The genomic epidemiology of *Campylobacter* from the Republic of South Africa



Melissa Jane Jansen van Rensburg

Merton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2015

ABSTRACT

As the leading cause of bacterial gastroenteritis, *Campylobacter* represents a significant public health burden; however, our knowledge of its epidemiology in low- and middle-income countries remains limited. Recent studies have demonstrated the power of whole-genome sequencing (WGS) for public health microbiology. The primary aim of this thesis was to exploit WGS to improve our understanding of the epidemiology of *Campylobacter* from the Republic of South Africa, a middle-income country.

In the first half of this thesis, *in silico* approaches were developed to evaluate diagnostic assays and methods of species identification. Large-scale analyses of publicly available WGS data identified a robust real-time PCR assay for the detection of *Campylobacter jejuni* and *Campylobacter coli*, the primary causes of human campylobacteriosis. Evaluation of *in silico* speciation methods demonstrated that the *atpA* gene and ribosomal multilocus sequence typing can be used to identify *Campylobacter* from WGS data.

The second half of this thesis extended concepts developed in the first half to investigate the epidemiology of *Campylobacter* from animals and humans from South Africa. Isolates from a study of *Campylobacter* from free-range broiler carcasses belonged to the agriculture-associated ST-828 lineage, but were atypically homogenous and differed at only 46/1,513 (3%) loci, providing novel insights into clonal infections in chickens. Analyses of human disease isolates collected in Cape Town in 1991, 2011, and 2012 confirmed that the local epidemiology of *Campylobacter* is distinct from that of high-income countries: in addition to major agriculture-associated *C. jejuni* and *C. coli* lineages, a putative novel *C. jejuni* subsp. *jejuni*/C. *jejuni* subsp. *doylei* hybrid clade and genetically diverse *C. jejuni* subsp. *doylei* and *C. upsaliensis* isolates were identified.

This work delivers further evidence of the utility of WGS for clinical microbiology, presents approaches that address general problems in *Campylobacter* diagnostics and public health microbiology, and provides insights into the epidemiology of this important group of pathogens in South Africa.

DECLARATION

The work presented in this thesis would not have been possible without the contributions of collaborators from the United Kingdom and South Africa. With respect to Chapter 2, I would like to acknowledge the contributions of: Dr Claire Jenkins, Public Health England (PHE), who initiated the work; Dr Alison Cody, University of Oxford, who provided genomic DNA and bacterial cultures; and Dr Craig Swift, PHE, who assisted with repeat testing of isolates with atypical results.

The isolates described in Chapter 4 were collected by Ms Laeeqa Basardien, University of the Western Cape (UWC), South Africa, during her MSc, which was supervised by Professor Al Lastovica, UWC, and Professor Pieter Gouws, formerly UWC, now at the University of Stellenbosch, South Africa. Similarly, the isolates described in Chapter 5 were collected by staff at the University of Cape Town (UCT), the Red Cross War Memorial Children's Hospital, and Groote Schuur Hospital, Cape Town, South Africa. Particular thanks are owed to: Professor Al Lastovica, who established and maintained the *Campylobacter* archive while he was at UCT; Ms Rachel Jacobson, currently at the University of Chicago, and Professor Mark Nicol, who oversaw the collection of *Campylobacter* isolates during 2011 and 2012; Ms Layla Hendricks and Dr Widaad Zemanay, who co-ordinated cataloguing of the archive and shipment of the study isolates; and Ms Slee Mbhele and Dr Justyna Wojno, who provided clinical data. WGS of the isolates described in Chapters 4 and 5 was carried out at the Wellcome Trust Sanger Institute.

All studies made use of BIGSDB software and additional scripts developed by Dr Keith Jolley, University of Oxford, and a VELVET pipeline, developed by Dr James Bray, University of Oxford, for the *de novo* assembly of whole-genome sequencing data. Mr Katriel Cohn-Gordon, University of Oxford, provided additional bioinformatics support, assisting with the generation of in-house Python scripts to improve the analysis of large datasets. Numerous analyses, particularly those described in Chapters 2 and 3, were contingent on publicly available WGS data from PubMLST databases, GenBank, or the Sequence Read Archive. Where available, peer-reviewed references have been provided, but this serves to acknowledge all those who released genomic data prior to publication.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Martin Maiden, for his advice and support during the preparation of this thesis. I would also like to acknowledge past and present members of the Maiden group who laid the foundations that made this work possible. Special thanks to Keith Jolley and James Bray for bioinformatics support, and Alison Cody, Frances Colles, and Noel McCarthy for technical assistance and many helpful discussions. I am also especially grateful to my South African collaborators, Al Lastovica, Laeeqa Basardien, Pieter Gouws, Mark Nicol, and Rachael Jacobson, without whom this thesis would not have been possible.

Regarding funding, I would like to acknowledge the Clarendon Fund and Merton College for the awards that enabled me to accept my DPhil offer. Thanks also to Funds for Women Graduates who provided additional financial assistance during my fourth year.

Lastly, I would like to thank my friends and family for their ongoing support. I am particularly grateful to Avra Cohn for proof-reading so much of this document – any mistakes that remain are mine. Thanks also to Kinda Al-Hourani, Nikki Bart, Harriet Keane, and Katriel Cohn-Gordon for their help and encouragement, especially in the eleventh hour. Most of all, thanks to my sister, Jennifer, and my mother, Alison, who, as always, did far more than expected.

CONTENTS

ABSTRACT ..... **ii**

DECLARATION ..... **iii**

ACKNOWLEDGEMENTS ..... **iv**

CONTENTS ..... **v**

LIST OF FIGURES ..... **viii**

LIST OF TABLES ..... **xiii**

ABBREVIATIONS, ACRONYMS, AND SYMBOLS ..... **xv**

PREFACE **1**

Diarrhoeal disease in the 21st century *2*

Applications of next-generation sequencing for the study of bacterial gastrointestinal pathogens *6*

Ecology of the gut microbiome 6

Persistence and pathogenesis 7

Epidemiology and evolution 9

CHAPTER 1 ..... **13**

Introduction *13*

1.1 Campylobacter *taxonomy, ecology, and microbiology* *13*

1.2 *Human campylobacteriosis* *17*

1.3 *Laboratory diagnosis of* Campylobacter *infections* *26*

1.4 *The descriptive epidemiology of* Campylobacter *30*

1.5 *Molecular typing of* Campylobacter *32*

1.6 *Project background and objectives* *65*

Part I: Using whole-genome sequencing to address problems in *Campylobacter* diagnostics 69

Part II: The genomic epidemiology of *Campylobacter* from South Africa 69

Part I: Using whole-genome sequencing to address problems in *Campylobacter* diagnostics 71

CHAPTER 2 72

Exploiting bacterial whole-genome sequencing data for the evaluation of diagnostic assays: Campylobacter *speciation as a case study 72*

2.1 Introduction 73

2.2 Materials and Methods 77

2.3 Results 82

2.4 Discussion 97

CHAPTER 3 101

Evaluation of typing schemes for in silico identification of Campylobacter *species 101*

3.1 Introduction 102

3.2 Materials and Methods 107

3.3 Results 111

3.4 Discussion 133

Part II: The genomic epidemiology of *Campylobacter* from South Africa 139

CHAPTER 4 140

The genetic diversity of Campylobacter *from two South African free-range broiler flocks at slaughter 140*

4.1 Introduction 142

4.2 Materials and Methods 148

4.3 Results 153

4.4 Discussion *166*

CHAPTER 5 ..... **173**

The genomic epidemiology of Campylobacter *from human disease samples from Cape Town, South Africa* *173*

5.1 Introduction *174*

5.2 Methods *179*

5.3 Results *185*

5.4 Discussion *201*

Final discussion **206**

REFERENCES **211**

APPENDICES **260**

Appendix A *260*

Appendix B *279*

Appendix C *280*

Appendix D *284*

Appendix E *292*

Appendix F *293*

Appendix G *294*

Appendix H *303*

Appendix I *304*

LIST OF FIGURES

Figure 1.1 Sources and clinical manifestations of *Campylobacter* infections in humans. 18

Figure 1.2 Transcellular and paracellular routes of *Campylobacter* migration across the intestinal epithelium during acute gastroenteritis in humans. 24

Figure 1.3 Multilocus sequence typing workflow. 40

Figure 1.4 Distribution of multilocus sequence typing loci (black) and antigen genes (orange) across the *C. jejuni* NCTC 11168 genome. 42

Figure 1.5 Growth of the *Campylobacter jejuni/coli* PubMLST database between 2001 and 2015. 42

Figure 1.6 Use of next-generation sequencing platforms for bacterial DNA sequencing. 50

Figure 1.7 Improvements in instrument output and read length of next-generation sequencing platforms produced by key manufacturers between 2005 and 2015. 51

Figure 1.8 The Illumina workflow. 52

Figure 1.9 Using the gene-by-gene approach to relate sequence-based typing data to taxonomy. 60

Figure 1.10 The gene-by-gene approach for analysis of whole-genome sequencing data as implemented with Bacterial Isolate Genome Sequence Database software. 62

Figure 1.11 Map of South Africa...... 66

Figure 2.1 Genetic diversity and speciation of *Campylobacter* isolates from the Oxfordshire sentinel surveillance collection. 83

Figure 2.2 Genetic diversity of concatenated *mapA* (A) and *ceuE* (B) primer and probe sequences. 86

Figure 2.3 Neighbor-joining trees based on 72 *mapA* (A) and 126 *ceuE* (B) unique gene sequences...... 91

Figure 2.4 Characterisation of introgression in *mapA*...... 93

Figure 2.5 Frequency distribution of *mapA* (A) and *ceuE* (B) nucleotide (black) and protein alleles (white) in *C. jejuni* and *C. coli*. 95

Figure 2.6 Distribution of synonymous and non-synonymous substitutions across *mapA* (A) and *ceuE* (B) in *C. jejuni* and *C. coli* as determined using SNAP (www.hiv.lanl.gov) (Korber, 2000)...... 96

Figure 3.1 Frequency distribution of putative orthologs identified by GENOME COMPARATOR (pink), and orthologous gene clusters delimited by OrthoMCL (blue) and Proteinortho (yellow), based on analyses of 70 genome sequences representative of 22 *Campylobacter* species. 118

Figure 3.2 Venn diagram indicating the number of unique and overlapping putative orthologs or orthologous gene clusters identified using GENOME COMPARATOR (pink), OrthoMCL (blue), and Proteinortho (yellow). 119

Figure 3.3 Mean *p*-distances calculated by GENOME COMPARATOR for combined core (blue) and non-core (red) coding sequences plotted relative to genome position in NCTC 11168. 120

Figure 3.4 Comparison of the functional distribution of coding sequences present in *C. jejuni* isolate NCTC 11168 and the *Campylobacter* core genome. 121

Figure 3.5 Evolutionary relationships among 70 isolates representative of 22 *Campylobacter* species. 123

Figure 3.6 Neighbor-joining phylogeny based on 16S rRNA gene sequences of 69 *Campylobacter* isolates. 126

Figure 3.7 Neighbor-joining phylogeny based on *atpA* sequences of 70 *Campylobacter* isolates. 127

Figure 3.8 Neighbor-joining phylogeny based on concatenated nucleotide sequences of six MLST loci from 70 *Campylobacter* isolates. 129

Figure 3.9 Neighbor-joining phylogeny based on concatenated nucleotide sequences of 52 rMLST loci from 70 *Campylobacter* isolates. 132

Figure 4.1 Genomic diversity among *Campylobacter* isolates from two free-range broiler flocks. 156

Figure 4.2 Comparison of the genomic diversity of ST-1587 isolates from chicken carcasses and human disease samples from South Africa and Denmark...... 158

Figure 4.3 Distribution of variable loci across the genome of *C. coli* 15-537360. 160

Figure 4.4 Assignment of variable loci to functional categories...... 161

Figure 4.5 The distribution of allelic variants at variable loci identified among *Campylobacter* from two free-range chicken flocks and a human disease isolate from the Western Cape, South Africa. 163

Figure 5.1 Recovery of genomic DNA from freeze-dried *Campylobacter* isolates..... 185

Figure 5.2 Neighbor-joining phylogeny based on *atpA* sequences of 159 *Campylobacter* isolates...... 188

Figure 5.3 Neighbor-joining phylogeny based on concatenated nucleotide sequences of 52 ribosomal protein genes from 159 *Campylobacter* isolates...... 190

Figure 5.4 Relationships among 83 *Campylobacter* isolates based on core genome diversity...... 192

Figure 5.5 Distribution of *Campylobacter* species among collections of human disease isolates from Cape Town, South Africa. 194

Figure 5.6 Frequency distribution of clonal complexes among *C. jejuni* and *C. coli* from the 1991 (black) and 2011/2012 (white) human disease isolate collections from Cape Town, South Africa. 196

Figure 5.7 Whole-genome MLST comparisons of South African and UK *C. jejuni* isolates according to clonal complex...... 198

Figure 5.8 Whole-genome MLST comparisons of South African and UK *C. coli* isolates...... 200

LIST OF TABLES

Table 1.1 Members of the *Campylobacter* genus and their known sources of isolation 15

Table 1.2 Infections associated with *Campylobacter* species..... 20

Table 1.3 Multilocus sequence typing schemes from the non-*jejuni*/*coli* *Campylobacter* PubMLST database 43

Table 2.1 Details of *ceuE* alleles with internal stop codons identified among *C. jejuni* isolates..... 84

Table 2.2 Details of *C. coli* isolates with atypical *mapA*/*ceuE* RT-PCR primer and probe sequences 87

Table 2.3 RT-PCR cycle thresholds for *mapA* and *ceuE* primer and probe combinations..... 89

Table 2.4 Intra- and interspecies diversity of *mapA* and *ceuE* gene and protein sequences..... 94

Table 3.1 Details of publicly available *Campylobacter* genomes included in this study..... 112

Table 3.2 Gene flow among eight *Campylobacter* species based on 657 core genes as measured by F_{ST}..... 122

Table 3.3 Gene flow among eight *Campylobacter* species as measured by F_{ST} based on the 16S rRNA gene (below the diagonal) and *atpA* (above the diagonal) 124

Table 3.4 Gene flow among eight *Campylobacter* species as measured by F_{ST} based on six MLST genes (below the diagonal) and 52 rMLST genes (above the diagonal) 130

Table 4.1 Variable loci with poor quality nucleotide sequence alignments . 162

Table 4.2 Details of underlying genetic diversity in variable loci with high quality nucleotide sequence alignments 164

Table 5.1 Species-specific properties of *Campylobacter* genomes from the 1991 archive collection 189

Table 5.2 Isolates from the 1991 and 2011/2012 collections with conflicting laboratory-based and *in silico* species assignments..... 193

Table 5.3 Identification of existing and novel sequence types (STs) among *C. jejuni* and *C. coli* isolates from the 1991 and 2011/2012 *Campylobacter* collections 196

ABBREVIATIONS, ACRONYMS, AND SYMBOLS

°C	Degrees Celsius
µg	Microgram
µl	Microlitre
µm	Micometer
AFLP	Amplified fragment length polymorphism
aspA	Aspartase A
AUS	Australian dollars
BHI	Brain-heart infusion
BIGSDB	Bacterial Isolate Genome Sequence Database
bp	Base pair(s)
CAD	Canadian dollars
CBA	Columbia horse blood agar
CC	Clonal complex
CDS	Coding sequence
CDT	Cytolethal distending toxin
cgMLST	Core genome multilocus sequence typing
CHBH	Chris Hani Baragwanath Hospital
CI	Confidence interval
COG	Clusters of Orthologous Groups of Proteins
C_T	Cycle threshold
CTP	Cape Town protocol
D	Simpson's index of diversity
DALY	Disability-adjusted life year
ddNTP	Dideoxynucleotide
d_N	Rate of non-synonymous mutations
dNTP	Deoxynucleotide
d_S	Rate of synonymous mutations
ERIC-PCR	Enterobacterial repetitive intergenic consensus PCR
GBS	Guillain-Barré syndrome
GDP	Gross domestic product
glnA	Glutamine synthetase
glmM	Phosphoglucosamine mutase
gltA	Citrate synthase
glyA	Serine hydroxymethyl-transferase
IBD	Inflammatory bowel disease
IBS	Irritable bowel syndrome
IL-8	Interleukin 8

k	Population number
LMICs	Low- and middle-income countries
LOS	Lipooligosaccharide
MALDI-TOF MS	Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry
mCCDA	Modified charcoal cefoperazone deoxycholate
MFS	Miller Fisher syndrome
MLEE	Multilocus enzyme electrophoresis
MLST	Multilocus sequence typing
NAAT	Nucleic acid amplification test
NGS	Next-generation sequencing
PCR	Polymerase chain reaction
PFGE	Pulsed-field gel electrophoresis
pgm	Phosphoglucomutase
RAPD	Random amplification of polymorphic DNA
RXH	Red Cross War Memorial Children's Hospital
Rep-PCR	Repetitive extragenic palindromic-PCR
RFLP	Restriction fragment length polymorphism
rMLST	Ribosomal multilocus sequence typing
rST	Ribosomal sequence type
RT-PCR	Real-time PCR
SNP	Single nucleotide polymorphism
SRA	Sequence Read Archive
SVR	Short variable region
ST	Sequence type
T	Type strain
TBA	Tryptose horse blood agar
tkt	Transketolase
uCMLST	Universal *Campylobacter* multilocus sequence typing
UCT	University of Cape Town
UK	United Kingdom
uncA	ATP synthase α subunit
USA	United States of America
UWC	University of the Western Cape
wgMLST	Whole-genome multilocus sequence typing
WGS	Whole-genome sequencing
YLDs	Years lived with disability
ZMW	Zero-mode waveguide

PREFACE

Campylobacter is a relative newcomer among bacterial gastrointestinal pathogens: although symptoms of campylobacteriosis were described as early as 1886 (Engberg, 2006), it was not until 1972 that *Campylobacter* was first isolated (Dekeyser *et al.*, 1972). *Campylobacter* has since been recognised as a leading bacterial cause of diarrhoeal disease and a major public health concern. In 2010, campylobacteriosis was associated with a global burden of 7.5 million disability-adjusted life years (DALYs), which accounted for 8.42% of all diarrhoeal disease and exceeded estimates for cholera (4.5 million), enterotoxigenic *Escherichia coli* (6.9 million), and *Shigella* (7.1 million) (Murray *et al.*, 2012). The international disease burden has provided the impetus for intensive research on *Campylobacter* over the past four decades, resulting in rapid advances in our knowledge; however, our understanding of the biology, epidemiology, and evolution of members of the genus remains incomplete.

Campylobacter epidemiology continues to be a topic of interest, particularly in reference to low- and middle-income countries (LMICs), as there is limited information from these regions. The primary aim of the work carried out for this thesis was to improve our understanding of the epidemiology of *Campylobacter* in the Republic of South Africa (from here on referred to as “South Africa”), a middle-income country. The secondary aim of the studies presented here was to use genomic approaches to address problems in diagnostic and public health microbiology, focusing on those relevant to *Campylobacter*. This preface contextualises the work presented in the body of the thesis by providing a brief

overview of diarrhoeal disease and current research on bacterial gastrointestinal pathogens.

Diarrhoeal disease in the 21st century

Although preventable and treatable, diarrhoeal disease remains an important public health issue (World Health Organisation, 2013). Viral, bacterial, and parasitic infections are the most common causes of diarrhoea worldwide (Dickinson & Surawicz, 2014), and the majority of the diarrhoeal disease burden is borne by LMICs, especially those in Africa and South East Asia (Boschi-Pinto *et al.*, 2008; Walker *et al.*, 2013a; Rahman *et al.*, 2014). After pneumonia, diarrhoea remains the second leading cause of mortality in children below the age of five years: approximately 760,000 children die from diarrhoeal disease each year, the majority of whom are from sub-Saharan Africa (World Health Organisation, 2013; Walker *et al.*, 2013a). The increased incidence and severity of infectious diarrhoea in LMICs has been attributed to poor sanitation, water quality, and personal hygiene, as well as malnutrition, zinc deficiency, vitamin A deficiency, and lack of access to treatment (Farthing *et al.*, 2013; Walker *et al.*, 2013a). Diarrhoeal disease is further compounded by the global HIV pandemic (Walker *et al.*, 2013a), particularly in Africa where HIV prevalence rates are highest (Hallett *et al.*, 2014). In contrast, infectious diarrhoea in high-income countries is typically non-fatal and predominantly due to the consumption of contaminated food or drink (Dickinson & Surawicz, 2014).

To reduce the diarrhoeal disease burden, programmes in LMICs have concentrated on improving sanitation, water quality, and personal hygiene, as well as nutrition and

access to treatment (Bhutta *et al.*, 2013). Declining diarrhoeal mortality in LMICs in Central and South America has also been attributed to rotavirus vaccination programmes; the recent implementation of the rotavirus vaccine in Africa and Asia is predicted to continue to reduce the diarrhoeal disease burden in these regions (Tate *et al.*, 2016). In contrast to programmes in LMICs, initiatives in high-income countries have focused on food safety (Newell *et al.*, 2010; Wielinga & Schlundt, 2013). These programmes have significantly reduced the global diarrhoeal disease burden, as evidenced by 19.1 and 35.7% decreases in the incidence of diarrhoea and the number of years lived with disability (YLDs), respectively, between 1990 and 2013 (Liu *et al.*, 2012; Walker *et al.*, 2013a; Vos *et al.*, 2015). Although interventions have been successful, disease incidence remains high: an estimated 2.7 billion cases of diarrhoea occurred in 2013, affecting the equivalent of 38% of the world's population. In the same year, diarrhoeal disease was associated with a global health burden of 6.9 million YLDs, approximately 0.8 times that of road injuries, and 1.7 and 4.1 times that of HIV/AIDS and meningitis, respectively (Vos *et al.*, 2015).

While the economic impact of infectious diarrhoea or gastroenteritis in LMICs has not been quantified extensively, the effects on health are known to be long-lasting: persistent diarrhoea, or multiple acute episodes per year during early childhood, may result in malnutrition (Guerrant *et al.*, 1992; Nel, 2010), growth stunting (Checkley *et al.*, 2008), and decreased cognitive function (Lorntz *et al.*, 2006; Walker *et al.*, 2011). In comparison, gastroenteritis typically does not have lasting effects in otherwise healthy children from high-income countries (Walker *et al.*, 2013a), but remains a concern because of its impact on society and the economy. Estimates of direct medical costs, loss of earnings, and pain and suffering suggest that gastroenteritis

costs the United Kingdom (UK) approximately £1.8 billion per annum (Food Standards Agency, 2013), equivalent to the total amount allocated by England's National Health Service to cover its countrywide public health responsibilities in 2013 (https://www.england.nhs.uk/allocations-2013-14/). The annual financial burden of food-borne illnesses is similarly high in other high-income countries, costing up to $77.7 billion per annum in the United States of America (USA) (Scharff, 2012), CAD 3.7 billion in Canada (Thomas *et al.*, 2008), €695 million in the Netherlands (Friesema *et al.*, 2012), and AUD 1.2 billion in Australia (Abelson *et al.*, 2006).

Given the public health and societal impact of diarrhoeal disease, improved control remains a global priority; however, political instability, conflict, and humanitarian emergencies impede progress in regions with the greatest disease burden (Walker *et al.*, 2013a; UN Inter-agency Group for Child Mortality Estimation, 2014). Co-operation of international organisations, governments, and local communities is critical if disease interventions are to be successful. It is also essential that we concurrently improve our understanding of the biology of gastrointestinal pathogens, as this knowledge can be harnessed to develop novel vaccines, therapeutic agents, diagnostic tools, and interventions.

Since the 1970s, advances in DNA sequencing technology have enabled researchers to address biological questions on a scale never before possible, particularly with respect to bacteria (Maiden, 2006; Wilson, 2012; McAdam *et al.*, 2014). The transition from chain termination or Sanger sequencing to next-generation sequencing (NGS) was an important milestone in the history of DNA

sequencing (Mardis, 2011), and is described in detail in Chapter 1 [1.5.2.2.3]. Although NGS encompasses several platforms with different chemistries [1.5.2.2.3], these methods all sequence multiple DNA fragments per sample in parallel, and can also be multiplexed, producing orders of magnitude more data per instrument run than Sanger sequencing (Figure 1) (Mardis, 2013). Concurrent decreases in sequencing costs have made NGS more accessible (Figure 1), with increasingly large volumes of bacterial whole-genome sequencing (WGS) data produced each year (Land *et al.*, 2015). For example, while it took several years to sequence the first *E. coli* genome (Blattner *et al.*, 1997), in 2012, the volume of sequencing data produced by the Wellcome Trust Sanger Institute, UK, alone was equivalent to approximately 3.3 million *E. coli* genomes, or one genome every ten seconds (http://www.sanger.ac.uk/resources/technologies/sequencing.html#t_3).

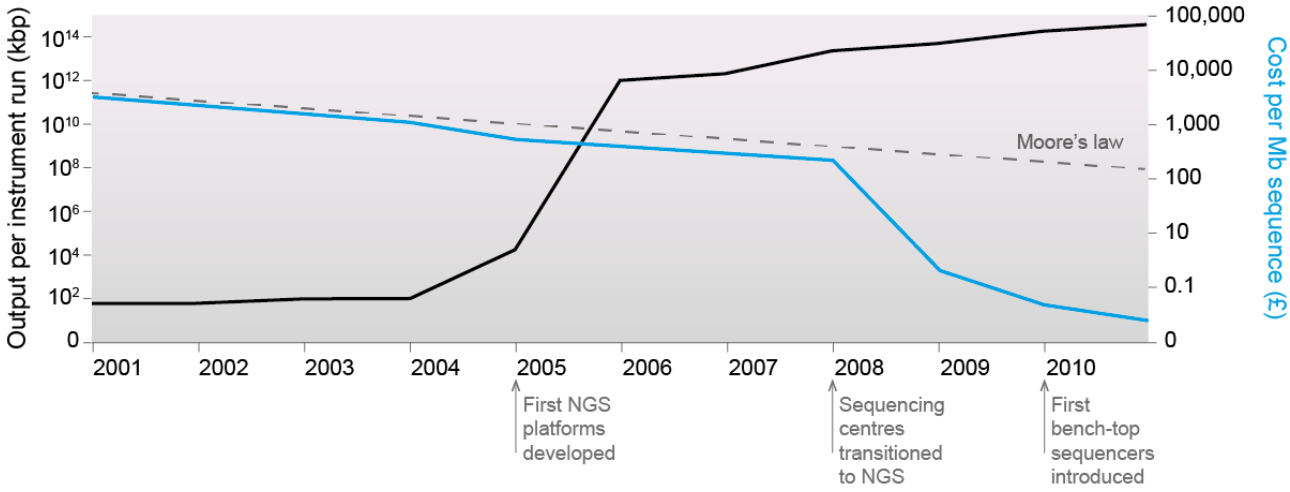


Figure 1 Changes in sequencing capacity (black) and cost (blue) between 2001 and 2010. Major developments in the field are indicated below the x-axis. The dashed grey line represents Moore's law, which predicts that the number of transistors on a microchip, often equated with computing power, will double each year (Carlson, 2003); this illustrates how, when seqencing centres transitioned to next-generation sequencing (NGS), the decrease in the cost of sequencing outstripped that of computing. Adapted from Mardis (2011) using data from the National Human Genome Research Institute website (http://www.genome.gov/sequencingcosts/).

Today, NGS has a broad range of applications for the study of bacteria (McAdam *et al.*, 2014). Major recent contributions to the field will be summarised here to illustrate how NGS can be used to study bacterial ecology, pathogenesis, epidemiology, and evolution, focusing on gastrointestinal pathogens.

Applications of next-generation sequencing for the study of bacterial gastrointestinal pathogens

Ecology of the gut microbiome

The advent of culture-independent NGS, typically based on 16S rRNA gene sequences, has facilitated large-scale studies of the microbial ecology of the human gut in health and disease (Lozupone *et al.*, 2012; Lepage *et al.*, 2013; Shreiner *et al.*, 2015). These studies have demonstrated that the gut microbiome helps to maintain human health by aiding digestion and nutrient extraction (Janssen & Kersten, 2015), modulating immune function (Belkaid & Hand, 2014), and protecting against gastrointestinal pathogens (Candela *et al.*, 2008; Fukuda *et al.*, 2011). Perturbations of the gut microbiota are predicted to adversely affect human health, and have been associated with inflammatory bowel disease (IBD) and other gastrointestinal disorders, such as acute diarrhoea, irritable bowel syndrome (IBS), Crohn's disease, and ulcerative colitis (Lozupone *et al.*, 2012; Lepage *et al.*, 2013; Wang *et al.*, 2014), in addition to malnutrition (Kau *et al.*, 2011), obesity (Rosenbaum *et al.*, 2015), cancer (Erdman & Poutahidis, 2015), and neurological conditions (Gonzalez *et al.*, 2011). Research in these areas remains a priority, as it is unclear whether changes in the gut microbiome are caused by or effect disease. In contrast, antibiotic treatment is an established cause of gut microbiome dysbiosis, which can

leave the human host susceptible to invasion by gastrointestinal pathogens. Antibiotics also enrich for a drug-resistant microbiome, which likely plays a role in the emergence of resistant human pathogens (Beaugerie & Petit, 2004; Salyers *et al.*, 2004; Pepin *et al.*, 2005; Sommer *et al.*, 2009; Dethlefsen & Relman, 2011; Modi *et al.*, 2013). An important example of the effects of antibiotic-associated gut microbiome dysbiosis is *Clostridium difficile* infection, a cause of hospital-acquired diarrhoea (Chang *et al.*, 2008). That *C. difficile* infection can be prevented with probiotics (Hickson *et al.*, 2007) and cured by faecal microbiota transplantation (Malnick & Melzer, 2015) highlights the importance of a healthy gut microbiome.

Persistence and pathogenesis

NGS technologies have applications beyond cataloguing nucleotide diversity, with methodologies developed for the study of DNA-protein interactions, gene expression and function, and regulatory elements (Mardis, 2008). One such technique, RNA-seq, generates a snapshot of the bacterial transcriptome: RNA is extracted from bacteria grown under conditions of interest, converted to cDNA by reverse transcription, and then sequenced using NGS technology. In addition to being used to study gene expression, RNA-seq protocols have been adapted to facilitate the identification of novel coding sequences, transcriptional start sites, operons, and non-coding RNAs (Croucher & Thomson, 2010).

RNA-seq studies have provided valuable insights into mechanisms of persistence and pathogenesis used by gastrointestinal bacteria. For instance, differential gene

expression within specialised sub-populations in *Salmonella enterica* serovar Typhimurium biofilms equips the community for invasion or environmental persistence as necessary (MacKenzie *et al.*, 2015). Relatedly, certain multidrug-resistant *S*. Typhimurium isolates upregulate the expression of invasion, attachment, and intracellular survival genes when exposed to sub-inhibitory antibiotic concentrations, which is thought to increase the severity and duration of infections that are already more likely to result in treatment failure due to resistance (Brunelle *et al.*, 2014). Although *S*. Typhimurium and *Salmonella enterica* serovar Typhi exhibit differences in host range and pathogenicity, characterisation of the OmpR regulon identified 31 orthologous OmpR-regulated genes involved in host colonisation, a subset of which may contribute to nutrient acquisition in the diseased intestine (Perkins *et al.*, 2013). Analogously, RNAseq studies have demonstrated that PerR is involved in oxidative stress resistance in *Campylobacter jejuni*, which promotes survival during transmission through the food chain (Handley *et al.*, 2015), while *Shigella flexneri* modulates gene expression in response to a lack of oxygen, promoting survival in the anaerobic environment of the intestine (Vergara-Irigaray *et al.*, 2014).

Another application of NGS that is starting to improve our understanding of bacterial pathogenesis is transposon-insertion sequencing, which combines parallel DNA sequencing with transposon mutagenesis for high-throughput study of gene function (van Opijnen & Camilli, 2013). A study of *S*. Typhimurium and *S*. Typhi identified 281 core genes 58 putative serovar-specific genes that likely reflect differences in host-adaptation (Barquist *et al.*, 2013). Other studies have identified genes essential for the colonisation of food-producing animals by *S*. Typhimurium (Chaudhuri *et al.*,

2013) and the survival of *S*. Typhi in bile (Langridge *et al.*, 2009). More recently, factors required for chicken colonisation have been identified in *C. jejuni* (Johnson *et al.*, 2014).

Epidemiology and evolution

Beyond high-throughput sequencing of individual loci, NGS can be used to rapidly sequence large numbers of bacterial genomes (Loman *et al.*, 2012a). WGS provides an unparalleled level of discriminatory power compared to other bacterial typing methods, which are described in Chapter 1 [1.5]. Consequently, there has been intense interest in the potential of WGS for use in public health microbiology (Didelot *et al.*, 2012; Köser *et al.*, 2012b; Fournier *et al.*, 2014), and the introduction of bench-top sequencers (Loman *et al.*, 2012b), in conjunction with decreasing WGS costs (Didelot *et al.*, 2012), is gradually making bacterial genomics accessible to a wider audience. Regarding bacterial gastrointestinal pathogens, recent retrospective studies have demonstrated the power of WGS for outbreak investigations, as well as long-term epidemiological surveillance and evolutionary studies (Wilson, 2012; McAdam *et al.*, 2014; Tang & Gardy, 2014).

One of the arguably best-known WGS-based epidemiological investigations took place during the large-scale *E. coli* O104:H4 outbreak that occurred in Germany in 2011, causing 3,816 cases and 54 deaths (Frank *et al.*, 2011). The first investigation of its kind, WGS data were released to the scientific community for crowd-sourced analyses. This led to the rapid identification of an enteroaggregative *E. coli* lineage

that, unusually, had acquired the gene encoding Shiga toxin 2, in addition to plasmid-borne antibiotic resistance genes (Rohde *et al.*, 2011). Likewise, WGS has been used to resolve single-organism outbreaks of milk- and waterborne *C. jejuni* (Revez *et al.*, 2014a, 2014b), healthcare-associated *C. difficile* and *Salmonella enterica* serovar Enteritidis (Eyre *et al.*, 2012; Quick *et al.*, 2015), food-borne *S. enterica* (Angelo *et al.*, 2015; Ashton *et al.*, 2015b; Hoffmann *et al.*, 2015), and *Vibrio cholerae* following the 2010 earthquake in Haiti (Mutreja *et al.*, 2011; Hendriksen *et al.*, 2011), as well as an uncommon multi-pathogen outbreak that included *Salmonella*, *Shigella*, and *E. coli* (Dallman *et al.*, 2014). These retrospective outbreak investigations have provided proof-of-concept, and the field is now moving towards using WGS in real-time.

WGS has also been used successfully for studies of long-term epidemiology: extended surveillance of *Campylobacter* and *E. coli* O157 at regional and national levels facilitated the detection of cryptic outbreaks in the UK (Cody *et al.*, 2013; Dallman *et al.*, 2015); an extended global outbreak of *S*. Typhimurium previously thought to be of zoonotic origin was found to represent two distinct outbreaks in humans and other animals (Mather *et al.*, 2013); and a large-scale cross-sectional study of global *S. flexneri* isolates obtained over a 19-year period provided important insights into transmission among men who have sex with men, highlighting the potential for sexual transmission of gastrointestinal pathogens (Baker *et al.*, 2015).

WGS has markedly improved our understanding of bacterial evolution, particularly with respect to the emergence and spread of pathogenic clones (McAdam *et al.*,

2014). A unifying theme across many studies is pathogen evolution in response to human behaviour. An exemplar of this is the change in incidence of *Shigella sonnei* and *S. flexneri* infections associated with industrialisation. Although improvements in water quality reduce the incidence of *S. flexneri*, it is thought that concomitant reduced exposure to *Plesiomonas shigelloides* results in increased incidence of *S. sonnei* due to loss of cross-protective immunity (Holt *et al.*, 2012; Connor *et al.*, 2015). As illustrated by the intercontinental dissemination of multidrug-resistant *S*. Typhi (Wong *et al.*, 2015), the emergence of antibiotic-resistant bacteria has become a global public health problem. Antibiotic resistance, in conjunction with increased HIV prevalence, has also been implicated in driving an epidemic of highly invasive non-typhoidal *Salmonella* in sub-Saharan Africa (Okoro *et al.*, 2012). Like antibiotics, agriculture is an innovation that has become essential for human survival, but it has created a new ecological niche, resulting in the emergence of a pathogenic agriculture-associated *Campylobacter coli* lineage (Sheppard *et al.*, 2013b). Finally, the role of human migration in the spread of infectious diseases is well-established (Wilson, 1995), and gastrointestinal pathogens are no exception; for instance, cholera was introduced into Haiti in 2010 by Nepalese United Nations peacekeepers (Hendriksen *et al.*, 2011), while World Wars I and II have been implicated in the dissemination of *Shigella dysenteriae* type 1 (Rohmer *et al.*, 2014).

The studies described in this preface demonstrate the potential of NGS for improving our understanding of the basic biology, pathogenesis, epidemiology, population genetics, and evolution of bacterial gastrointestinal pathogens; however, the preface also highlights the fact that most studies have focused on organisms that commonly cause severe disease, with studies on *Campylobacter* accounting for a smaller

proportion of the literature. With this in mind, this thesis explored the utility of WGS for *Campylobacter* diagnostics and epidemiology, both in a general context and, more specifically, in South Africa.

CHAPTER 1

Introduction

Campylobacter species occupy a wide range of ecological niches, colonising mammals, birds, shellfish, and reptiles (Man, 2011; Fitzgerald, 2015); however, the genus has risen to prominence because of the role a subset of species play as foodborne human pathogens (Kaakoush *et al.*, 2015). The incidence of *Campylobacter* infections in high-income countries has continued to rise over the past decade: campylobacteriosis was identified as the third leading cause of 9.4 million episodes of foodborne illness in the USA (Scallan *et al.*, 2011) and the most commonly-reported zoonosis in the European Union (European Food Safety Authority and European Centre for Disease Prevention and Control, 2014), while cases in the UK have been estimated to outnumber those caused by *Salmonella*, *Clostridium perfringens*, *E. coli* O157, and *Listeria monocytogenes* combined (Food Standards Agency, 2011; Tam *et al.*, 2012). In contrast, the incidence of campylobacteriosis in LMICs is not well defined, but is thought to be endemic and likely underappreciated (Kaakoush *et al.*, 2015). This introduction provides further information regarding key topics germane to the thesis, including *Campylobacter* taxonomy, pathology, diagnostics, and epidemiology. At the end of the introduction, the aims of this thesis, and the individual studies that comprise it, are described.

1.1 *Campylobacter* taxonomy, ecology, and microbiology

The genus *Campylobacter* belongs to the ecologically and physiologically diverse *Epsilonproteobacteria* (Gilbreath *et al.*, 2011), and was created in 1963 when it

became apparent that a subset of *Vibrio* species were sufficiently distinct from other members of the genus to warrant reassignment (Sebald & Veron, 1963). To date, 27 *Campylobacter* species, two provisional species, and nine subspecies have been described, which infect a variety of hosts (Table 1.1) (Fitzgerald, 2015; Kaakoush *et al.*, 2015). The major pathogens *Campylobacter jejuni* subsp. *jejuni* (referred to as *C. jejuni* from here on) and *C. coli*, in addition to the emerging pathogen *Campylobacter lari*, have broad host ranges, and have been isolated from agricultural, companion, and wild animals (Table 1.1). In contrast, the pathogen *Campylobacter jejuni* subsp. *doylei* has only been isolated from humans. Other species not typically associated with human disease, including *Campylobacter avium*, *Campylobacter canadensis*, *Campylobacter corcagiensis*, *Campylobacter cuniculorum*, *Campylobacter iguaniorum*, *Campylobacter peloridis*, *Campylobacter subantarcticus*, *Campylobacter troglodytis*, and *Campylobacter volucris*, appear to have narrow host ranges, many of which do not include common food animals (Table 1.1 and references therein). Given that most studies have focused on isolates obtained from humans, farm and companion animals, and agricultural sources, it is likely that broader sampling will continue to uncover additional species diversity and refine our understanding of host specificity among *Campylobacter* species.

Table 1.1 Members of the *Campylobacter* genus and their known sources of isolation

Species	Source																Original reference
	Poultry	Cattle	Pigs	Sheep	Horses	Cats	Dogs	Reptiles	Rabbits	Rodents	Non-human primates	Wild birds	Marine birds	Marine mammals	Shellfish	Environmental waters	
C. avium	■																(Rossi *et al.*, 2009)
C. canadensis												■					(Inglis *et al.*, 2007)
C. coli	■	■	■	■			■				■	■	■			■	(Veron & Chatelain, 1973)
C. concisus						■	■										(Tanner *et al.*, 1981)
C. corcagiensis											■						(Koziel *et al.*, 2014a)
C. cuniculorum									■								(Zanoni *et al.*, 2009)
C. curvus							■										(Vandamme *et al.*, 1991)
C. fetus subsp. *fetus*		■		■	■			■									(Veron & Chatelain, 1973)
C. fetus subsp. *testudinum*								■									(Fitzgerald *et al.*, 2014)
C. fetus subsp. *venerealis*		■		■													(Veron & Chatelain, 1973)
C. gracilis							■										(Vandamme *et al.*, 1995)
C. helveticus						■	■										(Stanley *et al.*, 1992)
C. hominis	Only isolated from humans																(Lawson *et al.*, 2001)
C. hyointestinalis subsp. *hyointestinalis*		■	■	■			■										(On *et al.*, 1995)
C. hyointestinalis subsp. *lawsonii*			■														(On *et al.*, 1995)
C. iguaniorum								■									(Gilbert *et al.*, 2015)
C. insulaenigrae														■			(Foster *et al.*, 2004)
C. jejuni subsp. *doylei*	Only isolated from humans																(Steele & Owen, 1988)
C. jejuni subsp. *jejuni*	■	■	■	■			■				■	■		■			(Veron & Chatelain, 1973)
C. lanienae		■	■	■													(Logan *et al.*, 2000)
C. lari subsp. *concheus*															■		(Debruyne *et al.*, 2009)
C. lari subsp. *lari*	■	■		■	■	■	■				■	■	■	■	■	■	(Debruyne *et al.*, 2009)

Table 1.1 cont.

Species	Poultry	Cattle	Pigs	Sheep	Horses	Cats	Dogs	Reptiles	Rabbits	Rodents	Non-human primates	Wild birds	Marine birds	Marine mammals	Shellfish	Environmental waters	Original reference
C. peloridis															■		(Debruyne *et al.*, 2009)
C. rectus							■								■		(Vandamme *et al.*, 1991)
C. showae							■										(Etoh *et al.*, 1993)
C. sputorum biovar *faecalis*		■		■													(Vandamme & On, 2001)
C. sputorum biovar *paraureolyticus*		■															(Vandamme & On, 2001)
C. sputorum biovar *sputorum*		■	■	■			■										(Vandamme & On, 2001)
C. subantarcticus													■				(Debruyne *et al.*, 2010a)
C. troglodytis											■						(Kaur *et al.*, 2011)
C. upsaliensis						■	■										(Sandstedt & Ursing, 1991)
C. ureolyticus					■												(Vandamme *et al.*, 2010)
C. volucris													■				(Debruyne *et al.*, 2010b)
"Campylobacter sp. Dolphin DP*"*														■			(Goldman *et al.*, 2011)
"Campylobacter sp. Prairie Dog*"*										■							(Beisele *et al.*, 2011)

Shaded blocks indicate known sources for each species. Adapted from Man (2011) and Fitzgerald (2015)

In spite of differences in their ecology and propensity to cause disease, *Campylobacter* species share several defining physiological and metabolic properties. Members of the genus are small, ranging from 0.2 to 0.8 μm by 0.5 to 5 μm, and are spiral, curved, or rod-shaped, but may become coccoid or spherical in old cultures. *Campylobacter* are typically motile, and the corkscrew motion characteristic of these species is due to the presence of a single flagellum at either or both ends of the bacterium, with the exception of *Campylobacter showae*, which has multiple flagella. The thermotolerant *Campylobacter* species, *C. jejuni*, *C. coli*, *C. lari*, and *Campylobacter upsaliensis*, grow optimally at 42°C, which corresponds to the temperature of the avian gut, while the remaining species grow best at 37°C. Most *Campylobacter* are microaerophilic, requiring oxygen for use as a terminal electron acceptor during respiration, while others thrive under anaerobic growth conditions. *Campylobacter* are nutritionally fastidious chemoorganotrophs, and they obtain energy from amino acids or intermediates of the tricarboxylic acid cycle, rather than carbohydrates (Vandamme *et al.*, 2005; Debruyne *et al.*, 2008).

1.2 Human campylobacteriosis

All but ten of the known *Campylobacter* species have been associated with human disease (Kaakoush *et al.*, 2015), and infections are predominantly caused by the consumption of contaminated meat, particularly chicken, milk, or water, or by environmental contact (Figure 1.1) (Kapperud *et al.*, 2003; Friedman *et al.*, 2004; Gormley *et al.*, 2011; Mughini-Gras *et al.*, 2012). *Campylobacter* infections typically result in acute gastroenteritis, but, in some cases, may also cause severe

extra-intestinal disease, or lasting neurological or gastrointestinal complications (Figure 1.1) (Kaakoush *et al.*, 2015).

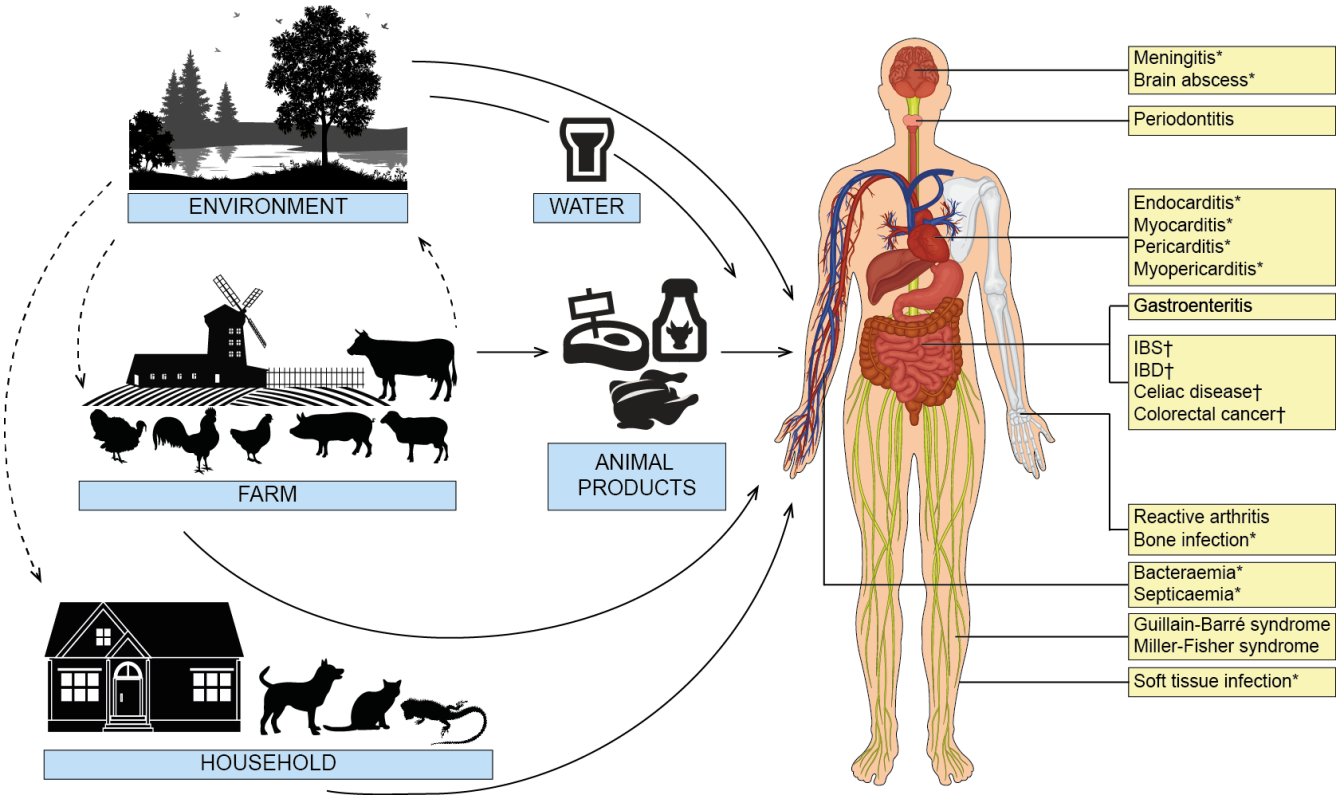


Figure 1.1 Sources and clinical manifestations of *Campylobacter* infections in humans. The left-hand side of the figure illustrates major sources and routes of *Campylobacter* transmission. Routes of transmission that remain incompletely understood are indicated with a dashed line. The right-hand side of the figure shows the range of clinical manifestations of *Campylobacter* infections reported in humans. *, rare complicated infections. †, conditions for which a disease association has been reported, but the precise role of *Campylobacter* is yet to be determined. IBS, irritable bowel syndrome; IBD, inflammatory bowel disease. Adapted from Young *et al*. (2007) and Kaakoush *et al*. (2015).

1.2.1 Gastroenteritis

Most *Campylobacter* infections manifest as self-limiting gastroenteritis and do not cause lasting complications (Butzler, 2004). Globally, *C. jejuni* is the predominant cause of *Campylobacter* gastroenteritis, followed by *C. coli* (Kaakoush *et al.*, 2015). Individuals with gastroenteritis due to *C. jejuni* or *C. coli* typically experience diarrhoea, abdominal cramping, and fever, while a smaller proportion report bloody diarrhoea and vomiting (Gillespie *et al.*, 2006). Approximately 10% of patients require hospitalisation (Gillespie *et al.*, 2002), and disease is often more severe in very young or immunocompromised individuals, or in cases where the infectious dose was large (Crushell *et al.*, 2004; Gillespie *et al.*, 2006). *Campylobacter concisus*, *C. lari*, *Campylobacter ureolyticus*, and *C. upsaliensis*, known collectively as the "emerging *Campylobacter* species", have also frequently been isolated from patients with gastroenteritis (Kaakoush *et al.*, 2015). Additional research is required to determine the precise role of the emerging and other non-*jejuni/coli Campylobacter* species associated with gastroenteritis (Table 1.2).

Table 1.2 Infections associated with *Campylobacter* species

Species	Clinical manifestations								
	Gastroenteritis	Meningitis	Bacteraemia or septicaemia	Abscesses	Soft tissue infections	Bone infections	Periodontal disease	Neurological sequelae	Gastrointestinal complications
C. coli	■	■	■						■
C. concisus	■		■	■			■		■
C. curvus	■		■	■					■
C. fetus subsp. *fetus*	■								
C. fetus subsp. *testudinum*	■				■				
C. fetus subsp. *venerealis*			■						
C. gracilis				■			■		■
C. helveticus	■								
C. hominis									■
C. hyointestinalis subsp. *hyointestinalis*	■		■						
C. insulaenigrae	■		■						
C. jejuni subsp. *doylei*	■		■						
C. jejuni subsp. *jejuni*	■	■	■					■	■
C. lari subsp. *lari*	■		■						
C. lari subsp. *concheus*	■								
C. mucosalis	■								
C. rectus	■		■	■	■		■		■
C. showae			■	■			■		■
C. sputorum biovar *sputorum*	■			■					
C. sputorum biovar *paraureolyticus*	■								
C. upsaliensis	■		■	■					
C. ureolyticus	■		■	■		■			■

Shaded blocks indicate known clinical manifestations. Adapted from Man (2011), Fitzgerald (2015), and Kaakoush *et al*. (2015).

1.2.2 Complicated infections

Although rare, complicated campylobacteriosis cases, including extra-intestinal infections and long-lasting sequelae, are major contributors to the overall disease burden caused by *Campylobacter* (Molbak & Havelaar, 2008). Extra-intestinal infections may manifest as local or systemic infections, and occur more frequently in those who are immunocompromised, pregnant, or elderly (Blaser & Engberg, 2008;

Lastovica & Allos, 2008). As recently reviewed in detail by Kaakoush *et al.* (2015), there have been case reports of *Campylobacter*-associated bacteraemia, septicaemia, meningitis, spontaneous abortion, neonatal sepsis, abscesses, soft-tissue infections, cardiovascular complications, and periodontal disease caused by a variety of species (Table 1.2).

A small proportion of patients with *Campylobacter* gastroenteritis develop long-lasting post-infectious complications, such as Guillain-Barré syndrome (GBS) (0.07%), reactive arthritis (2.86%), and IBS (4.01%) (Table 1.2) (Keithlin *et al.*, 2014). Although rarely quantified, estimates from individual countries suggest that the public health and economic impact of sequelae associated with *Campylobacter* infections is non-trivial: in the USA, 16,610 DALYs were attributed to post-infectious complications, equating to 74.1% of the total annual *Campylobacter* disease burden (Scallan *et al.*, 2015), while sequelae accounted for 51.5% of the €76 million annual cost of *Campylobacter* infections in the Netherlands (Mangen *et al.*, 2015). GBS is the most severe post-infectious complication, as it is characterised by acute flaccid paralysis, while patients with Miller Fisher syndrome (MFS), a clinical variant of GBS, experience opthalmoplegia and cerebellar-like ataxia (Wakerley & Yuki, 2015). Both GBS and MFS are associated with preceding *C. jejuni* infection; nerve damage is thought to arise due to antibody cross-reactivity, as *C. jejuni* lipooligosaccharides (LOS) mimic human gangliosides (Chiba *et al.*, 1992; Yuki, 1993; Yuki *et al.*, 2004; Koga *et al.*, 2005). While immune disorders are well-established lasting consequences of campylobacteriosis, there is also growing evidence of an association between *Campylobacter* infection and chronic gastrointestinal

conditions, including IBD, IBS, celiac disease, oesophageal disease, and colorectal cancer (Riddle *et al.*, 2012; Kaakoush *et al.*, 2015).

1.2.3 Asymptomatic carriage

Asymptomatic carriage of *Campylobacter* has been reported in experimental and clinical studies. In an experimental challenge study, 84% of healthy volunteers who ingested *C. jejuni* became infected, but only 33% of infected participants were symptomatic (Black *et al.*, 1988). It has been suggested that asymptomatic carriage is particularly prevalent in LMICs (Coker *et al.*, 2002). The results of studies carried out in Africa and South America indicate that the rates of asymptomatic *Campylobacter* infection in these countries are variable, ranging from 2.2 to 20.4% (Georges-Courbot *et al.*, 1987; Figueroa *et al.*, 1989; Samie *et al.*, 2007a; Fernández *et al.*, 2008; Samie *et al.*, 2009b; da Silva Quetz *et al.*, 2010; Mason *et al.*, 2013; Lee *et al.*, 2013; Komba *et al.*, 2015). Fernández *et al.* (2008) found that asymptomatic *Campylobacter* carriage was more frequent among malnourished (31.4%) than well-nourished children (9.1%), suggesting that nutritional status may play a role in infection. A recent study carried out in Peru demonstrated that children with asymptomatic *Campylobacter* infection experience growth stunting, which suggests that, contrary to previous assumptions, carriage may have long-lasting adverse effects (Lee *et al.*, 2013).

1.2.4 Mechanisms of pathogenesis

Our understanding of *Campylobacter* pathogenesis remains relatively limited, and much of our knowledge stems from studies of *C. jejuni* (van Putten *et al.*, 2009). Mechanisms of pathogenesis first characterised in *C. jejuni* have also been identified in other *Campylobacter* species (Man, 2011); however, for the sake of brevity, only *C. jejuni* pathogenesis, which is broadly applicable to the genus, will be described.

As is true of any bacterial pathogen, *C. jejuni* produces a variety of factors involved in adhesion, invasion, host cell subversion, and immune evasion, which collectively enable it to establish gastrointestinal infections and cause disease in humans (van Putten *et al.*, 2009). On arrival in the human gut, *C. jejuni* exhibits a strong chemotactic attraction towards mucin (Hugdahl *et al.*, 1988), and its spiral morphology, flagellar motility, and short LOS *O*-sidechain facilitate movement through the mucus layer to the intestinal epithelium below (McSweegan & Walker, 1986).

The mechanism of *C. jejuni* translocation across the intestinal epithelium is still poorly understood, as are the bacterial and host factors involved, but transcellular and paracellular models of migration have been proposed (Figure 1.2) (Backert *et al.*, 2013). Transcellular migration starts with bacterial adhesion to epithelial cells, followed by invasion with limited intracellular replication, and then exit at the basal membrane (van Putten *et al.*, 2009). Adhesion to epithelial cells is mediated by CadF (Konkel *et al.*, 1997; Monteville & Konkel, 2002; Monteville *et al.*, 2003), JlpA (Jin *et al.*, 2001), and Peb1-4 membrane proteins (Pei & Blaser, 1993; Kervella *et al.*, 1993;

Pei *et al.*, 1998; Asakura *et al.*, 2007), while bacterial cell structures, such as the flagellum, capsule, and LOS, may also play a role in adherence (McSweegan & Walker, 1986; Jones *et al.*, 2004; Bachtiar *et al.*, 2007). Epithelial cell invasion is dependent on microtubule polymerisation, and, in some cases, microfilament polymerisation (Oelschlaeger *et al.*, 1993; Hu & Kopecko, 1999; Biswas *et al.*, 2000, 2003; Monteville *et al.*, 2003). Other factors such as the flagellum, capsule, LOS, and CiaB are also thought to play a role in invasion, although this is an ongoing area of research (van Putten *et al.*, 2009). The paracellular route of migration is also incompletely understood, but involves the disruption of tight junctions between polarised epithelial cells, permitting *C. jejuni* access to the subepithelial tissue (Backert *et al.*, 2013).

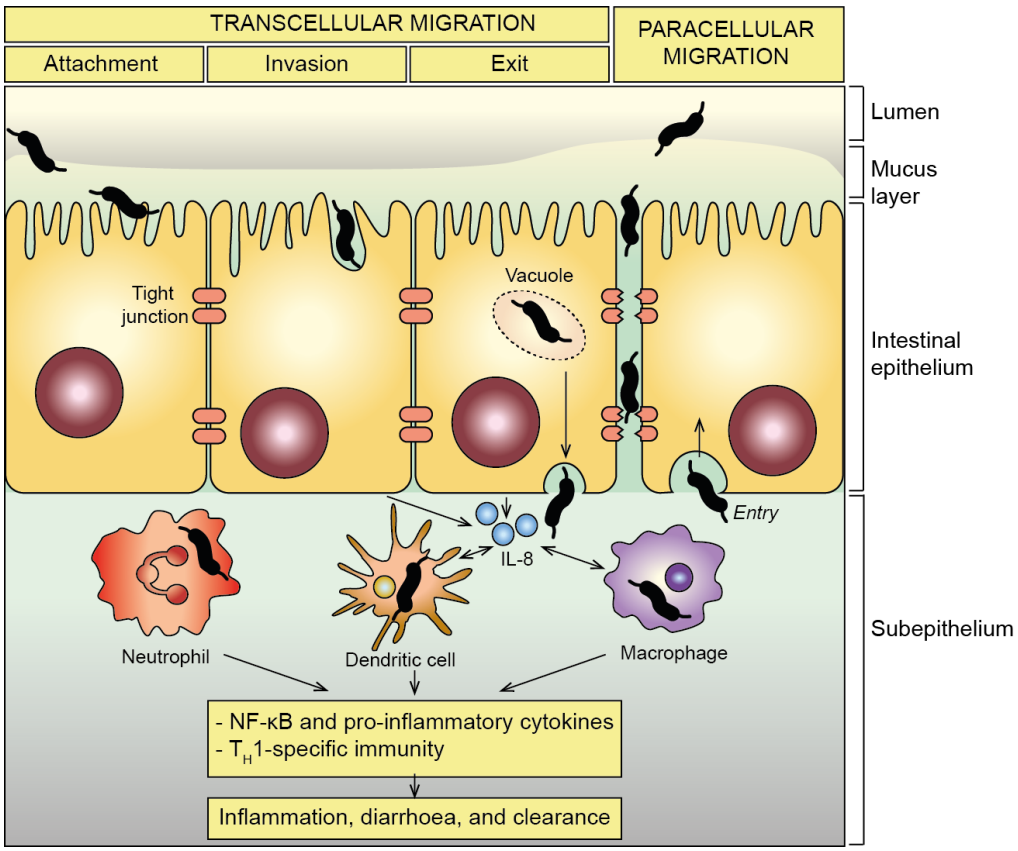


Figure 1.2 Transcellular and paracellular routes of *Campylobacter* migration across the intestinal epithelium during acute gastroenteritis in humans. Adapted from Young *et al*. (2007) and Backert *et al*. (2013).

Tissue damage occurs during acute infection as a result of the action of the only known *C. jejuni* toxin, cytolethal distending toxin (CDT), and the host's immune response (Young *et al.*, 2007). After host-cell uptake of CDT, the CdtB subunit translocates to the nucleus, where it causes double-stranded breaks in DNA, resulting in cytodistension and cell cycle arrest in G_2/M phase (Elwell & Dreyfus, 2000; Lara-Tejero & Galán, 2000, 2001). Furthermore, CDT is immunomodulatory, as it mediates the production of interleukin 8 (IL-8), which may also be produced by CDT-independent mechanisms during *C. jejuni* infection (Hickey *et al.*, 2000). IL-8 production leads to the recruitment of dendritic cells, macrophages, and neutrophils to the site of infection, resulting in a pro-inflammatory response and, ultimately, clearance of infection (Figure 1.2). An adaptive immune response also occurs during infection, with antibodies raised to flagella, LOS, outer membrane proteins, and CDT (Young *et al.*, 2007).

1.2.5 Clinical treatment

As most *Campylobacter* infections are self-limiting, only fluid and electrolyte replacement are required; however, antibiotic treatment is recommended for immunocompromised individuals, and for those with unusually severe gastroenteritis or complicated infections (Butzler, 2004). Fluoroquinolones and macrolides are the antimicrobial agents most commonly prescribed for treatment of severe *Campylobacter*-associated gastroenteritis, while intravenous aminoglycosides are recommended for systemic infections (Allos, 2001; Blaser & Engberg, 2008). As is true of other bacterial pathogens, the use of antibiotics in human medicine and animal husbandry has driven the emergence of antibiotic-resistant *Campylobacter*,

which represent a public health concern due to the increasing potential for treatment failure (Moore *et al.*, 2006; Luangtongkum *et al.*, 2009).

1.3 Laboratory diagnosis of *Campylobacter* infections

1.3.1 Conventional methods

Diagnosis of campylobacteriosis cases remains difficult because the isolation of these organisms is technically challenging (Fitzgerald, 2015), as indicated by the almost 90 year delay between the first description of *Campylobacter* (Engberg, 2006) and the development of suitable culture methods (Dekeyser *et al.*, 1972; Butzler *et al.*, 1973). To recover *Campylobacter* from stool specimens where the bacterial load is likely to be high, samples are most commonly plated directly onto complex media containing antibiotics that prevent the growth of other enteric bacteria (polymyxin B, trimethoprim, rifampicin, vancomycin, cephalothin, and cefoperazone), as well as yeast and fungal contaminants (amphotericin B) (Corry *et al.*, 1995; Fitzgerald *et al.*, 2008). For samples likely to have a low bacterial load, enrichment in selective broth is recommended prior to plating on solid media (Corry *et al.*, 1995; Kim *et al.*, 2009; Kaakoush *et al.*, 2011). Once inoculated, plates are incubated in a microaerobic atmosphere, with 5% oxygen, 10% carbon dioxide, and 85% nitrogen, at 42°C for two to seven days (Fitzgerald *et al.*, 2008).

Although most *C. jejuni* and *C. coli* strains grow well under routine culture conditions, other *Campylobacter* species are inhibited: several non-*jejuni*/*coli* *Campylobacter* species are susceptible to antibiotics added to selective media; species such as

C. concisus, *Campylobacter curvus*, *Campylobacter gracilis*, *Campylobacter mucosalis*, *Campylobacter rectus*, and *C. showae* require a hydrogen-enriched microaerobic atmosphere for growth; and growth of all species except the thermotolerant *Campylobacter* is inhibited at 42°C (Fitzgerald *et al.*, 2008; Lastovica & Allos, 2008). As a result, the disease burden caused by *Campylobacter* species other than *C. jejuni* and *C. coli* is underestimated.

Several filtration protocols have been developed to improve the recovery of non-*jejuni*/*coli Campylobacter* species (Butzler *et al.*, 1973; Goossens *et al.*, 1986; Engberg *et al.*, 2000; Lastovica, 2006). Briefly, a filter is placed on top of a blood agar plate with or without antibiotics and inoculated with a suspension of the clinical specimen. As *Campylobacter* are small and motile, they are able to pass through the small pores (0.45-0.65 μm) of the membrane (Man, 2011). While these methods have improved the isolation of non-*jejuni*/*coli Campylobacter* species (Engberg *et al.*, 2000; Lastovica & le Roux, 2000; Lastovica, 2006), non-motile species, such as *C. gracilis*, would not be recovered, highlighting the fact that there is no single protocol suitable for the recovery of all *Campylobacter* species (Lastovica & le Roux, 2000; Fitzgerald, 2015).

For the purposes of patient management and epidemiological surveillance, it is desirable to identify bacterial isolates to at least the species level. Until recently, diagnostic laboratories have been reliant on phenotypic testing for the identification of *Campylobacter* species (On, 2013). Biochemical tests used for *Campylobacter* speciation include: catalase, oxidase, and H_2S production; hippurate and indoxyl

acetate hydrolysis; and nitrate reduction. Other phenotypic tests, such as growth at different temperatures and in the presence of antibiotics, have also been used to distinguish *Campylobacter* species (Lastovica & Allos, 2008). As a consequence of their complex taxonomy and relative biochemical inertness, phenotypic testing of *Campylobacter* species is challenging, and, due to intra-species variability, results are not unequivocal (Lastovica & Allos, 2008; On, 2013).

1.3.2 Molecular diagnostics

Difficulties associated with the isolation and identification of *Campylobacter* have driven the development of molecular diagnostics, which fall into three broad categories, namely antigen-, protein-, and nucleic acid-based methods (Fitzgerald, 2015). Antigen-based methods of pathogen detection are rapid, simple, and culture-independent (Zhao *et al.*, 2014). Based on the results of evaluation studies, immunoassays were proposed as alternatives to culture for the detection of *C. jejuni* and *C. coli* (Tribble *et al.*, 2008; Granato *et al.*, 2010; Regnath & Ignatius, 2014), but there have been subsequent reports of low positive predictive values (36-80.6%) due to a lack of specificity (Tissari & Rautelin, 2007; Myers *et al.*, 2011; Floch *et al.*, 2012; Giltner *et al.*, 2013), as well as cross-reactivity with non-*jejuni*/*coli* *Campylobacter* species (Hindiyeh *et al.*, 2000; Couturier *et al.*, 2013; Platts-Mills *et al.*, 2014). Therefore, it has been recommended that guidelines be established for *Campylobacter* diagnostics and that immunoassay results be confirmed by culture (M'ikanatha *et al.*, 2012; Giltner *et al.*, 2013).

Concerning protein-based methods for bacterial diagnostics, matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) has recently been adapted for the identification of bacteria (Seng *et al.*, 2009). MALDI-TOF MS entails comparison of the peptide spectrum obtained from the bacterial cells of interest to those present in a reference database; the technique is becoming increasingly widely used because of its specificity and sensitivity, ease of use, and rapid 90 minute turnaround time (Seng *et al.*, 2009; Gaillot *et al.*, 2011; Tan *et al.*, 2012; Guo *et al.*, 2014). Additionally, although initial expenditure is high, MALDI-TOF reduces laboratory costs by approximately 70-90% when compared to standard testing methods (Seng *et al.*, 2009; Gaillot *et al.*, 2011; Tan *et al.*, 2012). MALDI-TOF has been validated for the identification of *C. jejuni*, *C. coli*, *C. curvus*, *Campylobacter fetus*, *Campylobacter helveticus*, *C. lari*, *Campylobacter sputorum*, and *C. upsaliensis*, but the method remains culture-dependent, which is a drawback in the case of *Campylobacter* (Mandrell *et al.*, 2005; Alispahic *et al.*, 2010; Bessède *et al.*, 2011; Martiny *et al.*, 2011).

Nucleic acid-based methods were first used to improve the speciation of *Campylobacter* isolates obtained from culture, but they have since been developed for culture-independent diagnostics (Fitzgerald, 2015; Zhang *et al.*, 2015). Numerous in-house and commercial nucleic acid amplification tests (NAATs), including single- and multiplex conventional and real-time PCR assays, have been developed for the identification of *Campylobacter* species (Best *et al.*, 2003; Fukushima *et al.*, 2003; On & Jordan, 2003; Schuurman *et al.*, 2007; Bullman *et al.*, 2012; Buchan *et al.*, 2013; de Boer *et al.*, 2013; Koziel *et al.*, 2013; Mengelle *et al.*, 2013; On *et al.*, 2013; Anderson *et al.*, 2014; Liu *et al.*, 2014; Perry *et al.*, 2014; Buss

et al., 2015; Van Lint *et al.*, 2015). The sensitivity and specificity of PCR-based NAATs is variable, and evaluation studies have also highlighted the need for assay re-validation on discovery of novel species (On & Jordan, 2003; On *et al.*, 2013). The importance of NAATs and assay evaluation in the post-genomic era are discussed further in Chapter 2.

Most PCR-based speciation assays detect *C. jejuni*, *C. coli*, and, in some cases, a limited number of other *Campylobacter* species (On & Jordan, 2003; Yamazaki-Matsune *et al.*, 2007; Bullman *et al.*, 2012; de Boer *et al.*, 2013); given the complex taxonomy of the genus, identification of all *Campylobacter* species in a single assay is not feasible. Instead, single target sequencing of loci such as 16S rRNA gene sequences (Lawson *et al.*, 1997; Kulkarni, 2002; Gorkiewicz *et al.*, 2003), *gyrB* (Gunther *et al.*, 2011), and *atpA* (Miller *et al.*, 2014b), has been used to detect a wider range of *Campylobacter* species. Sequence-based speciation of *Campylobacter* is described in more detail in Chapter 3.

1.4 The descriptive epidemiology of *Campylobacter*

For any pathogen, a thorough knowledge of its epidemiology is essential for the development of targeted interventions designed to reduce disease incidence. Much of our understanding of the descriptive epidemiology of *Campylobacter* is based on data from high-income countries where disease is largely sporadic. While outbreaks are not uncommon in these regions, they remain difficult to detect, and are typically caused by contaminated water or food, particularly poultry and dairy (Nichols *et al.*, 2012; Taylor *et al.*, 2013). Although Europe, North America, and Australia have

reported an increase in human campylobacteriosis over the past decade, the incidence of disease is highly variable (Kaakoush *et al.*, 2015). This is best illustrated by surveillance data from Europe, which indicated an average disease incidence of 64.8 cases per 100,000 population in 2013, with individual member state incidences ranging from 104 to 173.7 per 100,000 in the Czech Republic, Luxembourg, Slovakia, and the UK to < 2.0 per 100,000 in Bulgaria, Poland, Latvia, and Romania (European Food Safety Authority and European Centre for Disease Prevention and Control & European Centre for Disease Prevention and Control, 2015). However, differences in laboratory protocols, reporting, and healthcare systems make comparisons between countries difficult. In contrast to high-income countries, campylobacteriosis is thought to be endemic in LMICs, with higher rates of symptomatic and asymptomatic infections recorded in these regions (Oberhelman & Taylor, 2000; Platts-Mills & Kosek, 2014). This has been attributed to poor sanitation and nutrition, a higher prevalence of HIV, and people living in close proximity to animals (Coker *et al.*, 2002; Crushell *et al.*, 2004; Pfeiffer *et al.*, 2012).

Regardless of income status, most countries report a peak in disease incidence among children under the age of five, but LMICs do not report secondary peaks in older age groups, while high-income countries do (Coker *et al.*, 2002; Kaakoush *et al.*, 2015). It has been postulated that the lack of disease among adults in LMICs may be due to the development of protective immunity as a result of frequent exposure during early childhood (Rao *et al.*, 2001; Havelaar *et al.*, 2009). The summer peak in cases observed in countries with a temperate climate is also not detected in certain LMICs (Oberhelman & Taylor, 2000; Butzler, 2004; Humphrey *et al.*, 2007). Concerning the acquisition of *Campylobacter*, international travel,

consumption of poultry, contact with animals, and environmental exposure have been identified as important risk factors for disease (Kapperud *et al.*, 2003; Neimann *et al.*, 2003; Swaminathan *et al.*, 2009; Gautret *et al.*, 2012; Zinszer *et al.*, 2010; Ravel *et al.*, 2010; Domingues *et al.*, 2012; Mughini-Gras *et al.*, 2012; Mughini-Gras *et al.*, 2014; Ricotta *et al.*, 2014), while advances in molecular epidemiology have significantly improved our understanding of the sources of infections (Colles & Maiden, 2012), as discussed further in the following section [1.5.2.2.2].

1.5 Molecular typing of *Campylobacter*

Molecular typing methods, in conjunction with the improvements in laboratory diagnostics, have enhanced our understanding of the short- and long-term epidemiology of bacterial pathogens (Sabat *et al.*, 2013). Bacterial typing techniques can be separated into two broad categories, phenotypic and genotypic approaches, with the latter subdivided into band- and nucleotide sequence-based methods. Given their contribution to the overall campylobacteriosis burden, the majority of typing techniques have been developed for *C. jejuni* and *C. coli* (Ahmed *et al.*, 2012; Eberle & Kiess, 2012). Commonly used examples of phenotypic and genotypic techniques, and the impact of molecular typing on the field, are discussed here, with a focus on recent developments in genomic epidemiology.

1.5.1 Phenotypic techniques

Serotyping and phage typing were the first molecular typing schemes developed for *C. jejuni* and *C. coli* and were widely used during the 1980s and 1990s (Penner &

Hennessy, 1980; Lior *et al.*, 1982; Grajewski *et al.*, 1985; Patton *et al.*, 1991; Frost *et al.*, 1999; Wassenaar & Newell, 2000; Ahmed *et al.*, 2012; Eberle & Kiess, 2012). With reference to serotyping, the well-established Penner (Penner & Hennessy, 1980) and Lior (Lior *et al.*, 1982) schemes are based on the detection of heat-stable and heat-labile antigens, respectively, with the former subsequently found to correspond to components of the capsule (Karlyshev *et al.*, 2000). Serotyping and phage typing both suffer from low discriminatory power and a high level of non-typeability (Wassenaar & Newell, 2000). This is illustrated by two studies of over 2,400 *C. jejuni* and *C. coli* isolates from the UK: 53 and 60% of isolates corresponded to just ten serotypes and phage types, respectively, while 16-19% of isolates were non-typeable (Frost *et al.*, 1998, 1999). Furthermore, phenotypic typing methods are technically demanding, time consuming, and lack reproducibility, and reagents are also expensive and difficult to produce (Wassenaar & Newell, 2000; Ahmed *et al.*, 2012; Eberle & Kiess, 2012); therefore, phenotypic techniques have been superseded by genotypic typing (Ahmed *et al.*, 2012; Eberle & Kiess, 2012).

1.5.2 Genotypic techniques

1.5.2.1 Band-based typing

Following on from phenotypic typing techniques, a surfeit of band-based techniques emerged for bacterial typing. Band-based techniques are based on the comparison of banding patterns obtained after agarose gel electrophoresis of restriction fragments or PCR products (Sabat *et al.*, 2013). Methods developed for *Campylobacter* include pulsed-field gel electrophoresis (PFGE), amplified fragment

length polymorphism (AFLP), random amplification of polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), ribotyping, riboprinting, repetitive extragenic palindromic-PCR (Rep-PCR), and enterobacterial repetitive intergenic consensus PCR (ERIC-PCR) (On *et al.*, 2008). Among these techniques, PFGE, AFLP, and RAPD index genome-wide variation, while the remaining methods query particular subsets of the genome (Sabat *et al.*, 2013). PFGE, AFLP, and RAPD are highly discriminatory, and have been adapted for use in several *Campylobacter* species; however, RAPD has not been as widely used as PFGE and AFLP due to its poor intra- and inter-laboratory reproducibility (Wassenaar & Newell, 2000; On *et al.*, 2008; Ahmed *et al.*, 2012). For this reason, only PFGE and AFLP will be discussed further here.

1.5.2.1.1 Pulsed-field gel electrophoresis

PFGE has been described as the "gold standard" of molecular typing methods for several bacteria, including *Campylobacter* (Foley *et al.*, 2009). Briefly, PFGE entails the comparison of macro-restriction profiles obtained by using pulsed-field gel electrophoresis to separate high molecular weight fragments generated by restriction enzyme digestion of intact chromosomal DNA (Schwartz & Cantor, 1984). Since its adaptation for the characterisation of *C. jejuni*, *C. coli*, *C. fetus*, *Campylobacter hyointestinalis*, and *C. upsaliensis* (Wassenaar & Newell, 2000), PFGE has been used to support outbreak investigations (Suzuki *et al.*, 1994; Fitzgerald *et al.*, 2001; Hedberg *et al.*, 2001; Sails *et al.*, 2003; Kuusi, 2004; Mellmann *et al.*, 2004; Kuusi *et al.*, 2005), explore genotype diversity and persistence in humans and the environment (Ahmed *et al.*, 2012), and also to

examine relationships among isolates obtained from humans, broiler chickens, wild and domesticated animals, and retail meat (Hänninen *et al.*, 2000; Petersen *et al.*, 2001; Broman *et al.*, 2002; Nadeau *et al.*, 2002; Broman *et al.*, 2004; Parsons *et al.*, 2009; Zhang *et al.*, 2010b; Ku *et al.*, 2011; O'Leary *et al.*, 2011; Oporto *et al.*, 2011).

In spite of its power of discrimination and high level of typeability, PFGE has several drawbacks. Firstly, it is technically demanding, labour-intensive, time-consuming, low-throughput, and relatively subjective, although this has been improved by the introduction of software packages for data analysis (Sabat *et al.*, 2013). Secondly, studies have shown that *Campylobacter* PFGE genotypes are unstable: genetic changes, such as point mutations, insertions, deletions, horizontal genetic transfer, and genomic rearrangements, alter macro-restriction profiles (Hänninen *et al.*, 1999; Scott *et al.*, 2007). Finally, with the exception of PulseNet in the USA (Ribot *et al.*, 2001; Gerner-Smidt *et al.*, 2006), PFGE suffers from a lack of reproducibility and portability (Wassenaar & Newell, 2000; Sabat *et al.*, 2013).

1.5.2.1.2 Amplified fragment length polymorphism

Initially developed for fingerprinting plant genomes, AFLP combines genome-wide restriction enzyme digestion with PCR (Vos *et al.*, 1995). Genomic DNA is cleaved using two restriction enzymes, and adaptors are ligated to the resulting fragments. A subset of DNA fragments are amplified using selective primers complementary to the adaptor sequences, with a fluorescent marker incorporated to facilitate visualisation of PCR products following size-based separation (Vos *et al.*, 1995; Sabat *et al.*, 2013). AFLP was first adapted for *C. jejuni* and *C. coli* (Duim *et al.*, 1999; Desai *et al.*, 2001), but has since also been described for speciation and typing

of several non-*jejuni*/*coli* *Campylobacter* species (Duim *et al.*, 2001; Duim & Wagenaar, 2006). Like PFGE, AFLP has a high level of discrimination and typeability (Sabat *et al.*, 2013), and it has been used to explore relationships among *Campylobacter* isolates from different sources, in addition to investigate host and disease associations of particular genotypes (Duim *et al.*, 1999; Kokotovic & On, 1999; Duim *et al.*, 2000; Desai *et al.*, 2001; Hänninen *et al.*, 2001; Schouls *et al.*, 2003; Hopkins *et al.*, 2004; Siemer *et al.*, 2004; Wieland *et al.*, 2005; Koene *et al.*, 2009; Messens *et al.*, 2009; Gilbert *et al.*, 2014b). Although widely used, AFLP is labour intensive, suffers from similar problems to PFGE with respect to portability and subjectivity, and analysis of complex banding patterns can prove challenging (Ahmed *et al.*, 2012; Eberle & Kiess, 2012; Sabat *et al.*, 2013).

1.5.2.2 Sequence-based typing

While band-based typing techniques were unarguably an improvement over phenotypic methods, they were also found wanting; there was still a need for highly discriminatory techniques that met the performance and convenience criteria of an ideal typing system. Such a system should be applicable across the bacterial domain and able to differentiate closely related isolates, as well as detect distant genealogical relationships. On a practical level, it should be rapid, simple, scalable, reproducible, and have a high level of typeability. Interpretation should be objective, and the data generated, as well as protocols and reagents, should be portable and accessible. Most importantly, the method should be agreed on and widely adopted by the scientific community (van Belkum *et al.*, 2007; Maiden *et al.*, 2013). Although no widely used typing system currently meets all of these criteria, sequence-based techniques have come the closest (Maiden *et al.*, 2013). Several sequence-based

typing techniques have been developed for *Campylobacter*, including single locus, multilocus, and WGS-based methods (Taboada *et al.*, 2013).

1.5.2.2.1 Single locus sequence typing

The single locus methods most commonly used for *Campylobacter* sequence typing index variation in highly variable surface antigen genes, namely *flaA* and *flaB*, which encode flagellin, and *porA*, which encodes the major outer membrane protein. These techniques were developed for *C. jejuni* and *C. coli*, and entail PCR amplification of the ~320 bp short variable region (SVR) of *flaA* or *flaB* (Meinersmann *et al.*, 1997), or a ~630 bp internal fragment of *porA* (Dingle *et al.*, 2008), followed by direct sequencing. The resulting nucleotide sequences are then compared to variation catalogued in public databases, such as the *Campylobacter jejuni/coli* PubMLST database (http://pubmlst.org/campylobacter/) (Jolley & Maiden, 2010), and are assigned unique nucleotide and peptide allele numbers in order of discovery. At the time of writing, 1,654 *fla*-SVR nucleotide and 370 peptide alleles had been defined in the PubMLST database, and 2,065 and 1,887 *porA* nucleotide and peptide alleles, respectively.

Both *fla*-SVR and *porA* typing have been used to study the epidemiology of human campylobacteriosis, and, in combination with other typing methods, have been applied to investigate outbreaks (Meinersmann *et al.*, 1997; Dingle *et al.*, 2001b, 2002; Sails *et al.*, 2003; Mellmann *et al.*, 2004; Dingle *et al.*, 2005; Corcoran *et al.*, 2006; Dingle *et al.*, 2008; Cody *et al.*, 2009). However, a major drawback of single locus typing schemes is that relationships among isolates are easily distorted by

recombination. This is pertinent as *Campylobacter* are naturally competent (Wang & Taylor, 1990), with intra- and interspecies recombination well-established as drivers of genetic diversity, particularly among *C. jejuni* and *C. coli* (Sheppard *et al.*, 2008; Wilson *et al.*, 2009); studies have demonstrated *fla*-SVR and *porA* allele sharing among *C. jejuni* and *C. coli* (Dingle *et al.*, 2005; Cody *et al.*, 2009), while intraspecies and intragenomic recombination in the *fla* genes has also been observed, causing closely related isolates to appear distinct (Wassenaar *et al.*, 1995; Harrington *et al.*, 1997). As a result, *fla*-SVR and *porA* typing are best used in conjunction with methods that interrogate multiple genetically stable loci, such as multilocus sequence typing.

1.5.2.2.2 Multilocus sequence typing

1.5.2.2.2.1 The development of Campylobacter *multilocus sequence typing schemes*

Multilocus sequence typing (MLST) is a generic typing technique that was first developed for *Neisseria meningitidis* (Maiden *et al.*, 1998), and has since been adapted for over 100 taxa, predominantly corresponding to members of the bacterial domain (http://pubmlst.org/; http://www.mlst.net/; http://bigsdb.web.pasteur.fr/; http://mlst.warwick.ac.uk/mlst/). The method is based on the population genetics framework previously used by multilocus enzyme electrophoresis (MLEE), which assesses variation in genes that encode essential metabolic components, commonly known as housekeeping genes (Maiden, 2006). In MLEE, genotypic allelic variants are inferred based on differences in the electrophoretic mobility of isozymes of the metabolic enzymes of interest (Selander *et al.*, 1986). An extension of MLEE, MLST

indexes nucleotide diversity in housekeeping genes (Figure 1.3), which are under stabilising selection, thereby facilitating the study of distant genealogical relationships (Maiden, 2006). For MLST, internal fragments of 400-600 bp are amplified from 6-10 housekeeping genes by PCR and sequenced directly. The resulting nucleotide sequences and isolate provenance data are submitted to the appropriate online database (http://pubmlst.org/; http://www.mlst.net/; http://bigsdb.web.pasteur.fr/; http://mlst.warwick.ac.uk/mlst/), compared to existing sequences available for each locus, and assigned a unique allele number in order of discovery. Taken together, these numbers comprise an isolate's allelic profile or sequence type (ST), which is also summarised as a unique, arbitrary number (Figure 1.3) (Maiden *et al.*, 1998).

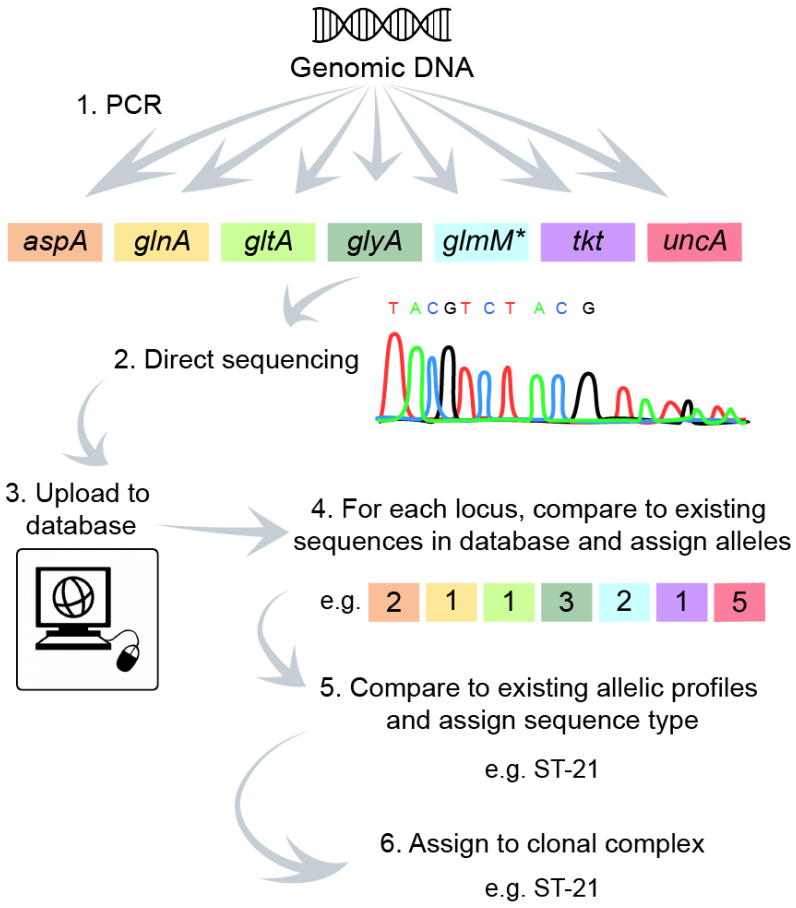


Figure 1.3 Multilocus sequence typing workflow. The MLST scheme developed for *C. jejuni* and *C. coli* (Dingle *et al.*, 2001a) is shown as an exemplar. *, *glmM* was thought to be *pgm* when the *C. jejuni* scheme was developed, but the name has not been changed in the database for consistency.

To describe relationships among isolates, STs can be grouped into clonal complexes (Figure 1.3), which are sets of genotypes likely to have descended from a common ancestor. In the case of *Campylobacter* MLST schemes, clonal complexes are defined informally by identifying a “central genotype” and its related genotypes. Central genotypes tend to occur at a high frequency, persist over time, and occupy a central position when STs from the population of interest are analysed using heuristic algorithms. Members of a particular clonal complex share four or more alleles with the central genotype. In addition to describing relationships among

isolates based on allele designations, STs, and assignment to clonal complexes, they can also be defined based on phylogenetic analyses of the underlying nucleotide sequences. The choice of analytical approach is dictated by the population structure of the organism of interest, with the former appropriate for studying weakly clonal populations with high levels of recombination, and the latter for clonal populations (Maiden, 2006; Colles & Maiden, 2012).

The first *Campylobacter* MLST scheme was developed for *C. jejuni*, including *C. jejuni* subsp. *doylei,* in 2001 (Dingle *et al.*, 2001a). The scheme was based on seven loci that are distributed across the *C. jejuni* genome, namely *aspA*, *glnA*, *gltA*, *glyA*, *pgm* (now known to be *glmM*), *tkt*, and *uncA* (also referred to as *atpA*) (Figure 1.4), and has since been applied to *C. coli* (Dingle *et al.*, 2005; Miller *et al.*, 2005). Although two other MLST schemes were developed for *C. jejuni* (Suerbaum *et al.*, 2001; Manning *et al.*, 2003), they have not been widely used (Colles & Maiden, 2012). As MLST is based on housekeeping genes which are under stabilising selection, it is well-suited to long-term epidemiological studies (Maiden, 2006). To increase the discriminatory power of MLST for studies of short-term *Campylobacter* epidemiology, the scheme may be supplemented with additional loci, typically antigenic genes, such as *flaA, flaB*, and *porA* (Dingle *et al.*, 2008) (Figure 1.3), which accumulate mutations more rapidly than housekeeping genes as they are under diversifying selection. As for *fla*-SVR and *porA*, *C. jejuni* and *C. coli* MLST data are submitted to the *Campylobacter jejuni/coli* PubMLST database (http://pubmlst.org/campylobacter/). The database has been widely used by the community and contains at least one representative of each ST, providing an exhaustive summary of known MLST diversity for these species. As of January 2016,

the *Campylobacter jejuni/coli* database contained 37,351 isolates, corresponding to 8,084 unique MLST profiles (Figure 1.5).

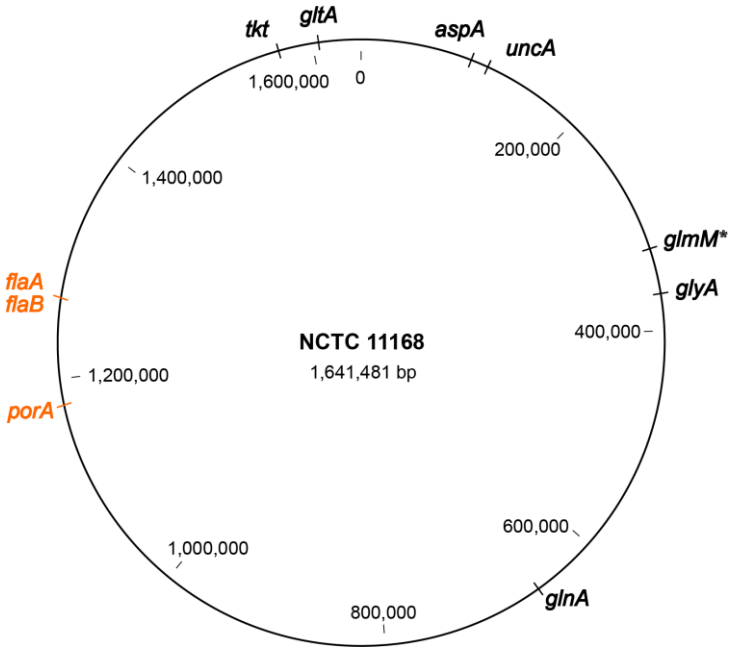


Figure 1.4 Distribution of multilocus sequence typing loci (black) and antigen genes (orange) across the *C. jejuni* NCTC 11168 genome. *, *glmM* was thought to be *pgm* when the *C. jejuni* scheme was developed, but the name has not been changed in the database for consistency.

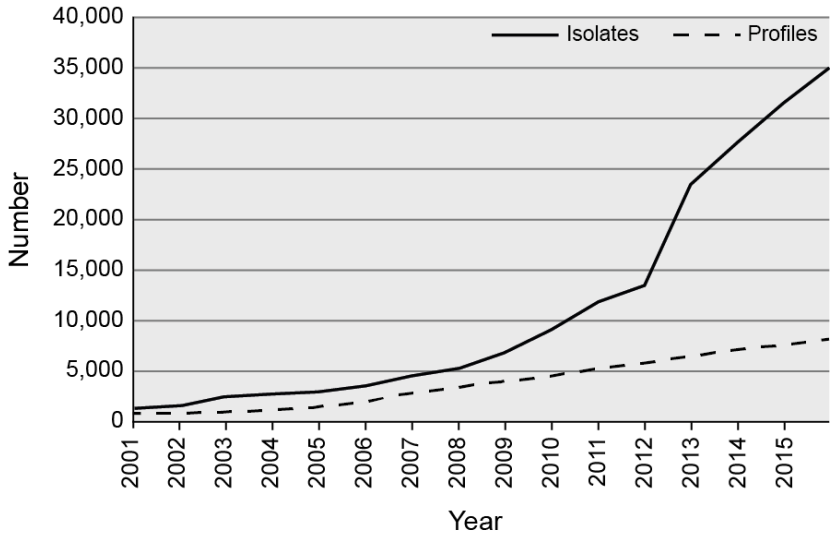


Figure 1.5 Growth of the *Campylobacter jejuni/coli* PubMLST database between 2001 and 2015. The cumulative number of isolates and allelic profiles (or STs) are indicated by the black and dashed lines, respectively. An up-to-date version of these data, as well as antigen typing data, can be obtained from: http://pubmlst.org/campylobacter/history.shtml.

Subsequent to the development of the *C. jejuni* and *C. coli* MLST scheme, additional schemes were introduced for *C. concisus*, *C. curvus*, *C. fetus* *C. helveticus*, *C. hyointestinalis*, *Campylobacter insulaenigrae*, *Campylobacter lanienae*, *C. lari*, *C. sputorum*, and *C. upsaliensis* (Miller *et al.*, 2005; van Bergen *et al.*, 2005; Stoddard *et al.*, 2007; Miller *et al.*, 2012). These schemes are also hosted on the PubMLST website, and can be found by accessing the non-*jejuni*/*coli* *Campylobacter* database (http://pubmlst.org/campylobacter/). The non-*jejuni*/*coli* *Campylobacter* database is smaller than the *Campylobacter jejuni*/*coli* database: as of January 2016, it contained 268 isolates and 762 MLST profiles across nine schemes (Table 1.3).

Table 1.3 Multilocus sequence typing schemes from the non-*jejuni*/*coli* *Campylobacter* PubMLST database

MLST[a] scheme	STs[b] (*n*)	Isolates (*n*)
C. concisus/*C. curvus*	130	0[c]
C. fetus	57	2
C. helveticus	9	9
C. hyointestinalis	131	0
C. insulaenigrae	43	75
C. lanienae	145	0
C. lari	28	20
C. sputorum	16	0
C. upsaliensis	203	162

[a] MLST, multilocus sequence typing.

[b] STs, MLST sequence types.

[c] Certain species have no isolates in the non-*jejuni*/*coli* *Campylobacter* PubMLST database as submitters have not provided representative isolates for novel STs.

1.5.2.2.2.2 The impact of multilocus sequence typing

As reviewed in depth by Colles and Maiden (2012), MLST has markedly improved our understanding of the epidemiology, population biology, and evolution of members of the *Campylobacter* genus, especially *C. jejuni* and *C. coli*. Critical to the success of MLST were concurrent advances in high-throughput sequencing and information technology. The former enabled studies of large isolate collections, and the latter made MLST data portable and accessible through online public databases (Maiden, 2006). MLST-based studies of *C. jejuni* and *C. coli* have demonstrated that although genotypes causing human disease are highly diverse, they overlap both nationally and internationally (Sopwith *et al.*, 2006; Litrup *et al.*, 2007; Mickan *et al.*, 2007; Dingle *et al.*, 2008; Kittl *et al.*, 2011; Cody *et al.*, 2012; Mughini-Gras *et al.*, 2012; Cody *et al.*, 2013) and are distinct from those of *C. jejuni* subsp. *doylei* (Parker *et al.*, 2007). Interestingly, although *C. jejuni* from high-income countries are genetically similar, isolates from Curaçao, a Caribbean island, were distinct from those from more industrialised countries, including the UK, Canada, and Australia (Duim *et al.*, 2003; Dingle *et al.*, 2008). This suggests that there may be *Campylobacter* lineages that are unique to LMICs.

The availability of large MLST collections including isolates from a variety of food animals facilitated the development of formal genetic methods to attribute human disease isolates to sources of infection (Pritchard *et al.*, 2000; Corander & Marttinen, 2006; Wilson *et al.*, 2008; Mullner *et al.*, 2009a, 2009b). These methods have consistently indicated that chicken is a major source of campylobacteriosis in many high-income countries (Mullner *et al.*, 2009a; Sheppard *et al.*, 2009b; Bessell *et al.*, 2012; Mughini-Gras *et al.*, 2012; Kittl *et al.*, 2013; Strachan *et al.*, 2013; Boysen *et*

al., 2014). Attribution studies have also revealed the seasonality of certain genotypes, as well as associations with age, sex, population density, and travel (Strachan *et al.*, 2009; Bessell *et al.*, 2012; Mughini-Gras *et al.*, 2012; Strachan *et al.*, 2013). Importantly, source attribution can be used to monitor the effects of interventions designed to reduce human disease. This is best illustrated by monitoring carried out in New Zealand following the introduction of a series of interventions in the poultry industry, which detected a 74% reduction in cases attributed to poultry, approximately halving the national human disease burden (Sears *et al.*, 2011).

A detailed understanding of the ecology of *Campylobacter* is a prerequisite for the development of effective intervention strategies, and MLST-based studies have revealed associations between certain lineages and particular animal hosts. For instance, clonal complex ST-257 is over-represented in chicken and ST-61 in cattle, while others, such as the ST-21 clonal complex, are considered "generalists" because they have a broad host range (Colles & Maiden, 2012). That the signal of host-association of certain genotypes is sufficiently strong that it obscures intercontinental geographic signals implies that there is a link between the global distribution of food and the international dissemination of *Campylobacter* genotypes (McCarthy *et al.*, 2007; Sheppard *et al.*, 2010b). Taken together, the results of MLST-based epidemiological investigations in humans and animals suggest that diet, agricultural practices, and food distribution networks all impact on the global epidemiology of *C. jejuni* and *C. coli* (Colles & Maiden, 2012).

With reference to the population biology of *C. jejuni* and *C. coli*, both species are weakly clonal, but their population structures are distinct: *C. jejuni* segregates into a large number of clonal complexes that have little evidence of a clonal frame (Dingle *et al.*, 2002; Maiden & Dingle, 2008), while *C. coli* population forms just three groups, known as clades 1, 2, and 3 (Dingle *et al.*, 2005; Sheppard *et al.*, 2008, 2010a). The majority of *C. coli* human disease isolates belong to clade 1, as do most agricultural isolates. In contrast, isolates from clades 2 and 3 predominantly correspond to environmental samples and are rarely associated with human disease (Sheppard *et al.*, 2010a). *Campylobacter* are naturally competent (Wang & Taylor, 1990), and MLST studies have provided evidence of progressive introgression (the introduction of alleles from one species into the gene pool of another due to horizontal genetic transfer) between *C. jejuni* and *C. coli*, suggesting that the species are converging (Sheppard *et al.*, 2008, 2011). It has been postulated that agriculture is driving this process, as it has generated a new niche in which *C. jejuni* and *C. coli* overlap (Sheppard *et al.*, 2008, 2010a).

Studies of non-*jejuni*/*coli Campylobacter* species have included smaller numbers of isolates compared to those of *C. jejuni* and *C. coli*, yet have still provided insights into the epidemiology and population structures of these organisms. For instance, although *Campylobacter fetus* subsp. *fetus* has been isolated from a variety of host sources and *Campylobacter fetus* subsp. *venerealis* appears to be restricted to the genital tract of cattle, these species were found to be highly homogeneous by MLST (van Bergen *et al.*, 2005). In contrast, *C. lari*, *C. upsaliensis*, and *C. helveticus* are genetically diverse, with recombination detected between MLST loci in *C. jejuni* and *C. lari*, and *C. upsaliensis* and *C. helveticus* (Miller *et al.*, 2005).

Although MLST has led to important advances in our knowledge of *Campylobacter*, especially its epidemiology, the technique is labour intensive and relatively expensive for the volume of sequencing data that is generated (Taboada *et al.*, 2013). It can also lack sufficient discriminatory power for outbreak investigations (Colles & Maiden, 2012). Inclusion of additional loci, specifically highly variable antigen genes, can provide the level of discrimination required to resolve outbreaks (Dingle *et al.*, 2008); however, this increases costs and labour requirements. As a result, there is still a need for highly discriminatory, portable, and objective typing techniques that can be easily incorporated into the clinical microbiology laboratory.

1.5.2.2.3 Whole-genome sequencing

As described in the preface to this thesis, recent studies have demonstrated the potential of WGS for use in clinical microbiology, for both public health and diagnostic applications. Although the term 'WGS' is a misnomer (the majority of bacterial genomes are draft assemblies, that is, they are incomplete and exist as sets of contiguous sequences or "contigs"), the discriminatory power of genome sequencing supersedes that of existing bacterial typing techniques. Unlike most other typing methods, WGS can be used for short- and long-term epidemiological investigations, as well as for studies of bacterial population genetics and evolution. An overview of the different NGS platforms and analytical approaches currently available is provided here, followed by a description of how WGS has been applied to study *Campylobacter*, focusing on epidemiological research.

1.5.2.2.3.1 Next-generation sequencing platforms

The bacterial genomic era began in the mid-1990s, when the genome sequence of *Haemophilus influenzae* was determined using Sanger sequencing based on a novel shotgun sequencing approach (Fleischmann *et al.*, 1995). As reviewed by Mardis (2013), Sanger sequencing has undergone several changes to improve accuracy, scalability, and instrument run time, but the underlying principle of chain termination remains the same. The cycle sequencing reaction contains a mixture of deoxynucleotides (dNTPs) and fluorescently labelled dideoxynucleotides (ddNTPs), and, once a ddNTP is incorporated into the growing strand of DNA, elongation is terminated due to the lack of a 3′ hydroxyl group (Sanger *et al.*, 1977; Smith *et al.*, 1986). This process generates DNA fragments of varying lengths, which are then size-separated by electrophoresis. The replacement of polyacrylamide gel electrophoresis by capillary electrophoresis during the late 1990s markedly improved sample throughput, base-calling, data extraction, and run time (Marsh *et al.*, 1997). On exit from the capillary, a laser excites the fluorescent tag of the terminal ddNTP, the signal is captured digitally, and the sequence of the fragment is gradually determined.

In 2005, the first NGS platforms were introduced (Mardis, 2008). These platforms built on the successes of Sanger sequencing, and, as outlined in the preface to this thesis, have a broad range of applications beyond bacterial WGS (Mardis, 2013). Competition among manufacturers has since resulted in the development of several new NGS platforms, as well as enhancements to existing technologies (Reuter *et al.*, 2015). Consequently, sequencing capacity and associated costs have improved, making bacterial WGS increasingly accessible (Mardis, 2011; Fournier *et al.*, 2014).

As of November 2015, the Sequence Read Archive (SRA), the main repository for raw NGS data, contained 297,354 records corresponding to bacterial DNA sequences. Filtering the records by NGS manufacturer indicated that Illumina platforms currently dominate the market, with 90.1% of sequences produced using these machines (Figure 1.6). Second to Illumina was Roche/454, the manufacturer of the first NGS platform (Margulies *et al.*, 2005), followed by Pacific Biosciences, which produced the first widely-used single-molecule sequencing platform (Eid *et al.*, 2009). These platforms will be described further in the sections that follow, with particular emphasis on Illumina, given its current position and that it was the technology of choice for the studies presented in this thesis. Other less frequently used NGS technologies represented in the SRA included the Ion Torrent and SOLiD platforms from Life Technologies (Valouev *et al.*, 2008; Rothberg *et al.*, 2011), the Heliscope single-molecule sequencer by Helicos BioSciences (Harris *et al.*, 2008), and Oxford Nanopore Technologies (Clarke *et al.*, 2009) (Figure 1.6). With the exception of Oxford Nanopore, these platforms will not be discussed here, but, for the interested reader, have been reviewed elsewhere (Zhou *et al.*, 2010; Reuter *et al.*, 2015).

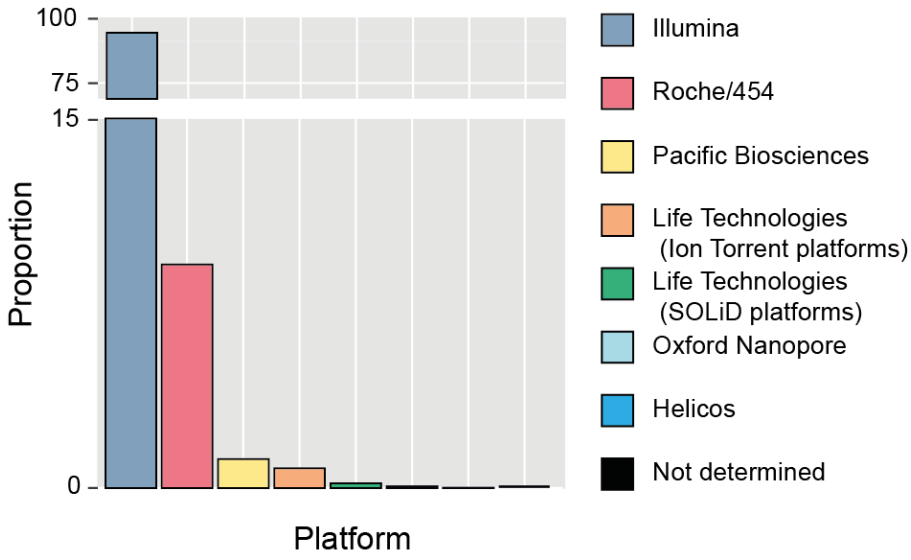


Figure 1.6 Use of next-generation sequencing platforms for bacterial DNA sequencing. The total number of bacterial DNA sequences in the SRA (297,354) (http://www.ncbi.nlm.nih.gov/sra) was determined using the search term 'bacteria AND "biomol DNA"'. To attribute records to NGS manufacturers, the SRA was iteratively searched using the term 'bacteria AND "biomol DNA" AND X', where 'X' denotes the names of major manufacturers and platforms shown in the legend. Platforms are shown along the x-axis in order of decreasing rank abundance, as reflected by the legend.

Early NGS platforms, including those from Illumina and Roche/454, were based on a sequencing-by-synthesis approach (Loman *et al.*, 2012a). Although the general principles behind these platforms remain unchanged, machine output and read-length have markedly improved, particularly in the case of Illumina (Figure 1.7) (Reuter *et al.*, 2015). The Illumina and Roche/454 workflows consist of three stages: (i) library preparation; (ii) template amplification; and (iii) sequencing. For both technologies, library preparation entails fragmentation of purified genomic DNA by mechanical or enzymatic methods to produce random, overlapping fragments. Platform-specific adaptors are ligated to the ends of the resulting fragments, which are separated by size in preparation for template amplification (Figure 1.8A). It is at this point that Illumina and Roche/454 technologies diverge.

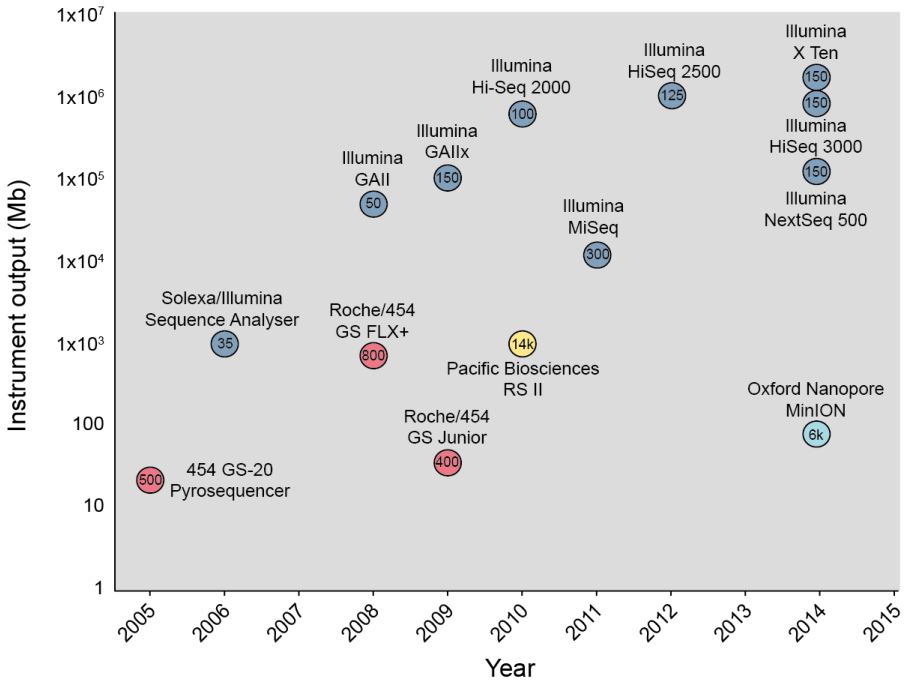


Figure 1.7 Improvements in instrument output and read length of next-generation sequencing platforms produced by key manufacturers between 2005 and 2015. Circles represent individual platforms and are coloured by manufacturer: Roche/454, red; Illumina, dark blue; Pacific Biosciences, yellow; Oxford Nanopore, turquoise. The number shown within each circle indicates sequence read length. Adapted from Reuter *et al.* (2015).

In preparation for Illumina sequencing (Figure 1.8), the genetic library is denatured, after which the single-stranded DNA fragments are immobilised at the 5′ end to the surface of a flow cell densely populated with oligonucleotides complementary to the terminal adaptor sequences (Figure 1.8B). Following the addition of DNA polymerase and dNTPs, “bridge amplification” is initiated. The free 3′ end of the template binds to a nearby complementary adaptor, forming the characteristic bridge-like structure for which the process is named. The adaptor functions as a primer, and the template is amplified, after which the double-stranded products are denatured. The process is repeated until, after multiple cycles, the flow cell is

covered with millions of surface-bound clusters or colonies of amplified template molecules, which are the foci for sequencing (Figure 1.8B).

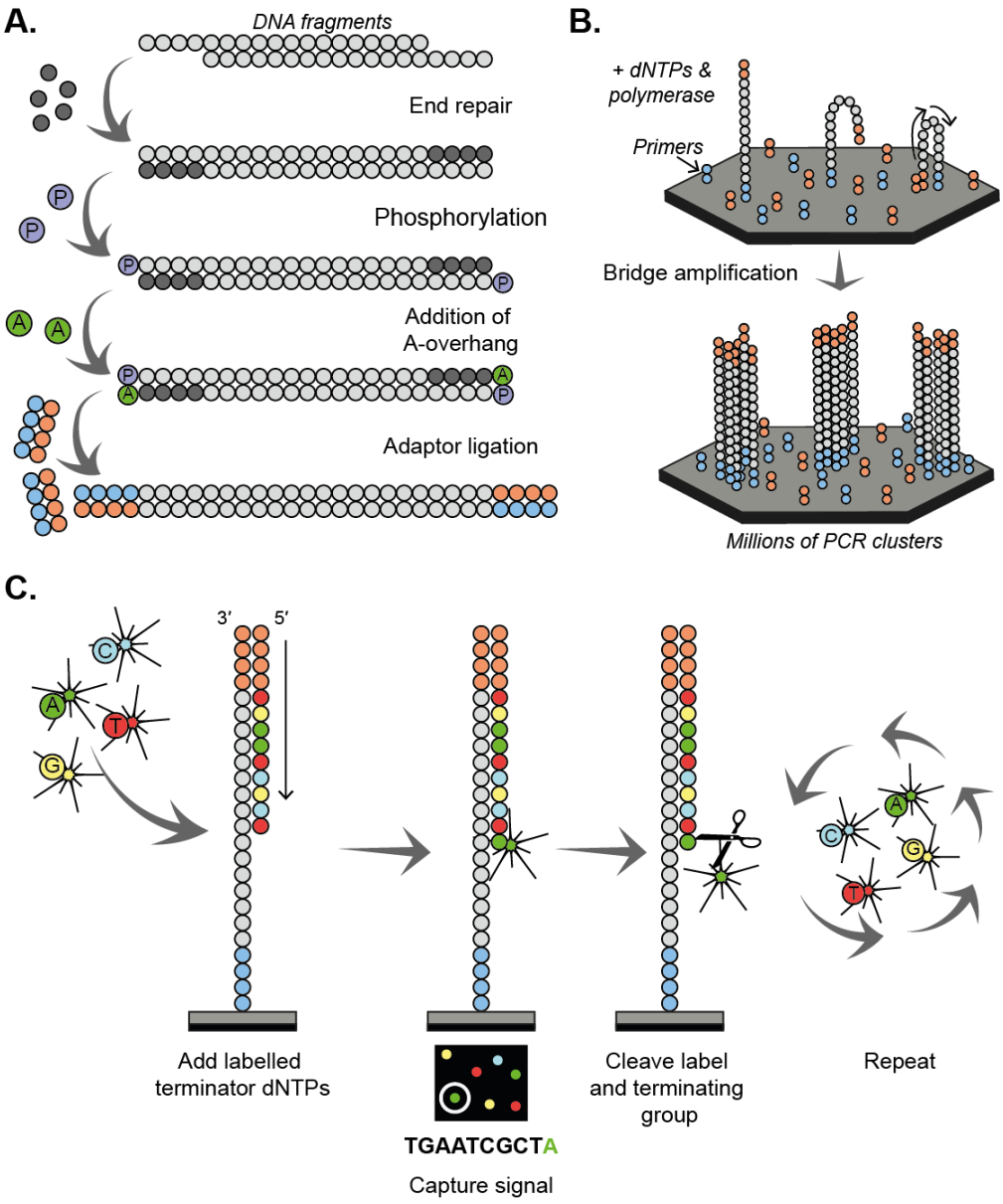


Figure 1.8 The Illumina workflow. (A) Library preparation. DNA fragments generated by mechanical or enzymatic shearing are ligated to adaptors. (B) Template amplification. DNA fragments bind to complementary to adaptors on the surface of the flow cell. Clusters of copies of the template DNA fragments are generated by bridge amplification. (C) Sequencing-by-synthesis. Reversible terminator nucleotides are added to the flow cell. A complementary nucleotide is incorporated and the signal captured. The fluorescent label and blocking group are removed, readying the strands for the next cycle. Adapted from Metzker (2010) and Mardis (2013).

The Illumina sequencing reaction mix contains four proprietary reversible terminator nucleotides, which have identifying fluorescent labels and a moiety at the 3′ hydroxyl that prevents incorporation of additional nucleotides (Figure 1.8C). Each cycle involves incorporation of a complementary terminator nucleotide, washing to remove unincorporated nucleotides, and fluorescence imaging, which is followed by cleavage of the fluorescent label and chemical removal of the blocking group to prepare for the next cycle. The process is repeated *n* times, where *n* is the read length, thereby gradually determining the template sequence for each cluster (Figure 1.8C) (Metzker, 2010; Mardis, 2013). Illumina platforms are characterised by low error rates (< 1%), with the majority corresponding to substitution errors at the ends of reads (Dohm *et al.*, 2008). As these errors are random, they may be overcome with increased coverage (Loman *et al.*, 2012a).

Although Roche/454 platforms are also based on a sequencing-by-synthesis approach, they exploit bioluminescence rather than fluorescence, and produce longer read lengths of up to 800 bp (Figure 1.7). In preparation for Roche/454 pyrosequencing, library fragments are denatured and immobilised onto the surface of agarose beads, such that each bead is bound to a single template molecule. Template amplification takes place in an oil-water emulsion: each bead is isolated in a single micelle, and the attached DNA fragment is amplified, generating several thousand copies that are also bound to the surface of the bead. The beads are then transferred to the wells of a PicoTiterPlate, to which enzyme and packing beads are added, followed by the four dNTPs in turn. When a dNTP is incorporated into the growing DNA strand, pyrophosphate is released, which initiates an enzyme cascade that effects the production of visible light by luciferase. The signal is proportional to

the number of dNTPs incorporated, and is captured as a flowgram; the underlying nucleotide sequence is determined based on the order and intensity of peaks in the flowgram. The Roche/454 error rate is approximately 1%, with errors most commonly corresponding to insertions and deletions in homopolymeric tracts (Mardis, 2008). As these errors are systematic, they typically cannot be overcome by increased coverage, instead requiring the use of an alternative technology to resolve the problematic sequence (Loman *et al.*, 2012a).

The most recent development in NGS has been the commercialisation of several single-molecule sequencing technologies. Instead of relying on template amplification, these platforms sequence library fragments directly, thereby decreasing processing time and eliminating the introduction of amplification artefacts. Single-molecule sequencers are characterised by long read lengths, reaching up to 14,000 bp (Figure 1.7). The Heliscope, released by Helicos BioSciences in 2009, was the first single-molecule sequencer; however, Helicos BioSciences filed for bankruptcy in 2012, and the platform has not been widely used. The RS II platform, delivered by Pacific Biosciences in 2010, has proved more successful, while the MinION from Oxford Nanopore Technologies, which was delivered this year, promises faster, cheaper single-molecule sequencing with read lengths averaging several thousand base pairs (Figure 1.7).

In the case of the RS II, sequencing is carried out on a chip covered with thousands of zero-mode waveguide (ZMW) detectors, which are nanophotonic chambers, each with a φ29 DNA polymerase molecule tethered to its base. Within each ZMW

detector, a single adaptor-bound library fragment is captured by the DNA polymerase. Incorporation of a complementary dye-labelled dNTP results in the release of a fluorescent signal, which, because of the physics of the chambers, can be detected against high background dNTP concentrations. The RS II is characterised by a high error rate of approximately 13-15%, with the majority of errors corresponding to random insertions (12%), followed by deletions (2%) and substitutions (1%) (Quail *et al.*, 2012; Carneiro *et al.*, 2012). That φ29 DNA polymerase is highly processive and strand-displacing facilitates resequencing of closed circular templates, which increases platform accuracy (Eid *et al.*, 2009).

In 2012, Oxford Nanopore Technologies announced the MinION, promising a sequencing platform with minimal sample preparation, rapid turnaround times, and long reads at low cost (Eisenstein, 2012). The long-awaited technology was made available through an early-access programme in 2014 (Bayley, 2015). In contrast with other NGS platforms, the MinION does not require complex sample preparation and labelling, or rely on enzyme-dependent amplification, as it exploits nanopore technology and electrochemistry. The MinION flow cell is covered with a synthetic bilayer embedded with hundreds of nanopores; translocation of a DNA molecule through one of these channels results in nucleotide-specific changes in the ionic current that flows through the pore, which are recorded and converted to sequencing data. At present, insertion, deletion, and substitution error rates remain high at 4.9%, 7.8%, and 5.1% (Quick *et al.*, 2014; Mikheyev & Tin, 2014; Jain *et al.*, 2015; Ashton *et al.*, 2015a); however, the MinION is in its infancy relative to other NGS platforms, and further developments to improve speed and accuracy are predicted (Check Hayden, 2015).

1.5.2.2.3.2 Trends in choice of NGS platform

The NGS platforms described here represent a subset of those that have been developed, all of which vary with respect to performance characteristics such as instrument capacity, read length, and error profiles. Platform choice is guided by the intended application, as well as practical considerations. To date, most studies on the genomic epidemiology of bacteria have been retrospective, and carried out primarily by academic institutions. In this context, researchers often batch samples to make use of high-end, high-throughput NGS platforms such as the Roche/454 GS FLX+, the RS II from Pacific Biosciences, and the Illumina Genome Analyser and HiSeq machines (Figures 1.6 and 1.7). However, due to their significant initial cost, the necessity of high throughput for cost efficiency, and their large laboratory footprints, these platforms are best suited to sequencing centres or core facilities. As the field moves towards real-time genomic epidemiology, there is a growing need for rapid, cost-effective in-house sequencing of smaller sets of isolates. Compact bench-top machines, such as the Roche/454 GS Junior and the MiSeq from Illumina, have started to meet this need. Concerning the ultimate goal of using WGS in the field during epidemiological investigations, the MinION, which is the size of a USB stick, represents the first portable NGS platform. The potential of this device for real-time outbreak investigations was demonstrated during the 2014 Ebola epidemic in West Africa and a *S. enterica* outbreak in a UK hospital (Check Hayden, 2015; Greninger *et al.*, 2015; Quick *et al.*, 2015).

1.5.2.2.3.3 Analysis of genomic data

The first studies to use WGS to investigate the epidemiology of bacterial pathogens developed an approach known as “read mapping” for the analysis of NGS data. In this approach, short reads generated by NGS are aligned to a reference genome to identify single nucleotide polymorphisms (SNPs), which are used to construct a phylogeny and visualise the relationships between the isolates (Loman *et al.*, 2012a). Read mapping has been applied to study the short- and long-term epidemiology of several clinically relevant bacteria, such as *Mycobacterium tuberculosis*, *Staphylococcus aureus, Streptococcus pneumoniae, C. difficile*, *V. cholerae*, *S.* Typhi and *S. flexneri* (Harris *et al.*, 2010; Croucher *et al.*, 2011; Mutreja *et al.*, 2011; McAdam *et al.*, 2012; Köser *et al.*, 2012a; Bryant *et al.*, 2013; Harris *et al.*, 2013; Holden *et al.*, 2013; Walker *et al.*, 2013b; Connor *et al.*, 2015; Wong *et al.*, 2015). However, an important drawback of the read mapping approach is that it cannot be used if a closely related reference genome is not available. Similarly, regions absent from the reference genome are not included in the analysis (Loman *et al.*, 2012a).

As read lengths increased, it became possible to assemble draft genomes *de novo*, generating contigs based on consensus sequences determined from overlapping short reads (Edwards & Holt, 2013). This allows for analysis of WGS data without the requirement for a reference genome. Numerous genome assembly methods have been developed, most of which make use of de Bruijn graphs (Compeau *et al.*, 2011), including VELVET (Zerbino & Birney, 2008), EULER-SR (Chaisson & Pevzner, 2008), ABySS (Simpson *et al.*, 2009), SOAPdenovo (Li *et al.*, 2010), and SPAdes (Bankevich *et al.*, 2012). Recently, “assemblathons” have been initiated to compare

the performance of existing genome assembly tools using standardised datasets, with the results released to the community (Earl *et al.*, 2011; Bradnam *et al.*, 2013). Among existing assembly tools, VELVET is one of the oldest and most widely used programs, as it is well-suited for use with Illumina data (Loman *et al.*, 2012a; Edwards & Holt, 2013). VELVET includes two programs, *velveth* and *velvetg* (Zerbino & Birney, 2008; Zerbino, 2010). In a preliminary step, short reads are converted to smaller fragments, known as *k*-mers, using *velveth*. Overlapping *k*-mers are then assembled into contigs with *velvetg* by means of a de Bruijn graph (Compeau *et al.*, 2011). A third program, VELVETOPTIMISER, is included in the VELVET package for the optimisation of assembly parameters (Zerbino, 2010).

Once assembled, draft genomes can be aligned using programs such as MUMmer (Kurtz *et al.*, 2004), Mugsy (Angiuoli & Salzberg, 2011), or Mauve (Darling *et al.*, 2004, 2010), and SNPs can be extracted for use in phylogenetic analyses. As recombination is known to distort phylogenetic signals, affected regions of the genome are typically removed prior to reconstructing the phylogeny, just as in read mapping analyses; however, this can exacerbate the misrepresentation of branch lengths (Hedge & Wilson, 2014). It is also problematic that read mapping and SNP-based analyses of *de novo* assemblies are computationally intensive, especially when working with large datasets. Furthermore, these approaches require extensive computational resources and bioinformatics expertise, or the purchase of commercial software. The gene-by-gene approach implemented using web-based platforms, such as the Bacterial Isolate Genome Sequence Database (BIGSDB), represents an alternative approach to SNP-based methods for the analysis of WGS data (Maiden *et al.*, 2013).

The gene-by-gene approach is “MLST-like”, defining relationships among isolates based on the comparison of alleles assigned for a particular set of loci. Just as in MLST [1.5.2.2.2], there is no requirement to identify and exclude those changes that are due to recombination, and analyses can be carried out using allele numbers or nucleotide sequences as appropriate. The gene-by-gene approach is flexible, in that the number of loci included in an analysis can be adjusted based on the level of resolution required. One of the strengths of this method is that it can be used to relate typing data to nomenclature (Figure 1.9). In this way, smaller sets of loci such as the 16S rRNA gene or the MLST loci can be used for low resolution typing, identifying isolates to the genus and species-levels, respectively. The number of loci can then be increased when higher resolution is required: ribosomal MLST (rMLST) can be used to identify isolates up to the lineage level (Jolley *et al.*, 2012b), core-genome MLST (cgMLST) can be used to examine relationships among related yet distant groups of isolates, and whole-genome MLST (wgMLST), encompassing all genes shared among isolates, can be used to distinguish closely related strains and clones (Figure 1.9) (Maiden, 2006; Maiden *et al.*, 2013).

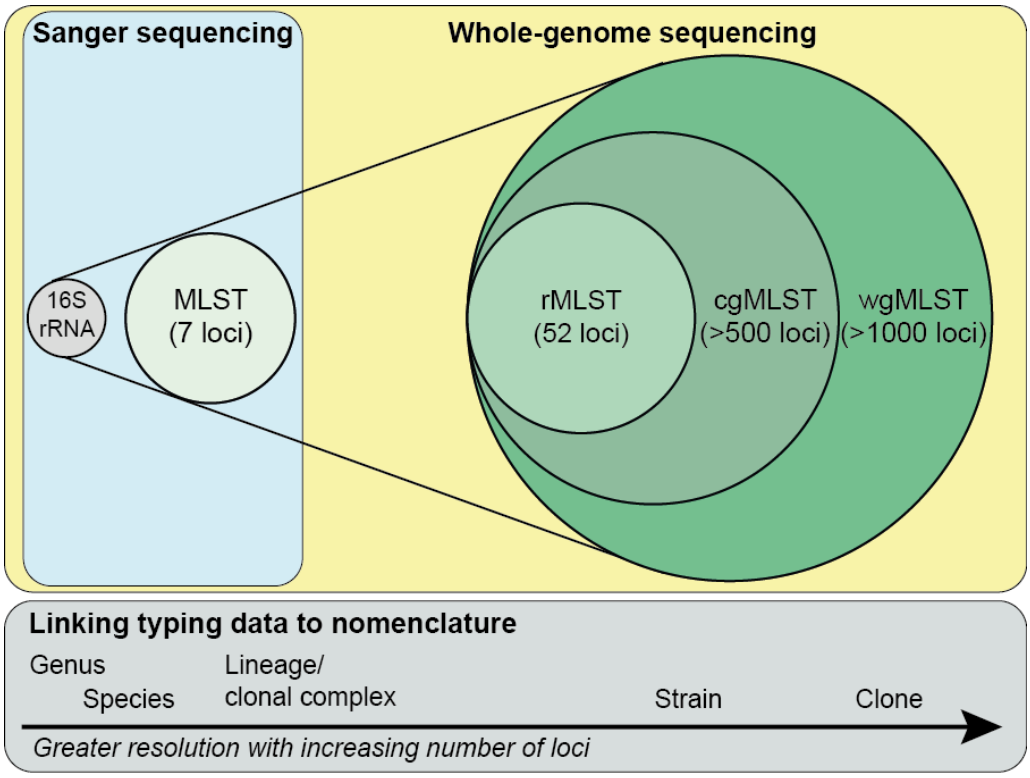


Figure 1.9 Using the gene-by-gene approach to relate sequence-based typing data to taxonomy. The hierarchical MLST franchise is shown in the top panel. The bottom panel demonstrates how typing schemes relate to bacterial nomenclature. MLST, multilocus sequence typing; rMLST, ribosomal MLST; cgMLST, core-genome MLST; wgMLST, whole-genome MLST. Adapted from Maiden *et al*. (2013).

The gene-by-gene approach is implemented using the open-source web-accessible BIGSDB software (Jolley & Maiden, 2010; Maiden *et al.*, 2013). An instance of BIGSDB comprises two databases, namely the isolate database and the sequence definitions database (Figure 1.10). The isolate database links provenance and phenotype data to nucleotide sequence data, while the sequence definitions database contains all known alleles for user-defined genes of interest. By means of a process known as “scanning and tagging”, contigs in the isolate database are searched for sequences similar to the genes of interest using the BLAST algorithm. When an existing allele is identified, the sequence is tagged in the isolate database,

indicating its position in the genome and allele designation. New alleles are first added to the sequence definitions database, and then the isolate is re-scanned and tagged in an iterative process (Figure 1.10) (Jolley & Maiden, 2010).

In addition to storing data and annotating genome sequences, BIGSDB also includes analysis tools, such as the GENOME COMPARATOR module (Figure 1.10). GENOME COMPARATOR can be used to compare isolates using user-defined loci, which can be grouped into schemes useful for typing or taxonomy, or to aid functional annotation. Alternatively, isolates can be compared to an annotated reference genome. In both cases, the resulting allelic profiles are converted to a distance matrix, which is automatically visualised as a NeighborNet graph (network) using SplitsTree. GENOME COMPARATOR also produces a list of variable loci and, if requested, additional outputs, such as genetic distances and nucleotide sequence alignments for use in further analyses (Figure 1.10) (Jolley & Maiden, 2010; Maiden *et al.*, 2013). Since its development, the gene-by-gene approach has been used to study the epidemiology of several bacterial pathogens, including *Campylobacter* (Cody *et al.*, 2013), *Neisseria* (Jolley *et al.*, 2012a; Bratcher *et al.*, 2014; Harrison *et al.*, 2015; Hill *et al.*, 2015; Brehony *et al.*, 2015; Oldfield *et al.*, 2016), *E. coli* (Mellmann *et al.*, 2011), and *Streptococcus pneumoniae* (Wyres *et al.*, 2012; van Tonder *et al.*, 2015).

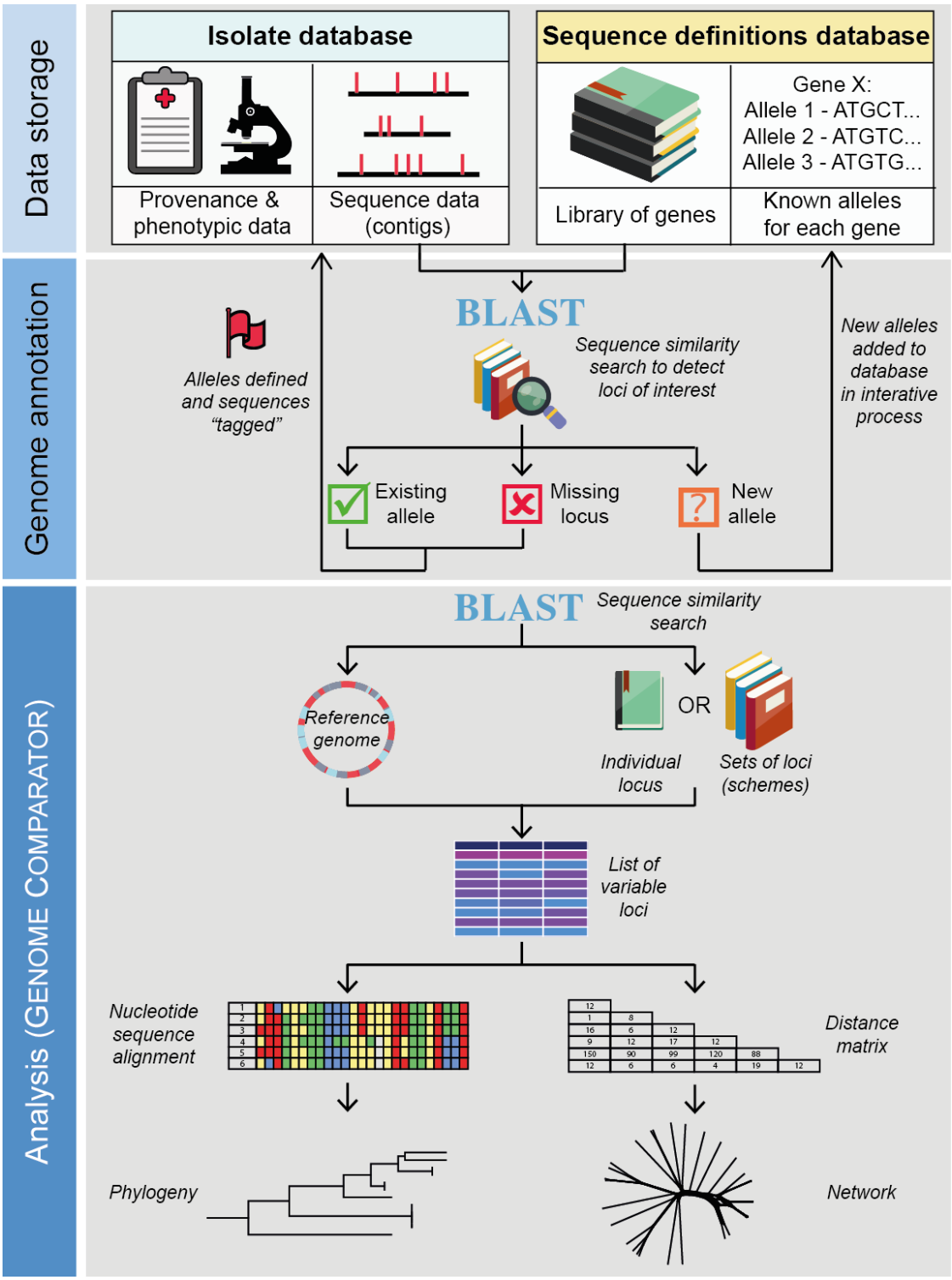


Figure 1.10 The gene-by-gene approach for analysis of whole-genome sequencing data as implemented with Bacterial Isolate Genome Sequence Database software. The top two panels illustrate how WGS data are stored and annotated in an instance of BIGSDB. The bottom panel shows how data can be analysed using GENOME COMPARATOR, which can also generate output files for use with external software (as in the case of phylogeny reconstruction, depicted here). Adapted from Maiden *et al*. (2013).

1.5.2.2.3.4 Whole-genome sequencing for Campylobacter *epidemiology*

The first finished *Campylobacter* genome was that of *C. jejuni* isolate NCTC 11168, which was completed in 2000 using a shotgun sequencing approach (Parkhill *et al.*, 2000). Since the sequencing of NCTC 11168, an additional 55 finished genomes corresponding to 16 *Campylobacter* species have been made publicly available via the GenBank Genome Resource (http://www.ncbi.nlm.nih.gov/genome/). An even greater volume of WGS data has been deposited in the SRA: 15,877 *Campylobacter* DNA records were present at the time of writing. The availability of large datasets of diverse *Campylobacter* genomes has facilitated numerous comparative genomics and evolutionary studies. These studies have demonstrated genome-wide introgression between *C. jejuni* and *C. coli* (Sheppard *et al.*, 2013b); explored mechanisms of host adaptation and inferred recombination hotspots in *C. jejuni* (Sheppard *et al.*, 2013a; Yahara *et al.*, 2014); revealed major structural differences among *C. jejuni*, *C. coli*, *C. lari*, and *C. upsaliensis* (Fouts *et al.*, 2005); provided insights into the biology of *C. lari* (Miller *et al.*, 2014a) and *C. fetus* (Ali *et al.*, 2012; Kienesberger *et al.*, 2014; van der Graaf-van Bloois *et al.*, 2014); and estimated the core genome of various subsets of *Campylobacter* species (Lefébure & Stanhope, 2009; Lefébure *et al.*, 2010; Zhang & Sievert, 2014; Iraola *et al.*, 2014; Méric *et al.*, 2014; Skarp-de Haan *et al.*, 2014).

Despite the growing interest in the application of WGS to problems in public health microbiology (Didelot *et al.*, 2012; Köser *et al.*, 2012b), it is not yet widely used for routine surveillance of *Campylobacter*. This is largely because, although sequencing costs have decreased (Mardis, 2011), the volume of *Campylobacter* samples

received makes such a programme prohibitively expensive in most laboratories (On, 2013). Recent studies have, however, demonstrated the potential of WGS for *Campylobacter* outbreak investigations, in-depth studies of lineages of clinical importance, and sentinel surveillance. For example, WGS was used to resolve milk- and water-borne *C. jejuni* outbreaks in Finland (Revez *et al.*, 2014a, 2014b), and a milk-borne *Campylobacter* outbreak in the UK that was caused by pasteurisation failure (Fernandes *et al.*, 2015). Similarly, a comparative genomics approach has been used on small datasets to better understand the emergence and genetic characteristics of important lineages, such as ST-474 in New Zealand (Biggs *et al.*, 2011) and ST-677 in Finland (Kivistö *et al.*, 2014; Skarp *et al.*, 2015). With reference to sentinel surveillance, WGS has been used to monitor the epidemiology of human campylobacteriosis in Oxfordshire, UK, since 2011. The first published reports on these data demonstrate the utility of WGS: as isolates can be compared to those from previous MLST-based studies, long-term epidemiological trends and changes can be monitored (Cody *et al.*, 2013, 2015); isolates can also be subjected to high-resolution analyses for the identification of potential outbreaks, which are typically difficult to detect (Cody *et al.*, 2013). As the costs of WGS continue to decrease and bench-top sequencers are introduced into more laboratories, it is likely that genomics will be used more frequently to study various aspects of the epidemiology of *Campylobacter*.

1.6 Project background and objectives

Although there is a paucity of information on the epidemiology of *Campylobacter* from LMICs, these organisms have been a research interest in South Africa since the 1970s. A PubMed search using the terms "*Campylobacter* AND 'South Africa'" identified 67 original research articles published between 1971 and 2015. These studies encompassed a broad range of topics, such as: (i) the isolation and characterisation of *Campylobacter* from animals, food, and water (Schutte *et al.*, 1971; van der Walt & van der Lugt, 1988; van der Walt *et al.*, 1988; Pefanis *et al.*, 1988; Odendaal *et al.*, 1994; Kunene *et al.*, 1999; Mosupye & von Holy, 1999, 2000; Cooper *et al.*, 2002; Obi *et al.*, 2003; Diergaardt *et al.*, 2004; van Nierop *et al.*, 2005; Potgieter *et al.*, 2005; Bartkowiak-Higgo *et al.*, 2006; Bester & Essack, 2008; Schmidt *et al.*, 2010; Jonker & Picard, 2010; Bester & Essack, 2012; Njiro *et al.*, 2011; Manyi-Loh *et al.*, 2014); (ii) the screening of novel medicinal compounds for activity against *Campylobacter* (Samie *et al.*, 2009a; Madikizela *et al.*, 2013); (iii) the development and testing of novel diagnostic assays and molecular typing methods (Owen *et al.*, 1988, 1990; Owen & Hernandez, 1990a; Bok *et al.*, 1991; Da Silva Tatley *et al.*, 1992; Fayos *et al.*, 1993; Matsheka *et al.*, 2001, 2002, 2006; Hoosain & Lastovica, 2009; Schmidt *et al.*, 2010); and (iv) the epidemiology of human campylobacteriosis. The latter corresponds to an important area of *Campylobacter* research in South Africa, which is discussed further in Chapter 5.

Situated at the southernmost tip of Africa (Figure 1.11), South Africa is a medium-sized country of 1.22 million square kilometres, which is approximately five times the size of the UK. The country is comprised of nine provinces and has three capital cities, Cape Town, Pretoria, and Bloemfontein (Figure 1.11). In 2015, the population

numbered approximately 55 million (Statistics South Africa, 2015), equivalent to ~86% of that of the UK. South Africa is currently recognised as an upper middle-income country and has the second-largest gross domestic product in Africa (The World Bank, 2015). However, widespread economic inequality remains a major problem. As of 2011, South Africa's Gini co-efficient, a measure of income inequality, was 0.69 (Statistics South Africa, 2014), identifying it as one of the most unequal countries in the world (The World Bank, 2016a).



Figure 1.11 Map of South Africa. The inset shows the location of South Africa (green) on the African continent. The main image depicts its provinces: the Western Cape, where the isolates described in this thesis were collected, is shown in green, and other provinces with a history of *Campylobacter* research are shown in yellow. The three capital cities are marked with black circles.

The socioeconomic challenges that South Africa faces can be attributed to policies implemented during apartheid (1948-1994), but also to the preceding three centuries of colonial rule, when the subjugation of indigenous people and the seizure of land and resources began (Thompson & Berat, 2014). When the National Party came to power in 1948, racial discrimination was institutionalised and the policies of apartheid were consolidated to ensure the dominance of "white" South Africans. To this end, legislation was passed curtailing the social, economic, and political freedom of people classified as "black", "coloured", or "Indian". These laws dictated where people of different racial groups could live, work, and attend school, who they could have relationships with, and whether they were entitled to citizenship or the right to vote. Similarly, resources allocated to healthcare, education, and basic services were unevenly distributed based on racial classification (Giliomee & Mbenga, 2007).

The policies of apartheid have had lasting repercussions, especially those that systematised geographic segregation and economic marginalisation based on race. The Group Areas Act of 1950 enforced racial segregation of residential areas, reserving developed areas for white people and requiring those classified as black, coloured, or Indian to move to segregated townships and informal settlements on the outskirts of urban settlements, or to rural areas. These regions typically lacked infrastructure and basic services, such as water, sanitation, and electricity (Giliomee & Mbenga, 2007). The majority of South Africans (64%) currently live in urban areas (The World Bank, 2016b), which, as indicated by 2011 census data, remain largely racially segregated (http://dotmap.adrianfrith.com/). Delivery and upkeep of basic services in townships and informal settlements continue to be poor, resulting in frequent protests by residents (Chen *et al.*, 2014). Just as race remains a predictor

of residential area, 2011 census data indicate that it is also a proxy for income (http://dotmap.adrianfrith.com/). Although poverty has declined since 1994, 45.5% of the South African population was still living below the poverty line in 2011, with 20.2% living in extreme poverty, that is, being unable to purchase sufficient food (Statistics South Africa, 2014). Importantly, these and other apartheid policies have had a significant impact on health and the healthcare system (Coovadia *et al.*, 2009; Harris *et al.*, 2011; Mayosi *et al.*, 2012; Mayosi & Benatar, 2014).

South Africa does not compare favourably with other middle-income countries in respect of measures such as life expectancy at birth, years lived with disability, and years lost due to premature death, with the poor worst affected (Mayosi & Benatar, 2014). Although the government is working towards universal health coverage, access to healthcare remains unequal (Coovadia *et al.*, 2009; Harris *et al.*, 2011; Mayosi *et al.*, 2012; Mayosi & Benatar, 2014). A two-tier system consisting of public and private sectors is currently in place, and human and financial resources are skewed towards the latter (Mayosi & Benatar, 2014). The time and cost to reach a healthcare facility, long waiting times, high out-of-pocket payments, and anticipation of disrespectful public healthcare workers have been identified as barriers to healthcare, particularly among those who are impoverished (Harris *et al.*, 2011).

South Africa has been described as facing four “concurrent epidemics” corresponding to: (i) infectious diseases, in particular HIV/AIDS and tuberculosis; (ii) non-communicable diseases; (iii) accidents and interpersonal violence; and (iv) childhood malnutrition and diarrhoea (Mayosi *et al.*, 2012). The latter highlights the

importance of epidemiological surveillance of gastrointestinal pathogens in the region. The primary aim of this thesis was to improve our understanding of the local epidemiology of *Campylobacter* and to contextualise it within a global framework. The secondary aim was to explore the potential of WGS for use in diagnostic and public health microbiology using *Campylobacter* as an exemplar. The individual studies that comprise the thesis form two parts, which are outlined briefly here to provide an overview of the material that follows.

Part I: Using whole-genome sequencing to address problems in *Campylobacter* diagnostics

The isolation and identification of *Campylobacter* remains a challenge, particularly in clinical microbiology laboratories. For this reason, high-income countries have begun to supplement or even replace conventional diagnostics with rapid molecular tests and, in some cases, WGS. In contrast, many LMICs do not routinely culture *Campylobacter*; however, molecular methods are gradually becoming more popular in these settings, and have the potential to improve *Campylobacter* diagnostics. In this context, Chapter 2 demonstrates how WGS data can be used to evaluate molecular diagnostic assays, while *in silico* methods for the identification of *Campylobacter* species are described in Chapter 3.

Part II: The genomic epidemiology of *Campylobacter* from South Africa

The second part of this thesis extended concepts developed in Part I, using WGS and the gene-by-gene approach implemented in BIGSDB software to study the

epidemiology of *Campylobacter* from South Africa. Chapter 4 describes the genetic diversity of *Campylobacter* isolates from free-range broiler chickens from the Western Cape province, while Chapter 5 describes the molecular epidemiology of human disease isolates from Cape Town, the capital city of that province (Figure 1.11). These studies built on previous work carried out at the University of Cape Town and the University of the Western Cape, providing novel insights into the epidemiology of *Campylobacter* from South Africa.

Part I: Using whole-genome sequencing to address problems in *Campylobacter* diagnostics

CHAPTER 2

Exploiting bacterial whole-genome sequencing data for the evaluation of diagnostic assays: *Campylobacter* speciation as a case study

ABSTRACT

Effective patient management and epidemiological surveillance are contingent on the rapid and accurate diagnosis of infectious diseases. In the future, identification, antimicrobial susceptibility testing, and epidemiological typing of bacterial pathogens could be carried out using WGS. While WGS is being introduced into routine practice, genomics and molecular diagnostics will have complementary roles to play in clinical microbiology. The wealth of WGS data that is already available represents a potential resource for the evaluation and development of molecular diagnostics. This study used publicly available WGS data to evaluate a duplex real-time PCR assay that targets *mapA* and *ceuE* for the detection of *C. jejuni* and *C. coli*. *In silico* analyses of *mapA* and *ceuE* primer and probe sequences from 1,713 genetically diverse *C. jejuni* and *C. coli* genomes, supported by RT-PCR testing, indicated that the assay was robust, with 1,707 (99.7%) isolates correctly identified. Rare instances of a lack of correlation were due to introgression in *mapA* or *ceuE* sequence diversity in *C. coli* isolates. The underlying reasons for the specificity of the *mapA*/*ceuE* assay were interspecies diversity and intraspecies conservation of the target genes in *C. jejuni* and *C. coli*. The results of this study demonstrate how WGS can be exploited to evaluate molecular diagnostic assays using publicly available WGS data, online databases, and open source software.

2.1 Introduction

Accurate and timely diagnosis of infectious diseases is a cornerstone of clinical microbiology. Notwithstanding the ongoing importance of culture in many settings, molecular diagnostics have markedly improved pathogen detection and identification (Buchan & Ledeboer, 2014). The application of WGS to problems in diagnostic and public health microbiology embodies the current clinical microbiology zeitgeist (Didelot *et al.*, 2012; Köser *et al.*, 2012b; Robinson *et al.*, 2013; Köser *et al.*, 2014). Although WGS has the potential to improve clinical microbiology, the transition is likely to be gradual, especially in resource-limited settings (Didelot *et al.*, 2012; Köser *et al.*, 2012b; Fricke & Rasko, 2014). Until there is widespread implementation of rapid WGS methods in routine laboratories, genomics and molecular diagnostic assays will have complementary roles to play in clinical microbiology. Additionally, molecular diagnostic assays have highlighted the benefits of using small targets, and there is an opportunity to draw on these successes to develop novel *in silico* approaches to address clinical problems.

Since their introduction in the 1980s, NAATs have arguably become the most widely used molecular diagnostic assays in bacteriology laboratories. The uptake of NAATs can be attributed to: (i) their broad applicability across the bacterial domain; (ii) their increased sensitivity and specificity compared to culture, particularly with respect to fastidious or non-viable organisms; (iii) the development of multiplex assays, which facilitate syndrome-driven diagnosis; and (iv) reduced turnaround time (Buchan & Ledeboer, 2014). However, many NAATs have been designed using nucleotide sequences from a relatively small number of isolates. During the pre-WGS era, the

accuracy of NAATs could not be extensively explored at the population level, partly because the requisite large isolate collections were challenging to assemble, and partly because primer sequences were difficult to determine using Sanger sequencing.

The rapid increase in WGS has generated a surfeit of publicly available genomic data that have the potential to improve the development and evaluation of NAATs and other diagnostic assays (Fournier *et al.*, 2014). As described in Chapter 1 [1.5.2.2.3], hundreds of thousands of bacterial genomes have now been sequenced. However, WGS data are most commonly available only as short reads in the SRA, which limits use to laboratories with bioinformatics expertise and resources. To address this issue, PubMLST databases (http://pubmlst.org) have begun to make large numbers of *de novo* assembled bacterial genomes publicly available (Jolley & Maiden, 2010; Maiden *et al.*, 2013). As of January 2016, the Ribosomal Multilocus Sequence Typing (rMLST) Database (http://pubmlst.org/rmlst/) (Jolley *et al.*, 2012b) contained over 130,000 genomes, corresponding to more than 3,400 bacterial species, and an increasing number of species-specific PubMLST databases are also being populated with WGS data.

The current study shows how diagnostic assays can be evaluated *in silico* by exploiting bacterial WGS data. Infectious gastroenteritis poses a particular diagnostic challenge, as causative agents include a wide range of bacteria, viruses, and parasites (Whyte *et al.*, 2015). As a result, molecular diagnostics for

gastrointestinal pathogens, particularly multiplex NAATs, are becoming increasingly popular because they are sensitive and specific, and they are also less labour-intensive, with decreased turnaround times (Platts-Mills *et al.*, 2013; Zhang *et al.*, 2015). This is especially true for the detection of *Campylobacter*, a leading cause of bacterial gastroenteritis (Kaakoush *et al.*, 2015). *Campylobacter* species are fastidious and difficult to identify (Fitzgerald, 2015); therefore, numerous NAATs have been developed for their detection (On & Jordan, 2003; Fitzgerald, 2015; Zhang *et al.*, 2015). It is of concern that clinical *Campylobacter* isolates are genetically diverse (Cody *et al.*, 2013) and that interspecies recombination occurs among members of the genus, particularly between the clinically relevant *C. jejuni* and *C. coli* (Sheppard *et al.*, 2008, 2011, 2013b). As the extent to which these factors affect NAATs is unknown, *Campylobacter* was selected as an exemplar for use in the current study.

A duplex Taqman real-time PCR (RT-PCR) used by Public Health England for the speciation of *C. jejuni* and *C. coli* was chosen as a case study (Best *et al.*, 2003). The assay and variations thereof have been widely used for routine speciation (Best *et al.*, 2003; Schuurman *et al.*, 2007; Van Lint *et al.*, 2015), studies of *Campylobacter* from humans (Sopwith *et al.*, 2006; Amar *et al.*, 2007; O'Brien *et al.*, 2010; Mason *et al.*, 2013), animals (Griggs *et al.*, 2005; Humphrey *et al.*, 2005; Bull *et al.*, 2006; Ridley *et al.*, 2008; Kwan *et al.*, 2008; Sheppard *et al.*, 2010b; Ridley *et al.*, 2011b, 2011a; Rapp *et al.*, 2012; Kalupahana *et al.*, 2013), and the environment (Oster *et al.*, 2014), and outbreak investigations (Edwards *et al.*, 2014). For *C. jejuni*, the RT-PCR target is *mapA,* which encodes a putative outer membrane lipoprotein

(Stucki *et al.*, 1995) shown to be immunogenic in chickens (Shoaf-Sweeney *et al.*, 2008; Johnson *et al.*, 2014). For *C. coli*, the target is *ceuE*, which encodes a periplasmic binding protein involved in iron scavenging (Richardson & Park, 1995). As *mapA* and *ceuE* are present in both organisms, the RT-PCR is contingent on the conservation of species-specific primer- and probe-binding sequences; however, the assay was developed in the pre-WGS era using single gene sequences (Best *et al.*, 2003). Accordingly, the *mapA*/*ceuE* RT-PCR assay provided an opportunity to demonstrate the utility of *in silico* evaluations of molecular diagnostics using only PubMLST databases and open-source software.

2.2 Materials and Methods

2.2.1 Whole-genome sequencing data

WGS data from 1,724 *Campylobacter* isolates were accessed via the *Campylobacter jejuni*/*coli* PubMLST database (http://pubmlst.org/campylobacter/). These *Campylobacter* comprised all single patient isolates recovered in Oxfordshire, UK, between June 2011 and June 2013, and were part of the Oxfordshire sentinel surveillance collection (Cody *et al.*, 2012, 2013).

2.2.2 Genome annotation and data extraction

Based on the gene-by-gene approach to population annotation described in Chapter 1 [1.5.2.2.3.3], BIGSDB software (Jolley & Maiden, 2010) implemented in the *Campylobacter jejuni*/*coli* PubMLST database was used to identify *mapA* and *ceuE*, as well as the seven MLST (Dingle *et al.*, 2001a, 2005) and 52 rMLST loci (Jolley *et al.*, 2012b), within the study genomes. This process was completed in part using the BIGSDB autotagger functionality, which automatically curated sequences with ≥ 99% identity to existing alleles for the loci of interest; less similar sequences were curated manually. Following annotation, two isolates were excluded from further analysis because they had multiple ribosomal sequence types (rSTs), suggesting a mixed culture. An additional nine isolates were excluded because they had incomplete MLST and/or rMLST profiles (n = 8), or misassembled *ceuE* sequences (n = 1), which suggested low quality WGS data. The 1,713 included isolates (Appendix A) can be found under the project "*mapA*/*ceuE* Evaluation" in the *Campylobacter jejuni*/*coli* PubMLST isolate database. Allelic data and

corresponding nucleotide sequences, STs, clonal complexes, and rSTs were extracted from the database using BIGSDB data export plugins (http://bigsdb.readthedocs.org/en/latest/) (Jolley & Maiden, 2010).

2.2.3 Isolate diversity and speciation

Allelic diversity of MLST and rMLST data was determined using the bias-corrected version of Simpson's index of diversity (*D*) (Simpson, 1949; Hunter & Gaston, 1988) with 95% confidence intervals (CIs) (Grundmann *et al.*, 2001). Possible values of *D* ranged from 0 (no diversity) to 1 (maximum diversity). The distribution of MLST clonal complexes was compared to that observed for 3,349 human disease isolates recovered in Oxfordshire between 2003 and 2009 (Cody *et al.*, 2012). Isolates were assigned to species groups using rMLST (Jolley *et al.*, 2012b): concatenated nucleotide sequences of unique rSTs (~20,780 bp) were aligned with MAFFT version 7.037b (Katoh & Standley, 2013), and a Neighbor-joining tree (Saitou & Nei, 1987) was generated in MEGA version 5.05 (Tamura *et al.*, 2011), using the Kimura 2-parameter model (Kimura, 1980) with 1,000 bootstrap replicates. Additional rMLST analyses were carried out to resolve the assignment of a subset of isolates to *C. coli* clades 1, 2, and 3: using the approach described above, rSTs of interest were compared to a reference set of 15 well-characterised publicly available genomes representative of the three clades (Appendix B).

2.2.4 *In silico* assay evaluation

Alignments of unique *mapA* and *ceuE* alleles were produced as for the rMLST phylogeny, and regions corresponding to the forward primer, probe, and reverse primer were extracted. Primer and probed nucleotide sequence fragments were concatenated, and unique combinations were assigned allele numbers in order of discovery. Alignments of unique combinations and published primer and probe sequences (Best *et al.*, 2003) were inspected to investigate inter- and intraspecies diversity.

2.2.5 RT-PCR confirmation of results of *in silico* evaluation

Archived genomic DNA and bacterial cultures were available for all *Campylobacter* isolates from the Oxfordshire sentinel surveillance collection sent for WGS (Cody *et al.*, 2013), which facilitated RT-PCR confirmation of the *in silico* evaluation results. Representative isolates (*n* = 124) were chosen for RT-PCR such that each unique *mapA* and *ceuE* forward primer, probe, and reverse primer combination was tested at least once, with the subset also representative of the genetic diversity of the study dataset. For isolates with insufficient archived genomic DNA (*n* = 5), glycerol stocks of single-colony cultures were inoculated onto Columbia agar with horse blood (CBA) (Oxoid Ltd., Basingstoke, UK) and incubated at 42°C in a microaerobic atmosphere for 48 hours. Boiled cell lysates were prepared from single colonies as previously described (Best *et al.*, 2003). RT-PCR was carried out according to the method of Best *et al*. (2003).

2.2.6 Introgression, diversity, and selection in RT-PCR targets

Additional analyses were carried out at the whole-gene level to assess the impact of introgression, genetic diversity, and selection on assay specificity. Individual *mapA* and *ceuE* alignments and gene phylogenies were generated as described for rMLST. STRUCTURE (Pritchard *et al.*, 2000), a Bayesian clustering algorithm previously used to characterise introgression in *C. jejuni* and *C. coli* (Sheppard *et al.*, 2008, 2011), was used to analyse *mapA* and *ceuE* nucleotide sequences at the population level. Isolates were probabilistically assigned to species using the linkage model, which adjusts for linkage disequilibrium between nucleotides (Falush *et al.*, 2003). The model was run with default settings for 10,000 burn-in iterations and 10,000 additional iterations, assuming a population number (*k*) of 2. Putative mosaic alleles were identified as those with ≤0.75 probability of belonging to either *C. jejuni* or *C. coli* (Sheppard *et al.*, 2011). Site-by-site frequencies generated by STRUCTURE were used to identify nucleotide sequence fragments with different ancestries in putative mosaic alleles (Falush *et al.*, 2003; Sheppard *et al.*, 2011).

Patterns of allelic and nucleotide diversity were examined for both *mapA* and *ceuE*. To examine diversity at the protein level, nucleotide sequences were translated using MEGA version 5.05 (Tamura *et al.*, 2011) and allele numbers were assigned to unique peptide sequences. After putative recombinant alleles were excluded, within- and between-group *p*-distances were calculated for *C. jejuni*- and *C. coli*-specific gene and protein sequences using DnaSP version 5.10 (Librado & Rozas, 2009). Species-specific synonymous and non-synonymous substitution rates (d_N/d_S)

were calculated for *mapA* and *ceuE* alleles encoding full-length protein sequences using SNAP version 2.1.1 (www.hiv.lanl.gov) (Korber, 2000).

2.3 Results

2.3.1 Isolate diversity and speciation

Complete nucleotide sequences for *mapA* and *ceuE*, as well as MLST and rMLST loci, were obtained from 1,713/1,724 (99.4%) *Campylobacter* isolates from the Oxfordshire sentinel surveillance collection. MLST and rMLST data indicated that the isolate collection was genetically diverse, including 293 STs [*D*=0.974 (95% CIs 0.972-0.976)] and 597 rSTs [*D*=0.989 (95% CIs 0.988-0.991)]. Isolates belonged to 33 MLST clonal complexes, with proportions similar to those previously observed in Oxfordshire (Cody *et al.*, 2013) (Figure 2.1A).

Isolates were assigned to species groups using rMLST, which differentiates bacterial species based on variation in the ribosomal protein (*rps*) genes (Jolley *et al.*, 2012b). Two clusters, corresponding to *C. jejuni* and *C. coli*, were visible on the rMLST phylogeny (Figure 2.1B). *C. jejuni* accounted for 1,521 (88.8%) isolates and *C. coli* for the remaining 192 (11.2%), which was comparable to proportions previously reported in Oxfordshire (Cody *et al.*, 2012, 2013). Of interest were two *C. coli* rSTs, rST-398 (*n* = 2) and rST-4701 (*n* = 1), that were distinct from other *C. coli* sequences and occurred at the tips of long branches (Figure 2.1B). At the population level, *C. coli* segregates into three clades. Clade 1 accounts for the majority of human disease and agricultural isolates and has the highest rates of introgression with *C. jejuni*, while isolates from clades 2 and 3 are generally from environmental sources (Sheppard *et al.*, 2008, 2010a). Further rMLST analyses including 15 well-characterised representatives of each of the three clades (Appendix B) suggested

that the rST-398 and rST-4701 isolates corresponded to clade 3 *C. coli* (Figure 2.1C).

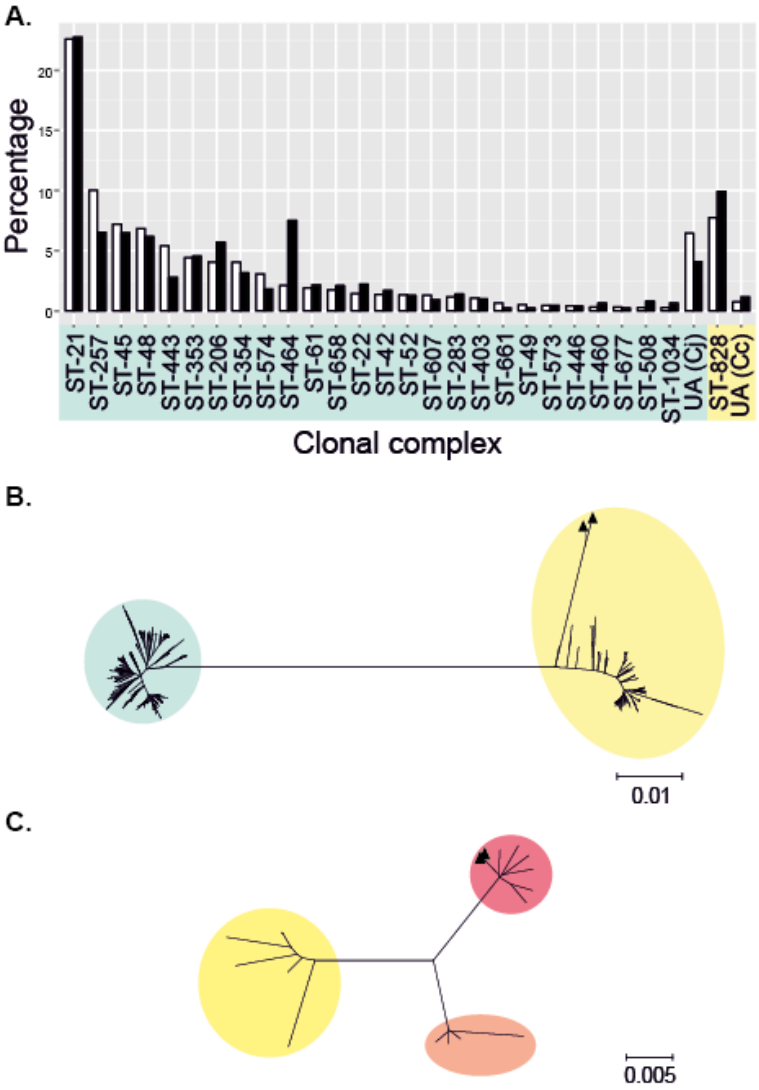


Figure 2.1 Genetic diversity and speciation of *Campylobacter* isolates from the Oxfordshire sentinel surveillance collection. (A) Frequency distribution of major MLST clonal complexes (defined as those with ≥ 10 isolates) present among 3,349 human disease isolates (collected 2003-2009) that were typed using MLST (white), and 1,713 *Campylobacter* genomes (2011-2013) that were included in the current study (black). UA, STs unassigned to a clonal complex; Cj, *C. jejuni* (blue); Cc, *C. coli* (yellow). (B) Neighbor-joining tree based on concatenated nucleotide sequences of unique rMLST profiles (*n* = 597) identified among the study genomes, showing the separation of *C. jejuni* (blue) and *C. coli* (yellow) genotypes. ▲, putative clade 3 *C. coli*. (C) Neighbor-joining tree based on concatenated nucleotide sequences of rMLST profiles of 15 representative isolates belonging to *C. coli* clades 1 (yellow), 2 (orange), and 3 (pink), and three putative clade 3 isolates from Oxfordshire (▲).

2.3.2 In silico assay evaluation

Both RT-PCR target genes were present in all study isolates. The collection included 72 *mapA* alleles of 645 bp and 126 *ceuE* alleles of 990-994 bp. Differences in *ceuE* allele lengths were mainly due to variation in three homopolymeric tracts, which resulted in internal stop codons in 11 *C. jejuni*-specific alleles (*n* = 20) (Table 2.1). Given that the specificity of the *mapA*/*ceuE* assay is dependent on conservation of species-specific primer- and probe-binding sequences, these regions were analysed in further detail.

Table 2.1 Details of *ceuE* alleles with internal stop codons identified among *C. jejuni* isolates

Polymorphism	Nucleotide position	Putative effect on protein	PubMLST gene allele	ST/CC/rST[a] (*n*)
T(8 -> 7)	34	Truncation	288	5756/UA[b]/263 (1)
			290	2844/ST-460/325 (2)
			291	48/ST-48/106 (2); 48/ST-48/98 (1); 48/ST-48/99 (1); 520/ST-21/377 (1)
			293	2274/UA/123 (1)
			294	443/ST-443/221 (1)
			296	5707/UA/3509 (1)
			298	1932/ST-460/4596 (2)
			300	464/ST-464/7025 (1)
T(8 -> 9)	34	Truncation	295	21/ST-21/538 (1)
A(5 -> 4)	202	Truncation	289	47/ST-21/510 (3); 3633/ST-21/510 (1)
T(6 -> 5)	483	Truncation	292	53/ST-21/460 (1)
Deletion (C)	675	Truncation	297	257/ST-257/186 (2)

[a] ST, sequence type; CC, clonal complex; rST, ribosomal ST.

[b] UA, not assigned to a clonal complex.

2.3.2.1 mapA

Twenty-three unique *mapA* primer and probe combinations were identified among the study isolates: twelve were present only in isolates designated as *C. jejuni*; nine in *C. coli*; and two were shared. Two distinct groups consistent with microbiological species were evident from the nucleotide sequence alignment of these unique combinations (Figure 2.2A). Primer and probe sequences were conserved among *C. jejuni* isolates. The predominant combination was identical to published primer and probe sequences (Best *et al.*, 2003) and occurred in 1,166 (76.7%) isolates. Nucleotide variation among divergent *C. jejuni* combinations was limited to between one and five polymorphisms across the three regions (Figure 2.2A).

Unlike *C. jejuni*, a subset of *C. coli* isolates (*n* = 4) carried complete or partial non-specific *mapA* primer and probe combinations (Figure 2.2A; Table 2.2). Two combinations, present in a single *C. coli* isolate each, corresponded to predominant *C. jejuni*-specific alleles 1 and 2. The remaining two non-specific primer and probe combinations were composites of *C. coli* and *C. jejuni* sequences; allele 22 included *C. jejuni*-specific forward primer sequence, while allele 23 included *C. jejuni*-specific forward primer and probe sequences (Figure 2.2A; Table 2.2). With the exception of putative introgressed sequences, *C. coli mapA* primer and probe combinations were conserved, yet divergent from *C. jejuni* sequences. Nucleotide differences included nine fixed mutations, as well as 11 *C. coli*-specific polymorphisms (Figure 2.2A).

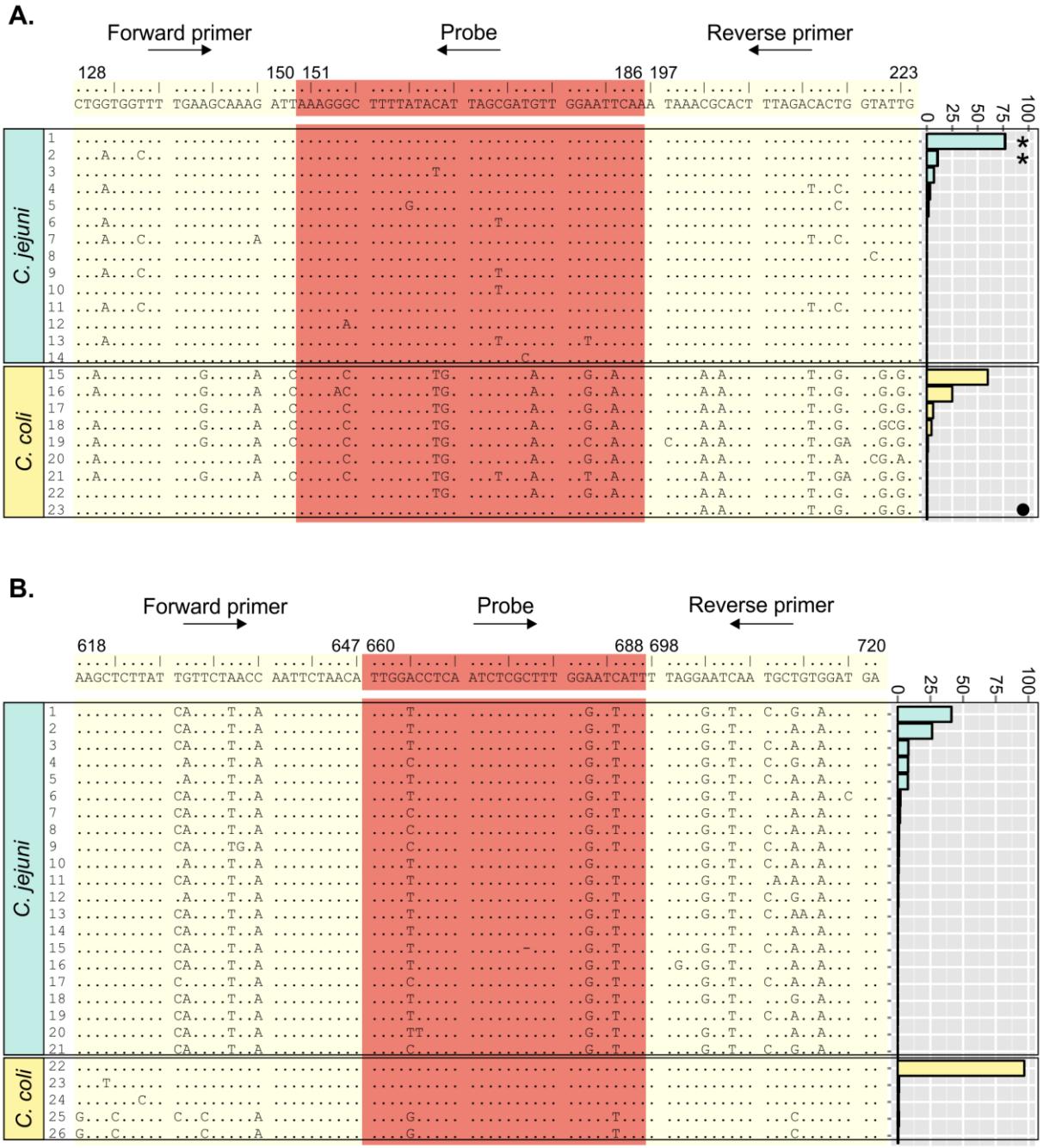


Figure 2.2 Genetic diversity of concatenated *mapA* (A) and *ceuE* (B) primer and probe sequences. Published sequences (Best *et al.*, 2003) are shown above nucleotide sequence alignments of unique primer (cream) and probe (orange) combinations, with dots indicating conserved nucleotides. Adjacent histograms indicate frequencies of combinations in *C. jejuni* (blue) and *C. coli* (yellow). *, complete *C. jejuni*-specific combination detected in a single *C. coli* isolate; ●, composite non-specific combination detected in a single *C. coli* isolate.

Table 2.2 Details of *C. coli* isolates with atypical *mapA*/*ceuE* RT-PCR primer and probe sequences

Atypical target	Isolate	ST (CC)[a]	Gene allele / primer and probe combination (RT-PCR[b] result)[c]		*mapA*/*ceuE* RT-PCR result
			mapA	*ceuE*	
mapA	OXC7352	6973 (ST-1150)	20/2 (+)	136/22 (+)	Mixed
	OXC6987	825 (ST-828)	88/1 (+)	3/22 (+)	Mixed
	OXC6395	1487 (ST-1150)	19/22[e] (-)	17/22 (+)	*C. coli*
	OXC7615	6760 (UA[d])	111/23[f] (late +)	139/22 (+)	Inconclusive
ceuE	OXC7241	6698 (UA)	96/19 (-)	153/25 (late +)	Inconclusive
	OXC7243	6698 (UA)	96/19 (-)	153/25 (late +)	Inconclusive
	OXC7653	6975 (UA)	114/21 (-)	183/26 (late +)	Inconclusive

[a] ST, sequence type; CC, clonal complex.

[b] RT-PCR, real-time PCR.

[c] (+), target detected; (-), target not detected; (late +), target detected after cycle 30.

[d] UA, ST unassigned to a clonal complex.

[e] Forward primer *C. jejuni*-specific; probe and reverse primer *C. coli*-specific.

[f] Forward primer and probe *C. jejuni*-specific; reverse primer *C. coli*-specific.

2.3.2.2 ceuE

In contrast to *mapA*, the 26 *ceuE* primer and probe combinations identified among the study isolates were all species-specific: 21 were present only in *C. jejuni* and five only in *C. coli*. Primer and probe combinations were also stratified by species at the nucleotide sequence level (Figure 2.2B). *C. coli* sequences were highly conserved, with 186 (96.9%) isolates identical to the published primer and probe sequences (Best *et al.*, 2003). Nucleotide variation was limited to between one and eight polymorphisms, the majority of which occurred in alleles 25 (n = 2) and 26 (n = 1)

(Figure 2.2B; Table 2.2). These alleles were found in the putative clade 3 *C. coli* isolates and were distinct from other *C. coli* and *C. jejuni* sequences. *C. jejuni* primer and probe combinations were divergent from the published sequences (Best *et al.*, 2003) and more diverse than those found in *C. coli*. *C. jejuni*-specific sequences contained 10 to 12 polymorphisms each, including five fixed differences and 11 species-specific mutations. Additionally, *C. jejuni* isolates were more evenly distributed across primer and probe combinations than *C. coli* isolates (Figure 2.2B).

2.3.2.3 Predicted performance and RT-PCR confirmation

Based on primer and probe combinations, the *mapA*/*ceuE* RT-PCR assay was predicted to speciate 1,707/1,713 (99.7%) isolates correctly, including 1,521 (100%) *C. jejuni* and 186 (96.9%) *C. coli*. RT-PCR testing of 124 representative isolates confirmed these predictions, with positive results defined as those with a maximum cycle threshold (C_T) value of 30 (Best *et al.*, 2003). As expected, the two *C. coli* isolates with complete *C. jejuni*-specific *mapA* primer and probe sequences were *mapA*-positive/*ceuE*-positive (Table 2.2). Results for the isolate with *C. jejuni*-specific *mapA* forward primer and probe sequences and the three clade 3 *C. coli* isolates were inconclusive due to late detection of *mapA* and *ceuE*, respectively (C_T values of 32-37) (Table 2.2; Table 2.3). While this study was not designed to quantify the effects of primer and probe mismatches on target detection, there was a clear correlation between the number of polymorphisms and C_T value (Table 2.3).

Table 2.3 RT-PCR cycle thresholds for *mapA* and *ceuE* primer and probe combinations

Gene	Primer and probe allele (*n*[a])	C_T[b]			Polymorphisms (*n*)[c]
		Maximum	Minimum	Average	
mapA	1 (68)	22	12	15	0-0-0
	2 (11)	20	15	17	2-0-0
	3 (3)	19	15	17	0-1-0
	4 (3)	17	16	16	1-0-2
	5 (4)	17	15	16	0-1-1
	6 (1)	-	-	14	1-1-0
	7 (2)	27	25	26	3-0-2
	8 (1)	-	-	12	0-0-1
	9 (2)	21	17	19	2-1-0
	10 (1)	-	-	14	0-1-0
	11 (1)	-	-	19	2-0-2
	12 (1)	-	-	14	0-1-0
	13 (1)	-	-	15	1-2-0
	14 (1)	-	-	15	0-1-0
	15 (9)	N/D[d]	N/D	N/D	4-6-6
	16 (3)	N/D	N/D	N/D	4-7-6
	17 (3)	N/D	N/D	N/D	3-6-6
	18 (2)	N/D	N/D	N/D	4-6-7
	19 (1)	N/D	N/D	N/D	4-6-8
	20 (1)	N/D	N/D	N/D	2-6-7
	21 (1)	N/D	N/D	N/D	4-7-7
	22 (1)	N/D	N/D	N/D	0-5-6
	23 (1)	-	-	33.29	0-0-6
ceuE	1 (18)	N/D	N/D	N/D	4-3-5
	2 (29)	N/D	N/D	N/D	4-3-4
	3 (8)	N/D	N/D	N/D	4-3-5
	4 (7)	N/D	N/D	N/D	3-3-5
	5 (12)	N/D	N/D	N/D	3-3-5
	6 (1)	N/D	N/D	N/D	4-3-5
	7 (1)	N/D	N/D	N/D	4-3-4
	8 (3)	N/D	N/D	N/D	4-3-5
	9 (1)	N/D	N/D	N/D	4-3-5

Table 2.3 cont.

Gene	Primer and probe allele (*n*[a])	C_T[b] Maximum	C_T[b] Minimum	C_T[b] Average	Polymorphisms (*n*)[c]
ceuE	10 (1)	N/D	N/D	N/D	3-2-5
	11 (1)	N/D	N/D	N/D	4-3-5
	12 (2)	N/D	N/D	N/D	3-3-5
	13 (1)	N/D	N/D	N/D	4-3-6
	14 (1)	N/D	N/D	N/D	4-3-3
	15 (1)	N/D	N/D	N/D	4-4-5
	16 (2)	N/D	N/D	N/D	4-3-5
	17 (2)	N/D	N/D	N/D	3-3-5
	18 (1)	N/D	N/D	N/D	4-3-4
	19 (1)	N/D	N/D	N/D	4-3-4
	20 (1)	N/D	N/D	N/D	4-4-4
	21 (1)	N/D	N/D	N/D	4-3-5
	22 (19)	22	12	15	0-0-0
	23 (1)	-	-	13	1-0-0
	24 (1)	-	-	6	1-0-0
	25 (1)	-	-	37	5-2-1
	26 (1)	-	-	32	4-2-1

[a] Number of representative isolates tested.

[b] C_T, cycle threshold.

[c] Number of polymorphisms in (forward primer - probe - reverse primer).

[d] N/D, not detected.

2.3.3 Introgression, diversity, and selection in RT-PCR targets

To explore the impact of introgression, diversity, and selection on assay specificity, additional analyses were carried out at the whole-gene level. Individual gene phylogenies confirmed that *mapA* and *ceuE* alleles were predominantly species-specific (Figure 2.3). Clade 3 *C. coli mapA* and *ceuE* alleles clustered with other *C. coli* alleles; however, they were distinct from clade 1 alleles and were on long branches, indicative of genetic divergence. The only sequences found in both species were *C. jejuni mapA* alleles 20 and 88, which accounted for two *mapA*-

positive/*ceuE*-positive *C. coli* isolates (Figure 2.3A; Table 2.2). Noteworthy were five *C. coli* alleles that occupied intermediate positions on the *mapA* phylogeny (Figure 2.3A). Taken together, these findings suggested introgression in *mapA*, which supported the *in silico* evaluation results.

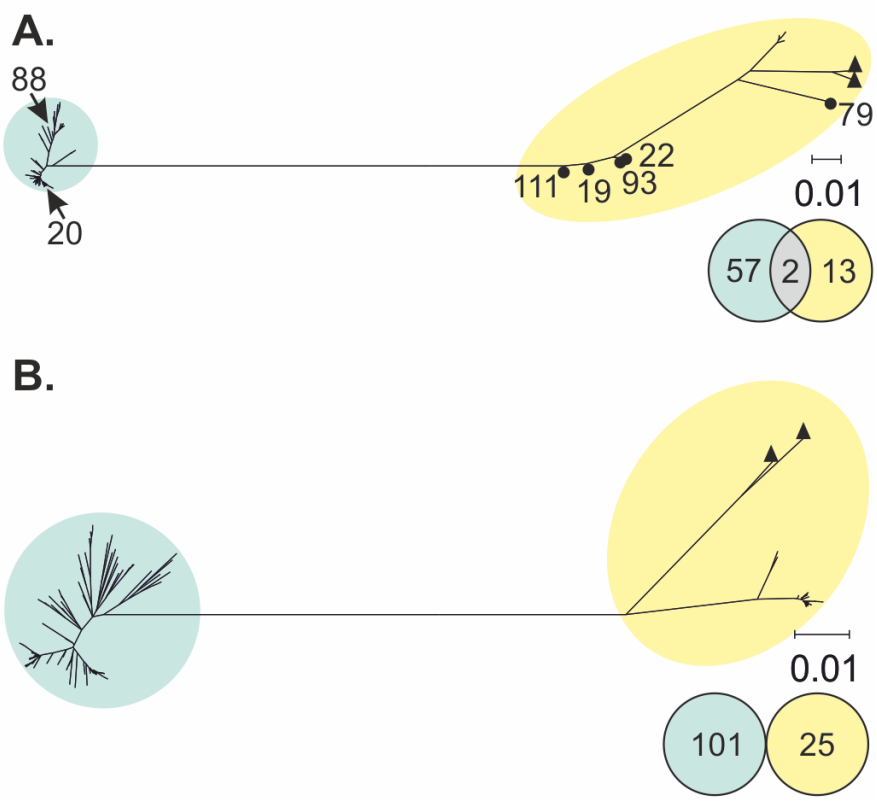


Figure 2.3 Neighbor-joining trees based on 72 *mapA* (A) and 126 *ceuE* (B) unique gene sequences. Adjacent Venn diagrams indicate the number of species-specific and shared alleles. →, *C. jejuni*-specific alleles detected in *C. coli* isolates; ●, putative introgressed alleles; ▲, putative clade 3 *C. coli* alleles. Key: *C. jejuni*, blue; *C. coli*, yellow; shared alleles, grey.

Introgression can result in the transfer of complete genes (whole-allele replacement) or the generation of mosaic alleles. Introgression in RT-PCR target genes was formally characterised using STRUCTURE, a model-based clustering algorithm, which

detected mixed ancestry only in the *mapA* gene of 17 (8.9%) *C. coli* isolates. In addition to the two previously identified whole-allele replacements, five putative mosaic alleles were detected (*n* = 15) (Figure 2.4A). All imported DNA was identical to the predominant *C. jejuni* sequence. Recombination breakpoints occurred within the RT-PCR target region in four introgressed alleles, of which alleles 19 and 111 corresponded to composite primer and probe sequences (Figure 2.4B). Alleles 22 (*n* = 11) and 93 (*n* = 1) were not identified as introgressed sequences during the *in silico* evaluation because the putative breakpoints occurred at the 5' end of the forward primer (Figure 2.4B).

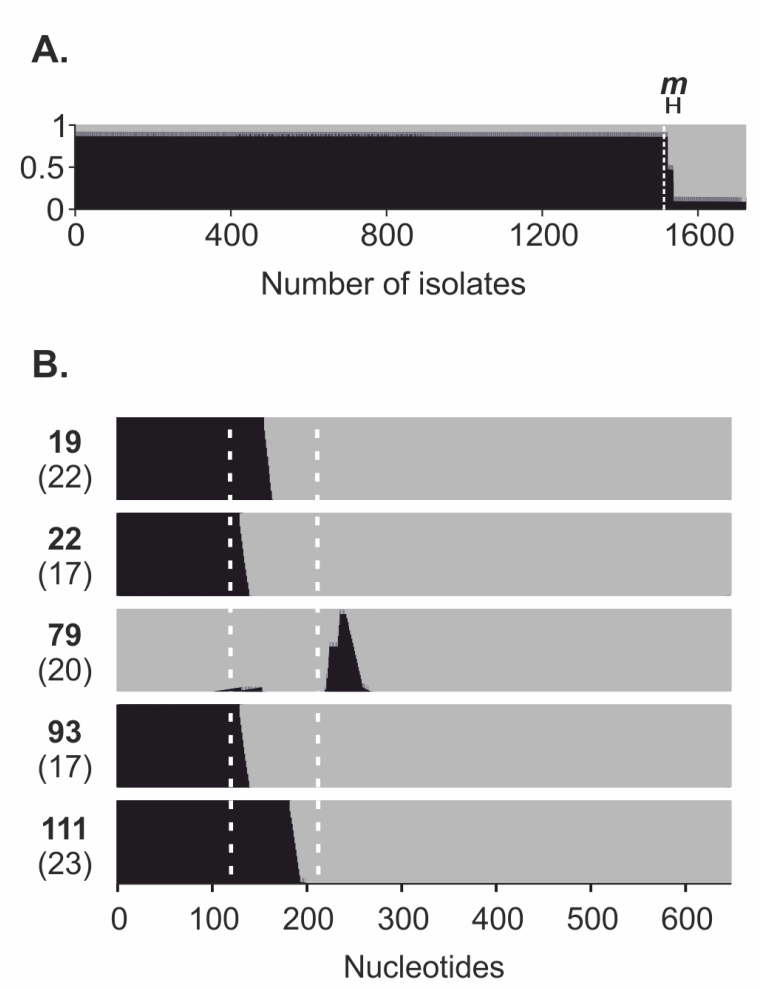


Figure 2.4 Characterisation of introgression in *mapA*. (A) Probabilistic assignment of study isolates to species based on STRUCTURE analysis of *mapA* nucleotide sequences using the linkage model. Putative mosaic sequences (*m*) were identified as those with ≤0.75 probability of belonging to either *C. jejuni* or *C. coli*. Each isolate is represented by a vertical line, with shading indicative of the proportion of *C. jejuni* (black) or *C. coli* (grey) ancestry. The dashed white line indicates the species boundary as determined by rMLST. (B) Recombination breakpoints in putative mosaic alleles were inferred using site-by-site nucleotide ancestries generated by STRUCTURE. Bar plots represent individual putative mosaic sequences, with whole-gene allele numbers and corresponding primer and probe combinations shown in bold and parentheses, respectively. Vertical lines represent individual nucleotides with shading indicative of ancestry as in (A). Dashed white lines demarcate the region amplified by *mapA* primers (Best *et al.*, 2003).

Overall *p*-distances within and between *C. jejuni*- and *C. coli*-specific sequences indicated the high levels of intraspecies conservation and interspecies diversity in *mapA* and *ceuE* (Table 2.4). That gene and protein diversity at the population level were largely due to an abundance of rare alleles provided further evidence of intraspecies conservation (Figure 2.5). Average d_N/d_S ratios for *mapA* and *ceuE* were <1 (0.077-0.14) in *C. jejuni* and *C. coli*, consistent with both genes being under stabilising selection; however, the distribution of synonymous and non-synonymous substitutions suggested species-specific differences in *mapA* and *ceuE* evolution (Figure 2.6).

Table 2.4 Intra- and interspecies diversity of *mapA* and *ceuE* gene and protein sequences

	mapA	***ceuE***
Gene sequence diversity		
C. jejuni	0.014	0.018
C. coli	0.023	0.005
C. jejuni/*C. coli*	0.232	0.13
Protein sequence diversity		
C. jejuni	0.009	0.015
C. coli	0.020	0.009
C. jejuni/*C. coli*	0.232	0.089

Values indicate *p*-distances. Putative recombinant sequences and alleles encoding truncated peptide sequences were excluded.

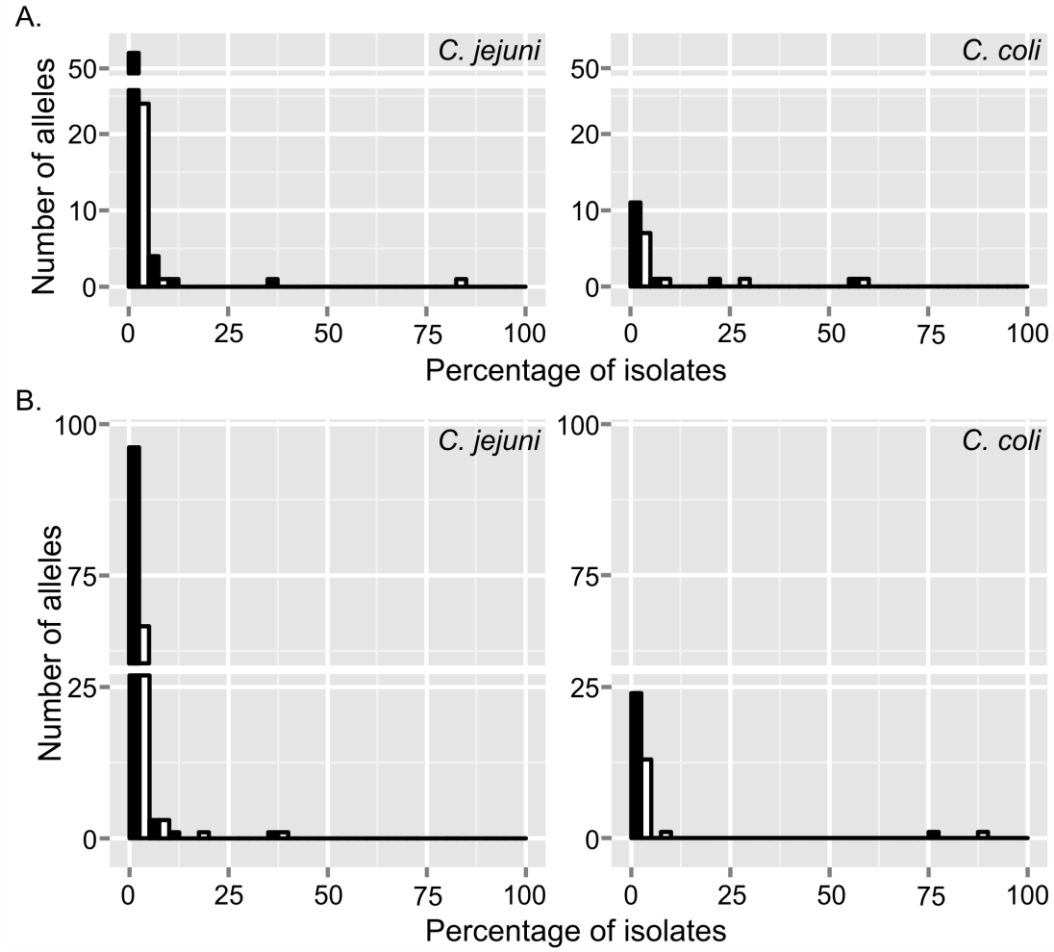


Figure 2.5 Frequency distribution of *mapA* (A) and *ceuE* (B) nucleotide (black) and protein alleles (white) in *C. jejuni* and *C. coli*.

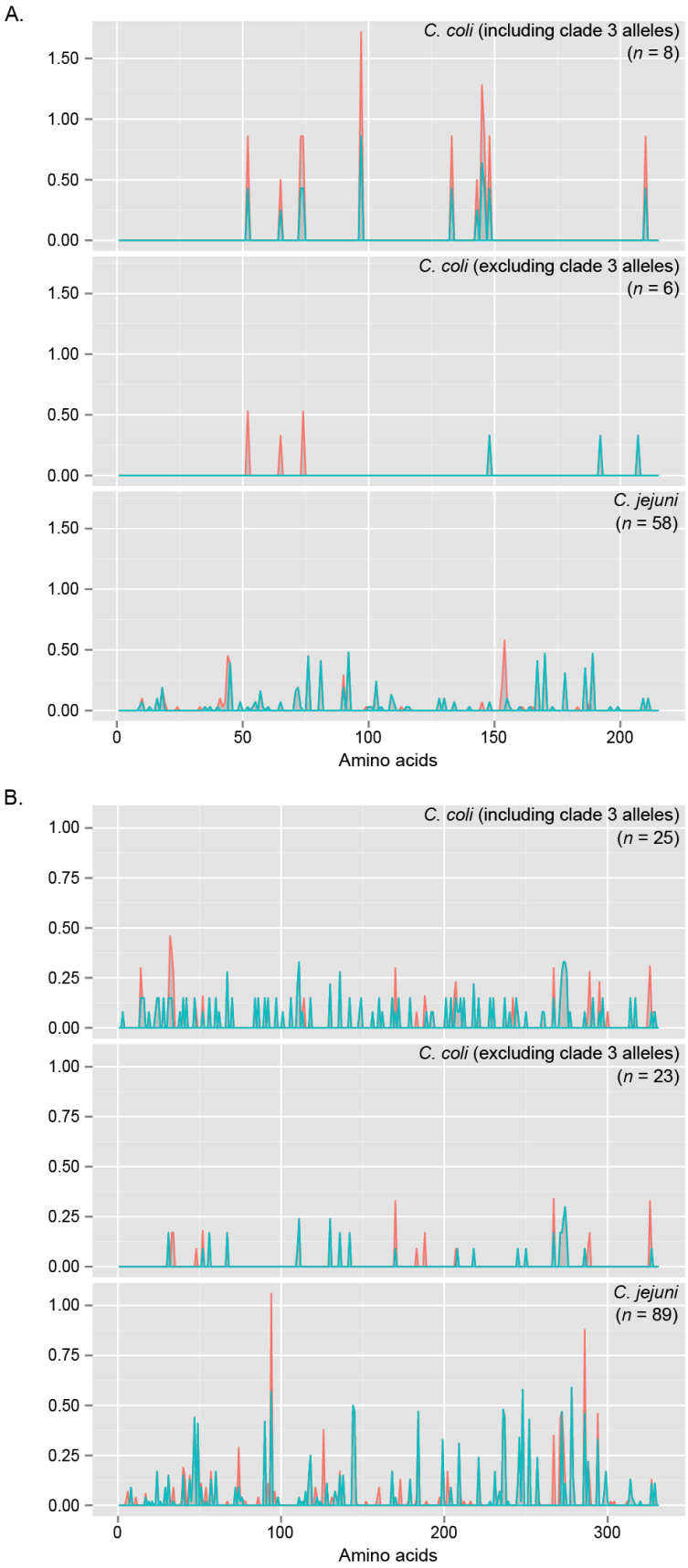


Figure 2.6 Distribution of synonymous and non-synonymous substitutions across *mapA* (A) and *ceuE* (B) in *C. jejuni* and *C. coli* as determined using SNAP (www.hiv.lanl.gov) (Korber, 2000)**.** The *y*-axis indicates the average rate of synonymous (blue) and non-synonymous (red) substitutions per codon across all pairwise comparisons. Sites under positive selection were defined as those with $d_N > d_S$.

2.4 Discussion

In the future, identification, antimicrobial susceptibility testing, and epidemiological typing of bacterial pathogens could, in principle, all be carried out in a single step with WGS (Didelot *et al.*, 2012; Köser *et al.*, 2012b); however, uptake of WGS by diagnostic laboratories will be gradual, particularly because bacterial culture will remain necessary while methods for direct testing of clinical specimens are in development (Lasken & McLean, 2014; Loman & Watson, 2015). In the interim, laboratories with and without access to WGS will likely continue to use NAATs as either preliminary or preferred tests for rapid diagnosis of infectious diseases. It is essential that NAATs be rigorously evaluated (Banoo *et al.*, 2010), and publicly available WGS data represent a valuable resource for this purpose.

The current study demonstrates how publicly available WGS data can be leveraged for *in silico* evaluations of NAATs, with the *mapA*/*ceuE* RT-PCR for speciation of *C. jejuni* and *C. coli* (Best *et al.*, 2003) chosen as a case study. Best *et al*. (2003) validated the *mapA*/*ceuE* assay using clinical *Campylobacter* isolates from the UK; therefore, because previous studies have shown that *Campylobacter* genotypes circulating in Oxfordshire are representative of the UK (Sopwith *et al.*, 2006; Sheppard *et al.*, 2009a; Cody *et al.*, 2012, 2013) and other high-income countries (http://pubmlst.org/campylobacter/), the Oxfordshire sentinel surveillance collection (Cody *et al.*, 2013) was identified as an appropriate source of WGS data for this study. The *in silico* evaluation included 1,713 genetically diverse *Campylobacter* isolates from Oxfordshire, which were confirmed to be representative of human disease genotypes from the UK (Figure 2.1).

Analyses of primer and probe sequences, in conjunction with RT-PCR, confirmed that the *mapA*/*ceuE* assay was robust. Overall, 1,707 (99.7%) isolates were correctly speciated, which was similar to the 97.7% level of accuracy reported during validation (Best *et al.*, 2003). Assay specificity could be attributed to a combination of interspecies diversity and marked intraspecies conservation within primer and probe sequences, as well as over-representation of combinations identical to published primers and probes (Figure 2.2).

Experimental evidence suggests that *Campylobacter mapA* and *ceuE* genes play roles in colonisation (Johnson *et al.*, 2014) and iron-acquisition (Miller *et al.*, 2009), respectively. For both genes, mutants showed reduced potential for chicken colonisation (Palyada *et al.*, 2004; Johnson *et al.*, 2014); therefore, it was unsurprising that intraspecies diversity was low at the gene and protein levels in *mapA* and *ceuE* (Table 2.4) and that both targets were under stabilising selection. Species-specific differences in the distribution of synonymous and non-synonymous substitutions in *mapA* and *ceuE* hint at divergent evolution in *C. jejuni* and *C. coli* post-speciation (Figure 2.6), perhaps due to host niche differences (Rosef *et al.*, 1983; Waldenström *et al.*, 2002; McCarthy *et al.*, 2007; Sheppard *et al.*, 2010b, 2010a). Interestingly, predicted protein sequences of 11 *C. jejuni*-specific *ceuE* alleles (n = 20) were truncated due to variation in three homopolymeric tracts, one of which occurred at the 5′ end of the gene (Table 2.1). Homopolymeric tracts are common in the *C. jejuni* genome and regulate gene expression through phase variation (Parkhill *et al.*, 2000). It is possible that *ceuE* may also be phase variable.

Importantly, although a small proportion of isolates could not be speciated using the *mapA*/*ceuE* assay, none was incorrectly identified. Only six (3.1%) *C. coli* could not be speciated, including two *mapA*-positive/*ceuE*-positive isolates and four isolates with inconclusive results due to late detection of *mapA* (*n* = 1) or *ceuE* (*n* = 3) (C_T values 32-37) (Table 2.2; Table 2.3). Introgression was the underlying cause of *mapA* detection among *C. coli* isolates (Figure 2.4). Whole-allele replacements were relatively uncommon (1%), but they accounted for the *mapA*-positive/*ceuE*-positive isolates. Mosaic alleles were more prevalent (7.8%), but caused only one inconclusive RT-PCR result. The apparent lack of introgression in *ceuE* may be due to functional and combinatorial epistasis, as the gene is part of the *ceuBCDE* operon (Richardson & Park, 1995), the products of which form part of a siderophore-mediated iron scavenging system (Miller *et al.*, 2009). Concerning inconclusive RT-PCR results due to late detection of *ceuE*, the affected isolates corresponded to clade 3 *C. coli*. Phylogenetic analyses showed that these *ceuE* sequences were divergent from other *C. coli*-specific alleles (Figure 2.3B); an accumulation of mutations in the forward primer region reduced amplification efficiency (Figure 2.2B; Table 2.3), indicating a lack of assay specificity for clade 3 *C. coli*.

Taken together, the results of the *in silico* evaluation showed that the *mapA*/*ceuE* RT-PCR assay reliably speciates *C. jejuni* and *C. coli*, while whole-gene analyses provided insights into underlying reasons for the assay's specificity. The value of these findings also extends to other diagnostic assays that use *mapA* or *ceuE* as targets (Schuurman *et al.*, 2007; Fukushima *et al.*, 2009; de Boer *et al.*, 2010, 2013; McAuliffe *et al.*, 2013; Van Lint *et al.*, 2015). In the UK and other high-income

countries, the impact of RT-PCR failures observed in this study would be limited because: (i) *C. coli* accounts for a small proportion of human campylobacteriosis cases (Kaakoush *et al.*, 2015); (ii) RT-PCR outcome was unaffected in the majority of isolates with introgressed *mapA* alleles; and (iii) clade 2 and 3 *C. coli* rarely cause human disease (Sheppard *et al.*, 2010a). Given that host association supersedes geographic signals (Sheppard *et al.*, 2010b), it is likely that the assay will perform well in other regions that consume similar food animals to the UK; however, laboratories in regions with marked differences in *Campylobacter* epidemiology should exercise caution and validate the assay prior to use.

Finally, the *in silico* approach to assay evaluation used here could be extended to other NAATs or molecular diagnostic tests. Compared to Sanger sequencing, WGS represents an attractive alternative for studying primer sequences, particularly those with mismatches that adversely affect amplification efficiency. The approach outlined in this study could be used to evaluate existing assays, or it could be used in conjunction with primer design software during assay development. Importantly, the analyses require only a personal computer with Internet access and publicly available software.

CHAPTER 3

Evaluation of typing schemes for *in silico* identification of *Campylobacter* species

ABSTRACT

Accurate identification of bacterial species remains a challenge in both research and clinical microbiology laboratories. *In silico* speciation based on WGS data has the potential to improve this aspect of microbiology, particularly in the case of *Campylobacter* species, which are difficult to identify using conventional phenotypic and biochemical tests. The gene-by-gene approach implemented in BIGSDB software was used to evaluate four typing schemes for the *in silico* speciation of *Campylobacter*. A genus-wide core genome, defined using 70 finished genomes encompassing 22/27 known *Campylobacter* species, served as the basis of comparison for these evaluations. The typing schemes that were evaluated included: (i) two single locus approaches based on the 16S rRNA and *atpA* genes; and (ii) two multilocus approaches, namely a novel universal *Campylobacter* MLST scheme developed for this study, and rMLST. The *atpA* and rMLST phylogenies were highly similar to that of the core genome, indicating that these methods could be used for the rapid speciation of *Campylobacter* from WGS data. The results also highlighted the limitations of the 16S rRNA gene, suggesting that those reliant on *in vitro* sequence-based typing methods for *Campylobacter* speciation should consider using the *atpA* assay. This study demonstrates the utility of *in silico* speciation and provides further evidence of the power of the gene-by-gene approach and BIGSDB.

3.1 Introduction

Campylobacter continues to be a leading cause of bacterial gastroenteritis and a major public health problem: in 2010, campylobacteriosis accounted for 8.4% of all diarrhoeal disease and was associated with a global disease burden of 7.5 million disability-adjusted life years (Murray *et al.*, 2012). The majority of infections are attributed to *C. jejuni* and *C. coli*, but a further 17 *Campylobacter* species have been isolated from human disease samples (Kaakoush *et al.*, 2015). Although there is a growing interest in non-*jejuni/coli Campylobacter* species, these organisms remain poorly understood, primarily because they are difficult to isolate and identify (Fitzgerald, 2015).

The *Campylobacter* genus currently includes 27 species (Kaakoush *et al.*, 2015; Fitzgerald, 2015), and identification by conventional phenotypic and biochemical testing is challenging due to their taxonomic complexity and relative biochemical inertness (On, 2013). Molecular methods, particularly NAATs, have improved the detection and identification of *Campylobacter* (Fitzgerald, 2015); however, these tests must be modified when novel species are described, or when taxa are revised, and at present there is no single assay available for the differentiation of all *Campylobacter* species. Consequently, sequence-based typing methods have become a popular alternative to NAATs because the nucleotide sequences generated are unambiguous and can be used to identify existing and novel taxa, and also for epidemiological typing (On, 2013; Taboada *et al.*, 2013).

Single locus sequence-based typing methods typically identify a large number of taxa through the use of primers with a broad amplification range that target universally present yet variable loci, such as the 16S rRNA gene (Lawson *et al.*, 1997; Kulkarni, 2002; Gorkiewicz *et al.*, 2003), *gyrB* (Gunther *et al.*, 2011), and *atpA* (also referred to as *uncA*) (Miller *et al.*, 2014b). Among these methods, 16S rRNA gene sequencing has been most widely used, but studies suggest that the technique has important limitations. Distinct species, such as *C. jejuni* and *C. coli*, and *C. upsaliensis* and *C. helveticus*, share indistinguishable sequences, while other taxa, for instance *C. hyointestinalis* and *C. lanienae*, are difficult to identify due to high levels of intra-species diversity (Gorkiewicz *et al.*, 2003; Miller *et al.*, 2012; On, 2013; Miller *et al.*, 2014b).

Multi-locus approaches, such as MLST, have higher resolution than single locus methods, but differences in gene content and high levels of sequence diversity have impeded the development of a genus-wide *Campylobacter* scheme. Instead, there are currently ten distinct *Campylobacter* MLST schemes (http://pubmlst.org/campylobacter/), which encompass only 12 species (Dingle *et al.*, 2001a, 2005; Miller *et al.*, 2005; van Bergen *et al.*, 2005; Stoddard *et al.*, 2007; Miller *et al.*, 2012). Furthermore, MLST is expensive, labour intensive, and has a slow turnaround time; therefore, single locus sequence-based typing remains the method of choice for rapid identification of *Campylobacter* isolates. With reference to the future of microbiology, these approaches are likely to be complemented with, and gradually superseded by, WGS.

Proof-of-concept studies have demonstrated that, in principle, species identification, antimicrobial susceptibility testing, and epidemiological typing of bacteria could all be carried out with WGS (Didelot *et al.*, 2012; Köser *et al.*, 2012b); however, widespread implementation of WGS in microbiology laboratories will be contingent upon the development of effective data repositories and simple, accessible analysis software. The open-source BIGSDB software, implemented on all PubMLST and Institut Pasteur databases (http://pubmlst.org; http://bigsdb.web.pasteur.fr/), addresses both of these issues (Jolley & Maiden, 2010). As described in Chapter 1 [1.5.2.2.3.3], BIGSDB links WGS data to provenance and phenotype data, which can be stored, queried, and analysed online or in local databases, using only a personal computer. The BIGSDB analytical framework is based on a scalable and flexible gene-by-gene approach, allowing the number and type of loci analysed to be tailored according to the level of resolution required. For instance, comparison of a small number of core loci is sufficient for rapid species identification. The gene-by-gene approach also allows WGS data to be mapped to existing nomenclature schemes, thereby enabling users to compare and communicate their results (Maiden *et al.*, 2013). Although other methods based on *k*-mer approaches have been developed for *in silico* speciation of bacteria (Larsen *et al.*, 2014), the software is not open-source, is not integrated with a database, and does not relate back to the biology of the organism of interest.

The utility of the gene-by-gene approach for bacterial taxonomy was first demonstrated for *Neisseria*: species groups and strains were identified from WGS data using rMLST (Bennett *et al.*, 2012), which led to the development of a single

locus sequence-based assay for rapid, accurate identification of *Neisseria* species (Bennett *et al.*, 2014). Although the gene-by-gene approach has been used to study the epidemiology of *C. jejuni* and *C. coli* (Cody *et al.*, 2013), and twelve species were included in the domain-wide validation of rMLST (Jolley *et al.*, 2012b), there has not yet been a detailed examination of the genomic taxonomy of *Campylobacter*. The purpose of the current study was to apply the gene-by-gene approach for *in silico* identification of *Campylobacter* species, with the aim of integrating appropriate loci into the hierarchical set of schemes available on the *Campylobacter jejuni*/*coli* PubMLST database (http://pubmlst.org/campylobacter/). Using publicly available WGS data, two single locus and two multilocus schemes were evaluated for the identification of 22 *Campylobacter* species. The former were based on complete sequences of the 16S rRNA gene and *atpA*, and the latter on a novel universal *Campylobacter* MLST (uCMLST) scheme developed for this study, and rMLST.

As a starting point, a core genome phylogeny was required as a benchmark for comparison. This provided an opportunity to assess three different approaches for the definition of a *Campylobacter* core genome: the BIGSDB GENOME COMPARATOR module (Jolley & Maiden, 2010), OrthoMCL (Li *et al.*, 2003), and Proteinortho (Lechner *et al.*, 2011). While these approaches are all sequence similarity-based methods that make use of the BLAST algorithm, they differ in how they define groups of orthologous genes. GENOME COMPARATOR defines putative orthologs as best-matches identified by unidirectional BLAST searches, using coding sequences (CDSs) from a reference genome to query each isolate sequentially. In contrast, OrthoMCL and Proteinortho are formal ortholog detection programs that begin with

an all-against-all BLAST comparison of protein sequences from the genomes of interest, and then, using the resulting similarity scores, define clusters of orthologous genes with a Markov cluster algorithm (Li *et al.*, 2003) or a novel best-hit heuristic (Lechner *et al.*, 2011), respectively. These three approaches were compared to generate a list of core CDSs to generate the required benchmark phylogeny, and also provided insights into characteristics of the *Campylobacter* core genome.

3.2 Materials and Methods

3.2.1 Whole-genome sequencing data

Assembled WGS data for *Campylobacter* species available from the GenBank Genome Resource (http://www.ncbi.nlm.nih.gov/genome/) as of August 2015 were considered for inclusion in this study. Finished genomes were chosen preferentially over draft assemblies, and duplicate isolates were excluded. In total, 70 assembled genome sequences were transferred from GenBank to the *Campylobacter jejuni*/*coli* PubMLST database. Associated taxonomic information, aliases, and provenance data were also added to the *Campylobacter jejuni*/*coli* PubMLST database, as were GenBank accession numbers and links to appropriate PubMed records (Appendix C). These data can be accessed by using the project name "*Campylobacter* genus diversity" when searching the *Campylobacter jejuni*/*coli* PubMLST database.

3.2.2 Core genome analyses

3.2.2.1 Ortholog detection

The BIGSDB GENOME COMPARATOR module (Jolley & Maiden, 2010), OrthoMCL (Li *et al.*, 2003), and Proteinortho (Lechner *et al.*, 2011) were used to detect putative orthologs and define a core genome for the *Campylobacter* genus. Concerning the GENOME COMPARATOR analysis, TBLASTX was used to identify sequences comparable to annotated coding sequences present in the complete *C. jejuni* NCTC 11168 genome (GenBank accession number AL111168) (Parkhill *et al.*, 2000; Gundogdu *et al.*, 2007). Search parameters included a minimum sequence identity

of 50%, minimum alignment of 30%, and a BLAST word size of 11. GENOME COMPARATOR also generated MAFFT nucleotide sequence alignments (Katoh & Standley, 2013) and mean *p*-distances for each locus.

In preparation for formal ortholog detection with OrthoMCL and Proteinortho, predicted protein sequences were extracted for all CDSs present in the GenBank file available for each genome. Genomes without sequence annotations ($n = 5$) were annotated with Prokka, a software package for the annotation of prokaryotic genomes (Seemann, 2014). CDSs were defined using the default settings, predicted amino acid sequences were extracted, and all CDSs were added to a relational database. OrthoMCL version 2.0.9 was used to perform an all-against-all BLASTP comparison of the CDSs, after which clusters of orthologous genes and putative paralogs were detected using default parameters with an *E*-value < $1e^{-5}$ and a 50% minimum match length (Li *et al.*, 2003). Similarly, groups of orthologous genes and putative paralogs were identified among all CDSs with Proteinortho version 5.11, using the following default settings: *E*-value < $1e^{-5}$; minimum sequence identity of 25% and minimum coverage of 50% for best BLAST alignments; adaptive best alignments similarity $f = 0.95$; and algebraic connectivity > 0.1 (Lechner *et al.*, 2011, 2014).

3.2.2.2 Core genome definition and functional characterisation

After the exclusion of putative paralogs, the core genome was defined as CDSs or orthologous gene clusters present in ≥ 90% of isolates. The intersection of the three

analyses comprised the combined core genome that was used in further analyses. Genes that belonged to the combined core genome were assigned to functional categories based on definitions for NCTC 11168 extracted from the Clusters of Orthologous Groups of proteins (COGs) database (Galperin *et al.*, 2015).

3.2.2.3 Core genome phylogeny

A *Campylobacter* core genome phylogeny was generated based on concatenated MAFFT alignments (Katoh & Standley, 2013) of combined core loci produced by GENOME COMPARATOR. The phylogeny was reconstructed in MEGA version 6.06 (Tamura *et al.*, 2013), using the Kimura 2-parameter model (Kimura, 1980), with 1,000 bootstrap replicates.

3.2.3 *In silico* speciation

3.2.3.1 16S rRNA

As is true of other bacteria, *Campylobacter* commonly contain multiple copies of the 16S rRNA gene (Klappenbach *et al.*, 2001). The BLAST-based search strategy used by BIGSDB for genome annotation cannot distinguish different copies of the 16S rRNA gene, and, as a result, these loci are not currently curated in the *Campylobacter jejuni*/*coli* PubMLST database. Instead, a BLAST search was carried out within the database, using default parameters, to extract a single copy of the 16S rRNA gene from each genome, with NCTC 11168 serving as a query sequence.

3.2.3.2 *atpA*, MLST, and rMLST

The novel uCMLST scheme was developed based on complete gene sequences of *atpA* and five other genes used in the *C. jejuni*/*coli* MLST scheme, namely *aspA*, *glmM* (*pgm*), *glyA*, *glnA*, and *tkt* (Dingle *et al.*, 2001a, 2005). MLST and rMLST loci were annotated using a combination of automated and manual curation, and MAFFT nucleotide and amino acid sequence alignments (Katoh & Standley, 2013) were generated for each locus to check the positions of start and stop codons, as well as overall sequence similarity. Allelic data and MAFFT nucleotide sequence alignments were extracted for *atpA*, the uCMLST scheme, and rMLST loci, and phylogenies were reconstructed as described for the core genome. For each speciation method, the fixation index, F_{ST}, was calculated at the nucleotide level between pairs of species with more than one isolate, using the PopGenome package (Hudson *et al.*, 1992; Pfeifer *et al.*, 2014) implemented in R version 3.2.0 (R Core Team, 2015).

3.3 Results

3.3.1 Taxonomic diversity

The study included seventy isolates, which encompassed the taxonomic diversity of assembled *Campylobacter* WGS data available from GenBank as of August 2015. The collection was representative of the genus, and comprised 22 species and eight subspecies, with 16 isolates corresponding to type strains (Table 3.1). At the time of writing, WGS data for *C. avium*, *C. canadensis*, *C. helveticus*, *C. lanienae*, *C. troglodytis*, and *Campylobacter hyointestinalis* subsp. *lawsonii* had not been deposited in GenBank. As expected, the distribution of isolates across species was skewed towards the predominant disease-causing species: *C. jejuni* subsp. *jejuni* and *C. coli* accounted for 25 (36%) and 6 (8.6%) isolates, respectively. Conversely, 16 taxa corresponded to only one isolate each (Table 3.1).

Table 3.1 Details of publicly available *Campylobacter* genomes included in this study

Species	Isolate	GenBank accession number	Size (Mb)	% GC	N_{50} (bp)	Contigs (*n*)	CDSs[a] (*n*)	References
Complete genomes								
C. coli	RM1875 (**T**[b])	CP007183.1	1.86	31.1	-	-	1,837	-
	RM4661	CP007181.1	1.87	31.1	-	-	1,806	-
	FB1	CP011015.1	1.73	31.4	-	-	1,672	-
	CVM N29710	CP004066.1	1.73	31.4	-	-	1,629	(Chen *et al.*, 2013)
	RM5611	CP007179.1	1.73	31.3	-	-	1,635	-
	15-537360	CP006702.1	1.69	31.4	-	-	1,588	(Pearson *et al.*, 2013)
C. concisus	13826	CP000792.1	2.10	39.2	-	-	2,024	-
C. curvus	525.92	CP000767.1	1.97	44.5	-	-	1,931	-
C. fetus subsp. *fetus*	04/554	CP008808.1	1.83	33.1	-	-	1,688	(van der Graaf-van Bloois *et al.*, 2014)
	82-40	CP000487.1	1.77	33.3	-	-	1,719	(van der Graaf-van Bloois *et al.*, 2014)
C. fetus subsp. *testudinum*	03-427 (**T**)	CP006833.1	1.78	33.1	-	-	1,695	(Tu *et al.*, 2004; Gilbert *et al.*, 2013; Fitzgerald *et al.*, 2014)
	pet-3	CP009226.1	1.78	33.1	-	-	1,652	(Wang *et al.*, 2015)
C. fetus subsp. *venerealis*	97/608	CP008810.1	2.00	33.2	-	-	1,874	(van der Graaf-van Bloois *et al.*, 2014)
	cfvi03/293	CP006999.1	2.00	33.1	-	-	1,773	(van der Graaf-van Bloois *et al.*, 2014)
	84-112	HG004426.1	1.99	33.2	-	-	1,981	(Kienesberger *et al.*, 2014; van der Graaf-van Bloois *et al.*, 2014)
C. hominis	ATCC BAA-381 (**T**)	CP000776.1	1.71	31.7	-	-	1,682	-
C. iguaniorum	1485E (**T**)	CP009043.1	1.75	35.9	-	-	1,645	(Gilbert *et al.*, 2014a)
C. insulaenigrae	NCTC 12927 (**T**)	CP007770.1	1.47	28.2	-	-	1,441	(Foster *et al.*, 2004; Miller *et al.*, 2014a)
C. jejuni subsp. *doylei*	269.97	CP000768.1	1.85	30.6	-	-	1,731	(Parker *et al.*, 2007; Brown *et al.*, 2015)
	NCTC 11168	AL111168.1	1.64	30.5	-	-	1,624	(Parkhill *et al.*, 2000; Gundogdu *et al.*, 2007)

C. jejuni subsp. *jejuni*	81116	CP000814.1	1.63	30.5	-	-	1,626	(Pearson *et al.*, 2007)
	81-176	CP000538.1	1.70	30.5	-	-	1,653	-
	00-1597	CP010306.1	1.74	30.4	-	-	1,722	-
	IA3902	CP001876.1	1.67	30.5	-	-	1,613	(Sahin *et al.*, 2008)
	F38011	CP006851.1	1.69	30.5	-	-	1,684	-
	RM1221	CP000025.1	1.78	30.3	-	-	1,838	(Fouts *et al.*, 2005)
	S3	CP001960.1	1.72	30.5	-	-	1,734	(Cooper *et al.*, 2011)
	M1	CP001900.1	1.62	30.6	-	-	1,622	(Friis *et al.*, 2010)
	32488	CP006006.1	1.70	30.5	-	-	1,689	-
	4031	HG428754.1	1.67	30.5	-	-	1,698	(Revez *et al.*, 2014a)
	MTVDSCj20	CP008787.1	1.65	30.5	-	-	1,616	(Taveirne *et al.*, 2014)
	YH001	CP010058.1	1.71	30.5	-	-	1,824	(He *et al.*, 2015)
	00-6200	CP010307.1	1.67	30.5	-	-	1,668	-
	01-1512	CP010072.1	1.83	30.4	-	-	1,787	-
	00-0949	CP010301.1	1.83	30.4	-	-	1,786	-
	35925B2	CP010906.1	1.61	30.7	-	-	1,613	-
	00-2425	CP006729.2	1.72	30.5	-	-	1,741	(Clark *et al.*, 2014b, 2014a)
	00-2426	CP006708.2	1.68	30.5	-	-	1,683	(Clark *et al.*, 2014b)
	00-2538	CP006707.2	1.72	30.5	-	-	1,741	(Clark *et al.*, 2014b)
	00-2544	CP006709.2	1.77	30.5	-	-	1,739	(Clark *et al.*, 2014b)
	CG8421	CP005388.1	1.64	30.5	-	-	1,635	(Poly *et al.*, 2008)
	ICDCCJ07001	CP002029.1	1.71	30.6	-	-	1,494	(Zhang *et al.*, 2010a)
	PT14	CP003871.3	1.64	30.5	-	-	1,607	(Brathwaite *et al.*, 2013; Hooton & Connerton, 2014)
	R14	CP005081.1	1.80	30.4	-	-	1,824	-
C. lari	Slaughter Beach[c]	CP011372.1	1.56	29.8	-	-	1,490	-
	CCUG 22395[c]	CP007776.1	1.52	29.9	-	-	1,481	(Miller *et al.*, 2014a)

	NCTC 11845[c]	CP007775.1	1.79	29.4	-	-	1,702	(Miller *et al.*, 2014a)
	RM16701[c]	CP007777.1	1.52	29.9	-	-	1,476	(Miller *et al.*, 2014a)
	RM16712[c]	CP007778.1	1.57	29.7	-	-	1,508	(Miller *et al.*, 2014a)
	RM2100	CP000932.1	1.57	29.6	-	-	1,504	(Tauxe *et al.*, 1985; Miller *et al.*, 2008, 2014a)
C. lari subsp. *concheus*	LMG 11760	CP007771.1	1.50	29.7	-	-	1,452	(Debruyne *et al.*, 2009; Miller *et al.*, 2014a)
C. peloridis	LMG 23910 (**T**)	CP007766.1	1.76	28.4	-	-	1,591	(Debruyne *et al.*, 2009; Miller *et al.*, 2014a)
C. subantarcticus	LMG 24377 (**T**)	CP007773.1	1.85	29.8	-	-	1,771	(Debruyne *et al.*, 2010a; Miller *et al.*, 2014a)
	LMG 24374	CP007772.1	1.78	29.9	-	-	1,675	(Debruyne *et al.*, 2010a; Miller *et al.*, 2014a)
C. volucris	LMG 24379	CP007774.1	1.52	28.6	-	-	1,480	(Debruyne *et al.*, 2010b; Miller *et al.*, 2014a)
Draft genomes								
C. corcagiensis	CIT045 (**T**)	JFAP00000000.1	1.67	31.9	281,888	21	1,725 [d]	(Koziel *et al.*, 2014a, 2014b)
C. cuniculorum	DSM 23162 (**T**)	JHZL00000000.1	1.87	31.3	83,241	46	1,837 [d]	(Zanoni *et al.*, 2009)
C. gracilis	RM3268 (**T**)	ACYG00000000.1	2.26	46.6	203,752	33	2,847	(Vandamme *et al.*, 1995)
C. hyointestinalis subsp. *hyointestinalis*	DSM 19053 (**T**)	JHQP00000000.1	1.73	33.9	224,275	26	1,533	(Gebhart *et al.*, 1985)
C. mucosalis	DSM 21682 (**T**)	JHQQ00000000.1	1.75	36.6	214,630	42	1,514	-
C. rectus	RM3267 (**T**)	ACFU00000000.1	2.51	44.8	65,738	89	2,971	(Vandamme *et al.*, 1991)
C. showae	RM3277 (**T**)	ACVQ00000000.1	2.07	45.7	141,033	33	2,361	(Etoh *et al.*, 1993)
	CC57C	AOTD00000000.1	2.19	45.4	16,873	274	2,230	(Warren *et al.*, 2013)
	UNSWCD	AMZQ00000000.1	2.13	45.1	209,023	23	2,478	(Tay *et al.*, 2013)
C. sputorum subsp. sputorum	INTA08/209	JMTI00000000.1	1.78	29.1	115,802	51	1,845[d]	(Iraola *et al.*, 2014)
C. upsaliensis	DSM 5365 (**T**)	JHZN00000000.1	1.62	35.0	118,240	44	1,636 [d]	
	JV21	AEPU00000000.1	1.65	35.0	88,025	49	1,705	

	RM3195	AAFJ00000000.1	1.77	34.3	153,933	20	1,934	(Fouts *et al.*, 2005)
C. ureolyticus	DSM 20703 (**T**)	ARGD00000000.1	1.74	28.9	304,567	37	1,771 [d]	
	ACS-301-V-Sch3b	AGYD00000000.1	1.70	29.4	306,603	19	1,732	

[a] CDSs, coding sequences.

[b] **T**, type strain.

[c] Urease-producing strain.

[d] Based on Prokka annotations carried out for this study.

3.3.2 Quality of whole-genome sequencing data

Although complete genome sequences were preferentially included, draft assemblies were the only WGS data available for ten species (*n* = 15) (Table 3.1). The quality of these genomes was assessed based on assembly statistics and the completeness of six MLST genes, *aspA*, *atpA*, *glmM* (*pgm*), *glyA*, *glnA*, and *tkt*, and 52 rMLST loci, which are essential genes that are evenly distributed across the genome. The draft genomes ranged from 1.6 to 2.5 Mb, which was comparable to the lengths of complete *Campylobacter* genomes (Table 3.1). The number of contigs in each assembly ranged from 19 to 274 (Table 3.1), with an average of 54 per genome. N_{50} length is a statistic commonly used to assess assembly quality, and it indicates the length of the smallest contig for which 50% of the assembly is contained in contigs that size or greater. Most draft genomes had a large N_{50} length, with an average length of 168,508 bp, ranging from 16,873 to 306,603 bp (Table 3.1). Curation of MLST and rMLST loci provided further evidence of the high quality of the draft assemblies: complete nucleotide sequences were obtained for all 58 MLST and rMLST genes from all genomes, except *C. showae* isolate CC57C and *C. upsaliensis* isolate RM3195, for which one MLST and two rMLST loci, respectively, were incomplete or not detected.

3.3.3 Core genome analyses

3.3.3.1 Definition of a Campylobacter *core genome*

The BIGSDB GENOME COMPARATOR module implemented in the *Campylobacter jejuni/coli* PubMLST database, OrthoMCL, and Proteinortho were

used to define core genomes for the *Campylobacter* genus. GENOME COMPARATOR was the simplest to run, as data were selected and parameters set using a simple web-interface. In contrast, Proteinortho and OrthoMCL are run on the command-line and require a Unix environment with BLAST, Perl, and Python installed, and, for OrthoMCL, a relational database. Proteinortho was the least computationally intensive and took four hours to run to completion, while OrthoMCL and GENOME COMPARATOR took 16 and 22 hours, respectively.

The GENOME COMPARATOR analysis was carried out using TBLASTX to search for sequences comparable to 1,643 loci present in the re-annotated *C. jejuni* subsp. *jejuni* NCTC 11168 genome (Parkhill *et al.*, 2000; Gundogdu *et al.*, 2007). TBLASTX, which compares a six-frame translation of both query and subject sequences, was used in conjunction with low stringency search parameters to improve detection of putative orthologs among the diverse members of the *Campylobacter* genus. Following exclusion of putative paralogs, 1,614 putative orthologs were detected among the study isolates. The distribution of these loci was bimodal, with the largest peak including 48.6% of NCTC 11168 CDSs, consistent with core loci that were detected in 90-100% of the isolates (Figure 3.1). The second peak accounted for 12.8% of reference CDSs, which were present in 60-70% of isolates, corresponding to the thermophilic *Campylobacter* species.

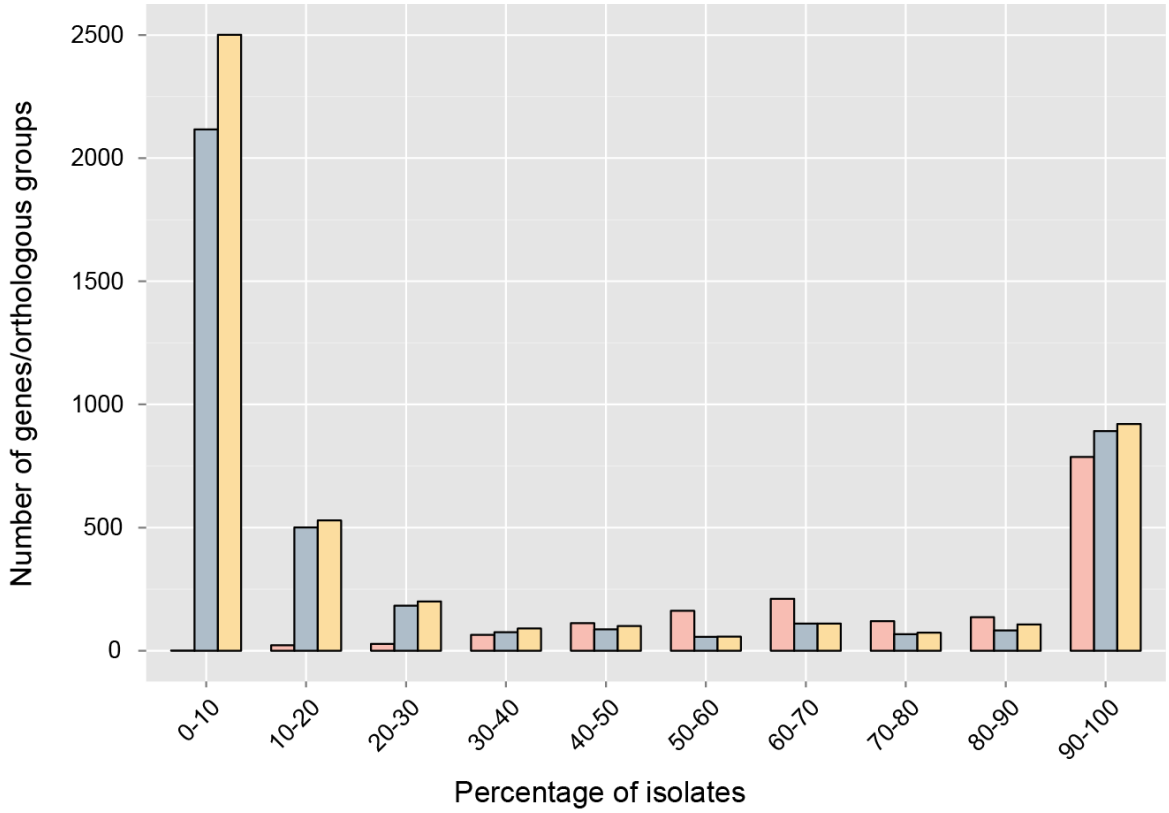


Figure 3.1 Frequency distribution of putative orthologs identified by GENOME COMPARATOR (pink), and orthologous gene clusters delimited by OrthoMCL (blue) and Proteinortho (yellow), based on analyses of 70 genome sequences representative of 22 *Campylobacter* species.

With respect to OrthoMCL and Proteinortho, 5,026 and 5,278 clusters of orthologous genes, respectively, were detected among the 122,513 CDSs present in the 70 study genomes. After gene clusters with putative paralogs were excluded, 4,170 and 4,688 groups remained in the OrthoMCL and Proteinortho datasets, respectively. The distribution of OrthoMCL and Proteinortho gene clusters was also bimodal: the largest peaks corresponded to accessory gene groups present in 0-10% of isolates, while the second, smaller peak corresponded to core loci detected in 90-100% of the isolates (Figure 3.1).

In all three analyses, CDSs or orthologous gene clusters detected in ≥ 90% of the isolates were considered part of the core genome; this flexible cut-off was used to allow for the possibility that a small subset of complete genomes may lack certain core genes or, more likely, that certain loci may not be included in draft assemblies. At the 90% threshold, the GENOME COMPARATOR core genome contained 784 genes, while 910 and 942 core gene clusters were detected using OrthoMCL and Proteinortho, respectively. Each OrthoMCL and Proteinortho gene cluster included a CDS from NCTC 11168, with one exception. In both analyses, the exception corresponded to *ctsF*, a putative type II secretion system F protein, which is a pseudogene in NCTC 11168. Comparison of the three core genomes identified 1,060 unique genes, 657 of which were present in all analyses, totalling 624,402 bp (38%) of the NCTC 11168 genome (Figure 3.2; Appendix D). Among the 141 genes unique to a single core genome, 67 (47.5%) were identified as paralogs in at least one other analysis. The 657 genes present in all three analyses were used as a provisional, catch-all core genome in the analyses that follow.

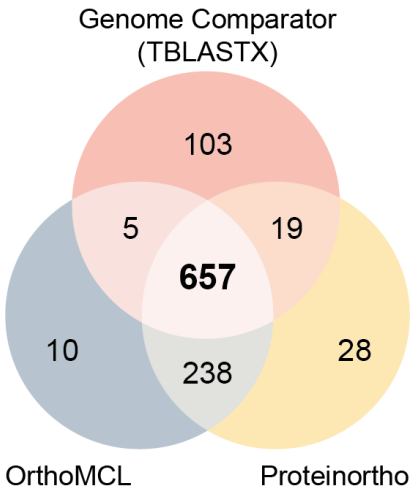


Figure 3.2 Venn diagram indicating the number of unique and overlapping putative orthologs or orthologous gene clusters identified using GENOME COMPARATOR (pink), OrthoMCL (blue), and Proteinortho (yellow).

3.3.3.2 *Characteristics of the* Campylobacter *core genome*

Loci from the combined core genome were evenly distributed across NCTC 11168 and had a mean *p*-distance of 0.24 (Figure 3.3). In comparison, non-core NCTC 11168 CDSs detected in the GENOME COMPARATOR analysis had a lower mean *p*-distance of 0.2, but the molecular distances of individual genes were more variable than those in the core genome (Figure 3.3). The functional distribution of core loci was determined based on COG assignments, and compared to that of NCTC 11168 (Figure 3.4). The core genome included a larger proportion of genes encoding proteins involved in information storage and processing (29.7 vs. 16%), particularly those with a role in translation, ribosomal structure and biogenesis (21 vs. 10%). The core genome also included slightly larger proportions of other essential genes (Figure 3.4), with a corresponding decrease in loci that were poorly characterised or lacking functional assignment (9.9 vs. 28.2%).

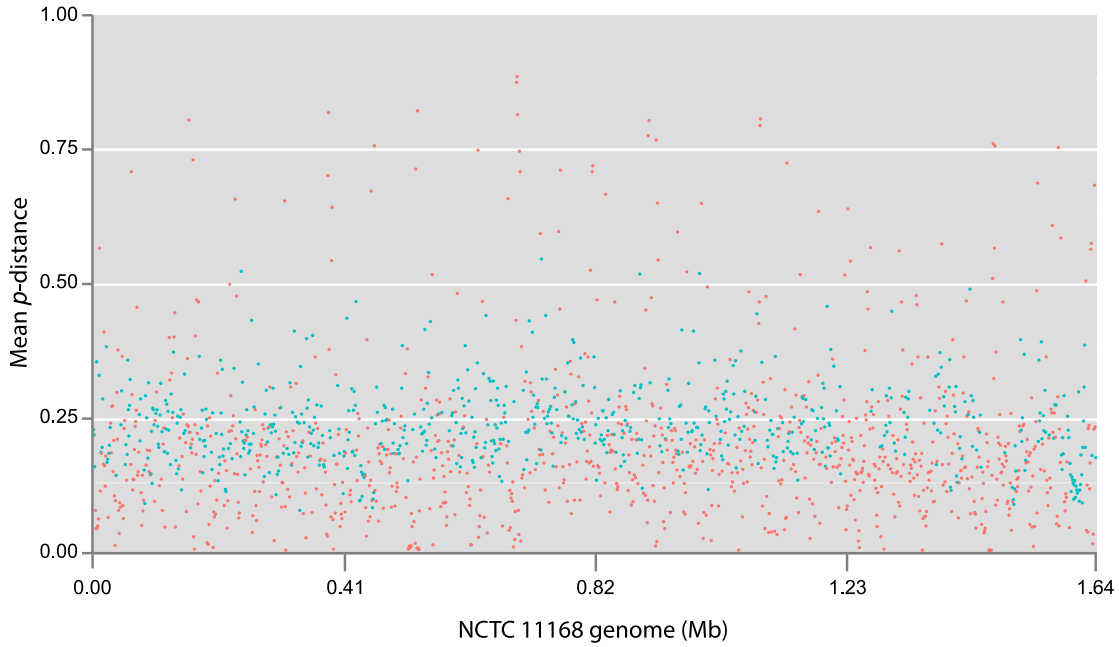


Figure 3.3 Mean *p*-distances calculated by GENOME COMPARATOR for combined core (blue) and non-core (red) coding sequences plotted relative to genome position in NCTC 11168.

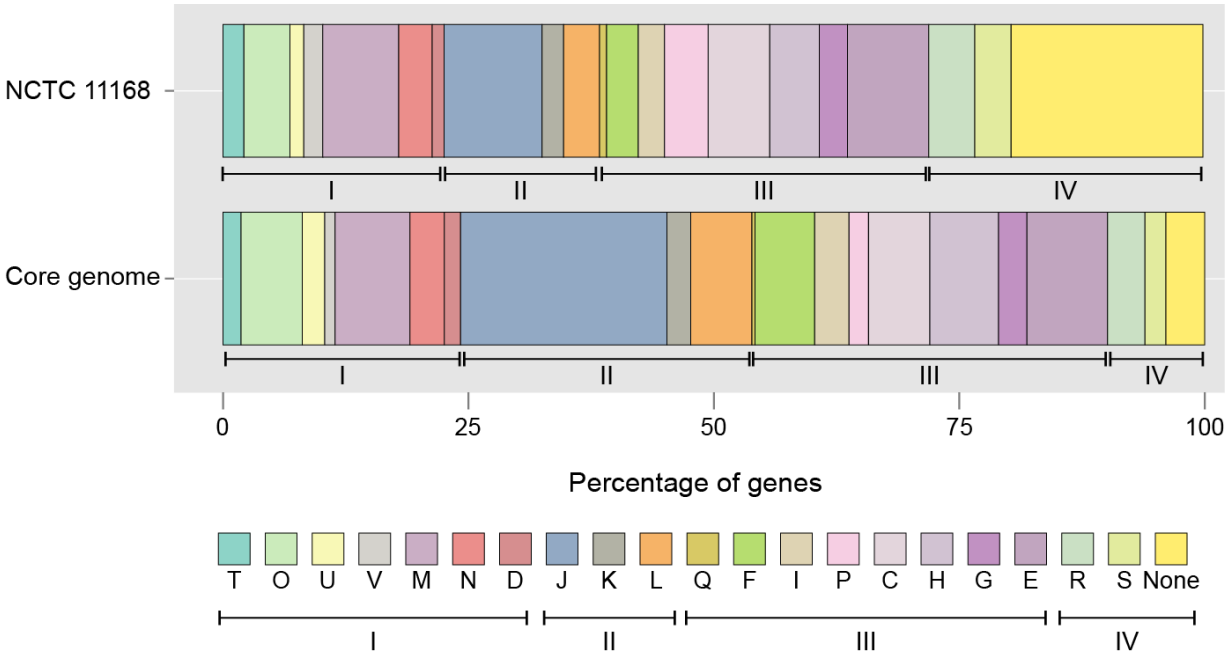


Figure 3.4 Comparison of the functional distribution of coding sequences present in *C. jejuni* isolate NCTC 11168 and the *Campylobacter* core genome. Categories correspond to those used in the COG database, which were combined into broader functional groups: (I), cellular processes and signalling; (II), information storage and processing; (III), metabolism; and (IV), poorly characterised or no functional information available. COG categories correspond to: (T), signal transduction mechanisms; (O), post-translational modification, protein turnover, chaperones; (U), intracellular trafficking, secretion, and vesicular transport; (V), defence mechanisms; (M), cell wall/membrane/envelope biogenesis; (N), cell motility; (D), cell cycle control, cell division, chromosome partitioning; (J), translation, ribosomal structure and biogenesis; (K), transcription; (L), replication, recombination and repair; (Q), secondary metabolites biosynthesis, transport and catabolism; (F), nucleotide transport and metabolism; (I), lipid transport and metabolism; (P), inorganic ion transport and metabolism; (C), energy production and conversion; (H), coenzyme transport and metabolism; (G), carbohydrate transport and metabolism; (E), amino acid transport and metabolism; (R), general function only; (S), function unknown; (None), no functional information available.

3.3.3.3 Core genome phylogeny

A Neighbour-joining phylogeny was reconstructed based on 395,838 unambiguously aligned nucleotides, equivalent to 24.1% of the NCTC 11168 genome (Figure 3.5A). All *Campylobacter* species were phylogenetically distinct, and those represented by multiple isolates formed monophyletic groups. Species groups were separated by long branch lengths, which were indicative of genetic divergence post-speciation. To estimate between-species gene flow, F_{ST} values were calculated for species with more than one representative isolate, with possible values ranging from 0, indicating panmixis, to 1, indicating complete differentiation. With the exception of *C. lari* and *C. subantarcticus*, F_{ST} values were generally high, providing supporting evidence of species divergence (Table 3.2). At the subspecies level, *C. jejuni* subsp. *jejuni* and *C. jejuni* subsp. *doylei* were phylogenetically distinct, as were *Campylobacter lari* subsp. *lari* isolate RM2100, *Campylobacter lari* subsp. *concheus* isolate LMG 11760, and urease-positive *C. lari* isolates (Figure 3.5A; Table 3.1). In contrast, *C. fetus* subsp. *fetus* and *C. fetus* subsp. *venerealis* were distinct from *Campylobacter fetus* subsp. *testudinum*, but they did not segregate into monophyletic groups (Figure 3.5B).

Table 3.2 Gene flow among eight *Campylobacter* species based on 657 core genes as measured by F_{ST}

	Cjej	***Ccol***	***Cfet***	***Clar***	***Csub***	***Csho***	***Cups***
Cjej[a]	-						
Ccol	0.874	-					
Cfet	0.934	0.931	-				
Clar	0.875	0.873	0.898	-			
Csub	0.939	0.936	0.934	0.641	-		
Csho	0.914	0.911	0.868	0.885	0.915	-	
Cups	0.905	0.902	0.920	0.871	0.923	0.899	-
Cure	0.918	0.917	0.882	0.879	0.918	0.873	0.906

[a] *Cjej, C. jejuni; Ccol, C. coli; Cfet, C. fetus; Clar, C. lari; Csub, C. subantarcticus; Cups, C. upsaliensis; Cure, C. ureolyticus.*

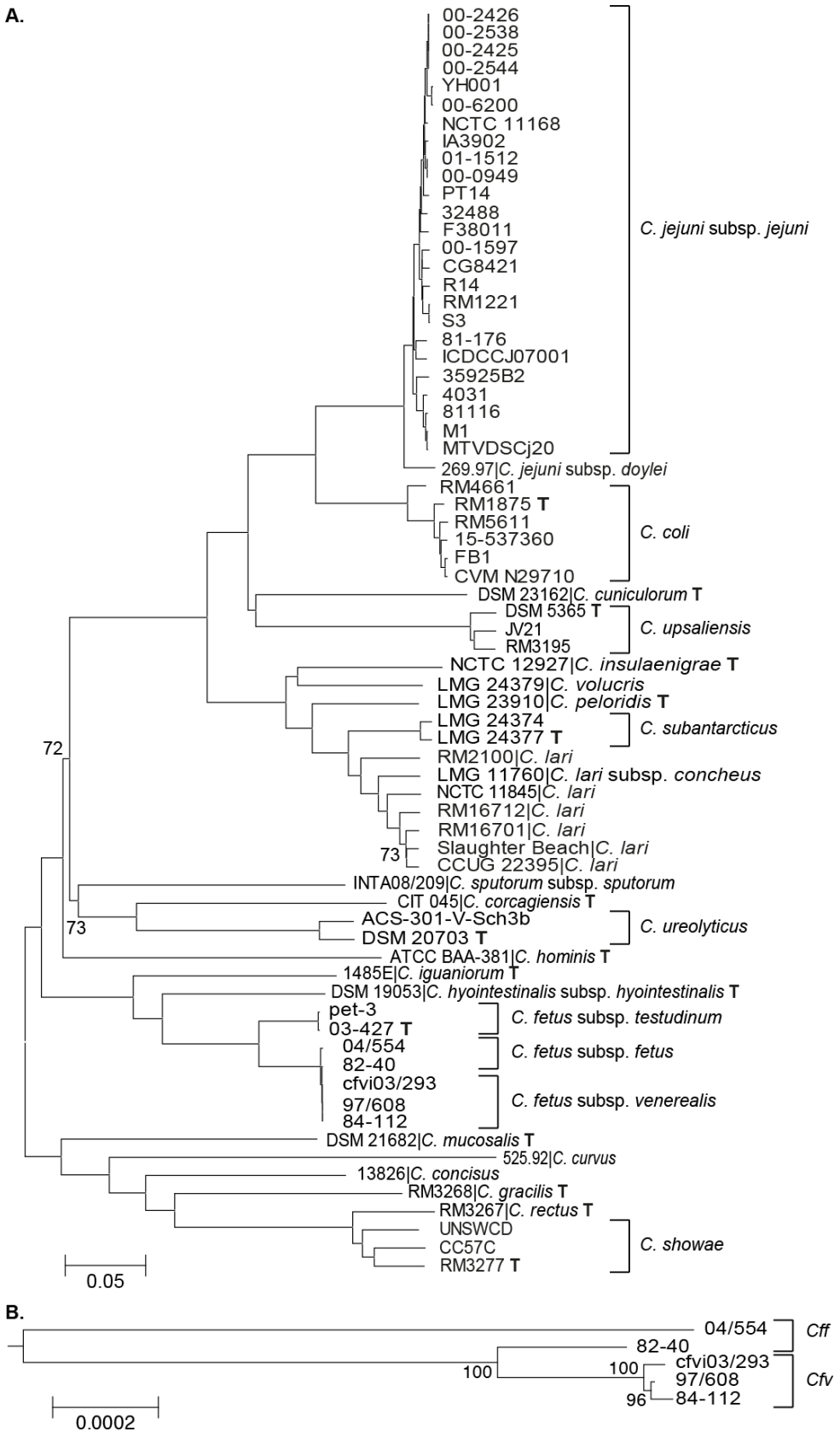


Figure 3.5 Evolutionary relationships among 70 isolates representative of 22 *Campylobacter* species. (A) The Neighbor-joining phylogeny was reconstructed based on 395,838 unambiguously aligned nucleotides from 657 core genes, with 1,000 bootstrap replicates; bootstrap values are shown for nodes supported by less than 80% of replicates. **T**, type strain. (B) Detailed view of the relationships among *C. fetus* subsp. *fetus* and *C. fetus* subsp. *venerealis* isolates.

3.3.4 Evaluation of approaches for *in silico* speciation of *Campylobacter*

3.3.4.1 The 16S rRNA gene and atpA

Single locus typing methods are commonly used for speciation of *Campylobacter* isolates in clinical and research laboratories; therefore, complete gene sequences of two target loci, the 16S rRNA gene and *atpA*, were evaluated for *in silico* identification of *Campylobacter* from WGS data. Complete 16S rRNA gene sequences were obtained from all study genomes, except *C. curvus* isolate 525.92. Forty-four unique sequences were identified and had a mean *p*-distance of 0.038, which was indicative of low diversity. Alleles were not species specific, with identical sequences detected in: *C. jejuni* subsp. *jejuni* and *C. coli*; *C. fetus* subsp. *fetus* and *C. fetus* subsp. *venerealis*; and urease-positive *C. lari* and *C. lari* subsp. *concheus*. Although between-species F_{ST} values were generally high (Table 3.3), those between *C. jejuni* and *C. coli*, and *C. lari* and *C. subantarcticus*, were very low.

Table 3.3 Gene flow among eight *Campylobacter* species as measured by F_{ST} based on the 16S rRNA gene (below the diagonal) and *atpA* (above the diagonal)

	Cjej	***Ccol***	***Cfet***	***Clar***	***Csub***	***Csho***	***Cups***	***Cure***
Cjej[a]	-	0.968	0.946	0.857	0.990	0.901	0.944	0.994
Ccol	0.001	-	0.946	0.859	0.990	0.896	0.945	0.994
Cfet	0.969	0.921	-	0.873	0.947	0.824	0.928	0.948
Clar	0.797	0.638	0.942	-	0.762	0.839	0.830	0.917
Csub	0.910	0.712	0.965	0.043	-	0.902	0.954	1.000
Csho	0.979	0.949	0.955	0.960	0.975	-	0.883	0.907
Cups	0.925	0.870	0.926	0.903	0.930	0.952	-	0.973
Cure	0.997	0.963	0.981	0.976	0.993	0.983	0.969	-

[a] *Cjej*, *C. jejuni*; *Ccol*, *C. coli*; *Cfet*, *C. fetus*; *Clar*, *C. lari*; *Csub*, *C. subantarcticus*; *Cups*, *C. upsaliensis; Cure, C. ureolyticus*.

The overall topology of the Neighbor-joining phylogeny based on 1501 unambiguously aligned nucleotides (Figure 3.6) was similar to that of the core genome phylogeny (Figure 3.5); however, the 16S rRNA phylogeny failed to differentiate *C. jejuni* subsp. *jejuni*, *C. jejuni* subsp. *doylei*, and *C. coli*, while *C. lari* isolates were polyphyletic. As was also true for the core genome phylogeny, *C. fetus* subsp. *fetus* and *C. fetus* subsp. *venerealis* could not be distinguished.

Complete *atpA* sequences were obtained for all isolates. Based on a nucleotide sequence alignment of 1502 bp, *atpA* was more diverse than the 16S rRNA gene, with an overall *p*-distance of 0.18. Forty-eight unique alleles were detected, all of which were species-specific. The topology of the *atpA* Neighbor-joining tree (Figure 3.7) was highly similar to that of the core genome phylogeny (Figure 3.5). Minor differences included the position of *C. hominis* and *C. iguaniorum*, as well as the branching order of *C. mucosalis*, *C. curvus*, *C. concisus*, and *C. gracilis*. Consistent with the high F_{ST} values observed for *atpA* (Table 3.3), all species were phylogenetically distinct, barring *C. showae* and *C. rectus* (Figure 3.7). At the subspecies level, *C. fetus* subsp. *fetus* and *C. fetus* subsp. *venerealis* could not be distinguished.

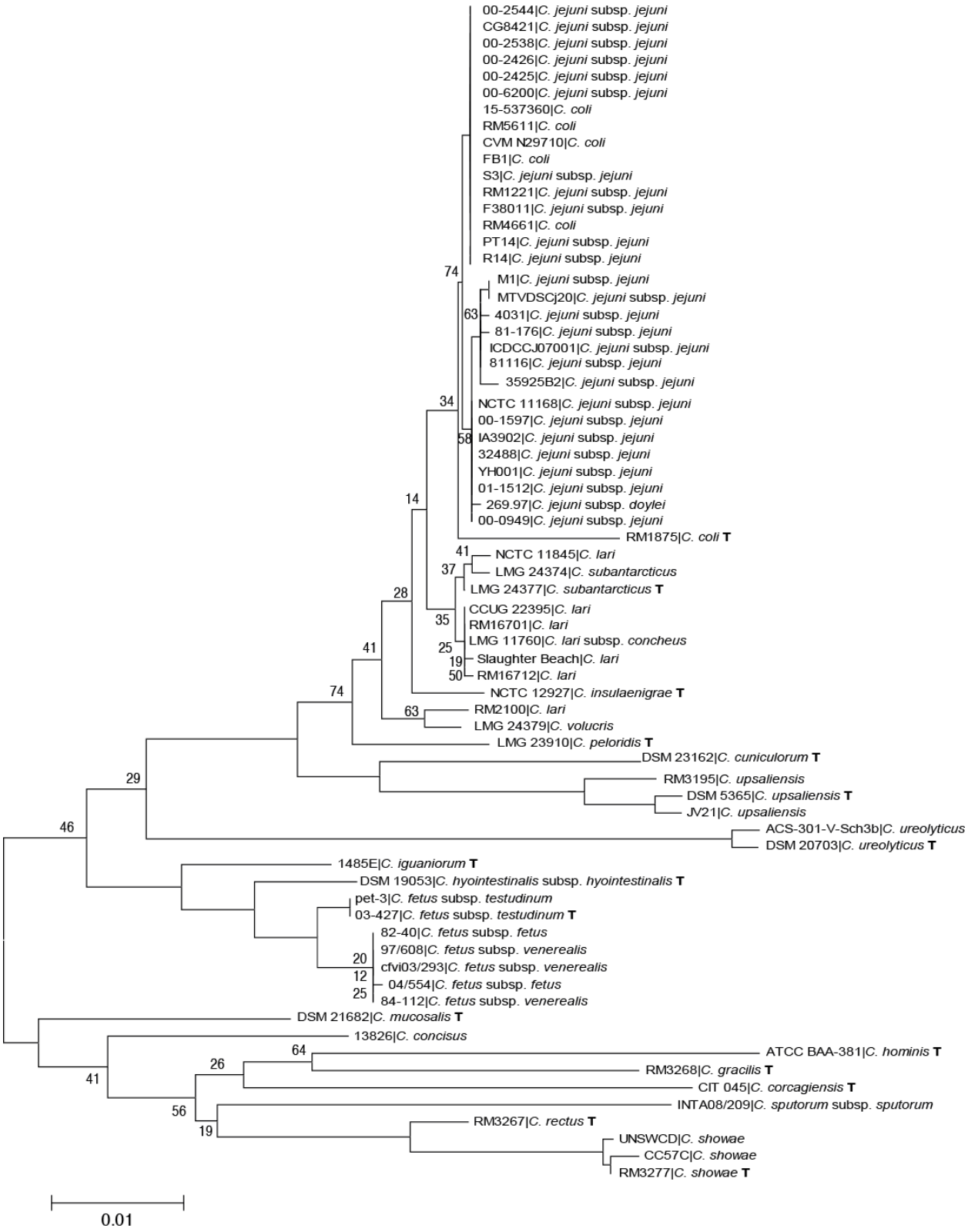


Figure 3.6 Neighbor-joining phylogeny based on 16S rRNA gene sequences of 69 *Campylobacter* isolates. Bootstrap values are shown for nodes supported by less than 80% of 1,000 replicates. **T**, type strain.

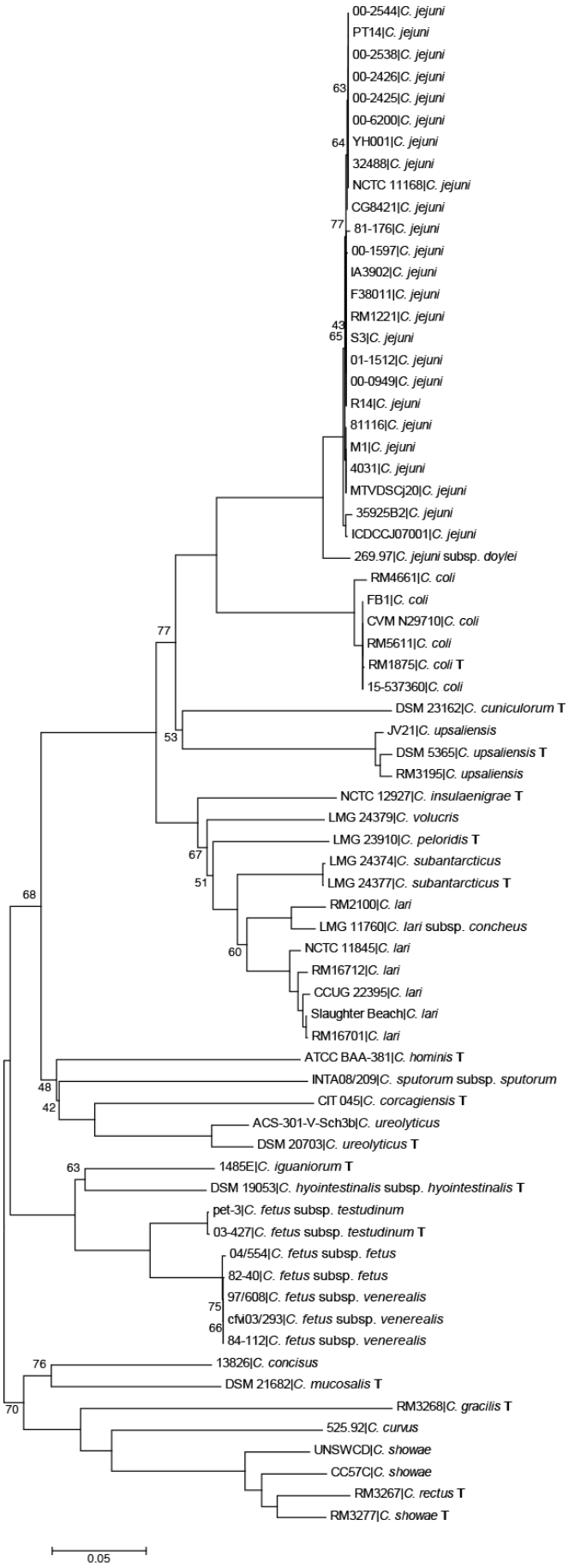


Figure 3.7 Neighbor-joining phylogeny based on *atpA* sequences of 70 *Campylobacter* isolates. Bootstrap values are shown for nodes supported by less than 80% of 1,000 replicates. **T**, type strain.

3.3.4.2 MLST and rMLST

As multilocus typing schemes provide greater resolution than single locus methods, the gene-by-gene approach was exploited to evaluate a novel universal MLST scheme uCMLST and rMLST for *in silico* speciation of *Campylobacter* isolates. The former was based on complete sequences of six of the seven genes used in the *C. jejuni*/*coli* MLST scheme, including *aspA*, *atpA*, *glmM* (*pgm*), *glyA*, *glnA*, and *tkt*; the seventh locus, *gltA*, was excluded as it could not be detected in members of the *C. lari* group, as previously described (Miller *et al.*, 2005). Fifty-two of the 53 rMLST loci were included, as *rpmD* appears to be absent among *Campylobacter* species (Cody *et al.*, 2013).

Complete uCMLST allelic profiles were obtained for all isolates, except *C. showae* isolate CC57C in which *aspA* was incomplete as it was at the end of a contig. The average number of alleles per locus was 52, ranging from 48 for *atpA* to 57 for *tkt*, all of which were species-specific. In total, 62 unique allelic profiles were identified, which had a mean *p*-distance of 0.21. Three allelic profiles accounted for multiple *C. jejuni* isolates (n = 10), while the *C. fetus* subsp. *venerealis* isolates (n = 3) were identical at all loci. The Neighbor-joining tree based on concatenated nucleotide sequences of the six MLST genes, totalling 7,373 bp or 0.45% of the NCTC 11168 genome, was similar to that produced for the core genome, although the position of *C. hominis* and branching order of *C. curvus*, *C. concisus*, and *C. gracilis* differed (Figure 3.8). As for *atpA*, all species were phylogenetically distinct, barring *C. showae* and *C. rectus*. This was consistent with the between-species F_{ST} values, which, with the exception of *C. lari* and *C. subantarcticus*, were high (Table 3.4). At

the subspecies level, *C. fetus* subsp. *fetus* and *C. fetus* subsp. *venerealis* remained indistinguishable.

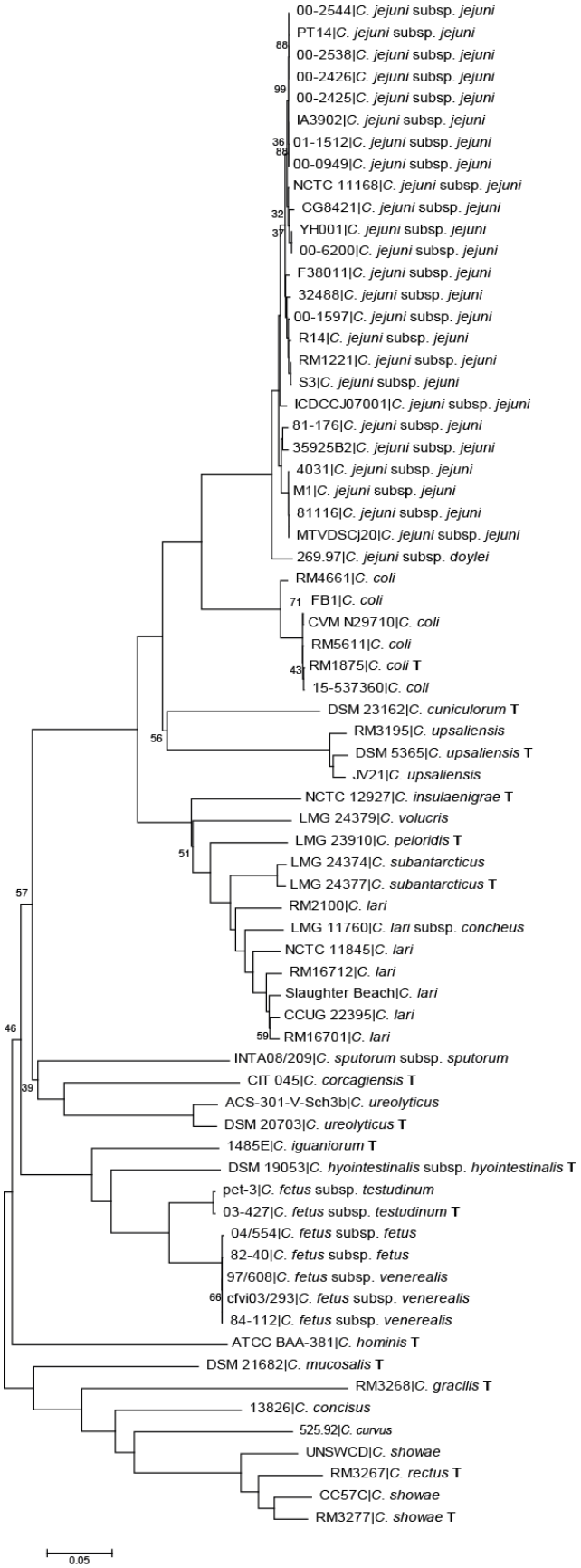


Figure 3.8 Neighbor-joining phylogeny based on concatenated nucleotide sequences of six MLST loci from 70 *Campylobacter* isolates. Bootstrap values are shown for nodes supported by less than 80% of 1,000 replicates. **T**, type strain.

Table 3.4 Gene flow among eight *Campylobacter* species as measured by F_{ST} based on six MLST genes (below the diagonal) and 52 rMLST genes (above the diagonal)

	Cjej	*Ccol*	*Cfet*	*Clar*	*Csub*	*Csho*	*Cups*	*Cure*
Cjej[a]	-	0.865	0.958	0.891	0.955	0.935	0.914	0.935
Ccol	0.921	-	0.952	0.880	0.944	0.930	0.904	0.930
Cfet	0.935	0.942	-	0.927	0.961	0.910	0.943	0.921
Clar	0.863	0.877	0.889	-	0.588	0.909	0.887	0.906
Csub	0.943	0.953	0.933	0.624	-	0.939	0.937	0.938
Csho	0.901	0.907	0.847	0.867	0.904	-	0.924	0.907
Cups	0.918	0.928	0.922	0.864	0.929	0.888	-	0.923
Cure	0.924	0.930	0.877	0.871	0.919	0.857	0.911	-

[a] *Cjej*, *C. jejuni*; *Ccol*, *C. coli*; *Cfet*, *C. fetus*; *Clar*, *C. lari*; *Csub*, *C. subantarcticus*; *Cups*, *C. upsaliensis; Cure, C. ureolyticus*.

Complete rMLST allelic profiles were obtained for all isolates except *C. upsaliensis* RM3195, in which *rpsI* was at the end of a contig and *rplM*, typically adjacent, was not detected. On average, there were 41 alleles per locus; *rpsA* and *rpsC* were most diverse with 51 alleles each, while *rpmJ* was least diverse with 25 alleles. Sixty-six unique allelic profiles were identified, sharing an overall *p*-distance of 0.17, and two profiles accounted for multiple *C. jejuni* subsp *jejuni* isolates ($n = 2$; $n = 4$). Allele sharing was observed within four sets of species: *C. jejuni* subsp. *jejuni*, *C. jejuni* subsp. *doylei*, and *C. coli* shared alleles at 11 loci; *C. rectus* and *C. showae*, at four loci; *C. lari* subspecies and *C. subantarcticus* at four loci; and *C. lari* subspecies and *C. peloridis* at a single locus. Unsurprisingly, allele sharing was more pronounced at the subspecies level: a single *rpmF* allele was shared by all *C. fetus* subspecies; *C. jejuni* subsp. *jejuni* and *C. jejuni* subsp. *doylei* shared alleles at three loci; *C. lari* subspecies shared alleles at nine loci; and *C. fetus* subsp. *fetus* and *C. fetus* subsp. *venerealis* shared alleles at all but two loci.

The rMLST phylogeny, based on an alignment of 19,552 bp or 1.2% of NCTC 11168, was highly similar to that of the core genome phylogeny. As for *atpA* and MLST, the positions of *C. hominis*, *C. mucosalis*, *C. curvus*, and *C. concisus* differed slightly compared to the core genome phylogeny, but almost all species were phylogenetically distinct (Figure 3.9). In contrast with *atpA* and MLST, *C. showae* and *C. rectus* were distinct, while *C. lari* subspecies and *C. subantarcticus* were not resolved, which was consistent with the low F_{ST} values (Table 3.4). As was true of the other methods evaluated here, *C. fetus* subsp. *fetus* and *C. fetus* subsp. *venerealis* could not be differentiated.

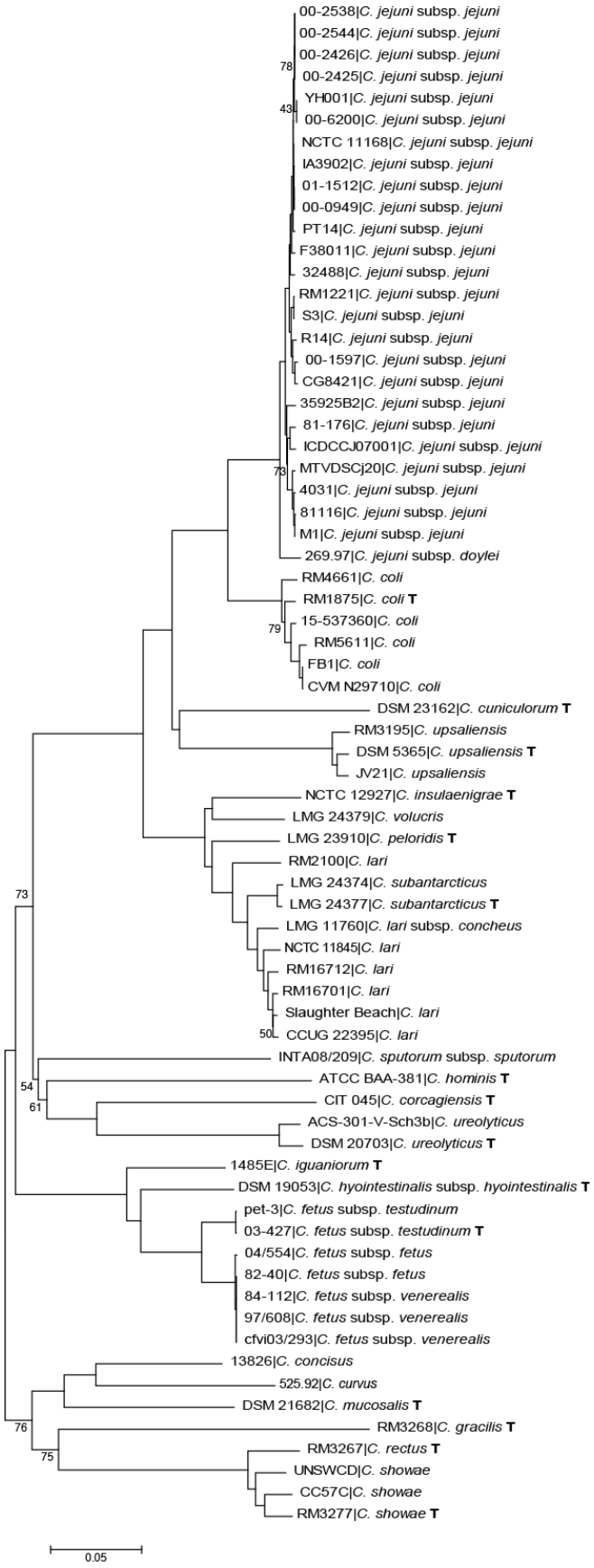


Figure 3.9 Neighbor-joining phylogeny based on concatenated nucleotide sequences of 52 rMLST loci from 70 *Campylobacter* isolates. Bootstrap values are shown for nodes supported by less than 80% of 1,000 replicates. **T**, type strain.

3.4 Discussion

The epidemiology and pathogenesis of emerging *Campylobacter* species are currently poorly understood, partly because identification of members of the genus is technically difficult. WGS has the potential to improve this aspect of microbiology, and, with this in mind, the current study used a diverse collection of 70 publicly available genomes to evaluate approaches for *in silico* speciation of *Campylobacter*, also generating a genus-wide core genome phylogeny that served as a benchmark for comparison. This study, and the integration of its findings into the *Campylobacter jejuni/coli* PubMLST database, represents a contribution towards an ongoing common endeavour: the development of a hierarchical, scalable *Campylobacter* WGS analysis workflow that is simple, transparent, and publicly accessible.

Three sequence similarity-based approaches for ortholog detection, the BIGSDB GENOME COMPARATOR module (Jolley & Maiden, 2010), OrthoMCL (Li *et al.*, 2003), and Proteinortho (Lechner *et al.*, 2011), were used to define a *Campylobacter* core genome. Among these programs, the web-based BIGSDB GENOME COMPARATOR module was easiest to use, but it was also the slowest, taking 22 hours to run. Although Proteinortho and OrthoMCL were faster at 4 and 16 hours, respectively, they require non-trivial bioinformatics expertise and resources; both programs are run in a Unix environment, and OrthoMCL requires a relational database and is comprised of 12 sequential scripts. Consequently, the BIGSDB GENOME COMPARATOR module represents a valuable core genome analysis tool for researchers with limited bioinformatics expertise or resources.

Although Proteinortho is a relatively new method that is less established in the field, this study suggests that it represents a powerful alternative to OrthoMCL. This is in agreement with the work of Lechner *et al*. (2011), who performed a benchmark comparison of Proteinortho and OrthoMCL and found the former less computationally intensive, with a significantly shorter runtime. The same study found that the two methods generated comparable clusters of orthologous genes; although beyond the scope of this thesis, work is ongoing to determine the level of concordance between orthologs defined using GENOME COMPARATOR, OrthoMCL and Proteinortho. These studies will facilitate the development of additional analysis schemes in the *Campylobacter jejuni*/*coli* PubMLST database, which will ultimately improve the identification and population annotation of *Campylobacter* species.

Previous studies on the *Campylobacter* core genome included six to 12 species (Lefébure & Stanhope, 2009; Zhang & Sievert, 2014; Iraola *et al.*, 2014), while others focused solely on *C. jejuni* and *C. coli* (Lefébure *et al.*, 2010; Méric *et al.*, 2014; Skarp-de Haan *et al.*, 2014). Using only publicly available assembled genomes, the current study encompassed 22 of the 27 known *Campylobacter* species, making it the most representative core genome study carried out to date. Comparison of putative orthologs identified by GENOME COMPARATOR, OrthoMCL, and Proteinortho yielded a provisional *Campylobacter* core genome of 657 genes, all of which were present in ≥ 90% of isolates (Figure 3.2). This estimate was similar in size to the 669-gene core genome defined for 12 *Campylobacter* species (Iraola *et al.*, 2014), but greater than the 571-ortholog core described for eight species (Lefébure *et al.*, 2010). That the core genome defined here was relatively large compared to previous

studies may appear counter-intuitive at first, as the core genome of a group of isolates typically diminishes with increasing sample size and diversity. This difference can, however, be attributed to the relatively low 90% cut-off that was applied in this study to compensate for the inclusion of draft genomes. Use of the standard stringent 100% inclusion threshold would have yielded core genomes of approximately 370-500 genes using the three approaches, in line with previous estimates.

The functional composition of the core genome was similar to that previously described for six diverse *Campylobacter* species (Zhang & Sievert, 2014): core loci predominantly corresponded to housekeeping genes, particularly those encoding proteins involved in transcription, translation, replication, and cell metabolism (Figure 3.4). In contrast, the proportion of core loci without functional information was reduced compared to the re-annotated *C. jejuni* subsp. *jejuni* NCTC 11168 genome (Parkhill *et al.*, 2000; Gundogdu *et al.*, 2007). That *Campylobacter* species share a common niche, the digestive tract of humans and other animals, may explain the bias of the core genome towards essential genes.

For over 300 years, microbiologists have strived to group bacteria into hierarchical categories based on stable, easily applicable rules, and these classifications remain important today, particularly in the field of public health (Doolittle & Papke, 2006; Georgiades & Raoult, 2010). Ideally, taxonomic definitions would correlate with evolutionary groups, but this is not always the case, as demonstrated for *Neisseria*

(Bennett *et al.*, 2012). This lack of congruence can be attributed to the fact that most existing bacterial species were defined in the pre-genomic era, based on morphological and phenotypic characteristics, or measures of genetic distance determined using techniques with limited discriminatory power, such as DNA-DNA hybridisation or 16S rRNA gene sequencing (Georgiades & Raoult, 2010). The core genome hypothesis postulates that the core genome is the unit that defines bacterial species, and that there will be limited lateral transfer of core loci, which will proceed to diverge post-speciation (Lan & Reeves, 1996, 2000, 2001). Phylogenetic analysis of the *Campylobacter* core genome demonstrated that, at the species level, taxa were equivalent to evolutionary groups. That species groups were generally separated by long branch lengths suggested substantial divergence post-speciation, with the exception of *C. lari* subspecies and *C. subantarcticus*, consistent with a low F_{ST} value of 0.641. The core genome phylogeny also resolved *Campylobacter* subspecies, barring *C. fetus subsp. fetus* and *C. fetus subsp. venerealis*. Previous whole-genome analyses of *C. fetus subsp. fetus* and *C. fetus subsp. venerealis*, including the isolates used in this study, found that phenotypic and genotypic characteristics for the two subspecies were not consistent; based on these findings, researchers called for a re-evaluation of the validity of phenotypic testing for the identification of *C. fetus* subspecies (van der Graaf-van Bloois *et al.*, 2014).

The core genome defined in this study will serve as a useful scheme for evolutionary analyses in the *Campylobacter jejuni/coli* PubMLST database; however, routine species identification based on the curation and analysis of 657 loci is impractical. For this reason, smaller groups of loci were evaluated for *in silico* identification of

Campylobacter species. With reference to single locus approaches, this study demonstrated that 16S rRNA gene sequences do not reliably identify *Campylobacter* species due to low levels of sequence diversity and inter-species allele sharing. As observed previously in studies based on amplified gene fragments (On, 2013), *C. jejuni* and *C. coli* could not be distinguished (Figure 3.6), which is problematic given their importance in human disease. In contrast, *atpA* gene sequences segregated according to species and the method only failed to differentiate *C. rectus* and *C. showae* (Figure 3.7). The findings presented here suggest that the *atpA* assay should be considered as an alternative to 16S rRNA gene sequencing in laboratories reliant on single locus sequencing for *Campylobacter* speciation, which was in agreement with the results of the *atpA* PCR assay evaluation (Miller *et al.*, 2014b). Furthermore, these findings have important implications for gut microbiome studies of humans and other animals, which are commonly based on 16S rRNA gene sequencing (Suau *et al.*, 1999).

Although *atpA* appears to reliably differentiate most *Campylobacter* species, there is evidence that the locus undergoes inter-species recombination (Miller *et al.*, 2005). In contrast to single locus typing methods, multilocus typing schemes that include loci distributed across the genome overcome the distortive impact of recombination and also provide increased resolution. For this reason, uCMLST and rMLST were also evaluated for *in silico* speciation of *Campylobacter*. The universal MLST scheme performed no better than *atpA*, also failing to differentiate *C. rectus* and *C. showae*. On the other hand, rMLST resolved all species except *C. lari* and *C. subantarcticus*, which shared alleles at several loci and were found to be highly

similar at the core genome level. A recent study of the comparative genomics of members of the *C. lari* group included in the current study showed that these organisms are highly similar (Miller *et al.*, 2014a). This suggests that it may be necessary to use cgMLST to differentiate members of the *C. lari* group.

In conclusion, the current study has provided insights into characteristics of the *Campylobacter* core genome and laid foundations for ongoing work that will improve the functionality of GENOME COMPARATOR and the *Campylobacter jejuni*/*coli* PubMLST database for genus-level evolutionary analyses. With respect to *in silico* identification of *Campylobacter*, *atpA* has the potential for rapid speciation of presumptive *Campylobacter* isolates, but rMLST represents a more powerful tool as it is applicable across the bacterial domain. These analyses highlight the utility of publicly available WGS data and show how it can be used to address problems in clinical microbiology. Finally, this work demonstrates how BIGSDB and PubMLST databases can be used to carry out genomic analyses without extensive bioinformatics resources or expertise.

Part II: The genomic epidemiology of *Campylobacter* from South Africa

CHAPTER 4

The genetic diversity of *Campylobacter* from two South African free-range broiler flocks at slaughter

ABSTRACT

Chicken remains the predominant source of human *Campylobacter* infections in many high-income countries. Recent research suggests that broiler flocks and retail meat from South Africa may have similar levels of *Campylobacter* contamination to those seen in Europe and the USA. The primary aim of this study was to use WGS to investigate the molecular epidemiology of *Campylobacter* isolates obtained at slaughter from two broiler flocks from a commercial free-range farm in South Africa. At the MLST level, all isolates corresponded to ST-1587, a member of the ST-828 clonal complex, which was also identified among contemporaneous human disease isolates from the same geographical region. Results of wgMLST analyses based on 1,513 coding sequences, equivalent to ~1.4×10^6 bp, indicated that the South African ST-1587 isolates were highly homogeneous. The isolates did not cluster according to source, flock, or carcass. Only 46 (3%) variable loci were identified, and most allelic variants occurred in individual isolates. Point mutations accounted for the majority of genetic differences in variable coding sequences, with 63.4 and 24.4% resulting in non-synonymous mutations and truncated products, respectively. Previous research has shown that *Campylobacter* from free-range broiler carcasses are genetically diverse, suggesting that this study represents a snapshot of a recently emerged, genetically stable *C. coli* clone blooming through the two flocks. More generally, this study demonstrates how WGS and the gene-by-gene approach

can be applied to improve our understanding of the epidemiology of *Campylobacter* from chickens, which is a prerequisite for the development of more effective intervention strategies.

4.1 Introduction

For any human pathogen, a thorough knowledge of the sources of infection is prerequisite for the development and monitoring of interventions designed to reduce disease incidence. The introduction of MLST advanced our understanding of the relative contributions of different food animals to human campylobacteriosis, as it facilitated the development of genetic models for the attribution of *Campylobacter* to sources of infection (Pritchard *et al.*, 2000; Corander & Marttinen, 2006; Wilson *et al.*, 2008; Mullner *et al.*, 2009a, 2009b). Population-based attribution studies of *C. jejuni* and *C. coli* have consistently implicated poultry, particularly chicken, as a primary source of human campylobacteriosis (Colles & Maiden, 2012). Between 58 and 78% of *Campylobacter* infections in Scotland, northwest England, the Netherlands, Denmark, Switzerland, and New Zealand have been attributed to chicken (Gormley *et al.*, 2008; Mullner *et al.*, 2009a; Sheppard *et al.*, 2009b; Mughini-Gras *et al.*, 2012; Kittl *et al.*, 2013; Boysen *et al.*, 2014). Consequently, research on the interplay between *Campylobacter* and chickens has become a priority, as our understanding remains incomplete.

Campylobacter is generally not detected in chickens during the first two to three weeks of life; however, once identified in a flock it spreads rapidly, probably via the faecal-oral route, as chickens shed the bacterium in their faeces and are coprophagic (Newell & Fearnley, 2003). Although *Campylobacter* is commonly described as a commensal of the avian gut, recent evidence suggests that it may in fact be pathogenic in chickens: infection in certain commercial broiler breeds results in intestinal inflammation, diarrhoea, and other signs of poor health, such as burns

to the legs, known as hock marks, and thickening and discolouration or lesions of the foot pad, known as pododermatitis (Humphrey *et al.*, 2014). These findings were consistent with the results of earlier studies of broiler flocks that found an association between *Campylobacter* infection and hock marks and pododermatitis, both of which are caused by a build-up of ammonia in the litter due to the accumulation of wet faeces, also a sign of poor gut health (Colles *et al.*, 2008; Bull *et al.*, 2008; Rushton *et al.*, 2009).

The prevalence of *Campylobacter* in broiler flocks varies among countries, and it is challenging to compare studies due to differences in sampling and isolation protocols. In a large UK-wide survey of commercial broilers, 79.2% of flocks were *Campylobacter*-positive (Lawes *et al.*, 2012), while the prevalence in other European countries has been shown to be highly variable, ranging from 47-84% in southern and central member states, and 4-12% in northern countries, particularly in Scandinavia (European Food Safety Authority and European Centre for Disease Prevention and Control, 2014). Similarly, a study of eight farms belonging to four major USA producers found that 87.5% of broiler flocks became *Campylobacter*-positive (Stern *et al.*, 2001). As is true of other areas of *Campylobacter* research, there is a lack of data from LMICs.

The contribution of chicken to the global campylobacteriosis burden is directly related to the high levels of contamination of retail meat (Gormley *et al.*, 2011). In a recent survey of over 4,000 chilled whole chickens obtained from major UK retailers,

73% tested positive for *Campylobacter*, and 19% carried high bacterial loads in excess of 1,000 colony-forming units per gram (Food Standards Agency, 2015). Similarly, the European Food Standards Authority found that up to 75.8% of broiler carcasses were *Campylobacter*-positive (EFSA Panel of Biological Hazards (BIOHAZ), 2011). Factors such as the large proportion of *Campylobacter*-positive birds in each flock, the high bacterial load in the chicken gut, and the automated slaughter process are thought to contribute to the contamination of retail chicken meat (Johnsen *et al.*, 2006), with genotypes showing differential survival rates the production process (Allen *et al.*, 2007; Berndtson *et al.*, 1996; Herman *et al.*, 2003; Klein *et al.*, 2007; Lienau *et al.*, 2007; Lindmark *et al.*, 2006). Voluntary and regulatory interventions have been introduced during and after production to attempt to reduce the incidence of disease, including surveillance, carcass freezing, and consumer education. Although these interventions have been successful in countries such as Norway, Sweden, Denmark, Iceland, and New Zealand (Stern *et al.*, 2003; Hofshagen & Kruse, 2005; Hansson *et al.*, 2007; Rosenquist *et al.*, 2009; Sears *et al.*, 2011), they have generally been impeded by consumer preferences and behaviours, or economic constraints (Wagenaar *et al.*, 2013).

Regarding the acquisition of *Campylobacter* by chickens, routes of transmission remain unclear. Vertical transmission, that is, the transfer of *Campylobacter* from breeder flocks to their offspring by transovarian transmission or faecal contamination of eggs, has proved controversial (Newell & Fearnley, 2003; Callicott *et al.*, 2006; Cox *et al.*, 2012; Agunos *et al.*, 2014). As reviewed by Agunos *et al.* (2014), 11 out of 32 studies showed evidence of vertical transmission, but conflicting

results reduce confidence in these findings. In comparison, horizontal transmission, that is, the acquisition of *Campylobacter* from agricultural, domestic, or wild animals, farm workers, or environmental sources, has been more widely accepted. Inadequate disinfection of the barn, the presence of neighbouring broiler flocks, and rapid turnover between flocks have been identified as the major risk factors for colonisation (Agunos *et al.*, 2014).

Given the potential role of the environment in transmission, there has been a strong emphasis on biosecurity to reduce colonisation of broiler flocks (Newell & Fearnley, 2003; Wagenaar *et al.*, 2013; Agunos *et al.*, 2014). Biosecurity entails limiting the exposure of broiler flocks to *Campylobacter* prior to and during slaughter, and measures include barriers to prevent contact with other animals, restrictions on human movements, as well as disinfection of houses, transport crates, and equipment (Agunos *et al.*, 2014). While biosecurity has had some success in housed broiler flocks, (Humphrey *et al.*, 1993; van de Giessen *et al.*, 1998; Hofshagen & Kruse, 2005), interventions have generally had a limited impact on human disease (Wagenaar *et al.*, 2013). Also problematic is the increasing demand from consumers for free-range and organic chicken. Free-range and organic flocks have high rates of *Campylobacter* colonisation and frequently carry multiple genotypes (Newell & Fearnley, 2003; Van Overbeke *et al.*, 2006); however, these farming systems are not amenable to biosecurity interventions.

The issues summarised here highlight the need for further information on the biology and epidemiology of *Campylobacter* in chicken. At present, the majority of our knowledge in this area is based on studies carried out in Europe and North America. Concerning previous research carried out in South Africa, a PubMed query using the terms "*Campylobacter* AND 'South Africa' AND chicken" retrieved seven studies, one of which focused on ostriches. A comparison of samples obtained from different farming systems in the KwaZulu-Natal province of South Africa (Figure 1.11) reported that 47% of commercial free-range and housed broilers, 68% of chickens from rural production systems, and 94% of industrial breeders were *Campylobacter*-positive (Bester & Essack, 2012). Two studies carried out in the Gauteng province (Figure 1.11) found that 24-32% of chicken carcasses, livers, and intestines, all of which are commonly consumed in South Africa, were contaminated (van Nierop *et al.*, 2005; Bartkowiak-Higgo *et al.*, 2006), while *Campylobacter* was not detected using culture to screen samples of cooked chicken, beef, gravy, and salad sold by street vendors in the same province (Mosupye & von Holy, 2000).

Taken together, the results of previous studies carried out in South Africa suggest that live chickens and retail meat are likely contaminated with *Campylobacter* at rates similar to those observed in high-income countries; however, none of the studies carried out molecular typing. As a result, there is no information available regarding the relationships among *Campylobacter* from South African chickens and local disease isolates or global collections. With this in mind, the primary aim of the current study was to use WGS to characterise the genetic diversity of chickens from

a South African commercial free-range farm in the Western Cape province (Figure 1.11).

4.2 Materials and Methods

4.2.1 Isolates

In a previous study, researchers at the University of the Western Cape (UWC), South Africa, compared the efficacy of two protocols for the detection of *Campylobacter* contamination on chicken carcasses at the end of the production process: the BS EN 5763 Horizontal Method for Detection of Thermotolerant *Campylobacter* (ISO 10272-1:2006) (Anon, 2006) and the Cape Town protocol (Lastovica, 2006). Meat and neck skin samples were obtained from free-range broiler flocks, which were slaughtered at age 38-41 days. The flocks were from a single commercial free-range farm located in the Western Cape province of South Africa (Figure 1.11). The farm supplied several retailers during the study period, with distribution limited to the Western Cape.

On arrival at UWC, samples were homogenised in Bolton selective enrichment broth and incubated at 37°C for four to six hours, and then at 42°C for a further 44 hours, after which each sample was subjected to both the ISO 10272-1:2006 and Cape Town protocols (Anon, 2006; Lastovica, 2006). As described in the ISO 10272-1:2006 protocol (Anon, 2006), enrichment cultures were inoculated onto modified charcoal cefoperazone deoxycholate agar (mCCDA) and incubated in 5% CO_2 at 42°C for 48 hours. Presumptive *Campylobacter* colonies were subcultured onto CBA and incubated in 5% CO_2 at 42°C for 48 hours. With respect to the Cape Town protocol (Lastovica, 2006), 200 µl of enrichment broth was inoculated onto a 0.45 µm filter placed in the centre of an antibiotic-free tryptose agar plate

supplemented with horse blood (TBA). After 15 minutes, the filters were removed and the plates were incubated in a hydrogen-enriched microaerobic atmosphere at 37°C for up to six days. The plates were checked for growth every 48 hours, and presumptive *Campylobacter* colonies were inoculated onto fresh TBA plates and incubated as for the primary cultures. Preliminary species identification of putative *Campylobacter* isolates was based on microscopy and biochemical tests as outlined in the ISO 10272-1:2006 method (Anon, 2006) and the Cape Town protocol (Lastovica, 2006). Isolates were subsequently stored at −70°C in Microbank cryogenic vials.

A random subset of 37 *Campylobacter* isolates, chosen based on culture viability, was included in the current study (Appendix E). The isolates were presumptive *C. coli* (*n* = 29) and *C. lari* (*n* = 8) collected from carcasses from two broiler flocks (designated flocks “A” and “B”) slaughtered eight weeks apart in September (*n* = 14) and November 2011 (*n* = 23). The collection included nine sets of duplicate isolates that were obtained from the same carcass from different sample types, or using different isolation protocols (Appendix E). Isolates were transported to the University of Oxford on sterile swabs in Amies transport medium with charcoal (VWR International Ltd., Lutterworth, UK). Swabs were streak plated onto CBA (Oxoid Ltd., Basingstoke, UK) and incubated at 42°C for 48 hours in a microaerobic atmosphere generated using the GENbox microaer system (bioMerieux UK Ltd., Basingstoke, UK). Plates without growth were incubated for a further 48 hours. In the case of isolates without growth at 96 hours, a second attempt was made to recover *Campylobacter* from the original swab following the same protocol. For long-term

storage of viable isolates, culture suspensions were made in brain-heart infusion (BHI) broth with 20% glycerol, and stored at −80°C.

4.2.2 Whole-genome sequencing

In preparation for WGS, genomic DNA was extracted from all viable isolates using the Wizard Genomic DNA Purification Kit (Promega, Southampton, UK), according to the manufacturer's instructions. Genomic DNA was sent to the Wellcome Trust Sanger Institute, Cambridge, UK, where Illumina paired-end multiplex libraries were prepared and WGS carried out as recommended by the manufacturer. Briefly, 1 µg of genomic DNA was sheared to produce 200-300 bp fragments using a Covaris E210 acoustic shearing device. The resulting DNA fragments were repaired to remove and fill in 3′ and 5′ overhangs, respectively, and phosphorylate the 5′ ends. A single adenosine residue was added to the blunt 3′ ends of the fragments, to which the Illumina multiplexing adaptor oligonucleotides were then ligated. Up to 96 libraries were pooled in equimolar ratios and sequenced on a single flow cell lane of an Illumina HiSeq 2000, producing 101 bp paired-end reads.

4.2.3 *De novo* assembly of whole-genome sequencing data and gene annotation

Short-read data were assembled with an in-house assembly pipeline (J. Bray, unpublished). The pipeline generated *de novo* assemblies using VELVET version 1.2.01 (Zerbino & Birney, 2008), with assembly parameters optimised using the

bundled VELVETOPTIMISER wrapper (Zerbino, 2010). The resulting contigs were automatically uploaded to the *Campylobacter jejuni/coli* PubMLST database and linked to European Nucleotide Archive accession numbers where available. Genes used for species identification, namely *atpA* and the rMLST loci, were annotated using a combination of automated and manual curation as outlined in Chapter 2 [2.2.2]. MLST loci were annotated following *in silico* species identification, carried out as described in Chapter 3 [3.2.3.2], which guided the choice of an appropriate MLST scheme.

4.2.4 Genomic analyses

Isolates were compared at the MLST, rMLST, and wgMLST levels using the BIGSDB GENOME COMPARATOR module [1.5.2.2.3.4] implemented in the *Campylobacter jejuni/coli* PubMLST database. MLST and rMLST comparisons were carried out using annotated sequences extracted from the database, while wgMLST was carried out using the finished genome of *C. coli* isolate 15-537360 (GenBank accession number CP006702) (Pearson *et al.*, 2013) as a reference sequence. Default BLAST parameters were used for all GENOME COMPARATOR analyses, corresponding to a minimum sequence identity of 70%, minimum alignment of 50%, and a word size of 20. Additional GENOME COMPARATOR options were selected such that putative paralogs identified in all isolates, as well as loci that were incomplete in at least one isolate, were excluded from the distance matrix calculations. The resulting distance matrices were visualised as networks using the NeighborNet algorithm (Bryant & Moulton, 2004) implemented in SplitsTree4 version 4.13.1 (Huson & Bryant, 2006).

4.2.5 Further analysis of variable loci

Variable loci identified by GENOME COMPARATOR were assigned to COG groups based on orthologous relationships among CDSs from *C. coli* 15-537360 and *C. jejuni* NCTC 11168 that were detected as described in Chapter 3 [3.2.2.2]. MAFFT (Katoh & Standley, 2013) nucleotide sequence alignments automatically generated by GENOME COMPARATOR were inspected to characterise underlying genetic differences among the alleles identified at each variable locus.

4.3 Results

4.3.1 Whole-genome sequencing data and species identification

WGS data were obtained for 28/37 (75.7%) *Campylobacter* isolates (Appendix E) obtained from two flocks originating from a single free-range farm in the Western Cape. The remaining nine isolates were not viable on arrival at the University of Oxford and consequently were not sequenced. Genome sequences were obtained for 10/14 (71.4%) isolates from flock A, and 18/23 (78.3%) from flock B. These genomes included six sets of duplicate isolates that were recovered from the same chicken carcass, with isolates from flocks A and B obtained from six and 14 carcasses, respectively (Appendix E). These data can be accessed via the *Campylobacter jejuni*/*coli* PubMLST database under the project name 'UWC'.

Assembly statistics, such as genome size, number of contigs, and N_{50}, were consistent across the study isolates. That the assemblies were approximately 1.7 Mb and had large N_{50} values, in combination with small numbers of contigs, suggested that the WGS data were of a high quality (Appendix E). As further evidence of the quality of the genomic data, complete MLST and rMLST profiles were obtained for all isolates. *In silico* speciation based on *atpA* and rMLST indicated that all isolates were *C. coli*, including six that were previously identified as *C. lari* based on positive indoxyl acetate hydrolysis test results (Appendix E).

4.3.2 Genomic analyses

The hierarchical gene-by-gene approach implemented in the BIGSDB GENOME COMPARATOR module was exploited to characterise diversity within and between isolates from flocks A and B at increasing levels of resolution, progressing from MLST to rMLST to wgMLST.

4.3.2.1 Multilocus sequence typing

All isolates were identical at the MLST level, and corresponded to ST-1587, which belongs to the ST-828 clonal complex. As of January 2016, only one ST-1587 isolate was present in the *Campylobacter jejuni*/*coli* PubMLST database: tico2087. This isolate was recovered from the stool of a patient with gastroenteritis in Denmark, 2001. A single ST-1587 isolate, UCT020, was also identified among *Campylobacter* collected for the study presented in Chapter 5 of this thesis, which describes the epidemiology of human disease isolates from Cape Town, South Africa. Cape Town is the capital city of the Western Cape (Figure 1.11), the only province supplied by the farm that provided the study samples (Figure 1.11). UCT020 was isolated from the stool of a 39-year-old female in-patient with sepsis in December 2011, approximately three weeks after flock B was sampled.

4.3.2.2 Ribosomal multilocus sequence typing

Given that rMLST indexes variation in 52 complete gene sequences, equivalent to ~20,670 bp, it has greater discriminatory power compared to that of MLST, which

includes ~3,300 bp across seven loci; however, only two rMLST profiles were detected among the study isolates, one of which, rST-3857, accounted for 27/28 (96.4%) isolates. The second profile, rST-14375, found in UWC023, differed only at *rplE*. Examination of the *rplE* gene sequences found that UWC023 carried a single synonymous substitution at nucleotide position 414.

4.3.2.3 Whole-genome multilocus sequence typing

The gene-by-gene approach is flexible and scalable, which allows users to catalogue genome-wide variation to resolve relationships among closely related isolates, as was necessary in the current study. The study isolates, as well as ST-1587 isolates UCT020 and tico2087 (provided by Eva Litrup, Statens Serum Institute, Denmark) were compared using wgMLST; *C. coli* 15-537360, also assigned to clonal complex ST-828, was chosen as a reference genome. Among the 1588 CDSs annotated in 15-537360, 75 were excluded from the GENOME COMPARATOR distance matrix calculations: 39 (2.5%) were not detected in any of the isolates; 26 (1.6%) were incomplete in at least one isolate; and 10 (0.6%) were identified as putative paralogs in all isolates.

Within-flock wgMLST comparisons revealed high levels of homogeneity in both sets of isolates (Figure 4.1). Only 23 variable loci were identified among isolates from flock A, and 34 among those from flock B; pairwise distances between isolates were very low, ranging from 3-15 and 1-16 loci for flocks A and B, respectively. Surprisingly, isolates obtained from the same carcass were, in some instances, less

similar to each other than they were to those from different carcasses (Figure 4.1; Appendix F).

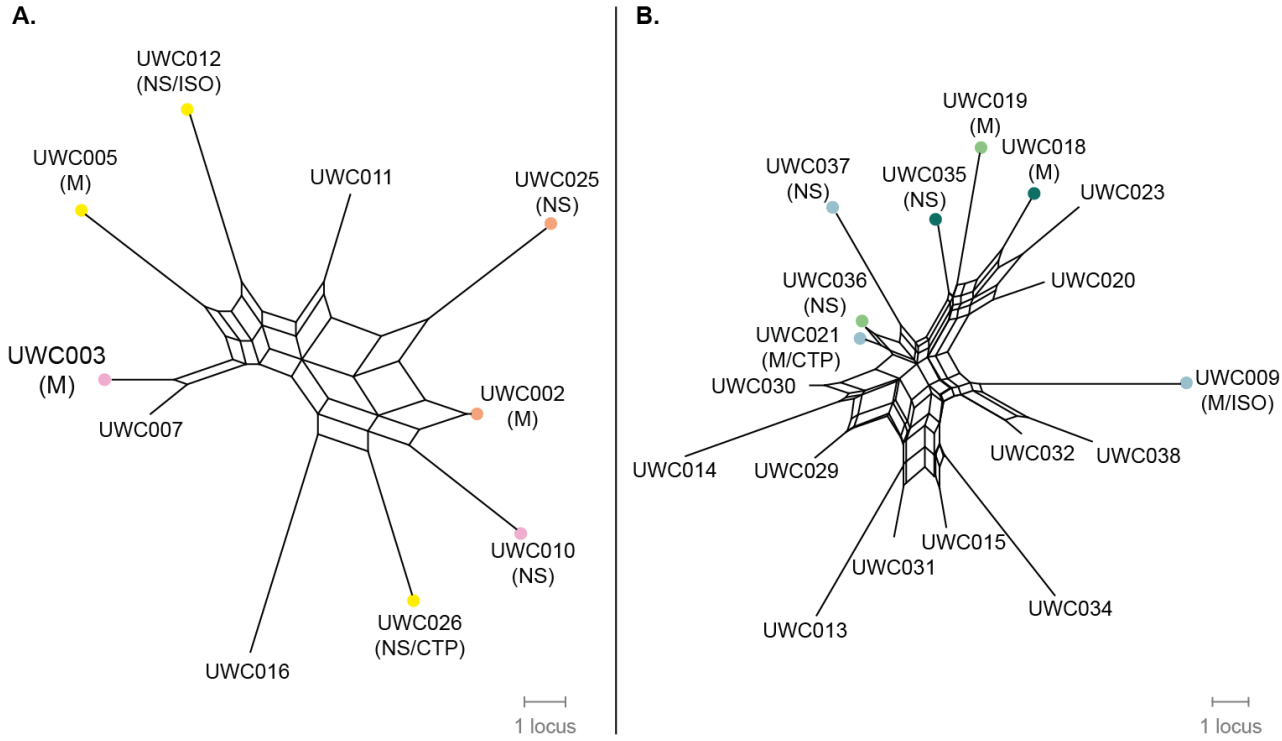


Figure 4.1 Genomic diversity among *Campylobacter* isolates from two free-range broiler flocks. Using GENOME COMPARATOR, isolates were compared at 1,513 loci. NeighborNet graphs were generated based on the 23 variable loci identified among isolates from flock A (**A**) and 34 from flock B (**B**). Duplicate isolates obtained from the same chicken carcass are indicated by coloured circles; additional provenance data are provided in parentheses for these isolates. (M), chicken meat sample; (NS), chicken neck skin sample; (ISO), isolated with the ISO 10272-1:2006 protocol; (CTP), isolated with the Cape Town protocol.

The combined comparison of isolates from flocks A and B, UCT020, and tico2087 indicated that the ST-1587 isolates were distinct from isolate 15-537360, differing at an average of 1154/1513 (76.3%) loci (Figure 4.2). Although the ST-1587 isolates were more similar to each other, tico2087 was separated from the South African isolates by an average of 594 loci, equivalent to ~39% of included CDSs. The South African isolates, including UCT020, were highly similar, with only 46 (3%) variable loci identified. The isolates did not segregate by flock (Figure 4.2). Distances between pairs of South African isolates ranged from 1-20 loci, with an average distance of ~10 loci, which was comparable to the levels of within-flock diversity. Duplicate isolates obtained from the same chicken carcass were separated by pairwise distances of 6-11 loci, with some more similar to isolates from other birds than each other (Figure 4.2; Appendix F), as observed in the intra-flock comparisons (Figure 4.1). Also noteworthy was the fact that UCT020, the clinical ST-1587 isolate from Cape Town, was nested within the diversity of the chicken isolates, differing by just 5-14 loci (Figure 4.2).

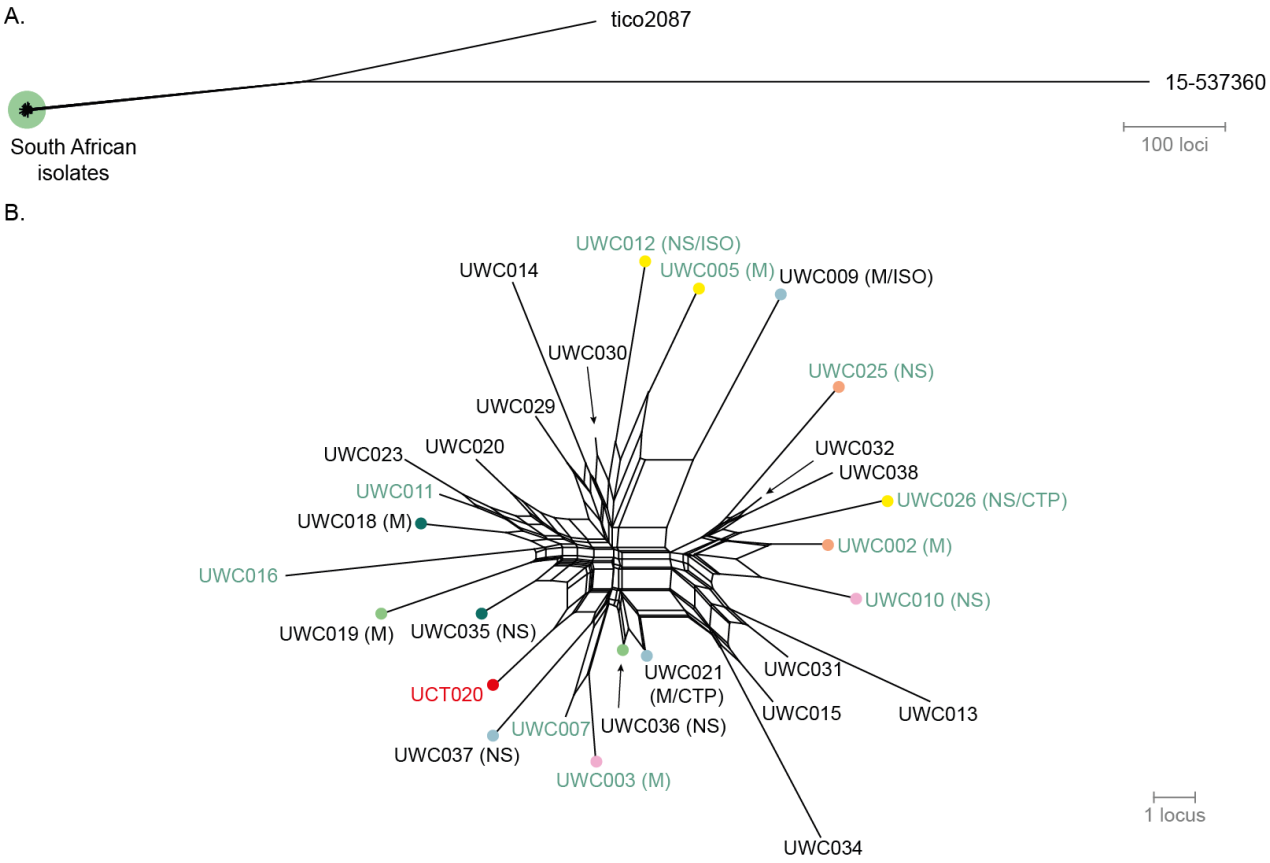


Figure 4.2 Comparison of the genomic diversity of ST-1587 isolates from chicken carcasses and human disease samples from South Africa and Denmark. Isolates were compared at 1,513 loci, using *C. coli* 15-537360 as a reference genome. (**A**) NeighborNet graph generated based on 1,178 loci that were variable among the South African ST-1587 isolates (green), tico2087 (an ST-1587 human isolate from Denmark), and 15-537360. (**B**) Detailed view of South African ST-1587 isolates, including those from flocks A (green text) and B (black text), and a human disease isolate from Cape Town (red). Duplicate isolates obtained from the same chicken carcass are marked with coloured circles and additional provenance data are provided in parentheses. (M), chicken meat sample; (NS), chicken neck skin sample; (ISO), isolated with the ISO 10272-1:2006 protocol; (CTP), isolated with the Cape Town protocol.

4.3.2.4 Further analysis of variable loci identified by whole-genome multilocus sequence typing

Variable genes detected among the South African isolates were evenly distributed across the reference genome (Figure 4.3). Assignment of these variable loci to COG groups, based on orthologous clusters detected between CDSs from 15-537360 and NCTC 11168, showed that the majority (54.3%) encoded proteins involved in cellular processes and signalling, or metabolism (Figure 4.4). Within these two broad functional categories, genes involved in cell wall, membrane, and envelope biogenesis, in addition to energy production and conversion, were most frequently observed. Genes encoding proteins involved in replication, transcription, and translation accounted for a relatively small proportion (10.9%) of the variable loci, while 16 (34.8%) genes were poorly characterised or lacked functional information (Figure 4.4).

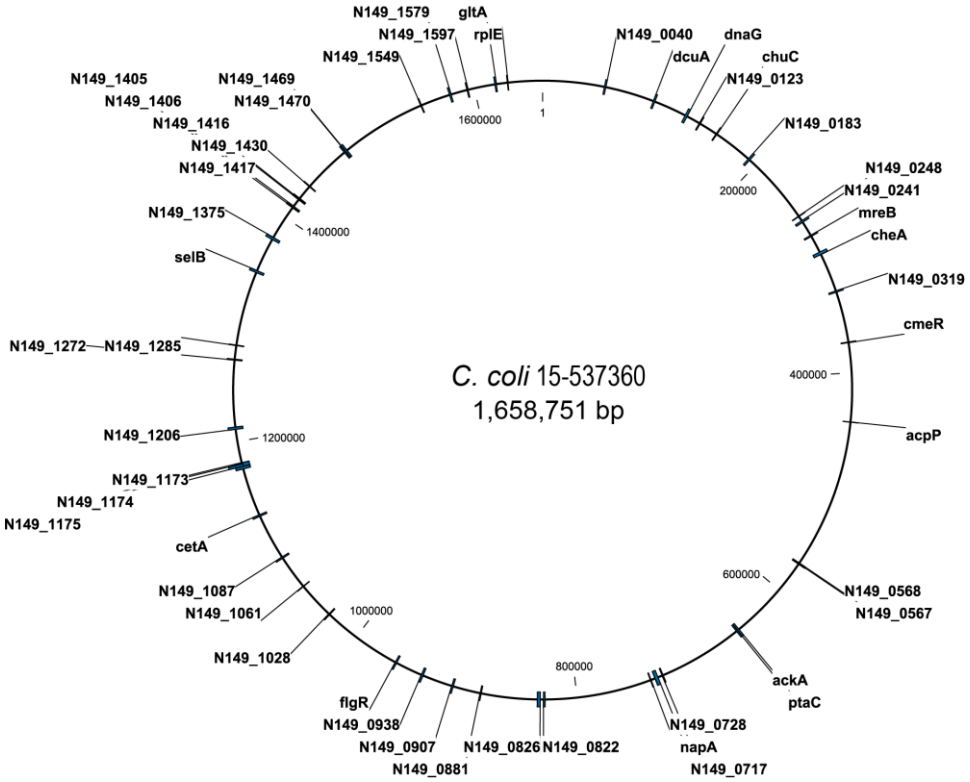


Figure 4.3 Distribution of variable loci across the genome of *C. coli* 15-537360.

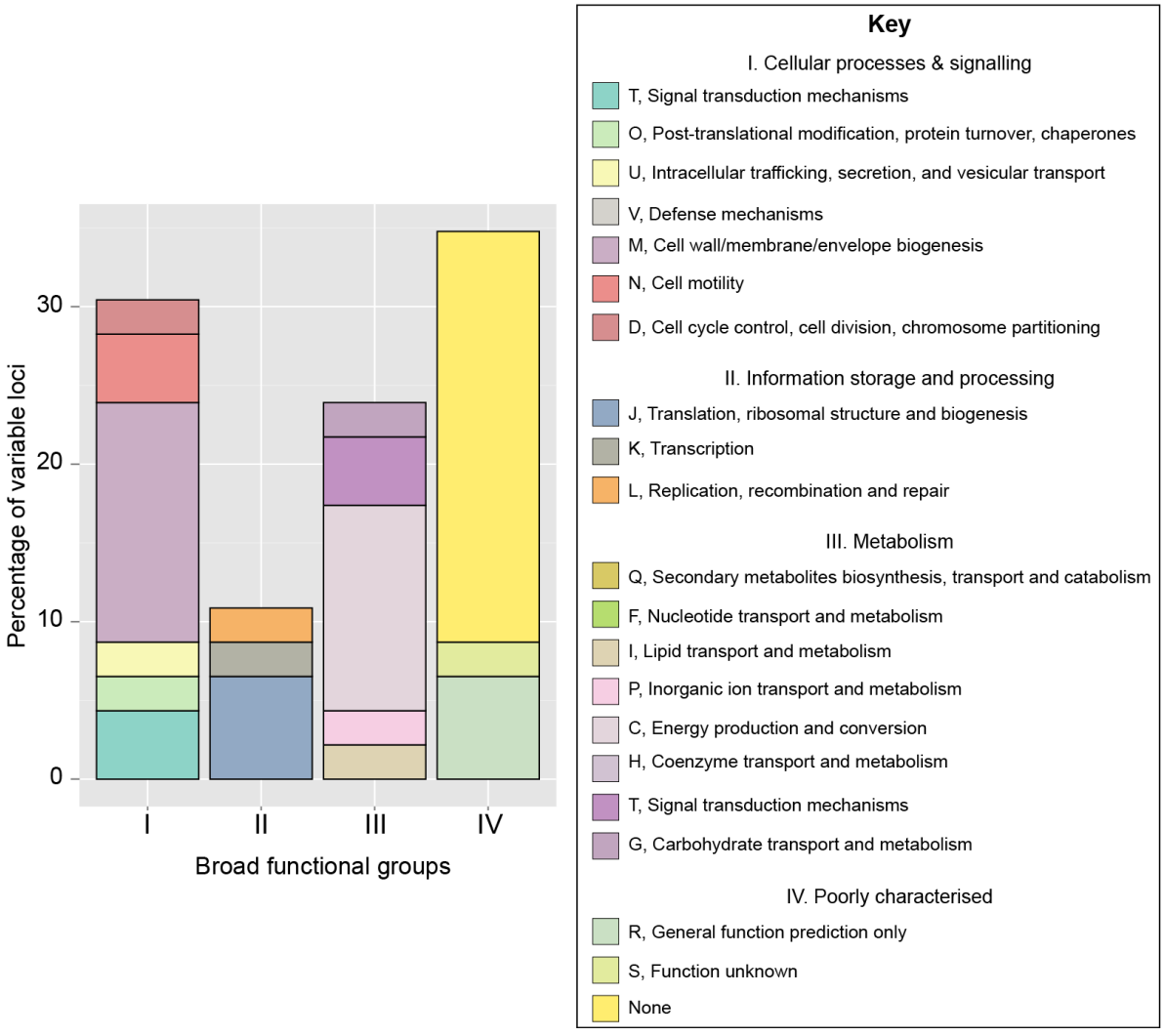


Figure 4.4 Assignment of variable loci to functional categories. Assignment to COG category was determined based on orthologous clusters detected between CDSs present in 15-537360 and NCTC 11168, with categories grouped as previously described by Zhang *et al*. (2014).

The number of alleles per variable gene ranged from two to four, with 43 (93.5%) loci including just two alleles. The distribution of alleles was uneven: variants at 34 (73.9%) loci were observed at a ratio of 28:1 (Figure 4.5). To explore the genetic variation underlying these allelic differences, nucleotide sequence alignments generated by GENOME COMPARATOR for all variable loci were examined. Eight (17.4%) of the alignments contained large internal gaps, or varied in length at the 5′ and 3′ ends, and so were excluded from further analyses (Table 4.1). Among the remaining loci, 31 (75.6%) variant alleles corresponded to single point mutations, 9 (22%) to differences in the lengths of homopolymeric tracts (defined as two or more adjacent nucleotides of the same type), and 1 (2.4%) to a deletion of 135 bp (Table 4.2). The majority of mutations (87.8%) altered the predicted protein sequence, resulting in an amino acid substitution (63.4%) or a truncated product (24.4%).

Table 4.1 Variable loci with poor quality nucleotide sequence alignments

Locus[a]	Predicted product	Broad functional category[b]
N149_0115 (*dnaG*)	DNA primase	II
N149_0123	Hypothetical protein	IV
N149_0728	AzlC family protein	IV
N149_1087	CDP-4-dehydro-6-deoxy-D-glucose 3-dehydratase	IV
N149_1173	Putative periplasmic protein	IV
N149_1174	Filamentous hemagglutinin domain protein	IV
N149_1285	Putative transferase	IV
N149_1430	Hypothetical protein	IV

[a] Locus tag and, where available, gene name derived from the 15-537360 reference genome
[b] (I), cellular processes and signalling; (II), information storage and processing; (III), metabolism; and (IV), poorly characterised or no functional information available.

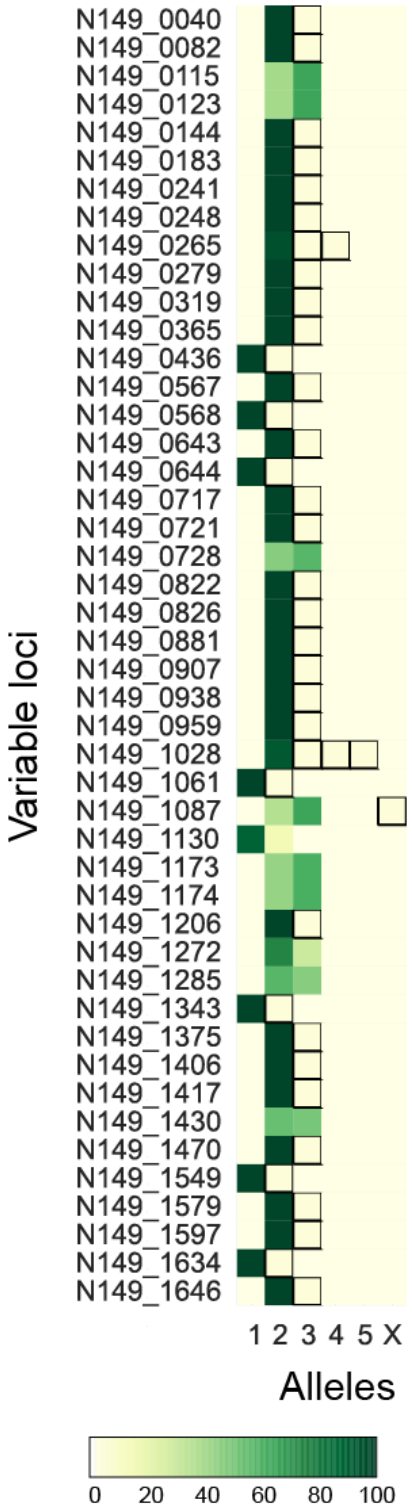


Figure 4.5 The distribution of allelic variants at variable loci identified among *Campylobacter* from two free-range chicken flocks and a human disease isolate from the Western Cape, South Africa. Allelic variants are shaded according to prevalence (%), as indicated by the gradient bar shown at the bottom of the figure. Alleles are numbered in order of discovery, with '1' used to denote sequences identical to that of the reference, and 'X' for a missing locus. Those variants present in a single isolate (3.4%) are shown with a black border for readability.

Table 4.2 Details of underlying genetic diversity in variable loci with high quality nucleotide sequence alignments

Locus[a]	Predicted product	Broad functional category[b]	Mutation(s)	Predicted effect on protein sequence
N149_0040	Putative efflux protein	III	5T/4T	Substitution
N149_0082 (*dcuA*)	Anaerobic C4-dicarboxylate transporter	III	GGA/AGA	Substitution
N149_0144 (*chuC*)	Hemin uptake system ATP-binding protein	III	GAC/GAT	None
N149_0183	Oligopeptide ABC transporter	III	GGG/CGG	Substitution
N149_0241	Hypothetical protein	IV	9A/8A	Truncated
N149_0248	Putative membrane protein	I	GTT/ATT	Substitution
N149_0265 (*mreB*)	Rod shape-determining protein	I	GCA/GTA	Substitution
			CTT/TTT	Substitution
N149_0279 (*cheA*)	Signal transduction histidine kinase	I	GTT/ATT	Substitution
N149_0319	Hypothetical protein	IV	GCT/ACT	Substitution
N149_0365 (*cmeR*)	Transcriptional repressor of CmeABC operon	II	ACC/ATC	Substitution
N149_0436 (*acpP*)	Acyl carrier protein	III	GCT/ACT	Substitution
N149_0567	Quinone-reactive Ni/Fe hydrogenase, cytochrome b subunit	III	TTT/TTA	Substitution
N149_0568	Hydrogenase maturation protease	III	GGT/AGT	Substitution
N149_0643 (*ptaC*)	Phosphate acetyltransferase	III	ACC/ATC	Substitution
N149_0644 (*ackA*)	Acetate kinase	III	GCT/GCC	None
N149_0717	Putative periplasmic protein	IV	Deletion (135 bp)	Truncated
N149_0721 (*napA*)	Periplasmic nitrate reductase precursor	III	GGT/GTT	Substitution
N149_0822	Putative lipoprotein	I	GGG/GAT	Substitution
N149_0826	Protein export cytoplasm protein SecA ATPase RNA helicase	I	GTG/ATG	Substitution
N149_0881	Putative periplasmic protein	I	CCT/TCT	Substitution
N149_0907	Putative transmembrane transport protein	IV	CCC/CCT	None
N149_0938	Cell division protein FtsH	I	GTA/GCA	Substitution
N149_0959 (*flgR*)	Signal-transduction regulatory protein	I	8A/7A	Truncated
N149_1028	Peptidase, M23/M37 family	I	GGA/AGA	Substitution
			2T/3T	Truncated
			CAA/TAA	Truncated
N149_1061	Chemotaxis regulator	I	7A/6A	Truncated
N149_1130 (*cetA*)	Signal transduction protein	I	8T/9T	Truncated
N149_1206	Outer membrane protein	I	ACT/ATT	Substitution
N149_1272	3-oxoacyl-[acyl-carrier-protein] synthase	IV	9G/8G (pseudogene)	Truncated

Table 4.2 cont.

Locus[a]	Predicted product	Broad functional category[b]	Mutation(s)	Predicted effect on protein sequence
N149_1343 (*selB*)	Selenocysteine-specific translation elongation factor	II	5T/4T	Truncated
N149_1375	Capsular polysaccharide biosynthesis/export	I	GCT/ACT	Substitution
N149_1406	tRNA pseudouridine 13 synthase	II	GAG/GAA	None
N149_1417	Putative membrane protein	IV	TCA/TAA	Truncated
N149_1470	Putative C4-dicarboxylate anaerobic carrier	IV	4A/3A	Substitution
N149_1549	Queuosine Biosynthesis QueE Radical SAM	IV	ACT/GCT	Substitution
N149_1579	Hypothetical protein	I	ATG/ATA	Substitution
N149_1597	Putative periplasmic protein	IV	ACA/ATA	Substitution
N149_1634 (*gltA*)	Citrate synthase	III	GCT/ACT	Substitution
N149_1646 (*rplE*)	Ribosomal protein L5p (L11e)	II	TTC/TTT	None

[a] Locus tag and, where available, gene name derived from the 15-537360 reference genome.
[b] (I), cellular processes and signalling; (II), information storage and processing; (III), metabolism; and (IV), poorly characterised or no functional information available.

4.4 Discussion

A detailed knowledge of the dynamics of *C. jejuni* and *C. coli* genotypes in chicken along the farm-to-fork continuum has proved elusive because of the inherent genetic diversity of these organisms, and their ubiquity in agricultural, domestic, and wild animals (Dingle *et al.*, 2001a). Although MLST-based studies have improved our understanding in this area (Colles & Maiden, 2012), research is ongoing, particularly with reference to defining routes of transmission in chickens. WGS, in conjunction with the gene-by-gene approach implemented in BIBSDB, has been used to study the epidemiology of human campylobacteriosis (Cody *et al.*, 2013; Revez *et al.*, 2014a, 2014b; Fernandes *et al.*, 2015), and in a small study of five *C. jejuni* isolates from chicken farms (Kivistö *et al.*, 2014); however, WGS has not yet been extensively exploited to address questions pertaining to *Campylobacter* infection in chickens. The current study demonstrates the utility of WGS and the gene-by-gene approach for characterising the genetic diversity of *Campylobacter* from chickens, providing insights into the global and local epidemiology of isolates obtained from carcasses from two broiler flocks from a single free-range farm in South Africa.

One of the advantages of the gene-by-gene approach to genomics is that it is compatible with existing MLST schemes. As a result, it was possible to compare the study isolates to those in the *Campylobacter jejuni*/*coli* PubMLST database, thereby contextualising the epidemiology of South African *Campylobacter* within a global framework. At the ST level, the study isolates were uniform and corresponded to ST-1587. Interestingly, a single ST-1587 isolate, UCT020, was detected among contemporaneous human disease isolates collected from Cape Town, which falls

within the distribution range of the farm that provided the chickens sampled for the current study (Figure 1.11). The ST was represented only once in the PubMLST database, by human disease isolate tico2087, which was cultured from the stool of a patient with gastroenteritis in Denmark, in 2001. ST-1587 was assigned to the ST-828 clonal complex, which is the predominant *C. coli* human disease lineage (Sopwith *et al.*, 2010). This generalist agricultural lineage forms part of *C. coli* clade 1 (Sheppard *et al.*, 2008, 2010a), and has likely been disseminated internationally by the global distribution of food animals (Sheppard *et al.*, 2010b).

Strikingly, the lack of diversity observed among the study isolates was not confined to the MLST level. Comparison of 1,513 loci indicated that the study isolates and UCT020 were distinct from tico2087, yet highly homogeneous: 23 and 34 variable loci were detected among isolates belonging to flock A and flock B, respectively, with a total of just 46 (3%) variable loci identified among the chicken isolates and UCT020 (Figure 4.2). It was also noted that, in some cases, isolates obtained from the same carcass were more similar to those from other carcasses within or between the two flocks than they were to each other (Appendix F). This lack of diversity contrasted sharply with the results of earlier studies, which found that free-range and organic flocks were genetically diverse, carrying multiple genotypes of *C. jejuni* and/or *C. coli* (Newell & Fearnley, 2003; El-Shibiny *et al.*, 2005; Van Overbeke *et al.*, 2006; Colles *et al.*, 2008, 2010, 2011, 2015). Additionally, it has been reported that *Campylobacter* isolates obtained after slaughter are more diverse than those from live birds, based on the recovery of up to eight STs from individual carcasses (Colles *et al.*, 2010). Instead, the results from the current study

were consistent with more recent WGS studies of *C. jejuni* ST-474 and *C. jejuni* ST-677 clonal complex isolates from chickens (Kivistö *et al.*, 2014; Friedrich *et al.*, 2015). Friedrich *et al*. (2015) challenged 21-day-old chicks with a single ST-474 strain, and then carried out WGS to explore the short-term evolution of this genotype. Mapping of short-read data from 27 isolates recovered from the caeca of 12 chicks over three days identified 15 SNPs, with no evidence of clustering by bird or isolation date. Similarly, Kivistö *et al*. (2014) used wgMLST to investigate the relationships among five ST-677 clonal complex isolates from two broiler farms in Finland, and found no differences among isolates with identical STs obtained from the same farm. Both of the studies described here focused on *C. jejuni* isolates, and similar work is yet to be carried out for *C. coli*.

Pairwise distances among the study isolates and UCT020 ranged from 1-20 loci, which was similar to the levels of diversity detected among *Campylobacter* isolates obtained from repeat sampling of individual patients, outbreaks, or re-sequencing of individual isolates. Same-patient *C. jejuni* isolates from the Oxfordshire sentinel surveillance genome collection differed at 3-14 loci (Cody *et al.*, 2013), while pairwise differences among same-patient ST-828 clonal complex *C. coli* isolates were slightly lower, ranging from 6-8 loci (A. Cody, unpublished data). Similarly, pairwise distances among *C. jejuni* isolates recovered from human disease samples, milk, and cattle faeces over eight weeks during a milk-borne *Campylobacter* outbreak ranged from 2-12 loci (Revez *et al.*, 2014b). The average number of differences between pairs of study isolates (~10 loci) was comparable to

the pairwise distances detected after re-sequencing individual isolates, which ranged from 1-7 loci (Cody *et al.*, 2013).

Further analysis of the 46 variable loci identified among the South African ST-1587 isolates showed that 54.3% encoded products involved in cellular processes and signalling, or metabolism (Figure 4.4; Table 4.2), with 73.9% of allelic variants observed in single isolates only (Figure 4.5). The underlying genetic variation typically corresponded to single nucleotide substitutions or changes in the lengths of homopolymeric tracts, which was consistent with the results of wgMLST analyses of same-patient and outbreak isolates (Revez *et al.*, 2014b; Cody *et al.*, 2013). Long G/C tracts (≥ 8 nucleotides) are common in the *C. jejuni* genome, and often occur in genes that encode products involved in synthesis or modification of surface antigens, regulating their expression through phase variation (Parkhill *et al.*, 2000); however, only 4/9 homopolymeric tracts identified in the current study included more than eight nucleotides, and all but one corresponded to poly(A/T) repeats (Table 4.2). Although poly(A/T) tracts are not as well characterised as their poly(G/C) counterparts, changes in the length of a poly(T) tract are known to result in phase variation in *flhA* in *C. coli* (Park *et al.*, 2000). Additionally, *flgR*, a variable locus that carried a poly(A) tract in this study (Table 4.2), has previously been shown to be phase variable in *C. jejuni* (Hendrixson, 2006). Given that most studies on phase variation in *Campylobacter* have focused on *C. jejuni*, further work in this area may elucidate signatures specific to *C. coli*.

That 87.8% of all allelic variants detected in this study corresponded to altered protein sequences, 24.4% of which were truncated, suggested the observed variants may represent rare, transient alleles unlikely to persist in the population. These findings were similar to those of Friedrich *et al*. (2015), who found that 14/15 SNPs identified among ST-474 isolates from experimentally challenged chicks were non-synonymous. Six of the non-synonymous mutations identified in that study were found in *mreB*, for which three alleles were detected in the current study (Table 4.2; Figure 4.5), suggesting the locus may be under selection.

Reasons for the genetic homogeneity observed in this study are unclear. The possibility that small sample size, the inclusion of an enrichment step, and subsequent follow-up of a subset of colonies might have obscured the true diversity of *Campylobacter* from the two flocks cannot be excluded. However, the age of the chickens at slaughter may have contributed to the observed homogeneity. Longitudinal MLST-based studies of free-range broiler breeder flocks have demonstrated that *Campylobacter* prevalence and diversity increase with flock age, with *C. coli* forming more stable populations over time than *C. jejuni* (Colles *et al.*, 2011, 2015). Isolates described in the current study were obtained from broiler flocks slaughtered at 38-41 days, which corresponds to the trough in free-range broiler breeder flock diversity observed by Colles *et al*. (2011). Accordingly, the current study may have captured a snapshot of a recently emerged clone as it was spreading through the young flocks. That the two flocks were slaughtered eight weeks apart hints at the genomic stability of *Campylobacter*. This observation

supports the findings of an earlier AFLP-based study that identified a *C. jejuni* strain that appeared to be stable over space and time (Manning *et al.*, 2001).

With respect to the source of infection in this study, vertical transmission from breeder flocks or horizontal on-farm transmission are more plausible than contamination of transport crates or abattoir machinery, given the time between slaughters and the fact that *Campylobacter* is not thought to survive well outside a host. While it is tempting to speculate that UCT020 originated from the same farm as the study isolates, vertical or environmental transmission could have resulted in the infection of flocks from other commercial chicken farms in the region, which also supply Cape Town retailers. The identification of UCT020 does, however, highlight the potential clinical importance of the South African ST-1587 clone, as the patient was severely ill with sepsis.

In conclusion, the current study has demonstrated the presence of the major agriculture associated *C. coli* clonal complex ST-828 in chicken and a human disease sample from South Africa. High-resolution WGS analyses revealed a high degree of genetic homogeneity among the study isolates, similar to the levels of diversity observed after experimental infection of chicks with a single *C. jejuni* strain (Friedrich *et al.*, 2015). However, the current study represents the first example of repeated sampling of a natural population of a recently emerged *C. coli* clone over time, providing novel insights into the evolution of this organism. This work highlights the importance of improving our understanding of the natural diversity and evolution of *Campylobacter* in chickens, and further sampling will help to contextualise the

WGS variation observed among human disease isolates. In turn, this information could be used to calibrate outbreak cut-offs, ultimately improving epidemiological surveillance and control of *Campylobacter*.

CHAPTER 5

The genomic epidemiology of *Campylobacter* from human disease samples from Cape Town, South Africa

ABSTRACT

The epidemiology of *Campylobacter* from LMICs is thought to be distinct from that of high-income countries; however, data from these regions are limited. South Africa, a middle-income country, is an exception, as human campylobacteriosis has been an active area of research since the 1970s, particularly in the city of Cape Town. The aim of this study was to use WGS to characterise the molecular epidemiology of *Campylobacter* isolated in Cape Town in the 1990s and between 2011 and 2012. To this end, WGS data were obtained from *Campylobacter* from patients at 15 public hospitals in Cape Town, including 108 freeze-dried isolates collected in 1991 and 56 isolates collected between 2011 and 2012. Hierarchical gene-by-gene analyses confirmed that the epidemiology of *Campylobacter* from Cape Town is similar to but distinct from that of high-income countries. *C. jejuni* subsp. *jejuni*, *C. jejuni* subsp. *doylei*, *C. upsaliensis*, and *C. coli* were present in both collections, while the 1991 collection also included *C. hyointestinalis* and *C. concisus*. Interestingly, a putative novel *C. jejuni* subsp. *jejuni*/*C. jejuni* subsp. *doylei* hybrid was also identified among the archive isolates. *C. jejuni* subsp. *jejuni* and *C. coli* isolates predominantly belonged to major livestock-associated clonal complexes and did not cluster spatially or temporally when compared to isolates from the UK. Taken together, the findings of this study suggest that the food chain is most likely a long-term source of disease caused by *C. jejuni* subsp. *jejuni* and *C. coli* in South Africa.

5.1 Introduction

Campylobacter remains the leading bacterial cause of gastroenteritis in high-income countries, where the majority of diarrhoeal disease is due to the consumption of contaminated food (Havelaar *et al.*, 2015). Our understanding of the epidemiology of campylobacteriosis in LMICs is incomplete, but it is likely that the burden of disease is underestimated in these regions (Platts-Mills & Kosek, 2014; Kaakoush *et al.*, 2015). As described in Chapter 1 [1.6], South Africa is an exception among LMICs, as local researchers have been working on *Campylobacter* since the 1970s. Human campylobacteriosis was first reported in South Africa in 1977: *C. "fetus"* (later reclassified as *C. jejuni*) was isolated from pericardial fluid and blood samples obtained from two adults and three children, one and two of whom died, respectively (Hallett *et al.*, 1977). In 1978, a second set of case studies was published, in which *C. "fetus"* was isolated from ten malnourished children from geographically distant parts of the country, all of whom had positive blood cultures (Schewitz & Roux, 1978). Given the serious nature of these infections and that diarrhoea was a leading cause of childhood morbidity and mortality in impoverished South African communities (Koornhof *et al.*, 1986), further studies were carried out to better understand the local epidemiology of *Campylobacter*.

The majority of *Campylobacter* research carried out in South Africa before 2000 took place in Johannesburg and Cape Town, the country's largest cities. Johannesburg, the larger of the two, is located in the Gauteng province (Figure 1.11) and is home to the Chris Hani Baragwanath Hospital (CHBH), a 3,200 bed public hospital in the township of Soweto. During apartheid, CHBH provided care to local black communities in which gastroenteritis was a leading cause of child mortality

(Koornhof *et al.*, 1986). Several small case-control studies were carried out in the Paediatric Division of CHBH in the 1970s and 1980s to: determine the aetiology of childhood gastroenteritis; study patterns of shedding of gastrointestinal pathogens; and explore socioeconomic factors driving disease (Bokkenheuser *et al.*, 1979; Robins-Browne *et al.*, 1980; Richardson *et al.*, 1983; Koornhof *et al.*, 1986). Similar studies were carried out in the late 1980s in Cape Town at the Red Cross War Memorial Children's Hospital (RXH), South Africa's dedicated public paediatric hospital that treats children from across the country (Household *et al.*, 1988, 1990). Since 2000, this body of work has been complemented by studies carried out in Limpopo (Figure 1.11), a poor, rural province where 23% of households lack access to safe drinking water (Obi & Bessong, 2002; Samie *et al.*, 2007a, 2007b, 2009b).

Studies carried out in Johannesburg and Cape Town identified *C. jejuni* as a cause of diarrhoea in children below the age of two years, but the organism was also detected among asymptomatic controls (6-19%) (Bokkenheuser *et al.*, 1979; Robins-Browne *et al.*, 1980; Richardson *et al.*, 1983; Koornhof *et al.*, 1986; Household *et al.*, 1988). In the first year of life, *Campylobacter* was significantly more prevalent among children with diarrhoea than controls. In contrast, between the ages of one and two, there was a decrease in *Campylobacter* in children with diarrhoea and a corresponding increase among controls, suggesting the development of protective immunity (Bokkenheuser *et al.*, 1979; Richardson *et al.*, 1983; Koornhof *et al.*, 1986; Household *et al.*, 1988). In Limpopo, *C. jejuni* and *C. coli* were the primary pathogens isolated from HIV-positive patients, hospitalised patients, and school children with diarrhoea (Obi & Bessong, 2002; Samie *et al.*, 2007b, 2007a, 2009b). Although the first reports of human campylobacteriosis in

South Africa were of serious or complicated infections (Hallett *et al.*, 1977; Schewitz & Roux, 1978), mild symptoms and generally good outcomes were reported in studies carried out in all three provinces (Bokkenheuser *et al.*, 1979; Richardson *et al.*, 1983; Househam *et al.*, 1990).

In addition to the studies described above, researchers in Cape Town developed methods for *Campylobacter* culture and typing that revealed features of local human disease epidemiology that appeared to be unique to South Africa. Arguably the most significant contribution made by local researchers was the development of the Cape Town protocol for the improved recovery of non-*jejuni*/*coli* *Campylobacter* species, as well as *C. jejuni* and *C. coli* (Kaakoush *et al.*, 2015). Briefly, the technique entails membrane filtration of samples onto antibiotic-free media, which is incubated in a hydrogen-enriched microaerophilic atmosphere at 37°C (Lastovica, 2006). The Cape Town protocol was introduced into the routine microbiology laboratory at RXH in 1990, resulting in a 14.7% increase in the detection of *Campylobacter* and related organisms (Lastovica & le Roux, 2000). Between 1990 and 2005, human disease isolates from RXH included a large proportion of non-*jejuni*/*coli* *Campylobacter* species. Among the 5,443 *Campylobacter* and related organisms recovered during this period, *C. jejuni* subsp. *jejuni* was most prevalent, accounting for 32.30% of isolates, followed by *C. concisus* (24.63%), *C. upsaliensis* (23.52%), and *C. jejuni* subsp. *doylei* (7.70%), with *C. coli* accounting for just 2.93% (Lastovica, 2006). This contrasts sharply with high-income countries, which typically attribute 90% of human disease to *C. jejuni* and 10% to *C. coli* (Cody *et al.*, 2010).

The atypical predominance of non-*jejuni/coli Campylobacter* species among human disease isolates from Cape Town prompted studies of the molecular epidemiology of these organisms. Typing techniques developed locally or as part of international collaborations included 1-dimensional polyacrylamide gel electrophoresis for *C. jejuni* subsp. *doylei* (Owen *et al.*, 1988), RAPD and PFGE for *C. concisus* (Matsheka *et al.*, 2002, 2006), DNA microarrays for *C. jejuni* (Quiñones *et al.*, 2008), and non-*jejuni/coli Campylobacter* MLST schemes (Miller *et al.*, 2005, 2012), which were used alongside existing methods, such as serotyping, plasmid profiling, MLEE, RFLP, and ribotyping (Lastovica *et al.*, 1986; Owen & Hernandez, 1990b; Da Silva Tatley *et al.*, 1992; Fayos *et al.*, 1993; Lastovica *et al.*, 1997; Wassenaar *et al.*, 2000; Engberg *et al.*, 2001; Nachamkin *et al.*, 2001). The results of these studies suggested that, with the exception of a *C. jejuni* O:41 clone associated with GBS (Lastovica *et al.*, 1997; Wassenaar *et al.*, 2000), South African *Campylobacter* isolates were genetically diverse and often distinct from isolates from other geographic locations. However, many studies were reliant on phenotypic or band-based typing techniques, which, as described in Chapter 1 [1.5], suffer from a lack of objectivity and portability, and have low discriminatory power.

The aim of the current study was use WGS to update our understanding of the molecular epidemiology of *Campylobacter* from Cape Town, focusing on isolates from the public healthcare system. Between 1982 and 1999, *Campylobacter* isolates collected by RXH were freeze-dried and stored, creating a valuable archive held by the University of Cape Town. By exploiting this collection, it was possible to compare *Campylobacter* isolates collected in 1991 to those collected in the city in

2011 and 2012, as well as to UK isolates, to describe the local and global epidemiology of *Campylobacter* from Cape Town.

5.2 Methods

5.2.1 Isolates

5.2.1.1 Archive collection

After the UCT *Campylobacter* archive was catalogued, 4,987 vials of freeze-dried cultures remained. These vials corresponded to 1,382 isolates that were recovered from blood and faecal samples collected from paediatric patients at RXH between 1982 and 1999. The decision was made to focus on isolates from 1991 for two main reasons: (i) the Cape Town protocol was introduced the year before, so the subset was likely to include a wide variety of *Campylobacter* species; and (ii) a large number of intact vials (*n* = 427) and isolates (*n* = 193) was available compared to other years.

At the University of Oxford, an attempt was made to revive isolates for which four or more vials of freeze-dried culture were available (*n* = 43). For each isolate, the biomaterial within one vial was resuspended in 200 µl of BHI broth and incubated at room temperature for 30 to 60 minutes, after which 100 µl was placed onto the centre of a CBA plate (Oxoid Ltd., Basingstoke, UK). Plates were incubated for up to seven days at 37°C in a microaerobic atmosphere generated using the GENbox microaer system (bioMerieux UK Ltd., Basingstoke, UK) and were checked for growth every 48 hours. Presumptive *Campylobacter* colonies were provisionally identified based on positive catalase and oxidase test results, and the detection of Gram-negative small, curved rods, or ring-shaped or coccoid cells, which have been observed in older cultures (Ng *et al.*, 1985).

5.2.1.2 2011-2012 collection

Between October 2011 and October 2012, all *Campylobacter* isolates cultured during the course of routine diagnostic activities by the National Health Laboratory Service Medical Microbiology Laboratory at Groote Schuur Hospital, Cape Town, were to be included in this study. Isolates were recovered using the Cape Town protocol (Lastovica, 2006), and were identified with the VITEK 2. During the study period, 218 *Campylobacter* isolates were recovered from stool samples collected from paediatric and adult patients at 15 healthcare facilities in the Cape Town metropolitan area, including 123 (56.4%) isolates from RXH. The isolates were stored in Microbank cryogenic vials at −80°C, and were subcultured in batches for shipment to the University of Oxford.

A random subset of 63 isolates, determined based on culture viability, was sent to the University of Oxford on sterile swabs in Amies transport medium with charcoal (VWR International Ltd., Lutterworth, UK). Swabs were streak plated onto CBA plates (Oxoid Ltd., Basingstoke, UK) and incubated at 37°C for up to seven days in a microaerobic atmosphere, using the GENbox microaer system (bioMerieux UK Ltd., Basingstoke, UK). Plates were checked for growth every 48 hours, and, in the case of isolates without growth, a second attempt was made to recover *Campylobacter* from the transport swab. Suspensions of viable bacterial cultures were made in BHI broth with 20% glycerol and stored at −80°C.

5.2.2 Generation of whole-genome sequencing data

As the archive collection included non-viable isolates, it was necessary to develop a protocol to extract genomic DNA directly from freeze-dried cultures. A pilot study carried out on 14 additional archive isolates collected during the 1980s demonstrated that the Wizard Genomic DNA Purification Kit (Promega, Southampton, UK) could be adapted for this purpose, producing DNA suitable for WGS. For each isolate, freeze-dried material was resuspended *in situ* in 600 µl of Nuclei Lysis Solution. In the case of isolates with multiple vials, the suspension was transferred from vial to vial until all available biomaterial was resuspended, after which the manufacturer's protocol was resumed. Genomic DNA was extracted from *Campylobacter* isolates from the 2011-2012 collection using the Wizard Genomic DNA Purification Kit without modification.

In preparation for WGS, genomic DNA extracts were quantified using a NanoDrop 1000 Spectrophotometer, and isolates with concentrations of ≤ 10 ng/µl were excluded. The remaining archive isolates were screened for the presence of *Campylobacter* DNA using a broad-range PCR assay for the amplification of *atpA* (Miller *et al.*, 2014b). Isolates from which *atpA* could not be amplified were also excluded from the study. In total, 114 and 56 isolates from the archive and 2011/2012 collections, respectively, were sent to the Wellcome Trust Sanger Institute for WGS. WGS, *de novo* assembly, and upload of the resulting contigs to the *Campylobacter jejuni*/*coli* PubMLST database were carried out as described in Chapter 4 [4.2.2-4.2.3].

5.2.3 Genome annotation and species identification

Loci required for species identification, namely *atpA* and the 52 *rps* genes, were annotated using a combination of automated and manual curation as outlined in Chapter 2 [2.2.2]. Isolates were assigned to species based on *atpA* and rMLST phylogenies, which were generated as described in Chapter 3 [3.2.3.2]. As basis for comparison, 70 reference genomes, encompassing 22/27 known *Campylobacter* species [3.3.1] and representative clade 2 (n = 4) and 3 (n = 5) *C. coli* genomes [2.2.3], were also included in the analyses. Those isolates that could not be assigned to a species based on *atpA* and rMLST sequences were subjected to higher resolution cgMLST analyses. The comparison included the same set of 79 taxonomically diverse reference genomes used for species identification, and was carried out using the TBLASTX-based core genome approach developed in Chapter 3 [3.2.2.1]. A Neighbor-joining tree was generated as previously described [3.2.2.3]. Once isolates were assigned to a species, loci belonging to the appropriate MLST scheme were annotated.

Following annotation, one archive isolate was excluded from further analyses as multiple alleles were detected at 29/52 (55.8%) rMLST loci, suggesting a mixed culture. A further five archive isolates were excluded as they appeared to correspond to poor quality draft genomes: the assemblies comprised 176 to 1,069 contigs and had low N_{50} values, ranging from 2,554 to 35,293 bp. This typically results in a large proportion of incomplete gene sequences, as predicted by the incomplete MLST (n = 5) and rMLST (n = 3) profiles obtained for these isolates.

5.2.4 Genomic analyses

Once segregated by species and, in the case of *C. jejuni*, subspecies, isolates were compared at the MLST level using allele designations and annotated sequences extracted from the database. Additional higher resolution analyses were carried out for *C. jejuni* subsp. *jejuni* and *C. coli* using the BIGSDB GENOME COMPARATOR module implemented in the *Campylobacter jejuni/coli* PubMLST database [2.2.2].

C. jejuni subsp. *jejuni* isolates were stratified by clonal complex, and those with $n > 1$ were compared using wgMLST. Reference-based comparisons were carried out using the complete genome of *C. jejuni* subsp. *jejuni* NCTC 11168 (GenBank accession number AL111168) (Parkhill *et al.*, 2000; Gundogdu *et al.*, 2007), with the exception of analyses carried out for clonal complexes ST-45, ST-52, and ST-353, for which isolates 4031 (HG428754) (Revez *et al.*, 2014a), CG8421 (CP005388) (Poly *et al.*, 2008), and R14 (CP005081), respectively, were used as reference genomes. Related genotypes from the UK were also included in these analyses: for each clonal complex, a representative of every unique rST present among the 1,713 *Campylobacter* genomes from the Oxfordshire sentinel surveillance collection described in Chapter 2 [2.3.1] was included in the analysis (Appendix G). Similarly, wgMLST analysis of *C. coli* isolates included a representative of each unique rST from ST-828 clonal complex members of the Oxfordshire sentinel surveillance collection (Appendix G), as well as representative clade 2 ($n = 4$) and 3 ($n = 5$) isolates described in Chapter 2 [2.2.3]. This comparison was carried out using *C. coli* isolate 15-537360 (CP006702) (Pearson *et al.*, 2013) as a reference genome.

Default parameters were used for all GENOME COMPARATOR analyses, corresponding to a minimum sequence identity of 70%, minimum alignment of 50%, and a word size of 20. GENOME COMPARATOR options were selected such that putative paralogs identified in all isolates, as well as incomplete loci identified in one or more isolates, were excluded from the distance matrix calculations. The resulting distance matrices were visualised as networks using the NeighborNet algorithm (Bryant & Moulton, 2004) implemented in SplitsTree4 version 4.13.1 (Huson & Bryant, 2006).

5.3 Results

5.3.1 Recovery of genomic DNA from freeze-dried *Campylobacter*

An attempt was made to revive archive isolates for which excess freeze-dried material was available; however, none of the 43/193 (22.3%) isolates tested was viable. Consequently, it was necessary to extract genomic DNA directly from the freeze-dried cultures. The Promega Wizard Genomic DNA Purification Kit was adapted for this purpose in a pilot study of 14 additional archive isolates from 1982 (*n* = 3), 1983 (*n* = 6), and 1989 (*n* = 5) (Appendix H). Genomic DNA was then extracted from 427 vials, which corresponded to 193 isolates with between one and six vials of freeze-dried culture each. Based on a threshold concentration of 10 ng/µl, sufficient DNA for WGS was obtained from 138 (71.5%) isolates (Figure 5.1). DNA concentration was generally proportional to the number of vials available per isolate, although this was not always a reliable predictor of final yield. Importantly, sufficient DNA was obtained from the majority of isolates with as few as one (50.8%) or two (72.9%) vials of freeze-dried material (Figure 5.1).

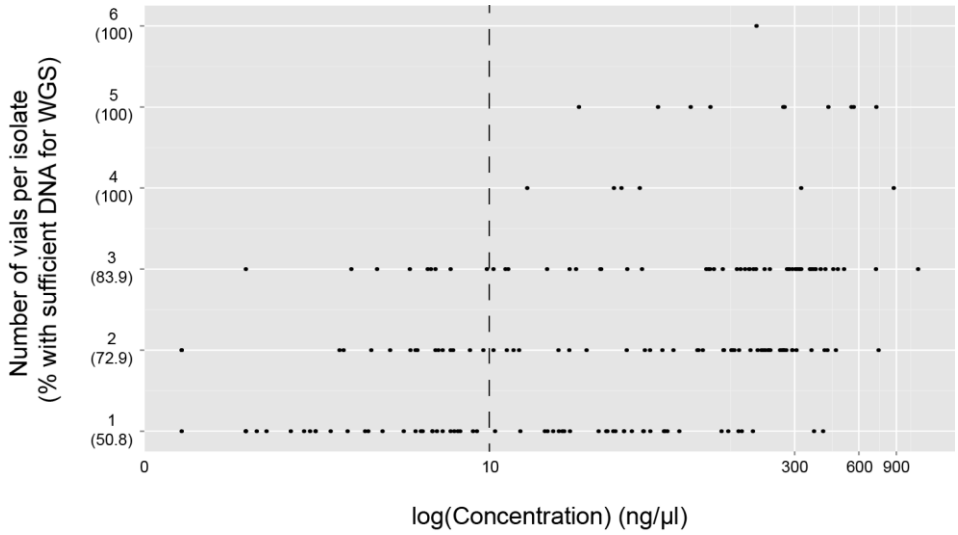


Figure 5.1 Recovery of genomic DNA from freeze-dried *Campylobacter* isolates. Each dot represents an isolate, with the number of vials of freeze-dried material available per isolate shown on the *y*-axis. The dashed line indicates the threshold concentration for whole-genome sequencing (WGS).

5.3.2 Whole-genome sequencing data

As the archive isolates had been subjected to freeze-drying and were stored at room temperature for over two decades, an additional screening step was carried out to confirm that the extracted material contained intact *Campylobacter* DNA. Isolates were screened using the *atpA* PCR assay developed by Miller and colleagues (Miller *et al.*, 2014b), in which a single degenerate primer pair is used to amplify all currently known taxa within the *Campylobacter* genus. In total, 114/138 (82.6%) archive isolates with DNA concentrations above the 10 ng/µl threshold were *atpA*-positive, all of which were sent for WGS. Based on assembly statistics, 108 (94.7%) of the resulting genomes met the requirements for inclusion in further analyses (Appendix I). Complete nucleotide sequences for all typing loci of interest, that is, *atpA*, in addition to the seven MLST and 52 rMLST loci, were recovered from 104 (96.3%) of the included genomes, with at most three loci absent from a single isolate (Appendix I).

Among the 2011/2012 collection, 56/63 (88.9%) *Campylobacter* isolates received by the University of Oxford were viable, all of which were sent for WGS (Appendix I). With respect to assembly statistics, all genomes met the requirements for inclusion in further analyses. As a further indication of the high quality of these genomes, complete *atpA*, MLST, and rMLST nucleotide sequences were obtained from all isolates.

5.3.3 Species identification

5.3.3.1 In silico *species identification*

Analyses based on *atpA* and rMLST nucleotide sequences indicated that all study isolates corresponded to members of the *Campylobacter* genus. The *atpA* and rMLST species assignments were in agreement, with the exception of one isolate each from the 1991 and 2011/2012 collections. Based on the *atpA* phylogeny, isolate UCT031 from the 2011/2012 collection clustered with *C. coli* isolates (Figure 5.2), whereas it was assigned to *C. jejuni* subsp. *jejuni* based on analysis of rMLST nucleotide sequences. This particular *atpA* sequence, designated allele 13, was identified in an additional 94 publicly available genomes in the *Campylobacter jejuni/coli* PubMLST database. Ninety-two (97.9%) of these isolates were assigned to the *C. jejuni* subsp. *jejuni* ST-61 clonal complex, as was UCT031. Further investigations revealed that *atpA* allele 13 was predominant among members of the ST-61 clonal complex, occurring in 92/99 (92.9%) genomes from the lineage that were present in the PubMLST database.

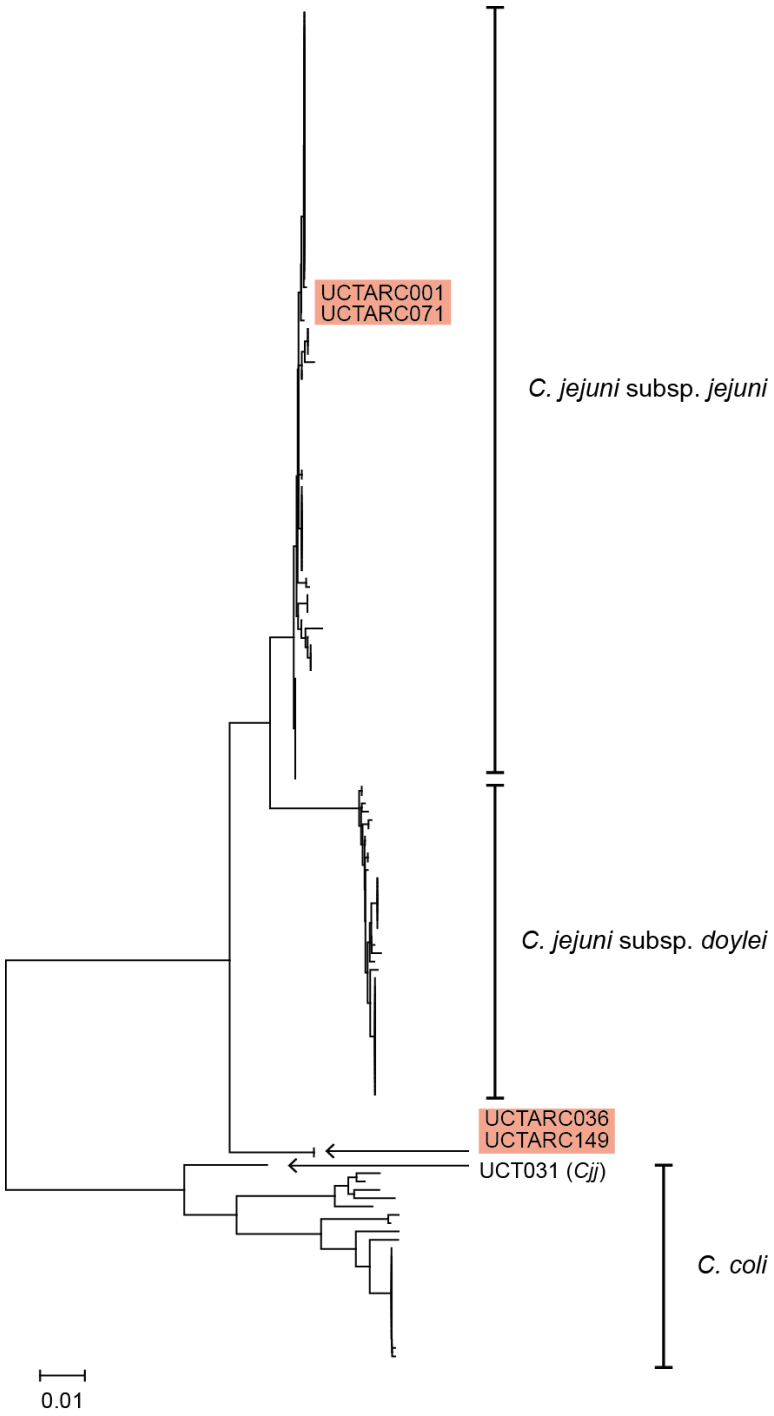


Figure 5.2 Neighbor-joining phylogeny based on *atpA* sequences of 159 *Campylobacter* isolates. Included in the phylogeny are *C. jejuni* subsp. *jejuni* (*Cjj*), *C. jejuni* subsp. *doylei*, and *C. coli* isolates from a global collection of finished genomes (*n* = 42) and draft assemblies from a South African archive (*n* = 117). South African archive isolates identified as belonging to a novel *C. jejuni* clade based on rMLST are highlighted orange. Also shown is isolate UCT031, which clustered with *C. coli atpA* sequences, but was identified as *C. jejuni* subsp. *jejuni* based on phylogenetic analysis of 52 concatenated rMLST loci.

In addition to UCT031, conflicting *atpA* and rMLST species assignments were obtained for isolate UCTARC071 from the 1991 archive collection. UCTARC071 clustered with *C. jejuni* subsp. *jejuni atpA* sequences (Figure 5.2); however, at the rMLST level it formed part of a cluster of isolates related to, yet distinct from, the known *C. jejuni* subspecies (Figure 5.3). This cluster contained three additional isolates, all of which were from the archive collection: UCTARC036 and UCTARC149, which were isolated in 1991 (Appendix I), and pilot study isolate UCTARC001, which was recovered in 1982 (Appendix H). Inspection of the *atpA* phylogeny revealed that isolates UCTARC036 and UCTARC149 also formed a distinct subgroup within the *C. jejuni* clade, while UCTARC001 clustered with UCTARC071 among the *C. jejuni* subsp. *jejuni* isolates (Figure 5.2). Although the GC content of these assemblies did not differ from those of isolates assigned to the known *C. jejuni* subspecies, they were notably larger by approximately 200,000 bp (Table 5.1). The archive isolates of interest were all obtained from males aged five years or younger, but there were no obvious epidemiological links among the three recovered in 1991, which were received in January, March, and August.

Table 5.1 Species-specific properties of *Campylobacter* genomes from the 1991 archive collection

Species	Average % GC content	Number of bases in assembly (Mb)	
		Average	Range
C. jejuni subsp. *jejuni*	30.4	1.68	1.60-1.81
C. jejuni subsp. *doylei*	30.5	1.71	1.60-1.80
Unidentified *C. jejuni* isolates[a]	30.4	1.90	1.84-1.96
C. coli	31.2	1.78	1.65-1.92
C. concisus	37.7	1.81	1.81-1.81
C. hyointestinalis	34.1	1.78	1.74-1.80
C. upsaliensis	34.4	1.73	1.56-1.87

[a] Including UCTARC001 from 1982.



Figure 5.3 Neighbor-joining phylogeny based on concatenated nucleotide sequences of 52 ribosomal protein genes from 159 *Campylobacter* isolates. (A) Relationships among *C. jejuni* subsp. *jejuni*, *C. jejuni* subsp. *doylei*, and *C. coli* isolates corresponding to finished genomes from a global reference collection (*n* = 42) and draft assemblies from the South African *Campylobacter* archive (*n* = 117). South African isolates belonging to a previously unidentified *C. jejuni* lineage are highlighted in orange. (B) A detailed view of the putative novel *C. jejuni* subspecies cluster.

Additional genomic analyses were carried out to resolve the species assignments of UCTARC001, UCTARC036, UCTARC071, and UCTARC149. Using the GENOME COMPARATOR approach for core genome analysis developed in Chapter 3 [3.2.2], 774 core genes were identified among the study isolates and a taxonomically diverse reference set of 79 genomes that accounted for 22/27 known *Campylobacter* species, including representatives of all three *C. coli* clades. The four study isolates again formed a distinct *C. jejuni* clade, which, at the higher level of resolution, was intermediate between *C. jejuni* subsp. *jejuni* and *C. jejuni* subsp. *doylei* (Figure 5.4).

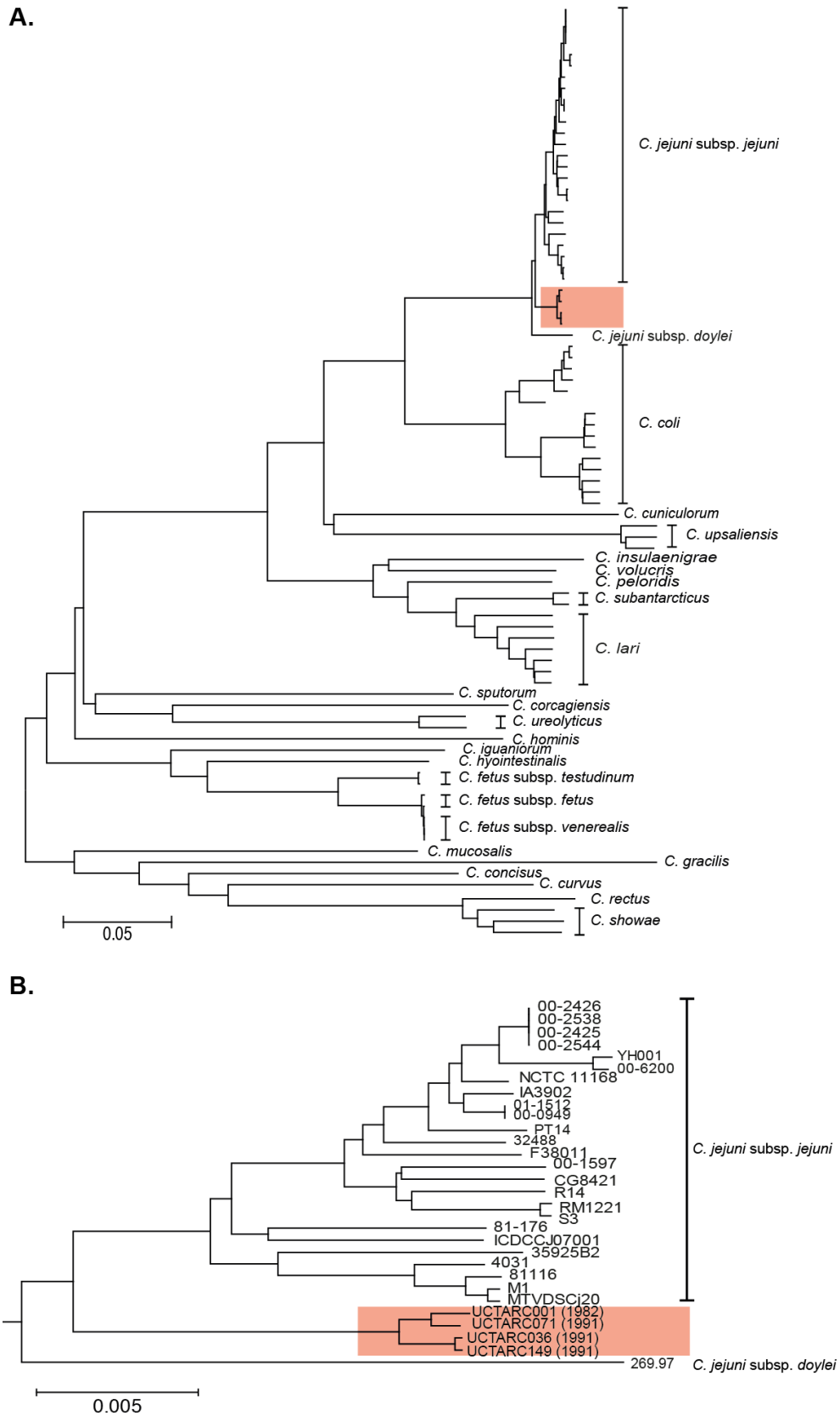


Figure 5.4 Relationships among 83 *Campylobacter* isolates based on core genome diversity. (A) The Neighbor-joining phylogeny reconstructed based on 399,138 unambiguously aligned nucleotides from 774 core genes that were identified among 79 taxonomically diverse *Campylobacter* and four South African isolates that could not be assigned to a species based on *atpA* or rMLST sequences (highlighted orange). (B) A detailed view of the relationships among *C. jejuni* subsp. *jejuni*, *C. jejuni* subsp. *doylei*, and South African isolates from the putative hybrid *C. jejuni* clade (orange).

5.3.3.2 Comparison of laboratory-based and in silico *species assignments*

In silico species assignments for isolates belonging to the 1991 and 2011/2012 collections were compared to those obtained using conventional phenotypic and biochemical tests, and the VITEK 2, respectively. Original species designations were available for 94/108 (87%) isolates from the 1991 archive isolates, only six (6.4%) of which were not in agreement with the *in silico* species assignments, including the three *C. jejuni* subsp. *jejuni*/*C. jejuni* subsp. *doylei* hybrid isolates (Table 5.2). With respect to VITEK 2 species assignments for *Campylobacter* from the 2011/2012 collection, 20/56 (35.7%) isolates were only identified to the genus level, accounting for *C. coli*, *C. jejuni* subsp. *jejuni*, and *C. upsaliensis*. Only three (8.3%) discrepancies were identified among the 36 isolates that were identified to the species level using the VITEK 2 (Table 5.2).

Table 5.2 Isolates from the 1991 and 2011/2012 collections with conflicting laboratory-based and *in silico* species assignments

Collection	PubMLST id	Isolate	Species assignments	
			Laboratory[a]	*In silico*
1991	32970	UCTARC031	*C. jejuni* subsp. *doylei*	*C. jejuni* subsp. *jejuni*
	32972	UCTARC036	*C. jejuni* subsp. *jejuni*	*Cjj/Cjd* hybrid[b]
	33003	UCTARC071	*C. jejuni* subsp. *jejuni*	*Cjj/Cjd* hybrid
	33008	UCTARC092	*C. coli*	*C. jejuni* subsp. *doylei*
	32980	UCTARC094	*C. jejuni* subsp. *jejuni*	*C. upsaliensis*
	33041	UCTARC149	*C. jejuni* subsp. *jejuni*	*Cjj/Cjd* hybrid
2011/2012	25284	UCT064	*C. jejuni*	*C. coli*
	25245	UCT036	*C. coli*	*C. jejuni* subsp. *jejuni*
	25282	UCT062	*C. jejuni*	*C. upsaliensis*

[a] For isolates from the 1991 and 2011/2012 collections, laboratory-based species assignments were based on conventional phenotypic and biochemical tests, and VITEK 2 results, respectively.

[b] Intermediate *C. jejuni* subsp. *jejuni* (*Cjj*) and *C. jejuni* subsp. *doylei* (*Cjd*) clade identified based on rMLST and cgMLST analyses.

5.3.3.3 Distribution of species

The majority of isolates in the 1991 and 2011/2012 collections corresponded to *C. jejuni* subsp. *jejuni*, *C. jejuni* subsp. *doylei*, *C. coli*, and *C. upsaliensis* (Figure 5.5). However, the relative abundances of these species differed, particularly in the case of *C. jejuni* subsp. *jejuni*, which accounted for 20.3 and 67.9% of isolates from the 1991 and 2011/2012 collections, respectively. Similarly, the 1991 collection included *C. concisus* and *C. hyointestinalis*, while the 2011/2012 collection did not (Figure 5.5). The distribution of species observed in this study was similar to that of *Campylobacter* recovered from stool samples obtained from paediatric patients at RXH between 1990 and 2000 (Lastovica & le Roux, 2000) (Figure 5.5). Noteworthy exceptions included differences in the proportions of *C. concisus*, *C. jejuni*, and *C. upsaliensis* isolates, and the absence of *C. curvus*, *C. fetus*, *C. lari*, *C. rectus*, and *C. sputorum* in the current study.

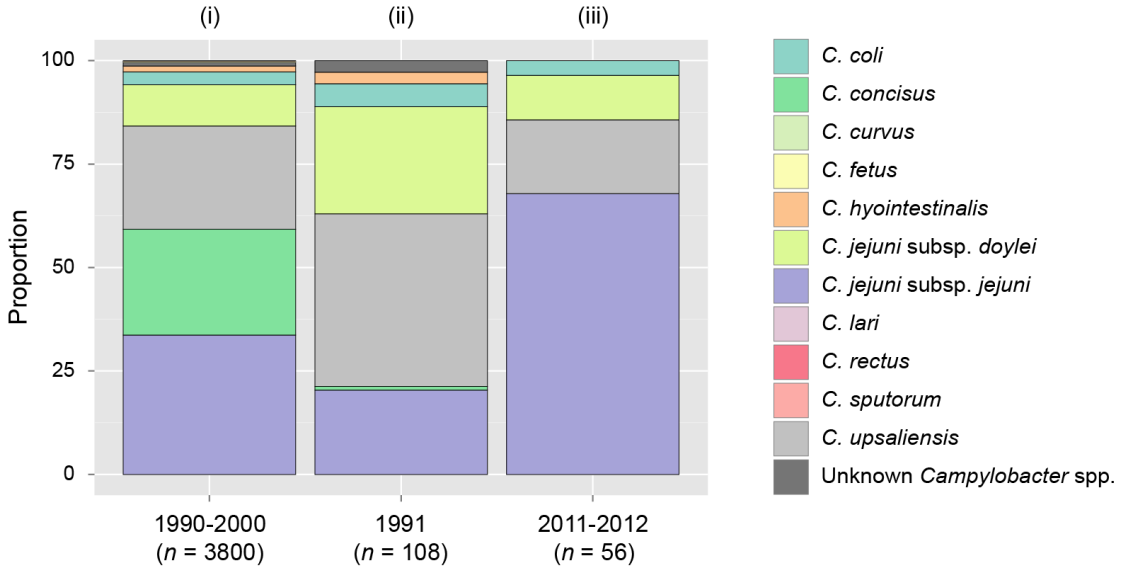


Figure 5.5 Distribution of *Campylobacter* species among collections of human disease isolates from Cape Town, South Africa. Collections are shown along the *x*-axis and correspond to isolates recovered from: (i) stool samples from children at RXH between 1990 and 2000 (Lastovica & le Roux, 2000); (ii) stool and blood samples from children at RXH in 1991 and included in this study; and (iii) stool samples from adults and children from 15 Cape Town healthcare facilities between October 2011 and October 2012, also included in the current study.

5.3.4 MLST

5.3.4.1 C. jejuni *and* C. coli

Sixty-two STs were identified among the 105 *C. jejuni* subsp. *jejuni, C. jejuni* subsp. *doylei,* hybrid *C. jejuni* subsp. *jejuni*/*C. jejuni* subsp. *doylei*, and *C. coli* isolates included in this study, corresponding to 32 (51.6%) novel allelic profiles (Table 5.3). Overall, 29/62 (46.8%) STs were assigned to 15 clonal complexes, known to correspond to globally relevant *C. jejuni* subsp. *jejuni* and *C. coli* lineages (Cody *et al.*, 2013) (Figure 5.6). STs assigned to these clonal complexes accounted for 54/60 (90%) *C. jejuni* subsp. *jejuni* and 6/8 (75%) *C. coli* isolates; with the exception of seven novel allelic profiles, all corresponded to STs present in the *Campylobacter jejuni*/*coli* PubMLST database. Comparison of the rank abundance distribution of clonal complexes across the 1991 and 2011/2012 collections indicated that, although there was a degree of overlap, the proportions of isolates assigned to lineages present in these datasets differed, particularly with respect to ST-206 and ST-353 (Figure 5.6). Additionally, the 2011/2012 collection included seven clonal complexes not represented in the 1991 collection, namely ST-179, ST-22, ST-283, ST-443, ST-48, ST-49, and ST-61 (Figure 5.6). Of the 33 STs unassigned to a clonal complex, the majority (75.8%) corresponded to *C. jejuni* subsp. *doylei* isolates, for which clonal complexes have not yet been defined (Figure 5.6). *C. jejuni* subsp. *doylei* isolates appeared to be genetically diverse with 25 STs identified among 34 isolates (Table 5.3). Also unassigned to a clonal complex were the three *C. jejuni* subsp. *jejuni*/*C. jejuni* subsp. *doylei* hybrid isolates from the 1991 collection (Table 5.3).

Table 5.3 Identification of existing and novel sequence types (STs) among *C. jejuni* and *C. coli* isolates from the 1991 and 2011/2012 *Campylobacter* collections

Species	Isolate collection (*n*)				Overlapping STs (*n*)
	1991		2011/2012		
	Isolates	STs[a]/ novel STs	Isolates	STs/ novel STs	
C. jejuni subsp. *jejuni*	22	13/1	38	21/5	5
C. jejuni subsp. *doylei*	28	21/19	6	6/6	2
C. jejuni subsp. *jejuni*/ *C. jejuni* subsp. *doylei* hybrid	3	2/1	0	0/0	0
C. coli	6	4/3	2	2/0	0

[a] ST, multilocus sequence typing-defined sequence type.

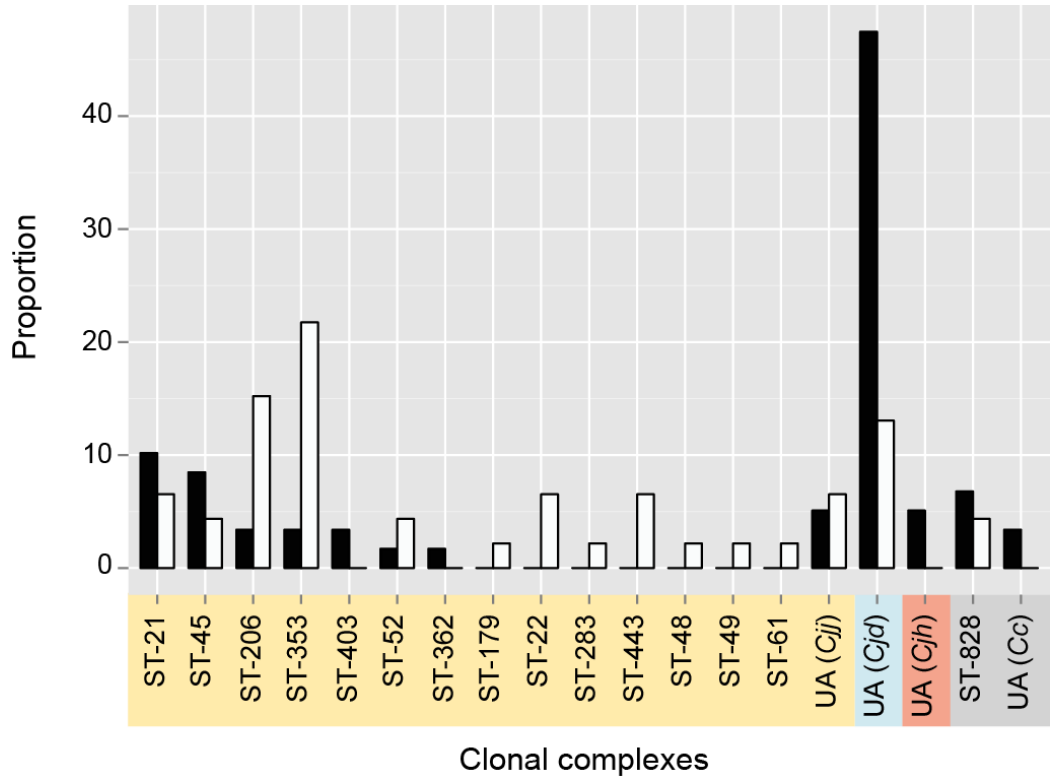


Figure 5.6 Frequency distribution of clonal complexes among *C. jejuni* and *C. coli* from the 1991 (black) and 2011/2012 (white) human disease isolate collections from Cape Town, South Africa. Clonal complexes are coloured by species for ease of reading: *C. jejuni* subsp. *jejuni*, yellow; *C. jejuni* subsp. *doylei*, turquoise; *C. jejuni* subsp. *jejuni*/*C. jejuni* subsp. *doylei* hybrid, orange; *C. coli*, grey. UA, unassigned; *Cjj*, *C. jejuni* subsp. *jejuni*; *Cjd*, *C. jejuni* subsp. *doylei*; *Cjh*, *C. jejuni* subsp. *jejuni*/*C. jejuni* subsp. *doylei* hybrid; *Cc*, *C. coli*.

5.3.4.2 Non-jejuni/coli species

C. upsaliensis was the predominant non-*jejuni*/*coli* *Campylobacter* species identified in this study (Figure 5.5). At the MLST level, *C. upsaliensis* isolates were highly diverse, with 47 STs identified among 55 isolates, including 43 (91.5%) novel allelic profiles. When segregated by dataset, 45 and 10 isolates from the 1991 and 2011/2012 collections were distributed across 39 and 8 STs, respectively, without overlap. Of these, 12/39 (30.8%) and 4/8 (50%) STs belonged to the ST-42 clonal complex, accounting for 21 (38.2%) of all *C. upsaliensis* isolates. The remaining 31 STs were not assigned to clonal complexes. Concerning the three *C. hyointestinalis* isolates, each corresponded to a unique ST, two of which were novel, and none was assigned to a clonal complex. Similarly, the single *C. concisus* isolate corresponded to a novel ST that was not assigned to a clonal complex.

5.3.5 Whole-genome MLST comparisons of *C. jejuni* subsp. *jejuni* and *C. coli*

Whole-genome MLST analyses were carried out to characterise the diversity of the *C. jejuni* subsp. *jejuni* and *C. coli* isolates at the whole-genome level. To explore the possibility that geographically restricted lineages have emerged in South Africa, the study isolates were also compared to UK *Campylobacter* representative of the same MLST clonal complexes. The results of the wgMLST analyses indicated that, although isolates from Cape Town were genetically diverse and often separated by long distances, they did not cluster according to space or time when compared to matched isolates from the UK (Figure 5.7; Figure 5.8).

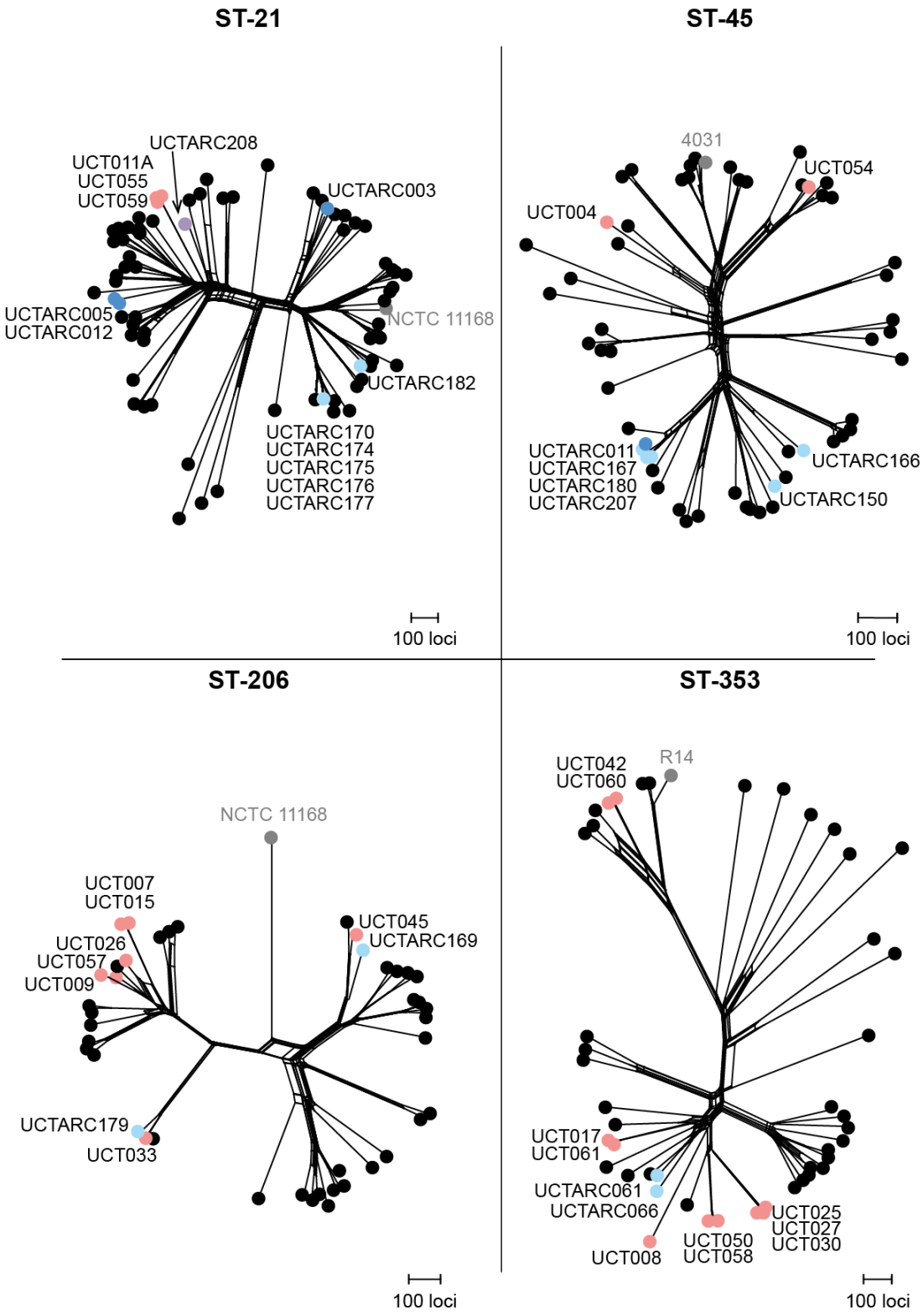


Figure 5.7 Whole-genome MLST comparisons of South African and UK *C. jejuni* isolates according to clonal complex. Coloured circles mark isolates from the 1991 archive collection (turquoise), the 2011-2012 collection (pink), and pilot study isolates from 1982 (purple), 1983 (dark blue), and 1989 (green). All UK human disease isolates are marked with a black circle. Reference genomes are shown in grey. Only clonal complexes with more than one South African isolate were included.

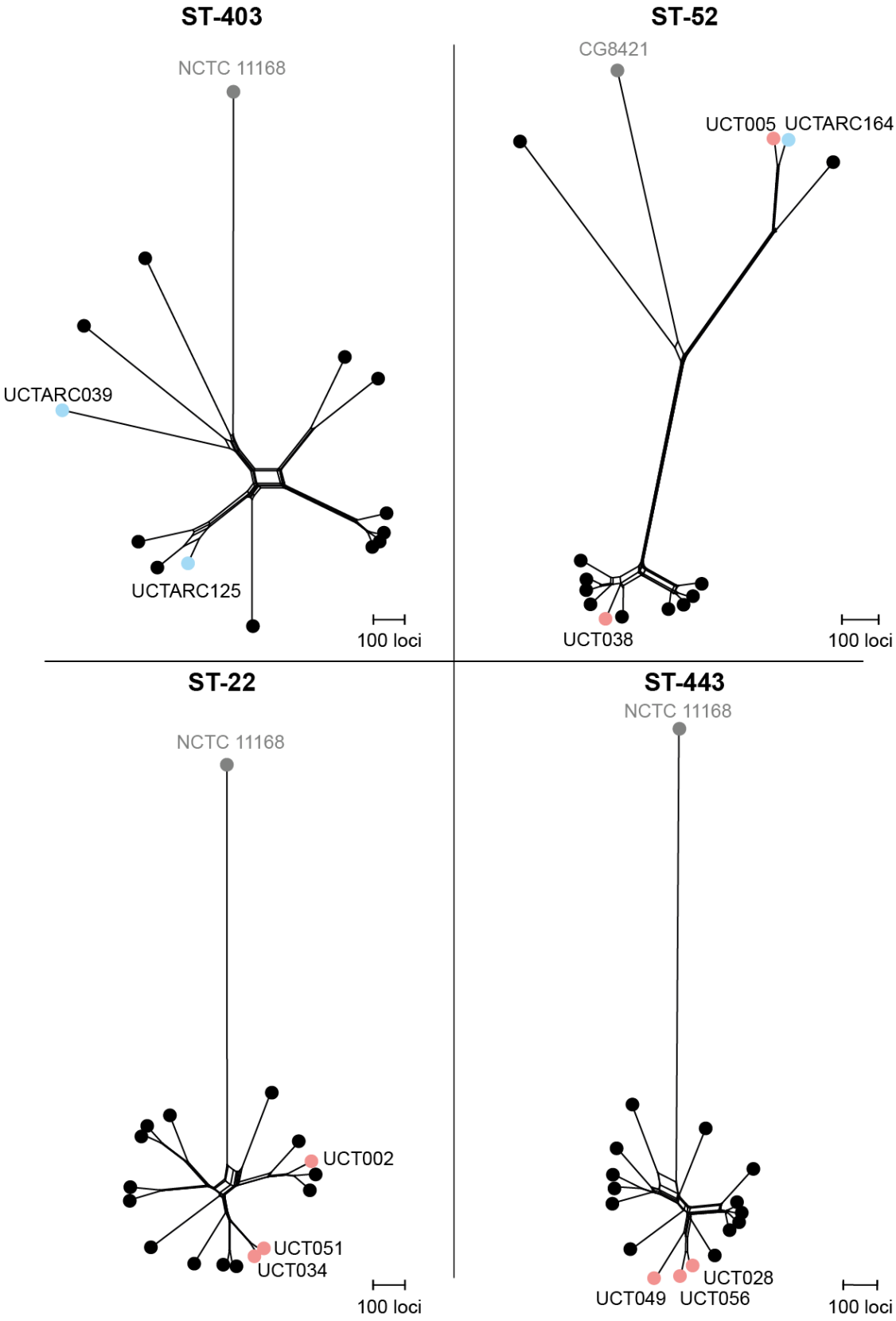


Figure 5.7 cont.

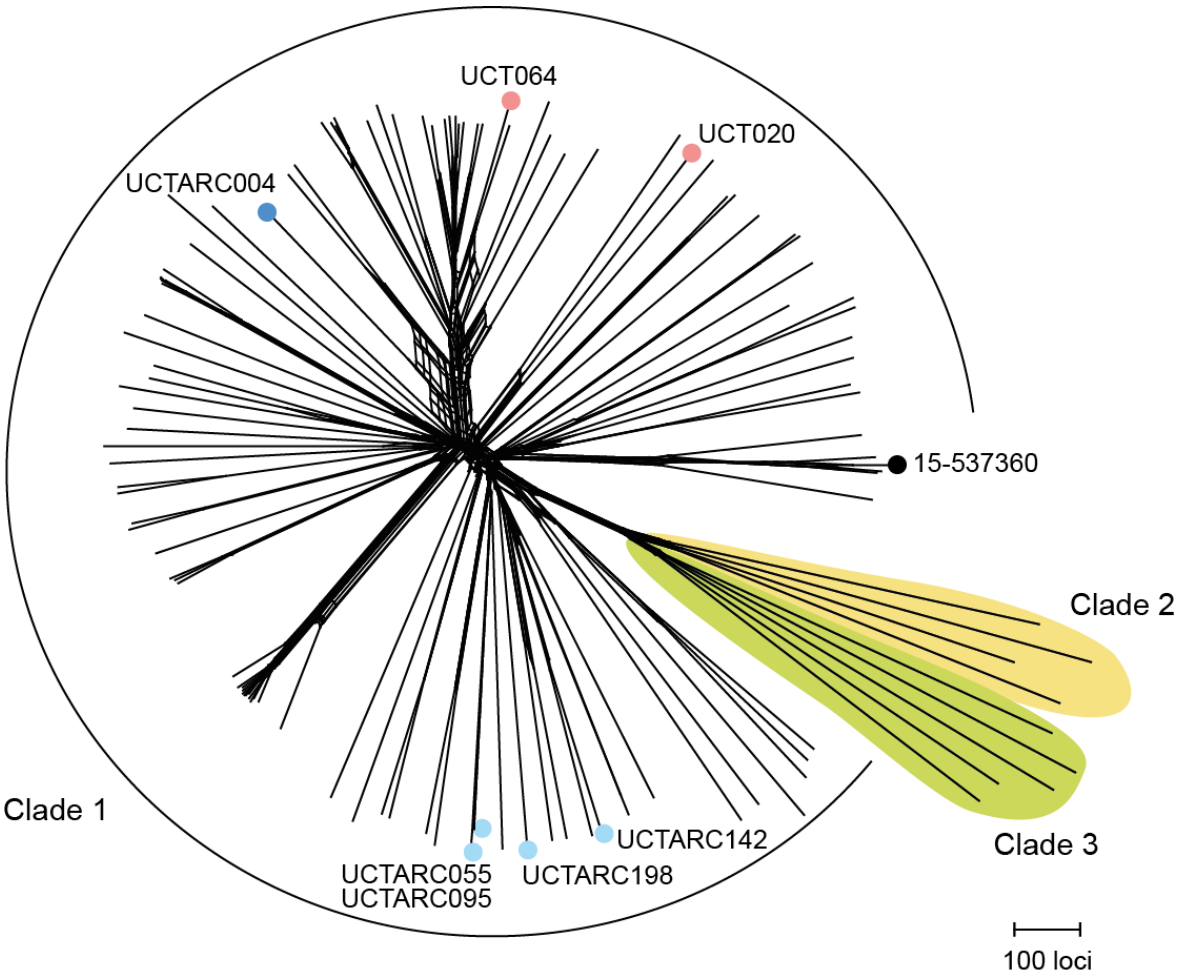


Figure 5.8 Whole-genome MLST comparisons of South African and UK *C. coli* isolates. Coloured circles mark isolates from the 1991 archive collection (turquoise), the 2011-2012 collection (pink), and pilot study isolates from 1982 (purple), 1983 (dark blue), and 1989 (green). The reference genome, *C. coli* 15-537360, is marked with a black circle. All unmarked clade 1 isolates correspond to UK *C. coli* from human disease samples, while the remaining unmarked isolates correspond to the clades 2 and 3 reference set isolates.

5.4 Discussion

Previous studies showed that *C. jejuni* from Curaçao, a Caribbean island, were genetically distinct from those from the UK, Canada, and Australia, suggesting that isolates from LMICs or newly industrialised regions may be distinct from those from high-income countries (Duim *et al.*, 2003; Dingle *et al.*, 2008). As outlined in Chapter 1 [1.6], South Africa represents an unusual context for *Campylobacter* research. Although classified as an upper middle-income country (The World Bank, 2015), wealth remains unevenly distributed as a result of policies implemented during apartheid and colonial rule. The healthcare system is stratified into private and public sectors, with those attending the latter typically corresponding to lower socioeconomic groups (Mayosi *et al.*, 2012; Mayosi & Benatar, 2014). Historical surveillance data from Cape Town public hospitals suggest that the local epidemiology of human campylobacteriosis is distinct from that of high-income countries, with larger proportions of non-*jejuni*/*coli* species detected among disease samples (Lastovica & le Roux, 2000; Lastovica, 2006). However, sequence-based data from local *Campylobacter* isolates are limited. Here, 164 human disease isolates from public hospitals in Cape Town were genome sequenced and analysed using the hierarchical gene-by-gene approach and BIGSDB to describe the local molecular epidemiology of *Campylobacter* between 1991 and 2011/2012, and also to place them in the global context.

This is the first recorded attempt to obtain WGS data from non-viable freeze-dried *Campylobacter* isolates. Based on genomic DNA concentrations alone, sufficient material was obtained from 138/193 (71.5%) isolates from the 1991 collection, including isolates with just one (50.8%) or two (72.9%) vials of biomaterial. In a

further screening step, *Campylobacter* were chosen for sequencing based on the detection of *atpA*, with high-quality WGS data generated for 108/114 (94.7%) isolates. Of these isolates, 96.3% had complete *atpA*, MLST and rMLST data. Similarly, high-quality WGS data were obtained for the 56 viable isolates from the 2011/2012 collection that were sent for sequencing. Although both isolate collections were incomplete, they provided valuable insights into the epidemiology of *Campylobacter* from Cape Town.

The results of the *atpA* and rMLST *in silico* speciation analyses confirmed that *C. jejuni* subsp. *jejuni*, *C. upsaliensis*, *C. jejuni* subsp. *doylei*, and *C. coli* accounted for the majority of isolates in both datasets, while *C. hyointestinalis* and *C. concisus* were also detected in the archive collection. The *in silico* results indicated the accuracy of microbiological testing in 1991: 93.6% of original species designations were in agreement with those based on WGS data. In contrast, the VITEK 2, used during the 2011-2012 study period, failed to identify the majority (64.3%) of isolates beyond the genus level, although 91.7% of those that were identified to the species level were in agreement with the *in silico* results. These results suggest that the VITEK 2 is not a reliable tool for identification of *Campylobacter*, which is in agreement with the results of an earlier evaluation (Martiny *et al.*, 2011). Importantly, the results of the current study also revealed that most members of the *C. jejuni* ST-61 clonal complex carry a single *C. coli*-specific *atpA* allele, as previously noted in an MLST-based study of mosaic alleles in *C. jejuni* and *C. coli* (Sheppard *et al.*, 2011). It is likely that an introgression event occurred in an ancestral ST-61 strain, which then underwent clonal expansion. This highlights the importance of using typing schemes with multiple loci.

For both collections, the distribution of species diverged from that observed at RXH between 1990 and 2000: the proportion of *C. concisus* was far greater over the ten year period, while both study collections had a smaller number of species than detected overall (Lastovica & le Roux, 2000). It is possible that these discrepancies reflect the incomplete nature of the study datasets. Alternatively, with reference to the 1991 collection, the Cape Town protocol had only been implemented the year before (Lastovica & le Roux, 2000), and it is possible that the isolation and identification of *C. concisus* improved over time. Additionally, it is plausible that the differences observed between 2011 and 2012 were at least partly due to the inclusion of samples from adults and children, rather than just children as in the RXH collection. There may also have been changes in the indications for stool culture requests that have affected the observed distribution. Although non-*jejuni*/*coli* *Campylobacter* species were commonly isolated from stool samples in Cape Town between 1990 and 2012, the pathogenic potential and reservoirs of these organisms remains unclear.

During species identification, a subset of archive isolates was found to be phylogenetically distinct from known *Campylobacter* species. Based on cgMLST analysis of 774 genes, the three isolates from the 1991 collection and one from 1982 corresponded to a separate cluster that fell between *C. jejuni* subsp. *jejuni* and *C. jejuni* subsp. *doylei.* A literature search identified another isolate with a matching ST that was included in a genomic characterisation of *C. jejuni* from South Africa (Quiñones *et al.*, 2008). Although not discussed in detail in that study, the isolate was also noticeably distinct from all others included in the analysis. Unfortunately, the hybrid isolates from the current study were not viable, so they could not be

characterised further. Although beyond the scope of this project, collections of publicly available WGS data could be screened to find related isolates to subject to phenotypic and biochemical testing.

The results of the MLST analyses suggested that *C. jejuni* subsp. *jejuni*, *C. jejuni* subsp. *doylei*, *C. coli*, and *C. upsaliensis* were genetically diverse, based on the high ratio of STs to isolates. No clonal complexes have been defined for *C. jejuni* subsp. *doylei*, but *C. upsaliensis* isolates included members of the ST-42 clonal complex, which has been identified previously in South Africa (Miller *et al.*, 2005). With respect to *C. jejuni* subsp. *jejuni* and *C. coli*, comparison with the *Campylobacter jejuni/coli* PubMLST database indicated the presence of major agriculture-associated clonal complexes (Colles & Maiden, 2012). These lineages have also been reported in Senegal (Kinana *et al.*, 2006) and Malawi (http://pubmlst.org/campylobacter/). Although it is difficult to draw comparisons between the Cape Town datasets as they are incomplete, it appears that the makeup of *C. jejuni* and *C. coli* lineages have changed over time in Cape Town, with a far greater number of clonal complexes detected in 2011 and 2012. Additional high-resolution wgMLST analyses indicated that the Cape Town isolates did not cluster spatially or temporally when compared to UK human disease isolates, providing support for the hypothesis that agriculture and food distribution drive the global dissemination of particular genotypes (McCarthy *et al.*, 2007; Sheppard *et al.*, 2010b).

While the WGS datasets presented here were not complete, the results of this study demonstrate the value of genomic data for rapid epidemiological analyses. The findings confirm that the epidemiology of *Campylobacter* from Cape Town shares certain features with high-income countries, in particular the presence of major agriculture-associated human disease clonal complexes. However, the detection of a large proportion of non-*jejuni*/*coli* *Campylobacter* species in both isolate collections, as well as the putative novel *C. jejuni* subsp. *jejuni*/*C. jejuni* subsp. *doylei* hybrid, indicate that the local epidemiology is also distinct. Finally, this study underscores the potential of isolate archives, with the Cape Town collection containing a further 1,189 *Campylobacter* and related organisms. The analyses described here are merely a starting point for how these data could be used to better understand the epidemiology, population biology, and evolution of *Campylobacter*.

FINAL DISCUSSION

Infectious intestinal diseases remain a major global public health problem (World Health Organisation, 2013). In high-income countries, gastroenteritis is typically caused by the consumption of contaminated food products (Havelaar *et al.*, 2015), and its impact is mainly economic (Abelson *et al.*, 2006; Thomas *et al.*, 2008; Friesema *et al.*, 2012; Scharff, 2012; Food Standards Agency, 2013). In contrast, diarrhoeal disease remains a leading cause of childhood morbidity and mortality in LMICs, and has lasting negative effects on health in these regions (World Health Organisation, 2013; Walker *et al.*, 2013a). *Campylobacter* is considered a leading bacterial cause of gastroenteritis worldwide; however, our understanding of the epidemiology of human campylobacteriosis in LMICs is incomplete (Kaakoush *et al.*, 2015). It is generally accepted that the epidemiology of *Campylobacter* in LMICs differs to that observed in high-income countries (Platts-Mills & Kosek, 2014), with some evidence of genetic differences among disease-causing genotypes from these regions (Duim *et al.*, 2003; Dingle *et al.*, 2008). The primary aim of this thesis was to use a high-resolution WGS approach to improve our understanding of the molecular epidemiology of *Campylobacter* from South Africa.

WGS is currently driving advances in the biomedical sciences and is considered particularly promising for use in clinical microbiology laboratories (Didelot *et al.*, 2012; Köser *et al.*, 2012b, 2014; Fournier *et al.*, 2014). The first half of this thesis demonstrated how WGS data and the gene-by-gene approach implemented in BIGSDB (Maiden *et al.*, 2013) can be used to address general problems in diagnostic microbiology. In Chapter 2, the commonly-used *mapA*/*ceuE* RT-PCR for the

detection of *C. jejuni* and *C. coli* (Best *et al.*, 2003) was evaluated, confirming that the assay is robust with a 99.7% accuracy rate. This is an important finding as the *mapA*/*ceuE* assay, and variations thereof, is widely used. Moving forward, RT-PCR has an important role to play in diagnostic microbiology, particularly in settings where culture is not feasible; therefore, it is important to develop novel cost-effective, rapid methods for rigorous evaluation of molecular diagnostic assays. To this end, the approach presented here could be developed further to create a plug-in for integration into BIGSDB, allowing others to carry out similar analyses using the large WGS datasets that are now available in PubMLST databases.

Over forty years have passed since methods were first developed for the culture of *Campylobacter* (Dekeyser *et al.*, 1972; Butzler *et al.*, 1973), yet these organisms remain difficult to isolate and identify (Fitzgerald, 2015). With this in mind, Chapter 3 explored the potential of WGS data for *in silico* speciation of *Campylobacter*. The results of the study showed that *atpA* and rMLST could be used for rapid identification of *Campylobacter* species; however, the detection of horizontal exchange of *atpA* alleles between *C. jejuni* and *C. coli* in Chapter 5 was cause for concern. This finding underscores the benefits of using multiple loci when typing recombinant organisms such as *Campylobacter*. In addition to identifying *in silico* approaches suitable for *Campylobacter* speciation, this study described the development of a genus-wide core-genome, using the most representative dataset to date. Integration of the core genome into the *Campylobacter jejuni*/*coli* PubMLST database will improve population annotation and evolutionary analyses of diverse members of the genus. Furthermore, the approach could also be used to develop

novel *in vitro* assays for the detection and identification of *Campylobacter*, as previously demonstrated for *Neisseria* (Bennett *et al.*, 2014).

The second half of this thesis examined the molecular epidemiology of *Campylobacter* from chickens and human disease samples from South Africa. The results of previous studies carried out in South Africa suggest that live chickens and retail meat are likely contaminated with *Campylobacter* at rates similar to those observed in high-income countries (Mosupye & von Holy, 2000; van Nierop *et al.*, 2005; Bartkowiak-Higgo *et al.*, 2006; Bester & Essack, 2012). The study described in Chapter 4 used WGS to characterise the genetic diversity among *Campylobacter* isolates obtained from two flocks from the same free-range farm that were slaughtered eight weeks apart. Surprisingly, the study isolates were highly homogeneous, which contrasted sharply with the results of previous studies (Newell & Fearnley, 2003; El-Shibiny *et al.*, 2005; Van Overbeke *et al.*, 2006; Colles *et al.*, 2008, 2010, 2011, 2015). Although these results may be due in part to the small sample size, this is the only example of repeated sampling of individual flocks to date. Further studies are required to determine whether or not the observed diversity is unique to the study dataset. Importantly, these results highlight our lack of knowledge of the genomic diversity of *Campylobacter* in the food chain. An improved understanding of this aspect of *Campylobacter* epidemiology is essential for proper monitoring of farmers and retailers, and also for outbreak investigations.

In spite of having a history of *Campylobacter* research, there was little sequence-based typing data available for South African isolates at the onset of this work.

However, the availability of a large archive of freeze-dried human campylobacteriosis isolates made it possible to use WGS to compare the epidemiology of *Campylobacter* collected in Cape Town in the 1990s and those obtained between 2011 and 2012, as described in Chapter 5. This study emphasises the value of historical isolate collections: it was possible to obtain high-quality WGS data from the majority of non-viable *Campylobacter* isolates. Hierarchical gene-by-gene analyses confirmed that the epidemiology of *Campylobacter* from Cape Town is similar to, yet distinct from, that of high-income countries. *C. jejuni* subsp. *jejuni*, *C. upsaliensis*, *C. jejuni* subsp. *doylei*, and *C. coli* were present in 1991 and 2011-2012, while the archive collection also included *C. hyointestinalis* and *C. concisus,* as well as a putative novel *C. jejuni* subsp. *jejuni*/*C. jejuni* subsp. *doylei* hybrid lineage. MLST analyses indicated that *Campylobacter* species present in Cape Town are genetically diverse, suggesting the presence of well-established reservoirs. In the case of *C. jejuni* subsp. *jejuni* and *C. coli*, although clonal complexes appear to have changed over time, the majority of isolates belonged to agriculture-associated lineages commonly found in global human disease isolate collections (Colles & Maiden, 2012). Importantly, high-resolution wgMLST analyses did not find these isolates to be distinct from UK *Campylobacter*. As suggested by previous studies (McCarthy *et al.*, 2007; Sheppard *et al.*, 2010b), it is likely that diet, agricultural practices, and food distribution have affected the epidemiology of *Campylobacter* in South Africa.

At present, the lack of easily understandable, open-source analysis tools is a limiting factor in the uptake of WGS for diagnostic and public health microbiology purposes (Fricke & Rasko, 2014). The studies presented throughout this thesis demonstrate

the utility, accessibility, flexibility, and scalability of the gene-by-gene approach and BIGSDB software, which has the added advantage of not requiring extensive bioinformatics resources or expertise. The continued development of BIGSDB and other similar open-source tools will improve uptake of WGS in routine laboratories able to afford NGS machines, and it will also make publicly available genomic data more accessible to clinical microbiologists worldwide. Finally, the studies presented here have raised further questions regarding the epidemiology and biology of *Campylobacter*, and this thesis is likely to act as a springboard for additional studies. That the unique collection of genomes generated for this thesis will be made publicly available will also enhance further research on *Campylobacter* epidemiology, population genetics, and evolution.

REFERENCES

Abelson, P., Potter Forbes, M. & Hall, G. (2006). *The annual cost of foodborne illness in Australia*. Canberra (Australia): Commonwealth Department of Health and Ageing.

Agunos, A. *et al.* (2014). A systematic review characterizing on-farm sources of *Campylobacter* spp. for broiler chickens. *PloS One*. 9 (8). p.p. e104905.

Ahmed, M.U., Dunn, L. & Ivanova, E.P. (2012). Evaluation of current molecular approaches for genotyping of *Campylobacter jejuni* strains. *Foodborne Pathogens and Disease*. 9 (5). p.pp. 375–85.

Ali, A. *et al.* (2012). *Campylobacter fetus* subspecies: comparative genomics and prediction of potential virulence targets. *Gene*. 508 (2). p.pp. 145–56.

Alispahic, M. *et al.* (2010). Species-specific identification and differentiation of *Arcobacter*, *Helicobacter* and *Campylobacter* by full-spectral matrix-associated laser desorption/ionization time of flight mass spectrometry analysis. *Journal of Medical Microbiology*. 59 (Pt 3). p.pp. 295–301.

Allen, V.M. *et al.* (2007). *Campylobacter* spp. contamination of chicken carcasses during processing in relation to flock colonisation. *International Journal of Food Microbiology*. 113 (1). p.pp. 54–61.

Allos, B.M. (2001). *Campylobacter jejuni* Infections: update on emerging issues and trends. *Clinical Infectious Diseases*. 32 (8). p.pp. 1201–1206.

Amar, C.F.L. *et al.* (2007). Detection by PCR of eight groups of enteric pathogens in 4,627 faecal samples: re-examination of the English case-control Infectious Intestinal Disease Study (1993-1996). *European Journal of Clinical Microbiology & Infectious Diseases*. 26 (5). p.pp. 311–23.

Anderson, N.W., Buchan, B.W. & Ledeboer, N.A. (2014). Comparison of the BD MAX enteric bacterial panel to routine culture methods for detection of *Campylobacter*, enterohemorrhagic *Escherichia coli* (O157), *Salmonella*, and *Shigella* isolates in preserved stool specimens. *Journal of Clinical Microbiology*. 52 (4). p.pp. 1222–4.

Angelo, K.M. *et al.* (2015). Outbreak of *Salmonella* Newport infections linked to cucumbers--United States, 2014. *MMWR. Morbidity and Mortality Weekly Report*. 64 (6). p.pp. 144–7.

Angiuoli, S. V & Salzberg, S.L. (2011). Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics*. 27 (3). p.pp. 334–42.

Anon (2006). *BS EN ISO 10272-1:2006 Microbiology of Food and Animal Feeding*

Stuffs-Horizontal Method for Detection and Enumeration of Campylobacter *spp.* London.

Asakura, H. *et al.* (2007). Deletion of *peb4* gene impairs cell adhesion and biofilm formation in *Campylobacter jejuni*. *FEMS Microbiology Letters*. 275 (2). p.pp. 278–85.

Ashton, P.M. *et al.* (2015a). MinION Nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nature Biotechnology*. 33 (3). p.pp. 296–300.

Ashton, P.M. *et al.* (2015b). Whole Genome Sequencing for the Retrospective Investigation of an Outbreak of *Salmonella* Typhimurium DT 8. *PLoS Currents*. 7.

Bachtiar, B.M., Coloe, P.J. & Fry, B.N. (2007). Knockout mutagenesis of the *kpsE* gene of *Campylobacter jejuni* 81116 and its involvement in bacterium-host interactions. *FEMS Immunology and Medical Microbiology*. 49 (1). p.pp. 149–54.

Backert, S. *et al.* (2013). Transmigration route of *Campylobacter jejuni* across polarized intestinal epithelial cells: paracellular, transcellular or both? *Cell Communication and Signaling*. 11 (1). p.p. 72.

Baker, K.S. *et al.* (2015). Intercontinental dissemination of azithromycin-resistant shigellosis through sexual transmission: a cross-sectional study. *The Lancet. Infectious Diseases*. 15 (8). p.pp. 913–21.

Bankevich, A. *et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*. 19 (5). p.pp. 455–77.

Banoo, S. *et al.* (2010). Evaluation of diagnostic tests for infectious diseases: general principles. *Nature Reviews Microbiology*. 8 (12). p.pp. S17–S29.

Barquist, L. *et al.* (2013). A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. *Nucleic Acids Research*. 41 (8). p.pp. 4549–64.

Bartkowiak-Higgo, A.J. *et al.* (2006). A pilot study on post-evisceration contamination of broiler carcasses and ready-to-sell livers and intestines (mala) with *Campylobacter jejuni* and *Campylobacter coli* in a high-throughput South African poultry abattoir. *Journal of the South African Veterinary Association*. 77 (3). p.pp. 114–9.

Bayley, H. (2015). Nanopore sequencing: from imagination to reality. *Clinical Chemistry*. 61 (1). p.pp. 25–31.

Beaugerie, L. & Petit, J.-C. (2004). Microbial-gut interactions in health and disease.

Antibiotic-associated diarrhoea. *Best Practice & Research Clinical Gastroenterology*. 18 (2). p.pp. 337–52.

Beisele, M. *et al.* (2011). *Helicobacter marmotae* and novel *Helicobacter* and *Campylobacter* species isolated from the livers and intestines of prairie dogs. *Journal of Medical Microbiology*. 60 (Pt 9). p.pp. 1366–74.

Belkaid, Y. & Hand, T.W. (2014). Role of the microbiota in immunity and inflammation. *Cell*. 157 (1). p.pp. 121–41.

Bennett, J.S. *et al.* (2012). A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*. *Microbiology*. 158 (Pt 6). p.pp. 1570–80.

Bennett, J.S. *et al.* (2014). Identifying *Neisseria* species by use of the 50S ribosomal protein L6 (*rplF*) gene. *Journal of Clinical Microbiology*. 52 (5). p.pp. 1375–81.

Berndtson, E., Danielsson-Tham, M.L. & Engvall, A. (1996). *Campylobacter* incidence on a chicken farm and the spread of *Campylobacter* during the slaughter process. *International Journal of Food Microbiology*. 32 (1-2). p.pp. 35–47.

Bessède, E. *et al.* (2011). Identification of *Campylobacter* species and related organisms by matrix assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry. *Clinical Microbiology and Infection*. 17 (11). p.pp. 1735–9.

Bessell, P.R. *et al.* (2012). Using sequence data to identify alternative routes and risk of infection: a case-study of *Campylobacter* in Scotland. *BMC Infectious Diseases*. 12. p.p. 80.

Best, E.L. *et al.* (2003). Applicability of a rapid duplex real-time PCR assay for speciation of *Campylobacter jejuni* and *Campylobacter coli* directly from culture plates. *FEMS Microbiology Letters*. 229 (2). p.pp. 237–41.

Bester, L.A. & Essack, S.Y. (2012). Observational study of the prevalence and antibiotic resistance of *Campylobacter* spp. from different poultry production systems in KwaZulu-Natal, South Africa. *Journal of Food Protection*. 75 (1). p.pp. 154–159.

Bester, L.A. & Essack, S.Y. (2008). Prevalence of antibiotic resistance in *Campylobacter* isolates from commercial poultry suppliers in KwaZulu-Natal, South Africa. *The Journal of Antimicrobial Chemotherapy*. 62 (6). p.pp. 1298–1300.

Bhutta, Z.A. *et al.* (2013). Interventions to address deaths from childhood pneumonia and diarrhoea equitably: what works and at what cost? *Lancet*. 381 (9875). p.pp. 1417–29.

Biggs, P.J. *et al.* (2011). Whole-genome comparison of two *Campylobacter jejuni* isolates of the same sequence type reveals multiple loci of different ancestral lineage. *PloS One*. 6 (11). p.p. e27121.

Biswas, D., Itoh, K. & Sasakawa, C. (2003). Role of microfilaments and microtubules in the invasion of INT-407 cells by *Campylobacter jejuni*. *Microbiology and Immunology*. 47 (6). p.pp. 469–73.

Biswas, D., Itoh, K. & Sasakawa, C. (2000). Uptake pathways of clinical and healthy animal isolates of *Campylobacter jejuni* into INT-407 cells. *FEMS Immunology and Medical Microbiology*. 29 (3). p.pp. 203–11.

Black, R.E. *et al.* (1988). Experimental *Campylobacter jejuni* infection in humans. *The Journal of Infectious Diseases*. 157 (3). p.pp. 472–9.

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

Blattner, F.R. *et al.* (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*. 277 (5331). p.pp. 1453–62.

Bok, H.E., Greeff, A.S. & Crewe-Brown, H.H. (1991). Incidence of toxigenic *Campylobacter* strains in South Africa. *Journal of Clinical Microbiology*. 29 (6). p.pp. 1262–4.

Bokkenheuser, V.D. *et al.* (1979). Detection of enteric campylobacteriosis in children. *Journal of Clinical Microbiology*. 9 (2). p.pp. 227–32.

Boschi-Pinto, C., Velebit, L. & Shibuya, K. (2008). Estimating child mortality due to diarrhoea in developing countries. *Bulletin of the World Health Organization*. 86 (9). p.pp. 710–7.

Boysen, L. *et al.* (2014). Source attribution of human campylobacteriosis in Denmark. *Epidemiology and Infection*. 142 (8). p.pp. 1599–608.

Bradnam, K.R. *et al.* (2013). Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species. *GigaScience*. 2 (1). p.p. 10.

Bratcher, H.B. *et al.* (2014). A gene-by-gene population genomics platform: *de novo* assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes. *BMC Genomics*. 15. p.p. 1138.

Brathwaite, K.J. *et al.* (2013). Complete Genome Sequence of Universal Bacteriophage Host Strain *Campylobacter jejuni* subsp. *jejuni* PT14. *Genome Announcements*. 1 (6).

Brehony, C. *et al.* (2015). Meningococcal vaccine antigen diversity in global databases. *Euro Surveillance*. 20 (49).

Broman, T. *et al.* (2002). *Campylobacter jejuni* in black-headed gulls (*Larus ridibundus*): prevalence, genotypes, and influence on *C. jejuni* epidemiology. *Journal of Clinical Microbiology*. 40 (12). p.pp. 4594–602.

Broman, T. *et al.* (2004). Diversities and similarities in PFGE profiles of *Campylobacter jejuni* isolated from migrating birds and humans. *Journal of Applied Microbiology*. 96 (4). p.pp. 834–43.

Brown, H.L. *et al.* (2015). Prevention of Biofilm Formation and Removal of Existing Biofilms by Extracellular DNases of *Campylobacter jejuni*. *PLoS One*. 10 (3). p.p. e0121680.

Brunelle, B.W., Bearson, B.L. & Bearson, S.M.D. (2014). Chloramphenicol and tetracycline decrease motility and increase invasion and attachment gene expression in specific isolates of multidrug-resistant *Salmonella enterica* serovar Typhimurium. *Frontiers in Microbiology*. 5. p.p. 801.

Bryant, D. & Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*. 21 (2). p.pp. 255–65.

Bryant, J.M. *et al.* (2013). Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: a retrospective cohort study. *Lancet*. 381 (9877). p.pp. 1551–60.

Buchan, B.W. *et al.* (2013). Clinical evaluation of a real-time PCR assay for identification of *Salmonella*, *Shigella*, *Campylobacter* (*Campylobacter jejuni* and *C. coli*), and shiga toxin-producing *Escherichia coli* isolates in stool specimens. *Journal of Clinical Microbiology*. 51 (12). p.pp. 4001–7.

Buchan, B.W. & Ledeboer, N.A. (2014). Emerging Technologies for the Clinical Microbiology Laboratory. *Clinical Microbiology Reviews*. 27 (4). p.pp. 783–822.

Bull, S.A. *et al.* (2008). Flock health indicators and *Campylobacter* spp. in commercial housed broilers reared in Great Britain. *Applied and Environmental Microbiology*. 74 (17). p.pp. 5408–13.

Bull, S.A. *et al.* (2006). Sources of *Campylobacter* spp. colonizing housed broiler flocks during rearing. *Applied and Environmental Microbiology*. 72 (1). p.pp. 645–52.

Bullman, S. *et al.* (2012). Molecular-based detection of non-culturable and emerging campylobacteria in patients presenting with gastroenteritis. *Epidemiology and Infection*. 140 (4). p.pp. 684–8.

Buss, S.N. *et al.* (2015). Multicenter evaluation of the BioFire FilmArray gastrointestinal panel for etiologic diagnosis of infectious gastroenteritis. *Journal of Clinical Microbiology*. 53 (3). p.pp. 915–25.

Butzler, J.P. (2004). *Campylobacter*, from obscurity to celebrity. *Clinical Microbiology and Infection*. 10 (10). p.pp. 868–76.

Butzler, J.P. *et al.* (1973). Related *vibrio* in stools. *The Journal of Pediatrics*. 82 (3). p.pp. 493–495.

Callicott, K.A. *et al.* (2006). Lack of evidence for vertical transmission of *Campylobacter* spp. in chickens. *Applied and Environmental Microbiology*. 72 (9). p.pp. 5794–8.

Candela, M. *et al.* (2008). Interaction of probiotic *Lactobacillus* and *Bifidobacterium* strains with human intestinal epithelial cells: adhesion properties, competition against enteropathogens and modulation of IL-8 production. *International Journal of Food Microbiology*. 125 (3). p.pp. 286–92.

Carlson, R. (2003). The pace and proliferation of biological technologies. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*. 1 (3). p.pp. 203–14.

Carneiro, M.O. *et al.* (2012). Pacific Biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics*. 13. p.p. 375.

Chaisson, M.J. & Pevzner, P.A. (2008). Short read fragment assembly of bacterial genomes. *Genome Research*. 18 (2). p.pp. 324–30.

Chang, J.Y. *et al.* (2008). Decreased diversity of the fecal microbiome in recurrent *Clostridium difficile*-associated diarrhea. *The Journal of Infectious Diseases*. 197 (3). p.pp. 435–8.

Chaudhuri, R.R. *et al.* (2013). Comprehensive assignment of roles for *Salmonella* typhimurium genes in intestinal colonization of food-producing animals. *PLoS Genetics*. 9 (4). p.p. e1003456.

Check Hayden, E. (2015). Pint-sized DNA sequencer impresses first users. *Nature*. 521 (7550). p.pp. 15–6.

Checkley, W. *et al.* (2008). Multi-country analysis of the effects of diarrhoea on childhood stunting. *International Journal of Epidemiology*. 37 (4). p.pp. 816–30.

Chen, L. *et al.* (2014). *What Does 'Service Delivery' Really Mean?* [Online]. 2014. Available from: http://www.worldpolicy.org/blog/2014/05/13/what-does-service-delivery-really-mean. [Accessed: 13 January 2016].

Chen, Y. *et al.* (2013). Whole-genome sequencing of gentamicin-resistant *Campylobacter coli* isolated from U.S. retail meats reveals novel plasmid-mediated aminoglycoside resistance genes. *Antimicrobial Agents and Chemotherapy*. 57 (11). p.pp. 5398–405.

Chiba, A. *et al.* (1992). Serum IgG antibody to ganglioside GQ1b is a possible

marker of Miller Fisher syndrome. *Annals of Neurology*. 31 (6). p.pp. 677–9.

Clark, C.G. *et al.* (2014a). DNA sequence heterogeneity of *Campylobacter jejuni* CJIE4 prophages and expression of prophage genes. *PloS One*. 9 (4). p.p. e95349.

Clark, C.G. *et al.* (2014b). The CJIE1 prophage of *Campylobacter jejuni* affects protein expression in growth media with and without bile salts. *BMC Microbiology*. 14 (1). p.p. 70.

Clarke, J. *et al.* (2009). Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology*. 4 (4). p.pp. 265–70.

Cody, A.J. *et al.* (2012). A longitudinal 6-year study of the molecular epidemiology of clinical *Campylobacter* isolates in Oxfordshire, United kingdom. *Journal of Clinical Microbiology*. 50 (10). p.pp. 3193–201.

Cody, A.J. *et al.* (2013). Real-time genomic epidemiological evaluation of human *Campylobacter* isolates by use of whole-genome multilocus sequence typing. *Journal of Clinical Microbiology*. 51 (8). p.pp. 2526–34.

Cody, A.J. *et al.* (2010). Where does *Campylobacter* come from? A molecular odyssey. *Advances in Experimental Medicine and Biology*. 659. p.pp. 47–56.

Cody, A.J. *et al.* (2015). Wild bird-associated *Campylobacter jejuni* isolates are a consistent source of human disease, in Oxfordshire, United Kingdom. *Environmental Microbiology Reports*. 7 (5). p.pp. 782–8.

Cody, A.J., Maiden, M.J.C. & Dingle, K.E. (2009). Genetic diversity and stability of the *porA* allele as a genetic marker in human *Campylobacter* infection. *Microbiology*. 155 (Pt 12). p.pp. 4145–54.

Coker, A.O. *et al.* (2002). Human campylobacteriosis in developing countries. *Emerging Infectious Diseases*. 8 (3). p.pp. 237–244.

Colles, F.M. *et al.* (2008). *Campylobacter* infection of broiler chickens in a free-range environment. *Environmental Microbiology*. 10 (8). p.pp. 2042–50.

Colles, F.M. *et al.* (2010). Comparison of *Campylobacter* populations isolated from a free-range broiler flock before and after slaughter. *International Journal of Food Microbiology*. 137 (2-3). p.pp. 259–64.

Colles, F.M. *et al.* (2015). The long-term dynamics of *Campylobacter* colonizing a free-range broiler breeder flock: an observational study. *Environmental Microbiology*. 17 (4). p.pp. 938–46.

Colles, F.M. *et al.* (2011). The prevalence of *Campylobacter* amongst a free-range broiler breeder flock was primarily affected by flock age. *PloS One*. 6 (12). p.p. e22825.

Colles, F.M. & Maiden, M.C.J. (2012). *Campylobacter* sequence typing: applications and future prospects. *Microbiology*. 158 (11). p.pp 2695-2709.

Compeau, P.E.C., Pevzner, P.A. & Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. *Nature Biotechnology*. 29 (11). p.pp. 987–91.

Connor, T.R. *et al.* (2015). Species-wide whole genome sequencing reveals historical global spread and recent local persistence in *Shigella flexneri*. *eLife*. 4.

Cooper, K.K. *et al.* (2011). Complete genome sequence of *Campylobacter jejuni* strain S3. *Journal of Bacteriology*. 193 (6). p.pp. 1491–2.

Cooper, R. *et al.* (2002). Genetic basis of quinolone resistance and epidemiology of resistant and susceptible isolates of porcine *Campylobacter coli* strains. *Journal of Applied Microbiology*. 93 (2). p.pp. 241–9.

Coovadia, H. *et al.* (2009). The health and health system of South Africa: historical roots of current public health challenges. *Lancet*. 374 (9692). p.pp. 817–34.

Corander, J. & Marttinen, P. (2006). Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology*. 15 (10). p.pp. 2833–43.

Corcoran, D. *et al.* (2006). Antimicrobial resistance profiling and *fla*-typing of Irish thermophillic *Campylobacter* spp. of human and poultry origin. *Letters in Applied Microbiology*. 43 (5). p.pp. 560–5.

Corry, J.E.L. *et al.* (1995). Culture media for the isolation of campylobacters. *International Journal of Food Microbiology*. 26 (1). p.pp. 43–76.

Couturier, B.A. *et al.* (2013). Detection of non-*jejuni* and -*coli Campylobacter* species from stool specimens with an immunochromatographic antigen detection assay. *Journal of Clinical Microbiology*. 51 (6). p.pp. 1935–7.

Cox, N.A. *et al.* (2012). Evidence for horizontal and vertical transmission in *Campylobacter* passage from hen to her progeny. *Journal of Food Protection*. 75 (10). p.pp. 1896–902.

Croucher, N.J. *et al.* (2011). Rapid pneumococcal evolution in response to clinical interventions. *Science*. 331 (6016). p.pp. 430–4.

Croucher, N.J. & Thomson, N.R. (2010). Studying bacterial transcriptomes using RNA-seq. *Current Opinion in Microbiology*. 13 (5). p.pp. 619–24.

Crushell, E. *et al.* (2004). Enteric *Campylobacter*: purging its secrets? *Pediatric Research*. 55 (1). p.pp. 3–12.

Dallman, T.J. *et al.* (2014). An investigation of the diversity of strains of enteroaggregative *Escherichia coli* isolated from cases associated with a large

multi-pathogen foodborne outbreak in the UK. *PloS One*. 9 (5). p.p. e98103.

Dallman, T.J. *et al.* (2015). Whole-Genome Sequencing for National Surveillance of Shiga Toxin-Producing *Escherichia coli* O157. *Clinical Infectious Diseases.* 61 (3). p.pp. 305–12.

Darling, A.C.E. *et al.* (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*. 14 (7). p.pp. 1394–403.

Darling, A.E., Mau, B. & Perna, N.T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS One*. 5 (6). p.p. e11147.

da Silva Quetz, J. *et al.* (2010). *Campylobacter jejuni* and *Campylobacter coli* in children from communities in Northeastern Brazil: molecular detection and relation to nutritional status. *Diagnostic Microbiology and Infectious Disease*. 67 (3). p.pp. 220–7.

da Silva Tatley, F.M., Lastovica, A.J. & Steyn, L.M. (1992). Plasmid profiles of '*Campylobacter upsaliensis'* isolated from blood cultures and stools of paediatric patients. *Journal of Medical Microbiology*. 37 (1). p.pp. 8–14.

de Boer, R.F. *et al.* (2013). Detection of *Campylobacter* species and *Arcobacter butzleri* in stool samples by use of real-time multiplex PCR. *Journal of Clinical Microbiology*. 51 (1). p.pp. 253–9.

de Boer, R.F. *et al.* (2010). Improved detection of five major gastrointestinal pathogens by use of a molecular screening approach. *Journal of Clinical Microbiology*. 48 (11). p.pp. 4140–6.

Debruyne, L. *et al.* (2010a). *Campylobacter subantarcticus* sp. *nov*., isolated from birds in the sub-Antarctic region. *International Journal of Systematic and Evolutionary Microbiology*. 60 (Pt 4). p.pp. 815–9.

Debruyne, L. *et al.* (2010b). *Campylobacter volucris* sp. *nov*., isolated from black-headed gulls (*Larus ridibundus*). *International Journal of Systematic and Evolutionary Microbiology*. 60 (Pt 8). p.pp. 1870–5.

Debruyne, L. *et al.* (2009). Novel *Campylobacter lari*-like bacteria from humans and molluscs: description of *Campylobacter peloridis* sp. *nov*., *Campylobacter lari* subsp. *concheus* subsp. *nov*. and *Campylobacter lari* subsp. *lari* subsp. *nov*. *International Journal of Systematic and Evolutionary Microbiology*. 59 (Pt 5). p.pp. 1126–32.

Debruyne, L., Gevers, D. & Vandamme, P. (2008). Taxonomy of the family. In: I. Nachamkin, C. Szymanski, & M. J. Blaser (eds.). *Campylobacter*. Washington, DC: ASM Press, pp. 3–25.

Dekeyser, P. *et al.* (1972). Acute enteritis due to related vibrio: first positive stool

cultures. *The Journal of Infectious Diseases*. 125 (4). p.pp. 390–2.

Desai, M. *et al.* (2001). Genome sequence-based fluorescent amplified fragment length polymorphism of *Campylobacter jejuni*, its relationship to serotyping, and its implications for epidemiological analysis. *Journal of Clinical Microbiology*. 39 (11). p.pp. 3823–9.

Dethlefsen, L. & Relman, D.A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proceedings of the National Academy of Sciences of the United States of America*. 108 Suppl . p.pp. 4554–61.

Dickinson, B. & Surawicz, C.M. (2014). Infectious diarrhea: an overview. *Current Gastroenterology Reports*. 16 (8). p.p. 399.

Didelot, X. *et al.* (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics*. 13 (9). p.pp. 601–12.

Diergaardt, S.M. *et al.* (2004). The occurrence of campylobacters in water sources in South Africa. *Water Research*. 38 (10). p.pp. 2589–95.

Dingle, K.E. *et al.* (2008). Extended sequence typing of *Campylobacter* spp., United Kingdom. *Emerging Infectious Diseases*. 14 (10). p.pp. 1620–2.

Dingle, K.E. *et al.* (2002). Molecular characterization of *Campylobacter jejuni* clones: a basis for epidemiologic investigation. *Emerging Infectious Diseases*. 8 (9). p.pp. 949–55.

Dingle, K.E. *et al.* (2001a). Multilocus sequence typing system for *Campylobacter jejuni*. *Journal of Clinical Microbiology*. 39 (1). p.pp. 14–23.

Dingle, K.E. *et al.* (2005). Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*. *Journal of Clinical Microbiology*. 43 (1). p.pp. 340–7.

Dingle, K.E. *et al.* (2001b). Sequence typing confirms that *Campylobacter jejuni* strains associated with Guillain-Barré and Miller-Fisher syndromes are of diverse genetic lineage, serotype, and flagella type. *Journal of Clinical Microbiology*. 39 (9). p.pp. 3346–9.

Dohm, J.C. *et al.* (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*. 36 (16). p.p. e105.

Domingues, A.R. *et al.* (2012). Source attribution of human campylobacteriosis using a meta-analysis of case-control studies of sporadic infections. *Epidemiology and Infection*. 140 (6). p.pp. 970–81.

Doolittle, W.F. & Papke, R.T. (2006). Genomics and the bacterial species problem. *Genome Biology*. 7 (9). p.p. 116.

Duim, B. *et al.* (2000). Amplified fragment length polymorphism analysis of *Campylobacter jejuni* strains isolated from chickens and from patients with gastroenteritis or Guillain-Barré or Miller Fisher syndrome. *Applied and Environmental Microbiology*. 66 (9). p.pp. 3917–23.

Duim, B. *et al.* (2001). Differentiation of *Campylobacter* species by AFLP fingerprinting. *Microbiology*. 147 (Pt 10). p.pp. 2729–37.

Duim, B. *et al.* (1999). High-resolution genotyping of *Campylobacter* strains isolated from poultry and humans with amplified fragment length polymorphism fingerprinting. *Applied and Environmental Microbiology*. 65 (6). p.pp. 2369–75.

Duim, B. *et al.* (2003). Molecular evidence for dissemination of unique *Campylobacter jejuni* clones in Curaçao, Netherlands Antilles. *Journal of Clinical Microbiology*. 41 (12). p.pp. 5593–7.

Duim, B. & Wagenaar, J. (2006). Amplified fragment-length polymorphism and protein profiling for identification of *Campylobacter lari* subgroups. *Methods in Molecular Biology*. 345. p.pp. 119–30.

Earl, D. *et al.* (2011). Assemblathon 1: a competitive assessment of *de novo* short read assembly methods. *Genome Research*. 21 (12). p.pp. 2224–41.

Eberle, K.N. & Kiess, A.S. (2012). Phenotypic and genotypic methods for typing *Campylobacter jejuni* and *Campylobacter coli* in poultry. *Poultry Science*. 91 (1). p.pp. 255–64.

Edwards, D.J. & Holt, K.E. (2013). Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microbial Informatics and Experimentation*. 3 (1). p.p. 2.

Edwards, D.S. *et al.* (2014). Campylobacteriosis outbreak associated with consumption of undercooked chicken liver pâté in the East of England, September 2011: identification of a dose-response risk. *Epidemiology and Infection*. 142 (2). p.pp. 352–7.

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

Eid, J. *et al.* (2009). Real-time DNA sequencing from single polymerase molecules. *Science*. 323 (5910). p.pp. 133–8.

Eisenstein, M. (2012). Oxford Nanopore announcement sets sequencing sector abuzz. *Nature Biotechnology*. 30 (4). p.pp. 295–6.

El-Shibiny, A., Connerton, P.L. & Connerton, I.F. (2005). Enumeration and diversity of campylobacters and bacteriophages isolated during the rearing cycles of

free-range and organic chickens. *Applied and Environmental Microbiology*. 71 (3). p.pp. 1259–66.

Elwell, C.A. & Dreyfus, L.A. (2000). DNase I homologous residues in CdtB are critical for cytolethal distending toxin-mediated cell cycle arrest. *Molecular Microbiology*. 37 (4). p.pp. 952–63.

Engberg, J. *et al.* (2001). Absence of clonality of *Campylobacter jejuni* in serotypes other than HS:19 associated with Guillain-Barré syndrome and gastroenteritis. *The Journal of Infectious Diseases*. 184 (2). p.pp. 215–20.

Engberg, J. (2006). Contributions to the epidemiology of *Campylobacter* infections. A review of clinical and microbiological studies. *Danish Medical Bulletin*. 53 (4). p.pp. 361–89.

Engberg, J. *et al.* (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 (1). p.pp. 286–91.

Erdman, S.E. & Poutahidis, T. (2015). Gut bacteria and cancer. *Biochimica et Biophysica Acta*. 1856 (1). p.pp. 86–90.

Etoh, Y. *et al.* (1993). *Campylobacter showae* sp. *nov*., isolated from the human oral cavity. *International Journal of Systematic Bacteriology*. 43 (4). p.pp. 631–9.

European Food Safety Authority and European Centre for Disease Prevention and Control (2014). The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2012. *EFSA Journal*. 12 (2). p.p. 2547.

European Food Safety Authority and European Centre for Disease Prevention and Control & European Centre for Disease Prevention and Control (2015). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013. *EFSA Journal*. 13 (1). p.p. 3991.

Eyre, D.W. *et al.* (2012). A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open*. 2 (3).

Falush, D., Stephens, M. & Pritchard, J.K. (2003). Inference of Population Structure Using Multilocus Genotype Data : Linked Loci and Correlated Allele Frequencies. *Genetics*. 164. p.pp. 1567–1587.

Farthing, M. *et al.* (2013). Acute diarrhea in adults and children: a global perspective. *Journal of Clinical Gastroenterology*. 47 (1). p.pp. 12–20.

Fayos, A. *et al.* (1993). Molecular subtyping by genome and plasmid analysis of *Campylobacter jejuni* serogroups O1 and O2 (Penner) from sporadic and

outbreak cases of human diarrhoea. *Epidemiology and Infection*. 111 (3). p.pp. 415–27.

Fernandes, A.M. *et al.* (2015). Partial Failure of Milk Pasteurization as a Risk for the Transmission of *Campylobacter* From Cattle to Humans. *Clinical Infectious Diseases.* 61 (6). p.pp. 903–9.

Fernández, H. *et al.* (2008). Occurrence of *Campylobacter* species in healthy well-nourished and malnourished children. *Brazilian Journal of Microbiology*. 39 (1). p.pp. 56–8.

Figueroa, G. *et al.* (1989). Prospective study of *Campylobacter jejuni* infection in Chilean infants evaluated by culture and serology. *Journal of Clinical Microbiology*. 27 (5). p.pp. 1040–4.

Fitzgerald, C. *et al.* (2014). *Campylobacter fetus* subsp. *testudinum* subsp. *nov*., isolated from humans and reptiles. *International Journal of Systematic and Evolutionary Microbiology*. 64 (Pt 9). p.pp. 2944–8.

Fitzgerald, C. (2015). *Campylobacter*. *Clinics in Laboratory Medicine*. 35 (2). p.pp. 289–98.

Fitzgerald, C. *et al.* (2001). Evaluation of methods for subtyping *Campylobacter jejuni* during an outbreak involving a food handler. *Journal of Clinical Microbiology*. 39 (7). p.pp. 2386–90.

Fitzgerald, C., Whichard, J. & Nachamkin, I. (2008). Diagnosis and antimicrobial susceptibility of *Campylobacter* species. In: I. Nachamkin, C. M. Szymanski, & M. J. Blaser (eds.). *Campylobacter*. Washington, D.C.: ASM Press, pp. 227–243.

Fleischmann, R.D. *et al.* (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*. 269 (5223). p.pp. 496–512.

Floch, P. *et al.* (2012). Evaluation of the positive predictive value of a rapid Immunochromatographic test to detect *Campylobacter* in stools. *Gut Pathogens*. 4 (1). p.p. 17.

Foley, S.L., Lynne, A.M. & Nayak, R. (2009). Molecular typing methodologies for microbial source tracking and epidemiological investigations of Gram-negative bacterial foodborne pathogens. *Infection, Genetics and Evolution*. 9 (4). p.pp. 430–40.

Food Standards Agency (2015). *A microbiological survey of* Campylobacter *contamination in fresh whole UK-produced chilled chickens at retail sale – February 2014 to February 2015*.

Food Standards Agency (2013). *Annual Report of the Chief Scientist 2012/13*.

Food Standards Agency (2011). *Foodborne disease strategy 2010-2015. An FSA*

programme for the reduction of foodborne disease in the UK.

Foster, G. *et al.* (2004). *Campylobacter insulaenigrae* sp. *nov*., isolated from marine mammals. *International Journal of Systematic and Evolutionary Microbiology*. 54 (Pt 6). p.pp. 2369–73.

Fournier, P.-E., Dubourg, G. & Raoult, D. (2014). Clinical detection and characterization of bacterial pathogens in the genomics era. *Genome Medicine*. 6 (11). p.p. 114.

Fouts, D.E. *et al.* (2005). Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. *PLoS Biology*. 3 (1). p.p. e15.

Frank, C. *et al.* (2011). Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *The New England Journal of Medicine*. 365 (19). p.pp. 1771–80.

Fricke, W.F. & Rasko, D.A. (2014). Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions. *Nature Reviews Genetics*. 15 (1). p.pp. 49–55.

Friedman, C.R. *et al.* (2004). Risk factors for sporadic *Campylobacter* infection in the United States: A case-control study in FoodNet sites. *Clinical Infectious Disease*. 38 Suppl 3. p.pp. S285–96.

Friedrich, A. *et al.* (2015). Short-term evolutionary dynamics of *Campylobacter* jejuni in chickens. In: *18th International Workshop on Campylobacter, Helicobacter and Related Organisms, Rotorua, New Zealand*. 2015, p. 120.

Friesema, I.H.M., Lugnér, A.K. & van Duynhoven, Y.T.H.P. (2012). Costs of gastroenteritis in the Netherlands, with special attention for severe cases. *European Journal of Clinical Microbiology & Infectious Diseases*. 31 (8). p.pp. 1895–900.

Friis, C. *et al.* (2010). Genomic characterization of *Campylobacter jejuni* strain M1. *PloS One*. 5 (8). p.p. e12253.

Frost, J.A. *et al.* (1998). Serotyping scheme for *Campylobacter jejuni* and *Campylobacter coli* based on direct agglutination of heat-stable antigens. *Journal of Clinical Microbiology*. 36 (2). p.pp. 335–9.

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 (1). p.pp. 47–55.

Fukuda, S. *et al.* (2011). *Bifidobacteria* can protect from enteropathogenic infection through production of acetate. *Nature*. 469 (7331). p.pp. 543–7.

Fukushima, H. *et al.* (2009). Comprehensive and rapid real-time PCR analysis of 21

foodborne outbreaks. *International Journal of Microbiology*. 2009. p.p. 917623.

Fukushima, H., Tsunomori, Y. & Seki, R. (2003). Duplex Real-Time SYBR Green PCR Assays for Detection of 17 Species of Food- or Waterborne Pathogens in Stools. *Journal of Clinical Microbiology*. 41 (11). p.pp. 5134–5146.

Gaillot, O. *et al.* (2011). Cost-effectiveness of switch to matrix-assisted laser desorption ionization-time of flight mass spectrometry for routine bacterial identification. *Journal of Clinical Microbiology*. 49 (12). p.p. 4412.

Galperin, M.Y. *et al.* (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Research*. 43 (Database issue). p.pp. D261–9.

Gautret, P. *et al.* (2012). Infectious diseases among travellers and migrants in Europe, EuroTravNet 2010. *Euro Surveillance*. 17 (26).

Gebhart, C.J. *et al.* (1985). '*Campylobacter hyointestinalis'* sp. *nov*.: a new species of *Campylobacter* found in the intestines of pigs and other animals. *Journal of Clinical Microbiology*. 21 (5). p.pp. 715–20.

Georges-Courbot, M.C. *et al.* (1987). Prospective study of enteric *Campylobacter* infections in children from birth to 6 months in the Central African Republic. *Journal of Clinical Microbiology*. 25 (5). p.pp. 836–9.

Georgiades, K. & Raoult, D. (2010). Defining pathogenic bacterial species in the genomic era. *Frontiers in Microbiology*. 1. p.p. 151.

Gerner-Smidt, P. *et al.* (2006). PulseNet USA: a five-year update. *Foodborne Pathogens and Disease*. 3 (1). p.pp. 9–19.

Gilbert, M.J. *et al.* (2015). *Campylobacter iguaniorum* sp. *nov*., isolated from reptiles. *International Journal of Systematic and Evolutionary Microbiology*. 65 (Pt 3). p.pp. 975–82.

Gilbert, M.J. *et al.* (2013). Complete Genome Sequence of *Campylobacter fetus* subsp. *testudinum* Strain 03-427T. *Genome Announcements*. 1 (6).

Gilbert, M.J. *et al.* (2014a). Complete Genome Sequence of *Campylobacter iguaniorum* Strain 1485ET, Isolated from a Bearded Dragon (*Pogona vitticeps*). *Genome Announcements*. 2 (4).

Gilbert, M.J. *et al.* (2014b). Occurrence, diversity, and host association of intestinal *Campylobacter*, *Arcobacter*, and *Helicobacter* in reptiles. *PloS One*. 9 (7). p.p. e101599.

Gilbreath, J.J. *et al.* (2011). Change is good: variations in common biological mechanisms in the epsilonproteobacterial genera *Campylobacter* and Helicobacter. *Microbiology and Molecular Biology Reviews*. 75 (1). p.pp. 84–132.

Giliomee, H. & Mbenga, B. (2007). *New History of South Africa*. Cape Town: Tafelberg.

Gillespie, I.A. *et al.* (2002). A case-case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: a tool for generating hypotheses. *Emerging Infectious Diseases*. 8 (9). p.pp. 937–42.

Gillespie, I.A. *et al.* (2006). Investigating vomiting and/or bloody diarrhoea in *Campylobacter* jejuni infection. *Journal of Medical Microbiology*. 55 (Pt 6). p.pp. 741–6.

Giltner, C.L. *et al.* (2013). Rapid detection of *Campylobacter* antigen by enzyme immunoassay leads to increased positivity rates. *Journal of Clinical Microbiology*. 51 (2). p.pp. 618–20.

Goldman, C.G. *et al.* (2011). Novel gastric helicobacters and oral campylobacters are present in captive and wild cetaceans. *Veterinary Microbiology*. 152 (1-2). p.pp. 138–45.

Gonzalez, A. *et al.* (2011). The mind-body-microbial continuum. *Dialogues in Clinical Neuroscience*. 13 (1). p.pp. 55–62.

Goossens, H. *et al.* (1986). Modified selective medium for isolation of *Campylobacter* spp. from feces: comparison with Preston medium, a blood-free medium, and a filtration system. *Journal of Clinical Microbiology*. 24 (5). p.pp. 840–3.

Gorkiewicz, G. *et al.* (2003). Species-specific identification of campylobacters by partial 16S rRNA gene sequencing. *Journal of Clinical Microbiology*. 41 (6). p.pp. 2537–46.

Gormley, F.J. *et al.* (2011). A 17-year review of foodborne outbreaks: describing the continuing decline in England and Wales (1992-2008). *Epidemiology and Infection*. 139 (5). p.pp. 688–99.

Gormley, F.J. *et al.* (2008). Has retail chicken played a role in the decline of human campylobacteriosis? *Applied and Environmental Microbiology*. 74 (2). p.pp. 383–90.

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 (1). p.pp. 13–8.

Granato, P.A. *et al.* (2010). Comparison of Premier CAMPY Enzyme Immunoassay (EIA), ProSpecT *Campylobacter* EIA, and ImmunoCard STAT! CAMPY Tests with Culture for Laboratory Diagnosis of *Campylobacter* Enteric Infections. *Journal of Clinical Microbiology*. 48 (11). p.pp. 4022–4027.

Greninger, A.L. *et al.* (2015). Rapid metagenomic identification of viral pathogens in

clinical samples by real-time nanopore sequencing analysis. *Genome Medicine*. 7 (1). p.p. 99.

Griggs, D.J. *et al.* (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 (2). p.pp. 699–707.

Grundmann, H., Hori, S. & Tanner, G. (2001). Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. *Journal of Clinical Microbiology*. 39 (11). p.pp. 4190–2.

Guerrant, R.L. *et al.* (1992). Diarrhea as a cause and an effect of malnutrition: diarrhea prevents catch-up growth and malnutrition increases diarrhea frequency and duration. *The American Journal of Tropical Medicine and Hygiene*. 47 (1 Pt 2). p.pp. 28–35.

Gundogdu, O. *et al.* (2007). Re-annotation and re-analysis of the *Campylobacter jejuni* NCTC11168 genome sequence. *BMC Genomics*. 8. p.p. 162.

Gunther, N.W. *et al.* (2011). GyrB versus 16s rRNA sequencing for the identification of *Campylobacter jejuni*, *Campylobacter coli*, and *Campylobacter lari*. *Journal of Nucleic Acids Investigation*. 2 (1) p.p. 7.

Guo, L. *et al.* (2014). Comparative study of MALDI-TOF MS and VITEK 2 in bacteria identification. *Journal of Thoracic Disease*. 6 (5). p.pp. 534–8.

Hallett, A.F., Botha, P.L. & Logan, A. (1977). Isolation of *Campylobacter fetus* from recent cases of human vibriosis. *The Journal of Hygiene*. 79 (3). p.pp. 381–9.

Hallett, T.B. *et al.* (2014). Embracing different approaches to estimating HIV incidence, prevalence and mortality. *AIDS*. 28 Suppl 4. p.pp. S523–32.

Handley, R.A. *et al.* (2015). PerR controls oxidative stress defence and aerotolerance, but not motility-associated phenotypes of *Campylobacter jejuni*. *Microbiology*. 161 (7). p.pp. 1524-1536.

Hänninen, M.L. *et al.* (2000). A three-year study of *Campylobacter jejuni* genotypes in humans with domestically acquired infections and in chicken samples from the Helsinki area. *Journal of Clinical Microbiology*. 38 (5). p.pp. 1998–2000.

Hänninen, M.L. *et al.* (2001). Genomic relatedness within five common Finnish *Campylobacter jejuni* pulsed-field gel electrophoresis genotypes studied by amplified fragment length polymorphism analysis, ribotyping, and serotyping. *Applied and Environmental Microbiology*. 67 (4). p.pp. 1581–6.

Hänninen, M.L., Hakkinen, M. & Rautelin, H. (1999). Stability of related human and chicken *Campylobacter jejuni* genotypes after passage through chick intestine

studied by pulsed-field gel electrophoresis. *Applied and Environmental Microbiology*. 65 (5). p.pp. 2272–5.

Hansson, I. *et al.* (2007). Summary of the Swedish *Campylobacter* program in broilers, 2001 through 2005. *Journal of Food Protection*. 70 (9). p.pp. 2008–14.

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 (9). p.pp. 2386–92.

Harris, B. *et al.* (2011). Inequities in access to health care in South Africa. *Journal of Public Health Policy*. 32 Suppl 1. p.pp. S102–23.

Harris, S.R. *et al.* (2010). Evolution of MRSA during hospital transmission and intercontinental spread. *Science*. 327 (5964). p.pp. 469–74.

Harris, S.R. *et al.* (2013). Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *The Lancet. Infectious Diseases*. 13 (2). p.pp. 130–6.

Harris, T.D. *et al.* (2008). Single-molecule DNA sequencing of a viral genome. *Science*. 320 (5872). p.pp. 106–9.

Harrison, O.B. *et al.* (2015). Genomic Analysis of the Evolution and Global Spread of Hyper-invasive Meningococcal Lineage 5. *EBioMedicine*. 2 (3). p.pp. 234–243.

Havelaar, A.H. *et al.* (2009). Immunity to *Campylobacter*: its role in risk assessment and epidemiology. *Critical Reviews in Microbiology*. 35 (1). p.pp. 1–22.

Havelaar, A.H. *et al.* (2015). World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. *PLOS Medicine*. 12 (12). p.p. e1001923.

He, Y. *et al.* (2015). Complete Genome Sequence of *Campylobacter jejuni* YH001 from Beef Liver, Which Contains a Novel Plasmid. *Genome Announcements*. 3 (1).

Hedberg, C.W. *et al.* (2001). Limitations of pulsed-field gel electrophoresis for the routine surveillance of *Campylobacter* infections. *The Journal of Infectious Diseases*. 184 (2). p.pp. 242–4.

Hedge, J. & Wilson, D.J. (2014). Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. *mBio*. 5 (6). p.p. e02158.

Hendriksen, R.S. *et al.* (2011). Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. *mBio*. 2 (4). p.pp. e00157–11.

Hendrixson, D.R. (2006). A phase-variable mechanism controlling the *Campylobacter jejuni* FlgR response regulator influences commensalism. *Molecular Microbiology*. 61 (6). p.pp. 1646–1659.

Herman, L. *et al.* (2003). Routes for *Campylobacter* contamination of poultry meat: epidemiological study from hatchery to slaughterhouse. *Epidemiology and Infection*. 131 (3). p.pp. 1169–80.

Hickey, T.E. *et al.* (2000). *Campylobacter jejuni* cytolethal distending toxin mediates release of interleukin-8 from intestinal epithelial cells. *Infection and Immunity*. 68 (12). p.pp. 6535–41.

Hickson, M. *et al.* (2007). Use of probiotic *Lactobacillus* preparation to prevent diarrhoea associated with antibiotics: randomised double blind placebo controlled trial. *BMJ (Clinical Research ed.)*. 335 (7610). p.p. 80.

Hill, D.M.C. *et al.* (2015). Genomic epidemiology of age-associated meningococcal lineages in national surveillance: an observational cohort study. *The Lancet. Infectious Diseases*. 15 (12). p.pp. 1420–8.

Hindiyeh, M. *et al.* (2000). Rapid detection of *Campylobacter jejuni* in stool specimens by an enzyme immunoassay and surveillance for *Campylobacter upsaliensis* in the greater Salt Lake City area. *Journal of Clinical Microbiology*. 38 (8). p.pp. 3076–9.

Hoffmann, M. *et al.* (2015). Tracing Origins of the *Salmonella* Bareilly strain causing a Foodborne Outbreak in the United States. *The Journal of Infectious Diseases*.

Hofshagen, M. & Kruse, H. (2005). Reduction in Flock Prevalence of *Campylobacter* spp. in Broilers in Norway after Implementation of an Action Plan. *Journal of Food Protection*. 10. p.pp. 2220–2223.

Holden, M.T.G. *et al.* (2013). A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Research*. 23 (4). p.pp. 653–664.

Holt, K.E. *et al.* (2012). *Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nature Genetics*. 44 (9). p.pp. 1056–9.

Hoosain, N. & Lastovica, A.J. (2009). An evaluation of the Oxoid Biochemical Identification System Campy rapid screening test for *Campylobacteraceae* and *Helicobacter* spp. *Letters in Applied Microbiology*. 48 (6). p.pp. 675–9.

Hooton, S.P.T. & Connerton, I.F. (2014). *Campylobacter jejuni* acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. *Frontiers in Microbiology*. 5. p.p. 744.

Hopkins, K.L. *et al.* (2004). Fluorescent amplified fragment length polymorphism

genotyping of *Campylobacter jejuni* and *Campylobacter coli* strains and its relationship with host specificity, serotyping, and phage typing. *Journal of Clinical Microbiology*. 42 (1). p.pp. 229–35.

Househam, K.C. *et al.* (1990). Factors influencing the duration of acute diarrheal disease in infancy. *Journal of Pediatric Gastroenterology and Nutrition*. 10 (1). p.pp. 37–40.

Househam, K.C., Mann, M.D. & Bowie, M.D. (1988). Enteropathogens associated with acute infantile diarrhoea in Cape Town. *South African Medical Journal*. 73 (2). p.pp. 83–7.

Hu, L. & Kopecko, D.J. (1999). *Campylobacter jejuni* 81-176 associates with microtubules and dynein during invasion of human intestinal cells. *Infection and Immunity*. 67 (8). p.pp. 4171–82.

Hudson, R.R., Slatkin, M. & Maddison, W.P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*. 132 (2). p.pp. 583–9.

Hugdahl, M.B., Beery, J.T. & Doyle, M.P. (1988). Chemotactic behavior of *Campylobacter jejuni*. *Infection and Immunity*. 56 (6). p.pp. 1560–6.

Humphrey, S. *et al.* (2014). *Campylobacter jejuni* is not merely a commensal in commercial broiler chickens and affects bird welfare. *mBio*. 5 (4). p.pp. e01364–14.

Humphrey, T., O'Brien, S. & Madsen, M. (2007). Campylobacters as zoonotic pathogens: a food production perspective. *International Journal of Food Microbiology*. 117 (3). p.pp. 237–257.

Humphrey, T.J. *et al.* (2005). Prevalence and subtypes of ciprofloxacin-resistant *Campylobacter* spp. in commercial poultry flocks before, during, and after treatment with fluoroquinolones. *Antimicrobial Agents and Chemotherapy*. 49 (2). p.pp. 690–8.

Humphrey, T.J., Henley, A. & Lanning, D.G. (1993). The colonization of broiler chickens with *Campylobacter jejuni*: some epidemiological investigations. *Epidemiology and Infection*. 110 (3). p.pp. 601–7.

Hunter, P.R. & Gaston, M.A. (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *Journal of Clinical Microbiology*. 26 (11). p.pp. 2465–6.

Huson, D.H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*. 23 (2). p.pp. 254–67.

Inglis, G.D. *et al.* (2007). *Campylobacter canadensis* sp. *nov*., from captive whooping cranes in Canada. *International Journal of Systematic and Evolutionary Microbiology*. 57 (Pt 11). p.pp. 2636–44.

Iraola, G. *et al.* (2014). Genomic evidence for the emergence and evolution of pathogenicity and niche preferences in the genus *Campylobacter*. *Genome Biology and Evolution*. 6 (9). p.pp. 2392–405.

Jain, M. *et al.* (2015). Improved data analysis for the MinION nanopore sequencer. *Nature Methods*. 12 (4). p.pp. 351–356.

Janssen, A.W.F. & Kersten, S. (2015). The role of the gut microbiota in metabolic health. *The FASEB Journal*. 29 (8). p.pp. 3111–3123.

Jin, S. *et al.* (2001). JlpA, a novel surface-exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial cells. *Molecular Microbiology*. 39 (5). p.pp. 1225–36.

Johnsen, G., Kruse, H. & Hofshagen, M. (2006). Genotyping of *Campylobacter jejuni* from broiler carcasses and slaughterhouse environment by amplified fragment length polymorphism. *Poultry Science*. 85 (12). p.pp. 2278–84.

Johnson, J.G., Livny, J. & Dirita, V.J. (2014). High-throughput sequencing of *Campylobacter jejuni* insertion mutant libraries reveals *mapA* as a fitness factor for chicken colonization. *Journal of Bacteriology*. 196 (11). p.pp. 1958–67.

Jolley, K.A. *et al.* (2012a). Resolution of a meningococcal disease outbreak from whole-genome sequence data with rapid web-based analysis methods. *Journal of Clinical Microbiology*. 50 (9). p.pp. 3046–3053.

Jolley, K.A. *et al.* (2012b). Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology*. 158 (Pt 4). p.pp. 1005–1015.

Jolley, K.A. & Maiden, M.C.J. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*. 11. p.p. 595.

Jones, M.A. *et al.* (2004). Adaptation of *Campylobacter jejuni* NCTC11168 to high-level colonization of the avian gastrointestinal tract. *Infection and Immunity*. 72 (7). p.pp. 3769–76.

Jonker, A. & Picard, J.A. (2010). Antimicrobial susceptibility in thermophilic *Campylobacter* species isolated from pigs and chickens in South Africa. *Journal of the South African Veterinary Association*. 81 (4). p.pp. 228–36.

Kaakoush, N.O. *et al.* (2015). Global Epidemiology of *Campylobacter* Infection. *Clinical Microbiology Reviews*. 28 (3). p.pp. 687–720.

Kaakoush, N.O. *et al.* (2011). The pathogenic potential of *Campylobacter concisus* strains associated with chronic intestinal diseases. *PloS One*. 6 (12). p.p. e29045.

Kalupahana, R.S. *et al.* (2013). Colonization of *Campylobacter* spp. in broiler chickens and laying hens reared in tropical climates with low-biosecurity

housing. *Applied and Environmental Microbiology*. 79 (1). p.pp. 393–5.

Kapperud, G. *et al.* (2003). Factors associated with increased and decreased risk of *Campylobacter* infection: a prospective case-control study in Norway. *American Journal of Epidemiology*. 158 (3). p.pp. 234–42.

Karlyshev, A. V *et al.* (2000). Genetic and biochemical evidence of a *Campylobacter jejuni* capsular polysaccharide that accounts for Penner serotype specificity. *Molecular Microbiology*. 35 (3). p.pp. 529–41.

Katoh, K. & Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*. 30 (4). p.pp. 772–80.

Kau, A.L. *et al.* (2011). Human nutrition, the gut microbiome and the immune system. *Nature*. 474 (7351). p.pp. 327–36.

Kaur, T. *et al.* (2011). *Campylobacter troglodytis* sp. *nov*., isolated from feces of human-habituated wild chimpanzees (*Pan troglodytes schweinfurthii*) in Tanzania. *Applied and Environmental Microbiology*. 77 (7). p.pp. 2366–73.

Keithlin, J. *et al.* (2014). Systematic review and meta-analysis of the proportion of *Campylobacter* cases that develop chronic sequelae. *BMC Public Health*. 14 (1). p.p. 1203.

Kervella, M. *et al.* (1993). Isolation and characterization of two *Campylobacter* glycine-extracted proteins that bind to HeLa cell membranes. *Infection and Immunity*. 61 (8). p.pp. 3440–8.

Kienesberger, S. *et al.* (2014). Comparative genome analysis of *Campylobacter fetus* subspecies revealed horizontally acquired genetic elements important for virulence and niche specificity. *PloS One*. 9 (1). p.p. e85491.

Kim, S.A. *et al.* (2009). Eight enrichment broths for the isolation of *Campylobacter jejuni* from inoculated suspensions and ground pork. *Letters in Applied Microbiology*. 49 (5). p.pp. 620–6.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*. 16 (2). p.pp. 111–20.

Kinana, A.D. *et al.* (2006). Genetic diversity and quinolone resistance in *Campylobacter jejuni* isolates from poultry in Senegal. *Applied and Environmental Microbiology*. 72 (5). p.pp. 3309–3313.

Kittl, S. *et al.* (2011). Comparison of genotypes and antibiotic resistance of *Campylobacter jejuni* isolated from humans and slaughtered chickens in Switzerland. *Journal of Applied Microbiology*. 110 (2). p.pp. 513–520.

Kittl, S. *et al.* (2013). Source Attribution of Human *Campylobacter* Isolates by MLST

and Fla-Typing and Association of Genotypes with Quinolone Resistance Y.-F. Chang (ed.). *PLoS One*. 8 (11). p.p. e81796.

Kivistö, R.I. *et al.* (2014). Evolution and comparative genomics of *Campylobacter jejuni* ST-677 clonal complex. *Genome Biology and Evolution*. 6 (9). p.pp. 2424–38.

Klappenbach, J.A. *et al.* (2001). rrndb: the Ribosomal RNA Operon Copy Number Database. *Nucleic Acids Research*. 29 (1). p.pp. 181–4.

Klein, G. *et al.* (2007). Predominant strains of thermophilic *Campylobacter* spp. in a German poultry slaughterhouse. *International Journal of Food Microbiology*. 117 (3). p.pp. 324–8.

Koene, M.G.J. *et al.* (2009). Strain variation within *Campylobacter* species in fecal samples from dogs and cats. *Veterinary Microbiology*. 133 (1-2). p.pp. 199–205.

Koga, M. *et al.* (2005). Antecedent infections in Fisher syndrome: a common pathogenesis of molecular mimicry. *Neurology*. 64 (9). p.pp. 1605–11.

Kokotovic, B. & On, S.L. (1999). High-resolution genomic fingerprinting of *Campylobacter jejuni* and *Campylobacter coli* by analysis of amplified fragment length polymorphisms. *FEMS Microbiology Letters*. 173 (1). p.pp. 77–84.

Komba, E.V.G. *et al.* (2015). Prevalence, Antimicrobial Resistance and Risk Factors for Thermophilic *Campylobacter* Infections in Symptomatic and Asymptomatic Humans in Tanzania. *Zoonoses and Public Health*.

Konkel, M.E. *et al.* (1997). Identification and molecular cloning of a gene encoding a fibronectin-binding protein (CadF) from *Campylobacter jejuni*. *Molecular Microbiology*. 24 (5). p.pp. 953–63.

Koornhof, H.J. *et al.* (1986). Bacterial enteritis of infancy and childhood in Soweto. *South African Medical Journal*. Suppl. p.pp. 50–4.

Korber, B. (2000). HIV Signature and Sequence Variation Analysis. In: A. G. Rodrigo & G. H. Learn (eds.). *Computational Analysis of HIV Molecular Sequences*. Dordrecht, Netherlands: Kluwer Academic Publisher, pp. 55–72.

Köser, C.U. *et al.* (2012a). Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *The New England Journal of Medicine*. 366 (24). p.pp. 2267–75.

Köser, C.U. *et al.* (2012b). Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathogens*. 8 (8). p.p. e1002824.

Köser, C.U., Ellington, M.J. & Peacock, S.J. (2014). Whole-genome sequencing to control antimicrobial resistance. *Trends in Genetics : TIG*. 30 (9). p.pp. 401–7.

Koziel, M. *et al.* (2014a). *Campylobacter corcagiensis* sp. *nov*., isolated from faeces of captive lion-tailed macaques (*Macaca silenus*). *International Journal of Systematic and Evolutionary Microbiology*. 64 (Pt 8). p.pp. 2878–83.

Koziel, M. *et al.* (2014b). Draft Genome Sequence of *Campylobacter corcagiensis* Strain CIT045T, a Representative of a Novel *Campylobacter* Species Isolated from Lion-Tailed Macaques (*Macaca silenus*). *Genome Announcements*. 2 (2).

Koziel, M. *et al.* (2013). Improved detection of bacterial pathogens in patients presenting with gastroenteritis by use of the EntericBio real-time Gastro Panel I assay. *Journal of Clinical Microbiology*. 51 (8). p.pp. 2679–85.

Ku, B.K. *et al.* (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 (3). p.pp. 381–6.

Kulkarni, S.P. (2002). Detection of *Campylobacter* species: a comparison of culture and polymerase chain reaction based methods. *Journal of Clinical Pathology*. 55 (10). p.pp. 749–753.

Kunene, N.F., Hastings, J.W. & von Holy, A. (1999). Bacterial populations associated with a sorghum-based fermented weaning cereal. *International Journal of Food Microbiology*. 49 (1-2). p.pp. 75–83.

Kurtz, S. *et al.* (2004). Versatile and open software for comparing large genomes. *Genome Biology*. 5 (2). p.p. R12.

Kuusi, M. *et al.* (2005). A large outbreak of campylobacteriosis associated with a municipal water supply in Finland. *Epidemiology and Infection*. 133 (4). p.pp. 593–601.

Kuusi, M. (2004). An outbreak of gastroenteritis from a non-chlorinated community water supply. *Journal of Epidemiology & Community Health*. 58 (4). p.pp. 273–277.

Kwan, P.S.L. *et al.* (2008). Longitudinal study of the molecular epidemiology of *Campylobacter jejuni* in cattle on dairy farms. *Applied and Environmental Microbiology*. 74 (12). p.pp. 3626–33.

Lan, R. & Reeves, P.R. (1996). Gene transfer is a major factor in bacterial evolution. *Molecular biology and evolution*. 13 (1). p.pp. 47–55.

Lan, R. & Reeves, P.R. (2000). Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends in Microbiology*. 8 (9). p.pp. 396–401.

Lan, R. & Reeves, P.R. (2001). When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends in Microbiology*. 9 (9). p.pp. 419–24.

Land, M. *et al.* (2015). Insights from 20 years of bacterial genome sequencing. *Functional & Integrative Genomics*. 15 (2). p.pp. 141–61.

Langridge, G.C. *et al.* (2009). Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Research*. 19 (12). p.pp. 2308–16.

Lara-Tejero, M. & Galán, J.E. (2000). A bacterial toxin that controls cell cycle progression as a deoxyribonuclease I-like protein. *Science*. 290 (5490). p.pp. 354–7.

Lara-Tejero, M. & Galán, J.E. (2001). CdtA, CdtB, and CdtC form a tripartite complex that is required for cytolethal distending toxin activity. *Infection and Immunity*. 69 (7). p.pp. 4358–65.

Larsen, M. V *et al.* (2014). Benchmarking of methods for genomic taxonomy. *Journal of Clinical Microbiology*. 52 (5). p.pp. 1529–39.

Lasken, R.S. & McLean, J.S. (2014). Recent advances in genomic DNA sequencing of microbial species from single cells. *Nature Reviews Genetics*. 15 (9). p.pp. 577–84.

Lastovica, A.J. *et al.* (1986). Distribution of sero-biotypes of *Campylobacter jejuni* and *C. coli* isolated from paediatric patients. *Journal of Medical Microbiology*. 21 (1). p.pp. 1–5.

Lastovica, A.J. (2006). Emerging *Campylobacter* spp.: The tip of the iceberg. *Clinical Microbiology Newsletter*. 28 (7). p.pp. 49–56.

Lastovica, A.J. & Allos, B.. (2008). Clinical significance of *Campylobacter* and related species other than *Campylobacter jejuni* and *Campylobacter coli*. In: I. Nachamkin, C. M. Szymanski, & M. J. Blaser (eds.). *Campylobacter*. Washington, D.C.: ASM Press, pp. 123–149.

Lastovica, A.J., Goddard, E.A. & Argent, A.C. (1997). Guillain-Barré syndrome in South Africa associated with *Campylobacter jejuni* O:41 strains. *The Journal of Infectious Diseases*. 176 Suppl . p.pp. S139–43.

Lastovica, A.J. & le Roux, E. (2000). Efficient isolation of campylobacteria from stools. *Journal of Clinical Microbiology*. 38 (7). p.pp. 2798–9.

Lawes, J.R. *et al.* (2012). Investigation of prevalence and risk factors for *Campylobacter* in broiler flocks at slaughter: results from a UK survey. *Epidemiology and Infection*. 140 (10). p.pp. 1725–37.

Lawson, A.J. *et al.* (2001). *Campylobacter hominis* sp. *nov*., from the human gastrointestinal tract. *International Journal of Systematic and Evolutionary Microbiology*. 51 (Pt 2). p.pp. 651–60.

Lawson, A.J. *et al.* (1997). Polymerase chain reaction detection and speciation of

Campylobacter upsaliensis and *C. helveticus* in human faeces and comparison with culture techniques. *Journal of Applied Microbiology*. 83 (3). p.pp. 375–80.

Lechner, M. *et al.* (2014). Orthology detection combining clustering and synteny for very large datasets. *PloS One*. 9 (8). p.p. e105015.

Lechner, M. *et al.* (2011). Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics*. 12. p.p. 124.

Lee, G. *et al.* (2013). Symptomatic and asymptomatic *Campylobacter* infections associated with reduced growth in Peruvian children. *PLoS Neglected Tropical Diseases*. 7 (1). p.p. e2036.

Lefébure, T. *et al.* (2010). Evolutionary dynamics of complete *Campylobacter* pan-genomes and the bacterial species concept. *Genome Biology and Evolution*. 2 (0). p.pp. 646–55.

Lefébure, T. & Stanhope, M.J. (2009). Pervasive, genome-wide positive selection leading to functional divergence in the bacterial genus *Campylobacter*. *Genome Research*. 19 (7). p.pp. 1224–32.

Lepage, P. *et al.* (2013). A metagenomic insight into our gut's microbiome. *Gut*. 62 (1). p.pp. 146–58.

Li, L., Stoeckert, C.J. & Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*. 13 (9). p.pp. 2178–89.

Li, R. *et al.* (2010). *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Research*. 20 (2). p.pp. 265–72.

Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 25 (11). p.pp. 1451–2.

Lienau, J.-A., Ellerbroek, L. & Klein, G. (2007). Tracing flock-related *Campylobacter* clones from broiler farms through slaughter to retail products by pulsed-field gel electrophoresis. *Journal of Food Protection*. 70 (3). p.pp. 536–42.

Lindmark, H. *et al.* (2006). Distribution of *Campylobacter* genotypes on broilers during slaughter. *Journal of Food Protection*. 69 (12). p.pp. 2902–7.

Lior, H. *et al.* (1982). Serotyping of *Campylobacter jejuni* by slide agglutination based on heat-labile antigenic factors. *Journal of Clinical Microbiology*. 15 (5). p.pp. 761–8.

Litrup, E., Torpdahl, M. & Nielsen, E.M. (2007). Multilocus sequence typing performed on *Campylobacter coli* isolates from humans, broilers, pigs and cattle originating in Denmark. *Journal of Applied Microbiology*. 103 (1). p.pp. 210–8.

Liu, J. *et al.* (2014). Development and assessment of molecular diagnostic tests for

15 enteropathogens causing childhood diarrhoea: a multicentre study. *Lancet Infectious Diseases*. 14 (8). p.pp. 716–24.

Liu, L. *et al.* (2012). Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet*. 379 (9832). p.pp. 2151–61.

Logan, J. *et al.* (2000). *Campylobacter lanienae* sp. *nov*., a new species isolated from workers in an abattoir. *International Journal of Systematic and Evolutionary Microbiology*. 50 (2). p.pp. 865–872.

Loman, N.J. *et al.* (2012a). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology*. 10 (9). p.pp. 599–606.

Loman, N.J. *et al.* (2012b). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*. 30 (5). p.pp. 434–9.

Loman, N.J. & Watson, M. (2015). Successful test launch for nanopore sequencing. *Nature Methods*. 12 (4). p.pp. 303–4.

Lorntz, B. *et al.* (2006). Early childhood diarrhea predicts impaired school performance. *The Pediatric Infectious Disease Journal*. 25 (6). p.pp. 513–20.

Lozupone, C.A. *et al.* (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*. 489 (7415). p.pp. 220–30.

Luangtongkum, T. *et al.* (2009). Antibiotic resistance in *Campylobacter*: emergence, transmission and persistence. *Future Microbiology*. 4 (2). p.pp. 189–200.

M'ikanatha, N.M. *et al.* (2012). Culturing stool specimens for *Campylobacter* spp., Pennsylvania, USA. *Emerging Infectious Diseases*. 18 (3). p.pp. 484–7.

MacKenzie, K.D. *et al.* (2015). Bistable expression of CsgD in *Salmonella enterica* serovar Typhimurium connects virulence to persistence. *Infection and Immunity*. 83 (6). p.pp. 2312–26.

Madikizela, B., Aderogba, M.A. & Van Staden, J. (2013). Isolation and characterization of antimicrobial constituents of *Searsia chirindensis L.* (*Anacardiaceae*) leaf extracts. *Journal of Ethnopharmacology*. 150 (2). p.pp. 609–13.

Maiden, M.C. *et al.* (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*. 95 (6). p.pp. 3140–5.

Maiden, M.C.J. *et al.* (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology*. 11 (10). p.pp. 728–36.

Maiden, M.C.J. (2006). Multilocus sequence typing of bacteria. *Annual Review of Microbiology*. 60. p.pp. 561–88.

Maiden, M.C.J. & Dingle, K.E. (2008). Population Biology of *Campylobacter jejuni* and Related Organisms. In: I. Nachamkin, C. M. Szymanski, & M. J. Blaser (eds.). *Campylobacter*. Washington, D.C.: ASM Press, pp. 27–40.

Malnick, S. & Melzer, E. (2015). Human microbiome: From the bathroom to the bedside. *World Journal of Gastrointestinal Pathophysiology*. 6 (3). p.pp. 79–85.

Man, S.M. (2011). The clinical importance of emerging *Campylobacter* species. *Nature reviews Gastroenterology & hepatology*. 8 (12). p.pp. 669–685.

Mandrell, R.E. *et al.* (2005). Speciation of *Campylobacter coli*, *C. jejuni*, *C. helveticus*, *C. lari*, *C. sputorum*, and *C. upsaliensis* by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Applied and Environmental Microbiology*. 71 (10). p.pp. 6292–307.

Mangen, M.-J.J. *et al.* (2015). Cost-of-illness and disease burden of food-related pathogens in the Netherlands, 2011. *International Journal of Food Microbiology*. 196. p.pp. 84–93.

Manning, G. *et al.* (2001). Evidence for a genetically stable strain of *Campylobacter jejuni*. *Applied and Environmental Microbiology*. 67 (3). p.pp. 1185–9.

Manning, G. *et al.* (2003). Multilocus sequence typing for comparison of veterinary and human isolates of *Campylobacter jejuni*. *Applied and Environmental Microbiology*. 69 (11). p.pp. 6370–9.

Manyi-Loh, C.E. *et al.* (2014). Inactivation of selected bacterial pathogens in dairy cattle manure by mesophilic anaerobic digestion (balloon type digester). *International Journal of Environmental Research and Public Health*. 11 (7). p.pp. 7184–94.

Mardis, E.R. (2011). A decade's perspective on DNA sequencing technology. *Nature*. 470 (7333). p.pp. 198–203.

Mardis, E.R. (2008). Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*. 9. p.pp. 387–402.

Mardis, E.R. (2013). Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*. 6. p.pp. 287–303.

Margulies, M. *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 437 (7057). p.pp. 376–80.

Marsh, M. *et al.* (1997). High-throughput DNA sequencing on a capillary array electrophoresis system. *Journal of Capillary Electrophoresis*. 4 (2). p.pp. 83–9.

Martiny, D. *et al.* (2011). Accuracy of the API Campy system, the Vitek 2 *Neisseria*-

Haemophilus card and matrix-assisted laser desorption ionization time-of-flight mass spectrometry for the identification of *Campylobacter* and related organisms. *Clinical Microbiology and Infection*. 17 (7). p.pp. 1001–6.

Mason, J. *et al.* (2013). *Campylobacter* infection in children in Malawi is common and is frequently associated with enteric virus co-infections. *PloS One*. 8 (3). p.p. e59663.

Mather, A.E. *et al.* (2013). Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science*. 341 (6153). p.pp. 1514–7.

Matsheka, M.I. *et al.* (2002). Genetic heterogeneity of *Campylobacter concisus* determined by pulsed field gel electrophoresis-based macrorestriction profiling. *FEMS Microbiology Letters*. 211 (1). p.pp. 17–22.

Matsheka, M.I. *et al.* (2006). The use of (GTG)5 oligonucleotide as an RAPD primer to type *Campylobacter concisus*. *Letters in Applied Microbiology*. 42 (6). p.pp. 600–5.

Matsheka, M.I., Lastovica, A.J. & Elisha, B.G. (2001). Molecular identification of *Campylobacter concisus*. *Journal of Clinical Microbiology*. 39 (10). p.pp. 3684–9.

Mayosi, B.M. *et al.* (2012). Health in South Africa: changes and challenges since 2009. *Lancet*. 380 (9858). p.pp. 2029–43.

Mayosi, B.M. & Benatar, S.R. (2014). Health and health care in South Africa--20 years after Mandela. *The New England Journal of Medicine*. 371 (14). p.pp. 1344–53.

McAdam, P.R. *et al.* (2012). Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences of the United States of America*. 109 (23). p.pp. 9107–12.

McAdam, P.R., Richardson, E.J. & Fitzgerald, J.R. (2014). High-throughput sequencing for the study of bacterial pathogen biology. *Current Opinion in Microbiology*. 19. p.pp. 106–13.

McAuliffe, G.N. *et al.* (2013). Systematic application of multiplex PCR enhances the detection of bacteria, parasites, and viruses in stool samples. *The Journal of Infection*. 67 (2). p.pp. 122–9.

McCarthy, N.D. *et al.* (2007). Host-associated genetic import in *Campylobacter jejuni*. *Emerging Infectious Diseases*. 13 (2). p.pp. 267–72.

McSweegan, E. & Walker, R.I. (1986). Identification and characterization of two *Campylobacter jejuni* adhesins for cellular and mucous substrates. *Infection*

and Immunity. 53 (1). p.pp. 141–8.

Meinersmann, R.J. *et al.* (1997). Discrimination of *Campylobacter jejuni* isolates by *fla* gene sequencing. *Journal of Clinical Microbiology*. 35 (11). p.pp. 2810–4.

Mellmann, A. *et al.* (2011). Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PloS One*. 6 (7). p.p. e22751.

Mellmann, A. *et al.* (2004). Sequence-based typing of *flaB* is a more stable screening tool than typing of *flaA* for monitoring of *Campylobacter* populations. *Journal of Clinical Microbiology*. 42 (10). p.pp. 4840–2.

Mengelle, C. *et al.* (2013). Simultaneous detection of gastrointestinal pathogens with a multiplex Luminex-based molecular assay in stool samples from diarrhoeic patients. *Clinical Microbiology and Infection*. 19 (10). p.pp. E458–65.

Méric, G. *et al.* (2014). A reference pan-genome approach to comparative bacterial genomics: identification of novel epidemiological markers in pathogenic *Campylobacter*. *PloS One*. 9 (3). p.p. e92798.

Messens, W. *et al.* (2009). Multiple typing for the epidemiological study of contamination of broilers with thermotolerant *Campylobacter*. *Veterinary Microbiology*. 138 (1-2). p.pp. 120–31.

Metzker, M.L. (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics*. 11 (1). p.pp. 31–46.

Mickan, L. *et al.* (2007). Multilocus sequence typing of *Campylobacter jejuni* isolates from New South Wales, Australia. *Journal of Applied Microbiology*. 102 (1). p.pp. 144–52.

Mikheyev, A.S. & Tin, M.M.Y. (2014). A first look at the Oxford Nanopore MinION sequencer. *Molecular Ecology Resources*. 14 (6).

Miller, C.E., Williams, P.H. & Ketley, J.M. (2009). Pumping iron: mechanisms for iron uptake by *Campylobacter*. *Microbiology*. 155 (Pt 10). p.pp. 3157–65.

Miller, W.G. *et al.* (2014a). Comparative genomics of the *Campylobacter lari* group. *Genome Biology and Evolution*. 6 (12). p.pp. 3252–66.

Miller, W.G. *et al.* (2005). Extended multilocus sequence typing system for *Campylobacter coli*, *C. lari*, *C. upsaliensis*, and *C. helveticus*. *Journal of Clinical Microbiology*. 43 (5). p.pp. 2315–29.

Miller, W.G. *et al.* (2012). Multilocus Sequence Typing Methods for the Emerging *Campylobacter* Species *C. hyointestinalis*, *C. lanienae*, *C. sputorum*, *C. concisus*, and *C. curvus*. *Frontiers in Cellular and Infection Microbiology*. 2. p.p. 45.

Miller, W.G. *et al.* (2008). The complete genome sequence and analysis of the human pathogen *Campylobacter lari*. *Foodborne Pathogens and Disease*. 5 (4). p.pp. 371–86.

Miller, W.G. *et al.* (2014b). Use of an improved *atpA* amplification and sequencing method to identify members of the *Campylobacteraceae* and *Helicobacteraceae*. *Letters in Applied Microbiology*. 58 (6). p.pp. 582–90.

Modi, S.R. *et al.* (2013). Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature*. 499 (7457). p.pp. 219–22.

Molbak, K. & Havelaar, A. (2008). *Burden of illness of campylobacteriosis and sequelae*. In: I. Nachamkin, C. M. Symanski, & M. J. Blaser (eds.). Washington, D.C.: ASM Press, pp. 151–162.

Monteville, M.R. & Konkel, M.E. (2002). Fibronectin-facilitated invasion of T84 eukaryotic cells by *Campylobacter jejuni* occurs preferentially at the basolateral cell surface. *Infection and Immunity*. 70 (12). p.pp. 6665–71.

Monteville, M.R., Yoon, J.E. & Konkel, M.E. (2003). Maximal adherence and invasion of INT 407 cells by *Campylobacter jejuni* requires the CadF outer-membrane protein and microfilament reorganization. *Microbiology*. 149 (Pt 1). p.pp. 153–65.

Moore, J.E. *et al.* (2006). The epidemiology of antibiotic resistance in *Campylobacter*. *Microbes and Infection*. 8 (7). p.pp. 1955–1966.

Mosupye, F.M. & von Holy, A. (2000). Microbiological hazard identification and exposure assessment of street food vending in Johannesburg, South Africa. *International Journal of Food Microbiology*. 61 (2-3). p.pp. 137–45.

Mosupye, F.M. & von Holy, A. (1999). Microbiological quality and safety of ready-to-eat street-vended foods in Johannesburg, South Africa. *Journal of Food Protection*. 62 (11). p.pp. 1278–84.

Mughini-Gras, L. *et al.* (2012). Risk factors for campylobacteriosis of chicken, ruminant, and environmental origin: a combined case-control and source attribution analysis. *PloS One*. 7 (8). p.p. e42599.

Mughini-Gras, L. *et al.* (2014). Campylobacteriosis in returning travellers and potential secondary transmission of exotic strains. *Epidemiology and Infection*. 142 (6). p.pp. 1277–88.

Mullner, P. *et al.* (2009a). Assigning the source of human campylobacteriosis in New Zealand: a comparative genetic and epidemiological approach. *Infection, Genetics and Evolution*. 9 (6). p.pp. 1311–9.

Mullner, P. *et al.* (2009b). Source attribution of food-borne zoonoses in New

Zealand: a modified Hald model. *Risk Analysis : An Official Publication of the Society for Risk Analysis*. 29 (7). p.pp. 970–84.

Murray, C.J.L. *et al.* (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 380 (9859). p.pp. 2197–223.

Mutreja, A. *et al.* (2011). Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature*. 477 (7365). p.pp. 462–5.

Myers, A.L., Jackson, M.A. & Selvarangan, R. (2011). False-positive results of *Campylobacter* rapid antigen testing. *The Pediatric Infectious Disease Journal*. 30 (6). p.p. 542.

Nachamkin, I. *et al.* (2001). Molecular population genetic analysis of *Campylobacter jejuni* HS:19 associated with Guillain-Barré syndrome and gastroenteritis. *The Journal of Infectious Diseases*. 184 (2). p.pp. 221–6.

Nadeau, E., Messier, S. & Quessy, S. (2002). Prevalence and comparison of genetic profiles of *Campylobacter* strains isolated from poultry and sporadic cases of campylobacteriosis in humans. *Journal of Food Protection*. 65 (1). p.pp. 73–8.

Neimann, J. *et al.* (2003). A case-control study of risk factors for sporadic *Campylobacter* infections in Denmark. *Epidemiology and Infection*. 130 (3). p.pp. 353–66.

Nel, E.. (2010). Diarrhoea and malnutrition. *South African Journal of Clinical Nutrition*. 23 (1).

Newell, D.G. *et al.* (2010). Food-borne diseases - the challenges of 20 years ago still persist while new ones continue to emerge. *International Journal of Food Microbiology*. 139 Suppl . p.pp. S3–15.

Newell, D.G. & Fearnley, C. (2003). Sources of *Campylobacter* colonization in broiler chickens. *Applied and Environmental Mcrobiology*. 69 (8). p.pp. 4343–4351.

Ng, L.K. *et al.* (1985). Morphological forms and viability of *Campylobacter* species studied by electron microscopy. *Journal of Bacteriology*. 164 (1). p.pp. 338–43.

Nichols, G.L. *et al.* (2012). *Campylobacter* epidemiology: a descriptive study reviewing 1 million cases in England and Wales between 1989 and 2011. *BMJ Open*. 2 (4). p.pp. 10.1136/bmjopen–2012–001179.

Njiro, S.M. *et al.* (2011). A study of some infectious causes of reproductive disorders in cattle owned by resource-poor farmers in Gauteng Province, South Africa. *Journal of the South African Veterinary Association*. 82 (4). p.pp. 213–8.

O'Brien, S.J. *et al.* (2010). Methods for determining disease burden and calibrating national surveillance data in the United Kingdom: the second study of infectious

intestinal disease in the community (IID2 study). *BMC Medical Research Methodology*. 10 (1). p.p. 39.

O'Leary, A.M. *et al.* (2011). Pulsed field gel electrophoresis typing of human and retail foodstuff Campylobacters: an Irish perspective. *Food Microbiology*. 28 (3). p.pp. 426–33.

Oberhelman, R.A. & Taylor, D.N. (2000). Campylobacter *infections in developing countries*. In: I. Nachamkin & M. J. Blaser (eds.). *Campylobacter*. Washington, D.C.: ASM Press, pp. 139 – 153.

Obi, C.L. *et al.* (2003). Scope of potential bacterial agents of diarrhoea and microbial assessment of quality of river water sources in rural Venda communities in South Africa. *Water Science and Technology*. 47 (3). p.pp. 59–64.

Obi, C.L. & Bessong, P.O. (2002). Diarrhoeagenic bacterial pathogens in HIV-positive patients with diarrhoea in rural communities of Limpopo Province, South Africa. *Journal of Health, Population, and Nutrition*. 20 (3). p.pp. 230–234.

Odendaal, M.W. *et al.* (1994). First isolation of *Campylobacter jejuni* from the vaginal discharge of three bitches after abortion in South Africa. *The Onderstepoort Journal of Veterinary Research*. 61 (2). p.pp. 193–5.

Oelschlaeger, T.A., Guerry, P. & Kopecko, D.J. (1993). Unusual microtubule-dependent endocytosis mechanisms triggered by *Campylobacter jejuni* and *Citrobacter freundii*. *Proceedings of the National Academy of Sciences of the United States of America*. 90 (14). p.pp. 6884–8.

Okoro, C.K. *et al.* (2012). Intracontinental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. *Nature Genetics*. 44 (11). p.pp. 1215–21.

Oldfield, N.J. *et al.* (2016). Genomic Analysis of Serogroup Y *Neisseria meningitidis* Isolates Reveals Extensive Similarities Between Carriage and Disease-Associated Organisms. *The Journal of Infectious Diseases*. 213 (11). p.pp. 1777-1785.

On, S.L.. *et al.* (2008). Molecular Epidemiology of *Campylobacter* Species. In: I. Nachamkin, C. M. Szymanski, & M. J. Blaser (eds.). *Campylobacter*. Washington, D.C.: ASM Press, pp. 191–211.

On, S.L.W. *et al.* (1995). *Campylobacter hyointestinalis* subsp. *lawsonii* subsp. *nov*., Isolated from the Porcine Stomach, and an Emended Description of *Campylobacter hyointestinalis*. *International Journal of Systematic Bacteriology*. 45 (4). p.pp. 767–774.

On, S.L.W. (2013). Isolation, identification and subtyping of *Campylobacter*: where to from here? *Journal of Microbiological Methods*. 95 (1). p.pp. 3–7.

On, S.L.W. *et al.* (2013). PCR revisited: a case for revalidation of PCR assays for microorganisms using identification of *Campylobacter* species as an exemplar. *Quality Assurance and Safety of Crops and Foods*. 5 (1). p.pp. 49–62.

On, S.L.W. & Jordan, P.J. (2003). Evaluation of 11 PCR assays for species-level identification of *Campylobacter jejuni* and *Campylobacter coli*. *Journal of Clinical Microbiology*. 41 (1). p.pp. 330–6.

Oporto, B. *et al.* (2011). Genetic diversity among *Campylobacter jejuni* isolates from healthy livestock and their links to human isolates in Spain. *Zoonoses and Public Health*. 58 (5). p.pp. 365–75.

Oster, R.J. *et al.* (2014). Bacterial pathogen gene abundance and relation to recreational water quality at seven Great Lakes beaches. *Environmental Science & Technology*. 48 (24). p.pp. 14148–57.

Owen, R.J. *et al.* (1990). Identification of catalase negative/weak *Campylobacter jejuni* from human blood and faecal cultures by numerical analysis of electrophoretic protein patterns. *FEMS Microbiology Letters*. 57 (3). p.pp. 329–35.

Owen, R.J., Costas, M. & Sloss, L.L. (1988). Electrophoretic protein typing of *Campylobacter jejuni* subspecies 'doylei' (nitrate-negative campylobacter-like organisms) from human faeces and gastric mucosa. *European Journal of Epidemiology*. 4 (3). p.pp. 277–83.

Owen, R.J. & Hernandez, J. (1990a). Genotypic variation in '*Campylobacter upsaliensis'* from blood and faeces of patients in different countries. *FEMS Microbiology Letters*. 60 (1-2). p.pp. 5–10.

Owen, R.J. & Hernandez, J. (1990b). Occurrence of plasmids in '*Campylobacter upsaliensis'* (catalase negative or weak group) from geographically diverse patients with gastroenteritis or bacteraemia. *European Journal of Epidemiology*. 6 (2). p.pp. 111–7.

Palyada, K., Threadgill, D. & Stintzi, A. (2004). Iron acquisition and regulation in *Campylobacter jejuni*. *Journal of Bacteriology*. 186 (14). p.pp. 4714–29.

Park, S.F., Purdy, D. & Leach, S. (2000). Localized reversible frameshift mutation in the *flhA* gene confers phase variability to flagellin gene expression in *Campylobacter coli*. *Journal of Bacteriology*. 182 (1). p.pp. 207–10.

Parker, C.T. *et al.* (2007). Common genomic features of *Campylobacter jejuni* subsp. *doylei* strains distinguish them from C. *jejuni* subsp. *jejuni*. *BMC Microbiology*. 7 (50).

Parkhill, J. *et al.* (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature*. 403 (6770). p.pp. 665–8.

Parsons, B.N. *et al.* (2009). Typing of *Campylobacter jejuni* isolates from dogs by use of multilocus sequence typing and pulsed-field gel electrophoresis. *Journal of Clinical Microbiology*. 47 (11). p.pp. 3466–71.

Patton, C.M. *et al.* (1991). Evaluation of 10 methods to distinguish epidemic-associated *Campylobacter* strains. *Journal of Clinical Microbiology*. 29 (4). p.pp. 680–8.

Pearson, B.M. *et al.* (2013). Complete Genome Sequence of the *Campylobacter coli* Clinical Isolate 15-537360. *Genome Announcements*. 1 (6).

Pearson, B.M. *et al.* (2007). The complete genome sequence of *Campylobacter jejuni* strain 81116 (NCTC11828). *Journal of Bacteriology*. 189 (22). p.pp. 8402–3.

Pefanis, S.M. *et al.* (1988). Trichomoniasis and campylobacteriosis in bulls in the Republic of Transkei. *Journal of the South African Veterinary Association*. 59 (3). p.pp. 139–40.

Pei, Z. *et al.* (1998). Mutation in the peb1A locus of *Campylobacter jejuni* reduces interactions with epithelial cells and intestinal colonization of mice. *Infection and Immunity*. 66 (3). p.pp. 938–43.

Pei, Z. & Blaser, M.J. (1993). PEB1, the major cell-binding factor of *Campylobacter* jejuni, is a homolog of the binding component in gram-negative nutrient transport systems. *The Journal of Biological Chemistry*. 268 (25). p.pp. 18717–25.

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 (6). p.pp. 732–7.

Pepin, J. *et al.* (2005). Emergence of Fluoroquinolones as the Predominant Risk Factor for Clostridium difficile-Associated Diarrhea: A Cohort Study during an Epidemic in Quebec. *Clinical Infectious Diseases*. 41 (9). p.pp. 1254–1260.

Perkins, T.T. *et al.* (2013). ChIP-seq and transcriptome analysis of the OmpR regulon of *Salmonella enterica* serovars Typhi and Typhimurium reveals accessory genes implicated in host colonization. *Molecular Microbiology*. 87 (3). p.pp. 526–38.

Perry, M.D., Corden, S.A. & Howe, R.A. (2014). Evaluation of the Luminex xTAG Gastrointestinal Pathogen Panel and the Savyon Diagnostics Gastrointestinal Infection Panel for the detection of enteric pathogens in clinical samples. *Journal of Medical Microbiology*. 63 (Pt 11). p.pp. 1419–26.

Petersen, L. *et al.* (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 (7).

p.pp. 3115–21.

Pfeifer, B. *et al.* (2014). PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*. 31 (7). p.pp. 1929–36.

Pfeiffer, M.L., DuPont, H.L. & Ochoa, T.J. (2012). The patient presenting with acute dysentery--a systematic review. *The Journal of Infection*. 64 (4). p.pp. 374–86.

Platts-Mills, J.A. *et al.* (2014). Detection of *Campylobacter* in stool and determination of significance by culture, enzyme immunoassay, and PCR in developing countries. *Journal of clinical microbiology*. 52 (4). p.pp. 1074–80.

Platts-Mills, J.A. & Kosek, M. (2014). Update on the burden of *Campylobacter* in developing countries. *Current Opinion in Infectious Diseases*. 27 (5). p.pp. 444–450.

Platts-Mills, J.A., Liu, J. & Houpt, E.R. (2013). New concepts in diagnostics for infectious diarrhea. *Mucosal Immunology*. 6 (5). p.pp. 876–85.

Poly, F. *et al.* (2008). Characterization of two *Campylobacter jejuni* strains for use in volunteer experimental-infection studies. *Infection and Immunity*. 76 (12). p.pp. 5655–67.

Potgieter, N. *et al.* (2005). Bacterial contamination of Vhuswa--a local weaning food and stored drinking-water in impoverished households in the Venda region of South Africa. *Journal of Health, Population, and Nutrition*. 23 (2). p.pp. 150–5.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). *Inference of Population Structure Using Multilocus Genotype Data*. *Genetics* 155 (2). p.pp. 945-959.

Quail, M.A. *et al.* (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*. 13 (1). p.p. 341.

Quick, J. *et al.* (2015). Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biology*. 16 (1). p.p. 114.

Quick, J., Quinlan, A.R. & Loman, N.J. (2014). A reference bacterial genome dataset generated on the MinION(TM) portable single-molecule nanopore sequencer. *GigaScience*. 3 (1). p.p. 22.

Quiñones, B. *et al.* (2008). Comparative genomic analysis of clinical strains of *Campylobacter jejuni* from South Africa. *PloS One*. 3 (4). p.p. e2015.

R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rahman, A.E. *et al.* (2014). Childhood diarrhoeal deaths in seven low- and middle-income countries. *Bulletin of the World Health Organization*. 92 (9). p.pp. 664–

71.

Rao, M.R. *et al.* (2001). Pathogenicity and convalescent excretion of *Campylobacter* in rural Egyptian children. *American Journal of Epidemiology*. 154 (2). p.pp. 166–73.

Rapp, D. *et al.* (2012). Differences in the fecal concentrations and genetic diversities of *Campylobacter jejuni* populations among individual cows in two dairy herds. *Applied and Environmental Microbiology*. 78 (21). p.pp. 7564–71.

Ravel, A. *et al.* (2010) Description and burden of travel-related cases caused by enteropathogens reported in a Canadian community. *Journal of Travel Medicine*. 18 (1). p.pp. 8–19.

Regnath, T. & Ignatius, R. (2014). Accurate detection of *Campylobacter* spp. antigens by immunochromatography and enzyme immunoassay in routine microbiological laboratory. *European Journal of Microbiology & Immunology*. 4 (3). p.pp. 156–8.

Reuter, J.A., Spacek, D. V. & Snyder, M.P. (2015). High-Throughput Sequencing Technologies. *Molecular Cell*. 58 (4). p.pp. 586–597.

Revez, J. *et al.* (2014a). Genome analysis of *Campylobacter jejuni* strains isolated from a waterborne outbreak. *BMC Genomics*. 15. p.p. 768.

Revez, J. *et al.* (2014b). Genomic variation between *Campylobacter jejuni* isolates associated with milk-borne-disease outbreaks. *Journal of Clinical Microbiology*. 52 (8). p.pp. 2782–6.

Ribot, E.M. *et al.* (2001). Rapid pulsed-field gel electrophoresis protocol for subtyping of *Campylobacter jejuni*. *Journal of Clinical Microbiology*. 39 (5). p.pp. 1889–94.

Richardson, N.J. *et al.* (1983). Age related susceptibility to *Campylobacter jejuni* infection in a high prevalance population. *Archives of Disease in Childhood*. 58 (8). p.pp. 616–9.

Richardson, P.T. & Park, S.F. (1995). Enterochelin acquisition in *Campylobacter coli*: characterization of components of a binding-protein-dependent transport system. *Microbiology*. 141 (Pt 1. p.pp. 3181–91.

Ricotta, E.E. *et al.* (2014). Epidemiology and antimicrobial resistance of international travel-associated *Campylobacter* infections in the United States, 2005-2011. *American journal of public health*. 104 (7). p.pp. e108–14.

Riddle, M.S. *et al.* (2012). The chronic gastrointestinal consequences associated with *Campylobacter*. *Current Gastroenterology Reports*. 14 (5). p.pp. 395–405.

Ridley, A. *et al.* (2011a). Potential sources of *Campylobacter* infection on chicken farms: contamination and control of broiler-harvesting equipment, vehicles and

personnel. *Journal of Applied Microbiology*. 111 (1). p.pp. 233–44.

Ridley, A.M. *et al.* (2011b). Longitudinal molecular epidemiological study of thermophilic campylobacters on one conventional broiler chicken farm. *Applied and Environmental Microbiology*. 77 (1). p.pp. 98–107.

Ridley, A.M. *et al.* (2008). Real-time PCR approach for detection of environmental sources of *Campylobacter* strains colonizing broiler flocks. *Applied and Environmental Microbiology*. 74 (8). p.pp. 2492–504.

Robins-Browne, R.M. *et al.* (1980). Summer diarrhoea in African infants and children. *Archives of Disease in Childhood*. 55 (12). p.pp. 923–8.

Robinson, E.R., Walker, T.M. & Pallen, M.J. (2013). Genomics and outbreak investigation: from sequence to consequence. *Genome Medicine*. 5 (4). p.p. 36.

Rohde, H. *et al.* (2011). Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *The New England Journal of Medicine*. 365 (8). p.pp. 718–24.

Rohmer, L. *et al.* (2014). Genomic analysis of the emergence of 20th century epidemic dysentery. *BMC Genomics*. 15. p.p. 355.

Rosef, O. *et al.* (1983). Isolation and characterization of *Campylobacter jejuni* and *Campylobacter coli* from domestic and wild mammals in Norway. *Applied and Environmental Microbiology*. 46 (4). p.pp. 855–9.

Rosenbaum, M., Knight, R. & Leibel, R.L. (2015). The gut microbiota in human energy homeostasis and obesity. *Trends in Endocrinology and Metabolism: TEM*. 26 (9). p.pp. 493-501.

Rosenquist, H. *et al.* (2009). Danish strategies to control *Campylobacter* in broilers and broiler meat: facts and effects. *Epidemiology and Infection*. 137 (12). p.pp. 1742–50.

Rossi, M. *et al.* (2009). *Campylobacter avium* sp. *nov*., a hippurate-positive species isolated from poultry. *International Journal of Systematic and Evolutionary Microbiology*. 59 (Pt 9). p.pp. 2364–9.

Rothberg, J.M. *et al.* (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 475 (7356). p.pp. 348–52.

Rushton, S.P. *et al.* (2009). *Campylobacter* in housed broiler chickens: a longitudinal study of risk factors. *Epidemiology and Infection*. 137 (08). p.p. 1099.

Sabat, A.J. *et al.* (2013). Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveillance*. 18 (4). p.p. 20380.

Sahin, O. *et al.* (2008). Emergence of a tetracycline-resistant *Campylobacter jejuni* clone associated with outbreaks of ovine abortion in the United States. *Journal of Clinical Microbiology*. 46 (5). p.pp. 1663–71.

Sails, A.D., Swaminathan, B. & Fields, P.I. (2003). Utility of multilocus sequence typing as an epidemiological tool for investigation of outbreaks of gastroenteritis caused by *Campylobacter jejuni*. *Journal of Clinical Microbiology*. 41 (10). p.pp. 4733–9.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. 4 (4). p.pp. 406–25.

Salyers, A.A., Gupta, A. & Wang, Y. (2004). Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in Microbiology*. 12 (9). p.pp. 412–6.

Samie, A. *et al.* (2009a). In-vitro cytotoxicity and antimicrobial activities, against clinical isolates of *Campylobacter* species and *Entamoeba histolytica*, of local medicinal plants from the Venda region, in South Africa. *Annals of Tropical Medicine and Parasitology*. 103 (2). p.pp. 159–70.

Samie, A. *et al.* (2007a). Prevalence of *Campylobacter* species, *Helicobacter pylori* and *Arcobacter* species in stool samples from the Venda region, Limpopo, South Africa: studies using molecular diagnostic methods. *The Journal of Infection*. 54 (6). p.pp. 558–566.

Samie, A. *et al.* (2009b). Prevalence of intestinal parasitic and bacterial pathogens in diarrhoeal and non-diarroeal human stools from Vhembe district, South Africa. *Journal of Health, Population, and Nutrition*. 27 (6). p.pp. 739–745.

Samie, A. *et al.* (2007b). Prevalence, haemolytic and haemagglutination activities and antibiotic susceptibility profiles of *Campylobacter* spp. isolated from human diarrhoeal stools in Vhembe District, South Africa. *Journal of Health, Population, and Nutrition*. 25 (4). p.pp. 406–13.

Sandstedt, K. & Ursing, J. (1991). Description of *Campylobacter upsaliensis* sp. nov. Previously Known as the CNW Group. *Systematic and Applied Microbiology*. 14 (1). p.pp. 39–45.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*. 74 (12). p.pp. 5463–7.

Scallan, E. *et al.* (2015). An assessment of the human health impact of seven leading foodborne pathogens in the United States using disability adjusted life years. *Epidemiology and Infection*. 143 (13). p.pp. 2795–804.

Scallan, E. *et al.* (2011). Foodborne illness acquired in the United States--major pathogens. *Emerging Infectious Diseases*. 17 (1). p.pp. 7–15.

Scharff, R.L. (2012). Economic burden from health losses due to foodborne illness in the United States. *Journal of Food Protection*. 75 (1). p.pp. 123–31.

Schewitz, I.A. & Roux, E. (1978). *Campylobacter* infections. First reports from Red Cross War Memorial Children's Hospital, Cape Town. *South African Medical Journal*. 54 (10). p.pp. 385–8.

Schmidt, T., Venter, E.H. & Picard, J.A. (2010). Evaluation of PCR assays for the detection of *Campylobacter fetus* in bovine preputial scrapings and the identification of subspecies in South African field isolates. *Journal of the South African Veterinary Association*. 81 (2). p.pp. 87–92.

Schouls, L.M. *et al.* (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 (1). p.pp. 15–26.

Schutte, A.P., McConnell, E.E. & Bosman, P.P. (1971). Vibrionic abortion in ewes in South Africa: preliminary report. *Journal of the South African Veterinary Medical Association*. 42 (3). p.pp. 223–6.

Schuurman, T. *et al.* (2007). Feasibility of a molecular screening method for detection of *Salmonella enterica* and *Campylobacter jejuni* in a routine community-based clinical microbiology laboratory. *Journal of Clinical Microbiology*. 45 (11). p.pp. 3692–700.

Schwartz, D.C. & Cantor, C.R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*. 37 (1). p.pp. 67–75.

Scott, A.E. *et al.* (2007). Genome dynamics of *Campylobacter jejuni* in response to bacteriophage predation. *PLoS Pathogens*. 3 (8). p.p. e119.

Sears, A. *et al.* (2011). Marked campylobacteriosis decline after interventions aimed at poultry, New Zealand. *Emerging Infectious Diseases*. 17 (6). p.pp. 1007–15.

Sebald, M. & Veron, M. (1963). Base DNA content and classification of vibrios. *Annales de l'Institut Pasteur*. 105. p.pp. 897–910.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 30 (14). p.pp. 2068–9.

Selander, R.K. *et al.* (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Applied and Environmental Microbiology*. 51 (5). p.pp. 873–84.

Seng, P. *et al.* (2009). Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clinical Infectious Diseases*. 49 (4). p.pp. 543–51.

Sheppard, S.K. *et al.* (2009a). *Campylobacter* genotypes from food animals,

environmental sources and clinical disease in Scotland 2005/6. *International Journal of Food Microbiology*. 134 (1-2). p.pp. 96–103.

Sheppard, S.K. *et al.* (2009b). *Campylobacter* genotyping to determine the source of human infection. *Clinical Infectious Diseases*. 48 (8). p.pp. 1072–8.

Sheppard, S.K. *et al.* (2008). Convergence of *Campylobacter* species: implications for bacterial evolution. *Science*. 320 (5873). p.pp. 237–9.

Sheppard, S.K. *et al.* (2010a). Evolution of an agriculture-associated disease causing *Campylobacter coli* clade: evidence from national surveillance data in Scotland. *PloS One*. 5 (12). p.p. e15708.

Sheppard, S.K. *et al.* (2013a). Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. *Proceedings of the National Academy of Sciences of the United States of America*. 110 (29). p.pp. 11923–7.

Sheppard, S.K. *et al.* (2010b). Host association of *Campylobacter* genotypes transcends geographic variation. *Applied and Environmental Microbiology*. 76 (15). p.pp. 5269–77.

Sheppard, S.K. *et al.* (2011). Introgression in the genus *Campylobacter*: generation and spread of mosaic alleles. *Microbiology*. 157 (Pt 4). p.pp. 1066–1074.

Sheppard, S.K. *et al.* (2013b). Progressive genome-wide introgression in agricultural *Campylobacter coli*. *Molecular Ecology*. 22 (4). p.pp. 1051–64.

Shoaf-Sweeney, K.D. *et al.* (2008). Identification of *Campylobacter jejuni* proteins recognized by maternal antibodies of chickens. *Applied and Environmental Microbiology*. 74 (22). p.pp. 6867–75.

Shreiner, A.B., Kao, J.Y. & Young, V.B. (2015). The gut microbiome in health and in disease. *Current Opinion in Gastroenterology*. 31 (1). p.pp. 69–75.

Siemer, B.L. *et al.* (2004). Genetic relatedness among *Campylobacter jejuni* serotyped isolates of diverse origin as determined by numerical analysis of amplified fragment length polymorphism (AFLP) profiles. *Journal of Applied Microbiology*. 96 (4). p.pp. 795–802.

Simpson, E.. (1949). Measurement of Diversity. *Nature*. 163. p.pp. 688–688.

Simpson, J.T. *et al.* (2009). ABySS: a parallel assembler for short read sequence data. *Genome Research*. 19 (6). p.pp. 1117–23.

Skarp, C.P.A. *et al.* (2015). Comparative genomics and genome biology of invasive *Campylobacter jejuni*. *Scientific Reports*. 5. p.p. 17300.

Skarp-de Haan, C.P.A. *et al.* (2014). Comparative genomics of unintrogressed *Campylobacter coli* clades 2 and 3. *BMC Genomics*. 15. p.p. 129.

Smith, L.M. *et al.* (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*. 321 (6071). p.pp. 674–9.

Sommer, M.O.A., Dantas, G. & Church, G.M. (2009). Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science*. 325 (5944). p.pp. 1128–31.

Sopwith, W. *et al.* (2006). *Campylobacter jejuni* multilocus sequence types in humans, northwest England, 2003-2004. *Emerging Infectious Diseases*. 12 (10). p.pp. 1500–7.

Sopwith, W. *et al.* (2010). Investigation of food and environmental exposures relating to the epidemiology of *Campylobacter coli* in humans in Northwest England. *Applied and Environmental Microbiology*. 76 (1). p.pp. 129–35.

Stanley, J. *et al.* (1992). *Campylobacter helveticus* sp. *nov*., a new thermophilic species from domestic animals: characterization, and cloning of a species-specific DNA probe. *Journal of General Microbiology*. 138 (11). p.pp. 2293–2303.

Statistics South Africa (2014). *Poverty trends in South Africa: an examination of absolute poverty between 2006 and 2011.* Pretoria.

Statistics South Africa (2015). *Statistical release P0302. Mid-year population estimates 2015*. Pretoria.

Steele, T.W. & Owen, R.J. (1988). *Campylobacter jejuni* subsp. *doylei* subsp. *nov*., a Subspecies of Nitrate-Negative Campylobacters Isolated from Human Clinical Specimens. *International Journal of Systematic Bacteriology*. 38 (3). p.pp. 316–318.

Stern, N.J. *et al.* (2003). *Campylobacter* spp. in Icelandic poultry operations and human disease. *Epidemiology and Infection*. 130 (1). p.pp. 23–32.

Stern, N.J. *et al.* (2001). Distribution of *Campylobacter* spp. in selected U.S. poultry production and processing operations. *Journal of Food Protection*. 64 (11). p.pp. 1705–10.

Stoddard, R.A. *et al.* (2007). *Campylobacter insulaenigrae* isolates from northern elephant seals (*Mirounga angustirostris*) in California. *Applied and Environmental Microbiology*. 73 (6). p.pp. 1729–35.

Strachan, N.J.C. *et al.* (2009). Attribution of *Campylobacter* infections in northeast Scotland to specific sources by use of multilocus sequence typing. *The Journal of Infectious Diseases*. 199 (8). p.pp. 1205–8.

Strachan, N.J.C. *et al.* (2013). Identifying the seasonal origins of human campylobacteriosis. *Epidemiology and Infection*. 141 (6). p.pp. 1267–75.

Stucki, U. *et al.* (1995). Identification of *Campylobacter jejuni* on the basis of a

species-specific gene that encodes a membrane protein. *Journal of Clinical Microbiology*. 33 (4). p.pp. 855–9.

Suau, A. *et al.* (1999). Direct Analysis of Genes Encoding 16S rRNA from Complex Communities Reveals Many Novel Molecular Species within the Human Gut. *Applied and Environmental Microbiology.* 65 (11). p.pp. 4799–4807.

Suerbaum, S. *et al.* (2001). Allelic diversity and recombination in *Campylobacter* jejuni. *Journal of Bacteriology*. 183 (8). p.pp. 2553–9.

Suzuki, Y. *et al.* (1994). Discrimination by means of pulsed-field gel electrophoresis between strains of *Campylobacter jejuni* Lior type 4 derived from sporadic cases and from outbreaks of infection. *Journal of Infection*. 29 (2). p.pp. 183–187.

Swaminathan, A. *et al.* (2009). A global study of pathogens and host risk factors associated with infectious gastrointestinal disease in returned international travellers. *The Journal of Infection*. 59 (1). p.pp. 19–27.

Taboada, E.N. *et al.* (2013). Current methods for molecular typing of *Campylobacter* species. *Journal of Microbiological Methods*. 95 (1). p.pp. 24–31.

Tam, C. *et al.* (2012). Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut*. 61 (1). p.pp. 69–77.

Tamura, K. *et al.* (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*. 28 (10). p.pp. 2731–9.

Tamura, K. *et al.* (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*. 30 (12). p.pp. 2725–9.

Tan, K.E. *et al.* (2012). Prospective Evaluation of a Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry System in a Hospital Clinical Microbiology Laboratory for Identification of Bacteria and Yeasts: a Bench-by-Bench Study for Assessing the Impact on Time to identification and cost-effectiveness. *Journal of Clinical Microbiology*. 50 (10). p.pp. 3301–3308.

Tang, P. & Gardy, J.L. (2014). Stopping outbreaks with real-time genomic epidemiology. *Genome Medicine*. 6 (11). p.p. 104.

Tanner, A.C.R. *et al.* (1981). *Wolinella* gen. *nov*., *Wolinella succinogenes* (*Vibrio succinogenes* Wolin *et al.*) comb. *nov*., and Description of *Bacteroides gracilis* sp. *nov*., *Wolinella recta* sp. *nov*., *Campylobacter concisus* sp. *nov*., and *Eikenella corrodens* from Humans with Periodontal Disease. *International Journal of Systematic Bacteriology*. 31 (4). p.pp. 432–445.

Tate, J.E. *et al.* (2016). Global, Regional, and National Estimates of Rotavirus

Mortality in Children <5 Years of Age, 2000-2013. *Clinical Infectious Diseases*. 62 (Suppl 2). p.pp. S96–S105.

Tauxe, R. V *et al.* (1985). Illness associated with *Campylobacter laridis*, a newly recognized *Campylobacter* species. *Journal of Clinical Microbiology*. 21 (2). p.pp. 222–5.

Taveirne, M.E. *et al.* (2014). Complete Genome Sequence and Annotation of a *Campylobacter jejuni* Strain, MTVDSCj20, Isolated from a Naturally Colonized Farm-Raised Chicken. *Genome Announcements*. 2 (4).

Tay, A.P. *et al.* (2013). Genome Sequence of *Campylobacter showae* UNSWCD, Isolated from a Patient with Crohn's Disease. *Genome Announcements*. 1 (1).

Taylor, E. V *et al.* (2013). Common source outbreaks of *Campylobacter* infection in the USA, 1997-2008. *Epidemiology and Infection*. 141 (5). p.pp. 987–96.

The World Bank (2015). *GDP Ranking*. [Online]. 2015. Available from: http://data.worldbank.org/data-catalog/GDP-ranking-table. [Accessed: 10 December 2015].

The World Bank (2016a). *GINI index (World Bank estimate)*. [Online]. 2016. Available from: http://data.worldbank.org/indicator/SI.POV.GINI. [Accessed: 13 January 2016].

The World Bank (2016b). *Urban Population*. [Online]. 2016. Available from: http://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS. [Accessed: 13 January 2015].

Thomas, M.K. *et al.* (2008). Burden of acute gastrointestinal illness in Canada, 1999-2007: interim summary of NSAGI activities. *Canada Communicable Disease Repor*. 34 (5). p.pp. 8–15.

Thompson, L. & Berat, L. (2014). *A history of South Africa*. Fourth edition. L. Berat (ed.). New Haven: Yale University Press.

Tissari, P. & Rautelin, H. (2007). Evaluation of an enzyme immunoassay-based stool antigen test to detect *Campylobacter jejuni* and *Campylobacter coli*. *Diagnostic Microbiology and Infectious Disease*. 58 (2). p.pp. 171–5.

Tribble, D.R. *et al.* (2008). Diagnostic approach to acute diarrheal illness in a military population on training exercises in Thailand, a region of *Campylobacter* hyperendemicity. *Journal of Clinical Microbiology*. 46 (4). p.pp. 1418–25.

Tu, Z.-C. *et al.* (2004). *Campylobacter fetus* of reptile origin as a human pathogen. *Journal of Clinical Microbiology*. 42 (9). p.pp. 4405–7.

UN Inter-agency Group for Child Mortality Estimation (2014). *Levels & Trends in Child Mortality*.

Valouev, A. *et al.* (2008). A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Research*. 18 (7). p.pp. 1051–63.

van Belkum, A. *et al.* (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clinical Microbiology and Infection*. 13 Suppl 3. p.pp. 1–46.

van Bergen, M.A.P. *et al.* (2005). Clonal nature of *Campylobacter fetus* as defined by multilocus sequence typing. *Journal of Clinical Microbiology*. 43 (12). p.pp. 5888–98.

Vandamme, P. *et al.* (2005). *Campylobacteraceae*. In: G. Garrity, D. Brenner, N. Krieg, & J. Staley (eds.). *Bergey's Manual of Systematic Bacteriology. Volume 2: The Proteobacteria*. New York: Springer Science, pp. 1147–1160.

Vandamme, P. *et al.* (1995). Chemotaxonomic Analyses of *Bacteroides gracilis* and *Bacteroides ureolyticus* and Reclassification of *B. gracilis* as *Campylobacter gracilis* comb. *nov*. *International Journal of Systematic Bacteriology*. 45 (1). p.pp. 145–152.

Vandamme, P. *et al.* (2010). Reclassification of *Bacteroides ureolyticus* as *Campylobacter ureolyticus* comb. *nov*., and emended description of the genus Campylobacter. *International Journal of Systematic and Evolutionary Microbiology*. 60 (Pt 9). p.pp. 2016–22.

Vandamme, P. *et al.* (1991). Revision of *Campylobacter*, *Helicobacter*, and *Wolinella* taxonomy: emendation of generic descriptions and proposal of *Arcobacter* gen. *nov*. *International Journal of Systematic Bacteriology*. 41 (1). p.pp. 88–103.

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 (Pt 2). p.pp. 719–21.

van de Giessen, A.W. *et al.* (1998). Reduction of *Campylobacter* infections in broiler flocks by application of hygiene measures. *Epidemiology and Infection*. 121 (01). p.pp. 57–66.

van der Graaf-van Bloois, L. *et al.* (2014). Inconsistency of phenotypic and genomic characteristics of *Campylobacter fetus* subspecies requires reevaluation of current diagnostics. *Journal of Clinical Microbiology*. 52 (12). p.pp. 4183–8.

van der Walt, M.L. & van der Lugt, J.J. (1988). The isolation of *Campylobacter hyointestinalis* from a pig in South Africa. *The Onderstepoort Journal of Veterinary Research*. 55 (2). p.pp. 85–7.

van der Walt, M.L., Spencer, B.T. & Loveday, R.K. (1988). The first isolations of *Campylobacter mucosalis* from pigs in South Africa. *The Onderstepoort Journal*

of Veterinary Research. 55 (3). p.pp. 165–8.

Van Lint, P. *et al.* (2015). Evaluation of a real-time multiplex PCR for the simultaneous detection of *Campylobacter jejuni*, *Salmonella* spp., *Shigella* spp./EIEC, and *Yersinia enterocolitica* in fecal samples. *European Journal of Clinical Microbiology & Infectious Diseases*. 34 (3). p.pp. 535–42.

van Nierop, W. *et al.* (2005). Contamination of chicken carcasses in Gauteng, South Africa, by *Salmonella*, *Listeria monocytogenes* and *Campylobacter*. *International Journal of Food Microbiology*. 99 (1). p.pp. 1–6.

van Opijnen, T. & Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Reviews Microbiology*. 11 (7). p.pp. 435–42.

Van Overbeke, I. *et al.* (2006). A comparison survey of organic and conventional broiler chickens for infectious agents affecting health and food safety. *Avian Diseases*. 50 (2). p.pp. 196–200.

van Putten, J.P.M. *et al.* (2009). Molecular mechanisms of *Campylobacter* infection. *Current Topics in Microbiology and Immunology*. 337. p.pp. 197–229.

van Tonder, A.J. *et al.* (2015). Genomics Reveals the Worldwide Distribution of Multidrug-Resistant Serotype 6E Pneumococci. *Journal of Clinical Microbiology*. 53 (7). p.pp. 2271–85.

Vergara-Irigaray, M. *et al.* (2014). RNA-seq analysis of the influence of anaerobiosis and FNR on *Shigella flexneri*. *BMC Genomics*. 15. p.p. 438.

Veron, M. & Chatelain, R. (1973). "Taxonomic study of the genus *Campylobacter* Sebald and Véron and designation of the neotype strain for the type species, *Campylobacter fetus* (Smith and Taylor) Sebald and Véron." *International Journal of Systematic Bacteriology*. 23 (2). p.pp. 122–134.

Vos, P. *et al.* (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*. 23 (21). p.pp. 4407–14.

Vos, T. *et al.* (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*. 386 (9995). p.pp. 743–800.

Wagenaar, J.A., French, N.P. & Havelaar, A.H. (2013). Preventing *Campylobacter* at the source: why is it so difficult? *Clinical Infectious Diseases*. 57 (11). p.pp. 1600–6.

Wakerley, B.R. & Yuki, N. (2015). Guillain–Barré syndrome. *Expert Review of Neurotherapeutics*. 15 (8). p.pp. 847-849.

Waldenström, J. *et al.* (2002). Prevalence of *Campylobacter jejuni*, *Campylobacter*

lari, and *Campylobacter coli* in different ecological guilds and taxa of migrating birds. *Applied and Environmental Microbiology*. 68 (12). p.pp. 5911–7.

Walker, C.L.F. *et al.* (2013a). Global burden of childhood pneumonia and diarrhoea. *Lancet*. 381 (9875). p.pp. 1405–16.

Walker, S.P. *et al.* (2011). Inequality in early childhood: risk and protective factors for early child development. *Lancet*. 378 (9799). p.pp. 1325–38.

Walker, T.M. *et al.* (2013b). Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infectious Diseases*. 13 (2). p.pp. 137–46.

Wang, C.-M. *et al.* (2015). Complete Genome Sequence of *Campylobacter fetus* subsp. *testudinum* Strain Pet-3, Isolated from a Lizard (*Hydrosaurus pustulatus*). *Genome Announcements*. 3 (1).

Wang, Y. & Taylor, D.E. (1990). Natural transformation in *Campylobacter* species. *Journal of Bacteriology*. 172 (2). p.pp. 949–55.

Wang, Z.-K. *et al.* (2014). Intestinal microbiota pathogenesis and fecal microbiota transplantation for inflammatory bowel disease. *World Journal of Gastroenterology*. 20 (40). p.pp. 14805–20.

Warren, R.L. *et al.* (2013). Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome*. 1 (1). p.p. 16.

Wassenaar, T.M. *et al.* (2000). Genetic characterization of *Campylobacter jejuni* O:41 isolates in relation with Guillain-Barré syndrome. *Journal of Clinical Microbiology*. 38 (2). p.pp. 874–6.

Wassenaar, T.M., Fry, B.N. & van der Zeijst, B.A.M. (1995). Variation of the flagellin gene locus of *Campylobacter* jejuni by recombination and horizontal gene transfer. *Microbiology*. 141 (1). p.pp. 95–101.

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

Whyte, L.A., Al-Araji, R.A. & McLoughlin, L.M. (2015). Guidelines for the management of acute gastroenteritis in children in Europe. *Archives of Disease in Childhood. Education and practice edition*.

Wieland, B. *et al.* (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 (4). p.pp. 183–9.

Wielinga, P.R. & Schlundt, J. (2013). Food Safety: at the center of a One Health approach for combating zoonoses. *Current Topics in Microbiology and Immunology*. 366. p.pp. 3–17.

Wilson, D.J. (2012). Insights from genomics into bacterial pathogen populations. *PLoS Pathogens*. 8 (9). p.p. e1002874.

Wilson, D.J. *et al.* (2009). Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. *Molecular Biology and Evolution*. 26 (2). p.pp. 385–97.

Wilson, D.J. *et al.* (2008). Tracing the source of campylobacteriosis. *PLoS Genetics*. 4 (9). p.p. e1000203.

Wilson, M.E. (1995). Travel and the emergence of infectious diseases. *Emerging Infectious Diseases*. 1 (2). p.pp. 39–46.

Wong, V.K. *et al.* (2015). Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events. *Nature Genetics*. 47 (6). p.pp. 632–9.

World Health Organisation (2013). *Diarrhoeal disease*. [Online]. 2013. Available from: http://www.who.int/mediacentre/factsheets/fs330/en/. [Accessed: 25 August 2015].

Wyres, K.L. *et al.* (2012). The multidrug-resistant PMEN1 pneumococcus is a paradigm for genetic success. *Genome Biology*. 13 (11). p.p. R103.

Yahara, K. *et al.* (2014). Efficient inference of recombination hot regions in bacterial genomes. *Molecular Biology and Evolution*. 31 (6). p.pp. 1593–605.

Yamazaki-Matsune, W. *et al.* (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 (Pt 11). p.pp. 1467–73.

Young, K.T., Davis, L.M. & Dirita, V.J. (2007). *Campylobacter jejuni*: molecular biology and pathogenesis. *Nature Reviews Microbiology*. 5 (9). p.pp. 665–79.

Yuki, N. (1993). A bacterium lipopolysaccharide that elicits Guillain-Barre syndrome has a GM1 ganglioside-like structure. *Journal of Experimental Medicine*. 178 (5). p.pp. 1771–1775.

Yuki, N. *et al.* (2004). Carbohydrate mimicry between human ganglioside GM1 and *Campylobacter jejuni* lipooligosaccharide causes Guillain-Barre syndrome. *Proceedings of the National Academy of Sciences*. 101 (31). p.pp. 11404–11409.

Zanoni, R.G. *et al.* (2009). *Campylobacter cuniculorum* sp. *nov*., from rabbits. *International Journal of Systematic and Evolutionary Microbiology*. 59 (Pt 7). p.pp. 1666–71.

Zerbino, D.R. (2010). Using the Velvet *de novo* assembler for short-read

sequencing technologies. *Current Protocols in Bioinformatics.* Chapter 11. p.p. Unit 11.5.

Zerbino, D.R. & Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research*. 18 (5). p.pp. 821–9.

Zhang, H., Morrison, S. & Tang, Y.-W. (2015). Multiplex Polymerase Chain Reaction Tests for Detection of Pathogens Associated with Gastroenteritis. *Clinics in Laboratory Medicine*. 35 (2). p.pp. 461–486.

Zhang, M. *et al.* (2010a). Genomic characterization of the Guillain-Barre syndrome-associated *Campylobacter jejuni* ICDCCJ07001 Isolate. *PloS One*. 5 (11). p.p. e15060.

Zhang, M. *et al.* (2010b). Molecular typing and antimicrobial susceptibility profiles of *Campylobacter jejuni* isolates from north China. *Journal of Medical Microbiology*. 59 (Pt 10). p.pp. 1171–7.

Zhang, Y. & Sievert, S.M. (2014). Pan-genome analyses identify lineage- and niche-specific markers of evolution and adaptation in *Epsilonproteobacteria*. *Frontiers in Microbiology*. 5. p.p. 110.

Zhao, X. *et al.* (2014). Advances in rapid detection methods for foodborne pathogens. *Journal of Microbiology and Biotechnology*. 24 (3). p.pp. 297–312.

Zhou, X. *et al.* (2010). The next-generation sequencing technology and application. *Protein & Cell*. 1 (6). p.pp. 520–36.

Zinszer, K. *et al.* (2010). The impact of domestic travel on estimating regional rates of human campylobacteriosis. *Epidemiology and Infection*. 138 (12). p.pp. 1735–43.

APPENDICES

Appendix A

Table A1 Isolates included in the *mapA/ceuE* RT-PCR evaluation

PubMLST id	Isolate	ENA accession
25430	OXC8106	ERR386294
25530	OXC8103	ERR386394
25529	OXC8101	ERR386393
25528	OXC8100	ERR386392
25527	OXC8099	ERR386391
25526	OXC8098	ERR386390
25612	OXC8097	ERR386476
25525	OXC8096	ERR386389
25428	OXC8095	ERR386292
25427	OXC8094	ERR386291
25611	OXC8093	ERR386475
25610	OXC8092	ERR386474
25524	OXC8091	ERR386388
25523	OXC8090	ERR386387
25426	OXC8089	ERR386290
25522	OXC8088	ERR386386
25521	OXC8087	ERR386385
25425	OXC8086	ERR386289
25424	OXC8085	ERR386288
25423	OXC8084	ERR386287
25422	OXC8083	ERR386286
25520	OXC8082	ERR386384
25421	OXC8081	ERR386285
25519	OXC8080	ERR386383
25420	OXC8079	ERR386284
25518	OXC8078	ERR386382
25517	OXC8077	ERR386381
25419	OXC8076	ERR386283
25609	OXC8075	ERR386473
25608	OXC8074	ERR386472
25607	OXC8073	ERR386471
25606	OXC8072	ERR386470
25605	OXC8071	ERR386469
25516	OXC8070	ERR386380
25418	OXC8069	ERR386282
25417	OXC8068	ERR386281
25604	OXC8067	ERR386468

PubMLST id	Isolate	ENA accession
25416	OXC8066	ERR386280
25515	OXC8065	ERR386379
25514	OXC8064	ERR386378
25513	OXC8063	ERR386377
25512	OXC8062	ERR386376
25603	OXC8061	ERR386467
25415	OXC8060	ERR386279
25602	OXC8059	ERR386466
25511	OXC8058	ERR386375
25601	OXC8057	ERR386465
25414	OXC8056	ERR386278
25600	OXC8055	ERR386464
25598	OXC8053	ERR386462
25510	OXC8052	ERR386374
28559	OXC8051	ERR483981
25596	OXC8050	ERR386460
25595	OXC8049	ERR386459
25509	OXC8048	ERR386373
25508	OXC8046	ERR386372
25413	OXC8044	ERR386277
25593	OXC8043	ERR386457
25507	OXC8042	ERR386371
25506	OXC8041	ERR386370
25505	OXC8040	ERR386369
25592	OXC8039	ERR386456
25504	OXC8038	ERR386368
25412	OXC8037	ERR386276
25411	OXC8036	ERR386275
25503	OXC8035	ERR386367
25591	OXC8034	ERR386455
25502	OXC8033	ERR386366
25500	OXC8031	ERR386364
25499	OXC8030	ERR386363
25498	OXC8029	ERR386362
25590	OXC8028	ERR386454
25497	OXC8027	ERR386361
25589	OXC8026	ERR386453

25496	OXC8025	ERR386360
25410	OXC8024	ERR386274
25588	OXC8023	ERR386452
25409	OXC8022	ERR386273
25495	OXC8021	ERR386359
25587	OXC8020	ERR386451
25494	OXC8019	ERR386358
25586	OXC8018	ERR386450
25408	OXC8017	ERR386272
25493	OXC8016	ERR386357
25585	OXC8015	ERR386449
25584	OXC8014	ERR386448
25583	OXC8013	ERR386447
25582	OXC8012	ERR386446
25581	OXC8011	ERR386445
25492	OXC8010	ERR386356
25580	OXC8009	ERR386444
25579	OXC8008	ERR386443
25491	OXC8007	ERR386355
25490	OXC8006	ERR386354
25489	OXC8005	ERR386353
25488	OXC8004	ERR386352
25407	OXC8003	ERR386271
25487	OXC8002	ERR386351
25406	OXC8001	ERR386270
25578	OXC8000	ERR386442
25486	OXC7999	ERR386350
25404	OXC7997	ERR386268
25485	OXC7996	ERR386349
25403	OXC7995	ERR386267
25484	OXC7994	ERR386348
25483	OXC7992	ERR386347
25401	OXC7991	ERR386265
25482	OXC7990	ERR386346
25481	OXC7989	ERR386345
25400	OXC7988	ERR386264
25399	OXC7987	ERR386263
25398	OXC7986	ERR386262
25577	OXC7985	ERR386441
25397	OXC7984	ERR386261
25576	OXC7983	ERR386440
25396	OXC7982	ERR386260
25395	OXC7981	ERR386259
25575	OXC7980	ERR386439
25394	OXC7979	ERR386258
25393	OXC7978	ERR386257
25392	OXC7977	ERR386256
25391	OXC7976	ERR386255
25390	OXC7975	ERR386254
25389	OXC7974	ERR386253
25388	OXC7973	ERR386252
25387	OXC7972	ERR386251
25386	OXC7971	ERR386250
25385	OXC7970	ERR386249
25384	OXC7969	ERR386248
25382	OXC7967	ERR386246
25381	OXC7966	ERR386245
25574	OXC7965	ERR386438
25380	OXC7964	ERR386244
25379	OXC7963	ERR386243
25480	OXC7962	ERR386344
25479	OXC7961	ERR386343
25378	OXC7960	ERR386242
25377	OXC7959	ERR386241
25478	OXC7958	ERR386342
25376	OXC7957	ERR386240
25375	OXC7956	ERR386239
25573	OXC7955	ERR386437
25374	OXC7954	ERR386238
25572	OXC7953	ERR386436
25477	OXC7952	ERR386341
25571	OXC7951	ERR386435
25570	OXC7950	ERR386434
25569	OXC7949	ERR386433
25652	OXC7948	ERR386516
25476	OXC7947	ERR386340
25456	OXC7946	ERR386320
25068	OXC7945	ERR348940
25067	OXC7944	ERR348932
25066	OXC7943	ERR348924
25568	OXC7942	ERR386432
25120	OXC7941	ERR348971
25065	OXC7940	ERR348916
25064	OXC7939	ERR348908
25063	OXC7938	ERR348900
25119	OXC7937	ERR348979
25567	OXC7936	ERR386431
25118	OXC7935	ERR348987
25061	OXC7933	ERR348884
25060	OXC7932	ERR348876
25059	OXC7931	ERR348868
25117	OXC7930	ERR348995

25058	OXC7929	ERR348955	25113	OXC7878	ERR349027
25057	OXC7928	ERR348947	25012	OXC7877	ERR348864
25056	OXC7927	ERR348939	25011	OXC7876	ERR348951
25055	OXC7926	ERR348931	25010	OXC7875	ERR348943
25054	OXC7925	ERR348923	25009	OXC7874	ERR348935
25053	OXC7924	ERR348915	25008	OXC7873	ERR348927
25051	OXC7922	ERR348899	25007	OXC7872	ERR348919
25116	OXC7921	ERR349003	25006	OXC7871	ERR348911
25050	OXC7920	ERR348891	25005	OXC7870	ERR348903
25049	OXC7919	ERR348883	25004	OXC7869	ERR348895
25048	OXC7918	ERR348875	25003	OXC7868	ERR348887
25047	OXC7917	ERR348867	25002	OXC7867	ERR348879
25046	OXC7916	ERR348954	25001	OXC7866	ERR348871
25045	OXC7915	ERR348946	25000	OXC7865	ERR348863
25044	OXC7914	ERR348938	24999	OXC7864	ERR348950
25043	OXC7913	ERR348930	24997	OXC7862	ERR348934
25042	OXC7912	ERR348922	24996	OXC7861	ERR348926
25041	OXC7911	ERR348914	24995	OXC7860	ERR348918
25040	OXC7910	ERR348906	24994	OXC7859	ERR348910
25039	OXC7909	ERR348898	24993	OXC7858	ERR348902
25038	OXC7908	ERR348890	24992	OXC7857	ERR348894
25037	OXC7907	ERR348882	24991	OXC7856	ERR348886
25069	OXC7906	ERR348874	24990	OXC7855	ERR348878
25035	OXC7904	ERR348953	24989	OXC7854	ERR348870
25034	OXC7903	ERR348945	24988	OXC7853	ERR348862
25033	OXC7902	ERR348937	24649	OXC7852	ERR330616
25032	OXC7901	ERR348929	24648	OXC7851	ERR330608
25031	OXC7900	ERR348921	24647	OXC7850	ERR330600
25030	OXC7899	ERR348913	24646	OXC7849	ERR330592
25028	OXC7897	ERR348897	24645	OXC7848	ERR330584
25027	OXC7896	ERR348889	24644	OXC7847	ERR330576
25026	OXC7895	ERR348881	24643	OXC7846	ERR330568
25025	OXC7894	ERR348873	24642	OXC7845	ERR330560
25023	OXC7892	ERR348952	24641	OXC7844	ERR330647
25022	OXC7891	ERR348944	24640	OXC7843	ERR330639
25021	OXC7890	ERR348936	24639	OXC7842	ERR330631
25020	OXC7889	ERR348928	24638	OXC7841	ERR330623
25019	OXC7888	ERR348920	24637	OXC7840	ERR330615
25018	OXC7887	ERR348912	24636	OXC7839	ERR330607
25115	OXC7886	ERR349019	24633	OXC7836	ERR330583
25114	OXC7885	ERR349011	24632	OXC7835	ERR330575
25017	OXC7884	ERR348904	24630	OXC7833	ERR330559
25016	OXC7883	ERR348896	24629	OXC7832	ERR330646
25015	OXC7882	ERR348888	24627	OXC7830	ERR330630
25014	OXC7881	ERR348880	24626	OXC7829	ERR330622
25013	OXC7879	ERR348872	24625	OXC7828	ERR330614

24624	OXC7827	ERR330606
24623	OXC7826	ERR330598
24622	OXC7825	ERR330590
24621	OXC7824	ERR330582
24620	OXC7823	ERR330574
24619	OXC7822	ERR330566
24618	OXC7821	ERR330558
24617	OXC7820	ERR330645
24616	OXC7819	ERR330637
24615	OXC7818	ERR330629
27885	OXC7817	ERR459482
24614	OXC7816	ERR330621
24612	OXC7814	ERR330605
24611	OXC7813	ERR330597
24610	OXC7812	ERR330589
24609	OXC7811	ERR330581
24608	OXC7810	ERR330573
24607	OXC7809	ERR330565
24606	OXC7808	ERR330557
24605	OXC7807	ERR330644
24604	OXC7806	ERR330636
25112	OXC7805	ERR349053
27884	OXC7804	ERR459481
24603	OXC7803	ERR330628
24602	OXC7802	ERR330620
25111	OXC7801	ERR349045
24601	OXC7800	ERR330612
24600	OXC7799	ERR330604
24599	OXC7798	ERR330596
24598	OXC7797	ERR330588
24597	OXC7796	ERR330580
27883	OXC7794	ERR459480
24596	OXC7793	ERR330572
24595	OXC7792	ERR330564
24594	OXC7791	ERR330556
24593	OXC7790	ERR330643
24592	OXC7789	ERR330635
24591	OXC7788	ERR330627
27882	OXC7787	ERR459479
24590	OXC7786	ERR330619
25455	OXC7785	ERR483978
25654	OXC7783	ERR386518
24589	OXC7782	ERR330611
24588	OXC7781	ERR330603
24587	OXC7780	ERR330595
24586	OXC7779	ERR330587
24585	OXC7778	ERR330579
24584	OXC7777	ERR330571
24583	OXC7776	ERR330563
24582	OXC7775	ERR330555
24581	OXC7773	ERR330642
24580	OXC7772	ERR330634
27881	OXC7771	ERR459478
24579	OXC7770	ERR330626
24578	OXC7769	ERR330618
24577	OXC7768	ERR330610
24576	OXC7767	ERR330602
24575	OXC7766	ERR330594
24574	OXC7765	ERR330586
24573	OXC7764	ERR330578
24572	OXC7761	ERR330570
24571	OXC7759	ERR330562
25110	OXC7758	ERR349037
25454	OXC7757	ERR386318
24570	OXC7756	ERR330554
24987	OXC7755	ERR343092
27870	OXC7754	ERR351541
27869	OXC7753	ERR351540
24986	OXC7752	ERR343071
25453	OXC7751	ERR386317
24985	OXC7750	ERR343063
24984	OXC7749	ERR343055
24983	OXC7748	ERR343047
24982	OXC7747	ERR343039
24981	OXC7746	ERR343031
24980	OXC7745	ERR343023
24979	OXC7744	ERR343016
24978	OXC7743	ERR343009
24977	OXC7742	ERR343091
24976	OXC7741	ERR343084
27868	OXC7740	ERR351539
25651	OXC7739	ERR386515
24975	OXC7738	ERR343070
27880	OXC7737	ERR459477
24974	OXC7736	ERR343062
24973	OXC7735	ERR343054
24972	OXC7734	ERR343046
24971	OXC7732	ERR343038
24970	OXC7731	ERR343030
25538	OXC7730	ERR386402
25630	OXC7729	ERR386494
24969	OXC7728	ERR343022

24968	OXC7727	ERR343015
24967	OXC7726	ERR343008
24966	OXC7725	ERR343090
24965	OXC7724	ERR343083
27879	OXC7723	ERR459476
24964	OXC7722	ERR343077
24963	OXC7721	ERR343069
24962	OXC7720	ERR343061
24961	OXC7719	ERR343053
24960	OXC7718	ERR343045
24959	OXC7717	ERR343037
24958	OXC7716	ERR343029
27867	OXC7715	ERR351538
27866	OXC7714	ERR351537
27865	OXC7713	ERR351536
24957	OXC7712	ERR343089
24956	OXC7711	ERR343082
25537	OXC7710	ERR386401
24955	OXC7709	ERR343076
25629	OXC7708	ERR386493
24954	OXC7707	ERR343068
24953	OXC7706	ERR343060
24952	OXC7705	ERR343052
24951	OXC7704	ERR343044
24950	OXC7702	ERR343036
24949	OXC7701	ERR343028
24948	OXC7700	ERR343021
24947	OXC7699	ERR343014
24946	OXC7698	ERR343007
24945	OXC7697	ERR343088
24944	OXC7696	ERR343081
24943	OXC7695	ERR343075
24942	OXC7694	ERR343067
24941	OXC7693	ERR343059
24940	OXC7692	ERR343051
25628	OXC7691	ERR386492
24939	OXC7689	ERR343043
25627	OXC7688	ERR386491
24938	OXC7687	ERR343035
24937	OXC7686	ERR343027
25465	OXC7685	ERR386329
24936	OXC7683	ERR343020
24935	OXC7682	ERR343013
27878	OXC7681	ERR459474
24934	OXC7680	ERR343006
24933	OXC7679	ERR343087
24932	OXC7678	ERR343080
24931	OXC7677	ERR343074
24930	OXC7676	ERR343066
24929	OXC7675	ERR343058
24928	OXC7674	ERR343050
24927	OXC7673	ERR343042
25650	OXC7671	ERR386514
24925	OXC7670	ERR343026
24924	OXC7669	ERR343019
24923	OXC7668	ERR343012
24922	OXC7667	ERR343005
24921	OXC7666	ERR343086
24920	OXC7665	ERR343079
24919	OXC7664	ERR343073
24918	OXC7663	ERR343065
24917	OXC7662	ERR343057
24916	OXC7661	ERR343049
24915	OXC7660	ERR343041
24914	OXC7659	ERR343033
24913	OXC7658	ERR343025
24912	OXC7657	ERR343018
24911	OXC7656	ERR343011
24910	OXC7655	ERR343004
24909	OXC7654	ERR343085
24908	OXC7653	ERR343078
25452	OXC7652	ERR386316
25555	OXC7651	ERR386419
24907	OXC7650	ERR343072
24906	OXC7649	ERR343064
24905	OXC7648	ERR343056
24904	OXC7647	ERR343048
24903	OXC7646	ERR343040
24901	OXC7644	ERR343024
24900	OXC7643	ERR343017
24899	OXC7642	ERR343010
24898	OXC7641	ERR343003
24569	OXC7640	ERR330553
24568	OXC7639	ERR330545
24567	OXC7638	ERR330537
24566	OXC7637	ERR330529
24565	OXC7636	ERR330521
24564	OXC7635	ERR330513
24563	OXC7634	ERR330505
24562	OXC7633	ERR330497
24561	OXC7632	ERR330489
24560	OXC7631	ERR330481

24559	OXC7630	ERR330473	24516	OXC7580	ERR330509
24558	OXC7629	ERR330465	25106	OXC7579	ERR349005
24557	OXC7628	ERR330552	24515	OXC7578	ERR330501
24556	OXC7627	ERR330544	24513	OXC7576	ERR330485
24555	OXC7626	ERR330536	24512	OXC7575	ERR330477
24554	OXC7625	ERR330528	24510	OXC7573	ERR330461
24553	OXC7624	ERR330520	24509	OXC7572	ERR330548
24552	OXC7623	ERR330512	24508	OXC7571	ERR330540
24551	OXC7622	ERR330504	24507	OXC7570	ERR330532
24550	OXC7621	ERR330496	24506	OXC7569	ERR330524
24549	OXC7620	ERR330488	24505	OXC7568	ERR330516
24548	OXC7619	ERR330480	24504	OXC7567	ERR330508
24547	OXC7618	ERR330472	24503	OXC7566	ERR330500
24546	OXC7617	ERR330464	24502	OXC7565	ERR330492
24545	OXC7616	ERR330551	24501	OXC7564	ERR330484
24544	OXC7615	ERR330543	24500	OXC7563	ERR330476
24543	OXC7614	ERR330535	24499	OXC7562	ERR330468
24542	OXC7613	ERR330527	24498	OXC7561	ERR330460
24541	OXC7612	ERR330519	24497	OXC7560	ERR330547
24540	OXC7611	ERR330511	24496	OXC7559	ERR330539
25109	OXC7610	ERR349029	24495	OXC7558	ERR330531
24539	OXC7609	ERR330503	24494	OXC7557	ERR330523
24538	OXC7608	ERR330495	24493	OXC7556	ERR330515
24537	OXC7607	ERR330487	25105	OXC7555	ERR348997
24536	OXC7606	ERR330479	24492	OXC7554	ERR330507
24535	OXC7604	ERR330471	24491	OXC7553	ERR330499
24534	OXC7603	ERR330463	24490	OXC7552	ERR330491
25108	OXC7602	ERR349021	24489	OXC7551	ERR330483
24533	OXC7601	ERR330550	25626	OXC7550	ERR386490
24532	OXC7600	ERR330542	24488	OXC7549	ERR330475
24531	OXC7599	ERR330534	24487	OXC7548	ERR330467
24530	OXC7598	ERR330526	24486	OXC7547	ERR330459
24529	OXC7597	ERR330518	25104	OXC7545	ERR348989
24528	OXC7596	ERR330510	24485	OXC7544	ERR330546
24527	OXC7595	ERR330502	24484	OXC7543	ERR330538
24526	OXC7594	ERR330494	24483	OXC7542	ERR330530
24525	OXC7593	ERR330486	24482	OXC7541	ERR330522
24524	OXC7592	ERR330478	24481	OXC7538	ERR330514
24523	OXC7589	ERR330470	24480	OXC7537	ERR330506
24522	OXC7588	ERR330462	24479	OXC7536	ERR330498
24521	OXC7587	ERR330549	24477	OXC7534	ERR330482
24520	OXC7586	ERR330541	24476	OXC7533	ERR330474
24519	OXC7585	ERR330533	24475	OXC7532	ERR330466
24518	OXC7584	ERR330525	24474	OXC7531	ERR330458
24517	OXC7583	ERR330517	24134	OXC7530	ERR278513
25107	OXC7582	ERR349013	24133	OXC7529	ERR278512

24132	OXC7528	ERR278511
24131	OXC7527	ERR278510
24130	OXC7526	ERR278509
24129	OXC7525	ERR278508
24128	OXC7524	ERR278507
25405	OXC7523	ERR494547
24126	OXC7522	ERR278505
24125	OXC7521	ERR278504
24124	OXC7520	ERR278503
24123	OXC7518	ERR278502
24122	OXC7517	ERR278501
24121	OXC7516	ERR278500
24120	OXC7515	ERR278499
24119	OXC7514	ERR278498
24038	OXC7513	ERR278415
24037	OXC7512	ERR278414
24036	OXC7511	ERR278413
24035	OXC7509	ERR278412
24034	OXC7508	ERR278411
24033	OXC7507	ERR278410
24032	OXC7506	ERR278409
24031	OXC7505	ERR278408
24030	OXC7504	ERR278407
24029	OXC7503	ERR278406
24028	OXC7502	ERR278405
24027	OXC7500	ERR278404
24026	OXC7499	ERR278403
24025	OXC7498	ERR278402
24024	OXC7496	ERR278401
24023	OXC7495	ERR278400
24022	OXC7494	ERR278399
24021	OXC7493	ERR278398
24019	OXC7491	ERR278396
24018	OXC7490	ERR278395
24017	OXC7489	ERR278394
24016	OXC7488	ERR278393
24015	OXC7487	ERR278392
24014	OXC7486	ERR278391
24013	OXC7485	ERR278390
24012	OXC7484	ERR278389
25624	OXC7483	ERR386488
24011	OXC7482	ERR278388
24010	OXC7481	ERR278387
24009	OXC7480	ERR278386
24008	OXC7479	ERR278385
24007	OXC7478	ERR278384
24006	OXC7477	ERR278383
24005	OXC7476	ERR278382
24004	OXC7475	ERR278381
24003	OXC7474	ERR278380
24002	OXC7473	ERR278379
24001	OXC7472	ERR278378
24000	OXC7471	ERR278377
23999	OXC7470	ERR278376
23998	OXC7469	ERR278375
23997	OXC7468	ERR278374
23996	OXC7467	ERR278373
23995	OXC7466	ERR278372
23994	OXC7465	ERR278371
23993	OXC7464	ERR278370
23992	OXC7463	ERR278369
23991	OXC7462	ERR278368
23990	OXC7461	ERR278367
25623	OXC7460	ERR386487
23989	OXC7459	ERR278366
23988	OXC7458	ERR278365
23987	OXC7457	ERR278364
23986	OXC7456	ERR278363
23985	OXC7455	ERR278362
23984	OXC7454	ERR278361
23983	OXC7453	ERR278360
23982	OXC7452	ERR278359
23981	OXC7451	ERR278358
23980	OXC7450	ERR278357
23979	OXC7449	ERR278356
23978	OXC7448	ERR278355
23977	OXC7447	ERR278354
23976	OXC7446	ERR278353
23975	OXC7444	ERR278352
23974	OXC7443	ERR278351
23973	OXC7442	ERR278350
23972	OXC7441	ERR278349
23971	OXC7440	ERR278348
23970	OXC7439	ERR278347
23969	OXC7438	ERR278346
23968	OXC7437	ERR278345
23967	OXC7436	ERR278344
25622	OXC7435	ERR386486
25621	OXC7434	ERR386485
23966	OXC7433	ERR278343
23965	OXC7432	ERR278342
23964	OXC7431	ERR278341

23963	OXC7430	ERR278340
23962	OXC7429	ERR278339
23961	OXC7428	ERR278338
25620	OXC7426	ERR386484
23959	OXC7425	ERR278336
23958	OXC7424	ERR278335
23957	OXC7423	ERR278334
23956	OXC7422	ERR278333
23955	OXC7421	ERR278332
23954	OXC7420	ERR278331
23952	OXC7418	ERR278329
23951	OXC7417	ERR278328
23950	OXC7416	ERR278327
23949	OXC7415	ERR278326
23948	OXC7414	ERR278325
23947	OXC7412	ERR278324
23946	OXC7411	ERR278323
23945	OXC7410	ERR278322
23944	OXC7409	ERR278289
23943	OXC7408	ERR278288
23942	OXC7407	ERR278287
23941	OXC7405	ERR278286
23940	OXC7404	ERR278285
23937	OXC7401	ERR278282
23936	OXC7400	ERR278281
25450	OXC7399	ERR386314
23935	OXC7398	ERR278280
23934	OXC7397	ERR278279
23933	OXC7395	ERR278278
23932	OXC7394	ERR278277
23930	OXC7392	ERR278275
23929	OXC7391	ERR278274
23928	OXC7390	ERR278273
23927	OXC7389	ERR278272
23926	OXC7388	ERR278271
23925	OXC7387	ERR278270
23924	OXC7386	ERR278269
23923	OXC7385	ERR278268
23922	OXC7384	ERR278267
23921	OXC7383	ERR278266
23920	OXC7382	ERR278265
23919	OXC7381	ERR278264
23918	OXC7380	ERR278263
23917	OXC7379	ERR278262
23916	OXC7378	ERR278261
23915	OXC7377	ERR278260
23914	OXC7376	ERR278259
23913	OXC7375	ERR278258
23912	OXC7374	ERR278257
23911	OXC7373	ERR278256
23910	OXC7372	ERR278255
23909	OXC7371	ERR278254
23908	OXC7370	ERR278253
23907	OXC7369	ERR278252
23906	OXC7367	ERR278251
23905	OXC7366	ERR278250
23904	OXC7365	ERR278249
23903	OXC7364	ERR278248
23902	OXC7363	ERR278247
23901	OXC7362	ERR278246
23900	OXC7361	ERR278245
23899	OXC7360	ERR278244
23898	OXC7359	ERR278243
23897	OXC7358	ERR278242
23896	OXC7357	ERR278241
23895	OXC7356	ERR278240
23894	OXC7355	ERR278239
23893	OXC7354	ERR278238
23892	OXC7353	ERR278237
27877	OXC7352	ERR459473
23891	OXC7351	ERR278236
23890	OXC7350	ERR278235
23889	OXC7349	ERR278234
23888	OXC7348	ERR278233
23887	OXC7347	ERR278232
23886	OXC7346	ERR278231
23885	OXC7345	ERR278230
23884	OXC7344	ERR278229
25449	OXC7343	ERR386313
23883	OXC7342	ERR278228
22378	OXC7341	ERR221475
22377	OXC7340	ERR221474
22376	OXC7339	ERR221473
22375	OXC7338	ERR221472
22374	OXC7336	ERR221471
22373	OXC7335	ERR221470
22372	OXC7334	ERR221469
22371	OXC7333	ERR221468
22370	OXC7332	ERR221467
22369	OXC7331	ERR221466
22368	OXC7330	ERR221465
22367	OXC7329	ERR221464

22366	OXC7328	ERR221463
22365	OXC7326	ERR221462
22364	OXC7325	ERR221461
22363	OXC7324	ERR221460
22362	OXC7323	ERR221459
22361	OXC7322	ERR221458
22360	OXC7321	ERR221457
22359	OXC7320	ERR221456
22358	OXC7319	ERR221455
25552	OXC7318	ERR386416
22357	OXC7316	ERR221454
22356	OXC7315	ERR221453
22355	OXC7314	ERR221452
22354	OXC7313	ERR221451
22353	OXC7312	ERR221450
22352	OXC7311	ERR221449
22351	OXC7310	ERR221448
22350	OXC7309	ERR221447
22349	OXC7308	ERR221446
22348	OXC7307	ERR221445
22347	OXC7306	ERR221444
22346	OXC7305	ERR221443
22345	OXC7304	ERR221442
22344	OXC7303	ERR221441
22343	OXC7302	ERR221440
22342	OXC7301	ERR221439
22341	OXC7300	ERR221438
22340	OXC7299	ERR221437
22339	OXC7298	ERR221436
22338	OXC7297	ERR221435
22337	OXC7296	ERR221434
22336	OXC7295	ERR221433
22335	OXC7294	ERR221432
22334	OXC7293	ERR221431
22333	OXC7292	ERR221430
22332	OXC7291	ERR221429
22331	OXC7290	ERR221428
22330	OXC7289	ERR221427
22329	OXC7288	ERR221426
22328	OXC7287	ERR221425
22327	OXC7286	ERR221424
22326	OXC7285	ERR221423
22325	OXC7284	ERR221422
22324	OXC7283	ERR221421
22323	OXC7282	ERR221420
22322	OXC7281	ERR221419
22321	OXC7280	ERR221418
22320	OXC7279	ERR221417
22319	OXC7278	ERR221416
22318	OXC7277	ERR221415
22317	OXC7276	ERR221414
22316	OXC7275	ERR221413
22315	OXC7274	ERR221412
22314	OXC7273	ERR221411
22313	OXC7272	ERR221410
22312	OXC7271	ERR221409
22311	OXC7270	ERR221408
22310	OXC7269	ERR221407
22309	OXC7268	ERR221406
22308	OXC7267	ERR221405
22307	OXC7266	ERR221404
22306	OXC7265	ERR221403
22305	OXC7264	ERR221402
22304	OXC7263	ERR221401
22303	OXC7262	ERR221400
22302	OXC7261	ERR221399
22301	OXC7260	ERR221398
22300	OXC7259	ERR221397
22299	OXC7258	ERR221396
22298	OXC7257	ERR221395
22297	OXC7256	ERR221394
22296	OXC7255	ERR221393
22295	OXC7253	ERR221392
22294	OXC7252	ERR221391
22293	OXC7251	ERR221390
22292	OXC7250	ERR221389
22291	OXC7249	ERR221388
22290	OXC7248	ERR221387
22289	OXC7247	ERR221386
22288	OXC7246	ERR221385
22287	OXC7245	ERR221384
22285	OXC7243	ERR221382
22284	OXC7242	ERR221381
22283	OXC7241	ERR221380
22730	OXC7240	ERR222753
22729	OXC7239	ERR222752
22728	OXC7238	ERR222751
22727	OXC7237	ERR222750
22726	OXC7236	ERR222749
22725	OXC7235	ERR222748
22724	OXC7234	ERR222747
22723	OXC7233	ERR222746

22722	OXC7232	ERR222745
22721	OXC7231	ERR222744
22720	OXC7230	ERR222743
22719	OXC7228	ERR222742
22718	OXC7227	ERR222741
22717	OXC7226	ERR222740
22716	OXC7225	ERR222739
22715	OXC7224	ERR222738
22714	OXC7223	ERR222737
22713	OXC7222	ERR222736
22712	OXC7221	ERR222735
22711	OXC7220	ERR222734
22710	OXC7219	ERR222733
22709	OXC7218	ERR222732
22708	OXC7217	ERR222731
22707	OXC7216	ERR222730
22706	OXC7215	ERR222729
22705	OXC7214	ERR222728
22704	OXC7213	ERR222727
22703	OXC7212	ERR222726
22702	OXC7211	ERR222725
22701	OXC7210	ERR222724
22700	OXC7209	ERR222723
22699	OXC7208	ERR222722
22698	OXC7207	ERR222721
22697	OXC7206	ERR222720
22696	OXC7205	ERR222719
22695	OXC7204	ERR222718
22694	OXC7203	ERR222717
22693	OXC7202	ERR222716
22692	OXC7201	ERR222715
22691	OXC7200	ERR222714
22690	OXC7199	ERR222713
22689	OXC7198	ERR222712
22688	OXC7197	ERR222711
22687	OXC7196	ERR222710
22686	OXC7195	ERR222709
22685	OXC7194	ERR222708
22684	OXC7193	ERR222707
22683	OXC7192	ERR222706
22682	OXC7191	ERR222705
22681	OXC7190	ERR222704
22680	OXC7189	ERR222703
22679	OXC7188	ERR222702
22678	OXC7187	ERR222701
22677	OXC7186	ERR222700
22676	OXC7185	ERR222699
22675	OXC7184	ERR222698
22674	OXC7183	ERR222697
22673	OXC7182	ERR222696
22672	OXC7181	ERR222695
22671	OXC7180	ERR222694
22670	OXC7179	ERR222693
22669	OXC7178	ERR222692
22668	OXC7177	ERR222691
22667	OXC7176	ERR222690
22666	OXC7175	ERR222689
22665	OXC7174	ERR222688
22664	OXC7173	ERR222687
22662	OXC7171	ERR222685
22661	OXC7170	ERR222684
21250	OXC7169	ERR193217
21249	OXC7168	ERR193216
21248	OXC7167	ERR193215
21247	OXC7166	ERR193214
21246	OXC7165	ERR193213
21245	OXC7164	ERR193212
21244	OXC7163	ERR193211
21243	OXC7162	ERR193210
21242	OXC7161	ERR193209
21241	OXC7160	ERR193208
21240	OXC7159	ERR193207
21239	OXC7158	ERR193206
21238	OXC7157	ERR193205
21237	OXC7156	ERR193204
21236	OXC7155	ERR193203
21235	OXC7154	ERR193202
21234	OXC7153	ERR193201
21232	OXC7151	ERR193199
21231	OXC7150	ERR193198
21230	OXC7149	ERR193197
21229	OXC7148	ERR193196
21228	OXC7147	ERR193195
21227	OXC7146	ERR193194
21226	OXC7145	ERR193193
21225	OXC7144	ERR193192
21224	OXC7143	ERR193191
21223	OXC7142	ERR193190
21222	OXC7141	ERR193189
21221	OXC7140	ERR193188
21220	OXC7139	ERR193187
21219	OXC7138	ERR193186

21217	OXC7136	ERR193184
21216	OXC7135	ERR193183
21215	OXC7134	ERR193182
21214	OXC7133	ERR193181
21213	OXC7132	ERR193180
21212	OXC7131	ERR193179
21211	OXC7130	ERR193178
21210	OXC7129	ERR193177
21209	OXC7128	ERR193176
21208	OXC7127	ERR193175
21207	OXC7125	ERR193174
21206	OXC7124	ERR193173
21205	OXC7123	ERR193172
21204	OXC7122	ERR193171
21203	OXC7121	ERR193170
21202	OXC7120	ERR193169
21201	OXC7119	ERR193168
21199	OXC7117	ERR193166
21198	OXC7116	ERR193165
21197	OXC7115	ERR193164
21196	OXC7114	ERR193163
21195	OXC7113	ERR193162
21194	OXC7112	ERR193161
21193	OXC7111	ERR193160
21192	OXC7110	ERR193159
21191	OXC7109	ERR193158
21190	OXC7108	ERR193157
21189	OXC7107	ERR193156
21188	OXC7106	ERR193155
21187	OXC7105	ERR193154
21186	OXC7104	ERR193153
21185	OXC7103	ERR193152
21184	OXC7102	ERR193151
21183	OXC7101	ERR193150
22282	OXC7100	ERR221379
22281	OXC7099	ERR221378
22280	OXC7098	ERR221377
22279	OXC7097	ERR221376
22278	OXC7096	ERR221375
22277	OXC7095	ERR221374
22275	OXC7093	ERR221372
22274	OXC7092	ERR221371
22273	OXC7091	ERR221370
22272	OXC7090	ERR221369
22660	OXC7089	ERR222683
22271	OXC7087	ERR221368
22270	OXC7086	ERR221367
22269	OXC7085	ERR221366
22268	OXC7084	ERR221365
22267	OXC7083	ERR221364
22265	OXC7081	ERR221362
22264	OXC7080	ERR221361
22263	OXC7079	ERR221360
22262	OXC7078	ERR221359
22261	OXC7077	ERR221358
22260	OXC7076	ERR221357
22259	OXC7075	ERR221356
22258	OXC7074	ERR221355
22257	OXC7073	ERR221354
22256	OXC7072	ERR221353
22255	OXC7071	ERR221352
22254	OXC7070	ERR221351
22253	OXC7069	ERR221350
22252	OXC7068	ERR221349
22251	OXC7067	ERR221348
22250	OXC7066	ERR221347
22249	OXC7065	ERR221346
22248	OXC7064	ERR221345
22247	OXC7063	ERR221344
22246	OXC7062	ERR221343
22245	OXC7061	ERR221342
22244	OXC7060	ERR221341
22243	OXC7058	ERR221340
22242	OXC7057	ERR221339
22241	OXC7056	ERR221338
22239	OXC7054	ERR221336
22238	OXC7053	ERR221335
22237	OXC7052	ERR221334
22235	OXC7050	ERR221332
22234	OXC7049	ERR221331
22233	OXC7048	ERR221330
22232	OXC7047	ERR221329
22231	OXC7046	ERR221328
22230	OXC7045	ERR221327
22229	OXC7044	ERR221326
22228	OXC7043	ERR221325
22226	OXC7041	ERR221323
22225	OXC7039	ERR221322
22224	OXC7038	ERR221321
22223	OXC7037	ERR221320
22222	OXC7036	ERR221319
22221	OXC7035	ERR221318

22220	OXC7034	ERR221317
22219	OXC7033	ERR221316
22218	OXC7032	ERR221315
22217	OXC7031	ERR221314
22216	OXC7030	ERR221313
22215	OXC7029	ERR221312
22214	OXC7027	ERR221311
22213	OXC7026	ERR221310
22212	OXC7025	ERR221309
22211	OXC7024	ERR221308
22209	OXC7022	ERR221306
22207	OXC7020	ERR221304
22206	OXC7019	ERR221303
22205	OXC7018	ERR221302
22204	OXC7017	ERR221301
22203	OXC7016	ERR221300
22202	OXC7015	ERR221299
22201	OXC7014	ERR221298
22200	OXC7013	ERR221297
22199	OXC7012	ERR221296
22198	OXC7011	ERR221295
22197	OXC7010	ERR221294
22196	OXC7009	ERR221293
22195	OXC7008	ERR221292
22194	OXC7007	ERR221291
22193	OXC7006	ERR221290
22192	OXC7005	ERR221289
22191	OXC7004	ERR221288
22190	OXC7003	ERR221287
22189	OXC7002	ERR221286
21172	OXC7001	ERR189869
21171	OXC7000	ERR189868
21170	OXC6999	ERR189867
21169	OXC6998	ERR189866
21168	OXC6997	ERR189865
21167	OXC6996	ERR189864
21165	OXC6994	ERR189862
21164	OXC6992	ERR189861
21163	OXC6991	ERR189831
21162	OXC6989	ERR189830
21161	OXC6988	ERR189829
21160	OXC6987	ERR189828
21158	OXC6985	ERR189826
21157	OXC6984	ERR189825
21156	OXC6983	ERR189824
21155	OXC6982	ERR189823
21154	OXC6981	ERR189822
21153	OXC6980	ERR189821
21152	OXC6979	ERR189820
21151	OXC6978	ERR189819
21150	OXC6977	ERR189818
21149	OXC6976	ERR189817
21147	OXC6974	ERR189815
21146	OXC6973	ERR189814
21143	OXC6969	ERR189811
21142	OXC6968	ERR189810
21141	OXC6967	ERR189809
21140	OXC6966	ERR189808
21139	OXC6965	ERR189807
21138	OXC6964	ERR189806
21137	OXC6963	ERR189805
21136	OXC6962	ERR189804
21135	OXC6961	ERR189803
21134	OXC6960	ERR189802
21133	OXC6959	ERR189801
21132	OXC6958	ERR189800
21131	OXC6957	ERR189799
21130	OXC6956	ERR189798
21129	OXC6955	ERR189797
21128	OXC6954	ERR189796
21127	OXC6953	ERR189795
21126	OXC6952	ERR189794
21125	OXC6951	ERR189793
21124	OXC6950	ERR189792
21123	OXC6949	ERR189791
21122	OXC6948	ERR189790
21121	OXC6947	ERR189789
21120	OXC6946	ERR189788
21119	OXC6945	ERR189787
21118	OXC6944	ERR189786
21117	OXC6943	ERR189785
21116	OXC6942	ERR189784
21115	OXC6941	ERR189783
21114	OXC6940	ERR189782
21113	OXC6939	ERR189781
21112	OXC6938	ERR189780
21111	OXC6937	ERR189779
21110	OXC6936	ERR189778
21109	OXC6935	ERR189777
21108	OXC6934	ERR189776
21107	OXC6933	ERR189775
21106	OXC6932	ERR189774

22125	OXC6930	ERR356131
22124	OXC6929	ERR356130
22123	OXC6928	ERR356129
22122	OXC6927	ERR356128
22121	OXC6926	ERR356127
22120	OXC6925	ERR356126
22119	OXC6924	ERR356125
22118	OXC6923	ERR356124
22117	OXC6922	ERR356123
22116	OXC6921	ERR356122
22115	OXC6920	ERR356121
22114	OXC6919	ERR356120
22113	OXC6918	ERR356119
22112	OXC6917	ERR356118
22111	OXC6916	ERR356117
22110	OXC6915	ERR356116
22109	OXC6914	ERR356115
22108	OXC6913	ERR356114
22107	OXC6912	ERR356113
22105	OXC6910	ERR356111
22104	OXC6909	ERR356110
22103	OXC6908	ERR356109
22102	OXC6907	ERR356108
22101	OXC6906	ERR356107
22100	OXC6905	ERR356106
22099	OXC6904	ERR356105
22098	OXC6903	ERR356104
22097	OXC6902	ERR356103
22096	OXC6901	ERR356102
22095	OXC6900	ERR356101
22094	OXC6899	ERR356100
22093	OXC6898	ERR356099
22092	OXC6897	ERR356098
22090	OXC6895	ERR356096
22089	OXC6894	ERR356095
22088	OXC6893	ERR356094
22087	OXC6892	ERR356093
22086	OXC6891	ERR356092
22085	OXC6890	ERR356091
22084	OXC6888	ERR356090
22083	OXC6887	ERR356089
22082	OXC6886	ERR356088
22081	OXC6885	ERR356087
22080	OXC6884	ERR356086
22079	OXC6883	ERR356085
22078	OXC6882	ERR356084
21600	OXC6881	ERR356083
21599	OXC6880	ERR356082
21598	OXC6879	ERR356081
21597	OXC6878	ERR356080
21596	OXC6877	ERR356079
21595	OXC6876	ERR356078
21594	OXC6875	ERR356077
21593	OXC6874	ERR356076
21592	OXC6873	ERR356075
21591	OXC6872	ERR356074
21590	OXC6871	ERR356073
21589	OXC6870	ERR356072
21588	OXC6869	ERR356071
21587	OXC6868	ERR356070
21586	OXC6867	ERR356069
21585	OXC6866	ERR356068
21584	OXC6865	ERR356067
21583	OXC6864	ERR356066
21582	OXC6863	ERR356065
21581	OXC6862	ERR356064
21580	OXC6861	ERR356063
21579	OXC6860	ERR356062
21578	OXC6859	ERR356061
21577	OXC6858	ERR356060
21576	OXC6857	ERR356059
21575	OXC6856	ERR356058
21574	OXC6855	ERR356057
21573	OXC6854	ERR356056
21572	OXC6853	ERR356055
21571	OXC6852	ERR356054
21570	OXC6851	ERR356053
21569	OXC6850	ERR356052
21568	OXC6849	ERR356051
21567	OXC6848	ERR356050
21566	OXC6847	ERR356049
21565	OXC6846	ERR356048
21564	OXC6845	ERR356047
21563	OXC6844	ERR356046
21562	OXC6843	ERR356045
21561	OXC6842	ERR356044
21560	OXC6841	ERR356043
21559	OXC6840	ERR356042
21558	OXC6839	ERR356041
21557	OXC6838	ERR356040
21556	OXC6837	ERR356039
21555	OXC6836	ERR356038

21554	OXC6835	ERR356037
18390	OXC6834	ERR137094
18389	OXC6833	ERR137093
18388	OXC6832	ERR137092
18387	OXC6831	ERR137091
18386	OXC6830	ERR137090
18385	OXC6828	ERR137089
18384	OXC6827	ERR137088
18383	OXC6826	ERR137087
18382	OXC6825	ERR137086
18381	OXC6824	ERR137085
18380	OXC6823	ERR137084
18378	OXC6821	ERR137082
18377	OXC6820	ERR137081
18376	OXC6819	ERR137080
18375	OXC6818	ERR137079
18374	OXC6817	ERR137078
18373	OXC6816	ERR137077
18372	OXC6815	ERR137076
18371	OXC6814	ERR137075
18370	OXC6813	ERR137074
18369	OXC6812	ERR137073
18368	OXC6811	ERR137072
18367	OXC6810	ERR137071
18366	OXC6809	ERR137070
18365	OXC6808	ERR137069
18364	OXC6807	ERR137068
18363	OXC6806	ERR137067
18362	OXC6805	ERR137066
18361	OXC6804	ERR137065
18360	OXC6803	ERR137064
18359	OXC6802	ERR137063
18358	OXC6801	ERR137062
18357	OXC6800	ERR137061
18356	OXC6799	ERR137060
18355	OXC6798	ERR137059
18354	OXC6797	ERR137058
18353	OXC6796	ERR137057
18352	OXC6795	ERR137056
18350	OXC6793	ERR137054
18348	OXC6791	ERR137052
18347	OXC6790	ERR137051
18346	OXC6789	ERR137050
18345	OXC6788	ERR137049
18344	OXC6787	ERR137048
18343	OXC6786	ERR137047
18342	OXC6785	ERR137046
18340	OXC6783	ERR137044
18338	OXC6781	ERR137042
18337	OXC6780	ERR137041
18336	OXC6779	ERR137040
18335	OXC6778	ERR137039
18334	OXC6777	ERR137038
18333	OXC6776	ERR137037
18331	OXC6774	ERR137035
18330	OXC6773	ERR137034
18329	OXC6772	ERR137033
18328	OXC6771	ERR137032
18327	OXC6770	ERR137031
18326	OXC6769	ERR137030
18325	OXC6768	ERR137029
18324	OXC6767	ERR137028
18323	OXC6766	ERR137027
18322	OXC6765	ERR137026
18321	OXC6764	ERR137025
18320	OXC6763	ERR137024
18319	OXC6762	ERR137023
18318	OXC6761	ERR137022
18317	OXC6760	ERR137021
18316	OXC6759	ERR137020
18314	OXC6757	ERR137018
18313	OXC6756	ERR137017
18312	OXC6755	ERR137016
18311	OXC6754	ERR137015
18310	OXC6753	ERR137014
18309	OXC6752	ERR137013
18308	OXC6751	ERR137012
18307	OXC6750	ERR137011
18306	OXC6749	ERR137010
18305	OXC6748	ERR137009
18304	OXC6747	ERR137008
18303	OXC6746	ERR137007
18302	OXC6745	ERR137006
18301	OXC6744	ERR137005
18300	OXC6743	ERR137004
18299	OXC6742	ERR137003
18298	OXC6741	ERR137002
18297	OXC6740	ERR137001
18296	OXC6739	ERR137000
18295	OXC6738	ERR136999
18294	OXC6737	ERR128822
18293	OXC6736	ERR128821

18292	OXC6735	ERR128820
18291	OXC6734	ERR128819
18290	OXC6733	ERR128818
18289	OXC6732	ERR128817
18288	OXC6731	ERR128816
18287	OXC6730	ERR128815
18286	OXC6729	ERR128814
18285	OXC6728	ERR128813
18284	OXC6727	ERR128812
18283	OXC6726	ERR128811
18282	OXC6725	ERR128810
18281	OXC6724	ERR128809
18280	OXC6723	ERR128808
18279	OXC6722	ERR128807
18278	OXC6721	ERR128806
18276	OXC6719	ERR128804
18275	OXC6718	ERR128803
18273	OXC6716	ERR128801
18272	OXC6715	ERR128800
18271	OXC6714	ERR128799
18270	OXC6713	ERR128798
18268	OXC6711	ERR128796
18267	OXC6710	ERR128795
18266	OXC6709	ERR128794
18265	OXC6708	ERR128793
18264	OXC6707	ERR128792
18263	OXC6706	ERR128791
18262	OXC6705	ERR128790
18260	OXC6703	ERR128788
18259	OXC6702	ERR128787
18258	OXC6701	ERR128786
18257	OXC6700	ERR128785
18256	OXC6699	ERR128784
18255	OXC6698	ERR128783
18254	OXC6697	ERR128782
18253	OXC6696	ERR128781
18252	OXC6695	ERR128780
18251	OXC6694	ERR128779
18250	OXC6693	ERR128778
18248	OXC6691	ERR128776
18247	OXC6690	ERR128775
18246	OXC6689	ERR128774
18245	OXC6688	ERR128773
18242	OXC6685	ERR128770
18241	OXC6684	ERR128769
18240	OXC6683	ERR128768
18239	OXC6682	ERR128767
18238	OXC6681	ERR128766
18237	OXC6680	ERR128765
18236	OXC6679	ERR128764
18235	OXC6678	ERR128763
18234	OXC6677	ERR128762
18233	OXC6676	ERR128761
18232	OXC6675	ERR128760
18231	OXC6674	ERR128759
18230	OXC6673	ERR128758
18229	OXC6672	ERR128757
18228	OXC6671	ERR128756
18226	OXC6669	ERR128754
18225	OXC6668	ERR128753
18224	OXC6667	ERR128752
18223	OXC6666	ERR128751
18222	OXC6665	ERR128750
18221	OXC6664	ERR128749
18220	OXC6663	ERR128748
18219	OXC6662	ERR128747
18218	OXC6661	ERR128746
18217	OXC6660	ERR128745
18216	OXC6659	ERR128744
18214	OXC6657	ERR128742
18213	OXC6656	ERR128741
18212	OXC6655	ERR128740
18211	OXC6654	ERR128739
18210	OXC6653	ERR128738
18209	OXC6652	ERR128737
18208	OXC6651	ERR128736
18207	OXC6650	ERR128735
18206	OXC6649	ERR128734
18205	OXC6648	ERR128733
18204	OXC6647	ERR128732
18203	OXC6646	ERR128731
18202	OXC6645	ERR128730
18201	OXC6644	ERR128729
18200	OXC6643	ERR128728
18199	OXC6642	ERR128727
16390	OXC6641	ERR108403
16389	OXC6640	ERR108402
16388	OXC6639	ERR108401
16387	OXC6638	ERR108400
16386	OXC6637	ERR108399
16385	OXC6636	ERR108398
16384	OXC6635	ERR108397

16383	OXC6634	ERR108396
16382	OXC6633	ERR108395
16381	OXC6632	ERR108394
16380	OXC6631	ERR108393
16379	OXC6630	ERR108392
16378	OXC6629	ERR108391
16377	OXC6628	ERR108390
16376	OXC6627	ERR108389
16375	OXC6626	ERR108388
16374	OXC6625	ERR108387
16373	OXC6624	ERR108386
16372	OXC6623	ERR108385
16371	OXC6622	ERR108384
16370	OXC6621	ERR108383
16369	OXC6620	ERR108382
16368	OXC6619	ERR108381
16367	OXC6618	ERR108380
16366	OXC6616	ERR108379
16365	OXC6615	ERR108378
16364	OXC6614	ERR108377
16363	OXC6613	ERR108376
16362	OXC6612	ERR108375
16361	OXC6611	ERR108374
16359	OXC6609	ERR108372
16358	OXC6608	ERR108371
16357	OXC6607	ERR108370
22905	OXC6605	ERR108369
16355	OXC6604	ERR108368
16354	OXC6603	ERR108367
16353	OXC6602	ERR108366
16352	OXC6601	ERR108365
16351	OXC6600	ERR108364
16350	OXC6599	ERR108363
16349	OXC6598	ERR108362
16348	OXC6597	ERR108361
16347	OXC6596	ERR108360
16346	OXC6595	ERR108359
16345	OXC6594	ERR108358
21507	OXC6593	ERR108357
16343	OXC6592	ERR108356
16342	OXC6591	ERR108355
16341	OXC6590	ERR108354
12924	OXC6587	ERR108351
12923	OXC6586	ERR108350
12922	OXC6585	ERR108349
12921	OXC6584	ERR108348
12920	OXC6583	ERR108347
12919	OXC6582	ERR108346
12918	OXC6581	ERR108345
12916	OXC6579	ERR108343
12913	OXC6578	ERR108342
12912	OXC6577	ERR108341
12911	OXC6576	ERR108340
12909	OXC6574	ERR108338
12908	OXC6573	ERR108337
12906	OXC6571	ERR108335
12905	OXC6570	ERR108334
12904	OXC6569	ERR108333
12903	OXC6568	ERR108332
12902	OXC6567	ERR108331
12901	OXC6566	ERR108330
12900	OXC6565	ERR108329
12899	OXC6564	ERR108328
12898	OXC6563	ERR108327
12897	OXC6562	ERR108326
12896	OXC6561	ERR108325
12895	OXC6560	ERR108324
12894	OXC6559	ERR108323
12893	OXC6558	ERR108322
12892	OXC6557	ERR108321
12891	OXC6556	ERR108320
12888	OXC6553	ERR108317
12887	OXC6552	ERR108316
12886	OXC6551	ERR108315
12885	OXC6550	ERR108314
12884	OXC6549	ERR108313
12883	OXC6548	ERR108312
12882	OXC6547	ERR108311
12881	OXC6546	ERR108310
12880	OXC6545	ERR108309
10935	OXC6544	ERR108308
16335	OXC6543	ERR084154
16334	OXC6542	ERR084153
16333	OXC6541	ERR084152
16332	OXC6540	ERR084151
16331	OXC6539	ERR084150
16330	OXC6538	ERR084149
16329	OXC6537	ERR084148
16328	OXC6536	ERR084147
16327	OXC6535	ERR084146
16326	OXC6534	ERR084145
16325	OXC6533	ERR084144

16324	OXC6532	ERR084143
16323	OXC6531	ERR084142
16322	OXC6530	ERR084141
16321	OXC6529	ERR084140
16320	OXC6528	ERR084139
16319	OXC6527	ERR084138
16318	OXC6526	ERR084137
16317	OXC6525	ERR084136
16316	OXC6524	ERR084135
16315	OXC6523	ERR084134
16314	OXC6522	ERR084133
16313	OXC6521	ERR084132
16312	OXC6520	ERR084131
16311	OXC6519	ERR084130
16310	OXC6518	ERR084129
16309	OXC6517	ERR084128
16308	OXC6516	ERR084127
16307	OXC6515	ERR084126
16306	OXC6514	ERR084125
16305	OXC6513	ERR084124
16304	OXC6512	ERR084123
16303	OXC6511	ERR084122
16302	OXC6510	ERR084121
16301	OXC6509	ERR084120
16300	OXC6508	ERR084119
16299	OXC6507	ERR084118
16298	OXC6506	ERR084117
16297	OXC6505	ERR084116
16296	OXC6504	ERR084115
16295	OXC6503	ERR084114
16294	OXC6502	ERR084113
16293	OXC6501	ERR084112
16292	OXC6500	ERR084111
16291	OXC6499	ERR084110
16290	OXC6498	ERR084109
16289	OXC6497	ERR084108
16288	OXC6496	ERR084107
16287	OXC6495	ERR084106
16286	OXC6494	ERR084105
16285	OXC6493	ERR084104
16284	OXC6492	ERR084103
16283	OXC6491	ERR084102
16282	OXC6490	ERR084101
16281	OXC6489	ERR084100
16280	OXC6488	ERR084099
16279	OXC6487	ERR084098
16278	OXC6486	ERR084097
16276	OXC6484	ERR084095
16275	OXC6483	ERR084094
16274	OXC6482	ERR084093
16273	OXC6481	ERR084092
16272	OXC6480	ERR084091
16271	OXC6479	ERR084090
16270	OXC6478	ERR084089
16269	OXC6477	ERR084088
16268	OXC6476	ERR084087
16267	OXC6475	ERR084086
16266	OXC6474	ERR084085
16265	OXC6473	ERR084084
16263	OXC6471	ERR084082
16262	OXC6470	ERR084081
16261	OXC6469	ERR084080
16260	OXC6468	ERR084079
16259	OXC6467	ERR084078
16258	OXC6466	ERR084077
16257	OXC6465	ERR084076
16256	OXC6464	ERR084075
16255	OXC6463	ERR084074
16253	OXC6461	ERR084072
16252	OXC6460	ERR084071
16251	OXC6459	ERR084070
16250	OXC6458	ERR084069
16249	OXC6457	ERR084068
16248	OXC6456	ERR084067
16247	OXC6455	ERR084066
16246	OXC6454	ERR084065
16245	OXC6453	ERR084064
16244	OXC6451	ERR084063
16243	OXC6450	ERR084062
16242	OXC6449	ERR084061
16241	OXC6448	ERR084060
16240	OXC6447	ERR084059
16239	OXC6446	ERR084058
16236	OXC6442	ERR084055
16235	OXC6441	ERR084054
16234	OXC6440	ERR084053
16233	OXC6438	ERR084052
16232	OXC6437	ERR084051
16231	OXC6436	ERR084050
16229	OXC6434	ERR084048
16228	OXC6433	ERR084047
16227	OXC6432	ERR084046

16226	OXC6431	ERR084045
16225	OXC6430	ERR084044
16224	OXC6429	ERR084043
16223	OXC6428	ERR084042
16222	OXC6427	ERR084041
16221	OXC6426	ERR084040
16220	OXC6425	ERR084039
16219	OXC6424	ERR084038
16218	OXC6423	ERR084037
16217	OXC6422	ERR084036
16216	OXC6421	ERR084035
16215	OXC6420	ERR084034
16214	OXC6419	ERR084033
16213	OXC6418	ERR084032
16212	OXC6417	ERR084031
16211	OXC6416	ERR084030
16210	OXC6415	ERR084029
16209	OXC6414	ERR084028
16208	OXC6413	ERR084027
16207	OXC6412	ERR084026
16206	OXC6411	ERR084025
16205	OXC6410	ERR084024
16204	OXC6409	ERR084023
16203	OXC6408	ERR084022
16202	OXC6407	ERR084021
16201	OXC6406	ERR084020
16200	OXC6405	ERR084019
16198	OXC6403	ERR084017
16197	OXC6402	ERR084016
16196	OXC6401	ERR084015
16195	OXC6400	ERR084014
16194	OXC6397	ERR084013
16193	OXC6396	ERR084012
16192	OXC6395	ERR084011
16191	OXC6394	ERR084010
16190	OXC6393	ERR084009
16188	OXC6391	ERR084007
16187	OXC6390	ERR084006
16186	OXC6389	ERR084005
16184	OXC6387	ERR084003
16183	OXC6386	ERR084002
16182	OXC6385	ERR084001
16181	OXC6384	ERR084000
16180	OXC6383	ERR083999
16179	OXC6382	ERR083998
16178	OXC6381	ERR083997
16177	OXC6380	ERR083996
16175	OXC6378	ERR083994
16173	OXC6376	ERR083992
16172	OXC6375	ERR083991
16171	OXC6374	ERR083990
16170	OXC6373	ERR083989
16169	OXC6372	ERR083988
16168	OXC6371	ERR083987
16167	OXC6370	ERR083986
16166	OXC6369	ERR083985
16165	OXC6368	ERR083984
16164	OXC6367	ERR083983
16163	OXC6366	ERR083982
16162	OXC6365	ERR083981
16161	OXC6364	ERR083980
16160	OXC6363	ERR083979
16159	OXC6362	ERR083978
16158	OXC6361	ERR083977
16157	OXC6360	ERR083976
16156	OXC6359	ERR083975
16155	OXC6358	ERR083974
16154	OXC6357	ERR083973
16153	OXC6356	ERR083972
16152	OXC6355	ERR083971
16151	OXC6354	ERR083970
16150	OXC6353	ERR083969
16149	OXC6352	ERR083968
16148	OXC6351	ERR083967
16147	OXC6350	ERR083966
16146	OXC6349	ERR083965
16145	OXC6348	ERR083964
16143	OXC6346	ERR083962
16141	OXC6344	ERR083960
16140	OXC6343	ERR083959
16139	OXC6342	ERR083958
16138	OXC6341	ERR083957
16137	OXC6340	ERR083956
16136	OXC6339	ERR083955
16135	OXC6338	ERR083954
16134	OXC6337	ERR083953
16132	OXC6335	ERR083951
16131	OXC6334	ERR083950
16130	OXC6333	ERR083949
16129	OXC6332	ERR083948
16128	OXC6331	ERR083947
16127	OXC6330	ERR083946

16126	OXC6329	ERR083945
16125	OXC6328	ERR083944
16124	OXC6327	ERR083943
16123	OXC6326	ERR083942
16122	OXC6325	ERR083941
16121	OXC6324	ERR083940
16120	OXC6323	ERR083939
16119	OXC6322	ERR083938
16118	OXC6321	ERR083937
16117	OXC6320	ERR083936
16116	OXC6319	ERR083935
16115	OXC6318	ERR083934
16114	OXC6317	ERR083933
16113	OXC6316	ERR083932
16112	OXC6315	ERR083931
16111	OXC6314	ERR083930
16110	OXC6313	ERR083929
16109	OXC6312	ERR083928
16108	OXC6311	ERR083927
16107	OXC6310	ERR083926
16106	OXC6309	ERR083925
16105	OXC6308	ERR083924
16104	OXC6307	ERR083923
16103	OXC6306	ERR083922
16102	OXC6305	ERR083921
16101	OXC6304	ERR083920
16100	OXC6303	ERR083919
16098	OXC6301	ERR083917
16097	OXC6300	ERR083916
16095	OXC6298	ERR083914
16094	OXC6297	ERR083913
16093	OXC6296	ERR083912
16091	OXC6294	ERR083910
16090	OXC6293	ERR083909
16088	OXC6291	ERR083907
16087	OXC6290	ERR083906
16086	OXC6289	ERR083905
16085	OXC6288	ERR083904
16084	OXC6287	ERR083903
16083	OXC6286	ERR083902
16082	OXC6285	ERR083901
16081	OXC6284	ERR083900
16080	OXC6283	ERR083899
16079	OXC6282	ERR083898
16078	OXC6281	ERR083897
16076	OXC6279	ERR083896
16075	OXC6278	ERR083895
16074	OXC6277	ERR083894
16073	OXC6276	ERR083893
16072	OXC6275	ERR083892
16071	OXC6274	ERR083891
16069	OXC6272	ERR083889
16068	OXC6271	ERR083888
16067	OXC6270	ERR083887
16066	OXC6269	ERR083886
16065	OXC6268	ERR083885
16064	OXC6267	ERR083884
16063	OXC6266	ERR083883
16062	OXC6265	ERR083882
16061	OXC6264	ERR083881
16060	OXC6263	ERR083880
16059	OXC6262	ERR083879
16058	OXC6261	ERR083878
16057	OXC6260	ERR083877
16056	OXC6259	ERR083876
16055	OXC6258	ERR083875
16054	OXC6257	ERR083874
16053	OXC6256	ERR083873
16336	OXC6255	ERR083872
16052	OXC6254	ERR083871
16051	OXC6253	ERR083870
16050	OXC6252	ERR083869
16049	OXC6251	ERR083868
16048	OXC6250	ERR083867

Appendix B

Table B1 Details of isolates used to determine membership of *C. coli* clades 1, 2, and 3

PubMLST id	Isolate	Clade	Reference
2473	RM4661	1	-
24266	8808	2	(Sheppard *et al.*, 2013b)
24267	2544	2	(Sheppard *et al.*, 2013b)
24268	6873	2	(Sheppard *et al.*, 2013b)
24269	dfvf1656	2	(Sheppard *et al.*, 2013b)
24270	FSA05.280042	3	(Sheppard *et al.*, 2013b)
24271	8096	3	(Sheppard *et al.*, 2013b)
24272	dfvf1912	3	(Sheppard *et al.*, 2013b)
24273	4944	3	(Sheppard *et al.*, 2013b)
24274	RM4931	3	(Sheppard *et al.*, 2013b)
32046	FB1	1	-
32237	CVM N29710	1	(Chen *et al.*, 2013)
32255	RM1875	1	-
32289	RM5611	1	-
32298	15-537360	1	(Pearson *et al.*, 2013)

Appendix C

Table C1 Provenance data for publicly available genomes included in the *Campylobacter* core genome study

Species	Isolate	PubMLST id	GenBank accession number	Culture collection designations	Year	Country	Source
C. coli	RM1875	32255	CP007183.1	ATCC 33559; CCUG 11283; DSM 4689; LMG 6440; NCTC 11366			Pig
	RM4661	2473	CP007181.1		2003	USA	Turkey
	FB1	32046	CP011015.1		1992	UK	Human (stool)
	CVM N29710	32237	CP004066.1		2011	USA	Chicken meat (retail)
	RM5611	32289	CP007179.1			USA	Cattle
	15-537360	32298	CP006702.1		2001	UK	Human (stool)
C. concisus	13826	34087	CP000792.1			Denmark	Human (stool)
C. corcagiensis	CIT 045	34127	JFAP00000000.1	CCUG 64942; LMG 27932	2013	Ireland	Lion-tailed macaque
C. cuniculorum	DSM 23162	34128	JHZL00000000.1	CCUG 56289; LMG 24588	2005	Italy	Rabbit
C. curvus	525.92	34088	CP000767.1		1992	South Africa	Human (stool)
C. fetus subsp. *fetus*	04/554	34089	CP008808.1			Argentina	Cattle (foetus)
	82-40	34090	CP000487.1		1981	USA	Human (blood culture)
C. fetus subsp. *testudinum*	03-427	34092	CP006833.1	ATCC BAA-2539; LMG 27499		USA	Human (blood culture)
	pet-3	34091	CP009226.1			Taiwan	Philippine sailfin lizard
C. fetus subsp. *venerealis*	97/608	34093	CP008810.1			Argentina	Cattle (placenta)
	cfvi03/293	34094	CP006999.1			Argentina	Cattle (foetus)
	84-112	34095	HG004426.1			USA	Cattle
C. gracilis	RM3268	34129	ACYG00000000.1	ATCC 33236; CCUG 27720; NCTC 12738		USA	Human (gingival crevice)
C. hominis	ATCC BAA-381	34096	CP000776.1	CCUG 45161; LMG 19568; NCTC 13146			Human (stool; carrier)

C. hyointestinalis subsp. *hyointestinalis*	DSM 19053	34126	JHQP00000000.1	ATCC 35217; CCUG 14169; LMG 7817; NCTC 11608	1985	USA	Pig
C. iguaniorum	1485E	34097	CP009043.1	CCUG 66346; LMG 28143			Bearded dragon
C. insulaenigrae	NCTC 12927	34098	CP007770.1	CCUG 48653		UK	Marine mammal
C. jejuni subsp. *doylei*	269.97	4312	CP000768.1		1997	South Africa	Human (blood culture)
C. jejuni subsp. *jejuni*	NCTC11168	48	AL111168.1		1977	UK	Human (stool)
	81116	1220	CP000814.1		1981	UK	Human (stool; general outbreak)
	81-176	1481	CP000538.1		1981	USA	Human (stool; general outbreak)
	00-1597	2245	CP010306.1		2000	Canada	Human (stool)
	IA3902	5293	CP001876.1		2006	USA	Sheep (abortion; general oubreak)
	F38011	5328	CP006851.1			USA	Human
	RM1221	18182	CP000025.1			USA	Chicken
	S3	18183	CP001960.1			USA	Chicken
	M1	18184	CP001900.1				Human (stool)
	32488	34099	CP006006.1			USA	Human
	4031	34100	HG428754.1		2000	Finland	Drinking water
	MTVDSCj20	34101	CP008787.1		2013	USA	Chicken
	YH001	34102	CP010058.1		2014	USA	Cattle (liver, retail)
	00-6200	34103	CP010307.1		2000	Canada	Human
	01-1512	34104	CP010072.1		2001	Canada	Human
	00-0949	34105	CP010301.1		2000	Canada	Human
	35925B2	34106	CP010906.1		1985	USA	Human
	00-2425	34107	CP006729.2		2000	Canada	Human (stool)
	00-2426	34108	CP006708.2		2000	Canada	Human (stool)
	00-2538	34109	CP006707.2		2000	Canada	Human (stool)
	00-2544	34110	CP006709.2		2000	Canada	Human (stool)
	CG8421	34111	CP005388.1		1999	Thailand	Human (stool)

	ICDCCJ07001	34112	CP002029.1			China	Human (stool; Guillain Barre Syndrome)
	PT14	34113	CP003871.3				
	R14	34114	CP005081.1				
C. lari	Slaughter Beach	34115	CP011372.1			USA	Environmental waters
	CCUG 22395	34116	CP007776.1			France	Human
	NCTC 11845	34117	CP007775.1			UK	Environmental waters
	RM16701	34118	CP007777.1		2013	USA	Environmental waters
	RM16712	34119	CP007778.1		2013	USA	Environmental waters
	RM2100	34120	CP000932.1		1982	USA	Human (stool)
C. lari subsp. *concheus*	LMG 11760	34121	CP007771.1			Canada	Human
C. mucosalis	DSM 21682	34130	JHQQ00000000.1	ATCC 43264; CCUG 6822; NCTC 11000	1974	UK	Pig
C. peloridis	LMG 23910	34122	CP007766.1	CCUG 55787		The Netherlands	Shellfish
C. rectus	RM3267	34131	ACFU00000000.1	ATCC 33238; CCUG 20446 B; DSM 3260; NCTC 11489			Human (periodontal pocket)
C. showae	RM3277	34134	ACVQ00000000.1	ATCC 51146; CCUG 30254; LMG 12635		Japan	Human (gingival crevice)
	CC57C	34132	AOTD00000000.1			Canada	Human (colon cancer tumour)
	UNSWCD	34133	AMZQ00000000.1			Australia	Human
C. sputorum subsp. *sputorum* biovar *sputorum*	INTA08/209	34135	JMTI00000000.1		2008	Argentina	Cattle (semen)
C. subantarcticus	LMG 24377	34124	CP007773.1	CCUG 38513		South Georgia, Antarctica	Grey-headed albatross
	LMG 24374	34123	CP007772.1			South Georgia, Antarctica	Gentoo penguin
C. upsaliensis	DSM 5365	34136	JHZN00000000.1	ATCC 43954; CCUG 14913; NCTC 11541	1980	Sweden	Dog
	JV21	34137	AEPU00000000.1				Human

	RM3195	34138	AAFJ00000000.1		1994	South Africa	Human (stool; Guillain Barre Syndrome)
C. ureolyticus	DSM 20703	34140	ARGD00000000.1	ATCC 33387; CCUG 7319; LMG 6451; NCTC 10941		Canada	Human (amniotic fluid)
	ACS-301-V-Sch3b	34139	AGYD00000000.1				Human
C. volucris	LMG 24379	34125	CP007774.1			Sweden	Black-headed gull

Appendix D

Table D1 Loci comprising the genus-wide *Campylobacter* core genome

PubMLST locus	NCTC 11168 identifier	PubMLST locus	NCTC 11168 identifier
CAMP0001	*dnaA*	CAMP0103	*slyD*
CAMP0002	*dnaN*	CAMP0104	*fabD*
CAMP0003	*gyrB*	CAMP0105	*pfs*
CAMP0006	Cj0006	CAMP0106	Cj0118
CAMP0010	*rnhB*	CAMP0108	Cj0120
CAMP0013	*ilvD*	CAMP0109	Cj0121
CAMP0015	Cj0015c	CAMP0111	Cj0123c
CAMP0016	Cj0016	CAMP0115	*accD*
CAMP0022	Cj0022c	CAMP0117	Cj0129c
CAMP0023	*purB*	CAMP0118	*tyrA*
CAMP0024	*nrdA*	CAMP0119	Cj0131
CAMP0026	*thyX*	CAMP0120	*lpxC*
CAMP0027	*pyrG*	CAMP0121	Cj0133
CAMP0028	*recJ*	CAMP0122	*thrB*
CAMP0038	*typA*	CAMP0125	*rbfA*
CAMP0045	*trmU*	CAMP0126	Cj0138
CAMP0046	Cj0054c	CAMP0129	Cj0141c
CAMP0052	*fliM*	CAMP0130	Cj0142c
CAMP0053	*fliA*	CAMP0134	*trxB*
CAMP0055	Cj0063c	CAMP0135	*trxA*
CAMP0056	*flhF*	CAMP0136	Cj0148c
CAMP0057	*folK*	CAMP0137	*hom*
CAMP0058	*aroQ*	CAMP0138	Cj0150c
CAMP0059	Cj0067	CAMP0141	Cj0153c
CAMP0060	*pspA*	CAMP0142	Cj0154c
CAMP0074	*ung*	CAMP0147	Cj0159c
CAMP0075	*aspA*	CAMP0148	Cj0160c
CAMP0076	*dcuA*	CAMP0149	*moaA*
CAMP0084	Cj0096	CAMP0156	*sodB*
CAMP0086	*fmt*	CAMP0158	Cj0172c
CAMP0088	Cj0100	CAMP0172	Cj0186c
CAMP0089	Cj0101	CAMP0173	*purN*
CAMP0093	*atpA*	CAMP0176	Cj0190c
CAMP0094	*atpG*	CAMP0177	*def*
CAMP0095	*atpD*	CAMP0178	*clpP*
CAMP0096	*atpC*	CAMP0179	*tig*
CAMP0097	*exbB3*	CAMP0180	*folE*
CAMP0098	*exbD3*	CAMP0181	*fliI*
CAMP0101	*pal*	CAMP0182	*purF*

CAMP0183	*dapB*
CAMP0184	Cj0198c
CAMP0192	*thrS*
CAMP0193	*infC*
CAMP0194	Cj0208
CAMP0199	*pcm*
CAMP0202	*nrdB*
CAMP0204	*pyrE*
CAMP0205	*frr*
CAMP0206	*secG*
CAMP0207	Cj0236c
CAMP0208	*cynT*
CAMP0209	Cj0238
CAMP0210	Cj0239c
CAMP0211	*iscS*
CAMP0219	Cj0249
CAMP0222	*moaC*
CAMP0225	*exoA*
CAMP0228	Cj0258
CAMP0238	Cj0268c
CAMP0239	*ilvE*
CAMP0240	Cj0270
CAMP0241	Cj0271
CAMP0242	Cj0272
CAMP0243	*fabZ*
CAMP0244	*lpxA*
CAMP0245	*clpX*
CAMP0246	*mreB*
CAMP0248	*carB*
CAMP0249	Cj0280
CAMP0251	*serB*
CAMP0252	*cheW*
CAMP0253	*cheA*
CAMP0254	*cheV*
CAMP0256	*greA*
CAMP0257	*lpxB*
CAMP0259	*surE*
CAMP0260	Cj0294
CAMP0267	*modB*
CAMP0277	*rplY*
CAMP0278	*pth*
CAMP0280	*lysA*
CAMP0282	*pheA*
CAMP0283	*hisC*
CAMP0285	*fliG*
CAMP0287	*dxs*
CAMP0288	*perR*
CAMP0290	*ubiE*
CAMP0291	*xseA*
CAMP0292	*serC*
CAMP0294	*fabH*
CAMP0295	*plsX*
CAMP0298	*ndk*
CAMP0299	*fdxA*
CAMP0300	*ahpC*
CAMP0304	*polA*
CAMP0308	*uvrA*
CAMP0314	*trpB*
CAMP0317	*fliN*
CAMP0319	Cj0353c
CAMP0321	Cj0355c
CAMP0323	Cj0357c
CAMP0325	*glmM*
CAMP0326	*lspA*
CAMP0327	Cj0362
CAMP0328	*hemN*
CAMP0331	*cmeB*
CAMP0334	Cj0369c
CAMP0336	Cj0371
CAMP0337	Cj0372
CAMP0338	Cj0373
CAMP0339	Cj0374
CAMP0342	Cj0377
CAMP0346	*pyrF*
CAMP0347	*nusB*
CAMP0348	*ribH*
CAMP0349	*kdsA*
CAMP0351	Cj0386
CAMP0353	*trpS*
CAMP0354	*serS*
CAMP0356	Cj0391c
CAMP0357	*pyk*
CAMP0358	*mqo*
CAMP0359	Cj0394c
CAMP0363	*gatC*
CAMP0365	*fur*
CAMP0366	*lysS*
CAMP0367	*glyA*
CAMP0368	Cj0403
CAMP0370	*aroE*

CAMP0372	*lgt*
CAMP0374	*frdA*
CAMP0375	*frdB*
CAMP0384	Cj0419
CAMP0385	Cj0420
CAMP0391	Cj0426
CAMP0397	*murD*
CAMP0399	*pgm*
CAMP0400	*fabG*
CAMP0406	*acpP*
CAMP0407	*fabF*
CAMP0408	*accA*
CAMP0411	Cj0449c
CAMP0413	*rep*
CAMP0414	*dnaQ*
CAMP0415	*thiC*
CAMP0419	Cj0457c
CAMP0420	*miaB*
CAMP0421	Cj0459c
CAMP0422	*nusA*
CAMP0423	Cj0461c
CAMP0424	Cj0462
CAMP0425	Cj0463
CAMP0426	*recG*
CAMP0432	*tuf*
CAMP0434	*secE*
CAMP0435	*nusG*
CAMP0440	*rpoB*
CAMP0455	*fusA*
CAMP0458	Cj0496
CAMP0460	*trpC*
CAMP0461	Cj0499
CAMP0466	*alaS*
CAMP0468	*pbpA*
CAMP0471	Cj0511
CAMP0472	*purC*
CAMP0473	*purS*
CAMP0474	*purQ*
CAMP0476	*plsC*
CAMP0485	*fliE*
CAMP0486	*flgC*
CAMP0487	*flgB*
CAMP0488	Cj0529c
CAMP0494	*oorD*
CAMP0495	*oorA*
CAMP0496	*oorB*
CAMP0497	*oorC*
CAMP0500	Cj0541
CAMP0501	*hemA*
CAMP0502	*proS*
CAMP0504	*hemC*
CAMP0505	*ubiD*
CAMP0508	*fliS*
CAMP0510	*efp*
CAMP0518	Cj0559
CAMP0521	*dnaB*
CAMP0530	*ribA*
CAMP0532	*ilvI*
CAMP0533	*ilvH*
CAMP0534	*lpxD*
CAMP0535	*queA*
CAMP0536	*tatC*
CAMP0537	*tatB*
CAMP0539	*nudH*
CAMP0540	*lysC*
CAMP0541	Cj0583
CAMP0543	*folP*
CAMP0544	*ligA*
CAMP0546	*tlyA*
CAMP0548	Cj0590
CAMP0553	*nth*
CAMP0555	*fba*
CAMP0559	Cj0601c
CAMP0565	Cj0607
CAMP0577	Cj0619
CAMP0581	*hypB*
CAMP0582	*hypC*
CAMP0583	*hypD*
CAMP0584	*hypE*
CAMP0588	Cj0631c
CAMP0589	*ilvC*
CAMP0591	*dprA*
CAMP0592	Cj0635
CAMP0595	*ppa*
CAMP0596	*adk*
CAMP0597	*aspS*
CAMP0598	Cj0641
CAMP0599	*recN*
CAMP0601	Cj0644
CAMP0603	Cj0646

CAMP0604	Cj0647
CAMP0607	Cj0650
CAMP0609	*pbpC*
CAMP0613	*era*
CAMP0614	*hslU*
CAMP0615	*hslV*
CAMP0617	*argG*
CAMP0618	Cj0667
CAMP0619	Cj0668
CAMP0620	Cj0669
CAMP0622	*dcuB*
CAMP0626	*uvrB*
CAMP0627	Cj0681
CAMP0630	*priA*
CAMP0632	*ispG*
CAMP0633	*flgH*
CAMP0634	*pta*
CAMP0637	Cj0691
CAMP0639	*mraW*
CAMP0641	*ftsA*
CAMP0642	*ftsZ*
CAMP0643	*flgG2*
CAMP0644	*flgG*
CAMP0645	*glnA*
CAMP0646	Cj0700
CAMP0647	Cj0701
CAMP0648	*purE*
CAMP0649	Cj0703
CAMP0650	*glyQ*
CAMP0651	Cj0705
CAMP0652	Cj0706
CAMP0653	*kdtA*
CAMP0654	Cj0708
CAMP0655	*ffh*
CAMP0657	Cj0711
CAMP0659	*trmD*
CAMP0662	Cj0716
CAMP0664	*dnaE*
CAMP0665	Cj0719c
CAMP0666	*flaC*
CAMP0667	Cj0721c
CAMP0671	Cj0725c
CAMP0680	*hisJ*
CAMP0693	*grpE*
CAMP0694	*dnaK*
CAMP0697	*aspB*
CAMP0698	*cysE*
CAMP0699	*speA*
CAMP0700	*hisS*
CAMP0702	*coaD*
CAMP0703	Cj0768c
CAMP0708	Cj0773c
CAMP0709	Cj0774c
CAMP0710	*valS*
CAMP0712	Cj0777
CAMP0714	*tpx*
CAMP0715	*napA*
CAMP0716	*napG*
CAMP0717	*napH*
CAMP0722	Cj0787
CAMP0723	Cj0788
CAMP0724	*cca*
CAMP0728	*flgS*
CAMP0730	*murF*
CAMP0731	Cj0796c
CAMP0732	Cj0797c
CAMP0733	*ddlA*
CAMP0734	*ruvA*
CAMP0736	Cj0801
CAMP0738	*msbA*
CAMP0739	*pyrD*
CAMP0740	Cj0805
CAMP0742	Cj0807
CAMP0744	Cj0809c
CAMP0745	*nadE*
CAMP0748	*kdsB*
CAMP0755	*fliP*
CAMP0756	*glmU*
CAMP0761	Cj0826
CAMP0763	*ilvA*
CAMP0766	*trmA*
CAMP0768	Cj0833c
CAMP0773	*metS*
CAMP0774	Cj0839c
CAMP0775	*fbp*
CAMP0776	*mobB*
CAMP0779	Cj0844c
CAMP0780	*gltX*
CAMP0782	*psd*
CAMP0783	Cj0848c

CAMP0785	Cj0850c
CAMP0787	Cj0852c
CAMP0788	*hemL*
CAMP0790	*folD*
CAMP0791	*lepP*
CAMP0793	*murA*
CAMP0808	Cj0880c
CAMP0809	Cj0881c
CAMP0810	*flhA*
CAMP0811	Cj0883c
CAMP0815	Cj0888c
CAMP0818	*serA*
CAMP0821	*ispH*
CAMP0822	*aroA*
CAMP0823	*pheT*
CAMP0824	*pheS*
CAMP0825	Cj0898
CAMP0828	Cj0901
CAMP0829	*glnQ*
CAMP0831	Cj0904c
CAMP0832	*alr*
CAMP0838	*cysM*
CAMP0840	*ciaB*
CAMP0841	Cj0915
CAMP0844	*prsA*
CAMP0851	*rpiB*
CAMP0852	Cj0926
CAMP0854	Cj0928
CAMP0855	Cj0929
CAMP0856	Cj0930
CAMP0858	*pckA*
CAMP0859	*pycB*
CAMP0862	*atpE*
CAMP0867	Cj0941c
CAMP0868	*secA*
CAMP0873	Cj0947c
CAMP0875	Cj0949c
CAMP0879	*purH*
CAMP0881	*purL*
CAMP0882	*trmE*
CAMP0883	Cj0957c
CAMP0884	Cj0958c
CAMP0885	Cj0959c
CAMP0886	*rnpA*
CAMP0899	Cj0976

CAMP0914	Cj0991c
CAMP0915	Cj0992c
CAMP0918	*hemB*
CAMP0924	*rpoD*
CAMP0928	Cj1005c
CAMP0929	Cj1006c
CAMP0930	Cj1007c
CAMP0931	*aroB*
CAMP0933	*tgt*
CAMP0945	Cj1022c
CAMP0946	*asd*
CAMP0947	*flgR*
CAMP0949	Cj1026c
CAMP0950	*gyrA*
CAMP0951	Cj1028c
CAMP0953	*lepA*
CAMP0958	*ate*
CAMP0960	*pycA*
CAMP0961	Cj1038
CAMP0962	*murG*
CAMP0967	*thiH*
CAMP0968	*thiG*
CAMP0970	*thiS*
CAMP0971	*dapE*
CAMP0975	*mutS*
CAMP0977	*murC*
CAMP0979	Cj1056c
CAMP0980	Cj1057c
CAMP0981	*guaB*
CAMP0982	*gatA*
CAMP0984	*ileS*
CAMP0988	*pgsA*
CAMP0989	Cj1068
CAMP0994	*lon*
CAMP0995	Cj1074c
CAMP1004	Cj1083c
CAMP1006	*mfd*
CAMP1013	*secF*
CAMP1014	*secD*
CAMP1015	Cj1094c
CAMP1017	*metK*
CAMP1018	Cj1097
CAMP1019	*pyrB*
CAMP1020	Cj1099
CAMP1022	Cj1101

CAMP1023	*truB*
CAMP1024	*csrA*
CAMP1026	*smpB*
CAMP1028	*clpS*
CAMP1030	*aat*
CAMP1034	Cj1113
CAMP1035	*pssA*
CAMP1037	*ftsH*
CAMP1038	*prmA*
CAMP1039	*cheY*
CAMP1042	*pglE*
CAMP1045	*pglC*
CAMP1047	*pglB*
CAMP1048	*pglJ*
CAMP1052	*gne*
CAMP1053	Cj1132c
CAMP1069	*hldE*
CAMP1070	*hldD*
CAMP1071	*gmhB*
CAMP1073	Cj1154c
CAMP1075	*rho*
CAMP1081	Cj1162c
CAMP1087	Cj1168c
CAMP1090	*ppi*
CAMP1091	Cj1172c
CAMP1094	*argS*
CAMP1096	*gmk*
CAMP1098	*fliR*
CAMP1099	Cj1180c
CAMP1100	*tsf*
CAMP1103	*petC*
CAMP1104	*petB*
CAMP1105	*petA*
CAMP1107	*gidA*
CAMP1115	*gpsA*
CAMP1116	*gatB*
CAMP1123	*atpB*
CAMP1124	*radA*
CAMP1125	*ftsY*
CAMP1128	Cj1209
CAMP1131	*rbn*
CAMP1132	*glcD*
CAMP1134	Cj1215
CAMP1139	*groES*
CAMP1140	*groEL*
CAMP1146	Cj1227c
CAMP1147	*htrA*
CAMP1148	*cbpA*
CAMP1149	*hspR*
CAMP1153	*glyS*
CAMP1162	Cj1243
CAMP1163	Cj1244
CAMP1165	*uvrC*
CAMP1167	*guaA*
CAMP1169	*purD*
CAMP1172	*pnp*
CAMP1179	*dnaJ*
CAMP1180	*racR*
CAMP1181	*racS*
CAMP1182	*recR*
CAMP1183	*hydD*
CAMP1184	*hydC*
CAMP1185	*hydB*
CAMP1188	*amiA*
CAMP1189	Cj1270c
CAMP1190	*tyrS*
CAMP1191	*spoT*
CAMP1192	*rpoZ*
CAMP1193	*pyrH*
CAMP1194	Cj1275c
CAMP1199	Cj1280c
CAMP1203	Cj1285c
CAMP1205	Cj1287c
CAMP1206	*gltX2*
CAMP1209	*accB*
CAMP1210	*dcd*
CAMP1260	Cj1344c
CAMP1262	Cj1346c
CAMP1275	*ppk*
CAMP1278	*ruvB*
CAMP1282	*glmS*
CAMP1284	Cj1368
CAMP1285	Cj1369
CAMP1286	Cj1370
CAMP1290	Cj1374c
CAMP1291	Cj1375
CAMP1293	Cj1377c
CAMP1294	*selA*
CAMP1298	*fldA*
CAMP1300	Cj1384c

CAMP1312	*tpiA*
CAMP1313	*pgk*
CAMP1314	*gapA*
CAMP1315	*nadD*
CAMP1318	Cj1407c
CAMP1319	*fliL*
CAMP1320	*acpS*
CAMP1354	*kpsF*
CAMP1361	*dut*
CAMP1364	Cj1454c
CAMP1365	*prfB*
CAMP1367	*truD*
CAMP1368	*thiL*
CAMP1371	Cj1461
CAMP1372	*flgI*
CAMP1373	*flgJ*
CAMP1377	Cj1467
CAMP1378	Cj1468
CAMP1385	Cj1477c
CAMP1386	*cadF*
CAMP1394	Cj1486c
CAMP1395	*ccoP*
CAMP1396	*ccoQ*
CAMP1397	*ccoO*
CAMP1398	*ccoN*
CAMP1402	*carA*
CAMP1403	Cj1495c
CAMP1406	*purA*
CAMP1411	*selD*
CAMP1416	*fdhC*
CAMP1419	Cj1513c
CAMP1421	Cj1515c
CAMP1423	*moaD*
CAMP1424	*moaE*
CAMP1425	*moeA2*
CAMP1429	*purM*
CAMP1434	Cj1534c
CAMP1435	*pgi*
CAMP1436	*galU*
CAMP1474	*nuoB*
CAMP1486	*infA*
CAMP1491	*rpoA*
CAMP1500	*hisI*
CAMP1501	*dapD*
CAMP1502	*mrp*
CAMP1503	*ispDF*
CAMP1506	*pgpA*
CAMP1508	*prfA*
CAMP1529	Cj1633
CAMP1530	*aroC*
CAMP1531	*rnc*
CAMP1532	*rnhA*
CAMP1534	*dnaG*
CAMP1535	Cj1639
CAMP1537	*murE*
CAMP1538	Cj1642
CAMP1540	*ispA*
CAMP1541	*tkt*
CAMP1547	*map*
CAMP1554	p19
CAMP1556	Cj1661
CAMP1558	Cj1663
CAMP1567	*eno*
CAMP1568	*recA*
CAMP1569	Cj1674
CAMP1570	*fliQ*
CAMP1571	*murB*
CAMP1577	Cj1684c
CAMP1578	*bioB*
CAMP1579	*topA*
CAMP1581	*secY*
CAMP1602	Cj1709c
CAMP1603	Cj1710c
CAMP1604	*ksgA*
CAMP1605	Cj1712
CAMP1606	Cj1713
CAMP1617	Cj1724c
CAMP1623	*ruvC*
CAMP1803	*rpsA*
CAMP1804	*rpsB*
CAMP1805	*rpsC*
CAMP1806	*rpsD*
CAMP1807	*rpsE*
CAMP1808	*rpsF*
CAMP1809	*rpsG*
CAMP1810	*rpsH*
CAMP1811	*rpsI*
CAMP1812	*rpsJ*
CAMP1813	*rpsK*
CAMP1814	*rpsL*

CAMP1815	*rpsM*
CAMP1816	*rpsN*
CAMP1817	*rpsO*
CAMP1818	*rpsP*
CAMP1819	*rpsQ*
CAMP1820	*rpsR*
CAMP1821	*rpsS*
CAMP1822	*rpsT*
CAMP1823	*rpsU*
CAMP1824	*rplA*
CAMP1825	*rplB*
CAMP1826	*rplC*
CAMP1827	*rplD*
CAMP1828	*rplE*
CAMP1829	*rplF*
CAMP1830	*rplL*
CAMP1831	*rplI*
CAMP1832	*rplJ*
CAMP1833	*rplK*
CAMP1834	*rplM*
CAMP1835	*rplN*
CAMP1836	*rplO*
CAMP1837	*rplP*
CAMP1838	*rplQ*
CAMP1839	*rplR*
CAMP1840	*rplS*
CAMP1841	*rplT*
CAMP1842	*rplU*
CAMP1843	*rplV*
CAMP1844	*rplW*
CAMP1845	*rplX*
CAMP1846	*rpmA*
CAMP1847	*rpmB*
CAMP1848	*rpmC*
CAMP1849	*rpmE*
CAMP1850	*rpmF*
CAMP1851	*rpmG*
CAMP1852	*rpmH*
CAMP1853	*rpmI*

Appendix E

Table E1 Details of *Campylobacter* isolates recovered at slaughter from two free-range broiler flocks from South Africa

Flock	Isolate	Carcass number	Sample type[a]	Isolation protocol	Duplicate isolates	Preliminary species	PubMLST id	WGS assembly statistics		
								Size (Mb)	Contigs (*n*)	*N50* (bp)
A	UWC001	1	M	ISO	-	*C. lari*	Not viable			
	UWC002	2	M	ISO	Set 1	*C. lari*	28261	1.67	51	142524
	UWC003	3	M	ISO	Set 2	*C. lari*	28262	1.67	48	142524
	UWC004	4	M	ISO	Set 3	*C. coli*	Not viable			
	UWC005	5	M	ISO	Set 4	*C. lari*	28263	1.67	47	142520
	UWC006	6	M	ISO	-	*C. lari*	Not viable			
	UWC007	7	M	ISO	-	*C. coli*	28264	1.67	56	124707
	UWC008	8	M	ISO	-	*C. coli*	Not viable			
	UWC010	3	NS	ISO	Set 2	*C. lari*	28266	1.67	49	142524
	UWC011	4	NS	ISO	Set 3	*C. lari*	28267	1.67	48	142524
	UWC012	5	NS	ISO	Set 4	*C. lari*	28268	1.66	36	212408
	UWC016	9	M	CTP	-	*C. coli*	28272	1.67	46	142520
	UWC025	2	NS	CTP	Set 1	*C. coli*	28278	1.67	52	142524
	UWC026	5	NS	CTP	Set 4	*C. coli*	28279	1.67	49	142524
B	UWC009	37	M	ISO	Set 9	*C. coli*	28265	1.67	45	142520
	UWC013	26	NS	ISO	-	*C. coli*	28269	1.67	42	142524
	UWC014	28	NS	ISO	-	*C. coli*	28270	1.67	50	142524
	UWC015	32	NS	ISO	Set 5	*C. coli*	28271	1.67	54	142524
	UWC017	33	M	CTP	Set 6	*C. coli*	Not viable			
	UWC018	34	M	CTP	Set 7	*C. coli*	28273	1.67	53	124699
	UWC019	35	M	CTP	Set 8	*C. coli*	28274	1.67	48	142524
	UWC020	36	M	CTP	-	*C. coli*	28275	1.67	53	142524
	UWC021	37	M	CTP	Set 9	*C. coli*	28276	1.67	44	147059
	UWC022	38	M	CTP	-	*C. coli*	Not viable			
	UWC023	39	M	CTP	-	*C. coli*	28277	1.67	47	142516
	UWC024	40	M	CTP	-	*C. coli*	Not viable			
	UWC028	22	NS	CTP	-	*C. coli*	Not viable			
	UWC029	23	NS	CTP	-	*C. coli*	28281	1.67	47	142516
	UWC030	27	NS	CTP	-	*C. coli*	28282	1.67	49	142524
	UWC031	29	NS	CTP	-	*C. coli*	28283	1.67	56	142528
	UWC032	30	NS	CTP	-	*C. coli*	28284	1.67	39	126610
	UWC033	32	NS	CTP	Set 5	*C. coli*	Not viable			
	UWC034	33	NS	CTP	Set 6	*C. coli*	28285	1.67	48	142524
	UWC035	34	NS	CTP	Set 7	*C. coli*	28286	1.67	51	147059
	UWC036	35	NS	CTP	Set 8	*C. coli*	28287	1.67	48	142524
	UWC037	37	NS	CTP	Set 9	*C. coli*	28288	1.67	47	142524
	UWC038	25	NS	CTP	-	*C. coli*	28289	1.67	44	142516

WGS, whole-genome sequencing; M, meat; NS, neck skin; ISO, ISO 10272-1:2006 protocol; CTP, Cape Town protocol.

[a] Highlighting indicates sets of isolates obtained from the same broiler carcass.

Appendix F

Table F1 Distance matrix generated by GENOME COMPARATOR for wgMLST comparison carried out to resolve ST-1587 isolates from South Africa and Denmark

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	15-537360 (R)	-																													
2	UWC002 (A)	1153	-																												
3	UWC003 (A)	1155	10	-																											
4	UWC005 (A)	1156	13	8	-																										
5	UWC007 (A)	1147	9	3	8	-																									
6	UWC009 (B)	1156	12	13	10	12	-																								
7	UWC010 (A)	1154	5	9	12	10	8	-																							
8	UWC011 (A)	1153	9	9	10	9	11	13	-																						
9	UWC012 (A)	1155	11	11	9	8	14	13	9	-																					
10	UWC013 (B)	1155	10	13	17	13	14	10	14	17	-																				
11	UWC014 (B)	1154	10	13	14	11	12	8	12	11	13	-																			
12	UWC015 (B)	1152	9	11	13	9	10	8	11	13	8	12	-																		
13	UWC016 (A)	1153	9	12	13	12	18	12	10	13	18	14	16	-																	
14	UWC018 (B)	1150	10	7	8	9	10	10	5	9	15	11	12	9	-																
15	UWC019 (B)	1155	11	10	13	11	15	11	8	15	13	14	11	10	7	-															
16	UWC020 (B)	1152	7	10	12	10	13	11	5	12	12	10	9	11	6	8	-														
17	UWC021 (B)	1153	9	6	9	5	11	9	6	10	9	10	5	12	9	8	6	-													
18	UWC023 (B)	1153	9	11	12	10	12	14	3	10	15	12	12	11	5	9	5	8	-												
19	UWC025 (A)	1152	7	13	14	10	15	13	9	14	12	16	9	15	12	12	9	8	11	-											
20	UWC026 (A)	1153	6	10	13	11	13	9	10	15	13	15	10	10	9	8	12	10	12	10	-										
21	UWC029 (B)	1153	8	11	11	8	12	7	9	9	9	7	7	11	10	9	6	5	9	12	12	-									
22	UWC030 (B)	1152	6	7	6	7	9	7	8	7	10	6	9	9	9	10	7	4	9	11	10	4	-								
23	UWC031 (B)	1153	7	10	11	9	12	5	12	11	7	11	4	12	11	9	10	6	13	8	7	5	7	-							
24	UWC032 (B)	1154	4	9	12	8	9	7	7	13	10	12	6	11	8	9	7	5	9	5	5	10	7	7	-						
25	UWC034 (B)	1155	13	13	17	12	13	11	14	18	11	16	7	20	14	14	11	8	15	12	14	11	12	9	9	-					
26	UWC035 (B)	1152	8	8	10	7	13	13	4	10	12	12	9	10	6	6	5	5	6	7	9	8	8	10	6	12	-				
27	UWC036 (B)	1153	8	5	8	4	10	8	7	9	10	9	6	11	8	9	7	1	9	9	9	6	3	7	4	9	6	-			
28	UWC037 (B)	1152	13	8	10	6	11	11	10	11	13	12	9	13	9	10	10	6	11	14	12	9	8	10	9	12	8	5	-		
29	UWC038 (B)	1156	6	11	14	10	9	8	8	14	12	13	8	13	9	11	9	7	10	7	7	13	9	9	3	11	8	6	11	-	
30	UCT020 (C)	1153	12	8	10	5	13	11	8	10	11	11	6	14	9	9	6	5	7	11	12	8	7	8	10	11	5	6	7	13	-
31	tico2087 (D)	1169	591	595	596	590	594	592	593	594	597	595	594	597	590	596	592	595	592	592	594	594	592	593	594	596	594	594	595	596	596

(R), reference genome; (A/B), flock A/B; (C), human disease isolate from Cape Town; (D), human disease isolate from Denmark. Highlighting indicates isolates obtained from the same carcass.

Appendix G

Table G1 Representatives from the Oxfordshire sentinel surveillance collection of *Campylobacter* included in genomic comparisons of UK and South African human disease isolates

Clonal complex	PubMLST id	Isolate	ENA accession number	ST[a]	rST[b]
ST-21	22252	OXC7068	ERR221349	19	376
ST-21	21249	OXC7168	ERR193216	19	459
ST-21	16152	OXC6355	ERR083971	19	466
ST-21	12883	OXC6548	ERR108312	19	472
ST-21	16191	OXC6394	ERR084010	19	473
ST-21	25378	OXC7960	ERR386242	19	484
ST-21	22295	OXC7253	ERR221392	19	512
ST-21	16100	OXC6303	ERR083919	19	513
ST-21	16167	OXC6370	ERR083986	19	514
ST-21	16330	OXC6538	ERR084149	19	515
ST-21	21598	OXC6879	ERR356081	21	441
ST-21	18246	OXC6689	ERR128774	21	443
ST-21	18259	OXC6702	ERR128787	21	444
ST-21	21234	OXC7153	ERR193201	21	446
ST-21	16067	OXC6270	ERR083887	21	464
ST-21	16114	OXC6317	ERR083933	21	465
ST-21	21591	OXC6872	ERR356074	21	494
ST-21	18288	OXC6731	ERR128816	21	496
ST-21	21228	OXC7147	ERR193195	21	497
ST-21	16386	OXC6637	ERR108399	21	500
ST-21	16049	OXC6251	ERR083868	21	502
ST-21	12897	OXC6562	ERR108326	21	503
ST-21	22368	OXC7330	ERR221465	21	504
ST-21	16054	OXC6257	ERR083874	21	519
ST-21	22347	OXC7306	ERR221444	21	520
ST-21	16300	OXC6508	ERR084119	21	529
ST-21	12898	OXC6563	ERR108327	21	535
ST-21	16353	OXC6602	ERR108366	21	538
ST-21	23891	OXC7351	ERR278236	21	3497
ST-21	24034	OXC7508	ERR278411	21	3575
ST-21	25374	OXC7954	ERR386238	21	4090
ST-21	24960	OXC7718	ERR343045	21	4541
ST-21	18224	OXC6667	ERR128752	44	486
ST-21	16313	OXC6521	ERR084132	47	510
ST-21	24568	OXC7639	ERR330545	50	350
ST-21	18221	OXC6664	ERR128749	50	448

Clonal complex	PubMLST id	Isolate	ENA accession number	ST[a]	rST[b]
ST-21	22288	OXC7246	ERR221385	50	449
ST-21	16123	OXC6326	ERR083942	50	470
ST-21	24037	OXC7512	ERR278414	50	474
ST-21	12906	OXC6571	ERR108335	50	475
ST-21	16281	OXC6489	ERR084100	50	479
ST-21	16285	OXC6493	ERR084104	50	481
ST-21	16074	OXC6277	ERR083894	50	482
ST-21	12900	OXC6565	ERR108329	50	516
ST-21	16253	OXC6461	ERR084072	50	517
ST-21	21165	OXC6994	ERR189862	50	518
ST-21	16363	OXC6613	ERR108376	50	524
ST-21	24898	OXC7641	ERR343003	50	4520
ST-21	22720	OXC7230	ERR222743	53	452
ST-21	12893	OXC6558	ERR108322	53	460
ST-21	16164	OXC6367	ERR083983	53	461
ST-21	24130	OXC7526	ERR278509	53	3521
ST-21	23908	OXC7370	ERR278253	53	3557
ST-21	18389	OXC6833	ERR137093	104	112
ST-21	16297	OXC6505	ERR084116	141	333
ST-21	18248	OXC6691	ERR128776	190	439
ST-21	22353	OXC7312	ERR221450	190	462
ST-21	16137	OXC6340	ERR083956	262	540
ST-21	22234	OXC7049	ERR221331	266	488
ST-21	21576	OXC6857	ERR356059	520	377
ST-21	21229	OXC7148	ERR193196	822	468
ST-21	16098	OXC6301	ERR083917	861	447
ST-21	18380	OXC6823	ERR137084	883	454
ST-21	22119	OXC6924	ERR356125	883	456
ST-21	16271	OXC6479	ERR084090	883	489
ST-21	24638	OXC7841	ERR330623	883	3504
ST-21	24030	OXC7504	ERR278407	982	3533
ST-21	22262	OXC7078	ERR221359	1949	445
ST-21	16097	OXC6300	ERR083916	2135	451
ST-21	18308	OXC6751	ERR137012	2135	487
ST-21	25404	OXC7997	ERR386268	2782	4544
ST-21	16306	OXC6514	ERR084125	3102	478
ST-21	23964	OXC7431	ERR278341	3102	3554
ST-21	24486	OXC7547	ERR330459	3574	4011
ST-21	12887	OXC6552	ERR108316	3769	477
ST-21	18355	OXC6798	ERR137059	4526	455
ST-21	16224	OXC6429	ERR084043	5018	525
ST-21	18284	OXC6727	ERR128812	5811	374
ST-21	21158	OXC6985	ERR189826	6135	467

Clonal complex	PubMLST id	Isolate	ENA accession number	ST[a]	rST[b]
ST-21	21186	OXC7104	ERR193153	6137	39
ST-21	25023	OXC7892	ERR348952	6891	4542
ST-22	18201	OXC6644	ERR128729	22	136
ST-22	16061	OXC6264	ERR083881	22	137
ST-22	18312	OXC6755	ERR137016	22	138
ST-22	16209	OXC6414	ERR084028	22	320
ST-22	16327	OXC6535	ERR084146	22	335
ST-22	18265	OXC6708	ERR128793	22	346
ST-22	21239	OXC7158	ERR193206	22	347
ST-22	22124	OXC6929	ERR356130	22	379
ST-22	22233	OXC7048	ERR221330	22	383
ST-22	16375	OXC6626	ERR108388	22	3480
ST-22	25500	OXC8031	ERR386364	22	4553
ST-22	25116	OXC7921	ERR349003	660	4550
ST-22	16201	OXC6406	ERR084020	1947	314
ST-45	22275	OXC7093	ERR221372	11	66
ST-45	21184	OXC7102	ERR193151	25	417
ST-45	16111	OXC6314	ERR083930	45	67
ST-45	16214	OXC6419	ERR084033	45	241
ST-45	16136	OXC6339	ERR083955	45	287
ST-45	16056	OXC6259	ERR083876	45	288
ST-45	16090	OXC6293	ERR083909	45	289
ST-45	21185	OXC7103	ERR193152	45	392
ST-45	22316	OXC7275	ERR221413	45	400
ST-45	22670	OXC7179	ERR222693	45	401
ST-45	22729	OXC7239	ERR222752	45	406
ST-45	22245	OXC7061	ERR221342	45	411
ST-45	22308	OXC7267	ERR221405	45	412
ST-45	16162	OXC6365	ERR083981	45	413
ST-45	16232	OXC6437	ERR084051	45	414
ST-45	22677	OXC7186	ERR222700	45	415
ST-45	24591	OXC7788	ERR330627	45	421
ST-45	16328	OXC6536	ERR084147	45	422
ST-45	16118	OXC6321	ERR083937	45	423
ST-45	16345	OXC6594	ERR108358	45	424
ST-45	18354	OXC6797	ERR137058	45	425
ST-45	24980	OXC7745	ERR343023	45	3507
ST-45	24611	OXC7813	ERR330597	45	3522
ST-45	25606	OXC8072	ERR386470	45	3973
ST-45	16148	OXC6351	ERR083967	45	3983
ST-45	25387	OXC7972	ERR386251	45	4060
ST-45	25491	OXC8007	ERR386355	45	4589
ST-45	25515	OXC8065	ERR386379	45	4591

Clonal complex	PubMLST id	Isolate	ENA accession number	ST[a]	rST[b]
ST-45	22369	OXC7331	ERR221466	137	418
ST-45	22661	OXC7170	ERR222684	137	427
ST-45	22322	OXC7281	ERR221419	137	428
ST-45	22251	OXC7067	ERR221348	137	429
ST-45	16343	OXC6592	ERR108356	137	432
ST-45	16241	OXC6448	ERR084060	137	4048
ST-45	25014	OXC7881	ERR348880	137	4588
ST-45	25508	OXC8046	ERR386372	230	4019
ST-45	16155	OXC6358	ERR083974	233	420
ST-45	23884	OXC7344	ERR278229	334	3494
ST-45	21195	OXC7113	ERR193162	538	431
ST-45	22218	OXC7032	ERR221315	583	419
ST-45	18376	OXC6819	ERR137080	583	426
ST-45	16091	OXC6294	ERR083910	583	433
ST-45	24518	OXC7584	ERR330525	583	4585
ST-45	16307	OXC6515	ERR084126	845	334
ST-45	21112	OXC6938	ERR189780	2109	438
ST-45	23909	OXC7371	ERR278254	2109	3491
ST-45	23915	OXC7377	ERR278260	2109	3515
ST-45	22213	OXC7026	ERR221310	2219	68
ST-45	21248	OXC7167	ERR193215	6886	416
ST-52	16165	OXC6368	ERR083984	52	116
ST-52	16235	OXC6441	ERR084054	52	118
ST-52	18294	OXC6737	ERR128822	52	120
ST-52	22229	OXC7044	ERR221326	52	357
ST-52	24132	OXC7528	ERR278511	52	3513
ST-52	24978	OXC7743	ERR343009	52	4621
ST-52	25421	OXC8081	ERR386285	52	4624
ST-52	25588	OXC8023	ERR386452	52	4628
ST-52	16227	OXC6432	ERR084046	614	264
ST-52	18226	OXC6669	ERR128754	775	117
ST-52	18325	OXC6768	ERR137029	2275	359
ST-206	22694	OXC7203	ERR222717	46	105
ST-206	21124	OXC6950	ERR189792	122	69
ST-206	16317	OXC6525	ERR084136	122	71
ST-206	22333	OXC7292	ERR221430	122	72
ST-206	16194	OXC6397	ERR084013	122	310
ST-206	22673	OXC7182	ERR222696	122	397
ST-206	23890	OXC7350	ERR278235	122	3582
ST-206	25570	OXC7950	ERR386434	122	4095
ST-206	25048	OXC7918	ERR348875	122	4448
ST-206	25427	OXC8094	ERR386291	122	4457
ST-206	16112	OXC6315	ERR083931	206	81

Clonal complex	PubMLST id	Isolate	ENA accession number	ST[a]	rST[b]
ST-206	25030	OXC7899	ERR348913	206	83
ST-206	21121	OXC6947	ERR189789	206	87
ST-206	16314	OXC6522	ERR084133	206	104
ST-206	25538	OXC7730	ERR386402	221	85
ST-206	16372	OXC6623	ERR108385	222	86
ST-206	16184	OXC6387	ERR084003	227	80
ST-206	24549	OXC7620	ERR330488	227	3503
ST-206	23972	OXC7441	ERR278349	227	3590
ST-206	16324	OXC6532	ERR084143	273	82
ST-206	12888	OXC6553	ERR108317	273	84
ST-206	16050	OXC6252	ERR083869	572	74
ST-206	16225	OXC6430	ERR084044	572	75
ST-206	16179	OXC6382	ERR083998	572	76
ST-206	21242	OXC7161	ERR193209	572	77
ST-206	16217	OXC6422	ERR084036	572	78
ST-206	18350	OXC6793	ERR137054	572	79
ST-206	18276	OXC6719	ERR128804	572	110
ST-206	24513	OXC7576	ERR330485	572	3496
ST-206	24612	OXC7814	ERR330605	572	3594
ST-206	25523	OXC8090	ERR386387	572	4470
ST-206	25492	OXC8010	ERR386356	2149	3976
ST-206	18233	OXC6676	ERR128761	5809	91
ST-257	16048	OXC6250	ERR083867	257	179
ST-257	16202	OXC6407	ERR084021	257	184
ST-257	16052	OXC6254	ERR083871	257	186
ST-257	16159	OXC6362	ERR083978	257	187
ST-257	16196	OXC6401	ERR084015	257	188
ST-257	21202	OXC7120	ERR193169	257	189
ST-257	21188	OXC7106	ERR193155	257	190
ST-257	12920	OXC6583	ERR108347	257	191
ST-257	22349	OXC7308	ERR221446	257	193
ST-257	18307	OXC6750	ERR137011	824	126
ST-257	24019	OXC7491	ERR278396	824	127
ST-257	24476	OXC7533	ERR330474	824	3540
ST-257	18304	OXC6747	ERR137008	990	181
ST-257	16318	OXC6526	ERR084137	990	182
ST-257	12919	OXC6582	ERR108346	2030	183
ST-353	22108	OXC6913	ERR356114	5	266
ST-353	22713	OXC7222	ERR222736	5	267
ST-353	12922	OXC6585	ERR108349	5	270
ST-353	24548	OXC7619	ERR330480	5	271
ST-353	12901	OXC6566	ERR108330	5	272
ST-353	16233	OXC6438	ERR084052	5	273

Clonal complex	PubMLST id	Isolate	ENA accession number	ST[a]	rST[b]
ST-353	21142	OXC6968	ERR189810	5	274
ST-353	22232	OXC7047	ERR221329	5	382
ST-353	24520	OXC7586	ERR330541	5	3534
ST-353	23978	OXC7448	ERR278355	5	3603
ST-353	16248	OXC6456	ERR084067	353	31
ST-353	16361	OXC6611	ERR108374	353	32
ST-353	16286	OXC6494	ERR084105	353	201
ST-353	24559	OXC7630	ERR330473	353	3495
ST-353	16357	OXC6607	ERR108370	353	4132
ST-353	24964	OXC7722	ERR343077	353	4477
ST-353	18371	OXC6814	ERR137075	356	291
ST-353	16116	OXC6319	ERR083935	356	292
ST-353	22219	OXC7033	ERR221316	356	458
ST-353	16115	OXC6318	ERR083934	356	490
ST-353	24903	OXC7646	ERR343040	356	4556
ST-353	23918	OXC7380	ERR278263	400	29
ST-353	22717	OXC7226	ERR222740	400	30
ST-353	12923	OXC6586	ERR108350	400	33
ST-353	24529	OXC7597	ERR330518	581	3523
ST-353	16371	OXC6622	ERR108384	595	265
ST-353	22312	OXC7271	ERR221409	2036	399
ST-353	24917	OXC7662	ERR343057	3696	4563
ST-353	16247	OXC6455	ERR084066	5730	47
ST-353	12921	OXC6584	ERR108348	5758	283
ST-353	18333	OXC6776	ERR137037	5815	387
ST-353	18363	OXC6806	ERR137067	5816	28
ST-353	23923	OXC7385	ERR278268	6701	3518
ST-353	24916	OXC7661	ERR343049	6888	4557
ST-403	18240	OXC6683	ERR128768	270	218
ST-403	22307	OXC7266	ERR221404	270	219
ST-403	24561	OXC7632	ERR330489	270	3591
ST-403	25486	OXC7999	ERR386350	270	4569
ST-403	18297	OXC6740	ERR137001	403	216
ST-403	16276	OXC6484	ERR084095	403	332
ST-403	24487	OXC7548	ERR330467	403	3558
ST-403	22102	OXC6907	ERR356108	932	378
ST-403	16146	OXC6349	ERR083965	933	215
ST-403	22079	OXC6883	ERR356085	985	217
ST-403	21588	OXC6869	ERR356071	5246	331
ST-443	21125	OXC6951	ERR189793	51	220
ST-443	16113	OXC6316	ERR083932	51	223
ST-443	18283	OXC6726	ERR128811	51	226
ST-443	16071	OXC6274	ERR083891	51	227

Clonal complex	PubMLST id	Isolate	ENA accession number	ST[a]	rST[b]
ST-443	22292	OXC7250	ERR221389	51	229
ST-443	18206	OXC6649	ERR128734	51	230
ST-443	16369	OXC6620	ERR108382	51	341
ST-443	16139	OXC6342	ERR083958	51	3476
ST-443	24123	OXC7518	ERR278502	51	3548
ST-443	22376	OXC7339	ERR221473	443	221
ST-443	22692	OXC7201	ERR222715	1703	228
ST-443	16273	OXC6481	ERR084092	2361	222
ST-443	21137	OXC6963	ERR189805	5794	224
ST-828	22239	OXC7054	ERR221336	825	2
ST-828	18367	OXC6810	ERR137071	825	5
ST-828	16094	OXC6297	ERR083913	825	8
ST-828	16219	OXC6424	ERR084038	825	12
ST-828	21160	OXC6987	ERR189828	825	14
ST-828	16195	OXC6400	ERR084014	825	311
ST-828	24581	OXC7773	ERR330642	825	3506
ST-828	23983	OXC7453	ERR278360	825	3562
ST-828	24629	OXC7832	ERR330646	825	3597
ST-828	24023	OXC7495	ERR278400	825	3606
ST-828	24966	OXC7725	ERR343090	825	3919
ST-828	22115	OXC6920	ERR356121	827	158
ST-828	22690	OXC7199	ERR222713	827	159
ST-828	16105	OXC6308	ERR083924	827	161
ST-828	18318	OXC6761	ERR137022	827	162
ST-828	22360	OXC7321	ERR221457	827	163
ST-828	18342	OXC6785	ERR137046	827	165
ST-828	22096	OXC6901	ERR356102	827	168
ST-828	18242	OXC6685	ERR128770	827	3482
ST-828	23932	OXC7394	ERR278277	827	3490
ST-828	24604	OXC7806	ERR330636	827	3524
ST-828	21167	OXC6996	ERR189864	827	3960
ST-828	25067	OXC7944	ERR348932	827	4673
ST-828	25507	OXC8042	ERR386371	827	4675
ST-828	21212	OXC7131	ERR193179	828	6
ST-828	16305	OXC6513	ERR084124	828	7
ST-828	21235	OXC7154	ERR193202	828	153
ST-828	24509	OXC7572	ERR330548	828	3520
ST-828	24565	OXC7636	ERR330521	828	3547
ST-828	24637	OXC7840	ERR330615	828	3605
ST-828	25021	OXC7890	ERR348936	828	4664
ST-828	21566	OXC6847	ERR356049	829	13
ST-828	16183	OXC6386	ERR084002	829	151
ST-828	16051	OXC6253	ERR083870	829	154

Clonal complex	PubMLST id	Isolate	ENA accession number	ST[a]	rST[b]
ST-828	22691	OXC7200	ERR222714	829	394
ST-828	24551	OXC7622	ERR330504	829	3564
ST-828	25064	OXC7939	ERR348908	829	4667
ST-828	27865	OXC7713	ERR351536	829	7039
ST-828	16060	OXC6263	ERR083880	830	9
ST-828	22719	OXC7228	ERR222742	830	10
ST-828	18322	OXC6765	ERR137026	832	173
ST-828	24541	OXC7612	ERR330519	854	3570
ST-828	12911	OXC6576	ERR108340	855	279
ST-828	22267	OXC7083	ERR221364	855	280
ST-828	24512	OXC7575	ERR330477	855	3596
ST-828	16268	OXC6476	ERR084087	860	17
ST-828	16055	OXC6258	ERR083875	860	285
ST-828	24500	OXC7563	ERR330476	860	3574
ST-828	25061	OXC7933	ERR348884	860	4665
ST-828	12903	OXC6568	ERR108332	872	157
ST-828	24525	OXC7593	ERR330486	872	3498
ST-828	24530	OXC7598	ERR330526	872	3517
ST-828	18241	OXC6684	ERR128769	962	175
ST-828	16140	OXC6343	ERR083959	962	176
ST-828	24639	OXC7842	ERR330631	962	3502
ST-828	24994	OXC7859	ERR348910	962	4660
ST-828	22214	OXC7027	ERR221311	1016	16
ST-828	21157	OXC6984	ERR189825	1055	19
ST-828	23887	OXC7347	ERR278232	1055	3532
ST-828	18282	OXC6725	ERR128810	1096	15
ST-828	16211	OXC6416	ERR084030	1145	11
ST-828	18292	OXC6735	ERR128820	1145	171
ST-828	18208	OXC6651	ERR128736	1181	364
ST-828	21192	OXC7110	ERR193159	1191	21
ST-828	16252	OXC6460	ERR084071	1541	328
ST-828	24621	OXC7824	ERR330582	1585	3607
ST-828	16134	OXC6337	ERR083953	1600	297
ST-828	22730	OXC7240	ERR222753	1614	337
ST-828	16352	OXC6601	ERR108365	1614	338
ST-828	24032	OXC7506	ERR278409	1614	3602
ST-828	16135	OXC6338	ERR083954	1628	298
ST-828	25400	OXC7988	ERR386264	1774	3885
ST-828	18262	OXC6705	ERR128790	1957	372
ST-828	24905	OXC7648	ERR343056	2178	4654
ST-828	21211	OXC7130	ERR193178	2183	389
ST-828	18382	OXC6825	ERR137086	2183	390
ST-828	18295	OXC6738	ERR136999	2273	373

Clonal complex	PubMLST id	Isolate	ENA accession number	ST[a]	rST[b]
ST-828	16221	OXC6426	ERR084040	2464	196
ST-828	24942	OXC7694	ERR343067	2689	4657
ST-828	24483	OXC7542	ERR330530	3016	3519
ST-828	21111	OXC6937	ERR189779	3753	18
ST-828	16175	OXC6378	ERR083994	4425	22
ST-828	22668	OXC7177	ERR222691	4425	23
ST-828	16169	OXC6372	ERR083988	4425	24
ST-828	24906	OXC7649	ERR343064	4956	4656
ST-828	16064	OXC6267	ERR083884	5642	167
ST-828	16315	OXC6523	ERR084134	5659	300
ST-828	16073	OXC6276	ERR083893	5733	3
ST-828	16106	OXC6309	ERR083925	5734	290
ST-828	16177	OXC6380	ERR083996	5735	304
ST-828	16093	OXC6296	ERR083912	5737	170
ST-828	12894	OXC6559	ERR108323	5757	262
ST-828	18267	OXC6710	ERR128795	5810	312
ST-828	18301	OXC6744	ERR137005	5813	146
ST-828	23946	OXC7411	ERR278323	6267	3578
ST-828	22709	OXC7218	ERR222732	6543	177
ST-828	24539	OXC7609	ERR330503	6705	3598
ST-828	24558	OXC7629	ERR330465	6706	3525
ST-828	24003	OXC7474	ERR278380	6755	3511

[a] ST, sequence type.
[b] rST, ribosomal sequence type.

Appendix H

Table H1 Details of isolates included in the pilot study on the recovery of whole-genome sequencing data from freeze-dried *Campylobacter* cultures

PubMLST id	Isolate	Year of isolation	Species	Assembly statistics				Proportion of complete typing loci (%)	
				Size (Mb)	Contigs (n)	N_{50} (bp)	GC content (%)	MLST	rMLST
25262	UCTARC001	1982	Putative novel *C. jejuni* subsp.	1.84	86	96266	30.5	100	100
33075	UCTARC208	1982	*C. jejuni* subsp. *jejuni*	1.68	285	10985	30.4	100	100
33076	UCTARC209	1982	*C. jejuni* subsp. *jejuni*	1.70	27	216425	30.4	100	100
25263	UCTARC003	1983	*C. jejuni* subsp. *jejuni*	1.67	56	106410	30.5	100	100
25264	UCTARC004	1983	*C. coli*	1.69	44	148120	31.3	100	100
25265	UCTARC005	1983	*C. jejuni* subsp. *jejuni*	1.62	46	105399	30.4	100	100
25238	UCTARC011	1983	*C. jejuni* subsp. *jejuni*	1.60	35	183955	30.5	100	100
25239	UCTARC012	1983	*C. jejuni* subsp. *jejuni*	1.62	44	119703	30.4	100	100
33077	UCTARC210	1983	*C. jejuni* subsp. *jejuni*	1.76	47	104867	30.3	100	100
25266	UCTARC006	1989	C. upsaliensis	1.74	124	43105	34.1	100	100
25267	UCTARC007	1989	C. *jejuni* subsp. *doyle*i	1.69	118	66603	30.5	100	100
25268	UCTARC008	1989	C. *jejuni* subsp. *doylei*	1.73	113	76877	30.5	100	100
25269	UCTARC009	1989	C. *jejuni* subsp. *doylei*	1.70	90	84667	30.5	100	100
25240	UCTARC013	1989	*C. upsaliensis*	1.73	150	38356	34.3	100	100

Appendix I

Table I1 Details of freeze-dried *Campylobacter* isolates from the 1991 archive collection for which WGS data were obtained

Isolate id	Isolate	Alias	Number of vials	Species	Size (Mb)	Contigs (n)	N50 (bp)	GC content (%)	MLST	rMLST
32967	UCTARC022	15.91	2	*C. upsaliensis*	1.76	173	35584	34.3	100	100
32968	UCTARC024	19.91B	5	*C. concisus*	1.81	252	15034	37.7	100	100
32969	UCTARC029	26.91	2	*C. jejuni* subsp. *doylei*	1.68	149	34021	30.5	100	100
32970	UCTARC031	30.91	2	*C. jejuni* subsp. *jejuni*	1.65	39	153587	30.5	100	98.1 (1)
34779	UCTARC032	16.91B	3	*C. jejuni* subsp. *doylei*	1.80	425	8455	30.5	100	100
32971	UCTARC033	32.91	3	*C. jejuni* subsp. *doylei*	1.69	123	76928	30.5	100	100
32972	UCTARC036	44.91	3	Putative novel *C. jejuni* subspecies	1.96	112	52335	30.4	100	100
32973	UCTARC039	48.91	5	*C. jejuni* subsp. *jejuni*	1.78	84	90762	30.1	100	100
32974	UCTARC041	52.91	2	*C. jejuni* subsp. *doylei*	1.68	116	60451	30.5	100	100
32975	UCTARC044	61.91	1	*C. hyointestinalis*	1.80	309	12556	34.1	100	100
32976	UCTARC052	84.91sm	3	*C. jejuni* subsp. *doylei*	1.73	111	80532	30.5	100	100
32977	UCTARC053	88.91	1	*C. jejuni* subsp. *doylei*	1.78	167	26642	30.4	100	98.1 (1)
32996	UCTARC054	89.91	2	*C. jejuni* subsp. *jejuni*	1.64	65	118134	30.4	100	100
32978	UCTARC055	91.91	2	*C. coli*	1.78	260	12972	31.3	100	100
32998	UCTARC061	84.91lg	2	*C. jejuni* subsp. *jejuni*	1.69	95	109603	30.3	100	100
32999	UCTARC063	103.91	3	*C. jejuni* subsp. *doylei*	1.69	115	78290	30.5	100	100
33000	UCTARC065	105.91	2	*C. upsaliensis*	1.67	148	33865	34.5	100	100
33001	UCTARC066	105.91lg	3	*C. jejuni* subsp. *jejuni*	1.73	61	148418	30.3	100	100
33002	UCTARC069	111.91	2	*C. jejuni* subsp. *doylei*	1.76	105	103653	30.5	100	100

33003	UCTARC071	113.91	5	Putative novel *C. jejuni* subspecies	1.90	92	65333	30.4	100	100
33004	UCTARC079	139.91	3	*C. upsaliensis*	1.77	182	33019	34.3	100	100
32979	UCTARC081	141.91sm	2	*C. jejuni* subsp. *doylei*	1.69	74	91463	30.5	100	100
33005	UCTARC082	142.91	3	*C. jejuni* subsp. *doylei*	1.68	97	82072	30.5	100	100
33006	UCTARC084	145.91	3	*C. upsaliensis*	1.76	149	38838	34.3	100	100
33007	UCTARC086	150.91	2	*C. jejuni* subsp. *doylei*	1.73	167	30581	30.5	100	100
33008	UCTARC092	166.91	1	*C. jejuni* subsp. *doylei*	1.74	137	72262	30.5	100	100
32980	UCTARC094	173.91	1	*C. upsaliensis*	1.75	133	39612	34.3	100	100
33009	UCTARC095	174.91	1	*C. coli*	1.81	56	124925	31.2	100	100
33010	UCTARC096	176.91	3	*C. upsaliensis*	1.69	146	33322	34.5	100	100
32981	UCTARC097	185.91	2	*C. upsaliensis*	1.70	118	49619	34.5	100	100
32982	UCTARC098	186.91	2	*C. upsaliensis*	1.76	148	35247	34.3	100	100
32983	UCTARC099	188.91	1	*C. upsaliensis*	1.76	129	39071	34.3	100	100
32984	UCTARC100	190.91	1	*C. upsaliensis*	1.76	114	39877	34.3	100	100
32985	UCTARC101	192.91	2	*C. jejuni* subsp. *doylei*	1.79	69	82660	30.5	100	100
33011	UCTARC103	199.91	1	*C. upsaliensis*	1.76	160	35262	34.3	100	100
33012	UCTARC104	201.91	3	*C. upsaliensis*	1.76	192	24227	34.3	100	100
32986	UCTARC105	203.91	2	*C. coli*	1.82	95	100352	31.1	100	100
33013	UCTARC106	204.91	1	*C. upsaliensis*	1.61	118	39298	35.1	100	100
33014	UCTARC107	205.91	3	*C. upsaliensis*	1.73	173	33025	34.4	100	100
33015	UCTARC108	210.91	1	*C. hyointestinalis*	1.74	135	32221	34.2	100	100
33016	UCTARC110	221.91	2	*C. upsaliensis*	1.87	210	35308	34.1	100	100
32987	UCTARC111	223.91A	2	*C. jejuni* subsp. *doylei*	1.72	134	54211	30.5	100	98.1 (1)
33017	UCTARC112	224.91	6	*C. jejuni* subsp. *doylei*	1.69	118	73605	30.5	100	100
34686	UCTARC113	234.91	1	*C. jejuni* subsp. *doylei*	1.60	645	4743	30.9	100	94.2 (3)
33018	UCTARC114	246.91	3	*C. jejuni* subsp. *doylei*	1.69	114	73237	30.5	100	100

33019	UCTARC115	247.91	1	*C. upsaliensis*	1.76	176	32703	34.3	100	100
33021	UCTARC118	258.91	2	*C. upsaliensis*	1.73	186	31203	34.3	100	100
33022	UCTARC119	259.91	1	*C. upsaliensis*	1.76	174	32820	34.3	100	100
33023	UCTARC120	223.91B	1	*C. upsaliensis*	1.84	215	27716	34.1	100	100
33024	UCTARC121	265.91	2	*C. jejuni* subsp. *doylei*	1.68	103	77310	30.5	100	100
33025	UCTARC124	274.91	1	*C. upsaliensis*	1.77	203	27466	34.3	100	100
34688	UCTARC125	277.91	1	*C. jejuni* subsp. *jejuni*	1.81	384	10946	30.3	100	100
33026	UCTARC126	280.91	1	*C. hyointestinalis*	1.78	55	108959	34.1	100	100
33027	UCTARC127	285.91	3	*C. upsaliensis*	1.76	203	27558	34.3	100	100
33028	UCTARC129	294.91	3	*C. jejuni* subsp. *doylei*	1.69	113	82279	30.5	100	100
33029	UCTARC130	296.91	3	*C. upsaliensis*	1.74	137	35476	34.4	100	100
33030	UCTARC131	312.91	2	*C. upsaliensis*	1.69	165	33381	34.5	100	100
33031	UCTARC132	315.91	2	*C. upsaliensis*	1.69	162	33326	34.5	100	100
33032	UCTARC133	325.91	3	*C. upsaliensis*	1.67	154	31639	34.5	100	100
33033	UCTARC134	327.91	1	*C. jejuni* subsp. *doylei*	1.73	107	75654	30.5	100	100
32988	UCTARC136	335.91	2	*C. upsaliensis*	1.67	173	31215	34.5	100	100
33034	UCTARC137	337.91	1	*C. upsaliensis*	1.75	211	22398	34.3	100	100
33035	UCTARC138	339.91	3	*C. upsaliensis*	1.81	190	32860	34.2	100	100
33036	UCTARC139	341.91	3	*C. jejuni* subsp. *jejuni*	1.61	33	143430	30.5	100	100
33037	UCTARC140	342.91	3	*C. jejuni* subsp. *doylei*	1.69	104	72952	30.5	100	100
32989	UCTARC141	348.91	2	*C. upsaliensis*	1.68	175	31863	34.5	100	100
32990	UCTARC142	351.91	1	*C. coli*	1.72	128	29220	31.3	100	100
33038	UCTARC144	355.91	3	*C. upsaliensis*	1.79	324	22214	34.3	100	100
32991	UCTARC145	356.91	1	*C. coli*	1.65	153	21399	31.4	100	100
33039	UCTARC146	359.91	3	*C. upsaliensis*	1.56	122	33340	35	100	100
33040	UCTARC147	360.91	3	*C. upsaliensis*	1.76	174	29153	34.3	100	100

33041	UCTARC149	366.91	3	Putative novel *C. jejuni* subspecies	1.91	159	26624	30.4	100	100
33042	UCTARC150	367.91	3	*C. jejuni* subsp. *jejuni*	1.65	41	143895	30.5	100	100
33043	UCTARC152	375.99	3	*C. upsaliensis*	1.66	225	18451	34.5	100	100
33044	UCTARC153	379.91	2	*C. upsaliensis*	1.67	145	32869	34.5	100	100
33045	UCTARC154	381.91	2	*C. upsaliensis*	1.81	203	29172	34.2	100	100
33046	UCTARC156	383.91	3	*C. upsaliensis*	1.74	133	33319	34.1	100	100
33047	UCTARC157	391.91	3	*C. upsaliensis*	1.75	134	39138	34.1	100	100
33048	UCTARC158	397.91	3	*C. upsaliensis*	1.69	150	33240	34.5	100	100
33049	UCTARC160	403.91	5	*C. upsaliensis*	1.76	172	28064	34.3	100	100
33050	UCTARC161	411.91	2	*C. jejuni* subsp. *doylei*	1.77	131	47372	30.5	100	100
33051	UCTARC163	415.91	4	*C. jejuni* subsp. *doylei*	1.72	96	69221	30.4	100	100
33052	UCTARC164	417.91	3	*C. jejuni* subsp. *jejuni*	1.73	70	111405	30.2	100	100
32992	UCTARC165	418.91	2	*C. jejuni* subsp. *doylei*	1.68	143	51910	30.5	100	100
33053	UCTARC166	419.91	3	*C. jejuni* subsp. *jejuni*	1.60	41	182218	30.5	100	100
33054	UCTARC167	420.91	3	*C. jejuni* subsp. *jejuni*	1.65	51	137817	30.5	100	100
33055	UCTARC168	425.91	5	*C. upsaliensis*	1.70	182	33112	34.5	100	100
32993	UCTARC169	430.91	2	*C. jejuni* subsp. *jejuni*	1.67	94	48094	30.4	100	100
32994	UCTARC170	431.92	2	*C. jejuni* subsp. *jejuni*	1.62	110	51941	30.5	100	100
33056	UCTARC171	432.91	5	*C. upsaliensis*	1.69	171	32785	34.4	100	100
33057	UCTARC172	433.91	3	*C. jejuni* subsp. *doylei*	1.69	101	83812	30.5	100	100
33059	UCTARC174	437.91	3	*C. jejuni* subsp. *jejuni*	1.62	81	46820	30.5	100	100
32995	UCTARC175	439.91	2	*C. jejuni* subsp. *jejuni*	1.71	90	60325	30.2	100	100
33060	UCTARC176	445.91	3	*C. jejuni* subsp. *jejuni*	1.67	57	127796	30.4	100	100
33061	UCTARC177	455.91	2	*C. jejuni* subsp. *jejuni*	1.62	65	112346	30.5	100	100
33062	UCTARC178	456.91	3	*C. upsaliensis*	1.75	171	28357	34.3	100	100
33063	UCTARC179	458.91	3	*C. jejuni* subsp. *jejuni*	1.64	53	62525	30.5	100	100

33064	UCTARC180	464.91	3	*C. jejuni* subsp. *jejuni*	1.64	44	100575	30.5	100	100
33065	UCTARC182	470.91	3	*C. jejuni* subsp. *jejuni*	1.75	57	113401	30.4	100	100
33066	UCTARC183	471.91	5	*C. upsaliensis*	1.73	157	35194	34.3	100	100
33067	UCTARC184	481.91	5	*C. jejuni* subsp. *jejuni*	1.75	61	185020	30.3	100	100
33068	UCTARC186	141.91lg?	3	*C. upsaliensis*	1.87	189	31985	34	100	100
33069	UCTARC191	70.91	1	*C. upsaliensis*	1.70	281	13234	34.4	100	100
33070	UCTARC198	115.91	1	*C. coli*	1.92	128	45842	30.9	100	100
33071	UCTARC201	153.91	1	*C. jejuni* subsp. *doylei*	1.69	101	76952	30.5	100	100
33072	UCTARC203	224.91 CephR da	1	*C. jejuni* subsp. *doylei*	1.68	101	65714	30.5	100	100
33073	UCTARC206	411.91	1	*C. jejuni* subsp. *doylei*	1.77	132	46832	30.5	100	100
33074	UCTARC207	464.91	1	*C. jejuni* subsp. *jejuni*	1.64	284	9764	30.5	100	100

Table I2 Details of *Campylobacter* isolates from the 2011-2012 collection for which WGS data were obtained

Isolate id	Isolate	Species	Size (Mb)	Contigs (n)	N50 (bp)	GC content (%)	MLST	rMLST
25261	UCT020	*C. coli*	1.67	46	142524	31.4	100	100
25284	UCT064	*C. coli*	1.70	41	142392	31.3	100	100
25222	UCT003	*C. jejuni* subsp. *doylei*	1.74	101	80042	30.4	100	100
25228	UCT011B	*C. jejuni* subsp. *doylei*	1.69	102	82990	30.5	100	100
25232	UCT023	*C. jejuni* subsp. *doylei*	1.74	113	82142	30.4	100	100
25248	UCT039	*C. jejuni* subsp. *doylei*	1.69	86	73467	30.5	100	100
25250	UCT041	*C. jejuni* subsp. *doylei*	1.69	99	83051	30.5	100	100
24316	UCT046	*C. jejuni* subsp. *doylei*	1.68	108	80277	30.5	100	100
25220	UCT001	*C. jejuni* subsp. *jejuni*	1.70	31	160992	30.4	100	100
25221	UCT002	*C. jejuni* subsp. *jejuni*	1.73	82	129899	30.4	100	100
25223	UCT004	*C. jejuni* subsp. *jejuni*	1.69	51	138574	30.4	100	100
25257	UCT005	*C. jejuni* subsp. *jejuni*	1.78	38	247149	30.2	100	100
25258	UCT007	*C. jejuni* subsp. *jejuni*	1.69	41	114737	30.4	100	100
25224	UCT008	*C. jejuni* subsp. *jejuni*	1.81	52	288652	30.2	100	100
25225	UCT009	*C. jejuni* subsp. *jejuni*	1.68	40	127414	30.4	100	100
25226	UCT011A	*C. jejuni* subsp. *jejuni*	1.66	38	153960	30.4	100	100
25227	UCT012	*C. jejuni* subsp. *jejuni*	1.72	75	173546	30.2	100	100
25229	UCT015	*C. jejuni* subsp. *jejuni*	1.69	47	114986	30.4	100	100
25259	UCT016	*C. jejuni* subsp. *jejuni*	1.65	43	148504	30.4	100	100
25260	UCT017	*C. jejuni* subsp. *jejuni*	1.77	45	117880	30.3	100	100
25230	UCT019	*C. jejuni* subsp. *jejuni*	1.65	32	183584	30.4	100	100
25233	UCT025	*C. jejuni* subsp. *jejuni*	1.73	40	194123	30.3	100	100
25234	UCT026	*C. jejuni* subsp. *jejuni*	1.64	37	114936	30.5	100	100
25235	UCT027	*C. jejuni* subsp. *jejuni*	1.73	33	194117	30.3	100	100
25236	UCT028	*C. jejuni* subsp. *jejuni*	1.67	31	220147	30.3	100	100
25237	UCT030	*C. jejuni* subsp. *jejuni*	1.73	45	157829	30.3	100	100
25241	UCT031	*C. jejuni* subsp. *jejuni*	1.75	35	291804	30.3	100	100
25242	UCT033	*C. jejuni* subsp. *jejuni*	1.79	64	114228	30.3	100	100
25243	UCT034	*C. jejuni* subsp. *jejuni*	1.61	33	175550	30.5	100	100
25244	UCT035	*C. jejuni* subsp. *jejuni*	1.69	20	160535	30.3	100	100
25245	UCT036	*C. jejuni* subsp. *jejuni*	1.63	32	180249	30.4	100	100
25247	UCT038	*C. jejuni* subsp. *jejuni*	1.76	72	121037	30.4	100	100
25251	UCT042	*C. jejuni* subsp. *jejuni*	1.84	79	183878	30.1	100	100
25253	UCT045	*C. jejuni* subsp. *jejuni*	1.76	46	157442	30.3	100	100
25270	UCT049	*C. jejuni* subsp. *jejuni*	1.75	66	490155	30.1	100	100
25271	UCT050	*C. jejuni* subsp. *jejuni*	1.74	50	155517	30.2	100	100
25272	UCT051	*C. jejuni* subsp. *jejuni*	1.65	42	151308	30.4	100	100
25274	UCT054	*C. jejuni* subsp. *jejuni*	1.64	44	239485	30.5	100	100
25275	UCT055	*C. jejuni* subsp. *jejuni*	1.66	40	153981	30.4	100	100
25276	UCT056	*C. jejuni* subsp. *jejuni*	1.64	31	220307	30.4	100	100

25277	UCT057	*C. jejuni* subsp. *jejuni*	1.75	61	114995	30.4	100	100
25278	UCT058	*C. jejuni* subsp. *jejuni*	1.75	45	147175	30.2	100	100
25279	UCT059	*C. jejuni* subsp. *jejuni*	1.69	36	168954	30.4	100	100
25280	UCT060	*C. jejuni* subsp. *jejuni*	1.84	87	183874	30.1	100	100
25281	UCT061	*C. jejuni* subsp. *jejuni*	1.77	50	142925	30.3	100	100
25283	UCT063	*C. jejuni* subsp. *jejuni*	1.60	30	156059	30.4	100	100
25231	UCT022	*C. upsaliensis*	1.77	156	39075	34.3	100	100
34777	UCT024	*C. upsaliensis*	1.77	213	32871	34.3	100	100
25246	UCT037	*C. upsaliensis*	1.76	147	39858	34.3	100	100
25249	UCT040	*C. upsaliensis*	1.76	137	39086	34.3	100	100
34776	UCT044	*C. upsaliensis*	1.82	212	33645	34.2	100	100
24317	UCT047	*C. upsaliensis*	1.77	161	38285	34.3	100	100
25256	UCT048	*C. upsaliensis*	1.77	158	39324	34.3	100	100
25273	UCT052	*C. upsaliensis*	1.77	151	38982	34.3	100	100
34778	UCT053	*C. upsaliensis*	1.76	170	33015	34.3	100	100
25282	UCT062	*C. upsaliensis*	1.76	136	38327	34.3	100	100