

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.

Campylobacter jejuni microevolution and phenotype:genotype relationships

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy in Veterinary Science
at Massey University, Palmerston North,
New Zealand,
by

Anja Friedrich

2014



Massey University
mEpiLab | Infectious Disease Research Centre (IDReC)
Institute of Veterinary, Animal & Biomedical Science
Palmerston North, New Zealand

Abstract

Campylobacter spp. are a major cause of human gastroenteritis. Their wide host range, environmental distribution and high genetic diversity contribute to the complex molecular epidemiology of campylobacteriosis. The aim of this multidisciplinary thesis is to investigate the phenotype:genotype relationships of *C. jejuni* and how they influence the micro-evolution of these bacteria in New Zealand.

The first study used a time series of genotyped human campylobacteriosis cases from a region of New Zealand to investigate if the clonal complexes (CCs) identified in human cases showed a seasonal pattern. The analysis revealed a prevalent clonal complex (CC-45) which showed a consistent summer peak.

The second study applied phylogenetic and population genetic tools to describe the population structure and host associated genotypes within the *C. jejuni* population in wild and agricultural animals. The findings showed that the *C. jejuni* isolates from non-agricultural animals exhibit a higher number of mosaic alleles and fewer shared sequence types (STs) between the host groups, whereas the *C. jejuni* in agricultural animals show a higher number of shared STs and fewer occurrences of admixture.

The third study tested the ability of a variety of *C. jejuni* isolates to utilise 95 substrates as carbon sources and tested their tolerance to different osmotic conditions using phenotypic microarray (PM) technology. These phenotypic expressions were correlated with their genomes and a genome wide association study was used to identify genes associated with the observed phenotype.

The last study made use of data from a dual isolate chicken challenge. The study showed the out-competition of one challenge strain and genetic variations of 15 core single nucleotide polymorphisms (SNPs), 14 of which were non-synonymous point mutations. These SNPs were confined to nine genes all of which were associated with cell shape, chemotaxis or motility of the bacteria.

This thesis has furthered our understanding of the seasonality of human campylobacteriosis in New Zealand, the existing population structure of *C. jejuni*, its biochemical requirements and tolerance to osmolytes and novel insights into short-term evolutionary dynamics *in vivo*. Based on these findings and the recommendations for future directions, this could lead to a greater understanding of host-association and new intervention strategies.

Acknowledgements

Firstly, I would like to thank my supervisors, Professor Nigel P. French, Dr Patrick J. Biggs, Dr Anne C. Midwinter, Professor Mike Hendy, and Professor Paul B. Rainey for their help, encouragement and all their guidance throughout the PhD project.

I am also grateful to the staff and students of the mEpiLab team for their technical support and helpful discussions, especially to Dr Jonathan Marshall for his help with the statistical analysis.

I would like to thank the Allan Wilson Centre for Molecular Ecology and Evolution for funding this project. I would like to show my gratitude to the Institute of Veterinary, Animal and Biomedical Sciences for funding several conference travels and research projects.

I am also grateful to Steve Richie for his help with the software Primer, Jukka Corander for helpful comments concerning the software BAPS, Professor V. Ravindran and colleagues for the animal work. I would also like to thank Mauro Tuglio for his expertise in biostructural analysis.

I would like to thank my parents for their love, support and encouragement.

Contents

1	Introduction	2
1.1	General background	2
1.2	The structure and aims of this thesis	5
2	Literature review	7
2.1	Introduction	7
2.2	History	8
2.3	Taxonomy	9
2.4	Morphological and biochemical characteristics	9
2.5	Mechanisms of colonisation, virulence and pathogenesis	11
2.5.1	Motility	12
2.5.2	Chemotaxis	12
2.5.3	Oxidative stress defence	13
2.5.4	Adhesion	14
2.5.5	Invasion	14
2.5.6	Toxin production	15
2.5.7	Iron acquisition	15
2.6	Characterisation of bacterial strains	16
2.6.1	Antibiotic sensitivity test	16
2.6.2	Phenotyping	16
2.6.3	Sequence based typing techniques	18
2.6.4	Whole genome sequencing technologies	22
2.6.5	Gene sequencing approaches	31
2.6.6	Phenotype MicroArray (PM) system	35
2.7	Epidemiology	38
2.7.1	Sources of <i>Campylobacter</i>	38
2.7.2	Source attribution	39
2.7.3	Seasonality of <i>Campylobacter jejuni</i>	40
2.8	Bacterial Evolution	41

2.8.1	Mechanisms of bacterial evolution	42
2.9	Bacterial core- and pan genome	49
2.10	Niche adaptation and host association of <i>Campylobacter jejuni</i>	52
3	Seasonality of <i>Campylobacter jejuni</i> isolates associated with human campylobacteriosis in the Manawatu, New Zealand	57
3.1	Introduction	58
3.2	Materials and methods	60
3.2.1	Data	60
3.2.2	Time series analysis using ARIMA methods	60
3.2.3	Linear Regression Models with Generalised Least Squares	61
3.3	Results	62
3.3.1	Time series modelled with ARIMA	64
3.3.2	Generalised Least Squares	67
3.4	Discussion	68
4	Evidence of host associated <i>Campylobacter jejuni</i> genotypes in wild and agricultural hosts in New Zealand	73
4.1	Introduction	74
4.2	Methods	75
4.2.1	Isolates and sequence type (ST) data sets	75
4.2.2	Bayesian clustering analysis and admixture	75
4.2.3	ST genealogy and association between defined clades and hosts based on ClonalFrame	76
4.2.4	Multidimensional scaling	77
4.3	Results	78
4.3.1	BAPS and ClonalFrame	79
4.3.2	Estimated admixture and gene flow between host associated clusters estimated by BAPS	82
4.3.3	Statistical testing of host-association based on ClonalFrame	85
4.4	Discussion	88
5	Use of phenotypic microarrays and Genome Wide Association Study (GWAS) to identify phenotype-genotype relationships in New Zealand <i>Campylobacter jejuni</i>	93
5.1	Introduction	95
5.2	Materials and Methods	96
5.2.1	Bacteria isolates	96
5.2.2	Preparation of PM 1 and PM 9 plates	97

5.2.3	Incubation and data recording	98
5.2.4	Statistical analysis of the kinetic data	98
5.2.5	Isolation of DNA and sequencing	99
5.2.6	Quality control and <i>de novo</i> assembly of sequence reads	100
5.2.7	Gene prediction and clustering	100
5.2.8	Gene-trait matching	100
5.2.9	Confirmation of growth on salt plates	100
5.3	Results	101
5.3.1	Utilisation of carbon sources at 38°C and 42°C	101
5.3.2	Tolerance to osmolytes (PM 9) at 38°C and 42°C	105
5.3.3	Linear mixed effects models for PM 1 and PM 9	107
5.3.4	REEMtrees	110
5.3.5	Gene-trait matching	112
5.4	Discussion	123
6	Short-term evolutionary dynamics of <i>Campylobacter jejuni</i> in chickens	129
6.1	Introduction	130
6.2	Materials and methods	132
6.2.1	Bacterial strains and challenge	132
6.2.2	Experimental infection	132
6.2.3	Motility agars	133
6.2.4	Isolation of DNA and sequencing	133
6.2.5	SNP detection methods	136
6.2.6	The probability of SNP co-locations	136
6.3	Results	137
6.3.1	Testing for recombination events between the ST-474 isolates	140
6.3.2	Identification of single nucleotide polymorphisms	141
6.4	Discussion	146
7	Discussion	151
7.1	General discussion	151
7.2	Limitations and future directions	155
A	Supplementary Material for chapter 5	159

List of Figures

2.2	SEM images of <i>C. jejuni</i> isolates	10
2.3	<i>Campylobacter</i> virulence factors	11
2.4	PacBio C2X1 raw read length distribution	28
2.5	Nanopore MinION	29
2.6	MLST genes	32
2.7	rMLST genes	34
2.8	Omnilog workflow	36
2.9	overview of the 20 PM plates	37
2.10	overview of the 3 mechanisms contributing to HGT	43
2.11	possible nucleotide substitutions	45
2.13	JC69	46
2.14	transition matrix for GTR	46
2.15	transitions and transversions	46
2.16	PAM250	48
2.17	comparison of BLOSUM and PAM	48
2.18	genome comparison of 15 <i>C. jejuni</i> strains	50
3.1	time series plot	63
3.2	Box plots of monthly human campylobacteriosis cases	64
3.3	acf and pacf for CC-45	65
3.4	acf for CC-45	66
4.1	Venn diagram of shared and unique STs in New Zealand <i>C. jejuni</i>	78
4.2	ClonalFrame genealogy of <i>C. jejuni</i>	81
4.3	Admixture plots	82
4.4	Gene flow between the clusters identified by BAPS	84
4.5	Density plots	86
4.6	MDS plots	87
5.1	Utilisation heatmap of the 15 isolates at 38°C in PM 1	102

5.2	Utilisation heatmap of the 15 isolates at 42°C in PM 1	103
5.4	enlarged REEMtree	112
5.3	REEMtree for PM 1	113
5.5	Algorithm for identification of gene importance	114
5.6	Gene cassette of type IV secretion system	116
5.7	Mapped type IV secretion system	117
5.8	Sequence of the type IV secretion system	118
5.9	PhenoLink result of the <i>ggt</i> gene cassette	119
5.10	Geneious extraction of the <i>ggt</i> cassette	120
5.11	Gene cassette of the L-fucose genomic island	121
5.12	Sequence extraction of the L-fucose genomic island cassette	122
6.1	Workflow of genomic analysis	135
6.2	BRIGs plot of the two inoculum strains and one experimental isolate	140
6.3	NJ three of the core SNPs	142
6.5	SEM images of <i>C. jejuni</i> isolates	144
6.6	3-D protein structure of MreB	144
6.7	Examples of motility assays	146

List of Tables

2.1	Comparison of Next Generation Sequencing technologies	30
2.2	Genetic diversity of MLST	32
3.1	Results for the ARIMA and SARIMA time series	66
3.2	Results of the gls analysis	68
4.1	Host-associated cluster assignment based on BAPS	80
5.1	Overview of isolates used in the study	97
5.2	Utilisation in PM 1 at 38°C and 42°C	104
5.3	Tolerance to osmolytes in PM 9 at 38°C and 42°C	106
5.4	lme model results for PM 1 at 38°C and 42°C	109
5.5	lme model results for PM 9 at 38°C and 42°C	111
6.1	Genome characteristics of the 29 isolates under investigation	138
6.2	Details of the non-synonymous core SNPs across the 28 isolates	143
6.3	Results table for linear model	145