

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

# 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 motility 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

$\mu$ l	.....	Microlitre
$^{\circ}$ C	.....	Degrees Celsius
$\times$ g	.....	Times gravity
bp	.....	Base pairs
CDC	.....	Centre for Disease Control and Prevention
<i>C. difficile</i>	.....	<i>Clostridium difficile</i>
CE	.....	Common Era
<i>C. jejuni</i>	.....	<i>Campylobacter jejuni</i>
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
<i>E. coli</i>	.....	<i>Escherichia coli</i>
ENA	.....	European Nucleotide Archive
ESR	.....	Institute of Environmental Science and Research

<b>EUCAST</b> .....	European Society of Clinical Microbiology and Infectious Diseases
<b>GC</b> .....	Guanine-cytosine
<b>GMRF</b> .....	Gaussian Markov Random Fields
<b>GTR</b> .....	Generalised time reversible
<b>GWRC</b> .....	Greater Wellington Regional Council
<b>HBRC</b> .....	Hawke's Bay Regional Council
<b>HKY</b> .....	Hasegawa, Kishino and Yano
<b>HPD</b> .....	Highest posterior density
<b>HIV</b> .....	Human immunodeficiency virus
<b>IgA</b> .....	Immunoglobulin A
<b>IgG</b> .....	Immunoglobulin G
<b>INDELS</b> .....	INsertions/DELetions
<b>IVABS</b> .....	Institute of Veterinary, Animal and Biological Sciences
<b>kb</b> .....	Kilo base pairs
<b>Mb</b> .....	Mega base pairs
<b>MCDHB</b> .....	MidCentral District Health Board
<b>MCMC</b> .....	Markov chain Monte Carlo
<b>MDS</b> .....	Multidimensional scaling
<b><i>m</i>EpiLab</b> .....	Massey University Molecular Epidemiology and Public Health Laboratory
<b>ml</b> .....	Millilitre
<b>mm</b> .....	Millimetre
<b>mM</b> .....	Millimolar
<b>MLST</b> .....	Multilocus sequence typing
<b>MSE</b> .....	Mean squared error
<b>MSS</b> .....	Mean sum of squares
<b>MSSS</b> .....	Manawatu sentinel surveillance site

<b>NCBI</b> .....	National Centre for Biotechnology Information
<b>NZGL</b> .....	New Zealand Genomics Limited
<b><i>P. aeruginosa</i></b> .....	<i>Pseudomonas aeruginosa</i>
<b>PBS</b> .....	Phosphate-buffered saline
<b>PCR</b> .....	Polymerase chain reaction
<b>PFGE</b> .....	Pulsed-field gel electrophoresis
<b>PNMRF</b> .....	Palmerston North Medical Research Fund
<b>SC</b> .....	Structured coalescent
<b><i>S. enterica</i></b> .....	<i>Salmonella enterica</i>
<b><i>S. Enteritidis</i></b> .....	<i>Salmonella enterica</i> serovar Enteritidis
<b>SIR</b> .....	Susceptible-Infected-Recovered
<b>SNP</b> .....	Single nucleotide polymorphism
<b>SS</b> .....	Sum of squares
<b>ST</b> .....	Sequence type
<b>STEC</b> .....	Shiga toxin-producing <i>Escherichia coli</i>
<b><i>S. Typhi</i></b> .....	<i>Salmonella enterica</i> serovar Typhi
<b><i>S. Typhimurium</i></b> .....	<i>Salmonella enterica</i> serovar Typhimurium
<b>UK</b> .....	United Kingdom
<b>US</b> .....	United States
<b>WHO</b> .....	World Health Organisation
<b>XML</b> .....	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. Phylogenomics 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

<b>Abstract</b> .....	i
<b>Acknowledgements</b> .....	iii
<b>Abbreviations</b> .....	iv
<b>Publications</b> .....	vii
<b>Contents</b> .....	viii
<b>List of figures</b> .....	xvi
<b>List of tables</b> .....	xxii
<b>Chapter 1: Introduction</b> .....	1
1.1 Epigraph .....	1
1.2 Introduction .....	1
1.3 Structure .....	1
1.4 Terminology .....	2
<b>Chapter 2: Literature review</b> .....	5
2.1 Introduction .....	5
2.2 Sources of bacterial enteritis .....	5
2.2.1 Pathogen reservoirs .....	6
2.2.1.1 Control measures for human reservoirs .....	6
2.2.1.2 Control measures for animal reservoirs .....	7
2.2.2 Disease pathway .....	8

2.2.2.1 Control measures for food pathways .....	8
2.2.2.2 Control measures for environmental pathways .....	9
2.2.3 Pathogen exposure .....	10
2.2.3.1 Control measures for food exposures .....	10
2.2.3.2 Control measures for environmental exposures .....	11
2.2.3.3 Control measures for direct exposures .....	11
2.2.4 Risk factors for bacterial enteritis .....	13
2.3 Outbreak analysis .....	14
2.3.1 Source strains .....	15
2.3.2 Phenotypic tests .....	15
2.3.3 Molecular tests .....	16
2.4 Bacterial evolution during outbreaks .....	17
2.4.1 Genotype evolution .....	17
2.4.2 Phenotype evolution .....	17
2.4.3 Related outbreaks .....	18
2.4.4 Date of common ancestor .....	18
2.4.5 Bacterial transmission .....	19
2.4.6 Effective population size .....	20
2.4.7 Genetic elements .....	20
2.5 Conclusion .....	20

<b>Chapter 3: Genomic analysis of <i>Salmonella enterica</i> serovar Typhimurium DT160 associated with a 14-year outbreak, New Zealand 1998-2012 .....</b>	<b>39</b>
3.1 Abstract .....	39
3.2 Introduction .....	39
3.3 Methods .....	41

3.3.1 Whole-genome sequencing .....	41
3.3.2 Genome assembly .....	42
3.3.3 Single-nucleotide polymorphism identification .....	42
3.3.4 Global DT160 strains .....	43
3.3.5 Phylogenetic inference and distances .....	43
3.3.6 Phylogenetic analysis .....	43
3.3.7 Protein-coding gene analysis .....	44
3.3.8 Antimicrobial susceptibility .....	45
3.3.9 Scripts .....	45
3.4 Results .....	45
3.4.1 Genomic DT160 comparison .....	45
3.4.2 DT160 introduction date .....	45
3.4.3 DT160 evolution .....	46
3.4.4 DT160 sources .....	49
3.4.5 Protein and gene analysis .....	51
3.4.6 Ancestral migration between hosts .....	54
3.4.7 DT160 antimicrobial susceptibilities .....	54
3.5 Discussion .....	54
3.6 Conclusion .....	56
Chapter 3 statement of contribution .....	61

<b>Chapter 4: Investigation of the validity of two ancestral state reconstruction models for estimating <i>Salmonella</i> transmission during outbreaks .....</b>	<b>62</b>
4.1 Abstract .....	62
4.2 Introduction .....	62
4.3 Methods .....	64

4.3.1	Outbreak simulations .....	64
4.3.2	Simulated outbreaks .....	64
4.3.3	Multiple variable simulations .....	65
4.3.4	DTA model .....	65
4.3.5	SC model .....	65
4.3.6	Model comparison .....	65
4.3.7	Model consistency .....	66
4.3.8	Disproportionate sampling .....	66
4.3.9	Equal-time sampling .....	66
4.3.10	Equal intra-population transmission and infectious periods .....	67
4.3.11	DT160 outbreak .....	67
4.3.12	Scripts .....	67
4.4	Results .....	67
4.4.1	Model consistency .....	67
4.4.2	Disproportionate sampling .....	70
4.4.3	Multiple variable simulations .....	73
4.4.4	Equal-time sampling .....	79
4.4.5	Equal intra-population transmission and infectious periods .....	79
4.4.6	DT160 outbreak .....	85
4.5	Discussion .....	86
4.6	Conclusion .....	88
<b>Chapter 5: Genomic comparison of two <i>Salmonella enterica</i> serovar Typhimurium</b>		
<b>strains responsible for consecutive salmonellosis outbreaks in New Zealand .....</b>		
5.1	Abstract .....	91
5.2	Introduction .....	91

5.3 Methods .....	94
5.3.1 Epidemiological data .....	94
5.3.2 Whole genome sequencing .....	95
5.3.3 Genome assembly .....	95
5.3.4 Single nucleotide polymorphism identification .....	95
5.3.5 Phylogenetic inference .....	96
5.3.6 Phylogenetic analysis .....	96
5.3.7 <i>Salmonella</i> Typhimurium comparison .....	96
5.3.8 Protein-coding gene analysis .....	97
5.3.9 Novel plasmid analysis .....	97
5.3.10 Scripts .....	97
5.4 Results .....	97
5.4.1 Epidemiology .....	97
5.4.2 Genome size and GC content .....	98
5.4.3 Strain evolution and comparison .....	98
5.4.4 Protein differences .....	99
5.4.5 Novel plasmid .....	100
5.5 Discussion .....	101
5.6 Conclusion .....	103
<b>Chapter 6: Long-term colonisation by <i>Campylobacter jejuni</i> within a human host: evolution, antimicrobial resistance and adaptation .....</b>	<b>109</b>
6.1 Abstract .....	109
6.2 Introduction .....	109
6.3 Methods .....	110
6.3.1 Ethics .....	110

6.3.2 Interview .....	110
6.3.3 Environmental sampling .....	110
6.3.4 PCR .....	111
6.3.5 Strains .....	111
6.3.6 Whole genome sequencing .....	111
6.3.7 Genomic assembly .....	112
6.3.8 Single nucleotide polymorphisms .....	112
6.3.9 ST45 comparison .....	112
6.3.10 NeighbourNet trees .....	113
6.3.11 Phylogenetic analysis .....	113
6.3.12 Protein-coding gene analysis .....	113
6.3.13 Genome degradation .....	114
6.3.14 Antimicrobial susceptibility .....	114
6.3.15 Motility .....	114
6.3.16 Chemotaxis .....	115
6.3.17 Scripts .....	115
6.4 Results .....	115
6.4.1 Patient background .....	115
6.4.2 <i>Campylobacter jejuni</i> isolates .....	116
6.4.3 Genomic ST45 comparison .....	116
6.4.4 Genetic distance .....	116
6.4.5 Phylogenetic analysis .....	117
6.4.6 ST45 comparison .....	117
6.4.7 Protein-coding gene analysis .....	118
6.4.8 Genome degradation .....	119

6.4.9 Antimicrobial susceptibility .....	120
6.4.10 Motility .....	122
6.4.11 Chemotaxis .....	123
6.5 Discussion .....	123
6.6 Conclusion .....	125
<b>Chapter 7: General discussion .....</b>	<b>131</b>
7.1 Introduction .....	131
7.2 Whole genome sequencing .....	131
7.3 Transmission .....	132
7.4 Evolution .....	133
7.5 Future work .....	135
7.6 Conclusion .....	137
<b>Appendix A: Supplementary material to Chapter 3 .....</b>	<b>141</b>
A.1 SNP comparison .....	141
A.2 Protein-coding gene analysis .....	144
A.3 Discrete trait analysis .....	149
A.4 Antimicrobial susceptibility testing .....	154
<b>Appendix B: Supplementary material to Chapter 4 .....</b>	<b>178</b>
B.1 Simulated outbreaks .....	178
B.1.1 Population size .....	178
B.1.2 Transmission rates .....	179
B.1.3 Infectious periods .....	179
B.1.4 Equal intra-transmission rates and infectious periods .....	180
B.1.5 Simulated outbreak size and length .....	180
B.2 Equal-time sampling .....	180

B.3 Model estimates versus sample proportions .....	186
<b>Appendix C: Supplementary material to Chapter 5 .....</b>	<b>194</b>
C.1 SNP analysis .....	194
C.2 Mobile elements .....	197
<b>Appendix D: Supplementary material to Chapter 6 .....</b>	<b>203</b>
D.1 SNP analysis .....	203
D.2 Effective population size .....	206
D.3 Protein-coding gene analysis .....	207
D.4 Antimicrobial susceptibility .....	208
D.5 Chemotaxis .....	210

## List of figures

<b>3.1</b>	Bar graph of the number of DT160 isolates reported from non-human sources, 1998-2012.	40
<b>3.2</b>	Line graphs of the number of bovine, human, poultry and wild bird cases reported each year, 1998-2012.	41
<b>3.3</b>	NeighborNet tree of 109 New Zealand and two United Kingdom DT160 isolates.	46
<b>3.4</b>	Line graph of the relative DT160 effective population size as estimated by the GMRF Bayesian Skyride model	47
<b>3.5</b>	A. NeighborNet tree of 109 DT160 isolates collected during an outbreak in New Zealand, 1998-2012. B. Scatterplot of the mean pairwise distance of 106 DT160 isolates from 2000-2011.	48
<b>3.6</b>	Maximum-likelihood tree of 109 DT160 isolates, coloured by source, and heatmap of Euclidean distances based on protein differences.	50
<b>3.7</b>	Multi-dimensional scaling of 107 DT160 isolates, coloured by date of collection and source.	52
<b>3.8</b>	Scatterplots of year of collection versus z-values for 25 poultry, 25 wild bird, 24 bovine and 33 human isolates.	53
<b>3.9</b>	Bar graph of the proportion of proteins shared by 107 DT160 isolates that differ in sequence for each COG functional group.	54
<b>4.1</b>	Scatterplots of the proportion of time spent in the animal and human populations as estimated by the SC and DTA models, for 10 random samples of the same simulated outbreak.	68
<b>4.2</b>	Scatterplots of the proportion of inter-population transmissions made up of animal-to-human and human-to-animal transmissions as estimated by the SC and DTA models, for 10 random samples of the same simulated outbreak.	69
<b>4.3</b>	Scatterplots of the proportion of time spent in the animal and human populations as estimated by the SC and DTA models versus the proportion of sampled isolates that are animal and human for the same outbreak.	71
<b>4.4</b>	Scatterplots of the proportion of inter-population transmissions made up of animal-to-human and human-to-animal transmissions as estimated by the SC and DTA models versus the proportion of isolates that are animal and human for the same outbreak.	72
<b>4.5</b>	Scatterplots of the proportion of time spent in the animal and human populations, versus the values estimated by the SC and DTA models for 23 simulated outbreaks that were randomly sampled.	74
<b>4.6</b>	Scatterplots of the proportion of inter-population transmissions made up of animal-to-human and human-to-animal transmissions, versus the values estimated by the SC and DTA models for 23 simulated outbreaks that were randomly sampled.	75

<b>4.7</b>	Scatterplots of the proportion of samples made up of animal and human isolates versus the proportion of time spent in the animal, and human populations and the proportion of inter-population transmissions made up of animal-to-human and human-to-animal transmissions, for 23 simulated outbreaks that were randomly sampled.	76
<b>4.8</b>	Sampled transmission tree, maximum clade credibility tree produced by the DTA model and maximum a posteriori tree produced by the SC model, for one of the 23 simulated outbreak that was randomly sampled.	78
<b>4.9</b>	Scatterplots of the proportion of time spent in the animal and human populations, versus the values estimated by the SC and DTA models for 12 simulated outbreaks with equal gamma and intra-population transmission rates between populations.	80
<b>4.10</b>	Scatterplots of the proportion of inter-population transmissions made up of animal-to-human and human-to-animal transmissions, versus the values estimated by the SC and DTA models for 12 simulated outbreaks with equal gamma and intra-population transmission rates between populations.	81
<b>4.11</b>	Sampled transmission tree, maximum clade credibility tree produced by the DTA model and maximum a posteriori tree produced by the SC model, for a simulated outbreak with equal gamma and intra-population transmission rates between populations.	82
<b>4.12</b>	Scatterplots of the proportion of samples made up of animal and human isolates versus the proportion of time spent in the animal and human populations, and the proportion of inter-population transmissions made up of animal-to-human and human-to-animal transmissions, for 12 simulated outbreaks with equal gamma and intra-population transmission rates between populations.	84
<b>4.13</b>	Estimates of the proportion of time spend in the animal and human populations, and the proportion of inter-population transmissions made up of animal-to-human and human-to-animal transmissions for the DT160 outbreak, as estimated by the SC and DTA models.	85
<b>4.14</b>	Maximum clade credibility tree produced by the DTA model and maximum a posteriori tree produced by the SC model, for the DT160 outbreak.	86
<b>5.1</b>	Line graph of the number of human DT160 and DT56 variant cases reported each year in New Zealand from 1998-2016.	93
<b>5.2</b>	Bar graph of the number of non-human DT160 and DT56 variant isolates reported in New Zealand from 1998-2016.	94
<b>5.3</b>	NeighbourNet tree of 109 DT160 isolates, eight DT56 variant isolates and 41 additional <i>S. Typhimurium</i> isolates in PATRIC.	99
<b>5.4</b>	Maximum likelihood tree of 109 DT160 and eight DT56 variant isolates, and coloured by mobile elements.	100
<b>5.5</b>	Bar graph of the predicted functions of the 38 CDS found on the novel DT56 variant plasmid.	101
<b>6.1</b>	NeighborNet tree of 16 ST45 isolates collected from the same patient and coloured by date of collection.	117
<b>6.2</b>	NeighborNet tree of 16 ST45 isolates collected from the same patient and four ST45 isolates collected from non-human sources.	118
<b>6.3</b>	Bar graph of the protein difference to quantity ratio for the 159 protein differences shared by 16 ST45 isolates, for each COG functional group, and the total number of proteins belonging to each COG functional group.	119

<b>6.4</b>	Barplot of the number of pseudogenes identified in the reference ST45 strain, ST45 isolates collected as part of the MSSS, and 16 ST45 isolate collected from the same patient.	120
<b>6.5</b>	Scatterplot of date of collection versus the minimum inhibitory concentration to ampicillin and ciprofloxacin, for 16 ST45 isolates collected from the same patient.	121
<b>6.6</b>	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.	122
<b>6.7</b>	Scatterplot of the distance travelled through motility agar for 10 ST45 isolates collected from the same episode.	123
<b>A.1</b>	Venn diagram of the number of unique and shared SNPs identified by Snippy and kSNP3 for the 109 New Zealand DT160 isolates.	142
<b>A.2</b>	Maximum likelihood tree of 109 DT160 isolates and presence-absence matrix of 773 core SNPs order on NC_016856.	143
<b>A.3</b>	Histogram of the number of protein differences found within the same protein sequence for 109 DT160 isolates.	144
<b>A.4</b>	Multi-dimensional scaling of 109 and 107 (minus two outliers) DT160 isolates.	145
<b>A.5</b>	Diagnostic plots of the regression model fitted to the z-values for 107 DT160 isolates.	147
<b>A.6</b>	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 and source.	148
<b>A.7</b>	Bar graph of the number of protein difference shared by 107 DT160 isolates for each COG functional groups.	149
<b>A.8</b>	Scatterplots of the number of animal and human Markov rewards estimated using the discrete trait analysis model for the real and ten randomly assigned datasets.	151
<b>A.9</b>	Scatterplots of the number of animal-to-human and human-to-animal Markov jumps estimated using the discrete trait analysis model for the real and ten randomly assigned datasets.	151
<b>A.10</b>	Scatterplots of the number of animal and human Markov rewards estimated using the discrete trait analysis model versus the proportion of samples assigned as human.	152
<b>A.11</b>	Scatterplots of the number of animal-to-human and human-to-animal Markov jumps estimated using the discrete trait analysis model versus the proportion of samples assigned as human.	153
<b>A.12</b>	Maximum clade credibility trees of 109 DT160 isolates placed through the discrete trait analysis model, with different proportions of isolates assigned as human and animal.	154
<b>A.13</b>	Scatterplots of date of collection versus amikacin antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates.	155
<b>A.14</b>	Diagnostic plots of the regression model fitted to the amikacin antimicrobial susceptibility results for 90 DT160 isolates.	156
<b>A.15</b>	Scatterplots of date of collection versus amoxicillin/clavulanate antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates.	157
<b>A.16</b>	Diagnostic plots of the regression model fitted to the amoxicillin/clavulanate antimicrobial susceptibility results for 90 DT160 isolates.	158
<b>A.17</b>	Scatterplots of date of collection versus ampicillin antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates.	159

<b>A.18</b>	Diagnostic plots of the regression model fitted to the ampicillin antimicrobial susceptibility results for 90 DT160 isolates.	160
<b>A.19</b>	Scatterplots of date of collection versus ceftiofur antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates.	161
<b>A.20</b>	Diagnostic plots of the regression model fitted to the ceftiofur antimicrobial susceptibility results for 90 DT160 isolates.	162
<b>A.21</b>	Scatterplots of date of collection versus cefepime antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates.	163
<b>A.22</b>	Diagnostic plots of the regression model fitted to the cefepime antimicrobial susceptibility results for 90 DT160 isolates.	164
<b>A.23</b>	Scatterplots of date of collection versus chloramphenicol antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates.	165
<b>A.24</b>	Diagnostic plots of the regression model fitted to the chloramphenicol antimicrobial susceptibility results for 90 DT160 isolates.	166
<b>A.25</b>	Scatterplots of date of collection versus ciprofloxacin antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates.	167
<b>A.26</b>	Diagnostic plots of the regression model fitted to the ciprofloxacin antimicrobial susceptibility results for 90 DT160 isolates.	168
<b>A.27</b>	Scatterplots of date of collection versus nalidixic acid antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates.	169
<b>A.28</b>	Diagnostic plots of the regression model fitted to the nalidixic acid antimicrobial susceptibility results for 90 DT160 isolates.	170
<b>A.29</b>	Scatterplots of date of collection versus oxytetracycline antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates.	171
<b>A.30</b>	Diagnostic plots of the regression model fitted to the oxytetracycline antimicrobial susceptibility results for 90 DT160 isolates.	172
<b>A.31</b>	Scatterplots of date of collection versus trimethoprim/sulfamethoxazole antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates.	173
<b>A.32</b>	Diagnostic plots of the regression model fitted to the trimethoprim/sulfamethoxazole antimicrobial susceptibility results for 90 DT160 isolates.	174
<b>A.33</b>	Scatterplots of date of collection versus tetracycline antimicrobial resistance for 30 human, 30 poultry and 30 wild bird isolates.	175
<b>A.34</b>	Diagnostic plots of the regression model fitted to the tetracycline antimicrobial susceptibility results for 90 DT160 isolates.	176
<b>B.1</b>	Scatterplots of the proportion of time spent in the animal and human populations, versus the values estimated by the SC and DTA models for 23 simulated outbreaks that were equally sampled over time.	182
<b>B.2</b>	Scatterplots of the proportion of inter-source transmissions made up of animal-to-human and human-to-animal transmissions, versus the values estimated by the SC and DTA models for 23 simulated outbreaks that were equally sampled over time.	183

<b>B.3</b>	Scatterplots of the proportion of samples made up of animal and human isolates versus the proportion of time spent in the animal and human populations, and the proportion of inter-source transmissions made up of animal-to-human and human-to-animal transmissions, for 23 simulated outbreaks that were equally sampled over time.	184
<b>B.4</b>	Sampled transmission trees, maximum clade credibility trees produced by the DTA model and maximum a posteriori trees produced by the SC model, for the same simulated outbreak that was sampled randomly and equally over time.	185
<b>B.5</b>	Scatterplots of the proportion of samples made up of animal and human isolates, versus the SC and DTA models' population estimates for 23 simulated outbreaks that were randomly sampled.	187
<b>B.6</b>	Scatterplots of the proportion of samples made up of animal and human isolates, versus the SC and DTA models' transmission estimates for 23 simulated outbreaks that were randomly sampled.	188
<b>B.7</b>	Scatterplots of the proportion of samples made up of animal and human isolates, versus the SC and DTA models' population estimates for 23 simulated outbreaks that were equally sampled over time.	189
<b>B.8</b>	Scatterplots of the proportion of samples made up of animal and human isolates, versus the SC and DTA models' transmission estimates for 23 simulated outbreaks that were equally sampled over time.	190
<b>B.9</b>	Scatterplots of the proportion of samples made up of animal and human isolates, versus the SC and DTA models' population estimates for 12 simulated outbreaks with equal gamma and intra-source transmission rates between populations.	191
<b>B.10</b>	Scatterplots of the proportion of samples made up of animal and human isolates, versus the SC and DTA models' transmission estimates for 12 simulated outbreaks with equal gamma and intra-source transmission rates between populations.	192
<b>C.1</b>	Venn diagram of the number of unique and shared DT160 and DT56 variant SNPs identified by Snippy and kSNP3.	195
<b>C.2</b>	Histogram of the number of DT160 and DT56 variant SNPs found within the same genes.	196
<b>C.3</b>	BRIG alignment of the 109 DT160 and eight DT56 variant <i>de novo</i> assembled genomes to the reference genome, <i>S. Typhimurium</i> 14028S (NC_016856).	197
<b>C.4</b>	Line graph of the mean read coverage of 109 DT160 and eight DT56 variant isolates across the ST160 phage (NC_014900).	198
<b>C.5</b>	Line graph of the mean read coverage of 109 DT160 and eight DT56 variant isolates across the pSTUK-100 phage (CP002615).	199
<b>C.6</b>	Line graph of the mean read coverage of 109 DT160 and eight DT56 variant isolates across the BTP1 region of <i>S. Typhimurium</i> strain D23580 (FN424405).	200
<b>C.7</b>	Line graph of the coverage of the DT56 variant isolate that contained an extra plasmid, the mean read coverage of 109 DT160 isolates and the mean read coverage of seven DT56 variant that do not contain an extra plasmid across the <i>Salmonella enterica</i> serovar Typhi plasmid, pBSSB1 (AM419040).	201
<b>D.1</b>	Venn diagram of the number of unique and shared ST45 SNPs identified by Snippy and kSNP3.	204
<b>D.2</b>	Histogram of the number of ST45 SNPs found within the same genes.	205
<b>D.3</b>	BRIG alignment of the 16 ST45 <i>de novo</i> assembled genomes to the reference genome, <i>C. jejuni</i> ST45 (NC_022529).	206

<b>D.4</b>	Line graphs of the relative ST45 effective population size, as estimated by the GMRF Bayesian Skyride model with and without sequence data.	207
<b>D.5</b>	Histogram of the number of protein differences found within the same protein sequence for the 16 ST45 isolates.	208
<b>D.6</b>	Scatterplots of date of collection versus the minimum inhibitory concentration to erythromycin, gentamycin, nalidixic acid, streptomycin, sulfamethoxazole, tetracycline and trimethoprim, for 16 ST45 isolates collected from the same patient.	209
<b>D.7</b>	Scatterplots of date of collection versus the minimum inhibitory concentration to ampicillin, ciprofloxacin, erythromycin, gentamycin, nalidixic acid, streptomycin, sulfamethoxazole, tetracycline and trimethoprim, for 10 ST45 isolates collected from the same faecal sample.	210
<b>D.8</b>	Scatterplots of date of collection versus distance travelled in motility agar around PBS, citrate, deoxycholate, pyruvate and serine discs for 16 ST45 isolates collected from the same patient, and the results for the control strains.	211
<b>D.9</b>	Scatterplots of date of collection versus distance travelled in motility agar around PBS and pyruvate hard-agar plugs for 16 ST45 isolates collected from the same patient, and the results for the control strains.	212
<b>D.10</b>	Scatterplots of the distance travelled in motility agar around PBS, citrate, deoxycholate, pyruvate and serine discs of 10 ST45 isolates collected from the same faecal sample.	213
<b>D.11</b>	Scatterplots of the distance travelled in motility agar around PBS and pyruvate hard-agar plugs of 10 ST45 isolates collected from the same faecal sample.	214

## List of tables

<b>4.1</b>	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.	70
<b>4.2</b>	Summary statistics of the SC and DTA models' estimates compared to known parameters for 23 randomly-sampled simulated outbreaks.	73
<b>4.3</b>	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.	76
<b>4.4</b>	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 populations.	79
<b>4.5</b>	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 intrapopulation transmission rates between host populations.	83
<b>A.1</b>	PERMANOVA output for 107 DT160 isolates, based on the presence of 684 protein differences and grouped by year of collection and source.	146
<b>B.1</b>	Initial population sizes for simulated outbreaks.	178
<b>B.2</b>	Beta values for simulated outbreaks.	179
<b>B.3</b>	Gamma values for simulated outbreaks.	180
<b>B.4</b>	Summary statistics of the SC and DTA models' estimates compared to the known parameters for 23 simulated outbreaks sampled equal-time.	181
<b>B.5</b>	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.	185
<b>B.6</b>	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.	186
<b>B.7</b>	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.	189
<b>B.8</b>	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.	191