Molecular Epidemiology of Campylobacteriosis and Evolution of *Campylobacter jejuni* ST-474 in New Zealand

A thesis presented

in partial fullfilment of the requirements

for the degree of Doctor of Philosophy

at Massey University

Vathsala Mohan

Institute of Veterinary, Animal and Biomedical Sciences

Massey University

Palmerston North, New Zealand

2011

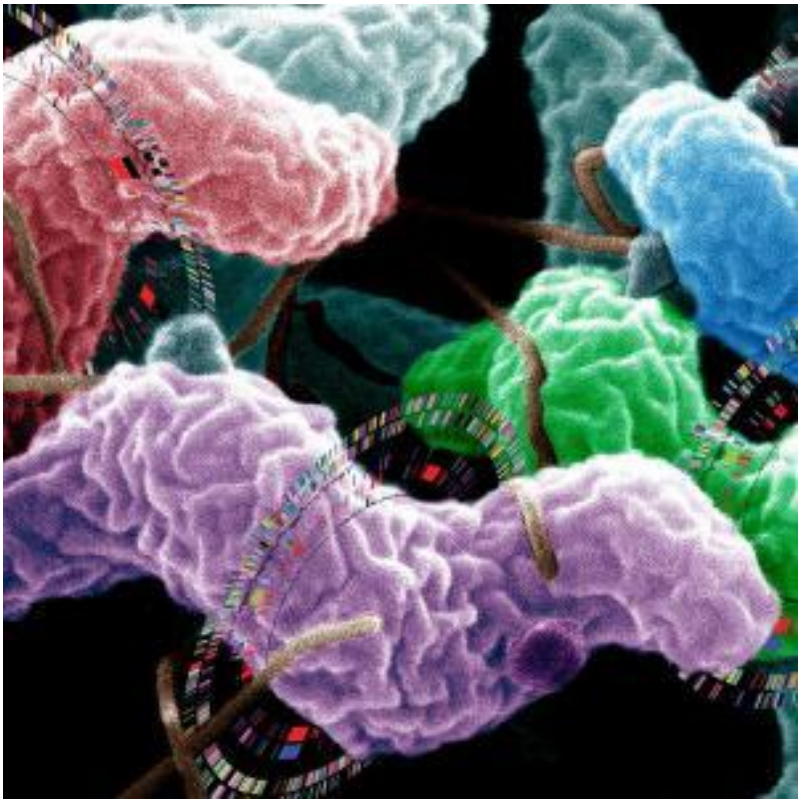
(Submitted July 14, 2011)

Institute of Veterinary, Animal and Biomedical Sciences

Massey University

Palmerston North, New Zealand

2011



— Abstract —

Population genetics and phylogenetics have the potential to provide enormous insights into the epidemiology and ecology of disease causing pathogens. Molecular datasets are the basis to infer population structure, gene flow (between host populations and between different geographical locations) and to predict the evolutionary dynamics of pathogens. *Campylobacter* colonisation in food producing animals has been extensively studied and the population structure and host association of *C. jejuni*, the most commonly reported gastro-enteric pathogen, has also been well defined. In contrast, host-pathogen relationships and the population structure of *C. jejuni* in urban wild birds and pets have not been well defined on a wide range of spatial and/or temporal scales. A greater understanding of these details should allow disease control authorities to track the transmission of pathogens from one host species to another, identify the origin of pathogens and to better understand environmental factors influencing underlying molecular mechanisms.

In the first study in this thesis the presence of *C. jejuni* in mallard ducks and starlings within five playgrounds in Palmerston North, New Zealand was studied. The prevalence of *Campylobacter* and *C. jejuni* in both species showed a bimodal seasonal pattern. The population structure and population differentiation of *C. jejuni* in these species were examined using multilocus sequence typing (MLST). Rarefaction analyses showed that the *C. jejuni* populations within mallard ducks were more diverse than starlings, particularly during the winter. Pairwise fixation indices showed that the population of *C. jejuni* in ducks was significantly different from that of starlings and that it differed over time. Conspicuous host association was evident with clonal complexes of *C. jejuni* such as ST-1034, ST-692 and ST-1332 specific to ducks and ST-177 and ST-682 specific to starlings. In addition, a larger proportion of *C. jejuni* genotypes that could not be assigned a clonal complex were found in both ducks and starlings, particularly during the winter.

In the second study, *C. jejuni* from domestic pets (dogs and cats) were characterised using MLST and by typing the cell surface antigens, *por*A and *fla*A. The ST-45 complex, a clonal complex predominantly reported in human campylobacteriosis cases, was found to be the predominant clone present in both species. These findings shed some light on the contribution of pets as a putative source of human campylobacteriosis cases in New Zealand.

In the third study, the ST-474 *C. jejuni* genotype, considered to be the endemic strain in New Zealand, was isolated from human cases and poultry carcasses from the Manawatu region from 2005 to 2009. Seven samples of ST-474 were sequenced and a subset of 50 full length genes were studied. These analyses demonstrated molecular differences between full length genes that were identical in the region used for MLST. Further, alleles characteristic of the ST-474 genome within the investigated metabolic housekeeping genes (n = 25) were identified. Our findings were that ST-474 genome is genetically distinct from other *C. jejuni* reference genomes with respect to certain alleles. In addition, MLST alleles were found to be robust predictors of the most recent common ancestors of a genome. The fourth study investigated the genetic stability and vulnerability of the informational genes to various evolutionary forces within the seven ST-474 genomes. Twenty five genes comprised of nucleotide metabolism, repair and ribosomal functions were investigated showing a high level of genetic diversity in the DNA repair as well as nucleotide metabolic genes such as *gid*A, *ogt*, *rec*J, *ssb*, *uvr*A, *uvr*B and *xse*A. In contrast, the ribosomal genes were stable and identical across the seven genomes. The insertion of selenocysteine in three of the 25 genes indicates the presence of horizontal gene transfer within the ST-474 genomes. It is hypothesised that the genetic uniqueness of ST-474 may have arisen due to the geographic isolation of New Zealand, its poultry industry and an absence of exchange of sequence types which might typically occur through international trade of fresh poultry meat.

Collectively, the studies presented in this thesis provide a better understanding of the dynamism of *C. jejuni* as a species and ST-474's adaptational capacity and evolutionary potential (within the investigated set of genes) in response to changing intracellular and extracellular environments. This thesis has introduced the idea of using individual full length gene analysis, demonstrating the molecular differences between genes that contained identical alleles at the MLST loci. The research approaches implemented in this thesis can be readily applied to any pathogenic bacteria, particularly foodborne and emerging pathogens such as *E. coli* and *Salmonella*. This, in turn should provide new opportunities for bacterial drug targets and vaccine candidates.

Acknowledgements

I sincerely thank my supervisors Professor Nigel French, Dr. Mark Stevenson and Dr. Jonathan Marshall for their valuable guidance without which this work would not have been possible. I thank my late supervisor Dr. Grant Hotter, who died suddenly during my second year of PhD, for his dedication, his enthusiasm and vision. I thank Dr. Deb Prattley, who came forward to help me during my difficult times and encouraging me and; Dr. Eve Pleydell for giving me moral support and encouragement. I owe my special thanks to Dr. Patrick Biggs, Massey Genome Service, IVABS, Massey University who had helped me learn certain valuable bioinformatics skills for my genome work and reading through my genome chapters.

I extend my sincere thanks to all of my laboratory colleagues from the mEpiLab and the staff from the Epicentre – Dr. Julie Collins, Dr. Anne Midwinter, Ms. Sarah Moore, Ms. Angie Reynolds, Mr. Errol Kwan, Ms. Rukshana Akther, Ms. Lynn, Ms. Rebecca Pattison, Mr. Simon Verschaffelt and all others for their timely help and suggestions. I thank Ms. Wendy Maharey and Mr. Neville Haack, for proof reading my chapters and references, respectively. I thank Dr. Eve Pleydell and Dr. Petra Mullner who gave me an opportunity to tutor in the MVM course that helped to cover some of my living expenses when my scholarship was finished.

I thank Vet. Nurse, Ms. Jill Hogan, for having been a great help to me during sample collection with a lot of patience and a friendly face. I thank my friends both past and present for their co-operation, help and fun they rendered me. I once again thank all the staff and students from the Epicentre, Massey University, for their support and help to run the journal club effectively and particularly, Ms. Christine Cunningham for being very caring and for giving me motherly affection.

I thank Dr. P. Ramadass (former HOD, Dept. of Animal Biotechnology, MVC, TANUVAS, India) who encouraged me to take up this PhD programme at the Massey University

and for visiting me during my candidature in New Zealand. I thank my best friend, Dr. T. V. Meenambigai, Associate Professor, Leptospira Diagnostic Laboratory, TANUVAS, India, for her constant valuable advises during my stay at India as well as New Zealand. I am much obliged to Dr. T. V. Meenambigai and Dr. Uma Arumugam (Associate Professor, TANUVAS) as well as all the ministerial staff members in the Registrar Office of Tamil Nadu University of Veterinary and Animal Science, Chennai, India, that helped me to get relieved from my duties to proceed with my higher studies.

I thank the New Zealand International Doctoral Scholarships (NZIDRS), funded by the New Zealand Government and Massey University for the Massey Doctoral Scholarship and the Massey Post-Graduate Research Fund that helped me procure lab consumables for my project. I thank Ms. Debbie Cresswell and Ms. Sally Iwakua who had been a great help in managing my scholarships and providing me prompt help and support.

I thank my parents for their continuous encouragement and blessings throughout my study period without which I would not have been able to overcome the difficult times during my candidature.

I am indebted to my family, my husband, Dr. Mohan Perumal and my children, Master. Honesh Vishva Bushan Mohan and Master. Vijayan Mohan for their co-operation, understanding and sacrifices during the tough times of my sampling and their patience during the entire period of my study.

Finally, I thank the supreme Lord of knowledge and wisdom, Lord Ganesha for removing my obstacles and giving me the strength to succeed.

Nomenclature

AFLP	Amplified fragment length polymorphism
AMOVA	Analysis of molecular variance
AT	Adenine – thymine
CBI	Codon usage bias index
CI	Confidence interval
CRISPR	Clustered regularly interspaced short palindromic repeats
DNA	De-oxyribonucleic acid
dN	Non-synonymous nucleotide substitution
dS	Synonymous nucleotide substitution
ESR	Environmental Science and Research Ltd
EU	European Union
FAO	Food and Agricultural Organization of the United Nations
GBS	Guillain-Barré syndrome
GC	Guanine – cytosine
GC3	Guanine – cytosine at the third codon position
HGT	Horizontal gene transfer
HL	Heat-labile (antigen)
HS	Heat-stable (antigen)
HK	Housekeeping
Ka	Non-synonymous nucleotide substitution

Ks	Synonymous nucleotide substitution
mCCDA	Modified cefoperazone charcoal desoxycholate agar
MLEE	Multilocus enzyme electrophoresis
MLSA	Multilocus sequence analysis
MLST	Multilocus sequence typing
MOMP	Major outer membrane protein
MST	Minimum spanning trees
NZ	New Zealand
NCBI	National Center for Biotechnology Information, USA
NZFSA	New Zealand Food Safety Authority
ORFs	Open reading frames
PCR	Polymerase chain reaction
*por*A	porin gene A
PFGE	Pulsed field gel electrophoresis
RAPD	Randomly amplified polymorphic DNA
rRNA	Ribosomal ribonucleic acid
RE	Restriction enzyme
REA	Restriction endonuclease analysis
RFLP	Restriction fragment polymorphism
spp.	Species (multiple)
ST	Sequence type
ST-U/A	Sequence type unassigned
SVR	Short variable region
TD	Tajima D
tRNA	Transfer ribonucleotide

WHO	World Health Organisation

List of Publications

Patrick J Biggs, Paul Fearnhead, Grant Hotter, **Vathsala Mohan**, Julie Collins-Emerson, Errol Kwan, Tom E Besser and Nigel P French., Whole-genome camparison of *Campylobacter jejuni* isolates indistinguishable on the basis of MLST and *fla*A SVR reveals multiple loci of different ancestral lineage. PLoS ONE, 2011. 6:e27121

Nigel French, Julie Collins-Emerson, Anne Midwinter, Patrick Biggs, **Vathsala Mohan**, Petra Muellner, Adrian Cookson, Patricia Jaros, Hamid Irshad, Eve Pleydell, Donald Campbell, Phil Carter. The evolution, epidemiology and control of enteric zoonoses in New Zealand. New Zealand Microbiological Society, New Zealand, 2011.

Vathsala Mohan, French Nigel, Biggs J. Patrick, Stevenson Mark, Marshall Jonathan, and Hotter Grant. Analysis of *por*A–MOMP genes of *C. jejuni* ST-474 genotypes show differential evolution. Federation of Infection Societies Annual Scientific Conference (FIS 2011), Manchester, UK, 2011.

V. Mohan, N. P. French, M. Stevenson, J. C. Marshall, and G. Hotter. Wild birds, pets and Campylobacter: A molecular understanding of their impact on the public health. 84th Scientific Meeting of the Wellington Health and Biomedical Research Society, New Zealand, 2011.

Vathsala Mohan, French Nigel, Stevenson Mark, Marshall Jonathan, and Hotter Grant. Characterisation of *Campylobacter jejuni* from pets in New Zealand. Molecular Diagnostics World Congress 2011, South San Francisco Conference Centre, USA, 2011

Vathsala Mohan, French Nigel, Biggs J. Patrick, Stevenson Mark, Marshall Jonathan, and Hotter Grant. Phylogenetic analysis of *Campylobacter jejuni* ST-474 genotypes. 2nd Annual Beyond The Genome Conference 2011, Washington DC, USA, 2011.

N. P. French, Marshall, J. and Vathsala, M. Campylobacter in Food and Environment. New and emerging data on typing of *Campylobacter* spp. strains in animals, environmen-

tal matrices and humans. Final report: 07-10436. Prepared for the New Zealand Food Safety Authority and Ministry for the Environment, 2010.

Vathsala, M., French, N., Stevenson, M., Marshall, J., and Hotter, G. *Campylobacter jejuni* colonisation and population structure in sympatric urban population of ducks and starlings. In advanced preparation for submission to *Environmental Microbiology*.

Vathsala, M., French, N., Stevenson, M., Marshall, J., and Hotter, G. Characterisation of *Campylobacter jejuni* from pets in New Zealand by combining multilocus sequence typing (MLST), *por*A and *fla*A typing. In advanced preparation for submission to *BMC Infectious Diseases*

Vathsala, M., French, N., Biggs, J. P., Stevenson, M., Marshall, J., and Hotter, G. Molecular phylogeny of seven *Campylobacter jejuni* ST-474 genomes based on an extended subset of housekeeping genes and *por*A/MOMP structural gene. In advanced preparation for submission to *BMC Microbiology*.

Vathsala, M., French, N., Biggs, J. P., Stevenson, M., Marshall, J., and Hotter, G. A genome wide comparison of selected subsets of ribosomal and DNA repair genes among seven *Campylobacter jejuni* ST-474 isolates. In advanced preparation for submission to *Journal of Evolutionary Biology*.

Contents

Acknowledgements **vii**

Nomenclature **ix**

List of Publications **xiii**

1 Introduction **1**

2 Review of literature **7**

2.1 Introduction 7

2.1.1 History 8

2.1.2 Taxonomy of *Campylobacter* 8

2.1.3 Morphological and biochemical characteristics 9

2.2 Epidemiology 11

2.2.1 Sources of *Campylobacter* 17

2.2.2 *Campylobacter* spp. in wild birds 18

2.2.3 *Campylobacter* spp. in pets 19

2.3 Molecular epidemiology 20

2.4 Typing techniques 22

2.4.1 Phenotyping 23

2.4.2 Genotyping 27

2.4.3 Macro-restriction mediated analysis 28

2.4.4 Polymerase chain reaction based assays 29
2.4.5 Multilocus sequence typing 34
2.4.6 Whole genome sequencing 39
2.5 Knowledge search 45

3 *Campylobacter jejuni* colonisation **47**

3.1 Introduction 48
3.2 Experimental procedures 51
3.2.1 Study design 51
3.2.2 Collection of faecal material 54
3.2.3 Bacterial Isolation and DNA preparation 55
3.2.4 Speciation and characterisation by Polymerase Chain Reaction . 56
3.2.5 Multilocus sequence typing 56
3.2.6 *flaA* and *por*A typing 59
3.2.7 Statistical analysis 59
3.3 Results 62
3.3.1 Prevalence of *Campylobacter* spp. and *C. jejuni* 62
3.3.2 Sequence type diversity by host 67
3.3.3 Sequence type diversity by sampling site 70
3.3.4 Sequence type diversity by sampling period 72
3.3.5 Analysis of molecular variance and Fst 76
3.3.6 *C. jejuni* cell surface antigens typing and diversity 79
3.3.7 Comparison with wider population of sequence types in the PubMLST database 79
3.4 Discussion 85

4 Characterisation of *Campylobacter jejuni* from pets **89**

4.1 Introduction 90

4.2 Experimental procedures 93

4.2.1 Study design 93

4.2.2 Collection of faecal material 95

4.2.3 Bacterial isolation and DNA preparation 95

4.2.4 Speciation by PCR 96

4.2.5 Multilocus sequence typing 97

4.3 Results 101

4.3.1 Prevalence of faecal *Campylobacter* spp. in dogs and cats . . . 101

4.3.2 *C. jejuni* genotypes and cell surface antigen types of dogs and cats 102

4.3.3 Comparison of *C. jejuni* genotypes with wider population . . . 106

4.4 Discussion 107

5 Phylogeny of housekeeping and *por*A genes **111**

5.1 Introduction 112

5.2 Experimental procedures 121

5.2.1 Reference *C. jejuni* genomes 122

5.2.2 *C. jejuni* isolates, DNA preparation and sequencing 124

5.2.3 *de novo* genome assembly and gene prediction 124

5.2.4 Selection of metabolic housekeeping genes 124

5.2.5 Retrieval of gene sequences from *C. jejuni* ST-474 gene predictions 126

5.2.6 Analysis of Guanine-Cytosine content, codon usage, selection pressure and recombination 127

5.2.7 Phylogenetic analysis 128

5.2.8 Analysis of *por*A gene 129

5.3 Results 129

5.3.1 Analysis of full length MLST housekeeping genes 129
5.3.2 Analysis of second subset of housekeeping genes: 143
5.3.3 Major outer membrane protein, the *por*A 161
5.4 Discussion 168

6 A genome wide comparison **177**
6.1 Introduction 178
6.1.1 Genes of genetic information processes in bacteria 179
6.1.2 Measures of evolution 189
6.1.3 New Zealand and *C. jejuni* 193
6.2 Experimental procedures 195
6.2.1 Reference *C. jejuni* genomes 195
6.2.2 Bacterial isolates, DNA preparation and sequencing 197
6.2.3 *de novo* genome assembly and gene prediction 197
6.2.4 Selection of repair and ribosomal genes 197
6.2.5 Retrieval of gene sequences from *C. jejuni* ST-474 gene predictions 202
6.2.6 Analysis of Guanine-Cytosine content, codon usage, selection pressure, recombination and evolutionary rates 203
6.2.7 Phylogenetic analysis 204
6.3 Results 205
6.3.1 Length and nucleotide composition 205
6.3.2 Analysis of Guanine-Cytosine contents 207
6.3.3 Codon usage bias 213
6.3.4 Selection pressure and evolutionary rate: 216
6.3.5 Phylogenetic analysis 223
6.3.6 Phylogenetic congruence between informational genes and MLST housekeeping alleles 233
6.4 Discussion 234

7 General discussion **243**

7.0.1 Conclusion 248

A Appendix A **A-1**

A.1 Summary A-1

A.2 Experimental procedures A-1

A.2.1 Sample collection A-1

A.2.2 Isolation of *Campylobacter* spp. : Laboratory procedures A-5

A.2.3 Cell surface antigen typing A-6

A.3 Results A-7

A.3.1 Multilocus sequence typing : Complete and partial *C. jejuni* MLST profiles from ducks and starlings A-7

A.3.2 Analysis of molecular variance and diversity indices A-11

A.3.3 Minimum spanning tree A-15

B Appendix B **B-1**

C Appendix C **C-1**

C.1 Introduction C-1

C.2 Materials and methods C-1

C.2.1 Retrieval of gene sequences C-1

C.3 Results C-6

C.3.1 Analysis of full length gene sequences of MLST seven house-keeping genes C-6

C.3.2 Codon usage C-9

D Appendix D **D-1**

D.1 Introduction D-1

D.2 Results D-2

D.2.1 Analysis of Guanine-Cytosine contents D-2

D.2.2 Codon usage bias D-4

D.2.3 Selection pressure, evolutionary rate and recombination: D-6

D.2.4 Phylogenetic analysis D-15

List of Figures

2.1 Seasonality of campylobacteriosis in developed countries 14

2.2 Circular genome of *C. jejuni* that shows the location of seven housekeeping genes on the chromosome 36

2.3 Sequences of *C. jejuni* ST-474 (poultry and human) with multiple mapping vs *C. jejuni* AL111168 46

3.1 Ducks and starlings sampling sites 53

3.2 Prevalence of *Campylobacter* during sampling occasions and in sampling sites 64

3.3 Prevalence of *C. jejuni* in mallard ducks and starlings 66

3.4 Frequency of *C. jejuni* sequence types (ST) in mallard ducks and starlings in different sampling sites 69

3.5 Species-based rarefaction curves of *C. jejuni* sequence types (ST) for mallard ducks and starlings 71

3.6 Distribution of *C. jejuni* sequence types (ST) in mallard ducks and starlings over time 74

3.7 Species-based rarefaction curves of *C. jejuni* sequence types (ST) for mallard ducks and starlings 75

3.8 Population differentiation of *C. jejuni* population in ducks and starlings during different seasons and at different sites 77

3.9 Minimum spanning tree (MST) of *C. jejuni* sequence types (ST) from mallard ducks, starlings and geese 78

4.1 The sampling sites of dogs 94

4.2 Seasonal prevalence of *Campylobacter* spp. and *C. jejuni* in dogs 102

5.1 *C. jejuni* circular genome showing metabolic genes 126

5.2 Comparison of nucleotide and amino acid sequences of *asp*A gene between ST-474 and reference genomes 133

5.3 Comparison of nucleotide and amino acid sequences of *tkt* gene between ST-474 and reference genomes 134

5.4 Alignment of *glt*A nucleotide sequences 135

5.5 Guanine-Cytosine variation in the MLST genes of *C. jejuni* genomes . . 137

5.6 Phylogenetic tree of concatenated MLST housekeeping genes. 142

5.7 Guanine-cytosine (GC) variation in the metabolic housekeeping genes – A 145

5.8 Guanine-cytosine (GC) variation in the metabolic housekeeping genes – B 146

5.9 Guanine-cytosine (GC) variation in the metabolic housekeeping genes – C 147

5.10 Relationship between the GC variance, length and the recombination sites in the metabolic genes 155

5.11 Maximum likelihood gene trees of individual metabolic genes 157

5.12 Maximum likelihood tree of concatenated metabolic genes 158

5.13 Schematic representation of the protein structure predicted by Zhang et al. (2000) 164

5.14 Nucleotide and protein recombination alignments of *por*A, and phylogenetic tree 165

5.15 DNA based phylogenetic tree of *por*A (MOMP) gene 166

5.16 Amino acid based recombination phylogenetic tree of *por*A (MOMP) gene 167

6.1 Replication and DNA repair system in prokaryotes 182

6.2 Mismatch repair system in prokaryotes 183

6.3 Base excision repair system in prokaryotes 184

6.4 Recombinational repair systems in prokaryotes –A 185

6.5 Recombinational repair systems in prokaryotes – B 186

6.6 *C. jejuni* circular genome showing DNA repair genes 202
6.7 Guanine-cytosine (GC) variance in the ribosomal and repair genes – A. 208
6.8 Guanine-cytosine (GC) variance in the ribosomal and repair genes – B. . 209
6.9 Guanine-cytosine (GC) variance in the ribosomal and repair genes – C. . 210
6.10 Guanine-cytosine (GC) variance in the ribosomal and repair genes – A. 211
6.11 Relationship between the GC variance, length and the recombination sites in the repair genes 222
6.12 Maximum likelihood gene tree of *gid*A 224
6.13 Maximum likelihood gene tree of *ogt* 225
6.14 Maximum likelihood gene tree of *ssb* 226
6.15 Maximum likelihood gene tree of *xse*A 227
6.16 Maximum likelihood gene tree of *rec*J 228
6.17 Maximum likelihood gene tree of *uvr*A 229
6.18 Maximum likelihood gene tree of *uvr*B 230
6.19 Phylogenetic tree of ribosomal, repair and nucleotide metabolic genes . 231
6.20 NeighbourNet tree of ribosomal, repair and nucleotide metabolic genes 232

A.1 Pictures of starling faecal material recorded during the sampling trial period A-2
A.2 Stages of laboratory procedures A-5
A.3 Flow chart of laboratory procedures A-5
A.4 The minimum spanning tree (MST) partitioned based on the clonal complexes (CC) A-16
A.5 The minimum spanning tree (MST) showing *fla*A and *por*A alleles clustered with the MLST profiles A-17

C.1 Comparison of primer sequences of *glm*M allele between *pgm* and *glm*M C-5
C.2 Maximum likelihood gene trees of metabolic genes – A C-21
C.3 Maximum likelihood gene trees of metabolic genes – B C-22

C.4 Maximum likelihood gene trees of metabolic genes – C C-23

D.1 Maximum likelihood gene trees of ribosomal and repair genes D-16

D.2 Maximum likelihood gene trees of ribosomal and repair genes D-17

D.3 Maximum likelihood gene trees of ribosomal and repair genes D-18

D.4 Maximum likelihood gene trees of ribosomal and repair genes D-19

List of Tables

2.1 Biochemical characteristics of *Campylobacter* spp. 10

2.2 Number of cases and incidence risk of campylobacteriosis, New Zealand 2005 – 2009. Adapted from Nicol et al. (2010). 13

2.3 Genes and gene positions used in a MLST typing scheme for *C. jejuni*. . 35

2.4 *Campylobacter jejuni subsp. jejuni* reference genomes. 41

3.1 The forward and reverse primer sequences used for the amplification of housekeeping genes of *C. jejuni* (Miller et al. 2005). 57

3.2 Prevalence estimates of *Campylobacter* and *C. jejuni* in mallard ducks, starlings and geese 62

3.3 MLST allelic profiles, *fla*A,*por*A nucleotide alleles and their frequency in mallard ducks 82

3.4 MLST allelic profiles, *fla*A,*por*A nucleotide alleles and their frequency in starlings. 83

3.5 Sequence types (STs) compared with the wider population of sequence types in the PubMLST database 84

4.1 Primers and target genes used for PCR 97

4.2 Primer sequences used for the amplifictaion of seven housekeeping alleles for *C.jejuni* MLST typing (Miller et al. 2006) 98

4.3 The prevalence of *Campylobacter* spp. and *C. jejuni* at the sampling sites 103

4.4 *C. jejuni* genotypes from dog and cat faecal samples 104

4.5 *C. jejuni* MLST partial profiles from dog faecal samples 104

4.6 Analysis of molecular variance of *C. jejuni* genotypes from pets 105

5.1 *Campylobacter jejuni* subsp. *jejuni* reference genomes and their MLST clonal complexes 123

5.2 Metabolic genes used in the *C. jejuni* MLST scheme (first subset of genes) 125

5.3 Metabolic genes selected from the MLST schemes of different bacterial species (second subset of genes) 125

5.4 MLST allelic profiles of reference genomes and *C. jejuni* ST-474 genomes 130

5.5 Metabolic housekeeping genes and their lengths 131

5.6 Synonymous and non-synonymous substitutions found within identical MLST alleles at full length gene level 132

5.7 Positions of MLST alleles covered on full length gene sequences 132

5.8 Tajima D values for the MLST housekeeping genes 138

5.9 Average of mean relative evolutionary rates of non-synonymous codons in the MLST housekeeping genes 139

5.10 Number of non-synonymous codons in MLST genes under purifying selection based on ω values 139

5.11 Recombination sites identified in the full length MLST genes of 19 genomes 141

5.12 Second subset of genes and their lengths (base pairs long) 144

5.13 Ancestry shared between reference and ST-474 genomes based on GC and GC3 contents 148

5.14 Ancestry shared with respect to codon usage between the reference and ST-474 genomes 149

5.15 Number of non-synonymous codons in the second subset of genes under purifying selection based on ω values 150

5.16 Tajima's D values for the second subset of housekeeping genes 151

5.17 Average of mean relative evolutionary rates of non-synonymous codons in the second subset of genes 152

5.18 Recombination sites identified in the second subset of genes 154

5.19 Ancestry shared between the reference and ST-474 genomes according to individual gene trees 160

5.20 Codon usage bias and guanine-cytosine contents of *por*A gene 162

6.1 *Campylobacter jejuni subsp. jejuni* reference genomes and their MLST clonal complexes 196

6.2 Genes of repair mechanism, their gene ID, name, function and the pathways they are involved in. 199

6.3 Gene lengths of repair genes : Bold fonts refer to genes with different lengths 206

6.4 Ancestry shared between reference and ST-474 genomes based on GC and GC3 contents 212

6.5 CBI and scaled chi square similarities shared amongst ST-474 genomes within the repair and ribosomal genes that differ 214

6.6 Ancestry shared between the reference and ST-474 genomes according to codon usage: 215

6.7 Tajima's D values of the ribosomal, repair genes and nucleotide metabolic genes 216

6.8 Number of non-synonymous codons in the repair genes under purifying selection based on ω values 217

6.9 Average of mean relative evolutionary rates of non-synonymous codons of repair, ribosomal and nucleotide metabolic genes 218

6.10 Recombination sites identified within ribosomal and repair genes 219

6.11 MLST allelic profiles of reference genomes and *C. jejuni* ST-474 genomes 233

A.1 Faecal sample details of ducks and starlings for the isolation of *Campylobacter* spp. A-3

A.2 Faecal sample details of ducks and starlings for the isolation of *Campylobacter* spp. continued A-4

A.3 *flaA* and *por*A typing A-6

A.4 Complete MLST profiles of *C. jejuni* from ducks and starlings A-8

A.5 Complete MLST profiles of *C. jejuni* from ducks and starlings - continued A-9

A.6 Partial MLST profiles of *C. jejuni* from ducks and starlings A-10

A.7 Analysis of molecular variance (AMOVA) A-12

A.8 Genetic differentiation of *C. jejuni* populations with respect to different sampling periods A-13

A.9 Diversity indices of *C. jejuni* populations of ducks and starlings from different sampling sites A-14

C.1 BLAST table of *C. jejuni* ST-474 metabolic genes C-2

C.2 Guanine-Cytosine (GC) content of seven MLST housekeeping genes of *C. jejuni* genomes C-7

C.3 GC3 content of metabolic genes C-8

C.4 Codon bias index (CBI) for twenty metabolic housekeeping genes . . . C-10

C.5 Codon usage bias (Scaled chisquare) indices for metabolic housekeeping genes from reference and ST-474 genomes C-11

C.6 Tajima D test values, synonymous and nonsynonymous sites of metabolic housekeeping genes C-12

D.1 GC contents of selected subset of ribosomal and repair genes D-2

D.2 GC3 contents of selected subset of ribosomal and repair genes D-3

D.3 Codon usage bias indices of ribosomal and repair genes D-4

D.4 Schaled chi square codon usage bias indices of ribosomal and repair genes D-5

D.5 Polymorphic synonymous, nonsynonymous sites and the Tajima D values of the ribosomal and repair D-6

‘The goal of mankind is knowledge ... knowledge is inherent in man. No knowledge comes from outside: it is all inside. What man ‘learns’ is really what he discovers by taking the cover off his own soul, which is a mine of infinite knowledge.’

Swami Vivekananda

A wise man is superior to any insults which can be put upon him, and the best reply to unseemly behavior is patience and moderation.

Moliere

Patience is the companion of wisdom.

Saint Augustine

xxx

CHAPTER 1

Introduction

Campylobacter spp. are members of the ϵ-proteobacteria, a diverse group of gram negative organisms found in a variety of habitats. Members of the *Campylobacterales* family are motile, non-spore forming curved rods demonstrating a surprising adaptability to various environmental niches (Lee & Newell 2006). The genus is comprised of a group of closely related organisms including *C. jejuni*, *C. coli*, *C. lari*, *C. upsaliensis*, *C. helviticus* and *C. curvisus* that colonise the gastrointestinal tracts of a wide variety of host species as commensals. In industrialised countries *Campylobacter* species, particularly *C. jejuni* and *C. coli* are considered the most important zoonotic enteric bacterial pathogens of humans (Wassenaar & Newell 2000). Furthermore, some *Campylobacter* spp. for example *C. fetus subsp. fetus* and *C. jejuni* can cause abortion in sheep (Allerberger et al. 1991, Sahin et al. 2008). The majority of reported human campylobacteriosis cases in developed countries are attributed primarily to *C. jejuni* followed by *C. coli*, *C.upsaliensis*, *C. lari* and *C. fetus* (Skirrow 1994, Friedman et al. 2000, Gillespie et al. 2002, Lopez et al. 2002, Sheppard et al. 2009). Due to the similarities in the clinical manifestations, disease histories (Gillespie et al. 2002) and lack of good biochemical markers for a differential diagnosis between *Campylobacter* spp., epidemiological studies exploring human campylobacteriosis have treated the disease caused by different *Campylobacter* species (*C. jejuni*, *C. coli* and other species) as a single clinical entity (On 1996, Siemer et al. 2005). In recent years molecular methods such as species specific polymerase chain reaction (PCR) and multilocus sequence typing (MLST) have made it easier for researchers to differentiate *Campylobacter* spp. (Dingle et al. 2005).

Ingestion of undercooked meat is thought to be the main source of campylobacteriosis in humans (Skirrow & Benjamin 1980, Kwan et al. 2008). The risk of campylobacteriosis

arising from consumption of contaminated food (particularly poultry meat) has been extensively studied (Eberhart-Phillips et al. 1997, Baker et al. 2006, 2007). Other important modes of transmission include exposure to faecal material from livestock, contact with animals (particularly ruminants) and recreational swimming (Savill et al. 2003, Mullner et al. 2009, Wilson et al. 2009). Understanding the relative contribution of each exposure pathway is critical if effective control measures are to be devised.

Structure of the thesis

This thesis is comprised of seven chapters. All except for this introduction, the literature review (Chapter 2) and the general discussion (Chapter 7) are presented in the form of manuscripts to be presented for peer-reviewed publication. As a result there is some repetition in the objectives and materials and methods. However, the results and the findings are unique to every chapter.

Chapter 1 provides an overview of the thesis and its overall objectives. Chapter 2 reviews the history, characteristics of *Campylobacter* spp., the disease, the disease status in developed and developing countries and the sources and typing techniques used for *Campylobacter* spp.

Urban wild birds, pets and *C. jejuni* populations

Campylobacter spp. have been isolated from both domestic and wild birds (Luechtefeld et al. 1980, Fricker & Metcalfe 1984, Frost 2001), and poultry as a source has been thoroughly investigated. Detailed knowledge of the molecular epidemiology of *C. jejuni* in wild birds is limited, however recent research has shown the genotypes of *C. jejuni* found in wild birds are similar to those isolated from the human campylobacteriosis cases (Colles et al. 2003, Broman et al. 2004, Colles et al. 2009, French et al. 2009a, Hughes et al. 2009). Similarly, pets that live in close proximity to humans have been shown to transmit *Campylobacter* spp. directly to their owners and a small number of reports have identified pets as a source of infection, particularly small children (Hald & Madsen 1997, Jimenez et al. 1999, Tenkate & Stafford 2001, Wolfs et al. 2001, Fullerton et al. 2007, Tsai et al. 2007, Chaban et al. 2010).

MLST is a high throughput nucleotide based technique which is widely used for molecular epidemiological studies of *Campylobacter* spp. The PubMLST database[1] has facilitated molecular epidemiological research into *Campylobacter* spp., making inter-laboratory comparisons of MLST data possible. Previous reports have shown that *Campylobacter* populations show large differences among host species and environmental niches (McCarthy et al. 2007) and that their lineages (clonal complexes, CC) undergo frequent recombinations (Fearnhead et al. 2005, Miller et al. 2006, McCarthy et al. 2007, Colles et al. 2008, 2009). Further, MLST has identified genotypes that are associated with different host sources and genotypes that are more broadly distributed among animal sources (Miller et al. 2006, Kwan et al. 2008, Carter et al. 2009). Genotypes with rarer allelic profiles that could not be assigned to a complex have been identified using MLST, particularly in wild birds (Colles et al. 2003, 2008, 2009, Hughes et al. 2009).

With this background, and given the scarcity of detailed epidemiological investigations of *C. jejuni* in wild birds and pets, the aims of this thesis are to:

- estimate the prevalence of *C. jejuni* in the faeces of wild birds (mallard ducks and starlings) and domestic pets (dogs and cats) in a provincial city in the lower North Island of New Zealand;
- identify the genotypes of *C. jejuni* in mallard ducks, starlings, dogs and cats and assess their relationship with human campylobacteriosis cases diagnosed in the same geographical area; and
- investigate the population structure and population differentiation of *C. jejuni* genotypes among mallard ducks and starlings at different sampling sites and at different times of the year in an effort to assess the potential risk of wild birds and pets as a source of campylobacter infection in humans.

Chapter 3 and Chapter 4 describe the findings from the molecular epidemiological investigations and population structure analyses carried out on *Campylobacter* spp. recovered from mallard ducks and starlings and domestic pets (dogs and cats). The prevalence of *Campylobacter* spp. and *C. jejuni* at different sampling sites and sampling periods, the population structure and population differentiation of *C. jejuni* in ducks and starlings at

[1]URL: (`http://pubmlst.org/Campylobacter`)

different sites and periods are described in Chapter 3. Similarly, Chapter 4 describes the prevalence and population structure of *C. jejuni* in dogs and cats.

C. jejuni in New Zealand

New Zealand has, until recently, ranked first in the world for human campylobacteriosis notification rates (Baker et al. 2007, Marler 2010). *C. jejuni* multilocus sequence type 474 (ST-474) is an internationally rare genotype that has accounted for 24% to 34% of human campylobacteriosis cases in New Zealand (French 2008). It is a genotype strongly associated with poultry, particularly in those living in urban areas (French 2008, Mullner et al. 2009, 2010, McTavish et al. 2008). Following the isolation of two sub-strains of *C. jejuni* ST-474 during the same time period and from the same geographical location, one from a fresh chicken carcass and the other from a human clinical case, the two isolates were sequenced fully using next generation Solexa sequencing technology and their draft genomes submitted to GenBank[2] (French et al. 2009b, Biggs et al. 2011). It was found that 93% of the genes from the two ST-474 isolates were identical and 103 genes differed by at least one single nucleotide with 72 of them showing non-synonymous substitutions. Furthermore, it was inferred that 5% of the differences were due to mutations and 95% were through recombination (Biggs et al. 2011). Additionally, a non-homologous recombination event was identified in the human isolate (H22082) with the insertion of two extra genes, whereas the chicken isolate (P110b) showed no evidence of such an event.

Is MLST enough to study the evolution of *C. jejuni*?

As the seven housekeeping alleles used in the MLST typing scheme cover only $<0.2\%$ of the entire genome (Dingle & Maiden 2005), the phylogeny and the ancestral lineages of individual genes may differ when considered at the whole genome level. Further, the variations between the two *C. jejuni* ST-474 isolates as inferred by Biggs et al. (2011) suggests that the seven housekeeping genes of the MLST typing system may not sufficiently reflect the evolutionary histories and phylogeny of every single gene in a genome. With this brief background, an additional subset of 50 full length housekeeping genes involved in various metabolic and repair functions from seven *C. jejuni* ST-474 isolates

[2]URL:(`http://www.ncbi.nlm.nih.gov/genbank/`)

were investigated in Chapters 5 and 6. Five poultry *C. jejuni* ST-474 isolates (P569a, P694a, P73020, P110b and P179a) and two human *C. jejuni* ST-474 isolates (H22082 and H704a) were sequenced using next generation Solexa resequencing techniques.

Evolution within genes in a genome has broadly classified them into two lineages: operational or housekeeping genes and informational genes that are involved in replication, transcription, translation and other related processes (Jain et al. 1999). Operational (metabolic) genes are housekeeping genes that have been shown to be evolutionarily dynamic compared with informational genes (Rivera et al. 1998). Population structure analyses in *C. jejuni* have generally been carried out using only a subregion of housekeeping genes (alleles). The questions asked here are: (1) how reliable are housekeeping alleles in predicting the evolutionary history and the ancestry of other peer genes in a genome? (2) do the evolutionary histories of housekeeping and informational genes differ within *C. jejuni*? and (3) does ST-474 possess any conspicuous genetic features? To answer these questions, a selected subset of housekeeping genes from the seven *C. jejuni* ST-474 genomes (n = 50) were investigated for characteristics such as:

- guanine-cytosine (GC) content and codon usage;
- single nucleotide polymorphisms (SNP);
- selection pressure;
- events of recombination; and
- phylogenetic relationship.

Gene sequences from 12 *C. jejuni* reference genomes (both completed genomes and genomes in draft forms) were used for a comparative evolutionary analysis with ST-474 genomes to identify ancestral lineages. It is important to be able to identify differences among these two lineages, particularly identical strains or genotypes. A better understanding of these differences will provide greater insight into characteristics of these genes and their functional importance. This, in turn will: (1) allow sites on the genome that might be targeted by novel chemotherapeutic agents, or (2) facilitate the selection of potential vaccine candidates.

Generally, whole genome comparison studies performed on fully sequenced genomes do not necessarily focus on individual genes. As a result, details such as the contribution of individual genes in the evolution of an organism, evolutionary history and phylogeny of individual genes have not been given a great deal of attention. To address this knowledge gap Chapters 5 and 6 analyse the evolutionary events, phylogeny and the most recent common ancestor of 50 selected genes from seven fully sequenced *C. jejuni* ST-474 genomes. Chapter 5 provides a description of the metabolic housekeeping genes while Chapter 6 describes the (informational) ribosomal and repair genes and nucleotide metabolic genes.

Chapter 7 provides a general discussion and a set of conclusions from the thesis as a whole. This thesis documents the evolutionary potential of *C. jejuni* as a species using a population genetic approach. It sheds light on the *in-vivo* evolution and formation of new variants of *C. jejuni* within the investigated host species. The phylogenetic approaches that have been used enhance our understanding of the two major lineages of genes that are important determinants of a pathogen's evolutionary potential. These investigative approaches have potential to be applied to other foodborne and emerging pathogens.

CHAPTER 2

Review of literature

2.1 Introduction

Campylobacter spp. are members of the epsilonproteobacteria which are a diverse group of gram-negative organisms found in a variety of habitats (Newell 2001). The members of the *Campylobacterales* family are ubiquitous and demonstrate a surprising adaptability to various environmental niches (Newell 2001). The *Campylobacter* genus comprises a group of gram-negative bacteria that are closely related and are found to primarily colonise the gastrointestinal tracts of a wide variety of animal host species (Wassenaar & Newell 2000). Some of these bacteria are commensals, but many, particularly *Campylobacter jejuni*, and *Campylobacter coli*, are zoonotic enteric pathogens of humans and domestic animals (Wassenaar & Newell 2000, Newell 2001). *Campylobacter* spp. are the most frequently isolated bacterial pathogens in human gastroenteritic cases in industrialised countries, where the majority of reported clinical cases were attributed to *Campylobacter jejuni* (90% to 95 %) and *Campylobacter coli* (5% to 10 %) (as reviewed by Tauxe 1992). Because of the apparent similar disease histories of the two *Campylobacter* species, the predominant occurrence of *C. jejuni* among human clinical cases, and the scarcity of specific biochemical markers for diagnostics (On 1996, Siemer et al. 2005), speciation is performed routinely in only a limited number of clinical laboratories around the world. As a consequence, most of the studies that explored the epidemiology of human *Campylobacter* infections have treated *Campylobacter* infections as a single source of infection in the past (Siemer et al. 2005). With the advent of molecular diagnostic tools, other *Campylobacter* species that include *C. upsaliensis*, *C. fetus* and *C. lari* have also been identified to cause human infection (Gillespie et al. 2002, Lopez et al. 2002,

Siemer et al. 2005, Sheppard et al. 2009). This review focuses on the following four areas:

1. General background and the characteristics of *Campylobacter* spp.
2. A brief account of the disease caused by *Campylobacter* spp. in humans.
3. The epidemiology of *Campylobacter* spp. particularly *C. jejuni*.
4. Typing methods currently used and the population structure of *C. jejuni*.

2.1.1 History

In 1909 scientists (McFadyean & Stockman 1913) isolated a new bacterial species, *Vibrio foetusovid* from aborted lamb foetuses and their dams. In 1938 a bovine strain of *Vibrio* spp. was isolated from humans in conjunction with a milk-borne disease outbreak in the United Sates of America (Levy 1946). Subsequently, *Vibrio* spp. was isolated from 11 patients with gastroenteritis in the United Sates of America, of which 7 strains were *V. fetus*, and 4 were a closely related species, designated as 'related vibrios' (King 1957). Following King's report, Sebald & Veron (1963) differentiated two groups of *Vibrio* spp. based on carbohydrate fermentation and DNA guanine-cytosine (GC) content. The group with the lesser GC content was assigned to a new genus *Campylobacter*, meaning 'curved rods' in Greek. Since its first isolation in 1909, our knowledge has expanded to include the complete genome sequences of *C. jejuni* NCTC11168 (Parkhill et al. 2000, Gundogdu et al. 2007), *C. jejuni* 81116 (NCTC11828) (Pearson et al. 2003) *C. jejuni* RM1221 (Parker et al. 2006) and *C. jejuni* 81-176 (Hofreuter et al. 2006) and several other *Campylobacter* genomes are now available in GenBank. Although several *Campylobacter* spp. have been fully sequenced, many details regarding their pathogenicity, host association, population diversity, and epidemiology remain unclear.

2.1.2 Taxonomy of *Campylobacter*

The genus *Campylobacter* has been classified as follows:

Domain:	Bacteria
Phylum:	Proteobacteria
Class:	Epsilonproteobacteria
Order:	*Campylobacterales*
Family:	*Campylobacteraceae*
Genus:	*Campylobacter*

The DNA-rRNA hybridisation study of sixty strains which represented *Campylobacter* species, *Campylobacter* like organisms, *Wolinella*, *Bacteroides* and *flexispira* species found that *Campylobacter*, *Wolinella* and *flexispira* represented a separate sixth rRNA superfamily sensu De Ley within the group of gram negative bacteria (Vandamme & De Ley 1991, Vandamme & On 2001). This lineage has been referred to as epsilon subdivision of the proreobacteria. At present, the genus *Campylobacter* contains 31 species and 13 subspecies.[1] In 2001 the species *Campylobacter hyolei* was transferred to *Campylobacter coli* (Vandamme & On 2001).

2.1.3 Morphological and biochemical characteristics

Members of the genus *Campylobacter* are gram-negative, non spore forming, curved, S-shaped or spiral rods 0.2 – 0.8 μm wide and 0.5 – 5.0 μm long where, they often display spiral forms when daughter cells are attached together (Hansson et al. 2007). Campylobacters are motile and they move by a characteristic rotating movement often referred to as rapid corkscrew-like motion, using their unipolar or bipolar flagella (Parkhill et al. 2000). Members of *Campylobacter* are microaerophiles: microorganisms that grow in low oxygen concentrations (Kelly 2001). *C. jejuni* hydrolyses hippurate, whilst *C. jejuni* subsp. *doylei* varies in its ability to hydrolyse hippurate (Vandamme & Goossens 1992). Therefore, hippurate hydrolysis (Hwang & Ederer 1975) has become the most widely used biochemical test to identify *C. jejuni*, and also to differentiate it from *C. coli* that are phenotypically and genotypically similar (Walder et al. 1983). However, variability in the hippurase reaction has been observed for some strains of *C. jejuni* (Morris et al. 1985, Totten et al. 1987, Fermer & Engvall 1999) and hence a number of other additional biochemical tests are employed to differentiate *Campylobacter* spp. The growth characteristics and biochemical tests that are routinely used to characterise *Campylobacter* spp. are summarised in Table 2.1.

[1] URL: `http://www.bacterio.cict.fr/` last accessed 10 August 2010.

Table 2.1: Biochemical characteristics of *Campylobacter* spp.

Characteristic	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
a-Haemolysis	-	-	(-)	(-)	+	(-)	-	V	-	+	NA	V	V	NA	+	+	+	V	-	+	+	+	+
Oxidase	+	+	+	V	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	V	+	+
Catalase	W	V	+	-	+	-	+	(+)	V	-	-	+	+	+	V	+	+	+	-	(-)	+	V	-
Characteristic	-	-	-	V	-	V	-	-	-	-	-	-	(-)	NA	-	-	+	-	(+)	-	-	-	-
c-Glutamyltranspeptidase	-	(+)	-	-	-	NA	-	NA	NA	-	NA	-	-	NA	-	-	NA	-	NA	NA	NA	-	-
Urease production	-	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	V*	-
Hippurate hydrolysis	+	-	-	-	-	(-)	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-
Indoxyl acetate hydrolysis	+	-	+	-	+	V	-	-	V	+	-	-	-	-	+	+	-	-	-	+	-	-	+
Nitrate reduction	+	V	+	(-)	+	+	+	+	(+)	+	-	+	+	+	-	+	+	+	-	+	+	+	+
Selenite reduction	-	NA	V	(-)	-	-	(+)	-	-	-	-	+	+	NA	-	+	+	+	-	+	+	+	+
TTC reduction	-	NA	+	-	V	V	-	-	-	-	NA	-	-	NA	V	+	NA	+	-	-	-	-	V
Trace H-S on TSI agar	-	V	-	-	-	(-)	-	-	-	-	-	+	+	-	-	-	-	-	+	-	V	+	-
Growth at/in/on:																							
-5 C (microaerobic)	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
37 C (microaerobic)	+	+	+	+	+	V	+	+	-	+	+	+	+	+	+	+	+	+	+	-	V	+	+
4- C (microaerobic)	+	+	+	(+)	(+)	V	(+)	-	V	+	(-)	+	+	-	-	+	+	+	+	(-)	V	+	+
37 C (anaerobic)	-	+	-	+	-	+	(-)	V	+	-	+	-	+	-	-	-	+	-	+	+	+	+	-
Nutrient agar	-	-	+	(-)	+	+	+	+	+	(+)	NA	+	+	NA	+	+	NA	+	+	(-)	V	+	+
CCDA	-	+	+	(-)	(+)	(+)	+	+	V	+	NA	+	+	NA	+	+	NA	+	+	-	+	(+)	+
MacConkey agar	-	+	V	-	-	(+)	(+)	V	(+)	-	-	V	V	NA	-	-	+	-	(+)	-	+	V	-
1% Glycine	-	V	+	(-)	-	+	+	-	+	V	+	+	V	+	(-)	+	-	+	V	+	V	+	+
-% NaCl	-	NA	-	(-)	-	V	-	-	V	-	+	-	-	-	-	-	-	(+)	+	V	+	+	-
1% Bile	V	NA	(+)	-	V	-	+	+	-	+	NA	+	(+)	NA	+	+	NA	+	+	-	-	V	+
Requirement for H2	V	-	-	+	-	+	-	-	+	-	+	V	V	NA	-	-	-	-	+	+	+	-	-
Resistance to:																							
Nalidixic acid	-	V	-	(+)	V	+	+	V	V	-	V	+	+	+	-	-	+	V	(+)	(+)	-	(+)	-
Cephalothin	+	-	+	-	(+)	-	-	-	-	-	-	(-)	-	+	-	+	+	+	-	-	-	-	(-)

Taxa: 1, *C. avium sp. nov.*; 2, *C. canadensis*; 3, *C. coli*; 4, *C. concisus*; 5, *C. cuniculorum*; 6, *C. curvus*; 7, *C. fetus subsp. fetus*; 8, *C. fetus subsp. venerealis*; 9, *C. gracilis*; 10, *C. helveticus*; 11, *C. hominis*; 12, *C. hyointestinalis subsp. hyointestinalis*; 13, *C. hyointestinalis subsp. lawsonii*; 14, *C. insulanigrae*; 15, *C. jejuni subsp. doylei*; 16, *C. jejuni subsp. jejuni*; 17, *C. lanienae*; 18, *C. lari*; 19, *C. mucosalis*; 20, *C. rectus*; 21, *C. showae*; 22, *C. sputorum*; 23, *C. upsaliensis*. Data for the biochemical properties were taken from Rossi et al. 2009. All taxa are negative for aerobic growth at 37 C.

+, 90100% of strains positive; -, 010% of strains positive; (+), 7589% strains positive; (-), 1125% of strains positive; V, 2674% of strains positive; W, weakly positive;

NA, no data available. CCDA, Charcoal cefoperazone deoxycholate agar (Oxoid); TTC, triphenyl tetrazolium chloride (Rossi et al. 2009)

2.2 Epidemiology

Illness caused by pathogenic *Campylobacter* spp. is collectively called campylobacteriosis which is the leading cause of foodborne bacterial gastroenteritis in humans (Altekruse et al. 1999). Since the first isolation of *Campylobacter* from humans during a milk-borne enteritis outbreak in 1938 in the United Sates of America (Levy 1946), the vast majority of human campylobacteriosis cases have been gastrointestinal infections. Campylobacteriosis is most often self-limiting, with the duration of illness lasting anywhere from a few days to up to two weeks. The incubation period is between 2 to 5 days, but may vary from 1 to 11 days. While watery or bloody diarrhoea is the most common symptom of campylobacteriosis in humans, other symptoms include abdominal cramps, fever, myalgia, headache, nausea and vomiting (Black et al. 1988). A small proportion of patients develop reactive arthritis following *Campylobacter* enteritis. Faecal excretion of the bacteria usually continues for two to three weeks post recovery (Skirrow 1994). An infrequent complication is an acute immune-mediated inflammation of the peripheral nerves known as Guillain-Barré syndrome (Nachamkin & Blaser 2000, chap. 2). Guillain-Barré syndrome (GBS) was first described in 1916 as an autoimmune-mediated disorder of the peripheral nervous system. Of those affected, 15% of patients recover completely, 3 to 8% die, and the remaining surviving patients suffer varying degrees of physiological, neurological and/or physical deficits (Smith 1995).

Campylobacteriosis represents a substantial burden to public health in developed countries (WHO 2000a). It has been estimated that 2.4 million cases of *Campylobacter* enteritis occur annually in the United Sates of America accounting for 5% to 7% of all human gastroenteritis cases (Melissa & Michael 2010).

It has been estimated that nearly 1% of the US population suffer from campylobacteriosis per year (WHO 2000a) and these infections result in around 13,000 hospitalisations and 124 deaths each year. In Canada in 2000 more than 2,300 people became infected with *Campylobacter* in Walkerton Ontario following a heavy rainfall event that resulted in bacteriological contamination of the town's water supply (BGOSHU 2000). In 2004 the incidence of campylobacteriosis in Canada increased to 9345 cases per 100,000 (Galanis 2007). Similarly, the incidence of reported human cases of campylobacteriosis in Northern European counries ranged from 60 to 90 cases per 100,000, (it has been estimated to

be substantially increasing in the last 20 years) a substantial increase over the previous 20 years (WHO 2000a).

In the United Kingdom in 1998, there were 58,059 laboratory confirmed cases in England and Wales whereas, during 2000 there were 1,388,772 cases of foodborne infection acquired in England and Wales of which *Campylobacter* accounted for 359,466 of these cases. There were 171,174 presentations to general practice due to *Campylobacter* infection, 16,946 hospital admissions (accounting for 62,701 hospital bed-days) and 86 deaths (Adak et al. 2005). In Denmark, the incidence of disease remained relatively unchanged from 1980 to 1990 (475 cases) (WHO 2000a) and then from 1992 to 1999 there was a three-fold increase (1,676) in incidence risk. In the case of Germany reported cases of campylobacteriosis in 2003 were 58 per 100,000 and this increased to 79 per 100,000 in 2009 (Gallay et al. 2003, ECDC 2009, Fitzenberger et al. 2010). The incidence of *Campylobacter* infection in Australia increased steadily from 1991 through to 2001. From 2001 to 2005 the incidence was relatively stable at 113 cases per 100,000 head of population (approximately 15,000 cases per year) (Stafford et al. 2008). Stafford et al. (2008) commented that *Campylobacter* is likely to be underreported estimating that there are around 223,000 *Campylobacter* infections occurring annually.

Although precise details are not available for all developed countries, the available data reflects a general trend of annual increase in campylobacteriois in developed countries over the past 20 years. This global increase in disease frequency could be explained by one or more of the following factors: changes in diagnostic criteria, improved surveillance and reporting systems, improved diagnostic facilities and public awareness. Besides the overall improvement in the reporting systems, surveillance and diagnostics, the scenario of global increase in campylobacteriosis could be a true increase rather than an artifactual.

New Zealand has one of the highest notification rates of human campylobacteriosis cases among all developed countries (Baker et al. 2007, Marler 2010). From 1998 to 2002 New Zealand's notification rates were twice that of England and three times that of Australia and Canada (Eberhart-Phillips et al. 1997, Lake et al. 2004). The incidence of reported infection in New Zealand increased from 14 cases to 120 cases per 100,000 between 1981 and 1990 and to 363 per 100,000 by 1998 (WHO 2000a). The incidence risk of campylobacteriosis in New Zealand for the period 2005 to 2009 are provided in Table 2.2. It should be noted that although the number of human campylobacteriosis cases decreased

Table 2.2: Number of cases and incidence risk of campylobacteriosis, New Zealand 2005 – 2009. Adapted from Nicol et al. (2010).

Year	Cases	IR [a]	95% CI
2005	13,836	335	329–340
2006	15,873	379	373–384
2007	12,778	302	296–307
2008	6,694	157	153–160
2009	7,176	166	162–169

[a] Incidence risk of campylobacteriosis, expressed as the number of cases per 100,000 head of population.

from 2007 to 2008 (Table 2.2) an increase in 2009 warrants further investigation.

In developing countries, campylobacteriosis has been reported to be hyperendemic however reliable estimates of the number of cases and the size of the population at risk tend not to be routinely recorded (Coker et al. 2002). Furthermore, priotities for surveillance and control of infections of public health importance have been focused more predominantly on diseases such as malaria, tuberculosis, trypanasomiasis, onchocercasis, shistosomiasis, diarrhoea and respiratory infections (WHO 2000a). Despite the increased burden of enteric disease associated with *Campylobacter* spp., the aetiological agent involved in clinical cases is often not identified (WHO 2000a). International reports (WHO 2000a) have cited a lack of suitable laboratory facilities and an absence of formal strategies to diagnose specific conditions (such as campylobacteriosis) as a reason for the absence of reliable disease frequency measures. The World Health Organisation has supported developing countries in the form of research grants for epidemiological studies, and the Public Health Service of Canada have provided Lior antisera for the detection of campylobacteriosis in many developing countries (WHO 2000a). After intervention by the WHO and the Public Health Service of Canada, estimates of the incidence of campylobateriosis have started to be recorded (WHO 2000a). The incidence of campylobacteriosis in developing countries from Africa, America, Eastern Mediterranean, Europe, South East Asia and Western Pacific for the period 1990 to 1995 ranged from 5000 to 20,000 cases per 100,000 head of population in children (Adegbola et al. 1990, Oberhelman & Taylor 2000, Raji et al. 2000, WHO 2000a). Compared with the situation in developing countries, the incidence of campylobacteriosis in developed countries seems negligible. In developing countries estimates of disease frequency in the general population have been

shown to be approximately 90 per 100,000 with varying orders of magnitude. Recovery rates of bacteria decrease with the age of the patient (Oberhelman & Taylor 2000, Raji et al. 2000, WHO 2000a, Coker et al. 2002) which implies the acquisition of immunity after exposure to the organism.

Seasonality of campylobacteriosis

Generally, *Campylobacter* infections have been shown to exhibit well defined spring and summer peaks in developed countries (WHO 2000a). Most campylobacteriosis cases are reported during the summer in the United Sates of America and other European countries (Tauxe 1992, Friedman et al. 2000, Vereen et al. 2007). Data recorded from 28 countries by European Centre for Disease Prevention and Control have shown that campylobacteriosis cases were mostly reported in the summer months between June and August (Figure 2.1, ECDC 2009).

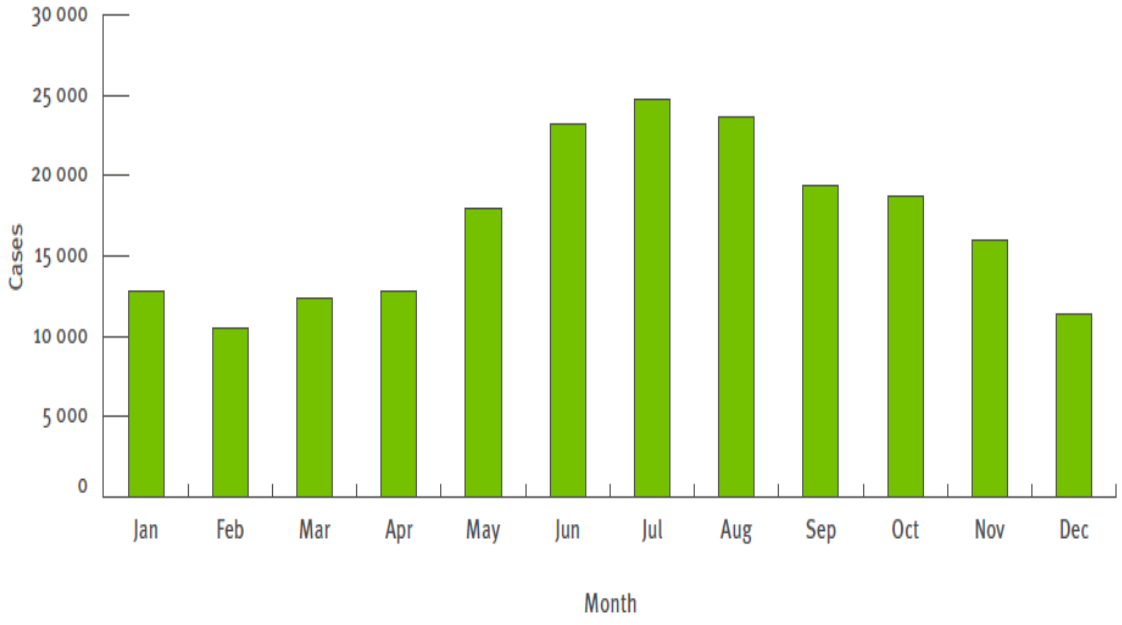


Figure 2.1: Seasonal distribution of human campylobacteriosis cases in EU and EEA / EFTA countries, 2007

Seasonal patterns vary among countries (Vereen et al. 2007, Louis et al. 2005, Nylen et al. 2002). For example, in Australia the temporal pattern of *Campylobacter* infections vary among temperate and sub-tropical regions (Bi et al. 2008). The weekly maximum and minimum temperatures were inversely associated with the weekly number of *Campy-*

lobacter cases in temperate regions but positively correlated with the number of cases in sub-tropical areas. Further, the seasonal distribution of *Campylobacter* infection in nine European countries and New Zealand (Nylen et al. 2002) shows no prominent seasonal pattern in Scotland and Austria. In contrast, the seasonal pattern was more prominent in Finland (more cases in summer) and all of the other countries showed consistent seasonal summer peaks. In New Zealand seasonality was less consistent (Nylen et al. 2002). Three broad groupings of seasonal campylobacteriosis patterns exist in New Zealand (Hearnden et al. 2003). Firstly, there is a marked difference between the North and South Islands. Rural areas in the North Island have a relatively low summer incidence and small inter-seasonal variation. Secondly, North Island urban areas (Auckland, Hamilton, Napier and their hinterlands) and some areas of the South Island have a higher summer incidence and more seasonality. Thirdly Christchurch, Dunedin, much of the South Island and cities such as Wellington and Upper Hutt experience a high summer incidence and strong inter-seasonal variation.

In developing countries, campylobacteriosis has not been shown to be seasonal in contrast to the defined seasonality evident in developed countries (Altekruse et al. 1999). This may reflect a lack of adequate recording or surveillance programs in developing countries. Peaks in the number of confirmed *Campylobacter* cases have been identified during the dry season in Mexico, Nigeria, Peru, Thailand and India (Varanasi) (Coker & Dosunmu-Ogunbi 1985, Taylor et al. 1992, WHO 2000a). Rainy season peaks have occurred in the Central African Republic, India (Calcutta) and Egypt (Coker & Dosunmu-Ogunbi 1985, Taylor et al. 1992). No significant seasonal variation has been observed in Egypt (Taylor et al. 1992).

Mechanisms behind the seasonality of campylobacteriosis remain unclear. The prevalence of *Campylobacter* in animal reservoirs is thought to be influenced by excretion rates from carrier animals, survival of the bacteria in the environment and the influence of climate or season on these factors (Hearnden et al. 2003, Kovats et al. 2005). The commonly known risk factors for acquiring the disease include: animal contact (related to farm practices and/or pet ownership), season, consumption of barbecued meals that are not properly cooked, recreational swimming and drinking water from streams or other natural water sources (Kovats et al. 2005). Besides increased exposure, the seasonal carriage of campylobacters by poultry has been shown to contribute significantly to seasonal

disease patterns (Kovats et al. 2005, Bi et al. 2008). It has been documented that the carriage rate of campylobacters was 100% in broiler flocks in summer and 50% in winter (Jacobs-Reitsma et al. 1994). The drinking water provided for broiler flocks was an important environmental factor influencing the colonisation of *Campylobacter* spp. in broiler flocks (Pearson et al. 1993, Ogden et al. 2007), while the treatment of water has been shown to significantly decrease the prevalence of *Campylobacter* in broilers (Hansson et al. 2007). In the United Kingdom and Sweden there was an inverse relationship betwen carriage rates of *Campylobacter* spp. in poultry and the incidence of human cases (Humphrey et al. 1993, Berndtson & Engvall 1994). Few studies have been carried out to determine the relationship between the sources and seasonality of campylobacteriosis in humans. No convincing co-variation between the prevalence of campylobacter in animal faeces (Stanley et al. 1998*a*,*b*) and human disease have been identified. A similar lack of co-variation has been identified with drinking water (Carter et al. 1987, Jones et al. 1990).

Risk factors for campylobacteriosis

Campylobacter infects humans of all ages. In the European Union countries, United Sates of America, Australia and New Zealand (WHO 2000a, Gillespie et al. 2009, 2008, Unicomb et al. 2008, Eberhart-Phillips et al. 1997) the disease shows a distinctive bimodal age distribution, affecting particularly children less than 4 years and young adults. Campylobacteriosis is associated with an increased risk of mortality in children and time lost from work and school (Sneyd & Baker 2003). Further, there is a distinctive distribution between males and females where, in males, the incidence is 1.2 to 1.5 times higher than females, particularly among young adults (WHO 2000a). The incidence of campylobacteriosis by age and gender was evaluated by Gillespie and The *Campylobacter* Sentinel Surveillance Scheme Collaborators (Gillespie et al. 2008). These authors showed differences among ethnic groups in England and Wales during May 2000 to April 2003. In New Zealand in 2002 the incidence of campylobacteriosis in 0 – 4 year olds and young adults (20 to 29 years of age) was 599 (95% CI: 566 to 631) and (484, 95% CI: 465 to 504) per 100,000, respectively (Sneyd & Baker 2003). For the 0 – 4 year old age group this pattern is thought to correspond to the time when children are weaned onto a completely solid food diet, while in the case of young adults, disease risk is thought to be associated with young adults setting up houses on their own and preparation of their own meals (Shane

1992).

The age dependent risk of campylobacteriosis in developing countries is almost similar to developed countries (WHO 2000a). Studies from different developing countries showed that *Campylobacter* was most commonly isolated from children with diarrhoea under 2 years of age and the incidence of disease decreased with advancing age (WHO 2000a). It has also been shown that recovery of *Campylobacter* is common in healthy children from countries such as India, Algeria, Northern Thailand, Nigeria and Guinea-Bissau (WHO 2000a). In developing countries gender associated infection rates are not as prominent as they are in developed countries (Coker et al. 2002). The age and gender associated infection rates may be attributed to predisposing factors such as the immune status of young children, general sanitation, occupation (particularly in males), recreational activities and meat consumption (particularly in developed countries).

2.2.1 Sources of *Campylobacter*

Raw milk and water were considered to be the main sources for human campylobacteriosis until 1990. A series of unresolved outbreaks between 1998 and 2004 led to the search for additional transmission routes and sources apart from milk and water (ESR 1993, 1996, CDC 2000, DEFRA 2000). As a result, the importance of food chains particularly those involving poultry were brought to light (Nachamkin & Blaser 2000, chap. 2). *Campylobacter* spp. has now been known to colonise the intestinal mucosa of a wide range of avian and animal hosts, including humans (Newell 2002). *Campylobacter* spp. are commensal organisms in birds. Livestock can carry the organism as asymptomatic carriers as do humans in endemically infected regions (Newell 2002), where wild rodents have been also identified as reservoirs of *Campylobacter* spp. (Williams et al. 2010)

Notably, *C. jejuni* has been identified to be a successful commensal in chickens, cattle and sheep (Newell 2002, Devane et al. 2005) while its sister species *C. coli* is more associated with pigs and sheep (Nesbakken et al. 2002, Brown et al. 2004). *C. fetus subsp. fetus*, a cause of genital campylobacteriosis in sheep and cattle, can also cause disease in humans (Allerberger et al. 1991, Krause et al. 2002, Cone et al. 2003, Herve et al. 2004, Monno et al. 2004). *C. hyointestinalis* and *C. mucosalis* in pigs (Minet et al. 1988, Gorkiewicz et al. 2002) and *C. upsaliensis* and *C. helviticus* in cats and dogs (Moreno et al. 1993, Hald

& Madsen 1997, Hald & Brondsted 2000, Moser et al. 2001, Gow et al. 2009, Chaban et al. 2010) are other identified important pathogens. Human sewage is also an established source of *Campylobacter* spp. in environmental as well as drinking water (Jones 2001).

2.2.2 *Campylobacter* spp. in wild birds

The optimum temperature for thermophilic *Campylobacter* spp. corresponds to the body temperature of birds more so than mammals. For this reason *Campylobacter* spp. are well adapted as commensals of the avian gut, and therefore birds have been widely regarded as natural hosts of these organisms (Lee & Newell 2006). *Campylobacter* spp. have been found in a variety of bird species, both domesticated and wild. Among domesticated birds, a high prevalence of *C. jejuni* and *C. coli* is often found in broiler chickens, breeder flocks and egg-laying hens (Cox et al. 2000, Hansson et al. 2004).

Campylobacter spp. have been isolated from a wide range of wild bird species (Luechtefeld et al. 1980, Fricker & Metcalfe 1984). The carriage rates vary from 0% to 100% (Colles et al. 2008), with differences arising from the ecology of bird species, feeding habits, habitat preferences and migration patterns. Differences could also arise from variations in the study regimen, sample size and the sensitivity of culture and detection methods. For example, studies conducted in Colorado, USA, reported that 35% of 445 migratory waterfowl carried *C. fetus subsp. jejuni* and the prevalence varied between species (Luechtefeld et al. 1980). In another study conducted in Alabama, United Sates of America, Oyarzabal et al. (1995) reported a *Campylobacter* spp. prevalence of 19% among 66 necropsied birds. In these two studies, particularly the study of Oyarzabal et al. (1995) the small sample size influences the precision of the prevalence estimate which, in turn, makes it difficult to distinguish differences that are artefactual or real.

A survey comprised of 1,794 individual migrating birds from 107 species in south-eastern Sweden found that the distribution of *Campylobacter* spp. among species was uneven, and that the feeding behaviour of the birds influenced colonisation rate (Waldenstrom et al. 2002). Although wild birds were found to carry *C. jejuni* and *C. coli* predominantly, a substantial proportion of isolates were identified as *C. lari* (Waldenstrom et al. 2002). Colonisation of *Campylobacter* spp. has been shown to be common in turkeys (Wallace et al. 1997, Borck 2003), geese (Aydin et al. 2001), ducks (Savill et al. 2003), ostriches,

quails, pigeons, waterfowl and parrots (Oyarzabal et al. 1995, Hughes et al. 2009). European starlings have a high carriage rate of *Campylobacter* spp. and have been identified to cause heavy faecal contamination in towns and park lands acting as a potential source of *Campylobacter* spp. in childrens' play areas (Waldenstrom et al. 2002, Colles et al. 2008, French et al. 2009a).

Environment and environmental water are considered potential sources while environmental contamination is likely to originate from faecal contamination by domestic livestock, wild mammals, wild birds and humans (Jones 2001). Wild birds such as geese have been implicated in outbreaks of campylobacteriosis arising from water supply contamination in Norway (Varslot et al. 1996, Colles et al. 2008) and similarly in northwest England where mallard ducks were reported to be an important source of river contamination (Obiri-Danso & Jones 1999). However, the role of wild birds in the causation of human campylobacteriosis has not been well explained except for the above-mentioned studies by Varslot et al. (1996), Obiri-Danso & Jones (1999).

2.2.3 *Campylobacter* spp. in pets

Pet ownership is a risk factor for zoonotic disease in humans, including campylobacteriosis (Skirrow & Benjamin 1980, Fernandez & Martin 1991, Damborg et al. 2004, Bender et al. 2005). Cats and dogs, both healthy and sick pets with diarrhoea, are frequent carriers of *Campylobacter* spp. (Moreno et al. 1993, Hald & Madsen 1997, Hald & Brondsted 2000, Gow et al. 2009). *C. upsaliensis* is the most frequently isolated species in cats and dogs although *C. jejuni*, *C. coli* and *C. helveticus* account for a substantial proportion of isolates (Stanley et al. 1992, Chaban et al. 2010). Carriage rates of *Campylobacter* spp. in healthy dogs and those with diarrhoea presented to Norwegian veterinarians were found to be 23% of 529 and 27% of 66, respectively. Carriage rates in cats ranged from 18% in 301 healthy cats and 16% in 31 cats with diarrhoea (Sandberg et al. 2002). No major differences in the prevalence of *Campylobacter* spp. between cats and dogs with and without diarrhoea were identified (Sandberg et al. 2002) which differs from findings reported in other studies that claim that an association between disease status and carriage of *Campylobacter* spp. exists (Nair et al. 1985, Burnens et al. 1992).

A prevalence study for *Campylobacter* spp. in cats and dogs was carried out in two ani-

mal shelters in Ireland (Acke et al. 2009). Prevalence did not significantly vary between healthy (87%) and diarrhoeic animals (86%) (Acke et al. 2009), while for both species prevalence was higher in animals less than one year of age compared with adults. Studies conducted in The Netherlands and Sweden identified simultaneous presence of multiple species of *Campylobacter* from dogs (Engvall et al. 2003, Koene et al. 2004, Acke et al. 2010). The species richness and the prevalence of *Campylobacter* spp. in healthy and diarrhoeic dogs was evaluated in Saskatoon, Canada (Chaban et al. 2010). *Campylobacter* spp. were detected in 56% (39 of 70) of healthy dogs and 97% (63 of 65) of diarrhoeic dogs. In the assay for species detection, diarrhoeic samples were positive for 11 of 14 *Campylobacter* species. The prevalence of *C. curvus*, *C. hyointestinalis* and *C. rectus* was constant between populations. In diarrhoeic dogs carriage rates for *C.upsaliensis* (85%, 55 of 65), *C. jejuni* (46%, 30 of 65) and *C. showae* (28%, 18 of 65) were high compared with healthy dogs where carriage rates for these species ranged between 6% and 7%. In contrast, *C. coli* was undetectable in the healthy dog population (0 of 70) whereas it was relatively high (25%, 16 of 65) in dogs with diarrhoea.

2.3 Molecular epidemiology

Three major factors have been identified as obstacles in tracing the sources of campylobacteriosis including the magnitude of the problem, pathways of infection and the population structure of *Campylobacter* spp. (Nachamkin & Blaser 2000, chap. 2). Sub-typing every campylobacter strain is not possible due to the time and expense involved and also due to the ubiquitous nature of the organism which comprises part of the normal intestinal microflora of almost all terrestrial species (Nachamkin & Blaser 2000, chap. 2).

With the advent of modern genotyping techniques, studies on source attribution have been carried out on foodborne campylobacteriosis cases. Since there is considerable genetic variation within *C. jejuni* populations in different host species (McCarthy et al. 2007), analytical epidemiological methods, such as case-control studies and risk assessments have not been successful. The recent advancements in genotyping techniques and typing schemes have been successfully employed to define host and host-associated infections (Sheppard et al. 2009). Source attribution modeling has also been conducted by several research groups (described below) and have provided detailed information on sources of

Campylobacter with special emphasis on *C. jejuni*.

A systematic model-based approach developed using details of 1,231 cases of *C. jejuni* infection in humans in Lancashire, England used multilocus sequence typing (MLST) to infer the source of infection of each patient by comparison with 1,145 animal and environmental *C. jejuni* isolates (Wilson et al. 2008). Wilson et al. (2008) found that the vast majority of human cases were attributable to *C. jejuni* carried by livestock, particularly poultry, as opposed to *C. jejuni* isolates from wild animals and the environment (Wilson et al. 2008). A longitudinal study conducted from 2005 to 2008 in the Manawatu region of New Zealand developed a source attribution model using *C. jejuni* as a prototype and MLST as a primary typing tool. In that study the majority of human clinical cases were attributable to poultry (French 2008, NZFSA 2008, Mullner et al. 2009), followed by cattle and sheep with a relatively small contribution from wild birds and water. The genotypes (also known as sequence types, STs) that are found in water as well as other environmental sources (ST-45, ST-137, ST- 583) have been reported from human clinical cases in New Zealand and in other developed countries. Amongst the STs, ST-45 was the most widely identified genotype among human cases and ST-137 and ST-177 that were isolated from water and wild birds have also been isolated from sporadic cases in humans globally.[2]

Sheppard et al. (2009) surveyed 15 health board regions in Scotland, to identify sources of confirmed human campylobacteriosis cases. This survey was conducted using MLST data collected for a period of 15 months from mid July 2005 through to mid October 2006. MLST genotypes of confirmed human clinical cases found that chickens were the dominant source for the majority of human cases followed by cattle and sheep, with the contribution from wild birds and the environment being relatively low. Sheppard et al. (2007) compared 379 human isolates in the United Kingdom with chicken and bovid isolates in an effort to determine the sources of human infection. The clonal complex ST-61 which is predominant in cattle was very rare in poultry and was uncommon among human isolates. ST-443, ST-574 and ST-353 complexes were found in poultry and in human clinical cases but not in bovine samples. In the United Kingdom Sopwith et al. (2006) studied the seasonality of campylobacteriosis using MLST techniques. These authors found that recovery rates of ST-45 and ST-21 increased whenever there was an increase in

[2]URL: (`http://pubmlst.org/Campylobacter`)

the number of human campylobacter cases. McCarthy et al. (2007), in a study comparing genotypes of *C. jejuni* recovered from chickens and cattle throughout the United Kingdom before 1997 and from 1998 to 2003 found a strong association between genotype and host species. In contrast, the number and composition of genotypes was not consistent over time and space.

Amplified fragment length polymorphism (AFLP) was used in a case report that investigated *C. jejuni* neonatal infection in a 3 week old girl with diarrhoea in The Netherlands (Wolfs et al. 2001). This confirmed the acquisition of *C. jejuni* from a newly arrived puppy. *C. jejuni* isolates from cats and dogs were sub-typed by *fla* and pulsed-field gel electrophoresis (PFGE) typing techniques in Ireland (Acke et al. 2010). A link between the isolates of *C. jejuni* in cats and dogs that shared a common environment was ruled out by the PFGE patterns which, in turn, indicating the possibility of interspecies transmission (Acke et al. 2010).

The number of studies using molecular epidemiological techniques has steadily increased since enzymatic, electrophoretic and other modern analytical and typing techniques have become widespread. Even though there has been enormous growth in the use of these methods in developed countries they are still in their infancy in developing countries (WHO 2000a,b). The real value of these techniques and their application will only be appreciated when they are able to be extended across countries and sources to better reflect the biology of *Campylobacter* spp. With the migration of people from one country to another, the generalisation of the techniques and the database would aid in better understanding of evolution of *Campylobacter* spp. and the gene flow from one region to another.

2.4 Typing techniques

Classification of bacterial strains at the species or subspecies level are generally known as bacterial typing or subtyping systems. The main purposes of bacterial subtyping are to evaluate taxonomy, define phylogenetic relationships, examine evolutionary mechanisms and to conduct epidemiological investigations (van Belkum et al. 2001). A plethora of typing techniques have been developed over recent years and can be broadly classified into two major categories: phenotyping and genotyping. Typing of infectious pathogens

was initially done based on phenotypic characteristics such as growth, morphology, biochemical, serological and functional properties. Genotyping came into existence with the advent of restriction enzymes, electrophoretic techniques and DNA sequencing (Riley 2004). Each one of these techniques has its own advantages and disadvantages and no single technique has been declared as universally acceptable and applicable (Sails et al. 2003b). Efficacy and efficiency are the two major properties that any typing system should possess to be adapted for further routine use (ECDC 2009). Efficacy of any typing technique can be assessed in terms of typeability, reproducibility, consistency and power of discrimination. Efficiency reflects the expertise required, time consumed or rapidity of the technique, flexibility and suitability to carry out a certain investigation (ECDC 2009).

Campylobacteriosis is most often sporadic in nature (WHO 2000a) and hence typing is done for tracing the source of infection to characterise the strain of an outbreak and, in some cases, for retrospective epidemiological investigations (Dingle et al. 2002). Typing and characterisation of source strain are most important to assess the degree of public health intervention required to design effective control measures at appropriate times. There are many typing methods performed for *Campylobacter* spp. where some are designed to identify the differences in their phenotypic characteristics whereas others are applied to genetic diversity. The methods that are in current and widespread use (only) are described in this review.

2.4.1 Phenotyping

Phenotyping of *Campylobacter* includes biotyping, serotyping and phage typing (Fitzgerald et al. 2001). Subtype or strain characterisation using phenotyping techniques becomes difficult and often ambiguous due to various reasons such as the lack of specific antisera when serotyping is employed; when there is unavailability of standard reagents; due to the presence of cross-reactivity between strains; and because of the emergence of high proportions of non-typeable strains (Jackson et al. 1996). Some of the widely used phenotyping techniques are described below.

Biotyping

A biotyping scheme was first developed utilising 12 biochemical tests for *Campylobacter* spp. that included growth at 28^{o}C, hippurate analysis and resistotyping tests (using antimicrobials or antibacterials) (Bolton et al. 1984). Generally the biotyping technique assesses the ability of the organism to utilise biochemical substrates and to grow in a difficult environment, for example in the presence of an antibiotic (Klena 2001). Although these methods are advantageous in terms of their ease of use, time, and interpretation, the inferences drawn from these tests are found to be too general for subtyping *Campylobacter* (Struelens & Members of European Study Group on Epidemiological Markers (ESGEM) of the European Society for Clinical Microbiology & Infectious Diseases (ESCMID) 1996). The reproducibility and stability of these methods are not very good and have low discriminatory power and hence are often used with serotyping to make the scheme more useful (Sails et al. 2003b). However, antibiotic sensitivity testing (or antibacterial resistance testing) has outgrown other biotyping techniques in recent years and has been established as a ideal typing tool in its own right.

Antibiotic resistance typing is often referred to as resistotyping. This involves testing an organism to examine its sensitivity or resistance to selected antibiotics (Klena 2001). The organism that shows resistance to antibiotics are known as resistotypes determined using agar dilution (Bolton et al. 1984, Huysmans & Turnidge 1997) or disc diffusion methods (Lior 1984). This is a component of biotyping that measures the phenotypic trait expressed by the organism for which there may be numerous reasons such as a mutation in a gene that codes for antibiotic sensitivity or resistance. For example, mutations in the *Campylobacter gyr*A gene have been shown to confer antibiotic resistance (Nachamkin & Blaser 2000, chap. 2).

A cross-sectional study was conducted in Nebraska, United Sates of America to determine the microbial profile and antibiotic susceptibility for *Campylobacter* spp. (Sanchez et al. 2002). This study evaluated the effect of immersion chilling and air chilling on microbial load on post-processed chicken carcases and found that immersion-chilled broilers had a higher incidence of *Campylobacter* spp. resistance to nalidixic acid (NAL) and other related fluoroquinolones, compared with isolates from air-chilled broilers. Similar studies on antibacterial and antimicrobial profiles have been conducted in various coun-

tries such as The Netherlands (Oza et al. 2003), United Sates of America (Luangtongkum et al. 2007), Trinidad (Rodrigo et al. 2007), India and Iran (Baserisalehi et al. 2007) and Ethiopia (Dadi & Asrat 2009).

The most frequent pattern of antibiotic resistance, multi-drug resistance (MDR) in *Campylobacter* spp. from imported chicken and human clinical cases was determined in a cross-sectional study conducted in South Korea (Ku et al. 2011). Ku et al. (2011) found a pattern of MDR resistance in *Campylobacter* spp. to four antimicrobials: ciprofloxacin, nalidixic acid, ampicillin, and tetracycline. These findings indicate the extent of MDR among *Campylobacter* spp. in Korea; it is plausible that similar patterns might also exist in other countries.

Serotyping

Serotyping of *Campylobacter* was initially based on the heat stable antigens (O) first described by Penner & Hennessy in 1980. This was later adopted as the gold standard for typing *Campylobacter* beyond the species level (Penner & Hennessy 1980). Subsequently, there was another typing scheme developed by Lior et al. based on heat labile antigens (Lior et al. 1982). Of these two schemes, the Penner typing scheme is the most frequently used technique in laboratories worldwide and has undergone further development with 66 different antisera being used for both *C. jejuni* and *C. coli* typing (McKay et al. 2001). Although Penner serotyping was considered the gold standard, the exact nature of the serotyping antigen was not known at the time the technique was first developed (Moran & Penner 1999). Later it was discovered that the capsular polysaccharide (CPS) was the contributory molecule for the serological reactions. Generally *C. jejuni* produces two different polysaccharide molecules; high molecular weight lipo-polisaccharide (LPS) and low molecular weight lipo-oligosaccharide (LOS) that contribute to the serological reactions (Moran & Penner 1999). Karyshev and his co-scientists discovered that genes in the CPS region contain intragenic homopolymeric tracts, which are likely to render their expression phase variable enabling rapid antigen variation of the capsular polysaccharide (Frost et al. 1998, Karlyshev et al. 2000). Therefore the serological typing is thought to suffer from limitations such as lack of discrimination and cross reactivity (Frost et al. 1998, Karlyshev et al. 2000). Often Penner serotyping is used in conjunction with other methods due to its low discriminatory power and also due to the lack of antisera stan-

dardisation. In these circumstances new serotypes remain untyped (Frost et al. 1998, Wassenaar & Newell 2000).

Phage typing

Due to the low discriminatory power of serotyping, phagetyping was initially performed to characterise *C. jejuni* and *C. coli* using 46 phage types (Grajewski et al. 1985, Khakhria & Lior 1992). Phagetyping is often used as an adjunct to serotyping and about 76 defined phage-types have been defined to date (Nachamkin & Blaser 2000, chap. 2). Briefly, the technique ultilises a set of virulent phages on a bacterial host irrespective of any receptors for attachment. If the phages are capable of attaching and infecting the bacterial hosts, they lyse the bacterial cells producing a characteristic lytic pattern on the cultured petri dishes, referred to as ‘plaques (Grajewski et al. 1985). A phage type is defined as two or more epidemiologically unrelated isolates giving the same phage reaction pattern (Frost et al. 1999, Nachamkin & Blaser 2000). Like serotyping, the usefulness of phagetyping is also limited by the occurrence of non-typeable isolates and problems with cross reactivity (Sails et al. 2003b). This technique is labour intensive and expensive rendering it unsuitable for most clinical laboratories.

Multi-locus enzyme electrophoresis

Multilocus enzyme electrophoresis (MLEE) has been used as the standard typing technique for eukaryotic population genetic studies (Ayala 1976) and was subsequently adapted to a variety of bacterial species to assess population diversity and the structure of bacterial populations including *C. jejuni* (Selander et al. 1986). This technique exploits the relative electrophoretic mobilities of a large number of water soluble cellular enzymes (Selander et al. 1986). The rate of migration of a given protein in an electric field is dependent on amino acid sequences. The mobility variants of an enzyme is then directly equated with the alleles at the corresponding structural gene locus. Studies have shown that electrophoresis can detect large proportions of amino acid substitutions, however some silent substitutions may not be evident phenotypically and may not be discriminatory (Selander et al. 1986). Multiple enzymes encoded by housekeeping genes are analysed simultaneously by MLEE (Maiden et al. 1998, Sails et al. 2003b). MLEE studies have been

performed to determine the clonal framework of *C. jejuni* (Meinersmann et al. 2002). This has provided insight into the nature of genome re-assortment and random exchange of DNA segments that contribute to the genetic diversity of *C. jejuni*. A number of clonal groups have also been reported within the *C. jejuni* population (Meinersmann et al. 2002). MLEE has also been utilised to study the congruence between other typing schemes used for *C. jejuni* (such as multilocus sequence typing [MLST] and pulse field gel electrophoresis [PFGE]) (Sails et al. 2003b). However, MLEE does suffer from a number of limitations: (1) it examines the electrophoretic mobility of enzymes rather than indexing the molecular source of variation and hence may not be comparable between laboratories; (2) maintenance of live cultures may be time consuming and costly and mutations may occur if isolates are subcultured or stored for long periods of time; (3) this technique is time consuming, expensive and requires a high level of technical expertise (Sails et al. 2003b). All of these factors have rendered MLEE unsuitable for regular typing. MLEE has been superseded by a nucleotide-based technique, MLST which essentially mimics the MLEE's multi loci principle.

2.4.2 Genotyping

Phenotypic traits form the basis of phenotyping while genes responsible for the production of those phenotypic characters form the foundation for genotyping. Genotyping methods measure differences in parts of the genome that are relatively stable (Wassenaar & Newell 2000). Detection of DNA variations to compare nucleotide sequences was not a practical possibility until the immense progress in DNA-based technologies over the last decade has made this technique more commonplace (Duim et al. 2000, 2003, Newell et al. 2000, Dingle et al. 2005). Often the emergence of new genotyping technologies takes place as a result of modification of an existing one (Duim et al. 2000). Most of them are computer assisted which makes analysis, interpretation and data sharing between laboratories easier (Duim et al. 2000). Even though various generations of technologies emerge (one succeeding the other) highly reliable and sensitive techniques are of paramount importance and the techniques that become widely accepted form the central part of molecular diagnostics (Duim et al. 2003). Molecular typing techniques are in two broad categories: (1) macro-restriction mediated analyses, based on separation of restriction enzyme digested nucleotide sequences; and (2) polymerase chain reaction (PCR) based assays, as

described below.

2.4.3 Macro-restriction mediated analysis

With the advent of restriction enzymes in the late 1960s, the first cleavage-site-specific restriction endonuclease (type II restriction enzyme) from a *Haemophilus influenzae* strain was discovered. This was shown to cleave DNA at specific sites (Smith & Wilcox 1970, Danna & Nathans 1971) allowing a more robust typing system with these enzymes to be developed (Ayala 1976).

Pulse field gel electrophoresis

Pulse field gel electrophoresis, also known as field alteration gel electrophoresis (FAGE) or macrorestriction profiling was originally developed to separate yeast chromosomal DNA (Carle & Olson 1984). This technique is an evolution of restriction enzyme analysis and gel electrophoresis that arose in response to difficulty in mobilising large DNA molecules through agarose gels (Carle & Olson 1984, Dawkins 1989). The inability of the large DNA molecules to move through agarose gel produced a zigzag movement of the larger DNA molecules and this was overcome by pulsing the electrical fields to form different orientations around the gel at given time intervals. This application eventually solved the problem of getting a straight line passage of large DNA molecules (Dawkins 1989, Townsend & Dawkins 1993). PFGE employs rare cutting restriction enzymes that digest the chromosomal DNA resulting in five to fifteen DNA fragments (ranging from 1 to 1000 kb pairs) depending on the chromosome and restriction enzymes used (Tenover et al. 1995, Wassenaar et al. 1998). The resultant digested DNA is electrophoresed in a pulse field within an agarose gel matrix to separate the fragments depending on size. PFGE is considered the 'gold standard' for epidemiological investigations due to its enormous discriminatory power (Sails et al. 2003b).

The reason this technique is unsuitable as a tool for routine use during outbreak investigations is due to its sensitivity to small amounts of nucleotide variations which eventually results in more complex restriction patterns (Sails et al. 2003b, Wassenaar et al. 1998). As a result the true relationship between strains can become obscure (Sails et al. 2003b). A similar finding was reported by de Boer et al. in The Netherlands in 2002. These authors

found that PFGE was not able to determine the genetic relatedness among the laboratory induced recombinant *C. jejuni* strains (de Boer et al. 2002). Similarly, a water-borne outbreak of *C. jejuni* in Canada was investigated using PFGE. This resulted in different banding patterns as there was an insertion of a 40 kb MU-like prophage (Barton et al. 2007). In addition, PFGE involves cumbersome procedures and it is highly time consuming to cast an agarose gel and, in some instances, deactivation of the DNAases is essential to get proper DNA digestion (Gibson et al. 1994). The Centers for Disease Control and Prevention in the United Sates of America introduced an initiative called PulseNet[3] to overcome the shortcomings of PFGE such as protocol differences and interlaboratory profile comparisons. Even then, due to uncompromising protocols that are labour intensive and requiring a high level of expertise, this technique has not been widely adopted in most countries (Sails et al. 2003b). However, PulseNet has been using PFGE extensively for characterising other bacterial species such as *E. coli*:O157, *Shigella*, *Listeria* as well as *Campylobacter* (Swaminathan et al. 2001). Despite these pitfalls PFGE remains a powerful tool to detect micro-evolution in bacterial species that may be indistinguishable by MLST or MLEE (Swaminathan et al. 2001). PFGE has been extensively used in genetic and epidemiological investigations of *C. jejuni* and *C. coli*. PFGE has also been applied in various fields including genomic mapping and sizing, exploration of diversity in sporadic infections, population structure studies and determination of the source of infection (Taylor et al. 1992, Owen et al. 1985, Petersen et al. 2001, Slader et al. 2002)

2.4.4 Polymerase chain reaction based assays

Polymerase chain reaction

The polymerase chain reaction (PCR) has certainly revolutionised molecular epidemiological studies with its versatility and ability to detect the presence or absence of an organism by detecting a gene of interest unique to the particular organism of interest (Oyarzabal et al. 1997). PCR is widely used to distinguish between isolates while it does not require an isolate to be cultured in pure or even to be a live culture (Wilson et al. 1990, Sails et al. 1998). Monoplex PCR assays were widely used for detection and differential diagnosis of *Campylobacter* spp. (Stucki et al. 1995, Linton et al. 1997). It has been replaced by

[3]URL:(`http://www.cdc.gov/`)

multiplex PCR assays which are used for simultaneous differentiation of *Campylobacter* spp. (Asakura et al. 2008, Yamazaki-Matsune et al. 2007). Apart from PCR being used as a diagnostic tool itself, most of the gentoyping techniques are PCR based which are simple, rapid and cost effective (Asakura et al. 2008). PCR methods also possess certain limitations such as reproducibility (in some instances) and inconsistencies in amplification which may be attributed to technical problems in the thermal cyclers such as ramp speed, cooling temperatures, maintenance of uniform temperature in the heating blocks, cycling conditions and biological issues such as purity of the target DNA, reagents and salt concentrations in the reaction mix.

Random amplified polymorphic DNA

This is a PCR-based molecular finger printing technique known as RAPD-PCR. RAPD-PCR employs arbitrary random primers that are ≈ 10 mer in length that bind to several regions over the chromosome generating random numbers of amplicons depending upon the source of DNA (Park & Kohel 1994, Meinersmann et al. 2002). The technique can be optimised to control the number of amplicons produced, by standardising the magnesium chloride concentration and annealing temperature to achieve more consistent results (Park & Kohel 1994). RAPD-PCR has been successfully employed for the characterisation of *C. jejuni* isolates from GBS and Miller-Fisher syndrome (MFS) patients (Hernandez et al. 1995, Endtz et al. 2000). This technique does not require a prior knowledge of the target DNA sequence as the random primers browse through the whole genome to generate amplified fragments from the target DNA. This provides good discriminatory power as well as making the technique cheaper and faster than PFGE (Meinersmann et al. 2002, Nielsen et al. 2006).

Despite the above-mentioned advantages, RAPD has certain limitations such as lack of standardisation, interlaboratory differences and difficulties in profile interpretations especially when they become complex and if the bands were weak. Moreover, techniques that are dependent on the electrophoretic banding patterns do not provide any information on the molecular variations or relatedness between strains (Leonard et al. 2003). Due to lack of consistency and discrepancies in laboratory protocols, RAPD has not been successfully used as a routine genotyping tool (Meunier & Grimont 1993).

Amplified length polymorphism

AFLP involves digestion of DNA with restriction enzymes and subsequent amplification to generate desired DNA products. Generally, two restriction enzymes are used to digest the target DNA and the restriction sites are ligated to specific adapters, where adapters provide the binding sites for primers and the DNA is subsequently PCR amplified. Primers are designed with fluorescent labels and are generally a few nucleotides longer extending beyond the restriction sites for hybridisation (Vos et al. 1995, Duim et al. 2000). AFLP has also been automated for the purpose of frequent usage as an epidemiological investigation tool in which an automated fluorescent DNA seququencer is used to deduce the nucleotide sequences (Vos et al. 1995). About 50 – 100 bands are generated under stringent PCR conditions which are discrimated based on the length of sequences. With the advent of automated sequencers, this technique has proven to be a high throughput tool for epidemiological investigations for *C. jejuni* and other bacterial species (Harrington et al. 1997). Although PFGE, RAPD and AFLP provide comparable levels of discrimination covering the entire genome, the exact basis of variation between strains still remains unclear. The complexity of the technique and equipment expenses have limited the application of the technique as a routine epidemiological tool (Newell et al. 2000).

Flagellin typing

Campylobacter has one or two polar flagella which impart motility and are one of the major immuno-dominant antigens (Wenman et al. 1985). Cloning analysis of the flagellar locus showed that the flagella are composed of many structural flagellin protein subunits encoded by highly homologous flagellin genes, the *fla*A and *fla*B which are separated by approximately 170 nucleotides (Logan et al. 1989, Guerry et al. 1990). The *fla*A and the *fla*B genes exhibit 92% homology and exhibit approximately 95% nucleotide variation among various isolates. This has provided the basis for the formation of the flagellin typing schemes

Because of the presence of both a short variable and a highly conserved region, standardisation of the flagellin typing technique is difficult. This means that flagellin type patterns using RFLP technique can be ambiguous and difficult to interpret. To overcome this, an

exclusive short variable region typing (SVR) scheme was developed based on sequencing (Wassenaar & Newell 2000). This was found to be useful in differentiating an outbreak strain from a sporadic strain when used in combination with MLST (Fitzgerald et al. 2001). In contrast, as a single typing technique, SVR cannot distinguish between *C. jejuni* and *C. coli*, because of the gene pools that are shared by these two species (Fitzgerald et al. 2001, Dingle et al. 2005). The level of discrimination for *fla*A SVR typing is greater than serotyping, but less than PFGE (Ribot et al. 2001) and it is often used in combination with other typing techniques mostly MLST (Dingle et al. 2005).

Real-time PCR assays and single nucleotide polymorphism profiling

In most instances single nucleotide polymorphism profiling (SNP) assays involve real-time PCR as part of the assay and, for this reason, these two typing techniques are being reviewed together.

A real-time PCR involves continuous monitoring of reactions, generally assisted by a computer attached to the thermal cycler (Higuchi et al. 1992). Monoplex real-time PCRs are generally used to detect *Campylobacter* spp. whereas multiplex real-time PCRs are used to differentiate between species. Monoplex and multiplex PCRs have been applied to detect and differentiate species of *Campylobacter* from chickens, cattle, milk, pets, environmental water and wild mammals (gorillas) (Sails et al. 2003a, Yang et al. 2003, Rudi et al. 2004, Lund et al. 2004, Bonjoch et al. 2010, Whittier et al. 2010). SNP profiling was developed for bacterial characterisation, which identifies single nucleotide variations or polymorphisms for strain comparison (Robertson et al. 2004). Most of the SNP-based studies for *Campylobacter* spp. have used comparative sequence data that were generated from MLST datasets to identify informative nucleotide polymorphisms (Robertson et al. 2004, Rudi et al. 2006).

A real-time taqman allelic discrimination assay was developed by Best et al. (2005) to delineate six major clonal complexes from MLST data available in the PubMLST database. These authors used the combined information from the alleles of six major clonal complexes (ST-21, ST-45, ST-48, ST-61, ST-206 and ST-257) to identify informative SNPs to detect the clonal complexes. Similarly, a bioinformatics-driven SNP genotyping assay using real-time PCR for *C. jejuni* and *C. coli* was developed and SNPs were identified

using the 'Minimum SNPs' software (Price et al. 2006) and the 'Minimum SNPs' software was updated as a part of this study. In this study 153 *C. jejuni* isolates were analysed using MLST, *fla*A, and real time PCR. The discriminatory power of SNP typing, MLST and *fla*A was 92%, 93% and 85%, respectively. When MLST was combined with *fla*A and SNP combined with *fla*A the discriminatory power for each method was 95%. Chan et al. (2008) used SNP and MLST to identify the presence of a hybrid *asp*A allele in a *C. coli* strain from a turkey that was shown to be derived from *C. jejuni*.

The emergence of new genotypes and the increasing availability of MLST data[4] means that the number of genotypes is constantly changing and increasing. Therefore, new analytical and data mining techniques are essential to make use of these vast data sources in developing a suitable SNP-real-time PCR package for routine use.

Clustered regularly interspaced short palindromic repeats

Numerous prokaryotic genomes contain structures known as clustered, regularly interspaced short palindromic repeats (CRISPRs). These are composed of 25 - 50 bp repeats separated by unique sequence spacers of similar length (Bolotin et al. 2005). CRISPR loci show a high level of polymorphism in different species of bacteria. CRISPRs have been exploited as a means for identifying clinical isolates of *Mycobacterium tuberculosis*, *Streptococcus pyogenes* and *C. jejuni* (Kamerbeek et al. 1997, Hoe et al. 1999, Schouls et al. 2003). These repeats are detected in the fully sequenced genomes of *C. jejuni* NCTC11168 and *C. jejuni* RM1221 (Fouts et al. 2005). These repeats are highly conserved within a species while they vary between species (Bolotin et al. 2005). Due to the large number of strains that are not able to be typed, this technique has not been used for regular bacterial species characterisation (Schouls et al. 2003), however it has been adopted for this purpose in some situations (for example routine typing of *Streptococcus* spp., (Horvath et al. 2008)).

DNA microarrays

DNA microarray is an intensive comparative genomic typing method which involves DNA-DNA hybridisation. It has provided considerable insights into intraspecies genetic

[4]URL: (`http://pubmlst.org/Campylobacter`)

diversity of many pathogens (Pearson et al. 2003, Taboada et al. 2005). This method measures the overall genetic relatedness between two isolates or strains by denaturing the query DNA and hybridising it with the known set of standard reference strains (Klena 2001). There are two basic methodologies involved in this technique: one based on cDNA expression arrays and the other on comparative genome (CGH) hybridisation arrays. The former is applied for whole genome characterisation and the latter is used for genotyping (Dorrell et al. 2001, Taboada et al. 2004). DNA microarrays are efficient tools for analysis of highly divergent and more conserved classes of genes among genomes (Leonard et al. 2003). Because this method targets genes more specifically, it can detect genomic differences in greater detail compared with other genotyping methods (Leonard et al. 2003). A comparative phylogenetic analysis of *C. jejuni* was carried out among whole genomes using microarrays where two different clades of *C. jejuni* were identified (Champion et al. 2005). Even though this technique is highly discriminatory at the whole genome level, it is not able to identify subtle mutations or nucleotide variations among *C. jejuni* isolates that occur due to evolutionary forces (Leonard et al. 2003). Therefore, this technique resides under a special category of genotyping methods that require specialised laboratory equipment and expertise (Klena 2001).

2.4.5 Multilocus sequence typing

MLST was first developed in 1991, and the technique uses comparative DNA sequencing of conserved housekeeping genes to characterise organisms (Maiden et al. 1998). Housekeeping genes are essential in the process of cellular metabolism of any life form. They are present in the core genome of all strains and encode proteins that are under stable selection for conservation of metabolic function (Maiden et al. 1998). In MLST, stretches of nucleotide sequences of approximately 400 – 600 bp from seven loci from a complete genome are chosen for analysis. The length of nucleotide sequences are chosen to give a reliable single run on automated sequencing instruments. MLST employs a universal nomenclature scheme for storing and interpreting nucleotide sequence data. Each allele fragment is assigned a unique number in the order of discovery. For example *asp*A-1 would be the first unique MLST allele identified for *asp*A locus (Maiden 2006). For each locus, distinct allelic sequences are assigned with allelic numbers and each isolate is therefore designated with seven numbers constituting an allelic profile which, in turn, is

given a sequence type (ST) or genotypic number. The isolates that share at least four alleles in common are grouped under a common central genotype, referred to as the founder ST or the known central ancestor, the clonal complex genotype.

MLST is reported to provide discrimination equivalent to 15 to 20 loci as examined by other techniques, such as MLEE (Dingle et al. 2005). An MLST system for *C. jejuni* was developed by Dingle et al. (2001) and is increasingly used in epidemiological studies (Manning et al. 2003, Clark et al. 2005) and population structure analysis of *Campylobacter* spp. The housekeeping genes for *Campyloabcter* spp. were chosen based on criteria such as chromosomal location, where the minimum distance was 70 kb between each gene, suitability to primer design and sequence diversity in the pilot studies employed by Dingle et al. (2001). A key advantage of MLST is that it can be used for population genetic studies as well as a typing tool for molecular epidemiological investigations (Maiden 2006).

Table 2.3 shows the housekeeping genes that were selected from the whole genome sequence available in the Genbank database (Parkhill et al. 2000) for the *C. jejuni* MLST scheme. Figure 2.2 shows the location of the seven housekeeping genes used for MLST on a circular genome of *C. jejuni*. The genome size of *C. jejuni* is 1.6 mega base pairs (mbp) and the locations of genes are dispersed around the genome.

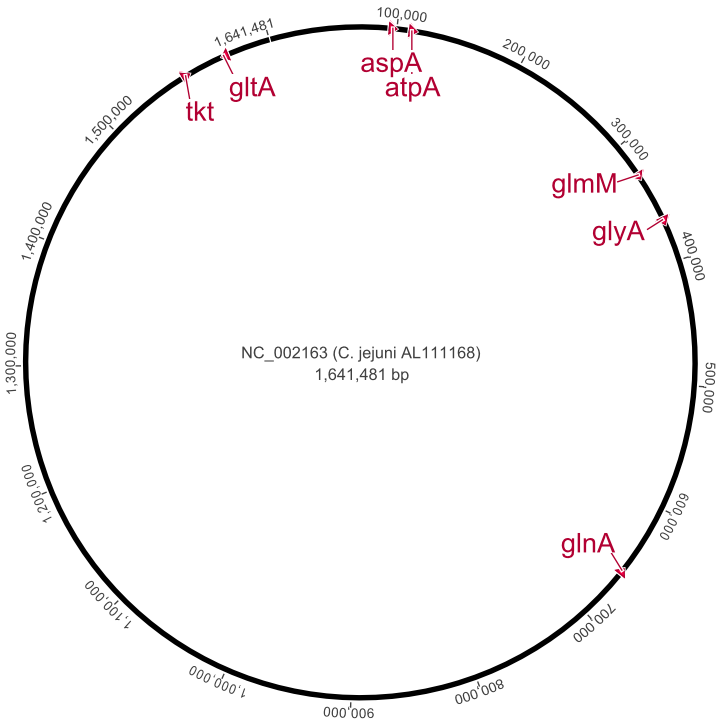
Table 2.3: Genes and gene positions used in a MLST typing scheme for *C. jejuni*.

Genes	Name	Function	Gene positions [a]
*asp*A	Aspartase	Amino acid metabolism	96074. . .97480
*gln*A	Glutamine synthetase	Amino acid metabolism	658331. . .656901
*glt*A	Citrate synthase	Tri carboxylic acid cycle	1605251. . .1603983
*gly*A	Serine hydroxy methyl transferase	Energy metabolism	367219. . .368463
*glm*M*	Phospho glucosamine mutase	Amino acid metabolism	327143. . .328480
tkt	Transketolase	Energy metabolism	1569190. . .1571088
*atp*A/*unc*A	ATP synthase a subunit	Energy metabolism	111488. . .112993

[a] Adapted from *C. jejuni* NCTC 11168 (Parkhill et al. 2000).

*:The *glm*M gene is still known as *pgm* as originally developed by Dingle et al. (2001).

Figure 2.2: Circular genome of *C. jejuni* that shows the location of seven housekeeping genes on the chromosome



The detection of outbreaks of gastrointestinal disease caused by *C. jejuni* has been facilitated by the use of MLST. Several reports are available on the use of MLST as an investigation tool (Dingle et al. 2002, Sails et al. 2003b, Urwin & Maiden 2003, McTavish et al. 2008). An important component of the MLST approach is the availability of databases (e.g. PubMLST) for use by public health and research communities. In turn researchers can submit the results of their findings to these databases (Maiden 2006). The following reports are examples that have utilised *C. jejuni* MLST for epidemiological investigation. Sails et al. (2003b) utilised MLST, PFGE and flagellin A gene typing in the investigation of a human campylobacteriosis outbreak in the United Sates of America. They investigated 47 isolates from 12 different outbreaks and analysed the STs that were associated with more than one outbreak. Similarly, MLST has been used to resolve a controversy related to the source of infection in Lancashire, England (Wilson et al. 2008). In this study Wilson et al. (2008) compared 1,145 animal and environmental *C. jejuni* isolates with 1,231 human *C. jejuni* isolates. They found that 20% of the animal STs accounted for 80% of human disease with ST-21 and ST-61 being the most frequent isolated genotypes. Sheppard et al. (2009) surveyed 5,247 clinical isolates from 28 diagnostic laboratories in 15 health boards in Scotland, from July 2005 to September 2006 using MLST to determine sources of human infection. The authors carried out a population structure analysis and found that ST-257 and ST-61 were more common in chickens and cattle, respectively.

In New Zealand, French (2008) compared the epidemiology of ruminant and poultry associated human cases of *C. jejuni* by performing a case-case comparison. Spatial and temporal attributes were compared among the two case groups, that is humans infected with poultry isolates *versus* humans infected with cattle isolated. In this study of 56 STs from 521 human samples they found ST-474 to be the dominant strain which was strongly associated with poultry. Similarly, Mullner et al. (2009) combined MLST and a modified Hald mathematical model to quantify the relative contribution of potential sources (poultry, cattle, sheep and environmental water) to human campybacteriosis identified in the Manawatu region of New Zealand. They inferred that the ruminant associated cases were most likely from environmental and occupational exposures rather than foodborne exposures (Mullner et al. 2009).

McCarthy & Giesecke (2001), McCarthy et al. (2007) employed MLST to analyse the host-association of *C. jejuni* genotypes between isolates obtained from chickens and cattle

and found that host association is stronger than temporal or geographic effects. Similarly, the natural populations of *C. jejuni* were examined by French et al. (2005) in a farmland ecosystem in the United Kingdom. They examined 172 isolates, and obtained 65 different sequence types (ST). There was an over representation of the ST-61 complex in the cattle isolates and the isolates from wildlife and water mostly belonged to the ST-45 complex. A cross-sectional study of *C. jejuni* populations in the same 100 km^2 area in Cheshire in the United Kingdom, characterised 327 *C. jejuni* isolates from cattle, wildlife and environmental sources using MLST (Kwan et al. 2008). Kwan et al. (2008) identified 91 STs and 18 clonal complexes (CC), with most of them belonging to ST-21, ST-45 and ST-61 complexes. These CCs have been shown to be frequently associated with human disease worldwide (Kwan et al. 2008, French et al. 2009a). In addition, Kwan et al. (2008) found that ST-21 and ST-61 were significantly associated with cattle and ST-45, ST-952 and ST-677 were associated with wild birds, wild rabbits and environmental water.

The population structure of *C. jejuni* in wild birds was studied by Colles et al. (2008) in Oxfordshire in the United Kingdom. This study was carried out between August 2002 and February 2003 within a mixed population of geese, starlings, lambs and free range chickens in the same farm ecosystem. A total of 331 faecal samples from geese, 954 samples from starlings and 975 samples from free range chickens were collected and compared with 540 *C. jejuni* human clinical isolates using MLST. Colles et al. (2008) found that CC-21 and CC-45 were the most abundant clonal complexes present in geese. These CCs were also shared by starlings and free range chickens. On a clonal frame tree (Prim 1957), isolates from geese and starlings formed separate clusters, and isolates from geese and chickens were closely related. This, in turn, was interpreted by the authors to mean that geese were a source of infection for poultry. In addition Colles et al. (2008) found that the geese genotypes were not monopyletic (not phylogenetically isolated) and shared common ancestors with genotypes from chickens and starlings. The authors, in addition, suggested that sequence data from more loci would need to be examined to provide a greater level of discrimination. Another serial cross-sectional survey was carried out in a wild bird population comprised of 2,084 individual birds (Hughes et al. 2009). Hughes et al. (2009) found that wild birds can carry both livestock and poultry associated genotypes (ST-42, ST-48 and ST-45) as well as novel genotypes (ST-3001, ST-3002, ST-3003, ST-3274, ST-3275 and ST-3276). Because of the apparent absence of unique *C. jejuni*

STs of wild birds in the livestock included in this study, Hughes et al. (2009) suggested that the direction of infection is from livestock to wild birds. In addition, the identification of unique STs in wild birds in this study was indicative of genetic recombination *in vivo* (Hughes et al. 2009). A longitudinal study was conducted in New Zealand to assess the potential risk of wild bird faecal contamination in children play areas using MLST (French et al. 2009a). This study was conducted between November 2004 and February 2005 where, one-half of the isolates recovered from wild bird faecal material belonged to ST-45, a genotype associated with many species of animals and human disease. The authors raised a possibility that the genotypes ST-177 and ST-682 isolated in this study might have originated from European birds during their introduction to New Zealand in the 19th century from the United Kingdom.

2.4.6 Whole genome sequencing

Despite the existence of numerous new generation genotyping techniques for *C. jejuni*, the molecular mechanisms for the pathogenesis, adaptation and the pattern of evolution still remains unclear (Fraser et al. 2009). Studies on bacterial population genetics have provided information on different genotypes found in different host species and population structures. However, analysis of genomes for single species or genera provides opportunities to better understand the details of evolution and adaptation (Rocha 2008). As bacteria can recombine within and between species and sometimes across kingdoms, molecular evolution in a given species can take place through recombination, horizontal gene transfer, mutation, deletion and duplication. This results in re-assortment of variants in an existing natural population (Lederberg & Tatum 1946, Heinemann & Sprague 1989, Rocha 2008). The potential of bacteria to transfer genes across species and/or kingdoms results in every individual gene potentially posing a different phylogenetic history (Rocha 2008). The genes that form the central part of a genome, the 'core genes', have been identified to provide significant phylogenetic signals about the inter-species and intra-species phylogenies (Rocha 2008). It is important and fundamental to study the patterns of individual genes in order to trace the history of cellular lineages which, in turn, frame evolutionary studies (Rocha 2008).

Chromosomal DNA and plasmid transformation in *C. jejuni* and *C. coli* were documented

under laboratory conditions using shuttle vectors to prove the natural competency of *Campylobacter* spp. (Wang & Taylor 1990). Subsequently, *C. jejuni* population diversity has also been demonstrated using a large number of MLST datasets from various host species. It has been demonstrated that the gene pools in different genotypes are overlapping in different host species (Dingle et al. 2001, Colles et al. 2003, Miller et al. 2006, McCarthy et al. 2007, Colles et al. 2008, Kwan et al. 2008, Wilson et al. 2008, Carter et al. 2009, Mullner et al. 2009), which in turn demonstrates the recombination potential of *C. jejuni*. Apart from MLST housekeeping genes, DNA uptake through homologous recombination has also been documented in virulence associated flagellin genes (Nuijten et al. 1990, Wassenaar et al. 1995) and the resistance gene (*tet*O gene) that confers tetracycline resistance *in vivo* in a study involving chickens (Avrain et al. 2004). Even though *Campylobacter* spp. is a highly recombining bacterial species, it has been reviewed that even highly recombining prokaryotic populations show identifiable patterns of phylogenetic relatedness (Spratt 2004). This observation has proven to be true in *C. jejuni* from a clonal context, which in turn, allows host associations to be identified. However, there may be several unprecedent diversities when isolates are looked at from a whole genome perspective (Section 2.4.4).

To date several *Campylobacter*[5] genomes from different species have been sequenced and details of twelve *C. jejuni* genomes that were available at the time of writing this thesis are provided in Table 2.4. Each one of the isolates has several characteristic features that differentiate them. For example, the genome of *C. jejuni* NCTC 11168 is devoid of any insertions, plasmids or transposons and this is a distinctive feature of this strain (Parkhill et al. 2000). In contrast a unique feature of the *C. jejuni* 81116 (NCTC11828) genome is the duplication of a 6.5 kb region which is not present in the genome sequences of *C. jejuni* strains NCTC11168 (Parkhill et al. 2000) and RM1221 (Parker et al. 2006). The genome of *C. jejuni* RM1221 is composed of 94% coding sequence (Fouts et al. 2005) and contains four genomic islands designated as CJIES i.e. *Campylobacter jejuni*-integrated elements and smaller gene clusters disrupting the whole genome (Fouts et al. 2005). The genome of *C. jejuni* 81-176 chromosome is almost a closed genome with only two remaining gaps located within highly repetitive regions (Hofreuter et al. 2006). *C. jejuni* 81-176 carries two large plasmids named p*Vir* and p*Tet* (Bacon et al. 2002, Batchelor et al. 2004).

[5]URL:(`http://www.ncbi.nlm.nih.gov/genbank/`)

Table 2.4: *Campylobacter jejuni subsp. jejuni* reference genomes.

Species	Strain	Size (Mb)	%GC	ORFs	Disease/source	Year and place of isolation	GenBank
Campylobacter jejuni subsp. jejuni	NCTC 11168	1.642	30.5	1643	Clinical food poisoning	1977, United Kingdom	AL111168
Campylobacter jejuni subsp. jejuni	RM1221	1.778	30.3	1838	Chicken	2005*, United Sates of America	CP000025
Campylobacter jejuni subsp. jejuni	81-176	1.17	30.6	1653	Clinical food poisoning	1981, United Sates of America	CP000538
Campylobacter jejuni subsp. jejuni	81116	1.628	30.5	1626	Clinical food poisoning	2007, United Kingdom	CP000814
Campylobacter jejuni subsp. jejuni	CG8421	1.609	30.4	1512	Clinical food poisoning	ns	ABGQ00000000
Campylobacter jejuni subsp. jejuni	HB93-13	1.695	30.6	1710	Clinical GBS	2006*, United Sates of America	AANQ00000000
Campylobacter jejuni subsp. jejuni	CG8486	1.598	30.4	1425	Clinical food poisoning	ns	AASY00000000
Campylobacter jejuni subsp. jejuni	CF93-6	1.676	30.5	1757	Clinical MFS	2006*, Japan	AANJ00000000
Campylobacter jejuni subsp. jejuni	84-25	1.672	30.4	1748	Clinical meningitis	2006*, United Sates of America	AANT00000000
Campylobacter jejuni subsp. jejuni	260.94	1.65	30.5	1716	Clinical GBS	ns	AANK00000000
Campylobacter jejuni subsp. jejuni	IA3902	1.64	30.5	1718	Sheep abortion	2010, United Sates of America	CP001876
Campylobacter jejuni subsp.doylei	269.97	1.845	30.6	2094	Human blood	2007, United Kingdom	CP000768

MLST: Multilocus sequence type.

ST: Sequence type.

CC: Clonal complex.

GBS: Guillain-Barré syndrome.

ORFs: Open reading frames.

Mb: Mega basepair.

GC: Guanine:cytosine.

ns: not stated.

*:Date of start of project

Genome plasticity and housekeeping genes

Housekeeping genes are those genes that are critical for cellular maintenance encoding proteins involved in metabolic pathways (Rivera et al. 1998). The genes encoding for metabolic pathways, cell wall components and ribosomal elements make up the vast majority of housekeeping genes (Riley 1993, Rivera et al. 1998). Since they are part of the core genome present in all strains, they are often used in population biology and to study the phylogenetic relationships between various species and subspecies of bacteria (Mahadevappa & Warrington 2002). The lineages of genes are broadly classified into informational and operational genes based on 'the complexity hypothesis' formulated by Rivera et al. (1998), Jain et al. (1999). The informational genes include genes of translation (T), transcription (S), and replication (R) and also the ATPases, GTPases (G) and tRNA synthetases. Operational genes are those involved in cell operations such as amino acid synthesis (A), biosynthesis of co-factors (B), cell envelope proteins (C), energy metabolism (E), intermediary metabolism (I), fatty acid and phospholipid biosynthesis (L), nucleotide biosynthesis (N), and regulatory genes (Z). Genes involved in transcription, translation and replication are grouped into a larger complex on the basis of their larger assemblies of gene products. Operational genes belong to the smaller complex, as they produce smaller assemblies of gene products (Coenye & Vandamme 2005). The operational genes are the most modular genes in the cells that are inclined to be horizontally transfered or recombined most often (Jain et al. 1999, Ma & Zeng 2004, Coenye & Vandamme 2005, Schumann 2005). However, the classification of genes and their lineages still remain unclear.

Genes involved in DNA repair, recombination and nucleotide metabolism pathways are some of the important key machineries of the genetic information processes in bacteria.[6] These machineries play a dual role of maintaining the genetic stability of an organism as well as in repairing the DNA lesions (Cann & Ishino 1999, Paques & Haber 1999, Zhou & Elledge 2000, Singh et al. 2010). These repair mechanisms take the advantage of incorporating DNA sequences or nucleotides from related (homologous) and sometimes unrelated (non-homologous) bacterial species to correct DNA damage (Schumann 2005). Metabolic housekeeping genes and DNA repair genes are therefore always vulnerable for nucleotide change (Coenye & Vandamme 2005) and such changes enable the bacteria to

[6]URL:(`http://www.genome.jp/dbget-bin/www_bget?ko+K03495`)

adapt to a new environment thus producing population diversity in accordance with the environment in which they are present. Metabolic housekeeping genes and DNA repair genes are under stringent purifying selection, but, the polymorphic regions or sites within the genes that provide selective advantages operate under positive selective pressure. For example, the DNA repair gene (this genes is also used in the MLST scheme for typing) whose nucleotide mutation is reported to be under advantageous positive selection in *C. jejuni* is *gyr*A (Snyder & Champness 1997, Han et al. 2008). The *gyr*A gene encodes for a DNA gyrase that is involved in DNA replication and DNA repair mechanisms, while a mutation at the 86th codon, confers fluoroquinolone resistance (Han et al. 2008). The impact of recombination (homologous, non-homologous or illegitimate), mutation and HGT has been shown to be enormous for the creation of microbial genome plasticity which, in turn, leads to evolution of microorganisms (Feil et al. 2001, Spratt et al. 2001, Hanage et al. 2005, Fraser et al. 2007, 2009). However, the rate of recombination may differ greatly amongst different bacterial species. While some species recombine more frequently to have multiple recombinational events than mutations that render them weakly clonal, in other species it appears to be a rare incident that leads to distinct clonal lineages (Spratt et al. 2001, Hanage et al. 2005, 2006, Fraser et al. 2007).

C. jejuni is a rapidly evolving species which is highly influenced by HGT and recombination (Sheppard et al. 2008, Duong & Konkel 2009). The MLST scheme for *Campylobacter* spp. demonstrated the massive evolutionary and recombination potential of *C. jejuni* (Dingle et al. 2001, 2005). The adaptational capacity of *C. jejuni* to different niches and the acquisition of definitive host signatures has been demonstrated in previous studies in greater detail using MLST datasets (Colles et al. 2003, McCarthy et al. 2007, Colles et al. 2008, 2009, Sheppard et al. 2009, 2011). Moreover, the antigenic genes of *C. jejuni* (the *fla*A, *fla*B and *por*A) have asp been shown to undergo frequent recombination and most often are considered alongside the MLST housekeeping genes to provide a high discriminatory ten locus typing system (Meinersmann et al. 1997, Meinersmann & Hiett 2000, Meinersmann et al. 2002, Dingle et al. 2008, Cody et al. 2009).

Generally, whole genome phylogenetic comparisons have been made and the ancestry of individual organisms have been examined (Dorrell et al. 2001, Coenye & Vandamme 2005, Pearson et al. 2003, Carrillo et al. 2004, Ma & Zeng 2004, Gressmann et al. 2005). Individual gene analyses to determine individual gene ancestry within an organism have

not been carried out to date. The cumbersome DNA sequencing processes, DNA library construction, contig analysis, time, and expertise required present constraints to achieving this objective. Whole genome analyses, by comparison, are more straightforward and have become more convenient with the advent of high throughput sequencing technologies.

Next generation sequencing — The Solexa

Determination of DNA sequences came into existence in 1977 through an invention by Sanger & Coulson (1975). This transformed biology by providing a tool for deciphering genes and whole genomes. A related method involving radioisotopes for DNA sequencing was developed in the same year by Maxam & Gilbert (1977). In contrast to the Maxam1977 method the Sanger & Coulson method and its subsequent improvements do not use radioisotopes or toxic chemicals. This has rendered this technique an extremely useful tool for DNA-based research (Schuster 2008). Human genome sequencing was a tremendous breakthrough, first carried out using an automated capillary driven sequencing protocol which eventually led to the creation of established laboratories for DNA sequencing (Schuster 2008). Sequencing by synthesis technology was developed in 2005 by a company called 454 Life Sciences (Margulies et al. 2005). This company used a strategy involving the arraying of several hundred thousand sequencing templates (Schuster 2008). Recent improvements include pyro-sequencing, a process that detects single nucleotide polymorphisms. Next generation sequencing tools have been applied to deduce all mutations at the cellular level in *Mycobacterium tuberculosis* to identify drug resistant alleles (Andries et al. 2005). Early approaches involved a combination of the Sanger method and pyrosequencing but this had the disadvantage of being expensive. More recently low cost non-Sanger sequencing methods such as Illumina's Solexa or Applied Biosystem's SOLiD platforms (Schuster 2008) are being used.

Illumina Inc. is a company founded in 1998 by a group of biotechnologists in San Diego, California, United Sates of America.[7] This company offers a range of genotyping services that includes SNP profiling, DNA microarray, gene expression, and protein analytical systems to cater for a broad range of academic, governmental, pharmaceutical and biotechnological institutions around the world. In 2007, this company acquired Solexa

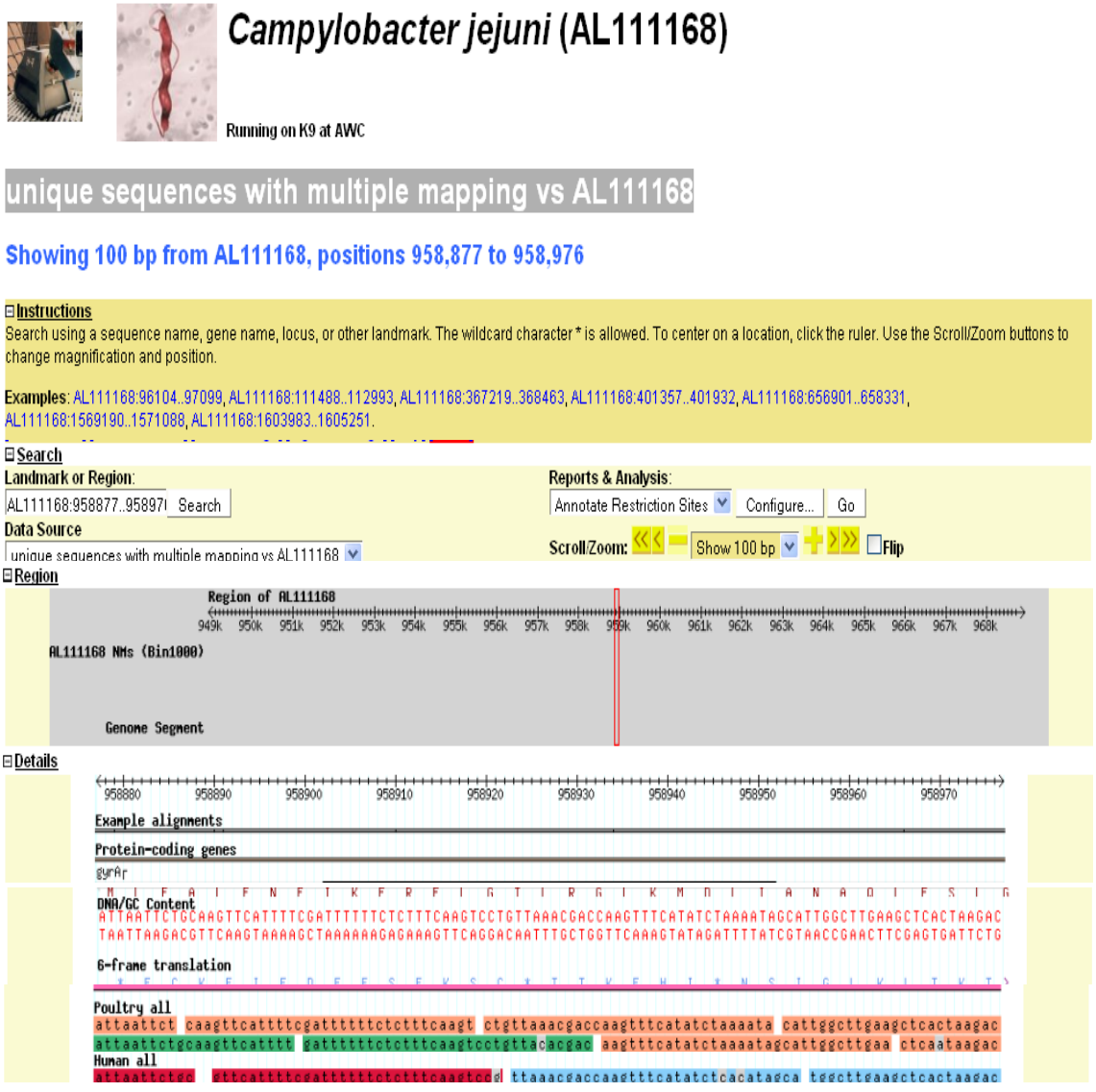
[7]URL:(`http://www.illumina.com`)

Inc., that develops genetic analysis technologies used for whole genome sequencing, ribonucleic acid (RNA), gene expression analysis. With the availability of non-Sanger sequencing methods, it has become possible to generate large amounts of sequence data that has renamed the application as re-sequencing. This technique is guided by a reference genome sequence to assemble the sequences that requires much less coverage than assembling genomes *de novo* (without reference) (Schuster 2008). Using this technology ten mammalian mitochondrial genomes were sequenced, enabling population genetics studies (Gilbert et al. 2007). In the Hopkirk Research Institute, IVABS, Massey University, New Zealand this technology has been used to index genomic variations in *C. jejuni* of the same strain, ST-474 (Chapters 5 and 6). Figure 2.3 is an example of the Solexa sequencing and assembly of the *C. jejuni* genomes of poultry and human matched against reference genome *C. jejuni* NCTC 11168 (GenBank ID: AL111168).

2.5 Knowledge search

Even though many *Campylobacter* genomes have been sequenced, many details of their pathogenicity, host association, population diversity and epidemiology remain unclear. Studies on individual housekeeping genes are scarce and most studies have either focused on the congruence of phylogenies interpreted by rRNA or a few housekeeping genes (Puhler et al. 1989, Bustamante et al. 1995, Bult et al. 1996, Martin 1999, Dorrell et al. 2001, Vandecasteele et al. 2001, Hinode et al. 2002, Velayudhan & Kelly 2002, Wolf et al. 2002, Coenye & Vandamme 2003, Fuglsang 2003, Pearson et al. 2003, Viscidi & Demma 2003, Carrillo et al. 2004, Karenlampi et al. 2004, Ma & Zeng 2004, Nakamura et al. 2004, Taboada et al. 2004, Venter et al. 2004, Foerstner et al. 2005, Fouts et al. 2005, Chen et al. 2006, Gupta 2006, Klancnik et al. 2006, Musto et al. 2006, Rooney et al. 2006, Miller et al. 2007, Sorek et al. 2007, Hughes et al. 2008, Rocha 2008, Liu et al. 2009, Lefebure & Stanhope 2009, Rocha & Feil 2010). Moreover, the molecular differences in housekeeping genes have not been studied at the individual gene level except for a sub-region among seven housekeeping genes in MLST (Section 2.4.4). Although the sequences of seven housekeeping genes illustrate the genetic diversity and population genetics in bacterial species, the underlying molecular variation when looked at a full gene level is obscure.

Figure 2.3: Sequences of *C. jejuni* ST-474 (poultry and human) with multiple mapping vs *C. jejuni* AL111168



CHAPTER 3

Campylobacter jejuni colonisation and population structure in sympatric urban population of ducks and starlings

Abstract – Wild birds such as mallard ducks and European starlings are potential sources of *Campylobacter* infection in humans and farm animals. A repeated cross-sectional study was conducted to determine the prevalence of *Campylobacter* in the faeces of mallard ducks and European starlings present in an urban area in provincial New Zealand. A total of 1,458 faecal samples were collected and cultured monthly from August 2008 to July 2009. Presumptive *Campylobacter* colonies were screened using two different monoplex PCRs: one for the genus *Campylobacter* and the other for the species *C. jejuni*.

The isolation rates of *C. jejuni* in ducks and starlings were 23% (165 of 716) and 21% (150 of 720), respectively. Characterisation of 124 *C. jejuni* isolates was performed by multilocus sequence typing and *fla*A and *por*A typing. This revealed evidence of host associated sequence types (ST). The ST-177 and ST-682 clonal complexes represented the starling associated complexes and the ST-1034, ST-692 and ST-1332 complexes predominantly represented the duck associated complexes. The prevalence of the ST-45 complex, a complex associated with many species of animals and humans with campylobacteriosis was high both in ducks (14%, 9 of 64) and starlings (28%, 16 of 57) particularly during summer. Further, the *C. jejuni* sequence types ST-1255, ST-137, ST-2026, ST-583, ST-526, ST-677, ST-696 and ST-710 that have been isolated from sporadic cases of campylobacteriosis in humans in other parts of the world were also isolated from ducks and starlings during the summer months of this study. This suggests that wild birds may be a source of human campylobacteriosis in the summer months. The *fla*A alleles were more diverse than the *por*A alleles particularly in ducks. Rarefaction analysis of MLST data showed that the *C. jejuni* sequence types from ducks were more diverse than that of starlings. Further, *C. jejuni* sequence type richness was higher during the winter months in ducks, where a higher proportion of STs that could not be assigned to a clonal complex were isolated. By calculating the pairwise Fst values and conducting an analysis of molecular variance (AMOVA), it was evident that the population of *C. jejuni* in ducks was significantly different from that of starlings and that the gene flow between these two species was limited. Further, the population of *C. jejuni* differed over time, whereas there was little evidence of differentiation between sampling sites.

Vathsala M, French N, Stevenson M, Marshall J, and Hotter G, (2011) *Campylobacter jejuni* colonisation and population structure in sympatric urban population of ducks and starlings

3.1 Introduction

Campylobacter is the leading cause of bacterial gastroenteritis in the industrialised world (Notermans 1994). *C. jejuni* and *C. coli*, in particular, are considered to be the most important human pathogens. *C. jejuni* accounts for approximately 90% of human cases of campylobacteriosis with *C. coli* accounting for the remainder. Although the disease is self limiting it occasionally produces severe sequelae such as Guillain-Barre syndrome (Altekruse et al. 1999). Consumption of contaminated and undercooked meat is generally thought to be the main source of infection and this exposure pathway has been studied extensively (Baker et al. 2006). However, contact with farm animals, pets and other environmental exposures including wild birds have also been implicated as potential sources of human campylobacteriosis (Friedman et al. 2000, Workman et al. 2005, French et al. 2009a).

Multi-locus sequence typing (MLST) has become widely used for molecular epidemiological studies of *Campylobacter* (Maiden et al. 1998, Dingle et al. 2001, Maiden 2006). The publicly accessible on-line MLST database provides definitive data and enables direct comparison of bacterial strains across the world (Jolley et al. 2001) allowing population based genetic analyses to be carried out (Maiden 2006). MLST has been extensively used to investigate the population biology of *C. jejuni* in various host species (Maiden 2006, McCarthy et al. 2007). The seven-locus MLST generates sequence types (ST) which can be grouped into clonal complexes (CC) on the basis of the sequence types sharing five or more alleles in common with a founder genotype or sequence type (Dingle et al. 2002). Recent studies have shown that *Campylobacter* populations show large differences among host species and environmental niches (McCarthy et al. 2007) and that their lineages are associated with different host sources (Colles et al. 2003, Fearnhead et al. 2005, Miller et al. 2006, Colles et al. 2008) using MLST. Further, rare allelic profiles (sequence types that were not assigned to a clonal complex) and their distribution among animal sources have been reported in previous investigations using MLST (Kwan et al. 2008, Carter et al. 2009).

The adaptation of *Campylobacter* spp. in birds is believed to be due to their thermophilic nature that corresponds more closely with the body temperature of birds, compared with mammals. As a result, birds have been widely regarded as the natural hosts for these

organisms (Lee & Newell 2006). *Campylobacter* spp. have been isolated in a number of birds, both domesticated and wild (Luechtefeld et al. 1980, Fricker & Metcalfe 1984, Frost 2001). The carriage rates among various species of wild birds vary from 0% to 100% (Waldenstrom et al. 2002), where differences in carriage rates are thought to be due to factors related to ecology, for example feeding habits, habitat preferences and migration patterns.

Mallard ducks and European starlings are potential sources of contamination of water for drinking and recreation purposes, gardens, parks and childrens' playgrounds in urban areas, often creating heavy faecal contamination (Odermatt et al. 1998, Colles et al. 2008). Starlings have been reported to have relatively high carriage rates of *Campylobacter* (40% was identified in the studies performed by Waldenstrom et al. (2002)) compared with other wild birds species. Starlings have also been shown to carry sequence types of *C. jejuni* similar to those isolated from human patients (Colles et al. 2003, Broman et al. 2004, French et al. 2009a). One recent study showed that none of the unique *C. jejuni* STs found in wild birds were identified in either domestic animals or humans (Hughes et al. 2009). Furthermore, Hughes et al. (2009) hypothesised that the livestock-associated strains that were detected in wild bird samples could have arisen from the shared environment that these two species cohabitat. An additional hypothesis was that if wild birds acted as reservoirs of *C. jejuni* for livestock, it should be expected that those strains found in wild birds should also be detected in livestock. The role of wild birds as potential reservoirs of human pathogenic *C. jejuni* STs could not be ascertained from this study. However, the CC ST-45 is a commonly reported strain of *C. jejuni* causing disease in humans and is also commonly isolated from (asymptomatic) animals including livestock and wild birds (Colles et al. 2008, Kwan et al. 2008, French et al. 2009a). This provides evidence that wild birds may also be a source of human campylobacteriosis.

A study investigating *C. jejuni* isolates from starling faeces recovered from children playgrounds in Palmerston North, New Zealand raised the possibility that starlings may harbour *C. jejuni* sequence types of European origin. This may have arisen when starlings were introduced into New Zealand by 'Acclimatisation Societies' (Societies that were created to enrich the fauna of New Zealand with different animal and plant species) that were active in the late 19th century (Thomson 1922, French et al. 2009a). Furthermore, this study identified that one-half (12/22) of the strains recovered from starling faecal material

were ST-45 strains, a ST associated with many species of animals, including wild birds as reviewed earlier (French et al. 2009a). French et al. (2009a) showed that the *Sma*I PFGE profile of three of the ST-45 isolates were indistinguishable from the profile of isolates recovered from human clinical cases that occurred in the same geographical area, in the same time frame. This provides evidence that wild bird faeces in playgrounds may be a source of human infection, particularly in children. This said, exposure to wild bird faecal material is unlikely to be a primary contributor to the overall burden of human clinical cases in New Zealand (French et al. 2009a).

The risk of campylobacteriosis arising from consumption of contaminated food has been extensively studied (Baker et al. 2006, 2007, Eberhart-Phillips et al. 1997). Other important modes of transmission include exposure to faecal material from livestock, including ruminants (Savill et al. 2003, Mullner et al. 2009, Wilson et al. 2009). From a public health perspective, understanding the relative contributions of other environmental exposure pathways is critical for designing appropriate control measures. As reviewed earlier, *C. jejuni* STs of wild birds have not been detected in livestock and also their role as a cause of human disease cannot be ascertained, primarily due to the limited number of studies that have explicitly investigated this association. It is important to investigate the population structure of *C. jejuni* STs among wild birds identifying their putative host preferences and genetic diversity which in turn will provide a better understanding of host-association and inter-host transmission. Common statistical methods for analysing population genetic structure include rarefaction analysis and the measure of molecular heterozygosity (Rocchini et al. 2009, Holsinger & Weir 2009)

Rarefaction is a method employed to assess the allelic richness in a set of populations (Siegel & Gelman 1982, Leberg 2002, Foulley & Ollivier 2006, Rocchini et al. 2009) where, initially, rarefaction analysis was employed to give the expected number of higher taxonomic groups, such as families or genera, represented in a random selection of lower taxonomic units, such as species or individuals (Siegel & Gelman 1982). Subsequently in ecology, rarefaction was employed for estimating the expected number of species within a given study area from local to regional scales where rarefaction curves were directly related to the heterogeneity in the species studied in the area of sampling (Rocchini et al. 2009). Wright (1951) described a common method to summarise the population structure of a given species that has become a widely used analysis in various fields of biology

where a series of hierarchical measures of heterozygosity such as Fst, Fsc and Fct are derived (Wright 1965, 1978). These analytical approaches are applied to gather insights into the geographic barriers to gene flow, the inter-site and inter-host transmission of bacterial populations.

With this background, the aim of this study was to investigate the prevalence of *C. jejuni* in the faeces of sympatric populations of mallard ducks and starlings found in public access areas in a provincial city in the lower North Island of New Zealand. Further, our aims were to characterise the *C. jejuni* isolates from the above species of birds, including a small sample of geese. A comparison of the *C. jejuni* STs from the sampled wild birds with those disease-causing human sequence types should be helpful when attempting to evaluate the possibility that these species are reservoirs of *C. jejuni* human pathogens in New Zealand. We also compared sequence types isolated from mallard ducks, starlings and geese with the wider population of *Campylobacter* sequence types available in the PubMLST database.[1] Comparison of STs with the wider global population provides an overview of the geographical distribution of these STs and the range of hosts in which they are found. Comparing isolates from ducks, starlings and geese can also provide insight into the population diversity of *C. jejuni* in these species, their host association and identification of molecular signatures associated with these hosts.

3.2 Experimental procedures

3.2.1 Study design

A repeated cross sectional study was conducted to determine the prevalence of *Campylobacter* in the faeces of mallard ducks (*Anas platyrhynchos*), European starlings (*Sturnus vulgaris*) and a small sample of Canadian geese (*Branta canadensis*) resident in the city of Palmerston North (longitude 175°, latitude -40°) in the lower North Island of New Zealand. Five public parkland sites within the city limits were selected for sampling: The Square, Hokowhitu, Memorial Park, Massey University and The Esplanade. A child's play area was present in four of the five locations and all sites had at least one duck pond (Figure 3.1).

[1] URL: (`http://pubmlst.org/Campylobacter`)

Assuming a prevalence in the order of 40% (Colles et al. 2008) calculations were conducted to determine the appropriate number of samples to be 95% certain that our estimate of *C. jejuni* prevalence was within 5% of the true population value (that is, from 35% to 45%). The number of samples that needed to be taken to achieve these objectives was estimated to be (Levy & Lemeshow 1999, pp. 258):

$$n \geqslant \frac{z^2(1-P_y)P_y}{\epsilon^2} \tag{3.1}$$

In Equation 3.1 P_y represents the unknown population prevalence, z is the reliability coefficient (1.96 for an alpha level of 0.05), and ϵ is the maximum absolute difference between the prevalence estimated from the sample and the true (unknown) population value. These calculations indicated that a total of 369 samples were required to meet the study objectives.

To account for the possibility that prevalence varied across the five sampling sites (that is, there was clustering of *Campylobacter* within sites) the estimated sample size was multiplied by a design effect of 2 (Levy & Lemeshow 1999, pp. 292). This means that the actual variance was twice that of the variance computed under the assumption of simple random sampling. The prevalence estimates and the confidence intervals were adjusted for design effect by including the design effect in the prevalence estimation.

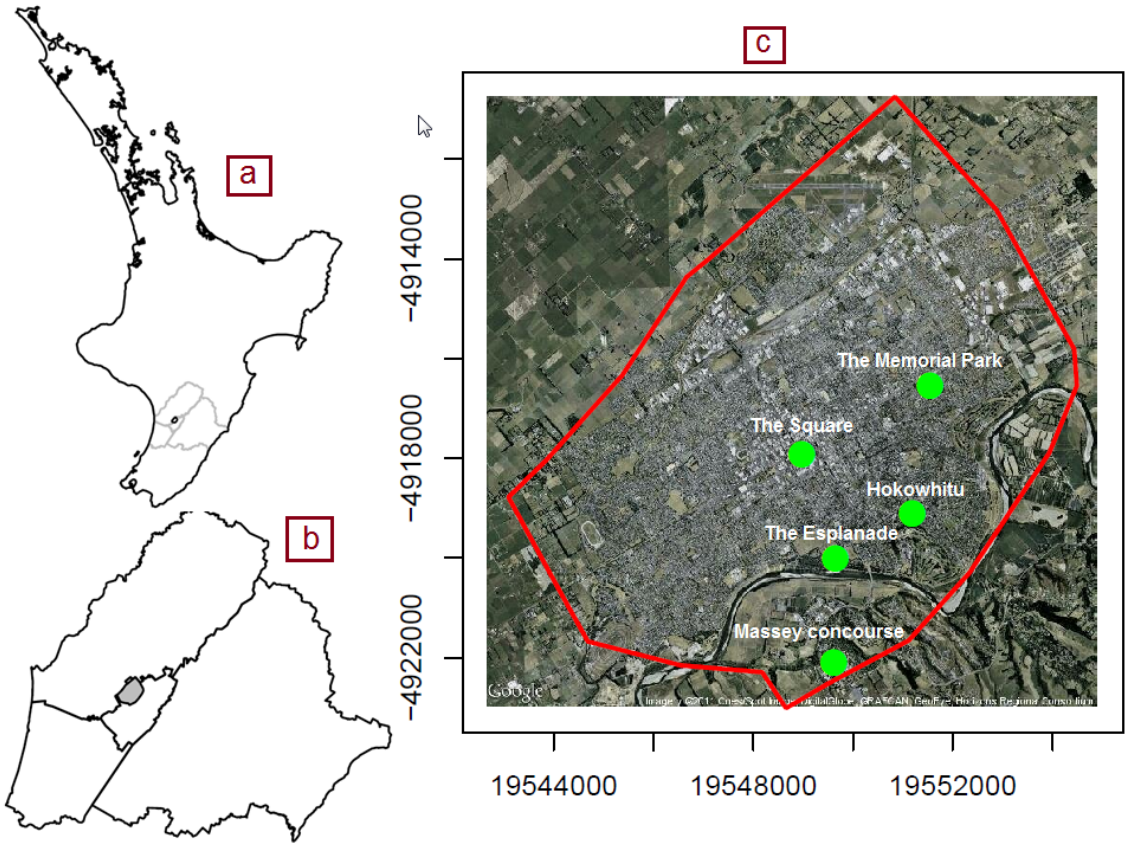


Figure 3.1: Ducks and starlings sampling sites. The map describes the five sampling sites from where the faecal materials were collected from mallard ducks and starlings for the isolation of *C. jejuni*. a. North Island of New Zealand; b. Manawatu – Palmerston North city; c. The five public parkland sites. (Geese faecal materials were collected from the Hokowhitu site only)

3.2.2 Collection of faecal material

Each of the five study sites was visited at monthly intervals for a period of 17 months, starting in March 2008. Standardisation of transport of faecal material and processing for *Campylobacter* isolation was carried out for a trial period of five consecutive months, from March to July 2008 (inclusive). The results reported in this paper are for the period August 2008 to July 2009 (inclusive). To facilitate comparison and interpretation of results, the months of sample collection were categorised into warmer times of the year (spring and summer, September 2008 to February 2009) and cooler times (autumn and winter, March to July 2009 and August 2008).[2] We classify these two as summer and winter.

At each sampling site fresh faecal material that was moist and slimy from ducks was collected from areas adjacent to the water sources where ducks rested. Nesting areas for starlings were identified at each sampling site and faecal samples under each nesting area collected. It is possible that other birds may have defecated under nesting areas, however, during the five months trial period, identification of starling faecal material was standardised by waiting and watching the birds defecating and the faecal characteristics were recorded as photographic images. Subsequently, during the main study only fresh faecal material with specific characteristics located directly under starling roosting areas were sampled (Photographic images of starling faecal material are provided in Appendix A, Figures A.1 a, b and c). Therefore, we are confident that the samples used in this study were from starlings. Samples were collected during the early mornings (0600 to 0800 hours) or late evenings (1800 to 2000 hours) on each sampling day. Each sampling site was divided into four quadrants and three samples from each quadrant collected from both ducks and starlings producing a total of 12 samples for each sampling day. The location of quadrants and the sampling sites did not vary between sampling days.

For the August 2008 sampling round it was not possible to collect a complete set of duck samples from two quadrants at the Hokowhitu site for the following reasons. One quadrant remained unsampled for duck faeces as that quadrant was full of water and for the second quadrant faecal samples were collected from two ducks and a single goose as there were only two duck samples. Thus, the total number of duck samples for the

[2]URL: (`http://www.fourcorners.co.nz/new-zealand/seasons-climate/`)

duration of the study was 716. Because Canadian geese were frequently sighted at the Hokowhitu site, samples (n = 22) were collected from this species for the months of August 2008 and February, March, April and July 2009. Data recorded for these months have not been used to calculate the *C. jejuni* prevalence estimates presented in this study.

3.2.3 Bacterial Isolation and DNA preparation

Faecal samples from all sources were collected simultaneously in transport media (Amies charcoal, Fort Richards, Auckland) and Bolton's enrichment broth (enrichment broth LAB-27.6 G; 50 mL lysed horse blood venous supplies; Antibiotics-LAB-10 mL, Auckland) and transported immediately to the Hopkirk Research Institute Laboratory on the Massey University campus at Palmerston North.

Faecal material collected in the transport media were directly streaked onto modified charcoal cefoperazone - deoxycholate (mCCDA) (Fort Richards, Auckland) plates. The inoculated mCCDA plates and Bolton's enrichment broth with faecal material were incubated for 48 hours at 42°C in a microaerophilic chamber (MACS VA500 Microaerophilic workstation, Don Whitley Scientific) with a gas composition of 5% oxygen, 10% carbon dioxide and 85% nitrogen. The plates were monitored for growth, and after 48 hours the colonies resembling *Campylobacter* spp. were sub-cultured onto blood agar plates (horse lysed blood agar, Fort Richards). After 48 hours the cultures from the Bolton's enrichment broth were swabbed onto mCCDA plates and then the inoculated plates were incubated for another 48 hours at 42°C in a microaerophilic chamber. Up to three colonies from the mCCDA plates were sub-cultured onto horse blood agar plates. The pure colonies isolated from the horse blood agar plates were tested for oxidase reduction (oxidase strips, Fort Richards, Auckland). The colonies that reduced oxidase within 5 seconds, as indicated by a purple colouration, were stored in glycerol and processed for DNA isolation. Three colonies of at least 3 mm in diameter were transferred to 1 mL of 2% (weight/volume) Chelex solution in distilled water and boiled at 100°C on heating blocks for 10 minutes. These were then cooled to room temperature, centrifuged at 13,000 rpm for 10 minutes and the supernatants collected in fresh sterile eppendorf tubes and stored at -20°C.

3.2.4 Speciation and characterisation by Polymerase Chain Reaction

Isolates were confirmed to be *Campylobacter* spp. and *C. jejuni* using monoplex PCR that targeted 16s rRNA (Linton et al. 1997) for *Campylobacter* spp. and the membrane associated protein A (*map*A) for *C. jejuni* (Stucki et al. 1995, Mullner et al. 2010), respectively. Genus primer sequences were: forward 5' GGATGACACTTTTCGGAGC 3'; reverse 3' CATTGTAGCACGTGTGTC 3'. *C. jejuni* primer sequences were: forward 5' CTTGGCTTGAAATTTGCTTG 3' and reverse 3' GCTTGGTGCGGATTGTAAA 5'. The targets were amplified at 96°C for 2 minutes for initial denaturation, 96°C for 30 seconds, primer annealing at 56°C for 30 seconds, and extension at 72°C for 60 seconds, for 35 cycles. The PCR reaction mix was comprised of 2 μL 10$\times$ PCR buffer (final concentration 1$\times$); 2 μL dNTPs (final concentration 2 mM); magnesium chloride 1 μL, (final concentration 2.5 mM); primers 2 μL each (final concentration 1 mM); Taq DNA polymerase 0.2 μL (final concentration 1 unit per reaction); DNA 2 μL (final concentration 10 ng per μL). The reaction mix was made up to 20 μL with distilled water. The amplicons were examined by agarose gel electrophoresis with results captured using a Bio-Rad gel documentation system (Life Science Group, Canada).

3.2.5 Multilocus sequence typing

Isolate selection for MLST screening

Where available, two representative *C. jejuni* positive isolates were chosen from each sampling site and each species, one from mallard ducks and the other from starlings, for complete MLST characterisation. A total of 120 isolates were characterised but it was not possible to isolate *C. jejuni* from some sampling sites on certain occasions (for example, November 2008 to January 2009 for Hokowhitu, December 2008 for The Square, and March 2009 for The Esplanade). Additional samples were chosen from the sites that had ample *C. jejuni* positive samples to increase the sample size for an analysis of genetic diversity of *C. jejuni*. A total of 140 of 315 *C. jejuni* positive samples from ducks, starlings and geese were therefore characterised by multilocus sequence typing (MLST).

MLST was carried out as described previously by Miller et al. (2005) using the primers shown in Table 3.1. Each 25μL amplification reaction mixture was comprised of 10

nanogram (ng) of *Campylobacter* chromosomal DNA per μL (2μL), 1 μL (5 picomoles) PCR primers (forward and reverse primers), 12.5 μL of 2$\times$ ABI Taq Gold Mastermix and 9.5 μL of distilled water. The reaction was carried out in 96 well plates for the 7 housekeeping genes. Each plate held 13 samples at one time. Amplification conditions for the PCR program consisted of a denaturation step at 94°C for 15 minutes on a 35 cycle run with initial denaturation at 94°C for 30 seconds, primer annealing at 50°C for 30 seconds and extension at 72°C for 90 seconds. The final elongation step was for 7 minutes. On completion, the plates were held at 4°C for further precipitation and purification of the amplicons.

Table 3.1: The forward and reverse primer sequences used for the amplification of housekeeping genes of *C. jejuni* (Miller et al. 2005).

Alleles	Forward and Reverse primers
*asp*A	F-5' GAGAGAAAAGCWGAAGAATTTAAAGAT 3'
	R-3' TTTTTTCATTWGCRSTAATACCATC 5'
*gln*A	F-5' TGATAGGMACTTGGCAYCATATYAC 3'
	R-3' ARRCTCATATGMACATGCATACCA 5'
*glt*A	F-5' GARTGGCTTGCKGAAAAYAARCTTT 3'
	R-3' TATAAACCCTATGYCCAAAGCCCAT 5'
*gly*A	F-5' ATTCAGGTTCTCAAGCTAATCAAGG 3'
	R-3' GCTAAATCYGCATCTTTKCCRCTAAA 5'
pgm	F-5' CATTGCGTGTDGTTTTAGATGTVGC 3'
	R-3' AATTTTCHGTBCCAGAATAGCGAAA 5'
tkt	F-5' GCAAAYTCAGGMCAYCCAGGTGC 3'
	R-3' TTTTAATHAVHTCTTCRCCCAAAGGT 5'
*unc*A	F-5' GCAAGGDGTTATYTGTATWTATGTTGC 3'
	R-3' TTTAADAVYTCAACCATTCTTTGTCC 5'

Purification of PCR amplified products

The amplified products were precipitated by mixing with 25 μL of PEG (20% Polyethylene glycol 8000 in 2.5 M NaCl) and incubated at 37°C for 15 minutes and the products were centrifuged at 2,500 rpm for 30 minutes. The PEG complex in the plate was cleaned by spinning the plate inverted onto tissue towels at 300 rpm for 2 minutes. The DNA in the plates was washed with 80% ethanol and centrifuged for 10 minutes at 2,500 rpm and the ethanol removed by inverting the plate once again onto tissue towels. The plates

were air dried in a dark clean chamber until the ethanol evaporated. The products were re-hydrated in sterile distilled water and were examined by agarose gel electrophoresis with results captured using a Bio-Rad gel documentation system (Life Science group, Canada). Quantification of DNA was done by visualising the bands on the agarose gels.

Sequencing

The sequencing reaction mix was comprised of 400 μL of distilled water, 200 μL of 5$\times$ sequencing buffer and 100 μL of BigDye Ready Reaction Mix (PE Biosystems). Seven μL of this mixture was dispensed to each well with 1 μL of forward primer (3.2 picomoles). Two μL of amplified DNA products from the amplification plate were then transferred to each well in the same order as that of the amplification plate and sequenced in the PCR machine. The sequencing cycle consisted of an initial denaturation step at 96°C for 3 minutes for 25 cycles with initial denaturation at 96°C for 15 seconds, primer annealing at 50°C for 15 seconds, and extension at 72°C for 4 minutes. The plates were then held at 4°C for further purification.

Precipitation and purification of sequenced products

The unincorporated dye terminators were removed by precipitation of the termination products with 95% ethanol, according to the manufacturer's instructions. The precipitation mixture prepared with 2 μL of 3 M sodium acetate, 10 μL distilled water and 50 μL of 95% ethanol. This mixture was made up for 96 wells and 62 μL added to each well and agitated and incubated for 15 minutes in a dark chamber without exposing the plate to light. The plate was then centrifuged at 2,500 rpm for 30 minutes. The precipitation mix was removed by inverting the plate onto tissue towels and spinning at 300 rpm for 2 minutes. The wells were then washed with prechilled (-20°C) 70% ethanol and washed by inverting the plate onto the tissue towels at 300 rpm for 2 minutes. The plates were air dried in dark chamber after washing and covered with plastic adhesive sealers and wrapped in aluminium foil for transportation. The sequenced plates were sent to the Institute of Environmental Science and Research (ESR), Wellington, New Zealand, for sequence reading and allele assignment. The clonal complexes were assigned using a

Campylobacter MLST database.[3]

3.2.6 *flaA* and *por*A typing

Those *C. jejuni* isolates that had complete MLST profiles were further typed by sequencing genes associated with cell surface antigens. The internal fragments in the *fla*A short variable region (SVR) and the internal fragments of *por*A gene were amplified and sequenced for assigning nucleotide allelic numbers. The DNA from the MLST-typed isolates were prepared and quantitated using Nanodrop spectrophotometer, to 50 ng/μL of distilled water and transported under chilled conditions to ESR, Porirua, Wellington for typing. The primers for the antigenic typing were adapted from the PubMLST database and the allelic numbers were assigned by referring to the *Campylobacter* PubMLST database. The isolates that were difficult to amplify were further processed in the Hopkirk Research Laboratory using primers from the PubMLST database. This was done by optimising the PCR reaction mix ($MgCl_2$ optimised to 2.5 M and primers: 3.2 picomoles) and the PCR program as described in Appendix A.

3.2.7 Statistical analysis

To understand the relationship between *Campylobacter* spp. and different sampling locations, and over time, a linear model was used to predict the influence of space and time. Simple linear regression analysis was carried out where, the total number of samples collected per day, per site was 12 throughout the entire study period for both ducks and starlings. Site and time of sampling were used as explanatory variables. The inferences on the significance of space and time were drawn based on the probability values obtained from the model. Means of *Campylobacter* prevalence in ducks and starlings and the means of *C. jejuni* prevalence in ducks and starlings were compared and tested with the function of analysis of variance (ANOVA).

Population diversity : Rarefaction

Rarefaction was performed by using the frequency of STs (as described by Gormley et al. 2008) found in ducks and starlings. The frequency distribution of each *C. jejuni* ST was

[3]URL: (`http://pubmlst.org/campylobacter`)

summarised at the sampling sites, species and seasonal level in an effort to describe how the *C. jejuni* population varied according to sampling sites, species and time of the year. The analysis was carried out using the contributed R package Vegan (*R: A Language and Environment for Statistical Computing, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2010, ISBN 3-900051-07-0, http://www.R-project.org* n.d.) that contains a rarefaction function. In rarefaction analysis, the horizontal axis of the plot represents the number of samples used for analysis and the vertical represents the diversity or the number of sequence types identified in the specified number of samples. Diversity indices such as Simpson index 1-D and Shannon index were measured (as described in the PAST, PAleontological STatistics reference manual v2.13 by Hammer 2009-2011 [page number 40]) to analyse the *C. jejuni* population in ducks and starlings. Simpson index 1-D measures the evenness of the community that has a scale from 0 to 1, where 0 indicates that all taxa are equally present and 1 indicates that one taxon dominates the community completely. Whereas, Shannon index takes the number of individuals as well as number of taxa into account that varies from value 0 for communities with only a single taxon to accounting for high values or frequency and values above 0 for communities with many taxa, each with few individuals.

Population differentiation:

Analysis of molecular variance and Fst

Analysis of molecular variance (AMOVA) was performed to compare the effect of sampling site, host species and sampling period on the population differentiation and population structure of *C. jejuni* Arlequin v3.11 (Excoffier et al. 2005). In AMOVA, the genetic structure of a given population is analysed by an analysis of variance framework. AMOVA analyses the variance of allelic frequencies within and between populations or groups. Analysis at different levels, referred to as 'hierarchical analysis' is carried out in AMOVA which divides the total variance into different covariance components such as within population, within groups among populations and inter-population differences (Excoffier et al. 2005). While analysing the genetic structure for different hierarchical levels, three hierarchical F-statistics are derived known as the fixation indices (expressed as components of AMOVA). The fixation indices include: Fst, a fixation index that measures the variance among subpopulations relative to the total variance, Fsc, that measures

variance among subpopulations within groups and Fct, the variance among groups relative to the total variance. Fst quantifies the genetic differentiation among populations under comparison using an index ranging from 0 to 1. A zero value implies that there is no differentiation between populations and a value of one implies the two populations are completely separate (Wright 1965, 1978, 1984). Wright (1978) has further suggested the qualitative guidelines for the interpretation of Fst such as: (1)the range 0.0 to 0.05 may be considered as indicating small or limited genetic differentiation; (2) the range 0.05 to 0.15 indicates moderate genetic differentiation (3) the range 0.15 to 0.25 indicates great genetic differentiation and (4) the values of Fst above 0.25 indicate very great genetic differentiation. The genetic distance of *C. jejuni* populations found in ducks and starlings at different time periods and at different sampling sites were calculated using MLST allelic profiles. In turn, all these fixation indices facilitate inference on the gene flow between populations compared.

Minimum spanning tree Phylogenetic relationships among the STs were analysed using Bionumerics v6.1.[4] A minimum spanning tree (MST) was constructed using the allelic profile data set for ducks and starlings collected from March 2008 (the trial period included) to July 2009. MST is an alternative approach to show the relationships among isolates from bacterial populations (Prim 1957). A MST links the STs that are closely related within their lineages or CCs, identifying the most likely extant ancestral sequence type for the STs analysed. The most likely extant ancestral sequence types are known as the 'consensus' clones or the founder STs from which the clonal variants (STs) have descended (Feil & Chan 2001). The radial spread of STs from the consensus clones are reflected by a series of circles where the size of the circles represent the number of isolates per ST (Prim 1957). The construction of a MST involves three main steps:[5]

1. The sub-division of the data into 'clonal complexes'.

2. The identification of ancestral sequence types. Inferring likely patterns of descent within each clonal complex.

Clonal complexes are multilocus sequence types in which every sequence type shares at least 5 loci in common with at least one other member of the group. Ancestral sequence types (or 'consensus clones' or founder strains) are identified based on the seven

[4]Applied Maths URL: `http://applied-maths.com/bionumerics/bionumerics.htm`
[5]URL: (`http://pubmlst.org/analysis/burst/burst.shtml`)

loci where an ancestral sequence type differs from the highest number of other sequence types in the clonal complex at only one locus out of seven that defines the highest number of single-locus variants, or SLVs. Single-locus variants are identical to the ancestral sequence type at 6 loci, but differ at the seventh while the double locus variants (DLVs) vary at two loci. Once the ancestral sequence types are assigned, strains are assigned according to their relationships with their respective ancestral sequence types of the clonal complex, where all the SLVs and DLVs are associated with their corresponding ancestral sequence types. Typically, strains that share at least 5 loci in common are included in a clonal complex.

3.3 Results

3.3.1 Prevalence of *Campylobacter* spp. and *C. jejuni*

Campylobacter spp.

The overall prevalence of *Campylobacter* spp. in the sampled wild bird faeces was 37% (95% CI 35 to 40%; 542 of 1,458)[N.B. All confidence intervals were adjusted for clustering] (Table 3.2). The prevalence of *Campylobacter* varied by month and site as shown in Figure 3.2. There was a trend of relatively high prevalence during the spring and winter and low prevalence during the summer.

Table 3.2: Prevalence estimates of *Campylobacter* and *C. jejuni* in mallard ducks, starlings and geese

Species	n	G.pos. (%)	Conf. int (95%)	Cj.pos. (%)	Conf. int. (95%)
Overall	1,458	37	35 to 40	22	20 to 24
Ducks	716	30	26 to 33	23	20 to 26
Starlings	720	46	42 to 50	21	18 to 23
Geese	22	9	11 to 29	9	11 to 29

n : Number of sample.

G.pos. : *Campylobacter* positive.

Cj.pos. : *C. jejuni* positive.

Conf. int. : confidence interval.

The prevalence of *Campylobacter* spp. in starlings (46% CI 42 to 50%) was significantly higher than that of ducks (30% CI 26 to 33%). In ducks, the overall prevalence of

Campylobacter spp. during the summer was 32% (95% CI 27 to 37%) and the prevalence in winter months was 28% (95% CI 23 to 33%). The Esplanade sampling site showed the highest prevalence of *Campylobacter* spp. in ducks throughout the study period, followed by Memorial Park, Palmerston North Square, Massey University Concourse and the Hokowhitu sampling site. In starlings, there was a bimodal prevalence pattern with an overall summer prevalence of 38% (95% CI 33 to 43%) and winter prevalence of 53% (95 CI 48 to 59%) (Figure 3.2). The Palmerston North Square and The Esplanade sampling sites showed distinctive bimodal prevalence patterns with increased prevalence during early spring and winter and a decrease over summer. In contrast, prevalence patterns at other sampling sites were not as distinctive.

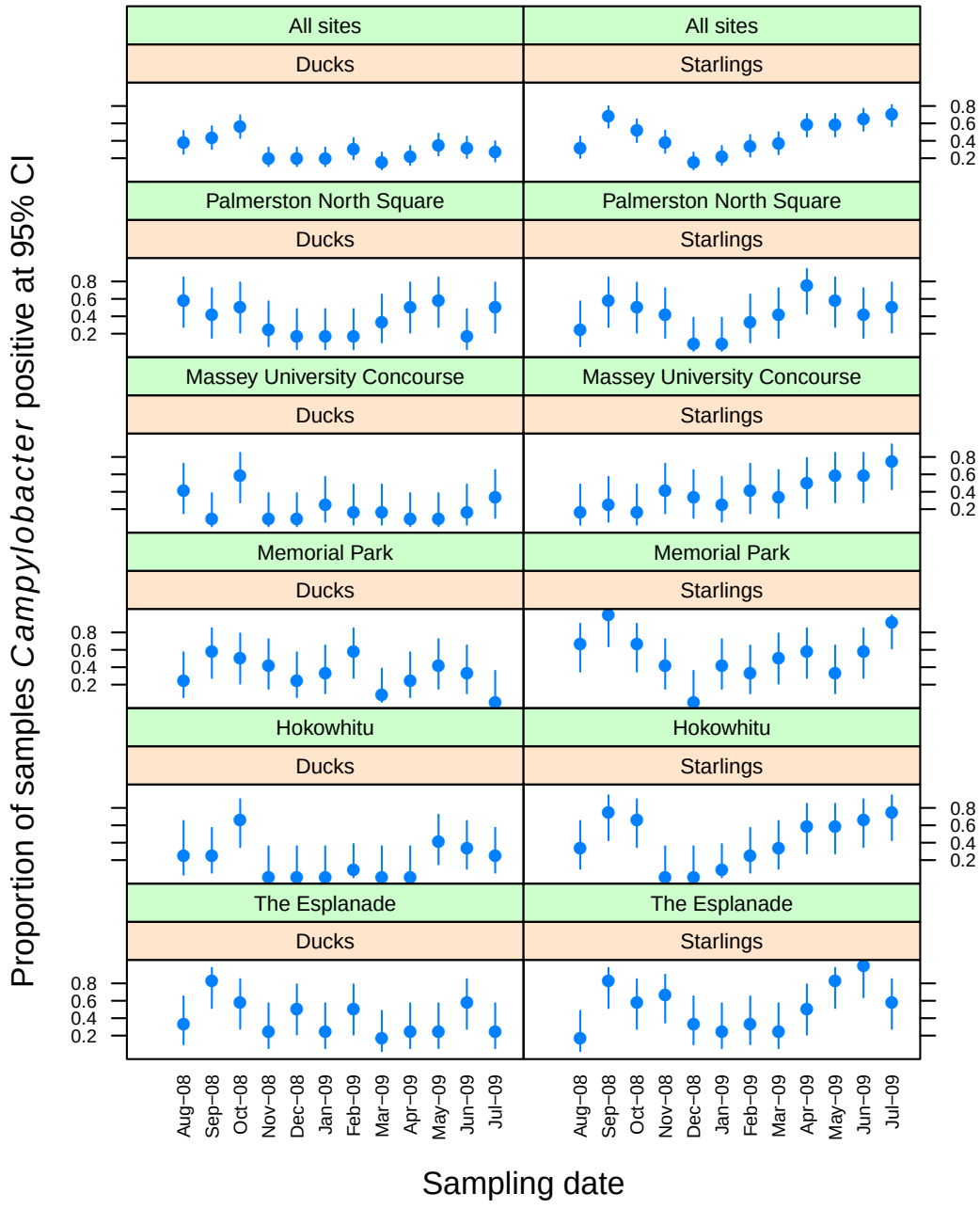


Figure 3.2: Prevalence of *Campylobacter* during sampling occasions and in sampling sites. The graph describes the prevalence of *Campylobacter* in the faecal materials of mallard ducks and starlings from August 2008 to July 2009 in each sampling site. CI in the figure refers to confidence interval.

C. jejuni isolates were recovered from 1,458 faecal samples from mallard ducks, starlings and geese giving an overall *C. jejuni* prevalence of 22% (95% CI 20 to 24%). The prevalence of *C. jejuni* in ducks, starlings and geese are provided in Table 3.2. The prevalence of *C. jejuni* showed significant differences by species (P = 0.008). There was no significant difference in the prevalence of *C. jejuni* between sampling sites (P = 0.07) but the prevalence did differ by sampling period (P = 0.01). Figure 3.3 shows the prevalence of *C. jejuni* by sampling periods and stratified by sampling site.

The prevalence of *C. jejuni* in both ducks and starlings showed a bimodal pattern. The prevalence was highest during the early spring (48%; 95% CI 36 to 61%) decreasing over the summer to reach a prevalence of 21% (95% CI 17 to 25%) during the winter. There was a distinctive bimodal prevalence pattern at the Palmerston North Square sampling site but at the other sampling sites temporal trends in prevalence were not very distinctive. In starlings, the prevalence of *C. jejuni* was highest in the early spring (45%; 95% CI 32 to 58%) across all sampling sites and the prevalence remained low throughout the winter (17%; 95% CI 13 to 21%). All of the sampling sites showed an increased spring prevalence except Massey University concourse. Likewise, The Esplanade showed an increased winter prevalence (30%; 95% CI 20 to 40%) that differed from all other sampling sites. In general, except Palmerston North Square, there were no distinctive temporal trends in prevalence in other sampling sites.

The prevalence of both *Campylobacter* and *C. jejuni* in Canadian geese was 9% (2 of 22), however geese were not included in the comparative analyses due to the small sample size. Further details on the number of samples and isolates of *Campylobacter* and *C. jejuni* for the entire study period are provided in Appendix A.

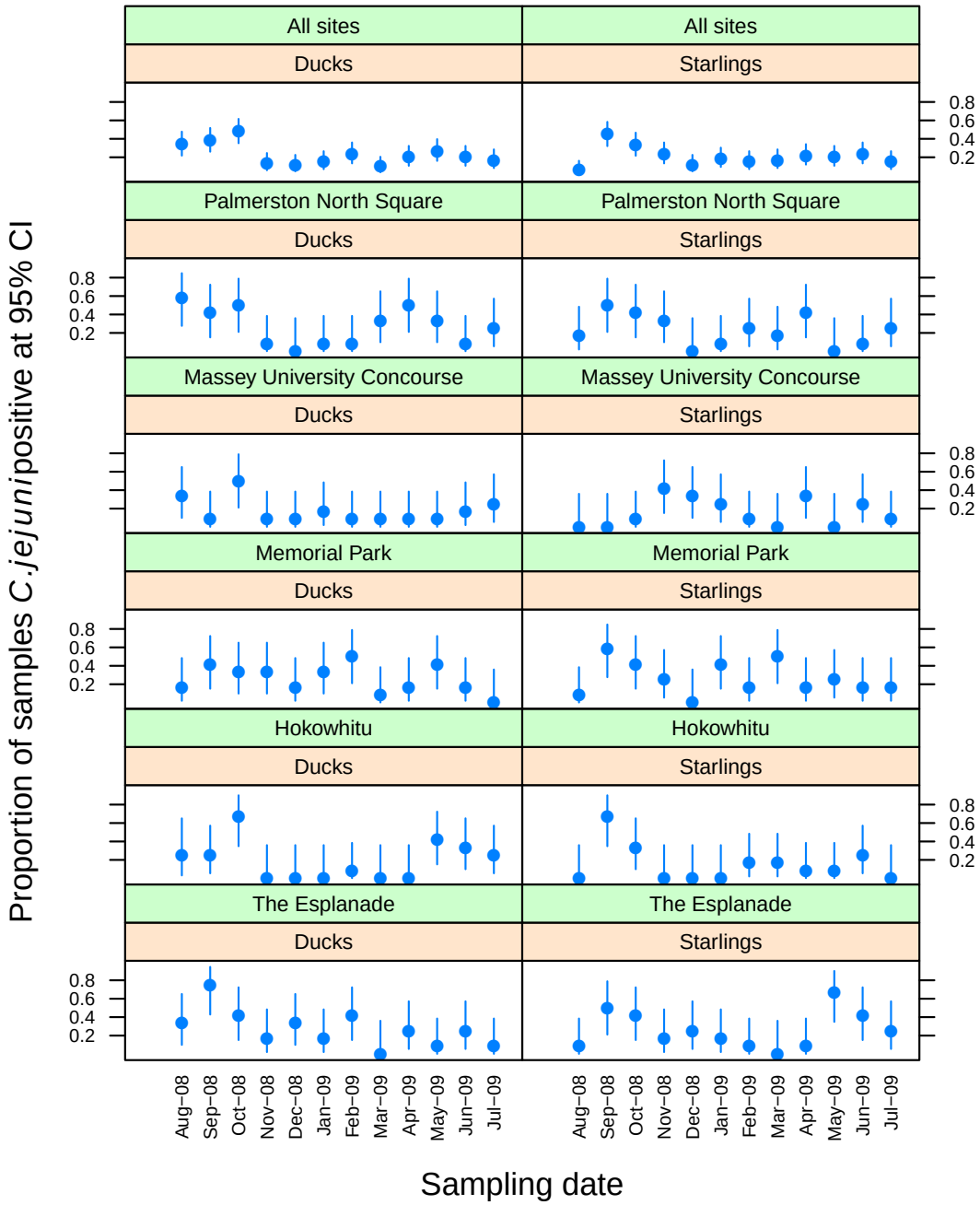


Figure 3.3: Prevalence of *C. jejuni* in mallard ducks and starlings. The graph describes the prevalence of *C. jejuni* in the faecal materials of mallard ducks and starlings during each sampling occasion and in each sampling site. CI in the figure refers to confidence interval.

3.3.2 Sequence type diversity by host

Forty three different sequence types (STs) were obtained by characterising 124 isolates from mallard ducks, starlings and geese, which were assigned to 11 clonal complexes (CC). The frequency of STs found in mallard ducks and starlings is shown in Figure 3.4 a. ST-45 was the most predominant sequence type found in the two species accounting for 21% (26 of 124) of the total isolates.

The most common CC in mallard ducks was ST-1034, accounting for 28% (18 of 64) of samples. The second largest was the ST-45 complex, accounting for 14% (9 of 64) of samples, followed by the ST-692 complex accounting for 11% (7 of 64) of samples. Twenty six of the 64 *C. jejuni* STs from mallard ducks could not be assigned to a CC (Unassigned complex = ST-U/A; to facilitate comparison, the STs not assigned to a clonal complex are referred to as U/A complex in this section)

The majority of the European starling isolates were assigned to the ST-45 complex, accounting for 28% (16 of 57) of starling isolates and the next largest complex was ST-177 (9%, 5 of 57). There were 23 isolates of the 57 isolates identified, belonging to different STs that were not assigned to a CC (ST-U/A) at the time of analysis. Other CCs were ST-677, ST-682, ST-692, ST-1034 and ST-42 each accounting for 4% (2 of 57 isolates) and ST-21, ST-1304 and ST-1332 accounted for 2% (1 of 57 isolates) each.

Three geese isolates were characterised, of which one belonged to ST-1034 complex and the other two STs could not be assigned to a complex. The STs that were present both in ducks and starling were ST-1324, ST-1342, ST-137, ST-2378, ST-3961, ST-45, ST-53, ST-583, ST-692, ST-991 and ST-992. Although ducks and starlings shared some STs in common, the frequency of each ST varied between these two species. The frequency distribution of STs in ducks and starlings is described in Figure 3.4.

Rarefaction curves showing the diversity of STs as a function of the number of isolates, by host are shown in Figure 3.5. In the plots shown in Figure 3.5 (a), a slope of zero in the rarefaction curves indicate that the maximum genetic diversity has been reached and that it is unlikely that more genetic diversity will be identified if more samples are analysed. A 95 % confidence interval with a bootstrap of 1000 random samples showed a Simpson index of 0.95 (95% CI 0.91 to 0.96) and a Shannon index of 3.3 (95% CI, 2.8 to 3.3) for *C. jejuni* population in ducks indicating that the *C. jejuni* population is more diverse in

ducks in comparison to the *C. jejuni* population in starlings (Simpson index = 0.9; 95% CI 0.9 to 0.95 and Shannon index = 2.8; 95% CI 2.8 to 3.3). Further details on the Simpson and Shannon indices are provided in Table A.9 in Appendix A. Although the confidence intervals of both the Simpson and Shannon indices overlap, the rarefaction analsyis does indicate a greater species richness in ducks.

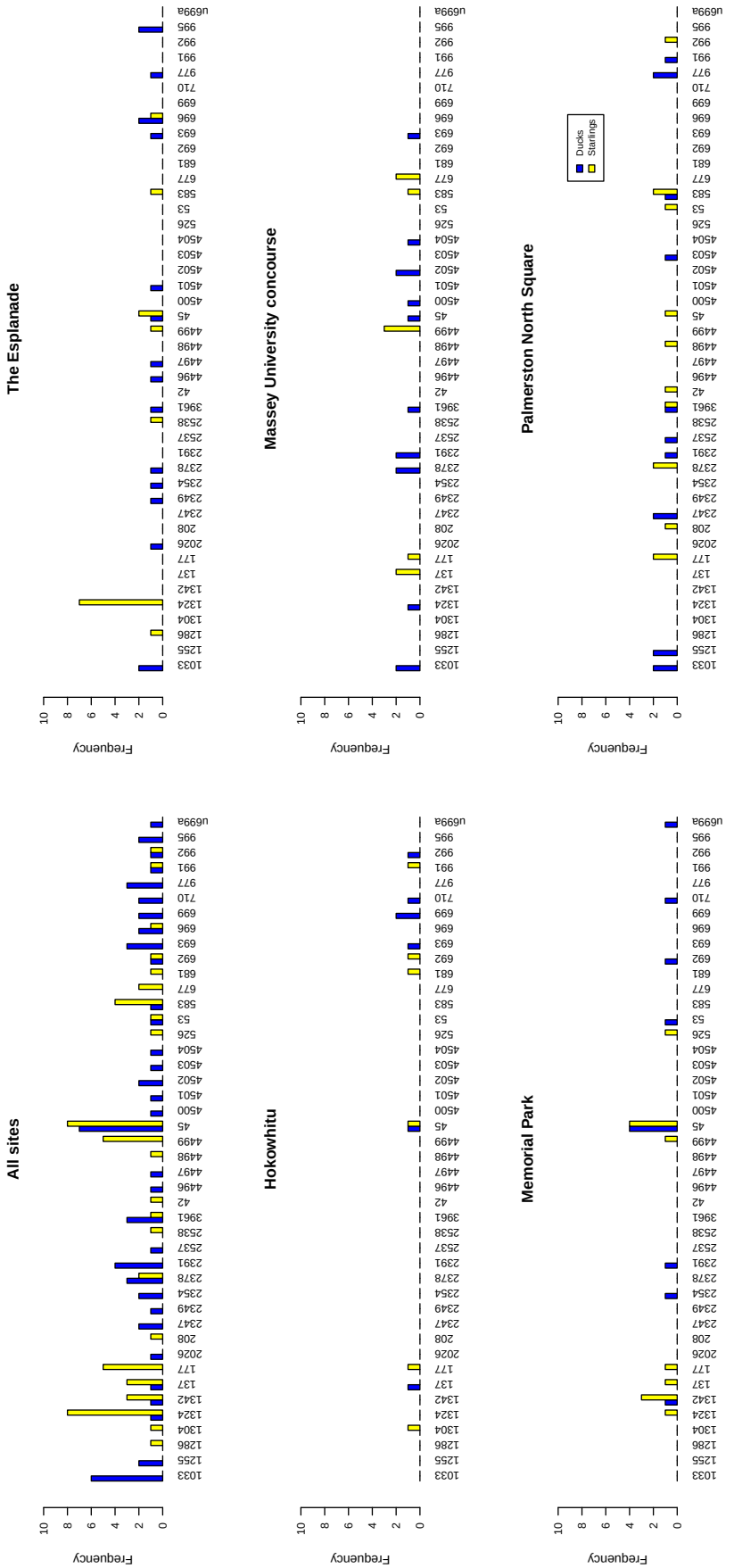
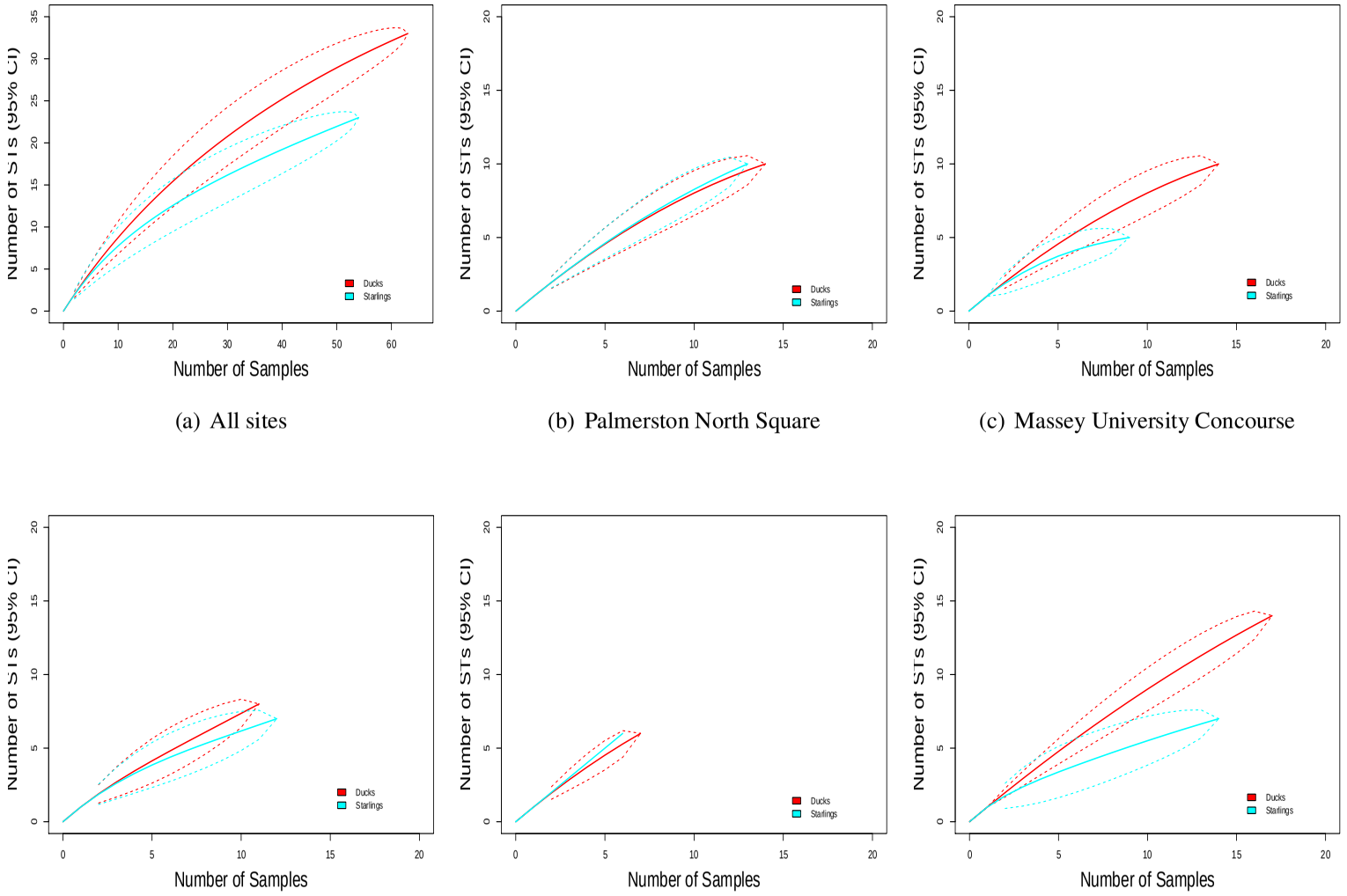


Figure 3.4: Frequency of *C. jejuni* sequence types (ST) in mallard ducks and starlings in different sampling sites.

3.3.3 Sequence type diversity by sampling site

The frequency of STs found in mallard ducks and starlings at different sampling sites is shown in Figure 3.4. Rarefaction curves of samples from ducks from The Esplanade showed more diversity, as indicated by taxonomic richness, in the *C. jejuni* ST population compared with starling samples from the same site (Figure 3.5 b). Starlings from the Hokowhitu site had relatively more diverse STs of *C. jejuni* although the number of isolates were small (starlings n = 6, ducks n = 7). The heterogeneity of *C. jejuni* STs was high in duck faecal material from the Massey University site compared with starling samples. Similarly, at Memorial Park the *C. jejuni* STs from ducks were more diverse compared with starlings. Genetic diversity at Palmerston North Square as similar for both ducks and starlings. Genetic diversity indices for all the sampling sites by species are provided in Table A.9 in Appendix A. Although the STs found in different sampling sites differed there was no evidence of population differentiation.



(a) All sites (b) Palmerston North Square (c) Massey University Concourse

(d) Memorial Park (e) Hokowhitu (f) The Esplanade

Figure 3.5: Species-based rarefaction curves of *C. jejuni* sequence types (ST) for mallard ducks and starlings. The genetic diversity of *C. jejuni* STs in mallard ducks and starlings is described in this graph . The overall population structure of *C. jejuni* sequence types in ducks is more diverse compared with starlings. The *C. jejuni* sequence types in ducks are more diverse than the starlings at all the sampling sites except The Palmerston North Square that shows a similar pattern of *C. jejuni* sequence type diversity

3.3.4 Sequence type diversity by sampling period

The prevalence of ST-45 complex was high in the summer (16%; 20 of 124 isolates) decreasing to 4% (5 of 124 isolates) in winter when estimated for the total number of isolates analysed in the study. The next predominant CC in the summer was ST-1034 which accounted for 8% (10 of 124 isolates), followed by other complexes such as ST-692 (3 of 124 isolates), ST-177 and ST-677 accounting for 2% each (2 of 124 isolates) and ST-1304 and ST-1332 each accounting for 1% each (1 of 124 isolates). The proportion of STs that were not assigned to a CC comprised 15% of the total number of isolates (18 of 124) in the summer. During the winter, the STs that were not assigned to a CC formed a highly prominent group that accounted for 25% (31 of 124) of isolates. The ST-U/A was followed by the ST-1034 (8%, 10 of 124 isolates), ST-692 (5%, 6 of 124 isolates), ST-45 (4%, 5 of 124 isolates), ST-177 (3%, 4 of 124 isolates), ST-21, ST-682 (2%, 2 of 124 isolates each) and finally by ST-403 (1%, 1 of 124 isolates).

The heterogeneity of the sequence types during summer and winter is shown in the rarefaction curves in Figure 3.7. It was observed that the diversity of the sequence types was high during winter in both species. However, the sequence types were more diverse in winter in ducks compared with starlings where, the starling faecal material showed only a little difference between summer and winter months in general. The frequency of the STs and their prevalence during different time periods are shown in Figure 3.6.

Distribution of sequence types in ducks and starlings in summer

The CCs that were prevalent during summer in ducks from all sites were ST-45 (n = 7), ST-U/A (n = 11), ST-1034 (n = 8) and ST-692(n = 3). The ST-45 complex was prevalent in September (n = 2), October (n = 2), November (n = 2) and December 2008 (n = 1). There were no isolates found in January and February 2009. The ST-1034 complex was found in November (n = 3), December 2008 (n = 3) and February 2009 (n = 2). ST-692 complex was found in January (n = 2) and February 2009 (n = 1). The STs of ST-U/A complex were isolated during September (n = 3), October 2008 (n = 2), January (n = 3) and February 2009 (n = 3).

The CCs that were found in starlings were ST-45 (n = 13), ST-U/A (n = 7), ST-1034 (n = 2), ST-177 (n = 2), ST-677 (n = 2), ST-1304 (n = 1), ST-1332 (n = 1). ST-45 was

found throughout the summer, however, the prevalence decreased gradually towards the end of summer (September 2008 = 2; October = 3; November = 5; December = 1; January 2009 = 1; February = 1). The ST-U/A complex was highly prevalent during the summer (October 2008 = 1; December = 3; January 2009 = 1; February = 1). In October 2008 (n = 4) and February 2009 (n = 4) there was more heterogeneity in STs among the summer months.

Distribution of sequence types in ducks and starlings in winter

ST-U/A complex formed the largest group (43%, 15 of 35 isolates) among the CCs that were found during the winter months in ducks. The STs of the ST-U/A complex were more frequent during the winter and particularly during the colder months of the year (May 2009 n = 3; June n = 6; July n = 2). The STs from ST-/U/A complex were also isolated during the trial periods of sample collection in April 2008 (ST-995), May 2008 (ST-3961, ST-696, ST-977), June 2008 (ST-977) and July 2008 (ST-2378) from ducks. The next predominant CC was ST-1034 (n = 10) that was found in March 2008, May and July 2009 followed by ST-692 (n = 4) which was found in May and July 2009. The ST-45 (n = 2), in contrast to summer, was a minor complex found in winter followed by ST-177, ST-21, ST-403 and ST-42 (n = 1 each).

The ST-U/A complex formed the major complex in starlings accounting for 55% (16 of 29 isolates). Although ST-U/A complex was found throughout winter, the prevalence was high in the colder months: August 2008 (n = 2), March 2009 (n = 2), April (n = 3), May (n = 3), June (n = 3) and July (n = 3). The other CCs were ST-177 (n = 3), ST-42 (n = 2), ST-45 (n = 3), ST-682 (n = 2), ST-692 (n = 2) and ST-21 (n = 1).

Of 124 isolates tested, it was found that isolates from a single sample or single bird; two samples from starlings and four from ducks belonged to the same sequence type while there were occasions where isolates from a single sample or single bird belonged to different sequence types; one sample from starlings and one from ducks (Appendix A, Table A.4).

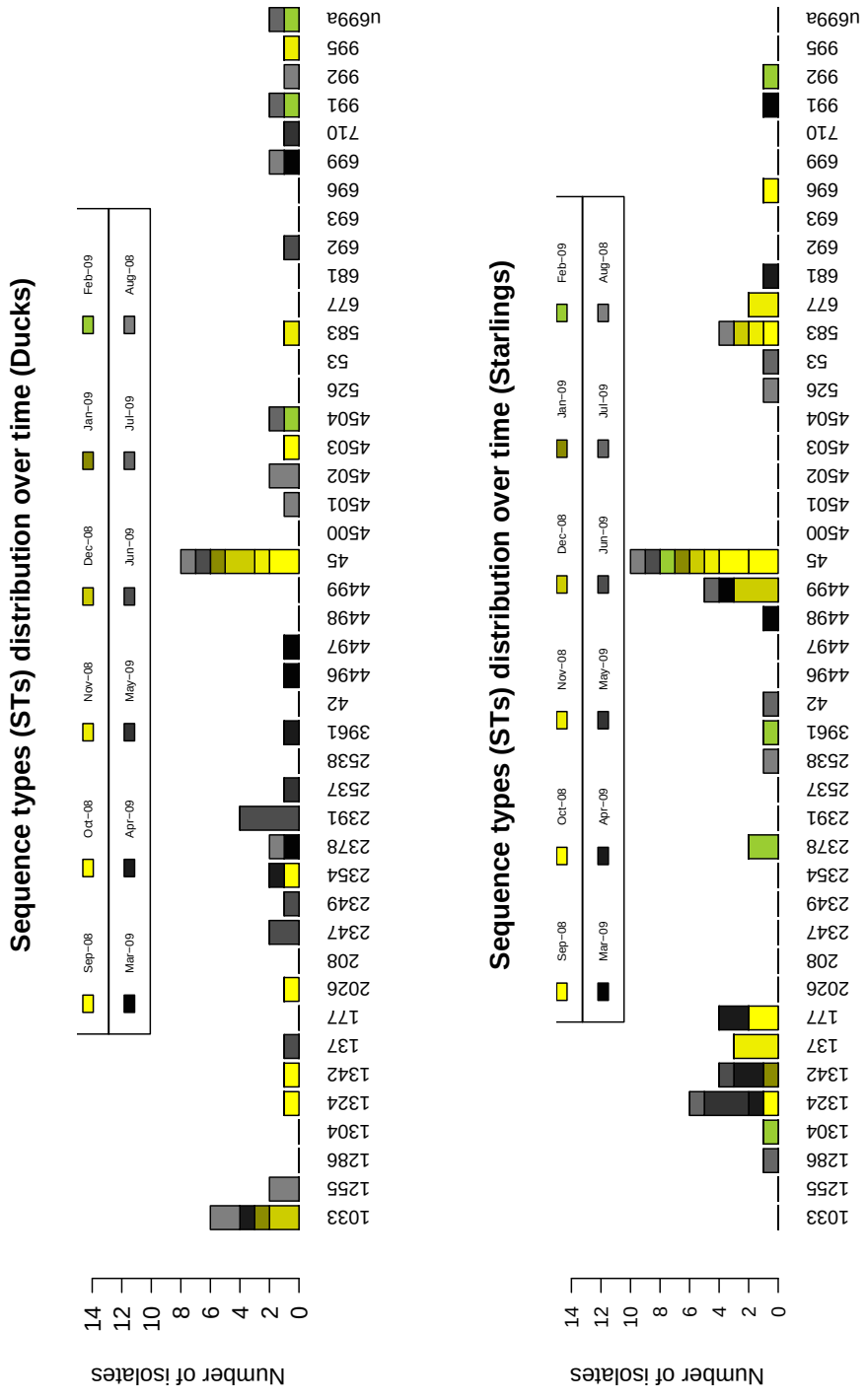


Figure 3.6: Distribution of *C. jejuni* sequence types (ST) in mallard ducks and starlings over time. The frequency distribution of *C. jejuni* STs during different time periods of sampling is represented in this graph.

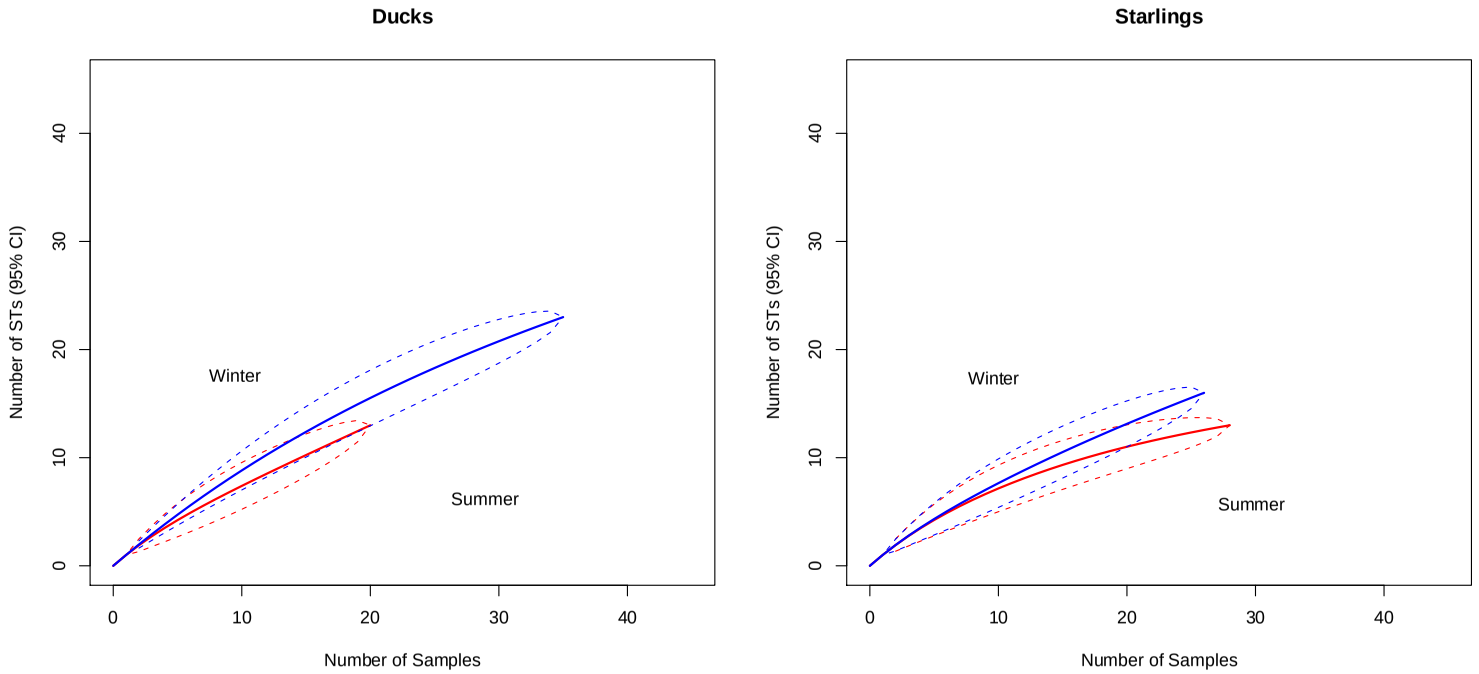


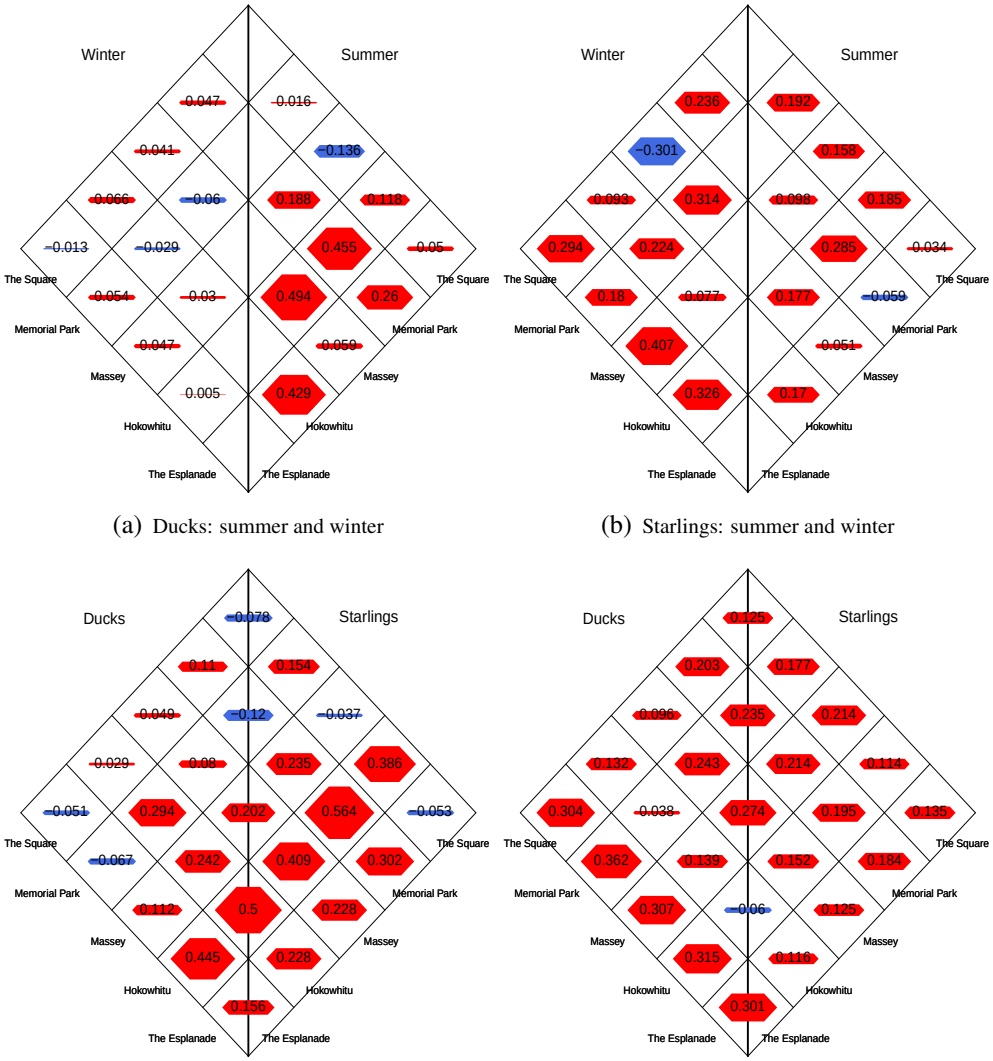
Figure 3.7: Species-based rarefaction curves of *C. jejuni* sequence types (ST) for mallard ducks and starlings. The genetic diversity of *C. jejuni* STs between mallard ducks and starlings is described in this graph during summer and winter months

3.3.5 Analysis of molecular variance and Fst

Pairwise Fsts were calculated using a hierarchical model (AMOVA) to estimate the components of variation to compare the genetic differentiation of *C. jejuni* populations between ducks and starlings from each sampling site and at different sampling periods. The figures 3.8 a, b, c and d show a diagrammatic representation of Fst values which denotes the population differentiation at different sampling sites and during different seasons in ducks and starlings. The overall Fst value between ducks and starlings was 0.08 (p value = 0.00), which means there is a significant level of differentiation between the *C. jejuni* populations of ducks and starlings.

AMOVA showed some evidence of differentiation between ducks and starlings, but there was no statistical evidence of differentiation between sites, within hosts (Fst values were very small). However, Table A.8 in Appendix A indicates that The Esplanade *C. jejuni* in winter were more different to other populations. Although there was no overall evidence for population differentiation between sites (Figure3.8 c), it does show some differentiation when stratified by season. The Hokowhitu site did appear to show stronger differences both within and between species in the summer but p values were border line significant (Table A.8 in Appendix A).

A minimum spanning tree (MST) was constructed to evaluate the relatedness among the *C. jejuni* STs isolated from the ducks, starlings and geese from the five sampling sites. The MST (Figure 3.9) identified the most likely extant ancestral sequence type for the STs analysed. The STs were partitioned based on the predominance of occurrence in ducks, starlings and geese. The partitioning identified the diversity in the STs from ducks and starlings which in turn indicated that the gene flow between these species is very limited.



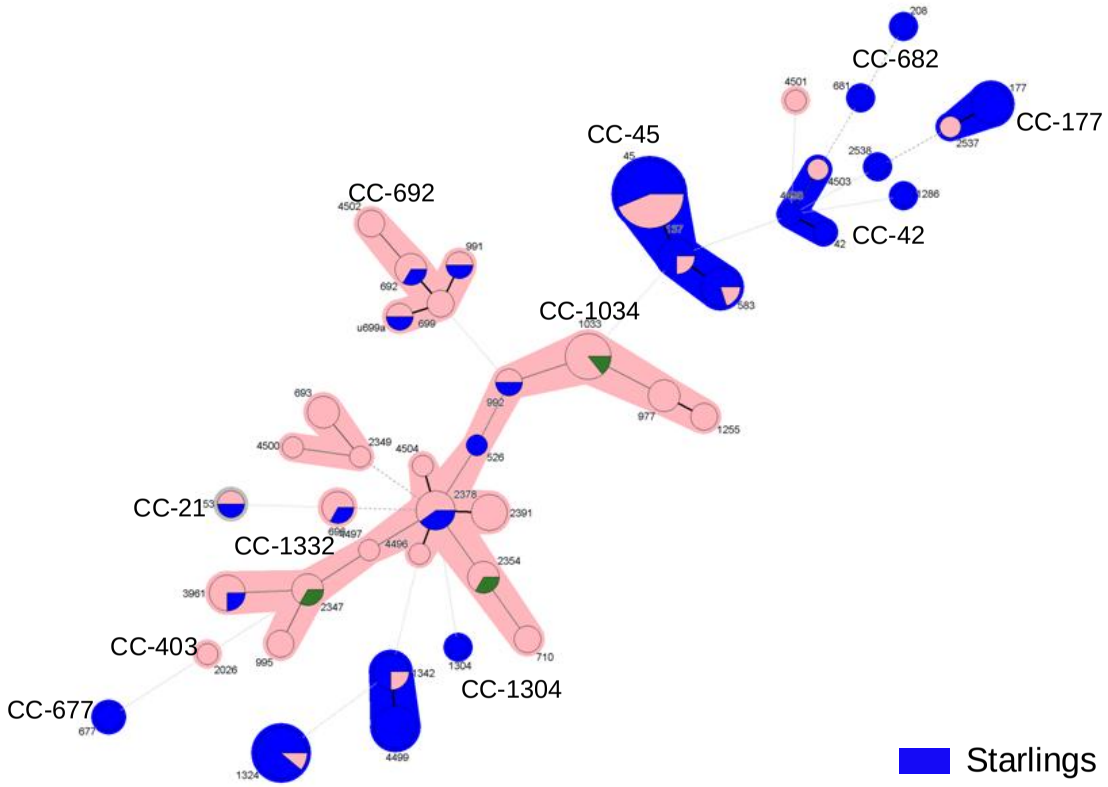
(a) Ducks: summer and winter

(b) Starlings: summer and winter

(c) Ducks and starlings: summer

(d) Ducks and starlings: winter

Figure 3.8: Population differentiation of *C. jejuni* population in ducks and starlings during different seasons and at different sites. The population differentiation of *C. jejuni* STs in mallard ducks and starlings is described in this graph. The Fst values of *C. jejuni* populations calculated using AMOVA are presented in a matrix format to show the population diversity across sampling sites and at different seasons in ducks and starlings. For example, in Figure (a), The Esplanade ducks are compared against Hokowhitu ducks, Massey ducks and others during the summer and winter seasons and Figure (b), (c) and (d) denote a similar pattern of comparison for the starlings at different sampling sites. It should be noted that the sample size of STs for this multilevel individual comparison was small, however, this figure is to identify the broader trends such as a relatively more differentiated *C. jejuni* population in ducks during the summer – Figure (a) and a relatively more differentiated *C. jejuni* population in The Esplanade starlings during the winter – Figure d. Further details on the Fst and p values are available in Appenix A



ST	CC
53	21
42	42
4498	
4503	
137	45
45	
583	
177	177
2537	
2026	403
677	677
208	682
681	
4502	692
692	
699	
991	
1033	1034
1255	
2378	
2391	
977	
1304	1304
696	1332
1286	U/A
1324	
1342	
2347	
2349	
2354	
2538	
3961	
4496	
4497	
4499	
4500	
4501	
4504	
526	
693	
710	
992	
995	
u699a	

Figure 3.9: Minimum spanning tree (MST) of *C. jejuni* sequence types (ST) from mallard ducks, starlings and geese. The STs are partitioned based on the occurrence of the sequence types in ducks, starlings and geese. The proportion of sequence types isolated from ducks, starlings and geese are shown in the respective STs represented by different colours. The table shows the STs and the CC they are assigned to and the STs that are unassigned (U/A)

3.3.6 *C. jejuni* cell surface antigens typing and diversity

All the 124 MLST characterised isolates were further typed by sequencing the genes associated with cell surface antigens, however despite repeated processing and several amplification attempts 23 of the alleles could not be sequenced for their antigens. The complete and incomplete MLST, the *fla*A and *por*A allelic profiles are provided in Appendix A. A total of 32 different *por*A nucleotide alleles were characterised from 101 MLST characterised *C. jejuni* isolates. *por*A nucleotide allele 8 was the predominant allele in ducks (8 of 58) and *por*A nucleotide allele 44 was found in equal numbers in both ducks and starlings (5 in ducks and 5 in starlings of a total of 101 isolates). *por*A nucleotide allele 970 was the most common allele among starlings accounting for 20% (9 of 44) isolates. There was a single *por*A nucleotide allele from the starlings that has yet to be assigned an allelic number. The *fla*A nucleotides were more diverse compared with the *por*A nucleotide alleles. Thirty seven different *fla*A nucleotide alleles were characterised from 101 MLST typed isolates from ducks and starlings. The duck *fla*A nucleotide alleles were more diverse (26 different alleles) than the alleles from starlings (17 alleles). The *fla*A nucleotide alleles 1170, 1219, 1221, 1222, 1235, 16, 209, 213, 219, 520, 56, 69, 73, 787, 85 and 89 were unique to ducks. The MLST allelic profiles, *fla*A and *por*A nucleotide allelic numbers of *C. jejuni* isolates of ducks and starlings are summarised in Tables 3.3 and 3.4.

3.3.7 Comparison with wider population of sequence types in the PubMLST database

The *C. jejuni* ST populations were compared and/or matched with global *C. jejuni* ST populations available in the PubMLST database. Of 43 *C. jejuni* STs, ten (ST-2538, ST-3961, ST-4496, ST-4497, ST-4498, ST-4499, ST-4500, ST-4501, ST-4502, ST-4503, ST-4504) were novel and were submitted to the *Campylobacter jejuni* PubMLST website. ST-991, ST-992, ST-995 isolated from both ducks and starlings in our study, have also been recovered from environmental water and wild birds in other parts of the world (Canada and Sweden). ST-208 (which has been isolated from a starling in this study) has been isolated from a bathing beach in the United Kingdom. ST-1304, ST-1324 and ST-1342 have been isolated from the environment (wild bird) in Sweden, whereas ST-1324 and

ST-1342 were isolated predominantly from starlings in our study.

In the United Kingdom ST-1033 has been predominantly isolated from geese, whereas in our study this ST was highly prevalent in ducks (5% of the total number of isolates). ST-1255 was isolated from a wild bird and a goose in Sweden and the United Kingdom respectively, and this isolate has also been reported to have caused sporadic outbreaks of disease in The Netherlands in 2002. In this study ST-137 was prevalent only in ducks however, it has been isolated from a range of species in other countries including starlings (United Kingdom), chickens (United Kingdom, Canada, Scotland, Senegal), sheep (United Kingdom), turkeys (USA) and water (Canada). This ST has also accounted for numerous sporadic campylobacteriosis cases in the United Kingdom, Scotland, Belgium, England, The Netherlands and Canada. ST-177 was prevalent only in starling faecal material in this study which in agreement as speculated in one of the previous New Zealand reports (French et al. 2009a). This ST has been isolated from starlings and bathing beaches in the United Kingdom in large numbers and in Spain from environmental water. ST-2026 was isolated from duck faecal material in this study and has been reported from a sporadic campylobacteriosis case in England which is the only record available for this ST in the PubMLST database.

ST-692, ST-693, ST-696, ST-699 and ST-710 were prevalent both in starlings and ducks. In comparison with the wider population, these STs were reported predominantly only from geese in the United Kingdom with an exception of ST-696 which was also isolated from cattle and a turkey in Germany. ST-696 was implicated in human campylobacteriosis cases in The Netherlands. ST-677 has been reported from few sources like cattle, environmental water, wild birds, starlings and chickens in the United Kingdom and has also been reported from sporadic human campylobacteriosis cases in the United Kingdom. Likewise, ST-583 has been isolated from a variety of sources including wild birds, turkeys, cattle and chickens from the United Kingdom, Denmark and Canada and has been implicated in sporadic cases of disease in humans predominantly in The Netherlands, England, Belgium and Germany. ST-42 is one of the predominantly reported STs in cattle, sheep, goats and chickens in the United Kingdom, The Netherlands, South Africa, USA, Australia and Germany and this ST has also been isolated from wild birds and environmental water in the United Kingdom. This ST, isolated exclusively from starlings in this study has been implicated as a cause of disease in humans both in New Zealand and overseas.

A small number of sequence types from ducks and starlings in this study (ST-2347, ST-2349, ST-2354 and ST-2391) were identical to the novel STs from environmental water reported previously by ESR in Wellington in 2001. In New Zealand ST-2379 has been previously isolated from a chicken and ST-2537 and ST-2538 from wild birds. All of these STs were prevalent in ducks and starlings except ST-2358, which was isolated only from starling faecal material in this study. On previous occasions ST-2378 has been isolated from chicken meat in New Zealand whereas in this study ST-2378 was found in both ducks and starlings. ST-526 has only previously been reported in Australia from a human gastroenteritis case in 1999 and no reports of this ST have been made since. This is the first time that ST-526 has been reported in New Zealand.

The results of comparison of STs from this study with the global population of *C. jejuni* STs, are tabulated in Table 3.5.

Table 3.3: MLST allelic profiles, *fla*A,*por*A nucleotide alleles and their frequency in mallard ducks

Frequency	*fla*A	*por*A	ST	CC	*asp*A	*gln*A	*glt*A	*gly*A	*pgm*	*tkt*	*unc*A
1	32	27	53	21	2	1	21	3	2	1	5
1	X	X	4503	42	1	2	9	5	5	9	21
2	21	44	45	45	4	7	10	4	1	7	1
1	85	53	137	45	4	7	10	4	42	7	1
1	21	14	45	45	4	7	10	4	1	7	1
1	X	14	45	45	4	7	10	4	1	7	1
1	X		583	45	4	7	10	4	42	51	1
1	2	14	45	45	4	7	10	4	1	7	1
2	8	44	45	45	4	7	10	4	1	7	1
1	1235	971	2537	177	17	2	8	5	8	2	143
1	1170	132	2026	403	10	1	16	19	10	5	7
1	571	6	991	692	37	52	57	26	107	29	23
1	89	6	699	692	37	52	57	26	129	29	23
1	1067	6	692	692	37	52	57	26	127	29	23
2	209	180	4502	692	37	52	21	388	127	29	23
1	X	X	692	692	37	52	57	26	127	29	23
1	1067	152	2378	1034	2	15	4	48	356	25	23
1	16	792	1033	1034	2	61	4	64	126	7	23
4	56	879	2391	1034	2	15	4	48	360	25	23
4	22	973	1033	1034	2	61	4	64	126	7	23
2	1222	975	1255	1034	22	146	4	64	74	25	23
2	X	X	1033	1034	2	61	4	64	126	7	23
1	15	6	977	1034	22	61	4	64	74	25	23
1	89	6	977	1034	22	61	4	64	74	25	23
1	X	X	977	1034	22	61	4	64	74	25	23
1	NEW	X	2378	1034	2	15	4	48	356	25	23
1	15	6	696	1332	2	1	4	28	58	25	58
1	56	886	696	1332	2	1	4	28	58	25	58
2	56	180	u699a	U/A	37	NEW	57	26	129	29	23
1	1067	152	4496	U/A	2	15	98	48	356	25	23
1	73	879	4497	U/A	2	4	4	48	358	25	280
2	56	886	3961	U/A	2	29	4	27	10	25	24
1	1219	209	2354	U/A	37	4	4	48	13	25	23
2	69	888	710	U/A	37	29	75	48	126	25	23
1	787	180	2349	U/A	2	59	4	48	131	24	57
2	1221	888	2347	U/A	2	4	4	105	10	25	57
1	X	188	693	U/A	2	29	4	48	13	24	57
1	520	188	4500	U/A	2	29	296	48	131	25	57
1	213	819	693	U/A	2	29	4	48	13	24	57
1	NEW	X	4501	U/A	237	2	254	340	433	349	290
1	219	188	992	U/A	2	59	4	27	126	29	23
1	1236	974	1324	U/A	99	128	91	125	170	146	111
1	22	973	2354	U/A	37	4	4	48	13	25	23
1	21	44	1342	U/A	98	122	98	125	180	150	113
2	X	152	995	U/A	2	4	84	105	126	25	57
1	X	X	4504	U/A	2	15	4	48	356	150	23
2	15	6	3961	U/A	2	29	4	27	10	25	24
1	X	152	995	U/A	2	4	84	105	126	25	57

Table 3.4: MLST allelic profiles, *fla*A,*por*A nucleotide alleles and their frequency in starlings.

Frequency	*fla*A	*por*A	ST	CC	*asp*A	*gln*A	*glt*A	*gly*A	*pgm*	*tkt*	*unc*
1	321	180	u699a	U/A	37	NEW	57	26	129	29	23
2	1067	152	2378	1034	2	15	4	48	356	25	23
1	1060	632	1304	1304	100	142	93	135	190	145	81
1	X	X	696	1332	2	1	4	28	58	25	58
3	322	971	177	177	17	2	8	5	8	2	4
1	X	X	177	177	17	2	8	5	8	2	4
1	341	236	177	177	17	2	8	5	8	2	4
1	X	X	53	21	2	1	21	3	2	1	5
1	1354	1074	4498	42	1	2	3	4	5	9	21
1	239	71	42	42	1	2	3	4	5	9	3
1	1237	970	45	45	4	7	10	4	1	7	1
1	21	14	45	45	4	7	10	4	1	7	1
4	X	X	45	45	4	7	10	4	1	7	1
1	1421	1185	583	45	4	7	10	4	42	51	1
2	2	44	45	45	4	7	10	4	1	7	1
1	239	73	583	45	4	7	10	4	42	51	1
1	21	44	45	45	4	7	10	4	1	7	1
1	8	44	45	45	4	7	10	4	1	7	1
1	X	X	137	45	4	7	10	4	42	7	1
1	321	53	137	45	4	7	10	4	42	7	1
1	1237	276	137	45	4	7	10	4	42	7	1
2	X	X	677	677	10	81	50	99	120	76	52
1	405	203	681	682	35	43	9	5	8	46	21
1	X	NEW	208	682	26	2	9	51	8	46	5
1	571	6	991	692	37	52	57	26	107	29	23
1	22	6	692	692	37	52	57	26	127	29	23
1	X	X	1342	U/A	98	122	98	125	180	150	113
1	21	14	992	U/A	2	59	4	27	126	29	23
1	X	X	3961	U/A	2	29	4	27	10	25	24
6	1237	970	4499	U/A	98	359	98	125	180	150	113
5	1236	974	1324	U/A	99	128	91	125	170	146	111
2	1237	970	1342	U/A	98	122	98	125	180	150	113
2	1236	972	1324	U/A	99	128	91	125	170	146	111
1	491	NEW	1286	U/A	95	2	94	127	172	144	114
1	X	X	2538	U/A	35	2	8	51	361	2	21
1	1421	1186	526	U/A	2	15	4	27	13	80	23
2	X	X	1324	U/A	99	128	91	125	170	146	111

Note: U/A = Unassigned to a clonal complex;

New = New allele for which an allelic number has not been assigned at the time of analysis

Table 3.5: Sequence types (STs) compared with the wider population of sequence types in the PubMLST database

STs found in the study		Sources compared											
ST	Species	Chicken	Turkey	Goose	Starling	Duck	Wild bird	Sheep	Cattle	Pig	Pets	Water	Human
1033	Ducks	-	-	+++	-	-	-	-	-	-	-	-	-
1255	Ducks	-	-	+	-	-	+	-	-	-	-	-	+
1286	Starlings	-	-	-	-	-	+	-	-	-	-	-	-
1304	Starlings	-	-	-	-	-	+	-	-	-	-	-	-
1324	Both	-	-	-	-	-	+	-	-	-	-	-	-
1342	Both	-	-	-	-	-	+	-	-	-	-	-	-
137	Both	++	+	+	+	-	-	+	-	-	-	+	+++
177	Starlings	-	-	-	+++	-	-	-	-	-	-	+	-
2026	Ducks	-	-	-	-	-	-	++	-	-	-	-	+
208	Starlings	-	-	-	-	-	-	-	-	-	-	+	-
2347	Ducks	-	-	-	-	-	-	-	-	-	-	+*	-
2349	Ducks	-	-	-	-	+*	-	-	-	-	-	-	-
2354	Ducks	-	-	-	-	+*	-	-	-	-	-	-	-
2378	Both	+*	-	-	-	-	-	-	-	-	-	-	-
2391	Ducks	-	-	-	-	-	-	-	-	-	-	+*	-
2537	Ducks	-	-	-	-	-	+*	-	-	-	-	-	-
2538	Starlings	-	-	-	-	-	+*	-	-	-	-	-	-
3961	Both	-	-	-	-	-	+**	-	-	-	-	-	-
42	Starlings	++	-	-	+	-	+	+++	+++	-	-	-	+++
4496	Ducks	-	-	-	-	+**	-	-	-	-	-	-	-
4497	Ducks	-	-	-	-	+**	-	-	-	-	-	-	-
4498	Starlings	-	-	-	+**	-	-	-	-	-	-	-	-
4499	Starlings	-	-	-	+**	-	-	-	-	-	-	-	-
45	Both	+++	+	-	-	-	+	+	++	-	+	+	+++
4500	Ducks	-	-	-	+**	-	-	-	-	-	-	-	-
4501	Ducks	-	-	-	+**	-	-	-	-	-	-	-	-
4502	Ducks	-	-	-	+**	-	-	-	-	-	-	-	-
4503	Ducks	-	-	-	+**	-	-	-	-	-	-	-	-
4504	Ducks	-	-	-	+**	-	-	-	-	-	-	-	-
526	Starlings	-	-	-	-	-	-	-	-	-	-	-	+
53	Both	+	-	-	+	-	+	+	+	-	-	-	+++
583	Both	+++	-	-	-	-	+	-	+	-	-	+	+++
677	Starlings	+	-	-	+	-	+	-	+	-	-	+	++
681	Starlings	+	-	-	-	-	-	-	-	-	-	-	-
692	Both	-	-	++	-	-	-	-	-	-	-	-	-
693	Ducks	-	-	+++	-	-	-	-	-	-	-	-	-
696	Starlings	-	+	+++	-	-	-	-	-	-	-	-	++
699	Ducks	-	-	+++	-	-	-	-	-	-	-	-	-
710	Ducks			+++	-	-	-	-	-	-	-	-	+
991	Both	-	-	-	-	-	-	-	-	-	-	+	-
992	Both	-	-	-	-	-	+	-	-	-	-	+	-
995	Ducks	-	-	-	-	-	+	-	-	-	-		-

+ - Number of reports in the database: Less than five.

++ - Number of reports in the database: Above five but under fifteen.

+++ - Number of reports in the database: Over fifteen.

+* - First reported in New Zealand

+** - Novel to the study

3.4 Discussion

Birds are considered potential natural reservoirs of *Campylobacter* spp. (Lee & Newell 2006) and particularly, wild birds have been identified as vectors of transmission of *Campylobacter* spp. to other species such as poultry, cattle and humans (Craven et al. 2000, Studer et al. 1999, Waldenstrom et al. 2002). The primary aim of this study was to estimate the prevalence of *Campylobacter* spp. and *C. jejuni* in ducks and starlings at different time periods and sites. These sites may represent a source of exposure for *Campylobacter* spp. and *C. jejuni* for the human population in Palmerston North. A secondary aim was to analyse the population structure of *C. jejuni* and to determine the population differentiation for different host species at different sampling sites and at different sampling periods.

The overall prevalence of *Campylobacter* spp. was 37% (95% CI 35 - 40%; 542 of 1,458) which is relatively high compared with estimates for migrating birds (22%) and relatively low compared with estimates for aquatic birds feeding on invertebrates (50%) as reported by Waldenstrom et al. (2002). In the European starlings in this study, the prevalence of *Campylobacter* spp. (48%) was relatively high compared with previous prevalence estimates of 33% to 40% (Waldenstrom et al. 2002, Colles et al. 2008, Hughes et al. 2009). It should be noted however that the prevalence of *Campylobacter* spp. in starlings is generally higher compared with other avian species (Colles et al. 2008, Hughes et al. 2009).

In contrast, the prevalence of *C. jejuni* in ducks (23%) and starlings (21%) were relatively low compared with previous prevalence estimates in *Anatidae* (ducks, 30.6%) and European starlings (29.9%) (Waldenstrom et al. 2002, Brown et al. 2004, Colles et al. 2008). Differences in these prevalence estimates may be due to several factors such as the risk of becoming colonised by the bacteria, sampling procedures, sample size and sensitivity of culture techniques. Further, differences may also arise due to the type of samples examined, for example, intestinal samples have been found to give a higher isolation rate compared with faecal samples (Stanley et al. 1998*a*). The bimodal seasonal prevalence pattern in *Campylobacter* spp. and *C. jejuni* in ducks and starlings has not been previously reported. The highest early spring prevalence in ducks was in agreement with one recent study (Colles et al. 2009), although that particular study estimated *Campylobacter*

spp. prevalence in swan and geese (not ducks). Few other studies (Waldenstrom et al. 2002, Broman et al. 2004) identified a greater level of shedding of *Campylobacter* spp. in the autumn. However, this trend of a seasonal bimodal prevalence was evidenced in drug resistant *E. coli* (Williams et al. 2011). This longitudinal study on drug resistant *E. coli* was conducted in the United Kingdom between two sympatric populations of bank voles and wood mice to study the seasonal cycle of drug resistant *E. coli*, where the authors found a peak prevalence in early- to mid-summer in mice whereas in voles, they found a late summer and early autumn peak.

There was marked seasonality in the prevalence of *C. jejuni* STs both in ducks and starlings. The observation of high prevalence of the ST-45 complex in spring and early summer was consistent with other reports and also was coincidentally consistent with increased human campylobacteriosis cases in summer (Colles et al. 2003, French et al. 2005, Colles et al. 2008, Sopwith et al. 2008, Colles et al. 2009, French et al. 2009a). Reports of the seasonality of *C. jejuni* populations are few and are limited to a small number of studies involving either starlings (French et al. 2009a) or geese (Colles et al. 2008). In this study the majority of STs from ducks (n = 24), starlings (n = 21) and geese (n = 2) remained unassigned (ST-U/A) to a complex which was in agreement with those identified in previous studies (French et al. 2005, Colles et al. 2008, French et al. 2009a).

The MLST data demonstrated high genetic diversity within the *C. jejuni* isolates from ducks, much more so than with starlings. The rarefaction analyses and the diversity indices such as Simpson index 1-D. and Shannon index showed that the *C. jejuni* populations of ducks and the starlings differ, indicating that the gene flow between these two species is limited. The population of *C. jejuni* STs was more diverse in winter and the ST-U/A complex formed the major group at this time. We hypothesise that the increased diversity and the high prevalence of the ST-U/A complex may be due to the flocking behaviour of birds at different times of the year. Aggregation of birds from different areas is thought to result in mixing of different strains of bacteria from different geographical locations. It may be hypothesised the physiological changes that occur in the avian gut at different times of the year (related to, for example, diet and feeding habits) would alter the composition of the gut micro-flora which, in turn, could influence the overall genetic diversity of *C. jejuni* in these species.

The MLST data from ducks and starlings further suggested the presence of a host associ-

ation between the *C. jejuni* STs and the species they were isolated from. For example, the ST-1034, ST-692 and ST-1332 complexes were the most predominant lineages found in ducks. Interrogation of the *Campylobacter* MLST database and previous reports (Colles et al. 2008) showed that these three complexes have only been reported from geese. However, there were two isolates from the ST-1034 complex and one isolate from the ST-1332 complex from European starlings in this study. Likewise, in starlings, the ST-177 complex formed the major lineage which is in agreement with the PubMLST database and previous reports on European starlings (Colles et al. 2008, 2009, French et al. 2009a). Still there is some controversy about the host association between *C. jejuni* sequence types and wild birds (Hughes et al. 2009). At the same time, authors that were questioning the association between *C. jejuni* sequence types and wild birds (Colles et al. 2003) have subsequently documented a relationship (Colles et al. 2008, 2009). In addition, French et al. (2009a) raised a possibility that, the resemblance between the clonal complexes found in New Zealand and the United Kingdom is suggestive of a common ancestor that originated from the European birds introduced by the 'Acclimatisation Societies' in the late 19th century into New Zealand. This present study lends support to the speculation made by French et al. (2009a) where, the majority of the sequence types isolated in this study are also found in Europe as outlined in the Results 3.3.7. Further, the heterogeneity in the *C. jejuni* populations and the diversity in the cell surface antigens may be due to the aquatic feeding habits of the ducks and the possibility that ponds could contain a variety of contaminants including the faecal materials of non-aquatic birds.

The molecular variance analyses (Figures 3.8a, b, c, and d) showed that variation resides both between species and between seasons. There was evidence of inter-site transmission and there was significant host-association. More importantly, apart from the population differentiation, it was evident that Fst values of sites that possessed STs of zoonotic importance did not show wide differences between them. Particularly during the summer, *C. jejuni* STs from starlings (ST-45, ST-137, ST-583, ST-677) that cause human infection were found in all sampling sites, which implies that these sites could be a potential source of infection for humans. Furthermore, ST-45 has been found to be strongly associated with the early summer seasonal peak of human campylobacteriosis and is well recognised to colonise domestic poultry and to survive outside the host compared with other STs (Sopwith et al. 2006, Karenlampi et al. 2007, Sopwith et al. 2008, French et al.

2009a, Habib et al. 2010). It may be hypothesised that the early summer seasonal peak in human camplyobacteriosis may be partly due to direct environmental exposures as well as the already established food-borne pathways. Additionally, the diversity in the *fla*A and the *por*A alleles found in this study from ducks and starlings supports previous observations that *C. jejuni* has a capacity to readily undergo homologous recombination of its virulence-associated genes (Nuijten et al. 1990, Wassenaar et al. 1995, Harrington et al. 1997, Nuijten et al. 2000).

In conclusion, this study has quantified the prevalence, population structure and genetic diversity of *C. jejuni* in mallard ducks and starlings at different sampling sites and time periods in an urban area in provincial New Zealand. We have demonstrated similarities between the *C. jejuni* sequence types isolated from ducks and starlings and humans which implies that ducks and starlings may play a role in human infection. Although the overall influence of ducks and starlings on human campylobacteriosis case numbers is likely to be small, we hypothesise that they may represent a risk for small children using contaminated public areas for play. The population structure and diversity of *C. jejuni* populations in ducks and starlings indicate some localised transmission between the two species and the presence of sequence types specific to ducks and starlings reflect host association. Further, these observed variations reflect the genetic dynamism of *C. jejuni* in terms of their capacity to recombine and also the evolutionary events due to migration and mixing of different species of animals or birds. As *Campylobacter* is a rapidly evolving species mainly through recombination Wang & Taylor (1990), Dingle et al. (2001) and intense purifying selection (de Boer et al. 2002, Sheppard et al. 2009) further investigations to determine the prevalence and population genetics of *C. jejuni* sequence types in wild birds in other parts of the country will provide a greater understanding of this organism's genetic diversity which, in turn, will be of importance for reducing the burden of wild bird-associated campylobacteriosis in humans.

CHAPTER 4

Characterisation of *Campylobacter jejuni* from pets in New Zealand by combining multilocus sequence typing (MLST), *por*A and *fla*A typing

Abstract – *Campylobacter* spp. is the major cause of bacterial gastroenteritis worldwide and pet ownership has been identified as a risk factor contributing to the burden of human disease. A repeated cross-sectional study was conducted over twelve months in Palmerston North, New Zealand, to estimate the prevalence of *C. jejuni* in dogs and cats and to examine the population structure of *C. jejuni* in pets. A total of 527 faecal samples were collected from ten dog walkways and 82 cat faecal samples were collected from the small animal veterinary clinic at Massey University. The prevalences of *Campylobacter* spp. and *C. jejuni* in dogs were 15% (95% CI 11% to 19%) and 7% (95% CI 5% to 11%) and in cats 9% (95% CI 4% to 17%) and 7% (95% CI 3% to 15%), respectively. A total of 31 *C. jejuni* positive isolates from dogs and cats were characterised by multilocus sequence typing, *fla*A, and *por*A typing of which 21 isolates were completely sequenced (17 from dogs and 4 from cats). The most predominant sequence type (ST) clonal complex was ST-45 (13 of 21) followed by ST-52 (2 of 21). Other STs identified were ST-50, ST-422, ST-474, ST-583, ST-696 and ST-3961. The most predominant *fla*A types were *fla* SVR 8 (3 of 21) and 21 (3 of 21). Following antigenic typing, it was noticed that the allelic combination of ST-45, *fla* SVR 8 and *por*A 44 and ST-45, *fla* SVR 21 and *por*A 44 were the predominant types found in cats and dogs with the *por*A allele 44 accounting for 29% of the total number of typed isolates (6 of 21). Eight of the dog *C. jejuni* isolates remained untyped due to different technical reasons. The high prevalence of ST-45, the genotype found predominantly in human campylobacteriosis in all parts of the world and the isolation of ST-474, the most pathogenic genotype exclusively found in New Zealand, emphasise the need for further investigations of the role of pets as vectors of human campylobacteriosis.

Vathsala M, French N, Stevenson M, Marshall J, and Hotter G, (2011) Characterisation of *Campylobacter jejuni* from pets in New Zealand by combining multilocus sequence typing (MLST), *por*A and *fla*A typing

4.1 Introduction

Campylobacter spp. is the major cause of bacterial gastroenteritis worldwide (Adak et al. 2002, CDC 2007). *Campylobacter jejuni* and *C. coli* are considered the most important human pathogens although other species such as *C. upsaliensis* have been isolated from a small proportion of human cases (Goossens et al. 1990, Labarca et al. 2002, Lastovica & Le Roux 2003). Ingestion of untreated water, undercooked meat, raw milk and cross contamination of foods are recognised risk factors for human campylobacteriosis (Fullerton et al. 2007, Marcus 2008). Various foodborne transmission pathways have been extensively studied to better understand the epidemiology of *C. jejuni* (Scott et al. 2000, Sneyd & Baker 2003, Baker et al. 2007). It has been hypothesised that pets that live in close proximity to humans have a greater opportunity to transmit pathogens directly to their owners (Hald & Madsen 1997) but this transmission pathway for campylobacteriosis has not been investigated in detail. Recently, studies from developed countries have identified pet ownership (particularly dogs and cats) as a risk factor for human campylobacteriosis particularly among small children and infants (Tenkate & Stafford 2001, Fullerton et al. 2007). Research on animal reservoirs for *Campylobacter* spp. have isolated *C. jejuni*, *C. coli*, *C. upsaliensis* and other *Campylobacter* spp. from pets (Tsai et al. 2007, Chaban et al. 2010). Transmission of *Campylobacter* from pets to humans, particularly children, has often been suspected and has also been proven in a small number of cases (Jimenez et al. 1999, Wolfs et al. 2001). The likelihood of healthy pets being infective to humans has been the subject of controversy (Gow et al. 2009). In contrast, pet ownership has been reported as protective for humans in a case-control study conducted in South Australian rural children in 2005, for a period of six weeks (Heyworth et al. 2006). The authors found, after adjusting for the effect of confounders such as hand washing habits and contact with sick animals, the incidence of highly credible gastrointestinal symptoms (HCGI) was 42% (n = 965) in children living in a household with a dog or a cat compared with 46% in those living without a pet. In addtion, the authors speculated that the decision to own a pet was often dictated by the presence or absence of health conditions in the child, such as asthma.

In a cross-sectional study of 214 dogs in southern Chile Fernandez & Martin (1991) found that *Campylobacter* carriage was more common in stray dogs (77 of 150, 51%) compared

with pet dogs (14 of 150, 9%) suggesting that environmental sanitary conditions play a role as a determinant of whether or not an animal carries the organism. Estimates of *Campylobacter* carriage rates among healthy pets have varied from 20% to 75% (Burnens et al. 1992, Baker et al. 1999, Engberg et al. 2000, Moser et al. 2001, Steinhauserova et al. 2000). The spectrum of species distribution and carriage rates vary considerably between studies and over time which is likely to be due to differences in the methodologies adopted for sample collection, isolation, screening and study design. A number of studies have found that *C. jejuni* was more frequently isolated from dogs that were less than 12 months of age compared with those that were older (Hald et al. 2004, Acke et al. 2009, Parsons et al. 2009). Others failed to identify such a relationship (Tsai et al. 2007, Wieland et al. 2005). In a cross-sectional study of 72 puppies and 42 kittens in Denmark were shown to have a relatively high prevalence of *C. jejuni* (76%; 54 of 72) followed by *C. upsaliensis* (19%; 4 of 72) and *C. coli* (5%; 1 of 72) (Hald & Madsen 1997).

While most studies of this type have focused on carriage rates and species differentiation, one particular longitudinal study investigated the shedding pattern of *Campylobacter* in 26 domestic pet dogs in Denmark (Hald et al. 2004). This study showed that the colonisation of *Campylobacter* lasts for long periods and dogs were healthy carriers for the first two years of their life. Persistent colonisation occurred with one or more strains of *C. upsaliensis* over several months and there was intermittent sporadic excretion of *C. jejuni* (Hald et al. 2004). Reports on the zoonotic enteric pathogens in cats and dogs have identified the asymptomatic presence of *Campylobacter* in puppies and kittens less than one year of age (Saeed et al. 1993, Hald et al. 2004). In contrast, studies conducted in Ireland have shown *Campylobacter* to be associated with gastroenteritis in cats and dogs, particularly young animals (Acke et al. 2006, 2009). A cross-sectional study conducted in the United Kingdom found that 7% of 47 kittens presented to a veterinary clinic for routine vaccination were *Campylobacter* positive (Gow et al. 2009). While the observational studies cited here provide a useful starting point in terms of indicating the frequency of *Campylobacter* carriage among pets, a general weakness relates to the representativeness of the dog and cat populations that have been sampled and consistency of the methodologies that have been applied. For example, in the study conducted by Gow et al. (2009), samples were collected and posted to the veterinary clinic by pet owners. Details of transportation, maintenance of refrigeration and time taken for transportation are obscure. In

the studies conducted in Ireland (Acke et al. 2006, 2009), the kennel or the cattery environment completely differs from that of a household with respect to overall management which casting some doubt on our ability to make inferences about the general pet population based on these data.

Molecular characterisation and the population structure of *C. jejuni* isolates from pets have, to date, not been investigated in any great detail. Genotyping has been carried out to diagnose neonatal sepsis caused by *C. jejuni* in a 3 week old infant who acquired infection from a household puppy (Wolfs et al. 2001). *C. jejuni* isolates from the puppy and the infant were characterised using amplified fragment length polymorphic fingerprinting (AFLP) which produced homologous AFLP patterns in both the puppy and the infant. This study was the first documented occurrence of transmission of *C. jejuni* from dogs to humans. Subsequently, the genetic relatedness amongst the *C. jejuni* isolates from both pets and humans was determined by pulsed field gel electrophoretic (PFGE) analysis by Damborg et al. and colleagues in 2004. Their study was carried out to investigate the occurrence of *Campylobacter* spp. in pets living along with patients infected with *C. jejuni*. The occurrence of *C. jejuni* was found to be higher in younger patients compared with those that were older. A cluster analysis of the canine strains showed 95% similarity between the human and canine strains of *C. jejuni* from different Danish counties (Damborg et al. 2004). Another important observation made in this study was that the cases were shown not to have recently travelled, ruling out the possibility that infection may have been acquired elsewhere. Similarly, genetic diversity of *C. jejuni* in pets was studied recently by Acke and others in Ireland (Acke et al. 2010). PFGE and flagellin typing (*fla*A) were performed to investigate the genetic diversity of *C. jejuni*. These authors detected 27 unique patterns of *fla*A, demonstrating a high level of heterogeneity in the *C. jejuni* populations in this group of pets. In addition to *C. jejuni*, the species *C. upsaliensis* has been shown to be prevalent in pets with carriage rates ranging from 64% to 82% (Owen & Hernandez 1990, Burnens et al. 1992, Baker et al. 1999, Newell et al. 2000). *C. upsaliensis* has been characterised using PFGE (Owen & Hernandez 1990, Moser et al. 2001, Lentzsch et al. 2004) and multilocus sequence typing (MLST) (Miller et al. 2005).

New Zealand has, until recently, ranked first in the world for campylobacteriosis notification rates (Baker et al. 2007, Marler 2010). Molecular epidemiological work has implicated poultry, cattle and recreational water as major contributing factors (Brown et al.

2004, Baker et al. 2007, French et al. 2009a, Mullner et al. 2009, 2010). Campylobacteriosis in humans is multi-factorial and for this reason, there is a need to better understand all possible exposure pathways. Given the substantial body of evidence linking household pets with human disease, as outlined above, it makes sense that this possibility should be assessed in a New Zealand context. With this background we conducted a repeated cross sectional study of a dog and cat population resident in the city of Palmerston North in the lower North Island of New Zealand. Our aims were to: (1) determine the prevalence of *C. jejuni* carriage among dogs and cats in an urban area in provincial New Zealand, (2) determine the population diversity of *C. jejuni* in this population and, (3) to examine the association between dog and cat *C. jejuni* STs and those found in confirmed human cases listed in the PubMED MLST database.[1]

4.2 Experimental procedures

4.2.1 Study design

A repeated cross sectional study was conducted to determine the prevalence of *Campylobacter* in the faeces of dogs and cats resident in the city of Palmerston North (longitude 175°, latitude -40°) in the lower North Island of New Zealand. Owing to the wide range in *C. jejuni* prevalence estimates from previous studies (Burnens et al. 1992, Baker et al. 1999, Engberg et al. 2000, Moser et al. 2001, Steinhauserova et al. 2000), we assumed a 50% design prevalence for our sample size calculations. This provided the largest sample size estimate required for a 95% level of confidence and a 6% margin of error (Levy & Lemeshow 1999, pp. 258).

Based on these calculations a total of 267 samples were required. Ten areas within the Palmerston North city limits commonly used for dog walking were selected as sampling sites: Hokowhitu, The Esplanade, Coronation Park, Milverton Park, Bledisloe Park, Albert Street, Vogel Street, Railway Road, Fitzerbhert Bridge and The Bridle Track (Figure 4.1). Five of the sampling sites included children playgrounds (Coronation Park, Bledisloe Park, Milverton Park, The Esplanade and Hokowhitu). Two were walkways within residential areas (Vogel Street and Railway Road). The Albert Street, Fitzherbert Bridge

[1]URL: (`http://pubmlst.org/Campylobacter`)

and The Bridle Track sampling sites were all situated near the banks of the Manawatu River. Faecal samples were also collected from dogs that attended the Massey University Small Animal Veterinary Clinic for routine procedures such as vaccination and deworming. The intention here was to ensure sampling of a wider population of dogs within the Palmerston North city limits.

For cats, faecal material was collected from those attending the Massey University Small Animal Veterinary Clinic using the same selection criteria used for dogs. In addition, faecal material was collected from cats belonging to staff and students of Massey University, one private veterinary clinic and a commercial cattery.

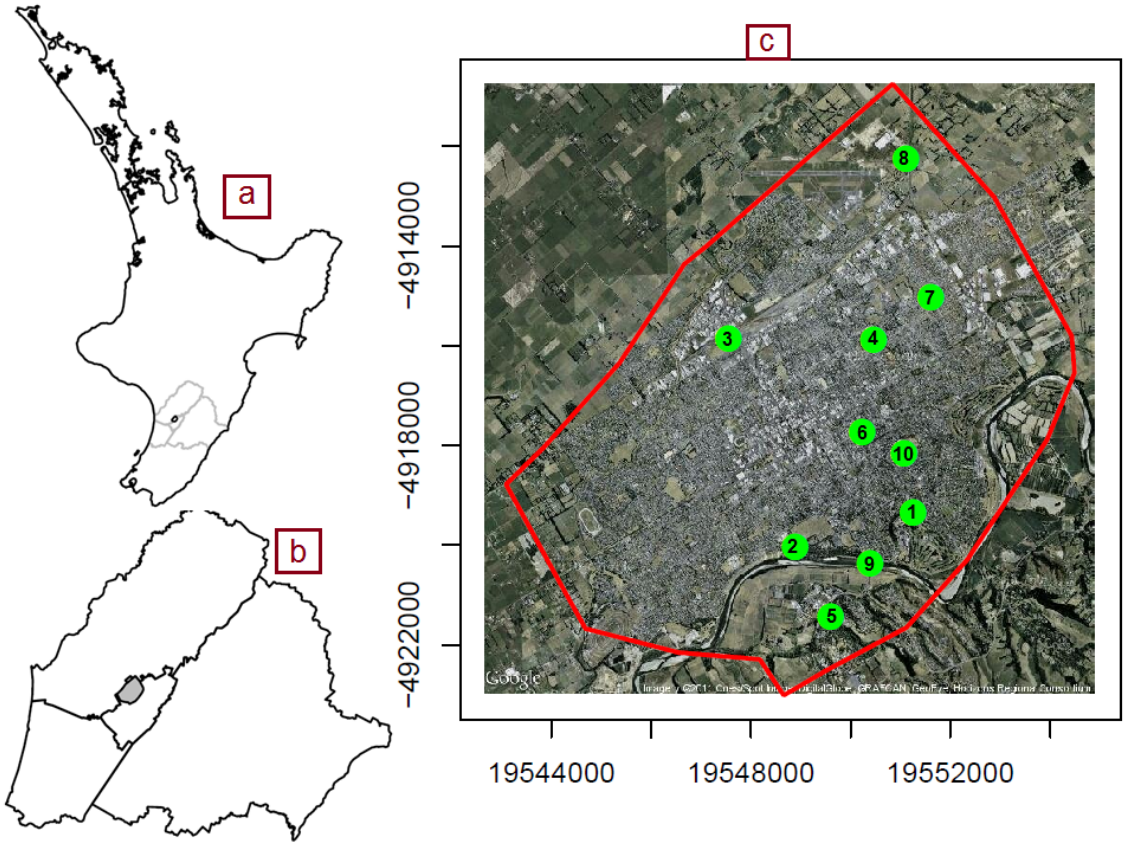


Figure 4.1: The sampling sites of dogs : a. North Island of New Zealand; b. Manawatu – Palmerston North city; c. Ten dog walkways with dog faecal bins placed in by the Palmerston North city council from where faecal samples were collected. 1=Hokowhitu, 2=The Esplanade, 3=Coronation Park, 4=Milverton Park, 5=Bledisloe Park, 6=Albert Street, 7=Vogel Street, 8=Railway Road, 8=Fitzerbhert Bridge and 10=The Bridle Track

4.2.2 Collection of faecal material

Each of the ten sampling sites was visited at monthly intervals for a period of 17 months, starting in March 2008. Standardisation of transport of faecal material and processing for *Campylobacter* isolation was carried out for a trial period of five consecutive months from March to July 2008 (inclusive). At each sampling site three faecal samples were collected from dog bins placed within the respective walkways by the city council. A total of 30 samples were collected each sampling month. Faecal material could not be collected from Milverton Park for the February 2009 sampling round and only a single sample could be collected from Coronation Park in June 2009. Samples were collected from intact dog faecal bags in order to avoid re-sampling the same faecal sample. Moreover, the sampling interval was around 25-30 days between each sampling round, hence we are confident that a new sample was retrieved during every sampling round.

Although faeces were collected over a period of 17 months, only data for the interval from August 2008 to July 2009 were used to estimate the prevalence of *C. jejuni*. The total number of faecal samples from dogs for the 12 month study period was 355 and the total number of samples collected (including the trial period) was 528. A total of 25 samples were collected from dogs presented to the Massey University Veterinary Clinic between March and July 2008 and 82 samples from cats presented between March and September 2010. The *C. jejuni* positive samples from the trial period were also used to examine the population structure of *C. jejuni*. Because the number of samples and the sampling pattern was not consistent in the case of cats, these samples were not used to estimate the prevalence of *C. jejuni* in cats, rather, the positive samples were used for the characterisation and population differentiation of *C. jejuni* in cats. However, in order to get an approximate idea about the prevalence of *C. jejuni* in cats, prevalence was estimated and it should be noted that this is unlikely to represent the cat population of Palmerston North.

4.2.3 Bacterial isolation and DNA preparation

All faecal samples from all sampling sites (except for the cat and dog faecal samples from Massey University Veterinary Clinic) were collected simultaneously in transport media (Amies charcoal, Fort Richards, Auckland) and Bolton's enrichment broth (enrichment

broth LAB-27.6 G; 50 mL lysed horse blood venous supplies; Antibiotics-LAB-10 mL, Auckland). Faecal bags were picked out from bins and sterile swabs (one for the Bolton enrichment and another for the transport media) were inserted into relatively fresh faecal materials and transported immediately to the Hopkirk Research Institute Laboratory on the Massey University campus at Palmerston North.

Faecal material collected in the transport media were directly streaked onto modified charcoal cefoperazone - deoxycholate (mCCDA) (Fort Richards, Auckland) plates. The inoculated mCCDA plates and Bolton's enrichment broth with faecal material were incubated for 48 hours at 42°C in a microaerophilic chamber (MACS VA500 Microaerophilic workstation, Don Whitley Scientific). The plates were monitored for growth, and after 48 hours the colonies resembling *Campylobacter* spp. were sub-cultured onto blood agar plates (horse lysed blood agar, Fort Richards). After 48 hours the cultures from the Bolton's enrichment broth were swabbed onto mCCDA plates and then the inoculated plates were incubated for another 48 hours at 42°C in a microaerophilic chamber. Up to three colonies from the mCCDA plates were sub-cultured onto horse blood agar plates. The pure colonies isolated from the horse blood agar plates were tested for oxidase reduction (oxidase strips, Fort Richards, Auckland). The colonies that reduced oxidase within 5 seconds, as indicated by purple colouration, were stored in glycerol and processed for DNA isolation. Three colonies of at least 3 mm in diameter were transferred to 1 mL of 2% Chelex solution in distilled water (weight/volume) and boiled at 100°C for 10 minutes on heating blocks. These were then cooled to room temperature, centrifuged at 13,000 rpm for 10 minutes and the supernatants collected in fresh sterile eppendorf tubes and stored at -20°C.

4.2.4 Speciation by PCR

Two separate PCR reactions were performed; one for the confirmation of genus *Campylobacter* using the 16s rRNA gene primer sequence and the species *C. jejuni* was confirmed using the membrane associated protein A (*map*A) gene primer sequence. Both primers were adapted from previous published reports (Stucki et al. 1995, Linton et al. 1997, Mullner et al. 2010) (Table 4.1).

The targets were amplified at 96°C for 2 minutes for initial denaturation, 96°C for 30 sec-

onds; the primer annealing was obtained at 56°C for 30 seconds for genus detection and the primer annealing was achieved at 60°C for 30 seconds for species detection with an extension temperature of 72°C for 60 seconds, for 35 cycles for both PCR reactions. The PCR reaction mix was comprised of 2 μL 10$\times$ PCR buffer (final concentration 1$\times$); 2 μL d-NTPs (final concentration 2 mM); magnesium chloride -1 μL, (final concentration 2.5 mM); primers -2 μL each (final concentration 1 mM); Taq DNA polymerase 0.2 μL (final concentration 1 unit per reaction); DNA 2 μL (final concentration 10 ng per μL). The reaction mix was made up to 20 μL with distilled water. The amplicons were examined by agarose gel electrophoresis with results captured using a Bio-Rad gel documentation system.

Table 4.1: Primers and target genes used for PCR

Target gene	Primer	Primer sequence	Adapted from
16S rRNA	Forward - 1	5' GGATGACACTTTTCGGAGC 3'	Stucki et al. (1995), Mullner et al. (2010)
(*Campylobacter* spp.)	Reverse - 2	3' CATTGTAGCACGTGTGTC 3'	
*map*A	Forward - 1	5' CTTGGCTTGAAATTTGCTTG 3'	Linton et al. (1997), Mullner et al. (2010)
Membrane associated protein A	Reverse - 2	3' GCTTGGTGCGGATTGTAAA 5'	
(*C. jejuni*)			

4.2.5 Multilocus sequence typing

All the *C. jejuni* positive isolates were chosen for MLST characterisation. A total of 25 isolates were subjected for amplification and characterisation. Some of the isolates could not be characterised fully for one or more of the following reasons: (1) some alleles were difficult to amplify more than twice even with superior polymerase enzymes and optimised magnesium chloride concentrations (2.5 mM); (2) some alleles could be amplified but could not be sequenced; (3) full profiles could not be obtained as one or two alleles could not be amplified and sequenced. The primers were adapted from the previous published reports on MLST characterisation (Table 4.2).

Each 25μL amplification reaction mixture comprised of 10 ng of *Campylobacter* chromosomal DNA per μL (2μL), 1 μL (5 picomoles) PCR primers (forward and reverse primers), 12.5 μL of 2$\times$ ABI Taq Gold Mastermix and 9.5 μL of distilled water. The reaction was carried out in 96 well plates for 7 house keeping genes. Each plate held 13 samples. The PCR program was comprised of a denaturation step at 94°C for 15 minutes,

on a 35 cycle run with initial denaturation done at 94°C for 30 seconds, primer annealing at 50°C for 30 seconds and extension at 72°C for 90 seconds and the final elongation step was for 7 minutes. On completion, the plates were held at 4°C for further precipitation and purification of the amplicons. In the case of isolates that were difficult to amplify, Platinum Taq polymerase enzyme was used with 2.5 mM concentration of magnesium chloride to enhance amplification.

Table 4.2: Primer sequences used for the amplifictaion of seven housekeeping alleles for *C.jejuni* MLST typing (Miller et al. 2006)

Alleles	Forward and Reverse primers
*asp*A	F-5' GAGAGAAAAGCWGAAGAATTTAAAGAT 3'
	R-3' TTTTTTCATTWGCRSTAATACCATC 5'
*gln*A	F-5' TGATAGGMACTTGGCAYCATATYAC 3'
	R-3' ARRCTCATATGMACATGCATACCA 5'
*glt*A	F-5' GARTGGCTTGCKGAAAAYAARCTTT 3'
	R-3' TATAAACCCTATGYCCAAAGCCCAT 5'
*gly*A	F-5' ATTCAGGTTCTCAAGCTAATCAAGG 3'
	R-3' GCTAAATCYGCATCTTTKCCRCTAAA 5'
pgm	F-5' CATTGCGTGTDGTTTTAGATGTVGC 3'
	R-3' AATTTTCHGTBCCAGAATAGCGAAA 5'
tkt	F-5' GCAAAYTCAGGMCAYCCAGGTGC 3'
	R-3' TTTTAATHAVHTCTTCRCCCAAAGGT 5'
*unc*A	F-5' GCAAGGDGTTATYTGTATWTATGTTGC 3'
	R-3' TTTAADAVYTCAACCATTCTTTGTCC 5'

Purification of PCR amplified products

The amplified products were precipitated by mixing with 25 μL of PEG (20% Poly Ethylene Glycol 8000 in 2.5 M NaCl) and incubated at 37°C for 15 minutes and the products were centrifuged at 2,500 rpm for 30 minutes. The PEG complex in the plate was removed by spinning the plate inverted onto tissue towels at 300 rpm for 2 minutes. The DNA in the plates were washed with 80% ethanol and centrifuged for 10 minutes at 2,500 rpm and the ethanol was removed by inverting the plate onto tissue towels. The plates were air dried in a clean dark chamber until the ethanol evaporated. The products were rehydrated in sterile distilled water and were examined by agarose gel electrophoresis with results captured using a Bio-Rad gel documentation system. Quantification of DNA was done by visualising the bands on the agarose gels.

Sequencing

The sequencing reaction mix contained 400 μL of distilled water, 200 μL of 5$\times$ sequencing buffer and 100 μL of BigDye Ready Reaction Mix (PE Biosystems). Seven μL of this mix was dispensed to each well with 1 μL of forward primer (3.2 picomoles) and 2 μL of amplified DNA products from the amplification plate were transferred to each well in the same order as that of the amplification plate and sequenced in the PCR machine. The sequencing cycle consisted of an initial denaturation step at 96 °C for 3 minutes for 25 cycles with initial denaturation done at 96°C for 15 seconds; primer annealing at 50°C for 15 seconds; and extension at 72°C for 4 minutes and the plates were held at 4 °C for further purification.

Precipitation and purification of sequenced products

The unincorporated dye terminators were removed by precipitation of the termination products with 95% ethanol, according to the manufacturer's instructions. The precipitation mixture prepared with 2 μL of 3M sodium acetate, 10 μL distilled water and 50 μL (95%) ethanol. This mixture was made up for 96 wells and 62 μL was added to each well and agitated and incubated for 15 minutes in a dark chamber without exposing the plate to light. The plate was then centrifuged at 2,500 rpm for 30 minutes. The precipitation mix was removed by inverting the plate onto tissue towels and spinning at 300 rpm for 2 minutes. The wells were washed with prechilled (-20°C) 70% ethanol and washed off by inverting the plate onto tissue towels at 300 rpm for 2 minutes. The plates were air dried in clean dark chamber after washing and covered with plastic adhesive sealers and wrapped in aluminium foil for transportation. The sequenced plates were sent to the Institute for Environmental Science and Research (ESR), Wellington, New Zealand, for sequence reading and allele assignment.

*fla*A and *por*A typing

Pure colonies of *C. jejuni* from dogs and cats are were transferred to 1 mL of 2% Chelex solution in distilled water (weight/volume) and boiled at 100°C for 10 minutes on heating blocks. These were then cooled to room temperature, centrifuged at 13,000 rpm for 20 minutes and the supernatants collected in fresh sterile eppendorf tubes. The DNA

preparations were quantiated in a Nanodrop-spectrophotometer (Australia) and reconstituted for 50 ng per μL and frozen and sent to ESR, Wellington, under cold conditions for sequencing and allele assignment.

Statistical analysis

Faecal samples were collected from dog walkways and therefore, the faecal samples may not represent that particular area from where the samples were collected. Hence, an analysis for the possibility that prevalence varied across dog walkways was not attempted. Prevalence and confidence intervals were estimated assumming that the dog faecal samples may not truly represent the respective sampling areas. Our rationale for such assumption was that there is a possibility for a dog from one area to be taken to another dog walkway.

Population differentiation: Analysis of molecular variance and Fst

Arlequin v3.11 (Excoffier et al. 2005) software was used to perform analysis of molecular variance (AMOVA) in an effort to analyse the population structure of *C. jejuni* in dogs and cats. The population structure is analysed in AMOVA through a variance framework. The allelic frequencies within and between populations are analysed at different 'hierarchical' levels, where AMOVA divides the total variance into different covariance components such as within population, within groups among populations and inter-population differences (Excoffier et al. 2005). Three hierarchical F-statistics, known as the fixation indices are obtained (expressed as components of AMOVA) namely, Fst : the variance among subpopulations relative to the total variance; Fsc : the variance among subpopulations within groups; and Fct : the variance among groups relative to the total variance. Fst is a widely used measure to quantify the differentiation between populations and a value of one implies the two populations are completely separate (Wright 1965, 1978, 1984). The following are some guidelines for interpreting the Fst values:

- the range 0.0 to 0.05 may be considered as indicating little genetic differentiation; the range 0.05 to 0.15 indicates moderate genetic differentiation;
- the range 0.15 to 0.25 indicates great genetic differentiation; and

- the values of Fst above 0.25 indicate very great genetic differentiation Wright (1978).

As faecal samples were collected from dog walkways where, dog bins at the sampling sites may not represent that particular area from where samples were collected or the samples in the bin may not belong to the dogs of those respective areas, population differentiation analysis was not carried out for individual sampling sites. Instead, an analysis of population differentiation within dog *C. jejuni* population; and between dogs and cats were carried out.

4.3 Results

4.3.1 Prevalence of faecal *Campylobacter* spp. in dogs and cats

Seventy two of the 355 (20%, 95% confidence interval [CI] 16 to 25%) dog faecal samples were positive for the *Campylobacter* spp. As faecal material was collected from dog bins, details of health status, diet and age were not available. The estimated prevalence of *C. jejuni* was 7% (95% CI 5 to 11%). In the case of cats, the estimated prevalence of *Campylobacter* spp. and *C. jejuni* was 9% (95% CI 4 to 17%) and 7% (95% CI 3 to 15%), respectively.

Details of the numbers of samples collected, the number of *Campylobacter* positive samples and prevalence estimates by sampling site are shown in Table 4.3. Figure 4.2 shows the same data by site and sampling month. *Campylobacter* spp. prevalence as a function of calendar time reveals a high prevalence in early spring (23%, 95% CI 8 to 38%) and a relatively low prevalence in summer (13%, 95% CI 1 to 25%). However, these estimates are associated with wide uncertainity. The prevalence estimates of *Campylobacter* spp. and *C. jejuni* are provided in Table 4.3.

C. jejuni prevalence during the summer was 20% (95% CI 7 to 34%) and during the early winter was 18% (95% CI 4 to 32%). Although the prevalence of *Campylobacter* spp. was highest in The Esplanade, The Bledisloe Park had the highest prevalence of *C. jejuni* followed by the Bridle Track, Fitzherbert Bridge and Railway Road (Table 4.3).

The overall prevalence of *Campylobacter* in faeces from dogs from Massey University Veterinary Clinic was 12% (CI 0.7 to 25%; 3 of 25 samples). None of the samples from dogs from Massey University Veterinary Clinic tested positive for *C. jejuni*.

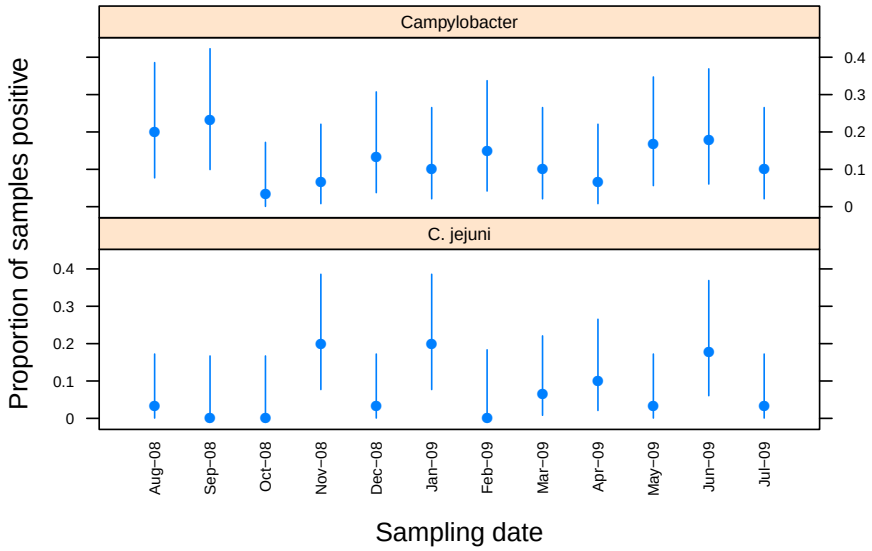


Figure 4.2: Seasonal prevalence of *Campylobacter* spp. and *C. jejuni* in dogs

4.3.2 *C. jejuni* genotypes and cell surface antigen types of dogs and cats

Of 25 *C. jejuni* isolates speciated from dogs, only 24 could be amplified and taken further for sequencing. Of the 24 successfully amplified isolates a total of 17 were sequenced completely and the remainder could not be typed completely despite repeated amplification and sequencing procedures (Table 4.5). The 17 sequence types (ST) were assigned to five clonal complexes (CC) and one isolate from ST-3961 could not be assigned to a complex. The majority of dog *C. jejuni* isolates were assigned to ST-45 complex accounting for 53% (9 of 17) isolates, three isolates were from the ST-52 complex, two from the ST-21 complex and the remainder belonged to ST-1332 (n = 1), ST-48 (n = 1) and an unassigned complex (ST-U/A)(n = 1).

For cats, 6 samples were speciated from a total of 82 faecal samples and all were subjected to MLST typing. Four *C. jejuni* isolates could be successfully typed and two of them could not be typed further. All four samples were assigned to the ST-45 complex. The

Table 4.3: The prevalence of *Campylobacter* spp. and *C. jejuni* at the sampling sites

		Campylobacter spp.		*C. jejuni*	
Sampling sites	n	n positive	95% Con. Int	n positive	95% Con. Int
The Esplanade	36	13 (36%)	20-52	1 (3%)	0.07-15
Coronation Park	34	3 (9%)	2-24	2 (6%)	0.7-20
Milverton Park	33	4 (12%)	3-28	1 (3%)	0.08-16
Hokowhitu	36	3 (8%)	2-22	3 (8%)	2-22
Bledisloe Park	36	6 (17%)	6-33	5 (14%)	5-29
Vogel Street	36	3 (8%)	2-22	1 (3%)	0.07-15
Albert Street	36	3 (8%)	2-22	3 (8%)	2-22
Bridle Track	36	4 (11%)	3-26	4 (11%)	3-26
Railway Road	36	1 (3%)	0.07-15	3 (8%)	2-22
Fitzherbert Bridge	36	6 (17%)	6.0-33	3 (8%)	2-22

n: Total number of samples
Con.Int: Confidence interval

cell surface antigen typing of the dog faecal samples showed six different types of *por*A alleles and seven different *fla*A alleles. It was difficult to type the cell surface antigens for some of the isolates. The *por*A alleles were 1, 44, 53, 60, 886, 905 and one allele could not be assigned an allelic number. The *fla*A alleles were 15, 21, 22, 32, 34, 56 and 57. The combinations of *fla*A allele 22, *por*A allele 53 and ST-45 complex (n = 2), *fla*A allele 21, *por*A allele 44 and ST-45 complex (n = 3) and, flaA allele 57, *por*A allele 905 and ST-52 complex (n = 3) were more frequent in dogs. In cats, the combination of *fla*A allele 8, *por*A allele 44 and ST-45 complex (n = 3) was predominant and *fla*A allele 239, *por*A allele 73 and ST-45 complex (n = 1) was found in a single isolate (Table 4.4). The majority of the ST-45 (n = 5) isolates were found during summer and the remainder of ST-45 and the other STs were found during autumn and winter (ST-52, ST-50, ST-422, ST-45, ST-474, ST-696 and ST-3961). The genotypes were more diverse during winter in general.

The p-value of differentiation was 0.04 which means the *C. jejuni* population of dogs did differ significantly from that of cats (Table 4.6). However, the sample size (genotypes) from cats (n = 4) was very small compared with dogs (n = 17).

Table 4.4: *C. jejuni* genotypes from dog and cat faecal samples

C. jejuni genotypes from dog faecal samples											
Frequency	*fla*A	*por*A	ST	CC	*asp*A	*gln*A	*glt*A	*gly*A	*glmM*	*tkt*	*unc*A
1	15	X	696	1332	2	1	4	28	58	25	58
2	X	X	45	45	4	7	10	4	1	7	1
2	22	53	45	45	4	7	10	4	1	7	1
1	X	X	45	45	4	7	10	4	1	7	1
3	21	44	45	45	4	7	10	4	1	7	1
1	32	60	474	48	2	4	1	2	2	1	5
1	34	NEW	422	21	2	1	5	3	2	5	5
1	56	886	3961	U/A	2	29	4	27	10	25	24
1	X	53	45	45	4	7	10	4	1	7	1
1	X	1	50	21	2	1	12	3	2	1	5
3	57	905	52	52	9	25	2	10	22	3	6
C. jejuni genotypes from cat faecal samples											
3	8	44	45	45	4	7	10	4	1	7	1
1	239	73	583	45	4	7	10	4	42	51	1

X:Allele could not be typed

NEW:Allelic number was not assigned at the time of analysis

U/A:Unassigned

Table 4.5: *C. jejuni* MLST partial profiles from dog faecal samples

Lab ID	Source Site	Month	*asp*A	*gln*A	*glt*A	*gly*A	*glm*M	*tkt*	*unc*A	ST	CC	*fla*A	*por*A
D274	Milverton Park	November-08	4	X	X	X	1	X	X	-	-	x	x
D277	Coronation Park	November-08	X	7	10	4	1	7	1	-	-	x	x
D404	Hokowhitu Lagoon	March-09	X	X	X	26	X	X	X	-	-	x	x
D408	Bledisloe Park	March-09	2	X	X	NEW	X	X	23	-	-	x	x
D486a	Bridle Track	June-09	2	4	1	2	2	X	5	-	-	32	60

X:Allele that could not be sequence typed
(-):Alleles that could not be amplified
Three dog and two cat isolates could not be amplified

Table 4.6: Analysis of molecular variance of *C. jejuni* genotypes from pets

Source of variation	df	Sum of squares	Variance components	Percentage
Among populations	1	2.8	0.12Va	5.36
Within populations	17	35.4	2.08 Vb	94.64
Total	18	38.3	2.2	
Fixation Index Fst : 0.05360				
P value : 0.04				

Among populations: *C. jejuni* populations from dogs and cats.

Within populations: *C. jejuni* populations within the isolates of dogs and cats

Va: variance among populations

Vb: variance within populations

4.3.3 Comparison of *C. jejuni* genotypes with wider population

The *C. jejuni* genotypes identified in this study were compared with the wider population of *C. jejuni* genotypes listed in the MLST database.[2] ST-3961 was novel to this study. ST-474 had been reported predominantly from poultry and human campylobacteriosis cases in New Zealand. ST-50 has been implicated most commonly in sporadic human campylobacteriosis cases in the United Kingdom, USA, The Netherlands, Finland, Korea, Canada, Scotland and Curaçao and has been recorded as a cause of disease in humans as early as 1990. ST-50 has predominantly been found in chicken meat and offal, beef, lamb and turkey. ST-52 has been implicated in sporadic cases of campylobacteriosis in humans in the United Kingdom, The Netherlands, Wales, Australia, New Zealand and Curaçao and is primarily isolated from sheep. ST-52 has also been recorded from chicken and this is first record of an isolation from dogs. ST-422 has been isolated from a chicken in the United Kingdom which is the only record of this ST in the MLST database, although this ST has been reported in human clinical cases in New Zealand (French et al. 2010). ST-696 has been reported predominantly from geese in the United Kingdom with an single isolates from cattle and a turkey in Germany. ST-696 was implicated in cases of human campylobacteriosis in The Netherlands and has been isolated from New Zealand ducks (Chapter 3). This is the first time ST-696 has been reported from dogs. ST-45 is one of the predominantly reported STs in a range of hosts and also highly implicated in human campylobacteriosis, accounting for 59% of the total STs in this study. ST-583 was isolated from a cat in this study and has also been isolated from a variety of sources including wild birds, turkeys, cattle and chickens from the United Kingdom, Denmark, Canada and New Zealand. ST-583 has been implicated in sporadic cases of disease in humans in The Netherlands, England, Belgium and Germany.

[2]URL: (`http://pubmlst.org/Campylobacter`)

4.4 Discussion

The aims of this study were to quantify the prevalence of *Campylobacter* spp. and *C. jejuni* in dogs and cats in Palmerston North City, New Zealand and to investigate the population structure of the *C. jejuni* isolates from those samples. The overall prevalence of *Campylobacter* spp. observed in faecal samples of dogs was 15% (95% CI 11% to 19%) which is relatively low compared with previous prevalence reports that ranged from 18% to 72% (Aydin et al. 2001, Sandberg et al. 2002, Engvall et al. 2003, Damborg et al. 2004, Hald et al. 2004, Bender et al. 2005, Acke et al. 2006, 2009, Parsons et al. 2009). The overall *C. jejuni* prevalence in this study was 7% (95% CI 5% to 11%) and the prevalence of *C. jejuni* in dogs in previous studies ranged from 3% to 40% (Hald & Madsen 1997, Lopez et al. 2002, Koene et al. 2004, Workman et al. 2005, Tsai et al. 2007).

In this study faecal samples were collected from dog bins, in contrast to other studies where samples have generally been collected directly from the rectum using swabs. This may be one explanation for the relatively low prevalence identified in this study compared with those from other work. The length of time taken for samples to reach the laboratory is also known to influence *Campylobacter* spp. recovery rates (Koene et al. 2004). In this study, even though samples reached the laboratory and processing started within 60 minutes, it was not known how long faecal material had actually been present in the bins at the time of collection. The existence of viable, but non-culturable forms of *Campylobacter* spp. in the faecal samples could have also influenced the recovery rate in this present study. This occurs when bacteria are exposed to adverse conditions outside the host (Persson & Olsen 2005, Murphy et al. 2006). This lends support to the explanations for the low prevalence estimates described above.

A direct PCR method has been used to detect small numbers of organisms (Parsons et al. 2009), however, inhibitory factors and the poor quality of DNA from faecal material have been found to influence the direct-PCR detection of bacteria from faeces where samples that were negative by direct-PCR were found to be positive by culture Lawson et al. (1999), Parsons et al. (2009). In contrast, isolation rates have been found to be higher from faecal samples compared with rectal swabs in dogs and *vice versa* in cats (Acke et al. 2006, 2009). In addition, *Campylobacter* spp. recovery rates may be influenced by other

factors such as the age of animals at the time of collection and the overall management of pets (which may vary from study to study and country to country).

The overall estimated prevalence of *Campylobacter* spp. in cats was 9% (95% CI 4% to 17%) which may not be truly representative of the Palmerston North pet cat population firstly because of selection bias (samples were taken from animals attending one particular veterinary hospital) and the relatively small number of samples (n = 82) would have decreased the precision of the prevalence estimate.

Over time the prevalence of *Campylobacter* spp. and *C. jejuni* was highest prevalence in the spring and a relatively low during the summer and winter, however the differences were not significant. Comparison of our findings with those from other studies is difficult because most have not sampled longitudinally and presented prevalence estimates at different times throughout the year. Although the differences were not significant, our findings in relation to seasonality are consistent with observations made by Tauxe (1992) in a review of epidemiology of *C. jejuni* infections in industrialised nations as well as in a cross sectional study conducted in domestic animals in Argentina by Lopez et al. (2002). Torre & Tello (1993) identified a high prevalence of *Campylobacter* spp. in dogs in the autumn. Factors influencing the high spring and summer prevalence of *Campylobacter* spp. in dogs are not clearly defined. However, high temperatures, humidity, overcrowding of animals and housing of pets with other animals have been identified to influence *Campylobacter* spp. transmission dynamics (Tauxe 1992, Torre & Tello 1993). Although pets have been identified as healthy carriers and sometimes with sub-clinical infections, dogs infected with other gastrointestinal infections were found to increase the prevalence and the species richness (presence of multiple species) of *Campylobacter* spp. (Lopez et al. 2002, Acke et al. 2006, 2009, Chaban et al. 2010).

The majority of dog *C. jejuni* isolates were assigned to the ST-45 complex and these accounted for 53% of the total number of isolates that were characterised. The only genotype found during the warmer months was ST-45 (n = 5) and the remainder of ST-45 and the other STs were found during the autumn and winter (ST-52, ST-50, ST-422, ST-45, ST-474, ST-696 and ST-3961). There is a possibility that the presence of the genotypes ST-3961 and ST-696 (duck associated genotypes) in the dog faecal material, could have resulted from the contamination of the dog faecal material with duck faecal material, while the owners were picking up the faecal material from the ground. The population

differentiation using AMOVA suggested that variation resides within the *C. jejuni* population of dogs and cats and among the cat and dog *C. jejuni* populations considered separately.

It can be hypothesised that the high spring and summer faecal prevalence of *Campylobacter* spp. and the high prevalence of ST-45 may contribute to the overall burden of campylobacteriosis in humans at this time of the year

In conclusion, this study has quantified the prevalence, the population structure and the genetic diversity of *C. jejuni* in dogs in an urban area in provincial New Zealand. Further, we have demonstrated the similarities between *C. jejuni* genotypes isolated from dogs, cats and humans which implies that dogs and cats may play a role in human infection. The *por*A and *fla*A alleles found in both dogs and cats in this study were identical to those from human campylobacteriosis recorded in the PubMLST database. This provides additional evidence that pets may represent a risk factor for human infection and that this risk may vary on a seasonal basis. Although the overall influence of dogs and cats on human campylobacteriosis case numbers is likely to be small, it is likely that they may represent a risk particularly for small children, elderly people and immuno-compromised individuals (Lopez et al. 2002).

Further research into the molecular epidemiology of campylobacteriosis combining *Campylobacter* spp. in pets and human clinical cases will improve our understanding of the transmission pathways and the dynamics of disease in humans. In addition, we suggest that in the absence of a standard isolation technique for *Campylobacter* spp. from faecal material, simultaneous application of direct-PCR and culture techniques should be employed for effective recovery of bacteria from faeces.

Acknowledgments This work is dedicated to the memory of my late supervisor Dr. Grant Hotter who died suddenly in 2009. He was actively involved in supervising me and helped me to carry out this work presented here.

CHAPTER 5

Phylogeny of housekeeping and *por*A genes from seven *Campylobacter jejuni* ST-474 genomes

Abstract – *C. jejuni* multilocus sequence type 474 (ST-474) is an internationally rare genotype that is endemic in New Zealand. The internal fragments of seven housekeeping genes and the structural gene, *por*A of *C. jejuni* have been widely studied using multilocus sequence typing (MLST). In this study seven ST-474 genomes (four from chickens and three from human clinical cases) and 12 *C. jejuni* reference genomes were compared. A subset of 25 housekeeping genes were analysed to investigate the extent of the similarity shared between the seven ST-474s which, in turn, were compared with the reference genomes. In addition, the structural gene, *por*A from the seven ST-474 genomes and the reference genomes were compared. The genes under investigation were divided into two subsets: the first was the genes that comprise the *C. jejuni* MLST scheme (n = 7) and second comprised a set of genes previously used in MLST schemes for other bacteria (n = 18). The MLST alleles *asp*A, *glt*A, *glm*M, *tkt* and *unc*A/*atp*A were found to be identical across the reference genomes CJJ11168, CJJ84-25, CJJIA3902 and CJJCF93-6 and the ST-474 genomes. In contrast, the full length genes which contain the identical MLST alleles showed variations at both the nucleotide and protein levels, at least between the reference and ST-474 genomes. Three genes, (*fum*C, *pyc*A and *trp*C) from the second subset showed significant molecular variations within the seven ST-474 genomes.

The genes investigated in this study showed a bimodal distribution of relative evolutionary rates and selection pressures that were indicative of functional constraints. The presence of recombination was evident both in the housekeeping and *por*A genes. There was a positive correlation between the guanine-cytosine (GC) content variance and the number of recombination sites. Genes that showed wider GC variance had a relatively high number of recombination sites. Genes that shared identical GC contents had identical codon bias indices. The majority of housekeeping genes shared their ancestry with the ST-21 complex as predicted by the MLST allelic profile whereas the *fts*Z, *gap*A, *hem*N, *sdh*A and *pyc*A alleles were found to be unique to ST-474 forming separate branches on the maximum likelihood phylogenetic trees. In addition, the *por*A gene sequences from the reference and ST-474 genomes showed extensive recombination, evidenced by variations in external loops one, four, six and eight. The human isolate H22082 was closer to the ST-22 complex while the remainder were closer to the ST-21 complex. This study demonstrates the robustness of MLST housekeeping alleles as predictors of the ancestry of the majority of genes in a genome and provides further insights into the molecular variation and similarities among the housekeeping genes within the seven *C. jejuni* ST-474 genotypes.

Vathsala M, French N, Biggs, P J, Stevenson M, Marshall J, and Hotter G, (2011). Phylogeny of housekeeping and *por*A genes from seven *Campylobacter jejuni* ST-474 genomes

5.1 Introduction

The impact of recombination (homologous, non-homologous or illegitimate) on the evolution of microorganisms has been evidenced as the major factor of microbial evolution (Feil et al. 2001, Spratt et al. 2001, Hanage et al. 2005, Fraser et al. 2007, 2009). However, the rate of recombination may differ greatly amongst different bacterial species; while some species recombine more frequently to have multiple recombinational events than mutations that render them weakly clonal, in other species it appears to be a rare incident leading to distinct clonal lineages (Spratt et al. 2001, Hanage et al. 2005, 2006, Fraser et al. 2007). Studies of genetic diversity in the bacterial kingdom have shown that bacteria form clusters of genetically related strains and that extensive recombinations among related clusters have been regarded as normal rather than exceptional events (Giovannoni 2004, Venter et al. 2004). In addition, compositional analytical studies of bacterial genomes have demonstrated considerable proportions of horizontally acquired genes in most bacterial genomes (Dauga 2002, Yang 2002, Nakamura et al. 2004, Gupta 2006, McQuiston et al. 2008). However, not every single gene is involved in recombination or horizontal gene transfer (Rivera et al. 1998, Jain et al. 1999).

Rivera and his co-workers 1998 used information on whole genome sequences of *Saccharomyces cerevisiae* (a eukaryote), *Synechocystis 6803* (a cyanobacterium), *Escherichia coli* (a proteobacterium) and *Methanococcus jannaschii* (a methanogen). Rivera et al. (1998) differentiated two striking inheritance patterns of genes (Riley 1993, Rivera et al. 1998). The lineages of the genes were broadly classified into informational and operational or housekeeping genes. The informational genes include genes of translation (T), transcription (S), and replication (R) and also the ATPases, GTPases (G) and tRNA synthetases whereas the operational genes are those involved in cell operations such as amino acid synthesis (A), biosynthesis of co-factors (B), cell envelope proteins (C), energy metabolism (E), intermediary metabolism (I), fatty acid and phospholipid biosynthesis (L), nucleotide biosynthesis (N), and regulatory genes (Z) (Rivera et al. 1998). The operational genes are the most modular genes in the cells that are inclined to be horizontally transfered or recombined most often (Jain et al. 1999, Ma & Zeng 2004); hence the use of single genes for comparing the populations and/or for comparing the phylogenies of bacterial species may not resolve the puzzle of phylogenetic relationships.

Multilocus enzyme electrophoresis (MLEE) and mulitlocus sequence typing (MLST) are techniques devised to sub-divide bacterial populations (Maiden et al. 1998). While MLEE targets the cellular enzymes encoded by the genes in a genome, MLST exploits the internal fragments that are approximately 470 to 500 base pairs of the housekeeping genes (Maiden et al. 1998, Maiden 2006). Further, MLST constructs allelic profiles based on variations in the nucleotide sequences of (usually) seven housekeeping genes, and bacterial isolates are grouped into different clusters (Maiden et al. 1998). The characterisation of closely related species using multiple loci has allowed the analysis of relationships between different bacterial species, which is often referred to as an extension of MLST, called multilocus sequence analysis (MLSA) (Hanage et al. 2005, Gevers et al. 2005). Some bacterial pathogens are monomorphic with little sequence diversity (e.g. *Salmonella enterica subsp.enterica* serovar Typhi) that serve as promising models of evolution. Some are polymorphic with greater sequence diversity (e.g. *Neisseria meningitis* and *Campylobacter jejuni*) (Linz et al. 2000, Dingle et al. 2001, Achtman & Wagner 2008) that have blurred genomic signals of phylogenetic history due to extensive recombinational events (Wirth et al. 2006, Achtman & Wagner 2008).

C. jejuni, a zoonotic pathogen that colonises the gut of a wide variety of birds and mammals, has been attributed to the majority of bacterial gastroenteritis cases in developed countries (Humphrey et al. 2007). In the majority of cases the disease caused by *C. jejuni* is self limiting however on rare occasions there can be serious sequelae such as Guillain-Barré syndrome and reactive arthritis (Zia et al. 2003). The natural competency and the plasticity of *C. jejuni* were not investigated until Dingle et al. (2001) designed an MLST scheme for *C. jejuni* which has subsequently been exploited to structure and investigate the association of *C. jejuni* populations with different hosts and the environment from which human clinical infection originated (Colles et al. 2003, Miller et al. 2006, Colles et al. 2008, Kwan et al. 2008, Wilson et al. 2008, Carter et al. 2009, Mullner et al. 2009). However, chromosomal DNA and plasmid transformation were documented in *C. jejuni* and *C. coli* under laboratory conditions using shuttle vectors (Wang & Taylor 1990). The *Campylobacter* PubMLST database has an extensive archive of these MLST data for *C. jejuni* and other bacterial species.[1] Further, Wilson et al. (2009) used the population genetics-phylogenetics approach to demonstrate the massive evolutionary potential of *C.*

[1]URL: (`http://pubmlst.org/Campylobacter`)

jejuni. These authors further inferred that a mutation at any site can occur somewhere in the population within the space of a week and that recombination plays a major role in the generation of diversity at twice the rate of mutation *per se*. Furthermore, the gene flow between *C. jejuni* and *C. coli* due to recombination has been demonstrated in gene sequences of these two species illustrating the rapidity of evolution in the *Campylobacter* genus (Sheppard et al. 2009, Wilson et al. 2009). Hybrid alleles were shown to be more common in four of the seven MLST loci (*tkt*, *asp*A, *glt*A and *gly*A) in *C. jejuni* (Sheppard et al. 2009). Similarly, the major outer membrane protein or MOMP gene of *C. jejuni* (interchangeably designated as *por*A) has been identified to show strong evidence of recombination (Clark et al. 2007).

MOMP is found on the surface of the bacterial cell and is important for the passive movement of nutrients between the external environment and the cell (Clark et al. 2007). The *Campylobacter* spp. MOMP is involved in adherence to host cells, antibiotic resistance and cytotoxicity (Page et al. 1989, Moser et al. 1997, Zhang et al. 2006a). MOMP plays a major role in the adaptation of *Campylobacter* spp. to various environments and hosts (Zhang et al. 2000, Huang et al. 2007). The *por*A/MOMP gene of *Campylobacter* spp. is comprised of seven highly variable regions interspersed among conserved regions, where the highly variable regions form the irregular external loops connected by 18 β strands and short periplasmic turns (Zhang et al. 2000, Labesse et al. 2001). These exposed surface epitopes have been identified to be the source of antigenic variation and this MOMP antigen has been identified as a potential vaccine candidate (Huang et al. 2007). Similarly, phylogenetic analysis of *por*A genes from Guillain-Barré Syndrome (GBS) strains has classified the *por*A alleles into three groups within *C. jejuni* and *C. coli* (Clark et al. 2007) where, the third group showed evidence of recombination. Further, it has been inferred that the *por*A alleles sequenced to date have demonstrated limited variation within the assigned three groups. Data from *por*A alleles may confer phylogenetic differentiation among *C. jejuni* isolates separating the isolates that differ according to functional, biological and virulence properties (Clark et al. 2007). Similarly, the combination of the *por*A gene with that of the MLST alleles has been shown to be highly discriminatory for *C. jejuni* and *C. coli* typing particularly in long-range comparisons and short-term epidemiological investigations (Dingle et al. 2008). A longitudinal study was conducted in a defined human population of $\approx$ 600,000 for one year in a Hospital Clinical Microbi-

ology Laboratory, in the UK. A total of 196 distinct *por*A variants were identified from *Campylobacter* culture positive samples and a similar clustering pattern was observed as the previous study in which the third group contained both *C. coli* and *C. jejuni* sequences suggesting recombination (Cody et al. 2009). In addition, the stability of the *por*A gene within human patients with prolonged infection (5 and 98 days) was investigated using repeated sampling from two of the patients in this study where, mutations were detected in the *por*A genes from both patients illustrating the potential of the gene to evolve under certain circumstances (Cody et al. 2009).

Although population genetics of *C. jejuni* provides important insights into the clonal relationships and serves as an efficient molecular epidemiological investigation tool, the recent availability of complete *C. jejuni* genome sequences from diverse strains has provided an opportunity to discover the functional mechanisms and pathways that are essential for the survival and growth of *C. jejuni*. This in turn will provide suggestions about how those genes involved in different metabolic pathways can evolve in response to different environments to render *C. jejuni* a successful survivor in the environment and the food chain.

Relationship between *C. jejuni* ST-474 and New Zealand

New Zealand has, until recently, ranked first in the world for campylobacteriosis notification rates (Baker et al. 2007, Marler 2010). *C. jejuni* multilocus sequence type 474 (ST-474) is an internationally rare genotype that accounted for 24% to 34% of human clinical cases in New Zealand (French 2008)[2] and was strongly associated with poultry, particularly, in urban areas (French 2008, Mullner et al. 2009, 2010). Further, among the *Campylobacter* clusters identified from five different district health boards in New Zealand in 2006, *C. jejuni* ST-474 accounted for 32 of 112 isolates and has been suggested to be an endemic sequence type present in New Zealand (McTavish et al. 2008). Two sub-strains of *C. jejuni* ST-474 *fla*SVR 14 (a short variable region within the *fla*A genes that are conserved among the *C. jejuni* virulent strains (Meinersmann et al. 1997)) were isolated during the same time period and from the same geographical location (August 2005, Palmerston North, New Zealand). One was isolated from a chicken and the

[2]URL:(http://www.foodsafety.govt.nz/elibrary/industry/enhancing-surveillance-potentially-research-projects-2/Campy_Attribution_Manawatu.pdf/)

other from a human clinical case, their genomes were sequenced using next generation Solexa sequencing technology. Their draft genomes were submitted to GenBank[3] (French et al. 2009b, Biggs et al. 2011). The authors found 93% of the genes to be identical between the two ST-474 isolates and 103 genes to differ by at least a single nucleotide with 72 of them showing non-synonymous substitutions. Furthermore, they inferred that 5% of the differences were due to mutations and 95% of them were as a result recombination (Biggs et al. 2011). Additionally, a non-homologous recombination event was identified in the human isolate (H22082) with the insertion of two extra genes, whereas the chicken isolate (P110b) did not show such an event. Although the housekeeping alleles of different genotypes of *C. jejuni* share 86% nucleotide sequence identity, the seven housekeeping genes used in the MLST scheme cover lesser than $< 0.2\%$ of the entire genome (Dingle & Maiden 2005). The phylogeny and the ancestral lineages of individual genes may differ when examined at the genome level, and these differences may not be detected even when MLST is combined with other antigenic genes such as *por*A. Moreover, studies on individual housekeeping genes are scarce and most studies have either focused on the congruence of phylogenies interpreted by rRNA, or a few housekeeping genes or whole genomes (Puhler et al. 1989, Bustamante et al. 1995, Bult et al. 1996, Martin 1999, Dorrell et al. 2001, Vandecasteele et al. 2001, Hinode et al. 2002, Jordan et al. 2001, Velayudhan & Kelly 2002, Whittam & Bumbaugh 2002, Wolf et al. 2002, Coenye & Vandamme 2003, Fuglsang 2003, Pearson et al. 2003, Viscidi & Demma 2003, Carrillo et al. 2004, Karenlampi et al. 2004, Ma & Zeng 2004, Nakamura et al. 2004, Taboada et al. 2004, Venter et al. 2004, Foerstner et al. 2005, Fouts et al. 2005, Chen et al. 2006, Gupta 2006, Klancnik et al. 2006, Musto et al. 2006, Rooney et al. 2006, Miller et al. 2007, Sorek et al. 2007, Hughes et al. 2008, Rocha 2008, Liu et al. 2009, Lefebure & Stanhope 2009, Rocha & Feil 2010). The molecular differences in housekeeping genes within genomes that belong to a single genetic type (sequence type) have not been studied at the individual gene level except for a sub-region among seven housekeeping genes in MLST (section 2.4.4). Although the sequences of seven housekeeping genes illustrate the genetic diversity that relfects past evolutionary events in those genes, the underlying molecular variations that can be expected due to the evolutionary forces in at least, other metabolic genes are obscure. There is no evidence to suggest that MLST alleles are iden-

[3]URL:(`http://www.ncbi.nlm.nih.gov/genbank/`)

tical at their full length level. It is not known if a detailed knowledge of MLST alleles can capture the true evolutionary history of the other genes in a genome.

The major evolutionary forces that shape the genetic diversity amongst organisms are genetic drift, mutation, migration and selection which directly influence the nucleotides in a DNA sequence (Rocha & Feil 2010). The genetic consequences brought about by these evolutionary forces are measured using different mathematical and statistical methods and models.

Models to study evolution

Evolution in genes and/or genomes are measured in terms of: the nucleotide composition (adenine, cytosine, guanine and thymine or uracil), particularly the guanine-cytosine (GC) content of a given gene or genome; the nucleotide substitutions or differences between genes, the selection pressures (negative or purifying, neutral and positive or advantageous); and amino acid usage and/or codon usage by the gene or genome (Rocha & Feil 2010). The evolutionary measures in turn provide opportunities to better understand the detailed mechanics of individual genes and the whole genome.

Guanine-cytosine content

The guanine-cytosine (GC) content of bacteria varies both at the genomic and the gene level. GC content has been shown to be driven by selection, mutational bias and biased recombination-associated DNA repair (Genereux 2002). The genetic stability, which is defined as 'a measure of the resistance to change, with time, of the sequence of genes within a DNA molecule or of the nucleotide sequence within a gene'[4], has been shown to be affected by the base pairing and stacking of GC and adenine – thymine (AT) while GC richness provides greater stability (Yakovchuk et al. 2006). The DNA stability is not only fundamental for the overall structure and thermal stability and energetics of DNA, but it also has significant influences on a number of biological processes within a cell (Gueron et al. 1987, Frank-Kamenetskii 1987, Yakovchuk et al. 2006). In bacteria, base composition variation is thought to affect both coding and non-coding sites within a gene

[4]URL:(`http://http://www8.nos.noaa.gov/coris_glossary/index.aspx?letter=g`)

and/or genome, as bacteria generally possess little intergenic DNA sequences (Hildebrand et al. 2010). The variation in base composition is not restricted to any particular group of bacteria, rather it is a consequence of differences in the patterns of evolutionary events (Freese 1962). Variation in the GC content is dependent on the mutational patterns and/or the evolutionary events that had occurred in a given nucleotide sequence (Sueoka 1961, Freese 1962). GC content has also been shown to be correlated with various factors such as genome size (Bentley & Parkhill 2004), the living nature of bacteria (bacteria that rely on their host for survival are rich in AT) (Rocha & Danchin 2002, Woolfit & Bromham 2003), the environment (Foerstner et al. 2005), aerobiosis (aerobic prokaryotes display a significant increase in genomic GC%) (Naya et al. 2002), nitrogen utilisation (nitrogen fixing bacteria are GC rich) (McEwan et al. 1998) and temperature (Galtier & Lobry 1997, Hurst & Merchant 2001, Musto et al. 2006). Further, the conversion of GC→AT has been found to be more common in comparison to AT→GC conversions in GC rich bacteria and an opposite pattern is found in AT-rich bacteria such as *C. jejuni* (Hildebrand et al. 2010). Although the causes for the differences in the GC content have not been clearly defined, GC content at the third codon position (GC3) and the conversion of GC→AT and AT→GC at this position has been reported to occur in favour of selection for GC content of the genome (Akashi 1995, Hildebrand et al. 2010) and it has been suggested this has had a great impact on evolution.

Nucleotide substitutions and selection pressures

Differences in rates of nucleotide substitutions in protein-coding genes occur as a result of three major selection processes such as (negative) purifying selection against deleterious nucleotide substitutions or mutations, random genetic drift of neutral mutations and (positive) adaptive selection (Patthy 2008). Synonymous or silent mutations/substitutions do not lead to an amino acid change in contrast to non-synonymous substitutions which do. The synonymous substitutions (d_S) (also interchangeably referred to as K_s) are neutral and do not cause any deleterious effects to a protein coding gene and are often fixed in a population (Kimura 1968, Patthy 2008). The non-synonymous substitutions (d_N) (also interchangeably referred to as (K_a) that lead to amino acid change play a major role in protein evolution. Several mathematical and statistical models have been developed to measure and infer evolution on the basis of nucleotide differences, initially in eukaryotes

that were subsequently adapted to analyse evolution in prokaryotes. The models that have been widely accepted and those that are in routine usage include:

1. Substitution models (nucleotide and amino acid) which measure the relative occurrence of nucleotide substitutions among the 4 nucleotides and the relative occurrence of substitutions of amino acids among 20 amino acids in a lineage. These models in turn have undergone several improvements (Jukes & Cantor 1969, Kimura 1980, Felsenstein 1981, Hasegawa et al. 1984, 1985, Nei & Gojobori 1986, Jones et al. 1992, Tamura & Nei 1993, Yang 1994b),

2. Codon substitution models are similar to nucleotide-substitution models but these models consider a sense codon as an unit of evolution (Hasegawa et al. 1985, Yang et al. 1998), where the relative codon frequencies are calculated as a result of nucleotide substitutions.

3. Site models are those that use methods for testing the effect of selection at individual sites in an alignment of protein coding DNA sequences (Suzuki & Gojobori 1999, Yang et al. 2000).

All of the three models mentioned above use the nucleotide frequencies as the basic requirement, where a relationship has been developed based on the frequencies and the type of nucleotide substitution, such as synonymous or non-synonymous, that had occurred in a given DNA sequence. The ratio of d_N to d_S (ω), is a measure of natural selection that has taken place in a protein coding sequence (Whelan et al. 2001). Values of ω less than 1.0, equal to 1.0, and greater than 1.0 signify purifying, neutral and positive Darwinian selections respectively (Tanaka & Nei 1989, McDonald & Kreitman 1991, Muse & Gaut 1994, Whelan et al. 2001). However, it should be noted that non-synonymous substitutions that favour positive selection are expected to occur only at sites that are critical and essential for an advantageous function (Patthy 2008). The number of such substitutions in turn affects the overall d_N/d_S ratio of a gene, where fewer numbers of non-synonymous substitutions may be overridden by purifying selections that have occurred in other sites within a protein coding gene (Muse & Gaut 1994, Patthy 2008). Therefore an overall d_N/d_S ratio might mask sites with positive selection on a gene. Hence identification of selection pressures at the individual codon and/or site level is required in most instances

when protein-coding sequences are analysed (Patthy 2008). Non-synonymous substitutions have been classified further into three distinct classes such as deleterious, neutral and advantageous (Ohta 2002, Hughes et al. 2008). Deleterious variants are removed by purifying selection and may lead to a false inference of positive selection (Hughes et al. 2008). Further, an ongoing purifying selection is said to be identified by the heterozygosity in gene sequences with lesser number of non-synonymous substitutions (Hughes et al. 2008). The substitution rates across sites are variable where each site has been assumed to have a different evolutionary rate in a gene (Jin & Nei 1990, Yang 1993). A gamma distribution with a scale parameter has been developed that measures the mean evolutionary rate variability among sites (Jin & Nei 1990, Tamura & Nei 1993). These rates are scaled such that the average evolutionary rate across all sites in a coding gene is 1.0. This means that sites showing a rate less than 1.0 are evolving slower than average, and those with a rate greater than 1 are evolving faster than average (Tamura & Nei 1993, Tamura & Dudley 2007).

Codon usage bias index

More than one codon can be used by bacteria to encode an amino acid where the same amino acid can be encoded by different codons, dependent on the preferences of individual bacteria (Ikemura 1985, Snyder & Champness 1997). The codon preferences, in turn, has been shown to be influenced by the tRNA concentrations which is related to the nucleotide base composition of the gene and/or an organism (Ikemura 1985, Snyder & Champness 1997, Kanaya et al. 1999, Tuller et al. 2011). The tRNA abundances in a genome is one of the major selective pressures that determines the synonymous codon usage and the differences in codon usage reflects the differential evolution in organisms (Lynn et al. 2002, Tuller et al. 2011). In addition, CBI is also a direct indicator of recombination and HGT (Marais et al. 2001, Fuglsang 2003). The indices that are generally estimated for examining the codon usage biases as measures of evolution are the overall codon usage bias index (CBI) and the scaled chi square codon usage bias index (Bennetzen & Hall 1982, Shields et al. 1988). CBI measures the extent to which a gene uses a subset of optimal codons. CBI will be equal to 1.0 in a gene with extreme codon bias and zero in a gene with random codon usage. A negative CBI value occurs when the number of optimal codons is less than expected by random change. The scaled chi square index is

a measure of bias in silent codon usage. It provides a measure of the general synonymous codon usage that is independent of gene length for genes that have more than 100 codons, and provides a measure of the third codon evolution (Shields et al. 1988).

Given the substantial evidence of the dynamic nature of *C. jejuni* and differences between identical MLST genotypes at the genome level as evidenced by Biggs et al. (2011), it is prudent to perform an extended analysis of multiple genomes with identical MLST genotypes in order to obtain an improved understanding of evolution which is an important strain of *C. jejuni* in New Zealand. With this background of information, the aims of the present study were to: (1) investigate the full length genes of MLST alleles and a selected subset of housekeeping genes involved in different metabolic functions; (2) examine the selected subset of housekeeping and *por*A genes from *C. jejuni* ST-474 and reference genomes for different evolutionary events, (3) compare and identify the lineage for selected individual housekeeping gene from seven *C. jejuni* ST-474 isolates; (4) evaluate the phylogenetic congruence and/or ancestry predictions obtained from the MLST housekeeping genes and a subset of full-length housekeeping genes; and (5) compare the *por*A gene across the seven ST-474 genomes and to investigate the concordance between the phylogenies estimated by the housekeeping genes and the *por*A gene. This work should provide useful insights into the evolution of housekeeping genes that are fundamental to our understanding of the process of evolution, species divergences and gene function (Yang 2002).

5.2 Experimental procedures

This study does not involve exhaustive statistical modelling and/or computer simulations instead, the wealth of specialised genetic software tools has been exploited to draw inferences on genetic parameters such as GC content, selection pressure (Tajima's D and omega values), codon usage, recombination events and the closest *C. jejuni* ancestor for the investigated genes of the ST-474 genome. Seven *C. jejuni* ST-474 genomes were fully sequenced at the Massey Genome Service, Palmerston North, New Zealand, using next generation Solexa sequencing technology.

5.2.1 Reference *C. jejuni* genomes

Details of the reference *C. jejuni* genomes used in this study are provided in Table 5.1. Twelve sequenced *C. jejuni* genomes were used to compare and identify the closest ancestors of 25 metabolic housekeeping genes (MLST housekeeping genes inclusive) from seven *C. jejuni* ST-474 genomes. The gene sequences were downloaded from the GenBank database.[5] Allele numbers for seven housekeeping genes and sequence types (ST) of reference *C. jejuni* genomes (n = 12) were identified by comparing the internal fragments of genes (*asp*A, *gln*A, *glt*A, *gly*A, *glm*M, *tkt* and *unc*A) by using the 'sequence query–query an allele sequence' option in the query database functionality available in the PubMLST database.

[5]URL:(`http://www.ncbi.nlm.nih.gov/genbank/`)

Table 5.1: *Campylobacter jejuni* subsp. *jejuni* reference genomes and their MLST clonal complexes

Species	Strain	Size (Mb)	%GC	ORFs	Disease/source	Year and place of isolation	GenBank	MLST : ST (CC)
Campylobacter jejuni subsp. *jejuni*	NCTC 11168	1.64	30.5	1643	Clinical Food poisoning	1977, UK	AL111168	43 (21)
Campylobacter jejuni subsp. *jejuni*	RM1221	1.78	30.3	1838	Chicken	2005*, USA	CP000025	354 (354)
Campylobacter jejuni subsp. *jejuni*	81-176	1.6	30.6	1653	Clinical Food poisoning	1981, USA	CP000538	604 (42)
Campylobacter jejuni subsp. *jejuni*	81116	1.63	30.5	1626	Clinical Food poisoning	2007, UK	CP000814	267 (283)
Campylobacter jejuni subsp. *jejuni*	CG8421	1.6	30.4	1512	Clinical Food poisoning	ns	ABGQ00000000	1919 (52)
Campylobacter jejuni subsp. *jejuni*	HB93-13	1.7	30.6	1710	Clinical GBS	2006*, USA	AANQ00000000	22 (22)
Campylobacter jejuni subsp. *jejuni*	CG8486	1.65	30.4	1425	Clinical Food poisoning	ns	AASY00000000	2943 (574)
Campylobacter jejuni subsp. *jejuni*	CF93-6	1.67	30.5	1757	Clinical MFS	2006*, Japan	AANJ00000000	883 (21)
Campylobacter jejuni subsp. *jejuni*	84-25	1.67	30.4	1748	Clinical Meningitis	2006*, USA	AANT00000000	21 (21)
Campylobacter jejuni subsp. *jejuni*	260.94	1.65	30.5	1716	Clinical GBS	ns	AANK00000000	362 (362)
Campylobacter jejuni subsp. *jejuni*	IA3902	1.64	30.5	1718	Sheep abortion	2010, USA	CP001876	8 (21)
Campylobacter jejuni subsp. *doylei*	269.97	1.85	30.6	2094	Human blood	2007, UK	CP000768	1845 (UA)

MLST: Multilocus sequence type.
ST: Sequence type.
CC: Clonal complex.
GBS: Guillain-Barré syndrome.
ORFs: Open reading frames.
Mb: Mega bases.
GC: Guanine:cytosine.
ns: not stated.
*:Date of start of project
U/A:Unassigned

5.2.2 *C. jejuni* isolates, DNA preparation and sequencing

We used seven *C. jejuni* samples belonging to ST-474 which were isolated and characterised at the Hopkirk Research Institute, IVABS, Massey University, Palmerston North, New Zealand. Five isolates (P179a, P569a, P694a and P110b) were derived from samples of poultry meat obtained from local supermarkets between August, 2005 and February, 2009. The remaining three isolates were from human clinical cases of campylobacteriosis obtained between August 2005 and February, 2009 (H22082, H704 and H73020). The genomic DNA from the isolates were extracted from pure bacterial colonies grown on blood agar plates (Fort Richard, Auckland, NZ) using a Wizard Genomic DNA Purification Kit (Promega) following the manufacturer's instructions. Nebulisation of genomic DNA and sequencing reactions were performed at the Massey Genome Service, Massey University, Palmerston North.

5.2.3 *de novo* genome assembly and gene prediction

Genome assembly and gene predictions were carried out at the Massey Genome Service by the third author (PJB). The nucleotide sequences of the housekeeping genes were retrieved from the gene predictions from the seven *C. jejuni* ST-474 genomes.

5.2.4 Selection of metabolic housekeeping genes

As there is no complete list of housekeeping (HK) genes described to date, the metabolic genes that are used in routine MLST typing schemes for different bacterial species were downloaded from the PubMLST database and from previous phylogenetic studies that used housekeeping genes for analysis (Wertz et al. 2003, Christensen et al. 2004, Margos et al. 2008). The gene names were matched with the *C. jejuni* NCTC 11168 genome (GenBank accession number NC_002163). Following the identification of metabolic genes, a total of 18 housekeeping genes were selected and these were further examined for their presence in other *C. jejuni* reference genomes. Full length nucleotide sequences of the selected set of housekeeping genes were retrieved in FASTA format. The list of genes (MLST/first subset and second subset), gene names and their function are provided in Tables 5.2 and 5.3, respectively. The positions of the selected subset of genes are shown

on the *C. jejuni* NCTC 11168 (GenBank accession number NC_002163) circular genome in Figure 5.1.

Table 5.2: Metabolic genes used in the *C. jejuni* MLST scheme (first subset of genes)

Gene	Gene_ID	Function	Name of the gene
*asp*A	Cj0087	Central intermediary metabolism	Aspartase
*gln*A	Cj0699c	Central intermediary metabolism	Glutamine synthetase
*glt*A	Cj1682c	Central intermediary metabolism	Citrate synthase
*gly*A	Cj0402	Central intermediary metabolism	Serine hydroxy methyl transferase
*glm*M	Cj0360	Central intermediary metabolism	Phospho glucosamine mutase
tkt	Cj1645	Central intermediary metabolism	Transketolase
*unc*A/*atp*A	Cj0105	Central intermediary metabolism	ATP synthase a subunit

Table 5.3: Metabolic genes selected from the MLST schemes of different bacterial species (second subset of genes)

Gene	Gene_ID	Function	Name of the gene
*arg*F	Cj0994c	Amino acid biosynthesis - Glutamate family	Probable ornithine carbamoyltransferase
*aro*E	Cj0405	Amino acid biosynthesis -Aromatic amino acid family	Probable shikimate 5-dehydrogenase
*atp*D	Cj0107	Energy metabolism	Probable ATP synthase F1 sector beta subunit
*dap*E	Cj1048c	Amino acid biosynthesis - Aspartate family	Probable succinyl-diaminopimelate desuccinylase
*fts*Z	Cj0696	Cell division	Probable cell division protein
*fum*C	Cj1364c	Energy met.TCA	Probable fumarate hydratase
*gap*A	Cj1403c	Energy metabolism	Glyceraldehyde 3-phosphate dehydrogenase
*glt*B	Cj0007	Energy met.TCA	Probable glutamate synthase (NADPH) large subunit
*gro*EL	Cj1221	Heat Shock	60 kD chaperonin
*hem*N	Cj0580c	Oxygen-independent coproporphyrinogen-III oxidases	Probable oxidoreductase
*ilv*D	Cj0013	Amino acid biosynthesis - Glutamate family	Probable dihydroxy-acid dehydratase
*inf*B	Cj0136	Protein translation	Probable translation initiation factor IF-2
*lys*A	Cj0314	Amino acid biosynthesis - Aspartate family	Probable diaminopimelate decarboxylase
*nuo*D	Cj1576c	Energy metabolism -Respiration - Aerobic	Probable NADH dehydrogenase I chain D
*pyc*A	Cj1037c	Central intermediary metabolism - Gluconeogenesis	Possible pyruvate carboxylase A subunit
*sdh*A	Cj0437	Energy metabolism -Tricarboxylic acid cycle	Probable succinate dehydrogenase flavoprotein subunit
*trp*B	Cj0348	Amino acid biosynthesis	Probable tryptophan synthase beta chain
*trp*C	Cj0498	Amino acid biosynthesis - Aromatic amino acid family	Probable indole-3-glycerol phosphate synthase

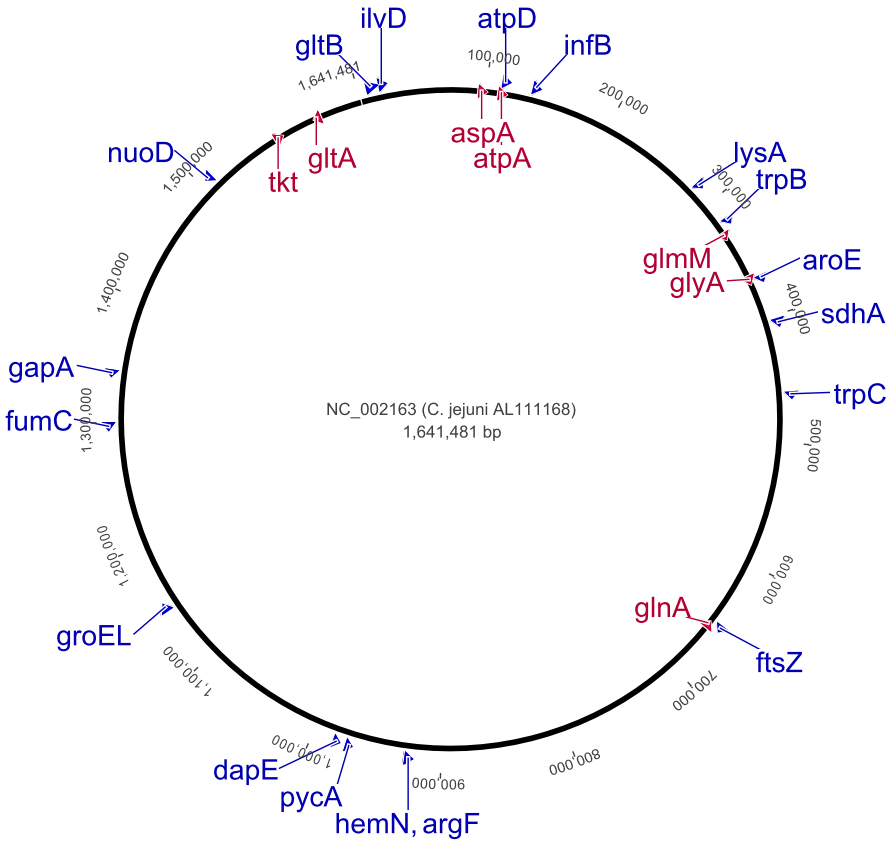


Figure 5.1: *C. jejuni* circular genome showing metabolic genes. The positions of the first and second subset of metabolic housekeeping genes are shown in this picture. The MLST housekeeping genes are shown in pink fonts and the second subset is shown in blue fonts. The genes were identified and positioned on the *C. jejuni* NCTC 11168(GenBank accession number NC_002163) using Geneious molecular genetic software v5.3.4

5.2.5 Retrieval of gene sequences from *C. jejuni* ST-474 gene predictions

The BLAST+ application was downloaded from The National Center for Biotechnology Information (NCBI)[6] to perform stand-alone BLAST searches on the ST-474 gene predictions in order to find the gene/nucleotide sequences from the *C. jejuni* NCTC 11168 genome. For a given metabolic housekeeping gene from the CJJ11168 genome, the BLAST hits with maximum identity, alignment length and highest bitscore were retained and the respective open reading frames (ORFs) were retrieved from the BLAST database in FASTA format. The best hits for all of the 25 genes that corresponded with the above selection criteria are consolidated and presented as a blast table in Appendix C. For the

[6]URL: (`ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST`)

reference genomes *C. jejuni doylei* 269.97 and CJJCG8486, the orthologs for the *sdh*A gene could not be obtained from their genome sequences. As a result the *sdh*A gene sequence from these genomes could not be included in the comparative analysis along with the other seventeen genomes.

While analysing the seven housekeeping genes, it was observed that the *glm*M and the *pgm* genes were two different genes involved in two different functions in the genome. The *pgm* gene is a phosphoglyceromutase involved in carbohydrate metabolism (in the interconversion of 2-phosphoglycerate) with a locus tag and gene ID of Cj0434 and 904759, respectively.[7] Whereas, the *glm*M gene is a bacterial phosphoglucosamine mutase (PNGM) involved in the interconversion of glucosamine-6-phosphate and glucosamine-1-phosphate in the biosynthetic pathway. The locus tag and gene ID of the *glm*M gene are Cj0360 and 904683, respectively. The primers used for the amplification and sequencing of the *pgm* gene in the MLST scheme of *C. jejuni* were retrieved and assembled against both *pgm* and *glm*M full length genes using Geneious software v5.3.4.[8]) There was a perfect assembly with *glm*M gene in contrast to *pgm*. Further details are provided in Figure C.1 in Appendix C. Therefore, *glm*M was chosen for analysis in place of *pgm* in the comparison of MLST housekeeping genes. There could have been a possibility that *pgm* was mispelt in the place of *pngm* in the MLST scheme for *C. jejuni* when it was originally established. In a recent report by Sheppard et al. (2011), the authors have acknowledged the renaming of the alleles *pgm* and *unc*A to *glm*M and *atp*A respectively in later genome annotations and their names (*pgm* and *unc*A) being retained in the MLST scheme to maintain consistency.

5.2.6 Analysis of Guanine-Cytosine content, codon usage, selection pressure and recombination

The overall GC and GC3 contents of individual housekeeping genes were compared using DnaSP v5 (Librado & Rozas 2009). The number of genes that shared identical GC and GC3 contents between ST-474 and the reference genomes were examined to predict the closer ancestor of ST-474 with respect to the GC contents of the investigated genes. Codon based maximum likelihood analyses were used to investigate selection pressures

[7]URL: (`http://www.ncbi.nlm.nih.gov/protein/218562018`)
[8]URL:(`www.biomatters.com/`)

on individual codons using the Muse & Gaut (1994) and Tamura & Nei (1993) methods implemented with the HyPhy software package within MEGA v5. A test statistic dN - dS was used for detecting codons that have undergone positive selection, where dS and dN denote the number of synonymous substitutions per site (s/S) and the number of non-synonymous substitutions per site (n/N) within a codon, respectively (Tamura & Dudley 2007). A positive value for the test statistic is an indication of an over abundance of non-synonymous substitutions. In addition, Tajima's D test was conducted to test the selection pressure on individual genes as well as individual sites using DnaSP v5 (Tajima 1989, Hudson & Kaplan 1985). CBI and the scaled chi square codon bias indices for individual genes from all *C. jejuni* genomes were estimated using DnaSP v5. The mean relative evolutionary rates of nucleotide sites and codons were estimated using MEGA v5 (Tamura & Dudley 2007) for all of the 25 genes using gene alignments from all the 19 genomes. The model used in MEGA analyses the first, second, third and the non-coding positions for their substitution rates within a codon, and the probability of all the four nucleotides (A, T, G and C) getting substituted within a site and/or a codon in order to predict the evolutionary rates of sites and codons (Jukes & Cantor 1969, Tamura & Dudley 2007).

Inferences on recombination within each gene under investigation were drawn using Dual-Brothers within Geneious v.5.3.4 (Minin et al. 2005). This function uses a double change-point model that detects spatial variation in the phylogenetic tree topology and spatial variation of the nucleotide substitution process (Minin et al. 2005). The gene sequences were aligned using Geneious v5.3.4 and the recombination detection functionality was applied on each gene alignment to infer changes in the topologies and substitution processes. In addition, the aligned sequences were examined using DnaSP v5 to identify the sites involved in recombination. The number of recombination events, referred to as R_m was estimated using DnaSP v5. The relationship between the GC variance and the recombination events was analysed using a linear model by having Rm as a dependent variable and log GC variance and length as independent variables.

5.2.7 Phylogenetic analysis

The nucleotide sequences of all the selected housekeeping genes from seven *C. jejuni* ST-474 genomes were aligned with their respective gene (nucleotide) sequences from twelve

reference genomes using Geneious v5.3.4 (Drummond et al. 2010). Maximum likelihood (bootstrapped to 500 replicates), NeighbourNet and Neighbour joining trees were generated using MEGA v5 (Tamura & Dudley 2007), SplitsTree v4 (Huson & Bryant 2006) and Geneious v5.3.4, respectively, for both individual genes and for the concatenated gene sequences. Full length nucleotide sequences of seven MLST housekeeping genes from the reference (CJJ84-25, CJJCF93-6, CJJIA3902 and CJJ11168) and ST-474 genomes were concatenated separately and a phylogenetic tree was constructed to determine the closest *C. jejuni* ancestor of ST-474 with respect to the seven MLST housekeeping genes. The entire subset of housekeeping genes investigated in this study was concatenated (genes arranged in an alphabetical order for the 19 genomes) and a phylogenetic tree was constructed. The ancestral states of individual genes as well as concatenated sequences of ST-474 genomes were determined based on distance matrices obtained in the process of phylogenetic tree construction.

5.2.8 Analysis of *por*A gene

The major outer membrane protein (MOMP), also known as *por*A, was analysed as a part of the comparison of ST-474 with the reference genomes. Parameters such as GC content, codon usage, selection pressures, relative evolutionary rate and phylogenetic analysis were examined for *por*A also.

5.3 Results

5.3.1 Analysis of full length MLST housekeeping genes

The results of the MLST allelic profiles of the reference genomes are provided in Table 5.4. MLST allelic profiles of ST-474 and reference genomes were compared as the first step of analysis of housekeeping genes. This showed that ST-474 genomes shared five alleles with *C. jejuni* 84-25 (ST-21), four alleles with *C. jejuni* CF93-6 (ST-883) and *C. jejuni* CFIA3902 (ST-8), and three alleles with *C. jejuni* 11168. When ST-474 (n = 7) and *C. jejuni* reference (n = 12) genomes were compared by placing them in a clonal context, *C. jejuni* 84-25 (ST-21) was found to be the closest ancestor of ST-474 genome followed

by other *C. jejuni* genomes that belonged to the ST-21 complex (CJJCF93-6, CJJIA3902 and CJJ11168). However, it should to be noted that this finding is based on a sub-region of seven housekeeping genes used in the *C. jejuni* MLST typing scheme. Therefore, the full length gene sequences of seven MLST housekeeping genes were compared amongst the seven ST-474 genomes (P110b, P569a, P694a, H73020, H22082 and H704) which showed that the genes were indeed identical in their lengths and nucleotide composition.

Table 5.4: MLST allelic profiles of reference genomes and *C. jejuni* ST-474 genomes

Genome	GenBank accession number	*asp*A	*gln*A	*glt*A	*gly*A	*glm*M	*tkt*	*unc*A/*atp*A	ST	CC
CJJ-ST-474		**2**	**4**	**1**	**2**	**2**	**1**	**5**	**474**	**48**
CJJCF93-6	AANJ00000000	**2**	17	2	3	**2**	**1**	**5**	**883**	**21**
CJJ84-25	AANT00000000	**2**	1	**1**	3	**2**	**1**	**5**	**21**	**21**
CJJ11168	AL111168	**2**	1	5	3	4	**1**	**5**	**43**	**21**
CJJRM1221	CP000025	8	10	2	**2**	11	12	6	354	354
CJJHB93-13	AANQ00000000	1	3	6	4	3	3	3	22	22
CJJ81-176	CP000538	1	2	3	27	5	9	3	604	42
CJJ260.94	AANK00000000	1	2	49	4	11	66	8	362	362
CJJ81116	CP000814	2	7	40	4	42	51	1	267	283
CJJCG84-86	AASY00000000	7	53	27	15	11	3	3	2943	574
CJJCG84-21	ABGQ00000000	9	2	2	10	10	3	5	1919	52
CJJIA3902	CP001876	**2**	1	**1**	3	**2**	**1**	6	**8**	**21**
C. Jejuni subsp *doylei*	CP000768	63	164	183	188	27	266	18	1845	U/A

ST:Sequence type
CC:Clonal complex
U/A:Unassigned
Alleles shared with ST-474 are shown in underlined bold fonts

Full length gene sequences of identical MLST alleles (*asp*A, *glt*A, *gly*A, *glm*M, *tkt* and *unc*A) from the respective reference genomes (CJJ84-25, CJJCF93-6, CJJIA3902 and CJJ11168) were compared with ST-474 genomes to examine the lengths and nucleotide composition. The length of genes were identical across the reference and ST-474 genomes (Table 5.5). However, the alleles *asp*A, *glt*A, *gly*A and *tkt* showed synonymous and non-synonymous nucleotide substitutions at the full length gene level. Differences between the six identical alleles amongst the ST-474 and reference genomes are summarised in Table 5.6. The regions (internal-fragments) covered by the MLST alleles on their respective full length gene sequences are presented in Table 5.7. Figures 5.2 a and b; 5.3 a, b, c, d, e and f; and 5.4 further illustrate the differences summarised in Table 5.6 at both nucleotide and amino acid level.

Table 5.5: Metabolic housekeeping genes and their lengths

Genomes	MLST genes and their lengths (base pairs long)						
	*asp*A	*gln*A	*glt*A	*gly*A	*glm*M	*tkt*	*unc*A
CJJ11168	1407	1431	1269	1245	1338	1899	1506
CJJ260.94	1407	1431	1269	1245	1338	1899	1506
CJJ81116	1407	1431	1269	1245	1338	1899	1506
CJJ81-176	1407	1431	1269	1245	1338	1899	1506
CJCG84-21	1407	1431	1269	1245	1338	1899	1506
CJJ84-25	1407	1431	1269	1245	1338	1899	1506
CJCG84-86	1407	1431	1308	1245	1338	1899	1506
CJHB93-13	1407	1431	1269	1245	1338	1899	1506
CJCF93-6	1407	1431	1269	1245	1338	1899	1506
CJD269.97	1407	1431	1269	1245	1338	1899	1506
CJIA3902	1407	1431	1269	1245	1338	1899	1506
CJRM1221	1407	1431	1269	1245	1338	1899	1506
H22082	1407	1431	1269	1245	1338	1899	1506
H704	1407	1431	1269	1245	1338	1899	1506
P110b	1407	1431	1269	1245	1338	1899	1506
P179a	1407	1431	1269	1245	1338	1899	1506
P569a	1407	1431	1269	1245	1338	1899	1506
P694a	1407	1431	1269	1245	1338	1899	1506
H73020	1407	1431	1269	1245	1338	1899	1506

Table 5.6: Synonymous and non-synonymous substitutions found within identical MLST alleles at full length gene level

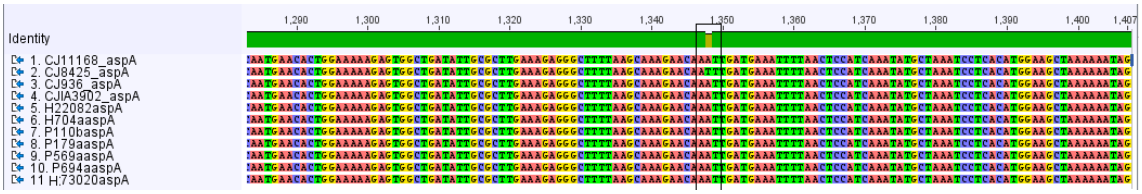
		CJJ11168		CJJCF93-6		CJJ84-25		CJJIA3902		CJJRM1221	
	AN	Syn	Non-syn	Syn	Non-syn	Syn	Non-syn	Syn	Non-syn	Syn	Non-syn
*asp*A	2	-	-	-	-	-	1	-	-	NA	-
*glt*A	1	NA	-	NA	-	5	-	-	-	NA	-
*gly*A	2	NA	-	NA	-	NA	-	NA	-	-	-
*glm*M	2	-	-	-	-	-	-	-	-	NA	-
tkt	1	-	1	3	1		1	-	1	-	-
*unc*A/*atp*A	5	-	-	-	-	-	-	-	-	NA	-

AN:Allele number
Syn:Synonymous
Non-syn:Non-synonymous
NA:No identical alleles found with respect to the genes compared in those genomes
(-):No differences found [Identical at full length level]

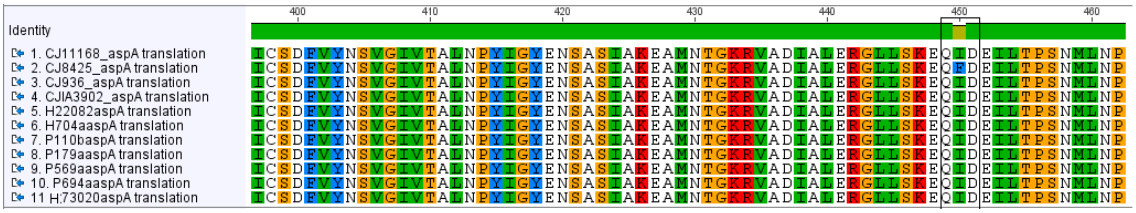
Table 5.7: Positions of MLST alleles covered on full length gene sequences

Gene	Gene_length (bp)	MLST_allele_position
*asp*A	1407	619 bp – 1095 bp
*gln*A	1431	247 bp – 724 bp
*glt*A	1269	361 bp – 763 bp
*gly*A	1245	355 bp – 862 bp
*glm*M	1338	649 bp – 1147 bp
tkt	1899	226 bp – 685 bp
*unc*A/*atp*A	1506	676 bp – 1165 bp

bp:base pair.

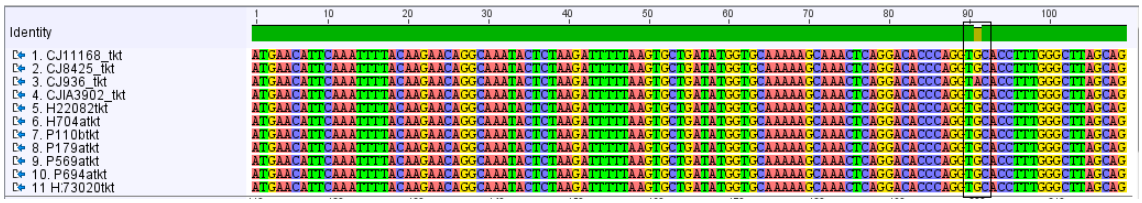


(a) *asp*A gene sequences

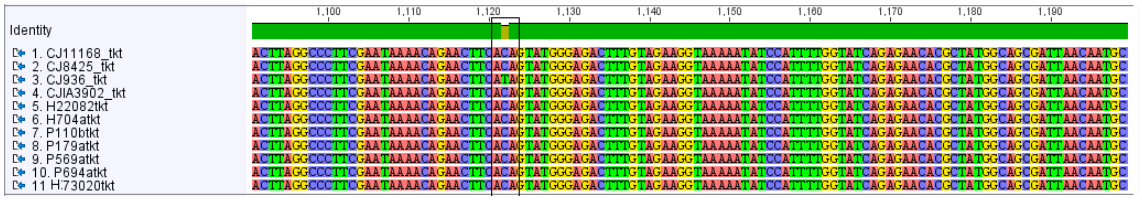


(b) *asp*A amino acid sequences

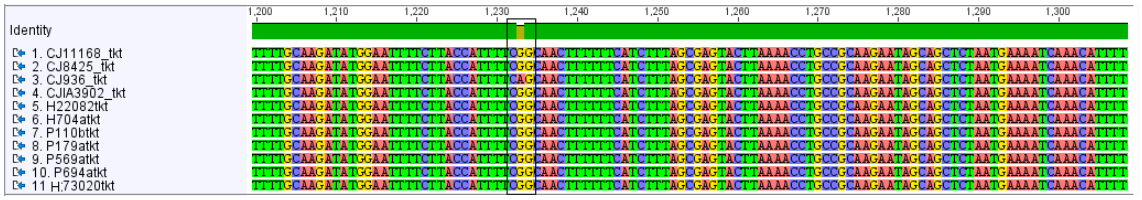
Figure 5.2: Comparison of nucleotide and amino acid sequences of *asp*A gene between ST-474 and reference genomes. The alignment of full length gene sequences of *asp*A from four reference genomes: CJJ84-25, CJJCF936, CJJIA3902, and CJJ11168 and ST-474 genomes shows a non-synonymous substitution in the CJJ84-25 genome



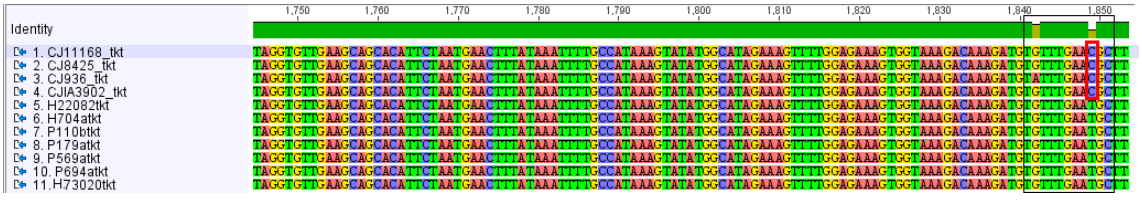
(a) *tkt* gene sequences-region one



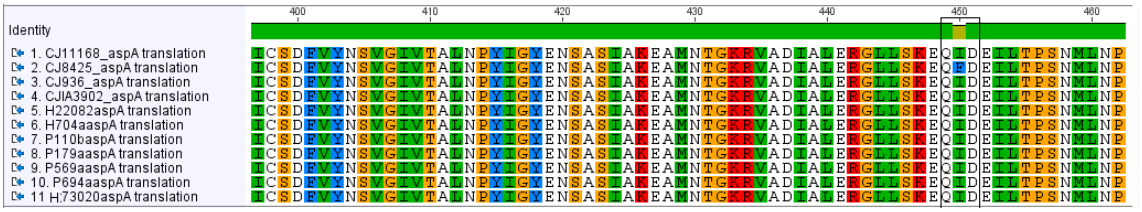
(b) *tkt* gene sequences-region two



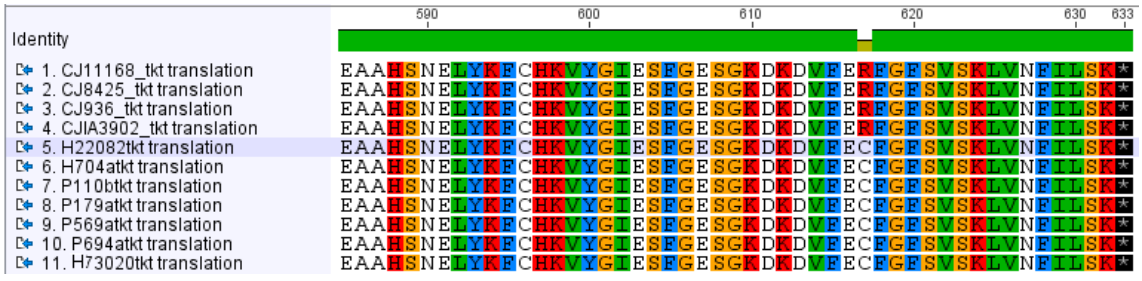
(c) *tkt* gene sequences-region three



(d) *tkt* gene sequences-region four



(e) *tkt* amino acid sequences-region one



(f) *tkt* amino acid sequences-region two

Figure 5.3: Comparison of nucleotide and amino acid sequences of *asp*A gene between ST-474 and reference genomes. The alignment of full length gene sequences of *tkt* from four reference genomes: CJJ84-25, CJJCF936, CJJIA3902, and CJJ11168 and ST-474 genomes that shows synonymous and non-synonymous substitutions

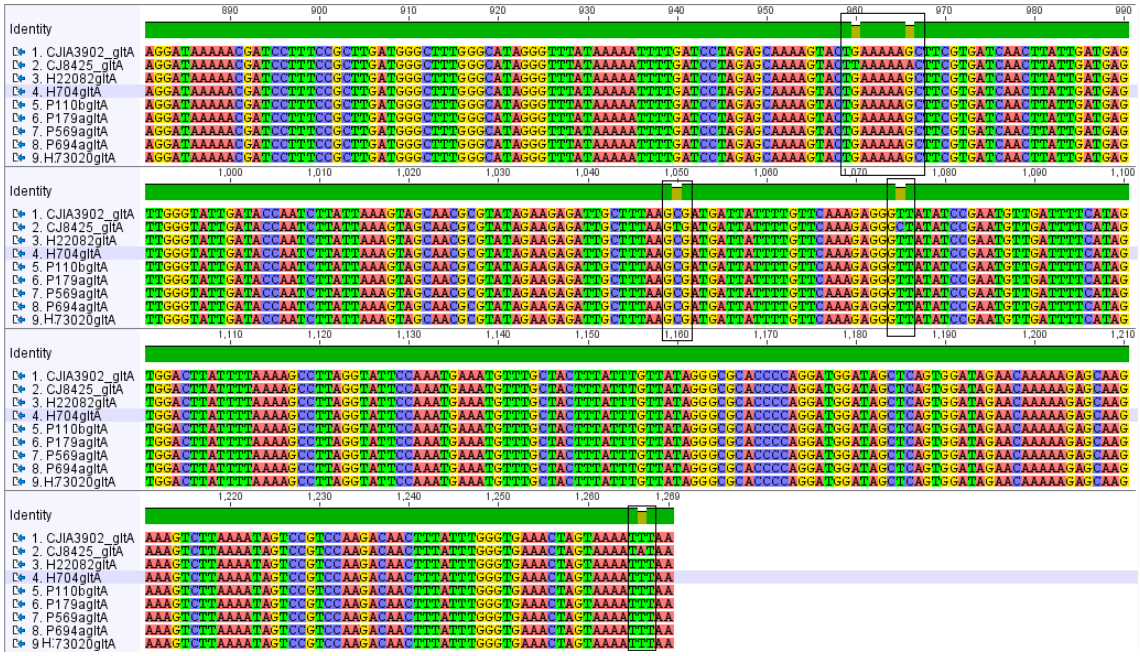


Figure 5.4: Alignment of *glt*A nucleotide sequences. Aligned *glt*A gene sequences from ST-474, CJJIA3902 and CJJ84-25 genomes show multiple synonymous nucleotide substitutions within CJJ84-25 genome

Analysis of Guanine-Cytosine content, codon usage, selection pressure, evolutionary rate and recombination:

Guanine-Cytosine content

Guanine-Cytosine (GC) content ranged between 31.9% and 36.4% across MLST housekeeping genes, where both the high and low ranges of the GC contents were evident in *C. jejuni* subsp. *doylei* (CJJD269.97269.97). The GC contents of individual MLST housekeeping genes across all genomes (12 reference and seven ST-474 genomes) is presented in Table C.2 in Appendix C. The GC content distribution within the seven MLST housekeeping genes across all the *C. jejuni* genomes (the reference and ST-474 genomes investigated in this study) are presented in Figure 5.5. The GC content of the housekeeping alleles varied among genomes with the *tkt* and *glt*A alleles showing a relatively wider variation followed by *gly*A, *glm*M, *asp*A, *gln*A and *unc*A alleles (Table C.2).

Codon usage bias

The overall codon usage bias index (CBI) estimates of CJJIA3902, CJJ84-25 and CJJCF93-6 and ST-474 genomes were shown to be identical between three MLST genes followed by CJJ11168 with two genes showing identical CBI estimates. When comparing the scaled chi square CBI for the seven MLST housekeeping genes between the ST-474 and the reference genomes, CJJ84-25 was found to be closer, sharing identical scaled chi square estimates for four MLST housekeeping genes followed by CJJIA3902, CJJ11168 and CJJCF93-6 (Table 5.14). However, when nucleotide sequences were compared, there were 18 mismatches found within the *gln*A gene sequences between the ST-474 and CJJ84-25 genomes while there were only 11 mismatches between the CJJCF93-6 and ST-474 genomes. It is interesting to note that the codon preferences remain similar in spite of several synonymous nucleotide mismatches between the CJJ84-25 and ST-474 MLST genes. Further details on the CBI and scaled chi square codon bias indices for individual housekeeping genes are provided in Appendix C, Tables C.5 and C.4.

Figure 5.5: Guanine-Cytosine variation in the MLST genes of *C. jejuni* genomes. The values in the parentheses denotes the estimated number of recombination sites.

Selection pressure and evolutionary rate

MLST housekeeping genes were found to be under stringent purifying selection that showed a negative Tajima D value for individual full length genes (Table 5.3.1). Tajima D (TD) values of non-synonymous over synonymous substitution (NonSyn/Syn) ratios were greater than 1 for *asp*A, *gly*A, *tkt* and *unc*A which indicated that, although the genes as a whole are under negative selection pressure, these sites are under positive selection pressure. A negative TD value for the NonSyn/Syn substitution was obtained for *gln*A that indicated an ongoing purifying selection (Tajima 1989, Hughes et al. 2008). Similarly, the ω values (tested for individual codons) for the test statistic d_N-d_S showed negative values for a few codons within *gln*A, *glt*A, *gly*A, *glm*M, *tkt* and *unc*A genes, indicative of an ongoing purifying selection over these codons.

The *tkt* gene showed one codon under positive selection pressure while four non-synonymous codons were found to be under an ongoing purifying selection pressure (Table 5.10). This created a level of uncertainity of purifying selection pressure and/or a positive selection prevailing within these codons. Mean relative evolutionary rates of individual sites were calculated using the Tamura & Nei (1993) model which showed that the polymorphic sites identified in each gene (synonymous and non-synonymous) evolve faster than non-polymorphic sites, while the non-polymorphic sites showed a mean relative evolutionary rate of 0.5. The mean evolutionary rate for these sites was greater than 1, ranging from 3.2 to 5.19 (Table 5.9).

Table 5.8: Tajima D values for the MLST housekeeping genes

Genes	Syn	Nonsyn	TD_gene	TD_nonsyn/syn
*asp*A	41	6	-0.59233	1.94469
*gln*A	40	7	-0.14032	**-2.58364**
*glt*A	46	6	-0.8	0.55457
*gly*A	51	12	-0.15614	13.2048
*glm*M	59	11	-0.66672	0.35281
tkt	83	24	-0.79287	1.30603
*unc*A/*atp*A	42	5	-0.79287	1.30603

Syn:Synonymous; Nonsyn:Non-synonymous; TD_gene:Tajima's D for the gene; TD_nonsyn/syn:Tajima's D for the Non-synonymous/synonymous ratio

Table 5.9: Average of mean relative evolutionary rates of non-synonymous codons in the MLST housekeeping genes

Genes	Evol_Rate_codons
*asp*A	5.023887
*gln*A	5.12454
*glt*A	5.192522
*gly*A	5.060395
*glm*M	5.10823
tkt	5.169835
*unc*A	3.230612

Evol_Rate_codons:Average of mean relative evolutionary rates of non-synonymous codons

Table 5.10: Number of non-synonymous codons in MLST genes under purifying selection based on ω values

Genes	Nonsyn_codons	NS_codons_purifying selection
*asp*A	6	-
*gln*A	7	2
*glt*A	6	1
*gly*A	12	2
*glm*M	11	2
tkt	24	4
*unc*A/*atp*A	5	1

Nonsyn_codons:Non-synonymous codons; NS_codons_purifying selection:Non-synonymous sites under purifying selection

Recombination and phylogenetic analysis

Recombination events (R_m [the minimum number of recombination sites]) and sites involved in recombination within MLST housekeeping genes are presented in Table 5.11. The number of recombination sites was highest in the *tkt* gene (n = 17) followed by *glt*A (n = 10), *gln*A (n = 9), *glm*M (n = 9), *gly*A (n = 7), *asp*A (n = 5) and *unc*A (n = 1). There was a positive correlation between the GC variance and the number of recombination sites, where the number of recombination sites were found to be higher in genes with high GC variance between genomes. The relationship is discussed in section 5.3.2. *tkt* was the longest gene among the seven MLST genes that showed the highest number of recombination sites where gene lengths could be a contributing factor for the observed differences in the number of recombination sites.

However, it should be remembered that only a limited number of genomes were examined to compare the above mentioned parameters (GC, CBI, selection pressure and recombi-

nation) in this study. Comparative analyses involving larger datasets of gene sequences and/or using MLST datasets by combining different species of *Campylobacter* may provide varied results with respect to the above mentioned parameters between different species of *Campylobacter*. A Maximum likelihood phylogenetic tree of the concatenated nucleotide sequences of MLST housekeeping genes (between ST-474 and CJJ84-25, CJJCF93-6, CJJIA3902 and CJJ11168 genomes) showed that CJJCF93-6 to be the closest ancestor of ST-474 (Figure 5.6).

Table 5.11: Recombination sites identified in the full length MLST genes of 19 genomes

Genes	Sites	Rm	GC range
*asp*A	(198-336) (336-414) (414-615) (627-897) (1203-1251)	5	0.331 – 0.333
*gln*A	(180-230) (231-258) (258-279) (358-448) (448-513) (711-768) (768-945) (945-1188) (1188-1260)	9	0.331 – 0.335
*glt*A	(312-325) (325-351) (351-372) (372-453) (453-654) (654-756) (780-921) (921-1089) (1089-1140) (1203-1305)	10	0.352 – 0.364
*gly*A	(165-219) (474-490) (492-562) (612-618) (618-657) (684-744) (966-1156)	7	0.329 – 0.336
*glm*M	(171-291) (291-366) (366-450) (492-573) (573-636) (689-798) (897-972) (972-1020) (1119-1286)	9	0.321 – 0.325
tkt	(54-102) (159-237) (237-297) (297-342) (342-366) (366-660) (966-1029) (1092-1107) (1125-1138) (1138-1254) (1254-1287) (1320-1486) (1525-1749) (1749-1761) (1761-1791) (1791-1827) (1827-1876)	17	0.357 – 0.361
*unc*A	(678-1065)	1	0.357 – 0.361

Rm:Number of recombination sites

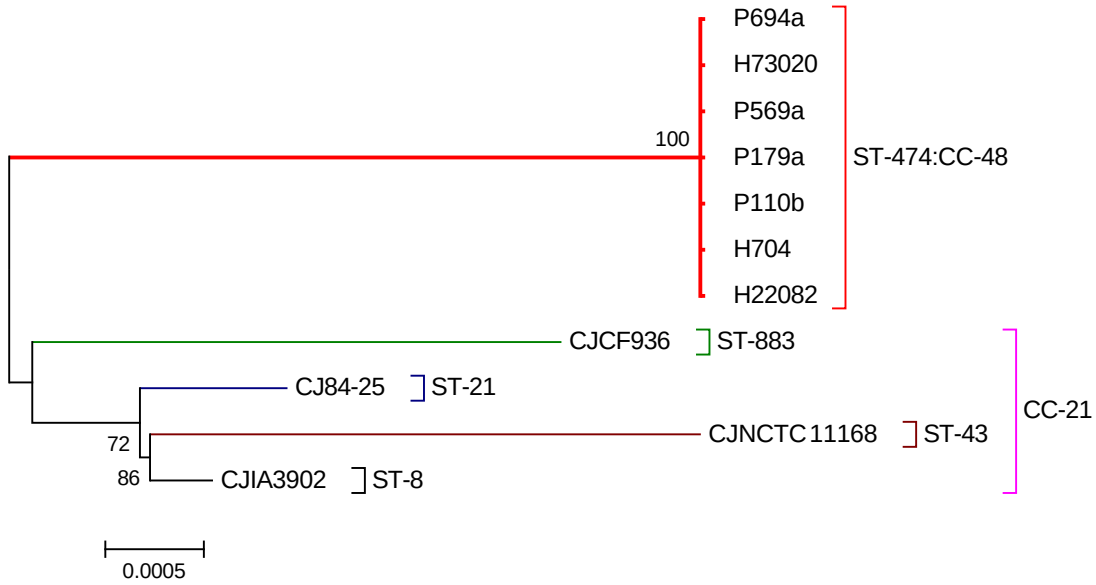


Figure 5.6: Phylogenetic tree of concatenated MLST housekeeping genes. Maximum likelihood tree of concatenated MLST full length gene sequences of the four reference and ST-474 genomes constructed in MEGA v5. The numbers refer to the bootstrap values corresponding to each branch. ST and CC refer to sequence type and clonal complex, respectively.

5.3.2 Analysis of second subset of housekeeping genes:

Analysis of the second subset of metabolic housekeeping genes was carried out following a similar pattern of analysis as that of the first subset of genes. Gene lengths (genes; n = 18) were identical across all ST-474 genomes whereas the lengths varied between genes among reference genomes (Table 5.12). The GC contents varied between genes and genomes. Figures 5.7, 5.8 and 5.9 illustrate the variation in the GC contents across all the *C. jejuni* genomes compared, and the number of recombination sites that had occurred within those genes. Further details on GC and GC3 contents of individual genes for all the genomes used in this study are provided in Tables C.2 and C.3 in Appendix C. GC contents were identical amongst all genes across all the ST-474 genomes except for two genes, namely *fum*C and *trp*C. CJJ84-25 and CJJCF93-6 were found to be closer to ST-474 by sharing their GC contents with 12 of 25 genes (MLST genes inclusive) followed by CJJ11168 (n = 10), CJJIA3902 (n = 10), CJJ81-176 (n = 8) and others. Whereas, CJJ11168 and CJJIA3902 shared identical GC3 contents with 13 genes of ST-474 followed by CJJ84-25 and CJJCF93-6 with 11 genes and others for which the information is provided in Table 5.13.

CBI was estimated for all 25 genes and further details are presented in Appendix C. CJJIA3902 was found to be closer to the ST-474 genomes with respect to CBI followed by CJJCF93-6, CJJ84-25 and CJJ11168 (Table 5.14). It was interesting to note that the GC3 and CBI showed an agreement in predicting the ancestry shared between the reference and ST-474 genomes, while it should be noted that the nucleotide composition of a gene is responsible for codon preferences in bacteria (Snyder & Champness 1997).

Table 5.12: Second subset of genes and their lengths (base pairs long)

Genomes	Genes and their lengths (base pairs long)																	
	*arg*F	*aro*E	*atp*D	*dap*E	*fts*Z	*fum*C	*gap*A	*glt*B	*gro*EL	*hem*N	*ilv*D	*inf*B	*lys*A	*nuo*D	*pyc*A	*sdh*A	*trp*B	*trp*C
CJNCTC11168	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	1446	1836	1179	777
CJJ260.94	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	**2583**	1209	1227	1446	1836	1179	777
CJJ81116	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	1446	1836	1179	777
CJJ81-176	921	789	1398	1098	1113	1392	999	**4488**	1638	1356	1677	2616	1209	1227	1446	1836	1179	777
CJJCG84-21	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	**1398**	1836	1179	777
CJJ84-25	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	1446	1836	1179	777
CJJCG84-86	921	**612**	1398	1098	1113	1392	999	4491	1638	1356	**1479**	**1668**	1209	1227	1446	**966**	1179	**789**
CJJHB93-13	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	1446	1836	1179	777
CJJCF93-6	**468**	789	1398	1098	1113	1392	999	4491	**910**	1356	1677	2616	1209	1227	1446	1836	1179	777
CJD269.97	921	789	**1455**	1098	1113	1392	**993**	**4488**	1638	1356	1677	2565	1209	1227	1446	**1992**	1179	777
CJJIA3902	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	**2628**	1209	1227	1446	1836	1179	777
CJJRM1221	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	1446	**1731**	1179	777
H22082	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	1446	1836	1179	777
H704	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	1446	1836	1179	777
P110b	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	1446	1836	1179	777
P179a	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	1446	1836	1179	777
P569a	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	1446	1836	1179	777
P694a	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	1446	1836	1179	777
H73020	921	789	1398	1098	1113	1392	999	4491	1638	1356	1677	2616	1209	1227	1446	1836	1179	777

Bold fonts refer to genes with different lengths

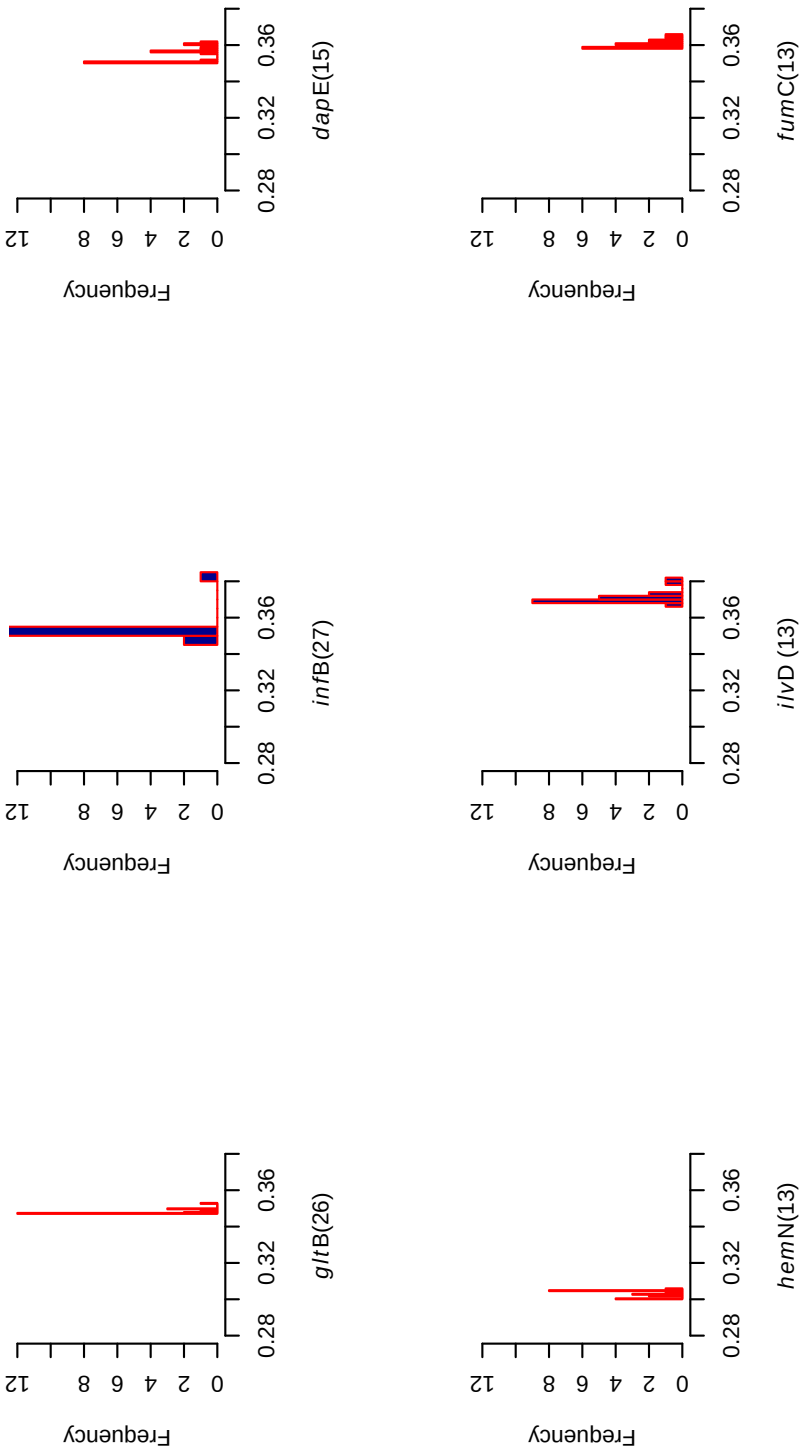


Figure 5.7: Guanine-cytosine (GC) variation in the metabolic housekeeping genes – A. The plots describe the distribution of GC across 19 *C. jejuni* genomes and the number inside the parenthesis denotes the number of recombination sites

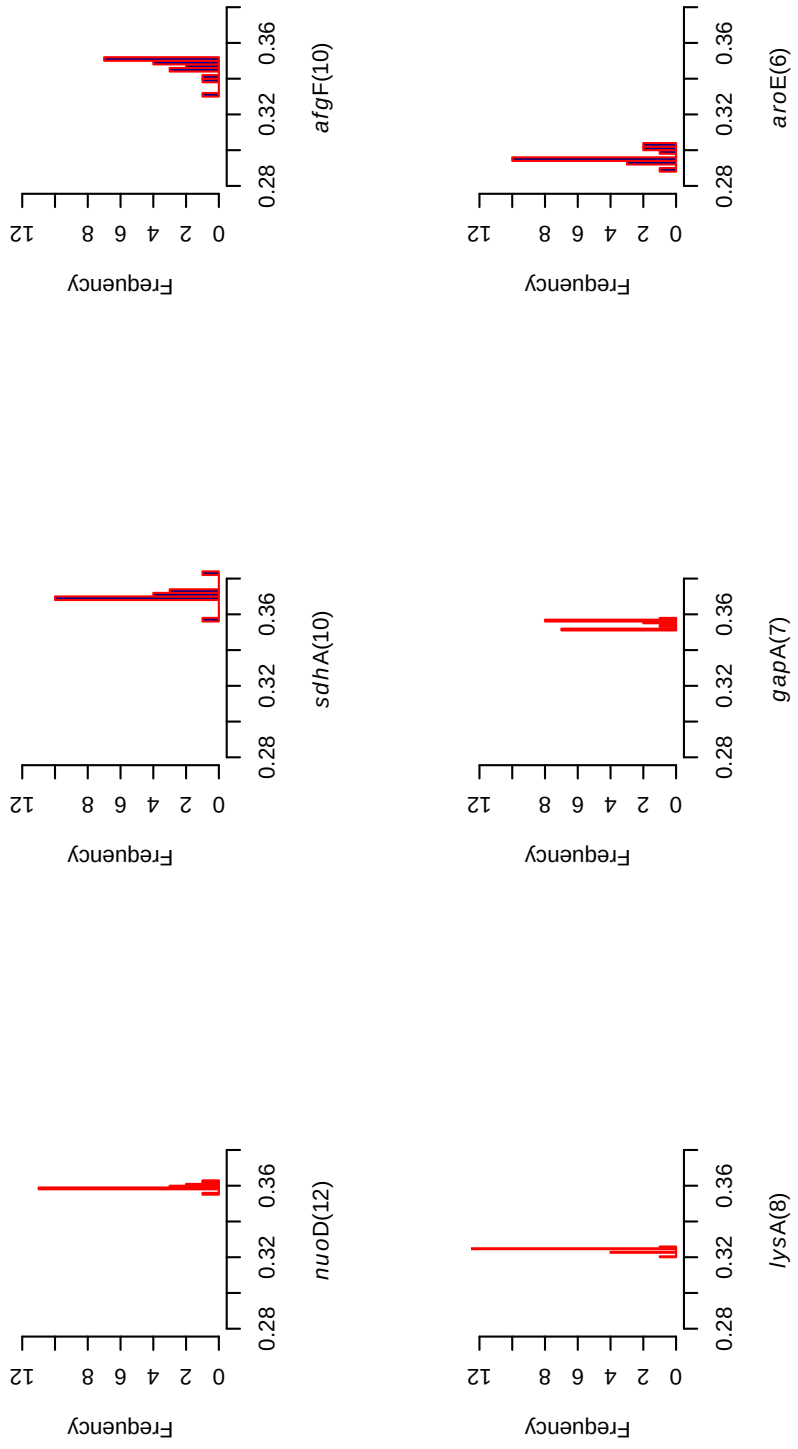


Figure 5.8: Guanine-cytosine (GC) variation in the metabolic housekeeping genes – B. The plots describe the distribution of GC across 19 *C. jejuni* genomes and the number inside the parenthesis denotes the number of recombination sites

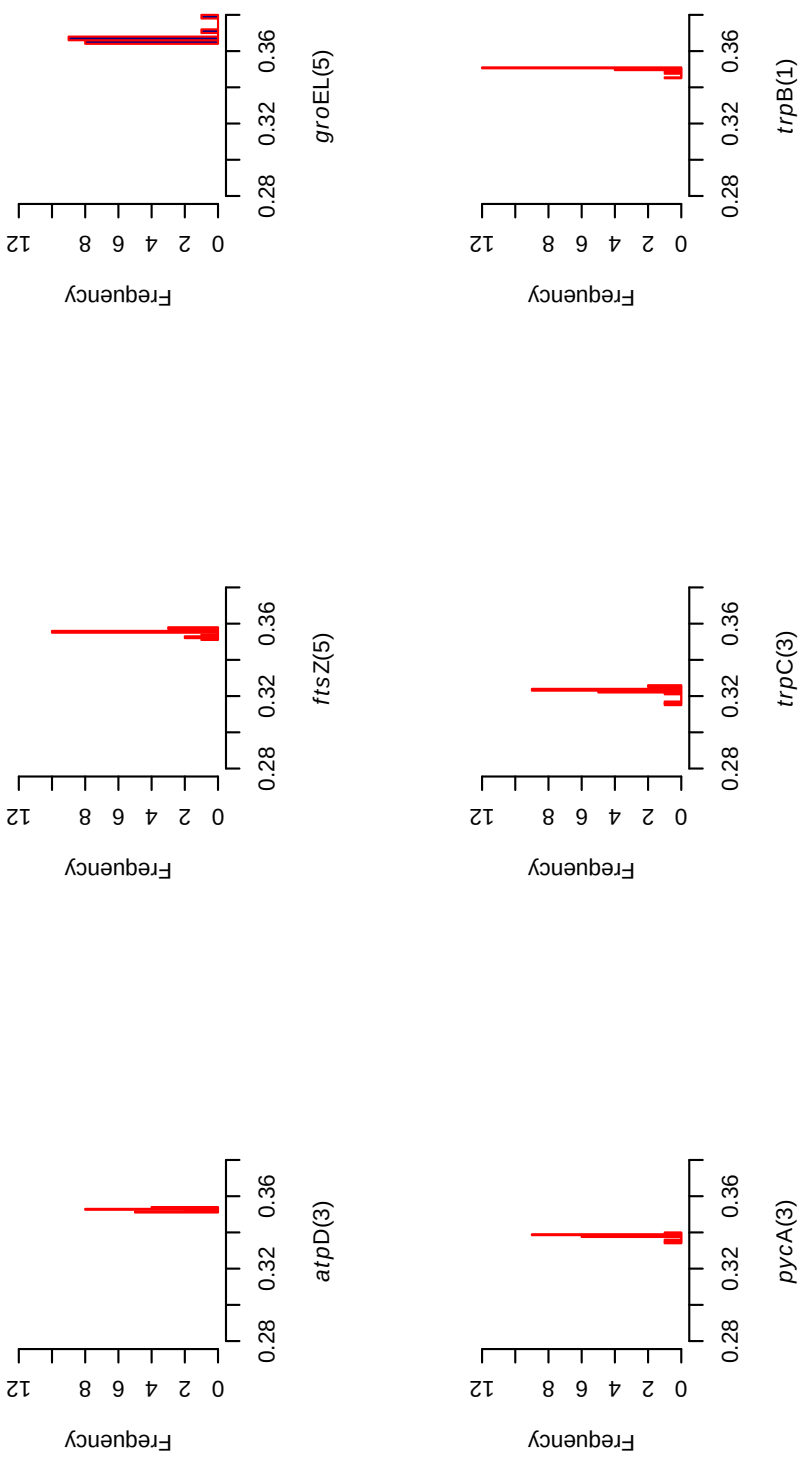


Figure 5.9: Guanine-cytosine (GC) variation in the metabolic housekeeping genes – C. The plots describe the distribution of GC across 19 *C. jejuni* genomes and the number inside the parenthesis denotes the number of recombination sites

Table 5.13: Ancestry shared between reference and ST-474 genomes based on GC and GC3 contents

	CJJ11168 (ST-43:CC-21)	CJJ260.94 (ST-362:CC-362)	CJ81116 (ST-267:CC-283)	CJJ81-176 (ST-604:CC-42)	CJJCG84-21 (ST-1919:CC-52)	CJJ84-25 (ST-21:CC-21)	CJJCG84-86 (ST-2943:CC-574)	CJJHB93-13 (ST-22:CC-22)	CJJCF93-6 (ST-883:CC-21)	CJD269.97 (ST-1845:CC-U/A)	CJJIA3902 (ST-8:CC-21)	CJJRM1221 (ST-354:CC-354)
Measures	n	n	n	n	n	n	n	n	n	n	n	n
GC	10	4	3	8	3	12	1	3	12	2	10	4
GC3	13	1	1	6	1	11	4	1	11	2	13	5

GC:Guanine and cytosine
n:Number of genes from ST-474 genomes shared identical estimates of GC and GC3 contents with the reference genomes.
This table includes the GC and GC3 contents of the MLST housekeeping genes for comparison.
ST:Sequence type
CC:Clonal complex
U/A:Unassigned

Table 5.14: Ancestry shared with respect to codon usage between the reference and ST-474 genomes

Measures	CJJ11168 (ST-43:CC-21)	CJJ260.94 (ST-362:CC-362)	CJ81116 (ST-267:CC-283)	CJJ81-176 (ST-604:CC-42)	CJJCG84-21 (ST-1919:CC-52)	CJJ84-25 (ST-21:CC-21)	CJJCG84-86 (ST-2943:CC-574)	CJJHB93-13 (ST-22:CC-22)	CJJCF93-6 (ST-883:CC-21)	CJD269.97 (ST-1845:CC-U/A)	CJJIA3902 (ST-8:CC-21)	CJJRM1221 (ST-354:CC-354)
	n	n	n	n	n	n	n	n	n	n	n	n
2^{nd} subset_CBI	6	0	2	5	2	6	3	1	9	0	10	1
MLST_CBI	2	0	0	0	1	3	1	0	3	0	3	1
2^{nd} subset_Scaled	6	1	2	3	0	8	3	0	8	2	7	3
MLST_Scaled	3	0	1	0	0	4	2	0	3	1	3	2

CBI:Codon usage bias index
Scaled:Scaled chi square index
ST:Sequence type
CC:Clonal complex
U/A:Unassigned

Analysis of selection pressures on individual codons provided further interesting insights into the nucleotide sites that could have occurred within the second subset of genes. The test statistics ω for synonymous substitutions within codons showed negative values across all genes, whereas few codons in two of the genes namely, *glt*B and *atp*D showed positive values below one. A considerable number of non-synonymous codons in few of the genes were found with negative ω values while others have shown values greater than one. The observation of negative values for non-synonymous codons is indicative of overabundance of non-synonymous substitutions in those codons leading to an amino acid change. This in turn indicates that there has been an ongoing purifying selection pressure operating on these codons. ω values greater than one indicate the positive selection pressure on these codons. Table 5.15 provides details on the number of non-synonymous codons within the second subset of housekeeping genes that obtained a negative ω value. The overall Tajima's D value for all the 18 full genes showed a negative value which means that, although some individual codons are under positive selection the genes as a whole are under purifying selection (Table 5.16).

Table 5.15: Number of non-synonymous codons in the second subset of genes under purifying selection based on ω values

Genes	Nonsyn_codons	NS_codons_purifying selection
*arg*F	7	3
*aro*E	12	-
*atp*D	5	-
*dap*E	31	7
*fts*Z	1	-
*fum*C	13	2
*gap*A	12	1
*glt*B	22	2
*gro*EL	2	-
*hem*N	25	4
*ilv*D	18	4
*inf*B	11	2
*lys*A	14	4
*nuo*D	8	1
*pyc*A	9	3
*sdh*A	11	-
*trp*B	24	3
*trp*C	6	-

Syn_sites:Synonymous sites
Nonsyn_sites:Non-synonymous sites
NS_codons_purifying selection:Non-synonymous that are under purifying selection

Table 5.16: Tajima's D values for the second subset of housekeeping genes

Genes	Syn	Nonsyn	TD_gene	TD_nonsyn/syn
*arg*F	44	7	-0.96811	1.23974
*aro*E	20	12	-1.43911	1.11983
*atp*D	34	5	-1.43984	1.63866
*dap*E	86	31	-1.28806	1.65624
*fts*Z	32	1	-0.24838	6.57937
*fum*C	68	13	-0.99988	0.82645
*gap*A	30	12	-0.38301	**-5.77967**
*glt*B	196	22	-1.47845	1.17223
*gro*EL	48	2	-1.16813	0.41075
*hem*N	61	25	-0.17368	0.42345
*ilv*D	79	18	-0.45085	2.19417
*inf*B	108	11	-0.84724	1.961
*lys*A	45	14	-0.65465	**-0.0504**
*nuo*D	65	8	-0.96403	0.35281
*pyc*A	71	9	-1.99257	0.93794
*sdh*A	30	11	-0.82915	**-0.02599**
*trp*B	48	14	-2.25126	0.99273
*trp*C	23	6	-0.97295	2.12579

Syn:Synonymous
Nonsyn:Non-synonymous
TD_gene:Tajima's D for the gene
TD_nonsyn/syn:Tajima's D for the Non-synonymous/synonymous ratio

The *hem*N, *ilv*D and *lys*A genes showed four codons to be under purifying selection followed by *arg*F, *pyc*A and *trp*B with three codons; *fum*C, *glt*B and *inf*B with two codons; and *gap*A and *nuo*D with one codon. The *dap*E gene showed seven non-synonymous codons to be under purifying selection (86 synonymous substitutions and 31 non-synonymous) with an average relative mean evolutionary rate of 5.373 (Table 5.17). The highest number of synonymous substitutions was found in *glt*B (n = 196) with *inf*B showing the second highest number of synonymous substitutions (n = 108) amongst all the genes investigated in the study. The *inf*B gene showed the highest mean evolutionary rate of 5.423 and the *trp*B gene showed the least mean relative evolutionary rate of 1.267 amongst the genes that were found to evolve faster than average (Table 5.17). It should be noted that the lengths between the investigated genes may confound the differences in the evolutionary rates. However, in general, genes with mean evolutionary rates greater than one can be regarded to evolve faster than average.

Table 5.17: Average of mean relative evolutionary rates of non-synonymous codons in the second subset of genes

Genes	Evol_Rate_codons
*arg*F	4.93627
*aro*E	4.944461
*atp*D	4.977999
*dap*E	5.373023
*fts*Z	5.075987
*fum*C	5.087526
*gap*A	5.027066
*glt*B	5.271881
*gro*EL	5.001216
*hem*N	5.237627
*ilv*D	5.045032
*inf*B	5.429229
*lys*A	5.18066
*nuo*D	5.297639
*pyc*A	4.572914
*sdh*A	4.99454
*trp*B	1.267403
*trp*C	5.06902

Evol_Rate_codons:Average of relative mean evolutionary rate of non-synonymous codons

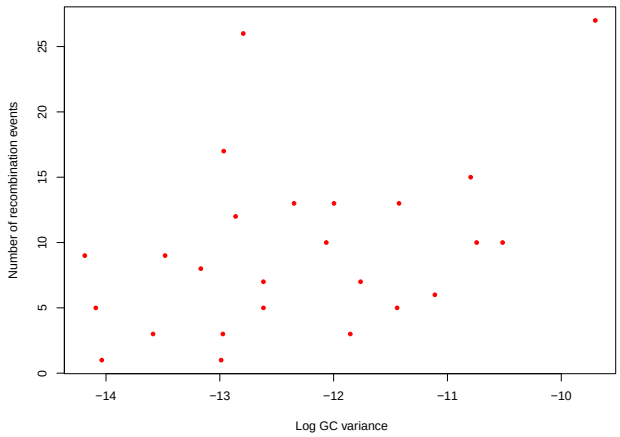
Recombination and its relationship with GC variance

An event of recombination was evident in all metabolic housekeeping genes where, *inf*B showed the highest number of 27 sites (Rm [the minimum number of recombination sites] = 27) (Table 5.18) and *trp*B and *unc*A showed the least number of sites (n = 1) to be involved in recombination. It was found that the number of recombination sites was positively correlated with the GC variance, where the genes that showed a wider GC variance showed a high number of recombination sites. However, the *inf*B and *glt*B genes showed the highest numbers (n = 27 and n = 26, respectively) of recombination sites while their GC contents did not vary widely among genomes as was observed in other genes. While the GC variance and the number of recombination sites were found to be positively correlated (p value = 0.009), the length was not found to influence the number of recombination sites significantly (p value = 0.7). Figures 5.10 a and b illustrate the relationship between the GC variance and the number of recombination sites and the relationship between the length and recombination sites, respectively.

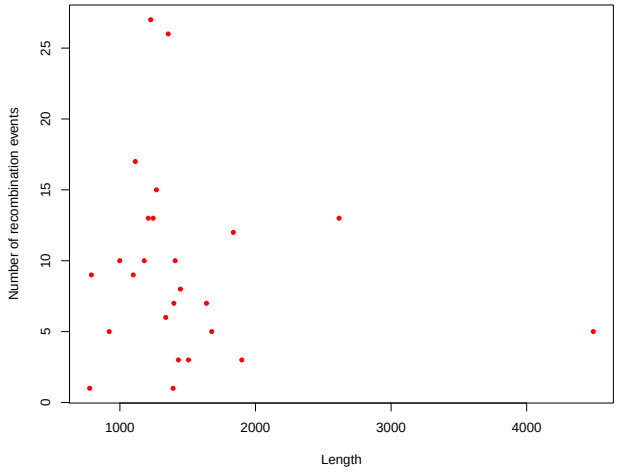
Table 5.18: Recombination sites identified in the second subset of genes

Genes	Sites	Rm	GC range
*arg*F	(114-163) (168-195) (195-207) (207-231) (231-282) (282-291) (291-303) (303-308) (308-339) (420-465)	10	0.341 – 0.351
*aro*E	(381-392) (392-435) (514-645) (645-654) (654-702) (702-717)	6	0.288 – 0.303
*atp*D	(348-408) (408-651) (924-1047)	3	0.351 – 0.354
*dap*E	(171-216) (216-230) (231-321) (321-351) (351-357) (357-387) (387-453) (474-522) (531-543) (543-560) (609-690) (701-867) (867-942) (942-945) (990-1035)	15	0.352 – 0.362
*fts*Z	(90-138) (348-561) (645-681) (840-903) (903-936)	5	0.351 – 0.358
*fum*C	(45-108) (168- 303) (354- 357) (576- 591) (624-687) (687- 711) (772-909) (930-987) (987-1038) (1038-1227) (1281-1311) (1311-1335) (1335-1380)	13	0.359 – 0.364
*gap*A	(288-381) (381-501) (501-567) (570-594) (594-732) (819-921) (948-961)	7	0.351 – 0.358
*glt*B	(159-675) (741-855) (1236-1401) (1401-1581) (1581-1605) (1605-1902) (1974-2067) (2151-2178) (2226-2325) (2349-2412) (2412-2418) (2418-2700) (2700-2838) (2853-2982) (3204-3210) (3210-3318) (3387-3462) (3462-3636) (3651-3861) (3885-4023) (4023-4092)(4092-4098) (4098-4119) (4119-4149) (4149-4305) (4324-4356)	26	0.347 – 0.353
*gro*EL	(1008-1023) (1023-1029) (1029-1035) (1200-1236) (1285-1290)	5	0.364 – 0.379
*hem*N	(82-174) (174-207) (210-372) (544-584) (585-783) (828- 903) (903-927) (930-975) (975-996) (996-1050) (1050-1065) (1104-1140) (1140-1163)	13	0.3 – 0.306
*ilv*D	(249- 300) (300- 399) (492- 576) (654-777) (777-816) (864-891) (1098-1137) (1170-1263) (1263-1296) (1323-1347) (1389-1416) (1416-1530) (1530-1665)	13	0.37 – 0.381
*inf*B	(1128-1140) (1140-1188) (1230-1263) (1263-1284) (1284-1365) (1365-1392) (1392-1416) (1476-1551) (1596-1692) (1743-1824) (1824-1845) (1845-1914) (1914-1938) (1938-1962) (1962-2124) (2124-2127) (2127-2235) (2235-2259) (2268-2286) (2286-2301) (2319-2376) (2376-2394) (2394-2436) (2436-2478) (2478-2487) (2487-2490) (2538-2562)	27	0.35 – 0.385
*lys*A	(198-222) (222-672) (706-790) (790-825) (825-944) (945-1035) (1035-1122) (1122-1171)	8	0.323 – 0.326
*nuo*D	(78-93) (105-249) (249-267) (393-453) (453-483) (543-566) (660-678) (780-852) (852-864) (875-894) (894-954) (1008-1050)	12	0.355 – 0.363
*pyc*A	(78-194) (194-546) (672-768)	3	0.338 – 0.339
*sdh*A	(684-686) (686-693) (693-700) (700-701) (775-787) (787-848) (1051-1066) (1066-1141) (1219-1291) (1291-1329)	10	0.37 – 0.384
*trp*B	(1053-1119)	1	0.35 – 0.0351
*trp*C	(138-195) (210-294) (567-750)	3	0.315 – 0.326

Rm:Number of recombination sites



(a) Relationship between the GC variance and the number of recombination sites



(b) Relationship between the length of genes and the number of recombination sites

Figure 5.10: Relationship between the GC variance, length and the recombination sites in the metabolic genes. This plot describes the relationship between the GC variance within individual genes and the number of recombination sites that had occurred within each gene and the relationship between the length of genes and the number of recombination sites.

Phylogenetic trees of individual genes were constructed using three methods, namely, maximum likelihood (MEGA v5), Neighbournet (SplitsTree v4) and neighbour joining methods (Geneious v5.3.4). The congruence and topographies between these three methods were found to be in agreement with each other and hence the maximum likelihood trees to represent the phylogenetic relationship of individual genes are presented in this study. Individual gene trees of 18 genes (MLST genes not included) are presented in Figures C.2, C.3 and C.4 in Appendix C while genes with salient genetic characteristics

with respect to ST-474 genomes such as *fts*Z, *fum*C, *hem*N, *sdh*A and *pyc*A are presented in Figure 5.11. While investigating the ancestry of ST-474 based on the individual gene trees, CJJ84-25 was found to be closer to ST-474 in 9 of the 18 gene trees followed by CJJ11168 (n = 9), CJJCF93-6 (n = 9) and CJJIA3902 (n = 9). Maximum likelihood tree of the concatenated sequences constructed from all the 19 *C. jejuni* genomes showed CJJIA3902 to be the closest ancestor of ST-474 genomes while two of the ST-474 isolates, P694a and H22082 were found to be placed in a separate branch from the remainder, Figure 5.12.

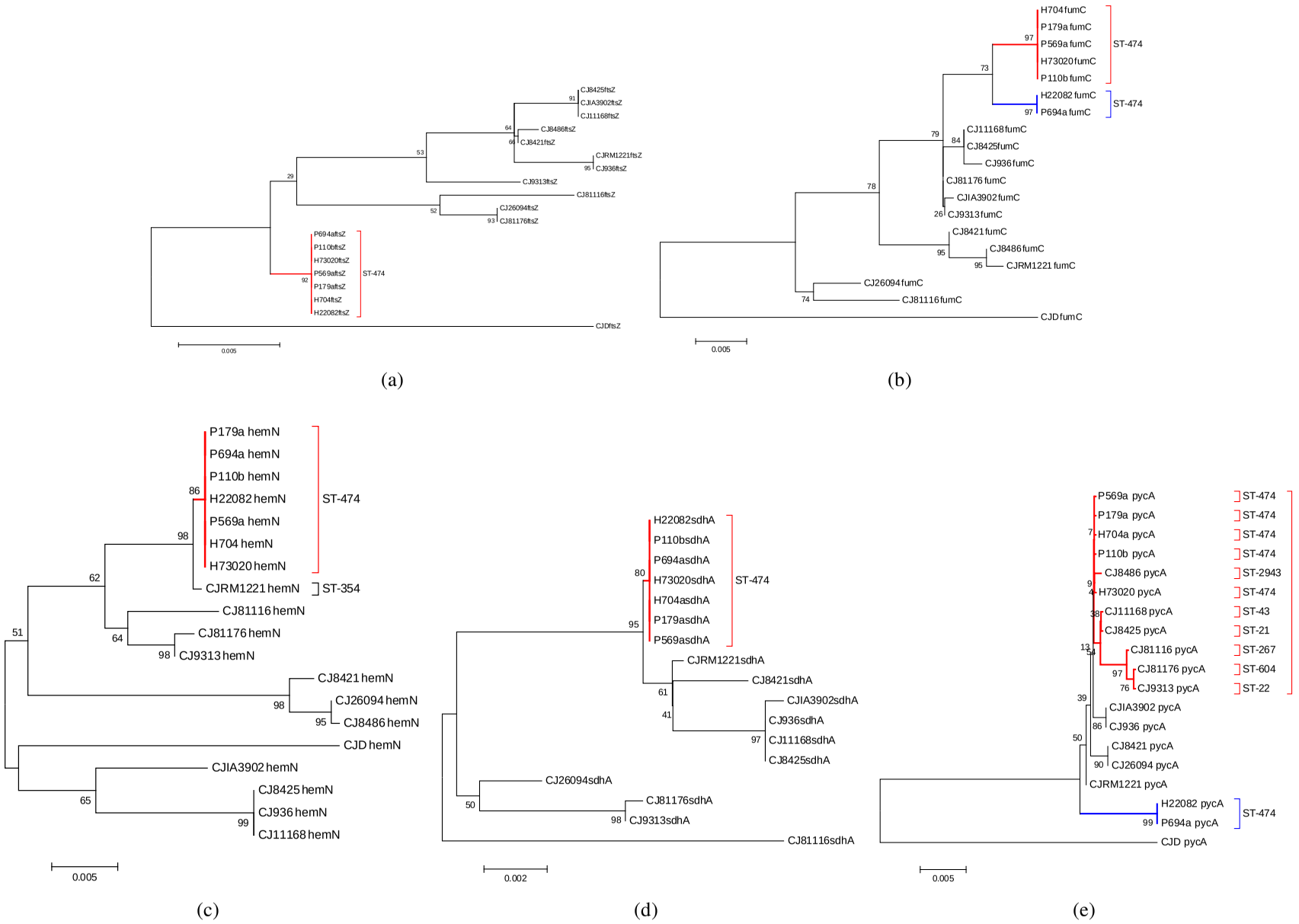


Figure 5.11: Maximum likelihood gene trees of individual metabolic genes. Gene trees of *fts*Z, *fum*C, *pyc*A, *sdh*A, and *pyc*A constructed using MEGA v5, bootstrapped to 500 times. ST refers to sequence type.

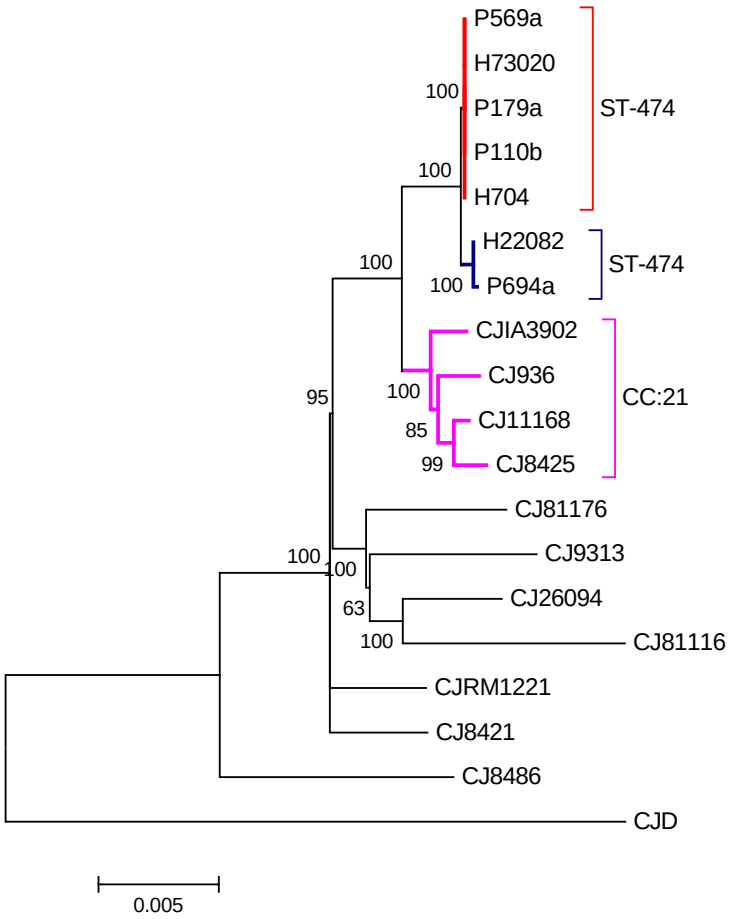


Figure 5.12: Maximum likelihood tree of concatenated metabolic genes. All 25 metabolic genes investigated in this study were concatenated and a maximum likelihood tree was constructed using MEGA v5. The numbers refer to bootstrap values corresponding to each branch. ST refers to sequence type and CC refers to clonal complex.

Comparison of clonal complex versus full length genes in the prediction of a common ancestor:

When the full length MLST genes were compared, there were nucleotide substitutions, both synonymous and non-synonymous among the MLST identical alleles as pointed out earlier in the results. The ancestry predicted by the clonal concept based on the seven alleles was found to have a minor shift from CJJ84-25 to CJJCF93-6 when full length genes were used for analysis with respect to seven MLST genes. There were distinct partitions between *C. jejuni* genomes compared in this study when a concatenated phylogenetic tree was constructed. The most striking finding was that the phylogenetic branching pattern was in accordance with the clonal relationship observed between the genomes. The ST-21 complex that comprised of CJJ11168, CJJIA3902, CJJ84-25 and CJJCF93-6 was the closest clonal complex to the ST-474 genomes (ST-48 complex) while CJJ84-25 and CJJ11168 were found to be the closest ancestors of the ST-474 genomes (Table 5.19).

Table 5.19: Ancestry shared between the reference and ST-474 genomes according to individual gene trees

	CJJ11168 (ST-43:CC-21)	CJJ260.94 (ST-362:CC-362)	CJ81116 (ST-267:CC-283)	CJJ81-176 (ST-604:CC-42)	CJJCG84-21 (ST-1919:CC-52)	CJJ84-25 (ST-21:CC-21)	CJJCG84-86 (ST-2943:CC-574)	CJJHB93-13 (ST-22:CC-22)	CJJCF93-6 (ST-883:CC-21)	CJD269.97 (ST-1845:CC-U/A)	CJJIA3902 (ST-8:CC-21)	CJJRM1221 (ST-354:CC-354)
	n	n	n	n	n	n	n	n	n	n	n	n
ST-474	9	1	1	3	1	9	0	1	9	0	9	1

n:Number of ST-474 genes investigated in this study that were closer to the respective reference genomes.
ST:Sequence type
CC:Clonal complex
U/A:Unassigned

5.3.3 Major outer membrane protein, the *por*A

The lengths of *por*A gene were identical between all the ST-474 genomes whereas, the CJJCG84-86 genome possessed a slightly longer gene sequence by 18 base pairs. The CBI and the scaled chi square codon usage of H22082 (ST-474 genome from human clinical case) genome differed from the remainder. CJJ84-25 was found to be closer to six of the ST-474 genomes, while in contrast, H22082 showed no similarities with any of the reference genomes. In addition, GC3 and the overall GC contents of H22082 also differed from the remaining ST-474 genomes, where CJJ84-25 and CJJHB93-13 were found to be closer to H22082 with respect to their GC3 contents, while CJJ81-176 was found to be closer to the rest of the six ST-474 genomes (Table 5.20). Nucleotide substitution analysis of *por*A showed an overabundance of non-synonymous substitutions (n = 160) compared with synonymous substitutions (n = 82). The number of recombination sites were detected to be 41 and further details on sites of recombination, synonymous and non-synonymous substitutions are provided in Appendix C. A total of 139 codons were found to have amino acid changes of which 32 codons were identified to have negative ω values which means that these codons are under ongoing purifying selection. There were 358 sites identified to be evolving faster than average where the mean evolutionary rate ranged from 1.34 to 3.49 across these sites. The *por*A gene was observed to be under a negative purifying selection (Tajima D = -1.2) while Tajima D for non-synonymous over synonymous substitutions was 1.3.

Table 5.20: Codon usage bias and guanine-cytosine contents of *por*A gene

Genomes	CBI	SChi2	GC3	GC
CJJ11168	0.276	0.255	0.361	0.361
CJJ260.94	0.304	0.273	0.367	0.378
CJJ81116	0.292	0.271	0.357	0.365
CJJ81-176	0.289	0.271	**0.356***	0.364
CJJCG8421	0.304	0.272	0.322	0.356
CJJ84-25	0.278	**0.245**	**0.359**	0.368
CJJCG8486	0.274	0.239	0.353	0.37
CJJHB93-13	0.3	0.269	**0.359**	0.375
CJJCF93-6	0.274	0.246	0.361	**0.366***
CJJD269.97	0.295	0.26	0.337	0.379
CJJIA3902	0.293	0.269	0.354	0.362
CJJRM1221	0.315	0.315	0.32	0.353
H22082	**0.297**	**0.265**	**0.359**	**0.374**
H704	0.277	**0.245***	**0.356***	**0.366***
P110b	0.277	**0.245***	**0.356***	**0.366***
P179a	0.277	**0.245***	**0.356***	**0.366***
P569a	0.277	**0.245***	**0.356***	**0.366***
P694a	0.277	**0.245***	**0.356***	**0.366***
H73020	0.277	**0.245***	**0.356***	**0.366***

*:identity shared between ST-474 and reference genomes.
CBI:Codon usage bias.
Schi2:Scaled chi square.
GC andGC3:Guanine-cytosine - overall and third position.

Schematic representation of the protein structure predicted by Zhang et al. (2000) is provided in Figure 5.13 to facilitate comparison and demonstration of variations that had occurred within the gene. The strands and the surface-exposed or extracellular loops on the amino acid alignment is provided in Figure 5.14 that illustrates the recombination sites and variations within the surface exposed loops. Unlike the housekeeping genes, *por*A was highly variable with marked differences at both the nucleotide and protein levels among the genomes. The differences are presented as a protein alignment in Figure 5.14 with regions that showed recombination sites within the gene. The H22082 isolate of ST-474 differed significantly in its *por*A sequence, at both the nucleotide and amino acid sequence levels. The amino acid sequence alignment suggested recombination sites at the extracellular exposed loops L1 to L8. This has led to a change in the ancestral lineage from ST-21 complex that was closer to ST-474 genomes, to ST-22 complex for the H22082 isolate. A HybridizationNetwork using nucleotide and protein sequences was built using Recomb2007 functionality in the SpitsTree v4 software program and the results are provided in Figures 5.15 and 5.16 at the DNA and protein level, respectively. It is evident from the phylogenetic network that the *por*A gene has undergone considerable amount of nucleotide exchanges to render it more variable where, the network obtained three distinctive groups in a HybridizationNetwork. The first group was formed between six ST-474 genomes (H73020, P110b, P569a, P694a, P179a and H704) and the reference genomes CJJ84-25, CJJCF93-6 and CJJCG8486. The second group comprised CJJIA3902, CJJ81-176 and CJJ81116, and the third group comprised *C. jejuni* doylei, CJJ260.94, CJJRM1221, CJJCG8421 and CJJHB93-13, where H22082 was found closer to CJJHB93-13. CJJ11168 was found to be placed near the root of the tree. While the protein HybridizationNetwork obtained a slightly different tree (Figure 5.16), but however, the predicted ancestry between ST-474 genomes did not change; CJJHB93-13 was still the closer ancestor to H22082 even on a protein tree.

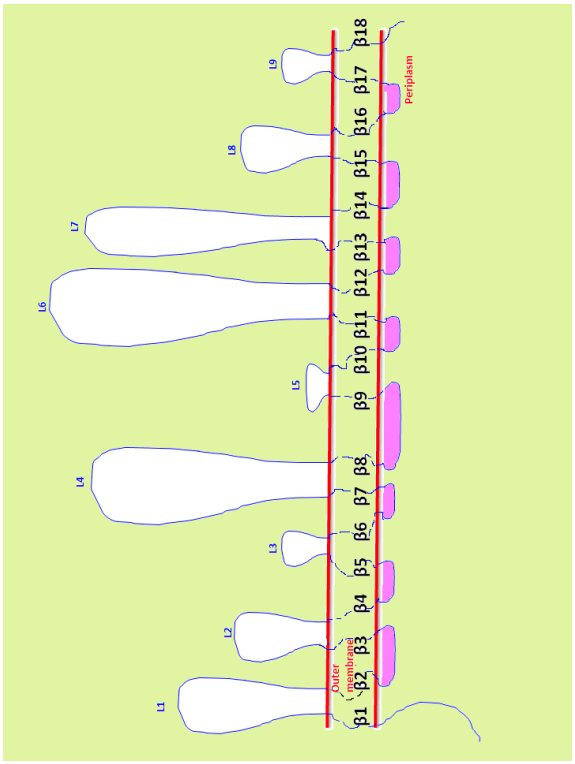


Figure 5.13: Schematic representation of the protein structure predicted by Zhang et al. (2000)

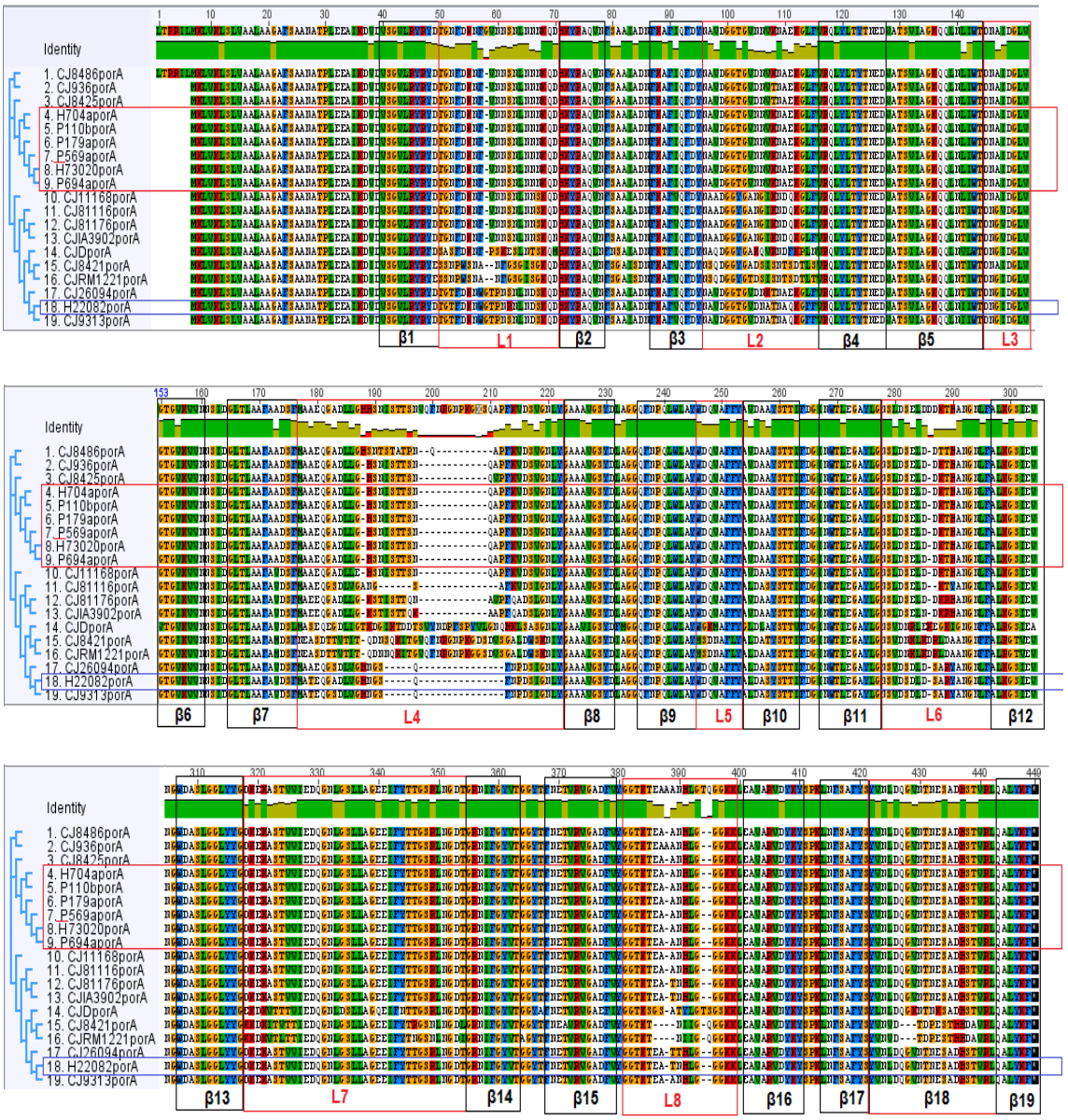


Figure 5.14: Nucleotide and protein recombination alignments of *por*A, and phylogenetic tree. External loops L1, L4, L6 and L8 of H22082 show evidence of recombination that differentiates it from other ST-474 genomes)

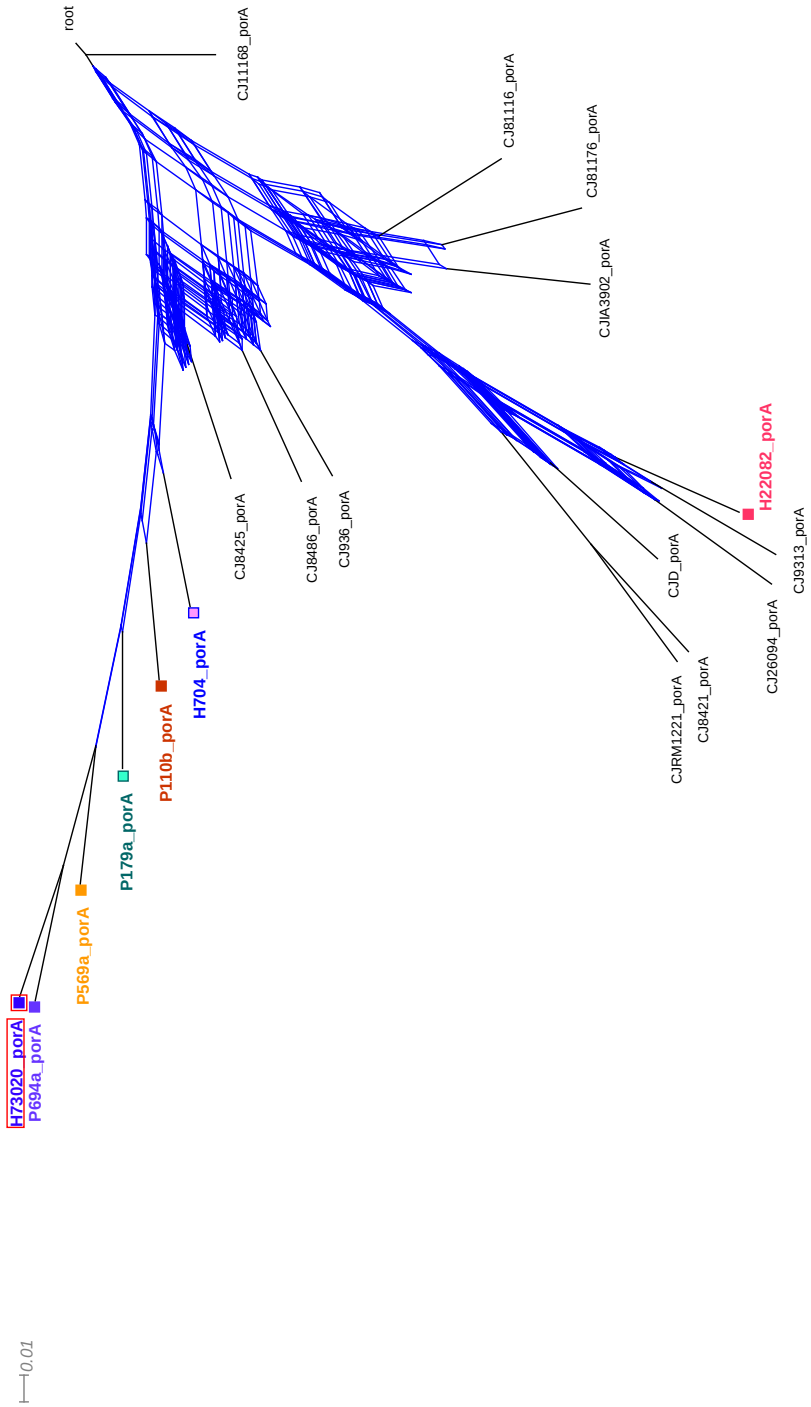


Figure 5.15: DNA based phylogenetic tree of *por*A (MOMP) gene. Recombination tree of *por*A gene constructed using Recomb – 2007 functionality within SplitsTree v4

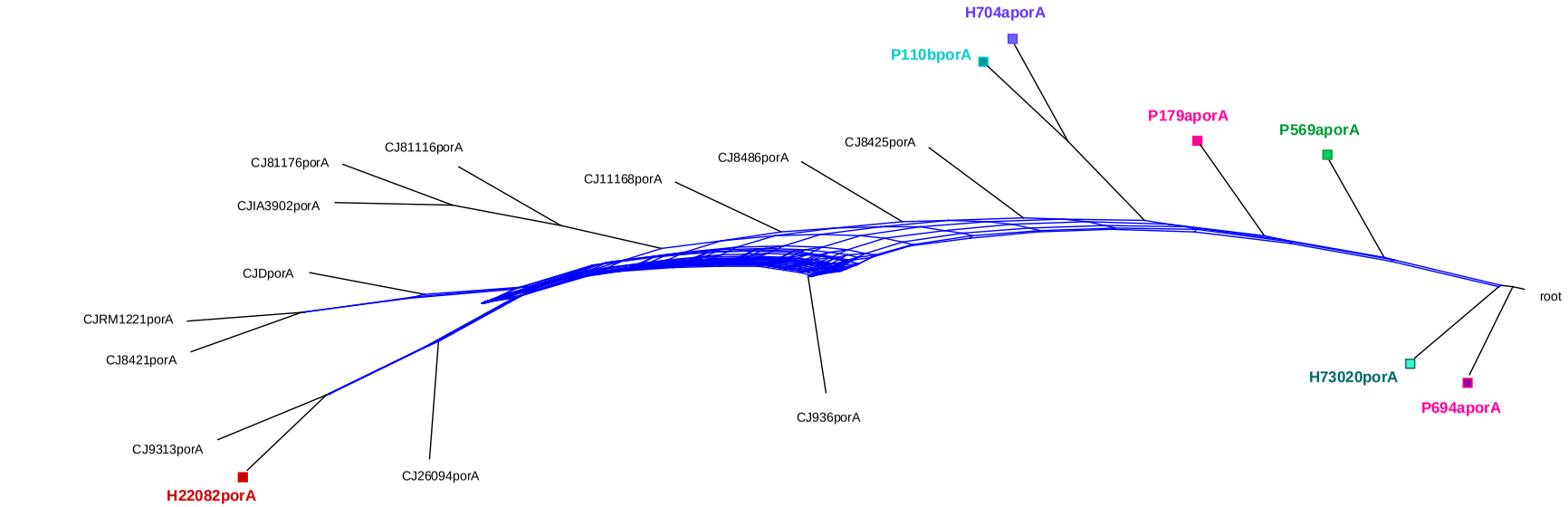


Figure 5.16: Amino acid based recombination phylogenetic tree of *por*A (MOMP) gene. Recombination tree of major outer membrane protein – MOMP (*por*A), constructed using Recomb – 2007 functionality within SplitsTree v4)

5.4 Discussion

Adaptive evolution plays a major role in the diversification of bacterial species that is brought about by variations that occur at both the DNA and protein levels (Lefebure & Stanhope 2009). Genes belonging to the housekeeping lineage have been identified to be highly influenced by major genetic events such as recombination and horizontal gene transfer (Riley 1993, Rivera et al. 1998). Further *C. jejuni* is a species that has been identified to possess a natural competency to undergo frequent DNA changes in response to the environments and/or hosts as a measure of adaptation. This has been demonstrated in the seven MLST housekeeping genes and virulence associated genes (Wang & Taylor 1990, Dingle et al. 2001, Colles et al. 2003, Miller et al. 2006, Colles et al. 2008, Kwan et al. 2008, Wilson et al. 2008, Carter et al. 2009, Wilson et al. 2009). Considering the evidences of general plasticity within the housekeeping genes and the natural competency of *C. jejuni*, the primary objective of this present study was to compare seven *C. jejuni* ST-474 genomes, the identical MLST genotypes, in order to obtain improved understanding about the extent of similarity they possess at full length gene level as a first level of analysis. Secondly, an extended subset of housekeeping genes were analysed to compare the seven ST-474 genomes addressing the same parameters as that of the first subset.

Full length genes from identical MLST housekeeping alleles showed variations at both the nucleotide and protein levels, at least between the reference and ST-474 genomes. This finding suggests that a sub-region within a gene may not possibly reflect the ancestry or lineage of a gene and this requires a full length gene analysis to better understand the evolutionary history of that gene. Differences in the nucleotide base composition of a gene and/or genome is the fundamental element shaping the genomic evolution which in turn, directly influences the GC contents of genes and/or genomes (Hurst & Merchant 2001). GC variation has been thought to be driven both by neutral mutational effects and adaptive selection pressures (Muto & Osawa 1987, Galtier & Lobry 1997, McEwan et al. 1998, Hurst & Merchant 2001, Naya et al. 2002, Foerstner et al. 2005, Musto et al. 2006)

In this study, GC contents of all the genes investigated in the study across the seven ST-474 genomes were identical except for two genes, the *fum*C and *trp*C, while the GC3 content varied in the *pyc*A gene (but showed an identical overall GC content across all the ST-474s). *fum*C is a fumarate hydratase that is involved in the tricarboxylic acid cycle

(TCA) and also an important enzyme that has been identified to be very stable during oxygen fluctuations (Pearson et al. 2003, Miller et al. 2007). H22082 and P694a isolates showed a slightly higher GC contents in their *fum*C gene where it was found that the *fum*C alleles varied greatly even within *C. jejuni* strains that obtained identical alleles with respect to other alleles used in a multi-loci typing scheme (Suerbaum et al. 2001). It is possible that the functional constraints, particularly the oxygen labilities within the hosts and/or environments under which the *fum*C gene functions may be a contributing factor to the underlying genetic variations. In contrast, the GC contents of *trp*C gene which is involved in amino acid biosynthesis belongs to an aromatic amino acid family (Indole-3-glycerol phosphate synthase) varied between the poultry isolates and human isolates. This is suggestive of the environmental influence on this particular gene where, *trp*C gene has been identified to be uncoupled from the *trp* operon (Parkhill et al. 2000) and has been identified as a relatively a divergent sequence in general.

Although, the *pyc*A gene showed an identical overall GC content across all the ST-474 genomes, the GC3 content varied between the H22082 and P694a isolates. It may be hypothesised that the GC3 variation found in the H22082 and P694a isolates within the *pyc*A gene could have occurred as a measure of co-adaptation or co-evolution of the gene to the different elements that were present in the host environment. The GC3 content in bacteria has been shown to play an important role in the evolution of bacteria (Bellgard & Gojobori 1999). It has been reported that the overall GC and the average GC3 contents of a genome are correlated (Bernardi & Bernardi 1985, Bellgard & Gojobori 1999) and it has been proposed that the GC3 changes with synonymous codons can be a determining force for the future change in the overall GC content of a genome (Bellgard & Gojobori 1999). In addition, the concentration of tRNA modification enzymes or the translational decoding system within genomes has been shown to influence the GC3 changes when there is a nucleotide compositional change through recombination (Grosjean et al. 2010). CBI of *pyc*A also varied in these two isolates which is suggestive of a differential evolution. However, analysis of larger datasets may provide support to the hypothesis of *pyc*A differential evolution. Comparison of the overall GC contents among the reference and ST-474 genomes in this study showed CJJ84-25 and CJJCF93-6 to be closer to ST-474 followed by CJJIA3902 and CJJ11168. This observation was in agreement with the MLST profile ancestry prediction.

There was an association found between the GC variance and the number of recombination sites that occurred within different housekeeping genes, where the majority of the genes (investigated in this study) that possessed a wider GC variance showed a higher number of recombination sites. It may be hypothesised that the variation in the GC contents might have arisen due an event of recombination within individual genes which in turn might have caused a GC variation. Hybrid alleles have been documented in the *tkt* and *glt*A genes by studying the MLST datasets elsewhere (Sheppard et al. 2009, 2011), while in this study, these two genes showed wider GC variance within genomes which might have resulted from frequent recombination. This in turn might have led to hybrid alleles as inferred by Sheppard et al. (2009, 2011). In the second subset of genes, *atp*D and *trp*B showed the least number of recombination sites where, the GC content did not vary much across genomes. Variation in the GC content within those genes analysed may be an indication of extensive recombination and/or DNA exchange that occurred within those genes which in turn, is reflected in their GC contents. However, there was one exception, the *inf*B gene that showed a high number of recombination sites with small GC variance. Hence, the hypothesis of association between GC variance and recombination events requires further investigation.

CBI has been identified as a direct indicator of selection, recombination and HGT (Marais et al. 2001, Fuglsang 2003, Tuller et al. 2011). The usage of amino acid is influenced by the nucleotide base composition, which in turn affects the codon usage by genes. Codon usage is highly dependent and/or positively correlated to the tRNA pools present between genomes that are recombining or undergoing lateral gene transfer or HGT (Ikemura 1985, Kanaya et al. 1999, Ochman et al. 2000, Tuller et al. 2011) and hence a distinguishable difference in CBI within genes in a genome can predict an event of recombination in a genome. CBI was another parameter evaluated in this study which showed greater similarity between ST-474 and CJJIA3902, followed by CJJ84-25, CJJCF93-6 and CJJ11168. Although the ancestry was swapped from CJJ84-25 to CJJIA3902, it should be noted that all these reference genomes belong to the ST-21 complex, where the clonal complex view was not completely changed. Collectively, the ST-474, CJJIA3902, CJJ84-25, CJJCF93-6 and CJJ11168 genomes were found to share identical GC contents and CBI within the selected subsets of genes which suggest that these genomes must have undergone a similar pattern of evolution. This in turn might have led the genomes to fit in to a cluster or a

group as predicted by the clonal frame concept.

The Tajima D values obtained for individual genes and codons in this study suggested a stringent purifying selection which is indicative of a population expansion driven by nucleotide substitutions (Tajima 1989). A significant number of codons across all the genes investigated in this study were found to evolve faster which supports the view that *C. jejuni* is a dynamic species evolving under certain situations and changing environment. The selection pressure was found to vary between regions within genes (some regions under purifying selection and others are under positive selection pressure), where, accounting for selective pressures among individual sites and codons has been more insightful. Detection of positive selection in a background of overwhelming purifying selection provides a clearer understanding of functional constraints of genes (Nielsen & Yang 1998, Yang et al. 2000, Yang 2002). It may be hypothesised that, as the metabolic housekeeping genes investigated in this study encode cellular proteins, a high degree of functional conservation and constraint may be expected. A breach in the amino acid composition may not be tolerated as it is regarded deleterious to the survivability of the organism and/or to the gene function (Yang 2002, Susko et al. 2003).

Individual gene trees from the second subset of housekeeping genes showed interesting ancestral lineages for the ST-474 genomes. Most of them obtained a unanimously shared ancestry with CJJ11168, CJJ84-25, CJJCF93-6 and CJJIA3902 whereas, genes such as *fts*Z, *gap*A, *hem*N, *sdh*A and *pyc*A showed differences in their ancestral lineages where other reference genomes were found to be more closer to these genes. The functional role or importance of these genes in a genome are discussed below. *fts*Z encodes a cell division protein which is a crucial protein that plays a prominent role in cytokinesis (de Boer et al. 1992, Zhang & Dong 2005). *fts*Z has been employed for studying the phylogenetic relationship between lactic acid bacterial species and the discriminative power of *fts*Z was greater than that of 16S rDNA (Zhang & Dong 2005). It has also been identified to be a potential target for generating anti-bacterial drugs and has been identified as one of the faster evolving and recombining genes in other bacterial species such as *Wolbachia* a bacterial symbiont (Jiggins et al. 2001). However, it has been found that *fts*Z is not essential for *C. jejuni* survival in contrast to most bacteria (Stahl & Stintzi 2011). In this study ST-474 *fts*Z genes were found to be separated from the reference genomes on a separate branch on the phylogenetic tree. On the other hand, the reference genomes formed

groups on the phylogenetic tree which suggests a few possible explanations for this clustering such as, the grouped reference genomes may have similar evolutionary histories and there might have been a heterogeneous nucleotide substitution rate in each group that might have arisen due to recombination within the reference genomes. Whereas, it seems that *fts*Z allele is a unique variant of ST-474 and a relative geographical separation of New Zealand might be a reason behind this feature. It may also be speculated that, as ST-474 is a rare international strain, the *fts*Z allele might even be a genetic characteristic of this strain itself. However, analysis of *fts*Z gene from *C. jejuni* isolates from different geographical locations within New Zealand and other parts of the world and hosts may shed further light on this observation.

The *gap*A and *pyc*A are two of the most important key bidirectional enzyme coding genes of the glycolysis-gluconeogenesis pathway (Parkhill et al. 2000, Velayudhan & Kelly 2002). The Embden–Meyerhof (glycolysis and gluconeogenesis) pathway is a gluconeogenic pathway that generates glucose-derived polysaccharides through which *C. jejuni* obtains carbon source for its metabolism (Mendz et al. 1997, Parkhill et al. 2000, Velayudhan & Kelly 2002). The *gap*A gene has been evidenced to be down regulated in the presence of chemicals such as sodium deoxycholate. In addition, *gap*A has also been identified to be involved in the successful colonisation of hosts where two variants of CJJ11168 isolates (between the original and the passaged strain) showed significant molecular variations between their *gap*A genes (Gaynor et al. 2004, Malik-Kale et al. 2008, Tunio et al. 2010). This reflects the rapidity of this gene to evolve under given circumstances. While in this study *gap*A was identified with relatively minimal numbers of recombination sites and was shown to be closer to CJJCG84-21, a safer vaccine strain that lacks ganglioside mimicry (Tribble et al. 2009). Phylogenetic attempts using wider molecular datasets of genes directly involved in the expression of similar traits or characteristics between ST-474 and CJJCG84-21 will provide further useful insights into the genetic properties that these two strains share in common. The *pyc*A gene encodes for an anaplerotic enzyme, the pyruvate carboxylase and this metabolic enzyme activity was evidenced to be affected in the *pyc*A mutants that in turn affect the colonisation and growth of *Campylobacter* (Velayudhan & Kelly 2002). The *pyc*A genes from ST-474 formed two different clusters, H22082 and P694a formed one cluster and the remainder formed another cluster. These findings indicate that these two genes have undergone a similar

physiological stress that might have probably led to their differential evolution while an anaplerotic enzyme is highly essential for the growth of *C. jejuni*.

*hem*N, a oxygen-independent coproporphyrinogen-III oxidase and *sdh*A, a succinate dehydrogenase flavoprotein subunit have been identified to be important for the growth of *C. jejuni* under varied growth conditions or environments (Guo et al. 2008, Reid et al. 2008, Wright et al. 2009). As strict anaerobiosis is a stress condition for *C. jejuni* that might be expected inside the guts of different hosts, *hem*N is one of the important genes required under oxygen-limited conditions as the inability to synthesise heme anaerobically due to an absence or a mutated *hem*N gene inhibits growth (Sellars et al. 2002). Notably, *sdh*A has been shown to be of paramount importance for successful colonisation and growth of *C. jejuni*. This gene has also been identified to be up-regulated in the chicken caecum and to sense and respond to decreasing environmental pH suggesting its role in adaptation (Grant et al. 2005, Guo et al. 2008, Reid et al. 2008). Bacteria are shown to use many different protective strategies to combat acid stress where they may involve alterations of cell surfaces and membranes in order to prevent the influx of hydrogen ions into the cell (Reid et al. 2008). Apart from *sdh*A, *glt*A, *glt*B and *ilv*D were also identified to possess a differential expression pattern to combat acid stress where these genes are down-regulated during acid stress conditions (Reid et al. 2008).

In this study, the *glt*B gene showed the second highest number of recombination sites which might possibly be due to the past evolutionary events as a measure of stress adaptation and on the phylogenetic tree all of the ST-474 alleles were found to be identical. Another interesting gene with special molecular characteristics found in this study was *sdh*A. The *sdh*A gene length differed between reference genomes while fragmentation of this gene has been reported in some regions of *sdh*A in *C. jejuni doylei* 269.97 (Nachamkin & Blaser 2000, chap. 5) raising a suspicion around the functionality of *sdh*A in this genome. Similarly the ortholog of *sdh*A from CJJCG84-86 was heavily fragmented and the gene sequences from both *C. jejuni doylei* 269.97 and CJJCG84-86 were not found complete to be included for analysis in this study. These observations reflect the possibility of this gene having been damaged by the environmental pH (these two strains were isolated from human clinical cases). In the phylogenetic analysis in this study, the *sdh*A and *hem*N gene sequences of ST-474 genomes were closer to the CJJRM1221 reference genome while it should be remembered that CJJRM1221 was isolated from a chicken carcass. This find-

ing might further be speculated that the *sdh*A from chicken may present a similar genetic characteristic, however it needs further data analysis to provide more support for this hypothesis.

While the *por*A cell surface antigen analyses in previous reports have identified three deep divisions within the *por*A alleles and the MOMP protein sequences (Zhang et al. 2000, Clark et al. 2007, Cody et al. 2009). The third division or group has been found to be extensively involved in recombination in contrast to the other two groups (Zhang et al. 2000, Clark et al. 2007, Cody et al. 2009). In this study, there was extensive recombination observed between the *por*A gene and MOMP protein sequences while the ancestry of ST-474 was swapped from ST-21 to ST-22 complex in the H22082 isolate (ST-474 genome from a human clinical case). CJJ81-176 was found to be closer to the rest of the six ST-474 genomes. The H22082 isolate differed in its CBI and the scaled chi square codon usage index.

The molecular variations and the differences in the selection pressures over this gene suggest the *por*A diversification and various evolutionary alterations the gene might have undergone under different circumstances. Further, the three distinctive groupings of *por*A gene sequences identified among the ST-474 and reference genomes suggest that, the group that possessed H22082 could belong to the third division of the *por*A allele that was found to undergo frequent recombination by previous studies (Zhang et al. 2000, Clark et al. 2007, Cody et al. 2009). The *por*A gene of *C. jejuni* subsp. *doylei* 269.97 was also found to be grouped with H22082. Furthermore, the variations exhibited in the external loops, particularly in loop one, four, six and eight suggest that the majority of molecular variations were concentrated towards the external loops which might be related to functional, biological and virulence properties of the protein as these loops are more exposed to the environment directly (Zhang et al. 2000, Clark et al. 2007, Huang et al. 2007, Cody et al. 2009). However, the association between the structural variations and disease, virulence properties, their role in ecological adaptation and their behaviour under stress have not been studied in detail so far. Studies relating to function and expression of MOMP proteins under various environments and/or hosts involving wider datasets will provide useful insights into the evolution of *por*A gene which in turn will serve as a potential vaccine candidate.

Conclusion

To conclude, this study has validated the reliability of a subregion within a gene to be used equivalent to a full length gene involving a limited number of genomes in an established typing scheme such as MLST. In addition, this study has introduced the idea of analysing identical alleles at their full length level that is essential for studying evolution of new variants. Even though there were molecular variations found amongst identical alleles at the level of full length genes, the phylogenetic prediction of the ST-21 complex being the closest ancestor of ST-474 genomes was still valid, even when 18 additional full length genes were employed. The closer ancestor of ST-474 was found to be CJJIA3902 (ST-21 complex) genome from the concatenated sequence of 25 genes which implies that the ST-474 and ST-21 complex genomes might have had similar evolutionary histories. While all of the *por*A gene sequences of ST-474 were closer to CJJ84-25, the H22082 isolate was found to be closer to CJJHB93-13. The Tajima'D values of whole genes indicated that the metabolic housekeeping genes are under stringent purifying selection. But there were regions within the genes that were under positive selection pressure. It was more insightful to consider the selection pressures operating within individual genes as it provides useful information about the functional constraints of those genes. This observation is in agreement with previous studies that demonstrated differences in the selective pressures operating within a single gene (Yang 2002, Susko et al. 2003). It was interesting to find that the *por*A gene was under purifying selection pressure, given its high variability and antigenic status.

The findings of genes (investigated in this study) differing at both nucleotide and amino acid levels between the seven ST-474 genomes lend support to the findings made by Biggs et al. (2011) who showed differences between two ST-474 genomes on a genomic scale. In addition, the special molecular signatures that the genes of ST-474 genomes possessed (as discussed earlier) may be hypothesised as a reflection of either the geographical isolation of ST-474 or it may just be the characteristic of ST-474 genome itself. Many of the novel biosynthesis pathways in *Campylobacter* spp. have been regarded to be still in its infancy (Nachamkin & Blaser 2000, chap. 4) and hence, this study suggests that functional exploration and gene expression assays of genes with special phylogenetic signals will provide further insights into the underlying functional mechanisms that may play a role in diversifying strains that belong to a single genetic type. Finally, analysis of wider datasets

of genes will provide more support to the current findings of this study.

CHAPTER 6

A genome wide comparison of selected subsets of ribosomal and DNA repair genes among seven *Campylobacter jejuni* MLST ST-474 isolates for evolution

Abstract – DNA repair and replication play a central role in shaping the evolution of an organism which provides enormous potential for evolutionary diversification of bacteria. *C. jejuni* multilocus sequence type 474 (ST-474) is an internationally rare genotype that was responsible for 24% to 34% of human clinical cases in New Zealand. In the present study, seven ST-474 genomes (four from the chickens and three from human clinical cases) that were collected between August, 2005 and February, 2009, were compared with 12 other *C. jejuni* reference genomes from the GenBank database. A subset of 25 genes (ribosomal, DNA repair and nucleotide metabolism genes) were analysed to understand the genetic similarities and differences amongst these genes within the seven ST-474 genomes in an effort to understand the evolution of the ST-474 genotype. Genes investigated in this study showed differential evolutionary histories, with seven of 25 genes differing in their nucleotide and amino acid compositions. The alleles *gid*A, *ogt*, *rec*J, *ssb*, *uvr*A, *uvr*B and *xse*A were found to be unique to the ST-474 genomes in this small data set. Further, recombination was evident in all of the genes. The majority of genes investigated in this study shared their ancestral lineage with the ST-21 complex. In addition, three genes *uvr*A, *gyr*A and *mut*S were found to have selenocysteine in their protein sequences which provided some evidence of horizontal gene transfer. This study provided a better understanding of the repair, ribosomal and nucleotide metabolism genes, their genetic characteristics and differences within a small selected set of genes in ST-474. In addition, the MLST scheme seven housekeeping alleles were found to be robust in providing a bird's eye view of the ancestral lineage of the majority of genes in a genome.

Vathsala M, French N, Biggs, P J, Stevenson M, Marshall J, and Hotter G, (2011). A genome wide comparison of selected subsets of ribosomal and DNA repair genes among seven *Campylobacter jejuni* MLST ST-474 isolates for evolution

6.1 Introduction

Living organisms are threatened by different environmental and biological agents where maintenance of their genome integrity or stability is a continual challenge. DNA repair mechanisms in organisms play a fundamental role in protecting the cells against damage as well as in the faithful transmission of genetic information from a mother cell to the daughter (Singh et al. 2010). Horizontal gene transfer (HGT) and recombination are the two major forces that shape the evolution of archaeal and bacterial genomes generating genetic diversity without losing too much genomic stability (Doolittle 1999, Feil et al. 2001, Spratt et al. 2001, Townsend et al. 2003, Fall et al. 2007). Interspecies recombination has been demonstrated between donor and recipient DNA molecules that differ by up to 25–30% of their nucleotide sites (Dowson et al. 1989, Reeves 1993, Bowler et al. 1994). However, the impact of these evolutionary forces may differ between genomes, and between genes in a genome, dependent on the prevailing biological and ecological factors. These include microbiota competitions, environmental conditions, the natural competence of bacteria and the DNA repair mechanisms that may vary from species to species, or even within sub-populations of the same species (Feil et al. 2001, Townsend et al. 2003, Mau et al. 2006, Fall et al. 2007, Gaasbeek et al. 2009a,b)

Population structures of bacterial species have been studied using techniques such as multilocus enzyme electrophoresis (MLEE) to index the allelic variations that targets the cellular enzymes encoded by genes in a genome (Maiden et al. 1998). This was eventually superseded by mulitlocus sequence typing (MLST) by taking the basic principle of multiloci analysis. MLST integrates the information from the internal fragments of seven housekeeping genes that are approximately 470 to 500 base pairs long (Maiden et al. 1998, Maiden 2006). Bacterial isolates are grouped into different clusters based on the allelic combination that provides an allelic profile of the seven housekeeping genes in MLST, that in turn places the strains in a clonal context. Monomorphic bacterial pathogens (with little sequence diversity) (Comas et al. 2009) serve as promising clonal models of evolution in contrast to polymorphic organisms with greater sequence diversity (Linz et al. 2000, Dingle et al. 2001, Achtman & Wagner 2008). The latter showed complex phylogenetic histories due to extensive recombinational events (Wirth et al. 2006, Achtman & Wagner 2008). *Campylobacter jejuni* is one example of polymorphic bacterial species

that is naturally competent to take up exogenous DNA and to transform its genome characteristics, as proven experimentally by Wang & Taylor in 1990. However, behaviour observed in a bacterium under laboratory conditions is not always identical to that experienced by the bacterium in the wild.

6.1.1 Genes of genetic information processes in bacteria

DNA repair, recombination and nucleotide metabolism pathways are some of the important key machineries of the genetic information processes in bacteria.[1] The DNA repair and recombination (DRR) pathways have contrasting roles in evolution, where they protect the genome from damage as well as accommodate a certain degree of mutation; thus balancing the two activities to maintain the genetic stability by correcting the deleterious mutations as well as enhancing the survivability of an organism in a fluctuating environment (Cann & Ishino 1999, Paques & Haber 1999, Zhou & Elledge 2000, Singh et al. 2010). The genetic stability in turn is defined as 'a measure of the resistance to change, with time, of the sequence of genes within a DNA molecule or of the nucleotide sequence within a gene'.[2] There has been an enormous increase in our understanding of the genetics and biochemistry of DNA repair mechanisms and pathways, where DNA repair in bacteria have evolved to correct DNA damage caused by a variety of agents such as ultraviolet (UV) irradiation (Cox 1998, Zhou & Elledge 2000, Lusetti & Cox 2002, Fall et al. 2007, Janion 2008, Fonville et al. 2010, Patel et al. 2010, Singh et al. 2010). However, most of the biochemical mechanisms remain obscure, requiring specialised experimental conditions to better understand every biochemical mechanisms of DNA repair (Swingle et al. 2010).

The DNA repair system involves four major pathways such as simple reversal of DNA modifications (e.g. *ada* methyltrasferase (Nieminuszczy & Grzesiuk 2007)), excision of damaged nucleotides from DNA, called base excision repair (BER) (Krwawicz et al. 2007), removal of whole damaged DNA fragment, called nucleotide excision repair (NER) and methylation-directed mismatch repair (MMR) (as reviewed by (Maddukuri et al. 2007, Nowosielska 2007)). An inducible DNA repair network was first discovered about

[1] URL:(http://www.genome.jp/dbget-bin/www_bget?ko+K03495)

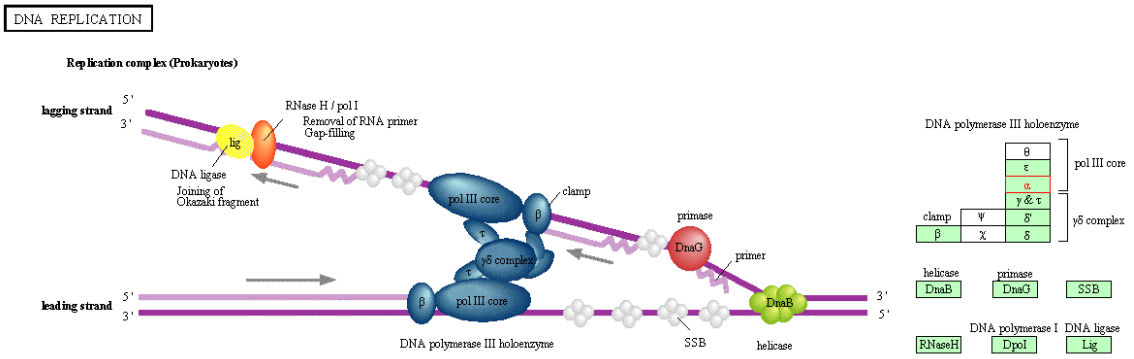
[2] URL:(http://http://www8.nos.noaa.gov/coris_glossary/index.aspx?letter=g)

30 years before in *Escherichia coli* and the term 'SOS response' was introduced to describe this network (Radman 1975, Michel 2005). The genes that were identified in the DNA repair mechanism as the first line defenders are *uvr*A, *uvr*B, and *uvr*D. These genes are involved in the nucleotide excision repair while the homologous recombination functions appear second in the line of defense against DNA lesions, including genes such as *rec*A from the RecBCD pathway (Michel 2005, Gaasbeek et al. 2009a). While DNA repair systems have undergone evolution in order to correct specific DNA modification and incorrectly paired bases, recombinational (homologous and non-homologous) repair systems play a major role in the recognition and repair of DNA breakage that have also evolved markedly, requiring extensive homologies for efficient homologous recombination (Lovett et al. 2002). In contrast, non-homologous recombination or non-homologous end joining (NHEJ) is used in situations where only one copy of the chromosome is available for repair requiring an approximation of the broken ends (Nowosielska 2007, Shuman & Glickman 2007).

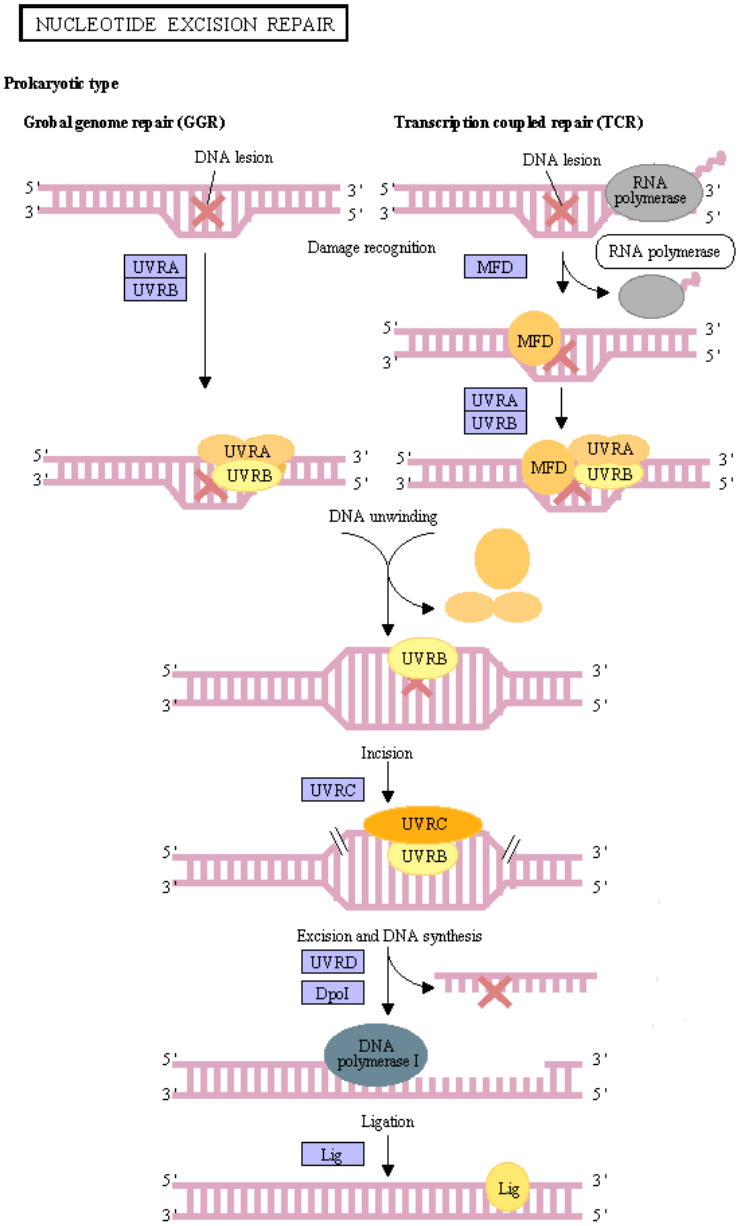
Homologous recombination (HR) is error free unlike NHEJ which is often mutagenic. However, NHEJ can be advantageous under certain situations as reported in eukaryotes in the diversification of immune repertoire (Lusetti & Cox 2002, Shuman & Glickman 2007). Bacterial NHEJ is an emerging field where the characterisation of an ATP dependent ligase in *Haemophilus influenzae* was a breakthrough in this field with subsequent identification and characterisation of NHEJ and associated specialised enzymes in *Mycobacterium tuberculosis*, *Mycobacterium smegmatis* and *Agrobacterium tumefaciens* (Cheng & Shuman 1997, Nowosielska 2007). The studies on NHEJ demonstrate the dynamic relationship between pathogenesis and DNA repair and their significant contribution in the emergence of new variants and resistant strains in bacterial species such as *Mycobacterium* spp. (Weller et al. 2002, Jacobs et al. 2003, Cirz et al. 2005, Gong et al. 2005, Martinez et al. 2005, Gandhi et al. 2006, Curti et al. 2007, Sinha et al. 2007), which is a potential public health concern. However, the field of NHEJ is regarded naive, with respect to the role of NHEJ in the disease ecology and impact of NHEJ on virulence of several bacterial species that require suitable genetic and biochemical tests (Nowosielska 2007). Schematic representation of the replication (as genes of replication are also involved in DNA repair, process of replication is included) and repair pathways are pro-

vided in Figures 6.1, 6.2, 6.3, 6.4 and 6.5, adapted from the KEGG pathway database.[3]

[3]URL: (`http://www.igs.cnrs-mrs.fr/mgdb-cgi/www_gene_catalog?rpr.ann.`)



(a)



(b)

Figure 6.1: Replication and DNA repair system in prokaryotes. Genes involved in the pathway of replication and replication process (a); Nucleotide excision (NER) pathway and genes involved in the repair process (b)

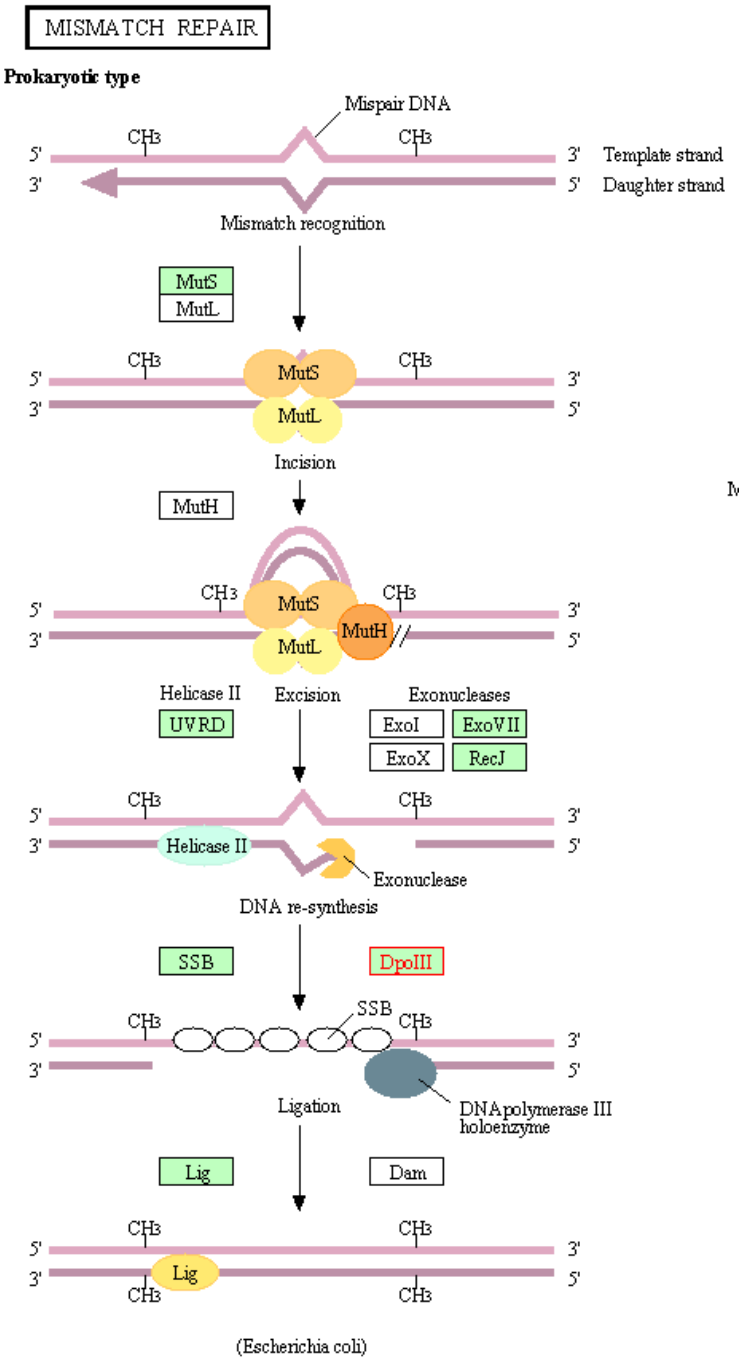


Figure 6.2: Mismatch repair system in prokaryotes. Genes involved in the mismatch repair mechanism (MMR)

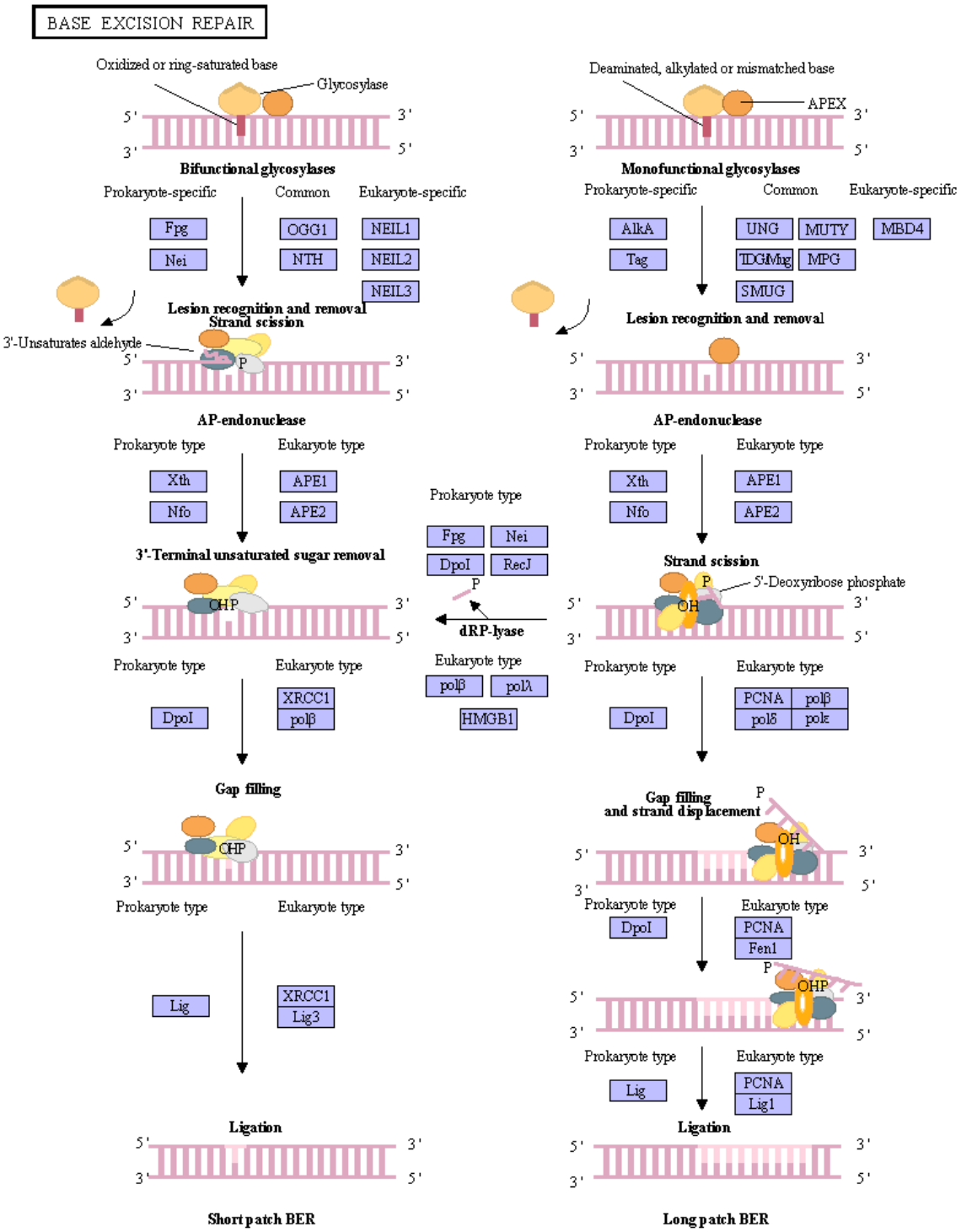


Figure 6.3: Base excision repair system in prokaryotes. Genes involved in the base excision repair mechanism

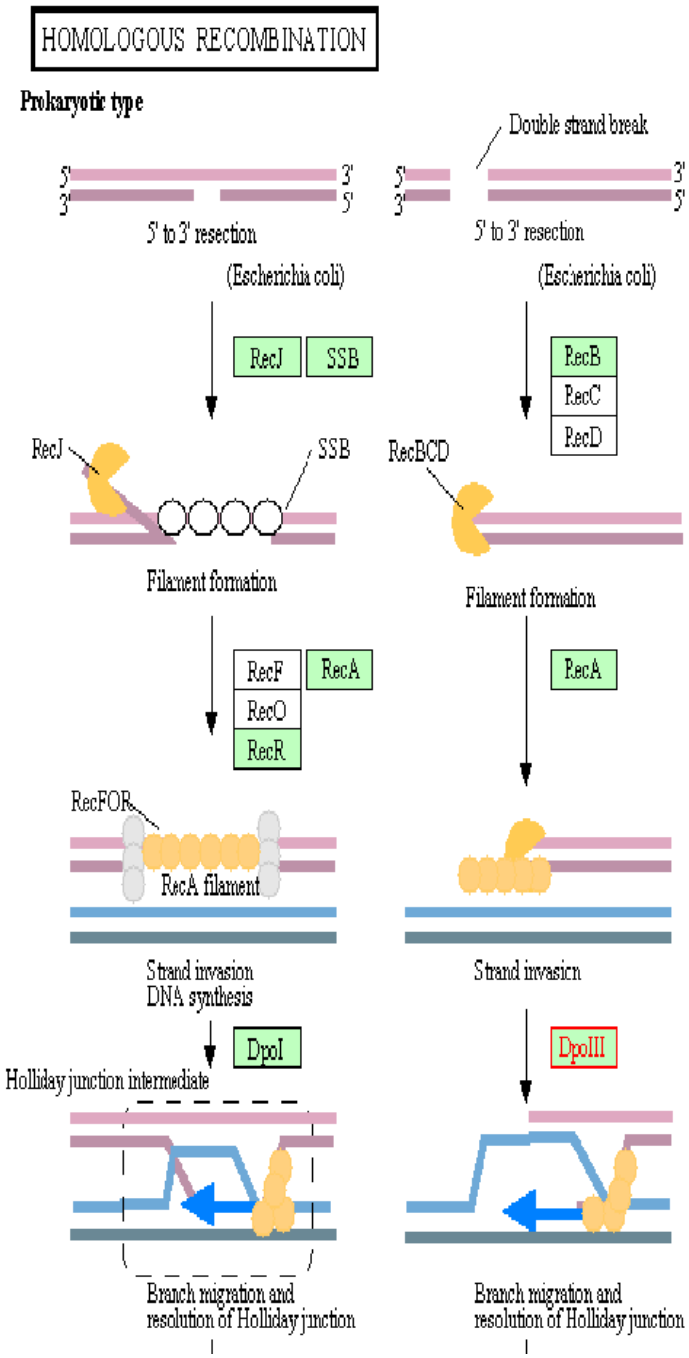


Figure 6.4: Recombinational repair systems in prokaryotes – A. Genes involved in the homologous recombinational repair mechanism

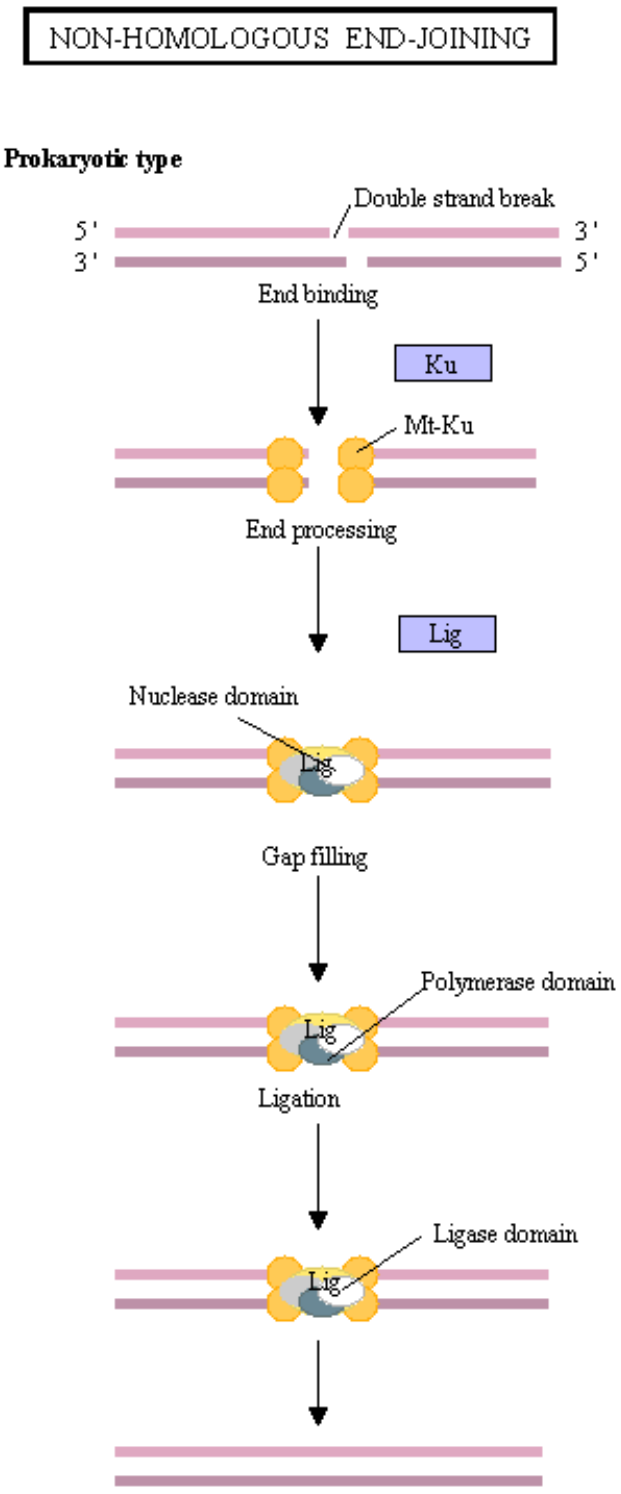


Figure 6.5: Recombinational repair systems in prokaryotes – B. Genes involved in the non-homologous recombinational repair mechanism

DNA repair genes and evolution

Exposure to several cycles of selection pressures and fluctuating levels of DNA-damaging factors such as chemicals, antibacterials and other environmental agents have been identified as shaping the DNA repair pathways in modern-day organisms (O'Brien 2006). Cellular localisation, protein–protein interactions and other regulatory mechanisms have been identified to be linked and/or to be influencing evolution of DNA repair pathways (O'Brien 2006).

The continuous need to identify new DNA damage and to repair new types of DNA lesion, act as a powerful selective force for the DNA repair system to evolve in response to changing intracellular and extracellular environments (O'Brien 2006). This in turn enables bacteria to survive stress conditions and to adapt to a new environment. For example, the *gyr*A gene that encodes a DNA gyrase involved in DNA replication and repair mechanisms has been identified to be a potential precursor for the emergence of drug resistant *Campylobacter* mutants. A mutation at the 86th codon, confers fluoroquinolone (FQ) resistance in *Campylobacter* spp. (Han et al. 2008). The primary action involved in this instance is to repair the damaged DNA strand caused by the antibacterial agent which in turn has led to the emergence of drug or FQ resistant *Campylobacter* mutants. Thus, DNA repair pathways and the genes involved in these pathways can potentially mediate evolution of an organism through selective DNA repair processes.

A unique feature observed in the FQ resistance development was the rapidity of emergence of mutants. FQ resistant mutants were found to emerge after an initial exposure of susceptible *Campylobacter* strains to FQ antimicrobials (under experimental conditions and/or when Campylobacter-infected humans or animals were treated with FQ antimicrobials) within 24-48 hours (Segreti et al. 1992, van Boven et al. 2003, Luo et al. 2003, Zhang et al. 2003, Griggs et al. 2005). Similarly, the *mfd* gene (mutation frequency decline gene) that encodes a transcription-repair coupling factor has been identified to confer FQ resistance in *Campylobacter* spp. (Han et al. 2008). The *mfd* gene is involved in DNA repair and has been shown to increase the frequency of emergence of spontaneous FQ mutants in *Campylobacter*. In contrast, mutation in the *mfd* gene has been shown to decrease the frequency of streptomycin-resistant *Campylobacter* mutants (Han et al. 2008). It has been thought that *mfd* might enhance replication of the non-repaired DNA strands through

a transcriptional bypass so as to maintain cell viability and/or to promote mutations for drug resistance (Han et al. 2008). Even though, it has been shown that *mfd* confers resistance to antibiotics, the exact mechanism through which this gene facilitates the increased frequency of emergence of drug resistant mutants is not clearly known. These two genes, the *gyr*A and *mfd* are both thought to be involved in DNA repair mechanism, where *mfd* enhances mutation in order to confer fitness to the affected bacteria. In contrast, *gyr*A repairs itself in order to confer fitness to the affected bacteria.

Campylobacter is the most commonly identified gastro-enteric pathogen causing diarrhoea in humans in developed countries (Nachamkin & Blaser 2000, Adak et al. 2005). The disease is usually self limiting and uncomplicated but the infection may lead to serious sequelae such as Guillain-Barré syndrome and reactive arthritis (Bremell et al. 1991, Zia et al. 2003). *C. jejuni* and *C. coli* are the recognised pathogenic species of the genus *Campylobacter*. The majority of the human campylobacteriosis cases are caused by *C. jejuni* with *C. coli*, *C. upsaliensis*, *C. fetus* and *C. lari* accounting for a relatively small number of cases (Gillespie et al. 2002, Lopez et al. 2002, Sheppard et al. 2009). *C. jejuni* can be retrieved from a wide variety of hosts and environments (Nachamkin & Blaser 2000, Jones 2001) and the bacterium has been reported to thrive well in the digestive tracts of birds and mammals, (Lee & Newell 2006) and in moist, cold dark environments (McCarthy et al. 2007). Human campylobacteriosis is commonly associated with the consumption of undercooked poultry, meat, contaminated water and through environmental exposures (French et al. 2005, Fullerton et al. 2007, French 2008, Marcus 2008, Sopwith et al. 2008, Mullner et al. 2009).

The design and application of the MLST scheme for *Campylobacter* spp. demonstrated the massive evolutionary and recombination potential of *C. jejuni* (Dingle et al. 2001, 2005). The adaptation of *C. jejuni* to different niches and the acquisition of definitive host signatures through recombination and the formation of new variants has been demonstrated in previous studies in greater detail (Colles et al. 2003, McCarthy et al. 2007, Colles et al. 2008, 2009). The antigenic genes of *C. jejuni* (the *fla*A, *fla*B and *por*A) were shown to undergo frequent recombination and are considered alongside the MLST housekeeping genes to provide a high discriminatory ten locus typing system (Meinersmann et al. 1997, Meinersmann & Hiett 2000, Meinersmann et al. 2002, Dingle et al. 2008, Cody et al. 2009).

Studies involving MLST datasets have shown that gene flow between species (between *C. jejuni* and *C. coli*) and intra-species is rapid (Sheppard et al. 2009, Wilson et al. 2009, Sheppard et al. 2011). Further, progressive fragmentation of imported DNA leading to the creation of mosaic alleles or hybrid alleles has been documented as an ongoing process of recombination between *C. coli* and *C. jejuni*. This has lead to recent speculation about the convergence of these two species recently (Sheppard et al. 2009, 2011). However, the molecular characteristics and the evolutionary potential of those genes apart from MLST housekeeping genes has received little attention.

6.1.2 Measures of evolution

Genetic diversity in a gene or genome is measured by estimating the nucleotide composition (adenine, cytosine, guanine and thymine or uracil), particularly the guanine-cytosine (GC) content, nucleotide substitutions, selection pressures and amino acid and/or codon usage (Rocha & Feil 2010) that are briefly described below.

Guanine-cytosine content

The guanine-cytosine (GC) content is positively correlated with the genetic stability of a gene and/or a genome (Yakovchuk et al. 2006, Hildebrand et al. 2010), where DNA stability significantly influences a number of biological processes within a cell (Gueron et al. 1987, Frank-Kamenetskii 1987, Yakovchuk et al. 2006). In bacteria, both coding and non-coding regions within a gene and/or a genome are affected due to the small number of intergenic DNA sequences (Hildebrand et al. 2010). DNA stability is mainly dependent on the base pairing and stacking of guanine-cytosine (GC) and adenine – thymine (AT) (Yakovchuk et al. 2006).

Nucleotide base composition variation is a consequence of differences in the patterns of evolutionary events (Sueoka 1961, Freese 1962). Further, GC content in bacteria is influenced by (1) genome size where larger sized genomes tend to be GC rich (Bentley & Parkhill 2004); (2) the habitat (e.g. ocean water bacterial samples have low GC and farm soil bacterial samples have high GC content) (Foerstner et al. 2005); (3) aerobiosis – aerobic organisms are GC rich) (Naya et al. 2002); (4) nitrogen utilisation – nitrogen fixing bacteria have high GC content (McEwan et al. 1998); (5) temperature where GC content

is thought to be positively correlated with temperature (Galtier & Lobry 1997, Hurst & Merchant 2001, Musto et al. 2006) and (6) the living nature or lifestyle of bacteria – bacterial symbionts that are dependent on their host are rich in AT (Rocha & Danchin 2002, Woolfit & Bromham 2003). The conversion of GC→AT and AT→GC differs according to the overall GC content of a bacterial species, where AT→GC conversion is common in GC rich bacteria and *vice versa* in AT rich bacterial species such as *Campylobacter* (Hildebrand et al. 2010). Of the three (one, two and three) codon positions, the third codon, GC3 has a positive correlation with the overall GC content of a genome while the conversion of GC→AT and AT→GC at this position is thought to occur as a measure of maintenance of the overall GC content of the genome (Akashi 1995, Hildebrand et al. 2010). However, the exact underlying causes that lead to the differences in GC content are not clearly defined.

Nucleotide substitutions and selection pressures

Three major selection processes influence the rates of nucleotide base substitutions in a protein-coding gene; these include negative or purifying selection against deleterious nucleotide substitutions or mutations, random genetic drift of neutral mutations and adaptive or positive selection (Patthy 2008). Substitutions can be synonymous or non synonymous, where a synonymous or silent mutation / substitution does not change the translated amino acid in contrast to a non-synonymous substitution that leads to an amino acid change. The synonymous substitutions (d_S) (also interchangeably referred to as K_s) are considered neutral and do not produce deleterious effects on a protein coding gene and are often fixed in a population (Kimura 1968, Patthy 2008). The non-synonymous substitutions, on the other hand (d_N) (also interchangeably referred to as K_a) play a major role in protein evolution and in the fitness of an organism. Mathematical and statistical models were developed to measure and infer evolution based on synonymous and non synonymous differences in genes. These models were initially applied for evolutionary studies in eukaryotes and were subsequently adapted to analyse evolution in prokaryotes. The models that have been widely accepted and used are: (1) the substitution models (nucleotide and amino acid); that measure the relative occurrence of nucleotide substitutions among the four nucleotides and the relative occurrence of substitutions of amino acids among 20 amino acids in a lineage. These models have subsequently undergone several

improvements and modifications (Jukes & Cantor 1969, Kimura 1980, Felsenstein 1981, Hasegawa et al. 1984, 1985, Nei & Gojobori 1986, Jones et al. 1992, Tamura & Nei 1993, Yang 1994b), (2) Codon substitution models; these models consider a sense codon that encodes an amino acid as an unit of evolution (Hasegawa et al. 1985, Yang et al. 1998) which in turn measures the relative codon frequencies that had resulted from nucleotide substitutions. (3) Site models; that test the effect of selection at individual sites in an aligned protein coding DNA sequence (Suzuki & Gojobori 1999, Yang et al. 2000).

A relationship was developed according to the nucleotide frequencies and the type of nucleotide substitution (synonymous or non synonymous) namely ω in all the above three mentioned models. ω is the ratio of d_N/d_S, is a measure of natural selection that had occurred in a coding gene sequence (Whelan et al. 2001). Variations in the ω values denote the type of selection pressure operating in a particular site or codon or a gene where, an ω value less than one signifies a purifying selection, ω values equal to one and greater than one represent neutral and positive Darwinian selection respectively (Tanaka & Nei 1989, McDonald & Kreitman 1991, Muse & Gaut 1994, Whelan et al. 2001). However, non-synonymous substitutions that influence the positive selection and/or fitness of an organism are expected to occur only at sites that are critical and essential for an advantageous function (Patthy 2008). As a consequence, the number of such substitutions affects the overall ratio of d_N/d_S of a gene. A small number of non-synonymous substitutions in a given gene may be overridden by purifying selection operating in other sites within a coding gene (Muse & Gaut 1994, Patthy 2008). Hence, analysis of selection pressures at the level of individual codon and/or site is absolutely necessary when protein-coding genes are analysed in order to avoid an overall d_N/d_S ratio masking the positive selection pressure operating within different sites in a gene (Patthy 2008). Non-synonymous substitutions have been further categorised into three distinct classes such as deleterious, neutral and advantageous (Ohta 2002, Hughes et al. 2008). Deleterious variants are removed immediately by purifying selection and in some instances it may lead to a false inference of positive selection by providing high positive values for a gene and/or for the sites or codons investigated (Hughes et al. 2008). An ongoing purifying selection is said to be identified by the heterozygosity in gene sequences with lower number of non-synonymous substitutions while an advantageous selection is a positive selection that increases the fitness of an organism (Hughes et al. 2008).

The substitution rates across sites are considered to vary independently and each site has been assumed to have a different evolutionary rate in any given gene (Jin & Nei 1990, Yang 1993). Therefore, a gamma distribution with a scale parameter was developed to measure the variability in the mean evolutionary rates among sites by Jin & Nei (1990) and Tamura & Nei (1993), where the rates have been scaled to represent an average evolutionary rate of one across all sites in a protein-coding gene. Thus, sites showing a rate lower than one are meant to evolve slower than average, and those that show a rate greater than one are considered to evolve faster than average (Tamura & Nei 1993, Tamura & Dudley 2007).

Codon usage bias index

Bacteria use more than one codon to encode an amino acid hence the same amino acid can be encoded by different codons (synonymous codons) (Snyder & Champness 1997). The codon preferences differ amongst different bacteria which has been shown to be dependent on the the tRNA concentrations in the genome (Snyder & Champness 1997, Ikemura 1985, Kanaya et al. 1999, Tuller et al. 2011). GC content and/or the nucleotide base composition of a gene and/or an organism is directly related to the codon usage preferences. Codon bias index (CBI) is a measure of the extent to which a gene uses a subset of optimal codons (Bennetzen & Hall 1982). CBI is expected to be equal to 1.0 in a gene with extreme codon bias and 0.0 in a gene with random codon usage. A negative CBI value is obtained when the number of optimal codons is less than expected by random change. The scaled chi square is a measure of codon usage bias in the silent codons and has been employed to measure the general synonymous codon usage bias in genes that have more than 100 codons (Shields et al. 1988). CBI within species and/or within genes from a species or sub-species has important uses, such as detection and timing of HGT where the introduced genes have atypical CBI compared to the recipient (Ermolaeva 2001), and to determine the recombination-dependent mutational patterns that show subtle variations in their CBI (Marais et al. 2001).

6.1.3 New Zealand and *C. jejuni*

In New Zealand, *C. jejuni* multilocus sequence type 474 (ST-474) strain has been identified as an endemic sequence type (ST) that accounted for 24% to 34% of human clinical cases (French 2008, McTavish et al. 2008). Further, ST-474 has been found to be internationally rare, strongly associated with poultry and to be predominantly associated with urban areas (French 2008, Mullner et al. 2009, 2010). Two sub-strains of *C. jejuni* ST-474 with *fla*A SVR 14 antigenic allele (a short variable region within the *fla*A genes that are conserved among the *C. jejuni* virulent strains (Meinersmann et al. 1997)) were isolated from Palmerston North, New Zealand, one from a chicken and the other from a human clinical case during the same time period (French et al. 2009b). The genomes of these two isolates were sequenced by next generation Solexa sequencing technology at the Massey Genome Service, Palmerston North, New Zealand and their draft genomes have been submitted to GenBank[4] (Biggs et al. 2011). In spite of the fact that these two *C. jejuni* ST-474 *fla*A SVR 14 isolates were identical at both the clonal frame and antigenic level, there were 103 genes that revealed single nucleotide polymorphisms with non-synonymous substitutions in 72 genes (Biggs et al. 2011). The impact of recombination was found to be massive (95%) in comparison with mutation (5%) where, loci with non-homologous recombinations were evident in the human isolate (H22082) with an insertion of two extra genes, which was not seen in the chicken isolate (P110b) (Biggs et al. 2011).

Housekeeping alleles represent less than $<0.2\%$ of the entire genome (Dingle & Maiden 2005) and may not adequately reflect the phylogeny of every single gene, particularly ribosomal genes, genes involved in repair meachanisms and information processing systems in a genome. Moreover, the evolutionary potential of *C. jejuni* has been studied only within a subregion of seven housekeeping alleles used in the MLST scheme where the insights into the genetic stability of the genes involved in the above mentioned systems are obscure. Further, the amplifying effects of recombination and HGT diversifying two indistinguishable *C. jejuni* MLST genotypes (Biggs et al. 2011) has provided an evidence that such identical genotypes do differ when viewed both at individual gene level as well as from a whole genome perspective.

[4]URL:(`http://www.ncbi.nlm.nih.gov/genbank/`)

This study builds on the fact that *C. jejuni* is a polymorphic bacteria and the evidence for the molecular variation identified between two previously indistinguishable genotypes/strains (Biggs et al. 2011) to gain insight into the molecular variation among seven isolates with an identical MLST genotype (ST-474). Given the evidence that the housekeeping genes of *C. jejuni* undergo frequent recombination generating new variants (section 6.1.1), this study focuses on the genes of ribosomes, information processing and DNA repair system to investigate the vulnerability of these genes to evolutionary forces. DNA repair is a primary mechanism through which the genetic integrity of a genome is maintained. Molecular diversity among genes involved in repair pathways are possible during repair, where a failure to remove abnormalities and/or accommodation of abnormalities as reviewed earlier favour diversity and fitness. Several studies have shown incredible amounts of diversity in individual ribosomal and repair genes from different species and their diversified evolutionary behaviour under different experimental conditions (Puhler et al. 1989, Woese 1987, Olsen & Woese 1993, Wang et al. 1993, Guerry et al. 1994, Bustamante et al. 1995, Taylor & Chau 1997, Doolittle 1999, Eisen & Hanawalt 1999, Thomas et al. 1999, Zirnstein et al. 2000, de Boer et al. 2002, Dauga 2002, Hakanen et al. 2002, Hinode et al. 2002, McIver et al. 2004, Dionisi et al. 2004, Karenlampi et al. 2004, Matsuda et al. 2004, Klancnik et al. 2006, Adekambi et al. 2009, Liu et al. 2009, Said et al. 2010). However, a comparative study of ribosomal and/or information processing genes and/or DNA repair genes involving indistinguishable genotypes (designated so by a particular typing system) has to the authors knowledge not been conducted. Hence, a subset of ribosomal, information processing and DNA repair genes from seven *C. jejuni* ST-474 genomes were analysed in this study:

1. to better understand the genetic characteristics of these genes;
2. to analyse the evolutionary events that had occurred in these genes and;
3. to identify the ancestral lineages of individual ribosomal and repair genes.

This was done in order to infer the pattern of evolution in ST-474 genomes by integrating the information on phylogeny and ancestral states. This in turn will help to understand the similarities and differences among the seven ST-474 genomes within the investigated genes and the possible underlying causes for differential evolution in this set of genes.

A better understanding of DNA repair and ribosomal genes will provide insights in to the genetic stability of these genes in *C. jejuni* which is of paramount importance for evolution and the emergence of new variants.

6.2 Experimental procedures

The wealth of specialised software tools on molecular genetics has been exploited to draw inferences on parameters such as GC content, selection pressure (Tajima's D for whole genes and polymorphic sites and omega values for codons and sites within codons), codon usage, recombination events and the closest *C. jejuni* ancestor for the investigated genes of the ST-474 genome. Hence, this study does not involve exhaustive statistical modelling and/or computer simulations. Seven *C. jejuni* ST-474 genomes were fully sequenced at the Massey Genome Service, Palmerston North, New Zealand, using next generation Solexa sequencing technology.

6.2.1 Reference *C. jejuni* genomes

Twelve fully sequenced *C. jejuni* genomes were used as references for comparative analysis and to identify the closest ancestor of genes within ST-474 genomes. The gene sequences for the selected set of repair genes from all these twelve reference genomes were downloaded from the GenBank database.[5] Details of reference *C. jejuni* genomes used in this study are provided in Table 6.1.

[5]URL:(`http://www.ncbi.nlm.nih.gov/genbank/`)

Table 6.1: *Campylobacter jejuni subsp. jejuni* reference genomes and their MLST clonal complexes

Species	Strain	Size (Mb)	%GC	ORFs	Disease/source	Year and place of isolation	GenBank	MLST : ST (CC)
Campylobacter jejuni subsp. *jejuni*	NCTC 11168	1.64	30.5	1643	Clinical Food poisoning	1977, UK	AL111168	43 (21)
Campylobacter jejuni subsp. *jejuni*	RM1221	1.78	30.3	1838	Chicken	2005*, USA	CP000025	354 (354)
Campylobacter jejuni subsp. *jejuni*	81-176	1.6	30.6	1653	Clinical Food poisoning	1981, USA	CP000538	604 (42)
Campylobacter jejuni subsp. *jejuni*	81116	1.63	30.5	1626	Clinical Food poisoning	2007, UK	CP000814	267 (283)
Campylobacter jejuni subsp. *jejuni*	CG84-21	1.6	30.4	1512	Clinical Food poisoning	ns	ABGQ00000000	1919 (52)
Campylobacter jejuni subsp. *jejuni*	HB93-13	1.7	30.6	1710	Clinical GBS	2006*, USA	AANQ00000000	22 (22)
Campylobacter jejuni subsp. *jejuni*	CG84-86	1.65	30.4	1425	Clinical Food poisoning	ns	AASY00000000	2943 (574)
Campylobacter jejuni subsp. *jejuni*	CF93-6	1.67	30.5	1757	Clinical MFS	2006*, Japan	AANJ00000000	883 (21)
Campylobacter jejuni subsp. *jejuni*	84-25	1.67	30.4	1748	Clinical Meningitis	2006*, USA	AANT00000000	21 (21)
Campylobacter jejuni subsp. *jejuni*	260.94	1.65	30.5	1716	Clinical GBS	ns	AANK00000000	362 (362)
Campylobacter jejuni subsp. *jejuni*	IA3902	1.64	30.5	1718	Sheep abortion	2010, USA	CP001876	8 (21)
Campylobacter jejuni subsp.*doylei*	269.97	1.85	30.6	2094	Human blood	2007, UK	CP000768	1845 (UA)

MLST: Multilocus sequence type.
ST: Sequence type.
CC: Clonal complex.
GBS: Guillain-Barré syndrome.
ORFs: Open reading frames.
Mb: Mega basepair.
GC: Guanine:cytosine.
ns: not stated.
*:Date of start of project
U/A:Unassigned

6.2.2 Bacterial isolates, DNA preparation and sequencing

A total of seven *C. jejuni* isolates, four from poultry carcasses (P179a, P569a, P694a and P110b) and three from human clinical cases of campylobacteriosis (H22082, H73020 and H704) obtained between August, 2005 and February, 2009, were MLST typed at the Hopkirk Research Institute, IVABS, Massey University, Palmerston North and these isolates were found to belong to the ST-474 genotype. The genomic DNA from these seven isolates were extracted from pure bacterial colonies grown on blood agar plates (Fort Richard, Auckland, NZ) using a Wizard Genomic DNA Purification Kit (Promega) by following the manufacturer's instructions. Nebulisation of genomic DNA, sequencing reactions and short read sequence generation were carried out at the Massey Genome Service, Massey University, Palmerston North.

6.2.3 *de novo* genome assembly and gene prediction

Genome assembly and gene predictions were carried out at the Massey Genome Service by Dr. Patrick Biggs, the bioinformatician involved in the sequencing project of *C. jejuni*. The gene sequences of twenty five repair genes from seven *C. jejuni* genomes were retrieved from the seven ST-474 genome-gene predictions and were analysed in this study.

6.2.4 Selection of repair and ribosomal genes

Some of the repair and ribosomal genes that are used for routine MLST typing schemes for some bacterial species (Wertz et al. 2003, Christensen et al. 2004, Margos et al. 2008), and genes that are involved in different repair mechanisms such as base excision, mismatch excision, nucleotide excision, recombinational, branch migration, replication and other repair processes (Eisen & Hanawalt 1999), were identified. In addition, three genes that are indirectly involved in replication such as nucleotide biosynthesis were identified. From these strata, a total of 25 genes were randomly selected from those involved in the above repair mechanisms. Full length nucleotide sequences for the selected set of genes from *C. jejuni* NCTC11168 genome (GenBank accession no. NC_002163) were retrieved in FASTA format to be used as reference sequences to retrieve the respective gene sequences from the ST-474 genome-gene predictions. The list of genes that were analysed

in this study with gene names, gene ID, their function and the pathways are provided in Table 6.2.[6] The positions of the selected subset of repair, ribosomal and nucleotide metabolic genes are shown on the *C. jejuni* NCTC 11168(GenBank accession number NC_002163) circular genome in Figure 6.6.

[6]URL: (`http:://www.igs.cnrs-mrs.fr/mgdb-cgi/www_gene_catalog?rpr.ann.`)

Table 6.2: Genes of repair mechanism, their gene ID, name, function and the pathways they are involved in

Genes	Gene_ID	Name of the gene	Function	Pathway
*dna*E*	Cj0718	DNA polymerase III subunit alpha	It is a main replicative polymerase (Catalyses DNA-template-directed extension of the 3'- end of a DNA strand by one nucleotide at a time)	Purine metabolism, pyrimidine metabolism, metabolic pathways, DNA replication, mismatch repair and homologous recombination.
*gid*A*	Cj1188c	tRNA uridine 5-carboxymethylaminomethyl modification enzyme	Involved in the modification of the wobble third base in tRNAs (5-carboxymethylaminomethyl modification (mnm(5)s(2)U) of the wobble uridine base in some tRNAs; also known as a glucose-inhibited cell division protein A)	tRNA modification and regulation
*gua*A*	Cj1248	Probable GMP synthase (glutamine-hydrolysing)	Purine ribonucleotide biosynthesis	Purine metabolism
*gyr*A*	Cj1027c	DNA gyrase subunit A	Negatively supercoils closed circular double-stranded DNA or prevents from super-coiling that is deleterious to bacterial survival	Direct DNA repair mechanism
*gyr*B*	Cj0003	DNA gyrase subunit B	Negatively super-coils closed circular double-stranded DNA	Direct DNA repair mechanism
*lig*A	Cj0586	NAD-dependent DNA ligase	Catalyses the formation of phosphodiester linkages and is essential for DNA replication and repair of damaged DNA (between 5'-phosphoryl and 3'-hydroxyl groups in double-stranded DNA using NAD as a coenzyme and as the energy source for the reaction)	Direct DNA repair mechanism, DNA replication, base excision repair, nucleotide excision repair and mismatch repair
mfd	Cj1085c	Transcription-repair coupling factor	Prevents or corrects mutagenic nucleotides by removing them from DNA strands during replication	Nucleotide excision repair – DNA repair
*mut*S	Cj1052c	Recombination and DNA strand exchange inhibitor protein	Involved in blocking homologous recombination and inhibits DNA strand exchange (has ATPase activity stimulated by recombination intermediates)	Mismatch excision repair – DNA repair
*mut*Y*	Cj1620c	Probable A/G-specific adenine glycosylase	Corrects incorrectly paired bases during DNA replication and involved in recombinational repair	Base excision repair – DNA repair

Table 6.2 (continued)

Genes	Gene_ID	Name of the gene	Function	Pathway
ogt	Cj0836	Methylated-DNA–protein-cysteine methyltransferase	Direct DNA repair by alkylation reversal	Direct DNA repair
*pol*A	Cj0338c	DNA polymerase I	Has 3'-5' exonuclease, 5'-3' exonuclease and 5'-3'polymerase activities, primarily functions to fill gaps during DNA replication and repair	Replication and DNA repair
*pyr*C*	Cj0259	Dihydroorotase	Pyrimidine ribonucleotide biosynthesis: catalyses the formation of N-carbamoyl-L-aspartate from (S)-dihydroorotate in pyrimidine biosynthesis	Pyrimidine nucleotide biosynthesis
*pyr*G*	Cj0027c	CTP synthetase	Catalyses the ATP-dependent amination of UTP to CTP with either L-glutamine or ammonia as the source of nitrogen	Pyrimidine nucleotide biosynthesis
*rec*A*	Cj1673c	Recombinase A	Involved in recombinational repair of DNA damage. Catalyses the hydrolysis of ATP in the presence of single-stranded DNA, the ATP-dependent uptake of single-stranded DNA by duplex DNA, and the ATP-dependent hybridisation of homologous single-stranded DNAs	Recombinational repair
*rec*J*	Cj0028	Putative single-stranded-DNA-specific exonuclease	Synthesis and modification of macromolecules - DNA replication,restriction/modification, recombination and repair	Recombinational repair
*rec*N	Cj0642	Putative DNA repair protein	DNA repair protein	Recombinational repair
*rec*R	Cj1263	Recombination protein	Involved in a recombinational process of DNA repair	Recombinational repair
*rpl*B*	Cj1704c	50S ribosomal protein L2	One of the primary rRNA-binding proteins; required for association of the 30S and 50S subunits to form the 70S ribosome, for tRNA binding and peptide bond formation	Ribosomal gene
*rpo*B*	Cj0479	DNA-directed RNA polymerase subunit beta'	RNA synthesis, RNA modification and DNA transcription	RNA polymerase

Table 6.2 (continued)

Genes	Gene_ID	Name of the gene	Function	Pathway
*rpo*D*	Cj1001	RNA polymerase sigma factor	Sigma factors are initiation factors that promote the attachment of RNA polymerase to specific initiation sites and are then released; this is the primary sigma factor of bacteria	RNA polymerase
*ruv*A	Cj0799c	Holliday junction DNA helicase	Plays an essential role in ATP-dependent branch migration of the Holliday junction	Branch migration repair mechanism
ssb	Cj1071	Single-stranded DNA-binding protein	Binds to single stranded DNA and may facilitate the binding and interaction of other proteins to DNA	Other repair mechanisms
*uvr*A*	Cj0342c	Excinuclease ABC subunit A	*uvr*A is an ATPase and a DNA-binding protein. A damage recognition complex composed of two *uvr*A and two *uvr*B subunits scans DNA for abnormalities. When the presence of a on has been verified by *uvr*B, the *uvr*A molecules dissociate.	Nucleotide excision repair
*uvr*B	Cj0680c	Excinuclease ABC subunit B	The UvrABC repair system catalyzes the recognition and processing of DNA lesions. The beta-hairpin of the Uvr-B subunit is inserted between the strands, where it probes for the presence of a lesion.	Nucleotide excision repair
*xse*A	Cj0325	Exo-deoxyribonuclease VII large subunit	Bidirectionally degrades single-stranded DNA into large acid-insoluble oligonucleotides	Mismatch excision repair

*:Repair, ribosomal and nucleotide metabolic genes employed for MLST typing of other bacterial species.

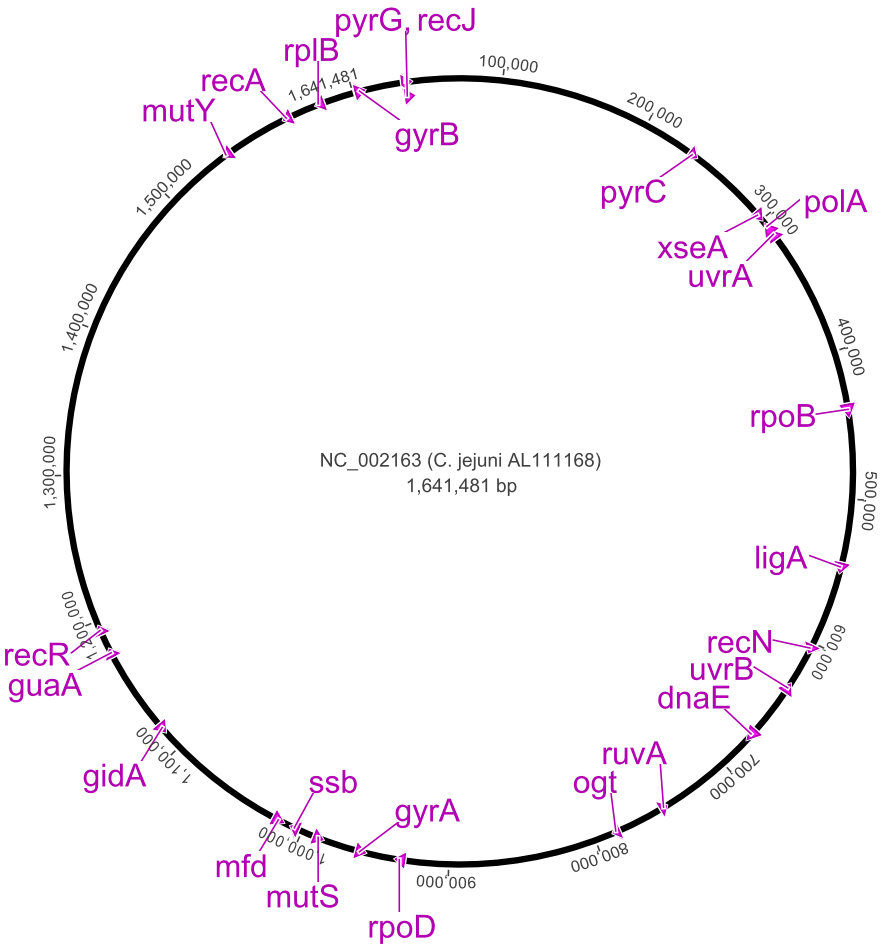


Figure 6.6: *C. jejuni* circular genome showing DNA repair genes. The positions of the repair, ribosomal and nucleotide metabolic genes are shown in this picture. The repair, ribosomal and nucleotide metabolic genes were identified and positioned on the *C. jejuni* NCTC 11168(GenBank accession number NC_002163) using Geneious molecular genetic software v5.3.4

6.2.5 Retrieval of gene sequences from *C. jejuni* ST-474 gene predictions

The BLAST+ application was downloaded from The National Center for Biotechnology Information (NCBI)[7] to perform stand-alone BLAST searches on the ST-474 gene predictions in order to find the gene/nucleotide sequences from the *C. jejuni* NCTC_11168 genome. For a given repair gene from NCTC_11168 genome, the BLAST hits with blast parameters such as maximum identity, alignment length and highest bitscore were selected. The respective open reading frames or the coding gene sequences were retrieved from the created BLAST db in FASTA format. For the reference genomes *C. jejuni doylei*

[7]URL: (`ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST`)

and CJCG84-86, CJCG8421, the orthologs for a few genes (*rpo*B for *C. jejuni doylei*; *mfd*, *pyr*C, *rec*N and *ruv*A for CJCG84-86; and *gyr*A for CJCG8421) could not be obtained from their genome sequences. Therefore, those missing gene sequences from above mentioned genomes could not be included in the comparative analysis along with other *C. jejuni* genomes.

6.2.6 Analysis of Guanine-Cytosine content, codon usage, selection pressure, recombination and evolutionary rates

The overall GC and GC3 contents of the selected subset of ribosomal, repair genes and nucleotide metabolic genes were compared (n = 25) using DnaSP v5 software (Librado & Rozas 2009). CBI and scaled chi square codon bias indices for individual genes from all *C. jejuni* genomes were also estimated using DnaSP v5.

Selection pressures operating on individual genes, codons and sites were analysed. To analyse the selection pressure on individual codons, methods developed by Muse & Gaut (1994) and Tamura & Nei (1993) that are available through the inbuilt functionality of HyPhy software package in MEGA v5 was used. A test statistic that gives the ratio of d_N to d_S (ω) was estimated using MEGA v5. dS denotes the number of synonymous substitutions per site (s/S) and dN denotes the number of non-synonymous substitutions per site (n/N) within a codon (Tamura & Dudley 2007). An overabundance of non-synonymous substitutions within a codon is indicated by a positive dN - dS value. In addition, the Tajima's D test was carried out to test the selection pressure on individual full length genes as well as non-synonymous sites over synonymous sites using DnaSP v5. Gene alignments from 19 genomes were used for estimating all the evolutionary measures.

Mean relative evolutionary rates were estimated using MEGA v5. The evolutionary rates have been scaled such that the average evolutionary rate across all sites in a given DNA sequence is assigned a value of 1.0 where, a site with a rate lower than 1.0 are considered to evolve slower than average, and those with a rate greater than one are regarded to evolve faster than average. The relative rates for all nucleotide sites, for all the 25 gene alignments across 19 genomes were estimated following the method of Jukes & Cantor (1969). Codons included in the estimation were 1st, 2nd, 3rd and the non-coding position.

Any nucleotide substitution at the second codon yields a 100% non-synonymous change while at the first codon position a synonymous nucleotide substitution can be expected. A discrete Gamma (+G) distribution was used to model evolutionary rate differences among sites (5 categories). The probability of classification of a site in each discrete rate category in Gamma is used where, all four nucleotide substitution (A, T, G and C) can be expected within a site of a codon (Jukes & Cantor 1969, Tamura & Dudley 2007).

DualBrothers (an advanced functionality that is available within Geneious v.5.3.4 (Minin et al. 2005)) was used to draw inferences on the recombination events within each gene under investigation. It uses a two change-point model to detect the spatial variation of the phylogenetic tree topology and nucleotide substitution variation (Minin et al. 2005). Nucleotide alignments from 19 genomes were used to detect a recombination event within individual genes. DnaSP v5 (Librado & Rozas 2009) was used to cross-check the sites involved in recombination along with Geneious and the specific sites of recombination were detected. R_m, the number of recombination sites was estimated using DnaSP v5. To understand the causal relationships between Rm, GC variance and the length, a linear model was used by having Rm as a dependent variable and GC variance and length as independent variables. The inferences were drawn based on the probability values obtained from the model.

6.2.7 Phylogenetic analysis

The nucleotide sequences (from the seven *C. jejuni* ST-474 and twelve reference genomes) were aligned using Geneious v5.3.4 (Drummond et al. 2010)[8]. Bootstrapped (500 replicates) maximum likelihood trees were generated using MEGA v5 (Tamura & Dudley 2007), and NeighbourNet and neighbour joining gene trees were generated using Splits Tree v4 (Huson & Bryant 2006) and Geneious v5.3.4 (Drummond et al. 2010), respectively. In turn, full length nucleotide sequences of 25 genes from ST-474 and the reference genomes were concatenated (genes arranged in an alphabetical order for all the 19 genomes) and the phylogenetic trees were constructed using the maximum likelihood method (MEGA), the NeighbourNet (SplitsTree) and neighbour joining methods (Geneious v5.3.4) separately. The topographies of the trees generated by these three

[8]URL:(www.biomatters.com/)

methods were compared to determine the closest *C. jejuni* ancestor of ST-474 genes. The ancestral lineages for all individual genes as well as concatenated gene sequences from ST-474 genomes were determined based on distance matrices that were produced in the process of phylogenetic tree construction.

6.3 Results

6.3.1 Length and nucleotide composition

The lengths of all 25 genes (repair and ribosomal) were identical across all the seven ST-474 genomes while five of the genes differed in their lengths amongst the reference genomes (Table 6.3). The *gyr*A, *mut*S, *rec*N, *rpo*B and *ssb* genes varied in their lengths where *rec*N was shorter by 18 nucleotides in the CJ260.94, CJJCF93-6 and CJJRM1221 genomes. Individual gene alignments showed that seven of 25 genes (*gid*A, *ogt*, *rec*J, *ssb*, *uvr*A, *uvr*B and *xse*A) differed in their nucleotide composition between the seven ST-474 genomes. The P694a ST-474 isolate showed nucleotide variations in four out of seven genes. The nucleotide substitutions were non-synonymous in *gid*A, *rec*J, *ssb* and *uvr*A genes while the substitutions in the remaining genes were synonymous.

Table 6.3: Gene lengths of repair genes : Bold fonts refer to genes with different lengths

Genes	*dna*E	*gid*A	*gua*A	*gyr*A	*gyr*B	*lig*A	*mfd*	*mut*S	*mut*Y	*ogt*	*pol*A	*pyr*C	*pyr*G	*rec*A	*rec*J	*rec*N	*rec*R	*rpl*B	*rpo*B	*rpo*D	*ruv*A	*ssb*	*uvr*A	*uvr*B	*xse*A
CJJ11168	3,603	1,860	1,536	**2,592**	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	4,137	1,869	552	552	2,826	1,974	1,164
CJJ260.94	3,603	1,860	1,536	2,589	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	**1,506**	573	831	**4,128**	1,869	552	**549**	2,826	1,974	1,164
CJJ81116	3,603	1,860	1,536	**2,592**	2,310	1,944	2,937	**2,208**	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	**4,128**	1,869	552	552	2,826	1,974	1,164
CJJ81-176	3,603	1,860	1,536	2,589	2,310	1,944	2,937	**2,208**	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	4,137	1,869	552	**549**	2,826	1,974	1,164
CJCG84-21	3,603	1,860	1,536	-	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	4,137	1,869	552	**549**	2,823	1,974	1,164
CJJ84-25	3,603	1,860	1,536	**2,592**	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	**4,128**	1,869	552	552	2,826	1,974	1,164
CJCG84-86	3,603	1,860	1,536	2,589	2,310	-	2,937	2,211	1,020	453	2,640	-	1,632	1,032	1,572	-	573	831	4,137	1,869	-	552	2,826	1,974	1,164
CJHB93-13	3,603	1,860	1,536	**2,592**	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	**4,128**	1,869	552	552	2,826	1,974	1,164
CJJCF93-6	3,603	1,860	1,536	**2,592**	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	**1,506**	573	831	**4,128**	1,869	552	552	2,826	1,974	1,164
CJD269.97	3,603	1,860	1,536	2,589	2,310	1,944	2,937	**2,208**	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	-	1,869	552	552	2,826	1,974	1,164
CJJIA3902	3,603	1,860	1,536	**2,592**	2,310	1,944	2,937	**2,208**	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	4,137	1,869	552	552	2,826	1,974	1,164
CJJRM1221	3,603	1,860	1,536	**2,592**	2,310	1,944	2,937	**2,208**	1,020	453	2,640	1,008	1,632	1,032	1,572	**1,506**	573	831	**4,128**	1,869	552	552	2,826	1,974	1,164
H22082 (ST-474)	3,603	1,860	1,536	2,589	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	**4,128**	1,869	552	552	2,826	1,974	1,164
H704 (ST-474)	3,603	1,860	1,536	2,589	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	**4,128**	1,869	552	552	2,826	1,974	1,164
P110 (ST-474)	3,603	1,860	1,536	2,589	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	**4,128**	1,869	552	552	2,826	1,974	1,164
P179a (ST-474)	3,603	1,860	1,536	2,589	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	**4,128**	1,869	552	552	2,826	1,974	1,164
P569a (ST-474)	3,603	1,860	1,536	2,589	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	**4,128**	1,869	552	552	2,826	1,974	1,164
P694a (ST-474)	3,603	1,860	1,536	2,589	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	**4,128**	1,869	552	552	2,826	1,974	1,164
H73020 (ST-474)	3,603	1,860	1,536	2,589	2,310	1,944	2,937	2,211	1,020	453	2,640	1,008	1,632	1,032	1,572	1,524	573	831	**4,128**	1,869	552	552	2,826	1,974	1,164

ST:Sequence type; CC:Clonal complex; U/A:Unassigned; Bold fonts refer to difference in the gene lengths
CJJ11168=ST-43:CC-21; CJJ260.94=ST-362:CC-362; CJJ81116=ST-267:CC-283; CJJ81-176=ST-604:CC-42; CJCG84-21=ST-1919:CC-52; CJJ84-25=ST-21:CC-21; CJCG84-86=ST-2943:CC-574; CJHB93-13=ST-22:CC-22; CJJCF93-6=ST-883:CC-21; CJD269.97=ST-1845:CC-U/A; CJJIA3902=ST-8:CC-21; CJJRM1221=ST-354:CC-354; ST-474:CC-48

6.3.2 Analysis of Guanine-Cytosine contents

The overall guanine-cytosine (GC) content varied greatly between genes in general, where genes such as *mfd*, *ogt*, *pol*A, *rec*J, *rec*N and *xse*A showed lower ranges of GC contents (from 0.265 to 0.298) whereas the remainder showed a relatively higher GC content with *rpl*B showing the highest GC content of 0.392. The overall GC and GC3 contents of individual ribosomal and repair genes across all genomes (the reference and ST-474 genomes) are provided in Table D.1 in Appendix D. CJJ93-6 was found to be closer to the ST-474 genomes that shared identical overall GC contents with 12 of 25 genes and identical GC3 values with 12 of 25 genes. Table 6.4 shows the number of genes from all the 12 reference genomes that shared identical GC and GC3 values with the ST-474 genomes. The distribution of GC contents within the ribosomal, repair and nucleotide metabolic genes across all the *C. jejuni* genomes (the reference and ST-474 genomes) are shown in Figures 6.7, 6.8, 6.9 and 6.10.

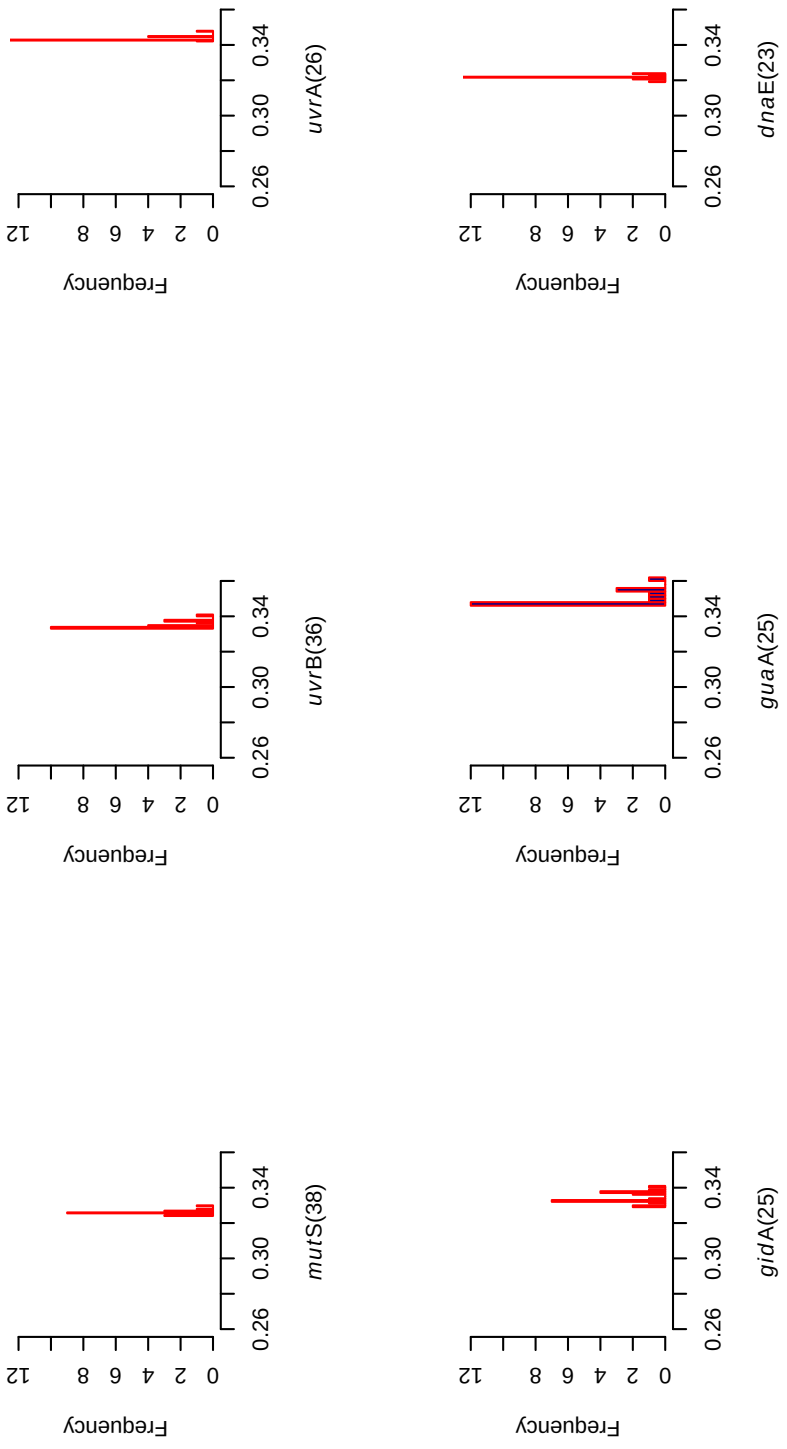


Figure 6.7: Guanine-cytosine (GC) distribution in the ribosomal and repair genes – A. This plot describes the GC distribution across all 19 *C. jejuni* genomes. The number of recombination sites are provided in parentheses.

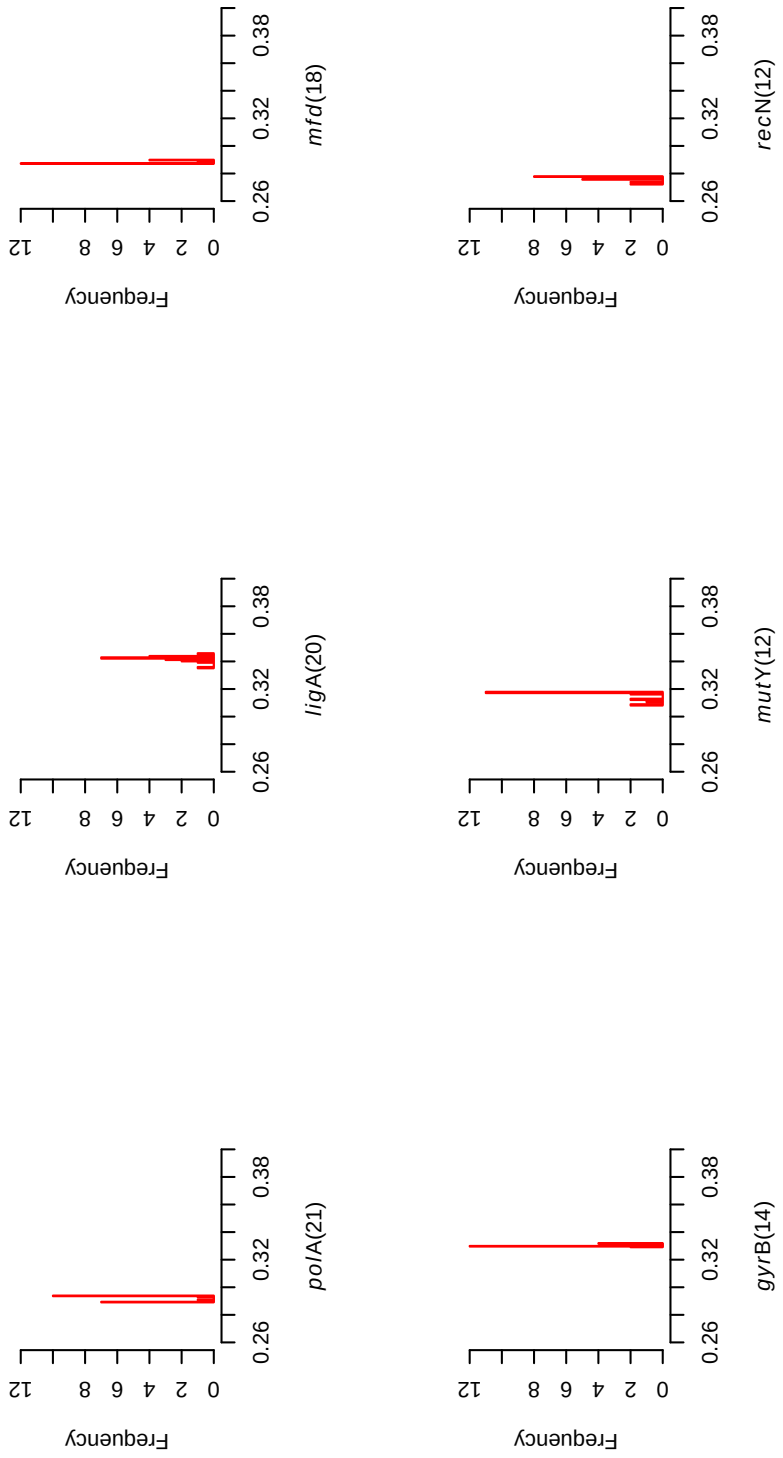


Figure 6.8: Guanine-cytosine (GC) distribution in the ribosomal and repair genes – B. This plot describes the GC distribution across all 19 *C. jejuni* genomes. The number of recombination sites are provided in parentheses.

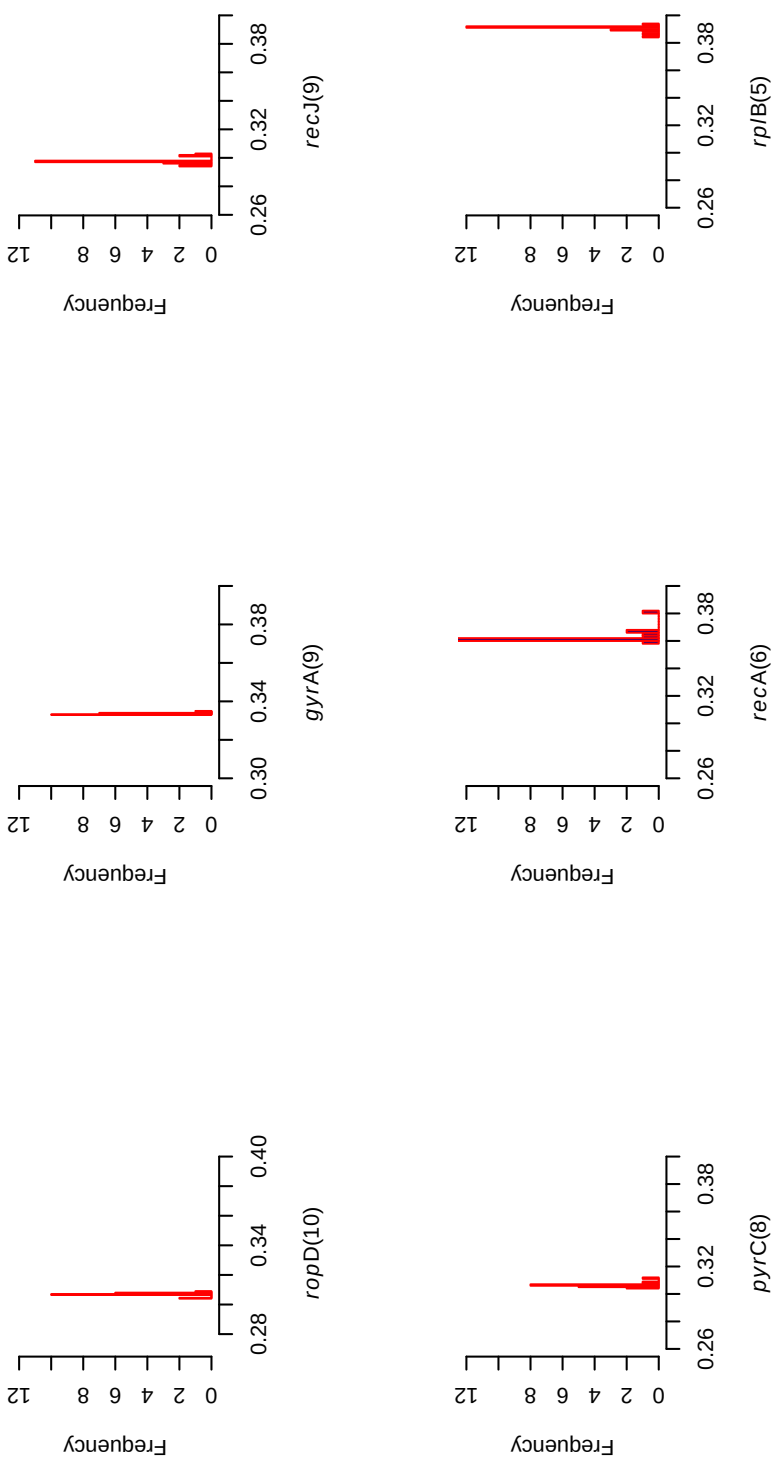


Figure 6.9: Guanine-cytosine (GC) distribution in the ribosomal and repair genes – C. This plot describes the GC distribution across all 19 *C. jejuni* genomes. The number of recombination sites are provided in parentheses.

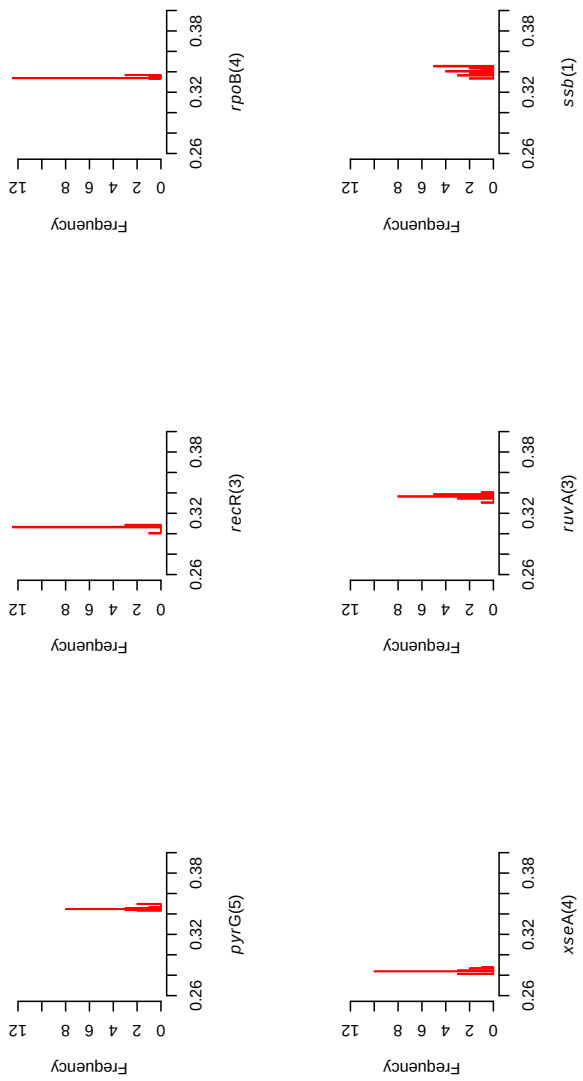


Figure 6.10: Guanine-cytosine (GC) distribution in the ribosomal and repair genes – A. This plot describes the GC distribution across all 19 *C. jejuni* genomes. The number of recombination sites are provided in parentheses.

Table 6.4: Ancestry shared between reference and ST-474 genomes based on GC and GC3 contents

Measures	CJJ11168 (ST-43:CC-21)	CJJ260.94 (ST-362:CC-362)	CJ81116 (ST-267:CC-283)	CJJ81-176 (ST-604:CC-42)	CJJCG84-21 (ST-1919:CC-52)	CJJ84-25 (ST-21:CC-21)	CJJCG84-86 (ST-2943:CC-574)	CJJHB93-13 (ST-22:CC-22)	CJJCF93-6 (ST-883:CC-21)	CJD269.97 (ST-1845:CC-U/A)	CJJIA3902 (ST-8:CC-21)	CJJRM1221 (ST-354:CC-354)
	n	n	n	n	n	n	n	n	n	n	n	n
GC	10	6	2	6	3	13	1	7	12	7	11	11
GC3	11	3	1	2	3	11	3	2	12	1	11	7

n:Number of genes that share identical GC and GC3 contents between each reference and the ST-474 genomes
GC:Guanine-cytosine content
GC3:Guanine-cytosine content at the third position
ST:Sequence type
CC:Clonal complex
U/A:Unassigned

6.3.3 Codon usage bias

The overall codon usage bias index (CBI) estimates showed that *gid*A, *ogt*, *rec*J, *ssb*, *uvr*A, *uvr*B and *xse*A genes varied between all seven of the ST-474 genomes. Amongst the poultry isolates, P110b and P179a isolates shared an identical overall CBI, while P179a differed by one gene (*xse*A) in its scaled chi square bias while the remainder differed both in the CBI and scaled chi square indices (Table 6.5). The human isolates showed identical CBI and scaled chi square indices in two of seven genes while the remainder varied between the two isolates. Table 6.6 shows the number of genes that shared identical CBI and scaled chi square indices with each of the reference and seven ST-474 genomes, where the genes that differed between ST-474 genomes shared identical values with one of the reference genomes and further details on individual genes and their CBI and scaled chi square indices are provided in Table D.3, D.4 in Appendix D, respectively. The *uvr*A did not show any differences in the CBI and scaled chi square indices. While comparing the reference and ST-474 genomes based on the codon usage indices, CJJIA3902 was found to be closer to ST-474 genomes that shared eight genes with identical CBI values with ST-474 genomes.

Table 6.5: CBI and scaled chi square similarities shared amongst ST-474 genomes within the repair and ribosomal genes that differ

Genes	P110b		P179a		P569a		P694a		H73020		H22082		H704	
	CBI	Schisq	CBI	Schisq	CBI	Schisq	CBI	Schisq	CBI	Schisq	CBI	Schisq	CBI	Schisq
*gid*A	a	**a**	a	**a**	a	**a**	b	**b**	a	**a**	a	**a**	a	**a**
ogt	a	**a**	a	**a**	b	**b**	c	**a**	a	**a**	b	**b**	b	**b**
*rec*J	a	**a**	a	**a**	b	**a**	a	**a**	a	**a**	a	**a**	a	**b**
ssb	a	**a**	a	**a**	a	**a**	b	**b**	a	**a**	b	**b**	a	**a**
*uvr*B	a	**a**	a	**a**	a	**a**	a	**a**	b	**b**	a	**a**	a	**a**
*xse*A	a	**a**	a	**b**	a	**a**	b	**a**	b	**b**	b	**b**	b	**b**

The scale of similarity index has been assigned a, b and c. Reading across the rows of the table the isolates with identical index have identical codon usage bias index (CBI) and scaled chi square codon usage index (Schisq) whereas those with a different index differ in their CBI and Schisq.

Table 6.6: Ancestry shared between the reference and ST-474 genomes according to codon usage:

Measures	CJJ11168 (ST-43:CC-21)	CJJ260.94 (ST-362:CC-362)	CJ81116 (ST-267:CC-283)	CJJ81-176 (ST-604:CC-42)	CJJCG84-21 (ST-1919:CC-52)	CJJ84-25 (ST-21:CC-21)	CJJCG84-86 (ST-2943:CC-574)	CJJHB93-13 (ST-22:CC-22)	CJJCF93-6 (ST-883:CC-21)	CJD269.97 (ST-1845:CC-U/A)	CJJIA3902 (ST-8:CC-21)	CJJRM1221 (ST-354:CC-354)
	n	n	n	n	n	n	n	n	n	n	n	n
CBI	7	1	2	2	1	7	2	0	7	0	8	6
Schisq	7	2	0	2	0	7	1	0	5	0	7	5

n:Number of genes that share identical CBI and Scaled chi square indices between each reference and the ST-474 genomes
CBI:Codon usage bias index; Schisq = Scaled chi square index
ST:Sequence type
CC:Clonal complex
U/A:Unassigned

6.3.4 Selection pressure and evolutionary rate:

Both ribosomal and repair genes were found to be under stringent purifying selection showing a negative Tajima's D value for individual full length genes (Table 6.7). However, the Tajima's D values for the non-synonymous sites over synonymous sites (NonSyn/Syn) were positive for the majority of the genes while some of the genes showed negative values. The *rpo*B gene showed the highest Tajima's D value of 36.9 compared with all of the genes investigated in this study. The NonSyn/Syn–Tajima D values of *gyr*B, *rec*N and *ssb* genes were negative which indicated an ongoing purifying selection pressure.

Table 6.7: Tajima's D values of the ribosomal, repair genes and nucleotide metabolic genes

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*dna*E	170	34	-1.37094	1.13669
*gid*A	104	26	-0.47209	1.14795
*gua*A	103	24	-0.95402	0.10576
*gyr*A	77	16	-0.50676	1.88165
*gyr*B	72	14	-0.4225	-0.42175
*lig*A	106	42	-1.28959	1.5028
mfd	126	32	-0.42339	0.88156
*mut*S	118	87	-0.80472	2.51068
*mut*Y	41	26	-0.94926	0.94322
ogt	17	11	-1.72611	1.36041
*pol*A	162	45	-0.2055	0.56685
*pyr*C	60	26	-1.40423	0.96958
*pyr*G	62	12	-1.72779	0.76546
*rec*A	49	6	-1.46434	1.54809
*rec*J	42	9	-0.71025	0.28068
*rec*N	56	21	-0.11234	-49.2621
*rec*R	14	6	-0.94561	1.38957
*rpl*B	35	5	-1.01716	1.7962
*rpo*B	43	11	-0.03347	36.90874
*rpo*D	69	11	-1.3889	1.48297
*ruv*A	19	6	-0.86384	0.61158
ssb	17	7	-0.87871	-0.17492
*uvr*A	180	28	-1.21604	0.93944
*uvr*B	172	29	-0.90449	2.60546
*xse*A	48	22	-1.61251	1.4809

TD_gene:Tajima's D value of full length gene
TD_nonsyn/syn:Tajima's D value for the non-synonymous sites over synonymous sites

Selection pressure on individual codons of all 25 genes was estimated. A total of 538 codons from 25 genes were found to be non-synonymous codons of which the ω values (test statistic of d_N-d_S) for 445 codons indicated a positive selection pressure. The over abundance of non-synonymous substitutions in these codons indicate a functional advantage, however this can only be confirmed by suitable genetic and biochemical tests. There was an indication of an ongoing purifying selection in the remaining codons (n = 95) that showed negative ω values. Table 6.8 provides the number of non-synonymous codons and the number of non-synonymous codons that are under purifying selection which were consolidates based on the ω values from the total number of codons of all the 25 genes.

Table 6.8: Number of non-synonymous codons in the repair genes under purifying selection based on ω values

Gene	Non-syn_codons	Non-syn_codons under purifying selection
*dna*E	34	6
*gid*A	26	5
*gua*A	22	3
*gyr*A	14	1
*gyr*B	14	2
*lig*A	39	10
mfd	32	2
*mut*S	73	10
*mut*Y	24	3
ogt	14	4
*pol*A	46	13
*pyr*C	27	5
*pyr*G	13	1
*rec*A	6	0
*rec*J	9	0
*rec*N	21	4
*rec*R	6	0
*rpl*B	5	1
*rpo*B	11	0
*rpo*D	10	0
*ruv*A	6	1
ssb	7	0
*uvr*A	23	5
*uvr*B	34	14
*xse*A	22	2

Non-syn_codons:Non-synonymous codons

The average of mean relative evolutionary rates of individual sites of all the 25 genes was estimated and it was found that both the synonymous and non-synonymous polymorphic sites evolve faster than that of the non-polymorphic sites. The mean evolutionary rates of all the polymorphic sites were greater than 1 which ranged from 2.1 to 5.6 (Table 6.9).

The genes *uvr*A, *mut*S and *gyr*A were identified to possess an insertion of a rare amino acid in the reference and ST-474 genomes, namely the selenocysteine when an analysis of codons was carried out using DnaSP software. However, these codons may also represent premature stop codons (Wong et al. 2008) which need further gene expression and protein analyses.

Table 6.9: Average of mean relative evolutionary rates of non-synonymous codons of repair, ribosomal and nucleotide metabolic genes. Average of evolutionary rates estimated from the gene sequence alignments of the ST-474 and reference genomes.

Genes	Non-syn. codons – evolutionary rates
*dna*E	5.096
*gid*A	5.655
*gua*A	4.991
*gyr*A	5.029
*gyr*B	5.281
*lig*A	5.157
mfd	5.051
*mut*S	5.452
*mut*Y	5.139
ogt	3.331
*pol*A	4.453
*pyr*C	2.576
*pyr*G	5.034
*rec*A	4.989
*rec*J	5.039
*rec*N	4.996
*rec*R	5.053
*rpl*B	5.102
*rpo*B	4.974
*rpo*D	4.983
*ruv*A	5.011
ssb	2.121
*uvr*A	5.010
*uvr*B	5.224
*xse*A	3.037

Nonsyn. codons:Non-synonymous codons

Relationship between GC variance and recombination

Analysis of ribosomal and repair genes for recombination events showed that *ogt* gene and *ssb* possessed a single recombination site, while *mut*S was found to possess the highest number of recombination sites (n = 38). Details on sites involved in recombination and the number of sites are provided in Table 6.10.

The linear model with Rm (the number of recombination sites) as a dependent variable and the GC variance and length as independent variables showed that Rm was dependent on the length of the gene (p value = 0.05) rather than on the GC variance (p value = 0.9). Recombination sites were measured using DnaSP v5 and Dual brothers (Geneious v.5.3.4) where, *ogt* and *ssb* showed a single recombination site while *uvr*B (n = 36) and *mut*S (n = 38) showed the highest numbers of recombination sites. Figures 6.11 a and b illustrate the relationship of Rm on the length of genes and the relationship between the GC variance and Rm.

Table 6.10: Recombination sites identified within ribosomal and repair genes

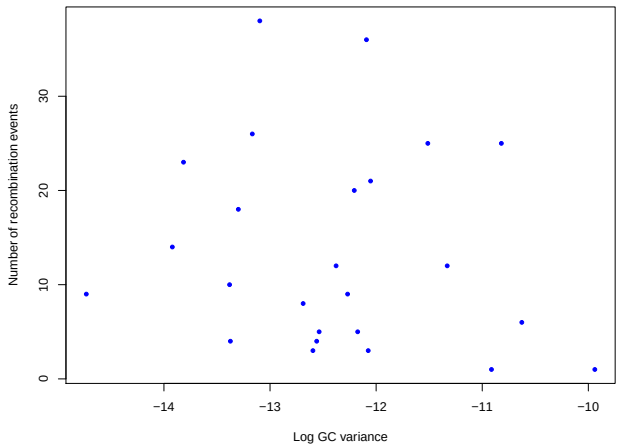
Genes	Sites	Rm
*dna*E	(210-376) (376-396) (396-501) (714-762) (828-917) (963-981) (981-1023) (1131-1347) (1347-1449) (1890-1923) (1992-2028) (2037-2048) (2048-2049) (2049-2217) (2226-2409) (2409-2679) (2688-2778) (2778-3016) (3177-3240) (3240-3285) (3285-3303) (3303-3423) (3426-3535)	23
*gid*A	(345-365) (365-399) (419-441) (444-546) (624-669) (669-687) (687-768) (768-771) (771-858) (912-930) (939-942) (942-955) (955-959) (959-1020) (1020-1127) (1127-1143) (1158-1233) (1233-1245) (1245-1281) (1281-1347) (1347-1389) (1446-1577) (1578-1585) (1602-1611) (1635-1746)	25
*gua*A	(132-174) (210-223) (226-232) (232-331) (331-351) (351-378) (402-471) (489-496) (496-516) (570-582) (663-687) (687-723) (723-765) (765-801) (825-858) (879-960) (960-1116) (1116-1140) (1140-1179) (1179-1230) (1230-1254) (1332-1392) (1398-1410) (1410-1428) (1482-1521)	25
*gyr*A	(243-257) (257-357) (891-1017) (1017-1476) (1770-1994) (1994-2142) (2283-2421) (2472-2478) (2478-2547)	9
*gyr*B	(165-324) (324-360) (360-375) (375-813) (852-924) (952-1053) (1053-1107) (1113-1143) (1167-1209) (1209-1360) (1449-1551) (1570-1662) (1662-1728) (2169- 2277)	14

Table 6.10 (continued)

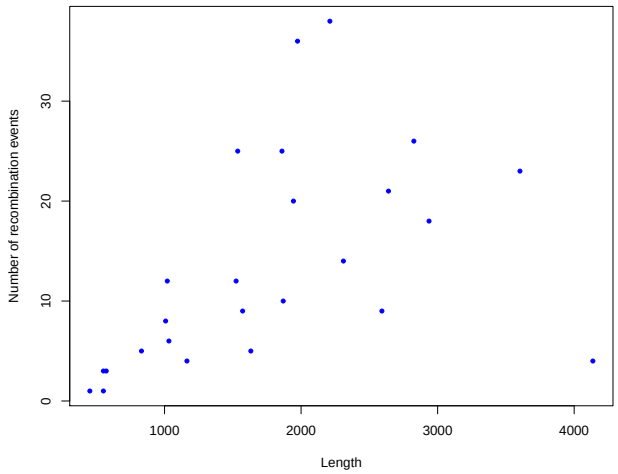
Genes	Sites	Rm
*lig*A	(125-188) (188- 558) (558- 579) (681- 708) (708- 720) (734-787) (819- 832) (832-1032) (1050-1104) (1104-1245) (1332-1371) (1371-1431) (1449-1500) (1500-1539) (1561-1587) (1587-1649) (1665-1717) (1743-1752) (1752-1815) (1815-1878)	20
mfd	(25- 87) (87-174) (195- 291) (465- 531) (693- 729) (729- 867) (867- 876) (888- 919) (936-1071) (1071-1185) (1350-1371) (1372-1410) (1464-1506) (1506-1512) (1536-1554) (1554-1767) (1767- 2070) (2080- 2713)	18
*mut*S	(120-132) (198-210) (210-225) (225-316) (339-372) (393-421) (421-423) (435-474) (474-489) (489-591) (599- 636) (636- 729) (774- 823) (843- 849) (852-867) (867-876) (1026-1028) (1028-1107) (1134-1243) (1243-1266) (1269-1305) (1305-1326) (1339-1365) (1497-1548) (1548-1596) (1596-1626) (1626-1680) (1710-1869) (1869-1914) (1975-1978) (1978-1983) (1984-1996) (1996- 2014) (2071-2077) (2077-2089) (2095-2116) (2152-2185) (2185-2203)	38
*mut*Y	(75-141) (141-261) (294-327) (327-336) (336-351) (363-393) (438-501) (501-553) (553-564) (613-670) (670-715) (766-847)	12
ogt	(75-249)	1
*pol*A	(21-142) (240-274) (274-327) (369-540) (600-649) (649-672) (1028-1083) (1083-1121) (1253-1278) (1278-1305) (1305-1347) (1347-1368) (1380-1464) (1578-1611) (1683-1720) (1721-1743) (1747-1860) (1893-1911) (2256-2262) (2472-2487) (2487-2523)	21
*pyr*C	(5-246) (252-312) (312-396) (396-402) (402-723) (723-747) (747-975) (975-987)	8
*pyr*G	(915-945) (945-1018) (1146-1392) (1419-1445) (1445-1479)	5
*rec*A	(226-351) (447-570) (570-612) (612-624) (807-825) (861-912)	6
*rec*J	(250-300) (318-355) (355-375) (678-688) (688-972) (972-1059) (1059-1062) (1062-1245) (1245-1530)	9
*rec*N	(87-111) (111-192) (243-396) (399-421) (421-539) (540-738) (759-834) (834-870) (870-879) (900-1068) (1257-1272) (1272-1311)	12
*rec*R	(78-123) (282-432) (432-447)	3
*rpl*B	(123-228) (240-258) (384-444) (468-588) (675-762)	5
*rpo*B	(606-1263) (1791-1812) (3363-3366) (3366-3417)	4

Table 6.10 (continued)

Genes	Sites	Rm
*rpo*D	(129-156) (156-174) (174-435) (495-735) (735-822) (822-858) (1272-1326) (1326-1758)	10
*ruv*A	(96-123) (147-264) (264-387)	3
ssb	(234- 448)	1
*uvr*A	(27-105) (105-132) (132-177) (222-243) (255-291) (291-294) (351-372) (372-384) (498-540) (579-621) (651-681) (681-804) (921-1031) (1101-1170) (1170-1255) (1326-1425) (1425-1671) (2058-2094) (2094-2139) (2202-2276) (2276-2397) (2433-2442) (2442-2532) (2610-2628) (2643-2700) (2700-2715)	26
*uvr*B	(36-72) (120-165) (165-168) (327-378) (420-594) (696-714) (714-751) (757-777) (777-784) (816-856) (876-924) (924-928) (936-948) (948-988) (988-1005) (1042-1068) (1068-1074) (1083-1086) (1086-1122) (1122-1182) (1182-1221) (1221-1248) (1248-1266) (1272-1284) (1284-1356) (1362-1383) (1437-1530) (1530-1533) (1533-1534) (1536-1593) (1701-1776) (1779-1785) (1785-1791) (1791-1866) (1881-1890) (1914-1938)	36
*xse*A	(90-186) (402-669) (832-930) (930-961)	4



(a) Relationship between the GC variance and the number of recombination sites in the repair genes



(b) Relationship between the length of genes and the number of recombination sites in the repair genes

Figure 6.11: Relationship between the GC variance, length and the recombination sites in the repair genes. This plot describes the relationship between the GC variance within individual genes and the number of recombination sites that had occurred within each gene and the relationship between the length of genes and the number of recombination sites.

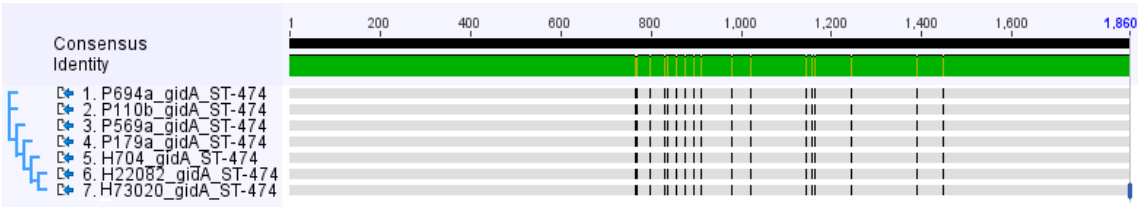
6.3.5 Phylogenetic analysis

Phylogenetic tress were constructed using the maximum likelihood method in MEGA, the NeighbourNet method in SplitsTree and the neighbour joining method in Geneious with a bootstrap replication of 500, for all the 25 individual genes investigated in this study. The tree topographies generated by these three methods were compared and were found to be identical and hence, the maximum likelihood trees of individual gene trees are presented in this manuscript. The maximum likelihood genes trees of those genes that differed between the seven ST-474 genomes are presented in this section and the individual gene trees of the remainder are provided in Figures D.1, D.2, D.3 and D.4 in Appendix D.

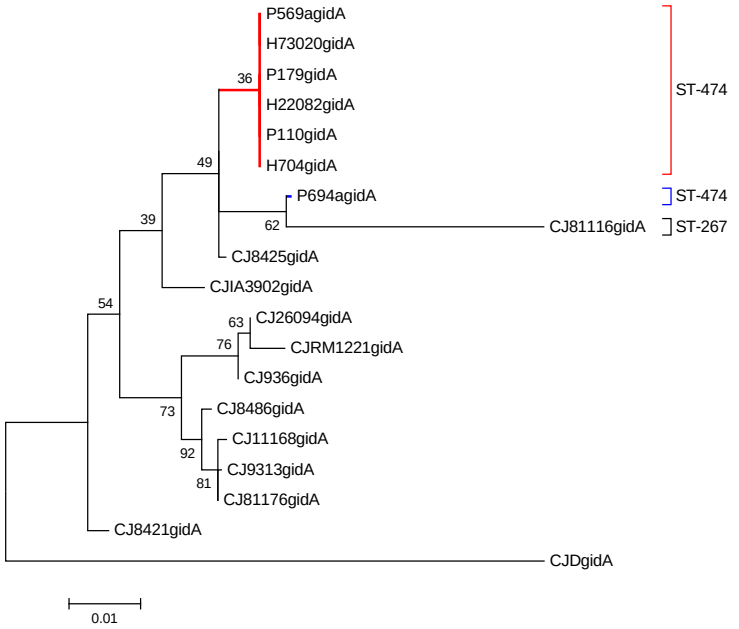
Isolates belonging to the clonal complex (CC) 21 represented the closest ancestors to the majority of the genes in ST-474; CJJ11168 was the closest ancestor for most of the genes, followed by CJJ84-25, CJJIA3902 and CJJCF93-6, while some genes differed in their ancestral lineage. However, there was a slight deviation of ST-474 ancestry found in some of the genes such as *mfd*, *pol*A, *rpo*D and *rec*J. Generally, the genes that differed within the seven ST-474 genomes were found to vary in their ancestry. For example, the *gid*A and the *ogt* genes of P694a isolate formed separate operational taxonomic units on individual gene trees (Figures 6.12 and 6.13); with *gid*A sharing its ancestry with the CJJ81116 genome (ST-283 complex) while *ogt* shared its ancestry with the ST-21 complex. In contrast, the *ssb* genes of the ST-474 genomes formed two separate clusters. The P694a and H22082 isolates formed one cluster that was closer to the CJJ84-25 genome whereas, the remaining ST-474 genomes formed another cluster that was closer to the CJJ11168 genome (Figure 6.14). However, it should be noted that both of these reference genomes belong to CC-21. The *xse*A gene showed another interesting phylogenetic relationship where P569a, P110b and P179a were identical and clustered separately on the gene tree (Figure 6.15) but the ancestor was common to all of the ST-474 genomes. Even though the *rec*J, *uvr*A and *uvr*B genes differed amongst the ST-474 genomes, the ancestor was common to all of the seven ST-474 genomes with subtle differences in the branch lengths of the corresponding ST-474 genome (Figures 6.16), 6.17 and 6.18).

The maximum likelihood tree of the concatenated gene sequences (ribosomal, repair and nucleotide metabolic genes) from all the 19 *C. jejuni* genomes identified the members of CC-21 to be the closest relatives of the ST-474 genomes, namely the CJJ11168, CJJ84-25,

CJJIA3902 and CJJCF93-6. The P110b, P179a, P569a and H73020 isolates were found to be identical, while the remainder were different which is illustrated in a maximum likelihood tree in Figure 6.19 and a NeighbourNet tree in Figure 6.20.

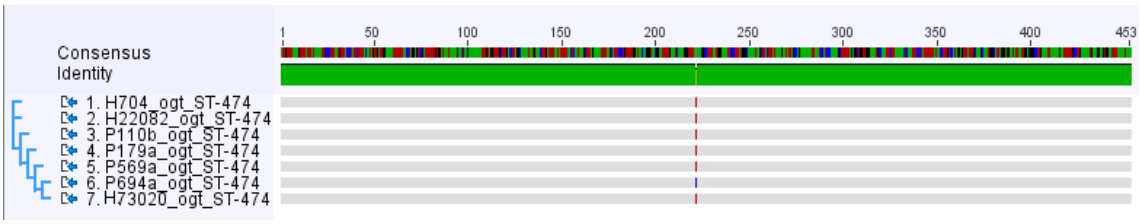


(a) Nucleotide alignment of *gid*A. This figure shows the identity bar (green bar above the sequences) and, the perpendicular small lines within the sequences refer to nucleotide differences or changes. The blue tree beside the isolate names illustrate the phylogenetic relationship among the seven strains

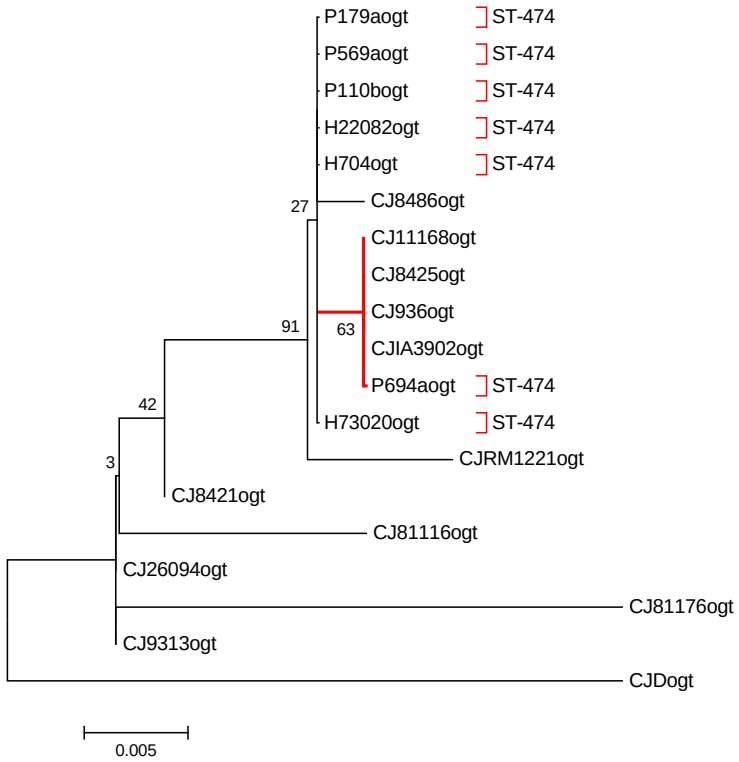


(b) Maximum likelihood tree of *gid*A

Figure 6.12: Maximum likelihood gene tree of *gid*A. Alignment of *gid*A gene from the seven *C. jejuni* ST-474 genomes showing multiple nucleotide changes. The numbers refer to the bootstrap values corresponding to each branch of the phylogenetic tree. ST refers to sequence type.

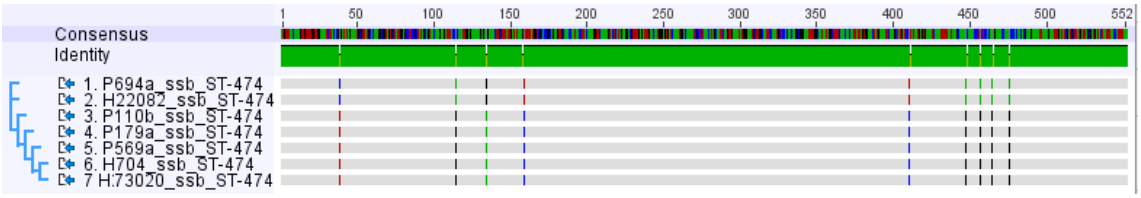


(a) Nucleotide alignment of *ogt*. This figure shows the identity bar (green bar above the sequences) and, the perpendicular small lines within the sequences refer to nucleotide differences or changes; different colours represent different nucleotides. The blue tree beside the isolate names illustrate the phylogenetic relationship among the seven strains

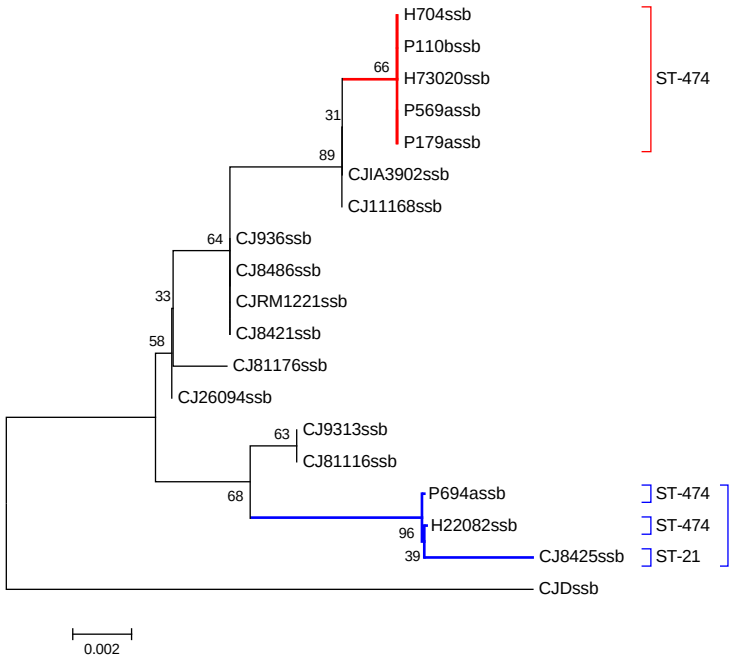


(b) Maximum likelihood tree of *ogt*

Figure 6.13: Maximum likelihood gene tree of *ogt*. Alignment of *ogt* gene from the seven *C. jejuni* ST-474 genomes showing a single nucleotide change. The numbers refer to the bootstrap values corresponding to each branch of the phylogenetic tree. ST refers to sequence type.

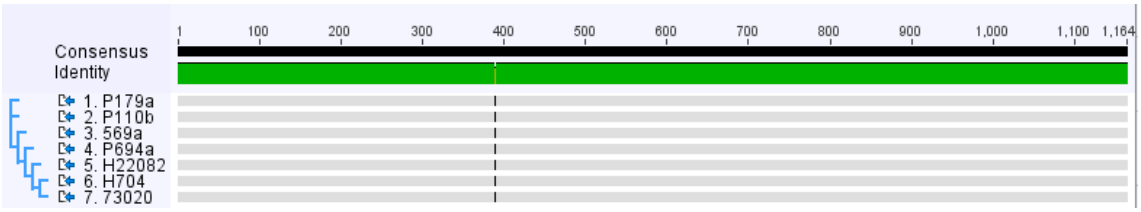


(a) Nucleotide alignment of *ssb*. This figure shows the identity bar (green bar above the sequences) and, the perpendicular small lines within the sequences refer to nucleotide differences or changes; different colours represent different nucleotides. The blue tree beside the isolate names illustrate the phylogenetic relationship among the seven strains

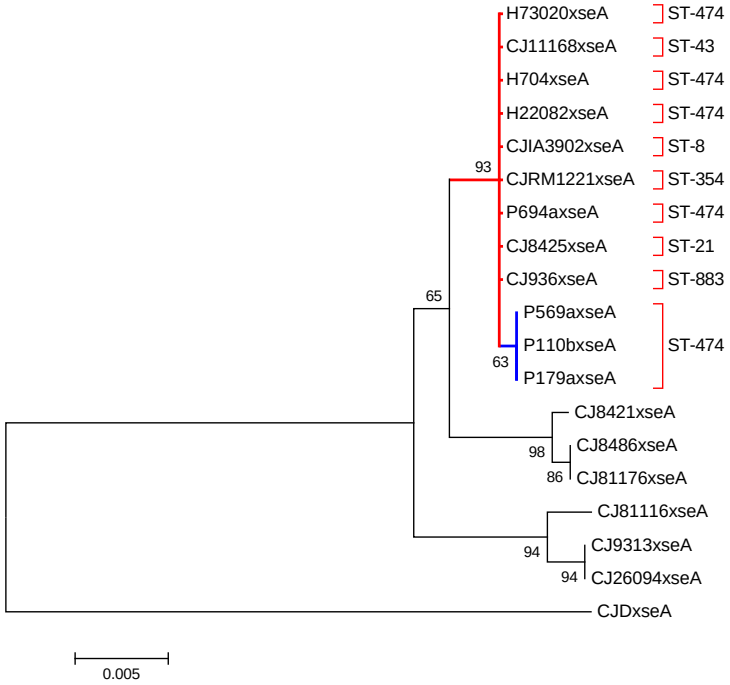


(b) Maximum likelihood tree of *ssb*

Figure 6.14: Maximum likelihood gene tree of *ssb*. Alignment of *ssb* gene from the seven *C. jejuni* ST-474 genomes showing multiple nucleotide changes. The numbers refer to the bootstrap values corresponding to each branch of the phylogenetic tree. ST refers to sequence type.

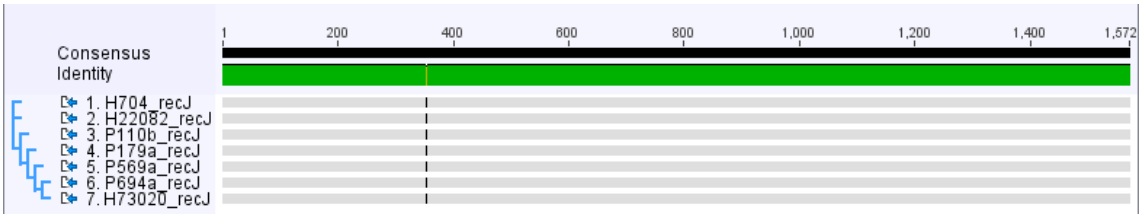


(a) Nucleotide alignment of *xse*A. This figure shows the identity bar (green bar above the sequences) and, the perpendicular small lines within the sequences refer to nucleotide differences or changes. The blue tree beside the isolate names illustrate the phylogenetic relationship among the seven strains.

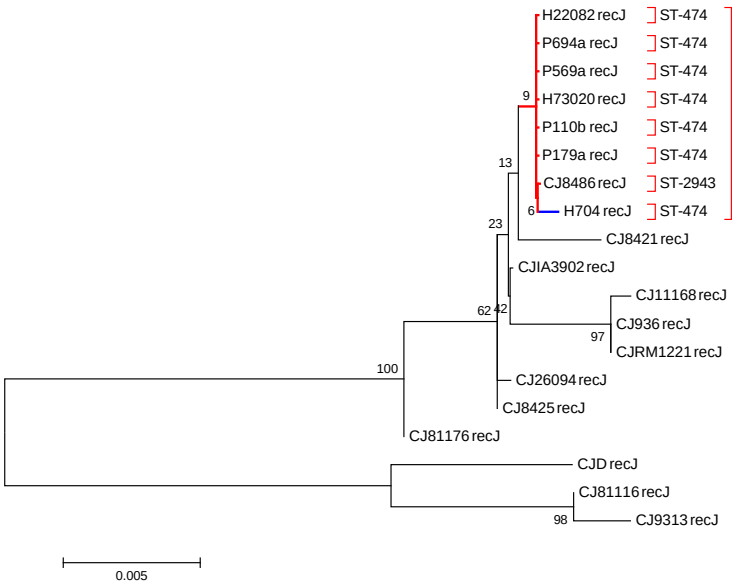


(b) Maximum likelihood tree of *xse*A

Figure 6.15: Maximum likelihood gene tree of *xse*A. Alignment of *xse*A gene from the seven *C. jejuni* ST-474 genomes showing a single nucleotide change. The numbers refer to the bootstrap values corresponding to each branch of the phylogenetic tree. ST refers to sequence type.

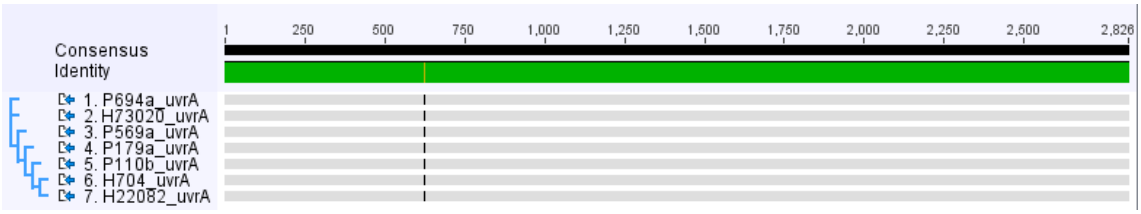


(a) Nucleotide alignment of *rec*J. This figure shows the identity bar (green bar above the sequences) and the perpendicular small lines within the sequences refer to nucleotide differences or changes; different colours represent different nucleotides. The blue tree beside the isolate names illustrate the phylogenetic relationship among the seven strains.

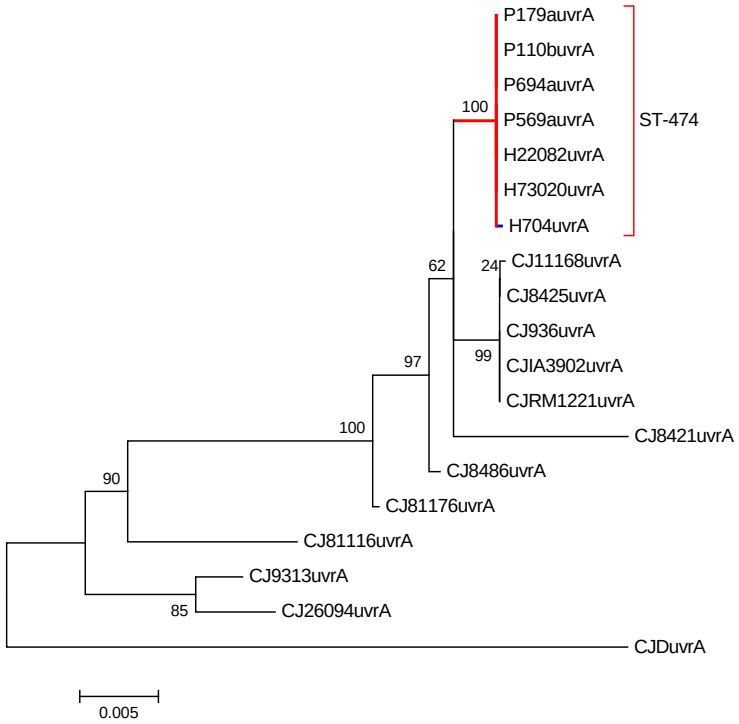


(b) Maximum likelihood tree of *rec*J

Figure 6.16: Maximum likelihood gene tree of *rec*J. Alignment of *rec*J gene from the seven *C. jejuni* ST-474 genomes showing a single nucleotide change. The numbers refer to the bootstrap values corresponding to each branch of the phylogenetic tree. ST refers to sequence type.

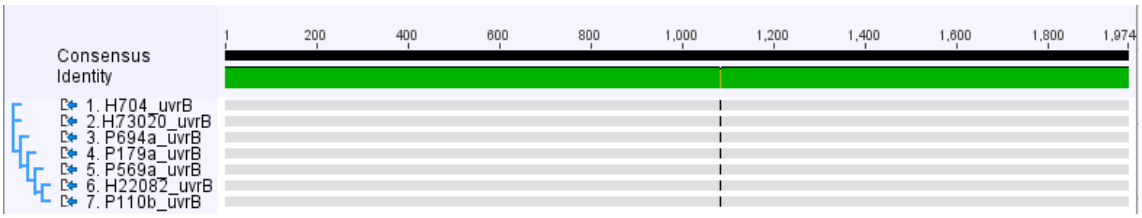


(a) Nucleotide alignment of *uvr*A. This figure shows the identity bar (green bar above the sequences) and the perpendicular small lines within the sequences refer to nucleotide differences or changes; different colours represent different nucleotides. The blue tree beside the isolate names illustrate the phylogenetic relationship among the seven strains.

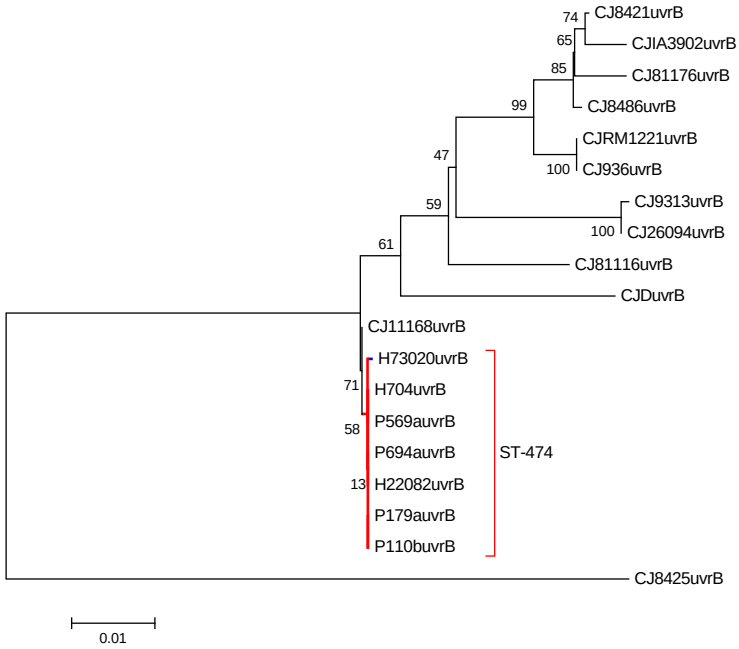


(b) Maximum likelihood tree of *uvr*A

Figure 6.17: Maximum likelihood gene tree of *uvr*A. Alignment of *uvr*A gene from the seven *C. jejuni* ST-474 genomes showing a single nucleotide change. The numbers refer to the bootstrap values corresponding to each branch of the phylogenetic tree. ST refers to sequence type.



(a) Nucleotide alignment of *uvr*B. This figure shows the identity bar (green bar above the sequences) and the perpendicular small lines within the sequences refer to nucleotide differences or changes; different colours represent different nucleotides. The blue tree beside the isolate names illustrate the phylogenetic relationship among the seven strains.



(b) Maximum likelihood tree of *uvr*B

Figure 6.18: Maximum likelihood gene tree of *uvr*B. Alignment of *uvr*B gene from the seven *C. jejuni* ST-474 genomes showing a single nucleotide change. The numbers refer to the bootstrap values corresponding to individual branch of the phylogenetic tree. ST refers to sequence type.

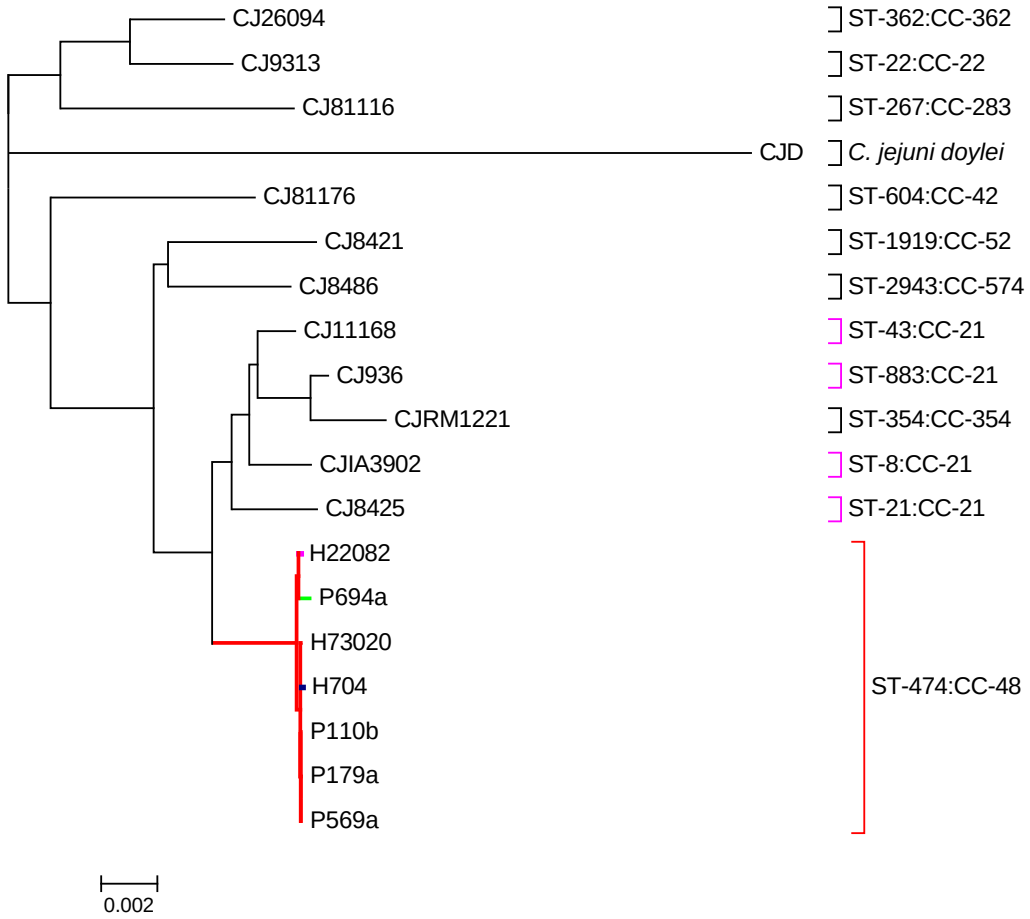


Figure 6.19: Phylogenetic tree of ribosomal, repair and nucleotide metabolic genes. Maximum likelihood gene tree constructed using concatenated DNA sequences of ribosomal, repair and nucleotide metabolic genes. ST and CC refer to the sequence type and clonal complex, respectively.

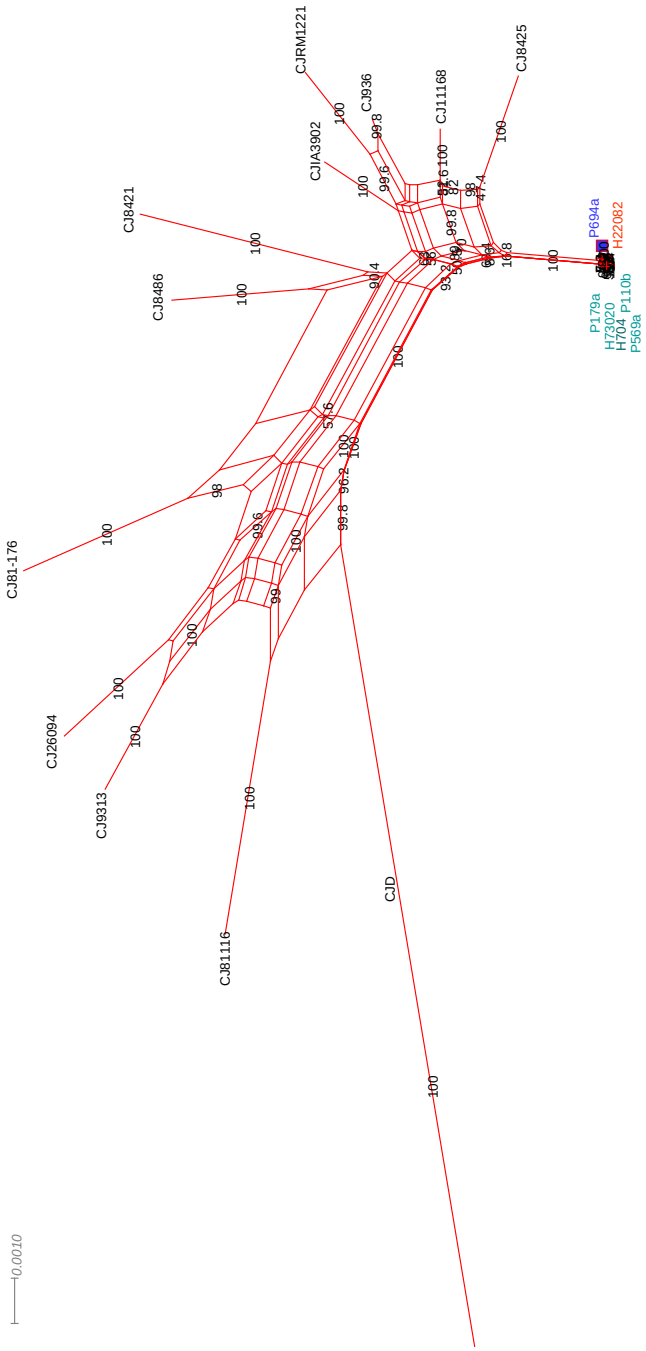


Figure 6.20: NeighbourNet tree of ribosomal, repair and nucleotide metabolic genes. NeighbourNet gene tree constructed using concatenated DNA sequences of ribosomal, repair and nucleotide metabolic genes using SplitsTree v4. The values corresponding to each node or branch denote the bootstrap values.

6.3.6 Phylogenetic congruence between informational genes and MLST housekeeping alleles

The MLST allelic profile of seven genes (*asp*A, *gln*A, *glt*A, *glm*M, *tkt* and *atp*A) originally established by Dingle et al. (2001) for the *C. jejuni* typing scheme found that CJJ84-25 to be the closest ancestral lineage of ST-474 followed by CJJCF93-6, CJJ11168 and CJJIA3902. Individual gene tree analysis and phylogenetic analysis using concatenated gene sequences in this study predicted that CJJ11168 to be the closest ancestor based on 25 full length gene sequences where, although all these four reference genomes, CJJ84-25, CJJCF93-6, CJJ11168 and CJJIA3902, belong to different sequence types they all belong to the clonal complex of ST-21. It was interesting to find that the phylogenetic prediction from 25 informational genes was in accordance with the clonal relationship as predicted by MLST allelic profiles that are provided in Table 6.11 and Figure 6.19.

Table 6.11: MLST allelic profiles of reference genomes and *C. jejuni* ST-474 genomes

Genome	GenBank accession number	*asp*A	*gln*A	*glt*A	*gly*A	*glm*M	*tkt*	*unc*A/*atp*A	ST	CC
CJJ-ST-474		**2**	**4**	**1**	**2**	**2**	**1**	**5**	**474**	**48**
CJJCF93-6	AANJ00000000	**2**	17	2	3	**2**	**1**	**5**	**883**	**21**
CJJ84-25	AANT00000000	**2**	1	**1**	3	**2**	**1**	**5**	**21**	**21**
CJJ11168	AL111168	**2**	1	5	3	4	**1**	**5**	**43**	**21**
CJJRM1221	CP000025	8	10	2	**2**	11	12	6	354	354
CJJHB93-13	AANQ00000000	1	3	6	4	3	3	3	22	22
CJJ81-176	CP000538	1	2	3	27	5	9	3	604	42
CJJ260.94	AANK00000000	1	2	49	4	11	66	8	362	362
CJJ81116	CP000814	2	7	40	4	42	51	1	267	283
CJJCG84-86	AASY00000000	7	53	27	15	11	3	3	2943	574
CJJCG84-21	ABGQ00000000	9	2	2	10	10	3	5	1919	52
CJJIA3902	CP001876	**2**	1	**1**	3	**2**	**1**	6	**8**	**21**
C. jejuni subsp *doylei*	CP000768	63	164	183	188	27	266	18	1845	U/A

CJJ:*C. jejuni* subsp. *jejuni*
ST:Sequence type
CC:Clonal complex
U/A:Unassigned
Alleles shared between the ST-474 and reference genomes are shown in underlined bold fonts

6.4 Discussion

The chemical structure of DNA is exceptionally stable and relatively simple but, the diversity encountered within cellular forms is enormous (O'Brien 2006). Evolution of DNA or genes, particularly, those involved in DNA repair result from the damage of DNA by a variety of endogenous and exogenous agents and a subsequent repair of those damaged DNA lesions. This in turn leads to the diversity in DNA structure and its composition (O'Brien 2006). Cellular life forms encode a wide range of cellular proteins and enzymes that are used to repair the damaged DNA (O'Brien 2006). The pathways of DNA repair system vary between organisms, where some may involve multiple pathways whereas, some may repair DNA damage using a single pathway (some viruses have single repair pathway) (O'Brien 2006). It has been generally accepted that mutation provides a selective advantage under certain circumstances. DNA repair and replication are the two identified systems that play a role in the alteration of a genome in order to achieve a selective advantage (O'Brien 2006).This vary widely between organisms.

C. jejuni is a human pathogen that possesses natural competency to uptake exogenous DNA and to transform its genome characteristics (Wang & Taylor 1990). The enormous adaptation potential and the genetic versatility of *C. jejuni* in response to differing environments and/or hosts has been extensively studied using the seven MLST housekeeping alleles and virulence associated genes (Dingle et al. 2001, Colles et al. 2003, Miller et al. 2006, Colles et al. 2008, Kwan et al. 2008, Sheppard et al. 2008, Wilson et al. 2008, Carter et al. 2009, Wilson et al. 2009, Sheppard et al. 2010, 2011). However, sufficient molecular signals in other genes such as genes involved in signal transduction, repair, and ribosomal functions have not been studied so far in *C. jejuni* where the DNA repair pathways have greater impact on the evolution of an organism. The primary objective of this study was to focus on the interface between a selected subset of repair and ribosomal genes from seven *C. jejuni* ST-474 genomes and the impact of evolution within this subset of genes. Comparative studies of repair and ribosomal genes from genotypes designated as identical, based on a subregion of multiple housekeeping alleles, can help to better understand the impact of evolution on the peer genes at least within the same strains. Generally genes from different species have been compared in similar studies (Zuckerkandl & Pauling 1965, Eisen & Hanawalt 1999).

There were variations in the lengths of genes investigated in this study amongst the reference genomes while the lengths did not differ amongst the ST-474 genomes. However, the nucleotide composition did differ in seven of 25 genes (*gid*A, *ogt*, *rec*J, *ssb*, *uvr*A, *uvr*B and *xse*A) which has provided the first evidence of dissimilarity between the ST-474 genomes in this study, while Biggs et al. (2011) found differences on a genomic scale. The P694a isolate of ST-474 was found to be very different from the remainder and the dissimilar nucleotide composition led to amino acid changes which reflect selection as well as diversification. However, the Tajima's D values of all genes (investigated in this study) showed a purifying selection pressure. Hence, further data analysis of P694a is required to confirm a diversification. While nucleotide composition forms the fundamental element of genomic evolution, it directly influences the GC contents of genes and/or genomes (Hurst & Merchant 2001). It was found that the overall GC content varied greatly among genes within the reference and the ST-474 genomes in this study that indicates genetic diversity was brought in by evolutionary forces.

Although there was variation in the GC contents within genes, the number of recombination sites was not associated with GC variance whereas it was related to the length of genes. Whereas, it was the opposite in the case of metabolic housekeeping genes (Chapter 5), where Rm was positively correlated with the GC variance. It may be hypothesised that an event of recombination in these genes may have occurred between sequences with high homologies which is an important genetic criteria for efficient homologous recombination (Lovett et al. 2002). Moreover, the repair and ribosomal genes belong to the lineages of higher complex Rivera et al. (1998), Jain et al. (1999), which may not allow them to compromise or tolerate GC variation during recombination. The recombination may have occurred between ST-474 (investigated in the study) and the genotypes or organisms belonging to CC-21 in the past, which was found to be closer to the majority of the genes (in this dataset). Besides the individual genome differences, the apparent differences in the overall GC contents amongst the repair and ribosomal genes investigated in this study (lower GC content in *mfd*, *ogt*, *pol*A, *rec*J, *rec*N and *xse*A genes and higher GC contents in the remainder) reflect the differences in the functional conservation and complexity. For example, *rpl*B showed the highest GC content of 39.2% which is relatively complex and conserved as it is a 50S ribosomal protein – L2, involved in several discrete steps of polypeptide synthesis such as peptidyl transferase activity, binding of aminoacyl-tRNA

to A and P-sites (Mikulik et al. 2001). Genes (investigated in this study) that differed between the seven ST-474 genomes and the possible hypothesis for the differences relating to their functional importance in a genome, are given more attention in the detailed discussion as provided below.

Situational regulation by *gid*A

The *gid*A gene of P694a was found to differ in all of the measured genetic parameters from the other genes that differed between the seven ST-474 genomes in this study. While its function was initially regarded to be fundamentally involved in the initiation of chromosome replication, recently its function was not considered to be as essential (Kinscherf & Willis 2002). Although *gid*A has been shown to be nonessential, it has been found to be profoundly conserved in nature and involved in global gene regulation (Kinscherf & Willis 2002). Replication activity at the origin by genes such as *dna*A are thought to be effectively transduced to other cellular processes through *gid*A expression (Kinscherf & Willis 2002). Further, *gid*A has been shown to be an essential gene in some organisms where there is a dependence of situational regulation, such as growth under unfavourable conditions including high temperatures (Karita et al. 1997), where under such situations, the *gid*A gene was found to be up-regulated. This raises a hypothesis that the differences in the *gid*A gene in P694a in this study could have evolved under a difficult environmental situation that accommodated certain degree of nucleotide variation. This in turn, may have resulted in the divergence of *gid*A gene from the remaining ST-474 genomes (Figure 6.12). However, it should be noted that the bootstrap values of the phylogenetic analysis of this gene are not high which raises an uncertainty around the situational evolution.

Direct DNA repair gene – *ogt*

The *ogt* gene is involved in a direct DNA repair pathway in alkylation reversal (Eisen & Hanawalt 1999). The alkyltransfer repair pathway is found in all three domains of life; the bacteria, Archaea and eukaryotes (Labahn et al. 1996). All alkyl transfer processes have been found to be catalysed by a suicide process where the protein will never be used again, while comparative sequence analyses have shown that all of the alkyl transferases contain a highly conserved core domain (Leclere et al. 1998, Skorvaga et al. 1998). This

gene has been regarded ancient and to be present in the last common ancestors of all organisms (Eisen & Hanawalt 1999). In addition, the absence of this repair pathway has been inferred as the loss of alkyl transferase activity due to gene loss during evolution. This is found in bacterial species such as *Deinococcus radiodurans*, *Rickettsia prowazekii*, *Synechocystis* sp., and *Borrelia borgdorferi* (Eisen & Hanawalt 1999) which indicates that *C. jejuni* has descended from ancient bacteria which still possesses this repair pathway.

In this study, there was only a single recombination site evident in the *ogt* gene spanning approximately 174 nucleotide base pairs while the length of the gene was only 453 base pairs (Table 6.3). In contrast, the GC content of this gene varied widely amongst genomes investigated in this study, which leads to the hypothesis that frequent recombination must have taken place in the *ogt* gene covering this particular region where recombination might not have occurred in the other parts of the gene due to its small size. The other interesting finding was that the *ogt* gene was found to differ between the ST-474 genomes. Two poultry isolates and one human isolate (P110b, P179a and H73020) were identical in their GC contents and CBI, while P694a showed greater diversity. Two human isolates and two of the poultry isolates P569a were identical in these measures, indicating the occurrence of differential evolutionary events within poultry isolates as well as between the poultry and human isolates. The Tajima's D value showed that *ogt* is under purifying selection pressure, but, the mean evolutionary rate for this gene was higher than average. This means that emergence of new variants of *ogt* can be expected in the population. From the results, we can speculate that this gene may evolve differentially under varied environmental conditions as a measure of repair through recombination. Phylogenetic analysis of *ogt* showed that CJJ11168 was closer to P694a isolate while CJJCG84-86 was the closest ancestor of the remainder.

Recombination repair and the *rec* genes

Recombination is used as a repair mechanism (Eisen & Hanawalt 1999). Traditionally the recombination pathways have been divided into RecBCD and the RecF pathways (Gillen et al. 1981, Clark et al. 1993). The primary components involved in the RecBCD pathway are *rec*A, *rec*B, *rec*C and *rec*D genes and are involved primarily in homologous recombination (Eisen & Hanawalt 1999, Skaar et al. 2002). While the other pathway mediated by RecF include the *rec*F, *rec*O, *rec*R, *rec*Q and *rec*J genes (Lovett & Kolodner 1989,

Umezu et al. 1990, Umezu & Kolodner 1994). Both the RecBCD and RecF pathway are involved in homologous recombination however, the role of the RecF pathway is limited in normal homologous recombination while RecF confers UV resistance (Horii & Clark 1973, Kolodner et al. 1985, Mahdi & Lloyd 1989), plasmid recombination (Kolodner1985, Mahdi1989) and conjugal recombination (Lloyd & Buckman 1985). *rec*J has been identified to be involved in illegitimate or non-homologous recombination (Ukita & Ikeda 1996).

In this study, all of the *rec* genes (investigated in this study) from the ST-474 genomes were found to be identical except for the H704 ST-474 isolate which differed due to a non-synonymous substitution, for which the CJJCG84-86 genome was the closest ancestor. The *rec*J gene was found to be evolving relatively faster than average compared to other *rec* genes investigated in this study except for *rec*R that was found to evolve even faster than *rec*J. While considering the positions of these two genes on the genome, *rec*R lies very close to the *gua*A gene (a nucleotide metabolic gene) that was evidenced with large number of recombination sites (n = 25). It is possible that an event of recombination may have spanned these two genes together that in turn may have led to an increased nucleotide diversity in *rec*R. However, this speculation needs further data analysis.

The single strand binding protein – *ssb*

Single strand binding protein (*ssb*) is one of the major contributing genes involved in both the DNA damage repair and replication mechanisms (Eisen & Hanawalt 1999). *ssb* is essential for DNA replication, recombination and repair (Lohman & Ferrari 1994) while this gene facilitates the binding of other DNA-binding proteins, particularly, Cj0011c (Jeon & Zhang 2007). The increased DNA-binding ability among the DNA-binding proteins has been reported to enhance natural transformation in *C. jejuni* that was not recognised previously (Jeon & Zhang 2007).

In this study, *ssb* is one of the genes that differed among the seven ST-474 genomes in that it varied in measures such as GC, GC3 and CBI. The GC content for this gene varied markedly among genomes. The overall GC content of this gene from two of the ST-474 isolates, the H22082 and P694a, were low compared with the other ST-474 isolates and the reference genomes except for CJJHB93-13 and CJ81116. The CJJHB93-13 and

CJ81116 genomes had a slightly lower GC content compared with these two ST-474 isolates. Although the GC content varied greatly among genomes, the *ssb* gene showed only a single recombination site which may be due to its small size (549-552 base pairs). In contrast, the region that was involved in recombination was longer, around 214 base pairs.

There was clustering between the H22082 and P694a isolates where, these two isolates were found to possess an insertion of *ykg*C gene in their genomes (Biggs et al. 2011). This gene encodes a probable pyridine nucleotide-disulfide oxidoreductase and has been reported to be present in most of the enteric bacteria including *E. coli* (Rozen & Belkin 2001, Bradley et al. 2007). Its presence in *C. jejuni* has not been reported previously. This gene has been evidenced to confer resistance to oxidative stress and has been demonstrated to promote survival in seawater in *E. coli* (Rozen & Belkin 2001, Bradley et al. 2007).

The *ykg*C and *ssb* genes were found to be positioned next to each other on the ST-474 genome in the study carried out by Biggs et al. (2011). The authors of this study inferred that the whole event of insertion or horizontal transfer of *ykg*C followed by recombination may have involved a partial region of the *ssb* gene in the ST-474 isolate. In this study, we believe that, a similar genetic event must have occurred in the P694a isolate. This in turn may have caused the clustering between these two isolates with respect to the *ssb* alleles. These two isolates share their ancestry with CJJ84-25, where this genome was found to possess the *ykg*C gene insertion (Biggs et al. 2011). The remaining ST-474 isolates were closer to CJJIA3902. It may be hypothesised that there may be a few other *C. jejuni* isolates in the population carrying this gene which may potentially be transferring this gene to the isolates that do not possess the *ykg*C gene.

Excision repair genes

The *uvr*A and *uvr*B genes differed between the seven ST-474 genomes and *uvr*A differed at the amino acid level. These genes are the first line DNA repair genes to appear when a DNA damage is induced (Michel 2005, Gaasbeek et al. 2009a). It may be hypothesised that the varied environmental threats that include different gut environments could have potentially influenced the subsequent evolution of these genes. However, further molecular analysis will provide stronger support for this hypothesis.

The H704a and H73020 isolates differed in these two genes in two different ways:

- the H73020 *uvr*B gene differed from the rest of the ST-474s and differed only at the nucleotide level;
- the H704a isolate showed differences in its *uvr*A gene and differed both at the nucleotide and amino acid level.

Both the *uvr*A and *uvr*B alleles were found to be unique to ST-474 where CJJ11168 was found to be closer to *uvr*B and CJJRM1221 was found to be closer to *uvr*A. It is interesting to note that CJJRM1221 was one of the closer relatives because, it is the only poultry reference strain while all others are from human clinical cases.

The *xse*A gene is a large subunit of exonulcease VII that was first characterised in *E. coli* and is involved in mismatch excision repair (Chase et al. 1986). However, the functional importance and the role of this particular gene has not been well studied and/or documented in *Campylobacter* spp. while the repair mechanism involving methyl-directed mismatch repair and recombinational repair has been studied experimentally (Gaasbeek et al. 2009b). Therefore, the importance of this gene in *Campylobacter* is not very clear. In this study, there was a considerable amount of DNA variation between the seven ST-474 genomes that clustered three poultry isolates together namely, P110b, P569a and P179a.

Selenosysteine – amino acids in repair genes

Selenocysteine is the 21^{st} amino acid encoded by the codon UGA. The codon UGA differs from the rest of the codons in the genetic code in that it plays a dual role either a stop codon or a codon of the rare protein, Selenocysteine (reviewed by (Low & Berry 1996, Bock 2000, Hatfield & Gladyshev 2002)). Recent phylogenetic analyses of bacterial genomes have explained the incorporation or evolution of genes with selenocysteine as a result of speciation, differential gene loss and horizontal gene transfer (HGT) while it has been stated that the loss or the acquisition of this particular trait is not reversible (Romero et al. 2005, Zhang et al. 2006b). Oxygen concentration, temperature and the environment are the factors identified to influence the evolution of selenium utilising traits (Zhang et al. 2006b). The Sec-decoding (selenocysteine decoding) system is a fundamental trait that greatly increases the catalytic selenoenzymes activity (selenium utilising

enzymes) and has been regarded as a selective advantage even though the amino acid is very rare Zhang et al. (2006b). Further this trait is found in most of the bacterial lineages that include species from Proteobacteria and Firmicutes, particularly, βproteobacteria, δproteobacteria, ϵproteobacteria, γproteobacteria and Firmicutes/Clostridia subdivisions (Zhang et al. 2006b), which includes *C. jejuni* in ϵproteobacteria. Analyses of prokaryotic selenoproteomes of both complete and incomplete genomes had revealed that although selenoproteins were found in most of the bacterial lineages, only 20% of the completed bacterial genomes were found to possess this trait (Kryukov & Gladyshev 2004). The SelW-like protein has been identified as a new selenoprotein in *C. jejuni* (Kryukov & Gladyshev 2004). However, the underlying molecular mechanisms involved and/or influencing this trait is not very clear.

In this study, *uvr*A, *mut*S and *gyr*A were identified to possess this rare amino acid in the reference and ST-474 genomes where this trait could have been acquired through HGT which in turn denotes the evolutionary action towards the environmental stress such as temperature, oxygen concentration and the environment itself, as may be expected inside different hosts. It also raises a possibility of the simultaneous presence of other bacterial species in the guts of the host that possessed this trait to pass on to the *Campylobacter* spp. or it can even be an innate presence of this trait in *Campylobacter*. However, as this codon plays a dual role, the results do not provide concrete evidence whether these codons encoded for a selenocysteine or a premature stop codon (Wong et al. 2008) which needs further investigation.

Conclusion

DNA repair mechanism is a vital system in bacteria that shapes their evolution under fluctuating environments. This study has analysed 25 genes (ribosomal, repair and nucleotide metabolic genes) from seven *C. jejuni* ST-474 genomes which in turn, were compared with 12 *C. jejuni* reference genomes. Even though the seven ST-474 genomes were found to belong to a single genetic type (sequence type) based on a subregion of seven housekeeping alleles, there were differences between seven of 25 informational genes between these genomes. These genes varied at nucleotide level while *gid*A, *rec*J, *ssb* and *uvr*A differed at the amino acid level also. Genes investigated in this study were found to be under purifying selection when using the Tajima's D test for whole genes but evidence of

positive selection pressure was found in a few individual codons of genes. The majority of genes showed a similar evolutionary pattern to isolates within CC-21 including CJJ11168 while the genetic characteristics of few genes/alleles were found to be unique to ST-474. The genetic dissimilarities found between the ST-474 genomes, particularly, within genes such as *ssb*, *rec*J and *uvr*A suggest the evolutionary potential of ST-474 as measure of adaptation to different niches. The presence of selenocysteine amino acid in three of 25 genes provides limited evidence of HGT which is suggestive of an added evolutionary measure to combat stress conditions. However, gene expression and protein analyses as well as analyses using wider molecular datasets are necessary to confirm the presence of this trait. The allelic uniqueness in some of the repair alleles indicates that ST-474 is genetically distinct from that of the reference genomes investigated in this study. Its evolution may be attributed to the relative geographical isolation of this country that separated this organism from the rest of the world. This isolation may have occurred through an absence of exchange of genotypes or strains that is typically mediated through international trade of fresh poultry meat where poultry is neither imported or exported from New Zealand. However, analysis of larger number of ST-474 samples and other strains of *C. jejuni* will lend additional support to the findings of this study. A better understanding of the relationship between the DNA repair pathway and evolution in the emergence of new strains will provide new opportunities for drug targeting and for identifying potential vaccine candidates.

CHAPTER 7

General discussion

Briefly, this thesis has combined molecular genetics and 'classical' epidemiological analytical techniques in an effort to study the prevalence, population structure and the evolutionary potential of *C. jejuni*. The colonisation or the prevalence of *C. jejuni*, the gene flow and divergence of *C. jejuni* populations among different host species (mallard ducks, starlings, dogs and cats) and between different sampling locations has been examined. Population genetics and phylogenetics have been employed to unravel the evolutionary events influencing the population-host species interface of *C. jejuni* as well as the evolutionary dynamics of a selected subset of genes from the genomes of identical MLST *C. jejuni* genotypes, namely the *C. jejuni* ST-474 found almost exclusively in New Zealand.

Molecular epidemiology combined population genetics

The analyses of the prevalence of *Campylobacter* and *C. jejuni* in mallard ducks and starlings over a 12-month period (Chapter 3) provided a unique opportunity to assess temporal changes in carriage frequency and also to compare these estimates among two different host species that share a common environment.

The overall prevalence of *Campylobacter* spp. was 37% (95% CI 35 to 40%) which is relatively high compared with estimates for migrating birds (22%) by Waldenstrom et al. (2002). The prevalence of *Campylobacter* spp. (46%, 95% CI 42 to 50%) was relatively high in European starlings compared with previous prevalence estimates that ranged between 33 to 40% (Waldenstrom et al. 2002, Colles et al. 2008, Hughes et al. 2009). In contrast, the prevalence of *C. jejuni* in ducks (23%, 95% CI 20 to 26%) and starlings (21%, 95% CI 18 to 23%) was relatively low. It should be noted that differences in study findings

may arise due to a number of factors such as an animal's risk of contracting the bacteria, sampling procedures, sample size, the type of samples retrieved and the sensitivity of the culture techniques that are applied. The bimodal seasonal pattern of *Campylobacter* spp. and *C. jejuni* prevalence in ducks and starlings (with distinct peaks in autumn and spring, Figures 3.2 and 3.3) has not previously been reported in the literature. In ducks in the early spring a relatively high prevalence of *C. jejuni* has been reported elsewhere (Colles et al. 2009), so too has an increased rate of shedding of *C. jejuni* in ducks in the autumn been reported by Waldenstrom et al. (2002), Broman et al. (2004).

The studies of the population structure of *C. jejuni* (Chapter 3) documented the association between host and genotype using minimum spanning trees and FST analyses. Population differentiation was studied over different time periods and different sampling locations. This, in turn, has provided an indication of host-pathogen interaction. In Chapter 3 evidence was provided to support the hypothesis that there was limited gene flow between host species and widespread transmission of genotypes between different locations. The limited gene flow and evidence of population differentiation between host species is thought to reflect the *in-vivo* evolution of *C. jejuni* as a measure of adaptation. Host adaptation is a subject of debate where, the majority of population genetic researchers have accepted the idea of host-adaptation or host-signatures expressed by *C. jejuni* populations, while others remain unconvinced. In Chapter 3 clear clustering of clonal complexes was discernible between hosts and over time. This lends support to those that argue towards host-adaptation. For example, the ST-1034, ST-692 and ST-1332 complexes were the most predominant lineages found in ducks while in starlings, the ST-177 complex was the prominent lineage. The cell surface antigens *por*A and *fla*A from both species (ducks and starlings) demonstrated a high degree of diversity, particularly in ducks. The high variability in genotypes and cell surface antigens in ducks suggests that the aquatic feeding habit of ducks could be a major contributing factor for this diversity.

The occurrence of new genotypes that could not be assigned to a clonal complex, particularly during winter (in both ducks and starlings) may be an indication of ongoing *in-vivo* evolution in *C. jejuni*. Possible explanations for the emergence of new variants may be : (1) an adaptation measure to suit temporal patterns, (2) the influence of gut microbiota and physiology of the hosts during different seasons (3) due to mixing up of new genotypes from newly arrived bird populations with genotypes in the existing population or

the older birds that nest during this season and (4) a selection process towards the host species that suit the survivability of the organism, but it could also be that they had not been sampled yet in New Zealand and this is the first study to the best of my knowledge.

In Chapter 4, prevalence of *Campylobacter* and *C. jejuni* in dogs for a 12 month period was quantified. The prevalences of *Campylobacter* spp. and *C. jejuni* in dogs were 15% (95% CI 11% to 19%) and 7% (95% CI 5% to 11%) and in cats 9% (95% CI 4% to 17%) and 7% (95% CI 3% to 15%), respectively. The population structure of *C. jejuni* in cats and dogs was not very complex unlike the mallard ducks and starlings and moreover, the genotype sample size from cats and dogs was small compared to the wild birds. The predominant clonal complex was ST-45 (13 of 21), a complex that is frequently reported in human campylobacteriosis cases throughout the world followed follwed by the sequence types ST-52 (2 of 21), ST-50, ST-422, ST-474, ST-583, ST-696 and ST-3961.

The presence of ST-45 in a wide range of hosts describes the adaptability of this genotype to a variety of animal hosts and environments (the ST-45 complex was predominant in the ducks as well as starlings, in this study). Although a reasonable number of *C. jejuni* isolates from dogs were characterised, many of the isolates remained partially characterised which raises speculation that these alleles could have become different in their nucleotide composition due to the evolutionary forces such as recombination or mutation that prevented them from being amplified or sequenced. The possible underlying reason may be the molecular differences that may have arisen due to different gut environments and/or errors in amplification and sequencing. This in turn may have prevented the primers from getting hybridised with the dog *C. jejuni* DNA. However, further amplification and sequencing using degenerate primers are being carried out on these isolates in order to obtain a complete profile.

Phylogenetics to estimate evolution

This thesis has further (Chapters 5 and 6) expanded the breadth of housekeeping gene analysis from a molecular epidemiology focused population genetics approach that was based on a sub-region of seven alleles (genes) to full length gene analysis in an effort to understand *C. jejuni* evolution. The increasing amount of MLST datasets[1] which describe the population expansion of *C. jejuni* raises important questions such as:

[1]URL: (`http://pubmlst.org/Campylobacter`)

1. can a sub-region in the MLST alleles potentially reflect the evolutionary events of the respective full length gene, particularly within the designated identical clones or genotypes;

2. what is the extent of molecular similarities or differences within a group of identical genotypes (designated so by MLST); and

3. does the ancestry of the MLST alleles reflect that of the majority of the genes within genotypes designated identical by MLST.

We addressed these questions in Chapters 5 and 6. Fifty loci or genes were examined to better understand the pattern of evolution in housekeeping and those genes involved in replication, information processes and nucleotide metabolism from seven *C. jejuni* ST-474 genomes (identical clones or genotypes designated so by MLST typing).

The housekeeping genes analysis presented in Chapter 5 demonstrates the robustness of the MLST alleles in predicting the last common ancestor for the majority of the genes investigated in the study. Further, the majority of genes were identical in their molecular characteristics and/or the evolutionary measures analysed in the study such as GC content, codon bias index (CBI) and phylogenetic relationship. In this study, GC and GC3 contents have provided valuable insights into the evolutionary dynamics of some genes within ST-474 such as *fum*C, *trp*C, *pyc*A, *tkt* and *glt*A that showed greater variations across the investigated genomes including the ST-474 genomes. GC content has been shown to influence the CBI and the number of recombination events in the genes and/or genomes (Marais et al. 2001, Genereux 2002, Fuglsang 2003). In this study (Chapter 5), however, evidence of positive correlation within these parameters such as a wider variation in the GC content within genes showed a relatively higher number of recombination events and the GC content did not vary much across genes where there were fewer recombination events (e.g. the *atp*D and *trp*B genes). Similarly, the ST-474, CJJIA3902, CJJ84-25, CJJCF93-6 and CJJ11168 genomes were identical in their GC contents and CBI within the selected subsets of genes, which suggests that these genomes must have undergone a similar pattern of evolution which had made them fit into a cluster or a group. It should be noted that CBI and GC content are positively correlated to the tRNA pools present between genomes that are recombining or undergoing lateral gene transfer (Ikemura 1985, Kanaya et al. 1999, Ochman et al. 2000, Tuller et al. 2011). Hence, it may be hypothesised

that the ST-474 genomes investigated in this study might have recombined with genomes that belonged to ST-21 complex in the past. The results from the cell surface antigen or structural gene (Chapter 5), the *por*A gene analysis demonstrated the hyper-variability regions within the gene and how recombination and evolution have influenced the change of ancestry between the ST-474 genomes. Analysis of wider datasets for the evolutionary parameters estimated in this study should provide support for the findings of this study.

Similarly, the results from Chapter 6 provide interesting insight into the ribosomal, repair genes, collectively known as information process genes and nucleotide metabolic genes. Once again MLST allelic prediction was found to be robust in predicting the evolutionary last common ancestor of the majority of the informational process genes of *C. jejuni* ST-474 genomes. In addition, seven of the 25 investigated genes within ST-474 genomes demonstrated surprising genetic uniqueness, raising speculation that *C. jejuni* ST-474's DNA repair system may be well equipped to cope up with the changing environment. Each ST-474 isolate demonstrated distinguishable levels of molecular differentiation in *gid*A, *ogt*, *rec*J, *ssb*, *uvr*A, *uvr*B and *xse*A genes with respect to all the measures of evolution and these alleles were found to be very specific to ST-474 genomes. Further, this study evidenced the presence of selenocysteine amino acid insertions in three genes which is an indication of horizontal gene transfer as a measure of combating the stress which in turn reflects the capability of *C. jejuni* in general to evolve for fitness. However, genetic analyses involving genomes from ST-48 complex apart from the seven ST-474 genomes from this study could not be carried out as there were no ST-48 genomes publicly available at the time of analysis. Such an approach in turn might have been beneficial to draw insightful inferences about the ST-48 complex itself.

Given the widespread availability of fully sequenced bacterial genomes and the increasing use of computational tools and statistical techniques for analysing DNA sequence data, there is a great deal of research that can be done not only for understanding the genetic epidemiology of *C. jejuni* but also for the other endemic and emerging disease pathogens such as *Salmonella* spp. and *E. coli* to understand the evolutionary mechanisms, gene flow, migration and transmission. Evolutionary research areas play a major role in the process of deciphering global health problems. Research in these areas can be translated to benefit a wide range of fields such as diagnostics, gene-based epidemiological investigations, source-pathogen relationships, trade and disease. Further, such studies

will provide opportunities to identify potential drug targets and candidates for vaccine production.

7.0.1 Conclusion

The application of population genetics and phylogenetic approaches to DNA sequence data of human pathogens such as *C. jejuni* provide a better understanding of disease transmission pathways, population structure, host-species relationships, gene flow and evolution. Descriptive analyses such as those applied in 3 and 4 have characterised *Campylobacter* colonisation by species, site and over time. Further, population genetic analyses such as rarefaction, molecular variance and fixation indices of *C. jejuni* genotypes are useful for describing genetic diversity within genotypes and differentiation between populations of pathogens. This general approach can be applied to just about any pathogen of humans and/or animals. Evolutionary measures such as those used in Chapters 5 and 6 are useful for providing a better understanding of the genetic evolution of pathogens. Further, comparison between metabolic housekeeping genes and the informational process genes showed that they differ in their relationship between GC variance and the number of recombination sites. The metabolic housekeeping genes showed a positive correlation between their GC variance and the number of recombination sites whereas, the later did not show such correlation. Collectively, the results presented in this thesis provide evidence that *C. jejuni* has the ability to evolve rapidly. The consequences of rapid evolution in terms of virulence is difficult to predict. Further insights into the genetics of *C. jejuni* are essential and this information, combined with the results of observational epidemiological studies, will help to improve preventive and control measures. Finally, it is important to make an effort to conduct such investigations in a number of geographic locations (where, for example, disease is endemic, emerging or re-emerging). This will facilitate a better understanding of the variation in pathogen populations, variations in disease expression, details of pathogen movement and transmission, and evolutionary characteristics.

Bibliography

Achtman, M. & Wagner, M. (2008), 'Microbial diversity and the genetic nature of microbial species', *Nature Reviews Microbiology* **6**, 433–443.

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

Acke, E., McGill, K., Lawlor, A., Jones, B. R., Fanning, S. & Whyte, P. (2010), 'Genetic diversity among *Campylobacter jejuni* isolates from pets in Ireland', *Veterinary Record* **166**, 102–106.

Acke, E., Whyte, P., Jones, B. R., McGill, K., Collins, J. D. & Fanning, S. (2006), 'Prevalence of thermophilic *Campylobacter* species in cats and dogs in two animal shelters in Ireland', *Veterinary Record* **158**, 51.

Adak, G. K., Long, S. M. & O'Brien, S. J. (2002), 'Trends in indigenous foodborne disease and deaths, England and Wales : 1992 to 2000', *Gut* **51**, 832–841.

Adak, G. K., Meakins, S. M., Yip, H., Lopman, B. A. & O'Brien, S. J. (2005), 'Disease risks from foods, England and Wales, 1996-2000', *Emerging Infectious Diseases* **11**, 365–372.

Adegbola, R. A., Alabi, S. A., Akinkuade, F. O., Coker, A. O. & Odugbemi, T. (1990), 'Correlation between human and animal bio-serogroups of *Campylobacter* isolates in Nigeria', *Journal of Tropical Medicine and Hygiene* **93**, 280–283.

Adekambi, T., Drancourt, M. & Raoult, D. (2009), 'The *rpo*B gene as a tool for clinical microbiologists', *Trends in Microbiology* **17**, 37–45.

Akashi, H. (1995), 'Inferring weak selection from patterns of polymorphism and divergence at 'silent' sites in Drosophila DNA', *Genetics* **139**, 1067–1076.

Allerberger, F., Kasten, M. J. & Anhalt, J. P. (1991), '*Campylobacter fetus* subspecies *fetus* infection', *Journal of Molecular Medicine* **69**, 813–816.

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

Andries, K., Verhasselt, P., Guillemont, J., Gohlmann, H. W. H., Neefs, J.-M., Winkler, H., Van Gestel, J., Timmerman, P., Zhu, M., Lee, E., Williams, P., de Chaffoy, D., Huitric, E., Hoffner, S., Cambau, E., Truffot-Pernot, C., Lounis, N. & Jarlier, V. (2005), 'A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*', *Science* **307**, 223–227.

Asakura, M., Samosornsuk, W., Hinenoya, A., Misawa, N., Nishimura, K., Matsuhisa, A. & Yamasaki, S. (2008), 'Development of a cytolethal distending toxin (cdt) gene-based species-specific multiplex PCR assay for the detection and identification of *Campylobacter jejuni*, *Campylobacter coli* and *Campylobacter fetus*', *FEMS Immunology and Medical Microbiology* **52**, 260–266.

Avrain, L., Vernozy-Rozand, C. & Kempf, I. (2004), 'Evidence for natural horizontal transfer of *tet*O gene between *Campylobacter jejuni* strains in chickens', *Journal of Applied Microbiology* **97**, 134–140.

Ayala, F. J. (1976), *Allozyme variation: its determinants in space and time. In Molecular evolution*, Sunderland, Mass. Sinauer Associates Inc.

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

Bacon, D. J., Alm, R. A., Hu, L., Hickey, T. E., Ewing, C. P., Batchelor, R. A., Trust, T. J. & Guerry, P. (2002), 'DNA sequence and mutational analyses of the p*Vir* plasmid of *Campylobacter jejuni* 81-176', *Infection and Immunity* **70**, 6242–6250.

Baker, J., Barton, M. D. & Lanser, J. (1999), '*Campylobacter* species in cats and dogs in South Australia', *Australian Veterinary Journal* **77**, 662–668.

Baker, M., Sneyd, G. E. & Wilson, N. A. (2007), 'Is the major increase in notified campylobacteriosis in New Zealand real?', *Epidemiology and Infection* **135**, 163–170.

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

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

Baserisalehi, M., Bahadur, N. & Kapadnis, B. P. (2007), 'A comparison study on antimicrobial susceptibility of *Campylobacter* spp. isolates from faecal samples of domestic animals and poultry in India and Iran', *Journal of Biological Sciences* **7**, 977–980.

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

Bellgard, M. I. & Gojobori, T. (1999), 'Significant differences between the G+C content of synonymous codons in orthologous genes and the genomic G+C content', *Gene* **238**, 33–37.

Bender, J. B., Shulman, S. A., Averbeck, G. A., Pantlin, G. C. & Stromberg, B. E. (2005), 'Epidemiologic features of *Campylobacter* infection among cats in the upper midwestern United States', *Journal of the American Veterinary Medical Association* **226**, 544–547.

Bennetzen, J. L. & Hall, B. D. (1982), 'Codon selection in yeast', *Journal of Biological Chemistry* **257**, 3026–3031.

Bentley, S. D. & Parkhill, J. (2004), 'Comparative genomic structure of prokaryotes', *Annual Reviews of Genetics* **38**, 771–792.

Bernardi, G. & Bernardi, G. (1985), 'Codon usage and genome composition', *Journal of Molecular Evolution* **22**, 363–365.

Berndtson, E. & Engvall, A. (1994), Control of *Campylobacter* in Swedish broiler flocks, Technical report, Report of a WHO consultation on epidemiology and control of campylobacteriosis. Bilthoven, The Netherlands.

Best, E. L., Fox, A. J., Frost, J. A. & Bolton, F. J. (2005), 'Real-time single-nucleotide polymorphism profiling using Taqman technology for rapid recognition of *Campylobacter jejuni* clonal complexes', *Journal of Medical Microbiology* **54**, 919–925.

BGOSHU (2000), Bruce-Grey-Owen Sound Health Unit. The investigative report of the Walkerton outbreak of waterborne gastroenteritis, May-June 2000, Technical report, Bruce-Grey-Owen Sound Health Unit, Walkerton, Canada. `http://water.sesep.drexel.edu/outbreaks/Walkerton/ReportOct2000`.

Bi, P., Cameron, S., Zhang, Y. & Kevin, P. A. (2008), 'Weather and notified *Campylobacter* infections in temperate and sub-tropical regions of Australia: an ecological study', *Journal of infection* **57**, 317–323.

Biggs, P. J., Fearnhead, P., Hotter, G., Mohan, V., Collins-Emerson, J., Kwan, E., Besser, T. E. & French, N. P. (2011), 'Whole-genome comparison of two *Campylobacter jejuni* isolates indistinguishable on the basis of MLST and *fla*A SVR reveals multiple loci of different ancestral lineage', *Under preparation* .

Black, R. E., Levine, M. M., Clements, M. L., Hughes, T. P. & Blaser, M. J. (1988), 'Experimental *Campylobacter jejuni* infection in humans', *Journal of Infectious Diseases* **157**, 472–479.

Bock, A. (2000), 'Biosynthesis of selenoproteins–an overview', *Biofactors* **11**, 77–78.

Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. (2005), 'Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin', *Microbiology* **151**, 2551–2561.

Bolton, F. J., Holt, A. V. & Hutchinson, D. N. (1984), '*Campylobacter* biotyping scheme of epidemiological value', *Journal of Clinical Pathology* **37**, 677–681.

Bonjoch, X., Calvo, L., Soler, M., Ruiz-Rueda, O. & Garcia-Gil, L. J. (2010), 'A new multiplexed real-time PCR assay to detect *Campylobacter jejuni, C. coli, C. lari*, and *C. upsaliensis*', *Food Analytical Methods* **3**, 40–46.

Borck, B. (2003), Campylobacter *spp. in the Danish turkey production*, The Royal Veterinary and Agricultural University, Copenhagen, 77 pp.

Bowler, L. D., Zhang, Q., Riou, J.-Y. & Spratt, B. G. (1994), 'Interspecies recombination between the *pen*A genes of *Neisseria meningitidis* and commensal *Neisseria* species during the emergence of penicillin resistance in *N. meningitidis*: natural events and laboratory simulation', *Journal of Bacteriology* **176**, 333–337.

Bradley, M. D., Beach, M. B., de Koning, A. P., Pratt, T. S. & Osuna, R. (2007), 'Effects of fis on *Escherichia coli* gene expression during different growth stages', *Microbiology* **153**, 2922–2940.

Bremell, T., Bjelle, A. & Svedhem, A. (1991), 'Rheumatic symptoms following an outbreak of campylobacter enteritis: a five year follow up', *Annals of the Rheumatic Diseases* **50**, 934–938.

Broman, T., Waldenstrom, J., Dahlgren, D., Carlsson, I., Eliasson, I. & Olsen, B. (2004), 'Diversities and similarities in PFGE profiles of *Campylobacter jejuni* isolated from migrating birds and humans', *Journal of Applied Microbiology* **96**, 834–843.

Brown, P. E., Christensen, O. F., Clough, H. E., Diggle, P. J., Hart, C. A., Hazel, S., Kemp, R., Leatherbarrow, A. J. H., Moore, A., Sutherst, J., Turner, J., Williams, N. J., Wright, E. J. & French, N. P. (2004), 'Frequency and spatial distribution of environmental *Campylobacter* spp.', *Applied and Environmental Microbiology* **70**, 6501–6511.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J.-F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M. & Venter, J. C. (1996), 'Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*', *Science* **273**, 1058–1073.

Burnens, A. P., Angeloz-Wick, B. & Nicolet, J. (1992), 'Comparison of *Campylobacter* carriage rates in diarrheic and healthy pet animals', *Zentralbl. Veterinarmed [B]* **39**, 175–180.

Bustamante, V. H., Puente, J. L., Sanchez-Lopez, F., Bobadilla, M. & Calva, E. (1995), 'Identification of *Campylobacter jejuni* and *C. coli* using the *rpo*B gene and a cryptic DNA fragment from *C. jejuni*', *Gene* **165**, 1–8.

Cann, I. K. O. & Ishino, Y. (1999), 'Archaeal DNA replication: identifying the pieces to solve a puzzle', *Genetics* **152**, 1249–1267.

Carle, G. F. & Olson, M. V. (1984), 'Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis', *Nucleic Acids Research* **12**, 5647–5664.

Carrillo, C. D., Taboada, E., Nash, J. H., Lanthier, P., Kelly, J., Lau, P. C., Verhulp, R., Mykytczuk, O., Sy, J., Findlay, W. A., Amoako, K., Gomis, S., Willson, P., Austin, J. W., Potter, A., Babiuk, L., Allan, B. & Szymanski, C. M. (2004), 'Genome-wide expression analyses of *Campylobacter jejuni* NCTC11168 reveals coordinate regulation of motility and virulence by *flh*A', *Journal of Biological Chemistry* **279**, 20327–20338.

Carter, A. M., Pacha, R. E., Clark, G. W. & Williams, E. A. (1987), 'Seasonal occurrence of *Campylobacter* spp. in surface waters and their correlation with standard indicator bacteria', *Applied and Environmental Microbiology* **53**, 523–526.

Carter, P. E., McTavish, S. M., Brooks, H. J. L., Campbell, D., Collins-Emerson, J. M., Midwinter, A. C. & French, N. P. (2009), 'Novel clonal complexes with an unknown animal reservoir dominate *Campylobacter jejuni* isolates from river water in New Zealand', *Applied Environmental Microbiology* **75**, 6038–6046.

CDC (2000), Foodborne disease outbreak line listing, 1990-1999, Technical report, Centers for Disease Control and Prevention, U.S. `http://www.cdc.gov/ncidod/dbmd/outbreak/us-outb.htm`.

CDC (2007), *Campylobacter jejuni* infection associated with unpasteurized milk and cheese–Kansas, 2007., Technical report, Centres for Disease Control and Prevention. January 2, 2009 / 57(51&52) ; 1377-1379. `http://www.cdc.gov/nczved/divisions/dfbmd/diseases/campylobacter/additional.html`.

Chaban, B., Ngeleka, M. & Hill, J. E. (2010), 'Detection and quantification of 14 *Campy-*

lobacter species in pet dogs reveals an increase in species richness in feces of diarrheic animals', *BioMedCentral Microbiology* **10**, 1–7.

Champion, O. L., Gaunt, M. W., Gundogdu, O., Elmi, A., Witney, A. A., Hinds, J., Dorrell, N. & Wren, B. W. (2005), 'Comparative phylogenomics of the food-borne pathogen *Campylobacter jejuni* reveals genetic markers predictive of infection source', *Proceedings of National Academy of Science* **102**, 16043–16048.

Chan, K., Elhanafi, D. & Kathariou, S. (2008), 'Genomic evidence for interspecies acquisition of chromosomal DNA from *Campylobacter jejuni* by *Campylobacter coli* strains of a turkey-associated clonal group (Cluster II)', *Foodborne Pathogens and Disease* **5**, 387–398.

Chase, J. W., Rabin, B. A., Murphy, J. B., Stone, K. L. & Williams, K. R. (1986), '*Escherichia coli* exonuclease VII. Cloning and sequencing of the gene encoding the large subunit (*xse*A)', *Journal of Biological Chemistry* **261**, 14929–14935.

Chen, S. L., Hung, C.-S., Xu, J., Reigstad, C. S., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R. R., Ozersky, P., Armstrong, J. R., Fulton, R. S., Latreille, J. P., Spieth, J., Hooton, T. M., Mardis, E. R., Hultgren, S. J. & Gordon, J. I. (2006), 'Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli:* a comparative genomics approach', *Proceedings of National Academy of Science* **103**, 5977–5982.

Cheng, C. & Shuman, S. (1997), 'Characterization of an ATP-dependent DNA ligase encoded by *Haemophilus influenzae*', *Nucleic Acids Research* **25**, 1369–1375.

Christensen, H., Kuhnert, P., Olsen, J. E. & Bisgaard, M. (2004), 'Comparative phylogenies of the housekeeping genes *atp*D, *inf*B and *rpo*B and the 16S rRNA gene within the *Pasteurellaceae*', *International Journal of Systematic and Evolutionary Microbiology* **54**, 1601–1609.

Cirz, R. T., Chin, J. K., Andes, D. R., de Crecy-Lagard, V., Craig, W. A. & Romesberg, F. E. (2005), 'Inhibition of mutation and combating the evolution of antibiotic resistance', *Public Library of Science Biology* **3**, 1024–1033.

Clark, A. J., Sharma, V., Brenowitz, S., Chu, C. C., Sandler, S., Satin, L., Templin, A., Berger, I. & Cohen, A. (1993), 'Genetic and molecular analyses of the C-terminal region of the *rec*E gene from the *Rac* prophage of *Escherichia coli* K-12 reveal the *rec*T gene', *Journal of Bacteriology* **175**, 7673–7682.

Clark, C. G., Beeston, A., Bryden, L., Wang, G., Barton, C., Cuff, W., Gilmour, M. W. & Ng, L.-K. (2007), 'Phylogenetic relationships of *Campylobacter jejuni* based on *por*A sequences', *Canadian Journal of Microbiology* **53**, 27–38.

Clark, C. G., Bryden, L., Cuff, W. R., Johnson, P. L., Jamieson, F., Ciebin, B. & Wang, G. (2005), 'Use of the Oxford multilocus sequence typing protocol and sequencing of the flagellin short variable region to characterize isolates from a large outbreak of water-borne *Campylobacter* sp. strains in Walkerton, Ontario, Canada', *Journal of Clinical Microbiology* **43**, 2080–2091.

Cody, A. J., Maiden, M. J. C. & Dingle, K. E. (2009), 'Genetic diversity and stability of the *por*A allele as a genetic marker in human *Campylobacter* infection', *Microbiology* **155**, 4145–4154.

Coenye, T. & Vandamme, P. (2003), 'Extracting phylogenetic information from whole-genome sequencing projects: the lactic acid bacteria as a test case', *Microbiology* **149**, 3507–3517.

Coenye, T. & Vandamme, P. (2005), 'Displacement of *epsilon*-proteobacterial core genes by horizontally transferred homologous genes', *Research in Microbiology* **156**, 738–747.

Coker, A. O. & Dosunmu-Ogunbi, O. (1985), 'Gastroenteritis due to *Campylobacter jejuni* in Lagos, Nigeria', *Central African Journal of Medicine* **31**, 72–74.

Coker, A. O., Isokpehi, R. D., Thomas, B. N., Amisu, K. O. & Obi, C. L. (2002), 'Human campylobacteriosis in developing countries', *Emerging Infectious Diseases* **8**, 237–244.

Colles, F. M., Dingle, K. E., Cody, A. J. & Maiden, M. C. J. (2008), 'Comparison of *Campylobacter* populations in wild geese with those in starlings and free-range poultry on the same farm', *Applied Environmental Microbiology* **74**, 3583–3590.

Colles, F. M., Jones, K., Harding, R. M. & Maiden, M. C. J. (2003), 'Genetic diversity of *Campylobacter jejuni* isolates from farm animals and the farm environment', *Applied Environmental Microbiology* **69**, 7409–7413.

Colles, F. M., McCarthy, N. D., Howe, J. C., Devereux, C. L., Gosler, A. G. & Maiden, M. C. J. (2009), 'Dynamics of *Campylobacter* colonization of a natural host, *Sturnus vulgaris* (European starling)', *Environmental Microbiology* **11**, 258–267.

Comas, I., Homolka, S., Niemann, S. & Gagneux, S. (2009), 'Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies', *Public Library of Science One* **4**, 1–11.

Cone, L. A., Dreisbach, P. B., Hirschberg, J., Shekar, C., Dreisbach, L. P. & Salatich, W. (2003), 'Cellulitis and septic arthritis caused by *Campylobacter fetus* and *Campylobacter jejuni* : report of 2 cases and review of the literature', *Journal of Clinical Rheumatology* **9**, 362–369.

Cox, M. M. (1998), 'A broadening view of recombinational DNA repair in bacteria', *Genes to Cells* **3**, 65–78.

Cox, N. A., Stern, N. J., Craven, S. E., Berrang, M. E. & Musgrove, M. T. (2000), 'Prevalence of *Campylobacter* and *Salmonella* in the cecal droppings of turkeys during production', *Journal of Applied Poultry Research* **9**, 542–545.

Craven, S. E., Stern, N. J., Line, E., Bailey, J. S., Cox, N. A. & Fedorka-Cray, P. (2000), 'Determination of the incidence of *Salmonella* spp., *Campylobacter jejuni*, and *Clostridium perfringens* in wild birds near broiler chicken houses by sampling intestinal droppings', *Avian Diseases* **44**, 715–720.

Curti, E., Smerdon, S. J. & Davis, E. O. (2007), 'Characterization of the helicase activity and substrate specificity of *Mycobacterium tuberculosis* UvrD', *Journal of Bacteriology* **189**, 1542–1555.

Dadi, L. & Asrat, D. (2009), 'Prevalence and antimicrobial susceptibility profiles of thermotolerant *Campylobacter* strains in retail raw meat products in Ethiopia', *Ethiopia Journal of Health Development* **2**, 195–200.

Damborg, P., Olsen, K. E., Moller Nielsen, E. & Guardabassi, L. (2004), 'Occurrence of *Campylobacter jejuni* in pets living with human patients infected with *C. jejuni*', *Journal of Clinical Microbiology* **42**, 1363–1364.

Danna, K. & Nathans, D. (1971), 'Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Haemophilus influenzae*', *Proceedings of the National Academy of Sciences* **68**, 2913–2917.

Dauga, C. (2002), 'Evolution of the *gyr*B gene and the molecular phylogeny of Enterobacteriaceae: a model molecule for molecular systematic studies', *International Journal of Systematic and Evolutionary Microbiology* **52**, 531–547.

Dawkins, H. J. S. (1989), 'Large DNA separation using field alteration agar gel electrophoresis', *Journal of Chromatography* **492**, 615–639.

de Boer, P., Crossley, R. & Rothfield, L. (1992), 'The essential bacterial cell-division protein *Fts*Z is a GTPase', *Nature* **359**, 254–256.

de Boer, P., Wagenaar, J. A., Achterberg, R. P., van Putten, J. P. M., Schouls, L. M. & Duim, B. (2002), 'Generation of *Campylobacter jejuni* genetic diversity *in vivo*', *Molecular Microbiology* **44**, 351–359.

DEFRA (2000), Zoonoses report United Kingdom 2003, Technical report, Department of Environment and Food and Rural Affairs. `http://archive.defra.gov.uk/foodfarm/farmanimal/diseases/atoz/zoonoses/documents/reports/zoonoses2003.pdf`.

Devane, M. L., Nicol, C., Ball, A., Klena, J. D., Scholes, P., Hudson, J. A., Baker, M. G., Gilpin, B. J., Garrett, N. & Savill, M. G. (2005), 'The occurrence of *Campylobacter* subtypes in environmental reservoirs and potential transmission routes', *Journal of Applied Microbiology* **98**, 980–990.

Dingle, K. E., Colles, F. M., Falush, D. & Maiden, M. C. J. (2005), 'Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*', *Journal of Clinical Microbiology* **43**, 340–347.

Dingle, K. E., Colles, F. M., Ure, R., Wagenaar, J. A., Duim, B., Bolton, F. J., Fox, A. J., Wareing, D. R. A. & Maiden, M. C. J. (2002), 'Molecular characterization of *Campylobacter jejuni* clones: A basis for epidemiologic investigation', *Emerging Infectious Diseases* **8**, 949–955.

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

Dingle, K. E. & Maiden, M. C. (2005), *In Campylobacter : Molecular and Cellular Biology*, J. M. Ketley, M. E. Konkel, Eds. (Horizon Bioscience, Wymondham, Norfolk, UK, 2005), pp. 43-58.

Dingle, K. E., McCarthy, N. D., Cody, A. J., Peto, T. E. A. & Maiden, M. C. J. (2008), 'Extended sequence typing of *Campylobacter* spp., United Kingdom', *Emerging Infectious Diseases* **14**, 1620–1622.

Dionisi, A. M., Luzzi, I. & Carattoli, A. (2004), 'Identification of ciprofloxacin-resistant *Campylobacter jejuni* and analysis of the *gyr*A gene by the LightCycler mutation assay', *Molecular and Cellular Probes* **18**, 255–261.

Doolittle, W. F. (1999), 'Phylogenetic classification and the universal tree', *Science* **284**, 2124–2128.

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

Dowson, C. G., Hutchison, A., Brannigan, J., George, R. C., Hansman, D., Linares, J., Tomasz, A., Maynard Smith, J. & Spratt, B. G. (1989), 'Horizontal transfer of penicillin binding protein genes in penicillin resistant clinical isolates of *Streptococcus pneumoniae*', *Proceedings of National Academy of Science* **86**, 8842–8846.

Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. & Wilson, A. (2010), *Geneious version 5.3*, Available from http://www.geneious.com.

Duim, B., Ang, C. W., van Belkum, A., Rigter, A., van Leeuwen, N. W. J., Endtz, H. P. & Wagenaar, J. A. (2000), 'Amplified fragment length polymorphism analysis of *Campylobacter jejuni* strains isolated from chickens and from patients with gastroenteritis or Guillain-Barre or Miller Fisher syndrome', *Applied Environmental Microbiology* **66**, 3917–3923.

Duim, B., Godschalk, P. C. R., van den Braak, N., Dingle, K. E., Dijkstra, J. R., Leyde, E., van der Plas, J., Colles, F. M., Endtz, H. P., Wagenaar, J. A., Maiden, M. C. J. & van Belkum, A. (2003), 'Molecular evidence for dissemination of unique *Campylobacter jejuni* clones in Curacao, Netherlands Antilles', *Journal of Clinical Microbiology* **41**, 5593–5597.

Duong, T. & Konkel, M. E. (2009), 'Comparative studies of *Campylobacter jejuni* genomic diversity reveal the importance of core and dispensable genes in the biology of this enigmatic food-borne pathogen', *Current Opinion in Biotechnology* **20**, 158–165.

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

ECDC (2009), Annual epidemiological report on communicable diseases in Europe 2009, Technical report, European Centre for Disease Prevention and Control, Stockholm: ECDC. `http://www.ecdc.europa.eu/en/publications/Publications/0910`.

Eisen, J. A. & Hanawalt, P. C. (1999), 'A phylogenomic study of DNA repair genes, proteins, and processes', *Mutation Research* **435**, 171–213.

Endtz, P. H., Ang, C. W., van den Braak, N., Duim, B., Rigter, A., Price, L. J., Woodward, D. L., Rodgers, F. G., Johnson, W. M., Wagenaar, J. A., Jacobs, B. C., Verbrugh, H. A. & van Belkum, A. (2000), 'Molecular characterization of *Campylobacter jejuni* from patients with Guillain-Barre and Miller Fisher syndromes', *Journal of Clinical Microbiology* **38**, 2297–2301.

Engberg, J., On, S. L. W., Harrington, C. S. & Gerner-Smidt, P. (2000), 'Prevalence of *Campylobacter*, *Arcobacter*, *Helicobacter*, and *Sutterella* spp. in human fecal samples

as estimated by a reevaluation of isolation methods for campylobacters', *Journal of Clinical Microbiology* **38**, 286 – 291.

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

Ermolaeva, M. D. (2001), 'Synonymous codon usage in bacteria', *Current Issues in Molecular Biology* **3**, 91–97.

ESR (1993), Environmental Science and Research : Communicable Disease New Zealand, 1992, Technical report, Prepared by the Institute of Environmental Science and Research for the Health Communicable Disease Centre, New Zealand.

ESR (1996), Environmental Science and Research : Health Communicable Disease Centre, Technical report, Prepared by the Institute of Environmental Science and Research for the New Zealand Public Health Report, New Zealand.

Excoffier, L., Laval, G. & Schneider, S. (2005), 'Arlequin (version 3.0): an integrated software package for population genetics data analysis', *Evolution and Bioinformatics Online* **1**, 47–50.

Fall, S., Mercier, A., Bertolla, F., Calteau, A., Gueguen, L., Perrie're, G., Vogel, T. M. & Simonet, P. (2007), 'Horizontal gene transfer regulation in bacteria as a 'Spandrel' of DNA repair mechanisms', *Public Library of Science One* **2**, 1–11.

Fearnhead, P., Smith, N. G. C., Barrigas, M., Fox, A. & French, N. P. (2005), 'Analysis of recombination in *Campylobacter jejuni* from MLST population data', *Journal of Molecular Evolution* **61**, 333–340.

Feil, E. & Chan, M.-S. (2001), *The BURST algorithm (Based Upon Related Sequence Types)*, Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK.

URL: *http://pubmlst.org/analysis/burst/burst.shtml*

Feil, E. J., Holmes, E. C., Bessen, D. E., Chan, M.-S., Day, N. P. J., Enright, M. C., Goldstein, R., Hood, D. W., Kalia, A., Moore, C. E., Zhou, J. & Spratt, B. G. (2001),

'Recombination within natural populations of pathogenic bacteria: Short-term empirical estimates and long-term phylogenetic consequences', *Proceedings of National Academy of Sciences* **98**, 182–187.

Felsenstein, J. (1981), 'Evolutionary trees from DNA sequences: a maximum likelihood approach', *Journal of Molecular Evolution* **17**, 368–376.

Fermer, C. & Engvall, E. O. (1999), 'Specific PCR identification and differentiation of the thermophilic campylobacters, *Campylobacter jejuni, C. coli, C. lari,* and *C. upsaliensis*', *Journal of Clinical Microbiology* **37**, 3370–3373.

Fernandez, H. & Martin, R. (1991), '*Campylobacter* intestinal carriage among stray and pet dogs', *Revista de Saude Publications* **25**, 473–475.

Fitzenberger, J., Uphoff, H., Gawrich, S. & Hauri, A. M. (2010), 'Urban-rural differences of age- and species-specific campylobacteriosis incidence, Hesse, Germany, July 2005 -June 2006', *Eurosurveillance* **15**, 1–7.

Fitzgerald, C., Helsel, L. O., Nicholson, M. A., Olsen, S. J., Swerdlow, D. L., Flahart, R., Sexton, J. & Fields, P. I. (2001), 'Evaluation of methods for subtyping *Campylobacter jejuni* during an outbreak involving a food handler', *Journal of Clinical Microbiology* **39**, 2386–90.

Foerstner, K. U., von Mering, C., Hooper, S. D. & Bork, P. (2005), 'Environments shape the nucleotide composition of genomes', *European Molecular Biology Organization Reports* **6**, 1208–1213.

Fonville, N. C., Bates, D., Hastings, P. J., Hanawalt, P. C. & Rosenberg, S. M. (2010), 'Role of RecA and the SOS response in thymineless death in *Escherichia coli*', *Public Library of Science Genetics* **6**, 1–10.

Foulley, J.-L. & Ollivier, L. (2006), 'Estimating allelic richness and its diversity', *Livestock Science* **101**, 150–158.

Fouts, D. E., Mongodin, E. F., Mandrell, R. E., Miller, W. G., Rasko, D. A., Ravel, J., Brinkac, L. M., DeBoy, R. T., Parker, C. T., Daugherty, S. C., Dodson, R. J., Durkin, A. S., Madupu, R., Sullivan, S. A., Shetty, J. U., Ayodeji, M. A., Shvartsbeyn, A.,

Schatz, M. C., Badger, J. H., Fraser, C. M. & Nelson, K. E. (2005), 'Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species', *Public Library of Science Biology* **3**, 72–85.

Frank-Kamenetskii, M. (1987), 'How the double helix breathes', *Nature* **17-18**, 328.

Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G. & Hanage, W. P. (2009), 'The bacterial species challenge: making sense of genetic and ecological diversity', *Science* **323**, 741–746.

Fraser, C., Hanage, W. P. & Spratt, B. G. (2007), 'Recombination and the nature of bacterial speciation', *Science* **315**, 476–480.

Freese, E. (1962), 'On the evolution of base composition of DNA', *Journal of Theoretical Biology* **3**, 82–101.

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', *Environmental Microbiology* **7**, 1116–1126.

French, N. P. (2008), Enhancing Surveillance of Potentially Foodborne Enteric Diseases in New Zealand : Human campylobacteriosis in the Manawatu, Technical Report Final report: FDI / 236 /2005, Massey University, New Zealand. http://www.foodsafety.govt.nz/elibrary/industry/enhancing-surveillance-potentially-research-projects-2/Campy$_A ttribution_M anawatu.pdf$.

French, N. P., Hotter, G. S. & Biggs, J. P. (2009b), Sequencing the genome of *Campylobacter jejuni* ST- 474 : New Zealand's most important bacterial pathogen of humans.BRCSRA project: RM12767, Technical report, Report prepared for TEC. Epicentre, Molecular Epidemiology and Veterinary Public Health Laboratory, Hopkirk Institute, Massey University.

French, N. P., Marshall, J. C. & Mohan, V. (2010), New and emerging data on typing of *Campylobacter* spp. strains in animals, environmental matrices and humans, Technical Report 07-10436, Molecular Epidemiology and Public Health Laboratory, Hopkirk Re-

search Institute, Institute of Veterinary, Animal and Biomedical Sciences, College of Sciences, Massey University, New Zealand.

French, N. P., Midwinter, A., Holland, B., Collins-Emerson, J., Pattison, R., Colles, F. & Carter, P. (2009a), 'Molecular epidemiology of *Campylobacter jejuni* isolates from wild-bird fecal material in childrens playgrounds', *Applied and Environmental Microbiology* **75**, 779–783.

Fricker, C. R. & Metcalfe, N. (1984), 'Campylobacters in wading birds *(Charadrii)* : incidence, biotypes and isolation techniques', *Zentralblatt fiir Bakteriologie und Hygiene (I. Abt. Originale B)* **179**, 469–475.

Friedman, C. R., Neimann, H. C. J., Wegener, H. C. & Tauxe, R. V. (2000), *Epidemiology of* Campylobacter jejuni *infections in the United States and other industrialized nations*, in i. nachamkin and m. j. blaser campylobacter edn, ASM Press, Washington, D.C.

Frost, J. A. (2001), 'Current epidemiological issues in human campylobacteriosis', *Journal of Applied Microbiology* **90**, 85–95.

Frost, J. A., Kramer, J. M. & Gillanders, S. A. (1999), 'Phage typing of *Campylobacter jejuni* and *Campylobacter coli* and its use as an adjunct to serotyping', *Epidemiology and Infection* **123**, 47–55.

Frost, J. A., Oza, A. N., Thwaites, R. T. & Rowe, B. (1998), 'Serotyping scheme for *Campylobacter jejuni* and *Campylobacter coli* based on direct agglutination of heat-stable antigens', *Journal of Clinical Microbiology* **36**, 335–339.

Fuglsang, A. (2003), 'The genome of *Campylobacter jejuni*: codon and amino acid usage', *Acta Pathologica, Microbiologica et Immunologica Scandinavica* **111**, 605–618.

Fullerton, K. E., Ingram, L. A., Jones, T. F., Anderson, B. J., McCarthy, P. V., Hurd, S., Shiferaw, B., Vugia, D., Haubert, N. & Hayes, T. (2007), 'Sporadic *Campylobacter* infection in infants: a population-based surveillance case-control study', *The Pediatric Infectious Disease Journal* **26**, 19–24.

Gaasbeek, E. J., van der Wal, F. J., van Putten, J. P. M., de Boer, P., van der Graaf-van Bloois, L., de Boer, A. G., Vermaning, B. J. & Wagenaar, J. A. (2009b), 'Functional

characterization of excision repair and *Rec*A-dependent recombinational DNA repair in *Campylobacter jejuni*', *Journal of Bacteriology* **191**, 3785–3793.

Gaasbeek, E. J., Wagenaar, J. A., Guilhabert, M. R., Wosten, M. M. S. M., van Putten, J. P. M., van der Graaf-van Bloois, L., Parker, C. T. & van der Wal, F. J. (2009a), 'A DNase encoded by integrated element CJIE1 inhibits natural transformation of *Campylobacter jejuni*', *Journal of Bacteriology* **191**, 2296–2306.

Galanis, E. (2007), 'Public health : *Campylobacter* and bacterial gastroenteritis', *Canadian Medical Association Journal* **11**, 570–571.

Gallay, A., Simon, F. & Megraud, F. (2003), 'Surveillance of human *Campylobacter* infections in France - part 1 - which data? A study of microbiological laboratories, 2000', *Eurosurveillance* **8**, 213–217.

Galtier, N. & Lobry, J. R. (1997), 'Relationships between genomic G+C content, RNA secondary structures and optimal growth temperature in prokaryotes', *Journal of Molecular Evolution* **44**, 632–636.

Gandhi, N. R., Moll, A., Sturm, A. W., Pawinski, R., Govender, T., Lalloo, U., Zeller, K., Andrews, J. & Friedland, G. (2006), 'Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa', *Lancet* **368**, 1575–1580.

Gaynor, E. C., Cawthraw, S., Manning, G., MacKichan, J. K., Falkow, S. & Newell, D. G. (2004), 'The genome-sequenced variant of *Campylobacter jejuni* NCTC 11168 and the original clonal clinical isolate differ markedly in colonization, gene expression, and virulence-associated phenotypes', *Journal of Bacteriology* **186**, 503–517.

Genereux, D. P. (2002), 'Evolution of genomic GC variation', *Genome Biology* **3**, 58.

Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., Stackebrandt, E., de Peer, Y. V., Vandamme, P., Thompson, F. L. & Swings, J. (2005), 'Re-evaluating prokaryotic species', *Nature Review Microbiology* **3**, 733–739.

Gibson, J. R., Sutherland, K. & Owen, R. J. (1994), 'Inhibition of DNAse activity in PFGE analysis of DNA from *Campylobacter jejuni*', *Letters in Applied Microbiology* **19**, 357–358.

Gilbert, M., Mandrell, R. E., Parker, C. T., Li, J. & Vinogradov, E. (2007), 'Structural analysis of the capsular polysaccharide from *Campylobacter jejuni* RM1221', *ChemBioChem* **8**, 625–631.

Gillen, J. R., Willis, D. K. & Clark, A. J. (1981), 'Genetic analysis of the *Rec*E pathway of genetic recombination in *Escherichia coli* K-12', *Journal of Bacteriology* **145**, 521–532.

Gillespie, I. A., O'Brian, S. J. & Bolton, F. J. (2009), 'Age patterns of persons with campylobacteriosis, England and Wales, 1990-2007', *Emerging Infectious Diseases* **15**, 2046–2048.

Gillespie, I. A., O'Brian, S. J., Frost, J. A., Adak, G. K., Horby, P., Swan, A. V., Painter, M. J., Neal, K. R. & the *Campylobacter* Sentinel Surveillance Scheme Collaborators (2002), 'A case-case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: a tool for generating hypotheses', *Emerging Infectious Diseases* **8**, 937–942.

Gillespie, I. A., O'Brian, S. J., Penman, C., Tomkins, D., Cowden, J., Humphrey, T. J. & on behalf of The *Campylobacter* Sentinel Surveillance Scheme Collaborators (2008), 'Demographic determinants for *Campylobacter* infection in England and Wales: Implications for future epidemiological studies', *Epidemiology and Infection* **136**, 1717–1725.

Giovannoni, S. (2004), 'Evolutionary biology: Oceans of bacteria', *Nature* **430**, 415–416.

Gong, C., Bongiorno, P., Martins, A., Stephanou, N. C., Zhu, H., Shuman, S. & Glickman, M. S. (2005), 'Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C', *Nature Structural and Molecular Biology* **12**, 304–312.

Goossens, H., Pot, B., Vlaes, L., Van den Borre, C., Van den Abbeele, R., Van Naelten, C., Levy, J., Cogniau, H., Marbehant, P. & Verhoef, J. (1990), 'Characterization and description of "*Campylobacter upsaliensis*" isolated from human feces', *Journal of clinical microbiology* **28**, 1039–46.

Gorkiewicz, G., Feierl, G., Zechner, R. & Zechner, E. L. (2002), 'Transmission of *Campylobacter hyointestinalis* from a pig to a human', *Journal of Clinical Microbiology* **40**, 2601–2605.

Gormley, F. J., MacRae, M., Forbes, K. J., Ogden, I. D., Dallas, J. F. & Strachan, N. J. C. (2008), 'Has retail chicken played a role in the decline of human campylobacteriosis?', *Applied and Environmental Microbiology* **74**, 383–390.

Gow, G. A., Gow, D. J., Hall, E. J., Langton, D., Clarke, C. & Papasouliotis, K. (2009), 'Prevalence of potentially pathogenic enteric organisms in clinically healthy kittens in the UK', *Journal of Feline Medicine and Surgery* **11**, 655–662.

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

Grant, A. J., Coward, C., Jones, M. A., Woodall, C. A., Barrow, P. A. & Maskell, D. J. (2005), 'Signature-tagged transposon mutagenesis studies demonstrate the dynamic nature of cecal colonization of 2-week-old chickens by *Campylobacter jejuni*', *Applied and Environmental Microbiology* **71**, 8031–8041.

Gressmann, H., Linz, B., Ghai, R., Pleissner, K.-P., Schlapbach, R., Yamaoka, Y., Kraft, C., Suerbaum, S., Meyer, T. F. & Achtman, M. (2005), 'Gain and loss of multiple genes during the evolution of *Helicobacter pylori*', *Proceedings of National Academy of Science : Genetics* **1**, 1:e43.

Griggs, D. J., Johnson, M. M., Frost, J. A., Humphrey, T., Jorgensen, F. & Piddock, L. J. V. (2005), 'Incidence and mechanism of ciprofloxacin resistance in *Campylobacter* spp. isolated from commercial poultry flocks in the United Kingdom before, during, and after fluoroquinolone treatment', *Antimicrobial Agents and Chemotherapy* **49**, 699–707.

Grosjean, H., de Crecy-Lagard, V. & Marck, C. (2010), 'Deciphering synonymous codons in the three domains of life: Co-evolution with specific tRNA modification enzymes', *FEBS Letters* **584**, 252–264.

Gueron, M., Kochoyan, M. & Leroy, J.-L. (1987), 'A single mode of DNA base-pair opening drives imino proton exchange', *Nature* **328**, 89–92.

Guerry, P., Logan, S. M., Thornton, S. & Trust, T. J. (1990), 'Genomic organization and expression of *Campylobacter* flagellin genes', *Journal of Bacteriology* **172**, 1853–1860.

Guerry, P., Pope, P. M., Burr, D. H., Leifer, J., Joseph, S. W. & Bourgeois, A. L. (1994), 'Development and characterization of *rec*A mutants of *Campylobacter jejuni* for inclusion in attenuated vaccines', *Infection and Immunity* **62**, 426–432.

Gundogdu, O., Bentley, S. D., Holden, M. T., Parkhill, J., Dorrell, N. & Wren, B. W. (2007), 'Re-annotation and re-analysis of the *Campylobacter jejuni* NCTC11168 genome sequence', *BioMedCentral Genomics* **8**, 162.

Guo, B., Wang, Y., Shi, F., Barton, Y.-W., Plummer, P., Reynolds, D. L., Nettleton, D., Grinnage-Pulley, T., Lin, J. & Zhang, Q. (2008), '*Cme*R functions as a pleiotropic regulator and is required for optimal colonization of *Campylobacter jejuni in vivo*', *Journal of Bacteriology* **190**, 1879–1890.

Gupta, R. S. (2006), 'Molecular signatures (unique proteins and conserved indels) that are specific for the epsilon proteobacteria *(Campylobacterales)*', *BioMedCentral Genomics* **7**, 2164–2167.

Habib, I., Uyttendaele, M. & De Zutter, L. (2010), 'Survival of poultry-derived *Campylobacter jejuni* of multilocus sequence type clonal complexes 21 and 45 under freeze, chill, oxidative, acid and heat stresses', *Food Microbiology* **27**, 829–834.

Hakanen, A., Jalava, J., Kotilainen, P., Jousimies-Somer, H. J., Siitonen, A. & Huovinen, P. (2002), '*gyr*A polymorphism in *Campylobacter jejuni*: detection of *gyr*A mutations in 162 *C. jejuni* isolates by single-strand conformation polymorphism and DNA sequencing', *Antimicrobial Agents and Chemotherapy* **46**, 2644–2647.

Hald, B. & Madsen, M. (1997), 'Healthy puppies and kittens as carriers of *Campylobacter* spp., with special reference to *Campylobacter upsaliensis*', *Journal of Clinical Microbiology* **35**, 3351–3352.

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

Hald, T. & Brondsted, T. (2000), Annual report on zoonoses in Denmark 2000, Technical report, Ministry of Food, Agriculture and Fisheries, Copenhagen, Denmark.

Hammer, O. (2009-2011), *PAST, PAleontological STatistics, Version 2.08. Reference manual.*

Han, J., Sahin, O., Barton, Y. W. & Zhang, Q. (2008), 'Key role of *Mfd* in the development of fluoroquinolone resistance in *Campylobacter jejuni*', *Public Library of Science Pathogens* **4**, e1000083.

Hanage, W. P., Fraser, C. & Spratt, B. G. (2005), 'Fuzzy species among recombinogenic bacteria', *BioMedCentral Biology* **6**, 1–7.

Hanage, W. P., Fraser, C. & Spratt, B. G. (2006), 'Sequences, sequence clusters and bacterial species', *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 1917–1927.

Hansson, I., Engvall, E. O., Lindblad, J., Gunnarsson, A. & Vagsholm, I. (2004), 'Surveillance programme for *Campylobacter* species in Swedish broilers, July 2001 to June 2002', *Veterinary Record* **155**, 193–196.

Hansson, I., Vagsholm, I., Svensson, L. & Engvall, E. O. (2007), 'Correlations between *Campylobacter* spp. prevalence in the environment and broiler flocks', *Journal of Applied Microbiology* **103**, 640–649.

Harrington, C. S., Thomson-Carter, F. M. & Carter, P. E. (1997), 'Evidence for recombination in the flagellin locus of *Campylobacter jejuni*: Implications for the flagellin gene typing scheme', *Journal of Clinical Microbiology* **35**, 2386–2395.

Hasegawa, M., Kishino, H. & Yano, T. (1985), 'Dating the human-ape splitting by a molecular clock of mitochondrial DNA', *Journal of Molecular Evolution* **22**, 160–174.

Hasegawa, M., Yano, T. & Kishino, H. (1984), 'A new molecular clock of mitochondrial DNA and the evolution of Hominoids', *Proceedings of Japanese Academy B* **60**, 95–98.

Hatfield, D. L. & Gladyshev, V. N. (2002), 'How selenium has altered our understanding of the genetic code', *Molecular and Cellular Biology* **22**, 3565–3576.

Hearnden, M., Skelly, C., Eyles, R. & Weinstein, P. (2003), 'The regionality of campylobacteriosis seasonality in New Zealand', *International Journal of Environmental Health Research* **13**, 337 – 348.

Heinemann, J. A. & Sprague, G. F. J. (1989), 'Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast', *Nature* **340**, 205–209.

Hernandez, J., Fayos, A., Ferrus, M. A. & Owen, R. J. (1995), 'Random amplified polymorphic DNA fingerprinting of *Campylobacter jejuni* and *C. coli* isolated from human faeces, seawater and poultry products', *Research in Microbiology* **146**, 685–696.

Herve, J., Aissa, N., Legrand, P., Sorkine, M., Calmette, M. J., Santin, A., Roupie, E. & Renaud, B. (2004), '*Campylobacter fetus* meningitis in a diabetic adult cured by imipenem', *European Journal of Clinical Microbiology and Infectious Diseases* **23**, 722–724.

Heyworth, J. S., Cutt, H. & Glonek, G. (2006), 'Does dog or cat ownership lead to increased gastroenteritis in young children in South Australia?', *Epidemiology and Infection* **134**, 926–934.

Higuchi, R., Dollinger, G., Walsh, P. S. & Griffith, R. (1992), 'Simultaneous amplification and detection of specific DNA sequences', *Nature Biotechnology* **10**, 413–417.

Hildebrand, F., Meyer, A. & Eyre-Walker, A. (2010), 'Evidence of selection upon genomic GC-content in bacteria', *Public Library of Science Genetics* **6**, e1001107.

Hinode, D., Yokoyama, M., Tanabe, S., Yoshioka, M. & Nakamura, R. (2002), 'Antigenic properties of the *gro*EL- like protein of *Campylobacter rectus*', *Oral Microbiology and Immunology* **17**, 16–21.

Hoe, N., Nakashima, K., Grigsby, D., Pan, X., Dou, S. J., Naidich, S., Garcia, M., Kahn, E., Bergmire-Sweat, D. & Musser, J. M. (1999), 'Rapid molecular genetic subtyping of serotype M1 group A *Streptococcus* strains', *Emerging Infectious Diseases* **43**, 254–263.

Hofreuter, D., Tsai, J., Watson, R. O., Novik, V., Altman, B., Benitez, M., Clark, C., Perbost, C., Jarvie, T., Du, L. & Galan, J. E. (2006), 'Unique features of a highly pathogenic *Campylobacter jejuni* strain', *Infection and Immunology* **74**, 4694 – 4707.

Holsinger, K. E. & Weir, B. S. (2009), 'Genetics in geographically structured populations: defining, estimating and interpreting Fst', *Nature Review Genetics* **10**, 639–650.

Horii, Z.-I. & Clark, A. J. (1973), 'Genetic analysis of the *rec*F pathway to genetic recombination in *Escherichia coli* K12: isolation and characterization of mutants', *Journal of Molecular Biology* **80**, 327–344.

Horvath, P., Romero, D. A., Coute-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C. & Barrangou, R. (2008), 'Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*', *Journal of Bacteriology* **190**, 1401–1412.

Huang, S., Sahin, O. & Zhang, Q. (2007), 'Infection-induced antibodies against the major outer membrane protein of *Campylobacter jejuni* mainly recognize conformational epitopes', *FEMS Microbiology Letters* **272**, 137–143.

Hudson, R. R. & Kaplan, N. L. (1985), 'Statistical properties of the number of recombination events in the history of a sample of DNA sequences', *Genetics* **111**, 147–164.

Hughes, A. L., Friedman, R., Rivailler, P. & French, J. O. (2008), 'Synonymous and nonsynonymous polymorphisms versus divergences in bacterial genomes', *Molecular Biology and Evolution* **25**, 2199–2209.

Hughes, L. A., Bennett, M., Coffey, P., Elliott, J., Jones, T. R., Jones, R. C., Lahuerta-Marin, A., Leatherbarrow, A. H., McNiffe, K., Norman, D., Williams, N. J. & Chantrey, J. (2009), 'Molecular epidemiology and characterisation of *Campylobacter* spp. isolated from wild bird populations in northern England', *Applied Environmental Microbiology* **75**, 3007–3015.

Humphrey, T. J., Henley, A. & Lanning, D. G. (1993), 'The colonization of broiler chickens with *Campylobacter jejuni*: some epidemiological investigations.', *Epidemiology and Infection* **110**, 601–607.

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

Hurst, L. D. & Merchant, A. R. (2001), 'High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes', *Proceedings of Biological Science* **268**, 493–497.

Huson, D. H. & Bryant, D. (2006), 'Application of phylogenetic networks in evolutionary studies', *Molecular Biology and Evolution* **23**, 254–267.

Huysmans, M. B. & Turnidge, J. D. (1997), 'Disc susceptibility testing for thermomphilic campylobacters', *Pathology* **29**, 209–216.

Hwang, M.-N. & Ederer, G. M. (1975), 'Rapid hippurate hydrolysis method for presumptive identification of group B Streptococci', *Journal of Clinical Microbiology* **1**, 114–115.

Ikemura, T. (1985), 'Codon usage and tRNA content in unicellular and multicellular organisms', *Molecular Biology and Evolution* **2**, 13–34.

Jackson, C. J., Fox, A. J., Wareing, D. R. A., Hutchinson, D. N. & Jones, D. M. (1996), 'The application of genotyping techniques to the epidemiological analysis of *Campylobacter jejuni*', *Epidemiology and Infection* **117**, 233–244.

Jacobs, M. A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C., Levy, R., Chun-Rong, L., Guenthner, D., Bovee, D., Olson, M. V. & Manoil, C. (2003), 'Comprehensive transposon mutant library of *Pseudomonas aeruginosa*', *Proceedings of National Academy of Science* **100**, 14339–14344.

Jacobs-Reitsma, W. F., Bolder, N. M. & Mulder, R. W. (1994), 'Cecal carriage of *Campylobacter* and *Salmonella* in Dutch broiler flocks at slaughter: a one-year study', *Poultry Science* **73**, 1260–1266.

Jain, R., Rivera, M. C. & Lake, J. A. (1999), 'Horizontal gene transfer among genomes: The complexity hypothesis', *Proceedings of the National Academy of Sciences* **96**, 3801–3806.

Janion, C. (2008), 'Inducible SOS response system of DNA repair and mutagenesis in *Escherichia coli*', *International Journal of Biological Sciences* **4**, 338–344.

Jeon, B. & Zhang, Q. (2007), 'Cj0011c, a periplasmic single- and double-stranded DNA-binding protein, contributes to natural transformation in *Campylobacter jejuni*', *Journal of Bacteriology* **189**, 7399–7407.

Jiggins, F. M., von Der Schulenburg, J. H., Hurst, G. D. & Majerus, M. E. (2001), 'Recombination confounds interpretations of *Wolbachia* evolution', *Proceedings of Royal Soceity London* **268**, 1423–1427.

Jimenez, S. G., Heine, R. G., Ward, P. B. & Robins-Browne, R. M. (1999), '*Campylobacter upsaliensis* gastroenteritis in childhood.', *The Pediatric Infectious Disease Journal* **18**, 988–992.

Jin, L. & Nei, M. (1990), 'Limitations of the evolutionary parsimony method of phylogenetic analysis', *Molecular Biology and Evolution* **7**, 82–102.

Jolley, K. A., Feil, E. J., Chan, M.-S. & Maiden, M. C. J. (2001), 'Sequence type analysis and recombinational tests (START)', *Bioinformatics* **17**, 1230–1231.

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992), 'The rapid generation of mutation data matrices from protein sequences', *Computer Applications in the Biosciences* **8**, 275–282.

Jones, K. (2001), 'Campylobacters in water, sewage and the environment', *Journal of Applied Microbiology* **90**, 68S–79S.

Jones, K., Betaieb, M. & Telford, D. R. (1990), 'Thermophilic campylobacters in surface waters around Lancaster, UK : negative correlation with *Campylobacter* infections in the community', *Journal of Applied Bacteriology* **69**, 758–764.

Jordan, I. K., Makarova, K. S., Spouge, J. L., Wolf, Y. I. & Koonin, E. V. (2001), 'Lineage-specific gene expansions in bacterial and archaeal genomes', *Genome Research* **11**, 555–565.

Jukes, T. H. & Cantor, C. R. (1969), *Evolution of protein molecules*, in H. N. Munro, ed. Mammalian protein metabolism, Academic Press, New York.

Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M. & van Embden, J. (1997), 'Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology', *Journal of Clinical Microbiology* **35**, 907–914.

Kanaya, S., Yamada, Y., Kudo, Y. & Ikemura, T. (1999), 'Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene

expression level and species-specific diversity of codon usage based on multivariate analysis', *Gene* **238**, 143–155.

Karenlampi, R. I., Tolvanen, T. P. & Hanninen, M. L. (2004), 'Phylogenetic analysis and PCR-restriction fragment length polymorphism identification of *Campylobacter* species based on partial *gro*EL gene sequences', *Journal of Clinical Microbiology* **42**, 5731–5738.

Karenlampi, R., Rautelin, H., Schonberg-Norio, D., Paulin, L. & Hanninen, M.-L. (2007), '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', *Applied and Environmental Microbiology* **73**, 148–155.

Karita, M., Etterbeek, M. L., Forsyth, M. H., Tummuru, M. K. & Blaser, M. J. (1997), 'Characterization of *Helicobacter pylori dap*E and construction of a conditionally lethal *dap*E mutant', *Infection and Immunity* **65**, 4158–4164.

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

Kelly, D. J. (2001), 'The physiology and metabolism of *Campylobacter jejuni* and *Helicobacter pylori*', *Journal of Applied Microbiology* **90**, 16S–24S.

Khakhria, R. & Lior, H. (1992), 'Extended phage-typing scheme for *Campylobacter jejuni* and *Campylobacter coli*', *Epidemiology and Infection* **108**, 403–414.

Kimura, M. (1968), 'Evolutionary rate at the molecular level', *Nature* **217**, 624–626.

Kimura, M. (1980), 'A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences', *Journal of Molecular Evolution* **16**, 111–120.

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

Kinscherf, T. G. & Willis, D. K. (2002), 'Global regulation by *gid*A in *Pseudomonas syringae*', *Journal of Bacteriology* **184**, 2281–2286.

Klancnik, A., Botteldoorn, N., Herman, L. & Mozina, S. S. (2006), 'Survival and stress induced expression of *gro*EL and *rpo*D of *Campylobacter jejuni* from different growth phases', *International Journal of Food Microbiology* **112**, 200–207.

Klena, J. (2001), A survey of phenotypic and genetic methods used to identify and differentiate thermotolerant *Campylobacter* spp. strains, Technical report, A report to the Ministry of Health . Department of Plant and Microbial Sciences. University of Canterbury.

Koene, M. G. J., Houwers, D. J., Dijkstra, J. R., Duim, B. & Wagenaar, J. A. (2004), 'Simultaneous presence of multiple *Campylobacter* species in dogs', *Journal of Clinical Microbiology* **42**, 819–821.

Kolodner, R., Fishel, R. A. & Howard, M. (1985), 'Genetic recombination of bacterial plasmid DNA: effect of RecF pathway mutations on plasmid recombination in *Escherichia coli*', *Journal of Bacteriology* **163**, 1060–1066.

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

Krause, R., Ramschak-Schwarzer, S., Gorkiewicz, G., Schnedl, W. J., Feierl, G., Wenisch, C. & Reisinger, E. C. (2002), 'Recurrent septicemia due to *Campylobacter fetus* and *Campylobacter lari* in an immunocompetent patient', *Infection* **30**, 171–174.

Krwawicz, J., Arczewska, K. D., Speina, E., Maciejewska, A. & Grzesiuk, E. (2007), 'Bacterial DNA repair genes and their eukaryotic homologues: 1. Mutations in genes involved in base excision repair (BER) and DNA-end processors and their implication in mutagenesis and human disease', *Acta Biochimica Polonica* **54**, 413–434.

Kryukov, G. V. & Gladyshev, V. N. (2004), 'The prokaryotic selenoproteome', *European Molecular Biology Organization Reports* **5**, 538–543.

Ku, B. K., Kim, H. J., Lee, Y. J., Kim, Y. I., Choi, J. S., Park, M. Y., Kwon, J. W., Nam, H.-M., Kim, Y. H., Jung, S.-C., Lee, S. J., Kim, S. H. & Kim, J. H. (2011), 'Genetic

characterization and antimicrobial susceptibility of *Campylobacter* spp. isolated from domestic and imported chicken meats and humans in Korea', *Foodborne Pathogens and Disease* **8**, 381–386.

Kwan, P. S. L., Barrigas, M., Bolton, F. J., French, N. P., Gowland, P., Kemp, R., Leatherbarrow, H., Upton, M. & Fox, A. J. (2008), 'Molecular epidemiology of *Campylobacter jejuni* populations in dairy cattle, wildlife, and the environment in a farmland area', *Applied Environmental Microbiology* **74**, 5130–5138.

Labahn, J., Scharer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L. & Ellenberger, T. E. (1996), 'Structural basis for the excision repair of alkylation-damaged DNA', *Cell* **86**, 321–329.

Labarca, J. A., Sturgeon, J., Borenstein, L., Salem, N., Harvey, S. M., Lehnkering, E., Reporter, R. & Mascola, L. (2002), '*Campylobacter upsaliensis*: Another pathogen for consideration in the United States', *Clinical Infectious Diseases* **34**, e59 – 60.

Labesse, G., Garnotel, E., Bonnel, S., Dumas, C., Pages, J. M. & Bolla, J. M. (2001), 'MOMP, a divergent porin from *Campylobacter* : cloning and primary structural characterization', *Biochemical and Biophysical Research Communications* **280**, 380–387.

Lake, R., Baker, M., Nicol, C. & Garrett, N. (2004), 'Lack of association between long-term illness and infectious intestinal disease in New Zealand', *New Zealand Medical Journal* **117**, 1194.

Lastovica, A. J. & Le Roux, E. (2003), 'Optimal detection of *Campylobacter* spp. in stools', *Journal of Clinical Pathology* **56**, 480.

Lawson, A. J., Logan, J. M., O'Neill, G. L., Desai, M. & Stanley, J. (1999), 'Large-scale survey of *Campylobacter* species in human gastroenteritis by PCR and PCR-enzyme linked immunosorbent assay', *Journal of Clinical Microbiology* **37**, 3860–3864.

Leberg, P. L. (2002), 'Estimating allelic richness: effects of sample size and bottlenecks', *Molecular Ecology* **11**, 2445–2449.

Leclere, M. M., Nishioka, M., Yuasa, T., Fujiwara, S., Takagi, M. & Imanaka, T. (1998), 'The O 6-methylguanine-DNA methyltransferase from the hyperthermophilic archaeon

Pyrococcus sp. KOD1: a thermostable repair enzyme', *Molecular and General Genetics* **258**, 69–77.

Lederberg, J. & Tatum, E. L. (1946), 'Gene recombination in *Escherichia coli*', *Nature* **158**, 558.

Lee, M. D. & Newell, D. G. (2006), '*Campylobacter* in poultry: filling an ecological niche', *Avian Diseases* **50**, 1–9.

Lefebure, T. & Stanhope, M. J. (2009), 'Pervasive, genome-wide positive selection leading to functional divergence in the bacterial genus *Campylobacter*', *Genome Research* **19**, 1224–1232.

Lentzsch, P., Rieksneuwohner, B., Wieler, L. H., Hotzel, H. & Moser, I. (2004), 'High-resolution genotyping of *Campylobacter upsaliensis* strains originating from three continents', *Journal of Clinical Microbiology* **42**, 3441–3448.

Leonard, E. E., Takata, T., Blaser, M. J., Falkow, S., Tompkins, L. S. & Gaynor, E. C. (2003), 'Use of an open-reading frame-specific *Campylobacter jejuni* DNA microarray as a new genotyping tool for studying epidemiologically related isolates', *The Journal of Infectious Diseases* **187**, 691–694.

Levy, A. J. (1946), 'A gastro-enteritis outbreak probably due to a bovine strain of vibrio', *Yale Journal of Biology and Medicine* **18**, 243–258.

Levy, P. S. & Lemeshow, S. (1999), *Sampling of Populations Methods and Applications*, Wiley Series in Probability and Statistics, London.

Librado, P. & Rozas, J. (2009), 'DnaSP v5: A software for comprehensive analysis of DNA polymorphism data', *Bioinformatics* **25**, 1451–1452.

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

Linz, B., Schenker, M., Zhu, P. & Achtman, M. (2000), 'Frequent interspecific genetic exchange between commensal neisseriae and *Neisseria meningitidis*', *Molecular Microbiology* **36**, 1049–1058.

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

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

Liu, H., Kovacs, E. & Lund, P. A. (2009), 'Characterisation of mutations in *gro*ES that allow *gro*EL to function as a single ring', *FEBS Letters* **583**, 2365–2371.

Lloyd, R. G. & Buckman, C. (1985), 'Identification and genetic analysis of sbcC mutations in commonly used recBC sbcB strains of *Escherichia coli* K-12', *Journal of Bacteriology* **164**, 836–844.

Logan, S. M., Trust, T. J. & Guerry, P. (1989), 'Evidence for post translational modification and gene duplication of *Campylobacter* flagellin', *Journal of Bacteriology* **171**, 3031–3038.

Lohman, T. M. & Ferrari, M. E. (1994), '*Escherichia coli* single-stranded DNA binding protein: multiple DNA-binding modes and cooperativities', *Annual Reviews of Biochemistry* **63**, 527–570.

Lopez, C. M., Giacoboni, G., Agostini, A., Cornero, F. J., Tellechea, D. M. & Trinidad, J. J. (2002), 'Thermotolerant campylobacters in domestic animals in a defined population in Buenos Aires, Argentina', *Preventative Veterinary Medicine* **55**, 193–200.

Louis, V. R., Gillespie, I. A., O'Brien, S. J., Russek-Cohen, E., Pearson, A. D. & Colwell, R. R. (2005), 'Temperature-driven *Campylobacter* seasonality in England and Wales', *Applied and Environmental Microbiology* **71**, 85–92.

Lovett, S. T., Hurley, R. L., Sutera, V. A., Jr, Aubuchon, R. H., & Lebedeva, M. A. (2002), 'Crossing over between regions of limited homology in *Escherichia coli*. RecA dependent and RecA-independent pathways', *Genetics* **160**, 851–859.

Lovett, S. T. & Kolodner, R. D. (1989), 'Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the *rec*J gene of *Escherichia coli*', *Proceedings of National Academy of Sciences* **86**, 2627–2631.

Low, S. C. & Berry, M. J. (1996), 'Knowing when not to stop: selenocysteine incorporation in eukaryotes', *Trends in Biochemical Science* **21**, 203–208.

Luangtongkum, T., Morishita, T. Y., El-Tayeb, A. B., Ison, A. J. & Zhang, Q. (2007), 'Comparison of antimicrobial susceptibility testing of *Campylobacter* spp. by the agar dilution and the agar disk diffusion methods', *Journal of Clinical Microbiology* **45**, 590–594.

Luechtefeld, N. A., Blaser, M. J., Reller, L. B. & Wang, W. L. (1980), 'Isolation of *Campylobacter fetus* subsp. *jejuni* from migratory waterfowl', *Journal of Clinical Microbiology* **12**, 406–408.

Lund, M., Nordentoft, S., Pedersen, K. & Madsen, M. (2004), 'Detection of *Campylobacter* spp. in chicken fecal samples by real-time PCR', *Journal of Clinical Microbiology* **42**, 5125–5132.

Luo, N., Sahin, O., Lin, J., Michel, L. O. & Zhang, Q. (2003), 'In vivo selection of *Campylobacter* isolates with high levels of fluoroquinolone resistance associated with *gyr*A mutations and the function of the cmeabc efflux pump', *Antimicrobial Agents and Chemotherapy* **47**, 390–394.

Lusetti, S. L. & Cox, M. M. (2002), 'The bacterial RecA protein and the recombinational DNA repair of stalled replication forks', *Annual Review of Biochemistry* **71**, 71–100.

Lynn, D. J., Singer, G. A. C. & Hickey, D. A. (2002), 'Synonymous codon usage is subject to selection in thermophilic bacteria', *Nucleic Acids Research* **30**, 4272–4277.

Ma, H. W. & Zeng, A. P. (2004), 'Phylogenetic comparison of metabolic capacities of organisms at genome level', *Molecular Phylogenetics and Evolution* **31**, 204–213.

Maddukuri, L., Dudzinska, D. & Tudek, B. (2007), 'Bacterial DNA repair genes and their eukaryotic homologues: 4. The role of nucleotide excision DNA repair (NER) system in mammalian cells', *Acta Biochimica Polonica* **54**, 469–482.

Mahadevappa, M. & Warrington, J. A. (2002), 'House keeping genes', *Encyclopedia of life sciences* **1**, 1–7.

Mahdi, A. A. & Lloyd, R. G. (1989), 'Identification of the *rec*R locus of *Escherichia coli* K-12 and analysis of its role in recombination and DNA repair', *Molecular and General Genetics* **216**, 503–510.

Maiden, M. C. J. (2006), 'Multilocus sequence typing of bacteria', *Annual Review of Microbiology* **60**, 561–588.

Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D. A., Feavers, I. M., Achtman, M. & Spratt, B. G. (1998), 'Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms', *Proceedings of the National Academy of Sciences* **95**, 3140–3145.

Malik-Kale, P., Parker, C. T. & Konkel, M. E. (2008), 'Culture of *Campylobacter jejuni* with sodium deoxycholate induces virulence gene expression', *Journal of Bacteriology* **190**, 2286–2297.

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

Marais, G., Mouchiroud, D. & Duret, L. (2001), 'Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes', *Proceedings of National Academy of Science* **98**, 5688–5692.

Marcus, R. (2008), 'New information about pediatric foodborne infections: the view from FoodNet', *Current Opinion in Pediatrics* **20**, 79 –84.

Margos, G., Gatewood, A. G., Aanensen, D. M., Hanincova, K., Terekhova, D., Vollmer, S. A., Cornet, M., Piesman, J., Donaghy, M., Bormane, A., Hurn, M. A., Feil, E. J., Fish, D., Casjens, S., Wormser, G. P., Schwartz, I. & Kurtenbach, K. (2008), 'MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*', *Proceedings of National Academy of Science* **24**, 8730–8735.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, Joseph, M.,

Gomes, X. V., Goodwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J.-B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F. & Rothberg, J. M. (2005), 'Genome sequencing in microfabricated high density picolitre reactors', *Nature* **437**, 376–380.

Marler, C. (2010), 'Surveillance and analysis on *Campylobacter* news and outbreaks'. `http:MarlerClarkNetwork://www.dailymail.co.uk/health/article-1350028/Food-poisoning-warning-Bacteria-OUTSIDE-chicken-packaging`.

Martin, W. (1999), 'Mosaic bacterial chromosomes: a challenge on route to a tree of genomes', *BioEssays* **21**, 99–104.

Martinez, J. J., Seveau, S., Veiga, E., Matsuyama, S. & Cossart, P. (2005), 'Ku70, a component of DNA-dependent protein kinase, is a mammalian receptor for *Rickettsia conorii*', *Cell* **123**, 1013–1023.

Matsuda, M., Tai, K., Moore, J. E., Millar, B. C. & Murayama, O. (2004), 'Nucleotide sequence analysis of the *rec*A gene and discrimination of the three isolates of urease-positive thermophilic *Campylobacter* (UPTC) isolated from seagulls (*Larus* spp.) in Northern Ireland', *Journal of Basic Microbiology* **44**, 215–219.

Mau, B., Glasner, J. D., Darling, A. E. & Perna, N. T. (2006), 'Genome-wide detection and analysis of homologous recombination among sequenced strains of *Escherichia coli*', *Genome Biology* **7**, R44.

Maxam, A. M. & Gilbert, W. (1977), 'A new method for sequencing DNA', *Proceedings of National Academy of Sciences : Biochemistry* **74**, 560–564.

McCarthy, N. D., Colles, F. M., Dingle, K. E., Bagnall, M. C., Manning, G., Maiden, M. C. J. & Falush, D. (2007), 'Host-associated genetic import in *Campylobacter jejuni*', *Emerging Infectious Diseases* **13**, 267–272.

McCarthy, N. & Giesecke, J. (2001), 'Incidence of Guillain-Barre syndrome following infection with *Campylobacter jejuni*', *American Journal of Epidemiology* **153**, 610–614.

McDonald, J. H. & Kreitman, M. (1991), 'Adaptive protein evolution at the *Adh* locus in *Drosophila*', *Nature* **351**, 652–654.

McEwan, C. E. A., Gatherer, D. & McEwan, N. R. (1998), 'Nitrogen-fixing aerobic bacteria have higher genomic GC content than non-fixing species within the same genus', *Hereditas* **128**, 173–178.

McFadyean, J. & Stockman, S. (1913), Abortion in sheep : Report of the Departmental Committee appointed by the Board of Agriculture and Fisheries to enquire into epizootic abortion, Technical Report Appendix Part II, 64 pp., His Majesty's Stationary Office, London.

McIver, C., Hogan, T., White, P. & Tapsall, J. (2004), 'Patterns of quinolone susceptibility in *Campylobacter jejuni* associated with different *gyr*A mutations', *Pathology* **36**, 166–169.

McKay, D., Fletcher, J., Cooper, P. & Thomson-Carter, F. M. (2001), 'Comparison of two methods for serotyping *Campylobacter* spp.', *Journal of Clinical Microbiology* **39**, 1917–1921.

McQuiston, J. R., Herrera-Leon, S., Wertheim, B. C., Doyle, J., Fields, P. I., Tauxe, R. V. & Logsdon, J. M. (2008), 'Molecular phylogeny of the *Salmonellae* : relationships among *Salmonella* species and subspecies determined from four housekeeping genes and evidence of lateral gene transfer events', *Journal of Bacteriology* **190**, 7060–7067.

McTavish, S. M., Pope, C. E., Nicol, C., Sexton, K., French, N. & Carter, P. E. (2008), 'Wide geographical distribution of internationally rare *Campylobacter* clones within New Zealand', *Epidemiology and Infection* **136**, 1244–1252.

Meinersmann, R. J., Helsel, O. L., Fields, P. I. & Hiett, K. L. (1997), 'Discrimination of *Campylobacter jejuni* isolates by *fla* gene sequencing', *Journal of Clinical Microbiology* **35**, 2810–2814.

Meinersmann, R. J. & Hiett, K. L. (2000), 'Concerted evolution of duplicate *fla* genes in *Campylobacter*', *Microbiology* **146**, 2283–2290.

Meinersmann, R. J., Patton, C. M., Evins, G. M., Wachsmuth, I. K. & Fields, P. I. (2002), 'Genetic diversity and relationships of *Campylobacter* species and subspecies', *International Journal of Systematic and Evolutionary Microbiology* **52**, 1789–1797.

Melissa, V. & Michael, L. (2010), *Campylobacter* enteritis, Technical report, Centres for Disease Control and Prevention : Chapter 5 : Travellers Health Yellow Book. `http://wwwnc.cdc.gov/travel/yellowbook/2010/chapter-5/campylobacter-enteritis.htm`.

Mendz, G. L., Ball, G. E. & Meek, D. J. (1997), 'Pyruvate metabolism in *Campylobacter* spp.', *Biochimica et Biophysica Acta* **1334**, 291–302.

Meunier, J. R. & Grimont, P. A. D. (1993), 'Factors affecting reproducibility of random amplified polymorphic DNA fingerprinting', *Research in Microbiology* **144**, 373–379.

Michel, B. (2005), 'After 30 years of study, the bacterial SOS response still surprises us', *Public Library of Science Biology* **3**, 1174–1176.

Mikulik, K., Man, P. & Halada, P. (2001), 'Characterization of the *rpl*B gene from *Streptomyces collinus* and its protein product by mass spectrometry', *Biochemical and Biophysical Research Communications* **285**, 1344–1349.

Miller, W. G., Englen, M. D., Kathariou, S., Wesley, I. V., Wang, G., Pittenger-Alley, L., Siletz, R. M., Muraoka, W., Fedorka-Cray, P. J. & Mandrell, R. E. (2006), 'Identification of host-associated alleles by multilocus sequence typing of *Campylobacter coli* strains from food animals', *Microbiology* **152**, 245–255.

Miller, W. G., On, S. L. W., Wang, G., Fontanoz, S., Lastovica, A. J. & Mandrell, R. E. (2005), 'Extended multilocus sequence typing system for *Campylobacter coli*, *C. lari*, *C. upsaliensis*, and *C. helveticus*', *Journal of Clinical Microbiology* **43**, 2315–2329.

Miller, W. G., Parker, C. T., Rubenfield, M., Mendz, G. L., Wosten, M. M. S. M., Ussery, D. W., Stolz, J. F., Binnewies, T. T., Hallin, P. F., Wang, G., Malek, J. A., Rogosin, A., Stanker, L. H. & Mandrell, R. E. (2007), 'The complete genome sequence and analysis of the epsilonproteobacterium *Arcobacter butzleri*', *Public Library of Science One* **12**, 1–21.

Minet, J., Grosbois, B., & Megraud, F. (1988), '*Campylobacter hyointestinalis* : an opportunistic enteropathogen?', *Journal of Clinical Microbiology* **26**, 2659–2660.

Minin, V. N., Dorman, K. S., Fang, F. & Suchard, M. A. (2005), 'Dual multiple change-point model leads to more accurate recombination detection', *Bioinformatics* **21**, 3034–3042.

Monno, R., Rendina, M., Ceci, G., Rizzo, C., Luzzi, I., Francavilla, A., Rizzo, G. & Ierardi, E. (2004), '*Campylobacter fetus* bacteremia in an immunocompromised patient: case report and review of the literature', *New Microbiologica* **27**, 281–285.

Moran, A. P. & Penner, J. L. (1999), 'Serotyping of *Campylobacter jejuni* based on heat stable antigens: relevance, molecular basis and implications in pathogenesis', *Journal of Applied Microbiology* **86**, 361–377.

Moreno, G. S., Griffiths, P. L., Conerton, I. F. & Park, R. W. A. (1993), 'Occurrence of campylobacters in small domestic and laboratory animals', *Journal of Applied Bacteriology* **75**, 49–54.

Morris, G. K., el Sherbeeny, M. R., Patton, C. M., Kodaka, H., Lombard, G. L., Edmonds, P., Hollis, D. G. & Brenner, D. J. (1985), 'Comparison of four hippurate hydrolysis methods for identification of thermophilic *Campylobacter* spp.', *Journal of Clinical Microbiology* **22**, 714–718.

Moser, I., Rieksneuwhner, B., Lentzsch, P., Schwerk, P., & Wieler, L. H. (2001), 'Genomic heterogeneity and O-antigenic diversity of *Campylobacter upsaliensis* and *Campylobacter helveticus* strains isolated from dogs and cats in Germany', *Journal of Clinical Microbiology* **39**, 2548–2557.

Moser, I., Schroeder, W. & Salnikow, J. (1997), '*Campylobacter jejuni* major outer membrane protein and a 59-kDa protein are involved in binding to fibronectin and INT 407 cell membranes', *FEMS Microbiology Letters* **157**, 233–238.

Mullner, P., Collins-Emerson, J. M., Midwinter, A. C., Carter, P., Spencer, S. E. F., van der Logt, P., Hathaway, S. & French, N. P. (2010), 'Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry', *Applied and Environmental Microbiology* **76**, 2145–2154.

Mullner, P., Jones, G., Noble, A., Spencer, S. E. F., Hathaway, S. & French, N. P. (2009), 'Source attribution of food-borne zoonoses in New Zealand: a modified Hald model', *Risk Analysis* **29**(7), 970–984.

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

Muse, S. V. & Gaut, B. S. (1994), 'A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome', *Molecular Biology and Evolution* **11**, 715–724.

Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valin, F., & Bernardi, G. (2006), 'Genomic GC level, optimal growth temperature, and genome size in prokaryotes', *Biochemical and Biophysical Research Communications* **347**, 1–3.

Muto, A. & Osawa, S. (1987), 'The guanine and cytosine content of genomic DNA and bacterial evolution', *Proceedings of National Academy of Science* **84**, 166–169.

Nachamkin, I. & Blaser, M. J. (2000), Campylobacter, third edition edn, American Society for Microbiology, ASM Press, Washington, DC, USA.

Nair, G. B., Sarkar, R. K., Chowdhury, S. & Pal, S. C. (1985), '*Campylobacter* infection in domestic dogs', *Veterinary Record* **116**, 237–238.

Nakamura, Y., Itoh, T., Matsuda, H. & Gojobori, T. (2004), 'Biased biological functions of horizontally transferred genes in prokaryotic genomes', *Nature Genetics* **36**, 760–766.

Naya, H., Romero, H., Zavala, A., Alvarez, B. & Musto, H. (2002), 'Aerobiosis increases the genomic guanine plus cytosine content (GC%) in prokaryotes', *Journal of Molecular Evolution* **55**, 260–264.

Nei, M. & Gojobori, T. (1986), 'Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions', *Molecular Biology and Evolution* **3**, 418–426.

Nesbakken, T., Eckner, K., Hoidal, H. K. & Rotterud, O.-J. (2002), 'Occurrence of *Yersinia enterocolitica* and *Campylobacter* spp. in slaughter pigs and consequences for meat inspection, slaughtering and dressing procedures', *International Journal of Food Microbiology* **80**, 231–240.

Newell, D. G. (2001), 'Animal models of *Campylobacter jejuni* colonization and disease and the lessons to be learned from similar *Helicobacter pylori* models', *Journal of Applied Microbiology* **90**, 57S–67S.

Newell, D. G. (2002), 'The ecology of *Campylobacter jejuni* in avian and human hosts and in the environment', *International Journal of Infectious Diseases* **6**, S16–S21.

Newell, D. G., Duim, B., Bergen, M. A. P. v., Grogono-Thomas, R. & Wagenaar, J. A. (2000), 'Speciation, subspeciation and subtyping of *Campylobacter* spp. associated with bovine infertility and abortion', *Cattle Practice* **8**, 421–425.

Nicol, C., King, N., Pirie, R. & Dufour, M. (2010), Diagnostic and public health management practices of foodborne bacterial diseases, Technical report, ESR : Report prepared for the New Zealand Food Safety Authority under project MRP/09/05 as part of overall contract for scientific services.

Nielsen, E. M., Fussing, V., Engberg, J., Nielsen, N. L. & Neimann, J. (2006), 'Most *Campylobacter* subtypes from sporadic infections can be found in retail poultry products and food animals', *Epidemiology and Infection* **134**, 758–767.

Nielsen, R. & Yang, Z. (1998), 'Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene', *Genetics* **148**, 929–936.

Nieminuszczy, J. & Grzesiuk, E. (2007), 'Bacterial DNA repair genes and their eukaryotic homologues: 3. *Alk*b dioxygenase and *Ada* methyltransferase in the direct repair of alkylated DNA', *Acta Biochimica Polonica* **54**, 459–468.

Notermans, S. H. (1994), Epidemiology and surveillance of campylobacter infections, Technical Report pp. 35-44, In Report on a WHO consultation on epidemiology and control of campylobacteriosis, Bilthoven: WHO.

Nowosielska, A. (2007), 'Bacterial DNA repair genes and their eukaryotic homologues: 5. The role of recombination in DNA repair and genome stability', *Acta Biochimica Polonica* **54**, 483–494.

Nuijten, P. J., van Asten, F. J., Gaastra, W. & van der Zeijst, B. A. (1990), 'Structural and functional analysis of two *Campylobacter jejuni* flagellin genes', *Journal of Biological Chemistry* **265**, 17798–17804.

Nuijten, P. J., van den Berg, A. J., Formentini, I., van der Zeijst, B. A. & Jacobs, A. A. (2000), 'DNA rearrangements in the flagellin locus of an *fla*A mutant of *Campylobacter jejuni* during colonization of chicken ceca', *Infection and Immunity* **68**, 7137–7140.

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

NZFSA (2008), *Campylobacter* Risk Management Strategy 2008 - 2011, Technical report, New Zealand Food Safety Authority. `www.nzfsa.govt.nz/.../strategy/Campylobacter-risk-management-strategy-2008-2011`.

Oberhelman, R. A. & Taylor, D. N. (2000), Campylobacter *infections in developing countries: In* Campylobacter, second edn, American Society for Microbiology, Washington, DC, USA.

Obiri-Danso, K. & Jones, K. (1999), 'Distribution and seasonality of microbial indicators and thermophilic campylobacters in two freshwater bathing sites on the River Lune in northwest England', *Journal of Applied Microbiology* **87**, 822–832.

O'Brien, P. J. (2006), 'Catalytic promiscuity and the divergent evolution of DNA repair enzymes', *Chemical Reviews* **106**, 720–752.

Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000), 'Lateral gene transfer and the nature of bacterial innovation', *Nature* **405**, 299–304.

Odermatt, P., Gautsch, S., Rechsteiner, D., Ewald, R., Haag-Wackernagel, D., Muhlemann, R. & Tanner, M. (1998), 'Swarms of starlings in Basel: a natural phenomenon, a nuisance or a health risk?', *Gesundheitswesen (In German; abstract in English)* **60**, 749–754.

Ogden, I. D., MacRae, M., Johnston, M., Strachan, N. J. C., Cody, A. J., Dingle, K. E. & Newell, D. G. (2007), 'Use of multilocus sequence typing to investigate the association between the presence of *Campylobacter* spp. in broiler drinking water and *Campylobacter* colonization in broilers', *Applied and Environmental Microbiology* **73**, 5125–5129.

Ohta, T. (2002), 'Near-neutrality in evolution of genes and gene regulation', *Proceedings of National Academy of Science* **99**, 16134–16137.

Olsen, G. J. & Woese, C. R. (1993), 'Ribosomal RNA : a key to phylogeny', *Federation of American Societies for Experimental Biology Journal* **7**, 113–123.

On, S. L. (1996), 'Identification methods for campylobacters, helicobacters, and related organisms', *Clinical Microbiology Reviews* **9**, 405–422.

Owen, R. J., Beck, A. & Borman, P. (1985), 'Restriction endonuclease digest patterns of chromosomal DNA from nitrate-negative *Campylobacter jejuni*-like organisms', *European Journal of Epidemiology* **1**, 281–287.

Owen, R. J. & Hernandez, J. (1990), 'Occurrence of plasmids in *'Campylobacter upsaliensis'* (catalase negative or weak group) from geographically diverse patients with gastroenteritis or bacteraemia', *European Journal of Epidemiology* **6**, 111–117.

Oyarzabal, O. A., Conner, D. E. & Hoerr, F. J. (1995), 'Incidence of campylobacters in the intestine of avian species in Alabama', *Avian Diseases* **39**, 147–151.

Oyarzabal, O. A., Wesley, I. V., Barbaree, J. M., Lauerman, L. H. & Conner, D. E. (1997), 'Specific detection of *Campylobacter lari* by PCR', *Journal of Microbiological Methods* **9**, 97–102.

Oza, A. N., McKenna, J. P., McDowell, S. W. J., Menzies, F. D. & Neill, S. D. (2003), 'Antimicrobial susceptibility of *Campylobacter* spp. isolated from broiler chickens in Northern Ireland', *Journal of Antimicrobial Chemotherapy* **52**, 220–223.

Page, W. J., Huyer, G., Huyer, M. & Worobec, E. (1989), 'Characterization of the porins of *Campylobacter jejuni* and *Campylobacter coli* and implications for antibiotic susceptibility', *Antimicrobial Agents and Chemotherapy* **33**, 297–303.

Paques, F. & Haber, J. E. (1999), 'Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*', *Microbiology and Molecular Biology Reviews* **63**, 349–404.

Park, Y. H. & Kohel, R. J. (1994), 'Effect of concentration of MgCl2 on random-amplified DNA polymorphism', *BioTechniques* **16**, 652–656.

Parker, C. T., Quinones, B., Miller, W. G., Horn, S. T. & Mandrell, R. E. (2006), 'Comparative genomic analysis of *Campylobacter jejuni* strains reveals diversity due to genomic

elements similar to those present in *C. jejuni* strain RM1221', *Journal of Clinical Microbiology* **44**, 4125–4135.

Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., Basham, D., Chillingworth, T., Davies, R. M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A. V., Moule, S., Pallen, M. J., Penn, C. W., Quail, M. A., Rajandream, M.-A., Rutherford, K. M., van Vliet, A. H. M., Whitehead, S. & Barrell, B. G. (2000), 'The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences', *Nature* **403**, 665–668.

Parsons, B. N., Cody, A. J., Porter, C. J., Stavisky, J. H., Smith, J. L., Williams, N. J., Leatherbarrow, A. J. H., Hart, C. A., Gaskell, R. M., Dingle, K. E. & Dawson, S. (2009), 'Typing of *Campylobacter jejuni* isolates from dogs by use of multilocus sequence typing and pulsed-field gel electrophoresis', *Journal of Clinical Microbiology* **47**, 3466–3471.

Patel, M., Jiang, Q., Woodgate, R., Cox, M. M. & Goodman, M. F. (2010), 'A new model for SOS-induced mutagenesis: how RecA protein activates DNA polymerase V', *Critical Reviews in Biochemistry and Molecular Biology* **45**, 171–184.

Patthy, L. (2008), *Evolution of protein-coding genes. In Protein evolution*, second edn, Blackwell publishing.

Pearson, A. D., Greenwood, M., Healing, T. D., Rollins, D., Shahamat, M., Donaldson, J. & Colwell, R. R. (1993), 'Colonization of broiler chickens by waterborne *Campylobacter jejuni*', *Applied and Environmental Microbiology* **59**, 987–996.

Pearson, B. M., Pin, C., Wright, J., I'Anson, K., Humphrey, T. & Wells, J. M. (2003), 'Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays', *Federation of European Biochemical Societies Letters* **554**, 224–230.

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

Persson, S. & Olsen, K. E. P. (2005), 'Multiplex PCR for identification of *Campylobacter coli* and *Campylobacter jejuni* from pure cultures and directly on stool samples', *Journal of Medical Microbiology* **54**, 1043–1047.

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

Price, E. P., Huygens, F. & Giffard, P. M. (2006), 'Fingerprinting of *Campylobacter jejuni* using resolution-optimized binary gene targets derived from comparative genome hybridization studies', *Applied and Environmental Microbiology* **72**, 7793–7803.

Prim, R. C. (1957), 'Shortest connection networks and some generalizations', *Bell System Technology Journal* **36**, 1389–1401.

Puhler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H. P., Lottspeich, F., Garrett, R. A. & Zillig, W. (1989), 'Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome', *Proceedings of National Academy of Science* **86**, 4569–4573.

Radman, M. (1975), 'SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis', *Basic Life Sciences* **5A**, 355–367.

Raji, M. A., Adekeye, J. O., Kwaga, J. K. & Bale, J. O. (2000), 'Bioserogroups of *Campylobacter* species isolated from sheep in Kaduna State, Nigeria', *Small Ruminant Research* **37**, 215–221.

R: A Language and Environment for Statistical Computing, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2010, ISBN 3-900051-07-0, http://www.R-project.org (n.d.).

Reeves, P. (1993), 'Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale', *Trends in Genetics* **9**, 17–22.

Reid, A. N., Pandey, R., Palyada, K., Whitworth, L., Doukhanine, E. & Stintzi, A. (2008), 'Identification of *Campylobacter jejuni* genes contributing to acid adaptation by transcriptional profiling and genome-wide mutagenesis', *Applied and Environmental Microbiology* **74**, 1598–1612.

Ribot, E. M., Fitzgerald, C., Kubota, K., Swaminathan, B. & Barrett, T. J. (2001), 'Rapid pulsed-field gel electrophoresis protocol for subtyping of *Campylobacter jejuni*', *Journal of Clinical Microbiology* **39**, 1889–1894.

Riley, L. W. (2004), *Molecular epidemiology of infectious diseases: principles and practices : Chapter II. Laboratory methods used for strain typing of pathogens: Conventional and Molecular techniques*, ASM Press, Washington, DC, USA.

Riley, M. (1993), 'Functions of the gene products of *Escherichia coli*', *Microbiology and Molecular Biology Reviews* **57**, 862–952.

Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. (1998), 'Genomic evidence for two functionally distinct gene classes', *Proceedings of National Academy of Science* **95**, 6239–6244.

Robertson, G. A., Thiruvenkataswamy, V., Shilling, H., Price, E. P., Huygens, F., Henskens, F. A. & Giffard, P. M. (2004), 'Identification and interrogation of highly informative single nucleotide polymorphism sets defined by bacterial multilocus sequence typing databases', *Journal of Medical Microbiology* **53**, 35–45.

Rocchini, D., Ricotta, C., Chiarucci, A., de Dominicis, V., Cirillo, I. & Macherrini, S. (2009), 'Relating spectral and species diversity through rarefaction curves', *International Journal of Remote Sensing* **30**, 2705–2711.

Rocha, E. P. C. (2008), 'Evolutionary patterns in prokaryotic genomes', *Current Opinion in Microbiology* **11**, 454–460.

Rocha, E. P. C. & Danchin, A. (2002), 'Base composition bias might result from competition for metabolic resources', *Trends in Genetics* **18**, 291–294.

Rocha, E. P. C. & Feil, E. J. (2010), 'Mutational patterns cannot explain genome composition: are there any neutral sites in the genomes of bacteria?', *Public Library of Science Genetics* **6**, e1001104.

Rodrigo, S., Adesiyun, A., Asgarali, Z. & Swanston, W. (2007), 'Antimicrobial resistance of *Campylobacter* spp. isolated from broilers in small poultry processing operations in Trinidad', *Food Control* **18**, 321–325.

Romero, H., Zhang, Y., Gladyshev, V. N. & Salinas, G. (2005), 'Evolution of selenium utilization traits', *Genome Biology* **6**, R66.

Rooney, A. P., Swezey, J. L., Friedman, R., Hecht, D. W. & Maddox, C. W. (2006), 'Analysis of core housekeeping and virulence genes reveals cryptic lineages of *Clostridium perfringens* that are associated with distinct disease presentations', *Genetics* **172**, 2081–2092.

Rossi, M., Debruyne, L., Zanoni, R. G., Manfreda, G., Revez, J. & Vandamme, P. (2009), '*Campylobacter avium* sp. nov., a hippurate-positive species isolated from poultry', *International Journal of Systematic and Evolutionary Microbiology* **59**, 2364–2369.

Rozen, Y. & Belkin, S. (2001), 'Survival of enteric bacteria in seawater', *FEMS Microbiology Reviews* **513-529**, 513–529.

Rudi, K., Hoidal, H. K., Katla, T., Johansen, B. K., Nordal, J. & Jakobsen, K. S. (2004), 'Direct real-time PCR quantification of *Campylobacter jejuni* in chicken fecal and cecal samples by integrated cell concentration and DNA purification', *Applied and Environmental Microbiology* **70**, 790–797.

Rudi, K., Zimonja, M. & Skanseng, B. (2006), 'Quencher extension for single nucleotide polymorphism quantification in bacterial typing and microbial community analyses', *Methods in Molecular Biology* **345**, 111–117.

Saeed, A. M., Harris, N. V. & DiGiacomo, R. F. (1993), 'The role of exposure to animals in the etiology of *Campylobacter jejuni / coli* enteritis', *Account American Journal of Epidemiology* **137**, 108–114.

Sahin, O., Plummer, P. J., Jordan, D. M., Sulaj, K., Pereira, S., Robbe-Austerman, S., Wang, L., Yaeger, M. J., Hoffman, L. J. & Zhang, Q. (2008), 'Emergence of a tetracycline-resistant *Campylobacter jejuni* clone associated with outbreaks of ovine abortion in the United States', *Journal of Clinical Microbiology* **46**, 1663–1671.

Said, M. M., Mohamady, H. E., Beih, F. M. E., Rockabrand, D. M., Ismail, T. F., Monteville, M. R., Ahmed, S. F., Klena, J. D. & Salama, M. S. (2010), 'Detection of *gyr*A mutation among clinical isolates of *Campylobacter jejuni* isolated in Egypt by MAMA-PCR', *The Journal of Infection in Developing Countries* **4**, 546–554.

Sails, A. D., Fox, A. J., Bolton, F. J., Wareing, D. R. A. & Greenway, D. L. A. (2003a), 'A real-time PCR assay for the detection of *Campylobacter jejuni* in foods after enrichment culture', *Applied and Environmental Microbiology* **69**, 1383–1390.

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

Sails, A. D., Wareing, D. R. A., Bolton, F. J., Fox, A. J. & Curry, A. (1998), 'Characterisation of 16 *Campylobacter jejuni* and *C. coli* typing bacteriophages', *Journal of Medical Microbiology* **47**, 123–128.

Sanchez, M. X., Fluckey, W. M., Brashears, M. M. & McKee, S. R. (2002), 'Microbial profile and antibiotic susceptibility of *Campylobacter* spp. and *Salmonella* spp. in broilers processed in air-chilled and immersion-chilled environments', *Journal of Food Protection* **65**, 948–956.

Sandberg, M., Bergsjo, B., Hofshagen, M., Skjerve, E. & Kruse, H. (2002), 'Risk factors for *Campylobacter* infection in Norwegian cats and dogs.', *Preventive Veterinary Medicine* **55**, 241–253.

Sanger, F. & Coulson, A. R. (1975), 'A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase', *Journal of Molecular Biology* **74**, 441–448.

Savill, M., Hudson, A., Devane, M., Garrett, N., Gilpin, B. & Ball, A. (2003), 'Elucidation of potential transmission routes of *Campylobacter* in New Zealand', *Water Science and Technology* **47**, 33–38.

Schouls, L. M., Reulen, S., Duim, B., Wagenaar, J. A., Willems, R. J. L., Dingle, K. E., Colles, F. M. & Van Embden, J. D. (2003), 'Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and recombination', *Journal of Clinical Microbiology* **41**, 15–26.

Schumann, W. (2005), *Dynamics of the bacterial chromosome. Structure and function*, 5 edn, Wiley-VCH Verlag GmbH & Co., Weinheim, Germany.

Schuster, S. C. (2008), 'Next-generation sequencing transforms today's biology', *Nature methods* **5**, 16–18.

Scott, W. G., Scott, H. M., Lake, R. J. & Baker, M. G. (2000), 'Economic cost to New Zealand of foodborne infectious disease', *The New Zealand Medical Journal* **113**, 281–284.

Sebald, E. R. & Veron, M. (1963), 'Base DNA content and classification of vibrios (In French; Teneur en bases da lADN et classification des Vibrions)', *Annales de L'institut Pasteur (Paris)* **105**, 897–910.

Segreti, J., Gootz, T. D., Goodman, L. J., Parkhurst, G. W., Quinn, J. P., Martin, B. A. & Trenholme, G. M. (1992), 'Highlevel quinolone resistance in clinical isolates of *Campylobacter jejuni*', *Journal of Infectious Diseases* **165**, 667–670.

Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N. & Whittam, T. S. (1986), 'Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics', *Applied and Environmental Microbiology* **51**, 873–884.

Sellars, M. J., Hall, S. J. & Kelly, D. J. (2002), 'Growth of *Campylobacter jejuni* supported by respiration of fumarate, nitrate, nitrite, trimethylamine-N-oxide, or dimethyl sulfoxide requires oxygen', *Journal of Bacteriology* **184**, 4187–4196.

Shane, S. M. (1992), 'The significance of *Campylobacter jejuni* infection in poultry: A review', *Avian Pathology* **21**, 189–213.

Sheppard, S. K., Colles, F., Richardson, J., Cody, A. J., Elson, R., Lawson, A., Brick, G., Meldrum, R., Little, C. L., Owen, R. J., Maiden, M. C. J. & McCarthy, N. D. (2010), 'Host association of *Campylobacter* genotypes transcends geographic variation', *Applied and Environmental Microbiology* **76**, 5269–5277.

Sheppard, S. K., Dallas, J. F., Strachan, N. J. C., MacRae, M., McCarthy, N. D., Wilson, D. J., Gormley, F. J., Falush, D., Ogden, I. D., Maiden, M. C. J. & Forbes, K. J. (2009), '*Campylobacter* genotyping to determine the source of human infection', *Clinical Infectious Diseases* **48**, 1072–1078.

Sheppard, S. K., McCarthy, N. D., Falush, D. & Maiden, M. C. J. (2008), 'Convergence of *Campylobacter* species: implications for bacterial evolution', *Science* **320**, 237–239.

Sheppard, S. K., McCarthy, N. D., Jolley, K. A. & Maiden, M. C. J. (2011), 'Introgression in the genus *Campylobacter*: generation and spread of mosaic alleles', *Microbiology* **157**, 1066–1074.

Sheppard, S., McCarthy, N., Maiden, M. C. J., Elson, R., Little, C. L., Owen, R. & Meldrum, R. (2007), Multilocus sequence typing and analysis of *Campylobacter* isolates from the current FSA retail poultry surveys, Technical report, Oxford University. `http://www.foodbase.org.uk/admintools/reportdocuments/513-1-903_Technical_Report_B15011.pdf`.

Shields, D. C., Sharp, P. M., Higgins, D. G. & Wright, F. (1988), ''Silent' sites in Drosophila genes are not neutral: evidence of selection among synonymous codons', *Molecular Biology and Evolution* **5**, 704–716.

Shuman, S. & Glickman, M. S. (2007), 'Bacterial DNA repair by non-homologous end joining', *Nature Reviews Microbiology* **5**, 852–861.

Siegel, A. F. & Gelman, R. Z. (1982), 'Rarefaction and taxonomic diversity', *Biometrics* **38**, 235–241.

Siemer, B. L., Nielsen, E. M. & On, S. L. W. (2005), 'Identification and molecular epidemiology of *Campylobacter coli* isolates from human gastroenteritis, food, and animal sources by amplified fragment length polymorphism analysis and penner serotyping', *Applied and Environmental Microbiology* **71**, 1953–1958.

Singh, S. K., Roy, S., Choudhury, S. R. & Sengupta, D. N. (2010), 'DNA repair and recombination in higher plants: insights from comparative genomics of arabidopsis and rice', *BioMedCentral Genomics* **11**, 443–455.

Sinha, K. M., Stephanou, N. C., Gao, F., Glickman, M. S. & Shuman, S. (2007), 'Mycobacterial UvrD1 is a Ku-dependent DNA helicase that plays a role in multiple DNA repair events, including double-strand break repair', *The Journal of Biological Chemistry* **282**, 15114–15125.

Skaar, E. P., Lazio, M. P. & Seifert, H. S. (2002), 'Roles of the *rec*J and *rec*N genes in homologous recombination and DNA repair pathways of *Neisseria gonorrhoeae*', *Journal of Bacteriology* **184**, 919–927.

Skirrow, M. B. (1994), 'Diseases due to *Campylobacter*, *Helicobacter* and related bacteria', *Journal of Comparative Pathology* **111**, 113–149.

Skirrow, M. B. & Benjamin, J. (1980), ''1001' Campylobacters: cultural characteristics of intestinal campylobacters from man and animals', *Journal of Hygiene* **85**, 427–442.

Skorvaga, M., Raven, N. D. H. & Margison, G. P. (1998), 'Thermostable archaeal O6-alkylguanine-DNA alkyltransferases', *Proceedings of National Academy of Sciences* **95**, 6711–6715.

Slader, J., Domingue, G., Jorgensen, F., McAlpine, K., Owen, R. J., Bolton, F. J. & Humphrey, T. J. (2002), 'Impact of transport crate reuse and of catching and processing on *Campylobacter* and *Salmonella* contamination of broiler chickens', *Applied and Environmental Microbiology* **68**, 713–719.

Smith, H. O. & Wilcox, K. W. (1970), 'A restriction enzyme from *Haemophilus influenzae*. I. Purification and general properties', *Journal of Molecular Biology* **51**, 379–391.

Smith, J. L. (1995), 'Arthritis, Guillain-Barre syndrome, and other sequelae of *Campylobacter jejuni* enteritis', *Journal of Food Protection* **58**, 1153–1170.

Sneyd, E. & Baker, M. (2003), Infectious diseases in New Zealand: 2002 annual surveillance summary, Technical report, Prepared by the Institute of Environmental Science and Research Limited (ESR). Wellington, New Zealand.

Snyder, L. & Champness, W. (1997), *Molecular genetics of bacteria*, 3 edn, American Society for Microbiology, Wasington, DC, USA.

Sopwith, W., Birtles, A., Matthews, M., Fox, A., Gee, S., Painter, M., Regan, M., Syed, Q. & Bolton, E. (2006), '*Campylobacter jejuni* multilocus sequence types in humans, northwest England, 2003-2004', *Emerging Infectious Diseases* **12**, 1500–1507.

Sopwith, W., Birtles, A., Matthews, M., Fox, A., Gee, S., Painter, M., Regan, M., Syed, Q. & Bolton, E. (2008), 'Identification of potential environmentally adapted *Campylobacter jejuni* strain, United Kingdom', *Emerging Infectious Diseases* **14**, 1769–1773.

Sorek, R., Zhu, Y., Creevey, C. J., Francino, M. P., Bork, P. & Rubin, E. M. (2007), 'Genome-wide experimental determination of barriers to horizontal gene transfer', *Science* **318**, 1449–1452.

Spratt, B. G. (2004), 'Exploring the concept of clonality in bacteria', *Methods in Molecular Biology* **266**, 323–352.

Spratt, B. G., Hanage, W. P. & Feil, E. J. (2001), 'The relative contributions of recombination and point mutation to the diversification of bacterial clones', *Current Opinion in Microbiology* **4**, 602–606.

Stafford, R. J., Schluter, P. J., Wilson, A. J., Kirk, M. D., Hall, G., Unicomb, L. & the OzFoodNet Working Group (2008), 'Population-attributable risk estimates for risk factors associated with *Campylobacter* infection, Australia', *Emerging Infectious Diseases* **6**, 895–901.

Stahl, M. & Stintzi, A. (2011), 'Identification of essential genes in *C. jejuni* genome highlights hyper-variable plasticity regions', *Functional and Integrative Genomics* **11**, 241–257.

Stanley, J., Burnens, A. P., Linton, D. P., On, S. L. W., Costas, M. & Owen, R. J. (1992), '*Campylobacter helveticus* sp. nov., a new thermophilic species from domestic animals: characterization and cloning of a species-specific DNA probe', *Journal of Genetic Microbiology* **138**, 2293–2303.

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

Stanley, K. N., Wallace, J. S., Currie, J. E., Diggle, P. J. & Jones, K. (1998*b*), 'Seasonal variation of thermophilic campylobacters in lambs at slaughter', *Journal of Applied Microbiology* **84**, 1111–1116.

Steinhauserova, I., Fojtikova, K. & Klimes, J. (2000), 'The incidence and PCR detection of *Campylobacter upsaliensis* in dogs and cats', *Letters in Applied Microbiology* **31**, 209–212.

Struelens, M. J. & Members of European Study Group on Epidemiological Markers (ESGEM) of the European Society for Clinical Microbiology & Infectious Diseases (ESCMID) (1996), 'Consensus guidelines for appropriate use and evaluation of microbial epidemiological typing systems', *Clinical microbiology and infection* **2**, 2–11.

Stucki, U., Frey, J., Nicolet, J. & Burnens, A. P. (1995), 'Identification of *Campylobacter jejuni* on the basis of a species- specific gene that encodes a membrane protein', *Journal of Clinical Microbiology* **33**, 855–859.

Studer, E., Luthy, J. & Hubner, P. (1999), 'Study of the presence of *Campylobacter jejuni* and *C. coli* in sand samples from four Swiss chicken farms', *Research in Microbiology* **150**, 213–219.

Sueoka, N. (1961), 'Correlation between base composition of deoxyribonucleic acid and amino acid composition of protein', *Proceedings of National Academy of Science* **47**, 1141–1149.

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

Susko, E., Field, C., Blouin, C. & Roger, A. J. (2003), 'Estimation of rates-across-sites distributions in phylogenetic substitution models', *Systematic Biology* **52**, 594.

Suzuki, Y. & Gojobori, T. (1999), 'A method for detecting positive selection at single amino acid sites', *Molecular Biology and Evolution* **16**, 1315–1328.

Swaminathan, B., Barrett, T. J., Hunter, S. B., Tauxe, R. V. & CDC PulseNet Task Force (2001), 'PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States', *Emerging Infectious Diseases* **7**, 382–389.

Swingle, B., Markel, E., Costantino, N., Bubunenko, M. G., Cartinhour, S. & Court, D. L. (2010), 'Oligonucleotide recombination in gram-negative bacteria', *Molecular Microbiology* **75**, 138–148.

Taboada, E. N., Acedillo, R. R., Carrillo, C. D., Findlay, W. A., Medeiros, D. T., Mykytczuk, O. L., Roberts, M. J., Valencia, C. A., Farber, J. M. & Nash, J. H. E. (2004), 'Large-scale

comparative genomics meta-analysis of *Campylobacter jejuni* isolates reveals low level of genome plasticity', *Journal of Clinical Microbiology* **42**, 4566–4576.

Taboada, E. N., Acedillo, R. R., Luebbert, C. C., Findlay, W. A. & Nash, J. H. E. (2005), 'A new approach for the analysis of bacterial microarray-based comparative genomic hybridization: insights from an empirical study', *BioMed Central Genomics* **6**, 65–78.

Tajima, F. (1989), 'Statistical method for testing the neutral mutation hypothesis by DNA polymorphism', *Genetics* **123**, 585–595.

Tamura, K. & Dudley, J. Nei M & Kumar, S. (2007), 'MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 5.0', *Molecular Biology and Evolution* **24**, 1596–1599.

Tamura, K. & Nei, M. (1993), 'Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees', *Molecular Biology and Evolution* **10**, 512–528.

Tanaka, T. & Nei, M. (1989), 'Positive Darwinian selection observed at the variable-region genes of immunoglobulins', *Molecular Biology and Evolution* **6**, 447–459.

Tauxe, R. V. (1992), Campylobacter jejuni *: current and future trends : Epidemiology of* Campylobacter jejuni *infections in the United States and other industrial nations*, American Society for Microbiology Press, Washington, DC, USA.

Taylor, D. E. & Chau, A. S. S. (1997), 'Cloning and nucleotide sequence of the *gyr*A gene from *Campylobacter fetus* subsp. *fetus* ATCC 27374 and characterization of ciprofloxacin-resistant laboratory and clinical isolates', *Antimicrobial Agents and Chemotherapy* **41**, 665–671.

Taylor, D. E., Eaton, M., Yan, W. & Chang, N. (1992), 'Genome maps of *Campylobacter jejuni* and *Campylobacter coli*', *Journal of Bacteriology* **174**, 2332–2337.

Tenkate, T. D. & Stafford, R. J. (2001), 'Risk factors for *Campylobacter* infection in infants and young children: a matched case-control study', *Epidemiology and Infection* **127**, 399 – 404.

Tenover, F. C., Arbeit, R. D., Goering, R. V., Mickelsen, P. A., Murray, B. E., Persing, D. H. & Swaminathan, B. (1995), 'Interpreting chromosomal DNA restriction patterns produced by pulsed- field gel electrophoresis : criteria for bacterial strain typing', *Journal of Clinical Microbiology* **33**, 2233–2239.

Thomas, C., Hill, D. J. & Mabey, M. (1999), 'Morphological changes of synchronized *Campylobacter jejuni* populations during growth in single phase liquid culture', *Letters in Applied Microbiology* **28**, 194–198.

Thomson, G. (1922), *The naturalisation of animals and plants in New Zealand Part II. Naturalisation of animals*, Cambridge University Press, Cambridge, United Kingdom.

Torre, E. & Tello, M. (1993), 'Factors influencing fecal shedding of *Campylobacter jejuni* in dogs without diarrhea', *American Journal of Veterinary Research* **54**, 260–262.

Totten, P. A., Patton, C. M., Tenover, F. C., Barrett, T. J., Stamm, W. E., Steigerwalt, A. G., Lin, J. Y., Holmes, K. K. & Brenner, D. J. (1987), 'Prevalence and characterization of hippurate-negative *Campylobacter jejuni* in King County, Washington', *Journal of Clinical Microbiology* **25**, 1747–1752.

Townsend, J. P., Nielsen, K. M., Fisher, D. S. & Hartl, D. L. (2003), 'Horizontal acquisition of divergent chromosomal DNA in bacteria: effects of mutator phenotypes', *Genetics* **164**, 13–21.

Townsend, K. M. & Dawkins, H. J. S. (1993), 'Field alternation gel electrophoresis - status quo', *Journal of Chromatography B : Biomedical Sciences and Applications* **618**, 223–249.

Tribble, D. R., Baqar, S., Carmolli, M. P., Porter, C., Pierce, K. K., Sadigh, K., Guerry, P., Larsson, C. J., Rockabrand, D., Ventone, C. H., Poly, F., Lyon, C. E., Dakdouk, S., Fingar, A., Gilliland, T. J., Daunais, P., Jones, E., Rymarchyk, S., Huston, C., Darsley, M. & Kirkpatrick, B. D. (2009), '*Campylobacter jejuni* strain CG8421: a refined model for the study of campylobacteriosis and evaluation of *Campylobacter* vaccines in human subjects', *Clinical Infectious Diseases* **49**, 1512–1519.

Tsai, H. J., Huang, H. C., Lin, C. M., Lien, Y. Y. & Chou, C. H. (2007), 'Salmonellae and

campylobacters in household and stray dogs in northern Taiwan', *Veterinary Research Communications* **31**, 931–939.

Tuller, T., Girshovich, Y., Sella, Y., Kreimer, A., Freilich, S., Kupiec, M., Gophna, U. & Ruppin, E. (2011), 'Association between translation efficiency and horizontal gene transfer within microbial communities', *Nucleic Acids Research* **39**, 4743–4755.

Tunio, S. A., Oldfield, N. J., Ala'Aldeen, D. A., Wooldridge, K. G. & Turner, D. P. (2010), 'The role of glyceraldehyde 3-phosphate dehydrogenase (GapA-1) in *Neisseria meningitidis* adherence to human cells', *BioMedCentral Microbiology* **10**, 280–290.

Ukita, T. & Ikeda, H. (1996), 'Role of the *rec*J gene product in UV-induced illegitimate recombination at the hotspot', *Journal of Bacteriology* **178**, 2362–2367.

Umezu, K. & Kolodner, R. D. (1994), 'Protein interactions in genetic recombination in *Escherichia coli*. Interactions involving *Rec*O and *Rec*R overcome the inhibition of *Rec*A by single-stranded DNA-binding protein', *The Journal of Biological Chemistry* **269**, 3005–3013.

Umezu, K., Nakayama, K. & Nakayama, H. (1990), '*Escherichia coli Rec*Q protein is a DNA helicase', *Proceedings of National Academy of Sciences* **87**, 5363–5367.

Unicomb, L. E., Dalton, C. B., Gilbert, G. L., Becker, N. G. & Patel, M. S. (2008), 'Age-specific risk factors for sporadic *Campylobacter* infection in regional Australia', *Foodborne Pathogens and Disease* **5**, 79–85.

Urwin, R. & Maiden, M. C. J. (2003), 'Multi-locus sequence typing: a tool for global epidemiology', *Trends in Microbiology* **11**, 479–487.

van Belkum, A., Struelens, M., de Visser, A., Verbrugh, H. & Tibayrenc, M. (2001), 'Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology', *Clinical Microbiology Reviews* **14**, 547–560.

van Boven, M., Veldman, K. T., de Jong, M. C. M. & Mevius, D. J. (2003), 'Rapid selection of quinolone resistance in *Campylobacter jejuni* but not in *Escherichia coli* in individually housed broilers', *Journal of Antimicrobial Chemotherapy* **52**, 719–723.

Vandamme, P. & De Ley, J. (1991), 'Proposal for a new family, *Campylobacteraceae*', *International Journal of Systematic Bacteriology* **41**, 451–455.

Vandamme, P. & Goossens, H. (1992), 'Taxonomy of *Campylobacter*, *Arcobacter*, and *Helicobacter* : a review', *Zentralblatt fur Bakteriologie* **276**, 447–472.

Vandamme, P. & On, S. L. (2001), 'Recommendations of the subcommittee on the taxonomy of *Campylobacter* and related bacteria', *International Journal of Systematic and Evolutionary Microbiology* **51**, 719–721.

Vandecasteele, S. J., Peetermans, W. E., Merckx, R. & Van Eldere, J. (2001), 'Quantification of expression of *Staphylococcus epidermidis* housekeeping genes with Taqman quantitative PCR during *in vitro* growth and under different conditions', *Journal of Bacteriology* **183**, 7094–7101.

Varslot, M., Resell, J. & Fostad, I. G. (1996), 'Water-borne *Campylobacter* infection–probably caused by pink-footed geese. Two outbreaks in Nord-Trondelag, Stjortdal in 1994 and Verdal in 1995', *Tidsskr. Nor. Laegeforen. (In Norwegian.)* **116**, 3366–3369.

Velayudhan, J. & Kelly, D. J. (2002), 'Analysis of gluconeogenic and anaplerotic enzymes in *Campylobacter jejuni*: an essential role for phosphoenolpyruvate carboxykinase', *Microbiology* **148**, 685–694.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H. & Smith, H. O. (2004), 'Environmental genome shotgun sequencing of the Sargasso Sea', *Science* **304**, 66–74.

Vereen, E. J., Lowrance, R. R., Cole, D. J. & Lipp, E. K. (2007), 'Distribution and ecology of campylobacters in coastal plain streams (Georgia, United States of America)', *Applied and Environmental Microbiology* **73**, 1395–1403.

Viscidi, R. P. & Demma, J. C. (2003), 'Genetic diversity of *Neisseria gonorrhoeae* housekeeping genes', *Journal of Clinical Microbiology* **41**, 197–204.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. & Zabeau, M. (1995), 'AFLP : a new technique for DNA fingerprinting', *Nucleic Acids Research* **23**, 4407–4414.

Waldenstrom, J., Broman, T., Carlsson, I., Hasselquist, D., Achterberg, R. P., Wagenaar, J. A. & Olsen, B. (2002), 'Prevalence of *Campylobacter jejuni*, *Campylobacter lari*, *and Campylobacter coli* in different ecological guilds and taxa of migrating birds.', *Applied and Environmental Microbiology* **68**, 5911–5917.

Walder, M., Sandstedt, K. & Ursing, J. (1983), 'Phenotypic characteristics of thermotolerant *Campylobacter* from human and animal sources', *Current Microbiology* **9**, 291–295.

Wallace, J. S., Stanley, K. N., Currie, J. E., Diggle, P. J. & Jones, K. (1997), 'Seasonality of thermophilic *Campylobacter* populations in chickens', *Journal of Applied Microbiology* **82**, 219–224.

Wang, Y., Huang, W. M. & Taylor, D. E. (1993), 'Cloning and nucleotide sequence of the *Campylobacter jejuni gyr*A gene and characterization of quinolone resistance mutations.', *Antimicrobial Agents and Chemotherapy* **37**, 457–463.

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

Wassenaar, T. M., Fry, B. N. & van der Zeijst, B. A. (1995), 'Variation of the flagellin gene locus of *Campylobacter jejuni* by recombination and horizontal gene transfer', *Microbiology* **141**, 95–101.

Wassenaar, T. M., Geilhausen, B. & Newell, D. G. (1998), 'Evidence of genomic instability in *Campylobacter jejuni* isolated from poultry', *Applied and Environmental Microbiology* **64**, 1816–1821.

Wassenaar, T. M. & Newell, D. G. (2000), 'Genotyping of *Campylobacter* spp.', *Applied and Environmental Microbiology* **66**, 1–9.

Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M., Devine, S. K., Day, J. P., Wilkinson, A., d'Adda di Fagagna, F., Devine, K. M., Bowater, R. P., Jeggo, P. A., Jackson, S. P. & Doherty, A. J. (2002), 'Identification of a DNA nonhomologous end-joining complex in bacteria', *Science* **297**, 1686–1689.

Wenman, W. M., Chai, J., Louie, T. J., Goudreau, C., Lior, H., Newell, D. G., Pearson, A. D. & Taylor, D. E. (1985), 'Antigenic analysis of *Campylobacter* flagellar protein and other proteins', *Journal of Clinical Microbiology* **21**, 108–112.

Wertz, J. E., Goldstone, C., Gordon, D. M. & Riley, M. A. (2003), 'A molecular phylogeny of enteric bacteria and implications for a bacterial species concept', *Journal of Evolutionary Biology* **16**, 1236–1248.

Whelan, S., Lio, P. & Goldman, N. (2001), 'Molecular phylogenetics: state-of-the-art methods for looking into the past', *Trends in Genetics* **17**, 262–272.

Whittam, T. S. & Bumbaugh, A. C. (2002), 'Inferences from whole-genome sequences of bacterial pathogens.', *Current Opinion in Genetics and Development* **12**, 719–725.

Whittier, C. A., Cranfield, M. R. & Stoskopf, M. K. (2010), 'Real-time PCR detection of *Campylobacter* spp. in free ranging mountain gorillas *(Gorilla beringei beringei)*', *Journal of Wildlife Diseases* **46**, 791–802.

WHO (2000a), World Health Organization. The Increasing Incidence of Human Campylobacteriosis. Report and Proccedings of a WHO Consultation of Experts, Technical report, World Health Organization. Department of Communicable Disease Surveillance and Response.WHO/CDS/CSR/APH/2000.4.

WHO (2000b), World Health Organization. Fact Sheet No. 255: *Campylobacter*, Technical report, World Health Organization. Department of Communicable Disease Surveillance and Response.WHO/CDS/CSR/APH/2000.4.

Wieland, B., Regula, G., Danuser, J., Wittwer, M., Burnens, A. P., Wassenaar, T. M. & Stirk, K. D. (2005), '*Campylobacter* spp. in dogs and cats in Switzerland: risk factor analysis and molecular characterization with AFLP', *Journal of Veterinary Medicine. B, Infectious Diseases and Veterinary Public Health* **52**, 183–189.

Williams, N. J., Jones, T. R., Leatherbarrow, H. J., Birtles, R. J., Lahuerta-Marin, A., Bennett, M. & Winstanley, C. (2010), 'Isolation of a novel *Campylobacter jejuni* clone associated with the bank vole, *Myodes glareolus*', *Applied and Environmental Microbiology* **76**, 7318–7321.

Williams, N. J., Sherlock, C., Jones, T. R., Clough, H. E., Telfer, S. E., Begon, M., French, N., Hart, C. A. & Bennett, M. (2011), 'The prevalence of antimicrobial-resistant *Escherichia coli* in sympatric wild rodents varies by season and host', *Journal of Applied Microbiology* **110**, 962–970.

Wilson, D. J., Gabriel, E., Leatherbarrow, A. J. H., Cheesbrough, J., Gee, S., Bolton, E., Fox, A., Fearnhead, P., Hart, C. A. & Diggle, P. J. (2008), 'Tracing the source of campylobacteriosis', *Public Library of Science Genetics* **4**, 100–113.

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

Wilson, K. H., Blitchington, R. B. & Greene, R. C. (1990), 'Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction', *Journal of Clinical Microbiology* **28**, 1942–1946.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., Karch, H., Reeves, P. R., Maiden, M. C. J., Ochman, H. & Achtman, M. (2006), 'Sex and virulence in *Escherichia coli* : an evolutionary perspective', *Molecular Microbiology* **60**, 1136–1151.

Woese, C. R. (1987), 'Bacterial evolution', *Microbiological Reviews* **51**, 221–271.

Wolf, Y. I., Rogozin, I. B., Grishin, N. V. & Koonin, E. V. (2002), 'Genome trees and the tree of life', *Trends in Genetics* **18**, 472–479.

Wolfs, T. F., Duim, B., Geelen, S. P., Rigter, A., Thomson-Carter, F., Fleer, A. & Wagenaar, J. A. (2001), 'Neonatal sepsis by *Campylobacter jejuni* : genetically proven transmission from a household puppy', *Clinical Infectious Diseases* **32**, e97–99.

Wong, T. Y., Fernandes, S., Sankhon, N., Leong, P. P., Kuo, J. & Liu, J. K. (2008), 'Role of premature stop codons in bacterial evolution', *Journal of Bacterial Evolution* **190**, 6718–6725.

Woolfit, M. & Bromham, L. (2003), 'Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes', *Molecular Biology and Evolution* **20**, 1545–1555.

Workman, S. N., Mathison, G. E. & Lavoie, M. C. (2005), 'Pet dogs and chicken meat as reservoirs of *Campylobacter* spp. in Barbados', *Journal of Clinical Microbiology* **43**, 2642–2650.

Wright, J. A., Grant, A. J., Hurd, D., Harrison, M., Guccione, E. J., Kelly, D. J. & Maskell, D. J. (2009), 'Metabolite and transcriptome analysis of *Campylobacter jejuni in vitro* growth reveals a stationary-phase physiological switch', *Microbiology* **155**, 80–94.

Wright, S. (1951), 'The genetical structure of populations', *Annals of Eugenics* **15**, 323–354.

Wright, S. (1965), 'The interpretation of population structure by F-statistics with special regard to systems of mating', *Evolution* **19**, 395–420.

Wright, S. (1978), *Evolution and the Genetics of Populations. Variability Within and Among Natural Populations*, Vol. 4, The University of Chicago, Chicago.

Wright, S. (1984), *Evolution and the Genetics of Populations: Genetics and Biometric Foundations (Genetic & Biometric Foundations)*, Vol. 1, The University of Chicago, Chicago.

Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. (2006), 'Base-stacking and base-pairing contributions into thermal stability of the DNA double helix', *Nucleic Acids Research* **34**, 564–574.

Yamazaki-Matsune, W., Taguchi, M., Seto, K., Kawahara, R., Kawatsu, K., Kumeda, Y., Kitazato, M., Nukina, M., Misawa, N. & Tsukamoto, T. (2007), 'Development of a multiplex PCR assay for identification of *Campylobacter coli*, *Campylobacter fetus*, *Campylobacter hyointestinalis* subsp. *hyointestinalis*, *Campylobacter jejuni*, *Campylobacter lari* and *Campylobacter upsaliensis*', *Journal of Medical Microbiology* **56**, 1467–1473.

Yang, C., Jiang, Y., Huang, K., Zhu, C. & Yin, Y. (2003), 'Application of real-time PCR for quantitative detection of *Campylobacter jejuni* in poultry, milk and environmental water', *FEMS Immunology & Medical Microbiology* **38**, 265–271.

Yang, Z. (1993), 'Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites', *Molecular Biology and Evolution* **10**, 1396–1401.

Yang, Z. (1994b), 'Estimating the pattern of nucleotide substitution', *Journal of Molecular Evolution* **39**, 105–111.

Yang, Z. (2002), 'Inference of selection from multiple species alignments', *Current Opinion in Genetics and Development* **12**, 688–694.

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A.-M. K. (2000), 'Codon-substitution models for heterogeneous selection pressure at amino acid sites', *Genetics* **155**, 431–449.

Yang, Z., Nielsen, R. & Hasegawa, M. (1998), 'Models of amino acid substitution and applications to mitochondrial protein evolution', *Molecular Biology and Evolution* **15**, 1600–1611.

Zhang, B. & Dong, X. Z. (2005), 'Partial sequence homology of *Fts*Z in phylogenetics analysis of lactic acid bacteria', *Wei Sheng Wu Xue Bao* **45**, 661–664.

Zhang, Q., Lin, J. & Pereira, S. (2003), 'Fluoroquinolone-resistant *Campylobacter* in animal reservoirs: dynamics of development, resistance mechanisms and ecological fitness', *Animal Health Research Reviews* **4**, 63–71.

Zhang, Q., Meitzler, J. C., Huang, S. & Morishita, T. (2000), 'Sequence polymorphism, predicted secondary structures, and surface-exposed conformational epitopes of *Campylobacter* major outer membrane protein', *Infection and Immunity* **68**, 5679–5689.

Zhang, Q., Sahin, O., McDermott, P. F. & Payot, S. (2006a), 'Fitness of antimicrobial-resistant *Campylobacter* and *Salmonella*', *Microbes and Infection* **8**, 1972 – 1978.

Zhang, Y., Romero, H., Salinas, G. & Gladyshev, V. N. (2006b), 'Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues', *Genome Biology* **7**, R94.

Zhou, B.-B. S. & Elledge, S. J. (2000), 'The DNA damage response: putting checkpoints in perspective', *Nature* **408**, 433–439.

Zia, S., Wareing, D., Sutton, C., Bolton, E., Mitchell, D. & Goodacre, J. A. (2003), 'Health problems following *Campylobacter jejuni* enteritis in a Lancashire population', *Rheumatology* **42**, 1083–1088.

Zirnstein, G., Helsel, L., Li, Y., Swaminathan, B. & Besser, J. (2000), 'Characterization of *gyr*A mutations associated with fluoroquinolone resistance in *Campylobacter coli* by DNA sequence analysis and MAMA PCR', *FEMS Microbiology Letters* **190**, 1–7.

Zuckerkandl, E. & Pauling, L. (1965), 'Molecules as documents of evolutionary history', *Journal of Theoretical Biology* **8**, 357–366.

APPENDIX A

Campylobacter jejuni colonisation and population structure in sympatric urban population of ducks and starlings

A.1 Summary

This appendix provides further information on the details described in Chapter 3. This section gives a detailed description of the sampling sites, the number of sample collected from ducks and starlings, the faecal sample processing protocols and the laboratory procedures performed in this study. Further, the MLST profiles of the isolates those with alleles that could not be amplified and/or sequenced are provided in this appendix in order to give a completeness to the work that was described in Chapter 3. Although the results and the patterns of population differentiation were presented and discussed in the manuscript of Chapter 3, the tables of Fst values and the p-values were not included in the manuscript, instead the tables are presented in this appendix in an effort to further illustrate the findings from the work presented in Chapter 3.

A.2 Experimental procedures

A.2.1 Sample collection

Identification of starling faecal material was standardised by waiting and watching the birds defecating and the faecal characteristics were recorded as photographic images that

are shown in Figures A.1. Details on number of samples collected from each sampling site, *Campylobacter* positive samples by PCR and *C. jejuni* positive samples are provided in the following Table A.2 as outlined in Chapter 3.

(a) (b) (c)

Figure A.1: Pictures of starling faecal material recorded during the sampling trial period. Note the characteristic yellowish and greenish tint covered with a whitish paste and slimy mucous

Table A.1: Faecal sample details of ducks and starlings for the isolation of *Campylobacter* spp.

Site	Month	Date	*C. jejuni*.duck	*Campylobacter*.duck	*C. jejuni*.starling	*Campylobacter*.starling	n.duck	n.starling
ES	1	1/08/2008	4	4	1	2	12	12
HO	1	1/08/2008	2	2	0	4	8	12
MM	1	1/08/2008	2	3	1	8	12	12
MS	1	1/08/2008	4	5	0	2	12	12
SQ	1	1/08/2008	7	7	2	3	12	12
TOTAL	1	1/08/2008	19	21	4	19	56	60
ES	2	1/09/2008	9	10	6	10	12	12
HO	2	1/09/2008	3	3	8	9	12	12
MM	2	1/09/2008	5	7	7	12	12	12
MS	2	1/09/2008	1	1	0	3	12	12
SQ	2	1/09/2008	5	5	6	7	12	12
TOTAL	2	1/09/2008	23	26	27	41	60	60
ES	3	1/10/2008	5	7	5	7	12	12
HO	3	1/10/2008	8	8	4	8	12	12
MM	3	1/10/2008	4	6	5	8	12	12
MS	3	1/10/2008	6	7	1	2	12	12
SQ	3	1/10/2008	6	6	5	6	12	12
TOTAL	3	1/10/2008	29	34	20	31	60	60
ES	4	1/11/2008	2	3	2	8	12	12
HO	4	1/11/2008	0	0	0	0	12	12
MM	4	1/11/2008	4	5	3	5	12	12
MS	4	1/11/2008	1	1	5	5	12	12
SQ	4	1/11/2008	1	3	4	5	12	12
TOTAL	4	1/11/2008	8	12	14	23	60	60
ES	5	1/12/2008	4	6	3	4	12	12
HO	5	1/12/2008	0	0	0	0	12	12
MM	5	1/12/2008	2	3	0	0	12	12
MS	5	1/12/2008	1	1	4	4	12	12
SQ	5	1/12/2008	0	2	0	1	12	12
TOTAL	5	1/12/2008	7	12	7	9	60	60
ES	6	1/01/2009	2	3	2	3	12	12
HO	6	1/01/2009	0	0	0	1	12	12
MM	6	1/01/2009	4	4	5	5	12	12
MS	6	1/01/2009	2	3	3	3	12	12
SQ	6	1/01/2009	1	2	1	1	12	12
TOTAL	6	1/01/2009	9	12	11	13	60	60

Table A.2: Faecal sample details of ducks and starlings for the isolation of *Campylobacter* spp. continued

Site	Month	Date	*C. jejuni*.duck	*Campylobacter*.duck	*C. jejuni*.starling	*Campylobacter*.starling	n.duck	n.starling
ES	7	1/02/2009	5	6	1	4	12	12
HO	7	1/02/2009	1	1	2	3	12	12
MM	7	1/02/2009	6	7	2	4	12	12
MS	7	1/02/2009	1	2	1	5	12	12
SQ	7	1/02/2009	1	2	3	4	12	12
TOTAL	7	1/02/2009	14	18	9	20	60	60
ES	8	1/03/2009	0	2	0	3	12	12
HO	8	1/03/2009	0	0	2	4	12	12
MM	8	1/03/2009	1	1	6	6	12	12
MS	8	1/03/2009	1	2	0	4	12	12
SQ	8	1/03/2009	4	4	2	5	12	12
TOTAL	8	1/03/2009	6	9	10	22	60	60
ES	9	1/04/2009	3	3	1	6	12	12
HO	9	1/04/2009	0	0	1	7	12	12
MM	9	1/04/2009	2	3	2	7	12	12
MS	9	1/04/2009	1	1	4	6	12	12
SQ	9	1/04/2009	6	6	5	9	12	12
TOTAL	9	1/04/2009	12	13	13	35	60	60
ES	10	1/05/2009	1	3	8	10	12	12
HO	10	1/05/2009	5	5	1	7	12	12
MM	10	1/05/2009	5	5	3	4	12	12
MS	10	1/05/2009	1	1	0	7	12	12
SQ	10	1/05/2009	4	7	0	7	12	12
TOTAL	10	1/05/2009	16	21	12	35	60	60
ES	11	1/06/2009	3	7	5	12	12	12
HO	11	1/06/2009	4	4	3	8	12	12
MM	11	1/06/2009	2	4	2	7	12	12
MS	11	1/06/2009	2	2	3	7	12	12
SQ	11	1/06/2009	1	2	1	5	12	12
TOTAL	11	1/06/2009	12	19	14	39	60	60
ES	12	1/07/2009	1	3	3	7	12	12
HO	12	1/07/2009	3	3	0	9	12	12
MM	12	1/07/2009	0	0	2	11	12	12
MS	12	1/07/2009	3	4	1	9	12	12
SQ	12	1/07/2009	3	6	3	6	12	12
TOTAL	12	1/07/2009	10	16	9	42	60	60

ES	The Esplanade
HO	Hokowhitu
MM	Memorial Park
MS	Massey University Concourse
SQ	The Square

A.2.2 Isolation of *Campylobacter* spp. : Laboratory procedures

Laboratory procedures involved in the isolation and speciation of *Campylobacter* spp. from ducks, starlings and geese are illustrated here. Once presumptive *Campylobacter* isolates were identified, the isolates were speciated and *C. jejuni* positive isolates were taken to the final stage of typing by multilocus sequence typing (MLST) (Figures A.3 and A.2).

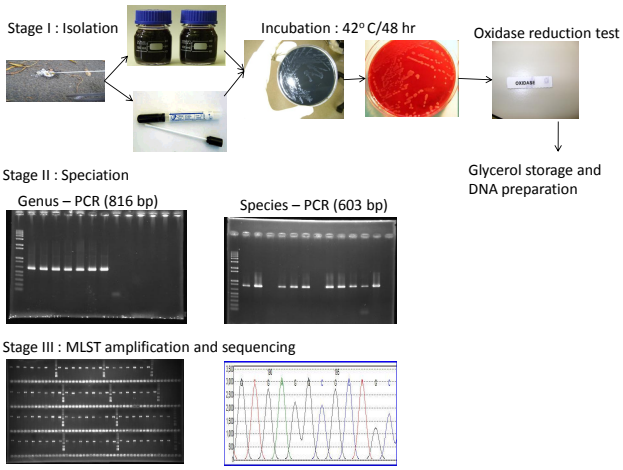


Figure A.2: Stages of laboratory procedures

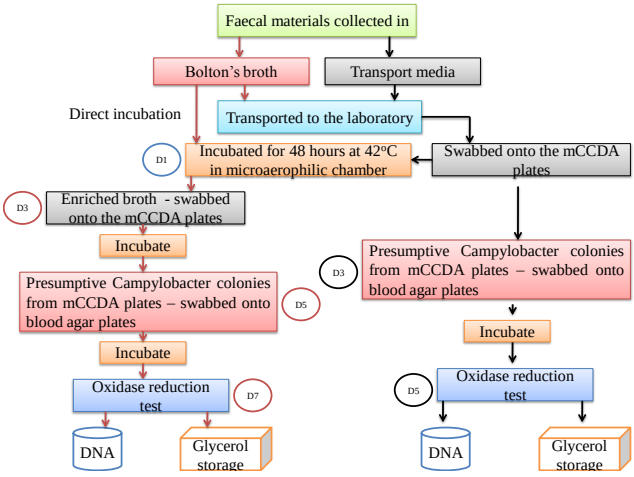


Figure A.3: Flow chart of laboratory procedures. The chart describes the day by day laboratory procedures carried out after collection of faecal materials. D = Day

A.2.3 Cell surface antigen typing

The isolates that were difficult to amplify were further processed in the Hopkirk Research Laboratory using primers from the PubMLST database. The PCR program was optimised to amplify the difficult *fla*A and *por*A sequences and further, the $MgCl_2$ was optimised to 2.5 M and primers to 3.2 picomoles. The modified PCR program used for amplifying and sequencing the difficult isolates is described in the Table A.3.

Table A.3: *flaA* and *por*A typing

*por*A PCR reaction and program:		*flaA* PCR reaction and program:	
Primers: Momp1 and Momp2 from the PUBMLST database:		Primers: flaA1 and flaA2 from the PUBMLST database:	
10× buffer: 2 microlit		10× buffer	2 μL
MgCl2	1 μL	MgCl2	1 μL
DNTPs	1 μL	DNTPs	1 μL
Primer1	2 μL	Primer1	2 μL
Primer2	2 μL	Primer2	2 μL
Taq(Platinum)	0.2 μL	Taq(Platinum)	0.2 μL
DNA (50 - 100ng)	2 μL	DNA (50 - 100ng)	2 μL
Made upto 20μL with distilled water		Made upto 20 μL with distilled water	
PCR program:		PCR program:	
Initial denaturation	95°C for 2 minutes	Initial denaturation	94 °Cfor 2 minutes
for 40 cycles:		for 37 cycles:	
Denaturation	94 °C : 30 seconds	Denaturation	94 °C : 30 seconds
Annealing	50 °C : 30 seconds	Annealing	64 °C : 30 seconds
Extension	72 °C : 30 seconds	Extension	72 °C : 45 seconds
Final extension: 72 °C for 2 minutes		Final extension: 72 °C for 2 minutes	

A.3 Results

A.3.1 Multilocus sequence typing : Complete and partial *C. jejuni* MLST profiles from ducks and starlings

The isolates of *C. jejuni* from mallard ducks and starlings were amplified and sequenced following the procedures described in the Chapter 3. However, the amplification of the alleles were optimised with the $MgCl_2$ was optimised to 2.5 M and primers to 3.2 pico-moles to get the difficult alleles work. Despite the efforts that were put in to amplify and / or sequence these isolates, some of the alleles were did not work. The isolates that had complete MLST profiles and those isolates with partial allelic profiles that could not be used for analysis are provided in Table A.4, Table A.5 and Table A.6

Table A.4: Complete MLST profiles of *C. jejuni* from ducks and starlings

Lab No	Sample Type	Source Site	*asp*A	*gln*A	*glt*A	*gly*A	*glm*M	*tkt*	*unc*A	ST	CC
B1133b	Duck faeces	Esplanade PN	2	15	4	48	356	25	23	2378	1034
B1135a	Starling faeces	PN The Square	2	15	4	48	356	25	23	2378	1034
B1135b	Starling faeces	PN The Square	2	15	4	48	356	25	23	2378	1034
B1255a	Duck faeces	PN The Square	2	61	4	64	126	7	23	1033	1034
B1278b	Goose faeces	Hokowhitu Lagoon	2	61	4	64	126	7	23	1033	1034
B1454a	Duck faeces	Massey University	2	15	4	48	360	25	23	2391	1034
B1454b	Duck faeces	Massey University	2	15	4	48	360	25	23	2391	1034
B1500c	Duck faeces	Memorial Park PN	2	15	4	48	360	25	23	2391	1034
B1527	Duck faeces	PN The Square	2	15	4	48	360	25	23	2391	1034
B156c	Duck faeces	Esplanade PN	22	61	4	64	74	25	23	977	1034
B1689a	Duck faeces	Esplanade PN	2	61	4	64	126	7	23	1033	1034
B1689c	Duck faeces	Esplanade PN	2	61	4	64	126	7	23	1033	1034
B1712a	Duck faeces	Massey University	2	15	4	48	356	25	23	2378	1034
B1732a	Duck faeces	PN The Square	22	146	4	64	74	25	23	1255	1034
B1732b	Duck faeces	PN The Square	22	146	4	64	74	25	23	1255	1034
B204a	Duck faeces	PN The Square	22	61	4	64	74	25	23	977	1034
B209a	Duck faeces	PN The Square	22	61	4	64	74	25	23	977	1034
B299c	Duck faeces	Massey University	2	15	4	48	356	25	23	2378	1034
B773b	Duck faeces	PN The Square	2	61	4	64	126	7	23	1033	1034
B830a	Duck faeces	Massey University	2	61	4	64	126	7	23	1033	1034
B847a	Duck faeces	Massey University	2	61	4	64	126	7	23	1033	1034
B1172c	Starling faeces	Hokowhitu Lagoon	100	142	93	135	190	145	81	1304	1304
B145c	Duck faeces	Esplanade PN	2	1	4	28	58	25	58	696	1332
B155	Duck faeces	Esplanade PN	2	1	4	28	58	25	58	696	1332
B647a	Starling faeces	Esplanade PN	2	1	4	28	58	25	58	696	1332
B1405a	Duck faeces	PN The Square	17	2	8	5	8	2	143	2537	177
B1410a	Starling faeces	PN The Square	17	2	8	5	8	2	4	177	177
B1417b	Starling faeces	PN The Square	17	2	8	5	8	2	4	177	177
B1581b	Starling faeces	Massey University	17	2	8	5	8	2	4	177	177
B692b	Starling faeces	Memorial Park PN	17	2	8	5	8	2	4	177	177
B715b	Starling faeces	Hokowhitu Lagoon	17	2	8	5	8	2	4	177	177
B1667a	Duck faeces	Memorial Park PN	2	1	21	3	2	1	5	53	21
B1754c	Starling faeces	PN The Square	2	1	21	3	2	1	5	53	21
B412c	Duck faeces	Esplanade PN	10	1	16	19	10	5	7	2026	403
B1267a	Starling faeces	PN The Square	1	2	3	4	5	9	21	4498	42
B1746b	Starling faeces	PN The Square	1	2	3	4	5	9	3	42	42
B436a	Duck faeces	PN The Square	1	2	9	5	5	9	21	4503	42
B1031b	Starling faeces	Memorial Park PN	4	7	10	4	1	7	1	45	45
B1139a	Starling faeces	PN The Square	4	7	10	4	1	7	1	45	45
B1541c	Duck faeces	Hokowhitu Lagoon	4	7	10	4	42	7	1	137	45
B1656	Starling faeces	Hokowhitu Lagoon	4	7	10	4	1	7	1	45	45
B1659a	Starling faeces	Hokowhitu Lagoon	4	7	10	4	1	7	1	45	45
B1783b	Duck faeces	Hokowhitu Lagoon	4	7	10	4	1	7	1	45	45
B447a	Starling faeces	PN The Square	4	7	10	4	42	51	1	583	45
B543b	Starling faeces	Esplanade PN	4	7	10	4	1	7	1	45	45
B547b	Starling faeces	Esplanade PN	4	7	10	4	1	7	1	45	45
B593b	Duck faeces	Memorial Park PN	4	7	10	4	1	7	1	45	45
B600a	Duck faeces	Memorial Park PN	4	7	10	4	1	7	1	45	45
B603a	Duck faeces	Massey University	4	7	10	4	1	7	1	45	45
B625b	Starling faeces	Massey University	4	7	10	4	42	51	1	583	45
B657a	Duck faeces	PN The Square	4	7	10	4	42	51	1	583	45
B689a	Starling faeces	Memorial Park PN	4	7	10	4	1	7	1	45	45
B695b	Starling faeces	Memorial Park PN	4	7	10	4	1	7	1	45	45
B727a	Duck faeces	Esplanade PN	4	7	10	4	1	7	1	45	45
B756a	Duck faeces	Memorial Park PN	4	7	10	4	1	7	1	45	45
B760b	Starling faeces	Memorial Park PN	4	7	10	4	1	7	1	45	45
B761a	Starling faeces	Memorial Park PN	4	7	10	4	42	7	1	137	45
B793b	Starling faeces	PN The Square	4	7	10	4	42	51	1	583	45
B835b	Starling faeces	Massey University	4	7	10	4	42	7	1	137	45
B840a	Starling faeces	Massey University	4	7	10	4	42	7	1	137	45
B894b	Duck faeces	Memorial Park PN	4	7	10	4	1	7	1	45	45
B951b	Starling faeces	Esplanade PN	4	7	10	4	42	51	1	583	45
B832b	Starling faeces	Massey University	10	81	50	99	120	76	52	677	677
B833b	Starling faeces	Massey University	10	81	50	99	120	76	52	677	677
B1432b	Starling faeces	Hokowhitu Lagoon	35	43	9	5	8	46	21	681	682

Table A.5: Complete MLST profiles of *C. jejuni* from ducks and starlings - continued

Lab No	Sample Type	Source Site	*asp*A	*gln*A	*glt*A	*gly*A	*glm*M	*tkt*	*unc*A	ST	CC
B1651a	Starling faeces	Hokowhitu Lagoon	37	52	57	26	127	29	23	692	692
B1711a	Duck faeces	Massey University	37	52	21	388	127	29	23	4502	692
B1711c	Duck faeces	Massey University	37	52	21	388	127	29	23	4502	692
B1782a	Duck faeces	Hokowhitu Lagoon	37	52	57	26	129	29	23	699	692
B973a	Duck faeces	Massey University	37	52	57	26	127	29	23	692	692
B1013a	Duck faeces	Memorial Park PN	37	NEW	57	26	129	29	23	u699a	U/A
B1031a	Starling faeces	Memorial Park PN	98	122	98	125	180	150	113	1342	U/A
B1126a	Duck faeces	Esplanade PN	2	15	98	48	356	25	23	4496	U/A
B1129a	Duck faeces	Esplanade PN	2	4	4	48	358	25	280	4497	U/A
B1139b	Starling faeces	PN The Square	2	59	4	27	126	29	23	992	U/A
B1142b	Starling faeces	PN The Square	2	29	4	27	10	25	24	3961	U/A
B1205a	Duck faeces	Massey University	2	29	4	27	10	25	24	3961	U/A
B1302b	Duck faeces	Memorial Park PN	37	4	4	48	13	25	23	2354	U/A
B1314b	Starling faeces	Memorial Park PN	98	359	98	125	180	150	113	4499	U/A
B1367b	Starling faeces	Esplanade PN	99	128	91	125	170	146	111	1324	U/A
B1381b	Duck faeces	Memorial Park PN	37	29	75	48	126	25	23	710	U/A
B1395a	Starling faeces	Memorial Park PN	98	122	98	125	180	150	113	1342	U/A
B1395b	Starling faeces	Memorial Park PN	98	122	98	125	180	150	113	1342	U/A
B1422b	Goose faeces	Hokowhitu Lagoon	37	4	4	48	13	25	23	2354	U/A
B142b	Goose faeces	Hokowhitu Lagoon	2	4	4	105	10	25	57	2347	U/A
B1475a	Duck faeces	Esplanade PN	2	59	4	48	131	24	57	2349	U/A
B1484	Starling faeces	Esplanade PN	99	128	91	125	170	146	111	1324	U/A
B1487	Starling faeces	Esplanade PN	99	128	91	125	170	146	111	1324	U/A
B1510b	Starling faeces	Memorial Park PN	99	128	91	125	170	146	111	1324	U/A
B1516a	Duck faeces	PN The Square	2	4	4	105	10	25	57	2347	U/A
B1516b	Duck faeces	PN The Square	2	4	4	105	10	25	57	2347	U/A
B1556c	Starling faeces	Hokowhitu Lagoon	37	NEW	57	26	129	29	23	u699a	U/A
B1569a	Duck faeces	Massey University	2	29	4	48	13	24	57	693	U/A
B1569c	Duck faeces	Massey University	2	29	296	48	131	25	57	4500	U/A
B156b	Duck faeces	Esplanade PN	2	29	4	27	10	25	24	3961	U/A
B1596a	Duck faeces	Esplanade PN	2	29	4	48	13	24	57	693	U/A
B1607a	Starling faeces	Esplanade PN	99	128	91	125	170	146	111	1324	U/A
B1610a	Starling faeces	Esplanade PN	99	128	91	125	170	146	111	1324	U/A
B1614a	Duck faeces	PN The Square	2	29	4	27	10	25	24	3961	U/A
B1643b	Duck faeces	Hokowhitu Lagoon	2	29	4	48	13	24	57	693	U/A
B1646a	Duck faeces	Hokowhitu Lagoon	37	29	75	48	126	25	23	710	U/A
B1692b	Duck faeces	Esplanade PN	237	2	254	340	433	349	290	4501	U/A
B1696a	Starling faeces	Esplanade PN	99	128	91	125	170	146	111	1324	U/A
B1700a	Starling faeces	Esplanade PN	98	359	98	125	180	150	113	4499	U/A
B1706a	Starling faeces	Esplanade PN	95	2	94	127	172	144	114	1286	U/A
B1780a	Duck faeces	Hokowhitu Lagoon	2	59	4	27	126	29	23	992	U/A
B339a	Starling faeces	Esplanade PN	35	2	8	51	361	2	21	2538	U/A
B391c	Duck faeces	Massey University	99	128	91	125	170	146	111	1324	U/A
B473c	Starling faeces	Memorial Park PN	2	15	4	27	13	80	23	526	U/A
B521b	Duck faeces	Esplanade PN	37	4	4	48	13	25	23	2354	U/A
B596a	Duck faeces	Memorial Park PN	98	122	98	125	180	150	113	1342	U/A
B628a	Duck faeces	Esplanade PN	2	4	84	105	126	25	57	995	U/A
B640a	Starling faeces	Esplanade PN	99	128	91	125	170	146	111	1324	U/A
B857b	Starling faeces	Massey University	98	359	98	125	180	150	113	4499	U/A
B859a	Starling faeces	Massey University	98	359	98	125	180	150	113	4499	U/A
B860a	Starling faeces	Massey University	98	359	98	125	180	150	113	4499	U/A
B94	Duck faeces	Esplanade PN	2	4	84	105	126	25	57	995	U/A
B968c	Duck faeces	Massey University	2	15	4	48	356	150	23	4504	U/A
B978a	Starling faeces	Massey University	98	359	98	125	180	150	113	4499	U/A
B1624b	Starling faeces	PN The Square	26	2	9	51	8	46	5	208	682
B1071b	Duck faeces	PN The Square	37	52	57	26	107	29	23	991	692
B1165a	Duck faeces	Hokowhitu Lagoon	37	52	57	26	129	29	23	699	692
B1295b	Starling faeces	Hokowhitu Lagoon	37	52	57	26	107	29	23	991	692
B1498a	Duck faeces	Memorial Park PN	37	52	57	26	127	29	23	692	692

Table A.6: Partial MLST profiles of *C. jejuni* from ducks and starlings

Lab ID	Sample Type	Source Site	*asp*A	*gln*A	*glt*A	*gly*A	*glm*M	*tkt*	*unc*A
B1045	Duck faeces	Esplanade PN	X	15	4	48	356	25	23
B1153a	Duck faeces	PN The Square	X	29	X	105	X	24	57
B1331b	Duck faeces	Massey University	2	15	X	27	13	80	23
B1345b	Starling faeces	Massey University	X	2	9	51	8	46	21
B1356c	Duck faeces	Esplanade PN	37	52	57	26	107	29	X
B1595a	Duck faeces	Esplanade PN	2	1	12	88	2	X	5
B407a	Duck faeces	Esplanade PN	2	29	4	28	58	X	58
B409a	Duck faeces	Esplanade PN	X	X	16	X	X	1	X
B410c	Duck faeces	Esplanade PN	10	1	16	19	10	X	7
B416c	Duck faeces	Hokowhitu Lagoon	98	NEW	98	125	X	150	113
B435a	Duck faeces	PN The Square	2	165	73	147	X	7	104
B443b	Duck faeces	PN The Square	4	4	4	48	X	25	23
B517d	Duck faeces	PN The Square	2	61	4	64	126	X	23
B566c	Starling faeces	PN The Square	26	2	9	51	8	X	21
B667b	Starling faeces	PN The Square	4	7	10	4	1	X	1
B699a	Duck faeces	Hokowhitu Lagoon	4	7	10	4	X	51	1
B156a	Duck faeces	Esplanade PN	2	1	4	28	X	25	X
B172	Starling faeces	Massey University	X	1	4	28	58	25	58
B190a	Starling faeces	PN The Square	X	18	NEW	NEW	X	X	NEW

A.3.2 Analysis of molecular variance and diversity indices

Population pairwise Fsts of *C. jejuni* populations between ducks and starlings from each sampling site for different sampling periods were calculated and the results are tabulated and presented in Tables A.7 and A.8. Genetic diversity indices for the genotypes found in ducks and starlings (overall) as well as the sampling sites by species is provided in Table A.9.

Table A.7: Analysis of molecular variance (AMOVA)

Population specific Fst indices for ducks and starlings from different sampling sites

	ESD	ESS	HD	HS	MD	MS	MD	MS	SD	SS
Fst	0.2130	0.2162	-0.0188	-0.0340	0.2001	0.1960	0.0217	0.0268	0.0103	0.0040

Population pairwise Fsts for ducks and starlings at different sampling sites

	ESS	HKS	MSS	MMS	SQS	ESD	HKD	MSD	MMD	SQD
ESS	0	**0.000**	0.063	0.117	**0.009**	**0.000**	**0.036**	**0.000**	**0.036**	**0.000**
HKS	0.224	0	0.072	**0.009**	0.135	**0.000**	0.396	**0.009**	0.072	**0.018**
MSS	0.226	0.220	0	0.162	0.261	**0.000**	0.144	**0.000**	0.108	**0.027**
MMS	0.075	0.156	0.069	0	0.135	**0.009**	0.216	0.063	0.811	**0.009**
SQS	0.162	0.052	0.048	0.053	0	**0.018**	0.261	0.153	0.333	0.144
ESD	0.276	0.187	0.292	0.172	0.060	0	0.090	0.396	**0.009**	0.775
HKD	0.203	0.014	0.135	0.060	0.015	0.069	0	0.144	0.748	**0.045**
MSD	0.248	0.208	0.321	0.150	0.043	-0.003	0.078	0	0.180	0.351
MMD	0.107	0.083	0.069	-0.051	0.005	0.086	-0.046	0.058	0	**0.009**
SQD	0.270	0.132	0.267	0.173	0.047	-0.015	0.061	0.007	0.087	0

Note: The values below the diagonal are the Fst values and the values above the diagonal are the p-values. The significant values are shown in bold fonts.

Populations compared:
ESS : Esplanade starlings; HKS : Hokowhitu starlings ; MSS : Massey starlings ; MMS : Memorial starlings ; SQS : The Square starlings ;
ESD : Esplanade ducks ; HKD : Hokowhitu ducks ; MSD : Massey ducks ; MMD : Memorial ducks ; SQD : The Square ducks.

Table A.8: Genetic differentiation of *C. jejuni* populations with respect to different sampling periods

	ESDS	ESSS	HKDS	HKSS	MSDS	MSSS	MMDS	MMSS	SQDS	SQSS	ESDW	ESSW	HKDW	HKSW	MSDW	MSSW	MMDW	MMSW	SQDW	SQSW
ESDS	0	0.090	**0.045**	0.072	0.360	**0.036**	**0.045**	**0.018**	0.288	0.640	0.063	**0.000**	0.234	**0.045**	0.387	0.991	0.595	**0.018**	0.153	**0.009**
ESSS	0.156	0	0.054	0.477	0.243	0.207	0.775	0.784	0.613	0.306	**0.018**	**0.027**	0.342	0.198	**0.009**	0.991	**0.027**	**0.018**	**0.000**	0.108
HKDS	0.429	0.445	0	0.360	0.054	0.054	0.063	0.081	0.243	**0.045**	**0.036**	**0.009**	0.225	0.099	**0.045**	0.991	0.135	0.090	**0.018**	**0.045**
HKSS	0.228	0.170	0.500	0	0.108	0.144	0.288	0.126	0.342	0.108	0.288	0.144	0.099	0.396	0.108	0.991	**0.054**	0.081	0.126	0.721
MSDS	0.059	0.112	0.494	0.242	0	0.036	**0.045**	0.054	0.748	0.523	0.477	**0.000**	0.396	0.252	0.081	0.991	0.072	**0.045**	0.126	**0.009**
MSSS	0.228	0.051	0.409	0.177	0.202	0	0.270	0.090	0.360	0.072	**0.027**	**0.000**	0.090	0.117	**0.000**	0.991	**0.000**	0.270	**0.000**	**0.036**
MMDS	0.260	-0.067	0.455	0.294	0.188	0.080	0	0.991	0.387	0.108	**0.000**	**0.000**	0.288	0.369	**0.009**	0.991	0.054	0.081	**0.000**	**0.000**
MMSS	0.302	-0.059	0.564	0.285	0.235	0.098	-0.120	0	0.063	**0.027**	**0.000**	**0.000**	0.108	0.099	**0.000**	0.991	**0.000**	**0.036**	**0.000**	0.054
SQDS	0.050	-0.051	0.118	0.029	-0.136	0.049	0.016	0.110	0	0.550	0.559	**0.000**	0.991	0.991	0.297	0.991	0.613	0.117	0.225	0.216
SQSS	-0.053	0.034	0.386	0.185	-0.037	0.158	0.154	0.192	-0.078	0	0.162	**0.000**	0.748	0.180	0.459	0.991	0.288	**0.009**	0.261	**0.036**
ESDW	0.071	0.143	0.326	0.093	-0.023	0.149	0.240	0.273	-0.036	0.038	0	**0.000**	0.441	**0.009**	0.135	0.991	0.171	**0.000**	0.477	**0.009**
ESSW	0.386	0.258	0.549	0.321	0.411	0.290	0.406	0.426	0.317	0.362	0.301	0	**0.000**	**0.000**	**0.009**	0.991	**0.000**	**0.045**	**0.000**	**0.000**
HKDW	0.019	-0.005	0.175	0.108	-0.022	0.099	0.035	0.108	-0.196	-0.072	0.005	0.315	0	0.685	0.243	0.991	0.595	**0.009**	0.108	0.054
HKSW	0.164	0.070	0.014	0.061	0.134	0.134	0.061	0.167	-0.201	0.099	0.116	0.326	-0.060	0	0.072	0.991	0.306	0.018	**0.027**	0.153
MSDW	-0.005	0.202	0.321	0.206	0.106	0.253	0.308	0.354	0.056	-0.002	0.047	0.307	0.030	0.139	0	0.991	0.748	0.018	0.144	**0.000**
MSSW	0.314	0.243	1.000	-1.000	0.357	0.240	0.400	0.357	0.048	0.257	0.125	0.407	0.152	0.077	0.274	0	0.126	0.324	0.459	0.991
MMDW	-0.029	0.207	0.220	0.170	0.104	0.241	0.289	0.357	-0.038	0.028	0.054	0.362	-0.029	0.038	-0.060	0.243	0	**0.045**	0.198	**0.009**
MMSW	0.258	0.225	0.479	0.222	0.266	0.075	0.285	0.332	0.174	0.229	0.184	0.180	0.195	0.224	0.214	0.314	0.235	0	**0.000**	**0.000**
SQDW	0.036	0.169	0.329	0.093	0.066	0.173	0.267	0.288	0.029	0.029	-0.013	0.304	0.066	0.132	0.041	0.096	0.047	0.203	0	**0.000**
SQSW	0.213	0.117	0.373	-0.075	0.212	0.155	0.209	0.185	0.064	0.158	0.135	0.294	0.114	0.093	0.214	-0.301	0.177	0.236	0.125	0

Note: The values below the diagonal are the Fst values and the values above the diagonal are the p-values. The significant values are shown in bold fonts.
Populations compared :
ESDS : Esplanade ducks summer ; ESSS : Esplanade starlings summer ; HKDS : Hokowhitu ducks summer ; HKSS : Hokowhitu starlings summer ;
MSDS : Massey ducks summer ; MSSS : Massey starlings summer ; MMDS : Memorial ducks summer ; MMSS : Memorial starlings summer ;
SQDS : The Square ducks summer ; SQSS : The Square starlings summer ; ESDW : Esplanade ducks winter ; ESSW : Esplanade starlings winter ;
HKDW : Hokowhitu ducks winter ; HKSW : Hokowhitu starlings winter ; MSDW : Massey ducks winter ; MSSW : Massey starlings winter ;
MMDW : Memorial ducks winter ; MMSW : Memorial starlings winter ; SQDW : The Square ducks winter ; SQSW : The Square starlings winter.

Table A.9: Diversity indices of *C. jejuni* populations of ducks and starlings from different sampling sites

Parameters	Ducks_STs	Lower	Upper	Star_STs	Lower	Upper	Es_ducks	Lower	Upper	Hk_ducks	Lower	Upper	Mas_ducks	Lower	Upper	Mem_ducks	Lower	Upper
Taxa	33	23	33	23	21	31	14	9	15	6	4	7	10	8	13	8	7	11
Individuals	63	63	63	54	54	54	17	17	17	7	7	7	14	14	14	11	11	11
Simpson_1-D	0.9529	0.9191	0.9559	0.9198	0.9102	0.9547	0.9204	0.8443	0.9273	0.8163	0.7347	0.8571	0.8878	0.8163	0.9184	0.8099	0.7934	0.9091
Shannon_H	3.283	2.847	3.287	2.799	2.772	3.257	2.589	2.038	2.67	1.748	1.352	1.946	2.243	1.909	2.54	1.894	1.768	2.398

Parameters	Sq_ducks	Lower	Upper	Es_star	Lower	Upper	Hk_star	Lower	Upper	Mas_star	Lower	Upper	Mem_star	Lower	Upper	Sq_star	Lower	Upper
Taxa	10	8	14	7	8	13	6	4	6	5	5	9	7	7	12	10	7	13
Individuals	14	14	14	14	14	14	6	6	6	9	9	9	12	12	12	13	13	13
Simpson_1-D	0.8878	0.8265	0.9286	0.7041	0.8163	0.9184	0.8333	0.6667	0.8333	0.7654	0.7407	0.8889	0.7917	0.8056	0.9167	0.8876	0.8047	0.9231
Shannon_H	2.243	1.946	2.639	1.567	1.909	2.54	1.792	1.242	1.792	1.523	1.581	2.197	1.748	1.792	2.485	2.245	1.818	2.565

Ducks_STs = Overall duck sequence types

Star_STs = Overall starling sequence types

STs = Sequence types

Star = Starlings

Es = The Esplanade

Hk = Hokowhitu

Mas = Massey concourse

Mem = Memorial Park

Sq = The Square

A.3.3 Minimum spanning tree

The minimum spanning tree (MST) that represents the clonal complex (CC) of the *C. jejuni* genotypes from ducks and starlings is provided in this Appendix while the manuscript contains the MST that shows the STs. The STs were partitioned based on the CC that describes the relatedness among the STs and also the STs that were not assigned to a complex and their positions in the tree (A.4). The *fla*A alleles and the *por*A alleles were not included in the MST representation in the manuscript to avoid too many directions in the analysis. A MST including the *fla*A and *por*A alleles with the MLST profiles were clustered and provided here for comparison in the Figure A.5.

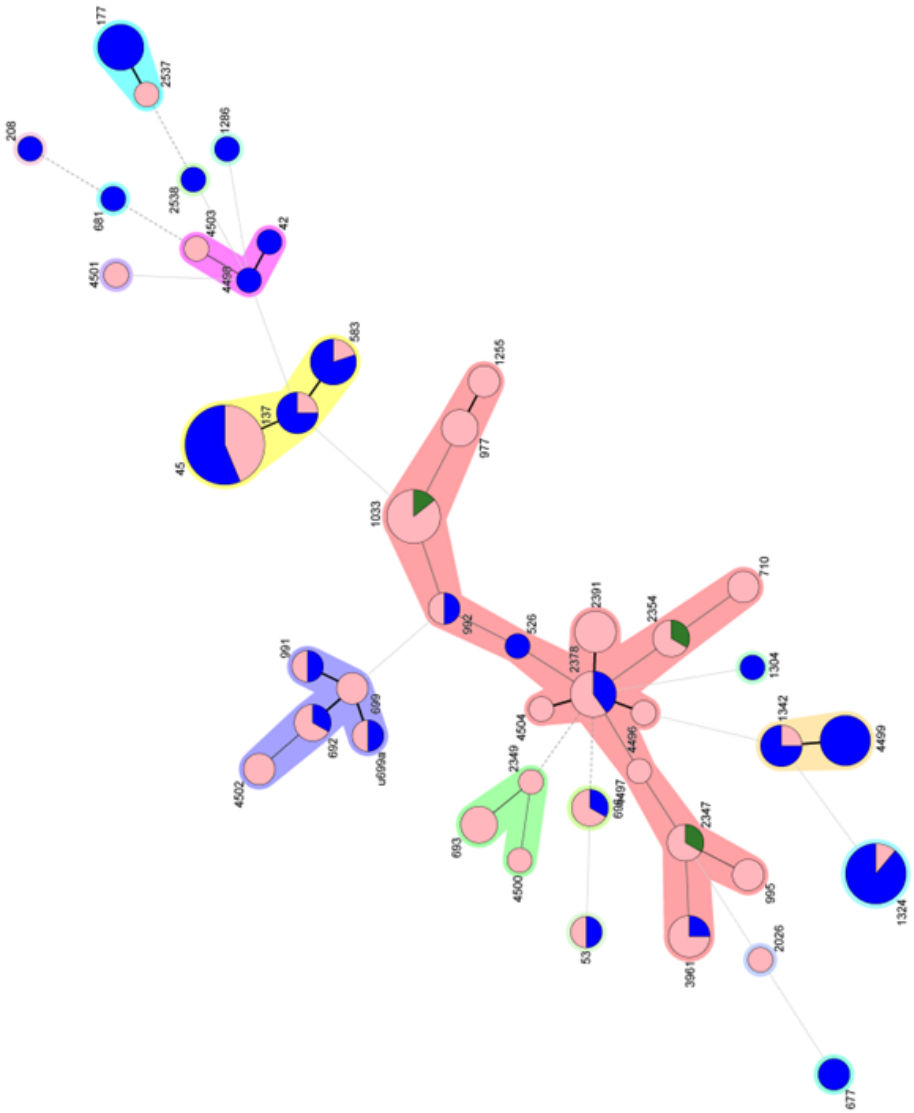


Figure A.4: The minimum spanning tree (MST) partitioned based on the clonal complexes (CC)

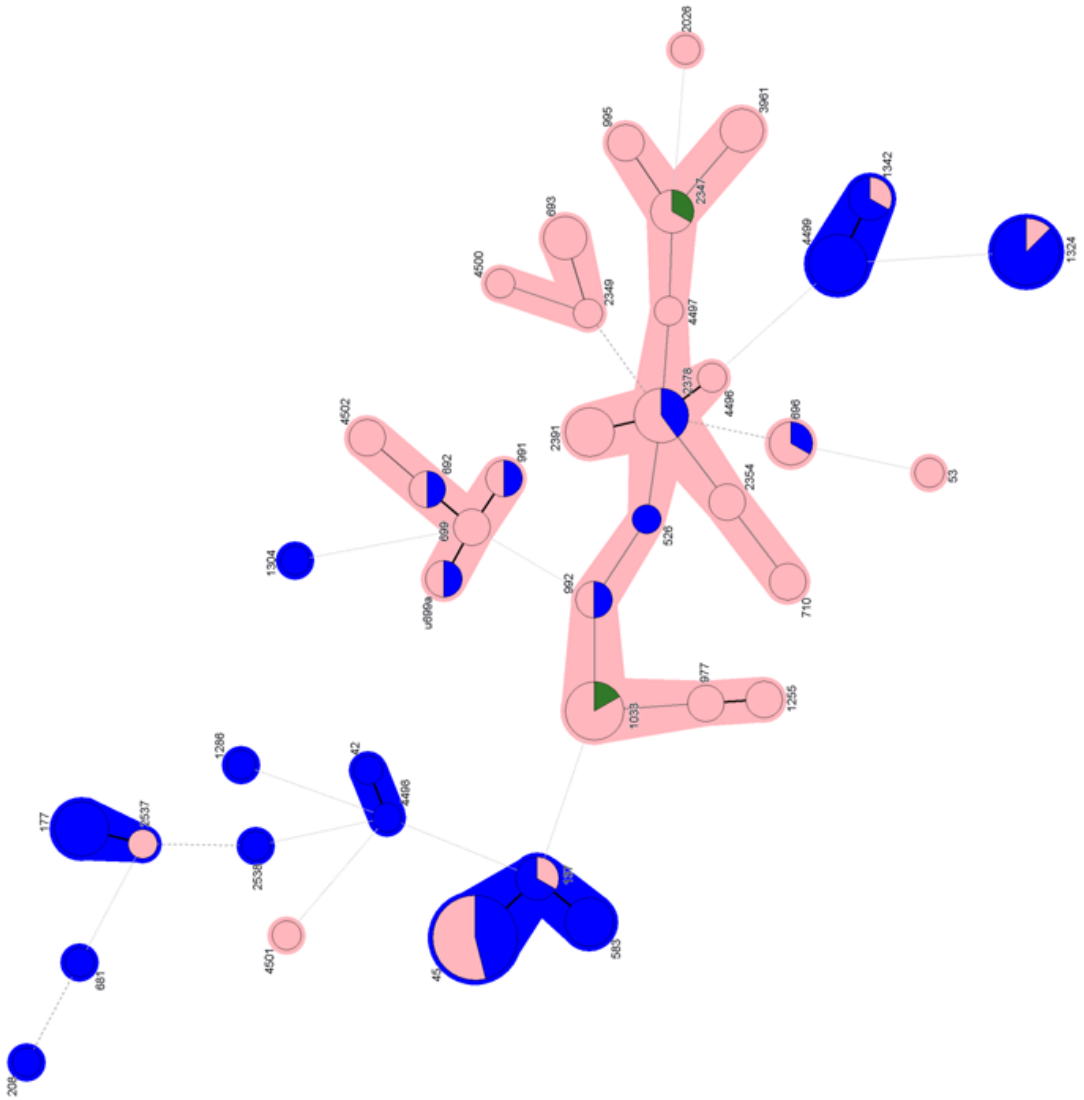


Figure A.5: The minimum spanning tree (MST) showing *flaA* and *porA* alleles clustered with the MLST profiles

APPENDIX **B**

Characterisation of *Campylobacter jejuni* from pets in New Zealand by combining multilocus sequence typing (MLST), *por*A and *fla*A typing

APPENDIX C

Phylogeny of housekeeping and *por*A genes from seven *Campylobacter jejuni* ST-474 genomes

C.1 Introduction

This appendix provides additional information on the details provided in Chapter 5. This section provides a detailed description of the the BLAST table and criteria adapted in retrieving the full length gene sequences from seven *C. jejuni* ST-474 genomes. A blast table created with consolidated information of blast hits and details about the parameters used to select the hits is shown in this section in addition to the description provided in Chapter 5. The results of the analyses conducted on seven housekeeping genes and other individual metabolic genes were presented and discussed in the manuscript of Chapter 5, however, the tables of GC content and codon usage were not included in the manuscript, instead the tables are presented in this appendix in order to provide a comprehensive report in Chapter 5.

C.2 Materials and methods

C.2.1 Retrieval of gene sequences

Table C.1 presented below provides a sample of the details of metabolic genes retrieval from the gene predictions of *C. jejuni* ST-474 genomes. The best hits for all the 25 genes were retained and the gene sequences were retrieved in FASTA format.

Table C.1: BLAST table of *C. jejuni* ST-474 metabolic genes

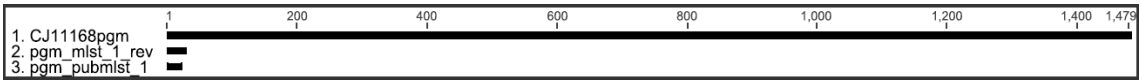
Isolate_gene_id	queryID	subjectID	% Identity	Align_length	Mis_Match	Gap_open	Query_start	Query_end	Sub_start	Sub_end	E_value	BitScore
*asp*A_P179a	15i11jejuni_P179a_orf00060_len1407	100	1407	0	0	1	1407	1	1407	0	2706	
*gln*A_P179a	15i11jejuni_P179a_orf00077_len1431	98.74	1431	18	0	1	1431	1	1431	0	2652	
*glt*A_P179a	15i11jejuni_P179a_orf00004_len1176	99.83	1176	2	0	94	1269	1	1176	0	2315	
*gly*A_P179a	15i11jejuni_P179a_orf00153_len1245	98.27	753	13	0	493	1245	493	1245	0	1306	
*glm*M_P179a	15i11jejuni_P179a_orf00004_len1479	98.97	1166	12	0	314	1479	314	1479	0	2145	
*tkt*_P179a	15i11jejuni_P179a_orf00028_len1899	99.89	943	1	0	957	1899	957	1899	0	1861	
*unc*A_P179a	15i11jejuni_P179a_orf00081_len1506	100	1506	0	0	1	1506	1	1506	0	2985	
*asp*A_P569a	15i07jejuni_569a_orf00002_len1407	100	1407	0	0	1	1407	1	1407	0	2706	
*gln*A_P569a	15i07jejuni_569a_orf00001_len1431	98.74	1431	18	0	1	1431	1	1431	0	2652	
*glt*A_P569a	15i07jejuni_569a_orf00032_len1176	99.83	1176	2	0	94	1269	1	1176	0	2315	
*gly*A_P569a	15i07jejuni_569a_orf00032_len1245	98.27	753	13	0	493	1245	493	1245	0	1306	
*glm*M_P569a	15i07jejuni_569a_orf00003_len1479	98.97	1166	12	0	314	1479	314	1479	0	2145	
*tkt*_P569a	15i07jejuni_569a_orf00014_len1899	99.89	943	1	0	957	1899	957	1899	0	1861	
*unc*A_P569a	15i07jejuni_569a_orf00023_len1506	100	1506	0	0	1	1506	1	1506	0	2985	
*asp*A_P694a	15i07jejuni_P694a_orf00099_len1407	100	1407	0	0	1	1407	1	1407	0	2706	
*gln*A_P694a	15i07jejuni_P694a_orf00028_len1431	98.74	1431	18	0	1	1431	1	1431	0	2652	
*glt*A_P694a	15i07jejuni_P694a_orf00032_len1176	99.4	1176	7	0	94	1269	1	1176	0	2276	
*gly*A_P694a	15i07jejuni_P694a_orf00035_len1245	98.27	753	13	0	493	1245	493	1245	0	1306	
*glm*M_P694a	15i07jejuni_P694a_orf00004_len1479	98.97	1166	12	0	314	1479	314	1479	0	2145	
*tkt*_P694a	15i07jejuni_P694a_orf00024_len1899	99.89	943	1	0	957	1899	957	1899	0	1861	
*unc*A_P694a	15i07jejuni_P694a_orf00121_len1506	100	1506	0	0	1	1506	1	1506	0	2985	
*asp*A_H704	15i08jejuni_H704_orf00013_len1407	100	1407	0	0	1	1407	1	1407	0	2706	
*gln*A_H704	15i08jejuni_H704_orf00047_len1431	98.74	1431	18	0	1	1431	1	1431	0	2652	
*glt*A_H704	15i08jejuni_H704_orf00094_len1176	99.83	1176	2	0	94	1269	1	1176	0	2315	

Table C.1 (continued)

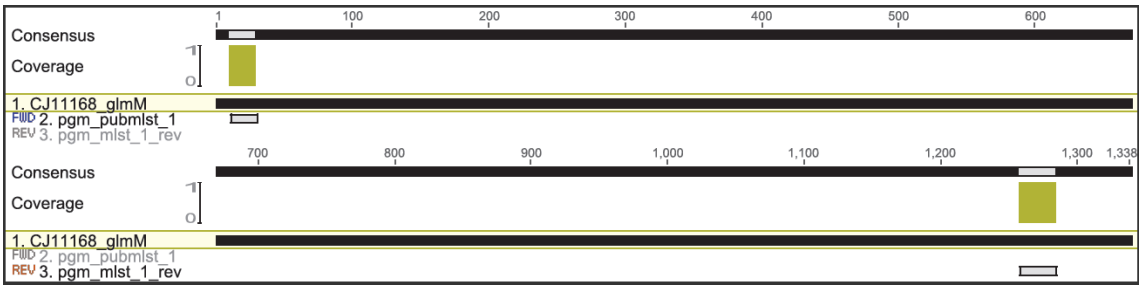
Isolate_gene_id	queryID	subjectID	% Identity	Align_length	Mis_Match	Gap_open	Query_start	Query_end	Sub_start	Sub_end	E_value	BitScore
*gly*A_H704	15i08jejuni_H704_orf00032_len1245	98.27	753	13	0	493	1245	493	1245	0	1306	
*glm*M_H704	15i08jejuni_H704_orf00152_len1479	98.97	1166	12	0	314	1479	314	1479	0	2145	
*tkt*_H704	15i08jejuni_H704_orf00036_len1899	99.89	943	1	0	957	1899	957	1899	0	1861	
*unc*A_H704	15i08jejuni_H704_orf00034_len1506	100	1506	0	0	1	1506	1	1506	0	2985	
*asp*A_H73020	13i04jejuni_73020_orf00058_len1407	100	1407	0	0	1	1407	1	1407	0	2706	
*gln*A_H73020	13i04jejuni_73020_orf00041_len1431	98.74	1431	18	0	1	1431	1	1431	0	2652	
*glt*A_H73020	13i04jejuni_73020_orf00004_len1176	99.83	1176	2	0	94	1269	1	1176	0	2315	
*gly*A_H73020	13i04jejuni_73020_orf00034_len1245	98.27	753	13	0	493	1245	493	1245	0	1306	
*glm*M_H73020	13i04jejuni_73020_orf00160_len1308	98.97	1166	12	0	314	1479	143	1308	0	2145	
*tkt*_H73020	13i04jejuni_73020_orf00034_len1899	99.89	943	1	0	957	1899	957	1899	0	1861	
*unc*A_H73020	13i04jejuni_73020_orf00079_len1506	100	1506	0	0	1	1506	1	1506	0	2985	
*asp*A_P110b	orf00058_238_44158_42752_len=1407	100	1407	0	0	1	1407	1	1407	0	2599	
*gln*A_P110b	orf00090_239_82615_81185_len=1431	98.74	1431	18	0	1	1431	1	1431	0	2543	
*glt*A_P110b	orf00137_232_112338_111070_len=1269	99.84	1269	2	0	1	1269	1	1269	0	2333	
*gly*A_P110b	orf00032_241_32671_31427_len=1245	98.15	1246	21	2	1	1245	1	1245	0	2172	
*glm*M_P110b	orf00003_236_2549_4027_len=1479	99.05	1479	14	0	1	1479	1	1479	0	2654	
*tkt*_P110b	orf00096_232_76278_78176_len=1899	99.95	1899	1	0	1	1899	1	1899	0	3502	
*unc*A_P110b	orf00037_238_28744_27239_len=1506	100	1506	0	0	1	1506	1	1506	0	2782	
*asp*A_H22082	orf00057_613_45303_46709_len=1407	100	1407	0	0	1	1407	1	1407	0	2599	
*gln*A_H22082	orf00091_612_82638_81208_len=1431	98.74	1431	18	0	1	1431	1	1431	0	2543	
*glt*A_H22082	orf00139_622_112247_111072_len=1176	99.83	1176	2	0	94	1269	1	1176	0	2161	
*gly*A_H22082	orf00032_610_32672_31428_len=1245	98.15	1246	21	2	1	1245	1	1245	0	2172	
*glm*M_H22082	orf00156_617_137365_135887_len=1479	99.05	1479	14	0	1	1479	1	1479	0	2654	
*tkt*_H22082	orf00097_622_76280_78178_len=1899	99.95	1899	1	0	1	1899	1	1899	0	3502	

Table C.1 (continued)

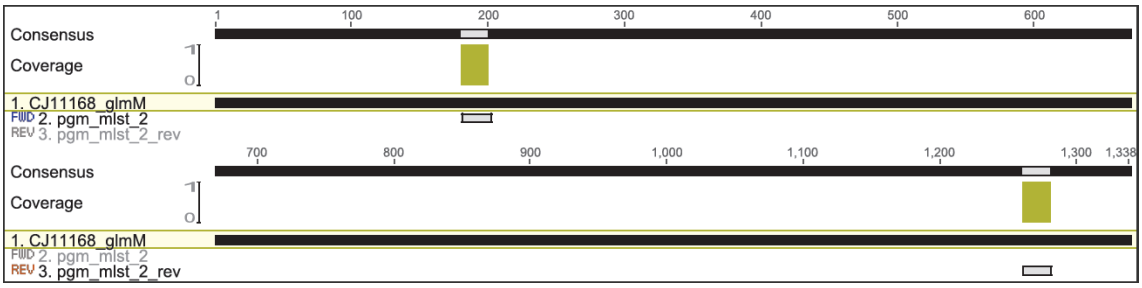
Isolate_gene_id	queryID	subjectID	% Identity	Align_length	Mis_Match	Gap_open	Query_start	Query_end	Sub_start	Sub_end	E_value	BitScore
*unc*A_H22082	orf00078_613_60717_62222_len=1506	100	1506	0	0	1	1506	1	1506	0	2782	



(a) *pgm* gene – primer assembly



(b) *glm*M gene – primer 1 assembly



(c) *glm*M gene – primer 2 assembly

(d) *glm*M gene – primers assembly report

Figure C.1: Comparison of primer sequences of *glm*M allele between *pgm* and *glm*M. The primer sequences used to amplify and sequence *pgm* allele were retrieved from the PubMLST database and assembled across *pgm* and *glm*M full length genes.

C.3 Results

C.3.1 Analysis of full length gene sequences of MLST seven housekeeping genes

GC content

Table C.2 provides the details on Guanine-Cytosine (GC) content of the seven MLST housekeeping genes of all the genomes analysed in Chapter 5. The *C. jejuni* ST-474 isolate P694a shows a relatively high proportion of GC content compared with other six ST-474 genomes.

Table C.2: Guanine-Cytosine (GC) content of seven MLST housekeeping genes of *C. jejuni* genomes

Genomes	*asp*A	*gln*A	*glt*A	*gly*A	*glm*M	*tkt*	*unc/atp*A	*arg*F	*aro*E	*atp*D	*dap*E	*fts*Z	*fum*C	*gap*A	*glt*B	*gro*EL	*hem*N	*ilv*D	*inf*B	*lys*A	*nuo*D	*pyc*A	*sdh*A	*trp*B	*trp*C
CJ11168	**0.333**	0.332	**0.358**	0.329	0.325	0.358	**0.352**	**0.35**	0.294	0.354	0.357	**0.356**	0.361	0.351	**0.347**	0.366	0.3	**0.37**	0.352	**0.325**	0.36	0.338	**0.37**	**0.351**	**0.324**
CJ26094	**0.333**	**0.333**	0.355	**0.332**	0.322	0.36	0.353	0.345	0.303	0.351	0.356	0.353	0.363	0.351	0.349	0.364	0.302	0.373	**0.351**	0.323	0.36	0.335	0.371	0.35	0.323
CJ81116	**0.33**	0.332	0.355	0.331	**0.321**	0.356	**0.352**	0.341	0.299	0.354	0.352	0.357	0.359	0.351	0.35	0.364	0.3	0.379	0.353	0.323	0.36	0.338	0.372	**0.351**	0.322
CJ81176	**0.333**	**0.333**	0.359	0.331	0.32	0.36	0.354	0.347	0.302	0.351	0.357	0.353	0.362	**0.357**	0.35	0.366	0.303	0.372	**0.351**	**0.325**	**0.359**	**0.339**	0.373	**0.351**	0.317
CJ8421	**0.333**	0.332	0.355	0.328	0.322	**0.357**	0.353	0.346	0.294	0.352	0.362	0.357	0.366	0.356	0.348	0.366	0.303	0.371	0.385	**0.325**	0.361	0.336	0.372	0.348	0.323
CJ8425	**0.333**	0.332	0.356	0.329	**0.321**	0.359	**0.352**	**0.35**	**0.295**	**0.353**	0.357	**0.356**	0.361	0.356	**0.347**	0.366	0.3	0.367	**0.351**	**0.325**	**0.359**	0.338	**0.37**	0.35	**0.324**
CJ8486	0.331	0.332	0.352	**0.332**	0.322	0.356	0.353	0.347	0.288	0.351	0.361	0.358	0.365	0.352	0.348	0.368	0.303	0.381	0.35	0.326	0.361	0.338	0.357	0.349	0.323
CJ9313	**0.333**	**0.333**	0.357	**0.332**	0.32	0.355	0.354	0.346	0.303	**0.353**	0.358	0.354	0.362	0.351	0.353	0.371	0.305	0.371	0.353	0.323	0.355	0.34	0.373	0.35	0.315
CJ936	**0.333**	0.331	0.356	0.329	**0.321**	0.361	**0.352**	0.34	**0.295**	0.354	0.359	0.358	0.361	0.358	**0.347**	0.379	0.3	**0.37**	**0.351**	**0.325**	**0.359**	**0.339**	**0.37**	**0.351**	0.323
CJD	0.332	0.335	0.364	0.336	0.319	0.358	0.355	0.331	0.302	0.351	0.357	0.351	0.361	0.352	**0.347**	0.364	0.302	0.374	0.355	0.32	0.363	0.334	0.384	0.345	**0.324**
CJIA3902	**0.333**	0.332	**0.358**	0.329	**0.321**	0.358	0.353	0.349	**0.295**	0.352	**0.35**	**0.356**	0.363	0.351	**0.347**	0.366	0.304	0.371	0.349	**0.325**	0.362	0.338	0.371	**0.351**	**0.324**
CJRM1221	0.334	0.332	0.356	**0.332**	0.322	0.359	0.353	**0.35**	0.293	0.351	0.361	0.358	0.364	0.354	0.35	0.368	0.306	0.372	**0.351**	0.323	**0.359**	0.338	0.373	0.35	0.323
H22082	**0.333**	**0.333**	**0.358**	**0.332**	**0.321**	**0.357**	**0.352**	**0.351**	**0.295**	0.354	**0.35**	**0.356**	**0.36***	**0.357**	**0.347**	**0.367**	**0.305**	**0.37**	**0.351**	**0.325**	**0.359**	**0.339**	**0.37**	**0.351**	**0.326***
H704	**0.333**	**0.333**	**0.358**	**0.332**	**0.321**	**0.357**	**0.352**	**0.351**	**0.295**	**0.353**	**0.35**	**0.356**	**0.358**	**0.357**	**0.347**	**0.367**	**0.305**	**0.37**	**0.351**	**0.325**	**0.359**	**0.339**	**0.37**	**0.351**	**0.326***
P110b	**0.333**	**0.333**	**0.358**	**0.332**	**0.321**	**0.357**	**0.352**	**0.351**	**0.295**	**0.353**	**0.35**	**0.356**	**0.358**	**0.357**	**0.347**	**0.367**	**0.305**	**0.37**	**0.351**	**0.325**	**0.359**	**0.339**	**0.37**	**0.351**	**0.324**
P179a	**0.333**	**0.333**	**0.358**	**0.332**	**0.321**	**0.357**	**0.352**	**0.351**	**0.295**	**0.353**	**0.35**	**0.356**	**0.358**	**0.357**	**0.347**	**0.367**	**0.305**	**0.37**	**0.351**	**0.325**	**0.359**	**0.339**	**0.37**	**0.351**	**0.324**
P569a	**0.333**	**0.333**	**0.358**	**0.332**	**0.321**	**0.357**	**0.352**	**0.351**	**0.295**	**0.353**	**0.35**	**0.356**	**0.358**	**0.357**	**0.347**	**0.367**	**0.305**	**0.37**	**0.351**	**0.325**	**0.359**	**0.339**	**0.37**	**0.351**	**0.324**
P694a	**0.333**	**0.333**	**0.358**	**0.332**	**0.321**	**0.357**	**0.352**	**0.351**	**0.295**	**0.353**	**0.35**	**0.356**	**0.36***	**0.357**	**0.347**	**0.367**	**0.305**	**0.37**	**0.351**	**0.325**	**0.359**	**0.339**	**0.37**	**0.351**	**0.324**
H73020	**0.333**	**0.333**	**0.358**	**0.332**	**0.321**	**0.357**	**0.352**	**0.351**	**0.295**	**0.353**	**0.35**	**0.356**	**0.358**	**0.357**	**0.347**	**0.367**	**0.305**	**0.37**	**0.351**	**0.325**	**0.359**	**0.339**	**0.37**	**0.351**	**0.324**

*:GC content differs between ST-474 genomes.

Table C.3: GC3 content of metabolic genes

Genomes	*asp*A	*gln*A	*glt*A	*gly*A	*glm*M	*tkt*	*unc*A	*arg*F	*aro*E	*atp*D	*dap*E	*fts*Z	*fum*C	*gap*A	*glt*B	*gro*EL	*hem*N	*ilv*D	*inf*B	*lys*A	*nuo*D	*pyc*A	*sdh*A	*trp*B	*trp*C
CJ11168	0.124	0.137	0.164	0.11	0.155	0.192	0.132	0.249	0.141	0.143	0.208	0.149	0.192	0.157	0.191	0.148	0.18	0.181	0.193	0.168	0.192	0.151	0.314	0.185	0.235
CJ26094	0.124	0.133	0.154	0.123	0.148	0.202	0.138	0.228	0.152	0.134	0.203	0.14	0.192	0.157	0.196	0.14	0.185	0.193	0.187	0.163	0.186	0.145	0.312	0.182	0.231
CJ81116	0.117	0.133	0.152	0.122	0.15	0.192	0.132	0.228	0.148	0.143	0.209	0.152	0.179	0.16	0.198	0.14	0.176	0.213	0.195	0.163	0.186	0.151	0.312	0.185	0.227
CJ81176	0.126	0.133	0.162	0.125	0.146	0.202	0.136	0.238	0.148	0.134	0.211	0.14	0.192	0.176	0.2	0.15	0.178	0.191	0.189	0.168	0.192	0.153	0.312	0.185	0.211
CJ8421	0.128	0.133	0.149	0.11	0.148	0.194	0.134	0.234	0.137	0.136	0.22	0.152	0.204	0.179	0.196	0.146	0.187	0.189	0.225	0.168	0.189	0.148	0.315	0.176	0.231
CJ8425	0.124	0.137	0.157	0.11	0.145	0.199	0.132	0.249	0.141	0.143	0.208	0.149	0.192	0.157	0.192	0.148	0.18	0.174	0.188	0.168	0.192	0.151	0.314	0.182	0.235
CJ8486	0.119	0.138	0.157	0.12	0.15	0.192	0.134	0.238	0.131	0.138	0.215	0.154	0.204	0.179	0.195	0.152	0.185	0.198	0.188	0.173	0.189	0.153	0.387	0.182	0.235
CJ9313	0.126	0.14	0.157	0.123	0.143	0.187	0.136	0.238	0.152	0.136	0.211	0.143	0.192	0.16	0.207	0.152	0.183	0.185	0.193	0.163	0.173	0.155	0.312	0.184	0.211
CJ936	0.124	0.133	0.157	0.11	0.145	0.202	0.132	0.23	0.141	0.134	0.211	0.157	0.192	0.157	0.191	0.128	0.18	0.181	0.19	0.168	0.192	0.155	0.314	0.185	0.235
CJD	0.126	0.142	0.177	0.13	0.15	0.192	0.136	0.199	0.148	0.13	0.214	0.138	0.192	0.189	0.194	0.139	0.18	0.206	0.2	0.158	0.199	0.143	0.336	0.171	0.235
CJIA3902	0.124	0.137	0.164	0.11	0.145	0.192	0.136	0.246	0.141	0.138	0.189	0.149	0.195	0.16	0.191	0.148	0.187	0.185	0.184	0.168	0.199	0.153	0.314	0.185	0.235
CJRM1221	0.128	0.133	0.157	0.12	0.148	0.197	0.136	0.241	0.133	0.134	0.215	0.157	0.201	0.172	0.2	0.15	0.196	0.194	0.19	0.168	0.189	0.151	0.333	0.182	0.231
H22082	0.124	0.14	0.164	0.12	0.145	0.192	0.132	0.246	0.141	**0.137**	0.189	0.149	**0.186**	0.176	0.192	0.15	0.192	0.181	0.19	0.168	0.192	**0.151**	0.314	0.185	**0.236**
H704	0.124	0.14	0.164	0.12	0.145	0.192	0.132	0.246	0.141	0.14	0.189	0.149	**0.186**	0.176	0.192	0.15	0.192	0.181	0.19	0.168	0.192	0.153	0.314	0.185	**0.236**
P110b	0.124	0.14	0.164	0.12	0.145	0.192	0.132	0.246	0.141	0.14	0.189	0.149	0.184	0.176	0.192	0.15	0.192	0.181	0.19	0.168	0.192	0.153	0.314	0.185	**0.235**
P179a	0.124	0.14	0.164	0.12	0.145	0.192	0.132	0.246	0.141	0.14	0.189	0.149	0.184	0.176	0.192	0.15	0.192	0.181	0.19	0.168	0.192	0.153	0.314	0.185	0.239
P569a	0.124	0.14	0.164	0.12	0.145	0.192	0.132	0.246	0.141	0.14	0.189	0.149	0.184	0.176	0.192	0.15	0.192	0.181	0.19	0.168	0.192	0.153	0.314	0.185	0.239
P694a	0.124	0.14	0.164	0.12	0.145	0.192	0.132	0.246	0.141	0.14	0.189	0.149	0.184	0.176	0.192	0.15	0.192	0.181	0.19	0.168	0.192	**0.151**	0.314	0.185	0.239
H73020	0.124	0.14	0.164	0.12	0.145	0.192	0.132	0.246	0.141	0.14	0.189	0.149	0.184	0.176	0.192	0.15	0.192	0.181	0.19	0.168	0.192	0.153	0.314	0.185	0.239

Differences in GC3 contents are shown in bold fonts

C.3.2 Codon usage

The codon usage bias indices (CBI) of all reference and ST-474 genomes are provided in this subsection to provide further information on the illustrations provided in Chapter 5.

Table C.4: Codon bias index (CBI) for twenty metabolic housekeeping genes

Genes	*asp*A	*gln*A	*glt*A	*gly*A	*glm*M	*tkt*	*unc/atp*A	*arg*F	*aro*E	*atp*D	*dap*E	*fts*Z	*fum*C	*gap*A	*glt*B	*gro*EL	*hem*N	*ilv*D	*inf*B	*lys*A	*nuo*D	*pyc*A	*sdh*A	*trp*B	*trp*C
CJ11168	**0.66**	0.643	0.59	0.703	0.63	0.561	**0.642**	0.554	**0.704**	0.642	0.558	0.647	0.548	0.61	0.563	0.644	0.597	**0.558**	0.562	**0.584**	0.584	0.624	0.601	**0.599**	0.586
CJ26094	0.658	0.647	0.607	0.679	0.629	0.544	0.636	0.553	0.674	0.649	0.569	0.65	0.559	0.615	0.558	0.661	0.589	0.543	0.571	0.598	0.569	0.628	0.596	0.603	0.581
CJ81116	0.673	0.641	0.61	0.676	0.631	0.549	0.644	0.564	0.679	0.644	**0.589**	0.633	0.571	0.619	0.553	0.671	0.6	0.551	0.557	0.593	0.569	0.622	0.606	**0.599**	0.57
CJ81176	0.656	0.647	0.594	0.682	0.638	0.543	0.64	0.539	0.692	0.655	0.557	0.65	0.547	**0.577**	0.548	**0.642**	0.594	0.56	**0.569**	**0.584**	0.587	0.62	0.598	**0.599**	0.611
CJ8421	**0.66**	0.648	0.604	0.703	0.639	0.567	0.641	0.558	0.706	0.65	0.555	0.641	0.539	0.581	0.559	0.651	0.583	0.546	0.537	0.586	**0.583**	0.627	0.596	0.605	0.581
CJ8425	**0.66**	0.643	0.601	0.703	**0.634**	0.561	**0.642**	0.554	**0.704**	0.642	0.558	0.647	0.548	0.61	0.563	0.643	0.597	0.571	0.563	**0.584**	**0.583**	0.624	0.601	0.603	0.586
CJ8486	0.665	**0.637**	0.591	0.687	0.625	0.56	0.643	0.55	0.701	**0.647**	0.559	0.635	0.54	0.576	0.557	0.64	0.589	0.537	0.568	0.579	**0.583**	0.617	-	0.602	0.574
CJ9313	0.656	0.638	0.601	0.679	0.642	0.552	0.64	0.539	0.674	0.649	0.558	**0.648**	0.547	0.612	0.545	0.664	0.593	0.549	0.561	0.598	0.584	0.617	0.6	0.595	0.611
CJ936	**0.66**	0.642	0.597	0.703	**0.634**	0.565	**0.642**	0.582	**0.704**	0.649	0.563	0.636	0.546	0.615	0.563	0.685	0.597	**0.558**	**0.569**	**0.584**	0.582	**0.619**	0.601	**0.599**	0.586
CJD	0.651	0.623	0.579	0.677	0.646	0.535	0.649	0.589	0.711	0.655	0.552	0.652	0.543	0.546	0.556	0.644	0.595	0.515	0.557	0.592	0.549	0.646	-	0.621	0.565
CJIA3902	**0.66**	0.643	**0.589**	0.703	**0.634**	0.561	0.638	**0.557**	**0.704**	**0.647**	**0.589**	0.647	0.545	0.612	0.563	0.644	0.586	0.549	0.573	**0.584**	0.573	0.622	**0.599**	**0.599**	0.586
CJRM1221	0.655	0.648	0.592	**0.689**	0.629	0.557	0.638	0.542	0.709	0.655	0.559	0.636	0.543	0.586	0.549	0.644	0.581	0.54	0.562	0.586	0.588	0.622	0.596	0.605	0.581
H22082	**0.66**	**0.637**	**0.589**	**0.689**	**0.634**	**0.558**	**0.642**	**0.557**	**0.704**	**0.647**	**0.589**	**0.648**	**0.555***	**0.577**	**0.564**	**0.642**	**0.578**	**0.558**	**0.569**	**0.584**	**0.583**	**0.615***	**0.599**	**0.599**	**0.593**
H704	**0.66**	**0.637**	**0.589**	**0.689**	**0.634**	**0.558**	**0.642**	**0.557**	**0.704**	**0.647**	**0.589**	**0.648**	**0.558**	**0.577**	**0.564**	**0.642**	**0.578**	**0.558**	**0.569**	**0.584**	**0.583**	**0.619**	**0.599**	**0.599**	**0.593**
P110b	**0.66**	**0.637**	**0.589**	**0.689**	**0.634**	**0.558**	**0.642**	**0.557**	**0.704**	**0.647**	**0.589**	**0.648**	**0.557**	**0.577**	**0.564**	**0.642**	**0.578**	**0.558**	**0.569**	**0.584**	**0.583**	**0.619**	**0.599**	**0.599**	**0.586***
P179a	**0.66**	**0.637**	**0.589**	**0.689**	**0.634**	**0.558**	**0.642**	**0.557**	**0.704**	**0.647**	**0.589**	**0.648**	**0.557**	**0.577**	**0.564**	**0.642**	**0.578**	**0.558**	**0.569**	**0.584**	**0.583**	**0.619**	**0.599**	**0.599**	**0.58**
P569a	**0.66**	**0.637**	**0.589**	**0.689**	**0.634**	**0.558**	**0.642**	**0.557**	**0.704**	**0.647**	**0.589**	**0.648**	**0.557**	**0.577**	**0.564**	**0.642**	**0.578**	**0.558**	**0.569**	**0.584**	**0.583**	**0.619**	**0.599**	**0.599**	**0.58**
P694a	**0.66**	**0.637**	**0.589**	**0.689**	**0.634**	**0.558**	**0.642**	**0.557**	**0.704**	**0.647**	**0.589**	**0.648**	**0.555***	**0.577**	**0.564**	**0.642**	**0.578**	**0.558**	**0.569**	**0.584**	**0.583**	**0.615***	**0.599**	**0.599**	**0.58**
H73020	**0.66**	**0.637**	**0.589**	**0.689**	**0.634**	**0.558**	**0.642**	**0.557**	**0.704**	**0.647**	**0.589**	**0.648**	**0.557**	**0.577**	**0.564**	**0.642**	**0.578**	**0.558**	**0.569**	**0.584**	**0.583**	**0.619**	**0.599**	**0.599**	**0.58**

*:Difference in CBI

–:Differences in CBI among ST-474 genomes

Table C.5: Codon usage bias (Scaled chisquare) indices for metabolic housekeeping genes from reference and ST-474 genomes

Genes	*arg*F	*aro*E	*asp*A	*atp*D	*dap*E	*fts*Z	*fum*C	*gap*A	*gln*A	*glt*A	*glt*B	*gly*A	*gro*EL	*hem*N	*ilv*D	*Inf*B	*lys*A	*nuo*D	*glm*M	*pyc*A	*sdh*A	*tkt*	*trp*B	*trp*C	*Unc*A
CJ11168	0.565	0.916	**0.734**	**0.829**	0.548	0.686	0.556	0.678	0.702	0.625	0.507	0.847	**0.76**	0.649	**0.56**	0.52	**0.627**	0.642	0.668	0.655	0.617	**0.571**	**0.605**	0.605	**0.739**
CJ26094	0.581	0.867	0.749	0.849	0.562	0.705	0.575	0.69	0.72	0.658	0.516	0.818	0.812	**0.611**	0.562	0.534	0.672	0.632	0.691	0.68	**0.613**	0.548	0.61	0.601	0.713
CJ81116	0.606	0.903	0.779	0.83	0.571	0.676	0.604	0.682	**0.7**	0.667	0.514	0.811	0.818	0.642	0.536	0.523	0.677	0.634	0.686	0.66	0.61	0.557	**0.605**	0.581	0.741
CJ81176	0.574	0.92	0.741	0.857	0.544	0.705	0.555	**0.638**	0.72	0.618	0.509	0.799	0.768	0.642	0.571	**0.528**	0.633	0.648	0.686	0.654	0.607	0.541	**0.605**	0.662	0.721
CJ8421	0.596	0.942	0.75	0.849	0.538	0.683	0.544	0.649	0.718	0.66	0.52	0.851	0.774	0.597	0.57	0.503	0.622	0.611	0.7	0.688	0.611	0.588	0.616	0.601	0.732
CJ8425	0.565	**0.907**	**0.734**	**0.829**	0.548	0.686	0.556	0.678	0.702	0.646	0.527	0.847	**0.76**	0.649	0.571	0.531	**0.627**	**0.641**	**0.695**	0.655	0.617	**0.571**	0.61	0.605	**0.739**
CJ8486	0.586	0.955	0.771	0.843	0.545	0.675	0.546	0.643	0.685	**0.626**	0.517	**0.836**	0.747	0.618	0.551	**0.528**	0.611	0.611	0.674	0.642	0.616	0.573	0.608	0.589	0.729
CJ9313	0.572	0.867	0.741	0.85	0.546	0.693	0.555	0.679	0.701	0.645	0.493	0.818	0.764	0.64	0.553	0.521	0.672	0.665	0.71	0.65	0.612	0.567	0.613	0.658	0.721
CJ936	0.782	**0.907**	**0.734**	0.849	0.556	0.658	0.557	0.69	0.709	0.64	**0.53**	0.847	0.826	0.649	**0.56**	**0.528**	**0.627**	0.636	**0.695**	**0.644**	0.617	0.577	**0.605**	0.604	**0.739**
CJD	0.612	0.895	**0.734**	0.837	0.529	0.709	0.556	0.605	0.674	0.606	0.516	0.791	0.846	0.615	0.503	0.522	0.682	0.591	0.728	0.69	0.643	0.537	0.625	0.574	0.706
CJIA3902	0.572	0.907	**0.734**	0.843	**0.601**	0.686	**0.543**	0.679	0.702	0.629	**0.53**	0.847	**0.76**	0.624	0.553	0.535	**0.627**	0.618	**0.695**	0.653	0.611	**0.571**	0.605	0.605	0.723
CJRM1221	0.587	0.958	0.742	0.854	0.545	0.658	0.549	0.649	0.718	0.639	0.507	**0.833**	0.757	**0.607**	0.564	0.522	0.638	0.64	0.691	0.66	0.602	**0.571**	0.612	0.601	0.723
H22082	**0.584**	**0.907**	**0.734**	**0.832**	**0.601**	**0.688**	**0.577**	**0.638**	**0.7**	**0.626**	**0.531**	**0.833**	**0.759**	**0.609**	**0.56**	**0.528**	**0.627**	**0.641**	**0.695**	**0.647**	**0.613**	**0.571**	**0.605**	**0.605**	**0.739**
H704	**0.584**	**0.907**	**0.734**	**0.832**	**0.601**	**0.688**	**0.565**	**0.638**	**0.7**	**0.626**	**0.531**	**0.833**	**0.759**	**0.609**	**0.56**	**0.528**	**0.627**	**0.641**	**0.695**	**0.648**	**0.613**	**0.571**	**0.605**	**0.605**	**0.739**
P110b	**0.584**	**0.907**	**0.734**	**0.832**	**0.601**	**0.688**	**0.565**	**0.638**	**0.7**	**0.626**	**0.531**	**0.833**	**0.759**	**0.609**	**0.56**	**0.528**	**0.627**	**0.641**	**0.695**	**0.648**	**0.613**	**0.571**	**0.605**	**0.605**	**0.739**
P179a	**0.584**	**0.907**	**0.734**	**0.832**	**0.601**	**0.688**	**0.565**	**0.638**	**0.7**	**0.626**	**0.531**	**0.833**	**0.759**	**0.609**	**0.56**	**0.528**	**0.627**	**0.641**	**0.695**	**0.648**	**0.613**	**0.571**	**0.605**	**0.605**	**0.739**
P569a	**0.584**	**0.907**	**0.734**	**0.832**	**0.601**	**0.688**	**0.565**	**0.638**	**0.7**	**0.626**	**0.531**	**0.833**	**0.759**	**0.609**	**0.56**	**0.528**	**0.627**	**0.641**	**0.695**	**0.648**	**0.613**	**0.571**	**0.605**	**0.605**	**0.739**
P694a	**0.584**	**0.907**	**0.734**	**0.832**	**0.601**	**0.688**	**0.577**	**0.638**	**0.7**	**0.626**	**0.531**	**0.833**	**0.759**	**0.609**	**0.56**	**0.528**	**0.627**	**0.641**	**0.695**	**0.647**	**0.613**	**0.571**	**0.605**	**0.605**	**0.739**
H73020	**0.584**	**0.907**	**0.734**	**0.832**	**0.601**	**0.688**	**0.565**	**0.638**	**0.7**	**0.626**	**0.531**	**0.833**	**0.759**	**0.609**	**0.56**	**0.528**	**0.627**	**0.641**	**0.695**	**0.648**	**0.613**	**0.571**	**0.60**		

Test of neutrality

Synonymous and non-synonymous sites with Tajima's D values are presented in Table C.6.

Table C.6: Tajima D test values, synonymous and nonsynonymous sites of metabolic housekeeping genes

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*arg*F	33, 39, 51, 63, 69, 73, 75, 114, 129, 138, 156, 156, 163, 165, 168, 183, 186, 195, 204, 207, 216, 228, 231, 234, 243, 255, 264, 282, 291, 294, 300, 303, 315, 330, 333, 339, 360, 363, 372, 390, 396, 405, 420, 465 (44)	259, 265, 280, 308, 364, 397, 423 (7)	-0.96811	1.23974
*aro*E	249, 261, 294, 315, 381, 408, 435, 456, 501, 564, 594, 630, 645, 654, 678, 699, 702, 717, 744, 771 (20)	274, 317, 337, 392, 418, 467, 469, 484, 514, 559, 577, 734 (12)	-1.43911	1.11983
*asp*A	39, 141, 198, 210, 234, 252, 252, 255, 276, 336, 414, 441, 459, 474, 615, 627, 663, 702, 774, 792, 897, 960, 1032, 1068, 1098, 1104, 1152,1176, 1203, 1218, 1251, 1275, 1290, 1305, 1317, 1338, 1368, 1378, 1383, 1386, 1389 (41)	538, 1094, 1117, 1348, 1369, 1393 (6)	-0.59233	1.94469
*atp*D	63, 231, 318, 348, 408, 522, 597, 651, 693, 708, 714, 823, 882, 903, 924, 969, 996, 1026, 1047, 1101, 1134, 1137, 1251, 1296, 1350, 1359, 1365, 1368, 1374, 1377, 1380, 1386, 1387, 1389 (34)	313 1268 1383 1385 1396 (5)	-1.43984	1.63866

Table C.6 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*dap*E	126, 129, 129, 138, 141, 144, 147, 153, 168, 171, 204, 207, 213, 216, 219, 231, 240, 252, 282, 285, 297, 300, 306, 309, 315, 315, 321, 333, 336, 345, 351, 352, 354, 354, 357, 372, 387, 390, 396, 402, 402, 405, 439, 453, 462, 462, 465, 468, 471, 474, 477, 498, 522, 531, 543, 561, 567, 603, 603, 606, 609, 642, 645, 678, 681, 690, 705, 714, 717, 747, 756, 786, 843, 864, 867, 897, 912, 921, 936, 942, 945, 990, 1035, 1041, 1059, 1065 (86)	49, 162, 227, 228, 230, 232, 235, 242, 242, 247, 247, 253, 259, 313, 322, 323, 328, 332, 334, 335, 392, 409, 424, 427, 431, 433, 486, 560, 676, 699, 701 (31)	-1.28806	1.65624
*fts*Z	90, 93, 138, 153, 216, 219, 237, 246, 252, 348, 432, 444, 543, 561, 612, 645, 681, 690, 696, 699, 705, 783, 798, 840, 903, 936, 963, 1044, 1053, 1062, 1071, 1086 (32)	1039 (1)	-0.24838	6.57937
*fum*C	9, 21, 27, 45, 69, 108, 168, 186, 297, 303, 354, 357, 366, 381, 399, 435, 453, 576, 586, 591, 603, 624, 654, 657, 669, 687, 690, 693, 702, 711, 744, 765, 819, 825, 834, 837, 843, 846, 861, 882, 889, 897, 900, 909, 930, 987, 1011, 1017, 1038, 1041, 1068, 1125, 1167, 1173, 1206, 1227, 1230, 1231, 1236, 1269, 1281, 1311, 1314, 1320, 1335, 1347, 1350, 1380 (68)	142 247 358 459 485 535 772 994 1031 1240 1312 1333 1381 (13)	-0.99988	0.82645
*gap*A	36, 153, 186, 288, 342, 378, 381, 399, 501, 552, 567, 570, 582, 594, 630, 660, 672, 678, 729, 732, 735, 750, 777, 819, 870, 876, 882, 921, 948, 961 (30)	167, 206, 262, 316, 340, 752, 759, 790, 808, 878, 901, 902 (12)	-0.38301	-5.77967
*gln*A	69, 84, 123, 147, 156, 180, 231, 258, 264, 279, 291, 372, 378, 399, 429, 489, 513, 537, 585, 615, 630, 711, 768, 777, 852, 870, 945, 951, 972, 978, 996, 1029, 1053, 1086, 1126, 1188, 1191, 1222, 1260, 1281 (40)	127, 230, 354, 358, 371, 448, 847 (7)	-0.14032	-2.58364

Table C.6 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*glt*A	105, 117, 129, 183, 192, 213, 234, 312, 325, 327, 351, 372, 399, 453, 495, 516, 528, 555, 561, 567, 579, 585, 654, 663, 708, 756, 780, 816, 828, 843, 894, 921, 942, 996, 999, 1005, 1011, 1077, 1080, 1089, 1114, 1140, 1158, 1203, 1236, 1305 (46)	337, 517, 560, 680, 782, 1233 (6)	-0.8	0.55457
*glt*B	48, 75, 111, 159, 201, 213, 285, 339, 465, 486, 492, 498, 528, 528, 531, 588, 591, 627, 642, 657, 675, 681, 684, 711, 717, 729, 732, 741, 750, 786, 792, 810, 822, 855, 858, 927, 927, 957, 993, 1014, 1044, 1047, 1068, 1147, 1149, 1170, 1203, 1233, 1236, 1239, 1353, 1389, 1401, 1431, 1443, 1467, 1482, 1560, 1575, 1581, 1590, 1605, 1635, 1656, 1677, 1689, 1698, 1809, 1866, 1899, 1902, 1962, 1968, 1974, 2034, 2067, 2106, 2118, 2148, 2151, 2178, 2224, 2226, 2229, 2235, 2241, 2280, 2313, 2316, 2325, 2328, 2349, 2412, 2418, 2436, 2511, 2550, 2553, 2604, 2607, 2637, 2649, 2700, 2788, 2802, 2829, 2838, 2853, 2895, 2898, 2928, 2934, 2937, 2964, 2967, 2973, 2982, 2991, 2992, 3003, 3084, 3084, 3087, 3123, 3156, 3156, 3177, 3180, 3189, 3204, 3210, 3258, 3279, 3318, 3342, 3387, 3417, 3420, 3429, 3441, 3447, 3462, 3489, 3540, 3573, 3579, 3612, 3636, 3639, 3645, 3648, 3651, 3684, 3687, 3717, 3723, 3726, 3735, 3753, 3765, 3768, 3807, 3837, 3855, 3861, 3867, 3876, 3885, 3906, 3942, 3945, 3945, 3954, 3957, 3999, 3999, 4020, 4023, 4026, 4029, 4032, 4092, 4098, 4119, 4128, 4146, 4149, 4161, 4167, 4239, 4260, 4263, 4281, 4284, 4305, 4356 (196)	172, 314, 518, 869, 1261, 1381, 2002, 2084, 2098, 2210, 2573, 2710, 2998, 3439, 3463, 3614, 3662, 4033, 4311, 4324, 4394, 4486 (22)	-1.47845	1.17223

Table C.6 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*gly*A	39, 66, 141, 165, 183, 219, 306, 357, 396, 405, 411, 429, 468, 474, 483, 490, 492, 504, 510, 552, 567, 570, 612, 618, 621, 639, 645, 657, 663, 666, 684, 690, 693, 744, 762, 789, 804, 822, 858, 858, 876, 933, 957, 960, 966, 1002, 1008, 1023, 1035, 1083, 1104 (51)	126, 223, 350, 562, 640, 718, 721, 835, 874, 1037, 1156, 1177 (12)	-0.15614	13.2048
*gro*EL	765, 771, 822, 825, 888, 903, 936, 951, 963, 1002, 1008, 1023, 1029, 1032, 1035, 1038, 1056, 1062, 1080, 1116, 1119, 1170, 1182, 1188, 1200, 1236, 1245, 1275, 1285, 1290, 1293, 1305, 1308, 1317, 1356, 1368, 1419, 1425, 1443, 1452, 1500, 1509, 1530, 1542, 1560, 1578, 1620, 1626 (48)	991, 1582 (2)	-1.16813	0.41075
*hem*N	48, 51, 123, 141, 174, 192, 198, 201, 207, 210, 258, 273, 288, 289, 291, 297, 315, 315, 318, 339, 372, 423, 429, 432, 486, 498, 525, 537, 585, 591, 591, 591, 624, 660, 684, 763, 783, 807, 816, 819, 825, 828, 843, 903, 927, 930, 957, 975, 981, 996, 1005, 1026, 1050, 1065, 1068, 1104, 1140, 1206, 1266, 1272, 1290 (61)	82, 269, 271, 356, 404, 443, 445, 455, 472, 544, 584, 601, 810, 865, 934, 955, 986, 1014, 1069, 1070, 1163, 1174, 1189, 1199, 1225 (25)	-0.17368	0.42345
*ilv*D	201, 225, 249, 276, 288, 300, 309, 345, 348, 366, 396, 399, 429, 444, 450, 468, 471, 492, 522, 576, 579, 600, 612, 642, 654, 660, 666, 721, 729, 744, 750, 765, 777, 816, 855, 864, 873, 891, 897, 903, 921, 951, 954, 963, 963, 969, 1029, 1032, 1035, 1071, 1077, 1098, 1107, 1128, 1131, 1137, 1158, 1170, 1260, 1263, 1275, 1293, 1296, 1320, 1323, 1347, 1377, 1389, 1401, 1416, 1437, 1518, 1530, 1548, 1560, 1569, 1599, 1644, 1665 (79)	226, 286, 346, 508, 781, 949, 982, 1051, 1159, 1216, 1276, 1543, 1552, 1561, 1564, 1583, 1603, 1671 (18)	-0.45085	2.19417

Table C.6 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*inf*B	993, 996, 1020, 1050, 1053, 1056, 1059, 1062, 1128, 1140, 1188, 1212, 1212, 1230, 1233, 1236, 1263, 1266, 1275, 1278, 1284, 1332, 1353, 1365, 1377, 1392, 1416, 1440, 1449, 1467, 1476, 1497, 1512, 1530, 1551, 1587, 1593, 1596, 1653, 1659, 1686, 1692, 1698, 1740, 1743, 1773, 1785, 1797, 1803, 1815, 1824, 1845, 1857, 1860, 1866, 1881, 1890, 1893, 1893, 1914, 1938, 1941, 1962, 1998, 2007, 2011, 2016, 2046, 2070, 2100, 2109, 2124, 2127, 2136, 2136, 2142, 2148, 2184, 2229, 2235, 2244, 2259, 2268, 2286, 2298, 2301, 2319, 2376, 2394, 2412, 2421, 2427, 2433, 2436, 2445, 2475, 2478, 2484, 2487, 2490, 2493, 2499, 2532, 2538, 2550, 2562, 2595, 2622 (108)	1100, 1252, 1547, 1697, 1705, 1729, 1732, 1892, 2119, 2230, 2533 (11)	-0.84724	1.961
*lys*A	21, 33, 36, 48, 63, 105, 126, 129, 153, 198, 222, 234, 363, 384, 474, 483, 534, 540, 582, 603, 636, 666, 672, 693, 702, 726, 729, 741, 753, 765, 790, 825, 876, 945, 960, 984, 990, 1017, 1023, 1026, 1035, 1077, 1089, 1113, 1122 (45)	73, 233, 472, 491, 586, 599, 694, 706, 718, 718, 944, 946, 1034, 1171 (14)	-0.65465	-0.0504
*nuo*D	18, 66, 78, 84, 90, 93, 102, 105, 114, 117, 132, 147, 156, 165, 183, 222, 249, 267, 285, 315, 330, 336, 336, 375, 381, 393, 402, 426, 438, 450, 453, 483, 498, 519, 543, 550, 660, 678, 684, 696, 706, 720, 756, 762, 771, 780, 792, 801, 813, 852, 864, 891, 894, 942, 954, 966, 993, 1008, 1032, 1035, 1041, 1050, 1089, 1167, 1209 (65)	370, 566, 584, 875, 892, 1009, 1069, 1085 (8)	-0.96403	0.35281

Table C.6 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*glm*M	81, 84, 171, 192, 219, 267, 291, 318, 336, 339, 363, 366, 372, 417, 441, 450, 456, 465, 492, 552, 573, 615, 636, 645, 681, 693, 705, 711, 729, 798, 810, 813, 813, 813, 816, 819, 864, 867, 879, 885, 897, 915, 939, 964, 966, 972, 990, 996, 1005, 1020, 1044, 1053, 1056, 1083, 1101, 1119, 1122, 1227, 1263 (59)	23, 130, 355, 689, 700, 874, 971, 1052, 1124, 1142, 1286 (11)	-0.66672	0.35281
*pyc*A	57, 78, 126, 150, 156, 174, 177, 204, 216, 237, 261, 270, 282, 300, 333, 339, 357, 360, 366, 367, 384, 402, 414, 417, 426, 429, 435, 456, 459, 477, 489, 495, 498, 504, 528, 546, 567, 576, 579, 615, 624, 657, 660, 672, 687, 699, 720, 726, 732, 738, 750, 768, 798, 804, 849, 870, 894, 918, 954, 984, 993, 1107, 1137, 1140, 1200, 1203, 1260, 1275, 1278, 1305, 1332 (71)	194, 295, 334, 376, 425, 427, 430, 436, 718 (9)	-1.99257	0.93794
*sdh*A	135, 219, 336, 366, 384, 432, 501, 591, 624, 672, 708, 762, 804, 945, 957, 987, 1005, 1122, 1152, 1242, 1260, 1302, 1302, 1338, 1462, 1464, 1551, 1569, 1599, 1683 (30)	118, 119, 172, 583, 610, 659, 946, 1198, 1355, 1397, 1533 (11)	-0.82915	-0.02599
tkt	54, 63, 102, 150, 159, 174, 174, 237, 240, 246, 253, 297, 342, 363, 366, 387, 399, 408, 409, 414, 450, 522, 531, 555, 660, 690, 756,783, 819, 843, 858, 876, 921, 957, 966, 1029, 1089, 1092, 1104, 1107, 1122, 1125, 1131, 1164, 1233, 1254, 1287, 1311, 1320, 1323, 1329, 1341, 1350, 1350, 1362, 1383, 1389, 1416, 1431, 1476, 1488, 1512, 1518, 1545, 1557, 1560, 1591, 1620, 1621, 1674, 1749, 1761, 1791, 1797, 1803, 1806, 1824, 1827, 1842, 1857, 1863, 1875, 1884 (83)	4, 91, 251, 415, 532, 592, 688, 935, 971, 977, 1012, 1123, 1138, 1480, 1486, 1525, 1711, 1804, 1807, 1849, 1850, 1856, 1876, 1878 (24)	-0.79287	1.30603

Table C.6 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*trp*B	12, 129, 132, 150, 153, 165, 189, 201, 240, 255, 264, 339, 351, 363, 396, 402, 408, 477, 498, 507, 528, 609, 684, 702, 705, 708, 771, 774, 780, 810, 843, 879, 918, 945, 954, 972, 990, 1008, 1014, 1041, 1044, 1053, 1056, 1110, 1119, 1131, 1161, 1164 (48)	61, 79, 80, 109, 128, 596, 736, 772, 782, 811, 987, 1051, 1105, 1152 (14)	-2.25126	0.99273
*trp*C	88, 111, 138, 195, 204, 207, 210, 240, 285, 294, 345, 417, 420, 429, 486, 516, 537, 552, 564, 567, 702, 750, 753 (23)	103, 268, 634, 697, 721, 754 (6)	-0.97295	2.12579
*unc*A	90, 204, 294, 315, 339, 357, 414, 438, 507, 537, 564, 615, 678, 732, 756, 792, 823, 840, 864, 876, 891, 942, 946, 948, 957, 1014, 1020, 1050, 1065, 1086, 1155, 1197 1206, 1239, 1257, 1269, 1275, 1356, 1362, 1392, 1419, 1425 (42)	361, 721, 1186, 1360, 1405 (5)	-0.79287	1.30603

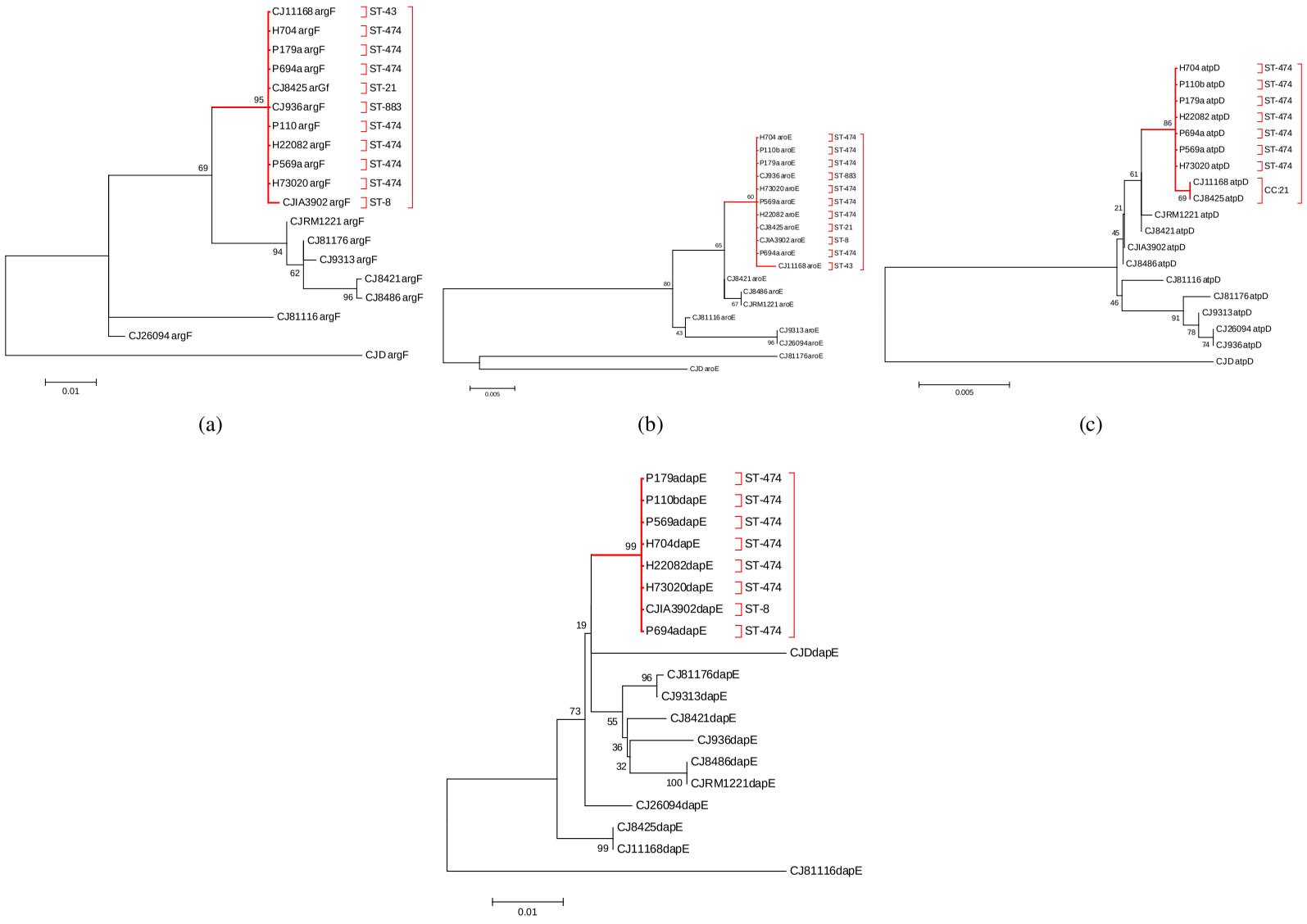


Figure C.2: Maximum likelihood gene trees of metabolic genes – A. Individual gene trees of *arg*F, *aro*E, *atp*D and *dap*E constructed using MEGA v5. The numbers refer to bootstrap values corresponding to each branch.

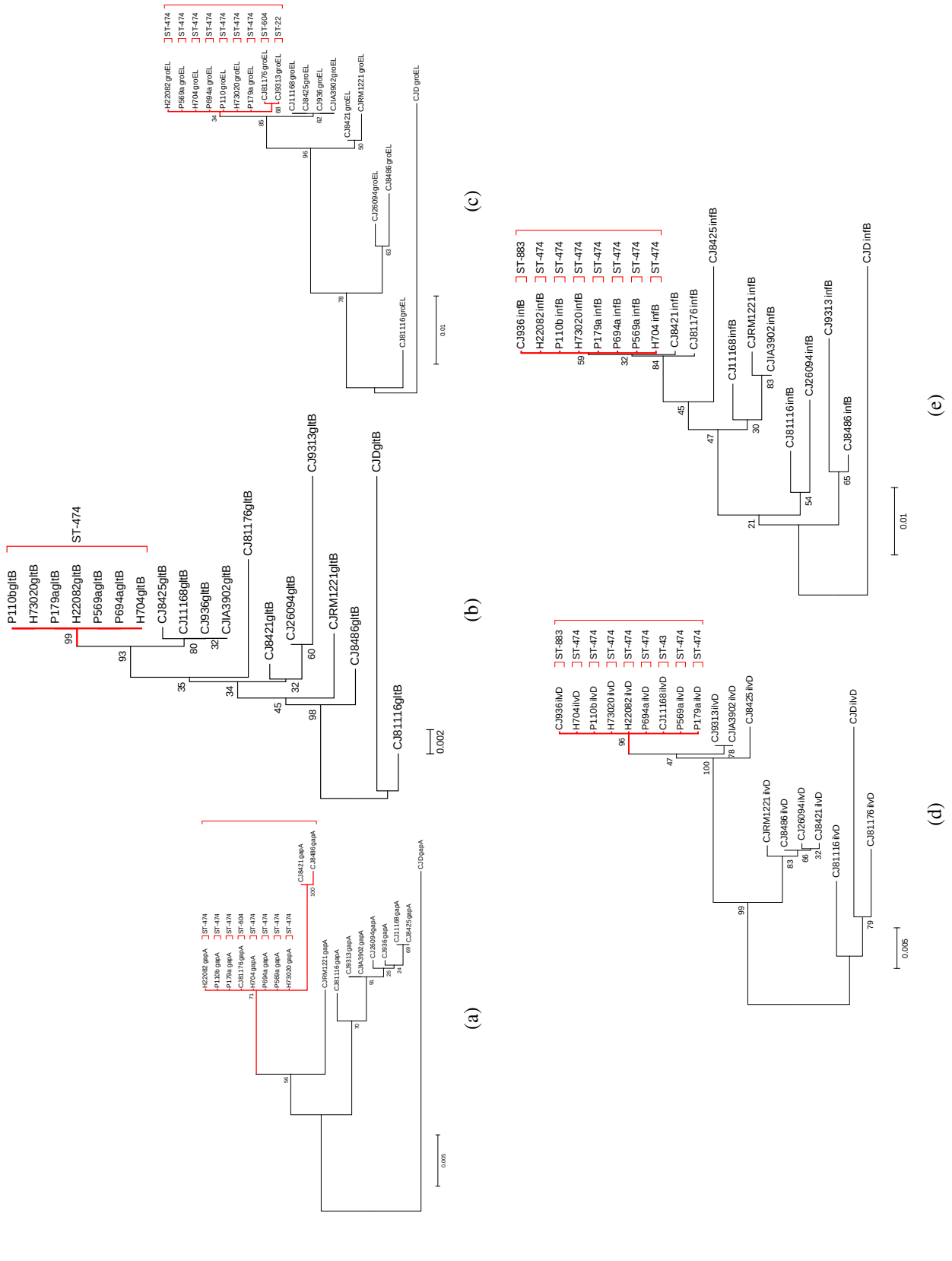


Figure C.3: Maximum likelihood gene trees of metabolic genes – B. Individual gene trees of *gap*A, *glt*B, *gro*EL, *ilv*D and *inf*B constructed using MEGA v5.

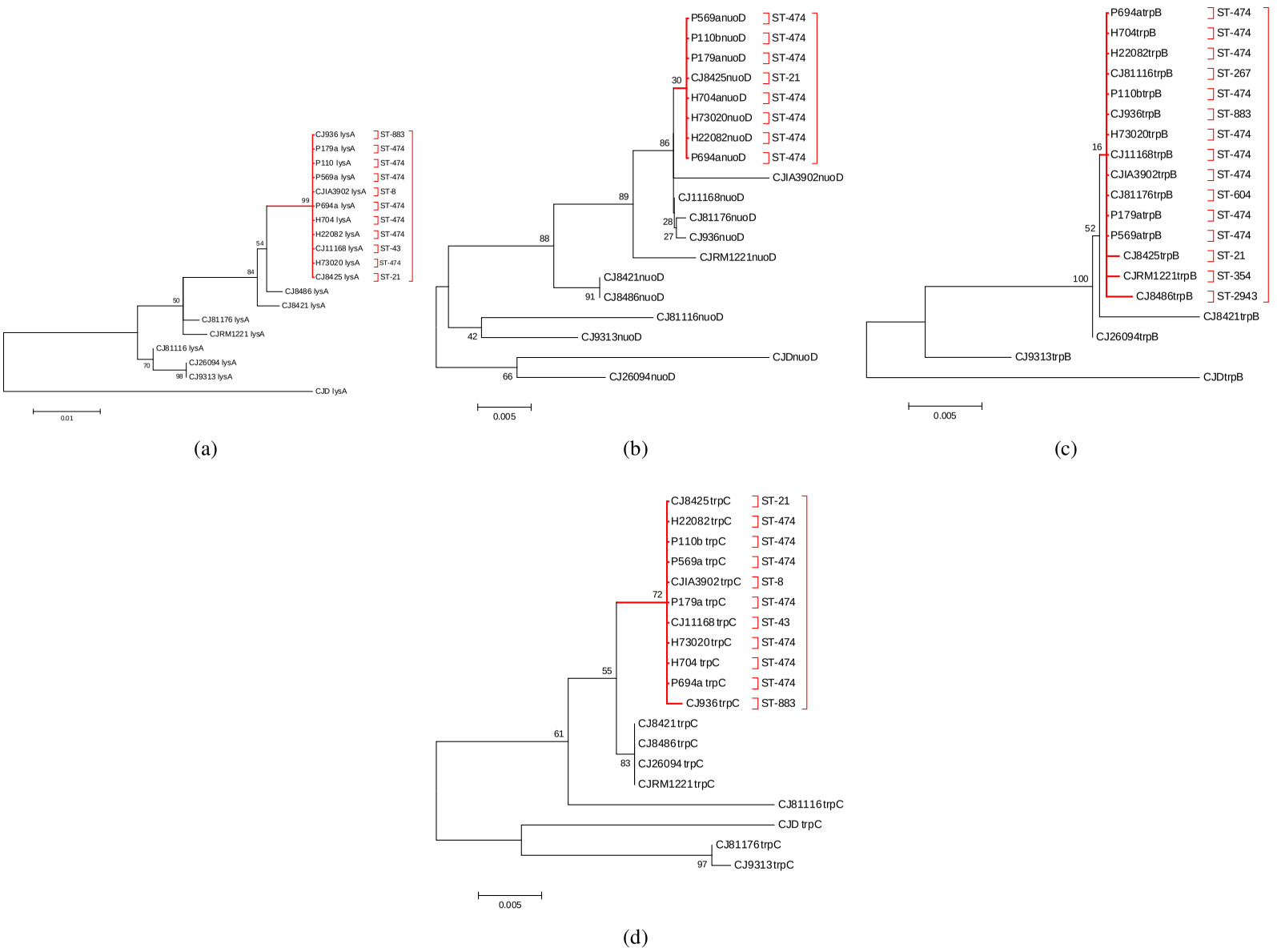


Figure C.4: Maximum likelihood gene trees of metabolic genes – C. Individual gene trees of *lys*A, *nuo*D, *trp*B and *trp*C constructed using MEGA v5. The numbers refer to bootstrap values corresponding to each branch.

APPENDIX **D**

A genome wide comparison of selected subsets of ribosomal genes and genes involved in DNA replication and repair among seven *Campylobacter jejuni* ST-474 isolates

D.1 Introduction

This appendix provides additional information on the details provided in Chapter 6. The results of the analyses conducted on 25 (repair, ribosomal and nucleotide metabolic genes) genes have been presented and discussed in the manuscript of Chapter 6, however, the tables of GC, GC3 contents and codon usage bias indices were not included in the manuscript, instead the tables are presented in this appendix in order to provide a comprehensive report in Chapter 6.

D.2 Results

D.2.1 Analysis of Guanine-Cytosine contents

GC content

Table D.1: GC contents of selected subset of ribosomal and repair genes

Genomes	*dna*E	*gid*A	*gua*A	*gyr*A	*gyr*B	*lig*A	*mfd*	*mut*S	*mut*Y	*ogt*	*pol*A	*pyr*C	*pyr*G	*rec*A	*rec*J	*rec*N	*rec*R	*rpl*B	*rpo*B	*rpo*D	*ruv*A	*ssb*	*uvr*A	*uvr*B	*xse*A
CJ11168	**0.322**	0.338	0.347	0.334	0.332	0.344	**0.287**	0.325	**0.318**	0.274	0.289	0.306	0.344	**0.362**	**0.298**	0.274	**0.307**	**0.392**	**0.334**	0.308	0.339	0.344	**0.343**	0.334	**0.284**
CJ26094	0.321	0.337	0.356	**0.333**	**0.33**	0.342	0.29	0.324	0.311	**0.272**	**0.294**	0.308	0.344	0.368	**0.298**	0.278	0.309	0.389	0.337	0.309	**0.337**	0.341	0.345	0.338	0.287
CJ81116	0.324	0.332	0.361	0.335	0.332	0.34	0.29	0.324	0.312	0.265	**0.294**	0.304	0.347	0.364	0.294	0.278	**0.307**	0.387	0.337	0.304	0.339	0.333	0.345	0.337	0.288
CJ81176	**0.322**	0.338	0.356	0.334	**0.33**	0.342	0.29	0.327	0.309	0.274	0.289	0.304	0.343	**0.362**	0.297	**0.276**	**0.307**	0.394	0.337	**0.307**	0.335	0.339	0.345	0.338	0.281
CJ8421	0.324	0.339	0.351	-	0.329	0.341	0.288	0.325	0.313	0.274	0.293	0.309	0.346	0.359	**0.298**	0.274	0.309	0.39	**0.334**	0.308	0.335	0.342	**0.343**	0.335	0.281
CJ8425	**0.322**	**0.333**	**0.348**	**0.333**	0.329	0.344	**0.287**	**0.326**	0.317	0.274	0.289	0.306	0.346	**0.362**	**0.298**	0.272	**0.307**	**0.392**	**0.334**	0.308	0.339	0.337	**0.343**	0.341	**0.284**
CJ8486	0.321	0.338	0.35	0.334	0.331	0.341	-	0.328	0.313	0.269	0.289	-	0.35	0.382	0.302	-	0.309	0.39	0.333	**0.307**	-	0.341	0.342	0.335	0.281
CJ9313	**0.322**	0.337	0.355	0.334	0.332	0.346	0.29	0.33	0.308	**0.272**	**0.294**	0.306	**0.345**	**0.362**	0.295	0.277	**0.307**	0.39	0.336	**0.307**	0.341	0.333	0.348	0.338	0.287
CJ936	**0.322**	0.338	**0.348**	0.334	0.332	0.344	**0.287**	0.327	**0.318**	0.274	0.289	0.306	0.344	**0.362**	**0.298**	**0.276**	**0.307**	**0.392**	**0.334**	0.308	0.339	0.341	**0.343**	0.335	**0.284**
CJD	0.319	0.33	**0.348**	**0.333**	**0.33**	0.335	0.289	**0.326**	0.317	0.283	0.291	0.312	0.343	**0.362**	0.297	**0.276**	0.3	0.384	-	0.304	0.33	0.339	0.345	0.334	**0.284**
CJIA3902	0.323	0.341	**0.348**	0.334	**0.33**	0.344	**0.287**	0.324	**0.318**	0.274	0.289	**0.307**	0.35	0.368	0.297	0.272	**0.307**	**0.392**	**0.334**	0.308	0.339	0.344	**0.343**	**0.333**	**0.284**
CJRM1221	**0.322**	0.334	0.353	0.334	**0.33**	0.342	**0.287**	0.327	**0.318**	0.267	0.289	0.306	0.346	0.366	**0.298**	**0.276**	**0.307**	**0.392**	**0.334**	0.308	0.335	0.341	**0.343**	0.335	**0.284**
H22082	**0.322**	**0.333**	**0.348**	**0.333**	**0.33**	**0.343**	**0.287**	**0.326**	**0.318**	**0.272**	**0.294**	**0.307**	**0.345**	**0.362**	**0.298**	**0.276**	**0.307**	**0.392**	**0.334**	**0.307**	**0.337**	**0.337**	**0.343**	**0.333**	**0.284**
H704	**0.322**	**0.333**	**0.348**	**0.333**	**0.33**	**0.343**	**0.287**	**0.326**	**0.318**	**0.272**	**0.294**	**0.307**	**0.345**	**0.362**	**0.302**	**0.276**	**0.307**	**0.392**	**0.334**	**0.307**	**0.337**	**0.346**	**0.343**	**0.333**	**0.284**
P110	**0.322**	**0.333**	**0.348**	**0.333**	**0.33**	**0.343**	**0.287**	**0.326**	**0.318**	**0.286**	**0.294**	**0.307**	**0.345**	**0.362**	**0.298**	**0.276**	**0.307**	**0.392**	**0.334**	**0.307**	**0.337**	**0.346**	**0.343**	**0.333**	**0.285**
P179a	**0.322**	**0.333**	**0.348**	**0.333**	**0.33**	**0.343**	**0.287**	**0.326**	**0.318**	**0.286**	**0.294**	**0.307**	**0.345**	**0.362**	**0.298**	**0.276**	**0.307**	**0.392**	**0.334**	**0.307**	**0.337**	**0.346**	**0.343**	**0.333**	**0.285**
P569a	**0.322**	**0.333**	**0.348**	**0.333**	**0.33**	**0.343**	**0.287**	**0.326**	**0.318**	**0.272**	**0.294**	**0.307**	**0.345**	**0.362**	**0.303**	**0.276**	**0.307**	**0.392**	**0.334**	**0.307**	**0.337**	**0.346**	**0.343**	**0.333**	**0.285**
P694a	**0.322**	**0.329**	**0.348**	**0.333**	**0.33**	**0.343**	**0.287**	**0.326**	**0.318**	*0.289*	**0.294**	**0.307**	**0.345**	**0.362**	**0.298**	**0.276**	**0.307**	**0.392**	**0.334**	**0.307**	**0.337**	**0.337**	**0.343**	**0.333**	**0.284**
P73020	**0.322**	**0.333**	**0.348**	**0.333**	**0.33**	**0.343**	**0.287**	**0.326**	**0.318**	**0.286**	**0.294**	**0.307**	**0.345**	**0.362**	**0.298**	**0.276**	**0.307**	**0.392**	**0.334**	**0.307**	**0.337**	**0.346**	**0.343**	**0.333**	**0.284**

GC3 content

Table D.2: GC3 contents of selected subset of ribosomal and repair genes

Genomes	*dna*E	*gid*A	*gua*A	*gyr*A	*gyr*B	*lig*A	*mfd*	*mut*S	*mut*Y	*ogt*	*pol*A	*pyr*C	*pyr*G	*rec*A	*rec*J	*rec*N	*rec*R	*rpl*B	*rpo*B	*rpo*D	*ruv*A	*ssb*	*uvr*A	*uvr*B	*xse*A
CJ11168	**0.175**	0.182	0.173	0.169	0.164	0.232	**0.156**	0.247	**0.215**	0.116	0.154	**0.158**	**0.188**	**0.153**	0.154	0.165	0.23	**0.169**	0.165	0.163	**0.218**	**0.175**	0.206	**0.195**	**0.133**
CJ26094	0.171	0.182	0.185	**0.167**	**0.16**	0.229	0.164	0.25	0.196	0.116	0.155	0.161	0.186	0.168	**0.152**	0.177	0.23	0.161	0.171	0.165	0.212	0.171	0.21	0.208	0.138
CJ81116	0.177	0.164	0.2	0.172	0.157	0.227	0.165	0.258	0.199	0.102	0.155	**0.158**	0.19	0.156	0.144	0.185	0.214	0.154	0.172	0.153	0.212	0.164	0.211	0.209	0.138
CJ81176	0.172	0.182	0.192	0.17	0.165	0.235	0.164	0.249	0.184	0.116	0.154	**0.158**	0.182	**0.153**	0.15	0.175	**0.225**	0.176	0.171	0.158	0.201	0.166	0.212	0.206	0.125
CJ8421	0.181	0.185	0.179	-	**0.16**	0.226	0.158	0.248	0.202	0.116	**0.161**	0.161	0.19	0.144	0.156	0.167	0.23	0.165	0.163	0.161	0.201	0.171	**0.208**	0.197	0.125
CJ8425	**0.175**	0.174	**0.175**	0.166	0.157	0.232	**0.156**	**0.255**	**0.215**	0.116	0.154	**0.158**	0.186	**0.153**	0.154	0.163	0.23	**0.169**	0.165	0.163	**0.218**	**0.175**	0.207	0.21	**0.133**
CJ8486	0.172	0.187	0.177	0.17	0.164	0.224		0.262	0.202	**0.109**	0.153		0.19	0.156	**0.152**		0.23	**0.169**	0.162	0.161		0.169	0.206	0.198	0.125
CJ9313	0.172	0.182	0.185	0.168	0.155	0.237	0.165	0.261	0.184	0.116	0.155	**0.158**	0.197	0.168	0.146	0.177	0.214	0.161	0.169	0.158	**0.218**	0.164	0.219	0.209	0.138
CJ936	**0.175**	0.183	**0.175**	0.169	**0.16**	0.232	**0.156**	0.262	**0.215**	0.116	0.154	**0.158**	0.184	**0.153**	**0.152**	0.17	0.23	**0.169**	**0.164**	0.163	**0.218**	0.169	0.207	0.198	**0.133**
CJD	0.171	0.166	0.171	0.168	0.159	0.211	0.17	0.258	0.208	*0.135*	0.155	0.171	0.197	0.221	0.156	0.179	0.193	0.15		0.158	0.19	0.186	0.216	0.193	0.141
CJIA3902	0.177	0.193	**0.175**	0.169	0.163	0.232	**0.156**	0.253	**0.215**	0.116	0.154	**0.158**	**0.188**	**0.153**	0.15	0.165	**0.225**	**0.169**	0.165	0.163	**0.218**	**0.175**	0.207	0.19	**0.133**
CJRM1221	0.176	0.174	0.185	0.169	0.164	0.229	**0.156**	0.262	**0.215**	0.102	0.154	**0.158**	0.19	0.162	**0.152**	0.17	0.23	**0.169**	**0.164**	0.163	0.201	0.169	0.207	0.198	**0.133**
H22082	**0.175**	**0.173**	**0.175**	**0.167**	**0.16**	**0.23**	**0.156**	**0.255**	**0.215**	**0.109**	**0.161**	**0.158**	**0.188**	**0.153**	**0.152**	**0.171**	**0.225**	**0.169**	**0.164**	0.161	**0.218**	**0.175**	**0.208**	**0.195**	**0.133**
H704	**0.175**	**0.173**	**0.175**	**0.167**	**0.16**	**0.23**	**0.156**	**0.255**	**0.215**	**0.109**	**0.161**	**0.158**	**0.188**	**0.153**	**0.154**	**0.171**	**0.225**	**0.169**	**0.164**	0.161	**0.218**	**0.181**	**0.208**	**0.195**	**0.133**
P110	**0.175**	**0.173**	**0.175**	**0.167**	**0.16**	**0.23**	**0.156**	**0.255**	**0.215**	**0.126**	**0.161**	**0.158**	**0.188**	**0.153**	**0.152**	**0.171**	**0.225**	**0.169**	**0.164**	0.161	**0.218**	**0.181**	**0.208**	**0.195**	**0.135**
P179a	**0.175**	**0.173**	**0.175**	**0.167**	**0.16**	**0.23**	**0.156**	**0.255**	**0.215**	**0.126**	**0.161**	**0.158**	**0.188**	**0.153**	**0.152**	**0.171**	**0.225**	**0.169**	**0.164**	0.161	**0.218**	**0.181**	**0.208**	**0.195**	**0.135**
P569a	**0.175**	**0.173**	**0.175**	**0.167**	**0.16**	**0.23**	**0.156**	**0.255**	**0.215**	**0.109**	**0.161**	**0.158**	**0.188**	**0.153**	**0.154**	**0.171**	**0.225**	**0.169**	**0.164**	0.161	**0.218**	**0.181**	**0.208**	**0.195**	**0.135**
P694a	**0.175**	**0.16**	**0.175**	**0.167**	**0.16**	**0.23**	**0.156**	**0.255**	**0.215**	*0.135*	**0.161**	**0.158**	**0.188**	**0.153**	**0.152**	**0.171**	**0.225**	**0.169**	**0.164**	0.161	**0.218**	**0.175**	**0.208**	**0.195**	**0.133**
P73020	**0.175**	**0.173**	**0.175**	**0.167**	**0.16**	**0.23**	**0.156**	**0.255**	**0.215**	**0.126**	**0.161**	**0.158**	**0.188**	**0.153**	**0.152**	**0.171**	**0.225**	**0.169**	**0.164**	0.161	**0.218**	**0.181**	**0.208**	0.193	**0.133**

D.2.2 Codon usage bias

Codon usage bias

Table D.3: Codon usage bias indices of ribosomal and repair genes

Genes	*dna*E	*gid*A	*gua*A	*gyr*A	*gyr*B	*lig*A	*mfd*	*mut*S	*mut*Y	*ogt*	*pol*A	*pyr*C	*pyr*G	*rec*A	*rec*J	*rec*N	*rec*R	*rpl*B	*rpo*B	*rpo*D	*ruv*A	*ssb*	*uvr*A	*uvr*B	*xse*A
CJ11168	**0.591**	0.554	0.589	0.589	0.6	0.503	0.649	0.484	**0.549**	0.731	0.626	0.651	0.589	**0.592**	0.634	0.635	0.508	**0.631**	0.594	0.621	**0.604**	0.667	0.527	**0.585**	**0.687**
CJ26094	0.601	0.556	0.578	0.59	0.597	0.503	0.642	0.48	0.594	0.74	0.629	0.646	0.588	0.57	0.639	0.628	**0.505**	0.643	0.587	0.621	0.587	0.662	0.543	0.56	0.673
CJ81116	**0.591**	0.586	0.574	0.586	0.595	0.508	0.637	**0.479**	0.584	0.726	0.63	0.66	0.582	0.572	0.646	0.603	0.536	0.65	0.586	0.623	0.58	0.667	0.524	0.569	0.675
CJ81176	0.6	0.554	0.575	0.591	0.593	0.49	0.642	0.488	0.606	0.734	0.627	0.66	0.593	0.585	0.642	0.629	**0.505**	0.622	0.586	0.628	0.584	0.656	**0.526**	0.556	0.695
CJ8421	0.585	0.553	0.587	-	0.609	0.506	0.646	0.481	0.555	0.74	0.626	0.663	0.585	0.598	0.631	0.635	**0.505**	0.632	0.597	0.608	0.584	0.66	0.527	0.571	0.695
CJ8425	**0.591**	0.57	0.586	0.594	0.6	0.503	0.649	**0.479**	0.553	0.731	0.626	0.651	0.587	**0.592**	0.635	0.639	0.508	**0.631**	0.593	0.621	**0.604**	0.641	**0.526**	0.56	**0.687**
CJ8486	0.6	0.55	0.587	0.588	0.605	0.517	-	0.474	0.572	0.739	0.629	-	0.585	0.587	0.635	-	**0.505**	0.628	0.6	**0.625**	-	0.664	0.53	0.566	0.695
CJ9313	0.6	0.554	0.591	0.587	0.602	0.499	0.64	0.475	0.605	0.74	0.628	0.661	0.592	0.57	0.644	0.627	0.536	0.643	0.587	0.628	0.597	0.667	0.53	0.557	0.673
CJ936	**0.591**	0.555	0.586	0.587	0.6	0.503	0.647	0.473	**0.549**	0.731	0.626	0.651	0.575	0.583	0.635	0.633	0.508	**0.631**	**0.595**	0.621	**0.604**	0.664	**0.526**	0.578	**0.687**
CJD	0.6	0.576	0.604	0.589	0.606	0.536	0.632	0.478	0.583	0.697	0.633	0.666	0.574	0.517	0.629	0.616	0.561	0.661	-	0.628	0.585	0.633	0.516	0.589	0.669
CJIA3902	0.588	0.54	0.586	0.589	0.6	0.503	0.654	0.473	**0.549**	0.731	0.626	**0.657**	**0.586**	**0.592**	0.637	0.639	0.516	**0.631**	0.594	0.621	**0.604**	0.667	**0.526**	0.58	**0.687**
CJRM1221	0.593	0.572	0.586	0.587	0.593	0.505	0.647	0.473	**0.549**	**0.747**	0.626	0.651	0.585	0.587	0.635	0.633	0.508	**0.631**	**0.595**	0.621	0.584	0.664	**0.526**	0.578	**0.687**
H22082	**0.591**	**0.568**	**0.582**	**0.595**	**0.606**	**0.504**	**0.653**	**0.479**	**0.549**	0.747	**0.622**	**0.657**	**0.586**	**0.592**	**0.636**	**0.629**	**0.505**	**0.631**	**0.595**	**0.625**	**0.604**	0.648	**0.526**	**0.585**	**0.687**
H704	**0.591**	**0.568**	**0.582**	**0.595**	**0.606**	**0.504**	**0.653**	**0.479**	**0.549**	**0.747**	**0.622**	**0.657**	**0.586**	**0.592**	**0.636**	**0.629**	**0.505**	**0.631**	**0.595**	**0.625**	**0.604**	**0.666**	**0.526**	**0.585**	**0.687**
P110b	**0.591**	**0.568**	**0.582**	**0.595**	**0.606**	**0.504**	**0.653**	**0.479**	**0.549**	0.735	**0.622**	**0.657**	**0.586**	**0.592**	**0.636**	**0.629**	**0.505**	**0.631**	**0.595**	**0.625**	**0.604**	**0.666**	**0.526**	**0.585**	**0.685**
P179a	**0.591**	**0.568**	**0.582**	**0.595**	**0.606**	**0.504**	**0.653**	**0.479**	**0.549**	0.735	**0.622**	**0.657**	**0.586**	**0.592**	**0.636**	**0.629**	**0.505**	**0.631**	**0.595**	**0.625**	**0.604**	**0.666**	**0.526**	**0.585**	**0.685**
P569a	**0.591**	**0.568**	**0.582**	**0.595**	**0.606**	**0.504**	**0.653**	**0.479**	**0.549**	**0.747**	**0.622**	**0.657**	**0.586**	**0.592**	**0.635**	**0.629**	**0.505**	**0.631**	**0.595**	**0.625**	**0.604**	**0.666**	**0.526**	**0.585**	**0.685**
P694a	**0.591**	0.582	**0.582**	**0.595**	**0.606**	**0.504**	**0.653**	**0.479**	**0.549**	0.716	**0.622**	**0.657**	**0.586**	**0.592**	**0.636**	0.631	**0.505**	**0.631**	**0.595**	**0.625**	**0.604**	0.648	**0.526**	**0.585**	**0.687**
P73020	**0.591**	**0.568**	**0.582**	**0.595**	**0.606**	**0.504**	**0.653**	**0.479**	**0.549**	0.735	**0.622**	**0.657**	**0.586**	**0.592**	**0.636**	**0.629**	**0.505**	**0.631**	**0.595**	**0.625**	**0.604**	**0.666**	**0.526**	**0.586**	**0.687**

Scaled chi square codon usage bias

Table D.4: Schaled chi square codon usage bias indices of ribosomal and repair genes

Genes	*dna*E	*gid*A	*gua*A	*gyr*A	*gyr*B	*lig*A	*mfd*	*mut*S	*mut*Y	*ogt*	*pol*A	*pyr*C	*pyr*G	*rec*A	*rec*J	*rec*N	*rec*R	*rpl*B	*rpo*B	*rpo*D	*ruv*A	*ssb*	*uvr*A	*uvr*B	*xse*A
CJ11168	**0.592**	0.559	0.668	0.603	0.596	0.523	0.655	0.439	**0.554**	0.997	0.641	0.738	0.607	**0.652**	0.634	0.663	0.496	**0.76**	0.597	0.612	0.636	0.744	0.496	**0.571**	**0.751**
CJ26094	0.602	0.557	0.64	0.607	0.604	**0.529**	0.639	0.427	0.622	0.991	0.657	0.737	0.608	0.636	0.639	0.657	**0.488**	0.805	0.586	0.631	0.617	0.744	0.518	0.544	0.718
CJ81116	0.578	0.614	0.633	0.593	0.601	0.518	0.635	0.44	0.625	0.969	0.658	0.75	0.603	0.644	0.661	0.582	0.549	0.832	0.585	0.631	0.636	0.752	0.489	0.551	0.718
CJ81176	0.601	0.56	0.645	0.61	0.601	0.512	0.639	**0.434**	0.642	1.027	0.641	0.75	0.612	0.651	0.636	0.66	**0.488**	0.779	0.584	0.639	0.619	0.74	0.495	0.523	0.772
CJ8421	0.576	0.555	0.666	*	0.619	0.527	0.651	0.429	0.59	0.992	0.624	0.751	0.6	0.662	0.632	0.649	0.486	0.752	0.601	0.6	0.619	0.743	0.496	0.543	0.77
CJ8425	**0.592**	**0.585**	0.66	**0.611**	0.596	0.523	0.655	**0.434**	0.566	0.997	0.641	0.738	0.606	**0.652**	0.634	0.667	0.496	**0.76**	0.596	0.612	0.636	0.71	0.492	0.536	**0.751**
CJ8486	0.593	0.55	0.661	0.606	0.613	0.55	*	0.423	0.607	0.998	0.645	*	0.6	0.641	0.652	*	0.486	0.777	0.604	**0.617**	*	0.747	0.503	0.543	0.772
CJ9313	0.601	0.56	0.659	0.603	0.616	0.522	0.636	0.422	0.646	0.991	0.656	0.752	0.588	0.636	0.661	0.649	0.549	0.807	0.586	0.639	0.632	0.752	0.497	0.538	0.718
CJ936	**0.592**	0.551	0.66	0.601	0.607	0.523	0.654	0.419	**0.554**	0.997	0.641	0.738	0.613	0.646	0.637	0.639	0.496	**0.76**	**0.598**	0.612	0.636	0.747	0.492	0.562	**0.751**
CJD	0.594	0.618	0.747	0.61	0.603	0.58	0.638	0.422	0.598	0.846	0.664	0.692	0.556	0.541	0.631	0.64	0.606	0.842	*	0.633	0.607	0.669	0.481	0.564	0.731
CJIA39	0.581	0.527	0.66	0.603	0.596	0.523	0.656	0.433	**0.554**	0.997	0.641	**0.739**	**0.601**	**0.652**	**0.644**	0.659	0.498	**0.76**	0.596	0.612	0.636	0.744	0.492	0.552	**0.751**
CJRM1221	0.584	0.578	0.635	0.601	0.599	**0.529**	0.654	0.419	**0.554**	1.014	0.641	0.738	0.6	0.638	0.637	0.639	0.496	**0.76**	**0.598**	0.612	0.619	0.747	0.492	0.562	**0.751**
H22082	**0.592**	**0.584**	**0.653**	**0.617**	**0.611**	**0.529**	**0.663**	**0.434**	**0.554**	**1.015**	**0.616**	**0.739**	**0.601**	**0.652**	**0.641**	**0.64**	**0.488**	**0.76**	**0.598**	**0.617**	**0.637**	**0.718**	**0.488**	**0.571**	**0.751**
H704	**0.592**	**0.584**	**0.653**	**0.617**	**0.611**	**0.529**	**0.663**	**0.434**	**0.554**	**1.015**	**0.616**	**0.739**	**0.601**	**0.652**	**0.644**	**0.64**	**0.488**	**0.76**	**0.598**	**0.617**	**0.637**	**0.737**	**0.488**	**0.571**	**0.751**
P110	**0.592**	**0.584**	**0.653**	**0.617**	**0.611**	**0.529**	**0.663**	**0.434**	**0.554**	**1.033**	**0.616**	**0.739**	**0.601**	**0.652**	**0.641**	**0.64**	**0.488**	**0.76**	**0.598**	**0.617**	**0.637**	**0.737**	**0.488**	**0.571**	**0.743**
P179a	**0.592**	**0.584**	**0.653**	**0.617**	**0.611**	**0.529**	**0.663**	**0.434**	**0.554**	**1.033**	**0.616**	**0.739**	**0.601**	**0.652**	**0.641**	**0.64**	**0.488**	**0.76**	**0.598**	**0.617**	**0.637**	**0.737**	**0.488**	**0.571**	**0.743**
P569a	**0.592**	**0.584**	**0.653**	**0.617**	**0.611**	**0.529**	**0.663**	**0.434**	**0.554**	**1.015**	**0.616**	**0.739**	**0.601**	**0.652**	**0.641**	**0.64**	**0.488**	**0.76**	**0.598**	**0.617**	**0.637**	**0.737**	**0.488**	**0.571**	**0.743**
P694a	**0.592**	**0.624**	**0.653**	**0.617**	**0.611**	**0.529**	**0.663**	**0.434**	**0.554**	**1.003**	**0.616**	**0.739**	**0.601**	**0.652**	**0.641**	**0.64**	**0.488**	**0.76**	**0.598**	**0.617**	**0.637**	**0.718**	**0.488**	**0.571**	**0.751**
P73020	**0.592**	**0.584**	**0.653**	**0.617**	**0.611**	**0.529**	**0.663**	**0.434**	**0.554**	**1.033**	**0.616**	**0.739**	**0.601**	**0.652**	**0.641**	**0.64**	**0.488**	**0.76**	**0.598**	**0.617**	**0.637**	**0.737**	**0.488**	**0.576**	**0.751**

D.2.3 Selection pressure, evolutionary rate and recombination:

Selection pressure – Nonsynonymous

Table D.5: Polymorphic synonymous, nonsynonymous sites and the Tajima D values of the ribosomal and repair

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*dna*E	6, 18, 21, 45, 63, 111, 141, 153, 210, 219, 225, 228, 246, 258, 285, 333, 336, 348, 381, 396, 420, 501, 502, 567, 618, 624, 636, 648, 654, 696, 702, 714, 741, 762, 771, 828, 921, 981, 987, 1002, 1014, 1023, 1026, 1029, 1038, 1047, 1065, 1089, 1131, 1152, 1161, 1167, 1269, 1323, 1347, 1353, 1449, 1467, 1518, 1530, 1578, 1585, 1599, 1608, 1614, 1647, 1656, 1686, 1692, 1704, 1764, 1776, 1785, 1785, 1788, 1851, 1854, 1887, 1890, 1923, 1944, 1953, 1962, 1965, 1977, 1992, 2001, 2028, 2034, 2037, 2049, 2067, 2088, 2118, 2133, 2142, 2151, 2163, 2175, 2181, 2208, 2217, 2226, 2259, 2280, 2298, 2340, 2349, 2361, 2379, 2388, 2409, 2421, 2436, 2511, 2544, 2565, 2571, 2574, 2589, 2613, 2622, 2658, 2661, 2673, 2676, 2679, 2688, 2742, 2778, 2793, 2835, 2844, 2850, 2874, 2880, 2910, 2958, 2973, 2985, 3006, 3027, 3066, 3072, 3156, 3162, 3177, 3183, 3198, 3210, 3240, 3249, 3255, 3282, 3285, 3294, 3303, 3330, 3345, 3372, 3384, 3409, 3423, 3426, 3474, 3492, 3522, 3535, 3549, 3579 (170)	91, 143, 198, 376, 380, 729, 788, 823, 846, 917, 963, 994, 1550, 2048, 2270, 2292, 2433, 2593, 2596, 2674, 2702, 2766, 2777, 2796, 2873, 2912, 3009, 3016, 3245, 3280, 3307, 3331, 3505, 3523 (34)	-1.37094	1.13669

Table D.5 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*gid*A	81, 84, 102, 117, 126, 132, 135, 141, 144, 150, 156, 198, 231, 243, 249, 258, 276, 297, 309, 312, 345, 372, 399, 417, 438, 441, 444, 471, 477, 489, 507, 525, 540, 544, 546, 549, 561, 570, 591, 594, 600, 624, 627, 645, 669, 687, 714, 720, 768, 771, 780, 783, 783, 798, 831, 837, 858, 876, 897, 909, 912, 930, 939, 942, 955, 960, 1020, 1107, 1143, 1158, 1164, 1179, 1233, 1245, 1278, 1281, 1293, 1347, 1389, 1437, 1440, 1443, 1446, 1455, 1476, 1482, 1521, 1563, 1578, 1590, 1593, 1602, 1611, 1632, 1635, 1638, 1677, 1719, 1746, 1764, 1800, 1809, 1815, 1824 (104)	190, 275, 307, 331, 359, 365, 385, 419, 608, 680, 832, 947, 959, 979, 1127, 1343, 1343, 372, 1507, 1569, 1577, 1585, 1625, 1756, 1842, 1856 (26)	-0.47209	1.14795
*gua*A	27, 36, 81, 111, 117, 120, 123, 129, 132, 159, 162, 165, 174, 192, 210, 222, 226, 249, 258, 261, 282, 288, 315, 324, 330, 333, 342, 345, 351, 372, 378, 402, 453, 471, 474, 489, 495, 496, 498, 516, 525, 564, 570, 576, 582, 597, 627, 630, 654, 660, 663, 672, 675, 678, 687, 711, 723, 747, 765, 801, 804, 810, 819, 825, 847, 858, 867, 879, 960, 963, 1008, 1080, 1098, 1116, 1117, 1122, 1128, 1140, 1149, 1158, 1173, 1179, 1191, 1212, 1230, 1236, 1254, 1257, 1308, 1332, 1341, 1347, 1371, 1377, 1392, 1398, 1410, 1416, 1428, 1443, 1482, 1500, 1521 (103)	118, 223, 224, 232, 331, 344, 415, 418, 420, 583, 644, 727, 766, 860, 869, 916, 1009, 1062, 1213, 1216, 1238, 1318, 1357, 1432 (24)	-0.95402	0.10576

Table D.5 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
gyrA	21, 72, 117, 243, 246, 276, 330, 351, 357, 360, 471, 483, 552, 622, 693, 717, 726, 738, 786, 891, 909, 957, 975, 1008, 1017, 1045, 1047, 1071, 1191, 1200, 1206, 1209, 1233, 1257, 1266, 1278, 1290, 1290, 1323, 1389, 1434, 1446, 1476, 1530, 1533, 1533, 1578, 1590, 1668, 1770, 1824, 1830, 1908, 1929, 2064, 2073, 2130, 2142, 2154, 2169, 2175, 2283, 2304, 2310, 2334, 2340, 2355, 2370, 2400, 2421, 2457, 2472, 2478, 2484, 2487, 2520, 2547 (77)	64, 257, 608, 616, 617, 781, 854, 1127, 1906, 1930, 1994, 2069, 2135, 2254, 2516, 2587 (16)	-0.50676	1.88165
gyrB	37, 46, 93, 99, 111, 165, 291, 324, 345, 360, 375, 423, 477, 489, 531, 555, 570, 597, 621, 645, 720, 738, 789, 807, 813, 852, 906, 921, 924, 942, 945, 969, 1023, 1047, 1053, 1107, 1113, 1125, 1126, 1128, 1131, 1143, 1167, 1185, 1209, 1296, 1308, 1318, 1360, 1362, 1404, 1413, 1449, 1497, 1497, 1527, 1551, 1570, 1581, 1626, 1650, 1662, 1728, 1743, 1902, 1986, 1998, 2154, 2169, 2250, 2277, 2295 (72)	409, 410, 442, 466, 674, 927, 952, 1084, 1112, 1552, 1823, 2000, 2029, 2276 (14)	-0.4225	-0.42175

Table D.5 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
ligA	39, 39, 54, 87, 111, 141, 171, 183, 189, 198, 249, 270, 345, 375, 486, 489, 513, 519, 549, 558, 579, 606, 612, 621, 669, 672, 675, 678, 681, 699, 708, 720, 732, 735, 741, 753, 756, 762, 787, 801, 819, 825, 859, 864, 912, 939, 939, 954, 954, 966, 984, 993, 1017, 1032, 1035, 1050, 1104, 1113, 1119, 1173, 1179, 1245, 1311, 1314, 1332, 1344, 1362, 1371, 1407, 1419, 1431, 1446, 1449, 1455, 1461, 1464, 1467, 1470, 1473, 1491, 1500, 1539, 1557, 1587, 1605, 1629, 1650, 1665, 1686, 1707, 1713, 1719, 1743, 1752, 1764, 1788, 1803, 1809, 1815, 1818, 1821, 1821, 1872, 1875, 1878, 1881 (106)	104, 123, 125, 151, 188, 224, 253, 259, 265, 368, 463, 609, 646, 684, 697, 705, 730, 734, 742, 743, 784, 785, 832, 940, 968, 1140, 1172, 1391, 1561, 1586, 1649, 1651, 1652, 1705, 1717, 1720, 1765, 1797, 1816, 1877, 1897, 1909 (42)	-1.28959	1.5028
mfd	25, 60, 87, 111, 114, 141, 144, 174, 195, 276, 279, 291, 306, 366, 444, 465, 489, 531, 537, 603, 624, 639, 645, 657, 657, 666, 669, 690, 693, 729, 735, 750, 798, 852, 867, 876, 888, 936, 963, 966, 1053, 1065, 1068, 1071, 1074, 1086, 1140, 1152, 1164, 1185, 1203, 1224, 1230, 1234, 1275, 1290, 1317, 1332, 1350, 1371, 1392, 1410, 1425, 1428, 1464, 1506, 1512, 1515, 1530, 1536, 1554, 1569, 1575, 1638, 1641, 1647, 1650, 1689, 1734, 1737, 1767, 1779, 1785, 1833, 1923, 1959, 1965, 2022, 2058, 2067, 2070, 2076, 2080, 2118, 2139, 2142, 2160, 2172, 2199, 2226, 2229, 2244, 2271, 2277, 2283, 2286, 2304, 2325, 2343, 2358, 2385, 2391, 2415, 2478, 2496, 2514, 2622, 2637, 2647, 2709, 2811, 2817, 2820, 2874, 2889, 2916 (126)	54, 214, 301, 325, 335, 541, 628, 682, 694, 745, 883, 913, 919, 1300, 1364, 1372, 1384, 1399, 1419, 1454, 1531, 1604, 1631, 1636, 1928, 1982, 2140, 2425, 2628, 2713, 2746, 2777 (32)	-0.42339	0.88156

Table D.5 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*mut*S	12, 48, 63, 72, 90, 93, 102, 114, 117, 120, 132, 147, 180, 198, 201, 201, 210, 225, 264, 277, 279, 291, 292, 294, 316, 318, 339, 342, 351, 366, 372, 378, 381, 393, 423, 427, 435, 441, 447, 474, 489, 507, 555, 555, 570, 591, 636, 729, 759, 774, 789, 795, 828, 831, 849, 852, 867, 876, 894, 948, 976, 981, 1020, 1020, 1026, 1041, 1062, 1065, 1068, 1095, 1107, 1128, 1134, 1137, 1167, 1188, 1227, 1239, 1245, 1266, 1269, 1272, 1272, 1290, 1305, 1326, 1332, 1338, 1359, 1365, 1374, 1380, 1383, 1410, 1410, 1446, 1488, 1491, 1494, 1497, 1548, 1587, 1626, 1635, 1680, 1701, 1710, 1722, 1740, 1779, 1818, 1869, 1900, 1914, 1920, 2080, 2082, 2205 (118)	86, 113, 115, 156, 168, 259, 286, 308, 335, 404, 412, 421, 446, 455, 475, 527, 599, 796, 809, 823, 1015, 1028, 1045,, 1148, 1243, 1283, 1289, 1301, 1339, 1355, 1470, 1495, 1596, 1628, 1631, 1656, 1657, 1703, 1704, 1732, 1860, 1918, 1931, 1948, 1951, 1954, 1955, 1960, 1960, 1966, 1966, 1968, 1969, 1975, 1978, 1981, 1981, 1983, 1984, 1996, 1999, 2014, 2021, 2023, 2026, 2054, 2056, 2062, 2068, 2071, 2077, 2078, 2083, 2086, 2089, 2095, 2096, 2107, 2116, 2131, 2140, 2152, 2164, 2185, 2203, 2207, 2211 (87)	-0.80472	2.51068
*mut*Y	27, 75, 141, 192, 198, 240, 261, 264, 285, 294, 297, 324, 327, 336, 351, 363, 375, 378, 379, 393, 396, 417, 417, 435, 438, 465, 477, 480, 501, 564, 573, 582, 609, 633, 759, 771, 789, 825, 916, 921, 1017 (41)	52, 217, 321, 350, 369, 376, 404, 532, 553, 592, 613, 631, 670, 679, 709, 715, 763, 764, 766, 779, 847, 871, 872, 946, 983, 996 (26)	-0.94926	0.94322
ogt	124, 126, 156, 180, 192, 201, 222, 249, 261, 276, 285, 294, 294, 312, 336, 375, 381 (17)	128, 167, 178, 190, 199, 235, 255, 257, 277, 289, 374 (11)	-1.72611	1.36041

Table D.5 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*pol*A	21, 51, 54, 57, 66, 114, 129, 138, 144, 150, 169, 171, 174, 177, 180, 198, 225, 237, 240, 270, 276, 297, 310, 312, 327, 339, 369, 384, 429, 435, 480, 498, 537, 540, 546, 588, 600, 609, 618, 624, 648, 654, 657, 672, 675, 684, 708, 720, 726, 729, 744, 753, 765, 822, 990, 1041, 1047, 1050, 1065, 1069, 1071, 1074, 1077, 1080, 1083, 1089, 1092, 1095, 1101, 1104, 1119, 1146, 1155, 1173, 1191, 1230, 1278, 1296, 1305, 1317, 1347, 1353, 1368, 1374, 1377, 1377, 1380, 1395, 1443, 1444, 1464, 1509, 1512, 1515, 1578, 1599, 1611, 1614, 1620, 1635, 1638, 1659, 1665, 1671, 1680, 1683, 1743, 1749, 1767, 1779, 1815, 1818, 1818, 1845, 1860, 1869, 1872, 1875, 1878, 1881, 1893, 1905, 1911, 1914, 1917, 1923, 1932, 2010, 2019, 2022, 2055, 2061, 2064, 2079, 2091, 2106, 2166, 2172, 2181, 2256, 2262, 2268, 2286, 2304, 2310, 2349, 2352, 2406, 2418, 2421, 2433, 2443, 2454, 2457, 2472, 2487, 2496, 2523, 2529, 2544, 2583, 2604 (162)	19, 100, 127, 142, 160, 212, 274, 311, 353, 364, 368, 476, 508, 545, 581, 649, 732, 748, 898, 1028, 1078, 1081, 1096, 1190, 1253, 1396, 1414, 1436, 1519, 1571, 1633, 1717, 1720, 1721, 1742, 1747, 1982, 2016, 2059, 2209, 2276, 2287, 2311, 2312, 2425 (45)	-0.2055	0.56685
*pyr*C	21, 48, 126, 252, 270, 300, 312, 348, 351, 385, 396, 402, 474, 492, 549, 552, 561, 576, 579, 580, 582, 594, 627, 645, 684, 693, 708, 714, 723, 732, 741, 747, 756, 756, 759, 762, 765, 771, 774, 777, 792, 795, 819, 822, 822, 846, 847, 848, 849, 855, 861, 867, 874, 876, 891, 900, 927, 960, 975, 987 (60)	5, 17, 80, 169, 220, 244, 244, 246, 287, 463, 573, 640, 695, 730, 763, 844, 856, 868, 871, 883, 886, 888, 898, 910, 913, 974 (26)	-1.40423	0.96958

Table D.5 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*pyr*G	51, 54, 60, 69, 72, 73, 74, 78, 102, 102, 138, 144, 153, 171, 180, 192, 270, 273, 285, 285, 333, 354, 399, 414, 429, 471, 543, 558, 600, 609, 657, 675, 687, 693, 777, 888, 891, 915, 945, 957, 1026, 1050, 1056, 1056, 1086, 1098, 1110, 1146, 1164, 1224, 1230, 1314, 1332, 1392, 1398, 1410, 1413, 1416, 1419, 1443, 1479, 1497 (62)	110, 776, 1018, 1202, 1219, 1231, 1388, 1445, 1459, 1502, 1591, 1594 (12)	-1.72779	0.76546
*rec*A	30, 60, 69, 75, 99, 135, 195, 220, 225, 226, 231, 246, 255, 261, 270, 276, 285, 288, 291, 300, 309, 318, 351, 366, 405, 420, 426, 441, 447, 561, 564, 570, 594, 612, 618, 624, 642, 651, 669, 732, 756, 795, 807, 813, 825, 840, 861, 912, 966 (49)	163, 181, 588, 602, 682, 1003 (6)	-1.46434	1.54809
*rec*J	165, 174, 249, 261, 300, 318, 375, 420, 432, 510, 528, 540, 543, 558, 561, 615, 651, 672, 678, 688, 741, 756, 954, 972, 981, 984, 987, 1051, 1059, 1062, 1092, 1137, 1215, 1245, 1275, 1287, 1425, 1530, 1533, 1536, 1548, 1563 (42)	250, 355, 427, 520, 565, 598, 1417, 1534, 1555 (9)	-0.71025	0.28068
*rec*N	51, 69, 75, 84, 87, 99, 108, 111, 123, 132, 180, 192, 210, 219, 243, 339, 396, 399, 426, 438, 529, 540, 609, 648, 687, 717, 738, 747, 750, 759, 825, 834, 852, 870, 879, 885, 894, 897, 900, 1068, 1080, 1230, 1257, 1272, 1296, 1311, 1317, 1362, 1371, 1401, 1410, 1413, 1449, 1491, 1503, 1512 (56)	178, 421, 481, 488, 539, 601, 715, 799, 851, 889, 922, 992, 1003, 1079, 1091, 1103, 1109, 1125, 1215, 1225, 1519 (21)	-0.11234	-49.2621
*rec*R	15, 78, 84, 96, 102, 123, 183, 195, 210, 243, 282, 432, 447, 510 (14)	5, 176, 226, 236, 367, 461 (6)	-0.94561	1.38957
*rpl*B	48, 90, 117, 123, 132, 228, 240, 258, 333, 339, 339, 345, 384, 387, 405, 444, 468, 474, 477, 486, 537, 540, 558, 588, 591, 594, 609, 654, 657, 672, 675, 747, 762, 768, 789 (35)	116, 167, 349, 610, 781 (5)	-1.01716	1.7962

Table D.5 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*rpo*B	147, 429, 441, 459, 498, 606, 966, 1059, 1233, 1263, 1278, 1419, 1422, 1611, 1635, 1656, 1791, 1812, 1869, 1887, 1956, 1986, 2004, 2031, 2037, 2082, 2154, 2172, 2286, 2337, 2514, 2571, 2595, 2622, 2655, 2658, 2664, 2682, 2706, 2745, 2754, 2910, 2961, 2991, 3033, 3258, 3294, 3297, 3303, 3336, 3351, 3360, 3363, 3366, 3378, 3417, 3432, 3438, 3459, 3462, 3468, 3486, 3522, 3699, 3714, 3744, 3747, 3750, 3756, 3774, 3801, 3807, 3816, 3825, 3846, 3852, 3906, 3909, 3948, 3969, 3972, 3981, 4005 (43)	1543, 2631, 2965, 2980, 3445, 3491, 3528, 3560, 3640, 3691, 4043 (11)	-0.03347	36.90874
*rpo*D	63, 66, 72, 108, 129, 156, 174, 180, 258, 273, 300, 402, 414, 435, 495, 537, 543, 603, 624, 675, 700, 717, 726, 729, 735, 759, 822, 831, 858, 948, 969, 990, 1029, 1036, 1047, 1056, 1089, 1095, 1104, 1119, 1173, 1215, 1227, 1233, 1254, 1272, 1326, 1350, 1353, 1383, 1407, 1437, 1443, 1497, 1509, 1536, 1545, 1602, 1629, 1662, 1680, 1698, 1704, 1710, 1743, 1755, 1758, 1797, 1854 (69)	19, 118, 577, 586, 598, 599, 730, 1057, 1453, 1489, 1614 (11)	-1.3889	1.48297
*ruv*A	96, 123, 147, 156, 186, 192, 228, 243, 264, 312, 345, 345, 363, 369, 387, 390, 429, 501, 510 (19)	283, 395, 418, 491, 508, 547 (6)	-0.86384	0.61158
ssb	39, 42, 60, 66, 87, 93, 135, 159, 171, 189, 231, 234, 240, 393, 411, 420, 465 (17)	115, 325, 427, 448, 457, 476, 479 (7)	-0.87871	-0.17492

Table D.5 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*uvr*A	27, 57, 69, 75, 81, 102, 105, 132, 144, 147, 162, 177, 183, 186, 207, 210, 222, 243, 255, 270, 282, 291, 294, 297, 318, 325, 334, 339, 348, 348, 351, 372, 384, 459, 498, 540, 552, 579, 582, 603, 621, 651, 675, 681, 720, 723, 759, 762, 765, 765, 789, 801, 804, 813, 837, 840, 843, 852, 864, 894, 912, 921, 951, 966, 972, 1032, 1047, 1056, 1068, 1071, 1101, 1113, 1125, 1170, 1194, 1203, 1224, 1224, 1236, 1239, 1248, 1257, 1272, 1299, 1326, 1335, 1341, 1362, 1365, 1389, 1425, 1434, 1437, 1452, 1506, 1509, 1512, 1638, 1662, 1671, 1692, 1698, 1716, 1722, 1728, 1749, 1767, 1803, 1818, 1872, 1887, 1917, 1936, 1965, 1968, 1971, 1992, 1995, 1998, 2031, 2058, 2094, 2130, 2139, 2160, 2160, 2181, 2181, 2184, 2187, 2202, 2259, 2274, 2289, 2292, 2310, 2319, 2322, 2328, 2334, 2376, 2397, 2409, 2427, 2433, 2442, 2472, 2499, 2506, 2511, 2532, 2550, 2580, 2610, 2628, 2637, 2643, 2652, 2670, 2700, 2712, 2712, 2715, 2718, 2724, 2757, 2760, 2763, 2769, 2770, 2772, 2775, 2778, 2784, 2791, 2796, 2797, 2811, 2815, 2817 (180)	85, 368, 398, 413, 457, 625, 764, 905, 1021, 1022, 1031, 1037, 1037, 1073, 1237, 1255, 1798, 2095, 2275, 2276, 2704, 2749, 2771, 2812, 2816, 2819, 2820, 2822 (28)	-1.21604	0.93944

Table D.5 (continued)

Genes	Synonymous sites	Non-synonymous sites	TD_gene	TD_nonsyn/syn
*uvr*B	15, 18, 30, 36, 60, 72, 99, 120, 168, 189, 201, 204, 216, 249, 297, 327, 333, 333, 336, 366, 378, 405, 420, 480, 508, 564, 594, 624, 630, 696, 714, 741, 744, 747, 756, 757, 768, 777, 780, 816, 856, 858, 876, 906, 915, 924, 928, 936, 948, 993, 1005, 1020, 1026, 1035, 1035, 1042, 1044, 1044, 1068, 1074, 1077, 1080, 1083, 1086, 1089, 1102, 1104, 1104, 1107, 1113, 1114, 1116, 1122, 1164, 1182, 1185, 1191, 1194, 1195, 1197, 1200, 1209, 1212, 1215, 1221, 1236, 1245, 1248, 1252, 1254, 1266, 1272, 1273, 1275, 1284, 1293, 1314, 1335, 1338, 1344, 1353, 1356, 1359, 1362, 1383, 1404, 1437, 1446, 1473, 1482, 1513, 1530, 1533, 1534, 1536, 1542, 1548, 1551, 1566, 1578, 1581, 1581, 1593, 1626, 1644, 1668, 1671, 1680, 1701, 1704, 1713, 1728, 1734, 1752, 1755, 1755, 1767, 1770, 1773, 1776, 1779, 1785, 1788, 1791, 1803, 1806, 1812, 1837, 1839, 1857, 1866, 1869, 1872, 1875, 1881, 1882, 1884, 1885, 1887, 1890, 1905, 1908, 1908, 1914, 1917, 1920, 1932, 1935, 1935, 1938, 1941, 1953 (172)	19, 165, 365, 367, 556, 751, 773, 784, 785, 907, 943, 953, 988, 1156, 1300, 1324, 1402, 1497, 1594, 1726, 1727, 1738, 1744, 1804, 1839, 1856, 1870, 1936, 1939 (29)	-0.90449	2.60546
*xse*A	6, 12, 15, 24, 27, 33, 37, 40, 42, 45, 48, 57, 63, 90, 108, 132, 144, 156, 186, 198, 204, 231, 246, 249, 252, 376, 390, 402, 408, 456, 600, 603, 669, 678, 741, 744, 762, 775, 777, 795, 816, 873, 930, 961, 963, 1086, 1089, 1116 (48)	7, 8, 11, 29, 47, 49, 57, 58, 64, 137, 148, 197, 205, 241, 362, 545, 571, 718, 819, 832, 840, 1136 (22)	-1.61251	1.4809

D.2.4 Phylogenetic analysis

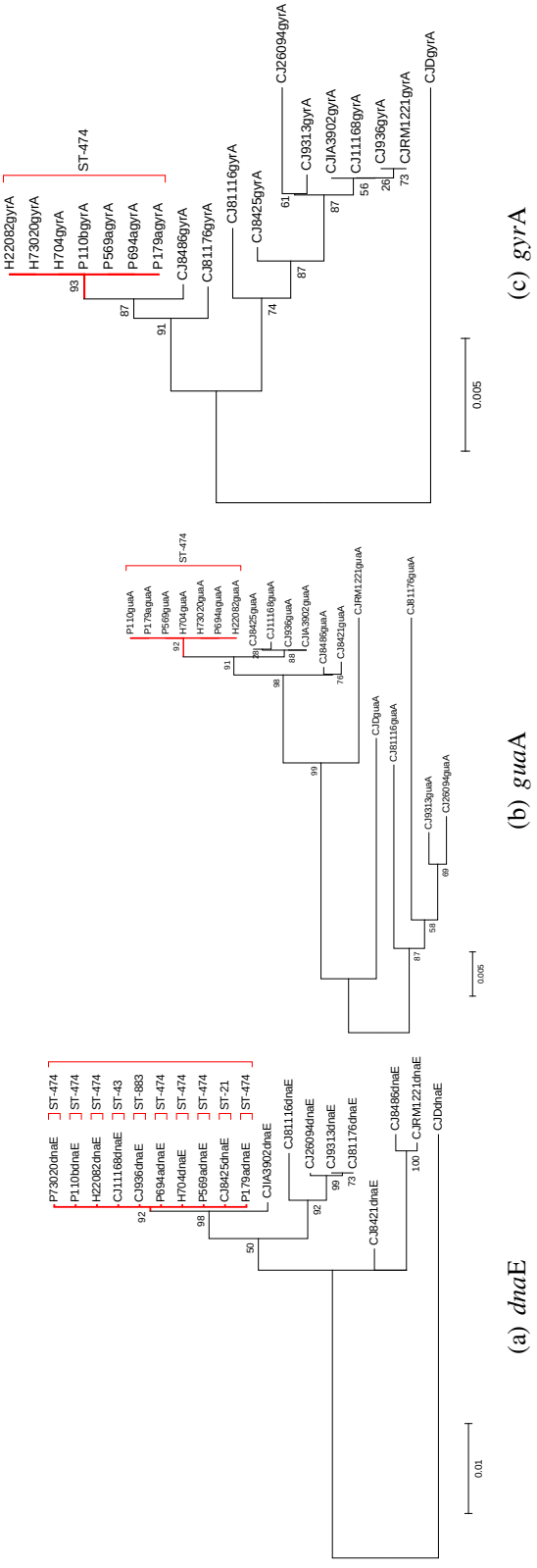
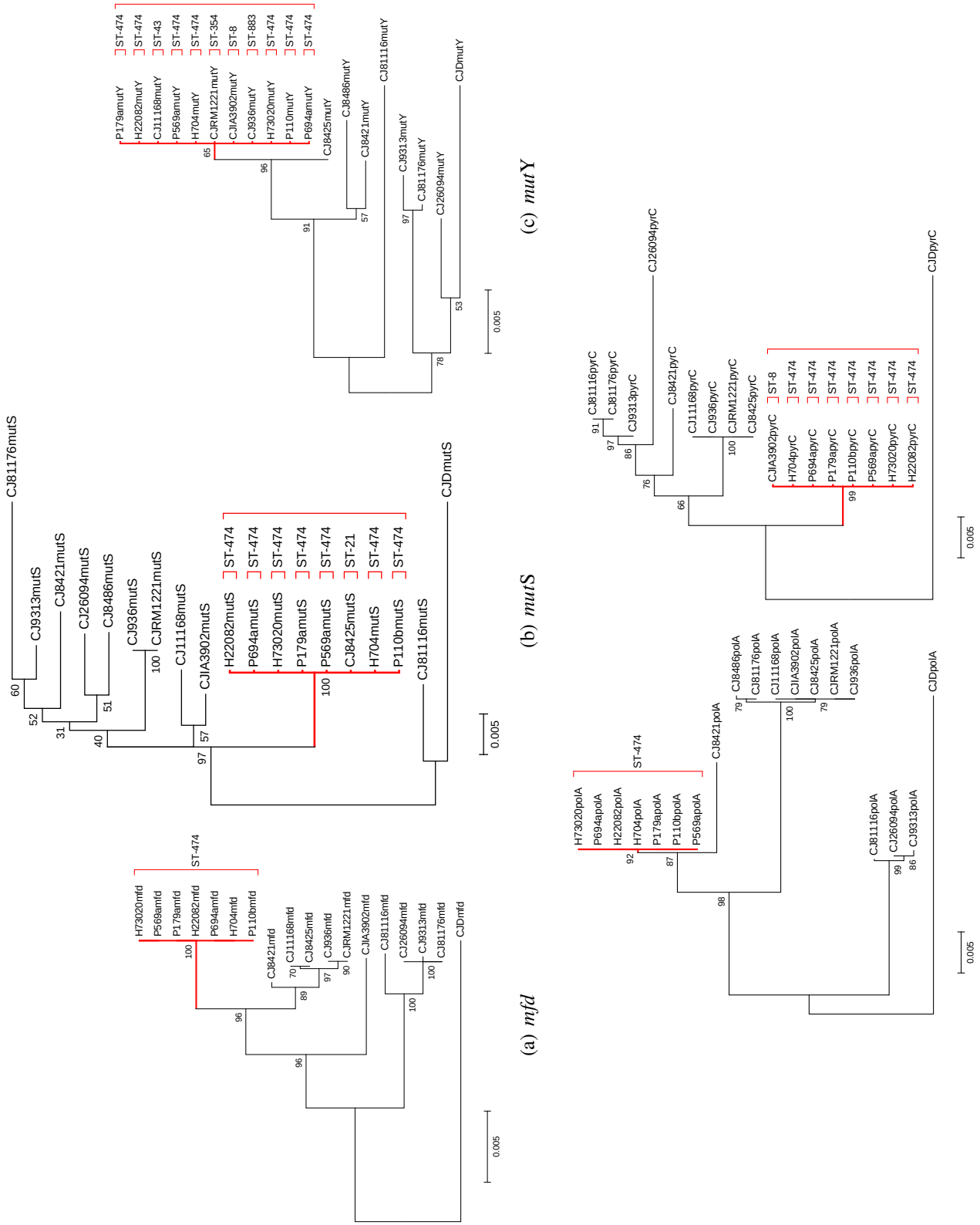


Figure D.1: Maximum likelihood gene trees of ribosomal and repair genes


(a) mfd
ST-474
H73020mfd
P569amfd
P179amfd
H22082mfd
P694amfd
H704mfd
P110bmfd
CJ8421mfd
CJ11168mfd
CJ8425mfd
CJ936mfd
CJRM1221mfd
CJIA3902mfd
CJ81116mfd
CJ26094mfd
CJ9313mfd
CJ81176mfd
CJDmfd
0.005
(b) mutS
CJ81176mutS
CJ9313mutS
CJ8421mutS
CJ26094mutS
CJ8486mutS
CJ936mutS
CJRM1221mutS
CJ11168mutS
CJIA3902mutS
H22082mutS
P694amutS
H73020mutS
P179amutS
P569amutS
CJ8425mutS
H704mutS
P110bmutS
CJ81116mutS
CJDmutS
ST-474
ST-21
0.005
(c) mutY
P179amutY
H22082mutY
CJ11168mutY
P569amutY
H704mutY
CJRM1221mutY
CJIA3902mutY
CJ936mutY
H73020mutY
P110mutY
P694amutY
CJ8425mutY
CJ8486mutY
CJ8421mutY
CJ81116mutY
CJ9313mutY
CJ81176mutY
CJ26094mutY
CJDmutY
ST-474
ST-43
ST-354
ST-8
ST-883
0.005
(d) polA
H73020polA
P694apolA
H22082polA
H704polA
P179apolA
P110bpolA
P569apolA
CJ8421polA
CJ8486polA
CJ81176polA
CJ11168polA
CJIA3902polA
CJ8425polA
CJRM1221polA
CJ936polA
CJ81116polA
CJ26094polA
CJ9313polA
CJDpolA
ST-474
0.005
(e) pyrC
CJ81116pyrC
CJ81176pyrC
CJ9313pyrC
CJ26094pyrC
CJ8421pyrC
CJ11168pyrC
CJ936pyrC
CJRM1221pyrC
CJ8425pyrC
CJIA3902pyrC
H704pyrC
P694apyrC
P179apyrC
P110bpyrC
P569apyrC
H73020pyrC
H22082pyrC
CJDpyrC
ST-8
ST-474
0.005

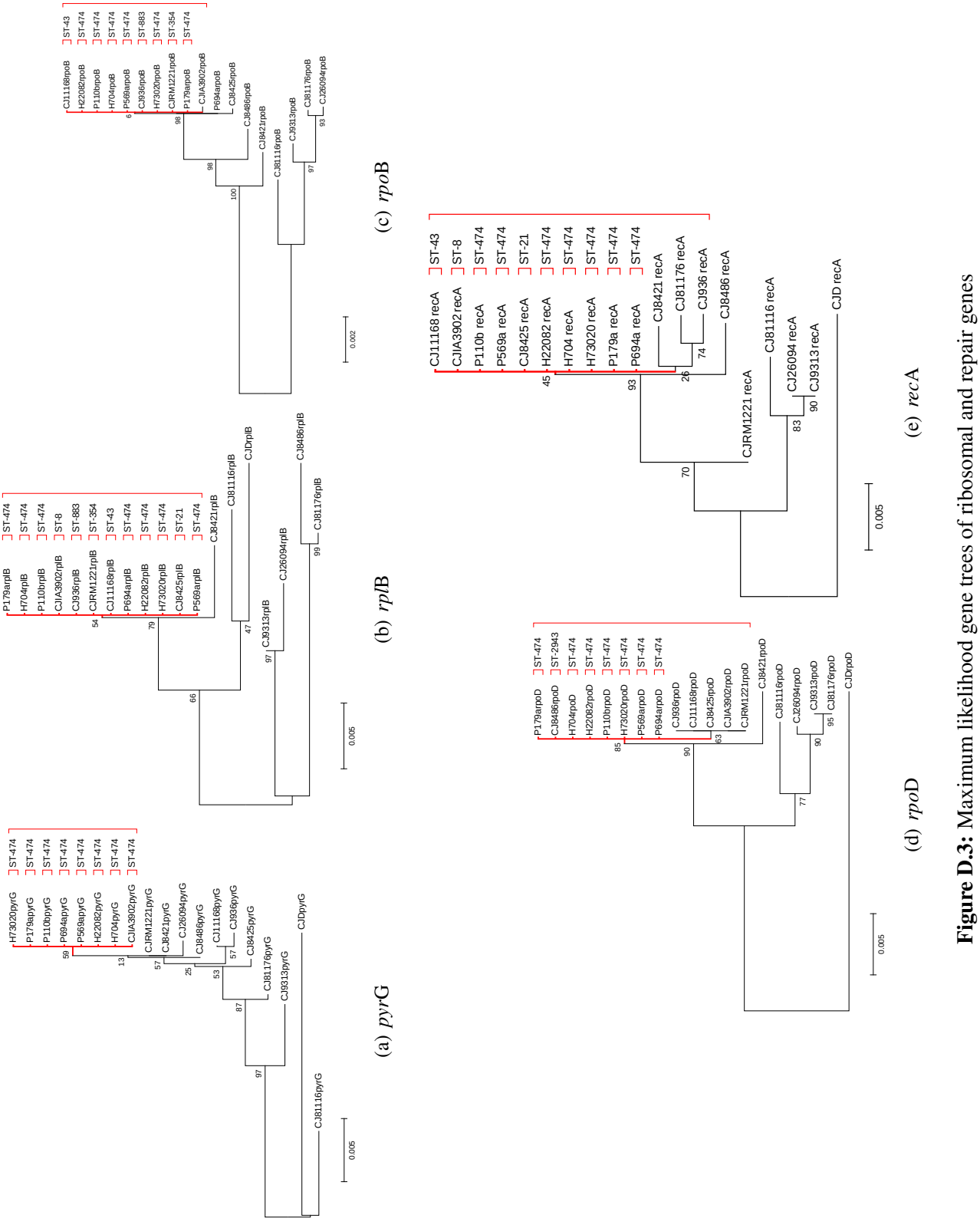


Figure D.3: Maximum likelihood gene trees of ribosomal and repair genes

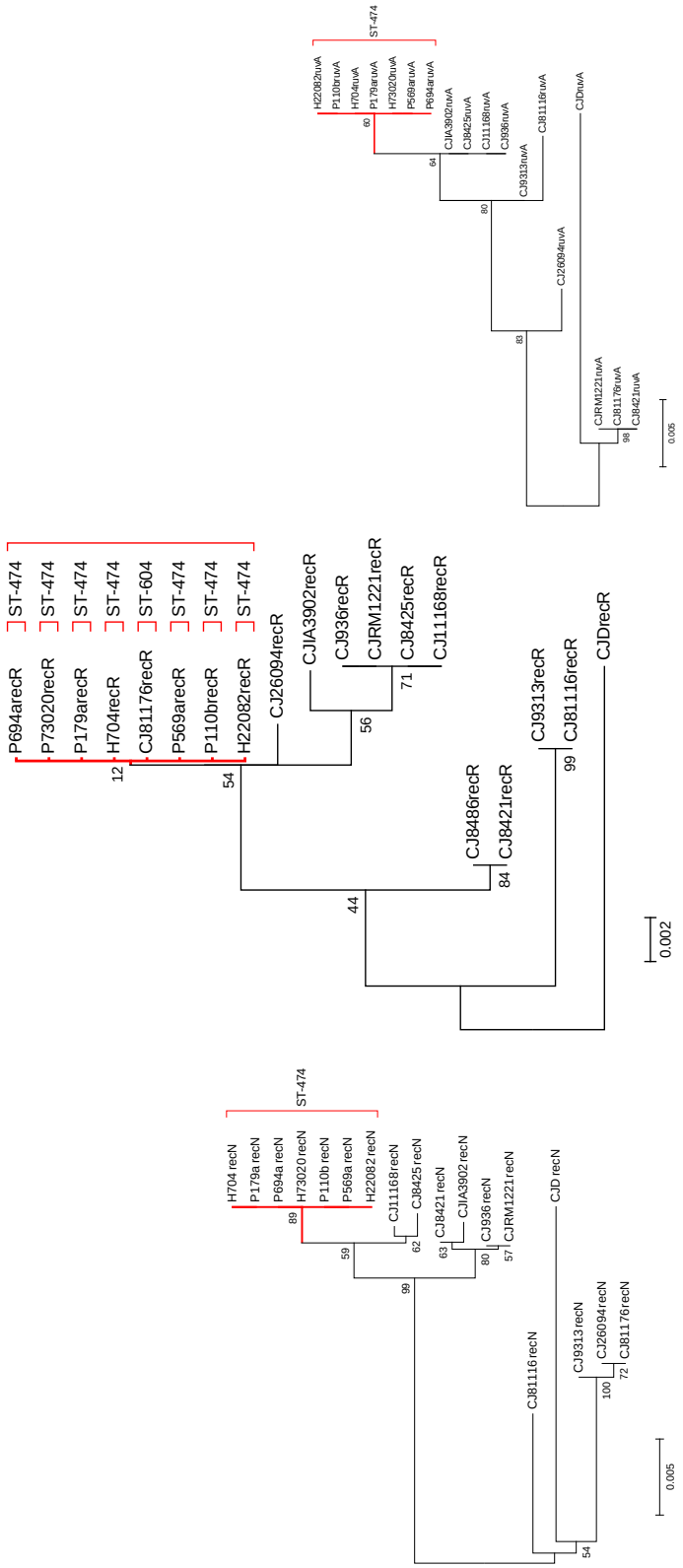


Figure D.4: Maximum likelihood gene trees of ribosomal and repair genes