Evaluation of the Genetic Differences Between Two Subtypes of *Campylobacter Fetus* (*Fetus and Venerealis*) in Canada

By: Lenah Mukhtar

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the degree of Master of Science
in Microbiology and Immunology

Department of Biochemistry, Microbiology and Immunology (BMI)
Faculty of Medicine
University of Ottawa

ABSTRACT

The pathogen *Campylobacter fetus* (CF) is classified into two subspecies, *Campylobacter fetus* subspecies *fetus* (CFF) and *Campylobacter fetus* subspecies v*enerealis* (CFV). Even though CFF and CFV are genetically closely related, they exhibit differences in their host adaptation; CFF inhabits the gastrointestinal tract of both humans and several animal species, while classical CFV is specific to the bovine genital tract and is of particular concern with respect to international bovine trade regulation. Traditionally, differentiation between the two subspecies has been achieved using a limited number of biochemical tests but more rapid and definitive genetic methods of discrimination are desired. A recent study suggested that the presence of a genomic island only in CFV could discriminate between the two sub-species but this hypothesis could not be confirmed on a collection of isolates originating in Canada.

To identify alternative gene targets that would support accurate subspecies discrimination, this study has applied several approaches including suppression subtractive hybridization and whole genome sequencing supplemented with optical mapping. A subtractive hybridization screen, using a well-characterized CFV isolate recovered during routine screening of bulls in an Artificial Insemination center in western Canada and that lacked much of the genomic island and a typical Canadian CFF isolate, yielded 50 clones; characterization of these clones by hybridization screening against selected CF isolates and by nucleotide sequence BLAST analysis identified three potentially CFV-specific clones that contained inserts originating from a second genomic island. Further screening using a larger CF sample set found that only Clone #35 was truly CFV-specific. Optical maps (NcoI digest) of the Canadian CFF and CFV isolates used for the subtractive hybridization showed that certain regions of these genomes were quite distinct from those of two reference strains. Whole genome sequencing of these two isolates identified two target genes (PICFV5_ORF548 and CFF_Feature #3) that appear to be selectively retained in the two subspecies. Screening of a collection of CF isolates by PCRs targeting these three loci (SSH_Clone #35, PICFV5_ORF548 and CFF_Feature #3) supported their use for subspecies discrimination. This work demonstrates the complex genomic diversity associated with these CF subtypes and the challenge posed by their discrimination using limited genetic loci.

ACKNOWLEDGMENTS

At first, I'd like to thank Allah, the all knowing, for giving me the strength to carry on. I would also like to dedicate my hard work on this project to the handful of people that helped make it possible.

Dr. Susan Nadin-Davis, my research supervisor, has played an important role in the success of this project and my thesis paper by providing a laboratory environment that is both educational and stimulating and by giving me valuable and constructive feedback, which have collectively helped me attain my goals. Thank you Dr. Susan, it has been such a pleasure being a part of this laboratory.

Dr. Craig Lee, thank you for being a member of the thesis advisory committee and for all the support given to me during this project. Dr. Catherine Carrillo, thank you for your input as a thesis advisory committee member. Your knowledge about bacterial genomics added a lot of value to this project, especially the whole genome sequencing data and the genome assembly. Dr. Kingsley Amoako, thank you for providing the optical mapping data, which I've analyzed with help from Dr. Dele Ogunremi, thank you both for your assistance.

I'm also extending my gratitude to the molecular genetics lab members at the Canadian Food Inspection Agency who I've had the pleasure to work with during my project. Mary Sheen, Adam Colville, and Louise Pope, thank you all for the great support and especially for the *de novo* genome assembly and analysis. Dr. John Devenish and Teresa Burke, thank you for providing the bacterial cultures and the biochemical analyses. I wouldn't be able to complete this research without the help of my colleagues at the CFIA.

I'd also like to thank my sponsor, The Ministry of Higher Education at Saudi Arabia, represented by The Royal Embassy and the Saudi Arabian Cultural Bureau in Canada, for the full scholarship and for providing all services needed to ease pursuing my studies abroad. Last but not least, I owe my loving thanks to my parents, my husband, sister, brothers and friends for their encouragement and support, which made it possible for me to finish this work.

DEDICATIONS

I dedicate this work to my lovely parents, Faimah and Emam, and my wonderful husband, Zeyad, Thank you all for your prayers and support.

Table of Contents

ABBREVIATIONS

Acrynom	Meaning
%	Percent
°C	Degrees Celsius
CT	Cycle Threshold
A	Absorbance
A260	Absorbance at 260 Nm
A280	Absorbance at 280 Nm
AB	Applied Biosystems
ADRI	Animal Disease Reseach Institute
AFLP	Amplified Fragment Length Polymorphism
AHML	Animal Health Microbiology Laboratory
AI	Artificial Insemination
AMP	Ampicillin
AP	Alkaline Phosphatase
ATCC	American Type Culture Collection
BC	British Colombia
BDT	Big Dye Terminator
BGC	Bovine Genital Campylobacteriosis
BLAST	Basic Local Alignment Search Tool
Bp	Base Pair
CRISPR	Clustered Regularly Interspaced Short Palindrome Repeats
CDT	Cytolethal distending toxin
CF	Campylobacter Fetus
CFF	Campylobacter Fetus subspecies Fetus
CFIA	Canadian Food Inspection Agency
CFV	Campylobacter Fetus subspecies Venerealis
CFVi	Campylobacter Fetus subspecies Venerealis Biotype Intermedius
ddH20	Double Distilled Water

Acrynom	Meaning
DEPC	Diethyl Pyrocarbonate
DIG	Digoxigenin
DNA	Deoxyribose Nucleic Acid
dNTPs	Deoxyribose Nucleoside 5’ Triphosphates
DTT	Dithiothreitol
dUTP	Deoxyuridine Triphosphate
E. coli	Escherichia Coli
EDTA	Ethylenediaminetetraacetic Acid
ELISA	Enzyme-Linked Immunosorbent Assay
EtBr	Ethidium Bromide
EtOH	Ethanol
FASTA	A text based format for storing nucleotide or peptide sequences
FASTQ	A text based format for storing usually nucleotide sequence and its corresponding quality
FAT	Fluorescent Antibody Test
g	Gram
gDNA	Genomic DNA
GI	Genomic Island
H_2S	Hydrogen Sulphide
HGT	Horizontal gene transfer
hr	Hour (S)
IDT	Integrated DNA Technology
IPTG	Isopropyl B-D-Thiogalacto Pyranoside
Int	Integrases
IS	Insertion Sequence element
KB	Kilo Base
L	Liter
LB	Luria-Bertani Media

Acrynom	Meaning
LPS	Lipopolysaccharide
M	Molar
MAUVE	Multiple Alignment Of Conserved Genomic Sequence With Rearrangements (software for efficient construction of multiple genome alignments)
mg	Milligram
$MgCl_2$	Magnesium Chloride
min	Minute(s)
ml	Millilitres
MOMP	Major outer membrane protein
MW	Molecular Weight Marker
N	Normal
NaCl	Sodium Chloride
NCBI	National Center For Biotechnology Information
NCTC	National Collection Of Type Cultures
NDK	Nucleoside diphosphate kinase
ng	Nanogram
NGS	Next Generation Sequencing
NH4OAc	Ammonium Acetate
nm	Nanometre
OD	Optical Density
OIE	Office International Des Epizooties (The World Organization For Animal Health)
OLF	Ottawa Laboratory Fallowfield
ORF	Open Reading Frame
oriT	Origin Of Transfer
PCR	Polymerase Chain Reaction
PFGE	Pulsed Field Gel Electrophoresis
PI	Pathogenicity Island
PIPS	Pathogenicity Island Prediction Software

Acrynom	Meaning
RAPD	Randomly Amplified Polymorphic DNA
RAST	Rapid Annotation Using Subsystem Technology
rpm	Rounds Per Minute
rRNA	Ribosomal RNA
RT	Room Temperature
RT-PCR	Real-Time Polymerase Chain Reaction
TE	Tris-EDTA Buffer
TEM	Transport Enrichment Medium
TMAO	Trimethylamine N-Oxide
TSI	Triple Sugar Iron
TTC	Triphenyl Tetrazolium Chloride
U	Unit
UTI	Urinary Tract Infection
UV	Ultraviolet
V	Volts
VF	Vibrio Fetus
v/v	Volume Per Volume
VMAT	Vaginal Mucus Agglutination Test
w/v	Weight Per Volume
WHO	World Health Organization
X-GAL	5-Bromo-4-Chloro-3-Indolyl- Β-D-Galactopyranoside
xg	Relative Centrifugal Force
µg	Microgram
µL	Microliter
µM	Micromolar
µm	Micrometer

LIST OF FIGURES

LIST OF TABLES

Chapter One: Introduction

1.1 Background

Campylobacter fetus (CF) is an important pathogen that has been recently recognized in veterinary medicine as comprising two subspecies, *Campylobacter fetus* subspecies *fetus* (CFF) and *Campylobacter fetus* subspecies *venerealis* (CFV) [25,174]. CFF is known to inhabit the intestinal tract of cattle, sheep and other animals, causing sporadic abortions; while CFV is strictly found in the genital area of cattle and is the primary cause of bovine genital campylobacteriosis (BGC) (also referred to as "bovine venereal campylobacteriosis"), which is a notifiable disease of the world organization for animal health (OIE) [25,118,174]. BGC causes infertility and conception failure; thus, animals and animal products destined for international trade must be certified as free from CFV [25,117,150]. *Campylobacter fetus* subspecies *venerealis biovar intermedius* (CFVi) is regarded as a sub-type of CFV despite having some phenotypic features typical of CFFs (i.e. hydrogen sulphide (H_2S) positive and tolerate high concentration of glycine) [138].

Usually the detection of CF is achieved through collecting bulls' preputial scrapings into normal saline, transporting the collected materials to the laboratory in transport enrichment medium (TEM), growing the microorganism microaerophilically in a selective culture that contains polymyxin B, and examining the isolated colonies by macroscopic and microscopic means [25,77,118].

Subspecies differentiation of CF is primarily based on the growth of the microorganism on a medium that contains 1% glycine, where CFF can be grown on this medium while CFV cannot [25]. This glycine may act as an antibiotic reducing the cell wall integrity of various bacteria [95], but the actual reason behind this different metabolic capability between CFF and CFV is currently unknown. In order to assure the credibility of this assay the laboratory must pay strict attention to the details of this assay.

In order to assure the credibility of this assay the laboratory must pay strict attention to the details of this assay. Direct immunofluorescence tests, enzyme linked immunoassays (ELISA), pulsed-field gel electrophoresis (PFGE), multilocus sequence typing (MLST), amplified fragment length polymorphism (AFLP), and polymerase chain reaction (PCR) are alternative approaches which have been used for CF detection and subspecies discrimination [25,36,78,79,97,100,120,138,166,168,169,178]. However, most of these tests are time-consuming and too laborious for routine application in diagnostic laboratories and some lack subspecies specificity. Presently, methods based on PCR are the most promising robust CFV detection methods for field samples, and many CFV-specific PCR protocols have been reported [25].

In Canada, CFV-positive cases are relatively rare [111]. Due to the highly prized genetics of Canadian cattle, a multimillion-dollar bovine germplasm industry has evolved to support breeding via artificial insemination (AI) using semen collected from prize bulls. The presence of either CFF or CFV in Artificial Insemination (AI) stations can have major economic consequences. The OIE has listed CFV, not CFF, as a class B pathogen (second highest priority) because of its importance in international trade. As such, bulls that donate semen to registered semen repositories must be free of CFV infection. In order to trade internationally, Canada must continue to prove its bovine germplasm industry CFV-free through routine testing. At the beginning of October 2008, an outbreak of CFV infection occurred in an AI station in Alberta. The outbreak involved 33 of the 100 bulls in the AI station. Two months prior to this outbreak, CFV was isolated from a bull in another AI station belonging to the same owner but which had originated form the United States; subsequently, CFV was also detected in a second animal at that facility.

Several of these CFV isolates were tested for the presence of the genomic island (GI) “PICFV8” which had previously been identified by Gorkiewicz et al as indicative of this subspecies [59]. However, from the five loci within the genomic island that were tested (i.e. GI 1-5 PCRs) only one locus (i.e. GI 5) was consistently present in these specimens (unpublished data, Dr. Nadin-Davis). Moreover, this laboratory has shown that the GI 5 PCR targeting the *ven* sequence was not always a reliable marker as sometimes CFF samples gave a positive result. This was subsequently shown to be due to the presence of a closely related pseudo-gene in these CFF isolates. A recent study in Canada that used the *ven sequence* as a CFV-specific target identified heterogeneity at this locus but did not do any biochemical or phenotypic characterization of their samples to clearly establish the subspecies nature of the isolates recovered; hence, the interpretation of their findings is questionable [25]. Accordingly, additional PCRs capable of discriminating CFF/CFV were needed to allow reliable differentiation between these two subspecies. In view of the lack of much of the GI in the CFV samples from the 2008 outbreak, one of these samples was selected as the tester in a Suppression Subtractive Hybridization (SSH) study. Therefore, this study focuses on the detection of the genetic differences between the CFF/CFV subspecies that could aid in the development of rapid diagnostic tools and might also have some relevance to the distinctive pathogenesis of these two CF subspecies. Other isolates, from Argentina, Australia, Sweden, USA, and UK, were used in parallel with the Canadian isolates for validation purposes.

1.2 Literature Review

1.2.1 The General Morphology and Physiology of the Genus *Campylobacter*

Members of the genus *Campylobacter* that belong to the Campylobacteraceae family are spiral, curved, rod-shaped, non-spore forming, gram-negative bacteria (Figure 1A) [114]. Generally, they are 0.5 – 8.0 µm long and 0.2 – 0.5 µm wide, motile with either a single flagellum (monotrichous) or two or more unsheathed flagella (amphitrichous) characterized by a darting, corkscrew-like motion under the phase-contrast microscope (Figure1B) [82,126]. *Campylobacter* species, a part of epsilon-proteobacteria, are microaerophilic, and highly adapted to mucosal host surfaces [73]. Currently, 17 species and 6 subspecies assigned to the genus *Campylobacter* are recognized by NCBI and WHO [114,180].

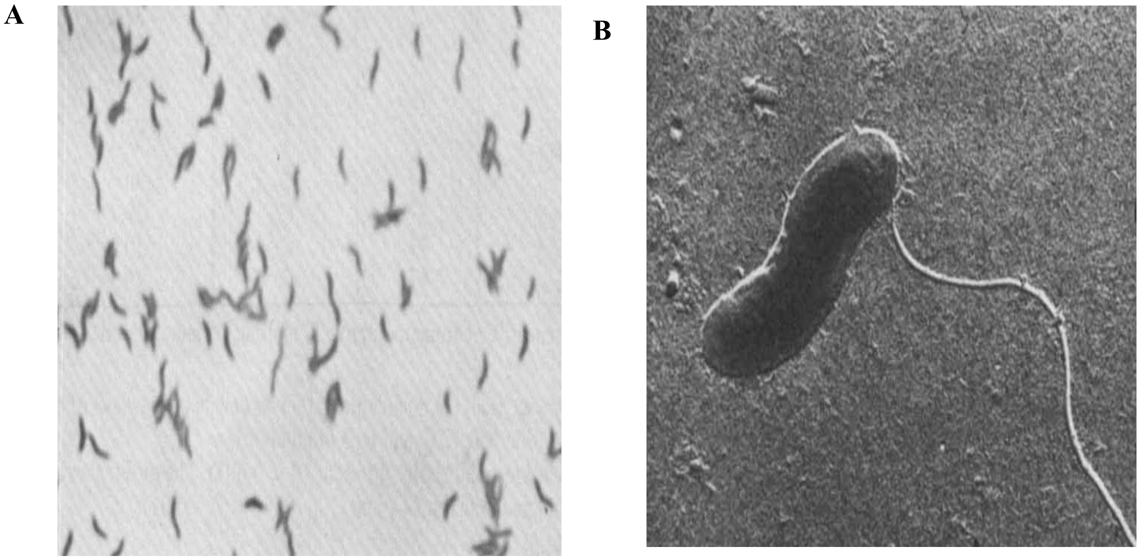


Figure 1. The morphology of the genus *Campylobacter*. (A) under Gram straining, (B) Electron micrograph of bacterium shape and corkscrew motility. Figure adapted from Journal of Bacteriology (Véron M, Chatelain R. (1973) Taxonomic study of the genus *Campylobacter* sebald and véron and designation of the phenotype strain for the type species, *Campylobacter fetus* (smith and taylor) sebald and véron. Int J Syst Bacteriol 23: 122-134). © with permission from ASM.

1.2.2 Complete bacterial genome sequence in C. fetus

The first published complete genome sequence of *Campylobacter* species is the genome of *C. jejuni* NCTC 11168 in 2000 [124]. It was 1.6 Mbp nucleotides in length and 94.3% of the genome codes for proteins, making *C. jejuni* the densest bacterial genome sequenced at that time [89,124]. Since then several complete and unfinished genome sequences have been added to the public databases [53,59,89,124,125,128,155].

Most *Campylobacter* species have a small circular genome with low GC content (30%) of approximately 1.6 - 1.8 Mbp, which is one-third the size of the *Escherichia coli* genome [26,158]; the *C. upsaliensis* genome, which is 2 Mbp, is an exception [88]. This small genome size reflects *Campylobacter*'s requirement for complex media for their growth as well as their inability to ferment or degrade complex sugars [60]. In addition, extra-chromosomal elements, conjugative plasmids and bacteriophage have been reported in *Campylobacter* species [158].

Similar to other *Campylobacter* genomes, CF has a small genome (1.8 Mbp); however, genome sizing is variable between CFF & CFV based on the pulsed field gel electrophoresis [138].

CFF 82-40 was completely sequenced in 2006; its circular DNA is 1.77 Mbp in length, has 33% G+C content, and it contains 1820 genes which comprise 90% of the genome (GenBank Accession number NC_008599) [3,155]. Draft genome sequences of CFV 84-112 (EBI Project ID: 42511) and CFV NCTC 10354 (GenBank Accession number CM001228 and AFGH01000000) [155] have recently become available. NCTC 10354 has a slightly larger genome (1.87 Mbp) and includes 1905 putative protein coding genes, 3 rRNA operons, 35 tRNA operons, and 185 pseudogenes [155]. The availability of complete genome sequences provides accessible resources for detailed analysis of *C. fetus* physiology, subspecies specific adaptation, and will stimulate efforts to identify mechanisms contributing to host-pathogen interactions and virulence [89,155].

1.2.3 Historical Classification of the species *Campylobacter fetus*

CF (formerly *Vibrio fetus* "VF") are serious veterinary and human pathogens, recognized as the main causative agents of abortions in cattle and sheep as documented in 1913 and cited in Tracy Schmidt (2008) [98,164].

A 1959 report divided CF into two subspecies. Florent observed a venereally transmitted enzootic infertility in cattle caused by one variant strain originating in the prepuce of the asymptomatic bull ("*Vibrio fetus*" *var. venerealis*), while sporadic abortion in cattle was caused by another variant of intestinal origin ("*Vibrio fetus*" *var. intestinalis*) [5,52]. Florent further described a few biochemical tests to differentiate between these two variants.

In 1973, Vèron and Chatelain published a taxonomic study and reclassified both VF subspecies into a new genus, *Campylobacter*. Thus, VF *intestinalis* became CFF and VF *venerealis* became CFV [174]. Accordingly, on the basis of the taxonomic divisions, clinical and biochemical differences, the *Campylobacter fetus* was split into two subspecies, CFF and CFV[168]. A summary of CF subspecies host preference and clinical importance is shown below in Table 1.

Table 1. An overview of the host preference niche and clinical pathology of *Campylobacter fetus* subspecies. Table adapted after modification from (Tracy Schmidt. (2008) Detection of *Campylobacter fetus* in bovine preputial scrapings using PCR and culture assays (Unpublished master dissertation). University of Pretoria, South Africa).

Subspecies	Targeted hosts(s)	Disease
Campylobacter fetus venerealis	Cattle	Bovine genital campylobacteriosis: infertility, early embryonic death and occasional abortion
Campylobacter fetus fetus	Cattle/Sheep	Gastroenteritis/ Sporadic abortion
	Humans	Sporadic infections mainly in immune-compromised patients

1.2.4 Virulence factors and determinants of *Campylobacter*

Bacterial virulence factors are secreted by many pathogens and have vital roles during bacterial pathogenesis. They help microorganisms to attach, enter and invade host cells. Also they interfere with host immunity or cell signaling and therefore help the bacteria to escape the host's defense mechanisms [31]. Some of these factors are involved in motility, toxin and protein production, cell wall and surface layer protein generation and the secretion system.

1.2.4.1 Cell wall and Surface layer proteins. The cell wall of *Campylobacte*r is composed of three layers as seen under cross section view: an outer lipoprotein, a middle lipopolysaccharide, and an inner mucopeptide. Lipopolysaccharide (LPS) is the major virulence factor for most enteric gram-negative pathogens [88]. It is composed of three distinct regions: (i) an inner lipid A membrane layer with endotoxic properties that contributes to serum resistance and resistance to the toxicity and killing activities of host phagocytic cells; (ii) A middle core oligosaccharide layer attached to the lipid A; and (iii) O antigen attached to the outer core that has a vital role in antigenic variation and bacterial adherence [8,88,114,130].

Most pathogenic gram-positive and gram-negative bacteria as well as Archaea have mechanisms for evading host immune systems, including destruction of the immune system, expression of anti-phagocytic polysaccharide protein capsules, regulation of cytokines, or antigenic variation of cell surface components targeted by the immune system [50,160]. One of these evasion mechanisms is manifested by surface-layer (S-Layer) proteins [160]. This S-layer is located externally in the outer membrane of *C. fetus* and *C. rectus* and can have multiple roles in immunity resistance[160,166]. Essentially, all *Campylobacter fetus* bacteria*,* in ungulates and immunocompromised humans, possess multiple homologues of the *sapA* gene, which encodes a 97 KDa crystalline capsule-like surface array (S-layer) proteins (SLPs) [11,13,61,160]. Examination of CF strains in *vitro* and in *vivo* shows that this S-layer is involved in two mechanisms of immune avoidance: (i) SLP confers complement resistance by preventing the binding of C3b complement factor to the cell surface; (ii) SLP protects against antibody-mediated clearance by high-frequency antigenic variation of SLPs and DNA rearrangements on the chromosome of SLP-encoding genes[160]. These two mechanisms are the main characteristics that permit the expression of heterogeneous *sap* and therefore facilitate persistent mucosal colonization and survival of *Campylobacter* under various environmental changes [41,45,160,161,177].

1.2.4.2 Secretion system: Bacterial virulence factors are often acquired through horizontal gene transfer (HGT) followed by point mutations, chromosome insertions, deletions and rearrangements that result in significant genome plasticity. Such genomic regions are referred to as genomic islands (GI) or pathogenicity islands (PI) and their presence in a bacterium can convert a non-pathogenic bacterial cell to a pathogenic one of the same species [116]. In general, a GI is divided into core genes that comprise chromosomal genes encoding essential cellular functions, and a flexible gene pool, which harbors mobile genetic elements such as plasmids, bacteriophages, integrases, transposons and insertion sequence elements (IS) [62,116]. These genetic elements are often flanked by tRNA, direct repeats and IS elements that stimulate integration and excision out of the bacterial genome through conjugation, transformation, or phage infection (Figure 2) [63,71,116]. PIs carry genes related to protein secretion systems, invasion systems, iron uptake systems, toxins, colonization and adhesion, and bacterial adaptation to different environmental conditions [116,144,152].

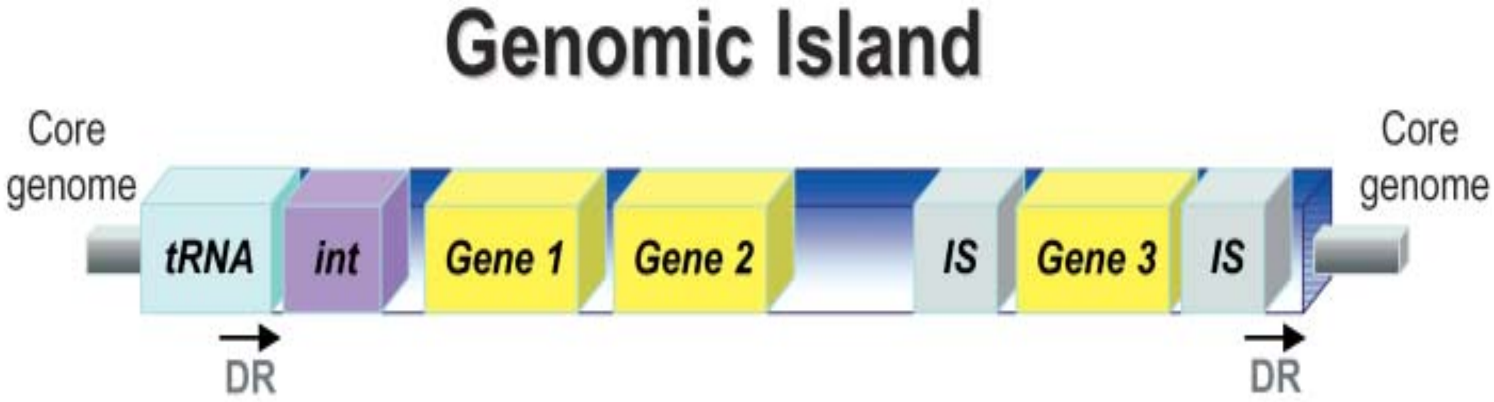


Figure 2. A linear map showing general characteristics of a model genomic island. Such a GI is normally located within a core chromosome of a pathogenic bacterium and contains discrete genetic units with various functions. The GI is often integrated into tRNA genes and flanked by direct repeat elements. Additional mobile genetic features such as integrases (*int*) and IS-elements (*IS*) are usually present. Figure adapted from APMIS (Oelschlaeger TA, Hacker J. (2004) Impact of pathogenicity islands in bacterial diagnostics, APMIS 112: 930-936). © with permission from John Wiley and Sons.

1.2.4.2.1 General features of Bacterial Type 4 Secretion System. Many bacterial pathogenic factors interact with the eukaryotic host cell either through surface presentation or by absorption into the host cell after secretion from the bacterial cell. Five distinct protein secretion systems, known as types I through V, have been documented (see Schmidt and Hensel 2004 for an overview) [142]. The Type IV extracellular secretion system (T4SS), which is present in all bacteria and some archaea, is the one most closely associated with PIs [142]. The T4SS has two fundamental functions; (i) exchange of genetic material, and (ii) delivery of effector molecules to eukaryotic cells [23,29,176]. This large protein complex, powered by cytoplasmic ATPases, contains a channel used mainly for transport of macromolecules such as DNA and proteins across the cell envelope [134,176]. T4SSs can be divided into three functional subfamilies, each having a distinct means of contributing to bacterial pathogenesis [4,23]. Details of the T4SS subfamilies are reviewed in Schmidt and Hensel (2004).

1.2.5 *C. fetus* as a model for studying pathogenicity

In an effort to better understand what genetic features discriminate the two CF subspecies, a subtractive hybridization approach identified a genomic island present in one CFV isolate (strain ATCC 19438); the sequence of this feature has been deposited in GenBank (Accession number EU443150) and its relationship to the pathogenicity of this subspecies has been further explored [59]. This genomic island was inserted into the core chromosome following a tRNA gene and contains a number of mobile genetic elements such as transposases and phage integrase with a different G+C content to the rest of the genome [59]. As illustrated in Figure 3, the sequences distal to ORF6 show strong homology to the VirB2-11 and VirD4 subunit proteins of the T4SS of *A. tumefaciens* [29].

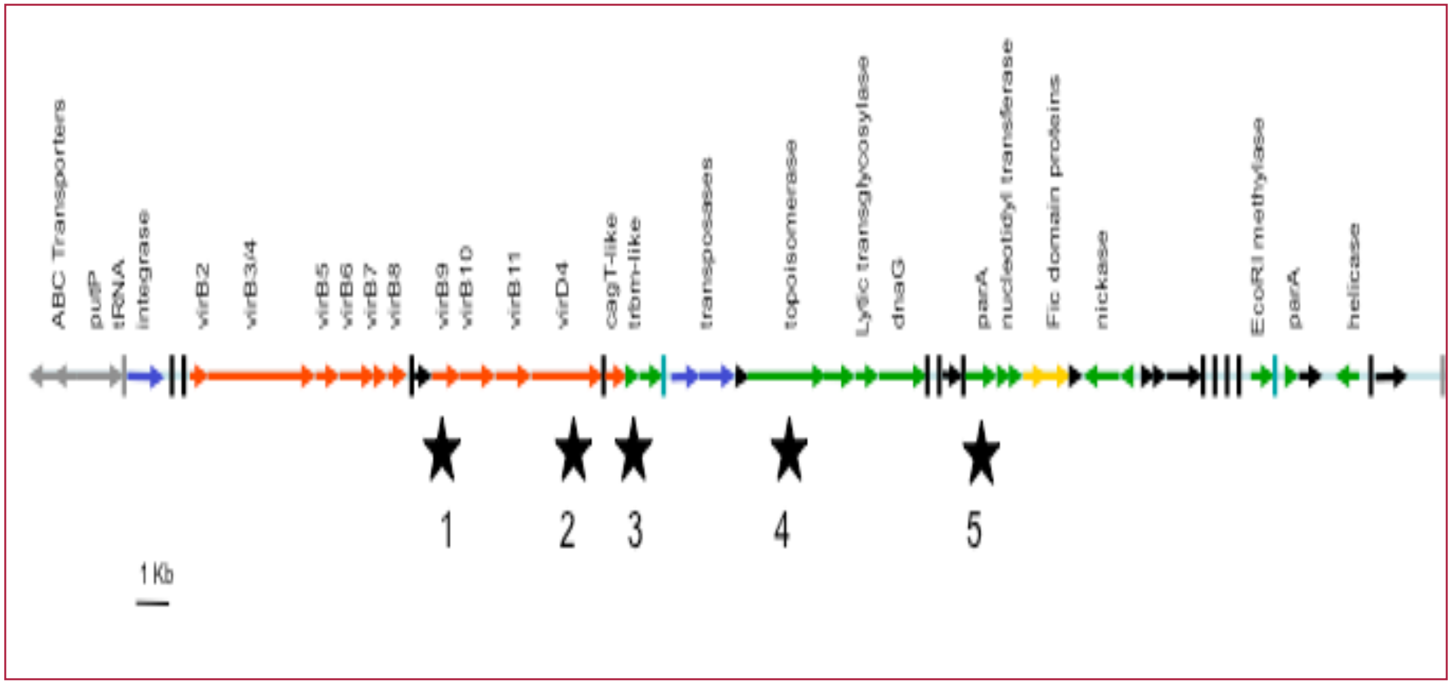


Figure 3. Map of Genomic Island identified in CFV isolate ATCC 19438. The island is inserted into a methionyl-tRNA gene. Functions of many genes were inferred based on the predicted sequences of the proteins encoded by open reading frames. Functional groups are indicated with color thus: mobility genes, such as transposases and integrases (blue), a type IV secretion system (red), putative effector molecules (yellow), genes of apparent plasmid origin (green), and genes of unknown function (black). Chromosomal core sequences are shown in (grey). The scale bar represents 1 kb. Figure adapted from Journal of Bacteriology (Gorkiewicz G, Kienesberger S, Schober C, Scheicher SR, Gully C, et al. (2010) A genomic island defines subspecies-specific virulence features of the host-adapted pathogen *Campylobacter fetus* subsp. *venerealis*. J Bacteriol 192: 502-517). © with permission from ASM.

★ Represents the locations of the five ‘GI-specific’ sequences targeted by PCR for screening of CF isolates.

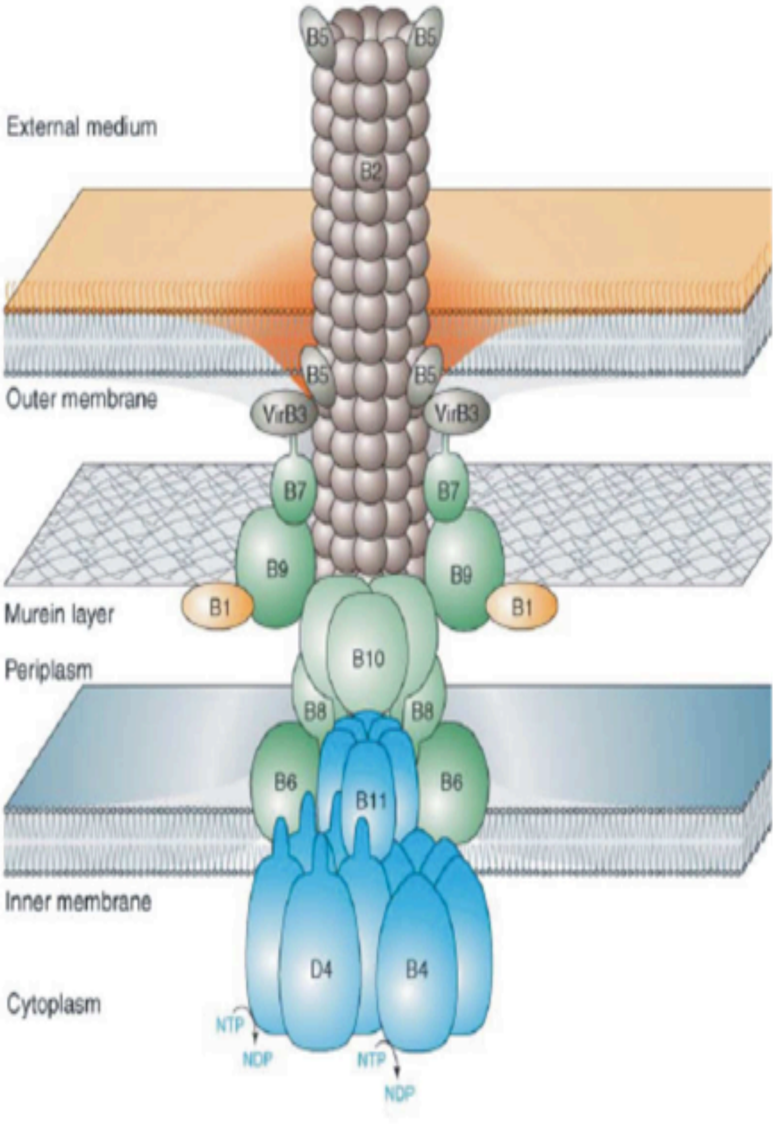


Figure 4. Model of assembled prototypical *A. tumefaciens* VirB/D4 T4SS machinery. It contains the coupling protein VirD4 and 11 structural mating-pore-formation components (VirB1 – VirB11). This multi-structural complex of proteins which spans the inner and outer membranes of gram-negative pathogens is required for substrate secretion; it transfers macromolecules either to the extracellular milieu or directly from bacterial cytoplasm into the cytoplasm of the infected host cell. Figure Adapted from current Opinion in Microbiology (Steffen Backert,Thomas F Meyer, (2006) Type IV secretion systems and their effectors in bacterial pathogenesis, Current Opinion in Microbiology 9:207–217). © with permission from Elsevier.

A very recent study in 2012 compared the whole genomes of both subspecies CFF_82-40 and CFV_NCTC10354 (GenBank Accession number NC_008599 and AFGH00000000, respectively) to get a better understanding of the virulence factors in their accessory genomes [3]. Using the pathogenicity island prediction software (PIPS) to predict the pathogenicity islands (PIs) in both strains, the authors identified 12 pathogenicity islands in NCTC10354 (PICFV) and 10 in CFF82-40 (PICFF) [3,152]. Three specific islands were of particular interest due to big deletion events between the two isolates [3]. The first island is PICFV5 that was partially deleted in the CFF, although it contains a large portion of surface array proteins (*sap*) conserved in both organisms [3]. These *sap* genes code for the surface layer proteins, which attach to the host's cell surface and disrupt immune defenses; therefore, they are good candidates for vaccine development [3]. The second island is PICFV8 that was completely absent in the 45 CFFs and found in 51 out of 67 CFVs as reported by Gorkiewicz et al. in 2010 [3,59]. This 33.4 kb island is flanked by methionyl-tRNA and *putP* genes, and encodes a conjugation-related type IV secretion system (T4SS), which is involved in the transportation of bacterial DNA and proteins into the target host [23,59,89,176]. The third island is PICFF6, which is partially deleted in the CFV, and contains Clustered Regularly Interspaced Short Palindrome Repeats (CRISPR) and a *dam* gene [3]. CRISPR elements form a "group of ~25–50 nucleotides of short direct repeats interspaced by unique sequences of similar size"; they are involved in prokaryotic immune defense against foreign DNA replication [3]. The *dam* gene codes for Dam enzyme that protects bacteria from alien DNA [3].

1.2.6 Campylobacter fetus Pathogenicity

1.2.6.1 *Campylobacter fetus fetus*. CFF is mainly found in the intestine of sheep and cattle; it can also be present in the genital tract of sheep and cattle, the placenta and stomachs of aborted animals, and it is a major cause of sporadic abortion in ruminants worldwide. Birds can also be infected as secondary reservoirs [6,56,126]. CFF is transmitted orally through the ingestion of food or water contaminated with bacterial content from feces thus causing a transient bacteremia, septicemia, and invasion of the uterus; due to their high affinity for the placental tissue this initiates abortion in pregnant animals. Generally, the infected fetus is aborted, but if live birth occurs, the newborn animal may survive 4 to 5 days [33,40,126].

Human infection with CFF is rare; it is mostly opportunistic and particularly affects immune-deficient or compromised and predisposed patients [19,54]. Infection is mainly acquired through drinking unpasteurized milk or ingestion of cottage cheese, raw beef or undercooked food leading to septicemia, meningitis, pericarditis, peritonitis, salpingitis and septic arthritis [80,105,126,136]. Further, it was shown that some cancer patients undergoing nutritional therapy could gain enteritis through consumption of raw beef liver contaminated with CFF [126].

1.2.6.2 *Campylobacter fetus venerealis*. CFV, in contrast to CFF*,* is highly adapted to the preputial genital area at the end of the penis between the penile gland and the fornix of the prepuce of the bull [46,126]. It is the major cause of Bovine Genital Campylobacteriosis (BGC), a sexually transmitted disease of major concern to the cattle industry that can lead to severe economic loss in endemic regions [118].

Since CFV is mainly present in the prepuce of the bull, it is transmitted to the cow during mating, causing chronic inflammation in the reproduction system and leading to infertility and abortion [70,139]. Infection of the bull causes no apparent symptoms or histological changes, so usually the BGC infection is unsuspected and subclinical until a decrease in calving rates is observed [12,30].

In cows and heifers, CFV mainly persists in the anterior vagina and the cervix. Usually, the organism invades the uterus causing severe vaginal discharge as a result of moderate endometritis, cervicitis, vaginitis and salpingitis [164,172]. However, the infection is less frequent due to the heavy presence of IgG antibodies dominating the cervix and the uterus [32,164]. CFV infection is usually self-limiting; the cow regains its fertility within 5 months following the elimination of the infection [69,164]; treatment with an antibiotic cocktail can help in elimination of the infection.

1.2.7 Epidemiology and Economic Significance of BGC

Bovine Genital Campylobacteriosis (BGC) is a worldwide disease reported by many countries including North America [77,109,110]. It can cause huge reproductive losses in the infected beef and dairy herds, due to decreased calving percentages or delayed calving, which represent large financial loss for producers [75,76,82].

In industrialized countries, the incidence of the BGC disease has declined due to the global use of artificial insemination for cattle[164]. However, it is still causing problems due to the use of contaminated semen collection equipment or poorly treated semen, or during mounting when large numbers of animals are enclosed together [30].

Campylobacteriosis is currently listed as “List B” according to OIE as a result of its socioeconomic and public health implications [105]. Subsequently, it influences the international trade of animals and animal products[100]. Thus, having herds free from CFV infection is an important factor for many countries for the import and export of bovine semen, and cattle, as well as for health certification of bulls in semen production and distribution centers [5]. For this reason, the availability of optimal, rapid, and reliable methods for the detection and the differentiation between CF subspecies is a priority [5].

1.2.8 Diagnosis of BGC

1.2.8.1 Sample collection. For the routine diagnosis of bovine genital campylobacteriosis, samples are collected from bulls, cows, or rarely from aborted fetuses. In bulls, smegma usually obtained by scraping, suction or by preputial washing; in cows, cervicovaginal mucus obtained by swabbing or vaginal washing; while foetal placenta or stomach contents might be taken from aborted fetus. However, the chance of recovering CFVs from testing preputial washes of bulls is greater than testing vaginal samples from cows [82,118,164]. The collected sample is preserved in a transport enrichment medium (TEM) and sent directly to the laboratory for culture [118,164].

1.2.8.2 Bacterial culture, and isolation. Although culture and morphological identification of CF is considered to be the golden standard for diagnosing campylobacteriosis, the organism is fragile and fastidious and specific nutritional and atmospheric growth conditions are required; moreover, overgrowth by other contaminants such as *Pseudomonas* subspecies and *Proteus* subspecies often make it difficult to detect CF subspecies [17,46,164]. It is recommended to plate out the samples onto two agar media; one selective, such as Skirrow's or Muller Hinton agar, and a non-selective agar but both agars must contain polymyxin B antibiotic to improve selectivity and recovery of CF [77,85,117,118,164]. Following incubation at 37°C under microaerophilic conditions (5% O_2, 10% CO_2, and 85% N_2), and after 2 to 5 days, CF becomes visible on the agar as pink, smooth, round, and shiny colonies with regular edges [118,164].

1.2.8.3 Biochemical and Immunological identification of *C. fetus*. Identification of *Campylobacter* species can be problematic since they react with few limited biochemical tests due to their inability to ferment or oxidize the common carbohydrate sugars utilized for routine microbial diagnosis; for example, they cannot phosphorylate or transport glucose so the energy required for their growth is obtained from the tricarboxylic acid cycle [56]. In addition, API Campy, a commercial kit that contains 21 various tests for proper identification of *Campylobacter*, has also been described; however, the efficacy and the integrity of the kit is questionable in light of misidentification of some *Campylobacter* species [126, 159].

In 1991, On and Holmes studied the influence of inoculum size on different biochemical tests and they defined a density of 10^6 CFU/ml of CF as being the optimal inoculum size required to ensure the reproducibility of test results [119]. Typically, the phenotypic characterization of CF isolates involves: (i) examining for growth at 25°C, 37°C, and 42°C; (ii) atmospheric growth needs (aerobic, anaerobic, or microaerophilic); (iii) growth on media containing glycine (ranging from 0.60% - 1.90%); (iv) growth in 1.0% Bile, 3.5% NaCl, and 0.04% TTC - triphenyl tetrazolium chloride; (v) reaction to catalase and oxidase; (vi) reduction of nitrate and selenite; (vii) hydrolysis of hippurate, indoxyl acetate, and urea; (viii) production of hydrogen sulphide on triple sugar iron agar or in lead acetate strips; and (ix) susceptibility to naladixic acid and cephalothin antibiotics

[47,126,164,173]. Generally, *Campylobacter* are catalase and oxidase positive and can reduce nitrates; the catalase reaction is one of the few tests for differentiating the species [126,162].

Alternative discriminatory tests include direct immunofluorescent tests and enzyme-linked immune-sorbent assays (ELISA) [79,97]. Fluorescent antibody test (FAT) is a rapid, specific, and sensitive immunological test that has been used for many years for the detection of CF in preputial washings [46,55,92,147]. It has been used as a confirmatory test following bacterial isolation, but its use is limited due to its detection rate (between 10^2 and 10^4 CFU/ml) and infrequent availability of commercial antisera. Moreover, FAT cannot distinguish between CFF and CFV [49,57,118,164,169].

Vaginal mucus agglutination test (VMAT), complement fixation, and ELISA are serological assays applied to diagnose campylobaceriosis by screening of the reproductive material for the presence of CF antibodies [147,164]. Both VMAT and ELISA detect secretory anti-CFV IgA antibodies in the vaginal mucus following abortion [30,77,117,118,164]. However, problems with specificity and sensitivity of these assays have been reported [30,77]. Interestingly, the concentration of these IgAs remains constant within vaginal secretion with a long half-life; they could be detected for up to several months [68,79,164]. Additionally, when testing vaccinated animals with this test, false positive reactions are not observed since the vaginal mucus of these animals contains only IgG isotope antibodies [79,164]. However, these serological tests can fail to detect CFV antibodies in the collected preputial materials from bulls; therefore, they cannot be used for routine diagnostic screening [185].

1.2.8.4 Differentiation between *C. fetus* subspecies. Although the identification of CF is straightforward, the discrimination between the two subspecies (i.e. CFF and CFV) is much more challenging [65]. Currently, the only OIE-recommended test and the one most widely used for subspecies differentiation is tolerance to 1% glycine; CFF is able to grow in the presence of the glycine, while CFV is not [118]. However, strict attention to detail is required to achieve reproducible results in this assay so that inadequate standardization of the inocula may lead to false positive or negative results [27,167]. In addition some CFV strains are sensitive to polymyxin B and their isolation is confounded by contamination of the selective media with a number of commensal bacteria [27]. In some cases a mixture of phenotypes has suggested the existence of a distinct CFV subtype that grows in 1% glycine media and which is identified as CFVi [25,77,92,138]. Additionally, it is suggested that glycine tolerance may be acquired by transduction or mutation [27,78,143,174].

Because both CF subspecies clearly meet the requirements of subspecies designation (i.e. being classified as genetically closely related organism diverging in phenotype), researchers are currently focusing on identifying additional differences between the two subspecies [179]. One of the methods used to help to discriminate between the two CF subspecies is based on the presence of surface array protein (*sap)* type A and B [127]. Each CF bacterium produces up to three different proteins with molecular mass ranging from 97 to 149 kDa [41,43,44]. These serotypes reflect the differences in the structure of lipopolysaccharide and the type of surface layer protein [168,169]. Whilst CFV is always type A, CFF can be either type A or B. According to Tu and others in 2001, based on a phylogenetic relationship study, some CFF and CFV type A strains were placed on the same branch; and since then it was proposed that the CF differentiation must have occurred after the type A – type B split [165].

1.2.9 Molecular Methods currently used to discriminate *Campylobacter fetus* subspecies

Although CFF and CFV have distinct habitat niches, host preferences, biological properties and clinical differences, they are genetically very closely related [126]. Due to the labour-intensive nature of bacterial culture and the application of phenotypic and biochemical tests for subspecies differentiation, the application of rapid molecular diagnostic methods has been actively pursued [164].

DNA probes targeting 16S and 23S rRNA genes of *C. fetus* successfully differentiated *C. fetus* from *C. hyointestinalis,* despite sharing of 98% sequence homology in the 16S rDNA gene; however, these probes were not able to differentiate at the subspecies level [14,122,164].

Analysis of CF DNA by pulsed-field gel electrophoresis (PFGE) following digestion by SmaI and SalI restriction endonuclease enzymes demonstrated that CFF and CFV have a genomic size of 1.1 Mbp and 1.3 Mbp respectively; additionally, the newly described group termed CFVi had a slightly larger genome at 1.5 Mbp [164,170]. However, these values were underestimated since the actual genome size of the CF species ranges between 1.6 – 1.8 Mbp, as reported by Salama in 1992 [3,89,138].

Hum et al (1997) demonstrated the use of a multiplex PCR-based assay in the differentiation between the two CF subspecies using two sets of primers (VenSF/VenSR & MG3F/MG4R), which had been developed previously; see descriptions of these primers in

Appendix I, Table 1 [78,123,164]. The MG3F and MG4R primers amplify a 750bp fragment (originally incorrectly labeled as a 960 bp fragment) of the chromosomal carbon starvation (*cstA*) gene that is present in both subspecies of CF, while VenSF and VenSR primers amplify a 142 bp fragment (originally designated as "ven") of the unrelated *parA* gene that is present uniquely in CFV [75,78,160,175,182]. Initially, it was believed that the *parA* gene was plasmid-encoded but later studies failed to generate this PCR product from plasmid DNA from different CFV isolates, thereby suggesting that the *parA* target is a chromosomal-encoded gene. Recently, different genomic studies have shown that this *parA* gene is located within a genomic island (PICFV8) specific to many CFV strains but not in CFF [3,59,108]. This simple PCR assay, which discriminates CFF/CFV based on the presence of a small amplicon, has been extensively used in different studies and the results were for the most part concordant with the phenotypic tests and with results based on other molecular assays such as PFGE, and random amplification polymorphic DNA (RAPD-PCR) [112,115,120,143,164,166,173,182].

Further refinement of the *parA* PCR method led to the development of a 5' Taq-nuclease real-time PCR targeting this same locus as Hum et al (1997). The assay proved to be more sensitive than the previous methods in the detection of CFV from both vaginal mucus and smegma and preputial washes directly from the cows and bulls respectively [100]. However in 2005, the specificity of the subspecies *ven* primers was questioned when studies in the United Kingdom and South Africa found a poor correlation between PCR and biochemical results [143,164,168,169,182]. Schulze subsequently proposed that these South African isolates were evolutionary distinct from CF obtained from other geographical areas [164].

An alternate multiplex PCR, employing the MG3F/MG4R primers to amplify the 750bp amplicon of the CF *cstA* gene and a primer pair (nC1165g4F/nC1165g4R) to amplify a 223bp product of the *virB11* gene unique to CFV, was recently reported [81]. However, when those primers were tested at the beginning of this study the results were not concordant with traditional typing methods. Two real-time PCR assays that amplify both *nahE* gene and insertion sequence *ISCfe1* in the CF isolates have been reported in 2013. Although both assays show 100% specificity and sensitivity in the identification of the CF at the species level, differentiation between CFF/CFV at the subspecies level was not 100% accurate [99,171].

Other molecular methods that have been evaluated for their use in subspecies differentiation include fluorescent amplified fragment length polymorphism fingerprinting (AFLP) and multilocus sequence typing (MLST). AFLP is a random genome fingerprinting technique used for whole bacterial genome identification for epidemiological and taxonomic purposes [141]. This method has been successfully applied for epidemiological typing of most of *Campylobacter* species found in veterinary infections, not for subspecies identification [42]. MLST has been found useful for studying the epidemiology and population genetics of the genus *Campylobacter*, and especially to examine the epidemiological links of outbreaks [38,48,96,103,168]. It involves comparison of the sequences of several common genes; typically, seven housekeeping genes are used to identify the allele sequence profiles for a given isolate [59,96,168]. MLST results were somewhat correlated with glycine and H_2S tests, AFLP, and the *parA* PCR; while they showed strong agreement with PFGE data in the identification of a cluster of CF isolates from known outbreaks [78,168,169]. A MLST study has suggested the clonal nature of CFV and given rise to the theory that this subspecies represents a bovine clone of CFF [38,89,168] that may lack the ability to continuously adapt to various habitats [35,181]. Despite their use for exploring the epidemiology and evolution of these organisms these methods are not highly amenable to routine use in the diagnostic laboratory [164,168,169]. Furthermore they have not proven useful for studying mechanisms of pathogenicity [21,22,38,104]. Despite the clinical importance of CF, the determinants responsible for its virulence are poorly understood perhaps due to the importance of multifactorial processes that contribute to CF pathogenicity [21,124].

1.2.10 Approaches to explore the genetic differences between *C. fetus* subspecies

Alternate reliable and rapid methods that discriminate between CFF and CFV are still needed. There are a number of approaches that may be used to identify genetic loci of value for this purpose and these are briefly described.

(I) Suppression Subtractive Hybridization: The suppression subtractive hybridization (SSH) technique has been applied widely to distinguish between two closely related DNA samples and to detect intra-species variation in many bacteria. It has proven useful for the study of bacterial pathogenicity and is a potentially powerful and rapid approach for the study of pathogenic potential in CF [21,37,163].

The original SSH technique, developed by Bautz and Reilly in 1996, was first used to identify mRNAs representing a region deleted from DNA of a mutant T4 bacteriophage [21,37,154]. The technique was first applied in bacteria to study *Helicobacter pylori* in 1998 [2,21] and subsequently employed in studies of many different pathogenic bacteria [15,21,84,133,184,188]. Although SSH can be applied to study the genomic differences

(DNA-based) or the differences in gene expression (cDNA/RNA-based), the basic principle remains the same [21,184].

Subtractive hybridization is based on a process that reduces the proportion of homologous DNA/cDNA sequences from two closely related strains or cell types and promotes selection and hence recovery of unique fragments found only in the strain of interest [10]. The SSH depends primarily on a suppression polymerase chain reaction technique that combines normalization in which fragmented DNAs are equalized within the target population followed by subtraction when all the common sequences are subtracted from the population under study [21,146]. Therefore, the probability of obtaining low-abundance genomic DNA fragments or differentially expressed cDNAs libraries is increased up to 1000-fold [146].

In the SSH protocol, the strain that contains target sequences with a characteristic of interest is often termed the "tester", while the "driver" (also known as the reference) contains the common sequences that are present in both strains [21,74]. Both tester and driver DNA preparations are first digested by a frequently cutting restriction endonuclease to generate relatively short fragments (~500 bp) and then ligated to synthetic adaptors which form the targets for PCR primers. The tester and driver DNAs are combined (driver DNA is always provided in excess), denatured and then hybridized; DNA sequences that are present only in the tester are selected and enriched by PCR before subsequent analysis [21,146].

During suppression PCR the primers used are engineered to prevent non-specific and unwanted amplification; during each denaturation/annealing step any DNA that has the same adaptor at each end will form a secondary "panhandle-like" structure because the intramolecular annealing of the long adaptor sequences is favoured over the intermolecular annealing of the short primer sequences; accordingly no exponential amplification occurs in these non-target sequences [24, 155]. The overall procedure of the SSH is shown in Figure 5.

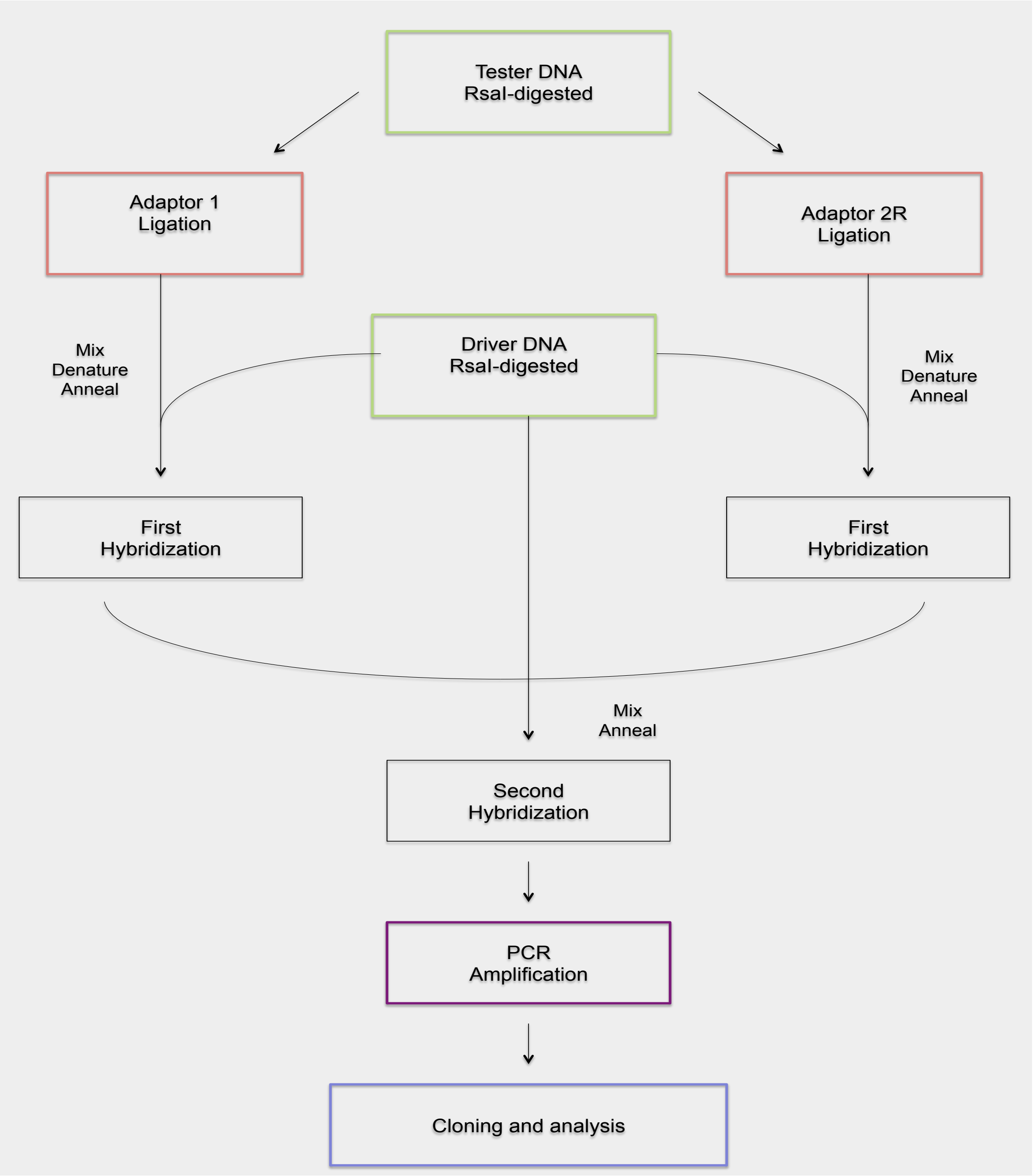


Figure 5. The overall workflow of the suppression subtractive hybridization (SSH) procedure. Briefly, both tester and driver DNAs are digested with the RsaI restriction endonuclease enzyme, following by separate ligation reactions of the digested tester DNA using adaptor 1 (Adp1) and adaptor 2R (Adp2R). Each tester/adaptor ligated DNA is separately hybridized with an excess digested driver, therefore any tester/driver or driver/driver hybrids should be removed from the reaction leaving tester/tester hybrids. Then, the two hybridized samples are eventually mixed with an addition of excess digested driver for further enrichment of tester/tester DNAs. Consequently, these sequences are subjected to PCR amplification using primer P1 designed to bind to both adaptor sequences. The two adaptors are supplied in the kit and the numbers (1 and 2R) are just identifiers. Details of the adaptors and primer are found in Table II (Appendix 2). Figure reproduced and adapted from (PCR-Select^{TM} Bacterial Genome Subtraction Kit User Manual, Protocol No. PT3170-1. ©2009 Clontech Laboratories, Inc. (www.clontech.com).

(II) Whole Genome Sequencing: Since the sequencing of the first *Haemophilus influenzae* genome in the early 1990s, more than 1800 complete bacterial genome sequences have been determined and reported in the Genomes Online Database [83,140,149,159], and a 2012 study indicated that over 4000 bacterial genome projects are ongoing [83,137]. The availability of whole genome sequence information yields a complete picture of the coding capacity for that organism that can subsequently be explored to answer comprehensive biological questions, such as understanding mechanisms of bacterial pathogenicity in the host [83].

The traditional approach to whole bacterial genome sequencing involved preparation of a genomic library using one of two approaches: (i) the shotgun *de novo* method, in which DNA template is randomly cut into fragments and cloned into a plasmid vector or (ii) targeted re-sequencing, in which a PCR reaction amplifies the targeted DNA template [51,83,140,149,156]. Once the DNA library is prepared it is subjected to standard Sanger sequencing chemistry often performed these days as a cycle sequencing reaction in which elongating chains are terminated with the incorporation of dideoxynucleotides that are fluorescently labeled [51,83,156]. These terminated fragments are separated based on their sizes by capillary electrophoresis and fluorescent dye traces are translated into DNA sequence [83,149,151,156]. The Sanger method has up to 99.999% accuracy and while it is still used extensively for sequencing of relatively small regions of genomic DNA, much higher throughput can be achieved with next generation sequencing technology [83].

Next generation sequencing (NGS) is an alternative strategy for DNA sequencing that generates hundreds of mega-to giga-bases of nucleotide sequences in a single run [87]. Of the three principal platforms currently in extensive use, only the Illumina platform, which is the most commonly used, is described in detail here. The Illumina workflow involves four steps: DNA extraction, preparation of DNA library, production of clonal clusters, and sequencing as illustrated in Figure 6. DNA is first prepared by random fragmentation, followed by in *vitro* ligation of common adaptor sequences; single DNA molecules are amplified on a solid surface or chip using bridge PCR and the adaptor sequences to generate clusters of identical DNA fragments; this is followed by a sequencing reaction in which the four bases are sequentially added to the chip and stepwise addition of a single base to each DNA cluster is documented fluorescently following electrophoretic analysis [24,39,87,106,107,148]. Initial read lengths with this technology ranged from 35 – 100 bp but currently longer read lengths of 250 bp are possible [87]. One advantage of the Illumina system is the use of a paired end strategy in which both ends of the DNA fragment are sequenced. This is advantageous during the assembly process in which large numbers of short DNA sequence reads are combined into long contiguous reads, a process that can be achieved using several assembler software programs [83,129]. Following assembly, the genome can be annotated at two main levels: (i) static level to obtain protein coding genes, GC content, genomic islands, codon usage, functional RNA products, and chemical and structural characteristics and subcellular localization of several proteins; and (ii) dynamic level to view gene context or order, regulatory, metabolic and protein interaction networks and comparative genomics [83,101]. Nowadays, bacterial genomes are often annotated using a web-based pipeline [101,153].

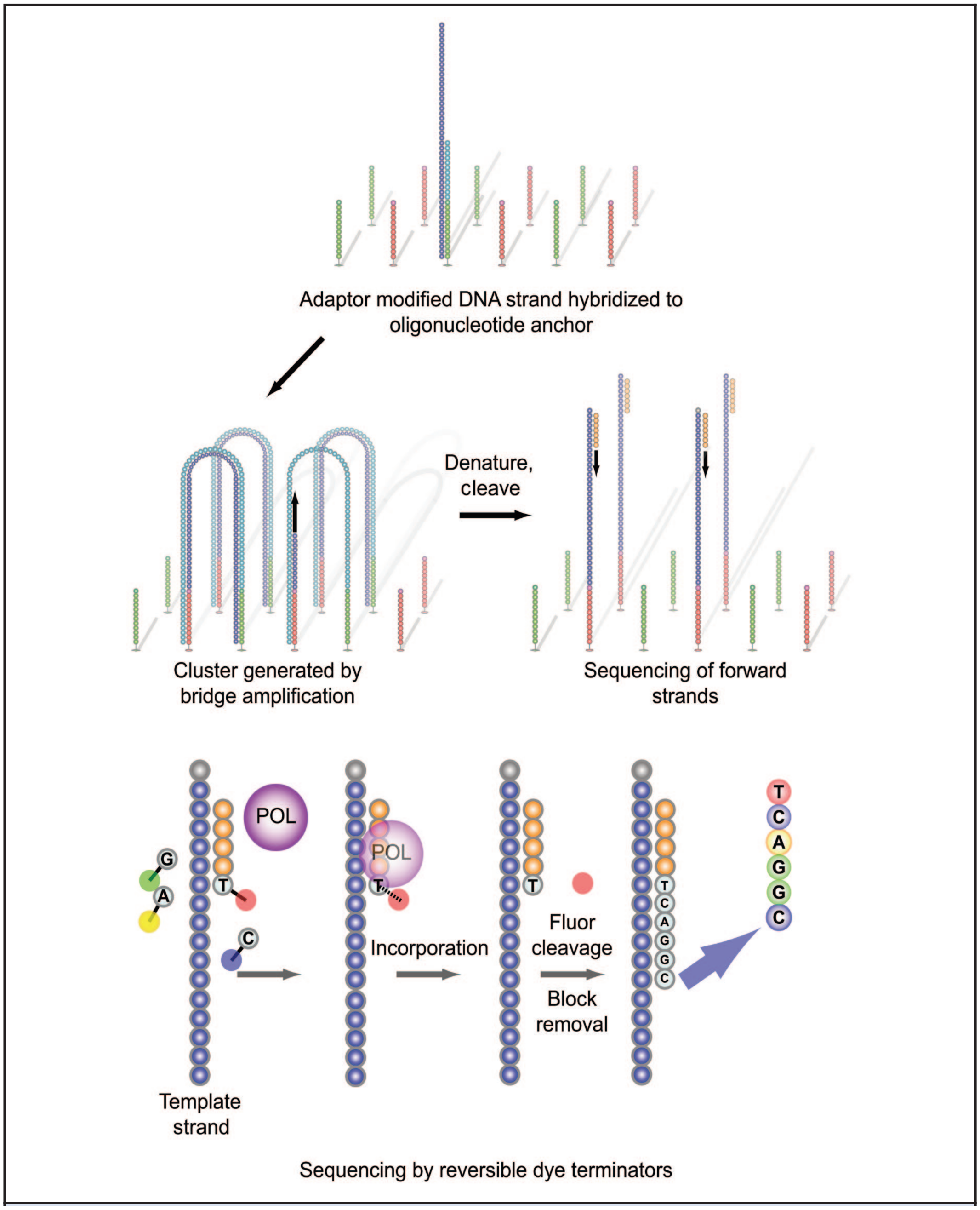


Figure 6. The overall workflow of Illumina Genome Analyzer Sequencing. Single stranded DNA is ligated to modified adaptors and immobilized on a solid surface. Using bridge PCR (aka 'cluster PCR') clonal clusters are generated for sequencing. The clusters undergo cycles of denaturation, cleavage, and sequencing by adding one nucleotide at a time using four fluorescently tagged reversible dye terminators. Figure adapted from HighWire Press (Karl V. Voelkerding, Shale A. Dames, and Jacob D. Durtschi. (2009) Next-generation sequencing: From basic research to diagnostics. Clin Chem. 55: 641-658). © with permission from American Association for Clinical Chemistry.

(III) Optical Mapping: Optical mapping is another novel and powerful comparative genomic technique that can be used together with the whole genome sequencing to compare two closely related genomes [28]. An optical map is an assemblage of a number of partial restriction fragment maps into a single complete genome restriction map [9]. It is a largely robust and an automated system, which generates high-resolution ordered restriction maps from entire bacterial chromosomes [20,28,94].

To construct a whole-genome restriction map, the purified single chromosomal DNA molecule is immobilized on a positively charged optical glass surface through microfluidic channels and then cut with a suitable restriction endonuclease; this maintains the order of the digested fragments [9,20,135,145]. The resulting DNA fragments are stained with a fluorochrome dye JOJO-1 (Invitrogen), visualized by fluorescence microscopy, and imaged using a digital camera. The intensity of the fluorescence is directly proportional to the mass and thus length of each digested fragment (Figure 7) [135,145,186]. A fully automated digital image acquisition system collects the images, processes them to create optical maps and then all the restriction maps are overlapped using optical map assembler software to generate a single whole genome map that indicates the position along the DNA strand of the recognition sites where the enzyme cleaves the DNA [9,113,135,145].

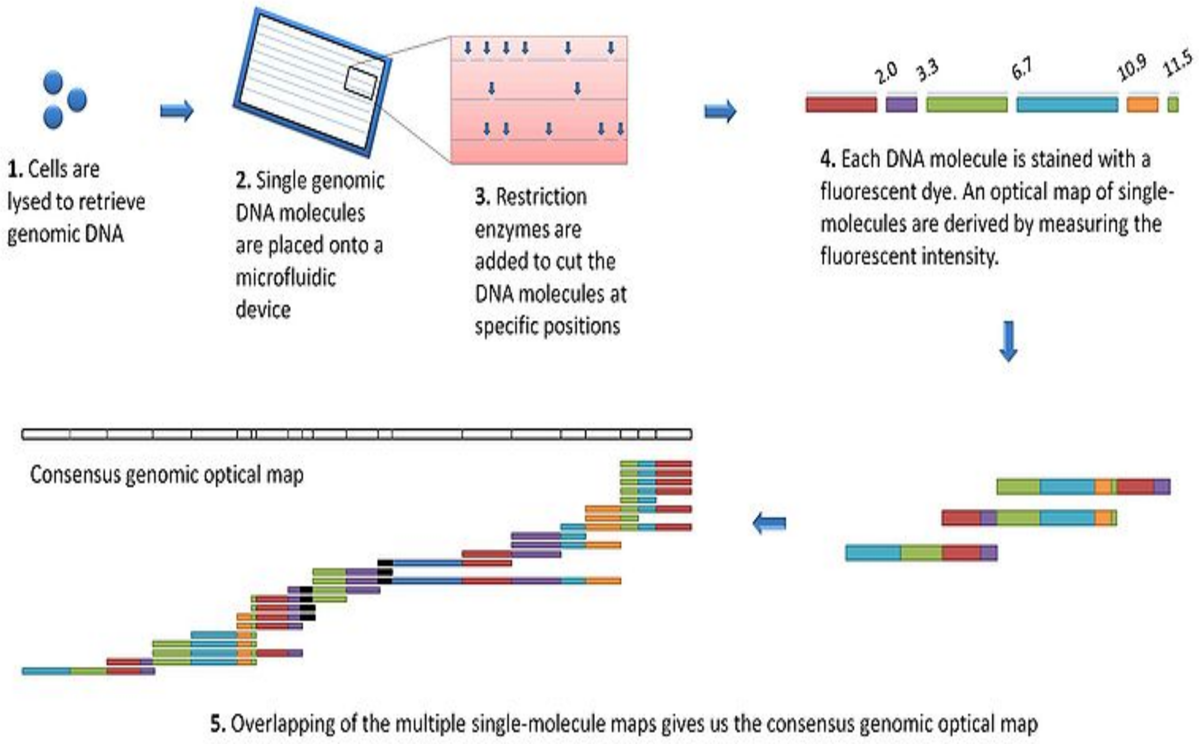


Figure 7. The Workflow of the optical mapping technique. A pure high molecular weight genomic DNA is isolated using detergent-based lysis protocol, adhered and elongated on the microfluidic device of the optical mapping surface via micro channel and electrostatic interaction, elongated following by fragmentation of the tested material by restriction endonuclease enzyme, and the digested DNAs are fluorescently stained and imaged on epi-fluorescent microscope. Figure adopted from (Wikipedia and Brian Teague 2010). This figure is licensed under the Creative Commons Attribution 3.0 Unported (//creativecommons.org/licenses/by/3.0/deed.en).

Comparison of optical maps between two closely-related organisms allows major structural variations (i.e. insertion, deletion, duplication, translocation, inversion, orientation and gross rearrangements) between the genomes to be detected. Generation of *in silico* restriction maps from the sequences of reference strains allows comparison and alignment with a strain of interest [189]. This methodology is also a valuable tool to assist in the finishing and checking of the assemblies of microbial whole genome sequencing projects [135,189]. Additionally, the use of genomic DNA as the source for mapping eliminates the need for PCR amplification, library construction, hybridization or separation on a pulse-field gel electrophoresis and their attending artifacts [135,186].

1.3 Hypothesis

Campylobacter fetus subspecies are considered to be useful models for studying the molecular basis of host adaptation due to their close genetic relationship and host preferences. It is presumed that the different life-styles of CF subspecies are reflected in their genetic composition. As more information becomes available regarding the genomic differences between these subspecies knowledge of the key genetic factors defining these life-styles will be acquired.

Thus, the hypothesis of this project is that better characterization of Canadian CFV isolates, which differ significantly from European and reference isolates with respect to the presence of a genomic island, will reveal genes potentially important for these phenotypic characteristics.

1.4 Objective of the study

1) Compare and examine the genetic differences between Canadian field isolates of the two CF subspecies, CFF and CFV, and use novel genetic loci to develop a robust molecular diagnostic test to accurately differentiate these two subspecies.
2) Identify candidate genes that may be linked to the distinct pathogenicity of these two subspecies.

1.5 Strategies for genome comparison

Two main strategies are employed in these studies:

1) A subtractive hybridization approach using a selected well-characterized Canadian field isolate of CFV recovered from a 2008 outbreak in Alberta that does not contain significant portions of the previously identified genomic island (PICFV8) to identify alternative CFV-specific genes.
2) A combination of whole genome sequencing and optical mapping to compare the complete genomes of CFF and CFV isolates.

Chapter Two: Materials and Methods

2.1 Bacterial Strains

The bacterial strains (49 isolates) used in this study are listed in Table 2.

Table 2. *Campylobacter fetus* strains and their relevant characteristics. All *Campylobacter* isolates were obtained from the Animal Health Microbiology (AHML) diagnostic unit, Ottawa Laboratory Fallowfield (OLF), Canadian Food Inspection Agency (CFIA), Ottawa, ON, Canada. Isolates recovered from routine diagnostic submissions were designated according to a specific identification system; the first two numbers indicate the year when the sample was taken, the A stands for Artificial Insemination, the following three digits indicate the specific lab number for that submission and the final number and letter represent the sample number in that submission. Isolates received from other sources were placed in an ADRI collection and given a unique number.

#	Strain number	Subspecies*	Description	Source of reference	Country of origin
1	82-40[1]	CFF	Preputial washings from Bulls	NC_008599.1	USA
2	NCTC 10354[2, 3]	CFV	Vaginal mucus of heifer	CM001228.1	UK
3	ATCC 27374[4, 5]	CFF	Brain of sheep fetus	ATCC	USA
4	ATCC 19438[4, 6]	CFV	Vaginal mucus of heifer	ATCC	UK
5	02A725-35A[7, 10, 11]	CFF	Preputial washings from Bulls	This study	Canada
6	08A1102- 42A[8, 10, 11]	CFV	Preputial washings from Bulls	This study	Canada
7	08A948-2A[10, 11]	CFV	preputial washings from Bulls	This study	Canada
8	08A110233B.5A[10]	CFV	Preputial washings from Bulls	This study	Canada
9	ADRI 1024[9, 11]	CFV	Preputial washings from Bulls	This study	Argentina
10	ADRI 520[9, 11]	CFV	Preputial washings from Bulls	This study	Sweden
11	08A1102 3A[11]	CFV	Preputial washings from Bulls	This study	Canada
12	08A1102 24A[11]	CFV	Preputial washings from Bulls	This study	Canada

13	08A1204 32A[11]	CFV	Preputial washings from Bulls	This study	Canada
14	08A1204 17B[11]	CFV	Preputial washings from Bulls	This study	Canada
15	08A1102 39A[11]	CFV	Preputial washings from Bulls	This study	Canada
16	08A1204 1A[11]	CFV	Preputial washings from Bulls	This study	Canada
17	08A1204 12A[11]	CFV	Preputial washings from Bulls	This study	Canada
18	ADRI 1345[9, 11]	CFVi	Preputial washings from Bulls	This study	Argentina
19	ADRI 510[9,] 11	CFVi	Vaginal mucous	This study	Canada
20	ADRI 545.1A[9, 10, 11]	CFVi	Vaginal mucous	This study	Australia
21	ADRI 546[9]	CFVi	Preputial washings from Bulls	This study	Australia
22	ADRI 1362.8A[9, 10,11]	CFF	Preputial washings from Bulls	This study	Argentina
23	ADRI 1346[9, 11]	CFF	Preputial washings from Bulls	This study	Argentina
24	09A9803.2B[10]	CFF	Preputial washings from Bulls	This study	Canada
25	Turbo serotypeB[10]	CFF	Preputial washings from Bulls	This study	Canada
26	STRAIN 802[11]	CFF	Preputial washings from Bulls	This study	Canada
27	01A988-2A[11]	CFF	Preputial washings from Bulls	This study	Canada
28	ADRI 553[9]	CFF	Preputial washings from Bulls	This study	Canada
29	ADRI 1032[9, 11]	CFF	Preputial washings from Bulls	This study	Argentina
30	ADRI 516[9, 11]	CFF	Preputial washings from Bulls	This study	Sweden
31	05A451 9B[11]	CFF	Preputial washings from Bulls	This study	Canada
32	07A621 15A[11]	CFF	Preputial washings from Bulls	This study	Canada
33	08A314 4B[11]	CFF	Preputial washings from Bulls	This study	Canada

34	09A376 - 1A[11]	CFF	Preputial washings from Bulls	This study	Canada
35	08A1242 2A[11]	CFF	Preputial washings from Bulls	This study	Canada
36	AIN458-3[11]	CFF	Preputial washings from Bulls	This study	Canada
37	98AIN525-1[11]	CFF	Preputial washings from Bulls	This study	Canada
38	01A603-82A[11]	CFF	Preputial washings from Bulls	This study	Canada
39	07A1157 #66[11]	CFF	Preputial washings from Bulls	This study	Canada
40	03A564-113A[11]	CFF	Preputial washings from Bulls	This study	Canada
41	04X189 1B[11]	CFF	Preputial washings from Bulls	This study	Canada
42	09A980 - 3A[11]	CFF	Preputial washings from Bulls	This study	Canada
43	09A980 -16B[11]	CFF	Preputial washings from Bulls	This study	Canada
44	06A1204 293A[11]	CFF	Preputial washings from Bulls	This study	Canada
45	06A1553 3A[11]	CFF	Preputial washings from Bulls	This study	Canada
46	ADRI 1359[11]	CFF	Preputial washings from Bulls	This study	Canada
47	01A1038 1B[11]	CFF	Preputial washings from Bulls	This study	Canada
48	08A1043 9B[11]	CFF	Preputial washings from Bulls	This study	Canada
49	09A9801 12B[11]	CFF	Preputial washings from Bulls	This study	Canada

Table 2 - Footnotes:

* Subspecies determined by the AHML, OLF, CFIA.

1. CFF82-40 GenBank # NC_008599.1 or CP000487.1. This strain submitted in 22-NOV-2006 from the Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA. This strain was used as a reference during genome assembly, WGS analysis and optical mapping.
2. NCTC - National Collection of Type Culture
3. NCTC10354 GenBankv# CM001228.1. Bacteria available from the NCTC, Health Protection Agency, Microbiology Services, Porton Down, Salisbury, SP4 0JG, UK. This strain was used as a reference during genome assembly, WGS analysis and optical mapping.
4. ATCC - American Type Culture Collection.
5. (ATCC® 27374™) *C. fetus fetus* (Smith and Tylor; Veron and Chateline, 1973), serotype B, strain designation: [NCTC 10842]. This isolate was used as a reference during the SSH and for PCR screening, as a control, using targeted CFV-specific genes (Clone #35, PICFV5_ORF548 and CFF Feature_3).
6. (ATCC® 19438™) *C. fetus venerealis* (Florent) Veron and Chateline, 1973, serotype A, strain designation: NCTC 10354 [CIP 6829, X/161/5]. This isolate was used as a reference during the SSH and for PCR screening, as a control, using targeted CFV-specific genes (SSH_Clone #35, PICFV5_ORF548 and CFF_Feature #3).
7. CFF strain used for the SSH as a driver DNA, and for WGS and optical mapping analyses.
8. CFV strain used for the SSH as a tester DNA, and for WGS and optical mapping analyses.
9. ADRI - Animal Disease Reseach Institute.
10. Following WGS of these strains, *de novo* assembly was performed by CLC genomics and used for BLAST analysis.
11. Strains were used for PCR screening using targeted CFV-specific genes (SSH_Clone #35, PICFV5_ORF548 and CFF_Feature #3).

2.2 Media, Buffers, and Solution preparations and sterilization

All the media, solutions, and glassware used in this study were sterilized by autoclaving at 121°C for 15 - 20 min. Antibiotics and other solutions that could not tolerate autoclaving were filter-sterilized by passing them through a 0.22 µm filer (Millipore type GS). All solutions were prepared using double-distilled water. Media and solution preparations are listed in Tables 3 and 4.

Table 3. Preparation of the media used for this study.

Bacterial Medium	Recipe	Usage
Muller-Hinton blood agar (MHA)	Dissolve 30g beef infusion, 17.5g acid hydrolysate casein, 1.5g starch, 17g Bacto-agar, 50ml defibrinated sheep blood in H_2O. Adjust pH to 7.4, bring to 1L and autoclave.	*Campylobacter fetus* isolation (Prepared by AHML, OLF, CFIA)
2X Yeast extract Tryptone (YT)	Dissolve 31g of 2X YT medium powder (DIFCO) in 1L H_2O, boil for 1 min, and autoclave at 121°C for 15 min.	Growth of *Escherichia coli* (TG-1 strain) to determine optimal bacterial cell density for DNA extraction
Luria Bertani (LB) broth + 100 µg/mL Amp	Dissolve 10g tryptone, 5g Bacto-yeast extract, 10g NaCl in H_2O. Adjust pH to 7.5, bring to 1L and autoclave. After autoclaving cool to ~ 55°C, add 100 µg/mL of filter-sterilized ampicillin.	Growth of *E. coli* clones
Luria Bertani (LB) + 100 µg/mL ampicillin plates	Prepare LB broth with 1.5% (w/v) Bacto-agar. After autoclaving cool to ~ 55°C, add 100 µg/mL of filter-sterilized ampicillin and pour into petri dishes.	Cloning & Transformation

Table 4. All solutions and buffers used in this study.

Solution	Constituent	Usage
6X DNA loading buffer	Mix 50% (v/v) Glycerol, 0.25% (w/v) Bromophenol blue, 0.25% (w/v) Xylene cyanol in10 ml H_2O	Provides density and color to DNA to be loaded in agarose gel electrophoresis
10 mg/ml Ethidium bromide	Dissolve 0.2 g ethidium bromide in 20 ml H_20. Store in a dark area at 4°C	Visualize DNA following agarose gel electrophoresis
50X TAE	Mix 242 g Tris base, 57.1 ml glacial acetic acid, 18.61 g Na_2EDTA in 1L H_2O (pH 8.0)	Used at a 1X dilution as running buffer for agarose gel electrophoresis
1XTE	Mix 10 mM Tris-HCl (pH 7.5) and 1 mM EDTA (pH 8.0)	General buffer and diluent for DNA Rehydration following extraction
1M Tris	Dissolve 121.1 g Tris base in 1L H_2O (pH 7.5)	General buffer and diluent
0.5M EDTA	Dissolve 186.1g EDTA in 1L H_2O (pH 8.0)	General buffer used at various dilutions
Phenol:Chloroform:Isoamyl Alcohol (25:24:1)	Mix 12.5 ml phenol, 12 ml chloroform, 0.5 ml isoamyl alcohol	DNA purification
Chloroform:Isoamyl Alcohol (24:1)	Mix 24 ml chloroform, 1 ml isoamyl alcohol	DNA purification
4M NH_4OAc	Dissolve 20.52 g NH_4OAc in 250 ml H_2O	DNA precipitation
2X Hybridization buffer	Equal volume of 4X hybridization buffer (Clontech) and deionized H_2O	Secondary Subtractive Hybridization
10X Maleic Acid buffer	Mix 116.1 g Maleic acid, 87.6 g NaCl in 1L H_2O (pH 7.5)	Used at 1X concentration
1X Blocking working solution	Dilute 10X blocking solution stock (Roche) 1:10 in 1X maleic acid buffer	Blocking nonspecific binding sites on the membrane
1X Anti-DIG working solution	Dilute anti-DIG-AP conjugate (Roche) 1:10000 (75 mU/ml) in 1X blocking solution	Binding to DIG-labeled probe
1X DIG Washing working buffer	Mix 50ml of 10X maleic acid buffer, 1.5ml Tween20 in 450ml H_2O (pH 7.5)	Removal of unspecific/unbound antibody
1X DIG Detection working buffer	Mix 50 ml 1 M Tris-HCl, 50 ml 1 M NaCl in 400ml H_2O (pH adjusted to 9.5)	To adjust pH to 9.5 for Alkaline phosphatase
Color-Substrate working solution	Mix 200 µl of NBT (BioRad), 200 µl BCIP (BioRad) in 10 ml 1X DIG detection buffer	Colorimetric substrate to visualize DNA/probe hybrid
CSPD ready-to-use	1 ml of 25 mM solution (Roche)	Chemiluminescent substrate to visualize DNA/probe hybrid
Depurination solution	0.25 M HCl	Transfer of the DNA out of the gel

Denaturation solution	1.5 M NaCl, 0.5 M NaOH	Unzip DNA into single strands for probe hybridization
Neutralization	1.5 M NaCl, 0.5 M Tris−HCl (pH 7.0)	To bring pH to < 9.0 during blotting procedures
Low stringency wash	2X SSC and 0.1% SDS	Post hybridization wash to remove non-specific DNA/probe hybrids
High stringency wash	0.5X SSC and 0.1% SDS	Post hybridization wash to remove non-specific DNA/probe hybrids
DIG Easy Hyb working solution	Add 64 ml sterile H_2O in two portions to the DIG Easy Hyb Granules (Roche), dissolve by stirring immediately for 5 min at 37°C	Pre-hybridization of DNA (Provided with DIG-High Prime kit)

2.3 Bacterial Culture Conditions

All CF isolates were cultured in Muller-Hinton Agar (MHA) containing 10% defibrinated sheep's blood, with different antibiotics to prevent growth of extraneous organisms such as Vancomycin 'for gram positive', Trimthoprim 'a bacteriostatic for UTI bacteria', and Cyclohexamide 'for eukaryotic protein synthesis inhibition'. The culture was incubated in an anaerobic jar at 37°C under microaerophilic condition (4% O_2, 9.5% CO_2, and 86.5% N_2) using GENbag microaer system from bioMèriux. Colonies of CF appear after 4 – 5 days on the plate as non-hemolytic, 1-3 mm round, smooth, translucent or grey colored. All suspicious colonies were picked and examined for motility by phase contrast microscopy and to determine presence of a polar flagellum responsible for the rapid corkscrew motility typical of this genus. Cellular morphology was observed as gram negative in gram strain, non-spore forming curved bacilli. Bacterial strains were supplied and cultured by the Animal Health Microbiology diagnostic unit (AHML), Ottawa Laboratory Fallowfield (OLF), Canadian Food Inspection Agency (CFIA).

2.4 Biochemical characterization of the isolates

ELISA and full biochemical workup were applied to confirm the identity of CF bacteria. Since there are no published standard phenotypic methods for the CF identification, the biochemical tests were performed according to general published studies and modified by the AHML (CFIA) [126,138,174]. The complete list of phenotypic tests used for the CF identification and subspecies differentiation is listed in Table 5.

Table 5. List of different biochemical tests used for CF identification and subspecies differentiation. All the biochemical identification information was supplied from the Animal Health Microbiology diagnostic unit, OLF, CFIA.

Identification Test	Expected outcome if *C. fetus* is present
Gram's Stain	Gram negative
Growth at 25°C, 35°C, and 42.5°C	Grow at 25°C and 35°C, not at 42°C
Catalase test	Positive
Oxidase test	Positive
Susceptibility to Naladixic acid and Cephalothin	Sensitive to Cephalothin and resistance to Naladixic acid
Growth in 1% bile (Bile tolerance)	Positive
Growth in 3.5% Sodium Chloride (NaCl)	Negative
Growth in 0.04% Triphenyl Tetrazolium Chloride (TTC)	Negative
Hippurate hydrolysis	Negative
Indoxyl acetate hydrolysis	Negative
Urea hydrolysis	Negative
Selenite reduction	Depend on the subspecies
Anaerobic growth in TMAO medium	Negative
Indirect Fluorescent Antibody (FAT)	Positive
Hydrogen Sulphide (H_2S) test	Depend on the subspecies
Glycine tolerance	Depend on the subspecies

2.5 Genomic DNA Isolation and Manipulation

2.5.1 Estimation of Bacterial Optical Density

Colonies of CF were scraped into 1.5 ml microcentrifuge tubes containing 480 μl of 50 mM EDTA. The optical density (OD) of this suspension was measured spectrophotometrically using a Genesys spectrophotometer at a wavelength of 600 nm. An illustration of the optimal starting amount of bacterial cells required for DNA extraction is found in Figure 8. An OD_{600} of 0.5 ($\sim 2 \times 10^9$ cells/ml) was the optimal value needed for the DNA extraction.

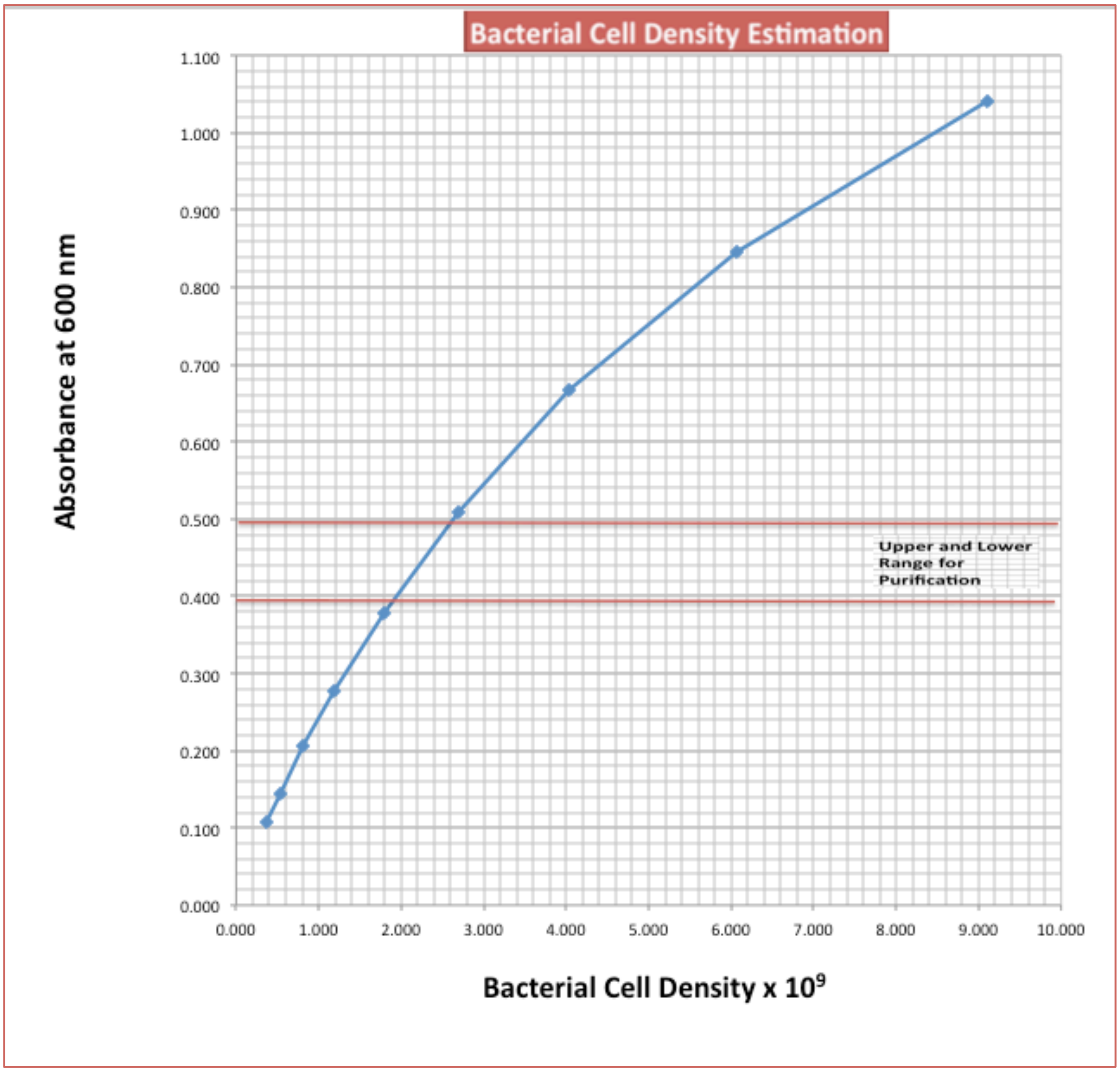


Figure 8. A standard curve illustrating the A600 versus the cell density. This curve was used to identify an OD_{600} value of 0.4 – 0.5 as equivalent to the optimal cell density of ~2 × 10^9 cells/ml required for DNA extraction. Experimental details are found in Appendix 1.

2.5.2 DNA Extraction

The isolation of genomic DNA from CF strains was performed from about 2 × 10^9 cells/ml of bacteria using a Wizard® Genomic DNA Purification kit, Cat# A1120 (Promega, Madison, WI, USA). The kit was employed using the manufacturer's recommendations. Briefly, bacterial cells were lysed, RNA was removed, proteins were precipitated, and the recovered DNA was resuspended in a 1X TE buffer to be used on the subsequent steps. Following the extraction, the concentration of the genomic DNA was measured using a GE Healthcare NanoVue™ Spectrophotometer, Buckinghamshire, UK, or Thermo Scientific NanoDrop™ 1000 Spectrophotometer, Wilmington, USA. Absorbance was determined at 260 nm and 280 nm and both concentration (based on A260) and purity (estimated by the ratio of A260/A280) were automatically calculated. An A260/A280 ratio between 1.8 – 1.9 indicated a highly purified DNA preparation.

2.6 Polymerase Chain Reaction

Amplification of genomic DNA fragments by PCR was used throughout the project for testing the integrity of DNA following genomic DNA extraction, testing the presence of the genomic islands, serotyping, subtractive hybridization, and for testing the specificity of the primers generated from whole genome sequencing analysis. All primers used in this study were synthesized by Integrated DNA Technologies (IDT, Coralville, USA) and are listed in Appendix 2, Table I. PCR was performed in a thermal Cycler GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA, USA). Amplification reactions were performed in a total volume of 50 µl containing 1X Taq polymerase buffer (provided with the enzyme), 1.5 mM $MgCl_2$, 0.5 µM forward and reverse primers, 0.2 mM dNTPs, 2.5U Taq DNA polymerase, and 20 ng – 2 µg genomic DNA. A standard PCR cycling program comprises an initial denaturation step of 95°C for 1 min followed by 35 cycles of DNA denaturation (94°C for 1 min), primer annealing (46°C – 55°C for 1 min), and primer-extension (72°C for 2 min) with a final extension at 72°C for 7 min. PCR amplicons were stored at 4°C or -20°C for temporary or permanent storage, respectively.

2.6.1 Testing Genomic Integrity

Following DNA extraction, the quality of genomic DNA was tested by amplifying the 16S and 23S rRNA genes. The primers used for this PCR testing assay were C412F-C1228R, for the 16S rRNA, and 23SF-23SR, for the 23S rRNA. PCR cycling conditions were slightly modified; the 16S and 23S rRNA primers were annealed at 46°C.

2.6.2 Screening for the PICFV8 Genomic Island

The genomic DNA of various CF isolates were screened and tested for the presence of the CFV-genomic island (PICFV8) previously identified by Gorkiewicz et al (2010) and Ali et al (2012), by amplifying different regions within PICFV8 (Figure 7) using 5 different primer sets (Appendix 2, Table I). PCR cycling conditions were slightly modified from the standard protocol; primer sets GI1 – GI4 were annealed at 70°C, while the GI5 (VenSF/R) set was annealed at 50 - 54°C.

2.6.3 PCR amplification of the CFF strains to determine *sap* gene type

Detection of the *sapA* and *sapB* genes was performed by PCR using gene-specific primers (Appendix 2, Table I). Cycling conditions were slightly modified as *sap* primers were annealed at 47°C.

2.6.4 CF Targeted Genes Screening

Screening a number of Canadian CF isolates for the presence of targeted genes (identified from SSH and WGS studies) was done using the standard PCR protocol. About 50 – 300 ng genomic DNA was used in a final volume of 50 µl PCR master mix, and all the primers (Appendix 2 Table I) were annealed at 50°C (for both cln35 F1/R1 primers and CFF3 F/R primers) or 57°C (for PI548 F/R).

2.6.5 Visualization of DNA using Agarose Gel Electrophoresis of DNA

Agarose solutions (0.8 - 2.0%) were prepared, depending on the level of resolution required, in 1X TAE buffer supplemented with 0.6 µg/ml ethidium bromide and cast into gels using a Bio-Rad sub-cell electrophoresis system. Gel wells were loaded with samples that had received 1/10 volume of 6X loading dye (Table 5) and the gel was run at 50-75 volts for 1 hr. Appropriate size markers (i.e. 100 bp / 1Kbp) were run in parallel for DNA size determination. DNA was visualized using a UV transIlluminator and gels were documented using a Bio-Rad Gel Doc system.

2.7 Suppression Subtractive Hybridization

Suppression Subtractive Hybridization was performed using the PCR-Select^{TM} Bacterial Genome Subtraction Kit (Cat# 637404, version# PT3170-1) obtained from Clontech Laboratories (Mountain View, CA), and followed according to the manufacturer's directions. DNA from CFV isolate 08A1102- 42A (Acc. # 6, Table 2) was used as tester and DNA from CFF isolate 02A725-35A (Acc. # 5, Table 2) as driver. The workflow as detailed in the kit manual and reproduced in Figure 9 was followed. All adaptors and primers used during the SSH approach are listed in Table II of Appendix 2.

The same procedure was performed using the control DNA (one copy of 3 ng/µl of HaeIII-digested φX174 DNA per 1 mg/ml *E. coli* genomic DNA as provided with the kit) to evaluate the efficiency of the subtractive hybridization technique. Slight modifications were done to the original DNA-based suppression subtractive hybridization protocol to overcome some frequent issues [2,37].

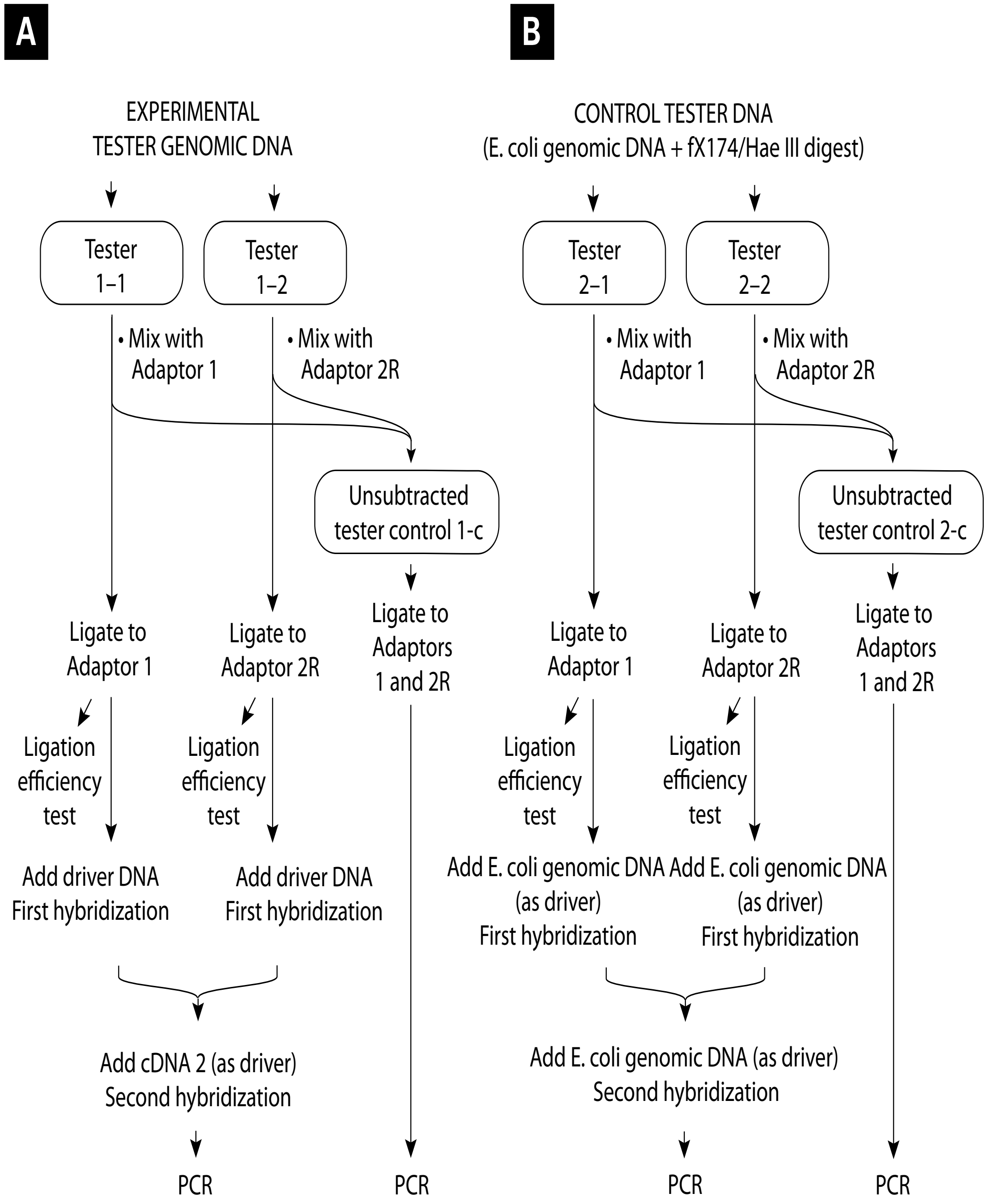


Figure 9. Workflow detailing the subtractive hybridization steps. Each tester DNA, experimental and control, is ligated with adaptors 1and 2R. Testers 1-1 and 1-2 refer to CFV-DNAs ligated to adaptors 1 and 2R, respectively (Panel A), while testers 2-1 and 2-2 refer to the mixture of *E. coli* genomic DNA + φX174/HaeIII-digest control DNA ligated with adaptors 1 and 2R, respectively (Panel B). When Adaptors 1 and 2R are ligated to *Rsa* I-digested DNA, the *Rsa* I site is restored. Samples 1-c and 2-c represent the negative control for the subtraction. Samples were subjected to 2 rounds of hybridization followed by 2 rounds of PCR amplification. Figure adapted from PCR-Select™ Bacterial Genome Subtraction Kit User Manual, Protocol No. PT3170-1. ©2012 Clontech Laboratories, Inc. (www.clontech.com).

2.7.1 Restriction Endonuclease Enzyme Digestion of Genomic DNA

Two micrograms of genomic DNA from tester (CFV) and driver (CFF), as well as the control *E. coli* genomic DNA, were completely digested overnight at 37°C with 1.5 µl RsaI enzyme, New England BioLabs, in a total volume of 50µl (2 µg of gDNA, 1X RsaI restriction buffer, 10U RsaI enzyme, and deionized H_2O). Five microliters of the digested DNAs were set aside for later analysis. The reaction was terminated by adding 2 µl of 0.2 M EDTA. Digested genomic DNA samples were then purified by performing phenol:chloroform extraction steps and ethanol precipitation as described in the manufacturer's instructions (Clontech, Mountain View, CA),

The dried pellet was resuspended in 6.5 µl deionized H2O and final concentration was measured using the GE Healthcare NanoVue™ Spectrophotometer. The digested products were stored at –20°C. Digest efficiency was examined by DNA gel electrophoresis to ensure a complete digestion of the DNA.

2.7.2 Ligation of Oligonucleotide Adaptors to Tester DNA

The digested tester DNAs of both experimental CFV (tester 1) and control DNA (tester 2), supplied with the kit, were aliquoted into two separate tubes and each underwent three adaptor ligations (Figure 10 and Table 8): adaptor 1 [Adp1] (tester 1-1 and tester 2-1), adaptor 2R [Adp2R] (tester 2-1 and tester 2-2), and unsubtracted tester controls (portions of each ligated tester DNA) which serve as a negative control for subtraction (1-c and 2-c). The ligation reactions (Table 6) were each set up in 10 µl final volume and contained 1X ligation buffer, 1µl of diluted digested-tester DNA, and 10µM adaptor 1 & 2R and 400U T4 DNA ligase (Clontech, Mountain View, CA); the mixture was incubated overnight at 16°C. Two microliters 0.2M EDTA was added the next day to terminate the reaction.

The efficiency of the ligation reaction was tested prior to proceeding with the hybridization, to verify that at least 25% of the fragmented testers have both adaptors at each end. During this experiment (Table 7), PCR amplified fragments that span DNA/adaptor junctions of both testers, by using primer P1 provided with the kit, and a primer set designed for amplifying specific gene sequence in the tester itself (primers for the 23S rRNA gene, that were included in the kit, were used for the *E. coli* control and *ven* primers were used for the CFV). Importantly, the primers should amplify DNA fragments that do not contain a RsaI site. Advantage®2 PCR kit (Protocol No. PT3281-1, Clontech Laboratories Inc, Mountain View, CA) was used for this PCR amplification and the protocol was employed using the manufacturer's recommendations. The reactions were placed in a thermal cycler and incubated for 5 min at 72°C to fill in gaps and extend adaptors sequences by adding the suitable nucleotides bases; thus creating binding sites for the PCR primer P1. Next, the reactions were subjected to 25 cycles of 94°C for 30 sec, 65°C for 30 sec, and 68°C for 1 min. Each PCR product was analyzed by electrophoresis in a 2% agarose/EtBr gel.

Table 6. Setting up the ligation reaction. Table adapted after modification from PCR-Select™ Bacterial Genome Subtraction Kit User Manual, Protocol No. PT3170-1· ©2012 Clontech Laboratories, Inc. (www.clontech.com).

	Test tube number					
	1	**2**	**3**	**4**	**5**	**6**
Component	**Tester 1-1**	**Tester 1-2**	**1 - C**	**Tester 2-1**	**Tester 2-2**	**2 - C**
Diluted tester	1µl	1µl	1.5 µl of tubes 1 & 2	1µl	1µl	1.5 µl of tubes 4 & 5
Adaptor1 (10µM)	2µl	-		2µl	-	
Adaptor2R (10µM)	-	2µl		-	2µl	
Ligation master mix	7µl	7µl		7µl	7µl	
Final volume	10µl	10µl	3µl	10µl	10µl	3µl

Table 7. Tester/adaptor ligation efficiency PCR set up. In case of CFV-tester DNA, VenSF and VenSR primers were used to amplify the *ven* gene, while in the case of E. coli control DNA genome, the 23SF and 23SR primers were used to amplify the 23S rRNA gene.

Component	Tube **1**	Tube **2**	Tube **3**	Tube **4**
Tester 1-1	1 µl	1 µl	-	-
Tester 1-2	-	-	1 µl	1 µl
10µM of Forward Primer	1 µl	1 µl	1 µl	1 µl
10µM of Reverse Primer	-	1 µl	-	1 µl
PCR Primer 1	1 µl	-	1 µl	-
Total Volume	3 µl	3 µl	3 µl	3 µl

* Tester 1-1 and 1-2 represent the ligated tester with Adp1 and Adp 2R, respectively. P1 PCR primer matches the long strands of adaptors 1 and 2 at their 5' ends.

2.7.3 Hybridization of Tester and Driver Genomic DNAs

Once tester-adaptor ligation had been verified, an excess of driver DNA was added to each tester-adaptor molecule, and samples were heat denatured and annealed allowing hybridization between the tester and the driver. Any tester-tester hybrids were enriched and underwent a second round of hybridization, while tester-driver and driver-driver hybrids were diluted out.

2.7.3.1 First Hybridization. During the first hybridization, 2 µl of each ligated tester (with adaptor 1 and adaptor 2R) were mixed with 1µl of digested driver DNA, and 1 µl of 4X hybridization buffer (Clontech, Mountain View, CA). The mixture was incubated at 98°C for 1.5 min in a thermal cycler and then annealed and hybridized at 63°C for 1.5 hr. After that, the reaction immediately proceeded to the second hybridization step.

2.7.3.2 Second Hybridization. The two samples from the first hybridization were mixed together (inside the thermal cycler) and added to 2 µl denatured mix of digested driver and 2X hybridization buffer (Table 1). The sample was incubated overnight at 63°C allowing complete hybridization. The following day, 200 µl of dilution buffer was added to the hybridized mixture, incubated at 63°C for 7 min in the thermal cycler and then stored at -20°C; this should promote formation of double stranded PCR product and eliminating any non-specific hybridization.

2.7.4 Amplification of Generated Tester-Specific DNAs

After the two hybridization steps, five different molecules were formed; three of them should not amplify because they are missing either one or both adaptor sites, or have formed a self-dimer molecule due to suppression PCR effect. The fourth type consists of double-stranded DNA that binds with the same adaptor-annealing site, so they would only amplify linearly. The fifth type contains the double stranded DNA with both adaptor sites attached at opposite ends allowing for exponential amplification. The mixture was briefly incubated at 72°C prior to primary thermal cycling to extend the adaptors and fill-in any missing strands creating binding sites for the PCR primers. A minimum five PCR reactions were set up for the primary PCR (Figure 10): (i) PCR control subtracted DNA (provided with the kit as positive PCR control and not shown in Figure 10; it contains a successfully subtracted mixture of *HaeIII*-DNA φX174 fragments), (ii) unsubtracted *E.coli* tester control, (iii) subtracted *E. coli* control DNA, (iv) unsubtracted experimental tester control, and (v) subtracted experimental tester DNA. The Advantage®2 PCR kit (Protocol No. PT3281-1), Clontech Inc, was used based on the manufacturer's recommendations for this amplification process.

2.7.4.1 Primary PCR. One microliter of each diluted subtracted, unsubtracted, and PCR control subtracted DNA was aliqouted into 5 different clean PCR tubes; and 24 μl of the primary PCR master mix (Table 8) was added to each tube. After incubation of the mixtures at 72°C for 2 min, the reactions were subjected to 30 cycles of denaturation at 94°C for 30 sec, annealing at 66°C for 30 sec, and extension at 72°C for 1.5 min. Seven microliters were aliqouted from each PCR reaction and set aside to be analyzed by 2% agarose/EtBr gel electrophoresis.

Table 8. Preparation of the primary PCR master mix. Table reproduced and adapted from (PCR-SelectTM Bacterial Genome Subtraction Kit User Manual, Protocol No. PT3170-1. ©2012 Clontech Laboratories, Inc. (www.clontech.com).

Component	Amount per Rxn (in µl)
Sterile H_2O	19.5 µl
10X PCR reaction buffer	2.5 µl
dNTP mix (10µM)	0.5 µl
PCR primer P1 (10µM)	1.0 µl
50X Advantage 2 polymerase mix	0.5 µl
Total volume	24.0 µl

2.7.4.2 Secondary PCR. Two microliters of each primary PCR mixture were diluted in 38 µl of H_2O, and then each diluted sample (1 µl/sample) was placed into a clean tube and used in a nested PCR. The primers used were NPI and NP2, which match the 3' ends of adaptor 1 and adaptor 2R, respectively. During this step 24 µl of the secondary PCR master mix (Table 9) was added to each tube. Tubes were placed in the thermal cycler and subjected to 15 cycles of denaturation at 94°C for 30 sec, annealing at 68°C for 30 sec, and extension at 72°C for 1.5 min. Seven microliters were aliqouted from each PCR reaction and set aside to be analyzed by 2% agarose/EtBr gel electrophoresis. The rest of the reaction products were stored at - 20°C for subsequent cloning and transformation.

Table 9. Preparation of the secondary PCR master mix. Table reproduced and adapted from (PCR-SelectTM Bacterial Genome Subtraction Kit User Manual, Protocol No. PT3170-1. ©2012 Clontech Laboratories, Inc. (www.clontech.com).

Component	Amount per Rxn (in µl)
Sterile H_2O	18.5 µl
10X PCR reaction buffer	2.5 µl
Nested primer NP1 (10µM)	1.0 µl
Nested primer NP2 (10µM)	1.0 µl
dNTP mix (10µM)	0.5 µl
50X Advantage 2 polymerase mix	0.5 µl
Total volume	24.0 µl

2.8 Real-Time PCR

Quanta biosciences kit (Cat# 95051-100, Gaithersburg, MD) was used for real-time PCR analysis of the subtracted samples (secondary PCR) to give an indication of sample selection. Forward and reverse primers for VenS and CstA were used with the appropriate probes. The real-time (qPCR) was performed on an ABI 7500 Sequence detector. Each reaction mixture contained 10 µl of PerfeCTa super mix containing 4X reaction buffer with optimized concentrations of $MgCl_2$, dNTPs, AccuFast Taq Polymerase, ROX reference dye and stabilizer. About 0.6µl of 10µM of both forward and reverse primers, and 0.2 µl of 10 µM probe were included in each reaction. Three master mixes were prepared: (i) for the CstA by using CstA F and R primers and CstA probe, (ii) for the CFV by using VenS F2 and R primers and CFV probe, and (iii) for the CFF by using VenS F2 and R primers and a CFF pseudogene probe. The final volume of 20 µl was denatured at 95°C for 5 min then subjected to a thermal cycle profile of repeated 40 cycles of denaturation at 95°C for 3 sec and annealing at 53°C for 30 sec. Having a threshold cycle (CT) of 15, a point where the qPCR detects the first fluorescence above background noise, is acceptable.

2.9 Rapid purification of PCR products

The amplicons from the secondary PCR were effectively purified to remove amplification primers, primer-dimers and any remaining PCR components using a Promega Wizard® PCR Preps DNA Purification kit with a Vacuum-Manifold System, Cat# A2180 (Promega, Madison, WI, USA). The kit was employed according to manufacturer's instructions. About 30 – 45 µl of each PCR product was used for the purification. The DNA was eluted from the PCR Preps DNA Purification Resin in 50 µl pre-warmed nuclease-free dH_2O and the purified DNA was stored at -20°C to be used for subsequent cloning.

2.10 Cloning and Transformation

Subtractive hybridization products from the secondary PCR were inserted and cloned into the PCR®2.1-TOPO® vector using a TOPO TA Cloning® Kit, with, (Cat# K4500-01, Invitrogen, Burlington, ON, CA) in order to generate a library of subtracted sequences. Four microliters of each secondary PCR were mixed with 2 µl of salt solution and 1 µl of TOPO cloning vector (Appendix 3), and the mixture was incubated for 30 min at RT. This mix was then used immediately for the transformation step. Four microliters of the TOPO cloning reaction was transformed and gently mixed with One Shot®Top10 chemically competent *E. coli* cells (Invitrogen, Burlington, ON, CA); the reaction was incubated on ice for 30 min and heat shocked at 42°C for 30 sec and then kept on ice. Instantly 250 µl of RT S.O.C. medium (Invitrogen, Burlington, ON, CA) was added to the transformed reaction and the mixture was capped and shaken at 37°C for 1hr. Next, 50 and 100 µl from each transformation was spread onto pre-warmed LB agar plates (Table 4), containing 100 µg/ml of either ampicillin or kanamycin, and overlaid with 100 mM of IPTG and 40 mg/µl of X-gal. Plates were incubated in an inverted position overnight at 37°C. On the following day, recombinant small white colonies were picked up using sterile toothpicks, and each colony was streaked onto an LB agar plate and sub-cultured in 3 ml new pre-warmed LB broth; both LB agar and broth contain100 µg/ml Amp (Table 4).

2.11 Colony PCR

E. coli cells from the LB liquid culture were lysed by incubation at 95°C for 5 min and 1 - 2 µl of the suspension were then used as DNA template in the colony PCR, using the universal M13 forward and reverse primers, to confirm the presence and the size of the inserts. PCRs were prepared and cycled using the standard PCR conditions as shown in Table 7 except the final volume was 25 µl. Five microliters of each PCR product was separated by agarose gel electrophoresis to assess DNA concentration and size of the fragment.

2.12 Differential Screening of the Subtracted Library

The inserts within the clones generated by the SSH approach were screened to determine their presence/absence in the two isolates used for SSH, CFF02A725-35A (Acc. #5, Table 2) and CFV08A1102- 42A (Acc. #6, Table 2). This was done by generating a DIG-labeled probe from each clone and examining its ability to hybridize to DNA from both CF subspecies. DNA probes were made using two methods: random primed labeling and PCR labeling techniques, due to availability of kits for these methods at the time.

2.12.1 DIG High Prime-DNA Labeling

DIG High Prime DNA Labeling and Detection Starter Kit II, Cat#11745832910 (Roche Diagnostics, Laval, QC, Canada) was used to generate DIG-labeled DNA probes according to the manufacturer's recommendations. Around 0.5 – 1 µg of DNA was diluted in 16 µl of ddH_2O as template for DIG labeling and mixed with 4 µl of 5X DIG High Prime mix, following by overnight incubation at 37^{o}C. On the next day, 0.2 M EDTA (pH 8.0) was added to terminate the reaction, and the efficiency of the labeled DNA was tested and compared to the DIG-labeled control DNA (provided with the kit) to determine the optimal yield of DIG-labeled DNA prior to hybridization; high probe concentration would cause high background level, while too low concentration would lead to weak signals. About 1 ng/µl of both DNA probes and DIG-labeled control DNA were serially diluted according to the manual protocol. These dilutions were applied to a positively charged nylon membrane, fixed by UV crosslinking and the DIG-labeled DNAs were then subjected to immunological and chemiluminescence detection using anti-digoxigenin-AP conjugate and ready-to-use, substrate CSPD (Roche Applied Science substrate), respectively. The intensities of each diluted DIG-labeled DNA were compared to the control DNA, and the optimal concentration determined; about 0.1 pg/μl probe concentration is considered the lowest concentration recommended in the hybridization.

2.12.2 DIG PCR-DNA Labeling

The PCR DIG Probe Synthetic Kit, Cat#11636090910, (Roche Diagnostics, Laval, QC, Canada) was used as an alternative method for probe generation according to the manufacturer's protocol. It contains a 2:1 ratio of dTTP:DIG-dUTP, which allows generation of highly sensitive hybridization probes. The DIG is attached to dUTP via an alkali-labile ester bond, which allows the labeled dUTP to be easily incorporated into the DNA synthesis using DNA polymerase. To evaluate DIG incorporation with this method, PCRs were performed using both PCR DIG labeling mix and dNTP mix, to generate labeled and unlabeled DNAs respectively. The 50 μl PCR mixes contained 100 pg DNA, 5 μl of 10X PCR buffer with $MgCl_2$, 0.5 μl of 10 μM M13 forward and reverse primers, 0.75 μl of 2.6U Expand High Fidelity enzyme mix, and 5 μl of 200 μM either PCR DIG probe synthesis mix (in case of DIG-labeled DNA) or dNTP stock solution (in case of unlabeled DNA). Thermal cycling was performed using the following conditions: initial denaturation at 95^{o}C for 1 min and 25 cycles of 1 min at 95^{o}C, 1 min at 55^{o}C, 2 min at 72^{o}C, followed by a final elongation at 72^{o}C for 7 min. The efficiency of DIG incorporation was examined by up-shift analysis in which the band size of the unlabeled DNA was compared to that of the labeled product on an agarose mini gel since the presence of DIG in DNA makes the DNA run more slowly in the gel.

2.12.3 DNA Slot Blot Preparation

For each sample of CFF/CFV, about 1 µg DNA was diluted in 1X TE (pH 7.5) to a final volume of 400 µl; 40 µl of 3 N NAOH was added and the mixture was incubated at 65°C for 45 min. After cooling 440 µl of 2 M NH4OAc (pH7.0) was added. This solution was transferred to nylon membrane (Cat# 11209272001, Roche) using a Minifold™ II slot blot device (Cat# 10447800, Whatman, GE Healthcare). After transfer using a vacuum system, DNA was fixed to the membrane by UV crosslinking (Fisher FB UV XL-100 crosslinker, Fisher Scientific Biotech).

2.12.4 Southern Blot Analysis

Additional screening was done for the selected clones to test their presence/absence among other CFFs/CFVs. Genomic DNAs of various CF isolates were digested with SauIII, electrophoresed through an agarose gel and transferred to a nylon membrane by capillary electrophoresis according to the Southern blot protocol. DNA was fixed to the membrane by UV crosslinking. Probes were hybridized with these blots. An oligonucleotide corresponding to a sequence of the carbon starvation (*cstA*) gene was tail-labeled with DIG-dUTP/dATP using a DIG Oligonucleotide Tailing Kit (Cat# 03353583910, Roche Diagnostics, Laval, QC, Canada) and included in the screening analysis as a positive control for the CF strains.

2.12.5 Probe-DNA Hybridization

DNA membranes were subjected to a pre-hybridization step by adding 10 ml/100cm^2 filter of a DIG Easy Hyb solution (Table 2) to each membrane. Membranes were incubated at 41°C (hybridization temperature for *Campylobacter fetus*) for 30 min in a sealed hybridization bag placed in a roller-bottle in an Inter Science Hybaid Mini- hybridization oven (Inter Science). Then a hybridization mix was prepared by mixing 25 ng of DNA-DIG labeled probe in 1 ml DIG Easy Hyb buffer and this was denatured at 95°C for 5 min and rapidly cooled on ice. The pre-hybridization solution was recovered and replaced by the probe/hybridization mixture (3.5 ml/100cm^2). The membranes were allowed to hybridize at 41°C overnight with rotation in the hybridization oven. The following day, the hybridization mix was removed and the membranes were washed with a low stringency buffer (Table 4) twice for 5 min at RT with continuous shaking followed by washes in a high stringency buffer pre-warmed to 65°C (Table 4) twice for 15 min with constant rotation. The recovered probes were stored at -20°C to be reused again.

2.12.6 Immunological Detection

Following high stringency washes all membranes were subjected to immunological and chemiluminescent detection assays. All incubations were carried out at RT on a platform shaker (VWR Signature™ 3-D Rotator Waver, cat# 12620-916, CA). First, the membrane was rinsed with DIG washing buffer (Table 4) for 5 min at RT and then incubated in 100 ml of 1X DIG working blocking solution (Table 4) for 30 min to block nonspecific DNA binding sites, followed by incubation in 20ml of 1X blocking solution containing sheep-derived Anti-DIG-AP-Fab fragment antibody (diluted 1:10000 in 1X blocking solution

(Table 5) for 30 min to bind anti-DIG antibody Fab to the DIG-labeled probe. Membranes were washed twice with 100 ml washing buffer for 15 min to remove unbound antibody, and then equilibrated in 20 ml of 1X DIG detection buffer (Table 4) for 3 - 5 min. Finally, DNA hybrids were visualized in 1 ml chemiluminscence, ready-to-use, substrate CSPD (Roche Applied Science substrate) following incubation in the dark. The enzymatic dephosphorylation of CSPD by antibody-conjugated alkaline phosphatase led to light emission which was imaged with a digital Kodak camera.

2.13 Plasmid DNA preparation

Plasmid DNA was extracted for use in DNA sequencing with a Wizard® Plus Minipreps DNA Purification System, Cat# A7500 (Promega, Madison, WI, USA). About 3 ml of an overnight culture (from section 2.11) was used for the extraction, performed according to the manufacturer's instructions. Briefly, bacterial cells were lysed and neutralized, followed by selective plasmid DNA recovery using a column containing silica gel from which the plasmid was eluted using 50 µl pre-warmed nuclease-free H_2O. The concentration of the plasmid DNA was measured using a GE Healthcare NanoVue™ Spectrophotometer prior to use in sequencing.

2.14 DNA sequencing and BLAST analysis

Sanger DNA sequencing was performed on plasmid DNA from selected clones identified by SSH using the BigDye® Terminator v3.1 Cycle Sequencing Kit, PN#4337035, (Applied Biosystems, Foster City, CA) which employs four different fluorescent dyes to identify the four dideoxy nucleotides. Reagents of the kit were used at a 1/8 dilution as directed, together with M13 forward or reverse primers and DNA template, and thermocycling was performed using the following profile: initial denaturing at 96ºC for 1 min, then 25 cycles sequencing of 96ºC for 10 sec, 50ºC for 5 sec, and 60ºC for 4 min, followed by a 4ºC hold. Sequencing products were purified using the BigDye® XTerminator™ Purification Kit, PN#4374408, (Applied Biosystems, Foster City, CA) and analyzed on an ABI 3500xl Genetic analyzer. Variant Reporter software v1.1 (Applied Biosystems®) was used to generate a consensus sequence from the forward and reverse reads for each clone and this was exported in FASTA format for subsequent BLAST analysis. The sequences of the RsaI and adaptors were removed from the nucleotide sequences of the selected clones to facilitate bioinformatics analysis. Additional analysis was done for the selected clones in order to know which of the clones correspond with the ORFs of the putative pathogenicity island of the NCTC10354.

2.15 Whole Genome Illumina Sequencing and data analysis

At least 10 μg DNA was prepared from each of eight CF isolates (Ac. c# 5, 6, 7, 8, 20, 22, 24 and 25 – Table 2) using the Wizard Genomic DNA Prep kit. All DNA preparations had a 260/280 ratio between 1.6 to 1.8. These samples were submitted to the BC Genome Centre for Illumina sequencing using a paired end strategy to generate 100 base reads from both ends of 500 bp inserts. Raw sequence data were provided as two FASTQ files per sample.

The reads generated from Illumina-NGS were assembled in two ways: one was a reference-based assembly using the Burrows-Wheeler Alignment (BWA) Tool [93], and the second was by a *de novo* approach using CLC Genomics software [18]. Since NGS data generates thousands of short reads and portions of the genome may remain uncovered, the *de* *novo* approach generates large numbers of contigs that need to be ordered and oriented correctly and gaps frequently cannot be filled in without further analysis. In this project the order of the generated contigs was confirmed using optical mapping data with OpGen MapSolver software [9], and alignments with reference sequences using the Mauve software [34]. Bacterial genomes were annotated using the RAST software [7]. Regions of particular interest were also analyzed and compared to CFF/CFV reference sequences and several additional *de novo* assembled *C. fetus* genomes using both BLAST and Mega version 5 software [157]. Following the bioinformatics analysis, primers for conventional PCR were designed using the IDT PrimerQuest^{SM} [121].

2.16 Optical Mapping

Three CF isolates (Acc. # 5, 6, and 19 – Table 2) were submitted to the laboratory of Dr. K. Amoako (CFIA Lethbridge Laboratory, Alberta, Canada.) for optical mapping analysis using the Argus^{TM} Optical Mapping System (OpGen Inc., Gaithersburg, MD) [102]. Maps based on restriction digestion by NcoI were generated and provided as XML files, which could be analyzed using the Argus^{TM} OPGen MapSolver^{TM} software (OpGen Inc., Gaithersburg, MD) [102]. For comparative purposes this software could also generate maps in *silico* from nucleotide sequence data.

Chapter Three: Results

Table 10. Summary of experimental approaches used during the study and their outcomes.

Technique	Principle of the technique	Strains	Results
Phenotype (1% Glycine)	Inoculation the cell suspension onto a medium containing 1% glycine (OIE recommended).	5, 6, 7, 9, 19, 22, 23	Positive for CFFs (3, 5, 22, 23) and for CFVi (19), while Negative for CFVs (4, 6, 7, 9),
gDNA integrity following extraction	PCR amplification of both 16S and 23S rRNA genes.	5, 6, 7, 9, 19, 22, 23	All CF amplifies both genes with amplicon sizes 816bp and 650bp for the 16S rRNA and 23S rRNA, respectively.
Genomic Island (PICFV8) screening	PCR amplification of GI1 – GI5 within PICFV8.	5, 6, 7, 9, 19, 22, 23	▪ All CFFs isolates (5, 22, 23) were negative for GI1 - GI5. ▪ All CFVs (6, 7, 9) were negative for GI1 – GI4, except GI5 (positive). ▪ CFVi (19) was positive for all GIs.
***sap* PCR**	PCR amplification of *sapA* and *sapB* genes in the CFF isolates.	5, 22, 23	All CFFs have *sapA* gene.
SSH	Identification of target genes in the CFV genome but absent from the CFF.	5, 6	SSH_Clone #35 is a useful CFV-specific target located within PICFV10 of NCTC10354 reference strain. It codes for hypothetical protein.
Whole genome sequencing and Optical mapping	Using Illumina-MiSeq platform for whole genome sequencing and Argus MapSolver system for the optical mapping.	5, 6, 7, 8, 20, 22, 24, 25,	▪ CFF_Feature #3 is a useful target for CFF that located within PICFF10 of CFF82-40 reference strain. It codes for UDP-galactopyranoside, an enzyme for LOS biosynthesis in the cell wall of gram-negative bacteria. ▪ ORF548 is a CFV-specific target region that located in the PICFV5 of NCTC10354 reference strain. It codes for hypothetical protein.

3.1 Phenotypic Characterization of *C. fetus* Isolates

For preliminary studies seven CF field isolates were selected for analysis based on their biochemical and genetic attributes as determined previously. Staff of the Microbiology Diagnostic Unit at the Canadian Food Inspection Agency performed all biochemical testing on these isolates using 72 hr bacterial cultures with comparison to the reference ATCC strains (CFF 27374 and CFV 19438). The results of all tests are summarized in Table 12. According to OIE recommendation, CF is classified as CFF, CFV, or as CFV *biovar intermedius* (CFVi), based on tolerance to 1% glycine and the ability to metabolize the sulfur-containing amino acid cysteine to produce hydrogen sulfide, detected as a black precipitation on TSI (Triple Sugar Iron) or lead acetate strips. By confirming the results with the CFF/CFV reference strains, 3 of the 7 isolates tested were able to grow on a blood medium containing 1% glycine (glycine tolerance) and accordingly identified as CFF, while other 3 strains were unable to grow on the glycine-supplemented medium and consequently classified as CFV. The last strain, which originated from the bovine genital tract, was relatively glycine tolerant and thus recognized as a CFVi isolate.

Table11. Differential biochemical characteristics of Campylobacter fetus species isolated from the bovine genital tract and aborted fetuses. The data were provided by the diagnostic AHML, OLF, CFIA.

		Tests that identify the C. fetus																						Tests that differentiate between C. fetus subspecies								
Reference Assigned Number	Source	Gram Strain	Motility (Wet Mount)	Morphology	Growth in/at			Catalase	Oxidase	Nitrate Reduction	Nitrite Reduction	Susceptibility to		Growth in			Hydrolysis of			Selenite Reduction	Growth on MacConkey Agar	Anaerobic Growth in TMAO	Fluorescent Antibody Test (FAT)	C-ELISA on pure culture (Serotype A, B, or AB)	H2S Production		Growth in Glycine						Isolate ID
					Aerobic growth	25°C	42.5°C					Nalidixic Acid NA-30	Cephalothin KF-30	Bile 1.0%	NaCl 3.5%	TTC 0.04%	Hippurate	Indoxyl Acetate	Urea						TSI	Lead Acetate	0.60%	1.00%	1.10%	1.30%	1.50%	1.90%	
3	ATCC 27374	GNR	+	Curved Rod	-	+	V	+	+	+	-	R	S	+	-	-	-	-	-	+	+	-	very good	B	-	+	+	+	+	+	+	+	CFF (Reference strain)
4	ATCC 19438	GNR	+		-	+	-	+	+	+	-	R	S	+	-	-	-	-	-	-	+	-	good	A	-	-	+	-	-	-	-	-	CFV (Reference strain)
5	02A725-35A	GNR	+		-	+	-	+	+	+	-	R	S	+	-	-	-	-	-	+	+	-	very good	No info	-	+	+	+	+	+	-	-	CFF
22	ADRI 1362	GNR	Darting		-	+	-	+	+	+	-	R	S	+	-	-	-	-	-	+	+	-	very good	A	-	+	+	+	+	+	+	-	CFF
23	ADRI 1346	GNR	Darting		-	+	-	+	+	+	-	R	S	+	-	-	-	-	-	+	+	-	good	A	-	+	+	+	+	+	-	-	CFF
6	08A1102- 42A	GNR	Darting		-	+	-	+	+	+	-	R	S	+	-	-	-	-	-	-	+	-	good	A	-	-	+	-	-	-	-	-	CFV
7	08A948-2A	GNR	Darting		-	+	-	+	+	+	-	R	S	+	-	-	-	-	-	-	+	-	very good	A	-	-	+	-	-	-	-	-	CFV
9	ADRI 1345	GNR	Darting		-	+	+	+	+	+	-	R	S	+	-	-	-	-	-	w+	+	-	good	A	-	-	+	-	-	-	-	-	CFV
19	ADRI 510	GNR	N/A		N/A	N/A	N/A	+	+	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	slant -/stab +	N/A	N/A	N/A	A	N/A	+	+/+	-/+	-/-	-/-	N/A	N/A	CFVi

R: Resistance, **S:** Sensitive, **+**: positive, **-:** negative, w+: weak positive, **N/A**: non-available
CFF: Campylobacter fetus fetus, **CFV:** Campylobacter fetus venerealis, **CFVi:** Campylobacter fetus venerealis biovar intermedius

3.2 Assessment of genomic DNA integrity following DNA extraction

Following DNA extraction using the Wizard Genomic DNA Purification Kit as described, the genomic integrity and suitability of the DNA from the seven selected isolates for molecular analysis was confirmed by amplifying the 16S and 23S rRNA genes that code for small and large subunits of bacterial ribosomal RNA. These genes are conserved in all *C. fetus* isolates. The expected products were 815bp for the 16S rRNA and 650bp for the 23S rRNA (Figure 10). The targeted amplicon production indicated the suitability of genomic DNAs for molecular analysis.

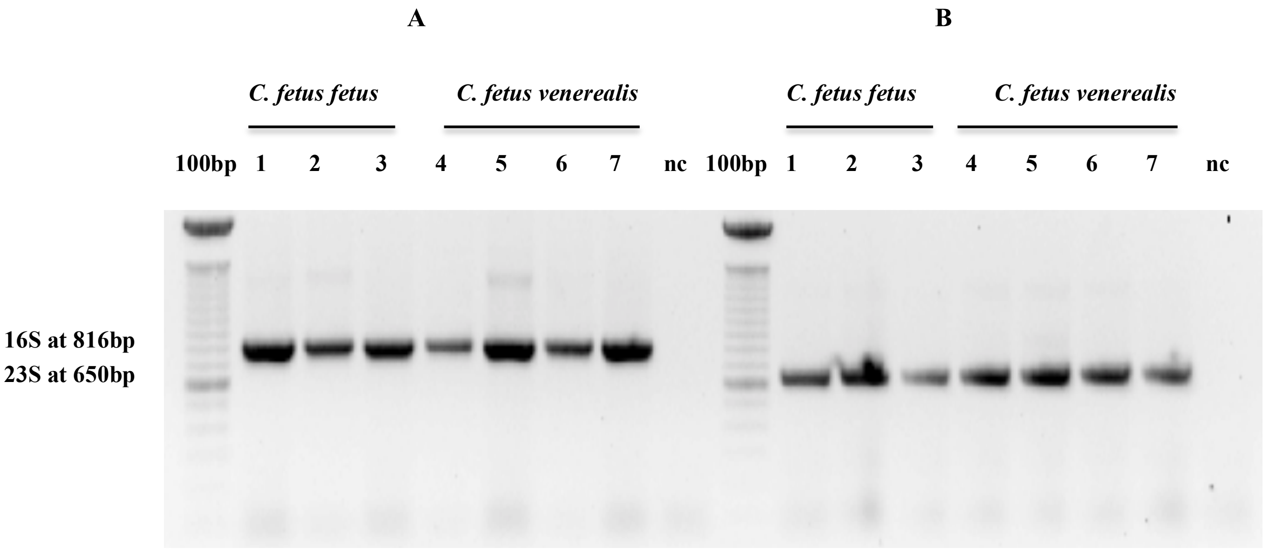


Figure 10. Check of *C. fetus* genomic DNA for suitability in PCR. PCRs targeting the rRNA genes were performed and analyzed by agarose gel electrophoresis. Lanes 1 – 3 represent CFF isolates (Acc. #5, 22, and 23 as in Table 2), while lanes 4 – 7 are CFV strains (Acc. # 6, 7, 9 and 19 as mentioned in Table 2). The gel was run at 100V for 30 min, and the image of the DNA bands were obtained using an UV transIlluminator Gel Doc system. The DNA molecular marker (100bp) is shown on each panel. Length of amplicons is 816bp for the 16S rRNA gene (panel A), and 650bp for the 23S rRNA gene (panel B). ATCC strains of CFF/CFV were used as positive controls for this assay (data not shown) while the negative control (nc) is shown on each panel.

3.3 Screening for the CFV genomic island (PICFV8-T4SS)

Five different PCRs directed to several locations within the CFV genomic island (PICFV8) were employed to screen for the presence of this sequence. PCR cycling conditions were optimized for each primer set based on the positive control DNA (CFF and CFV of ATCC strains); where GI fragments 1-4 were amplified at 70°C, while GI5 was amplified at 54°C. A major portion of this PICFV8 codes for the Type Four Secretion System (T4SS) and it is located uniquely within the bacterial chromosome of the subspecies CFV [59]. G1 primers (VirB9-1 and CFVvirB11F), bind a locus specific for *virB9/virB11* (712bp) that code for T4SS components; G2 primers (TaxB3 and nc1165g6F) amplify the *virD4* gene (468bp) that codes for a T4SS component; G3 primers (3DV3' and 3DV5') amplify the *ORF21* (262bp) that codes for an apparent plasmid origin; G4 primers (Top/Tra#4 and TraE7) amplify the *ORF27* (841bp) that codes for an apparent plasmid origin; while G5 primers (VenSF and VenSR) target the *parA* gene (142bp) that codes for an apparent plasmid origin. Table 13 summarizes the PCR results using the five different primer sets (Table I in Appendix I). All the selected isolates exhibit differences from the reference ATCC strains and from many other strains, especially those of European origin, which had been investigated by Gorkiewicz and colleagues in 2010 (Table 12). Interestingly, the four CFV isolates all retained the *ven* sequence (*parA gene)* (Figure 11), which was previously identified as a potential target to discriminate CFF and CFV (see section 1.2.9), but some CFV isolates lack other regions of this GI.

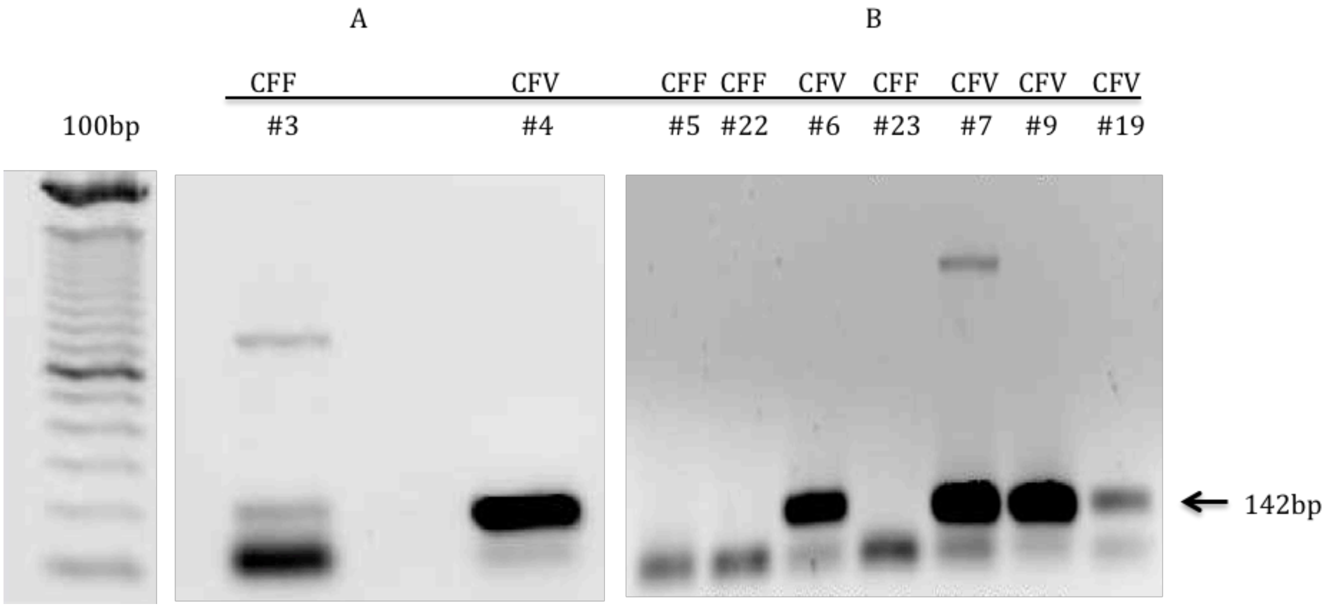


Figure 11. Agarose gel picture of PI G5 (ven) amplicons generated from CF genomic DNAs. In panel A, lanes 1 and 2 represent the products of ATCC CFF/CFV reference strains used as negative and positive controls respectively (Acc. #3 & 4 in Table 2). In panel B, lanes 1, 2 and 4 represent the results using CFF isolates (Acc#5, 22, and 23 as in Table3), while lanes 3, 5 – 7 show the amplicons generated by CFV strains (Acc. # 6, 7, 9 and 19 as in Table 2). The image of the DNA bands was obtained using UV transIlluminator Gel Doc system. A 100bp DNA molecular marker, shown on the left of the gel, confirms the size of the 142bp amplicon.

Table 12. Table summarizing PCR results to test for the presence of the genomic island PICFV8. Five distinct regions of the GI were targeted by PCR as illustrated in Figure 6 uniquely present in all CFV, not the CFF strains.

Amplicon #	Expected amplicon size (bp	Reference Strains		Selected isolates						
		3 CFF	**4** CFV	**5** CFF	**22** CFF	**23** CFF	**6** CFV	**7** CFV	**9** CFV	**19** CFVi
GI 1	712	NEG	POS	NEG	NEG	NEG	NEG	NEG	NEG	POS
GI 2	468	NEG	POS	NEG	NEG	NEG	NEG	NEG	NEG	POS
GI 3	262	NEG	POS	NEG	NEG	NEG	NEG	NEG	NEG	POS
GI 4	841	NEG	POS	NEG	NEG	NEG	NEG	NEG	NEG	POS
GI 5	142	NEG	POS	NEG	NEG	NEG	POS	POS	POS	POS

*Genomic Island, POS: Positive, NEG: Negative.

3.4 Determination of the sap genotype of CFF strains by PCR

CF can be designated as serotype A or B based on the presence of proteins encoded by *sapA* or *sapB* genes (see section 1.2.4.1). Since all CFV strains are serotype A whereas CFF strains can be either type A or B, or rarely AB, use of a serotype A CFF for the subtractive hybridization technique would be preferred. Accordingly the genotype of all CFF strains was determined by PCR using primer pairs SAF01-SAR01 (for *sapA*) and SBF01-SBR01 for (*sapB*). Expected amplicon sizes are 531bp for (*sapA*) and 505bp (for s*apB*). Figure 12 shows the validation of the primers for typing using the ATCC reference strains, while Table 13 summarizes the PCR results of the three CFF test strains which all have the *sapA* gene.

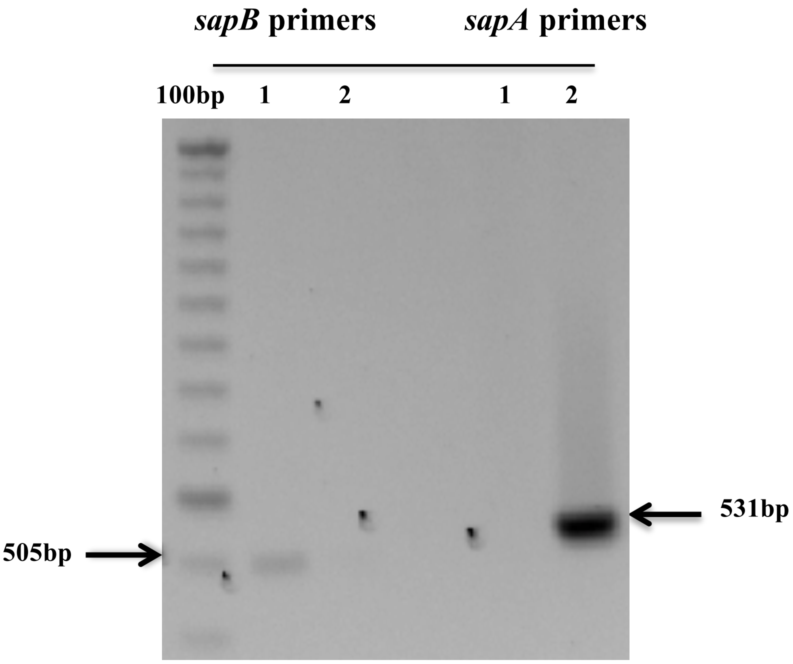


Figure 12. Validation of PCR for serotype testing. Primers SAF01/SAR01 were used for amplifying *sapA* and primers SBF01/SBR01 for the *sapB* gene. Lane 1 represents the ATCC strain CFF27374 and lane 2 is the ATCC CFV19438. The expected sizes are 531bp (for *sapA*) and 505bp (for *sapB*). DNA ladder is on the left side of the gel.

Table 13. Serotype of the CFF strains as determined using primers specific for serotype A (SAF01 and SAR01) and serotype B (SBF01 and SBR01).

Amplicon #	Expected amplicon size (bp)	Name of the organism		
		5 CFF	22 CFF	23 CFF
SAF01-SAR01	531	POS	POS	POS
SBF01-SBR01	505	NEG	NEG	NEG

* CFF isolates (Acc#5, 22, and 23 as in Table3). POS: positive, NEG: negative.

3.5 Suppression Subtractive Hybridization between *C. fetus fetus* and *C. fetus venerealis*

Based on the biochemical and genetic characteristics of the seven CF isolates examined in detail, genomic DNAs from two isolates were selected for SSH analysis. The Canadian isolate CFV08A1102-42A (Acc. #6, Table 2), which originated from an outbreak in Alberta, was specifically chosen as the tester because it lacked much of the GI PICFV8 thought to define the CFV subspecies. The typical Canadian isolate CFF02A725-35A (Acc. #5, Table 2) was selected as driver. The SSH workflow was followed using these two CF isolates together with an *E. coli* genomic DNA control provided with the kit. This control contains 1 copy of Hae III-digested φX174 DNA per *E. coli* genome and its use in parallel with the DNA of interest allows evaluation of the success of each stage in the SSH procedure.

3.5.1 Testing the Efficiency of CFF-RsaI and CFV-RsaI Digestion

Complete digestion of tester and driver DNAs with RsaI is necessary for an efficient SSH. Cleaving genomic DNA into multiple small fragments with defined ends allows attachment of adaptors for subsequent amplification in which only hybrid molecules are efficiently recovered [2,37,146]. The efficiency of the digested tester and driver (CFV and CFF) was confirmed and compared to the digested *E.coli* DNA control by running 5 µl (0.2µg) of the digested products, along with 0.2 µg undigested DNA side by side, on a 1.0 % agarose-EtBr gel. The undigested genomic DNA appears as a high-molecular weight band at the top of the gel, while Rsa-I digested DNA shows drastic decrease in size and appears as a smear from 0.1 to 2 kb. The results of the digested experimental and control DNAs are shown in Figure 13.

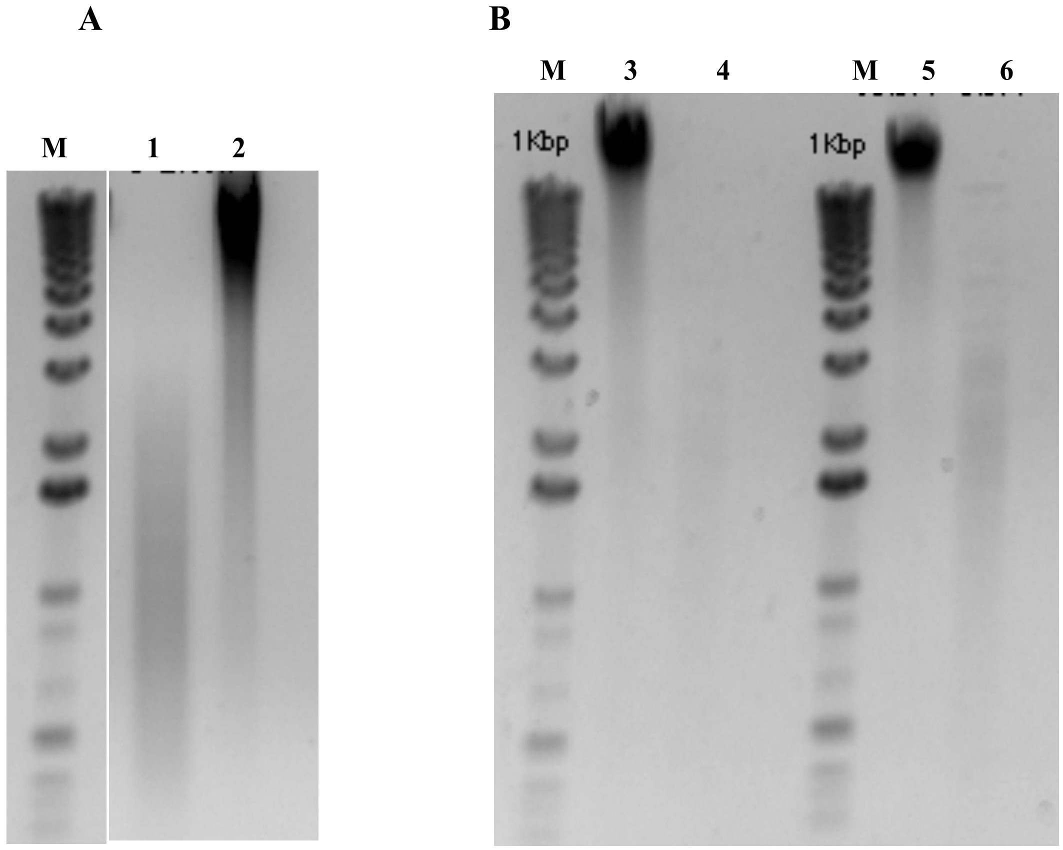


Figure13. Agarose gel image representing the efficiency of genomic DNA digestion using RsaI restriction enzyme of both tester and driver of CF, as well as the control DNA. Lane M is 1 kb DNA marker, lanes 1 and 2 of panel A are the *E. coli* genome digested and undigested DNA, respectively. Lanes 3 and 4 of panel B show the undigested and digested CFF *(*driver DNA), while lanes 5 and 6 of the same panel represent the undigested and digested CFV (tester DNA).

3.5.2 Testing the Efficiency of CFV DNA-Adaptors Ligation

The digested tester DNA was split into two portions and each was ligated to a double stranded adaptor (Adp1 and Adp2R). The end of each adaptor is not phosphorylated; therefore only one strand of both adaptors would covalently bind to the 5' end of the digested tester [2,37,146]. In addition, Adp1 and Adp2R contain stretches of identical nucleotide sequences at both ends, which facilitate the annealing of PCR primer P1 after the recessed ends have been filled in [2,37,146]. Therefore, they provide a means of selectively enriching target sequences [2,163]. The efficiency of the adaptor ligations were tested by determining if there was a detectible shift in band size when ligated DNAs were amplified using particular primer pairs as described in Table 7. The reactions in tubes 2 and 4 contain primers against a specific gene which will amplify whether or not the ligation worked while those of tubes 1 and 3 include one primer against the adaptor and one against a specific bacterial gene that will only amplify if the ligation worked. The ratio of band intensities between tubes 1 & 2 and between 3 & 4 indicates the ligation efficiency; a minimum 25% of the control-tester and CFV-tester DNAs should have the two adaptors, 1 and 2R, on both ends of the DNA fragments.

During this experiment, each ligated tester DNA (as in Figure 9: tester 1-1, 1-2, 2-1, and 2-2) was diluted and subjected to PCR using primer P1, the sequence of which is included within both adaptors, and a primer targeting a specific gene sequence in the tester itself. This was a *ven* gene primer for the CFV and a primer targeting the 23S rRNA gene in the case of the *E. coli* control. Importantly, the primers should amplify DNA fragments that do not contain an RsaI site; absence of any RsaI site within the GI5 amplicon was confirmed by analysis of the amplicon sequence using the NEB cutter tool. Five microliters of each

PCR product were analyzed and examined by electrophoresis on a 2% agarose/EtBr gel (Figure 14).

Part A of the figure represents the *E. coli* DNA control. Lanes 1 and 2 employed tester 2-1 (Adaptor 1 ligated DNA) as the template. PCR was performed with 23S rRNA Forward primer and PCR Primer 1 (lane 1) and 23S rRNA Forward and Reverse primers (lane 2). Lanes 3 and 4 used tester 2-2 (Adaptor 2R ligated DNA) as the template with the primer sets 23S rRNA Forward primer and PCR Primer 1 (lane 3) and 23S rRNA Forward and Reverse primers (lane 4). The 23S rRNA primers amplify a 270 bp region of the 23S rRNA gene of *E. coli*, while the PCR primer 1 /rRNA primers should amplify a 374bp product providing the adaptor ligation was successful.

Part B of the figure represents the experimental tester CFV DNA. Lanes 1 and 2 employed tester 1-1 (Adaptor 1 ligated DNA) as the template; PCR was performed with VenS Forward Primer and PCR Primer 1 (lane 1) and VenS Forward and Reverse primers (lane 2). Lanes 3 and 4 employed tester 1-2 (Adaptor 2R ligated DNA) as the template; PCR was performed using VenS Forward primer and PCR Primer 1 (lane 3) and VenS Forward and Reverse primers (lane 4). As expected the VenS primers amplified a 142 bp region of the Ven gene (GI5 PCR) of CFV, while the amplicons generated by the PCR Primer 1 / Ven primers were larger. The detectible band size shifts in lanes 1 and 3 compared to lanes 2 and 4 of each panel indicate successful ligation.

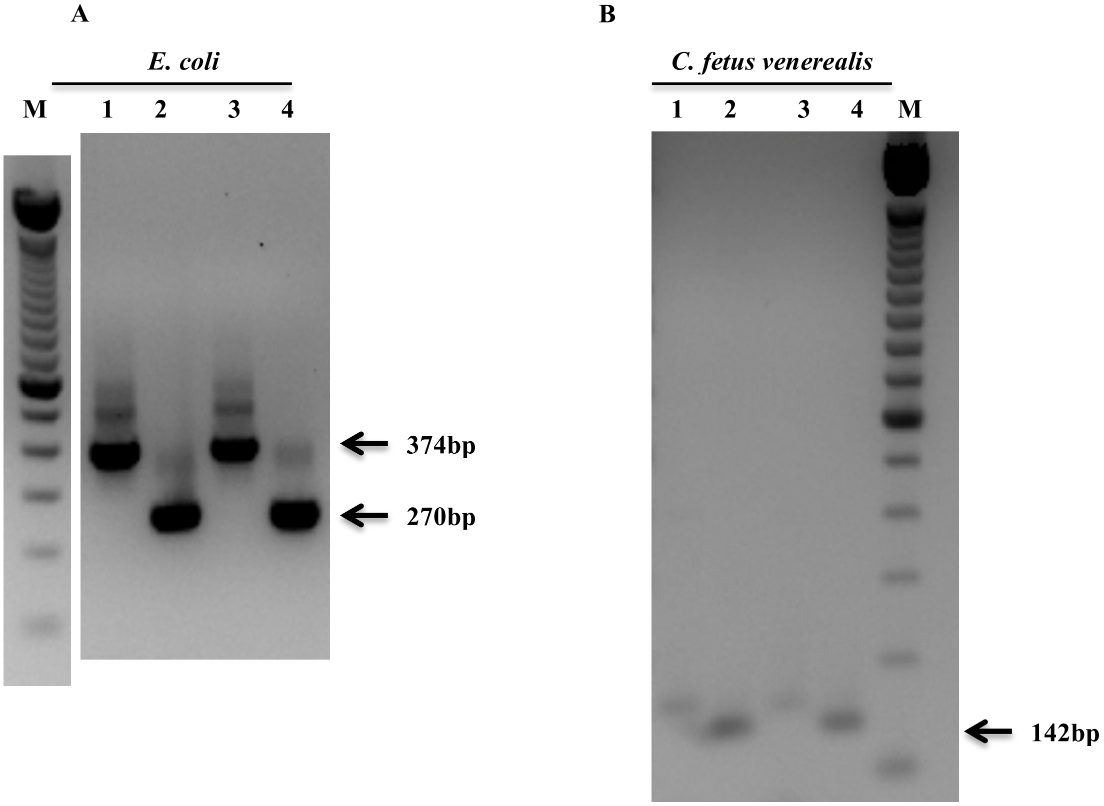


Figure 14. Typical results of ligation efficiency analysis for the control *E. coli*-tester DNA (panel A) and the experimental CFV-tester DNA (panel B). PCRs were analysed by electrophoresis on a 2% agarose gel. In both figures, lane M is the 100bp DNA marker; lanes 1 & 2: PCR products obtained using adaptor 1 ligated tester 1-1 (panel A) and 2-1 (panel B) as the template. Lanes 3 & 4: PCR products obtained using adaptor 2R ligated tester 1-2 (panel A) and 2-2 (panel B) as the template. Description of the primers used is provided in the text.

3.5.3 Analysis of PCR Products Following Subtractive Hybridization

Tester-adaptor (control and experimental) DNAs obtained from the ligation reactions were hybridized with digested driver DNA. After 2 rounds of subtractive hybridization, the hybridized products were PCR amplified to enrich any unique tester sequences. Then, PCR products were subjected to nested-PCR, to try to remove any background and non-specific sequences. The nested primer sequences (NP1 and NP2R) are present within the respective adaptors but internal to the P1 primer; production of amplicons using this primer pair would indicate generation of hybrid sequences from fragments flanked by the adaptors.

Following two rounds of amplification, the subtracted and unsubtracted libraries obtained from primary PCR and secondary PCR were run side-by-side with the PCR Control Subtracted DNA by 2% agarose/EtBr gel electrophoresis to check the efficiency of the subtraction hybridization. The subtracted library contains subtracted tester/tester hybrids, while unsubtracted library includes the diluted genomic DNA ligated with both Adaptors 1 and 2R. PCRs were performed on hybridizations products under slightly different annealing conditions to identify the optimal hybridization temperature; and it was found that the subtracted library with 63°C hybridization temperature (Lane 6 of Figure15) was the best and thus chosen for the subsequent analyses. The resulting bands of the subtracted control DNA (Figure 15A) were similar to the PCR control subtracted DNA, while the subtracted experimental (CFV) DNA (Figure 15B) appeared as a smear with a size from 0.8 – 5 kb after primary PCR, but some additional distinct bands were apparent after secondary (nested) PCR. Differences in the band patterns between the unsubtracted and subtracted libraries indicated successful subtraction.

The products from the secondary PCR were then subjected to Quanta RT-PCR to check the integrity of the subtracted libraries, and the subtracted CFV library included the *ven* sequence as expected, with a CT around 20, following the amplification with *venS* primers and *venS* probe (data not shown).

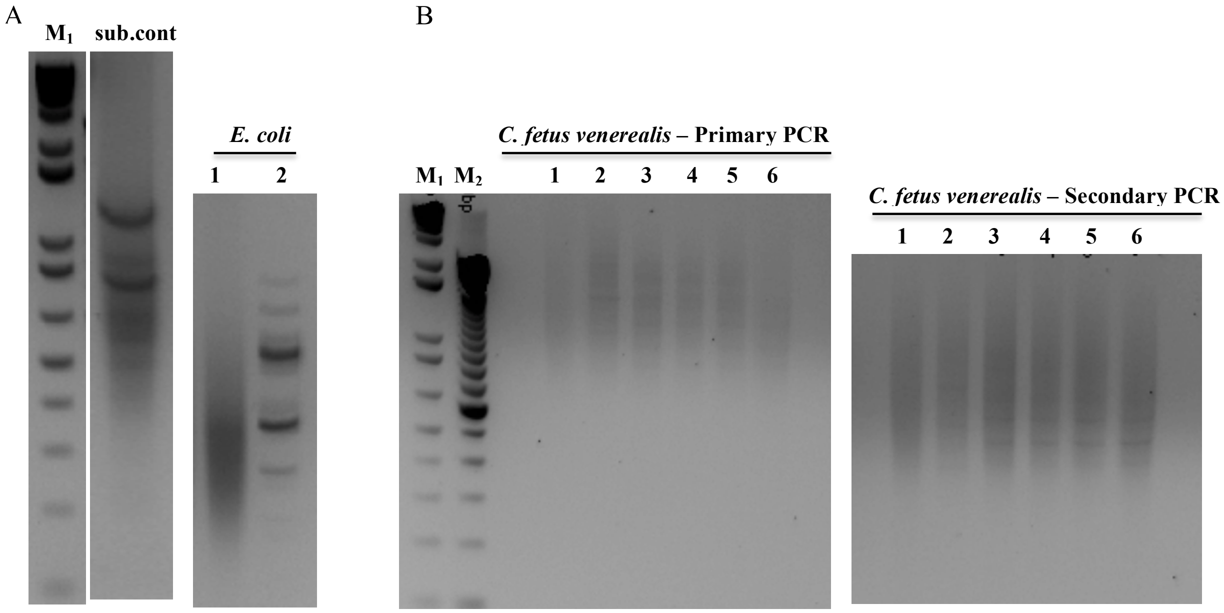


Figure 15. Gel image showing typical results of PCR control, subtracted and unsubtracted DNA libraries following two rounds of PCR amplification. Results were run on a 2.0% agarose/EtBr gel for both control DNA (panel A) and experimental DNA (panel B). In panel A: lane M is φX174/*Hae* III DNA size markers, lane sub.cont: is the PCR control subtracted DNA, while lane 1: unsubtracted *E. coli* DNA and lane 2: subtracted tester *E. coli* genomic DNA. In panel B: lanes M_1 & M_2 are the 1kb plus and 100bp DNA markers, respectively; lanes 1 & 2 are the unsubtracted CFV library with 2 sets of dilution, 1:10 and 1:100; while lanes 3 – 6 represent the subtracted CFV library generated with different hybridization temperatures (60°C, 61°C, 62°C, and 63°C).

3.6 Colony PCR and Slot blot screening

The secondary PCR products from the subtractive hybridization (Lane 6 from Figure 15) were inserted by T/A cloning into the PCR^{TM} 2.1 - TOPO vector to generate a subtracted CFV-DNA library. From the clones thus generated, only the white colonies were subjected to colony PCR using M13 Forward / Reverse primers to confirm the presence of an insert. From this analysis, 50 clones were confirmed to have inserts ranging in size between 0.5kb – 2kb (Figure 16). These clones were selected for further characterization using Slot blotting. Probes corresponding to the cloned insert sequences were generated by random DIG labeling and used in slot blot analysis for their ability to hybridize to genomic DNA from both the CFF and CFV isolates used for the SSH procedure (Table 14). Thirty-eight clones were either CFV specific or gave higher signal with CFV than with CFF.

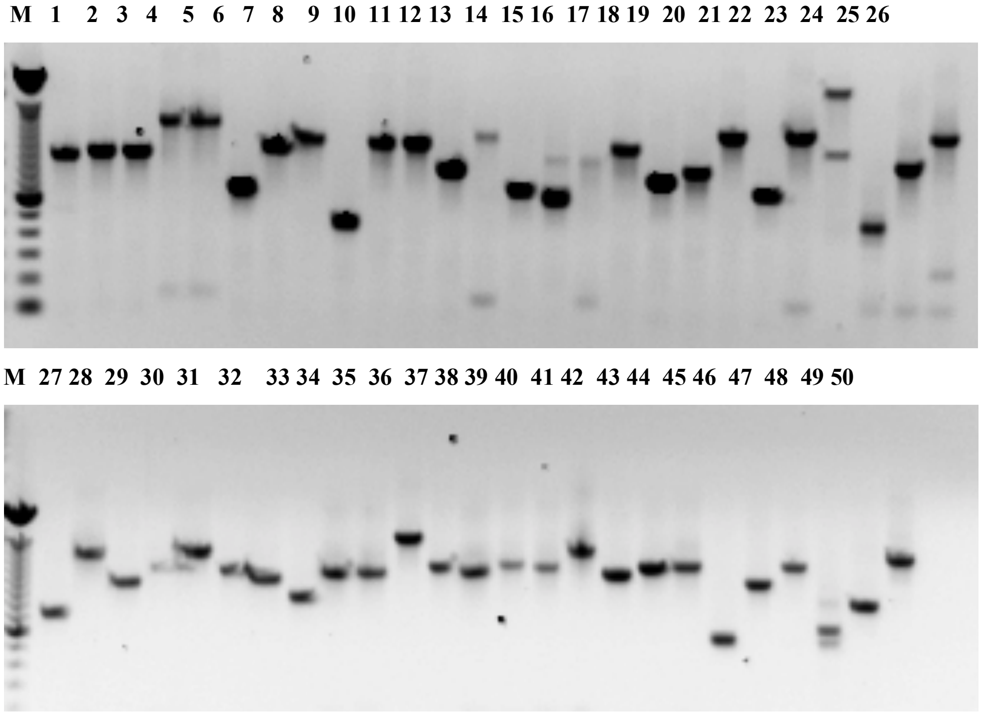


Figure 16. Gel image illustrating insert sizes of clones recovered from the SSH procedure. Colony PCR using universal M13 forward / reverse primers was performed on 50 clones. The sizes of the cloned inserts (lanes 1 – 50) ranges between 500bp – 2.0 kb. Lane M is 100bp molecular weight DNA marker.

Table 14. Slot blot hybridization screening of the 50 clones generated from the subtractive hybridization against genomic CFF/CFV. Positive results were obtained based on visual observation of the hybridization signals.

	Hybridization against			Hybridization against	
Clone#	**CFV.42A**	**CFF.35A**	**Clone#**	**CFV.42A**	**CFF.35A**
1*	+	-	**26***	+	w/+
2*	+	w/+	**27**	+	+
3*	+	w/+	**28***	+	-
4*	+	w/+	**29**	+	+
5	+	+	**30***	+	w/+
6	+	+	**31**	+	+
7*	+	w/+	**32***	+	-
8*	+	w/+	**33**	+	+
9*	+	w/+	**34***	+	-
10*	+	w/+	**35***	+	-
11	+	+	**36***	+	-
12*	+	w/+	**37***	+	w/+
13*	+	w/+	**38***	+	w/+
14	+	+	**39***	+	w/+
15*	+	w/+	**40***	+	w/+
16	+	+	**41***	+	w/+
17	+	+	**42***	+	-
18*	+	w/+	**43***	+	w/+
19*	+	w/+	**44***	+	-
20*	+	w/+	**45***	+	w/+
21*	+	w/+	**46***	+	-
22	+	+	**47***	+	-
23*	+	w/+	**48***	+	w/+
24*	+	w/+	**49***	+	w/+
25*	+	w/+	**50**	+	+

*indicates clones that were selected for further sequencing analysis. w = weak positive.

3.7 Clonal analysis by DNA Sequencing

The inserts of these 38 clones were further characterized by Sanger sequencing using M13 Forward and Reverse primers to generate bidirectional reads, which were combined to generate consensus sequences for each clone insert. Clones containing inserts with the same sequences were recognized using sequence alignment ClustalW software and excluded from the subsequent analysis. It was shown that clones 1 and 2 were identical; the same applied to clones 4 and 14; clones 7 and 9; clones 10 and 17; as well as clones 28, 32, and 46. Thus, clones 1, 9, 14, 17, 32, and 46 were excluded from the analysis.

The identity of the remaining 34 sequences was examined by BLAST analysis of the GenBank sequence database (NCBI website). This search revealed 26/34 cloned sequences had 100% nucleotide homology with the genome of CFF reference strain CFF82-40 (Table I of Appendix 4). On the other hand, 8/34 cloned sequences did not show a good match over a significant portion of their length to the CFF82-40, but did with *Campylobacter* species other than *fetus* (Table I of Appendix 4). However, there was no match to any CFV sequences in the database at the time when the analyses were done. Sequencing results were compared to the probe/DNA hybridization screenings (Table 1I of Appendix 4) to validate the obtained results. The nucleotide sequences of the 8 clones, after removal of the RsaI and adaptor sequences, were subjected to further analysis. Probes of these clones were used to screen genomes of 6 CFV DNAs (Acc. # 7, 9, 11, 12, 13, and 20 as in Table 2) to explore the extent to which these sequences were retained and to assess how specific these sequences are for CFV (Table 15). A carbon starvation (*cstA*) gene probe was included as a positive control since it is present in all CF strains. Based on the presence/absence of the selected cloned sequences among CFV isolates, another 5 clones which did not hybridize to all CFVs were excluded from the subsequent analyses while sequences of the remaining 3 clones (#2, 28, and 35) were present in all the selected CFVs. The nucleotide sequences of the 3 clones, with RsaI and adaptor sequences removed, are in Appendix 5. Further analysis by screening against 2 CFF DNAs (Acc. # 22 and 23 as in Table 2) confirmed that the 3 clones (#2, 28, and 35) appeared to be highly CFV-specific since they hybridized with the CFV genomes but not with the CFFs (see Table 15).

Table15. DIG-labelled probes were generated from the inserts of the eight selected clones by DIG-High Prime DNA Labeling and Detection Starter Kit, Roche, Cat#11585614910 and used to hybridize against blots (slot and Southern) of genomic DNA from several selected CFF and CFV isolates.

		Probe of Clone								
Strain Acc#	**Strain type**	**CstA**	**#2**	**#3**	**#28**	**#34**	**#35**	**#36**	**#38**	**#47**
5	CFF (SSH)	POS	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
6	CFV (SHH)	POS	POS	POS	POS	POS	POS	POS	POS	POS
7	CFV	POS	POS	POS	POS	POS	POS	POS	POS	POS
9	CFV	POS	POS	NEG	POS	NEG	POS	NEG	NEG	NEG
11	CFV	POS	POS	NEG	POS	NEG	POS	NEG	NEG	NEG
12	CFV	POS	POS	POS	POS	POS	POS	NEG	POS	POS
13	CFV	POS	POS	POS	POS	POS	POS	NEG	POS	POS
20	CFVi	POS	POS	N/A	POS	N/A	POS	N/A	N/A	N/A
22	CFF	POS	POS	N/A	POS	N/A	POS	N/A	N/A	N/A
23	CFF	POS	NEG	N/A	NEG	N/A	NEG	N/A	N/A	N/A

* *cstA*: Carbon starvation gene probe was used as a positive control, POS: positive, NEG: negative, SSH: suppression subtractive hybridization, N/A: non-available (no screening done for this isolate).

3.8 Bioinformatics analyses of the three SSH-clones

The 3 clones identified from the CFV-subtracted library were subjected to further analyses to explore their nature. First, the clonal sequences were aligned against genomic DNA of the CFF82-40 and CFV NCTC10354 reference strains, as reported by Ali el (2012) [3] by using DNASTAR Lasergene® software: SeqBuilder and Megalign (DNASTAR Inc., Madison, WI); and from the alignment analysis all 3 clones perfectly matched the CFV reference genome sequence but not that of the CFF (Table 16). BLAST analysis was done to locate the sequences of the 3 selected clones within the reference CFV genome (NCTC10354) and it was discovered that all three mapped within the pathogenicity island #10 (PICFV10); the ORFs of this putative pathogenicity island all code for hypothetical proteins of no known function (Table 17).

Table 16. BLAST analysis of the 3 cloned sequences against CFV NCTC10354 and CFF82-40. Results obtained from NCBI.

Alignment with CFV NCTC 10354 (GenBank Accession # CM001228.1)

Clone#	Length (bp)	Max score	Query coverage	E-value	Max indent.	Gaps	Strands	Query Start	Query End	Subject Start	Subject End
2	749	1310	100%	0	99%	0	+/+	1	749	1495002	1494254
28	764	951	100%	0	100%	0	+/-	1	764	1483085	1483617
35	825	1353	95%	0	100%	0	+/-	1	825	1484066	1484820

Alignment with CFF 82-40 (GenBank Accession # CP000487.1)

Clone#	Length (bp)	Max score	Query coverage	E-value	Max indent.
2	764	44.4	41%	0.001	100%
28	764	39.2	33%	0.0062	100%
35	825	86	18%	5E-16	100%

Table 17. BLAST analysis of the 3 cloned sequences to PICFV10 of NCTC10354. Results obtained from NCBI.

Alignment with NCTC10354 PICFV10

Clone#	Alignment start	Alignment End	Correspond ORF	Correspond gene
2	1495002	1494254	N/A	hypothetical protein
28	1483085	1483617	PI10_1550 - PI10_1552	hypothetical protein
35	1484066	1484820	PI10_1553	hypothetical protein

3.8.1 Analysis of Sequencing Data – BLAST Analysis of 3 SSH Subtracted Clones Against De Novo Assembled Genomes

Once the results from the whole genome sequencing were available, the presence of sequences for the 3 clones obtained from the subtractive hybridization (Clones #2, 28, and 35) was checked and confirmed by BLAST comparison of their sequences with both reference assembled and *de novo* assembled CFV08A1102-42A (Acc. #6, Table 2) and CFF02A725-35A (Acc. #5, Table 2). Surprisingly, the full length sequence of SSH_Clone #2 was found within both CFF and CFV, and it aligned with contig #5 in the *de novo* assembled CFF02A725-35A but was absent from the reference assembly; while the majority of SSH_Clone #28, (with only a few gaps ~ 300bp) was found in contig #5 of *de novo* assembled CFF. On the other hand, SSH_Clone #35 was uniquely present in both reference and *de novo* assembled CFV08A1102-42A and absent from CFF02A725-35A (Table 18). Consequently, the specificity of SSH_clone35 was tested by BLAST analysis of the clonal sequence with *de novo* assembled genomes from 8 CF isolates; it was present in all CFVs and absent from all CFFs, except CFF ADRI1362.8A (Table 19). Furthermore, Mega5 software was used to confirm the nucleotide alignment of the clonal sequence 35 and the 8 CF genomes listed in Table 20 (data not shown).

Table 18. BLAST of SSH_Clones # 2, 28, and 35 with *de novo* assembled CFV/CFF isolate used for subtractive hybridization.

		***De novo* assembled**		**Reference assembled**	
Clone#	**Length (bp)**	**BLAST with CFF 02A725-35A**	**BLAST with CFV 08A1102-42A**	**BLAST with CFF 02A725-35A**	**BLAST with CFV 08A1102-42A**
2	764	Yes, with contig 5	Yes	NO	Yes
28	825	Yes, with contig 5	Yes	NO	Yes
35	764	NO	Yes	NO	Yes

Table 19. BLAST of SSH_Clone35 with *de novo* assembled Canadian CFV/CFF isolates.

Assigned accession #

Clone	5	6	7	8	20	22	24	25
SSH_Clone # 35	BLAST with contig sequences CFF 02A2725.35A (paird) de novo assembly	BLAST with contig sequences CFV 08A1102.42A (paird) de novo assembly	BLAST with contig sequences CFV 08A948.2A (paird) de novo assembly	BLAST with contig sequences CFV 08A110233B.5A (paird) de novo assembly	BLAST with contig sequences CFVi ADRI 545.1A (paird) de novo assembly	BLAST with contig sequences CFF ADRI 1362.8A (paird) de novo assembly	BLAST with contig sequences CFF 09A9803.2B (paird) de novo assembly	BLAST with contig sequences CFFTurbo serotypeB (paird) de novo assembly
	No	yes, with contig_3	Yes, with contig_36	Yes, with contig_36	Yes, with contig_44 & 60	yes, with contig_33 & 46	No	NO

3.9 Whole Genome Sequencing and Optical Mapping Data Evaluation

3.9.1 Analysis of Sequencing Data – Reference and *De Novo* Assemblies

Whole genome sequencing (WGS) is an alternative approach used during the project to assess the genetic differences between CFF and CFV. Genomic DNA was prepared from the two *C. fetus* isolates used previously in the subtractive hybridization approach, Acc. # 5 & 6 of Table 2 (CFF02A725-35A and CFV08A1102-42A*)*, and six other *C. fetus* isolates (Acc. # 7, 8, 20, 22, 24, and 25 in Table 2). After checking the genomic integrity, concentration, and purity these samples were sent to Health Canada where they were batched with additional samples and submitted to the BC genome centre for whole genome sequencing performed using an Illumina MiSeq - next generation sequencing system. Two FASTQ files (representing 100 bp paired-end reads for fragments of ~500 bp) were generated for each isolate and provided to us for the subsequent assembly. The sequencing reads, with up to 250x coverage, were assembled by two different approaches: Reference assisted and *de novo.* The subsequent analyses were focused on the two CF isolates used for the SSH.

In the reference-assisted assembly, CF reference genomes (CFF82-40 and CFV NCTC10354) were used to assemble the generated reads using Burrows-Wheeler Alignment tool (BWA). The length of the produced genomes was 1,753,992-bp for the CFF02A725-35A and 1,824,368 bp for the CFV08A1102-42A. A *de novo* assembly, generated using CLC Genomics Workbench software (version 6.0), in which no preliminary information about the genome is required was also performed. The number of the contigs produced from the *de novo* assembly was 472 contigs in the CFV and 105 contigs in the CFF. The sequencing outputs are presented in Table 20.

Table 20. Summary of the *de novo* assembly of Canadian CFF02A725-35A (Acc. #5) and CFV08A1102-42A (Acc. #6). (A) Nucleotide distribution among the genome, and (B) Contig measurements.

A

Nucleotide distribution – CFV.42A de novo assembly

Nucleotide	Count	Frequency
Adenine (A)	660,769	33.70%
Cytosine (C)	326,584	16.60%
Guanine (G)	322,326	16.40%
Thymine (T)	650,917	33.20%
Any nucleotide (N)	1,958	0.10%

Nucleotide distribution - CFF.35A de novo assembly

Nucleotide	Count	Frequency
Adenine (A)	680,619	33.50%
Cytosine (C)	338,533	16.70%
Guanine (G)	331,681	16.30%
Thymine (T)	677,322	33.30%
Any nucleotide (N)	4,579	0.20%

B

Contig measurements - CFV.42A de novo assembly

Length	
N75	56,417
N50	177,121
N25	279,172
Minimum	68
Maximum	404,775
Average	4,158
Count	47
total	1,962,554

Contig measurements - CFF.35A de novo assembly

Length	
N75	53,337
N50	109,169
N25	38,168
Minimum	71
Maximum	280,978
Average	1,296
Count	1,569
total	2,032,734

3.9.2. Analysis of Sequencing Data – Optical Mapping

While awaiting results of the WGS, DNA aliquots of the same samples used for the SSH, CFF02A725-35A (Acc. #5, Table 2) and CFV08A1102-42A (Acc. #6, Table 2), were submitted to CFIA Lethbridge Laboratory for optical mapping. All the optical maps were generated on the basis of restriction digestion by NcoI enzyme and provided as XML files. Additional in *silico* maps predicted from the nucleotide sequence data of the same isolates and from the reference genomes were also created for comparative purposes and analyzed using the Argus^{TM} OpGen software (Figures 18 and 19). In these figures the blue shaded area indicates that there is a good match of restriction patterns between the strains, the red area indicates the presence of multiple alignment sites, while the white region represents poorly matched areas. Vertical lines represent matching NcoI restriction cut sites so that the regions between cut sites illustrate estimated fragment sizes along the genome. The dark lines (thicker bands) represent higher density of cut sites in that locus, while the faint lines (thinner bands) indicate lower density of cut sites in that locus.

The first comparison was to align the NcoI-genome maps of the isolates used for SSH with the in *silico* NcoI maps of the reference strain genome sequences (Figures 17A & 17B). From the figures, we can conclude that for both the CFV (Figure 17A) and CFF (Figure 17B) comparisons, when the Canadian isolates are compared with the reference genomes there are regions of high and low homology. In particular, both Canadian isolates are different from the reference strains with respect to the genome length and in regions where particular putative pathogenicity islands are located. The map of CFF02A725-35A is distinctive from reference strain 82-40 in the region corresponding to the 2 PIs PICFF6 and 7. The map of CFV08A1102-42A is different from that of the CFV reference in regions that contain several

PIs including PICFVs 5, 6, 8, 9 and 10. Particularly notable is the difference in the maps over the region encoding PICFV8 as predicted at the start of the project by PCR screening for this GI. These data are concordant with the project hypothesis in which the Canadian and the reference strains differ in regards to the presence/absence of pathogenicity islands.

Another comparison was done to compare the NcoI-genome maps of the SSH isolates with those derived *in silico* using both reference-based and *de novo* assembled sequences of the same isolates. These NcoI-genome maps were used to guide and validate the genome sequence assembly of the selected isolates [204]. Based on these maps it is clear that differences in the NcoI restriction site patterns are most evident in the regions where the Canadian isolates differ most from the reference sequences. This suggests a bias in the reference-assisted assemblies towards the template (Figure 18A & 18B). On the other hand, the genome coverage of the contigs recovered from the *de novo* assembly (472 contigs from CFV08A1102-42A and 105 contigs from CFF02A725-35A) was 72.1% in the case of CFV, while only 35.7% in the case of CFF (data not shown). Therefore it can be concluded that both assemblies were needed for proper alignment and better coverage of the CF genomes.

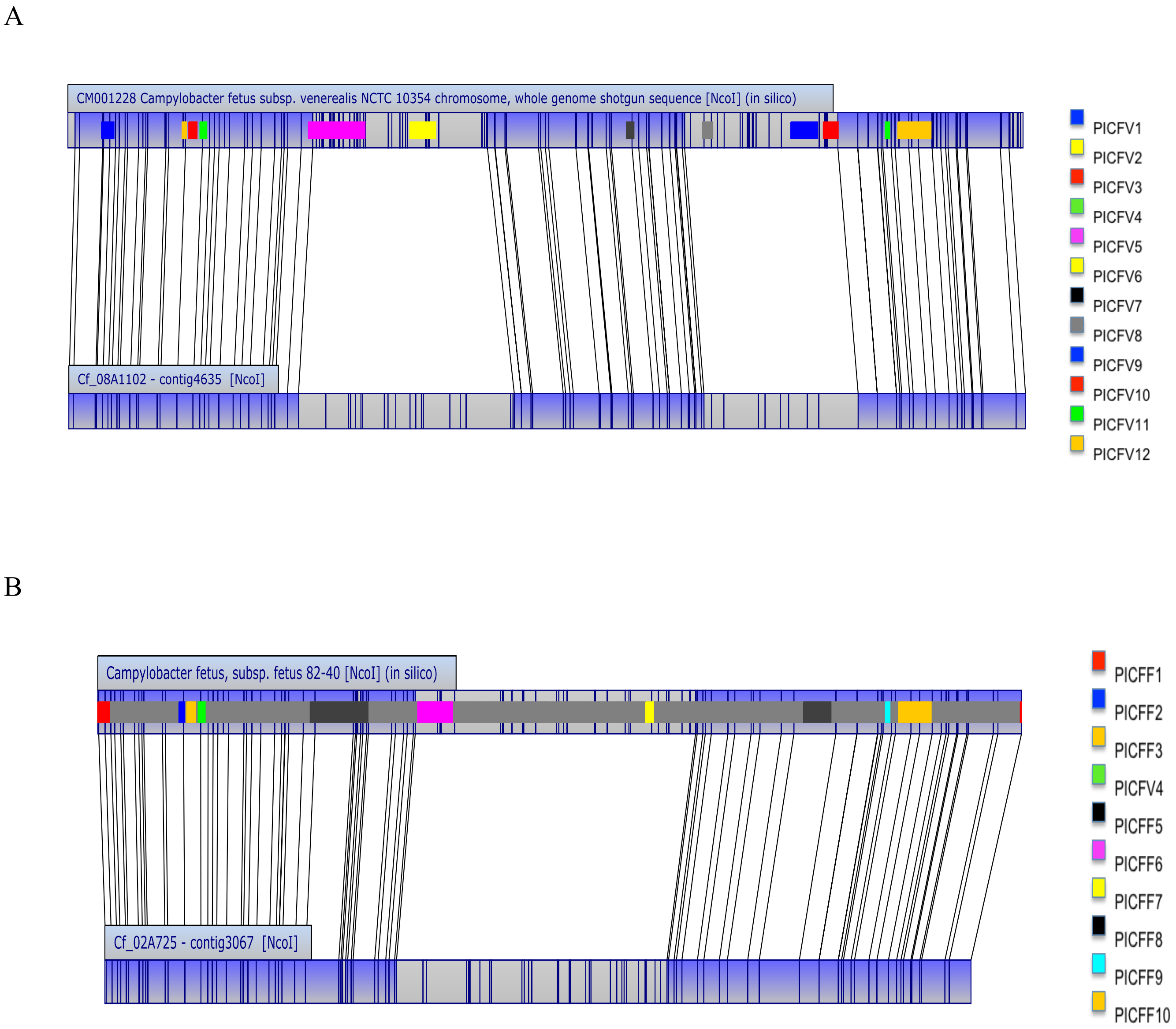


Figure 17. Comparison of optical maps generated *in silico* from CF reference genome sequences with genomic maps of the Canadian CF isolates used for SSH. The *in silico* map determined from the sequence of the reference genome CFV NCTC10354 (GenBank CM_001228) is compared to the genomic map of the Canadian isolate CFV08A1102-42A (Panel A); the in *silico* map of the reference genome CFF82-40 (GenBank NC_008599) is compared to the genomic optical map of the Canadian isolate CFF02A725-35A (Panel B). The locations of the putative pathogenicity islands are indicated in the colored boxes. Maps generated by MapSolver software, OpGen. The grey bar across the reference genome (CFF82-40) indicates complete annotation is available for this sequence in GenBank.

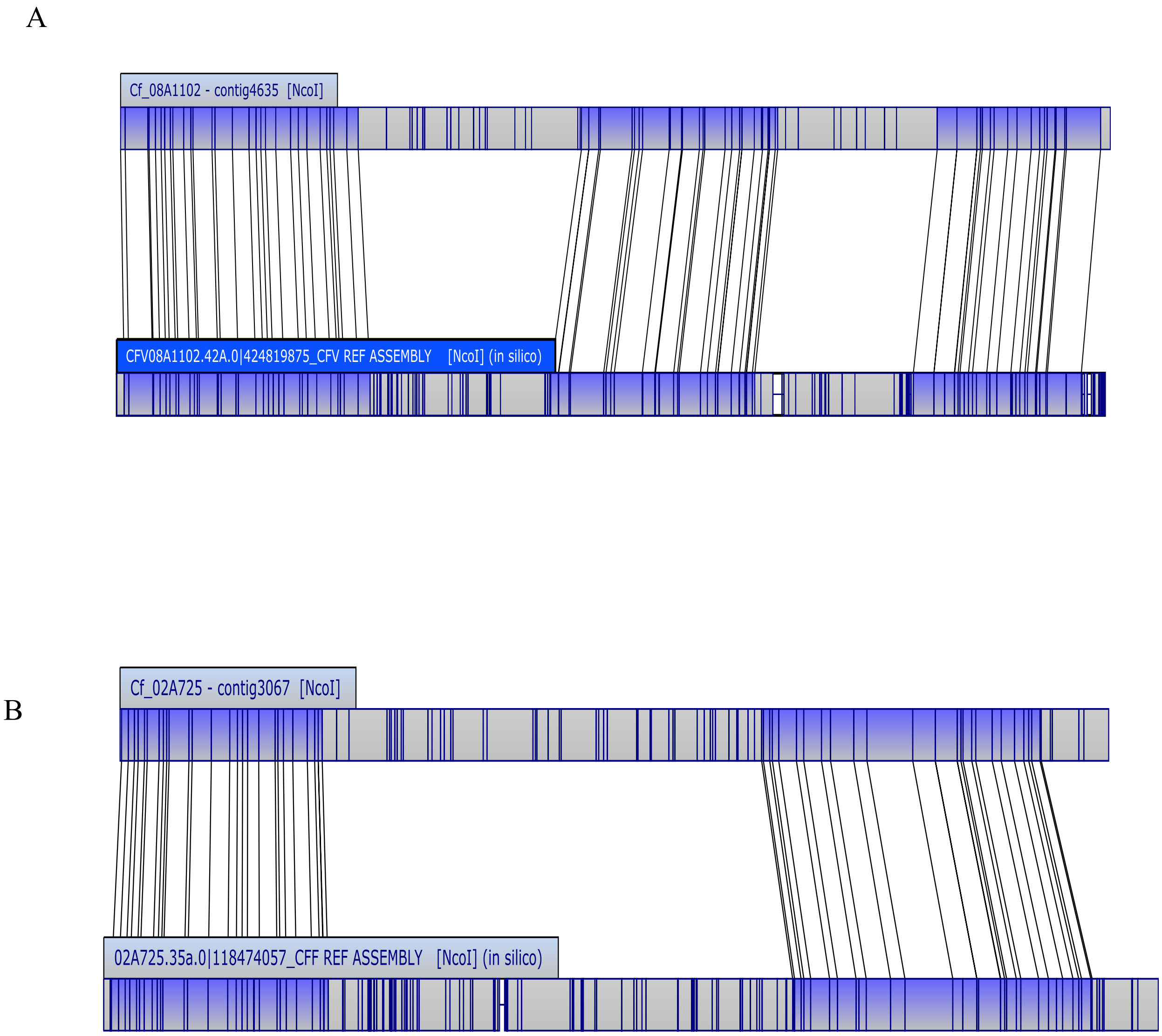
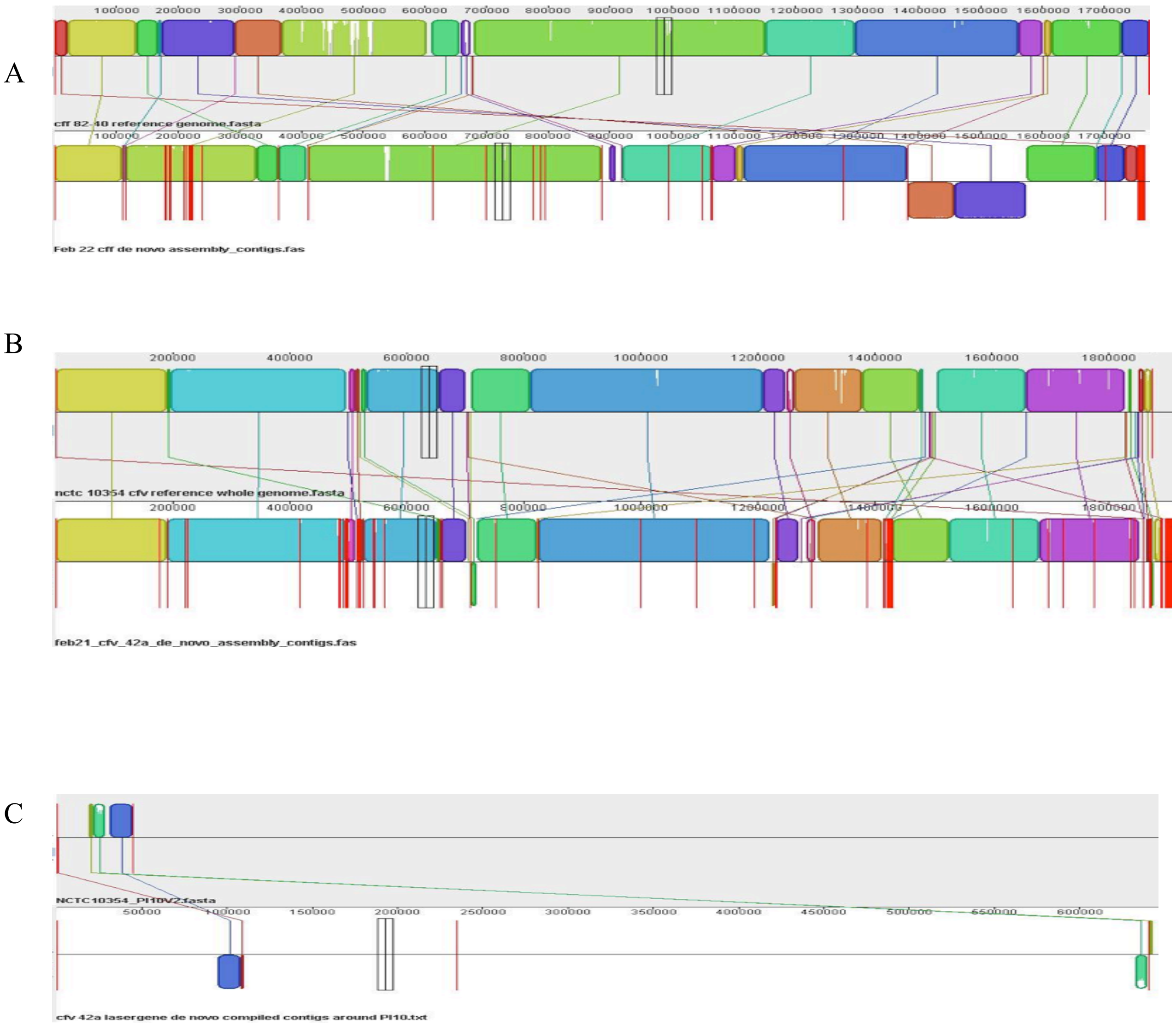


Figure 18. Evaluation of reference directed CF sequence assemblies by optical mapping. The optical genome maps of the Canadian CFV08A1102-42A (Panel A) and CFF02A705-35A (Panel B) isolates are compared to their respective in *silico* maps generated from the reference-based sequence assembly of each isolate. Maps generated by MapSolver software, OpGen.

3.9.3 Analysis of Sequencing Data – MAUVE Analysis

Additional analysis was done using Mauve software to confirm the order of the collection of contigs obtained from the *de novo* assemblies. By Mauve software, the genome of both the CFF and CFV reference isolates were aligned to the contigs of the *de novo* assembled genomes of the Canadian CFF and CFV sequences respectively to examine the predicted order of the contigs (Figure 19A and 19B). From the figure below one can infer that there are some rearrangements of the predicted genes in the reference–based assemblies compared to the template references. Therefore, the order of the genes may not be correct in some cases. Furthermore, when the analysis was restricted to the sequence corresponding to the pathogenicity island PICFV10 (since all 3 SSH clones are located within this region) in the Canadian CFV genome this region was split into two locations about 400kb distant, so the genes present in this island are in different locations whereas in the reference genome this island is contiguous (Figure 19C). In conclusion, the distinctions between the two genomes represented by the difference in the order of the contigs as predicted by Mauve analysis, together with the optical mapping data, imply that there are some issues with the reference based sequence assemblies.

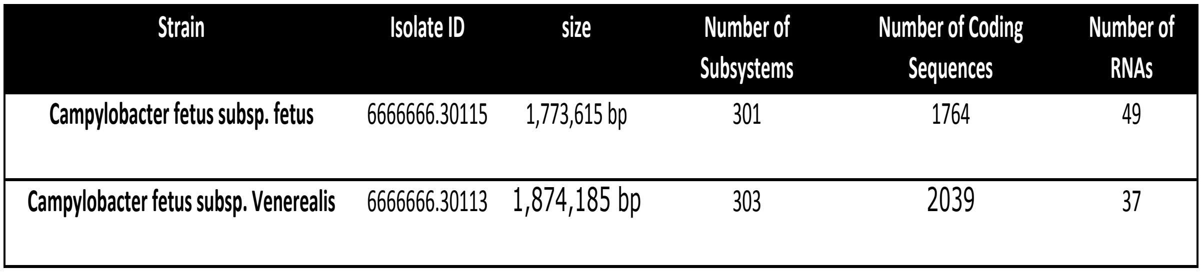
Figure 19. Analysis of sequence alignments using Mauve. Alignment of (A) CFF82-40 reference genome (top) with CFF02A725-35A contigs (bottom); (B) CFV NCTC10354 reference genome (top) with CFV08A1102-42A contigs (bottom); and (C) PICFV10 of CFV NCTC10354 reference genome (top) with CFV08A1102-42A contigs (bottom). The alignment of the genome sequence, that contains the name and the coordinates of that genome, is ordered into one horizontal panel. The colored boxes indicate homologous regions present in each bacterial genome that are connected by separate lines. Any blocks above the centerline represent sequences in the forward direction, while any blocks underneath the centerline of each genome indicate regions with inverse orientation (reverse complement of the contig). In each block DNA sequence profiles are similar, and any white area represents unique sequence in that genome. Red lines symbolize junction between two contigs that needs to be validated by PCR or re-sequencing. The scale is in base pairs.



3.9.4 Analysis of Sequencing Data – Rapid Annotation

The reference-based assembled bacterial genomes of both CF isolates were annotated by using RAST software. Although it is acknowledged that the order of genes predicted in the reference–based assemblies may not be correct in some cases, for the most part the prediction of the genes themselves should be generally quite accurate; furthermore the *de novo* assembled genomes were not chosen for the genome annotation because of the low (<60%) coverage for the CFF. The RAST system, together with SEED-viewer analysis software, allows online fully automated bacterial genome annotation. The circular genome of CFV08A1102-42A is composed of 1,874,185 bp with a G+C content of 33 % and includes 2,039 putative protein-coding genes or open reading frames. While the genome of CFF02A725-35A is composed of 1,773,615 bp with a G+C content of 33 % and includes 1764 putative protein-coding genes or open reading frames. The overall results of the annotated CFF/CFV reference-based assembled genomes, together with coding functional genes, are found in Figure 20 and Table 21 below.

Table 21. Organism Overview annotation of SSH CFF02A725-35A and CFV08A1102-42A using SEED viewer – RAST (http://rast.nmpdr.org).



Strain	Isolate ID	size	Number of Subsystems	Number of Coding Sequences	Number of RNAs
Campylobacter fetus subsp. fetus	6666666.30115	1,773,615 bp	301	1764	49
Campylobacter fetus subsp. Venerealis	6666666.30113	1,874,185 bp	303	2039	37

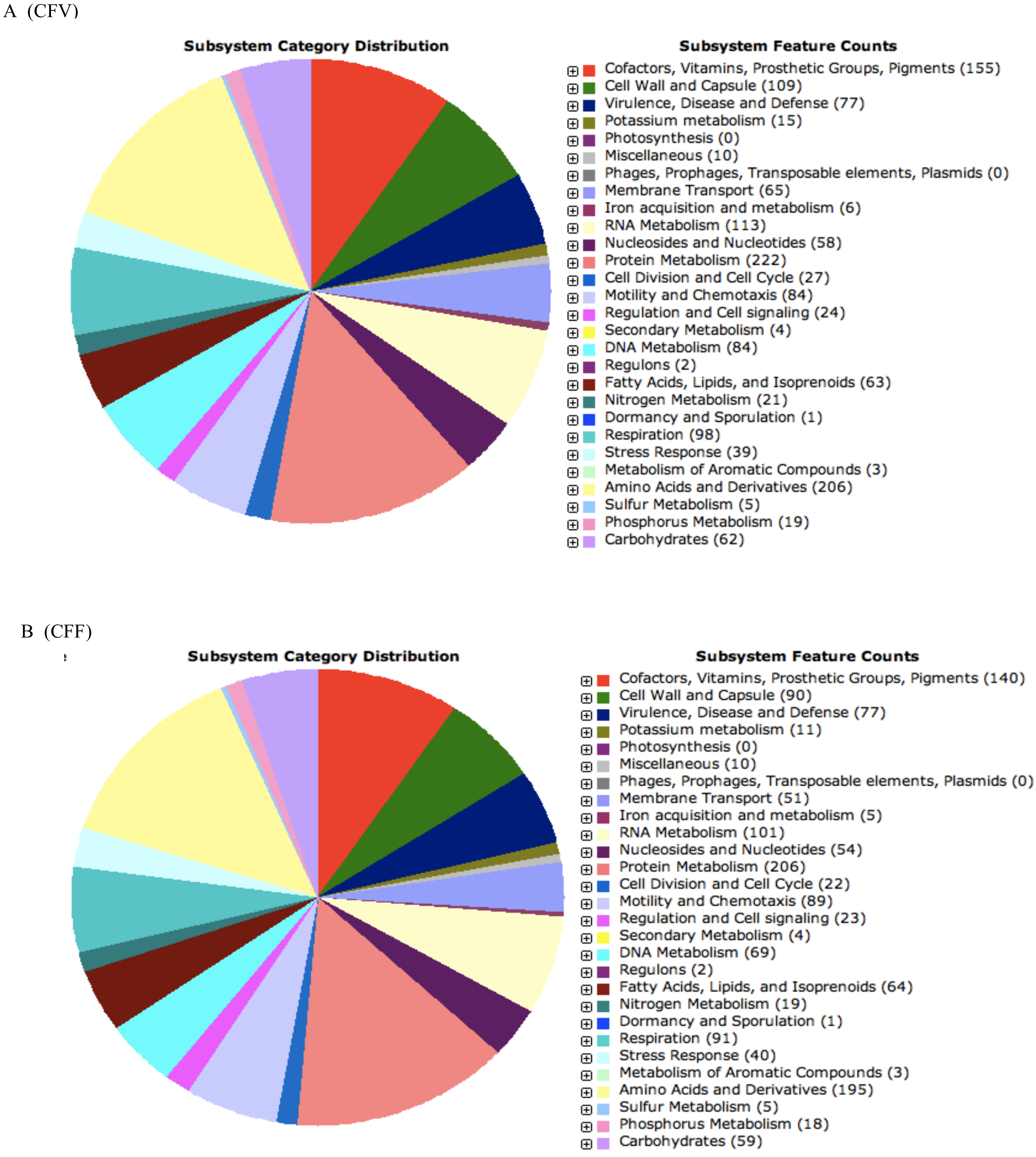


Figure 20. Genes connected to subsystems and their distribution in different categories within the bacterial genomes of CFV08A1102-42A (A) and CFF02A725-35A (B). Results obtained using SEED viewer – RAST (http://rast.nmpdr.org).

Comparison of the functional genes in both CFF and CFV genomes identified 7 unique features in CFV08A1102-42A and 4 unique features in CFF02A725-35A (Table 22). A BLAST search of each identified unique feature using both Canadian isolates, CFF02A725-35A and CFV08A1102-42A, as well as the reference genomes of CFF82-40 and CFV NCTC10354, was undertaken to confirm its presence or absence among these isolates (Table 23). Further BLAST analyses were done to confirm the specificity of these unique features, by aligning each unique feature with *de novo* assembled genomes of 6 additional *C. fetus* isolates (Table 23). From these analyses it was found that only CFF_Feature #3, which is located within pathogenicity island PICFF10 in the CFF82-40 reference genome, was uniquely present in the CFF reference and all Canadian CFF isolates, but absent from all CFVs; thus it is considered a promising discriminatory target.

Table 22. Unique features in the (A) Canadian CFV08A1102-42Aand (B) CFF02A725-35A after comparison of all functional genes in both isolates.

A

Unique features in CFV 08A1102-42A

Feature#	start	stop	strand	Category	Subcategory	Subsystem	function
1	1770651	1771388	+	Membrane Transport	Protein and nucleoprotein secretion system, Type IV	Conjugative_transfer	IncQ plasmid conjugative transfer protein TraQ (RP4 TrbM homolog)
2	1797811	1799148	+	Membrane Transport	Protein and nucleoprotein secretion system, Type IV	Conjugative_transfer	IncQ plasmid conjugative transfer DNA nicking endonuclease TraR (pTi VirD2 homolog)
3	1815661	1816887	+	Membrane Transport	Protein and nucleoprotein secretion system, Type IV	Conjugative_transfer	Conjugative transfer protein TrbE
4	1819875	1820303	+	Membrane Transport	Protein and nucleoprotein secretion system, Type IV	Conjugative_transfer	Conjugative transfer protein TrbL
5	1820467	1821150	+	Membrane Transport	Protein and nucleoprotein secretion system, Type IV	Conjugative_transfer	Conjugative transfer protein TrbF
6	1822531	1823505	+	Membrane Transport	Protein and nucleoprotein secretion system, Type IV	Conjugative_transfer	Conjugative transfer protein TrbI
7	1821169	1822083	+	Membrane Transport	Protein and nucleoprotein secretion system, Type IV	Conjugative_transfer	Conjugative transfer protein TrbG

B

Unique features in CFF 02A725-35A

Feature#	start	stop	strand	Category	Subcategory	Subsystem	function
1	382429	383781	+	Cofactors, Vitamins, Prosthetic Groups, Pigments	no subcategory	Thiamin_biosynthesis	Thiamin biosynthesis protein ThiC
2	1470118	1470981	+	Motility and Chemotaxis	Flagellar motility in Prokaryota	Flagellar_motility isu;Flagellum	Flagellar synthesis regulator FleN
3	1549125	1550234	+	Cell Wall and Capsule	Gram-Negative cell wall components	LOS_core_oligosaccharide_biosynthesis	UDP-galactopyranose mutase (EC 5.4.99.9)
4	1722858	1722930	+	Protein Metabolism	Protein biosynthesis	tRNAs	tRNA-Ala-GGC

Table 23. BLAST of unique CFF/CFV features with Canadian and reference CFV/CFF isolates

		Assigned accession #									
		1	**2**	**5**	**6**	**7**	**8**	**20**	**22**	**24**	**25**
Feature#	**Function**	**BLAST with CFF 82-40 reference genome sequence**	**BLAST with CFV NCTC10354 reference genomesequence**	**BLAST with contig sequences CFF 02A2725.35A (paird) de novo assembly**	**BLAST with contig sequences CFV 08A1102.42A (paird) de novo assembly**	**BLAST with contig sequences CFV 08A948.2A (paird) de novo assembly**	**BLAST with contig sequences CFV 08A110233B.5A (paird) de novo assembly**	**BLAST with contig sequences CFVi ADRI 545.1A (paird) de novo assembly**	**BLAST with contig sequences CFF ADRI 1362.8A (paird) de novo assembly**	**BLAST with contig sequences CFF 09A9803.2B (paird) de novo assembly**	**BLAST with contig sequences CFFTurbo serotypeB (paird) de novo assembly**
CFV feature_1	IncQ plasmid conjugative transfer protein TraQ (RP4 TrbM homolog)	NO	Yes	No	yes, with contig_21	Yes, with contig_1	Yes, with contig_33	Yes, with contig_5	yes, with contig_45	No	Yes, with contig_14
CFV feature_2	IncQ plasmid conjugative transfer DNA nicking endonuclease TraR (pTi VirD2 homolog)	NO	Yes	No	yes, with contig_20	yes, with contig_37	yes, with contig_33	Yes, with contig_61	No	Yes, with contig_4	No
CFV feature_3	Conjugative transfer protein TrbE	NO	Yes	No	yes, with contig_9	yes, with contig_44	yes, with contig_40	Yes, with contig_5 & 80	yes, with contig_45	No	Yes, with contig_14
CFV feature_4	Conjugative transfer protein TrbL	NO	Yes	No	yes, with contig_10	No	yes, with contig_32	Yes, with contig_5	yes, with contig_55	No	Yes, with contig_14
CFV feature_5	Conjugative transfer protein TrbF	NO	Yes	No	yes, with contig_11	yes, with contig_44	yes, with contig_33	Yes, with contig_5	yes, with contig_56	No	Yes, with contig_14
CFV feature_6	Conjugative transfer protein TrbG	NO	Yes	No	yes, with contig_12	yes, with contig_45	yes, with contig_34	Yes, with contig_5	yes, with contig_57	No	Yes, with contig_14
CFV feature_7	Conjugative transfer protein TrbI	NO	Yes	No	yes, with contig_13	yes, with contig_46	yes, with contig_35	Yes, with contig_5	yes, with contig_58	No	Yes, with contig_14
CFF feature_1	Thiamin biosynthesis protein ThiC	Yes	NO	yes, with contig 4	No	yes, with contig 4	yes, with contig 6	Yes, with contig 17	yes, with contig 2	Yes, with contig_5	Yes, with contig_5
CFF feature_2	Flagellar synthesis regulator FleN	Yes	NO	yes, with contig 3	No	yes, with contig 14	yes, with contig 13	Yes, with contig 28	yes, with contig 3	No	Yes, with contig_3
CFF feature_3	UDP-galactopyranose mutase (EC 5.4.99.9)	Yes	NO	yes, with contig 20	No	No	No	NO	No	Yes, with contig_2	Yes, with contig_21
CFF feature_4	tRNA-Ala-GGC	Yes	NO	No	No	No	yes, with contig 1	Yes, with contig 21	No	No	No

3.9.6 Analysis of Sequencing Data – BLAST of ORFs from PICFV5 and PICFV10 Against *De novo* Assembled Genomes

Additional bioinformatics analyses were done to test the presence of the ORFs, identified previously by Ali et al (2012), of the pathogenicity islands 5 and 10 of the reference CFV NCTC10354 (PICFV5 and PICFV10) in the Canadian CFV. The two islands were selected specifically because: (1) In the case of PICFV10, all the previous 3 clones obtained from the SSH were located in the PICFV10; and (2) In the case of PICFV5, according to Ali et al (2012) this island was partially deleted in the CFF. Thus, a BLAST search of all the ORFs of the two islands was performed on the *de novo* assembled genomes of all sequenced CF isolates (data not shown); the analysis of the eight genomes resulted in the identification of several genes in the PICFV5 that are found in the four CFVs but not in the CFFs. However, only one gene of the PICFV10 was found exclusively in the CFVs. Based on this information the use of 4 ORFs (3 in the PICFV5 and 1 in the PICFV10) were as potential discriminatory targets was explored. Those 4 ORFs are: (i) PI5_CFV354_548, (ii) PI5_CFV354_576, (iii) PI5_CFV354_581, and (iv) PI10_CFV354_1581; details of the nucleotide sequences of these ORFs are found in Table 24 and Appendix 6.

Table 24. BLAST of PICFV5 and PICFV10 - ORFs with *de novo* assembled Canadian CFV isolates.

C.fetus denevo assembly	ORF	PICFV	Contig #	% identity	alignment length	mismatches	gap opens	Query start	Query end	Subject start	Subject end	e value	bit score
CFV2008A1102.42A (paired)	CFV354_0548: 523218-523649	5	contig 34	99.77	433	0	1	1	433	2728	3160	0	793
	CFV354_576:543461-544717		contig_25	100%	1257	0	0	1	1257	9340	8084	0	2322
	CFV 354_581:548170-549096		contig_25	100%	927	0	0	1	927	4631	3705	0	1712
	CFV354_1581.1505706-1506275	10	contig 10	72.73	517	123	14	15	522	10062	10569	2.00E-41	158
08A948.2A (paired)	CFV354_0548: 523218-523649	5	contig 7	99.77	433	0	1	1	433	2729	3161	0	793
	CFV354_576:543461-544717		contig_26	100%	1257	0	0	1	1257	10534	11790	0	2322
	CFV354_581:548170-549096		contig_26	100%	927	0	0	1	927	15243	16169	0	1712
	CFV354_1581.1505706-1506275	10	contig 34	72.73	517	123	14	15	522	2285	1778	2.00E-41	158
08A110233B.5a (paired)	CFV354_0548: 523218-523649	5	contig 18	99.77	433	0	1	1	433	2724	3156	0	793
	CFV354_576:543461-544717		contig_22	100%	1257	0	0	1	1257	10535	11791	0	2322
	CFV354_581:548170-549096		contig_22	100%	927	0	0	1	927	15244	16170	0	1712
	CFV354_1581.1505706-1506275	10	contig 23	72.73	517	123	14	15	522	10063	10570	2.00E-41	158
ADRI545.1a (paired)	CFV354_0548: 523218-523649	5	contig 26	98.38	433	6	1	1	433	643	1075	0	760
	CFV354_576:543461-544717		contig_37	100%	1257	0	0	1	1257	10536	11792	0	2322
	CFV354_581:548170-549096		contig_37	100%	927	0	0	1	927	15245	16171	0	1712
	CFV354_1581.1505706-1506275	10	contig 36	72.73	517	123	14	15	522	2285	1778	7.00E-42	158

3.9.7 Analysis of Sequencing Data – PCR Screening of the Targeted Genes

Primer sets for the promising targeted genes, from sections 3.9.4, 3.9.5 and 3.9.6 (CFF_Feature#3, SSH_clone#35, PICFV10_ORF1581, and the 3 ORFs in PICFV5, ORF548, ORF576, and ORF581) were designed. Initial PCR trials of the targeted primers were done using the CFF/CFV isolates used for the SSH. From the initial screening only three PCRs out of several tested are in fact CFV-specific. These three genomic regions are the SSH_Clone #35 identified by the subtractive hybridization approach but note that only parts of this sequence are indeed restricted to CFV, ORF548 in Pathogenicity Island PICFV5 and CFF_Feature #3 in PICFF10, which were identified from WGS and genome annotation studies. The presence of the faint band for ORF548 in the CFF (lane 9, Figure 21B) was subsequently eliminated by raising the annealing temperature of the PCR cycling profile to 57 ^{o}C (Figure 21C).

A collection of the 42 selected CF isolates (Acc. # 5 – 49 as in Table 2, except #8, 21, 25, and 28), previously characterized biochemically and phenotypically as CFV or CFF, was tested using the three PCRs targeting each of these regions. From the screening (Figure 21 and Table 25) PCRs targeting both SSH_Clone #35 and PICFV5_ORF548 appeared to be highly specific for the CFVs (PICFV5 was positive for all but one CFV, while SSH_Clone #35 was positive for all and one of the CFF). On the other hand, The CFF_Feature #3 target is consistently negative with all CFV isolates but not positive for all CFFs. To confirm that these negative results for the CFF feature_3 were not a result of poor template quality, some DNA preparations were reassessed by PCRs using a universal primer set (see below in Table 26) that amplifies all CFs.

DNA from all isolates did support amplification of rRNA genes indicating that the integrity of the template was not a problem. Accordingly the negative results obtained with the PCR targeting the CFF_Feature #3 appears to be a true negative result.

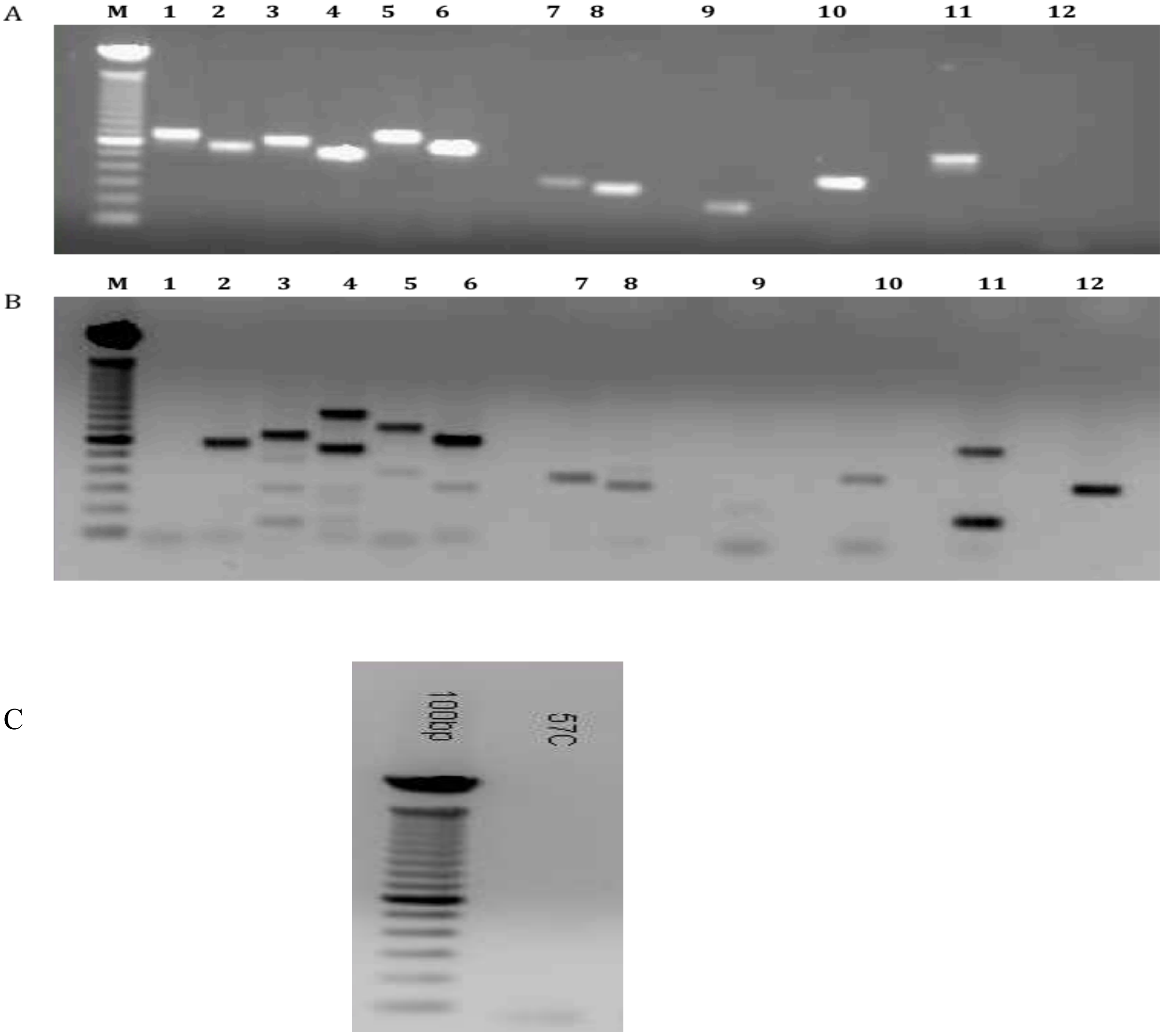


Figure 21. Evaluation of PCRs targeting selected sequences for CF subspecies discrimination. Gel image of PCR products on an ethidium bromide stained 1% agarose gel following amplification using different sets of primers against (A) CFV08A1102-42A and (B) CFF02A725-35A. Lane M: 100bp molecular weight marker, the first six lanes represent DNA template with SSH_clone#35 primers (lane 1: cln35F1/R1, lane2: cln35F1/R2, lane3: cln35F2/R1, lane4: cln35F2/R2, lane5: cln35F3/R1, and lane6: cln35 F3/R2); lanes 7 and 8 represent primers of PICFV10_ORF1581 F1/R1 and F2/R1, respectively; lane9 – 11 represent PICFV5_ORFs 548, ORF576, and ORF581, respectively; while lane12 indicate primers of CFF_feature3. The expected product sizes are: 687 bp in lane 1, 558 bp in lane 2, 660 bp in lane 3, 561 bp in lane 4, 670 bp in lane 5, 570 bp in lane 6, 394 bp in lane 7, 365 bp in lane 8, 247 bp in lane 9, 424 bp in lane 10, 657 bp in lane 11, and 400 bp in lane 12. Panel C shows the elimination of the PI548 amplicon in the CFF02A725-35A isolate using an elevated annealing temperature (57°C).

Table 25. PCR screening of the 3 targeted regions against 42 *C. fetus* strains.

Isolate Acc#	subspecies*	Ref_Number	Clone35 F1/R1 primer sets	PI5_548 F/R primer sets	CFF3 F/R primer sets
4	CFV	ATCC 19438	POS	POS	NEG
6	CFV	08A1102 42A	POS	POS	NEG
7	CFV	08A948 - 2A	POS	POS	NEG
9	CFV	ADRI 1345	POS	NEG	NEG
10	CFV	ADRI 1024	POS	POS	NEG
11	CFV	ADRI 520	POS	w.POS	NEG
12	CFV	08A1102 3A	POS	POS	NEG
13	CFV	08A1102 24A	POS	POS	NEG
14	CFV	08A1204 32A	POS	POS	NEG
15	CFV	08A1204 17B	POS	POS	NEG
16	CFV	08A1102 39A	POS	POS	NEG
17	CFV	08A1204 1A	POS	POS	NEG
18	CFV	08A1204 12A	POS	POS	NEG
19	CFVi	ADRI 510	POS	POS	NEG
20	CFVi	ADRI 545.1A	POS	POS	NEG
21	CFVi	ADRI 546	POS	POS	NEG
3	CFF	ATCC 27374	NEG	NEG	POS
5	CFF	2002A725-35A	NEG	NEG	POS
22	CFF	ADRI 1362.8A	POS	NEG	NEG
23	CFF	ADRI 1346	w.POS	NEG	NEG
27	CFF	01A988-2A	NEG	NEG	w. POS
29	CFF	ADRI 1032	NEG	NEG	NEG
30	CFF	ADRI 516	NEG	NEG	POS
31	CFF	05A451 9B	NEG	NEG	w. POS
32	CFF	07A621 15A	NEG	NEG	POS
33	CFF	08A314 4B	NEG	NEG	POS
34	CFF	09A376 - 1A	NEG	NEG	NEG
35	CFF	08A1242 2A	NEG	NEG	w. POS
36	CFF	AIN458-3	NEG	NEG	POS
37	CFF	98AIN525-1	NEG	NEG	NEG
38	CFF	01A603-82A	NEG	NEG	POS
39	CFF	07A1157 #66	NEG	NEG	NEG
40	CFF	03A564-113A	NEG	NEG	NEG
41	CFF	04X189 1B	NEG	NEG	w. POS
42	CFF	09A980 - 3A	NEG	NEG	POS
43	CFF	09A980 -16B	NEG	NEG	POS
44	CFF	06A1204 293A	NEG	NEG	POS
45	CFF	06A1553 3A	NEG	NEG	POS

* CF subspeciation determined biochemically and phenotypically

Table 26. PCR screening confirming suitability of genomic preparations of selected *C. fetus fetus* isolates for PCR.

ACC#	CFF isolate	Testing against 16S rRNA	Testing against 23S rRNA
22	ADRI 1362	POS	POS
23	ADRI 1346	POS	POS
29	ADRI 1032	POS	POS
34	09A376-1A	POS	POS
37	098AIN525-1	POS	POS
39	07A1157#66	POS	POS
40	03A564-113A	POS	POS
46	ADRI 1359	POS	POS

Chapter Four: Discussion

Pathogenic CF, which belongs to the ε-proteobacteria, is an important emerging veterinary pathogenic bacterium that targets both humans and animals causing significant morbidity [3,150]. It is highly adapted to the host's mucosal surface [73]. It has two subspecies that deserve special attention based on their habitat and disease they cause. CFV is the major cause of a serious venereal disease in cattle (i.e. bovine genital campylobacteriosis), and leads to severe loss in cattle productivity worldwide; consequently, strict international animal trade regulation requires cattle to be CFV-free. On the other hand, CFF can infect many mammalian species including humans where it is the causative agent of human gastroenteritis and affects immunocompromised patients in particular. As such these bacteria are of interest for study at two levels: (i) to develop robust and rapid discriminatory tests so as to limit their socio-economic impacts and (ii) to understand the fundamental genetic differences that result in such different pathologies.

Based on the phylogenetic analysis of ribosomal RNA, both subspecies (CFF and CFV) are clustered together, due to their high genomic similarity; findings supported by other molecular analyses such as PFGE which indicated that the two subspecies have genomes that are 86% related [130,150] consistent with other microbiological and molecular findings [120,138]. The logical reason for their similarity is due to their small genome size (~1.8Mb) compared to other bacteria, thus large proportions of their genes are expected to be in the core genome, and relatively few genes are expected to be part of the dispensable accessory genome [3,66].

To date, only two annotated whole genome sequences of the CF are available in the GenBank (NCBI). One complete genome sequence of the CFF strain 82-40 (NC_008599) represents a polished and finished sequence while a CFV sequence; strain NCTC10354 (CM001228), remains as a draft genome. In addition another incomplete genome sequence of CFV, strain Azul-94, is available only as several contigs and has yet to be assembled [3,108]. These limited sequence data show that despite much genetic relatedness the CF genome is characterized by significant flexibility, as observed in other *Campylobacter* species exhibiting rapid evolution [183,187]. Based on a recent study, about 428 gene families were found uniquely in the CFV subspecies, while 88 gene families were found in the CFF subspecies [3]. Many of these genes were found exclusively in genomic islands, and they don't show any noticeable similarity to genes in other *Campylobacter* genomes available in the database [3,72]; accordingly their functions remain unknown.

Despite this information, the molecular mechanisms responsible for these subspecies traits remain mostly unknown [89]. A number of molecular typing methods such as PFGE, MLST and RAPD-PCR, have been used to study intra-species variation. However, these methods have successfully identified CF strains for epidemiological purposes only, but did not reveal features pertinent to their pathogenicity [21,22,38,184].

A comparison of the two sequenced and assembled genomes of CFF and CFV subspecies (see Ali et al 2012) yielded some understanding of their genomic architecture. Most notably a number of distinct putative pathogenicity islands were identified in both subspecies [3]. However, comparison between only two genomes generates limited information, which may be a poor basis for identification of the definitive genomic differences between CFF and CFV.

An alternative approach utilized by Gorkiewicz et al (2010), attempted to overcome these limitations by applying a subtractive hybridization approach. The suppression subtractive hybridization technique is one of the few methods that has been applied previously for several comparative and functional genomic studies such as distinguishing between pathogenic and non-pathogenic microorganisms, identification of certain diseases like cancer, or other differentially expressed genes associated with cellular growth and development [64,146,184]. SSH is a particularly important tool in the bacteriology field for the identification of transposable mobile genetic elements, in the identification of strain-specific or clone-specific markers, and the identification of pathogenic/genomic islands implicated in bacterial virulence [21,146,163].

This approach, which was initially applied to two reference genomes and then extended to an analysis of a larger collection of isolates, successfully identified the presence of a genomic island (PICFV8) exclusively in many CFV strains [59]. This genomic island contains several genes that are thought to be associated with horizontal transfer of molecules and the genes that are necessary for the bacterial type IV secretion system. In bacteria, the process of horizontal gene transfer provides genetic diversity that leads to virulence variations, such as antibiotic resistance and niche preferences [16,58,59,63,86,91]. A genome analysis in CFV isolates 84-112 and AZUL-94 revealed that this pathogenicity island contains the *par*A gene, a locus previously considered to be unique to the subspecies *venerealis* [3,25,59,108]. While many CFV strains contain the PICFV8 region, this analysis was heavily biased towards European and Australian samples with relatively few samples from the Americas. Furthermore, they were not able to demonstrate complete concordance between the CFV phenotype and PICFV8 presence, thereby raising the possibility that other factors play a role in defining these subspecies. This is the basis on which this project started and has attempted to develop additional genetic tests to discriminate between the CF subspecies.

By reapplying the subtractive hybridization approach to analyze differences between Canadian CFF and CFV isolates, I hoped to identify CFV-specific features and to explore their involvement with the pathogenicity of the organism taking into account information available from prior studies. Cases of CFV are occasionally reported in Canada [25,111], but the only diagnostic method to indicate the prevalence of CFV in Canada is the time-consuming culture technique. A better understanding of the genetic differences between the two Canadian subspecies would provide more robust and rapid discriminatory tools.

At the beginning of the study, a collection of CF isolates, mostly from Canada and Argentina, was tested by PCRs targeting several locations within PICFV8; several of the CFV isolates were lacking much of this island. Accordingly the isolate CFV08A1102-42A, a classical CFV that was isolated from an outbreak in Alberta in 2008 and lacked much of the PICFV8, was selected as the tester organism for SSH and CFF02A725-35A, a typical CFF strain, was selected as the driver.

Out of 50 clones recovered from the SSH protocol, inserts ranged between 500bp – 2.0 kb. After preliminary screening a total of 34 of these clones were sequenced. However, the Blast searches and sequence alignments revealed that 26 of these cloned sequences were 100% identical with the genome of CFF82-40. Eventually it was determined that only one clonal sequence, SSH_Clone #35, was a useful target for a CF subspecies discriminatory assay. Portions of this sequence were absent in all of the CFF isolates used for the blotting/sequencing analyses except for one CFF strain ADRI 1362 (Acc. #22 in Table 2). However, the identity of this isolate is questionable since there was a contradiction in its identification. This strain was isolated from the cervico-vaginal mucus of a heifer with a syndrome of infertility; the heifer belonged to a herd with a low rate of pregnancy. It was identified in Argentina as CFV, while in Canada it was classified biochemically as a CFF strain. The sequence of clone #35 was located within the pathogenicity island (PICFV10), previously identified by Ali et al (2012); an ORF contained around this sequence codes for a hypothetical protein so its function is unknown.

There are many reasons behind obtaining so few useful clones from this SSH study. First, the DNA concentration of tester and driver employed for the hybridizations is very important to ensure a large excess of driver. Errors in these measurements would influence the tester:driver ratio, and might result in incomplete removal of tester DNA that was homologous to the driver. This would increase the rate of obtaining false positives clones [21,74]. In these experiments genomic DNA was measured using a Nanovue® spectrophotometer, but these measurements can be influenced by contaminants in the preparation and thus provide inaccurate DNA concentrations. Indeed some genomic DNA extracts used in these experiments did exhibit high 260/280 ratios, above 2.0, indicative of the presence of some residual protein contamination in the DNA. Such DNA contamination could prevent complete RsaI-restriction endonuclease digestion of the DNA, and result in DNA unsuitable for adapter ligation [2,21,37,154]. Further purification of such DNA preparations might be helpful. Moreover, the use of alternative methods of measuring DNA concentration, which are less prone to being influenced by components other than dsDNA might have been beneficial.

Secondly, since the results of the SSH depend on the efficiency of the ligation of the two adaptors (Adp1 and Adp2R), incomplete ligation can result in loss of tester-specific fragments. As shown in Figure 14, although the ligation reaction was successful, the intensity of the bands representing the ligated CFV-tester DNA (lanes 2 and 4) were not as strong as the bands representing the *ven* amplicons (lanes 1 and 3), thereby suggesting that ligation may not have been 100% efficient.

Another important limitation of any subtractive hybridization method is that there will always be a certain proportion of false positives [2,21], and it has been reported that the number of false positive clones can exceed the number of targeted clones in certain subtracted libraries [74,184]. In an effort to evaluate the success of two consecutive rounds of subtractive hybridization in this study, electrophoresis of secondary PCR products (lane 6 of Figure 15) indicated a positive subtractive hybridization between CFV and CFF. As expected, differential amplicon patterns were visualized upon comparing unsubtracted with subtracted library of CFV genomic DNA. Despite this observation the SSH success rate was very low. It is clear that if a high percentage of single strand sequences homologous to the CFF driver evade the elimination step and then hybridize to produce dsDNA bearing adaptor sequences they can form amplified products. Also, some unincorporated adapters could randomly be annealed to non-target driver DNA during subtractive hybridization, thereby raising the background of false positives [1,74,132].

A key factor for successful hybridization results is the stringency (hybridization temperature) at which the hybridization takes place; some tester strain sequences, which have low sequence homology with the driver DNA, might still be removed by low stringency subtraction if these conditions favor hybridization of poorly matched sequences; adversely, perfectly matched sequence could be removed by high stringency subtraction since high hybridization temperature would favor the annealing of longer PCR products and therefore limit overpopulation of the needed short fragments [2,21,67,74].

A second trial of the same technique but using the CFF as tester and CFV as a driver failed; that may be due to the smaller genome size of the CFF strain (1.7MB) and the near complete hybridization of all fragments by the larger CFV genome (1.8MB) and thus the removal of any unique features for amplification. Another possible reason might be the purity of the tester DNA; after the genomic DNA extraction from the CFF for the second trial the OD260/280 ratio was high at 2.08 (data not shown).

To supplement the SSH results, whole genome sequencing and Optical Mapping approaches (which complement each other) were alternative techniques applied to assist in these studies. The genomic DNA of both subspecies of the CF, used for the SSH, were sent for whole genome sequencing analysis using an Illumina platform. The generated reads were assembled using both reference-based and *de novo* assemblies. The application of the optical mapping helps to confirm and corroborate the order of the assembled contigs. By comparing the genomic maps of both subspecies to sequenced *in silico* reference maps, major differences between the optical maps of the reference and Canadian isolates were evident thereby indicating that these genomes exhibit significant differences especially with respect to genome length and the positions of the pathogenicity islands in the CFV (i.e. PICFV8 and PICFV10 are located in regions where the maps are rather different). Additionally, the difference in the maps between the one determined by direct genomic restriction analysis and the one inferred from the reference based assembly of both subspecies revealed that the reference-based assemblies were not that accurate. Therefore, the reliability of the reference genomes (82-40 and NCTC10354) used for the assemblies was not optimal; indeed as shown by the Mauve analysis some rearrangements were observed in the Canadian CFV genome compared to the reference genome. On the other hand, based on the optical mapping analysis, the genome coverage of the CFV contigs generated by the *de novo* assembly was 71%, which is considered a reasonable coverage. However the genome coverage of the CFF contigs obtained from the same method was very low (35%). The presence of SSH_clones #2 & #28 in contig 5 from the *de novo* assembled CFF is questionable since both were absent from the previous molecular analyses and only found in this contig after genome assembly. The analysis indicates that none of these assemblies was recommended over the other.

Although the optical mapping system creates high-resolution physical maps of bacterial genomes based on ordered restriction maps of individual DNA molecules and so should be accurate, however it does have limitations; when many small fragments are generated they can get lost in the analysis. On the other hand due to the small read length of Illumina WGS, assembly of these data can be fraught with difficulties, especially when the genome contains many copies of certain sequence motifs eg. IS elements pseudogenes, repetitive elements etc. High sequence coverage can help to reduce the complexity of the assembly process.

Despite these misgivings it was decided to use the genome sequences of both subspecies generated by reference-based assembly for annotation by using RAST software. It was hoped that this analysis might identify sequences of interest that could be further evaluated empirically. The identified functional genes of both CFF and CFV were compared to each other using RAST; a total of seven unique features were found in the genome of CFV08A1102-42A and four presented in the CFF02A725-35A genome. From the BLAST analysis of those unique features against the *de novo* assembled genomes of six selected CF isolates, it was revealed that only CFF feature #3, present in CFF02A725-35A, was the most promising target. Surprisingly, this feature matched within a region of the putative pathogenicity island (PICFF10) of the reference CFF82-40, as identified by Ali et al (2012). This region codes for UDP-galactopyranose mutase enzyme, which plays a role in the biosynthesis of the LOS_core_oligosaccharide cell wall of gram-negative bacteria [3]. Another CFV-specific target region (ORF548) was identified from the BLAST analysis of the ORFs from the PICFV5 (as identified by Ali et al (2012) [3] against the *de novo* assembled genomes of six selected CF isolates. This region was uniquely present in all selected CFVs and absent from all the CFFs.

With the use of next generation sequencing technology in combination with the optical mapping, the following benefits were achieved in this project: (i) showing some differences between Canadian CFF/CFV after comparing the genomic sequences of the two selected isolates, (ii) proving that both assemblies were needed for proper alignment and better coverage of the CF genomes; (iii) observing the dissimilarity in the pathogenicity islands in both CFF_82-40 and CFV_NCTC10354, which is in agreement with Ali et al, 2012, (data not shown), and (iv) identifying significant differences between the Canadian and reference CF strains with respect to the presence of pathogenicity islands, (v) there was a correlation between culture and biochemical results and the PCR screening of the targeted genes, except for a few CFF isolates (as shown in Table 25). Remarkably, comparing the project data to those of Ali et al, part of pathogenicity island PICFV5 (coordinates from 548170 to 549096) is uniquely found in the Canadian CFVs, not the CFFs, while much of the other islands (PICFV8 and PICFF6) are absent in the Canadian CFs. Many of the genes within PICFV5, including ORF548, code for hypothetical proteins whose function is unknown. Interestingly, these 3-targeted genes are also differing from genes identified recently by McGoldrick et al (2013). Therefore, that proves the variation in the CF genome, which is reflected by the life-style of the bacteria and by the geographical area.

To account for the observation that few CFV subspecies-specific sequences could be recovered despite significant differences in genome structure, as demonstrated by optical mapping, would suggest that the major difference between these two subspecies is not in their overall genomic content but in the arrangement of their genomes. The high plasticity of the genomes of these organisms makes identification of genes responsible for their distinct pathogenicities extremely challenging.

Conclusion and Future Direction

This thesis project describes the application of a PCR-based subtractive hybridization method together with the whole genome sequencing and optical mapping to a better understanding of the genomic differences between Canadian isolates of *Campylobacter fetus*. The findings from this study have added some new information to our general understanding of the genomic variation between *C. fetus feus* and *C. fetus venerealis.* Although it appears that the subtractive hybridization technique is not suitable for evaluating the genetic differences between Canadian CFF and CFV strains, optimizing SSH conditions might increase its potential to identify differences between these strains. However, this approach is being increasingly overshadowed by development in WGS technologies. Further sequence analyses of a broader collection of CF strains would be beneficial in more clearly identifying unique features responsible for the distinct pathogenicity of these organisms. In particular better appreciation of the functions of the hypothetical proteins associated with several of the pathogenicity islands is required. In the meantime the use of the sequences (SSH_clone35, PI5_ORF548, and CFF_feature3) identified by the SSH and WGS analyses, together with the *parA* locus, for robust CF subspecies discrimination is being actively explored by the laboratory of Dr. Nadin-Davis at OLF. Finally this work has demonstrated the complexity and the flexibility of the *Campylobacter fetus* genome that contribute to the evolution of the organism.

References

1. Agron PG, Macht M, Radnedge L, Skowronski EW, Miller W, Andersen GL. (2002) Use of subtractive hybridization for comprehensive surveys of prokaryotic genome differences. FEMS Microbiol Lett. 11: 175-182.

2. Akopyants NS, Fradkov A, Diatchenko L, Hill JE, Siebert PD, et al. (1998) PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. Proc Natl Acad Sci U S A 95: 13108-13113.

3. Ali A, Soares SC, Santos AR, Guimaraes LC, Barbosa E, et al. (2012) *Campylobacter fetus* subspecies: Comparative genomics and prediction of potential virulence targets. Gene 508: 145-156.

4. Alvarez-Martinez CE CP. (2009) Biological diversity of prokaryotic type IV secretion systems. Microbiol Mol Biol Rev. 73: 775-808.

5. Astrid F. (2010) Characterization of the putative *Campylobacter fetus venerealis* effector proteins Fic1 and Fic2 (unpublished master dissertation). Retrieved from University Graz, Austria.

6. Atabay, H.I, and Corry, J.E. (1998) The isolation and prevalence of *Campylobacters* from dairy cattle using a variety of methods. J.Appl.Microbiol 84: 733-740.

7. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, et al. (2008) The RAST server: Rapid annotations using subsystems technology. BMC Genomics 9: 1471-2164.

8. Bacon, D. J., AIm, R. A., Burr, D. H. Hu, L., Kopecko, et al. (2000) Involvement of a plasmid in the virulence of*Campylobacter jejuni*. Infect. Immun. 68: 4384-4390.

9. Bannantine JP, Wu CW, Hsu C, Zhou S, Schwartz DC, et al. (2012) Genome sequencing of ovine isolates of *Mycobacterium avium* subspecies *paratuberculosis* offers insights into host association. BMC Genomics. 13: 1471-2164.

10. Bjourson AJ, Stone CE, Cooper JE. (1992) Combined subtraction hybridization and polymerase chain reaction amplification procedure for isolation of strain-specific *Rhizobium* DNA sequences. Appl Environ Microbiol 58: 2296-2301.

11. Blaser MJ. (1998) *Campylobacter fetus-emerging* infection and model system for bacterial pathogenesis at mucosal surfaces. Clin Infect Dis 27: 256-258.

12. Blaser, M.J., Newell, D.G., Thompson, S.A., and Zechner, E.L. (2008) Pathogenesis of *Campylobacter fetus* infections. In: I. Nachamkin, C.M. Szymanski, and M.J. Blaser, editor. Washington, D.C: ASM Press. pp. 401-428.

13. Blaser MJ, Gotschlich EC. (1990) Surface array protein of *Campylobacter fetus* cloning and gene structure. J Biol Chem 265: 14529-14535.

14. Blom K, Patton CM, Nicholson MA, Swaminathan B. (1995) Identification of *Campylobacter fetus* by PCR-DNA probe method. J Clin Microbiol 33: 1360-1362.

15. Bogush ML, Velikodvorskaya TV, Lebedev YB, Nikolaev LG, Lukyanov SA, et al. (1999) Identification and localization of differences between *Escherichia coli* and *Salmonella typhimurium* genomes by suppressive subtractive hybridization. Mol.Gen.Genet 262: 721-729.

16. Boto L. (2010) Horizontal gene transfer in evolution: Facts and challenges. Proc Biol Sci 227: 819-827.

17. Brooks BW, Devenish J, Lutze-Wallace CL, Milnes D, Robertson RH, et al. (2004) Evaluation of a monoclonal antibody-based enzyme-linked immunosorbent assay for detection of *Campylobacter fetus* in bovine preputial washing and vaginal mucus samples. Vet Microbiol. 103: 77-84.

18. Brown CT, Fishwick LK, Chokshi BM, Cuff MA, Jackson JM 4th, et al. (2011) Whole-genome sequencing and phenotypic analysis of *Bacillus subtilis* mutants following evolution under conditions of relaxed selection for sporulation. Appl.Environ.Microbiol. 77: 6867-6877.

19. Butzler J. (1978) Infection with *Campylobacters*. In: J. D. Williams, editor. Heineman, London: Modem topics in infection. pp. 214-239.

20. Cai W, Jing J, Irvin B, Ohler L, Rose E, et al. (1998) High-resolution restriction maps of bacterial artificial chromosomes constructed by optical mapping. Proc Natl Acad Sci U S A 95: 3390-3395.

21. Carolyn V. Mander. (2001) The application of subtractive hybridisation to detect intra-species variation in *Campylobacter jejuni*. (unpublished master dissertation). University of Canterbury, New Zealand.

22. Carvalho AC, Ruiz-Palacios GM, Ramos-Cervantes P, Cervantes LE, Jiang X, et al. (2001) Molecular characterization of invasive and noninvasive *Campylobacter jejuni* and *Campylobacter coli* isolates. J Clin Microbiol 39: 1353-1359.

23. Cascales E, Christie PJ. (2003) The versatile bacterial type IV secretion systems. Nature reviews.Microbiology 1: 137-149.

24. Céline A, Gilles M, Guidon A, Gerardo T, Jean-Jacques M, et al. (2000) Solid phase DNA amplification: Characterisation of primer attachment and amplification mechanisms. Nucleic Acids Res 28: e87.

25. Chaban B, Chu S, Hendrick S, Waldner C, Hill JE. (2012) Evaluation of a *Campylobacter fetus* subspecies venerealis real-time quantitative polymerase chain reaction for direct analysis of bovine preputial samples. Can J Vet Res 76: 166-173.

26. Chang, N., Taylor, D.E. (1990) Use of pulsed-field gel electrophoresis to size genomes of *Campylobacter* species and to construct a SalI map of *Campylobacter jejuni* UA580s. J Bacteriol 172: 5211-5217.

27. Chang W, Ogg JE. (1971) Transduction and mutation to glycine tolerance in *Vibrio fetus*. Am J Vet Res 32: 649-653.

28. Chen Q, Savarino SJ, Venkatesan MM. (2006) Subtractive hybridization and optical mapping of the *Enterotoxigenic Escherichia coli* H10407 chromosome: Isolation of unique sequences and demonstration of significant similarity to the chromosome of *E. coli* K-12. Microbiology. 152: 1041-1054.

29. Christie, P. J., K. Atmakuri, V. Krishnamoorthy, S. Jakubowski, and E. Cascales. (2005) Biogenesis, architecture, and function of bacterial type IV secretion systems. Annu. Rev. Microbiol. 59: 451-485.

30. Clark BL. (1971.) Review of bovine vibriosis. Aust.Vet.J. 47: 103-107.

31. Clark, V.L., and Bavoil, P.M., editor. (1994) Methods in enzymology. Bacterial pathogenesis, part A: Identification and regulation of virulence factors. volume 235: Academic Press.

32. Coebeil LB WA. (1978) Animal model for the study of genital secretory immune mechanism: Venereal vibriosis in cattle. Immunobiology of *Neisseria gonorrehoeae*. Brooks GF, Gotschlich EC, Holmes KK, Sawyer WD, Young FE, editor. Washington, DC.: Am Soc Microbiol. pp. 293-299.

33. Cottral GE. (1978) Manual of standardized methods for veterinary microbiology. In: Cottral GE, editor. Ithaca, N.Y: Cornell University Press. pp. 461-471.

34. Darling AC, Mau B, Blattner FR, Perna NT. (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14: 1394-1403.

35. de Boer, P., Wagenaar, J.A., Achterberg, R.P., van Putten, J.P., Schouls, L.M., and Duim, B. (2002) Generation of *Campylobacter jejuni* genetic diversity in vivo. Mol Microbiol 44: 351-359.

36. Devenish J, Brooks B, Perry K, Milnes D, Burke T, et al. (2005) Validation of a monoclonal antibody-based capture enzyme-linked immunosorbent assay for detection of *Campylobacter fetus*. Clin Diagn Lab Immunol 12: 1261-1268.

37. Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, et al. (1996) Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci U S A 93: 6025-6030.

38. Dingle KE, Blaser MJ, Tu ZC, Pruckler J, Fitzgerald C, et al. (2010) Genetic relationships among reptilian and mammalian *Campylobacter fetus* strains determined by multilocus sequence typing. J Clin Microbiol 48: 977-980.

39. Dressman, D., Yan, H., Traverso, G., Kinzler, K.W. & vogelstein, B. (2003) Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. USA. Proc. Natl. Acad. Sci. 100: 8817-8822.

40. Drost, M. & Thomas, P.G.A. (1996) Diseases of the reproductive system in large animal, Internal medicine, In: B.P. Smith., editor. St. Louis: Mosby pp. 1519-1565.

41. Dubreuil JD, Kostrzynska M, Austin JW, Trust TJ. (1990) Antigenic differences among *Campylobacter fetus* S-layer proteins. J Bacteriol 172: 5035-5043.

42. Duim B, Vandamme PA, Rigter A, Laevens S, Dijkstra JR, et al. (2001) Differentiation of *Campylobacter* species by AFLP fingerprinting. Microbiology 147: 2729-2737.

43. Dworkin J, Blaser MJ. (1997) Molecular mechanisms of campylobacter fetus surface layer protein expression. Mol Microbiol 26: 433-440.

44. Dworkin J, Shedd OL, Blaser MJ. (1997) Nested DNA inversion of *Campylobacter fetus* S-layer genes is recA dependent. J Bacteriol 179: 7523-7529.

45. Dybvig K. (1993) DNA rearrangements and phenotypic switching in prokaryotes. Mol Microbiol 10: 465-471.

46. Eaglesome, M.D. & Garcia, M.M. (1992) Microbial agents associated with bovine genital tract infections and semen. part I. *Brucella abortus, leptospira, Campylobacter fetus* and *Tritrichomonas foetus*. Veterinary Bulletin 62: 743-755.

47. Edmonds P, Patton CM, Barrett TJ, Morris GK, Steigerwalt AG, et al. (1985) Biochemical and genetic characteristics of atypical *Campylobacter fetus* subsp. fetus strains isolated from humans in the united states. J Clin Microbiol 21: 936-940.

48. Enright MC, Spratt BG. (1998) A multilocus sequence typing scheme for *Streptococcus pneumoniae*: Identification of clones associated with serious invasive disease. Microbiology 144: 3049-3060.

49. Figueiredo et al. (2002) Evaluation of direct fluorescent antibody test for the diagnosis of bovine genital campylobacteriosis. Microbiologia 44: 118-123.

50. Finlay BB FS. (1997) Common themes in microbial pathogenicity revisited. Microbiol Mol Biol Rev 61: 136-169.

51. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae*. Science 269: 496-498.

52. Florent A. (1959) Le dex vibriosis genitales: La vibriose due à *V. fetus venerealis* at la bivriose d'origine intestinale due à *V. fetus intestinalis*. mededelinge veearts school. 3: 1-60.

53. Fouts DE, Mongodin EF, Mandrell RE, Miller WG, Rasko DA, et al. (2005) Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. PLoS Biol 3: 1-15.

54. Francioli P, Herzstein J, Grob JP, Vallotton JJ, Mombelli G, Glauser MP. (1985) *Campylobacter fetus* subspecies*fetus* bacteremia. Arch Intern Med 145: 289-292.

55. Garcia Guerra D. (1993) The theory of contagion in French pathology prior to the work of Pasteur. Cuad Complut Hist Med Cienc 1: 239-261.

56. Garcia, M.M., Eaglesome, M.D. & Rigby, C. (1983) *Campylobacters* important in veterinary medicine. Veterinary Bulletin 53: 793-811.

57. Garcia MM, Ruckerbauer GM, Eaglesome MD, Boisclair WE. (1983) Detection of *Campylobacter fetus* in artificial insemination bulls with a transport enrichment medium. Can J Comp Med 47: 336-340.

58. Gogarten, J.P., Doolittle, W.F., and Lawrence, J.G. (2002) Prokaryotic evolution in light of gene transfer. Mol Biol Evol 19: 2226-2238.

59. Gorkiewicz G, Kienesberger S, Schober C, Scheicher SR, Gully C, et al. (2010) A genomic island defines subspecies-specific virulence features of the host-adapted pathogen *Campylobacter fetus* subsp. *venerealis*. J Bacteriol 192: 502-517.

60. Griffiths, P. L. & Park, R. W. A. (1990) *Campylobacters* associated with human diarrhoea1 disease. J Appl Bacteriol 69: 281-301.

61. Grogono-Thomas, R., M. J. Blaser, M. Ahmadi, and D. G. Newell. (2003) Role of S-layer protein antigenic diversity in the immune responses of sheep experimentally challenged with *Campylobacter fetus* subsp.*fetus*. Infec Immun 71: 147-154.

62. Hacker, J., and J. B. Kaper. (2000) Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. 54: 641-679.

63. Hacker J, Blum-Oehler G, Muhldorfer I, Tschape H. (1997) Pathogenicity islands of virulent bacteria: Structure, function and impact on microbial evolution. Mol Microbiol 23: 1089-1097.

64. Hara E, Kato T, Nakada S, Sekiya S, Oda K. (1991) Subtractive cDNA cloning using oligo(dT)30- latex and PCR: Isolation of cDNA clones specific to undifferentiated human embryonal carcinoma cells. Nucleic Acids Res. 19: 7097-7104.

65. Harvey, S.M., and Greenwood, J.R. (1983) Relationships among catalase-postive *Campylobacter* determined by deoxyribonucleic acid-deoxyribonucleic acid hybridization. J.Syst.Bacteriol 33: 275-284.

66. Hepworth PJ, et al. (2011) Genomic variations define divergence of water/wildlife- associated *Campylobacter jejuni* niche specialists from common clonal complexes. Environ. Microbiol. 13: 1549-1560.

67. Hepworth PJ, Leatherbarrow H, Hart CA, Winstanley C. (2007) Use of suppression subtractive hybridisation to extend our knowledge of genome diversity in *Campylobacter jejuni*. BMC Genomics 8: 110.

68. Hewson PI, Lander KP, Gill KP. (1985) Enzyme-linked immunosorbent assay for antibodies to *Campylobacter fetus* in bovine vaginal mucus. Res Vet Sci 38: 41-45.

69. Hoerlein, A.B., T. Kramer and E.J. Carroll. (1981) Vibriosis in range cattle. J.Am.Vet.Med.Assoc. 144: 146-151.

70. Hoffer MA. (1981) Bovine campylobacteriosis: A review. Can.Vet.J. 22: 327-330.

71. Hou YM. (1999) Transfer RNAs and pathogenicity islands. Trends Biochem Sci 24: 295-298.

72. Hsiao WWL, Ung K, Aeschliman D, Bryan J, Brett Finlay B, et al. (2005) Evidence of a large novel gene pool associated with prokaryotic genomic islands. PLoS genet 1: 540-550.

73. Hu, L., and D. J. Kopecko. (2000) Interactions of *Campylobacter* with eukaryotic cells: Gut luminal colonization and mucosal invasion mechanisms. In: I. Nachamkin and M. J. Blaser, editor. Washington, DC.: ASM Press. pp. 191-215.

74. Huang X, Li Y, Niu Q, Zhang K. (2007) Suppression subtractive hybridization (SSH) and its modifications in microbiological research. Appl Microbiol Biotechnol 76: 753-760.

75. Hum S, Hornitzky M, Berg T. (2009) Bovine genital campylobacteriosis. Australia and New Zealand standard diagnostic procedures . In Press.

76. Hum S. (2007) Vibriosis of cattle. [NSW]: Department of primary industries, Australia. Primefact no. 451.

77. Hum S, Brunner J, McInnes A, Mendoza G, Stephens J. (1994) Evaluation of cultural methods and selective media for the isolation of *Campylobacter fetus* subsp *venerealis* from cattle. Aust Vet J 71: 184-186.

78. Hum S, Quinn K, Brunner J, On SL. (1997) Evaluation of a PCR assay for identification and differentiation of *Campylobacter fetus* subspecies. Aust Vet J 75: 827-831.

79. Hum S, Stephens LR, Quinn C. (1991) Diagnosis by ELISA of bovine abortion due to *Campylobacter fetus*. Aust Vet J 68: 272-275.

80. Ichiyama et al. (1998) *Campylobacter fetus* subspecies*fetus* cellulitis associated with bacteremia in debilitated hosts. clin infect Dis*27*: 252-255.

81. Iraola G, Hernandez M, Calleros L, Paolicchi F, Silveyra S, et al. (2012) Application of a multiplex PCR assay for *Campylobacter fetus* detection and subspecies differentiation in uncultured samples of aborted bovine fetuses. J Vet Sci 13: 371-376.

82. Irons PC, Schutte AP, van der Walt, M L. (2004) Genital campylobacteriosis in cattle. infectious diseases of livestock with special reference to southern Africa. Coetzer J A W, Tustin R C, editor. pp. 1459-1468.

83. J. Duan, J.J. Heikkila, B.R. Glick. (2010) Sequencing a bacterial genome: An overview, Current research, technology and education topics in applied microbiology and microbial biotechnology, department of biology, University of Waterloo, Canada. In: Mendez-Vilas A, editor: Formatex Research Center. pp. 1443-1451.

84. Janke B, Dobrindt U, Hacker J, Blum-Oehler G. (2001) A subtractive hybridisation analysis of genomic differences between the *Uropathogenic E. coli* strain 536 and the *E. coli* K-12 strain MG1655. FEMS Microbiol Lett 199: 61-66.

85. Jones, R. L., M. A. Davis, and H. Vonbyern. (1985) Cultural procedures for the isolation of campylobacter fetus subsp. venerealis from preputial secretions and the occurrence of antimicrobial resistance. Proc. Annu. Meet. Am. Assoc. Vet. Lab. Diagn. 28: 225-238.

86. Juhas, M., van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W., and Crook, D.W. (2009) Genomic islands: Tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev 33: 376-393.

87. Karl V. Voelkerding, Shale A. Dames, and Jacob D. Durtschi. (2009) Next-generation sequencing: From basic research to diagnostics. Clin Chem. 55: 641-658.

88. Ketley JM. (1997) Pathogenesis of enteric infection by*Campylobacter*.. Microbiology 143: 5-21.

89. Kienesberger S, Gorkiewicz G, Wolinski H, Zechner EL. (2011) New molecular microbiology approaches in the study of *Campylobacter fetus*. Microb Biotechnol 4: 8-19.

90. Kienesberger S, Schober Trummler C, Fauster A, Lang S, Sprenger H, et al. (2011) Interbacterial macromolecular transfer by the *Campylobacter fetus* subsp. *venerealis* type IV secretion system. J Bacteriol 193: 744-758.

91. Koonin, E.V., and Wolf, Y.I. (2008) Genomics of bacteria and archaea: The emerging dynamic view of the prokaryotic world. Nucleic Acids Res 36: 6688-6719.

92. Lander KP. (1990) The application of a transport and enrichment medium to the diagnosis of *Campylobacter fetus* infections in bulls. Br Vet J 146: 334-340.

93. Li H DR. (2009) Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 25: 1754-1769.

94. Lin J, Qi R, Aston C, Jing J, Anantharaman TS, et al. (1999) Whole-genome shotgun optical mapping of *Deinococcus radiodurans*. Science 285: 1558-1562.

95. Luengpailin J, Doyle RJ. (2000) Glycine prevents the phenotypic expression of *Streptococcal* glucan-binding lectin. Biochim Biophys Acta 1474: 212-218.

96. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, et al. (1998) Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A. 95: 3140-3145.

97. Martinez AH, Bardon JC, Nosoda BP, Cordeviola JM, Sarmiento F, et al. (1986) Herd diagnosis on trichomoniasis and campylobacteriosis in bovine utilizing the empty cow as indicator. Vet.Arg 111: 962-966.

98. McFadyean, J. & Stockman, S. (1913) Report of the departmental committee appointed by the board of agriculture and fisheries to inquire into epizootic abortion, appendix to part III, abortion in sheep.

99. McGoldrick A, Chanter J, Gale S, Parr J, Toszeghy M, Line K. (2013) Real time PCR to detect and differentiate *Campylobacter fetus* subspecies *fetus* and *Campylobacter fetus* subspecies *venerealis*. J Microbiol Methods 94: 199-204.

100. McMillen L, Fordyce G, Doogan VJ, Lew AE. (2006) Comparison of culture and a novel 5' taq nuclease assay for direct detection of *Campylobacter fetus* subsp. *venerealis* in clinical specimens from cattle. J Clin Microbiol 44: 938-945.

101. Médigue C MI. (2007) Annotation, comparison and databases for hundreds of bacterial genomes. Research in Micobiology. 158: 724-736.

102. Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, et al. (2011) Prospective genomic characterization of the german *Enterohemorrhagic Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. PloS one 6: e22751.

103. Miller WG, On SL, Wang G, Fontanoz S, Lastovica AJ, et al. (2005) Extended multilocus sequence typing system for *Campylobacter coli*, *C. lari*, *C. upsaliensis*, and *C. helveticus* . J Clin Microbiol 43: 2315-2329.

104. Misawa N, Allos BM, Blaser MJ. (1998) Differentiation of *Campylobacter jejuni* serotype O19 strains from non-O19 strains by PCR. J Clin Microbiol 36: 3567-3573.

105. Mishu, B., Patton, C.M. & Tauxe, R.V. (1992) Clinical and epidemiologic features of non-jejuni, non-coli *Campylobacter* species, in *Campylobacter jejuni:* current status and future trends. In: I. Nachamkin, M.J. Blaser & L.S. Tompkins, editor. Washington: American Society of Microbiology. pp. 31-41.

106. Mitra, R.D. & Church, G.M. (1999) In situ localized amplification and contact replication of many individual DNA molecules. Nucleic Acids Res 27: e34-e34.

107. Mitra RD, Shendure J, Olejnik J, Edyta-Krzymanska-Olejnik, Church GM. (2003) Fluorescent in situ sequencing on polymerase colonies. Anal Biochem 320: 55-65.

108. Moolhuijzen PM, Lew-Tabor AE, Wlodek BM, Aguero FG, Comerci DJ, et al. (2009) Genomic analysis of *Campylobacter fetus* subspecies: Identification of candidate virulence determinants and diagnostic assay targets. BMC Microbiol 9.

109. Mshelia GD. (2008) Bovine venereal campylobacteriosis and reproductive efficiency in the lake chad basin of Nigeria. (Unpublished doctoral dissertation). University of Maiduguri, Nigeria.

110. Mshelia GD, Amin JD, Woldehiwet Z, Egwu GO, Murray RD,. (2007) Bovine venereal campylobacteriosis: An overview. CAB rev. Retrieved from

http://Www.cabi.org/cabreviews/?loadmodule=review&page=4051&reviewid=80000&site=167. Perspect Agr Vet Sci Nutr Nat Resour 2: 14.

111. Mshelia GD, Amin JD, Woldehiwet Z, Murray RD, Egwu GO. (2010) Epidemiology of bovine venereal campylobacteriosis: Geographic distribution and recent advances in molecular diagnostic techniques. Reproduction in Domestic Animals 45: e221-e230.

112. Muller W, Hotzel H, Schulze F. (2003) Identification and differentiation of *Campylobacter fetus* subspecies by PCR. Dtsch Tierarztl Wochenschr 110: 55-59.

113. Myers EW HX. (1992) An O (N2 log N) restriction map comparison and search algorithm. Bull Math Biol. 54: 599-618.

114. Nachamkin I, Allos BM, Ho T. (1998) *Campylobacter* species and guillain-barré syndrome. Clin Microbiol Rev : 555-567.

115. Newell, D. G. & 6 other authors. (2000) New developments in the subtyping of *Campylobacter* strains. Washington, DC: American Society for Microbiology.

116. Oelschlaeger TA, Hacker J. (2004) Impact of pathogenicity islands in bacterial diagnostics. APMIS 112: 930-936.

117. OIE. (2010) Bovine genital campylobacteriosis. Paris: Office international des épizooties, 2010. Rep Carcinog Backgr Doc (10-5981).

118. OIE. (2008) Bovine genital campylobacteriosis. OIE terrestrial manual, chapter 2. Rep Carcinog Backgr Doc (8-5978): 661-670.

119. On SL. (2001) Taxonomy of *Campylobacter*, *Arcobacter*, *Helicobacter* and related bacteria: Current status, future prospects and immediate concerns. Symp Ser Soc Appl: 1S-15S.

120. On SL, Harrington CS. (2001) Evaluation of numerical analysis of PFGE-DNA profiles for differentiating *Campylobacter fetus* subspecies by comparison with phenotypic, PCR and 16S rDNA sequencing methods. J Appl Microbiol 90: 285-293.

121. Owczarzy R, Tataurov AV, Wu Y, Manthey JA, McQuisten KA, et al. (2008) IDT SciTools: A suite for analysis and design of nucleic acid oligomers. Nucleic Acids Res 36: W163-W169.

122. Oyarzabal OA, Wesley IV, Harmon KM, Schroeder-Tucker L, Barbaree JM, et al. (1997) Specific identification of *Campylobacter fetus* by PCR targeting variable regions of the 16S rDNA. Vet Microbiol 58: 61-71.

123. Panaccio, M., Stephens, L. and Widders, P. (1993) Field evaluation of a PCR-based test for the diagnosis of bovine campylobacteriosis. In: Panaccio, M., Stephens, L. and Widders, P., editor. Australia: Dairy Research and Development Corporation. pp. 1-39.

124. Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, et al. (2000) The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyper variable sequences. Nature 403: 665-668.

125. Pearson, B.M., Gaskin, D.J., Segers, R.P., Wells, J.M., Nuijten, P.J., and van Vliet, A.H. (2007) The complete genome sequence of *Campylobacter jejuni* strain 81116 (NCTC11828). J Bacteriol 189: 8402-8403.

126. Penner JL. (1988) The genus *Campylobacter*: A decade of progress. Clin Microbiol Rev: 157-172.

127. Perez-Perez, G.I., Blaser, M.J., and Bryner, J.H. (1986) Lipopolysaccharide structures of *Campylobacter fetus* are related to heat-stable serogroups. Infect Immun 51: 209-212.

128. Poly, F., Read, T., Tribble, D.R., Baqar, S., Lorenzo, M., and Guerry, P. (2007) Genome sequence of a clinical isolate of *Campylobacter jejuni*from thailand. Infec Immun 75: 3425-3433.

129. Pop M, Kosack DS, Salzberg SL. (2004) Hierarchical scaffolding with bambus. Genome Research 14: 149-159.

130. Prendergast, M. M., Willison, H. J., and Moran, A. P. (1999) Human monoclonal immunoglobin M antibodies to ganglioside GMI show diverse cross-reactivities with lipopolysaccharides of *Campylobacter jejuni* Strains associated with guillain-barre syndrome. Infect Immun 67: 3698-3701.

131. Quini, P.J., Carter, M.E., Markey, B. & Carter, G.R. (1994) *Campylobacter* species. Clin Vet Microbiol : 268-272.

132. Rebrikov DV, Britanova OV, Gurskaya NG, Lukyanov KA, Tarabykin VS, and Luckanov SA. (2000) Mirror orientation selection (MOS): A method for eliminating false positive clones from libraries generated by suppression subtractive hybridization. Nucleic Acids Res 28: e90.

133. Reckseidler SL, DeShazer D, Sokol PA, Woods DE. (2001) Detection of bacterial virulence genes by subtractive hybridization: Identification of capsular polysaccharide of *Burkholderia pseudomallei* as a major virulence determinant. Infect Immun 69: 34-44.

134. Rêgo AT, Chandran V, Waksman G. (2010) Two-step and one-step secretion mechanisms in gram-negative bacteria: Contrasting the type IV secretion system and the chaperone-usher pathway of pilus biogenesis. Biochem 425: 475-488.

135. Reslewic S, Zhou S, Place M, Zhang Y, Briska A, et al. (2005) Whole-genome shotgun optical mapping of *Rhodospirillum rubrum*. Appl Environ Microbiol. 71: 5511-5522.

136. Rettig PJ. (1979) Campylobacter infections in human beings. J. Periatr 94: 855-864.

137. Ribeiro FJ, Przybylski D, Yin S, Sharpe T, Gnerre S, et al. (2012) Finished bacterial genomes from shotgun sequence data. Genome Res. 22: 2270-2277.

138. Salama SM, Garcia MM, Taylor DE. (1992) Differentiation of the subspecies of *Campylobacter fetus* by genomic sizing. Int J Syst Bacteriol 42: 446-450.

139. Samuelson, J.D., and Winter, A.J. (1996) Bovine vibriosis: The nature of the carrier state in the bull. J Infect Dis. 116: 581-592.

140. Sanger F. (1988) Sequences, sequences, and sequences. Annu Rev Biochem 57: 1-28.

141. Savelkoul PH, Aarts HJ, de Haas J, Dijkshoorn L, Duim B, et al. (1999) Amplified-fragment length polymorphism analysis: The state of an art. J Clin Microbiol 37: 3083-3091.

142. Schmidt H, Hensel M. (2004) Pathogenicity islands in bacterial pathogenesis. Clin Microbiol Rev 17: 14-56.

143. Schulze F, Bagon A, Muller W, Hotzel H. (2006) Identification of *Campylobacter fetus* subspecies by phenotypic differentiation and PCR. J Clin Microbiol 44: 2019-2024.

144. Schumann W. (2007) Thermosensors in Eubacteria: Role and evolution. J Biosci 32: 549-557.

145. Schwan WR, Briska A, Stahl B, Wagner TK, Zentz E, et al. (2010) Use of optical mapping to sort *Uropathogenic Escherichia coli* strains into distinct subgroups. Microbiol 156: 2124-2135.

146. Sergey L, Denis R, Anton B. (2007) Nucleic acids hybridization chapter 3:Suppression subtractive hybridization by springer. In: A. Buzdin and S. Lukyanov, editor. pp. 53-84.

147. Seyyal, A.K., Turan, N., Kemal, A.K., Ileri, K. & Ilgaz, A. (2000) Use of ELISA, IFA, and avidin-biotin staining for the diagnosis of bovine genital campylobacteriosis. Turk J Vet Anim Sci 24: 113-121.

148. Shendure Jea. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309: 1728-1732.

149. Shendure J, Ji H. (2008) Next-generation DNA sequencing. Nat Biotechnol 26: 1135-1145.

150. Skirrow MB. (1994) Diseases due to *Campylobacter*, *Helicobacter* and related bacteria. J Comp Pathol 111: 113-149.

151. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, et al. (1986) Fluorescence detection in automated DNA sequence analysis. Nature. 321: 674-679.

152. Soares SC, AbreuViní VAC, Ramos RTJ, Cerdeira L, Silva A, et al. (2012) PIPS: Pathogenicity island prediction software. PLoS ONE 7: e30848.

153. Stothard P WD. (2006) Automated bacterial genome analysis and annotation. Curr Opin Microbiol. 9: 505-510.

154. Straus, D., and Ausubel, F. (1990) Genomic subtraction for cloning DNA corresponding to deletion mutations. Proc Natl Acad Sci U S A. 87: 1888-1893.

155. Stynen APR, Lage AP, Moore RJ, Rezende AM, de Resende VDda S, et al. (2011) Complete genome sequence of type strain *Campylobacter fetus* subsp. *venerealis* NCTC 10354T. J Bacteriol 193: 5871-5872.

156. Swerdlow, H., Wu, S.L., Harke, H. & Dovichi, N.J. (1990) Capillary gel electrophoresis for DNA sequencing. laser-induced fluorescence detection with the sheath flow cuvette. J. Chromatogr. 516: 67.

157. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739.

158. Taylor DE. (1992) Genetics of *Campylobacter and Helicobacter*. Annu Rev Microbiol 46: 35-64.

159. The DOE Joint Genome Institute. GOLD: Genome online database. Retrieved from http://www.genomesonline.org.

160. Thompson SA. (2002) *Campylobacter* surface-layers (S-layers) and immune evasion. Ann Periodontol 7: 43-53.

161. Thompson SA, Shedd OL, Ray KC, Beins MH, Jorgensen JP, et al. (1998) *Campylobacter fetus* surface layer proteins are transported by a type I secretion system. J Bacteriol 180: 6450-6458.

162. Timoney, J.F., Gillespie, J.H., Scott, F.W. & Barlough, J.E. (1988) The genus *Campylobacter*, in Hagan and bruner's microbiology and infectious diseases of domestic animals. Timoney, J.F., Gillespie, J.H., Scott, F.W. & Barlough, J.E, editor. Ithaca: Cornell University Press pp. 153-160.

163. Tinsley C., and Nassif, X. (1996) Analysis of the genetic differences Between *Neisseria meningitidis* and *Neisseria gonorrhoeae:* Two closely related bacteria expressing two different pathogenicities. Proc Natl Acad Sci U S A. 93: 11109-11114.

164. Tracy Schmidt. (2008) Detection of *Campylobacter fetus* in bovine preputial scrapings using PCR and culture assays (Unpublished master dissertation). University of Pretoria, South Africa.

165. Tu ZC, Dewhirst FE, Blaser MJ. (2001) Evidence that the *Campylobacter fetus* sap locus is an ancient genomic constituent with origins before mammals and reptiles diverged. Infect Immun 69: 2237-2244.

166. Tu ZC, Eisner W, Kreiswirth BN, Blaser MJ. (2005) Genetic divergence of *Campylobacter fetus* strains of mammal and reptile origins. J Clin Microbiol 43: 3334-3340.

167. Van Bergen MAP. (2005) Subspecies differentiation and typing of *Campylobacter fetus*. (Unpublished doctoral dissertation), University of Utrecht, Netherland.

168. van Bergen MA, Dingle KE, Maiden MC, Newell DG, van der Graaf-Van Bloois L, et al. (2005) Clonal nature of *Campylobacter fetus* as defined by multilocus sequence typing. J Clin Microbiol 43: 5888-5898.

169. van Bergen MA, Simons G, van der Graaf-van Bloois L, van Putten JP, Rombout J, et al. (2005) Amplified fragment length polymorphism based identification of genetic markers and novel PCR assay for differentiation of *Campylobacter fetus* subspecies. J Med Microbiol 54: 1217-1224.

170. van Bergen MA, van der Graaf-van Bloois L, Visser IJ, van Putten JP, Wagenaar JA. (2006) Molecular epidemiology of *Campylobacter fetus* subsp. fetus on bovine artificial insemination stations using pulsed field gel electrophoresis. Vet Microbiol 112: 65-71.

171. van der Graaf-van Bloois L, van Bergen MA, van der Wal FJ, de Boer AG, Duim B, Schmidt T, Wagenaar JA. (2013) Evaluation of molecularassays for identification *Campylobacter fetus* species and subspecies and development of a *C. fetus* specific real-time PCR assay. J Microbiol Methods S0167: 180-200.

172. Vandeplassche, M., Florent, A., Bouters, R., Huysmin, A., Brone, E. & Dekeyser. (1963) The pathogenesis, epidemiology and treatment of *Vibrio fetus* infection in cattle. C.R. Recherches, 29: 1-90.

173. Vargas AC, Costa MM, Vainstein MH, Kreutz LC, Neves JP. (2003) Phenotypic and molecular characterization of bovine *Campylobacter fetus* strains isolated in Brazil. Vet Microbiol 93: 121-132.

174. Véron M, Chatelain R. (1973) Taxonomic study of the genus *Campylobacter* sebald and véron and designation of the neotype strain for the type species, *Campylobacter fetus* (smith and taylor) sebald and véron. Int J Syst Bacteriol 23: 122-134.

175. Wagenaar JA, van Bergen MA, Newell DG, Grogono-Thomas R, Duim B. (2001) Comparative study using amplified fragment length polymorphism fingerprinting, PCR genotyping, and phenotyping to differentiate *Campylobacter fetus* strains isolated from animals. J Clin Microbiol 39: 2283-2286.

176. Wallden K, Rivera-Calzada A, Waksman G. (2010) Type IV secretion systems: Versatility and diversity in function. Cell Microbiol 12: 1203-1212.

177. Wang E, Garcia MM, Blake MS, Pei Z, Blaser MJ. (1993) Shift in S-layer protein expression responsible for antigenic variation in *Campylobacter fetus*. J Bacteriol 175: 4979-4984.

178. Wang G, Clark CG, Taylor TM, Pucknell C, Barton C, et al. (2002) Colony multiplex PCR assay for identification and differentiation of *Campylobacter jejuni*, *C. coli*, *C. lari*, *C. upsaliensis*, and *C. fetus* subsp. *fetus*. J Clin Microbiol 40: 4744-4747.

179. Wayne et al. (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int.J.Syst.Bacteriol 37: 463-646.

180. WHO. (2011) Campylobacter. Retrieved from http://Www.who.int/mediacentre/factsheets/fs255/en/index.html

181. Wiesner, R.S., Hendrixson, D.R., and DiRita, V.J. (2003) Natural transformation of *Campylobacter jejuni* requires components of a type II secretion system. J Bacteriol 185: 5408-5418.

182. Willoughby K, Nettleton PF, Quirie M, Maley MA, Foster G, et al. (2005) A multiplex polymerase chain reaction to detect and differentiate *Campylobacter fetus* subspecies fetus and *Campylobacter fetus* -species *venerealis*: Use on UK isolates of *C. fetus* and other *Campylobacter* spp. J Appl Microbiol 99: 758-766.

183. Wilson DJ, Gabriel E, Leatherbarrow AJH, Cheesbrough J, Gee S, et al. (2009) Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. Mol Biol Evol 26.

184. Winstanley C. (2002) Spot the difference: Applications of subtractive hybridisation to the study of bacterial pathogens. J Med Microbiol 51: 459-467.

185. Winter AJ. (1982) Microbial immunity in the reproductive tract. J Am Vet Med Assoc 181: 1069-1073.

186. Wu CW, Schramm TM, Zhou S, Schwartz DC, Talaat AM. (2009) Optical mapping of the *Mycobacterium avium* subspecies *paratuberculosis* genome. BMC Genomics 10: 10-25.

187. Yu S, Fearnhead P, Holland BR, Biggs P, Maiden M, French N. (2012) Estimating the relative roles of recombination and point mutation in the generation of single locus variants in *Campylobacter jejuni* and *Campylobacter coli*. J Mol Evol. 74: 273-280.

188. Zhang YL, Ong CT, Leung KY. (2000) Molecular analysis of genetic differences between virulent and avirulent strains of *Aeromonas hydrophila* isolated from diseased fish. Microbiol 146: 999-1009.

189. Zhou S, Kile A, Bechner M, Place M, Kvikstad E, et al. (2004) Single- molecule approach to bacterial genomic comparisons via optical mapping. J Bacteriol 186: 7773-7782.

Appendices

1. Cell Density experiment Instructions

This experiment was done to assist in choosing the appropriate culture OD prior to bacterial genomic extraction. Briefly, TG-1 *E. coli* strain was inoculated in 3ml 2YT liquid medium and cultured overnight. The next day 1ml of the overnight culture was inoculated into 100ml media, and at the end of the day, when bacteria was in the mid log phase, the OD was measured at 600Å. When OD reached 0.5, the cultured media were placed in the fridge overnight. Next morning, bacterial cells were counted using a haemocytometer and the total number of cells was calculated. Next, the cultured media were transferred into centrifuge tube and underwent centrifugation at 5000 rpm for 10 mins. The supernatant was removed and the pellet was re-suspended into 1ml media. A two-fold dilution series for a total of 8 dilutions was generated in culture media and OD was measured at 600Å for each dilution. A minimum of three OD readings was recorded; though, the cell counts was taken only from the first reading. The data were input into Excel and used to generate a graph of absorbance vs. cell density.

2. Lists of Primers

Table I. PCR primers used for this study.

Primer name	Sequnce (5' --> 3')	Base Length	Tm °C	Accession#	Locus specificity	Strain specificity	Amplicon	Comment
C412F	GGATGACACTTTTCGGAGC	19	53.6	AL111168	16S rRNA	All Campy sp.	816 bp	For checking gDNA integrity
C1228R	CATTGTAGCACGTGTGTC	18	51.2	AL111168				
23SF	TATACCGGTAAGGAGTGCTGGAG	23	57.6	Z29326	23S rRNA	All Campy sp.	650bp	For checking gDNA integrity
23SR	ATCAATTAACCTTCGAGCACCG	22	55.4	Z29326				
MG3F	GGTAGCCGCAGCTGCTAAGAT	21	59.6	EU443150	cstA	C. fetus	764 bp	Confirming C. fetus identification. Amplicon size NOT 960 bp as originally reported. Note that in complete genome forward primer is in reverse orientation (Hut et al 1997)
MG4R	TAGCTACAATAACGACAACT	20	48.4	EU443150				
CFF	GCAAATATAAATGTAAGCGGAGAG	24	52.1	AF048699	sapB2	CFF	435 bp	Reported to be specific for CFF but not confirmed in this study
CFR	TGCAGCGGCCCCACCTAT	18	62.5	AF048699				
SAF01	5ATGTTAAACAAAACRGATGT	20	46.7	AY45039	sapA	C. fetus	531 bp	Serotype determination
SAR01	ATCAAGATCACTAGCACTA	19	47.2	AY45039				
SBF01	TTCAGAGCTATTTATAGTTC	20	44	AY450400	sapB	C. fetus	505 bp	Serotype determination
SBR01	TCAACACTACTACTATYACTA	21	45.7	AY45040				
VirB9-1	CTTTGGGATATTTATTGCGTGC	22	52.6	EU443150	virB9/virB11	CFV	262bp	PICFV8 detection (GI 1 locus)
VirB11F	CCTGAGTTTGTAACAGATGAAGAGAT	26	54.6	EU443150				
nc1165g6F	ATGTTCTAGCMGAGCTTGG	19	53.4	EU443150	virD4	CFV	468 bp	PICFV8 detection (GI 2 locus)
TaxB3	CCATCAGGCTGATTTGCTTC	20	54.2	EU443150				
3DV3'	CGGTTGCGATGCTAATACACTA	22	54.9	EU443150	ORF21	CFV	262bp	PICFV8 detection (GI 3 locus)
3DV5'	TTTATCGCTTGGTAGCGTATTG	22	53.1	EU443150				
Top/traE4	ATTGCAAAAGAGTGTGAAGATA	22	50.2	EU443150	ORF27	CFV	841 bp	PICFV8 detection (GI 4 locus)
TraE7	GATCTAGTGGCAGGTGTTCC	20	55.1	EU443150				
VenSF	CTTAGCAGTTTGCGATATTGCCATT	25	56.3	EU443150	parA	CFV	142bp	PICFV8 detection (GI 5 locus within an ABC transporter permease gene)
VenSR	GCTTTTGAGATAACAATAAGAGCTT	25	51.8	EU443150				
C35F1	CTTAGTACGCTTAGGCAAGTA	21	51.6	This study	Clone35	CFV	668bp and 569bp	Identified by SSH
C35F2	AAGCGGCTATGAGTGATCTG	20	54.6	This study				
C35F3	AGATAGATGCTAAAGGCAGG	20	51.5	This study				
C35R1	TTCGCATCGTCTCTAGCGAT	20	55.9	This study				
C35R2	TTGGGACGCTTATCTCTAGT	20	52.6	This study				
1581F1	CGGTATGTAGCYGGWTCTCTAAAC	24	55.8	This study	PI10-ORF1581	CFV	449bp	Identified by WGS
1581F2	ATCTACAGGCACAACAGGCTT	21	56.6	This study				
1582R1	TTTGGARARATGGCAAACAAGA	22	53.3	This study				
548F	TTGAATGGATCTGCGATGATT	21	52.3	This study	PI5-ORF548	CFV	385bp	Identified by WGS
548R	GGCTACAAGCTCAGCAGT	18	54.9	This study				
576F	TGGACGAAGCAAGCCGATAT	20	56.6	This study	PI5-ORF576	CFV	415bp	Identified by WGS
576R	CTGCGACTAATCCAGTGACCA	21	56.7	This study				
581F	GCTCATAGCACTCCCAGCCTTA	22	58.7	This study	PI5-ORF581	CFV	250bp	Identified by WGS
581R	GAGAGCGTGGTCGTCCCTGAA	20	60.8	This study				
CFF3-F	ATGACGGATGGGCGATTGATTA	22	56.5	This study	CFF feature#3	CFF	520bp	Identified by WGS
CFF3-R	ATCTAGATTGGTTGGGTATGCGT	23	56.0	This study				
M13F	GTAAAACGACGGCCAG	16	54.0	This study	Colony PCR & Sequencincing	All Campy sp.	variabe	Universal sequencing primers. M13F/R also used for colony PCR to check presence of inserts
M13R	CAGGAAACAGCTATGAC	17	53.0	This study				

Table II. Sequences of adaptors and PCR primers used for subtractive hybridization.

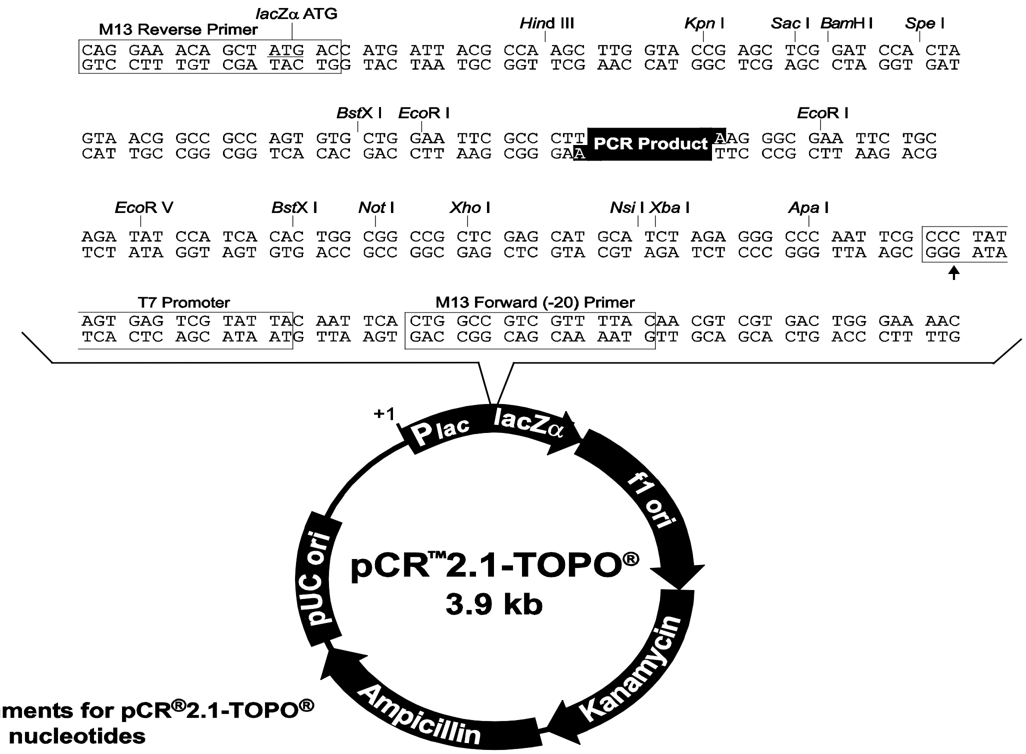
Adaptor / Primer name	Nucleotide Sequence (5’ → 3’)
Adaptor 1-A	CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT
Adaptor 1-B	ACCTGCCCGG
Adaptor 2R-A	CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGT
Adaptor 2R-B	ACCTCGGCCG
Primer P1[1]	CTAATACGACTCACTATAGGGC
Nested primer NP1[2]	TCGAGCGGCCGCCCGGGCAGGT
Nested primer NP2[3]	AGCGTGGTCGCGGCCGAGCT

[1] P1 PCR primer matches the long strands of adaptors 1 and 2 at their 5’ ends [2].

[2] NP1 nested primer matches the 3’ terminal sequence of adaptor 1 [2].

[3] NP2 nested primer matches the 3’ terminal sequence of adaptor 2R [2].

3. Map of PCR^{TM} 2.1 - TOPO Vector. Map adopted from TOPO TA cloning kit, Invitrogen (www.lifetechnologies.com/support).



Comments for pCR®2.1-TOPO®
3931 nucleotides

*LacZ*α fragment: bases 1-547
M13 reverse priming site: bases 205-221
Multiple cloning site: bases 234-357
T7 promoter/priming site: bases 364-383
M13 Forward (-20) priming site: bases 391-406
f1 origin: bases 548-985
Kanamycin resistance ORF: bases 1319-2113
Ampicillin resistance ORF: bases 2131-2991
pUC origin: bases 3136-3809

4. Blotting/ Sequencing analyses

Table I. Sequence and BLAST analysis of gene fragments identified during SSH. Most clones contained inserts with high or complete identity to CFF strain 82-40. Eight clones were selected based on more limited identity. Some clones show better homology with other *Campylobacter* species, but not with *C. fetus*.

Alignment with CFF82-40 (GenBank Accession # CP000487.1)					
Clone#	**Sequence length (bp)**	**Max score**	**Max identity**	**Query coverage**	**E value**
2	749	44.4	100%	41%	0.001
3	519	59	92%	29%	4E-04
4	995	1737	100%	97%	0.0
7	708	1189	99%	93%	0.0
8	319	480	99%	84%	2E-135
10	504	1613	98%	90%	0.0
12	216	228	100%	90%	2E-50
13	659	908	99%	92%	0.0
15	561	789	99%	91%	0.0
18	367	567	99%	86%	2E-158
19	817	1059	100%	92%	0.0
20	755	1337	100%	100%	0.0
21	936	1627	99%	97%	0.0
23	948	563	99%	33%	8E-160
24	817	1063	100%	92%	0.0
25	942	1631	100%	97%	0.0
26	552	906	100%	90%	0.0
28	764	39.2	100%	33%	0.062
30	1022	1779	100%	97%	0.0
34	894	44.6	100%	10%	0.002
35	825	86	100%	18%	5E-16
36	749	35.6	100%	35%	0.026
37	923	1604	99%	96%	0.0
38	874	690	96%	51%	0.0
39	954	1236	96%	94%	0.0
40	939	1604	99%	94%	0.0
41	938	1620	100%	96%	0.0
42	875	86	100%	17%	6E-16
43	221	228	100%	88%	2E-59
44	890	744	95%	93%	0.0
45	358	551	100%	86%	2E-156
47	973	41	93%	10%	0.023
48	358	551	100%	86%	2E-156
49	638	1065	100%	92%	0.0

Table II. Comparing the Sequencing/BLAST results of the 34 clones with Hybridization screening results.

Clone#	Hybridization results	Sequencing/BLAST results
2	CFV	No match with CFF
3	CFV	No match with CFF
4	Both CFV/CFF, but high in CFV	CFF
7	Both CFV/CFF, but high in CFV	CFF
8	Both CFV/CFF, but high in CFV	CFF
10	Both CFV/CFF, but high in CFV	CFF
12	Both CFV/CFF, but high in CFV	CFF
13	Both CFV/CFF, but high in CFV	CFF
15	Both CFV/CFF, but high in CFV	CFF
18	Both CFV/CFF, but high in CFV	CFF
19	Both CFV/CFF, but high in CFV	CFF
20	Both CFV/CFF, but high in CFV	CFF
21	Both CFV/CFF, but high in CFV	CFF
23	Both CFV/CFF, but high in CFV	CFF
24	Both CFV/CFF, but high in CFV	CFF
25	Both CFV/CFF, but high in CFV	CFF
26	Both CFV/CFF, but high in CFV	CFF
28	CFV	No match with CFF
30	Both CFV/CFF, but high in CFV	CFF
34	CFV	No match with CFF
35	CFV	No match with CFF
36	CFV	No match with CFF
37	Both CFV/CFF, but high in CFV	CFF
38	Both CFV/CFF, but high in CFV	CFF
39	Both CFV/CFF, but high in CFV	CFF
40	Both CFV/CFF, but high in CFV	CFF
41	Both CFV/CFF, but high in CFV	CFF
42	CFV	No match with CFF
43	Both CFV/CFF, but high in CFV	CFF
44	Both CFV/CFF, but high in CFV	CFF
45	Both CFV/CFF, but high in CFV	CFF
47	CFV	No match with CFF
48	Both CFV/CFF, but high in CFV	CFF
49	Both CFV/CFF, but high in CFV	CFF

5. Nucleotide Sequences of “CFV-specific” clones identified by SSH:

>Clone_2

GTACAAAAAGGGAATTTAACAAAGCTAGAAAAAGCTACTGCTTTAAATGCCAAAAAAATCGAACTAGAA
AAAGCTACTGCTGCAGAAATAAAAAGAATAAACGAAGATGCCATAAACGACGAATTAAACAAGCTCGAT
GAGAATATAGAGCTTTATAAAATTACAGCGCAATACGAGAAAATGGCTGCAGAGCAAAAGAAAAAAGC
AAGGTTAGAAGCGCAGGAGTTTATAAAAAAATATAAAAATCTAAGCCAAAAAGAGATATCTGAAATTTA
TAAATTTTATGAGCTAGAAGCAGCTAGAAAAGCAGATAGCGCAGTAGATAAATTTAAAAATAGCTTTGA
AGAGATAAAAGATAGCTGGAATGTTACAGTATCTGCTATGGGCAAAAACATAGAAGACAATTTATTTGA
TTTTATAACAGGTAAAACAACATCTTTGAGCAACGCTTTTAGGCAGATGGGCAAAGAGATGTTTGCAAGC
TTTATAAGTCCATATGCAAGGAGTTTTAGCAGTAGTGCAGCAGGATTATTTACAGGGCTTTTAGGTGGCG
GTAATGGTGGAGTTGCTAGTTTTGCAAATAGCTATGGCTTAGAGTTAAAAAACGGAGTTTATAGCGGAGA
TATAAAAGGTAGCCGAGTTGAAATACTTAGCGATGGAACAGTAAAGAGTGGCGGAAATGTTTTAGGAGA
GATACTAAGTGTAGGAAGCAGCGTTACTAGCATTGCAAACTTAGCAAATGGAAGTAC

>Clone_28

GTACAAAGTTTAAATGTAAAAACTCTTTGGCGTGTGTCTTGTAATAAACTCTTGCGGACTCCAAAAACTG
TTCTTTGCTTTGCTCAAACTCAAATTTTGCCATATTGTGGCTAAAGACTATTGTTAGTTTTTCATCTTCACA
TATCACATTTTTTATAAATTTATGCAAGCTCTCTTTTAGTCCTGTTTTTACTAATCCAAGACCTAGCTTATG
CCTATCAAGCTTAGGCCTTATATTGCCAAGTTCATAATCCATTTTTATACTCCAAAAACTCATTTATCAAG
GCTTTAAAAACAATTTAAAGCCTTTGTAAATTAGTCTAAATTTAGCTTCTTTTTTGTCTTAATATAATACTC
AACAAGCTCTTTGGCGCTTTTATATATCTCTTCTTCATCCTCCACTTCTATAGCTTCACATAACTATCTATC
ATCTCATCAAGATCGCTTAGAATCTCAGCTAGATCAGACAATCTTTTTGATTTTTTTTCTTTTTCTATCTCT
TTTTCCATCATATAACACATTTAAGCTCCTTTAAAATAAGACTTAATTAAGTAGTCCCTAACCACCCCAAA
TTGCCGCCGAAAACTCCTTCGAGTTTTCTTTTGCGGGAATTTTATATCAGGCTAACTAACGGTGCTTTTTCT
TTTATCTCTAGACACTCTTTTAGCCCATCATCTTCAAGAGCTAGCTCTTTTAAACGTTTTTCAGGCTTGTAA
ATTTTCTCTTCTTTGACAAGCAAACTAAATTTATCCCCCAAAAGTAC

>Clone_35

GTACGCCTTTAAACTTCTCTTTTAAGATCTCTGCCATGTTTGCTCTTGTTATCTCGCCATTAAAGACTTTGG
CCGTGATAGCTTTTAGATCATTACTTATTTTTTCGTTGTAATCATTTTTCATCCAGTAGAAGTTTTTTCGCA
TCGTCTCTAGCGATTTTATATCTATTTCATCAAACGCAAATTTAGCCCCAGCACTACTAGCTAAGCTCTCG
TATATCGTCTCGAAGTTTATATTTGGGACGCTTATCTCTAGTGCGTTTATCTTGTTTTCTACTAAGATTTTT
AAGTTTGCATTATCATAGTTATCATAAATCATTAAAAATAGCTCATCTAATGCTGCTTTTAATGACTTAGC
GTTTTGCTTGCTAAGTCCCATTAAAAGCTCATCTATAATCTCATCTTTGCTTTTTGTCTTTGCTATTTTGCTC
TTAAGCAGTCTTGTAAGTAGATCTTTATCATACATCTTTTATCTCCTCTTTATATTCCTTTAGCTCTTGCCA
ATCACGCTCTTTTATAAGAACGGTTTCGTTTGTTTCAAACGTTTTTTTACATTTAAAACAATAGCGGAAGC
GTCTATTTTCCAAGCCTTTTATGGTTGCTACCACTCTTGTTTTATCGTTGGCACAAAACGGGCATATCATTT
TTTACCATCTTTCGCAATTTTATCTAAAATATAGATAATTTTGCTTGTGTTTTTAACATTTAGCTTAAAAGG
GTCTGCGATTATCTTTGTTTTTTTATAATAAACATTTTGAGCTGGCTCTCATTCCAACCCAGATCACTCATA
GCCGCTTTTAATCTATCTACTTGCCTAAGCGTAC

6. Nucleotide Sequences of potentially useful genes identified by WGS analysis:

```
>Feature3_ORF_PI10_CFF8240_1602: 1567996_1569105
ATGAAAAAGGCTATTATAATAGGTGGAGGATTTGCAGGGTGTACTGCTGCATATTTTGAAAAATA
AAGGATTTGATGTAACGATATTTGAAGCTAGCAATGAGTTAGGAGCCGATAAGAACTCATTTTTA
CAAAGGGCATCCTTACACATTTGGACCTCATCACTTACTAAGACATAAATGACGGATGGGCGATT
GATTATTTTAAAGGTTTTTTAGATCTTAGGCACTTACTCATCAGGTATTAACCACACCTTATGGAG
AAAATAGTTTTTATAATTATCCACCGTTAAAAGCGATATCGACACATTTAAAAATAAAGAACGAA
TTTATAGAGAACTTAAGAAAGAAACACTAAAGCTAAACCAAAAAATTTAGAAGAATACTGGGAA
AATTCTATGGTAATACACTATACAATATGTTTGTAAATAGCTATTCAAAAAAAATGTGGAATATA
AAAACAATACGGAGCTAGATGAAGATTTATGTTTTTCGGTAAAAAAAGAACCTTTAAAAATAAA
CCTGGAGATTATTTTAGTGGAGCTATTTATAACGCATACCCAACCAATCTGATGGATATAATATAT
ATTTTGATAAATGCGTGTATGGCGTTAAAGTACAACTCAATCAAAAATATCTAAATTCGATTTAG
AAAATAAAACTATATGGATAGACGAGCAAAAGATAATGCAGATATAATAATTTCAACTATATCG
TTAGATGATTTATTTGAATATCACTAGGAGAATTGAAGTATATGGGTAGAGACTTTATGAAGATA
TTATTACCTGTTGAAAGATAACTCCTGAAGGATACTACTTTTTATATTATGCCGGAGACGAGCCAT
ATACTAGGAAGTTGAATATAAAACTCTAACCGGTTATAAATCGAAAGATACTCTAATAGGAATAG
AATACCATCAAACTCAAATAGACTATATCCATATCCAACAAAAGCAGAGGTAGAGAGGCGTATA
GATACAAAAATCTTATACCAAAAGATGTTTATAGCATAGGTAGACTTGGAAATATAAATATAAAG
ATATGTACTGGATACTAAAAGATATAAAAGAAATTATGAATCAATTTGA
```

```
>ORF_PI10_CFV354_1581.1505706-1506275
TTAGCACCCTTCTTTAGTAAATGCATAAGGTTGTTGGTCAAATTTTGTTATTGTTATAACCTTGGTT
TATAATAGCGGTATGTAGCTGGATCTCTAAACTTTATAAATTCTAAATCTACAGGCACAACAGGC
TTATTTGGCAAAGTTAGATTATATTTATCCTCTATAAATTCTTCATCTTCAAGTTTGTAGTCATGGG
TGCCAAAAAGGAAATAATGGTTGTCCCCCACTTTTTCCCACTCTAAATCCTCTTTAAACTCCGGAC
GGTCCTTTCTTCTTACTAGCTGCTTACCACTTAAGTAGGCTTTTAGTACTTTTATAATTTTGGCTGC
GTCTTCTTTTTCGCGTTTTCTGAAATAGTTTTCCCTGTGCTCATCTGTTATTTTATACTCTACCCCCG
AGTAAAAAATTCTACTATATCACTGCAGACTTTCCACTCTACTTTGTCTTGTTTGCCATTTCTCCAA
AAGCCCCAGGTAAGCCAATATACTAATTTATTATCTAAAAAAGCCTCCAAAGCCGAAATCCAGGC
TTCCATTCTTGCAGGGTCTTTTCTAAGTTCATCCAT
```

```
>ORF_PI5_CFV354_548:523218-523649
TTATTTACCTTTTTTTATTTTCTCAAGAATAAAAATTAGTTTGCTAGTATTCTTTACATCATTTTGA
ATGGATCTGCGATGATTTTTAGCTTTTTAACCACAAAACTTTTAAGCATTCTTCACTCCAGCCAAG
GCGAACCATAGCGTTTTTTAGCCTATTTGCTTGACTTGGGCGCTAAGCTCTTTTTTAGCTCTTGCG
ATTAGCTCTCTGCCTTTTAGGTCTATGTCATTTAGGTGGTACTTTTACCATACAACAAATCAAGAC
AATTTTCAAGCTCTTTGATACTTAAACTGCTGAGCTTGTAGCCCCAAAATAACAGTCCAAAAATG
TCTCCCAAGCATCATTTTCTTTTAGCTCTTTATAGCGTTTTGTAGTGTGGATTTTACTTAATAGTTG
TCTTCTTAAAACGTTTGTTTCTCGGTCAT
```

```
>ORF_PI5_CFV354_576:543461-544717
TTATTCCTTTATAATAGCATAGCAATATGGGCTAATTAGTTCATTTCCAGCAGTTACATCTTGTAA
TTATGTCCATCACAGCTCACATCATCACCCACACTTACACCACTTCCAATATGATAGCTATAATAC
CATCCATGTCGCCTAAGAGTTTGCTTTCTACATCTTTATACAATATAAAGAGAGAGCTCTAGCCCT
TTCTTGTATGGTAGAGCCAAGAGAACTACTGCATTGTATTTACCAAAATAACAGACTATAAGAGG
CGTACAAAATGGTTCATCAAGGTCGCCATAAAGATTGTCTTTAGTAGGCGGATAAGCCCAGCACG
GATGAGTTATACACTGCTGATATTCATTATTATACTTAGCATATGGCAAAGATATATGCTCTTTTG
ATGCAAGGCTGCTACTCTACTAGTCCCACTAGCAAAACACAAAAATGGATCAGGGGATATCTCTT
GCCATATGGACGAAGCAAGCCGATATAAAAAGGCTGCCAAGTTATATTATTGCCATTTGCATACT
AAAAGCCCCACAAATACGTCTATCTTGTAAGCTTATATAGTAGTTATTAAACCTTTATTTAGCATT
GCTATAGCCATATTTGGCATCGCTCCAATTTGAGCGTAGCTATCTATTATCGCTTTTTTGTTCTCAT
TAAAACTCTTTAATGGACTCATATGCAAGAAGTATTTGTCGCCATAATCTTCTAAAATAAACATAT
ACAAAAGCTCTTTGTTGCCAACTGCTTTCTAAGTATAAGCTCACGGTATATATAGCTCCCGTTTTC
TAAAGTGATATCTACTTATATTGACCATTTGCTTTTATATGTATACCTCTGTATTTTTTATTAGTAG
GTGATTCTTGGTCACTGGATTAGTCGCAGAGACTATATCTGTGATTTTTTCTGTTGAATTGTCTTCT
TTATGCCAATAATTTCGCCATTTATAGTTGATAGATAATTTTTAATTACAATTTTTTTAATAACACA
TCTTTGGTAAAAATGATATTTGATTGGTAATCATAATAAGAAGAGTTAATATCTTTAGCGGTATTT
AGGTAACGACTTGCTCCACTTTTCACATTATCAATCTGCGTTCAAACCCGCCAATATCAAAATAAT
CATTGACGTTTGTGACATCTAATGTTCTATCACCACACTCTCATATCCAGCCTCTAATGCTTTTGCT
TTTAAAACATCTAAAAACTCAGGACACTACTCACACTTCCTTCAAAATACAT
```

```
>ORF_PI5_CFV354_581:548170-549096
TTATGCAAAATATGTTAGCGTGAAGTTATCATCGCCTTTATCACTAACTGCCATAAAGTTCGCTCT
ATCACAAATATTCCGTCGCTATCTTCAAATTTAGGTTGTTCGCTAAGCTAGCTCTTGGAATTTCTA
GCTTAAATATTGTCTTTGCTTCACTGTCTTTAAACTCAGCAAAAAGCTCATAGCACTCCCAGCCTT
AAACTCATCAAATCCATCACTAGCTCCAGCTTCTAAACGCATTTTTAAAGTGAGTTTTGGCTCATA
ATCACTCATATAAAACTCGCCAATATTGTAAATTTAGAAAAACTTATTTTATTGCCCATTTCAAAG
CTAAACTCAGAGAGATTAACTGGAGTACCATTTAATGTCATTCCACCAAGTCTTCTTATAAGCAG
TGCTTCTCCACTCCTACATCTGTAATATTTTGATTTTCTATCACTCTATCTACATAAGCCCCGCTGA
TATAAACTCAGCATTTATTTTCTCGCCGACCTTACCACTGAGCTTAAAATCACTCTTTGCCCTTGG
ATTTTAAAAAGTCTATCTGGTAAAGCTAGATCCATACTAGCTGTAGCGCTGCATGAGTATCTGGG
CTAAATGTTTTGCTTACTAGATCCTCATCCTCACCAACTGGATTTCTAACGCTTTTTAGGTTGCATA
TCTTAAAAAGATTTTCGTAAAAATCTAAATTTTTGTATAAACTTGTTTTGATTTGCAGTTTTGCTTG
ACCCCAATCAGCTAATGTTACAACGTTTTGGCACCACCAAAAAAGCTAAAATTTGGTATTTCTAC
CTTATCAAACTCTATTTCAGGGACGACCACGCTCTCAAGCTTAAGCACTTGAGATGGGGCTTCAT
AAGTCCCAGCCCGCTTTCTACGCCAATTTTTACACCAGCTCTTGATAATACTCTTGCCAT
```