human illness.

Comparative genomic sequencing of the clinical isolate 81-176 (Bacon *et al.*, 2000) has revealed some genomic elements that may contribute to virulence. Most strains of *C. jejuni* carry an ~35 kb plasmid, pVir, which is believed to contribute to virulence (Fouts *et al.*, 2005). Some of the genes encoded by this plasmid appear to be similar to those of the *cag* pathway, which is located within a pathogenicity island found in *Helicobacter pylori*. Genes from the *cag* pathway are involved in type IV secretion, and are associated with virulence in *H. pylori* (Bacon *et al.*, 2000). Although the exact mechanism is not known, the pVir plasmid has been shown to be associated with bloody stools and diarrhea (Tracz *et al.*, 2005). As bloody stools serve as an indicator of severe invasive infection, it may suggest that the presence of pVir could be used as a predictor for virulence (Tracz *et al.*, 2005). Similar in size to pVir, pTet carries *tet*O which encodes resistance to tetracycline (Bacon *et al.*, 2000) and the plasmid may also play a role in pathogenicity. In pVir, mutation of the *com*B3 operon encoded on this plasmid dramatically impairs the natural transformation ability of *C. jejuni* 81-176 suggesting that the encoded proteins enhance natural transformation, and promote DNA binding, perhaps through surface changes (Bacon *et al.*, 2000). This may be advantageous for pathogenic strains as they could acquire genes specific to host environmental conditions.

Analysis of strains 81-176 (Hofreuter *et al.*, 2006) and CG8486 (Poly *et al.*, 2007), which induce inflammatory diarrhea in humans and have been shown to be virulent in tissue culture and in the ferret model, has been used to examine additional genomic features that may contribute to *C. jejuni* virulence. Both strains have a highly similar gene content and share genes that are absent from other sequenced *C. jejuni* strains. This suggests that the genes required for inflammatory diarrhea are shared by these strains. Subtle differences between the two strains in their ability to invade epithelial cells *in vitro* are likely to be due to changes in carbohydrate loci and/or more subtle changes in other genes, which may modulate virulence (Poly *et al.*, 2007).

Taboada *et al.* (2008) have recently shown that highly clonal bacterial isolates that would be indistinguishable by MLST nevertheless harbour significant genomic differences between them. This suggests that while the analysis of a small number of genetic loci could prove useful for establishing strain to strain relationships, MLST may fail to elucidate potential genetic differences between strains that could be of clinical importance. Analysis of strains by comparative genomics goes beyond current methods of strain comparison and also attempts to account for HGT in bacteria. Comparative genomic sequencing and microarray-based comparative genomic hybridization have been used in the analysis of diverse pathogenic and non-pathogenic strains. This has led to identification of putative virulence-associated markers that can be used in the analysis of additional strains in order to confirm such associations (Bolton, Hutchinson & Parker, 1987; Fayos, Owen, Desai & Hernandez, 1992; Johnson *et al.,* 2003; Olsen *et al.,* 2001). Despite the usefulness of the comparative genomics approach and recent advances in high-throughput genomic sequencing, comparative genomic analysis remains slow, costly

and labour intensive. As a result, comparative genomic approaches have not found a use in routine bacterial typing and molecular epidemiology.

1.2.7 Comparative genomic fingerprinting

Comparative genomic fingerprinting (CGF) is a novel method of comparative genomics-based bacterial characterization developed by our research group. It has been used in this study in an attempt to circumvent the main problems associated with genome sequencing and microarray-based comparative genomics; *i.e.*, cost, ease and throughput.

Microarray studies have demonstrated extensive genomic variability among *C. jejuni* strains and have led to the identification of a pool of accessory genes that are present in some but not all strains (Taboada *et al.*, 2004; Taboada *et al.*, 2008). CGF is based on the concept that differential carriage of these accessory genes can be used to generate unique genomic fingerprints for genotyping purposes. A CGF assay for the analysis of *E. coli* has recently been developed by our research group and shows great promise as a high-throughput comparative genomics-based method for genotyping that yields epidemiologically relevant information (Laing *et al.*, 2008). As part of this thesis, we have developed a method of Comparative Genomic Fingerprinting (CGF) for *C. jejuni*, a rapid genotyping method that assesses the conservation status of a small number of genes previously described as having high intraspecies variability based on comparative genomics studies.

1.3 Overview of the thesis

Campylobacteriosis is a common occurrence in the United States and Canada. In the US and Canada, incidence rates range between 13 and 20 cases per 100,000 individuals (http://www.cdc.gov/nczved/dfbmd/disease_listing/campylobacter_gi.html, http://www.wrongdiagnosis.com/c/campylobacter_food_poisoning/prevalence.htm) and affects millions of people each year. Southern Alberta represents a local 'hot spot' as incidence rates range between 30 and 45 cases per 100,000 (Population Health, 2007) with the highest rates in the Chinook Regional Health Authority, an area to the south of Calgary and bordering with Montana (Alberta Health and Wellness, 2005).

As part of this thesis, 590 *C. jejuni* isolates were obtained from over 2000 environmental, animal and human clinical samples collected in the region of southern Alberta during a 3-year study (2004-2006). These isolates were subsequently analyzed using CGF in a molecular epidemiological framework aimed at examining: the identification of reservoirs that are significant sources of human infection, the identification of subtypes that may be associated with higher incidence of human infection, and the examination of persistence of prevalent subtypes in animal/environmental reservoirs.

1.3.1 Research Objectives

This research for this thesis was focused on the development of CGF as a novel genotyping method, and the development of CGF for application to molecular epidemiological studies. Several smaller objectives were addressed to achieve this purpose:

1. A dataset of *C. jejuni* isolated from a common geographical region over several sampling years was obtained. The acquisition and confirmation of 641 *Campylobacter jejuni* isolates obtained from water, animal fecal and human clinical samples was done for southern Alberta using methods of acquisition discussed in detail in Chapter 2. A list of isolates used for this study can be found in Appendix A.

2. Three CGF assays targeting 20, 35, and 119 markers (*i.e.,*CGF-20, CGF-35, and CGF-119) were developed and validated. These assays are introduced in Chapter 2, and validated in Chapter 3. The "*functional equivalence"* of the CGF-20 to CGF-35 and CGF-119 assays is demonstrated with Chapter 3. The results of assigning clusters based on the CGF-20 fingerprint is found in Appendix B.

3. A "proof of concept" study using CGF-20 to demonstrate that this approach can be used effectively in molecular epidemiological studies. A discussion of the CGF-20 genotyping results in combination with available epidemiological data is described in Chapter 4.

4. To further provide "proof of concept" of the effectiveness of the CGF-20 genotyping method, the dataset of 641 *C. jejuni* isolates was typed using a *fla*A-RFLP genotyping method and the epidemiological relevance of the

clusters generated by both methods compared. The *fla*A-RFLP method is described in Chapter 2 with epidemiological evaluation of both methods in Chapter 4.

5. Assessment of the use of additional molecular markers and advantages/disadvantages of using a high-throughput electrophoresis platform was compared to more traditional laboratory methods of data collection as described in Chapter 2 and evaluated in Chapter 3.

6. Development of CGF and its practical application to problems in molecular epidemiology are discussed in Chapter 5.

Chapter 2

2. Materials and Methods

2.1 Isolation of *Campylobacter* spp. from water

Water reservoirs along the Little Bow River, part of the Old Man River Basin, were routinely tested for the presence of *Campylobacter jejuni, Campylobacter coli,* and other *Campylobacter* spp. The Oldman river passes through a portion of southern Alberta with a watershed of 22 641 km^2 and a population of approximately 200 000 in rural farms, villages, and the city of Lethbridge. Additionally, the Oldman water basin contains the highest livestock density in southern Alberta (Hyland *et al.,* 2003). Routine testing also took place within the Sumas watershed in southern British Columbia, South Nation watershed in southern Ontario and the Grand Falls watershed in New Brunswick. The Sumas and Grand Falls watersheds are relatively small with 343 km^2 and 340 km^2 whereas the South Nation watershed is 3 900 km^2. The Sumas watershed is among the most intensive areas for poultry farms in Canada, the South Nation watershed is intensively an intensively farmed area of cropland and dairy operations, and the Grand Falls watershed is used for agriculture, primarily potatoes. These external watersheds provide a basis of comparison between water samples from Albertan and foreign water samples, and to test the ability of the typing method to cluster geographically related samples more closely than unrelated. The majority of water samples were taken weekly or bi-weekly. All samples consisted of 500 mL to 1 L of liquid which was collected upstream into several 250 mL sterile bottles using a pole system placed 30 cm below the

surface of the water. Samples were stored on ice for transport and shipped within coolers via ground transportation. After receipt they were vacuum filtered using a 0.22 µm sterile membrane filter. The filter was placed in Bolton's Broth (Oxoid Inc., Nepean, Ontario) and incubated in a microaerobic environment (5% oxygen, 10% carbon dioxide, 85% nitrogen) for 24-48 h at 42°C ± 0.5°C. Bacterial cultures were then obtained by plating a loop-full of culture onto blood-free modified charcoal cefoperzone deoxycholate agar including CCDA selective supplement SR155 containing cefoperazone (32mg/L) and amphotericin B (10mg/L) (mCCDA) plates (20.0 mL/plate) and incubated again in a microaerobic environment for 24-72 h at 42°C ± 0.5°C. Colonies that grew on these plates were considered *Campylobacter* spp. "suspects". *Campylobacter* spp. "suspects" were confirmed through biochemical analysis; *i.e.*, if both oxidase and catalase tests (Oxoid Inc) were positive, the samples were still considered to be *Campylobacter* suspects. These biochemically confirmed *Campylobacter* suspects were transferred to an overnight culture in 5 mL of Bolton's Broth and grown at 42°C ± 0.5°C under microaerophilic conditions. After 25 h, 800 µL of Bolton's Broth was placed in an equal volume of 60% glycerol and stored at -80°C. An additional aliquot of Bolton's Broth was taken and used for DNA extraction. Biochemically confirmed *Campylobacter* colonies finally were confirmed as *Campylobacter* bacterial isolates by PCR (described in Section 2.5).

2.2 Isolation of *Campylobacter* spp from animal fecal samples

Fecal samples were obtained from farmlands located within the Oldman River Basin of southern Alberta with permission of land owners and with the understanding

that exact sampling locations would be kept confidential. Fecal samples were obtained in sterile containers and stored on ice until arrival in the laboratory. Fecal samples were mixed with phosphate buffered saline (PBS) composed of 0.01M Phosphate, 0.027M Potassium chloride, and 0.137 Sodium chloride in a ratio of 1 g of fecal material to 10 mL of PBS. Samples were thoroughly agitated to obtain liquid PBS-fecal slurry. For detection of *Campylobacter*, 1 mL of slurry was added to 9 mL of Bolton's Broth and incubated microaerobically at 42°C ± 0.5°C for 24 hours. After incubation, one loopful of Bolton's Broth was streak-plated on to modified charcoal cefoperazone deoxycholate agar including CCDA selective supplement SR155 containing cefoperazone (32mg/L) and amphotericin B (10mg/L) (mCCDA) plates and incubated at 42°C ± 0.5°C, microaerophilically for 24-72 hours. Following incubation, the presence of colonies dictated that samples were *Campylobacter* spp. "suspect". Three *Campylobacter* spp. colonies were screened by biochemical analysis using oxidase and catalase tests (Oxoid Inc, Nepean, Ontario). *Campylobacter* "suspects" that were positive for both oxidase and catalase production were transferred to 5mL of Bolton's Broth. Overnight cultures that were grown at 42°C ± 0.5°C under microaerophilic conditions were combined with an equal volume of 60% glycerol and stored at -80°C. An additional aliquot from each culture was taken, and used for DNA extraction. Confirmation of *Campylobacter* colonies was completed using the PCR methods described in section 2.5

.

2.3 Isolation of *Campylobacter* spp. from clinical isolates

A total of 260 clinical isolates were obtained through collaboration with Clinical Laboratory Services (Calgary, Alberta) and the Calgary Health Region. These isolates

were obtained from fecal samples of individuals experiencing symptoms of enteritis that had been isolated by CLS and confirmed as *Campylobacter* spp. prior to arrival in Lethbridge, Alberta.

2.4 Extraction of *Campylobacter* DNA

Blood particles found in Bolton's Broth can interfere with PCR reactions. To avoid this, 1.8 mL of biochemically confirmed *Campylobacter* isolates were pelleted from solution by centrifugation in 2.0 mL microcentrifuge tubes in a conventional table top microcentrifuge at 8000 rpm for 10 min. The supernatant was removed and the pellet was resuspended in 200 µL sterile PBS or a "Q2" buffer supplied by the manufacturer (Qiagen Inc., Mississauga, Ontario). In cases where the amount of harvested cells was visibly less than that typically obtained, an additional 1.5 mL of Bolton's Broth was pelleted in the same centrifuge tube after supernatant removal. Pelleted cells were resuspended in 200 µL sterile PBS or Q2 buffer (Qiagen). DNA was extracted manually from the resuspended cells using a DNeasy tissue kit (Qiagen) according to the manufacturer's instructions and eluted in 200 µL of "elution buffer" or extracted mechanically using the Qiagen EZ1 BioRobot system, the EZ1 DNA Tissue Kit, and an EZ1 DNA blood card (Qiagen). The program included the optional ethanol wash, and a final elution volume of 100-200 µL. DNA was quantified using the programmed double-stranded DNA quantification settings on a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA).

In cases where samples were analyzed by ArrayTube™ microarrays, large scale extraction was performed using a phenol/chloroform method (Taboada *et al.*, 2004). In

this case, a bacterial lawn was grown overnight on mCCDA, Mueller Hilton, or blood agar dishes using microaerobic conditions. Cells were harvested from one-half to two-thirds of the plate and resuspended in TE 100:10 (100 mM Tris (pH 8.0), 10 mM EDTA) containing 0.5 mg lysozyme and 0.5 µg of RNase A per reaction. After 10 min incubation at room temperature, 0.4 mg of Proteinase K were added, and the samples were incubated at 37°C for 1 h; 10% SDS was added to a final concentration of 0.1%, and mixed until the sample formed a clear lysate. The lysate was transferred to phase lock gel light 2 mL tubes (5 Prime Inc., Gaithersburg, MD), with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). Mixed samples were centrifuged at 3000 rpm for 10 min. Samples were repeatedly extracted with phenol until the aqueous phase was clear. DNA was precipitated overnight by addition of 1/10 volume of sodium acetate (NaOAc, pH=7.0) and 0.6 volumes of isopropanol and collected by spooling. Spooled DNA was washed with 500 µL of 70% ethanol, and transferred to absolute ethanol. The absolute ethanol wash was decanted immediately and the DNA allowed to air dry. DNA was re-suspended in 200 µL of TE 10:1 (10 mM Tris (pH 8.0), 1 mM EDTA).

2.5 Confirmation and species identification of *C. jejuni* and *C. coli*

The identity of *C. jejuni and C. coli* isolates was confirmed by PCR targeting three genes coding for 16S rDNA, *map*A and *ceu*E (Linton, Lawson, Owen, & Stanley, 1997). Nucleotide sequence of the primer sets used for gene amplifications are outlined in Table 1. Amplification reactions contained 2.0 units of Taq polymerase (Qiagen), with a final concentration of 1X Qiagen PCR buffer, 1.5 mM $MgCl_2$, 0.1 µM of each of the 16S primers, and 0.2 µM each of the *map*A and *ceu*E primer sets (Table 1). Each reaction

Table 1: Primer sequences used to confirm *Campylobacter* spp.*. Confirmation and specification was completed using a multiplex PCR. Isolates determined to be *C. jejuni* were further tested using the *fla*A-RFLP typing method.

PCR Target	**Gene / Fragment Size (bp)**	**Forward Primer**	**Reverse Primer**
Campylobacter Species	16S rDNA / 857	ATCTAATGGCTTAACCATTAAAC	GGACGGTAACTAGTTTAGTATT
Campylobacter jejuni	*map* F / 589	CTATTTTATTTTTGAGTGCTTGTG	GCTTTATTTGCCATTTGTTTTATT
Campylobacter coli	*ceu* E / 462	ATTTGAAAATTGCTCCAACTATG	TGATTTTATTATTTGTAGCAGCG
fla A Amplicon for RFLP	*fla* A / 1700	GGATTTCGTATTAACACAAATGGTGC	CTGTAGTAATCTTAAAACATTTTC

* Primers used for confirmation and specification of the *Campylobacter* isolates were as described in (Linton *et al.*, 1997); the *fla*A protocol was as described in (Harrington, Moran, Ridley, Newell, & Madden, 2003).

mixture contained 2 μL of template DNA (~5 ng) which was brought up to a final volume of 50 μL with distilled deionized water. The amplification reactions were carried out using an Applied Biosystems thermocycler (Model 9700, Applied Biosystems, Foster City, CA). Amplifications were done using an initial denaturation step of 1 min at 95°C, followed by 30 amplification cycles consisting of a denaturation step at 95°C for 30 s, an annealing step at 59°C for 90 s and an elongation step at 72°C for 1 min. A final elongation step was done for an additional 5 min at 72°C. PCR products were maintained at 4°C until required. PCR amplicons were run on 1% agarose gel (Invitrogen Life Technologies, Carlsbad, CA) prepared with 1X Tris-borate-EDTA (TBE) buffer and containing 1 μg/ml of ethidium bromide. Gels were run at 110 V until 2 cm from the bottom edge of the gel, and visualized with ultra-violet (UV) light to confirm the presence of amplified DNA fragments 857 bp (*Campylobacter* genus), 589 bp (*C. jejuni* species) and 462 bp (*C. coli* species) long. All gels were photographed using a Syngene transilluminator and GeneSnap by Syngene software with an internal UV wavelength setting for 'EtBr/UV' gels (Synoptics Ltd., Cambridge, England).

2.6 Sample Subsets / Strain composition

The complete southern Alberta *Campylobacter jejuni* data set consisted of 260 (40.6%) clinical isolates, 241 (37.6%) animal fecal isolates, and 91 (14.2%) water isolates; the remaining 49 (7.6%) samples consisted of isolates previously evaluated in-house by the Public Health Agency of Canada (PHAC), Laboratory for Foodborne Zoonosis (LFZ) in Guelph, Ontario, using selected typing methods and fulfilled the purpose of known reference strains. Data identifying source of the isolates used in this

study is presented in Appendix A; clinical cases were separated based on individuals who had traveled prior to onset of symptoms and those who had not; animal fecal isolates were separated based on the species of animal from which the isolate was obtained; water isolates were separated based on the watershed from which the original sample was obtained. All isolates were also identified by the month beginning in June 2002 (month 1) to August 2006 (month 43) in which a sample was isolated when the information was available.

Subsets of isolates used for Comparative Genomic Fingerprinting (CGF) each consisted of similar proportions of clinical isolates, animal fecal isolates, water isolates, and reference strains; *i.e.*, subsets of *Campylobacter* isolates used to perform CGF with 20 DNA probes (CGF-20), 35 DNA probes (CGF-35) and 119 DNA probes (CGF-119) all were chosen to contain the same % water, % clinical, % environmental, and % reference strains relative to the larger sample set, in order to avoid introducing bias to the data but otherwise were selected at random.

2.7 Subtyping of strains using *fla*A-RFLP:

All *Campylobacter* strains were sub-typed using a *fla*A restriction fragment length polymorphism (RFLP) method. The methodology used to type *C. jejuni* isolates was a combination of previously described methods (Ertas *et al.*, 2004; Nachamkin, Bohachick, & Patton, 1993) and modifications (Harrington *et al.*, 2003). PCR amplification of the *fla*A allele was performed on extracted DNA using an Applied Biosystems thermocycler (Model 9700). The PCR primers used are described in Table 1. Amplification began with an initial denaturation step for 1 min at 94°C followed by 35 amplification cycles each

consisting of a denaturation step at 94°C for 15 s, an annealing step at 45°C for 15 s, and an elongation step at 72°C for 105 s. The amplified DNA was extended at 72°C for 5 min and then maintained at 4°C until needed. Presence of a 1.7 kb band in the amplified samples was confirmed by electrophoresis of 2 µL of PCR product in 1% agarose (Ultra-pure Agarose, electrophoresis grade, Invitrogen Life Technologies) made up in 1X Tris-Borate-EDTA (TBE) buffer containing 1 µg/mL ethidium bromide. All gels were photographed using a Syngene transilluminator equipped with a GeneSnap camera and software with an internal UV light source set for use with 'EtBr/UV' gels (Synoptics Ltd.).

A single enzyme digest was used for restriction analysis of purified *fla*A amplicons. Each reaction consisted of 1 unit of *Dde*-I enzyme (New England Biolabs, Ispwich, MA), enzyme buffer to a final concentration of 1X, and 10 µL of the PCR product. Samples were digested at 37°C for 2 h and the products analyzed by electrophoresis in a 2% agarose gel (Ultra-pure Agarose, electrophoresis grade, Invitrogen Life Technologies) containing 1 µg/ml ethidium bromide. All gels were run with 10 µL of a 0.1 µg/µL 1 kb plus molecular weight ladder (Promega). Gels were run at 110 V for 80 min and then visualized and photographed using a Syngene transilluminator (Synoptic Ltd Beacon House, Cambridge, England). Cluster analysis of *Dde*-I *fla*A profiles was performed using BioNumerics v4.2 software (Applied Maths, Austin, TX).

2.8 Subtyping using Comparative Genome Fingerprinting with 20 genes (CGF-20)

The presence/absence of the 20 genes used to produce a comparative genomic fingerprint in the CGF-20 assay was evaluated using four 5-plex PCR sets of primers

A

```
                                      Forward Primer
11168    534    TTATACAGGGTTTGGTTATAATCAATCAAGAAAATTTAAAACCCTTAAAATAGAGTATGG  593
81-176   83183  .........A..................................................  83242
269.97   76565  ......C..T..................................................  76624
RM1221   71546  ......C..C...........C......................................  71605

11168    594    AAAATAGTATTTATGTGGGTGGGATTTATCTATTATTCTAAGTCCTAAAATTACTTTAGA  653
81-176   83243  ....................C.......................................  83302
269.97   76625  ....................C.......................................  76684
RM1221   71606  ............................................................  71665

11168    654    TTTAGGAGCAGAGCAAAGATTTCAAAT-GAAACAAAAAATCAATGGCTATCAAAATCTG   712
81-176   83303                             -                                  83361
269.97   76685  ...............A...........-.................T.............  76743
RM1221   71666  .........................CC..G-........T.....T.............  71724
                 Reverse Primer
11168    713    AAGTGCGTTCTATTCCAACTTTAAGTTTAGGTTCTACTTATAGCATTAATTCAGATACCG  772
81-176   83362  ............................................................  83421
269.97   76744  ............................................................  76803
RM1221   71725  ............................................................  71784
```

B

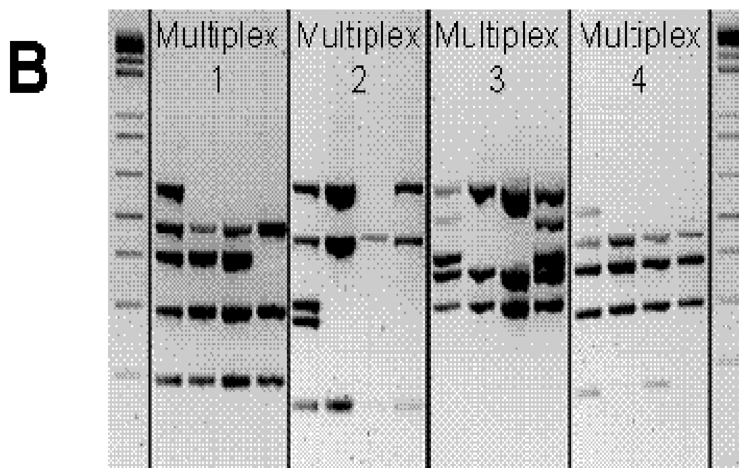
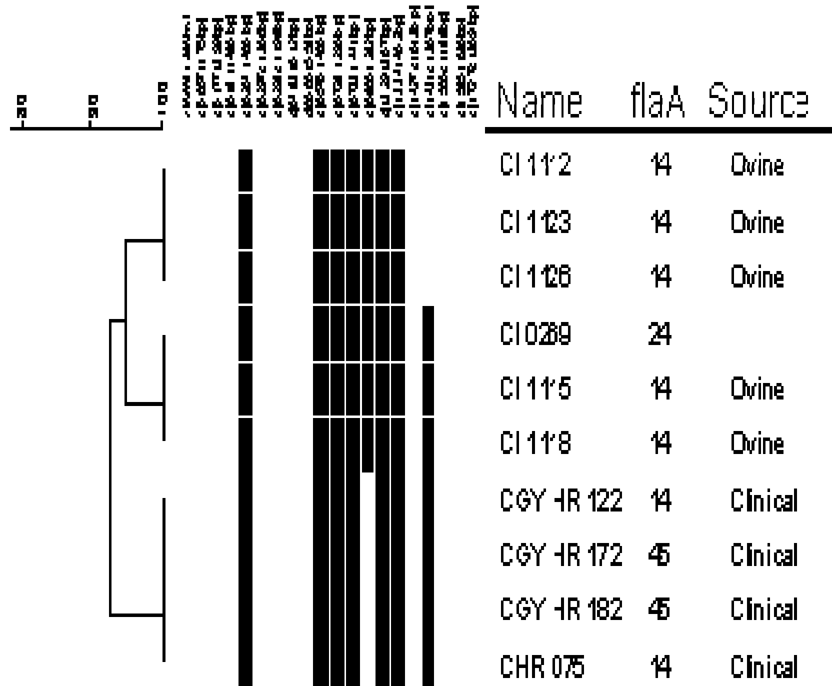


Figure 3: Development of the CGF-20 Assay. A) Primer sequences were designed in SNP free regions based on all available sequences. B) CGF-20 fingerprints were obtained from four multiplex PCRs, each targeting five genes. Sizing ladders are shown to the left and right of the four multiplex PCRs contains bands at 100bp, 200bp, 300bp, 400bp, 500bp, 700bp, 1000bp and 1500bp C) Fingerprint clustering was performed using BioNumerics software to determine similarity of fingerprints. Similarity can be visualized using program generated dendrograms based on the presence and absence of 20 highly variable loci (the unique CGF-20 genetic fingerprint) and combined with strain information including name, *fla*A-RFLP type, and available epidemiological data including source and date of isolation. Part C can be found on the following page.


C
Name
flaA
Source
CI 11'2
14
Ovine
CI 1123
14
Ovine
CI 1126
14
Ovine
CI 0289
24
CI 11'5
14
Ovine
CI 11'8
14
Ovine
CGY HR 122
14
Clinical
CGY HR 172
45
Clinical
CGY HR 182
45
Clinical
CHR 075
14
Clinical

with each gene discriminated by a signature amplicon, and each 5-plex capable of producing a unique five band fingerprint (Figure 3). Primers were designed using all available (eight) *C. jejuni* genome sequences (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi, accessed March 2007) in order to avoid single nucleotide polymorphisms (SNPs). For each target gene, the corresponding DNA sequence was retrieved from each genome strain bearing the gene in order to construct a multiple sequence alignment using the program ClustalX (downloaded from http://bips.ustrasbg.fr/fr/Documentation/ClustalX/). A consensus sequence was generated from each multiple sequence alignment, with SNP sites replaced by an unrecognizable place holder, thus, ensuring that the primers would be designed in SNP-free regions. The altered sequence was used for primer design using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) with a desired length of 30 nucleotides and a melting temperature of 60°*C* (PCR set 1-4, Table 2). These 20 genes were able to target 10 out of the 16 identified hyper-variable regions identified in *Campylobacter* spp. (Taboada *et al.*, 2004) (Figure 4) and were estimated to produce a fair representation of the genetic diversity associated with the entire genome.

PCR reactions for each of the 5-plex PCR sets were set up with final concentrations of 1X Qiagen buffer, 1.0 mM $MgCl_2$ (Qiagen), 0.2 µM of each dNTP (*i.e.,* dATP, dCTP, dGTP and dTTP), 0.5 µM of each primer, and 0.4 µl Taq DNA polymerase(Qiagen) per reaction. For the CGF-20, PCR primer sets 1-4 were used to generate amplicons (Table 2). Approximately 5 ng of template DNA was added to each reaction mixture. Distilled deionized water was added to bring the mixture to a final volume of 50 µL. PCR cycles were completed on a GeneAmp PCR Systems 9700

Table 2: Primer sequences for PCR sets 1-7 for CGF of *Campylobacter jejuni*. PCR sets 1-4 were used for the CGF-20, whereas sets 1-7 were used for the enhanced CGF (CGF-35) analysis. In order for PCR fragments to be amplified equally, PCR primer concentrations for the additional 15 genes used in the CGF-35 were altered as listed.

PCR Set	Gene	Forward Primer	Reverse Primer	Fragment Size	Final Concentration (pmol/uL)
1	cj0298c	CTGTTTCTCGCGGAGCTAAA	CAACGACGGGTATTCCTGAT	198	5
	cj0728	AGCGATAAGCCTGATGGAGA	TCTCTTTCTATGATGATGCCGTA	296	5
	cj0570	CCAAACGATCACACCTATGC	AAAAGATTTGCCTTGGACATT	405	5
	cj0181	AGGAGCGGAATTTACTTCCAT	GCTGCATAAGGGCAAGATTT	486	5
	cj0483	CAGAATGCGGTGGAAGTGAT	GCTCATCACGCACTTCTTCTT	612	5
2	cj0057	CAATCAAGAAAATTTAAAACCCTTA	TTGGAATAGAACGCACTTCAGA	175	5
	cj0860	ATAATGGGTTGGGCGAGTCT	AAAAGCAAAAATATCAAGCAAGC	282	5
	cj1431c	GAAGGGATGATGGATTTGGA	CACCCCCATTCTTTTTCGTA	307	5
	cj0733	TGGCTTAAATTTTGTGTTGAAAAA	CATCCATTTTTCATAATCTTTACCC	441	5
	cj1427c	GCTCCACTTTGTAAAAGAAATCC	CTGGTTTCCAACCTGTAGCC	613	5
3	cj0297c	GGCGTTGATATGGTTTTTATCC	GCTTTTAGCTAAGCCATCTTCTTC	300	5
	cj1727c	GGGCTTTTGGAATTCCTGTT	CCACGCCATGTTTAAAAGGT	369	5
	cj0264c	GCCTACACAGAAAATGGCAAA	TGTGGCAAAGTTGAGAGTGC	406	5
	cj0008	TGAAAGCAAGACAGCTTGGA	CCGCTTTATCTGGATCTTGTTC	524	5
	cj1585	TTGGAATTTAAGGCCTGTGC	TGTGCCGTGTTCAGCTTTTA	630	5
4	cj1550c	AGGCACTGCGTTTGAAAAGT	TTTTAGTTGCACCACCAGAATG	188	5
	cj1329	GGCTTAGGCTCTCGCTTAAAA	GCTCGGTTAAAATGTCTGCAT	307	5
	cj0177	TTTATGTTGAGCGCTTGTGC	AAAAACCACATTGACAAATTGC	399	5
	cj1334	TGCTTGTACATCGCCCTTTA	GAAATTTTGGAAAATCAAAAGGT	462	5
	cj0566	GTTTCCACAAGCCTTAGAAAAA	TCCCCAAATCCTAGTAACTCCA	558	5
5	cj 0421c	ATCCAAATTTTCTTCACTTTCTTGC	TTTAGAATCCACCTTGAAAAACACG	100	10
	cj 0033	AATTGGGATAAAAGGGGTGAGAAT	TGAAGCCAAGTAAAACCAAAAACA	206	20
	cj 0486	TGCATTGGTAGCTAATATGGGAAA	AAAATAATAAGCACCTTGCGCATC	301	10
	cj 0569	CTGATTGGTGTGGATCTAGTGGAG	CAGAAAGGAAAACTACCGGATTTG	399	10
	cj 0625	AGCGATCATGAATGATGAAAATGT	CTATAGGACTTCTTGGCGTACAAG[illegible]	498	10
6	cj 0755	TGAGCTTACTTTAAAAGGCGGTGT	GAAATCTGCCTTGGCCACTATAAC	101	10
	cj 0763	GGATCTGTAGCTTTTGATGGTGAA	CGATATTGGCTGAATCGATAAGAA	205	10
	cj 0967	CCGAACAAGGACAAGCCTTAGATA	TTTGAGAAAGATCGGTTGTGGTAA	301	10
	cj 1141	GCCAAAAGAGCAGGAGCTAAGATA	TGAAGCAAAACAAAAGGTATTTCA	413	20
	cj 1136	TATCTCGTTTAGCTGGGGTGGTAT	TGCCAAATTTTCCAAAGATCATAA	510	20
7	cj 1134	TACACGCTTGGATGAAATACCCTA	TCTTCAACGCATTTATTGTCAAGC	152	10
	cj 1552	GCCTGAAAGAACTTGCCATAATTT	GAGCTTGAAAGTTGGTTGCATAAA	222	20
	cj 1439	TCCCCCAAAATATACATTACTCTCG	TGATTATCGATATGGAGCTTTGGA	307	10
	cj 1721	AGCCGTAATTTGTTTGTCTTGGAT	GCGGGTTATCGTTTTAATGATTTG	415	10
	cj 1679	CAAGTTTTAATGGTCTTATGCTTGC	TTGGCTTCTTGTAGTGATTGAAAAA	529	10

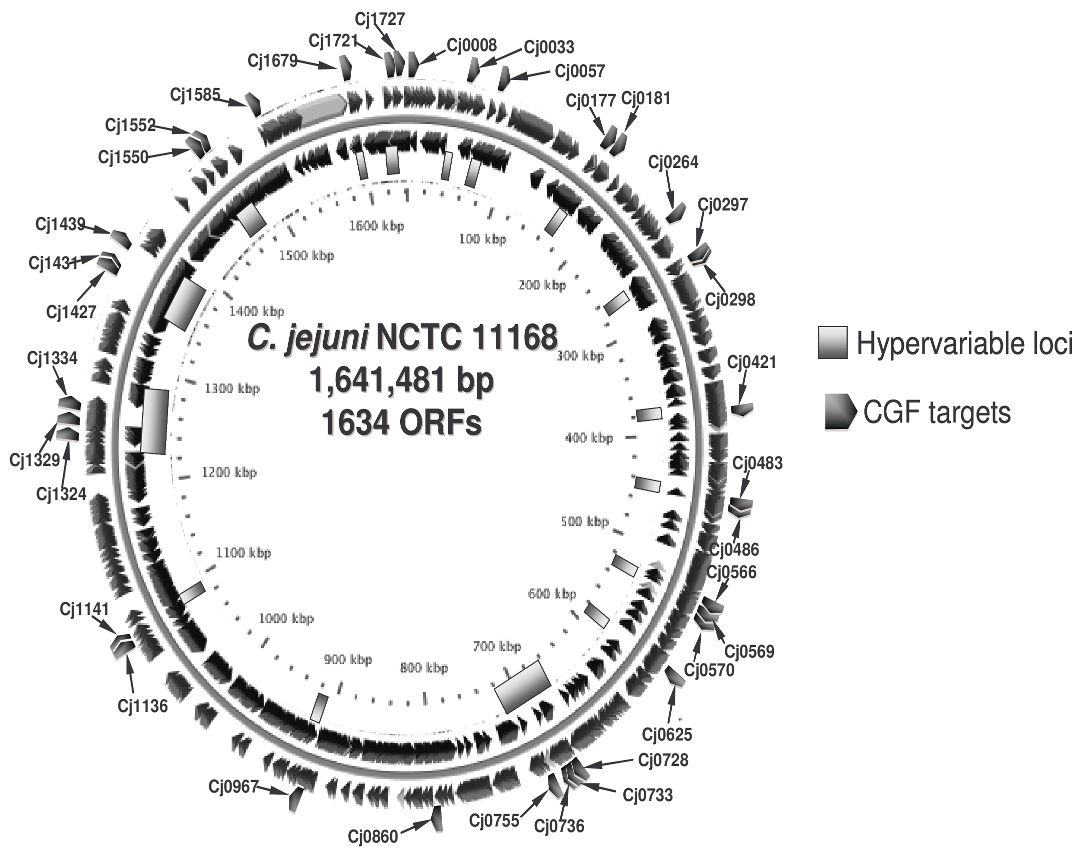


Figure 4: Location of hyper-variable regions and the CGF targets within the *C. jejuni* genome. Hypervariable regions previously identified by Taboada *et al.* (2004) are defined by boxes while arrowheads indicate ORFs in the genome (the direction of the arrowheads indicates gene orientation). Grey arrows represent the CGF targets within the genome and are placed in the approximate genome location.

thermocycler (Applied Biosystems, Foster City, CA) with an initial denaturation step at 94°C for 5 min, followed by 34 amplification cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s and elongation at 72°C for 30 s. A final elongation step at 72°C was done for 5 min before storing the sample at 4°*C* for short term storage, or at -20°C for long term storage.

PCR amplicons were viewed in 2.5% agarose gels (Ultra-pure Agarose, electrophoresis grade, Invitrogen Life Technologies) containing 1 μg/ml ethidium bromide. All gels were run with molecular weight markers to quantify band sizes. Gels were ran at 100 V until the loading dye reached a distance of 8 cm from the loading well. Gels were visualized using UV light (Figure 5) and photographed using Syngene transilluminator (Synoptic Ltd Beacon House, Cambridge, England). The presence or absence of each gene was scored as a "1" or "0" respectively, which generated unique binary fingerprint for each strain. Cluster analysis of gene fingerprints was performedusing BioNumerics v4.2 software (Applied Maths, Austin, TX) via the simple matching coefficient and UPGMA algorithm.

2.9 Subtyping using Comparative Genome Fingerprinting with 35 genes (CGF-35)

A second comparative genomic fingerprint (CGF-35) was performed based on an extended fingerprint that includes all of the genes used for the CGF-20 assay plus 15 additional genes (Table 2). These 15 genes provide additional discriminatory power by providing additional coverage of the hypervariable regions (Figure 4), and allow for 2^{35} or over 34 trillion unique fingerprints to be generated.

Figure 5: Visualization of CGF-20 targets using agarose gel electrophoresis. Lane 1 contains positive control strain NCTC 11168, Lane 2 contains an representative sample strain, NZ_0346b obatined from an outbreak in New Zealand and used as a PCR control for the southern Alberta isolates, and lane 3 contains a PCR blank. GeneRuler 1kb DNA ladder plus (Fermentas, Burlington, ON) is in unmarked lanes.

The original four 5-plex PCR sets were completed as previously described above for the CGF-20 (Section 2.8). The additional genes were distributed into three additional 5-plex PCR reaction sets with primers designed in SNP free regions. The reaction setup was identical to the PCR setup for the CGF-20 assay. Each of the additional PCR sets required optimization of primer concentrations for successful multiplexing (Table 2). Each PCR was done using a final concentration of 2.5 mM $MgCl_2$, 0.2 μM of each dNTP, and 2.0 units of Taq DNA polymerase (Qiagen) per reaction in a 1X PCR Buffer (Qiagen).

All PCR products for the CGF-35 assay were separated using a QIAxcel system (Qiagen) which consists of a QIAxcel capillary electrophoresis device and BioCalculator v3.0 (Qiagen) software for analysis. All PCR reactions were run once using the AM320 method with a 20 second injection time. The 15 bp – 3000 bp alignment marker was used as the internal standard marker and band sizes were determined using the QX 100 bp to 3 kb DNA size marker (Qiagen). The presence or absence of each gene was scored as a "1" or "0", respectively, which generated a unique binary fingerprint for each strain. The QIAxcel was able to provide a visualization of bands with a size similar to what would been seen if the samples were run on a conventional agarose gel (Figure 6). Cluster analysis of gene fingerprints was performed using BioNumerics v4.2 software (Applied Maths). Both the 20 gene fingerprint obtained from PCR sets 1-4 and the 35 gene fingerprint from PCR sets 1-7 were evaluated using Bionumerics software (Applied Maths).

PCR Set	Gene	Fragment Size
1	cj0483	612
	cj0181	486
	cj0570	405
	cj0728	296
	cj0298c	198

PCR Set	Gene	Fragment Size
2	cj1427c	613
	cj0733	441
	cj1431c	307
	cj0860	282
	cj0057	175

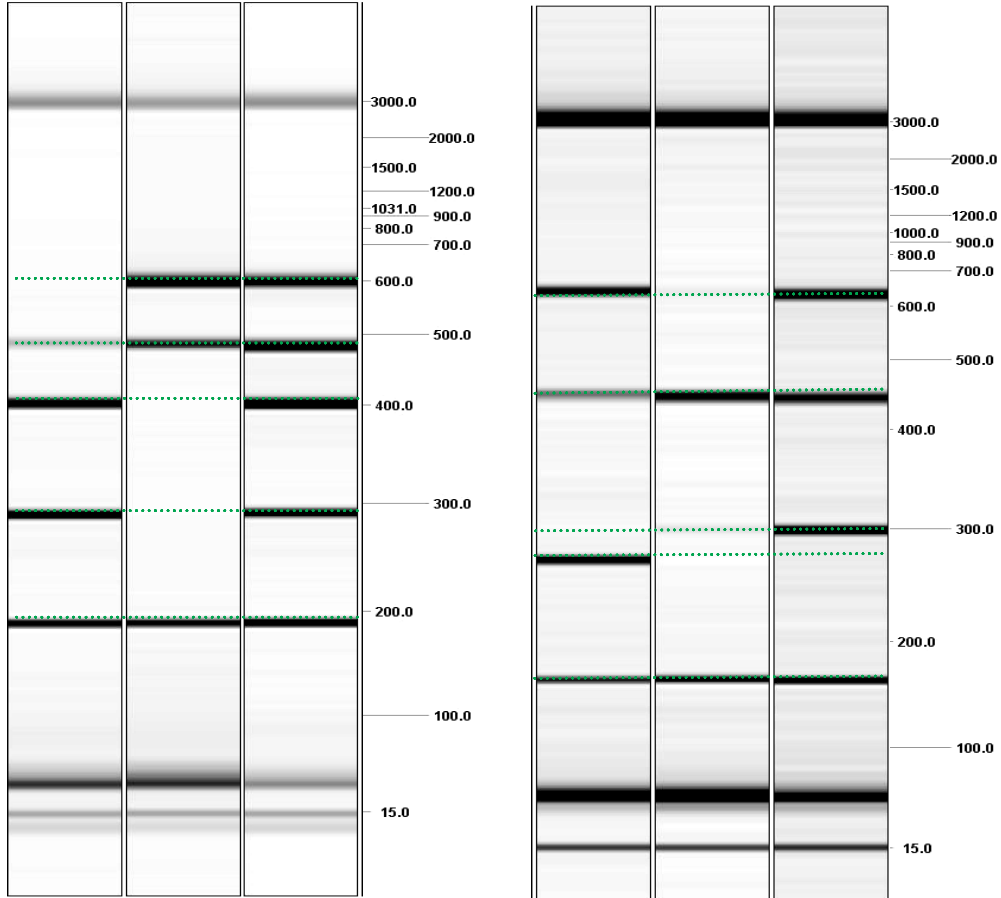
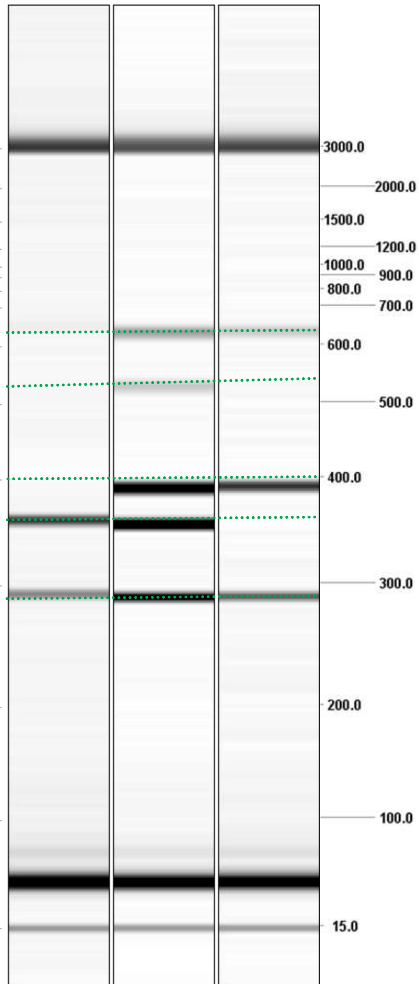
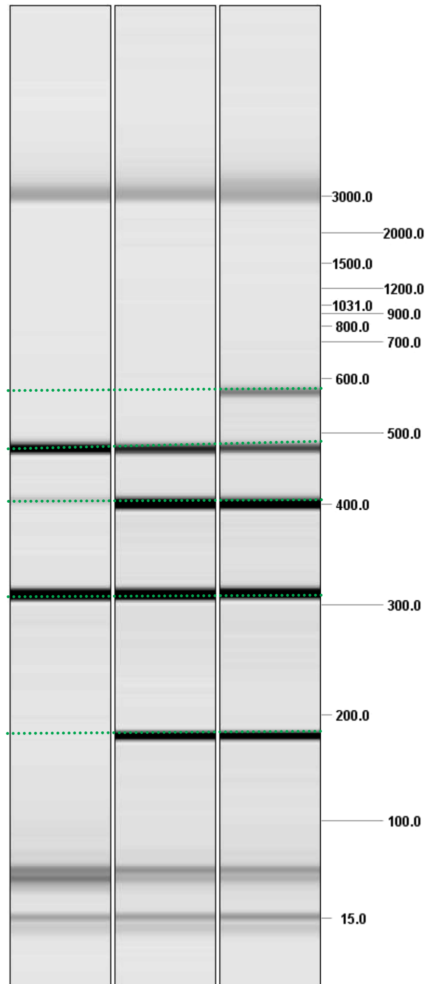


Figure 6: Visual representation of the CGF-35 images using the QIAxcel high-throughput capillary gel electrophoresis system. All seven multiplex PCRs of five genes are shown as images generated using the BioCalculator software (Qiagen, Mississauga, ON). Note that for each lane, the software adjusts band sizes using the 15 bp and 3000 bp alignment markers, which serve as internal controls. Based on the internal controls, estimated band sizes are given to the right of each set of three lanes as determined by the QIAxcel software, and lanes are normalized and aligned at 15 bp and 3000 bp. Bands were detected at concentrations as low as 0.1 ng/µL using the software, with a resolution as high as 3-5bp, allowing for identification of expected bands (shown by the green lines), and differentiation from amplification products resulting from non-specific binding. PCR Sets 3-7 can be found on the following three pages.

PCR Set	Gene	Fragment Size
3	cj1585	630
	cj0008	524
	cj0264c	406
	cj1727c	369
	cj0297c	300

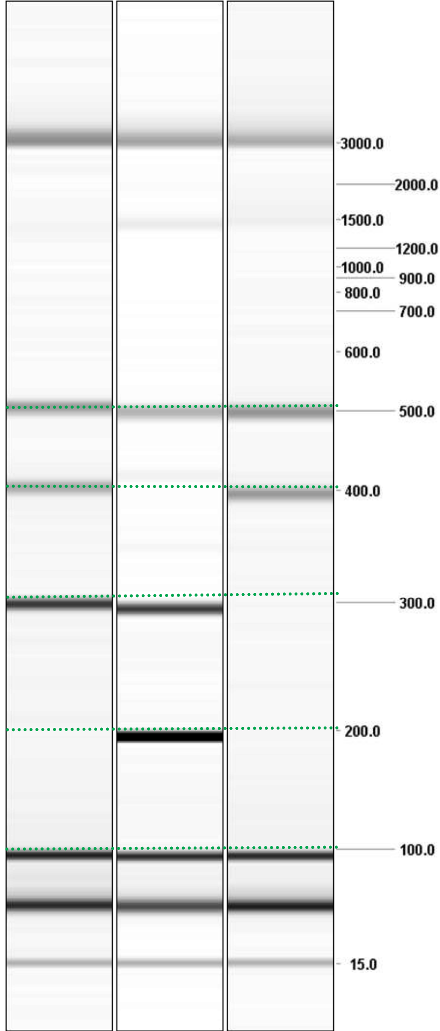
PCR Set	Gene	Fragment Size
4	cj0566	558
	cj1334	462
	cj0177	399
	cj1329	307
	cj1550c	188

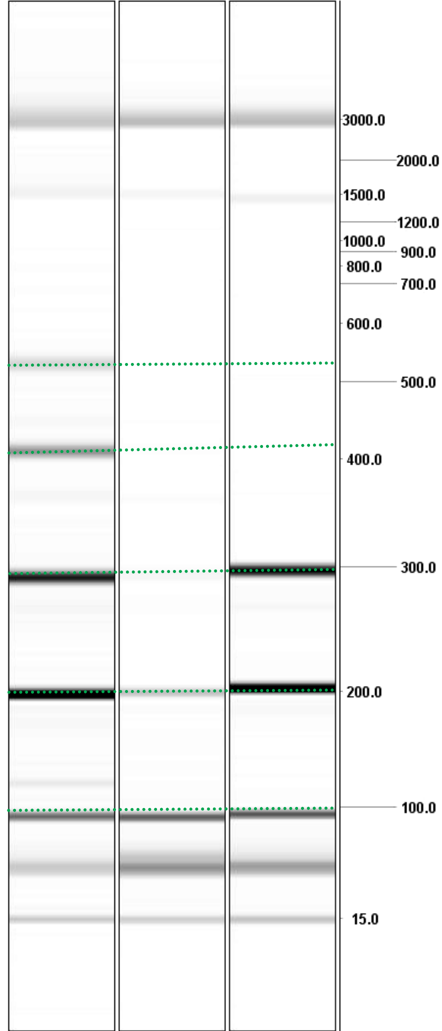




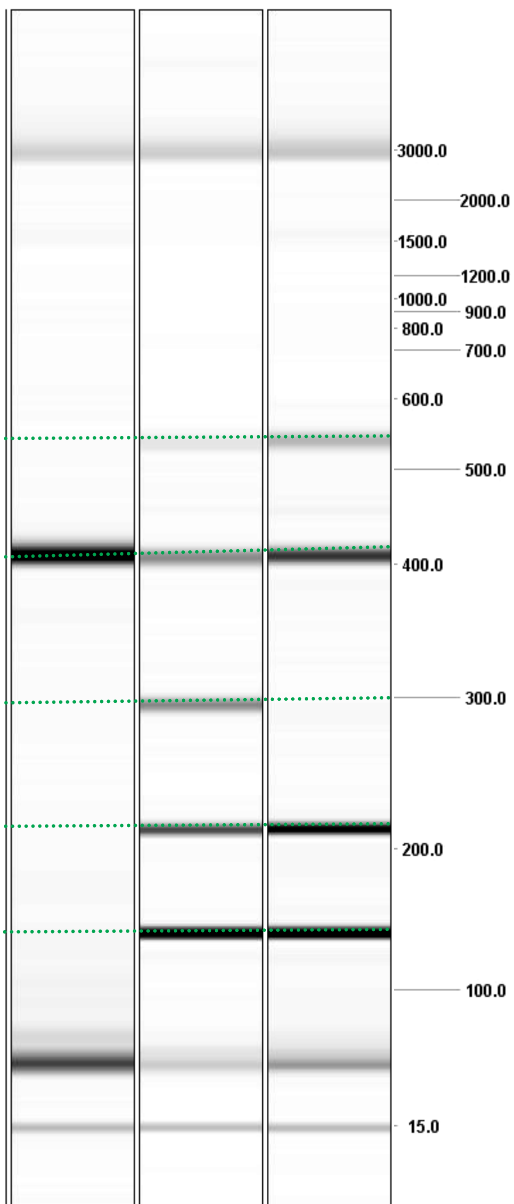
PCR Set	Gene	Fragment Size
5	cj 0625	498
	cj 0569	399
	cj 0486	301
	cj 0033	206
	cj 0421c	100

PCR Set	Gene	Fragment Size
6	cj 1136	510
	cj 1141	413
	cj 0967	301
	cj 0763	205
	cj 0755	101





PCR Set	Gene	Fragment Size
7	cj 1679	529
	cj 1721	415
	cj 1439	307
	cj 1552	222
	cj 1134	152



2.10 Subtyping using Comparative Genome Fingerprinting with 119 genes (CGF-119)

A third, high resolution comparative genomic fingerprint (CGF-119) was obtained based on conservation of 119 variably absent or present genes in *C. jejuni* using the ArrayTube platform developed by CLONDIAG Chip Technologies (Jena, Germany) (Table 3). These genes represent roughly half of the variable genes in *C. jejuni* NCTC 11168 and targeted all of the hyper-variable regions in *C. jejuni* (Taboada *et al.*, 2004). All 119 30-mer probes were designed using the program, OligoArray 2.0 (Rouillard, Zuker, & Gulari, 2003), and were designed to be SNP-free as in the CGF-20 and CGF-35 assays. Probes were produced and spotted in duplicate by CLONDIAG Chip Technologies (Jena, Germany) onto miniaturized microarrays, or ArrayTubes.

Genomic DNA was extracted from *Campylobacter* strains using an EZ1 DNA Tissue Kit and BioRobot EZ1 (Qiagen) according to the manufacturer's instructions or by a phenol/chloroform method (Taboada *et al.*, 2004) as described in Section 2.4 of the methods. DNA was quantified spectrophotometrically using a Nanodrop® ND-1000 (Nanodrop Technologies, Inc.). Four µg of genomic DNA from each *Campylobacter* strain was biotinylated during individual random amplification procedures using a BioPrime® Array CGH Genomic Labeling System (Invitrogen) and Biotin-16-dUTP (Roche) according to the manufacturer’s Instructions. Labeled genomic DNA was purified using a purification module provided with the BioPrime® kit according to directions provided by the manufacturer, and quantified spectrophotometrically. Hybridizations were initially carried out using varying concentrations (2-8 µg) of *C. jejuni* NCTC 11168 and 81-176 genomic DNA to determine the optimal amount for detection using the ArrayTubes. In the optimized version of the protocol, 5 µg of labeled

Table 3: Oligonucleotide probe sequences for CGF-119. Each 30mer probe is shown with the corresponding NCTC 11168 gene name. Continued on next page.

Gene	Probe Sequence
Cj0008	CTTAATTCCTGATGTTGATTTATCTTTGGA
Cj0014c	AGCTCCGAAAATACCCGCTAATATACCTAT
Cj0030	ACAGGAAGCATATAAAGGTATTAATGATAT
Cj0032	GATGAAATTTTAAAAGCCAAAGAACAAGAC
Cj0033	AGCTTTAGTTCTTTTCTCACTCCAAAAGAC
Cj0057	CCTGATAGCATATTTGGGATAAGCTTATGG
Cj0058	CTAAAACCTTTTAAGTTAAACACTCCAAAA
Cj0100	GTTGCTTATCAAAATTTGGCGTATTCTATA
Cj0122	TAATAAAATTCTTTCTCAAAACACTTCAGC
Cj0139	GCGAGAGAAGTATTTATAATATTGCTTCAT
Cj0170	TGTGGAAATGGAGTGCATTCTAAGTATTTT
Cj0171	GGAGCACTTTATTTATATAGGACAAAAGTA
Cj0177	AGCTATGGAGAAATAGATTTAAAAGACAGT
Cj0178	CGGTCGTACTTTTATAGTAAGTTTTGAATA
Cj0179	TAGATGAGCTTGCAAATTTTGTCTATATAC
Cj0180	CCTAGAAAATCTTTCAAGTGCTTTAGATAT
Cj0181	CAGTTTTGTTATGCCTATTGATTACAATAT
Cj0247c	GCTTGATTTTATTTGTGCAAAGATTTTCTA
Cj0259	GGCTGGAGAAATTTTGAAATTTCAATTAAA
Cj0263	GCTTTAAGTTTAAATTTGCTTGGGCAATAG
Cj0283c	ATCGACATCAGGACCAGCCATCTGAGTTTG
Cj0295	AAACAAGGTTATCAAAGAGGGACAAAATTT
Cj0299	GCTTTCGCGCTAAATTTAGATATTGATAAA
Cj0334	AACCCTAAAGGCGTGGCTGAATATCTTGGC
Cj0410	AGACAACTTGTCGCTCAAAGAAATAAATAA
Cj0422c	ATGTCTAGTTCATATGCCATATCACGCTTG
Cj0423	TTGTTTGTTTAATATGGTCTTTTACGGGAA
Cj0427	GTGGAGCATTTGCCTACTCACCATGATTAA
Cj0480c	ATTTAACACTCTAAGAGTGGGCTGATGCAT
Cj0481	AAACACTCGCTTTGATGAAGCAAAAGAGCT

Gene	Probe Sequence
Cj0483	GGTAAAAGCAGAGCTTAACGAGCAAAATGA
Cj0484	ACTTATAGTGTTAGCATTGATAGCAATTTT
Cj0485	TTTAGATAGAGCATTAAATTGGGATGAAAA
Cj0486	TTGGTATGGCTTTATAGGCTCAAAAGTGAG
Cj0487	GGGAGCAATTATCCAGTGGCAAAGATAACT
Cj0488	TAAATATGGAAGAAGTATTTCACTTAGACT
Cj0490	CCATGTGGTATATTTGCAGTATAATACTAA
Cj0555	GGCGATTTTAGCTACTATTATAATGAGTTT
Cj0566	GCAATGACAATCTTTCTTCGTCGGATTTAG
Cj0568	AGTTGGAAATCCTAAAGATAGCTATCTAGG
Cj0569	GCTCATAATAAATCTAAGCCTTTTAAAGAA
Cj0570	AATTTTAGGCATAGAAGGCGTAAGTTTGTT
Cj0617	TAGAACGATTTTGTTTGGAAATTCATATGG
Cj0618	GTGTATACTTTCAAAGAAGAAAATCATCAA
Cj0628	TTTAATAATCAAGGTATCATAGGTAATGGT
Cj0629	TAATGGAATGTTTGATAAAGATGGCAATAC
Cj0677	GCTTGGATACTTAGAATTTTAGGAGTGTGA
Cj0727	AAAGCTAAACTTTTACCAAGTGAGCAATAC
Cj0728	AGAGAACCTAATAAAACTGATGATAATGAA
Cj0730	GTTGTTACCTTTATAGCCAATTTCCTTTCT
Cj0731	ACTTTTATAGCTAGTTTGATTAAGGAGTGA
Cj0732	CAATCTTTCTTTAGCAAAAGAGTTAAAATG
Cj0733	TTAGCAAAGCCTTTGTATGTAATGGTGATG
Cj0736	AAATTTTAGCTTTGAAGTTTCAGATGATGC
Cj0737	AAAATTGCTTCAACTGGTTATGATGCTTAG
Cj0738	TAAGTATAGTAGCAAAGAAGGTTCTTATGT
Cj0741	GACAAACCTTCTGTATTGCCGCCTGAAATT
Cj0755	AGCACTTGGGTAAATACCTATAACCGTATA
Cj0818	GCTTATCAAGAAGATAGAGCTATACCTTTA
Cj0842	TCAGAAGATGTCGATACACAATATAATTAG

Gene	Probe Sequence
Cj0859c	GTTGACTTTAACATATTGACTTAAGCTATT
Cj0860	TATGATTGCCTCTATAGAGCCTGTAGCAGC
Cj0876c	CCACATAACAACAAAGTTAAAACTATAGAA
Cj0913c	AACTAATGAAATGAAATCTGCTTTAGTCAT
Cj0967	TACGATGCTCCTATTAATATTTCTAGCTAG
Cj0970	TGTCATTCGTGGTCCTGCATCAAATACAAT
Cj0971	ACAAAGGAAATTCTTTCTCAAAATTCCTAA
Cj0972	GTGCAAACTATAAGCCAGTATCCATTATAA
Cj0975	AGCTAGAGCTTTAAAGAAATATAAACTAGA
Cj1136	GGGTATTTTAGATAAAATCAACTTTGCAAA
Cj1138	CCTTTGTGTATTAAAGAGAAATTGCAACTA
Cj1140	TTCCAAGTGACATAAAACACTACTTAAAGG
Cj1141	ATACAACTAACATGGGATGATTTTGAATGA
Cj1142	GAAGCGATTTTAAATACTAGTAAATGCACT
Cj1143	TAGTTTGGATATAGATACAACAGAGGATTT
Cj1145c	CAATACCTTTATAACCTCTAGGTGTATACC
Cj1158c	TTGAGCATCTAAATTTCTAGCTTTAAACAT
Cj1160c	CTAACGATACAACGGTAAGCAAAATAATAA
Cj1208	TAGTCGGACTCGGTATAGGATATTTAGTTG
Cj1211	TTTATTCCCTTCATTATGATTGTGATTTAG
Cj1225	CTTAAAGATGGGGATTATCAAAAGGCAAGT
Cj1297	TTTGCCATATATGAAAAGCTTCTTTTATGA
Cj1298	CAGGAGATCATATAAATATTTACAACGCTA
Cj1300	ATCACTATGGCAGTAGTTTTAAATGCTTAA
Cj1306c	ATCGATTTCAAAGCTTTGGTTTATGAGCAT
Cj1307	CGACAAATTTAAGCTTAATCAAAATGGTAA
Cj1318	AGAAAATTAGATAAATGGATGGAAAGGGTG
Cj1319	GAGACAAAGCATAGAGTATTTTAAAGAAAA
Cj1320	AATTTACCTAGTAGCGTTAGAATAGCAAAT
Cj1321	ACTTAGAAGATAACTGCATAGTAAGCAGGT

Gene	Probe Sequence
Cj1322	GCTTGACTATAGAAAGCTTAAATGAAGATT
Cj1323	GACACTATAGGCATAGAACTTTGTAATAAA
Cj1324	TCATAAGCTTGACAATAAATGCGTTGAAGA
Cj1325	TTAAAAGAAATTTGGGAGCAAAATATTAGC
Cj1326	TCGGCACACCATTATATTTATATAGGTATT
Cj1327	TACAAAGAAGATGAGCTTATCCGTGAGTAA
Cj1328	ATCATTGATTCCGAAATCAACGATCTTGAT
Cj1329	TTATTGGATTGACATTGGAAGGCCAGATGA
Cj1330	TAAAAGTGCTTGAACTTTGTGATGAAGTAA
Cj1331	TGATAGCGAGCTTGATTTTAAAATCGTAGA
Cj1332	CTTAGCGATGAGAGTAAATTTATCACAGGG
Cj1333	AGAAAATTAGATAAATGGATGGAAAGGGTG
Cj1334	GATATCTTGATACTCAAAACAAAACCATAG
Cj1336	AGAAAATTAGATAAATGGATGGAAAGGGTG
Cj1337	ATTTTACCTTTGCTAGACTTCTTAGAAAAG
Cj1376	AGTGTCATATCTAAAGCTATGCTGTGGATG
Cj1419c	CTTGGTATAATCCCACACTTGTTCTACTAG
Cj1434c	GGTACAACAATAGAAACTTTTGGAGTATTA
Cj1520	GGTAAAATCTACAAGAATGAGGATGATGAT
Cj1521c	AAGCACACGCATAAATTTATCTTCAATCAT
Cj1553c	TTTCCATAACGCATCTTCTAGTTTTATTTC
Cj1556	ATCTTTAGGCAATAGTTTAGAATCCATCTT
Cj1560	TTATTTTAGGCGTGGCTATATCTTTTGGAA
Cj1563c	ATTCTATCCATTTTACCCATTCTATATCGC
Cj1602	TTGATGAAATTAAAAGCGATGAAGAAATGG
Cj1630	AATGGAGATTTAAGAATTACACTACCTATT
Cj1677	GTGGTTTTATTAGTGCAAATAAGGGTGTAG
Cj1678	TAATGGAATGTTTGATAAAGATGGCAATAC
Cj1679	GGCCAAAAGCAAAATGTTTTATCATAAAAG

genomic DNA from each sample was concentrated to a volume of 10 µl using a Speedvac centrifuge (Thermo-Savant SpeedVac Concentrator SPD111V, Holbrook, NY, USA).

All probes were assessed using *C. jejuni* NCTC 11168 as a positive control strain. ArrayTubes were then validated using a set of 19 strains that were previously characterized by whole genome microarray analysis and MLST (Taboada *et al.*, 2008). Once the *Campylobacter* ArrayTubes were validated, the 74 strains selected for additional analysis (Appendix A, as indicated) were assessed. *C. jejuni* NCTC 11168 was used as a positive control strain for all experimental analyses.

The ArrayTubes were washed with 500 µl hybridization buffer (1M NaPOi, 20% SDS, 0.5M EDTA, 1X SSC, pH 7.2) for 5 min at 45°C while shaking at 550 rpm in a thermomixer (Eppendorf, Mississauga, ON). Ten µl of labeled *Campylobacter* genomic DNA (5 µg) was combined with 90 µl hybridization buffer, heat denatured for 5 min at 95°C using a T Gradient thermocycler (Biometra, Montreal Biotech Inc., Montreal, QC) and added to the ArrayTubes for a 60 min hybridization at 45°C and 550 rpm. The ArrayTubes were then washed with 500 µl 2X SSC (Invitrogen) 0.01% Triton X-100 (BioRad, Mississauga, ON), then with 500 µl 2X SSC (Invitrogen), both for 5 min at 40°*C* and 550 rpm, followed by a wash with 500 µl 0.2X SSC (Invitrogen) for 5 min at 30°*C* and 550 rpm. The ArrayTubes were blocked with 100 µl 6X SSPE (Invitrogen)/0.005% Triton X-100 (BioRad)/2% skim milk powder (Oxoid, Nepean, ON) for 15 min at 30°C and 550 rpm. Streptavidin-horseradish peroxidase conjugate (Sigma-Aldrich Canada, Ltd, Oakville, ON) was diluted 1:2500 in 6X SSPE (Invitrogen)/0.005% Triton X-100 (Bio-Rad) and 100 µl of the diluted mixture was added to the ArrayTubes

for 15 min at 30°C and 550 rpm. The ArrayTubes were then washed with 500 µl of 2X SSC (Invitrogen) 0.01% Triton X-100 (Bio-Rad) for 5 min at 30°C at 550 rpm, then 500 µl 2X SSC (Invitrogen) and 0.2X SSC (Invitrogen) both for 5 min at 20°C and 550 rpm. One hundred microlitres of True Blue peroxidase substrate (Kirkegaard & Perry Laboratories, Gaithersburg, MD) was added to each ArrayTube for 5 min at 25°C to facilitate visualization.

Signals were detected using the ATR01 ArrayTube reader (CLONDIAG) and extracted using IconoClust software, version 2.5 (CLONDIAG). A Microsoft Excel macro was used to calculate mean signal values for the duplicate spots for each probe. Spots flagged due to poor spot morphology were not included for further analysis. Once individual reports were created for each *Campylobacter* sample, a metadata report was created to include all samples in the dataset. The data set was analyzed by calculating the "percent of positive" (PoP) for each probe compared to the average *C. jejuni* NCTC 11168 positive control signal (*i.e.*, average sample signal per average *C. jejuni* NCTC 11168 signal in six replicates). Spots were considered conserved if the PoP was greater than 55%, divergent if the PoP was between 25-55%, and absent if the PoP was less than 25%.

2.11 Molecular epidemiological analysis

Primary data analysis was completed using Microsoft Excel. Binary fingerprints were imported into Bionumerics (Applied Maths, and clustered as described in Section 2.12. For ease of analysis, and to define a temporal cluster, isolate dates were recorded by month, and assigned a numerical digit starting from the earliest date in our data set,

which was assigned the first value. Two approaches were used to obtain information from the data. Initially isolates were screened according to primary composition (animal, water, clinical) or secondary composition (actual animal species or water source). Alternately, data summaries were produced using pivot tables in Microsoft Excel.

2.12 Analysis of fingerprints using Bionumerics

Present/absent variable gene results were recorded into Microsoft Excel as a binary fingerprint before import into Bionumerics v5.1. Fingerprints were incorporated as an assay panel and each fingerprint type (CGF-20 or CGF-35) was analyzed separately. Clustering was performed using the simple matching coefficient using the UPGMA clustering algorithm. Cluster numbers were automatically assigned at 65%, 75%, 80%, 85%, 90%, 93%, 95% and 100% identity using an add-on internet script available from within the Bionumerics program. Data was exported to Microsoft Excel for further analysis and evaluation.

2.13 Statistical analysis of clustering results

The discriminatory ability of the different typing methods, CGF-20 and *flaA*-RFLP, were addressed using Simpson's Index of Diversity (ID). The Simpson's ID is used to determine the probability of any two strains randomly sampled from a population to belong to two different types (Simpson, 1949) and has been adapted for evaluating typing methods (Hunter & Gaston, 1988) and improved for objective assessment of identification of efficient and discriminatory methods (Grundmann, Hori, & Tanner, 2001). The Simpson's ID and Confidence Intervals (C.I.) were calculated online for *fla*A-

RFLP, and CGF-20 using the Comparing Partitions website (http://comparingpartitions.info) (Carrillo *et al.*, 2004). Additional analysis of clusters was performed using Microsoft Excel through the use of pivot Tables. Clusters were evaluated based on composition and sample types.

Similarity between strains was defined using the CGF-119 fingerprint. One point was given for each identical match (*i.e.*, for a gene scored as absent, present or divergent in the same way in both assays). One half point was given for each divergent to present, or present to divergent match identified when the two assays were compared. Points for each strain were totaled, and used to calculate a percentage of total possible matches; *i.e.*, 119 points, or 100%.

Chapter 3

3. Development and validation of a comparative genomic fingerprint for *Campylobacter jejuni*

3.1 Introduction

Southern Alberta has nearly twice the incidence of campylobacteriosis compared to both northern Alberta and across Canada. Sub-species typing methods play an important role in the classification of strains, surveillance to identify trends in disease transmission, outbreak detection, and monitoring of prevention/control efforts. *Campylobacter* is no exception and significant efforts have been targeted at the development of sub-typing methods for use in molecular epidemiology. However, comparative genomics, the study of genetic similarity among organisms at the whole-genome level, has yet to be developed and applied to exploring the molecular epidemiology of *C. jejuni* within southern Alberta.

Comparative genomic fingerprinting (CGF) of *Campylobacter jejuni* is based on the selection of genes identified as highly variable after extensive whole genome microarray comparison of a dataset comprised of 150 *C. jejuni* strains (Taboada, personal communication). It has been used in this study in an attempt to circumvent the main problems associated with genome sequencing and microarray-based comparative genomics; *i.e.*, cost, ease and throughput. Microarray studies have demonstrated extensive genomic variability among *C. jejuni* strains and have led to the identification of a pool of accessory genes that are present in some but not all strains (Taboada et al.,

2004; Taboada et al., 2008). The concept that differential carriage of these accessory genes can be used to generate unique genomic fingerprints for genotyping purposes is the basis for CGF.

Three variations of the CGF assays (CGF-20, CGF-35 and CGF-119) were developed (Sections 2.8, 2.9 and 2.10) based on the number of variable genes evaluated in each method. Additionally, a high-throughput alternative using capillary electrophoresis in place of traditional horizontal agarose gels for determining PCR product sizes to facilitate data analysis was also employed (Section 2.6). The CGF-20 method was tested using a sample set of 641 *C. jejuni* isolates including isolates from water, animal fecal, human clinical samples and known reference strains that were previously evaluated using *fla*A-RFLP. This chapter addresses the results of CGF-20, adaptation of the CGF-20 to a high-throughput methodology, the reproducibility of CGF-20 results, validation of the CGF-20 clustering results using both the CGF-35 and CGF-119 assays, and the overall functionality of CGF to define clusters and discriminate among *C. jejuni* strains.

3.2 Results

3.2.1 Analysis and validation of the CGF-20 assay

A total of 268 unique CGF-20 fingerprints or clusters were obtained by analysis of all 641 *C. jejuni* isolates in the collection (Figure 7). These clusters resolved to an average of 2.4 ± 3.8 members per cluster, with the largest cluster containing 29 isolates (4.5% of all isolates). A total of 184 clusters contained a single isolate (29% of all

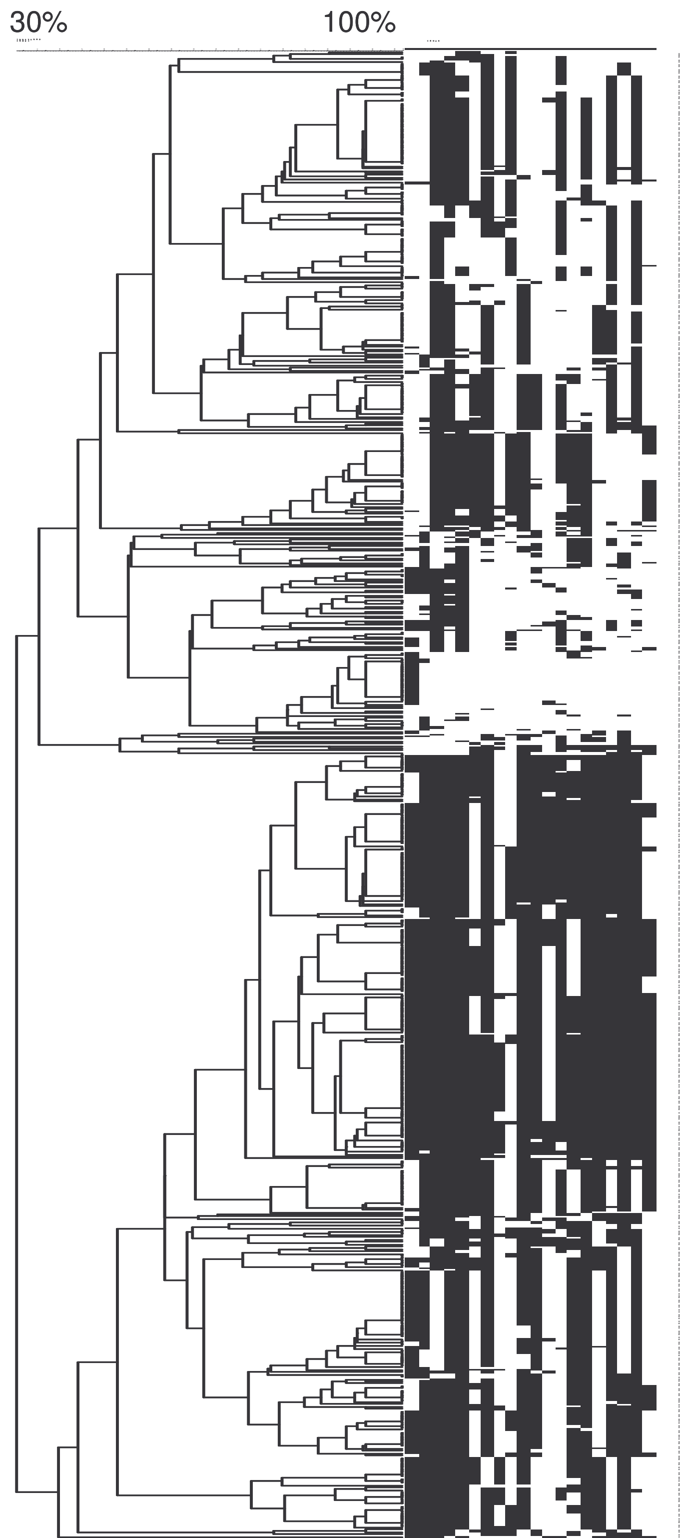

% Similarity
CGF-20
Fingerprint
30%
100%

Figure 7: Dendrogram of the CGF-20 data from the southern Alberta data set and includes a set of 49 reference strains. Binary fingerprints (shown to the right of the dendrogram), are visualized as black indicating gene presence, with white indicating gene absence, and were clustered in BioNumerics v 5.1 (Applied Maths) using the simple matching coefficient and UMPGA clustering algorithm. Each unique cluster (*i.e.,* groups of strains with 100% fingerprint identity and each strain with no matching isolates – *i.e.,* "singletons") was assigned a CGF-20 cluster number. Clusters resolve to an average of 2.4 ± 3.8 members per cluster with the largest cluster containing 29 isolates. A detailed version of the dendrogram which allows visualization of the CGF-20 fingerprints is available in Appendix B.

isolates) and were considered as ‘singletons’. A more detailed epidemiological analysis of the CGF-20 clustering results is presented in Chapter 4.

3.2.2 Development of a high-throughput CGF assay

Although traditional PCR-to-gel method has reasonably fast throughput (*i.e.,* 96 samples can be processed in 3-4 days), a method with enhanced throughput was developed and evaluated. To overcome the agarose gel electrophoresis ‘bottleneck’, an alternative method of visualizing PCR fragments was developed using the Qiagen QIAxcel capillary electrophoresis system. After PCR, samples in 96-well plates were loaded, separated and visualized directly by the instrument, eliminating the step-by-step process of manual pouring, loading, running and scoring agarose gels and reducing handling of toxic materials. Samples were sequentially analyzed and visualized after each 12-sample lane was run, and immediately scored with the exact DNA fragment size estimated by the machine’s software. Initially, a subset of 96 samples had PCR results visualized by both traditional gel electrophoresis and capillary gel electrophoresis to determine the reproducibility of amplicon detection and size estimation by the two methods (*i.e.,* the same PCR amplicons were run using both methods). The CGF-20 was then tested for assay reproducibility by repeating all multiplex PCRs using a subset composed of 93 isolates (*i.e.,* using the same strains and comparing fingerprints generated on two separate occasions).

3.2.2.1 Analysis of reproducibility: A comparison of agarose gel electrophoresis vs. capillary gel electrophoresis

The four multiplex PCR reactions used in the CGF-20 assay were initially resolved using agarose gel electrophoresis. These same multiplex PCR products were subsequently subjected to capillary electrophoresis using the QIAxcel (Qiagen, Mississauga, ON) to determine the technical reproducibility between both methods of amplicon visualization.

A preliminary analysis of multiplex concordance was performed in order to identify whether any individual multiplex was more prone to variant results when comparing the original results to those obtained using capillary electrophoresis. A majority, 344 out of 384 (89.5%), of 5-plex PCR sets (*i.e.*, four 5-plex PCR sets x 96 samples) gave the same result using both methods (Figure 8), with each of the four 5-plexes yielding similar levels of concordance, ranging from 88.5 to 90.6%. Out of the 96 samples analyzed 76 (79.2%) had identical CGF-20 fingerprints when replicated using capillary gel electrophoresis. When the 20 samples with variant fingerprints were examined for mismatches, the majority, 15 out of 20 (75.0%), had a single gene mismatch. Only 5 out of 20 samples had multiple gene mismatches (Figure 9). The overall concordance was therefore assessed by examining the number of genes that yielded identical data using both methods. A total of 1920 observations were performed (*i.e.,* 96 samples x 4 multiplexes x 5 genes per multiplex) of which 1893 (98.6%) generated identical data.

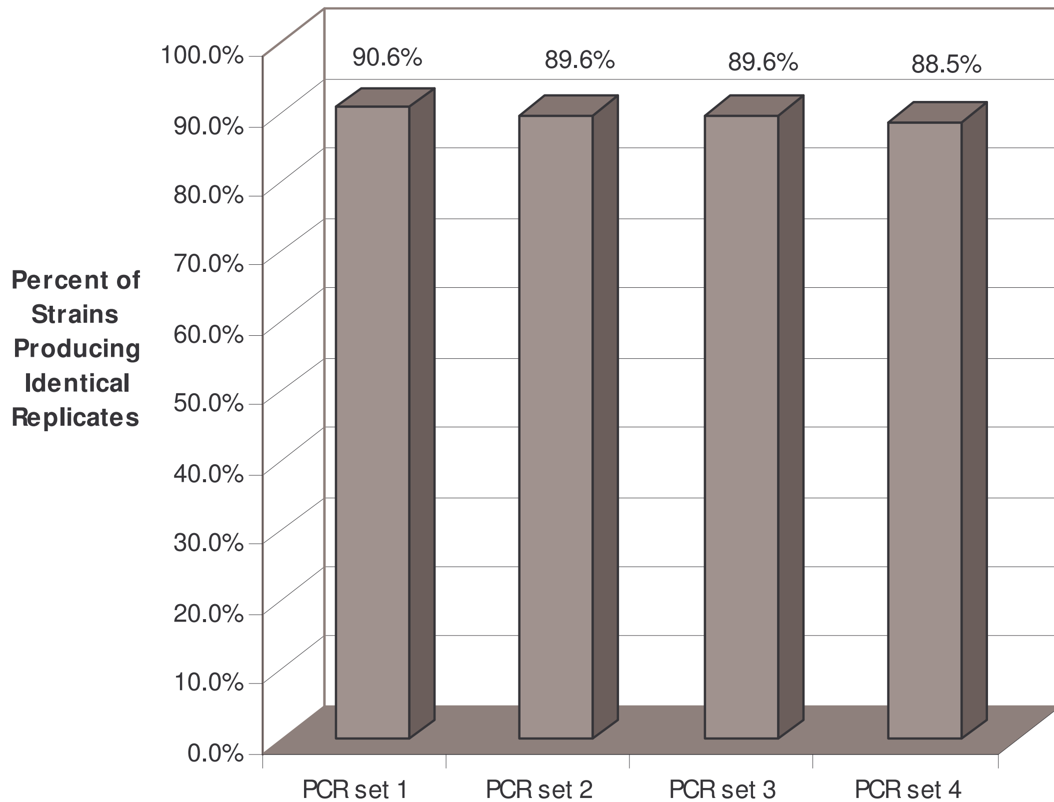


Figure 8: A comparison of CGF-20 amplification products analyzed by capillary gel electrophoresis using the QIAxcel instrument with CGF-20 amplifications analyzed using agarose gel electrophoresis. Each multiplex PCR generated very similar results ~89% reproducibility (n=96) regardless of the method used for visualization of PCR amplicons.

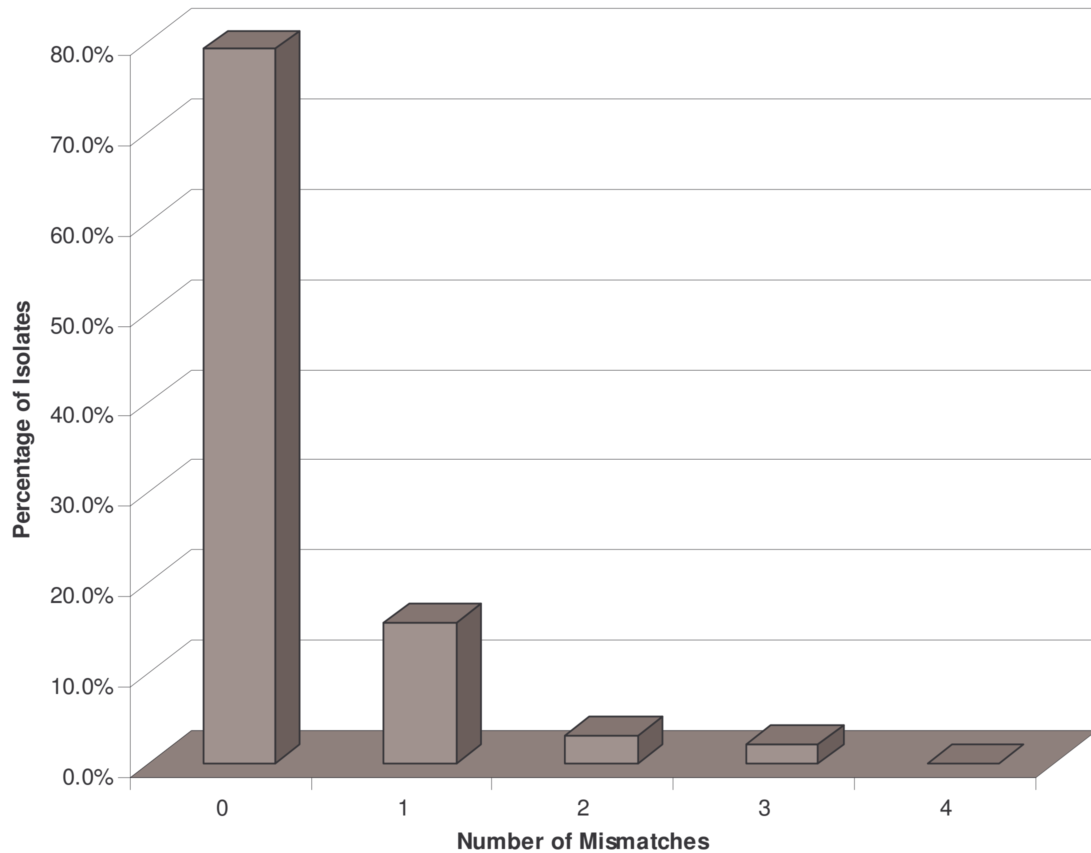


Figure 9: Distribution of mismatches obtained for CGF-20 fingerprints when the amplicons were measured by capillary gel electrophoresis using a QIAxcel instrument. The number of mismatches (*i.e.*, non-identical results for genes present or absent) between capillary gel electrophoresis and agarose gel electrophoresis was calculated. The majority of samples (94.8%, n=96) had identical fingerprints or only one mismatch. Overall, a total of 1893 out of 1920 (98.6%) individual gene observations yielded identical data using both methods.

3.2.2.2 Analysis of assay reproducibility using PCR replicates

In order to determine the reproducibility of the CGF-20 assay, a subset of 93 strains were re-evaluated by performing the assay a second time, and comparing the resulting CGF fingerprints to the original.

A preliminary analysis of multiplex concordance was performed in order to identify whether any individual multiplex was more prone to variant results when comparing the results from both replicates. The majority, 345 out of 384 (89.8%), of 5-plex PCR sets (*i.e.*, four 5-plex PCR sets x 96 samples) gave the same result using either methods (Figure 10), with each of the four 5-plexes yielding similar levels of concordance, ranging from 86.0 to 93.5%. Out of the 96 samples analyzed, 66 (68.8%) had identical CGF-20 fingerprints when replicated using capillary gel electrophoresis. When the 30 samples with variant fingerprints were examined for mismatches, the majority, 18 out of 30 (60.0%), had a single gene mismatch (*i.e.,* an amplicon of the expected size would be present in the initial PCR, but not in the replicate or vice versa). Only 12 out of 30 samples had multiple gene mismatches (Figure 11). The overall concordance was therefore assessed by examining the number of genes that yielded identical data using both methods. A total of 1920 observations were performed (*i.e.,* 96 samples x 4 multiplexes x 5 genes per multiplex) of which 1849 (96.3%) generated identical data.

3.2.3 Development of High Resolution CGF Assays: CGF-35 and CGF-119

A second CGF assay (CGF-35) was developed based on an extended fingerprint that includes all of the genes used for the CGF-20 assay plus 15 additional genes. These

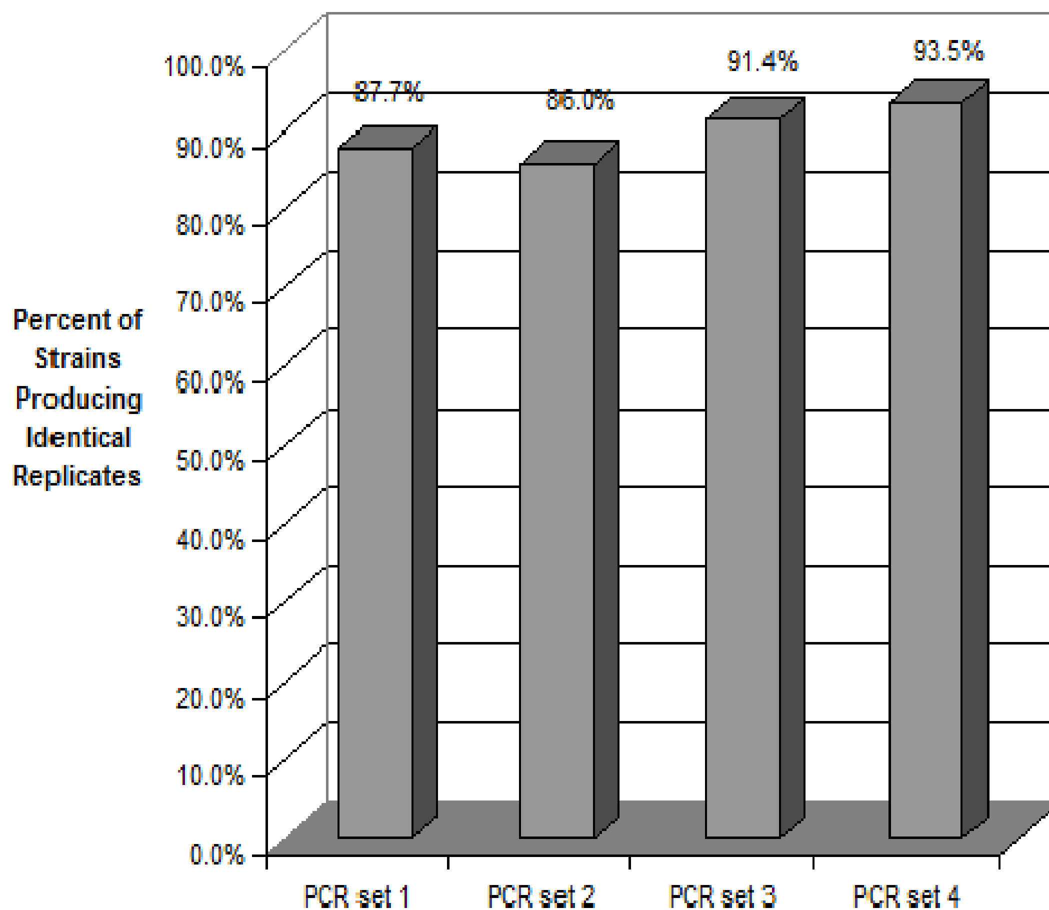


Figure 10: Reproducibility of CGF-20 fingerprints by PCR replication. CGF-20 fingerprints were compared from two sets of PCRs to establish the reproducibility of method. Each PCR generated very similar results, with ~2/3 of all strains generating identical fingerprints at all 20 loci when replicated (n=93). Among the remaining strains, most were identical in three of four PCR sets.

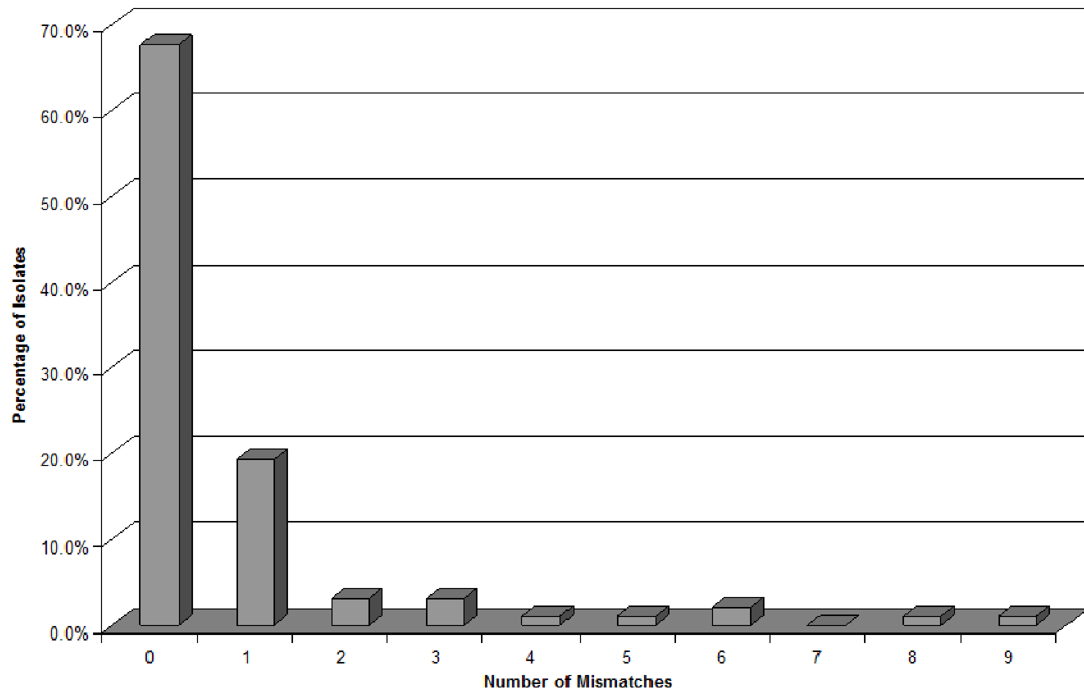


Figure 11: Distribution of mismatches obtained for CGF-20 fingerprints between replications of PCR. The number of mismatches (*i.e.*, non-identical results for genes present or absent) was calculated when the CGF-20 assay was replicated. The majority of isolates (87.1%, n=96) had identical fingerprints or only one mismatch. Overall, a total of 1849 out of 1920 (96.3%) individual gene observations yielded identical data when the PCRs were replicated.

15 genes provide additional discriminatory power by providing additional coverage of the hypervariable regions in the *C. jejuni* genome. A subset of isolates was selected to evaluate this method.

A third CGF assay, targeting 119 variably absent/present genes using the ArrayTube platform, was also developed. A subset of 74 isolates was evaluated using the CGF-119 assay in order to determine the concordance of clustering obtained with this assay and the CGF-20 assay, and to assess the extent to which the CGF-20 assay can be used to predict overall genomic similarity.

3.2.3.1 Validation of CGF-20 clusters using CGF-35 and CGF-119

Nearly all CGF-20 clusters examined showed a strong propensity for high within-cluster similarity at both the CGF-35 and CGF-119 levels (Figure 12). Strains displayed a high *intra-cluster similarity* (*i.e.*, strains within a CGF-20 cluster) at the higher resolutions with an average similarity of 91.6% in CGF-35 fingerprints and 80.1% in CGF-119 fingerprints (Table 4). The overall average *inter-cluster similarity* (*i.e.*, similarity to strains external to the CGF-20 cluster) was dramatically lower at 60.6% between CGF-35 fingerprints, and 54.7% between CGF-119 fingerprints (Table 4).

3.4 Discussion

Comparative genomic fingerprinting has been developed as a platform-independent approach to genotyping based on selecting genes that are variably present or absent (VAP) in the genome as determined from whole-genome comparative genomic analysis. CGF builds on the genetic diversity of *C. jejuni* demonstrated in the literature

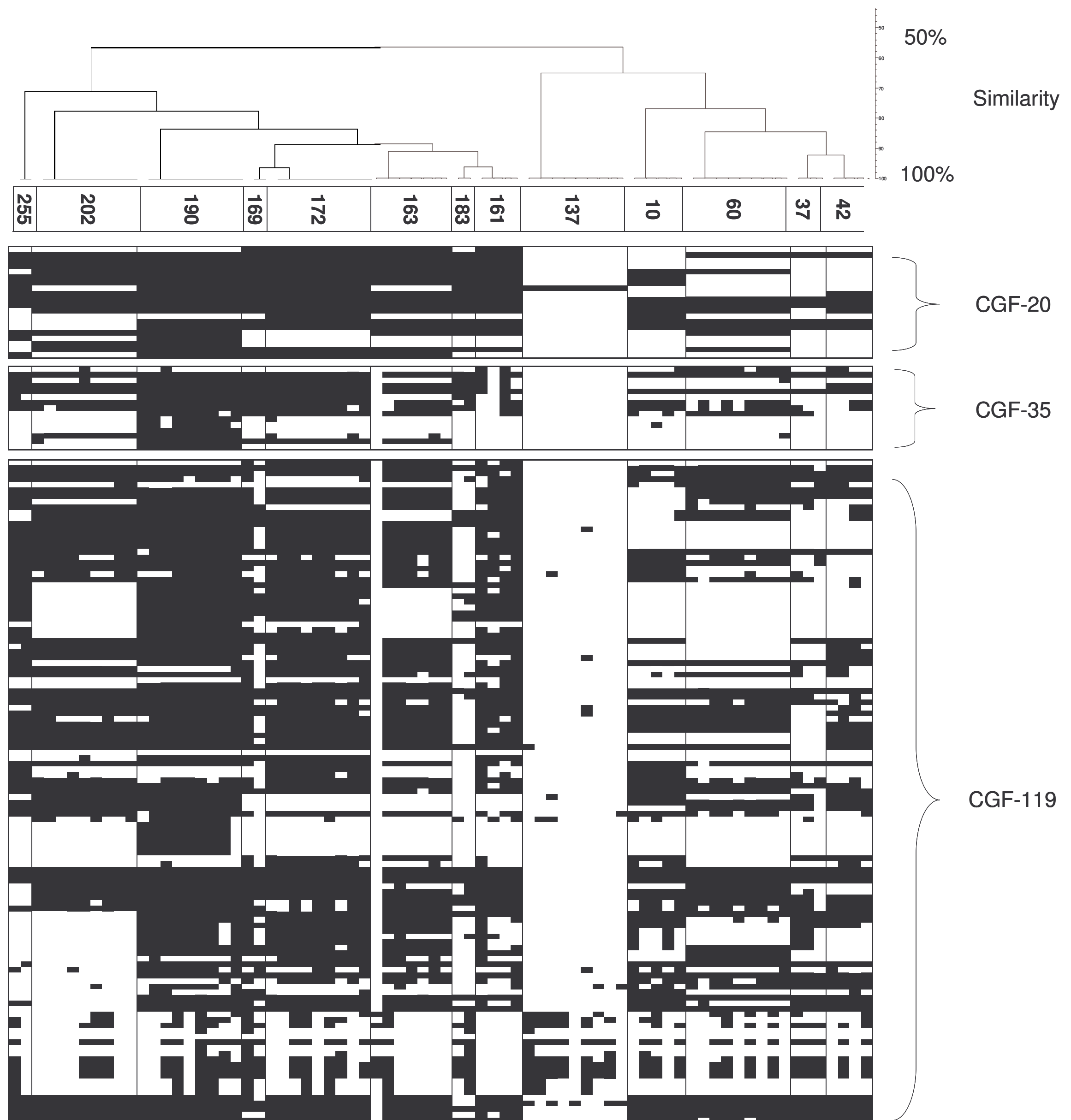


Figure 12: Detailed genomic analysis of select CGF-20 clusters using CGF-35 and CGF-119 assays. CGF-20 clusters, shown with the calculated genetic similarity, are shown with the corresponding CGF-20, CGF-35 and CGF-119 fingerprints, and below the CGF-20 cluster number. Black is used to indicate that a gene is present or divergent; white is used to indicate that the gene is absent.

Table 4: Average intra-cluster and inter-cluster genetic similarities identified for CGF-20 clusters based on CGF-35 and CGF-119 data. Pair-wise similarities were computed for each isolate versus all other isolates within the CGF-20 cluster (intra-cluster) and all isolates outside of the CGF-20 cluster (inter-cluster) for the CGF-20, CGF-35 and CGF-119 fingerprints is given with the standard deviation between all intra-cluster or inter-cluster isolates. Matches for CGF-20 and CGF-35 fingerprints were calculated based on the simple matching coefficient (score of 1.0 per match, as a percentage of 20 or 35 possible matches, respectively). Matches for CGF-119 were determined using the scoring scheme described in the methods section (score of 1.0 per match, 0.5 per divergent/present mismatches, as a percentage out of a total 119 possible matches).

		Intra-Cluster			Inter-Cluster		
Cluster Number	# Members	CGF-20 Average Similiarity	CGF-35 Average Similiarity	CGF-119 Average Similiarity	CGF-20 Average Similiarity	CGF-35 Average Similiarity	CGF-119 Average Similiarity
10	5	100.0% ± 0.0%	78.5% ± 7.5%	85.4% ± 4.6%	56.1% ± 9/3%	63.5% ± 11.0%	58.2% ± 8.7%
37	3	100.0% ± 0.0%	96.2% ± 1.5%	82.2% ± 6.0%	54.5% ± 18.0%	61.2% ± 16.8%	56.8% ± 9.2%
42	4	100.0% ± 0.0%	94.3% ± 4.2%	82.0% ± 4.1%	58.7% ± 12.7%	63.1% ± 12.7%	59.1% ± 8.9%
60	9	100.0% ± 0.0%	94.0% ± 5.2%	86.8% ± 6.7%	59.7% ± 10.5%	62.2% ± 10.8%	58.8% ± 8.4%
137	9	100.0% ± 0.0%	100.0% ± 0.0%	91.3% ± 3.3%	32.9% ± 16.9%	43.0% ± 19.7%	40.8% ± 15.7%
161	4	100.0% ± 0.0%	63.8% ± 13.3%	71.4% ± 8.0%	67.4% ± 20.8%	62.8% ± 14.9%	57.0% ± 15.1%
163	7	100.0% ± 0.0%	85.6% ± 12.1%	72.3% ± 24.4%	64.8% ± 23.1%	65.0% ± 13.1%	59.2% ± 15.6%
169	2	100.0% ± 0.0%	97.1% ± 0.0%	42.9% ± 0.0%	66.5% ± 22.2%	65.3% ± 18.2%	52.1% ± 13.4%
172	9	100.0% ± 0.0%	94.8% ± 4.0%	81.3% ± 7.5%	59.4% ± 21.2%	59.7% ± 17.6%	54.8% ± 15.9%
183	2	100.0% ± 0.0%	97.1% ± 0.0%	84.5% ± 0.0%	70.4% ± 18.7%	66.5% ± 14.1%	46.5% ± 14.2%
190	9	100.0% ± 0.0%	94.6% ± 4.3%	83.9% ± 9.1%	59.1% ± 22.5%	49.6% ± 20.0%	51.5% ± 17.1%
202	9	100.0% ± 0.0%	97.5% ± 2,6%	91.1% ± 5.1%	59.1% ± 15.5%	61.9% ± 9.2%	59.4% ± 8.7%
255	2	100.0% ± 0.0%	97.1% ± 0.0%	86.6% ± 0.0%	53.2% ± 12.1%	64.0% ± 11.7%	56.7% ± 10.4%
Overall Average Similiarity		**100.0%**	**91.6%**	**80.1%**	**58.6%**	**60.6%**	**54.7%**

through analysis of whole-genome microarrays and whole-genome comparative genomic sequencing, which have provided insight into genome evolution, and the development of hypotheses related to virulence, pathogenicity, and host specificity (Dassanayake *et al.*, 2005; Dobrindt & Hacker, 2001; Dorrell *et al.*, 2001; Pearson *et al.*, 2003; Taboada *et al.*, 2004). Microarray-based comparative genomic hybridization remains labor intensive and expensive, and is not used to any significant extent in molecular typing and molecular epidemiology. CGF represents a possible alternative to microarray-based comparative genomic hybridization in the estimation of whole-genome similarity and in the assessment of genetic relatedness using comparative genomics-based genotyping.

In this study, isolates have been analyzed using CGF-20, CGF-35, and CGF-119 assays, which target 20, 35, and 119 VAP genes respectively, and have shown that the PCR-based CGF assays are flexible assays and can be easily modified by the addition of genes, with no more additional work than optimizing a PCR, while leaving the remaining fingerprinting information easily accessible for strain to strain or inter-laboratory comparison. Although the term 'comparative genomic fingerprinting' has not yet been universally applied, the importance of CGF as an emerging method for genotyping has been established through work carried out in our laboratory on vero-toxigenic *Escherichia coli* (Laing *et al.*, 2008).

In this thesis, the technical reproducibility of the CGF-20 assay was tested two-fold: a) by comparing results using traditional agarose gel electrophoresis versus capillary electrophoresis and b) by comparing results of technical replicates for a subset of strains. Both comparisons showed that the PCR-based CGF assay is extremely robust, with greater than 97% of signal calls being identical in both tests (a: 1893/1920 calls, or 98.6%

concordance; b: 1881/1920 calls, or 97.9% concordance). The small discrepancies observed may be explained by the higher resolution (within 3-5 bp) and enhanced detection sensitivity (as low as 0.1 ng/µL of PCR product) of the QIAxcel instrument compared to traditional agarose gel electrophoresis and ethidium bromide/UV band detection. Although the gel-based CGF-20 can be completed relatively quickly in most laboratory settings (*i.e.,* 96 samples in 4-5 days), the capillary electrophoresis system increases the accuracy of signal calls due to enhanced resolution and sensitivity, and greatly increases the throughput of the method (*i.e.,* 96 samples in 1-2 days).

A small difference between CGF-20 fingerprints between PCR replicates was observed. This difference was small (~3.7% of PCR amplicons) and is likely attributed to minute differences between PCR reactions including concentrations of PCR components including primers, and temperature depending on location in heat block. These variations between PCRs are unavoidable but may still lead to changes within the assay affecting the ability of primers to anneal, reduce extension or otherwise affect the overall yield of PCR product. As the majority of differences were in a single PCR, it is important to note that reduction of cluster stringency may compensate and still provide effective clustering of closely related strains.

In this study, CGF assays targeting a larger number of loci that deliver increased sub-typing resolution have also been examined. These assays were used to assess the ability of the CGF-20 assay to rapidly cluster strains based on high levels of genetic similarity. The results of the study show that strains within the same CGF-20 cluster have much higher levels of genetic similarity, as determined by similarity in CGF-35 and CGF-119 fingerprints, compared to strains outside of the cluster. The data suggests that

the CGF-20 assay groups isolates with high genetic similarity and is highly correlated with additional genetic content. Interestingly, the high resolution also revealed genomic differences consistent with epidemiological information. For example, CGF-20 Cluster #202 contained 6 clinical isolates that show a high overall similarity (90.0%) in CGF-119 fingerprints. Three of the clinical strains showed significant differences at 8 loci between cj1337 and cj1630. These strains were obtained between one-to-two months after the other clinical isolates in the cluster were sampled.

Our analysis of genetic similarity using CGF-119 data demonstrates that the CGF-20 method is able to reliably cluster isolates with high genetic similarity and is thus a useful typing method for *Campylobacter jejuni.* The CGF-20 assay was shown to generate *functionally equivalent* data to the CGF-119 assay in that strains that are similar using the CGF-20 assay also have extremely similar CGF-119 fingerprints. The CGF-119 assay is based on targeting nearly half of all the variably absent/present genes in the *C. jejuni* genome identified by whole-genome comparative genomic hybridization using microarrays (Taboada *et al.*, 2004) and has been designed to replicate whole-genome microarray data (E. Taboada, personal communication). Since there is high correlation between CGF-20 and CGF-119 data, it can be inferred that the CGF-20 assay can be used to rapidly assess whole genome genetic similarity between strains. Consequently, I will concentrate on epidemiological analysis of the CGF-20 data in subsequent sections of the thesis.

An advantage to CGF compared to other methods and, in particular, to band-based fingerprinting methods, is that analysis at reduced stringencies can retain biologically meaningful ‘inferred genetic relatedness’. Although CGF fingerprints can

only be used to estimate the genetic content of a strain, they appear to provide a relevant assessment of overall genetic relatedness. For example, strains with 90% matching CGF-35 fingerprints are identical at 32 out of 35 loci and, by extrapolation, would also be expected to have an overall genetic similarity (*i.e.,* at higher levels of resolution) on the order of 90%. By contrast, it is much more difficult to extrapolate the overall genetic similarity between strains that have 90% matching fingerprints based on the analysis of data from band-based methods such as PFGE and *fla*A-RFLP or based on methods that interrogate few loci, such as *fla*A-SVR, sequencing because of lower levels of genome sampling and the relatively low level of information content generated in those methods. The ability to analyze data at various stringencies of clustering while retaining biologically and genetically meaningful information may give CGF a wider range of applicability.

CGF-20 can be performed at a fraction of the time and cost of other molecular typing methods, such as MLST and PFGE. CGF-20 fingerprints tend to be non-ambiguous (*i.e.*, the gene is present or absent) compared to other methods (*i.e.*, restriction fragment-based fingerprints requiring less transparent algorithms to establish relationships). The reproducibility testing demonstrates that, not only are the results robust and easily reproducible, but that binary fingerprints based on presence/absence data could easily be compared among institutions.

Chapter 4

4. Molecular epidemiology of *Campylobacter jejuni* in southern Alberta

4.1 Introduction

Southern Alberta has nearly twice the incidence of campylobacteriosis compared to that found both in northern Alberta and across Canada. This region contains extensive livestock production which may contribute to the greater levels of infection with livestock-associated pathogens than in other regions. Cases of campylobacteriosis are thought to be generally sporadic, and of foodborne origin (Kapperud et al., 1992; Mead et al., 1999). To gain insight into the high incidence rate of campylobacteriosis in southern Alberta, 592 isolates of *Campylobacter jejuni* were obtained from water, animal fecal and clinical samples in the region over a three year period (2004-2006), and analyzed using a *fla*A-RFLP molecular typing assay and were also analyzed using the 20 gene comparative genomic fingerprinting assay (CGF-20) previously described and validated in Chapter 3. The purpose of this study was to evaluate the sample set by analyzing the genotyping data within an epidemiological framework to identify potential environmental reservoirs of *C. jejuni*, potential routes of transmission into the human population, and to identify prevalent or endemic genotypes (*i.e.,* clones), if any, present within the region using CGF. This study aims to provide the proof-of-concept for the application of CGF in molecular epidemiological studies.

4.2 Results

4.2.1 Evaluation of samples from southern Alberta

C. jejuni were isolated from water, animal fecal and clinical samples obtained in the southern Alberta; *i.e.*, out of 834 water samples processed, 91 (10.9%) were culture positive for *C. jejuni*, as compared to the 30.0% isolated from animal fecal samples (241 out of 803 samples). The combined data set also contained 260 clinical isolates from Clinical Laboratory Services (Calgary, Alberta). Due to confidentially issues, the exact date of isolation for some samples could not be obtained; *i.e.*, isolation dates for 16 of the 2005 clinical isolates and all 34 isolates from 2006 were not made available. In addition 49 reference strains used in-house, most of which have been characterized by whole-genome comparative genomic hybridization (Taboada *et al.*, 2004; Taboada *et al.*, 2008), were included in the final data set of 641 *C. jejuni* isolates as controls.

Isolation rates for *C. jejuni* varied between animal species (Table 5). Almost half (44%) of chicken fecal samples contained *C. jejuni*. Most other animal species sampled also had similar isolation rates for *C. jejuni*. These included buffalo, pig, deer, duck, cattle and sheep with isolation rates of 57%, 46%, 45%, 45%, 37% and 33% respectively. Other avian sources, in addition to chicken, have been identified as significant contributors to *Campylobacter* infections. Surprisingly, *Campylobacter* isolates were obtained from only 20% of goose samples and 9% of other avian species including migratory birds such as seagulls and pelicans.

Table 5: Isolation rates of *C. jejuni* from animal fecal samples. Total numbers of fecal samples tested for each animal source are listed, as well as the number of *C. jejuni* isolates obtained, and the percentage of samples that were culture positive for *C. jejuni.* Only one *C. jejuni* isolate for each sample was used for CGF-20 genotyping although up to three *C. jejuni* suspect colonies were obtained during the isolation procedure until confirmation by PCR.

Animal	Number of Samples	Number of C. jejuni Isolates	Percent of Samples Tested
Cattle	218	80	36.7%
Chicken	98	43	43.9%
Sheep	82	27	32.9%
Pig	59	28	47.5%
Duck	38	17	44.7%
Goose	81	17	21.0%
Horse	81	8	9.9%
Goat	33	5	15.2%
Deer	11	5	45.5%
Other Avian	45	4	8.9%
Buffalo	7	4	57.1%
Human	2	2	100.0%
Domestic Dog	34	1	2.9%
Domestic Cat	14	1	7.1%
Total	803	214	26.7%

4.2.2 Epidemiological analysis of *C. jejuni* isolates using a CGF-20 assay

A total of 268 unique 20 gene fingerprints or profiles were obtained by analysis of all 641 isolates of *C. jejuni* included in our collection (Figure 7 or Appendix B). Using a fingerprint similarity threshold of 100% for clustering (*i.e.,* all 20 genes evaluated must be identical in conservation status), a total of 268 clusters were obtained. These clusters resolved to an average of 2.4 ± 3.8 members per cluster, with the largest cluster containing 29 isolates (4.5% of all isolates). Overall, 184 clusters contained a single isolate (29% of all isolates) and were identified as 'singletons'. Of the 49 reference strains, 18 appeared as singletons, reducing the total number of unique fingerprints observed among the 592 southern Alberta isolates to 250.

Of the total 268 clusters obtained for *C. jejuni*, 226 were exclusively composed of one or more isolates from the same source; *i.e.*, clinical, animal and water sources. Each had 89, 87, and 50 respective clusters composed exclusively of that sample type (Figure 13). Animal isolates shared profiles exclusively with clinical isolates in 28 clusters, and shared profiles exclusively with water isolates in four clusters. Water isolates shared profiles exclusively with clinical isolates in two clusters. A total of eight clusters were composed of isolates from all three sources (clinical, animal fecal and water).

Many of the *C. jejuni* isolates from water (51.6%, n=91) had unique fingerprints (Figure 13). However, isolates from several clusters were isolated at a similar time and/or geographical region. Out of 8 clusters (containing 32/91 water isolates) that contain two or more isolates from multiple water samples, seven clusters (representing 30/91 isolates) contained isolates from the same location obtained within a two month period, or that had been isolated from the same location at the same time. The majority of isolates from

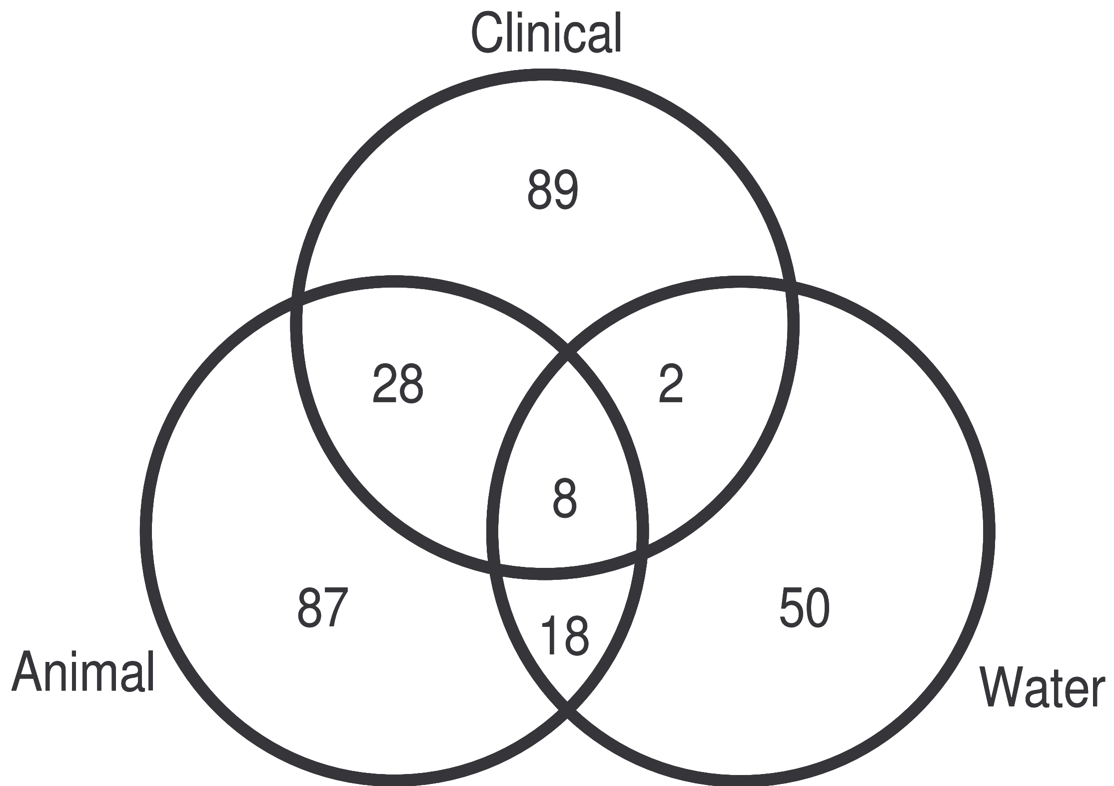


Figure 13: Analysis of CGF-20 cluster composition by source type. Strains from the three source types (Animal fecal, water or clinical sample) were sub-divided into 250 CGF-20 clusters (including singletons). Clusters resolved to an average 2.4 ± 3.8 members per cluster, with the largest cluster containing 29 isolates (4.5% of all isolates). A total of 184 clusters contained a single isolate (29% of all isolates) and were considered as 'singletons'. Among the 66 multi-strain clusters, 42 were composed of strains from more than one source, and eight clusters were composed of strains from all three sources.

water grouped according to sampling site and date as would have been expected epidemiologically.

Several of the animal and water isolates appeared linked based upon similarities in their CGF-20 profiles. The water isolates were similar to geese (35.3%, n=17 isolates), and duck (35.3%, n=16 isolates) isolates in support of the idea that migratory birds contribute significantly to *Campylobacter* spp. contamination of the natural environment, and specifically bodies of water. In addition, water isolates were similar to isolates from cattle (25.0%, n=80), deer (40.0%, n=5), and horses (37.5%, n=8). Feces from these species may have entered water through run-off from pastures or feedlots following snow melt or precipitation. Other bird and animal species sharing CGF-20 profiles with water isolates in this study were isolates from chicken (16.3%, n=43 isolates), sheep (14.8%, n=27 isolates), and other avian species (25.0%, n=4).

Isolates obtained from several animal sources had identical CGF-20 fingerprints to clinical isolates (Figure 14). Animal isolates with CGF-20 fingerprints identical to those from clinical specimens included those from chicken (58.1% of all chicken isolates), cattle (51.3% of all cattle isolates), sheep (59.3% of all sheep isolates), pig (44.0% of all pig isolates), goose (37.5% of all geese isolates) and ducks (35.3% of all duck isolates). Overall, 36.7% of cattle, 43.9% of chicken, 32.9% of sheep, 45.8% of pig, 19.8% of goose, and 44.7% of duck fecal samples contained *C. jejuni* isolates with an identical CGF-20 fingerprint to a clinical isolate in this study.

Several large clusters of clinical isolates also were detected based upon similarities in their CGF-20 profiles (Figure 15). Eight of these clusters contained 39.2%

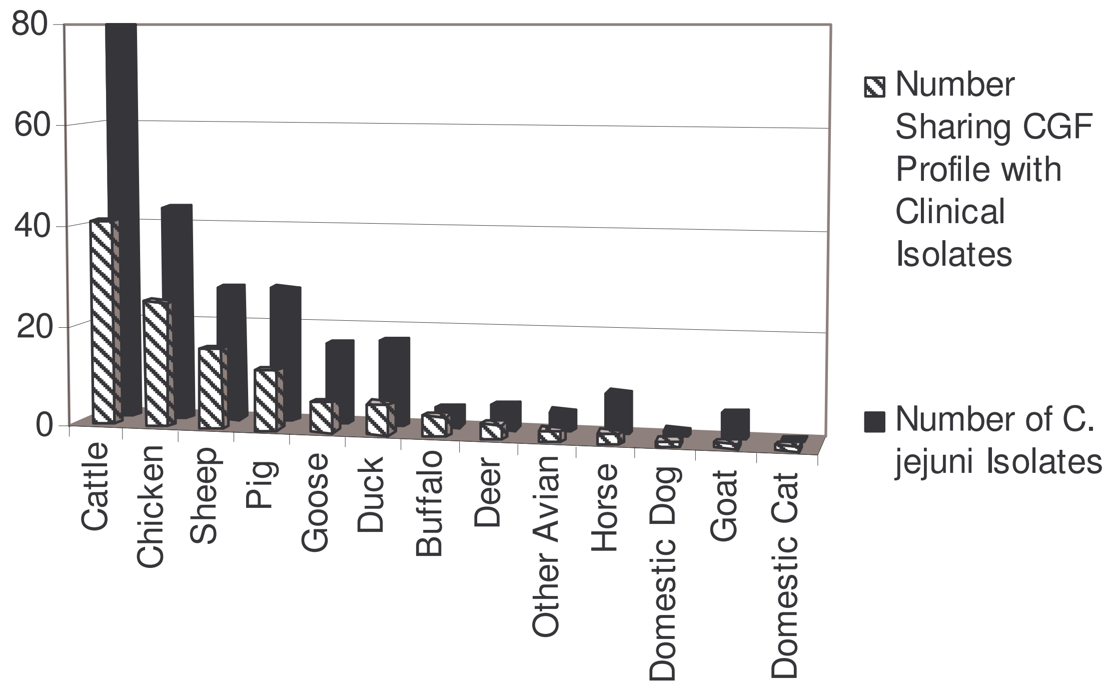


Figure 14: *C. jejuni* isolates from animal sources with matching CGF-20 fingerprints to human clinical isolates. Nearly 50% of *C. jejuni* isolates from animal fecal samples share identical CGF profiles to human clinical isolates.

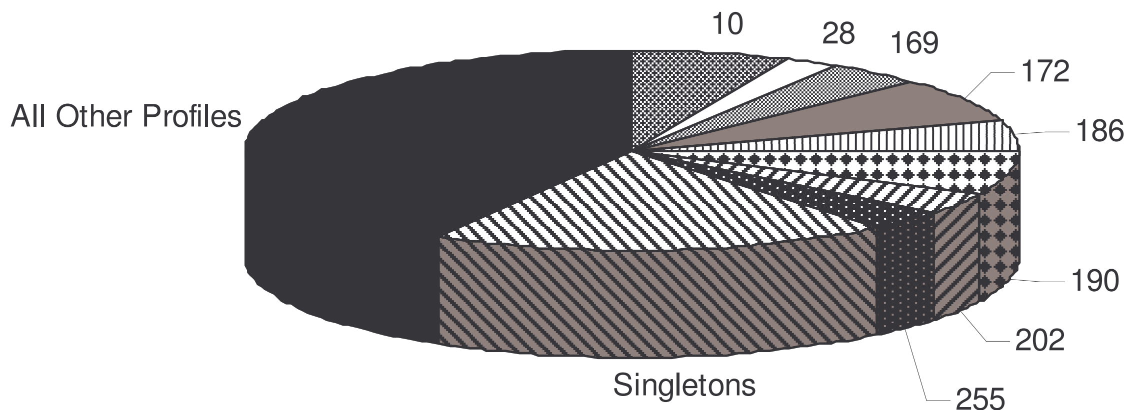


Figure 15: Proportion of clinical isolates by CGF-20 cluster represented as a percentage of all clinical isolates in the dataset (n=260).Eight CGF-20 clusters (CGF-20 Clusters 10, 28, 169, 172, 186, 190, 202 and 255) contain 39.2% of total clinical isolates, while 19.2% of all the clinical isolates appeared as singletons.

(n=260) of all clinical isolates, while 19.2% (n=260) of all clinical isolates appeared as singletons**.**

4.2.3 Epidemiological analysis of *C. jejuni* isolates using a *fla*A-RFLP assay

A total of 182 unique *fla*A-RFLP fingerprints were obtained by analysis of the 641 isolates of *C. jejuni* in our collection. The largest cluster contained 66 members (10% of all isolates) with 122 clusters containing only a single member – *i.e.,* "singletons" (19% of all isolates). On average, analysis of the *fla*A-RFLP data resulted in approximately 3.5 ± 7.2 members per cluster.

Of the 182 *fla*A-RFLP clusters, 146 were composed exclusively of one sample type. Animal, clinical and water sources each had 61, 46, and 39 respective clusters composed of that sample type (Figure 16). Animal isolates shared profiles exclusively with clinical isolates in 19 clusters, and shared profiles exclusively with water isolates in 8 clusters. Water isolates shared profiles exclusively with clinical isolates in a single cluster. A total of eight clusters were composed of isolates from all three sources (clinical, animal fecal and water).

Many of the *C. jejuni* isolates from water (38.5%, n=91) had unique fingerprints (Figure 16). However, several clusters were isolated at a similar time and/or geographical region. Out of 13 clusters of the bacteria obtained from multiple water samples, ten clusters (containing 45/91 isolates) contained *C. jejuni* isolated from the same location within a 2 month period, or that had been isolated from the same location at the same time. Seven clusters, representing seven of 91 isolates, contained a single water isolate that grouped with isolates from different sources and 35 water isolates had unique

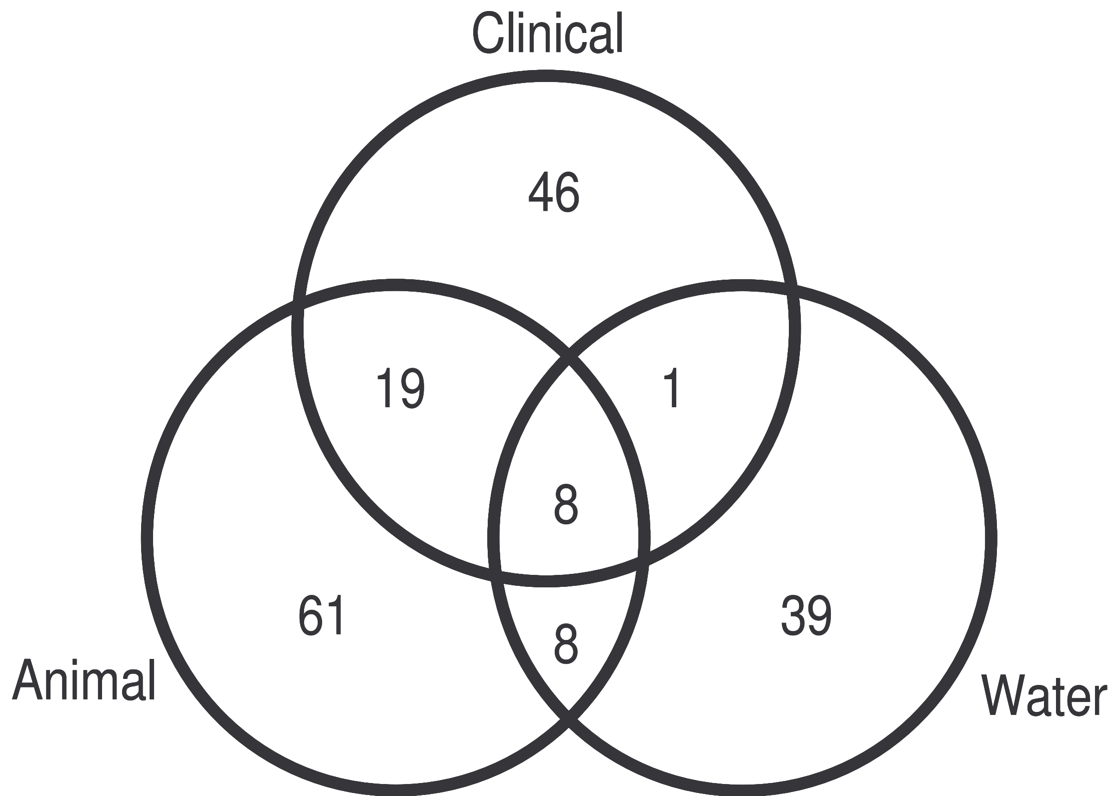


Figure 16: Analysis of *flaA* RFLP cluster composition by source type. Strains from the three source types (Animal fecal, water or clinical sample) were sub-divided into 182 *flaA*-RFLP clusters (including singletons). Clusters resolve to average 3.5 ± 7.2 members per cluster, with a total of 146 clusters contain singletons. Among the remaining 60 multi-strain clusters, thirty-six were composed of strains from more than one source, and eight clusters were composed of strains from all three sources.

fingerprints. The majority of water isolates grouped according to site and sampling date.

Several of the animal and water isolates appeared linked based upon similarities in their *fla*A-RFLP profiles. Host species with isolates most commonly similar *fla*A-RFLP profile to water isolates were geese (64.7%, n=17 isolates) and ducks (50.0%, n=16 isolates). Similarly, cattle (26.3%, n=80 isolates) commonly matched *fla*A-RFLP profile to water isolates. As mentioned above (see 4.2.2) this may be related to water contamination with cattle manure following run-off associated with snow melting or precipitation. Other species sharing *fla*A-RFLP profiles with water isolates were chicken (14.0%, n=43 isolates), sheep (11.1%, n=27 isolates), pig (7.1%, n=28 isolates) and cat (100.0%, n=1 isolate) fecal samples.

Isolates obtained from several animal fecal sources had identical *fla*A-RFLP fingerprints to clinical isolates (Figure 17). Animal isolates with *fla*A-RFLP fingerprints identical to those from clinical specimens included chicken (67.4% of all chicken isolates), cattle (56.3% of all cattle isolates), sheep (96.3% of all sheep isolates), pig (46.4% of all pig isolates), goose (29.4% of all geese isolates) and duck (29.4% of all duck isolates). Overall, 20.6% of cattle, 29.6% of chicken, 31.7% of sheep, 22.0% of pig, 28.6% of buffalo, and 27.3% of deer fecal samples contained a *C. jejuni* isolate in the same *fla*A-RFLP cluster.

Several large clusters of clinical isolates also were detected based upon similarities in their *fla*A-RFLP profiles (Figure 18). Two of these clusters contained 28.5% (n=260) of all the clinical isolates, and seven additional clusters contained 51.1% (n=260) of all the clinical isolates.

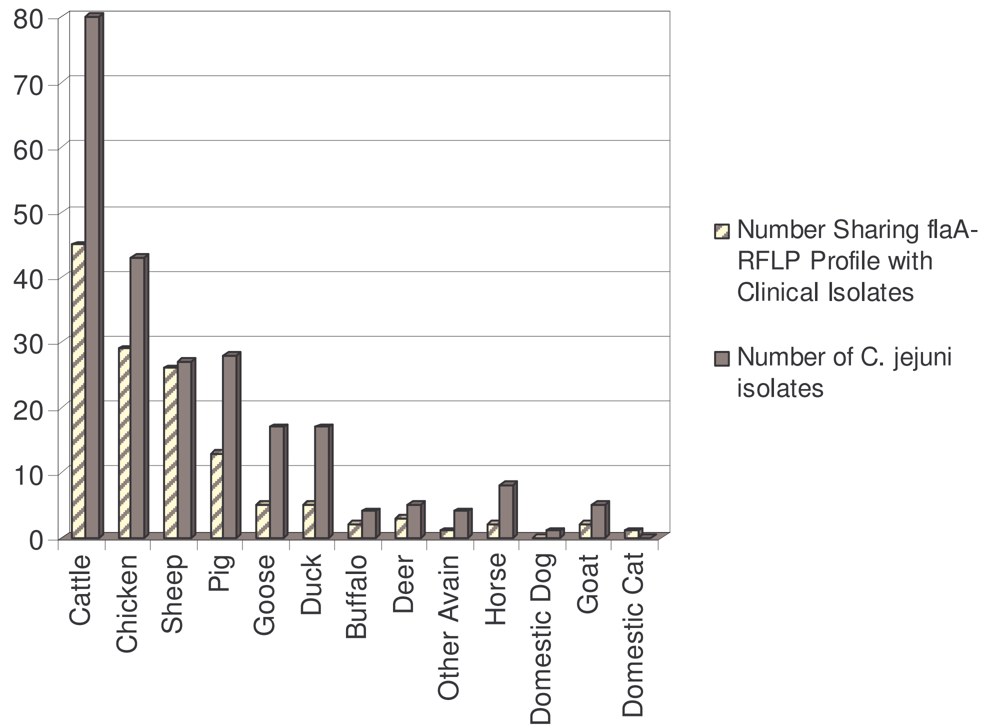


Figure 17: *C. jejuni* isolates from animal sources with matching *fla*A-RFLP fingerprints to human clinical isolates. Nearly 50% of *C. jejuni* isolates from animal fecal samples share identical CGF profiles to human clinical isolates.

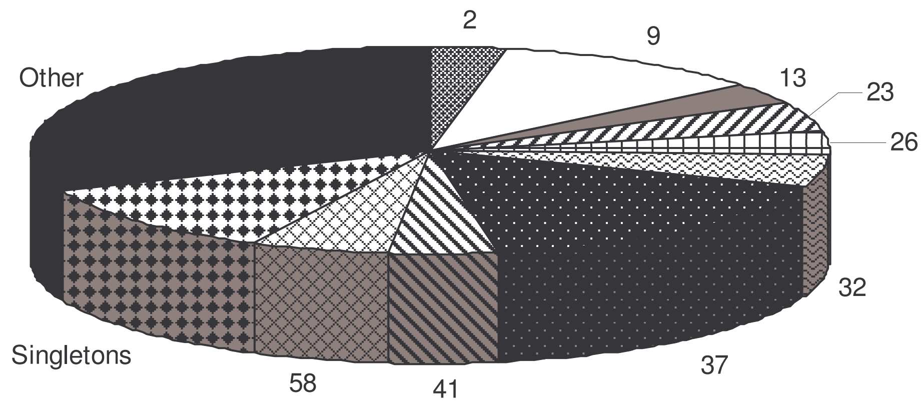


Figure 18: Clinical isolates arranged by *fla*A-RFLP type. The majority of the 260 clinical isolates are found in nine *fla*A-RFLP types (57.0%) and 'singletons' (11.5%), isolates that do not cluster with any other isolates from the data set of 641 isolates including isolates from water and animal fecal samples in southern Alberta. The remaining isolates (31.5%) comprise a total of 29 *fla*A-RFLP types.

4.2.4 Comparison of CGF-20 and *fla*A-RFLP typing results

4.2.4.1 Statistical comparison of CGF-20 and *fla*A-RFLP

The Index of Diversity (ID), based on Simpson's Index of Diversity is used for quantification of the biodiversity of a system, was calculated in order to provide an objective assessment of the level of diversity suggested by the CGF-20 and *fla*A-RFLP genotyping results. Both methods produced a high ID, indicating that on average, members were assigned to small clusters, with a high likelihood of genetic relatedness, while non-related isolates could be discriminated and allocated into separate clusters. Despite some differences in the average and largest cluster size, both the CGF-20 and *fla*A-RFLP methods had essentially indistinguishable IDs; *i.e.*, the ID for CGF-20 was 98.86 while that for *fla*A-RFLP was 97.31.

In order to further compare the results obtained from the CGF-20 and *fla*A-RFLP genotyping methods, concordance was assessed by evaluating the corresponding *fla*A-RFLP types for isolates within a particular CGF-20 cluster and vice-versa. Sixty *fla*A-RFLP clusters consisted of more than one isolate. Of these, 9 (15.0%) were composed of isolates with a single CGF-profile. These nine clusters were very small and ranged from 2 to 3 members averaging 2.2 members per cluster. When the reverse analysis was performed, 27 (32.1%) of the 84 multi-member CGF clusters in the dataset were composed of isolates with a single corresponding *fla*A-RFLP profile (Table 6). These twenty-seven CGF-20 clusters were also small, ranging from 2 to 9 members, and averaging 3.3 members per cluster. Thus, despite sharing similar discriminatory power,

Table 6: Distribution of *fla*A-RFLP types among CGF-20 clusters composed of a single *fla*A-RFLP type. Twenty-seven of eighty-four multi-member CGF-20 clusters were composed of isolates with a single *fla*A-RFLP type; five of these *fla*A-RFLP types appear in multiple CGF-20 clusters.

***fla*A-RFLP Type**	**CGF-20 Cluster(s)**
7	5
9	255, 256, 260, 261, 262
11	102
12	3
13	6, 8
17	40
23	22
26	26
27	9, 248
30	221
37	171, 195
41	164
46	223
57	23
58	20, 74, 75, 81
59	237
70	70

CGF-20 appeared to generate clusters that were more likely to be composed of isolates indistinguishable by a secondary typing method (*i.e*., *fla*A-RFLP).

There were several CGF-20 clusters that were comprised of members sharing the same *fla*A-RFLP profile, but where the *fla*A type was also observed in other CGF-20 These *fla*A-RFLP types (type 9, 13, 27, 37 and 58) demonstrate that, while some CGF-20 clusters are homogenous for a single *fla*A allele, the alleles need not be exclusive to a given CGF-20 profile. Larger CGF-20 clusters, such as cluster #60, 70, 71, 202 and 250, were often composed of isolates sharing a dominant *fla*A allele, but which may also include isolates with secondary *fla*A-RFLP alleles (Table 7). These dominant *fla*A alleles also appear in several different CGF-20 clusters.

Although certain CGF-20 clusters were composed of isolates with a single or a dominant *fla*A-RFLP allele, other CGF-20 clusters (*i.e.,* CGF-20 cluster #10, 37, 42, and 137) were composed of isolates with a variety of different *fla*A-RFLP alleles (Table 7). For example, CGF-20 cluster #10 contains 27 members with nine distinct *fla*A-RFLP types, CGF cluster # 37 contains four members, each with a distinct *fla*A-RFLP type, and CGF #137 contains 16 members, each with a distinct *fla*A-RFLP type that is unique in the dataset and appears as a singleton.

Taken together, these findings suggest that while there is a significant association between *fla*A alleles and the genetic background as assessed by the CGF-20 profile, the same *fla*A allele can be found in isolates with very different genetic backgrounds and that, conversely, isolates with similar genetic backgrounds can have different *fla*A alleles.

Table 7: *fla*A-RFLP composition of multi-member CGF-20 clusters. Several large CGF-20 clusters contained a dominant *fla*A-RFLP type as well as additional secondary *fla*A-RFLP types that appear to be unrelated whereas other clusters were composed of several *fla*A-RFLP types, none of which appeared to be dominant.

CGF Cluster	Number of isolates	Dominant *fla* A-RFLP type	Percent Dominant *fla* A-RFLP type	Additional *fla* A-RFLP Types
10	27	23	85.2%	14, 20, 23, 25, 48, 49, 50, 54, 102
37	4	47	25.0%	38, 47, 66, 96
42	14	38	28.6%	17, 23, 34, 37, 38, 47, 90
60	11	41	63.6%	22, 28, 67
70	9	70	100.0%	
71	10	59	60.0%	58, 60
137	16	131	6.3%	131, 132, 133, 134, 135, 143, 145, 146, 147, 148, 149, 150, 151, 152, 160, 161
202	16	2	87.5%	82, 136
250	7	15	71.4%	29, 77

4.2.5 Assessment of the ability of CGF-20 and *fla*A-RFLP assays to predict overall genetic similarity among *C. jejuni* isolates

Overall genetic similarity among a subset of *C. jejuni* isolates was assessed using the CGF-119 fingerprinting method, which has been determined to provide an excellent picture of overall genome content of strains (E. Taboada, personal communication). Similarity in CGF-119 genetic profiles was calculated as a proportion of matches at the 119 genes represented in the ArrayTube hybridization assay. In order to assess the ability of the CGF-20 and the *fla*A-RFLP assays to predict overall genetic similarity, the average CGF-119 similarity between members of the same cluster (*intra-cluster similarity*) was then compared to the average similarity versus all other strains in the dataset (*inter-cluster similarity*) for groups of strains with the same *fla*A and/or CGF-20 type.

The large multi-member *fla*A-RFLP clusters (*fla*A-RFLP types 2, 37, 41 and 52) had an overall intra-cluster similarity of 76.3%, with an inter-cluster similarity of 58.0% (Table 8). A similar low level of inter-cluster similarity (54.7%) was observed for strains analyzed by CGF-20 (Table 9). When examining the strains within the same CGF-20 cluster an average intra-cluster similarity of 85% was observed. Taken together, this suggests that while strains of different CGF-20 or *fla*A-RFLP clusters tend to have a similarly low level of genetic similarity (*i.e.,* ~ 60% identical CGF-119 profiles), strains within a given CGF-20 or *fla*A-RFLP cluster are likely to have much higher levels of genetic similarity (*i.e.,* ~ 75-85% identical CGF-119 profiles), with strains within a CGF-20 cluster having a marginally higher level of genetic similarity than strains within the same *fla*A-RFLP cluster.

Table 8: Average intra-cluster and inter-cluster genetic similarities for multi-member CGF-20 clusters. Pair-wise similarities in CGF-119 fingerprints were evaluated for intra-cluster (comparisons made between isolates within a cluster) and inter-cluster (comparisons made between isolates external to the cluster). Standard deviations are given as calculated using Microsoft Excel.

		Intra-Cluster			**Inter-Cluster**		
Cluster Number	# Members	# Pair-wise Matches	Average Matches	% Average Matches	# Pair-wise Matches	Average Matches	% Average Matches
10	5	20	100.4 ± 5.8	84.4	345	66.8 ± 11.0	56.1
37	3	6	96.0 ± 7.8	80.6	213	64.8 ± 11.5	54.5
42	4	12	94.3 ± 6.7	79.2	280	66.7 ± 11.0	56.0
60	9	72	99.1 ± 12.0	83.2	585	66.4 ± 11.1	55.8
137	9	72	109.6 ± 4.0	92.1	585	48.8 ± 19.0	41.0
161	4	12	71.5 ± 16.2	60.0	280	61.3 ± 18.4	51.5
163	7	42	82.4 ± 28.5	69.2	469	68.4 ± 21.4	57.5
169	2	2	33.0 ± 0.0	27.7	145	100.0 ± 20.4	84.0
172	9	72	88.8 ± 6.4	74.5	585	60.1 ± 19.0	50.5
183	2	2	100 ± 0.0	84.0	145	48.7 ± 20.4	40.9
190	9	72	96.1 ± 16.1	80.7	585	57.1 ± 20.1	48.0
202	9	72	107.1 ± 8.1	90.0	585	67.9 ± 10.8	57.1
255	2	2	102.0 ± 0.0	85.7	145	63.9 ± 12.0	53.7

Table 9: Average intra-cluster and inter-cluster genetic similarities for multi-member *fla*A-RFLP clusters. Pair-wise similarities in CGF-119 fingerprints were evaluated for intra-cluster (comparisons made between isolates within a cluster) and inter-cluster (comparisons made between isolates external to the cluster). Standard deviations are given as calculated using Microsoft Excel.

		Intra-Cluster			Inter-Cluster		
Cluster Number	# Members	# Pair-wise Matches	Average Matches	% Average Matches	# Pair-wise Matches	Average Matches	% Average Matches
2	7	42	107.5 ± 5.9	90.3	469	71.8 ± 11.9	60.3
9	2	2	103.0 ± 0.0	86.6	144	67.5 ± 12.3	56.7
17	2	2	105.5 ± 0.0	88.7	144	71.2 ± 11.3	59.9
21	2	2	98.0 ± 0.0	82.4	144	72.3 ± 15.3	60.7
23	2	2	65.0 ± 0.0	54.6	144	68.1 ± 15.3	57.2
26	4	12	71.3 ± 28.8	59.9	280	62.3 ± 19.6	52.4
28	2	2	84.5 ± 0.0	71.0	144	62.3 ± 19.6	63.0
32	2	2	107.0 ± 0.0	89.9	144	70.6 ± 22.8	59.3
37	10	90	87.6 ± 10.9	73.7	640	64.8 ± 20.5	54.4
38	3	6	86.8 ± 4.6	73.0	213	69.7 ± 11.3	58.6
41	10	90	93.1 ± 14.7	78.3	630	71.9 ± 13.7	60.4
52	6	30	65.8 ± 22.6	55.3	408	64.2 ± 18.6	53.9
56	2	2	105.5 ± 0.0	88.7	144	65.2 ± 23.9	54.8

Although the average intra-cluster similarities between strains of the same CGF-20 or *fla*A-RFLP cluster are similar, the similarity distributions of all intra-cluster pair-wise comparisons shows that there is a significant difference between both methods (Figure 19). The CGF-20 intra-cluster similarity distribution is skewed to the right, with most strains being distributed around a mean of ~ 90%. By contrast, the *fla*A-RFLP intra-cluster similarity distribution appears to be bi-modal, with one peak at ~ 90% and an equally prominent peak at ~ 70%. This secondary peak is due to a small number of cases in which strains that are quite genetically distinct (*i.e.,* low CGF-119 similarity) share the same *fla*A-RFLP allele. Similarly, when the inter-cluster pair-wise similarity distributions were compared, it is possible to see that although both the CGF-20 and *fla*A-RFLP assays generate nearly identical distributions, both peaking at ~60% (Figure 20), the *fla*A-RFLP inter-cluster distribution had a slight right-handed tail. This tail could be due to a small number of cases in which strains with unrelated *fla*A-RFLP alleles share significant genetic similarity.

4.2.6 Epidemiological comparison of the CGF-20 assay and *fla*A-RFLP assay

An important consideration in molecular epidemiological studies is the concept of genotyping resolution because sufficient resolution is necessary in order to link epidemiologically related strains while excluding unrelated strains. Although the CGF-20 assay and the *fla*A-RFLP assay both have similar levels of ID (Index of Diversity), the CGF-20 assay tends to yield more clusters of smaller size and thus, appears to have increased resolution. Consequently, it is important to determine if this increased level of resolution is useful within an epidemiological context.

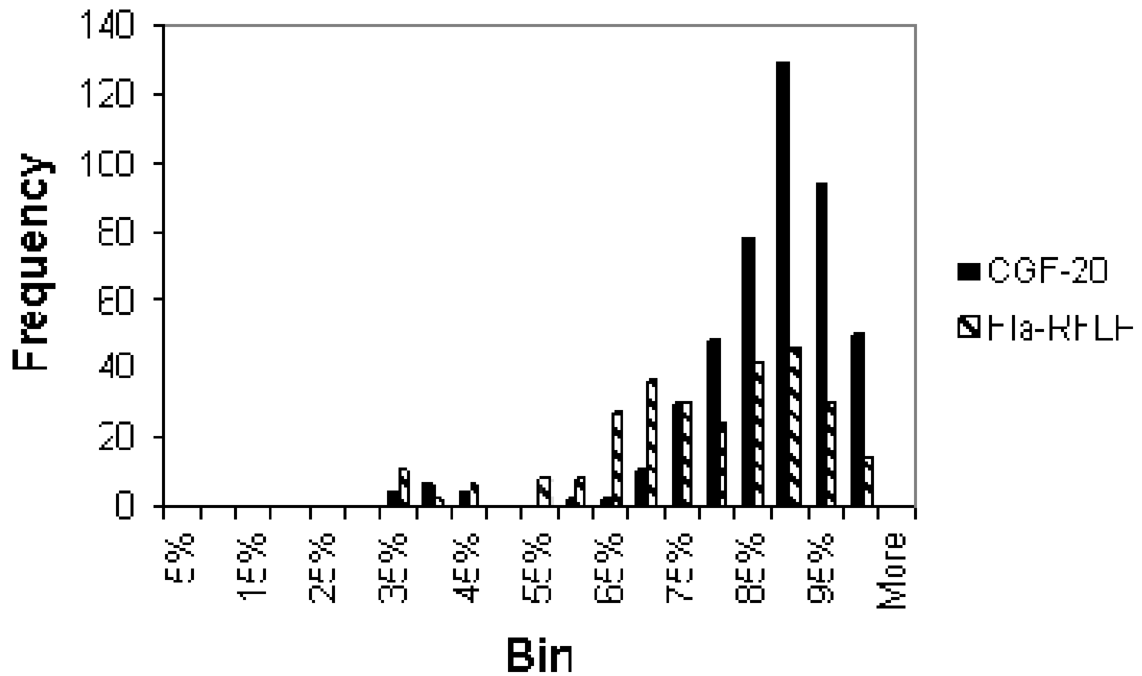


Figure 19: Distribution of intra-cluster CGF-119 similarities for all pair-wise comparisons between the 74 isolates selected for CGF-119 analysis. The distribution of pair wise comparisons of the CGF-119 shows a higher frequency of high similarity matches which peaked at ~ 90% for the CGF-20 clusters. Although a similar peak was seen for the *fla*A-RFLP clusters, a secondary peak of low similarity matches was also seen at ~ 70%.

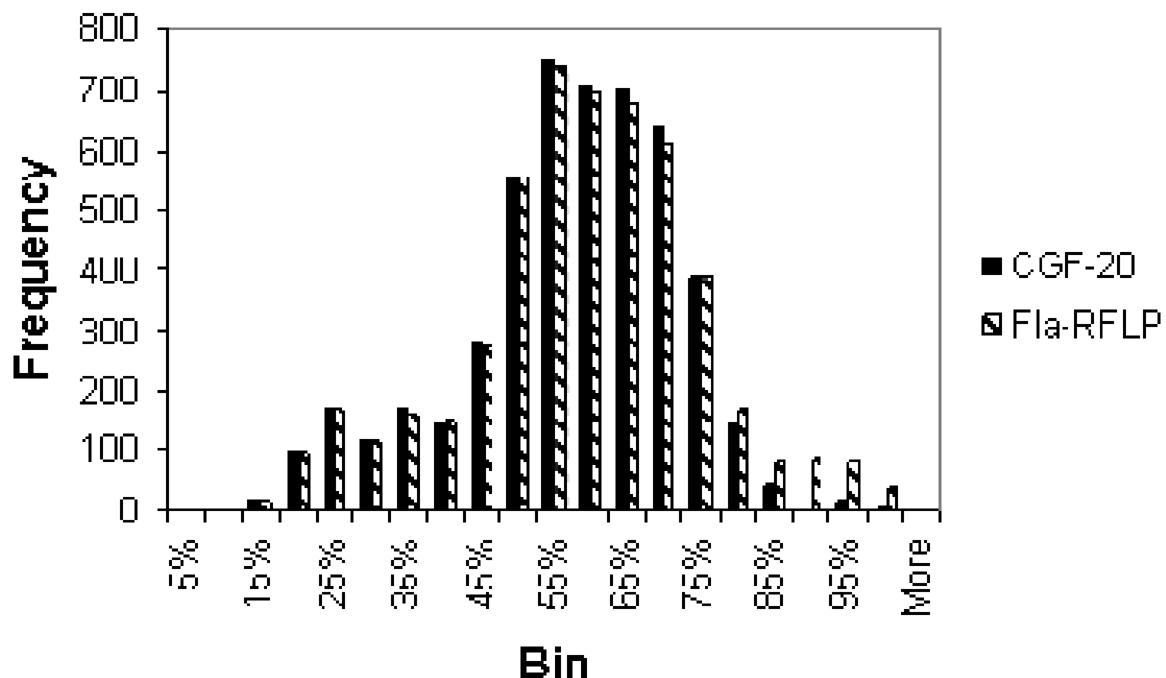


Figure 20: Distribution of inter-cluster CGF-119 similarities for all pair-wise comparisons between the 74 isolates selected for CGF-119 analysis. Although the distribution of percent similarity of pair wise comparisons between inter-cluster isolates was nearly identical for both CGF-20 and *fla*A-RFLP clusters, the right-hand of the distribution revealed a number of high-similarity matches between isolates of unrelated *fla*A-RFLP profiles (*i.e.*, in different clusters).

An example where the CGF-20 resolution correlates with epidemiological data is the small six member *fla*A-RFLP cluster #15, containing five chicken isolates from 2004 and one human clinical isolate from 2005 (Figure 21). Despite their unlikely epidemiological linkage, all of these isolates share the same *fla*A-RFLP allele and would be considered indistinguishable from each other. Based on their CGF-20 profiles however, the single clinical isolate (CJ339) differs from the chicken isolates at five of the 20 loci (75% similarity), which all share the same CGF-20 fingerprint (cluster CGF-20 #250). Given the underlying epidemiology of these isolates, the genetic differences observed between isolate CJ339 (cluster CGF-20 #181) and the chicken isolates would be best explained by lateral transfer of the *fla*A allele #16 into the CGF-20 #181 background. Interestingly, six of the seven remaining CGF-20 #181 isolates in the dataset share the same *fla*A-RFLP fingerprint (*fla*A-RFLP #42).

Evidence for lateral transfer, or *fla*A *allelic transfer*, is further supported by CGF-119 data. In a specific example, two strains with *fla*A-RFLP type #23 but from different CGF-20 clusters (#10 and #172) were evaluated using the CGF-35 and CGF-119 assays (Figure 22). Based on *fla*A-RFLP assay results alone, one would predict that the isolates share high genetic similarity. However, examination of the CGF-119 data shows differences at 41 loci, demonstrating an example of isolates with identical *fla*A alleles but low CGF-119 similarity, suggesting that the *fla*A similarity is likely to have arisen from allelic transfer of *fla*A between unrelated isolates.

Conversely, evidence can also be found of *fla*A *allelic replacement* between highly related *C. jejuni* strains (Figure 23). For example, CGF-20 cluster #202 is

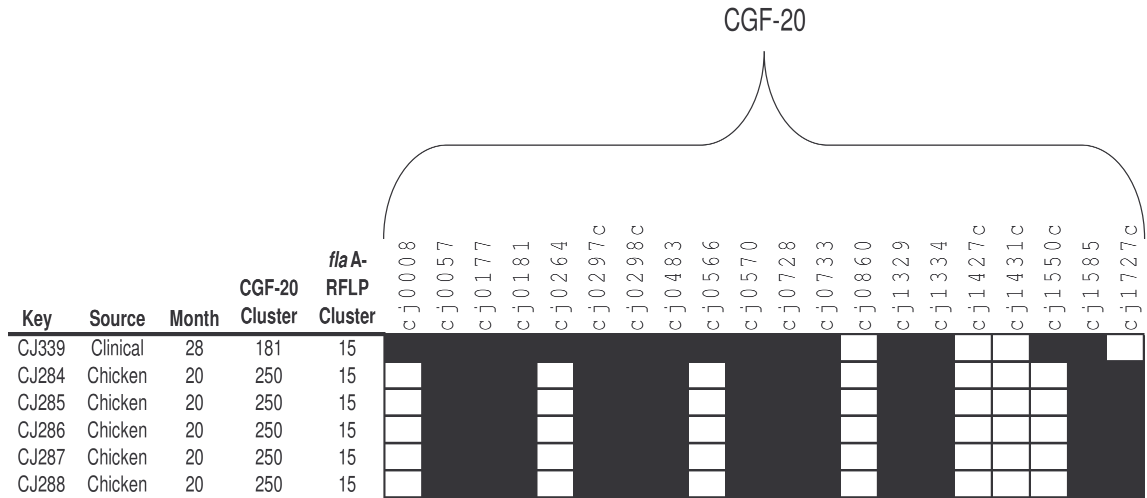


Figure 21: CGF-20 relationships and fingerprints of the 6 isolates in *fla*A-RFLP cluster #15. Although these isolates have a common *fla*A-RFLP type, the CGF-20 fingerprints between the chicken isolate and the clinical isolate differ at 5 out of 20 loci resulting in assignment to different CGF-20 clusters.

Figure 22: Evidence of lateral transfer, or *flaA allelic transfer*, is observed when strains with very different CGF profiles share the same *fla*A-RFLP allele. In this example isolates CJ389 and CJ127, members of separate CGF-20 clusters (#172 and #10, respectively) and with very different CGF profiles, share the same *fla*A-RFLP allele. Genes present or conserved are indicated in black, and genes that are absent are indicated in white.

Key	CJ279	CJ278	CJ277	CJ272	CJ271	CJ267	CJ266	CJ265	CJ264
Type	Cattle	Cattle	Water	Human	Human	Human	Human	Human	Human
Month	40	28	20	27.5	29	28	28	27	27
Fla A RFLP type	2	136	2	82	2	2	2	2	2
CGF-20 cluster	202	202	202	202	202	202	202	202	202

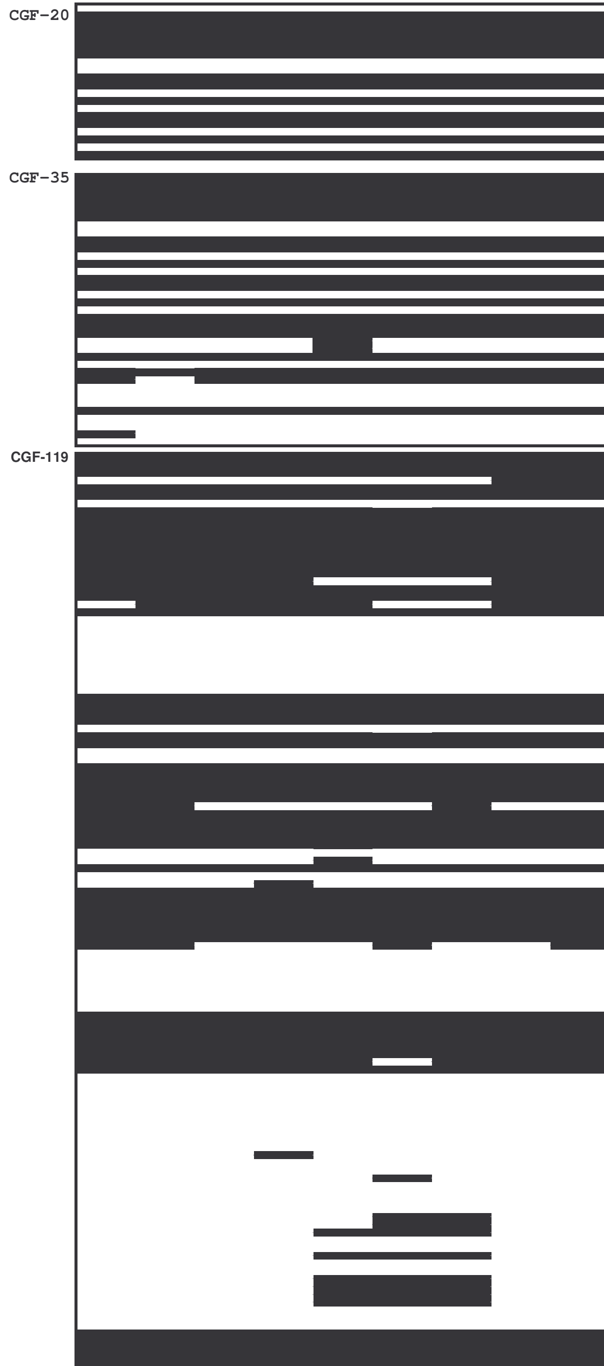


Figure 23: Evidence of *flaA allelic replacement* or exchange of *fla*A alleles between highly related *C. jejuni* isolates. CGF-20 cluster #202 was composed of isolates with three distinct *fla*A-RFLP alleles despite high similarity in CGF-20, CGF-35 and CGF-119 profiles. Genes present or divergent are indicated in black, absent genes in white.

composed of nine isolates, seven of which had the same *fla*A-RFLP type (*fla*A-RFLP type #2), with two isolates from additional *fla*A-RFLP types (*fla*A-RFLP types #82 and 136). Detailed genomic analysis using the CGF-119 assay shows very little variation in all nine isolates; *i.e.*, 105 out of 119 loci (88%) had an identical gene conservation status in all nine isolates. As only 14 loci (12%) varied between all nine strains, the evidence strongly suggests that the *fla*A allele has been replaced in strains that otherwise (*i.e.*, based on the majority of the hypervariable loci assayed by the CGF-119 assay) would appear to be highly related. Interestingly, three of the isolates (CJ266, CJ267, and CJ271) shared several additional polymorphisms not found in the other isolates.

Although these examples represent only anecdotal evidence for the usefulness of the increased resolution provided by the CGF-20 assay, and to an even greater extent the CGF-119 assay, the importance of validating such findings with relevant datasets with more detailed epidemiological information cannot be overstated.

4.2.7 Examination of highly prevalent CGF-20 genotypes among the southern Alberta data set

Although 111 distinct CGF-20 fingerprints were found among the 260 clinical isolates in the dataset (*i.e.,* the 260 clinical isolates segregate into 111 distinct CGF-20 clusters), six CGF-20 clusters (#10, 169, 172, 190, 202 and 255) contained 31.3% (81 of 260) of all clinical isolates in the dataset (Figure 24).

Selected isolates from these six CGF-20 clusters were analyzed using the CGF-119 assay in order to confirm genetic relationships and examine the genetic content of

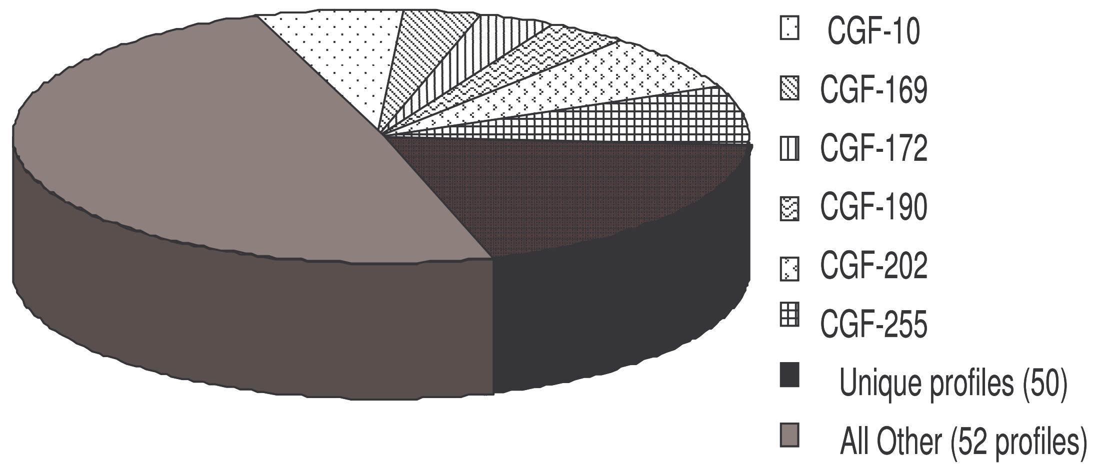


Figure 24: Proportion of clinical isolates by CGF-20 cluster represented as a percentage of all clinical isolates in the dataset (n=260). Approximately one-quarter of all strains appear as unique genotypes based on the CGF-20 fingerprint, while one third of the dataset is found within six fingerprints appearing over multiple time periods, which may suggest that these fingerprints could be persistent within the environment.

these isolates with a greater degree of resolution. CGF-119 data reveals that the isolates from a given CGF-20 cluster are more similar to one another than to isolates outside the cluster, suggesting that they represent genetically distinct clonal groups or strains of clinical importance in southern Alberta (Figure 25). The additional markers in the CGF-119 assay, however, can further discriminate the isolates within a CGF-20 cluster into sub-types based on the presence/absence of certain genes.

Interestingly, these six prominent "clinically-associated" CGF-20 clusters contain isolates from chicken, cattle, bison, and water. As these clinical isolates share CGF-20 profiles with isolates from each of these sources and can share high levels of genetic similarity when examined with the CGF-119 assay, it is likely that these non-human isolates share phenotypic properties, including the ability to cause illness in humans, with the human clinical isolates. This strongly suggests these various sources represent possible reservoirs for *C. jejuni* strains of public health importance.

Although all of the genotypes examined appear throughout the sampling period, several of the fingerprints (CGF-20 clusters #169, #172, #190, and #202) appeared to occur more often (Figure 26). For example, isolates belonging to CGF-20 cluster #10 appeared most often and were isolated in 12 different months and across multiple sampling years. It would appear that these clinical *C. jejuni* genotypes are at least persistent, if not endemic to the southern Alberta region.

Key	Source	Month	CGF-20 Cluster	*Fla* A-RFLP Type
CJ250	Human	29	255	9
CJ249	Human	29	255	9
CJ279	Cattle	40	202	2
CJ278	Cattle	28	202	136
CJ277	Water	20	202	2
CJ272	Human	27.5	202	82
CJ271	Human	29	202	2
CJ267	Human	28	202	2
CJ266	Human	28	202	2
CJ265	Human	27	202	2
CJ264	Human	27	202	2
CJ442	Chicken	39	190	37
CJ436	Cattle	17	190	37
CJ433	Human	27.5	190	80
CJ427	Human	29	190	56
CJ426	Human	29	190	56
CJ422	Human	29	190	26
CJ421	Human	29	190	26
CJ420	Human	29	190	37
CJ419	Human	29	190	37
CJ409	Chicken	21	172	37
CJ402	Human	30	172	37
CJ401	Human	30	172	32
CJ396	Human	29	172	37
CJ394	Human	29	172	32
CJ393	Human	29	172	37
CJ391	Human	28	172	37
CJ390	Human	28	172	37
CJ389	Human	28	172	23
CJ357	Dog	18	169	26
CJ355	Human	27.5	169	26
CJ145	Water	29	10	49
CJ142	Chicken	18	10	102
CJ128	Human	29	10	48
CJ127	Human	29	10	23
CJ126	Human	27	10	54

CGF-20

CGF-119

Figure 25: Detailed genomic analysis of suspected endemic clones in southern Alberta. A subset of clusters containing the six suspected endemic CGF-20 fingerprints were evaluated using the CGF-119 assay. Isolates with these six genotypes comprise ~31% of all clinical isolates in the dataset. While *fla*A-RFLP type usually correlates with CGF-20 cluster, several CGF-20 cluster that have highly similar CGF-119 profiles have multiple *fla*A-RFLP types. Black is used to indicate genes that are conserved or divergent, and white to indicate genes that are absent.

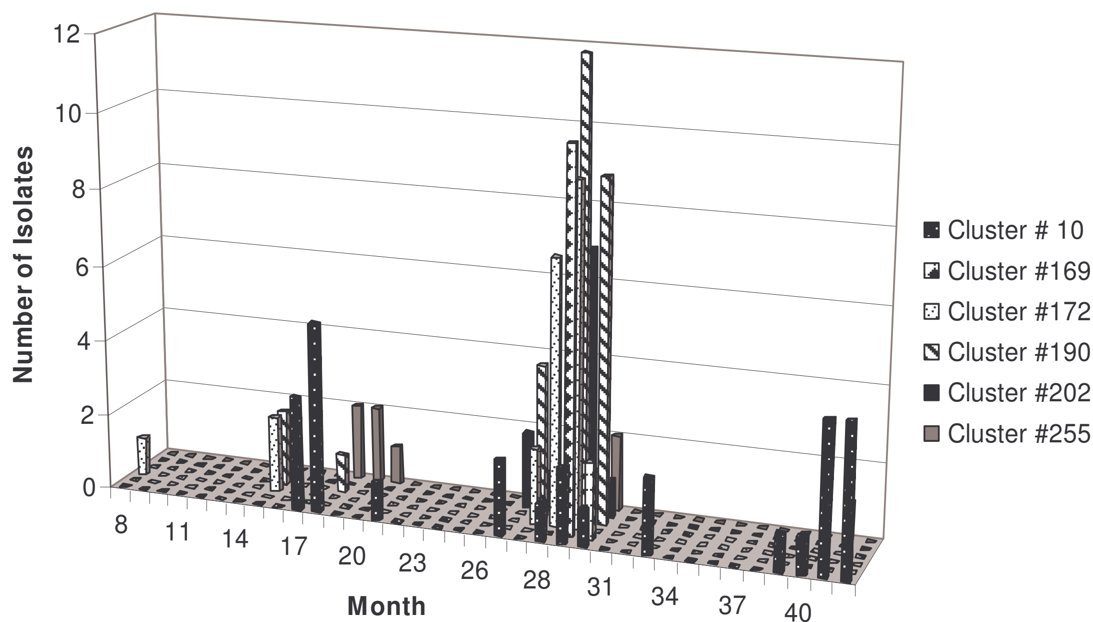


Figure 26: Temporal distribution of the six CGF-20 clusters containing a high proportion of clinical isolates in the dataset. Several CGF-20 clusters (#10, 169, 172, 190, 202 and 155) containing clinical isolates appeared at multiple times over the sampling period from September 2003 to August 2006. These profiles were not restricted to one particular source and were isolated from animal, environmental and clinical isolates (source of isolation not shown) suggesting that these profiles may be endemic to southern Alberta.

4.3 Discussion

Comparison of CGF-20 and *fla*A-RFLP clustering results shows that although the CGF-20 method tends to generate more clusters with a smaller average cluster size, that both methods ultimately have a similar Simpson's Index of Diversity. Therefore, it is important to examine the performance of the methods in an epidemiological context.

Analysis of the overall genetic similarity between strains within *fla*A-RFLP and CGF-20 clusters (*i.e.*, intra-cluster similarity) using CGF-119 data showed that strains with the same CGF-20 fingerprint tend to have a higher level of genetic similarity (~87%) compared to strains with the same *fla*A-RFLP fingerprint (~77%).

Interestingly, when examining genetic similarity among strains outside *fla*A-RFLP and CGF-20 clusters (*i.e.*, inter-cluster similarity) both methods had a similar level of inter-cluster similarity (~66%). This level of similarity appears to represent the genetic relatedness expected for any two random *C. jejuni* strains when these strains are assessed using the CGF-119 assay. It is important to note, that the loci included in the CGF-119 assay were selected from the ~20% of the genes that represent the most variable portions of the *C. jejuni* genome. As the assay is designed to uncover genetic variability and to provide enhanced typing resolution, it presents an over-estimation of genetic differences among strains because it does not include any of the *C. jejuni* genes that are conserved in all of the strains. Thus, the intra-cluster genetic similarity among strains within a CGF-20 cluster, and to a lesser extent within a *fla*A-RFLP cluster, is underestimated by the CGF-119 data. The assay, however, yields a comparable level of sub-typing resolution to whole-genome comparative genomic analysis (E. Taboada, personal communication) and

therefore represents a benchmark against which the CGF-20 and *fla*A-RFLP assays can be measured.

While the rates of intra-cluster and inter-cluster CGF-119 similarity vary between different CGF-20 clusters, in 12 of 13 clusters examined the intra-cluster similarities were significantly higher than inter-cluster similarities (~ 90% vs. 60%). When the distributions of pair-wise intra-cluster similarities were compared, while a similar peak was observed at 90% for both methods, the *fla*A-RFLP intra-cluster had a prominent secondary peak at 70%. This suggests that while *fla*A-RFLP and CGF-20 clusters can be highly predictive of overall genetic similarity, in many instances *fla*A-RFLP can cluster isolates that share only marginally higher similarity (~ 10%) than the background 'noise' of genetic relatedness observed between isolates from different clusters. Interestingly, although the pair-wise inter-cluster CGF-119 similarity distributions for both methods had a nearly identical shape with a prominent peak at ~ 60%, the *fla*A-RFLP distribution differed from that of the CGF-20 by the presence of a small right-handed peak, which suggests that in a small number of cases, isolates can share significant overall genetic similarity despite having different *fla*A-RFLP alleles. Taken together, both of these observations (isolates of the same *fla*A-RFLP type with low CGF-119 similarity; isolates of different *fla*A-RFLP type with high CGF-119 similarity) are consistent with potential lateral transfer or allelic replacement of *fla*A alleles. The *fla*A gene is known to be a 'hotspot' for recombination in the *C. jejuni* genome and is known to be subjected to high rates of lateral transfer (Mellmann *et al.*, 2004). One effect of these high rates of recombination is a reduction in the accuracy of inferred strain to strain relationships due to the 'uncoupling' of the *fla*A locus from the rest of the genome: strains that are

genetically related can have different *fla*A alleles and strains that are genetically unrelated can have the same *fla*A allele. Consistent with this evidence, several prominent *fla*A-RFLP types which appear to be segregated into multiple CGF-20 clusters and prominent CGF-20 clusters comprised of strains with diverse *fla*A-RFLP types were observed. The most parsimonious explanation for the *fla*A type composition in these clusters would be for *allelic replacement* of the *fla*A locus by recombination in strains that originally had the same *fla*A allele, and highlights problems with only using one or a small number of loci in the genome for predicting overall genetic content.

Both the CGF-20 method and the *fla*A-RFLP typing method were able to uncover possible genotypic connections between isolates that were missed by the alternate method. One weakness of this dataset is that there is a lack of sufficiently detailed epidemiologic data (due in part to privacy restrictions of potentially personal information about the clinical isolates) to fully assess the likelihood of these connections in a quantitative manner. The CGF-119 data, however, can provide an appropriate context with which to examine the likelihood of clustering results obtained with either method because of its enhanced sub-typing resolution. If the typing method generates clusters of genetically related isolates, the *intra-cluster* CGF-119 similarities are expected to be high. Whereas the CGF-119 similarity data shows that ~22% (31 of 142) of *fla*A-RFLP *intra-cluster* pair-wise comparisons have low similarity levels (*i.e.,* ≤ 65%) that are no better than those obtained from isolates outside the cluster, for CGF-20 the proportion is significantly lower (~4%, or 9 of 229). Conversely, if the typing method is able to generate clusters that are genetically distinct from one another, the *inter-cluster* CGF-119 similarities are expected to be low. Whereas the proportion of *fla*A-RFLP *inter-cluster*

pair-wise comparisons that have the high similarity levels (*i.e.,* $\geq$ 90%), *i.e.,* that might be expected for isolates within a cluster is ~2.4% (62 of 2559), for CGF-20 it is considerably lower (~0.5%, or 12 of 2472). The analysis of CGF-119 data for *fla*A-RFLP and CGF-20 clusters would suggest that more meaningful fingerprints were produced using the latter method, although a thorough analysis of "epidemiological fit" should be the focus of future studies.

Although the *fla*A-RFLP and the CGF-20 typing methods generated different numbers of clusters, both methods produced approximately eight large clusters containing isolates obtained from each of the various sources sampled. A comparison of cluster composition, however, revealed that the eight *fla*A-RFLP clusters contain 216 isolates compared to 87 isolates in the eight CGF-20 clusters, with only 50 isolates common to both sets of clusters. Although, based on the available epidemiological data, it is impossible to determine the likelihood that the isolates not identified by both typing methods truly represent clusters of genetically related organisms found in the various sources sampled. Our data suggests that the genotypes represented by the 50 overlapping isolates are likely to be common in this geographical region, can circulate between the various reservoirs, and are of public health significance since they are able to cause illness in humans.

Chapter 5

5. Review of Thesis

Campylobacter jejuni has a highly variable "accessory" pool of genes consisting of 30-40% of the genome. As a result, genome content varies among strains of *C. jejuni*, with a select number of genetic regions in the *C. jejuni* genome being highly plastic (Pearson *et al.*, 2003; Taboada *et al.*, 2004). Whole genome comparative genomic methods (*i.e.,* sequencing, comparative genomic hybridization) therefore represent the most appropriate approach to fully address the genetic relatedness of two strains. However, comparative genomic analyses remain slow, costly and labor intensive and have yet to find a use in routine bacterial typing and molecular epidemiology.

In response to these challenges, our group has developed an advanced genotyping method, comparative genomic fingerprinting or CGF, based on the theoretical foundation of comparative genomics but utilizing conventional molecular biology methods available to most laboratories. CGF is based on assessing the gene conservation profile of genes known to be variably absent or present (VAP) in the species of interest and utilizing this information as a genetic fingerprint for genotyping purposes. The resulting fingerprints can be compared and used to assess the degree of genetic relatedness of strains based on the percentage of conserved loci shared. An advantage of comparative genomic fingerprinting compared to other methods and, in particular, to band-based fingerprinting methods, is that analysis at reduced stringencies may retain biologically meaningful 'inferred genetic relatedness'. Although CGF fingerprints can only be used to estimate the genetic content of a strain, they appear to provide a relevant assessment of overall

genetic relatedness. Comparative genomic fingerprinting at 20 VAP loci (CGF-20) can be performed at a fraction of the time and cost of other molecular typing methods, including MLST and PFGE. As demonstrated in Chapter 3, CGF-20 fingerprints are non-ambiguous, robust and easily reproducible, which make them amenable to inter-laboratory comparison.

Possible applications of the CGF-20 assay could be to provide a basis for rapid and low-cost identification of prominent lineages that could then be verified with the slower but higher resolution CGF-35 or CGF-119 assays. Although it may be intuitive to assume that the highest levels of resolution are always preferable, because of the significant level of correlation between the CGF-20 and CGF-119 data, clustering obtained with the CGF-20 assay could be used to rapidly identify lineages that may represent groups that have a shared epidemiology, that share certain biological characteristics, or that may have become adapted to a particular environmental niche. Conversely, the increased resolution provided through the addition of supplementary loci in the CGF-35 and, to a greater extent, CGF-119 assays could refine the clusters to provide an enhanced ability to discriminate between strains for which differences in underlying geographical, temporal or source parameters may be associated with the acquisition of additional genetic differences, thus enabling confirmation of genetic/epidemiological links with a higher confidence.

In this thesis, a dataset comprised of 641 *C. jejuni* isolates from human clinical, animal, and environmental sources was analyzed using the CGF-20 assay in order to address basic questions about the molecular epidemiology of *C. jejuni* in southern Alberta. The data indicates that in addition to the traditionally recognized reservoirs of

infectious *C. jejuni* (*i.e.* commercial poultry), several common farm animals and wild animals native to southern Alberta can also harbor humanpathogenic *C. jejuni.* The reoccurrence of highly related CGF-20 fingerprints among species over time may indicate that not only are campylobacters being shed into the environment; but that they persist while acquiring subtle genetic differences over time. That several species of animals harbor and shed campylobacters that share identical CGF-20 profiles to human clinical isolates represents a public health concern. These matching profiles do not appear to be due to random chance, as validation using CGF-119 data identified several groups of strains from animal and human sources that were nearly identical as assessed by the CGF-119 assay. Since this assay was designed to "amplify" genetic differences between strains by targeting the most variable loci in the *C. jejuni* genome, it therefore appears that these animal species may shed *C. jejuni*, some of which appear to have pathogenic capabilities. Interestingly, four animal species often shared CGF profiles with clinical isolates: chicken, cattle, sheep and pigs. As many of the bacterial isolates were of cattle and sheep origin, two animals not commonly investigated for *Campylobacter* spp. carriage, they must be considered as potential links in the chain leading to human infection. Current campylobacteriosis rates in southern Alberta are double the rates found in other regions of Canada. New Zealand also has been facing rates of campylobacteriosis with epidemic proportions (Baker *et al.*, 2006). It is interesting to note that both southern Alberta and New Zealand are known for both sheep and cattle production. Our data also suggest that bison also require more investigation as a potential reservoir of pathogenic campylobacters in southern Alberta and elsewhere. Although bison ranches are rare, several are located in southern Alberta. Out of seven bison fecal samples obtained in this

study, four were positive for *C. jejuni*, and each of these matched with at least one clinical isolate in our dataset using both *fla*A-RFLP typing and CGF typing. Although bison meat has been tested for other microbes, association of campylobacters with this meat has not been established (Gill, 2007).

Several instances where animal, water and clinical isolates shared profiles were found which suggests that a chain of transmission between all three reservoirs is possible. This supports the concept that agriculture, including livestock production and spreading of manure on crops leads to an increase in fecal coliforms in surface waters (Johnson *et al.*, 2003), making it likely that campylobacters from animal manure enter the southern Alberta waterways. The most common manure spread in southern Alberta is from cattle, seconded by swine, and use of composted manure as a carbon credit is likely to increase in the future with pending carbon emission legislation (Coulter, 2008; Papworth & Olson, 2002). In addition, cattle are estimated to produce 4.9 to 5.4 kg of manure per head each day (Ormann, 2005) and Alberta has both a beef cattle population of over 5.5 million and possesses an extensive feedlot production system with 2.36 million head of finished cattle processed in 2002 (*http://www.growingalberta.com/features/default.asp?id=93*). This suggests that at least 26.9 to 29.7 million kg of manure are produced from beef cattle in Alberta daily. These and other animal sources of fecal matter are likely important sources of *Campylobacter* spp., and other bacterial pathogens, into the natural waters of southern Alberta. Another route of transmission that will require investigation in southern Alberta and that has been recently suggested based on ongoing work on campylobacters in Iceland, could be from insects such as flies acting as vectors to transmit campylobacters

from manure directly to food or into the food production environments at various points along the farm to fork continuum (Guerin *et al.*, 2007).

An important finding from this study is the detection of a small number of genotypic lineages comprising a significant proportion of clinical isolates in our dataset. This suggests that some strains of *C. jejuni* with increased virulence may be endemic to the region, and may be maintained in asymptomatic animal hosts as isolates with identical profiles were obtained from human clinical and animal fecal samples. For example, among the 4 isolates obtained from buffalo in the study, the first two isolates obtained were identical to clinical isolates obtained several months later, whereas the later two samples obtained near the end of the study were identical to clinical samples that were obtained early, but persisted over several months; these isolates also were observed months later in chickens. Although it appears that some of these genotypes and relevant clusters are present in the environment and are of clinical concern during certain months, this is likely to be an artifact due to seasonal sampling bias as water samples were not collected when water surfaces were frozen and fecal sample collection declined during poor weather conditions. The temporal distribution of some of these genotypes is such that they appear in isolates obtained over several different sampling years and consequently, may be endemic to southern Alberta. The high prevalence of six genotypes among the clinical isolates in our dataset (~31%) suggests that they represent genotypes with an increased risk to human health that should be investigated further in order to identify possible genetic commonalities that could be targeted for diagnostic purposes.

Due to the plasticity seen in the genome of *C. jejuni*, genetically related strains identified on the basis of estimating whole genome genetic content should define more

complete and useable genetic relationships for multiple applications rather than methods based on one or a limited number of genes. The CGF-20 and CGF-35 assays were validated with the CGF-119 assay, which additional ongoing work in our laboratory has shown it to be a robust predictor of overall genome similarity between strains (E. Taboada, personal communication). CGF methods allow for rapid visualization of genes present throughout the entire genome allowing for patterns of gene transfer, loss and acquisition to be observed and analyzed in the context of underlying epidemiology and phenotype. The methodology also enables a comparison of the genetic relatedness of strains to be assessed with an unprecedented level of resolution, even for strains that have similar genotypes. This is in contrast to band-based methods such as *fla*A-RFLP or PFGE, which can be used to infer genetic similarity in the case of identical or highly related banding patterns, but in which increasing levels of fingerprint divergence can quickly translate into an inability to assess genetic relatedness with any level of accuracy. Although multi-locus sequence typing (MLST) has been shown to predict overall genetic similarity among strains (Taboada *et al.*, 2008), it would not be able to predict the pattern of conservation within the hyper-variable regions of the genome, which could result in a change to the phenotype between strains of the same MLST type. Natural competence and high rates of lateral gene transfer in *C. jejuni* pose an additional challenge in inferring genetic relatedness of strains since these events occur readily and genetic divergence can therefore arise rapidly. By evaluating 20 different loci interspersed across the genome and targeting most of the hyper-variable loci in the *C. jejuni* genome, the CGF-20 assay is better able to compensate for these effects by providing sufficient data at many loci,

hopefully identifying changes that might not have been observed if only one locus is studied.

A number of issues limited our ability to identify and validate epidemiological linkages among matching isolates obtained from clinical, animal and environmental sources. For instance, although the majority of water and animal samples were obtained from within the Oldman River Basin in the Chinook Health Region, during this study, I was unable to obtain clinical samples from the same corresponding region. The clinical samples, which were from the Calgary Health Region, lacked exact diagnoses and isolation dates due to privacy restrictions. For animal and water sources, sampling was not done year round, since water collection did not occur during the winter months, and animal fecal collection occurred sporadically throughout the Oldman River watershed. Despite these concerns, the strong connections observed based on the genotyping data and the available epidemiological data suggest that many *C. jejuni* isolates from animals and contaminated water in this region represent a public health risk.

5.1 Thesis Conclusions

A method of comparative genomic fingerprinting (CGF) for genotyping *Campylobacter jejuni* based on the presence or absence of 20 highly variable genes (CGF-20) located throughout the *C. jejuni* genome has been presented in the analysis of 641 isolates of human, animal, and environmental origin obtained in southern Alberta. The CGF-20 data was validated two-fold: all isolates were also genotyped using *fla*A-RFLP, and a subset of these isolates was also analyzed using higher genetic resolution CGF-35 and CGF-119 assays, which are based on 35 and 119 highly variable genes in the

C. jejuni genome, respectively. Based on the work presented in this thesis, the following conclusions can be reached:

1. The CGF-20 assay was shown to be a highly robust assay based on validation using the CGF-35 assay using a subset of isolates from the dataset, since ~98% of all loci queried yielded matching data when the assay was repeated.

2. The QIAxcel-based assay presents a high-throughput CGF assay capable of high-resolution genotyping (*i.e.,* using the CGF-35 assay) of 96 samples per day at a cost of ~ $10 CAD per sample; the lower resolution CGF-20 assay can achieve twice the throughput of the CGF-35 assay at half the cost. This is in contrast with MLST, the current gold standard for *C. jejuni* genotyping, which anecdotal evidence from research collaborators shows to have a throughput of 96 samples in 4 weeks or more at a cost of ~$50-75 per sample.

3. The CGF-20 assay was found to be *functionally equivalent* to the CGF-119 assay since isolates with a given CGF-20 cluster shared high levels of similarity in CGF-119 fingerprints. The CGF-119 assay is highly predictive of whole-genome genetic similarity among strains as it incorporates more than half of known variably absent/present genes while targeting all the highly variable regions in the *C. jejuni* genome. The CGF-20 assay thus represents a rapid and inexpensive method for genotyping *C. jejuni* that can be used to rapidly establish groups or clusters of strains with high levels of whole-genome genetic similarity. In

addition, for the majority of purposes, the resolution of 35 and 119 gene fingerprints would be too high when analyzed at the level of 100% identity, and may need to be evaluated using a lower level of stringency to increase the possibility of finding matches between closely related although not identical strains. This is particularly important for an organism such as *C. jejuni*, which is known to undergo rapid genetic change.

4. The CGF-20 assay, when applied to a molecular epidemiological study of *C. jejuni* isolates from southern Alberta was able to generate epidemiologically relevant clusters. CGF-20 results were also compared to *fla*A-RFLP data. Both methods were shown to yield similar results by generating clusters comprised of isolates with high genetic similarity while excluding unrelated isolates. The CGF-20 assay, however, produced smaller clusters with a higher within-cluster, or *intra-cluster*, similarity when compared to *fla*A-RFLP.

5. The data presents strong evidence for two types of lateral transfer of the *fla*A allele. The first type, *fla*A *allelic replacement*, was observed when strains deemed to be highly similar based on CGF-119 profiles had unrelated *fla*A-RFLP profiles. The second type, *fla*A *allelic transfer,* occurred in cases where two strains appeared to be vastly different based on CGF-119 profiles but nevertheless shared the same *fla*A alleles. Since the CGF-119 fingerprints are based on data on a large number of loci spread across the genome (in contrast to the *flaA*-RFLP data, which is based on a single locus known to be undergoing constant

recombination), the most parsimonious explanation would suggest that in these cases the *fla*A allele was transferred between unrelated strains or replaced between related strains. CGF provides a better assessment of whole-genomic similarity as numerous genes from hypervariable areas of the genome are evaluated.

6. Our data indicates that in addition to the currently recognized reservoir of poultry, other livestock such as cattle, sheep, and pigs can harbour *C. jejuni* strains with genotypes that match those from human clinical cases. As there were found several instances where animal, water and clinical isolates shared genotypic profiles, this would suggest that transmission can occur between all three reservoirs. Determining the mechanism of transmission is an extremely important subject for subsequent investigations so that it is possible to determine how to best implement control measures (*i.e.,* changes to the manufacturing, handling or consumption of meat, treatment of water, or treatment of manure) so as to reduce campylobacteriosis in southern Alberta.

7. At least six 'endemic' clinically-associated CGF-20 genotypes were identified. The high prevalence of these six genotypes among the clinical isolates in our dataset (~31%) would suggest that they represent genotypes with an increased risk to human health and should be investigated further. Whole genomic analysis of strains with these genotypes could be useful in identifying genes that may important in virulence and are worthwhile targets for additional investigation.

8. This study establishes the *C. jejuni* CGF-20 as a high-throughput molecular typing method is able to predict genetic similarity, yield epidemiologically relevant clusters, and be suitable for phylogenetic and molecular epidemiological applications. The methodology is robust and can enable rapid and inexpensive molecular epidemiological analysis, and straightforward collaboration between members of the national or international public health and research laboratory communities.

9. A combined approach using CGF assays of varying resolution (*i.e.,* using a different number of variably present/absent genes) coupled with detailed epidemiological information could prove useful in future genomics-based molecular epidemiological studies of *C. jejuni*. Although for the majority of purposes the 20 gene fingerprint has enough resolution to allow for rapid determination of important genotypic clusters, the increased resolution of the 35 and 119 gene fingerprints may be useful for further refining genotypic clusters in the context of epidemiological investigations. This is particularly important for an organism such as *C. jejuni*, which is known to undergo rapid genetic change.

References

Açik, M. N., & Çetinkaya, B. (2006). Heterogeneity of *Campylobacter jejuni* and *Campylobacter coli* strains from healthy sheep. *Veterinary Microbiology, 115*, 370-375.

Alberta Health and Wellness. (2005). *Notifiable disease in Alberta: 2004 Annual Report*. Edmonton, AB: Alberta Health and Wellness.

Allos, B. M. (2001). *Campylobacter jejuni* infections: update on emerging issues and trends. *Clinical Infectious Diseases, 32*, 1201-1206.

Altekruse, S. F., Stern, N. J., Fields, P. I., & Swerdlow, D. L. (1999). *Campylobacter jejuni* - an emerging foodborne pathogen. *Emerging infectious diseases, 5*, 28-35.

Alter, T., Guall, F., Kasimir, S., Gurtler, M., Miekle, H., Linnebur, M., *et al.* (2005). Prevalences and transmission routes of *Campylobacter* spp strains within multiple pig farms. *Veterinary Microbiology, 108*, 251-261.

Alexandrina, A. and D. Botus. (2008). The contamination of pork meat with *Campylobacter* germs during the technological flow. *Veterinary Medicine, 65*(1), 234-237.

Andrés, S., Jiménez, A., Sánchez, J., Alonso, J. M., Gómez, L., López, F., et al. (2007). Evaluation of some etiological factors prediposing to diarrhea in lambs in "La Serena" (Southwest Spain). *Small Ruminant Research, 70*, 272-275.

Arsenault, J., Letellier, A., Quessy, S., Normand, V., & Boulianne, M. (2007). Prevalence and risk factors for *Salmonella* spp. and *Campylobacter* spp. caecal colonization in broiler chicken and turkey flocks in Quebec, Canada. *Preventive Veterinary Medicine, 81*, 250-264.

Auld, H., MacIver, D., & Klaassen, J. (2004). Heavy rainfall and waterborne disease outbreaks: The Walkerton example. *Journal of Toxicology and Environmental Health, Part A, 67*(20), 1879-1887.

Avrain, L., Humbert, F., L'Hospitalier, R., Sanders, P., Vernozy-Rozand, C., & Kempt, I. (2003). Antimicrobial resistance in *Campylobacter* from broilers: associate with production type and antimicrobial use. *Veterinary Microbiology, 96*(3), 267-276.

Aydin, F., Atabay, H. I., & Akan, M. (2001). The isolation and characterization of *Campylobacter jejuni* subsp. *jejuni* from domestic geese (*Anser anser). Journal of Applied Microbiology, 90*, 637-642.

Bacon, D. J., Alm, R. A., Burr, D. H., Hu, L., Kopecko, D. J., Ewing, C. P., et al. (2000). Involvement of a plasmid in virulence of *Campylobacter jejuni* 81-176. *Infection and Immunity, 68*(8), 4384-4390.

Baker, M., Wilson, N., Ikram, R., Chambers, S., Shoemack, P., & Cook, G. (2006). Regulation of chicken contamination urgently needed to control New Zealand's serious campylobacteriosis epidemic. *Journal of the New Zealand Medical Association, 119*(1243), U2264.

Barrios, P. R., Reiersen, J., Lowman, R., Bisaillon, J. R., Michel, P., Fridriksdottir, V., et al. (2006). Risk factors for *Campylobacter* spp. colonization in broiler flocks in Iceland. *Preventive Veterinary Medicine, 74*, 264-278.

Barton, C., Ng, L. K., Tyler, S. D., & Clark, C. G. (2007). Temperate bacteriophages affect pulsed-field gel electrophoresis patterns of *Campylobacter jejuni*. *Journal of Clinical Microbiology, 45*(2), 386-391.

Bigwood, T., Hudson, J. A., Billington, C., Carey-Smith, G. V., & Heinemann, J. A. (2008). Phage inactivation of foodborne pathogens on cooked and raw meat. *Food Microbiology, 25*(2), 400-406.

Blaser, M. J. & Reller, L. B. (1981). *Campylobacter enteritis*. *New England Journal of Medicine. 305*(24). 1444-1452.

Blaser, M. J., & Engberg, J. (2008). Clinical aspects of *Campylobacter jejuni* and *Campylobacter coli* infections. In I. Nachamkin, C. M. Szymanski & M. J. Blaser (Eds.), *Campylobacter, 3rd edition* (pp. 99-121). Washington, DC: ASM Press.

Boer, P., Duim, B., Rigter, A., van der Plas, J., Jacobs-Reitsma, W. F., & Wagenaar, J. A. (2000). Computer-assisted analysis and epidemiological value of genotyping methods for *Campylobacter jejuni* and *Campylobacter coli. Journal of Clinical Microbiology, 38*(5), 1940-1946.

Boerlin, P. (1999). Evolution of virulence factors in Shiga-toxin-producing *Escherichia coli*. *Cellular and Molecular Life Sciences (CMLS), 56*(9), 735-741.

Bolton, F. J., Hutchison, D. N., & Parker, G. (1987). Isolation of *Campylobacter;* what are we missing? *Journal of Clinical Pathology, 40*(6), 702-703.

Callicott, K. A., Frioriksdottir, V., Rejersen, J., Lowman, R., Bisaillon, J. R., Gunnarsson, E., et al. (2006). Lack of evidence for vertical transmission of *Campylobacter* spp. in chicken. *Applied and Environmental Microbiology, 72*(8), 5794-5798.

Carrillo, C. D., Taboada, E., Nash, J. H. E., Lanthier, P., Kelly, J., Lau, P. C., et al. (2004). Genome-wide expression analyses of *Campylobacter jejuni* NCTC1168 reveals coordinate regulation of motility and virulence by *flh*A. *The Journal of Biological Chemistry, 279*(19), 20327-20338.

Centers for Disease Control and Prevention. (1991). *Campylobacter enteritis - New Zealand, 1990*: U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition.

Chai, L. C., Robin, T., Ragavan, U. M., Gunsalam, J. W., Baker, F. A., Ghazali, F. M., et al. (2007). Thermophilic *Campylobacter* spp. in salad vegetable in Malaysia. *International Journal of Food Microbiology, 117*, 106-111.

Clark, C. G., Beeston, A., Brygen, L., Wang, G., Barton, C., Cuff, W. R., et al. (2007). Phylogenetic relationships of *Campylobacter jejuni* based on *por*A sequences. *Canadian Journal of Microbiology, 53*, 27-38.

Clark, C. G., Brygen, L., Cuff, W. R., Johnson, P. L., Jamieson, F., Ciebin, B., et al. (2005). Use of the oxford multilocus sequence typing protocol and sequencing of the flagellin short variable region to characterize isolates from a large outbreak of waterborne *Campylobacter* sp. strains in Walkerton, Ontario, Canada. *Journal of Clinical Microbiology, 43*(5), 2080-2091.

Clark, C. G., Price, L., Ahmed, R., Woodward, D., Melito, P. L., Rodgers, F. G., et al. (2003). Characterization of a waterborne outbreak associated *Campylobacter jejuni* in Walkerton, Ontario. *Emerging infectious diseases, 9*(10), 1232-1241.

Coulter, B. (2008, August 01, 2008). The emissions trader. *Alberta Venture*.

Cox, N. A., Stern, N. J., Hiett, K. L., & Berrang, M. E. (2002). Identification of a new source of *Campylobacter* contamination in poultry: transmission from breeder hens to broiler chickens. *Avian Disease, 46*(3), 535-541.

Dassanayake, R. P., Zhou, Y., Hinkley, S., Stryker, C. J., Plauche, G., Borda, J. T., et al. (2005). Characterization of cytolethal distending toxin of *Campylobacter* species isolated from captive Macaque monkeys. *Journal of Clinical Microbiology, 43*(2), 641-649.

Denes, A. S., Lutze-Wallace, C. L., Cormier, C., & Manzano, M. (1995). Detection and identification of *Campylobacter fetus* using cloned constructs of ribosomal RNA and surface array proteins. *Veterinary Microbiology, 54*, 185-193.

Diker, K. S., Esendal, O. M., & Akan, M. (2000). Epidemiology of ovine *Campylobacter* infection determined by numberical analysis of electrophoretic protein profiles. *Journal of Veterinary Medicine, 47*, 739-743.

Dingle, K. E., Van Den Braak, N., Colles, F. M., Price, L. J., Woodward, D. L., Rodgers, F. G., et al. (2001). Sequence typing confirms the *Campylobacter jejuni* strains associated with Gullian-Barre and Miller-Fisher syndromes are of diverse genetic lineage, serotype, and flagella type. *Journal of Clinical Microbiology, 39*(9), 3346-2249.

Dobrindt, U., & Hacker, J. (2001). Whole genome plasticity in pathogenic bacteria. *Current Opinion in Microbiology, 4*, 550-557.

Dorrell, N., Mangan, J. A., Laing, K. G., Hinds, J., Linton, D., Al-Ghusein, H., *et al.* (2001). Whole genome comparision of *Campylobacter jejuni* using low-cost microarray reveals extensive genetic diversity. *Genome Research, 11*, 1706-1715.

Duim, B., Wassenaar, T. M., Rigter, A., & Wagenaar, J. A. (1999). High-resolution genotyping of *Campylobacteri* strains isolated from poultry and humans with amplified fragment length polymorphism fingerprinting. *Applied and Environmental Microbiology, 65*(6), 2369-2375.

Duim, B., P. A. R. Vandamme, A. Rigter, S. Laevens, J. R. Dijkstra, and J. A Wagenaar. (2001). Differentiation of *Campylobacter* species by AFLP fingerprinting. *Microbiology, 147,* 2729-2737.

Eriksson, J., Lofstrom, C., Aspan, A., Gunnarsson, A., Karlsson, I., Borch, E., *et al.* (2005). Comparison of genotyping methods by application to *Salmonella* livingstone strains associated with an outbreak of human salmonellosis. *International Journal of Food Microbiology, 104*, 93-103.

Ertas, H. B., Cetinkaya, B., Muz, A., & Ongor, H. (2004). Genotyping of broiler-originated *Campylobacter jejuni* and *Campylobacter coli* isolates using fla typing and random amplified polymorphic DNA methods. *International Journal of Food Microbiology, 94*(2), 203-209.

Evans, M. R., Ribeiro, C. D., & Salmon, R. L. (2003). Hazards of healthy living: Bottled water and salad vegetables as risk factors for *Campylobacter* infection. *Emerging Infectious Diseases, 9*(10), 1219-1225.

Evans, S. (1997). Introduction and the spread of thermophilic campylobacters in broiler flocks. *Veterinary Record, 131*, 574-576.

Farber, J. M. (1996). An introduction to the hows and whys of molecular typing. *Journal of Food Protection, 59*(10), 1091-1101.

Fayos, A., Owen, R. J., Desai, M., & Hernandez, J. (1992). Ribosomal RNA gene restriction fragment diversity amongst Lior biotypes and Penner serotypes of *Campylobacter jejuni* and *Campylobacter coli*. *FEMS Microbiology Letters, 74*(1), 87-93.

Finney, M. (1993). Pulsed-field electrophoresis. In F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith & K. Struhl (Eds.), *Current Protocols in Molecular Biology* (Vol. 1, pp. 2.5.9-2.5.17). New York, N.Y.: Greene-Wiley.

Fitzgerald, C., Helsel, L. O., Nicholson, M. A., Olson, S. J., Swerdlow, D. L., Flahart, R., et al. (2001). Evaluation of methods for subtyping *Campylobacter jejuni* during an outbreak involving a food handler. *Journal of Clinical Microbiology, 39*(7), 2386-2390.

Fitzgerald, C., Owen, R. J., & Stanley, J. (1996). Comprehensive ribotyping scheme for heat-stable serotypes of *Campylobacter jejuni*. *Journal of Clinical Microbiology, 34*(2), 265-269.

Fouts, D. E., Monogodin, E. F., Mandrell, R. E., Miller, W. G., Rasko, D. A., Ravel, J., et al. (2005). Major structural differences and novel potential virulence mechanism from the genomes of multiple *Campylobacter* species. *Public Library of Science Biology, 3*(1), 72-85.

Fujimoto, S., Allos, B. M., Misawa, N., Patton, C. M., & Blaser, M. J. (1997). Restriction fragment length polymorphism analysis and random amplified polymorphic DNA analysis of *Campylobacter jejuni* strains isolated from patients with Guillain-Barre syndrome. *Journal of Infectious Diseases, 176*(4), 1105-1108.

Gadewar, S., & Fasano, A. (2005). Current concepts in the evaluation, diagnosis and management of acute infectious diarrhea. *Current Opinion in Pharmacology, 5*(6), 559-565.

Gannon, V. P. J., Duke, G. D., Thomas, J. E., VanLeeuwen, J., Byrne, J., Johnson, D., et al. (2005). Use of in-stream reservoirs to reduce bacterial contamination of rural watersheds. *Science of the Total Environment, 348*, 19-31.

Gannon, V. P. J., Graham, T. A., Read, S., Ziebell, K., Muckle, A., Mori, J., et al. (2004). Bacterial pathogens in rural water supplies in southern Alberta, Canada. *Journal of Toxicology and Environmental Health, Part A, 67*, 1643:1653.

Gibson, J. R., Fitzgerald, C., & Owen, R. J. (1995). Comparison of PFGE, ribotyping and phage-typing in the epidemiological analysis of *Campylobacter jejuni* serotype HS2 infections. *Epidemiology and Infection, 115*(2), 215-225.

Gill, C. O. (2007). Microbiological conditions of meats from large game animals and birds. *Meat Science, 2007*, 149-160.

Godschalk, P. C. R., Bergman, M. P., Gorkink, R. F. J., Simons, G., van der Braak, N., Lastovica, A. J., et al. (2006). Identification of DNA sequence variation in *Campylobacter jejuni* strains associated with the Gullain-Barie syndrome by high-throughput ALFP analysis. *BMC Microbiology, 6*, 32-43.

Goehler, L. E., Lyte, M., & Gaykema, R. P. A. (2007). Infection-induced viscerosensory signals from the gut enhance anxiety: Implications for pyschoneuroimmunology. *Brain, Behavior and Immunity, 21*(6), 721-726.

Grundmann, H. S., Hori, S., & Tanner, G. (2001). Determing confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. *J Clin Microbiol, 39*, 4190-4192.

Guerin, M. T., Martin, W., Reiersen, J., Berke, O., McEwan, S. A., Bisaillon, J.-R., et al. (2007). House-level risk factors associated with the colonization of broiler flocks with *Campylobacter* spp. in Iceland, 2001-2004. *BMC Veterinary Research, 3*(30), 264-278.

Hacker, J., & Carniel, E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity. *European Molecular Biology Organization Reports, 2*(5), 376-381.

Hansson, I., Ederoth, M., Andersson, L., Vagsholm, I., & Olsson Engvall, E. (2005). Transmission of *Campylobacter* spp. to chickens during transport to slaughter. *Journal of Applied Microbiology, 99*, 1149-1157.

Harrington, C. S., Moran, L., Ridley, A. M., Newell, D. G., & Madden, R. H. (2003). Inter-laboratory evaluation of three flagellin PCR/RFLP methods for typing *Campylobacter jejuni* and *C. coli*: The CAMPYNET experience. *Journal of Applied Microbiology, 95*, 1321-1333.

Hofreuter, D., Tsai, J., Watson, R. O., Novik, V., Altman, B., Benitez, M., et al. (2006). Unique Features of a Highly Pathogenic *Campylobacter jejuni* Strains. *Infection and Immunity, 74*(8), 4694-4707.

Hrudey, S. E., Payment, P., Huck, P. M., Gillham, R. W., & Hrudey, E. J. (2003). A fatal waterborne disease epidemic in Walkerton, Ontario: comparison with other

waterborne outbreaks in the developed world. *Water Science & Technology, 47*(3), 7-14.

Humphrey, T., & Jorgensen, F. (2006). Pathogens on meat and infection in animals - Establishing a relationship using *Campylobacter* and *Salmonella* as examples. *Meat Science, 74*, 89-97.

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

Hunter, P. R., & Gaston, M. A. (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *Journal of Clinical Microbiology, 26*, 2465-2466.

Hyland, R., Byrne, J., Selinger, B., Graham, T., Thomas, J., Townshend, I., et al. (2003). Spatial and temporal distribution of fecal indicator bacteria within the Oldman river basin of southern Alberta, Canada. *Water Quality Research Journal Canada, 38*(1), 15-32.

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

Jacobs, B. C., Rothbarth, P. H., van der Meche, F. G., Herbrink, P., Schmitz, P. I., de Klerk, M. A., et al. (1998). The spectrum of antecedent infections in Gullian-Barre syndrome. *Neurology, 51*, 1110-1115.

Janda, M. J., & Abbott, S. L. (2006). New gram-negative entropathogens: fact or fancy? *Reviews in Medical Microbiology, 17*, 27-37.

Jensen, A. N., Dalsgarrd, A., Baggesen, D. L., & Nielson, E. M. (2006). The occurrence and characterization of *Campylobacter jejuni* and *C. coli* in organic pigs and their outdoor environment. *Veterinary Microbiology, 116*, 96-105.

Johnson, J. Y. M., Thomas, J. E., Graham, T. A., Townshend, I., Byrne, J., Selinger, L. B., et al. (2003). Prevalence of *Escherichia coli* O157 and *Salmonella* spp. in surface waters of southern Alberta and its relation to manure sources. *Canadian Journal of Microbiology, 49*, 326-335.

Kapperud, G., Skjerve, E., Bean, N. H., Ostroff, S. M., & Lassen, J. (1992). Risk factors for sporadic *Campylobacter* infections: Results of a case-control study in southeastern Norway. *Journal of Clinical Microbiology, 30*(12), 3117-3121.

Karlyshev, A. V., Linton, D., Gregson, N. A., Lastovica, A. J., & Wren, B. W. (2000). Genetic and biochemical evidence of a *Campylobacter jejuni* capsular polysaccharide that accounts for Penner serotype specificity. *Molecular Microbiology, 35*(3), 529-541.

Keller, J., & Perreten, V. (2006). Genetic diversity in fluoroquinolone and macrolide-resistant *Campylobacter coli* from pigs. *Veterinary Microbiology, 113*, 103-108.

Kiehlbauch, J. A., Plikaytis, B. D., Swaminathan, B., Cameron, D. N., & Wachsmuth, I. K. (1991). Restriction fragment length polymorphisms in the ribosomal sequences for speices identification and subtyping of aerotolerant *Campylobacter* species. *Journal of Clinical Microbiology, 29*, 1670-1676.

Koene, M. G. J., Houwers, D. J., Dijikstra, J. R., Duim, B., & Wagenaar, J. A. (2009). Strain variation within *Campylobacter* species in fecal samples from dogs and cats. *Veterinary Microbiology*. *133*(1-2), 199-205.

Koga, M., Gilbert, H., Koike, S., Takahashi, M., Furukawa, K., Hirata, K., et al. (2005). Antecedent infections in Fisher syndrome: a common pathogenesis of molecular mimicry. *Neurology, 64*, 1605-1611.

Laing, C., Pegg, C., Yawney, D., Ziebell, K., Steele, M., Johnson, R., et al. (2008). Rapid Determination of *Escherichia coli* O157: H7 lineage types and molecular subtypes by using comparative genomic fingerprinting. *Applied and Environmental Microbiology, 74*(21), 6606-6615.

Lawrence, J. G. (1999). Gene transfer, speciation, and the evolution of bacterial genomes. *Current Opinion in Microbiology, 2*, 519-523.

Levesque, S., Frost, E., Arbeit, R. D., & Michaud, S. (2008). Multilocus sequence typing of *Campylobacter jejuni* isolates from humans, chickens, raw milk, and environmental water in Quebec, Canada. *Journal of Clinical Microbiology, 46*(10), 3404-3411.

Linton, D., Lawson, A. J., Owen, R. J., & Stanley, J. (1997). PCR detection, identification to species level, and fingerprinting of *Campylobacter jejuni* and *Campylobacter coli* direct from diarrheic samples. *Journal of Clinical Microbiology, 35*(10), 2568-2572.

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

Logue, C. M., Sherwood, J. S., Elijah, L. M., Olah, P. A., & Dockter, M. R. (2003). The incidence of *Campylobacter* spp. on processed turkey from processing plants in the midwestern United States. *Journal of Applied Microbiology, 95*, 234-241.

Maiden, M. C. J. (2006). Multi-locus sequence typing of Bacteria. *Annual Reviews in Microbiology, 60*, 561-588.

Manfreda, G., De Cesare, A., Bondioli, V., & Franchini, A. (2003). Ribotyping characterization of *Campylobacter* isolates randomly collected from different sources in Italy. *Diagnostic Microbiology and Infectious Disease, 47*(2), 385-392.

Mangen, M. J., Havelaar, A. H., Bernsen, R., Koningsveld, R. C., & Wit, G. A. D. (2005). The cost of human *Campylobacter* infections and sequelae in the Netherlands: A DALY and cost-of-illness approach. *Acta Agriculture Scandinavica, Section C - Economy, 2*(1), 35-17.

Marri, P. R., Hao, W., & Golding, G. B. (2007). The role of laterally transferred genes in adaptive evolution. *BMC Evolutionary Biology, 7*(Supplement I), S6.

Martinez-Rodriguez, A., & Mackey, B. M. (2005). Factors affecting the pressure resistance of some *Campylobacter* species. *Letters in Applied Microbiology, 31*, 321-326.

Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., et al. (2000). Food-related illness and death in the United States. *Journal of Environmental Health, 62*(7), 9-18.

Mead, P. S., Slutsker, L., Dietz, V., McCraig, L. F., Bresee, J. S., Shapiro, C., et al. (1999). Food-related deaths in the United States. *Emerging Infectious Diseases, 5*, 607-625.

Mellmann, A., Mosters, J., Bartelt, E., Roggentin, P., Ammon, A., Friedrich, A. W., et al. (2004). Sequence-based typing of *fla*B is more stable screening tool than typing of *fla*A for monitoring *Campylobacter* populations. *Journal of Clinical Microbiology, 42*(10), 4840-4842.

Mera, V., Lopez, T., & Serralta, J. (2007). Take traveller's diarrhea to heart. *Travel Medicine and Infectious Disease, 5*, 202-203.

Misawa, N., Shinohara, S., Satoh, H., Itoh, H., Shinohara, K., Shimomura, K., et al. (2000). Isolation of *Campylobacter* species from zoo animals and polymerase chain reaction-based random amplified polymorphism DNA analysis. *Veterinary Microbiology, 71*, 59-68.

Nachamkin, I., Bohachick, K., & Patton, C., M. (1993). Flagellin gene typing of *Campylobacter jejuni* by restriction fragment length polymorphism analysis. *Journal of Clinical Microbiology, 31*(6), 1531-1536.

Olsen, G. J., & Woese, C. R. (1997). Archaeal genomes: An overview. *Cell, 89*, 991-994.

Olsen, S. J., Hansen, G. R., Bartlett, L., Fitzgerald, C., A., S., Manjrekar, R., et al. (2001). An outbreak of *Campylobacter jejuni* infections associated with food handler contamination: The use of pulsed field gel electrophoresis. *The Journal of Infectious Diseases, 183*, 164-167.

Ono, K., Kurazono, T., Niwa, H., & Itoh, K. (2003). Comparison of three methods for epidemiological typing of *Campylobacter jejuni* and *C. coli*. *Current Microbiology, 47*(5), 364-371.

Ormann, T. (2005). *Manure nutrient value: Wisdom gained from experience in southern Alberta*. Paper presented at the Manure Management Planning: The Essentials. Retrieved from: http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/epw9921/$file/Troy_Ormann.pdf?OpenElement

Palmer, S. R., Gully, P. R., White, J. M., Pearson, A. D., Suckling, W. G., Jones, D. M., et al. (1983). Water-borne outbreak of *Campylobacter* gastroenteritis. *Lancet, i*, 287-290.

Papworth, L., & Olson, B. M. (2002). Manure applications on forages in southern Alberta. *Advances in Pork Production, 13*.

Parker, C. T., Miller, W. G., Horn, S. T., & Lastovica, A. J. (2007). Common genomic feature of *Campylobacter jejuni* subsp. *doylei* strains distinguish themselves from *C. jejuni* subsp. *jejuni*. *BMC Microbiology, 7*(50). Retrieved from: http://www.biomedcentral.com/content/pdf/1471-2180-7-50.pdf

Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., Basham, D., et al. (2000). The genome sequence of food-bourne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature, 403*, 665-668.

Patton, C., M., Barrett, T. J., & Morris, G. K. (1985). Comparison of the penner and lior methods for serotyping *Campylobacter* spp. *Journal of Clinical Microbiology, 22*(4), 558-565.

Patton, C. M., Wachsmuth, I. K., Evins, G. M., Kiehlbauch, J. A., Plikaytis, B. D., Troup, N., et al. (1991). Evaluation of 10 methods to distinguish epidemic-associated *Campylobacter* strains. *Journal of Clinical Microbiology, 29*(4), 680-688.

Payment, P. (2001). Transmission of gastrointestinal disease: Hygiene as the final barrier. *American Journal of Infection Control, 29*(4), 218-221.

Payne, R. E., Lee, M. D., Dreesen, D. W., & Barnhart, H. M. (1999). Molecular epidemiology of *Campylobacter jejuni* in broiler flocks using randomly amplified polymorphic DNA-PCR and 23S rRNA-PCR and role of litter in its transmission. *Applied Environmental Microbiology, 65*(1), 260-263.

Pearson, B., Pin, C., Wright, J., I'Anson, K., Humphrey, T., & Wells, J. (2003). Comparative genomic analysis of *Campylobacter jejuni* using whole genome DNA microarrays. *FEBS Letters, 554*(1), 224-230.

Pebody, R. G., Ryan, M. J., & Wall, P. G. (1997). Outbreaks of *Campylobacter* infection: rare events for a common pathogen. *Communicable Disease Report, 7*(3), R33-R37.

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

Peterson, M. C. (1994). Clinical aspects of *C. jejuni* infection in adults. *Western Journal of Medicine, 161*, 148-152.

Poly, F., & Guerry, P. (2008). Pathogenesis of *Campylobacter*. *Current Opinion in Gastroenterology, 24*, 27-31.

Poly, F., Read, T., Tribble, D., Baqar, S., Lorenzo, M., & Guerry, P. (2007). Genome sequences of *Campylobacter* from Thailand. *Infection and Immunity, 75*(7), 3425-3433.

Population Health. (2007). *Health Report 2007*. Red Deer, AB: David Thompson Health Region.

Rees, J. H., Soudain, S. E., Gregson, N. A., & Hughes, R. A. (1995). *Campylobacter jejuni* infection and Gullian-Barre syndrome. *New England Journal of Medicine, 333*, 1374-1379.

Reiersen, J., Briem, H., Hardardotti, H., Gunnarsson, E., Georgsson, F., Gudmundsdottir, E., et al. (2002). *Human campylobacteriosis epidemic in Iceland 1998-2000 and*

effect of interventions aimed at poultry and humans. Paper presented at the Food and Agriculture Organization of the United Nations and The World Health Organization Global Forum of Food Safety Regulators, Marrakech, Morocco.

Robinson, D. A. (1981). Infective dose of *Campylobacter jejuni* in milk. *British Medical Journal, 282*, 1584.

Rouillard, J. M., Zuker, M., & Gulari, E. (2003). OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Research, 31*(12), 3057-3062.

Sahin, O., Kobalka, P., & Zhang, Q. (2003). Detection and survival of *Campylobacter* in chicken eggs. *Journal of Applied Microbiology, 95*, 1017-1079.

Sails, A. D., Swaminathan, B., & Fields, P. I. (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*(10), 4733-4739.

Sanders, J. W., Putnam, S. D., Riddle, M. S., & Tribble, D. R. (2004). Military importance of diarrhea: lessons from the Middle East. *Current Opinion in Gastroenterology, 21*, 9-14.

Schrotz-King, P., Prokhorova, T. A., Nielsen, P. N., Crawford, J. S., & Morsczeck, C. (2007). *Campylobacter jejuni* proteomics for new travellers' diarrhea vaccines. *Travel Medicine and Infectious Disease, 5*, 106-109.

Simpson, E. H. (1949). Measurement of species diversity. *Nature, 163*, 688.

Skirrow, M. B. (1997). *Campylobacter* and *Helicobacter*. In D. Greenwood, R. C. Slack & J. F. Peutherer (Eds.), *Medical Microbiology* (15 ed.). Edinburgh: Churchill Livingstone.

Snelling, W. J., Matsuda, M., Moore, J. E., & Dooley, J. S. (2005). Under the microscope: *Campylobacter jejuni*. *Letters in Applied Microbiology, 41*, 297-302.

Spiller, R. C., D. Jenkins, J. P. Thornley, J. M. Hebden, T. Wright, M. Skinner, and K. R. Neat. Increased rectal mucosal enterendorine cells, T lymphocytes, and increased gut permeability following acute *Campylobacter* enteritis and in post-dysenteric irritable bowel syndrome. *Gut*, 47*,* 804-811.

Steele, M., McNab, B., Fruhner, L., DeGrandis, S., Woodward, D., & Odumeru, J. A. (1998). Epidemiological typing of *Campylobacter* isolates from meat processing

plants by pulsed-field gel electrophoresis, fatty acid profile typing, serotyping and biotyping. *Applied and Environmental Microbiology, 64*(7), 2346-2349.

Stehr-Green, J. K., Nicholls, C., McEwan, S., Payne, A., & Mitchell, P. (1991). Waterborne outbreak of *Campylobacter jejuni* in Christchurch; the importance of a combined epidemiological and microbiological investigation. *New Zealand Medical Journal, 104*(918), 356-358.

Stern, N. J., Hiett, K. L., Alfredsson, G. A., Kristinsson, K. G., Reiersen, J., Hardardottir, H., et al. (2003). *Campylobacter* spp. in Icelandic poultry operations and human disease. *Epidemiology and Infection, 130*, 23-32.

Swaminathan, B., Barrett, T. J., Hunter, S. B., Tauxe, R. V., & Cdc, P. (2001). PulseNet: The Molecular Subtyping Network for Foodborne Bacterial Disease Surveillance, United States. *Emerging Infectious Diseases, 7*(3), 382-389.

Taboada, E., van Belkum, A., Yuki, N., Acedillo, R. R., Godschalk, P. C. R., Koga, M., et al. (2007). Comparative genomics analysis of *Campylobacter jejuni* associated with Guillian-Barre and Miller Fisher syndromes: neuropathogenic and enteritis-associated isolated can share high levels of genomic similarity. *BMC Genomics, 8*, 359-370.

Taboada, E. N., Acedillo, R. R., Carrillo, C. D., Findlay, W. A., Medeiros, D. T., Mykytezuk, O., et al. (2004). Large-scale comparative genomics meta-analysis of *Campylobacter jejuni* isolates reveals low level of genome plasticity. *Journal of Clinical Microbiology, 42*(10), 4566-4576.

Taboada, E. N., MacKinnon, J. M., Luebbert, C. C., Gannon, V. P. J., Nash, J. H. E., & 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.

Tenkate, T., & Stafford, R. (2001). Risk factors for campylobacter infection in infants and young children: a matched case-control study. *Epidemiology and Infection, 127*, 399-404.

Terletski, V., Brenner, G., & Schwarz, S. (2004). Subtracted restriction fingerprinting–a new technique using magnetic capture of tagged restriction fragments. *FEMS Immunology and Medical Microbiology, 41*, 1-8.

Thomas, C., Gibson, H., Hill, D. J., & Mabey, M. (1999). *Campylobacter* epidemiology: an aquatic perspective. *Journal of Applied Microbiology, 85*, 168S-177S.

Thomas, K., Majowicz, S., Sockett, P., Fazil, A., Pollari, F., Doré, K., et al. (2006). Estimated numbers of community cases of illness due to *Samonella, Campylobacter* and Verotoxigenic *Eschichia coli:* Pathogen specific community rates. *Canadian Journal Infectious Diseases Microbiology, 17*(4), 229-234.

Tracz, D. M., Keelan, M., Ahmed-Bentley, J., Gibreel, A., Kowalewska, K., & Taylor, D. E. (2005). pVir and bloody diarrhea in *Campylobacter jejuni* enteritis. *Emerging Infectious Diseases, 11*(6), 838-843.

Van Pelt, W., Wannet, W. J. B., van de Giessen, A. W., Mevius, D. J., & van Duynhoven, Y. T. H. P. (2004). Trends in gastroenteritis (GE) in the Netherlands, 1996-2003. *Infectieziekten Bull, 15*, 335-341.

Vellinga, A., & Van Loock, F. (2002). The dioxin crisis as experiment to determine poultry-related *Campylobacter enteritis*. *Emerging Infectious Diseases, 8*(1), 19-22.

Wagenaar, J. A., Mevius, D. J., & Havelaar, A. H. (2006). *Campylobacter* in primary animal production and control strategies to reduce the burden of human campylobacteriosis. *Rev. sci. tech. Off. int. Epiz., 25*(2), 581-594.

Wagenaar, J. A., Van Bergen, M. A. P., Mueller, M. A., Wassenaar, T. M., & Carlton, R. M. (2005). Phage therapy reduces *Campylobacter jejuni* colonization in broiler. *Veterinary Microbiology, 109*, 275-283.

Waldenstrom, J., Broman, T., Carlsson, I., Hasselquist, D., Achterberg, R. P., Wagenaar, J. A., et al. (2002). Prevalence of *Campylobacter jejuni, Campylobacter lari, and Campylobacter coli* in different ecological guilds and taxa of migrating birds. *Applied and Environmental Microbiology, 68*(12), 5911-5917.

Waldenstrom, J., On, S. L., Ottvall, R., Hasselquist, D., & Olsen, B. (2007). Species diversity of campylobacteria in a wild bird community in Sweden. *Journal of Applied Microbiology, 102*(2), 424-432.

Walker, R. I., Caldwell, M. B., Lee, E. C., Guerry, P., Trust, T. J., & Ruiz-Palacios, G. M. (1986). Pathophysiology of *Campylobacter* enteritis. *Microbiological Reviews, 50*(1), 81-94.

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

Wassenaar, T. M., Geilhausen, B., & Newell, D. G. (1998). Evidence of genomic instability in *Campylobacter jejuni* isolated from poultry. *Applied and Environmental Microbiology, 64*(5), 1816-1821.

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

Whyte, R., Hudson, J. A., & Graham, C. (2006). *Campylobacter* in chicken livers and their destruction by pan frying. *Letters in Applied Microbiology, 43*, 591-595.

Woodward, D. L., & Rodgers, F. G. (2002). Identification of *Campylobacter* heat-stable and heat-labile antigens by combining the Penner and Lior serotyping schemes. *Journal of Clinical Microbiology, 40*(3), 741-745.

Yates, J. A., & Stetz, L. C. (2006). Reiter's syndrome (reactive arthritis) and travelers' diarrhea. *Journal of Travel Medicine, 12*(1), 54-56.

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

Ziprin, R. L., Hume, M. E., Young, C. R., & Harvey, R. B. (2002). Cecal colonization of chicks by porcine strains of *Campylobacter coli*. *Avian Diseases, 46*(2), 473-477.

Ziprin, R. L., Sheffield, C. L., Hume, M. E., Drinnon, D. L., & Harvey, R. B. (2003). Cecal colonization of chicks by bovine-derived strains of *Campylobacter*. *Avian Diseases, 47*(4), 1429-1433.

Appendix A: Table of 641 isolates that compose the southern Alberta *C. jejuni* data set. As well as the sample type, the source, 2-digit representation for the month of isolation, location of travel if travel occurred prior to clinical symptoms for strains isolated from patients. Month of isolation begins with month 1 set as June, 2002 and month 43 as August, 2006. The 74 isolates selected for further genetic analysis by CGF-35 and CGF-119 are indicated with an asterisk (*). Due to privacy concerns, several clinical isolates were only provided with year of isolation. This isolates were treated as month "27.5" to allow for identification and exclusion from concurrent month calculations. Similarly, not all water locations could be released and are listed as "water" under source. Isolates previously evaluated in-house in Guelph, Ontario using several typing methods and fulfilled the purpose of known reference strains, and are labeled as "reference" to separate these from isolates obtained in Alberta. Continued on the following pages.

Key	**Sample Type**	**Source $_{Travel}$**	**Month**	**CGF-35/119**
CJ001	Animal	Pig	29	
CJ002	Animal	Pig	29	
CJ003	Water	Water	27	*
CJ004	Water	Water	27	*
CJ005	Water	Water	27	*
CJ006	Water	Water	27	
CJ007	Water	Water	27	
CJ008	Water	Water	29	*
CJ009	Water	Water	29	
CJ010	Water	Water	29	*
CJ011	Animal	Cattle	29	*
CJ012	Animal	Cattle	29	*
CJ013	Animal	Sheep	29	*
CJ014	Animal	Cattle	29	
CJ015	Animal	Cattle	29	
CJ016	Animal	Cattle	29	
CJ017	Animal	Horse	31	*
CJ018	Animal	Horse	31	
CJ019	Water	LB4W2	20	
CJ020	Water	LB4-14	20	
CJ021	Animal	Cattle	31	
CJ022	Animal	Horse	31	
CJ023	Animal	Pig	41	
CJ024	Animal	Pig	41	
CJ025	Clinical	Human	29	
CJ026	Clinical	Human	30	
CJ027	Reference	Chicken	18	
CJ028	Reference	Chicken	18	
CJ029	Clinical	Human	29	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ030	Clinical	Human	29	
CJ031	Clinical	Human	29	
CJ032	Clinical	Human	30	
CJ033	Clinical	Human	30	
CJ034	Reference	Human	17	
CJ035	Animal	Cattle	18	
CJ036	Reference	Chicken	N/A	
CJ037	Clinical	Human	27.5	*
CJ038	Water	Water	15	*
CJ039	Water	Grand Falls	20	
CJ040	Animal	Goose	41	*
CJ041	Clinical	Human	28	
CJ042	Clinical	Human	27.5	
CJ043	Animal	Goose	42	
CJ044	Animal	Duck	42	
CJ045	Water	LB4W3	20	
CJ046	Animal	Cattle	31	
CJ047	Animal	Duck	19	
CJ048	Water	B7	19	
CJ049	Reference	Wild avian	17	
CJ050	Water	SU-53	42	
CJ051	Clinical	Human	30	
CJ052	Clinical	Human	30	
CJ053	Clinical	Human	30	
CJ054	Water	B5	19	
CJ055	Water	LB4W	19	
CJ056	Clinical	Human	30	
CJ057	Animal	Sheep	16	
CJ058	Animal	Cattle	17	
CJ059	Animal	Cattle	17	
CJ060	Animal	Sheep	40	
CJ061	Animal	Cattle	33	
CJ062	Clinical	Human	29	
CJ063	Clinical	Human	29	
CJ064	Reference	Reference	19	

Key	Sample Type	Source $_{Travel}$	Month	CGF-35/119
CJ065	Animal	Duck	42	
CJ066	Clinical	Human	27.5	
CJ067	Water	Water	1	
CJ068	Water	Water	15	
CJ069	Water	Water	16	
CJ070	Animal	Goose	19	*
CJ071	Animal	Goose	19	
CJ072	Animal	Goose	19	
CJ073	Water	LB4W	19	
CJ074	Water	LB4	20	*
CJ075	Animal	Duck	20	
CJ076	Animal	Duck	41	
CJ077	Animal	Duck	41	*
CJ078	Animal	Goose	42	*
CJ079	Animal	Duck	41	
CJ080	Animal	Sheep	31	
CJ081	Animal	Sheep	33	
CJ082	Animal	Sheep	33	
CJ083	Clinical	Human	29	
CJ084	Clinical	Human	30	
CJ085	Clinical	Human	30	
CJ086	Clinical	Human	22	
CJ087	Water	Water	40	
CJ088	Water	SU-52	42	
CJ089	Reference	Human	8	
CJ090	Reference	Human	9	
CJ091	Water	B4	19	
CJ092	Clinical	Human	30	
CJ093	Clinical	Human	30	
CJ094	Clinical	Human	30	
CJ095	Clinical	Human	30	
CJ096	Clinical	Human	30	
CJ097	Clinical	Human	29	
CJ098	Water	LB4W	20	
CJ099	Water	LB4W2	20	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ100	Reference	Chicken	17	
CJ101	Animal	Sheep	31	
CJ102	Animal	Sheep	31	
CJ103	Clinical	Human	27	
CJ104	Clinical	Human	27	
CJ105	Clinical	Human	27	
CJ106	Animal	Sheep	31	
CJ107	Water	B7	19	
CJ108	Water	B5	19	
CJ109	Clinical	Human	27.5	
CJ110	Clinical	Human	28	*
CJ111	Clinical	Human	16	*
CJ112	Clinical	Human	29	*
CJ113	Animal	Cattle	17	*
CJ114	Animal	Cattle	17	*
CJ115	Animal	Sheep	29	*
CJ116	Animal	Sheep	29	*
CJ117	Animal	Sheep	33	*
CJ118	Animal	Goat	33	*
CJ119	Animal	Cattle	43	*
CJ120	Clinical	Human	27.5	
CJ121	Animal	Cattle	17	
CJ122	Animal	Cattle	17	
CJ123	Animal	Cattle	30	
CJ124	Animal	Cattle	17	
CJ125	Clinical	Human	27	
CJ126	Clinical	Human	27	*
CJ127	Clinical	Human	29	*
CJ128	Clinical	Human	29	*
CJ129	Clinical	Human	29	
CJ130	Clinical	Human	29	
CJ131	Clinical	Human	29	
CJ132	Clinical	Human	29	
CJ133	Clinical	Human	29	
CJ134	Clinical	Human	29	

Key	Sample Type	Source $_{Travel}$	Month	CGF-35/119
CJ135	Clinical	Human	29	
CJ136	Clinical	Human	29	
CJ137	Clinical	Human	29	
CJ138	Clinical	Human	29	
CJ139	Clinical	Human	30	
CJ140	Clinical	Human	30	
CJ141	Clinical	Human	27.5	
CJ142	Animal	Chicken	18	*
CJ143	Animal	Chicken	18	
CJ144	Animal	Chicken	18	
CJ145	Water	Pine Creek	29	*
CJ146	Animal	Goose	41	
CJ147	Animal	Pelican	41	
CJ148	Animal	Duck	42	
CJ149	Animal	Chicken	42	
CJ150	Animal	Chicken	42	
CJ151	Animal	Chicken	42	
CJ152	Reference	Reference	N/A	
CJ153	Reference	Human	9	
CJ154	Animal	Pig	41	
CJ155	Animal	Pig	41	
CJ156	Animal	Chicken	20	
CJ157	Animal	Chicken	20	
CJ158	Clinical	Human	27.5	
CJ159	Clinical	Human	15.5	
CJ160	Animal	Pig	18	
CJ161	Animal	Pig	18	
CJ162	Animal	Cattle	18	
CJ163	Animal	Sheep	29	
CJ164	Clinical	Human	30	
CJ165	Clinical	Human	30	
CJ166	Clinical	Human	30	
CJ167	Water	Water	18	
CJ168	Animal	Sheep	16	
CJ169	Animal	Sheep	16	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ170	Animal	Cattle	29	
CJ171	Animal	Chicken	41	
CJ172	Clinical	Human	29	
CJ173	Reference	Human	9	
CJ174	Animal	Cattle	17	
CJ175	Animal	Cattle	18	
CJ176	Animal	Cattle	33	
CJ177	Animal	Cattle	33	
CJ178	Animal	Cattle	33	
CJ179	Animal	Cattle	33	
CJ180	Animal	Cattle	33	
CJ181	Animal	Chicken	41	
CJ182	Animal	Cattle	41	
CJ183	Clinical	Human	27	
CJ184	Clinical	Human	29	
CJ185	Animal	Cattle	17	
CJ186	Animal	Cattle	40	
CJ187	Animal	Cattle	40	
CJ188	Animal	Cattle	40	
CJ189	Clinical	Human	27	
CJ190	Clinical	Human	27	
CJ191	Clinical	Human	29	
CJ192	Clinical	Human	29	
CJ193	Clinical	Human	29	
CJ194	Animal	Cattle	18	
CJ195	Reference	Human	17	
CJ196	Reference	Human	18	
CJ197	Animal	Cattle	30	
CJ198	Animal	Cattle	17	
CJ199	Animal	Cattle	17	
CJ200	Animal	Cattle	17	
CJ201	Clinical	Human	29	
CJ202	Clinical	Human	29	
CJ203	Water	Water	18	
CJ204	Reference	Pig	N/A	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ205	Clinical	Human	27.5	
CJ206	Animal	Pig	1	
CJ207	Animal	Pig	1	
CJ208	Animal	Turkey	18	
CJ209	Animal	Pig	18	
CJ210	Animal	Pig	18	
CJ211	Animal	Pig	17	
CJ212	Animal	Chicken	39	
CJ213	Animal	Chicken	39	
CJ214	Animal	Chicken	39	
CJ215	Animal	Chicken	39	
CJ216	Animal	Chicken	39	
CJ217	Animal	Chicken	39	
CJ218	Animal	Chicken	39	
CJ219	Animal	Pig	41	*
CJ220	Animal	Chicken	41	
CJ221	Animal	Pig	41	*
CJ222	Animal	Pig	41	
CJ223	Animal	Pig	41	
CJ224	Animal	Pig	41	
CJ225	Animal	Pig	41	
CJ226	Animal	Pig	41	
CJ227	Clinical	Human	27	
CJ228	Clinical	Human	28	
CJ229	Clinical	Human	29	
CJ230	Clinical	Human	29	
CJ231	Water	Water	1	
CJ232	Animal	Cattle	31	
CJ233	Clinical	Human	29	
CJ234	Clinical	Human	29	
CJ235	Clinical	Human	29	
CJ236	Animal	Cattle	29	
CJ237	Animal	Cattle	29	
CJ238	Clinical	Human	27.5	
CJ239	Clinical	Human	15.5	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ240	Clinical	Human	29	
CJ241	Clinical	Human	29	
CJ242	Clinical	Human	29	
CJ243	Clinical	Human	29	
CJ244	Clinical	Human	29	
CJ245	Clinical	Human	29	
CJ246	Clinical	Human	29	
CJ247	Clinical	Human	29	
CJ248	Clinical	Human	29	
CJ249	Clinical	Human	29	*
CJ250	Clinical	Human	29	*
CJ251	Clinical	Human	29	
CJ252	Clinical	Human	29	
CJ253	Clinical	Human	30	
CJ254	Clinical	Human	30	
CJ255	Clinical	Human	30	
CJ256	Clinical	Human	27	
CJ257	Clinical	Human	29	
CJ258	Clinical	Human	30	
CJ259	Clinical	Human	30	
CJ260	Clinical	Human	15.5	
CJ261	Water	Casselman	20	
CJ262	Water	B5	20	
CJ263	Water	Water	29	
CJ264	Clinical	Human	27	*
CJ265	Clinical	Human	27	*
CJ266	Clinical	Human	28	*
CJ267	Clinical	Human	28	*
CJ268	Clinical	Human	28	
CJ269	Clinical	Human	28	
CJ270	Clinical	Human	28	
CJ271	Clinical	Human	29	*
CJ272	Clinical	Human	27.5	*
CJ273	Water	Water	18	
CJ274	Water	Water	18	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ275	Water	B4	19	
CJ276	Water	B4	19	
CJ277	Water	B4	20	*
CJ278	Animal	Cattle	28	*
CJ279	Animal	Cattle	40	*
CJ280	Animal	Chicken	18	
CJ281	Animal	Chicken	18	
CJ282	Clinical	Human	28	
CJ283	Clinical	Human	31	
CJ284	Animal	Chicken	20	
CJ285	Animal	Chicken	20	
CJ286	Animal	Chicken	20	
CJ287	Animal	Chicken	20	
CJ288	Animal	Chicken	20	
CJ289	Clinical	Human	27	
CJ290	Clinical	Human	28	
CJ291	Clinical	Human	27.5	
CJ292	Clinical	Human Mexico	27	
CJ293	Clinical	Human	27.5	
CJ294	Clinical	Human Mexico	27	
CJ295	Clinical	Human	28	
CJ296	Clinical	Human	28	
CJ297	Clinical	Human	28	
CJ298	Clinical	Human	28	
CJ299	Reference	Chicken	18	
CJ300	Animal	Chicken	18	
CJ301	Animal	Chicken	18	
CJ302	Clinical	Human	30	
CJ303	Animal	Cattle	18	
CJ304	Animal	Cattle	18	
CJ305	Clinical	Human	30	
CJ306	Clinical	Human	15.5	
CJ307	Clinical	Human	27.5	
CJ308	Clinical	Human	27.5	
CJ309	Clinical	Human	29	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ310	Clinical	Human	15.5	*
CJ311	Animal	Cattle	18	*
CJ312	Animal	Sheep	30	
CJ313	Animal	Sheep	30	
CJ314	Animal	Sheep	30	
CJ315	Animal	Sheep	30	
CJ316	Animal	Sheep	31	
CJ317	Clinical	Human	27.5	
CJ318	Clinical	Human	29	*
CJ319	Clinical	Human	30	*
CJ320	Animal	Buffalo	17	*
CJ321	Animal	Buffalo	17	*
CJ322	Animal	Horse	19	*
CJ323	Animal	Cattle	30	*
CJ324	Animal	Turkey	31	*
CJ325	Clinical	Human	27	
CJ326	Clinical	Human	27	
CJ327	Clinical	Human	28	
CJ328	Clinical	Human	27.5	
CJ329	Clinical	Human	29	
CJ330	Clinical	Human	27.5	
CJ331	Clinical	Human	15.5	
CJ332	Reference	Human	8	
CJ333	Reference	Human	9	
CJ334	Reference	Reference	17	
CJ335	Reference	Human	17	
CJ336	Reference	Human	17	
CJ337	Reference	Human	17	
CJ338	Reference	Reference	19	
CJ339	Clinical	Human	28	
CJ340	Water	Water	1	
CJ341	Animal	Cattle	16	
CJ342	Animal	Cattle	16	
CJ343	Water	LB4	20	
CJ344	Water	LB4W	20	

Key	Sample Type	Source $_{Travel}$	Month	CGF-35/119
CJ345	Water	LB4W3	20	
CJ346	Water	LB4-14	20	
CJ347	Clinical	Human	28	
CJ348	Clinical	Human	28	
CJ349	Clinical	Human	28	
CJ350	Clinical	Human	29	
CJ351	Clinical	Human	29	
CJ352	Clinical	Human	29	
CJ353	Clinical	Human	15.5	
CJ354	Clinical	Human	15.5	
CJ355	Clinical	Human	27.5	*
CJ356	Reference	Human	8	
CJ357	Reference	Dog	18	*
CJ358	Reference	Chicken	18	
CJ359	Clinical	Human	27	
CJ360	Clinical	Human	28	
CJ361	Clinical	Human	30	
CJ362	Animal	Cattle	16	
CJ363	Reference	Reference	17	
CJ364	Animal	Sheep	18	
CJ365	Animal	Cattle	31	
CJ366	Animal	Cattle	31	
CJ367	Animal	Cattle	31	
CJ368	Animal	Cattle	31	
CJ369	Animal	Horse	31	
CJ370	Animal	Cattle	33	
CJ371	Animal	Cattle	33	
CJ372	Clinical	Human	28	
CJ373	Clinical	Human	30	*
CJ374	Clinical	Human	15.5	*
CJ375	Animal	Cattle	30	*
CJ376	Animal	Cattle	30	*
CJ377	Clinical	Human	29	
CJ378	Clinical	Human	29	
CJ379	Clinical	Human	29	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ380	Reference	Reference	17	
CJ381	Animal	Sheep	19	
CJ382	Animal	Sheep	19	
CJ383	Clinical	Human	27.5	
CJ384	Animal	Cattle	17	
CJ385	Clinical	Human	27	
CJ386	Clinical	Human	30	
CJ387	Animal	Cattle	17	
CJ388	Animal	Deer	19	
CJ389	Clinical	Human Mexico	28	*
CJ390	Clinical	Human	28	*
CJ391	Clinical	Human	28	*
CJ392	Clinical	Human	28	
CJ393	Clinical	Human	29	*
CJ394	Clinical	Human	29	*
CJ395	Clinical	Human	29	
CJ396	Clinical	Human	29	*
CJ397	Clinical	Human	29	
CJ398	Clinical	Human	29	
CJ399	Clinical	Human	29	
CJ400	Clinical	Human	29	
CJ401	Clinical	Human	30	*
CJ402	Clinical	Human	30	*
CJ403	Clinical	Human	30	
CJ404	Clinical	Human	30	
CJ405	Clinical	Human	30	
CJ406	Clinical	Human	15.5	
CJ407	Clinical	Human	15.5	
CJ408	Reference	Human	18	
CJ409	Animal	Chicken	21	*
CJ410	Clinical	Human Philippines	28	
CJ411	Clinical	Human	???	
CJ412	Animal	Chicken	21	
CJ413	Animal	Chicken	21	
CJ414	Clinical	Human	30	

Key	Sample Type	Source $_{Travel}$	Month	CGF-35/119
CJ415	Clinical	Human	30	
CJ416	Clinical	Human	30	
CJ417	Clinical	Human	27.5	
CJ418	Clinical	Human	28	
CJ419	Clinical	Human	29	*
CJ420	Clinical	Human	29	*
CJ421	Clinical	Human	29	*
CJ422	Clinical	Human	29	*
CJ423	Clinical	Human	29	
CJ424	Clinical	Human	29	
CJ425	Clinical	Human	29	
CJ426	Clinical	Human	29	*
CJ427	Clinical	Human	29	*
CJ428	Clinical	Human	29	
CJ429	Clinical	Human	30	
CJ430	Clinical	Human	30	
CJ431	Clinical	Human	30	
CJ432	Clinical	Human	30	
CJ433	Clinical	Human	27.5	*
CJ434	Clinical	Human	27.5	
CJ435	Clinical	Human	27.5	
CJ436	Animal	Cattle	17	*
CJ437	Reference	Reference	17	
CJ438	Reference	Reference	17	
CJ439	Animal	Cattle	30	
CJ440	Animal	Buffalo	33	
CJ441	Animal	Buffalo	33	
CJ442	Animal	Chicken	39	*
CJ443	Animal	Chicken	41	
CJ444	Animal	Chicken	41	
CJ445	Animal	Chicken	41	
CJ446	Clinical	Human	30	
CJ447	Reference	Chicken	18	
CJ448	Reference	Human	18	
CJ449	Reference	Human	18	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ450	Clinical	Human	30	
CJ451	Clinical	Human	30	
CJ452	Clinical	Human	30	
CJ453	Clinical	Human	27	
CJ454	Clinical	Human Costa Rica	28	
CJ455	Clinical	Human	29	
CJ456	Water	Water	16	
CJ457	Animal	Deer	18	
CJ458	Animal	Deer	19	
CJ459	Reference	Sheep	17	
CJ460	Reference	Human	N/A	
CJ461	Clinical	Human	15	
CJ462	Clinical	Human	15	
CJ463	Water	Sewage	1	
CJ464	Animal	Cattle	31	
CJ465	Water	Water	16	
CJ466	Animal	Cattle	31	
CJ467	Animal	Sea Gull	19	
CJ468	Animal	Cattle	17	
CJ469	Water	Grand Falls	19	
CJ470	Animal	Sheep	31	
CJ471	Clinical	Human	28	
CJ472	Clinical	Human	15.5	
CJ473	Clinical	Human	29	
CJ474	Animal	Cattle	40	
CJ475	Clinical	Human	18	
CJ476	Animal	Human	31	
CJ477	Water	LB4-14	19	
CJ478	Water	B7	20	
CJ479	Water	LB4W3	19	
CJ480	Animal	Horse	29	
CJ481	Water	B6	20	
CJ482	Clinical	Human	29	
CJ483	Reference	Human	18	
CJ484	Water	Grand Falls	20	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ485	Animal	Human	31	
CJ486	Animal	Cattle	40	
CJ487	Animal	Goose	19	
CJ488	Reference	Wild avian	18	
CJ489	Water	Water	18	
CJ490	Animal	Duck	42	
CJ491	Animal	Pelican	31	
CJ492	Water	Water	29	
CJ493	Water	SU-54	42	
CJ494	Animal	Goose	30	
CJ495	Animal	Chicken	20	
CJ496	Water	Water	17	
CJ497	Animal	Duck	19	
CJ498	Animal	Chicken	41	
CJ499	Water	B4	20	
CJ500	Animal	Cattle	18	
CJ501	Animal	Duck	19	
CJ502	Animal	Duck	19	
CJ503	Water	LB4W	19	
CJ504	Water	Water	20	
CJ505	Animal	Goose	42	
CJ506	Clinical	Human	27.5	
CJ507	Water	Water	16	
CJ508	Water	Water Treatment	21	
CJ509	Animal	Goose	30	
CJ510	Water	Water	18	
CJ511	Animal	Goose	19	
CJ512	Animal	Goat	33	
CJ513	Animal	Duck	20	
CJ514	Animal	Goose	20	
CJ515	Water	LB-4	1	
CJ516	Water	LB4-14	20	
CJ517	Animal	Sheep	40	
CJ518	Animal	Deer	18	
CJ519	Animal	Chicken	18	

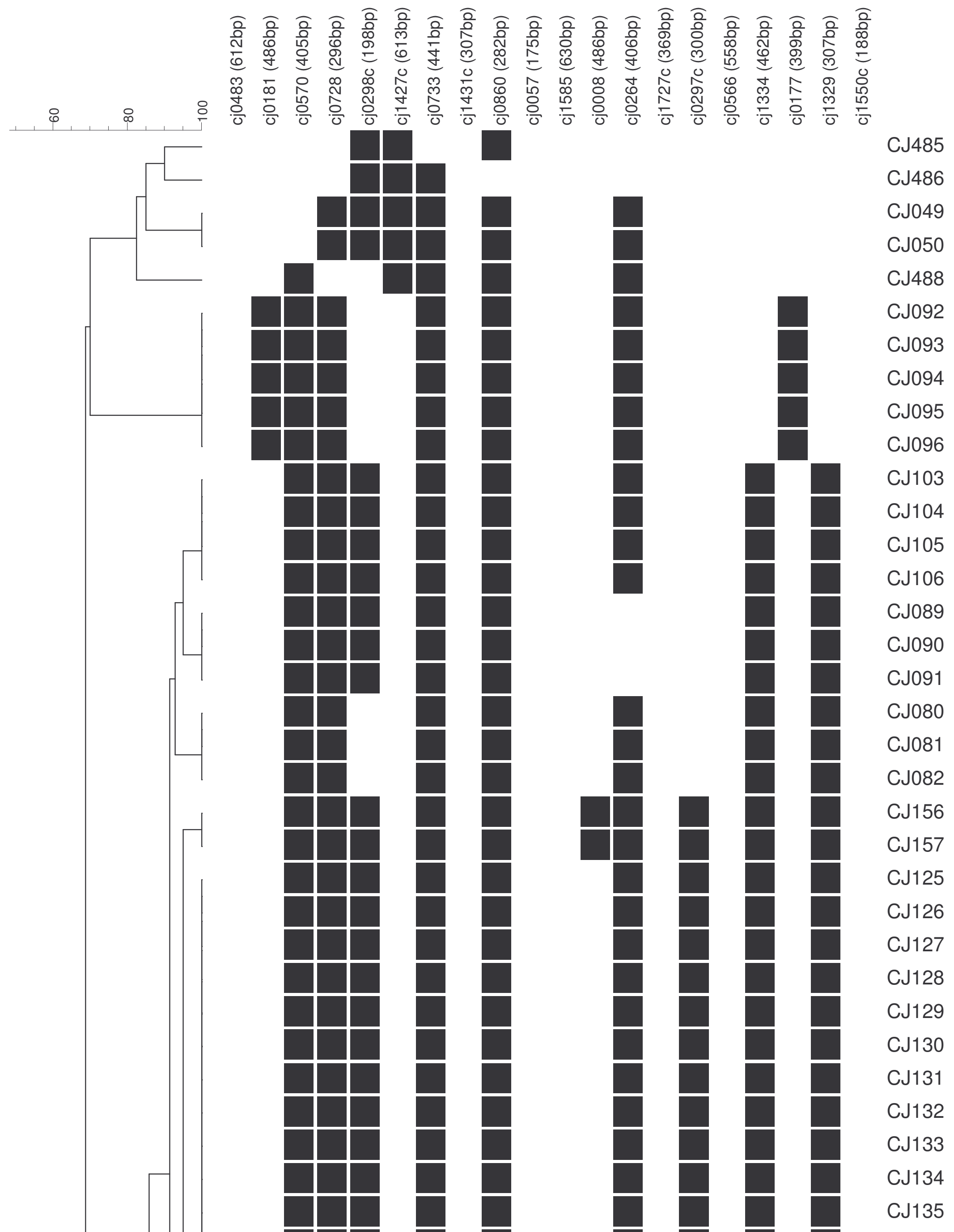
Key	Sample Type	Source Travel	Month	CGF-35/119
CJ520	Clinical	Human Mexico	27	
CJ521	Reference	Human	9	
CJ522	Reference	Reference	19	
CJ523	Animal	Cattle	29	
CJ524	Animal	Cattle	29	
CJ525	Animal	Sheep	29	
CJ526	Animal	Deer	22	
CJ527	Animal	Pig	41	
CJ528	Animal	Horse	29	
CJ529	Animal	Goat	41	
CJ530	Animal	Pig	1	
CJ531	Water	B5	19	
CJ532	Animal	Horse	29	
CJ533	Animal	Cattle	18	
CJ534	Animal	Duck	41	
CJ535	Animal	Cattle	17	
CJ536	Water	Water	42	
CJ537	Water	B4	19	
CJ538	Animal	Goose	42	
CJ539	Animal	Goose	41	
CJ540	Clinical	Human	27.5	
CJ541	Water	LB4W2	20	
CJ542	Clinical	Human	29	
CJ543	Clinical	Human	30	
CJ544	Animal	Pig	17	
CJ545	Clinical	Human	27	
CJ546	Water	Sewage	21	
CJ547	Animal	Duck	42	
CJ548	Water	Grand Falls	19	
CJ549	Animal	Duck	42	
CJ550	Clinical	Human	27.5	
CJ551	Animal	Goose	42	
CJ552	Reference	Human	18	
CJ553	Clinical	Human	27	
CJ554	Water	LB-4	19	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ555	Clinical	Human	27.5	
CJ556	Clinical	Human	15.5	
CJ557	Clinical	Human	28	
CJ558	Animal	Chicken	20	
CJ559	Clinical	Human	29	
CJ560	Reference	Human	17	
CJ561	Water	LB4W2	19	
CJ562	Clinical	Human	28	
CJ563	Clinical	Human	27.5	
CJ564	Clinical	Human	30	
CJ565	Animal	Cat	19	
CJ566	Animal	Cattle	41	
CJ567	Animal	Cattle	18	
CJ568	Animal	Cattle	42	
CJ569	Animal	Cattle	18	
CJ570	Animal	Goat	41	
CJ571	Water	Water	17	
CJ572	Water	LB4W2	19	
CJ573	Water	B6	19	
CJ574	Water	B5	20	
CJ575	Animal	Chicken	39	
CJ576	Animal	Pig	19	
CJ577	Water	SU-51	42	
CJ578	Animal	Pig	41	
CJ579	Reference	Human	9	
CJ580	Animal	Goose	41	
CJ581	Clinical	Human Guatemala	28	
CJ582	Clinical	Human	27.5	
CJ583	Animal	Pig	1	
CJ584	Animal	Pig	41	
CJ585	Animal	Pig	17	
CJ586	Clinical	Human	29	
CJ587	Clinical	Human	15.5	
CJ588	Animal	Cattle	18	
CJ589	Animal	Goat	18	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ590	Animal	Cattle	40	
CJ591	Reference	Cattle	17	
CJ592	Reference	Human	8	
CJ593	Clinical	Human Mexico	28	
CJ594	Clinical	Human Mexico	27	
CJ595	Clinical	Human	30	
CJ596	Clinical	Human Costa Rica	27	
CJ597	Clinical	Human		
CJ598	Water	B6	20	
CJ599	Clinical	Human	27.5	
CJ600	Clinical	Human	27	
CJ601	Clinical	Human	27	
CJ602	Clinical	Human	30	
CJ603	Clinical	Human	30	
CJ604	Clinical	Human	27.5	
CJ605	Animal	Chicken	39	
CJ606	Clinical	Human	27.5	
CJ607	Clinical	Human	15.5	
CJ608	Clinical	Human	28	
CJ609	Clinical	Human	15.5	
CJ610	Water	B5	19	
CJ611	Water	Water	18	
CJ612	Water	Water	18	
CJ613	Water	Water	1	
CJ614	Clinical	Human	29	
CJ615	Water	B5	20	
CJ616	Water	B5	19	
CJ617	Water	Water	16	
CJ618	Water	LB4-14	19	
CJ619	Clinical	Human	28	
CJ620	Animal	Sheep	31	
CJ621	Clinical	Human	27	
CJ622	Animal	Chicken	20	
CJ623	Clinical	Human	29	
CJ624	Animal	Cattle	29	

Key	Sample Type	Source Travel	Month	CGF-35/119
CJ625	Reference	Dog	17	
CJ626	Animal	Chicken	21	
CJ627	Clinical	Human	29	
CJ628	Clinical	Human	29	
CJ629	Water	Water	18	
CJ630	Clinical	Human	27.5	
CJ631	Clinical	Human	27.5	
CJ632	Clinical	Human	29	
CJ633	Reference	Human	8	
CJ634	Water	LB4W2	20	
CJ635	Animal	Chicken	21	
CJ636	Reference	Reference	17	
CJ637	Animal	Cattle	43	
CJ638	Reference	Human	18	
CJ639	Animal	Cattle	17	
CJ640	Clinical	Human	30	
CJ641	Clinical	Human	30	

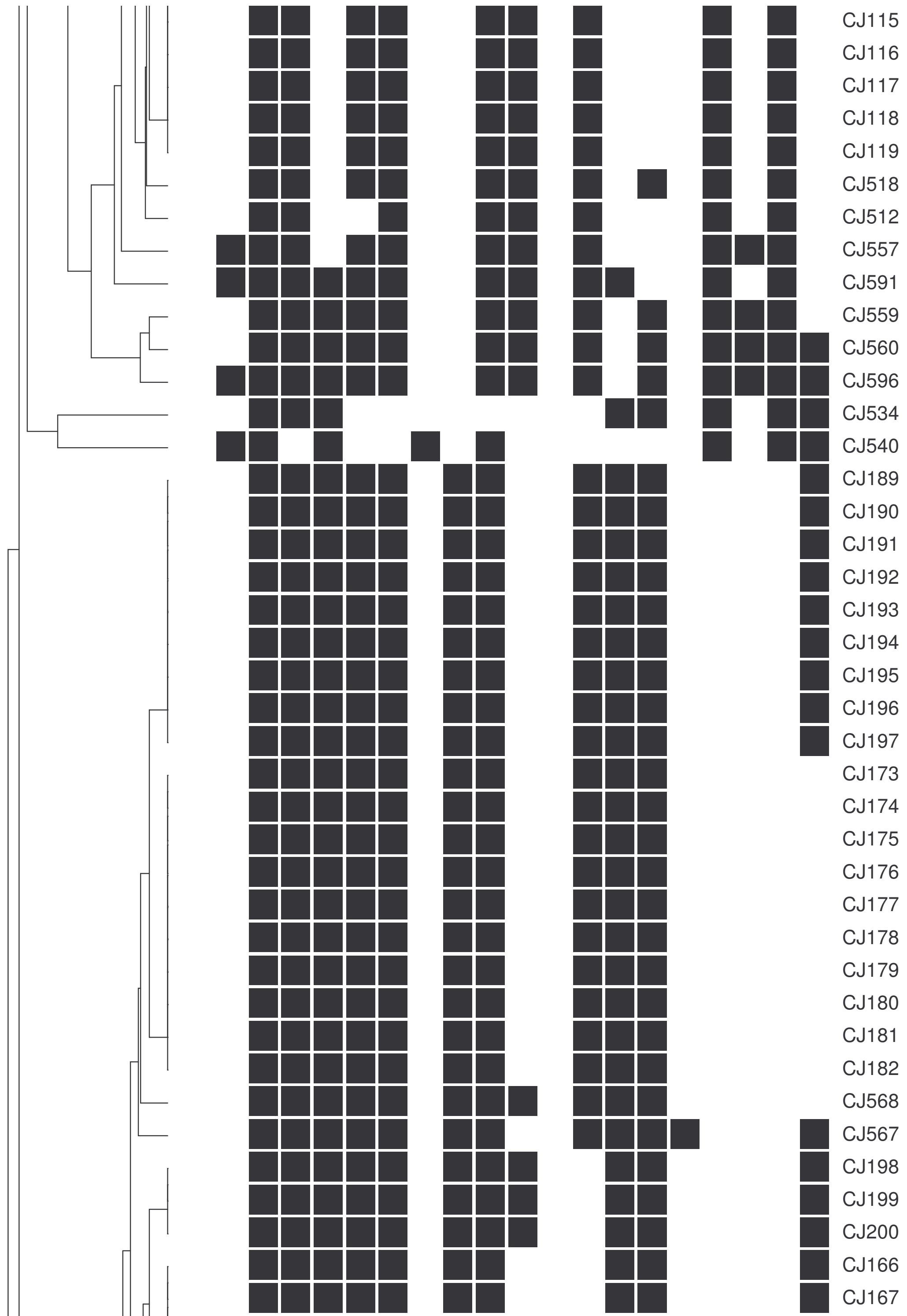
Appendix B: Detailed Clustering of the southern Alberta Data Set using Bionumerics v5.0 (Applied Maths, Austin, TX). Fingerprint clustering was performed using BioNumerics software to determine and present similarity of fingerprints using program generated dendrograms based on the unique CGF-20 genetic fingerprint. This data was then combined with available information and epidemiological data such as name, *fla*A-RFLP type, and source of sample and date of isolation. Continued on the following pages.

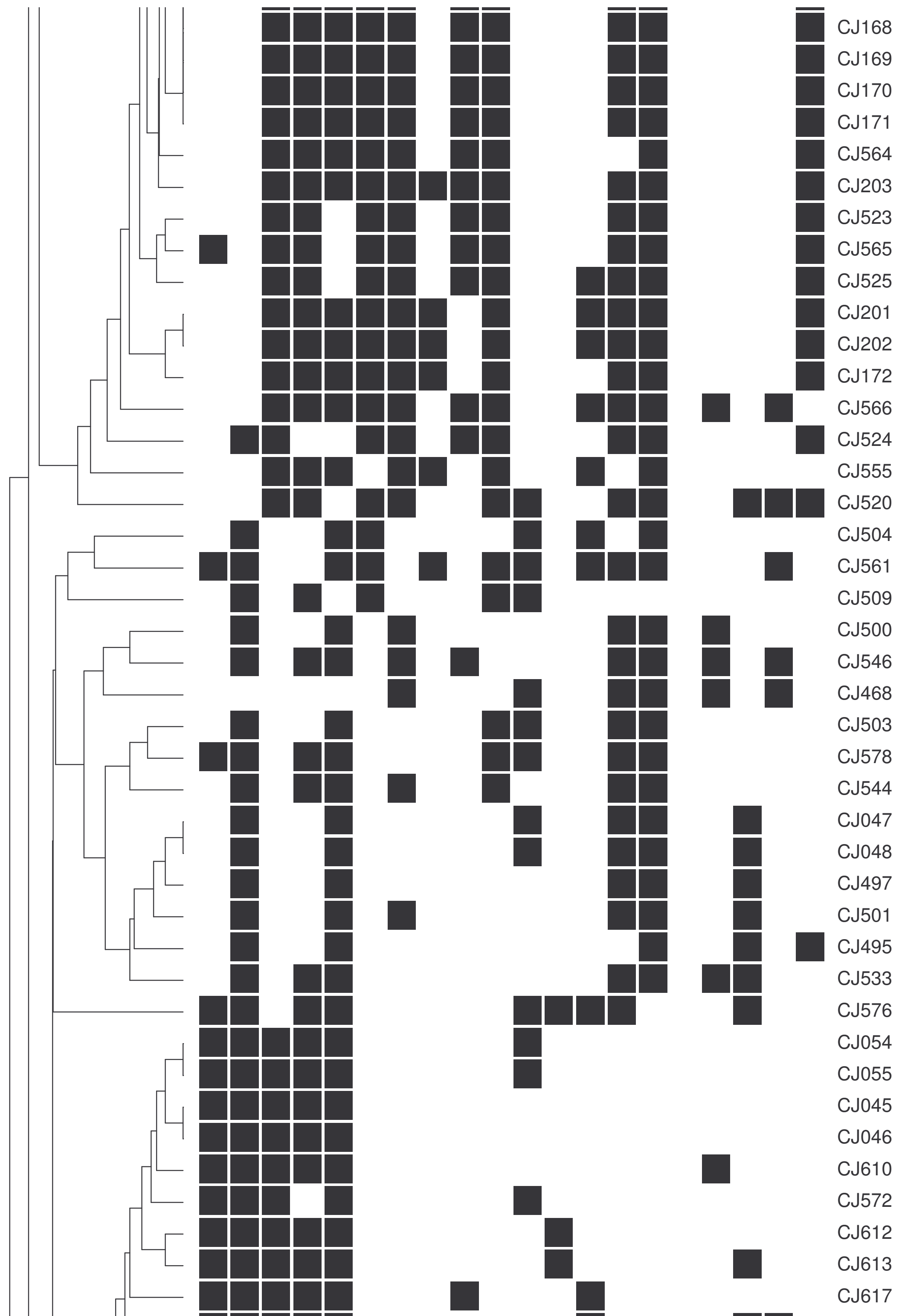


CJ136
CJ137
CJ138
CJ139
CJ140
CJ141
CJ142
CJ143
CJ144
CJ145
CJ146
CJ147
CJ148
CJ149
CJ150
CJ151
CJ556
CJ547
CJ548
CJ558
CJ538
CJ545
CJ542
CJ620
CJ097
CJ098
CJ099
CJ051
CJ052
CJ053
CJ543
CJ164
CJ165
CJ056
CJ057
CJ058
CJ059
CJ060
CJ061

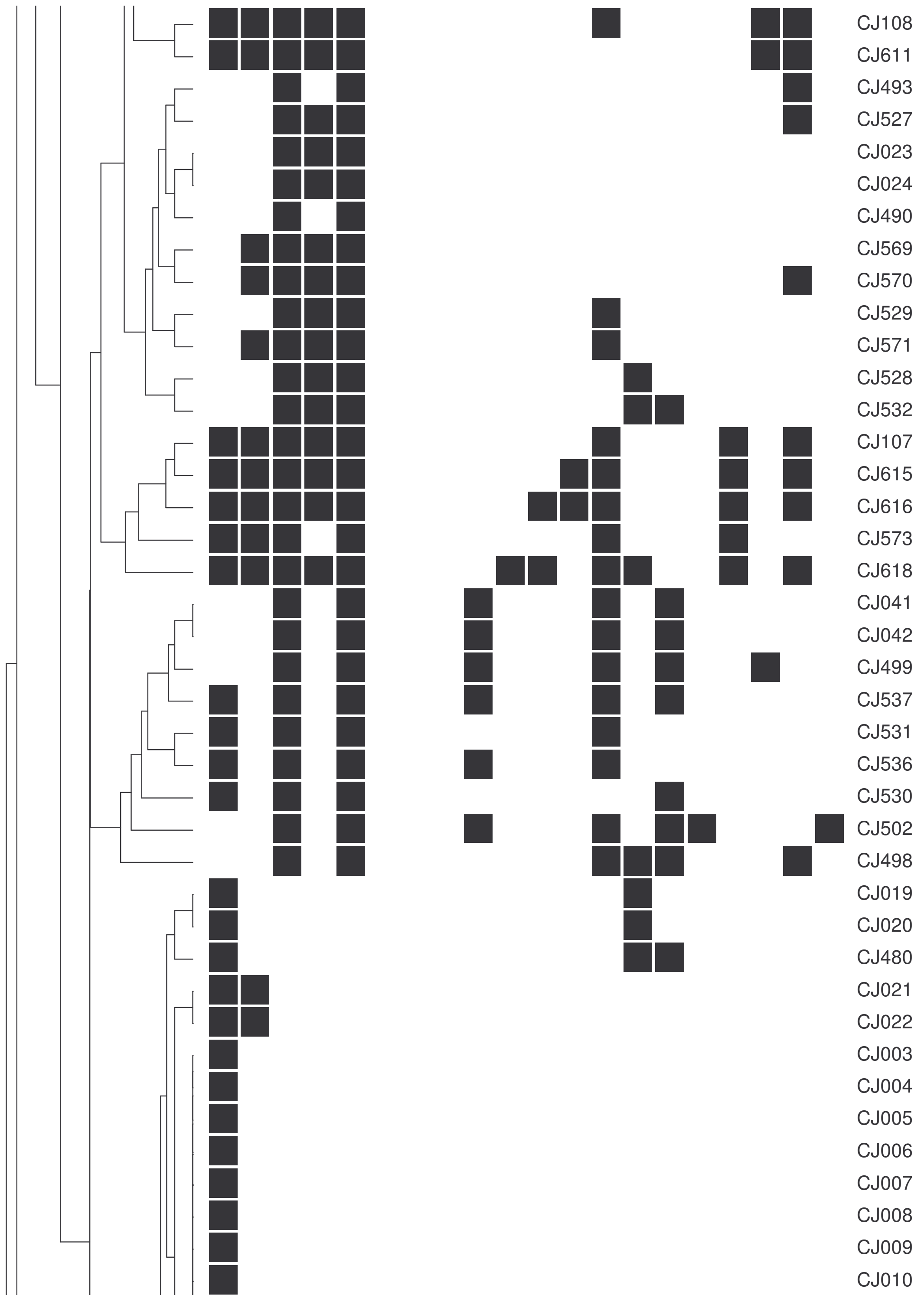
CJ470
CJ100
CJ101
CJ102
CJ083
CJ084
CJ085
CJ086
CJ029
CJ030
CJ031
CJ032
CJ033
CJ034
CJ035
CJ036
CJ025
CJ026
CJ027
CJ028
CJ483
CJ062
CJ063
CJ064
CJ065
CJ496
CJ467
CJ478
CJ087
CJ088
CJ506
CJ037
CJ038
CJ039
CJ040
CJ515
CJ553
CJ043
CJ044
CJ510

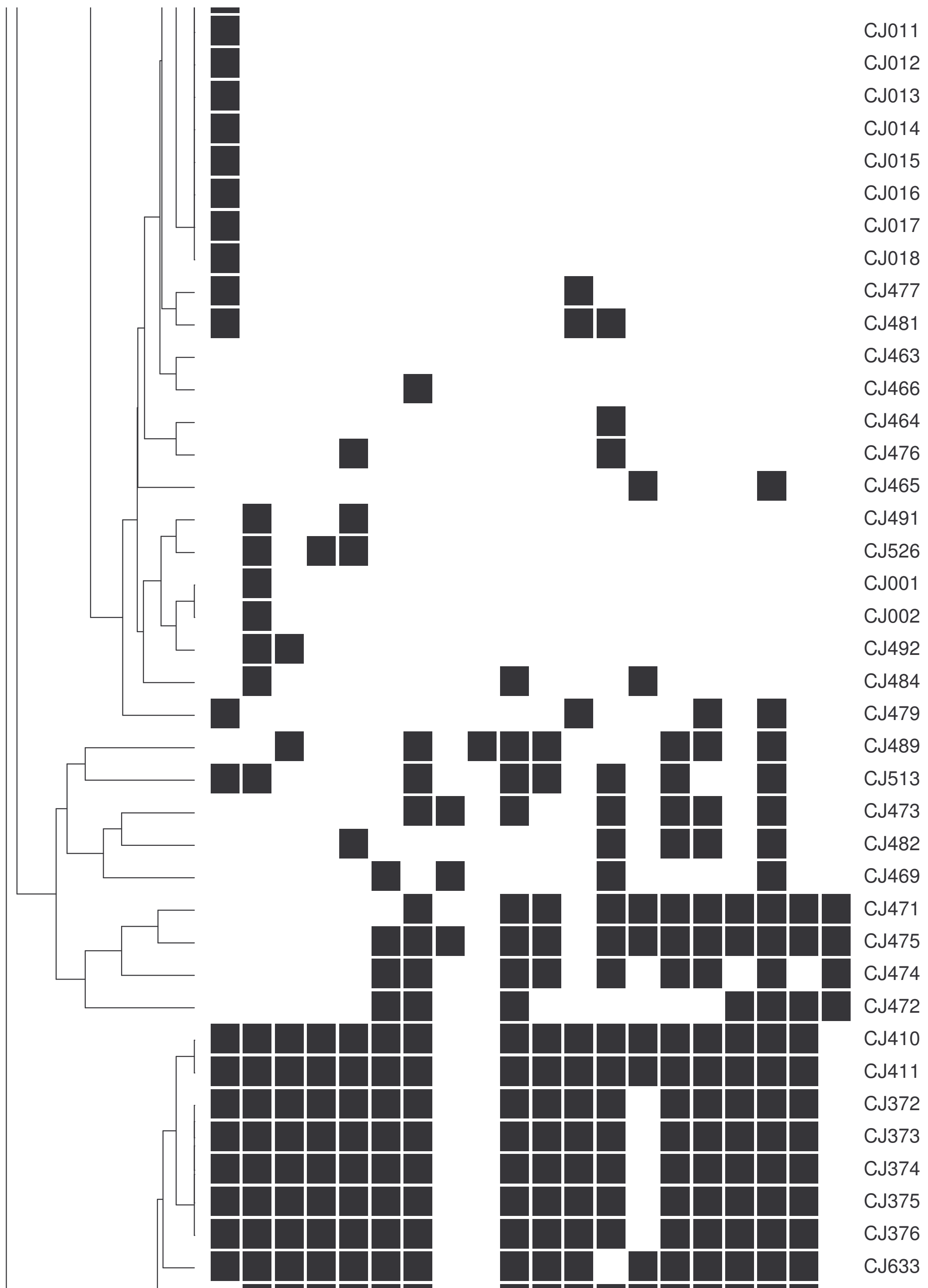
CJ066
CJ067
CJ068
CJ069
CJ070
CJ071
CJ072
CJ073
CJ074
CJ075
CJ076
CJ077
CJ078
CJ079
CJ505
CJ541
CJ539
CJ516
CJ554
CJ580
CJ487
CJ508
CJ507
CJ494
CJ511
CJ577
CJ535
CJ152
CJ153
CJ521
CJ517
CJ162
CJ163
CJ109
CJ110
CJ111
CJ112
CJ113
CJ114

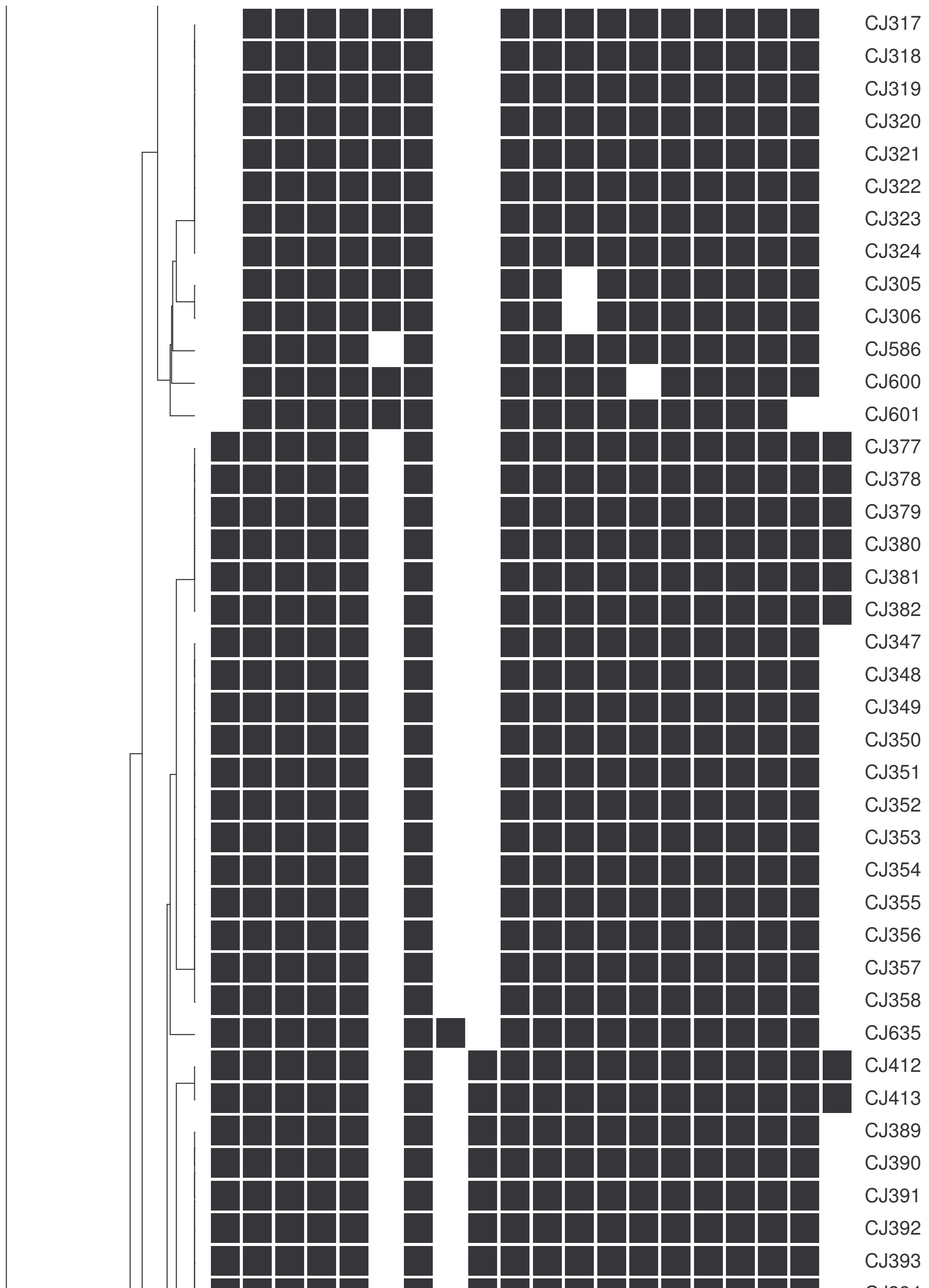

CJ115
CJ116
CJ117
CJ118
CJ119
CJ518
CJ512
CJ557
CJ591
CJ559
CJ560
CJ596
CJ534
CJ540
CJ189
CJ190
CJ191
CJ192
CJ193
CJ194
CJ195
CJ196
CJ197
CJ173
CJ174
CJ175
CJ176
CJ177
CJ178
CJ179
CJ180
CJ181
CJ182
CJ568
CJ567
CJ198
CJ199
CJ200
CJ166
CJ167



CJ168
CJ169
CJ170
CJ171
CJ564
CJ203
CJ523
CJ565
CJ525
CJ201
CJ202
CJ172
CJ566
CJ524
CJ555
CJ520
CJ504
CJ561
CJ509
CJ500
CJ546
CJ468
CJ503
CJ578
CJ544
CJ047
CJ048
CJ497
CJ501
CJ495
CJ533
CJ576
CJ054
CJ055
CJ045
CJ046
CJ610
CJ572
CJ612
CJ613
CJ617


CJ108
CJ611
CJ493
CJ527
CJ023
CJ024
CJ490
CJ569
CJ570
CJ529
CJ571
CJ528
CJ532
CJ107
CJ615
CJ616
CJ573
CJ618
CJ041
CJ042
CJ499
CJ537
CJ531
CJ536
CJ530
CJ502
CJ498
CJ019
CJ020
CJ480
CJ021
CJ022
CJ003
CJ004
CJ005
CJ006
CJ007
CJ008
CJ009
CJ010

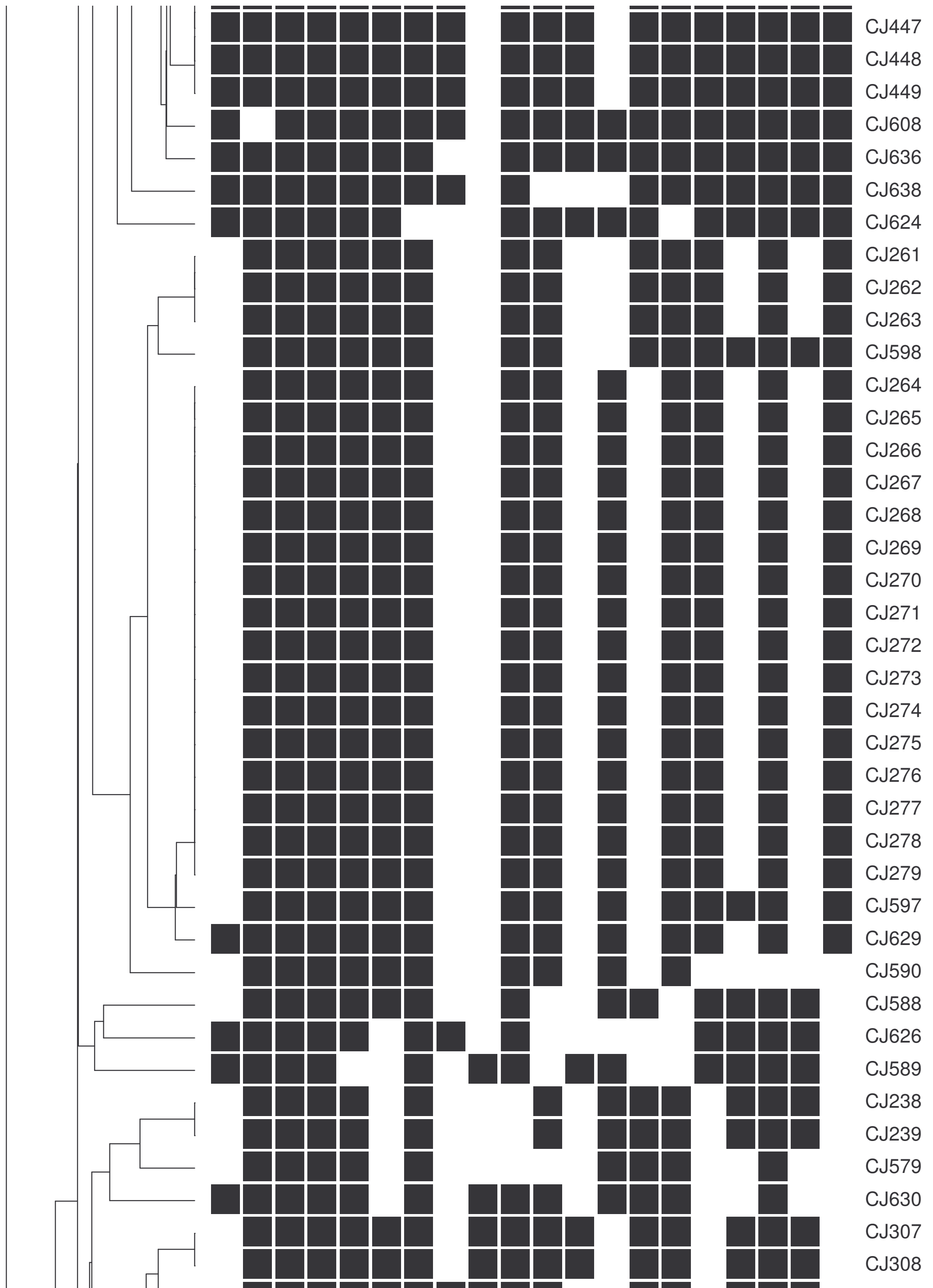

CJ011
CJ012
CJ013
CJ014
CJ015
CJ016
CJ017
CJ018
CJ477
CJ481
CJ463
CJ466
CJ464
CJ476
CJ465
CJ491
CJ526
CJ001
CJ002
CJ492
CJ484
CJ479
CJ489
CJ513
CJ473
CJ482
CJ469
CJ471
CJ475
CJ474
CJ472
CJ410
CJ411
CJ372
CJ373
CJ374
CJ375
CJ376
CJ633


CJ317
CJ318
CJ319
CJ320
CJ321
CJ322
CJ323
CJ324
CJ305
CJ306
CJ586
CJ600
CJ601
CJ377
CJ378
CJ379
CJ380
CJ381
CJ382
CJ347
CJ348
CJ349
CJ350
CJ351
CJ352
CJ353
CJ354
CJ355
CJ356
CJ357
CJ358
CJ635
CJ412
CJ413
CJ389
CJ390
CJ391
CJ392
CJ393

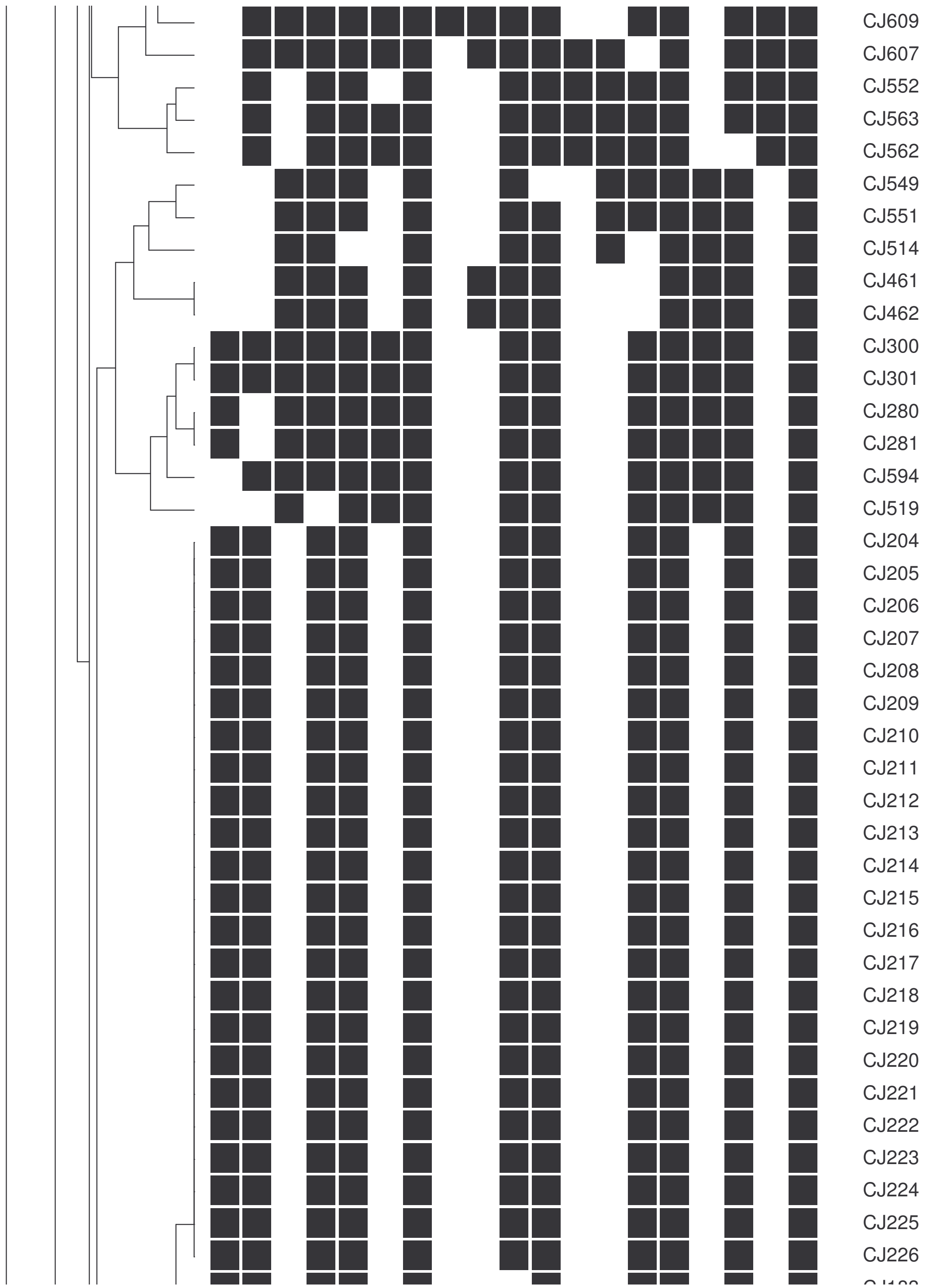
CJ394
CJ395
CJ396
CJ397
CJ398
CJ399
CJ400
CJ401
CJ402
CJ403
CJ404
CJ405
CJ406
CJ407
CJ408
CJ409
CJ632
CJ625
CJ631
CJ302
CJ303
CJ304
CJ599
CJ522
CJ450
CJ451
CJ452
CJ634
CJ339
CJ340
CJ341
CJ342
CJ343
CJ344
CJ345
CJ346
CJ359
CJ360
CJ361
CJ362

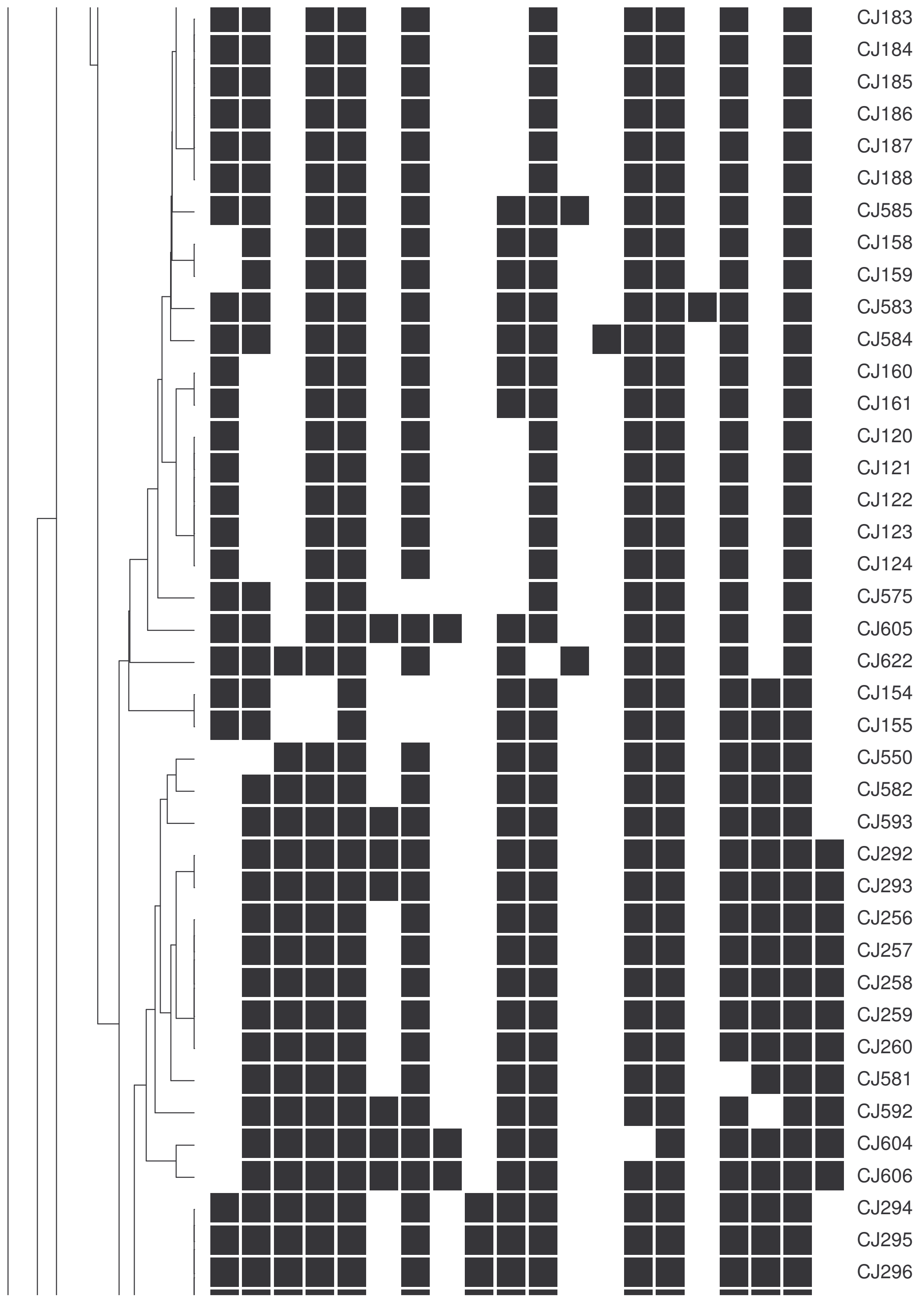
CJ363
CJ364
CJ365
CJ366
CJ367
CJ368
CJ369
CJ370
CJ371
CJ310
CJ311
CJ312
CJ313
CJ314
CJ315
CJ316
CJ640
CJ383
CJ384
CJ325
CJ326
CJ327
CJ328
CJ329
CJ330
CJ331
CJ332
CJ333
CJ334
CJ335
CJ336
CJ337
CJ338
CJ619
CJ414
CJ415
CJ416
CJ641
CJ417
CJ418
CJ419

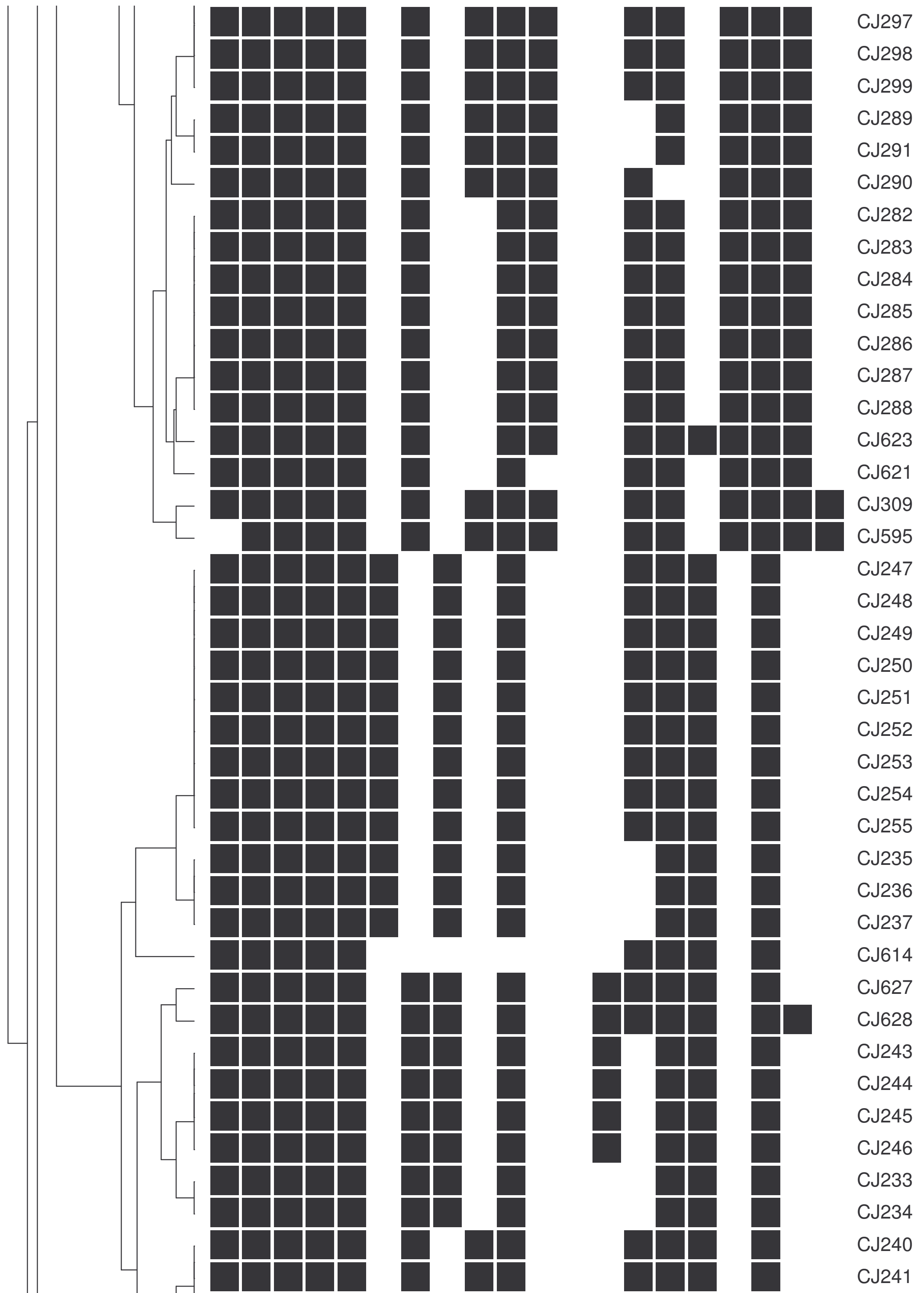
CJ420
CJ421
CJ422
CJ423
CJ424
CJ425
CJ426
CJ427
CJ428
CJ429
CJ430
CJ431
CJ432
CJ433
CJ434
CJ435
CJ436
CJ437
CJ438
CJ439
CJ440
CJ441
CJ442
CJ443
CJ444
CJ445
CJ385
CJ386
CJ387
CJ388
CJ459
CJ460
CJ453
CJ454
CJ455
CJ456
CJ457
CJ458
CJ639
CJ446

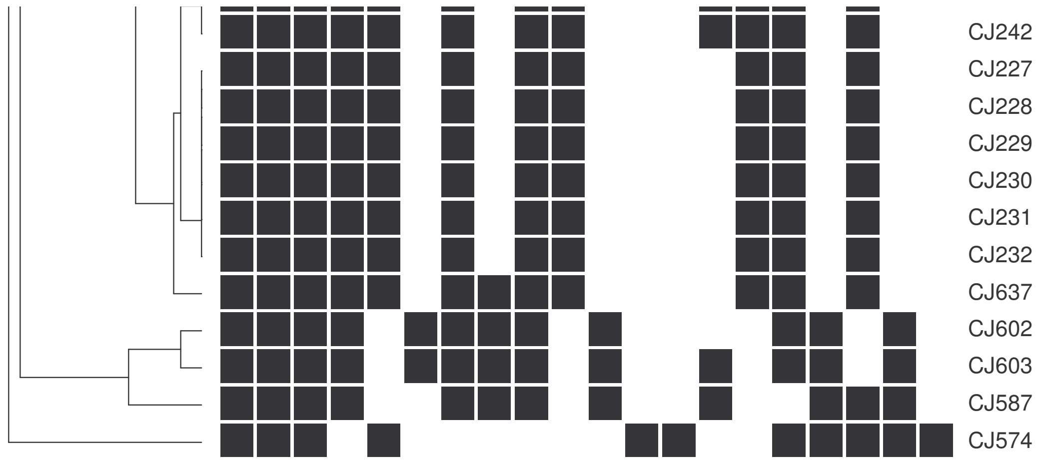


CJ447
CJ448
CJ449
CJ608
CJ636
CJ638
CJ624
CJ261
CJ262
CJ263
CJ598
CJ264
CJ265
CJ266
CJ267
CJ268
CJ269
CJ270
CJ271
CJ272
CJ273
CJ274
CJ275
CJ276
CJ277
CJ278
CJ279
CJ597
CJ629
CJ590
CJ588
CJ626
CJ589
CJ238
CJ239
CJ579
CJ630
CJ307
CJ308


CJ609
CJ607
CJ552
CJ563
CJ562
CJ549
CJ551
CJ514
CJ461
CJ462
CJ300
CJ301
CJ280
CJ281
CJ594
CJ519
CJ204
CJ205
CJ206
CJ207
CJ208
CJ209
CJ210
CJ211
CJ212
CJ213
CJ214
CJ215
CJ216
CJ217
CJ218
CJ219
CJ220
CJ221
CJ222
CJ223
CJ224
CJ225
CJ226


CJ183
CJ184
CJ185
CJ186
CJ187
CJ188
CJ585
CJ158
CJ159
CJ583
CJ584
CJ160
CJ161
CJ120
CJ121
CJ122
CJ123
CJ124
CJ575
CJ605
CJ622
CJ154
CJ155
CJ550
CJ582
CJ593
CJ292
CJ293
CJ256
CJ257
CJ258
CJ259
CJ260
CJ581
CJ592
CJ604
CJ606
CJ294
CJ295
CJ296


CJ297
CJ298
CJ299
CJ289
CJ291
CJ290
CJ282
CJ283
CJ284
CJ285
CJ286
CJ287
CJ288
CJ623
CJ621
CJ309
CJ595
CJ247
CJ248
CJ249
CJ250
CJ251
CJ252
CJ253
CJ254
CJ255
CJ235
CJ236
CJ237
CJ614
CJ627
CJ628
CJ243
CJ244
CJ245
CJ246
CJ233
CJ234
CJ240
CJ241


CJ242
CJ227
CJ228
CJ229
CJ230
CJ231
CJ232
CJ637
CJ602
CJ603
CJ587
CJ574