Department of Food Hygiene and Environmental Health
Faculty of Veterinary Medicine
University of Helsinki
Helsinki, Finland

# Molecular Epidemiology of *Campylobacter jejuni* in the Genomic Era

## Sara Kovanen

**ACADEMIC DISSERTATION**

To be presented, with the permission of the Faculty of Veterinary Medicine of the University of Helsinki, for public examination in Walter Sali, Agnes Sjöberginkatu 2, Helsinki, on 22nd of September 2017, at 12 noon.

Helsinki 2017

| | |
|---|---|
| Supervising Professors | Professor Emerita Marja-Liisa Hänninen, DVM, PhD |
| | Professor Maria Fredriksson-Ahomaa, DVM, PhD |
| | Department of Food Hygiene and Environmental Health |
| | Faculty of Veterinary Medicine |
| | University of Helsinki |
| | Helsinki, Finland |
| | |
| Supervised by | Professor Emerita Marja-Liisa Hänninen, DVM, PhD |
| | Department of Food Hygiene and Environmental Health |
| | Faculty of Veterinary Medicine |
| | University of Helsinki |
| | Helsinki, Finland |
| | |
| | Docent Rauni Kivistö, MSc, PhD |
| | Department of Food Hygiene and Environmental Health |
| | Faculty of Veterinary Medicine |
| | University of Helsinki |
| | Helsinki, Finland |
| | |
| Reviewed by | Professor Thomas Alter, DVM, PhD |
| | Institute of Food Safety and Food Hygiene |
| | Freie Universität Berlin |
| | Berlin, Germany |
| | |
| | Frank Pollari, DVM, MPH, D.V. Sc |
| | Branch of Infectious Disease Prevention and Control |
| | Public Health Agency of Canada |
| | Ottawa, Ontario, Canada |
| | |
| Opponent | Saara Salmenlinna, MSc, PhD |
| | Department of Health Security |
| | Expert Microbiology Unit |
| | National Institute for Health and Welfare |
| | Helsinki, Finland |

# TABLE OF CONTENTS

# ABSTRACT

Campylobacteriosis, caused especially by *Campylobacter jejuni*, is a leading cause of human bacterial gastroenteritis in the industrialized world, including Finland. With a wide range of mammal and avian hosts, this zoonotic pathogen transmits to the human gastrointestinal tract mainly indirectly, via consumption of contaminated food and drinking water or unpasteurized milk. The infection can originate also from environmental sources, such as surface waters or directly through contacts with animals. In Finland, approximately 4500 confirmed campylobacteriosis cases are registered annually, and most of the infections occur during the seasonal peak in the summer months of July and August. Approximately half of the infections are estimated to be associated with foreign travel, but, especially in the summer, the proportion of domestically acquired infections increases.

Molecular epidemiology of *C. jejuni* isolates has been widely investigated using different molecular typing methods and especially multilocus sequence typing (MLST), which assigns isolates to sequence types (STs) and clonal complexes (CCs), has provided valuable information about *C. jejuni* isolates in different sources worldwide. MLST has unified nomenclature and a curated public website, where a broad compilation of *C. jejuni* STs isolated from different sources and countries is available. MLST has revealed a diverse and weakly clonal population structure of *C. jejuni* isolates, however, certain predominant lineages are frequently found worldwide. In this thesis, we studied MLST types among Finnish human, poultry and water *C. jejuni* isolates and the most common types comprised ST-45 and ST-230 (ST-45 CC), ST-50 (ST-21 CC), ST-267 (ST-283 CC), ST-677 (ST-677 CC) and ST-3272 (unassigned [UA]). From these, ST-45 CC and ST-21 CC are common generalist lineages present worldwide, while ST-677 CC has been found more commonly in Finland than elsewhere. These three lineages have been also the most persistent in Finland, during the last decade when MLST types have been investigated, and they have been isolated from various sources including human patients, poultry, bovine, wild birds and environmental waters.

While MLST has advantages in congruence and ability to compare larger data worldwide, it has limitations in further distinguishing isolates within STs for epidemiological purposes. In recent years, whole-genome sequencing (WGS) has become more reasonable and also affordable to study the molecular epidemiology of bacterial isolates, offering the ability to explore and compare whole bacterial genomes. In this thesis, whole-genome (wg) MLST of *C. jejuni* isolates representing same STs revealed genetically distinct isolates and entire sub-lineages, suggesting that MLST itself is insufficient in differentiating closely related isolates. WgMLST also identified genetically highly related *C. jejuni* isolates among sporadic human infections,

originating from different geographical locations, suggesting the same infection source.

Poultry meat has been identified as a major source of human campylobacteriosis in many countries. However, European Food Safety Agency Panel on Biological Hazards has estimated that only 20-30% of the total *Campylobacter* infections associate directly with the consumption of poultry meat, thus, other potential sources should not be underestimated. In this thesis, we studied for the first time highly discriminatory wgMLST together with temporal relationship data to identify both genetically and epidemiologically related isolates. Our results revealed that chicken was a source in only 24% of domestically acquired human infections. Hence, the origin of more than 70% of the infections remained unidentified, suggesting other potential transmission routes, such as other domestic animals or environment-associated sources including drinking and surface water, wild animals and wild birds, as sources of human infections in Finland. In this thesis, we used the cut-off limit of ≤ 5 SNPs to identify genetically related isolates in wgMLST. This cut-off value was succesfully tested for the chicken isolates, representing the same ST, that originated from different batches reared on the same farms and was also used in our previous outbreak studies that identified outbreak-related strains.

Further, WGS data can also be applied to characterize genomic features among certain bacterial strains or clonal lineages. In Finland, *C. jejuni* isolates representing ST-677 CC have been frequently detected among human patients and associated also with more severe disease. In this thesis, WGS observation of genomic features among ST-677 CC isolates revealed highly similar genome maintenance among both epidemiologically related and unrelated ST-677 CC isolates. In addition, several putative virulence-associated characteristics that could possibly be linked to the association of this lineage with human bacteraemia infections were identified.

All in all, MLST provides valuable information especially for larger sets of isolates and is essential in the primary discrimination of the isolates for targeting further analysis. However, WGS and its applications are necessary in identifying genetically related isolates among outbreak investigations and source tracking, and for further exploring the genetic features among specific bacterial strains or lineages of interest.

# ACKNOWLEDGEMENTS

I'm deeply grateful to my supervisor, Professor Emerita Marja-Liisa Hänninen, for taking me into this research group and for the endless inspiration and encouragement to study this organism. I also thank my second supervisor, Docent Rauni Kivistö, for patient support and great advice. Professor Maria Fredriksson-Ahomaa is acknowledged for taking care of the supervising part at the final stages of my thesis. I thank Professor Hannu Korkeala and Docent Mari Nevas for providing high-quality research facilities and a good working environment. Johanna Laine and Mia Ketoharju are acknowledged for help in administrative area and Timo Haapanen for IT support.

Professor Thomas Alter and PhD Frank Pollari are highly acknowledged for reviewing my thesis and Carol-Ann Pelli, HonBSc, for editing the language. Professor Hilpi Rautelin and MDs Ulla-Maija Kärkkäinen, Tamara Tuuminen and Jaakko Uksila are warmly thanked for providing the clinical isolates and being as co-authors. Evira is acknowledged for providing the isolates from the chicken monitoring program and PhD Marjaana Hakkinen for being a co-author. Municipal health authorities are thanked for delivering water samples.

Assistant Professor Mirko Rossi is warmly thanked for generous support and help especially in bioinformatics and collaboration in the studies. I'm also grateful to Ann-Katrin Llarena, PhD, Astrid Skarp de Haan, PhD, Ji Zhang, PhD, Thomas Schott, PhD and Marjatta Rahkio, PhD, as being co-authors in these studies. I thank other members of our research group, Satu Olkkola, PhD, Alejandra Culebro, MSc and Joana Revez, PhD, as well as all personnel in our department, particularly Erika Pitkänen, Esa Penttinen, Heimo Tasanen, Jari Aho, Suvi Joutsen, Jenni Luukkanen, Karoliina Kettunen, Riitta Rahkila, Jenni Hultman and Hanna Korpunen, for good discussions and great laughs. In addition, Rauha Mustonen, Urszula Hirvi and Anneli Luoti are warmly thanked for technical help and media preparation.

My dear friends and my family are thanked for consistent support and inspiring discussions. Finally, I thank my loving spouse Lauri for endless support and encouragement and our son Veikko, the love of my life.

# LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, which are referred to in the text by Roman numerals I to IV:

I        Kovanen S.M., Kivistö R., Rossi M., Schott T., Kärkkäinen U, Tuuminen T., Uksila J., Rautelin H., and Hänninen M-L. Multilocus sequence typing (MLST) and whole-genome MLST of *Campylobacter jejuni* isolates from human infections in three districts during a seasonal peak in Finland. Journal of Clinical Microbiology. 2014 Dec 52(12):4147–54.

II      Kovanen S.M., Kivistö R., Llarena A. K, Zhang J., Kärkkäinen U., Tuuminen T., Uksila J., Hakkinen M., Rossi M., and Hänninen M-L. Tracing isolates from domestic human *Campylobacter jejuni* infections to chicken slaughter batches and swimming water using whole-genome multilocus sequence typing. International Journal of Food Microbiology. 2016 Jun 2; 226:53–60.

III    Kovanen S.M., Kivistö R., Rossi M., and Hänninen M-L. A combination of MLST and CRISPR typing reveals dominant *Campylobacter jejuni* types in organically farmed laying hens. Journal of Applied Microbiology. 2014 Jul; 117(1):249–57.

IV    Kivistö R.I., Kovanen S., Skarp-de Haan A., Schott T., Rahkio M., Rossi M., and Hänninen M-L. Evolution and comparative genomics of *Campylobacter jejuni* ST-677 clonal complex. Genome Biology and Evolution. 2014 Sep 4; 6(9):2424–38.

These publications were reprinted with the permission of their copyright holders. In addition, some unpublished material is presented.

# ABBREVIATIONS

| | |
|---|---|
| aa | Amino acid |
| BIOHAZ | Panel of Biological Hazards |
| BLAST | Basic Local Alignment Search Tool |
| bp | Base pair |
| CC | Clonal complex |
| CDS | Coding sequence |
| CJIE | *Campylobacter jejuni* integrated element |
| CRISPR | Clustered Regularly Interspaced Palindromic Repeat |
| (d)dNTP | (Di)deoxynucleotide |
| DNA | Deoxyribonucleic acid |
| ECDC | European Centre of Disease Control |
| EFSA | European Food Safety Authority |
| EVIRA | Finnish Food Safety Authority |
| FIMM | Institute for Molecular Medicine Finland |
| LOS | Lipo-oligosaccharide |
| mCCDA | Modified Charcoal Cefoperazone Deoxycholate Agar |
| MLST | Multilocus sequence typing |
| NCTC | National Collection of Type Cultures |
| NGS | Next-generation sequencing |
| ORF | Open reading frame |
| PCR | Polymerase chain reaction |
| PFGE | Pulsed-field gel electrophoresis |
| RAST | Rapid annotation using subsystems technology |
| SNP | Single-nucleotide polymorphism |
| SPSS | Statistical Package for Social Sciences |
| ST | Sequence type |
| THL | National Institute for Health and Welfare |
| UA | Unassigned (clonal complex) |
| wgMLST | Whole-genome multilocus sequence typing |
| WGS | Whole-genome sequencing |

# 1. INTRODUCTION

*Campylobacter* spp., and especially *Campylobacter jejuni* (causes > 90 % of human *Campylobacter* infections), have been the leading causes of human bacterial gastroenteritis worldwide (WHO, EFSA 2015). With more than 190 000 reported cases annually in the European Union (EU), *Campylobacter* spp. have been estimated to be responsible of 8.4% of the total burden of diarrhoeal diseases (Murray et al. 2010), with annual costs exceeding 2400 million € to the public health system (www.efsa.europa.eu/en/topics/topic/campylobacter). In Finland, more than 4000 human infections are registered every year (4589 in 2015), and campylobacteriosis has been continuously recognized as the most frequent bacterial gastroenteritis since 1998 (THL 2016).

*C. jejuni* is a zoonotic pathogen that is able to colonize a wide range of warm-blooded animals, including mammals and birds, and it may transmit to humans via several routes, but mainly via contaminated food items. Poultry and poultry products are considered major reservoirs and sources of human *C. jejuni* infections (EFSA 2005), and since 2007, the EU member states (MSs) have been monitoring the poultry production chain for campylobacters. However, the role of several other potential transmission sources, including direct contact with animals and faeces, drinking unpasteurized milk or faecally contaminated drinking water and ingestion of surface water while engaging in recreational activities should not be underestimated (Shönberg-Norio et al. 2004, Jore et al. 2010, Pitkänen et al. 2013, Ravel et al. 2016). Large outbreaks caused by *C. jejuni* are rare; also in Finland, most of the human infections are sporadic. Often the source, e.g. certain food products or contaminated drinking water, has been consumed several days before onset of symptoms, complicating the possibility of its identification (www.thl.fi).

For more than 30 years, several approaches have been developed to study the molecular epidemiology of *C. jejuni* isolates to connect the related isolates between human patients and different hosts and sources. Penner heat-stable serotyping was developed for *C. jejuni* in 1980 (Penner & Hennessy 1980), but its popularity diminished after molecular methods evolved. During the late 1990s and 2000s pulsed-field gel electrophoresis (PFGE) was widely used, and a standardized PFGE method for *C. jejuni* was described (On et al. 1998). In 2001, multilocus sequence typing (MLST) was described for *C. jejuni* (Dingle et al. 2001), which has since produced valuable knowledge about the molecular epidemiology of *C. jejuni* in different sources worldwide. The common nomenclature and public database of MLST types allow comparison between studies and countries. MLST has, however, limitations in further discriminating epidemiologically related isolates from unrelated ones. Also, genetically distinct strains may occur within the same sequence type (ST), and hence, methods with higher discriminative resolution are needed. In recent years, whole-genome sequencing (WGS), using next-generation sequencing (NGS) technologies, has become

more reasonable and affordable to use (Loman et al. 2015, Llarena et al. 2017). WGS data and implemented bioinformatics tools produce information about bacterial genomes, relationships and phylogeny between different isolates with a much higher resolution than previous molecular typing methods (Sheppard et al. 2012, Cody et al. 2013, Zhang et al. 2015). To better understand the role of different sources in posing a risk for human health, WGS data with suitable bioinformatics, combined with other epidemiological information, are essential for recognizing genetically closely related isolates and for tracing the sources of human infections (Sheppard et al. 2012, Carrillo et al. 2012, Cody et al. 2013, Revez et al. 2014a&b, Zhang et al. 2015).

The main aim of this thesis was to study molecular epidemiology of Finnish *C. jejuni* isolates in different sources and to elucidate their role in domestically acquired human infections using both conventional molecular typing methods and newly developed WGS tools.


# 2. REVIEW OF THE LITERATURE

## 2.1 *Campylobacter*

### 2.1.1 Historical aspects

*Campylobacter* infection was described for the first time by Theodor Escherich in 1886, when he found spiral-shaped bacteria in the colons of infants who had died from cholera-like symptoms. This infection manifesting as a "summer complaint" was initially named *Cholera infantum* (Kirst et al. 1985). The genus *Campylobacter* was created in 1963, when two *Vibrio* species (spp.) were transferred into *Campylobacter* spp. based on their low DNA base composition, microaerophilic growth and non-fermentative metabolism (Sebald & Veron 1963, Butzler et al. 2004). Further, in the 1970s, the filtration method enabled the isolation of *Campylobacter* spp. from the stools of diarrhoeic human patients (Butzler 1973), and a few years later, *Campylobacter* research reached a true milestone when cultivation procedures improved as selective antimicrobial supplements were added into growing media and the importance of these bacteria in human health was noted (Skirrow 1977). Finally, in the 1980s, almost one hundred years after first being noticed, *Campylobacter* research increased remarkably, resulting in the detection of *Campylobacter* spp. from different animal, human and environmental sources, description of several novel *Campylobacter* spp. (Lawson 1981, Benjamin et al. 1983, Marshall 1984, Gebhart et al. 1985) and growing knowledge about the zoonotic nature of *Campylobacter* and its role in human gastrointestinal infections (Skirrow et al. 1991).

## 2.1.2 Taxonomy and special features

*Campylobacter* spp. belong to the class Epsilonproteobacteria and to the order *Campylobacterales*. Together with *Arcobacter* and *Sulfurospirillum*, *Campylobacter* form the family *Campylobacteraceae* and further the genus *Campylobacter* (Vandamme et al. 1991, Vandamme et al. 2010). At present, 26 *Campylobacter* spp. and 11 subspecies (subsp.) have been described, 16 of which have been associated with human disease (www.bacterio.net/campylobacter.html.). However, the two most remarkable species that are responsible for over 95% of all human *Campylobacter* infections worldwide are *C. jejuni* and *C. coli* (Lastovica & Allos 2008).

The name "*Campylobacter*" originates from the Greek and means the curved rod. These slender, spiral-shaped rod cells are Gram-negative and oxidase-positive, with a length from 0.5 to 5 μm and a width from 0.2 to 0.8 μm (Snelling et al. 2005). *Campylobacter* spp. require a decreased amount of ambient oxygen and grow under microaerobic conditions (5-10% $O_2$ and 5-12% $CO_2$) at temperatures ranging from 37°C to 42°C, which are typical for mammal and avian intestines. Some species also require added hydrogen for optimal growth (Vandamme et al. 2010). These bacteria usually have flagella at one or both ends of the cells, enabling high motility and rapid movements (Vandamme 2000).

## 2.1.3 Genome characteristics

The genome of *Campylobacter jejuni* comprises a circular chromosome, with size of ca. 1.6 Mbp, G/C content of around 30% and approximately 1600 coding sequences (NCTC 11168; 1,641,481 bp, 30.6% G+C and 1,654 CDS (Parkhill et al. 2000)). The genome is relatively small compared to, for example, the 4.6 Mbp genome of *Escherichia coli* (Blattner et al. 1997), and it lacks the classical operon system and repetitive DNA sections. All *C. jejuni* strains share similar core genomes and strain-dependent variable accessory genes. The genome of *C. jejuni* contains hypervariable sequences that are short, homopolymeric nucleotide runs commonly found in accessory genes associated with biosynthesis or modification of surface structures (Parkhill et al. 2000). These highly variable sequences are usually present in phase-variable genes encoding cell surface structures like capsular polysaccharides (CPS), lipo-oligosaccharide (LOS) locus, flagella and also restriction-modification systems and metabolism (Aidley & Bayliss 2014). By changing the lengths of these homopolymeric tracts (polyG or polyC) in these genes, *C. jejuni* can switch the gene to ON or OFF phase depending on the environmental pressure (Aidley & Bayliss 2014). Other genomic features found in certain *C. jejuni* strains are integrated elements (CJIEs), from which four were first described in strain RM1221 (Fouts et al. 2005) and CJIE5 was recently described (Skarp et al. 2015). Some *C. jejuni* strains have been shown to carry plasmids, and pVir and pTet plasmids, which are considered to affect

the virulence of *C. jejuni*, were described in strain 81-176 (Bacon et al. 2000, Batchelor et al. 2004). Among other bacteria, *C. jejuni* may gain or lose genetic material by DNA rearrangements, such as insertions, deletions and recombination, the latter of which includes transformation, conjugation and transduction (Tenover et al. 1995, Barton et al. 2007, Fearnhead et al. 2014). In addition, genetic material can be acquired through horizontal gene transfer (Wang et al. 1989, Miller et al. 2005).

After ingestion, *C. jejuni* requires different mechanisms to survive the low pH in the stomach and bile in the duodenum before being transmitted to the human gastrointestinal tract. In gastrointestinal tract, *C. jejuni* colonize especially the colon and caecum (Flint et al. 2014). The pathogenicity of *C. jejuni* strains has been considered to rely on adhesion and invasion to the epithelial cells in the human intestines (Newell et al. 1985). Rapid movements and spiral shape allow these bacteria to penetrate through gastrointestinal mucus and to adhere to the epithelial cells (Baig & Manning 2014). Pathogenicity and ability to colonize the gut and cause infection depend on the strain, each of which have different mechanisms and genomic features. In general, *C. jejuni* strains that are detected in human patients seem to be more adapted to the human host and colonize human epithelial cells better than some other non-clinical isolates, which may not necessarily cause any disease in humans (Newell et al. 1985, Baig & Manning 2014). Motility of bacterial cells has been linked to invasion, and thus, the flagella and related genes have been proposed to play a role in virulence of *C. jejuni* (Wassenaar et al. 1991). In addition to flagella, genes encoding other cell surface structures, such as LOS (Louwen et al. 2008) and the capsule (Bachtiar et al. 2007, Keo et al. 2011), have also been associated with *C. jejuni* virulence. Furthermore, several metabolic pathways, like sulfite metabolism (Tareen et al. 2011), ferrous iron uptake (Raphael & Joens 2003), toxins such as cytolethal distending toxin (*cdt*) (Lee et al. 2003), as well as certain specific genetic features, such as pVir or pTet plasmids (Bacon et al. 2000, Batchelor et al. 2004) and integrated elements (Fouts et al. 2005), have been suggested to act as potential virulence factors and to be linked to pathogenicity of the strain. Although several studies have suggested and identified factors involved in pathogenicity of *C. jejuni*, the main determinants affecting virulence and pathogenicity remain poorly known.


## 2.2 Campylobacteriosis

Clinical significance of *Campylobacter* spp. is based on their ability to cause acute, gastrointestinal inflammatory enteritis, known as campylobacteriosis, in humans. *C. jejuni* has been reported as the most frequent causative agent, responsible for 90-95% of human campylobacteriosis cases, followed by *Campylobacter coli*. Several other *Campylobacter* species, such as *C. fetus*, *C. lari*, *C. upsaliensis* and *C. hyointestinalis,* have been infrequently linked to human campylobacteriosis (Lastovica & Allos 2008).

People of all ages and socio-economic backgrounds may acquire this disease, but it occurs most commonly in children under four years and young adults (ECDC). The incubation period varies from 2 to 5 days (range 1-10 days), with a mean of 3 days before onset of symptoms. The infectious dose is low, as only 500 bacterial cells are sufficient to cause the disease in a person (Robinson 1981). Also, the low infectious dose is supported by the epidemiological data from outbreaks combined with the fact that the organism does not grow outside the gut (Blaser et al. 1979). Acute, diarrhoeic symptoms usually last 3 to 5 days, but abdominal pain and stomach rumbling may last for weeks. After the clinical recovery, patient can still excrete *C. jejuni* in faeces for several weeks (Blaser & Engberg 2008).

Symptoms of campylobacteriosis vary from mild to more severe and include diarrhoea, which may contain blood (15% of cases), abdominal cramps, fever, nausea, muscle pain, headache and sometimes vomiting. For very young, elderly or immunocompromised patients or for patients with a chronic disease, such as diabetes, the symptoms are usually more severe (Blaser & Engberg 2008). However, usually campylobacteriosis is self-limiting and antimicrobial treatment is not needed. In case the symptoms are severe or patient is at higher risk, fluoroquinolones like ciprofloxacin or macrolides like erythromycin are used (Aarestrup & Engberg 2001, Blaser & Engberg 2008).

Sometimes, severe post-infectious symptoms may occur as late-onset complications, including Guillain Barré syndrome (GBS), Miller Fisher syndrome (MFS) and reactive arthritis (ReA). Campylobacteriosis is also associated with post-infectious irritable bowel syndrome (Blaser & Engberg 2008). Of these, GBS and MFS develop in 0.1-0.3% of patients (McCarthy 2001, Tam et al. 2006) and can lead to disorders affecting the autoimmune response of the peripheral nervous system, causing acute and rapidly developing weakness of the muscles, with a fatal outcome. Of other sequelae, ReA (joint inflammation) has been estimated to develop in 3% (Keithlin et al. 2014) to 7% of patients (Hannu et al. 2002). Irritable bowel syndrome, instead, was shown to develop in 4% (Keithlin et al. 2014) to 36% of patients (Spiller et al. 2009). In addition to these, certain strains of *C. jejuni* have been occasionally reported to cause bacteraemia among patients (Campbell et al. 2006, Feodoroff et al. 2011, Skarp et al. 2015).

## 2.3 Ecology of *Campylobacter jejuni*

To better understand how *C. jejuni* has adapted to its specific ecological niche, it is essential to know how this bacterium colonizes different host animals and, on the other hand, how these fragile organisms survive and remain infective in the environment.

### 2.3.1 Reservoirs of *C. jejuni*

*C. jejuni* has been found in the gastrointestinal tract of most warm-blooded animal species, including both mammals and birds. Poultry (i.e. broilers, hens and turkeys) and wild birds often get colonized by *C. jejuni*, which is supported by the temperature of the avian intestines (42°C), which is optimal for the growth of these bacteria (Williams et al. 2014). In broiler production, when chickens are colonized by *C. jejuni,* they can carry up to $10^9$ CFU/g of the bacteria in caeca and excrete them in their faeces until slaughtered. Because as few as 40 CFUs have been reported to be required to colonize chicken (Cawthraw et al. 1996), *C. jejuni* spreads efficiently to the entire or at least most of the flock through a faecal-oral route in a relatively short time (Vidal et al. 2014). Chicken flocks are proposed to get colonized mainly via horizontal transmission from external sources. *C. jejuni* can access broiler houses through contaminated environment originating from other livestock and animals such as rodents, pets and wild birds living in the farm surroundings (Hermans et al. 2012). Also personnel, equipment, flies and other insects, feed, litter and water are potential vehicles to transmit *C. jejuni* to rearing houses (Vidal et al. 2014). Therefore, strict biosecurity, including barriers, is important to prevent transmission.

Ruminants, including sheep and especially bovines, are also reservoirs of *C. jejuni* (Stanley et al. 1998a, Stanley et al. 1998b, Sproston et al. 2011). Bovines have frequently been shown to be colonized by *C. jejuni* (Humphrey & Beckett 1987, Hakkinen et al. 2007, Sproston et al. 2011), whereas pigs more often carry *Campylobacter coli* (Alter et al. 2005, Malakauskas et al. 2006, Juntunen et al. 2010). Also, pet animals, including cats and dogs, may act as reservoirs of *C. jejuni;* however, *C. helveticus* has been reported as the most common finding in cats (Rossi et al. 2008), while in dogs *C. jejuni* and *C. upsaliensis* prevail (Rossi et al. 2008, Parsons et al. 2010, Olkkola et al. 2015).

Livestock is colonized by *C. jejuni* by transfer from other domestic animals or via environmental pathways. Wild birds are considered a remarkable reservoir of *C. jejuni*. Especially migrating birds can distribute and spread *C. jejuni* to geographically wide areas. Several studies have reported wild birds, such as starlings, shore birds, corvids, gulls, pigeons, thrushes, black birds, geese and ducks, to often carry *C. jejuni* (Waldenström et al. 2002, Keller et al. 2011, Sippy et al. 2012, Griekspoor et al. 2013, Griekspoor et al. 2015, Llarena et al. 2015a). Wild birds, particularly those living in agricultural habitats, may potentially transmit *C. jejuni* also to domestic animals, including cattle and poultry (Waldenström & Griekspoor 2014, Cody et al. 2015). In addition to wild birds, other wildlife, such as rabbits, rats, mice and bats, which live in farm environments, may carry and spread *C. jejuni* to, for example, pasture environments (Adhikari et al. 2004, Kwan et al. 2008, Hatta et al. 2016).

When colonizing animal intestines, *C. jejuni* usually act as a commensal, resulting in asymptomatic carriage of these bacteria. Some previous reports have, however, shown

that chickens colonized by *C. jejuni* have had raised levels of *Campylobacter*-specific antibodies and a pro-inflammatory response (Cawthraw et al. 1994, Smith et al. 2008) and that young dogs were more often diarrhoeic when carrying *C. jejuni* (Amar et al. 2014).

### 2.3.2 *C. jejuni* in the environment

Colonized animals excrete huge numbers of *C. jejuni* in their faeces, consequently causing dispersion widely into the environment. After raining, the bacteria will spread from soil and pasture as run-offs into water systems. Therefore, *C. jejuni* has been detected from several environmental sources, including surface and well water, agricultural run-offs, farm surroundings and animal slurries (Colles et al. 2003, Hörman et al. 2004, Bronowski et al. 2014). The survival of this sensitive organism depends on prevailing temperature and humidity, amount of oxygen, pH and UV radiation (Park et al. 2002). Previous studies have shown the enhanced ability of *C. jejuni* to survive in the cool temperatures of well water and fresh food products (Kärenlampi et al. 2004, González et al. 2009, González et al. 2012). In addition to ambient temperature and UV radiation, natural predation of planktonic organisms can decrease the number of *C. jejuni* in environmental waters (Schallenberg et al. 2005). However, depending on the water system, *C. jejuni* may remain infective from weeks to months (Rollins et al. 1986, Korhonen et al. 1991), and certain strains seem to survive better in the environment than others (French et al. 2005, Bronowski et al. 2014).

Although *C. jejuni* cannot replicate outside of the intestines of its warm-blooded host and it requires microaerobic conditions to grow, it can survive prolonged periods in unfavourable environmental conditions. The organism has developed several genetic mechanisms to respond to such environmental pressures as oxidative stress. By modifying the lengths of the homopolymeric tracts of phase-variable contingency genes, *C. jejuni* is able to switch genes affecting, for example, its surface structure ON or OFF phase depending on the selective pressure (van Belkum et al. 1998). Increased survival in the environment is most probably associated with biofilm formation or interaction with free-living protozoa (Kalmokoff et al. 2006, Bronowski et al. 2014). Furthermore, in old cultures or stressful conditions, *C. jejuni* may go through morphological changes to form coccoid or filamentous cell shapes to save energy and enhance survival (Bronowski et al. 2014, Ghaffar et al. 2015).

## 2.4 Epidemiology of *C. jejuni* infections

Although *Campylobacter jejuni* was recognized as an emerging human pathogen already 40 years ago (Skirrow 1977), better knowledge of the transmission and significance of different sources of infections is still needed (Colles & Maiden 2012). In consequence of the sporadic nature of the human cases and several possible routes of transmission, the infection source often remains unclear. Also, due to a rather long incubation period and the required microbiological diagnostics, the time lap between exposure and recognized illness is too long to determine a reliable connection between a potential infection source and the patient. This section describes epidemiology and common transmission routes of *C. jejuni* in human infections.

### 2.4.1 *C. jejuni* infections in the EU

Reported campylobacteriosis cases in EU member states, including Finland, in 2011-2015 are shown in Table 1. Campylobacteriosis has remained the most frequently reported foodborne illness in the EU, with approximately 200 000 cases reported every year and the trend has been increasing since 2008, when registration of laboratory-confirmed cases began (EFSA 2016). In 2015, a total of 229 213 human campylobacteriosis cases were reported by the EU member states, with an incidence of 65.5 per 100 000 inhabitants (EFSA 2016). As this number only includes confirmed cases, the true annual number of cases is much higher, estimated at nine million in the EU alone (EFSA 2011).

Table 1. Laboratory-confirmed, registered campylobacteriosis cases in 2011-2015 in EU member states (EFSA 2016).

| Country | 2015 | | 2014 | | 2013 | | 2012 | | 2011 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Cases | Rate | Cases | Rate | Cases | Rate | Cases | Rate | Cases | Rate |
| Austria | 6258 | 73.0 | 6514 | 76.6 | 5731 | 67.8 | 4710 | 56.0 | 5129 | 61.0 |
| Belgium[a] | 6096 | - | 8098 | - | 8148 | - | 6607 | - | 7716 | - |
| Bulgaria | 227 | 3.2 | 144 | 2.0 | 124 | 1.7 | 97 | 1.3 | 73 | 1.0 |
| Croatia | 1393 | 33.0 | 1647 | 38.8 | 0 | 0.0 | 0 | 0.0 | - | - |
| Cyprus | 29 | 3.4 | 40 | 4.7 | 56 | 6.5 | 68 | 7.9 | 62 | 7.4 |
| Czech Republic | 20 960 | 198.9 | 20 750 | 197.4 | 18 267 | 173.7 | 18 287 | 174.1 | 18 743 | 178.7 |
| Denmark | 4327 | 76.5 | 3773 | 67.0 | 3772 | 67.3 | 3720 | 66.7 | 4060 | 73.0 |
| Estonia | 318 | 24.2 | 285 | 21.7 | 382 | 28.9 | 268 | 20.2 | 214 | 16.1 |
| **Finland** | **4588** | **83.8** | **4889** | **89.7** | **4066** | **74.9** | **4251** | **78.7** | **4267** | **79.4** |
| France[b] | 6074 | 45.7 | 5958 | 45.2 | 5198 | 39.6 | 5079 | 38.9 | 5538 | 42.6 |
| Germany | 69 829 | 86.0 | 70 571 | 87.4 | 63 280 | 78.6 | 62 548 | 77.9 | 70 811 | 88.3 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Greece[c] | - | - | - | - | - | - | - | - | - | - |
| Hungary | 8342 | 84.6 | 8444 | 85.5 | 7247 | 73.5 | 6367 | 64.4 | 6121 | 62.4 |
| Ireland | 2453 | 53.0 | 2593 | 56.3 | 2288 | 49.8 | 2391 | 52.2 | 2433 | 53.2 |
| Italy[a] | 1014 | - | 1252 | - | 1178 | - | 774 | - | 468 | - |
| Latvia | 74 | 3.7 | 37 | 1.8 | 9 | 0.4 | 8 | 0.4 | 7 | 0.3 |
| Lithuania | 1186 | 40.6 | 1184 | 40.2 | 1139 | 38.3 | 917 | 30.5 | 1124 | 36.8 |
| Luxembourg | 254 | 45.1 | 873 | 158.8 | 675 | 125.7 | 581 | 110.7 | 704 | 137.5 |
| Malta | 248 | 57.8 | 288 | 67.7 | 246 | 58.4 | 220 | 52.7 | 220 | 53.0 |
| Netherlands[d] | 3778 | 43.0 | 4159 | 47.5 | 3702 | 42.4 | 4248 | 48.8 | 4408 | 50.9 |
| Poland | 653 | 1.7 | 650 | 1.7 | 552 | 1.4 | 431 | 1.1 | 354 | 0.9 |
| Portugal | 271 | 2.6 | - | - | - | - | - | - | - | - |
| Romania | 311 | 1.6 | 256 | 1.3 | 218 | 1.1 | 92 | 0.5 | 149 | 0.7 |
| Slovakia | 6949 | 128.2 | 6744 | 124.5 | 5845 | 108.0 | 5704 | 105.5 | 4565 | 84.7 |
| Slovenia | 1328 | 64.4 | 1184 | 57.4 | 1027 | 49.9 | 983 | 47.8 | 998 | 48.7 |
| Spain[e] | 13 227 | 63.3 | 11 481 | 54.9 | 7064 | 50.4 | 5548 | 47.4 | 5469 | 46.9 |
| Sweden | 9180 | 94.2 | 8288 | 85.9 | 8114 | 84.9 | 7901 | 83.3 | 8214 | 87.2 |
| United Kingdom | 59 846 | 92.2 | 66 716 | 103.7 | 66 382 | 103.9 | 72 500 | 114.2 | 72 139 | 114.5 |
| **EU Total** | **229 213** | **65.5** | **236 818** | **69.5** | **214 710** | **64.3** | **214 300** | **65.4** | **223 986** | **69.2** |
| Iceland | 119 | 36.2 | 142 | 43.6 | 101 | 31.4 | 60 | 18.8 | 123 | 38.6 |
| Norway | 2318 | 44.9 | 3386 | 66.3 | 3291 | 65.2 | 2933 | 58.8 | 3005 | 61.1 |
| Switzerland[f] | 7055 | 85.3 | 7565 | 92.9 | 7481 | 93.1 | 8432 | 106.0 | 7963 | 101.2 |

(a): Sentinel surveillance; no information on estimated coverage. Thus, notification rate cannot be estimated.

(b): Sentinel surveillance; notification rates calculated on estimated coverage of 20%.

(c): No surveillance system.

(d): Sentinel surveillance; notification rates calculated on estimated coverage 52%.

(e): Sentinel system; notification rates calculated with an estimated population coverage of 45% in 2014-2015, 30% in 2013 and 25% in 2009-2012.

(f): Switzerland provided data directly to EFSA. The human data for Switzerland also include the ones from Liechtenstein.

Seasonal variation is typical in incidence of *Campylobacter* infections. In areas with a temperate climate, like Europe, a major peak in infections can be seen during the summer months (Nylen et al. 2002). Besides the summer peak in the EU data set, a slight peak in the number of infections has occurred in January since 2011 (EFSA 2016). Similar peaks do not appear in tropical countries, however, infections may be more frequent during the rainy period than during the dry season (Taylor et al. 1992).

### 2.4.2 *C. jejuni* infections in Finland

Since 1994, Finnish clinical laboratories have reported all culture-positive *Campylobacter* findings to the National Infectious Diseases Register (NIDR), and from 1998, campylobacteriosis has been the most frequently reported bacterial cause of gastroenteritis among human patients (Vierikko et al. 2004). Approximately 4000

*Campylobacter* infections have been registered annually since 2005, with an increasing trend (www.thl.tilasto), and about 95% of these infections are caused by *C. jejuni* (THL 2016). In 2015, a total of 4589 *Campylobacter* infections were registered to NIDR (4887 in 2014), and the incidence for the whole population was 84 per 100 000 inhabitants (THL 2016). From the registered infections, 55% were in men, and most frequently the infected patients were aged 25–29 years (incidence 148/100 000). The distribution in different hospital districts varied, showing the highest incidence (121/100 000) in the Helsinki metropolitan hospital district (THL 2016). Approximately half of the total annual infections are estimated to be associated with travelling abroad, but the real number is unclear since information regarding travelling history is frequently missing.

According to NIDR, all registered *Campylobacter* cases (and *Salmonella* cases for comparison) from 2006–2015 are shown in Figure 1. Seasonal variation is evident in the Nordic countries, including Finland (Nylen et al. 2002), and the incidence of *Campylobacter* infections is usually highest in July (de Haan et al. 2014). Infections appearing in summertime have been reported to often have domestic origin (Rautelin et al. 2000, Vierikko et al. 2004) and are associated with recreational activities during summertime, which are discussed in more detail in Section 2.4.4.
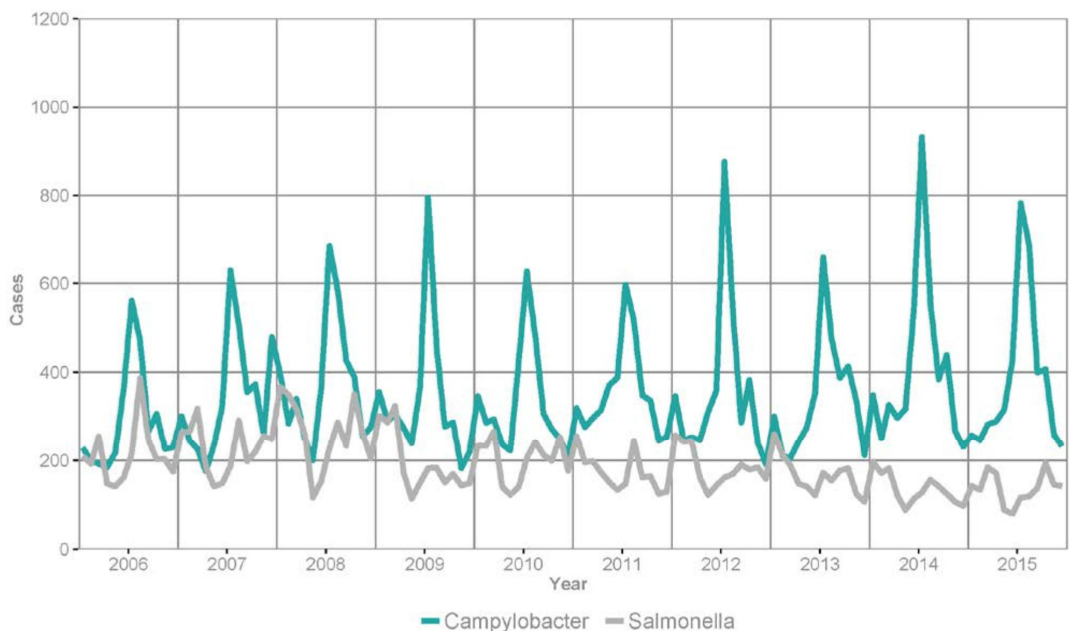


Figure 1. Registered *Campylobacter* and *Salmonella* infections monthly from 2006 to 2015 in Finland (THL 2016).

### 2.4.3 Outbreaks caused by *C. jejuni*

Most of the *C. jejuni* infections are sporadic and large outbreaks seldom occur. In 2015, from the total of 4362 food- and waterborne outbreaks reported to EFSA, *Campylobacter* was the second most common bacterial agent after *Salmonella,* causing 8.9% of the outbreaks (EFSA 2016). Outbreaks caused by *Campylobacter* have typically been associated with drinking unpasteurized milk or water contaminated by agricultural run-offs originating from ruminants. In the first reported outbreak in the USA, 3000 people became ill after drinking water contaminated by wild bird faeces (Vogt et al. 1982). In Canada in 2000, *C. jejuni*, together with bovine-originated *Escherichia coli* O157, was associated with a large waterborne outbreak originating from municipal groundwater source contaminated by bovine faeces, causing seven deaths and infections for more than 2300 patients (Mackay et al. 2002). Also, unpasteurized milk has been reported to cause several *Campylobacter* outbreaks in the USA and the UK (Pebody et al. 1997, Headrick et al. 1998). Poultry-associated *C. jejuni* rarely cause larger outbreaks. However, small outbreaks are relatively common, and recently, as the genomic typing methods have evolved, there has been increasing evidence of diffuse *C. jejuni* outbreaks often caused by poultry retail meat (Strachan & Forbes 2014).

Amongst other industrialized countries, including Finland, larger *Campylobacter* outbreaks are rare. However, waterborne outbreaks associated with drinking water are more common in countries, such as Finland, where non-chlorinated groundwater is used in water supplies (Miettinen et al. 2001, Hänninen et al. 2003). In 1998–2015, *C. jejuni* caused 19 foodborne and 15 waterborne outbreaks among the recorded data where the infective agent was identified (EVIRA). Foodborne outbreaks were most commonly associated with consumption of poultry meat (broiler salad and turkey soup) and unpasteurized milk, whereas waterborne outbreaks were linked to drinking of contaminated source or well water (EVIRA, Hänninen et al. 2003). The largest waterborne outbreak occurred in Nokia in 2007, where 400 m$^3$ of purified waste water contaminated the municipal drinking water system, resulting in thousands of cases of illnesses. Several pathogens, including *C. jejuni*, *Giardia,* norovirus and *Salmonella,* were detected from both the tap water and the patients (Laine et al. 2011).

### 2.4.4 Sources and transmission routes of human infections

Humans acquire *C. jejuni* via the faecal-oral route either directly by contact with faecal material (animal or human) or more commonly indirectly by consumption of contaminated food items or water.

The major single source of human campylobacteriosis is considered to be poultry and poultry meat, and the occurrence of *C. jejuni* is shown to be higher among both humans and chickens at the same time of the year (Jore et al. 2010). In poultry production, the

carcasses of *C. jejuni*-positive chickens are contaminated due to slaughtering processes, and thus, the meat reaching retail and consumed by customers often contains variable levels of *C. jejuni* (EFSA 2014). The European Food Safety Authority (EFSA) Panel of Biohazards (BIOHAZ) has estimated that handling, preparation and consumption of chicken meat may directly account for 20-30% of human infections (EFSA 2011). Also, epidemiological studies performed in several countries have shown that consumption of chicken, i.e. handling and eating improperly cooked chicken meat, poses an increased risk for the consumer to acquire the infection (Schönberg-Norio et al. 2004, Mughini-Gras et al. 2012, Levesque et al. 2013, Mossong et al. 2016, Pintar et al. 2016). Furthermore, 50-80% of the human infections can be associated with chickens indirectly, suggesting alternative routes of transmission (EFSA 2011). Since estimated as the most important single source of human campylobacteriosis, the member states (MSs) of the EU have been required to monitor their broiler production chain for *Campylobacter* and report to the EU Commission since 2007 (EFSA 2005). The prevalence of *C. jejuni* in chicken slaughter batches or in retail meat has been high in several EU MSs, however, in Finland, the number of *C. jejuni*-positive batches has remained low (EFSA 2016). In 2015, according to the EFSA monitoring programme, fourteen MSs reported that 47% of the 6707 tested items, including retail meat and carcasses in slaughterhouses, were positive for *Campylobacter* (EFSA 2016). Among broiler batch samples, the prevalence of positive slaughter batches varied from 3.8% in Finland to 69.9% in the UK (EFSA 2016). Since the prevalence of positive slaughter batches reflects the presence of *Campylobacter* in meat (Skarp et al. 2016), the meat sold in retail has also been investigated. The results have revealed the presence of *C. jejuni* in 49% of chicken meat sold in France (Guyard-Nicodème et al. 2015), in 20% in Poland (Korsak et al. 2015), in 24% in Turkey (Ozbey et al. 2014) and in 11% in Finland (EFSA 2014).

In addition to poultry, ruminants and especially bovines may carry *C. jejuni* and due to slaughter processes, carcasses and finally the retail meat may be contaminated. In previous studies, 20% of faecal samples of Finnish beef and dairy cattle at slaughter (Hakkinen et al. 2007) and 42% of dairy cattle in Japan were positive for *C. jejuni* (Sasaki et al. 2013). From retail meat, however, low numbers of positive findings have been reported, with 3.5% (Wong et al. 2007) and 3.9% (Trokhymchuk et al., 2014) of retail beef found to be positive for *C. jejuni*. In previous epidemiological studies, contacts with animals, barbequing, tasting of undercooked red meat and consumption of tripe were recognized as risk factors for campylobacteriosis (Schönberg-Norio et al. 2004, Mughini-Gras et al. 2012). Besides cattle, *C. jejuni* colonizes sheep, resulting in a risk for humans to acquire campylobacteriosis from lamb meat; *C. jejuni* was detected in 6.5% of lamb and mutton sold in retail in New Zealand (Wong et al. 2007).

Human infection can also be acquired via environmental pathways and in addition to domestic and wild animals, wild birds spread *C. jejuni* in their feces. From the environment, *C. jejuni* may transmit to humans directly via consumption of food or

water contaminated by bird faeces or when the birds are consumed as game. Transmission may also be indirect, e.g. when *C. jejuni* first transmits from the environment to domestic animals and then further to humans (Waldenström & Griekspoor 2014). In previous studies, eating strawberries and vegetables straight from the fields was shown to be associated with human infections (Shönberg-Norio et al. 2004, Kwan et al. 2014). Also, surface waters have been shown to contain *C. jejuni,* among several other pathogens in many countries including Finland (Hörman et al. 2004, de Haan et al. 2013, Pitkänen et al. 2013, Bronowski et al. 2014). Lakes, streams and the sea are often used for swimming and other recreational activities, especially during the summertime. This may pose a risk to humans to acquire campylobacteriosis through ingestion of water during swimming in natural waters, as shown in previous case-control studies (Shönberg-Norio et al. 2004, MacDonald et al. 2015, Ravel et al. 2016). Other factors that raise the risk of acquiring campylobacteriosis include visits to farms and contacts with animals (Ravel et al. 2016), barbequing and consumption of water from private wells (Shönberg-Norio et al. 2004, Jore et al. 2010, MacDonald et al. 2015) and occupational work at slaughterhouses (Ellström et al. 2014).

In addition to zoonotic transmission, campylobacteriosis may be acquired from infected persons in, for instance, the same family, although this is considered to occur rarely (Rotario et al. 2010). Also, besides the *C. jejuni* originating from animal faeces, environmental waters may be contaminate with human-derived *C. jejuni,* which can access surface waters through municipal wastewater plants (Hokajärvi et al. 2013).

## 2.5 Molecular epidemiology and genotyping methods

To understand the relative importance of different hosts and transmission routes of *C. jejuni* to humans, molecular epidemiology and the use of suitable genotyping methods are essential. The most commonly used genotyping methods of *C. jejuni* usually rely on comparison of the patterns of DNA fragments or DNA sequence contents between different isolates.

### 2.5.1 Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis (PFGE) has been considered as "a gold standard" to support isolate characterization in epidemiological studies of pathogenic organisms including *C. jejuni* and the method was validated for *C. jejuni* already in the 1990's (Yan et al. 1991, On et al. 1998). PFGE is a highly discriminatory method using restriction enzymes that specifically digest genomic DNA into fragments, which are then run on gel electrophoresis and separated according to their size. Variable PFGE patterns allow separation of different strains of *C. jejuni* (Tenover et al. 1995). PFGE has been broadly applied in molecular epidemiology studies of *C. jejuni* to compare genotypes of isolates

from different sources and to identify similar isolates in outbreak investigations (Hänninen et al. 2003, Kärenlampi et al. 2003, Revez et al. 2014a&b, Llarena et al. 2015b). The discriminatory power of PFGE depends on the restriction enzyme used. *Sma*I has been commonly used for PFGE typing of *C. jejuni*, however, *Kpn*I, has been shown to be more discriminatory (Michaud et al. 2001, Kärenlampi et al. 2007b). By comparing the PFGE profiles of the isolates detected from different suspected sources and from human patients, the studies have identified the source in *C. jejuni* outbreaks (Hänninen et al. 2003, Pitkänen et al. 2008, Revez et al. 2014a&b,). Although widely used and suitable for certain analyses, the difficulties in comparing results between different laboratories and countries and the lack of consistent typing nomenclature are major limitations of PFGE typing.

## 2.5.2 Multilocus sequence typing

Most of the existing knowledge about *C. jejuni* genotypes occurring in various sources and among human cases in different countries is based on multilocus sequence typing (MLST), more widely used for *C. jejuni* from the beginning of the 2000s (Dingle et al. 2001). MLST has been frequently used in molecular epidemiology and population genetic studies of *C. jejuni* (Colles & Maiden 2012, Taboada et al. 2013, de Haan et al. 2014). In MLST, alleles of seven housekeeping genes, including aspartase A (*asp*A), glutamine synthetase (*gln*A), citrate synthase (*glt*A), serine hydroxymethyltransferase (*gly*A), phosphoglucomutase (*pgm*), transketolase (*tkt*) and adenosinetriphosphate synthase subunit α (*unc*A), are sequenced. Obtained allele sequences are numbered and further assigned to sequence types (STs) and clonal complexes (CCs) (Dingle et al., 2001). Sequence types are assigned to common lineages, CCs, if they share identical alleles in at least four loci with the founder ST, measured using the program BURST (Dingle et al. 2001). If ST has less than four identical alleles with the founder ST, it is unassigned (UA) to a CC. The main advantage of MLST, relative to, for example, PFGE, is a common nomenclature of MLST types and a curated web-based database that allows comparison of the results between different countries and previous studies (pubMLST.org/campylobacter).

To date (22.03.2017), a total of 8855 MLST profiles, including both *C. jejuni* and *C. coli*, are assigned in the pubMLST database (pubMLST.org/campylobacter). Despite the high diversity of STs, several studies concerning MLST types in different sources have revealed that some MLST types are frequently found and some are rare. The two most common clonal complexes, ST-21 CC and ST-45 CC, have been detected from numerous sources in different parts of the world (Dingle et al. 2001, Sails et al. 2003, Kärenlampi et al. 2007a, Lévesque et al. 2008, Mullner et al. 2010, Gripp et al. 2011). These widely spread generalist lineages have been adapted to a broad range of animal hosts and are frequently detected also in infected patients. In addition to ST-21 CC and ST-45 CC, other lineages present in several countries and sources include, among

others, ST-48 CC, ST-52 CC, ST-206 CC, ST-257 CC and ST-353 CC (Dingle et al. 2001, Sails et al. 2003, Sheppard et al. 2009a, De Haan et al. 2013, Levesque et al. 2008, Kittl et al. 2011). Besides the lineages colonizing several hosts, certain STs may also be strongly adapted to a particular host, e.g. ST-61 (ST-61 CC) to cattle and ST-403 (ST-403 CC) to sheep (Sheppard et al. 2009a, de Haan et al. 2010b). In addition, some of the lineages show differences in their geographical distribution such as ST-677 and ST-794 (ST-677 CC), which have been common in Nordic countries (Kärenlampi et al. 2007a, Griekspoor et al. 2013, de Haan et al. 2014), and ST-474 (ST-48 CC), which has been a typical ST in New Zealand (McTavish et al. 2008, Mullner et al. 2009).

In Finland, several studies concerning the MLST types in human patients, chicken, bovine, wild birds, zoo animals and environmental waters have been performed (Kärenlampi et al. 2007a, de Haan et al., 2010a&b, de Haan et al. 2013, Llarena et al. 2015a&b). Studies have revealed a diverse MLST population in Finland including prevalent STs and, on the other hand, STs that are detected only rarely and usually from a single source. Most of the Finnish *C. jejuni* isolates belong to clonal complexes that have persisted for years. These predominant STs include ST-11, ST-45, ST-137 and ST-230 (ST-45 CC), ST-50 and ST-451 (ST-21 CC), ST-677 and ST-794 (ST-677 CC) and ST-267 (ST-283 CC), and they have been detected most frequently in human patients, bovines, chickens and domesticated animals. Most of them have also been detected in environmental waters, wild birds and zoo animals (de Haan et al. 2013, de Haan et al. 2014, Llarena et al. 2015a). From among these lineages, especially interesting is ST-677 CC, which has been commonly detected among human patients that require longer hospitalization (Kärenlampi et al. 2007a) and linked frequently to human sepsis cases (Feodoroff et al. 2013).

### 2.5.3 Population structure analyses

To understand why certain lineages are either adapted to special niches or multi-host generalists, and why some STs are more predominant in human infections, it is important to examine the population structure of *C. jejuni*. Sequence-based data, such as MLST data, can be used as an input in several computer-based programs assessing population structure (Sheppard et al. 2015).

#### *2.5.3.1 Genetic distance and phylogenetic analyses*

Sequence data, and especially MLST data, have been widely used in population structure analyses, which can be applied to estimate distance-based or phylogenetic relationships between STs and CCs, or to determine the niche adaptation of specific MLST lineages. Population structure of *C. jejuni* can be measured by analysing the

genetic distances of the isolates using NeighborNet algorithm, which utilizes MLST data to determine genetic distances of the alleles. This algorithm creates a distance matrix from the allelic profiles and constructs a split network from the isolates that can be visualized using Splitstree (Huson & Bryant 2006). Genetic distance networks are especially suitable for large MLST data and can be applied also to whole-genome sequence (WGS) data, the latter of which is described in more detail in Section 2.6.3.

Phylogenetic analyses combined with source data can be applied to identify niche adaptation of specific MLST lineages and to estimate the phylogeny and clonal relationships between the sequences of, for example, MLST loci. ClonalFrame is a widely used evolutionary model that takes into consideration both the mutation rate and the recombination of the sequences and creates a tree-like genealogy of the isolates (Didelot & Falush 2007). Previous studies using ClonalFrame have revealed relationships of different STs and CCs that either are adapted to a specific niche, such as sheep, bovines or wild birds, or are lineages that colonize a wide range of host animals (Sheppard et al. 2009a, Gripp et al. 2011, Griekspoor et al. 2013). Since the MLST data are limited to a small fraction of the genome, in recent years, population structure models, such as ClonalFrame, have been refined to utilize WGS data (Didelot et al. 2010), providing information about phylogenetic relationships of bacterial isolates with an extremely high resolution (Croucher et al. 2013).

### 2.5.3.2 Source attribution modelling

MLST sequence data can also be used in source attribution, which utilizes mathematical modelling to connect *C. jejuni* isolates from different sources to human isolates. Such programs as STRUCTURE (Falush et al. 2003), Bayesian Association of Population Structure (BAPS) (Corander et al. 2003) and asymmetric island model (Wilson et al. 2008) use sequence-based data, e.g. MLST data, to depict the population structure of *C. jejuni* in different sources. These programs differ from each other in modelling calculations, as was shown in a Scottish study in which 58% of the human infections were attributed to chicken meat using STRUCTURE and 78% using asymmetric island model (Sheppard et al. 2009b). In previous studies, all of these source attribution models have depicted chicken as the predominant source and cattle as the second most important source in human *C. jejuni* infections (Wilson et al. 2008, Sheppard et al. 2009b, de Haan 2010b, Mughini-Gras et al. 2012, Lévesque et al. 2013).

### 2.5.4 CRISPR sequences

Clustered regularly interspaced short palindromic repeats (CRISPRs) are short sequence repeats that consist of direct or almost direct repeats (DRs) interspersed with a varying number of non-repetitive spacer sequences (Price et al. 2007). CRISPRs,

together with CRISPR-associated proteins (Cas), form the CRISPR-Cas system, which is widely spread among different bacterial species. The CRISPR-Cas system is linked to the adaptive immune system protecting bacteria against outside threats, such as bacteriophages and plasmids (Rath et al. 2015), and spacer sequences are reported to have external phage or plasmid origin (Bolotin et al. 2005). Thus, CRISPR analyses may provide information about previous phage contacts of bacteria (Deveau et al. 2010). CRISPR sequences were originally tested as a molecular typing method of *C. jejuni*, however, the results revealed highly variable repeat and spacer combinations (Schouls et al. 2003). As several spacer sequences have been associated with specific MLST types, analyses of CRISPRs may be used to complement the molecular epidemiology investigations of *C. jejuni* (Pearson et al. 2015).

## 2.6 Next-generation sequencing

The basis of all DNA sequencing originates from Sanger sequencing, developed by Frederick Sanger in 1977 (Sanger et al. 1977). As the first method to discover the base pair (bp) order of DNA fragments, Sanger sequencing used DNA polymerase enzyme during DNA replication for selective incorporation of chain-terminating dideoxynucleotides (ddNTPs). The classical Sanger method was laborious and time-consuming, as four different DNA reactions, with their own radioactively labelled ddNTPs, were needed and the sizes of sequenced DNA fragments were separated in four lanes in gel electrophoresis (Sanger et al. 1977). Subsequently, techniques developed and marked improvements, such as fluorescent labelling of ddNTPs and automated sequencing machines, were implemented in Sanger sequencing (Smith et al. 1986). However, Sanger sequencing, as a suitable method for sequencing DNA fragments with size ranging from 100 to 1000 bp, was insufficient in sequencing longer DNA strands.

In the 1990s, conventional Sanger sequencing together with a new, rapid shotgun sequencing technology enabled sequencing for longer DNA strands, which led to the establishment of the first whole-genome DNA sequence of the bacterium *Haemophilus influenzae* (Fleischmann et al. 1995) and few years later, also to the whole-genome sequencing (WGS) of *C. jejuni* NCTC 11168 (Parkhill et al. 2000). In shotgun sequencing, DNA is cut into random fragments, clone library is prepared and the fragments are sequenced using chain termination and the multiple reads are aligned. Finally, overlapping reads are reconstructed to one continuous DNA strand using computer-based programs (Loman et al. 2015, Fleischmann et al. 1995). Since 2005, massively parallel DNA sequencing platform techniques have emerged and have started to replace the costly and laborious shotgun sequencing feasible for observation of only one or few strains of interest (Parkhill et al. 2000, Biggs et al. 2011, Hepworth et al. 2011, Cooper et al. 2011). These new NGS technologies are faster and much more affordable, enabling numerous isolates to be analysed. Also, since nowadays several

biotech companies offer NGS services at reasonable price, special skills for individual researchers and clinical workers working with whole-genome sequence data are no longer required (Illumina 2010, Cody et al. 2013, Loman et al. 2015).

## 2.6.1 Next-generation sequencing (NGS) technologies

At present, several NGS (also called high-throughput sequencing) platforms, with strengths and limitations, are available including Illumina (Illumina Inc., San Diego, CA, USA), Ion Torrent (Thermo Fisher Scientific, WA, USA) and PacBio (Pacific Biosciences, Menlo Park, CA, USA). All of these technologies use extracted genomic DNA as a template, with a specified concentration, and have variation in throughput of the samples and produced read lengths and error rates (Illumina 2010, Loman et al. 2012, Loman et al. 2015). From these platforms, Illumina and Ion Torrent rely on fragmentation of the DNA, tagging the template with adaptors, library preparation, template amplification and sequencing of the amplified products (Illumina 2010, Loman et al. 2012). For example, in Illumina technology, fragmented DNA is first tagged with adaptors and loaded to a chip consisting of separated lanes, each with its own library preparations. Templates of each lane are amplified using DNA polymerase and synthetic oligomers in a phase called cluster generation and amplified DNA templates are sequenced by synthesis. Illumina sequencing chemistry uses special reversible terminating nucleotides with fluorescent tags. The computer measures the fluorescent signals in different wavelengths and record the dNTP data, finally creating the sequence of a DNA template (Illumina 2010). As Illumina technology uses bridge amplification and terminating nucleotides in sequencing, Ion Torrent rely on emulsion PCR (polymerase chain reaction) and utilizes detection and release of hydrogen ions in sequencing chemistry (Loman et al. 2012). In addition to platforms that use amplification processes, single-molecule sequencing technologies that do not require template amplification, such as PacBio, are available. PacBio introduced a 'real-time' sequencing platform, where dyed nucleotides are attached to an extending DNA strand using special polymerase and zero-mode wavelength detector, which determines the labelled nucleotides, thus reconstructing the sequence (Loman et al. 2012). Platforms have differences in quality of provided sequence data and also in costs. PacBio, for instance, produces extensively longer read lengths, but is much more expensive and has a higher rate of errors than Illumina, which produces short reads and allows massive throughput of samples (Illumina 2010). Sequencing errors may impact the assembly, and errors that occur randomly can be solved by increasing the sequencing coverage. However, in case the errors appear systematically among homopolymeric runs, the combined use of two different technologies may be justified (Loman et al. 2012).

## 2.6.2 High-throughput sequence data: assembly and annotation

The raw nucleotide sequence reads provided by NGS sequencing consist of huge amounts of data, which need to be trimmed and extracted before being suitable for analyses. Raw reads can be trimmed to provide a better quality (Smeds & Kunster 2011) and if passing quality threshold, assembled into contigs that are longer, continuous nucleotide sequences. Reads can be assembled to contigs either by mapping them against a known reference genome or more commonly *de novo*, where no reference is required and assembly is based on mathematical algorithms that use k-mers (short overlapping reads of k length) to construct the contigs (Illumina 2010, Paszkiewicz et al. 2010). Numerous assembling programs using reference genome or *de novo* or both have been developed, including Velvet (Zerbino et al. 2008), MIRA (Mimicking Intelligent Read Assembly) (Chevreux et al. 2004), Abyss (Assembly by Short Sequences) (Simpson et al. 2009) and SPAdes (Bankevich et al. 2012). Different programs are optimized to assemble reads created by certain technologies. The widely used De Bruijn graph-based Velvet, for example, allows both mix-end and pair-end assembly, while MIRA relies on sequence editors, which allow repair of sequencing errors and use of quality data in generating assemblies, and Abyss is a *de novo* parallel paired-end assembler (Dark et al. 2013). SPAdes, by contrast, has improved its De Bruijn graph algorithm to assemble both the single-cell and multi-cell sequences (Bankevich et al. 2012). The final quality of assembly depends on the coverage (error-free read cover of the genome), the N50 value (shortest contig length at 50% of the genome) and the number of internal gaps in contigs (Illumina 2010). Assembled contigs that contain information of the whole genomic DNA sequence of the isolate can be utilized in several approaches and tools using whole-genome sequence (WGS) data. Assembled data can be used, for example, as an input for whole-genome (wg) MLST, which creates allelic profiles among the genes found in analysed isolates.

For deeper inspection of the gene content and gene functions, assembled contigs can be annotated either manually, supported by visualization tools such as Artemis (Rutherford et al. 2000) or automatically, using annotation programs, such as RAST (Rapid Annotation using Subsystem technology) web server (Aziz et al. 2008). Automated annotation is more commonly used nowadays and RAST is a program that uses fully automated annotation for *de novo* sequences, which is based on the assumption of several subsystems (Aziz et al. 2008). For forecasting function of a predicted gene, RAST performs BLAST searches against reference dataset of functional annotated genes and public annotated genomes (Aziz et al. 2008).

## 2.6.3 Approaches using whole-genome sequence data

WGS data provides the most accurate discriminatory power for molecular typing in epidemiological studies since the bacterial isolates can be distinguished from each

other only by single-nucleotide polymorphisms (SNPs) (Taboada et al. 2013). Comparative genomic tools, for example, whole-genome MLST (wgMLST), produce information from bacterial genomes with a much higher resolution than traditional molecular typing methods such as MLST (Sheppard et al. 2012, Cody et al. 2013, Revez et al. 2014a). While MLST relies on sequence loci of seven housekeeping genes, wgMLST is a gene-by-gene approach covering the shared genes among the studied isolates and can be exploited for example in studies exploring the genomic relationships between the isolates originating from the same or different sources (Sheppard et al. 2012, Cody et al. 2013). In wgMLST, shared genes found from all isolates analysed are defined to create the allelic profiles. The number of shared genes depends on the size of the population (analysed isolates) and genetic similarity of the isolates. The analysis can be performed locally using programs, such as Genome profiler (GeP) (Zhang et al. 2015), or public Bacterial Isolate Genome Sequence database (BIGSdb) in pubMLST.org website (Jolley & Maiden 2010). In the latter, the assembled contigs are uploaded into the website, which then determines the allelic profile of the genes that have been defined in the BIGS database (pubMLST.org/campylobacter). WgMLST can be performed also using a highly accurate GeP algorithm (Zhang et al. 2015) that requires genome annotation for one isolate under analysis, and, by using this genome as a reference, GeP performs BLAST searches against all of the draft genomes being investigated, finally creating allelic profiles of the analysed isolates. Allelic profiles, i.e. the wgMLST results, can be visualized in phylogenetic networks using a distance-based NeighborNet algorithm (Huson & Bryant 2006), which reveals the genetically closely related isolates that can be further examined as a new subset of isolates.

WGS data can be used to describe genomic features of certain isolates and to compare full genomes of individual isolates and, for example, to identify bacterial clones among outbreaks (Cody et al. 2013, Revez et al. 2014a&b, Llarena et al. 2017). Although most human campylobacteriosis cases are sporadic, comparative genomics of the closely related isolates may reveal diffuse outbreaks among a larger population of sporadic isolates (Strachan & Forbes 2014). Previous studies have utilized WGS data to determine the genome differences between isolates originating from the same outbreak. In two Finnish studies, the WGS comparisons revealed three SNPs between the *C. jejuni* isolates associated with a milk-borne outbreak (Revez et al. 2014a) and three SNPs also between human and water isolates originating from a water-borne outbreak (Revez et al. 2014b). In another study, 0 to 15 SNVs (single-nucleotide variants) were detected using high-quality core SNV analysis, among four *C. jejuni* isolates originating from a waterborne outbreak in Canada (Clark et al. 2016). WGS data can also be applied to depict the population structure among isolates representing, for example, a certain sequence type of clonal lineage. Genome comparisons of isolates of a common generalist, ST-45, revealed several smaller populations, i.e. sub-lineages, among the ST-45 populations available in public databases (Llarena et al. 2016). The study further showed that the isolates among

certain sub-lineages were neither spatially nor temporally dependent but were detected from multiple sources in several countries for over a decade inferring the stability of certain clones (Llarena et al. 2016). More data are needed to elucidate how common this kind of clonality is among other *C. jejuni* STs.

Recently, programs that utilize WGS data as an input have been developed to be faster and more user-friendly, with the end result that processing of WGS data no longer necessarily requires special bioinformatics expertise (Loman et al. 2015). Thus, as a highly accurate method used previously by mainly researchers, WGS data is becoming increasingly applicable also to clinical practices in routine diagnostics (Cody et al. 2013, Loman et al. 2015).

# 3. AIMS OF THE STUDY

The major aim of this thesis was to study the molecular epidemiology of Finnish *Campylobacter jejuni* isolates using genomic tools varying from more conventional genotyping methods to highly accurate genome comparisons at the whole-genome sequence level. Specific aims of the four original publications of the thesis were as follows:

1. To investigate and compare *C. jejuni* isolates occurring in domestic human infections in three districts during the summer peak using MLST and whole-genome (wg) MLST (I, II)

2. To trace the human *C. jejuni* infections to chicken or swimming water source by combining recently developed, highly accurate wgMLST with the temporal relationships of the isolates (II)

3. To determine which MLST types are present in Finnish organic laying hens compared with other sources and to evaluate the application of CRISPR sequencing in characterization of the isolates (III)

4. To further investigate ST-677 CC isolates by comparing the whole-genome sequences among associated chicken farm isolates and also against unrelated clinical isolates (IV)

5. To characterize the genomic features typical for ST-677 CC (IV)

# 4. MATERIALS AND METHODS

## 4.1 Bacterial isolates

An overview of all *C. jejuni* isolates included in Studies I to IV is presented in Table 2.

Table 2. Source of isolates used in Studies I-IV.

| Study | Origin of the isolates | Year | Country | Reference |
|---|---|---|---|---|
| I | Human gastroenteritis (n=95) | 2012 | Finland | Study I |
| II | Human gastroenteritis (n=95) | 2012 | Finland | Study II |
| | Chicken caeca (n=83) | 2012 | Finland | Llarena et al. 2015b |
| | Swimming water (n=20) | 2012 | Finland | Study II |
| III | Organic laying hens (n=147) | 2003, 2004 | Finland | Sulonen et al. 2007 |
| IV | Chicken fecal swab (n=2) | 2003, 2004 | Finland | Study IV |
| | Boot sock sample (n=2) | 2003, 2004 | Finland | Study IV |
| | Water puddle (n=1) | 2004 | Finland | Study IV |
| | Human meningitis (n=1) | 1979 | Sweden | Norrby et al. 1980 |
| | Human gastroenteritis (n=5) | 2010–2012 | UK | BIGS db (pubMLST.org) |

### 4.1.1 Human isolates (I, II and IV)

In Studies I and II, *Campylobacter* spp. isolates from human patients (n=109) were obtained during the seasonal peak, from June to September 2012. Isolates were collected and stored in case the patient had not travelled in the two weeks before onset of the symptoms to include only infections acquired from domestic sources. The primary isolation and collection of isolates were performed by the clinical laboratories of four hospital districts, located in Central and Eastern Finland. Two sparsely populated and closely located hospital districts were merged into one, resulting in three larger districts, with populations of 279 000 (district 1), 121 000 (district 2, two hospital districts) and 265 000 (district 3). In further analysis, 95 *C. jejuni* isolates were included in Studies I and II, as 13 of the isolates did not grow after primary isolation and delivery and one isolate was identified as *C. coli* (see Section 4.1.5).

Isolates were coded and numbered indicating the district from which they were collected as K1, K2, etc. (district 1), M1, M2, etc. (district 2) and J1, J2, etc. (district 3).

In Study IV, the genome sequences of additional human *C. jejuni* ST-677 CC isolates were obtained from GenBank (accession number NZ_AIPM00000000) and from Bacterial Isolate Genomic Sequence database (BIGSdb), available in *Campylobacter* PubMLST website (pubMLST.org/campylobacter). One ST-677 isolate (LMG9872) originated from the cerebrospinal fluid of a meningitis patient, detected in Sweden in 1979 (Norrby et al. 1980). Four clinical isolates representing ST-677 (OXC6332, OXC7095, OXC7345 and OXC7358) and one ST-794 isolate (OXC5341) originated from Oxford, UK, and were collected from gastroenteritis patients from 2010 to 2012.

### 4.1.2 Poultry isolates (II-IV)

#### 4.1.2.1 Chicken slaughter batch isolates (II)

The chicken-derived *C. jejuni* isolates (n=83) included in Study II were collected in 2012 during the Finnish chicken *Campylobacter* monitoring programme (MMM 10/EEO2007) and sent to Evira (Finnish Food Safety Agency). The isolates have been described in more detail in a previous study (Llarena et al. 2015b). In brief, according to the national poultry monitoring programme, all chicken batches slaughtered between June and October were tested, and from November to May, sampling was randomized using an expected target prevalence at the expected precision of 5%. Isolation of *C. jejuni* was performed following the Evira Protocol 3512/5 (EVIRA 3512/5), in which 10 intact caeca samples from each slaughter batch are pooled and cultivated on mCCDA agar plates. A total of 83 chicken slaughter batches (1534 batches tested during June to October) were positive for *C. jejuni,* and from these, 75 were detected during the same time period as human isolates, from June to September. These isolates originated from 37 farms (from the total of 124 farms delivering batches to slaughter), indicating that a single farm could have produced more than one positive batch during the study period. In addition, eight isolates were detected from slaughter batches beyond the human sampling period, i.e. the seasonal peak. All 83 isolates were included in whole-genome sequence analyses. For MLST frequency measurement, only 51 isolates were included because 32 isolates representing the same STs and detected from different slaughter batches reared simultaneously on the same farm (in different broiler houses) and slaughtered on successive days were excluded.

### 4.1.2.2 Organically farmed laying hen isolates (III)

In Study III, 147 *C. jejuni* isolates from organically farmed laying hens, collected from a total of 18 farms in 2003 and 2004 and genotyped using PFGE, were analysed using MLST and CRISPR sequences. Sampling of individual hens and PFGE typing (using *Kpn*I restriction enzyme) have been described in detail elsewhere (Sulonen et al. 2007). In our study, we chose 50 isolates, by selecting one or two isolates from each of the 42 distinct PFGE types, for further MLST analyses. In addition, 86 isolates representing distinct PFGE types occurring on a single farm in each sampling were selected for CRISPR sequence analysis.

### 4.1.2.3 Chicken farm isolates (IV)

Chicken farm isolates in Study IV originated from two farms (farms A and C) and were part of a study investigating the occurrence of *C. jejuni* in farm environments (unpublished data, 2004). Three ST-677 isolates from farm A were collected during a one-day farm visit in 2003 and included a chicken faecal swab (5070), a boot sock sample from inside the chicken house (5071) and a sample from a rain water puddle outside the chicken house (5072). Two ST-794 isolates were collected during one day in 2004 and included a chicken cloacal swab (3515) and a boot sock sample from the surroundings of the chicken house (3516).

### 4.1.3 Water isolates (II)

Twenty *C. jejuni* isolates from swimming water were included in Study II. Seven of them originated from EU beaches (EU directive: 2006/7/EY, Ministry of Social Affairs and Health: 711/2014), in lakes located in the same districts as the human isolates. The samples were collected and delivered by the municipal authorities of the three districts from June to August 2012. In addition, 13 *C. jejuni* isolates collected from small beaches on a lake and river located in the Helsinki Metropolitan area were included in Study II as controls.

Water samples were analysed within 24 h from the collection by filtrating 100 ml or 100 ml and 1000 ml volumes through 0.45 µm pore size membrane filters (Millipore, Billerica, MA, USA), which were subsequently transferred to Bolton enrichment broth (Bolton broth, supplement SR0183, Oxoid Ltd., Basingstoke, Hampshire, UK) with 5% horse blood, and incubated microaerobically at 37°C for 44 ± 4 h. After incubation, 10 µl and 100 µl of enrichment broth were cultivated on mCCDA agar plates and incubated as described below (Section 4.1.4). Colonies showing typical growth for *Campylobacter* were preliminarily confirmed by Gram staining and aerobic incubation at 37°C for 44 ± 4 h, and, if typical for *Campylobacter*, were stored as described below.

### 4.1.4 DNA extraction (I-IV) and species confirmation (I and II)

Prior to further analyses, all *Campylobacter* isolates were cultivated on selective modified charcoal cefoperatsone deoxycholate agar (mCCDA, Oxoid Ltd., Basingstoke, Hampshire, UK) plates and incubated microaerobically (5% $O_2$, 10% $CO_2$, 85% $N_2$) at 37°C for 24-48 h. After incubation, isolates were sub-cultivated on non-selective Nutrient agar plates (Oxoid Ltd., Basingstoke, Hampshire, UK), with 5% horse blood, incubated microaerobically at 37°C or at 42°C for 24 h. After incubation, bacterial mass was transferred into a skimmed milk or nutrient broth glycerol stock (85% Nutrient broth, 15% glycerol) and stored at -70°C.

For DNA extraction, isolates were obtained from the freezer, melted on ice and cultivated on Nutrient blood agar plates. After overnight microaerobic incubation at 37°C or at 42°C, genomic DNA was isolated using a commercial Wizard Genomic DNA purification Kit (Promega, Mannheim, Germany). DNA concentration and purity (260/280 and 260/230 values) were measured using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). For isolates delivered to whole-genome sequencing, concentration and quality parameters were measured using a Qubit fluorometer (Life Technologies, Invitrogen, CA, USA) according to the recommendations of the sequencing laboratory of the FIMM (Institute for Molecular Medicine).

Human and water *C. jejuni* isolates in Studies I and II were distinguished from other species (i.e. *C. coli*) using multiplex PCR (Denis et al. 1999) and verified by running the PCR products on a 1.5% TAE agarose gel.

## 4.2 Molecular typing and sequencing methods

### 4.2.1 MLST (I and III)

Conventional multilocus sequence typing was performed for 50 organic laying hen isolates in Study III and for 21 human isolates in Study I. PCR conditions and primers used for the seven MLST loci were as described before (Miller et al. 2005, Korczak et al. 2009), and PCR products were verified by running on a 1.5% agarose gel. Sanger sequencing was carried out by the BigDye terminator version 3.1 (Applied Biosystems, Foster City, CA, USA) chemistry, and amplified PCR products were run on an ABI Prism3130XL genetic analyser or ABI 3730 DNA analyser (Applied Biosystems, Carlsbad, CA, USA). MLST sequence assemblies were edited with BioNumerics version 5.1 software (Applied Maths, Kortijk, Belgium).

### 4.2.2 CRISPR typing (III)

Primers and PCR conditions used for analysis of CRISPR sequences were as described earlier (Price et al. 2007). PCR products were sequenced using Sanger sequencing as described in the previous section.

### 4.2.3 Whole-genome sequencing (I, II and IV)

Whole-genome sequencing was performed in Study I (74 human isolates), Study II (remaining 21 human isolates, 83 chicken and 20 water isolates) and Study IV (5 chicken isolates) using Illumina HiSeq sequencing technology with 100 cycles and paired-end Nextera library preparation. Sequencing was carried out in Studies I and II by the Institute for Molecular Medicine (FIMM), with high genome coverage (>40 fold). In Study IV, sequencing was carried out in Base-Clear BV (Leiden, Netherlands). Raw data reads were filtered using ConDeTri Perl script (Smeds & Kunster 2006) with default parameters and minimum read length of 75 nucleotides. Draft genomes passing the quality threshold were assembled into contigs using Abyss 1.3.5 (Simpson et al. 2009) with default parameters in Studies I and IV, and using SPAdes 3.2.1 (Bankewich et al. 2012) with default parameters in Study II. In addition to NGS by Illumina, the sequence gaps of the draft genome of isolate 5070 in Study IV, which were located in the genome regions of interest, were closed using Sanger sequencing as described above (Section 4.2.1).

## 4.3 Data analyses

### 4.3.1 MLST (I, II and III) and CRISPR typing (III)

Different sequence types (STs) and clonal complexes (CCs) of all isolates used in Studies I to III were assigned by querying either the sequences of the seven MLST loci individually or the draft genome as a whole genome in the pubMLST web site for *Campylobacter* (Jolley & Maiden 2012).

CRISPR sequences, consisting of direct repeats and spacers, were obtained using the CRISPRFinder web server (Grissa et al. 2007). Each spacer sequence and the combination of spacers, i.e. CRISPR type, were numbered.

## 4.3.2 Whole-genome MLST (I, II and IV)

In this thesis, wgMLST was performed for isolates representing each ST and further for closely related isolates forming distinct clusters in wgMLST analysis of each ST. In Studies I and IV, wgMLST was performed by querying the draft genomes (i.e assembled contigs as fasta file) against the pubMLST website, which then automatically annotated all of the defined loci and numbered the alleles (including MLST type) present in the BIGS database. Allele data were processed in Excel (Microsoft Excel 2010) separately for the isolates representing the same ST, and further for the closely related isolates that formed clusters inside STs. Numbers of allelic differences among the shared loci were recorded as a distance-based matrix, and NeighborNet algorithm was used to construct a phylogenetic network based on the distance matrix.

In Study II, four isolates were excluded from the wgMLST analyses (three human and one water) since the assembled draft genomes were too fragmented (>100 contigs), and wgMLST was performed on the remaining isolates using a recently developed gene-by-gene approach Genome profiler (GeP) (Zhang 2015). GeP uses BLAST searches of all loci/ORFs in the chosen reference genome, instead of defined alleles in the BIGS database, to determine the genes present in the genome sequence of the query isolates. GeP was performed for isolates representing each ST and again for isolates that formed clusters within STs, and thus, were considered as sub-lineages of closely related isolates. Split Decomposition network (Study II) was used to construct a phylogenetic network based on the allelic distance matrix of the shared loci of the isolates created by GeP.

Phylogenetic networks in Studies I, II and IV were visualized using SplitsTree 4 (Huson & Bryant 2006) and further edited in CorelDRAW X6 (Corel Corporation, Ottawa, Ontario, Canada).

In Study II, wgMLST was done for the isolates representing the same ST and again for closely related isolates that formed clusters within STs. Isolates were considered to be genetically highly related, thus representing the same bacterial clone, if they differed from each other by 0 to 5 SNPs. The selected cut-off limit (≤ 5 SNPs) is based on data from our previous studies where 0 to 3 SNPs were identified in the genomes of isolates associated with a milk- and waterborne outbreak (Revez et al. 2014a&b). Changes in the lengths of homopolymeric tracts (i.e. poly G or poly C runs) were not included in counting SNP differences. However, differences in the lengths of homopolymeric tracts were counted among the total allelic differences.

### 4.3.3 ClonalFrame genealogies (III and IV)

Both the MLST and WGS data were used to describe the phylogenetic relationships among *C. jejuni* isolates using ClonalFrame genealogy (Didelot & Falush 2007). In Study III, the MLST sequence data were used as an input, and ClonalFrame was run for 14 STs detected from organic laying hens and in addition for 197 STs collected from seven different Finnish sources (de Haan et al. 2010a&b, de Haan et al. 2013) with 100 000 iterations and 50 000 burn-in iterations.

In Study IV, ClonalFrame was run for the WGS data of ST-677 CC isolates. Draft genomes were first aligned using progressive Mauve (Darling et al. 2010), and collinear blocks (> 500 bp) were filtered using the perl script available in the ClonalFrame package (Didelot et al. 2010). ClonalFrame was run for the sequences with 10 000 burn-in iterations followed by 10 000 data collection iterations.

ClonalFrame genealogies were constructed from the strict consensus trees of three independent runs, which were then displayed in MEGA 5.1 and labelled using CorelDRAW X5 (Corel Corporation, Ottawa, Ontario, Canada).

### 4.3.4 Comparative genomics (I, II and IV)

In Studies I, II and IV, genetic variation between closely related isolates, recognized using wgMLST, was carefully analysed and alignments of loci that had multiple alleles were manually inspected. Variation included single-nucleotide polymorphisms (SNPs) and changes in the lengths of the homopolymeric tracts (mainly poly G and poly C tracts). Homopolymeric tract changes were not considered true SNPs (Bayliss et al. 2012), however, they were accounted for in the total numbers of allelic differences.

In Study IV, gene calling and automated annotation were performed using RAST web server (Aziz et al. 2008). Sequence comparisons and genome browsing were done using the SEED Viewer (Overbeek et al. 2014), Artemis (Rutherford et al. 2000) and BLAST Ring Image Generator (BRIG) (Alikhan et al. 2011). Alignment of genomic regions was performed using progressive Mauve (Darling et al. 2004, Darling et al. 2010), and the pair-wise SNP analysis between the chicken isolates was performed using the breseq pipeline with default settings (Barrick et al. 2009). Novel open reading frames (ORFs) were analysed using InterProScan (Quevillon et al. 2005) and by BLASTP searches against the non-redundant protein sequences in GenBank (Benson et al. 2014).

### 4.3.5 Temporal association analysis (II)

In Study II, the temporal relationships of human, chicken and water *C. jejuni* isolates were defined. The temporal connection between human and chicken isolates was considered if the slaughter of the chicken batch preceded human illness by 2 to 23 days, taking into account the time the meat products are available in retail and the incubation time of the illness as well as the delay in contacting the doctor after start of symptoms. The model has been described in more detail in a previous study (Kärenlampi et al. 2003, Hakkinen et al. 2009). The time period varying from 1 to 10 days was used to determine the temporal relationship between swimming water and human isolates, taking into consideration the incubation time of the infection and time delay in contacting the doctor.

### 4.3.6 Statistical analyses (I and III)

In Study I, SPSS software (IBM SPSS Statistics 21) was used to compare the human MLST data, collected in three districts in 2012, with MLST data collected in previous years (1996, 1999, 2002, 2003 and 2006) (Kärenlampi et al. 2007a, de Haan et al. 2010a). The frequencies of CCs were measured using Pearson chi-square and Fischer's exact tests, which were carried out by cross-tabulations for each CC in districts 1, 2 and 3. Frequency differences were considered significant at $P < 0.05$. In Study III, Simpson's index of diversity (*D*) (Simpson 1949), with values ranging from 0 (no diversity) to 1 (unlimited diversity), was used to determine the diversity of STs among organic laying hen isolates.

### 4.3.7 WGS data deposition (I, II and IV)

Draft genomes of all isolates used in wgMLST analysis in Studies I and II were deposited to PubMLST database (pubmlst.org/campylobacter). The genome data of five chicken isolates used in Study IV were deposited to the European Nucleotide Archive (ENA, EBBL) with the accession number PRJEB5964.

# 5. RESULTS

## 5.1 Molecular typing of *C. jejuni* isolates

### 5.1.1 MLST types among human, poultry and water *C. jejuni* isolates (I, II and III)

An overview of all MLST types detected in Studies I, II and III is provided in Table 3, and MLST frequencies among isolates collected in 2012 are shown in Figure 2. In Studies I and II, concerning the isolates collected in 2012, 75% of the human isolates

(n=95) and 64% of the chicken isolates (n=51) belonged to the four most frequent STs: ST-45 and ST-230 (ST-45 CC), ST-267 (ST-283 CC) and ST-677 (ST-677 CC). In addition, ST-11, ST-794 and ST-1276 were detected among both human and chicken isolates. Of the 95 human *C. jejuni* isolates, 75 had MLST types found also in chickens, resulting in 79% overlap between human and chicken MLST types. Eleven (55%) of the swimming water *C. jejuni* isolates (n=20) had ST-45, ST- 230, ST-677 and ST-945, which were also detected among human isolates. Furthermore, 49% of the human isolates had overlapping STs with swimming water isolates. Finally, 20% of human, 17% of chicken and 45% of water isolates had unique STs that were not found in the other two sources (Fig. 2).

Table 3. MLST types and frequencies among human, poultry and water isolates in Studies I, II and III.

| CC | ST | Human isolates (n=95) No. of isolates/(frequency) | Chicken isolates (n=51) No. of isolates/(frequency) | Water isolates (n=20) No. of isolates/(frequency) | Organic hen isolates (n=147) No. of isolates/(frequency) | Overall ST frequency |
|---|---|---|---|---|---|---|
| ST-21 | 19 | 2 (2.1%) | | | | 0.6% |
| | 50 | | | | 39 (26.5%) | 12.5% |
| | 451 | | | | 8 (5.4%) | 2.6% |
| ST-42 | 42 | | 1 (2%) | | | 0.3% |
| ST-45 | 11 | 1 (1.1%) | 7 (13.7%) | | 4 (2.7%) | 3.8% |
| | 45 | 20 (21.1%) | 22 (43.1%) | 6 (30%) | 17 (11.6%) | 20.8% |
| | 137 | | | | 3 (2%) | 1.0% |
| | 230 | 13 (13.7%) | 3 (5.9%) | 3 (15%) | 1 (0.7%) | 6.4% |
| | 538 | 1 (1.1%) | | | | 0.3% |
| | 583 | | 1 (2%) | | | 0.3% |
| ST-52 | 52 | | | | 8 (5.4%) | 2.6% |
| ST-61 | 61 | 2 (2.1%) | | | | 0.6% |
| ST-283 | 267 | 20 (21.1%) | 4 (7.8%) | | | 7.7% |
| | 383 | 1 (1.1%) | | | | 0.3% |
| ST-353 | 356 | | | | 17 (11.6%) | 5.4% |
| | 2801 | | | | 1 (0.7%) | 0.3% |
| ST-403 | 403 | | | | 5 (3.4%) | 1.6% |
| ST-446 | 446 | | | | 5 (3.4%) | 1.6% |
| ST-677 | 677 | 18 (18.9%) | 3 (5.9%) | 1 (5%) | | 7.0% |
| | 794 | 2 (2.1%) | 2 (3.9%) | | | 1.3% |
| ST-692 | 991 | 1 (1.1%) | | | | 0.3% |
| ST-952 | 3492 | 1 (1.1%) | | | | 0.3% |
| | 5987 | 1 (1.1%) | | | | 0.3% |
| | 6471 | | 1 (2%) | | | 0.3% |
| ST-1287 | 945 | 1 (1.1%) | | 1 (5%) | | 0.6% |
| | 6555 | | 1 (2%) | | | 0.3% |

41

| ST-1332 | 1276 | | | | |
|---|---|---|---|---|---|
| UA | | | | | |
| 1276 | 1 (1.1%) | 1 (2%) | | | 0.6% |
| 951 | 1 (1.1%) | | | | 0.3% |
| 1030 | 1 (1.1%) | | | | 0.3% |
| 1080 | 2 (2.1%) | | | | 0.6% |
| 1286 | | | 1 (5%) | | 0.3% |
| 1294 | | | 1 (5%) | | 0.3% |
| 1367 | 2 (2.1%) | | | | 0.6% |
| 1721 | | 1 (2%) | | | 0.3% |
| 1972 | | | | 2 (1.4%) | 0.6% |
| 2068 | 1 (1.1%) | | | 7 (4.8%) | 2.6% |
| 3272 | | 1 (2%) | 1 (5%) | 30 (20.4%) | 9.9% |
| 3322 | | | 1 (5%) | | 0.3% |
| 4001 | | 1 (2%) | | | 0.3% |
| 6460 | | 1 (2%) | | | 0.3% |
| **6515** | | | 1 (5%) | | 0.3% |
| **6517** | | | 2 (10%) | | 0.6% |
| **6518** | | | 1 (5%) | | 0.3% |
| **6519** | | | 2 (10%) | | 0.6% |
| 6556 | | 1 (2%) | | | 0.3% |
| 6591 | 1 (1.1%) | | | | 0.3% |
| 6626 | 1 (1.1%) | | | | 0.3% |
| **7007** | 1 (1.1%) | | | | 0.3% |

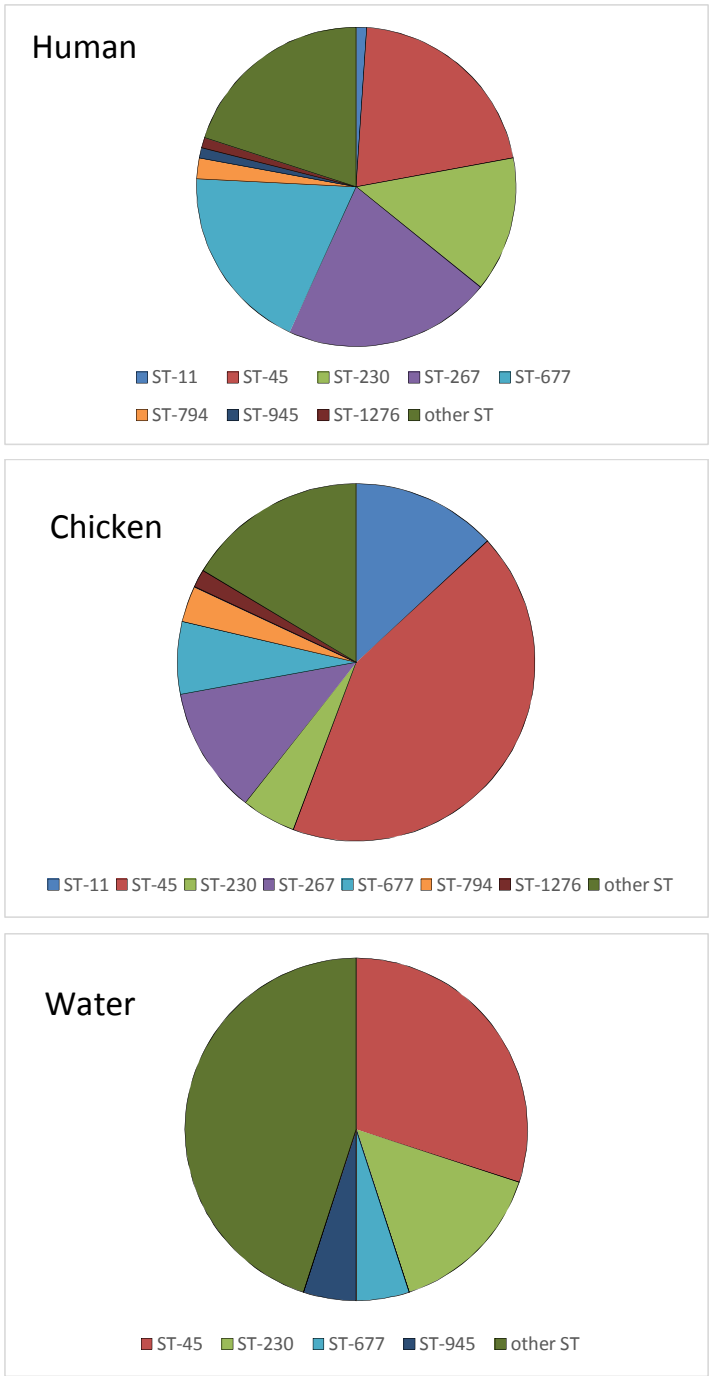Bolded STs; ST-6515-6519 (water) and ST-7007 (human), were described for the first time in these studies.

Figure 2. Frequencies (%) of the overlapping MLST types detected among domestically acquired human infections (n=95), chicken slaughter batches (n=51) and swimming water (n=20) collected in 2012 (Studies I and II). Other ST refers to unique STs found in humans, chickens or water.

In Study III, in which the *C. jejuni* isolates were collected from organically farmed laying hens in 2003 and 2004, 71% of the isolates belonged to ST-50 (ST-21 CC), ST-45 (ST-45 CC), ST-356

(ST-353 CC) and ST-3272 (UA). The latter was detected also in chicken slaughter batches in 2012 (Table 3), and the other three STs mentioned before were found in previous studies in poultry, humans and environmental waters (Fig. 3).

## 5.1.2 CRISPR sequences and association with MLST and PFGE types (III)

In Study III, a total of 20 CRISPR types, with unique spacer sequence combinations, were detected among 86 organically farmed laying hen isolates. CRISPR types were frequently associated with the MLST types. For example, among ST-50 isolates with 10 different PFGE types, all CRISPRs were identical. Two highly similar CRISPR types were also detected among ST-3272 isolates, representing nine different PFGE types (Table 2 in Study III). Similar or highly similar (one spacer difference) CRISPR types were seen also among ST-52 and ST-446 isolates. ST-45 isolates, by contrast, were most diverse, with seven different CRISPR and PFGE types. Variation in CRISPR types was seen also among ST-356, ST-11 and ST-137 isolates. On the other hand, among isolates of ST-230, ST-1972 and ST-2068, no CRISPR sequences were detected.

## 5.1.3 Phylogeny of MLST types (III)

In Study III, the ClonalFrame genealogy tree (Fig. 3) of the 14 STs detected among organic laying hens and 197 STs found from several sources in our previous studies (Kärenlampi et al. 2007a, de Haan et al. 2013) showed that isolates belonging to clonal complexes ST-21 CC and ST-45 CC clustered separately as expected. However, ST-3272 (20% of the organic laying hen isolates) and ST-1972, both unassigned to a clonal complex, clustered with ST-2801 (ST-353 CC) and were close to ST-356 (ST-353 CC), suggesting phylogenetic relatedness of these STs to ST-353 CC.
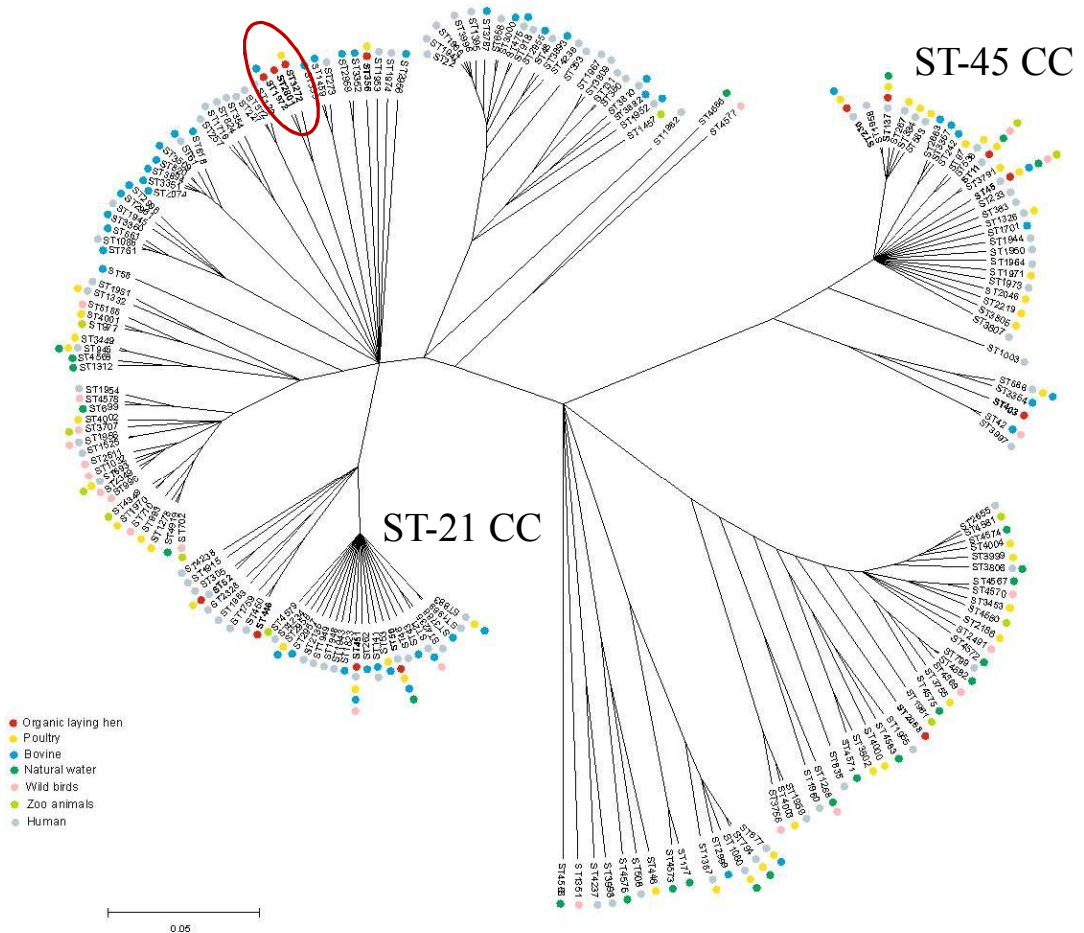
Figure 3. ClonalFrame genealogy of the 197 STs detected from seven different sources; organic laying hens, poultry, bovines, water, wild birds, zoo animals and human patients (Kärenlampi et al. 2007a, de Haan et al. 2010a&b, de Haan et al. 2013) are indicated with different colours. ST-3272 (UA) and ST-1972 (UA) found among organic laying hens seem to be closely related to ST-2801 (marked with an ellipse) and other isolates belonging to ST-353 clonal complex. Major clonal complexes ST-45 CC and ST-21 CC are indicated.

## 5.2 Epidemiological analyses

### 5.2.1 Epidemiology of human infections (I and II)

The overall incidences, including both imported and domestic human infections, during the seasonal peak in 2012 were 39.1 (district 1), 30.2 (district 2) and 40.4 (district 3) per 100 000 inhabitants. In Studies I and II, the incidences covering human infections (n=109) from only domestically acquired infections were approximately half of the total incidences, resulting in incidences of 15.4 (district 1), 20.6 (district 2) and 15.5 (district 3) per 100 000 inhabitants. Men were more commonly infected, comprising 55% of the cases in district 1 and 63% of the cases in districts 2 and 3. The proportion of young adults (20-29 years) comprised 9% and patients older than 60 years comprised 40% (district 1), 42% (district 2) and 34% (district 3) of the total cases. Only three cases (one in each district) were obtained from children under 10 years.

### 5.2.2 Temporal association of human, chicken and water isolates (I and II)

In Study I, most of the human isolates that represented the most frequent STs (ST-45, ST-230, ST-267 and ST-677) were detected from early July to mid-August in districts 1 and 3. The highest peaks were seen in July during weeks 29-31. In district 2, by contrast, the human infections were distributed more evenly from June to August, with no distinct peaks seen.

In Study II, when the temporal connection (chicken slaughter batch tested 2-23 days before human patient sampling) of the human and chicken isolates was taken into consideration, in addition to the ST similarity, the MLST overlap of the isolates decreased from 79% to 48%. Most chicken (82.5%) and human (92.4%) isolates, representing ST-45, ST-230 and ST-267, were detected from mid-July to mid-August. The swimming water ST-45, ST-230 and ST-677 isolates from June to mid-July, instead, were in most cases (71.4%) detected prior to the human and chicken isolates. Furthemore, almost all human ST-677 isolates (16 of 18) were detected before the positive chicken slaughter batches (Fig. S1 in Study II).

### 5.2.3 *C. jejuni* on poultry farms (II and III)

In Study II, a total of 1534 chicken slaughter batches (originating from 124 farms) were analysed from June to October, resulting in *C. jejuni* prevalence of 5.3% and a prevalence of 1.6% in the randomized sampling from November to May. From June to October, 37 of the 124 farms produced *C. jejuni*-positive chicken slaughter batches from 1 to 4 times during the study period. On the farms that produced *C. jejuni*-positive chicken batches more than once, the batches were colonized in different rearing cycles by either similar (1 farm) or different STs (5 farms).

In Study III, sampling of organically farmed laying hens was performed on 18 farms in autumn 2003 and spring 2004. On four farms, the hen flock had changed between samplings and two of the farms were only once positive for *C. jejuni*. However, ST-50 was detected in both samplings at one farm and ST-45 in autumn and ST-50 in spring at another farm. Ten of the 18 farms had the same flock of hens in both samplings, and in these hens, ST-50 and ST-3272 were detected in both samplings on four out of seven and eight farms, respectively, indicating that these STs often persisted among chicken flocks. By contrast, ST-45 and ST-356 were detected in both samplings only once out of seven and five farms, respectively. ST-1972 was also detected in both samplings on one farm, but other rarely detected STs were found only in autumn or in spring.

## 5.3 WGS analyses of human, chicken and water *C. jejuni* isolates

### 5.3.1 WgMLST and the association of human, chicken and water isolates (I, II and IV)

Discrimination of human, chicken and water isolates and association between the sources was determined using wgMLST implemented by phylogenetic networks and by additional manual inspection of allelic differences between closely related isolates. Two different methods were used, and in Studies I and IV, wgMLST was performed using BIGSdb, which detected a smaller number of shared genes than GeP, used in Study II, which also accounted for the changes in the homopolymeric tracts (poly C or poly G tracts) that were not observed using BIGSdb.

In Study II, human, chicken and water isolates representing ST-45 shared 1391 gene loci and formed three distinct clusters, differing from each other by 400 to 550 alleles. Clusters were seen also among isolates of ST-230, ST-267 and ST-11 isolates, however, they were not as diverse as clusters of ST-45, and had allelic differences between 63 (ST-11) and 120 (ST-230). ST-677 isolates did not form distinguished clusters and isolates differed from each other by 22 to 113 alleles, except for three chicken isolates, originating from batches reared on the same farm, which were more similar to each other and had seven allelic differences.

Further wgMLST analyses revealed three human isolates in cluster 1 (1644 shared loci) and one human isolate in cluster 2 (1605 shared loci) that were both temporally and genetically related to chicken isolates. Among ST-230 (1541 shared loci), one cluster of closely related isolates was identified and in further wgMLST analysis (1612 shared loci), two human isolates were found to be genetically and temporarily associated with chicken isolates. Fifteen out of 18 ST-267 human isolates formed a genetically highly similar cluster (0-4 SNPs) with the chicken isolates, within 1566 shared loci, and a temporal connection was also verified. A human isolate was both temporally and genetically associated with a chicken isolate within the cluster of ST-11 isolates (1615 shared loci). Human ST-677 isolates that had no temporal connection to chicken or water isolates were also not genetically related,

and dozens (from 22 to 103) of allelic differences occurred between the isolates. Eleven farms produced several *C. jejuni*-positive slaughter batches (two to four) during a single rearing cycle. All isolates from different slaughter batches that originated from the same farm and had the same ST were also closely related in wgMLST, differing only by 0 to 5 SNPs, except in one case where two different clones of ST-45 with 70 allelic differences were observed. In addition to related chicken and human isolates, wgMLST revealed also human isolates that were genetically highly similar, suggesting a common unknown source of infection, which was seen in both Studies I and II.

By combining wgMLST results (≤ 5 SNPs between genetically related isolates) with the temporal connection (chicken slaughter preceded patient sampling by 2-23 days), we found that altogether only 22 of the 92 human *C. jejuni* isolates were both genetically and temporally associated with the chicken isolates, and thus, chicken could be a potential source of human infection in 24% of the cases. None of the swimming water isolates were genetically related to the human isolates, and temporal connection was also lacking in most cases.

In Study IV, chicken farm isolates of the two STs; ST-677 and ST-794 (ST-677 CC), diverged from each other by 170-180 alleles in wgMLST (1286 shared loci). Isolates that were collected from a single farm (farm A or farm C) were highly similar, as wgMLST showed no allelic differences. Unrelated clinical isolates from the UK, differed by dozens (9 to 41) of alleles from farm isolates, however, one human isolate (LGM9872) that was isolated in 1979 from Sweden, showed only 9 allelic differences (SNPs) with the farm ST-677 isolates (1319 shared loci) collected from Finland in 2003 (Fig. 2 in Study IV).

### 5.3.2 ClonalFrame genealogies (I and IV)

ClonalFrame genealogy was performed for assembled draft genomes of the same sets of isolates that were analysed using wgMLST. In Study I, human isolates that were identical in wgMLST, performed using BIGSdb, showed minor variation in ClonalFrame (Fig. S2 in Study I), which depicts the genetic variation occurring in the whole genome, not only among the shared genes of isolates. Also, in Study IV, the unrelated human isolate (LGM9872) was more distinct to chicken isolates from farm A, which, in wgMLST, showed only 9 SNP differences. However, three ST-677 isolates (farm A) and two ST-794 isolates (farm C) that had no allelic variation in wgMLST were similar also in ClonalFrame genealogy.

### 5.3.3 Genomic characterization of ST-677 CC isolates (IV)

In Study IV, wgMLST was performed using BIGSdb and the allelic profiles defined in the database showed no allelic differences within ST-677 (farm A) or ST-794 (farm B) isolates. However, the Breseq pipeline showed allelic variation in poly C or poly G homopolymeric

tracts, indicating possible phase variability (ON or OFF phase) of these genes. Changes in the lengths of homopolymeric tracts occurred in genes encoding proteins associated with capsule biosynthesis, LOS locus, flagellin modification and putative periplasmic proteins. In addition, a SNP was observed between the two ST-794 isolates in a hypothetical protein located in *C. jejuni* integrated element (CJIE1) region.

Most of the genomic variation was observed in CJIE1- and CJIE2-like integrated elements between the Finnish chicken farm ST-677 CC isolates (farm A and farm B) and unrelated clinical isolates from the UK and Sweden. CJIE1 was located in the same genomic region in all ST-677 and ST-794 isolates; however, among ST-677 isolates (farm A), a deletion of 24 open reading frames (ORFs) occurred compared with one clinical isolate (OX7358). Otherwise, CJIE1 seemed to have conserved gene content among all ST-677 CC isolates, but when compared with CJIE1 of the reference strain RM1221 (Fouts et al. 2005), 55 and 35 ORFs were missing in farm A and farm C isolates, respectively. CRISPR sequences were not conserved in ST-677 and ST-794 isolates, suggesting a degenerated CRISPR-Cas system among ST-677 CC.

Putative virulence-associated genes in ST-677 CC isolates, located in LOS locus, flagella and capsule, were further investigated comparing the genome sequence of chicken ST-677 isolate 5070 (farm A) with known reference genomes from GenBank. ST-677 CC isolates had LOS class O biosynthesis locus, which showed high genomic similarity (90-99% amino acid identity) with the corresponding class O reference strain RM3423 (Parker et al. 2008). Also, the flagellar glycosylation gene locus of ST-677 CC isolates showed high similarity to *C. jejuni* strain M1 (ST-137, ST-45 CC) (99-100% aa identity). The whole capsular polysaccharide gene locus (size of 37 071 bp) had high gene content similarities with *C. jejuni* strain 32488 (ST-1460, ST-48 CC) (97% full-length identity), Penner serotype HS4 type strain ATCC43432 (92-100% aa identity) and some similarities with *C. jejuni* subsp. *doylei* strain 269.97, from which the two latter have been recognized as highly virulent strains. Also, the phase-variable gene *wcb*K (GDP-mannose 4,6-dehydratase), located in the capsule locus, was more closely related to *C. jejuni* subsp. *doylei* strain 269.97 than several other *C. jejuni* reference strains, indicating that this gene has not evolved in strict association with MLST lineages or different subspecies. Furthermore, this gene was completely missing from several reference strains and lineages, and the phase-variable poly G tract was present only in one cluster containing the ST-677 CC. In addition, several other putative virulence-associated genes and features were identified, including the highly degenerated cytolethal distending toxic operon in all ST-677 CC isolates.

# 6. DISCUSSION

## 6.1 MLST and other typing methods as tools in molecular epidemiology of *C. jejuni* (I, II and III)

Since 2001, when the MLST scheme was developed for *C. jejuni* (Dingle 2001), this molecular typing method, which is suitable for large sets of bacterial isolates, has provided valuable information about the molecular epidemiology, population structure and evolution of different genetic lineages of this organism at a global scale. All defined MLST types (8855 STs at pubMLST database in 20/03/2017) and a variety of different isolates, including information about the source, year of collection and country of origin, is available at the pubMLST webpage for *Campylobacter* (pubMLST.org/campylobacter). MLST data can be utilized, for example, to estimate population structure of *C. jejuni* isolates and to connect human infections to their potential sources, as shown in numerous studies performed in different countries (Lévesque et al. 2008, Mullner et al. 2009, Habib et al. 2009, Sheppard et al. 2009a, de Haan et al. 2010a&b, Griekspoor et al. 2013, Meinersmann et al. 2013, Llarena et al. 2015a&b). Studies have often indicated chickens, amongst other poultry, as a major source of human infections, which is supported by the knowledge that similar MLST types are frequently found among both human patients and chickens (Colles & Maiden 2012, de Haan et al. 2014). Further, mathematical source attribution models applying MLST data have emphasized the importance of poultry as a source of human infections (Wilson et al. 2008, Sheppard et al. 2009b, Mughini-Gras et al. 2012, de Haan et al. 2013, Lévesque et al. 2013). In this thesis, MLST typing was applied to human, poultry and water isolates analysed in Studies I, II and III.

MLST has revealed that among two common clonal complexes, generalist ST lineages, such as ST-45 (ST-45 CC) and ST-50 (ST-21 CC), exist. These STs colonize a large variety of host animals, including both domestic and wild animals (mammals and birds), indicating a wide ecological adaptation (Gripp et al. 2011, Llarena et al. 2016). In addition, these multi-host STs have been commonly found worldwide at different times, indicating temporal persistence as well (Kärenlampi et al. 2007a, Sheppard et al. 2009a, Habib et al. 2009, Mullner et al. 2009, Griekspoor et al. 2010). Some CCs and STs, by contrast, have been described to colonize mostly specific host animals such as ST-61 CC (especially ST-61) in bovines (Colles et al. 2003, Kwan et al. 2008, de Haan et al. 2010b), ST-1034 CC in geese (Colles et al. 2008, Llarena et al. 2015a) and ST-403 CC in sheep (Sheppard et al. 2009a). Also, certain lineages seem to be more common in certain geographical areas, e.g. ST-677 CC (especially ST-677 and ST-794) common among *C. jejuni* isolates collected from several sources, including human patients, poultry, water and wild birds, in Finland and other Nordic countries (Griekspoor et al. 2013, de Haan et al. 2014). However, limited availability of MLST data collected from several geographical regions prevents understanding of the global distribution of certain STs such as ST-677.

In this thesis, ST-45 was found in high frequencies among human patients (21%), chicken batches (43%), swimming water (30%) and organic laying hens (12%), confirming this ST as a predominant genotype and a major generalist ST in Finland (Studies I, II and III). The same observation was made in previous Finnish MLST studies concerning isolates collected since 1996 (Kärenlampi et al. 2007a, de Haan et al. 2010a&b, de Haan et al. 2013, Llarena et al. 2015b). Also, ST-230 (ST-45 CC) was found among human, poultry and water isolates (total frequency 6.4%). The second most common ST among human patients, occurring with a frequency of 21%, was ST-267 (ST-283 CC), which was also found among 8% of chicken slaughter batch isolates in 2012 (Study II), but was not detected among organically farmed laying hens in 2003 and 2004 (Study III). The third most common ST among human patients was ST-677, accounting for 14% of the isolates, and it was also found with minor proportions among chickens and swimming water in 2012. This ST, besides being common among domestic human *C. jejuni* infections in Finland (de Haan et al. 2014), has been connected to more severe invasive bacteraemia infections among Finnish patients (Feodoroff et al. 2013). As a rare finding in previous studies, ST-3272 (UA) was represented by 20% of the organically farmed laying hen isolates (Study III), and it was also found among conventionally reared chicken slaughter batches in 2012 (Study II), suggesting optimal adaptation to the poultry host. Interestingly, ST-50 (ST-21 CC), which has previously been commonly found in several sources in Finland (de Haan et al. 2014), was overrepresented among organic laying hens, but not found among human, chicken or swimming water in 2012, suggesting that the reservoir of this ST was not available at that time.

Sequence types are assigned to clonal complexes if they share at least four alleles with the founder ST of the clonal complex (Dingle et al. 2001). If common alleles are missing, the ST remains unassigned to any known CCs, even though it would seem related to certain clonal complexes in population structure analyses. In Study III, the ClonalFrame genealogy showed the phylogenetic relationships of 197 *C. jejuni* STs, revealing the close connection of ST-3272 (UA) and ST-1972 (UA) with ST-356 and ST-2801 (ST-353 CC); however, due to the absence of four joint alleles with the founder ST of this clonal complex (ST-353), ST-3272 and ST-1972 are classified unassigned. The controversy between definition of founder genotypes and CCs may be more common in the MLST typing scheme. This phenomenon has been found also in analyses of WGS data, revealing genetically distinct subpopulations among isolates within certain STs. For example, among ST-45, distinct subpopulations have been described (Llarena et al. 2016), underlying the hypothesis that MLST typing itself is insufficient for discriminating isolates in outbreaks (Clark et al. 2012). However, the PubMLST database and the defined, common nomenclature of MLST types provide essential primary tools for analysing the population genetics and global molecular epidemiology of the isolates, followed by deeper investigation of isolates of special interest.

In addition to MLST, other genotyping methods have been developed and applied to distinguish *C. jejuni* isolates, such as PFGE, which has been used for *C. jejuni* since the 1990s (On et al. 1998, Hänninen et al. 1998, Hänninen et al. 2003, Kärenlampi et al. 2007b). PFGE has been widely used for discriminating isolates and identifying similar strains in, for

example, outbreak studies (Hänninen et al. 2003, Hakkinen et al. 2009, Revez et al. 2014b), however, it has limitations as a unified typing method for comparing larger data and genotypes between studies (Wassenaar & Newell 2000). In Study III, PFGE and MLST types were compared in their ability to characterize *C. jejuni* isolates from organically farmed laying hens, and PFGE was found to be more discriminatory, yielding 42 distinct types compared with 14 STs obtained using MLST. The most diverse PFGE types occurred among ST-50 (10 PFGE types), ST-3272 isolates (9 PFGE types) and ST-45 isolates (7 PFGE types), and differences in PFGE patterns were seen in one to four bands among those PFGE types that represented the same ST. PFGE has been shown to be more discriminatory than MLST (Kärenlampi et al. 2007b), and *Kpn*I, which was used in this study, is highly discriminatory (Michaud et al. 2001).

Knowledge about the role of the CRISPR region in the bacterial immunity system controlling genome maintenance has accumulated more recently, and the mechanisms are still only partly understood (Rath et al. 2015). In Study III, CRISPR typing among *C. jejuni* isolates representing certain STs revealed highly variable spacer combinations, similarly as described in a previous study using CRISPR typing for *C. jejuni* (Schouls et al. 2003). Among organic laying hen isolates, CRISPR types were strongly associated with ST-50, ST-3272, ST-52 and ST-446 isolates. However, especially among ST-45 isolates, several diverse CRISPR types occurred, which may reflect the genetic diversity of ST-45 described recently (Llarena et al. 2016). In Study III, CRISPR sequencing provided 20 CRISPR types, and thus, was more discriminatory than MLST, which is reasonable based on the function of the CRISPR-Cas system in reflecting bacteriophage contacts (Bolotin et al. 2005, Deveau et al. 2010). In conclusion, CRISPR sequences seem to be either highly variable or stable depending on the ST that they represent, providing supportive information about the CRISPR-Cas system in the genomes of different STs and CCs.


## 6.2 WgMLST and tracing sources of human *C. jejuni* infections (I and II)

High-throughput whole genome sequencing of larger collections of bacterial genomes has become more affordable and feasible for studies on molecular epidemiology and genomic evolution of bacteria (Sheppard et al. 2012, Cody et al. 2013, Taboada et al. 2013, Loman et al. 2015, Llarena et al. 2017). The applications utilizing WGS data provide information of bacterial isolates with much higher resolution than the conventional sequence-based methods such as MLST.

In study I, during the seasonal peak, *C. jejuni* isolates from domestic infections covered approximately half of the total number of registered infections. Common knowledge about the seasonal distribution of *Campylobacter* infections, acquired either from domestic sources or during foreign travel, is lacking (THL 2016), but our estimate is concordant with the results of a previous study (Vierikko et al. 2004). Distinct peaks were seen in the

occurrence of infections in districts 1 and 3, whereas in district 2, infections were dispersed more evenly. This result may indicate that in district 2 sources of infections were different or more continuous compared with districts 1 and 3. WgMLST revealed genetically related human isolates among ST-45, ST-230, ST-267 and ST-677 isolates originating from the same and also different districts, suggesting these sporadic infections could have the same infection source.

In Study II, we used gene-by-gene algorithm GeP, recently developed in our group (Zhang et al. 2015), to compare human, chicken and water isolates. WgMLST and phylogenetic Split decomposition networks revealed three highly distinct clusters among ST-45 isolates, indicating that distinct sub-lineages exist under this ST. Also, a recent study using large, international WGS data of ST-45 isolates confirmed the presence of monomorphic sub-lineages within this ST (Llarena et al. 2016). Human isolates clustered similarly in both Studies I and II; however, Study II, where wgMLST was performed using GeP, revealed more shared loci and allelic differences between the isolates relative to BIGSdb, which only detects the alleles defined in the database (pubMLST.org/campylobacter).

Most of the genetic variation between highly similar human and chicken isolates was observed in single-nucleotide polymorphisms (SNPs) and changes in the lengths of homopolymeric tracts of certain phase-variable genes, the latter of which was described by Bayliss et al. (2012). The homopolymeric tract changes were not detected between genetically related human isolates in Study I, however GeP, used in Study II, utilizes BLAST searches of chosen reference genome against the query genomes under investigation and produces information with higher resolution than wgMLST performed using BIGSdb. In Study II, we were able to confirm this high discriminatory power of GeP in comparative genomics using a larger collection of temporarily related *C. jejuni* isolates from three sources, i.e. human patients, chicken and water, by observing changes of homopolymeric tract lengths and also the higher numbers of shared genes among the isolates.

In addition to WGS comparisons, epidemiological data, such as information of the isolation dates of *C. jejuni* from patients and different sources of infections, as well as information about travelling status (domestic versus foreign) of the patients, are essential in tracing the sources of domestically acquired human infections. All human isolates were selected to represent acquisition from domestic sources during the summer peak similarly as chicken slaughter batch isolates. In Study II, the highly accurate wgMLST combined with the knowledge of temporal association revealed that only 24% of the domestic human infections could be traced back to chicken slaughter batches.

In Study II, genetically and temporally associated human and chicken isolates were recognized among both ST-45 and ST-230. Most of the human ST-267 isolates were associated with chicken isolates both genetically and temporally, and isolates formed a highly clonal cluster with only 0 to 4 SNP differences. Besides human isolates that originated from three differents districts, also chicken ST-267 isolates were genetically highly similar to each other, although originating from three different farms. This result suggests that ST-

267 is highly clonal, warranting further WGS investigation to study the ecology of this ST. Interestingly, lack of both genetic and temporal associations between human and chicken ST-677 isolates raises a question about the reservoir of this ST. Most of the human infections appeared one to two weeks before the positive chicken batches were detected. This phenomenon has been seen also earlier in Finland, when the peak in the occurrence of ST-677 was detected in humans prior to chickens (de Haan et al. 2014, Llarena et al. 2015b). A common reservoir of ST-677 for humans and chickens cannot be excluded since chicken flocks have been shown to be colonized by *C. jejuni* during the first weeks of the rearing period (Hermans et al. 2012), resulting in a more simultaneous colonizing time of humans and chickens. However, as the human and chicken isolates in Study II were not genetically related the source may also be different. These results suggest that chickens, despite being colonized by ST-677, are not the source of human ST-677 infections in Finland. Since ST-677 has previously been detected also from wild birds in Sweden and UK (Griekspoor et al. 2013, Griekspoor et al. 2015) and from environmental waters (pubMLST.org), environmental sources in human infections in Finland might have been underestimated.

In principle, in Study II, all chicken meat reaching markets was tested for *C. jejuni* as each slaughter batch between June and October was sampled according to EU legislation (MMMa 2007). According to the official monitoring programme, only a single colony was chosen from the pooled sample for further analyses. Therefore, there is a possibility that all different genotypes present in the batches were not detected due to the sampling procedures causing a bias in our analysis. However, a targeted Finnish study has been shown that once the chicken flock is colonized by *C. jejuni*, the majority of the birds are colonized by a single PFGE type (Hakkinen & Kaukonen, 2009, presented at the 15th International Workshop on *Campylobacter*, *Helicobacter* and related organisms, Niigata, Japan). Further, in Study II, wgMLST revealed that *C. jejuni* isolates from chicken slaughter batches that were reared simultaneously on the same farm but in different broiler houses were mostly colonized by a single wgMLST genotype. This was verified by only 0 to 5 SNPs that occurred between these isolates in all cases except one, where two ST-45 isolates, detected from different batches reared at the same farm, had a total of 70 allelic differences, revealing two distinct wgMLST genotypes.

Altogether, our result that only 24% of human domestically acquired infections during the summer peak could be associated with chickens differ from earlier source attribution modelling studies performed in Finland, where 45.5% (de Haan et al. 2013), in Canada, where 65% (Lévesque et al. 2013), or in UK, where up to 97% (Wilson et al. 2008) of patient isolates were attributed to chickens. On the other hand, our result is concordant with the estimation of EFSA BIOHAZ panel (EFSA 2011) that in general 20-30% of human infections may be directly associated with the consumption of chicken meat. The low association in Study II can be explained by several facts. First, due to a centralized chicken meat production chain and a high biosecurity level in primary production in Finland, the prevalence of *C. jejuni*-positive batches has been continuously much lower than in other EU countries (5.3% in Finland versus 62% in Spain in 2012) (EFSA 2014). Second, epidemiological data was well

characterized, allowing measurement of the temporal connection between our human and chicken isolates. Also, the decreased overlap between human and chicken isolates was seen in previous Finnish studies, where the PFGE similarity between human and chicken isolates decreased from 70% to 31% (Hakkinen et al. 2009) and the similarity of combined PFGE and Penner serotypes decreased from 46% to 31% (Kärenlampi et al. 2003) when the temporal relationships were taken into consideration. Third, the highly discriminatory wgMLST method was used for genome comparison, allowing much higher resolution than the traditional MLST analysis.

Three common STs among human patients (ST-45, ST-230 and ST-677) were also found in swimming water isolates, suggesting the wide presence of these STs also in the environment. However, since none of the water isolates were genetically associated with the human isolates, we were unable to confirm the results of a previous case–control study, which recognized swimming in natural waters as a significant risk factor for acquiring campylobacteriosis during the summer peak in Finland (Schönberg-Norio et al. 2004). One reason for this is the complexity of environmental sampling, which is rather sporadic, consisting of monthly sampling by the municipal inspectors. To increase the possibility of detecting isolates from water, which could reveal an association with human infections, sampling should be performed more often and should cover more sites.

In conclusion, wgMLST used in Studies I and II revealed distinct clusters within certain sequence types, especially among ST-45. This was also revealed in a recent study where distinct sub-lineages, which included spatially and temporally unrelated isolates, were found among the ST-45 (Llarena et al. 2016), suggesting that monomorphic clonal sub-lineages, persisting in different times and locations, may exist within STs. Methods utilizing WGS data, such as wgMLST, can be used to provide accurate genomic information of the isolates with a much higher resolution than before. Isolates can be distinguished from each other by a single SNP and together with epidemiological data, such as temporal information, genetically similar sporadic isolates can be recognized and sources of human infections identified.

## 6.3 Comparative genomics of ST-677 CC (IV)

*C. jejuni* isolates belonging to ST-677 CC, and especially ST-677, are more common in Finland than elsewhere (McCarthy et al. 2012, de Haan et al. 2014). Also, ST-677 isolates have been associated with more severe symptoms of the disease, including frequent hospitalization and longer stay in hospital (Kärenlampi et al. 2007a) as well as bacteraemia (Feodoroff et al. 2013). Several genomic features and putative virulence-associated factors have been discovered for different strains of *C. jejuni* (Bacon et al. 2000, Fouts et al. 2005, Bolotin et al. 2005, Skarp et al. 2015). In Study IV, we characterized genomic features occurring among ST-677 CC isolates originating from Finnish chicken farms and clinical

isolates from Sweden and the UK. Altogether, ST-677 and ST-794 (ST-677 CC) were fairly closely related and had only 170-180 allelic differences in their 1286 shared genes, measured in wgMLST. Isolates within ST-677 were highly clonal and between the related farm isolates and unrelated clinical isolates, only 9 to 41 SNPs occurred among the 1319 shared genes in wgMLST, even though in some cases the isolates had been collected 25 years apart.

Most of the genomic variation between the isolates was observed in genes associated with *C. jejuni* integrated elements CJIE1 and CJIE2. In Study IV, CJIE1 was completely or partially present in all studied ST-677 CC isolates and it was found also among clinical *C. jejuni* ST-677 CC isolates in a recent study (Skarp et al. 2015). Furthermore, the 24 ORF deletion in CJIE1 region, which was detected among some isolates in our study, was found also among part of the clinical isolates originating from human patients with bacteremia or gastroenteritis (Skarp et al. 2015). In Study IV, CJIE2-like element was found from all farm ST-677 CC isolates but was present in only one clinical isolate. Similar result was obtained in a study by Skarp et al. (2015), in which CJIE2 was present in part of the blood and faecal isolates. This finding may suggest that CJIE2 is more recently obtained than the more conserved CJIE1, which was found in both farm and clinical isolates. One explanation for ST-677 CC remaining relatively stable might be the presence of CJIE1, which may prevent the natural transformation of the strains. This was shown in a previous study in which the DNase activity, encoded by *cje0256* present in CJIE1, was found in non-naturally competent cells, whereas naturally competent strains lacked this activity (Gaasbeek et al. 2009). All ST-677 CC isolates in this study contained a 100% amino acid (aa) homologue of *cje0256*. Also, homologues (with 90% aa identity) were found from DNA non-specific endonuclease gene *cje0566,* associated with decreased efficiency of natural transformation (Gaasbeek et al. 2010), between six ST-677 isolates of this study and with *C. jejuni* reference strain RM1221. The CRISPR-Cas system, which has been proposed to be involved to immune defense against exogenous genetic elements, such as phages and plasmids (Deveau et al. 2010, Louwen et al. 2013, Rath et al. 2015), was found to be degenerated and most likely non-functional in all ST-677 CC isolates.

Among the farm isolates, only a single SNP in one gene was observed, and in addition, changes in homopolymeric tracts in a few genes associated with capsular biosynthesis, lipo-oligosaccharide (LOS) locus and flagella. Phase variation in genes encoding these surface structures has been linked to colonization, adhesion and increased virulence of *C. jejuni* (Jerome et al. 2011, Keo et al. 2011). Homopolymeric tract changes were observed in phase-variable capsule-associated gene *wcb*K, encoding GDP-mannose 4,6-dehydratase, which we hypothesized to have a major role in pathogenesis by conferring acid and serum resistance in different phases during infection. Also, in a recent study by Skarp et al. (2015), the higher occurrence of ST-677 CC in bacteremia was associated with serum resistance and phase variation *wcbK*. In that study, ST-677 CC isolates originating from patients with bacteremia and gastroenteritis were investigated and isolates with 9-G homopolymeric tract were found to encode intact *wcb*K resulting in lower serum resistance, while isolates with 8-G and 10-G

homopolymeric tract, expressed higher serum resistance and encoded most likely truncated *wcb*K (Skarp et al. 2015).

Capsule polysaccharide (CPS) of *C. jejuni* has been recognized as a major component in Penner serotyping scheme (Karlyshev et al. 2000) and this region has been found to be a highly variable between strains (Karlyshev et al. 2005). However, in this study, the whole capsular locus (37 071 bp) of ST-677 isolate 5070 showed high genetic similarities with *C. jejuni* subsp. *doyley* 269.97 and with a type strain of Penner serotype HS4 (ATCC43432), both of which have often been found among human bacteraemia cases (Skirrow et al. 1993, Morey et al. 1996). Further, in a study by Skarp et al. (2015) capsular locus among clinical ST-677 CC isolates showed high similarities to Penner serotype HS4 type strain. Cytolethal distending toxin (*cdt*) operon that has previously been connected to the virulence of *C. jejuni* (Poly & Guerry 2008), was found to be degenerated among ST-677 CC isolates in our study and also among clinical isolates in a study by Skarp et al. (2015). The *cdt* operon has been found to be conserved in many *C. jejuni* isolates and a highly degraded among *C. jejuni* subsp. *doylei* isolates and association between the absence or truncation of *cdt* genes and bacteremia has been suggested (Parker et al. 2007).

Overall, the genomic findings identified in Study IV that are typical for ST-677 CC isolates may elucidate why ST-677 CC isolates cause more severe human infections in Finland. However, more research is still needed to verify our findings and to define the role of different genes and their functions in more detail under different environmental conditions and during human infections.

# 7. CONCLUSIONS

1. Prevalent MLST types (ST-45, ST-230, ST-267 and ST-677) were detected among human domestic *C. jejuni* infections collected from three districts during the seasonal peak in 2012. The highly accurate wgMLST revealed distinct sub-lineages within certain STs and identified genetically related isolates originating from different districts, suggesting the same source among sporadic human infections (I).

2. In 2012, 79% of the MLST types of human isolates overlapped with chicken isolates but when temporal association was included, the overlap decreased to 48%. Further, when wgMLST was combined with temporal data, the overlap decreased to 24%, suggesting other, uncharacterized and possibly environment-associated sources of human infections (II). This is supported also by the result that ST-677 occurred in human patients prior to chicken slaughter batches, indicating a potential environmental source of ST-677 both for humans and chickens (II).

3. Furthermore, 20% of human isolates were represented by unique, rarely detected STs, also suggesting potential environmental sources (I, II), however, we were unable to confirm swimming water as a source of human *C. jejuni* infections probably due to sporadic sampling (II).

4. Predominant MLST types (ST-50, ST-45 and ST-3272) were found also among organically farmed laying hens. CRISPR sequences were in most cases associated with the MLST types, suggesting that CRISPRs likely depict genomic features of the isolates representing certain STs (III).

5. The genome sequences of ST-677 CC were highly conserved and only nine SNPs occurred between unrelated ST-677 isolates collected 25 years apart. The cut-off value of 0 to 5 SNPs, which was used to define the genetically similar isolates (used in Studies I and II), was supported by the results of Study IV.

6. The high frequency of ST-677 among human bacteraemia may be linked to the capsule, phase-variation observed in the *wcb*K and novel putative virulence-associated genes identified among the ST-677 isolates (IV).

# 8. REFERENCES

Aarestrup FM and Engberg J. 2001. Antimicrobial resistance of thermophilic *Campylobacter*.Vet Res. 2001 May-Aug;32(3-4):311-21.

Adhikari B, Connolly JH, Madie P, Davies PR. 2004. Prevalence and clonal diversity of *Campylobacter jejuni* from dairy farms and urban sources. N Z Vet J. 2004 Dec;52(6):378-83.

Aidley J and Bayliss CD. 2014. Repetitive DNA: A major source of genetic diversity in *Campylobacter* populations. In: Sheppard SK, editor. *Campylobacter* Ecology and Evolution. 1st edition. Norfolk, UK: Caister academic Press; 2014. pp. 55-72.

Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011, 12: 402-2164.

Alter T, Gaull F, Kasimir S, Gürtler M, Mielke H, Linnebur M, Fehlhaber K. 2005. Prevalences and transmission routes of *Campylobacter* spp. strains within multiple pig farms. Vet Microbiol. 2005 Jul 1;108(3-4):251-61.

Amar C, Kittl S, Spreng D, Thomann A, Korczak BM, Burnens AP, Kuhnert P. 2014. Genotypes and antibiotic resistance of canine *Campylobacter jejuni* isolates.Vet Microbiol. 2014 Jan 10;168(1):124-30.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008 Feb 8;9:75

Bachtiar BM, Coloe PJ, Fry BN. 2007. Knockout mutagenesis of the *kps*E gene of *Campylobacter jejuni* 81116 and its involvement in bacterium-host interactions. FEMS Immunol Med Microbiol. 2007 Feb;49(1):149-54.

Bacon DJ, Alm RA, Burr DH, Hu L, Kopecko DJ, Ewing CP, Trust TJ, Guerry P. 2000. Involvement of a plasmid in virulence of *Campylobacter jejuni* 81-176. Infect Immun. 2000 Aug;68(8):4384-90.

Baig A and Manning G. 2014. *Campylobacter* association with human host. 2014. In: Sheppard SK, editor. *Campylobacter* Ecology and Evolution. 1st edition. Norfolk, UK: Caister academic Press; 2014. pp. 99- 110.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA,

Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012 May;19(5):455-77.

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature. 2009 Oct 29;461(7268):1243-7.

Barton C, Ng LK, Tyler SD, Clark CG. 2007. Temperate bacteriophages affect pulsed-field gel electrophoresis patterns of *Campylobacter jejuni*. J Clin Microbiol. 2007 Feb;45(2):386-91. Epub 2006 Nov 29.

Batchelor RA, Pearson BM, Friis LM, Guerry P, Wells JM. 2004. Nucleotide sequences and comparison of two large conjugative plasmids from different *Campylobacter* species. Microbiology. 2004 Oct;150 (Pt 10):3507-17.

Bayliss CD, Bidmos FA, Anjum A, Manchev VT, Richards RL, Grossier JP, Wooldridge KG, Ketley JM, Barrow PA, Jones MA, Tretyakov MV. 2012. Phase variable genes of *Campylobacter jejuni* exhibit high mutation rates and specific mutational patterns but mutability is not the major determinant of population structure during host colonization. Nucleic Acids Res. 2012 Jul;40(13):5876-89.

Benjamin J, Leaper S, Owen RJ and Skirrow MB. 1983. Description of *Campylobacter laridis*, a new species comprising the nalidixic-acid resistant thermophilic (NARTC) group. Curr. Microbiol.8:231-238.

Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2014. GenBank.Nucleic Acids Res. 2014 Jan;42(Database issue):D32-7.

Biggs PJ, Fearnhead P, Hotter G, Mohan V, Collins-Emerson J, Kwan E, Besser TE, Cookson A, Carter PE, French NP. 2011. Whole-genome comparison of two *Campylobacter jejuni* isolates of the same sequence type reveals multiple loci of different ancestral lineage. PLoS One, 6, p. e27121

Blattner. 1997. The complete genome sequence of Escherichia coli K-12. Science. 1997 Sep 5;277(5331):1453-62.

Blaser MJ, Berkowitz ID, LaForce FM, Cravens J, Reller LB, Wang WL. 1979. *Campylobacter* enteritis: clinical and epidemiologic features. Ann Intern Med. 1979 Aug;91(2):179-85.

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

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPR) have spacers of extrachromosomal origin. Microbiology 151:2551–61

Butzler JP. 1973. Related vibrio in stools. J. Pediatr.82:493-495.

Butzler JP. 2004. *Campylobacter*, from obscurity to celebrity. Clin Microbiol. Infect. 10:868-876.

Bronowski C, James CE, Winstanley C. 2014. Role of environmental survival in transmission of *Campylobacter jejuni*.FEMS Microbiol Lett. 2014 Jul;356(1):8-19.

Campbell LK, Havens JM, Scott MA, Lamps LW. 2006. Molecular detection of *Campylobacter jejuni* in archival cases of acute appendicitis. Mod Pathol. 2006 Aug;19(8):1042-6.

Carrillo CD, Kruczkiewicz P, Mutschall S, Tudor A, Clark C, Taboada EN. 2012. A framework for assessing the concordance of molecular typing methods and the true strain phylogeny of *Campylobacter jejuni* and *C. coli* using draft genome sequence data.Front Cell Infect Microbiol. 2012 May 1;2:57.

Cawthraw SA, Wassenaar TM, Ayling R, Newell DG. 1996. Increased colonization potential of *Campylobacter jejuni* strain 81116 after passage through chickens and its implication on the rate of transmission within flocks. Epidemiol Infect. 1996 Aug;117(1):213-5.

Cawthraw S, Ayling R, Nuijten P, Wassenaar T, Newell DG. 1994. Isotype, specificity, and kinetics of systemic and mucosal antibodies to *Campylobacter jejuni* antigens, including flagellin, during experimental oral infections of chickens. Avian Dis. 1994 Apr-Jun;38(2):341-9.

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WE, Wetter T, Suhai S. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 2004 Jun;14(6):1147-59. Epub 2004 May 12.

Clark CG, Taboada E, Grant CC, Blakeston C, Pollari F, Marshall B, Rahn K, Mackinnon J, Daignault D, Pillai D, Ng LK. 2012. Comparison of molecular typing methods useful for detecting clusters of *Campylobacter jejuni* and *C. coli* isolates through routine surveillance. J Clin Microbiol. 2012 Mar;50(3):798-809

Clark CG, Berry C, Walker M, Petkau A, Barker DO, Guan C, Reimer A, Taboada EN. 2016. Genomic insights from whole genome sequencing of four clonal outbreak *Campylobacter jejuni* assessed within the global *C. jejuni* population. BMC Genomics. 2016 Dec 3;17(1):990.

Colles FM, Jones K, Harding RM, Maiden MC. 2003. Genetic diversity of *Campylobacter jejuni* isolates from farm animals and the farm environment. Appl Environ Microbiol. 2003 Dec;69(12):7409-13.

Colles FM, Dingle KE, Cody AJ, Maiden MC. 2008. Comparison of *Campylobacter* populations in wild geese with those in starlings and free-range poultry on the same farm. Appl Environ Microbiol. 2008 Jun;74(11):3583-90.

Colles FM and Maiden MC. 2012. *Campylobacter* sequence typing databases: applications and future prospects. Microbiology. 2012 Nov;158(Pt 11):2695-709.

Cody AJ, McCarthy ND, Jansen van Rensburg M, Isinkaye T, Bentley SD, Parkhill J, Dingle KE, Bowler IC, Jolley KA, Maiden MC. 2013. Real-time genomic epidemiological evaluation of human *Campylobacter* isolates by use of whole-genome multilocus sequence typing.J Clin Microbiol. 2013 Aug;51(8):2526-34.

Cody AJ, McCarthy ND, Bray JE, Wimalarathna HM, Colles FM, Jansen van Rensburg MJ, Dingle KE, Waldenström J, Maiden MC. 2015. Wild bird-associated *Campylobacter jejuni* isolates are a consistent source of human disease, in Oxfordshire, United Kingdom. Environ Microbiol Rep. 2015 Oct;7(5):782-8.

Cooper KK, Cooper MA, Zuccolo A, Law B, Joens LA. 2011. Complete genome sequence of *Campylobacter jejuni* strain S3. J. Bacteriol., 193, pp. 1491–1492

Corander J, Waldmann P, Sillanpää MJ. 2003. Bayesian analysis of genetic differentiation between populations. Genetics. 2003 Jan;163(1):367-74.

Corander J, Marttinen P. 2006. Bayesian identification of admixture events using multilocus molecular markers. Mol Ecol. 2006, 15: 2833-2843.

Croucher NJ, Harris SR, Grad YH, Hanage WP. 2013. Bacterial genomes in epidemiology-- present and future. 2013. Philos Trans R Soc Lond B Biol Sci. 2013 Feb 4;368(1614):20120202. Review.

Dark MJ. 2013. Whole-genome sequencing in bacteriology: state of the art. Infect Drug Resist. 2013 Oct 8;6:115-23. Review.

Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004, 14 (7): 1394-1403

Darling AE, Mau B, Perna NT. 2010. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010 Jun 25;5(6):e11147

De Haan CP, Kivistö R, Hakkinen M, Rautelin H, Hänninen ML. 2010a. Decreasing trend of overlapping multilocus sequence types between human and chicken *Campylobacter jejuni* isolates over a decade in Finland. Appl Environ Microbiol. 2010 Aug;76(15):5228-36.

De Haan CP, Kivistö RI, Hakkinen M, Corander J, Hänninen ML. 2010b. Multilocus sequence types of Finnish bovine *Campylobacter jejuni* isolates and their attribution to human infections. BMC Microbiol. 2010 Jul 26;10:200.

De Haan CP, Lampén K, Corander J, Hänninen ML. 2013. Multilocus sequence types of environmental *Campylobacter jejuni* isolates and their similarities to those of human, poultry and bovine *C. jejuni* isolates. Zoonoses Public Health. 2013 Mar;60(2):125-33.

De Haan CP, Kivistö RI, Rautelin H and Hänninen ML. 2014. How Molecular Typing has changed our Understanding on Sources and Transmission Routes of *Campylobacter*iosis in Finland, In: Sheppard SK, editor. *Campylobacter* Ecology and Evolution. 1st edition. Norfolk, UK: Caister academic Press; 2014. pp. 241.

Denis M, Soumet C, Rivoal K, Ermel G, Blivet D, Salvat G, Colin P. 1999. Development of a m-PCR assay for simultaneous identification of *Campylobacter jejuni* and *C. coli*. Lett Appl Microbiol. 1999 Dec;29(6):406-10.

Deveau H, Garneau JE and Moineau S. 2010. CRISPR/Cas system and its role in phage-bacteria interactions. Annu. Rev. Microbiol. 64:475–93.

Didelot X and Falush D. 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics. 2007 Mar;175(3):1251-66.

Didelot X, Lawson D, Darling A, Falush D. 2010. Inference of homologous recombination in bacteria using whole-genome sequences. Genetics. 2010 Dec;186(4):1435-49.

Dingle KE, Colles FM, Wareing DR, Ure R, Fox AJ, Bolton FE, Bootsma HJ, Willems RJ, Urwin R, Maiden MC. Multilocus sequence typing system for *Campylobacter jejuni*. J Clin Microbiol. 2001 Jan;39(1):14-23.

ECDC.http://ecdc.europa.eu/en/healthtopics/campylobacteriosis/basic_facts/Pages/Fact sheet_generalpublic.aspx

EFSA, www.efsa.europa.eu/en/topics/topic/campylobacter

EFSA 2005. Opinion of the Scientific Panel on biological hazards (BIOHAZ) related to *Campylobacter* in animals and foodstuffs. Tabs Panel on Biological Hazards. EFSA Journal 2005; 3(4).

EFSA 2010. Analysis of the baseline survey on the prevalence of *Campylobacter* in broiler batches and of *Campylobacter* and Salmonella on broiler carcasses in the EU, 2008, part A: *Campylobacter* and Salmonella prevalence estimates. EFSA Journal 2010; 8(03):1503.

EFSA 2011. Panel of Biological Hazards (BIOHAZ), 2011. Scientific opinion on *Campylobacter* in broiler meat production: Control opinions and performance objectives and/or targets at different stages of the food chain. EFSA Journal 2011:9(4):2105.

EFSA 2014. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013. EFSA Journal 2015;13(1):3991 [165 pp.].

EFSA 2016. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2015. European Food Safety Authority. EFSA Journal: EFSA Journal 2016;14(12):4634 [231 pp.].

Ellström P, Hansson I, Söderström C, Engvall EO, Rautelin H. 2014. A prospective follow-up study on transmission of *Campylobacter* from poultry to abattoir workers.Foodborne Pathog Dis. 2014 Sep;11(9):684-8.

EVIRA.RuokamyrkytyksetSuomessa.https://www.evira.fi/elaimet/zoonoosikeskus/ruoka myrkytykset/ruokamyrkytysepidemiat-suomessa/

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics, 164, 1567–1587.

Fearnhead P, Biggs PJ, French N. 2014. Learning about recombination in *Campylobacter*. In: Sheppard SK, editor. *Campylobacter* Ecology and Evolution. 1st edition. Norfolk, UK: Caister academic Press; 2014. pp. 9-22.

Feodoroff B, Lauhio A, Ellström P, Rautelin H. 2011 A nationwide study of *Campylobacter jejuni* and *Campylobacter* coli bacteremia in Finland over a 10-year period, 1998-2007, with special reference to clinical characteristics and antimicrobial susceptibility. Clin Infect Dis. 2011 Oct;53(8):e99-e106.

Feodoroff B, de Haan CP, Ellström P, Sarna S, Hänninen ML, Rautelin H. 2013. Clonal distribution and virulence of *Campylobacter* jejuni isolates in blood. Emerg Infect Dis. 2013 Oct;19(10):1653-5.

Fleischmann R, Adams M, White O, Clayton R, Kirkness E, Kerlavage A, Bult C, Tomb J, Dougherty B, Merrick J et al. 1995. "Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd". Science 1995 Jul 28;269(5223):496-512.

Flint A, Butcher J, Stintzi A. 2014. *Campylobacter* strategies for colonizing the host gastrointestinal tract. In: Sheppard SK, editor. *Campylobacter* Ecology and Evolution. 1st edition. Norfolk, UK: Caister academic Press; 2014. pp. 75-88.

Fouts DE, Mongodin EF, Mandrell RE, Miller WG, Rasko DA, Ravel J, Brinkac LM, DeBoy RT, Parker CT, Daugherty SC, Dodson RJ, Durkin AS, Madupu R, Sullivan SA, Shetty JU, Ayodeji MA, Shvartsbeyn A, Schatz MC, Badger JH, Fraser CM, Nelson KE. 2005. Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. PLoS Biol. 2005 Jan;3(1):e15. Epub 2005 Jan 4.

French N, Barrigas M, Brown P, Ribiero P, Williams N, Leatherbarrow H, Birtles R, Bolton E, Fearnhead P, Fox A. 2005. Spatial epidemiology and natural population structure of *Campylobacter jejuni* colonizing a farmland ecosystem. Environ Microbiol. 2005 Aug;7(8):1116-26.

Gaasbeek EJ, Wagenaar JA, Guilhabert MR, Wösten MM, van Putten JP, van der Graaf-van Bloois L, Parker CT, van der Wal FJ. 2009. A DNase encoded by integrated element CJIE1 inhibits natural transformation of *Campylobacter jejuni*. J Bacteriol. 2009 Apr;191(7):2296-306.

Gaasbeek EJ, Wagenaar JA, Guilhabert MR, van Putten JP, Parker CT, van der Wal FJ. 2010. Nucleases encoded by the integrated elements CJIE2 and CJIE4 inhibit natural transformation of *Campylobacter jejuni*.J Bacteriol. 2010 Feb;192(4):936-41.

Gebhart CJ, Edmonds P, Ward GE, Kurtz HJ, Brenner DJ. 1985. *Campylobacter hyointestinalis*" sp. nov.: a new species of *Campylobacter* found in the intestines of pigs and other animals. J Clin Microbiol. 1985 May;21(5):715-20.

Ghaffar NM, Connerton PL, Connerton IF. Filamentation of *Campylobacter* in broth cultures. Front Microbiol. 2015 Jun 30;6:657.

González, M., Skandamis, P.N. and Hänninen, M.L. 2009. A modified Weibull model for describing the survival of *Campylobacter jejuni* in minced chicken meat. Int J Food Microbiol 136, 52–58.

González M., Hänninen M-L. 2012. Effect of temperature and antimicrobial resistance on survival of *Campylobacter jejuni* in well water: application of the Weibull model. J Appl. Microbiol. Aug;113(2):284-93.

Grant CC, Konkel ME, Cieplak W Jr, Tompkins LS. 1993. Role of flagella in adherence, internalization, and translocation of *Campylobacter jejuni* in nonpolarized and polarized epithelial cell cultures. Infect Immun. 1993 May;61(5):1764-71.

Griekspoor P, Engvall EO, Olsen B, Waldenström J. 2010. Multilocus sequence typing of *Campylobacter jejuni* from broilers. Vet Microbiol. 2010 Jan 6;140(1-2):180-5.

Griekspoor P, Colles FM, McCarthy ND, Hansbro PM, Ashhurst-Smith C, Olsen B, Hasselquist D, Maiden MC, Waldenström J. 2013. Marked host specificity and lack of phylogeographic population structure of *Campylobacter jejuni* in wild birds. Mol Ecol. 2013 Mar;22(5):1463-72

Griekspoor P, Hansbro PM, Waldenström J, Olsen B. 2015. *Campylobacter jejuni* sequence types show remarkable spatial and temporal stability in Blackbirds. Infect Ecol Epidemiol. 2015 Dec 2;5:28383.

Gripp E, Hlahla D, Didelot X, Kops F, Maurischat S, Tedin K, Alter T, Ellerbroek L, Schreiber K, Schomburg D, Janssen T, Bartholomäus P, Hofreuter D, Woltemate S, Uhr M, Brenneke B, Grüning P, Gerlach G, Wieler L, Suerbaum S, Josenhans C. 2011. Closely related *Campylobacter jejuni* strains from different sources reveal a generalist rather than a specialist lifestyle. BMC Genomics. 2011 Nov 28;12:584.

Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W52-7. Epub 2007 May 30.

Guyard-Nicodème M, Rivoal K, Houard E, Rose V, Quesne S, Mourand G, Rouxel S, Kempf I, Guillier L, Gauchard F, Chemaly M. 2015. Prevalence and characterization of *Campylobacter jejuni* from chicken meat sold in French retail outlets. Int J Food Microbiol. 2015 Jun 16;203:8-14.

Habib I, Louwen R, Uyttendaele M, Houf K, Vandenberg O, Nieuwenhuis EE, Miller WG, van Belkum A, De Zutter L. 2009 Correlation between genotypic diversity, lipooligosaccharide gene locus class variation, and caco-2 cell invasion potential of *Campylobacter jejuni* isolates from chicken meat and humans: contribution to virulotyping. Appl Environ Microbiol. 2009 Jul;75(13):4277-88.

Hakkinen M, Heiska H, Hänninen ML. 2007. Prevalence of *Campylobacter* spp. in cattle in Finland and antimicrobial susceptibilities of bovine *Campylobacter jejuni* strains. Appl Environ Microbiol. 2007 May;73(10):3232-8. Epub 2007 Mar 16.

Hakkinen M, Nakari UM, Siitonen A. 2009. Chickens and cattle as sources of sporadic domestically acquired *Campylobacter jejuni* infections in Finland. Appl Environ Microbiol. 2009 Aug;75(16):5244-9.

Hannu T, Mattila L, Rautelin H, Pelkonen P, Lahdenne P, Siitonen A, Leirisalo-Repo M. 2002. Campylobacter-triggered reactive arthritis: a population-based study. Rheumatology (Oxford). 2002 Mar;41(3):312-8.

Hatta Y, Omatsu T, Tsuchiaka S, Katayama Y, Taniguchi S, Masangkay JS, Puentespina R Jr, Eres E, Cosico E, Une Y, Yoshikawa Y, Maeda K, Kyuwa S, Mizutani T. 2016. Detection of *Campylobacter jejuni* in rectal swab samples from Rousettus amplexicaudatus in the Philippines. J Vet Med Sci. 2016 Sep 1;78(8):1347-50.

Headrick ML, Korangy S, Bean NH, Angulo FJ, Altekruse SF, Potter ME, Klontz KC. 1998. The epidemiology of raw milk-associated foodborne disease outbreaks reported in the United States, 1973 through 1992. Am J Public Health. 1998 Aug;88(8):1219-21.

Hepworth PJ, Ashelford KE, Hinds J, Gould KA, Witney AA, Williams NJ, Leatherbarrow H, French NP, Birtles RJ, Mendonca C, Dorrell N, Wren BW, Wigley P, Hall N, Winstanley C. 2011. Genomic variations define divergence of water/wildlife-associated *Campylobacter jejuni* niche specialists from common clonal complexes. Environ Microbiol. 2011 Jun;13(6):1549-60.

Hermans D, Pasmans F, Messens W, Martel A, Van Immerseel F, Rasschaert G, Heyndrickx M, Van Deun K, Haesebrouck F. 2012. Poultry as a host for the zoonotic pathogen *Campylobacter jejuni*. Vector Borne Zoonotic Dis. 2012 Feb;12(2):89-98.

Hofreuter D, Tsai J, Watson RO, Novik V, Altman B, Benitez M, Clark C, Perbost C, Jarvie T, Du L, Galán JE. 2006. Unique features of a highly pathogenic *Campylobacter jejuni* strain. Infect Immun. 2006 Aug;74(8):4694-707.

Hokajärvi AM, Pitkänen T, Siljanen HM, Nakari UM, Torvinen E, Siitonen A, Miettinen IT. 2013. Occurrence of thermotolerant *Campylobacter* spp. and adenoviruses in Finnish bathing waters and purified sewage effluents.J Water Health. 2013 Mar;11(1):120-34.

Humphrey TJ and Beckett P. 1987. *Campylobacter jejuni* in dairy cows and raw milk. Epidemiol Infect. 1987 Jun;98(3):263-9.

Huson DH and Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006 Feb;23(2):254-67. Epub 2005 Oct 12.

Hänninen ML, Pajarre S, Klossner ML, Rautelin H. 1998. Typing of human *Campylobacter jejuni* isolates in Finland by pulsed-field gel electrophoresis. J Clin Microbiol. 1998 Jun;36(6):1787-9.

Hänninen ML, Haajanen H, Pummi T, Wermundsen K, Katila ML, Sarkkinen H, Miettinen I, Rautelin H. 2003. Detection and typing of *Campylobacter jejuni* and Campylobacter coli and analysis of indicator organisms in three waterborne outbreaks in Finland. Appl Environ Microbiol. 2003 Mar;69(3):1391-6.

Hörman A, Rimhanen-Finne R, Maunula L, von Bonsdorff CH, Torvela N, Heikinheimo A, Hänninen ML. 2004. *Campylobacter* spp., *Giardia* spp., *Cryptosporidium* spp.,

noroviruses, and indicator organisms in surface water in southwestern Finland, 2000–2001. Appl. Environ. Microbiol. 70, 87–95

Illumina.2010.https://www.illumina.com/Documents/products/technotes/technote_deno vo_assembly_ecoli.pdf

Jerome JP, Bell JA, Plovanich-Jones AE, Barrick JE, Brown CT, Mansfield LS. 2011. Standing genetic variation in contingency loci drives the rapid adaptation of *Campylobacter jejuni* to a novel host. PLoS One. 2011 Jan 24;6(1):e16399.

Jerome JP and Mansfield LS. 2014. Within-host evolution of Campylobacter jejuni. In: Sheppard SK, editor. *Campylobacter* Ecology and Evolution. 1st edition. Norfolk, UK: Caister academic Press; 2014. p. 23-32.

Jolley KA and Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11: 595.

Jore S, Viljugrein H, Brun E, Heier BT, Borck B, Ethelberg S, Hakkinen M, Kuusi M, Reiersen J, Hansson I, Engvall EO, Lofdahl M, Wagenaar JA, van Pelt W, Hofshagen M. 2010. Trends in *Campylobacter* incidence in broilers and humans in six European countries, 1997–2007. Prev. Vet. Med. 93, 33–41

Juntunen P, Heiska H, Olkkola S, Myllyniemi AL, Hänninen ML. 2010. Antimicrobial resistance in *Campylobacter coli* selected by tylosin treatment at a pig farm. Vet Microbiol. 2010 Nov 20;146(1-2):90-7.

Kalmokoff M, Lanthier P, Tremblay TL, Foss M, Lau PC, Sanders G, Austin J, Kelly J, Szymanski CM. 2006. Proteomic analysis of *Campylobacter jejuni* 11168 biofilms reveals a role for the motility complex in biofilm formation.J Bacteriol. 2006 Jun;188(12):4312-20.

Karlyshev AV, Linton D, Gregson NA, Lastovica AJ, Wren BW. 2000. Genetic and biochemical evidence of a *Campylobacter jejuni* capsular polysaccharide that accounts for Penner serotype specificity. Mol Microbiol. 2000 Feb;35(3):529-41.

Karlyshev AV, Ketley JM, Wren BW. 2005. The *Campylobacter jejuni* glycome. FEMS Microbiol Rev. 2005 Apr;29(2):377-90. Review.

Keithlin J, Sargeant J, Thomas MK, Fazil A. 2014. Systematic review and meta-analysis of the proportion of *Campylobacter* cases that develop chronic sequelae. BMC Public Health. 2014 Nov 22;14:1203. Review.

Keller JI, Shriver WG, Waldenström J, Griekspoor P, Olsen B. 2011. Prevalence of *Campylobacter* in wild birds of the mid-Atlantic region, USA.J Wildl Dis. 2011 Jul;47(3):750-4.

Keo T, Collins J, Kunwar P, Blaser MJ, Iovine NM. 2011. *Campylobacter* capsule and lipooligosaccharide confer resistance to serum and cationic antimicrobials. Virulence. 2011 Jan-Feb;2(1):30-40. Epub 2011 Jan 1.

Kittl S, Kuhnert P, Hächler H, Korczak BM. 2011. Comparison of genotypes and antibiotic resistance of *Campylobacter jejuni* isolated from humans and slaughtered chickens in Switzerland.J Appl Microbiol. 2011 Feb;110(2):513-520.

Kirst M. 1985. The historical background to *Campylobacter* infection: new prospect. In: B. Rowe, editor. Campylobacter III. Public Health Laboratory Service, United Kingdom; 1985:23–27

Korhonen LK, Martikainen PJ. 1991. Survival of *Escherichia coli* and *Campylobacter jejuni* in untreated and filtered lake water. J Appl Bacteriol. 1991 Oct;71(4):379-82.

Korczak BM, Zurfluh M, Emler S, Kuhn-Oertli J, Kuhnert P. 2009. Multiplex strategy for multilocus sequence typing, *fla* typing, and genetic determination of antimicrobial resistance of *Campylobacter jejuni* and *Campylobacter coli* isolates collected in Switzerland. J Clin Microbiol. 2009 Jul;47(7):1996-2007.

Korsak D, Maćkiw E, Rożynek E, Żyłowska M. 2015. Prevalence of *Campylobacter* spp. in Retail Chicken, Turkey, Pork, and Beef Meat in Poland between 2009 and 2013. J Food Prot. 2015 May;78(5):1024-8.

Kwan PS, Barrigas M, Bolton FJ, French NP, Gowland P, Kemp R, Leatherbarrow H, Upton M, Fox AJ. 2008. Molecular epidemiology of *Campylobacter jejuni* populations in dairy cattle, wildlife, and the environment in a farmland area.Appl Environ Microbiol. 2008 Aug;74(16):5130-8.

Kwan PS, Xavier C, Santovenia M, Pruckler J, Stroika S, Joyce K, Gardner T, Fields PI, McLaughlin J, Tauxe RV, Fitzgerald C. 2014. Multilocus sequence typing confirms wild birds as the source of a *Campylobacter* outbreak associated with the consumption of raw peas. Appl Environ Microbiol. 2014 Aug;80(15):4540-6.

Kärenlampi R, Rautelin H, Hakkinen M, Hänninen ML. 2003. Temporal and geographical distribution and overlap of Penner heat-stable serotypes and pulsed-field gel electrophoresis genotypes of *Campylobacter jejuni* isolates collected from humans and chickens in Finland during a seasonal peak. J Clin Microbiol. 2003 Oct;41(10):4870-2.

Kärenlampi, R and Hänninen, M.L. 2004. Survival of *Campylobacter jejuni* on various fresh produce. Int J Food Microbiol. 2004 Dec 15;97(2):187-95.

Kärenlampi R, Rautelin H, Schönberg-Norio D, Paulin L, Hänninen ML. 2007a. Longitudinal study of Finnish *Campylobacter jejuni* and *C. coli* isolates from humans, using

multilocus sequence typing, including comparison with epidemiological data and isolates from poultry and cattle. Appl Environ Microbiol. 2007 Jan;73(1):148-55.

Kärenlampi R, Rautelin H, Hänninen ML. 2007b. Evaluation of genetic markers and molecular typing methods for prediction of sources of *Campylobacter jejuni* and *C. coli* infections. Appl Environ Microbiol. 2007 Mar;73(5):1683-5.

Laine J, Huovinen E, Virtanen MJ, Snellman M, Lumio J, Ruutu P, Kujansuu E, Vuento R, Pitkänen T, Miettinen I, Herrala J, Lepistö O, Antonen J, Helenius J, Hänninen ML, Maunula L, Mustonen J, Kuusi M. 2011. An extensive gastroenteritis outbreak after drinking-water contamination by sewage effluent, Finland. Epidemiol Infect. 2011 Jul;139(7):1105-13.

Lastovica AJ and Allos BM. 2008. Clinical significance of Campylobacter and related species other than *Campylobacter jejuni* and *Campylobacter coli*. In: Nachamkin I, Szymanski C, Blaser M, editors. *Campylobacter*. 3rd edition. Washington DC: ASM Press; 2008. p. 123-149.

Lawson. 1981. Some features of *Campylobacter sputorum* subsp. mucosalis subsp. nov., nom. rev. and their taxonomic significance. Int. J. System. Bacteriolog. 31:385-391.

Lee RB, Hassane DC, Cottle DL, Pickett CL. 2003. Interactions of *Campylobacter jejuni* cytolethal distending toxin subunits *Cdt*A and *Cdt*C with HeLa cells. Infect Immun. 2003 Sep;71(9):4883-90.

Lévesque S, Frost E, Arbeit RD, Michaud S. 2008. Multilocus sequence typing of *Campylobacter jejuni* isolates from humans, chickens, raw milk, and environmental water in Quebec, Canada. J Clin Microbiol. 2008 Oct;46(10):3404-11.

Lévesque S, Fournier E, Carrier N, Frost E, Arbeit RD, Michaud S. 2013. Campylobacteriosis in urban versus rural areas: a case-case study integrated with molecular typing to validate risk factors and to attribute sources of infection. PLoS One. 2013 Dec 26;8(12):e83731.

Llarena AK, Skarp-de Haan CP, Rossi M, Hänninen ML. 2015a. Characterization of the *Campylobacter jejuni* population in the barnacle geese reservoir. Zoonoses Public Health. 2015 May;62(3):209-21.

Llarena AK, Huneau A, Hakkinen M, Hänninen ML. 2015b. Predominant *Campylobacter jejuni* sequence types persist in Finnish chicken production. PLoS One. 2015 Feb 20;10(2):e0116585.

Llarena AK, Zhang J, Vehkala M, Välimäki N, Hakkinen M, Hänninen ML, Roasto M, Mäesaar M, Taboada E, Barker D, Garfolo G, Cammà C, Di Giannatale E, Corander J, Rossi

M. 2016. Monomorphic genotypes within a generalist lineage of *Campylobacter jejuni* show signs of global dispersion. Microb Genom. 2016 Oct 21;2(10):e000088

Llarena AK, Taboada E, Rossi M. 2017. Whole-genome sequencing in the epidemiology of *Campylobacter jejuni* infections. J Clin Microbiol. 2017 Mar 1. pii: JCM.00017-17.

Louwen R, Heikema A, van Belkum A, Ott A, Gilbert M, Ang W, Endtz HP, Bergman MP, Nieuwenhuis EE. 2008. The sialylated lipooligosaccharide outer core in *Campylobacter jejuni* is an important determinant for epithelial cell invasion. Infect Immun. 2008 Oct;76(10):4431-8.

Louwen R, Horst-Kreft D, de Boer AG, van der Graaf L, de Knegt G, Hamersma M, Heikema AP, Timms AR, Jacobs BC, Wagenaar JA, Endtz HP, van der Oost J, Wells JM, Nieuwenhuis EE, van Vliet AH, Willemsen PT, van Baarlen P, van Belkum A. 2013. A novel link between *Campylobacter jejuni* bacteriophage defence, virulence and Guillain-Barré syndrome.Eur J Clin Microbiol Infect Dis. 2013 Feb;32(2):207-26.

Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, Robinson ER, Pallen MJ. 2012. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat Rev Microbiol. 2012 Sep;10(9):599-606.

Loman NJ, Pallen MJ. 2015. Twenty years of bacterial genome sequencing. Nat Rev Microbiol. 2015 Dec;13(12):787-94.

Maa- ja Metsätalousministeriö. 2007. Maa- ja Metsätalousministeriön asetus broilereiden kampylobakteerivalvonnasta.10/EEO/2007.Available:http://mmm.fi/documents/1410837 /1818689/10_eeo_2007.pdf/02941e7e-eb46-4f6f-8555-ad01609cfd02, accessed 29.3.2017.

MacDonald E, White R, Mexia R, Bruun T, Kapperud G, Lange H, Nygård K, Vold L. 2015. Risk Factors for Sporadic Domestically Acquired *Campylobacter* Infections in Norway 2010-2011: A National Prospective Case-Control Study. PLoS One. 2015 Oct 2;10(10):e0139636.

Mackay B. 2002. Walkerton, 2 years later: "Memory fades very quickly". CMAJ. 2002 May 14;166(10):1326.

Malakauskas M, Jorgensen K, Nielsen EM, Ojeniyi B, Olsen JE. 2006. Isolation of *Campylobacter* spp. from a pig slaughterhouse and analysis of cross-contamination.Int J Food Microbiol. 2006 May 1;108(3):295-300.

Marshall. 1984. Original isolation of *Campylobacter pyloridis* from human gastric mucosa. Microbios Letter 25:83-88.

McCarthy N, Giesecke J. 2001. Incidence of Guillain-Barré syndrome following infection with *Campylobacter jejuni*. Am J Epidemiol. 2001 Mar 15;153(6):610-4.

McCarthy ND, Gillespie IA, Lawson AJ, Richardson J, Neal KR, Hawtin PR, Maiden MC, O'Brien SJ. 2012. Molecular epidemiology of human *Campylobacter jejuni* shows association between seasonal and international patterns of disease. Epidemiol Infect. 2012 Dec;140(12):2247-55.

Meinersmann RJ, Berrang ME, Little E. 2013. *Campylobacter* spp. recovered from the Upper Oconee River Watershed, Georgia in a 4-year study. Microb Ecol. 2013 Jan;65(1):22-7.

Michaud S, Menard S, Gaudreau C, Arbeit RD. 2001. Comparison of *Sma*I-defined genotypes of *Campylobacter jejuni* examined by *Kpn*I: a population-based study. J Med Microbiol. 2001 Dec;50(12):1075-81.

Miettinen IT, Zacheus O, von Bonsdorff CH, Vartiainen T. 2001. Waterborne epidemics in Finland in 1998–1999. Water Sci Technol. 2001;43(12):67-71.

Miller WG, On SL, Wang G, Fontanoz S, Lastovica AJ, Mandrell RE. 2005. Extended multilocus sequence typing system for *Campylobacter coli*, *C. lari*, *C. upsaliensis*, and *C. helveticus*.J Clin Microbiol. 2005 May;43(5):2315-29.

Morey F. 1996. Five years of *Campylobacter* bacteraemia in Central Australia. In Newell DG, Ketley JM, Feldman RA, editors. Campylobacters, Helicobacter, and related organisms. New York: Plenum Press. p. 491-494.

Mossong J, Mughini-Gras L, Penny C, Devaux A, Olinger C, Losch S, Cauchie HM, van Pelt W, Ragimbeau C. 2016. Human Campylobacteriosis in Luxembourg, 2010-2013: A Case-Control Study Combined with Multilocus Sequence Typing for Source Attribution and Risk Factor Analysis.Sci Rep. 2016 Feb 10;6:20939.

Mughini-Gras L, Smid JH, Wagenaar JA, de Boer AG, Havelaar AH, Friesema IH, French NP, Busani L, van Pelt W. 2012. Risk factors for campylobacteriosis of chicken, ruminant, and environmental origin: a combined case-control and source attribution analysis. PLoS One. 2012;7(8):e42599. Epub 2012 Aug 3.

Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. 2010. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2197-223. Epub 2012/12/19.

Müllner P, Spencer SE, Wilson DJ, Jones G, Noble AD, Midwinter AC, Collins-Emerson JM, Carter P, Hathaway S, French NP. 2009. Assigning the source of human campylobacteriosis

in New Zealand: a comparative genetic and epidemiological approach. Infect Genet Evol. 2009 Dec;9(6):1311-9.

Müllner P, Collins-Emerson JM, Midwinter AC, Carter P, Spencer SE, van der Logt P, Hathaway S, French NP. 2010. Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry. Appl Environ Microbiol. 2010 Apr;76(7):2145-54.

Newell DG, McBride H, Saunders F, Dehele Y, Pearson AD. The virulence of clinical and environmental isolates of *Campylobacter jejuni*.J Hyg (Lond). 1985 Feb;94(1):45-54.

Norrby R, McCloskey RV, Zackrisson G, Falsen E. Meningitis caused by *Campylobacter fetus* ssp jejuni. Br Med J. 1980 May 10;280(6224):1164. No abstract available.

Nylen G, Dunstan F, Palmer SR, Andersson Y, Bager F, Cowden J, Feierl G, Galloway Y, Kapperud G, Megraud F, Molbak K, Petersen LR, Ruutu P. 2002. The seasonal distribution of *Campylobacter* infection in nine European countries and New Zealand. Epidemiol Infect. 2002 Jun;128(3):383-90.

Olkkola S, Kovanen S, Roine J, Hänninen ML, Hielm-Björkman A, Kivistö R. 2015. Population Genetics and Antimicrobial Susceptibility of Canine *Campylobacter* Isolates Collected before and after a Raw Feeding Experiment.PLoS One. 2015 Jul 14;10(7):e0132660.

On SL, Nielsen EM, Engberg J, Madsen M. Validity of *Sma*I-defined genotypes of *Campylobacter jejuni* examined by *Sal*I, *Kpn*I, and *Bam*HI polymorphisms: evidence of identical clones infecting humans, poultry, and cattle. Epidemiol Infect. 1998 Jun;120(3):231-7.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014 Jan;42(Database issue):D206-14.

Ozbey G, Tasdemi B. 2014. Seasonality and antibiotic resistance of *Campylobacter* in Turkish chicken meat.Vet Ital. 2014 Dec 29;50(4):277-83.

Park SF. The physiology of *Campylobacter* species and its relevance to their role as foodborne pathogens. 2002. Int J Food Microbiol. 2002 Apr 5;74(3):177-88. Review.

Parker CT, Miller WG, Horn ST, Lastovica AJ. 2007. Common genomic features of *Campylobacter jejuni* subsp. *doylei* strains distinguish them from *C. jejuni* subsp. *jejuni*. BMC Microbiol. 2007 May 29;7:50.

Parker CT, Gilbert M, Yuki N, Endtz HP, Mandrell RE. 2008. Characterization of lipooligosaccharide-biosynthetic loci of *Campylobacter jejuni* reveals new lipooligosaccharide classes: evidence of mosaic organizations. J Bacteriol. 2008 Aug;190(16):5681-9.

Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D, Chillingworth T, Davies RM, Feltwell T, Holroyd S, Jagels K, Karlyshev AV, Moule S, Pallen MJ, Penn CW, Quail MA, Rajandream MA, Rutherford KM, van Vliet AH, Whitehead S, Barrell BG. 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature. 2000 Feb 10;403(6770):665-8.

Parsons BN, Porter CJ, Ryvar R, Stavisky J, Williams NJ, Pinchbeck GL, Birtles RJ, Christley RM, German AJ, Radford AD, Hart CA, Gaskell RM, Dawson S. 2010. Prevalence of *Campylobacter* spp. in a cross-sectional study of dogs attending veterinary practices in the UK and risk indicators associated with shedding.Vet J. 2010 Apr;184(1):66-70.

Paszkiewicz K, Studholme DJ. 2010. *De novo* assembly of short sequence reads. Brief Bioinform. 2010 Sep;11(5):457-72.

Pearson BM, Louwen R, van Baarlen P, van Vliet AH. 2015. Differential Distribution of Type II CRISPR-Cas Systems in Agricultural and Nonagricultural *Campylobacter coli* and *Campylobacter jejuni* Isolates Correlates with Lack of Shared Environments. Genome Biol Evol. 2015 Sep 2;7(9):2663-79.

Pebody RG, Ryan MJ, Wall PG. 1997. Outbreaks of *Campylobacter* infection: rare events for a common pathogen. Commun Dis Rep CDR Rev. 1997 Mar 7;7(3):R33-7.

Penner JL, Hennessy JN. Passive hemagglutination technique for serotyping *Campylobacter fetus* subsp. *jejuni* on the basis of soluble heat-stable antigens. J Clin Microbiol. 1980 Dec;12(6):732-7.

Pintar KD, Thomas KM, Christidis T, Otten A, Nesbitt A, Marshall B, Pollari F, Hurst M, Ravel A. 2016. A Comparative Exposure Assessment of *Campylobacter* in Ontario, Canada. Risk Anal. 2016 Sep 18.

Pitkänen T, Miettinen IT, Nakari UM, Takkinen J, Nieminen K, Siitonen A, Kuusi M, Holopainen A, Hänninen ML. 2008. Faecal contamination of a municipal drinking water distribution system in association with *Campylobacter jejuni* infections. J Water Health. 2008 Sep;6(3):365-76.

Pitkänen T. 2013. Review of *Campylobacter* spp. in drinking and environmental waters. J Microbiol Methods. 2013 Oct;95(1):39-47. Review.

Poly F and Guerry P. 2008. Pathogenesis of *Campylobacter*. Curr Opin Gastroenterol. 2008 Jan;24(1):27-31. Review.

Price EP, Smith H, Huygens F, Giffard PM. 2007. High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of *Campylobacter jejuni*. Appl. Environ. Microbiol. 73:3431–36

PubMLST Database for *Campylobacter*. http://pubmlst.org/campylobacter/ Last accessed 1.4.2017

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W116-20.

Ramonaite S, Kudirkiene E, Tamuleviciene E, Leviniene G, Malakauskas A, Gölz G, Alter T, Malakauskas M. 2014. Prevalence and genotypes of *Campylobacter jejuni* from urban environmental sources in comparison with clinical isolates from children. J Med Microbiol. 2014 Sep;63(Pt 9):1205-13.

Raphael BH, Joens LA. 2003. *Feo*B is not required for ferrous iron uptake in *Campylobacter jejuni*. Can J Microbiol. 2003 Nov;49(11):727-31.

Rath D, Amlinger L, Rath A, Lundgren M. 2015. The CRISPR-Cas immune system: biology, mechanisms and applications.Biochimie. 2015 Oct;117:119-28. Review.

Rautelin H, Hänninen ML. 2000. Campylobacters: the most common bacterial enteropathogens in the Nordic countries. Ann Med. 2000 Oct;32(7):440-5. Review.

Ravel A, Pintar K, Nesbitt A, Pollari F. 2016. Non food-related risk factors of campylobacteriosis in Canada: a matched case-control study. BMC Public Health. 2016 Sep 27;16(1):1016.

Revez J, Schott T, Llarena AK, Rossi M, Hänninen ML. 2013. Genetic heterogeneity of *Campylobacter jejuni* NCTC 11168 upon human infection. Infect Genet Evol. 2013 Jun;16:305-9.

Revez J, Zhang J, Schott T, Kivistö R, Rossi M, Hänninen ML. 2014a. Genomic variation between *Campylobacter jejuni* isolates associated with milk-borne-disease outbreaks. J Clin Microbiol. 2014 Aug;52(8):2782-6

Revez J, Llarena AK, Schott T, Kuusi M, Hakkinen M, Kivistö R, Hänninen ML, Rossi M. 2014b. Genome analysis of *Campylobacter jejuni* strains isolated from a waterborne outbreak. BMC Genomics. 2014 Sep 8;15:768.

Robinson DA. 1981. Infective dose of *Campylobacter jejuni* in milk. Br Med J (Clin Res Ed). 1981 May 16;282(6276):1584.

Rossi M, Hänninen ML, Revez J, Hannula M, Zanoni RG. 2008. Occurrence and species level diagnostics of *Campylobacter* spp., enteric *Helicobacter* spp. and *Anaerobiospirillum* spp. in healthy and diarrheic dogs and cats. Vet Microbiol. 2008 Jun 22;129(3-4):304-14.

Rollins, D. M., Colwell, R. R.1986. Viable but non-culturable stage of *Campylobacter jejuni* and its role in the survival in the natural aquatic environment. Appl. Environ. Microbiol. 52, 531–538.

Rotariu O, Smith-Palmer A, Cowden J, Bessell PR, Innocent GT, Reid SW, Matthews L, Dallas J, Ogden ID, Forbes KJ, Strachan NJ. 2010. Putative household outbreaks of campylobacteriosis typically comprise single MLST genotypes.Epidemiol Infect. 2010 Dec;138(12):1744-7.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. Bioinformatics. 2000 Oct;16(10):944-5.

Sails AD, Swaminathan B, Fields PI. 2003. Clonal complexes of *Campylobacter jejuni* identified by multilocus sequence typing correlate with strain associations identified by multilocus enzyme electrophoresis. J Clin Microbiol. 2003 Sep;41(9):4058-67.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977 Dec;74(12):5463-7.

Sasaki Y, Murakami M, Haruna M, Maruyama N, Mori T, Ito K, Yamada Y. 2012. Prevalence and characterization of foodborne pathogens in dairy cattle in the eastern part of Japan. J Vet Med Sci. 2013 May 2;75(4):543-6. Epub 2012 Dec 3.

Schallenberg M, Bremer PJ, Henkel S, Launhardt A, Burns CW. 2005. Survival of *Campylobacter jejuni* in water: effect of grazing by the freshwater crustacean *Daphnia carinata* (Cladocera). Appl Environ Microbiol. 2005 Sep;71(9):5085-8.

Schönberg-Norio D, Takkinen J, Hänninen ML, Katila ML, Kaukoranta SS, Mattila L, Rautelin H. 2004. Swimming and *Campylobacter* infections. Emerg Infect Dis. 2004 Aug;10(8):1474-7.

Schouls LM, Reulen S, Duim B, Wagenaar JA, Willems RJ, Dingle KE, Colles FM, Van Embden JD. 2003. Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and recombination. J Clin Microbiol. 2003 Jan;41(1):15-26.

Sebald M. and Veron M. 1963. Base DNA content and Classification of Vibrios. Ann Inst Pasteur (Paris) 1963 Nov;105:897-910.

Sheppard SK, Dallas JF, MacRae M, McCarthy ND, Sproston EL, Gormley FJ, Strachan NJ, Ogden ID, Maiden MC, Forbes KJ. 2009a. *Campylobacter* genotypes from food animals, environmental sources and clinical disease in Scotland 2005/6.Int J Food Microbiol. 2009 Aug 31;134(1-2):96-103.

Sheppard SK, Dallas JF, Strachan NJ, MacRae M, McCarthy ND, Wilson DJ, Gormley FJ, Falush D, Ogden ID, Maiden MC, Forbes KJ. 2009b. *Campylobacter* genotyping to determine the source of human infection. Clin Infect Dis. 2009 Apr 15;48(8):1072-8.

Sheppard SK, Jolley KA, Maiden MC. 2012. A Gene-By-Gene Approach to Bacterial Population Genomics: Whole Genome MLST of *Campylobacter*. Genes (Basel). 2012 Apr 12;3(2):261-77.

Sheppard SK, Maiden MC. 2015. The evolution of *Campylobacter jejuni* and *Campylobacter coli*. Cold Spring Harb Perspect Biol. 2015 Jun 22;7(8):a018119. Review.

Simpson, E.H. 1949. Measurement of diversity. Nature 163, 688.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. 2009 Jun;19(6):1117-23.

Sippy R, Sandoval-Green CM, Sahin O, Plummer P, Fairbanks WS, Zhang Q, Blanchong JA. 2012. Occurrence and molecular analysis of *Campylobacter* in wildlife on livestock farms.Vet Microbiol. 2012 Jun 15;157(3-4):369-75.

Skarp CP, Akinrinade O, Nilsson AJ, Ellström P, Myllykangas S, Rautelin H. 2015. Comparative genomics and genome biology of invasive *Campylobacter jejuni*. Sci Rep. 2015 Nov 25;5:17300.

Skarp CP, Hänninen ML, Rautelin HI. 2016. Campylobacteriosis: the role of poultry meat. Clin Microbiol Infect. 2016 Feb;22(2):103-9. Review.

Skirrow MB. 1977. *Campylobacter* enteritis: a "new" disease. British medical journal 2, 9-11

Skirrow MB. 1991. Epidemiology of *Campylobacter* enteritis. Int J Food Microbiol. Jan;12(1):9-16

Skirrow MB, Jones DM, Sutcliffe E, Benjamin J. 1993. *Campylobacter* bacteraemia in England and Wales, 1981-91.Epidemiol Infect. 1993 Jun;110(3):567-73.

Smeds L, Künstner A. 2011. ConDeTri--a content dependent read trimmer for Illumina data. PLoS One. 2011;6(10):e26314.

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE.1986. Fluorescence detection in automated DNA sequence analysis. Nature. 1986 Jun 12-18;321(6071):674-9.

Smith CK, Abuoun M, Cawthraw SA, Humphrey TJ, Rothwell L, Kaiser P, Barrow PA, Jones MA. 2008. *Campylobacter* colonization of the chicken induces a proinflammatory response in mucosal tissues. FEMS Immunol Med Microbiol. 2008 Oct;54(1):114-21.

Snelling WJ, Matsuda M, Moore JE, Dooley JS. 2005. *Campylobacter jejuni*. Lett Appl Microbiol. 2005;41(4):297-302.

Spiller R, Garsed K. 2009. Postinfectious irritable bowel syndrome. Gastroenterology. 2009;136(6):1979-88.

Sproston EL, Ogden ID, MacRae M, Dallas JF, Sheppard SK, Cody AJ, Colles FM, Wilson MJ, Forbes KJ, Strachan NJ. 2011. Temporal variation and host association in the *Campylobacter* population in a longitudinal ruminant farm study. Appl Environ Microbiol. 2011 Sep;77(18):6579-86.

Stanley KN, Wallace JS, Currie JE, Diggle PJ, Jones K. 1998a. The seasonal variation of thermophilic campylobacters in beef cattle, dairy cattle and calves. J Appl Microbiol. 1998 Sep;85(3):472-80.

Stanley KN, Wallace JS, Currie JE, Diggle PJ, Jones K. 1998b. Seasonal variation of thermophilic campylobacters in lambs at slaughter. J Appl Microbiol. 1998 Jun;84(6):1111-6.

Strachan NJC, Forbes K. 2014. Extensive spatial and temporal clustering of *Campylobacter* infections. In: Sheppard SK, editor. *Campylobacter* Ecology and Evolution. 1st edition. Norfolk, UK: Caister academic Press; 2014. pp. 253.

Sulonen J, Kärenlampi R, Holma U, Hänninen ML. 2007. *Campylobacter* in Finnish organic laying hens in autumn 2003 and spring 2004. Poult Sci. 2007 Jun;86(6):1223-8.

Taboada EN, Clark CG, Sproston EL, Carrillo CD. 2013. Current methods for molecular typing of *Campylobacter* species. J Microbiol Methods. 2013 Oct;95(1):24-31. Review.

Tam CC, Rodrigues LC, Petersen I, Islam A, Hayward A, O'Brien SJ. 2006. Incidence of Guillain-Barré syndrome among patients with *Campylobacter* infection: a general practice research database study.J Infect Dis. 2006 Jul 1;194(1):95-7. Epub 2006 May 2

Tareen AM, Dasti JI, Zautner AE, Gross U, Lugert R. 2011. Sulphite: cytochrome c oxidoreductase deficiency in *Campylobacter jejuni* reduces motility, host cell adherence and invasion. Microbiology. 2011 Jun;157(Pt 6):1776-85.

Taylor DN. 1992. *Campylobacter* infections in developing countries. In *Campylobacter jejuni*: Current status and future needs. eds Nachamkin I., Blaser M.J. and Tompkins L.S. pp, 20–30 Washington DC: American Society for Microbiology.

Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, Swaminathan B. 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol. 1995 Sep;33(9):2233-9. Review.

THL,2016.http://www.julkari.fi/bitstream/handle/10024/131379/URN_ISBN_978-952-302-710-7.pdf?sequence=1

THL.Tilastotietokanta.https://sampo.thl.fi/pivot/prod/fi/ttr/shp/fact_shp?row=area12260&column=time-12059&filter=reportgroup-

Trokhymchuk A, Waldner C, Chaban B, Gow S, Hill JE. 2014. Prevalence and diversity of *Campylobacter* species in Saskatchewan retail ground beef. J Food Prot. 2014 Dec;77(12):2106-10.

Van Belkum A., Scheher S, van Alphen L, Verbrugh H. 1998. Short-sequence DNA repeats in prokaryotic genomes. Microbiol. Mol. Biol.Rev. 62:275-293.

Vandamme and De Ley. 1991. Proposal for a new family, Campylobacteraceae. Int. J. Syst. Bacteriol. 41:451-455.

Vandamme P. 2000. Taxonomy of the family *Campylobacteracea*. In: Nachamkin I, Szymanski C, Blaser M, editors. *Campylobacter*. 2nd edition. Washington DC: ASM Press; 2008. pp. 3-26.

Vandamme P, Debruyne L, De Brandt E, Falsen E. 2010. Reclassification of *Bacteroides ureolyticus* as *Campylobacter ureolyticus* comb. nov., and emended description of the genus *Campylobacter*. Int J Syst Evol Microbiol. 2010 Sep;60(Pt 9):2016-22.

Vidal AB, Davies RH, Rodgers JD, Ridley A, Clifton-Hadley F. 2014. Epidemiology and control of *Campylobacter* in modern broiler production. In: Sheppard SK, editor. *Campylobacter* Ecology and Evolution. 1st edition. Norfolk, UK: Caister academic Press; 2014. pp. 297-312.

Vierikko A, Hänninen ML, Siitonen A, Ruutu P, Rautelin H. 2004. Domestically acquired *Campylobacter* infections in Finland. Emerg Infect Dis. 2004 Jan;10(1):127-30.

Vogt RL, Sours HE, Barrett T, Feldman RA, Dickinson RJ, Witherell L. 1982. *Campylobacter* enteritis associated with contaminated water. Ann Intern Med. 1982 Mar;96(3):292-6.

Waldenström J, Broman T, Carlsson I, Hasselquist D, Achterberg RP, Wagenaar JA, Olsen B. 2002. Prevalence of *Campylobacter jejuni*, *Campylobacter lari*, and *Campylobacter coli* in different ecological guilds and taxa of migrating birds. Appl Environ Microbiol. 2002 Dec;68(12):5911-7.

Waldenström J and Griekspoor P. 2014. Ecology and host associations of Campylobacter in wild birds. In: Sheppard SK, editor. *Campylobacter* Ecology and Evolution. 1st edition. Norfolk, UK: Caister academic Press; 2014. p. 265-284.

Wang Y, Taylor DE. 1989. Natural transformation in *Campylobacter* species. Journal of bacteriology.1989;172(2):949.

Wassenaar TM, Newell DG. 2000. Genotyping of *Campylobacter* spp. Appl Environ Microbiol. 2000 Jan;66(1):1-9. Review.

Williams LK, Trantham EK, Cogan TA. 2014. Conditional commensalism of *Campylobacter* in Chickens. In: Sheppard SK, editor. *Campylobacter* Ecology and Evolution. 1st edition. Norfolk, UK: Caister academic Press; 2014.p. 313-319.

Wilson DJ, Gabriel E, Leatherbarrow AJ, Cheesbrough J, Gee S, Bolton E, Fox A, Fearnhead P, Hart CA, Diggle PJ. 2008. Tracing the source of campylobacteriosis. PLoS Genet. 2008 Sep 26;4(9):e1000203.

Wilson DL, Rathinam VA, Qi W, Wick LM, Landgraf J, Bell JA, Plovanich-Jones A, Parrish J, Finley RL, Mansfield LS, Linz JE. 2010. Genetic diversity in Campylobacter jejuni is associated with differential colonization of broiler chickens and C57BL/6J IL10-deficient mice. Microbiology 156: 2046-57.

Wong TL, Hollis L, Cornelius A, Nicol C, Cook R, Hudson JA. Prevalence, numbers, and subtypes of *Campylobacter jejuni* and *Campylobacter coli* in uncooked retail meat samples. J Food Prot. 2007 Mar;70(3):566-73.

WHO, World Health Organization. http://www.who.int/mediacentre/factsheets/fs255/en/

Zerbino DR, Birney E. 2008.Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 2008 May;18(5):821-9.

Zhang J, Halkilahti J, Hänninen ML, Rossi M. Refinement of whole-genome multilocus sequence typing analysis by addressing gene paralogy. J Clin Microbiol. 2015 May;53(5):1765-7.

Yan W, Chang N, Taylor DE. Pulsed-field gel electrophoresis of *Campylobacter jejuni* and *Campylobacter coli* genomic DNA and its epidemiologic application.J Infect Dis. 1991 May;163(5):1068-72.