View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Massey Research Online

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Molecular Epidemiology of Waterborne Zoonoses in the North Island of New Zealand

A thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Veterinary Science
(Epidemiology and Public Health)
at
Institute of Veterinary, Animal and Biomedical Sciences (IVABS),
Massey University, Palmerston North,
New Zealand

Rima Devi Shrestha

2016

(Submitted October 2015)

Abstract

Campylobacter, *Cryptosporidium* and *Giardia* species are three important waterborne zoonotic pathogens of global public health concern. This PhD opens with an interpretive overview of the literature on *Campylobacter*, *Cryptosporidium* and *Giardia* spp. in ruminants and their presence in surface water (Chapter 1), followed by five epidemiological studies of *Campylobacter*, *Cryptosporidium* and *Giardia* spp. in cattle, sheep and aquatic environment in New Zealand (Chapters 2-6).

The second chapter investigated four years of retrospective data on *Campylobacter* spp. (n=507) to infer the source, population structure and zoonotic potential of *Campylobacter jejuni* from six high-use recreational rivers in the Wanganui-Manawatu region of New Zealand through the generalised additive model, generalised linear/logistic regression model, and minimum spanning trees. This study highlights the ubiquitous presence of *Campylobacter* spp. in both low and high river flows, and during winter months. It also shows the presence of *C. jejuni* in 21% of samples containing highly diverse strains, the majority of which were associated with wild birds only. These wild birds-associated *C. jejuni* have not been detected in human, suggesting they may not be infectious to human. However, the presence of some poultry and ruminant-associated strains that are potentially zoonotic suggested the possibility of waterborne transmission of *C. jejuni* to the public. Good biosecurity measures and water treatment plants may be helpful in reducing the risk of waterborne *Campylobacter* transmission

In the third study, a repeated cross-sectional study was conducted every month for four months to investigate the source of drinking source-water contamination. A total of 499 ruminant faecal samples and 24 river/stream water samples were collected from two rural town water catchments (Dannevirke and Shannon) in the Manawatu-Wanganui region of New Zealand, and molecular analysis of those samples was performed to determine the occurrence of *Campylobacter*, *Cryptosporidium*, and *Giardia* spp. and their zoonotic potential. The major pathogens found in faecal samples were *Campylobacter* (n=225 from 7/8 farms), followed by *Giardia* (n=151 from 8/8 farms), whereas *Giardia* cysts were found in many water samples (n=18), followed by *Campylobacter* (n=4). On the contrary, *Cryptosporidium* oocysts were only detected in a few faecal (n=18) and water (n=3) samples. *Cryptosporidium* and *Giardia* spp. were detected in a higher number of faecal samples from young animals

(≤ 3 months) than juvenile and adult animals, whereas *Campylobacter* spp. were highly isolated in the faecal samples from juvenile and adult ruminants. PCR-sequencing of the detected pathogens indicated the presence of potentially zoonotic *C. jejuni* and *C. coli*, *Cryptosporidium parvum* (gp60 allelic types IIA18G3R1 and IIA19G4R1) and *Giardia duodenalis* (assemblages AII, BII, BIII, and BIV) in cattle and sheep. In addition, potentially zoonotic *C. jejuni* and *Giardia duodenalis* assemblages AII, BI, BII, and BIV were also determined in water samples. These findings indicate that these three pathogens of public health significance are present in ruminant faecal samples of farms and in water, and may represent a possible source of human infection in New Zealand.

In the fourth study, PCR-sequencing of *Cryptosporidium* spp. isolates obtained from the faeces of 6-week- old dairy calves (n=15) in the third study were investigated at multiple loci (18S SSU rDNA, HSP70, Actin and gp60) to determine the presence of mixed *Cryptosporidium* spp. infections. *Cryptosporidium parvum* (15/15), *C. bovis* (3/15) and *C. andersoni* (1/15), and two new genetic variants were determined along with molecular evidence of mixed infections in five specimens. Three main *Cryptosporidium* species of cattle, *C. parvum, C. bovis and C. andersoni,* were detected together in one specimen. Genetic evidence of the presence of *C. Anderson* and two new *Cryptosporidium* genetic variants are provided here for the first time in New Zealand. These findings provided additional evidence that describes *Cryptosporidium* parasites as genetically heterogeneous populations and highlighted the need for iterative genotyping at multiple loci to explore the genetic makeup of the isolates.

The *C. jejuni* and *C. coli* isolates (n=96) obtained from cattle, sheep and water in the third study were subtyped to determine their genetic diversity and zoonotic potential using a modified, novel multi-locus sequence typing method ("massMLST"; Chapter 5). Primers were developed and optimised, PCR-based target-MLST alleles' amplification were performed, followed by next generation sequencing on an Illumina MiSeq machine. A bioinformatics pipeline of the sequencing data was developed to define *C. jejuni* and *C. coli* multi-locus sequence types. This study demonstrated the utility and potential of this novel typing method, massMLST, as a strain typing method. In addition to identifying the possible *C. jejuni/coli* clonal complexes or sequence types of 68/96 isolates from ruminant faeces and water samples, this study reported three new *C. jejuni* strains in cattle in New Zealand, along with many strains, such as CC-61, CC-828 and CC-21, that have also been found in humans, indicating the public health significance of these isolates circulating on the farms in the two water catchment areas. Automation of the massMLST method and

may allow a cost-effective high-resolution typing method in the near future for multi-locus sequence typing of large collections of *Campylobacter* strains.

In the final study (Chapter 6), a pilot metagenomic study was carried out to obtain a snapshot of the microbial ecology of surface water used in the two rural towns of New Zealand for drinking purposes, and to identify the zoonotic pathogens related to waterborne diseases. Fresh samples collected in 2011 and 2012, samples from the same time that were frozen, and samples that were kept in the preservative RNA*later* were sequenced using whole-genome shotgun sequencing on an Illumina MiSeq machine. *Proteobacteria* was detected in all the samples characterised, although there were differences in the genus and species between the samples. The microbial diversity reported varied between the grab and stomacher methods, between samples collected in the year 2011 and 2012, and among the fresh, frozen and RNA*later* preserved samples. This study also determined the presence of DNA of potentially zoonotic pathogens such as *Cryptosporidium, Campylobacter* and *Mycobacterium* spp. in water. Use of metagenomics could potentially be used to monitor the ecology of drinking water sources so that effective water treatment plans can be formulated, and for reducing the risk of waterborne zoonosis.

As a whole, this PhD project provides new data on *G. duodenalis* assemblages in cattle, sheep and surface water, new information on mixed *Cryptosporidium* infections in calves, a novel “massMLST” method to subtype *Campylobacter* species, and shows the utility of shotgun metagenomic sequencing for drinking water monitoring. Results indicate that ruminants (cattle and sheep) in New Zealand shed potentially zoonotic pathogens in the environment and may contribute to the contamination of surface water. A better understanding of waterborne zoonotic transmission would help in devising appropriate control strategies, which could reduce the shedding of *Campylobacter*, *Cryptosporidium*, and *Giardia* spp. in the environment and thereby reduce waterborne transmission.

Preface

This PhD thesis aimed to study the molecular epidemiology of waterborne zoonosis in New Zealand, focussing on top three notifiable diseases: campylobacteriosis, cryptosporidiosis, and giardiasis. The project aimed to determine the presence of *Campylobacter*, *Cryptosporidium*, and *Giardia* spp. in ruminants (cattle and sheep) on farms and surface water in two catchment areas in the North Island: Dannevirke and Shannon. In addition to providing relevant epidemiological data, this project also developed a novel typing method, "massMLST" and applied state of the art metagenomic approaches using next generation sequencing technology.

Acknowledgements

"Cultivate the habit of being grateful for every good thing that comes to you, and to give thanks continuously. And because all things have contributed to your advancement, you should include all things in your gratitude."
— Ralph Waldo Emerson

First and foremost, I would like to express enormous gratitude to my supervisors for their excellent help, guidance, and encouragement to accomplish the research work presented in this thesis. Thank you Dr. Alexandro Grinberg for your inspiring guidance to help me understand the concept of molecular studies, preparing manuscript, including this thesis. I am also enormously thankful to Dr. Patrick Biggs for bringing the concept of a novel "massMLST" method, helping me to understand the concept of next-generation sequencing, and assisting in bioinformatics data analysis. I wish to express my profound gratitude to Prof. Dr. Nigel P. French for helping me to find the research funds, providing amicable suggestions during the study period, particularly in the statistical analysis of chapter 2. Dr. Evelyn J. Pleydell, thank you for your valuable suggestions, especially for the chapter 1, 2 and 3 of this PhD study and supporting me morally when I was on stress. Thanks also go to Dr. Deborah J. Prattley for her assistance in improving my writing techniques.

This thesis would not have been prepared without the help of molecular epidemiology laboratory (mEpiLab) staffs: Dr. Anne Midwinter, Dr. V.S.R Dukkipati, Lynn Rogers, Sarah Moore, Anthony Pita, Niluka Velathanthri, Angy Reynolds, Errol Kwan, Tessy George, Rukhshana Akhtar, and Neville Haack; and Massey genome services staffs Trish McLenachan and Richard Fong. Thank you all for your technical assistance and laboratory expertise.

I am also thankful to Dr. Jonathan Marshall for providing help in analysing the results, Dr. Mark Stevenson, Dr Julie Collins-Emerson and Dr. Jackie Benschop for their valuable suggestions and encouragement during my study period.

I am immensely grateful to farmers involved in this study, Horowhenua and Tararua District Councils for allowing me permission to take faecal/water samples from their premises.

I gratefully acknowledge the support of New Zealand Development Scholarship, International Student Support Office, McGeorge Research Funding Agency, Ministry of Health, and IVABS' Postgraduate Research Fund Committee, without which I could not have undertaken the PhD studies. Thanks also go to mEpiLab and IVABS administrative staffs, Wendey Mahrey, Simon Verschaft, and Debbie Hill.

I am heartily indebted to my fellow students from mEpiLab, Epicentre and IVABS, especially Tessy, Springer, Patricia, Hamid, Fang, Shoukai, Barbara, Anja, Ben, Julanda, Peter, Emilie, Masako, Anou, Nelly, and Late Mbvuto for their friendship, sharing thoughts, widening my horizon, and being kind to me during the moments of stress. I am also thankful to Nepalese in the Palmerston North for their support during the first year of my study.

I am thankful to my parents, in-laws, and all my family members who directly and indirectly provided moral supports for completing my study.

Finally yet importantly, I am indebted to my husband, Saroj Yadav, for his continuous support during my study period, especially for driving vehicle for sampling, cooking delicious food, creating and performing music, and counselling and understanding me when I was in stress.

List of Contents

Abstract ii
Preface v
Acknowledgements vi
List of Contents viii
List of Tables xiii
List of Figures xvi
List of Abbreviations xx

Chapter 1 .... **1**

Campylobacter, *Cryptosporidium* and *Giardia* spp. in ruminants and their contamination to surface water: A literature review. 1

1.1. Background 1
- 1.1.1. Faecal pollution source 2
- 1.1.2. Pathogens in water and molecular studies 3

1.2. *Campylobacter* spp. infection in humans and ruminants 4
- 1.2.1. Campylobacteriosis in humans 4
- 1.2.2. Campylobacteriosis in farmed ruminants 6

1.3. *Cryptosporidium* spp. in humans and ruminants 10
- 1.3.1. Cryptosporidiosis in humans 11
- 1.3.2. Cryptosporidiosis in farmed ruminants 13

1.4. *Giardia* spp. in humans and ruminants 19
- 1.4.1. Giardiasis in humans 19
- 1.4.2. Giardiasis in farmed ruminants 21

1.5. Pathogen contamination of waterways 24
- 1.5.1. Faecal loading rate of pathogens in the environment 25
- 1.5.2. Transport of the pathogens to the waterways 26
- 1.5.3. Survival of pathogens in the environment 27
- 1.5.4. Occurrence of pathogens in waterways 29

1.6. Detection of pathogens in faeces and/or water 31
- 1.6.1. Pathogen isolation 31
- 1.6.2. Identification of pathogens 33

1.7. Mitigation of pathogens' presence in water 46

1.8. Waterborne zoonoses in New Zealand context: current knowledge Zealand 47
- 1.8.1. Epidemiology of *Campylobacter*, *Cryptosporidium* and *Giardia* infections in humans 47
- 1.8.2. Epidemiology of *Campylobacter*, *Cryptosporidium* and *Giardia* in farmed-ruminants in New Zealand 51
- 1.8.3. Survival and abundance studies of *Campylobacter*, *Cryptosporidium* and *Giardia* spp. in New Zealand environment 54

1.8.4. Occurrences of *Campylobacter*, *Cryptosporidium* and *Giardia* spp. in New Zealand surface water 55
1.9. Conclusion 56

Chapter 2 58

Dynamics, Source and Population Structure of *Campylobacter jejuni* Isolated from Six High-Use Recreational Rivers in New Zealand 58

2.1. Preamble 58
2.2. Abstract 58
2.3. Introduction 59
2.4. Materials and methods 62
2.4.1. Sources of *Campylobacter* data 62
2.4.2. Statistical analysis 65
2.4.3. Population structure analysis 68
2.5. Results 68
2.5.1. Descriptive analysis 68
2.5.2. Relationship of presumptive *Campylobacter* spp. with standardised river flow, month, and river sites 76
2.5.3. Relationship of ruminant-associated *Campylobacter* jejuni with standardised river flow, month, and river sites 80
2.5.4. Population structure analysis 80
2.6. Discussion 83
2.7. Conclusion 86
2.8. Highlights of the study 87
2.9. Acknowledgments 87

Chapter 3 88

Epidemiology and molecular characterisation of *Cryptosporidium*, *Giardia* and *Campylobacter* species in farmed ruminant faeces and pre-treatment drinking water in two rural town water catchments in New Zealand 88

3.1. Preamble 88
3.2. Abstract 88
3.3. Introduction 89
3.4. Materials and methods 93
3.4.1. Study design and sampling period 93
3.4.2. Drinking water catchment area and farm selection 94
3.4.3. Sample size computations 97
3.4.4. Farmer interviews 98
3.4.5. Ethics approval 99
3.4.6. Faecal sample collection and processing 99
3.4.7. Water sampling and processing 102

3.4.8. Polymerase chain reaction (PCR) and DNA sequence analysis ... 105
3.4.9. Statistical analysis ... 109
3.4.10. Genotyping and subgenotyping analysis of pathogens ... 111
3.5. Results ... 111
3.5.1. Questionnaire survey ... 112
3.5.2. Pathogens in faecal samples ... 114
3.5.3. Pathogens in water samples ... 119
3.5.4. Genotyping and subtyping analysis ... 121
3.6. Discussion ... 128
3.6.1. Pathogens in faeces ... 129
3.6.2. Pathogens in water samples ... 136
3.7. Conclusion ... 138
3.8. Highlights of the study ... 139
3.9. Acknowledgments ... 139

Chapter 4 ... 140

Infections with multiple *Cryptosporidium* species and new genetic variants in young dairy calves on a farm located within a drinking water catchment area in New Zealand ... 140

4.1. Preamble ... 140
4.2. Abstract ... 140
4.3. Introduction ... 141
4.4. Materials and methods ... 142
4.4.1. Study design and faecal samples ... 142
4.4.2. Laboratory analysis ... 143
4.5. Results ... 146
4.5.1. Identification of *Cryptosporidium* oocysts ... 146
4.5.2. Identification of *Cryptosporidium* species and subgenotypes ... 146
4.6. Discussion ... 150
4.7. Conclusion ... 151
4.8. Highlights of the study ... 151
4.9. Acknowledgments ... 152

Chapter 5 ... 153

Molecular characterisation of *Campylobacter jejuni* and *Campylobacter coli* isolates from faecal and water samples using next-generation sequencing technology: Introduction of a novel method "massMLST" ... 153

5.1. Preamble ... 153
5.2. Abstract ... 153
5.3. Introduction ... 154
5.4. Materials and methods ... 158
5.4.1. Sequencing concept and massMLST primer design ... 158

5.4.2. PCR optimisation ... 163
5.4.3. Amplification, normalisation, purification and NGS library preparation ... 165
5.4.4. Sequence data analysis ... 174
5.5. Results ... 178
5.5.1. PCR optimisation and sequencing of isolates ... 178
5.5.2. Sequence data analysis ... 178
5.6. Discussion ... 189
5.6.1. Primer design, PCR optimisation and sequencing of isolates ... 189
5.6.2. Sequence data analysis ... 190
5.7. Conclusion ... 194
5.8. Highlights of the study ... 195
5.9. Acknowledgments ... 196

Chapter 6 ... 197

Microbial communities present in the surface water destined for drinking purposes: use of next-generation sequencing technology ... 197

6.1. Preamble ... 197
6.2. Abstract ... 197
6.3. Introduction ... 198
6.4. Materials and methods ... 199
6.4.1. Water sampling and concentration by filtration ... 199
6.4.2. Genomic DNA extraction ... 201
6.4.3. NGS library preparation ... 201
6.4.4. Metagenome data analysis ... 202
6.5. Results ... 204
6.5.1. Sample analysis ... 204
6.5.2. Whole genome shotgun sequencing output ... 206
6.6. Discussion ... 219
6.6.1. Sample analysis ... 219
6.6.2. Whole genome shotgun sequencing output ... 220
6.7. Conclusion ... 223
6.8. Highlights of the study ... 223
6.9. Acknowledgments ... 224

Chapter 7 ... 225

General Discussion ... 225

7.1. Introduction ... 225
7.2. Use of molecular epidemiology in disease surveillance ... 228
7.3. Lessons learnt and future research ... 230
7.4. Conclusion ... 234

Appendix A **236**
Supplementary materials for the Chapter 1 236

Appendix B **239**
Supplementary materials for the Chapter 2 239

Appendix C **240**
Supplementary materials for the Chapter 3 240

Appendix D **255**
Supplementary materials for the Chapter 4 255

Appendix E **270**
Supplementary materials for the Chapter 5 270

Bibliography **290**

List of Tables

Chapter 2 .... **58**

Table 2.1: Information available in the data sets used for building logistic regression models to identify the relationship between the isolation of *Campylobacter* spp. from freshwater samples and various covariates 64

Table 2.2: Total number of water samples collected from six recreational river sites in the Manawatu region of New Zealand, and the number and percentage of presumptive *Campylobacter* spp. and *C. jejuni* in those samples 69

Table 2.3: Clonal complexes (CC), sequence types (ST), seven housekeeping genes, total number and relative frequency of *C. jejuni* isolated from six recreational rivers. Blue, green, yellow colour respectively denotes the first, second and third highest frequency of sequence types obtained in this study 70

Table 2.4: Results of two multivariate logistic regression models (A and B) of culturing presumptive *Campylobacter* spp. from recreational water samples; the two models differed only in the measures of month use that were fitted 79

Chapter 3 .... **88**

Table 3.1: Campylobacteriosis, cryptosporidiosis and giardiasis cases reported in New Zealand and other developed countries in the year 2010. 'NA' indicates data not available 91

Table 3.2: Primers and conditions used for PCR amplification of *Campylobacter* spp. loci in this study 107

Table 3.3: Primers and conditions used for PCR amplification of *Cryptosporidium* and *Giardia* loci in this study 108

Table 3.4: Information on farm management variables from Dannevirke and Shannon farms that were obtained through questionnaire interview of farmers 113

Table 3.5: Percentage of presumptive Campylobacter, Cryptosporidium and Giardia species detected alone or co-occurred in ruminant faecal samples collected from the two catchment areas 115

Table 3.6: Apparent prevalence of presumptive *Campylobacter*, *Cryptosporidium* and *Giardia* in three different species' faecal samples collected from the ground on farms in the two catchment areas: Dannevirke and Shannon of the Manawatu-Wanganui region in New Zealand (N: number of samples positive and CI: 95% confidence interval). A "-"denotes the pathogens not detected in the samples 116

Table 3.7: Percentage of regular and storm water samples found positive for presumptive *Campylobacter*, *Cryptosporidium* and *Giardia*. The water samples were collected from an

abstraction point for the Tamaki River in Dannevirke and from the Mangaore Stream at the Shannon water treatment plant 120

Table 3.8: Sensitivity (recovery rate) and probable number of *Cryptosporidium* oocysts and *Giardia* cysts present in the Tamaki River, Dannevirke and Mangaore Stream, Shannon 120

Table 3.9: Total number and percentage of presumptive *Campylobacter*, *C. jejuni*, *C. coli,* in three different species' faecal samples collected from the ground on farms of the two catchment areas: Dannevirke and Shannon of the Manawatu region in New Zealand. A "-"denotes the pathogens are not detected in the samples 122

Table 3.10: Number and percentages of *Cryptosporidium* genotypes determined by PCR and DNA sequencing of immunofluorescence assay (IFA) positive isolates detected in dairy, beef cattle, and sheep faecal samples collected from the ground on farms within the two catchment areas: Dannevirke and Shannon of the Manawatu region in New Zealand. A "-"denotes the pathogens were not detected in the samples (in brackets, %)........ 124

Table 3.11: Number of *Giardia* assemblages identified through PCR and DNA sequencing analysis of by immunofluorescence assay (IFA) positive *Giardia* isolates detected in dairy, beef cattle, and sheep faecal samples collected from the ground on farms within the two catchment areas: Dannevirke and Shannon of the Manawatu region in New Zealand. A "-"denotes the pathogens were not detected in the samples (in brackets, %). Only numbers are shown for the identified assemblages of *Giardia..* 125

Table 3.12: Number of *Cryptosporidium parvum* gp60 allelic types and *Giardia* assemblages identified in this study, and that are found in bovine, and humans of New Zealand (retrieved from New Zealand *Cryptosporidium* and *Giardia* sequence database). The A "-"denotes the specified subtypes not detected yet 126

Table 3.13: Number of *Campylobacter*, *Cryptosporidium* and *Giardia* genotypes identified through PCR and DNA sequencing analysis of isolates from the collected water samples (sites: Tamaki River, Dannevirke and Mangaore Stream, Shannon). A "-"denotes the pathogens were not detected in the samples (in brackets, %) 127

Chapter 4 140

Table 4.1: Primers and PCR conditions used for amplification of *Cryptosporidium* loci. 145

Table 4.2: *Cryptosporidium* taxa and subgenotypes identified in cattle faecal specimens by sequence analysis of PCR products. In the same column, identical sequences are denoted by superscripts 148

Chapter 5 153

Table 5.1: Origin and number of *C. jejuni* and *C. coli* isolates used in this study 158

Table 5.2: The primer sequences for seven housekeeping genes for use in massMLST of *Campylobacter* isolates. These primers were modified from the conventional *Campylobacter* MLST primers used to sequence seven housekeeping genes. Each gene's primer sequences

contain two fragments along with transposome adapters at the 5' end. The table also shows melting temperature (Tm), sequence size and PCR product size for each locus 161

Table 5.3: Preparation of reaction master mixes containing PCR buffer, Mg^{2+}, dNTPs, Taq polymerase enzyme, and water for amplification of the isolates……………………………………...172

Table 5.4: PCR conditions used in this study to amplify the two fragments of seven housekeeping multi-locus sequence typing alleles of *Campylobacter* species. Temperatures (Temp.) stated are in Celsius (°C) and times are in minutes (min) or seconds (s)……………...173

Table 5.5: The number of samples (N=96) of *Campylobacter* that were successfully amplified for sequencing using NGS for each of the seven housekeeping genes, and the number of unsuccessful samples (N=45) that were amplified for traditional MLST using Sanger sequencing……………………………………………………………………………………………………...179

Table 5.6: Table 5.6: Examples showing the assigned allele numbers for each housekeeping gene of "processed" and "processed trimmed" sequences generated for each massMLST sample. The complete profile with defined sequence type (ST-best case based on non-ambiguous alleles) and clonal complex (CC) is highlighted in green. Red indicates the discrepant alleles between "processed" and "processed trimmed" sequences and yellow shows the extra allele determined based on the highest bit-scores for a given sample. (#) indicates the number of best bit-score determined for that sample, (*) denotes the possible ST and CC determined for some samples, and UD** represents the ST/CC that could not be determined 187

Table 5.7: Numbers of *Campylobacter jejuni* and *C. coli* clonal complexes (CC) and sequence types (ST) determined in the massMLST of the 96 samples from various origins. (-) denotes the absence of those CC/ST in the given isolates 188

Chapter 6 196

Table 6.1: The minimum DNA concentration (ng/μL) required for whole-genome shotgun metagenomics library-preparation method, and the DNA concentration present in the water samples collected from the two sites in the year 2011 and 2012. The libraries were run on an Illumina MiSeq machine at Massey Genome Service 205

Table 6.2. The sequences generated from different samples used in the whole-genome shotgun metagenomics study. (-) indicate sequencing was not conducted for that sample. .. 205

List of Figures

Chapter 1 .... **1**

Figure 1.1: Campylobacteriosis (black), cryptosporidiosis (red) and giardiasis (orange) cases notified in people of New Zealand between the year 1988 and 2013 (ESR, 2014). Cryptosporidiosis and giardiasis were included in notification list from 1996 only 51

Chapter 2 .... **58**

Figure 2.1: The recreational swimming water study sampling sites within the Manawatu, New Zealand. Manawatu A refers to the Albert Street section of the Manawatu River, and Manawatu H refers to the Hopelands Picnic Reserve section of the Manawatu River 63

Figure 2.2: A bar plot showing the monthly variations in presumptive *Campylobacter* spp. positive water samples from six recreational water sites in the Manawatu region combining the data collected between December 2005 and April 2009. The vertical lines represent the 95% confidence intervals 72

Figure 2.3: Notched Box-and-whiskers plot showing the temporal variation of standardised river flow rates (cubic meter per second) for four recreational water sites in the Manawatu between December 2005 and April 2009. Top and bottom of the box represent 25th and 75th percentile, a notch with a dark line in the box shows median value and the two ends of the whiskers depict minimum and maximum value of river flow rates. Round dots depict the outliers of the data 73

Figure 2.4: Error bar plot illustrating the variation in the proportion of presumptive *Campylobacter* spp. positive water samples (red dot) across each decile of the standardised river flow rates for the four recreational water sites. The vertical lines represent the 95% confidence intervals. 74

Figure 2.5: A histogram depicting the number of isolates that were attributed to the probability of being ruminant-associated *C. jejuni*.. 75

Figure 2.6: Generalised additive model plots demonstrating the relationship between standardised river flow and the response (the presence of *Campylobacter* spp. in the water samples) when adjusted for sampling sites and sampling months. The solid line shows the fitted model and the dashed lines represent the 95% confidence intervals. The rungs at the X-axis indicate the individual data points 77

Figure 2.7: A model-fitted relationship of the probability of detecting presumptive *Campylobacter*-positive water samples in each month from four rivers, adjusted for the site and river flow. The solid coloured line with shade is the mean probability and 95% confidence

interval. Manawatu A refers to the Albert Street section of the Manawatu River, and Manawatu H refers to the Hopelands Picnic Reserve section of the Manawatu River 78

Figure 2.8: Generalised additive model plots demonstrating the relationship between sampling months and the response (the probability of obtaining ruminant-associated *Campylobacter* in the water samples) when adjusted for river flow and sampling months. The solid line shows the fitted model and the dashed lines represent the 95% confidence intervals. Each line and number at the X-axis denote the month from March (3) to December (12) 81

Figure 2.9: A minimum spanning tree of subtypes of *C. jejuni* found in the six recreational rivers in the Manawatu region of New Zealand. Each pie chart represents one subtype and the colours represent different river sites. Larger pie charts indicate a greater number of isolates of that subtype present. The thicker connecting lines show greater similarities between subtypes. For example, a thick black solid line between ST 2381 and ST 3656 illustrate that they are different in only one allele of seven housekeeping genes 82

Chapter 3 88

Figure 3.1: The Dannevirke (a) and Shannon (b) catchment areas along with river/streams, water abstraction (intake) points and farms within the catchment. 96

Figure 3.2: Distribution of *Campylobacter*, *Cryptosporidium* and *Giardia* detected in a) each farm from location Dannevirke (DF) and Shannon (SF) (Upper graph) and b) faecal samples from different age groups of animals (lower graph) 117

Figure 3.3: Proportion of *Campylobacter*, *Cryptosporidium* and *Giardia* detected in faecal samples collected in different months from August to November. The lines show the number of water samples positive for each pathogen on that farms 118

Chapter 4 140

Figure 4.1: Variable repeat region of HSP70 (PCR 1) allele found in this study. Type 1 and 1a contain 12 repeat units of 12 bps each, whereas type 2 has only 11 repeat units. Each colour represents a 12-bp repeat unit type. Note also the difference in the position of the repeat unit types between type 1 and 1a. 149

Chapter 5 153

Figure 5.1: The concept behind the re-designed Multi-Locus Sequence Typing (MLST) primers and obtaining the massMLST sequences… 160

Figure 5.2: Exemplification of the method used to identify unamplified or problematic primers by using different forward and reverse primers combinations (1 - 6), including the original MLST primer, for a new amplification. 165

Figure 5.3: The amplification, purification, quantification, and normalisation steps involved in the first round of PCR (PCR 1) in the massMLST method……………………………………………..169

Figure 5.4: Amplification of PCR1 amplicons in order to add the adapters, and the purification, quantification, and normalisation steps involved in the second round of PCR (PCR2) in the massMLST method........170

Figure 5.5: Steps involved in sequencing the amplicons from the prepared library of 96 samples using an Illumina MiSeq machine........171

Figure 5.6: A flow diagram showing the steps involved from processing the resultant sequences to determining the sequence types and clonal complexes........181

Figure 5.7: Violin plots of the overall coverage of nucleotides sequences for each of the seven *Campylobacter* MLST alleles using the data from full-length PCR products to show SNP data. The upper and lower graph represents the processed and processed trimmed data........200

Figure 5.8: Violin plots of the overall coverage of nucleotides sequences for each of the seven *Campylobacter* MLST alleles using the data from full-length PCR products to show consensus coverage. The upper and lower graph represent the processed and processed trimmed data. 182

Figure 5.9: Violin plots of nucleotide sequences coverage in a log scale for each of 96 samples by plate location for all seven *Campylobacter* MLST alleles. The A01 to H12 represents wells in a 96-well plate, each of which contains a separate sample. The white dot is the median nucleotide sequences coverage and the dark black lines are the interquartile range along with whiskers........183

Figure 5.10: Violin plots of nucleotide sequences coverage in log scale for each of 96 samples by plate location for the *gltA Campylobacter* MLST allele. The A01 to H12 represents wells in a 96-well plate, each of which contains a separate sample. The white dot is the median nucleotide sequences coverage and the dark black lines are the interquartile range along with whiskers. 184

Chapter 6........196

Figure 6.1: This bar chart is a MEGAN output on whole genome sequencing data. The bar chart shows various organisms that were obtained in NGS of the seven samples processed with different methods that are compared for their taxonomic kingdom classification........208

Figure 6.2a: This bricks chart is a MEGAN output on whole genome sequencing data. The bricks chart show predominant phyla with increased number of tiles for those phyla, and the seven samples processed with different methods are compared for their taxonomic phyla classification........209

Figure 6.2b: This bricks chart is a MEGAN output on whole genome sequencing data. The bricks chart show predominant phyla with increased number of tiles for those phyla, and the seven samples processed with different methods are compared for their taxonomic phyla classification........210

Figure 6.3: This word cloud is a MEGAN output on whole genome sequencing data. The word cloud show predominant class with a larger size text for those classes, and the seven samples processed with different methods are compared for their taxonomic class classification. 211

Figure 6.4: This word cloud is a MEGAN output on whole genome sequencing data. The word cloud show predominant family with a larger size text for those family, and the seven samples processed with different methods are compared for their taxonomic family classification. . 212

Figure 6.5: This word cloud is a MEGAN output on whole genome sequencing data. The word cloud show predominant genera with a larger size text for those genus, and the seven samples processed with different methods are compared for their taxonomic genus classification. .. 213

Figure 6.6: This word cloud is a MEGAN output on whole genome sequencing data. The word cloud shows microbial richness between Dannevirke samples processed with Nextera and TruSeq sequencing methods at the taxonomic family (upper) and species (lower) levels of classification. 214

Figure 6.7: This word cloud is a MEGAN output on whole genome sequencing data. The word cloud shows microbial richness between Dannevirke samples that were processed fresh, after 3 months of freezing and of preservation in RNA*later* at the taxonomic family (upper) and species (lower) levels of classification. 215

Figure 6.8: This two-dimensional principal co-ordinate analyse (PCoA) is a MEGAN output on whole genome sequencing data. 216

Figure 6.9: This dendogram tree is a MEGAN output on whole genome sequencing data, and shows microbial diversities between the seven different samples. 217

List of Abbreviations

18S SSU rDNA /18S rRNA	Small Subunit 18S Ribosomal RNA
AIC	Akaike Information Criterion
BA	Blood Agar
bg	Beta-Giardin
bp	Base Pairs
CC	Clonal Complex
CDC	Centre For Disease Control
DAPI	4-,6-Diam,Idino-2-Phenylindole
ELISA	Enzyme Linked Immunosorbent Assay
ESR	The Institute Of Environmental Science And Research
FAO	Food And Agriculture Organisation
GAM	Generalised Additive Model
Gdh	Glutamate Dehydrogenase
GIS	Geographical Information System
GLM	Generalised Linear Model
GP60	Glycoprotein (Or 60-Kda Glycoprotein)
HSP70	70 kDa Heat Shock Protein Gene
IFA	Immunofluorescence Assay
IMS	Immunomagnetic Separation
IVABS	Institute Of Veterinary, Animal And Biomedical Sciences
mCCDA	Modified Charcoal, Cefoperazone Desoxycholate Agar
MEGAN	Metagenome Analyser
MGS	Massey Genome Service
MGW	Molecular Grade Water
MLST	Multilocus Sequence Typing
MSSP	Manawatu Sentinel Surveillance Program
MST	Minimum Spanning Tree

MU	Massey University
mEpiLab	Molecular Epidemiology And Public Health Laboratory
NCBI	National Centre For Biotechnology Information
NGS	Next-Generation Sequencing
NIWA	National Institute Of Water And Atmospheric Research
NZGL	New Zealand Genomics Limited
OPG	Oocysts Per Gram Of Faeces
OR	Odds Ratio
PAUDA	Protein Alignment Using A DNA Aligner
PCoA	Principal Co-Ordinate Analyses
PCR	Polymerase Chain Reaction
pDNA	Pseudo-DNA
PEG	Polyethylene Glycol
PFGE	Pulsed Field Gel Electrophoresis
PRU	Protozoa Research Unit
QC	Quality Check
RFLP	Restriction Fragment Length Polymorphism
SNPs	Single Nucleotide Polymorphisms
spp	Species
ST	Sequence Types
Tpi	Triosephosphate Isomerase
UPGMA	Unweighted Pair Group Method With Arithmetic Mean
USEPA	United States Environmental Protection Agency
UV	Ultra Violet
VBNC	Viable But Non-Culturable
WGS	Whole Genome Shotgun Sequencing
WHO	World Health Organisation
WINZ	Water Information New Zealand