# Transmission and evolution of bacteria during the course of enteritis outbreaks

A thesis submitted in partial fulfilment of the

requirements for the degree of Doctor of Philosophy

Massey University,

Palmerston North,

New Zealand

Samuel Bloomfield

2017

# Abstract

Bacterial enteritis outbreaks are a worldwide problem. They are hard to investigate as the bacterial agents are often associated with multiple sources, closely-related bacteria often co-colonise these sources, highly discriminatory tests are often required to distinguish between these bacteria, and bacteria are continuously evolving, changing how they behave. In this thesis I investigated the transmission and evolution of bacteria over the course of enteritis outbreaks by integrating genomic, phenotypic and antibiotic susceptibility testing, and phylogenetic modelling in four studies.

The aim of the first study was to investigate the origin, evolution and transmission of *Salmonella enterica* serovar Typhimurium DT160 over a 14-year long outbreak in New Zealand. Genomic analysis of 109 DT160 isolates collected over this timeframe established that the DT160 strain was introduced into New Zealand approximately a year before the first human isolate was reported; there were frequent transmissions between the source groups investigated (human, wild bird, poultry and bovine); and there was no evidence of specific selective pressures imposed on DT160. This study demonstrated how genomic analyses can be used to investigate extended outbreaks of bacterial diseases.

The aim of the second study was to investigate whether two ancestral state reconstruction models (the discrete trait analysis and structured coalescent models) were applicable to salmonellosis outbreak investigations. Both models were used to estimate transmission and population parameters of simulated salmonellosis outbreaks. Comparisons between the models' estimates and the true transmission and population values for the simulations revealed that both models made assumptions that did not apply to outbreaks and prevented them from accurately predicting these parameters. This study highlighted the need for outbreak-specific phylogenetic transmission models.

The aim of the third study was to investigate the relationship between two strains of *Salmonella* that were the predominant causes of human salmonellosis in New Zealand in the 2000s (*S.* Typhimurium DT160 and *S.* Typhimurium DT56 variant), and identify potential reasons for one strain declining (DT160) as the other emerged (DT56 variant). This study demonstrated how genomic analyses can be used to compare *Salmonella* strains and identify genetic elements that may influence strain behaviour.

The aim of the fourth study was to investigate a patient that had presented excreting the same genotype of *Campylobacter*, *C. jejuni* ST45, on multiple occasions over a 10-year period. Genomic analyses, phenotypic testing and antimicrobial susceptibility testing of sixteen *Campylobacter* isolates collected from the

patient found that the patient was persistently colonised with *Campylobacter* over this period, and that the *Campylobacter* had adapted to long-term colonisation by altering its motily and developing resistance to the antibiotics the patient had been prescribed. This study demonstrated how genomic analyses can be used to investigate a patient's infection history.

These studies demonstrated the applicability and limitations of genomic analyses when investigating bacterial enteritis outbreaks, how genetics and the environment influence bacterial evolution, and highlighted areas in the fields of microbiology, phylogenetics, epidemiology and public health that require further research.

# Acknowledgements

I sincerely thank my supervisory team (Dr Jackie Benschop, Dr Anne Midwinter, Assoc Prof Patrick Biggs, Assoc Prof David Hayman, Dr Jonathan Marshall, Dr Philip Carter and Prof Nigel French) for their guidance and help throughout my PhD. Specifically, I thank: Dr Jackie Benschop for continually encouraging and guiding me over the course of my PhD, and helping me maintain strong relationships with individuals involved in the studies outlined in this thesis; Dr Anne Midwinter for proof-reading all my first drafts (regardless of quality), helping with laboratory work and always having an ear for my opinions, ideas and thoughts, even the work-related ones; Assoc Prof Patrick Biggs for guiding me in all work bioinformatics-related, and helping me make my work as aesthetically-pleasing as possible; Assoc Prof David Hayman for helping me with modelling and for guidance regarding the future-directions of my projects and career; Dr Jonathan Marshall for helping with all work statistics- and mathematics-related; Dr Philip Carter for helping me obtain isolates and for providing insight into the *Salmonella*-collection and phage-typing processes at ESR; and finally Prof Nigel French for helping with the directions of the studies outlined in this thesis and helping me network with other experts to improve the studies.

I extend my thanks to all other experts that helped with the studies outlined in this thesis: Dr Tim Vaughan for his expertise on phylogenetic analyses, Dr Alison Mather for her expertise on salmonellosis outbreaks, and Dr Craig Thornley, Dr Rudyard Yap and Tui Shadbolt for helping set up the *Campylobacter* project.

I thank all $^{m}$EpiLab laboratory staff that have helped me perform the tests outlined in this thesis. Specifically, I thank Lynn Rogers for helping set up antimicrobial susceptibility tests and PCR reactions, and Tania Buwalda for helping with media preparation. To all the staff and students that I have shared an office with over the course of my PhD I thank you for your support and for putting up with my quirks.

I thank Massey University for allowing me to undertake my PhD and providing me with a Massey Doctoral Scholarship that funded my living costs, the Allan Wilson Centre that funded the *Salmonella* projects outlined in this thesis, the Institute of Veterinary and Biological Sciences (IVABS) post-graduate fund and Palmerston North Medical Research Fund (PNMRF) that funded the *Campylobacter* project outlined in this thesis, and the IVABS travel fund that funded conference expenses.

Finally, I thank all my family, friends and flatmates, particularly my parents, for supporting me emotionally, and for putting up with me during the highs and lows of my PhD.

# Abbreviations

**µl** ..... Microlitre

**°C** ..... Degrees Celsius

**×g** ..... Times gravity

**bp** ..... Base pairs

**CDC** ..... Centre for Disease Control and Prevention

***C. difficile*** ..... *Clostridium difficile*

**CE** ..... Common Era

***C. jejuni*** ..... *Campylobacter jejuni*

**CI** ..... Confidence interval

**COG** ..... Cluster of Orthologous Group

**CRC** ..... Canterbury Regional Council

**CVID** ..... Common variable immune deficiency

**Df** ..... Degrees of freedom

**DNA** ..... Deoxyribonucleic acid

**DT** ..... Definitive type

**DTA** ..... Discrete trait analysis

**EFSA** ..... European Food Safety Authority

***E. coli*** ..... *Escherichia coli*

**ENA** ..... European Nucleotide Archive

**ESR** ..... Institute of Environmental Science and Research

**EUCAST** ..... European Society of Clinical Microbiology and Infectious Diseases

**GC** ..... Guanine-cytosine

**GMRF** ..... Gaussian Markov Random Fields

**GTR** ..... Generalised time reversible

**GWRC** ..... Greater Wellington Regional Council

**HBRC** ..... Hawke's Bay Regional Council

**HKY** ..... Hasegawa, Kishino and Yano

**HPD** ..... Highest posterior density

**HIV** ..... Human immunodeficiency virus

**IgA** ..... Immunoglobulin A

**IgG** ..... Immunoglobulin G

**INDELS** ..... INsertions/DELetions

**IVABS** ..... Institute of Veterinary, Animal and Biological Sciences

**kb** ..... Kilo base pairs

**Mb** ..... Mega base pairs

**MCDHB** ..... MidCentral District Health Board

**MCMC** ..... Markov chain Monte Carlo

**MDS** ..... Multidimensional scaling

**$^{m}$EpiLab** ..... Massey University Molecular Epidemiology and Public Health Laboratory

**ml** ..... Millilitre

**mm** ..... Millimetre

**mM** ..... Millimolar

**MLST** ..... Multilocus sequence typing

**MSE** ..... Mean squared error

**MSS** ..... Mean sum of squares

**MSSS** ..... Manawatu sentinel surveillance site

**NCBI** ..... National Centre for Biotechnology Information

**NZGL** ..... New Zealand Genomics Limited

***P. aeruginosa*** ..... *Pseudomonas aeruginosa*

**PBS** ..... Phosphate-buffered saline

**PCR** ..... Polymerase chain reaction

**PFGE** ..... Pulsed-field gel electrophoresis

**PNMRF** ..... Palmerston North Medical Research Fund

**SC** ..... Structured coalescent

***S. enterica*** ..... *Salmonella enterica*

***S.* Enteritidis** ..... *Salmonella enterica* serovar Enteritidis

**SIR** ..... Susceptible-Infected-Recovered

**SNP** ..... Single nucleotide polymorphism

**SS** ..... Sum of squares

**ST** ..... Sequence type

**STEC** ..... Shiga toxin-producing *Escherichia coli*

***S.* Typhi** ..... *Salmonella enterica* serovar Typhi

***S.* Typhimurium** ..... *Salmonella enterica* serovar Typhimurium

**UK** ..... United Kingdom

**US** ..... United States

**WHO** ..... World Health Organisation

**XML** ..... Extensible Markup Language

# Publications

## Journals

Bloomfield, S.J., Benschop, J., Biggs, P.J., Marshall, J.C., Hayman, D.T.S., Carter, P.E., Midwinter, A.C., Mather, A.E. and French, N.P. (2017) Genomic analysis of *Salmonella enterica* serovar Typhimurium DT160 associated with a 14-year outbreak, New Zealand, 1998-2012. *Emerging Infectious Diseases* 23: 906-913

## Conferences

Bloomfield, S.J., Benschop, J., Biggs, P.J., Marshall, J.C., Hayman, D.T.S., Carter, P.E., Midwinter, A.C., Mather, A.E. and French, N.P. (2017) Genomic analysis of a decade-long outbreak of *Salmonella enterica* serovar Typhimurium DT160 in New Zealand. ASM Microbe 2017, New Orleans, LA, USA

Bloomfield, S.J., Benschop, J., Midwinter, A.C., Biggs, P.J., Marshall, J.C., Hayman, D.T.S., Carter, P.E., and French, N.P. (2017) Evolution of *Campylobacter jejuni* within a long-term human host. Phylogeneomics 2017, Waiheke Island, New Zealand

Bloomfield, S.J., Benschop, J., Midwinter, A.C., Biggs, P.J., Marshall, J.C., Hayman, D.T.S., Carter, P.E., and French, N.P. (2016) Evolution of *Campylobacter* in a persistently colonised human host. IDReC symposium 2016, Wellington, New Zealand

Bloomfield, S.J., Benschop, J., Midwinter, A.C., Biggs, P.J., Hayman, D.T.S., Marshall, J.C., Carter, P.E., and French, N.P. (2015) Evolution of *Campylobacter* in a persistently infected human host. CHRO 2015, Rotorua, New Zealand

Bloomfield, S.J., Benschop, J., Midwinter, A.C., Biggs, P.J., Marshall, J.C., Jaros, P., Carter, P.E., Hayman, D.T.S., and French, N.P. (2015) Death, disease and supersession: new approaches to salmonellosis outbreak analysis using molecular techniques. Allan Wilson Centre meeting 2015, Palmerston North, New Zealand

Bloomfield, S.J., Benschop, J., Midwinter, A.C., Biggs, P.J., Marshall, J.C., Carter, P.E., Hayman, D.T.S., and French, N.P. (2014) Transmission and evolution of *Salmonella* outbreaks. Allan Wilson Centre meeting 2014, Palmerston North, New Zealand

# Contents

# List of figures

# List of tables

# Chapter 1

# Introduction

## 1.1 Epigraph

'In the eighteenth century, no such thing [germs], nada, nothing. No one ever imagined such a thing. No sane person, anyway. Ah! Ah! Along comes this doctor, uh, uh, uh Semmelweis, Semmelweis. Semmeilweis comes along. He's trying to convince people, well, other doctors mainly, that there's these teeny tiny invisible bad things called germs that get into your body and make you sick. Ah? He's trying to get doctors to wash their hands. What is this guy? Crazy? Teeny, tiny, invisible? What do you call it? Uh-uh, germs? Huh? What? Now, cut to the $20^{\text{th}}$ century. Last week, as a matter of fact, before I got dragged into this hellhole. I go in to order a burger in this fast food joint, and the guy drops it on the floor. Jim, he picks it up, he wipes it off, he hands it to me like it's all OK. "What about the germs?" I say. He says, "I don't believe in germs. Germs is just a plot they made up so they can sell you disinfectants and soaps." Now he's crazy, right? See?' -Jeffrey Groines (Twelve Monkeys, 1995).

## 1.2 Introduction

Bacterial enteritis is inflammation of the intestines due to the actions of bacteria and the host's response. Outbreaks of bacterial enteritis continue to be a problem in New Zealand and worldwide, as evidenced by the 2016 *Campylobacter* outbreak in Havelock North, New Zealand that affected more than 5,500 individuals and was associated with three deaths (Department of Internal Affairs, 2017), and the 2011 *Escherichia coli* O104:H4 outbreak in Europe that affected 3,950 individuals and was associated with 53 deaths (EFSA, 2012). In this thesis I investigated bacterial enteritis outbreaks, how the bacteria evolve and are transmitted over the course of the outbreaks, and the methods used to analyse them.

## 1.3 Structure

The aim of this thesis was to investigate the transmission and evolution of bacteria over the course of enteritis outbreaks. This is a very broad topic and this thesis is not intended to cover all the complexities associated

with this topic. Instead it focuses on several case studies of bacterial enteritis outbreaks and then pulls these studies together in the general discussion. To achieve this, the chapters of this thesis were written by publication with a complete description of the necessary background, methods (even if similar), results, and discussion for each data set.

This current chapter (Chapter 1) is intended to introduce the reader to the topic, the structure of the thesis and the terminology used. Chapter 2 is a literature review that critiques the literature regarding bacterial enteritis outbreaks, their origin, the transmission and evolutionary dynamics of bacterial populations over the course of outbreaks, and the methods used to analyse them. Chapter 3 uses several of the genomic methods described in Chapter 2 to investigate the transmission and evolution of *Salmonella enterica* serovar Typhimurium DT160 over the course of a 14-year outbreak in New Zealand. Chapter 4 uses simulated salmonellosis outbreaks to investigate the applicability of two ancestral state reconstruction models for predicting bacterial transmission and population parameters of these outbreaks. Chapter 5 uses genomic analyses to investigate two *Salmonella* strains (*S.* Typhimurium DT160 (from Chapter 3) and *S.* Typhimurium DT56 variant) that were subsequently the largest causes of salmonellosis in New Zealand, to determine their relationship, to identify differences between the strains and to identify potential reasons why one strain declined as the other emerged. Chapter 6 uses multiple genomic analyses described in Chapter 2 to investigate a patient who had repeatedly presented excreting the same strain of *Campylobacter* (*C. jejuni* ST45) over a 10-year period, to provide insight into the patient's infection history. Chapter 7 is the general discussion that compares the different research chapters, how they contribute to our current knowledge of bacterial enteritis outbreaks and outlines future directions for this research.

## 1.4 Terminology

Enteritis is "Inflammation of the intestine, especially the small intestine, usually accompanied by diarrhoea" (Oxford, 2017). The term 'enteritis' was used in this thesis to broadly describe the diseases resulting from *Campylobacter* and non-typhoidal *Salmonella* intestinal infections, as described in Chapters 3-6. In this thesis I could have used 'gastroenteritis' instead of 'enteritis', as it has been defined as "...a catchall term for infection or irritation of the digestive tract, particularly the stomach and intestines' (Medical Dictionary, 2017). However, gastroenteritis has also been defined as "...inflammation of the lining membrane of the stomach and the intestines characterised especially by nausea, vomiting, diarrhoea and cramps" (Merriam-Webster, 2017b). This definition does not apply to non-typhoidal *Salmonella* and *Campylobacter* infections that do not usually affect the stomach and are not invariably associated with vomiting (Fitzgerald and Nachamkin, 2007; Nataro et al., 2007). Therefore, enteritis was used in this thesis to describe the diseases caused by non-typhoidal *Salmonella* strains and *Campylobacter*.

A disease outbreak is "...the occurrence of disease in excess of what would normally be expected in a defined community, geographical area or season. An outbreak may occur in a restricted geographical area, or

may extend over several countries. It may last for a few days or weeks, or for several years" (WHO, 2016). In Chapter 6 I discuss the continued excretion of *Campylobacter* from a single patient as an outbreak, for it is in excess of what would normally be expected (Porter and Reid, 1980; Rao et al., 2001), but traditionally the term 'outbreak' has referred to an increase in the number of cases as opposed to an extended case (Bingham et al., 2004). For the purposes of this thesis, I will regard this case as an outbreak.

This thesis primarily focuses on intestinal infections caused by non-typhoid *Salmonella* and *Campylobacter*. Unless otherwise stated, salmonellosis and campylobacteriosis are used to describe intestinal infections caused by these bacteria in this thesis, respectively (CDC, 2015; Oxford, 2017). However, it is understood that salmonellosis and campylobacteriosis have been described as any infection caused by non-typhoid *Salmonella* and *Campylobacter*, respectively (Merriam-Webster, 2017b,a).

# References


Bingham, P., Verlander, N. Q., and Cheal, M. J. (2004). John Snow, William Farr and the 1849 outbreak of cholera that affected London: A reworking of the data highlights the importance of the water supply. *Public Health*, 118(6):387–394.

CDC (2015). Antibiotic treatment: Recommendations for the use of antibiotics for the treatment of cholera. Retrieved 2017-06-29, from: `http://www.cdc.gov/cholera/treatment/antibiotic-treatment.html`, 1-1.

Department of Internal Affairs (2017). Report of the Havelock North drinking water inquiry: Stage 1. Retrieved 2017-06-18, from: `https://www.dia.govt.nz/vwluResources/Report-Havelock-North-Water-Inquiry-Stage-1/$file/Report-Havelock-North-Water-Inquiry-Stage-1.pdf`, 1–230.

EFSA (2012). *E. coli*: Rapid response in a crisis. Retrieved 2017-05-06, from: `http://www.efsa.europa.eu/en/press/news/120711`, 1–1.

Fitzgerald, C. and Nachamkin, I. (2007). *Campylobacter* and *Arcobacter*. In Murray, P. R., Baron, E. J., Jorgensen, M. L., Landry, M. L., and Pfaller, M. A., editors, *Manual of Clinical Microbiology*, chapter 59, pages 933–946. Washington DC, 9th edition.

Medical Dictionary (2017). Gastroenteritis. Retrieved 2017-05-04, from: `http://medical-dictionary.thefreedictionary.com/gastroenteritis`, 1–1.

Merriam-Webster (2017a). Evolutionary biology. Retrieved 2017-05-04, from: `https://www.merriam-webster.com/dictionary/evolutionarybiology`, 1–1.

Merriam-Webster (2017b). Gastroenteritis. Retrieved 2017-05-04, from: `https://www.merriam-webster.com/dictionary/gastroenteritis`, 1–1.

Nataro, J. P., Bopp, C. A., Fields, P. I., Kaper, J. B., Strockbine, Nancy, A., Strockbine, B., and Strockbine, N. A. (2007). *Escherichia*, *Shigella* and *Salmonella*. In Murray, P. R., Baron, E. J., Jorgensen, J. H., Landry, M. L., and Pfaller, M. A., editors, *Manual of Clinical Microbiology*, chapter 43, pages 670–687. ASM Press, Washington DC, 9th edition.

Oxford (2017). Enteritis. Retrieved 2017-05-04, from: `https://en.oxforddictionaries.com/definition/enteritis`, 1–1.

Porter, I. A. and Reid, T. M. S. (1980). A milk-borne outbreak of *Campylobacter* infection. *Journal of Hygiene*, 84(3):415–419.

Rao, M. R., Naficy, A. B., Savarino, S. J., Abu-Elyazeed, R., Wierzba, T. F., Peruski, L. F., Abdel-Messih, I., Frenck, R., and Clemens, J. D. (2001). Pathogenicity and convalescent excretion of *Campylobacter* in rural Egyptian children. *American Journal of Epidemiology*, 154(2):166–173.

WHO (2016). Disease outbreaks. Retrieved 2017-06-29, from: `http://www.who.int/topics/disease_outbreaks/en/`, 1–1.

# Chapter 2

## Literature review

### 2.1 Introduction

Bacterial enteritis is inflammation of the intestines due to the actions of bacteria and the host's response. In New Zealand bacterial enteritis is an important infectious disease, with an annual incidence of 183 cases per 100,000. The predominant causes in New Zealand are *Campylobacter* and non-typhoid *Salmonella* (ESR, 2015). However, a significant number of other bacteria can cause bacterial enteritis, including: *Shigella*, Shiga toxin-producing *Escherichia coli* (STEC), *Yersinia enterocolitica*, *Vibrio parahaemolyticus* and *Clostridium difficile*. Each of these agents is associated with different biochemistries, genetics and behaviours (Fitzgerald and Nachamkin, 2007; Nataro et al., 2007).

An outbreak of disease is defined as "...the occurrence of disease in excess of what would normally be expected in a defined community, geographical area or season. An outbreak may occur in a restricted geographical area, or may extend over several countries. It may last for a few days or weeks, or for several years" (WHO, 2016a). Bacterial enteritis outbreaks are a problem worldwide (Byrne et al., 2014; Howie et al., 2003). From 1998-2008, bacteria were responsible for 3,613 outbreaks of enteritis in the United States, resulting in 92,093 illnesses, 6,446 hospitalisations and 148 deaths (Gould et al., 2013). However, 30% of the outbreaks investigated in this paper had an unknown aetiology, and bacterial enteritis cases are often underreported (Mellou et al., 2013). Therefore, bacteria were probably responsible for a larger number of outbreaks, hospitalisations and deaths than reported. To decrease the number and severity of bacterial enteritis outbreaks, a large amount of research has been put into understanding how they are initiated, how they behave and what factors influence them (Nyberg et al., 2011; Yoon et al., 2004). The aim of this literature review was to critique the literature regarding bacterial enteritis outbreaks, their origin, the transmission and evolutionary dynamics of the bacterial agents that cause them, and the methods used to analyse them.

### 2.2 Sources of bacterial enteritis

Sources of bacterial enteritis are the humans, animals, objects or places via which susceptible individuals are exposed to bacterial enteritis agents. Identifying the source of an infection can be difficult, as multiple

sources may harbour the agent and distinguishing between isolates from each source can be difficult (Dyet et al., 2011). However, identifying the sources of infection allows protocols and practices to be put in place to limit infections from the source (Sears et al., 2011). Sources of bacterial enteritis can be divided into reservoirs, pathways, exposures and risk factors for control and prevention purposes (Wagenaar et al., 2013).

### 2.2.1 Pathogen reservoirs

Infectious disease reservoirs are "...one or more epidemiologically connected populations or environments in which the pathogen can be permanently maintained and from which infection is transmitted to the defined target population"(Haydon et al., 2002). Most bacterial enteritis agents are zoonotic with one or more animals acting as the reservoir, e.g. *Campylobacter* is found in the gastrointestinal tract of chickens, cows and various other land-dwelling animals. Some bacterial enteritis agents are not found in animals. These agents are directly transmitted by the faecal-oral route between humans who act as their reservoir. Multiple control measures have been designed to decrease the number of bacterial enteritis agents in reservoirs and prevent transmission from reservoirs, but their effectiveness varies with the infectious agent and reservoir (Bolder et al., 1999).

#### 2.2.1.1 Control measures for human reservoirs

Bacterial enteritis agents with human reservoirs can be controlled by treating infected humans (Mahoney et al., 1993). Treatment varies with the infectious agent, its antimicrobial susceptibility profile, the age and gender of the patient, and the severity of symptoms (CDC, 2015). Outbreaks of these diseases are usually associated with marginalised groups in developed counties that have limited access to medical resources (Hines et al., 2016). They are more common in developing countries, where it is more difficult to obtain medical attention, leaving a large number of untreated humans reservoirs and outbreaks (Kotloff et al., 1999).

Asymptomatic carriers excrete the bacterial enteritis agent but do not display any disease symptoms, e.g. *Salmonella enterica* serovar Typhi carriers. Asymptomatic carriers do not usually seek treatment and instead continue to spread the bacterial agent (Lynch et al., 2009). They are more prevalent in devloping countries due to increased exposure from suboptimal hygiene, and, possibly, undernutrition-induced immunosuppression (Lee et al., 2013). Much research effort has been placed into creating sensitive, specific tests that can help identify asymptomatic carriers before quarantining or treating them (Pratap et al., 2013). It is often impractical to screen an entire population. Instead, individuals that are more likely to spread the bacteria are usually targeted for screening, e.g. those that work with food (Feglo et al., 2004).

#### 2.2.1.2 Control measures for animal reservoirs

Animal reservoirs of human bacterial enteritis agents may not be affected by the agent. Animal hosts that are clinically affected can be controlled with medical treatment, similar to human reservoirs, or quarantining or culling infected animals (Sternberg et al., 2008). When the bacteria has no effect on the animal host, other processes are required to control the agent.

Phages are viruses that infect and replicate in bacteria. Phages may be used to decrease the carriage of bacterial enteritis agents in animal reservoirs. For example, Borie et al. (2008) was able to decrease the concentration of *Salmonella* in chickens by feeding them a cocktail of phages. One of the advantages of phage treatment is that the concentration of bacteria in the reservoir has little effect on the treatment, as the phages will continue to replicate and lyse bacteria until most susceptible isolates are lysed in the animal host (El-Shibiny et al., 2009). One of the disadvantages is that phage treatment selects for phage resistance, so may require continued reassessment of efficacy and sourcing of novel phages. However, phage resistance is associated with decreased virulence in some bacteria (Loc Carrillo et al., 2005). Phages are also very specific, which prevents them altering an animal's microbiota, but means that they are only effective against a narrow range of strains (Atterbury et al., 2007). Further research is required to create phage treatments that are effective against a wide range of bacterial enteritis strains and species, and to investigate phage treatment's long-term effectiveness.

Probiotics are live microorganisms that once consumed, may provide health benefits to the host. Probiotics may modulate the immune system by normalising the ratio of anti-inflammatory to pro-inflammatory cytokines (O'Mahony et al., 2005); compete with pathogens for resources and space in the gastrointestinal tract (Ahmadova et al., 2013); help the gastrointestinal tract develop, such that it can perform gastric and immune functions effectively (Babińska et al., 2005); help with the digestion of food, such that it does not linger and cause irritation (Lahti et al., 2013); promote the gastrointestinal tract's barrier function by increasing the number of tight junctions between intestinal epithelial cells (Ukena et al., 2007); or promote the production of various defensive compounds (e.g. $\alpha$-defensin and immunoglobulin A) (He et al., 2007). Probiotics are often used to help prevent and treat various gastrointestinal diseases, and to improve general health (Brenner and Chey, 2009; De Milliano et al., 2012). Their ability to outcompete and inhibit pathogen growth may be applicable to pathogen reservoirs.

Probiotics may be used to decrease the carriage of bacterial enteritis agents in animal reservoirs. For example, Ghareeb et al. (2012) found that feeding chickens *Enterococcus faecium*, *Pediococcus acidilactici*, *Lactobacillus salivarius* or *Lactobacillus reuteri* decreased *Campylobacter* carriage, whilst Zhao et al. (1998) found that feeding cows specific *E. coli* and *Proteus mirabilis* strains decreased *E. coli* 0157 carriage. In addition, these probiotics have been shown to improve the health and food yield of many animal reservoirs (Zhang and Kim, 2013). However, probiotic effectiveness varies with the specific probiotics used, dose, how they are administered and the animal they are used on (Higgins et al., 2008; Willis and Reid, 2008). Further

research is required to identify optimal probiotic treatments for decreasing the carriage by animal reservoirs of agents that cause bacterial enteritis in humans.

Vaccines are biological materials that stimulate an acquired immune response in animals or humans. This acquired immune response allows the animal to recognise the biological material and respond to it more quickly in the future. Vaccines may be used to prevent infectious agents colonising animals. For example, Vilte et al. (2012) found that cows vaccinated with *E. coli* 0157 bacterial ghosts (empty cell envelopes), developed an immune response to this bacterium, decreasing *E. coli* O157 colonisation and excretion, whilst Wyszyńska et al. (2004) found that chickens vaccinated with an avirulent *Salmonella* strain that expressed *Campylobacter* proteins, developed an immune response to these bacteria, decreasing *Salmonella* and *Campylobacter* excretion. These vaccines did not prevent excretion, they only decreased it, and were only effective against a narrow range of strains that may vary with different animals and geographical locations (House et al., 2001). Therefore, effective vaccine use may rely on identifying the bacterial enteritis strains present in a group of animals and tailoring the vaccine for these strains.

#### 2.2.2 Disease pathway

Disease pathways are the route by which disease-causing organisms are transmitted from reservoirs to susceptible individuals. Bacterial enteritis agents are usually transmitted to humans via foods that animal reservoirs have contaminated (Mughini-Gras et al., 2012). Humans can also be exposed to bacterial enteritis agents via environments that human and animal reservoirs have contaminated, e.g. contaminated waterways and soils (Dekker et al., 2015). It is difficult to eliminate bacterial enteritis agents from all environments and food chains, as bacterial enteritis reservoirs are associated with multiple environments and foods (Palhares et al., 2014). However, the number of foods and environments that are contaminated with these agents can be decreased.

##### 2.2.2.1 Control measures for food pathways

Humans consume many foods produced by or derived from animals that act as bacterial enteritis agent reservoirs. To combat food-borne bacterial enteritis, multiple methods have been developed to decrease the amount of bacterial enteritis agents entering the food chain.

Food processing is the preparation of food for consumption by humans and animals. It is used to decrease the number of microorganisms found on the food, decrease food spoilage, increase the shelf life of food and decrease the risk of disease. Multiple food processing techniques have been developed, including: heating (Bhavsar et al., 2007), freezing (Harrison et al., 2013), salting (Mol et al., 2010), irradiating (Kudra et al., 2011a) and pickling. However, food processing can alter the nutrient content of foods and/or taste (Ferraro et al., 2014). The effectiveness of different food processing techniques also varies with different foods (Montero

et al., 2007), and some methods are more effective at inhibiting certain bacteria than others (Ekhtiarzadeh et al., 2012). Therefore, effective food processing relies on prior knowledge of the food's chemical and physical properties, and what bacteria they are mostly likely to be colonised with.

After processing, foods must be stored until they are consumed. Food storage aims to increase the shelf-life of foods by preventing food contamination and inhibiting microorganism growth on food. Multiple food storage methods are available, including: refrigeration (Nguyen et al., 2014), freezing (Yoon et al., 2004), vacuum and modified atmosphere packing (Kudra et al., 2011b), and the addition of preservatives (Er et al., 2014). As with food processing, the effectiveness of different food storage techniques varies with the food (Zhang et al., 2012), and some methods are more effective at preventing the growth of certain bacteria (Nair et al., 2015). Therefore, effective food storage relies on prior knowledge of the food's chemical and physical properties, how it will be transported, and what bacteria the food is most likely to be colonised with.

#### 2.2.2.2 Control measures for environmental pathways

Bacterial enteritis agents have been isolated from multiple environments (Karp et al., 2015). Environmental contamination is usually the result of exposure to excrement from colonised or infected animals (Clark et al., 2003; Van Dyke et al., 2010). Environmental sources may not contribute to as many human cases as food sources. However, they allow other animals and plants to be exposed to these agents that can lead to human infections (Islam et al., 2004). Therefore, multiple methods have been designed to decrease or limit the amount of bacterial enteritis agents in the environment.

The amount of bacterial enteritis agents in the environment can be decreased by lowering bacterial survival. Strachan et al. (2002) found that the concentration of *E. coli* O157 in agricultural land was significantly reduced if the land was left for at least four weeks before recreational use. However, Ma et al. (2014) found that the survival of STEC strains varied with different types of soil, whilst Islam et al. (2004) found that the survival of *Salmonella* in soil varied with different fertilisers. Therefore, the amount of bacterial enteritis agents in an environment usually decreases over time, but the rate of decrease varies with different environmental conditions.

Multiple procedures have been designed to decrease the amount of bacterial enteritis agents in the environment. Ibekwe et al. (2007) decreased the concentration of *E. coli* O157 in soil by fumigation with methyl bromide and methyl iodide. However, these chemicals are known ozone-depleting chemicals and carcinogens, respectively. Also, fumigation effectiveness varies with soil composition (Ibekwe et al., 2010). Nyberg et al. (2011) decreased the concentration of *Salmonella* in soil by increasing the soil pH to 12 with calcium hydroxide. However, pH values this high are dangerous to humans and animals. Overall, new treatments need to be designed that decrease the amount of bacterial enteritis agents, but do not harm the environment.

Bacterial gene expression may hold the clue to environmentally-friendly treatments. Gene expression is the process in which information encoded in a gene is converted into a gene product. Bacterial enteritis

agents have thousands of genes, some of which are important for living in hosts, whilst others are important for surviving in soil and water (Piveteau et al., 2011). Bacterial enteritis agents alter their gene expression in response to different environments (Duffitt et al., 2011). Research into these genes has identified enzymes and pathways required for bacterial enteritis agent survival in the environment (Vivant et al., 2015). Treatments that target these pathways may decrease the concentration of bacterial enteritis agents whilst having little impact on the environment.

It is difficult to prevent animal reservoirs contaminating the environments they inhabit (Donnison et al., 2004). However, it is possible to prevent bacterial enteritis agents spreading from contaminated to adjacent environments. Buffer zones are sections of land that separate environments free of bacterial enteritis agents from potential reservoirs, e.g. waterways and agricultural soil. Strawn et al. (2013) found that produce (e.g. vegetables) fields surrounded by buffer zones were associated with decreased *Salmonella* and *Listeria* prevalence in the produce, whilst Wilkes et al. (2013) found that rivers lined with riparian zones that restricted livestock access were associated with a decreased concentration of *Campylobacter* and *E. coli* O157:H7 compared to sections of the same river where livestock had access to the river. Karp et al. (2015) also found that the effectiveness of buffer zones varied with their composition. Therefore, maintained buffer and riparian zones between bacterial enteritis-free environments and contaminated environments can be used to limit environmental contamination, but their composition must be considered.

#### 2.2.3 Pathogen exposures

Pathogen exposures are the specific ways in which susceptible individuals are exposed to infectious agents via disease pathways. Individuals are usually exposed to bacterial enteritis agents in the food chain via contaminated foods and drinks, or via recreational activities on contaminated environments, e.g. camping on agricultural land (Magnússon et al., 2011; Mughini-Gras et al., 2014). Susceptible individuals may also be exposed via direct contact with colonised humans and animals (Younus et al., 2010).

##### 2.2.3.1 Control measures for food exposures

Most susceptible individuals are exposed to bacterial enteritis agents via a contaminated food chain. Food exposure is usually via the consumption of unprocessed foods originating from animals (e.g. poultry and raw milk), but can arise from the consumption of vegetables and fruits exposed to animal excrement (Bayer et al., 2014; Gould et al., 2013; Waitt et al., 2014). Exposure can also arise from handling these contaminated foods and drinks (Pitout et al., 2003).

Food exposures can be difficult to control as susceptible individuals are often unaware of what foods and practices expose them to bacterial enteritis agents. Prevention may be achieved through interventions and campaigns that inform the public about these exposures, the risks associated with these exposures and

which practices will reduce transmission (Dharod et al., 2004). Campaigns usually involve identifying at risk individuals, informing this group about risky food practices (e.g. advertisements, skits or video games), and evaluating the effectiveness of the campaign (Crovato et al., 2016). Multiple food safety campaigns have been run, but few studies have evaluated the effectiveness of these campaigns (Abbot et al., 2012). Furthermore, none have evaluated the long-term effects of these campaigns. Further research is required into food safety campaigns, their effectiveness, and their long-term effects.

#### 2.2.3.2 Control measures for environmental exposures

Contaminated environments can expose multiple susceptible individuals to bacterial enteritis agents. Environmental exposure is usually through recreational activities involving contaminated soil or water (CDC, 2015; Howie et al., 2003), but outbreaks have resulted from the inhalation of aerosols in contaminated environments (Varma et al., 2003) and bacterial enteritis agents have been found in wild bird faecal samples on children's playgrounds (French et al., 2009).

Exposure to environmental bacterial enteritis agents can be prevented by informing the public about environments that contain bacterial enteritis agents and are not safe for use. Regional and national councils often advertise rivers that are safe for recreational use and those that are not (GWRC, 2016; HBRC, 2016). However, the larger problem for councils is identifying contaminated areas, as it involves persistent testing of environments (CRC, 2013).

As with food exposure, public health campaigns can be used to inform at-risk individuals about environmental bacterial exposure. These campaigns usually focus on informing individuals who are continually exposed to environments contaminated with bacterial enteritis agents, e.g. farmers (Klumb et al., 2013). However, Jones et al. (2011) found that rural residents in the United Kingdom were more aware of environmental exposure to *E. coli* O157 than visitors to rural areas, whilst Belongia et al. (2003) found that farm-resident children in Wisconsin, US, contained higher titres of anti-*E. coli* O157 and anti-*Campylobacter jejuni* antibodies than other rural children, possibly preventing clinical disease. This may explain why environmental outbreaks of bacterial enteritis often involve individuals that are irregularly exposed to these environments (Howie et al., 2003; Varma et al., 2003). Therefore, public health campaigns aimed at environmental bacterial enteritis exposure should focus on individuals who are regularly exposed, but should not exclusively target these individuals.

#### 2.2.3.3 Control measures for direct exposures

Susceptible individuals can be exposed to bacterial enteritis agents via direct exposure to contaminated or infected humans and animals.

Most bacterial enteritis agents are zoonotic with one or more animal reservoirs. Direct contact with these animal reservoirs can contaminate body parts with bacterial enteritis agents that in turn can transport these agents to the oral cavity, e.g. working on a farm, contacting pets or handling dead wild birds (Thornley et al., 2003; Warshawsky et al., 2002). It is difficult to prevent individuals directly contacting animals, especially if they contact them daily, e.g. farmers, veterinarians and pet owners. However, animal practices can be modified to minimise exposure.

Exposure to bacterial enteritis agents via direct animal contact can be decreased by wearing protective clothing when dealing with animals and by washing body parts exposed to animals. Washing decreases the number of bacterial enteritis agents on people, preventing consumption or transmission of the agent. There is a lot of variation in hygiene practices amongst professionals working with animals and what practices are considered acceptable (McMillian et al., 2007). It is also difficult to encourage hygienic practices around animals. For example, Anderson et al. (2014) investigated hand hygiene practices in veterinary clinics in Ontario, Canada. They found that hand hygiene practices were poor, even in the presence of posters that encouraged hand hygiene, but surveys indicated that the posters raised awareness of hand hygiene. Therefore, although washing body parts exposed to animals can decrease bacterial enteritis agent exposure, active campaigns are required to inform individuals at risk about the benefits of washing exposed body parts, and to measure the effectiveness of campaigns.

Bacterial enteritis agents may be transmitted between humans via direct exposure to infected or colonised humans, such that body parts become contaminated with bacterial enteritis agents and can transmit the agent to the oral cavity. Direct transmission can only transmit a small number of agents between humans. Therefore, bacterial enteritis agents with a low infectious dose (e.g. STEC) are more likely to cause an infection via direct contact than those with a higher infectious dose (e.g. non-typhoid *Salmonella*) (Tuttle et al., 1999). However, Hara-Kudo and Takatori (2011) found that in certain salmonellosis outbreaks, *Salmonella* had an infectious dose as low as STEC, and Mather et al. (2013) investigated a salmonellosis outbreak in Scotland using whole genome sequencing and found that more human cases were the result of exposure to human sources than from exposure to animal sources. The ancestral state reconstruction model used by Mather et al. has been shown to both over- and under-estimate transmission rates (De Maio et al., 2015). Therefore, further analysis is required to determine if direct exposure to infected humans contributes to these salmonellosis outbreaks.

As with direct animal exposure, direct human exposure to bacterial enteritis agents can be decreased by washing body parts exposed to infectious humans, and sterilizing environments that infectious humans have contaminated (Horn and Otter, 2015). Multiple methods have been developed for washing and decontaminating exposed body parts and environments, but they vary in effectiveness. For example, Stone et al. (2012) found that increased soap usage in hospitals in England and Wales decreased the incidence of *C. difficile*, but increased alcohol hand rub use had no effect. Therefore, exposure to bacterial enteritis agents via direct human contact can be decreased by improved hygiene, but the effectiveness of the methods used to improve

hygiene must be evaluated.

### 2.2.4 Risk factors for bacterial enteritis

Risk factors are attributes, characteristics or exposures that increase the likelihood of developing a disease or injury (WHO, 2016b). The risk factors for bacterial enteritis differ with the agent, the reservoir and the disease pathway (Mughini-Gras et al., 2012, 2014).

The human microbiota consists of all the microorganisms that live in or on an individual. The microbiota helps prevent bacterial enteritis by competing with pathogens for resources and priming the immune system (Taur and Pamer, 2014). Microbiota disruptions prevent this competition, allowing bacterial enteritis agents to colonise the gastrointestinal tract and harm the individual (O'Loughlin et al., 2015). Antibiotic use (Pérez-Cobas et al., 2014), diet and gastrointestinal diseases can affect the microbiota (David et al., 2015). Antibiotics are used to treat various infections, so it is hard to restrict their use. However, some antibiotics disrupt the microbiota more than others, whilst some alternatives to antibiotics, such as bacteriocins, are effective at treating certain infections whilst having little effect on the microbiota (Rea et al., 2011). Therefore, bacterial enteritis risk can be decreased by maintaining a healthy microbiota via diet, limited antibiotic use and the development of non-disruptive therapies.

Abattoirs are where animals are butchered for human consumption. Most of the animals butchered at abattoirs are animal reservoirs for bacterial enteritis agents. Therefore, working at an abattoir is a risk factor for bacterial enteritis (Ellström et al., 2014). Bacterial agents are often spread throughout abattoirs (Okraszewska-Lasica et al., 2014), making it difficult to prevent exposure to these agents. However, it is possible to identify high risk areas in abattoirs and modify working protocols to decrease the risk (De Perio et al., 2013; Cossi et al., 2014).

The consumption of certain foods is associated with an increased risk of bacterial enteritis. In an Australian case-control study, Unicomb et al. (2008) found that consumption of restaurant-prepared meats and fast-food were associated with an increased risk of campylobacteriosis. Doorduyn et al. (2006) conducted a case-control study in the Netherlands and found that the consumption of raw eggs was associated with an increased risk of salmonellosis from *S. enterica* serovar Enteritidis, whilst raw meat consumption was associated with an increased risk of salmonellosis from *S. enterica* serovar Typhimurium. Restaurant, fast-food, agriculture and food-processing standards vary between countries (Jol et al., 2006), as do diets and the presence of specific bacterial strains (Public Health Surveillance, 2017). Therefore, the consumption of certain foods is a risk factor for bacterial enteritis agents, but the risks vary between countries. Nevertheless, the consumption of some foods is a risk factor worldwide.

Pasteurisation is the process of heating a liquid for a short length of time to reduce the number of pathogens present. Consuming milk that has not been pasteurised (i.e. raw milk) is a risk factor for bacterial enteritis (Kirchner et al., 2013). Raw milk is usually consumed due to convenience, taste and perceived health

benefits. However, pasteurisation has been shown to have minimal effects on the nutritional content of milk (Claeys et al., 2013; Lacroix et al., 2006). There is also a lack of awareness of the risks regarding raw milk consumption. For example, Jayarao et al. (2006) found that most dairy farmers that consumed raw milk in Pennsylvania, United States, were unaware of its health risks, but those aware of the risks were less likely to consume it. Therefore, further work is required to inform the public of the risks of raw milk.

Humans are primarily exposed to animals as pets or via work. Many of these animals harbour bacterial enteritis agents and may transmit them to human hosts (Chaban et al., 2010). Therefore, working with these animals (e.g. farmers) or keeping them as pets (e.g. lizards) are risk factors for bacterial enteritis (Wikström et al., 2014). Strict regulations can be placed on animals, limiting their occupational exposure and preventing their use as pets. However, it is difficult to prevent all animal exposure in the agricultural and veterinarian sectors and restricted animals can be illegally purchased as pets (Walters et al., 2016). Also, few individuals are aware of the risks associated with keeping or working with these animals (Villar et al., 1998). Therefore, further work is required to inform the public of the risks associated with keeping pets or working with animals.

The immune system is a series of biological processes and structures that defends individuals from pathogens, such as bacterial enteritis agents. Immunosuppression is a reduced immune system and increases the risk of bacterial enteritis (Oksenhendler et al., 2008). Immunosuppression can be genetic (e.g. Omenn syndrome) or the result of medication (e.g. cyclosporine), malnutrition, certain cancers (e.g. leukaemia), ageing or certain infections (e.g. Human Immunodeficiency Virus (HIV)). Immunosuppression also increases the risk of bacterial enteritis infections developing into invasive diseases (Vaezirad et al., 2017). However, this risk can be minimised by informing immunosuppressed individuals regarding the risks of bacterial enteritis (Hlady and Klontz, 1996).

Vaccines can elicit immune responses that protect against infectious diseases. Multiple cholera and typhoid vaccines have been developed that are safe to use and elicit seroconversion (Valera et al., 2009; Lyon et al., 2010). However, there is still some debate regarding the level of seroconversion that is required for protection. Attempts have been made to develop vaccines to protect humans from other bacterial enteritis agents (Prendergast et al., 2004), but vaccine candidates that are effective against a broad range of strains have been difficult to identify (Meunier et al., 2016), and vaccines formed from potential candidates were unable to elicit protective immune responses in humans (Zeng et al., 2014). Therefore, further work is required to develop vaccines that elicit a protective immune response to a broad range of bacterial enteritis agents.

## 2.3 Outbreak analysis

In 1854, there was an outbreak of cholera in London, England. During this epidemic the physician, John Snow, discovered that most patients that died from cholera had been living in the vicinity of the Broad Street pump. Snow's investigation led to the removal of the pump's handle, a subsequent decrease in cholera mortality,

and the discovery that cholera was transmitted through faecal-contaminated water (McLeod, 2000). Case-control studies similar to Snow's, are still used today, where the putative cause of an outbreak is identified by comparing sources common to patients (cases) but not to healthy individuals (controls) (Young et al., 2014). However, this can be laborious, requiring selection of a control group to compare cases with, patients to be contacted after diagnosis, and cases and controls to have a good memory of all foods consumed, water sources and environmental exposures (Harker et al., 2014).

Matching a strain of bacteria found in several patients with a strain from a putative source, provides strong evidence that the patients may have obtained the infectious agent via this source (Kim et al., 2014). Therefore, this technique is often used to determine the cause of outbreaks (Hald and Pires, 2010). This process relies on tests that can distinguish the outbreak strain from closely related strains, and knowledge of the strains associated with each source.

### 2.3.1 Source strains

Bacteria can be sourced from multiple animals and environments worldwide. The presence of specific bacteria and the predominant strain present in each source can change over time (Bang et al., 2003; Haley et al., 2009). Therefore, outbreak analysis relies on active surveillance of potential sources to determine if sources contain harmful bacteria that could contribute to outbreaks, and to identify the strains of bacteria associated with each source (Magnússon et al., 2011; Raufu et al., 2013).

Bacterial enteritis outbreaks are usually only associated with a single strain of bacterium (Cummings et al., 2014; Imanishi et al., 2014). The presence of more than one strain usually represents multiple outbreaks with slightly different behaviours (Angelo et al., 2015). However, bacterial enteritis outbreaks involving more than one strain have been described (Clark et al., 2003; Newton, 2016). Regardless, multiple isolates should be collected from bacterial enteritis outbreaks and characterised to help determine if they are representative of a single or multiple bacterial strains.

### 2.3.2 Phenotypic tests

The phenotype of an organism consists of all its physical, biochemical and behavioural traits. In microbiology, phenotypic tests are used to describe and distinguish between closely related microorganisms (Debruyne et al., 2010). Phenotypic tests can help identify:

- What enzymes microorganisms contain (enzymatic tests)
- What carbon and nitrogen sources microorganisms utilise (utilisation tests)
- What surface molecules microorganisms exhibit (serological tests)

- What conditions microorganisms replicate in (growth characteristics)
- What molecules microorganisms are inhibited by (susceptibility testing)
- What phages microorganisms are susceptible to (phage typing)
- What macromolecules microorganisms consist of (lipid and protein profiles)

Multiple phenotypic tests are available to analyse and distinguish different bacterial enteritis agents. However, many of these tests are complex and/or expensive (Hazeleger et al., 1992; Howe et al., 1995; Mandrell et al., 2005). Therefore, effective phenotypic testing relies on prior knowledge of the microorganisms being tested and how they are mostly likely to differ in phenotype.

Phenotypic tests were initially used to distinguish different bacterial strains, e.g. phage typing to distinguish between *S.* Typhimurium strains (Anderson et al., 1977). However, many species and strains of bacteria have very similar phenotypes, and are unable to be distinguished from each other using phenotypic tests (Stoddard et al., 2007). Therefore, molecular techniques are growing in popularity as they are usually more discriminatory than phenotypic tests (Wang et al., 2006).

### 2.3.3 Molecular tests

The genome consists of all the genetic material that an organism contains, whilst the proteome consists of all the proteins expressed by the genome. Molecular tests are used to identify and compare sections of an organism's genome or proteome (Cooke et al., 2008), distinguish isolates based on differences in their genome or proteome (Linton et al., 1996), and determine the function of genetic elements and proteins (Lin et al., 2005).

Molecular tests can be used to distinguish microorganisms by: testing whether an organism contains a genetic element (e.g. using polymerase chain reaction (PCR)) (Magnússon et al., 2011), determining the length and location of genetic elements (e.g. using pulsed field gel electrophoresis (PFGE)) (Praakle-Amin et al., 2007), and determining the sequence of genetic elements (e.g. using multi-locus sequence typing (MLST)) (Nielsen et al., 2010). These molecular techniques vary in cost and discriminatory power. Therefore, when distinguishing strains of bacteria, the appropriate test depends on how closely related the strain of interest is to the other strains present.

Molecular tests have allowed microorganisms to be discriminated into strains that are indistinguishable by phenotypic tests (Kim et al., 2014; Young et al., 2014). However, even combinations of phenotypic and molecular techniques have failed to discriminate certain bacterial strains and identify the cause of outbreaks (Dyet et al., 2011). Therefore, multiple outbreaks are now being analysed with the highly discriminatory whole genome sequencing (Bronowski et al., 2013).

Whole genome sequencing involves sequencing the entire genome of an organism. It is discriminatory enough to distinguish between different isolates in an outbreak and provides a large amount of additional

information on the organism that can be used for further analysis, e.g. determining where a bacterial strain originated from (Eppinger et al., 2014), modelling how the bacteria evolved over the course of the outbreak (Mather et al., 2013), and identifying genetic elements specific to the causative strain (Kingsley et al., 2009). The costs and time required to perform whole genome sequencing has continued to decrease and now whole genome sequencing can be used to analyse outbreaks in real-time (Quick et al., 2015).

## 2.4 Bacterial evolution during outbreaks

Bacteria are constantly evolving through point mutations and the acquisition and loss of genetic material. During an outbreak, infected individuals, treatment and the environment place multiple selection pressures on the infectious agent that may select for specific genotypes and phenotypes (Lieberman et al., 2011). Investigating the evolution of bacteria over the course of an outbreak may provide information on these selective pressures, where the bacteria originated from and how one outbreak superseded another.

### 2.4.1 Genotype evolution

Most studies on bacterial evolution focus on changes in the agent's genotype. There are multiple molecular tests available for determining an isolates genotype, but isolates from the same outbreak often have very similar genotypes that can only be distinguished using highly discriminatory tests, e.g. whole genome sequencing (Mather et al., 2013). Changes in genotypic data can be used to make inferences on the agents effective population size (Minin et al., 2008), transmission of the agent between different populations (Vaughan et al., 2014), and when and where the outbreak originated from (Lemey et al., 2009). However, these genotypic tests often produce a large amount of information (e.g. sequence data) that must be sifted through before these inferences can be made.

### 2.4.2 Phenotype evolution

Bacterial phenotypes evolve/change over time. Phenotypic tests can be used to monitor how an agents phenotype changes over the course of an outbreak (Lieberman et al., 2011). They can be used to identify selective pressures imposed by the environment, hosts and/or treatment in an outbreak (Betancor et al., 2009), and if there are any associations between different phenotypes and groups of isolates (e.g. isolates collected from specific sources) (Osek, 2004). Most studies on phenotypic evolution only utilise antimicrobial susceptibility tests (Hocquet et al., 2003), as there are a large number of potential ways bacteria may differ in phenotype, isolates are more likely to differ in antibiotic susceptibilities than other phenotypes (Oh et al., 2003), and antimicrobial susceptibility tests are more straightforward (Turner et al., 2013). Automatic

analysers have been developed that can perform multiple phenotypic tests on bacteria (e.g. OmniLog) or can perform highly sensitive phenotypic tests (e.g. MALDI-TOF protein analysis) (Christner et al., 2014; Fox and Jordan, 2014). These analysers have made it easier to distinguish isolates based on phenotype and may promote further research into phenotypic evolution of outbreaks.

### 2.4.3 Related outbreaks

Bacteria can be transported between geographical locations on vectors or fomites. This allows outbreaks to spread from one geographical area to another. If a strain of bacteria causes two outbreaks in close succession, then it is often assumed that one outbreak led to the other. However, the same strain of bacteria may cause multiple outbreaks in different geographies at almost the same time (Stine et al., 2008). In these circumstances, it is difficult to determine the relationship between the outbreaks. Molecular and phenotypic tests can be used to determine the relationship of bacteria collected from separate outbreaks. This in turn can be used to determine how bacteria spread from one geographical area to another, and help identify potential vectors or fomites (Eppinger et al., 2014).

### 2.4.4 Date of common ancestor

All organisms share a common ancestor (Darwin, 1860). Identifying the date of the most recent common ancestor for isolates in an outbreak can help identify when the outbreak strain was introduced into an area. In turn, identifying key events that occurred at the time of the most recent ancestor can help identify potential ways in which the outbreak strain was introduced (Eppinger et al., 2014).

The date of common ancestor is usually estimated using a molecular clock. Molecular clocks utilise the mutation rates of molecular or phenotypic data to time when groups of isolates shared a common ancestor (coalescent events) (Sarich and Wilson, 1967). Strict molecular clocks assume that the mutation rate is constant, whilst relaxed molecular clocks assume the mutation rate can vary between lineages. Relaxed molecular clocks often give more accurate estimations of phylogenetic relationships and rates, than strict clocks, as they can accommodate both constant and variable mutation rates. However, this accommodation often results in a large amount of uncertainty and possible tree shapes for specific lineages within the tree (Drummond et al., 2006).

The mutation rate is the rate at which a set of molecular or phenotypic data mutates over time. Mutation rates can be calculated from genetic or phenotypic data if isolates were collected at different time points (Mather et al., 2013). In the absence of data from different time points, mutation rates are acquired from previous studies (Sheppard et al., 2010). However, mutation rates vary with different data sets and organisms (Nobusawa and Sato, 2006). Therefore, appropriate mutation rates can be difficult to find.

### 2.4.5 Bacterial transmission

The success with which bacteria may be transmitted between two individuals is dependent on the strain of bacteria, the health of the individuals, and how the agent is transmitted (Lawley et al., 2008). Outbreaks can be prevented or their size diminished, by identifying how bacteria are primarily transmitted and putting in policies to prevent transmission (Sears et al., 2011). Bacterial transmission can be investigated at the individual or population level.

Dense sampling of bacteria over the course of local outbreaks has been used to determine the probability that an agent was directly transmitted between individuals (Croucher and Didelot, 2015). Dependent on the agent and type of infection, individuals may contain a diverse population of the bacteria that can lead to inaccurate transmission predictions if a single isolate is taken from each host (Worby et al., 2014). Accurate transmission estimates often rely on analysing multiple isolates from each individual to estimate the diversity of the agent population within hosts (Jiang et al., 2015). Sequencing multiple isolates from an infividual is more expensive than sequencing a single isolate, but can provide a large amount of additional information on the bacterial agent, how it evolves within the host and the nature of the disease (Okoro et al., 2012).

Predicting transmission at the population-wide level relies on a different approach to investigating transmission at the individual level, as it is often logistically and financially problematic to obtain isolates from all individuals in a population. Instead a sample of isolates is analysed from each population and transmission between populations is investigated (Mather et al., 2013).

Source attribution models estimate the proportion of cases attributed to sources. They are primarily used on infectious diseases involving multiple strains, as opposed to outbreaks of a single strain (Mullner et al., 2009), but they can be used to identify likely sources for outbreaks. The Hald (Hald et al., 2004) and Dutch (Van Pelt et al., 1999) models subtype the microorganism, and attribute cases to sources based on the frequency of subtypes in the source compared to the affected population. Models, such as the asymmetric island model (Wilson et al., 2008) go a step further taking into account the evolutionary relationship of the subtypes to obtain population genetic measurements that are compared between the sources and affected populations (Mather et al., 2015). A problem with these models is that they assume transmission is unidirectional from the source to the affected population, and do not consider transmission between the sources.

Ancestral state reconstruction models were designed to estimate the ancestral state of organisms based on their evolutionary history. Ancestral state reconstruction models, such as the discrete trait analysis (Lemey et al., 2009) and structured coalescent models (Vaughan et al., 2014) have been used to accurately predict infectious disease transmission between distinct geographies, over the course of outbreaks. The discrete trait analysis model has also been used to predict the amount of transmission between different sources over the course of outbreaks (Mather et al., 2013). Unlike other source attribution models, these models allow for bidirectional transmission. However, they make many population and transmission-related assumptions that may not apply to outbreaks with sources occupying similar geographies. Therefore, further work is required

to determine if these models can be used for outbreak transmission studies.

#### 2.4.6 Effective population size

The effective population size is the number of individuals in a population that contribute to the next generation. Effective population size estimates can help identify when the population of an organism changed and help identify factors responsible for population change (Glaser et al., 2016).

The effective population size affects the timing of coalescent events for randomly sampled individuals. For example, if two individuals were randomly sampled from a large population then I would expect that they have a more distant ancestor than if they were randomly sampled from a smaller population. Models such as the Extended Bayesian Skyline Plot (Heled and Drummond, 2008) and the Gaussian Markov Random Field (GMRF) Bayesian Skyride model (Minin et al., 2008), estimate coalescent events over time and predict the effective population size for a population based on the timing of these events. However, changes in the effective population size can be difficult to estimate, relying on a large amount of data to accurately time coalescent events.

#### 2.4.7 Genetic elements

Mobile genetic elements are pieces of DNA that can move within an organisms genome and/or between genomes from different organisms, e.g. transposons, plasmids and viruses (phages). They can change a bacteria's characteristics by the addition and removal of genes (Nedialkova et al., 2016). Mobile genetic elements from various bacteria can be analysed and compared using molecular techniques. This can help: identify where mobile genetic elements originated from, when the mobile genetic element was obtained, how the mobile genetic elements have evolved over time, and what strains of bacteria interact with each other, either directly or via phages (Yamamoto and Taneike, 2000). However, mobile genetic element analysis can be difficult to perform, as they rely on an extensive database of genomes from different organisms to identify these mobile genetic elements and where they originated from.

## 2.5 Conclusion

Bacterial enteritis is a clinically important disease that affects a large number of humans worldwide. Multiple agents cause bacterial enteritis and each is associated with different sources and behaviours. There sources can be subdivided into reservoirs, disease pathways, disease exposures and risk factors for control purposes. Multiple control measures have been designed to prevent bacterial enteritis. However, they vary in price, environmental and clinical sustainability, and effectiveness.

Outbreaks of bacterial enteritis are common worldwide. They are challenging to investigate as a large number of potential sources must often be considered; highly discriminatory molecular and phenotypic tests are often required to distinguish between isolates; and these tests often produce a large amount of information that must be sifted through. However, bacterial enteritis outbreak analyses may provide insight into the agent, how the agent evolves, transmission of the agent, what individuals are at risk, and how agents from different outbreaks are related. This information can then be used to inform prevention strategies and decrease the incidence of bacterial enteritis.

# References


Abbot, J. M., Policastro, P., Bruhn, C., Schaffner, D. W., and Byrd-Bredbenner, C. (2012). Development and evaluation of a university campus-based food safety media campaign for young adults. *Journal of Food Protection*, 75(6):1117–1124.

Ahmadova, A., Todorov, S. D., Hadji-Sfaxi, I., Choiset, Y., Rabesona, H., Messaoudi, S., Kuliyev, A., de Melo Franco, B., Chobert, J.-M., and Haertlé, T. (2013). Antimicrobial and antifungal activities of *Lactobacillus curvatus* strain isolated from homemade Azerbaijani cheese. *Anaerobe*, 20:42–49.

Anderson, E. S., Ward, L. R., de Saxe, M. J., and de Sa, J. D. H. (1977). Bacteriophage-typing designations of *Salmonella typhimurium*. *Journal of Hygiene*, 78(2):297–300.

Anderson, M. E. C., Sargeant, J. M., and Weese, J. S. (2014). Video observation of hand hygiene practices during routine companion animal appointments and the effect of a poster intervention on hand hygiene compliance. *BMC Veterinary Research*, 10:1–16.

Angelo, K. M., Chu, A., Anand, M., Nguyen, T., Bottichio, L., Wise, M., Williams, I., Seelman, S., Bell, R., Fatica, M., Lance, S., Baldwin, D., Shannon, K., Lee, H., Trees, E., Strain, E., and Gieraltowski, L. (2015). Outbreak of *Salmonella* Newport infections linked to cucumbers United States, 2014. *Morbidity and Mortality Weekly Report*, 64(6):144–147.

Atterbury, R. J., Van Bergen, M. A. P., Ortiz, F., Lovell, M. A., Harris, J. A., De Boer, A., Wagenaar, J. A., Allen, V. M., and Barrow, P. A. (2007). Bacteriophage therapy to reduce *Salmonella* colonization of broiler chickens. *Applied and Environmental Microbiology*, 73(14):4543–4549.

Babińska, I., Rotkiewicz, T., and Otrocka-Domagała, I. (2005). The effect of *Lactobacillus acidophilus* and *Bifidobacterium* spp. administration on the morphology of the gastrointestinal tract, liver and pancreas in piglets. *Polish Journal of Veterinary Sciences*, 8(1):29–35.

Bang, D. D., Nielsen, E. M., Knudsen, K., and Madsen, M. (2003). A one-year study of campylobacter carriage by individual Danish broiler chickens as the basis for selection of *Campylobacter* spp. strains for a chicken infection model. *Epidemiology and Infection*, 130(2):323–333.

Bayer, C., Bernard, H., Prager, R., Rabsch, W., Hiller, P., Malorny, B., Pfefferkorn, B., Frank, C., De Jong, A., Friesema, I., Stark, K., and Rosner, B. M. (2014). An outbreak of *Salmonella* Newport associated with mung bean sprouts in Germany and the Netherlands, October to November 2011. *Eurosurveillance*, 19(1):1–9.

Belongia, E. A., Chyou, P. H., Greenlee, R. T., Perez-Perez, G., Bibb, W. F., and DeVries, E. O. (2003). Diarrhea incidence and farm-related risk factors for *Escherichia coli* O157:H7 and *Campylobacter jejuni* antibodies among rural children. *Journal of Infectious Diseases*, 187(9):1460–1468.

Betancor, L., Yim, L., Fookes, M., Martinez, A., Thomson, N. R., Ivens, A., Peters, S., Bryant, C., Algorta, G., Kariuki, S., Schelotto, F., Maskell, D., Dougan, G., and Chabalgoity, J. A. (2009). Genomic and phenotypic variation in epidemic-spanning *Salmonella enterica* serovar Enteritidis isolates. *BMC Microbiology*, 9:1–16.

Bhavsar, S. P., Augustine, S. K., and Kapadnis, B. P. (2007). Effect of physical and chemical treatments on *Campylobacter* spiked into food samples. *Food Science and Technology International*, 13(4):277–283.

Bolder, N. M., Wagenaar, J. A., Putirulan, F. F., Veldman, K. T., and Sommer, M. (1999). The effect of flavophospholipol (Flavomycin®) and salinomycin sodium (Sacox®) on the excretion of *Clostridium perfringens*, *Salmonella enteritidis*, and *Campylobacter jejuni* in broilers after experimental infection. *Poultry Science*, 78(12):1681–1689.

Borie, C., Albala, I., Sànchez, P., Sánchez, M. L., Ramírez, S., Navarro, C., Morales, M. A., Retamales, J., and Robeson, J. (2008). Bacteriophage treatment reduces *Salmonella* colonization of infected chickens. *Avian Diseases*, 52(1):64–67.

Brenner, D. M. and Chey, W. D. (2009). *Bifidobacterium infantis* 35624: A novel probiotic for the treatment of irritable bowel syndrome. *Reviews in Gastroenterological Disorders*, 9(1):7–15.

Bronowski, C., Fookes, M. C., Gilderthorp, R., Ashelford, K. E., Harris, S. R., Phiri, A., Hall, N., Gordon, M. A., Wain, J., Hart, C. A., Wigley, P., Thomson, N. R., and Winstanley, C. (2013). Genomic characterisation of invasive non-typhoidal *Salmonella enterica* subspecies *enterica* serovar Bovismorbificans isolates from Malawi. *PLoS Neglected Tropical Diseases*, 7(11):1–12.

Byrne, L., Fisher, I., Peters, T., Mather, A., Thomson, N., Rosner, B., Bernard, H., McKeown, P., Cormican, M., Cowden, J., Aiyedun, V., Lane, C., and Team, I. O. C. (2014). A multi-country outbreak of *Salmonella* Newport gastroenteritis in Europe associated with watermelon from Brazil, confirmed by whole genome sequencing: October 2011 to January 2012. *Euro surveillance : bulletin Européen sur les maladies transmissibles = European communicable disease bulletin*, 19(31):6–13.

CDC (2015). Antibiotic treatment: Recommendations for the use of antibiotics for the treatment of cholera. Retrieved 2017-06-29, from: http://www.cdc.gov/cholera/treatment/antibiotic-treatment.html, 1-1.

Chaban, B., Ngeleka, M., and Hill, J. E. (2010). Detection and quantification of 14 *Campylobacter* species in pet dogs reveals an increase in species richness in feces of diarrheic animals. *BMC Microbiology*, 10:1–7.

Christner, M., Trusch, M., Rohde, H., Kwiatkowski, M., Schlüter, H., Wolters, M., Aepfelbacher, M., and Hentschke, M. (2014). Rapid MALDI-TOF mass spectrometry strain typing during a large outbreak of Shiga-toxigenic *Escherichia coli*. *PLoS ONE*, 9(7):1–11.

Claeys, W. L., Cardoen, S., Daube, G., De Block, J., Dewettinck, K., Dierick, K., De Zutter, L., Huyghebaert, A., Imberechts, H., Thiange, P., Vandenplas, Y., and Herman, L. (2013). Raw or heated cow milk consumption: Review of risks and benefits. *Food Control*, 31(1):251–262.

Clark, C. G., Price, L., Ahmed, R., Woodward, D. L., Melito, P. L., Rodgers, F. G., Jamieson, F., Ciebin, B., Li, A., and Ellis, A. (2003). Characterization of waterborne outbreak-associated *Campylobacter jejuni*, Walkerton, Ontario. *Emerging Infectious Diseases*, 9(10):1232–1241.

Cooke, F. J., Brown, D. J., Fookes, M., Pickard, D., Ivens, A., Wain, J., Roberts, M., Kingsley, R. A., Thomson, N. R., and Dougan, G. (2008). Characterization of the genomes of a diverse collection of *Salmonella enterica* serovar Typhimurium definitive phage type 104. *Journal of Bacteriology*, 190(24):8155–8162.

Cossi, M. V. C., Burin, R. C. K., Camargo, A. C., Dias, M. R., Lanna, F., Pinto, P., and Nero, L. A. (2014). Low occurrence of *Salmonella* in the beef processing chain from Minas Gerais state, Brazil: From bovine hides to end cuts. *Food Control*, 40(1):320–323.

CRC (2013). Contaminated land. Retrieved 2017-06-29, from: http://ecan.govt.nz/publications/Plans/canterbury-regional-policy-statement.pdf, 177-182.

Croucher, N. and Didelot, X. (2015). The application of genomics to tracing bacterial pathogen transmission. *Current Opinion in Microbiology*, 23:62–67. cited By 21.

Crovato, S., Pinto, A., Giardullo, P., Mascarello, G., Neresini, F., and Ravarotto, L. (2016). Food safety and young consumers: Testing a serious game as a risk communication tool. *Food Control*, 62:134–141.

Cummings, K. J., Rodriguez-Rivera, L. D., Mitchell, K. J., Hoelzer, K., Wiedmann, M., McDonough, P. L., Altier, C., Warnick, L. D., and Perkins, G. A. (2014). *Salmonella enterica* serovar oranienburg outbreak in a veterinary medical teaching hospital with evidence of nosocomial and on-farm transmission. *Vector-Borne and Zoonotic Diseases*, 14(7):496–502.

Darwin, C. (1860). *On the origin of species by natural selection: Or the preservation of the favoured races in the struggle for life, pages 1–490.*

David, L. A., Materna, A. C., Friedman, J., Campos-Baptista, M. I., Blackburn, M. C., Perrotta, A., Erdman, S. E., and Alm, E. J. (2015). Host lifestyle affects human microbiota on daily timescales. *Genome Biology*, 15(7):1–15.

De Maio, N., Wu, C., O'Reilly, K., and Wilson, D. (2015). New routes to phylogeography: A Bayesian structured coalescent approximation. *PLoS Genetics*, 11(8):1–22.

De Milliano, I., Tabbers, M. M., Van Der Post, J. A., and Benninga, M. A. (2012). Is a multispecies probiotic mixture effective in constipation during pregnancy? 'A pilot study'. *Nutrition Journal*, 11(1):1–6.

De Perio, M. A., Niemeier, R. T., Levine, S. J., Gruszynski, K., and Gibbins, J. D. (2013). *Campylobacter* infection in poultry-processing workers, Virginia, USA, 2008-2011. *Emerging Infectious Diseases*, 19(2):286–288.

Debruyne, L., Broman, T., Bergström, S., Olsen, B., On, S. L. W., and Vandamme, P. (2010). *Campylobacter volucris* sp. nov., isolated from black-headed gulls (*Larus ridibundus*). *International Journal of Systematic and Evolutionary Microbiology*, 60(8):1870–1875.

Dekker, D. M., Krumkamp, R., Sarpong, N., Frickmann, H., Boahen, K. G., Frimpong, M., Asare, R., Larbi, R., Hagen, R. M., Poppert, S., Rabsch, W., Marks, F., Sarkodie, Y. A., and May, J. (2015). Drinking water from dug wells in rural Ghana *Salmonella* contamination, environmental factors, and genotypes. *International Journal of Environmental Research and Public Health*, 12(4):3535–3546.

Dharod, J. M., Pérez-Escamilla, R., Bermúdez-Millán, A., Segura-Pérez, S., and Damio, G. (2004). Influence of the Fight BAC! Food safety campaign on an urban Latino population in Connecticut. *Journal of Nutrition Education and Behavior*, 36(3):128–134.

Donnison, A., Ross, C., and Thorrold, B. (2004). Impact of land use on the faecal microbial quality of hill-country streams. *New Zealand Journal of Marine and Freshwater Research*, 38(5):845–855.

Doorduyn, Y., Van Den Brandhof, W. E., Van Duynhoven, Y., Wannet, W. J. B., and Van Pelt, W. (2006). Risk factors for *Salmonella* Enteritidis and Typhimurium (DT104 and non-DT104) infections in The Netherlands: Predominant roles for raw eggs in Enteritidis and sandboxes in Typhimurium infections. *Epidemiology and Infection*, 134(3):617–626.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5):699–710.

Duffitt, A. D., Reber, R. T., Whipple, A., and Chauret, C. (2011). Gene expression during survival of *Escherichia coli* O157:H7 in soil and water. *International Journal of Microbiology*, pages 1–12.

Dyet, K. H., Turbitt, E., and Carter, P. E. (2011). Multiple-locus variable-number tandem-repeat analysis for discriminating within *Salmonella enterica* serovar Typhimurium definitive types and investigation of outbreaks. *Epidemiology and Infection*, 139(7):1050–1059.

Ekhtiarzadeh, H., Akhondzadeh Basti, A., Misaghi, A., Sari, A., Khanjari, A., Rokni, N., Abbaszadeh, S., and Partovi, R. (2012). Growth response of *Vibrio parahaemolyticus* and *Listeria monocytogenes* in salted fish fillets as affected by zataria multiflora boiss. essential oil, nisin, and their combination. *Journal of Food Safety*, 32(3):263–269.

El-Shibiny, A., Scott, A., Timms, A., Metawea, Y., Connerton, P., and Connerton, I. (2009). Application of a group II *Campylobacter* bacteriophage to reduce strains of *Campylobacter jejuni* and *Campylobacter coli* colonizing broiler chickens. *Journal of Food Protection*, 72(4):733–740.

Ellström, P., Hansson, I., Söderström, C., Engvall, E. O., and Rautelin, H. (2014). A prospective follow-up study on transmission of *Campylobacter* from poultry to abattoir workers. *Foodborne Pathogens and Disease*, 11(9):684–688.

Eppinger, M., Pearson, T., Koenig, S. S. K., Pearson, O., Hicks, N., Agrawal, S., Sanjar, F., Galens, K., Daugherty, S., Crabtree, J., Hendriksen, R. S., Price, L. B., Upadhyay, B. P., Shakya, G., Fraser, C. M., Ravel, J., and Keim, P. S. (2014). Genomic epidemiology of the Haitian cholera outbreak: A single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. *mBio*, 5(6):1–8.

Er, B., Demirhan, B., Onurda, F. K., Özgacar, S. Ö., and Öktem, A. B. (2014). Antimicrobial and antibiofilm effects of selected food preservatives against *Salmonella* spp. isolated from chicken samples. *Poultry Science*, 93(3):695–701.

ESR (2015). Notifiable diseases. Retrieved 2017-03-02, from: `http://www.nzpho.org.nz/NotifiableDisease.aspx`, 1–1.

Feglo, P. K., Frimpong, E. H., and Essel-Ahun, M. (2004). Salmonellae carrier status of food vendors in Kumasi, Ghana. *East African Medical Journal*, 81(7):358–361.

Ferraro, V., Cruz, I. B., Ferreira Jorge, R., Pintado, M. E., and Castro, P. M. L. (2014). Kinetics of release of water and nutrients from codfish (*Gadus Morhua L.*) through a heavy-salting. *Journal of Food Processing and Preservation*, 38(4):1772–1778.

Fitzgerald, C. and Nachamkin, I. (2007). *Campylobacter* and *Arcobacter*. In Murray, P. R., Baron, E. J., Jorgensen, M. L., Landry, M. L., and Pfaller, M. A., editors, *Manual of Clinical Microbiology*, chapter 59, pages 933–946. Washington DC, 9th edition.

Fox, E. M. and Jordan, K. (2014). High-throughput characterization of *Listeria monocytogenes* using the OmniLog phenotypic microarray. *Methods in Molecular Biology*, 1157:103–108.

French, N. P., Midwinter, A., Holland, B., Collins-Emerson, J., Pattison, R., Colles, F., and Carter, P. (2009). Molecular epidemiology of *Campylobacter jejuni* isolates from wild-bird fecal material in children's playgrounds. *Applied and Environmental Microbiology*, 75(3):779–783.

Ghareeb, K., Awad, W. A., Mohnl, M., Porta, R., Biarnés, M., Böhm, J., and Schatzmayr, G. (2012). Evaluating the efficacy of an avian-specific probiotic to reduce the colonization of *Campylobacter jejuni* in broiler chickens. *Poultry Science*, 91(8):1825–1832.

Glaser, L., Carstensen, M., Shaw, S., Robbe-Austerman, S., Wunschmann, A., Grear, D., Stuber, T., and Thomsen, B. (2016). Descriptive epidemiology and whole genome sequencing analysis for an outbreak of bovine tuberculosis in beef cattle and white-tailed deer in northwestern Minnesota. *PLoS ONE*, 11(1):1–21.

Gould, L. H., Walsh, K. A., Vieira, A. R., Herman, K., Williams, I. T., Hall, A. J., and Cole, D. (2013). Surveillance for foodborne disease outbreaks - United States, 1998-2008. *MMWR Surveillance Summaries*, 62(1):1–34.

GWRC (2016). Recreastional water quality. Retrieved 2017-06-29, from: `http://mapping.gw.govt.nz/GW/RecWaterQualityMap/RecWaterQualityMap.htm`, 1-1.

Hald, T. and Pires, S. M. (2010). Attributing the burden of foodborne disease to specific sources of infection. In Brul, S., Fratamico, F. P., and McMeekin, T. A., editors, *Tracing pathogens in the food chain*, chapter 5, pages 89–113. Woodhead Publishing, United Kingdom.

Hald, T., Vose, D., Wegener, H. C., and Koupeev, T. (2004). A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. *Risk Analysis*, 24(1):255–269.

Haley, B. J., Cole, D. J., and Lipp, E. K. (2009). Distribution, diversity, and seasonality of waterborne salmonellae in a rural watershed. *Applied and Environmental Microbiology*, 75(5):1248–1255.

Hara-Kudo, Y. and Takatori, K. (2011). Contamination level and ingestion dose of foodborne pathogens associated with infections. *Epidemiology and Infection*, 139(10):1505–1510. cited By 22.

Harker, K. S., Lane, C., Gormley, F. J., and Adak, G. K. (2014). National outbreaks of *Salmonella* infection in the UK, 2000-2011. *Epidemiology and Infection*, 142(3):601–607.

Harrison, D., Corry, J. E. L., Tchórzewska, M. A., Morris, V. K., and Hutchison, M. L. (2013). Freezing as an intervention to reduce the numbers of campylobacters isolated from chicken livers. *Letters in Applied Microbiology*, 57(3):206–213.

Haydon, D. T., Cleaveland, S., Taylor, L. H., and Laurenson, M. K. (2002). Identifying reservoirs of infection: A conceptual and practical challenge. *Emerging Infectious Diseases*, 8(12):1468–1473.

Hazeleger, W. C., Beumer, R. R., and Rombouts, F. M. (1992). The use of latex agglutination tests for determining *Campylobacter* species. *Letters in Applied Microbiology*, 14(4):181–184.

HBRC (2016). Recreational water quality map. Retrieved 2017-06-29, from: `http://www.hbrc.govt.nz/hawkes-bay/swimming/water-quality/`, 1–1.

He, B., Xu, W., Santini, P. A., Polydorides, A. D., Chiu, A., Estrella, J., Shan, M., Chadburn, A., Villanacci, V., Plebani, A., Knowles, D. M., Rescigno, M., and Cerutti, A. (2007). Intestinal bacteria trigger T cell-independent immunoglobulin A2 class switching by inducing epithelial-cell secretion of the cytokine APRIL. *Immunity*, 26(6):812–826.

Heled, J. and Drummond, A. J. (2008). Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, 8(1):1–15.

Higgins, S. E., Higgins, J. P., Wolfenden, A. D., Henderson, S. N., Torres-Rodriguez, A., Tellez, G., and Hargis, B. (2008). Evaluation of a *Lactobacillus*-based probiotic culture for the reduction of *Salmonella* Enteritidis in neonatal broiler chicks. *Poultry Science*, 87(1):27–31.

Hines, J. Z., Pinsent, T., Rees, K., Vines, J., Bowen, A., Hurd, J., Leman, R. F., and Hedberg, K. (2016). Shigellosis outbreak among men who have sex with men and homeless persons Oregon, 20152016. *Morbidity and Mortality Weekly Report*, 65(31):812–813.

Hlady, W. G. and Klontz, K. C. (1996). The epidemiology of *Vibrio* infections in Florida, 1981-1993. *Journal of Infectious Diseases*, 173(5):1176–1183.

Hocquet, D., Bertrand, X., Köhler, T., Talon, D., and Plésiat, P. (2003). Genetic and phenotypic variations of a resistant *Pseudomonas aeruginosa* epidemic clone. *Antimicrobial Agents and Chemotherapy*, 47(6):1887–1894.

Horn, K. and Otter, J. A. (2015). Hydrogen peroxide vapor room disinfection and hand hygiene improvements reduce *Clostridium difficile* infection, methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant *Enterococci*, and extended-spectrum $\beta$-lactamase. *American Journal of Infection Control*, 43(12):1354–1356.

House, J. K., Ontiveros, M. M., Blackmer, N. M., Dueger, E. L., Fitchhorn, J. B., McArthur, G. R., and Smith, B. P. (2001). Evaluation of an autogenous *Salmonella* bacterin and a modified live *Salmonella* serotype Choleraesuis vaccine on a commercial dairy farm. *American Journal of Veterinary Research*, 62(12):1897–1902.

Howe, R. A., Clarke, T., Wilcox, M. H., Vandamme, P., and Spencer, R. C. (1995). *Campylobacter fetus* subspecies *fetus* septicaemia: SDS-PAGE as an aid to speciation. *Journal of Infection*, 31(3):229–232.

Howie, H., Mukerjee, A., Cowden, J., Leith, J., and Reid, T. (2003). Investigation of an outbreak of *Escherichia coli* O157 infection caused by environmental exposure at a scout camp. *Epidemiology and Infection*, 131(3):1063–1069.

Ibekwe, A. M., Grieve, C. M., and Yang, C. H. (2007). Survival of *Escherichia coli* O157:H7 in soil and on lettuce after soil fumigation. *Canadian Journal of Microbiology*, 53(5):623–635.

Ibekwe, A. M., Papiernik, S. K., Grieve, C. M., and Yang, C. H. (2010). Influence of fumigants on soil microbial diversity and survival of *E. coli* O157:H7. *Journal of Environmental Science and Health - Part B Pesticides, Food Contaminants, and Agricultural Wastes*, 45(5):416–426.

Imanishi, M., Rotstein, D. S., Reimschuessel, R., Schwensohn, C. A., Woody, D. H., Davis, S. W., Hunt, A. D., Arends, K. D., Achen, M., Cui, J., Zhang, Y., Denny, L. F., Phan, Q. N., Joseph, L. A., Tuite, C. C., Tataryn, J. R., and Behravesh, C. B. (2014). Public veterinary medicine: Public health outbreak of *Salmonella enterica* serotype Infantis infection in humans linked to dry dog food in the United States and Canada, 2012. *Journal of the American Veterinary Medical Association*, 244(5):545–553.

Islam, M., Morgan, J., Doyle, M. P., Phatak, S. C., Millner, P., and Jiang, X. (2004). Fate of *Salmonella enterica* serovar Typhimurium on carrots and radishes grown in fields treated with contaminated manure composts or irrigation water. *Applied and Environmental Microbiology*, 70(4):2497–2502.

Jayarao, B. M., Donaldson, S. C., Straley, B. A., Sawant, A. A., Hegde, N. V., and Brown, J. L. (2006). A survey of foodborne pathogens in bulk tank milk and raw milk consumption among farm families in Pennsylvania. *Journal of Dairy Science*, 89(7):2451–2458.

Jiang, Y., Wei, Z., Wang, Y., Hua, X., Feng, Y., and Yu, Y. (2015). Tracking a hospital outbreak of kpc-producing st11 klebsiella pneumoniae with whole genome sequencing. *Clinical Microbiology and Infection*, 21(11):1001–1007. cited By 6.

Jol, S., Kassianenko, A., Oggel, J., and Wszol, K. (2006). A country-by-country look at regulations and best practices in the global cold chain. *Food Safety magazine*, pages 1–1.

Jones, C. D. R., Hunter, C., Williams, A. P., Strachan, N. J. C., and Cross, P. (2011). *Escherichia coli* O157: Comparing awareness of rural residents and visitors in livestock farming areas. *Epidemiology and Infection*, 139(10):1522–1530.

Karp, D. S., Gennet, S., Kilonzo, C., Partyka, M., Chaumont, N., Atwill, E. R., and Kremen, C. (2015). Comanaging fresh produce for nature conservation and food safety. *Proceedings of the National Academy of Sciences of the United States of America*, 112(35):11126–11131.

Kim, J., Hyeon, J.-Y., Lee, E., Lee, D., Kim, Y.-J., Kim, Y.-J., and Kim, S. (2014). Molecular epidemiological analysis of five outbreaks associated with *Salmonella enterica* serovar Enteritidis between 2008 and 2010 on Jeju Island, Republic of Korea. *Foodborne Pathogens and Disease*, 11(1):38–42.

Kingsley, R. A., Msefula, C. L., Thomson, N. R., Kariuki, S., Holt, K. E., Gordon, M. A., Harris, D., Clarke, L., Whitehead, S., Sangal, V., Marsh, K., Achtman, M., Molyneux, M. E., Cormican, M., Parkhill, J.,

MacLennan, C. A., Heyderman, R. S., and Dougan, G. (2009). Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Research*, 19(12):2279–2287.

Kirchner, M., Dildei, C., Runge, M., Brix, A., Claussen, K., Weiss, U., Fruth, A., Prager, R., Mellmann, A., Beutin, L., Miko, A., Wichmann-Schauer, H., Pulz, M., and Dreesman, J. (2013). Outbreak of non-sorbitol-fermenting shiga toxin-producing *E. Coli* 0157:H7 infections among school children associated with raw milk consumption in Germany. *Archiv fur Lebensmittelhygiene*, 64(3):68–74.

Klumb, C., Saunders, S., and Smith, K. (2013). *E. coli* O157:H7 surveillance in agricultural populations in Minnesota. In *North American Agricultural Safety Summit*, pages 221–221.

Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., Sansonetti, P. J., Adak, G. K., and Levine, M. M. (1999). Global burden of *Shigella* infections: Implications for vaccine development and implementation of control strategies. *Bulletin of the World Health Organization*, 77(8):651–666.

Kudra, L., Sebranek, J., Dickson, J., Mendonca, A., Zhang, Q., Jackson-Davis, A., and Prusa, K. (2011a). Control of *Salmonella enterica* typhimurium in chicken breast meat by irradiation combined with modified atmosphere packaging. *Journal of Food Protection*, 74(11):1833–1839. cited By 2.

Kudra, L. L., Sebranek, J. G., Dickson, J. S., Mendonca, A. F., Larson, E. M., Jackson-Davis, A. L., and Lu, Z. (2011b). Effects of vacuum or modified atmosphere packaging in combination with irradiation for control of *Escherichia coli* O157:H7 in ground beef patties. *Journal of Food Protection*, 74(12):2018–2023.

Lacroix, M., Léonil, J., Bos, C., Henry, G., Airinei, G., Fauquant, J., Tomé, D., and Gaudichon, C. (2006). Heat markers and quality indexes of industrially heat-treated [ 15N] milk protein measured in rats. *Journal of Agricultural and Food Chemistry*, 54(4):1508–1517.

Lahti, L., Salonen, A., Kekkonen, R. A., Salojärvi, J., Jalanka-Tuovinen, J., Palva, A., Orešič, M., and de Vos, W. M. (2013). Associations between the human intestinal microbiota, *Lactobacillus rhamnosus* GG and serum lipids indicated by integrated analysis of high-throughput profiling data. *PeerJ*, 2013(1):1–25.

Lawley, T. D., Bouley, D. M., Hoy, Y. E., Gerke, C., Relman, D. A., and Monack, D. M. (2008). Host transmission of *Salmonella enterica* serovar Typhimurium is controlled by virulence factors and indigenous intestinal microbiota. *Infection and Immunity*, 76(1):403–416.

Lee, G., Pan, W., Peñataro Yori, P., Paredes Olortegui, M., Tilley, D., Gregory, M., Oberhelman, R., Burga, R., Chavez, C. B., and Kosek, M. (2013). Symptomatic and asymptomatic *Campylobacter* infections associated with reduced growth in Peruvian children. *PLoS Neglected Tropical Diseases*, 7(1):1–9.

Lemey, P., Rambaut, A., Drummond, A., and Suchard, M. (2009). Bayesian phylogeography finds its roots. *PLoS Computational Biology*, 5(9):1–16.

Lieberman, T., Michel, J.-B., Aingaran, M., Potter-Bynoe, G., Roux, D., Davis Jr., M., Skurnik, D., Leiby, N., Lipuma, J., Goldberg, J., McAdam, A., Priebe, G., and Kishony, R. (2011). Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nature Genetics*, 43(12):1275–1280.

Lin, J., Akiba, M., Sahin, O., and Zhang, Q. (2005). CmeR functions as a transcriptional repressor for the multidrug efflux pump CmeABC in *Campylobacter jejuni*. *Antimicrobial Agents and Chemotherapy*, 49(3):1067–1075.

Linton, D., Owen, R. J., and Stanley, J. (1996). Rapid identification by PCR of the genus *Campylobacter* and of five *Campylobacter* species enteropathogenic for man and animals. *Research in Microbiology*, 147(9):707–718.

Loc Carrillo, C., Atterbury, R. J., El-Shibiny, A., Connerton, P. L., Dillon, E., Scott, A., and Connerton, I. F. (2005). Bacteriophage therapy to reduce *Campylobacter jejuni* colonization of broiler chickens. *Applied and Environmental Microbiology*, 71(11):6554–6563.

Lynch, M. F., Blanton, E. M., Bulens, S., Polyak, C., Vojdani, J., Stevenson, J., Medalla, F., Barzilay, E., Joyce, K., Barrett, T., and Mintz, E. D. (2009). Typhoid fever in the United States, 1999-2006. *JAMA - Journal of the American Medical Association*, 302(8):859–865.

Lyon, C. E., Sadigh, K. S., Carmolli, M. P., Harro, C., Sheldon, E., Lindow, J. C., Larsson, C. J., Martinez, T., Feller, A., Ventrone, C. H., Sack, D. A., DeNearing, B., Fingar, A., Pierce, K., Dill, E. A., Schwartz, H. I., Beardsworth, E. E., Kilonzo, B., May, J. P., Lam, W., Upton, A., Budhram, R., and Kirkpatrick, B. D. (2010). In a randomized, double-blinded, placebo-controlled trial, the single oral dose typhoid vaccine, M01ZH09, is safe and immunogenic at doses up to 1.7  1010 colony-forming units. *Vaccine*, 28(20):3602–3608.

Ma, J., Mark Ibekwe, A., Crowley, D. E., and Yang, C.-H. (2014). Persistence of *Escherichia coli* O157 and non-O157 strains in agricultural soils. *Science of the Total Environment*, 490:822–829.

Magnússon, S. H., Guomundsdóttir, S., Reynisson, E., Rúnarsson, Á. R., Haroardóttir, H., Gunnarson, E., Georgsson, F., Reiersen, J., and Marteinsson, V. T. (2011). Comparison of *Campylobacter jejuni* isolates from human, food, veterinary and environmental sources in Iceland using PFGE, MLST and fla-SVR sequencing. *Journal of Applied Microbiology*, 111(4):971–981.

Mahoney, F. J., Farley, T. A., Burbank, D. F., Leslie, N. H., and McFarland, L. M. (1993). Evaluation of an intervention program for the control of an outbreak of shigellosis among institutionalized persons. *Journal of Infectious Diseases*, 168(5):1177–1180.

Mandrell, R. E., Harden, L. A., Bates, A., Miller, W. G., Haddon, W. F., and Fagerquist, C. K. (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):6292–6307.

Mather, A., Reid, S., Maskell, D., Parkhill, J., Fookes, M., Harris, S., Brown, D., Coia, J., Mulvey, M., Gilmour, M. o., Petrovska, L., De Pinna, E., Kuroda, M., Akiba, M., Izumiya, H., Connor, T., Suchard, M. l., Lemey, P., Mellor, D., Haydon, D., and Thomson, N. (2013). Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science*, 341(6153):1514–1517.

Mather, A. E., Vaughan, T. G., and French, N. P. (2015). Molecular approaches to understanding transmission and source attribution in nontyphoidal *Salmonella* and their application in Africa. *Clinical Infectious Diseases*, 61:S259–S265.

McLeod, K. S. (2000). Our sense of Snow: The myth of John Snow in medical geography. *Social Science and Medicine*, 50(7-8):923–935.

McMillian, M., Dunn, J. R., Keen, J. E., Brady, K. L., and Jones, T. F. (2007). Risk behaviors for disease transmission among petting zoo attendees. *Journal of the American Veterinary Medical Association*, 231(7):1036–1038.

Mellou, K., Sideroglou, T., Kallimani, A., Potamiti-Komi, M., Pervanidou, D., Lillakou, E., Georgakopoulou, T., Mandilara, G., Lambiri, M., Vatopoulos, A., and Hadjichristodoulou, C. (2013). Evaluation of under-reporting of salmonellosis and shigellosis hospitalised cases in Greece, 2011: Results of a capture-recapture study and a hospital registry review. *BMC Public Health*, 13(1):1–6.

Meunier, M., Guyard-Nicodème, M., Hirchaud, E., Parra, A., Chemaly, M., and Dory, D. (2016). Identification of novel vaccine candidates against *Campylobacter* through reverse vaccinology. *Journal of Immunology Research*, 2016:1–9.

Minin, V., Bloomquist, E., and Suchard, M. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, 25(7):1459–1471.

Mol, S., Cosansu, S., Ucok Alakavuk, D., and Ozturan, S. (2010). Survival of *Salmonella* Enteritidis during salting and drying of horse mackerel (*Trachurus trachurus*) fillets. *International Journal of Food Microbiology*, 139(1-2):36–40.

Montero, P., Gómez-Estaca, J., and Gómez-Guillén, M. C. (2007). Influence of salt, smoke, and high pressure on growth of *Listeria monocytogenes* spoilage microflora in cold-smoked dolphinfish (*Coryphaena hippurus*). *Journal of Food Protection*, 70(2):399–404.

Mughini-Gras, L., Enserink, R., Friesema, I., Heck, M., Van Duynhoven, Y., and Van Pelt, W. (2014). Risk factors for human salmonellosis originating from pigs, cattle, broiler chickens and egg laying hens: A combined case-control and source attribution analysis. *PLoS ONE*, 9(2):1–9.

Mughini-Gras, L., Smid, J. H., Wagenaar, J. A., de Boer, A. G., Havelaar, A. H., Friesema, I. H. M., French, N. P., Busani, L., and van Pelt, W. (2012). Risk factors for campylobacteriosis of chicken, ruminant, and environmental origin: A combined case-control and source attribution analysis. *PLoS ONE*, 7(8):1–13.

Mullner, P., Spencer, S. E. F., Wilson, D. J., Jones, G., Noble, A. D., Midwinter, A. C., Collins-Emerson, J. M., Carter, P., Hathaway, S., and French, N. P. (2009). Assigning the source of human campylobacteriosis in New Zealand: A comparative genetic and epidemiological approach. *Infection, Genetics and Evolution*, 9(6):1311–1319.

Nair, D. V. T., Kiess, A., Nannapaneni, R., Schilling, W., and Sharma, C. S. (2015). The combined efficacy of carvacrol and modified atmosphere packaging on the survival of *Salmonella*, *Campylobacter jejuni* and lactic acid bacteria on Turkey breast cutlets. *Food Microbiology*, 49:134–141.

Nataro, J. P., Bopp, C. A., Fields, P. I., Kaper, J. B., Strockbine, Nancy, A., Strockbine, B., and Strockbine, N. A. (2007). *Escherichia*, *Shigella* and *Salmonella*. In Murray, P. R., Baron, E. J., Jorgensen, J. H., Landry, M. L., and Pfaller, M. A., editors, *Manual of Clinical Microbiology*, chapter 43, pages 670–687. ASM Press, Washington DC, 9th edition.

Nedialkova, L. P., Sidstedt, M., Koeppel, M. B., Spriewald, S., Ring, D., Gerlach, R. G., Bossi, L., and Stecher, B. (2016). Temperate phages promote colicin-dependent fitness of *Salmonella enterica* serovar Typhimurium. *Environmental Microbiology*, 18(5):1591–1603.

Newton, K. (2016). Timeline: Havelock North's water contamination crisis. Retrieved 2017-05-04, from: `http://www.radionz.co.nz/news/national/311404/timeline-nz's-worst-waterborne-outbreak`, 1–1.

Nguyen, T. P., Friedrich, L. M., and Danyluk, M. D. (2014). Fate of *Escherichia coli* O157: H7 and salmonella on whole strawberries and blueberries of two maturities under different storage conditions. *Journal of Food Protection*, 77(7):1093–1101.

Nielsen, L. N., Sheppard, S. K., McCarthy, N. D., Maiden, M. C. J., Ingmer, H., and Krogfelt, K. A. (2010). MLST clustering of *Campylobacter jejuni* isolates from patients with gastroenteritis, reactive arthritis and Guillain-Barré syndrome. *Journal of Applied Microbiology*, 108(2):591–599.

Nobusawa, E. and Sato, K. (2006). Comparison of the mutation rates of human influenza A and B viruses. *Journal of Virology*, 80(7):3675–3678.

Nyberg, K. A., Vinnerås, B., Lewerin, S. S., Kjellberg, E., and Albihn, A. (2011). Treatment with Ca(OH)2 for inactivation of *Salmonella* Typhimurium and *Enterococcus faecalis* in soil contaminated with infected horse manure. *Journal of Applied Microbiology*, 110(6):1515–1523.

Oh, J. Y., Yu, H. S., Kim, S. K., Seol, S. Y., Cho, D. T., and Lee, J. C. (2003). Changes in patterns of antimicrobial susceptibility and integron carriage among *Shigella sonnei* isolates from southwestern Korea during epidemic periods. *Journal of Clinical Microbiology*, 41(1):421–423.

Okoro, C., Kingsley, R., Quail, M., Kankwatira, A., Feasey, N., Parkhill, J., Dougan, G., and Gordon, M. (2012). High-resolution single nucleotide polymorphism analysis distinguishes recrudescence and reinfection in recurrent invasive nontyphoidal salmonella typhimurium disease. *Clinical Infectious Diseases*, 54(7):955–963. cited By 41.

Okraszewska-Lasica, W., Bolton, D. J., Sheridan, J. J., and McDowell, D. A. (2014). Airborne *Salmonella* and *Listeria* associated with Irish commercial beef, sheep and pig plants. *Meat Science*, 97(2):255–261.

Oksenhendler, E., Gérard, L., Fieschi, C., Malphettes, M., Mouillot, G., Jaussaud, R., Viallard, J.-F., Gardembas, M., Galicier, L., Schleinitz, N., Suarez, F., Soulas-Sprauel, P., Hachulla, E., Jaccard, A., Gardeur, A., Théodorou, I., Rabian, C., and Debré, P. (2008). Infections in 252 patients with common variable immunodeficiency. *Clinical Infectious Diseases*, 46(10):1547–1554.

O'Loughlin, J. L., Samuelson, D. R., Braundmeier-Fleming, A. G., White, B. A., Haldorson, G. J., Stone, J. B., Lessmann, J. J., Eucker, T. P., and Konkel, M. E. (2015). The intestinal microbiota influences *Campylobacter jejuni* colonization and extraintestinal dissemination in mice. *Applied and Environmental Microbiology*, 81(14):4642–4650.

O'Mahony, L., Mccarthy, J., Kelly, P., Hurley, G., Luo, F., Chen, K., O'Sullivan, G. C., Kiely, B., Collins, J. K., Shanahan, F., and Quigley, E. M. M. (2005). *Lactobacillus* and *Bifidobacterium* in irritable bowel syndrome: Symptom responses and relationship to cytokine profiles. *Gastroenterology*, 128(3):541–551.

Osek, J. (2004). Phenotypic and genotypic characterization of *Escherichia coli* O157 strains isolated from humans, cattle and pigs. *Veterinarni Medicina*, 49(9):317–326.

Palhares, J. C. P., Kich, J. D., Bessa, M. C., Biesus, L. L., Berno, L. G., and Triques, N. J. (2014). *Salmonella* and antimicrobial resistance in an animal-based agriculture river system. *Science of the Total Environment*, 472:654–661.

Pérez-Cobas, A. E., Artacho, A., Ott, S. J., Moya, A., Gosalbes, M. J., and Latorre, A. (2014). Structural and functional changes in the gut microbiota associated to *Clostridium difficile* infection. *Frontiers in Microbiology*, 5(JULY):1–15.

Pitout, J. D. D., Reisbig, M. D., Mulvey, M., Chui, L., Louie, M., Crowe, L., Church, D. L., Elsayed, S., Gregson, D., Ahmed, R., Tilley, P., and Hanson, N. D. (2003). Association between handling of pet treats and infection with *Salmonella enterica* serotype Newport expressing the AmpC $\beta$-lactamase, CMY-2. *Journal of Clinical Microbiology*, 41(10):4578–4582.

Piveteau, P., Depret, G., Pivato, B., Garmyn, D., and Hartmann, A. (2011). Changes in gene expression during adaptation of *Listeria monocytogenes* to the soil environment. *PLoS ONE*, 6(9):1–10.

Praakle-Amin, K., Roasto, M., Korkeala, H., and Hänninen, M.-L. (2007). PFGE genotyping and antimicrobial susceptibility of *Campylobacter* in retail poultry meat in Estonia. *International Journal of Food Microbiology*, 114(1):105–112.

Pratap, C. B., Kumar, G., Patel, S. K., Verma, A. K., Shukla, V. K., Kumar, K., and Nath, G. (2013). Targeting of putative fimbrial gene for detection of *S.* Typhi in typhoid fever and chronic typhoid carriers by nested PCR. *Journal of Infection in Developing Countries*, 7(7):520–527.

Prendergast, M. M., Tribble, D. R., Baqar, S., Scott, D. A., Ferris, J. A., Walker, R. I., and Moran, A. P. (2004). *In vivo* phase variation and serologic response to lipooligosaccharide of *Campylobacter jejuni* in experimental human infection. *Infection and Immunity*, 72(2):916–922.

Public Health Surveillance (2017). Non-human Salmonella isolates, 2003-2016. Retrieved 2017-05-07, from: `https://surv.esr.cri.nz/enteric_reference/nonhuman_salmonella.php`, 1–1.

Quick, J., Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J., Nair, S., Neal, K., Nye, K., Peters, T., De Pinna, E., Robinson, E., Struthers, K., Webber, M., Catto, A., Dallman, T. J., Hawkey, P., and Loman, N. J. (2015). Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biology*, 16(1):1–14.

Raufu, I., Bortolaia, V., Svendsen, C. A., Ameh, J. A., Ambali, A. G., Aarestrup, F. M., and Hendriksen, R. S. (2013). The first attempt of an active integrated laboratory-based *Salmonella* surveillance programme in the north-eastern region of Nigeria. *Journal of Applied Microbiology*, 115(4):1059–1067.

Rea, M. C., Dobson, A., O'Sullivan, O., Crispie, F., Fouhy, F., Cotter, P. D., Shanahan, F., Kiely, B., Hill, C., and Paul Ross, R. (2011). Effect of broad- and narrow-spectrum antimicrobials on *Clostridium difficile* and microbial diversity in a model of the distal colon. *Proceedings of the National Academy of Sciences of the United States of America*, 108(SUPPL. 1):4639–4644.

Sarich, V. M. and Wilson, A. C. (1967). Immunological time scale for hominid evolution. *Science*, 158(3805):1200–1203.

Sears, A., Baker, M. G., Wilson, N., Marshall, J., Muellner, P., Campbell, D. M., Lake, R. J., and French, N. P. (2011). Marked campylobacteriosis decline after interventions aimed at poultry, New Zealand. *Emerging Infectious Diseases*, 17(6):1007–1015.

Sheppard, S. K., Dallas, J. F., Wilson, D. J., Strachan, N. J. C., McCarthy, N. D., Jolley, K. A., Colles, F. M., Rotariu, O., Ogden, I. D., Forbes, K. J., and Maiden, M. C. J. (2010). Evolution of an agriculture-associated disease causing *Campylobacter coli* clade: Evidence from national surveillance data in Scotland. *PLoS ONE*, 5(12):1–9.

Sternberg, S., Johnsson, A., Aspan, A., Bergstrm, K., Kallay, T., Szanto, E., and Johnson, A. (2008). Outbreak of *Salmonella* thompson infection in a swedish dairy herd. *Veterinary Record*, 163(20):596–599. cited By 6.

Stine, O. C., Alam, M., Tang, L., Nair, G. B., Siddique, A. K., Faruque, S. M., Huq, A., Colwell, R., Sack, R. B., and J Glenn Jr., M. (2008). Seasonal cholera from multiple small outbreaks, rural Bangladesh. *Emerging Infectious Diseases*, 14(5):831–833.

Stoddard, R. A., Miller, W. G., Foley, J. E., Lawrence, J., Gulland, F. M. D., Conrad, P. A., and Byrne, B. A. (2007). *Campylobacter insulaenigrae* isolates from northern elephant seals (*Mirounga angustirostris*) in California. *Applied and Environmental Microbiology*, 73(6):1729–1735.

Stone, S. P., Fuller, C., Savage, J., Cookson, B., Hayward, A., Cooper, B., Duckworth, G., Michie, S., Murray, M., Jeanes, A., Roberts, J., Teare, L., and Charlett, A. (2012). Evaluation of the national Cleanyourhands campaign to reduce *Staphylococcus aureus* bacteraemia and *Clostridium difficile* infection in hospitals in England and Wales by improved hand hygiene: Four year, prospective, ecological, interrupted ti. *BMJ (Online)*, 344(7858):1–11.

Strachan, N. J. C., Dunn, G. M., and Ogden, I. D. (2002). Quantitative risk assessment of human infection from *Escherichia coli* O157 associated with recreational use of animal pasture. *International Journal of Food Microbiology*, 75(1-2):39–51.

Strawn, L. K., Gröhn, Y. T., Warchocki, S., Worobo, R. W., Bihn, E. A., and Wiedmann, M. (2013). Risk factors associated with *Salmonella* and *Listeria monocytogenes* contamination of produce fields. *Applied and Environmental Microbiology*, 79(24):7618–7627.

Taur, Y. and Pamer, E. (2014). Harnessing microbiota to kill a pathogen: Fixing the microbiota to treat *Clostridium difficile* infections. *Nature Medicine*, 20(3):246–247. cited By 16.

Thornley, C., Simmons, G., Callaghan, M., Nicol, C., Baker, M., Gilmore, K., and Garrett, N. (2003). First incursion of *Salmonella enterica* serotype Typhimurium DT160 into New Zealand. *Emerging Infectious Diseases*, 9(4):493–495.

Turner, C. E., Dryden, M., Holden, M. T. G., Davies, F. J., Lawrenson, R. A., Farzaneh, L., Bentley, S. D., Efstratiou, A., and Sriskandan, S. (2013). Molecular analysis of an outbreak of lethal postpartum sepsis caused by *Streptococcus pyogenes*. *Journal of Clinical Microbiology*, 51(7):2089–2095.

Tuttle, J., Gomez, T., Doyle, M., Wells, J., Zhao, T., Tauxe, R., and Griffin, P. (1999). Lessons from a large outbreak of *Escherichia coli* o157:h7 infections: Insights into the infectious dose and method of widespread contamination of hamburger patties. *Epidemiology and Infection*, 122(2):185–192. cited By 200.

Ukena, S. N., Singh, A., Dringenberg, U., Engelhardt, R., Seidler, U., Hansen, W., Bleich, A., Bruder, D., Franzke, A., Rogler, G., Suerbaum, S., Buer, J., Gunzer, F., and Westendorf, A. M. (2007). Probiotic *Escherichia coli* Nissle 1917 inhibits leaky gut by enhancing mucosal integrity. *PLoS ONE*, 2(12):1–9.

Unicomb, L. E., Dalton, C. B., Gilbert, G. L., Becker, N. G., and Patel, M. S. (2008). Age-specific risk factors for sporadic *Campylobacter* infection in regional Australia. *Foodborne Pathogens and Disease*, 5(1):79–85.

Vaezirad, M., Keestra-Gounder, A., de Zoete, M., Koene, M., Wagenaar, J., and van Putten, J. (2017). Invasive behavior of *Campylobacter jejuni* in immunosuppressed chicken. *Virulence*, 8(3):248–260. cited By 1.

Valera, R., García, H. M., Díaz Jidy, M., Mirabal, M., Armesto, M. I., Fando, R., García, L., Fernández, R., Año, G., Cedré, B., Ramírez, M., Bravo, L., Serrano, T., Palma, S., González, D., Miralles, F., Medina, V., Nuñez, F., Plasencia, Y., Martínez, J. C., Mandarioti, A., Lugones, J., Rodríguez, B. L., Moreno, A., González, D., Baro, M., Solis, R. L., Sierra, G., Barbera, R., Domínguez, F., Gutiérrez, C., Kouri, G., Campa, C., and Menéndez, J. (2009). Randomized, double-blind, placebo-controlled trial to evaluate the safety and immunogenicity of live oral cholera vaccine 638 in Cuban adults. *Vaccine*, 27(47):6564–6569.

Van Dyke, M. I., Morton, V. K., McLellan, N. L., and Huck, P. M. (2010). The occurrence of *Campylobacter* in river water and waterfowl within a watershed in southern Ontario, Canada. *Journal of Applied Microbiology*, 109(3):1053–1066.

Van Pelt, W., Van De Giessen, A. W., Van Leeuwen, W. J., Wannet, W., Henken, A. M., Evers, E., De Wit, M. A. S., and Van Durnhoven, Y. T. H. P. (1999). Oorspong van humane salmonellose met betrekking tot varken, rund, kip, ei en overige bronnen. *Infectieziekten Bulletin*, 10:240–243.

Varma, J. K., Greene, K. D., Reller, M. E., DeLong, S. M., Trottier, J., Nowicki, S. F., DiOrio, M., Koch, E. M., Bannerman, T. L., York, S. T., Lambert-Fair, M.-A., Wells, J. G., and Mead, P. S. (2003). An outbreak of *Escherichia coli* O157 infection following exposure to a contaminated building. *Journal of the American Medical Association*, 290(20):2709–2712.

Vaughan, T. G., Kühnert, D., Popinga, A., Welch, D., and Drummond, A. J. (2014). Efficient Bayesian inference under the structured coalescent. *Bioinformatics*, 30(16):2272–2279.

Villar, R. G., Connick, M., Barton, L. L., Meaney, F. J., and Davis, M. F. (1998). Parent and pediatrician knowledge, attitudes, and practices regarding pet-associated hazards. *Archives of Pediatrics and Adolescent Medicine*, 152(10):1035–1036.

Vilte, D. A., Larzábal, M., Mayr, U. B., Garbaccio, S., Gammella, M., Rabinovitz, B. C., Delgado, F., Meikle, V., Cantet, R. J. C., Lubitz, P., Lubitz, W., Cataldi, A., and Mercado, E. C. (2012). A systemic vaccine based on *Escherichia coli* O157:H7 bacterial ghosts (BGs) reduces the excretion of *E. coli* O157:H7 in calves. *Veterinary Immunology and Immunopathology*, 146(2):169–176.

Vivant, A.-L., Garmyn, D., Gal, L., Hartmann, A., and Piveteau, P. (2015). Survival of *Listeria monocytogenes* in soil requires AgrA-mediated regulation. *Applied and Environmental Microbiology*, 81(15):5073–5084.

Wagenaar, J. A., French, N. P., and Havelaar, A. H. (2013). Preventing *Campylobacter* at the source: Why is it so difficult? *Clinical Infectious Diseases*, 57(11):1600–1606.

Waitt, J. A., Kuhn, D. D., Welbaum, G. E., and Ponder, M. A. (2014). Postharvest transfer and survival of *Salmonella enterica* serotype enteritidis on living lettuce. *Letters in Applied Microbiology*, 58(2):95–101.

Walters, M. S., Simmons, L., Anderson, T. C., De Ment, J., Van Zile, K., Matthias, L. P., Etheridge, S., Baker, R., Healan, C., Bagby, R., Reporter, R., Kimura, A., Harrison, C., Ajileye, K., Borders, J., Crocker, K., Smee, A., Adams-Cameron, M., Joseph, L. A., Tolar, B., Trees, E., Sabol, A., Garrett, N., Bopp, C., Bosch, S., and Behravesh, C. B. (2016). Outbreaks of salmonellosis from small turtles. *Pediatrics*, 137(1):1–9.

Wang, T. K. F., Yam, W.-C., Yuen, K.-Y., and Wong, S. S. Y. (2006). Misidentification of a mucoid strain of *Salmonella enterica* serotype choleraesuis as *Hafnia alvei* by the Vitek GNI+ card system. *Journal of Clinical Microbiology*, 44(12):4605–4608.

Warshawsky, B., Gutmanis, I., Henry, B., Dow, J., Reffle, J., Pollett, G., Ahmed, R., Aldom, J., Alves, D., Chagla, A., Ciebin, B., Kolbe, F., Jamieson, F., and Rodgers, F. (2002). Outbreak of *Escherichia coli* 0157:H7 related to animal contact at a petting zoo. *Canadian Journal of Infectious Diseases*, 13(3):175–181.

WHO (2016a). Disease outbreaks. Retrieved 2017-06-29, from: `http://www.who.int/topics/disease_outbreaks/en/`, 1–1.

WHO (2016b). Risk factors. Retrieved 2017-06-29, from: `http://www.who.int/topics/risk_factors/en/`, 1-1.

Wikström, V. O., Fernström, L.-L., Melin, L., and Boqvist, S. (2014). *Salmonella* isolated from individual reptiles and environmental samples from terraria in private households in Sweden. *Acta Veterinaria Scandinavica*, pages 1–6.

Wilkes, G., Brassard, J., Edge, T. A., Gannon, V., Jokinen, C. C., Jones, T. H., Neumann, N., Pintar, K. D. M., Ruecker, N., Schmidt, P. J., Sunohara, M., Topp, E., and Lapen, D. R. (2013). Bacteria, viruses, and parasites in an intermittent stream protected from and exposed to pasturing cattle: Prevalence, densities, and quantitative microbial risk assessment. *Water Research*, 47(16):6244–6257.

Willis, W. L. and Reid, L. (2008). Investigating the effects of dietary probiotic feeding regimens on broiler chicken production and *Campylobacter jejuni* presence. *Poultry Science*, 87(4):606–611.

Wilson, D. J., Gabriel, E., Leatherbarrow, A. J. H., Cheesbrough, J., Gee, S., Bolton, E., Fox, A., Fearnhead, P., Hart, C. A., and Diggle, P. J. (2008). Tracing the source of campylobacteriosis. *PLoS Genetics*, 4(9):1–9.

Worby, C. J., Lipsitch, M., and Hanage, W. P. (2014). Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. *PLoS Computational Biology*, 10(3):1–10.

Wyszyńska, A., Raczko, A., Lis, M., and Jagusztyn-Krynicka, E. K. (2004). Oral immunization of chickens with avirulent *Salmonella* vaccine strain carrying *C. jejuni* 72Dz/92 *cjaA* gene elicits specific humoral immune response associated with protection against challenge with wild-type *Campylobacter*. *Vaccine*, 22(11-12):1379–1389.

Yamamoto, T. and Taneike, I. (2000). The sequences of enterohemorrhagic *Escherichia coli* and *Yersinia pestis* that are homologous to the enteroaggregative *E. coli* heat-stable enterotoxin gene: Cross-species transfer in evolution. *FEBS Letters*, 472(1):22–26.

Yoon, K. S., Burnette, C. N., Abou-Zeid, K. A., and Whiting, R. C. (2004). Control of growth and survival of *Listeria monocytogenes* on smoked salmon by combined potassium lactate and sodium diacetate and freezing stress during refrigeration and frozen storage. *Journal of Food Protection*, 67(11):2465–2471.

Young, N. J., Day, J., Montsho-Hammond, F., Verlander, N. Q., Irish, C., Pankhania, B., and Oliver, I. (2014). *Campylobacter* infection associated with consumption of duck liver pâté: A retrospective cohort study in the setting of near universal exposure. *Epidemiology and Infection*, 142(6):1269–1276.

Younus, M., Wilkins, M. J., Davies, H. D., Rahbar, M. H., Funk, J., Nguyen, C., Siddiqi, A. E., Cho, S., and Saeed, A. M. (2010). The role of exposures to animals and other risk factors in sporadic, non-typhoidal *Salmonella* infections in Michigan children. *Zoonoses and Public Health*, 57(7-8):e170–e176.

Zeng, X., Brown, S., Gillespie, B., and Lin, J. (2014). A single nucleotide in the promoter region modulates the expression of the $\beta$-lactamase OXA-61 in *Campylobacter jejuni*. *Journal of Antimicrobial Chemotherapy*, 69(5):1215–1223.

Zhang, L., Moosekian, S. R., Todd, E. C. D., and Ryser, E. T. (2012). Growth of *Listeria monocytogenes* in different retail delicatessen meats during simulated home storage. *Journal of Food Protection*, 75(5):896–905.

Zhang, Z. F. and Kim, I. H. (2013). Effects of probiotic supplementation in different energy and nutrient density diets on performance, egg quality, excreta microflora, excreta noxious gas emission, and serum cholesterol concentrations in laying hens. *Journal of Animal Science*, 91(10):4781–4787.

Zhao, T., Doyle, M. P., Harmon, B. G., Brown, C. A., Mueller, P. O. E., and Parks, A. H. (1998). Reduction of carriage of enterohemorrhagic *Escherichia coli* O157:H7 in cattle by inoculation with probiotic bacteria. *Journal of Clinical Microbiology*, 36(3):641–647.

# Chapter 3

## Genomic analysis of *Salmonella enterica* serovar Typhimurium DT160 associated with a 14-year outbreak, New Zealand, 1998-2012

### 3.1 Abstract

During 1998-2012, an extended outbreak of *Salmonella enterica* serovar Typhimurium definitive type 160 (DT160) affected more than 3,000 humans and killed wild birds in New Zealand. However, the relationship between DT160 within these two host groups and the origin of the outbreak are unknown. Whole-genome sequencing was used to compare 109 DT160 isolates from sources throughout New Zealand. Genomic analyses provided evidence that DT160 was introduced into New Zealand around 1997 and rapidly propagated throughout the country, becoming more genetically diverse over time. The genetic heterogeneity was evenly distributed across multiple predicted functional protein groups, and there was no evidence of host group differentiation between isolates collected from human, poultry, bovine, and wild bird sources, indicating ongoing transmission between these host groups. This study demonstrates how a comparative genomic approach can be used to gain insight into outbreaks, disease transmission, and the evolution of a multihost pathogen after a probable point-source introduction.

### 3.2 Introduction

Non-typhoidal *Salmonella* strains, which cause salmonellosis, are responsible for an estimated 93.8 million illnesses and 155,000 deaths among humans worldwide each year (Majowicz et al., 2010). In New Zealand, these strains are the second largest cause of bacterial enteritis, annually causing 23 cases per 100,000 population (ESR, 2015). Non-typhoid *Salmonella* strains vary in host specificity and are usually transmitted to humans via direct contact or consumption of foods originating from animals (Baker et al., 2007; King et al., 2011). In New Zealand, salmonellosis incidence among humans peaks in the warm summer months, probably in association with increased multiplication of *Salmonella* in animal and food sources, and with increased participation in higher risk outdoor activities (e.g. activities that increase contact with wild-life) (Lal et al., 2012). Climate change is expected to increase summer temperatures, potentially increasing salmonellosis incidence in New Zealand (Lal et al., 2013).

During 1998-2012, an extended outbreak of *Salmonella enterica* serovar Typhimurium definitive type 160 (DT160) occurred in New Zealand (Alley et al., 2002; ESR, 2003; Public Health Surveillance, 2017a). During the outbreak, DT160 was the predominant *Salmonella* strain isolated from human salmonellosis patients and sick wild birds. DT160 was also isolated from other animals and the environment (Figure 3.1), but it was not the main *Salmonella* strain isolated from these sources (ESR, 2003; Public Health Surveillance, 2017b). DT160 has been isolated from animals and environments worldwide (Penfold et al., 1979; Tizard et al., 1979) and is usually associated with moribund birds (Lawson et al., 2010; Piccirillo et al., 2010). Before the 1998-2012 outbreak, DT160 had not been reported in New Zealand. In 2009, an outbreak of DT160 involving humans and wild birds was reported in Tasmania, Australia (Grillo and Post, 2010); however, as with the outbreak in New Zealand, the relationship between DT160 within the bird and human host groups of Tasmania was unknown.

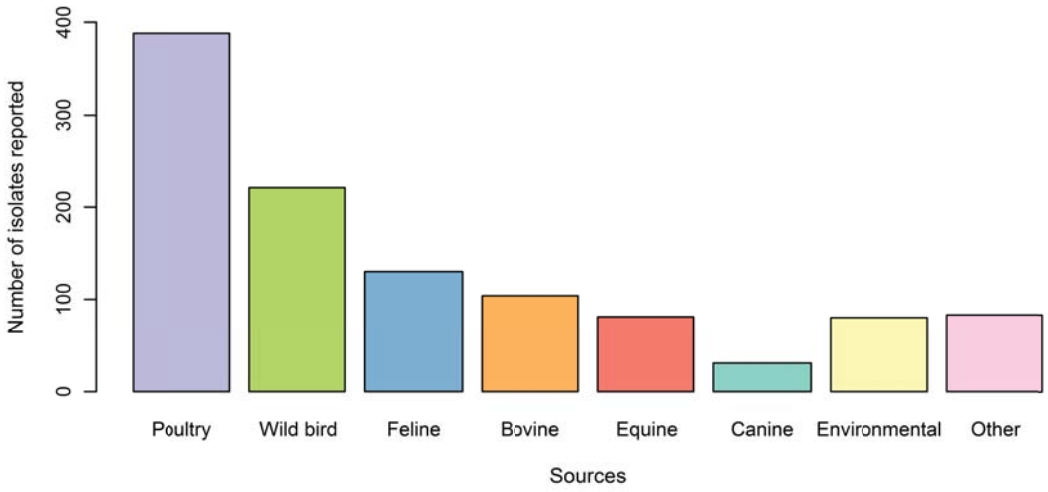


**Figure 3.1.** Bar graph of the number of DT160 isolates reported from non-human sources from 1998-2012 (ESR, 2003; Public Health Surveillance, 2017b).

During the 1998-2012 outbreak of DT160 in New Zealand, human incidence displayed a typical epidemic curve: prevalence increased from 1999 to 2000, before peaking at 791 patients in 2001, and then slowly decreased from 2002 to 2012. DT160 incidence within the bovine, poultry and wild bird populations displayed similar epidemic curves (Figure 3.2). The aim of this study was to use genomic epidemiologic approaches to characterise the origin, evolution, and transmission of DT160 in New Zealand.

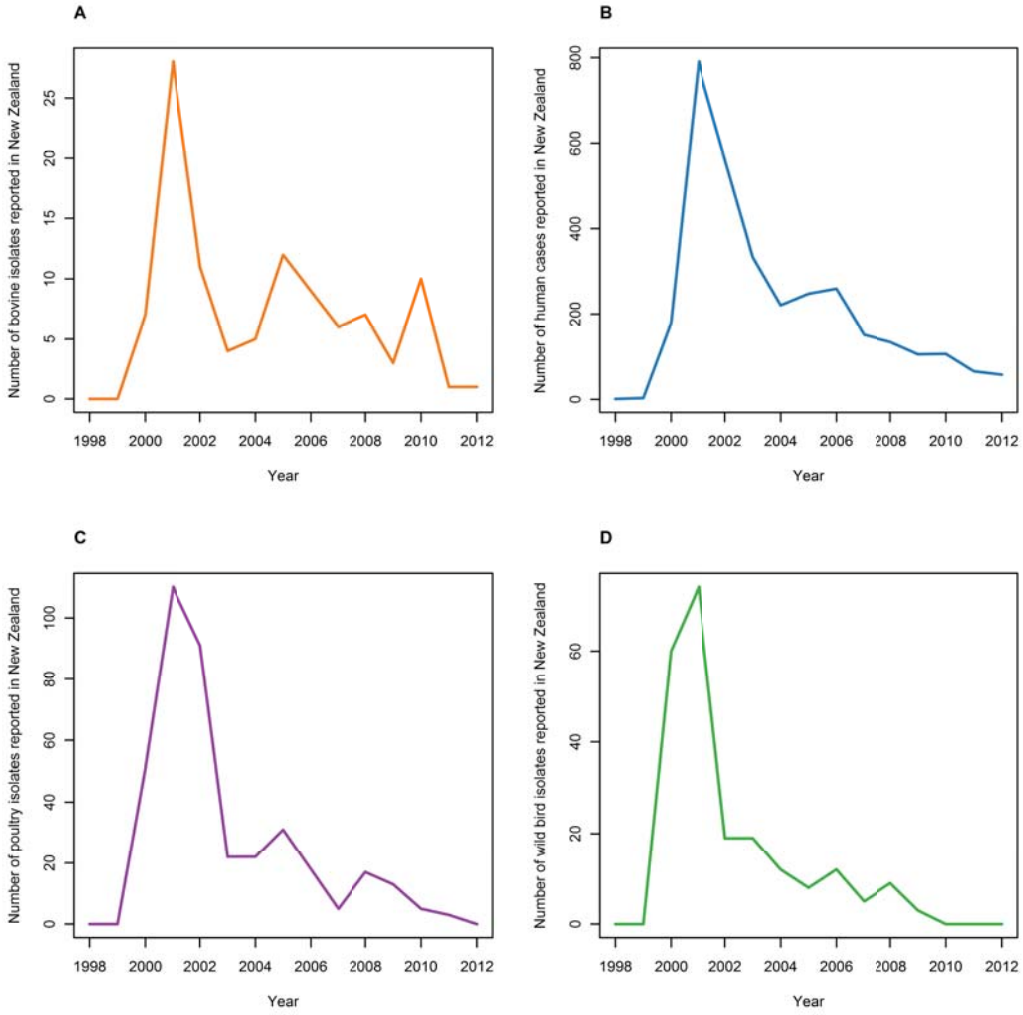


**Figure 3.2.** Line graphs of the number of bovine (A: orange), human (B: blue), poultry (C: purple) and wild bird (D: green) DT160 cases reported each year in New Zealand from 1998-2012 (ESR, 2003; Public Health Surveillance, 2017b,a).

## 3.3 Methods

### 3.3.1 Whole-genome sequencing

After stratifying the *Salmonella* strain collection at the Enteric Reference Laboratory of the Institute of Environmental Science and Research Ltd. (Wallaceville, New Zealand) by age and host, 35 human, 25 wild bird, 25 poultry, and 24 bovine DT160 isolates were randomly selected from 1998-2012. Genomic DNA was extracted from these isolates using a QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) (Qiagen, 2012). Genome extracts were whole genome sequenced by New Zealand Genomics Limited (NZGL) at the Massey Genome Service (Massey University, Palmerston North, New Zealand). A library was prepared for each isolate

using an Illumina TruSeq™DNA PCR-Free kit (Illumina, Scorsby, Victoria, Australia) and sequenced using an lllumina MiSeq (Illumina, Scorsby, Victoria, Australia) as 2x250 bp paired-end runs (120-150x average genome coverage). After sequencing and barcode demultiplexing, PhiX control library reads and adapter sequences were removed using FASTQ-MCF (Aronesty, 2013). The raw reads for the 109 DT160 isolates are available in the European Nucleotide Archive (http://www.ebi.ac.uk/ena; accession no. PRJEB18077).

### 3.3.2 Genome assembly

Each isolates genome was assembled *de novo*. An in-house Perl script was used to trim reads at an error probability of 0.01 using solexaQA++ (Cox et al., 2010) and generate random subsets of paired reads from 750,000 to 1,200,000 paired reads in increments of 150,000, varying the average coverage. Each of the random sets was assembled using the *de novo* assembler Velvet v1.1 (Zerbino and Birney, 2008) at a variety of kmers (from 55 to 245) in increments of 10. *De novo* assembly resulted in multiple genome assemblies for each isolate. Metrics for each of four parameters (longest genome length, fewest number of contigs, largest $N_{50}$ value, and longest contig length) were ranked in numeric order and an overall equally summed ranking score was calculated. The assemblies with the lowest total rank were used for further analyses. QUAST (Gurevich et al., 2013), a quality assessment tool for evaluating and comparing genome assemblies, was used to analyse the *de novo* assemblies and determine their GC content (i.e. the percentage of a DNA sequence made up of guanine and cytosine bases).

### 3.3.3 Single-nucleotide polymorphism identification

Snippy v2.6 (https://github.com/tseemann/snippy) and kSNP v3.0 were used (Gardner et al., 2015) to identify core single-nucleotide polymorphisms (SNPs). Snippy is a pipeline that uses the Burrows-Wheelers Aligner (Li and Durbin, 2009) and SAMtools v1.3.1 (Li, 2011) to align reads from different isolates to a sequence and uses FreeBayes (Garrison and Marth, 2012) to identify variants among the alignments. kSNP was used to analyse *de novo* assembled genomes, along with the reference genome, *S.* Typhimurium 14028S (NC_016856). An in-house Python script was used to determine the read coverage of all the SNPs identified via kSNP. Snippy was used to align reads from each isolate to the reference genome (NC_016856) before identifying SNPs. SNPs were accepted if they had a greater than 10 read depth and a greater than 90% consensus for each isolate. The position of the SNP on the reference genome was used to determine if both methods identified the SNP or if they were unique to the method (Appendix A.1). This method identified 793 core SNPs shared by the 109 New Zealand DT160 isolates.

### 3.3.4 Global DT160 strains

Using the genomic assembly and SNP identification methods described above, two DT160 strains from the United Kingdom were compared with the 109 DT160 isolates from New Zealand: 1,521 core SNPs were identified. The UK strains were previously published by Petrovska et al. (2016), and downloaded from the European Nucleotide Archive (ERS015626 and ERS015627).

### 3.3.5 Phylogenetic inference and distances

RAxML v8.2.4 (Stamatakis, 2014) was used to construct a maximum-likelihood tree based of the 793 core SNPs shared by the 109 New Zealand DT160 isolates; EvolView v2 (He et al., 2016) was used to visualise and edit the tree. SplitsTree (Huson and Bryant, 2006) was used to form a NeighborNet tree of the 109 New Zealand DT160 isolates based on the 793 core SNPs they shared and to compare the New Zealand and UK isolates based on the 1,521 core SNPs they shared. MEGA6 (Tamura et al., 2013) and the maximum composite likelihood model (Tamura et al., 2004) were used to predict the pairwise distance between the 109 New Zealand DT160 isolates, based on the 793 core SNPs they shared, and the 109 New Zealand and two UK isolates, based on the 1,521 core SNPs they shared.

### 3.3.6 Phylogenetic analysis

An in-house Perl script was used to split the 793 SNPs shared by the 109 New Zealand DT160 isolates into five groups: those associated with the first, second, or third codon; those contained in overlapping coding regions; and those found in intergenic regions. The script was also used to determine whether the SNPs were synonymous or non-synonymous. The partitioned SNPs were exported into BEAUti to create an XML (Extensible Markup Language) file for BEAST 1.8.3 (Drummond et al., 2012).

To allow for variation in base substitution among codon positions, the five SNP groups were given separate Hasegawa Kishino Yano models (Hasegawa et al., 1985); to allow for and estimate changes in the effective population size, the Gaussian Markov random field (GMRF) Bayesian skyride model (Minin et al., 2008) was used; to allow for variation in mutation rates among lineages, an uncorrelated relaxed molecular clock (Drummond et al., 2006) was used, which was calibrated by the tip dates. The XML file was run in BEAST for 40 million steps a total of 3 times with different starting seeds before LogCombiner (http://beast.bio.ed.ac.uk/LogCombiner) was used to combine the runs with a 10% burn-in. The results were visualised and the relative change in effective population size was calculated using Tracer v1.6 (Rambaut et al., 2014).

The mutation rate for the DT160 genome was calculated by multiplying the mutation rate estimated by BEAST by the number of SNPs analysed (793 bp) before dividing the product by the mean genome size

of the analysed isolates (4,884,485 bp). The discrete trait analysis model (Lemey et al., 2009) was used to predict ancestral migrations between host groups over the course of the outbreak.

### 3.3.7 Protein-coding gene analysis

Prokka (Seemann, 2014) was used to annotate *de novo* assembled genomes, and Roary (Page et al., 2015) was used to cluster proteins and identify those that were found only in a subset of isolates and those that differed in length between the isolates. ClustalW v2.1 (Thompson et al., 1994) was used to align amino acid sequences, and an in-house Perl script was used to determine if these alignments contained mismatches. The nucleotide sequences encoding all proteins that differed were extracted from the assembled genomes, along with 500-bp flanks on either side of the sequence, using an in-house Perl script. 500-bp flanks could not be obtained for some genes because they were located at the end of contigs. For those genes, the flank was cut short, but their length was annotated. Flanks were extracted to help with read alignment. This extraction left a pool of nucleotide sequences from each isolate, for every protein that potentially differed in sequence. For each protein, all nucleotide variants were extracted from the pool by using an in-house Perl script. SRST2 v2, a read mapping-based tool (Inouye et al., 2014), was used to align reads from each isolate to the sequence variants, and SAMtools v1.3.1 (Li, 2011) was used to form a consensus sequence from the aligned reads. The consensus cutoff was set as the default for SAMtools (read depth of >8 and a consensus of >80%). The flanks were removed from the consensus sequences, and the sequence variants were translated into amino acid sequences using an in-house Perl script. Protein differences were identified by comparing the amino acid sequences from each isolate and were combined with the non-synonymous SNPs identified by SNP analysis. The position of non-synonymous SNPs within proteins was used to prevent repeats.

The protein sequences shared by the 109 DT160 isolates were compared with those in the Clusters of Orthologous Groups of proteins (COGs) database (Tatusov et al., 2000) to predict protein functions. For each functional group, the proportion of proteins that differed in sequence was calculated, and a Fisher exact test, computed via Markov chain Monte Carlo of $10^9$ iterations, was used to determine if there were any differences between these proportions.

An in-house Perl script was used to form a presence-absence matrix of all the protein differences. Primer-E v6 (Clarke and Gorley, 2006) was used to predict the Euclidian distance between the isolates based on the presence-absence matrix. The centroid is the arithmetic mean for a group of data points in an n-dimensional space. PERMANOVA (http://www.primer-e.com/permanova.htm) was used to assess differences in centroids among isolates collected from different sources or time periods. To assess differences in dispersions between different groups, PermDisp (Anderson, 2006) was used to compute dispersions (z-values) that were modelled using a regression model with date of collection and source as the explanatory variables.

### 3.3.8 Antimicrobial susceptibility

The antimicrobial susceptibility profiles of 90 DT160 isolates collected in New Zealand from human (n=30), poultry (n=30) and wild bird (n=30) sources were obtained from Omar (2010), who measured each isolate's susceptibility to amikacin amoxicillin/clavulanate, ampicillin, cefoxitin, cefpodoxime, chloramphenicol, ciprofloxacin, nalidixic acid, oxytetracycline, trimethoprim/sulfamethoxazole and tetracycline, using disc diffusion. The zone of inhibitions for the antimicrobials were modelled using regression models with date of collection and source as the explanatory variables.

### 3.3.9 Scripts

The in-house scripts used for genomic analyses in this study were specifically designed for this dataset. The scripts are available from GitHub (https://github.com/samuelbloomfield/Scripts-for-genomic-analyses).

## 3.4 Results

### 3.4.1 Genomic DT160 comparison

The *de novo* assembled DT160 genomes were 4.8-4.9 Mb in length and had a GC content of 52.11%-52.16%. This is within the normal GC content range for *S. enterica*: 50%-53% (Popoff and Le Minor, 2005). 793 core SNPs were identified amongst the 109 DT160 isolates from New Zealand.

### 3.4.2 DT160 introduction date

Ancestral date reconstruction analysis predicted that the 109 New Zealand DT160 isolates shared a date of common ancestor in approximately August 1997 (95% highest posterior density (HPD) interval: June 1996-August 1998). Comparative analysis indicated that the two DT160 isolates collected from the United Kingdom (ERS015626 from a horse in 1998 and ERS015627 from a bird in 1997) were genetically distinct from the 109 New Zealand DT160 isolates (Figure 3.3). The average pairwise SNP distance between the two UK DT160 isolates and the New Zealand isolates was 0.0287, compared with an average pairwise distance of 0.0151 between New Zealand isolates.

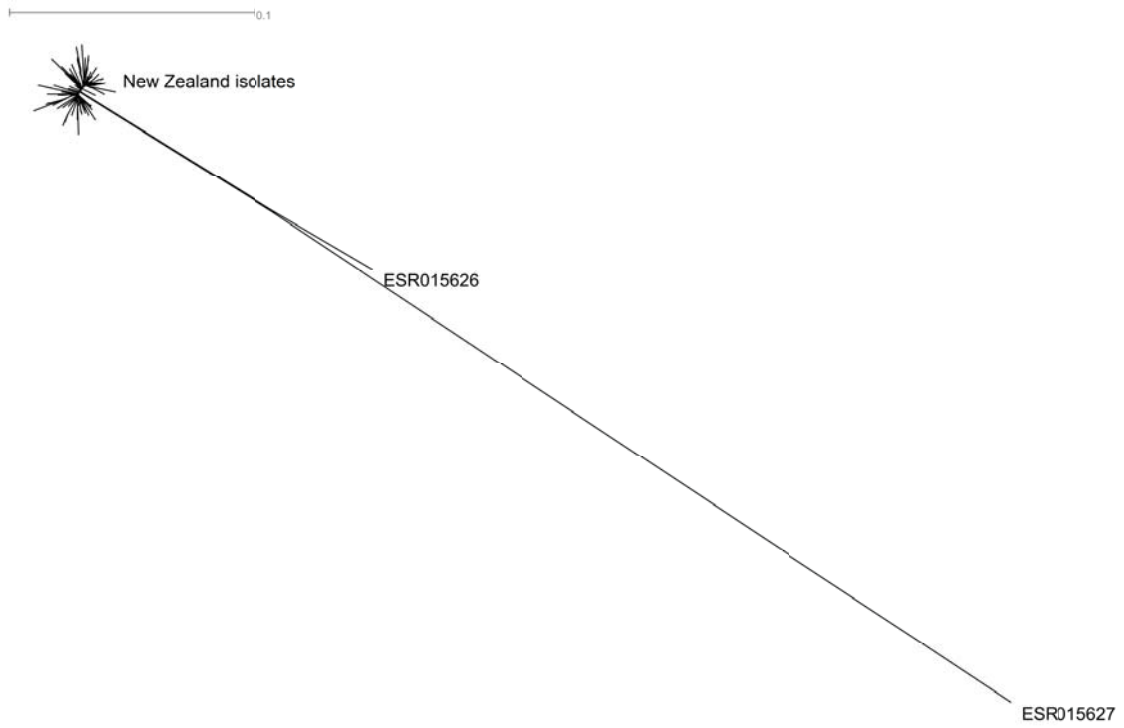


**Figure 3.3** NeighborNet tree of 111 DT160 isolates (based on 1,521 core SNPs): 109 from New Zealand and two from the United Kingdom (ERS015626 and ERS015627). The scale bar represents the number of nucleotide substitutions per site.

### 3.4.3 DT160 evolution

Phylogenetic analysis predicted that the 109 DT160 isolates mutated at a rate of 3.3-4.3 x $10^{-7}$ substitutions $site^{-1}$ $year^{-1}$ (95% HPD) and that the DT160 effective population size increased from 1998 to 2003 (Figure 3.4). Over the course of the outbreak, DT160 also increased in genetic diversity (Figure 3.5).

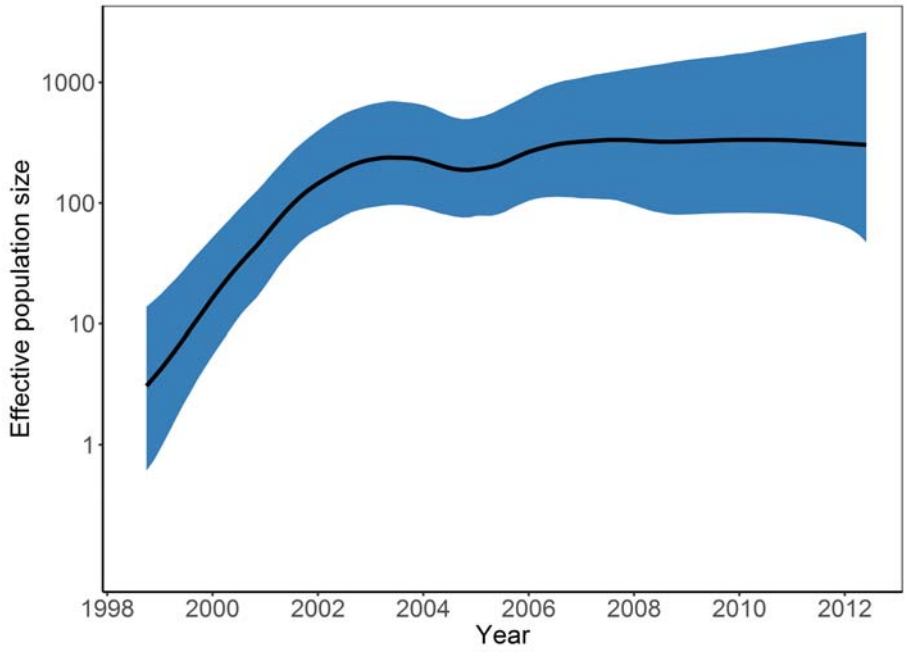


**Figure 3.4.** Relative effective population size (log scale) of DT160 during an outbreak in New Zealand, 1998-2012. The black line represents the median effective population size estimate; blue represents the 95% HPD interval.

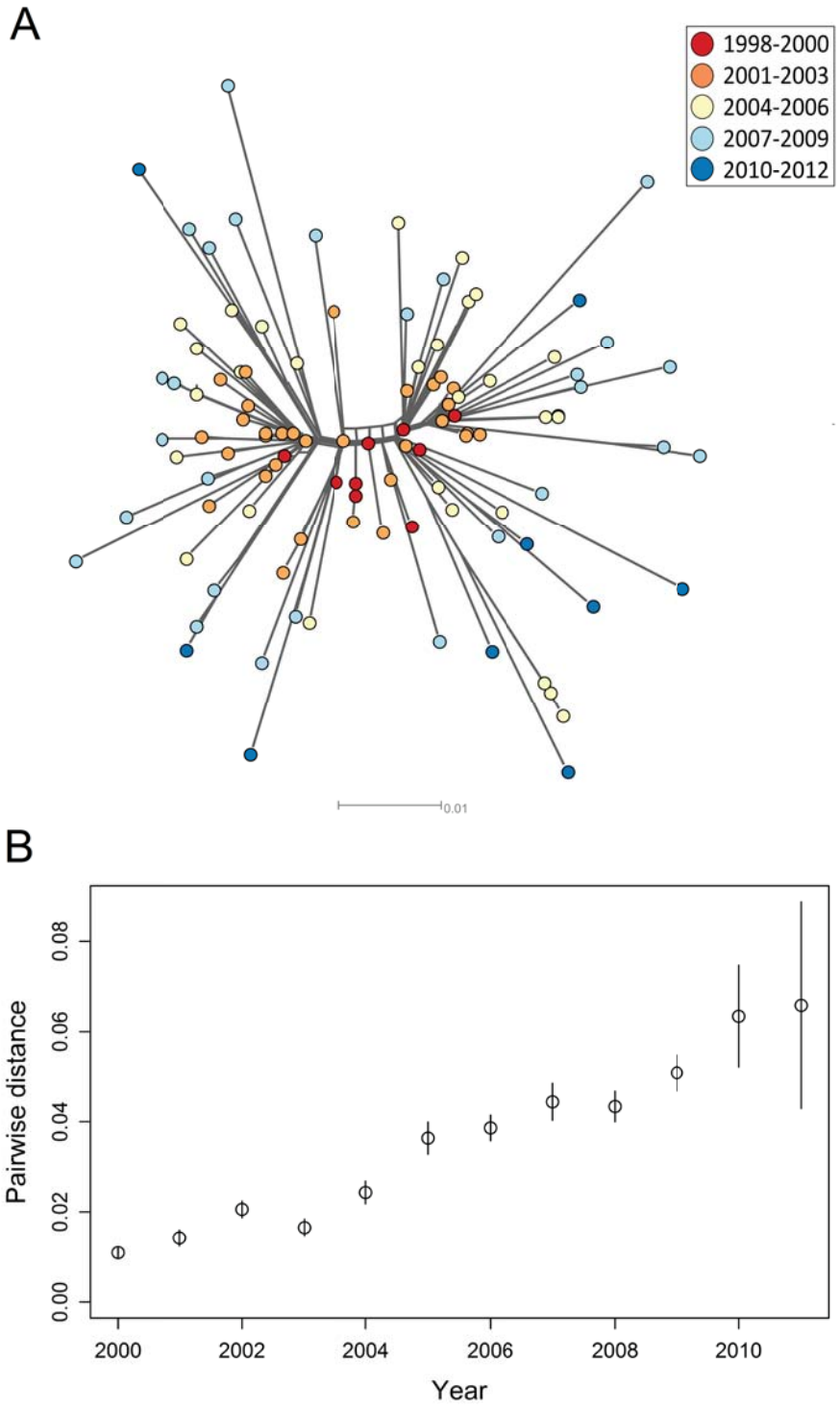


**Figure 3.5.** A) NeighborNet tree of 109 DT160 isolates collected during an outbreak in New Zealand, 1998-2012 (based on 793 core SNPs). Colours indicate date of isolate collection. The scale bar represents the number of nucleotide substitutions per site. B) Scatterplot of the mean pairwise distance of 106 DT160 isolates from 2000-2011. Error bars represent 95% confidence intervals.

### 3.4.4 DT160 sources

PFGE (pulsed-field gel electrophoresis) was previously used to compare New Zealand DT160 isolates from humans, poultry, and wild birds (Omar, 2010); however, PFGE could not distinguish DT160 from these sources. In this study, whole-genome sequencing was able to distinguish DT160 at the isolate level. However, no distinct DT160 clades associated with any one source were identified (Figure 3.6).

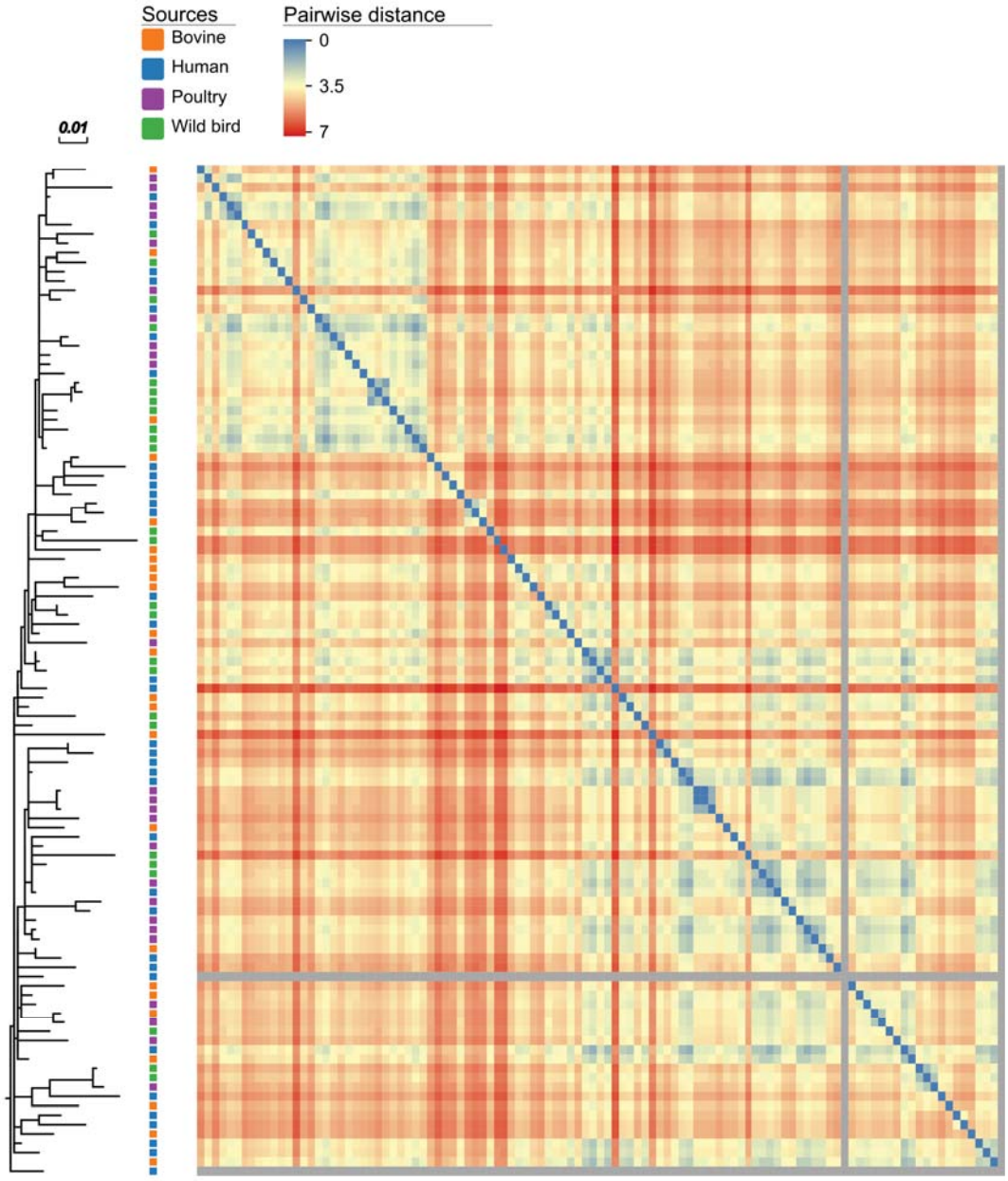


**Figure 3.6.** Maximum-likelihood tree of 109 DT160 isolates collected during an outbreak in New Zealand, 1998-2012 (based on 793 core SNPs). Coloured squares to the right of the branches indicate the source of isolates. The scale bar represents the number of nucleotide substitutions per site. The heat map represents the Euclidean pairwise distance between isolates (based on the presence of 684 protein differences). Isolates that shared a small number of protein differences contained small Euclidean distances and are closer to blue in colour on the heat map; isolates that shared a large number of protein differences contained large Euclidean distances and are closer to red in colour. The grey squares represent the two outliers missing a large number of genes. The diagonal array of blue squares represents the pairwise distance for the same isolates.

### 3.4.5 Protein and gene analysis

Protein annotation identified 5,096 coding-DNA sequences (CDSs) contained by the 109 New Zealand DT160 isolates, of which 4,983 (98%) were found in all of the isolates, 108 (2%) were found in 95%-99% of isolates, and 3 (<1%) were found in 1%-5% of the isolates. Protein-coding gene analysis also identified 477 non-synonymous SNPs, of which 27 were nonsense mutations and 96 were INDELs (insertions/deletions). The nonsense SNPs and INDELs were responsible for 123 proteins that differed in length. In total, 684 differences in 604 protein sequences were identified. Two isolates were excluded from protein-coding gene analysis because they were missing a large number of proteins (Appendix A.2).

PERMANOVA found that centroids based on the 684 protein differences were indistinguishable among groups of DT160 isolates collected from different sources and time periods (Figure 3.7). PERMANOVA's inability to distinguish centroids appears to be due to the fact that DT160 isolates radiated out from a point source. The z-value is the distance from an isolate to the centroid of a group of isolates; the z-value for 107 DT160 isolates were calculated based on the 684 protein differences. Regression modeling showed that the z-values were associated with the date, but not source, of collection (Figure 3.8).

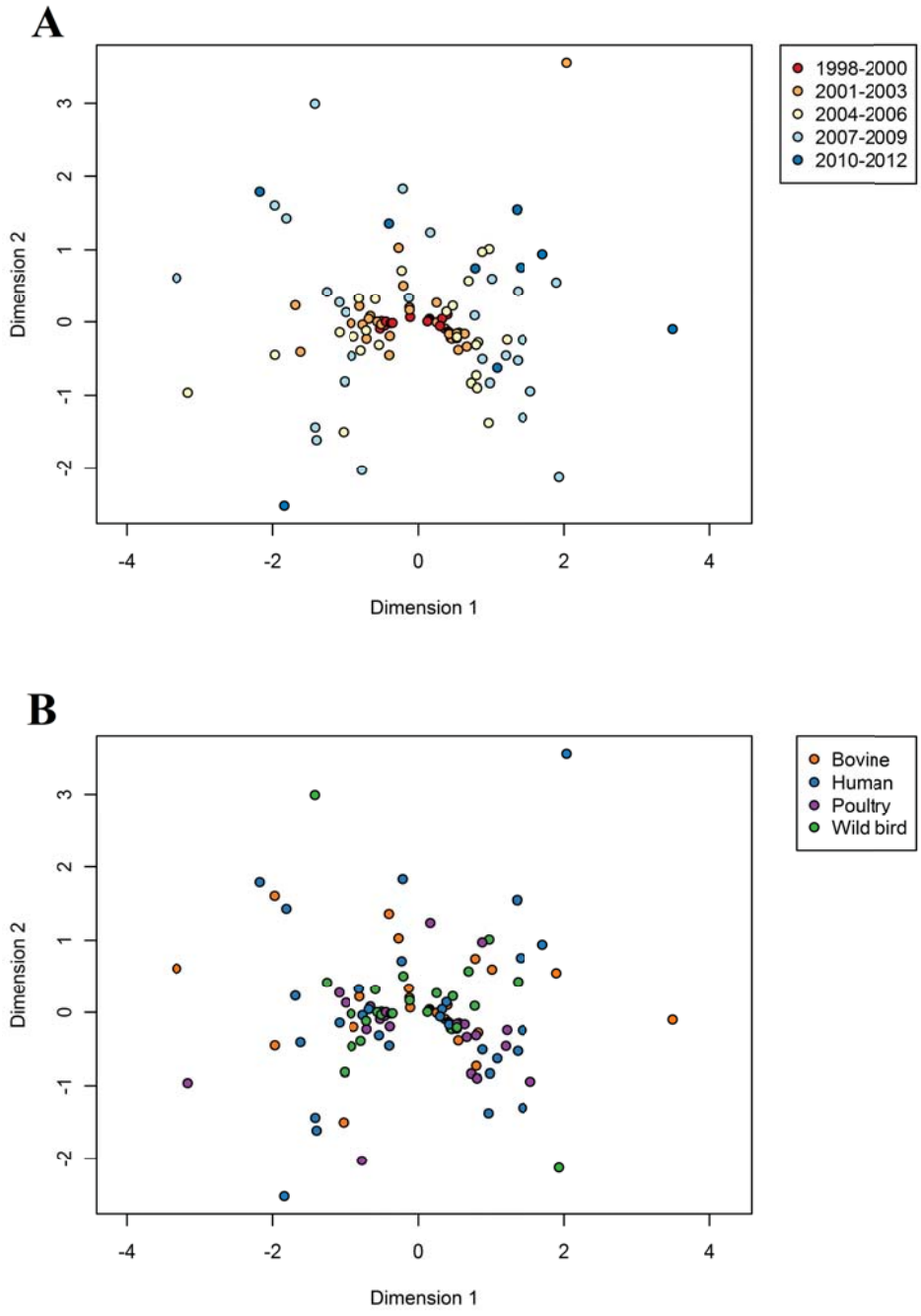


**Figure 3.7** Multi-dimensional scaling of 107 DT160 isolates, based on the presence of 684 protein differences and coloured by date of collection (A) and source (B).

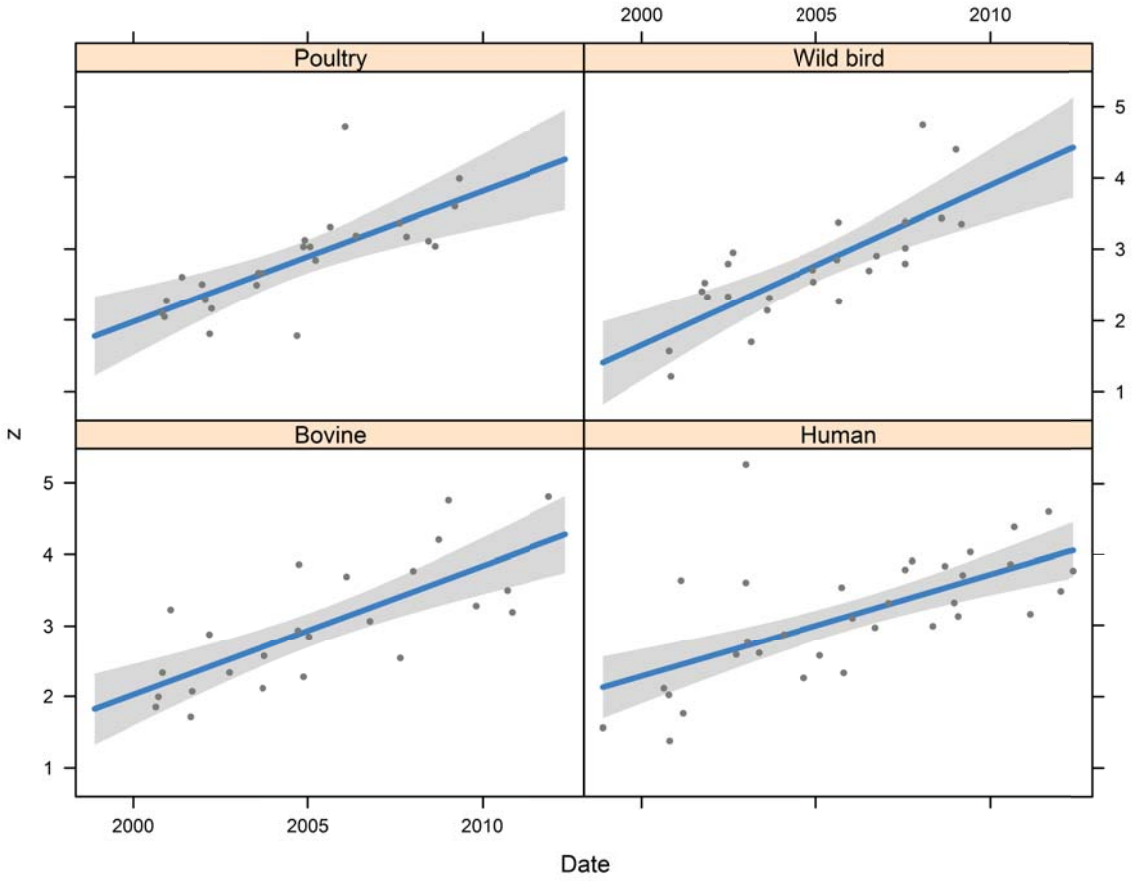


**Figure 3.8.** Scatterplots of year of collection versus z-values for 107 DT160 isolates collected during an outbreak in New Zealand, 1998-2012. Of the 107 isolates, 25 were from poultry (A), 25 from wild birds (B), 24 from bovines (C), and 33 from humans (D). Black lines represent the regression equation; blue represents the standard error for this equation. Date of collection was significantly associated with z-values in this model ($p<2 \times 10^{-16}$). There was insufficient evidence to suggest that source was associated with z-values ($p = 0.558$), and the interaction between source and date of collection was not significant ($p = 0.458$).

The 684 protein differences shared by the DT160 isolates were associated with a large number of COG functional groups. The proportion of proteins that contained sequence differences differed between functional groups ($p = 0.00002$). The proportions varied from 0.06 to 0.18, although most were between 0.09 and 0.13 (Figure 3.9). In addition, there was insufficient data to model the effects of source or date of collection on the number of protein differences associated with each group (Appendix A.2).

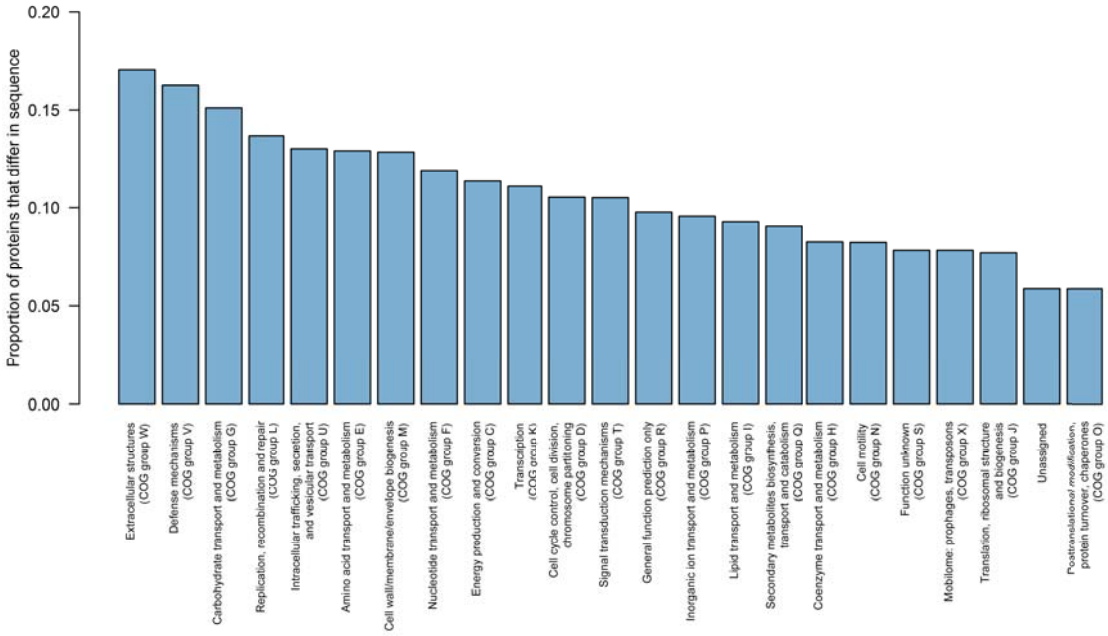


**Figure 3.9** Bar graph of the proportion of proteins shared by 107 DT160 isolates that differ in sequence for each COG functional group.

### 3.4.6 Ancestral migration between hosts

The discrete trait analysis model was used to predict ancestral migration of DT160 between the animal and human host groups, similar to Mather et al. (2013). However, no signal could be detected that could not be attributed to different sampling fractions in the host groups (Appendix A.3). Therefore, an alternate method, larger sample size, or both are required to predict these ancestral migrations.

### 3.4.7 DT160 antimicrobial susceptibilities

The 90 DT160 isolates published by Omar (2010) had very similar antimicrobial resistance profiles. No associations between these profiles and date of collection or source were found (Appendix A.4).

## 3.5 Discussion

In New Zealand, DT160 was first reported in Christchurch in 1998 from a human with salmonellosis (Thornley et al., 2003) (an isolate from this patient was included as part of this study). The New Zealand DT160 isolates analysed were estimated to share a common ancestor 0-2 years before this isolate was collected and were distinct from the UK isolates analysed, suggesting that DT160 was probably introduced into New Zealand

as a single incursion within this time period. Worldwide comparative studies are required to track DT160 migration and validate this hypothesis.

The mutation rate estimated for the DT160 outbreak was similar to rates reported by Mather et al. (2013) for an outbreak of *S.* Typhimurium DT104 in Scotland from 1990-2012 and by Okoro et al. (2012) for invasive *S.* Typhimurium strains in sub-Saharan Africa. The similarity of these mutation rates suggests consistency between outbreaks caused by *S.* Typhimurium and has implications for modeling the evolution of future outbreaks caused by this serovar.

In bacteriology, the effective population size is the number of bacteria that contribute to the next generation. The increase in the DT160 effective population size during 1998-2003 coincided with an increased prevalence of DT160 among human and non-human hosts during this time. However, the subsequent levelling-off of the effective DT160 population size was probably an artifact as the effective population size was calculated from the timing of coalescent events for randomly sampled bacteria (Minin et al., 2008), and as the outbreak proceeded, fewer coalescent points were available for estimation. Overall, phylogenetic analyses suggested that the DT160 population increased dramatically in the first few years following introduction. As the DT160 population increased, it acquired multiple SNPs, resulting in a progressive increase in diversity over time.

Identifying the source of a salmonellosis outbreak can be difficult because multiple potential sources must be considered (Gieraltowski et al., 2013). Probable sources of *Salmonella* can be identified by comparing isolates from infected humans with those from other human, non-human, and environmental sources (Byrne et al., 2014). No distinct DT160 clades associated with any one source were identified, suggesting that after its introduction into New Zealand, DT160 was transmitted between multiple hosts, resulting in large epidemics among humans and wild birds. It also suggests that humans obtained DT160 from multiple sources over the course of the outbreak. This finding is consistent with a case-control study performed by Thornley et al. (2003), which found that human DT160 patients were associated with multiple risk factors involving different sources: handling dead wild birds, contact with persons with diarrhoea, and consumption of fast-food.

Bacteria often adapt to new environments by altering (changing or losing) genes that are not essential for colonising that environment (Hottes et al., 2013). Gene loss can result in an increase in bacterial fitness, as fewer genes and processes need to be maintained within the bacteria (Koskiniemi et al., 2012). Multiple protein changes were identified amongst the New Zealand DT160 isolates, and these changes occurred in multiple COG functional groups as the epidemic progressed. However, no evidence of host group differentiation was found, suggesting that most of the evolution was due to random genetic drift rather than adaptive evolution.

Antibiotic usage selects for antimicrobial resistance (Wiuff et al., 2003). The lack of any associations between source groups, date of collection and antimicrobial resistance for 90 DT160 isolates suggests that antimicrobial usage in the source groups analysed was insufficient to select for DT160 resistance or DT160 was moving too quickly between these source groups and others for any antimicrobial usage to have an effect.

## 3.6 Conclusion

In this study the evolution and emergence of DT160 within New Zealand was described. Genomic analyses suggest that DT160 was introduced into New Zealand on a single occasion between 1996 and 1998, before propagating throughout the country and becoming more genetically diverse over time. In addition, DT160 isolates collected from human, poultry, bovine and wild bird sources were highly similar, indicating a large number of transmission episodes between these host groups.

# References


Alley, M. R., Connolly, J. H., Fenwick, S. G., Mackereth, G. F., Leyland, M. J., Rogers, L. E., Haycock, M., Nicol, C., and Reed, C. E. M. (2002). An epidemic of salmonellosis caused by *Salmonella* Typhimurium DT160 in wild birds and humans in New Zealand. *New Zealand Veterinary Journal*, 50(5):170–176.

Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, 62(1):245–253.

Aronesty, E. (2013). Comparison of sequencing utility programs. *Open Bioinformatics Journal*, 7(1):1–8.

Baker, M., Thornley, C., Lopez, L., Garrett, N., and Nicol, C. (2007). A recurring salmonellosis epidemic in New Zealand linked to contact with sheep. *Epidemiology and Infection*, 135(1):76–83.

Byrne, L., Fisher, I., Peters, T., Mather, A., Thomson, N., Rosner, B., Bernard, H., McKeown, P., Cormican, M., Cowden, J., Aiyedun, V., Lane, C., and Team, I. O. C. (2014). A multi-country outbreak of *Salmonella* Newport gastroenteritis in Europe associated with watermelon from Brazil, confirmed by whole genome sequencing: October 2011 to January 2012. *Euro surveillance : bulletin Européen sur les maladies transmissibles = European communicable disease bulletin*, 19(31):6–13.

Clarke, K. R. and Gorley, R. N. (2006). *PRIMER v6:User manual/tutorial, pages 1–192.* PRIMER-E, Plymouth, United Kingdom.

Cox, M. P., Peterson, D. A., and Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*, 11(485):1–6.

Drummond, A., Suchard, M., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8):1969–1973.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5):699–710.

ESR (2003). Lablink Annual Summaries, 2000-2003. Retrieved 2016-09-27, from: `https://surv.esr.cri.nz/PDF_surveillance/Lablink/`, 1–1.

ESR (2015). Notifiable diseases. Retrieved 2017-03-02, from: http://www.nzpho.org.nz/NotifiableDisease.aspx, 1–1.

Gardner, S. N., Slezak, T., and Hall, B. G. (2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics*, 31:2877–2878.

Garrison, E. and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. Retrieved 2017-02-02, from: https://arxiv.org/abs/1207.3907, 1–1.

Gieraltowski, L., Julian, E., Pringle, J., MacDonald, K., Quilliam, D., Marsden-Haug, N., Saathoff-Huber, L., Von Stein, D., Kissler, B., Parish, M., Elder, D., Howard-King, V., Besser, J., Sodha, S., Loharikar, A., Dalton, S., Williams, I., and Behravesh, C. (2013). Nationwide outbreak of *Salmonella* Montevideo infections associated with contaminated imported black and red pepper: Warehouse membership cards provide critical clues to identify the source. *Epidemiology and Infection*, 141(6):1244–1252.

Grillo, T. and Post, L. (2010). *Salmonella* Typhimurium DT160 outbreak in Tasmania. *Animal Health Surveillance Quarterly Reports*, 14(4):8–8.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8):1072–1075.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2):160–174.

He, Z. L., Zhang, H. K., Gao, S. H., Lercher, M. J., Chen, W. H., and Hu, S. N. (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Research*, 44:W236–W241.

Hottes, A., Freddolino, P., Khare, A., Donnell, Z., Liu, J., and Tavazoie, S. (2013). Bacterial adaptation through loss of function. *PLoS Genetics*, 9(7):1–13.

Huson, D. and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2):254–267.

Inouye, M., Dashnow, H., Raven, L., Schultz, M., Pope, B., Tomita, T., Zobel, J., and Holt, K. (2014). SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Medicine*, 6(11):1–16.

King, N., Lake, R., and Campbell, D. (2011). Source attribution of non-typhoid salmonellosis in New Zealand using outbreak surveillance data. *Journal of Food Protection*, 74:438–445.

Koskiniemi, S., Sun, S., Berg, O., and Andersson, D. (2012). Selection-driven gene loss in bacteria. *PLoS Genetics*, 8(6):1–7.

Lal, A., Baker, M., French, N., Dufour, M., and Hales, S. (2012). The epidemiology of human salmonellosis in New Zealand, 1997-2008. *Epidemiology and Infection*, 140(9):1685–1694.

Lal, A., Ikeda, T., French, N., Baker, M., and Hales, S. (2013). Climate variability, weather and enteric disease incidence in New Zealand: Time series analysis. *PLoS ONE*, 8(12):1–11.

Lawson, B. b., Howard, T., Kirkwood, J., MacGregor, S., Perkins, M., Robinson, R., Ward, L., and Cunningham, A. (2010). Epidemiology of salmonellosis in garden birds in England and Wales, 1993 to 2003. *EcoHealth*, 7(3):294–306.

Lemey, P., Rambaut, A., Drummond, A., and Suchard, M. (2009). Bayesian phylogeography finds its roots. *PLoS Computational Biology*, 5(9):1–16.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21):2987–2993.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754–1760.

Majowicz, S., Musto, J., Scallan, E., Angulo, F., Kirk, M., O'Brien, S., Jones, T., Fazil, A., and Hoekstra, R. (2010). The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clinical Infectious Diseases*, 50(6):882–889.

Mather, A., Reid, S., Maskell, D., Parkhill, J., Fookes, M., Harris, S., Brown, D., Coia, J., Mulvey, M., Gilmour, M. o., Petrovska, L., De Pinna, E., Kuroda, M., Akiba, M., Izumiya, H., Connor, T., Suchard, M. l., Lemey, P., Mellor, D., Haydon, D., and Thomson, N. (2013). Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science*, 341(6153):1514–1517.

Minin, V., Bloomquist, E., and Suchard, M. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, 25(7):1459–1471.

Okoro, C. K., Kingsley, R. A., Connor, T. R., Harris, S. R., Parry, C. M., Al-Mashhadani, M. N., Kariuki, S., Msefula, C. L., Gordon, M. A., De Pinna, E., Wain, J., Heyderman, R. S., Obaro, S., Alonso, P. L., Mandomando, I., MacLennan, C. A., Tapia, M. D., Levine, M. M., Tennant, S. M., Parkhill, J., and Dougan, G. (2012). Intracontinental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. *Nature Genetics*, 44(11):1215–1221.

Omar, S. (2010). *Molecular epidemiology of Salmonella Typhimurium DT160 in New Zealand*. PhD thesis, Massey University.

Page, A., Cummins, C., Hunt, M., Wong, V., Reuter, S., Holden, M., Fookes, M., Falush, D., Keane, J., and Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31(22):3691–3693.

Penfold, J. B., Amery, H. C., and Peet, P. J. (1979). Gastroenteritis associated with wild birds in a hospital kitchen. *British medical journal*, 2(6193):802.

Petrovska, L., Mather, A. E., Abuoun, M., Branchu, P., Harris, S. R., Connor, T., Hopkins, K. L., Underwood, A., Lettini, A. A., Page, A., Bagnall, M., Wain, J., Parkhill, J., Dougan, G., Davies, R., and Kingsley, R. A. (2016). Microevolution of monophasic *Salmonella* Typhimurium during epidemic, United Kingdom, 20052010. *Emerging Infectious Diseases*, 22(4):617–624.

Piccirillo, A., Mazzariol, S., Caliari, D., and Menandro, M. L. (2010). *Salmonella* Typhimurium phage type DT160 infection in two Moluccan cockatoos (*Cacatua moluccensis*): clinical presentation and pathology. *Avian diseases*, 54(1):131–135.

Popoff, M. Y. and Le Minor, L. E. (2005). Genus XXXIII. *Salmonella*. In Brenner, D. J. and Staley, J. T., editors, *Bergey's Manual of Systematic Bacteriology*, pages 764–799. Springer, New York.

Public Health Surveillance (2017a). 2003-2016 Human Salmonella serotypes. Retrieved 2017-05-07, from: `https://surv.esr.cri.nz/enteric_reference/human_salmonella.php`, 1–1.

Public Health Surveillance (2017b). Non-human Salmonella isolates, 2003-2016. Retrieved 2017-05-07, from: `https://surv.esr.cri.nz/enteric_reference/nonhuman_salmonella.php`, 1–1.

Qiagen (2012). QIAamp DNA mini and blood mini handbook. Retrieved 2017-06-29, from: `\nhttps://www.qiagen.com/us/resources/resourcedetail?id=62a200d6-faf4-469b-b50f-2b59cf738962&lang=en`, 33-56.

Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014). Tracer 1.6. Retrieved 2016-09-27, from: `http://beast.bio.ed.ac.uk/Tracer`, 1–1.

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30(14):2068–2069.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9):1312–1313.

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30):11030–11035.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12):2725–2729.

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28(1):33–36.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). ClustalW - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673–4680.

Thornley, C., Simmons, G., Callaghan, M., Nicol, C., Baker, M., Gilmore, K., and Garrett, N. (2003). First incursion of *Salmonella enterica* serotype Typhimurium DT160 into New Zealand. *Emerging Infectious Diseases*, 9(4):493–495.

Tizard, I. R., Fish, N. A., and Harmeson, J. (1979). Free flying sparrows as carriers of salmonellosis. *Canadian Veterinary Journal*, 20(5):143–144.

Wiuff, C., Lykkesfeldt, J., Svendsen, O., and Aarestrup, F. M. (2003). The effects of oral and intramuscular administration and dose escalation of enrofloxacin on the selection of quinolone resistance among *Salmonella* and coliforms in pigs. *Research in Veterinary Science*, 75(3):185–193.

Zerbino, D. R. and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18(5):821–829.

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

# STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

**Name of Candidate:** Samuel Bloomfield

**Name/Title of Principal Supervisor:** Dr Jackie Benschop

**Name of Published Research Output and full reference:**

Emerging Infectious Diseases: Bloomfield, S.J., Benschop J., Biggs P.J., Marshall J.C., Hayman, D.T.S., Carter, P.E., Midwinter, A.C., Mather, A.E., and French, N.P. Genomic analysis of Salmonella enterica serovar Typhimurium DT160 associated with a 14-year outbreak, New Zealand, 1998-2012. Emerging Infectious Diseases, 2017; 23(6):906–913

**In which Chapter is the Published Work:** 3

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate 85%

  and / or

- Describe the contribution that the candidate has made to the Published Work:

  Samuel prepared the samples for sequencing, analysed the genomic data, wrote most of the scripts used to analyse the data, wrote and edited the manuscript, and prepared the manuscript for publication.

Samuel Bloomfield — Digitally signed by Samuel Bloomfield Date: 2017.07.17 10:28:26 +12'00'

Candidate's Signature — 17/07/2017 Date

Jackie Benschop — Digitally signed by Jackie Benschop Date: 2017.07.20 09:11:24 +12'00'

Principal Supervisor's signature — 17/07/2017 Date

# Chapter 4

# Investigation of the validity of two Bayesian ancestral state reconstruction models for estimating *Salmonella* transmission during outbreaks

## 4.1 Abstract

Ancestral state reconstruction models use genetic data to estimate the ancestral states of organisms. They have been applied to salmonellosis outbreaks to estimate the number of transmissions between different animals that share similar geographical locations, with animal host as the state. However, no studies have validated these models for outbreak analysis. In this study, salmonellosis outbreaks were simulated using a stochastic Susceptible-Infected-Recovered model, and the host population and transmission parameters of these simulated outbreaks were estimated using Bayesian ancestral state reconstruction models (discrete trait analysis (DTA) and structured coalescent (SC)). These models were unable to accurately estimate the number of transmissions between the host populations or the amount of time spent in each host population. The DTA model was inaccurate because it assumed the number of isolates sampled from each host population was proportional to the number of individuals infected within each host population. The SC model was inaccurate possibly because it assumed that each host population's effective population size was constant over the course of the simulated outbreaks. This study highlights the need for outbreak-specific phylodynamic models that can take into consideration factors that influence the behaviour of outbreaks, e.g. changing effective population sizes, variation in infectious periods, intra-population transmissions, and disproportionate sampling.

## 4.2 Introduction

Ancestral state reconstruction models estimate the ancestral states of organisms based on their evolutionary history. Outbreaks are "...the occurrence of disease in excess of what would normally be expected in a defined community, geographical area or season" (WHO, 2016). Ancestral state reconstruction models have

been used to investigate the transmission of infectious agents between animal populations over the course of outbreaks, with host population as the state (Mather et al., 2013). However, no studies have validated these models for this type of analysis.

The discrete trait analysis (DTA) and structured coalescent (SC) models are ancestral state reconstruction models. A trait is "...a characteristic feature or quality distinguishing a particular person or thing" (Collins, 2017). Both models treat each host population as a discrete trait and can be approximated using Markov chains Monte Carlo (Lemey et al., 2009; Vaughan et al., 2014).

There are many differences between the two ancestral state reconstruction models. The DTA model uses a substitution model to model the transmission between host populations (Lemey et al., 2009). The pruning method (Felsenstein, 1981) is often used in phylogenetic analysis to integrate all possible mutation histories to be computationally efficient, but the DTA model uses the method to integrate all possible migration histories (De Maio et al., 2015). The SC model assumes that each host population is evolving under a Wright-Fisher model (Wright, 1931) and each individual in time can remain in its current host population or migrate to another at a fixed rate (Vaughan et al., 2014). In bacteriology, the effective population size is the number of bacteria that contribute to the next generation. The DTA model can allow for variation in the overall effective population size over time, whilst the SC model assumes that each host population has a constant effective population size. The DTA model assumes that the number of offspring an individual is likely to produce is independent of its host population, whilst the SC model allows for variation between host populations (Vaughan et al., 2014). The DTA model assumes that the proportion of isolates sampled from each host population is proportional to the number of infected individuals within that host population, whilst the SC model allows for variation (De Maio et al., 2015). Some of these assumptions are applicable to the investigation of outbreaks (e.g. varying effective population size), whilst others are not (e.g. isolate proportionality).

Salmonellosis is an intestinal infection caused by non-typhoidal *Salmonella* strains. Salmonellosis outbreaks vary in size and can involve one or more host populations (Gould et al., 2013). Identifying the amount of time *Salmonella* spends in a host population over an outbreak and the amount of transmission between host populations can inform control measures to limit salmonellosis outbreaks, e.g. if human cases are primarily from exposure to poultry sources than control measures that limit human exposure to poultry or decrease the amount of *Salmonella* in poultry may be beneficial. However, it is often difficult to distinguish between isolates collected from different host populations and there is growing evidence that exposure to human sources contributes more to salmonellosis outbreaks than previously thought (Wikström et al., 2014). Therefore, methods and models are required that can approximate the number of cases that are the result of exposure to animal and/or human sources. The aim of this study was to use simulated outbreaks to investigate whether the DTA or SC models could be applied to infer transmission dynamics in outbreaks involving multiple hosts, motivated by non-typhoidal *Salmonella*.

## 4.3 Methods

### 4.3.1 Outbreak simulations

The MASTER package (Vaughan and Drummond, 2013) in BEAST2 (Bouckaert et al., 2014) can simulate structured and unstructured populations, and inheritance trees and networks for these population. It was used to simulate salmonellosis outbreaks using a Gillespie direct method stochastic Susceptible-Infected-Recovered (SIR) model. In this model, susceptible individuals become infectious by exposure to other infected individuals:

$$S_i + I_i \xrightarrow{\beta_{ii}} 2I_i \tag{1}$$

$$S_i + I_j \xrightarrow{\beta_{ji}} I_i + I_j \tag{2}$$

Equation 1 represents the transmission of the infectious agent from an infected individual to a susceptible individual of the same host population. Equation 2 represents the transmission of the infectious agent from an infected individual to a susceptible individual of another host population. Here, $S_i$ represents a susceptible individual from one host population, $I_i$ represents an infectious individual from the same host population, $I_j$ represents an infectious individual from another host population, and $\beta_{ii}$ and $\beta_{ji}$ represents the infectious rate per susceptible individual per infectious individual.

In this model, infectious individuals also recover or are removed over time:

$$I_i \xrightarrow{\gamma_i} R_i \tag{3}$$

Equation 3 represents the infectious period for an infectious individual. Here, $I_i$ represents an infectious individual in one host population, $R_i$ represents a recovered/removed individual in the same host population, and $\gamma_i$ represents the recovery/removal rate per infectious individual for this host population.

### 4.3.2 Simulated outbreaks

Due to limited computational time, 23 salmonellosis outbreaks were simulated using the MASTER package, hereinafter 'outbreak simulations'. This created a transmission tree consisting of all the transmissions that took place over the course of the outbreak. These simulations consisted of two host populations: human and animal. The initial susceptible host population size, $\gamma$ and $\beta$ values were allowed to vary between these simulations, but represented possible values for salmonellosis outbreaks in New Zealand (Appendix B.1).

### 4.3.3 Multiple variable simulations

One hundred isolates were randomly sampled from each outbreak simulation, after stratifying for host population, using an in-house Perl script, hereinafter 'random sampling'. For each outbreak simulation, the transmission tree was simplified to only include nodes common to the 100 isolates using an in-house Perl script. The sampled transmission trees were treated as phylogenetic trees and used to simulate genetic data for the simulated outbreaks using the MASTER package, hereinafter 'sequence simulations'. 800 SNPs were simulated in total for the 100 isolates, similar to the 793 core SNPs shared by 109 *Salmonella enterica* serovar Typhimurium DT160 isolates during an extended outbreak in New Zealand (Chapter 3). In-house Perl and R scripts were used to analyse the sampled transmission tree and to calculate the amount of time spent in each host population and quantify the number of transmissions.

### 4.3.4 DTA model

For the DTA model, the 800 SNPs were imported into BEAUti 1.8.3 to create an XML file for BEAST 1.8.3 (Drummond et al., 2012). The generalised time reversible (GTR) model was used to model base substitutions (Tavare, 1986), the Gaussian Markov random field (GMRF) Bayesian skyride model was used to allow for changes in the effective population size (Minin et al., 2008), and a strict molecular clock was used to estimate the mutation rate, which was calibrated by the tip date. The XML file was run in BEAST for 10 million steps as a single run with a 10% burn-in.

### 4.3.5 SC model

For the SC model, the 800 SNPs were imported into BEAUti 2.4 with the MultiTypeTree package (Vaughan et al., 2014) to create an XML file for BEAST 2.4 (Bouckaert et al., 2014). The GTR model was used to model base substitution and a strict molecular clock was used to estimate the mutation rate, which was calibrated by the tip date. The XML file was run in BEAST for 250 million steps as a single run with a 10% burn-in. The SC model was run for a larger number of steps than the DTA model as its population and transmission parameters took longer to converge.

### 4.3.6 Model comparison

The SC and DTA models were used to estimate the amount of time spent in each host population (population parameters) and the amount of transmissions between the host populations (transmission parameters). However, the models' raw outputs were not directly comparable as the SC model allowed transmissions along

branches, whilst the DTA could not. Therefore, the relative amount of time (i.e. proportion) spent in each host population and the relative number of inter-population transmissions made up of each transmission were compared. The two models' abilities were compared using four parameters:

1. The proportion of outbreak simulations that a model included the known parameter within their 95% highest posterior density (HPD) intervals.
2. The mean squared error between a known parameter and a model's mean estimates.
3. The mean size of a model's 95% HPD intervals.
4. The correlation coefficient between a known parameter and a model's mean estimates.

#### 4.3.7 Model consistency

To investigate variation in model estimates between different samples (i.e. model consistency), one of the simulated outbreaks was randomly sampled 10 times after stratifying for host population. For each sample, sequence simulations were used to create genetic data, which was modelled using the SC and DTA models.

#### 4.3.8 Disproportionate sampling

To investigate the effect of the relative number of isolates from each source on model estimates (i.e. disproportionate sampling), as expected during the outbreaks, one of the simulated outbreaks was randomly sampled 10 times with different numbers of animal and human isolates. For each sample, 100 isolates were analysed, but the proportion of isolates that were from each host population were systematically ranged from 5-95% in 10% intervals. For each sample, sequence simulations were used to create genetic data, which was modelled using the SC and DTA models.

#### 4.3.9 Equal-time sampling

To investigate an alternative sampling method, 'equal-time sampling', an in-house Perl script was used to stratify the isolates from the initial 23 simulated outbreaks by host population, before randomly sampling an equal number of isolates from each year of the simulated outbreaks, to a total of 100 isolates. Sequence simulations were used to create genetic data for the samples, which were modelled using the SC and DTA models.

#### 4.3.10 Equal intra-population transmission and infectious periods

To investigate if different intra-population transmission rates and infectious periods had any effect on model estimates, twelve additional outbreaks were simulated with equal intra-population transmission rates and infectious periods between host populations, but inter-population transmission rates and initial susceptible host population sizes that varied. For each simulation, 100 isolates were sampled using random sampling, and sequence simulations were used to create genetic data, which was modelled using the SC and DTA models.

#### 4.3.11 DT160 outbreak

The DTA and SC models were used to analyse a previously-described salmonellosis outbreak in New Zealand. The outbreak was caused by *Salmonella enterica* serovar Typhimurium DT160 (Chapter 3). 109 DT160 isolates from animal (n=74) and human (n=35) host populations over 14 years were investigated using the 793 core SNPs they shared. The simulated outbreaks were in part based on this DT160 outbreak.

#### 4.3.12 Scripts

The in-house scripts used in this study are available from GitHub (https://github.com/samuelbloomfield/Scripts-for-outbreak-simulations).

## 4.4 Results

#### 4.4.1 Model consistency

There was some variation in the DTA and SC models' population and transmission mean estimates for the same simulated outbreak that was randomly sampled ten times (Figures 4.1 and 4.2). The SC model's 95% HPD intervals predicted known population parameters more frequently, whilst the DTA model's 95% HPD intervals predicted known transmission parameters more frequently. There was some variation in the sampled transmission trees, resulting in slight differences between the known transmission and population parameters for the same sampled outbreak.

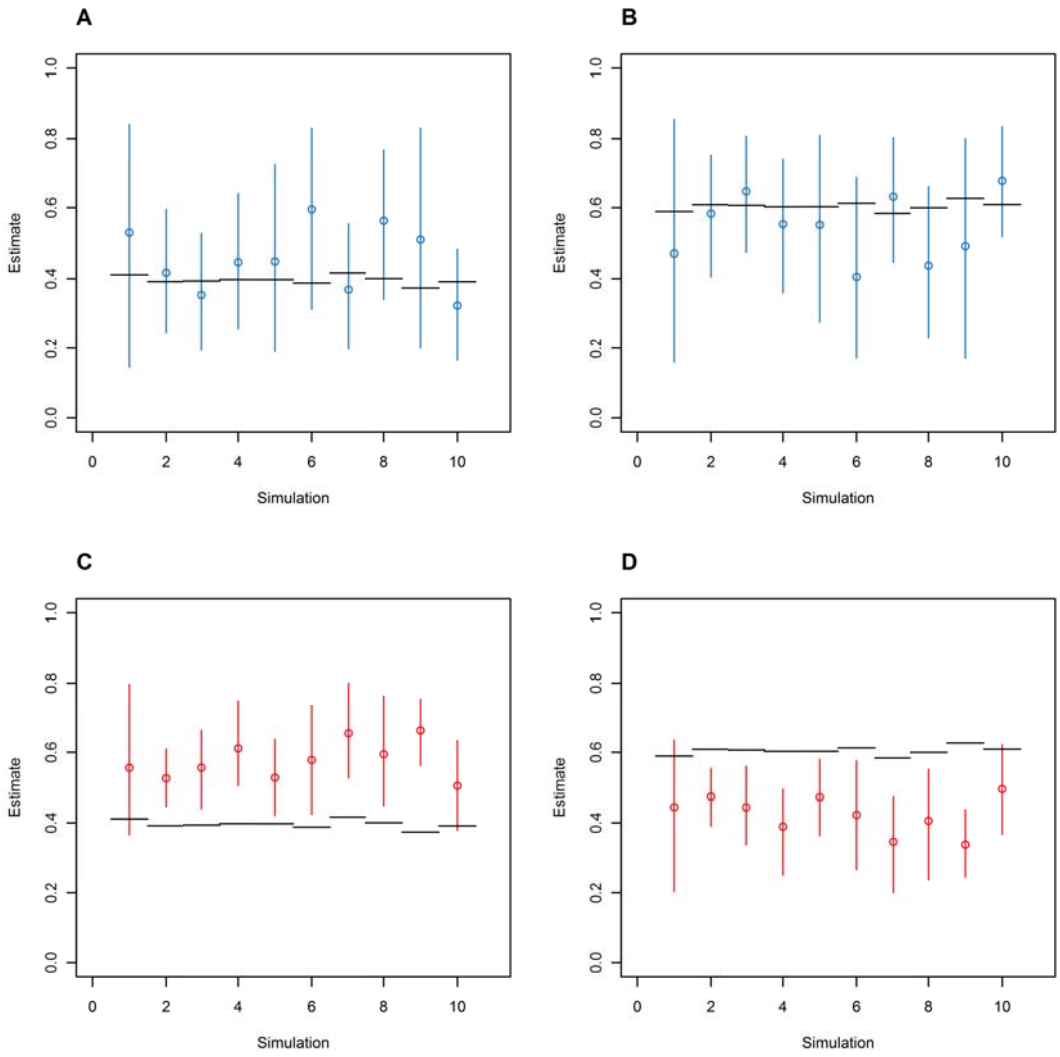


**Figure 4.1.** Scatterplots of the proportion of time spent in the animal (A and C) and human (B and D) host populations as estimated by the SC (blue: A and B) and DTA (red: C and D) models, for 10 random samples of the same simulated outbreak. The horizontal lines represent the parameters for the sampled outbreaks, the circles represent the mean and the error bars represent the 95% HPD interval.

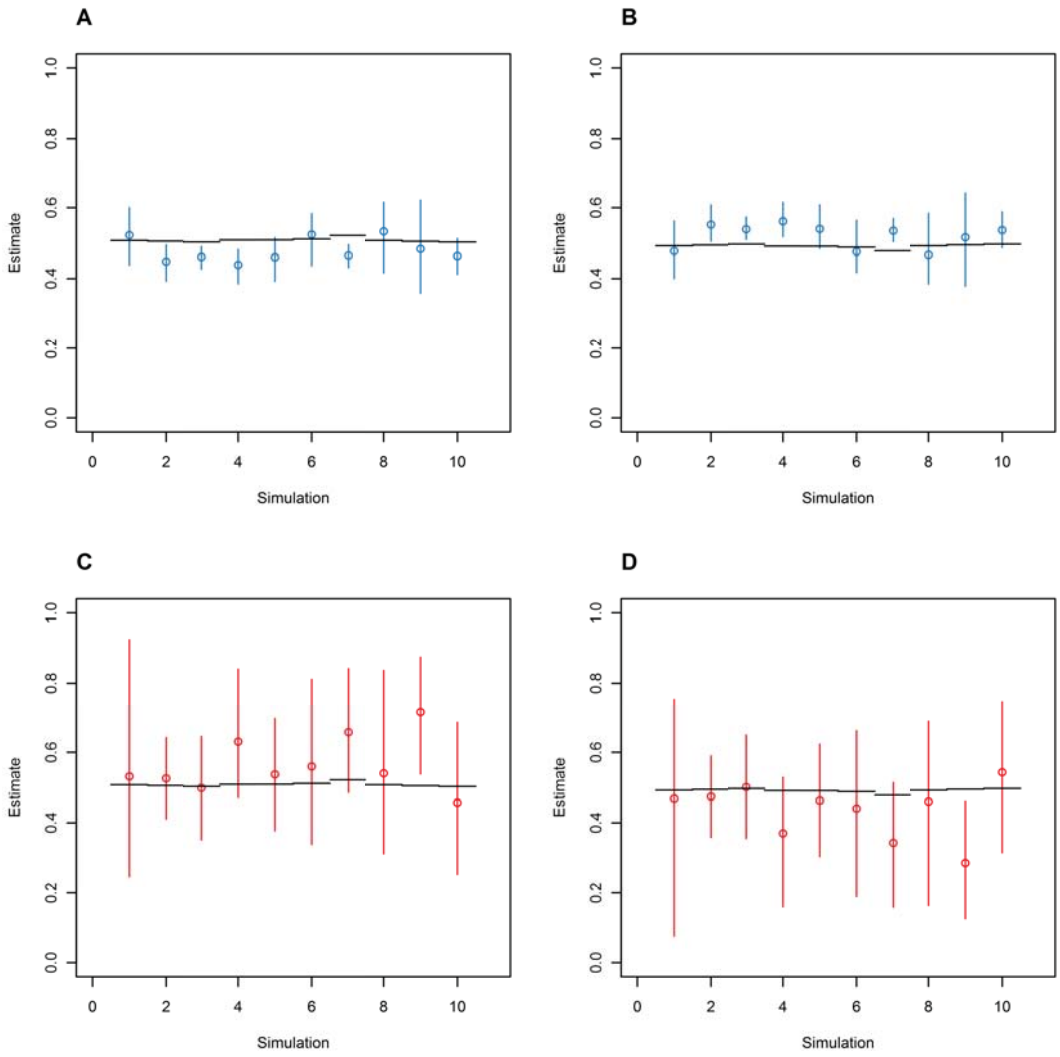


**Figure 4.2.** Scatterplots of the proportion of inter-population transmissions made up of animal-to-human (A and C) and human-to-animal (B and D) transmissions as estimated by the SC (blue: A and B) and DTA (red: C and D) models, for 10 random samples of the same simulated outbreak. The horizontal lines represent the parameters for the sampled outbreaks, the circles represent the mean and the error bars represent the 95% HPD interval.

### 4.4.2 Disproportionate sampling

The DTA and SC models responded to variation in sample proportions for the same simulated outbreak differently. The DTA model's estimates showed a much stronger positive correlation with the proportion of isolates sampled from each host population than the SC models's mean estimates (Table 4.1). The DTA model's mean estimates displayed a sigmoid-like association with the proportion of isolates sampled from each host population. (Figures 4.3 and 4.4).

**Table 4.1** Correlation coefficients of the SC and DTA models' mean estimates and the proportion of isolates sampled from each host population for a simulated outbreak that was disproportionately sampled.

| | Population | | Transmissions | |
|---|---|---|---|---|
| | SC | DTA | SC | DTA |
| Correlation coefficient | 0.593 | 0.983 | 0.400 | 0.964 |

Population, the proportion of time spent in a host population; Transmission, the proportion of inter-population transmissions; Correlation coefficient, the correlation coefficient between the proportion of samples from each host population and the model's mean estimates

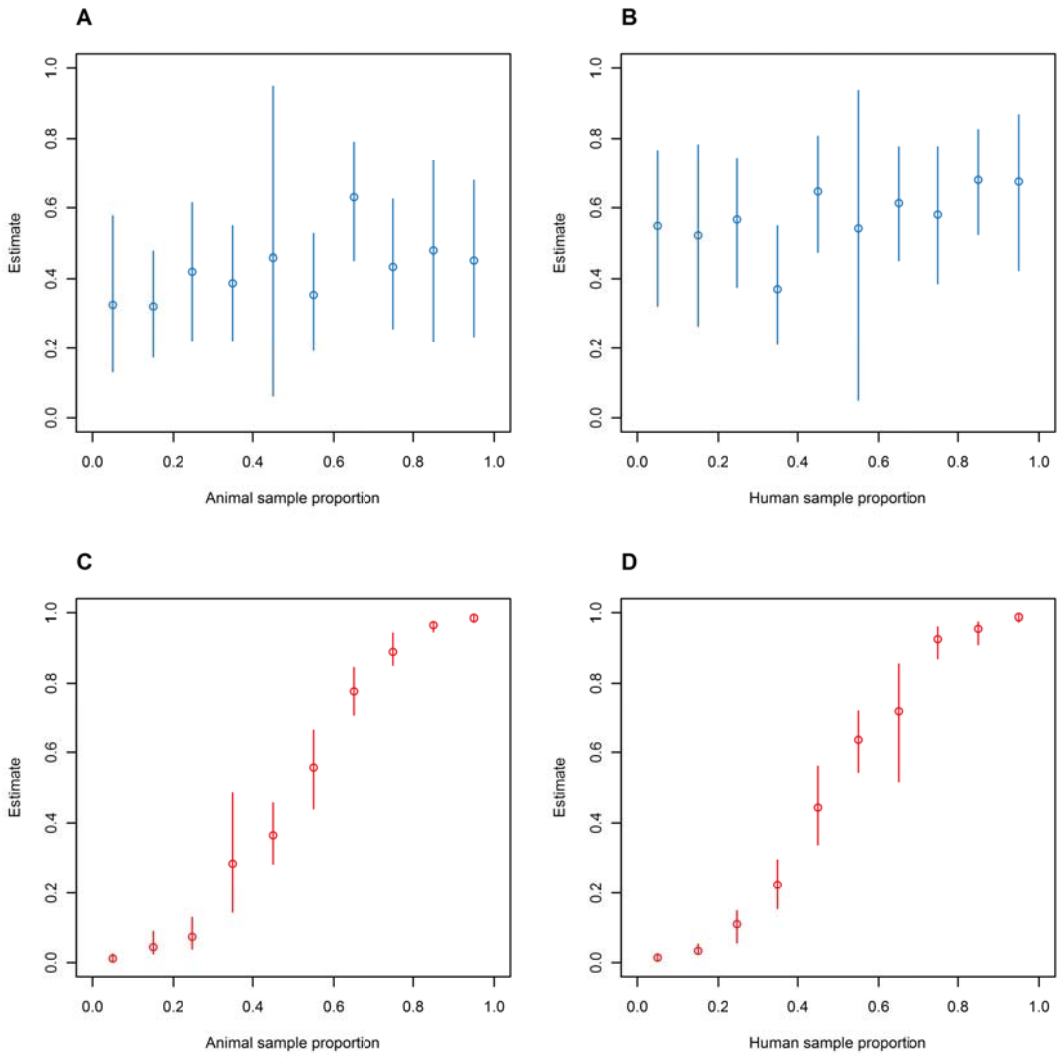


**Figure 4.3.** Scatterplots of the proportion of time spent in the animal (A and C) and human (B and D) host populations as estimated by the SC (blue: A and B) and DTA (red: C and D) models versus the proportion of sampled isolates that are animal (A and C) and human (B and D) for the same outbreak. The circles represent the mean and the error bars represent the 95% HPD interval.

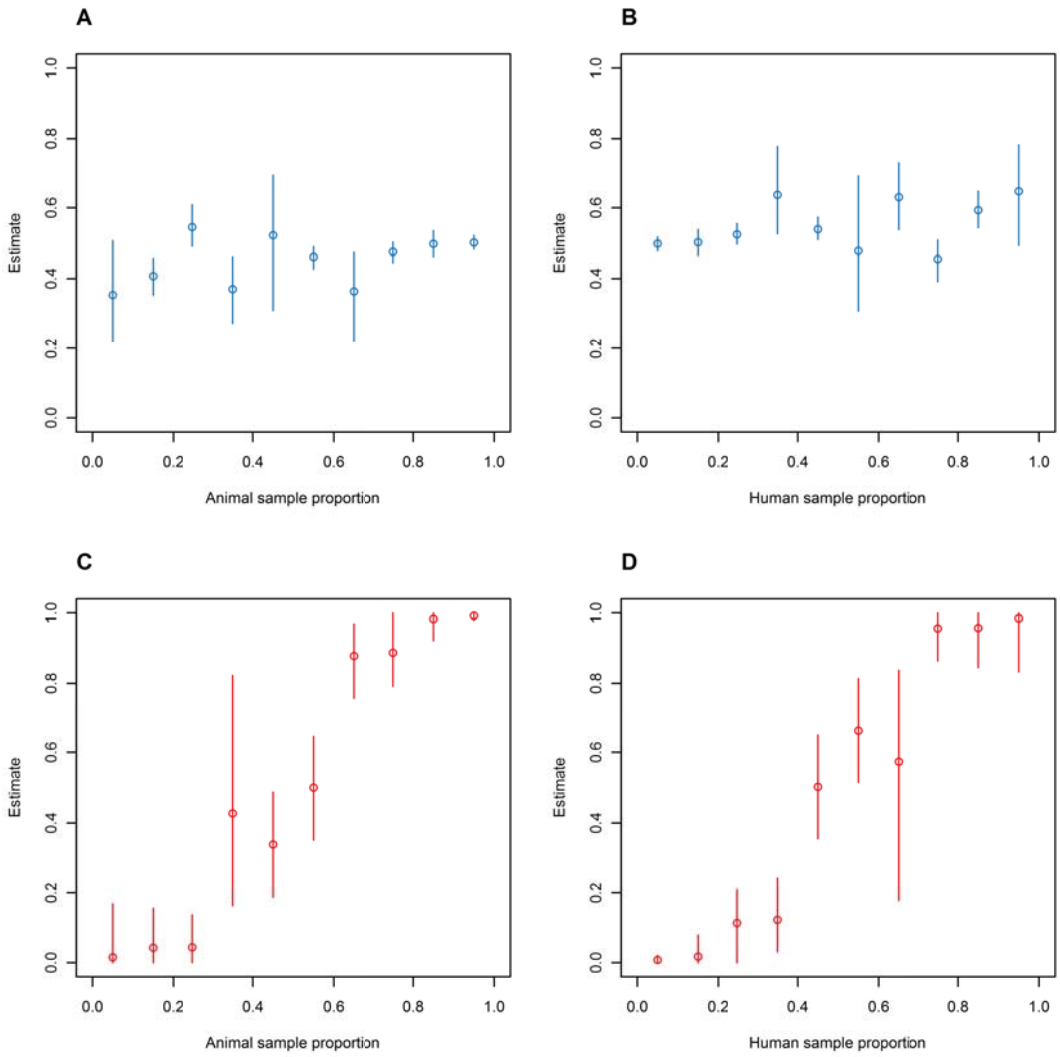


**Figure 4.4.** Scatterplots of the proportion of inter-population transmissions made up of animal-to-human (A and C) and human-to-animal (B and D) transmissions as estimated by the SC (blue: A and B) and DTA (red: C and D) models versus the proportion of isolates that are animal (A and C) and human (B and D) for the same outbreak. The circles represent the mean and the error bars represent the 95% HPD interval.

### 4.4.3 Multiple variable simulations

The DTA and SC models showed different estimate patterns for the 23 simulated outbreaks. The SC model predicted a larger proportion of known population and transmission parameters within its 95% HPD interval compared to the DTA model (Table 4.2). However, its mean 95% HPD interval sizes were larger and the DTA model's mean estimates showed a stronger positive correlation with the known paramater values than the SC model's mean estimates. Both models had similar mean squared errors between the known parameters and the models' mean estimates. However, the SC model's mean population estimates were all within the 0.2-0.8 interval and its mean transmission rates were all within the 0.35-0.65 interval, whilst the DTA model made mean estimates outside of these ranges (Figures 4.5 and 4.6).

**Table 4.2** Summary statistics of the SC and DTA models' estimates compared to known parameters for 23 randomly-sampled simulated outbreaks.

| | Population | | Transmissions | |
|---|---|---|---|---|
| | SC | DTA | SC | DTA |
| Estimate accuracy | 0.609 | 0.217 | 0.565 | 0.348 |
| Mean squared error | 0.127 | 0.164 | 0.052 | 0.099 |
| 95% HPD size | 0.609 | 0.168 | 0.233 | 0.267 |
| Correlation coefficient | 0.377 | 0.593 | 0.232 | 0.637 |

Population, the proportion of time spent in a host population; Transmission, the proportion of inter-population transmissions; Estimate accuracy, the proportion of outbreak simulations that a model included the known parameter within their 95% HPD intervals; Mean squared error, the mean squared error between the known parameters and a model's mean estimates; 95% HPD size, the mean 95% HPD interval size; Correlation coefficient, the correlation coefficient between the known parameter and a model's mean estimates

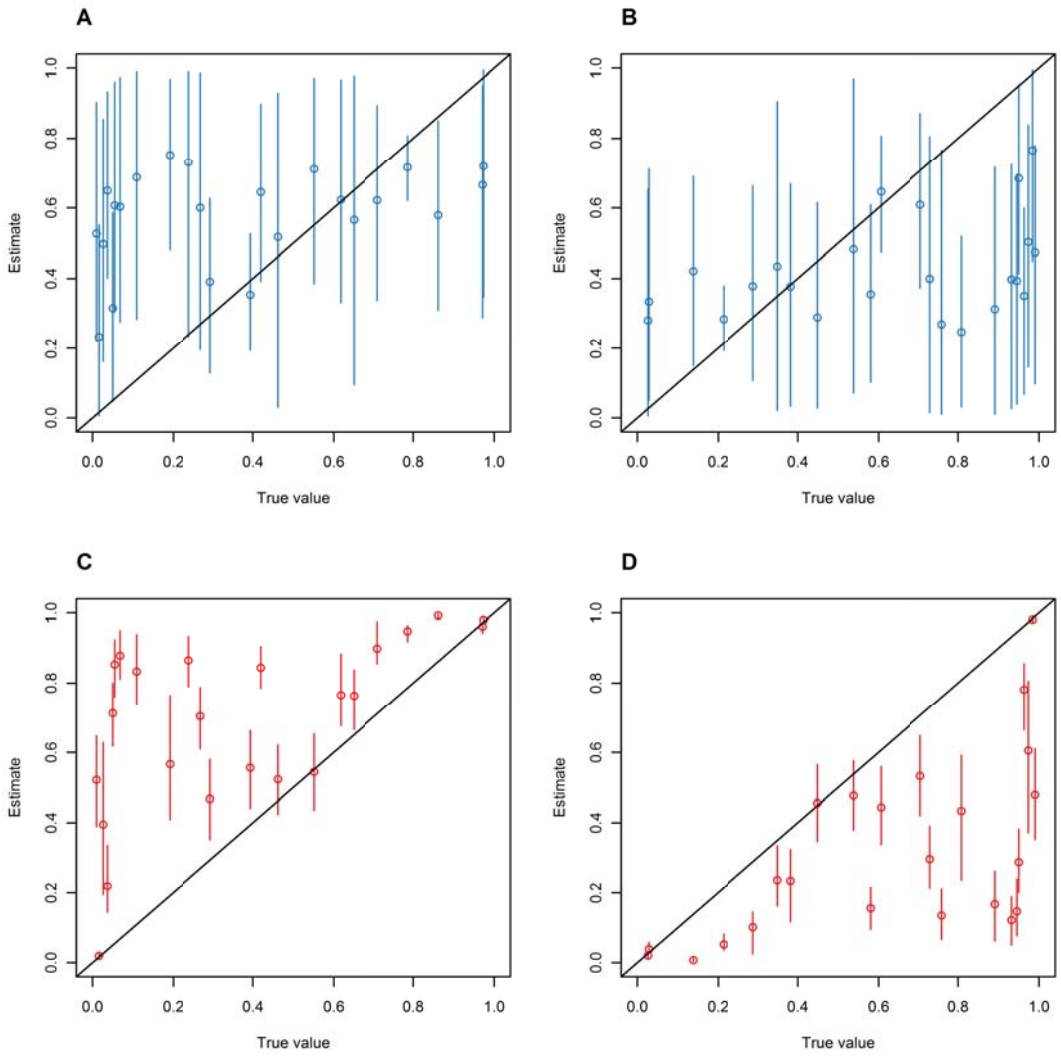


**Figure 4.5.** Scatterplots of the proportion of time spent in the animal (A and C) and human (B and D) host populations, versus the values estimated by the SC (blue: A and B) and DTA (red: C and D) models for 23 simulated outbreaks that were randomly sampled. The diagonal line represents accurate estimates of the sampled outbreaks, the dots represent the mean, and the error bars represent the 95% HPD interval.

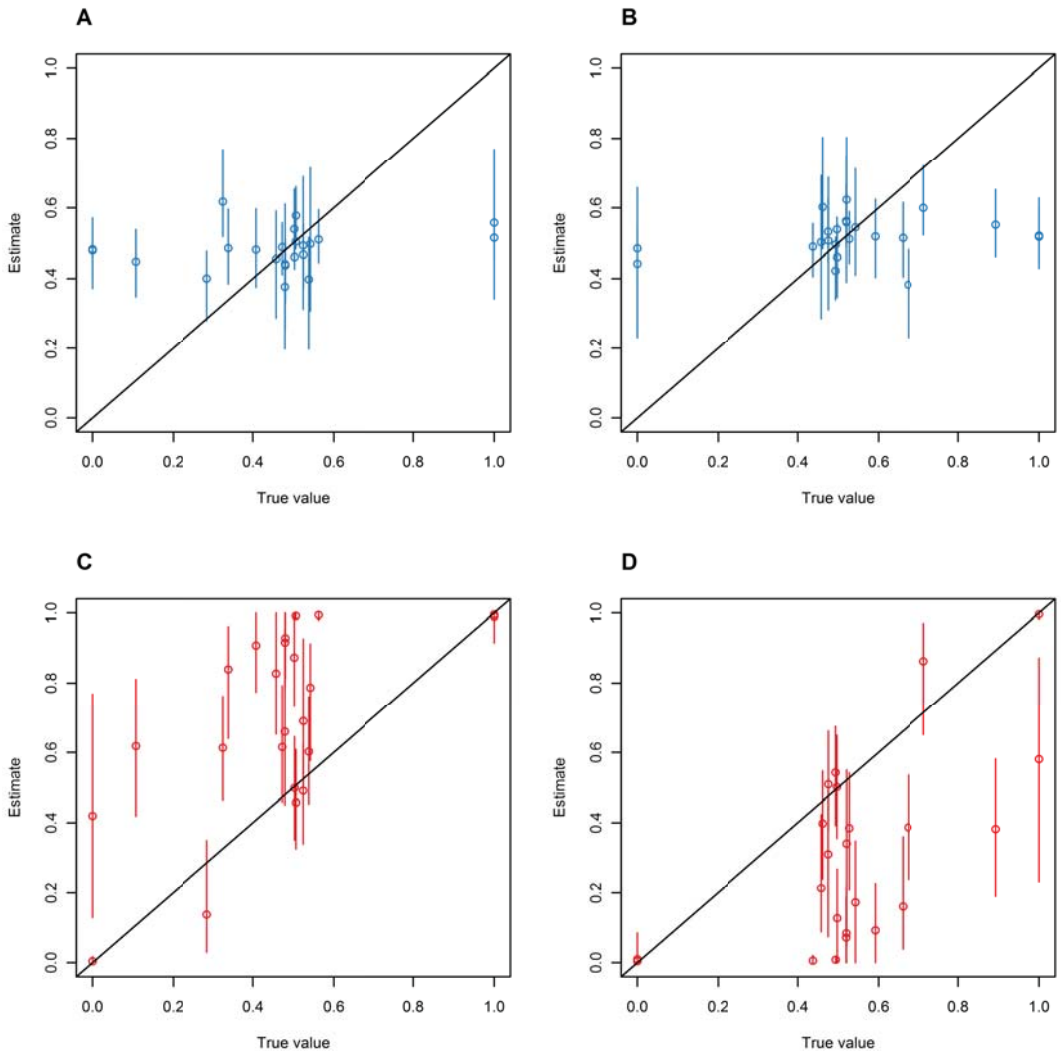


**Figure 4.6.** Scatterplots of the proportion of inter-population transmissions made up of animal-to-human (A and C) and human-to-animal (B and D) transmissions, versus the values estimated by the SC (blue: A and B) and DTA (red: C and D) models for 23 simulated outbreaks that were randomly sampled. The diagonal line represents accurate estimates of the sampled outbreaks, the dots represent the mean, and the error bars represent the 95% HPD interval.

There was a weak positive correlation between the known population and transmission parameters for the 23 simulated outbreaks and the number of animal and human isolates sampled (Figure 4.7; Table 4.3). The DTA model's mean estimates were more positively correlated with the proportion of samples from each host population than the SC model's mean estimates for the 23 simulated outbreaks (Appendix B.3).

**Table 4.3** Correlation coefficients of the proportion of isolates sampled from each host population versus the known transmission and population parameters for 23 simulated outbreaks that were randomly sampled.

| | Population | Transmissions |
|---|---|---|
| Correlation coefficient | 0.702 | 0.720 |

Population, the proportion of time spent in a host population; Transmission, the proportion of inter-population transmissions; Correlation coefficient, the correlation coefficient between the known parameter and the proportion of samples isolated from each host population

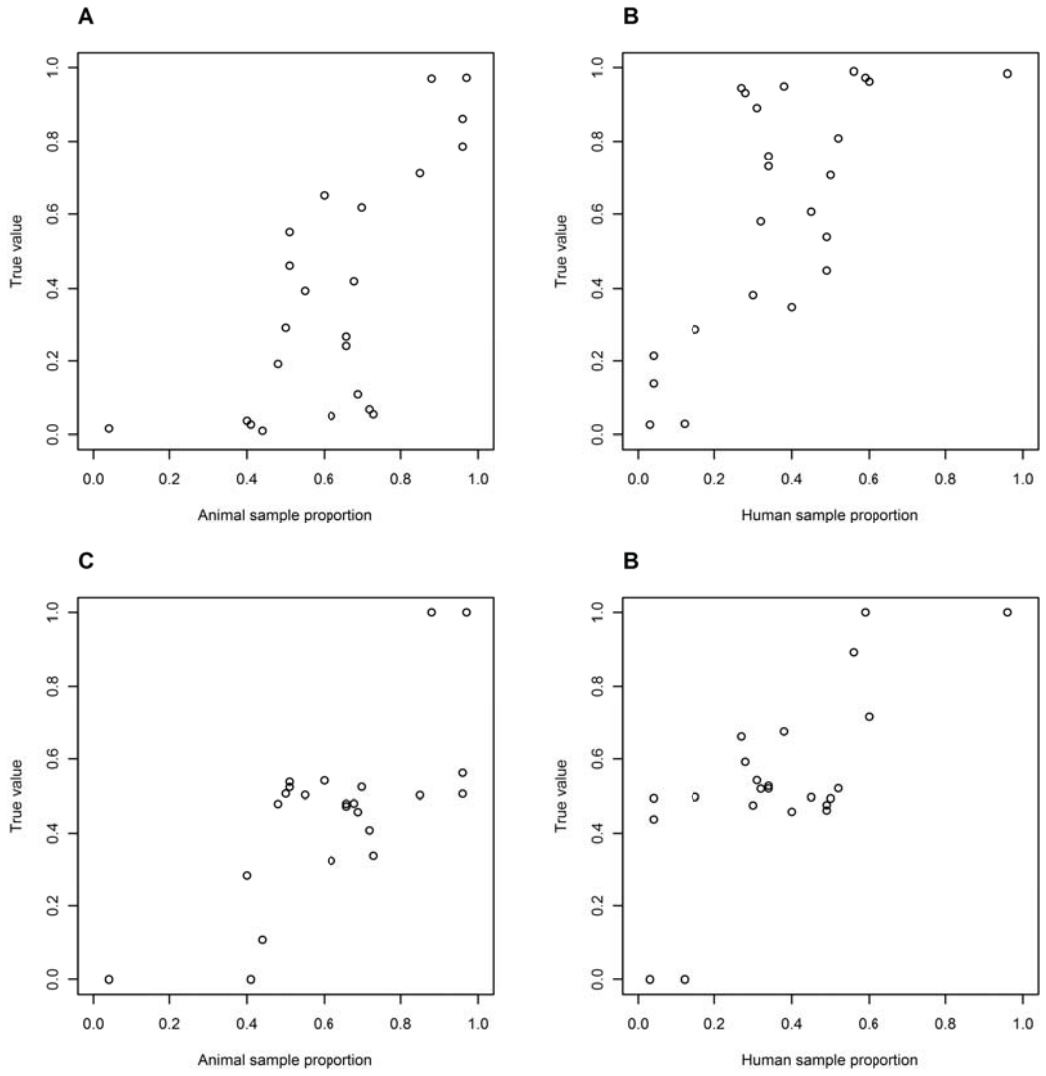


**Figure 4.7.** Scatterplots of the proportion of samples made up of animal (A and C) and human (B and D) isolates (x-axis) versus the proportion of time spent in the animal (A) and human (B) host populations and the proportion of inter-population transmissions made up of animal-to-human (C) and human-to-animal (D) transmissions (y-axis), for 23 simulated outbreaks that were randomly sampled.

The phylogenetic trees produced by the DTA and SC models for the 23 simulated outbreaks poorly reflected the sampled transmission trees (Figure 4.8). The DTA model was unable to predict transmissions along branches, which occurred in the transmission trees. The SC model could predict transmissions along branches, but often over-estimated the amount of transmissions compared to the transmission tree. The SC model also often predicted that one host population was predominant along the branches but involved in few coalescent events, whilst the other host population took up a minority of the branches but was involved in most coalescent events, even though this was not evident in the transmission tree. The phylogenetic trees in Figure 4.8 represent the most likely trees estimated using the DTA and SC models for one simulated outbreak, not the variation amongst each model, as each model estimated thousands of phylogenetic trees.

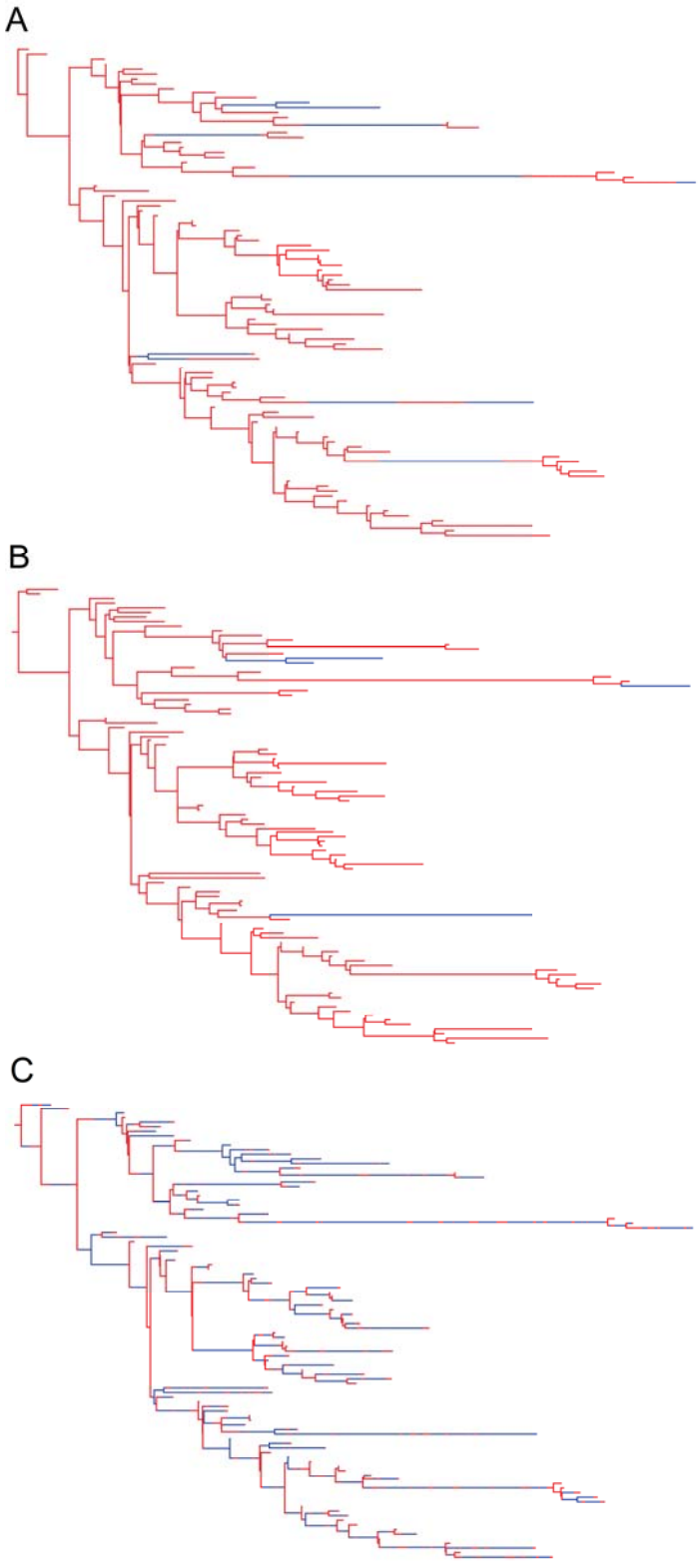


**Figure 4.8.** Sampled transmission tree (A), maximum clade credibility tree produced by the DTA model (B) and maximum a posteriori tree produced by the SC model (C), for one of the 23 simulated outbreaks that was randomly sampled.

### 4.4.4 Equal-time sampling

The 23 simulated salmonellosis outbreaks were re-sampled with equal numbers of isolates sampled from each year of the outbreaks. The SC and DTA models gave similar estimates for both sampling methods (Appendix B.2).

### 4.4.5 Equal intra-population transmission rates and infectious periods

The DTA and SC models gave more accurate population estimates for the 12 simulated outbreaks with equal intra-population transmission rates and infectious periods between host populations (Figure 4.9; Table 4.4) than the 23 simulations where these parameters varied (Figure 4.5; Tables 4.2), with smaller mean squared errors, a higher proportion of known parameter within their 95% HPD intervals and mean estimates that were more positively correlated with the known parameters. The DTA model's mean population estimates displayed a sigmoid shape, similar to the simulated outbreak that was disproportionately sampled (Figure 4.4). The DTA and SC models gave less accurate transmission estimates for the 12 outbreaks with equal intra-population transmission rates and infectious periods between host populations (Figure 4.10; Table 4.4) than for the 23 simulations where these parameters varied (Figure 4.6; Table 4.2), with larger mean squared errors, a lower proportion of known parameter within their 95% HPD intervals and mean estimates that were less positively correlated or negative correlated with the known parameters.

**Table 4.4** Summary statistics of the SC and DTA models' results compared to known parameters for 12 simulated outbreaks with equal infectious periods and intra-population transmission rates between host populations.

| | Population | | Transmissions | |
|---|---|---|---|---|
| | SC | DTA | SC | DTA |
| Estimate accuracy | 0.75 | 0.333 | 0.333 | 0.167 |
| Mean squared error | 0.032 | 0.027 | 0.103 | 0.152 |
| 95% HPD size | 0.563 | 0.119 | 0.272 | 0.252 |
| Correlation coefficient | 0.823 | 0.938 | -0.426 | 0.233 |

Population, the proportion of time spent in a host population; Transmission, the proportion of inter-population transmissions; Estimate accuracy, the proportion of outbreak simulations that a model included the known parameter within their 95% HPD intervals; Mean squared error, the mean squared error between the known parameters and a model's mean estimates; 95% HPD size, the mean 95% HPD interval size; Correlation coefficient, the correlation coefficient between the known parameter and a model's mean estimates

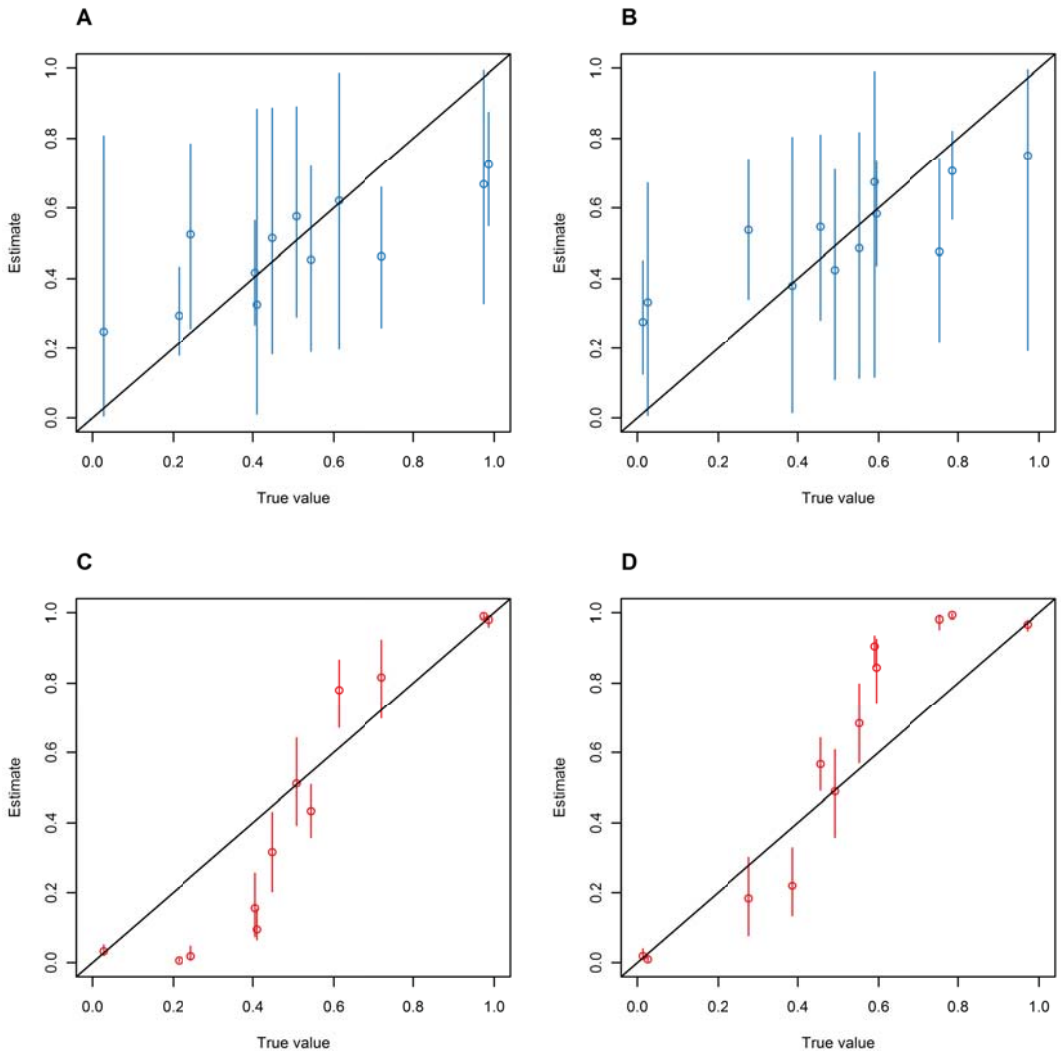


**Figure 4.9.** Scatterplots of the proportion of time spent in the animal (A and C) and human (B and D) host populations, versus the values estimated by the SC (blue: A and B) and DTA (red: C and D) models for 12 simulated outbreaks with equal gamma and intra-population transmission rates between host populations. The diagonal line represents accurate estimates of the sampled outbreaks, the dots represent the mean, and the error bars represent the 95% HPD interval.

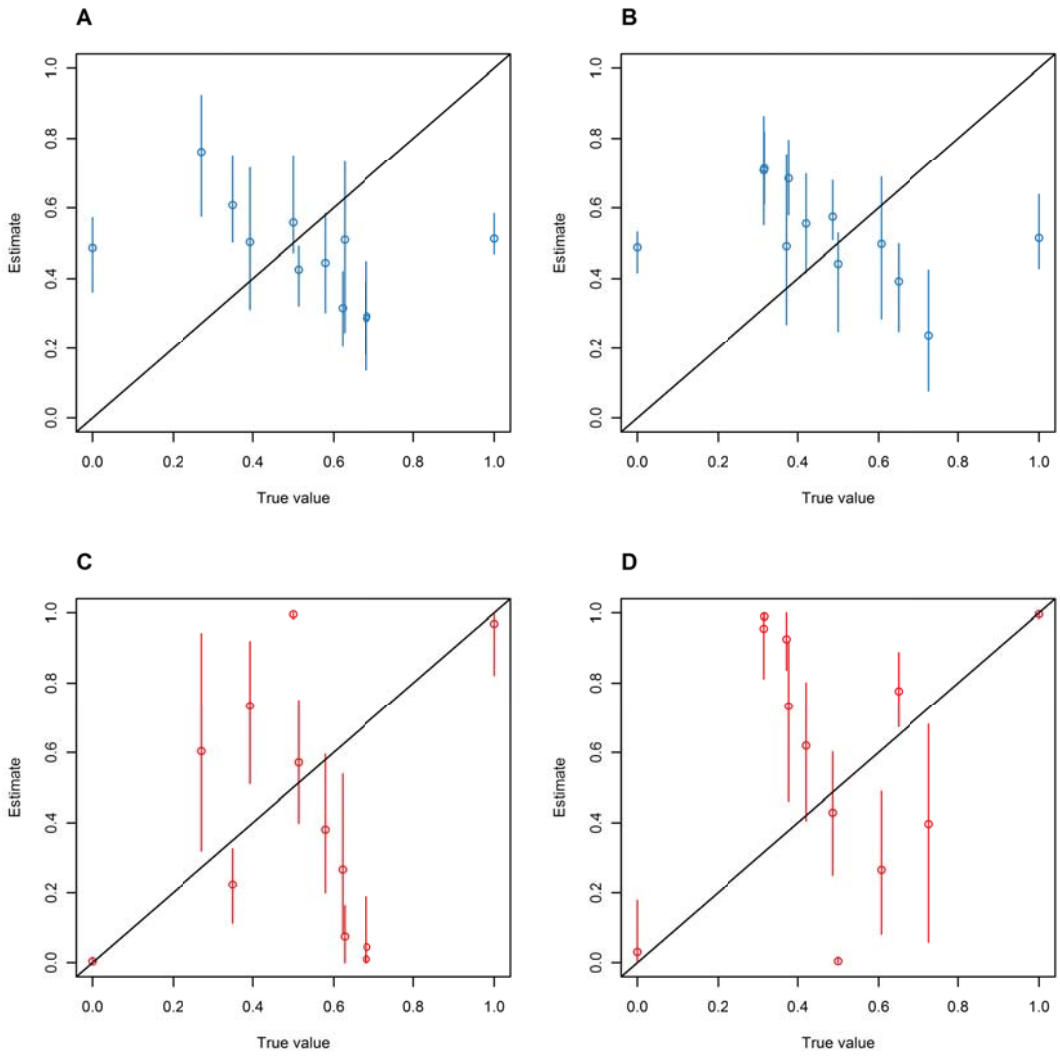


**Figure 4.10.** Scatterplots of the proportion of inter-population transmissions made up of animal-to-human (A and C) and human-to-animal (B and D) transmissions, versus the values estimated by the SC (blue: A and B) and DTA (red: C and D) models for 12 simulated outbreaks with equal gamma and intra-population transmission rates between host populations. The diagonal line represents accurate estimates of the sampled outbreaks, the dots represent the mean, and the error bars represent the 95% HPD interval.

The phylogenetic trees estimated for the 12 outbreaks with equal intra-transmission rates and infectious periods between host populations (Figure 4.11) were like those of previous simulated outbreaks (Figure 4.7). They also demonstrated that the DTA model was unable to predict ancestral branch states that were a different host population to daughter branches and tips. The SC model could predict the state of ancestral branches that differed to the tips, but often predicted these branches inaccurately.

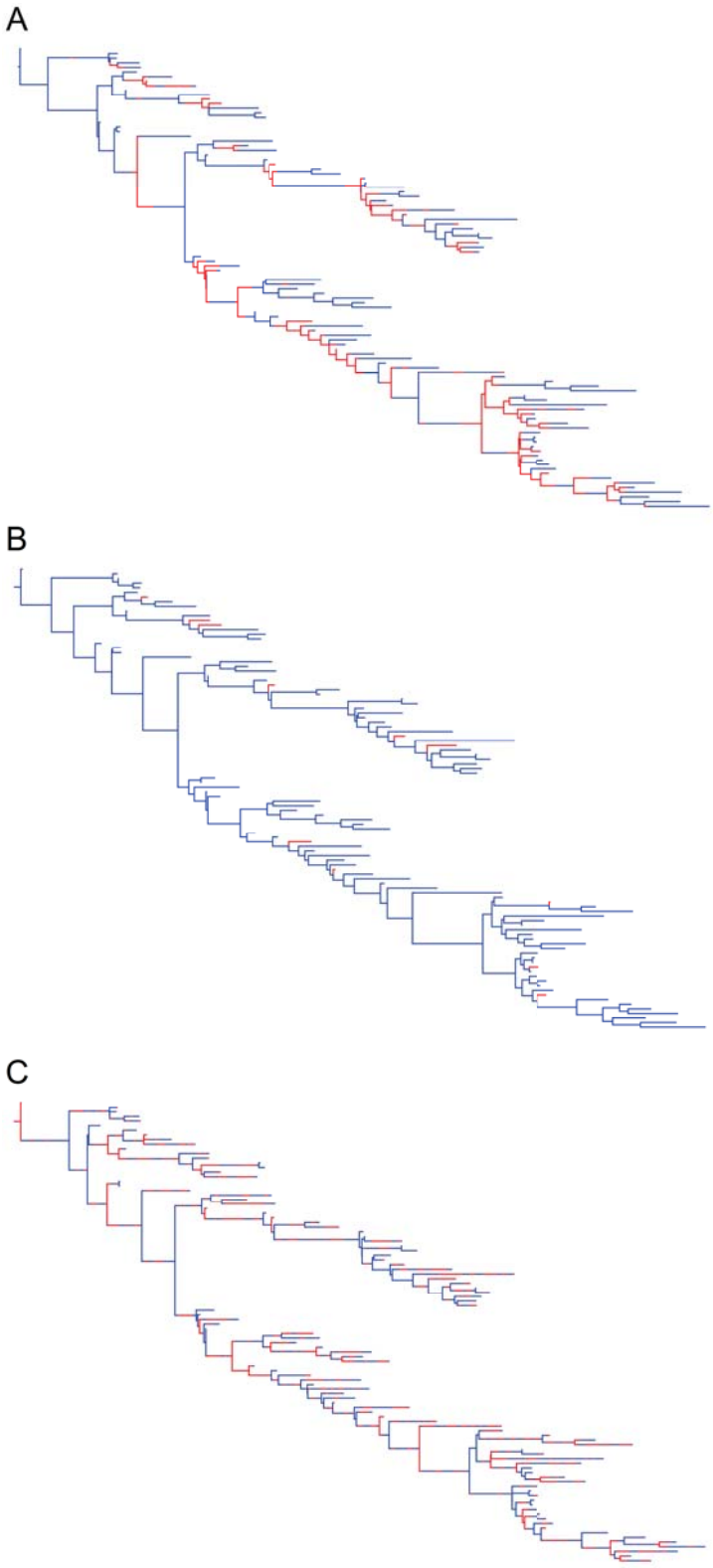


**Figure 4.11.** Sampled transmission tree (A), maximum clade credibility tree produced by the DTA model (B) and maximum a posteriori tree produced by the SC model (C), for a simulated outbreak with equal gamma and intra-population transmission rates between host populations.

For the 12 simulated outbreaks with equal intra-population transmission rates and infectious periods between host populations, the population parameters were more positive correlated with the proportion of samples from each host population than the transmission parameters (Figure 4.12 and Table 4.5). The DTA model's mean estimates were more positively correlated with the proportion of samples from each host population than the SC model's mean estimates (Appendix B.3).

**Table 4.5** Correlation coefficients of the proportion of isolates sampled from each host population versus the known transmission and population parameters for 12 simulated outbreaks with equal gamma and intra-population transmission rates between host populations.

| | Population | Transmissions |
|---|---|---|
| Correlation coefficient | 0.973 | 0.169 |

Population, the proportion of time spent in a host population; Transmission, the proportion of inter-population transmissions; Correlation coefficient, the correlation coefficient between the known parameter and the proportion of samples isolated from each host population

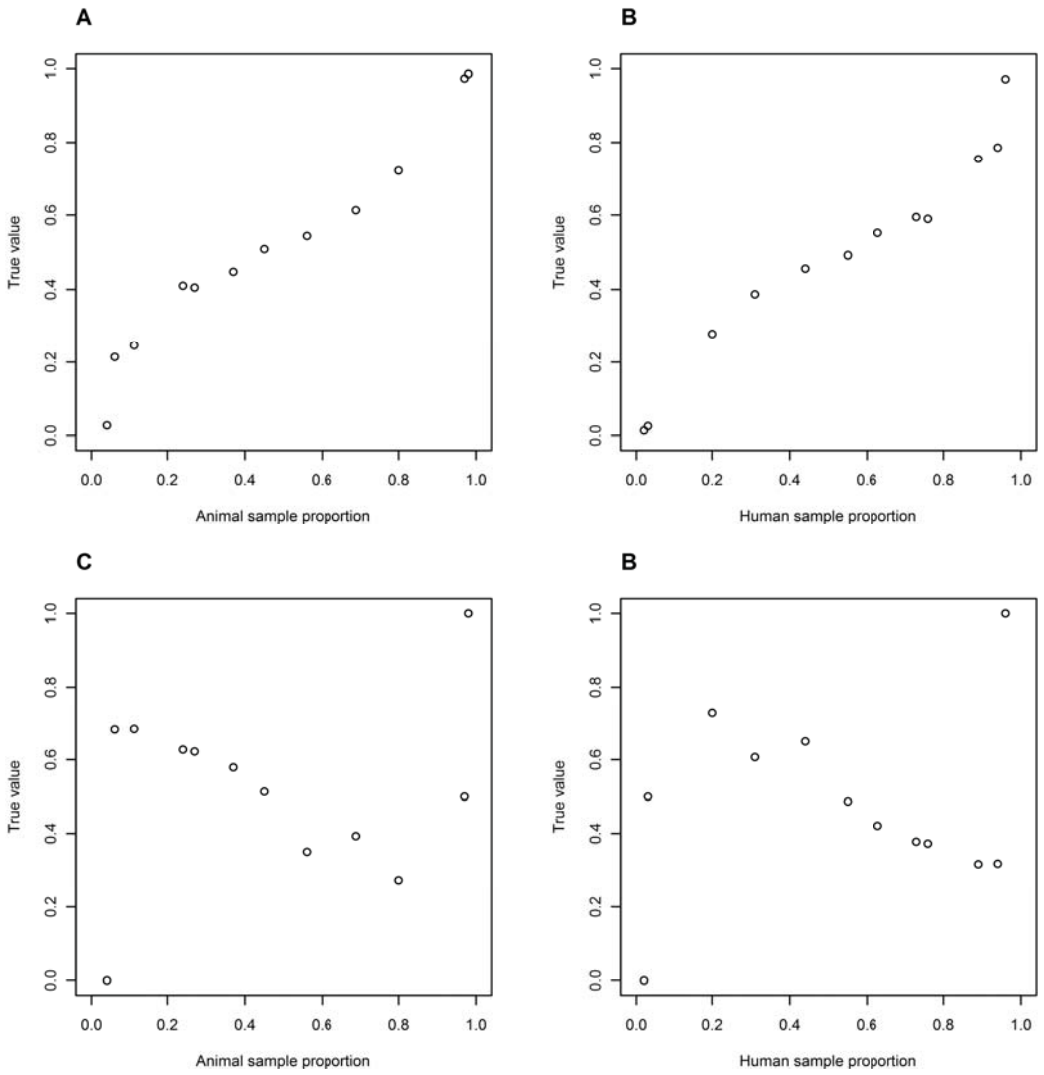


**Figure 4.12.** Scatterplots of the proportion of samples made up of animal (A and C) and human (B and D) isolates versus the proportion of time spent in the animal (A) and human (B) host populations and the proportion of inter-population transmissions made up of animal-to-human (C) and human-to-animal (D) transmissions, for 12 simulated outbreaks with equal gamma and intra-population transmission rates between host populations.

### 4.4.6 DT160 outbreak

The SC and DTA models both predicted that DT160 spent most of the time in the animal host population over the course of the DT160 outbreak in New Zealand (Figure 4.13). However, the SC model predicted that there were relatively equal amounts of transmission between the animal and human host populations, whilst the DTA model predicted that there was a large amount of animal-to-human transmission and relatively less human-to-animal transmission. The phylogenetic trees estimated for the DT160 outbreak also displayed larger intervals between coalescent events later in the outbreak compared to the outbreaks simulated in this study (Figure 4.14).

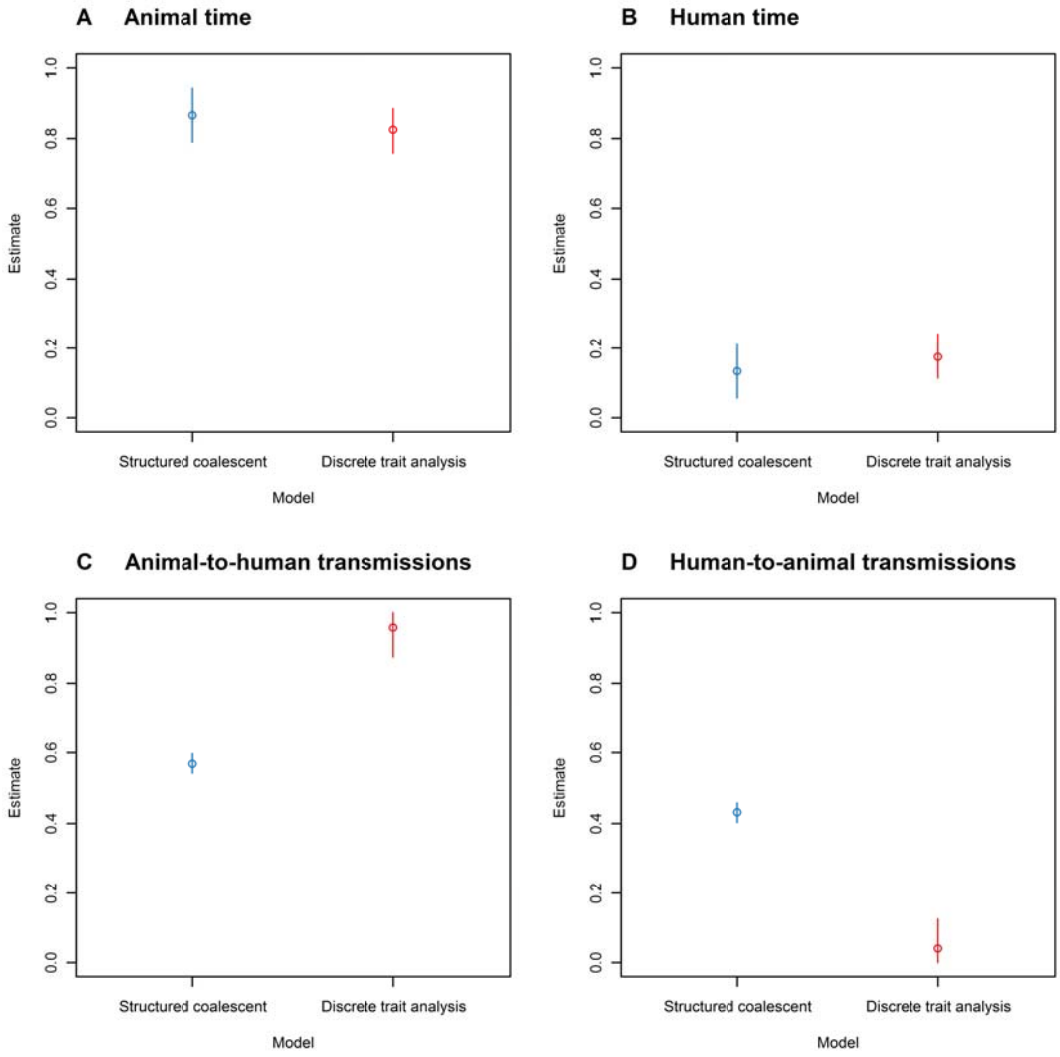


**Figure 4.13.** Estimates of the proportion of time spend in the animal (A) and human (B) host populations, and the proportion of inter-population transmissions made up of animal-to-human (C) and human-to-animal (D) transmissions for the DT160 outbreak, as estimated by the SC (blue) and DTA (red) models. The circles represent the mean and the error bars represent the 95% HPD interval.

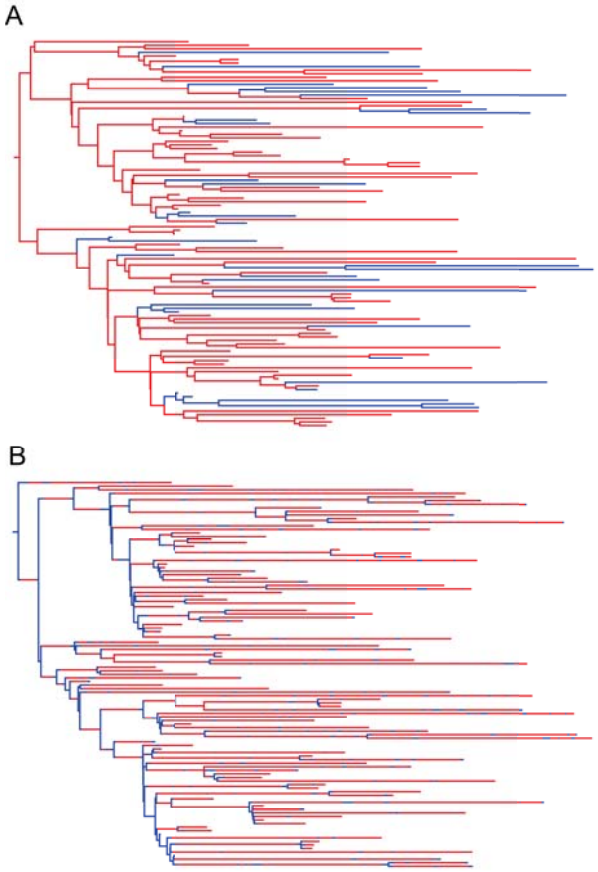


**Figure 4.14.** Maximum clade credibility tree produced by the DTA model (A) and maximum a posteriori tree produced by the SC model (B), for the DT160 outbreak.

## 4.5 Discussion

The DTA and SC models are ancestral state reconstruction models that were designed to estimate the ancestral state of a group of organisms based on their evolutionary history (Lemey et al., 2009; Vaughan et al., 2014). In this study, I demonstrated using simulated and actual salmonellosis outbreaks that neither of these models could accurately estimate known population and transmission parameters.

The DTA model assumes that the proportion of samples from each host population is proportional to its relative size (De Maio et al., 2015). This is a problem for outbreaks involving multiple host populations, as the host populations may be sampled at different rates, resulting in samples disproportional to the number of individuals infected within each host population. The simulated outbreaks in this study were stratified by host population before random sampling to compensate for this assumption. However, differing intra-population transmission rates and infectious periods resulted in sample sizes disproportionate to the known simulated outbreak parameters. This explains why the DTA model consistently over-estimated the length of time in the animal host population and the number of animal-to-human transmissions for the initial 23

simulated outbreaks, as the human host populations of these outbreaks were simulated to have a longer infectious period than the animal host populations that resulted in longer periods spent in the human host population and a larger number of human-to-animal transmissions relative to the number of humans sampled.

The DTA model estimated population parameters accurately when the parameter was directly proportional to the number of isolates from each host population sampled. In these instances, the population estimates and simulated outbreak parameters shared a sigmoid-like relationship due to the model's ancestral branch estimates: the DTA model usually predicts that all the ancestral branches are one host population, until the majority of the tips are another host population, where all the ancestral branches switch (Appendix A.3). This relied on simulating outbreaks with equal intra-population transmission rates and infectious periods, parameters that usually differ between *Salmonella* host populations (Alexander et al., 2009; Murase et al., 2000). However, even in these instances the DTA model inaccurately estimated ancestral host population states and transmission parameters.

The SC model gave similar estimates for all the simulated outbreaks. Occasionally it accurately estimated a simulation's parameters, but this usually only occurred when these parameters were within the range that it consistently estimated. This suggests that the model was unable to obtain a clear signal from the sequence data to indicate that *Salmonella* was in one host population for longer or there was uneven transmission between the host populations. Possibly this is due to the model's assumption that the effective population size of the host populations were consistent throughout the outbreak (Vaughan et al., 2014), which does not apply to salmonellosis outbreaks whose effective population size varies over the course of the outbreak (Chapter 3.4.3). There may be other reasons why the SC model was unable to detect a signal, but it is impossible to test for these without first accounting for the model's effective population size assumption.

The inability of the SC and DTA models to accurately estimate salmonellosis outbreak parameters highlights the need for outbreak-specific models. These models would need to be able to take into consideration variable sampling between host populations, like the SC model, and changes in the effective population size, like the DTA model. In addition, they would need to be able to take into consideration variation in infectious periods and intra-population transmission rates.

The MASTER package of BEAST2 allowed many salmonellosis outbreaks to be simulated using the Gillespie direct method stochastic SIR model. The simulated outbreaks contained a large amount of variation in the amount of time spent in the animal and human host populations, but less variation in inter-population transmissions. The lack of inter-population variation was due to the simulations only considering two host populations, as when an infectious agent moves between host populations it can either stay in the host population or move back to the previous host population, leaving relatively even inter-population transmission values. It was only possible to simulate outbreaks with uneven transmission values by using one very high and one very low inter-population transmission value. This in part explains why the SC model was better at predicting transmission parameters, as it always gave similar estimates with mean estimates around the 0.35-0.65 range, which most of the known transmission parameters for the simulated outbreaks were within.

In addition the simulated outbreaks did not take into consideration variation in sample sizes or the amount of sequence data. Further work with these SIR models may determine how these factors influence ancestral state reconstruction model estimates.

The DTA and SC models' estimates of the DT160 outbreak outline some of the limitations of this study. The DTA model estimated that DT160 spent most of its time in the animal host population and that there was a larger amount of animal-to-human transmission than human-to-animal transmission, which is to be expected as the DTA model is affected by sample size and a larger number of animal isolates were analysed than human. The SC model estimated similar amounts of animal-to-human transmission than human-to-animal transmission, which is also to be expected as it usually gave similar transmission rates between two host populations. However, the SC model estimated that DT160 spent over 90% of its time in the animal host population and less than 10% of its time in the human host population, outside the 20-80% range estimated for simulated outbreaks, and both models produced phylogenetic trees with larger distances between coalescent events towards the later part of the outbreak than simulated outbreaks. The effective population size affects the timing of coalescent events for randomly sampled individuals (Heled and Drummond, 2008). Therefore, the DT160 outbreak had a much larger effective population size then any of the simulated outbreaks in this study. It also indicates that the SC model's estimates maybe influenced by branch length. Simulations using larger outbreaks are required to test this.

## 4.6 Conclusion

In this study, the applicability of the SC and DTA models to salmonellosis outbreaks were investigated. Comparisons between the known parameters of simulated outbreaks and the models' estimates suggest neither model is appropriate for this analysis and highlights the need for outbreak-specific models that can also take into consideration intra-population transmission rates, infectious periods, disproportionate sampling and changes in the effective population size.

# References

Alexander, K. A., Warnick, L. D., Cripps, C. J., McDonough, P. L., Grohn, Y. T., Wiedmann, M., Reed, K. E., James, K. L., Soyer, Y., and Ivanek, R. (2009). Fecal shedding of, antimicrobial resistance in, and serologic response to *Salmonella* Typhimurium in dairy calves. *Journal of the American Veterinary Medical Association*, 235(6):739–748.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., and Drummond, A. J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, 10(4).

Collins (2017). Trait. Retrieved 2017-06-22, from: `https://www.collinsdictionary.com/dictionary/english/trait`, 1-1.

De Maio, N., Wu, C., O'Reilly, K., and Wilson, D. (2015). New routes to phylogeography: A Bayesian structured coalescent approximation. *PLoS Genetics*, 11(8):1–22.

Drummond, A., Suchard, M., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8):1969–1973.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, 17(6):368–376.

Gould, L. H., Walsh, K. A., Vieira, A. R., Herman, K., Williams, I. T., Hall, A. J., and Cole, D. (2013). Surveillance for foodborne disease outbreaks - United States, 1998-2008. *MMWR Surveillance Summaries*, 62(1):1–34.

Heled, J. and Drummond, A. J. (2008). Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, 8(1):1–15.

Lemey, P., Rambaut, A., Drummond, A., and Suchard, M. (2009). Bayesian phylogeography finds its roots. *PLoS Computational Biology*, 5(9):1–16.

Mather, A., Reid, S., Maskell, D., Parkhill, J., Fookes, M., Harris, S., Brown, D., Coia, J., Mulvey, M., Gilmour, M. o., Petrovska, L., De Pinna, E., Kuroda, M., Akiba, M., Izumiya, H., Connor, T., Suchard, M. l., Lemey, P., Mellor, D., Haydon, D., and Thomson, N. (2013). Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science*, 341(6153):1514–1517.

Minin, V., Bloomquist, E., and Suchard, M. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, 25(7):1459–1471.

Murase, T., Yamada, M., Muto, T., Matsushima, A., and Yamai, S. (2000). Fecal excretion of *Salmonella enterica* serovar Typhimurium following a food-borne outbreak. *Journal of Clinical Microbiology*, 38(9):3495–3497.

Tavare, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *American Mathematical Society*, 17:57–86.

Vaughan, T. G. and Drummond, A. J. (2013). A stochastic simulator of birth-death master equations with application to phylodynamics. *Molecular Biology and Evolution*, 30(6):1480–1493.

Vaughan, T. G., Kühnert, D., Popinga, A., Welch, D., and Drummond, A. J. (2014). Efficient Bayesian inference under the structured coalescent. *Bioinformatics*, 30(16):2272–2279.

WHO (2016). Disease outbreaks. Retrieved 2017-06-29, from: http://www.who.int/topics/disease_outbreaks/en/, 1–1.

Wikström, V. O., Fernström, L.-L., Melin, L., and Boqvist, S. (2014). *Salmonella* isolated from individual reptiles and environmental samples from terraria in private households in Sweden. *Acta Veterinaria Scandinavica*, pages 1–6.

Wright, S. (1931). Evolution in Medelian populations. *Genetics*, 16(2):97–159.

# Chapter 5

## Genomic comparison of two *Salmonella enterica* serovar Typhimurium strains responsible for consecutive salmonellosis outbreaks in New Zealand

### 5.1 Abstract

*Salmonella enterica* serovar Typhimurium DT160 was the predominant cause of notified human salmonellosis cases in New Zealand from 2000-2010, before it was superseded by another *S.* Typhimurium strain identified as DT56 variant. Whole genome sequencing was used to compare 109 DT160 isolates with eight DT56 variant isolates from animal and human New Zealand source groups to determine the time of common ancestor of both strains, and identify potential reasons for the decline in DT160 and emergence of DT56 variant. Phylogenetic analysis provided evidence that DT160 and DT56 variant strains were distantly related with an estimated date of common ancestor around the 16$^{th}$ century Common Era. Genomic comparisons confirmed that all DT160 isolates contained the ST160 phage and a plasmid similar to the pSTUK-100 plasmid, whilst the DT56 variant isolates contained the BTP1 phage, which may have contributed to its emergence. It also identified a novel plasmid in one of the DT56 variant isolates analysed. The study demonstrates how comparative genomics can help identify strain-specific elements and factors that may have influenced the emergence and decline of bacterial strains of public health importance.

### 5.2 Introduction

Salmonellosis is inflammation of the intestinal tract attributed to infection with non-typhoidal *Salmonella*. Non-typhoidal *Salmonella* consists of multiple serovars and strains, which are associated with different sources (Magwedere et al., 2015). Some non-typhoidal *Salmonella* strains are host specific, associated with a single source, whilst others are generalists, associated with a spectrum of hosts (Xu et al., 2009). A source may contain one or multiple strains of non-typhoidal *Salmonella* and the strains associated with the source may change over time (Van Duijkeren et al., 2002). This change in non-typhoidal *Salmonella* strains may alter the dynamics of salmonellosis in a geographical area (Yang et al., 2015), complicating control measures.

Multiple factors influence the prevalence of non-typhoidal *Salmonella* strains in animal reservoirs and human clinical cases. For example, interventions aimed at eradicating one strain of *Salmonella* from a population may allow another strain to fill the ecological niche (Bäumler et al., 2000); changes in the behaviour of a population (e.g. changed dietary habits) may alter the transmission of a *Salmonella* strain between the population and a source (Tavechio et al., 1996); natural infection to a *Salmonella* strain may lead to herd immunity, allowing an immunologically distinct strain to replace it (Lu et al., 2013); and vaccination against a *Salmonella* strain can decrease its prevalence within a population (Collard et al., 2008). However, with so many potential causes it may be hard to determine the specific reasons why one strain of *Salmonella* supersedes another, replacing it as the most prevalent cause of salmonellosis in a geographical region.

In New Zealand, *Salmonella enterica* serovar Typhimurium definitive type (DT) 160 replaced *S.* Typhimurium DT135 as the predominant cause of human salmonellosis cases in 2000; a situation that persisted until 2010. It was hypothesised that the introduction of the ST160 phage into New Zealand may have precipitated the decline in DT135, which was susceptible to this phage, allowing resistant strains, such as DT160 to increase in prevalence (Price-Carter et al., 2011). *S.* Typhimurium DT160 was then superseded by *S.* Typhimurium DT56 variant (formerly RDNC-May06) as the predominant cause of human salmonellosis in 2011 (Figure 5.1). Previous research found that the strains were isolated from the same animal and environmental sources (Figure 5.2), and pulsed-field gel electrophoresis indicated that these strains were closely related (French et al., 2013). The aim of this study was to investigate the ancestral relationship between these two strains and identify potential reasons for one strain superseding the other.

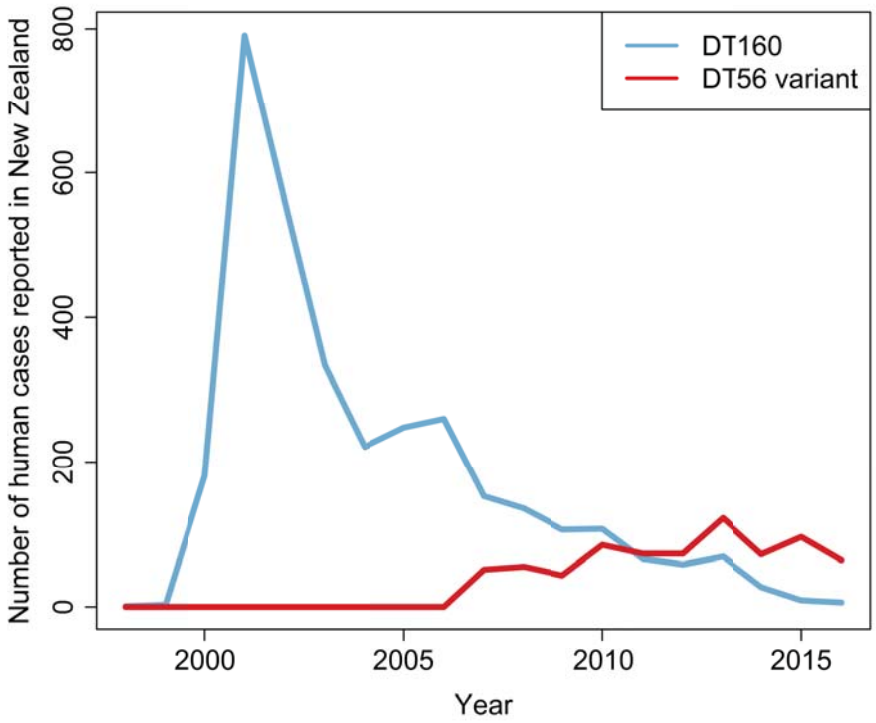


**Figure 5.1.** Line graph of the number of human DT160 (blue) and DT56 variant (red) cases reported in New Zealand each year from 1998-2016 (ESR, 2003; Public Health Surveillance, 2017a).

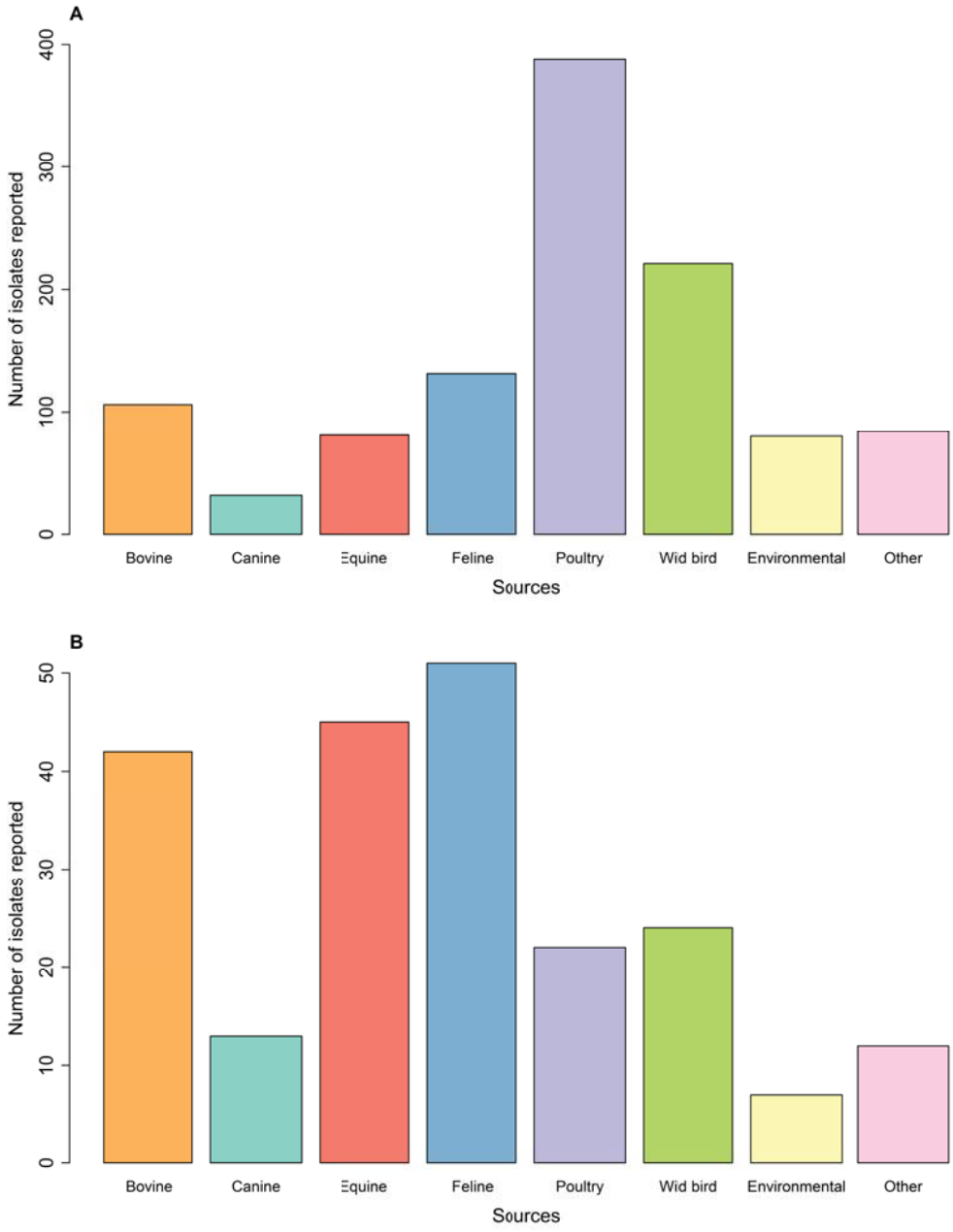


**Figure 5.2.** Bar graphs of the number of non-human DT160 (A) and DT56 variant (B) isolates reported in New Zealand from 1998-2016 (ESR, 2003; Public Health Surveillance, 2017b). The y-axes use different scales as a larger number of DT160 isolates were reported than DT56 variant isolates.

## 5.3 Methods

### 5.3.1 Epidemiological data

Epidemiological data on the DT160 and DT56 variant strains was obtained from the annual Institute of Environmental Science and Research Limited (ESR), and Public Health Surveillance reports (ESR, 2003;

Public Health Surveillance, 2017a,b). These reports contain information on the number of *Salmonella* strains reported in human and non-human sources in New Zealand.

### 5.3.2 Whole genome sequencing

Eight DT56 variant isolates collected from 2007-2013 were supplied from the Enteric Reference Laboratory at ESR, Wallaceville, New Zealand. These isolates were obtained from human (n=4), bovine (n=2), wild bird (n=1) and feline (n=1) sources. Genomic DNA was extracted from these isolates using a QIAamp DNA mini kit (Qiagen, Hilden, Germany) (Qiagen, 2012). Genome extracts were whole genome sequenced by New Zealand Genomics Limited (NZGL) at the Massey Genome Service (Massey University, Palmerston North, New Zealand). NZGL prepared a library for each isolate using an Illumina TruSeq™DNA PCR-Free kit (Illumina, Scorsby, Victoria, Australia) and sequenced the libraries using an lllumina MiSeq (Illumina) as 2x250 bp paired-end runs (120-150 mean nucleotide coverage). After sequencing and standard barcode demultiplexing, NZGL performed quality control procedures to remove any PhiX control library reads and adapter sequences using FASTQ-MCF (Aronesty, 2013).

The eight DT56 variant isolates were compared to 109 previously described DT160 isolates (Chapter 3). A smaller number of DT56 variant isolates were investigated as the aim of this study was to investigate its relationship with the DT160 strain, rather than investigate the transmission and evolution of the DT56 variant strain, which was previously performed on the DT160 strain *per se*.

### 5.3.3 Genome assembly

Each isolate's genome was assembled *de novo*. An in-house Perl script was used to trim reads at an error probability of 0.01 using solexaQA++ (Cox et al., 2010) and generate random subsets of paired reads from 750,000 to 1,200,000 paired reads in increments of 150,000, varying the mean nucleotide coverage. Each of the random sets was assembled with the *de novo* assembler Velvet v1.1 (Zerbino and Birney, 2008) at a variety of kmers from 55 to 245 in increments of 10. This resulted in multiple genome assemblies for each isolate. The metrics for each of four parameters (longest genome length, fewest number of contigs, largest $N_{50}$ value, and longest contig length) were ranked in numerical order, and an overall equally summed ranking score was calculated for each assembly. The assemblies with the lowest total rank were used for further analysis. QUAST (Gurevich et al., 2013) was used to analyse the *de novo* assemblies and determine their GC content, the percentage of a DNA sequence made up of guanine and cytosine bases.

### 5.3.4 Single nucleotide polymorphism identification

Core single nucleotide polymorphisms (SNPs) were identified using kSNP3 v3.0 (Gardner et al., 2015) and Snippy v2.6 (https://github.com/tseeman/snippy). Snippy is a pipeline that uses the Burrows-Wheeler

Aligner (Li and Durbin, 2009) and SAMtools (Li, 2011) to align reads from different isolates to a sequence, and FreeBayes (Garrison and Marth, 2012) to identify variants among the alignments. kSNP3 was used to analyse *de novo* assembled genomes, along with the reference genome, *S.* Typhimurium 14028S (NC_016856). An in-house Python script was used to determine the read coverage of all the SNPs identified via kSNP. Snippy was used to align reads from each isolate to the reference genome, NC_016856, before identifying SNPs. SNPs were accepted if each isolate had a greater than 10 read depth and a greater than 90% consensus. The position of the SNP on the reference genome was used to determine if both methods identified the SNP or if they were unique to the method (Appendix C.1).

### 5.3.5 Phylogenetic inference

MEGA6 (Tamura et al., 2004) was used to form a maximum likelihood tree of the 117 DT160 and DT56 variant isolates based on the 1,709 core SNPs they share. Evolview v2 (He et al., 2016) was used to visualise and edit the tree.

### 5.3.6 Phylogenetic analysis

An in-house Perl script was used to split the 1,709 core SNPs shared by the 117 DT160 and DT56 variant isolates into groups based on whether they were associated with the 1st, 2nd or 3rd codon position of a gene, contained in overlapping coding regions, or found in intergenic regions. The partitioned SNPs were exported into BEAUti to create an Extensive Markup Language (XML) file for BEAST 1.8.3 (Drummond et al., 2012). The five SNP groups were given separate Hasegawa Kishino Yano (HKY) substitution models (Hasegawa et al., 1985), while their tree and clock models were linked to allow for variation in base substitution among codon positions. The effective population size for the isolates was assumed to be constant (Kingman, 1982) and a relaxed uncorrelated molecular clock (Drummond et al., 2006) was used to allow for variation in the mutation rate among different lineages and was calibrated by the tip dates. The XML file was run in BEAST for 20 million steps, three times with different starting seeds, before LogCombiner was used to combine the runs with a 10% burn-in. Tracer v1.6 (Rambaut et al., 2014) was used to visualise the results.

### 5.3.7 *Salmonella* Typhimurium comparison

The 117 *de novo* assembled genomes were uploaded into PATRIC (Wattam et al., 2014) and annotated with RAST (Aziz et al., 2008). PLfams (Davis et al., 2016) was used to cluster the 117 annotated genomes and 41 completed *S.* Typhimurium genomes in PATRIC. An in-house perl script was used to identify SNPs among core genes of equal size and found once within all analysed genomes. SplitsTree was used to form a NeighbourNet tree from the 2,156 SNPs identified.

### 5.3.8 Protein-coding gene analysis

Missing proteins amongst the 109 DT160 and eight DT56 variant isolates were identified using a previously described protein-coding gene analysis method (Chapter 3). However, only those proteins that were missing from isolates were investigated. BLAST (Altschul et al., 1990) was used to determine what genetic elements the strain-specific proteins were found within. SRST2 v2 (Inouye et al., 2014) was used to align the reads from each isolate to the genetic elements identified via BLAST.

### 5.3.9 Novel plasmid analysis

plasmidSPAdes (Antipov et al., 2016) was used to *de novo* assemble reads from isolates suspected of containing novel plasmids (i.e isolates that contained genes that did not align to any known sequences via BLAST searches). *De novo* assembled plasmids were annotated using Prokka (Seemann, 2014) and the function of their proteins was predicted by comparing them to protein sequences in the Clusters of Orthologous Groups of proteins (COGs) database (Tatusov et al., 2000).

### 5.3.10 Scripts

The in-house scripts used in this study were previously described (Chapter 3). They are available from GitHub (https://github.com/samuelbloomfield/Scripts-for-genomic-analyses).

## 5.4 Results

### 5.4.1 Epidemiology

The DT160 and DT56 variant outbreaks displayed distinct epidemic curves in the human population of New Zealand (Figure 5.1). DT160 displayed a typical point source epidemic curve: the first case was reported in 1998, from 1999-2000 there was a rapid increase in prevalence that peaked in 2001, and from 2002 onwards the prevalence decreased. On the other hand, DT56 variant was first reported in 2006 and slowly increased in prevalence from 2007-2013, replacing DT160 as the predominant cause of salmonellosis in New Zealand in 2011. Since 2013, the prevalence of DT56 variant has slowly decreased. To date the prevalence of DT56 variant remains lower than the peak prevalence of DT160.

The DT160 and DT56 variant strains were isolated from a multitude of sources in New Zealand, but the predominant sources differed. DT160 was predominantly isolated from poultry and wild bird sources, whilst the DT56 variant strain was predominantly isolated from feline, equine, bovine and to a lesser extent wild bird sources (Figure 5.2).

### 5.4.2 Genome size and GC content

Assembled DT160 and DT56 variant genomes differed in length. DT160 isolates were 4.88 - 4.89 Mb in length (95% confidence interval (CI)) and on average 78 kb longer than DT56 variant isolates (95% CI: 4.80 - 4.82 Mb). All isolates contained a GC content of 52.11-52.16%, apart from one DT56 variant isolate that had a GC content of 52.02%. This isolate also had a genome that was 20 kb larger than the other DT56 variant isolates analysed by 20 kb.

### 5.4.3 Strain evolution and comparison

Phylogenetic analysis estimated that the 117 D160 and DT56 variant isolates mutated at a rate of 2.8-5.4 x $10^{-7}$ substitutions $site^{-1}$ $year^{-1}$ (95% HPD interval), and shared a date of common ancestor between the years 1419-1803 CE (95% HPD interval). The substitution rate estimated for the DT160 and DT56 variant isolates is similar to rates reported by Mather et al. (2013) for an outbreak of *S.* Typhimurium DT104 in Scotland from 1990-2012, Okoro et al. (2012) for invasive *S.* Typhimurium strains in sub-Saharan Africa, and from previous studies on DT160 in New Zealand (Chapter 3).

*S.* Typhimurium comparative studies revealed that the 109 DT160, eight DT56 variant and 41 other *S.* Typhimurium strains contained a pan-genome of 7,874 coding DNA sequences (CDS). The core genome consisted of 3,872 CDS, and 1,864 of these were of equal size and found only once within each isolate. The DT160 and DT56 variant strains were distinct from previously published *S.* Typhimurium strains (Figure 5.3). The DT56 variant strain was also more closely related to *S.* Typhimurium str. 22495 (CP017617), an isolate collected from a wild sea gull in Canada (Ogunremi et al., 2017), than the DT160 strain.

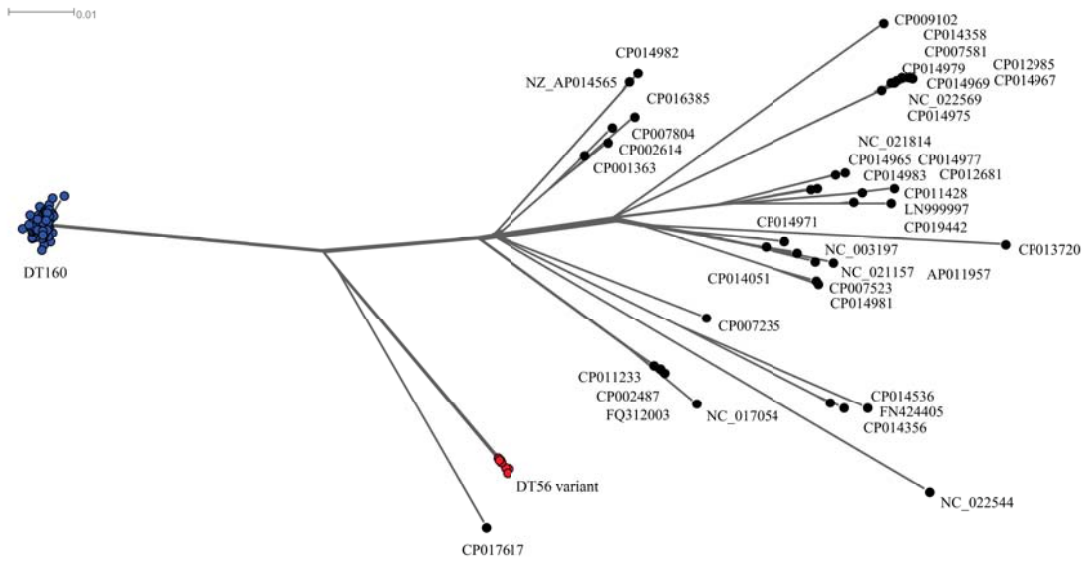


**Figure 5.3.** NeighbourNet tree of 109 DT160 isolates (blue), eight DT56 variant isolates (red) and 41 *S.* Typhimurium isolates in PATRIC, based on 2,156 SNPs in 1,864 CDS. The scale bar represents the number of nucleotide substitutions per site.

### 5.4.4 Protein differences

Protein-coding gene analysis identified 5,131 CDS contained by the 117 DT160 and DT56 variant isolates analysed. 211 of these were strain-specific: 176 DT160-specific and 35 DT56 variant-specific CDS. BLAST searches found that the DT160-specific CDS were found in the ST160 phage (NC_014900) and the pSTUK-100 plasmid (CP002615), whilst the DT56-variant CDS were found in *S.* Typhimurium strain D23580 (FN424405) (Figure 5.4). Read alignments between these strains confirmed this (Appendix C.2).

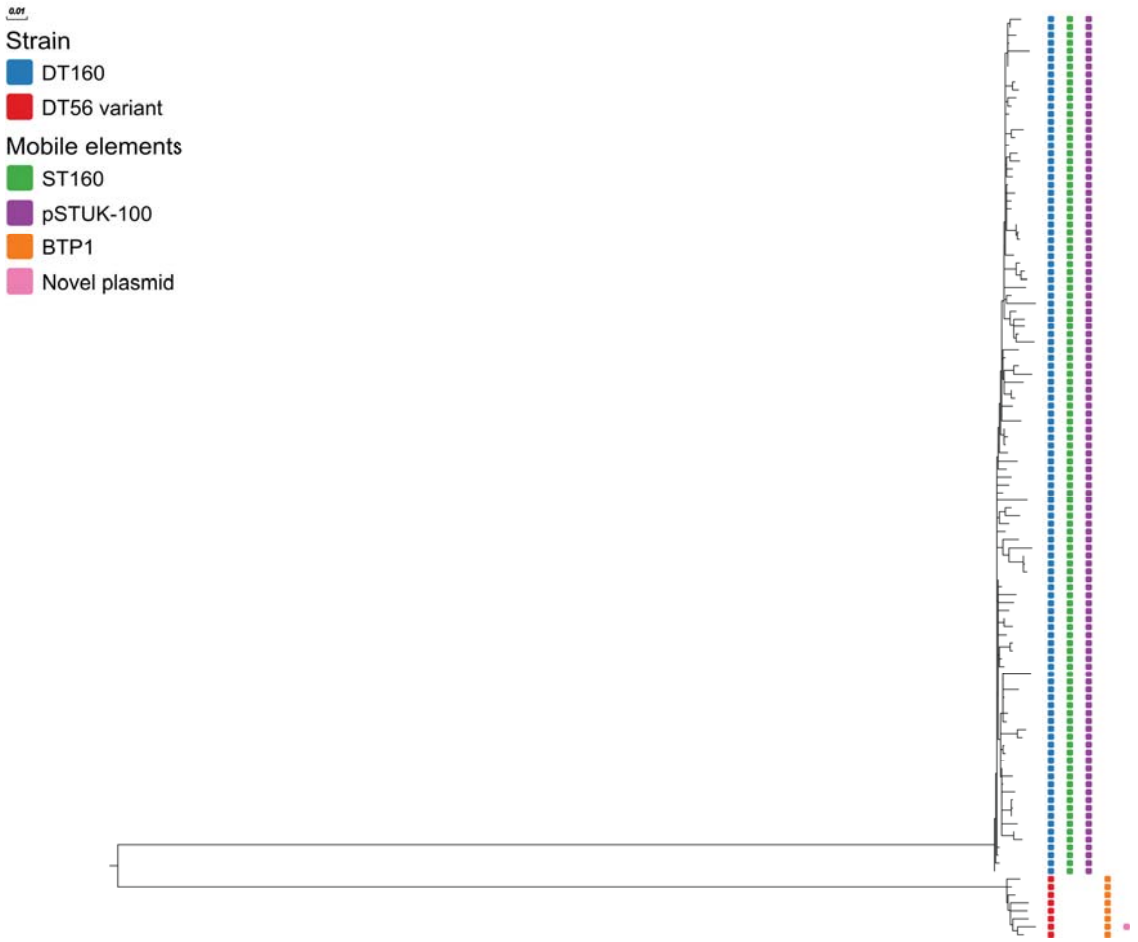


**Figure 5.4.** Maximum likelihood tree of 109 DT160 (blue) and eight DT56 variant (red) isolates (based on 1,709 core SNPs), and coloured by strain and the presence of mobile elements. The scale bar represents the number of nucleotide substitutions per site.

### 5.4.5 Novel plasmid

One of the eight DT56 variant isolates analysed contained a larger genome with an additional 42 genes compared to the other DT56 variant isolates. This DT56 variant isolate also had a lower GC content to the other DT56 variant isolates. BLAST was unable to identify any matches for 8 out of the 42 genes, and partially matched the remaining 34 genes to the linear *Salmonella enterica* serovar Typhi plasmid, pBSSB1 (AM419040) (Baker et al., 2007). However, it also identified a large amount of variation between these 34 genes and those of pBSSB1. When reads from each isolate were aligned to pBSSB1, only reads from the isolate that contained the extra genes, aligned to the sequence. However, the reads contained a large number of SNPs and did not align to the entire sequence (Appendix C.2). This suggests that the DT56 variant isolate contained a novel plasmid that may be distantly related to pBSSB1.

*De novo* plasmid assembly predicted that the novel plasmid was 28.7 kb in length. Prokka identified 38 CDS on the plasmid. Comparisons with the COG database could only identify the functional groupings of 15 of these 38 CDSs (Figure 5.5).

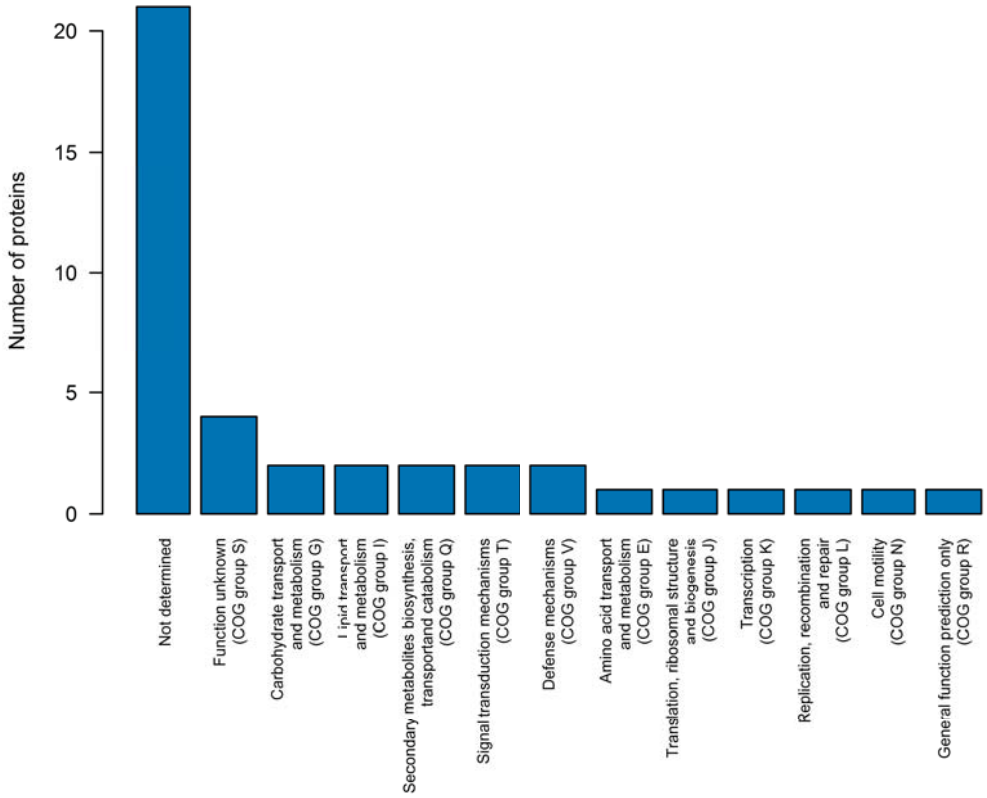


**Figure 5.5.** Bar graph of the predicted functions of the 38 CDS found on the novel DT56 variant plasmid.

## 5.5 Discussion

DT56 variant replaced DT160 as the predominant cause of human salmonellosis cases in New Zealand in 2010. Both strains were predominantly isolated from different animal and environmental sources. In New Zealand, *Salmonella* isolates are sent to ESR from medical, veterinary and food-testing laboratories throughout the country to be reported, serotyped and, depending on the serotype, phage typed. *Salmonella* isolates are cultured from clinical samples from both human and animals, and as part of routine surveillance of various sources (e.g. environmental testing during outbreaks). *Salmonella* strains vary in virulence (McWhorter et al., 2015), such that those collected as part of routine surveillance are more likely to contain both virulent and avirulent strains whilst those collected from clinical samples are more likely to contain virulent strains (Gebreyes et al., 2009). These less virulent strains may not cause a disease or may cause a milder disease that is less likely to be notified. In addition, methodologies for isolating *Salmonella* vary between laboratories,

which can disproportionately favour the isolation of certain *Salmonella* strains (Gorski, 2012). Finally, clinical samples are collected from potential *Salmonella* sources at different rates, e.g. companion animal owners will seek veterinary care for their pets and thus pet faecal samples will likely be cultured for *Salmonella* for a much milder disease than might be seen if wild birds were under observation. These factors may have biased the estimates of DT160 and DT56 variant prevalence in each source.

From 1998-2016, no interventions were conducted to decrease the prevalence of DT160 or DT56 variant in New Zealand. Thornley et al. (2003) suggested several strategies for preventing DT160 exposure, e.g. routine treatment of roof-collected rainwater, hygienic disposal of dead birds, and stricter fast-food protocols. However, there is no evidence of these strategies being deployed. In 2006, the New Zealand Food Safety Authority (NZFSA) released a risk management strategy to decrease the prevalence of poultry-associated campylobacteriosis, but this was not associated with a decline in the incidence of salmonellosis in New Zealand (Sears et al., 2011). Therefore, no external factors were identified that could account for the differences in DT160 and DT56 variant prevalence amongst the various animal and human populations.

The doubling time is the length of time required for a bacterial population to double the number of viable cells present. A strain of bacteria may be able to outcompete another strain if its doubling time is faster. However, it is difficult to use genomics to predict what strain replicates at a faster rate. It was once believed that the bacterial genome size affected the doubling time of bacteria (Riley and Anilionis, 1978), which would mean that DT56 variant with its smaller genome would replicate faster than DT160. However, Mira et al. (2001) found no correlation between genome size and doubling time amongst strains of *Escherichia coli*, and cases of *Salmonella* plasmid acquisition with no metabolic burden have been reported (Aviv et al., 2016). Therefore, it is unlikely that the smaller genome of the DT56 variant strain resulted in a faster replication rate compared to the DT160 strain. Nevertheless, DT56 variant may have a faster replication rate for a multitude of other reasons, such as better nutrient uptake, etc. Cell growth and competitive growth assays could help determine if the DT56 variant strain has a faster doubling time and could outcompete the DT160 variant strain in animal sources.

The DT160 and DT56 variant strains were reported in New Zealand at different times and places. DT160 was first reported in 1998, from a human salmonellosis patient in Christchurch (Thornley et al., 2003), whilst DT56 variant was first reported in 2006, from a human salmonellosis patient in Auckland (Public Health Surveillance, 2017a). However the strains were estimated to share a date of common ancestor in the $16^{\text{th}}$ century Common Era. Care must be taken when interpreting the date of common ancestor, particularly when using samples collected over a small time period (15 years) to predict a date of common ancestor centuries before. Regardless, phylogenetic analysis demonstrated that the two strains were not as closely related as suggested by French et al. (2013) and it is highly unlikely that the DT56 variant strain evolved from the DT160 strain in New Zealand.

The DT160 and DT56 variant strains were associated with different mobile elements that may have influenced their phenotypes. ST160 is a phage that is found in multiple *S.* Typhimurium strains. The

presence of ST160 in DT160 isolates but not DT56 variant isolates (previously referred to as RDNCMay06) is supported by Price-Carter et al. (2011)'s analysis of ST160 evolution within *S.* Typhimurium strains. In this paper, Price-Carter et al. also proposed that the ST160 phage allowed DT160 to emerge as the largest cause of human salmonellosis in New Zealand as it provided the strain with a selective advantage over the large number of ST160-susceptible strains circulating in New Zealand at the time. The DT56 variant strain was obtained from similar sources to the DT160 strain, but superseded it as the largest cause of human salmonellosis in New Zealand, suggesting that it is most likely resistant to the ST160 phage. Phage susceptibility testing is required to confirm this.

The pSTUK-100 plasmid was found in DT160 genomes but not DT56 variant genomes. This plasmid was initially isolated from *S.* Typhimurium UK-1 (Luo et al., 2011), but similar plasmids have been found in a large number of other *S.* Typhimurium strains (McClelland et al., 2001). These plasmids are large (approximately 90 kb in length) and contain approximately 110 genes, many of which are involved in virulence (Gonzalo-Asensio et al., 2013; Herrero et al., 2008). Mather et al. (2016) previously investigated the genomes of DT56 variant isolates in England and Wales, and also found that they lacked this plasmid. The absence of this plasmid in the DT56 variant strain may indicate that it is less virulent than the DT160 strain, unable to affect as many humans and other animals, or causing a milder disease that is not detected as frequently by clinical intervention or health-seeking behaviour in New Zealand (Figures 5.1 and 5.2). The role played by these plasmids in determining virulence could be determined by virulence assays involving strains with and without these plasmids.

The DT56 variant-specific genes were identified within the BTP1 prophage section of the genome of *S.* Typhimurium D23580; a multi-drug resistant isolate belonging to the ST313 sequence type. ST313 are the predominant cause of invasive non-typhoidal salmonellosis in Malawi and Kenya (Kingsley et al., 2009). BTP1 contains the $gtrC^{\mathrm{BTP1}}$ gene, which is involved in O-antigen modification and prevents superinfection with phages that use the O-antigen as a co-receptor (Kintz et al., 2015). The presence of $gtrC^{\mathrm{BTP1}}$ within the DT56 variant strain could in part explain why it does not contain the ST160 phage and how the DT56 variant strain superseded DT160, as DT160-specific anti-O antibodies may not cross-react against DT56-variant anti-O antibodies. Serological studies could determine if anti-DT160 and -DT56 variant antibodies cross-react.

The novel plasmid identified in one of the DT56 variant isolates contained a large number of proteins of unknown function. Phenotypic tests involving strains with and without these the novel plasmid could help determine the functions of these proteins and the plasmid as a whole.

## 5.6 Conclusion

In this study, the evolution and genomics of *S.* Typhimurium DT160 and DT56 variant were described. Genomic analyses suggest that the two strains are distantly related and outlined possible reasons for the

emergence of DT56 variant and decline of DT160: the DT56 variant strain contained a BTP1 prophage that possibly made it serologically distinct from DT160 and/or resistant to ST160, and DT56 variant may replicate at a faster rate.

# References


Altschul, S. F., Gish, W., Miller, W., Meyers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215(3):403–410.

Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A., and Pevzner, P. A. (2016). plasmidSPAdes: assembling plasmds from whole genome sequencing data. *Bioinformatics*, 32(22):3380–3387.

Aronesty, E. (2013). Comparison of sequencing utility programs. *Open Bioinformatics Journal*, 7(1):1–8.

Aviv, G., Rahav, G., and Gal-Mor, O. (2016). Horizontal transfer of the *Salmonella enterica* serovar Infantis resistance and virulence plasmid pESI to the gut microbiota of warm-blooded hosts. *mBio*, 7(5):1–12.

Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008). The RAST server: Rapid annotations using subsystems technology. *BMC Genomics*, 9(75):1–15.

Baker, S., Hardy, J., Sanderson, K. E., Quail, M., Goodhead, I., Kingsley, R. A., Parkhill, J., Stocker, B., and Dougan, G. (2007). A novel linear plasmid mediates flagellar variation in *Salmonella* Typhi. *PLoS pathogens*, 3(5):1–7.

Bäumler, A. J., Hargis, B. M., and Tsolis, R. M. (2000). Tracing the origins of *Salmonella* outbreaks. *Science*, 287(5450):50–52.

Collard, J. M., Bertrand, S., Dierick, K., Godard, C., Wildemauwe, C., Vermeersch, K., Duculot, J., Van immerseel, F., Pasmans, F., Imberechts, H., and Quinet, C. (2008). Drastic decrease of *Salmonella* Enteritidis isolated from humans in Belgium in 2005, shift in phage types and influence on foodborne outbreaks. *Epidemiology and Infection*, 136(6):771–781.

Cox, M. P., Peterson, D. A., and Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*, 11(485):1–6.

Davis, J. J., Gerdes, S., Olsen, G. J., Olson, R., Pusch, G. D., Shukla, M., Vonstein, V., Wattam, A. R., and Yoo, H. (2016). PATtyFams: Protein families for the microbial genomes in the PATRIC database. *Frontiers in Microbiology*, 7(FEB):1–12.

Drummond, A., Suchard, M., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8):1969–1973.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5):699–710.

ESR (2003). Lablink Annual Summaries, 2000-2003. Retrieved 2016-09-27, from: `https://surv.esr.cri.nz/PDF_surveillance/Lablink/`, 1–1.

French, N., Pleydell, E., Marshall, J., Carter, P., Thornley, C., and Campbell, D. (2013). Source attribution of salmonellosis using microbial subtyping. Retrieved 2017-01-18, from: `https://www.mpi.govt.nz/document-vault/13137%0A`, 1–1.

Gardner, S. N., Slezak, T., and Hall, B. G. (2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics*, 31:2877–2878.

Garrison, E. and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. Retrieved 2017-02-02, from: `https://arxiv.org/abs/1207.3907`, 1–1.

Gebreyes, W. A., Thakur, S., Dorr, P., Tadesse, D. A., Post, K., and Wolf, L. (2009). Occurrence of spvA virulence gene and clinical significance for multidrug-resistant *Salmonella* strains. *Journal of Clinical Microbiology*, 47(3):777–780.

Gonzalo-Asensio, J., Ortega, A. D., Rico-Pérez, G., Pucciarelli, M. G., and García-Del Portillo, F. (2013). A novel antisense RNA from the *Salmonella* virulence plasmid pSLT expressed by non-growing bacteria inside eukaryotic cells. *PloS one*, 8(10):1–14.

Gorski, L. (2012). Selective enrichment media bias the types of *Salmonella enterica* strains isolated from mixed strain cultures and complex enrichment broths. *PLoS ONE*, 7(4).

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8):1072–1075.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2):160–174.

He, Z. L., Zhang, H. K., Gao, S. H., Lercher, M. J., Chen, W. H., and Hu, S. N. (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Research*, 44:W236–W241.

Herrero, A., Mendoza, M. C., Rodicio, R., and Rodicio, M. R. (2008). Characterization of pUO-StVR2, a virulence-resistance plasmid evolved from the pSLT virulence plasmid of *Salmonella enterica* serovar typhimurium. *Antimicrobial Agents and Chemotherapy*, 52(12):4514–4517.

Inouye, M., Dashnow, H., Raven, L., Schultz, M., Pope, B., Tomita, T., Zobel, J., and Holt, K. (2014). SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Medicine*, 6(11):1–16.

Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and their Applications*, 13(3):235–248.

Kingsley, R. A., Msefula, C. L., Thomson, N. R., Kariuki, S., Holt, K. E., Gordon, M. A., Harris, D., Clarke, L., Whitehead, S., Sangal, V., Marsh, K., Achtman, M., Molyneux, M. E., Cormican, M., Parkhill, J., MacLennan, C. A., Heyderman, R. S., and Dougan, G. (2009). Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Research*, 19(12):2279–2287.

Kintz, E., Davies, M. R., Hammarlöf, D. L., Canals, R., Hinton, J. C. D., and van der Woude, M. W. (2015). A BTP1 prophage gene present in invasive non-typhoidal *Salmonella* determines composition and length of the O-antigen of the lipopolysaccharide. *Molecular Microbiology*, 96(2):263–275.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21):2987–2993.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754–1760.

Lu, Z., Mitchell, R., Smith, R., Karns, J., van Kessel, J., Wolfgang, D., Schukken, Y., and Grohn, Y. (2013). Invasion and transmission of *Salmonella* kentucky in an adult dairy herd using approximate bayesian computation. *BMC Veterinary Research*, 9. cited By 3.

Luo, Y., Kong, Q., Yang, J., Golden, G., Wanda, S.-Y., Jensen, R. V., Ernst, P. B., and Curtiss, R. (2011). Complete genome sequence of the universal killer *Salmonella enterica* serovar typhimurium UK-1 (ATCC 68169). *Journal of Bacteriology*, 193(15):4035–4036.

Magwedere, K., Rauff, D., De Klerk, G., Keddy, K., and Dziva, F. (2015). Incidence of nontyphoidal *Salmonella* in food-producing animals, animal feed, and the associated environment in South Africa, 2012-2014. *Clinical Infectious Diseases*, 61:S283–S289.

Mather, A., Reid, S., Maskell, D., Parkhill, J., Fookes, M., Harris, S., Brown, D., Coia, J., Mulvey, M., Gilmour, M. o., Petrovska, L., De Pinna, E., Kuroda, M., Akiba, M., Izumiya, H., Connor, T., Suchard, M. l., Lemey, P., Mellor, D., Haydon, D., and Thomson, N. (2013). Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science*, 341(6153):1514–1517.

Mather, A. E., Lawson, B., de Pinna, E., Wigley, P., Parkhill, J., Thomson, N. R., Page, A. J., Holmes, M. A., and Paterson, G. K. (2016). Genomic analysis of *Salmonella enterica* serovar Typhimurium from wild passerines in England and Wales. *Applied and Environmental Microbiology*, 82(22):6728–6735.

McClelland, M., Sanderson, K. E., Spieth, J., Clifton, S. W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R., and Wilson, R. K. (2001). Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature*, 413(6858):852–856.

McWhorter, A. R., Davos, D., and Chousalkar, K. K. (2015). Pathogenicity of *Salmonella* strains isolated from egg shells and the layer farm environment in Australia. *Applied and Environmental Microbiology*, 81(1):405–414.

Mira, A., Ochman, H., and Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends in Genetics*, 17(10):589–596.

Ogunremi, D., Blais, B., Huang, H., Wang, L., Elmufti, M., Allain, R., Hazelwood, J., Grenier, C., Amoako, K., Savic, M., and Ghazi, N. F. (2017). Draft genome sequences of two strains of *Salmonella enterica* serovar Typhimurium displaying different virulence in an experimental chicken model. *Genome Announcements*, 5(6):1–2.

Okoro, C. K., Kingsley, R. A., Connor, T. R., Harris, S. R., Parry, C. M., Al-Mashhadani, M. N., Kariuki, S., Msefula, C. L., Gordon, M. A., De Pinna, E., Wain, J., Heyderman, R. S., Obaro, S., Alonso, P. L., Mandomando, I., MacLennan, C. A., Tapia, M. D., Levine, M. M., Tennant, S. M., Parkhill, J., and Dougan, G. (2012). Intracontinental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. *Nature Genetics*, 44(11):1215–1221.

Price-Carter, M., Roy-Chowdhury, P., Pope, C. E., Paine, S., De Lisle, G. W., Collins, D. M., Nicol, C., and Carter, P. E. (2011). The evolution and distribution of phage ST160 within *Salmonella enterica* serotype Typhimurium. *Epidemiology and Infection*, 139(8):1262–1271.

Public Health Surveillance (2017a). 2003-2016 Human Salmonella serotypes. Retrieved 2017-05-07, from: `https://surv.esr.cri.nz/enteric_reference/human_salmonella.php`, 1–1.

Public Health Surveillance (2017b). Non-human Salmonella isolates, 2003-2016. Retrieved 2017-05-07, from: `https://surv.esr.cri.nz/enteric_reference/nonhuman_salmonella.php`, 1–1.

Qiagen (2012). QIAamp DNA mini and blood mini handbook. Retrieved 2017-06-29, from: `\nhttps://www.qiagen.com/us/resources/resourcedetail?id=62a200d6-faf4-469b-b50f-2b59cf738962&lang=en`, 33-56.

Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014). Tracer 1.6. Retrieved 2016-09-27, from: `http://beast.bio.ed.ac.uk/Tracer`, 1–1.

Riley, M. and Anilionis, A. (1978). Evolution of the bacterial genome. *Annual Review of Microbiology*, 32:519–560.

Sears, A., Baker, M. G., Wilson, N., Marshall, J., Muellner, P., Campbell, D. M., Lake, R. J., and French, N. P. (2011). Marked campylobacteriosis decline after interventions aimed at poultry, New Zealand. *Emerging Infectious Diseases*, 17(6):1007–1015.

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30(14):2068–2069.

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30):11030–11035.

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28(1):33–36.

Tavechio, A. T., Fernandes, S. A., Neves, B. C., Dias, A. M. G., and Irino, K. (1996). Changing patterns of *Salmonella* serovars: Increase of *Salmonella enteritidis* in São Paulo, Brazil. *Revista do Instituto de Medicina Tropical de Sao Paulo*, 38(5):315–322.

Thornley, C., Simmons, G., Callaghan, M., Nicol, C., Baker, M., Gilmore, K., and Garrett, N. (2003). First incursion of *Salmonella enterica* serotype Typhimurium DT160 into New Zealand. *Emerging Infectious Diseases*, 9(4):493–495.

Van Duijkeren, E., Wanner, W. J. B., Houwers, D. J., and Van Pelt, W. (2002). Serotype and phage type distribution of *Salmonella* strains isolated from humans, cattle, pigs, and chickens in The Netherlands from 1984 to 2001. *Journal of Clinical Microbiology*, 40(11):3980–3985.

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., Gillespie, J. J., Gough, R., Hix, D., Kenyon, R., MacHi, D., Mao, C., Nordberg, E. K., Olson, R., Overbeek, R., Pusch, G. D., Shukla, M., Schulman, J., Stevens, R. L., Sullivan, D. E., Vonstein, V., Warren, A., Will, R., Wilson, M. J. C., Yoo, H. S., Zhang, C., Zhang, Y., and Sobral, B. W. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research*, 42(D1):D581–D591.

Xu, T., Maloy, S., and McGuire, K. (2009). Macrophages influence *Salmonella* host-specificity in vivo. *Microbial Pathogenesis*, 47(4):212–222. cited By 8.

Yang, J., Barrila, J., Roland, K. L., Kilbourne, J., Ott, C. M., Forsyth, R. J., and Nickerson, C. A. (2015). Characterization of the invasive, multidrug resistant non-typhoidal *Salmonella* strain D23580 in a murine model of infection. *PLoS Neglected Tropical Diseases*, 9(6):1–17.

# Chapter 6

# Long-term colonisation by *Campylobacter jejuni* within a human host: evolution, antimicrobial resistance and adaptation

## 6.1 Abstract

Campylobacteriosis is inflammation of the gastrointestinal tract as a result of *Campylobacter* infection. Most campylobacteriosis cases are acute and self-limiting, with symptoms and *Campylobacter* excretion ceasing after a few weeks. Through routine sentinel-site surveillance a common variable immune deficiency patient was identified that had tested faecally-positive for the same multilocus sequence type, *Campylobacter jejuni* (ST45), intermittently for ten years. In order to determine whether this was likely to be a persistent infection, or multiple reinfections, sixteen *Campylobacter* isolates collected from the patient were whole genome sequenced and had their antimicrobial susceptibility patterns and motility determined. Phylogenetic analyses estimated that the isolates shared a date of common ancestor between the years 1998-2006, coinciding with the onset of symptoms for the patient. Genomic analysis identified selection for changes in motility, and antimicrobial susceptibility testing suggested that the *Campylobacter* population developed resistance to several antibiotics coinciding with periods of antibiotic therapy. This indicated that the patient was consistently colonised with *Campylobacter* that adapted to the internal environment of the patient. The results also demonstrated how genomic and phylogenetic analyses can give insight into a patient's infection history and the effect of antimicrobial treatment on *Campylobacter* populations in this unusual situation of long-term colonisation of an individual.

## 6.2 Introduction

*Campylobacter* are the prime bacterial cause of human enteritis (WHO, 2011). Campylobacteriosis is associated with diarrhoea, abdominal pain and fever. Most campylobacteriosis cases are acute and self-limiting, with symptoms ceasing after a week (Kirkpatrick et al., 2013). Campylobacteriosis patients begin excreting *Campylobacter* at the onset of symptoms and continue to excrete for 2-3 weeks after symptoms subside (Porter and Reid, 1980). Longer periods of excretion can occur in a number of conditions. In cases of

chronic campylobacteriosis, patients are unable to clear Campylobacter, and suffer from continuous episodes of diarrhoea and Campylobacter-positive faeces (Molina et al., 1995; Pignata et al., 1984). Cases of reinfection with Campylobacter can arise from subsequent exposure and result in symptomatic or asymptomatic campylobacteriosis (Tribble et al., 2010). Asymptomatic campylobacteriosis cases are faecally-positive for Campylobacter but do not display any symptoms. They are more prevalent in developing countries, due to increased exposures and, possibly, under nutrition-induced immunosuppression (Lee et al., 2013).

Common variable immune deficiency (CVID) is a primary immunosuppressive disease characterized by reduced serum IgA and IgG concentrations (Conley et al., 1999). CVID patients have a higher risk of bacterial infections, including *Campylobacter*, and non-infectious gastrointestinal diseases (Resnick et al., 2012). However, *Campylobacter* has been isolated from CVID patients with and without diarrhoea (Dionisi et al., 2011).

The Manawatu sentinel surveillance site (MSSS) at $^{m}$EpiLab, Massey University, in collaboration with Mid-Central District Health Board (MCDHB), collects *Campylobacter* isolates from human campylobacteriosis cases, food and environmental sources in the Mid-Central region of New Zealand's North Island. Isolates are genotyped, allowing inferences to be made about the sources of human cases (Muellner et al., 2011). The MSSS identified an individual who tested faecally-positive for *Campylobacter jejuni* ST45 seven times over ten years (2006-2016). The aim of this study was to determine whether the patient was persistently infected or re-infected on multiple occasions using whole genome sequencing and phenotypic testing.

## 6.3 Methods

### 6.3.1 Ethics

This study was approved by Central Health and Disease Ethics Committee (16/CEN/13).

### 6.3.2 Interview

The patient was interviewed to identify potential occupational and domestic sources of *Campylobacter* and to gain further details of their medical history. The patient stated that they had been on multiple medications since the onset of diarrhoea. After obtaining consent from the patient, their medical records were obtained and investigated for antibiotic use.

### 6.3.3 Environmental sampling

Potential domestic sources identified in the interview were cultured for *Campylobacter*. Faecal samples were obtained from the patients pets, soil samples from their vegetable gardens and water samples from their

water supply. Faecal samples were directly plated onto two CAT and two mCCDA agar plates (Fort Richard, Auckland, New Zealand). One set was incubated microaerobically (3% $O_2$, 5% $H_2$, 10% $CO_2$, 82% $N_2$) at 37°C for 48 hours and the other was incubated microaerobically (5% $O_2$, 10% $CO_2$, 85% $N_2$) at 42°C for 48 hours to promote the growth of *Campylobacter* that grow at 37°C, 42°C and require hydrogen gas. In addition, approximately 100 g of each faecal sample was inoculated in 10 ml of Bolton broth (Lab M Ltd., Heywood, Bury, UK) and incubated microaerobically (5% $O_2$, 10% $CO_2$, 85% $N_2$) at 42°C for 48 hours.

Approximately 1 g of each soil sample was inoculated in 10 ml of Bolton broth and incubated microaerobically (5% $O_2$, 10% $CO_2$, 85% $N_2$) at 42°C for 48 hours. A litre of the water sample was filtered through a 0.45 µm nitrocellulose filter (Merck Millipore, Billerica, Massachusetts, USA). The filter was inoculated in 10 ml of Bolton broth and incubated microaerobically (5% $O_2$, 10% $CO_2$, 85% $N_2$) at 42°C for 48 hours.

From each Bolton broth, a loopful was inoculated onto two mCCDA and two CAT agar plates. One set was incubated microaerobically (3% $O_2$, 5% $H_2$, 10% $CO_2$, 82% $N_2$) at 37°C for 48 hours and the other was incubated microaerobically (5% $O_2$, 10% $CO_2$, 85% $N_2$) at 42°C for 48 hours. All cultured oxidase-positive, Gram-negative colonies underwent species typing by use of a *C. jejuni* PCR.

### 6.3.4 PCR

A loopful of 24-hour *Campylobacter* growth was inoculated into an Eppendorf tube containing 1 ml of 2% Chelex (Bio-Rad, Hercules, California, USA) and incubated at 100°C for 10 minutes. After incubation, each tube was centrifuged for 3 minutes at 12,000 × g. 2 $\mu$l of supernatant was removed from each tube and used for PCR. *C. jejuni* was tested for using a PCR method that tests for the presence of the *hipO* gene (Wang et al., 2002).

### 6.3.5 Strains

Sixteen ST45 isolates were collected from seven episodes of diarrhoea reported to the MSSS from the same patient (2006-2016). *C. jejuni* NCTC 11168 and NCTC 11531 were obtained from the New Zealand Reference Culture Collection (ESR, Porirua, New Zealand). A non-motile *C. jejuni* strain, previously isolated from retail poultry through the MSSS, was used as a control in motility and chemotaxis assays (Friedrich, 2014).

### 6.3.6 Whole genome sequencing

Genomic DNA was extracted from the 16 ST45 isolates using a QIAamp DNA mini kit (Qiagen, Hilden, Germany) (Qiagen, 2012). Genome extracts were whole genome sequenced by New Zealand Genomics Limited (NZGL) at Massey Genome Services (Massey University, Palmerston North, New Zealand). A library was

prepared for each isolate using an Illumina TruSeq™DNA PCR-Free kit (Illumina, Scorsby, Victoria, Australia) and sequenced using an lllumina MiSeq (Illumina, Scorsby, Victoria, Australia) as 2x250 bp paired-end runs (120-150x average genome coverage). After sequencing and barcode demultiplexing, PhiX control library reads and adapter sequences were removed using FASTQ-MCF (Aronesty, 2013). The raw reads are available at the European Nucleotide Archive (http://www.ebi.ac.uk/ena; accession number: PRJEB18520).

### 6.3.7 Genomic assembly

Genome were assembled *de novo*. An in-house Perl script trimmed reads at an error probability of 0.01 using solexaQA++ (Cox et al., 2010) and generated random subsets of paired reads from 500,000 to 800,000 paired reads in 100,000 increments, varying average genome coverage. Each random sets was assembled with the *de novo* assembler Velvet v1.1 (Zerbino and Birney, 2008) using kmers from 55 to 245 in increments of 10. This resulted in multiple genome assemblies for each isolate. Metrics for each of four parameters (longest genome length, fewest number of contigs, largest $N_{50}$ value, and longest contig length) were ranked in numerical order, and an overall equally summed ranking score was calculated. The assemblies with the lowest total rank were used. QUAST (Gurevich et al., 2013) was used to analyse the ST45 *de novo* assembly GC contents.

### 6.3.8 Single nucleotide polymorphisms

Core single nucleotide polymorphisms (SNPs) were identified using kSNP v3.0 (Gardner et al., 2015) and Snippy v2.6 (https://github.com/tseemann/snippy). The Snippy pipeline uses the Burrows-Wheelers Aligner (Li and Durbin, 2009) and SAMtools (Li, 2011) to align reads from different isolates to a sequence and FreeBayes (Garrison and Marth, 2012) to identify variants among the alignments. kSNP was used to analyse *de novo* assembled genomes, along with the reference genome, *Campylobacter jejuni* str. 4031 (NC_022529). An in-house Python script was used to determine the read coverage of all the SNPs identified via kSNP. Snippy was used to align reads from each isolate to the reference genome, NC_022529, before identifying SNPs. SNPs were accepted if they had a greater than 10 read depth and a greater than 90% consensus for each isolate. The position of the SNP on the reference genome was used to determine if both methods identified the SNP or if they were unique to the method (Appendix D.1). BRIG (Alikhan et al., 2011) was used to compare the *de novo* assembled genomes to the reference (Appendix D.1).

### 6.3.9 ST45 comparison

Four ST45 isolates from the MSSS were previously sequenced (Fearnhead et al., 2014) and uploaded to BIGSdb (https://pubmlst.org/software/database/bigsdb). These isolates were collected from 2005-2008 from

poultry (P104a and P544b), wild bird (R68c) and bovine (S150a) sources. The *de novo* assembled genomes of these isolates were downloaded from BIGSdb (access available on request) and compared with the *de novo* assembled genomes of the 16 ST45 isolates using kSNP3.

### 6.3.10 NeighborNet trees

SplitsTree (Huson and Bryant, 2006) was used to form a NeighborNet tree of the 16 ST45 isolates collected from the same patient based on the 196 core SNPs they share, and the 20 ST45 isolates from the patient and other sources based on the 5,216 core SNPs they share.

### 6.3.11 Phylogenetic analysis

An in-house Perl script was used to split the 196 core SNPs into groups based on whether they were associated with the 1st, 2nd or 3rd codon, contained in overlapping coding regions, or found in intergenic regions. It was also used to determine if the SNPs were synonymous or non-synonymous. The partitioned SNPs were exported into BEAUti to create an Extended Markup Language (XML) file for BEAST 1.8.3 (Drummond et al., 2012). The five SNP groups were given separate Hasegawa Kishino Yano (HKY) substitution models (Hasegawa et al., 1985), while their tree and clock models were linked to allow for variation in base substitution among codon positions. The Gaussian Markov Random Field (GMRF) Bayesian Skyride model (Minin et al., 2008) was used to allow for and estimate changes in the effective population size. An uncorrelated relaxed molecular clock (Drummond et al., 2006) was used to allow for variation in mutation rates among lineages and was calibrated by tip dates. The XML file was run in BEAST for 40 million steps, three times with different starting seeds, before LogCombiner was used to combine the runs with a 10% burn-in. Tracer v1.6 (Rambaut et al., 2014) was used to visualise the results and the relative change in effective population size.

The GMRF Bayesian Skyride model was re-run without sequencing data to determine if the model was picking up a signal or simply sampling from priors. The mean mutation rate estimated by BEAST was multiplied by the number of core SNPs analysed (196 bp) and divided by the mean genome size of the isolates analysed (1,641,217 bp), to give the mutation rate for the ST45 genome.

### 6.3.12 Protein-coding gene analysis

Proteins and protein differences shared by the 16 ST45 isolates were identified using a previously described method (Chapter 3). The functions of the proteins were predicted by comparing their sequences to those in the Clusters of Orthologous Groups of proteins (COGs) database (Tatusov et al., 2000). The number of protein differences in each functional group was divided by the total number of ST45 proteins in each functional group to give the protein difference to quantity ratio.

### 6.3.13 Genome degradation

The *de novo* assembled genomes of the 16 ST45 isolates collected from the patient were uploaded to PATRIC (Wattam et al., 2014), along with the four *C. jejuni* ST45 isolates that had been sequenced as part of the MSSS, and were annotated using RAST (Aziz et al., 2008). PLfams (Davis et al., 2016) was used to cluster these 20 *de novo* assembled genomes, along with the ST45 reference genome, str. 4031 (NC_022529). All genes that were found once within all 21 genomes and differed in length were extracted from PATRIC and their positions within the genomes were used to determine if the genes differed in length or were the result of assembly gaps. Pseudogenes were classified as genes that were less than 90% the length of those found in the other isolates analysed.

### 6.3.14 Antimicrobial susceptibility

Trek Sensititre™plates (Thermo Fisher Scientific, Waltham, Massachusetts, USA) were used to measure each isolates minimum inhibitory concentration (MIC) for: erythromycin, ciprofloxacin, tetracycline, gentamycin, nalidixic acid and streptomycin (TREK, 2012). E-tests strips (Liofilchem, Roseto degli Abruzzi, Italy) were used to measure MICs for: ampicillin, sulfamethoxazole and trimethoprim (Liofilchem, 2014). EUCAST (EUCAST, 2016) breakpoints were used to classify the *Campylobacter* isolates as susceptible or resistant.

### 6.3.15 Motility

*Campylobacter* motility was measured based on a method described by Hänel et al. (2009). Briefly, motility media was made by adding 5.6 g of Brucella broth (Becton Dickinson, Franklin Lakes, New Jersey, USA) and 0.8 g of agar (Merck, Darmstadt, Germany) to 200 ml of MilliQ water. The media was autoclaved and left to cool, before 1 ml of 2% 2,3,5-tetrazolium chloride (Sigma-Aldrich, St. Louis, Missouri, USA) was added and the media was poured into petri dishes.

Single day *Campylobacter* growth was inoculated into Mueller-Hinton broth (Fort Richard, Auckland, New Zealand) and incubated microaerobically (3% $O_2$, 5% $H_2$, 10% $CO_2$, 82% $N_2$) at 37°C for 24 hours. Cultured broths were diluted to a 0.5 McFarland standard with additional Mueller-Hinton broth, before 1 µl of the growth was inoculated into the centre of a motility agar plate and incubated microaerobically at 37°C for 24 hours. The distance travelled through the motility media was measured with callipers.

The *Campylobacter* isolates from the first six episodes had their motility tested 6-8 times, whilst the 10 isolates collected from the final episode and the controls had their motility tested 3-5 times. The first six isolates displayed more variation in motility compared to the other isolates so were repeated more.

### 6.3.16 Chemotaxis

*Campylobacter* chemotaxis was measured using a disc and hard agar plug method. The disc diffusion method was based on a method described by Hugdahl et al. (1988). Briefly, 100 mM citrate, 100 mM L-serine, 100 mM pyruvate and 250 mM deoxycholate were formed by adding 0.2491 g of trisodium citrate (BDH, Poole, England), 0.1051 g of L-serine (Sigma-Aldrich, St. Louis, Missouri, USA), 0.1101 g of sodium pyruvate (BDH, Poole, England), and 1.0365 g of deoxycholate (Sigma-Aldrich, St. Louis, Missouri, USA), to separate 10 ml volumes of 10 mM pH 7.4 PBS. For each chemical and 10 mM pH 7.4 PBS, 6 mm filter discs were inoculated with 20 µl volumes. 24 hour *Campylobacter* growth was added to 10 ml of 40°C molten 0.4% PBS agar, adjusted to a 0.9 absorbance and left to set in a petri dish. Inoculated filter discs were placed onto the set agar, before the agar was incubated microanaerobically (3% $O_2$, 5% $H_2$, 10% $CO_2$, 82% $N_2$) at 37°C for 10 hours. The distance between the disc and the edge of the halo of growth was measured with callipers.

The hard agar plug method was based on a method described by Algamoudi and Ketley (2015). Briefly, 100 mM pyruvate hard plugs were formed by adding 7 ml of 214.3 mM sodium pyruvate in 10 mM pH 7.4 PBS to 0.5 ml of 1% tetrazolium chloride and 7.5 ml of 4% PBS agar. The mixture was left to set in a petri dish, before 5-6 mm diameter hard plugs were removed. Negative controls were formed by repeating the procedure but replacing sodium pyruvate with 10 mM pH 7.4 PBS. 24 hour *Campylobacter* growth was added to 7.5 ml of 10 mM pH 7.4 PBS and adjusted to a 0.5 absorbance. 7.5 ml of 50°C molten 0.8% PBS agar was added to the inoculated PBS, before the mixture was placed in a petri dish containing a set of hard plugs and left to set. Set agar was incubated microanaerobically (3% $O_2$, 5% $H_2$, 10% $CO_2$, 82% $N_2$) at 37°C for 20 hours. The distance between the disc and the edge of the halo of growth was measured with callipers.

### 6.3.17 Scripts

The in-house scripts used in this study were previously described (Chapter 3). They are available from GitHub (https://github.com/samuelbloomfield/Scripts-for-genomic-analyses).

## 6.4 Results

### 6.4.1 Patient background

In 1992, the patient was diagnosed with common variable immune deficiency (CVID), an immune disorder associated with an increased risk of bacterial infections (Resnick et al., 2012), and in 2000 they started to suffer from daily episodes of diarrhoea. Since 1992, they had also been on multiple antibiotic and immunosuppressive therapies for unrelated disorders. From 2000 to 2016, all faecal samples collected from the patient

tested positive for *Campylobacter*. Faecal specimens tested negative for *Shigella*, *Salmonella*, *Yersinia enterocolitica*, *Giardia* and *Cryptococcus* during this time. Endoscopies, gastroscopies and gastrointestinal biopsies performed from 2000-2016 varied considerably and were consistent with a wide range of gastrointestinal disorders at various times.

The patient had multiple jobs since 2000, making it difficult to identify occupational sources. Domestic samples obtained from the patients pets, vegetable gardens and water source all tested negative for *C. jejuni*.

### 6.4.2 *Campylobacter jejuni* isolates

Sixteen *C. jejuni* ST45 isolates were cultured from seven faecal samples from 2006-2016. One *Campylobacter* isolate was collected from each of the first six faecal samples (2006-2013), whilst ten were collected from the last sample to assess genomic variability at a single time point (2016).

### 6.4.3 Genomic ST45 comparison

Assembled genomes were 1.64-1.66 Mb in length and contained a GC content of 30.45-30.62%. This is within the normal GC content range for *C. jejuni*: 28-33% (Vandamme et al., 2005).

### 6.4.4 Genetic distance

196 core SNPs were identified among the closely related 16 ST45 isolates collected from the patient, with no genetically distinct isolates observed (Figure 6.1). There was an association between the date of collection of the isolates and genetic distance, with a larger genetic distance between isolates collected at distant time points than those collected at closer time points. There was some diversity in the 10 isolates collected from the same episode in 2016, but this was relatively small compared to the diversity between isolates collected at different time points.

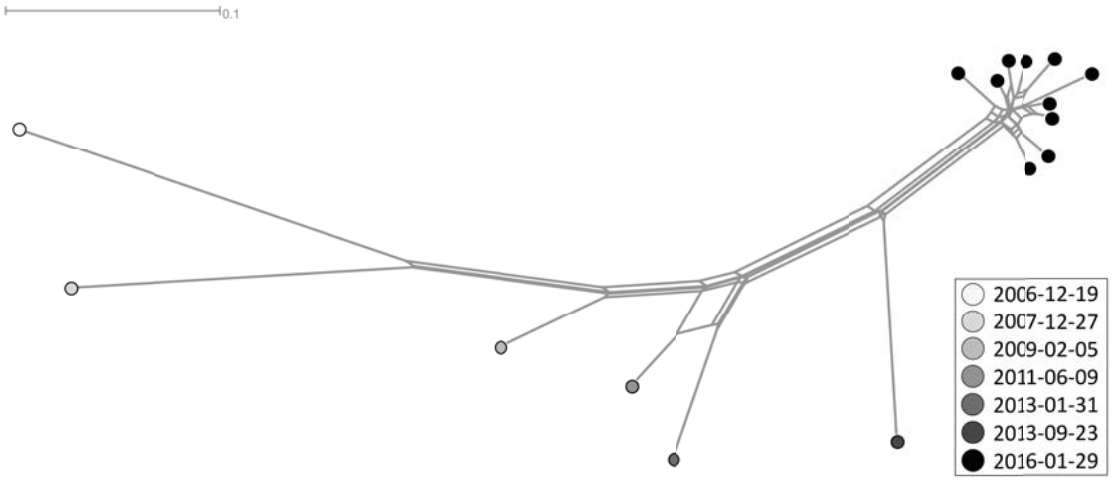


**Figure 6.1.** NeighborNet tree of 16 ST45 isolates collected from the same patient, based on 196 core SNPs. The isolates are coloured by their date of collection and the scale bar represents the number of nucleotide substitutions per site.

### 6.4.5 Phylogenetic analysis

Phylogenetic analysis predicted that the 16 ST45 isolates shared a mean date of common ancestor in June 2002 (95% HPD interval: July 1998 - January 2006). The estimated mutation rate was 2.8-8.3 x $10^{-6}$ substitutions $site^{-1}$ $year^{-1}$ (95% HPD interval). The GMRF Bayesian Skyride model was unable to identify any changes in effective population size (Appendix D.2).

### 6.4.6 ST45 comparison

The four MSSS ST45 genomes from non-human sources were distinct from the 16 isolates collected from the patient (Figure 6.2). However, they represent a small sample size. Further genomic analysis with ST45 isolates collected from the MSSS are required to determine likely sources for the patient. The raw reads for the four MSSS ST45 isolates were unobtainable, preventing Snippy and SNP coverage analysis. However, kSNP3 analysis was sufficient for determining how the 16 ST45 isolates collected from the same patient relate to these isolates.

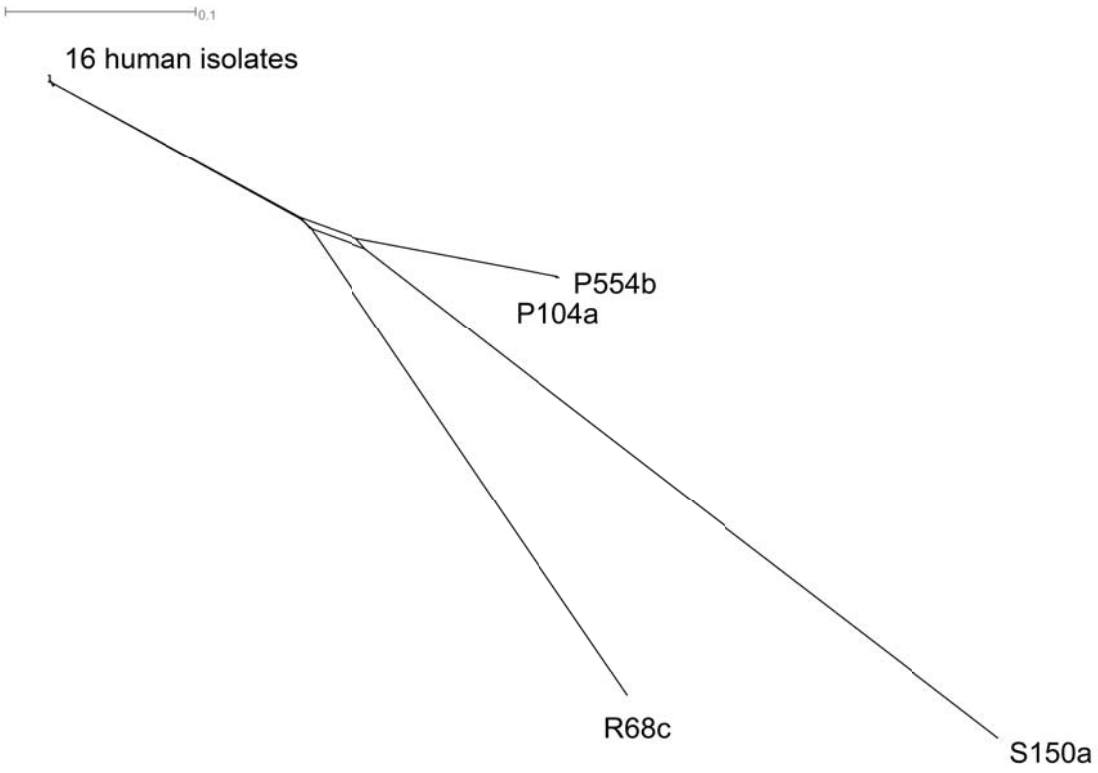


**Figure 6.2.** NeighborNet tree of 16 ST45 isolates collected from the same patient and four ST45 isolates collected from non-human sources, based on 5,216 core SNPs. The scale bar represents the number of nucleotide substitutions per site.

### 6.4.7 Protein-coding gene analysis

Protein annotation identified 1,785 coding-DNA sequences (CDS) contained among the 16 ST45 isolates, one of which was missing in one isolate. Protein analysis also identified 129 non-synonymous SNPs and 29 indels, altogether giving 159 protein differences among the 16 ST45 isolates. These 159 protein differences were associated with a large number of COG functional groups, but a disproportionate number were involved in cell motility (COG group N) and signal transduction (COG group T) (Figure 6.3).

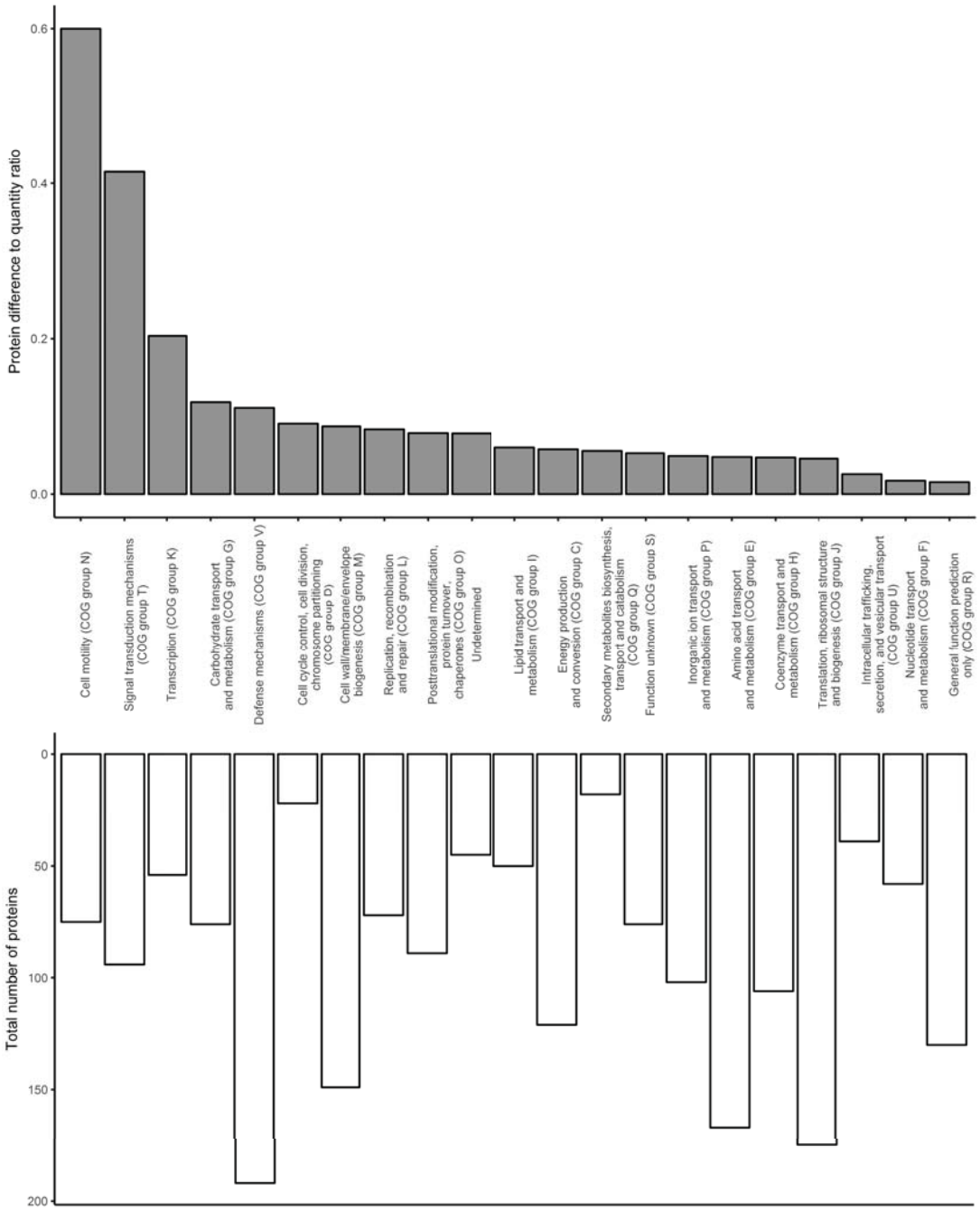


**Figure 6.3.** Bar graph of the protein difference to quantity ratio for the 159 protein differences shared by 16 ST45 isolates, for each COG functional group (grey), and the total number of proteins belonging to each COG functional group (white).

### 6.4.8 Genome degradation

The 16 ST45 isolates collected from the patient, four ST45 isolates from the MSSS and ST45 reference genome (NC_022529) contained a pan-genome of 1,908 CDS. Their core genome consisted of 1,427 CDS, and 151 of

these differed in size and were found only once within each isolate's genome. The isolates collected from the same patient contained 1-4 more peudogenes than the reference genome or the other isolates collected as part of the MSSS (Figure 6.4).

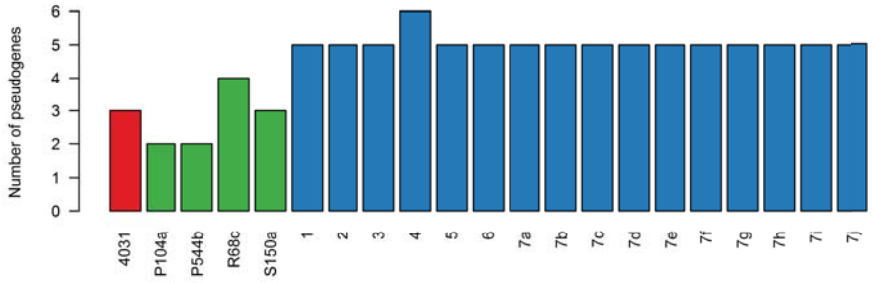


**Figure 6.4.** Barplot of the number of pseudogenes identified in the reference ST45 strain (red), ST45 isolates collected as part of the MSSS (green), and 16 ST45 isolate collected from the same patient (blue).

### 6.4.9 Antimicrobial susceptibility

From 2006-2016 the patient was on multiple courses of antibiotics for various conditions. However, the patient could not recall whether any of these courses alleviated diarrhoea.

The ST45 isolates became less susceptible to ampicillin between December 2007 and February 2009, prior to cefuroxime treatment in March 2009 and amoxicillin-clavulanic acid treatment in May 2009 (Figure 6.5). The medical records of the patient state that in September 2008 they mentioned being on antibiotics. However, the specific nature of these antibiotics was not reported. The ampicillin resistant ST45 isolates collected from February 2009 onwards contained a mutation in the promoter region of the blaOXA-61 gene. This gene encodes a beta-lactamase gene that degrades beta-lactam antibiotics and the mutation is associated with increased expression of the gene and high-level beta-lactam resistance (Zeng et al., 2010).

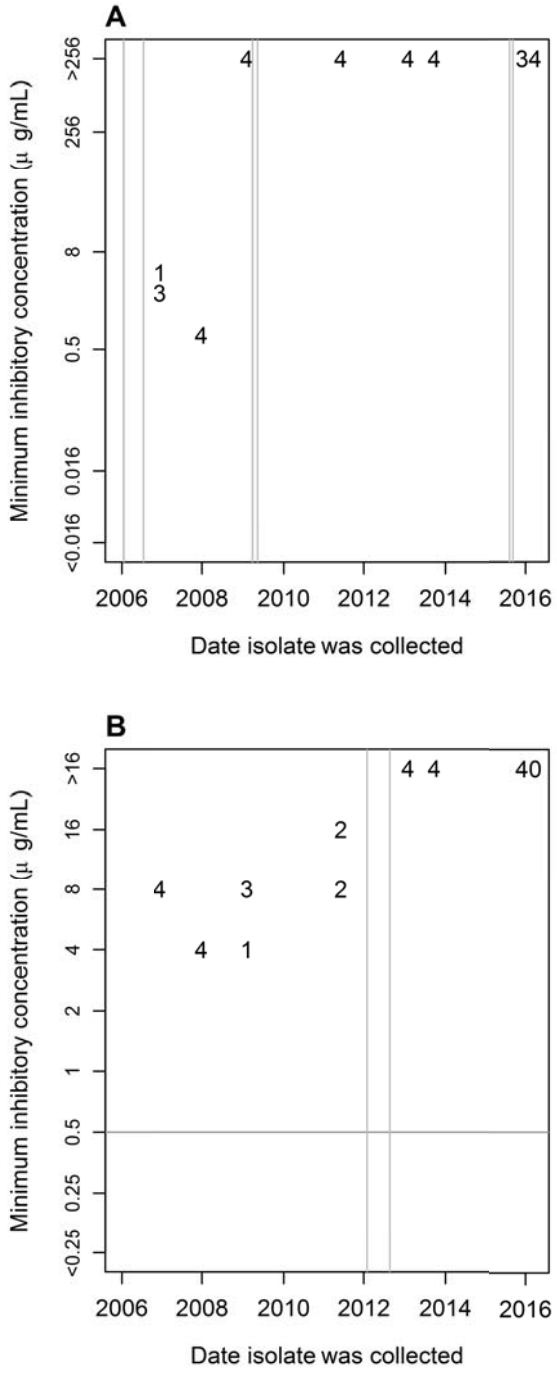


**Figure 6.5.** Scatterplot of date of collection versus the minimum inhibitory concentration to ampicillin (A) and ciprofloxacin (B), for 16 ST45 isolates collected from the same patient. The points represent the number of repeats that had the same value and the vertical lines represent times in the patients medical records when they were prescribed or mention taking antibiotics belonging to the same class. The horizontal line represents the EUCAST breakpoint (EUCAST, 2016).

All ST45 isolates collected from the patient were resistant to nalidixic acid and ciprofloxacin according to EUCAST breakpoints (EUCAST, 2016), as a result of a T86I mutation in the DNA gyrase A subunit gene (Wang et al., 1993). They also became more resistant to ciprofloxacin between June 2011 and January 2013, coinciding with ciprofloxacin treatment in January and August of 2012. These isolates contained an additional S460F amino acid change in their DNA gyrase B subunit gene. This mutation has not previously been reported in *Campylobacter*. An equivalent mutation (S464F) is associated with fluoroquinolone resistance in *Pseudomonas aeruginosa* (Mouneimné et al., 1999) and *Escherichia coli* (Kohanski et al., 2010).

The antimicrobial susceptibility profiles to the other antimicrobial agents tested were similar for the 16 ST45 isolates (Appendix D.4). All the isolates contained the A2058T mutation in the 23S rRNA gene associated with erythromycin resistance (van der Beek et al., 2010). The isolates were also resistant to sulfamethoxazole and trimethoprim by an unknown mechanism (Appendix D.4).

### 6.4.10 Motility

The first isolate collected from the patient (2006) was the most motile, whilst the second isolate (2008) was the least motile. The rest of the isolates displayed less variation in motility (Figure 6.6). There was no difference in motility between isolates collected from the same occurrence of diarrhoea (Figure 6.7).

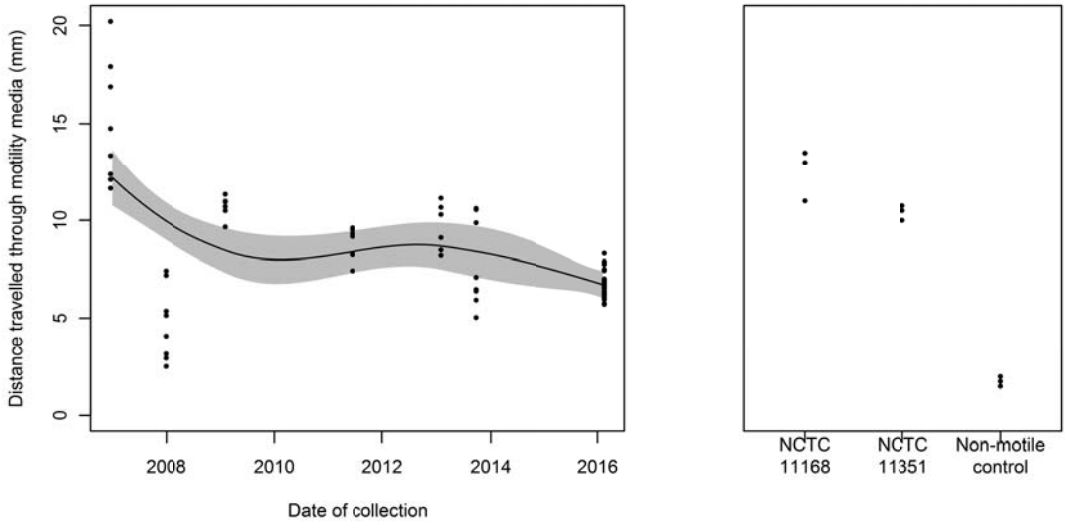


**Figure 6.6.** Scatterplot of date of collection versus distance travelled in motility agar for 16 ST45 isolates collected from the same patient, and the results for the control strains. The grey area represents the 95% confidence interval surrounding the Loess line.

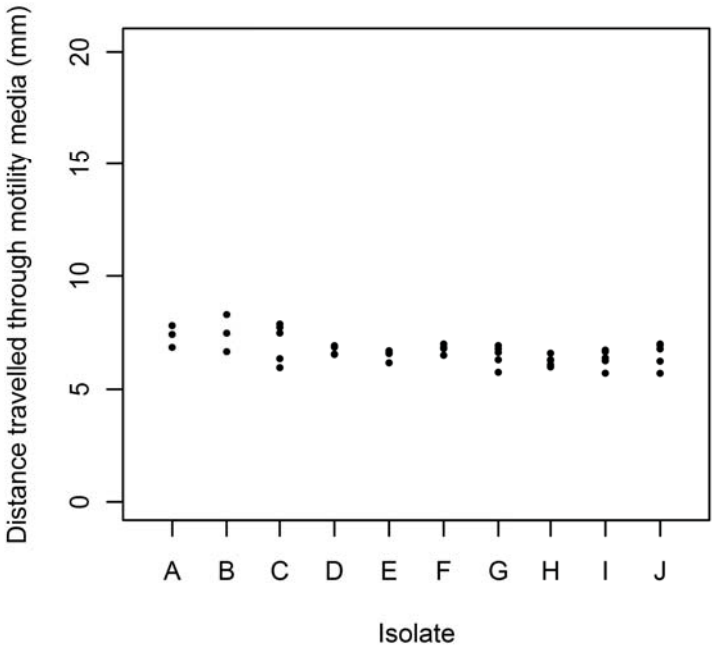


**Figure 6.7.** Scatterplot of the motility of 10 ST45 isolates collected from the same episode.

### 6.4.11 Chemotaxis

The results from the two chemotaxis assays trialled in this study were inconclusive, as the negative (non-motile) control displayed chemotaxis (Appendix D.5).

## 6.5 Discussion

Long-term *Campylobacter* excretion may arise from continued or recurring colonisation. Recurrent colonisation may result from exposure to multiple *Campylobacter* sources or continued exposure to a single source. The isolates collected from the patient were closely related, with a date of common ancestor coinciding with the date that the patient began suffering from daily episodes of diarrhoea. If the patient had been exposed to multiple *Campylobacter* sources, I would have expected the collected isolates to be more distantly related, with an earlier date of common ancestor. In addition, I was unable to identify any *Campylobacter* sources in the patients environment and the *Campylobacter* developed resistance to antibiotics during times that they were prescribed similar antibiotics. This suggests that the patient's long-term excretion of *Campylobacter* was the result of continued colonisation, and demonstrates how genomic and epidemiological analyses can help shed light on a patients infection history.

It is unknown how long-term *Campylobacter* excretion has affected the patient. Since diarrhoea began, all faecal samples collected from the patient have been positive for *Campylobacter*. This suggests a chronic infection. Previous studies on recurrent *Campylobacter* excretion were able to diagnose chronic campylobacteriosis by subsequent symptom alleviation following *Campylobacter* removal or antibiotic usage (Molina et al., 1995; Pignata et al., 1984). In the current study, the patient was continuously colonised with *Campylobacter* and it is unknown whether previous antibiotic usage alleviated symptoms. In addition, endoscopies, gastroscopies and gastrointestinal biopsies performed on the patient have at various times been consistent with a wide range of gastrointestinal disorders, and *Campylobacter* has been isolated from CVID patients, such as the patient, with and without gastrointestinal symptoms (Dionisi et al., 2011). Therefore, it is possible that the patient has an underlying gastrointestinal disorder that has allowed them to be colonised with *Campylobacter* for this length of time, similarly to chronic obstructive pulmonary disease (COPD) and persistent *P. aeruginosa* lung colonisation (Valderrey et al., 2010).

The mutation rate estimated for the 16 ST45 isolates was lower than the mutation rate estimated by Wilson et al. (2008) for *Campylobacter jejuni* and *coli* based on the sequences of several house-keeping genes. This may be because the ST45 isolates were under a large amount of selection pressure within the host, leaving little room for genetic variation; the house keeping genes investigated by Wilson et al. mutate at a faster rate compared to the rest of the genome; or individual *C. jejuni* strains may show a large amount of variation in mutation rates that are not reflected in Wilson et al.'s broad estimate. To date there have been few studies describing variation in the mutation rates of *Campylobacter* strains and the potential effects different environments have on these rates.

Genome degradation usually accompanise host-adaptation as genes that are not required to colonise the host are lost (Kingsley et al., 2013). The 16 ST45 isolates collected from the patient contained 1-4 more pseudogenes than other ST45 isolates analysed, suggesting adaption to the long-term human host via genome degradation. Klemm et al. (2016) performed a similar method on recurrent blood-borne *Salmonella enterica* serovar Enteriditis isolates collected from a single patient, but were able to identify a larger number of pseudogenes (approximately 70), as: 1. they had isolates from the start of the infection and did not have to rely on *S.* Enteriditis genomes collected from other sources as references to compare with; 2. they sequenced their isolates with larger reads, removing assembly gaps that decreased the core genome; and 3. *Salmonella* has a larger genome with more potential pseudogenes than *Campylobacter*. Studies with isolates collected earlier in the infection and/or sequenced with longer reads may have identified more pseudogenes shared by the ST45 isolates.

Flagella allow *Campylobacter* to move through the viscous mucus layer and colonise the intestinal tract (Riazi et al., 2013). Therefore, motile *Campylobacter* are often selected for on passage through human (Black et al., 1988) and poultry (Jones et al., 2004) hosts. However, these studies investigated the immediate selection pressures on *Campylobacter*. Studies on COPDs patients found that long-term *P. aeuriginosa* colonisation selected for non-motile isolates (Martínez-Solano et al., 2008). The loss of motility enabled *P.*

*aeruginosa* to evade phagocytes (Patankar et al., 2013). In this study there was a large amount of variation in motility between the first two *Campylobacter* isolates obtained from the patient, but less variation in the other 14 *Campylobacter* isolates. This may suggest that the *Campylobacter* isolates reached a trade-off between decreased motility to evade the immune response and the need for motility to colonise the gastrointestinal tract.

The patients medical records revealed that they had been prescribed multiple courses of antibiotics from 2006-2016. Two of these courses coincided with an increased resistance to ampicillin and ciprofloxacin. This suggests that the patient's antibiotic therapies acted as evolutionary bottlenecks, removing all susceptible ST45 variants within the host as a result of selective sweeps (Didelot et al., 2016). This hypothesis is supported by the relatively small amount of genetic diversity between the 10 isolates collected from the same faecal sample relative to the amount of time the patient has been colonised. Similar findings were found by Rodrigo-Troyano et al. (2016), with *P. aeruginosa* isolates that developed resistance to the antibiotics COPD patients were prescribed. However, they also observed the co-existence of isolates with different antimicrobial susceptibilities within the same host. Further work is required to investigate within-host diversity and how it is influenced by antibiotic therapy and other factors.

## 6.6 Conclusion

In this study, the evolution of *Campylobacter jejuni* ST45 within a persistently colonised human host was described. Genomic analyses, antimicrobial susceptibility and motility testing suggest that the patient was colonised with ST45 between 1998-2006 and that ST45 evolved within the patient's gastrointestinal tract, developing resistance to antibiotics the patient was prescribed and selected for changes in motility. The study also demonstrates how genomic and phylogenetic analyses can give insight into a patient's infection history.

# References


Algamoudi, B. A. and Ketley, J. M. (2015). Improved assays to determine the chemotactic behaviour of *Campylobacter jejuni*. In *Campylobacter, Helicobacter and Related Organisms*, pages 1–1, Rotorua, New Zealand.

Alikhan, N., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genomics*, 12(402):1–10.

Aronesty, E. (2013). Comparison of sequencing utility programs. *Open Bioinformatics Journal*, 7(1):1–8.

Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V.,

Wilke, A., and Zagnitko, O. (2008). The RAST server: Rapid annotations using subsystems technology. *BMC Genomics*, 9(75):1–15.

Black, R. E., Levine, M. M., Clements, M. L., Hughes, T. P., Blaser, M. J., and Black, R. E. (1988). Experimental *Campylobacter* infection in humans. *Journal of Infectious Diseases*, 157(3):472–479.

Conley, M. E., Notarangelo, L. D., and Etzioni, A. (1999). Diagnostic criteria for primary immunodeficiencies. *Clinical Immunology*, 93(3):190–197.

Cox, M. P., Peterson, D. A., and Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*, 11(485):1–6.

Davis, J. J., Gerdes, S., Olsen, G. J., Olson, R., Pusch, G. D., Shukla, M., Vonstein, V., Wattam, A. R., and Yoo, H. (2016). PATtyFams: Protein families for the microbial genomes in the PATRIC database. *Frontiers in Microbiology*, 7(FEB):1–12.

Didelot, X., Walker, A. S., Peto, T. E., Crook, D. W., and Wilson, D. J. (2016). Within-host evolution of bacterial pathogens. *Nature Reviews Microbiology*, 14(3):150–162.

Dionisi, A. M., Milito, C., Martini, H., Pesce, A. M., Mitrevski, M., Granata, G., Lucarelli, C., Parisi, A., Luzzi, I., and Quinti, I. (2011). High prevalence of intestinal carriage of *Campylobacter coli* in patients with primary antibody deficiencies. *Journal of Clinical Gastroenterology*, 45(5):474–475.

Drummond, A., Suchard, M., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8):1969–1973.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5):699–710.

EUCAST (2016). *Campylobacter jejuni* and *coli* EUCAST clinical breakpoint tables. Retrieved 2017-06-29, from: `http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_6.0_Breakpoint_table.pdf`, 80–80.

Fearnhead, P., Biggs, P. J., and French, N. (2014). Learning about recombination in *Campylobacter*. In Sheppard, S. K. and Meric, G., editors, *Campylobacter ecology and evolution*, pages 9–22. Caister Academic Press, Norfolk, England.

Friedrich, A. (2014). *Campylobacter jejuni microevolution and phenotype:genotype relationships*. PhD thesis, Massey University.

Gardner, S. N., Slezak, T., and Hall, B. G. (2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics*, 31:2877–2878.

Garrison, E. and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. Retrieved 2017-02-02, from: `https://arxiv.org/abs/1207.3907`, 1–1.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8):1072–1075.

Hänel, I., Borrmann, E., Müller, J., Müller, W., Pauly, B., Liebler-Tenorio, E. M., and Schulze, F. (2009). Genomic and phenotypic changes of *Campylobacter jejuni* strains after passage of the chicken gut. *Veterinary Microbiology*, 136(1-2):121–129.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2):160–174.

Hugdahl, M. B., Beery, J. T., and Doyle, M. P. (1988). Chemotactic behavior of *Campylobacter jejuni*. *Infection and Immunity*, 56(6):1560–1566.

Huson, D. and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2):254–267.

Jones, M. A., Marston, K. L., Woodall, C. A., Maskell, D. J., Linton, D., Karlyshev, A. V., Dorrell, N., Wren, B. W., and Barrow, P. A. (2004). Adaptation of *Campylobacter jejuni* NCTC 11168 to high-level colonization of the avian gastrointestinal tract. *Infection and Immunity*, 72(7):3769–3776.

Kingsley, R. A., Kay, S., Connor, T., Barquist, L., Sait, L., Holt, K. E., Sivaraman, K., Wileman, T., Goulding, D., Clare, S., Hale, C., Seshasayee, A., Harris, S., Thomson, N. R., Gardner, P., Rabsch, W., Wigley, P., Humphrey, T., Parkhill, J., and Dougan, G. (2013). Genome and transcriptome adaptation accompanying emergence of the definitive type 2 host-restricted *Salmonella enterica* serovar Typhimurium pathovar. *mBio*, 4(5):1–14.

Kirkpatrick, B. D., Lyon, C. E., Porter, C. K., Maue, A. C., Guerry, P., Pierce, K. K., Carmolli, M. P., Riddle, M. S., Larsson, C. J., Hawk, D., Dill, E. A., Fingar, A., Poly, F., Fimlaid, K. A., Hoq, F., and Tribble, D. R. (2013). Lack of homologous protection against *Campylobacter jejuni* CG8421 in a human challenge model. *Clinical Infectious Diseases*, 57(8):1106–1113.

Klemm, E. J., Gkrania-Klotsas, E., Hadfield, J., Forbester, J. L., Harris, S. R., Hale, C., Heath, J. N., Wileman, T., Clare, S., Kane, L., Goulding, D., Otto, T. D., Kay, S., Doffinger, R., Cooke, F. J., Carmichael, A., Lever, A. M. L., Parkhill, J., MacLennan, C. A., Kumararatne, D., Dougan, G., and Kingsley, R. A. (2016). Emergence of host-adapted *Salmonella* Enteritidis through rapid evolution in an immunocompromised host. *Nature Microbiology*, 1(3):1–6.

Kohanski, M. A., DePristo, M. A., and Collins, J. J. (2010). Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis. *Molecular Cell*, 37(3):311–320.

Lee, G., Pan, W., Peñataro Yori, P., Paredes Olortegui, M., Tilley, D., Gregory, M., Oberhelman, R., Burga, R., Chavez, C. B., and Kosek, M. (2013). Symptomatic and asymptomatic *Campylobacter* infections associated with reduced growth in Peruvian children. *PLoS Neglected Tropical Diseases*, 7(1):1–9.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21):2987–2993.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754–1760.

Liofilchem (2014). MIC test strip tehcnical sheet *Campylobacter* spp. Retrieved 2017-06-29, from: `http://www.liofilchem.net/login.area.mic/technical_sheets/MTS16.pdf`, 1–1.

Martínez-Solano, L., Macia, M. D., Fajardo, A., Oliver, A., and Martinez, J. L. (2008). Chronic *Pseudomonas aeruginosa* infection in chronic obstructive pulmonary disease. *Clinical Infectious Diseases*, 47(12):1526–1533.

Minin, V., Bloomquist, E., and Suchard, M. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, 25(7):1459–1471.

Molina, J.-M., Casin, I., Hausfater, P., Giretti, E., Welker, Y., Decazes, J.-M., Garrait, V., Lagrange, P., and Modai, J. (1995). *Campylobacter* infections in HIV-infected patients: Clinical and bacteriological features. *AIDS*, 9(8):881–885.

Mouneimné, H., Robert, J., Jarlier, V., and Cambau, E. (1999). Type II topoisomerase mutations in ciprofloxacin-resistant strains of *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, 43(1):62–66.

Muellner, P., Marshall, J. C., Spencer, S. E. F., Noble, A. D., Shadbolt, T., Collins-Emerson, J. M., Midwinter, A. C., Carter, P. E., Pirie, R., Wilson, D. J., Campbell, D. M., Stevenson, M. A., and French, N. P. (2011). Utilizing a combination of molecular and spatial tools to assess the effect of a public health intervention. *Preventive Veterinary Medicine*, 102(3):242–253.

Patankar, Y. R., Lovewell, R. R., Poynter, M. E., Jyot, J., Kazmierczak, B. I., and Berwin, B. (2013). Flagellar motility is a key determinant of the magnitude of the inflammasome response to *Pseudomonas aeruginosa*. *Infection and Immunity*, 81(6):2043–2052.

Pignata, C., Guandalini, S., Guarino, A., De Vizia, B., Capano, G., and De Ritis, G. (1984). Chronic diarrhea and failure to thrive in an infant with *Campylobacter jejuni*. *Journal of Pediatric Gastroenterology and Nutrition*, 3(5):812–814.

Porter, I. A. and Reid, T. M. S. (1980). A milk-borne outbreak of *Campylobacter* infection. *Journal of Hygiene*, 84(3):415–419.

Qiagen (2012). QIAamp DNA mini and blood mini handbook. Retrieved 2017-06-29, from: \nhttps://www.qiagen.com/us/resources/resourcedetail?id=62a200d6-faf4-469b-b50f-2b59cf738962&lang=en, 33-56.

Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014). Tracer 1.6. Retrieved 2016-09-27, from: http://beast.bio.ed.ac.uk/Tracer, 1–1.

Resnick, E. S., Moshier, E. L., Godbold, J. H., and Cunningham-Rundles, C. (2012). Morbidity and mortality in common variable immune deficiency over 4 decades. *Blood*, 119(7):1650–1657.

Riazi, A., Strong, P. C. R., Coleman, R., Chen, W., Hirama, T., Van Faassen, H., Henry, M., Logan, S. M., Szymanski, C. M., MacKenzie, R., and Ghahroudi, M. A. (2013). Pentavalent single-domain antibodies reduce *Campylobacter* motility and colonization in chickens. *PLoS ONE*, 8(12):1–12.

Rodrigo-Troyano, A., Suarez-Cuartin, G., Peiró, M., Barril, S., Castillo, D., Sanchez-Reus, F., Plaza, V., Restrepo, M. I., Chalmers, J. D., and Sibila, O. (2016). *Pseudomonas aeruginosa* resistance patterns and clinical outcomes in hospitalized exacerbations of COPD. *Respirology*, 21(7):1235–1242.

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28(1):33–36.

TREK (2012). Sensititre susceptibility plates for *Campylobacter*. Retrieved 2017-06-29, from: http://www.trekds.com/techDocs/techInsert/009_CAMPY_GB__V3.1.CID8522.pdf, 1–5.

Tribble, D. R., Baqar, S., Scott, D. A., Oplinger, M. L., Trespalacios, F., Rollins, D., Walker, R. I., Clements, J. D., Walz, S., Gibbs, P., Burg, E. F. I., Moran, A. P., Applebee, L., and Bourgeois, A. L. (2010). Assessment of the duration of protection in *Campylobacter jejuni* experimental infection in humans. *Infection and Immunity*, 78(4):1750–1759.

Valderrey, A. D., Pozuelo, M. J., Jiménez, P. A., Maciá, M. D., Oliver, A., and Rotger, R. (2010). Chronic colonization by *Pseudomonas aeruginosa* of patients with obstructive lung diseases: Cystic fibrosis, bronchiectasis, and chronic obstructive pulmonary disease. *Diagnostic Microbiology and Infectious Disease*, 68(1):20–27.

van der Beek, M. T., Claas, E. C. J., Mevius, D. J., van Pelt, W., Wagenaar, J. A., and Kuijper, E. J. (2010). Inaccuracy of routine susceptibility tests for detection of erythromycin resistance of *Campylobacter jejuni* and *Campylobacter coli*. *Clinical Microbiology and Infection*, 16(1):51–56.

Vandamme, P., Dewhirst, F. E., Paster, B. J., and On, S. L. W. (2005). Genus 1. *Campylobacter*. In Brenner, D. J., Kreig, N. R., Staley, J. T., and Garrity, G. M., editors, *Manual of Systematic Bacteriology*, pages 1147–1160. Springer, New York.

Wang, G., Clark, C. G., Taylor, T. M., Pucknell, C., Barton, C., Price, L., Woodward, D. L., and Rodgers, F. G. (2002). Colony multiplex PCR assay for identification and differentiation of *Campylobacter jejuni*, *C. coli*, *C. lari*, *C. upsaliensis*, and *C. fetus subspecies fetus*. *Journal of Clinical Microbiology*, 40(12):4744–4747.

Wang, Y., Huang, W. M., and Taylor, D. E. (1993). Cloning and nucleotide sequence of the *Campylobacter jejuni* gyrA gene and characterization of quinolone resistance mutations. *Antimicrobial Agents and Chemotherapy*, 37(3):457–463.

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., Gillespie, J. J., Gough, R., Hix, D., Kenyon, R., MacHi, D., Mao, C., Nordberg, E. K., Olson, R., Overbeek, R., Pusch, G. D., Shukla, M., Schulman, J., Stevens, R. L., Sullivan, D. E., Vonstein, V., Warren, A., Will, R., Wilson, M. J. C., Yoo, H. S., Zhang, C., Zhang, Y., and Sobral, B. W. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research*, 42(D1):D581–D591.

WHO (2011). *Campylobacter*. Retrieved 2016-12-07, from: `http://www.who.int/mediacentre/factsheets/fs255/en/`, 1–1.

Wilson, D. J., Gabriel, E., Leatherbarrow, A. J. H., Cheesbrough, J., Gee, S., Bolton, E., Fox, A., Fearnhead, P., Hart, C. A., and Diggle, P. J. (2008). Tracing the source of campylobacteriosis. *PLoS Genetics*, 4(9):1–9.

Zeng, X., Xu, F., and Lin, J. (2010). Development and evaluation of CmeC subunit vaccine against *Campylobacter jejuni*. *Journal of Vaccines and Vaccination*, 1(3):1–21.

Zerbino, D. R. and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18(5):821–829.

# Chapter 7

# General discussion

## 7.1 Introduction

Bacterial enteritis outbreaks are a worldwide problem (Okoro et al., 2012; Byrne et al., 2014). A large number of studies have investigated these outbreaks (Kingsley et al., 2009), how they behave (Stine et al., 2008), their origin (Eppinger et al., 2014), and how the causative bacterial agents are transmitted and evolve (Bingham et al., 2004; Mather et al., 2013), to identify ways to limit or prevent them (Mahoney et al., 1993). The studies in this thesis investigated the transmission and evolution of bacteria over the course of outbreaks. The aim of this chapter was to compare and contrast the studies performed in this thesis, how they fit within our current knowledge of bacterial transmission and evolution, and highlight areas that require further work.

## 7.2 Whole genome sequencing

Whole genome sequencing involves sequencing the entire genome of an organism. In this thesis I investigated multiple bacterial enteritis outbreaks using whole genome sequencing and demonstrated many of the benefits and limitations of this approach.

Identifying the date of the most recent common ancestor for isolates in an outbreak can help identify when the outbreak strain was introduced into an area. In turn, comparing global isolates collected at this date can help determine where the isolates originated from (Eppinger et al., 2014). In Chapter 3, I was able to use whole genome sequencing and phylogenetic analysis to demonstrate that *Salmonella enterica* serovar Typhimurum DT160 was introduced into New Zealand around the year 1997, before it propogated throughout the country, accumulating mutations and becoming more genetically diverse. However, in the absence of a large database of other DT160 isolates I was unable to determine where the strain originated from. Similarly, Dyson et al. (2017) sequenced the genomes of 44 historical *Salmonella enterica* serovar Typhi between 1973 and 1992 from Thailand. They were able to predict the movement *S.* Typhi between Thailand and other countries after 1977, but a lack of published *S.* Typhi prior to 1977 prevented *S.* Typhi geographical movement estimates. Therefore, whole genome sequencing can distinguish between isolates collected from

a country, and may assist in determining the movement of bacteria globally given appropriate additional isolates.

The phenotype of an organism consists of all its physical, biochemical and behavioural traits, and is influenced by an organism's genome. In chapter 5, I was able to identify genotypic differences between the DT160 and *S.* Typhimurium DT56 variant strains that may explain why one strain emerged in New Zealand as the other declined. However, in the absence of phenotypic results I was unable to validate any of these hypotheses. Therefore, whole genome sequencing can be used to identify potential ways in which bacterial isolates may differ phenotypically, but phenotypic test are required to validate these hypotheses.

Campylobacteriosis patients usually begin excreting *Campylobacter* at the onset of symptoms and continue to excrete for 2-3 weeks after symptoms subside (Porter and Reid, 1980). Longer *Campylobacter* excretion periods have been described in association with chronic campylobacteriosis (Molina et al., 1995; Pignata et al., 1984), reinfections and asymptomatic campylobacteriosis cases (Tribble et al., 2010). However, in these studies *Campylobacter* excretion did not last longer than 2 years and insufficient typing was performed to demonstrate that the *Campylobacter* strains obtained from the patients were the same. In chapter 6, I was able to use whole genome sequencing and antimicrobial susceptibility testing to demonstrate that 16 *Campylobacter* isolates collected from a patient over 10 years were closely related and that the patient was continuously colonised with *Campylobacter* over this time. However, unlike the long-term *Campylobacter* excretion studies previously mentioned I had insufficient medical information to conclude why the patient had been excreting *Campylobacter* for this length of time. Therefore, whole genome sequencing can be used to estimate how long a patient had been colonised with a bacteria, but cannot determine why the patient was colonised with the bacteria without additional medical information.

In this thesis, whole genome sequencing was able to provide a large amount of information on groups of bacterial isolates and how they related to each other. However, a lack of phenotypic data, medical information and global isolates, limited the conclusions that could be drawn from the genomic data.

## 7.3 Transmission

Bacteria are continuously transmitted between individuals within the same species, between different species, and between species and their environment. Studies on bacterial transmission usually involve investigating bacteria from different sources, determining how the bacterial isolates are related and deducing transmission likelihood from relatedness (Mughini-Gras et al., 2014). Previously, the biggest road block to predicting bacterial transmission was the inability to distinguish between bacterial isolates collected from different sources (Omar, 2010). However, with the advent of highly discriminatory genomic sequencing, isolates from different sources can be distinguished, but this has led to other problems when predicting bacterial transmission. In this thesis, I describe several of these problems.

In Chapter 3, 109 DT160 isolates were collected from human, bovine, poultry and wild bird sources over 14 years. Whole genome sequencing was able to distinguish the 109 DT160 isolates from each other.

However, there was no clear distinction between isolates collected from each source. Therefore, it was deduced that there were frequent transmissions between these sources. Mather et al. (2013) had previously used the discrete trait analysis model to investigate *S.* Typhimurium DT104 transmission over the course of a 20 year Scottish outbreak. Chapter 4 investigated the applicability of the discrete trait analysis model and the structured coalescent model for investigating salmonellosis outbreaks and found that both inaccurately predicted transmission parameters. Chapters 3 and 4 highlighted the need for better quality control when bacterial transmission is inferred from models not intended for this purpose, and, more specifically, the need for outbreak-specific phylodynamic models that can infer transmission parameters from highly discriminatory genomic data.

In Chapter 6, as part of the Manawatu Sentinel Surveillance Site (MSSS) study, 16 *Campylobacter jejuni* ST45 isolates from the same patient were investigated using whole genome sequencing, antimicrobial susceptibility testing and motility assays. These tests showed that the patient had been continuously colonised with *Campylobacter* for at least 10 years. Potential sources in the patient's domestic environment were cultured for *Campylobacter*, but all tested negative. Workplace sources were unobtainable as the patient had multiple jobs over this period of time. Assuming the source of *Campylobacter* was domestic or work-related, it was unlikely it would still be present after 10 years. Potential sources were also investigated by comparing the isolates collected from the patient with other ST45 isolates collected as part of the MSSS. However, only four other ST45 isolates had been whole genome sequenced and these isolates were genetically distinct from those collected from the patient. Multiple studies have attributed uncertainty in transmission estimates to a lack of isolates from potential sources. For example, Eyre et al. (2013) sequenced the genomes of 1,223 *Clostridium difficile* genomes from 957 hospital- and community-acquired cases in Oxfordshire, United Kingdom over three years. This enabled them to determine if repeated infections were due to repeated exposure or recrudescence of previous infections. However, they had insufficient *C. difficile* genomes derived from non-human sources to identify likely sources for 45% of cases. Therefore, whole genome sequencing can provide a large amount of information on the nature of infections, but is unable to determine the likely source of infections without sufficient source data for comparison.

In this thesis, genomics was able to provide a large amount of information on clinically significant *Campylobacter* and *Salmonella* isolates. However, without applicable outbreak transmission models and source data, limited bacterial transmission inferences could be made.

## 7.4 Evolution

Bacteria are constantly evolving to adapt to new and changing environments. In this thesis I outlined some of the ways in which bacteria evolve in clinical and ecological contexts.

Evolution is influenced by changes in the environment (Baronchelli et al., 2013). In Chapter 3, I described DT160 evolution in New Zealand, how it was introduced between 1996 and 1998, how it quickly spread

throughout the country accumulating mutations, and how the predicted proteins in each functional group contained similar proportions of protein differences. DT160's introduction into New Zealand was an example of a 'founder effect' (Mayr, 1942), as a small number of DT160 isolates established the population in New Zealand, and the bacteria evolved via neutral evolution (Kimura, 1968), as there was no evidence of selective pressures imposed on DT160. On the other hand, in Chapter 6 I described the evolution of ST45 within a human host over 10 years, how the patient became infected with the bacteria between 1998 and 2006, how the diversity within the host was small, and how less motile, antibiotic-resistant variants were selected for. ST45's initial colonisation of the patient was another example of a 'founder effect', as a small number of ST45 isolates formed the population within the host, but the bacteria evolved by natural selection (Darwin, 1860), as the human host selected for specific *Campylobacter* phenotypes. The difference in evolution between these isolates is most likely the result of their different environments. DT160 was introduced into an environment with multiple host species and little competition as most other *S.* Typhimurium species circulating in New Zealand at the time were sensitive to the ST160 phage contained within the DT160 genome (Price-Carter et al., 2011). This likely allowed them to rapidly spread between species and accumulate mutations via genetic drift. On the other hand, the ST45 isolates were introduced to an environment that could host a much smaller number of bacteria and were imposed on by multiple selective pressures from the human host's immune system and antibiotic treatments. Therefore, ST45 isolates that could efficiently colonise the host under these conditions were continually being selected. These examples demonstrate how different environments can influence how bacteria evolve over the course of outbreaks.

Genetics influences bacterial evolution (Nobusawa and Sato, 2006). In this thesis I described the evolution of two clinically significant bacterial strains, *S.* Typhimurium DT160 (Chapter 3) and *C. jejuni* ST45 (Chapter 6). The ST45 isolates analysed were found to mutate at a much faster rate than the DT160 isolate analysed (almost 10 times faster) and contained a slightly larger indel-to-SNP ratio. As the *Campylobacter* isolates mutated at a much faster rate, based on the generated SNP data, I can assume that their slightly higher indel-to-SNP ratio correlated to a much higher indel rate. It should be pointed out that the method used to identify indels only investigated those found within genes, whilst the method used to identify SNPs identified them throughout the genome. Nevertheless, this bias should be constant between the strains. The difference in these mutation rates could be the result of the different environments the strains inhabited. This is hard to prove, as to date all the papers that have described genome-wide *S.* Typhimurium mutation rates investigated isolates from similar outbreaks to the DT160 outbreak (Mather et al., 2013; Okoro et al., 2012), and no genome-wide mutation rates have been published for *Campylobacter*. It is likely that the differences are due to the different genetics of the *Campylobacter* and *Salmonella* genera. Unlike *Salmonella*, *Campylobacter* does not contain the *mutH*, *mutL* or *mutS* genes, which are used for nucleotide mismatch repair and whose absence is associated with a higher mutation rate (Ambur et al., 2009; Pang et al., 1985). In addition, *Campylobacter* contain a larger density of Simple Sequence Repeats (SSRs) within their genome compared to *Salmonella*, which are associated with higher mutation rates (Lin and Kussell, 2012). Combined, these

genetic differences account for many of the differences in the DT160 and ST45 mutation rates. They also highlight how genetic differences can affect bacterial evolution.

In this thesis, genomics was used to study the evolution of multiple bacterial strains over the course of outbreaks. This research demonstrated how bacterial evolution is influenced by both the environments that the bacteria inhabit and the underlying genetics of the bacterial strains.

## 7.5 Future work

The findings of this thesis add to our current knowledge of microbiology, phylogenetics, epidemiology and public health. They also highlight areas within these fields that require further work.

In Chapter 3, an outbreak of *S.* Typhimurium DT160 in New Zealand was investigated using whole genome sequencing. One of the limitations of this study was the lack of comparative studies with DT160 genomes outside of New Zealand. At the time of writing, only two DT160 isolates from the United Kingdom had been published (Petrovska et al., 2016). Therefore, comparative studies with DT160 isolates from worldwide sources are required to track the migration of this pathogen and identify where the DT160 strain in New Zealand originated from. In 2009, an outbreak of DT160 began in Tasmania, Australia (Grillo and Post, 2010). $^{m}$Epilab is currently collaborating with Dr Deborah Williamson, who is investigating this outbreak using genomics at the Doherty Institute, Melbourne, Australia. It is hoped that this collaboration will help determine how these outbreaks relate and further our understanding of the DT160 strain and how it behaves.

In Chapter 5, the genomes of two *Salmonella* strains that were subsequently the predominant causes of human salmonellosis in New Zealand were compared using whole genome sequencing. This chapter found that the isolates were distantly related and described possible reasons why one strain may have declined whilst the other emerged, e.g. phage-resistance, quicker doubling time and different serologies. Phenotypic tests could be used to investigate these hypotheses:

- Competitive growth assays may help determine if one strain can out-compete the other. However, these assays require a test to easily distinguish between the strains and a suitable medium to imitate their sources (e.g. an animal's gastrointestinal tract) (Portal-Celhay and Blaser, 2012). Competitive growth assays also rely on further testing to determine the reason for any competitive advantage.

- Cell growth assays may help determine if there are any differences in the growth rates of either strain. If one strain can out-grow another strain than it may be able to outcompete it when colonising the same source. However, as with competitive growth assays, cell growth assays rely on a suitable medium to imitate an animal's gastrointestinal tract (Sant'Ana et al., 2012).

- Phage assays may help determine if either strain is susceptible to the phage associated with the other (Price-Carter et al., 2011).

- Immunological assays could be used to determine if antibodies raised against defined antigens of one strain cross-react with those of the other strain (Biswas et al., 2010). If antibodies do not cross-react then an immune response raised against one strain may not protect against the other.

In Chapter 6, an individual was investigated who had been excreting the same strain of *Campylobacter* for over 10 years. The patient had been suffering from daily episodes of diarrhoea that varied in severity for approximately 16 years. Genomic and phenotypic analyses of the isolates demonstrated that the patient had most likely been continually colonised with *Campylobacter* over this period and that the *Campylobacter* had adapted to selective pressures imposed by the human host. However, it could not be determined if the *Campylobacter* were contributing to the patient's gastrointestinal ailment. Monthly faecal and blood samples from the patient during times of severe and mild diarrhoea could help investigate *Campylobacter*'s effect:

- The concentration of serum and faecal inflammatory markers (e.g., C-reactive protein and calprotectin) could be measured. If there was an association between inflammatory markers and diarrhoea severity, then this may indicate that diarrhoea is the result of intestinal inflammation. If there was no association then this may indicate that the diarrhoea is from inadequate electrolyte and water absorption (i.e. secretory diarrhoea).

- The concentration of anti-*Campylobacter* antibodies in the patient's serum could be measured. If there was an association between anti-*Campylobacter* antibodies, and the inflammatory markers and/or diarrhoea severity, then this may indicate that *Campylobacter* contributes to enteritis and/or diarrhoea severity. If there was no association, then this may indicate that the patient is an asymptomatic carrier of *Campylobacter*.

- *Campylobacter* isolates from each faecal sample could have their genomes sequenced. This would help determine how diarrhoea severity, enteritis or any immune responses influence *Campylobacter* evolution.

- 16S rRNA metabarcoding approaches could be used to determine what microorganisms make up the patient's microbiome. This could help determine if diarrhoea severity was associated with any changes in the microbiota.

In Chapter 6, the *C. jejuni* ST45 isolates collected from the patient via the MSSS were compared to all other ST45 isolates that had been collected as part of this programme. Only four other ST45 isolates had been sequenced and these isolates were distinct from those collected from the patient, so no potential sources were found. ST45 is a generalist sequence type associated with a large number of sources and responsible for 8% of human campylobacteriosis cases in New Zealand (Müllner et al., 2010). Despite this, very little is known about the strain. Dearlove et al. (2016) found that ST45 displayed frequent host switching between different species. However, this finding was based on the output of the discrete trait analysis model without taking into consideration the effect of different sample proportions, so it may not be correct. Whole genome sequencing 100-200 of the approximately 700 ST45 isolates that have been collected as part of the MSSS could help

determine if the entire ST45 sequence type is generalist or if the sequence type is made of host-specific and generalist strains, how the sequence type is structured, and may help identify the most likely source(s) of ST45 for the continually-excreting patient.

## 7.6 Conclusion

In this thesis I investigated the transmission and evolution of bacteria over several outbreaks. The results demonstrate many of the benefits and limitations of whole genome sequencing, indicate that current ancestral state reconstruction models are not applicable for predicting bacterial transmission over the course of outbreaks, highlight how genetics and the environment influence the evolution of bacteria, and identify areas within the scope of microbiology, phylogenetic, epidemiology and public health that require further research.

# References


Ambur, O. H., Davidsen, T., Frye, S. A., Balasingham, S. V., Lagesen, K., Rognes, T., and Tønjum, T. (2009). Genome dynamics in major bacterial pathogens. *FEMS Microbiology Reviews*, 33(3):453–470.

Baronchelli, A., Chater, N., Christiansen, M. H., and Pastor-Satorras, R. (2013). Evolution in a Changing Environment. *PLoS ONE*, 8(1):1–8.

Bingham, P., Verlander, N. Q., and Cheal, M. J. (2004). John Snow, William Farr and the 1849 outbreak of cholera that affected London: A reworking of the data highlights the importance of the water supply. *Public Health*, 118(6):387–394.

Biswas, D., Herrera, P., Fang, L., Marquardt, R. R., and Ricke, S. C. (2010). Cross-reactivity of anti-*Salmonella* egg-yolk antibodies to *Salmonella* serovars. *Journal of Environmental Science and Health - Part B Pesticides, Food Contaminants, and Agricultural Wastes*, 45(8):824–829.

Byrne, L., Fisher, I., Peters, T., Mather, A., Thomson, N., Rosner, B., Bernard, H., McKeown, P., Cormican, M., Cowden, J., Aiyedun, V., Lane, C., and Team, I. O. C. (2014). A multi-country outbreak of *Salmonella* Newport gastroenteritis in Europe associated with watermelon from Brazil, confirmed by whole genome sequencing: October 2011 to January 2012. *Euro surveillance : bulletin Européen sur les maladies transmissibles = European communicable disease bulletin*, 19(31):6–13.

Darwin, C. (1860). *On the origin of species by natural selection: Or the preservation of the favoured races in the struggle for life, pages 1–490.*

Dearlove, B. L., Cody, A. J., Pascoe, B., Méric, G., Wilson, D. J., and Sheppard, S. K. (2016). Rapid host switching in generalist *Campylobacter* strains erodes the signal for tracing human infections. *ISME Journal*, 10(3):721–729.

Dyson, Z., Thanh, D., Bodhidatta, L., Mason, C., Srijan, A., Rabaa, M., Vinh, P., Thanh, T., Thwaites, G., Baker, S., and Holt, K. (2017). Whole qenome sequence analysis of *Salmonella* typhi isolated in thailand before and after the introduction of a national immunization program. *PLoS Neglected Tropical Diseases*, 11(1). cited By 1.

Eppinger, M., Pearson, T., Koenig, S. S. K., Pearson, O., Hicks, N., Agrawal, S., Sanjar, F., Galens, K., Daugherty, S., Crabtree, J., Hendriksen, R. S., Price, L. B., Upadhyay, B. P., Shakya, G., Fraser, C. M., Ravel, J., and Keim, P. S. (2014). Genomic epidemiology of the Haitian cholera outbreak: A single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. *mBio*, 5(6):1–8.

Eyre, D., Cule, M., Wilson, D., Griffiths, D., Vaughan, A., O'Connor, L., Ip, C., Golubchik, T., Batty, E., Finney, J., Wyllie, D., Didelot, X., Piazza, P., Bowden, R., Dingle, K., Harding, R., Crook, D., Wilcox, M., Peto, T., and Walker, A. (2013). Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *New England Journal of Medicine*, 369(13):1195–1205. cited By 226.

Grillo, T. and Post, L. (2010). *Salmonella* Typhimurium DT160 outbreak in Tasmania. *Animal Health Surveillance Quarterly Reports*, 14(4):8–8.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626.

Kingsley, R. A., Msefula, C. L., Thomson, N. R., Kariuki, S., Holt, K. E., Gordon, M. A., Harris, D., Clarke, L., Whitehead, S., Sangal, V., Marsh, K., Achtman, M., Molyneux, M. E., Cormican, M., Parkhill, J., MacLennan, C. A., Heyderman, R. S., and Dougan, G. (2009). Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Research*, 19(12):2279–2287.

Lin, W. H. and Kussell, E. (2012). Evolutionary pressures on simple sequence repeats in prokaryotic coding regions. *Nucleic Acids Research*, 40(6):2399–2413.

Mahoney, F. J., Farley, T. A., Burbank, D. F., Leslie, N. H., and McFarland, L. M. (1993). Evaluation of an intervention program for the control of an outbreak of shigellosis among institutionalized persons. *Journal of Infectious Diseases*, 168(5):1177–1180.

Mather, A., Reid, S., Maskell, D., Parkhill, J., Fookes, M., Harris, S., Brown, D., Coia, J., Mulvey, M., Gilmour, M. o., Petrovska, L., De Pinna, E., Kuroda, M., Akiba, M., Izumiya, H., Connor, T., Suchard, M. l., Lemey, P., Mellor, D., Haydon, D., and Thomson, N. (2013). Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science*, 341(6153):1514–1517.

Mayr, E. (1942). *Systematics and the origin of species, pages 1-237*. Columbia University Press, New York.

Molina, J.-M., Casin, I., Hausfater, P., Giretti, E., Welker, Y., Decazes, J.-M., Garrait, V., Lagrange, P., and Modai, J. (1995). *Campylobacter* infections in HIV-infected patients: Clinical and bacteriological features. *AIDS*, 9(8):881–885.

Mughini-Gras, L., Enserink, R., Friesema, I., Heck, M., Van Duynhoven, Y., and Van Pelt, W. (2014). Risk factors for human salmonellosis originating from pigs, cattle, broiler chickens and egg laying hens: A combined case-control and source attribution analysis. *PLoS ONE*, 9(2):1–9.

Müllner, P., Collins-Emerson, J. M., Midwinter, A. C., Carter, P., Spencer, S. E. F., Van Der Logt, P., Hathaway, S., and French, N. P. (2010). Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry. *Applied and Environmental Microbiology*, 76(7):2145–2154.

Nobusawa, E. and Sato, K. (2006). Comparison of the mutation rates of human influenza A and B viruses. *Journal of Virology*, 80(7):3675–3678.

Okoro, C. K., Kingsley, R. A., Connor, T. R., Harris, S. R., Parry, C. M., Al-Mashhadani, M. N., Kariuki, S., Msefula, C. L., Gordon, M. A., De Pinna, E., Wain, J., Heyderman, R. S., Obaro, S., Alonso, P. L., Mandomando, I., MacLennan, C. A., Tapia, M. D., Levine, M. M., Tennant, S. M., Parkhill, J., and Dougan, G. (2012). Intracontinental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. *Nature Genetics*, 44(11):1215–1221.

Omar, S. (2010). *Molecular epidemiology of Salmonella Typhimurium DT160 in New Zealand*. PhD thesis, Massey University.

Pang, P. P., Lundberg, A. S., and Walker, G. C. (1985). Identification and characterization of the *mutL* and *mutS* gene products of *Salmonella typhimurium* LT2. *Journal of Bacteriology*, 163(3):1007–1015.

Petrovska, L., Mather, A. E., Abuoun, M., Branchu, P., Harris, S. R., Connor, T., Hopkins, K. L., Underwood, A., Lettini, A. A., Page, A., Bagnall, M., Wain, J., Parkhill, J., Dougan, G., Davies, R., and Kingsley, R. A. (2016). Microevolution of monophasic *Salmonella* Typhimurium during epidemic, United Kingdom, 20052010. *Emerging Infectious Diseases*, 22(4):617–624.

Pignata, C., Guandalini, S., Guarino, A., De Vizia, B., Capano, G., and De Ritis, G. (1984). Chronic diarrhea and failure to thrive in an infant with *Campylobacter jejuni*. *Journal of Pediatric Gastroenterology and Nutrition*, 3(5):812–814.

Portal-Celhay, C. and Blaser, M. J. (2012). Competition and resilience between founder and introduced bacteria in the *Caenorhabditis elegans* gut. *Infection and Immunity*, 80(3):1288–1299.

Porter, I. A. and Reid, T. M. S. (1980). A milk-borne outbreak of *Campylobacter* infection. *Journal of Hygiene*, 84(3):415–419.

Price-Carter, M., Roy-Chowdhury, P., Pope, C. E., Paine, S., De Lisle, G. W., Collins, D. M., Nicol, C., and Carter, P. E. (2011). The evolution and distribution of phage ST160 within *Salmonella enterica* serotype Typhimurium. *Epidemiology and Infection*, 139(8):1262–1271.

Sant'Ana, A. S., Franco, B., and Schaffner, D. W. (2012). Modeling the growth rate and lag time of different strains of *Salmonella enterica* and *Listeria monocytogenes* in ready-to-eat lettuce. *Food Microbiology*, 30(1):267–273.

Stine, O. C., Alam, M., Tang, L., Nair, G. B., Siddique, A. K., Faruque, S. M., Huq, A., Colwell, R., Sack, R. B., and J Glenn Jr., M. (2008). Seasonal cholera from multiple small outbreaks, rural Bangladesh. *Emerging Infectious Diseases*, 14(5):831–833.

Tribble, D. R., Baqar, S., Scott, D. A., Oplinger, M. L., Trespalacios, F., Rollins, D., Walker, R. I., Clements, J. D., Walz, S., Gibbs, P., Burg, E. F. I., Moran, A. P., Applebee, L., and Bourgeois, A. L. (2010). Assessment of the duration of protection in *Campylobacter jejuni* experimental infection in humans. *Infection and Immunity*, 78(4):1750–1759.

# Appendix A

## Supplementary material to Chapter 3

### A.1 SNP comparison

SNPs (single nucleotide polymorphisms) are single base pairs that differ between isolates. Two software programs were used to identify SNPs shared by the 109 New Zealand DT160 isolates: kSNP3 (Gardner et al., 2015) and Snippy (https://github.com/tseeman/snippy). Snippy was used to align reads from each isolate to a reference genome, in this case *S.* Typhimurium strain 14028s (NC_016856), and then to compare the alignment results and identify single base pairs that were found in all isolates but differed in sequence (core SNPs). kSNP3 was used to identify kmers of a fixed length that differed in one nucleotide between *de novo*-assembled genomes and NC_016856. kSNP3 identified 731 SNPs shared by the 109 DT160 isolates, while Snippy identified 771 SNPs (Figure A.1). 709 SNPs were identified by both methods, leaving 22 kSNP-unique and 62 Snippy-unique SNPs. The kSNP-unique SNPs mostly consisted of SNPs found on reads that did not align to the reference genome, while the Snippy-unique SNPs mostly consisted of SNPs that were in close vicinity, unable to be picked up by kSNP as kmers of a fixed length would differ in more than one nucleotide. By using both methods a larger number of SNPs were identified than if a single method alone was used.

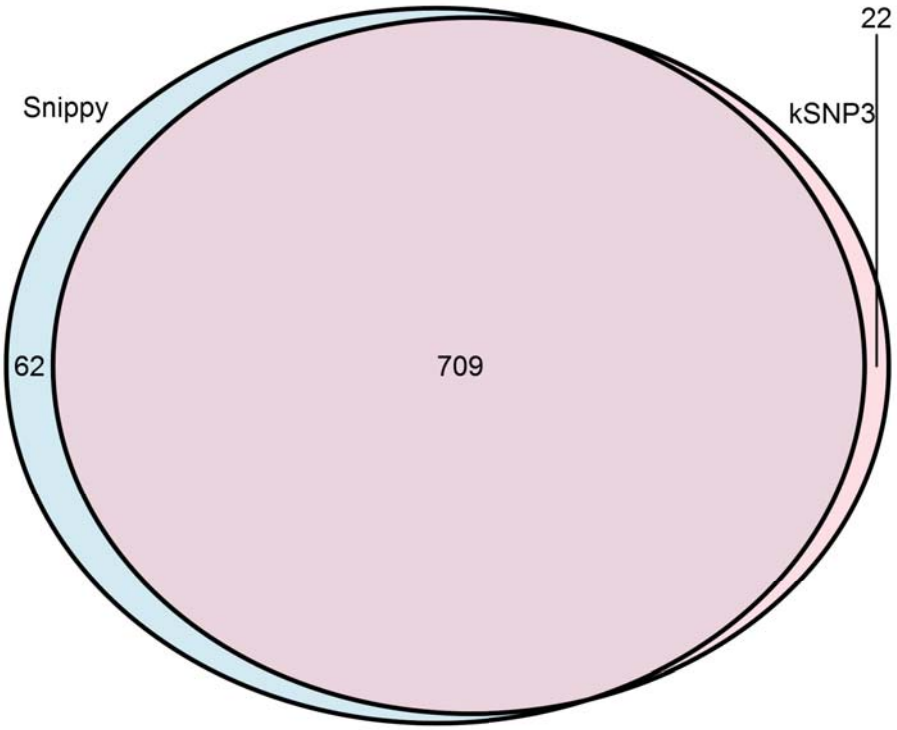


**Figure A.1** Venn diagram of the number of unique and shared SNPs identified by Snippy and kSNP3 for the 109 New Zealand DT160 isolates.

773 out of the 793 core SNPs shared by the 109 DT160 isolates were located on the reference genome, NC_016856. The order of these SNPs on the reference genome identified several small clades associated with close clusters of SNPs (Figure A.2). However, most of the SNPs in these clusters were synonymous and unlikely to result from selection pressures. The order of these SNPs also identified the non-synonymous SNPs responsible for the formation of two distinct DT160 clades and the proteins they were located within: glycogen debranching enzyme (A), 2-dehydro-3-deoxyphosphooctonate aldolase (B), a YggT family protein (C), galactose-1-epimerase (D), uvrABC system protein B (E) and acrylyl-coA reductase (F). Many of these proteins are involved in carbohydrate metabolism, suggesting that the two DT160 clades may have distinct carbohydrate metabolism phenotypes.

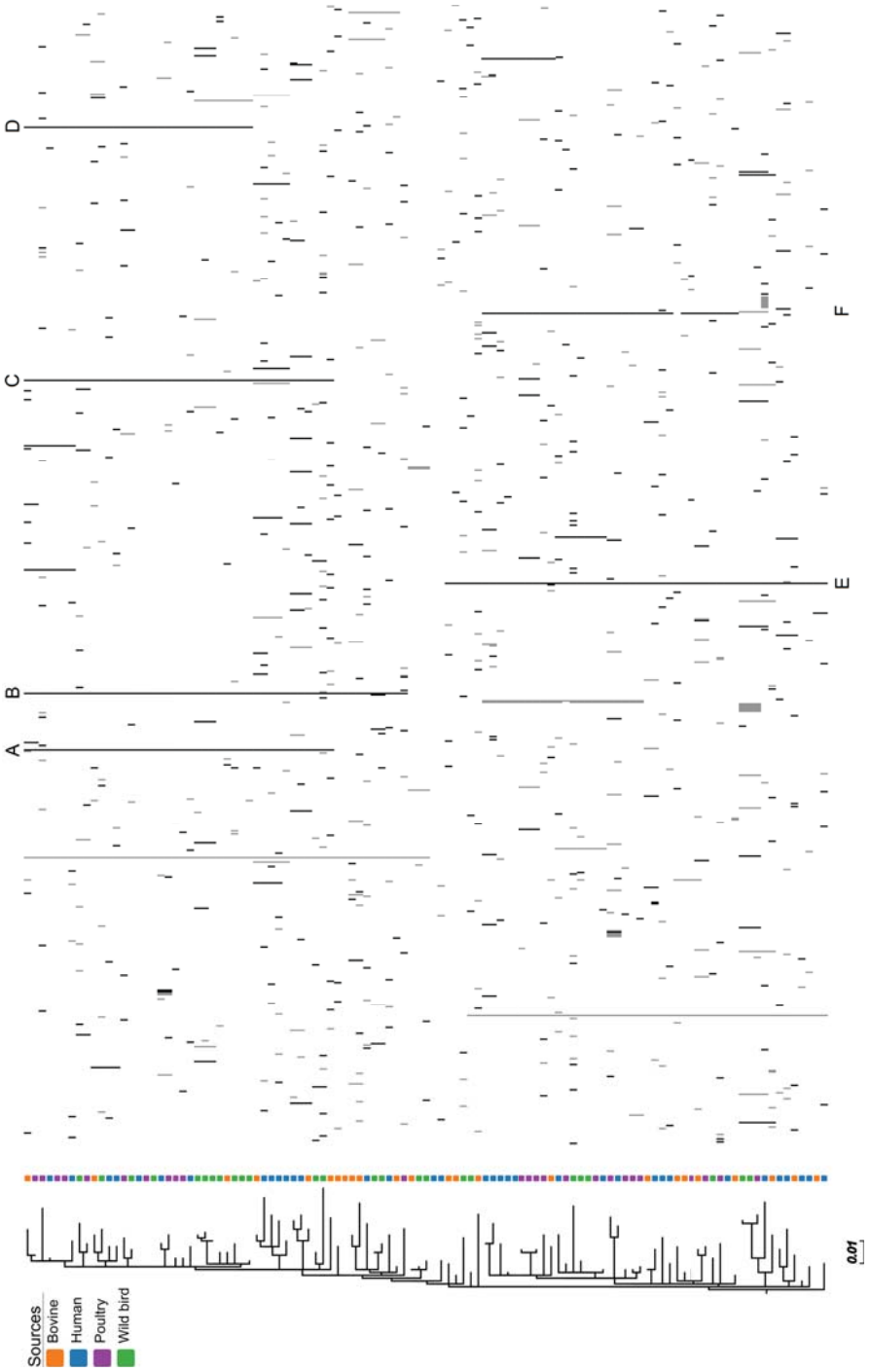


**Figure A.2** Maximum likelihood tree of 109 DT160 isolates (based on 793 core SNPs). The scale bar represents the number of nucleotide substitutions per site. The coloured squares represent the sources of the isolates. The presence-absence matrix represents the presence of the 773 core SNPs located on the reference genome, NC_016856. The SNPs were arranged in the order they appear on the reference genome. Black bars represent non-synonymous SNPs and grey bars represent synonymous SNPs. The non-synonymous SNPs responsible for the formation of the major DT160 clades were found in: a glycogen debranching enzyme (A), 2-dehydro-3-deoxyphosphooctonate aldolase (B), a YggT family protein (C), Galactose-1-epimerase (D), uvrABC system protein B (E) and Acrylyl-coA reductase (F).

A.2 Protein-coding gene analysis

The 109 DT160 isolates shared 684 protein differences. Primer-E v6 (Clarke and Gorley, 2006) was used to predict the Euclidian distance matrix based on the presence of these protein differences.

Of the 684 proteins that differed in sequence, 546 (93%) contained a single protein difference (SNP, indel or presence), 53 (7%) contained two protein differences, and 5 (<1%) contained more than two (Figure A.3).

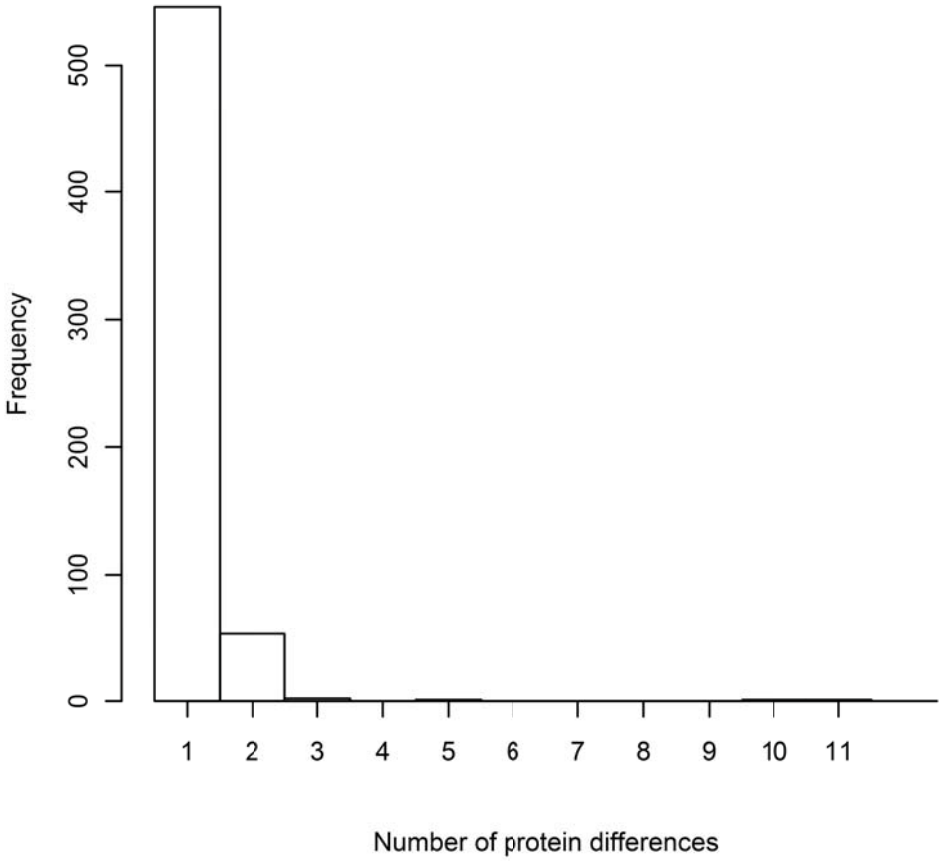


**Figure A.3** Histogram of the number of protein differences found within the same protein sequence for 109 DT160 isolates.

Two isolates were excluded from protein analyses as they lacked a large number of genes and were skewing the multi-dimensional scaling, functional plots and PermDisp calculations (Figure A.4). These outliers shared similar epidemiologic information: collected from human sources from 2004-2006. However, they were missing different sets of genes.

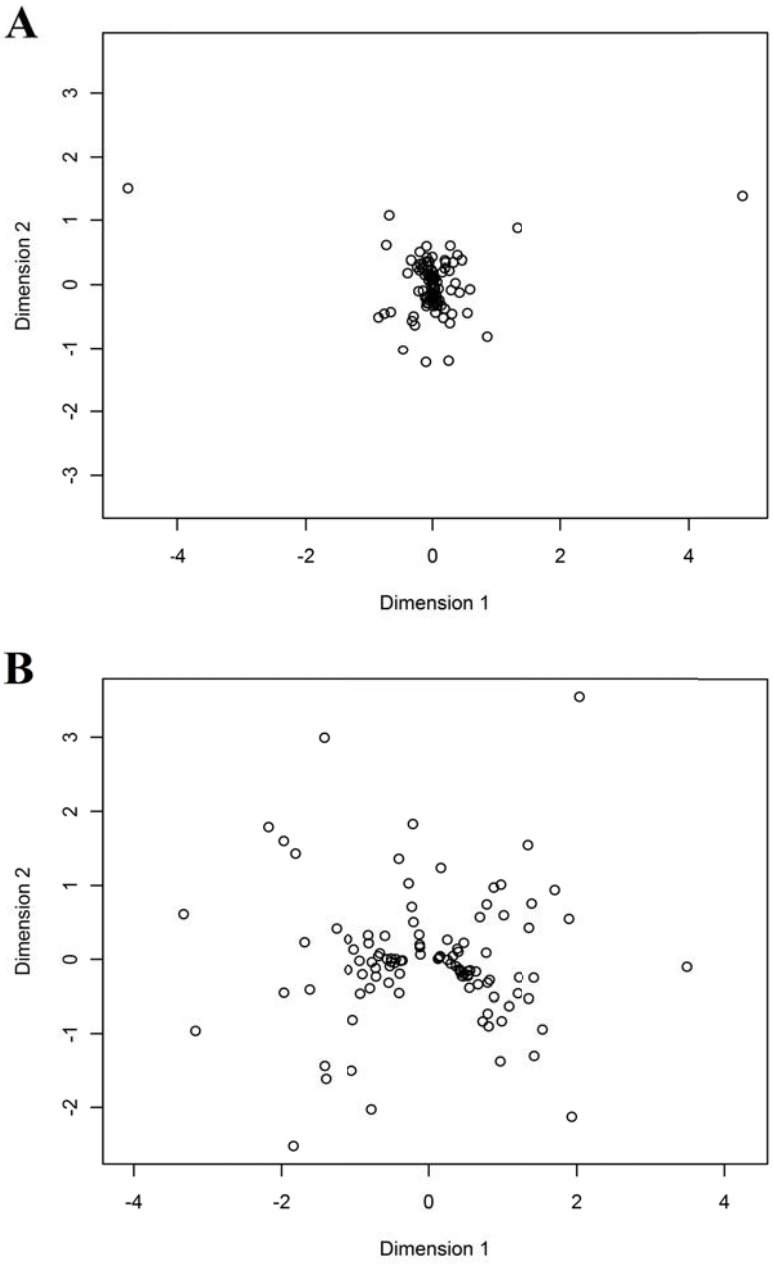


**Figure A.4** Multi-dimensional scaling of 109 (A) and 107 (minus two outliers) (B) DT160 isolates based on the presence of 684 protein differences.

Multidimensional scaling helps visualise the amount of similarity or dissimilarity between data points. In multi-dimensional scaling, the centroid is the central point for a group of data points. PERMANOVA found that the centroids were indistinguishable between isolates collected from different sources or time periods (Table A.1), as these isolates appeared to radiate out from a point source.

**Table A.1** PERMANOVA (http://www.primer-e.com/permanova.htm) output for 107 DT160 isolates, based on the presence of 684 protein differences and grouped by year of collection and source.

| Coefficient | Df | SS | MSS | Psuedo-F | P(perm) | Unique perms |
|---|---|---|---|---|---|---|
| Year | 4 | 42.26 | 10.57 | 1.143 | 0.121 | 998 |
| Source | 3 | 26.9 | 8.968 | 0.97 | 0.515 | 997 |
| YearxSource[a] | 10 | 99.9 | 9.99 | 1.081 | 0.187 | 996 |
| Residuals | 89 | 822.8 | 9.245 | | | |
| Total | 106 | 1,002 | | | | |

Df, degrees of freedom; SS, sum of squares; MSS, mean sum of squares; Pseudo-F, F-value from the data; P(perm), proportion of permuted datasets whose F-value exceeds Pseudo-F; Unique perms, number of unique permutations. [a]Coefficient interaction.

The distance from the centroid to each isolate (z-value) is a measure of dispersion and equivalent to the accumulation of protein differences. The z-values were calculated using PermDisp (Anderson, 2006) and were modeled using a regression model. The residuals for this model lacked normality (Figure A.5). To normalise the residuals, the z-values could have been transformed. However, with such a low p-value for the date of collection, this would not have changed the conclusions and would have made interpretation more difficult.

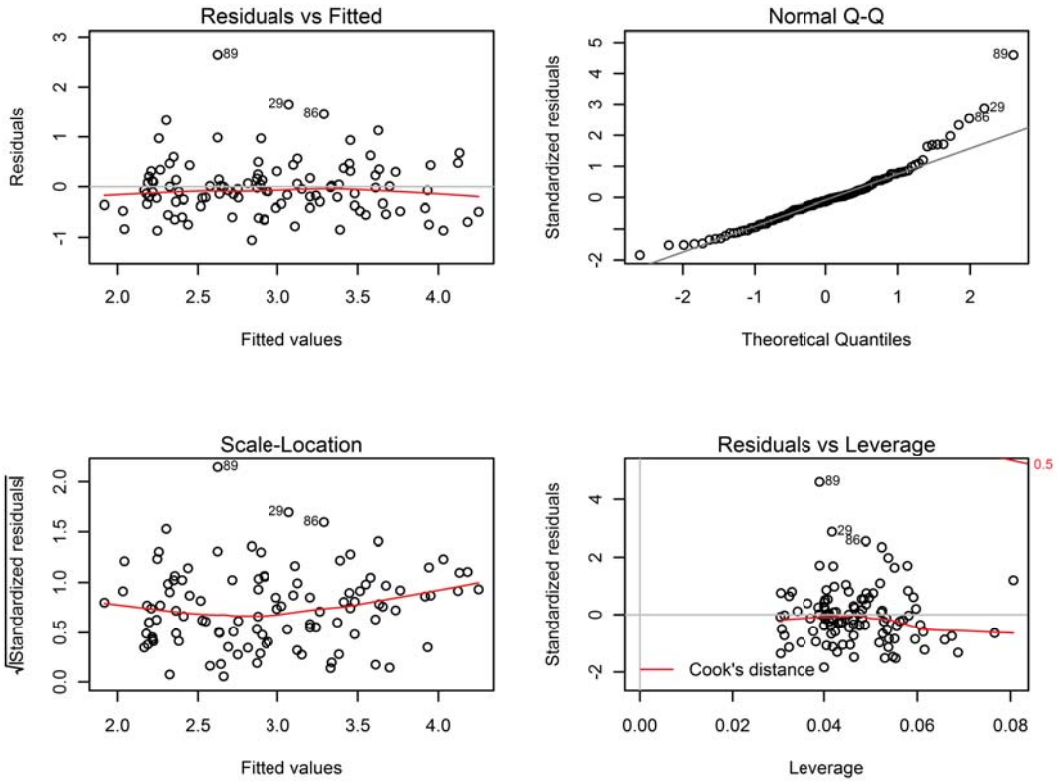


**Figure A.5** Diagnostic plots of the regression model fitted to the z-values for 107 DT160 isolates.

The mean proportion of proteins that differ in sequence for each functional group within each time period and source was calculated by dividing the proportion of proteins that differed in sequence among each source and time period in each functional group by the number of samples in each group (Figure A.6). Year of collection and source seemed to have a significant effect on the mean proportion of proteins that differ in sequence within each functional group: the proportion within each functional group tended to increase over time, and certain functional groups (e.g., Extracellular structures (COG group W), Cell cycle control, cell division and chromosome partitioning (COG group D), Signal transduction mechanisms (COG group T), Lipid transport and metabolism (COG group I), and Cell motility (COG group N)) had higher proportions in the bovine and human host groups compared to the poultry and wild bird. However, the total number of protein differences within each functional group was smaller than the total number of samples (Figure A.7). Therefore, a regression model could not be used to model the effect of source and date of collection on the number of differences in each functional group, as a large number of isolates would have the same z-value.

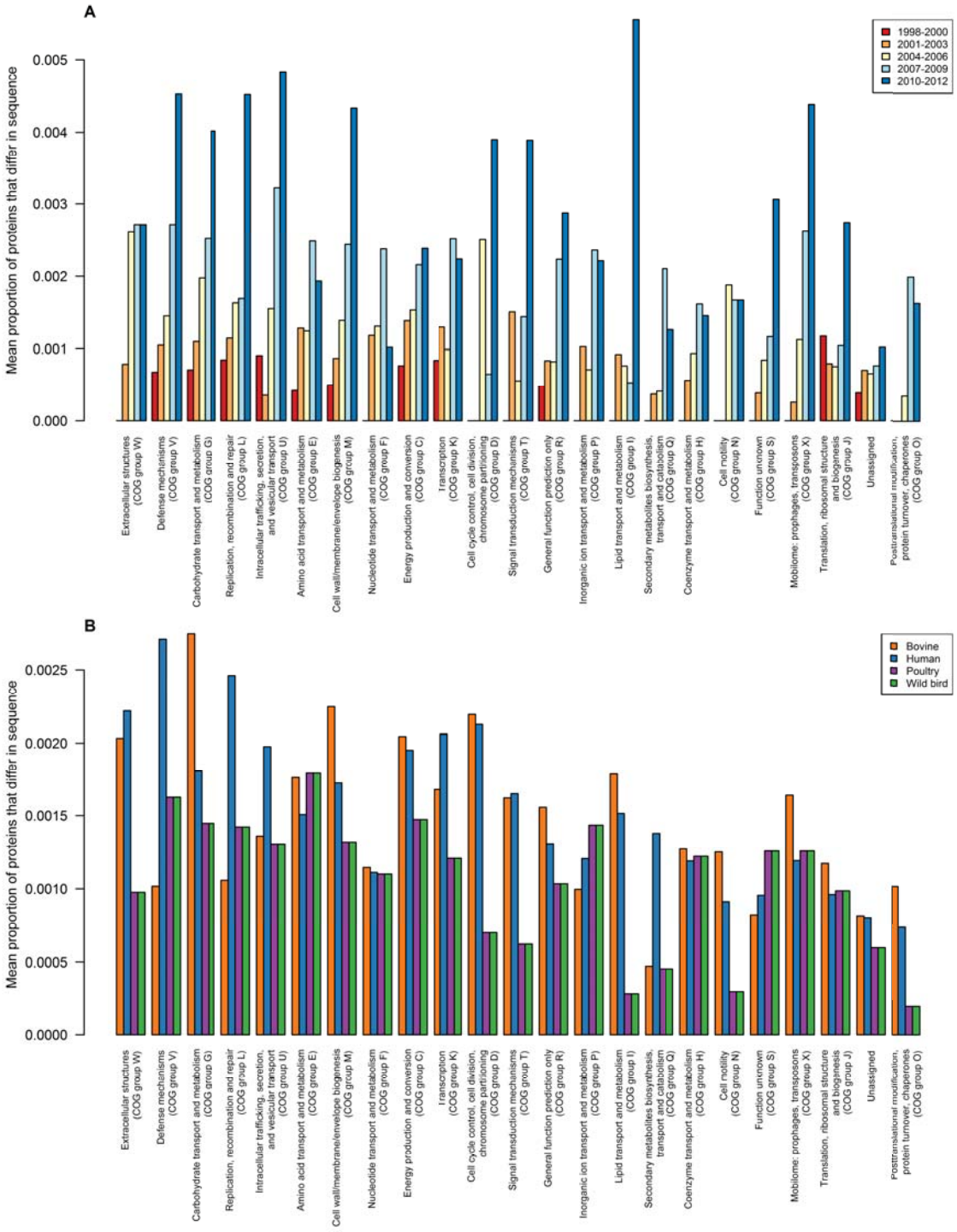


**Figure A.6** Bar graph of the mean proportion of proteins shared by 107 DT160 isolates that differ in sequence for each COG functional group within each time period (A) and source (B).

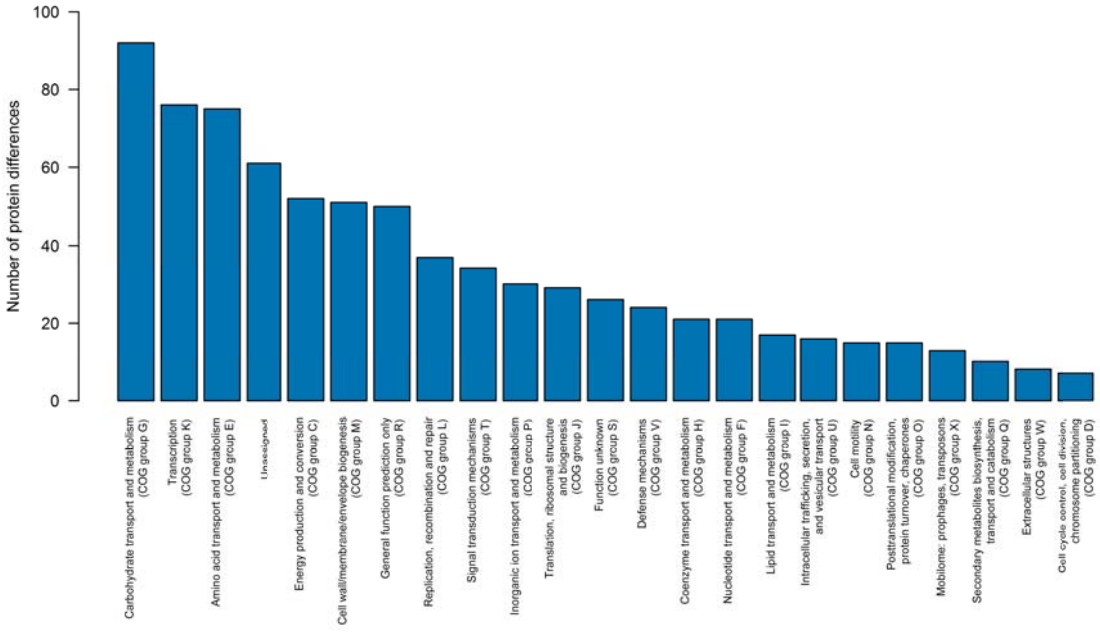


**Figure A.7** Bar graph of the number of protein difference shared by 107 DT160 isolates for each COG functional groups.

### A.3 Discrete trait analysis

The discrete trait analysis model was designed to use phenotypic or molecular data to predict the ancestral state of organisms (Lemey et al., 2009). However, the model has been applied to outbreaks to predict transmission between distinct host groups that share the same geography (Mather et al., 2013). Twenty-two datasets were formed from the 109 DT160 isolates and the 793 core SNPs they share, to determine if the discrete trait analysis model was appropriate for investigating this outbreak. The real dataset consisted of the 109 isolates split into those from animal sources (n = 74) and those from human sources (n = 35) (real dataset). Ten datasets were formed by randomly assigning the 109 isolates as animal or human, while keeping the total number of animal and human isolates the same (datasets A-J). Eleven datasets were formed by randomly assigning one of the isolates as human, while assigning the rest as animal, before progressively assigning random isolates as human, until a range of data was formed with different numbers of human and animal isolates. Each dataset was exported into BEAUti to create an XML file for BEAST 1.8.3 (Drummond et al., 2012). For simplicitys sake, each dataset was given a separate Hasegawa Kishino Yano (HYK) substitution model (Hasegawa et al., 1985) and strict molecular clock. The GMRF Bayesian skyride model (Minin et al., 2008) was used to allow for variation in the effective population size of each model and the discrete trait analysis model (Lemey et al., 2009) was used to predict the time spent in the animal and human host groups (Markov rewards) over the course of the outbreak, and the number of transmission between these host groups (Markov jumps). Each XML file was run in BEAST for 10 million steps.

The discrete trait analysis model predicted that DT160 spend most of the time in the animal host group, and that there was a larger amount of transmission from the animal to the human host group than the

reciprocal. However, the same result was obtained when the isolates were randomly assigned as human or animal, but the sample proportions were kept the same (Figures A.8 and A.9). In addition, the proportion of samples assigned as human had a significant effect on the Markov rewards and jumps (Figures A.10 and A.11). This indicates that the results obtained from the discrete trait analysis model are the result of an uneven sample size and not true migration events.

The proportion of samples that are human and Markov rewards shared a step-like or sigmoid association (Figure A.10). This is due to the deep DT160 branches that are predominantly one source until the proportion of samples that are human meets a threshold (30%-40% of samples are human), where they suddenly all switch (Figure A.12). However, the relationship between the proportion of samples that are human and Markov jumps is more complex (Figure A.11). As the proportion of samples that are human increases, the number of human branches increases, but the ancestral branches remain animal, resulting in an increase in the number of animal-to-human Markov jumps. There are no human-to-animal Markov jumps up until the threshold, as there are no ancestral branches that are human. However, after the human proportion threshold is meet, the ancestral branches switch to human, resulting in no animal-to-human Markov jumps and a large number of human-to-animal Markov jumps that decrease as the human sample proportion increases and the number of animal tips decrease. If there were no deep branches or coalescent events, I would expect the correlation between the human proportion and Markov rewards to be more linear. In addition, I would expect there to be a positive linear relationship between the human proportion and the number of each Markov jump up to the threshold and a negative linear relationship afterwards.

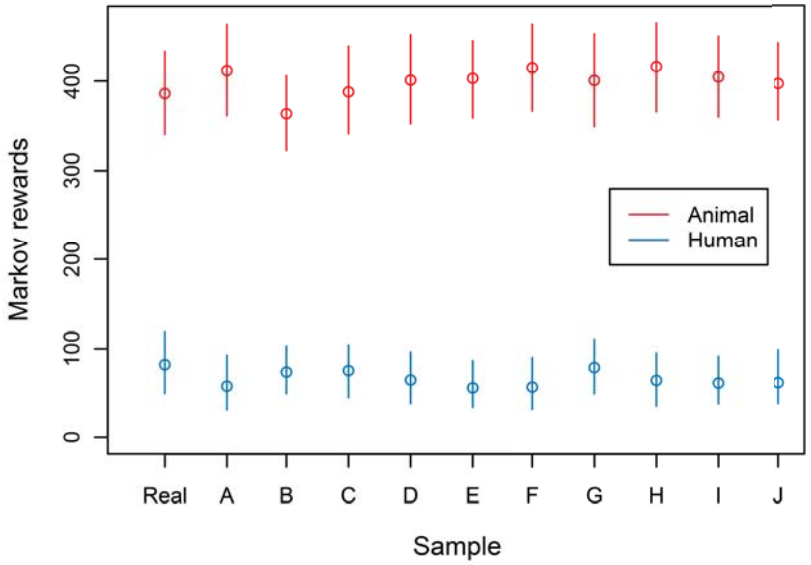


**Figure A.8** Scatterplots of the number of animal (red) and human (blue) Markov rewards estimated using the discrete trait analysis model for the real and ten randomly assigned (A-J) datasets. The circles represent the mean Markov reward value and the error bars represent the 95% HPD interval.

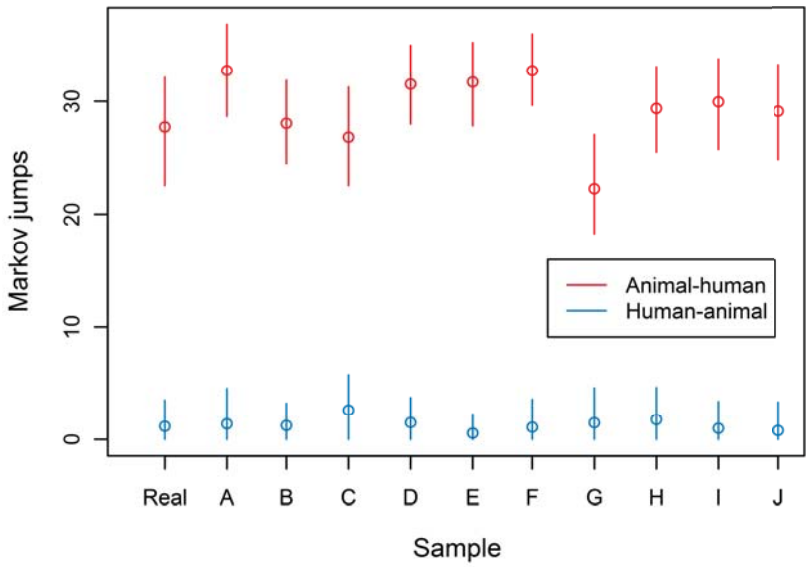


**Figure A.9** Scatterplots of the number of animal-to-human (red) and human-to-animal (blue) Markov jumps estimated using the discrete trait analysis model for the real and ten randomly assigned (A-J) datasets. The circles represent the mean Markov reward value and the error bars represent the 95% HPD interval.

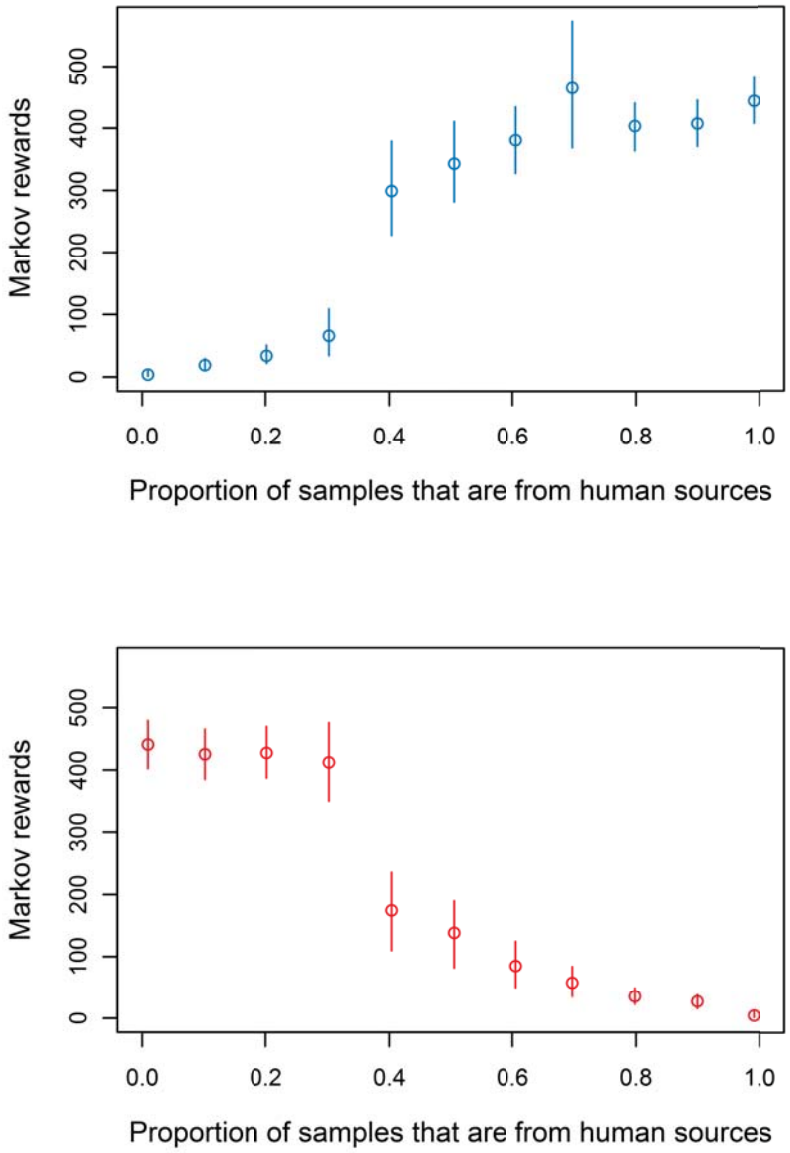


**Figure A.10** Scatterplots of the number of animal (blue) and human (red) Markov rewards estimated using the discrete trait analysis model versus the proportion of samples assigned as human. The circles represent the mean Markov reward value and the error bars represent the 95% HPD interval.

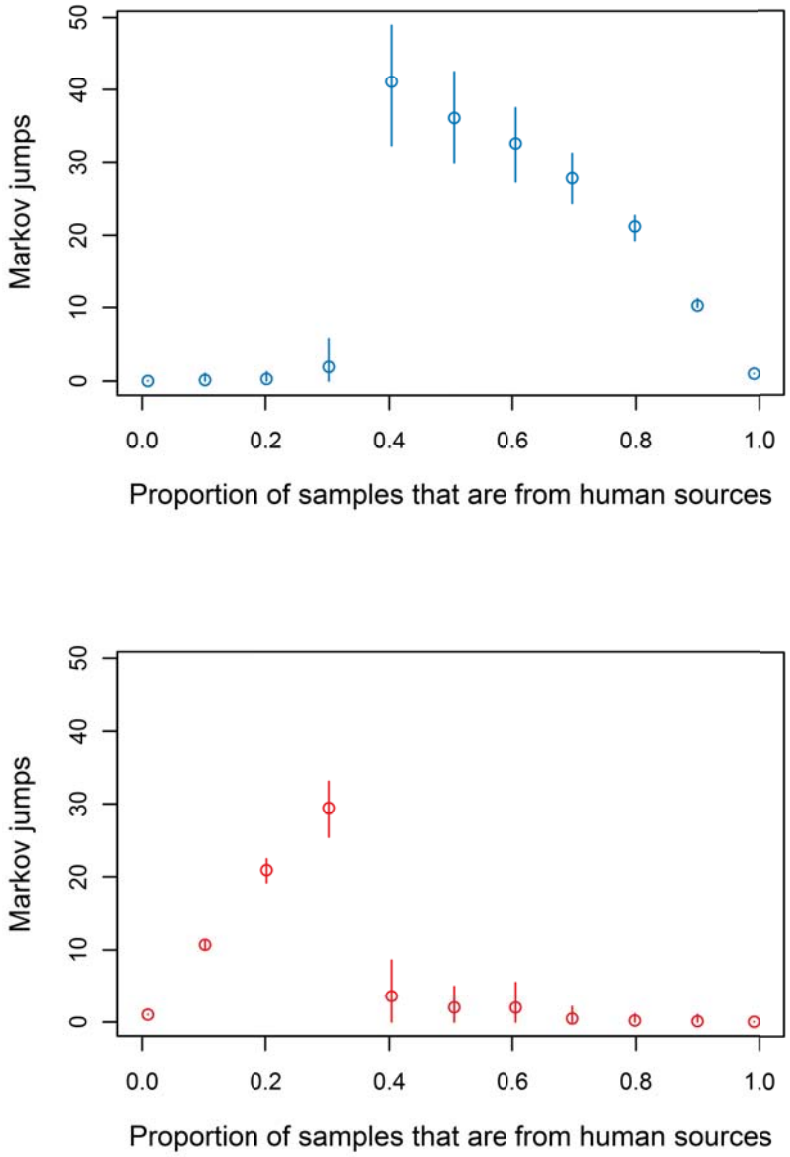


**Figure A.11** Scatterplots of the number of animal-to-human (blue) and human-to-animal (red) Markov jumps estimated using the discrete trait analysis model versus the proportion of samples assigned as human. The circles represent the mean Markov jump value and the error bars represent the 95% HPD interval.


**Figure A.12** Maximum clade credibility trees of 109 DT160 isolates placed through the discrete trait analysis model, with different proportions of isolates assigned as human (blue) and animal (red).

## A.4 Antimicrobial susceptibility testing

Omar (2010) initially investigated the New Zealand DT160 outbreak as part of their Master's project. Part of this project involved investigating the antimicrobial susceptibility profiles of 90 DT160 isolates using disc diffusion: 30 human, 30 wild bird and 30 poultry isolates. Most of these isolates were re-analysed in this thesis using whole genome sequencing (Chapter 3)

Omar's antimicrobial susceptibility results were obtained and re-analysed using regression models, with date of collection and source as explanatory variables. There was insufficient evidence to suggest that date of collection or source had any effect on the antimicrobial susceptibility results (Figures A.13, A.15, A.17, A.19, A.21, A.25, A.27, A.29, A.31, A.33). The only association was between chloramphenicol zone of inhibition values and year of collection (Figure A.23). However, the coefficient for date of collection was very small and the there was a lack of residual normality for this model, as with the other models tested (Figures A.14, A.16, A.18, A.20, A.22, A.24, A.26, A.28, A.30, A.32, A.34). In addition, disc diffusion is not a very sensitive assay (Sjölund-Karlsson et al., 2014). Therefore, further testing is required to determine the significance of these results.

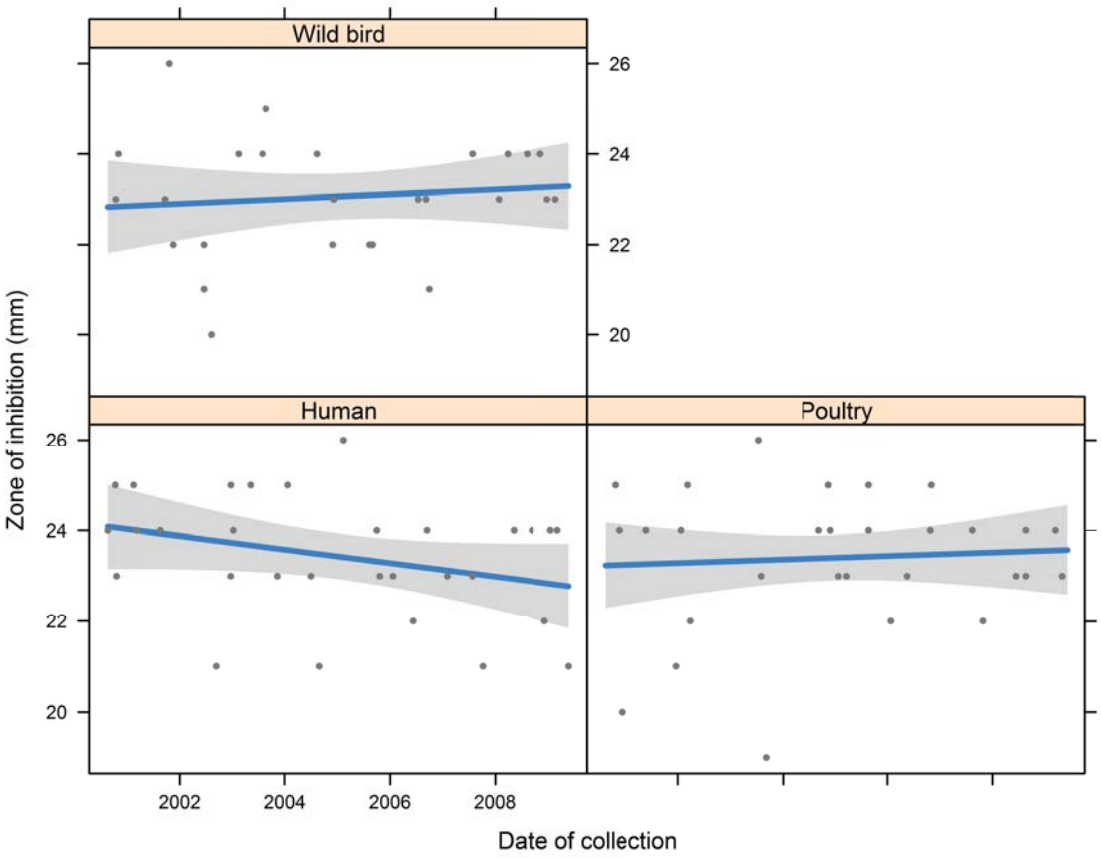


**Figure A.13** Scatterplots of date of collection versus amikacin antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates. The blue line represents the regression equation and the grey area represents the standard error for this equation. There was insufficient evidence to suggest that date of collection (p-value = 0.6058) or source (p-value = 0.5508) was associated with amikacin antimicrobial resistance, or there was an interaction between source and date of collection (p = 0.2304).

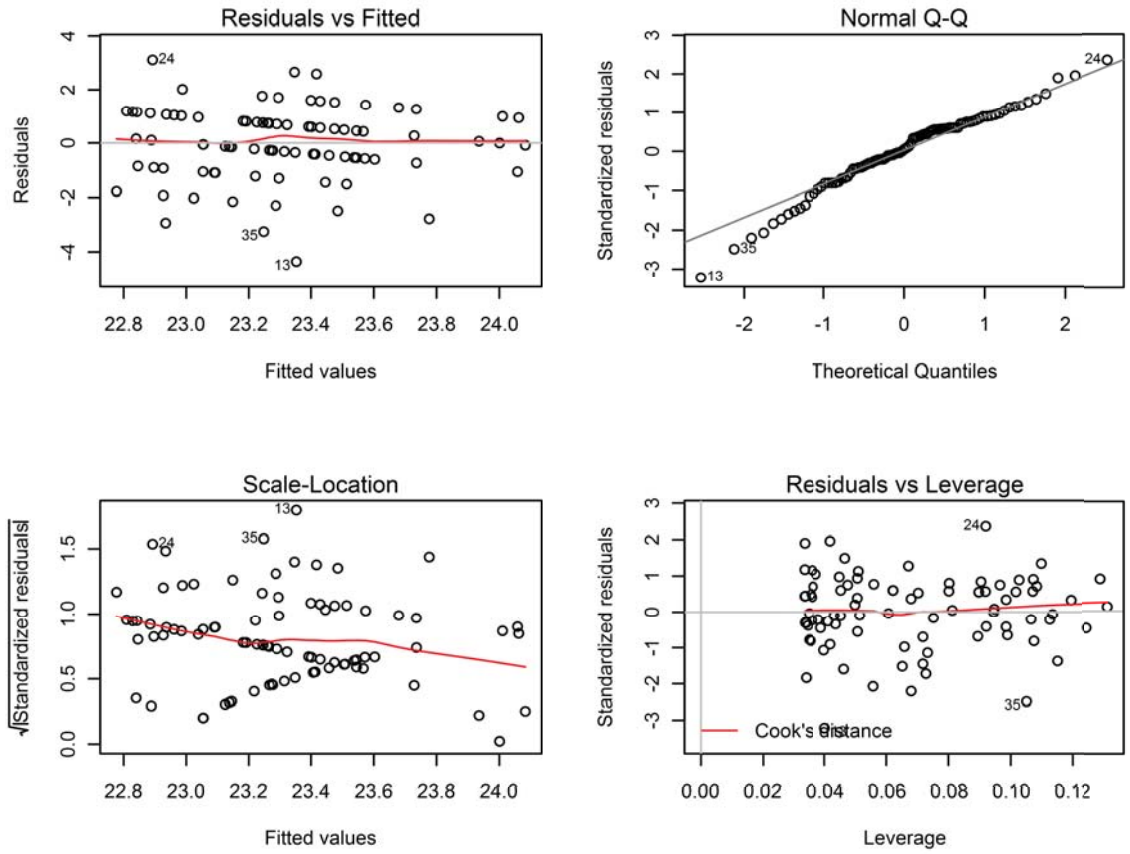


**Figure A.14** Diagnostic plots of the regression model fitted to the amikacin antimicrobial susceptibility results for 90 DT160 isolates.

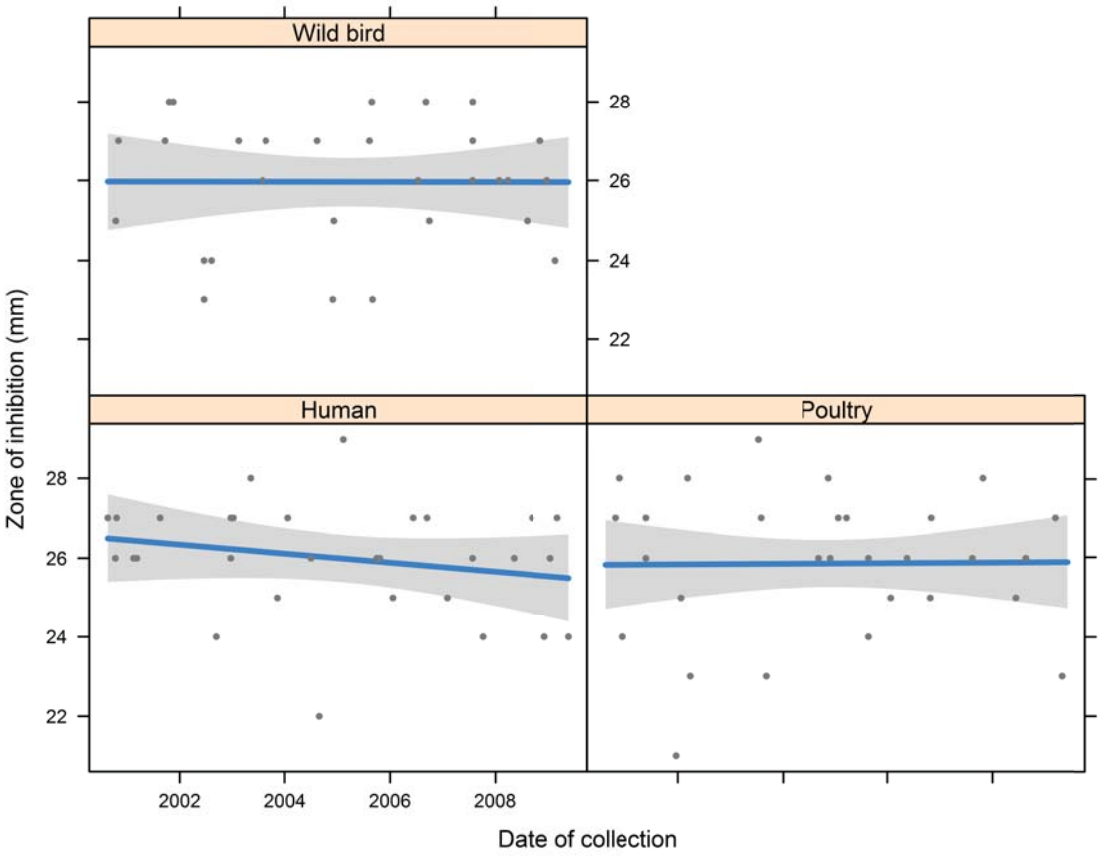


**Figure A.15** Scatterplots of date of collection versus amoxicillin/clavulanate antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates. The blue line represents the regression equation and the grey area represents the standard error for this equation. There was insufficient evidence to suggest that date of collection (p-value = 0.5458) or source (p-value = 0.9416) was associated with amoxicillin/clavulanate antimicrobial resistance, or there was an interaction between source and date of collection (p = 0.6779).

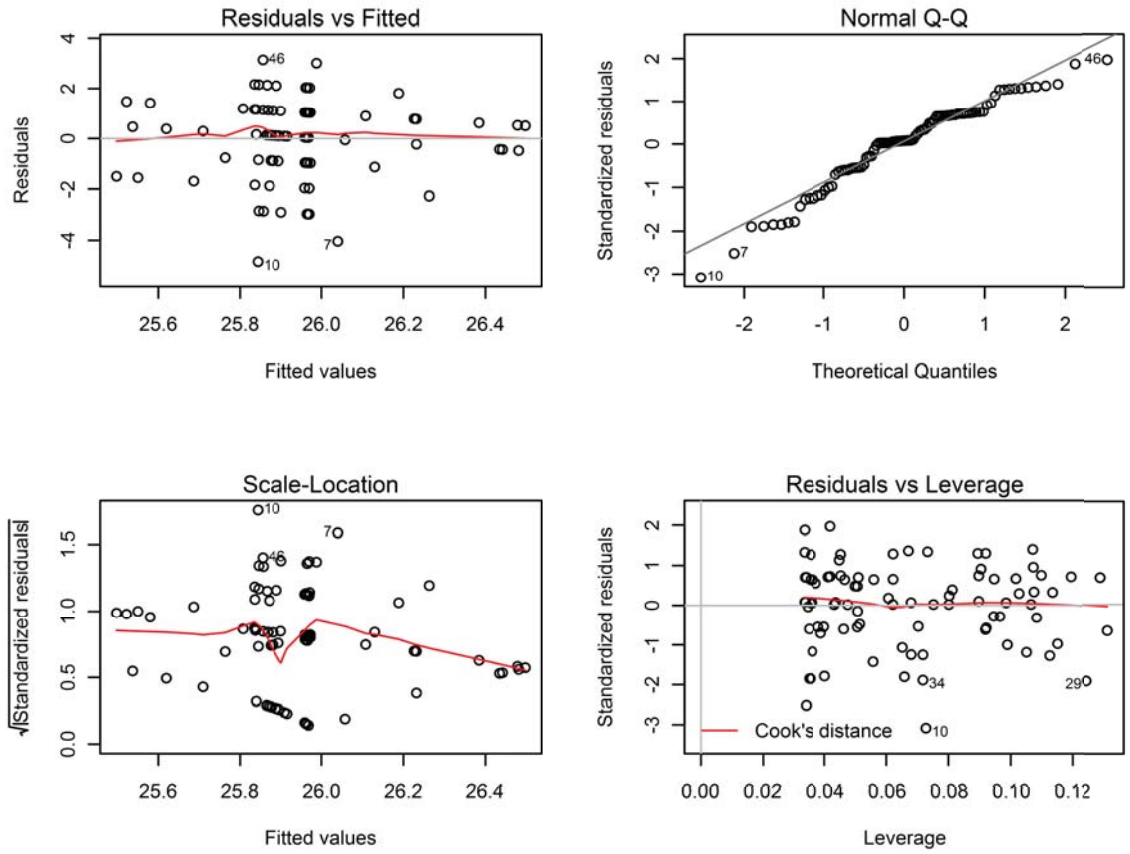


**Figure A.16** Diagnostic plots of the regression model fitted to the amoxicillin/clavulanate antimicrobial susceptibility results for 90 DT160 isolates.

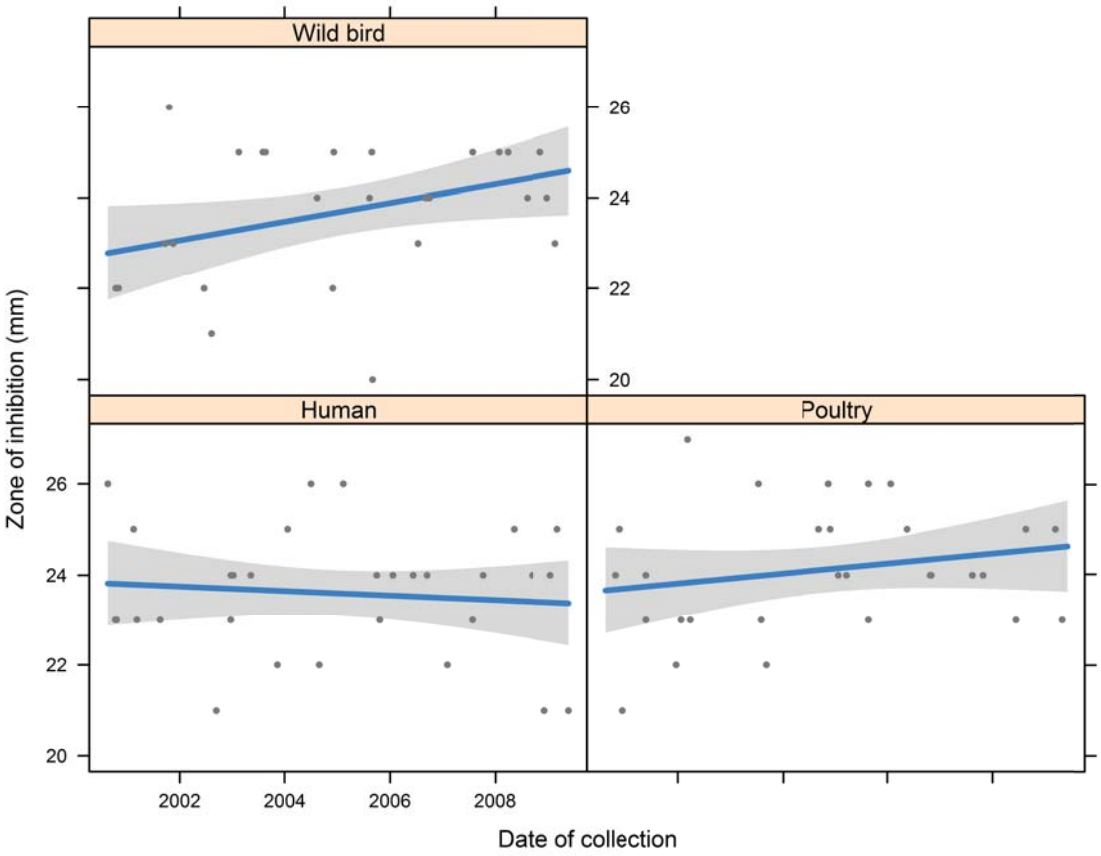


**Figure A.17** Scatterplots of date of collection versus ampicillin antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates. The blue line represents the regression equation and the grey area represents the standard error for this equation. There was insufficient evidence to suggest that date of collection (p-value = 0.1549) or source (p-value = 0.2888) was associated with ampicillin antimicrobial resistance, or there was an interaction between source and date of collection (p = 0.1592).

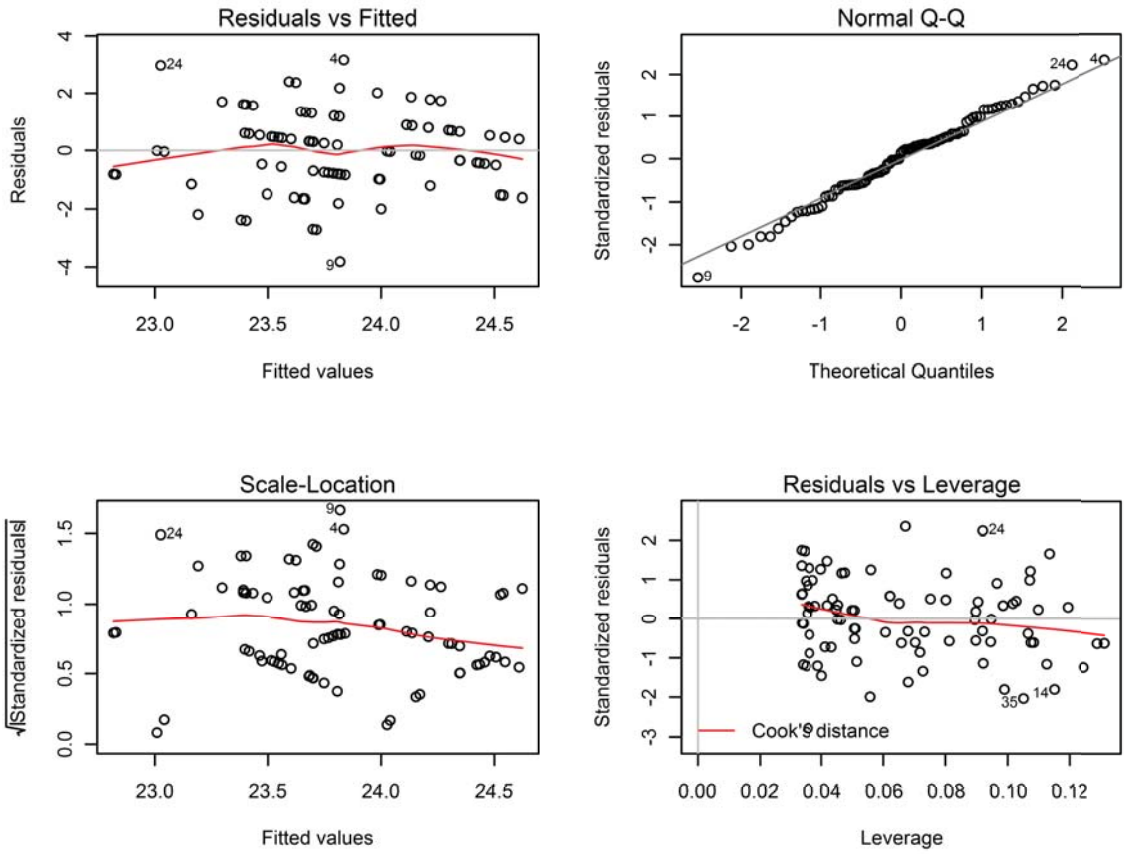


**Figure A.18** Diagnostic plots of the regression model fitted to the ampicillin antimicrobial susceptibility results for 90 DT160 isolates.

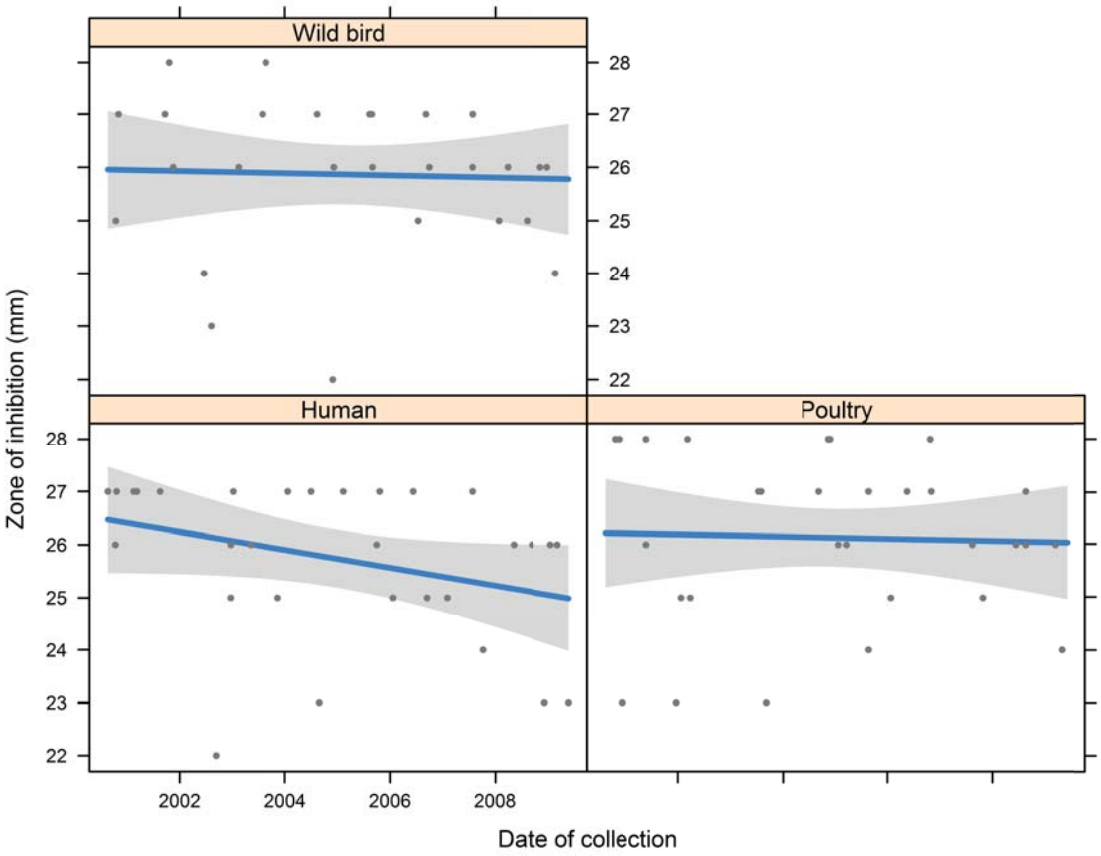


**Figure A.19** Scatterplots of date of collection versus cefoxitin antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates. The blue line represents the regression equation and the grey area represents the standard error for this equation. There was insufficient evidence to suggest that date of collection (p-value = 0.1983) or source (p-value = 0.5945) was associated with cefoxitin antimicrobial resistance, or there was an interaction between source and date of collection (p = 0.4784).

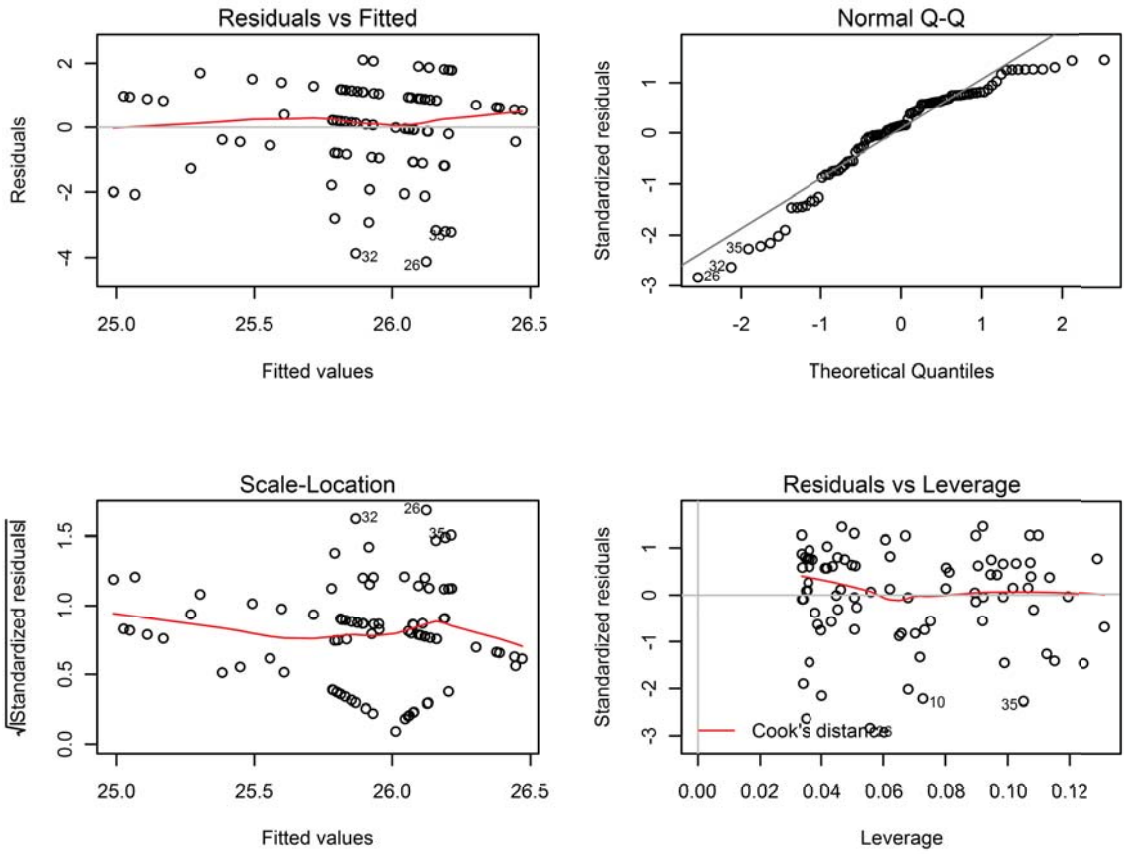


**Figure A.20** Diagnostic plots of the regression model fitted to the cefoxitin antimicrobial susceptibility results for 90 DT160 isolates.

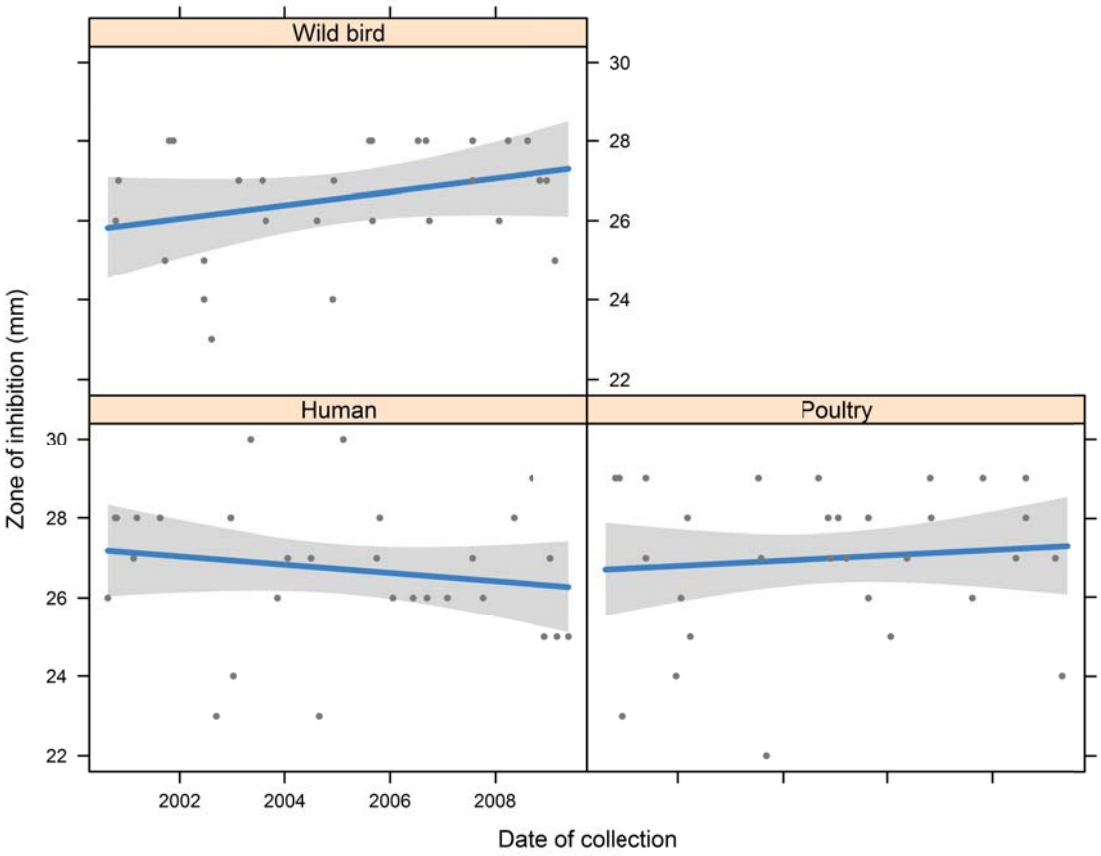


**Figure A.21** Scatterplots of date of collection versus cefpodoxime antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates. The blue line represents the regression equation and the grey area represents the standard error for this equation. There was insufficient evidence to suggest that date of collection (p-value = 0.6261) or source (p-value = 0.6291) was associated with cefpodoxime antimicrobial resistance, or there was an interaction between source and date of collection (p = 0.2453).

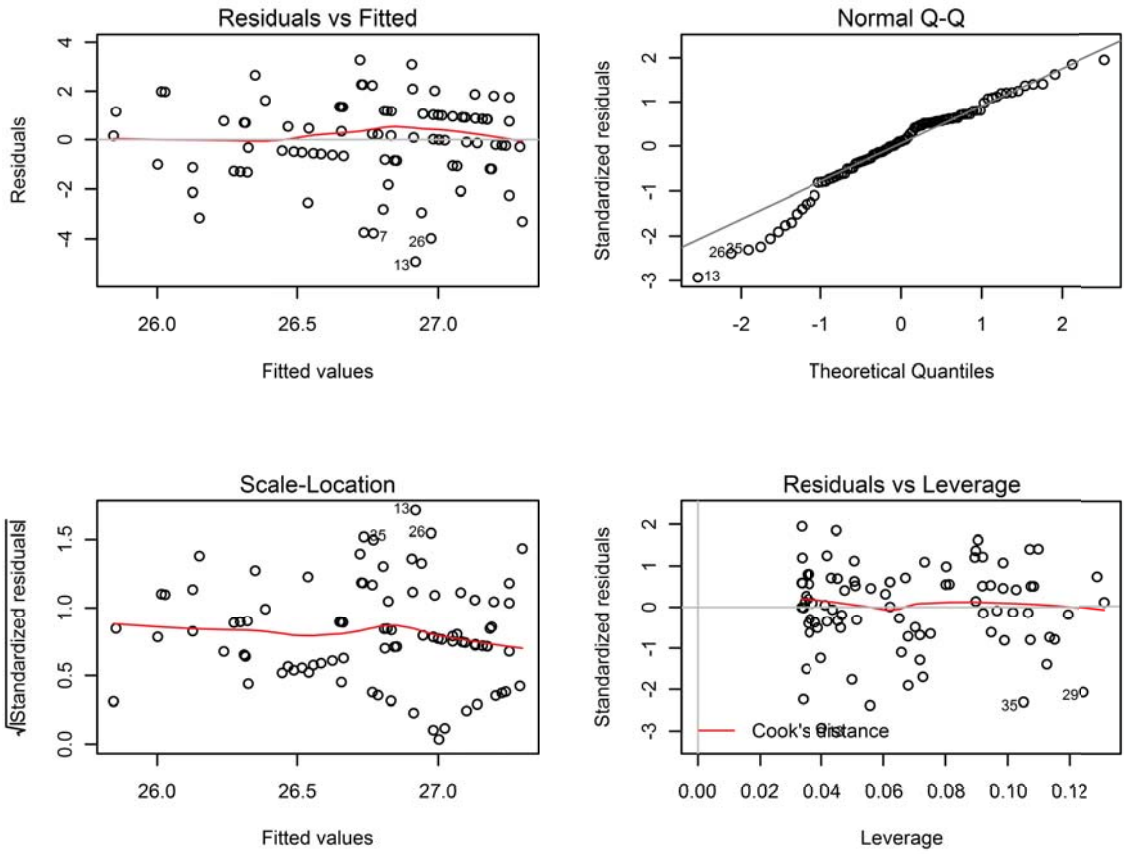


**Figure A.22** Diagnostic plots of the regression model fitted to the cefpodoxime antimicrobial susceptibility results for 90 DT160 isolates.

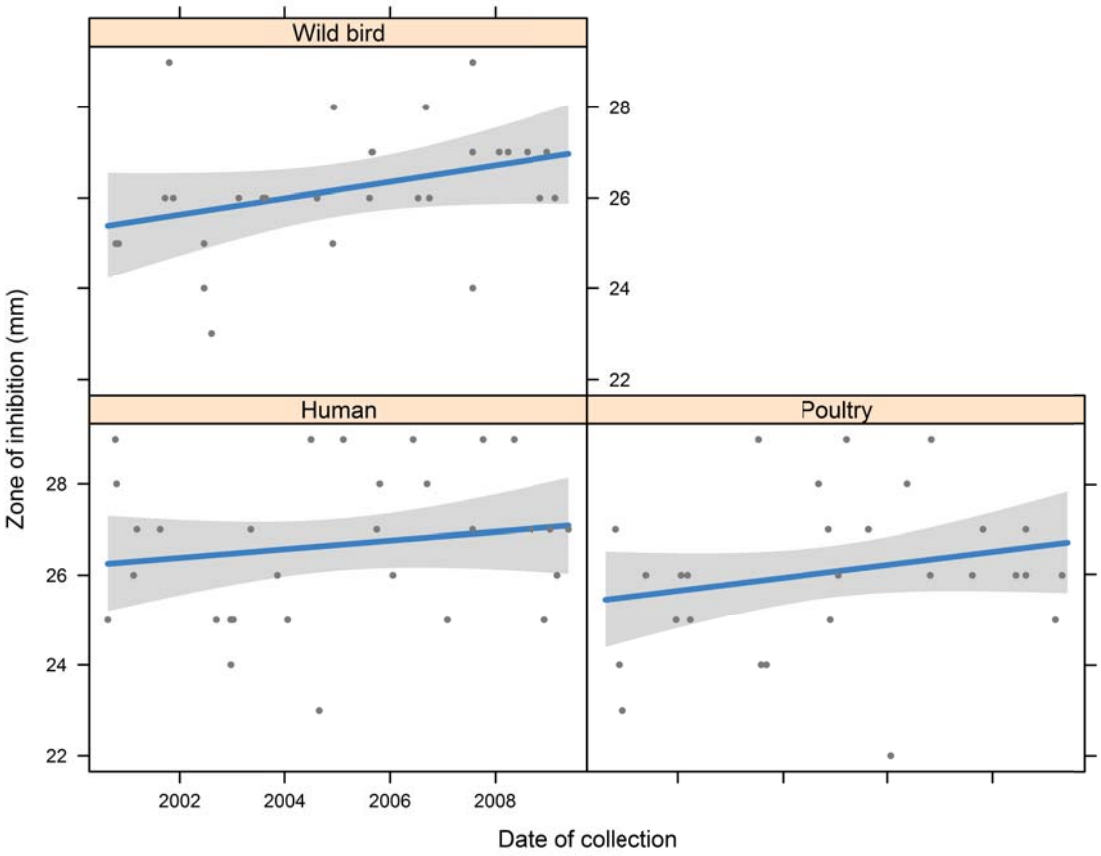


**Figure A.23** Scatterplots of date of collection versus chloramphenicol antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates. The blue line represents the regression equation and the grey area represents the standard error for this equation. Date of collection was significantly associated with chloramphenicol antimicrobial resistance in this model (p-value = 0.0275), but there was insufficient evidence to suggest that soource (p-value = 0.3066) was associated with chloramphenicol antimicrobial resistance, or there was an interaction between source and date of collection (p = 0.8514).

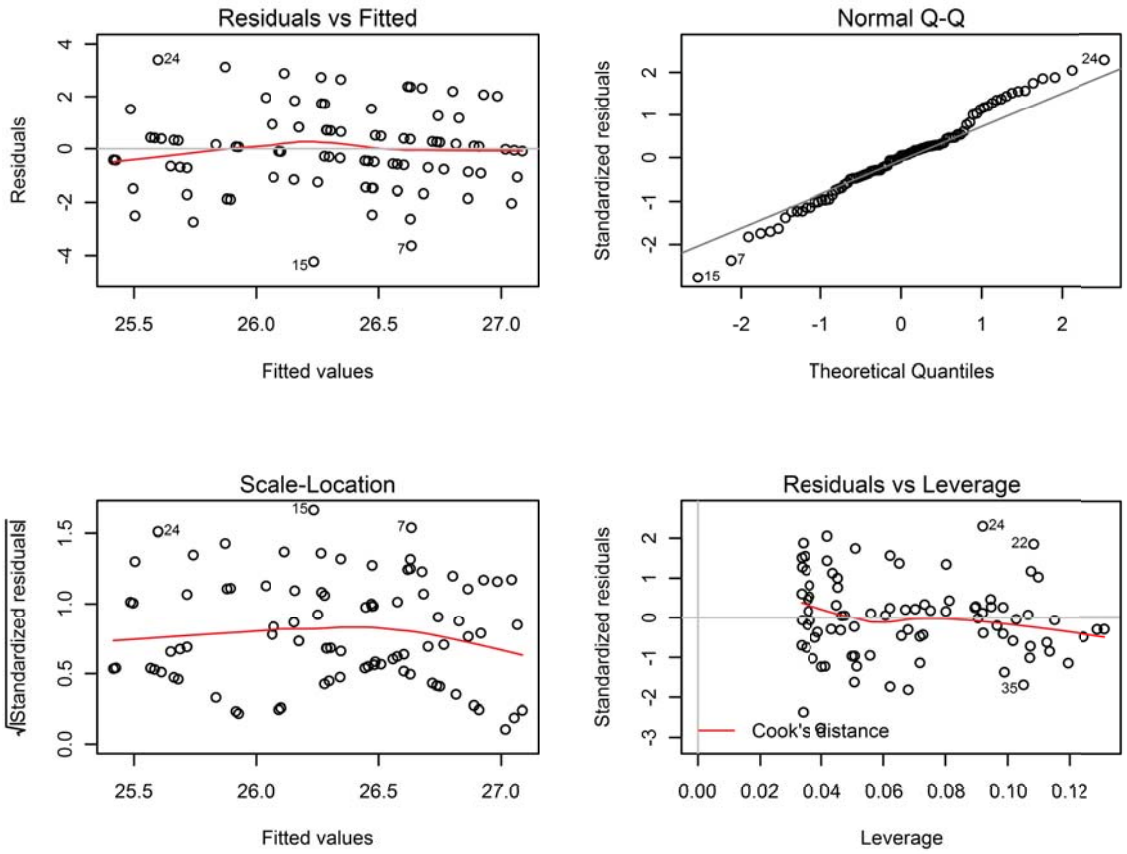


**Figure A.24** Diagnostic plots of the regression model fitted to the chloramphenicol antimicrobial susceptibility results for 90 DT160 isolates.

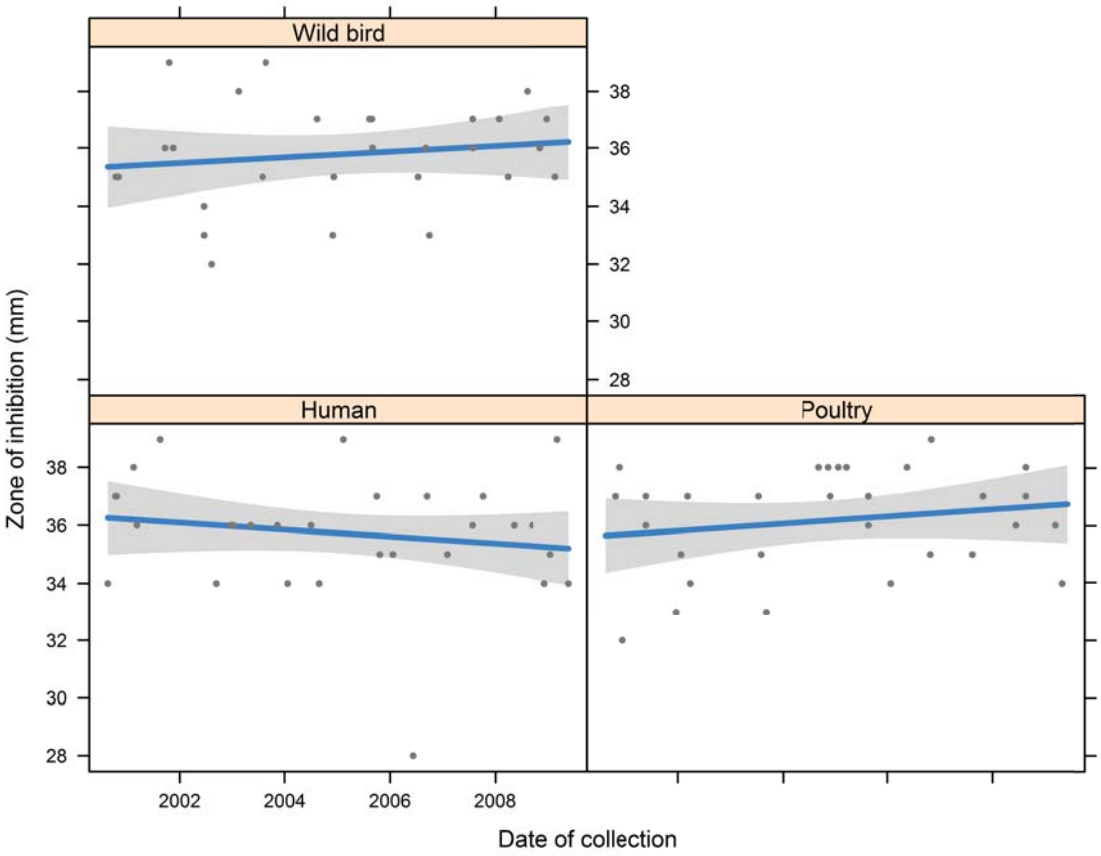


**Figure A.25** Scatterplots of date of collection versus ciprofloxacin antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates. The blue line represents the regression equation and the grey area represents the standard error for this equation. There was insufficient evidence to suggest that date of collection (p-value = 0.7385) or source (p-value = 0.6244) was associated with ciprofloxacin antimicrobial resistance, or there was an interaction between source and date of collection (p = 0.3264).

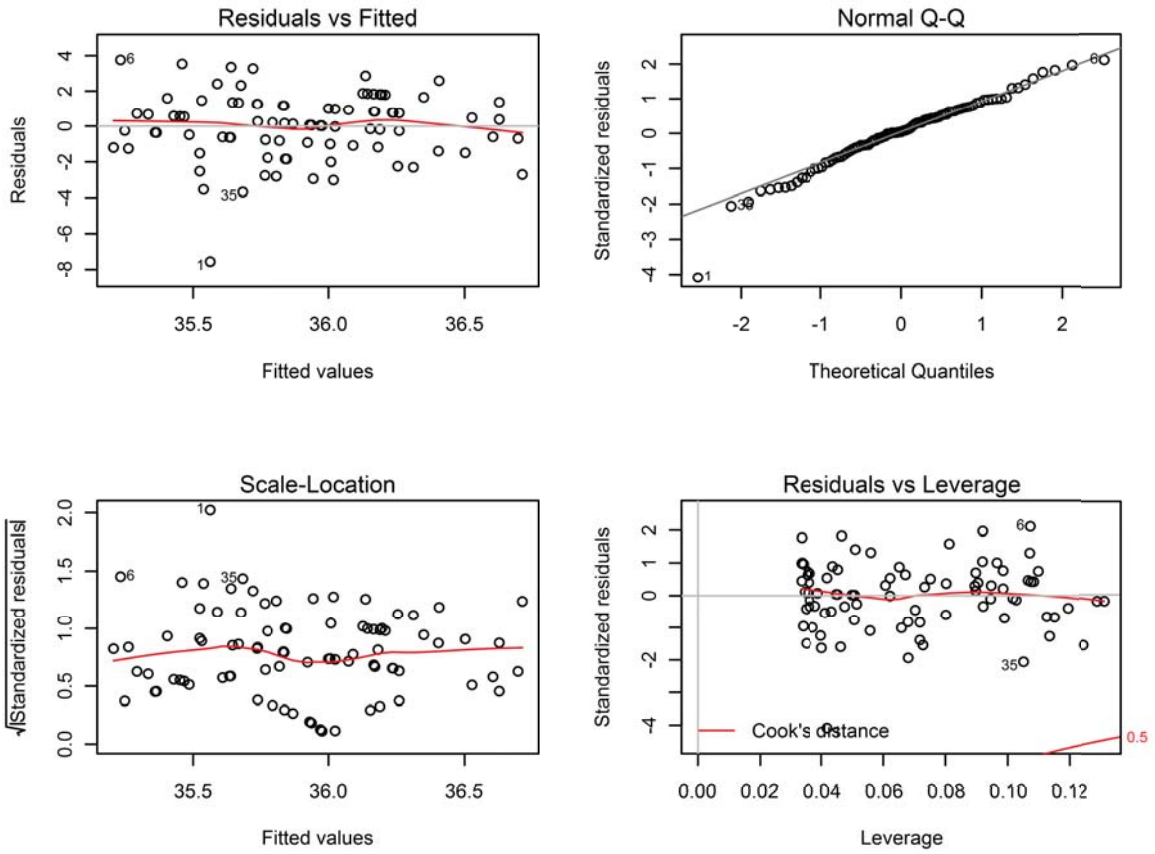


**Figure A.26** Diagnostic plots of the regression model fitted to the ciprofloxacin antimicrobial susceptibility results for 90 DT160 isolates.

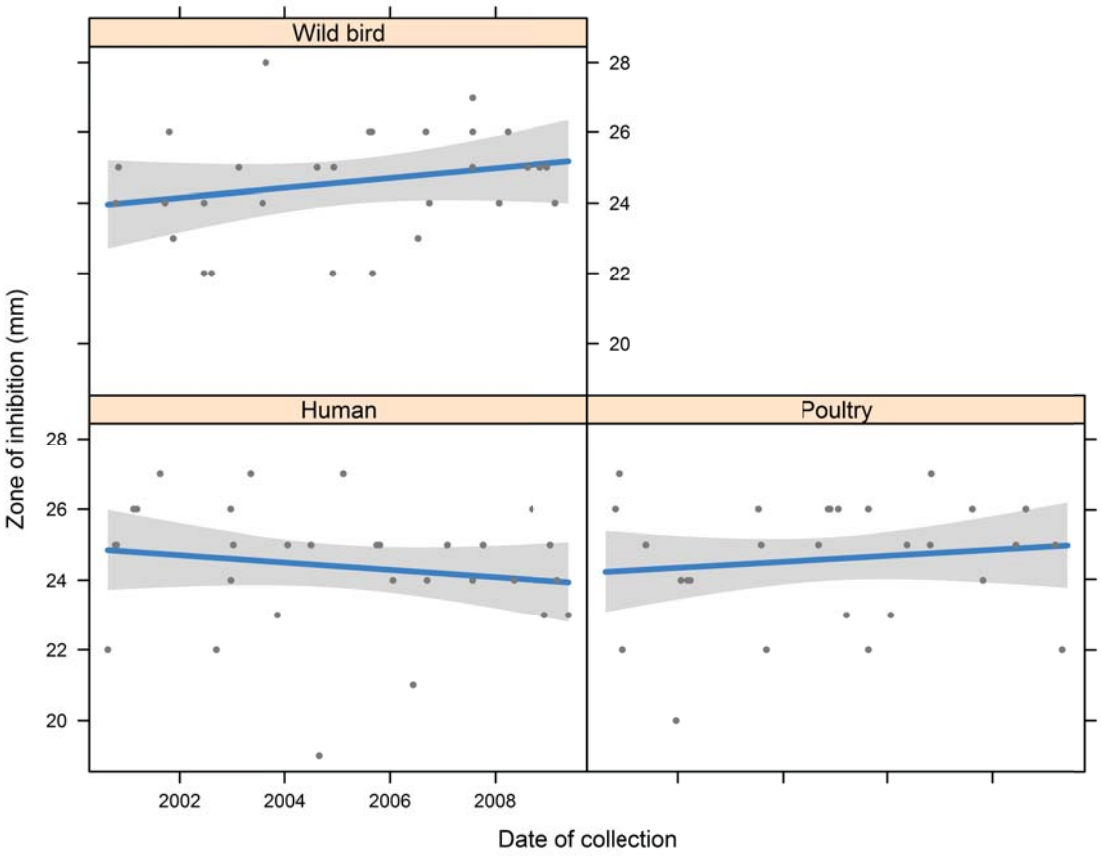


**Figure A.27** Scatterplots of date of collection versus nalidixic acid antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates. The blue line represents the regression equation and the grey area represents the standard error for this equation. There was insufficient evidence to suggest that date of collection (p-value = 0.6219) or source (p-value = 0.8775) was associated with nalidixic acid antimicrobial resistance, or there was an interaction between source and date of collection (p = 0.2842).

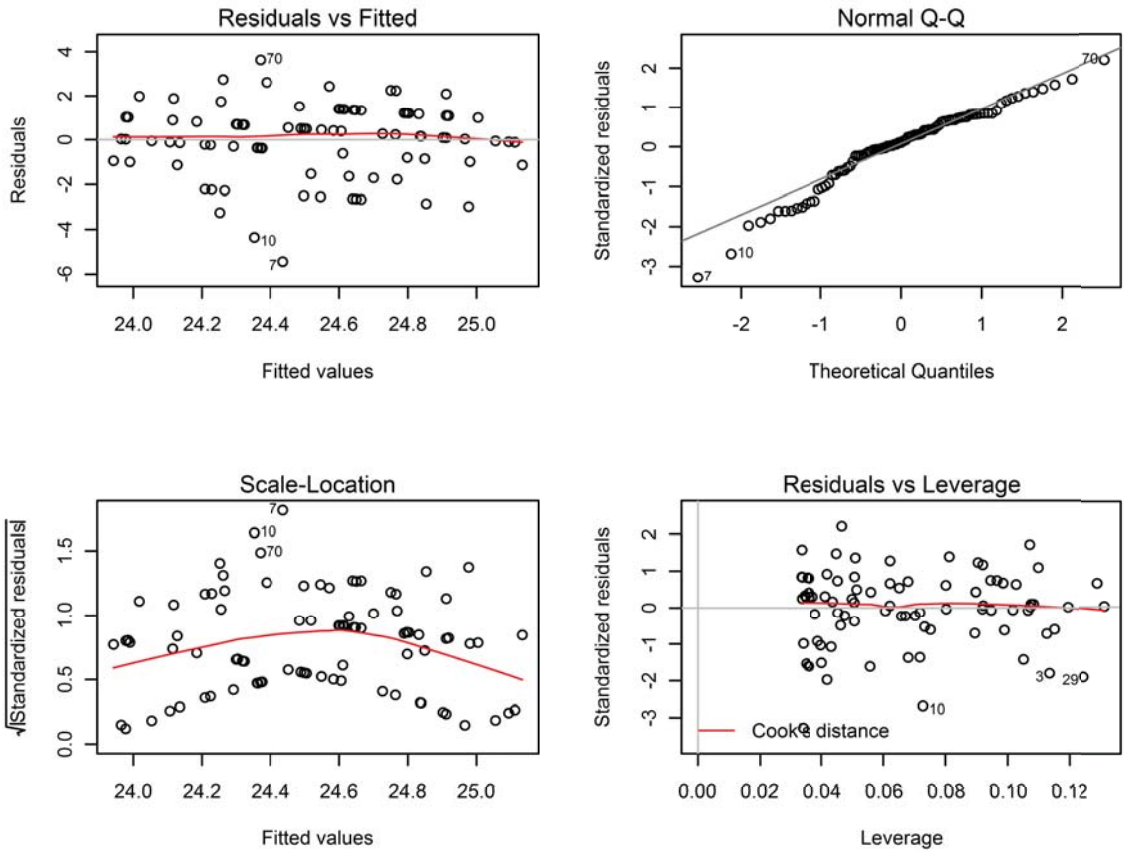


**Figure A.28** Diagnostic plots of the regression model fitted to the nalidixic acid antimicrobial susceptibility results for 90 DT160 isolates.

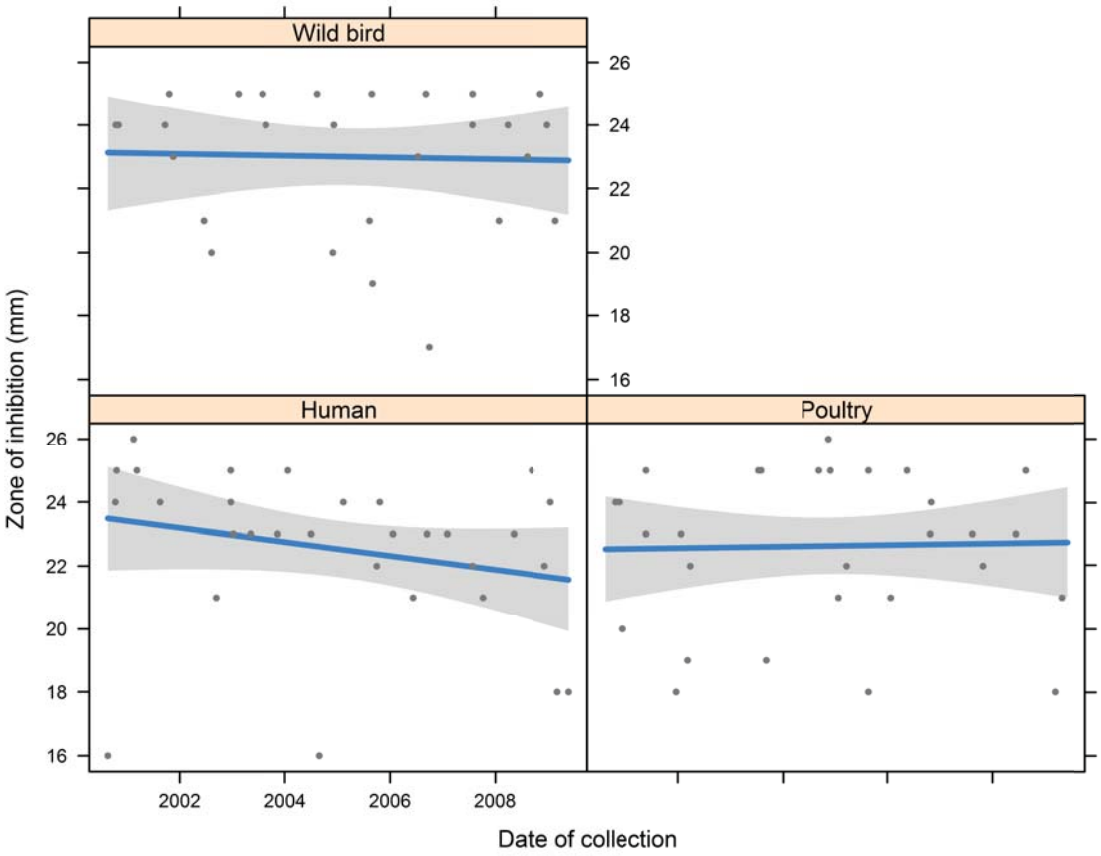


**Figure A.29** Scatterplots of date of collection versus oxytetracycline antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates. The blue line represents the regression equation and the grey area represents the standard error for this equation. There was insufficient evidence to suggest that date of collection (p-value = 0.4195) or source (p-value = 0.7219) was associated with oxytetracycline antimicrobial resistance, or there was an interaction between source and date of collection (p = 0.5350).

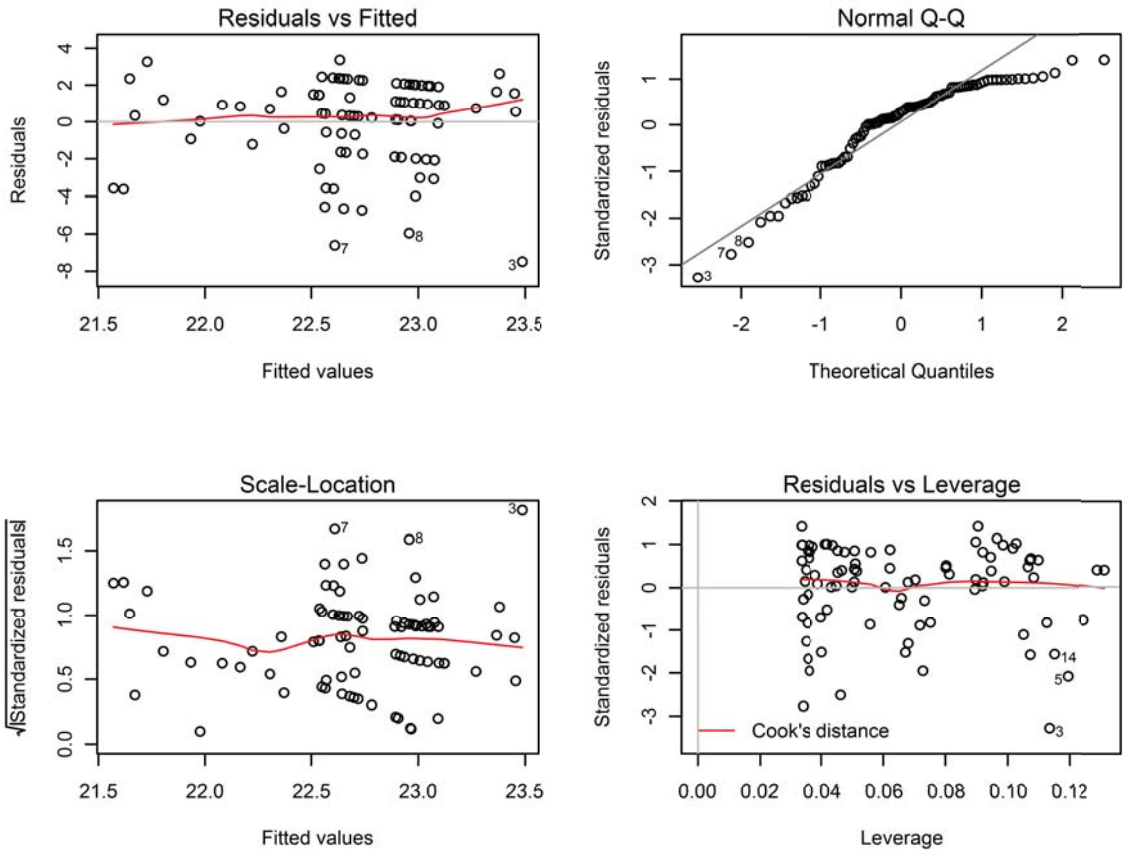


**Figure A.30** Diagnostic plots of the regression model fitted to the oxytetracycline antimicrobial susceptibility results for 90 DT160 isolates.

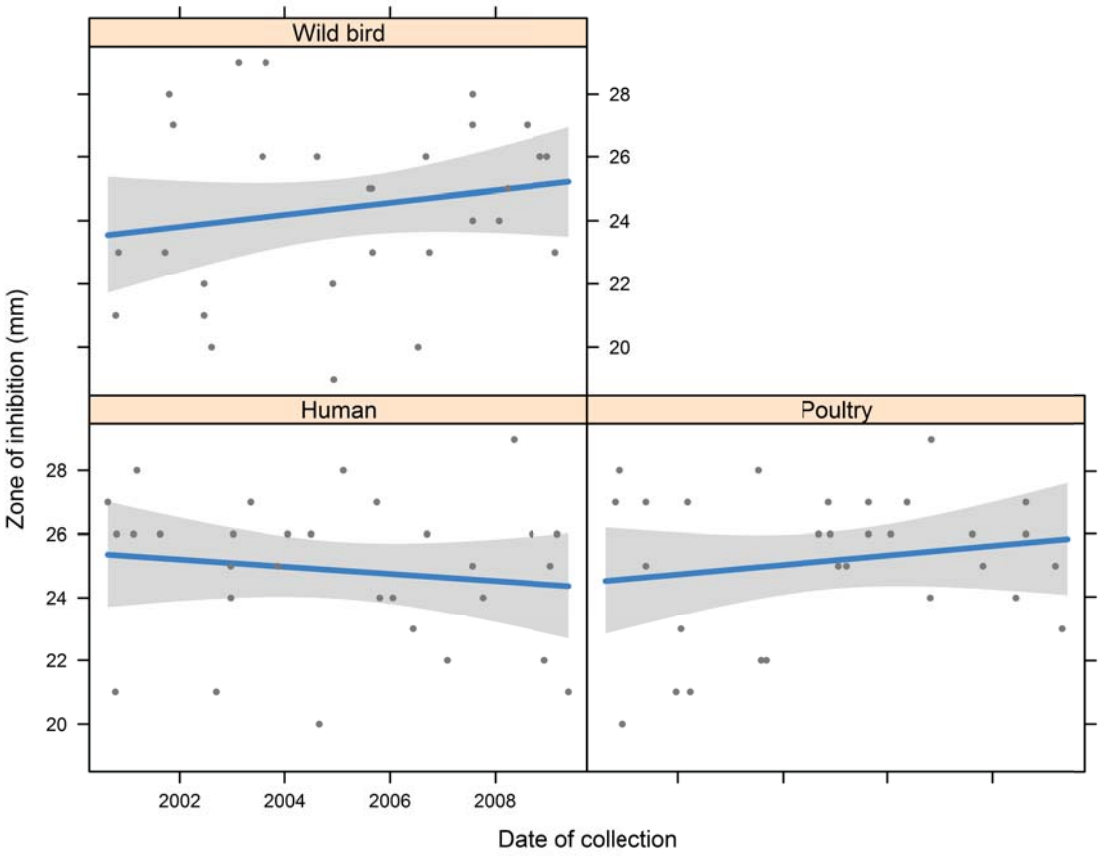


**Figure A.31** Scatterplots of date of collection versus trimethoprim/sulfamethoxazole antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates. The blue line represents the regression equation and the grey area represents the standard error for this equation. There was insufficient evidence to suggest that date of collection (p-value = 0.6058) or source (p-value = 0.5508) was associated with trimethoprim/sulfamethoxazole antimicrobial resistance, or there was an interaction between source and date of collection (p = 0.2304).

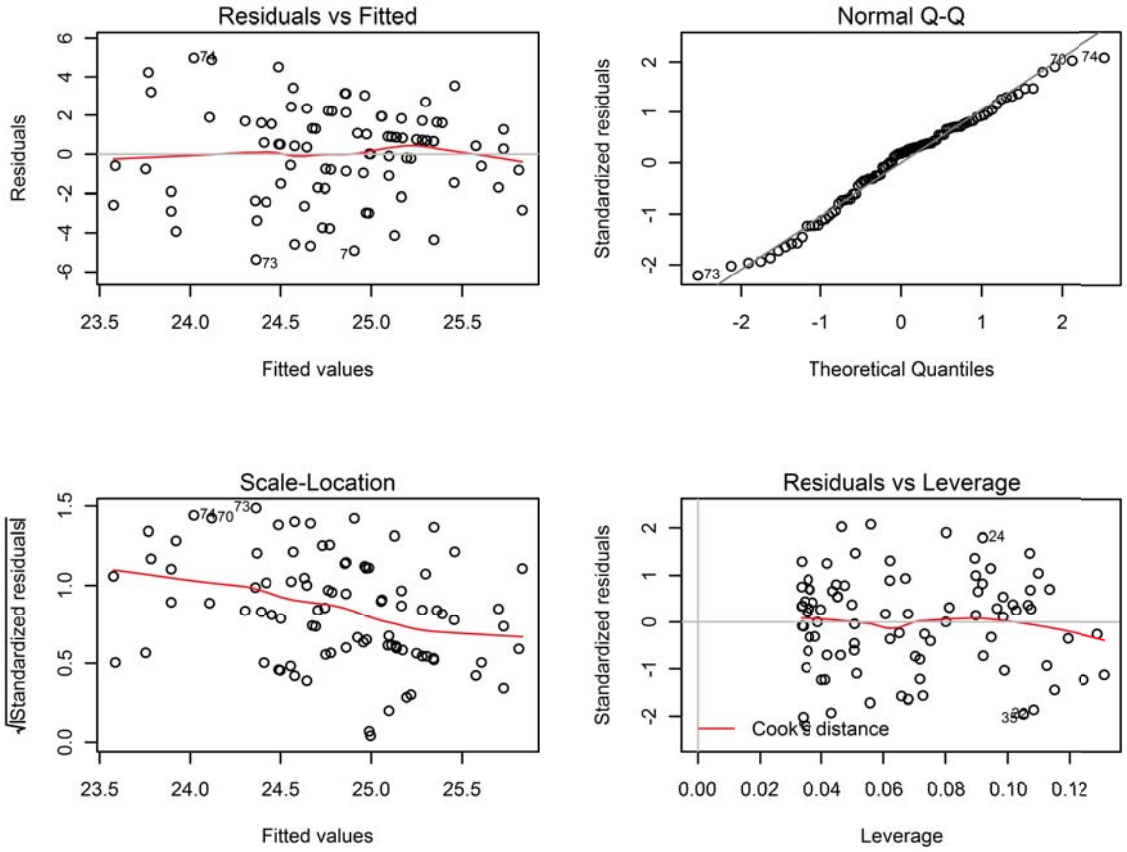


**Figure A.32** Diagnostic plots of the regression model fitted to the trimethoprim/sulfamethoxazole antimicrobial susceptibility results for 90 DT160 isolates.

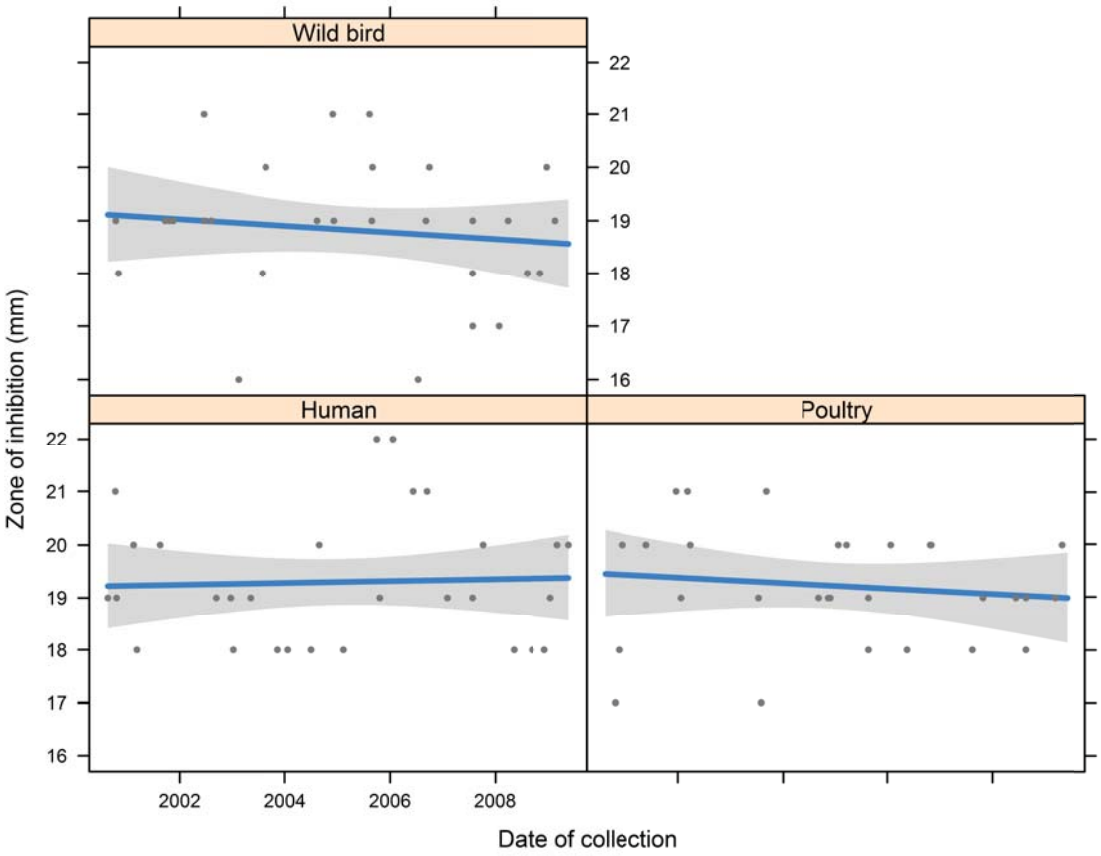


**Figure A.33** Scatterplots of date of collection versus tetracycline antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates. The blue line represents the regression equation and the grey area represents the standard error for this equation. There was insufficient evidence to suggest that date of collection (p-value = 0.4842) or source (p-value = 0.2841) was associated with tetracycline antimicrobial resistance, or there was an interaction between source and date of collection (p = 0.7446).

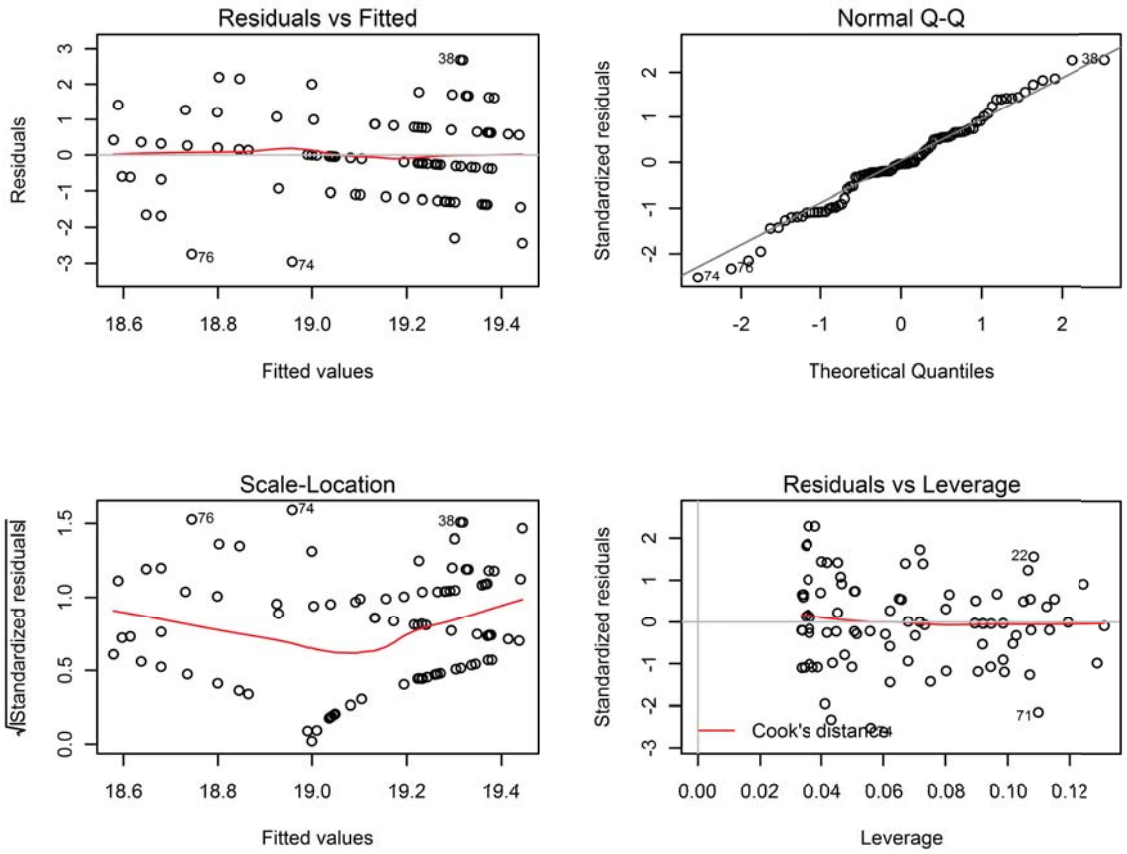


**Figure A.34** Diagnostic plots of the regression model fitted to the tetracycline antimicrobial susceptibility results for 90 DT160 isolates.

# References


Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, 62(1):245–253.

Clarke, K. R. and Gorley, R. N. (2006). *PRIMER v6:User manual/tutorial, pages 1–192*. PRIMER-E, Plymouth, United Kingdom.

Drummond, A., Suchard, M., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8):1969–1973.

Gardner, S. N., Slezak, T., and Hall, B. G. (2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics*, 31:2877–2878.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2):160–174.

Lemey, P., Rambaut, A., Drummond, A., and Suchard, M. (2009). Bayesian phylogeography finds its roots. *PLoS Computational Biology*, 5(9):1–16.

Mather, A., Reid, S., Maskell, D., Parkhill, J., Fookes, M., Harris, S., Brown, D., Coia, J., Mulvey, M., Gilmour, M. o., Petrovska, L., De Pinna, E., Kuroda, M., Akiba, M., Izumiya, H., Connor, T., Suchard, M. l., Lemey, P., Mellor, D., Haydon, D., and Thomson, N. (2013). Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science*, 341(6153):1514–1517.

Minin, V., Bloomquist, E., and Suchard, M. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, 25(7):1459–1471.

Omar, S. (2010). *Molecular epidemiology of Salmonella Typhimurium DT160 in New Zealand*. PhD thesis, Massey University.

Sjölund-Karlsson, M., Howie, R. L., Crump, J. A., and Whichard, J. M. (2014). Fluoroquinolone susceptibility testing of *Salmonella enterica*: Detection of acquired resistance and selection of zone diameter breakpoints for levofloxacin and ofloxacin. *Journal of Clinical Microbiology*, 52(3):877–884.

# Appendix B

## Supplementary material to Chapter 4

### B.1 Simulated outbreaks

The outbreaks simulated in this study were designed to represent possible salmonellosis outbreaks in New Zealand. This was achieved by using host population and infectious period parameters of humans and animals in New Zealand.

#### B.1.1 Population size

The human population in New Zealand increased linearly from 3.8 million in 1998 to 4.6 million in 2015 (Statistics New Zealand, 2017b). The susceptible human host population size was taken from this statistic (3.8-4.6 million).

Cows and chickens are the primary sources of salmonellosis outbreaks (Gould et al., 2013). The number of cows in New Zealand fluctuated from 7.7-9.3 million in New Zealand from 1971-1999, whilst the number of chickens in New Zealand fluctuated from 18.3-21.9 million from 2002-2014 in New Zealand (Statistics New Zealand, 2017a). Population records outside these time periods are scant for these animals. The susceptible host animal population size was taken by summing up these populations (25-34 million).

**Table B.1** Initial host population sizes for simulated outbreaks

| Population | Minimum | Maximum |
|---|---|---|
| Susceptible animal | 25,000,000 | 34,000,000 |
| Susceptible human | 3,800,000 | 4,600,000 |
| Infected animal/human | 1 | 1 |

### B.1.2 Transmission rates

A large range of transmission rates between the animal and human host populations were used as these were the parameters being estimated in this study.

**Table B.2** Beta values for simulated outbreaks

| Population | | $\beta$ (year$^{-1}$ I$^{-1}$ S$^{-1}$) | |
|---|---|---|---|
| Infectious (I) | Susceptible (S) | Minimum | Maximum |
| Animal | Animal | 1 x $10^{-9}$ | 1.22 x $10^{-6}$ |
| Human | Human | 1 x $10^{-9}$ | 2.97 x $10^{-6}$ |
| Animal | Human | 1 x $10^{-9}$ | 6.13 x $10^{-6}$ |
| Human | Animal | 1 x $10^{-9}$ | 2.18 x $10^{-6}$ |

### B.1.3 Infectious periods

The infectious period values used for the animal and human host populations were estimated from studies that measured the average length of *Salmonella* excretion after exposure. Buchwald and Blaser (1984) reviewed 32 articles on human non-typhoid *Salmonella* excretion and found that patients excreted *Salmonella* for approximately 5 weeks. Murase et al. (2000) investigated an outbreak of *Salmonella enterica* serovar Typhimurium in Japan and found that on average asymptomatic and symptomatic patients shed *Salmonella* for 4 weeks. The papers that Buchwald and Blaser reviewed focused on symptomatic salmonellosis patients, whilst Murase et al. considered asymptomatic patients that excreted *Salmonella* for a shorter length of time. This explains why Murase et al. predicted a lower average *Salmonella* excretion rate than Buchwald and Blaser. However, Murase et al.'s study only represents a single outbreak. Therefore, the human *Salmonella* infectious period was taken from both studies (4-5 weeks).

The animal infectious period was estimated using studies on cow and chicken excretion periods. Gast et al. (2015) exposed hens to different *S. enterica* serovar Enteritidis strains under different living conditions and found that on average the hens excreted *S.* Enteritidis for 2.0 weeks in conventional cages and 1.5 weeks in enriched colony cages. Barrow et al. (2004) exposed multiple chicken lines to a strain of *S.* Typhimurium and a strain of *S.* Enteritidis and found that the average excretion varied between the strains and chicken lines (0.9-3.8 weeks). Alexander et al. (2009) investigated two *S.* Typhimurium outbreaks in cows and found they shed Salmonella for 1.3-2 weeks. A broad range of animal infectious period values (0.9-3.8 weeks) were included to account for the large amount of variation in these studies.

**Table B.3** Gamma values for simulated outbreak

| | $\gamma$ (year$^{-1}$ I$^{-1}$) | |
|---|---|---|
| Population | Minimum | Maximum |
| Animal | 14 | 58 |
| Human | 10 | 13 |

### B.1.4 Equal intra-transmission rates and infectious periods

Twelve outbreaks were simulated with identical infectious periods and intra-transmission rates between the host populations, but inter-transmission rates and host population sizes that varied. For these outbreaks, all the transmission rates and host population sizes were within the ranges in Table B.1 and B.2. However, the animal and human host populations shared infectious periods between 3.7-5.2 weeks (10-14 year$^{-1}$ I$^{-1}$).

### B.1.5 Simulated outbreak size and length

The simulated salmonellosis outbreaks in this study varied in size and length of time. The mean size was 5.2 x $10^5$ infected individuals (range: 4.7 x $10^4$ - 8.8 x $10^6$) and the mean length was 13.3 years (range: 9.3 - 17.1).

## B.2 Equal-time sampling

The DTA and SC models gave similar population and transmission estimates for the 23 simulated outbreaks with random (Table 3.1; Figures 3.5 and 3.6) and equal-time sampling (Table B.4; Figures B.1 and B.2). Random sampling estimated more known simulation parameters within its 95% HPD interval, but equal-time sampling had smaller mean squared errors between known simulation parameters and the mean estimates, and smaller 95% HPD intervals. As with random sampling, there was a weak positive correlation between the known simulation parameters and the proportion of each host population sampled using equal-time sampling (Figure B.3 and Table B.5). The SC and DTA models also estimated similar phylogenetic trees for simulated outbreaks that were sampled using random and equal-time sampling (Figure B.4). This suggests that neither sampling method was more suitable for these ancestral state reconstruction models.

**Table B.4** Summary statistics of the SC and DTA models' estimates compared to known parameters for 23 simulated outbreaks sampled equal-time.

| | Population | | Transmissions | |
|---|---|---|---|---|
| | SC | DTA | SC | DTA |
| Estimate accuracy | 0.304 | 0.174 | 0.304 | 0.217 |
| Mean squared error | 0.106 | 0.176 | 0.058 | 0.112 |
| 95% HPD size | 0.491 | 0.170 | 0.135 | 0.266 |
| Correlation coefficient | 0.408 | 0.654 | -0.048 | 0.798 |

Population, the proportion of time spent in a host population; Transmission, the proportion of inter-population transmissions; Estimate accuracy, the proportion of outbreak simulations that a model included the known parameter within their 95% HPD intervals; Mean squared error, the mean squared error between the known parameters and a model's mean estimates; 95% HPD size, the mean 95% HPD interval size; Correlation coefficient, the correlation coefficient between the known parameter and a model's mean estimates

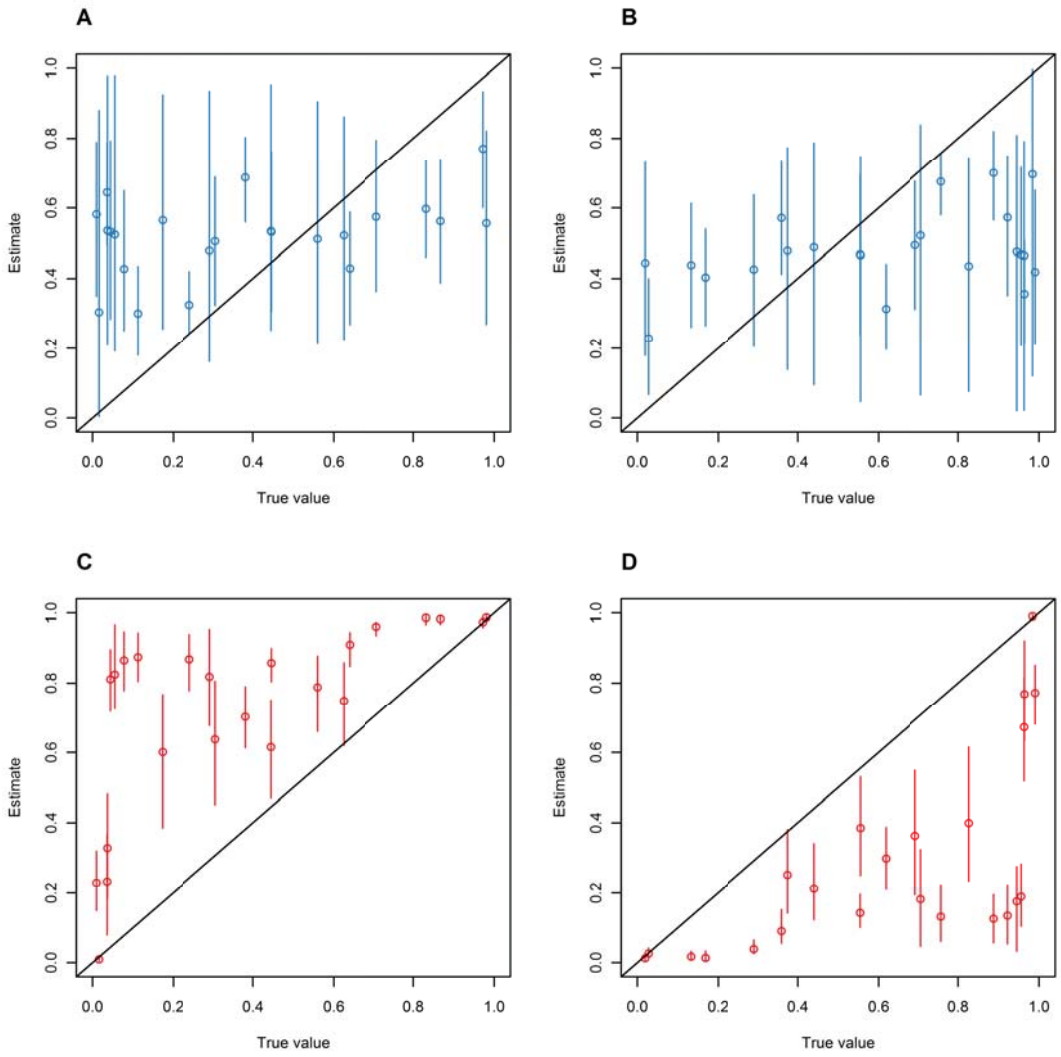


**Figure B.1.** Scatterplots of the proportion of time spent in the animal (A and C) and human (B and D) host populations, versus the values estimated by the SC (blue: A and B) and DTA (red: C and D) models for 23 simulated outbreaks that were equally sampled over time. The diagonal line represents accurate estimates of the sampled outbreaks, the dots represent the mean, and the error bars represent the 95% HPD interval.

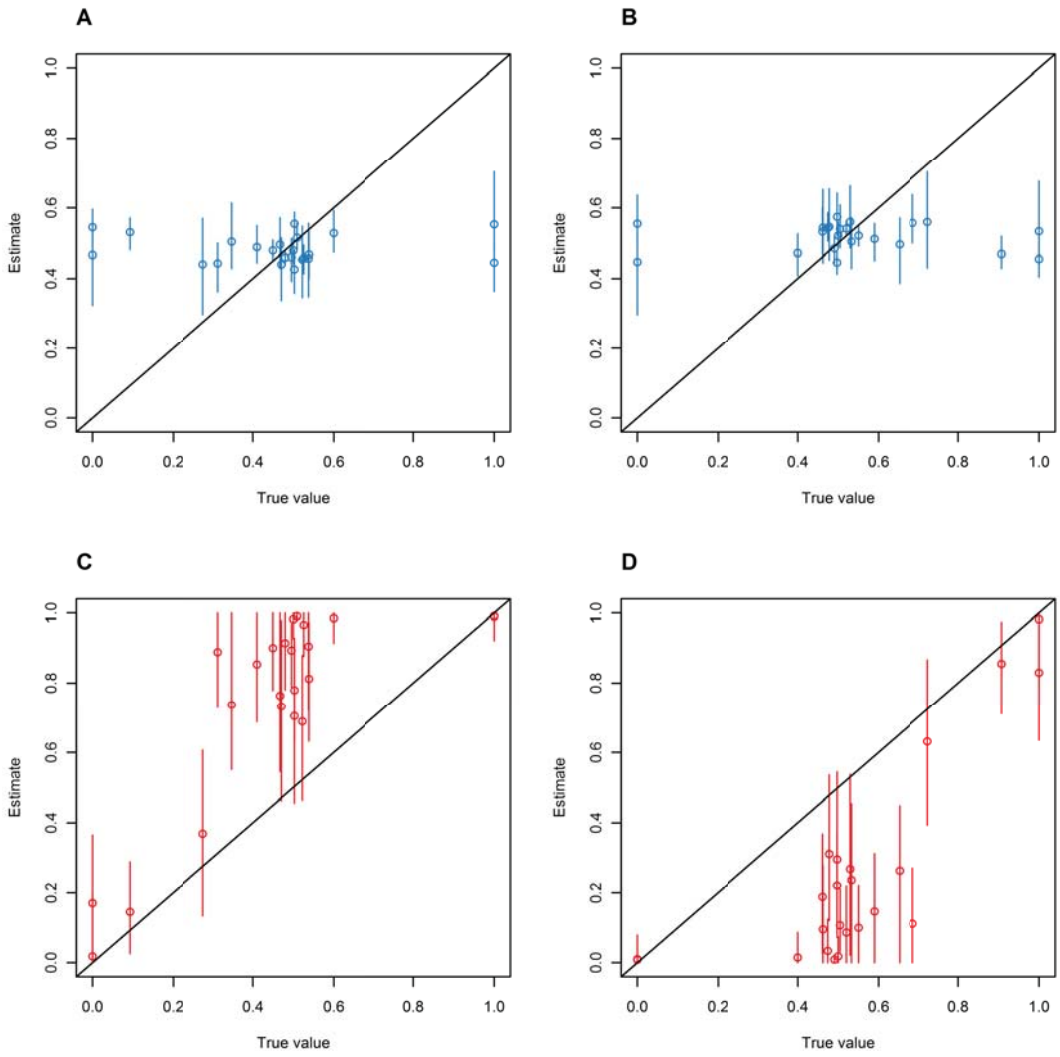


**Figure B.2.** Scatterplots of the proportion of inter-source transmissions made up of animal-to-human (A and C) and human-to-animal (B and D) transmissions, versus the values estimated by the SC (blue: A and B) and DTA (red: C and D) models for 23 simulated outbreaks that were equally sampled over time. The diagonal line represents accurate estimates of the sampled outbreaks, the dots represent the mean, and the error bars represent the 95% HPD interval.

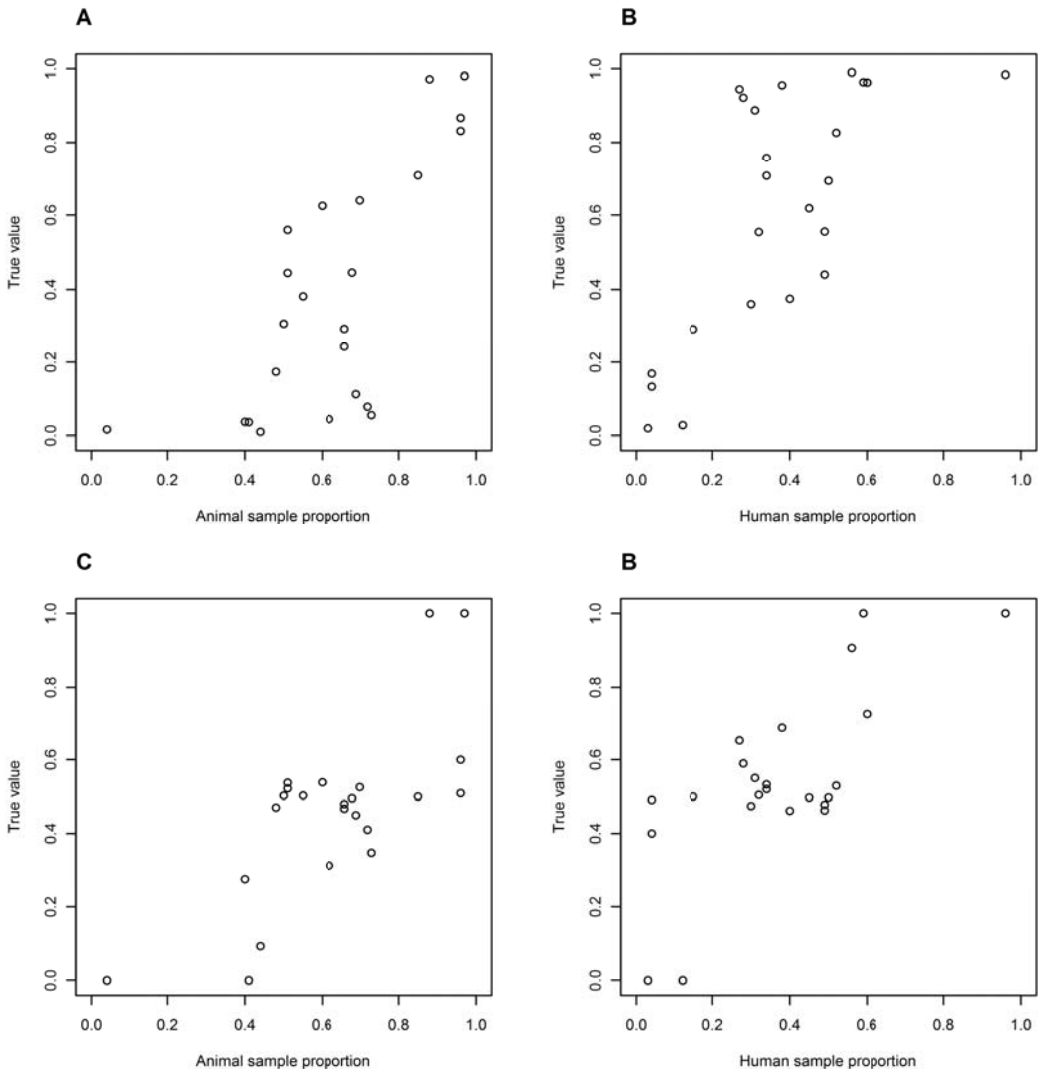


**Figure B.3.** Scatterplots of the proportion of samples made up of animal (A and C) and human (B and D) isolates versus the proportion of time spent in the animal (A) and human (B) host populations and the proportion of inter-source transmissions made up of animal-to-human (C) and human-to-animal (D) transmissions, for 23 simulated outbreaks that were equally sampled over time.


**Figure B.4.** Sampled transmission trees (A and D), maximum clade credibility trees produced by the DTA model (B and E) and maximum a posteriori trees produced by the SC model (C and F), for the same simulated outbreak that was sampled randomly (A-C) and equally over time (D-F).

**Table B.5** Correlation coefficients of the proportion of isolates sampled from each host population versus the known transmission and population parameters for 23 simulated outbreaks that were sampled equal-time.

| | Population | Transmissions |
|---|---|---|
| Correlation coefficient | 0.712 | 0.732 |

Population, the proportion of time spent in a host population; Transmission, the proportion of inter-population transmissions; Correlation coefficient, the correlation coefficient between the known parameter and the proportion of samples isolated from each host population

B.3 Model estimates versus sample proportions

The DTA model's mean estimates were more positively correlated with the proportion of samples from each host population than the SC model's mean estimates for the 23 simulated outbreaks, sampled randomly (Figure B.5 and B.6; Table B.6) and equally over time (Figures B.7 and B.8; Table B.7), and the 12 simulated outbreaks with equal intra-population transmission rates and infectious periods between host populations (Figures B.9 and B.10; Table B.8). The length of time spent in each host population showed a stronger positive correlation with the proportion of samples from each host population than the inter-population transmission rates.

**Table B.6** Correlation coefficients of the SC and DTA models' mean estimates and the proportion of isolates sampled from each host population for 23 simulated outbreaks that were randomly sampled.

| | Population | | Transmissions | |
|---|---|---|---|---|
| | SC | DTA | SC | DTA |
| Correlation coefficient | 0.559 | 0.947 | 0.358 | 0.902 |

Population, the proportion of time spent in a host population; Transmission, the proportion of inter-population transmissions; Correlation coefficient, the correlation coefficient between the proportion of samples from each host population estimated and a model's mean estimates

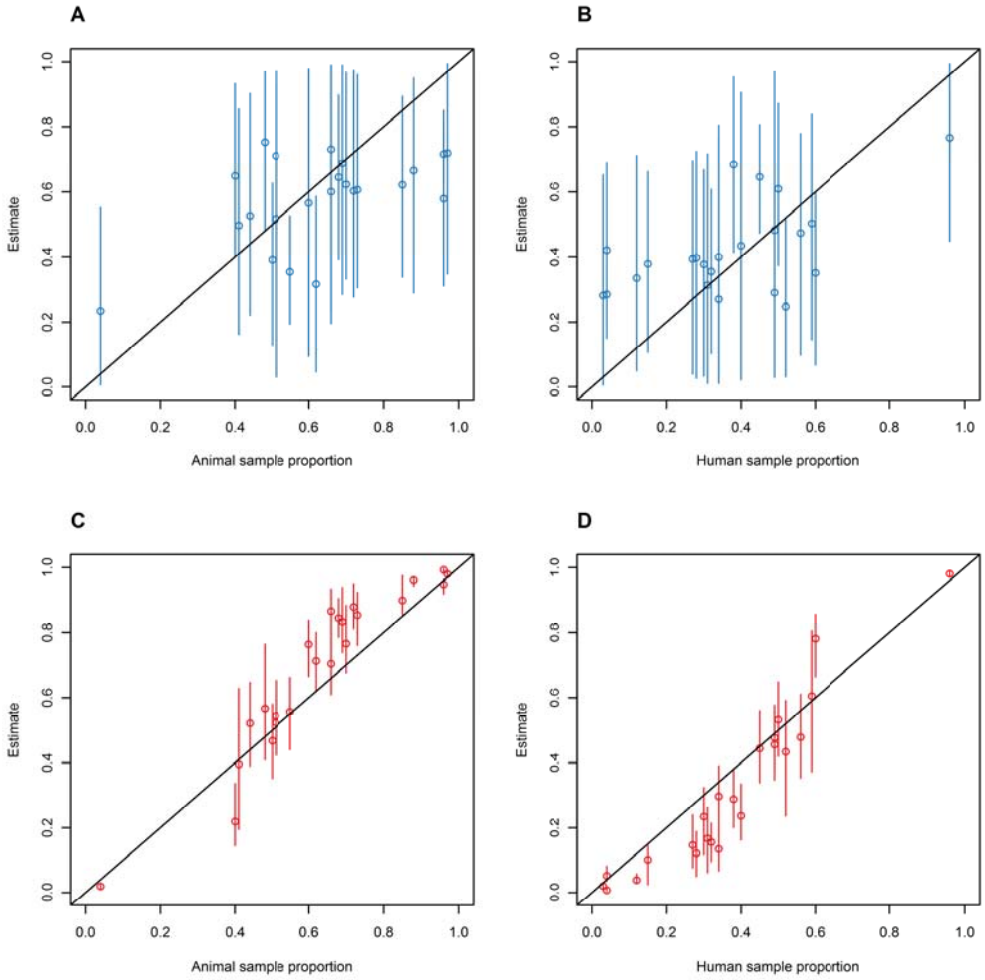


**Figure B.5.** Scatterplots of the proportion of samples made up of animal (A and C) and human (B and D) isolates, versus the SC (blue: A and B) and DTA (red: C and D) models' population estimates for 23 simulated outbreaks that were randomly sampled. The diagonal line represents estimates directly proportional to the number of each host population sampled, the dots represent the mean, and the error bars represent the 95% HPD interval.

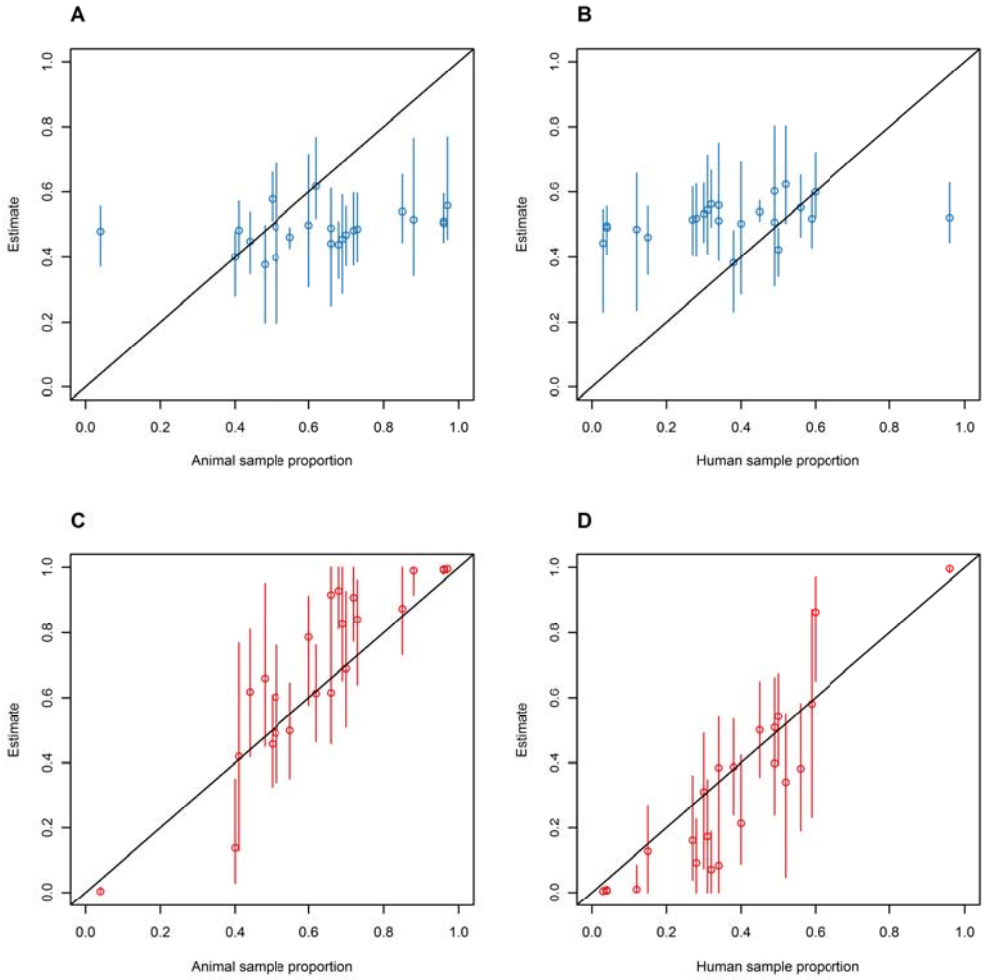


**Figure B.6.** Scatterplots of the proportion of samples made up of animal (A and C) and human (B and D) isolates, versus the the SC (blue: A and B) and DTA (red: C and D) models' transmission estimates for 23 simulated outbreaks that were randomly sampled. The diagonal line represents estimates directly proportional to the number of each host population sampled, the dots represent the mean, and the error bars represent the 95% HPD interval.

**Table B.7** Correlation coefficients of the SC and DTA models' mean estimates and the proportion of isolates sampled from each host population for 23 simulated outbreaks that were sampled equal-time.

| | Population | | Transmissions | |
|---|---|---|---|---|
| | SC | DTA | SC | DTA |
| Correlation coefficient | 0.295 | 0.903 | 0.271 | 0.830 |

Population, the proportion of time spent in a host population; Transmission, the proportion of inter-population transmissions; Correlation coefficient, the correlation coefficient between the proportion of samples from each host population estimated and a model's mean estimates

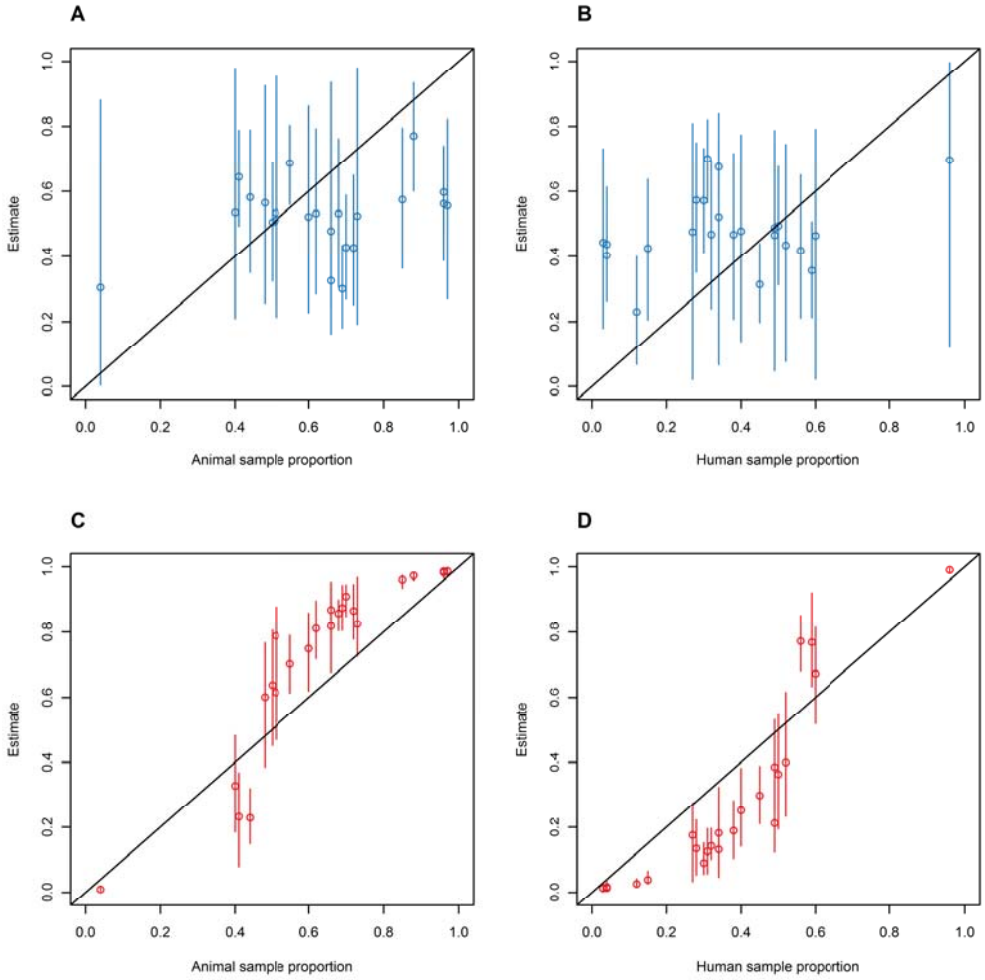


**Figure B.7.** Scatterplots of the proportion of samples made up of animal (A and C) and human (B and D) isolates, versus the SC (blue: A and B) and DTA (red: C and D) models' population estimates for 23 simulated outbreaks that were equally sampled over time. The diagonal line represents estimates directly proportional to the number of each host population sampled, the dots represent the mean, and the error bars represent the 95% HPD interval.

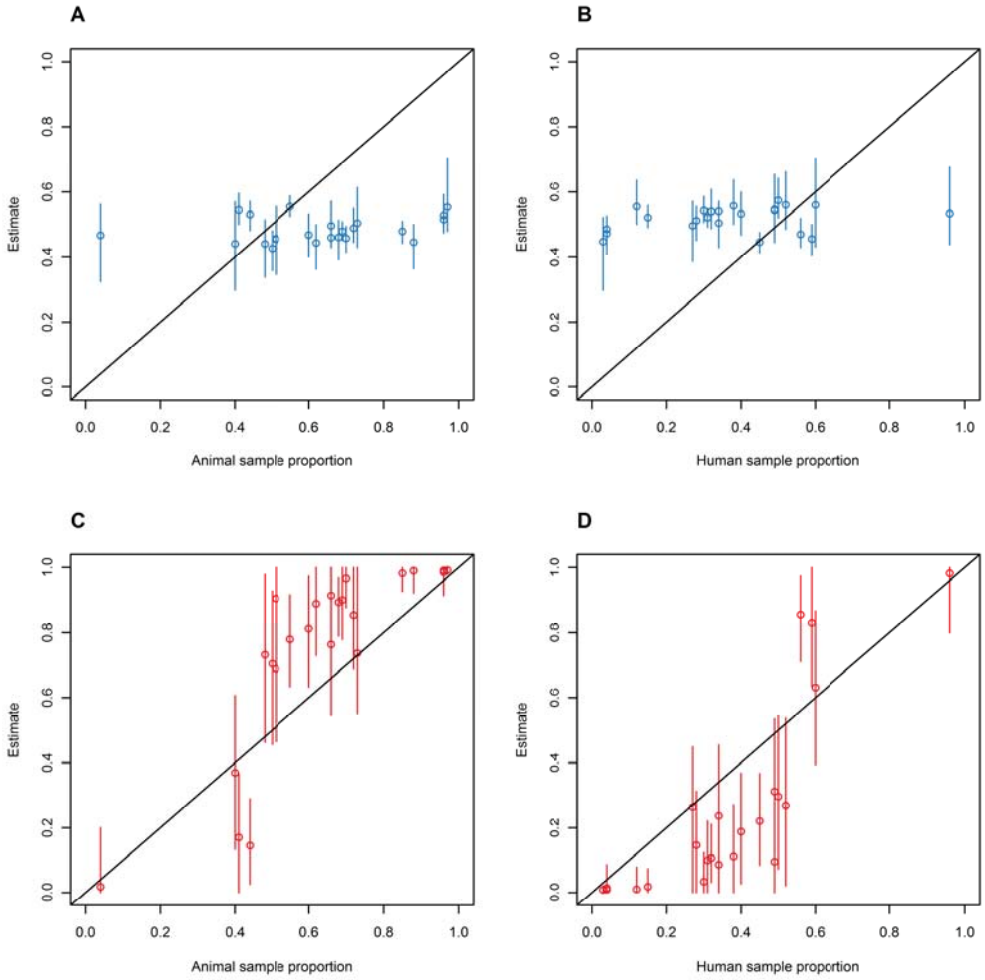


**Figure B.8.** Scatterplots of the proportion of samples made up of animal (A and C) and human (B and D) isolates, versus the SC (blue: A and B) and DTA (red: C and D) models' transmission estimates for 23 simulated outbreaks that were equally sampled over time. The diagonal line represents estimates directly proportional to the number of each host population sampled, the dots represent the mean, and the error bars represent the 95% HPD interval.

**Table B.8** Correlation coefficients of the SC and DTA models' mean estimates and the proportion of isolates sampled from each host population for 12 simulated outbreaks with equal gamma and intra-population transmission rates between host populations.

| | Population | | Transmissions | |
|---|---|---|---|---|
| | SC | DTA | SC | DTA |
| Correlation coefficient | 0.801 | 0.985 | 0.664 | 0.940 |

Population, the proportion of time spent in a host population; Transmission, the proportion of inter-population transmissions; Correlation coefficient, the correlation coefficient between the proportion of samples from each host population estimated and a model's mean estimates

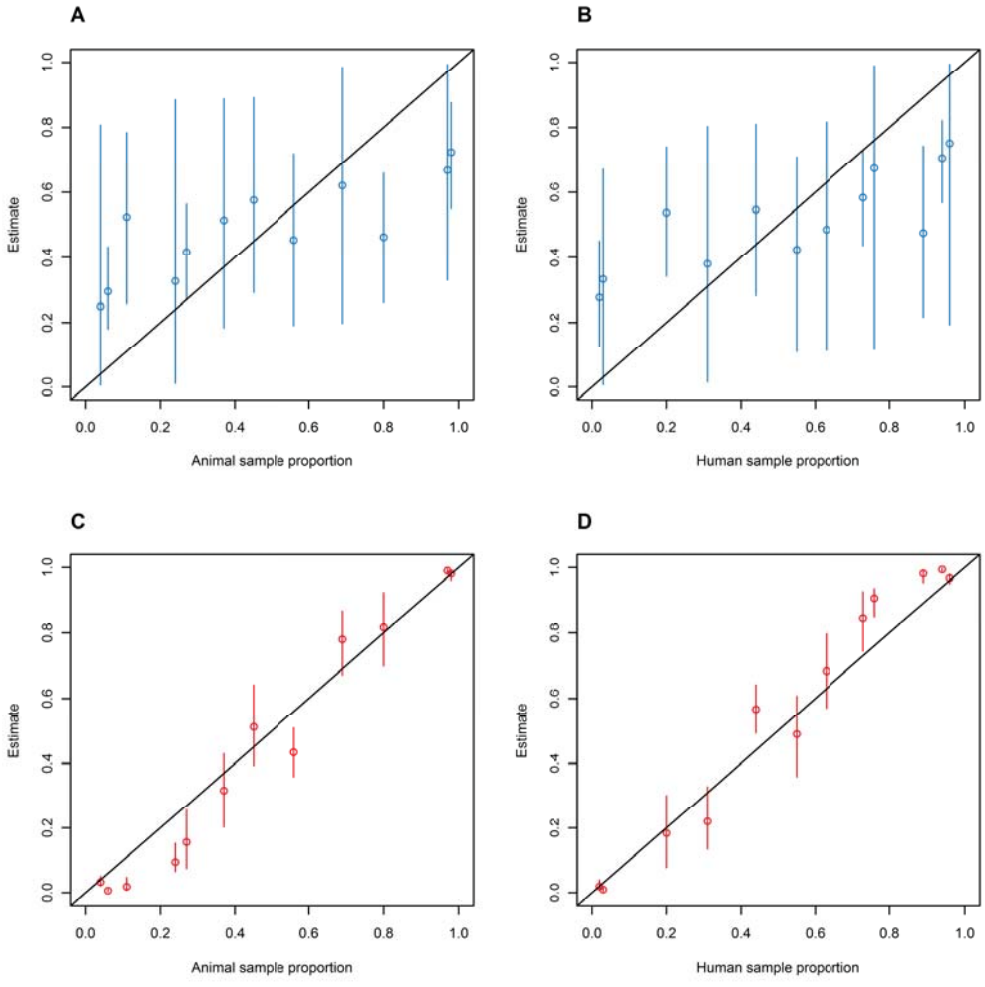


**Figure B.9.** Scatterplots of the proportion of samples made up of animal (A and C) and human (B and D) isolates, versus the SC (blue: A and B) and DTA (red: C and D) models' population estimates for 12 simulated outbreaks with equal gamma and intra-source transmission rates between host populations. The diagonal line represents estimates directly proportional to the number of each host population sampled, the dots represent the mean, and the error bars represent the 95% HPD interval.

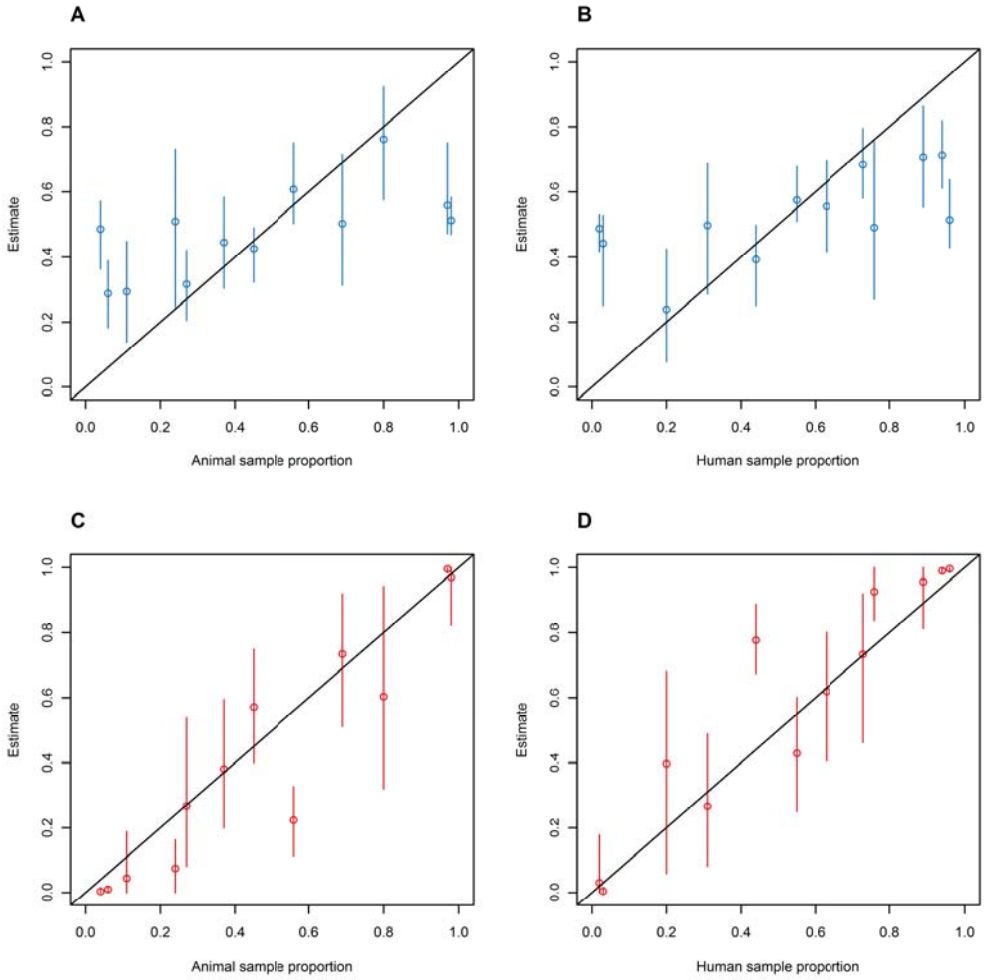


**Figure B.10.** Scatterplots of the proportion of samples made up of animal (A and C) and human (B and D) isolates, versus the SC (blue: A and B) and DTA (red: C and D) models' transmission estimates for 12 simulated outbreaks with equal gamma and intra-source transmission rates between host populations. The diagonal line represents estimates directly proportional to the number of each host population sampled, the dots represent the mean, and the error bars represent the 95% HPD interval.

# References

Alexander, K. A., Warnick, L. D., Cripps, C. J., McDonough, P. L., Grohn, Y. T., Wiedmann, M., Reed, K. E., James, K. L., Soyer, Y., and Ivanek, R. (2009). Fecal shedding of, antimicrobial resistance in, and serologic response to *Salmonella* Typhimurium in dairy calves. *Journal of the American Veterinary Medical Association*, 235(6):739–748.

Barrow, P. A., Bumstead, N., Marston, K., Lovell, M. A., and Wigley, P. (2004). Faecal shedding and intestinal colonization of *Salmonella* enterica in in-bred chickens: The effect of host-genetic background. *Epidemiology and Infection*, 132(1):117–126.

Buchwald, D. S. and Blaser, M. J. (1984). A review of human salmonellosis: II. Duration of excretion following infection with nontyphi *Salmonella*. *Reviews of Infectious Diseases*, 6:345–356.

Gast, R. K., Guraya, R., Jones, D. R., and Anderson, K. E. (2015). Persistence of fecal shedding of *Salmonella* Enteritidis by experimentally infected laying hens housed in conventional or enriched cages. *Poultry Science*, 94(7):1650–1656.

Gould, L. H., Walsh, K. A., Vieira, A. R., Herman, K., Williams, I. T., Hall, A. J., and Cole, D. (2013). Surveillance for foodborne disease outbreaks - United States, 1998-2008. *MMWR Surveillance Summaries*, 62(1):1–34.

Murase, T., Yamada, M., Muto, T., Matsushima, A., and Yamai, S. (2000). Fecal excretion of *Salmonella enterica* serovar Typhimurium following a food-borne outbreak. *Journal of Clinical Microbiology*, 38(9):3495–3497.

Statistics New Zealand (2017a). Agriculture. Retrieved 2017-03-17, from: `http://www.stats.govt.nz/infoshare/SelectVariables.aspx?pxID=5a6c6aee-c5d6-47da-afdd-1b22625c1bda`, 1–1.

Statistics New Zealand (2017b). Population estimates. Retrieved 2017-03-14, from: `http://www.stats.govt.nz/infoshare/ViewTable.aspx?pxID=8bca9269-7647-47c9-bd0c-9a153a6c6d24`, 1–1.

# Appendix C

## Supplementary material to Chapter 5

### C.1 SNP analysis

SNPs (single nucleotide polymorphisms) are single base pairs that differ between isolates. Two software programs were used to identify SNPs shared by the 117 DT160 and DT56 variant isolates: kSNP3 (Gardner et al., 2015) and Snippy (https://github.com/tseeman/snippy). Snippy aligns sequence reads from an isolate to a reference genome, in this case *Salmonella enterica* serovar Typhimurium strain 14028s (NC_016856), and then was used to compare the alignments from multiple isolates to identify single base pairs that are found in all isolates but differ in sequence (core SNPs). kSNP3 identifies kmers of a fixed length that differ in one nucleotide between isolates, and was applied to the *de novo*-assembled genomes and NC_016856 to identify core SNPs. kSNP3 identified 1,638 SNPs shared by the 117 DT160 and DT56 variant isolates, whilst Snippy identified 1,502 SNPs (Figure C.1). 1,431 SNPs were identified by both methods, leaving 207 kSNP3-unique, 71 Snippy-unique SNPs and 1,709 SNPs in total.


**Figure C.1.** Venn diagram of the number of unique and shared DT160 and DT56 variant SNPs identified by Snippy and kSNP3.

The kSNP3-unique SNPs mostly consisted of SNPs found on reads that did not align to the reference genome. However, a low number of SNPs were found in close vicinity, with the majority found in separate genes (Figure C.2), allowing kSNP to detect a large number of SNPs. The Snippy-unique SNPs mostly consisted of SNPs that were in close vicinity, unable to be picked up by kSNP as kmers of a fixed length would differ in more than one nucleotide. The reference genome used was also similar to the DT160 and 56 variant strains (Figure C.3), allowing Snippy to detect a large number of SNPs. By using both methods a larger number of SNPs were identified than if a single method alone was used.

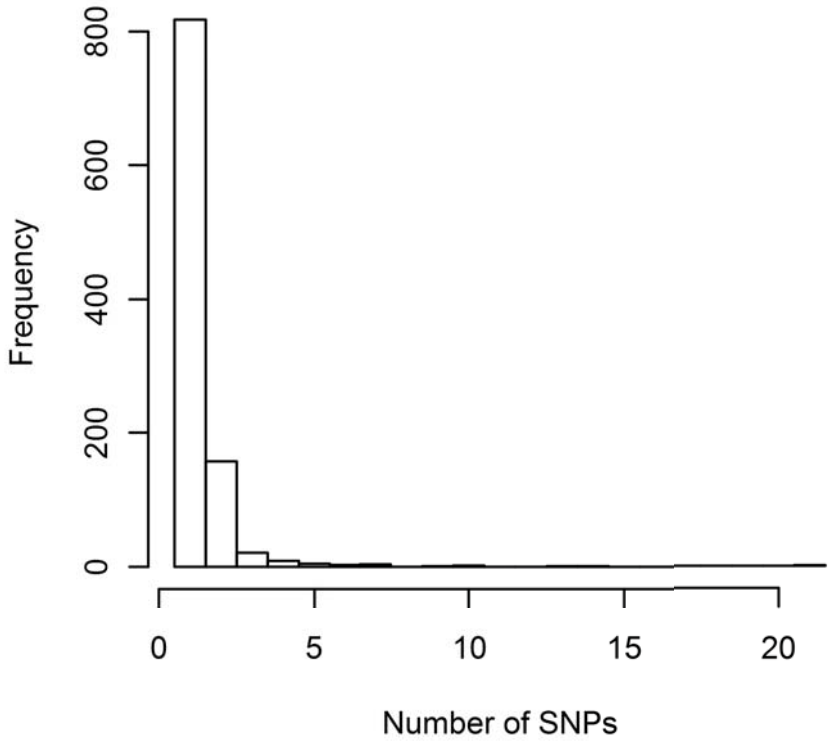


**Figure C.2.** Histogram of the number of DT160 and DT56 variant SNPs found within the same genes.

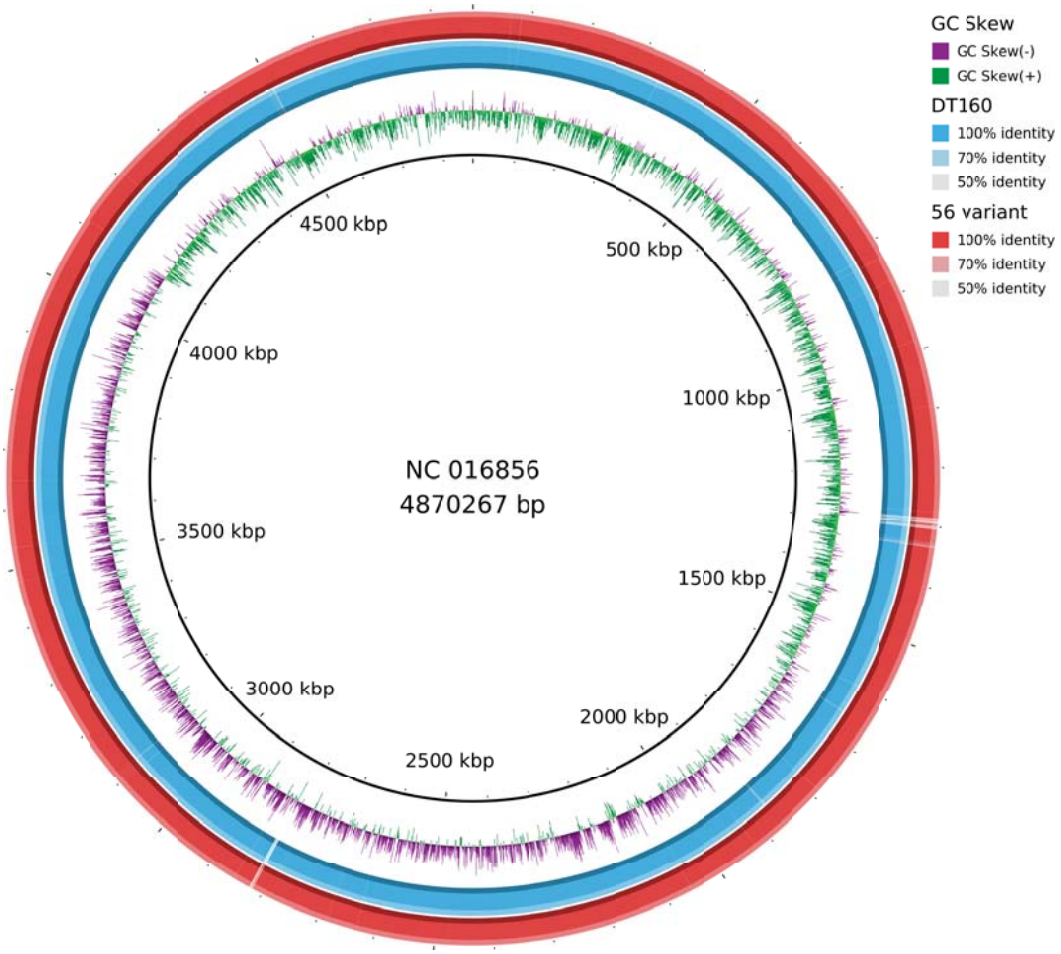


**Figure C.3.** BRIG alignment of the 109 DT160 and eight DT56 variant *de novo* assembled genomes to the reference genome, *S.* Typhimurium 14028S (NC_016856). The rings are coloured by strain.

C.2 Mobile elements

Reads from each of the 117 isolate were aligned to the mobile elements hypothesised to be associated with the DT160 and DT56 variant strains. For each alignment, one strain contained a larger read depth over the length of the sequence compared to the other strain (Figures C.4-C.6). However, reads from both strains aligned to certain sections of these sequences. In these instances, one strain contained a larger number of SNPs, suggesting that these reads were from homologous genetic elements. The exception is the pSSB1 plasmid (Figure C.7), as only one DT56 variant isolate had reads that aligned to this sequence, but these reads did not align well, as suggested by the large amount of SNPs.

The DT160 isolates contained a much higher read depth around the 10 kb position of the ST160 phage compared to the rest of the sequence (Figure C.4). This may indicate a recombination event has occurred, repeating this section of the phage and increasing the read depth. Further work on the ST160 phage with longer reads is required to investigate this.

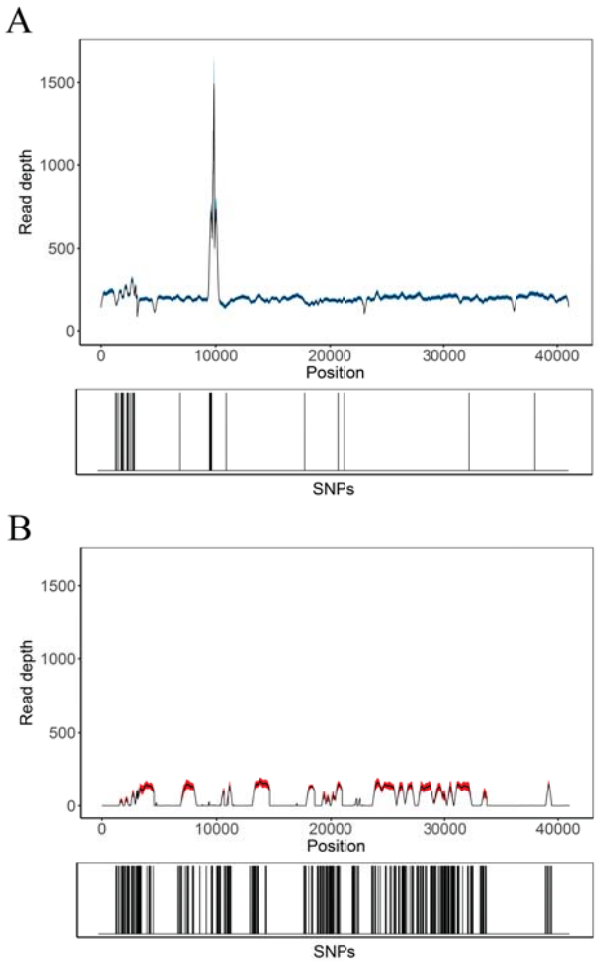


**Figure C.4.** Line graph of the mean read coverage of 109 DT160 (A) and eight DT56 variant (B) isolates across the ST160 phage (NC_014900). The 95% confidence intervals are represented by the blue and red ribbons, and the location of SNPs between the reads and the consensus sequence are demonstrated on the lower bars.

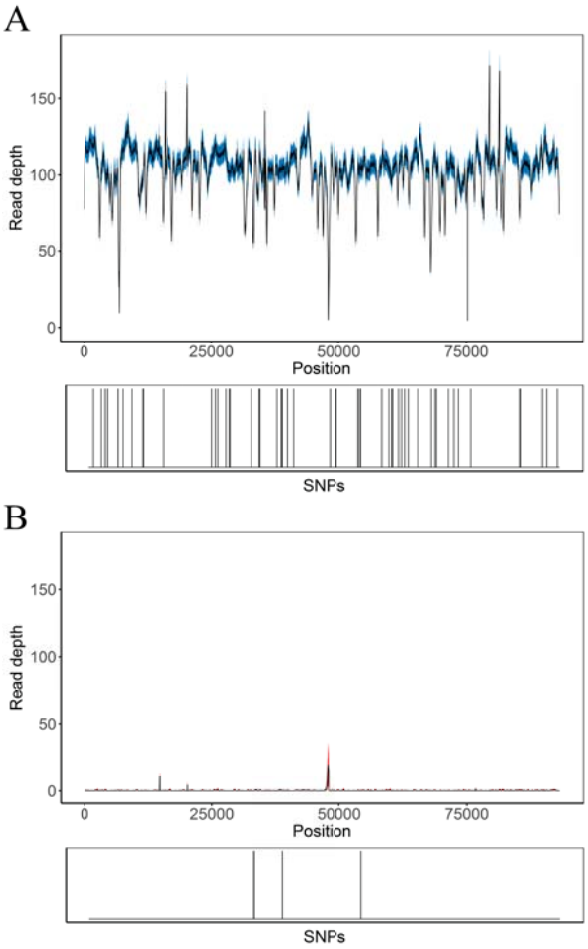


**Figure C.5.** Line graph of the mean read coverage of 109 DT160 (A) and eight DT56 variant (B) isolates across the pSTUK-100 phage (CP002615). The 95% confidence intervals are represented by the blue and red ribbons, and the location of SNPs between the reads and the consensus sequence are demonstrated on the lower bars.

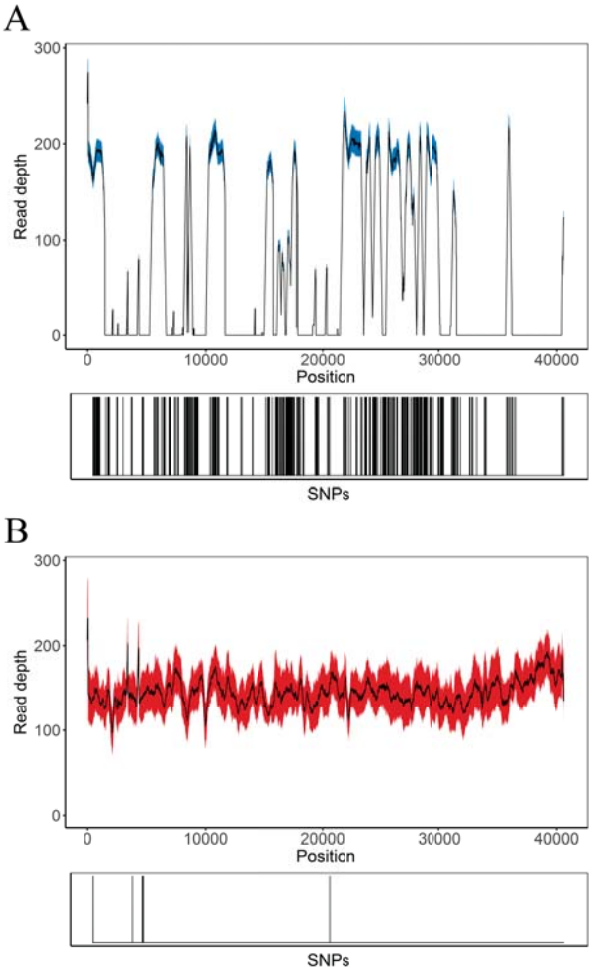


**Figure C.6.** Line graph of the mean read coverage of 109 DT160 (A) and eight DT56 variant (B) isolates across the BTP1 region of *S.* Typhimurium strain D23580 (FN424405). The 95% confidence intervals are represented by the blue and red ribbons, and the location of SNPs between the reads and the consensus sequence are demonstrated on the lower bars.

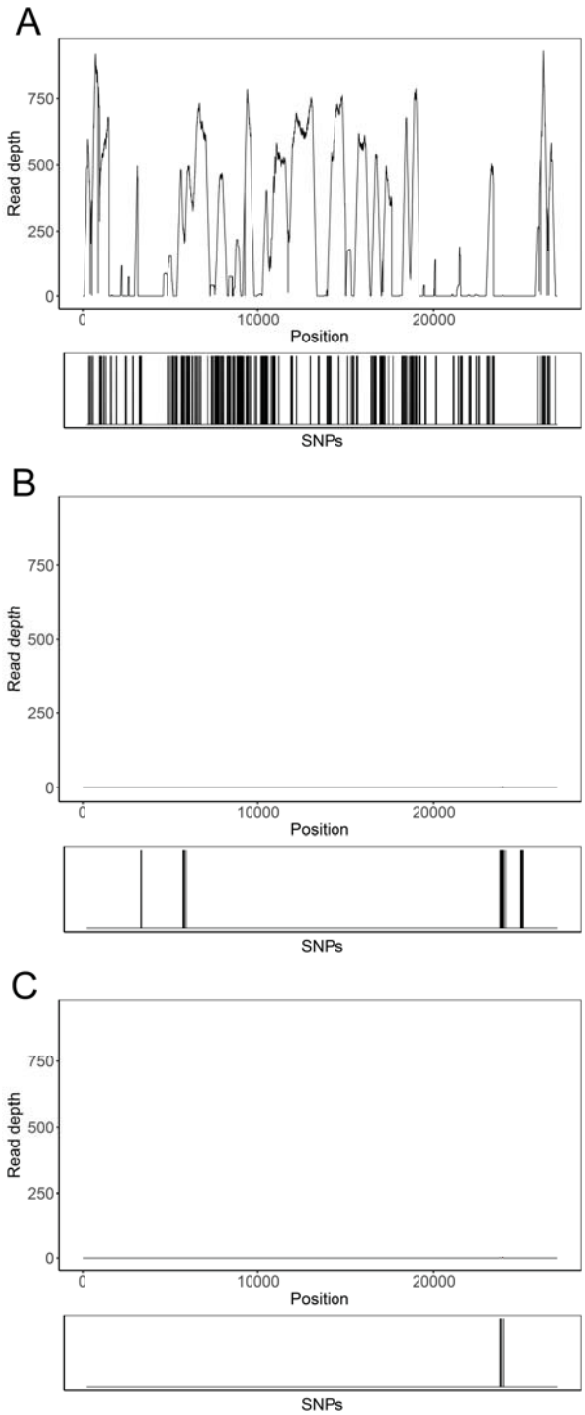


**Figure C.7.** Line graph of the coverage of the DT56 variant isolate that contained an extra plasmid (A), the mean read coverage of 109 DT160 isolates (B) and the mean read coverage of seven DT56 variant that do not contain an extra plasmid (C) across the *Salmonella enterica* serovar Typhi plasmid, pBSSB1 (AM419040). The 95% confidence intervals are represented by the blue (DT160) and red (56 variant) ribbons, and the location of SNPs between the reads and the consensus sequence are demonstrated on the lower bars.

# References

Gardner, S. N., Slezak, T., and Hall, B. G. (2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics*, 31:2877–2878.

# Appendix D

## Supplementary material to Chapter 6

### D.1 SNP analysis

SNPs are single base pairs that differ between isolates. kSNP3 identified 99 SNPs shared by the 16 *Campylobacter jejuni* ST45 isolates, whilst Snippy identified 178 SNPs (Figure D.1). 81 SNPs were identified by both methods, leaving 18 kSNP3-unique and 97 Snippy-unique SNPs. The kSNP3-unique SNPs mostly consisted of SNPs found on reads that did not align to the reference genome, whilst the Snippy-unique SNPs mostly consisted of SNPs that were in close vicinity, unable to be picked up by kSNP3 as the k-mers would differ in more than one nucleotide. A large number of SNPs were found in close vicinity, with multiple SNPs found within the same gene (Figure D.2), explaining why Snippy identified a larger proportion of SNPs than did kSNP3. In addition, the reference genome was similar to the ST45 isolates analysed (Figure D.3), which allowed a large proportion of the reads to align. In total, 196 core SNPs were identified by Snippy and kSNP3, and were used for further genetic analyses.

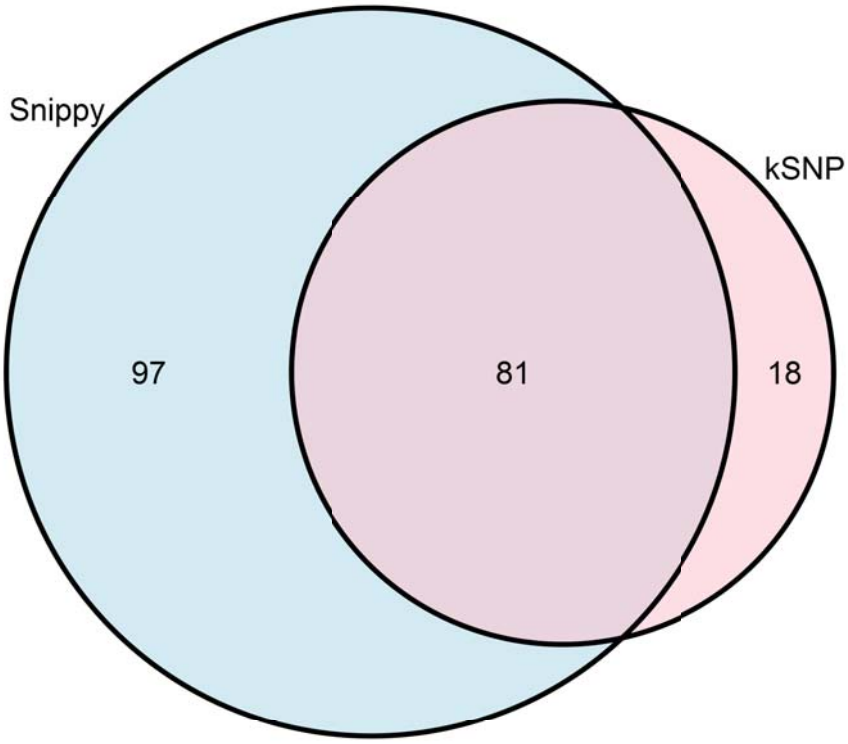


**Figure D.1.** Venn diagram of the number of unique and shared ST45 SNPs identified by Snippy (blue) and kSNP3 (pink).

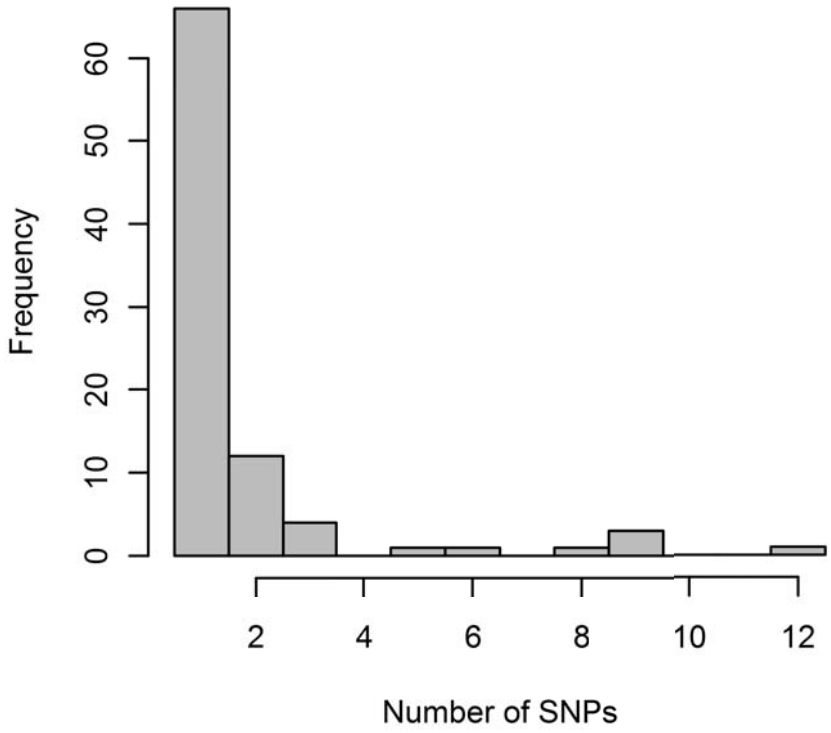


**Figure D.2.** Histogram of the number of ST45 SNPs found within the same genes.

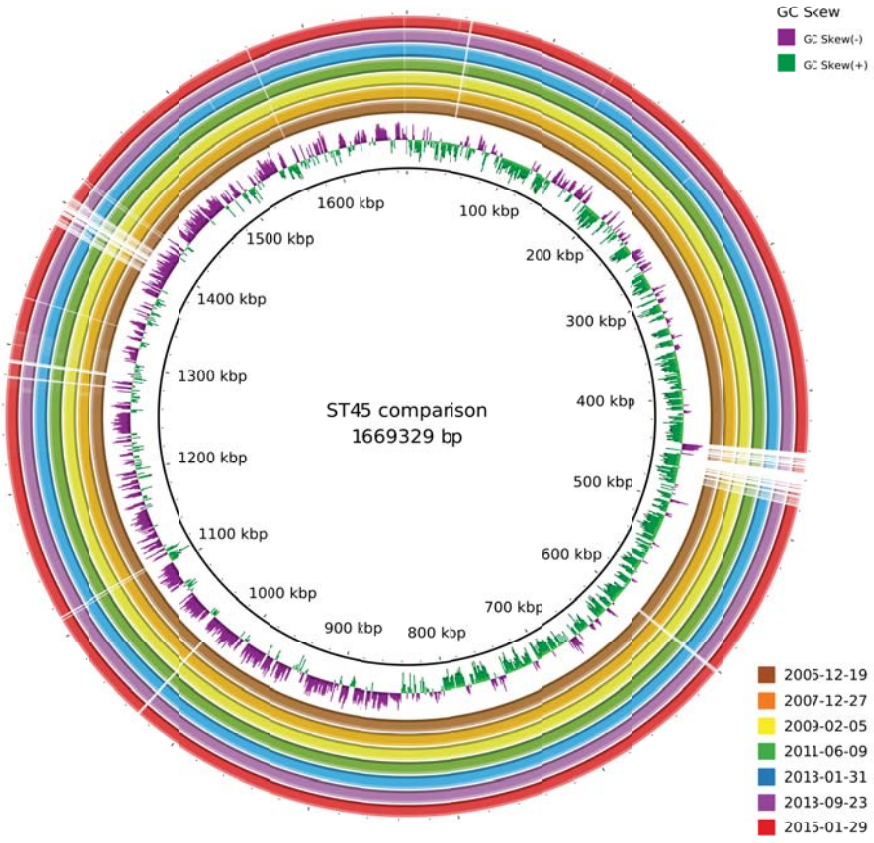


**Figure D.3.** BRIG alignment of the 16 ST45 *de novo* assembled genomes to the reference genome, *C. jejuni* ST45 (NC_022529). The rings are colored by date of collection and the 10 isolates collected from the same occurrences were combined.

## D.2 Effective population size

The effective population size is the number of individuals in a population that contribute to the next generation. The GMRF Bayesian Skyride model suggested that there was no change in the effective population size of ST45 from 2006 to 2016 (Figure D.4). There was a lot more uncertainty in the effective population size estimate in the absence of genetic data, suggesting that the model could detect a signal.

The gastrointestinal tract can only accommodate a set number of *Campylobacter*. Therefore, the uniform effective population size detected is consistent with persistent colonisation of the patients gastrointestinal tract. However, the antimicrobial susceptibility results suggest that the bacteria became less susceptible to certain antibiotics during antimicrobial therapies. These therapies would have acted as bottlenecks, killing all susceptible isolates and leaving resistant variants. This should have decreased the ST45 effective population

size. However, the GMRF Bayesian Skyride model did not pick up any of these changes. There are two potential reasons for this. 1) the resistant ST45 isolates quickly repopulated the gastrointestinal tract and accumulated point mutations, maintaining the effective population size; and 2) an insufficient number of ST45 isolates were collected to detect the change in effective population size. Further analysis is required to determine why no change in effective population size was detected.

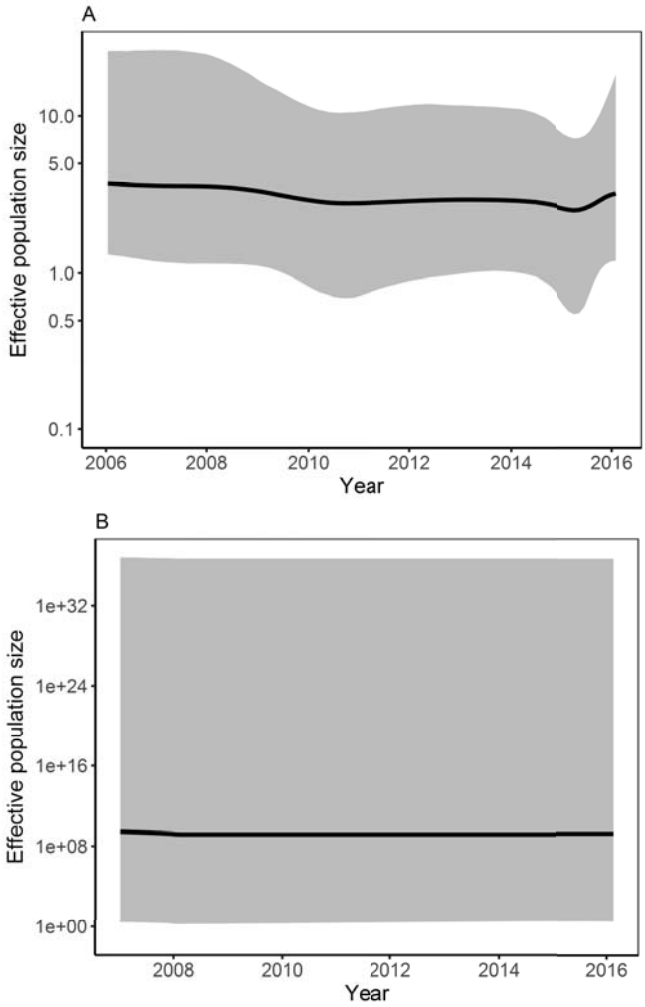


**Figure D.4.** Line graphs of the relative ST45 effective population size (log scale), as estimated by the GMRF Bayesian Skyride model with (A) and without (B) sequence data. The black line represents the mean effective population size estimate and the grey area represents the 95% HPD interval.

### D.3 Protein-coding gene analysis

Of the 90 proteins that contained sequence differences, 62 (69%) contained a single protein difference (SNP, indel or protein presence), 17 (19%) contained two protein differences, 6 (7%) contained three protein differences and 5 (5%) contained more than 3 (Figure D.5). The high number of proteins containing more than

one difference supports the hypothesis that the human host was selecting for protein differences in specific genes.

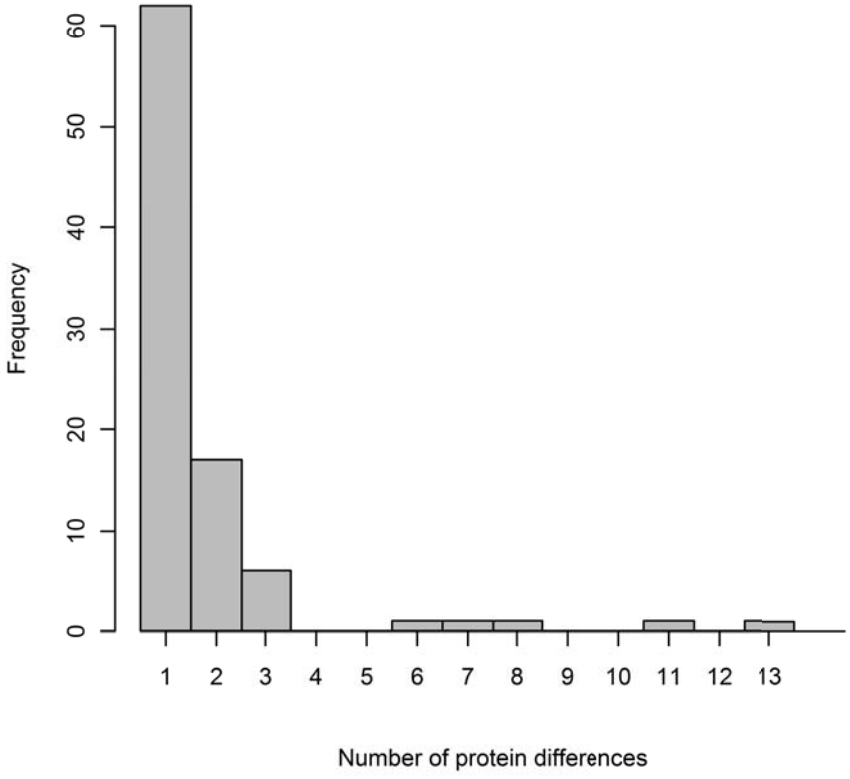


**Figure D.5.** Histogram of the number of protein differences found within the same protein sequence.

### D.4 Antimicrobial susceptibility

There was some evidence of the ST45 isolates becoming less susceptivle to gentamycin and streptomycin over time (Figure D.6). Antimicrobial susceptibility patterns for the other tested antibiotics were reasonably constant over time. The exception is the 2008 isolate, which was more susceptible to sulfamethoxazole than the rest of the isolates. Isolates collected from the same episode shared identical antimicrobial susceptibility patterns (Figure D.7).

All ST45 isolates collected from the patient were resistant to sulfamethoxazole and trimethoprim, but the mechanisms for resistance are unknown. Gibreel and Sköld (1999) found that several mutations in the *folP* gene of *C. jejuni* were associated with sulphonamide resistance. However, the *folP* genes found in the ST45 isolates were distinct from the sequence published by Gibreel and Scold, such that these mutations could not

be identified. BLAST revealed that the published *folP* sequence was primarily found in *Campylobacter coli* genomes and a few *C. jejuni* genomes. Gibreel and Sköld (1998) also found that trimethoprim resistance in *Campylobacter* was associated with the presence of the *dfr1* and *dfr9* genes. However, the sequences of these genes were not published. To date, no other studies on *Campylobacter* sulphonamide resistance mechanisms have been investigated. Further work on *Campylobacter* sulphonamide resistance mechanisms is required to determine how the ST45 isolates were resistant to sulfamethoxazole and trimethoprim.

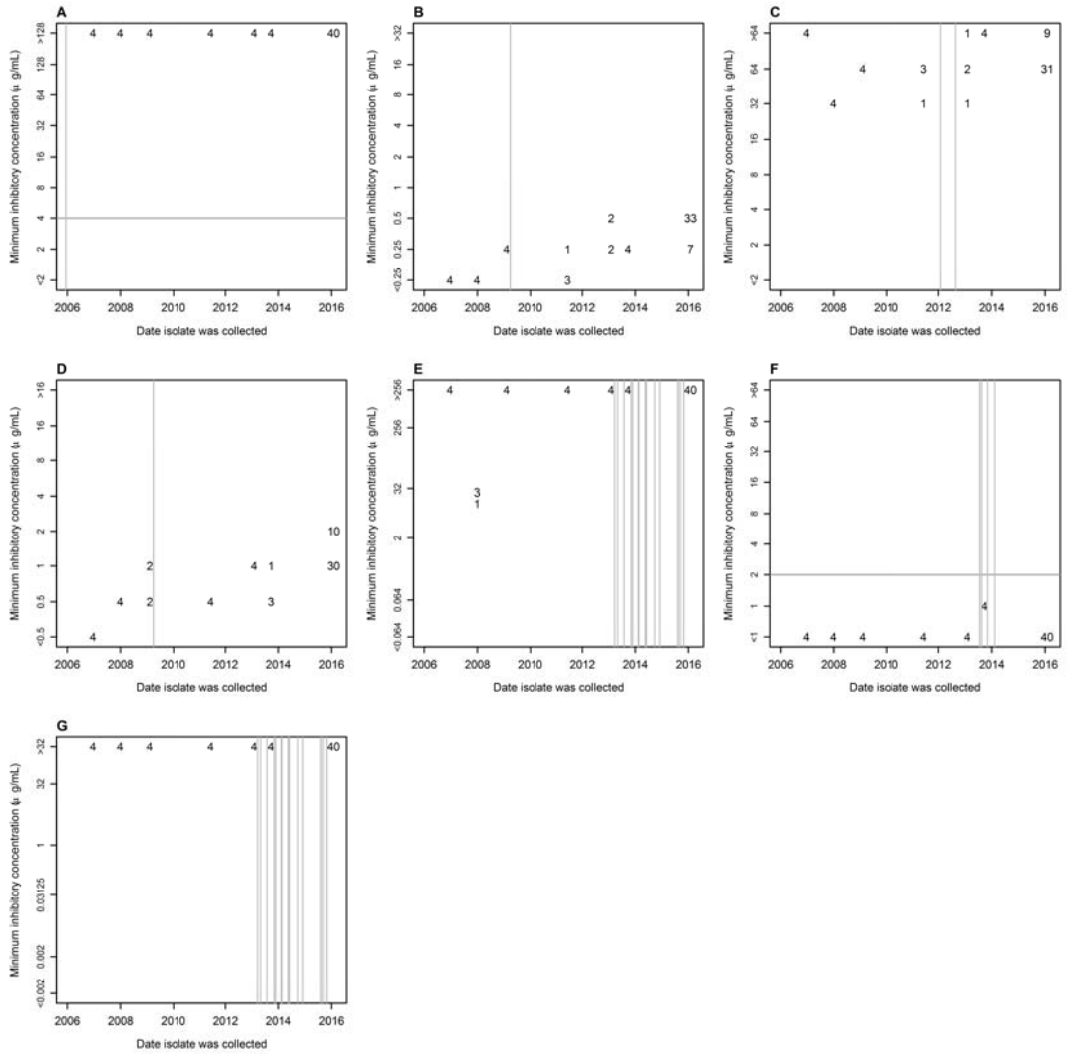


**Figure D.6.** Scatterplots of date of collection versus the minimum inhibitory concentration to erythromycin (A), gentamycin (B), nalidixic acid (C), streptomycin (D), sulfamethoxazole (E), tetracycline (F) and trimethoprim (G), for 16 ST45 isolates collected from the same patient. The points represent the number of repeats that had the same value and the vertical lines represent times in the patient's medical records when they were prescribed or mention taking antibiotics belonging to the same class. The horizontal lines represent the EUCAST breakpoints (EUCAST, 2016).

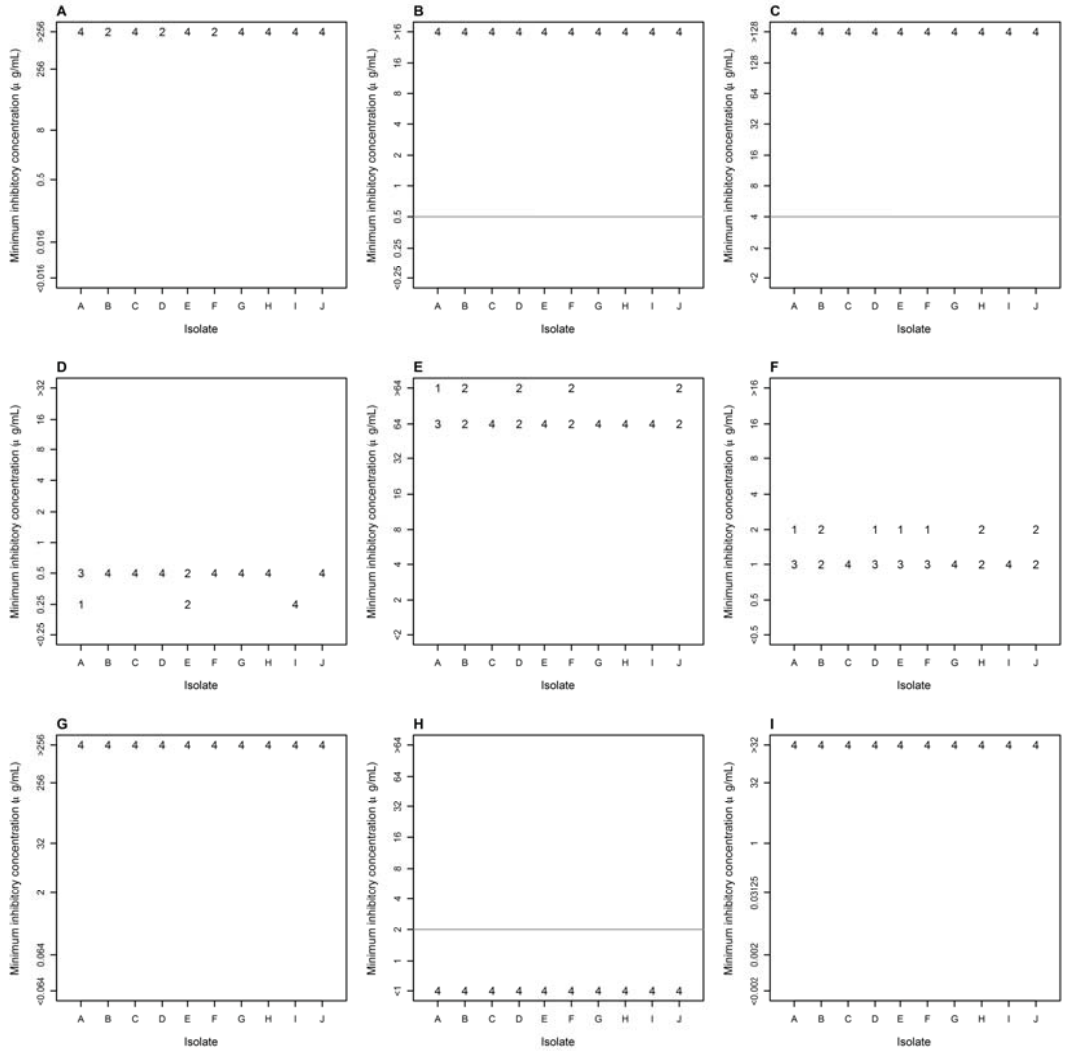


**Figure D.7.** Scatterplots of the minimum inhibitory concentration of 10 ST45 isolates (A-J) collected from the same faecal sample to ampicillin (A), ciprofloxacin (B), erythromycin (C), gentamycin (D), nalidixic acid (E), streptomycin (F), sulfamethoxazole (G), tetracycline (H) and trimethoprim (I). The points represent the number of repeats that had the same value. The horizontal lines represent the EUCAST breakpoints (EUCAST, 2016).

### D.5 Chemotaxis

The two chemotaxis methods trialled in this study showed similar results. The disc diffusion method demonstrated no significant difference in chemotactic ability among the 16 ST45 isolates tested (Figure D.8), whilst the hard agar plug demonstrated a slight decrease in chemotactic ability over time (Figures D.9). There was no difference in chemotactic ability among isolates collected from the same episode (Figures D.10-11). However, the non-motile, negative control strains displayed chemotaxis using both methods. Therefore, it is uncertain whether these methods were measuring the ST45 isolate's chemotactic abilities.

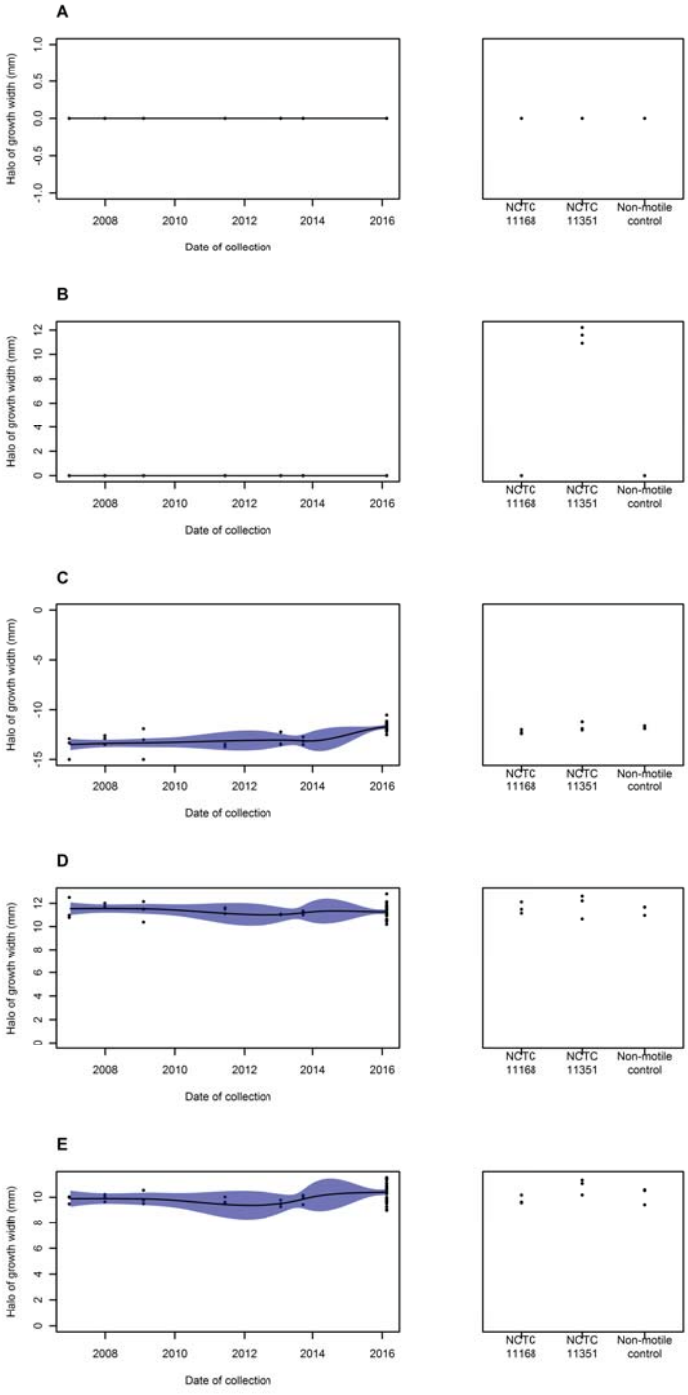


**Figure D.8.** Scatterplots of date of collection versus distance travelled in motility agar around PBS (A), citrate (B), deoxycholate (C), pyruvate (D) and serine (E) discs for 16 ST45 isolates collected from the same patient, and the results for the control strains. The purple area represent the 95% confidence interval surrounding the Loess line.

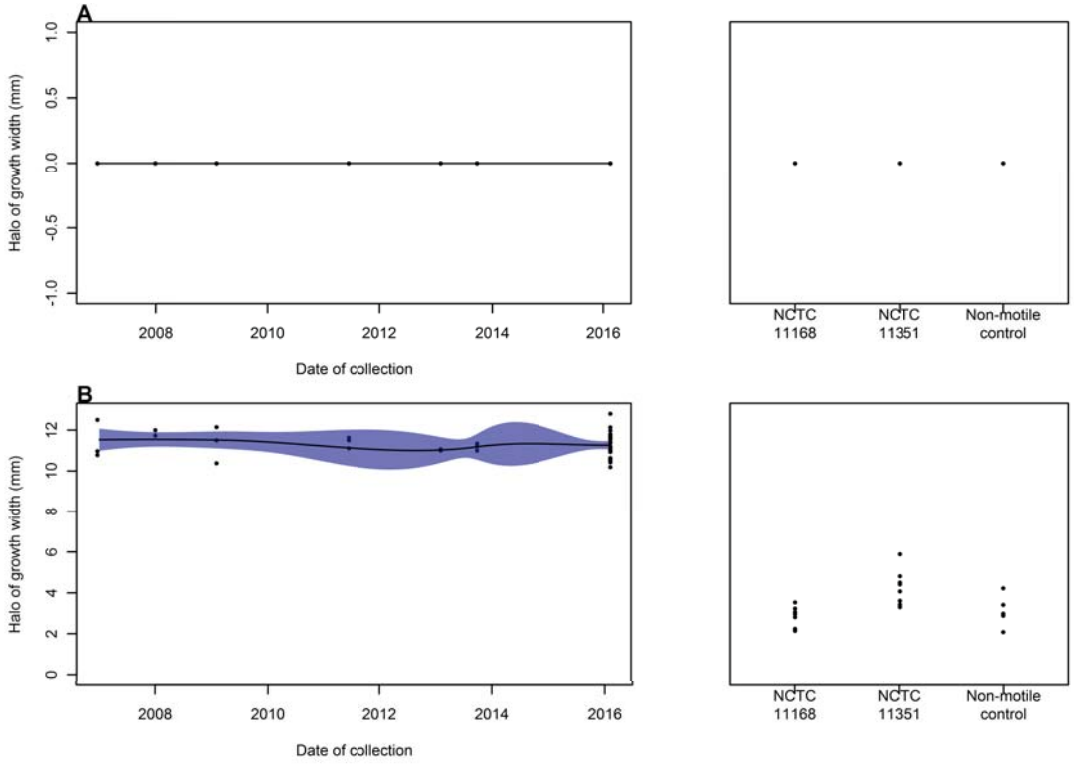


**Figure D.9.** Scatterplots of date of collection versus distance travelled in motility agar around PBS (A) and pyruvate (B) hard-agar plugs for 16 ST45 isolates collected from the same patient, and the results for the control strains. The purple area represent the 95% confidence interval surrounding the Loess line.

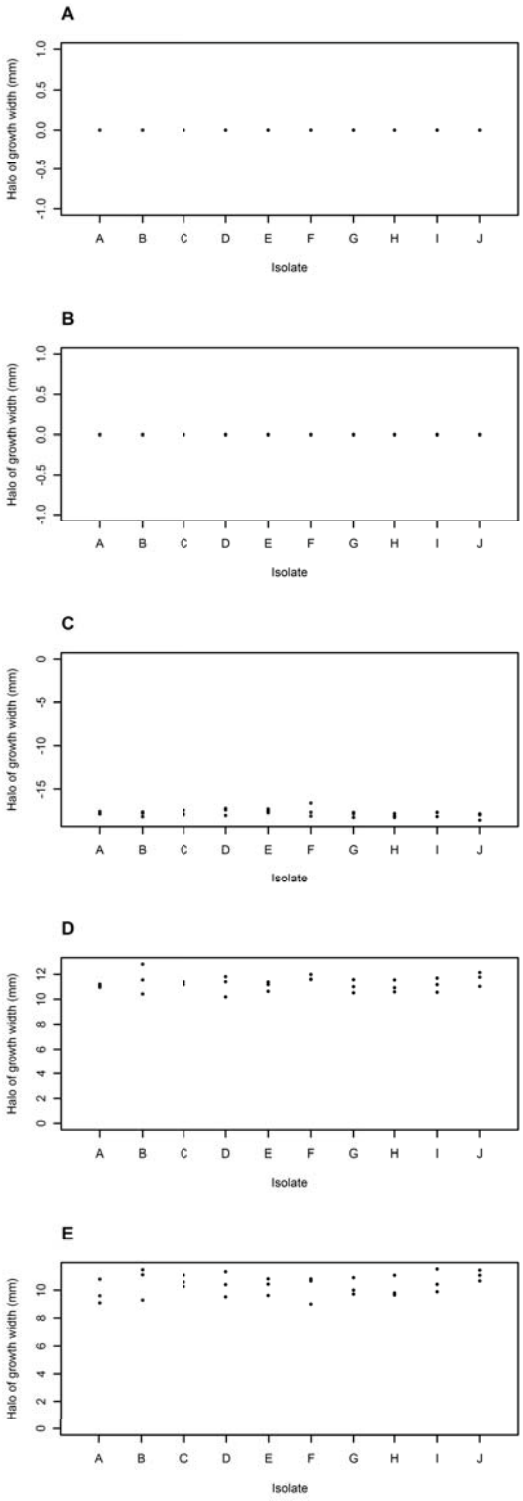


**Figure D.10.** Scatterplots of the distance travelled in motility agar around PBS (A), citrate (B), deoxycholate (C), pyruvate (D) and serine (E) discs of 10 ST45 isolates collected from the same episode.

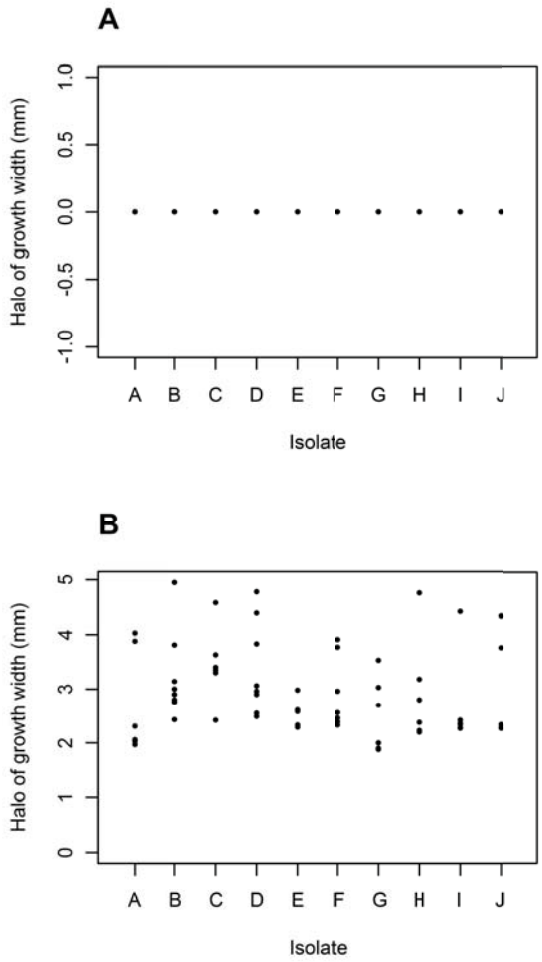


**Figure D.11.** Scatterplots of the distance travelled in motility agar around PBS (A) and pyruvate (B) hard-agar plugs of 10 ST45 isolates collected from the same episode.

# References

EUCAST (2016). *Campylobacter jejuni* and *coli* EUCAST clinical breakpoint tables. Retrieved 2017-06-29, from: `http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_6.0_Breakpoint_table.pdf`, 80–80.

Gibreel, A. and Sköld, O. (1998). High-level resistance to trimethoprim in clinical isolates of *Campylobacter jejuni* by acquisition of foreign genes (dfr1 and dfr9) expressing drug-insensitive dihydrofolate reductases. *Antimicrobial Agents and Chemotherapy*, 42(12):3059–3064.

Gibreel, A. and Sköld, O. (1999). Sulfonamide resistance in clinical isolates of *Campylobacter jejuni*: Mutational changes in the chromosomal dihydropteroate synthase. *Antimicrobial Agents and Chemotherapy*, 43(9):2156–2160.