Microbial Ecology and Metagenomics of a Foodborne Pathogen

Agata Helena Dziegiel

100333106

Quadram Institute Bioscience

University of East Anglia

This thesis is submitted for the degree of Doctor of Philosophy

September 2024

Word count: 58,059 words

Abstract

Foodborne illness represents a major global health burden. *Campylobacter* is a major bacterial cause of foodborne illness, and chicken meat is a leading source of infection. Outbreaks are rarely reported, though this may partly be due to culturing difficulties and limited isolate sampling from food sources. Culture-independent methods like shotgun metagenomics can facilitate direct investigation of retail chicken microbiomes, potentiating direct *Campylobacter* characterisation and identification of microbial signatures associated with its presence and absence to inform intervention strategies.

This work utilised a combination of culture approaches, whole genome sequencing (WGS) and shotgun metagenomics to characterise *Campylobacter* diversity on retail chicken and assess the viability of culture-independent sequencing for pathogen surveillance and detection of microbial signatures associated with *Campylobacter* persistence.

Campylobacter recovery from individual samples was affected by the culture method, indicating that common approaches and media can fail to recover the pathogen. Within-sample diversity was observed at the species, sequence type (ST), single nucleotide polymorphism and antimicrobial resistance genotype level, with significant implications for public health investigations. Evidence of plasmid sharing between STs indicated that the genomic diversity extends beyond the chromosome.

Detection and characterisation of *Campylobacter* directly with metagenomics was difficult due to its low abundance, and organisms associated with *Campylobacter* presence and absence could not be reliably identified. Detection may be less challenging in infection scenarios whereby the pathogen is colonising the host; as a proof-of-concept, this was demonstrated by the identification of *Salmonella enterica* from faecal samples representing culture-confirmed cases.

While metagenomics is a promising method for microbiome profiling and pathogen identification, optimisation in sequencing, analysis pipelines and databases is needed to improve efficacy for food safety applications, where culture and WGS remains important for *Campylobacter* typing. In clinical scenarios, metagenomic pathogen detection is more effective, though sample preparation and analysis pipeline development can improve characterisation efficacy.

Acknowledgements

I would firstly like to express a huge thanks to my primary supervisor, Prof Alison Mather, for the opportunity to work on this PhD project, her endless support and guidance and opportunities that I have been given over the last four years. I was told many times that the key to PhD success is having a good supervisory team, and I can honestly say I have been very fortunate to have had such a knowledgeable and inspiring mentor. Thank you to my secondary supervisor, Prof John Wain, for always encouraging me to think outside the box, for all of his support and insights that have been highly beneficial throughout the project. I also want to thank Dr Nicol Janecko for her guidance and feedback, for passing on her knowledge from experience in *Campylobacter* research, and the times spent discussing the struggles of working with Campy! A big thanks also goes to Dr Samuel Bloomfield, without whom I would not have got to where I am today – his technical help and support has been hugely valuable to me, and I really appreciate the words of reassurance during (frequent) times of panic!

To other past and present members of the Mather group – Raphaëlle Palau, Dr Roxana Zamudio-Zea, Dr Leonardo de Oliveira Martins, Dr Ricardo Castellanos, Dr Carol Davies-Sala, Dr Min Yap, Dr Cailean Carter, Gab Astorga and Sam Mellor – thank you for everything. I feel very fortunate to have been a part of such a friendly and multidisciplinary lab. I am also grateful for support from Dr George Savva, Dr Noemi Tejera-Hernandes, Dr Bilal Djeghout, Dr Dipali Singh, Dr Sumeet Tiwari, Dr Thanh Le Viet, Prof Mark Webber, David Baker and the core sequencing team, as well as the core laboratory support team.

To external collaborators, and past and current contacts at the Food Standards Agency – Dr Aldert Zomer, Dr Roosmarijn Luiken, Dr Linda van der Graaf-van Bloois, Prof Jaap Wagenaar, Prof Stephen Baker, Dr Vu Thuy Duong, Victoria Cohen, Dr Daniel Lloyd and Dr Ben Goodall – thank you so much for your input, help and feedback.

I feel very lucky to have met amazing friends during my PhD. To Maria, Iliana, Emma H, Ho Yu, Heather, Oleksii, Gab, Min and Sam (again) – thank you for motivating me, making me laugh, distracting me from my PhD woes when needed, and for being generally great people and fantastic scientists.

Endless thanks to my fiancé Jack for keeping me somewhat sane, always supporting me in my endeavours and for listening to me complain every day and still sticking around. To my tortoise Brahms for being a constant in my life over the last 26 years and entertaining me every day. To my family, especially my Mum for being an inspiration and shaping me into the person I am. Thank you also to my friends from undergrad and my lifelong friends in Portsmouth for making me feel like I always have a home to return to and for keeping me going.

Finally, thank you to the Biotechnology and Biological Sciences Research Council, the Norwich Research Park Biosciences Doctoral Training Partnership and the Food Standards Agency for funding the project. The whole PhD process at the Quadram has been very rewarding. It would take a whole thesis in itself to fully express my gratitude to everyone that has been part of this journey, but I truly appreciate the opportunity to have worked alongside amazing scientists and everything I was able to get involved in over the last four years. It has been a wild ride, but it has been great!

Table of Contents

1. **Chapter 1**: General introduction 20
   1.1 *Campylobacter* taxonomy, pathogenicity and epidemiology 20
      1.1.1 Human disease 20
      1.1.2 Epidemiology of *Campylobacter* 21
      1.1.3 Infection sources 21
   1.2 *Campylobacter* identification and typing methods 22
      1.2.1 *Campylobacter* growth and survival requirements 22
      1.2.2 Isolation from chicken samples 23
      1.2.3 Immunological identification and characterisation methods 30
      1.2.4 Molecular identification and characterisation methods 31
      1.2.5 Whole genome sequencing of *Campylobacter* isolates 33
      1.2.6 Direct sequencing for *Campylobacter* detection 34
         1.2.6.1 Amplicon sequencing – metabarcoding 34
         1.2.6.2 Shotgun metagenomics 35
   1.3 *Campylobacter* ecology 39
      1.3.1 Genetic drivers of diversity and niche adaptation 39
      1.3.2 *Campylobacter* diversity in chickens and on chicken meat 42
      1.3.3 Sampling strategies and limitations in light of *Campylobacter* diversity 46
   1.4 Other organisms and association with *Campylobacter* presence 47
      1.4.1 Poultry colonisation and effect of the chicken gut microbiota 47
      1.4.2 *Campylobacter* and retail chicken meat microbiota 52
      1.4.3 Culture-independent sequencing for microbial community profiling 53
   1.5 Aims and objectives 55
2. **Chapter 2**: General materials and methods 56
   2.1 Retail chicken sample collection 56
   2.2 *Campylobacter* detection and isolation from retail chicken 56
   2.3 DNA extraction and genome sequencing 60
   2.4 Genome analysis 61
   2.5 Retail chicken metagenome extraction and sequencing 61
   2.6 Metagenome analysis 62
   2.7 Data manipulation, visualisation and statistical analysis 63
3. **Chapter 3**: Investigation of *Campylobacter* diversity on retail chicken 64
   3.1 Introduction 64
   3.2 Aims and objectives 66
   3.3 Materials and methods 66
      3.3.1 Genome analysis 66

3.3.2 Phylogenetic analysis 67
3.3.3 Further investigation of sequence types obtained with individual methods 68
3.3.4 Statistical comparisons of culture methods 69
3.3.5 Simulation studies 69
3.4 Results 70
3.4.1 Retail chicken samples processed 70
3.4.2 Comparison of culture approaches for the isolation of *Campylobacter* 70
3.4.3 Intra-sample species and sequence type diversity and potential implications on source attribution 75
3.4.4 Diversity of individual sequence types inferred with pairwise single nucleotide polymorphism analysis 80
3.4.5 Diversity of individual sequence types inferred through antimicrobial resistance genotype analysis 82
3.4.6 Further investigation of sequence types obtained with individual methods 84
3.5 Discussion 86
3.5.1 Comparing culture methods for *Campylobacter* recovery 86
3.5.2 *Campylobacter* species and sequence type diversity on retail chicken and potential impacts on source attribution and outbreak investigation 87
3.5.3 Evaluation of the diversity of individual sequence type groups within samples 88
3.5.4 Evaluation of the recovery of *Campylobacter* species and sequence types with specific culture methods 90
3.6 Conclusions 90
4. **Chapter 4**: Investigation of *Campylobacter* plasmid diversity in retail chicken using long read sequencing 92
4.1 Introduction 92
4.2 Aims and objectives 93
4.3 Materials and methods 93
4.3.1 Long read sequencing 93
4.3.2 Long read genome quality control 94
4.3.3 Screening of short read assemblies for plasmid contigs 95
4.3.4 Short read sequencing repeats and final assembly selection 96
4.3.5 Plasmid analysis 96
4.3.6 Further investigation of plasmids in an epidemiologically important sequence type 97
4.4 Results 98
4.4.1 Selection of isolates for long read sequencing 98
4.4.2 Plasmid clustering and diversity 108
4.4.3 Further investigation of plasmids in an epidemiologically important sequence type 111
4.5 Discussion 113

4.5.1 *Campylobacter* plasmid diversity and characteristics 114
4.5.2 Investigation of plasmids in an epidemiologically relevant sequence type 115
4.6 Conclusions 116
5. **Chapter 5**: Using metagenomics for the identification of *Campylobacter* and organisms associated with *Campylobacter* presence and absence on retail chicken 118
5.1 Introduction 118
5.2 Aims and objectives 120
5.3 Materials and methods 120
5.3.1 *Campylobacter* genome analysis 120
5.3.2 *Campylobacter* quantitative polymerase chain reaction trials 121
5.3.3 Chicken metagenome analysis 123
5.3.3.1 Trimming, filtering, detection of antimicrobial resistance genes and classification 123
5.3.3.2 Identification of sequence types, read assembly and metagenome assembled genomes 123
5.3.3.3 Alpha and beta diversity analysis and visualisation of the most abundant families and phyla 124
5.3.3.4 Differential abundance analysis 125
5.3.3.5 Metabolic pathway and reaction analysis 125
5.3.4 Analysis of metagenome negative controls 126
5.4 Results 126
5.4.1 Retail chicken samples processed 126
5.4.2 *Campylobacter* quantitative polymerase chain reaction trials 127
5.4.3 *Campylobacter* culture results and metagenome sequencing 130
5.4.4 Identification of *Campylobacter* from chicken metagenomes 132
5.4.5 Alpha and beta diversity 136
5.4.6 Metagenome controls 144
5.4.7 Differential abundance analysis 147
5.4.7.1 DESeq2 147
5.4.7.2 ANCOM-BC2 152
5.4.7.3 ALDEx2 152
5.4.8 Metabolic pathway and reaction abundance comparisons 155
5.5 Discussion 159
5.5.1 Identification of *Campylobacter* in retail chicken metagenomes 159
5.5.2 Microbial signatures associated with *Campylobacter* presence and absence on retail chicken 161
5.5.3 Comparison of metabolic pathway and reaction abundances of *Campylobacter* culture positive and negative retail chicken samples 164
5.6 Conclusions 164

6. **Chapter 6**: Metagenomics-based identification of disease-causing *Salmonella enterica* serovars and antimicrobial resistance genes from faecal samples 166
6.1 Introduction 166
6.2 Aims and objectives 168
6.3 Materials and methods 168
6.3.1 Faecal sample collection and ethical approval 168
6.3.2 *Salmonella* isolation from faecal samples 168
6.3.3 DNA extraction and sequencing 169
6.3.4 Genome analysis 169
6.3.5 Metagenome analysis 171
6.3.6 Statistical analysis 175
6.4 Results 175
6.4.1 *Salmonella* genomes 175
6.4.2 Faecal metagenomes 177
6.4.2.1 Taxonomic profiling 179
6.4.2.2 Metagenome resistomes 181
6.4.3 Detection of *Salmonella enterica* and the isolated serovars in the associated faecal metagenomes 183
6.4.4 Detection of antimicrobial resistance genes identified in the associated *Salmonella enterica* genomes with metagenomics 184
6.4.5 Consideration of the wider sample microbiome in the clinical context 192
6.5 Discussion 195
6.5.1 Metagenomics for the identification of *Salmonella enterica* and the serovar isolated from salmonellosis cases 196
6.5.2 Metagenomics for the identification of *Salmonella enterica* associated antimicrobial resistance genes and comparison with healthy controls 198
6.5.3 Differences in the microbiome between the salmonellosis cases and controls 199
6.6 Conclusions 201
7. **Chapter 7**: General discussion 202
7.1 Key findings and contributions to the field 203
7.1.1 Investigation of *Campylobacter* diversity on retail chicken 203
7.1.2 Investigation of *Campylobacter* plasmid diversity 203
7.1.3 Metagenomic identification of *Campylobacter* and organisms associated with *Campylobacter* presence and absence on retail chicken 204
7.1.4 The potential of metagenomics for pathogen detection and characterisation: *Salmonella* and culture-confirmed salmonellosis cases 204
7.2 Future directions 205
7.3 Final remarks 206
8. **References** 207

Appendices 240

Appendix 1: Retail chicken samples collected, and the chapters they are referenced in 240
Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST) 242
Appendix 3: Allelic profiles of the novel sequence types (STs) identified 262
Appendix 4: Pairwise non-recombinogenic single nucleotide polymorphism (SNP) distances for *Campylobacter* sequence types (STs) identified in the study within samples and overall 263
Appendix 5: Antimicrobial resistance genotypes of *Campylobacter* sequence type (ST) groups within samples 266
Appendix 6: Plate counts and colony forming units (CFU)/mL values for comparisons of growth of four *Campylobacter* isolates representing four STs initially identified through one condition only (Bolton broth/cefoperazone, amphotericin B, teicoplanin (CAT) broth)..... 271
Appendix 7: Genes within each cluster of orthologous groups (COG) category identified in different plasmid clusters based on eggNOG analysis, summarised by preferred name and available description 274
Appendix 8: DESeq2 results showing differentially abundant taxa at the 0.05 p-value threshold (Benjamini-Hochberg corrected) 277
Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR 282
Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange..... 299
Appendix 11: Proportion of *Salmonella enterica* and *Salmonella enterica* serovars determined through genome analysis identified in the associated metagenomes; values highlighted in green indicate that the proportion represents the most likely serovar present..... 320
Appendix 12: Proportion of *Salmonella enterica* identified in the healthy control metagenomes and the top serovar result 323
Appendix 13: *Salmonella* metagenome assembled genomes (MAGs) identified in the case metagenomes 325

List of Figures

Figure 1.1: General overview of commonly used standard methods for *Campylobacter* recovery from chicken meat, including enrichment broths and agar types (Center for Food Safety and Applied Nutrition, 2000; Rosenquist, Bengtsson and Hansen, 2007; Health Canada, 2014; Standards Australia, 2015; International Organization for Standardization, 2017a, 2017b; Harrison *et al.*, 2022; United States Department of Agriculture, Food Safety and Inspection Service, 2022) 25

Figure 1.2: Phylogenetic tree based on 16S rRNA genes of representative strains within the *Campylobacter* genus using *Helicobacter pylori* as an outgroup; red rectangles represent species that have been isolated from chicken. Adapted from Miller *et al*. (2024)..... 43

Figure 2.1: Culture condition combinations used for the recovery of *Campylobacter* from retail chicken samples 58

Figure 3.1: Number of *Campylobacter* isolates recovered from 45 retail chicken samples through direct plating and enrichment in Bolton and cefoperazone, amphotericin B, teicoplanin (CAT) broth at 37°C and 42°C, split by the theoretical maximum number of isolates taken per sample 71

Figure 3.2: The observed proportion of samples testing positive for *Campylobacter* under each condition on modified charcoal-cefoperazone-deoxycholate agar (mCCDA) plates (A) and the estimated marginal mean proportions of samples testing positive under each condition in the final model (B); error bars represent 95% confidence intervals 74

Figure 3.3: Maximum likelihood trees displaying the major sequence types (STs) and their pairwise single nucleotide polymorphism (SNP) difference ranges, the chicken meat sample of origin, the culture method for recovery, and the number of antimicrobial resistance (AMR) determinants identified in the 499 *C. jejuni* genomes (A) and the 228 *C. coli* genomes (B) 76

Figure 3.4: Number of *Campylobacter* sequence types (STs) identified in the retail chicken meat samples, coloured by species..... 78

Figure 3.5: Sequence type (ST) rarefaction curve without replacement (A) and with replacement (B); simulations performed followed the principle of rarefying with replacement to determine the number of isolates required to observe 95% of the ST diversity within each chicken meat sample compared to the number of isolates that were collected (C) and to determine the average probability of detecting a randomly selected ST as the number of isolates increased, with the dashed line representing an average probability of 95% (D) 79

Figure 3.6: Highest non-recombinogenic pairwise single nucleotide polymorphism (SNP) difference between isolates of individual sequence types (STs)* within individual chicken samples (A) and overall (B), with STs displaying differences >200 SNPs annotated 81

Figure 3.7: Comparison of growth (log colony forming units (CFU)/mL, using the natural log scale) of four *Campylobacter* isolates representing four sequence types (STs) initially identified through

one condition only (Bolton/cefoperazone, amphotericin B, teicoplanin (CAT) broth) in different conditions tested based on Columbia blood agar (CBA) and modified charcoal-cefoperazone-deoxycholate agar (mCCDA) plate counts 85
Figure 4.1: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0314, the genomes long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing 98
Figure 4.2: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0317, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genome marked with X was selected for the second batch of long read sequencing 99
Figure 4.3: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0319, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genome marked with X was selected for the second batch of long read sequencing 100
Figure 4.4: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0321, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing 101
Figure 4.5: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0326, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing 102
Figure 4.6: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0331, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing 103
Figure 4.7: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0333, the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; this sample

was additionally selected as it contained the highest chromosomal ST diversity, and genomes marked with X were selected for the second batch of long read sequencing 103
Figure 4.8: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0336, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing 104
Figure 4.9: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0337, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing 105
Figure 4.10: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0338, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing 106
Figure 4.11: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0340, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing 107
Figure 4.12: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0358, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing 107
Figure 4.13: *Campylobacter* plasmid dendrogram displaying Ward clusters of Jaccard distances calculated from Roary gene presence/absence, the species of the genome, the chromosome sequence type (ST), eggNOG clusters of orthologous groups (COG) categories* from Bakta annotations, and antimicrobial resistance (AMR) genes 109
Figure 5.1: Quantitative polymerase chain reaction (qPCR) amplification curves from the 16S rRNA qPCR run that were visually inspected to evaluate the uncertain result of CH-0336 (A) and CH-0383 (B), by comparing to the curve of the positive *Campylobacter coli* control (C) and qPCR negative sample CH-0382 (D)..... 127

Figure 5.2: Relative abundance of the *Campylobacter* genus in each chicken meat metagenome determined following genus-level Bracken analysis* (A) and the relative abundance of the two most abundant *Campylobacter* species in each chicken meat metagenome determined following species-level Bracken analysis, with the rest summarised into "other"* 133
Figure 5.3: Alpha diversity indices (observed richness, Shannon and Simpson) and distributions of the *Campylobacter* culture positive and negative chicken meat metagenomes based on estimation from bacterial species-level Bracken results, after rarefying to the smallest library size 136
Figure 5.4: Rarefaction curve displaying the number of bacterial taxa identified as the number of sequences increased* 137
Figure 5.5: Non-metric multidimensional scaling (NMDS) ordination plot showing differences in beta diversity using Bray-Curtis dissimilarity in bacterial species composition based on Bracken results between *Campylobacter* culture positive and negative samples* 138
Figure 5.6: Non-metric multidimensional scaling (NMDS) ordination plot showing differences in beta diversity using Bray-Curtis dissimilarity in bacterial taxa composition based on species-level Bracken results agglomerated at the genus level between *Campylobacter* culture positive and negative samples* 139
Figure 5.7: Non-metric multidimensional scaling (NMDS) ordination plot showing differences in beta diversity using Bray-Curtis dissimilarity in bacterial taxa composition based on species-level Bracken results agglomerated at the family level between *Campylobacter* culture positive and negative samples* 140
Figure 5.8: Non-metric multidimensional scaling (NMDS) ordination plot showing differences in beta diversity using Bray-Curtis dissimilarity in bacterial taxa composition based on species-level Bracken results agglomerated at the phylum level between Campylobacter culture positive and negative samples* 141
Figure 5.9: Relative abundance of the five most abundant bacterial phyla in each of the chicken meat metagenomes based on species-level Bracken results agglomerated at the phylum level, with the rest of the phyla grouped into the "other" category, split by *Campylobacter* culture result (positive and negative) 142
Figure 5.10: Relative abundance of the top 10 most abundant bacterial families in the *Campylobacter* culture positive and negative chicken meat metagenomes grouped by phylum, using species-level Bracken results agglomerated at the family level, with the remaining families grouped into the "other" category 143
Figure 5.11: Alpha diversity indices (observed richness, Shannon and Simpson) and distributions of the *Campylobacter* culture positive and negative chicken meat samples and negative controls estimated from bacterial species-level Bracken results, after rarefying to the lowest library size 144

Figure 5.12: Alpha diversity indices (observed richness, Shannon and Simpson) and distributions of the *Campylobacter* culture positive and negative chicken meat samples and negative controls estimated from bacterial species-level Bracken results, without rarefying..... 145
Figure 5.13: Non-metric multidimensional scaling (NMDS) ordination plot showing differences in beta diversity using Bray-Curtis dissimilarity in bacterial species composition based on Bracken results between *Campylobacter* culture positive and negative chicken meat samples and negative controls* 146
Figure 5.14: Comparison of the resistome of the chicken meat metagenomes and the negative control (blank) sample associated with each processing run, based on KMA results 147
Figure 5.15: Species determined to be significantly more (log_2fold change >0) or less (log_2fold change <0) abundant in *Campylobacter* culture positive samples compared to *Campylobacter* culture negative chicken meat samples, determined with DESeq2 at the 0.01 p-value threshold based on maximum likelihood log_2fold change (Benjamini-Hochberg corrected) 148
Figure 5.16: DESeq2 normalised read counts for species found to be differentially abundant ($p<0.01$, Benjamini-Hochberg corrected) in the filtered input dataset, grouped according to *Campylobacter* culture result..... 150
Figure 5.17: Relative abundance, calculated as the percentage of the overall metagenome reads, for species identified as differentially abundant with DESeq2 ($p<0.01$, Benjamini-Hochberg corrected) in the filtered input dataset, grouped according to *Campylobacter* culture result 151
Figure 5.18: ALDEx2 Bland-Altman plot showing the relationship between taxa median relative abundance and median difference between *Campylobacter* culture positive and negative samples (left) and effect plot showing the relationship between median difference (between-group variation) and dispersion (median of the maximum within-group difference) (right)* 153
Figure 5.19: ALDEx2 Bland-Altman plot showing the relationship between taxa median relative abundance and median difference between *Campylobacter jejuni* culture positive and negative samples (left) and effect plot showing the relationship between median difference (between-group variation) and dispersion (median of the maximum within-group difference) (right)* 154
Figure 5.20: ALDEx2 Bland-Altman plot showing the relationship between taxa median relative abundance and median difference between *Campylobacter coli* culture positive and negative samples (left) and effect plot showing the relationship between median difference (between-group variation) and dispersion (median of the maximum within-group difference) (right)* 154
Figure 5.21: Three-dimensional non-metric multidimensional scaling (NMDS) ordination plot based on Bray-Curtis dissimilarity in metabolic pathway composition determined with Humann3 between *Campylobacter* culture positive and negative samples* 156
Figure 5.22: Two-dimensional non-metric multidimensional scaling (NMDS) ordination plot based on Bray-Curtis dissimilarity in metabolic pathway composition determined with Humann3

between *Campylobacter* culture positive and negative samples, after removing the unclassified proportions and renormalising the abundances* 157
Figure 5.23: Three-dimensional non-metric multidimensional scaling (NMDS) ordination plot based on Bray-Curtis dissimilarity in metabolic reaction composition determined with Humann3 between *Campylobacter* culture positive and negative samples* 158
Figure 5.24: Two-dimensional non-metric multidimensional scaling (NMDS) ordination plot based on Bray-Curtis dissimilarity in metabolic reaction composition determined with Humann3 between *Campylobacter* culture positive and negative samples, after removing the unclassified proportions and renormalising the abundances* 159
Figure 6.1: The 10 most abundant families in each case and control metagenome determined with Bracken, with remaining classifications grouped into the 'other' category..... 180
Figure 6.2: Comparison of the distribution of the number of antimicrobial resistance (AMR) variants (A) and unique genes (B), and the number of antimicrobial classes represented by the AMR genes (C) identified in the case and control metagenomes with the median indicated in red, as well as the percentage of metagenomes containing genes conferring resistance to specific antimicrobial classes (D) 182
Figure 6.3: Maximum likelihood tree based on core gene alignment of the 28 *Salmonella enterica* genomes, displaying their serovars, sequence types (STs) and the antimicrobial resistance (AMR) determinants identified with KMA; the AMR determinant matrix also indicates whether the gene was identified in the associated faecal metagenome, the relative abundance of *S. enterica* determined with Centrifuge and overall number of reads in the metagenome 186
Figure 6.4: Antimicrobial resistance (AMR) genes associated with the *Salmonella enterica* genomes in the control faecal metagenomes 188
Figure 6.5: Relative abundances of *Escherichia coli* across the case and control metagenomes, determined with Bracken..... 193
Figure 6.6: Comparison of the number of unique *Salmonella enterica* associated antimicrobial resistance (AMR) genes and other genes present in the associated faecal metagenomes according to KMA (A) and the relative abundance of unique *S. enterica* AMR genes compared to the overall AMR gene repertoire in the metagenomes (B)..... 195

List of Tables

Table 1.1: Associations identified between *Campylobacter* and other taxa identified in chicken faecal or caecal samples, *in vitro* and *in vivo* models* 48

Table 3.1: Comparison table for the agar plate types used for *Campylobacter* recovery 72

Table 3.2: Comparison of the main effects logistic regression model using modified charcoal-cefoperazone-deoxycholate agar (mCCDA) plate data only (mCCDA_main) and a logistic regression model with an interaction term between Broth and Temperature (mCCDA_interaction), with the chicken sample modelled as a random effect; the values represent odds ratios, confidence intervals and p-values 73

Table 3.3: Model comparisons using analysis of variance (ANOVA) 74

Table 3.4: Antimicrobial resistance determinants identified in the *Campylobacter jejuni*, *Campylobacter coli* and *Campylobacter lari* genomes obtained from retail chicken meat 83

Table 4.1: Plasmid clusters (based on Jaccard distance and Ward clustering of Roary gene presence/absence) identified in each retail chicken sample, and the chromosomal sequence types (STs) of the genomes in which they were identified 110

Table 4.2: Unique genes identified in all plasmids within cluster 1, associated with ST-6175 genomes 113

Table 5.1: Primer and probe sets trialled on the chicken metagenome samples 122

Table 5.2: *Campylobacter* quantitative polymerase chain reaction (qPCR) results; crossing point (Cp) values >40 represent negative results 128

Table 5.3: *Campylobacter* species and sequence types (STs) found in the analysed chicken metagenomes 131

Table 5.4: Quality control metrics for the *Campylobacter* metagenome assembled genomes (MAGs) obtained from sample CH-0317 using Metabat2 and Maxbin2 135

Table 6.1: Accession information for the *Salmonella enterica* genomes 170

Table 6.2: Accession information for the faecal metagenomes 172

Table 6.3: *Salmonella enterica* serovars and sequence types (STs) of the 28 genomes generated 176

Table 6.4: Metadata associated with the faecal metagenomes analysed 178

Table 6.5: Antimicrobial resistance (AMR) genes identified in the Metabat2 metagenome assembled genomes (MAGs) from the case metagenomes 190

Table 6.6: Antimicrobial resistance (AMR) genes identified in the Maxbin2 metagenome assembled genomes (MAGs) from the case metagenomes and BLASTn hits for the contigs 191

List of Abbreviations

AHB	Abeyta-Hunt-Bark agar
AIC	Akaike information criterion
AMR	Antimicrobial resistance
ANOVA	Analysis of variance
BAM	Bacteriological Analytical Manual
BIC	Bayes Information Criterion
BPW	Buffered peptone water
CARD	Comprehensive Antibiotic Resistance Database
CAT	Cefoperazone, amphotericin B, teicoplanin
CBA	Columbia blood agar
CC	Clonal complex
CFU	Colony forming units
cgMLST	Core genome multi-locus sequence typing
clr	Centred log ratio
COG	Cluster of orthologous groups
Cp	Crossing point
Df	Degrees of freedom
DNA	Deoxyribonucleic acid
ELISA	Enzyme-linked immunosorbent assay
ENA	European Nucleotide Archive
FDA	US Food and Drug Administration
FSA	Food Standards Agency
FSIS	Food Safety and Inspection Service
HGT	Horizontal gene transfer
HPC	High performance computing
HTS	High-throughput sequencing
ICC	Intraclass information criterion
ISO	International Organization for Standardization
LMIC	Low- and middle-income countries
logLik	Log likelihood
MAG	Metagenome assembled genome
mCCDA	Modified charcoal-cefoperazone-deoxycholate agar

MDR	Multidrug resistant
MFLP	Canadian Standard Microbiology Food Laboratory Procedure
MGE	Mobile genetic element
MLG	Microbiology Laboratory Guidebook
MLST	Multi-locus sequence typing
NCBI	National Center for Biotechnology Information
NCTC	National Collection of Type Cultures
NMKL	Nordic Committee of Food Analysis
NTS	Nontyphoidal *Salmonella enterica*
Num.Obs	Number of observations
ONT	Oxford Nanopore Technologies
ORF	Open reading frame
OTU	Operational taxonomic unit
PacBio	Pacific Biosciences
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PERMANOVA	Permutational analysis of variance
PFGE	Pulsed-field gel electrophoresis
QC	Quality control
qPCR	Quantitative polymerase chain reaction
R2 Cond.	Conditional R-squared
R2 Marg.	Marginal R-squared
RMSE	Root square mean error
rRNA	Ribosomal ribonucleic acid
SCFA	Short chain fatty acid
SNP	Single nucleotide polymorphism
SRA	Sequence Read Archive
ST	Sequence type
T4SS	Type IV secretion system
T6SS	Type VI secretion system
UK	United Kingdom
UKHSA	UK Health Security Agency
u-mCCDA	Modified charcoal-cefoperazone-deoxycholate agar without supplements
USDA	United States Department of Agriculture

wgMLST	Whole genome multi-locus sequence typing
WGS	Whole genome sequencing

1. Chapter 1: General introduction

Foodborne illness represents a major global health burden. Based on recent reports by the World Health Organization Foodborne Disease Burden Epidemiology Reference group, it was estimated that 1 in 10 people would experience foodborne disease each year, and the number of deaths resulting from such illness is similar to major health threats such as malaria and tuberculosis (Pires and Devleesschauwer, 2021). Diarrheagenic pathogens in particular were found to represent the majority (90%) of the disease burden. The emphasis on the importance of foodborne disease by the World Health Organization has driven an increase in studies to provide accurate estimates of the burden of disease at the local scale, though also highlighted that more research is needed to account for data gaps, including those pertaining to diarrhoeal pathogens (Pires and Devleesschauwer, 2021). This thesis is about developing methods that may improve the tracking of such organisms, with a primary focus on *Campylobacter*, one of the least well understood foodborne pathogens.

1.1 *Campylobacter* taxonomy, pathogenicity and epidemiology

Campylobacter is a genus of Gram-negative, spiral-shaped, oxidase-positive bacteria (Vandamme *et al.*, 2010). The *Campylobacter* genus lies within the order of Campylobacterales (Garrity, Bell and Lilburn, 2005), in the class of Epsilonproteobacteria (Oren and Garrity, 2021). Classically Epsilonproteobacteria were classified within the Proteobacteria phylum, which has more recently been reclassified as Pseudomonadota (Oren and Garrity, 2021). As of July 2024, the *Campylobacter* genus consists of 48 species recognised in the List of Prokaryotic names with Standing in Nomenclature (Parte *et al.*, 2020), with some further divided into subspecies.

1.1.1 Human disease

Campylobacter spp. are recognised as the leading bacterial cause of gastroenteritis worldwide (Kaakoush *et al.*, 2015). In the majority of individuals, *Campylobacter* infection leads to acute gastroenteritis, which is self-limiting and does not require medical treatment. However, immunocompromised individuals may develop systemic infection, chronic infection (Bloomfield *et al.*, 2018), or serious post-infectious sequelae (Allos, 1997; Barrett *et al.*, 2018). Post-infectious events like Guillain-Barré syndrome are linked to the similarity between *Campylobacter* lipooligosaccharide structures and ganglioside oligosaccharide structures on human nerve cells that result in autoimmunity; this is also part of the reason why vaccine development has been difficult (Poly *et al.*, 2018). Disease duration and severity depends on the strain and species

implicated (Nielsen *et al.*, 2012). The species most commonly implicated in disease is *C. jejuni*, accounting for approximately 90% of cases in England, followed by *C. coli* (~10%), with other *Campylobacter* species making up a small proportion of cases (UK Health Security Agency, 2024).

1.1.2 Epidemiology of *Campylobacter*

The epidemiology of *Campylobacter* differs between high and low-and-middle-income countries (LMICs). In LMICs where *Campylobacter* infections are commonly encountered, there is suggestion of adaptive immunity following encounter early in life (Rao *et al.*, 2001), whereas this is not observed in high income countries (Tribble *et al.*, 2010). Asymptomatic shedding is evident in LMICs (Pazzaglia *et al.*, 1991), facilitating human-to-human spread (Pascoe *et al.*, 2020). This has allowed adaptation of certain lineages in communities in which *Campylobacter* is able to persist. Asymptomatic shedding also makes reservoirs and transmission pathways difficult to determine. Host factors, like malnutrition (Rao *et al.*, 2001), may have an effect on disease presentation (Pascoe *et al.*, 2020).

The *Campylobacter* investigations presented in this thesis were conducted in the UK, where the median annual cost of campylobacteriosis has been estimated at £50 million (Tam and O'Brien, 2016), and 96.3 cases per 100,000 population were reported in England in 2022 (UK Health Security Agency, 2024). Outbreaks of campylobacteriosis are considered to be rare, with approximately 1% of cases in Europe being outbreak attributed (Liu *et al.*, 2022). Underreporting is a major issue in outbreak tracing and estimation of true incidence rates (de Wit *et al.*, 2001; Wagenaar, French and Havelaar, 2013). In the UK, it is estimated that the number of community cases is over nine times higher than the number of cases reported to national surveillance (Tam *et al.*, 2012). Imported cases associated with recent travel are sometimes reported, although domestic infections make up the majority of known cases (European Centre for Disease Prevention and Control, 2024). There is evidence of seasonal spikes in campylobacteriosis, with a peak in infections typically occurring between mid-June and mid-July in England and Wales (Louis *et al.*, 2005). Risk factors include age (Nichols *et al.*, 2012), animal contact (Doorduyn *et al.*, 2010) as well as increased dissemination by insects and other vectors (Nichols, 2005).

1.1.3 Infection sources

Campylobacter has been isolated from a range of hosts, including wild birds (Kwon *et al.*, 2017), pigs (Kempf *et al.*, 2017), cattle (An *et al.*, 2018) and domesticated animals (Pölzler, Stüger and Lassnig, 2018). Poultry, specifically chicken meat, is reported as the leading infection source (Skarp, Hänninen and Rautelin, 2016; European Food Safety Authority and European Centre for

Disease Prevention and Control, 2019), linked to 20-83% of human infections (EFSA Panel on Biological Hazards (BIOHAZ), 2010; Cody *et al.*, 2019). The prevalence of contamination of chickens available at retail in the UK was 56% according to the most recent Food Standards Agency (FSA) report that sampled chickens across different store types (Jorgensen *et al.*, 2019), although this varied significantly between retailers. In smaller shops, 75% whole chickens were positive for *Campylobacter*, of which 15% had >1000 CFU/10g neck skin. Of these, approximately 75% were positive for *C. jejuni* and the remainder were positive for either *C. coli* only or both species. The FSA monitoring studies took place between 2014-2020, sampling major and non-major retailer meat between 2014-2017 and non-major retailer meat only for the remaining years. The prevalence of *Campylobacter* on retail chicken derived from major retailers, which hold the most market share (Kantar, 2024), may therefore be outdated. The survey also focused on neck skins of whole chickens only, thus prevalence rates may not be representative of other chicken products available at retail. Reported prevalence can also potentially be affected by the detection methods used.

1.2 *Campylobacter* identification and typing methods

1.2.1 *Campylobacter* growth and survival requirements

The current, widely adopted method for the identification of *Campylobacter* relies on culture using *Campylobacter* selective media. Growth requirements can be species-specific, though most standard isolation methods are mainly optimised for *C. jejuni* isolation.

Amino acids, citric acid cycle intermediates including 2-oxoglutarate, fumarate and succinate and organic acids such as lactate, pyruvate and acetate are the primary energy sources for *C. jejuni* growth and survival (Hofreuter, 2014), as carbohydrates are not typically utilised by the bacteria, except for a small proportion of strains with a functional Entner-Doudoroff pathway (Vegge *et al.*, 2016) or L-fucose pathway (Stahl *et al.*, 2011). Model *C. jejuni* strain 81-176 utilises serine as the preferred energy source, which is converted to pyruvate that feeds into the citric acid cycle (Gao *et al.*, 2017). This substrate appears to be necessary for colonisation of the chicken gut, as mutation in the serine transporter *sdaC* results in colonisation hindrance (Gao *et al.*, 2017). Along with serine, influx of aspartate, asparagine, glutamate, glutamine and proline may also occur, although proline utilisation becomes more important once other amino acids have been mostly metabolised, and glutamine and asparagine uptake appears to be strain-dependent (Stahl, Butcher and Stintzi, 2012). These amino acids are abundant in chicken excreta (Parsons, Potter and Brown, 1983), highlighting potential *C. jejuni* niche adaptation.

However, it has been found that specific *C. jejuni* strains exhibit differing auxotrophies for substrates. A metabolic model utilising the annotated genome of a strain epidemiologically linked to poultry isolates (M1cam) identified a preference for methionine, alongside pantothenate and niacinamide in the culture medium (Tejera *et al.*, 2020). On the other hand, some *C. jejuni* strains are auxotrophic for cysteine instead of methionine (Vorwerk *et al.*, 2014). This suggests that absence of some substrates from culture media may result in reduced ability to isolate certain *C. jejuni* strains, and therefore culture media used to isolate *Campylobacter* must contain a variety of substrates to facilitate detection of different strains.

C. jejuni require oxygen for biomass production, as growth is typically reduced or not possible in anaerobic conditions (Bolton and Coates, 1983); however, aerobic concentrations of oxygen can be toxic, particularly in low-density cultures (Kaakoush *et al.*, 2007). As a result, in the laboratory, microaerophilic conditions are typically used for *C. jejuni* isolation; this comprises of approximately 5% oxygen, 10% carbon dioxide and 85% nitrogen (Harrison *et al.*, 2022). Although not adapted to purely aerobic conditions, many *Campylobacter* spp. possess a number of factors that diminish oxidative stress, including a cytochrome c peroxidase homologue (Cj0020c) (Hendrixson and DiRita, 2004) and katA catalase (Grant and Park, 1995), thus providing tolerance to more oxygen exposure that may be important for survival in unfavourable conditions, such as on the surface of chicken meat. Hydrogen can be used as an electron donor in oxidative phosphorylation, and therefore addition of hydrogen gas can be beneficial for isolation; it has also been demonstrated that *C. jejuni* can survive, but not grow, using hydrogen as a sole substrate (Weerakoon *et al.*, 2009; Stoakes *et al.*, 2022).

1.2.2 Isolation from chicken samples

A number of methods have been employed to isolate *Campylobacter* from chicken meat samples. There are multiple widely applied national and international standard methods that have been developed and validated for *Campylobacter* detection and sometimes enumeration (Figure 1.1). These include the International Organization for Standardization (ISO) standard (International Organization for Standardization, 2017a, 2017b) used across Europe, the Nordic Committee of Food Analysis (NMKL) method (Rosenquist, Bengtsson and Hansen, 2007) for Nordic and Baltic regions, the United States Department of Agriculture, Food Safety and Inspection Service Microbiology Laboratory Manual (FSIS USDA MLG) (United States Department of Agriculture, Food Safety and Inspection Service, 2022) and US Food and Drug Administration Bacteriological Analytical Manual (FDA BAM) (Center for Food Safety and Applied Nutrition, 2000) in the United States, the Australian standard applied in Australia and New Zealand (Standards Australia, 2015), and the Canadian Standard Microbiology Food Laboratory Procedure (MFLP) (Health Canada,

2014) used for monitoring of *Campylobacter* in food in Canada. Other local and national *Campylobacter* monitoring studies either utilise these methods directly or with minor modifications; for example, the UK Health Security Agency (UKHSA; formerly Public Health England) and FSA *Campylobacter* monitoring studies utilise a modified version of the ISO method (Food Standards Agency, 2016; Public Health England, 2020). Similarly, a modified version of the ISO standard has been applied to recover a wider range of *Campylobacter* species in Ireland (Duffy, Cagney and Lynch, 2007; Lynch *et al.*, 2011). Although there are similarities between the methods, they differ in various aspects, reducing comparability between studies, and suggesting that there is currently no ideal method for the identification of all *Campylobacter* strains and species from chicken meat (Harrison *et al.*, 2022). It is therefore an important area that requires evaluation and improvement.

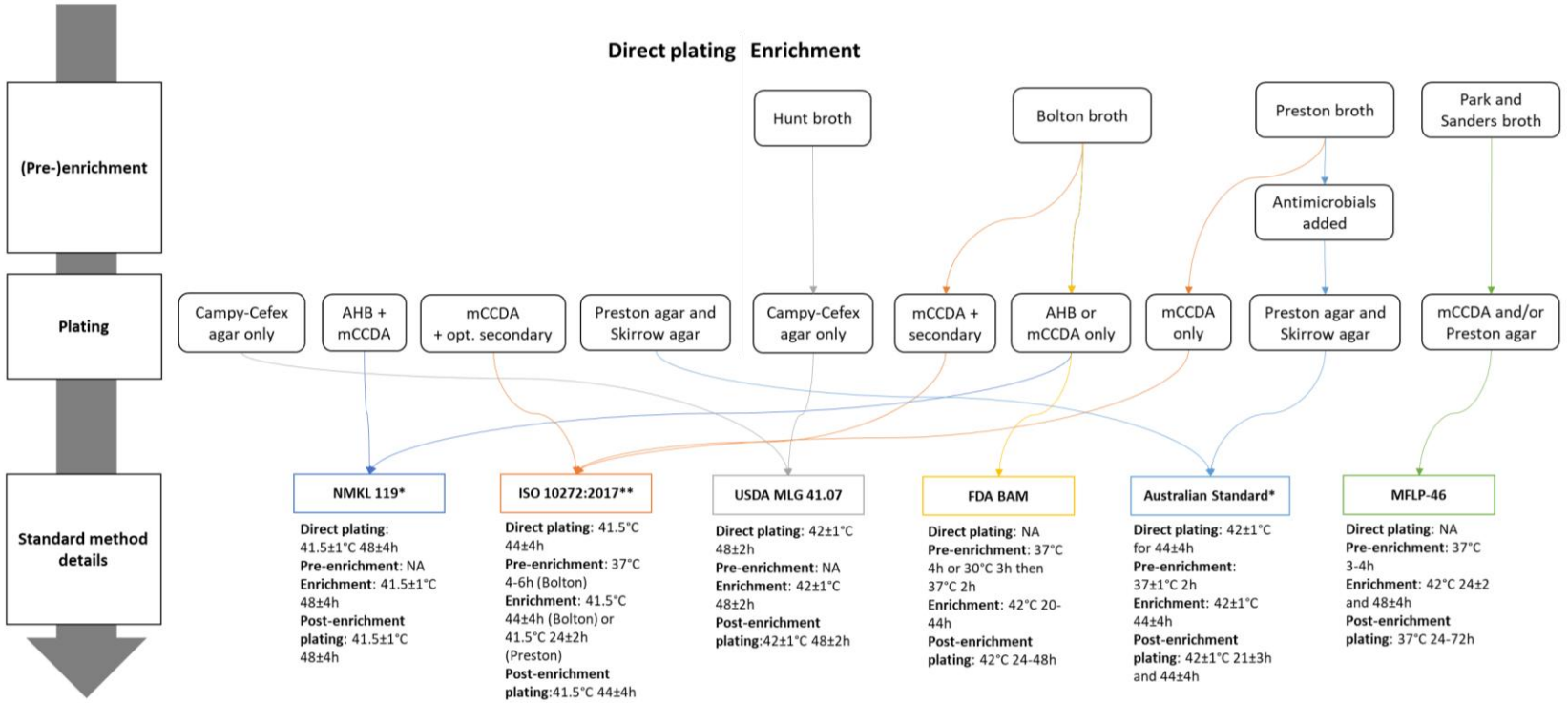

Figure 1.1: General overview of commonly used standard methods for *Campylobacter* recovery from chicken meat, including enrichment broths and agar types (Center for Food Safety and Applied Nutrition, 2000; Rosenquist, Bengtsson and Hansen, 2007; Health Canada, 2014; Standards Australia, 2015; International Organization for Standardization, 2017a, 2017b; Harrison *et al.*, 2022; United States Department of Agriculture, Food Safety and Inspection Service, 2022)

opt.=optional, AHB=Abeyta-Hunt-Bark agar, mCCDA=modified charcoal-cefoperazone-deoxycholate agar, NMKL=Nordic Committee on Food Analysis, ISO=International Organization for Standardization, USDA MLG=United States Department of Agriculture Microbiology Laboratory Guidebook, FDA BAM=US Food and Drug Bacteriological Analytical Manual, MFLP=Canadian Standard Microbiology Food Laboratory Procedure

*this standard has separate streams for qualitative vs quantitative detection and/or expected sample contamination levels; **this standard has separate streams for expected sample contamination levels and expected levels of competing flora; direct plating can also be used for enumeration

Culture methods can be generally split into direct plating and enrichment approaches. Direct plating of poultry samples is recommended by FSIS USDA (United States Department of Agriculture, Food Safety and Inspection Service, 2022) and ISO standards (International Organization for Standardization, 2017a) for the detection of *Campylobacter* from samples with high abundances of the pathogen. The ISO (International Organization for Standardization, 2017b), NMKL (Rosenquist, Bengtsson and Hansen, 2007) and Australian standard (Standards Australia, 2015) methods also contain enumeration techniques for quantifying *Campylobacter* using a direct plating approach. Such approaches have been utilised by the UKHSA and in the FSA monitoring of *Campylobacter* on retail chicken (Food Standards Agency, 2016; Jorgensen *et al.*, 2019; Public Health England, 2020). Campy-Cefex, Abeyta-Hunt-Bark (AHB), Preston, Skirrow and modified charcoal-cefoperazone-deoxycholate agar (mCCDA) are commonly used for this purpose (Rosenquist, Bengtsson and Hansen, 2007; Standards Australia, 2015; International Organization for Standardization, 2017a, 2017b; Public Health England, 2020), containing antimicrobials such as cefoperazone, rifampicin, vancomycin, trimethoprim, polymyxin B and amphotericin B or cycloheximide, as well as either blood or charcoal to reduce accumulation of oxygen derivatives that are toxic to *Campylobacter* (Bolton, Coates and Hutchinson, 1984). *C. jejuni* and other common species are typically resistant to the antimicrobials in *Campylobacter* selective media (Corry *et al.*, 1995), allowing growth of the organism to a detectable level and a reduction of background flora. Agars also often contain additional aerotolerance- and growth-enhancing substrates such as sodium pyruvate (Center for Food Safety and Applied Nutrition, 2000), and sometimes dyes such as trimethyl tetrazolium chloride added to AHB in the NMKL method (Rosenquist, Bengtsson and Hansen, 2007). Direct plating can potentially enhance detection of *Campylobacter* in the case of high contamination with other organisms that can bloom in enrichment media (Jasson *et al.*, 2009). However, *Campylobacter* is often present at low abundance on chicken meat (Jorgensen *et al.*, 2019), likely due to the fact that growth is not supported in the storage and packaging conditions, thus the pathogen is contaminating rather than colonising the meat (Murphy, Carroll and Jordan, 2006; Kim *et al.*, 2021). An enrichment step may therefore be required, as the limit of detection of direct plating approaches is often lower than enrichment (Harrison *et al.*, 2022). There is also some evidence that recovery of certain species, such as *C. coli*, may be improved with the use of enrichment compared to direct plating (Ladely *et al.*, 2017).

Enrichment broths for *Campylobacter* detection include Bolton broth, Preston broth, Hunt broth and Park and Sanders broth (Center for Food Safety and Applied Nutrition, 2000; Rosenquist, Bengtsson and Hansen, 2007; Health Canada, 2014; Standards Australia, 2015; International Organization for Standardization, 2017a; United States Department of Agriculture, Food Safety and Inspection Service, 2022), usually supplemented with blood and antimicrobials, though the

specific composition of these broths can vary by method. Bolton broth facilitates the recovery of injured cells derived from unfavourable environments (Public Health England, 2020), and is thus often used to enrich chicken meat samples (Williams, Ebel and Nyirabahizi, 2021; Poudel *et al.*, 2022). It is typically supplemented with cefoperazone, trimethoprim, vancomycin and amphotericin B or cycloheximide (Public Health England, 2020; Harrison *et al.*, 2022). In some cases natamycin is used as an alternative antifungal agent (Harrison *et al.*, 2022). Bolton broth is the recommended option by ISO and NMKL for samples with low *Campylobacter* abundance, and potential presence of sub-lethally injured cells (Rosenquist, Bengtsson and Hansen, 2007; International Organization for Standardization, 2017a). Other standards, such as the FDA BAM, as well as the UKHSA method and Irish method for enhanced recovery of different *Campylobacter* species recommend Bolton broth as the sole isolation medium (Center for Food Safety and Applied Nutrition, 2000; Public Health England, 2020). The USDA MLG method traditionally utilised Bolton broth, but the most recent version of the standard recommends Hunt broth, which has a different basal formula but the same selective supplements as Bolton broth (Hardy Diagnostics, 2020). Park and Sanders broth is used for *Campylobacter* isolation by the Canadian MFLP; this consists of Brucella broth base supplemented with vancomycin, trimethoprim, cefoperazone and optionally cycloheximide (Health Canada, 2014), providing similar selective properties to Bolton broth. Preston broth is recommended by ISO in situations where a high abundance of competing flora is suspected (International Organization for Standardization, 2017a), and is the sole isolation broth recommended by the Australian standard (Standards Australia, 2015). Preston broth is supplemented with polymyxin B, rifampicin, trimethoprim and cycloheximide or amphotericin B, providing different selective properties to Bolton broth. The enrichment step in standard methods usually ranges between 24 and 48 hours (with acceptable upper and lower limits), and can also include a pre-enrichment of two to six hours in the supplemented broth (Center for Food Safety and Applied Nutrition, 2000; Health Canada, 2014; International Organization for Standardization, 2017a). Alternatively, the supplements are added to the broth after the pre-enrichment (Standards Australia, 2015). The pre-enrichment step is performed at 37°C, aiming to resuscitate sub-lethally injured cells, after which the temperature is increased to 42°C for the remainder of the isolation process (Center for Food Safety and Applied Nutrition, 2000; Health Canada, 2014; Standards Australia, 2015; International Organization for Standardization, 2017a).

Following enrichment, the sample is cultured on selective agar plates as a secondary selective step. Common selective media used include mCCDA, Campy-Cefex agar, AHB agar, Preston and Skirrow agars (Center for Food Safety and Applied Nutrition, 2000; Health Canada, 2014; Standards Australia, 2015; International Organization for Standardization, 2017a; United States Department of Agriculture, Food Safety and Inspection Service, 2022). The plates are usually

incubated for one to two days (Harrison *et al.*, 2022). The use of membrane filters can enhance isolation of *Campylobacter*, given sufficient time and volume for infiltration on dried plates. This has been evidenced in enriched chicken meat samples filtered in 20 µL volume portions onto 0.65 µm cellulose filters on modified Campy-Cefex agar for 15 minutes (Speegle *et al.*, 2009).

Despite their extensive validation, standard methods can still fail to recover *Campylobacter* (Ugarte-Ruiz *et al.*, 2012; Seliwiorstow *et al.*, 2016; Jo *et al.*, 2017). Changing the combination of enrichment broths and agar types, as well as altering the sample to broth ratio, can be beneficial for *Campylobacter* recovery, and this has been investigated by a number of studies. A study by Junhyung Kim *et al*. (2019) found that a sample to broth ratio of $1:10^3$ was optimal for *Campylobacter* recovery from highly contaminated chicken faecal samples, performing better than $1:10^1$ and $1:10^2$. In the same study, enrichment in Preston broth followed by plating on mCCDA yielded the highest abundance of *C. jejuni* compared to Bolton broth with mCCDA/Preston agar and Preston broth with plating on Preston agar. This was likely due to the presence of competing organisms that are also resistant the supplemented antimicrobials, such as *Proteus*, *Enterococcus* and extended-spectrum beta-lactamase producing *E. coli*. Such organisms are also commonly found on chicken meat (Overdevest *et al.*, 2011; Manson *et al.*, 2019), thus different broth-plate combinations, different antibiotic combinations and sample to broth ratios may be beneficial in identification of *Campylobacter* from retail chicken. In fact, a comparison of Bolton and Preston broth enrichment of chicken carcasses followed by selective plating identified differing microbiomes between the two conditions, with evidence that certain taxa, like *Escherichia*, may negatively affect recovery of *Campylobacter* (Jinshil Kim *et al.*, 2019). Although Preston broth may be more effective in removing non-target contaminants due to its increased selectivity, use of Preston selective media can also reduce recovery of sensitive or injured *Campylobacter* cells (Baylis *et al.*, 2000), and polymyxin B has been shown to inhibit *C. coli* growth (Ng, Stiles and Taylor, 1985). This can potentially reduce the identification of the full diversity of *Campylobacter* strains in the sample. Instead, the selectivity of Bolton broth can be improved. Supplementation of Bolton broth with additional antimicrobials has been explored, which may be beneficial for enhanced recovery in the presence of contaminants. For example, addition of triclosan, tazobactam or potassium clavulanate has been shown to increase *Campylobacter* recovery from poultry meat and carcass rinses (Chon *et al.*, 2014, 2018; Seliwiorstow *et al.*, 2016). Further evaluations of combinations of culture media or use of alternative supplements for the isolation of *Campylobacter* from different chicken cuts may be necessary.

Isolation of species other than *C. jejuni* and *C. coli* can be challenging and often not possible with the standard methods, due to differing growth requirements (Lastovica, 2006). The Cape Town protocol, utilising direct plating on tryptic blood agar with a filtration followed by anaerobic incubation as opposed to microaerophilic incubation recommended by standard methods (Center

for Food Safety and Applied Nutrition, 2000; Rosenquist, Bengtsson and Hansen, 2007; Health Canada, 2014; Standards Australia, 2015; International Organization for Standardization, 2017a; United States Department of Agriculture, Food Safety and Inspection Service, 2022), has been effective in the identification of a number of different species from stools (Jacob, Mdegela and Nonga, 2011). Although an enrichment temperature of 42°C is suitable for the isolation of thermotolerant species like *C. jejuni* and *C. coli*, this may affect recovery of non-thermotolerant species that require a lower culture temperature of 37°C (Center for Food Safety and Applied Nutrition, 2000). However, culturing at the lower temperature of 37°C can result in a significant decrease in inhibition of competing flora (Gee *et al.*, 2002), which may in turn reduce sensitivity. Increasing the post-enrichment plate incubation time to as long as six days can also enhance isolation of other *Campylobacter* species and subtypes (Duffy, Cagney and Lynch, 2007; Lynch *et al.*, 2011). Commonly used methods adapted for *C. jejuni* and *C. coli* recovery may therefore be inefficient for the recovery of other *Campylobacter* species. Moreover, there may also be an isolation bias for predominating subtypes of the species that the standard methods are tailored to isolate, as there is evidence that more abundant *C. jejuni* subtypes are likely to be preferentially isolated following an enrichment step compared to less abundant subtypes, thus presenting an inaccurate representation of the sample ecology (Hetman *et al.*, 2020). Establishment of an enrichment and post-enrichment plating technique to allow identification of the maximum diversity of *Campylobacter* species and strains to reflect the true diversity on chicken meat is therefore needed.

The vast majority of current isolation methods are lengthy, requiring up to five days to obtain *Campylobacter* colonies on plates (Health Canada, 2014), which then need to be purified and confirmed. The discourse around *Campylobacter* culture methods also suggests that no individual method is optimal for *Campylobacter* recovery, as highlighted by the differences between standards; there is thus still a requirement for optimised culture methods for reliable *Campylobacter* detection. Time-to-result limitations and differences between methods can be suboptimal in critical scenarios in which a pathogen needs to be identified or characterised quickly and reliably, such as in clinical settings, or source attribution studies for products with a relatively short shelf-life, such as chicken meat. Time-to-result can be improved using molecular or immunological methods used in pathogen detection that may facilitate direct identification of *Campylobacter*. Alternatively, some of these methods can be applied for further characterisation of the pathogen following isolation.

1.2.3 Immunological identification and characterisation methods

Immunological methods, relying on the detection of markers present on the outer membrane or within bacterial cells, can be applied to detect *Campylobacter* and other foodborne pathogens (Ricke *et al.*, 2019). These methods include latex agglutination, flow cytometry, enzyme-linked immunosorbent assays (ELISAs) or derivatives of these that rely on antibodies specific to bacterial antigens. A number of targets have been described, including flagellin proteins, lipooligosaccharides and hippurate hydrolase (HipO) proteins (Ricke *et al.*, 2019).

Agglutination assays most often utilise latex beads attached to antibodies specific to an antigen on the bacterial cell surface that clump together if bound by the target (Molina-Bolívar and Galisteo-González, 2005). These assays target *Campylobacter* outer membrane proteins including flagellar proteins, and can be used to confirm presumptive *Campylobacter* isolates following culture-based isolation (Oyarzabal and Battie, 2012). ELISA relies on the use of mono- or polyclonal antibodies that bind to a specific target. In its most simple form, ELISA utilises antibodies attached to an enzyme specific to a target antigen, so that when the antigen is bound and the enzyme substrate is added, a detectable reaction takes place (Crowther, 2008). For the purpose of *Campylobacter* identification from complex samples like food, sandwich ELISA-based assays are most often used (Oyarzabal and Battie, 2012). These are comprised of antibodies fixed to a surface that bind a bacterial antigen that is also bound by a secondary antibody attached to an enzyme; addition of the substrate then facilitates a reaction that can be measured. Flow cytometry is another detection method, which works by passing a sample through a laser beam in a single-cell manner, resulting in light scattering in a specific pattern. The cells need to be labelled with a fluorescent marker, which is often attached to an antibody specific for a target, to facilitate detection (Picot *et al.*, 2012). Modifications of this system utilise beads or magnetic microspheres attached to antibodies with affinities for different antigens for simultaneous detection of different targets; such assays have been tested for simultaneous identification of *Campylobacter*, *Salmonella*, *Listeria*, *Escherichia coli* and enterotoxin B of *Staphylococcus* (Kim *et al.*, 2010).

Major drawbacks of the routine use of immunological assays include a high limit of detection that can be larger than the natural contamination level of food samples (Oyarzabal and Battie, 2012). Although chicken faecal samples can contain over 10^6 *Campylobacter* colony forming units (CFU)/g (Seliwiorstow *et al.*, 2015) resulting from replication of the organism in the gut, chicken meat samples that are merely contaminated with the pathogen often display *Campylobacter* abundances of less than 1,000 CFU/g (Jorgensen *et al.*, 2019), which may be missed when testing without any prior enrichment due to the limit of detection. As a result, many of these methods, particularly the latex agglutination kits, are often only recommended for isolate confirmation following culture-based isolation (Center for Food Safety and Applied Nutrition, 2000; United

States Department of Agriculture, Food Safety and Inspection Service, 2022). Depending on the target, the antibodies used in immunological assays can cross-react with targets present on closely related bacteria, thus reducing the ability to characterise particular species (Ricke *et al.*, 2019). As a result, the usefulness of such assays for *Campylobacter* characterisation is limited.

1.2.4 Molecular identification and characterisation methods

Molecular methods can be used for confirmation and characterisation of isolates following culture or *Campylobacter* identification directly from samples of interest. Some of the common methods utilised in the context of confirmation and direct detection from food samples are discussed below.

Polymerase chain reaction (PCR) based methods are sensitive and can be utilised for pathogen identification, or isolate confirmation and genotyping (Liu *et al.*, 2023). PCR relies on primers that attach to specific nucleotide sequences, allowing amplification of a specific stretch of DNA using a polymerase enzyme. In quantitative PCR (qPCR), a fluorescent probe is used to detect the specific DNA sequence and indicate the amount present in a sample (Dymond, 2013). PCR is often used for confirmation or speciation of isolates that have been obtained through culture-based isolation (Jribi *et al.*, 2017; Public Health England, 2020).

PCR-based detection of pathogens directly from samples of interest has been described but can be complex. A number of targets have been used, including the *Campylobacter* 16S rRNA gene (de Boer *et al.*, 2013) for the detection of the *Campylobacter* genus. Other targets include conserved gene sequences common to particular species, such as *cadF* for *C. jejuni* and *C. coli* (Platts-Mills *et al.*, 2014), *hipO* specific for *C. jejuni* and partial sequences of *asp* for *C. coli* (Persson and Olsen, 2005). Primer design is informed by what is expected to be present in the sample, thus introducing limitations in terms of resolution, or the requirement to run multiple assays. This can be somewhat overcome by using multiplex qPCR containing multiple primer and probe sets, allowing simultaneous identification of multiple targets (Shen, 2019). This potentiates simultaneous typing of multiple species, for example, without the requirement to run the reaction multiple times with different gene targets. This method has been applied to identify different *Campylobacter* species from clinical and food production samples (Ricke *et al.*, 2019). However, designing multiplex qPCR assays targeting multiple templates is difficult; there is a limitation in the number of fluorescent probes that can be simultaneously detected by PCR machines (European Commission, 2021), and primer sets need to be designed such that they have similar target affinity, similar melting temperatures whilst retaining specificity and not binding to one another (Elnifro *et al.*, 2000). This can limit the number of targets and thus the ability to fully characterise the pathogen diversity in a sample. There are also other limitations of (q)PCR-based

methods. Samples can contain PCR inhibitors that can decrease detection of the organisms of interest, and the limit of detection can be an issue for samples with low pathogen abundance (Park *et al.*, 2014). This has been shown to be the case in a study by Lanzl *et al*. (2022), whereby 40 hours of enrichment in Bolton broth was required for qPCR detection of *Campylobacter* in all of the retail chicken samples tested, as the initial abundance of *Campylobacter* was below the limit of detection. This complicates detection in many cases, as PCR-based detection after enrichment can be affected by inhibitors in blood that is often added to enrichment broths (Josefsen *et al.*, 2004).

There are a number of molecular subtyping methods that allow differentiation beyond the species level. Methods such as pulsed-field gel electrophoresis (PFGE) can be used to separate large genomic DNA fragments using an electrical current after cleavage with restriction enzymes, which creates distinct patterns on agarose gel (Neoh *et al.*, 2019). Single locus sequence typing and multi-locus sequence typing (MLST) can also be performed. These methods are often also based on PCR for amplification of the gene(s) of interest, followed by either gel-based separation of products or sequencing. The flagellin gene, *flaA*, has been classically used for *Campylobacter* single locus typing. In *flaA*-restriction fragment length polymorphism (RFLP) analysis, a PCR-amplified product is digested by restriction enzymes and the fragments separated with gel electrophoresis, followed by Southern blotting to determine fragment length. The genetic diversity can then be determined by comparing lengths of fragments (Yadav *et al.*, 2018). In the sequencing-based approach, a short variable region of the PCR product is sequenced, providing higher resolution (Taboada *et al.*, 2013). However, the usefulness of this approach may be reduced by the inherent instability of *flaA* genes due to commonly occurring DNA transfer between *Campylobacter* species, and presence of the same alleles in multiple species (de Boer *et al.*, 2002; Dingle *et al.*, 2005).

Such issues are partly overcome by looking at multiple loci, hence an MLST scheme was developed for *C. jejuni* and *C. coli*, followed by schemes for other *Campylobacter* species (Dingle *et al.*, 2001; Miller *et al.*, 2005; Jolley, Bray and Maiden, 2018). As with single locus sequence typing, the procedure classically entailed amplification of genes with PCR and sequencing (Taboada *et al.*, 2013). This approach clusters strains based on sequence similarity of seven housekeeping *Campylobacter* genes and assigns a sequence type (ST) for each seven-allele combination (Dingle *et al.*, 2001); if STs have four or more identical alleles, they belong to the same clonal complex (CC). This approach has been widely applied to characterise *Campylobacter* isolates from various sources including retail chicken (Tedersoo *et al.*, 2022), and has facilitated source attribution of human campylobacteriosis cases (Cody *et al.*, 2019). MLST can be considered a reductionistic approach, as differences elsewhere in the genome may be ignored. Such limitations make it difficult to establish infection source, though it is still a useful method for

determining general population structure. More recent developments in the scheme, including core genome MLST (cgMLST) and whole genome MLST (wgMLST), facilitated by whole genome sequencing (WGS) approaches, have increased the utility and discriminatory power of *Campylobacter* typing methods for more detailed strain typing (Cody *et al.*, 2013, 2017).

1.2.5 Whole genome sequencing of *Campylobacter* isolates

WGS technologies have evolved in recent decades; improvements in computational power and automation have paved the way for high-throughput sequencing (HTS), allowing rapid simultaneous sequencing of large quantities of DNA sequences (Heather and Chain, 2016). Short read HTS methods utilise an input of a library of cleaved sequences with adapters attached that are amplified prior to loading onto the instrument flow cell and sequenced as single or paired reads of up to 400 base pairs (bp) (Satam *et al.*, 2023). These fragments can then be computationally assembled or aligned into contiguous sequences (Mardis, 2017). Popular short-read sequencing technologies include the Illumina MiSeq and NextSeq (Slatko, Gardner and Ausubel, 2018) and Life Technologies Ion Torrent (Buermans and den Dunnen, 2014). Although short-read sequencing platforms feature low error rates (Kchouk, Gibrat and Elloumi, 2017), the short length of the reads generated can be insufficient for resolving repetitive or variable regions (Adewale, 2020). Long read sequencing, sometimes termed third-generation sequencing, can generate reads of up to 30,000 bp without amplification, allowing resolution of variable and repetitive regions and making assembly more straightforward, which can allow generation of closed chromosome and plasmid sequences (Taylor *et al.*, 2019). Oxford Nanopore (ONT) and Pacific Biosciences (PacBio) technologies are currently the leading long read sequencing providers. However, the lower precision associated with long-read sequencing approaches, particularly in the case of ONT, often requires short read data to polish genome assemblies; alternatively, algorithms developed for detecting and correcting errors based on the long read sequencing data alone can be used (Morisse *et al.*, 2021). Recent developments in long read sequencing approaches have focused on improving accuracy; for example, the PacBio circular consensus sequencing approach can correct reads based on repeated template passing and allow variant calling (Wenger *et al.*, 2019). Similarly, recent updates to the Nanopore sequencing platforms and sequencing kits have focused on accuracy of sequencing output (Espinosa *et al.*, 2024).

WGS is becoming less expensive and thus more extensively used for genotyping purposes (Adewale, 2020; Uelze *et al.*, 2020). It can facilitate *Campylobacter* subtyping by extracting the MLST alleles of interest, for instance, though beyond this, WGS enables strain differentiation due to the high resolution provided by sequencing of the whole genome (Quainoo *et al.*, 2017). Genomic bioinformatics analysis methods can be used to infer evolutionary history and determine

Campylobacter diversity, for example by quantifying single nucleotide polymorphisms (SNPs) between genomes within lineages. Additional information on characteristics such as antimicrobial resistance (AMR) genotypes can also be inferred, which may be important in clinical applications, or to monitor resistance prevalence amongst pathogens in major infection sources (Rantsiou *et al.*, 2018). WGS is therefore useful for pathogen monitoring and the determination of infection sources in campylobacteriosis outbreaks, as demonstrated in large outbreak investigation studies (Joensen *et al.*, 2020, 2021).

1.2.6 Direct sequencing for *Campylobacter* detection

While useful, WGS requires prior *Campylobacter* isolation, which can be difficult and time-consuming, with the added risk of some species and subtypes being missed (1.2.2). Culture-independent sequencing approaches may thus be a more optimal detection and characterisation method for *Campylobacter*. This may also allow investigations of the pathogenomics of emerging *Campylobacter* species that have differing culture requirements, particularly as currently the amount of WGS data available for emerging species is limited.

1.2.6.1 Amplicon sequencing – metabarcoding

Pathogen detection can be achieved by sequencing DNA directly from samples of origin (Rodríguez-Valera, 2004). One widely applied method is amplicon sequencing or metabarcoding, which relies on PCR to target and amplify fragments of conserved genes in a group of organisms. Common targets include the 16S rRNA gene for detection of bacteria and archaea, the 18S gene for eukaryotes, and internal transcribed spacer rRNA for fungi (Y.-X. Liu *et al.*, 2021). For 16S rRNA sequencing, PCR primers can target a conserved region situated upstream of a variable region that is then amplified either on its own or in conjunction with typically up to two additional adjacent variable regions (Fuks *et al.*, 2018). The amplicons can then be short read sequenced, which is relatively cost effective and high throughput (Johnson *et al.*, 2019). The sequences of the variable fragments are then clustered into operational taxonomic units (OTUs) to differentiate organisms present using bioinformatics tools (Y.-X. Liu *et al.*, 2021).

Amplicon sequencing for bacterial pathogen identification can be advantageous when dealing with complex samples comprised of both target and non-target (i.e. host) DNA, due to the selective amplification of fragments of genes specific to bacteria (Y.-X. Liu *et al.*, 2021). This can be particularly useful for the analysis of samples of human or animal origin, such as clinical samples and retail meat that can contain high proportions of host DNA (Shi *et al.*, 2022; Bloomfield *et al.*, 2023). The direct sequencing of amplicons from samples can also remove the requirement for

culture-based identification of pathogens that is often time consuming; and it may be particularly useful for the detection of organisms that can be difficult to isolate, like *Campylobacter*. Indeed, such approaches have been shown to offer improved identification of fastidious bacteria compared to culture methods (Muhamad Rizal *et al.*, 2020).

However, the resolution offered by amplicon approaches like 16S rRNA sequencing may not be sufficient for epidemiological applications. Generally, short read amplicon sequencing may offer genus level classification at best (Johnson *et al.*, 2019), which is suboptimal in outbreak investigation scenarios that require strain-level classification, or in surveillance. Although clustering to higher sequence similarity, or denoising using pipelines such as DADA2 can potentially improve resolution (Callahan, McMurdie and Holmes, 2017), it has been argued that targeting of one to three variable 16S rRNA regions may not be sufficient to distinguish closely related organisms (Johnson *et al.*, 2019). There are also concerns around bias of PCR primers used to amplify 16S rRNA gene regions, as although these aim to target conserved regions of the gene, the target regions may not be truly universal and thus potentially not suitable for the detection of all bacteria (Martinez-Porchas *et al.*, 2017). The particular variable regions chosen for amplification can also introduce bias (Kumar *et al.*, 2011), which can lead to underrepresentation of sequences of interest, and thus reduction in sensitivity for pathogens of interest. In fact, previous research has highlighted an underrepresentation of a number of bacterial phyla in 16S rRNA amplicon sequences of chicken caeca and crop samples (Durazzi *et al.*, 2021). Identification of low abundance organisms may also be difficult (Durazzi *et al.*, 2021), thus making the identification of *Campylobacter* on retail meat challenging (Lanzl *et al.*, 2022).

Some of these issues can be potentially addressed with long read amplicon sequencing using ONT or PacBio, which facilitate sequencing of the whole target gene or even the entire operon. This can result in increased classification resolution and reduced bias in identified taxa (Johnson *et al.*, 2019). Although recent improvements have increased accuracy (Karst *et al.*, 2021), they require more thorough evaluation in regards to application to complex samples before they can be widely implemented (Walsh *et al.*, 2024). Furthermore, a single target approach like 16S rRNA amplicon sequencing may not provide additional functional information about an identified pathogen that may be epidemiologically relevant (Y.-X. Liu *et al.*, 2021). An alternative approach that can potentially overcome issues around the identification of low abundance organisms, amplicon bias and resolution whilst also potentiating functional inference is shotgun metagenomic sequencing.

1.2.6.2 Shotgun metagenomics

Shotgun metagenomics is an untargeted sequencing approach, potentiating study of the entire genetic content within a sample. The potential of shotgun metagenomics for pathogen detection

is also being explored to identify organisms associated with central nervous system, bloodstream and respiratory infections, with potential clinical applications (d'Humières *et al.*, 2021). In fact, a clinical pilot study of metagenomic respiratory infection diagnostics in intensive care units is currently underway in the UK following a successful trial at Guy's and St. Thomas' hospital (Charalampous *et al.*, 2024). Some previous work has indicated that metagenomics could be used to infer microbial causes of gastrointestinal infection, potentially allowing characterisation of outbreak strains (Loman *et al.*, 2013; Zhou *et al.*, 2016; Royer *et al.*, 2024), though many studies highlight the concurrent requirement of isolate or WGS data to confirm pathogen presence, and the diagnostic uncertainty related to the abundance of different organisms in stool samples calls for more detailed evaluation (Fourgeaud *et al.*, 2024). Outside of the clinic, shotgun metagenomics could potentially be used in food surveillance for the detection of pathogens (Grützke *et al.*, 2019), circumventing the need to target specific organisms individually, but the efficacy of this needs to be further investigated (Ko, Chng and Nagarajan, 2022).

Because shotgun metagenomics utilises an untargeted sequencing approach, often a large amount of the sequenced DNA is derived from the host (Li *et al.*, 2020; Peterson *et al.*, 2022). This can have implications on the identification of target organisms, as the coverage of the organisms of interest is reduced (Pereira-Marques *et al.*, 2019). As a result, a number of host depletion methods have been developed to increase the resolution of bacterial genome sequences, which often exploit the differences between bacterial and host cells (Shi *et al.*, 2022). For human DNA depletion prior to DNA extraction, lysis of host cells and degradation of the released DNA can be achieved with treatment with a lysis buffer or detergent and addition of DNase (Charalampous *et al.*, 2019; Shi *et al.*, 2022). For samples with low microbial concentrations, commercial kits have been developed that allow amplification of the microbial genetic material (Shi *et al.*, 2022). Host DNA can also be captured based on methylation patterns for exclusion after DNA extraction, or microbial DNA selected using methyl-directed restriction nucleases bound to magnetic beads that recognise motifs present in bacterial DNA (Shi *et al.*, 2022). Host depletion protocols have also been coupled with DNA extraction to produce commercial kits that allow concurrent host DNA depletion and isolation of microbial DNA for sequencing, for example by coupling host cell lysis with enzymes specific for microbial DNA capture (Shi *et al.*, 2022). Host depletion protocols for food samples have not been as widely applied, however recent developments have shown effective depletion of animal and plant DNA from food samples by coupling detergent-based host cell lysis, DNase degradation of the released DNA and protease treatment to remove free proteins (Bloomfield *et al.*, 2023). This facilitates greater coverage of microorganisms in the food sample, and may potentially allow pathogen detection.

DNA extraction protocols are an important consideration for all direct sequencing approaches but particularly shotgun sequencing. For a truly untargeted approach, the method should allow extraction of different organisms within the sample. In reality, different extraction protocols can be better or worse at obtaining DNA from specific organisms (Wesolowska-Andersen *et al.*, 2014). A number of commercial kits have been developed to improve the breadth of genomic content that can be extracted, utilising methods such as solid phase extraction or chemical or mechanical lysis, though this latter method can result in large amounts of fragmented DNA (Billington, Kingsbury and Rivas, 2022). This can in turn lead to a reduction in the quality of sequences produced, which can impact the power to identify organisms of interest, thus the selection of an appropriate extraction kit is important prior to metagenomics-based pathogen identification. For *C. jejuni* identification from faecal samples, MO Bio PowerLyzer PowerSoil DNA extraction kit extraction has been shown to recover the highest overall DNA yield, but extraction with MP Biomedicals FastDNA Spin kit for soil achieved optimal *C. jejuni* PCR amplification (Josefsen *et al.*, 2015). The efficacy of these kits for metagenomic *Campylobacter* detection from food samples has not been evaluated. Importantly, for surveillance purposes, one may wish to opt for DNA extraction kits that capture the full diversity of the sample to be able to identify different pathogens from the sample. Amongst DNA extraction kits tailored to microbial DNA extraction from food samples in the context of food microbiome studies, the Macherey-Nagel Nucleospin Food, Quiagen DNeasy Blood and Tissue, Zymo HostZERO Microbial DNA kit and Promega Maxwell PureFood Pathogen kits have been shown to produce high quality DNA for metagenomic sequencing (Buytaers *et al.*, 2020; Bloomfield *et al.*, 2023).

As with amplicon sequencing and WGS, both short and long read solutions are available for shotgun metagenomics. Short read sequencing, particularly using Illumina technology, has been a popular choice due to reducing costs and low error rates (Illumina, 2024), though long read sequencing technologies for shotgun metagenomics have improved in recent years, and are thus becoming more commonly applied (Kim, Pongpanich and Porntaveetus, 2024). A major drawback is the relatively high DNA concentration and quality that are required for long read sequencing (Trigodet *et al.*, 2022). This can be a particular challenge for sequencing food samples, which after host depletion may display very low DNA concentrations unless the microbial DNA is somewhat enriched or the protocol optimised (Bloomfield *et al.*, 2023). As a result, although long read sequencing has high potential for pathogen surveillance, the lower DNA concentration requirement and high throughput of short read sequencing may make this approach more useful for such applications.

Following sequencing, metagenomic reads can be classified using taxonomic mapping tools, which allow profiling of the sample microbiome and potentially identification of specific organisms present. Metagenome reads can be aligned to a custom set of reference genomes to identify and

locally assemble specific sequences, or alternatively bioinformatics tools can be used to estimate taxonomic profiles and relative abundance of taxa using extensive databases of either sequences or marker genes (Quince *et al.*, 2017). This can potentiate the identification of specific organisms in samples up to the strain level (D. Kim *et al.*, 2016), which may be highly useful in surveillance studies. A *Campylobacter* spike-in study showed that a concentration as low as 1 x 10^3 CFU/g could be detected in chicken faeces (Andersen *et al.*, 2017), whereas another study was able to identify as few as 200 CFU of *Campylobacter* from air samples (Haverkamp *et al.*, 2021). The sample type may affect the limit of detection, and there is currently a lack of consensus on the efficacy of metagenomics for *Campylobacter* detection on retail chicken meat.

Reads can also be assembled into longer contiguous sequences, or contigs, to form assemblies. Contig assembly is most often performed either by overlap layout consensus or De Bruijn Graph based algorithms, by either finding overlaps between reads, performing a layout and identifying a consensus, or by separating the reads into k-mers to find overlaps and form a graph and infer connections (Quince *et al.*, 2017). Contigs can then be clustered to construct metagenome assembled genomes (MAGs), with the aim to reconstruct individual organism genomes. This may be particularly useful for characterisation of pathogens, and for finding previously uncharacterised species or strains thanks to the lack of reliance on reference databases (Quince *et al.*, 2017). Binning algorithms to construct MAGs rely on k-mer and tetranucleotide frequency and read depth (Quince *et al.*, 2017). These MAGs can then be subjected to more detailed study to determine their taxonomy and gene content, thus in theory facilitating genome-scale investigations, which can be a useful approach for organisms that are difficult to isolate from food like *Campylobacter*. Metagenomics may additionally facilitate the identification of viable but nonculturable organisms, which may be a common feature of organisms on food (Fakruddin, Mannan and Andrews, 2013; Ayrapetyan and Oliver, 2016).

As metagenomics is still an evolving field, there are a number of limitations. A major one of these is the lack of standardised analysis pipelines for metagenomes (Yap *et al.*, 2022). The existing tools are also subject to limitations, such as the risk of false positive hits when using read-based taxonomic classification tools in the presence of low abundance organisms, which can arise due to the presence of common sequences between different organisms (Doster *et al.*, 2019; Ye *et al.*, 2019) or contaminants within reference genomes (Mukherjee *et al.*, 2015). Assembly approaches, on the other hand, can lead to the underrepresentation of low abundance organisms, and assemblies can be fragmented or mixed (Yorki *et al.*, 2023), with subsequent effects on MAG construction. MAGs are built based on consensus, which can lead to the underrepresentation of diversity of individual organisms (Royer *et al.*, 2024) and plasmids can be missed (Maguire *et al.*, 2020).

There has been limited metagenomic sequencing effort for the investigation of the presence and ecology of *Campylobacter* on retail chicken, despite previous studies indicating the possibility of the identification and characterisation of pathogens from foods (Yap *et al.*, 2022). By combining traditional detection methods with shotgun metagenomics of retail chicken samples, the feasibility of shotgun sequencing for pathogen identification and characterisation can be investigated, in order to determine whether or not it can be used in place of current standard methods and provide sufficient resolution for ecological investigation.

1.3 *Campylobacter* ecology

1.3.1 Genetic drivers of diversity and niche adaptation

Diversity in *Campylobacter* is driven by a number of mechanisms. *Campylobacter* spp. are phase variable, harbouring homopolymeric tracts within reading frames (ORFs) or promoter regions of genes encoding external structures or enzymes that enable alteration of morphology in unfavourable environments, or enhanced colonisation and virulence. Insertions or deletions in mononucleotide polyG, PolyC, PolyA or PolyT repeats within ORFs result in on/off switching, whereas changes in the number of mononucleotides in promoter regions can alter the transcription levels (Cayrou *et al.*, 2021). In *C. jejuni*, the main mechanism allowing phase variation is slipped strand mispairing resulting from errors in DNA replication, and the changes introduced are maintained because of a lack of functional mismatch repair system (Gaasbeek *et al.*, 2009; Cayrou *et al.*, 2021). For example, the hypervariable region of the *flgR* gene affects motility of *C. jejuni* and is important for chicken colonisation in phase "on" state, whereas a phase "off" state in a proportion of the population may facilitate shedding and spread of the organism (Hendrixson, 2006).

Gene pool diversification can occur by random mutation over time, although it is believed that horizontal gene transfer (HGT) primarily drives *Campylobacter* lineage evolution (Golz and Stingl, 2021). HGT can occur via transformation, conjugation and transduction (Golz and Stingl, 2021).

Among these, transformation encompasses natural competence, whereby *Campylobacter* can uptake DNA from the environment and incorporate it into the genome mainly through homologous recombination. Although intragenomic recombination events have been documented, facilitating the diversification of individual genes (Harrington, Thomson-Carter and Carter, 1997), these are less common than recombination following transformation, which is thought to be to be the primary evolutionary force in *Campylobacter* (Golz and Stingl, 2021). This can lead to increased genetic exchange between different strains within the chicken host (Samarth and Kwon, 2020), and provide capability for enhanced survival in this niche, allowing

subsequent contamination of retail meat. Introgression events have also been documented, representing an extensive genetic exchange whereby a large amount of DNA moves from one species to another and can eventually lead to species convergence. This has been observed between *C. jejuni* and agriculture-associated *C. coli* clade 1, likely due to occupation of the same niche (Sheppard *et al.*, 2013). The effect of introgression is thought to be dependent on downstream persistence of the mosaic gene in the lineage, which is widespread in *C. coli* hybrids containing DNA from *C. jejuni*. The transfer also occurs in the opposite direction, from *C. coli* to *C. jejuni*, although persistence of such hybrids is less common (Sheppard *et al.*, 2011). Nonetheless, such events have been observed; a recent study in the United States identified MLST profiles suggestive of introgression events whereby two *C. jejuni* genomes obtained from retail chicken contained MLST alleles typically associated with *C. coli* MLST loci (Hull *et al.*, 2021), further highlighting that the agricultural niche, representing farm animals and their environment, may foster such large genetic exchange events in both directions.

Other forms of HGT include phage transduction, facilitating the introduction of genomic fragments in the chromosome, and conjugation, which facilitates diversification as well as survival in unfavourable environments, whereby genetic material is exchanged through direct contact with a donor cell (Golz and Stingl, 2021). This is the most common method for transfer of mobile genetic elements (MGEs), including transposons, integrons and plasmids, though it can also occur through transformation and transduction (San Millan and MacLean, 2017). Plasmids can carry AMR genes and virulence or colonisation factors, which can enhance survival and propagation within the host. However, the understanding of *Campylobacter* plasmids is still rather limited, with no typing scheme akin to other pathogen plasmids currently publicly available. Nonetheless, some existing work has tried to cluster *Campylobacter* plasmids into groups, identifying four general types of plasmids in *C. jejuni* and *C. coli*: the pTet, *C. coli* specific, pVir-like and small plasmids (Marasini *et al.*, 2018). Among these, the pTet plasmids are considered most prevalent (Marasini *et al.*, 2018). Although named *C. coli* specific plasmids due to their prevalence in *C. coli*, this group was also identified in a selection of *C. jejuni* genomes (Marasini *et al.*, 2018). Most of the studied plasmids are thought to be conjugative, as indicated by the presence of components of the type IV secretion system (T4SS) that facilitates plasmid movement, though the rate of this may be strain dependent (Golz and Stingl, 2021). Previous studies have also identified a number of megaplasmids (>80 kb) in isolates from retail meats (Marasini and Fakhr, 2014, 2016; Hull *et al.*, 2023). Some of these have been found to carry genes previously associated with interspecies recombination, as well as AMR genes and virulence factors that have been suggested to enhance pathogenicity or survival on chicken meat, such as toxin components and genes encoding proteins involved in adherence and chemotaxis (Hull *et al.*, 2023). The presence of different plasmids in both *C. jejuni* and *C. coli* species suggests that plasmid sharing may occur between the species

(Marasini *et al.*, 2018; Hull *et al.*, 2023), though the extent of this, as well as the extent of plasmid sharing between different strains within individual retail meat samples, has not been widely investigated. Some of the existing studies have utilised short read sequencing to study plasmid sequences, which can cause potential misclassifications. Long read characterisation of plasmid sequences from retail chicken may be necessary for full plasmid resolution and verification (Hull *et al.*, 2023).

Plasmid carriage of AMR genes indicates that AMR determinants may be readily spread among *Campylobacter* populations. Increases in antimicrobial resistant *Campylobacter* in the food chain are concerning, as although antimicrobial treatment is not typically recommended, it may be necessary in severe disease (Blaser and Engberg, 2008). Presence of AMR on MGEs also potentiates spread to other bacteria within the host, affecting future treatment outcome (Zilhao, Papadopoulou and Courvalin, 1988). Tetracycline resistance can be gained through the acquisition of the often plasmid-borne *tet(O)* gene, which forms part of the core genome of pTet plasmids, some of which also carry aminoglycoside resistance genes (Marasini *et al.*, 2018; Abraham *et al.*, 2020). Nonetheless, AMR in *Campylobacter* is generally considered to be synergistic and usually both chromosomally- and plasmid-encoded (Iovine, 2013). Where treatment is required, patients are usually prescribed fluoroquinolone or macrolide antibiotics (Blaser and Engberg, 2008). Quinolone and fluoroquinolone resistance in *Campylobacter* is achieved by mutation to the *gyrA* gene, most often a T86I substitution (Marotta *et al.*, 2019). Rates of macrolide resistance are usually lower than fluoroquinolone resistance, as the ribosomal subunit containing 23S rRNA encoded by genes responsible for this phenotype is associated with a low mutation rate (Iovine, 2013), making this the preferred treatment. However, macrolide resistance can be gained by HGT of *erm(b)* genes (Mourkas *et al.*, 2019) or efflux (Iovine, 2013) that can also facilitate resistance to other drug groups. The plasmid-borne CmeABC efflux pump can facilitate multidrug resistance (MDR) (Guo *et al.*, 2008), which is usually defined as resistance to three or more antimicrobial classes (Magiorakos *et al.*, 2012), when activated (Guo *et al.*, 2008). Active efflux, alongside reduced permeability (Iovine, 2013), also makes *Campylobacter* intrinsically resistant to most beta-lactams, though beta-lactamase genes are often found in *Campylobacter* genomes, particularly in the case of chicken-derived isolates (Kramer *et al.*, 2000).

There are several barriers to HGT that affect the level of genetic exchange, including biogeography, homology dependence and fitness of the new genotype (Golz and Stingl, 2021). Nonetheless, existing evidence shows that HGT facilitates a significant increase in diversity. Movement of large genetic segments, for example during introgression, may also have an impact on genotyping, as many current methods rely on the amplification of specific gene fragments that are common to a species (Golz *et al.*, 2020). This further highlights the value of sequencing methods for that facilitate detailed studies of pathogen populations whilst accounting for

recombination events, and metagenomic methods for direct detection that don't rely on specific markers potentially affected by genetic exchange events. These genetic drivers of diversity are reflected in the *Campylobacter* populations observed in the chicken host, as discussed in the following section.

1.3.2 *Campylobacter* diversity in chickens and on chicken meat

C. jejuni and *C. coli* are the two most common species identified in chickens, and consequently on chicken meat (Bull *et al.*, 2006; Jorgensen *et al.*, 2019). However, other species linked to human disease including *Campylobacter upsaliensis* (Kuana *et al.*, 2008), *Campylobacter fetus* (Kuana *et al.*, 2008; Sinulingga *et al.*, 2020), *Campylobacter hyointestinalis* (Zhang *et al.*, 2020), *Campylobacter concisus* (Lynch *et al.*, 2011), *Campylobacter mucosalis* (Figura *et al.*, 1993; Lynch *et al.*, 2011), *Campylobacter lanienae* (Acik *et al.*, 2013; Fornefett *et al.*, 2021) and *Campylobacter lari* (Sheppard, Dallas, *et al.*, 2010; Acik *et al.*, 2013) have also been identified on poultry sources (Figure 1.2). There are also a number of other species present in chickens that have not thus far been identified as pathogenic for humans, including *Campylobacter hepaticus* (Van *et al.*, 2016), *Campylobacter bilis* (Van *et al.*, 2023) and *Campylobacter avium* (Rossi *et al.*, 2009). As most current identification methods are tailored to the growth requirements of *C. jejuni* and *C. coli*, it is more difficult to isolate and study the less-common species for which there is also a shortage of genome sequencing data (Costa and Iraola, 2019). Although subtyping schemes, including MLST, exist for non-*jejuni* and *coli* species, the relatively small amount of data available impedes studies of genetic diversity of emerging species from chicken. Most genetic diversity studies to date have thus focussed *C. jejuni* and *C. coli*. *C. jejuni* can be further split into two subspecies, *C. jejuni* subsp. *jejuni* and *C. jejuni* subsp. *doylei* (Parker *et al.*, 2007). Most evidence suggests that *C. jejuni* subsp. *doylei* is phylogenetically distinct from *C. jejuni* subsp. *jejuni*, may be linked to systemic disease, and represents a relatively small proportion of isolates obtained from livestock carcasses and chicken meat (2%) (Flynn, Blair and Mcdowell, 1994; Parker *et al.*, 2007; Karikari *et al.*, 2017). However, a study by Kaakoush *et al*. (2014) identified the subspecies in 61.3% of chicken faecal samples using direct HTS, which suggests that culture method limitations could be responsible for the lack of identification of certain *Campylobacter* subspecies, further highlighting the advantages of nonselective sequencing approaches.

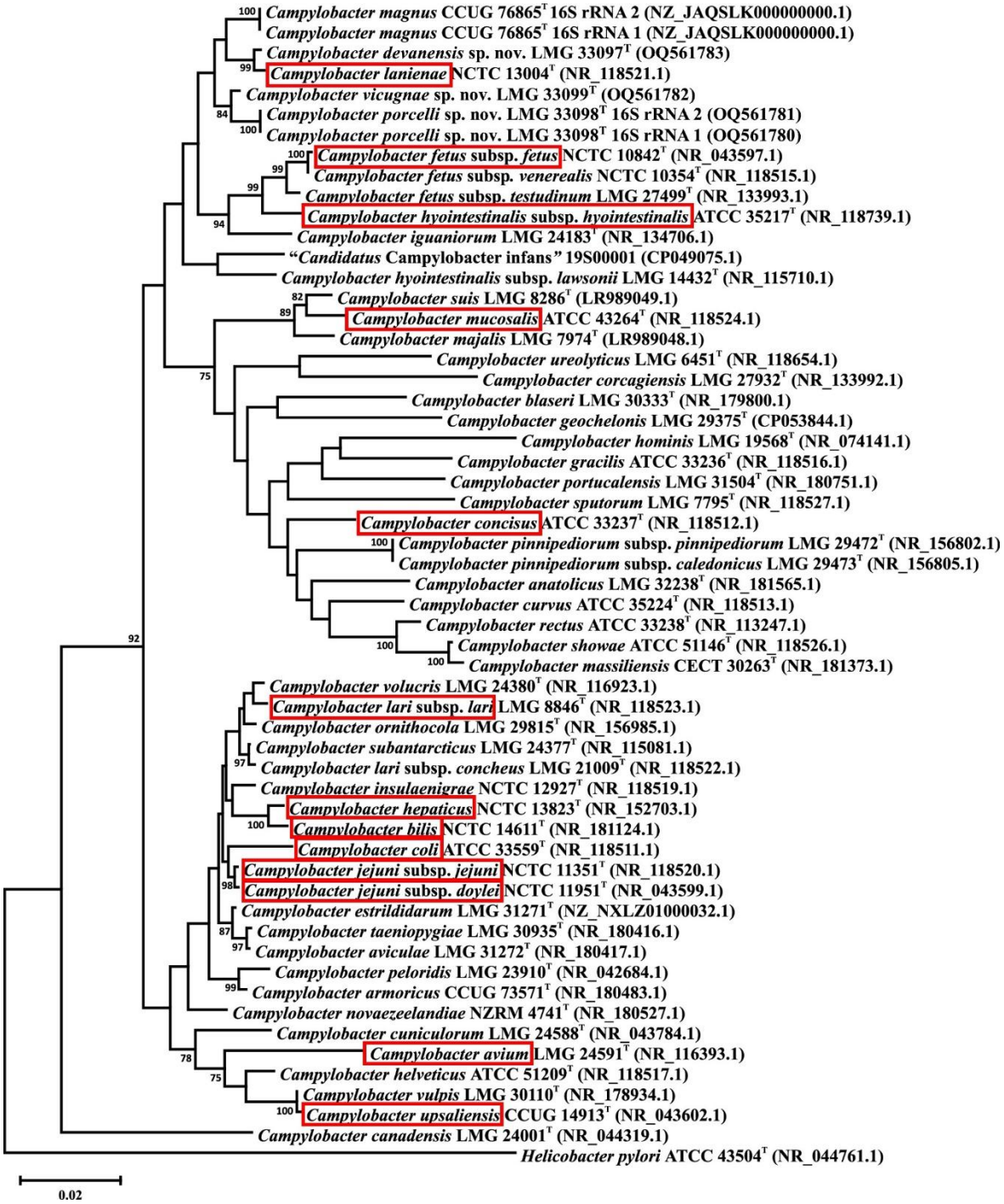

Figure 1.2: Phylogenetic tree based on 16S rRNA genes of representative strains within the *Campylobacter* genus using *Helicobacter pylori* as an outgroup; red rectangles represent species that have been isolated from chicken. Adapted from Miller *et al*. (2024)

Chickens become colonised with *Campylobacter* by two weeks after hatching, following a reduction in maternal antibodies (Hendrixson and DiRita, 2004; Rawson, Dawkins and Bonsall, 2019). Despite the passing of antibodies to the chicks being suggestive of the production of an immune response, investigations have indicated that *Campylobacter* colonisation in the chicken gut does not result in disease (Burnham and Hendrixson, 2018). However, it has been shown that *Campylobacter* introduction does elicit an innate immune response, and modern broiler breeds

may suffer from active inflammation due to lack of subsequent regulatory control mediated through Treg cell action (Humphrey *et al.*, 2014). An adaptive response also occurs, marked by production of antibodies, though these have limited role in clearance, leading to persistence in the gut lumen (Gilroy *et al.*, 2024). Carriage rates in a flock can be up to 100% (Stern *et al.*, 1995), indicating fast spread that is thought to most likely occur via horizontal transmission and the faecal-oral route (Adkin *et al.*, 2006). The diversity of *Campylobacter* observed in flocks can be complex as different strains can be introduced and spread over time (Colles and Maiden, 2012), and individual strains can display differing colonisation dynamics and ecologies (Chaloner *et al.*, 2014). Although a small number of dominant strains is typically found against a background of other lower abundance strains in individual birds, the removal of dominating strains can provide opportunity for others to expand (Rawson, Dawkins and Bonsall, 2019). This complexity, along with limited antibody-facilitated clearance (Lacharme-Lora *et al.*, 2017), are major reasons why an effective vaccine against *Campylobacter* in poultry is yet to be developed.

The seven-locus MLST scheme for *C. jejuni* and *C. coli* typing has been useful for determining the population structure of these prevalent species in the poultry niche. As of July 2024, the PubMLST isolate collection contains 4,631 assigned, unambiguous *C. jejuni* and *C. coli* STs and 41 recognised CCs for isolates obtained from the chicken or chicken meat/offal source. ST-828, ST-45 and ST-21 CCs have been shown to contain some of the most prevalent *Campylobacter* STs in chickens and on chicken meat as well as in clinical samples, demonstrating the role of poultry in infection (Sheppard, Colles, *et al.*, 2010; Colles and Maiden, 2012; Guyard-Nicodème *et al.*, 2015; Zhang *et al.*, 2020), though the identification of these CCs from a number of other sources including geese and starlings indicates that they are host generalists, which makes it difficult to attribute infection sources based on MLST typing alone (Colles *et al.*, 2008; Dearlove *et al.*, 2016). However, a number of CCs are considered to be more specific to chicken hosts, including the CCs of ST-353, ST-354, ST-443, ST-257 and ST-464 (Epping *et al.*, 2021). *C. coli* isolated from agricultural environments are considered to be generally less diverse than *C. jejuni*, with the generalist ST-828 CC representing the majority of isolates (Sheppard, Dallas, *et al.*, 2010).

The *Campylobacter* subtypes found in chickens and their environment are important as they may later be found on retail chicken products. A number of retail chicken meat surveys have investigated *Campylobacter* ST and CC diversity, in attempt to determine the strains found at the end of the farm to fork continuum, the point closest to the consumer. One study investigating *Campylobacter* on UK chicken meat from three surveillance studies between 2001-2006 found 194 STs, with ST-257, ST-45, ST-827, ST-51, ST-21 and ST-573 most prevalent (Wimalarathna *et al.*, 2013). A similarly high diversity and some of the same STs were observed in the most recent FSA report of *Campylobacter* obtained from non-major retailer derived retail chicken (2018-2020), with ST-5136, ST-50, ST-354, ST-6175, ST-21, ST-51, ST-573, ST-122, ST-48, ST-2066 and ST-828,

ST-825, ST-1595, ST855 being the most prevalent *C. jejuni* and *C. coli* STs, respectively, out of a total 135 STs identified (Jorgensen *et al.*, 2021). Alongside known STs, previously undescribed STs are routinely identified (Jorgensen *et al.*, 2021), highlighting the high diversity of *Campylobacter* in this niche.

Although MLST is a useful scheme for identifying closely related isolates, individual STs can also harbour substantial diversity (Epping, Antão and Semmler, 2021). Individual STs can vary by thousands of SNPs (Ghielmetti *et al.*, 2023) due to the relatively low resolution provided by looking at only seven genes through the MLST scheme. For example, a study of isolates from different countries belonging to the commonly isolated ST-50 revealed up to 2,826 core genome SNPs, with highly variable SNP ranges within regions, indicative of both closely and distantly related sub-lineages (Wallace *et al.*, 2021). An investigation of a selection of STs from clinical samples found the generalist ST-45 to separate into three distinct clusters that were more genetically distinct from one another than the clusters of ST-230, ST-267 and ST-677 based on wgMLST analysis (Kovanen *et al.*, 2014), though another study found that certain sub-lineages of ST-45 have shown very little genetic variation despite differences in location and time of sampling (Llarena *et al.*, 2016; El-Adawy *et al.*, 2023), further highlighting the complex population structure of *Campylobacter*. This highlights the importance of WGS-based analyses for poultry surveillance and outbreak detection purposes.

It has been suggested that presence of *Campylobacter* in the caecum of birds prior to slaughter may be an important determinant of *Campylobacter* presence and diversity on the carcass after processing, largely due to contamination with faecal matter following slaughter (Prendergast *et al.*, 2022). There are a number of steps in the meat processing chain that can reduce *Campylobacter* loads throughout processing, including scalding and chilling (Rasschaert *et al.*, 2020). The abundance and diversity of *Campylobacter* can thus reduce as carcasses move through the abattoir processing points, as the conditions of the processing stages may provide a bottleneck for many *Campylobacter* strains (Mohamed, Williams and van Klink, 2021). Nonetheless, certain strains may be able to persist within meat processing establishments (García-Sánchez *et al.*, 2017). The processing environment can actually foster opportunities for cross-contamination between birds, which can be a result of direct contact between birds or the carcasses and slaughterhouse staff handling them, as well as contaminated equipment (Rasschaert *et al.*, 2020). In fact, there is evidence to suggest that there may actually be a large increase in *Campylobacter* prevalence, abundance and subtype diversity at abattoirs, which can persist through to retail (Colles *et al.*, 2010; Corry *et al.*, 2017; Würfel *et al.*, 2019; Inglis *et al.*, 2021; Faverjon, Cameron and De Nardi, 2022), indicating that cross-contamination events during meat processing can increase any existing diversity of individual birds and have a potentially significant impact on retail products to which the consumer is later exposed.

1.3.3 Sampling strategies and limitations in light of *Campylobacter* diversity

Standard isolation methods require picking of multiple isolates for confirmation of *Campylobacter* presence. However, rarely are all of these colonies typed, sequenced or stored. Feasibility constraints of many studies, including national surveillance programmes often mean that a very limited number of isolates, often as few as one, is taken or characterised from individual food samples (Food Standards Agency, 2016; Hull *et al.*, 2021). This may be suboptimal if individual samples are contaminated with multiple species or strains. There is some limited data indicating high within-sample *Campylobacter* diversity. For example, up to eight STs were identified on individual free-range chicken carcasses post-slaughter (Colles *et al.*, 2010). Another study of retail chicken products in Brazil found up to three PFGE pulsotypes per product (Würfel *et al.*, 2019). However, such information at the retail level in the UK is currently limited, thus it is unclear if this level of diversity persists through to retail.

The recovery or detailed investigation of only single isolates to represent individual chicken samples means within-sample diversity is rarely inferred, despite the evidence of cross contamination events during processing that can result in complex *Campylobacter* populations on individual meat samples. This can lead to underestimation of diversity, and may be part of the reason why most *Campylobacter* infections are considered to be sporadic (Liu *et al.*, 2022), as genotyping of a single isolate per sample may lead to epidemiologically relevant strains being missed during source attribution and outbreak investigation. Generation of more data is needed to understand *Campylobacter* diversity on epidemiologically relevant infection matrices, especially retail chicken meat due to its high association with infections.

The indication of high diversity means that a lot of culture and genome sequencing effort may be required, leading to concerns around feasibility of such efforts, particularly in time-sensitive situations such as outbreak investigations. However, the elucidation of the full sample diversity is important, otherwise outbreaks cannot be traced. Such issues could be potentially overcome with direct metagenomic sequencing, but the feasibility of metagenomics-based identification of diverse *Campylobacter* populations needs to be evaluated. Metagenomics may additionally allow study of other organisms within the poultry meat microbiome, some of which may be associated with *Campylobacter* survival, which could pave the way for novel intervention strategies.

1.4 Other organisms and association with *Campylobacter* presence

1.4.1 Poultry colonisation and effect of the chicken gut microbiota

Campylobacter colonisation efficiency can be linked to the chicken gut microbiome. Mathematical models have been applied to study the dynamics of *Campylobacter* within a broiler flock to determine which factors have the highest effect on persistence, with findings indicating that growth and death rates of other organisms present in the flora have the greatest influence (Rawson, Dawkins and Bonsall, 2019). A study investigating changes associated with *C. jejuni* colonisation in caecal samples found that *C. jejuni* colonisation had no significant effect on caecal alpha diversity (within-sample diversity), but the beta diversity (between-sample diversity) was affected (Thibodeau *et al.*, 2015). This inspires consideration of which organisms are responsible for potentially aiding *Campylobacter* survival, and which are potentially antagonistic to its survival. A number of studies have investigated associations between *Campylobacter* presence or abundance and the presence or abundance of other taxa in chicken caeca or faecal samples, as well as using *in vitro* and *in vivo* models (Table 1.1).

Table 1.1: Associations identified between *Campylobacter* and other taxa identified in chicken faecal or caecal samples, *in vitro* and *in vivo* models*

Taxon	Association	Reference
Firmicutes (Bacillota)	Higher abundance in absence of *Campylobacter*	Sofka *et al*. (2015)[a]
	Higher abundance associated with increased *Campylobacter* abundance/*C. jejuni* presence	Connerton *et al*. (2018); Sakaridis *et al*. (2018)[a,b]
Unclassified Clostridia/ Clostridiales/Clostridiaceae	Both higher and lower abundance of different OTUs observed in presence of *C. jejuni*	Thibodeau *et al*. (2015); Connerton *et al*. (2018)[b]
Mogibacteriaceae	Higher abundance in presence of *C. jejuni*	Thibodeau *et al*. (2015)[b]
Christensenellaceae	Lower abundance in presence of *C. jejuni*	Thibodeau *et al*. (2015)[b]
Defluviitaleaceae genus	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Eubacteriaceae	Higher abundance in presence of *C. jejuni*	Connerton *et al*. (2018)[b]
Lachnospiraceae (classified and unclassified)	Higher abundance in presence of *C. jejuni* or lower abundance in absence of *Campylobacter*	Thibodeau *et al*. (2015); Connerton *et al*. (2018); Hertogs *et al*. (2021)[b]
	Higher abundance in absence of *C. jejuni*	Connerton *et al*. (2018)[b]
Fusicatenibacter	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Anaerostipes, Anaerostipes butyraticus	Higher abundance in absence of *C. jejuni*	Connerton *et al*. (2018)[b]
Butyricicoccus	Higher abundance in presence of *C. jejuni*	Connerton *et al*. (2018)[b]
Blautia	Higher abundance in presence of *C. jejuni*	Kaakoush *et al*. (2014); Connerton *et al*. (2018)[b]
	Lower abundance in presence of *C. jejuni*	Thibodeau *et al*. (2015)[b]
Ruminococcaceae	Higher abundance in presence of *C. jejuni*	Kaakoush *et al*. (2014); Connerton *et al*. (2018)[b]
	Higher abundance in absence of *Campylobacter/C. jejuni*	Connerton *et al*. (2018); Hertogs *et al*. (2021)[b]
Ruminococcus 1	Lower abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Ruminococcus gauveauii group	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Enterococcaceae	Higher abundance in absence of *C. jejuni*	Connerton *et al*. (2018)[b]
Enterococcus	Higher abundance in presence of *C. jejuni*	Kaakoush *et al*. (2014)
	Higher abundance in absence of *Campylobacter/C. jejuni*	Connerton *et al*. (2018); Hertogs *et al*. (2021)[b]
Streptococcus	Lower abundance in presence of *C. jejuni*	Thibodeau *et al*. (2015)[b]
	Higher abundance in presence of *C. jejuni* or lower in absence of *Campylobacter*	Kaakoush *et al*. (2014); Hertogs *et al*. (2021)

Table 1.1: Associations identified between *Campylobacter* and other taxa identified in chicken faecal or caecal samples, *in vitro* and *in vivo* models*

Taxon	Association	Reference
Faecalibacterium	Higher abundance in presence of *C. jejuni*	Kaakoush *et al*. (2014); Thibodeau *et al*. (2015)[b]
Faecalibacterium prausnitzii	OTUs co-associated with *Campylobacter* (Clstr97)	Oakley *et al*. (2013)[b]
Oscillibacter	Higher abundance in absence of *C. jejuni*	Connerton *et al*. (2018)[b]
Anaerofustis	Higher abundance in presence of *C. jejuni*	Connerton *et al*. (2018)[b]
Tyzzerella 3/Lachnoclostridium	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Clostridium	Higher abundance in presence of *Campylobacter*	Kaakoush *et al*. (2014); Thibodeau *et al*. (2015)[b]
Clostridium IV	Both higher and lower abundance of different OTUs observed in absence of *C. jejuni*	Connerton *et al*. (2018)[b]
Clostridium XIVa	Higher abundance in absence of *C. jejuni*	Connerton *et al*. (2018)[b]
Lactobacillus	Lower abundance in presence of *C. jejuni/*higher in absence of *C. jejuni*	Kaakoush *et al*. (2014); Thibodeau *et al*. (2015)[b]
Globicatella	Lower abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Megamonas	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Megamonas hypermegale	OTUs co-associated with *Campylobacter* (Clstr97)	Oakley *et al*. (2013)[b]
Coprobacillus	Lower abundance in presence of *C. jejuni*	Thibodeau *et al*. (2015)[b]
Aliicoccus	Lower abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Gallicola	Lower abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Anaerococcus	Lower abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Atopostipes	Lower abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Facklamia	Lower abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Aerococcus	Lower abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Marvinbryantia	Lower abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Subdoligranulum	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
	Higher abundance in presence of *Campylobacter*	Connerton *et al*. (2018)[b]
Erysipelatoclostridium	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Oceanobacillus	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Anaerofilum	Lower abundance in presence of *C. jejuni*	Thibodeau *et al*. (2015)[b]

Table 1.1: Associations identified between *Campylobacter* and other taxa identified in chicken faecal or caecal samples, *in vitro* and *in vivo* models*

Taxon	Association	Reference
Bacteroidetes (Bacteroidota)	Higher abundance in presence of *Campylobacter*	Sofka *et al*. (2015)[a]
	Lower abundance linked to increased *Campylobacter* abundance	Sakaridis *et al*. (2018)[a]
Alistipes	Higher abundance in presence of *C. jejuni*	Kaakoush *et al*. (2014)
	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Bacteroides	Higher abundance in presence of *C. jejuni*	Kaakoush *et al*. (2014)
Barnesiella	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Proteobacteria (Pseudomonadota)	Higher abundance in presence/higher abundance of *Campylobacter*	Sofka *et al*. (2015); Sakaridis *et al*. (2018)[a]
Epsilonproteobacteria/Campylobacterales/ Campylobacteraceae/*Campylobacter*	Higher abundance in presence of *C. jejuni*	Kaakoush *et al*. (2014); Connerton *et al*. (2018)[b]
Escherichia	Higher abundance in presence of *C. jejuni*	Kaakoush *et al*. (2014)
Shigella	Higher abundance in presence of *C. jejuni*	Kaakoush *et al*. (2014)
Gallibacterium	Higher abundance in presence of *C. jejuni*	Kaakoush *et al*. (2014)
Helicobacter	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Enterobacter	Higher abundance in presence of *C. jejuni*	Kaakoush *et al*. (2014)
Sutterella	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Tenericutes (Mycoplasmatota)	Lower abundance in presence of *C. jejuni*	Thibodeau *et al*. (2015)[b]
Mollicutes	Lower abundance in presence of *C. jejuni*	Thibodeau *et al*. (2015)[b]
Anaeroplasmatales, Anaeroplasmataceae	Lower abundance in presence of *C. jejuni*	Thibodeau *et al*. (2015)[b]
Anaeroplasma	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
	Lower abundance in presence of *C. jejuni*	Thibodeau *et al*. (2015)[b]
Corynebacterium	Lower abundance in presence of *C. jejuni*	Kaakoush *et al*. (2014)
Eggerthella	Higher abundance in absence of *C. jejuni*	Connerton *et al*. (2018)[b]
Dietzia	Lower abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Gordonibacter	Higher abundance in absence of *Campylobacter*	Hertogs *et al*. (2021)
Bifidobacterium	Associated with *Campylobacter* presence	Thibodeau *et al*. (2015)[b]
Bifidobacterium longum strain PCB 133	Administration of strain reduced *C. jejuni* abundance	Santini *et al*. (2010); Baffoni *et al*. (2012)

*Not an exhaustive list

[a]=difference not statistically significant/not formally tested in at least one of the studies; [b]=associations pertain to specific OTUs in at least one study

Both positive and negative associations have been indicated between *Campylobacter* and specific phyla or taxa within them, including the Firmicutes (Oakley *et al.*, 2013; Kaakoush *et al.*, 2014; Sofka *et al.*, 2015; Thibodeau *et al.*, 2015; Connerton *et al.*, 2018; Sakaridis *et al.*, 2018; Hertogs *et al.*, 2021), as well as Bacteroidetes (Kaakoush *et al.*, 2014; Sofka *et al.*, 2015; Sakaridis *et al.*, 2018; Hertogs *et al.*, 2021), Proteobacteria (Kaakoush *et al.*, 2014; Sofka *et al.*, 2015; Sakaridis *et al.*, 2018; Hertogs *et al.*, 2021), Tenericutes (Thibodeau *et al.*, 2015; Hertogs *et al.*, 2021), and to a smaller extent taxa within other phyla (Santini *et al.*, 2010; Baffoni *et al.*, 2012; Kaakoush *et al.*, 2014; Thibodeau *et al.*, 2015; Connerton *et al.*, 2018; Hertogs *et al.*, 2021). It is clear that there are contrasting results between studies at individual taxonomic levels, and there are also differences within individual studies. In most cases, the latter may be attributed to different conditions within the study or the fact that the associations relate to specific OTUs (Oakley *et al.*, 2013; Thibodeau *et al.*, 2015; Connerton *et al.*, 2018), thus indicating that different populations within individual taxonomic levels can have differing effects on *Campylobacter* colonisation or survival, or alternatively, *Campylobacter* presence may be affecting persistence of specific OTUs. Equally, some studies have noted that the associations may only be valid for specific *Campylobacter* subpopulations (Oakley *et al.*, 2013). It is also difficult to draw conclusions based on the data collated from different studies due to differences in study design (such as the use of antimicrobials or other additives in some studies) and sample types (Thibodeau *et al.*, 2015; Hertogs *et al.*, 2021; Marmion *et al.*, 2021). Many of the existing studies have utilised 16S rRNA amplicon sequencing (Oakley *et al.*, 2013; Kaakoush *et al.*, 2014; Sofka *et al.*, 2015; Thibodeau *et al.*, 2015; Connerton *et al.*, 2018; Sakaridis *et al.*, 2018; Hertogs *et al.*, 2021), limiting the resolution of classification. Potential differences in analysis methods between these studies, particularly relating to clustering algorithms to obtain OTUs and the statistical methods to infer associations, are additional confounders that reduce comparability. Despite this, it appears that microbiome changes related to *Campylobacter* presence and absence in the chicken gut can be identified, and approaches allowing higher resolution, such as shotgun sequencing, may provide more clarity on this.

The associations may not be entirely related to the presence of specific organisms themselves but rather to production of specific metabolites that can help or hinder *Campylobacter* survival and subsequent colonisation. There is evidence that *Campylobacter* may be able to utilise hydrogen produced by organisms within genera like *Clostridium* in oxidative phosphorylation (Kaakoush *et al.*, 2014; Stoakes *et al.*, 2022). Short chain fatty acids (SCFAs), such as acetate and butyrate also produced by members of the Clostridiales in the gut, can be utilised as an energy source (Kaakoush *et al.*, 2014), or increase the expression of genes facilitating amino acid transport; this is needed for biomass production and effective colonisation, particularly in the lower gut regions where the concentrations of these SCFAs are high (Luethy *et al.*, 2017). Conversely, lactate can

potentially repress the expression of these genes, thus production of lactate by organisms like Lactobacilli can reduce the establishment of *Campylobacter*. This may also explain why *Campylobacter* does not typically colonise upper gut regions, as although lactate producers have been found in all gut regions, they are most abundant in the crop and gizzard (Luethy *et al.*, 2017; Marmion *et al.*, 2021), though contrasting evidence suggests that lactate can be used as an energy source by *Campylobacter* (Masanta *et al.*, 2013). Propionic acid, produced by genera such as *Megamonas* (Polansky *et al.*, 2016), has also been shown to have bactericidal effects on *Campylobacter* in hydrated poultry feed, though this was largely dependent on pH (Chaveerach *et al.*, 2002). The existing data indicates that interactions of bacteria and metabolites in the gut are complex and require further study.

1.4.2 *Campylobacter* and retail chicken meat microbiota

An important consideration is that the chicken meat microbiota differs considerably from chicken gut microbiota (Marmion *et al.*, 2021). Studies conducted to assess microbial diversity in gut regions and faeces are likely not representative of the skin and meat flora, making comparisons difficult. This suggests the possibility that there may be different microbial communities present on chicken meat that aid *Campylobacter* survival.

Propionic acid producing bacteria have been associated with *Campylobacter* absence in the chicken gut (1.4.1). Propionic acid has also been shown to inhibit *C. jejuni* on chicken meat (González-Fandos, Maya and Pérez-Arnedo, 2015), indicating that presence of this creates an unfavourable environment for *C. jejuni*, and thus the presence propionic acid producers on retail meat could negatively affect *Campylobacter* survival. Some investigations have incorporated carcass rinse and weep (rinse stored at 4°C for 48 hours) samples from processing plants and retail weep (exudate from retail products) samples alongside chicken caecal samples to identify associations with *Campylobacter* (Oakley *et al.*, 2013), finding certain *Faecalibacterium prausnitzii* and *Megamonas hypermegale* subtypes to co-occur with specific *Campylobacter* OTUs (Oakley *et al.*, 2013). However, there were differences in the microbiome of the different sample types (caecal, carcass and weep obtained from processing and retail weep), with these taxa being seemingly absent in retail meat samples; a limitation of this study was the low sample size at retail (n = 6), and pooling of samples from the different niches sampled, which can also invalidate investigation of inter-sample differences. Nonetheless, the study highlighted differences in the microbial composition between the different sample types. In the UK, retail chicken products are packaged in different gas mixtures, with the majority packed in atmospheric gas concentrations or increased oxygen (70-80%) and carbon dioxide (20-30%) concentrations, though this can vary between producers (Burfoot *et al.*, 2015). Chicken meat microbiomes may thus lack

representation of obligate anaerobic genera seen in gut microbiomes, such as *Clostridium*, instead showing a predominant abundance of organisms able to survive in modified atmosphere packaging and cold storage conditions such as *Pseudomonas, Aeromonas, Carnobacterium, Enterobacter, Serratia, Lactobacillus* and *Leucnostoc* (Li *et al.*, 2020; Marmion *et al.*, 2021; Bloomfield *et al.*, 2023). Although the microbial diversity of chicken meat may be vastly different to the chicken gut, some of these genera have been identified to potentially interact with *Campylobacter* in the chicken gut (Kaakoush *et al.*, 2014), thus leading to speculation that similar patterns could be observed on chicken meat. Equally, other organisms not identified to be associated with *Campylobacter* in the chicken gut may have influence on *Campylobacter* survival on retail meat.

Due to the unfavourable conditions on retail meat, *Campylobacter* may in part rely on biofilm formation for survival. *C. jejuni* biofilm formation in monoculture (Ica *et al.*, 2012) is increased by addition of meat exudate, which facilitates *Campylobacter* binding to abiotic surfaces like glass and stainless steel, also allowing survival in aerobic conditions (Brown *et al.*, 2014). However, a study by Teh *et al*. (2010) has shown that *Campylobacter* is also able to form biofilms with other bacteria, like *Staphylococcus simulans* and *Enterococcus faecalis*, though as the absence of *C. jejuni* did not affect biofilm formation by other organisms, it is likely that it is a secondary biofilm coloniser.

There is currently a lack of research looking at specific microbial communities associated with *Campylobacter* presence and absence on chicken meat that should be addressed, given its relevance as a leading source of *Campylobacter* infection. Identification of organisms that support or hinder *Campylobacter* survival could inform novel intervention strategies, as the addition or exclusion of such organisms could reduce *Campylobacter* prevalence on retail meat. Such investigations can be achieved with direct sequencing of retail chicken samples. It is important to note that the addition of microorganisms to food requires regulatory approval, unless the agents are already generally considered safe, according to guidelines from national and international bodies (Burdock and Carabin, 2004; European Food Safety Authority, 2007). Criteria determining the safety of microbial food additives require the evaluation of pathogenicity, genetic stability, toxicity and AMR, therefore reducing the viability of some potential candidates.

1.4.3 Culture-independent sequencing for microbial community profiling

Direct sequencing, and particularly metagenomic sequencing, can facilitate detailed study of the microbial composition of retail food samples, therefore allowing investigations of organisms associated with *Campylobacter* presence and absence on retail meat. Although still a developing

area, the power of culture-independent sequencing has been demonstrated in different contexts. Culture-independent sequencing has been applied to study differences in microbiomes of healthy individuals and those with diarrhoeal disease, finding a general reduction in bacterial diversity, decreased abundance of strictly anaerobic organisms and increases in the abundance of rapidly growing facultative anaerobes (The and Le, 2022). Although there is evidence indicating that pathogens, such as *Salmonella* and *Campylobacter*, can interact with other organisms during infection and potentially affect the gut microbiome composition (Indikova, Humphrey and Hilbert, 2015; Aljahdali *et al.*, 2020), the identification of microbiome patterns associated with specific pathogens has been shown to be difficult (The and Le, 2022) and requires further evaluation. Factors such as age, nutrition and geography have been found to also result in observable microbiome differences (The *et al.*, 2018).

At the food production level, metagenomic methods have been applied to study microbial composition in different broiler gut regions (Huang *et al.*, 2018), uncovering substantial differences in taxonomic diversity. The effect of diets on the chicken microbiota composition has also been investigated (De Cesare *et al.*, 2019). In addition, a limited number of studies have utilised metagenomic sequencing to characterise the resistomes of various environments, such as gut microbiomes of animals for human consumption (Ma *et al.*, 2016; Kumar *et al.*, 2020) including poultry (Munk *et al.*, 2018), as well as retail foods (Bloomfield *et al.*, 2023), highlighting the power of metagenomics for community profiling.

Metagenomic sequencing allows reconstruction of the microbiota, potentiating the identification of specific pathogens as well as other organisms, including those previously undescribed, and investigation of associations between microbiome composition and pathogen presence. However, there has been limited effort in metagenomic sequencing of retail chicken specifically to look for microbial differences associated with the presence and absence of pathogens for potential intervention. Some sequencing data representing retail chicken meat is available on public repositories; the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) (Leinonen *et al.*, 2011) contained 931 results for “"metagenomic"[Source] AND "food metagenome"[Organism] AND "chicken"[All Fields]” as of 13th August 2024. Although labelled as metagenomic, a proportion of this represents data obtained with amplicon sequencing. As shotgun metagenomics overcomes issues associated with bias and resolution that are often noted for amplicon sequencing approaches (Durazzi *et al.*, 2021), shotgun sequencing may be more appropriate for detailed study of microbiome features. Other issues with using publicly available data include differences in the sequencing instruments and sample preparation applied, and lack of experimental validation, particularly in the case of *Campylobacter* presence or absence. This necessitates sequencing of more retail chicken meat metagenomes, alongside culture-based validation.

1.5 Aims and objectives

This project aimed to address the knowledge gaps outlined throughout this chapter, relating to:

- Optimal culture condition combinations for *Campylobacter* recovery.
- The diversity of *Campylobacter* on retail chicken meat, at the point closest to the consumer.
- The potential of metagenomics for pathogen identification and characterisation, to overcome culturing difficulties and sampling intensity requirements in presence of high pathogen diversity, and reduce time-to-result for time-sensitive applications.
- The presence of organisms or microbial communities that may help or hinder *Campylobacter* survival on retail meat, which differs in microbial composition to the live chicken gut.

The specific aims and objectives were:

1. To explore different culture conditions that may enhance isolation of *Campylobacter* to reflect true prevalence and diversity on retail chicken.
2. To assess *Campylobacter* diversity on retail chicken samples using WGS.
3. To explore the efficacy of shotgun metagenomics for pathogen identification and characterisation, focusing on *Campylobacter* on retail chicken, and comparing this to *Salmonella* in human infection cases.
4. To investigate microbial signatures associated with *Campylobacter* presence and absence on retail chicken meat.

2. Chapter 2: General materials and methods

Chapter contributions: Method development for the *Campylobacter* culturing was done by AH Dziegiel, SJ Bloomfield, N Janecko, J Wain, AE Mather and AC Midwinter. Samples were collected by R Palau. Sample processing and *Campylobacter* culturing was done by AH Dziegiel. Metagenome extraction and host/bacterial qPCR was done by AH Dziegiel, SJ Bloomfield and R Palau. Metagenome libraries were formed by AH Dziegiel, SJ Bloomfield and the Quadram core sequencing team. Metagenomic sequencing was performed externally. Bioinformatics analysis was done by AH Dziegiel, with advice and training from SJ Bloomfield, in discussion with AE Mather.

2.1 Retail chicken sample collection

Fresh pre-packaged retail chicken meat samples were collected from small and large chain stores in Norwich, Norfolk, UK by Raphaëlle Palau. A total of 67 fresh retail chicken meat samples were purchased between March 2021 and March 2022 (Appendix 1). The samples were kept at 2-8°C and processed within 72 hours of receipt at Quadram Institute Bioscience, Norwich, UK.

2.2 *Campylobacter* detection and isolation from retail chicken

Approximately 100g of each chicken sample was placed into FBAG-03 filter blender bags (Corning). Samples were homogenised (Seward stomacher 400C laboratory blender) in 225 mL of buffered peptone water (BPW) (Southern Group Laboratory). Samuel Bloomfield combined 3.5 mL of each homogenised sample with 1.5 mL Brucella Broth (Becton Dickinson) + 17.5% glycerol (Fisher Scientific) for sample preservation at -70°C. For each sampling run, a positive and negative control was processed alongside the samples. The positive control consisted of 225 mL of BPW inoculated with a loopful of National Collection of Type Cultures (NCTC) strain 13367 *C. jejuni* culture, and the negative control consisted of 225 mL of uninoculated BPW.

Two aliquots of 35 mL of the homogenised liquid were taken into two 50 mL centrifuge tubes per sample for 30 minutes of centrifugation at 4,000 rpm at 6°C (Eppendorf Centrifuge 5810 R). The supernatant was removed, and the pellets were resuspended in 5 mL of BPW. The two replicate aliquots of each sample were pooled together.

The samples were subjected to 12 culture condition combinations for the recovery of *Campylobacter* (Figure 2.1). The methods tested included three broth conditions (none – direct plating, Bolton broth and cefoperazone, amphotericin B, teicoplanin (CAT) broth), two

temperature conditions (37°C and 42°C) and two agar plate types (mCCDA and mCCDA without selective supplements (u-mCCDA)).

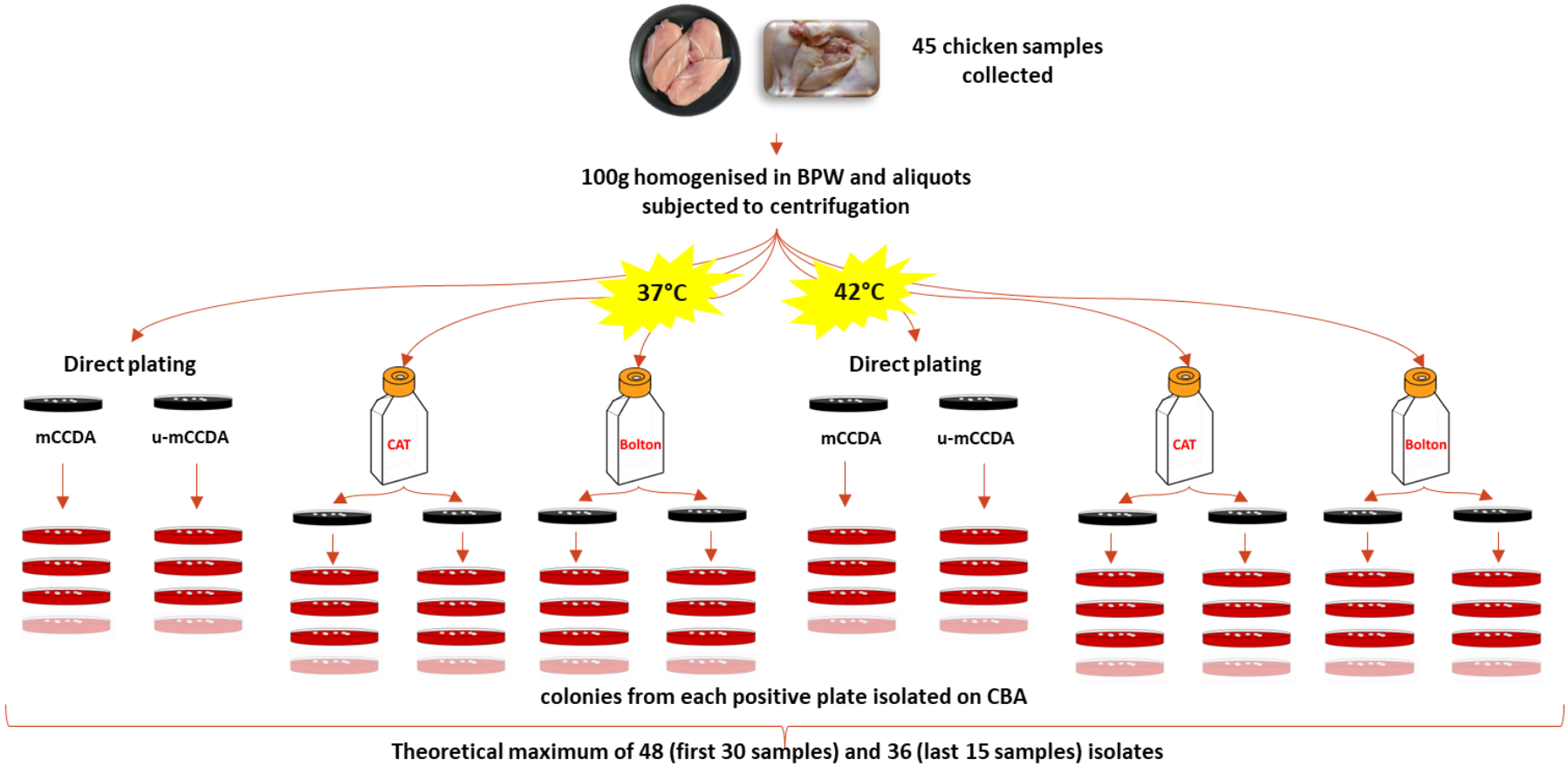

Figure 2.1: Culture condition combinations used for the recovery of *Campylobacter* from retail chicken samples

BPW=buffered peptone water; CAT=cefoperazone, amphotericin B, teicoplanin broth; mCCDA=modified charcoal-cefoperazone-deoxycholate agar; u-mCCDA=modified charcoal-cefoperazone-deoxycholate agar without supplements

Method 1: Direct plating

Approximately 10 µL of each pooled sample was inoculated onto mCCDA plates (Oxoid) prepared according to manufacturer instructions and mCCDA plates without selective supplements (u-mCCDA) in duplicate. One plate of each type was incubated at 37°C and the remaining plates at 42°C for 48 hours in microaerophilic conditions (85% nitrogen, 10% carbon dioxide, 5% oxygen) using an anaerobic cabinet (Don Whitley Scientific MAC 1000 Microaerophilic Workstation) and Anoxomat jars (Mart Microbiology B.V Anoxomat System AN2CTS), respectively, filled with an 85% nitrogen, 10% carbon dioxide, 5% oxygen gas mix.

Method 2: Enrichment in Bolton broth

For the first 15 samples (CH-0312-CH-0326), 1 mL of each pooled sample was inoculated into 20 mL of Bolton broth (Oxoid) supplemented with laked horse blood (Fisher Scientific) and Bolton selective supplement (SR0183; Oxoid) in duplicate in sterile universal bottles, with the lids loosened. One of the bottles was incubated at 37°C and the other at 42°C for 48 hours in microaerophilic conditions, as described previously. Sterile tissue flasks (Geiner Bio-One) were used instead of sterile universal bottles for the remaining samples, to reduce the risk of spills. Approximately 10 µL of sample from each bottle or flask was inoculated onto mCCDA and u-mCCDA and the plates incubated for 44-48 hours at the same temperature as the respective enrichment broth, using microaerophilic conditions as previously.

Method 3: Enrichment in CAT broth

CAT broth was prepared using Bolton broth base (Oxoid) supplemented with laked horse blood (Fisher Scientific) and CAT supplement (Oxoid). Use of this broth type was suggested by Anne Midwinter (Massey University, New Zealand). For the first 15 samples (CH-0312-CH-0326), 1 mL of each centrifuged pooled sample was inoculated into 20 mL of broth in duplicate in sterile universal bottles, with the lids loosened. After the first 15 samples, sterile flasks were used. The broths were incubated and the samples subcultured on agar plates as described in method 2.

Following direct plating, enrichment and post-enrichment plating, a maximum of four colonies appearing small and grey to translucent with an oily or metallic sheen were typically selected from the first 30 samples processed (CH-0312 to CH-0341) and a maximum of three colonies were selected from the next 15 samples (CH-0347 to CH-0361). This resulted in a theoretical maximum of 48 or 36 isolates per sample, respectively. For the first 30 samples, if individual culture

combinations displayed no growth or contained fewer than four typical colonies, additional colonies were selected from other combinations, though the potential maximum number of isolates per sample remained the same. For the last 22 samples (CH-0362 to CH-0383), one colony was selected from each culture positive plate; at this stage the focus shifted from investigation of *Campylobacter* diversity within the samples to determining whether the samples were culture positive or negative for *Campylobacter*.

The selected colonies were subcultured onto Columbia blood agar (CBA) (Oxoid) containing 5% horse blood (Trafalgar Scientific) and incubated in microaerophilic conditions for 48 hours at 37°C. Typical colonies were tested for oxidase production (Oxoid). A representative selection of typical, oxidase-positive isolates (at least one isolate per culture condition combination per sample) was also examined under the microscope to confirm the typical curved rod morphology and corkscrew motility of *Campylobacter* using an Olympus CX41 microscope (Olympus Solutions). Isolates were preserved in 1.5 mL Brucella broth + 20% glycerol at -70°C and confirmed with WGS.

2.3 DNA extraction and genome sequencing

Each presumptive *Campylobacter* isolate was subjected to DNA extraction using the Maxwell RSC Cultured Cells DNA kit (Promega). DNA extraction was performed according to manufacturer instructions. The bacterial inoculum for extraction was prepared as follows. For sequencing of genomes described in Chapter 3 and 5, isolates were cultured from frozen glycerol stocks on CBA for 48 hours in microaerophilic conditions at 37°C, using either the anaerobic cabinet (Don Whitley Scientific) or Anoxomat jars filled with a microaerophilic gas mixture (Mart Microbiology B.V). Colonies were either directly suspended in 400 µL phosphate buffered saline (PBS), which was loaded into the Maxwell cartridge, or inoculated into Brucella Broth made with ultrapure water and cultured in microaerophilic conditions at 37°C on a shaker at 300 rpm overnight before suspending 400 µL of the overnight culture directly into the Maxwell cartridge. DNA extraction was performed in part by Samuel Bloomfield. The DNA concentration was determined with the Qubit fluorimeter and broad range kit (Thermo Fisher Scientific). Presumptive isolates that appeared atypical during phenotypic confirmation testing or those that did not grow during culture for extraction were excluded from the dataset.

For short read sequencing (Chapter 3 and 5), the Nextera Flex DNA library preparation kit was used for paired-end library preparation and the libraries were sequenced on an Illumina NextSeq (Illumina) as 150 bp paired-end reads by the Quadram core sequencing team. During library preparation, the Kapa 2G PCR kit (Merck) was used for library barcoding for the majority of genomes; due to supply issues during the COVID-19 pandemic, the NEB Q5 (New England Biolabs)

polymerase was used instead in two sequencing batches representing 32.3% of genomes described in Chapter 3.

Long read sequencing methods are described within Chapter 4. The genome sequence data are available under the accessions outlined within the chapters or upon request.

2.4 Genome analysis

The common short read genome analysis methods used are outlined here. Genomic analysis for the long read genomes discussed in Chapter 4 is detailed within that chapter.

Analyses were performed in Galaxy (Afgan *et al.*, 2018), or on the QIB Cloud (based on the Cloud Infrastructure for Microbial Bioinformatics (CLIMB)) (Connor *et al.*, 2016) server.

Paired reads were trimmed using fastp (Chen *et al.*, 2018), with specific tool versions outlined within the chapters. Contigs were assembled using the SPAdes assembler (Prjibelski *et al.*, 2020) in Shovill v1.1.0+galaxy0 (https://github.com/tseemann/shovill), with the minimum coverage to count as part of a contig set to 0 [AUTO]. QUAST v5.0.2 (Gurevich *et al.*, 2013) and CheckM v1.0.11 (Parks *et al.*, 2015) were used to assess quality of the assemblies. Assemblies were generally accepted if they displayed <500 contigs over 500 bp and <50 duplicate marker genes. Coverage was estimated using BWA-MEM Galaxy v0.7.17.1 (Li and Durbin, 2009), using trimmed reads and assembled contigs as inputs. The resulting .bam files were analysed with CoverM v0.3.2 (https://github.com/wwood/CoverM). Assemblies were accepted if the mean read depth of the four largest contigs was over 30.

Sequence types (STs) were classified with MLST v2.16.1 (https://github.com/tseemann/mlst). ARIBA (Hunt *et al.*, 2017) was used to detect AMR genes or mutations associated with quinolone and macrolide resistance, with details and tool versions outlined in the chapters. StarAMR (Bharat *et al.*, 2022) was used to verify ARIBA results or for the identification of fluoroquinolone resistance mutations, with details and tool versions outlined in the chapters.

Details of other specific downstream analysis methods are outlined within the chapters.

2.5 Retail chicken metagenome extraction and sequencing

The procedure for chicken host DNA depletion was performed according to a published method (Bloomfield *et al.*, 2023). Raw and depleted chicken metagenomes were extracted using the Maxwell RSC Purefood Pathogen Kit (Promega). DNA quantification was performed using a Qubit

fluorometer and high sensitivity kit (Thermo Fisher Scientific) and qPCR was performed to quantify bacterial and host DNA in the depleted and undepleted samples to ensure that host DNA was removed and bacterial DNA retained, using the LightCycler 480 II instrument (Roche) according to a published method (Bloomfield *et al.*, 2023). qPCR quantification was performed partly by Raphaëlle Palau. The Nextera DNA flex library kits (Illumina) were used for library preparation. Library preparations for metagenomes with concentrations ≥5 ng/µL were performed by the Quadram core sequencing team. Library preparation was also attempted for all samples with concentrations below 5 ng/µL using the same method. Negative controls were spiked with phiX174 RF1 DNA (New England Biolabs) (maximum concentration of 0.5 ng/µL) for sequencing due to undetectable DNA concentration.

Metagenome libraries were pooled and the DNA concentration of the pooled samples was quantified with Qubit and library sizes were determined using the TapeStation 2200 instrument (Agilent) according to manufacturer instructions. The metagenomes were externally sequenced by Novogene and Source Bioscience using Illumina NovaSeq to generate 150 bp paired-end reads (target sequencing depth of 8 Gb per metagenome). The metagenome sequence data are available upon request.

2.6 Metagenome analysis

The analysis methods commonly applied to the metagenomes are outlined below. Analyses were performed in Galaxy, on the QIB Cloud server and on the Norwich Bioscience Institutes high performance computing (HPC) cluster.

Metagenome reads were pre-processed with fastp, with specific tool versions outlined within the chapters.

AMR genes were identified using KMA (Clausen, Aarestrup and Lund, 2018) using the ResFinder (Florensa *et al.*, 2022) database, with specific tool versions outlined within the chapters. KMA results were filtered at the 60% template coverage and 90% query identity thresholds. Metagenomes were classified with Kraken2 (Wood, Lu and Langmead, 2019) and Bracken (Lu *et al.*, 2017), with the specific details, tool versions and databases used outlined within the chapters.

Metagenomes were assembled with MEGAHIT v1.1.2 (Li *et al.*, 2015) and coverage estimated with Bowtie2 v2.3.4.1 (Langmead and Salzberg, 2012) and CoverM. Metagenome assembled genomes (MAGs) were assembled using Metabat2 v2.14 (Kang *et al.*, 2019) and Maxbin2 v2.2.4 (Wu, Simmons and Singer, 2016). MAG quality was assessed with QUAST and CheckM to determine

completeness, contamination, GC percentage and total length. MAGs were taxonomically classified with CAT bins v5.0.3 (von Meijenfeldt *et al.*, 2019).

Other specific analyses are outlined within the chapters.

2.7 Data manipulation, visualisation and statistical analysis

Data manipulation, visualisation and statistical analysis was performed in RStudio v2022.02.3+492 (RStudio Team, 2022) running R v4.2.3 (R Core Team, 2023) unless stated otherwise. Data was uploaded using the readxl v1.4.2 (Wickham and Bryan, 2023) and data.table v1.14.8 (Dowle and Srinivasan, 2023) packages, and manipulated using the tibble v3.2.0 (Müller and Wickham, 2023), tgutil v0.1.16 (Chomsky and Lifshitz, 2024), tidyverse v2.0.0 (Wickham *et al.*, 2019) and reshape2 v1.4.4 (Wickham, 2007) packages. Visualisation was performed using the ggplot2 v3.4.2 (Wickham, 2016), ggtree v3.6.2 (Yu *et al.*, 2017), ggtreeExtra v1.8.1 (Xu, Dai, *et al.*, 2021), treeio v1.22.0 (Wang *et al.*, 2020), ggstar v1.0.4 (Xu, 2022), ggnewscale v0.4.8 (Campitelli, 2022), ggridges v0.5.4 (Wilke, 2022), ggpattern v1.0.1 (FC and Davis, 2022), patchwork v1.1.2 (Pedersen, 2022), ggbreak v0.1.1 (Xu, Chen, *et al.*, 2021), ggpubr v0.6.0 (Kassambara, 2023), and phytools v1.5-1 (Revell, 2012) packages. Specific methods and any additional packages used are outlined in individual research chapters.

3. Chapter 3: Investigation of *Campylobacter* diversity on retail chicken

Chapter contributions: The study was conceptualised by AE Mather. Method development for the *Campylobacter* culturing was done by AH Dziegiel, SJ Bloomfield, N Janecko, J Wain, AE Mather and AC Midwinter. Samples were collected by R Palau. Sample processing and *Campylobacter* culturing was done by AH Dziegiel. Genome extraction was done by AH Dziegiel and SJ Bloomfield. Genome sequencing was performed by the Quadram core sequencing team. Bioinformatics analysis was done by AH Dziegiel and SJ Bloomfield, in discussion with AE Mather. Statistical method development and analysis was done by AH Dziegiel (Wilcoxon signed-rank tests) and GM Savva (regression models, McNemar tests and simulations). Visualisation was done by AH Dziegiel; in addition, GM Savva provided code for figures resulting from the regression models and simulations performed in the study. The work was discussed throughout by AH Dziegiel, SJ Bloomfield, GM Savva, N Janecko, J Wain and AE Mather.

A manuscript based on the work described in this chapter, entitled "**High *Campylobacter* diversity in retail chicken – epidemiologically important strains may be missed with current sampling methods**", has been published in Epidemiology & Infection (Dziegiel *et al.*, 2024)**.**

3.1 Introduction

Campylobacter spp. are the most common causative agents of bacterial gastroenteritis in the UK and worldwide (Tam *et al.*, 2012). The epidemiology of *Campylobacter* is complex and not fully understood, and it is believed that only a fraction of the overall cases are reported (Tam *et al.*, 2012). The species most often implicated in infection are *C. jejuni* and *C. coli*; however, other emerging *Campylobacter* species such as *C. lari* and *C. fetus* have also been attributed to infections, particularly in immunocompromised hosts (Tauxe *et al.*, 1985; Krause *et al.*, 2002; Wagenaar *et al.*, 2014). The majority of campylobacteriosis cases feature self-limiting gastroenteritis, although chronic infections and post-infectious sequelae such as Guillain-Barré syndrome are a risk in a small proportion of cases (Allos, 1997; Bloomfield *et al.*, 2018).

Campylobacter is a highly prevalent organism in poultry, and thus one of the most common sources of *Campylobacter* in humans is chicken meat (Skarp, Hänninen and Rautelin, 2016). The most common *Campylobacter* species isolated from poultry is *C. jejuni* followed by *C. coli*, but other species including *C. upsaliensis*, *C. concisus* and *C. lari* have also been detected (European Food Safety Authority, 2010; Lynch *et al.*, 2011; Kaakoush *et al.*, 2014). The standard method of *Campylobacter* detection is through culture, with the majority of methods proposed by

internationally recognised regulatory bodies tailored to *C. jejuni* and *C. coli* isolation. However, these methods can be subject to isolation bias, as the growth requirements can vary by sample type, individual species and subtypes (Williams *et al.*, 2012; Hetman *et al.*, 2020; Harrison *et al.*, 2022). In addition, generally a single colony is selected as representative of the *Campylobacter* population within a particular sample (Food Standards Agency, 2016; Hull *et al.*, 2021); if multiple *Campylobacter* species or strains are present, this can have implications on outbreak investigations and source attribution. Indeed, *Campylobacter* outbreaks are considered to be relatively rare – the European Food Safety Authority estimated only 1% of the overall cases were attributed to outbreaks in 2020 (Liu *et al.*, 2022). This may be in part due to underreporting (Wagenaar, French and Havelaar, 2013), though there is also some evidence to suggest that chicken meat processing may increase the diversity of *Campylobacter* present on chicken (Inglis *et al.*, 2021), which may not be captured with limited sampling, and thus the outbreak burden may be underestimated. This issue may be further exacerbated through the use of culture methods that can miss *Campylobacter* positive samples. Establishment of an optimal isolation method, or combination of methods, to maximise the isolation of different *Campylobacter* species and subtypes present on chicken meat is therefore important to facilitate understanding of the transmission and pathogenic capabilities of *Campylobacter* subtypes.

The *Campylobacter* MLST scheme (Dingle *et al.*, 2001; Jolley, Bray and Maiden, 2018) has been useful for clustering genetically related *Campylobacter* isolates without the need for sequencing of the whole genome, allowing grouping of potential outbreak and source isolates based on sequence differences in seven housekeeping genes. Limited research sampling multiple isolates has found that up to eight STs can be recovered from individual free range chicken carcasses at the slaughterhouse and such high diversity is not found at the farm level (Colles *et al.*, 2010). Although it has been suggested that the diversity of *Campylobacter* can generally increase through processing and be maintained through to retail (Inglis *et al.*, 2021), it is unclear whether or not such diversity persists on individual meat samples. This is important to establish as this is the point closest to the consumer.

Although MLST provides a useful diversity measure, it has been shown that many of the major *Campylobacter* STs implicated in disease display a host-generalist lifestyle (Dearlove *et al.*, 2016), facilitating colonisation of a number of hosts and reducing host signals required for source attribution. Whole genome sequencing provides a higher resolution than MLST, allowing for the investigation of multiple genomic features to infer within-lineage diversity, including SNPs and AMR genes. While investigation of AMR in *Campylobacter* has clinical significance, particularly regarding mutations associated with quinolone and macrolide resistance (1.3.1), differences in AMR genotypes for chromosomally located genes can potentially suggest presence of multiple

clones. It is important to establish the extent of the diversity of *Campylobacter* spp. on major infection sources such as chicken meat in order to understand the success of this pathogen within this niche and establish potential intervention strategies.

3.2 Aims and objectives

The work outlined in this chapter aimed to:

- Use a combination of culture methods to maximise the isolation of *Campylobacter* from retail chicken meat.
- Compare *Campylobacter* recovery between the culture combinations used.
- Take multiple isolates per sample to evaluate the diversity of the *Campylobacter* isolates obtained using WGS.
- Illustrate the implications of high intra-sample *Campylobacter* diversity on source attribution and outbreak investigation.
- For STs isolated with only one enrichment broth type or direct plating, determine if these STs can be enumerated in alternative conditions, or if their growth is method-dependent.

3.3 Materials and methods

Details of sample collection and *Campylobacter* detection, isolation, genome extraction and sequencing are outlined in the Materials and Methods chapter, under sections 2.1-2.3. The sequence data are available in the NCBI SRA (Bioproject accession PRJNA1022324).

3.3.1 Genome analysis

The common genome analysis methods are outlined in the Materials and Methods chapter (2.4).

Paired reads were trimmed using fastp v0.19.5+galaxy1 (Chen *et al.*, 2018). Centrifuge v1.0.4_beta (nt_2018_3_3 database) (D. Kim *et al.*, 2016) was applied to paired trimmed reads for species classification. The estimated genome size was set to 2,000,000 for assembly with Shovill v1.1.0+galaxy0 (https://github.com/tseemann/shovill).

AMR genes were detected in genome assemblies with ABRicate v0.9.7 (https://github.com/tseemann/abricate) using the ResFinder (Florensa *et al.*, 2022) database (built 5th November 2021) and 90% coverage and identity thresholds. ARIBA v2.14.6 (Hunt *et al.*, 2017) was used to identify quinolone and macrolide resistance determinants, using *gyrA* genes

extracted from the *C. jejuni* SAMEA1705929, *C. coli* SAMN11056450 and *C. lari* SAMN02604025 reference genomes (for quinolone resistance mutations) and *C. jejuni* (NR_076226.1), *C. coli* (NR_121940.1) and *C. lari* (NR_076560.1) 23S rRNA genes (for macrolide resistance mutations). The 23S rRNA gene database was obtained from Samuel Bloomfield. StarAMR v0.5.1 (Bharat *et al.*, 2022) was used to verify ARIBA results for the identification of quinolone resistance mutations with 98% BLAST identity and 95% PointFinder (Zankari *et al.*, 2017) (database v050218) BLAST overlap thresholds.

In cases where StarAMR did not identify any T86I and/or P104S mutations in the *gyrA* gene, ARIBA reports were manually screened to confirm their presence; this was because in some cases ARIBA indicated that the gene was incomplete, however the truncation presented away from the mutation of interest. Manual screening also aimed to avoid PointFinder missing mutations in genomes of species other than *C. jejuni*. For macrolide resistance mutations in the 23S rRNA gene, the ARIBA results were summarised with ARIBA summary. Mutations at the 2074 and 2075 position in the *C. jejuni* 23S rRNA gene, 2075 or 2076 in the *C. lari* gene and 2232 or 2233 in the *C. coli* assembled gene were searched for, regardless of the species of the genome. For genomes in which none of the genes assembled, the genome assemblies were subjected to annotation with Bakta v1.6.1+galaxy0 (Schwengers *et al.*, 2021) and the nucleotide sequences of the 23S rRNA gene manually extracted. These sequences were aligned to the references using muscle v3.8.31 in SeaView v5.0.4 (Edgar, 2004; Gouy *et al.*, 2021) to manually screen for mutations at these positions.

3.3.2 Phylogenetic analysis

The method for ST classification is outlined in the Materials and Methods chapter (2.4).

Isolates were grouped by species and the trimmed reads analysed with snippy and snippy-core v4.4.3 (https://github.com/tseemann/snippy), using the SAMEA1705929, SAMN02743854 and SAMN02604025 reference genome chromosome sequences for *C. jejuni*, *C. coli* and *C. lari*, respectively. Gubbins v2.4.1 (Croucher *et al.*, 2015) was then used to remove putative recombination regions, and the filtered polymorphic sites used to construct maximum-likelihood phylogenetic trees using IQ-tree v1.6.12 with 1,000 ultrafast bootstrap replicates (Minh, Nguyen and von Haeseler, 2013; Nguyen *et al.*, 2015). Additionally, the snippy-core full alignments were split by ST by Samuel Bloomfield. These individual alignment groups were subjected to analysis with Gubbins and snp-dists v0.6.3+galaxy0 (https://github.com/tseemann/snp-dists) to quantify non-recombinogenic pairwise SNP distances between genomes within ST groups. Alignments of STs that comprised of three genomes or less were combined with alignments of a closely related

ST for removal of putative recombination regions as Gubbins required alignment inputs of at least three sequences.

3.3.3 Further investigation of sequence types obtained with individual methods

Five isolates belonging to four STs that were initially only isolated with either Bolton broth enrichment or CAT broth enrichment were selected for further investigation. The aim of this was to determine if the growth of these STs was method dependent.

The isolates were cultured from frozen glycerol stocks on CBA for 48 hours at 37°C in microaerophilic conditions in the anaerobic cabinet filled with microaerophilic gas mixture. An inoculum equivalent to 0.5 McFarland standard was prepared in sterile PBS for each isolate, measured using a Fisher Scientific Cell Density Meter Model 40 (Fisher Scientific) to a value of 0.1 (±0.01).

For direct plating, approximate 1 in 100 serial dilutions were performed in PBS by adding 100 µL to 10 mL PBS, and 100 µL of the 10^{-4}, 10^{-6} and 10^{-8} dilutions were spread on four mCCDA and four CBA plates. Two plates of each type and each dilution were incubated at 37°C and the remaining plates at 42°C for 48 hours in microaerophilic conditions using an anaerobic cabinet filled with microaerophilic gas mixture and Anoxomat jars filled with microaerophilic gas mixture, respectively.

Two tissue flasks containing 20 mL Bolton broth and two tissue flasks containing 20 mL CAT broth were inoculated with 200 µL of the undiluted initial inoculum. One flask of each type was incubated at 37°C and the other at 42°C in microaerophilic conditions, as previously. After 48 hours, approximate 1 in 100 serial dilutions were performed in PBS as previously, and 100 µL of the 10^{-4}, 10^{-6} and 10^{-8} dilutions were spread on mCCDA and CBA, as previously.

During the first trial involving isolate 330-6-5, the post-enrichment plates from the 42°C enrichments were also incubated at 42°C. Due to large amounts of swarming on the plates, the post-enrichment plates for the remaining isolates tested were incubated at 37°C instead. Plates were incubated for 48 hours in microaerophilic conditions in the anaerobic cabinet.

Colonies on the plates were counted after incubation, and the average number of colonies per condition combination calculated. The average number of colonies, approximate dilution factor (10^{-4}, 10^{-6} or 10^{-8}) and dilution volume (0.1 mL) spread on the plates were used to calculate the concentration of *Campylobacter* (CFU/mL) for each condition combination. The values were log-transformed (using the natural log scale) for plotting and statistical analysis.

The log CFU/mL values obtained from mCCDA and CBA plates in this experiment were compared using a Wilcoxon signed-rank test.

3.3.4 Statistical comparisons of culture methods

Methods and code for statistical analysis were generated by George Savva. Statistical analyses were performed to determine if there were significant differences in *Campylobacter* detection between the culture method combinations used for isolation. As each temperature and broth combination was plated on both mCCDA and u-mCCDA, initially the effect of plate type on *Campylobacter* detection was tested using a McNemar test (for a paired binary outcome).

To test the effect of the broth and temperature predictor variables, mixed effect logistic regression modelling was applied. A multivariable model was constructed using data from mCCDA plates only in R using the glmmTMB v1.1.7 package (Brooks *et al.*, 2017), with the unique sample modelled as a random effect. The outcome variable modelled was *Campylobacter* growth or no growth (defined through recovery of *Campylobacter* isolates confirmed with sequencing). Another multivariable model with an interaction term between the broth and temperature conditions was also constructed, and the two models compared using a log likelihood ratio test (analysis of variance; ANOVA). The emmeans v1.8.5 package (Lenth, 2023) was used to estimate marginal means for proportions along with asymptotic confidence intervals from the models.

3.3.5 Simulation studies

Simulations were used to examine the potential implications of selecting a limited number of isolates per sample in the presence of high *Campylobacter* ST diversity. The method and code for this was generated by George Savva. The vegan v2.6.4 (Oksanen *et al.*, 2022) package was used to produce a rarefaction curve without replacement, reflecting the diversity of STs across samples. Another rarefaction curve was generated by resampling of the observed ST distribution (with replacement).

In each simulation, random subsamples of size N (with replacement) of the observed isolates were selected, with the estimated diversity from the subsample compared to the total observed diversity from all isolates. The observed ST distribution across culture conditions was considered to represent the true distribution of STs within each sample for each simulation.

First, as the number of isolates selected was increased, the number and proportion of different STs identified in each sample were measured. This was used to determine the number of isolates

needed for the expected number of different STs identified in the subsample to be at least 95% of the observed number in the whole sample.

Then, one ST from each sample was randomly selected as an ST of interest – i.e. marked as a potential outbreak causing ST – and the probability of its detection was estimated as the number of isolates sampled increased. This was averaged over each ST in each sample in turn to determine the average number of isolates required for the probability of detecting a specific ST to reach 95%.

3.4 Results

3.4.1 Retail chicken samples processed

Details of the retail chicken samples processed for *Campylobacter* detection and assessment of diversity are outlined in Appendix 1. For this part of the project, 45 chicken samples were collected between March and November 2021, from chain stores in Norwich, Norfolk, UK.

3.4.2 Comparison of culture approaches for the isolation of *Campylobacter*

From the 45 chicken samples processed, 39 (86.7%) were culture-positive for *Campylobacter*. A total of 743 isolates were recovered from the tested samples, with a median of 16 and a range of 0-45 isolates per sample. The majority of isolates (n = 705, 94.9%) were obtained with enrichment compared to direct plating (n = 38, 5.11%). The highest number of isolates was recovered using CAT broth (n = 367), followed by Bolton broth (n = 338), direct plating on mCCDA (n = 29) and u-mCCDA (n = 9). There was also a difference in the number of isolates recovered at 37°C (n = 330) and 42°C (n = 413).

A total of 28 (62.2%) samples were culture positive for *Campylobacter* through Bolton broth enrichment, 37 (82.2%) through CAT broth enrichment and 10 (22.2%) through direct plating. *Campylobacter* was recovered from 33 (73.3%) samples at 37°C and 38 (84.4%) at 42°C.

Only two samples (CH-0317 and CH-0335) out of 39 were culture positive through all broth and temperature combinations (Figure 3.1) and only one was positive through all broth, temperature, and plate combinations (CH-0317). One sample was positive through direct plating on mCCDA only (CH-0315), and Bolton broth was the only medium to yield growth in another sample (CH-0358). CAT broth enrichment resulted in the isolation of *Campylobacter* in eight samples that did not yield growth with Bolton broth or direct plating.

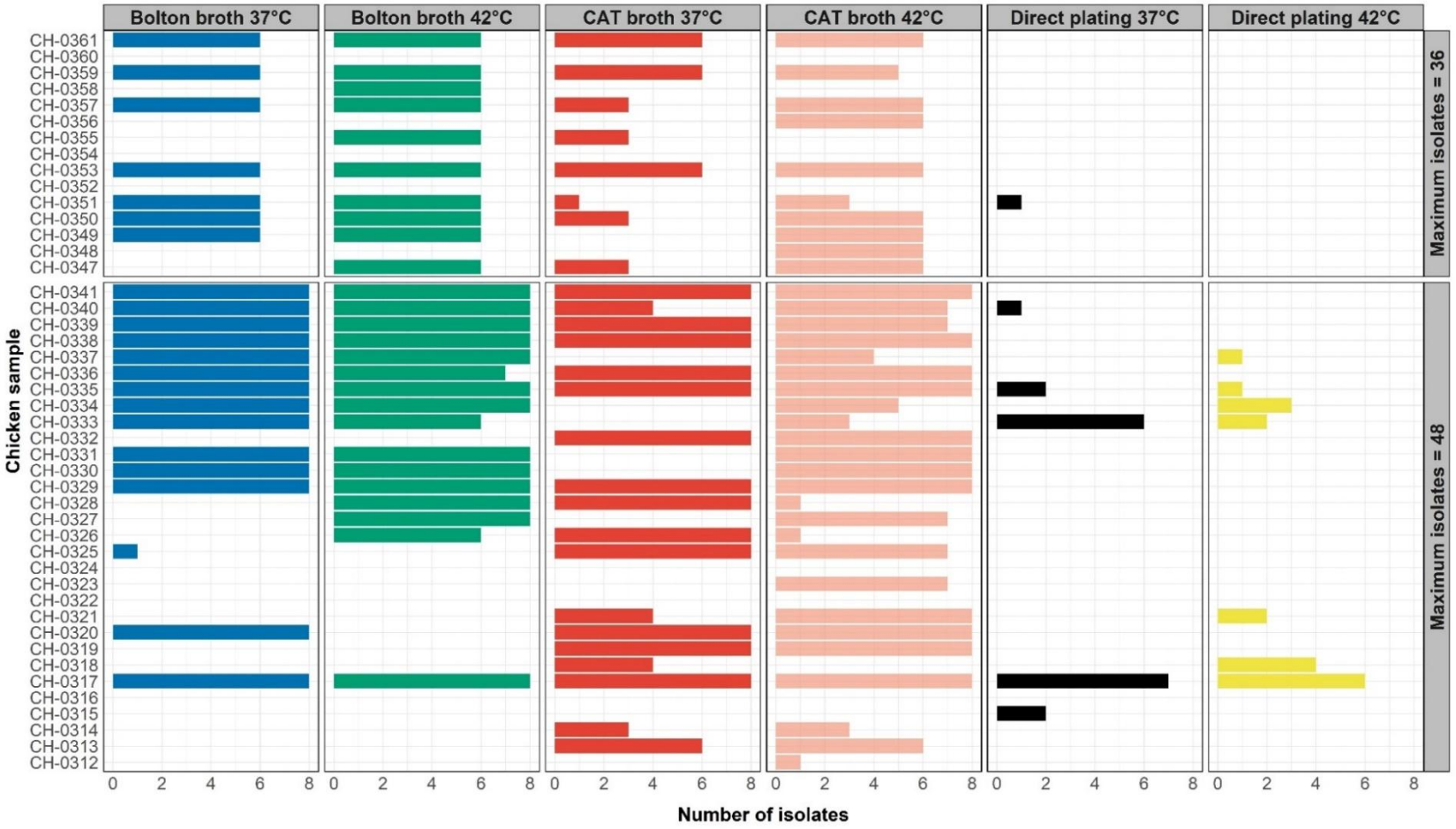

Figure 3.1: Number of *Campylobacter* isolates recovered from 45 retail chicken samples through direct plating and enrichment in Bolton and cefoperazone, amphotericin B, teicoplanin (CAT) broth at 37°C and 42°C, split by the theoretical maximum number of isolates taken per sample

A summarised dataset of *Campylobacter* growth (1) or no growth (0) by sample and culture method component was used as input for statistical analysis and modelling (Supplementary Table S1).

The concordance between agar types was high. In 94 cases, *Campylobacter* was detected using both agar types, and in 148 cases *Campylobacter* was not detected using either agar type (Table 3.1). Nevertheless, mCCDA was the more sensitive agar type – in 24 cases, *Campylobacter* was recovered with mCCDA but not u-mCCDA, whereas u-mCCDA recovered *Campylobacter* in four cases that did not yield *Campylobacter* on mCCDA (McNemar test p = <0.001).

Table 3.1: Comparison table for the agar plate types used for *Campylobacter* recovery

		u-mCCDA	
		0	**1**
mCCDA	**0**	148	4
	1	24	94

mCCDA=modified charcoal-cefoperazone-deoxycholate agar; u-mCCDA= modified charcoal-cefoperazone-deoxycholate agar without supplements

As mCCDA was determined to be the more sensitive agar type condition, multivariable mixed effects models testing the effect of temperature and broth were made using summarised *Campylobacter* growth/no growth data from mCCDA plates only. The final model indicated a significantly higher recovery of *Campylobacter* at 42°C compared to 37°C and using CAT broth enrichment compared to Bolton broth enrichment. *Campylobacter* recovery was also significantly lower with direct plating compared to Bolton broth enrichment (Table 3.2, Figure 3.2).

Table 3.2: Comparison of the main effects logistic regression model using modified charcoal-cefoperazone-deoxycholate agar (mCCDA) plate data only (mCCDA_main) and a logistic regression model with an interaction term between Broth and Temperature (mCCDA_interaction), with the chicken sample modelled as a random effect; the values represent odds ratios, confidence intervals and p-values

	mCCDA_main	**mCCDA_interaction**
BrothCAT	2.253 [1.040, 4.880] p = 0.039 *	1.531 [0.536, 4.369] p = 0.43
BrothNone	0.042 [0.015, 0.123] p = <0.001 ***	0.050 [0.012, 0.203] p = <0.001 ***
Temperature42	2.264 [1.134, 4.520] p = 0.020 *	1.768 [0.616, 5.071] p = 0.29
BrothCAT x Temperature42		2.359 [0.501, 11.114] p = 0.28
BrothNone x Temperature42		0.745 [0.124, 4.494] p = 0.75
Num.Obs.	270	270
R2 Marg.	0.325	0.325
R2 Cond.	0.657	0.664
AIC	279.1	281.2
BIC	297.1	306.4
ICC	0.5	0.5
RMSE	0.32	0.31

Num.Obs=number of observations; R2 Marg.=marginal R-squared; R2 Cond.=conditional R-squared; AIC=Akaike Information Criterion; BIC=Bayes Information Criterion; ICC=intraclass correlation criterion; RMSE=root square mean error

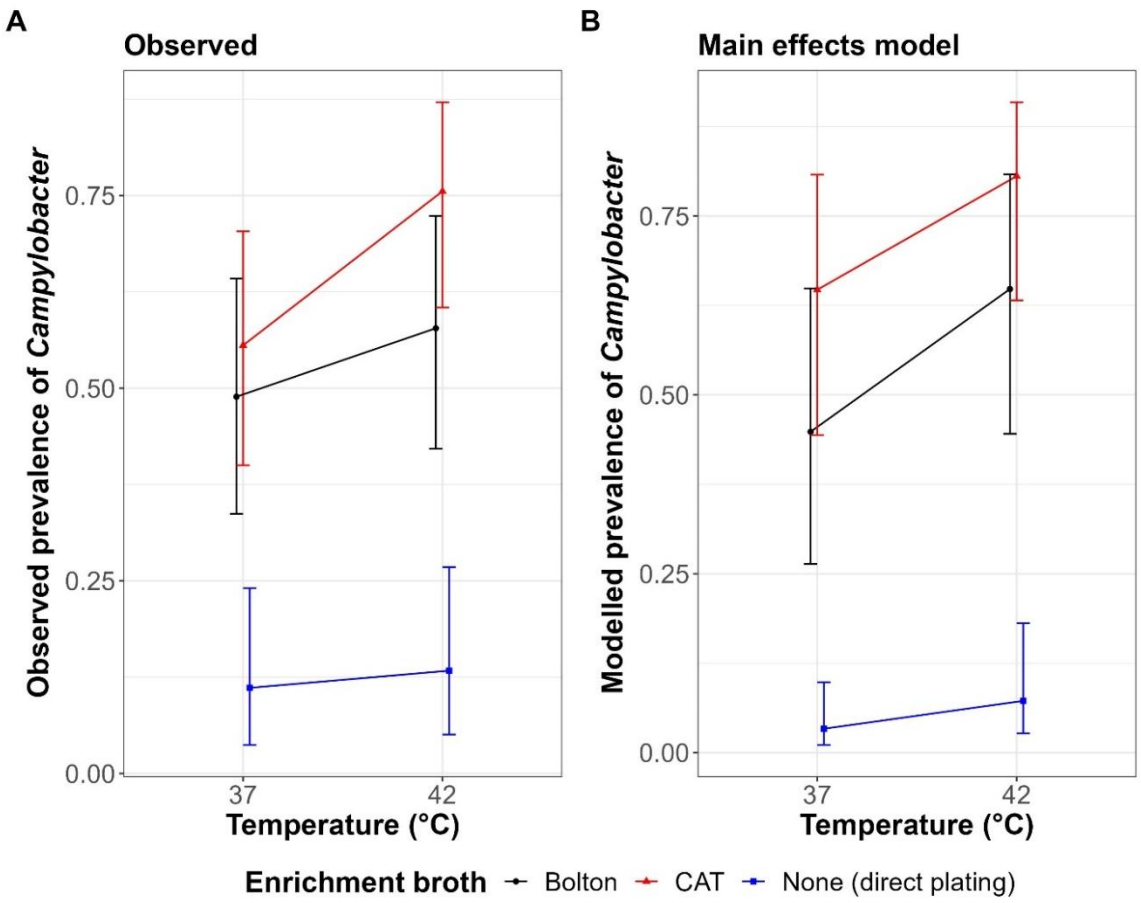

Figure 3.2: The observed proportion of samples testing positive for *Campylobacter* under each condition on modified charcoal-cefoperazone-deoxycholate agar (mCCDA) plates (A) and the estimated marginal mean proportions of samples testing positive under each condition in the final model (B); error bars represent 95% confidence intervals

Addition of an interaction term between broth and temperature did not significantly improve the model (p = 0.39; Table 3.3), indicating no evidence that the effect of temperature varied by broth.

Table 3.3: Model comparisons using analysis of variance (ANOVA)

Model	Model Type	Df	AIC	logLik	*P*
mCCDA_main	Main effects model	5	279.08	-134.54	
mCCDA_interaction	Interaction model	7	281.22	-133.61	0.39

mCCDA=modified charcoal-cefoperazone-deoxycholate agar; u-mCCDA= modified charcoal-cefoperazone-deoxycholate agar without supplements; Df=degrees of freedom; AIC=Akaike information criterion; logLik=log likelihood

3.4.3 Intra-sample species and sequence type diversity and potential implications on source attribution

The genomes obtained were analysed to determine their species and ST. Of the 743 *Campylobacter* isolates recovered in this study, 499 (67.2%) were *C. jejuni*, 228 (30.7%) were *C. coli* and 16 (2.15%) were *C. lari*. *C. jejuni* isolates were recovered through all of the culture methods tested, and *C. coli* isolates were recovered using Bolton broth and CAT broth at both temperatures, with only one isolate recovered with direct plating at 37°C on mCCDA.

C. jejuni was recovered from 36 (80.0%) samples and *C. coli* from 17 (37.8%) samples, with 14 (31.1%) samples containing both species. All 16 *C. lari* isolates were obtained through enrichment at 37°C, equally with CAT broth and Bolton broth from one (2.22%) sample (CH-0320) that also contained *C. jejuni*. Overall, 15 (33.3%) samples were positive for more than one *Campylobacter* species. The phylogenies of the *C. jejuni* and *C. coli* isolates are displayed in Figure 3.3.

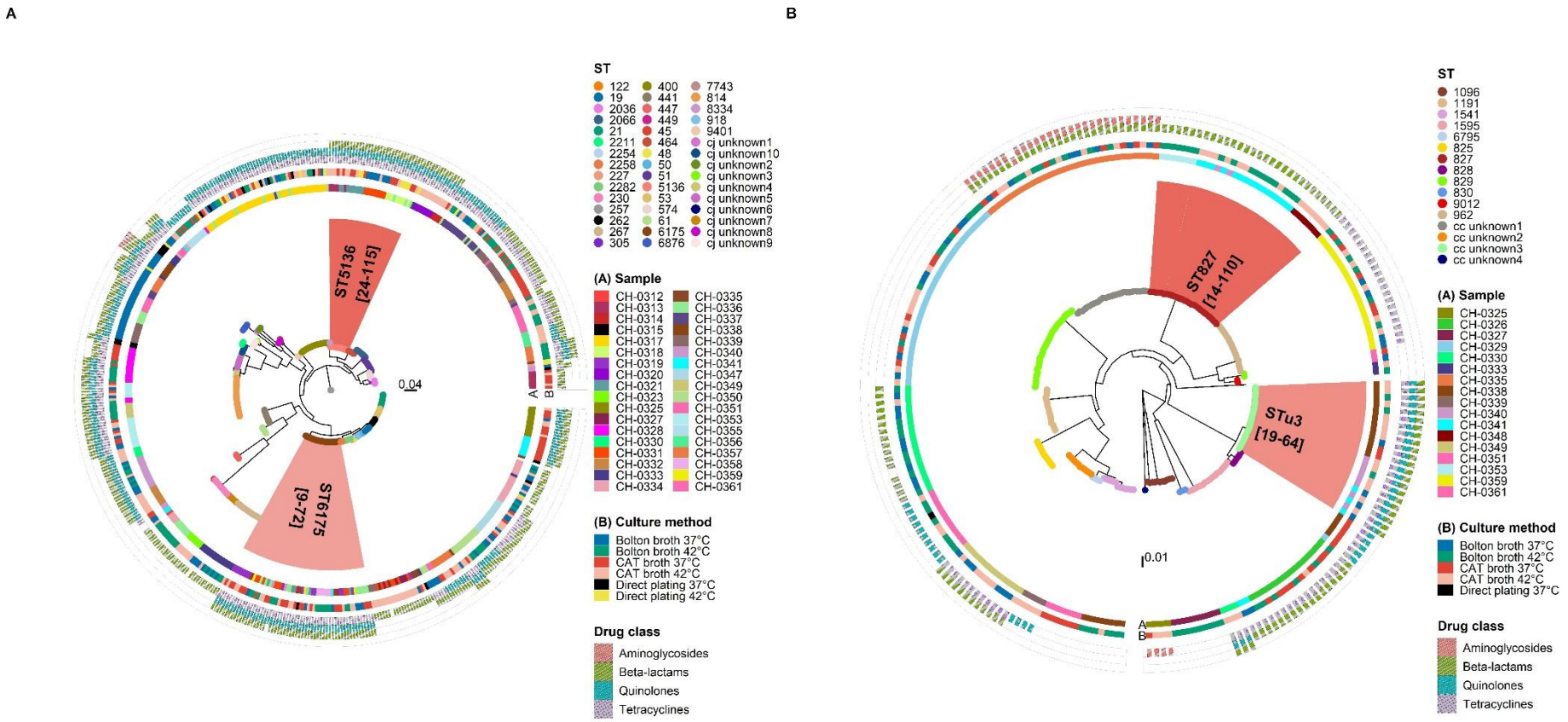

Figure 3.3: Maximum likelihood trees displaying the major sequence types (STs) and their pairwise single nucleotide polymorphism (SNP) difference ranges, the chicken meat sample of origin, the culture method for recovery, and the number of antimicrobial resistance (AMR) determinants identified in the 499 *C. jejuni* genomes (A) and the 228 *C. coli* genomes (B)

A total of 62 different STs were identified amongst the three *Campylobacter* species, 14 of which were novel STs (Appendix 2 and Appendix 3). Of these, 16 were identified with CAT broth only, out of 47 STs isolated with CAT broth in total. Enrichment in Bolton broth facilitated the identification of 41 STs, nine of which were identified using this broth type only. Direct plating yielded 14 STs, five of which were not identified with enrichment. Incubation at 37°C resulted in the identification of 43 STs and incubation at 42°C resulted in the recovery of 48 STs, with 15 identified at 37°C only and 19 at 42°C only. Fifty-nine STs were identified using mCCDA compared to 54 using u-mCCDA, with eight identified exclusively with mCCDA and three with u-mCCDA. The major *C. jejuni* STs identified were ST-6175 and ST-5136, identified in 11 and six samples, respectively; whereas ST-827 (four samples) and one of the novel STs (*C. coli* unknown3; three samples) were the major *C. coli* STs.

Twenty-six STs were identified in more than one sample. The number of STs in a single culture positive sample ranged between 1-8 (Figure 3.4) and up to seven STs of a single *Campylobacter* species were identified in a single sample. Sample CH-0333 displayed the highest ST diversity, followed by ST-0341 with seven STs; both of these samples contained both *C. jejuni* and *C. coli*. The highest number of isolates was recovered from sample CH-0317, with all culture method combinations isolating *Campylobacter* but all the isolates belonged to the same ST (ST-400).

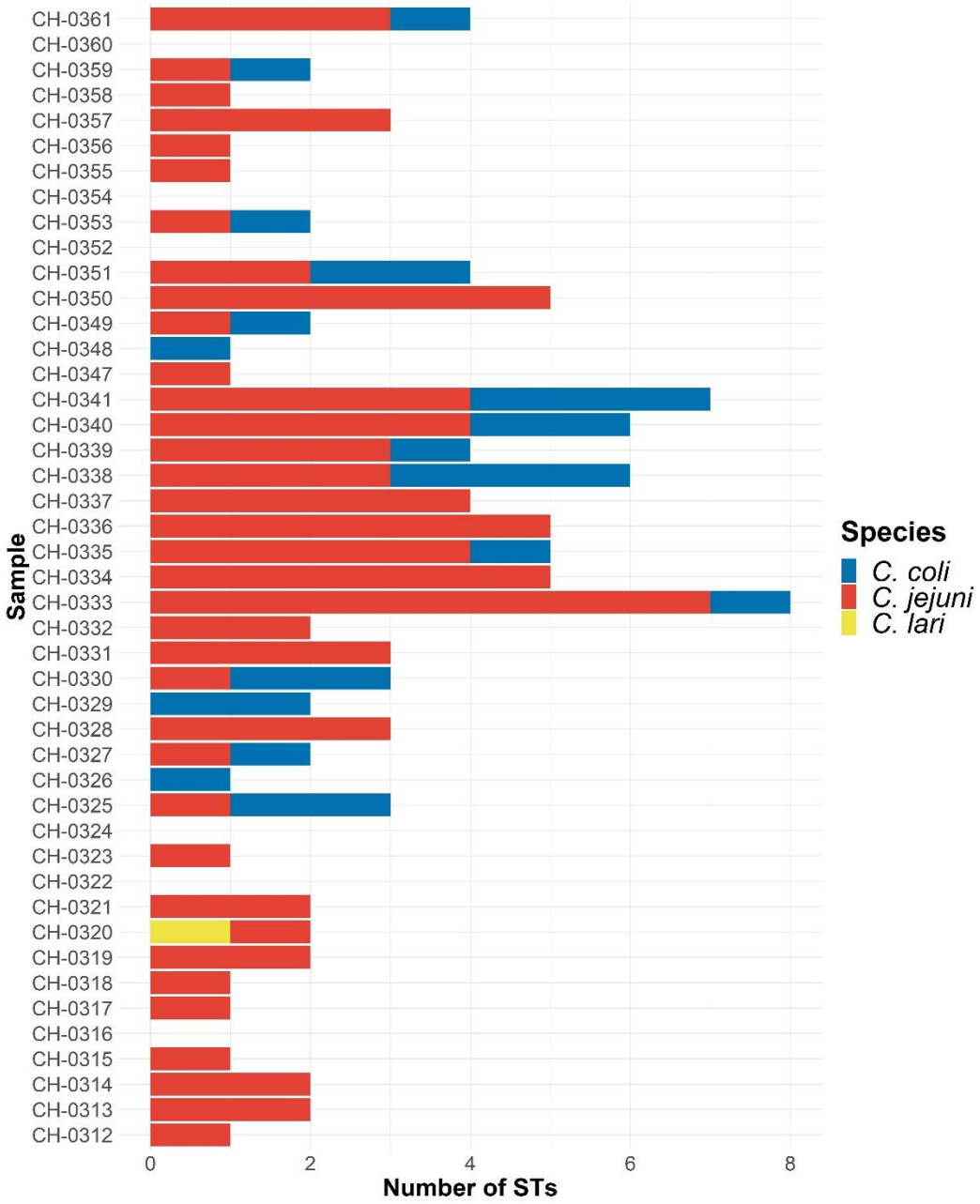

Figure 3.4: Number of *Campylobacter* sequence types (STs) identified in the retail chicken meat samples, coloured by species

Rarefaction curves of STs for each sample (with and without replacement) demonstrated a wide range in the number of expected STs recovered with increased sampling intensity (Figure 3.5A and 3.5B). The data obtained were used to perform simulations to determine the number of isolates required to obtain 95% of the sample ST diversity on average based on random resampling. The number of isolates required ranged between one, if only a single ST was detected, to 87 (median = 8) when very high diversity or very rare STs (represented by a low isolate count) were discovered (Figure 3.5C).

In an outbreak scenario, a limited number of isolates is usually taken from each potentially implicated food sample. High diversity of *Campylobacter* present on chicken meat can therefore affect source attribution. Isolate counts for each ST present within the chicken samples analysed in this study were used to calculate the probability of detecting the ST implicated in a theoretical outbreak from the sample depending on the number of isolates taken. This was repeated for all of the STs to obtain average probabilities. If a single isolate was sampled, the average probability of detecting the correct ST was 53%, while 26 isolates per sample would be required for the probability that the correct ST is among those discovered to reach 95% (Figure 3.5D).

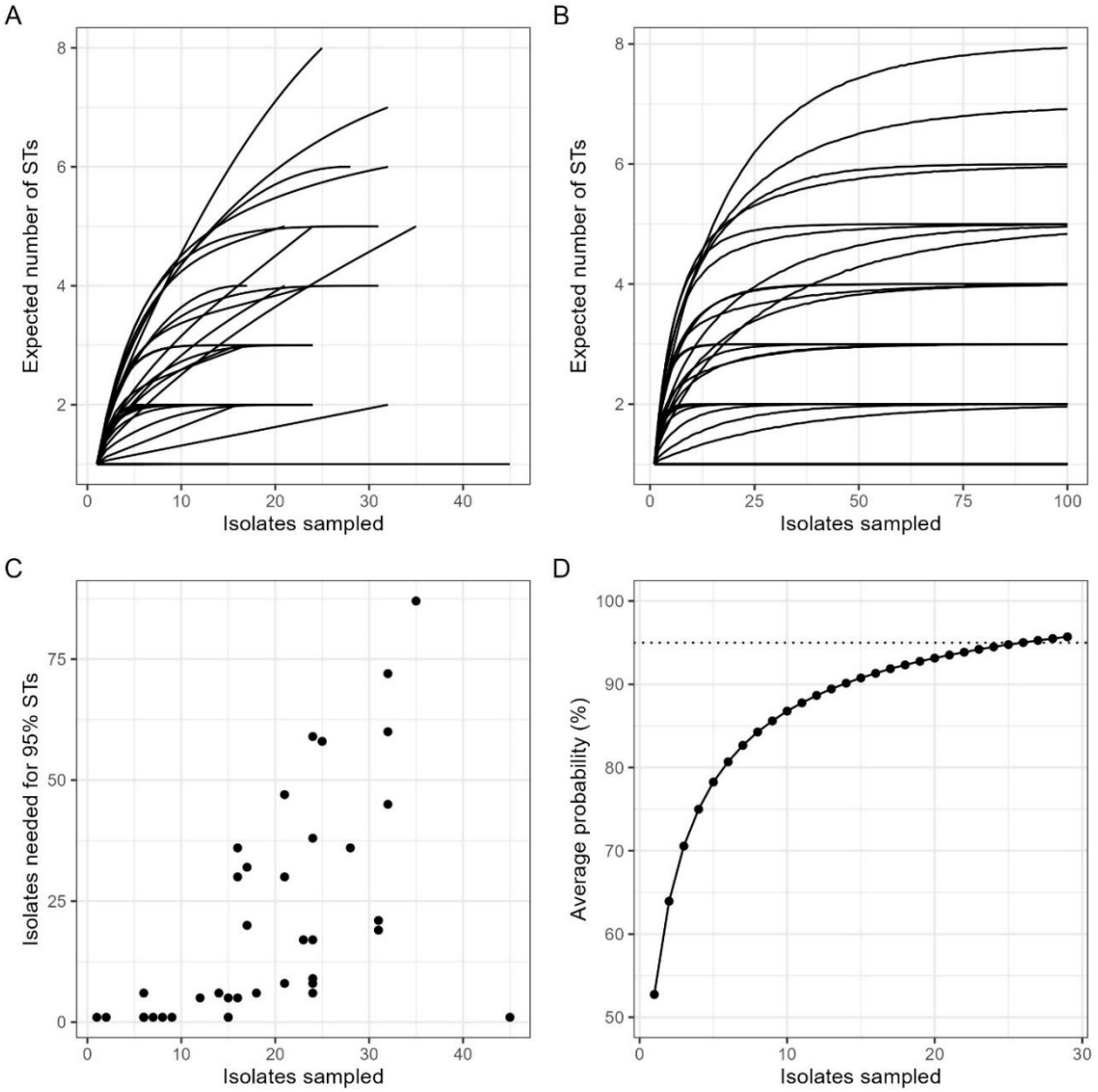

Figure 3.5: Sequence type (ST) rarefaction curve without replacement (A) and with replacement (B); simulations performed followed the principle of rarefying with replacement to determine the number of isolates required to observe 95% of the ST diversity within each chicken meat sample compared to the number of isolates that were collected (C) and to determine the average probability of detecting a randomly selected ST as the number of isolates increased, with the dashed line representing an average probability of 95% (D)

3.4.4 Diversity of individual sequence types inferred with pairwise single nucleotide polymorphism analysis

Fifty-five STs consisted of more than one isolate, allowing determination of pairwise SNP distances. The maximum SNP distances for STs with >1 isolate ranged between 17-244 (median = 54) SNPs within samples and 22-2413 (median = 86) SNPs overall (Figure 3.6, Appendix 4).

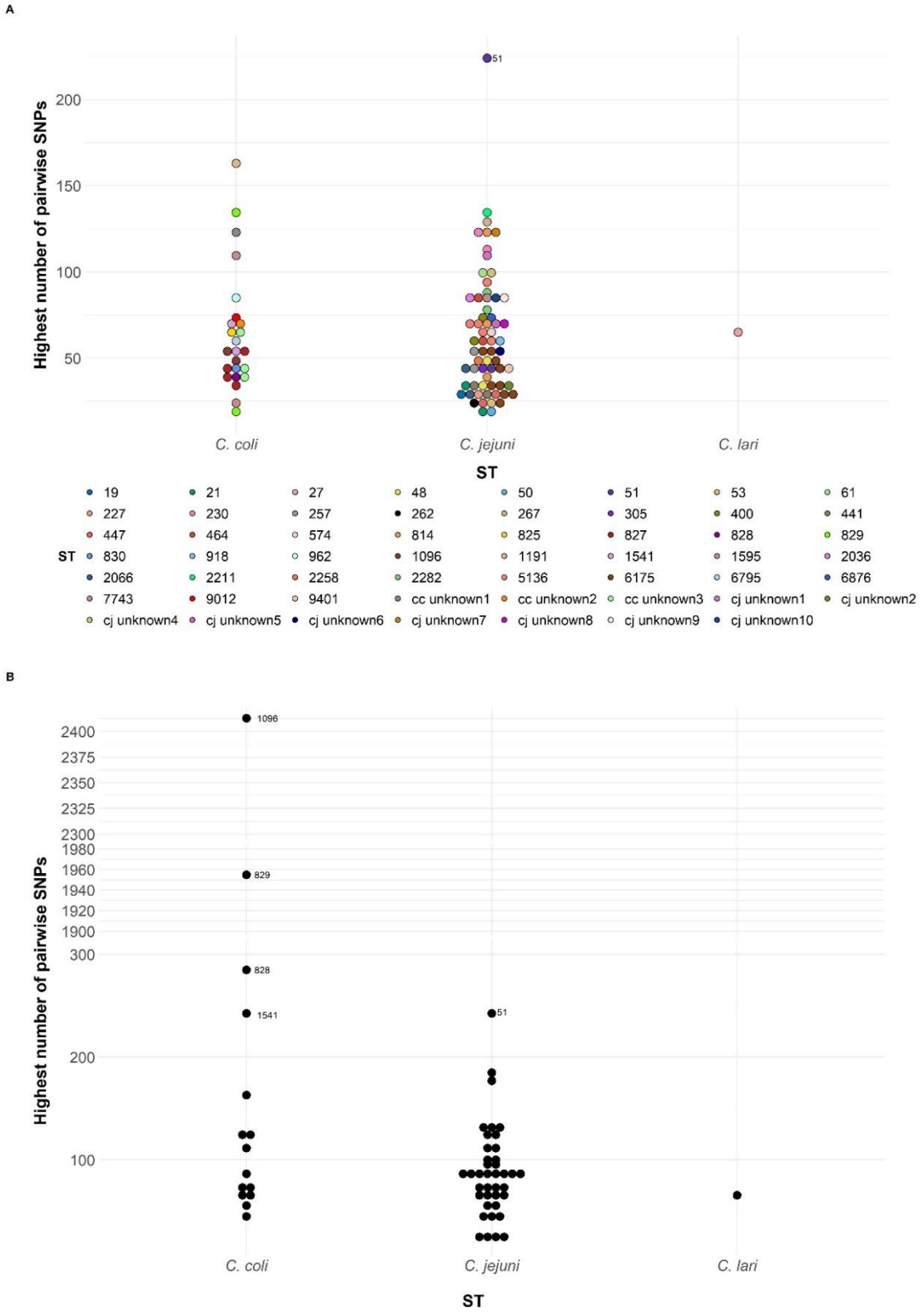

Figure 3.6: Highest non-recombinogenic pairwise single nucleotide polymorphism (SNP) difference between isolates of individual sequence types (STs)* within individual chicken samples (A) and overall (B), with STs displaying differences >200 SNPs annotated
*only ST groups consisting of more than one isolate displayed

The range of the maximum SNP distances for *C. coli* STs within samples was 17-163 (median = 53.5) and for *C. jejuni* 17-244 (median = 54.5). The highest within-sample SNP range was observed

amongst *C. jejuni* ST-51 genomes in sample CH-0339 (maximum 244 SNPs) and *C. coli* ST-1191 in sample CH-0330 (maximum 163 SNPs). The *C. lari* ST-27 isolates from sample CH-0320 did not exceed 65 pairwise SNPs.

The range of the maximum SNP distances for *C. coli* STs between samples was 46-2413 (median = 110). The highest maximum SNP distance was observed for ST-1096. On the other hand, the range of maximum SNP distances for *C. jejuni* isolates from all samples was 22-244 (median = 86) SNPs, similar to the within-sample range.

3.4.5 Diversity of individual sequence types inferred through antimicrobial resistance genotype analysis

As another measure of within-ST diversity, the genomes were screened for AMR determinants. The number of AMR determinants ranged from 0-4 per genome (Figure 3.3). *C. jejuni* and *C. coli* genomes (n = 727) displayed genotypes conferring resistance to beta-lactams (73.7%), aminoglycosides (5.36%), tetracyclines (53.9%) and quinolones (52.5%) (Table 3.4). The *C. lari* genomes (all recovered from one sample) contained the beta-lactamase gene $bla_{\text{OXA-493}}$ only. MDR genotypes (indicating resistance to at least three different classes of antibiotics) were identified in both *C. jejuni* (33.5%) and *C. coli* (11.0%) genomes, from 23 and four samples, respectively. None of the genomes contained macrolide resistance mutations.

Table 3.4: Antimicrobial resistance determinants identified in the *Campylobacter jejuni*, *Campylobacter coli* and *Campylobacter lari* genomes obtained from retail chicken meat

Species	Number of isolates	Antimicrobial group	Number of isolates with at least one resistance determinant in the antimicrobial group	AMR determinant	Nonsynonymous mutations identified	Number (%) of isolates positive for determinant	% all isolates
C. jejuni	499	Beta-lactam	409	bla_{OXA-61}		6 (1.2)	0.8
				$bla_{OXA-184}$		65 (13.0)	8.7
				$bla_{OXA-185}$		18 (3.6)	2.4
				$bla_{OXA-193}$		310 (62.1)	41.7
				$bla_{OXA-447}$		1 (0.2)	0.1
				$bla_{OXA-465}$		9 (1.8)	1.2
		Tetracycline	313	*tet(O)*		204 (40.9)	27.5
				tet(O/32/O)		109 (21.8)	14.7
		Aminoglycoside	6	*ant(6)-Ia*		6 (1.2)	0.8
		Quinolone	338	*gyrA* mutation	T86I/T86I+P104S	338 (67.7)	45.5
C. coli	228	Beta-lactam	127	$bla_{OXA-193}$		36 (15.8)	4.8
				$bla_{OXA-452}$		12 (5.3)	1.6
				$bla_{OXA-453}$		22 (9.6)	3.0
				$bla_{OXA-489}$		57 (25.0)	7.7
		Tetracycline	79	*tet(O)*		76 (33.3)	10.2
				tet(O/32/O)		3 (1.3)	0.4
		Aminoglycoside	33	*aadE-Cc*		33 (14.5)	4.4
		Quinolone	44	*gyrA* mutation	T86I	44 (19.3)	5.9
C. lari	16	Beta-lactam	16	$bla_{OXA-493}$		16 (100)	2.2

Within-ST differences in AMR genotype were observed in eight STs within individual samples (Appendix 5). In six STs, there was variation in the presence of *bla*$_{OXA-193}$ genes, whereby the coverage of the gene in one genome in the group was below the threshold for identification (90%). In one of five ST-7743 genomes from sample CH-0336, *bla*$_{OXA-184}$ was identified instead of *bla*$_{OXA-185}$. These genes only differed by one SNP. ST-51 isolates from sample CH-0339 displayed two different AMR genotypes, whereby *bla*$_{OXA-193}$ and *tet(O)* genes were identified in 14 isolates, and *bla*$_{OXA-184}$ in two isolates without *tet(O)*.

3.4.6 Further investigation of sequence types obtained with individual methods

A total of 23 unique STs consisting of more than one isolate were identified through a single culture method (direct plating or Bolton broth enrichment or CAT broth enrichment only). In order to determine whether or not the growth of the isolates of these STs was dependent solely on the culture conditions used, five isolates representing four STs (ST-830, ST-19, ST-230, ST-464) initially identified through either CAT (314-6-1, 330-6-5, 330-6-7) or Bolton broth (334-5-2, 350-2-1) enrichment only were selected and subjected to all of the condition combinations used for initial isolation, with minor modifications (3.3.3). Isolate 330-6-5 was removed from analysis due to swarming growth on all direct plating plates, meaning enumeration was not possible.

Despite initial recovery of the isolates in only Bolton or CAT broth enrichment, enumerable growth was observed in all of the conditions tested (Figure 3.7, Appendix 6). The level of growth varied between the conditions, as indicated by the log CFU/mL values; the method yielding the highest log CFU/mL was not consistently the method with which the isolates were initially isolated. There was a significant difference in the log CFU/mL values obtained between mCCDA and CBA ($p = <0.001$).

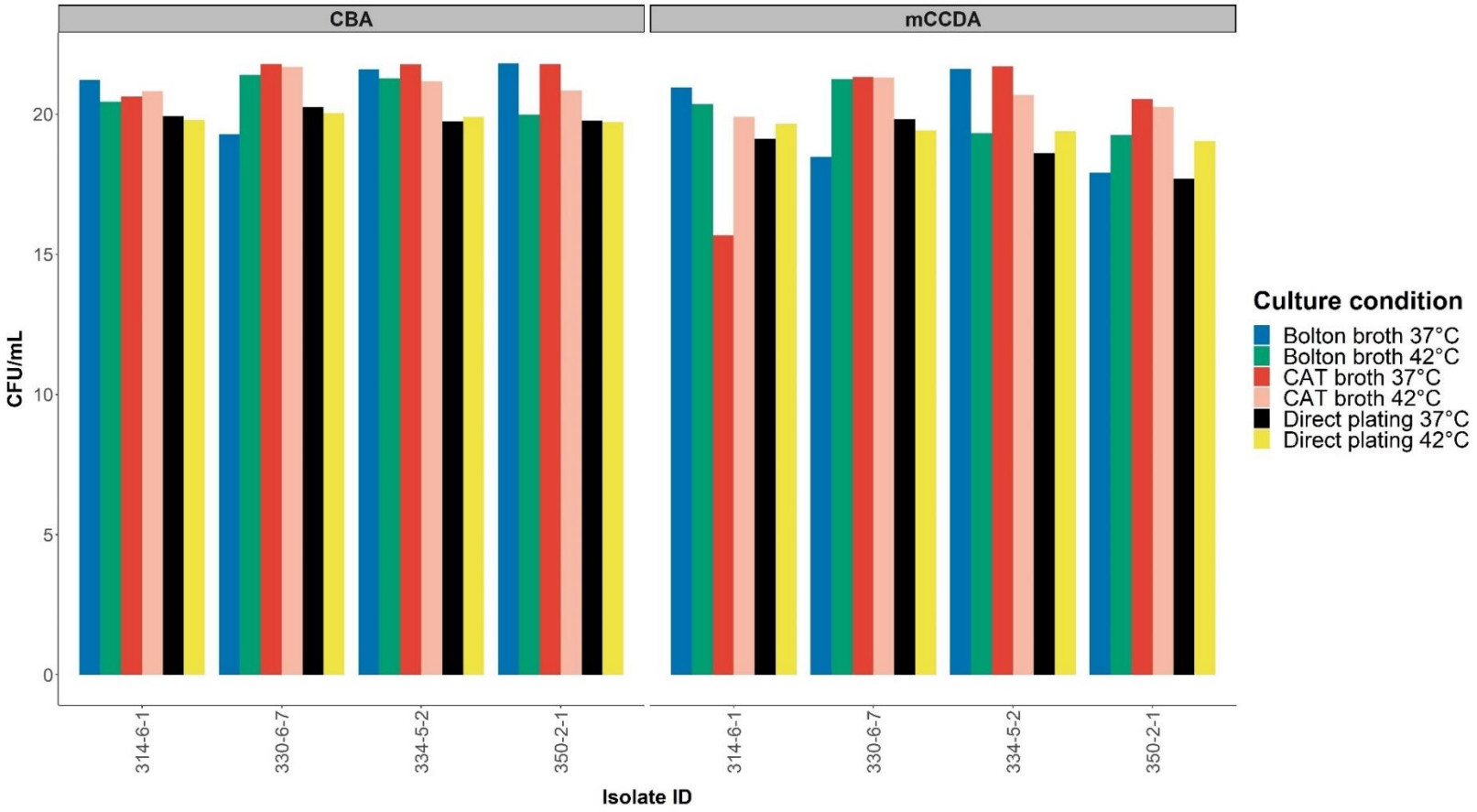

Figure 3.7: Comparison of growth (log colony forming units (CFU)/mL, using the natural log scale) of four *Campylobacter* isolates representing four sequence types (STs) initially identified through one condition only (Bolton/cefoperazone, amphotericin B, teicoplanin (CAT) broth) in different conditions tested based on Columbia blood agar (CBA) and modified charcoal-cefoperazone-deoxycholate agar (mCCDA) plate counts

3.5 Discussion

This chapter aimed to utilise a combination of methods to culture *Campylobacter* from retail chicken meat collected in Norwich, Norfolk, UK, and compare the effect of culture method combinations on *Campylobacter* recovery. The diversity of the *Campylobacter* isolates recovered was assessed with WGS, and selected isolates representing STs identified through individual methods investigated for growth in the other conditions tested. The distribution of STs within samples was used to illustrate potential implications of *Campylobacter* diversity on source attribution and outbreak investigation scenarios.

3.5.1 Comparing culture methods for *Campylobacter* recovery

Various *Campylobacter* isolation methods have been described, although each have limitations. Many microbiological standards recommend Bolton broth as either one of or the sole enrichment medium for the isolation of *Campylobacter* from chicken meat, including ISO and FDA BAM (Center for Food Safety and Applied Nutrition, 2000; International Organization for Standardization, 2017a; Harrison *et al.*, 2022). As a result, this enrichment broth has been widely utilised in previous studies (Piddock *et al.*, 2003; El Baaboua *et al.*, 2021). However, this has been suggested to be inefficient in the recovery of *Campylobacter* from samples with low *Campylobacter* abundance and a high abundance of competing microbes (Jo *et al.*, 2017; Jinshil Kim *et al.*, 2019). Bolton broth enrichment has also been associated with overgrowth of non-target organisms including *Escherichia*, which is often resistant to cefoperazone, vancomycin and trimethoprim in the selective supplement (J. Kim *et al.*, 2016; Jinshil Kim *et al.*, 2019) and highly abundant on chicken meat, thus repressing *Campylobacter* isolation. The relatively high concentration of antimicrobials in Bolton broth has also been reported to affect recovery of injured bacterial cells on food samples (Ugarte-Ruiz *et al.*, 2012). However, supplementation of Bolton broth base with alternative antimicrobial agents can result in enhanced *Campylobacter* recovery from retail meats (Moran *et al.*, 2011; Seliwiorstow *et al.*, 2016; Chon *et al.*, 2018), as evidenced in this chapter. While the majority of the samples tested were culture positive for *Campylobacter* (86.7%), only two of 45 samples yielded growth under all the broth and temperature combinations examined (Figure 3.1). Modification of Bolton broth with alternative selective supplements can enhance its efficacy, as evidenced by the isolation of *Campylobacter* with CAT broth in eight samples that were culture negative through direct plating and Bolton broth enrichment. Logistic regression modelling in this study found that *Campylobacter* recovery was significantly higher with CAT broth compared to Bolton broth (OR = 2.253; p = 0.039).

These findings indicate that culture method can have implications for reported contamination rates, as false negative results may be obtained. The *Campylobacter* contamination rate in this study was higher than the 56.0% recently reported by the FSA in a surveillance study of retail chickens obtained from major retailers in the UK (Jorgensen *et al.*, 2019). There were a number of differences between the two studies that may be in part or jointly responsible for the difference in results. The current study sampled major retailer derived meat only, with the sampling area restricted to Norfolk; thus, although samples were obtained from multiple suppliers and from retailers with national supply chains, the *Campylobacter* population identified in these samples may not necessarily be representative of all chicken meat samples available at retail in the UK. The FSA study processed neck skin samples derived from whole chickens only, whereas the current study sampled different chicken cuts. The FSA study utilised direct plating on mCCDA to isolate and enumerate *Campylobacter* from chicken. Direct plating can introduce false negative results in samples containing low abundances of *Campylobacter*, injured cells, or high abundances of competing flora (Ladely *et al.*, 2017), such as chicken meat (Ugarte-Ruiz *et al.*, 2012). In this study, only 10 (22.2%) samples were *Campylobacter* culture positive with direct plating, and recovery of *Campylobacter* with direct plating was significantly lower compared to Bolton broth enrichment (OR = 0.042, p = <0.001).

Campylobacter species commonly found on chicken and most commonly causing disease in humans are thermophilic and can therefore grow at 42°C (Summers *et al.*, 2024). Indeed, 42°C was associated with significantly higher *Campylobacter* recovery in this study (OR = 2.264; p = 0.020), consistent with previous data showing that culture at 42°C can inhibit growth of competing organisms, whereas culturing at 37°C can result in overgrowth of competing microbiota and thus reduce *Campylobacter* recovery (Gee *et al.*, 2002). This chapter strongly suggests that the isolation method used can affect *Campylobacter* recovery from retail chicken samples, thus potentially hindering source attribution and outbreak investigations, as initial isolation is the first necessary step prior to downstream typing.

3.5.2 *Campylobacter* species and sequence type diversity on retail chicken and potential impacts on source attribution and outbreak investigation

A high proportion of samples were contaminated with *C. jejuni* (80.0%), followed by *C. coli* (37.8%) and *C. lari* (2.22%). Moreover, 33.3% of samples were positive for two *Campylobacter* species. A total of 62 STs were identified, 14 of which were not seen previously, with individual samples containing between 1-8 STs. This highlights the importance of the recovery of multiple isolates per sample in order to determine both inter- and intra-sample diversity of *Campylobacter* in this important reservoir, and to facilitate effective investigations of outbreaks and attribution of

cases to the infection source. Based on the data available and random resampling, it was estimated that the number of isolates required to identify at least 95% of the total STs in the sample ranged between 1-87, indicating that samples with high ST diversity and presence of rare STs, indicated by low counts, require more intensive sampling. Common protocols for *Campylobacter* sampling may thus underestimate the diversity of this pathogen in individual food samples due to a limited number of isolates being taken.

Simulations based on the diversity and distribution of STs across samples observed in this study were used to determine the number of isolates required for the detection of a specific ST in an outbreak scenario. This revealed that up to 26 isolates per chicken meat sample would be required for the average probability of detecting the ST of interest to reach 95%. If only one isolate was sampled, the average probability of that isolate being the ST of interest was reduced to 53%. The high *Campylobacter* ST diversity observed on retail chicken meat can therefore affect the identification of food samples causing outbreaks if a limited number of isolates is sampled. It is also important to note that these are likely to be conservative estimates, as the study was limited in the number of isolates taken per culture condition, with the possibility that this caused an underestimation of the diversity in the samples.

3.5.3 Evaluation of the diversity of individual sequence type groups within samples

STs within individual samples can also be diverse, thus requiring further evaluation in order to identify specific strains in surveillance or outbreak scenarios. In this work, the SNP distances of STs within samples varied for *C. jejuni* (17-224) and *C. coli* (17-163) STs. The SNP distances of *C. jejuni* STs overall, including isolates across different samples also did not exceed 244 SNPs. *C. jejuni* SNP acquisition over time varies between lineages, with some estimations indicating accumulation of 2-8 SNPs per year (Bloomfield *et al.*, 2021; Djeghout *et al.*, 2022). The current work reported SNP distances exceeding the likely natural accumulation of SNPs within lineages present in individual samples, suggesting multiple potential contamination sources or contaminations with a genetically diverse population of *Campylobacter* belonging to the same ST.

Differences in AMR genotypes were also investigated. The *C. lari* genomes contained $bla_{\text{OXA-493}}$ genes only, though it is worth noting that the wildtype *gyrA* gene of *C. lari* ssp. *lari* may confer intrinsic resistance to quinolones (Jurinović *et al.*, 2023). Resistance to beta-lactams (73.7%), aminoglycosides (5.36%), tetracyclines (53.9%) and quinolones (52.5%) was observed in the *C. jejuni* and *C. coli* isolates recovered in this study, and MDR genotypes were evident in 33.5% of *C. jejuni* and 11.0% of *C. coli* genomes. These results are similar to a recent FSA study investigating AMR in *Campylobacter* from UK slaughterhouse and retail chicken, which found 52% of *C. jejuni* in

2018 and 50% of *C. coli* in 2017 to be resistant to nalidixic acid (quinolone) and 59.1% of *C. jejuni* between 2012-2020 and over 55% *C. coli* after 2013 to be resistant to tetracycline (Jorgensen *et al.*, 2022). Resistance to aminoglycosides was generally low (below 2%) for both species. The higher prevalence identified in the current work may be related to the collection of multiple isolates per sample, potentially selecting more aminoglycoside resistant clones from individual samples. On the other hand, the FSA study made random selections of isolates for testing, with single isolates generally representing individual samples. Although this provides insight into general resistance rates, there are gaps in knowledge surrounding the diversity in resistance profiles within individual products, which the work presented in this chapter was able to address.

In this work, the number of AMR determinants ranged between 0-4, with eight STs displaying within-ST differences in AMR genotype within samples. For six of these, the differences resulted from the gene identified falling below the ABRicate threshold (≥90% identity and coverage), which can be an artefact of the genome assembly and analysis pipeline (Hodges *et al.*, 2021; Juraschek *et al.*, 2021). The remaining two within-sample ST groups displayed differences in the bla_{OXA} gene identified (ST-7743 in CH-0336), or the presence of an additional *tet(O)* gene with a different bla_{OXA} gene (ST-51 in CH-0339). In the case of the ST-7743 genomes, the difference between the bla_{OXA} genes identified was only one SNP, which again could be a sequencing or assembly artefact. The AMR genotype differences in ST-51 from CH-0339 can potentially represent the presence of multiple strains, though differences in AMR genotype can also arise due to recombination events (Samarth and Kwon, 2020). A relatively high SNP range (11-244 SNPs) was observed in this ST; as these were non-recombinogenic SNPs, it is likely that this sample contained more than one ST-51 strain. Beta-lactam resistance genes are often not considered in national surveys (Jorgensen *et al.*, 2022), as *Campylobacter* is intrinsically resistant to most beta-lactams (Iovine, 2013), though the current chapter highlights the benefit of considering such genes for the evaluation of diversity.

Diversity of individual ST groups within samples suggests that the number of isolates required to identify a particular lineage on a chicken sample may further exceed current estimates based on ST resampling. This highlights the importance of extensive sampling of isolates from food samples implicated in outbreaks; this is not often done, with many protocols isolating or characterising one isolate per sample or a only subset of isolates obtained from the overall dataset (Food Standards Agency, 2016; Hull *et al.*, 2021; Royden *et al.*, 2021). This could be a potential reason why the majority of *Campylobacter* cases are considered to be sporadic (Liu *et al.*, 2022); if the source contains multiple strains, individuals exposed to the source may be infected with different strains that do not appear epidemiologically related.

3.5.4 Evaluation of the recovery of *Campylobacter* species and sequence types with specific culture methods

The majority of isolates recovered in this study were *C. jejuni* (67.2%), with lower numbers of *C. coli* (30.7%) and *C. lari* (2.15%) also identified. *C. jejuni* were isolated through all method combinations, whereas *C. lari* were isolated from one sample with enrichment in Bolton and CAT broth at 37°C only, despite their thermophilic nature (Summers *et al.*, 2024). The majority of *C. coli* isolates were also recovered with enrichment only, with only one isolate recovered using direct plating; enrichment has previously been shown to be more effective in the isolation of *C. coli* compared to direct plating (Ladely *et al.*, 2017). It is also possible that other species may be present in the sample that were not identified due to limitations in the method, including incubation time and media used.

It has been suggested that the dominance of certain STs, such as ST-45, ST-50 and ST-21, in the PubMLST database and in certain studies may be in part a result of isolation bias introduced by culture methods (Hetman *et al.*, 2020), providing additional support that a combination of methods is optimal for understanding the true intra-sample diversity of *Campylobacter*. In this work, 23 STs consisting of more than one isolate were only recovered using one method: either direct plating, Bolton broth or CAT broth enrichment. However, it is important to note that a random selection of isolates was picked from the agar plates during *Campylobacter* isolation, and therefore the identification of certain STs with individual methods may have occurred by chance. The growth of four isolates representing four STs initially identified either with Bolton or CAT broth enrichment only in the other isolation conditions was therefore evaluated, to determine if the growth of these STs was culture method dependent. All isolates displayed enumerable growth in all conditions, suggesting that these STs are capable of growth or at least survival in all of the tested culture conditions and therefore all of the conditions should support recovery. However, the varying counts observed may indicate that the chance of initial isolation can differ by specific method combinations. The different selective properties of the media used may have had different effects in repression of growth of non-target contaminants, leading to reduced initial isolation of certain STs in a given condition.

3.6 Conclusions

The culture method used to isolate *Campylobacter* can affect its recovery from retail chicken meat. Conventional, widely used isolation methods can yield false negative results. The recovery of multiple isolates per sample allows insight into the diversity of *Campylobacter* on retail chicken, resulting in observations with potentially significant public health impact; a high number of STs or

presence of rare STs can affect the identification of STs of interest in an outbreak scenario if a limited number of isolates is recovered. Investigation of individual ST groups within samples at the SNP and AMR genotype level indicates further diversity of individual lineages, which can have additional implications on isolate sampling during source attribution and outbreak investigation.

The findings presented here sparked interest into the extent of *Campylobacter* diversity on retail meat; while this chapter focused on diversity in context of the *Campylobacter* chromosome (through metrics such as ST and pairwise SNPs), the following chapter extends the concept to examine *Campylobacter* plasmid diversity in a selection of retail chicken meat samples. The following chapter utilises long read sequencing to resolve plasmid diversity, determine if plasmids are shared between STs, and whether or not epidemiologically relevant lineages carry genes associated with survival on retail meat or enhanced pathogenicity.

4. Chapter 4: Investigation of *Campylobacter* plasmid diversity in retail chicken using long read sequencing

Chapter contributions: The study was conceptualised by AE Mather. *Campylobacter* culturing and additional genome extraction for long read sequencing was done by AH Dziegiel. Sequencing was done by the Quadram core sequencing team. Bioinformatics analysis was done by AH Dziegiel and SJ Bloomfield, in discussion with AE Mather. The plasmid typing database was obtained from L van der Graaf-van Bloois. Visualisation was done by AH Dziegiel. The work was discussed throughout by AH Dziegiel, SJ Bloomfield, N Janecko, J Wain and AE Mather.

4.1 Introduction

Plasmids play an important role in HGT, allowing pathogens to acquire genes conferring beneficial traits. These include virulence factors and AMR genes that can enhance survival in a given niche, such as chicken meat, and subsequent human infection (Bacon *et al.*, 2000; Morita *et al.*, 2023). Although plasmids can be acquired by transformation and transduction, most often they are transferred through conjugation (San Millan and MacLean, 2017).

Campylobacter plasmids are understudied, with no typing scheme currently publicly available. Existing knowledge is thus based on comparisons of available sequences or small datasets from individual studies. Previous research comparing locally sequenced as well as publicly available plasmid sequences identified four types of *Campylobacter* plasmids, with the most prevalent being the pTet-type plasmids, with *C. coli* specific, pVir-like and small plasmids also identified (Marasini *et al.*, 2018).

It has been suggested that there may be plasmid exchange occurring between *C. jejuni* and *C. coli*, as similar plasmids have been found in both species (Marasini *et al.*, 2018; Hull *et al.*, 2023). Regarding retail meat specifically, previous work suggests that *Campylobacter* may carry megaplasmids (Marasini and Fakhr, 2014, 2016; Hull *et al.*, 2023) that can contain AMR genes as well as factors associated with enhanced chicken colonisation and pathogenicity in humans (Hull *et al.*, 2023). pTet plasmids have been found to contain *tet(O)* genes within their core genome (Marasini *et al.*, 2018; Abraham *et al.*, 2020). There is also evidence to suggest that plasmid-mediated HGT can facilitate transfer of macrolide resistance genes (Mourkas *et al.*, 2019) and multidrug efflux pumps such as CmeABC (Guo *et al.*, 2008).

However, the diversity of *Campylobacter* plasmids in individual samples is currently unknown. *Campylobacter* chromosomes display considerable diversity, even within individual samples

(Chapter 3). This chapter extends on the previous chapter to examine *Campylobacter* plasmid diversity in a selection of retail chicken meat samples, while also looking for potential niche adaptation signals or factors associated with pathogenicity. While some previous studies have utilised short read data to assemble plasmids (Hull *et al.*, 2023), this can be error prone as tools can struggle to differentiate between chromosomal and plasmid contigs, particularly if the contigs are small or contain features that can be found both on the chromosome and on plasmids (van der Graaf-van Bloois, Wagenaar and Zomer, 2021; Hull *et al.*, 2023). The current chapter thus utilises long read sequencing to fully resolve plasmids and determine their diversity amongst *Campylobacter* recovered from retail chicken.

4.2 Aims and objectives

The work outlined in this chapter aimed to:

- Identify and characterise *Campylobacter* plasmids from selected retail chicken isolates.
- Investigate plasmid diversity between STs, within and between samples.
- Characterise potential unique signatures of host adaptation or virulence factors in plasmids from an epidemiologically relevant ST.

4.3 Materials and methods

Details of sample collection, *Campylobacter* detection, isolation, genome extraction and short read sequencing are outlined in the Materials and Methods chapter (2.1-2.3). An initial selection of 42 isolates from the previous chapter, representing 31 samples, was subjected to long read sequencing.

4.3.1 Long read sequencing

Libraries were prepared using the Native barcoding kit (Oxford Nanopore Technologies) and sequenced on the MinION (Oxford Nanopore Technologies) by the Quadram core sequencing team.

4.3.2 Long read genome quality control

Filtlong v0.2.0 (https://github.com/rrwick/Filtlong) was used to filter out low quality reads, with the minimum length threshold set to 1,000 and 90% of the best reads kept. Porechop v0.2.3 (https://github.com/rrwick/Porechop) was used to trim adapters with the barcode kit specified as EXP-NBD196. The genomes were assembled with Canu v2.2 (mode: Nanopore raw, estimated genome size: 2m) (Koren *et al.*, 2017), Flye v2.9.3+galaxy0 (estimated genome size: 2m, number of polishing iterations: 5, rescue of short unassembled plasmids enabled) (Kolmogorov *et al.*, 2019), Raven v1.1.10 (https://github.com/lbcb-sci/raven) and Unicycler v0.5.1 (Wick *et al.*, 2017). The latter method used both long and short reads (trimmed with fastp v0.19.5 (Chen *et al.*, 2018)) for assembly. All assemblies were polished by aligning trimmed short reads to the assembled contigs using BWA-MEM v0.7.17-r1188 (Li and Durbin, 2009) and correcting with Pilon v1.22 (Walker *et al.*, 2014) five times.

BWA-MEM (v0.7.17-r1188) was used to align long and short reads to the long read assemblies using the -a option (output all read alignments), and samtools v1.5 (Danecek *et al.*, 2021) was used to obtain Illumina and MinION coverage, respectively. Illumina coverage represented the percentage of the assembly with a short read depth of 10 or more.

Only assemblies containing solely circular contigs were accepted. The assemblies were evaluated with ABRicate v0.9.7 (https://github.com/tseemann/abricate) using the ResFinder (Florensa *et al.*, 2022) database (built 5th November 2021) with 90% coverage and identity thresholds and results compared to the ABRricate results of the short read assemblies from the previous chapter to ensure no AMR genes were missing. One of the long read genomes (CA21CH-0336-2-1) contained a *bla*$_{\text{OXA-185}}$ gene, whereas the short read assembly contained a *bla*$_{\text{OXA-184}}$ gene; these genes differed by only one SNP, thus the long read assembly was accepted. The assemblies were analysed with Socru v2.2.4 (Page, Ainsworth and Langridge, 2020) to determine the presence of all three chromosomal fragments without repeats. For *C. coli* genomes, assemblies were accepted if they had a genome structure identifier of GS1.0 and the Socru colour was green only. For *C. jejuni* genomes, assemblies were accepted if they had a genome structure identifier of GS1.0 regardless of the Socru colour. Unlike *C. coli*, *C. jejuni* has the origin of replication and terminus on the same chromosomal fragment, which Socru calls red to indicate that the genome structure is not possible, even though it is normal for this species. Contig repeats were evaluated either as part of the Flye output, or if there were differences in the number of contigs between assemblers, BLAST (Altschul *et al.*, 1990) was used to compare the extra contigs to the others to determine if they were a repeat of a particular region. If a genome contained more than one circular contig, the largest contig was defined as the chromosome and additional contigs were assumed to be

presumptive plasmids. Filtering, assembly, Socru and BLAST analysis were performed by Samuel Bloomfield.

4.3.3 Screening of short read assemblies for plasmid contigs

To increase the number of long-read sequenced genomes available to investigate plasmid diversity, a second set of isolates were selected for long-read sequencing. This was achieved by interrogating the dataset that passed quality control (QC) and contained presumptive plasmids. The short read assemblies (Chapter 3) representing samples from which these genomes were obtained were revisited. In addition, short read assemblies from sample CH-0333 were also evaluated, as this was the sample with the highest ST diversity (Chapter 3).

The short read assemblies were analysed with RFplasmid v0.0.19 (van der Graaf-van Bloois, Wagenaar and Zomer, 2021) to identify presumptive plasmid contigs. Presumptive plasmid contigs were defined as those estimated as having 0.6 or greater proportion of "votes plasmid". The presumptive plasmid contigs were extracted from the short read assemblies. The plasmid contigs for each isolate were then concatenated to form one fasta file per genome. Mashtree v1.2.0 (Katz *et al.*, 2019) was used to build trees of concatenated plasmid contigs for isolates within each selected sample. The concatenated presumptive plasmid contigs were analysed with ABRicate v1.0.1 using a plasmid typing database (v68) developed by Linda van der Graaf-van Bloois, with 80% identity and coverage thresholds. At the time of analysis, this database consisted of sequences representing initiator and replication proteins, replication regulatory proteins, winged helix-turn-helix DNA-binding proteins, RepL domain-containing proteins and proteins of unknown function.

The mashtree trees were imported into Rstudio v2022.02.3+492 (RStudio Team, 2022). Heatmaps were added to display the genome sequenced in the first long read sequencing run, the chromosomal ST for each isolate (Chapter 3) and the plasmid typing genes identified. A second batch of isolates was selected for long read sequencing, prioritising those with evidence of diverse plasmids (differing branch length) or different plasmid types based on the ABRicate results, and to represent different STs within samples.

A total of 48 isolates were selected, one of which was in the initial long read dataset but failed QC. The isolates were cultured on CBA for 48 hours at 37°C in microaerophilic conditions as described previously (2.3). Colonies were inoculated into 10 mL Brucella broth and cultured in microaerophilic conditions at 37°C on a shaker at 300 rpm in the microaerophilic cabinet before suspending 400 µL of the overnight culture into the Maxwell RSC Cultured Cells DNA kit. DNA was extracted following manufacturer instructions and quantified (2.3). In cases where insufficient

DNA concentration was obtained (<20 ng/µL), the isolates were cultured again, this time using CampyGen sachets (Oxoid) to maintain microaerophilic atmosphere in anaerobic boxes, and the Brucella cultures were incubated statically. DNA extracts that still exhibited <20 ng/µL concentrations were repeated again, culturing the isolates in the same way, but the Brucella broth cultures were subjected to centrifugation 4,000 rpm for 30 minutes at 20°C in 15 mL centrifuge tubes. The supernatant was discarded and pellets resuspended in 400 µL of PBS, which was then loaded directly into the Maxwell cartridges for genomic extraction.

The genomes were long read sequenced on the PromethION (Oxford Nanopore Technologies) by the Quadram core sequencing team and analysed as previously.

4.3.4 Short read sequencing repeats and final assembly selection

Fifteen long read genomes containing presumptive plasmid contigs displayed low (<98%) Illumina coverage based on BWA-MEM alignment (4.3.2), indicating short read coverage gaps particularly for presumptive plasmid contigs. These genomes were re-sequenced on Illumina, and the new short reads used in Unicycler assembly and for Flye, Canu, Raven and Unicycler assembly polishing, as described above.

All of the assembly results were compiled to select one assembly per genome. Chosen assemblies had only circular contigs, no missing AMR genes when compared to short read assemblies, all chromosome fragments present only once according to Socru (and GREEN colour for *C. coli* genomes only), and contained at least one presumptive plasmid contig. If more than one assembly for each genome matched these criteria, the assembly with the highest presumptive plasmid contig coverage was selected, or the assembly containing the contigs with the smallest size was selected.

One genome (CA21CH-0336-5-2) was excluded due to having only three short reads mapping to the presumptive plasmid Flye assembly, whereas the Raven and Canu assemblies comprised of only the chromosomal contig, and the Unicycler assembly comprised of the chromosomal contig as well as linear contigs.

4.3.5 Plasmid analysis

Presumptive plasmid contigs were extracted from the long read assemblies and analysed with RFplasmid to confirm that they were plasmids. The sequences were then uploaded into Galaxy (Afgan *et al.*, 2018) for analysis with Bakta v1.9.3 (Schwengers *et al.*, 2021) using database V5.1_2024-01-19 and Roary v3.13.0 (Page *et al.*, 2015) with 95% and 99% thresholds for minimum

identity and minimum percentage of isolates a gene needs to be in to be core, respectively. The annotation files were additionally analysed with eggNOG-mapper v2.1.12 (http://eggnog-mapper.embl.de/) to obtain clusters of orthologous groups (COG). ABRicate analysis was performed on the plasmids using the VFDB database (B. Liu *et al.*, 2021) (built 5th Nov 2021) to identify virulence genes.

The gene presence/absence output of Roary was uploaded into RStudio. A distance matrix was created using Jaccard distance from vegan v2.6.4 (Oksanen *et al.*, 2022), and the plasmids clustered using ward.D linkage to create a dendrogram. Silhouette scores were calculated using the cluster v2.1.4 package (Maechler *et al.*, 2022) to obtain the optimal number of clusters. The cluster information, as well as chromosome ST (Chapter 3), COG groups and AMR genes were plotted alongside the dendrogram. For genes with multiple COG categories, the COG categories were split such that the gene was classified into both individual categories. The eggNOG data was summarised by plasmid cluster, COG category, description and preferred name.

4.3.6 Further investigation of plasmids in an epidemiologically important sequence type

Trimmed short reads of the 56 isolates belonging to ST-6175 (Chapter 3) were aligned to the long read plasmids obtained from ST-6175 isolate genomes using BWA-MEM v0.7.17.1 in Galaxy. The plasmid coverage in the short read genomes was estimated using Samtools v1.6 to determine how many of the ST-6175 genomes in the whole dataset contained the plasmid.

The Roary output (4.3.5) was combined with the cluster information to find genes unique to the cluster of plasmids from ST-6175 genomes that were present in all of the plasmids in that cluster. The hits were evaluated to determine if genes with the same annotations, suggestive of the same function, were found in other plasmid clusters. For the gene with a unique annotation, the nucleotide sequence of the gene (obtained from Flye_CA21CH-0314-3-4_final_0) was input into BLASTx v2.16.0+ (Altschul *et al.*, 1997) (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) for translation and mapping to publicly available proteins. The first 100 results with at least 95% coverage and identity were considered.

4.4 Results

4.4.1 Selection of isolates for long read sequencing

An initial subset of *Campylobacter* genomes representing a range of retail chicken samples from Chapter 3 was selected for long read sequencing (Appendix 1). From the first batch of long read genomes that passed quality control, 12 genomes representing 11 samples were determined to contain putative plasmid sequences. The samples from which these genomes were obtained were revisited to determine which of the other genomes from those samples also contained plasmid contigs, based on short read analysis. Genomes from sample CH-0333 were additionally screened, as this was the sample with the highest ST diversity. Short read contigs predicted as putative plasmids were concatenated for each individual genome, and mashtree used to visualise similarities between plasmids within individual retail chicken samples (Figure 4.1-Figure 4.12). Additional isolates were selected based on differences in chromosomal ST, plasmid typing genes identified, and branch length, in attempt to capture the full plasmid diversity within the selected samples.

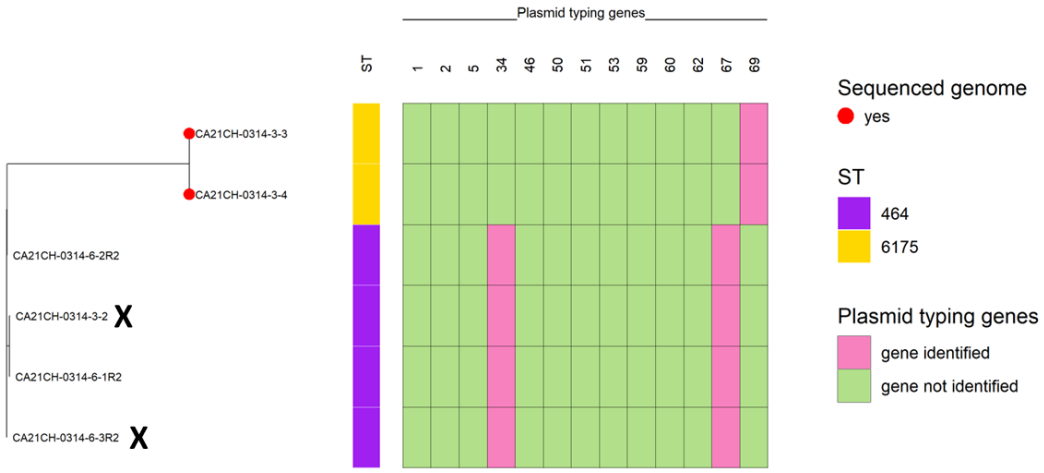

Figure 4.1: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0314, the genomes long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing

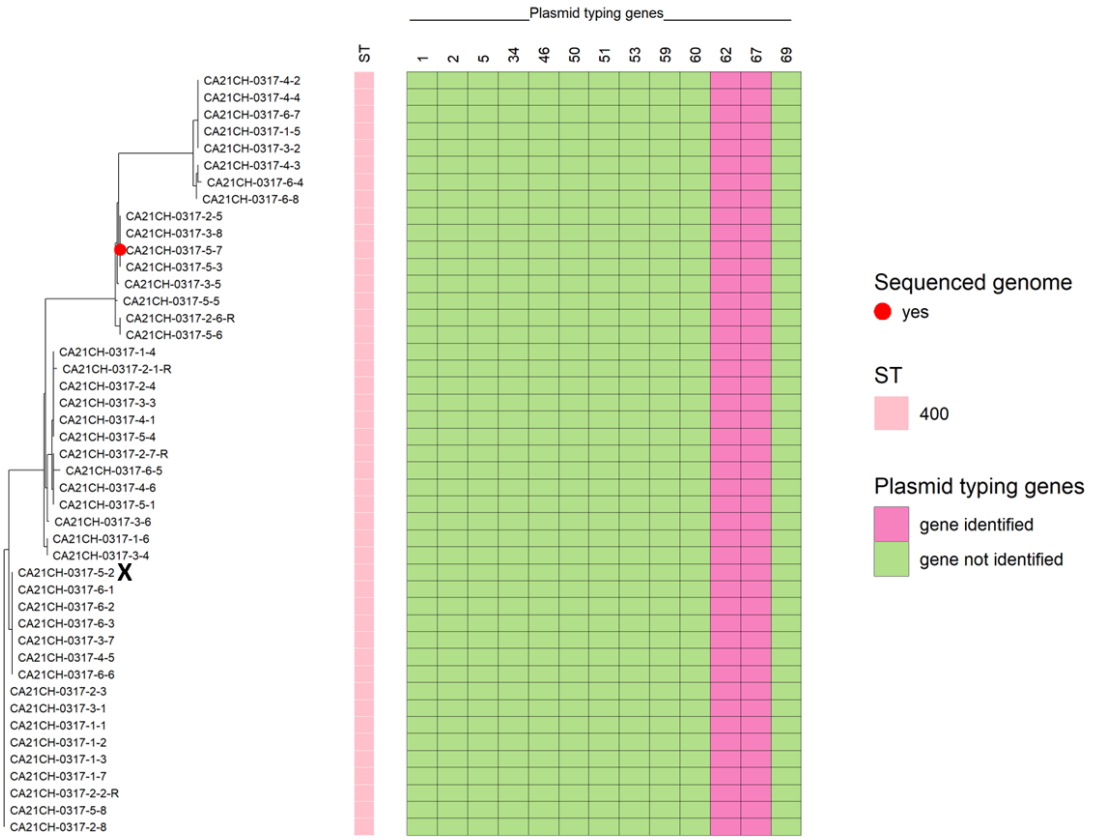

Figure 4.2: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0317, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genome marked with X was selected for the second batch of long read sequencing

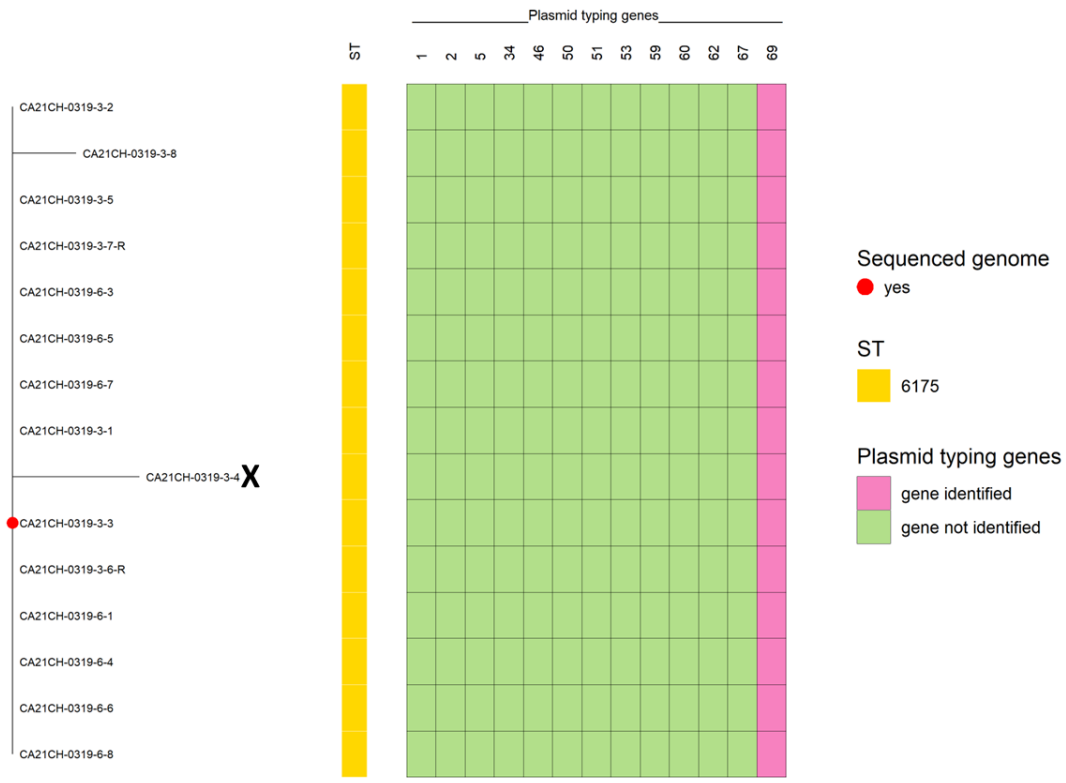

Figure 4.3: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0319, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genome marked with X was selected for the second batch of long read sequencing

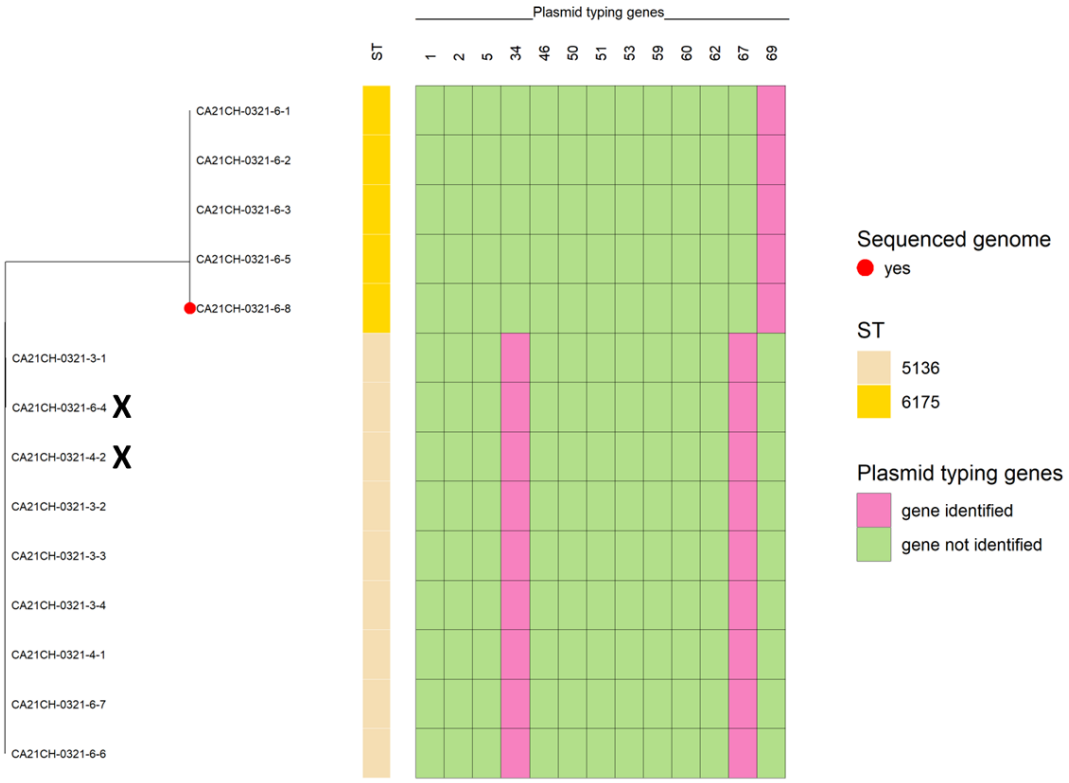

Figure 4.4: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0321, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing

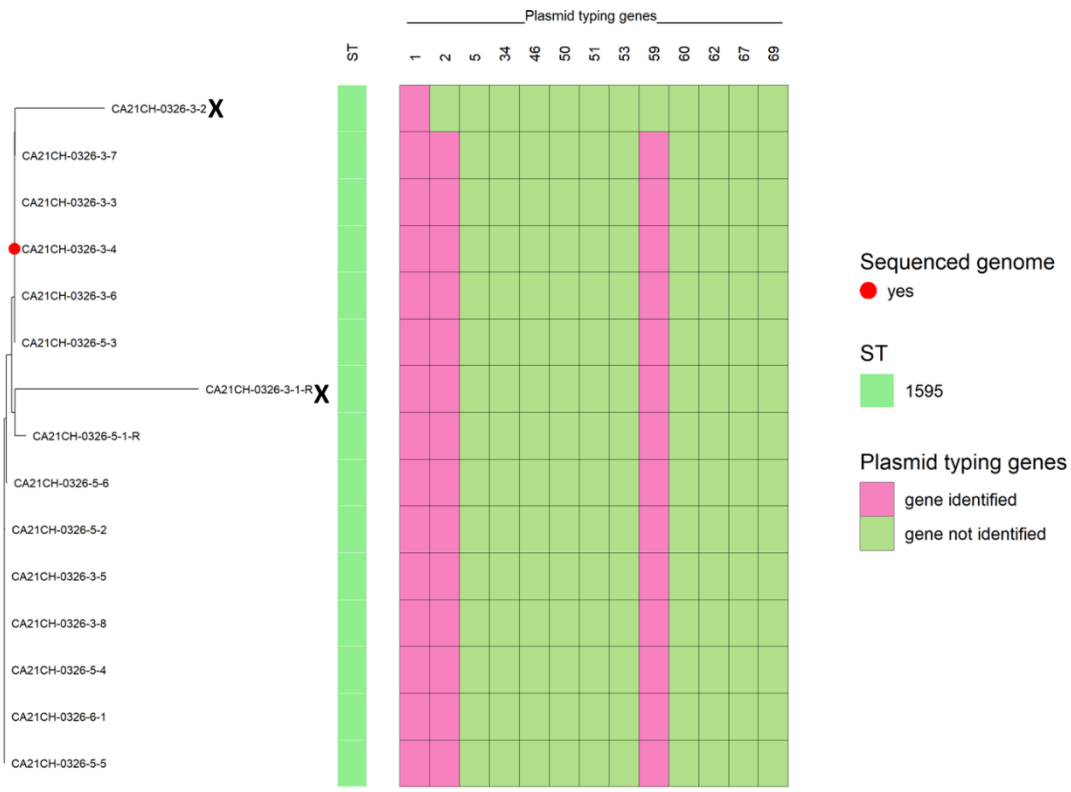

Figure 4.5: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0326, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing

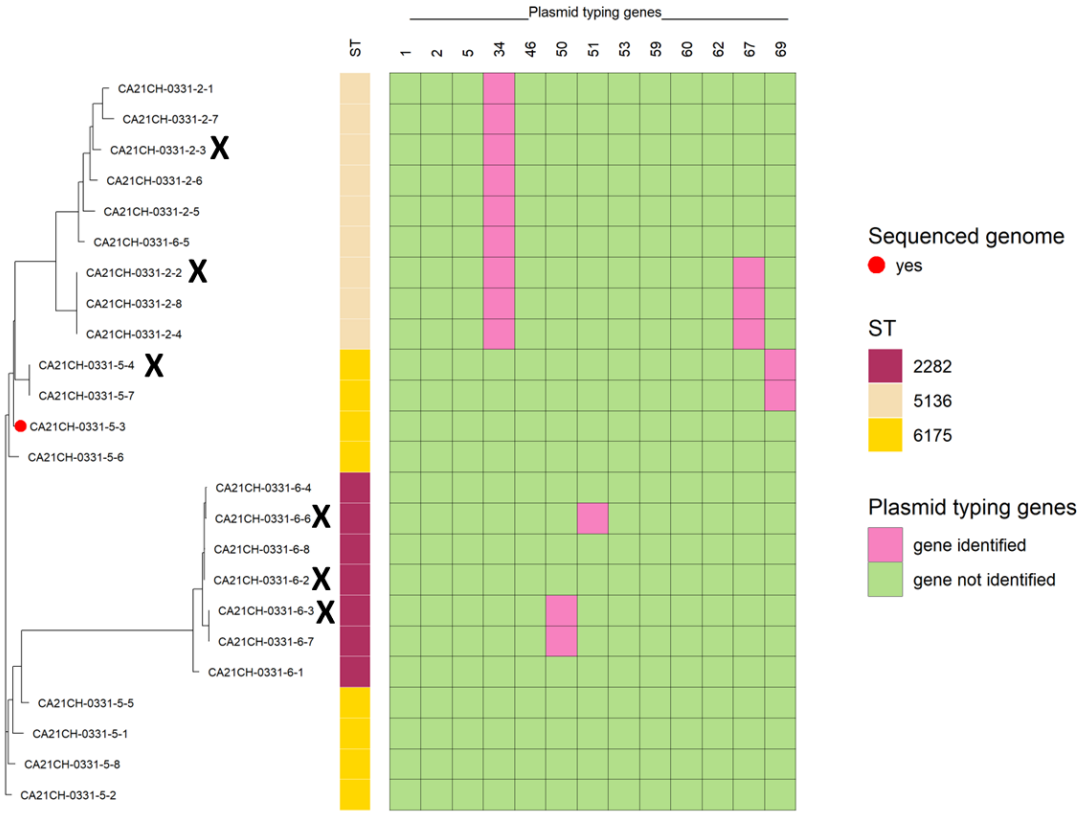

Figure 4.6: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0331, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing

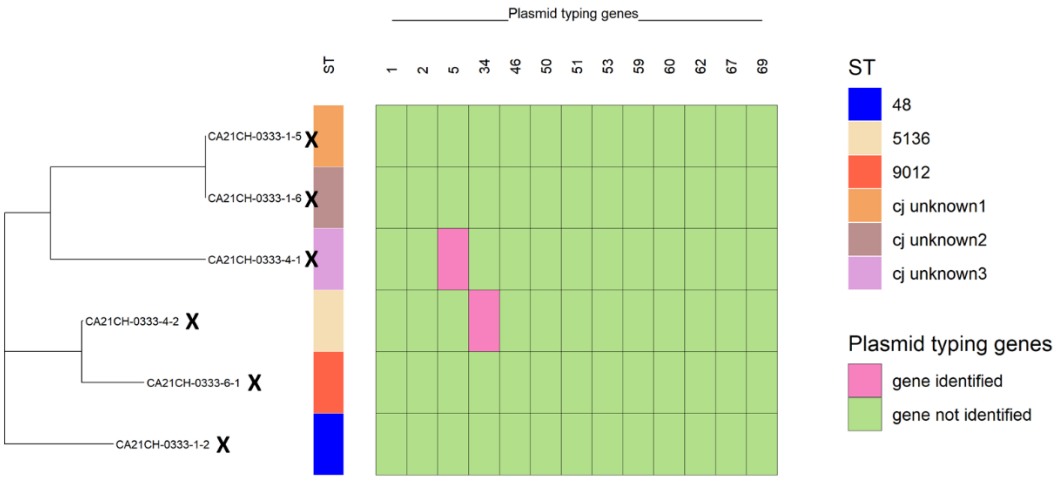

Figure 4.7: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0333, the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; this sample was additionally selected as it contained the highest chromosomal ST diversity, and genomes marked with X were selected for the second batch of long read sequencing

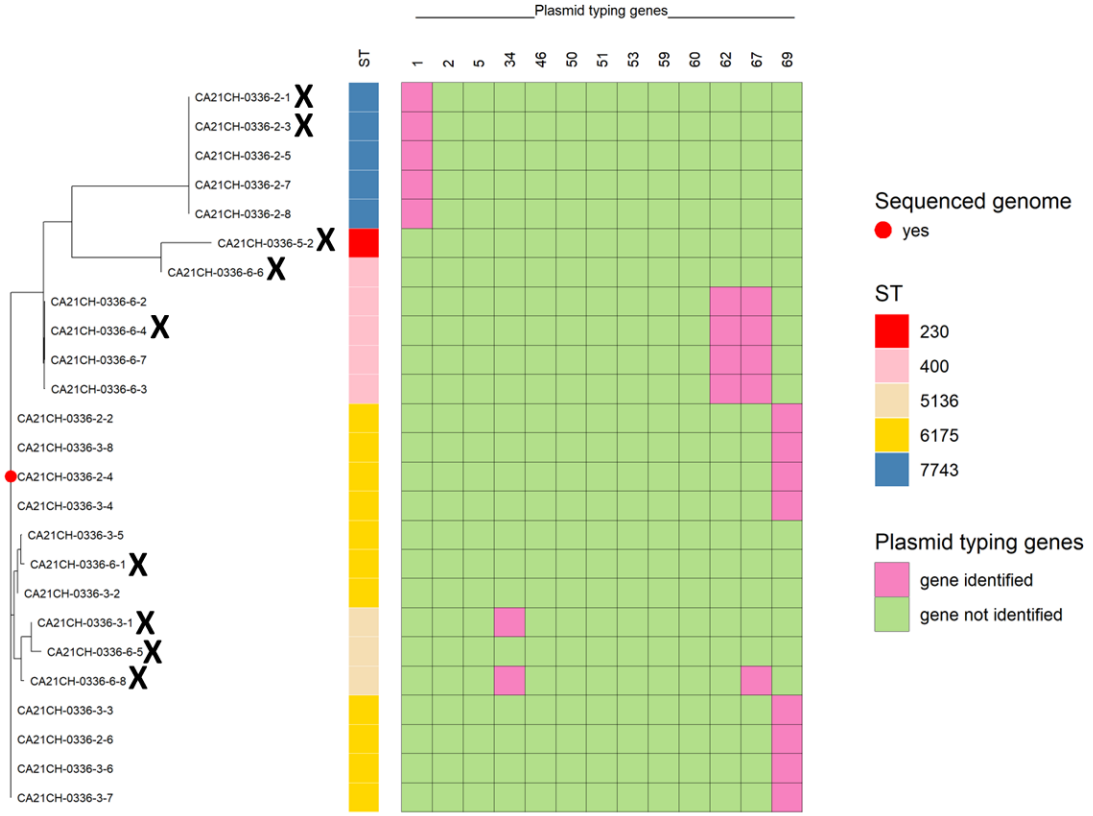

Figure 4.8: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0336, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing

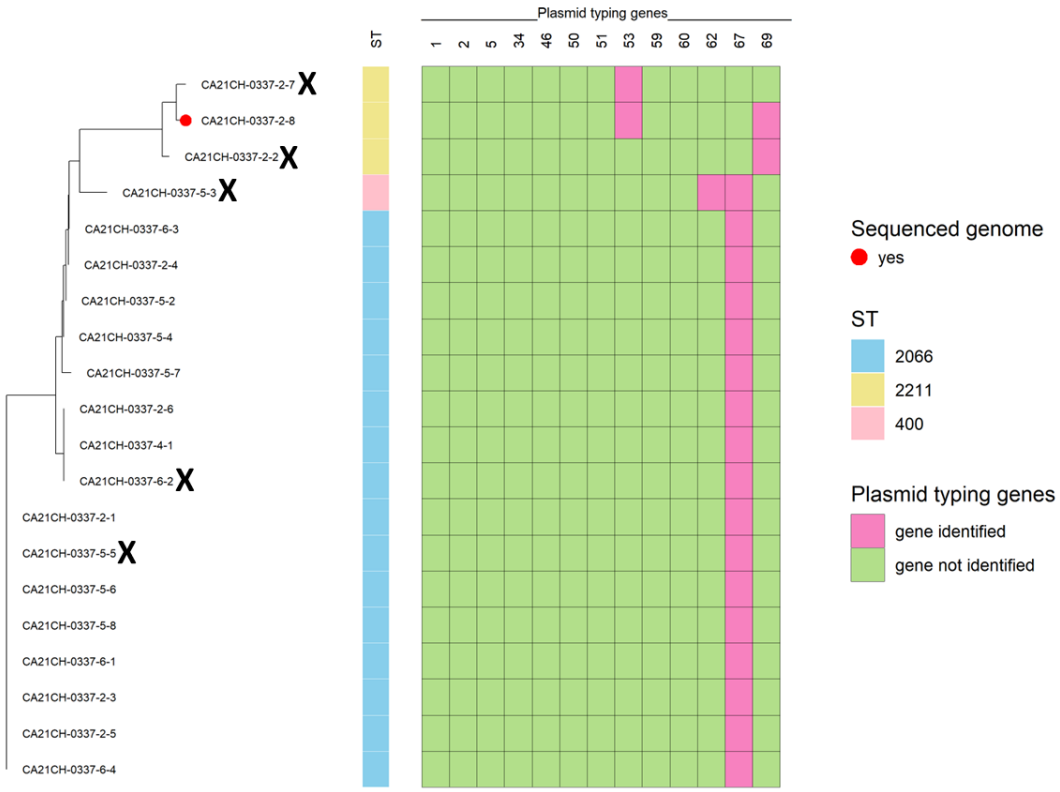

Figure 4.9: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0337, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing

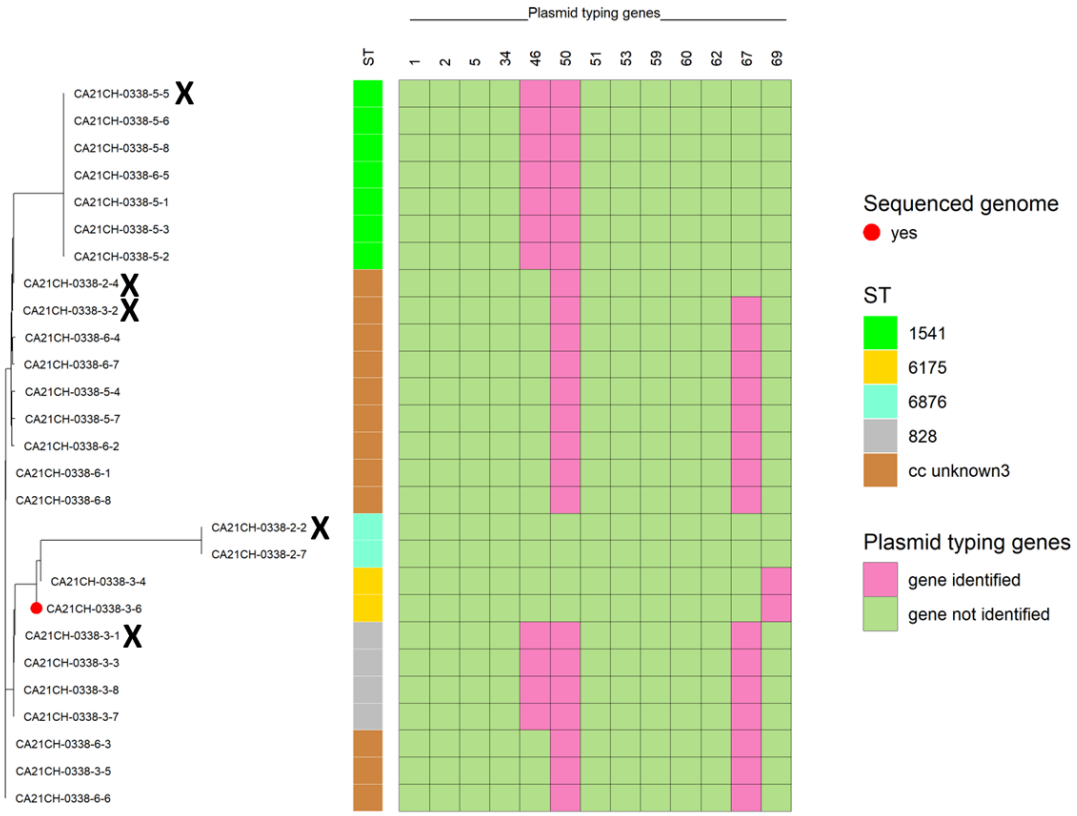

Figure 4.10: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0338, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing

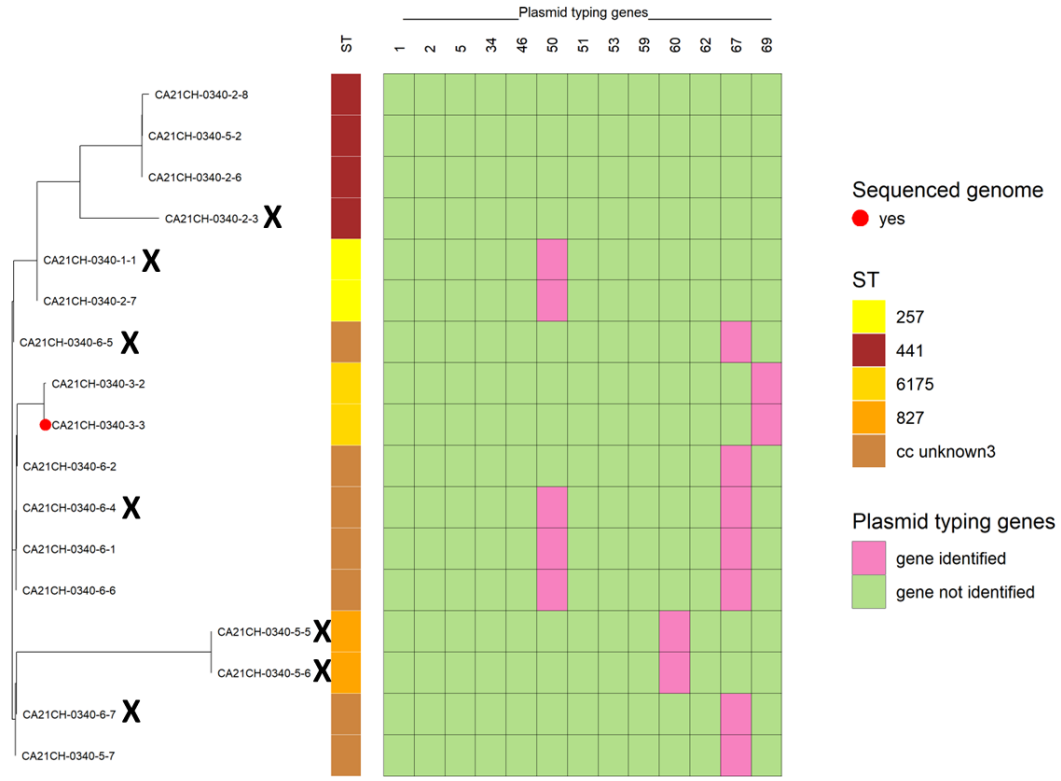

Figure 4.11: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0340, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing

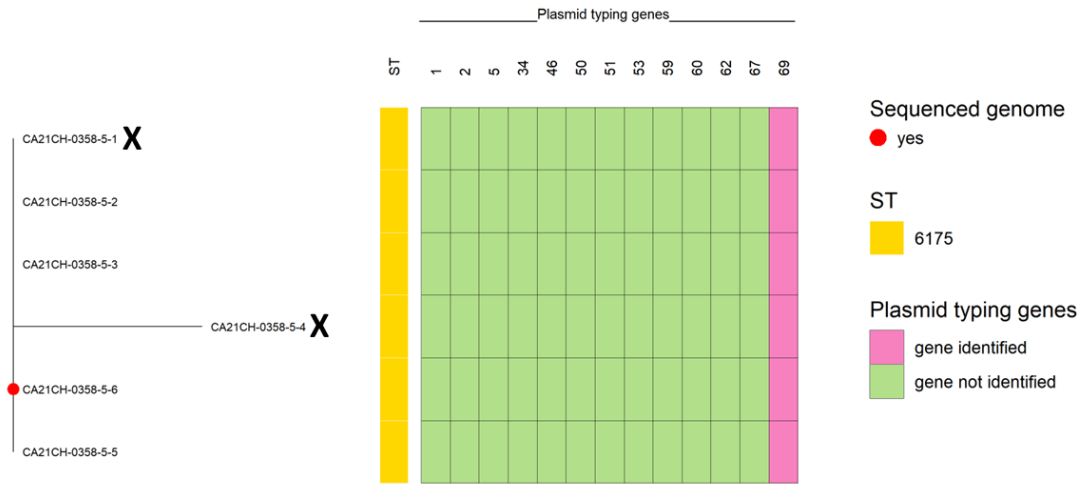

Figure 4.12: Concatenated putative plasmid contigs for each genome determined to contain them with RFplasmid within retail chicken sample CH-0358, the genome long read sequenced in the initial sequencing selection (sequenced genome), the chromosomal sequence type (ST) of the isolate, and plasmid typing genes identified in the concatenated plasmid sequences; genomes marked with X were selected for the second batch of long read sequencing

Based on the results, an additional 48 genomes were long read sequenced, including a repeat (CA21CH-0333-1-6) from the initial set that initially failed QC. Assemblies that passed QC representing 17 genomes were determined not to contain plasmid contigs at all. One additional genome (CA21CH-0336-5-2) was excluded due to having only three short reads mapping to the only assembly containing more than one circular contig. Combining all long read genomes, after QC a total of 41 long read genome assemblies with putative plasmid contigs were retained, of which six were Flye assemblies, 12 Raven assemblies and 23 Unicycler assemblies. Each assembly contained 1-3 plasmid contigs, with one assembly containing three plasmid contigs, three containing two and the rest containing one. A total of 46 plasmid sequences were obtained. The plasmid sizes ranged between 2,984-158,153 bp, with six plasmids larger than 80 kbp.

4.4.2 Plasmid clustering and diversity

The plasmid sequences were clustered based on gene presence/absence, revealing 13 clusters (Figure 4.13).

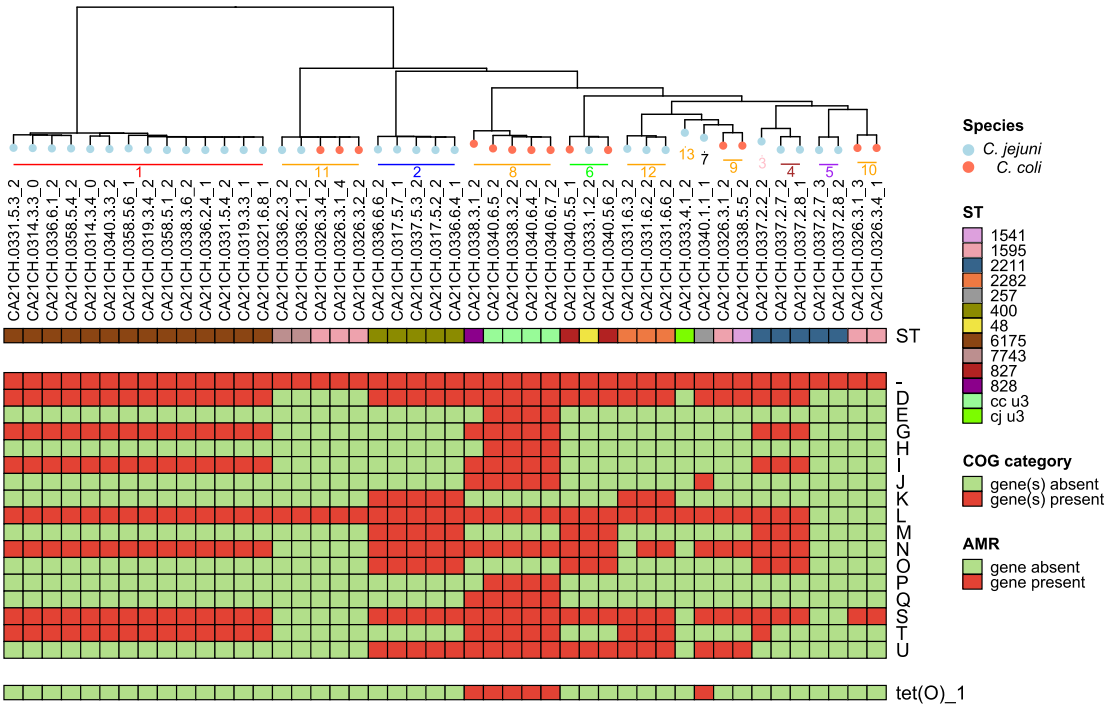

Figure 4.13: *Campylobacter* plasmid dendrogram displaying Ward clusters of Jaccard distances calculated from Roary gene presence/absence, the species of the genome, the chromosome sequence type (ST), eggNOG clusters of orthologous groups (COG) categories* from Bakta annotations, and antimicrobial resistance (AMR) genes

*COG category definitions: D=Cell cycle control, cell division, chromosome partitioning; E=Amino acid transport and metabolism; G=Carbohydrate transport and metabolism; H=Coenzyme transport and metabolism; I=Lipid transport and metabolism; J=Translation, ribosomal structure and biogenesis; K=Transcription; L=Replication, recombination and repair; M=Cell wall/membrane/envelope biogenesis; N=Cell motility; O=Posttranslational modification, protein turnover, chaperones; P=Inorganic ion transport and metabolism; Q=Secondary metabolites biosynthesis, transport and catabolism; S=Function unknown; T=Signal transduction mechanisms; U=Intracellular trafficking, secretion, and vesicular transport; -=No COG category assigned

Of the 13 clusters identified, two clusters contained more than one species and four contained more than one ST. A total of six clusters contained sequences from more than one retail chicken sample.

Within individual retail chicken samples, 1-4 plasmid clusters were identified (Table 4.1). Individual clusters were found in genomes belonging to 1-2 STs that were obtained from the same sample. For three of the 12 samples where one or more *Campylobacter* isolate was long read sequenced, only one ST was identified within the sample (Chapter 3); for another sample (CH-0319), short read analysis predicted plasmid contigs in only one ST within the sample, and in two additional samples (CH-0314 and CH-0321) plasmid sequences could only be recovered from one ST within the sample, despite RFplasmid prediction of plasmid presence in the other STs in the samples. The ability to evaluate plasmid sharing between STs within samples was thus limited for these samples.

Table 4.1: Plasmid clusters (based on Jaccard distance and Ward clustering of Roary gene presence/absence) identified in each retail chicken sample, and the chromosomal sequence types (STs) of the genomes in which they were identified

Sample	Plasmid cluster(s)	Number of sequences in cluster	Chromosome ST(s) within cluster
CH-0314	1	2	6175
CH-0317	2	2	400
CH-0319	1	2	6175
CH-0321	1	1	6175
CH-0326	9	1	1595
	10	2	1595
	11	3	1595
CH-0331	1	2	6175
	12	3	2282
CH-0333	6	1	48
	13	1	*C. jejuni* unknown3
CH-0336	1	2	6175
	2	2	400
	11	2	7743
CH-0337	2	1	400
	3	1	2211
	4	2	2211
	5	2	2211
CH-0338	1	1	6175
	8	2	828, *C. coli* unknown3
	9	1	1541
CH-0340	1	1	6175
	6	2	827
	7	1	257
	8	3	*C. coli* unknown3
CH-0358	1	3	6175

The plasmid sequence gene annotations were classified based on COG categories to identify groups of genes that may enhance survival on chicken meat or virulence. The plasmids within individual clusters generally contained the same COG categories, except for clusters 8 and 12, in which one plasmid did not contain genes representing all of the COG categories identified in the other plasmids in that cluster. All clusters contained genes that could not be assigned to a COG category, and 10 clusters contained genes of unknown function (S). All of the clusters contained genes without an assigned COG category without any description (Appendix 7), but some of these genes had a preferred name assigned. From these, a mob gene was found in cluster 13, and *tssE* genes were identified in clusters 1, 3 and 8 (Appendix 7). Among the S COG category, genes encoding bacterial mobilisation proteins (cluster 10), conjugative transfer system proteins (clusters 3 and 4), as well as type VI secretion system (T6SS) components (clusters 1, 3, 4, 8), Cag pathogenicity island proteins (clusters 7, 8, 9, 12), virulence associated protein D (clusters 8, 9, 12), plasmid stabilisation proteins (clusters 3 and 4), endonucleases (clusters 2 and 8), HicB-like

antitoxin of bacterial toxin-antitoxin system (clusters 1, 3, 4, 7, 8, 9 and 12) and RelE StbE family addiction module toxin (clusters 3, 4 and 8) were identified.

Eleven clusters contained genes involved in replication, recombination and repair (L), nine clusters contained genes involved in cell cycle control, cell division or chromosome partitioning (D), as well as genes involved in cell motility (N) and six clusters contained genes involved in intracellular trafficking, secretion and vesicular transport (U). The U COG category included genes encoding conjugation proteins and type IV secretion system (T4SS) proteins (clusters 2, 6, 7, 8, 9 and 12) and relaxases (clusters 2, 6, 7, 8, 9 and 12). Four clusters contained genes involved in carbohydrate transport and metabolism (G; clusters 1, 3, 4 and 8), as well as genes involved in lipid transport and metabolism (I; clusters 1, 3, 4 and 8), cell wall/membrane/envelope biogenesis (M; clusters 2, 3, 4 and 6), posttranslational modification, protein turnover or chaperones (O; clusters 2, 3, 4 and 6) and signal transduction mechanisms (T; clusters 1, 3, 8 and 12). Two clusters contained genes involved in translation, ribosomal structure and biogenesis (J; clusters 7 and 8) and transcription (K; clusters 2 and 12). One cluster (cluster 8) contained genes involved in amino acid transport and metabolism (E), coenzyme transport and metabolism (H), inorganic ion transport and metabolism (P), and secondary metabolites biosynthesis, transport and catabolism (Q).

Three of four plasmid clusters that contained plasmids from genomes representing more than one ST contained both T4SS and mobilisation proteins indicating conjugative plasmids (clusters 6, 8 and 9).

The plasmids were screened for AMR genes and virulence factors. Six plasmids were found to carry *tet(O)* genes, five belonging to cluster 8 and one to cluster 7. Virulence factors were not identified with VFDB in any of the plasmids.

4.4.3 Further investigation of plasmids in an epidemiologically important sequence type

ST-6175 was the most prevalent *Campylobacter* ST identified during the diversity investigations (Chapter 3), and it was also one of the most common STs in clinical samples during the sample collection period according to UK Health Security Agency (C. Swift, personal communication, January 8, 2024).

The investigations carried out here indicated that plasmids originating from ST-6175 genomes formed a distinct cluster (cluster 1; Figure 4.13). In order to determine whether or not other ST-6175 genomes carried this plasmid, the trimmed short reads from genomes belonging to ST-6175 from the previous chapter were aligned to each of the cluster 1 plasmids. The coverage ranged

between 70.7-100% (median = 99.7%), indicating presence of the plasmids in all of the ST-6175 genomes sampled. Coverage below 98% was observed for 11 genomes. This was likely due to using a different polymerase enzyme during library preparation for short read sequencing (Materials and Methods; 2.3) for two sequencing batches in which these genomes were sequenced, resulting in lower plasmid coverage.

This cluster did not feature any unique COG groups compared to the other clusters. As a result, the Roary output was investigated to identify specific genes unique to this plasmid cluster. This revealed 31 genes unique to the cluster, 15 of which were identified in all of the plasmids within the cluster (Table 4.2). Two genes encoded hypothetical proteins. Genes with the same annotations were found in other plasmid clusters for 12 of the remaining genes. The only unique gene with an annotation not seen in other clusters was putative prophage LambdaCh01 recombination protein Bet. BLASTx analysis of a representative sequence of this gene revealed that the translated sequence was most similar to recombinases found in *Campylobacter* or *C. jejuni*, sometimes further annotated as recombinase RecT or phage recombination protein Bet, indicating possible misclassification as a Lambda recombination protein. Recombinases were also found in clusters 3, 4 and 8.

Table 4.2: Unique genes identified in all plasmids within cluster 1, associated with ST-6175 genomes

Gene	Annotation	Clusters containing gene(s) with the same annotation
bdrR	BdrR	8
DUF2974 domain-containing protein	DUF2974 domain-containing protein	3, 8
DUF3800 domain-containing protein	DUF3800 domain-containing protein	1, 3, 4, 8
DUF459 domain-containing protein	DUF459 domain-containing protein	3, 4, 8
eexN	EexN family lipoprotein	2, 6, 7, 8, 9, 12
group_103		
group_106	Methyltransferase	1, 3, 4, 8
group_107	hypothetical protein	NA
group_165	hypothetical protein	NA
group_168	conjugal transfer protein TraG N-terminal domain-containing protein	1, 3, 4
group_186	Thioredoxin reductase	1, 3, 4, 8
group_187	Integral membrane protein	3, 4, 7, 8, 12
group_190	Lysozyme	1, 3, 4, 8
Putative prophage LambdaCh01, recombination protein Bet	Putative prophage LambdaCh01, recombination protein Bet	None
Small hydrophobic protein	Small hydrophobic protein	4, 8

4.5 Discussion

This chapter aimed to identify and characterise *Campylobacter* plasmids from selected retail chicken samples and investigate their diversity with long read sequencing. After an initial round of long read sequencing, 12 isolates representing 11 samples from Chapter 3 were determined to contain plasmids. A further selection of isolates was made from each of these samples, as well as another sample displaying the highest ST diversity (CH-0333; Chapter 3), to represent the plasmid diversity predicted by short read genome analysis with RFplasmid and a draft plasmid typing scheme. Interestingly, a proportion of genomes in the second long read sequencing batch were found to not contain plasmid contigs. This could potentially suggest plasmid loss, which can occur due to fitness costs associated with maintaining the plasmid (San Millan and MacLean, 2017). As these isolates were cultured again specifically for long read sequencing, it is possible that the plasmids were lost during passage. Alternatively, this could represent errors associated with plasmid prediction in short read sequencing data, which can occur when contigs are small, contain phage-encoding genes or transposases (van der Graaf-van Bloois, Wagenaar and Zomer, 2021) thus highlighting the importance of long read sequencing for plasmid investigation.

4.5.1 *Campylobacter* plasmid diversity and characteristics

From the final dataset of 46 plasmid sequences, up to three plasmids were identified from a single isolate, and these were spread across different clusters based on gene presence/absence. There was limited evidence of within-sample plasmid sharing, except for cluster 8 plasmids from sample CH-0338 that were found in two *C. coli* STs. The plasmid found in the ST-828 genome within this plasmid cluster did not contain COG category E, H and P genes identified in the plasmids from the other ST represented in that cluster, indicating variability between the STs. The current work was limited by the small sample size, and long read sequencing of additional isolates representing more retail chicken samples is required to determine the extent of within-sample plasmid sharing between STs, particularly as some samples were only represented by one ST.

However, three additional clusters were identified that were found in multiple STs, some from different species, though also from different samples. This suggests that plasmid sharing does occur between STs, as is common for other bacteria (Zamudio *et al.*, 2024), indicating genetic diversity amongst *Campylobacter* lineages that extends beyond the chromosome. This represents an underappreciated mechanism of *Campylobacter* evolution, and highlights the need for wider application of long read sequencing to understand this more fully. For three out of four plasmid clusters that were found in multiple *Campylobacter* STs, there was evidence of genes encoding T4SS machinery and mobilisation proteins suggestive of conjugation capacity. The presence of COG categories containing genes without further annotation, and the presence of hypothetical proteins, could mean that analogues are present across additional clusters. Future work should focus on identifying mobilisation and conjugation factors specifically, for example using MOB-suite tools (Robertson *et al.*, 2020) to provide further evidence of transmissibility.

The tetracycline resistance gene *tet(O)* was identified in two clusters. Previously, these genes have been found within the core genome of pTet plasmids (Marasini *et al.*, 2018), which are considered the most prevalent in *Campylobacter*. Although virulence genes were not identified using the VFDB database, there were a number of genes identified through annotation and COG clustering that suggest pathogenic capability or niche adaptation. For example, T4SS components and Cag proteins have been associated with pVir plasmids that were shown to enhance virulence in *C. jejuni* (Bacon *et al.*, 2000). Genes encoding proteins within the T6SS, including *tssE*, were also identified in four clusters. *Campylobacter* T6SS have been associated with bacterial competition and enhanced chicken colonisation that may provide a competitive advantage (Liaw *et al.*, 2019). A previous study investigating *C. jejuni* and *C. coli* plasmids from retail meat isolates also found T6SS genes amongst some megaplasmids (>80 kb) in the pTet plasmid group (Marasini *et al.*, 2018). Here, cluster 8 containing *tet(O)* genes that form the core genome of pTet plasmids (Marasini *et al.*, 2018) also consisted of megaplasmids and was one of the clusters containing T6SS

genes. However, these were also identified in other clusters and megaplasmids were also found within cluster 3 in this chapter. Genes encoding virulence-associated protein D have also been previously identified on *Campylobacter* pTet plasmids (Marasini *et al.*, 2018; Garcia-Fernandez *et al.*, 2024), though the exact function of this gene in *Campylobacter* is unclear. Here, such genes were identified in three clusters, including the suspected pTet cluster. This highlights the limitations of existing databases for the identification of virulence genes in *Campylobacter* plasmids, and the need for further investigation of the gene content of these MGEs.

The plasmids analysed in this chapter displayed considerable diversity, suggestive of multiple distinct groups, though they were not mapped against publicly available plasmid sequences. As a result, it was not possible to infer how many of them fall into the previously categorised groups (Marasini *et al.*, 2018), despite the highlighted similarities in gene content. As a lot of previous research has utilised short read sequencing only or small sample sizes (Marasini *et al.*, 2018; Hull *et al.*, 2023), the diversity of *Campylobacter* plasmids may have been previously underrepresented. This work highlights the value of long-read sequencing to determine the diversity of *Campylobacter* plasmids, though further work is needed to put this into context of existing publicly available data.

4.5.2 Investigation of plasmids in an epidemiologically relevant sequence type

The work in this chapter also aimed to identify potential signatures of host adaptation or virulence in plasmids from ST-6175, which was the most prevalent ST on retail chicken during the *Campylobacter* diversity investigations carried out in 2021 (Chapter 3). In May of that year, the UKHSA noted an increase in reported campylobacteriosis cases and launched a three-month enhanced surveillance study to characterise the *Campylobacter* strains causing infections, also finding ST-6175 to be the most prevalent ST (C. Swift, personal communication, January 8, 2024). The high prevalence of this ST on retail chicken and in submitted clinical cases suggests that there may be unique genomic signatures associated with increased persistence in the food chain, and perhaps virulence, resulting in higher prevalence in campylobacteriosis cases. The plasmids from retail chicken-derived ST-6175 genomes were all represented in one defined cluster (cluster 1), and when the plasmid assemblies were mapped to the ST-6175 short read genomes from Chapter 3, all of the isolates were found to contain the plasmid. This plasmid cluster contained genes encoding T6SS components, which could enhance persistence on chicken, though this was not unique to this cluster. As there were many genes without any COG category assigned, the Roary results were filtered for genes unique to this plasmid cluster, and present in all of the plasmids within it to identify any essential genes that might set this plasmid group apart. Although 15 unique genes were identified, the annotations for most of these genes suggested equivalent

genes were present in other clusters. Roary clusters genes based on identity, thus similar genes that display less than 95% identity cluster separately, but likely still have the same function. This could not be evaluated for two of the unique genes that encoded hypothetical proteins. As these genes are not characterised, it was not possible to determine their importance in persistence or virulence, thus further highlighting the need for laboratory characterisation and database improvements. It is also possible that this plasmid cluster contained a combination of non-unique genes that contributed to the persistence of this ST, and exploration of the wider gene repertoire could help to identify these combinations.

Importantly, clinical ST-6175 isolates were not evaluated in this chapter. Determining whether or not clinical ST-6175 isolates also contained the plasmid identified in the retail chicken isolates could help to elucidate the importance of this plasmid in the clinical setting. If the plasmid was present in clinical isolates in the same time period, this could indicate the carriage of factors important for both persistence in retail chicken and fitness within the human host. However, if the plasmid was only found in retail chicken isolates and not clinical isolates, this could indicate that the plasmid may be important for survival on retail chicken meat, with no positive effect on virulence. Nonetheless, enhanced survival on retail chicken meat may increase the chance of identification in clinical cases due to the importance of chicken meat in infection. Alternatively, the persistence of the plasmid may be facilitated through the toxin-antitoxin systems that prevent plasmid loss (Shen *et al.*, 2016), and the presence of the plasmid may not be implicated in enhanced persistence. Here, genes encoding HicB-like antitoxin were identified in seven clusters including cluster 1 associated with ST-6175. However, HicB and its associated toxin HicA have been associated with survival under stress more so than plasmid maintenance (Wadie *et al.*, 2021). This may still be significant for the persistence of the ST, and there may be other toxin-antitoxin systems present in this plasmid cluster that have not yet been characterised. It is also possible that the selective advantages of this ST may not be related to the plasmid but rather factors present on the chromosome, which should be investigated further.

4.6 Conclusions

Based on the small long read genome dataset in this chapter, there is limited evidence of plasmid sharing between STs within individual retail chicken samples. However, plasmid clusters containing multiple STs across samples were identified, suggesting that plasmid sharing does occur; long read sequencing of more isolates covering a larger number of samples is required to fully elucidate the extent of this. The identification of different plasmids, and sometimes multiple plasmids, within STs highlights high genetic diversity of the pathogen that extends beyond the chromosome. The most prevalent *Campylobacter* ST at the time of retail chicken meat sampling,

which was also the most prevalent clinically at the time, was found to carry a distinct plasmid type, suggesting possible association with persistence and enhanced pathogenicity. However, it was difficult to link any unique features to this, and the prevalence and importance of the plasmid in clinical isolates requires further study.

The high diversity and associated intensive sampling requirement demonstrated in Chapter 3 and the current chapter, alongside the difficulty of recovering *Campylobacter* from retail chicken samples, as exemplified by differences between the culture methods used (Chapter 3), strongly suggest that alternative identification methods are needed for the reliable identification of *Campylobacter* from common sources. The previous chapter also elucidated a high prevalence of the pathogen on retail chicken meat, which is a leading source of infection; this indicates that intervention strategies to reduce the burden of campylobacteriosis are required. The next chapter therefore explores the use of metagenomics for the identification of *Campylobacter* and organisms associated with its presence and absence for potential intervention.

5. Chapter 5: Using metagenomics for the identification of *Campylobacter* and organisms associated with *Campylobacter* presence and absence on retail chicken

Chapter contributions: The study was conceptualised by AE Mather. Method development for the *Campylobacter* culturing was done by AH Dziegiel, SJ Bloomfield, N Janecko, J Wain, AE Mather and AC Midwinter. Samples were collected by R Palau. Sample processing and *Campylobacter* culturing was done by AH Dziegiel. Metagenome extraction and host/bacterial qPCR was done by AH Dziegiel, SJ Bloomfield and R Palau. *Campylobacter* qPCR information and/or primer sets were obtained from B Djeghout and L van der Graaf-van Bloois. Metagenome libraries were formed by AH Dziegiel, SJ Bloomfield and the Quadram core sequencing team. Metagenomic sequencing was performed externally. Bioinformatics analysis was done by AH Dziegiel, with some example code provided by REC Luiken and AL Zomer. Visualisation was done by AH Dziegiel. The work was discussed throughout by AH Dziegiel, SJ Bloomfield, AL Zomer, REC Luiken, N Janecko, J Wain and AE Mather.

The work presented in this chapter is described in a manuscript entitled "***Campylobacter* and associated organisms on retail chicken – using metagenomics to investigate the potential of using co-resident microbes for pathogen control**" (In preparation).

5.1 Introduction

Campylobacter identification from food, such as chicken meat, is typically achieved using traditional culture methods. This is usually followed by genotyping for further characterisation of isolates, for example to determine species and STs. However, identification of *Campylobacter* with culturing can be laborious, and international standard methods can fail to recover *Campylobacter* (Ugarte-Ruiz *et al.*, 2012; Seliwiorstow *et al.*, 2016). This was discussed in Chapter 3, whereby Bolton broth enrichment and direct plating methods often used for *Campylobacter* isolation (Center for Food Safety and Applied Nutrition, 2000; International Organization for Standardization, 2017a, 2017b), were significantly less effective for *Campylobacter* recovery than CAT broth. These issues regarding initial *Campylobacter* recovery can make detailed genomic studies for the purpose of outbreak tracking and source attribution challenging (Joensen *et al.*, 2021; McCarthy *et al.*, 2021). In addition, previous chapters have strongly indicated that retail chicken samples contain diverse *Campylobacter* populations, meaning that intensive isolate sampling is needed to capture this diversity, otherwise epidemiologically important strains may be missed. This can be potentially overcome by using metagenomics for direct *Campylobacter*

detection and characterisation from chicken samples. Previous enumeration investigations have indicated that *Campylobacter* abundance on retail chicken meat is generally low (Jorgensen *et al.*, 2019), but the efficacy of detecting the pathogen with shotgun metagenomics has not been widely tested.

Metagenomics is a growing field with increasing potential and many existing applications in microbiome studies (Quince *et al.*, 2017). Shotgun metagenomic sequencing is particularly useful as it allows an unbiased insight into the entire microbial population, potentiating examination of the sample microbial composition, specific microorganisms, or functional diversity of the microorganisms present (1.2.6.2 and 1.4.3). The complex metabolic requirements of *Campylobacter* indicate that long-term survival in unfavourable conditions, such as those present on retail chicken meat, should not be viable, yet this pathogen is very commonly found in such environments. This leads to the hypothesis that its survival may be linked to the presence of other organisms or metabolites produced by the bacterial community. On the other hand, the presence of certain taxa or certain metabolic by-products may prevent *Campylobacter* survival. Metagenomics can be used for the identification of such associations, to inform potential intervention strategies to reduce the prevalence of *Campylobacter* on retail chicken.

Previous research has focused on investigations of organisms associated with *Campylobacter* in the chicken gut *in vitro* and *in vivo*, with high variability in results and seemingly a lack of consensus (Kaakoush *et al.*, 2014; Sofka *et al.*, 2015; Thibodeau *et al.*, 2015; Sakaridis *et al.*, 2018; Šikić Pogačar *et al.*, 2020). The contrasting results may be partly related to differences in methods and sample types between studies, especially as abundances of organisms can vary between different areas of the chicken gut and caecal compared to faecal samples (Stanley *et al.*, 2015).

Current literature has focused mostly on the chicken gut, and associations of different microbial taxa with *Campylobacter* presence or absence on chicken meat specifically are currently unknown. Although some researchers have included chicken carcass rinse samples from processing plants, carcass weeps representing samples after cold storage, and retail products alongside caecal samples (Oakley *et al.*, 2013), the microbiome of the samples derived from live chickens and their environment differed largely from those after processing and at retail, as not all organisms originally present survive the processing chain and subsequent storage conditions (Marmion *et al.*, 2021). Retail chicken meat represents the point closest to the consumer, thus further investigations of organisms associated with *Campylobacter* on retail chicken meat specifically are important to inform potential novel intervention strategies. These could involve treatment of meat with antagonistic bacteria, or elimination of organisms supporting *Campylobacter* survival on chicken meat, subject to regulatory approval unless the organisms are generally considered safe (European Food Safety Authority, 2007). Shotgun metagenomic

sequencing of chicken samples that have been tested for *Campylobacter* presence can therefore be a useful approach to identify any potential organisms that support or prevent *Campylobacter* survival in this epidemiologically important niche.

5.2 Aims and objectives

The work outlined in this chapter aimed to:

- Determine the efficacy of metagenomics for the direct identification of *Campylobacter* from retail chicken samples, using direct taxonomic classification as well as assembly and binning approaches.
- Determine whether or not the *Campylobacter* STs isolated from the samples can be identified in the associated metagenomes.
- Determine whether or not there is a difference in the alpha and beta diversity of *Campylobacter* culture positive and negative chicken samples, indicative of the presence or absence of microbial communities that may be enhancing or hindering *Campylobacter* survival.
- Use differential abundance analysis approaches to identify specific taxa enriched or reduced in *Campylobacter* culture positive and negative samples, and thus potentially associated with *Campylobacter* survival.
- Compare metabolic reactions and pathways in *Campylobacter* culture positive and negative metagenomes to identify any potential community-wide effects that may support or hinder *Campylobacter* survival.

5.3 Materials and methods

Details of sample collection, *Campylobacter* detection, isolation and chicken metagenome extraction and sequencing are outlined in the Materials and Methods chapter sections 2.1, 2.2 and 2.5.

5.3.1 *Campylobacter* genome analysis

Methods for *Campylobacter* DNA extraction and sequencing are outlined in the Materials and Methods chapter (2.3). General methods for the analysis of *Campylobacter* genomes obtained from the samples discussed in this chapter are outlined in the Materials and Methods chapter (2.4). Genomes obtained from the first 45 chicken samples were fully characterised, as outlined in

Chapter 3. Genomes from the remaining 22 samples were characterised to the species and ST level using Centrifuge v1.0.4_beta (nt_2018_3_3 database) (D. Kim *et al.*, 2016) and MLST v2.16.1 (https://github.com/tseemann/mlst), respectively, in order to confirm whether or not the samples were *Campylobacter* culture positive or negative and allow screening for the isolated *Campylobacter* STs in the associated metagenomes.

5.3.2 *Campylobacter* quantitative polymerase chain reaction trials

To confirm presence or absence of *Campylobacter* in the chicken samples, two primer and probe sets were trialled for qPCR of the chicken metagenomes. This included a *Campylobacter* 16S rRNA qPCR primer and probe set and a *cadF* gene primer and probe set (Table 5.1). Results of two qPCR runs using each of the primer/probe sets are reported. The primer and probe sets were obtained from Eurofins Genomics and qPCR carried out using the LightCycler 480 II instrument, using the second derivative maximum absolute quantification method. The FAM filter was applied for both analyses. Either *C. coli* (CA21CH-0340-6-6) or *C. jejuni* (CA21CH-0351-2-3) genomes were used as positive controls in the qPCR runs; both of these were genomes obtained from the chicken samples analysed in this chapter and confirmed with sequencing (Chapter 3). An isolate of *Escherichia coli* obtained from retail chicken from a previous food survey study, isolated and characterised by Raphaëlle Palau and Samuel Bloomfield, was used as a negative control. The first 16S rRNA and *cadF* qPCR runs were performed on 1 in 10 dilutions by diluting 2 µL DNA in 18 µL molecular water (Merck). For the second 16S rRNA qPCR run, the same volumes were used unless less than 2 µL was available for dilution, in which case the dilution comprised of 1 µL DNA in 9 µL molecular water or 0.5 µL DNA in 4.5 µL molecular water. The samples that were negative with the first *cadF* qPCR run were repeated using undiluted DNA. For the 16S qPCR assay, each well contained 10 µL of Precision Plus 2X qPCR Master Mix STD (Primer Design, Eastleigh, UK), 0.6 µL of the forward and reverse primers (10 µM), 0.3 µL of the probe (10 µM), 6.5 µL of molecular water (Merck) and 2 µL of sample. The cycling conditions were: initiation at 95°C for five minutes; amplification for 45 cycles with denaturation at 95°C for 10 seconds; and annealing and extension for 32 seconds at 55°C. For the *cadF* qPCR, each well contained 10 µL of the master mix, 0.4 µL of the forward and reverse primers (10 µM), 0.2 µL of the probe (10 µM) and 7 µL of molecular water with 2 µL of sample. The cycling conditions consisted of the following: pre-amplification at 95°C for 10 minutes; amplification for 45 cycles with denaturation at 95°C for 15 seconds; annealing and extension for one minute at 55°C; and cooling at 40°C for 30 seconds. Samples were considered qPCR positive if crossing point (Cp) values up to and including 40 were obtained; negative results were denoted by Cp values >40. The Cp values were corrected for diluted samples.

Table 5.1: Primer and probe sets trialled on the chicken metagenome samples

Target	Primer/ probe	Sequence	Gene target	Reference	Samples trialled
Campylobacter spp.	16S_Lv1_F	5'-CCTGAMGCAGCAACGCC-3'	16S rRNA	de Boer *et al*. (2013)	CH-0314, CH-0315, CH-0317, CH-0318, CH-0319, CH-0320, CH-0321, CH-0322, CH-0323, CH-0324, CH-0325, CH-0326, CH-0327, CH-0328, CH-0329, CH-0330, CH-0331, CH-0332, CH-0333, CH-0334, CH-0335, CH-0336, CH-0337, CH-0338, CH-0339, CH-0340, CH-0341, CH-0348, CH-0349, CH-0350, CH-0351, CH-0352, CH-0353, CH-0354, CH-0355, CH-0356, CH-0357, CH-0358, CH-0360, CH-0361, CH-0362, CH-0363, CH-0364, CH-0365, CH-0366, CH-0367, CH-0368, CH-0369, CH-0370, CH-0371, CH-0372, CH-0373, CH-0374, CH-0375, CH-0377, CH-0379, CH-0380, CH-0381, CH-0382, CH-0383
	16S_Lv1_R	5'-CGGAGTTAGCCGGTGCTTATT-3'			
	16S_Lv1_ Probe	FAM-CTCCGAAAAGTGTCATCCT-MGBEQ			
C. jejuni + *C. coli*	cadF_F	5´-CTGCTAAACCATAGAAATAAAATTTCTCAC-3'	*cadF*	Platts-Mills *et al*. (2014)	CH-0314, CH-0317, CH-0322, CH-0323, CH-0324, CH-0331, CH-0340, CH-0349, CH-0352, CH-0354
	cadF_R	5´-CTTTGAAGGTAATTTAGATATGGATAATCG-3'			
	cadF_Probe	HEX-CATTTTGACGATTTTTGGCTTGA-BHQ2			

5.3.3 Chicken metagenome analysis

5.3.3.1 Trimming, filtering, detection of antimicrobial resistance genes and classification

Paired reads were pre-processed in Galaxy (Afgan *et al.*, 2018) using fastp v0.19.5+galaxy1 (Chen *et al.*, 2018), then uploaded onto the klif2.uu.nl Utrecht University server and analysed with the FastDeMe (https://github.com/aldertzomer/FastDeMe) pipeline (release v0.1). Example code was provided by Aldert Zomer. This pipeline involved trimming with fastp v0.19.6 (removing reads with less than 80% of the average read length and PHRED scores below 20), screening for and filtering out host reads with mash screen v2.2 (Ondov *et al.*, 2016, 2019) using a mash database comprised of the vertebrate_mammalian and vertebrate_other NCBI RefSeq (https://ftp.ncbi.nlm.nih.gov/refseq/release/) databases and BioBloom tools v2.1.2-5-g8a47 (Chu *et al.*, 2014) with default parameters. These trimmed, host depleted reads were used for further analysis. For samples which contained chicken DNA proportions below the filtering threshold, the trimmed reads were used in further analysis instead. AMR genes were also identified with KMA v1.4.2 (Clausen, Aarestrup and Lund, 2018) using the ResFinder (Florensa *et al.*, 2022) database (built 25th January 2019). KMA results were filtered at the 60% template coverage and 90% query identity thresholds.

The metagenomes were classified with Kraken2 v2.1.2 (Wood, Lu and Langmead, 2019) using the k2_pluspf_20220607 database and Bracken v2.2 (Lu *et al.*, 2017) at the species and genus level, setting the read length to 150 and read threshold to include as part of Bracken report to 10. Kraken-style Bracken reports were also produced based on Bracken analysis at the species level, which contained the re-estimated abundance counts for taxa with at least 10 reads as well as the remaining taxa below the threshold. A BIOM file was created with kraken-biom v1.0.1 (https://github.com/smdabdoub/kraken-biom) to combine the reports for all samples, with the minimum taxonomic level set to species and maximum to domain. Relative abundances of *Campylobacter* and individual *Campylobacter* species were obtained from Bracken genus and species-level reports (thus retaining taxa represented by at least 10 reads) by dividing the number of reads (new_est_reads) by the total number of reads (classified and unclassified in original Kraken2 reports) in the samples, and multiplying by 100 to obtain percentages for visualisation.

5.3.3.2 Identification of sequence types, read assembly and metagenome assembled genomes

The metagenomes were uploaded into Galaxy for identification of STs with metaMLST v1.2.2 (Zolfo *et al.*, 2017). Assembly of the reads into contigs and subsequent assembly into MAGs was also performed, and the MAG quality checked as described in the Materials and Methods chapter

(2.6). The species of the MAGs were additionally classified using GTDB-tk v0.3.2 (https://github.com/Ecogenomics/GTDBTk).

5.3.3.3 Alpha and beta diversity analysis and visualisation of the most abundant families and phyla

Example code for alpha and beta diversity analysis was provided by Roosmarijn Luiken. The phyloseq v1.42.0 package (McMurdie and Holmes, 2013) was used to import the BIOM file containing the species-level counts into RStudio v2022.02.3+492 running R v4.2.3 (RStudio Team, 2022; R Core Team, 2023) and to add sample metadata. The dataset was split by domain to obtain a separate phyloseq object containing only bacterial classifications. *Shigella* species were renamed to *Escherichia coli* in the taxonomy table, and the taxa were agglomerated at the species level (with NArm=FALSE) to prevent presence of multiple *E. coli* groups in the phyloseq object. The OTU table was rarefied to obtain equal library sizes for alpha diversity analysis (using the library size of the smallest sample, resulting in normalization to 443,615 reads with 288 taxa removed). Alpha diversity was estimated using estimate_richness() from phyloseq to determine observed richness, Simpson's and Shannon's diversity indices. The alpha diversity of *Campylobacter* culture positive and negative samples was compared using Wilcoxon rank sum tests. To determine whether the low rarefaction threshold affected the alpha diversity results, alpha diversity indices were also calculated using the non-rarefied data, and compared in the same way. The microbiome v1.20.0 package (Lahti and Shetty, 2022) was used to transform counts to relative abundances for beta diversity analysis, using non-rarefied data. Ordination was performed using non-metric multidimensional scaling (NMDS) and the Bray-Curtis dissimilarity index. An overall stress value was obtained for the ordination to determine the goodness of fit, with values <0.3 considered acceptable. Permutation ANOVA (PERMANOVA) adonis2 tests were performed using the vegan v2.6-4 package (Oksanen *et al.*, 2022) to determine significant differences in the species composition between *Campylobacter* culture positive and negative samples. Multivariate homogeneity of group dispersions was assessed using betadisper and permutest from vegan. Ordination was also performed after agglomerating the data at the genus, family and phylum level (NArm=FALSE), and the groups compared in the same way. Proportions of different phyla in the samples were visualised using relative abundances. The top 10 most abundant families in the *Campylobacter* culture positive and negative samples were visualised using fantaxtic v0.2.0 (Teunisse, 2022).

5.3.3.4 Differential abundance analysis

Differential abundance analysis was carried out using non-rarefied data. Bacterial taxa not seen in at least 5% of samples were removed. Using relative abundances of the taxa in the samples (calculated in the phyloseq object containing bacterial classifications), any taxa with combined relative abundance below 0.001% across samples were also filtered out, resulting in 7,183 taxa remaining for analysis. DESeq2 v1.38.3 (Love, Huber and Anders, 2014) (local fit type of dispersions to the mean intensity), ALDEx2 v1.30.0 (Fernandes *et al.*, 2013, 2014) (Welch's t and Wilcoxon tests) and ANCOM-BC2 (using the ANCOMBC v2.0.1 package) (Lin and Peddada, 2020) were used to infer differentially abundant taxa between *Campylobacter* culture positive and negative samples. DESeq2 example code was provided by Roosmarijn Luiken. ALDEx2 analysis was also performed to identify any differentially abundant species between *C. jejuni* culture positive and negative as well as *C. coli* culture positive and negative samples. ANCOM-BC2 analysis was performed on a QIB Cloud server running R v4.2.2. For ALDEx2 and ANCOM-BC2 analysis, the filtered phyloseq object was converted to a treesummarizedexperiement object using the mia v1.6.0 (Ernst *et al.*, 2022) package. ANCOM-BC2 analysis was performed at the species level, and also at the family and phylum levels by agglomerating the data within the tool function. Significant taxa were determined by an adjusted p-value of less than 0.05 (following correction for multiple testing by each tool, using the default alpha value). DESeq2 results were also filtered at the $p<0.01$ level for visualisation. Significant differences in abundance of species determined to be differentially abundant were verified by examining box plots of the tool's normalised abundance of the taxa of interest, and their calculated relative abundance (percentage of the overall metagenome reads represented by the taxon of interest) in the filtered data.

5.3.3.5 Metabolic pathway and reaction analysis

In addition to differential abundance analysis, differences in metabolic pathway and reaction compositions were investigated. The forward and reverse metagenome reads files were concatenated, then subjected to analysis with Humann3 v3.6 using the ChocoPhlAn and uniref90_201901b translated search databases (downloaded May 2023) for functional profiling (Beghini *et al.*, 2021), producing abundances of MetaCyc metabolic pathways. The metabolic pathway outputs were converted to relative abundance (proportions adding up to 1) using humann_renorm_table, results for individual samples combined with humann_join_tables, and imported into RStudio. A phyloseq object was made using the combined pathway abundance file and sample metadata. The species-specific pathway proportions were removed, and the "unmapped" and "undesignated" proportions combined into one "unclassified" group. To obtain

reaction abundances, the gene families files containing gene abundances generated by Humann3 were converted to MetaCyc reactions (uniref90_rxn) using the humann_regroup_table script. The outputs were again converted to relative abundance and combined. The "unmapped" and "ungrouped" proportions were again collapsed into one "unclassified" group for further analysis and species-specific abundances removed. NMDS ordination was performed using Bray-Curtis dissimilarity in three dimensions for both pathway and reaction relative abundance data. Adonis2 as well as betadisper and permutest from vegan were used to determine significant differences between *Campylobacter* culture positive and negative samples.

To compare only the classified pathways and reactions, ordination was also performed on only the classified pathway and reaction proportions. For this purpose, the "unclassified" group was removed, and the remaining proportions renormalised (to sum to 1 in each sample). NMDS ordinations were performed using Bray-Curtis dissimilarity in two dimensions, and Adonis2 and betadisper permutest used for comparisons between *Campylobacter* culture positive and negative samples.

5.3.4 Analysis of metagenome negative controls

Negative controls were processed alongside each retail chicken sampling run, with details available in the Materials and Methods chapter (2.2). These negative controls were also sequenced with the chicken metagenomes (Materials and Methods chapter section 2.5). Reads of the metagenome control sequences were aligned to the phiX174 genome (SAMN04281799) using BBsplit v38.75 (https://sourceforge.net/projects/bbmap/) by Samuel Bloomfield. The reads that mapped to the genome were filtered out.

The remaining reads were subjected to pre-processing with fastp in Galaxy, the FastDeMe pipeline and classification in the same way as the chicken metagenomes. The species classifications were used to form a separate BIOM file with the metagenome samples for alpha and beta diversity visualisation. AMR genes identified in the negative controls and chicken samples were also compared.

5.4 Results

5.4.1 Retail chicken samples processed

Details of the retail chicken samples processed are outlined in Appendix 1. All 67 chicken samples were collected between March 2021 and March 2022, from chain stores in Norwich, Norfolk, UK.

5.4.2 *Campylobacter* quantitative polymerase chain reaction trials

From the 67 chicken samples obtained, 53 (79.1%) were culture positive for *Campylobacter*. Two qPCR primer and probe sets were trialled on a selection of the metagenome samples (Table 5.1), though the results were not consistent with the culture data. Two results were flagged as uncertain and were subject to visual inspection of the amplification curves to determine the result. One of the Cp values obtained was changed to >40 due to the curve resembling a negative result (Figure 5.1). For some samples for which 16S rRNA qPCR was repeated, the results were inconsistent. Importantly, sixteen out of 47 samples that were culture positive and screened with the 16S assay were consistently qPCR negative and two of six culture positive samples that were screened with the *cadF* assay were qPCR negative (Table 5.2). The negative *E. coli* control also displayed a positive result for one of the qPCR runs, thus suggesting potentially low specificity of the assay, though this was not the case in the other qPCR runs. Due to the conflicting qPCR results, comparisons in downstream analysis were made based on the *Campylobacter* culture results.

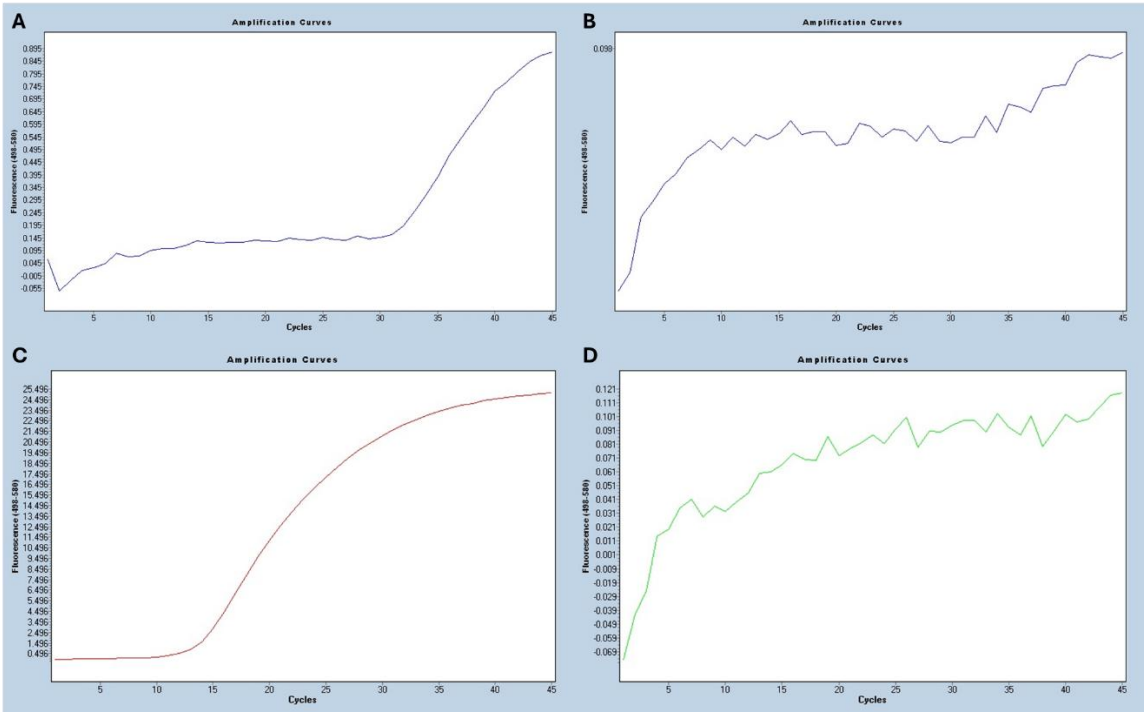

Figure 5.1: Quantitative polymerase chain reaction (qPCR) amplification curves from the 16S rRNA qPCR run that were visually inspected to evaluate the uncertain result of CH-0336 (A) and CH-0383 (B), by comparing to the curve of the positive *Campylobacter coli* control (C) and qPCR negative sample CH-0382 (D)

Table 5.2: *Campylobacter* quantitative polymerase chain reaction (qPCR) results; crossing point (Cp) values >40 represent negative results

qPCR gene target	Sample	Cp	*Campylobacter* culture result
16S rRNA run 1	CH-0362	>40	Negative
	CH-0363	>40	Negative
	CH-0364	30.88	Positive
	CH-0365	>40	Negative
	CH-0366	>40	Negative
	CH-0367	29.72	Positive
	CH-0368	31.82	Positive
	CH-0369	31.69	Positive
	CH-0370	>40	Positive
	CH-0371	31.38	Positive
	CH-0372	31.92	Positive
	C. jejuni control	11.98	-
	E. coli control	31.42	-
16S rRNA run 2	CH-0314	32.75	Positive
	CH-0315	32.85	Positive
	CH-0317	27.67	Positive
	CH-0318	31.05	Positive
	CH-0319	32.32	Positive
	CH-0320	>40	Positive
	CH-0321	31.51	Positive
	CH-0322	>40	Negative
	CH-0323	29.35	Positive
	CH-0324	30.36	Negative
	CH-0325	>40	Positive
	CH-0326	>40	Positive
	CH-0327	>40	Positive
	CH-0328	>40	Positive
	CH-0329	28.51	Positive
	CH-0330	>40	Positive
	CH-0331	27.1	Positive
	CH-0332	33.03	Positive
	CH-0333	30.59	Positive
	CH-0334	30.49	Positive
	CH-0335	>40	Positive
	CH-0336	28.31**	Positive
	CH-0337	29.45	Positive
	CH-0338	30.24	Positive
	*CH-0339	33.74	Positive
	CH-0340	32.69	Positive
	CH-0341	29.22	Positive
	CH-0348	32.26	Positive
	CH-0349	33.73	Positive
	*CH-0350	>40	Positive
	CH-0351	32.55	Positive
	CH-0352	34.65	Negative
	CH-0353	>40	Positive
	CH-0354	33.02	Negative
	CH-0355	>40	Positive

Table 5.2: *Campylobacter* quantitative polymerase chain reaction (qPCR) results; crossing point (Cp) values >40 represent negative results

qPCR gene target	Sample	Cp	*Campylobacter* culture result
16S rRNA run 2	CH-0356	28.49	Positive
	CH-0357	23.55	Positive
	*CH-0358	>40	Positive
	*CH-0360	>40	Negative
	*CH-0361	>40	Positive
	CH-0363	32.85	Negative
	CH-0364	30.6	Positive
	CH-0365	>40	Negative
	CH-0367	30.22	Positive
	CH-0368	32.44	Positive
	CH-0369	31.63	Positive
	CH-0370	27.84	Positive
	CH-0371	30.95	Positive
	CH-0372	>40	Positive
	CH-0373	>40	Negative
	CH-0374	31.07	Negative
	CH-0375	>40	Positive
	CH-0377	25.55	Positive
	CH-0379	>40	Positive
	CH-0380	>40	Negative
	CH-0381	>40	Positive
	CH-0382	>40	Negative
	CH-0383	>40**	Positive
	C. coli control	10.39	-
	E. coli control	>40	-
cadF run 1	CH-0314	>40	Positive
	CH-0317	30.42	Positive
	CH-0322	34.71	Negative
	CH-0323	33.98	Positive
	CH-0324	>40	Negative
	CH-0331	34.77	Positive
	CH-0340	>40	Positive
	CH-0349	36.68	Positive
	CH-0352	>40	Negative
	CH-0354	>40	Negative
	C. coli control	11.45	-
	E. coli control	>40	-
cadF run 2	CH-0314 (undiluted)	>40	Positive
	CH-0324 (undiluted)	>40	Negative
	CH-0340 (undiluted)	>40	Positive
	CH-0352 (undiluted)	40	Negative
	CH-0354 (undiluted)	40	Negative
	C. coli control	11.33	-
	E. coli control	>40	-

*less than 2µL of the sample was taken for the dilution; **uncertain results (visually inspected)

5.4.3 *Campylobacter* culture results and metagenome sequencing

A total of 58 metagenomes were sequenced. Of these, 47 (81.0%) were *Campylobacter* culture positive and 11 (19.0%) were *Campylobacter* culture negative. The total number of reads after host removal ranged between 944,935-53,941,872.

In order to allow inference of whether *Campylobacter* can be characterised with metagenomics, the metagenome data was compared to the genomes obtained from the samples. The *Campylobacter* culture positive samples were confirmed to contain *C. jejuni* (43 samples), *C. coli* (22 samples) and *C. lari* (one sample) with culture and genomics (Table 5.3). A total of 19 samples contained two species; 18 samples contained a combination of *C. jejuni* and *C. coli* and one sample contained *C. jejuni* and *C. lari*. A total of 63 different STs were identified in the sample set, 14 of which were novel STs. Allelic profiles of novel STs are outlined in Appendix 3.

Table 5.3: *Campylobacter* species and sequence types (STs) found in the analysed chicken metagenomes

Sample	*Campylobacter* culture result	*C. jejuni* result	*C. coli* result	*C. lari* result	STs identified
CH-0314	Positive	Positive	Negative	Negative	464, 6175
CH-0315	Positive	Positive	Negative	Negative	cj unknown6
CH-0317	Positive	Positive	Negative	Negative	400
CH-0318	Positive	Positive	Negative	Negative	5136
CH-0319	Positive	Positive	Negative	Negative	122, 6175
CH-0320	Positive	Positive	Negative	Positive	27, 464
CH-0321	Positive	Positive	Negative	Negative	5136, 6175
CH-0322	Negative	Negative	Negative	Negative	-
CH-0323	Positive	Positive	Negative	Negative	cj unknown7
CH-0324	Negative	Negative	Negative	Negative	-
CH-0325	Positive	Positive	Positive	Negative	21, 1096, cc unknown4
CH-0326	Positive	Negative	Positive	Negative	1595
CH-0327	Positive	Positive	Positive	Negative	2282, 1096
CH-0328	Positive	Positive	Negative	Negative	814, 9401, cj unknown4
CH-0329	Positive	Negative	Positive	Negative	828, 829
CH-0330	Positive	Positive	Positive	Negative	441, 830, 1191
CH-0331	Positive	Positive	Negative	Negative	2282, 5136, 6175
CH-0332	Positive	Positive	Negative	Negative	51, 61
CH-0333	Positive	Positive	Positive	Negative	48, 53, 267, 5136, 9012, cj unknown1, cj unknown2, cj unknown3
CH-0334	Positive	Positive	Negative	Negative	53, 230, 257, 449, cj unknown2
CH-0335	Positive	Positive	Positive	Negative	21, 227, 2254, 6175, cc unknown1
CH-0336	Positive	Positive	Negative	Negative	230, 400, 5136, 6175, 7743
CH-0337	Positive	Positive	Negative	Negative	51, 400, 2066, 2211
CH-0338	Positive	Positive	Positive	Negative	50, 828, 1541, 6175, 6876, cc unknown3
CH-0339	Positive	Positive	Positive	Negative	51, 2258, 6795, cj unknown5
CH-0340	Positive	Positive	Positive	Negative	257, 441, 447, 827, 6175, cc unknown3
CH-0341	Positive	Positive	Positive	Negative	21, 447, 827, 1595, 6175, cc unknown3, cj unknown8
CH-0348	Positive	Negative	Positive	Negative	827
CH-0349	Positive	Positive	Positive	Negative	814, cc unknown2
CH-0350	Positive	Positive	Negative	Negative	19, 48, 50, 2066, 8334
CH-0351	Positive	Positive	Positive	Negative	48, 825, 829, cj unknown8
CH-0352	Negative	Negative	Negative	Negative	-
CH-0353	Positive	Positive	Positive	Negative	814, 827
CH-0354	Negative	Negative	Negative	Negative	-
CH-0355	Positive	Positive	Negative	Negative	9401
CH-0356	Positive	Positive	Negative	Negative	305

Table 5.3: *Campylobacter* species and sequence types (STs) found in the analysed chicken metagenomes

Sample	*Campylobacter* culture result	*C. jejuni* result	*C. coli* result	*C. lari* result	STs identified
CH-0357	Positive	Positive	Negative	Negative	918, 2036, cj unknown9
CH-0358	Positive	Positive	Negative	Negative	6175
CH-0360	Negative	Negative	Negative	Negative	-
CH-0361	Positive	Positive	Positive	Negative	51, 574, 1541, cj unknown10
CH-0363	Negative	Negative	Negative	Negative	-
CH-0364	Positive	Positive	Negative	Negative	51, cj unknown8
CH-0365	Negative	Negative	Negative	Negative	-
CH-0367	Positive	Positive	Negative	Negative	cj unknown8
CH-0368	Positive	Positive	Negative	Negative	354
CH-0369	Positive	Positive	Negative	Negative	6175, cj unknown2
CH-0370	Positive	Positive	Negative	Negative	354, cj unknown8
CH-0371	Positive	Positive	Positive	Negative	400, 855
CH-0372	Positive	Positive	Positive	Negative	19, 855
CH-0373	Negative	Negative	Negative	Negative	-
CH-0374	Negative	Negative	Negative	Negative	-
CH-0375	Positive	Negative	Positive	Negative	-
CH-0377	Positive	Positive	Positive	Negative	47, 855, 2254
CH-0379	Positive	Positive	Positive	Negative	574, 828
CH-0380	Negative	Negative	Negative	Negative	-
CH-0381	Positive	Positive	Positive	Negative	828, 2258
CH-0382	Negative	Negative	Negative	Negative	-
CH-0383	Positive	Positive	Negative	Negative	1231

5.4.4 Identification of *Campylobacter* from chicken metagenomes

One of the aims of this chapter was to determine whether or not metagenomics can be used to identify *Campylobacter* in retail chicken samples. In an assembly-free approach, metagenome reads were classified with Kraken2 and Bracken and relative abundance calculated. The proportions of Kraken2 unclassified reads ranged between 6.45-55.9%. The proportion of reads classified as *Campylobacter* with Bracken analysis at the genus level ranged between 0.000862-0.466% (Figure 5.2). The highest relative abundance of *Campylobacter* in a culture negative sample was 0.00886%, and the lowest relative abundance of *Campylobacter* in a culture positive sample was 0.000862%, with 48 samples displaying proportions between these values.

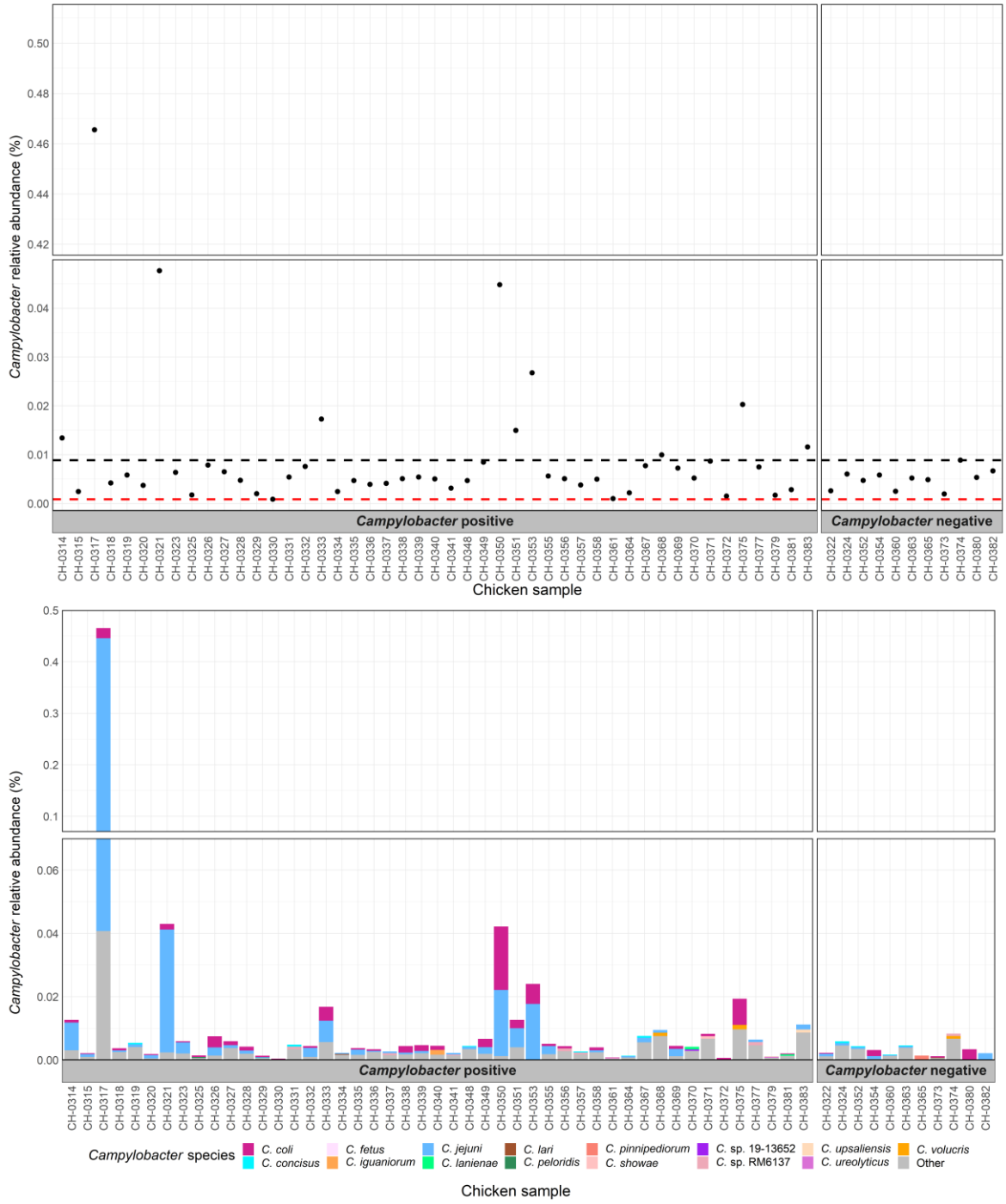

Figure 5.2: Relative abundance of the *Campylobacter* genus in each chicken meat metagenome determined following genus-level Bracken analysis* (A) and the relative abundance of the two most abundant *Campylobacter* species in each chicken meat metagenome determined following species-level Bracken analysis, with the rest summarised into "other"*

*the red dashed line represents the lowest relative abundance of *Campylobacter* in a culture positive sample and the black dashed line represents the highest relative abundance of *Campylobacter* in a culture negative sample; **only species with at least 10 reads included as part of the Bracken report

Bracken analysis at the species level revealed that the most abundant species within each sample represented 0.000170-0.405% of the metagenome reads. Between 1-51 (median = 14.5) species were identified within individual samples (Supplementary Table S2). For 27 samples, the most

abundant species in the metagenome was not identified through culture and genomics. These species were *C. jejuni*, *C. coli*, *C. iguanorium*, *C. pinnipediorum*, *C. volucris*, *C. lanienae*, *C. showae, C. ureolyticus* and *C.* sp. RM6137.

To determine if metagenomics can be used for further characterisation of *Campylobacter* at the ST level, the metagenomes were screened for bacterial STs. MetaMLST identified *C. jejuni* ST-400 in sample CH-0317, which also contained the highest proportion of *Campylobacter*. This ST was also identified through culture and genomics. This was the only sample that contained a *Campylobacter* MAG, identified with both Metabat2 and Maxbin2 (Table 5.4).

Table 5.4: Quality control metrics for the *Campylobacter* metagenome assembled genomes (MAGs) obtained from sample CH-0317 using Metabat2 and Maxbin2

Tool	Completeness (%)	Contamination*	Size (bp)	GC (%)	CAT bins classification	GTDB-tk classification
Metabat2	99.48	0.45	1,622,502	30.42	*Campylobacter*	*Campylobacter*_D *jejuni*
Maxbin2	99.48	3.55	1,998,775	30.09	*Campylobacter*	*Campylobacter*_D *jejuni*

*higher value indicates increased abundance of multi-copy marker genes

5.4.5 Alpha and beta diversity

The large range in the relative abundance of *Campylobacter* and overlap between culture positive and negative samples further indicated the requirement to use the culture results to define samples as either *Campylobacter* positive or negative in downstream analysis. This involved the investigation into microbial signatures associated with *Campylobacter* presence and absence. To determine whether there were any differences in the bacterial composition and diversity between *Campylobacter* culture positive and negative samples at the global level, alpha and beta diversity metrics were calculated. The total number of bacterial reads for the samples ranged between 443,615-47,737,177.

The observed alpha diversity measure indicated a range of 1,670-5,963 (median = 3,117) taxa in *Campylobacter* culture positive and 2,187-4,471 (median = 3,681) taxa in *Campylobacter* culture negative chicken metagenomes after subsampling (Figure 5.3). There was no significant difference in the observed richness (p = 0.39), Shannon (p = 0.64) or Simpson (p = 0.67) diversity indices between the two groups according to Wilcoxon rank sum tests.

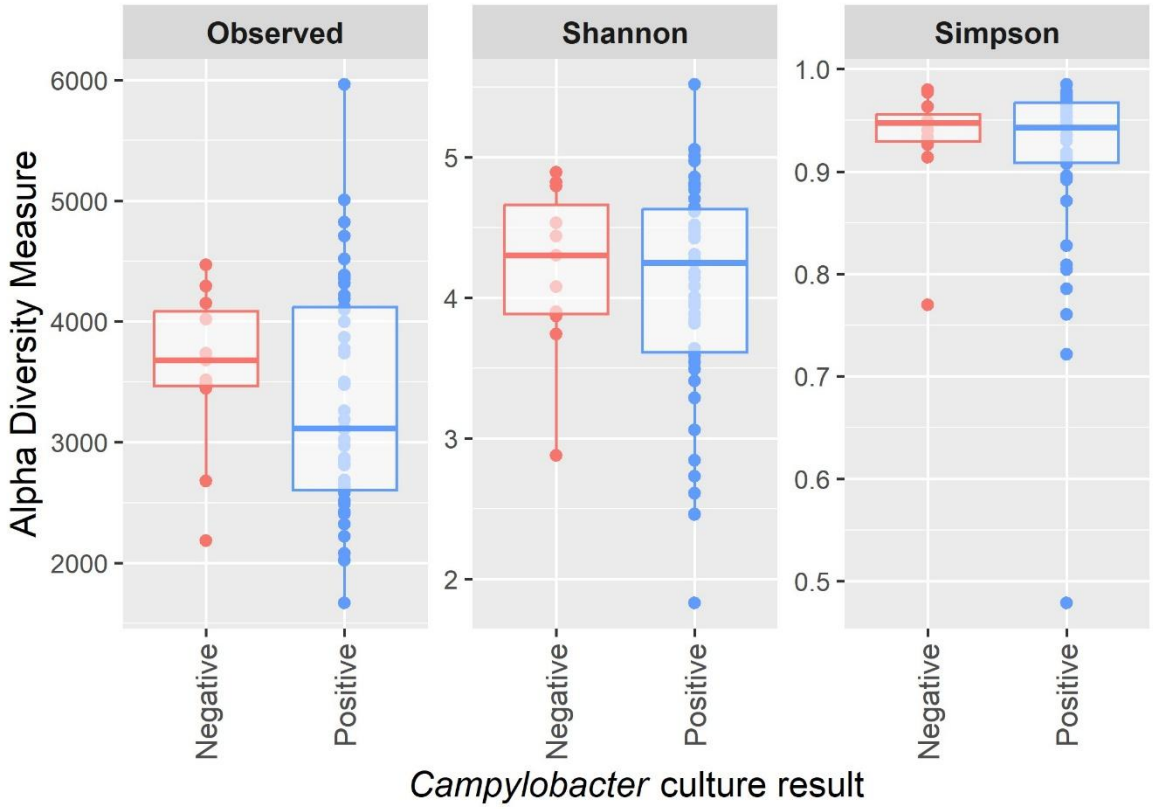

Figure 5.3: Alpha diversity indices (observed richness, Shannon and Simpson) and distributions of the *Campylobacter* culture positive and negative chicken meat metagenomes based on estimation from bacterial species-level Bracken results, after rarefying to the smallest library size

The rarefaction threshold was low, which can result in the underrepresentation of diversity in some samples (Figure 5.4). Alpha diversity indices were also calculated using non-rarefied data as input, to determine whether the low rarefaction threshold affected the results. There was no significant difference in observed richness (p= 0.61), Shannon (p= 0.62) or Simpson (p= 0.67) diversity indices between *Campylobacter* culture positive and negative groups based on non-rarefied data.

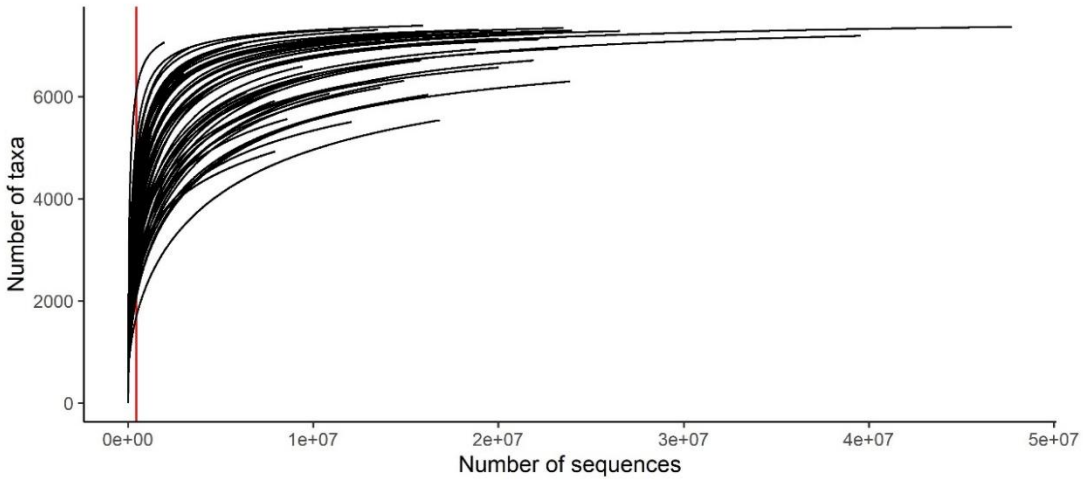

Figure 5.4: Rarefaction curve displaying the number of bacterial taxa identified as the number of sequences increased*
*the sample with the lowest number of sequences (rarefaction threshold - 443,615 reads) marked in red

Despite there being lack of evidence of significant difference within *Campylobacter* culture positive and negative samples, there may be differences in the bacterial composition between the different samples. Sample beta diversity was investigated using non-rarefied data, to enable comparisons of microbial compositions between samples. NMDS ordination revealed that the sample microbial compositions were diverse, but there was no significant difference in the beta diversity of *Campylobacter* culture positive and negative samples (Adonis2 p = 0.66, betadisper permutest p = 0.22; Figure 5.5).

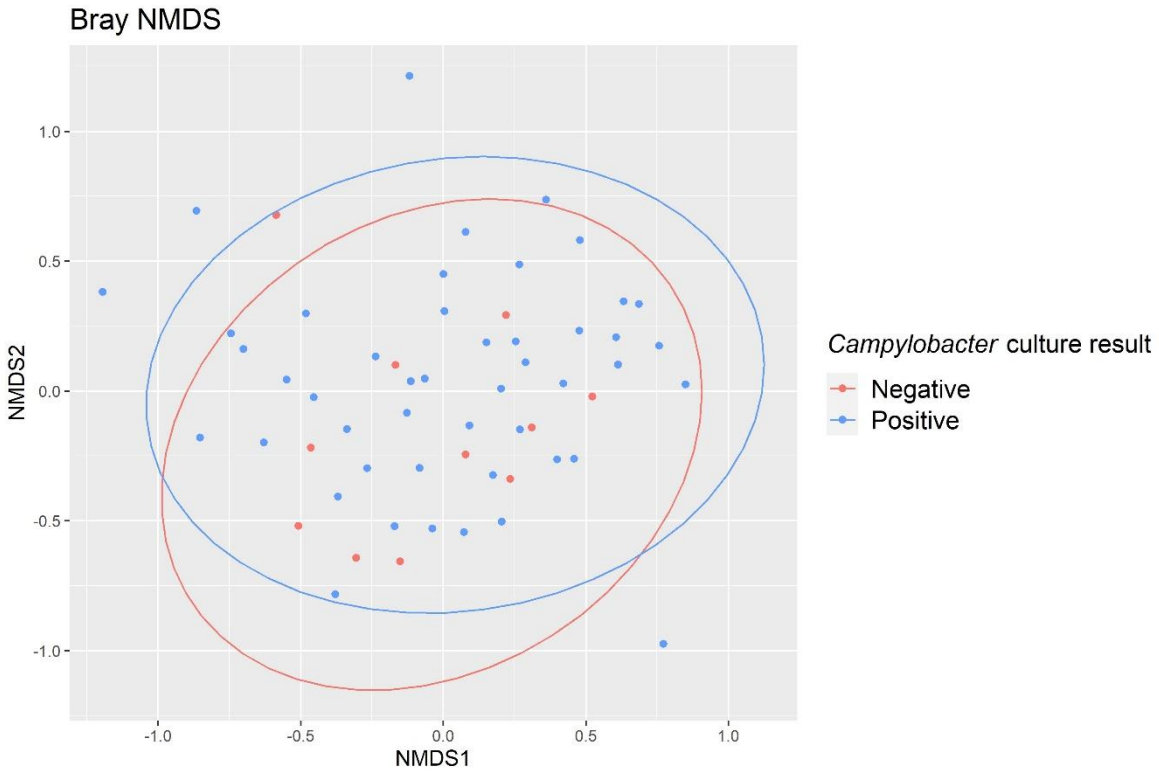

Figure 5.5: Non-metric multidimensional scaling (NMDS) ordination plot showing differences in beta diversity using Bray-Curtis dissimilarity in bacterial species composition based on Bracken results between *Campylobacter* culture positive and negative samples*
*stress = 0.215

Beta diversity was also compared at higher taxonomic levels. There was also no significant difference in beta diversity based on agglomerated data at the genus (Figure 5.6; Adonis2 p = 0.43, betadisper permutest p = 0.14), family (Figure 5.7; Adonis p = 0.44, betadisper permutest p = 0.22) and phylum (Figure 5.8; Adonis2 p = 0.79, betadisper permutest p = 0.57) level.

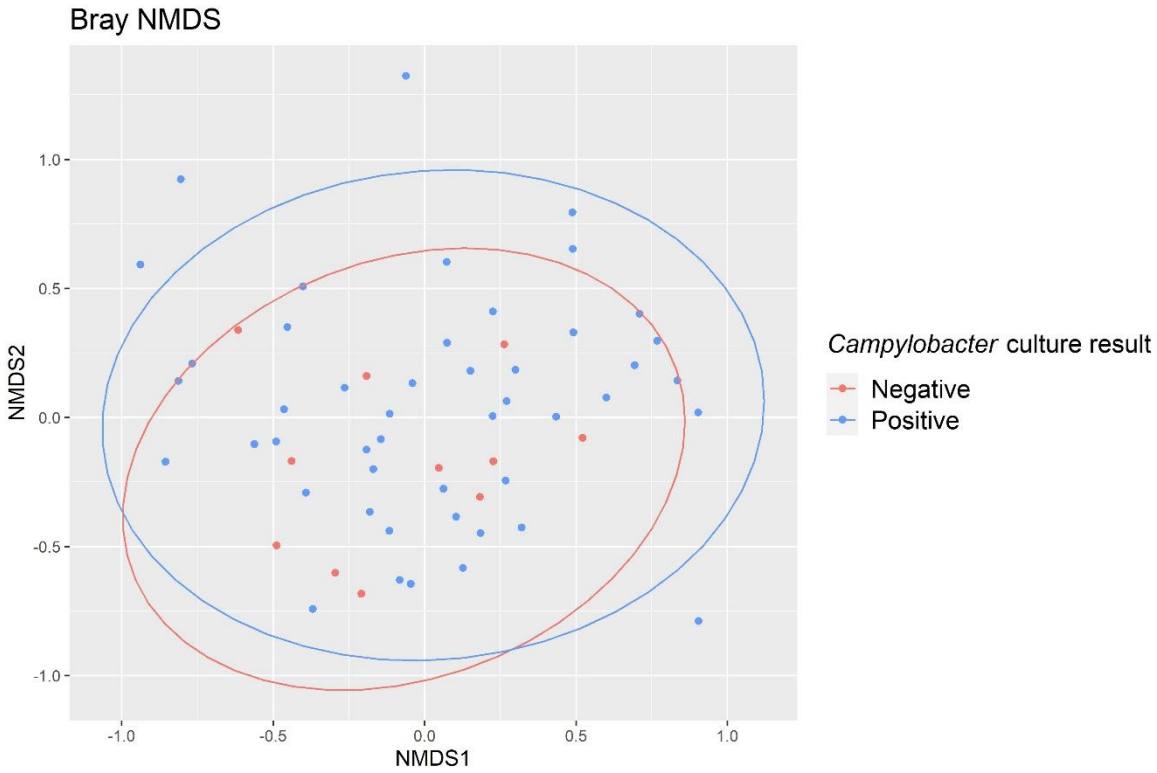

Figure 5.6: Non-metric multidimensional scaling (NMDS) ordination plot showing differences in beta diversity using Bray-Curtis dissimilarity in bacterial taxa composition based on species-level Bracken results agglomerated at the genus level between *Campylobacter* culture positive and negative samples*

*stress = 0.201

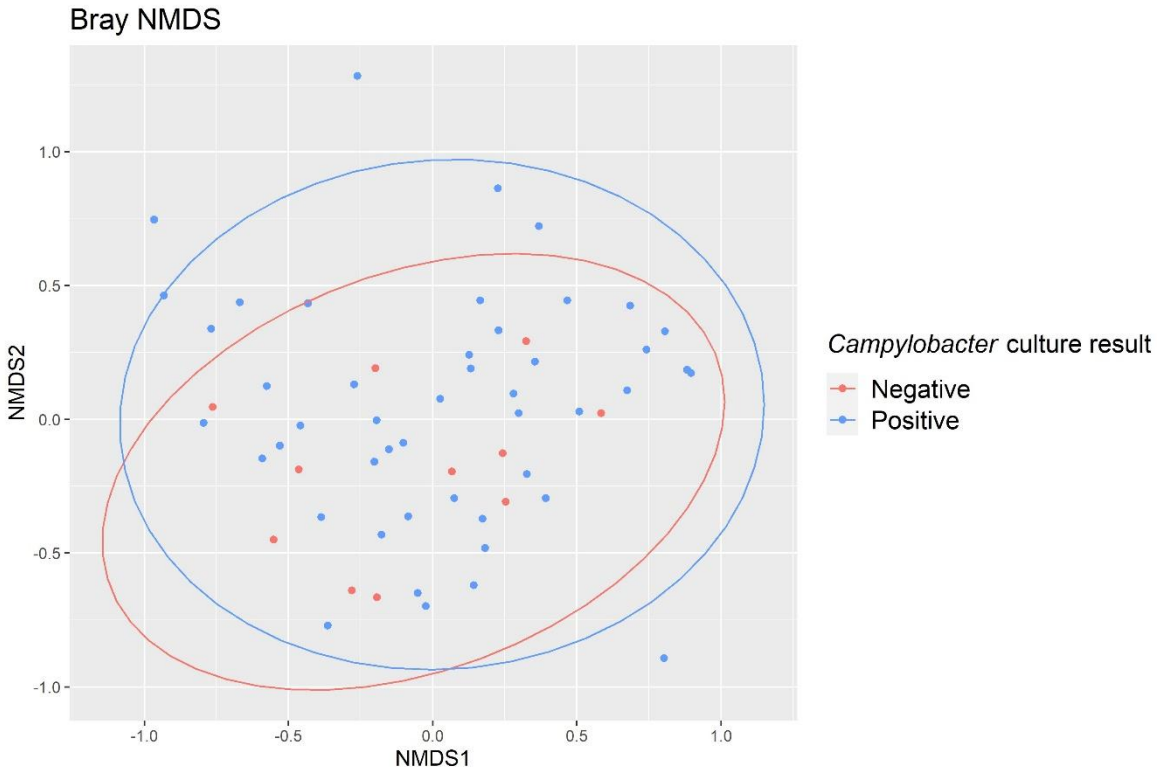

Figure 5.7: Non-metric multidimensional scaling (NMDS) ordination plot showing differences in beta diversity using Bray-Curtis dissimilarity in bacterial taxa composition based on species-level Bracken results agglomerated at the family level between *Campylobacter* culture positive and negative samples*

*stress = 0.187

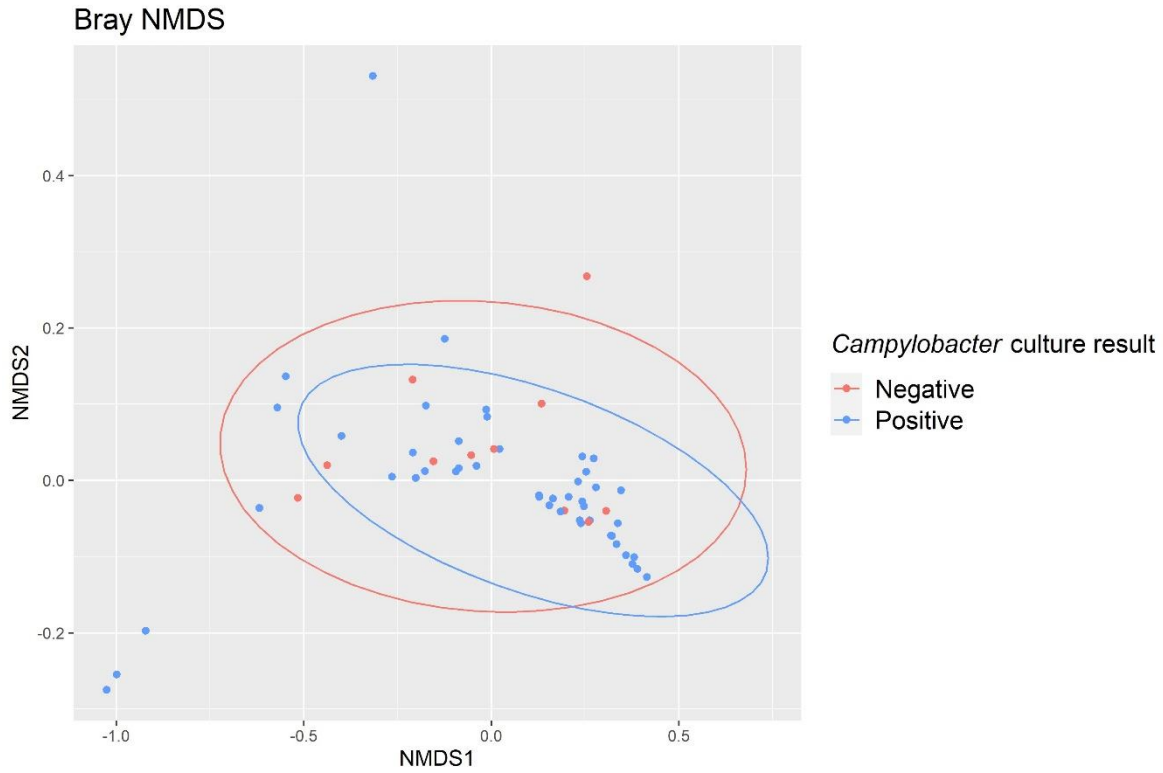

Figure 5.8: Non-metric multidimensional scaling (NMDS) ordination plot showing differences in beta diversity using Bray-Curtis dissimilarity in bacterial taxa composition based on species-level Bracken results agglomerated at the phylum level between Campylobacter culture positive and negative samples*

*stress = 0.018

Visualisation of the most abundant bacterial phyla identified in the samples (Figure 5.9) revealed that the samples were enriched in Proteobacteria and Firmicutes, regardless of *Campylobacter* presence or absence. Investigation of the top 10 most abundant families across the samples in both groups revealed three Firmicutes families and nine families within Proteobacteria to be most abundant in the samples (Figure 5.10).

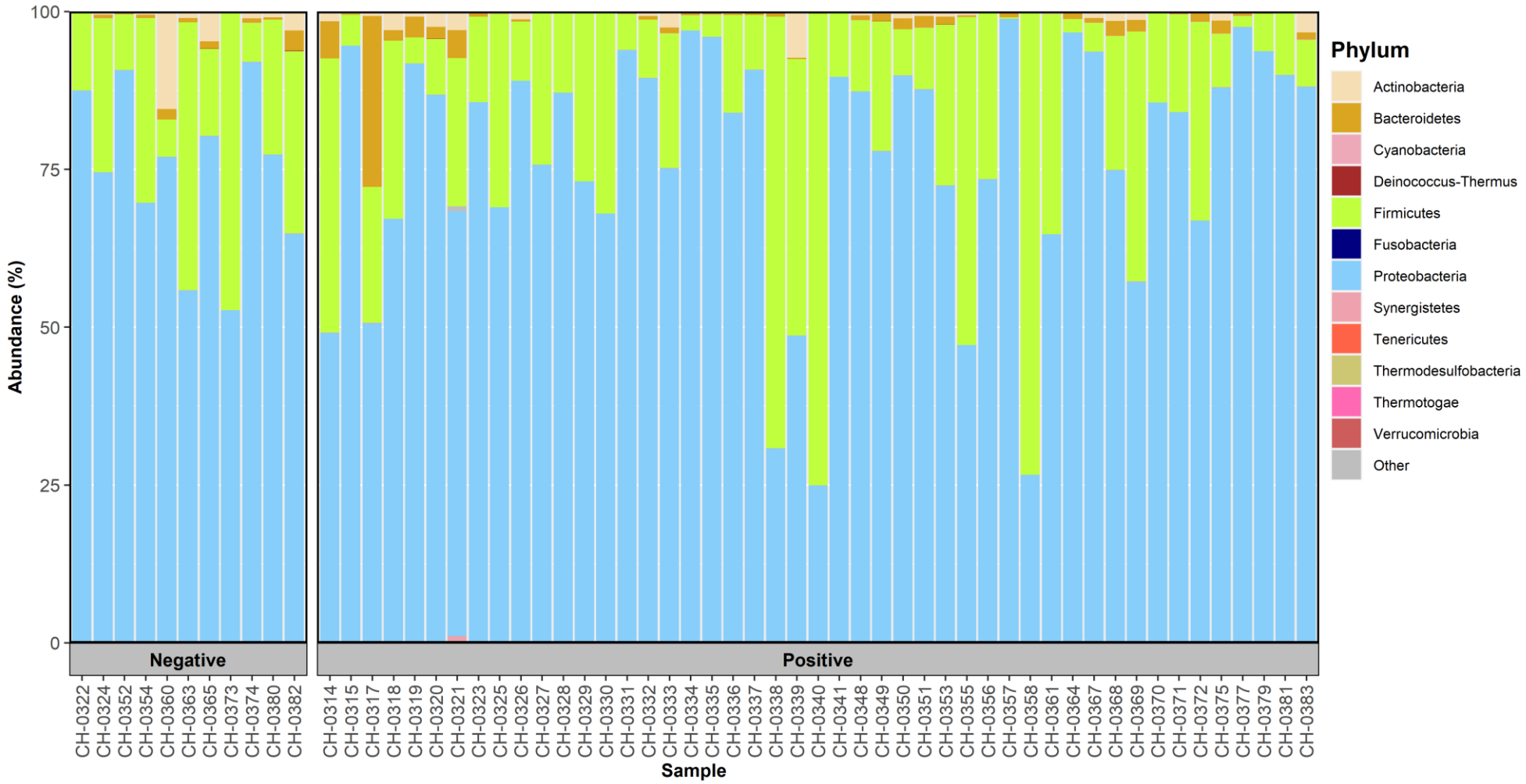

Figure 5.9: Relative abundance of the five most abundant bacterial phyla in each of the chicken meat metagenomes based on species-level Bracken results agglomerated at the phylum level, with the rest of the phyla grouped into the "other" category, split by *Campylobacter* culture result (positive and negative)

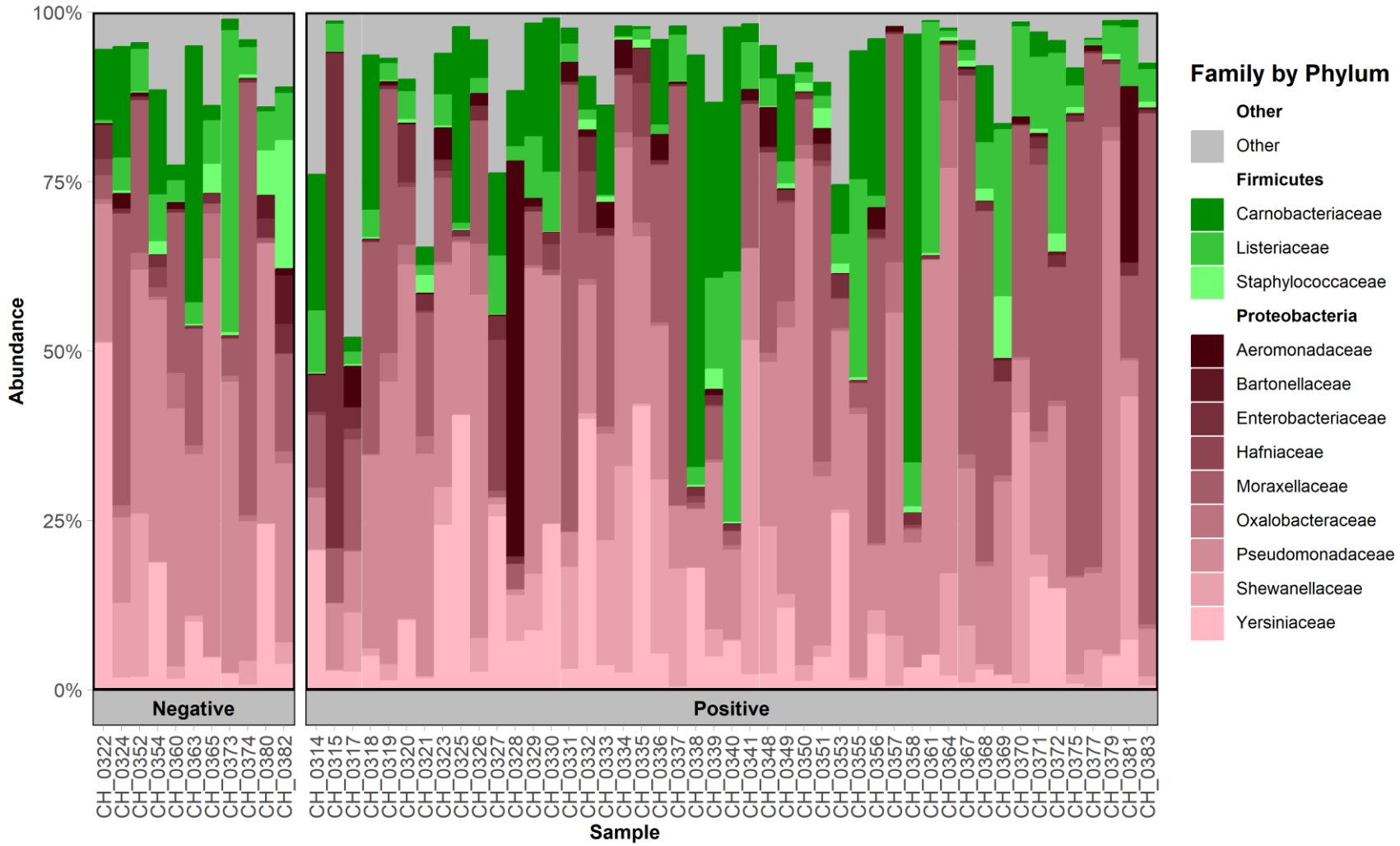

Figure 5.10: Relative abundance of the top 10 most abundant bacterial families in the *Campylobacter* culture positive and negative chicken meat metagenomes grouped by phylum, using species-level Bracken results agglomerated at the family level, with the remaining families grouped into the "other" category

5.4.6 Metagenome controls

Each retail chicken processing run was accompanied by a negative control (225 mL BPW) that was processed as a sample, sequenced and analysed to ensure there was no cross-contamination between samples during processing.

Alpha and beta diversity analyses were performed with the control metagenomes included (Figure 5.11-Figure 5.13). The alpha diversity of controls was considerably lower than the samples (110-484 taxa after rarefying, median = 150). Beta diversity ordination showed that the negative control samples generally clustered separately from the chicken samples. One chicken sample was more similar to one of the negative controls than the rest; this sample contained a high relative abundance of *E. coli*. This was the most abundant classification in the control metagenomes due to the addition of phiX174, which originates from *E. coli*. Although effort was made to filter this out of the negative control metagenomes, the filtering likely missed many associated reads. The AMR genes identified in the chicken samples and control samples were also compared (Figure 5.14). Only three samples displayed a common gene between the metagenome and the negative control (blank) associated with it.

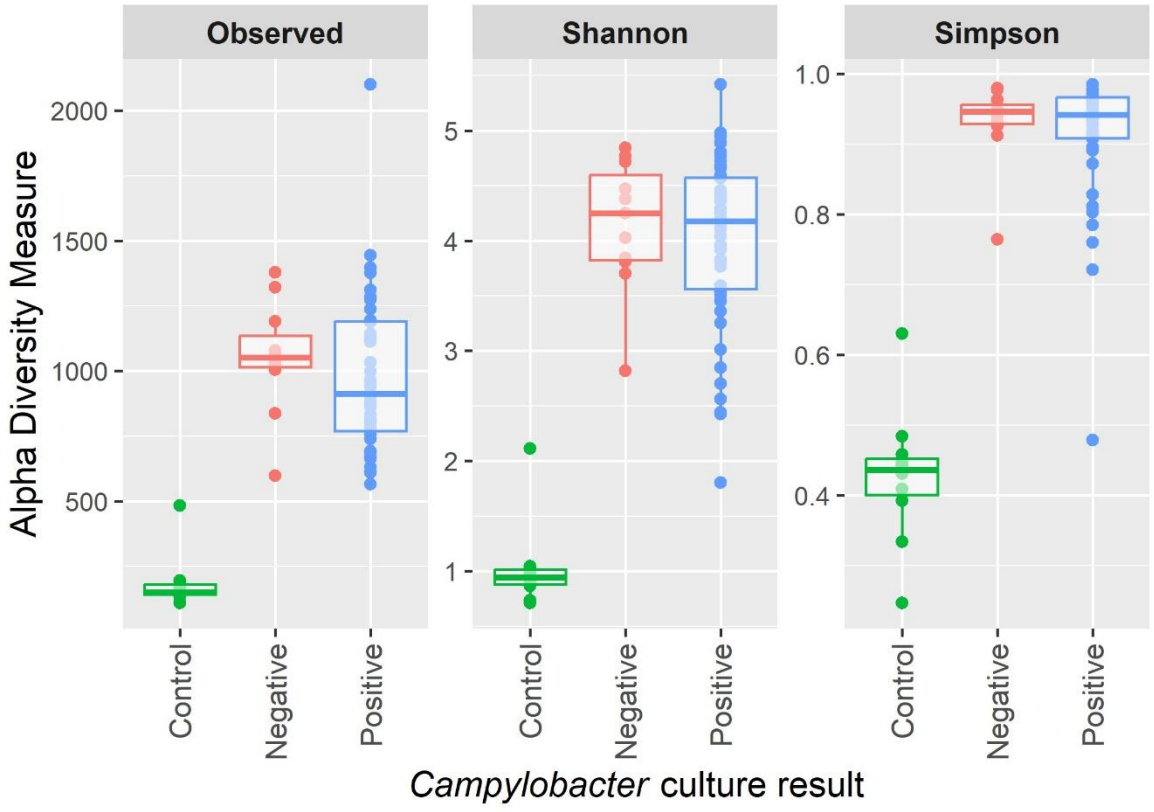

Figure 5.11: Alpha diversity indices (observed richness, Shannon and Simpson) and distributions of the *Campylobacter* culture positive and negative chicken meat samples and negative controls estimated from bacterial species-level Bracken results, after rarefying to the lowest library size

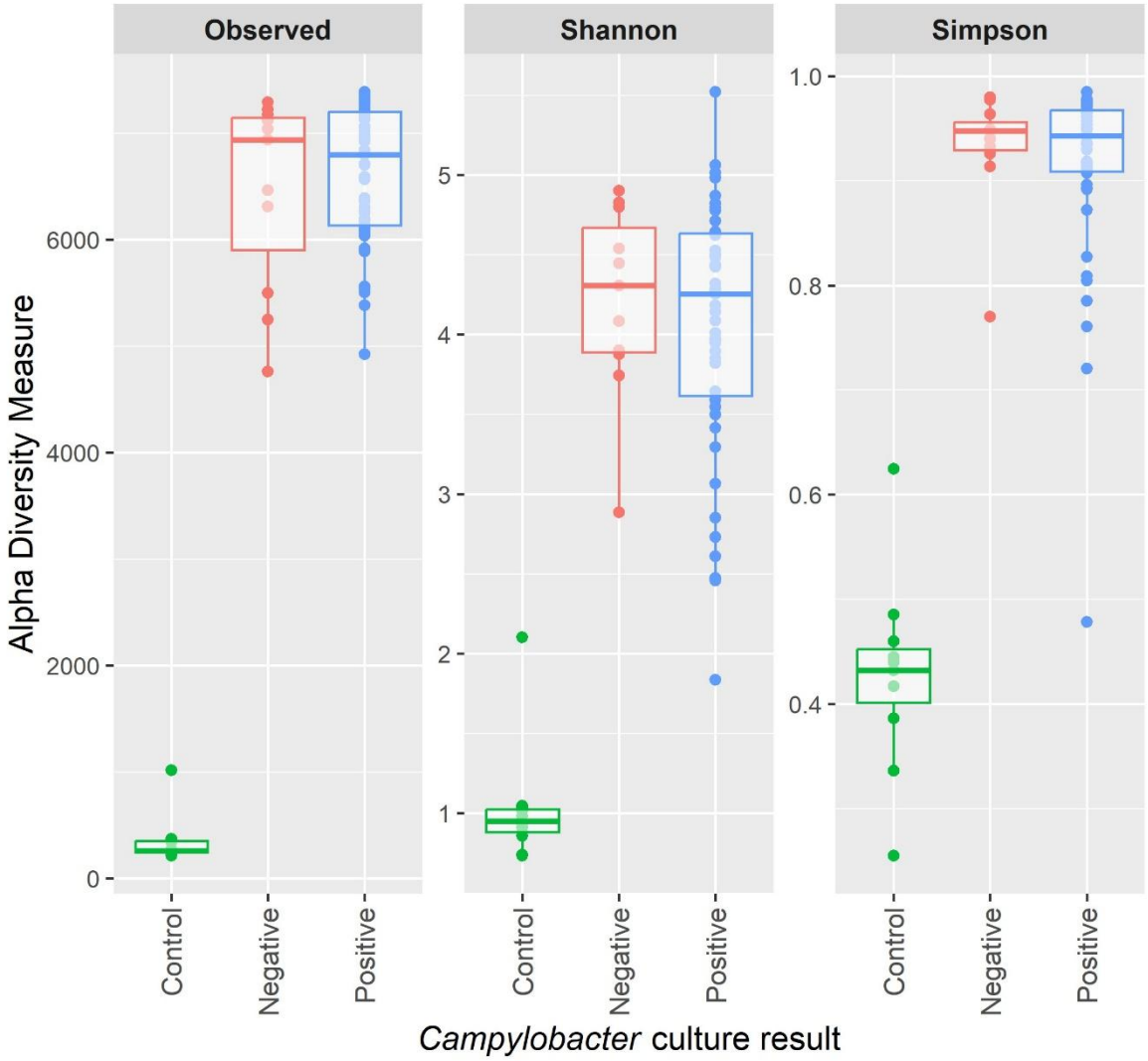

Figure 5.12: Alpha diversity indices (observed richness, Shannon and Simpson) and distributions of the *Campylobacter* culture positive and negative chicken meat samples and negative controls estimated from bacterial species-level Bracken results, without rarefying

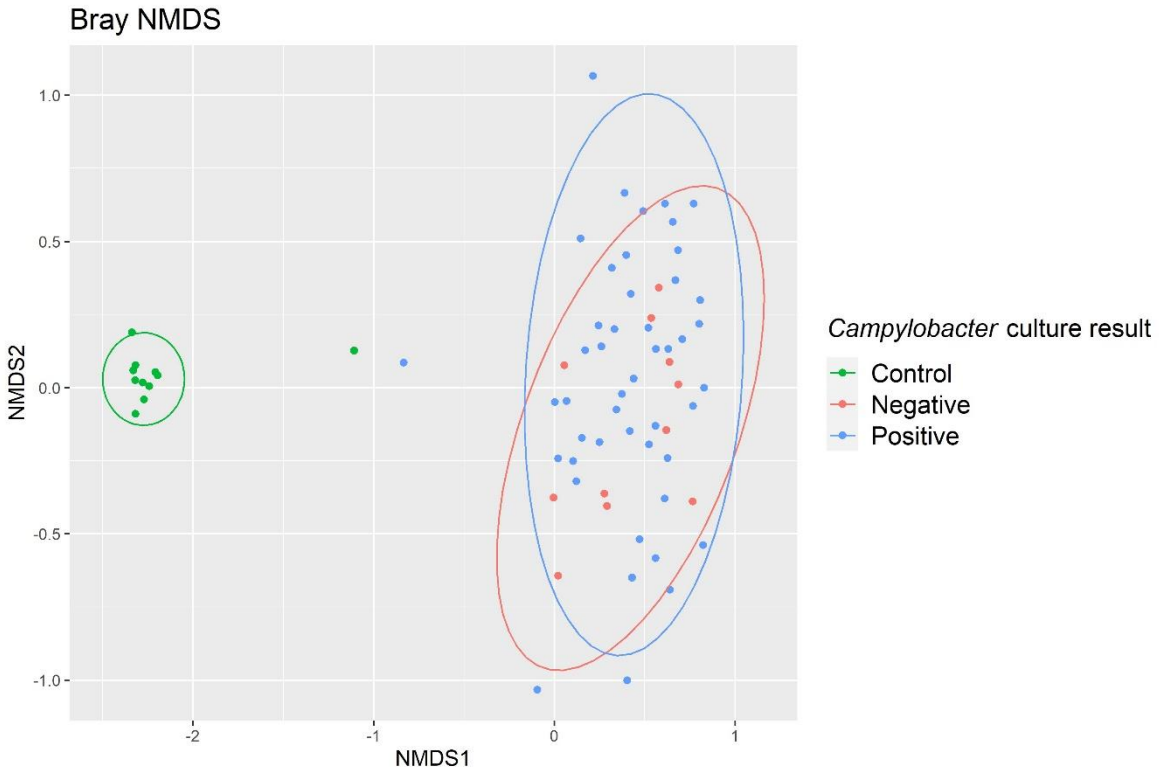

Figure 5.13: Non-metric multidimensional scaling (NMDS) ordination plot showing differences in beta diversity using Bray-Curtis dissimilarity in bacterial species composition based on Bracken results between *Campylobacter* culture positive and negative chicken meat samples and negative controls*

*stress = 0.142

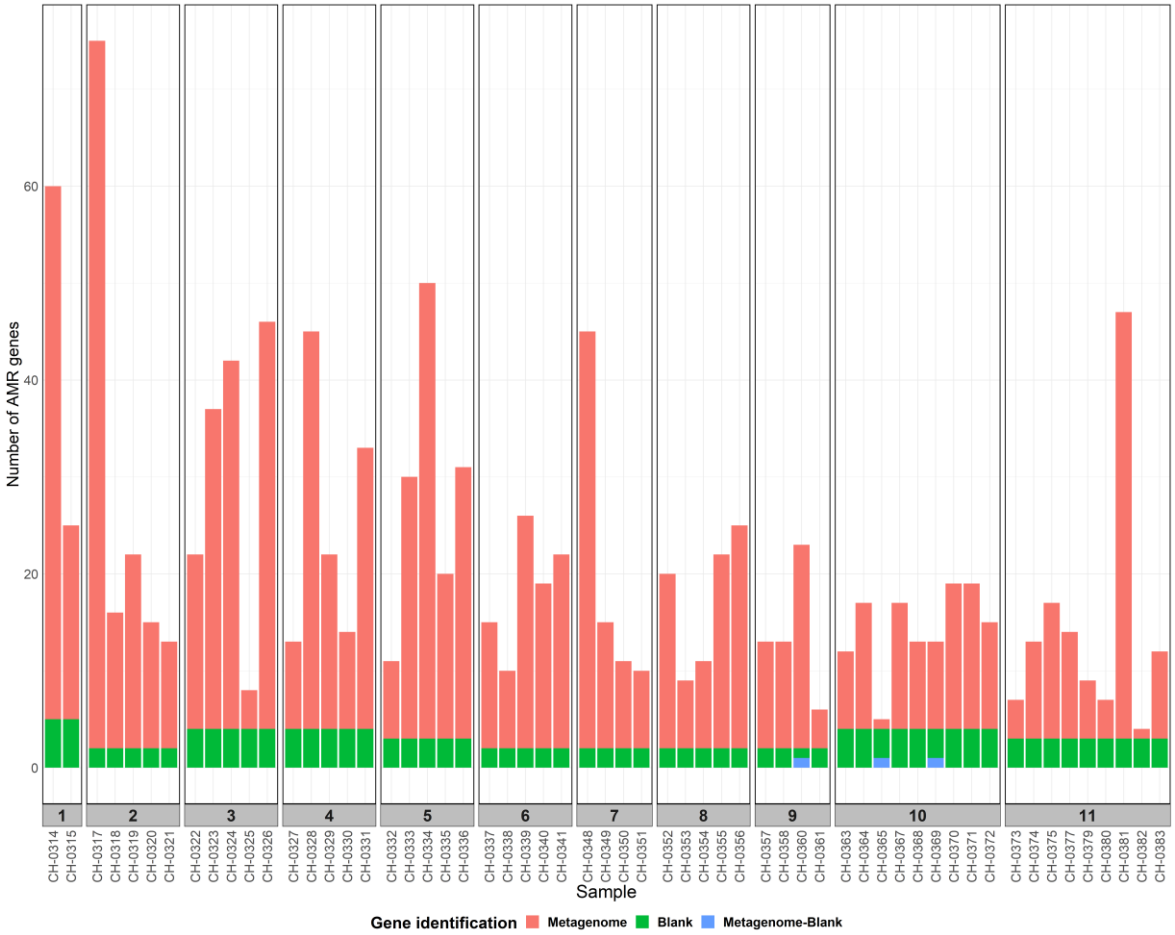

Figure 5.14: Comparison of the resistome of the chicken meat metagenomes and the negative control (blank) sample associated with each processing run, based on KMA results

5.4.7 Differential abundance analysis

Strong relationships between diversity and *Campylobacter* culture result could not be identified. Therefore, further analysis focused on whether specific individual taxa were differentially abundant in the presence or absence of *Campylobacter*. In attempt to identify such differences, three differential abundance analysis tools commonly used in microbiome studies were applied, and their results evaluated. Prior to this, the dataset was filtered to exclude rare and very low abundance taxa.

5.4.7.1 DESeq2

A total of 204 species were determined to be differentially abundant at the $p < 0.05$ significance level after correcting for multiple testing; 100 of these were significantly more abundant in the *Campylobacter* culture positive samples and 104 were significantly less abundant. The $\log_2$fold change ranged from -5.81 to 10.9 (median = -1.3), with the negative values indicating significantly

less abundant organisms in *Campylobacter* culture positive samples and the positive values representing significantly more abundant organisms (Appendix 8). The p-value threshold was reduced to 0.01 for visualisation (Figure 5.15); at this significance level, 23 species were significantly more abundant in *Campylobacter* culture positive samples and 31 were less abundant.

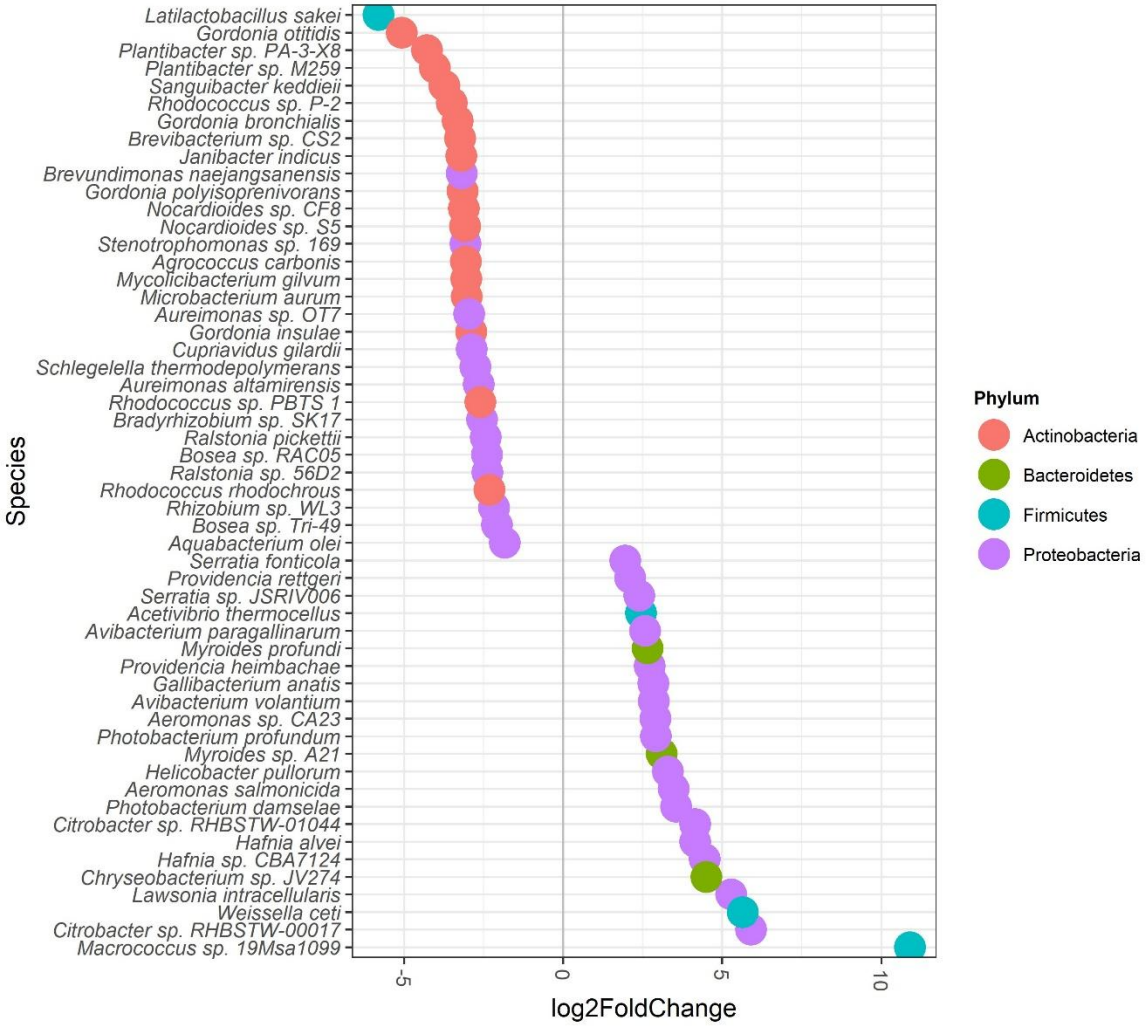

Figure 5.15: Species determined to be significantly more (log_2fold change >0) or less (log_2fold change <0) abundant in *Campylobacter* culture positive samples compared to *Campylobacter* culture negative chicken meat samples, determined with DESeq2 at the 0.01 p-value threshold based on maximum likelihood log_2fold change (Benjamini-Hochberg corrected)

The distribution of relative abundance of the species determined to be differentially abundant at the p <0.01 significance level in the *Campylobacter* culture positive and negative samples after filtering were visualised using both DESeq2 normalised counts and percentage of the overall metagenome reads (Figure 5.16 and Figure 5.17). The relative abundance of these species in the filtered dataset ranged greatly between samples but was below 1% for 45 out of the 54

differentially abundant species. Species that displayed relative abundances above 1% displayed a high within-group range (up to 39.8%).

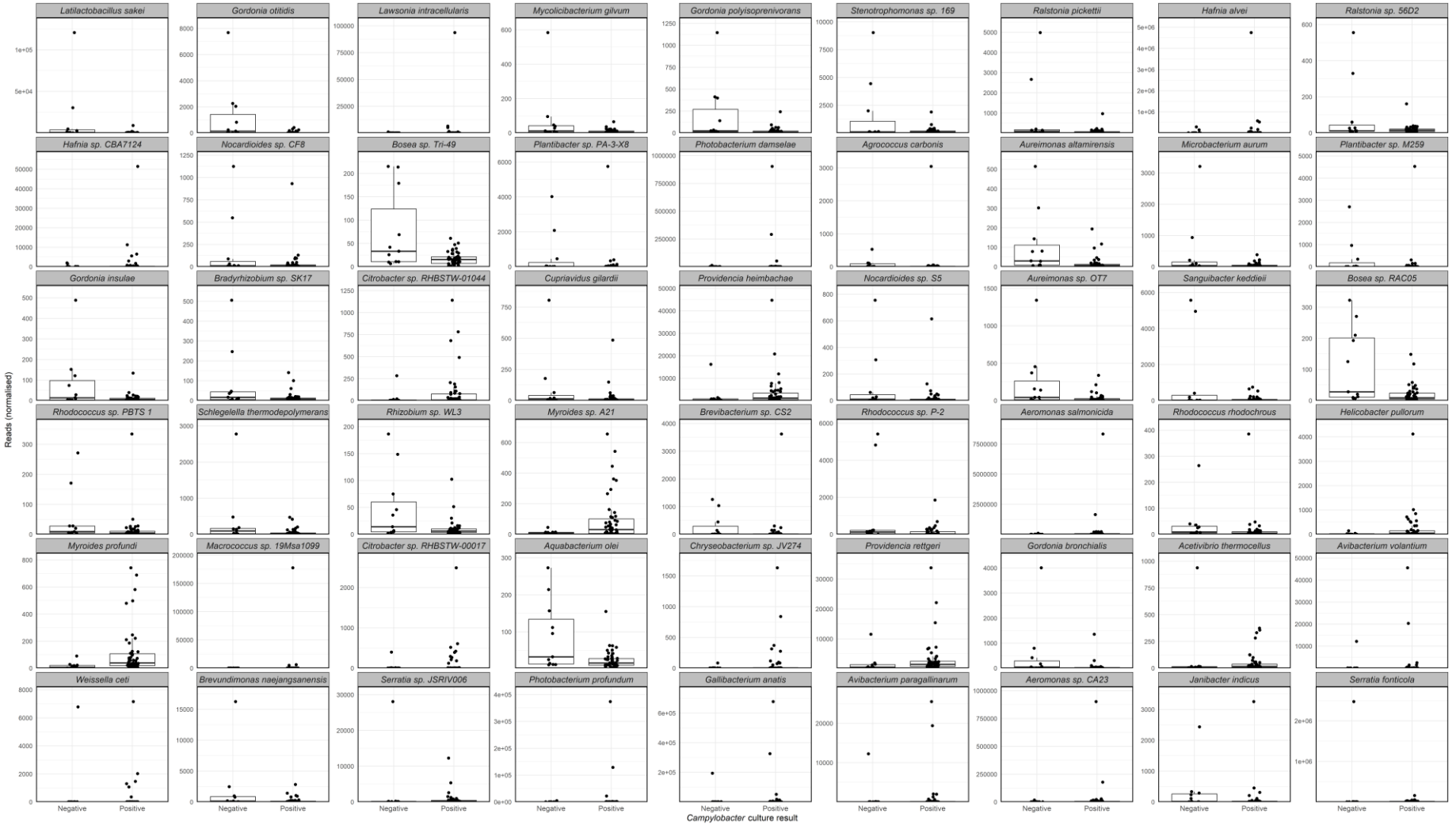

Figure 5.16: DESeq2 normalised read counts for species found to be differentially abundant (p<0.01, Benjamini-Hochberg corrected) in the filtered input dataset, grouped according to *Campylobacter* culture result

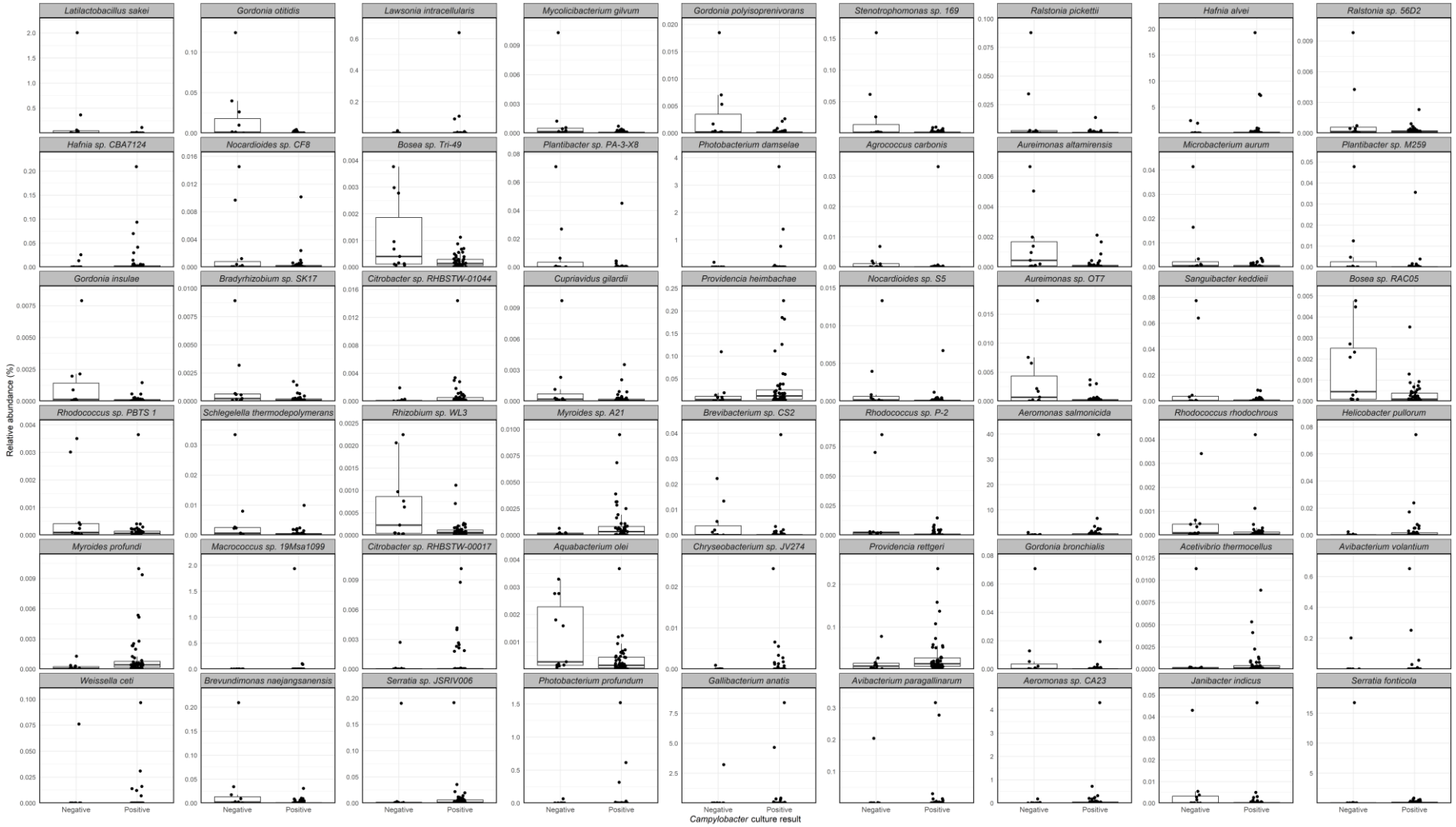

Figure 5.17: Relative abundance, calculated as the percentage of the overall metagenome reads, for species identified as differentially abundant with DESeq2 (p<0.01, Benjamini-Hochberg corrected) in the filtered input dataset, grouped according to *Campylobacter* culture result

5.4.7.2 ANCOM-BC2

ANCOM-BC2 analysis revealed no differentially abundant species between *Campylobacter* culture positive and negative chicken meat samples. The analysis was also performed with agglomeration at the family and phylum level, again revealing no significant differences between *Campylobacter* culture positive and negative samples. The log fold change ranged from -2.11 to 1.84 (median = 0.0196) at the species level, -1.58 to 1.23 (median = 0.150) at the family level and -0.433 to 0.664 (median = 0.184) at the phylum level, based on comparisons of the *Campylobacter* culture positive samples to *Campylobacter* culture negative samples, the reference group (Supplementary Tables S3-S5).

5.4.7.3 ALDEx2

Despite initial filtering, ALDEx2 results indicated the presence of many rare species and abundant species that were not differentially abundant (Figure 5.18). There was overall a large variation in median centred log-ratio (clr) transformed relative abundance (-5.75 to 14.6; median = -0.406) (Supplementary Table S6). The median difference in clr values between groups ranged between -3.18 to 2.89 (median = -0.0323). The taxa that displayed larger differences in abundance between groups also displayed high dispersion. The dispersion (within-group variation, measured as the median of the largest difference in clr values within groups) varied between 0.982-11.0 (median = 2.70), indicating generally higher variation in the relative abundance of taxa within the *Campylobacter* culture positive and negative sample groups than between groups. This high variation within groups resulted in the inability to determine significant differences in relative abundance between groups, as the within-group variation was greater than between-group variation. On the contrary, the few taxa with low within-group variation in relative abundance did not display high variation in relative abundance between groups, thus they were not deemed differentially abundant.

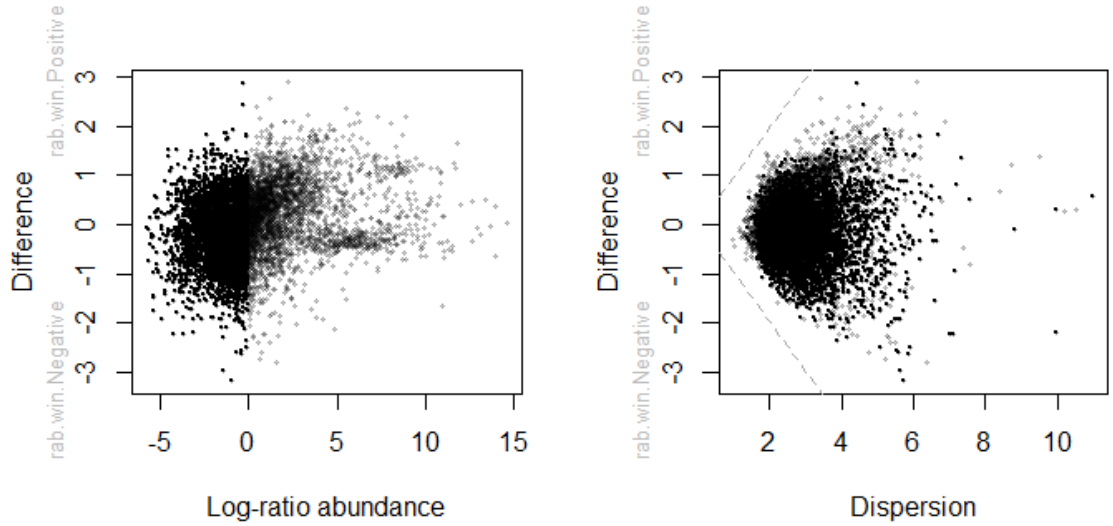

Figure 5.18: ALDEx2 Bland-Altman plot showing the relationship between taxa median relative abundance and median difference between *Campylobacter* culture positive and negative samples (left) and effect plot showing the relationship between median difference (between-group variation) and dispersion (median of the maximum within-group difference) (right)*
*non-differentially abundant taxa are marked with grey or black dots (significance determined with Benjamini-Hochberg corrected p-value threshold of 0.05); the dashed line represents an effect size threshold of 1

ALDEx2 was also used to determine if any species were differentially abundant between *C. jejuni* culture positive and negative as well as *C. coli* culture positive and negative samples, as the survival of these species may be associated with different organisms (Figure 5.19 and Figure 5.20). Similar to previous results, differentially abundant taxa were not identified (Supplementary Tables S7-S8).

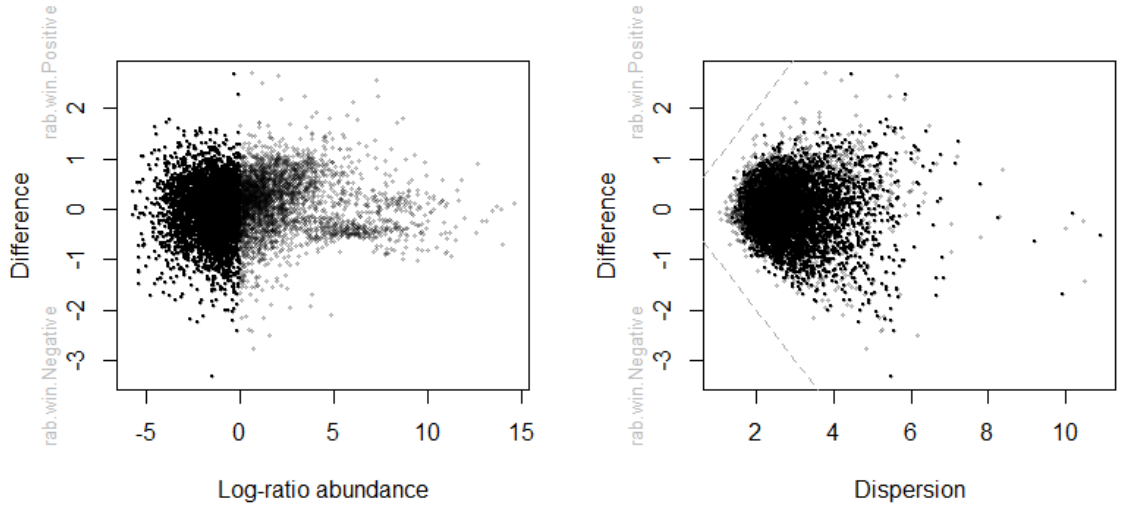

Figure 5.19: ALDEx2 Bland-Altman plot showing the relationship between taxa median relative abundance and median difference between *Campylobacter jejuni* culture positive and negative samples (left) and effect plot showing the relationship between median difference (between-group variation) and dispersion (median of the maximum within-group difference) (right)*
*non-differentially abundant taxa are marked with grey or black dots (significance determined with Benjamini-Hochberg corrected p-value threshold of 0.05); the dashed line represents an effect size threshold of 1

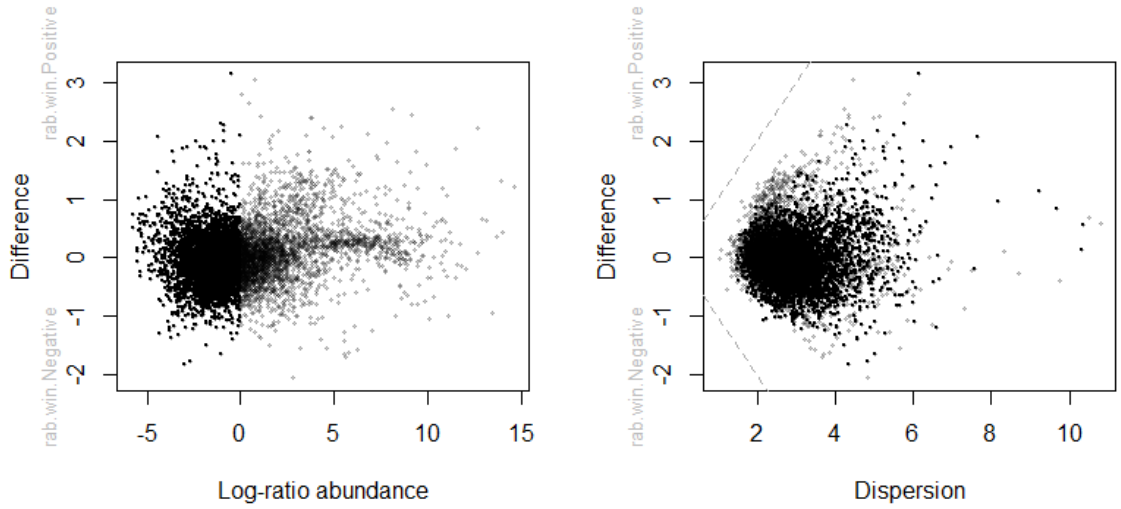

Figure 5.20: ALDEx2 Bland-Altman plot showing the relationship between taxa median relative abundance and median difference between *Campylobacter coli* culture positive and negative samples (left) and effect plot showing the relationship between median difference (between-group variation) and dispersion (median of the maximum within-group difference) (right)*
*non-differentially abundant taxa are marked with grey or black dots (significance determined with Benjamini-Hochberg corrected p-value threshold of 0.05); the dashed line represents an effect size threshold of 1

5.4.8 Metabolic pathway and reaction abundance comparisons

Metabolic pathways were classified and quantified, to allow investigation of whether or not metabolic pathway composition varied significantly between *Campylobacter* culture positive and negative samples; such variation could indicate production of metabolites by the microbial community that either supports or hinders *Campylobacter* survival. This revealed large unmapped (0.0844-0.605, median = 0.262) or unintegrated (0.353-0.846, median = 0.670) proportions, indicating that a large percentage of the reads could not be assigned to either gene families or pathways. The unmapped and unintegrated proportions were combined into the "unclassified" group; this ranged between 0.899-0.958 (median = 0.937) in individual samples. Comparisons of Bray-Curtis dissimilarity of the pathway relative abundances revealed no significant difference in pathway abundance (Adonis2 $p = 0.24$, betadisper permutest $p = 0.93$) between *Campylobacter* culture positive and negative samples (Figure 5.21). After removing the "unclassified" proportion, 562 pathways remained for analysis, and the relative abundances of classified pathways were renormalised. Nonetheless, there was no significant difference between the groups (Adonis2 $p = 0.65$, betadisper permutest $p = 0.58$) (Figure 5.22).

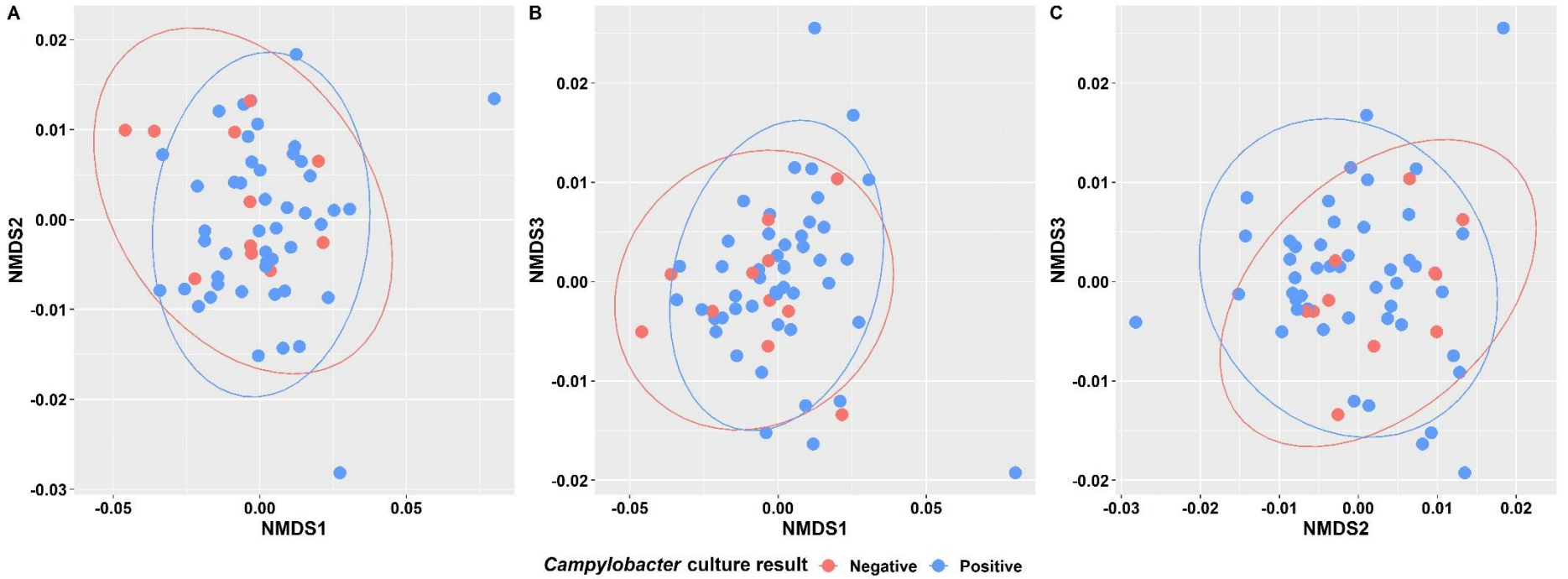

Figure 5.21: Three-dimensional non-metric multidimensional scaling (NMDS) ordination plot based on Bray-Curtis dissimilarity in metabolic pathway composition determined with Humann3 between *Campylobacter* culture positive and negative samples*

*stress = 0.063

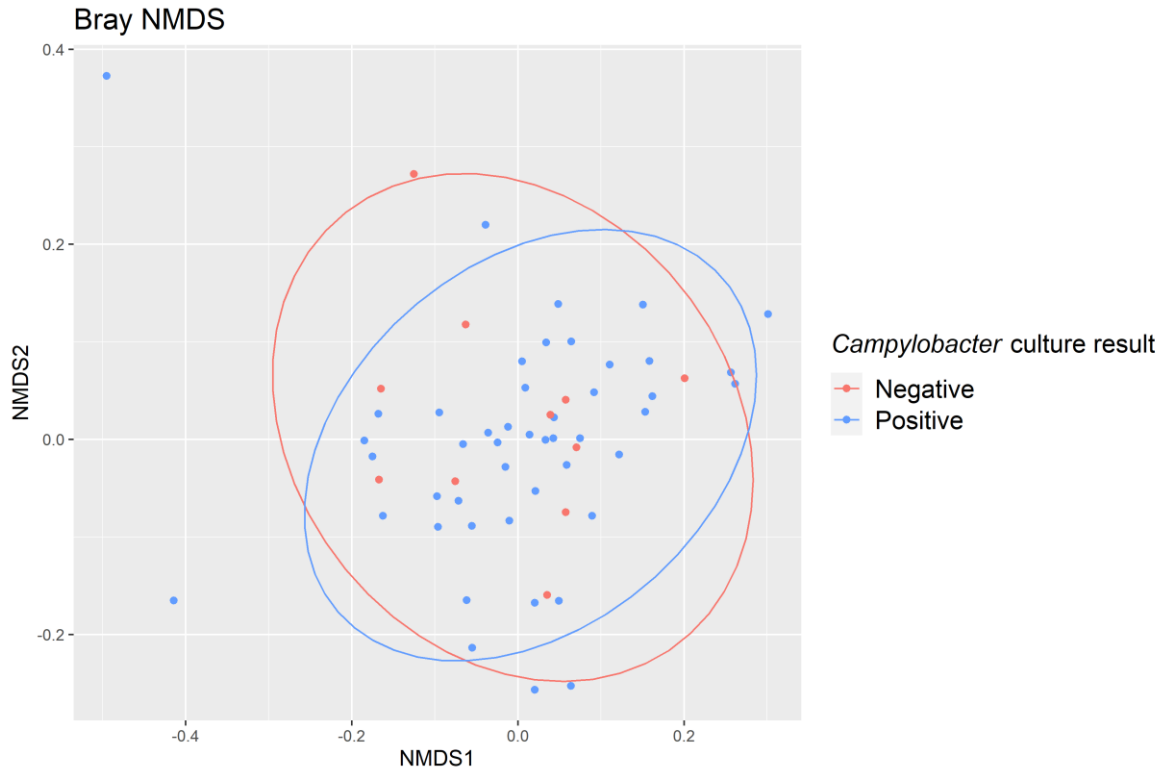

Figure 5.22: Two-dimensional non-metric multidimensional scaling (NMDS) ordination plot based on Bray-Curtis dissimilarity in metabolic pathway composition determined with Humann3 between *Campylobacter* culture positive and negative samples, after removing the unclassified proportions and renormalising the abundances*
*stress = 0.149

The high proportion of unclassified pathways may be linked to the method by which Humann3 calculates pathway abundances, as the abundance of a pathway is limited by the least abundant reaction within it. Gene families were therefore also converted to reactions to determine whether this resulted in a higher proportion of classified reads. The unmapped proportion ranged between 0.0708-0.545 (median = 0.226) and the ungrouped proportion ranged between 0.306-0.665 (median = 0.530). The proportion of unclassified reads (unmapped and ungrouped) was lower (0.698-0.851; median = 0.767) than previously observed for the pathways. Adonis2 and betadisper permutest analysis of calculated Bray-Curtis dissimilarity suggested no significant difference between the *Campylobacter* culture positive and negative sample groups (Figure 5.23; betadisper permutest p = 0.75, Adonis2 p = 0.25). Two-dimensional ordination after removing the "unclassified" proportions and renormalising the remaining relative abundances (4,363 reactions) also indicated no significant difference between the groups (Adonis2 p = 0.43, betadisper permutest p = 0.67) (Figure 5.24).

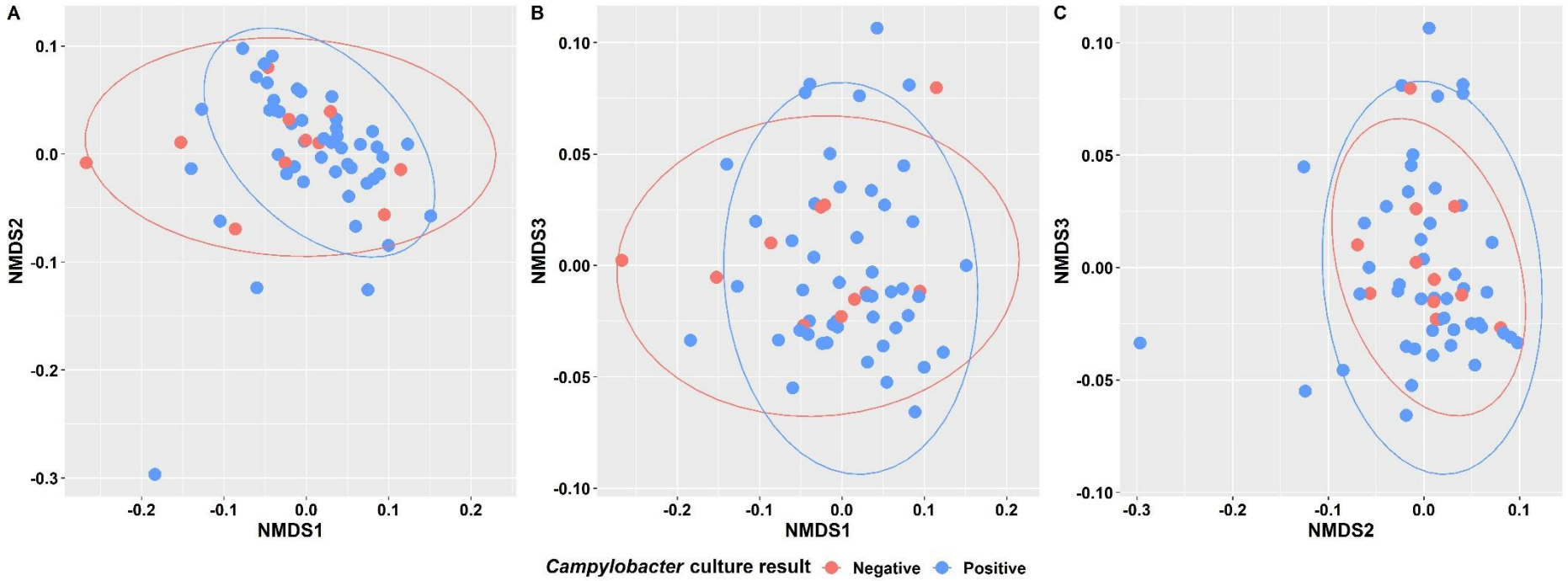

Figure 5.23: Three-dimensional non-metric multidimensional scaling (NMDS) ordination plot based on Bray-Curtis dissimilarity in metabolic reaction composition determined with Humann3 between *Campylobacter* culture positive and negative samples*

*stress = 0.068

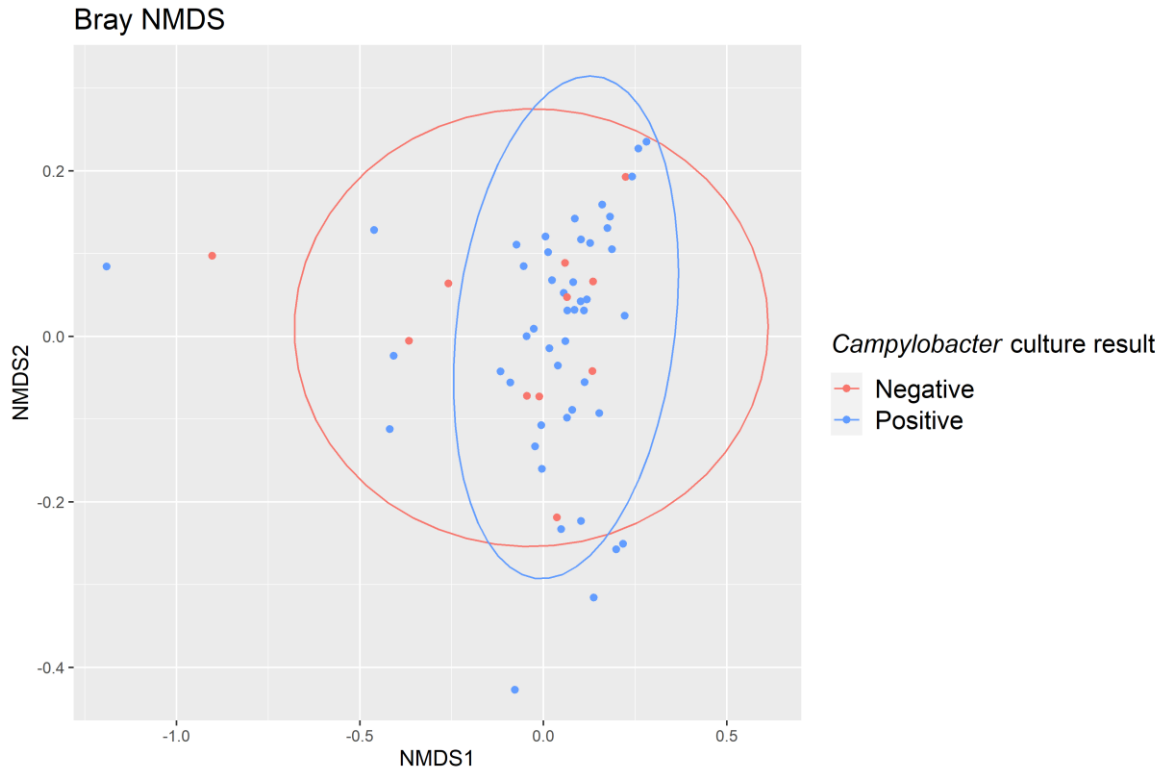

Figure 5.24: Two-dimensional non-metric multidimensional scaling (NMDS) ordination plot based on Bray-Curtis dissimilarity in metabolic reaction composition determined with Humann3 between *Campylobacter* culture positive and negative samples, after removing the unclassified proportions and renormalising the abundances*
*stress = 0.118

5.5 Discussion

The work described in this chapter aimed to evaluate the use of metagenomics for *Campylobacter* identification and characterisation directly from retail chicken samples, and to identify signatures associated with *Campylobacter* presence and absence. Identification of potential *Campylobacter* antagonists could allow establishment of novel intervention strategies to reduce *Campylobacter* prevalence in this epidemiologically important niche.

5.5.1 Identification of *Campylobacter* in retail chicken metagenomes

The relative abundance of *Campylobacter* in the metagenomes was variable but generally very low (0.000862-0.466%). Identification of low abundance organisms such as *Campylobacter* in metagenomes can be difficult, as the presence of host DNA or other taxa results in reduced coverage of organisms of interest, and classification tools are subject to bias (Peabody *et al.*, 2015; Couto *et al.*, 2018). *Campylobacter* abundance on retail meat may be generally low due to the pathogen contaminating rather than colonising the food sample, as the processing and storage conditions may prohibit replication (Murphy, Carroll and Jordan, 2006; Kim *et al.*, 2021).

The ability to recover high quality MAGs for low abundance organisms like *Campylobacter* is also limited due to low abundance and coverage, as evidenced in this study. The overlap in relative abundance between culture positive and culture negative samples, as well as the identification of up to 51 *Campylobacter* species in individual samples, could be linked to the culture methods in this study being inefficient for the isolation of all species and subtypes, or presence of viable but not culturable cells, which can be common on food matrices (Fakruddin, Mannan and Andrews, 2013; Ayrapetyan and Oliver, 2016). However, the identification of a large number of species is likely largely linked to high sequence similarity between taxa resulting in false positive classifications (Peabody *et al.*, 2015).

Molecular methods such as qPCR can be used for *Campylobacter* detection, although this does not provide any genomic information, and detection of *Campylobacter* with qPCR can be difficult in mixed samples without any prior enrichment (Lanzl *et al.*, 2022). This was also observed in the current work, whereby sixteen out of 47 samples that were culture positive and screened with the 16S rRNA assay were consistently qPCR negative and two of six culture positive samples that were screened with the *cadF* primer mix were qPCR negative. The presence of blood in many of the recommended *Campylobacter* enrichment broths can introduce inhibitors that reduce the efficacy of PCR-based methods unless the sample is purified (Josefsen *et al.*, 2004), hence performing qPCR on the enriched broths directly would likely not yield robust results. It is important to note that other features, such as similar melting temperatures between the primers and probe may have an effect on amplification and detection (Rodríguez *et al.*, 2015), which, if investigated, could potentially explain the inconsistency in results for the samples that were repeated. Optimisation of the method and primer and probe sets used here could potentially improve detection, but this was outside of the scope of this work.

Higher relative abundance of *Campylobacter* in the metagenome appeared to facilitate further characterisation at the ST level. *C. jejuni* ST-400 was identified in sample CH-0317, which also displayed the highest relative abundance of *Campylobacter*. The metagenomic identification of *Campylobacter* in this sample aligned with the results obtained through culture and genomics, as all of the isolates recovered from this sample belonged to ST-400. Nonetheless, the results indicate that for most samples reliable detection and characterisation may not be possible. As a result of this, as well as the unreliable qPCR results, the culture results were used for further analysis to identify organisms associated with *Campylobacter* presence and absence.

5.5.2 Microbial signatures associated with *Campylobacter* presence and absence on retail chicken

Previous studies indicate that *Campylobacter* persistence in chicken flocks is largely determined by the growth and death rates of other organisms (Rawson, Dawkins and Bonsall, 2019). Research has thus been conducted to identify organisms associated with *Campylobacter* in broiler chickens. The presence of *C. jejuni* has been correlated with a reduction in the abundance of a number of taxa including *Corynebacterium*, *Lactobacillus* (Kaakoush *et al.*, 2014), and *Blautia* (Thibodeau *et al.*, 2015) and raised abundance of *Escherichia* (Kaakoush *et al.*, 2014) and *Clostridium* (Thibodeau *et al.*, 2015), although contradictory results have been reported by other studies (Kaakoush *et al.*, 2014; Connerton *et al.*, 2018). There is a shortage of investigations of organisms associated with *Campylobacter* on retail chicken meat specifically, which has a significantly different microbial composition to the chicken gut (Marmion *et al.*, 2021).

Alpha diversity varied largely within the *Campylobacter* culture positive sample group (1,670-5,963 taxa) and culture negative sample group (2,187-4,471 taxa). There were no significant differences between the groups based on any of the metrics tested. These results could potentially be affected by a relatively low rarefaction threshold (443,615 reads) in light of the large range in the number of reads in samples (443,615-47,737,177). However, rarefying at a higher read threshold would result in loss of samples for comparison. Given the relatively small number of *Campylobacter* culture negative samples (n = 11) compared to culture positive samples (n = 47), this could lead to loss of power in statistical testing. Alpha diversity metrics were also calculated based on nonrarefied data, again indicating no significant difference between *Campylobacter* culture positive and negative samples, thus indicating that the low rarefaction threshold did not affect the results. The samples also did not cluster based on the *Campylobacter* culture result in beta diversity investigations, suggesting that the culture positive and negative samples did not significantly differ in bacterial composition.

Although there were no significant differences in alpha and beta diversity between *Campylobacter* culture positive and negative samples, there may be more subtle differences in proportions of individual taxa between the two groups. Three tools were used to determine whether any species were differentially abundant between the *Campylobacter* culture positive and negative samples.

DESeq2 identified 54 species as differentially abundant at the 0.01 p-value significance cutoff. The highest $\log_2$fold changes were observed for *Macrococcus* sp. 19Msa1099, significantly more abundant in *Campylobacter* culture positive samples, whereas *Latilactobacillus sakei* and *Gordonia otitidis* were amongst the significantly less abundant species in *Campylobacter* culture positive samples and thus potential *Campylobacter* antagonists. The relative abundances of the

organisms in the filtered metagenome data determined to be differentially abundant with DESeq2 at the 0.01 significance level were visualised, revealing characteristics indicative of false positive hits.

Firstly, low relative abundances were observed for many of the significant species. Species present at low abundance could indicate false positive classifications upstream of the analysis, which can occur with k-mer based classification approaches like Kraken2 (Govender and Eyre, 2022), as mentioned previously. One way to account for this is by filtering the data prior to analysis. In the current study, taxa that were not observed in at least 5% of samples and those with total relative abundance across samples below 0.001% were removed. Other studies have applied higher thresholds (Nearing *et al.*, 2022; Patel *et al.*, 2021), though the optimal cutoff for filtering is difficult to define, and may depend on the data being analysed (McMurdie and Holmes, 2014; Nearing *et al.*, 2022). Here, selection of the sample prevalence filter was related to the low number of *Campylobacter* culture negative samples, as a higher prevalence threshold could lead to removal of taxa in a large proportion of *Campylobacter* culture negative samples and reduce the ability to compare between groups. The relative abundance filter aimed to remove very rare taxa whilst retaining the majority of the data for comparison, to ensure important differences were not missed. Alongside low abundance taxa, within-group variation in relative abundance was observed for some of the DESeq2 differentially abundant taxa, indicating high dispersion that can also cause an issue (Fernandes *et al.*, 2013). DESeq2 aims to account for overdispersion through modelling the variance of counts for each feature, combining the dispersion estimates to fit a trend curve and shrinking the estimates towards the curve. In early implementations of the tool, this was followed by shrinkage of the $\log_2$fold change estimates for features with high dispersion, as well as features with low abundance, such that the uncertainty around these would be accounted for. However, more recent versions of the tool, including the version used here, do not shrink $\log_2$fold changes prior to hypothesis testing (https://bioconductor.org/packages/release/bioc/html/DESeq2.html); instead, this can be done after significant features are determined for visualisation and ranking purposes. This increases the false positivity rate, and therefore some of the $\log_2$fold changes observed in this chapter could be a result of low information available for these taxa, as evidenced in the relative abundance plots. Importantly, earlier versions of DESeq2 have also been shown to produce false positive results (Fernandes *et al.*, 2014; Hawinkel *et al.*, 2019). This may be because despite its common use in microbiome studies, DESeq2 was designed for differential expression analysis of RNA-seq data; the tool's assumptions may therefore not be suitable for analysing complex microbiome data (Fernandes *et al.*, 2014; Weiss *et al.*, 2017).

To address issues with false positive hits from low abundance taxa and high dispersion within groups, ALDEx2 and ANCOM-BC2 were used. These tools did not identify any differentially abundant taxa, which is likely linked to better control of false positivity and false discovery rate in microbiome analysis (Lin and Peddada, 2020; Wallen, 2021; Nearing *et al.*, 2022). These tools are considered to be more conservative (Nearing *et al.*, 2022; Lin and Peddada, 2024), and sensitivity can be reduced with smaller sample sizes, which can reduce power to detect differentially abundant features at lower abundance. The low sample size in this study, particularly in the *Campylobacter* culture negative sample group, is an important limitation that may reduce the ability to reliably identify differentially abundant organisms. There may also be a considerable loss of power associated with the high number of comparisons being performed due to a high number of taxa present in the feature tables, particularly at the species level, resulting in a lack of differentially abundant organisms identified after p-value adjustments performed by ANCOM-BC2 as well as ALDEx2. The trade-off between control of false discovery rate and sensitivity is the reason why it has been argued that multiple tools should be used, as features that are consistently identified as differentially abundant between different tools are most likely to be true positives (Nearing *et al.*, 2022). In the current study, a number of species were identified as differentially abundant with DESeq2, but the results could not be reproduced by ALDEx2 and ANCOM-BC2. The concurrent lack of sample clustering by the culture result group as indicated by beta diversity analysis could indicate that there may be no true difference in this dataset. Alternatively, any microbial signatures associated with *Campylobacter* presence or absence may be very subtle and difficult to identify with tools considered to be more appropriate for microbiome analysis, due to their conservative nature.

It is important to note that the rigorous trimming of the reads in this study can affect the amount of data retained for analysis. However, the extra quality trimming parameters of the FastDeMe pipeline, particularly pertaining to removal of reads with less than 80% of the average read length and reads with low PHRED scores (Vermeulen, 2019), are important for achieving high quality input data for downstream analysis, and can actually reduce false positive results. Perhaps more importantly, the samples analysed in this work displayed variable proportions of unclassified reads (up to 55.9%), which reduces the amount of data available for comparison. There may be features present in the unclassified proportion that differ in abundance between *Campylobacter* culture positive and negative samples, which are currently not possible to identify. This highlights the importance of further sampling and sequencing of retail chicken meat to help curate databases and enable such investigations.

5.5.3 Comparison of metabolic pathway and reaction abundances of *Campylobacter* culture positive and negative retail chicken samples

Campylobacter survival on chicken meat may perhaps not be related to any organism in particular, but rather to other factors such as metabolites that aid survival. Previous research suggests that *C. jejuni* may be able to utilise metabolic by-products such as short chain fatty acids that are produced by members of the microbiome including Clostridiales bacteria (Kaakoush *et al.*, 2014) for expression of genes facilitating amino acid transport, thus supporting survival and chicken gut colonisation. Conversely, the presence of other metabolites, such as propionic acid, can have a negative effect on *Campylobacter* (Chaveerach *et al.*, 2002). It was therefore hypothesised that the presence of specific metabolites on chicken meat can also affect *Campylobacter* survival. To investigate this, Humann3 was used to classify and quantify reactions and metabolic pathways to enable comparison of compositions between *Campylobacter* culture positive and negative samples. The relative abundances of unmapped and unintegrated (0.899-0.958) or unmapped and ungrouped (0.698-0.851) reads in the pathway and reaction combined outputs were high, which limited the ability to identify significant metabolic signals between the sample groups. Indeed, ordination of these data indicated overlap between *Campylobacter* culture positive and negative samples. This again highlights database limitations that require further sampling of different niches, including chicken meat, to facilitate identification of features that differ between *Campylobacter* culture positive and negative samples.

5.6 Conclusions

This chapter demonstrated that the use of metagenomics for the identification of pathogens directly from food samples can be difficult, as the pathogen of interest is often present at very low abundance. This is likely a consequence of the pathogen contaminating rather than colonising the surface of food samples. There may be very limited to no direct modulation of the sample microbiome as a result of *Campylobacter* presence, or modulation of *Campylobacter* presence by other organisms, as *Campylobacter* culture positive and negative samples did not significantly differ in beta diversity. Differential abundance testing using a combination of approaches failed to reliably identify differentially abundant organisms between *Campylobacter* culture positive and negative samples, suggesting that there may be no true microbial signatures associated with *Campylobacter* presence and absence, or such signatures may be subtle and difficult to identify in this dataset.

Although the current chapter showed that pathogens like *Campylobacter* cannot be reliably identified and characterised using metagenomic sequencing of retail chicken samples, clinical

infection cases in which a pathogen colonises and replicates within the host may be marked with altered microbiome compositions and higher abundances of the pathogen that facilitate metagenomic detection and characterisation. The following chapter therefore explores the potential of metagenomics for pathogen detection and characterisation in a clinical scenario, using *Salmonella* genomes derived from paediatric enterocolitis cases and associated metagenomes.

6. Chapter 6: Metagenomics-based identification of disease-causing *Salmonella enterica* serovars and antimicrobial resistance genes from faecal samples

Chapter contributions: The study was conceptualised by AE Mather, NR Thomson, DJ Maskell and S Baker. Method development for sampling was developed by VT Duong, AE Mather and S Baker. Samples were collected and processed by VT Duong. Genomic and metagenomic extraction and sequencing was performed at the Wellcome Sanger Institute. Analysis was done by AH Dziegiel and SJ Bloomfield, in discussion with AE Mather. Visualisation was done by AH Dziegiel. The work was discussed throughout by AH Dziegiel, VT Duong, SJ Bloomfield, N Janecko, J Wain, S Baker and AE Mather.

The work presented in this chapter is described in a manuscript entitled "**Metagenomic identification of disease-causing *Salmonella enterica* serovars and antimicrobial resistance genes from paediatric faecal samples**" (In preparation).

6.1 Introduction

It is difficult to identify low abundance organisms using metagenomics, as shown in Chapter 5 with *Campylobacter* on chicken meat. Although this limits the efficacy of metagenomic sequencing for pathogen surveillance on food, metagenomics may still be a promising potential diagnostic tool for pathogen identification in clinical settings. This is because stool samples from clinical cases of diarrhoea are expected to contain higher proportions of the pathogen potentially responsible for the symptoms, which can be related to proliferation and pathogen-induced changes in the gut microbiome. This is well documented for salmonellosis infections caused by non-typhoidal *Salmonella enterica* (NTS) (Aljahdali *et al.*, 2020), which is another common and epidemiologically important foodborne pathogen (Ehuwa, Jaiswal and Jaiswal, 2021). The direct identification of NTS from clinical cases of diarrhoea is therefore explored in the current chapter.

S. enterica are facultative anaerobes belonging to the Enterobacteriaceae family. *S. enterica* consists of six subspecies, and each of these can be further classified into serovars (Brenner *et al.*, 2000). NTS represent all *S. enterica* serovars that do not cause enteric fever (Gal-Mor, Boyle and Grassl, 2014); most NTS infections are instead characterised as mild to moderate enterocolitis (Gal-Mor, Boyle and Grassl, 2014). However, paediatric and immunocompromised patients are at higher risk of complications and invasive disease, particularly in LMICs (Rosanova *et al.*, 2002; Lan *et al.*, 2016; Ugboko *et al.*, 2020), where NTS infections often outnumber those caused by

Campylobacter and *Shigella* due to differing pathogen epidemiology (Li *et al.*, 2014). The samples analysed in the current chapter are derived from paediatric cases of diarrhoea in Vietnam, which is an LMIC with a high NTS burden in children (Thompson *et al.*, 2013, 2015). Common clinically relevant serovars of *S. enterica* identified in Vietnam include *S.* Typhimurium, *S.* Weltevreden, and *S.* Stanley (Duong *et al.*, 2020). Clinical isolates are commonly associated with resistance to clinically important drug groups including beta-lactams, phenicols, fluoroquinolones and macrolides (Duong *et al.*, 2020; Parisi *et al.*, 2020). AMR in Vietnam may be associated with high antimicrobial consumption, that has historically been unregulated (Van Kinh, 2010). Agricultural use of antimicrobials in animal feed has been common, although this is being gradually phased out (Carrique-Mas *et al.*, 2020, 2023). MDR is particularly concerning, as it may reduce options for treatment of invasive NTS disease and affect future therapeutic outcomes. In Vietnam, 50% of NTS derived from humans and up to 55% of NTS derived from agricultural animals and meat were found to be MDR (Van *et al.*, 2012; Nhung *et al.*, 2024).

Akin to food samples, culture-based techniques for pathogen identification and characterisation from faecal samples are also commonly used in clinical settings (Gilligan, 2013). In recent years, genomic approaches have simplified pathogen typing and prediction of AMR, reducing the requirement for phenotyping (Uelze *et al.*, 2020). However, culturing and sequencing can be a long and expensive process, which makes it less feasible in time-sensitive clinical situations. Metagenomic sequencing is increasingly being evaluated for the analysis of clinical samples to infer aetiology of infections (d'Humières *et al.*, 2021), and therefore could be a viable alternative, though its application for diagnostic use in gastrointestinal infection requires further evaluation (Fourgeaud *et al.*, 2024). *Salmonella* infections have been associated with a decrease in the gut microbial diversity and increase in the abundance of Enterobacteriaceae (Aljahdali *et al.*, 2020), which in itself may be a useful diagnostic signal that can be identified with shotgun metagenomics. As previously discussed, metagenomic sequencing can also facilitate the identification of specific organisms, potentially allowing detection of the causative agents of diarrhoea, and study of AMR genes carried by pathogens as well as the overall bacterial community resistome in order to guide treatment. While its current cost may preclude routine clinical use, sequencing costs continue to decrease (Illumina, 2024), and metagenomics may therefore be a viable diagnostic option in the future. The applicability of metagenomics to clinical samples therefore requires investigation, to determine whether or not it is a viable method for pathogen detection and characterisation in absence of culturing and for the wider characterisation of the microbiome that may facilitate disease diagnosis.

6.2 Aims and objectives

The work outlined in this chapter aimed to:

- Investigate whether or not *S. enterica* can be identified directly from stool metagenomes obtained from culture- and WGS-confirmed salmonellosis cases, using accompanying genomes and comparisons to healthy control metagenomes.
- Determine if the serovar isolated from the samples, confirmed through WGS, can be identified in the accompanying metagenomes.
- Infer whether or not the AMR genes identified in the isolated *S. enterica* through genomics can be identified in the associated faecal metagenomes.
- Compare the resistome of salmonellosis cases to healthy controls, to determine whether or not samples from salmonellosis cases can be distinguished from those obtained from healthy individuals.

6.3 Materials and methods

6.3.1 Faecal sample collection and ethical approval

Faecal samples were collected by Vu Thuy Duong from paediatric patients under 5 years of age at Children's Hospital No. 1 in Ho Chi Minh City, Vietnam. Samples were collected either on the day of admission (before initiation of antibiotic treatment) or up to three days after admission. Ethical approval for this study was obtained from the institutional review board of the hospital and the University of Oxford Tropical Research Ethics Committee (OxTREC No. 1045-13).

6.3.2 *Salmonella* isolation from faecal samples

Salmonella isolation and confirmation was performed by Vu Thuy Duong. Faecal samples were cultured on MacConkey agar (Oxoid), xylose-lysine-deoxycholate agar (Oxoid) and in selenite broth (Oxoid) and incubated at 37°C for 18–24 hours. Colonies that appeared colourless and 2-3 mm in diameter on MacConkey agar or red colonies with a black centre on xylose-lysine-deoxycholate agar were selected and confirmed as *Salmonella* using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (Bruker) and API20E (bioMérieux), following the manufacturer's instructions.

6.3.3 DNA extraction and sequencing

The *Salmonella* genomic DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega) and sequenced on the Illumina HiSeq X Ten platform according to the manufacturer's protocols to generate 150 bp paired-end reads.

The metagenome DNA of the faecal samples was extracted using the FastDNA Spin Kit for Soil (MP Biomedicals), according to manufacturer instructions, and sequenced on the Illumina HiSeq 2500, generating 125 bp paired-end reads.

Whole genome sequencing and metagenome sequencing was performed at the Wellcome Sanger Institute, and the reads deposited in the NCBI SRA and European Nucleotide Archive (ENA) under project PRJEB21259.

6.3.4 Genome analysis

The methods for read trimming, assembly and quality control are outlined in the Materials and Methods chapter (2.4).

The Illumina paired reads were downloaded from the NCBI SRA project (Table 6.1) using the SRA toolkit v2.10.8 (Leinonen *et al.*, 2011).

Table 6.1: Accession information for the *Salmonella enterica* genomes			
Sample ID	**Project**	**ERR (Sample)**	**ERS**
AMR_metagenomic_sequencing6916594	PRJEB21259	ERR2711050	ERS1783778
AMR_metagenomic_sequencing6916595	PRJEB21259	ERR2711051	ERS1783779
AMR_metagenomic_sequencing6916596	PRJEB21259	ERR2711052	ERS1783780
AMR_metagenomic_sequencing6916597	PRJEB21259	ERR2711053	ERS1783781
AMR_metagenomic_sequencing6916598	PRJEB21259	ERR2711054	ERS1783782
AMR_metagenomic_sequencing6916599	PRJEB21259	ERR2711055	ERS1783783
AMR_metagenomic_sequencing6916600	PRJEB21259	ERR2711056	ERS1783784
AMR_metagenomic_sequencing6916601	PRJEB21259	ERR2711057	ERS1783785
AMR_metagenomic_sequencing6916602	PRJEB21259	ERR2711058	ERS1783786
AMR_metagenomic_sequencing6916603	PRJEB21259	ERR2711059	ERS1783787
AMR_metagenomic_sequencing6916604	PRJEB21259	ERR2711060	ERS1783788
AMR_metagenomic_sequencing6916605	PRJEB21259	ERR2711061	ERS1783789
AMR_metagenomic_sequencing6916606	PRJEB21259	ERR2711062	ERS1783790
AMR_metagenomic_sequencing6916607	PRJEB21259	ERR2711063	ERS1783791
AMR_metagenomic_sequencing6916608	PRJEB21259	ERR2711064	ERS1783792
AMR_metagenomic_sequencing6916609	PRJEB21259	ERR2711065	ERS1783793
AMR_metagenomic_sequencing6916610	PRJEB21259	ERR2711066	ERS1783794
AMR_metagenomic_sequencing6916611	PRJEB21259	ERR2711067	ERS1783795
AMR_metagenomic_sequencing6916612	PRJEB21259	ERR2711068	ERS1783796
AMR_metagenomic_sequencing6916613	PRJEB21259	ERR2711069	ERS1783797
AMR_metagenomic_sequencing6916614	PRJEB21259	ERR2711070	ERS1783798
AMR_metagenomic_sequencing6916615	PRJEB21259	ERR2711071	ERS1783799
AMR_metagenomic_sequencing6916616	PRJEB21259	ERR2711072	ERS1783800
AMR_metagenomic_sequencing6916617	PRJEB21259	ERR2711073	ERS1783801
AMR_metagenomic_sequencing6916618	PRJEB21259	ERR2711074	ERS1783802
AMR_metagenomic_sequencing6916619	PRJEB21259	ERR2711075	ERS1783803
AMR_metagenomic_sequencing6916620	PRJEB21259	ERR2711076	ERS1783804
AMR_metagenomic_sequencing6916621	PRJEB21259	ERR2711077	ERS1783805

Paired reads were trimmed using fastp v0.23.2 (Chen *et al.*, 2018), and the estimated genome size set to 5,000,000 for assembly with Shovill v1.1.0+galaxy0 (https://github.com/tseemann/shovill). Assemblies were annotated with Prokka v1.14.5 (Seemann, 2014). Roary v3.13.0 (Page *et al.*, 2015) (95% identity threshold for BLASTp and 99% threshold for isolates a gene must be in to be considered a core gene) was used for core gene alignment. A maximum likelihood phylogeny was constructed in IQ-TREE v1.6.11 (Trifinopoulos *et al.*, 2016) using a general time-reversible model on the web server (https://www.hiv.lanl.gov/content/sequence/IQTREE/iqtree.html) with ultrafast bootstrap approximation (Hoang *et al.*, 2018) and sH-like approximate likelihood ratio test (1,000 replicates) (Anisimova *et al.*, 2011).

The serovars were determined with SISTR v1.0.2 (Yoshida *et al.*, 2016). The methods for ST classification are outlined in the Materials and Methods chapter (2.4). ARIBA v2.13.2 (Hunt *et al.*, 2017) was used for initial screening to identify AMR genes in the *Salmonella* genomes, using the ResFinder (Florensa *et al.*, 2022) database (built 20th May 2019) and a ≥90% alignment identity

threshold, and the results summarised using ARIBA summary with a 90% identity threshold. KMA v1.4.3 (Clausen, Aarestrup and Lund, 2018) was additionally used for AMR gene identification, using the ResFinder database (built 25th January 2019) with 90% query identity and template coverage thresholds. If a gene was identified by one tool but not the other, the ResFinder database of the tool that did not identify the gene was checked to ensure that this did not arise as a result of the gene being absent from the database. StarAMR v0.4.0 (Bharat *et al.*, 2022) was used for the identification of fluoroquinolone resistance mutations with the PointFinder database (Zankari *et al.*, 2017) (v050218), with 90% BLAST identity and hit overlap thresholds.

6.3.5 Metagenome analysis

Illumina paired reads of 28 case faecal metagenomes, each associated with one *S. enterica* genome, were downloaded from the ENA project (6.3.3) and the files for individual runs concatenated. “Control” faecal metagenomes (n = 21) from healthy children up to 5 years of age from the same area were obtained from a study by Pereira-Dias *et al*. (2021) and downloaded from ENA project PRJEB22032 (Table 6.2).

Table 6.2: Accession information for the faecal metagenomes

Sample	Project	Study	ERS
AMR_metagenomic_sequencing6916498	PRJEB21259	ERP023489	ERS1783777
AMR_metagenomic_sequencing6916499	PRJEB21259	ERP023489	ERS7912519
AMR_metagenomic_sequencing6916500	PRJEB21259	ERP023489	ERS7912520
AMR_metagenomic_sequencing6916501	PRJEB21259	ERP023489	ERS7912521
AMR_metagenomic_sequencing6916502	PRJEB21259	ERP023489	ERS7912522
AMR_metagenomic_sequencing6916503	PRJEB21259	ERP023489	ERS7912523
AMR_metagenomic_sequencing6916504	PRJEB21259	ERP023489	ERS7912524
AMR_metagenomic_sequencing6916505	PRJEB21259	ERP023489	ERS7912525
AMR_metagenomic_sequencing6916506	PRJEB21259	ERP023489	ERS7912526
AMR_metagenomic_sequencing6916507	PRJEB21259	ERP023489	ERS7912527
AMR_metagenomic_sequencing6916508	PRJEB21259	ERP023489	ERS7912528
AMR_metagenomic_sequencing6916509	PRJEB21259	ERP023489	ERS7912529
AMR_metagenomic_sequencing6916510	PRJEB21259	ERP023489	ERS7912530
AMR_metagenomic_sequencing6916511	PRJEB21259	ERP023489	ERS7912531
AMR_metagenomic_sequencing6916512	PRJEB21259	ERP023489	ERS7912533
AMR_metagenomic_sequencing6916513	PRJEB21259	ERP023489	ERS7912532
AMR_metagenomic_sequencing6916514	PRJEB21259	ERP023489	ERS7912534
AMR_metagenomic_sequencing6916515	PRJEB21259	ERP023489	ERS7912535
AMR_metagenomic_sequencing6916516	PRJEB21259	ERP023489	ERS7912538
AMR_metagenomic_sequencing6916517	PRJEB21259	ERP023489	ERS7912536
AMR_metagenomic_sequencing6916518	PRJEB21259	ERP023489	ERS7912537
AMR_metagenomic_sequencing6916519	PRJEB21259	ERP023489	ERS7912539
AMR_metagenomic_sequencing6916520	PRJEB21259	ERP023489	ERS7912542
AMR_metagenomic_sequencing6916521	PRJEB21259	ERP023489	ERS7912541
AMR_metagenomic_sequencing6916522	PRJEB21259	ERP023489	ERS7912540
AMR_metagenomic_sequencing6916523	PRJEB21259	ERP023489	ERS7912543
AMR_metagenomic_sequencing6916524	PRJEB21259	ERP023489	ERS7912544
AMR_metagenomic_sequencing6916525	PRJEB21259	ERP023489	ERS7912545
22EN-C-01	PRJEB22032	ERP024351	ERS1865481
22EN-C-02	PRJEB22032	ERP024351	ERS1865482
22EN-C-03	PRJEB22032	ERP024351	ERS1865484
22EN-C-04	PRJEB22032	ERP024351	ERS1865485
22EN-C-05	PRJEB22032	ERP024351	ERS1865490
22EN-C-06	PRJEB22032	ERP024351	ERS1865491
22EN-C-07	PRJEB22032	ERP024351	ERS1865492
22EN-C-08	PRJEB22032	ERP024351	ERS1865493
22EN-C-09	PRJEB22032	ERP024351	ERS1865494
22EN-C-10	PRJEB22032	ERP024351	ERS1865495
22EN-C-11	PRJEB22032	ERP024351	ERS1865497
22EN-C-13	PRJEB22032	ERP024351	ERS1865496
22EN-C-14	PRJEB22032	ERP024351	ERS1865498
22EN-C-15	PRJEB22032	ERP024351	ERS1865499
22EN-C-16	PRJEB22032	ERP024351	ERS1865500
22EN-C-17	PRJEB22032	ERP024351	ERS1865501
22EN-C-18	PRJEB22032	ERP024351	ERS1865502
22EN-C-19	PRJEB22032	ERP024351	ERS1865503
22EN-C-20	PRJEB22032	ERP024351	ERS1865505
22EN-C-21	PRJEB22032	ERP024351	ERS1865504
22EN-C-12	PRJEB22032	ERP024351	ERS1865519

Stool metagenome reads were trimmed with fastp v0.23.4 and host reads removed with Hostile v0.1.0 (Constantinides, Hunt and Crook, 2023) using the T2T-CHM13v2.0 + IPD-IMGT/HLA v3.51 human reference genome.

Taxa classification was performed with Kraken2 v2.1.1+galaxy0 (Wood, Lu and Langmead, 2019) and Bracken v2.8 (Lu *et al.*, 2017) using the k2_nt_20230502 database and 0.1 confidence level, 100mer k-mer distribution and read threshold to include as part of Bracken report set to 10. Bracken analysis was performed at the species and family level. Centrifuge v1.0.3 (nt_2018_3_3 database, with the maximum primary assignments for each read pair set to 1) (D. Kim *et al.*, 2016) was also used for the identification of *S. enterica* and *S. enterica* serovars. To determine the number of reads assigned to the taxa of interest, Kraken-style reports were produced from the Centrifuge outputs using centrifuge-kreport. The relative abundance (proportion of the overall faecal microbiome represented by *S. enterica*) was calculated using the number of *S. enterica* reads according to the classifiers used (number of reads in the Centrifuge kreports and new_est_reads in Bracken) and the number of overall metagenome reads in each sample obtained from the Kraken reports (sum of unclassified and root), and reported as a percentage. The Centrifuge kreports for the case metagenomes were screened for the serovar of the isolated *S. enterica*, and relative abundance calculated in the same way. For the detection of monophasic *S*. Paratyphi B var. Java, the serovar was considered to be present regardless of monophasic or biphasic *S*. Java identification. However, for the monophasic variant of *S*. Typhimurium (I 4,[5],12:i:-), the metagenomes were specifically screened for *S*. I 4,[5],12:i:-. For the control metagenomes, the relative abundance of most abundant *S. enterica* serovar was calculated. The 20 most abundant species and 10 most abundant families in the metagenomes were determined using the Bracken reports and relative abundances calculated in the same way.

Contig and MAG assembly and MAG classification methods are outlined in the Materials and Methods chapter (2.6). MAGs classified as *Salmonella* were additionally classified with Kmerfinder v3.0.2+galaxy0 (https://cge.cbs.dtu.dk/services/KmerFinder-3.2/) (Hasman *et al.*, 2014; Clausen, Aarestrup and Lund, 2018; Khachatryan *et al.*, 2020) to determine the serovar of the MAGs. *Salmonella* MAGs displaying less than 10% completeness were considered to be of poor quality and were removed from further analysis.

NCBI (https://www.ncbi.nlm.nih.gov/) was searched for *Salmonella* reference genomes representing different serovars (May 2021). A total of 117 complete reference genomes were identified. Plasmid contigs were removed and the chromosome contigs were used to form a database. BBsplit v38.75 (https://github.com/BioInfoTools/BBMap/blob/master/sh/bbsplit.sh) was used to align the reads of each metagenome against this database and extract the *Salmonella* reads. The reads were assembled using SPAdes v3.14.1 (Prjibelski *et al.*, 2020), and the assemblies

were aligned back to the database with Nucmer v3.1 (Marçais *et al.*, 2018) using 95% identity and coverage thresholds. The coverage (percentage of each chromosome in the database to which the contigs aligned) was calculated. The serovar displaying the highest proportion of reference covered was defined as the most likely serovar present. The assemblies were also analysed with SISTR cgMLST to predict the serovars. The mapping and assembly analysis was performed by Samuel Bloomfield.

AMR genes were identified with KMA using 60% template coverage and 90% query identity thresholds, to allow direct comparison with the genome results. The genome and metagenome KMA results were compared to determine whether or not *S. enterica* AMR genes were identified in the associated metagenomes. AMR gene variant names were collapsed at the root (e.g. sul2_2_AY034138, sul2_3_HQ840942, sul2_6_FN995456 and sul2_9_FJ197818 results were summarised as *sul2*) to identify unique genes. A gene was considered to be present in the associated metagenome regardless of the specific variant. Fluoroquinolone resistance *gyrA* mutations were excluded when evaluating the proportion of genome AMR determinants identified in the associated metagenome, as these could not be inferred from the metagenomes. The relative abundance of each AMR gene in the metagenome resistome was calculated by first multiplying the KMA template length and depth to obtain the base depth, then dividing this by the total of the base depths for all AMR genes in that sample, and multiplying by 100 to obtain a percentage. Relative abundances of genes identified in the associated *S. enterica* genomes were added up to determine the relative abundance of *S. enterica* associated genes in each metagenome. For genes with multiple variants, the relative abundances of all variants were summed up. The classes of antimicrobials that the genes confer resistance to were reported according to an in-house curated database (Supplementary Table S9) based on the drug classes associated with the genes in the Comprehensive Antibiotic Resistance Database (CARD) (Alcock *et al.*, 2023), or from published literature in cases where the genes were not found in CARD. The drug class for *mcr* genes was changed to polymyxins instead of the peptide classification given by CARD, as this is a more distinct representation of the antimicrobial class to which these genes confer resistance (Hussein *et al.*, 2021). The antimicrobial class for all beta-lactamase genes was simplified to “beta-lactams”. For genes conferring resistance to multiple classes, the summary of the gene classes was set to “multiple”. The class summary was used for plotting and in downstream analysis unless stated otherwise.

The accepted *Salmonella* MAGs were screened for AMR genes with ABRicate v0.9.7 (https://github.com/tseemann/abricate) using the ResFinder database (built 5th November 2021) and 90% identity and 60% coverage thresholds. Contigs containing AMR genes were mapped against publicly available sequences with BLASTn v2.15.0+ (megablast) (Zhang *et al.*, 2000) using

the nucleotide collection (nr/nt, updated 2024/01/12) (Morgulis *et al.*, 2008) on the online server (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with default parameters. Contigs containing *aac(6')-Iaa* genes were not mapped as these genes are considered endogenous to *Salmonella* (Magnet, Courvalin and Lambert, 1999; Salipante and Hall, 2003). The first 100 chromosome and plasmid hits with ≥90% identity were considered.

6.3.6 Statistical analysis

The *S. enterica* and Enterobacteriaceae relative abundances in the case and control metagenomes were compared using Mann-Whitney U (Wilcoxon rank-sum) tests. The number of unique AMR genes and antimicrobial classes to which the genes confer resistance in the case and control metagenomes were compared in the same way. Fisher's exact tests were used to compare the proportions of case and control metagenomes containing genes conferring resistance to AMR classes commonly used in salmonellosis treatment; these comparisons also considered genes conferring resistance to multiple classes, if the given gene(s) conferred resistance to macrolides, beta-lactams or fluoroquinolones.

Kendall's correlation tests were performed to investigate associations between the relative abundance of *S. enterica* and the overall number of reads in the case metagenomes; relative abundance of the isolated *S. enterica* serovar and the overall number of reads in the case metagenomes; and the overall number of reads and proportion of AMR genes identified in the case metagenomes that were also identified in the associated genome. The purpose of this was to investigate whether or not the ability to detect *S. enterica* or the isolated serovar and its AMR determinants in the metagenomes was associated with sequencing depth.

Higher relative abundance of *S. enterica* in the metagenome may increase the likelihood of detection of the AMR determinants identified in the *S. enterica* genome. To test this formally, the relative abundance of *S. enterica* in the metagenomes and the proportion of *S. enterica* AMR genes (the AMR genes identified in the associated *S. enterica* genome obtained from the same sample) identified in the case metagenomes were compared.

6.4 Results

6.4.1 *Salmonella* genomes

All 28 genomes were classified as *S. enterica* subsp. *enterica* using SISTR. Further classification of the genomes resulted in the identification of 11 serovars and 13 STs (Table 6.3).

Table 6.3: *Salmonella enterica* serovars and sequence types (STs) of the 28 genomes generated

Serogroup	Serovar	ST	Number of genomes
B	Typhimurium	34	2
		19	2
		36	1
	I 4,[5],12:i:-	34	11
	Stanley	29	1
		2615	1
	Paratyphi B var. Java monophasic	423	1
	Saintpaul	50	2
C2-C3	Newport	4157	1
	Albany	292	1
D1	Enteritidis	11	2
E1	Weltevreden	365	1
E4	Meleagridis	463	1
I	Hvittingfoss	446	1

The genomes were initially screened for AMR genes with ARIBA (Supplementary Table S10), though to facilitate direct comparisons with the metagenome data, KMA was also used. KMA and StarAMR analysis revealed 35 unique AMR determinants amongst the 28 genomes (Appendix 9). The AMR determinants identified with KMA conferred resistance to beta-lactams (82.1% of genomes), aminoglycosides (100%), tetracyclines (82.1%), polymyxins (14.3%), sulphonamides (82.1%), macrolides (21.4%), lincosamides (14.3%), phosphonic acids (3.57%), phenicols (67.9%), diaminopyrimidines (53.6%) and rifamycins (17.9%). Fluoroquinolone resistant genotypes were identified in 78.6% of genomes, based on KMA and StarAMR results. A high proportion of genomes (82.1%) displayed MDR genotypes. The number of unique AMR determinants per genome ranged between 1-18 (median = 9).

Both *S.* Stanley genomes, the *S.* Paratyphi B var. Java, *S.* Weltevreden and *S.* Hvittingfoss genomes displayed susceptible genotypes except for the presence of *aac(6')-Iaa*. Although this is a genotypic aminoglycoside resistance determinant and thus counted as an aminoglycoside resistance gene, it is considered to be a cryptic gene that does not confer phenotypic resistance (Magnet, Courvalin and Lambert, 1999). *S.* Saintpaul genomes shared the same AMR genotype, containing rifamycin, aminoglycoside, beta-lactam, diaminopyrimidine, phenicol, lincosamide, macrolide, fluoroquinolone, sulphonamide and tetracycline resistance determinants. The *S.* Enteritidis genomes also contained AMR genes conferring resistance to the same antimicrobial classes, namely aminoglycosides, beta-lactams, sulphonamides and tetracyclines.

All of the *S. enterica* I 4,[5],12:i:- genomes belonged to ST-34. Most (90.9%) contained tetracycline resistance genotype *tet(A)* genes and 54.5% contained *tet(B)* genes. Phenicol resistant genotypes

were evident in 81.8% of the genomes, and a smaller proportion of genomes displayed additional genotypic resistance to rifamycins conferred by *ARR* genes (18.2%), diaminopyrimidines (*dfr* genes; 63.6%), polymyxins (*mcr* genes; 27.3%), lincosamides (*lnu(F)*; 9.09%) and macrolides (*mph(A)*; 18.2%).

Two of the *S.* Typhimurium genomes belonged to ST-34, the same ST as the monophasic genomes. Their AMR genotypes were similar to those of *S.* I 4,[5],12:i:-, both containing *aac(6)-Iaa*, *aph(6)-Id, aph(3'')-Ib, bla*$_{\text{TEM-1B}}$, *floR*, *sul2*, *tet(A)* and *tet(B)* genes. Some variation in the presence of other genes was observed, similar to the other ST-34 isolates. The two ST-19 *S.* Typhimurium genomes displayed similar AMR genotypes to one another, except for an additional *ant(3'')-Ia* gene identified in one of the genomes.

6.4.2 Faecal metagenomes

This study investigated faecal metagenomes from children with culture confirmed salmonellosis and healthy controls. The metadata associated with the faecal samples analysed in this chapter, including the sample collection year, group (case/control), and associated genomes is outlined in Table 6.4.

Table 6.4: Metadata associated with the faecal metagenomes analysed			
Sample	**Collection year**	**Group**	**Associated genome**
metagenome_6916498	2016	Case	ERR2711050
metagenome_6916499	2016	Case	ERR2711051
metagenome_6916500	2016	Case	ERR2711052
metagenome_6916501	2016	Case	ERR2711053
metagenome_6916502	2016	Case	ERR2711054
metagenome_6916503	2016	Case	ERR2711055
metagenome_6916504	2016	Case	ERR2711056
metagenome_6916505	2016	Case	ERR2711057
metagenome_6916506	2016	Case	ERR2711058
metagenome_6916507	2016	Case	ERR2711059
metagenome_6916508	2016	Case	ERR2711060
metagenome_6916509	2016	Case	ERR2711061
metagenome_6916510	2016	Case	ERR2711062
metagenome_6916511	2016	Case	ERR2711063
metagenome_6916512	2016	Case	ERR2711064
metagenome_6916513	2016	Case	ERR2711065
metagenome_6916514	2016	Case	ERR2711066
metagenome_6916515	2017	Case	ERR2711067
metagenome_6916516	2017	Case	ERR2711068
metagenome_6916517	2017	Case	ERR2711069
metagenome_6916518	2017	Case	ERR2711070
metagenome_6916519	2017	Case	ERR2711071
metagenome_6916520	2017	Case	ERR2711072
metagenome_6916521	2017	Case	ERR2711073
metagenome_6916522	2017	Case	ERR2711074
metagenome_6916523	2017	Case	ERR2711075
metagenome_6916524	2017	Case	ERR2711076
metagenome_6916525	2017	Case	ERR2711077
ERR9904448	2017	Control	-
ERR9904449	2017	Control	-
ERR9904450	2017	Control	-
ERR9904451	2017	Control	-
ERR9904456	2017	Control	-
ERR9904457	2017	Control	-
ERR9904458	2017	Control	-
ERR9904459	2017	Control	-
ERR9904460	2017	Control	-
ERR9904461	2017	Control	-
ERR9904462	2017	Control	-
ERR9904463	2017	Control	-
ERR9904464	2017	Control	-
ERR9904465	2017	Control	-
ERR9904466	2017	Control	-
ERR9904467	2017	Control	-
ERR9904468	2017	Control	-
ERR9904469	2017	Control	-
ERR9904470	2017	Control	-
ERR9904471	2017	Control	-
ERR9904485	2017	Control	-

There was a larger range in the overall number of reads after trimming and host read depletion amongst the case metagenomes (30,576-94,890,832 reads) compared to controls (10,821,636-15,177,561).

6.4.2.1 Taxonomic profiling

The 10 most abundant taxa in the faecal samples in the case and control metagenomes determined with Bracken were compared at the family level (Figure 6.1). The 10 most abundant families represented 64.6-99.6% (median 97.7%) and 57.9-92.5% (median 78.2%) of the overall case and control metagenomes, respectively. Focusing on microbial families specifically, the most abundant family in the case metagenomes were Enterobacteriaceae (14 samples), Bacteroidaceae (five samples), Bifidobacteriaceae (four samples), Enterococcaceae (two samples), Herpesviridae (one sample), Lactobacillaceae (one sample) and Staphylococcaceae (one sample). In the controls, the most abundant microbial families were Lachnospiraceae (eight samples), Bacteroidaceae (seven samples), Bifidobacteriaceae (four samples) and Enterobacteriaceae (two samples).

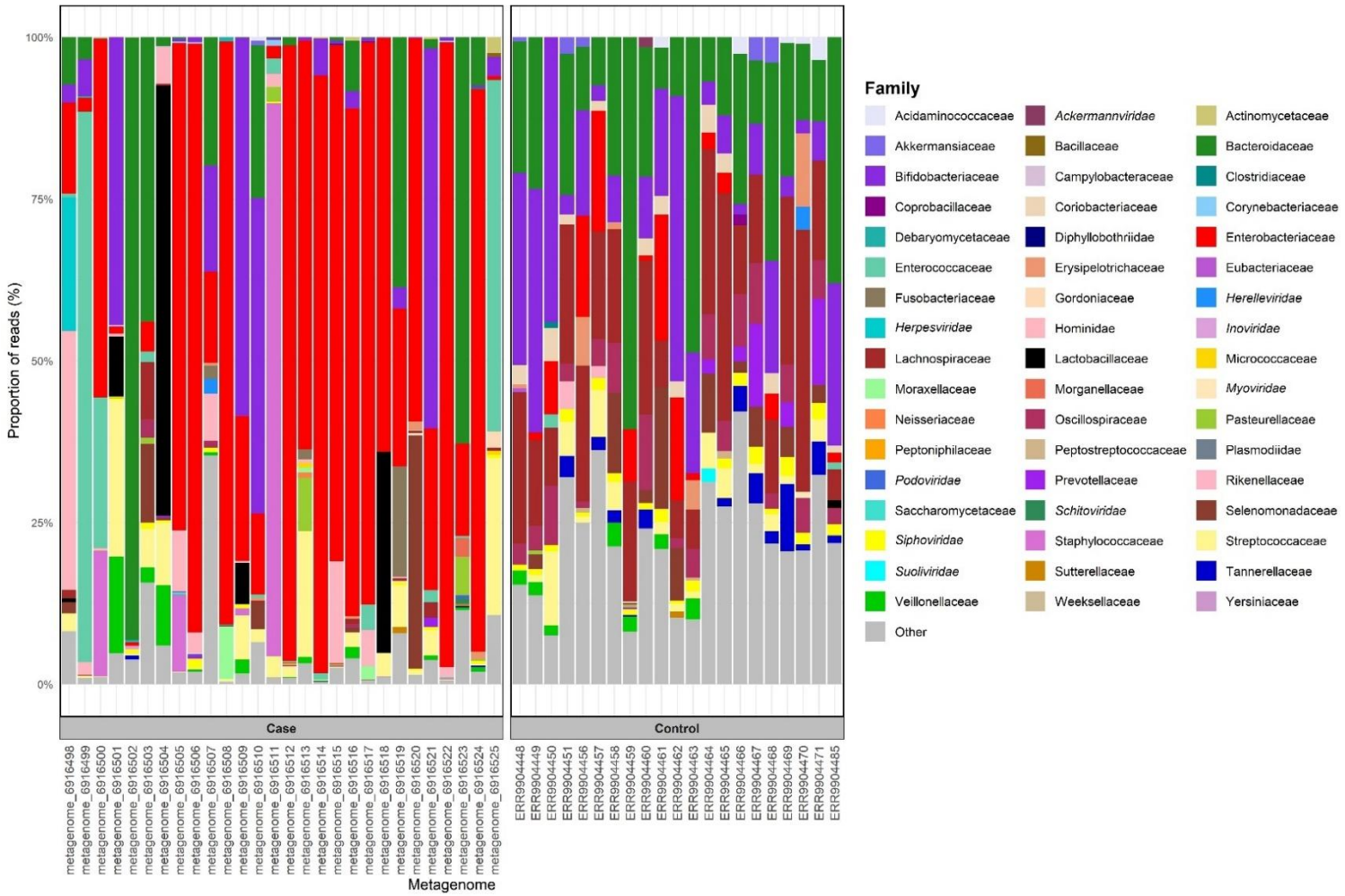

Figure 6.1: The 10 most abundant families in each case and control metagenome determined with Bracken, with remaining classifications grouped into the 'other' category

The species compositions were also compared. the top 20 most abundant species represented 63.3-99.6% (median = 97.4%) of the case metagenome reads and 44.8-90.5% (median = 69.6%) of the control metagenome reads (Appendix 10).

Focusing on microbial species, the most abundant species in the case metagenomes was *Escherichia coli* in 12 samples, *Bifidobacterium longum* in three samples, *Enterococcus faecium* in two samples, *Bacteroides fragilis* in two samples, and human gammaherpesvirus 4, *Bifidobacterium pseudocatenulatum, Phocaeicola vulgatus, Phocaeicola dorei, Limosilactobacillus fermentum, Klebsiella* sp. KP20-425-1*, Staphylococcus haemolyticus, Klebsiella pneumoniae and Bifidobacterium breve* in one sample each. The most abundant microbial species in the control samples were *Blautia wexlerae* (four samples), *Bifidobacterium longum* (three samples), *Ruminococcus gnavus* (three samples), *Bacteroides uniformis* (two samples), *Bacteroides fragilis* (two samples), *Prevotella copri* (two samples), *Phocaeicola dorei* (one sample)*, Bifidobacterium bifidum* (one sample)*, Bifidobacterium breve* (one sample)*, Megamonas funiformis* (one sample) and *Phocaeicola vulgatus* (one sample).

6.4.2.2 Metagenome resistomes

The resistomes (AMR gene variants and unique genes, and classes of antimicrobials to which the genes confer resistance) of cases and controls were compared. The number of AMR variants identified ranged between 1-119 (median = 41) and 36-89 (median = 62) for case and control metagenomes, respectively (Figure 6.2). The number of unique genes ranged between 1-82 (median = 33.5) for cases and 23-64 (median = 45) for controls, whereas the number of antimicrobial classes to which these genes confer resistance ranged between 1-13 (median = 10) in cases and 6-12 (median = 9) in controls. Mann-Whitney U tests revealed no significant difference between the number of unique AMR genes ($p = 0.132$) and the number of antimicrobial classes to which these genes confer resistance ($p = 0.697$) between the case and control metagenomes (Figure 6.2). The genes identified in both sample groups confer resistance to aminoglycoside, beta-lactam, diaminopyrimidine, fluoroquinolone, glycopeptide, lincosamide, macrolide, polymyxin, phenicol, phosphonic acid, rifamycin, sulphonamide and tetracycline antimicrobials. Both groups also contained genes that confer resistance to multiple classes. One case metagenome additionally contained a fusidane resistance gene. In contrast, aminocoumarin and nitroimidazole resistance genes were only identified in the control group.

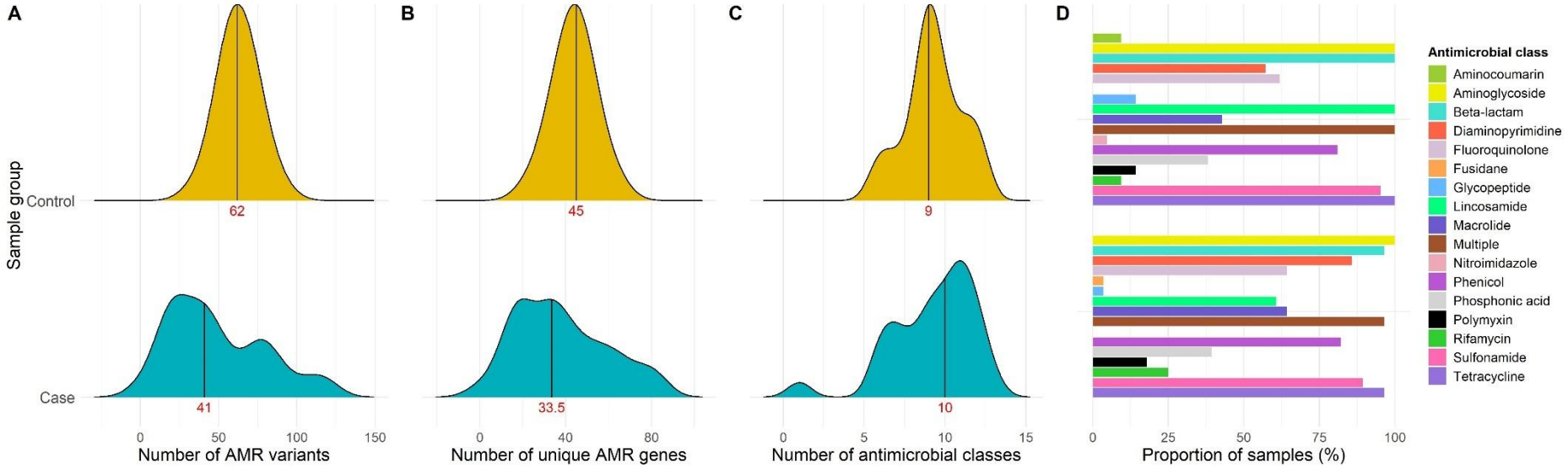

Figure 6.2: Comparison of the distribution of the number of antimicrobial resistance (AMR) variants (A) and unique genes (B), and the number of antimicrobial classes represented by the AMR genes (C) identified in the case and control metagenomes with the median indicated in red, as well as the percentage of metagenomes containing genes conferring resistance to specific antimicrobial classes (D)

6.4.3 Detection of *Salmonella enterica* and the isolated serovars in the associated faecal metagenomes

The relative abundance of *S. enterica* in the case metagenomes was compared to *S. enterica* relative abundance in healthy control faecal samples to assess the suitability of metagenomics for the identification of *Salmonella* directly from culture-confirmed faecal samples.

S. enterica was identified in all of the faecal metagenomes, though the relative abundance varied. Specifically, the relative abundance of *S. enterica* in the case metagenomes ranged between 0.00432-18.3% (median = 0.173%) with Centrifuge and 0.00259-27.7% (median = 0.735%) with Bracken (Appendix 11). Kendall's correlation tests determined that there was no significant correlation (p = 0.150-0.298) between the relative abundance of *S. enterica* and the overall metagenome reads after trimming and host depletion, regardless of the classification method.

S. enterica reads were identified in all of the control metagenomes with Centrifuge and 81% of the control metagenomes with Bracken. Relative abundance ranged between 0.00381-0.0546% (median = 0.00719%) according to Centrifuge and 0.00-0.230% (median = 0.00330%) of total metagenome reads according to Bracken (Appendix 12). Mann-Whitney U tests revealed that the relative abundance of *S. enterica* in the cases and controls varied significantly for both Centrifuge (p = <0.001) and Bracken (p = <0.001). However, some overlap between the cases and controls was evident. Relative abundances of at least 0.0581% were associated exclusively with case metagenomes following analysis with Centrifuge, with 21 samples at or above this threshold. For Bracken, relative abundances of 0.598% or higher were associated exclusively with case metagenomes, with 16 samples at or above the threshold. All of the metagenomes above the Bracken threshold were also above the Centrifuge threshold. However, the relative abundances for individual samples differed following analysis with the two tools (Appendix 11 and Appendix 12).

The serovar of the *S. enterica* genomes was identified in 22 (78.6%) of the associated case metagenomes with Centrifuge. Amongst the samples in which the serovar of the isolated *Salmonella* was identified, the relative abundance ranged between 0.00000208-1.82% (median = 0.0000745%). There was no significant association between the relative abundance of the genome serovar in the metagenome and the overall metagenome reads (p = 0.953).

The serovar reference displaying the highest chromosome coverage through alignment to a *Salmonella* reference database, assembly and realignment ranged between 0.243%-92.8% (median = 3.89%), and the isolated serovar was predicted as the most likely serovar present in 10 metagenomes. SISTR cgMLST analysis on the assemblies resulted in a serovar prediction for all metagenomes, although the majority of assignments issued a QC warning or fail due to missing

cgMLST loci. This was expected, as the tool was not designed for use with fragmented metagenome assemblies, and thus the results should be treated with caution. The serovar of the associated genome was predicted in eight metagenomes (Appendix 11). *Salmonella* MAGs displaying over 10% completeness were identified in six metagenomes with Maxbin2 and five metagenomes with Metabat2 (Appendix 13). Metabat2 assembled 1-3 *S. enterica* MAGs in five samples, whereas Maxbin2 analysis resulted in one *S. enterica* MAG in six samples. Kmerfinder classifications of the MAGs revealed the associated genome serovar in four and three metagenomes for Metabat2 and Maxbin2, respectively. Altogether, alignment and assembly methods identified or predicted presence of the associated genome serovar in 12 out of 28 samples, including two samples in which reads associated with the isolated serovar were not identified with Centrifuge.

S. Meleagridis was not identified in the associated metagenome at all. Identification was not possible with the serovar alignment method as a reference genome was not available in the database, thus limiting the ability to test the efficacy of metagenomics for serovar characterisation in this sample. Another metagenome (metagenome_6916498) also did not produce any MAGs; this was the smallest sample in the dataset in terms of total read number (30,576 total reads).

The control samples were also screened for *S. enterica* serovars. According to Centrifuge, the most abundant *S. enterica* serovar represented 0.000172-0.00806% (median = 0.000356%) of the metagenome reads (Appendix 12). The serovar predicted through alignment, assembly and realignment displayed coverage between 0.0396-6.87% (median = 2.26%), again indicating overlap with the case samples. SISTR cgMLST also gave a serovar prediction for all control metagenomes, indicating presence of *S*. Enteritidis in nine metagenomes, *S*. Bredeney in three metagenomes, *S*. Rissen in three metagenomes, *S*. Typhimurium in three metagenomes, *S*. Choleraesuis in two metagenomes and *S*. Mikawasima in one metagenome.

6.4.4 Detection of antimicrobial resistance genes identified in the associated *Salmonella enterica* genomes with metagenomics

Characterisation of *S. enterica* AMR genotypes directly in stool metagenomes would be valuable for guiding treatment in severe salmonellosis or invasive disease. The case metagenomes were therefore screened for AMR genes identified in the isolated *Salmonella*. The proportion of AMR genes in the genomes identified in the associated metagenomes ranged between 0.00-100% (Figure 6.3, Appendix 9, Supplementary Table S11). All of the unique AMR genes identified in the genomes were also identified in seven (25.0%) of the associated metagenomes. Kendall's

correlation analysis revealed that there was a significant correlation between the proportion of unique genome AMR genes identified in the associated metagenomes and the relative abundance of *S. enterica* in the metagenomes based on Centrifuge ($p = 0.005$) and Bracken ($p = <0.001$) analysis. However, the overall number of metagenome reads was not statistically significantly correlated with the proportion of AMR genes identified ($p = 0.113$).

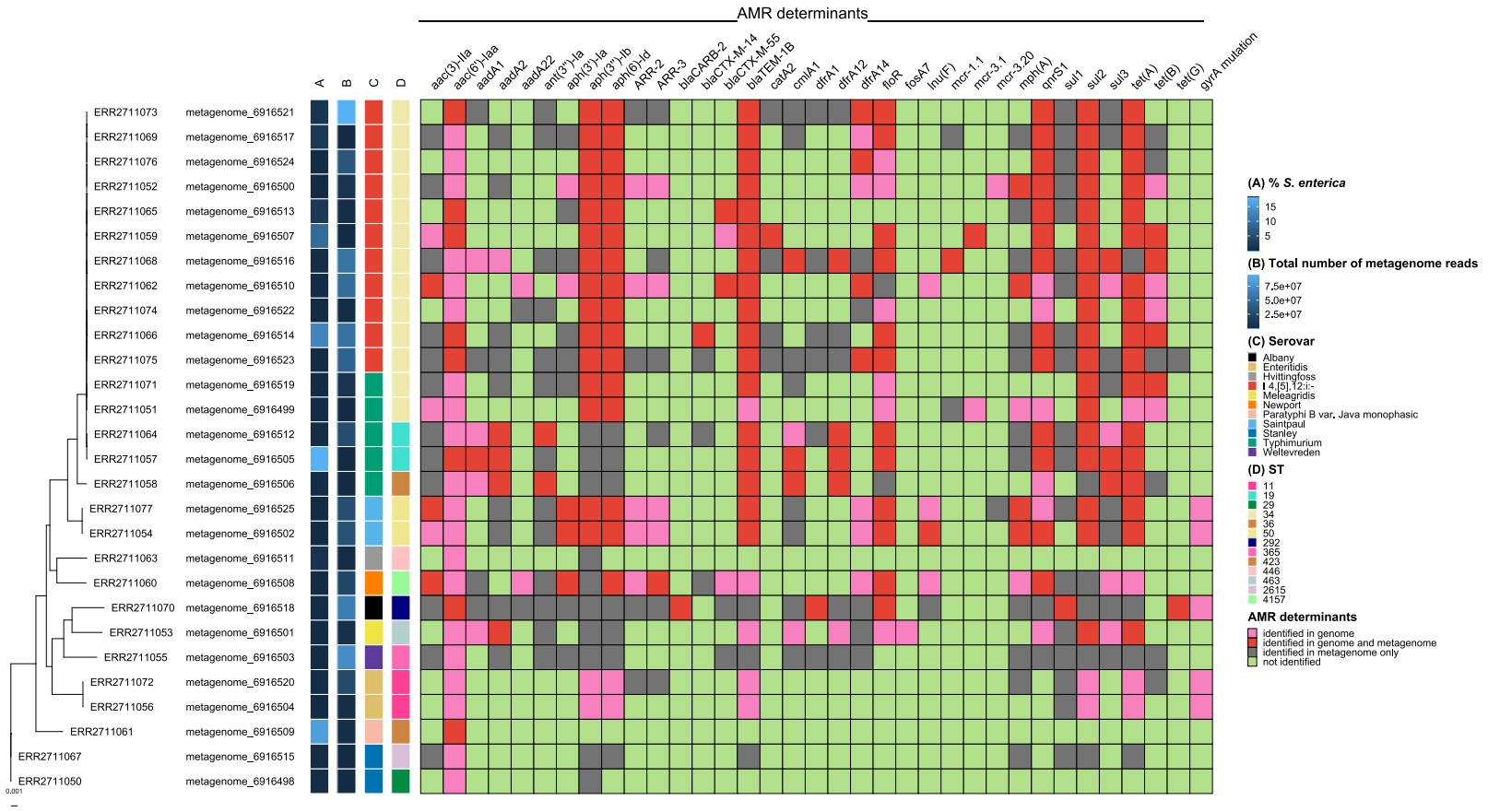

Figure 6.3: Maximum likelihood tree based on core gene alignment of the 28 *Salmonella enterica* genomes, displaying their serovars, sequence types (STs) and the antimicrobial resistance (AMR) determinants identified with KMA; the AMR determinant matrix also indicates whether the gene was identified in the associated faecal metagenome, the relative abundance of *S. enterica* determined with Centrifuge and overall number of reads in the metagenome

Some genes commonly identified in the *S. enterica* genomes and the case metagenomes from this study, such as *aph(3")-Ib*, *aph(6)-Id*, bla_{TEM-1B}, *sul2* and *tet(A)*, were also commonly identified in the control metagenomes (Figure 6.4, Supplementary Table S12). On the other hand, *aac(6')-Iaa*, *aadA22*, $bla_{CTX-M-55}$, *lnu(F)*, *mcr-3.1* and *mcr-3.20* were not identified in the control metagenomes. The *aac(6')-Iaa* gene, endogenous to *Salmonella* and identified in all of the *Salmonella* genomes analysed, was only identified in a small proportion (28.6%) of the case metagenomes.

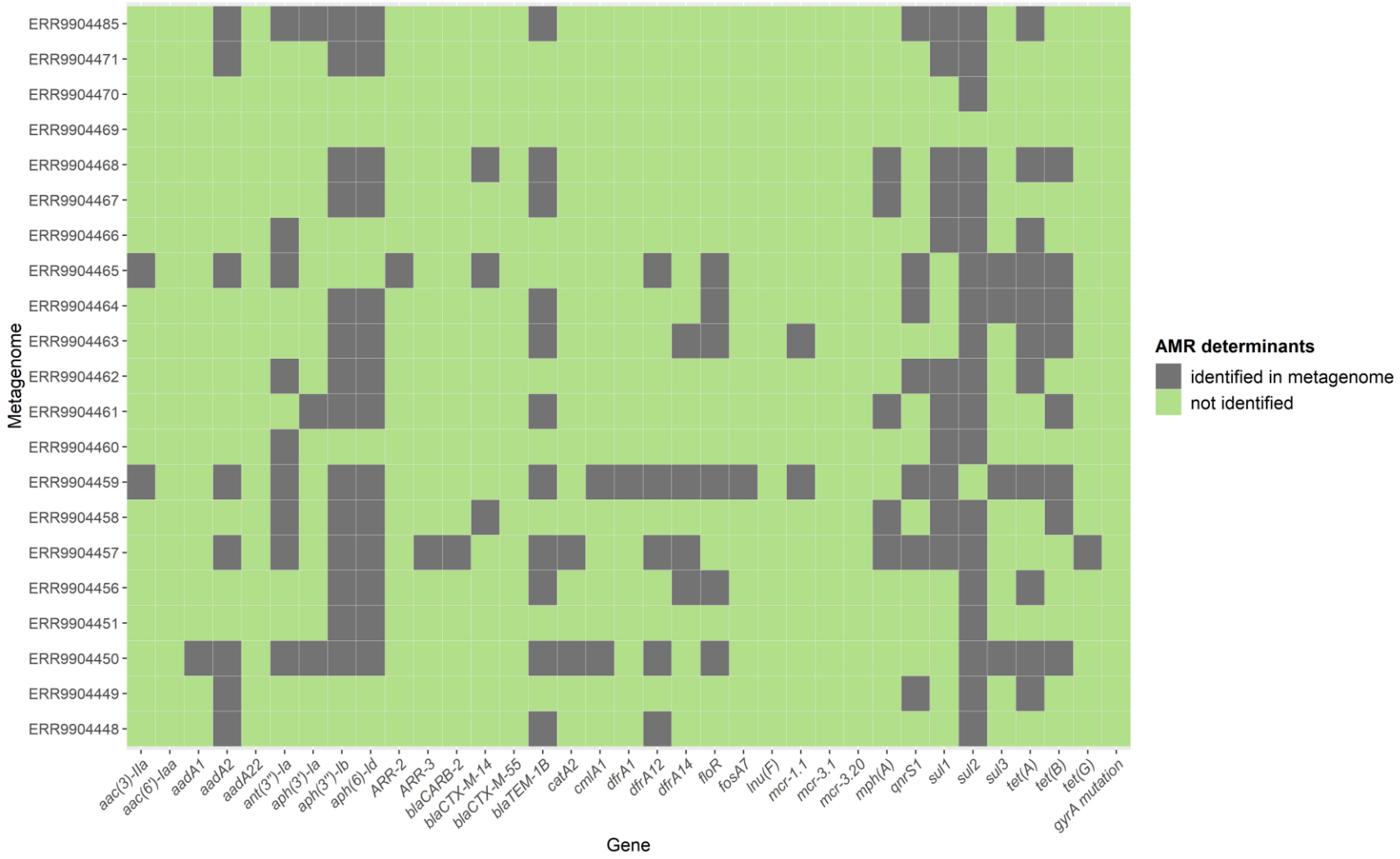

Figure 6.4: Antimicrobial resistance (AMR) genes associated with the *Salmonella enterica* genomes in the control faecal metagenomes

The proportions of samples containing genes conferring resistance to drug groups specifically used in the treatment of *S. enterica* infection were similar between the case and control groups. The proportion of control samples with genes conferring resistance to beta-lactams (100%) was not significantly different (Fisher's exact test; p = 1) to the case metagenomes (96.4%). Similarly, the lower proportion of case metagenomes containing genes conferring resistance to macrolides (92.9%) compared to controls (100%) was not significantly different (p = 0.5). Genes conferring resistance to fluoroquinolones were identified in a higher proportion of controls (71.4%) than case metagenomes (67.9%), but the difference was not significant (p = 1). Genes conferring resistance to multiple drug classes were included in these comparisons, if the drug classes they conferred resistance to included beta-lactams, macrolides and fluoroquinolones. It is important to note that known fluoroquinolone resistance is often conferred through single nucleotide polymorphisms (SNPs) in DNA gyrase and topoisomerase genes (Jacoby, 2005); such SNPs could not be identified in the metagenomes, which may affect the proportions of samples with fluoroquinolone resistance determinants reported.

The metagenome reads were assembled, and contigs binned into MAGs using two different tools (6.3.5). The *Salmonella* MAGs obtained were also screened for AMR genes, to determine whether or not the genes identified in the isolated and sequenced *Salmonella* were also detected on *Salmonella* MAGs from the associated samples. The Metabat2 MAGs contained either no AMR genes or one AMR gene and the Maxbin2 MAGs contained 2-10 AMR genes (Table 6.5 and Table 6.6). The AMR gene identified in the Metabat2 MAGs (samples metagenome_6916509 and metagenome_6916514) was also identified in the *Salmonella* isolated from those samples. However, the Maxbin2 MAGs from these samples contained *erm* genes that were not identified in the associated genomes. This was also the case with some of the genes identified in the other Maxbin2 MAGs. The contigs of the Maxbin2 MAGs containing AMR genes were interrogated using BLASTn to investigate to which bacterial genomes they map. This revealed that the contigs containing *mecA* and *erm* genes were mostly associated with Gram positive bacteria. The *mecA* gene was exclusively associated with *Staphylococcus*. Other genes (*dfrA12, tet(M), qnrB4*, *ant(3'')-Ia*, *bla*$_{\text{DHA-1}}$, and *bla*$_{\text{CTX-M-15}}$, *bla*$_{\text{CTX-M-27}}$, *aph(3')-Ia*, *lnu(F)*, *mef(B)* and *bla*$_{\text{CTX-M-55}}$) were associated with plasmids or chromosomes of Enterobacterales, including *Salmonella*, or other Gram negative bacteria. The AMR genes identified in the isolated *S. enterica* were always associated with plasmids or chromosomes of Enterobacterales or Vibrionales.

Table 6.5: Antimicrobial resistance (AMR) genes identified in the Metabat2 metagenome assembled genomes (MAGs) from the case metagenomes

Metagenome	MAG	Contig	Gene	Coverage (%)	Identity (%)	Identified in genome
metagenome_6916509	metabat_509_bin.1	-	-	-	-	-
	metabat_509_bin.8	k141_5602	*aac(6')-Iaa*_1	100	98.63	Yes
metagenome_6916513	metabat_513_bin.6	-	-	-	-	-
metagenome_6916514	metabat_514_bin.8	-	-	-	-	-
	metabat_514_bin.9	-	-	-	-	-
	metabat_514_bin.10	k141_9075	*aac(6')-Iaa*_1	100	100	Yes
metagenome_6916518	metabat_518_bin.21	-	-	-	-	-
metagenome_6916521	metabat_521_bin.13	-	-	-	-	-
	metabat_521_bin.34	-	-	-	-	-

Table 6.6: Antimicrobial resistance (AMR) genes identified in the Maxbin2 metagenome assembled genomes (MAGs) from the case metagenomes and BLASTn hits for the contigs

Metagenome	MAG	Contig	Gene	Coverage (%)	Identity (%)	Identified in genome	BLASTn hits (C=chromosome, P=plasmid)
metagenome_6916505	maxbin_505_002	k141_2425	*mecA*_6	63.73	100	No	*Staphylococcus* (C)
	maxbin_505_002	k141_2658	*sul1*_5	100	99.89	No	Enterobacterales (C,P), Aeromonadales (C,P)
	maxbin_505_002	k141_3063	*tet(M)*_8	100	96.15	Yes (ARIBA)	Enterobacterales (C,P)
	maxbin_505_002	k141_3416	*qnrB4*_1	91.01	100	No	Enterobacterales (C,P)
	maxbin_505_002	k141_6254	*ant(3'')-Ia*_1	82.51	99.75	No	Enterobacterales (C,P)
	maxbin_505_002	k141_6254	*cmlA1*_1	100	99.92	Yes	
	maxbin_505_002	k141_6254	*aadA2*_1	97.92	100	Yes (KMA)	
	maxbin_505_002	k141_6254	*dfrA12*_8	100	100	Yes	
	maxbin_505_002	k141_6935	$bla_{\text{DHA-1}}$_1	90.53	100	No	Enterobacterales (C,P)
	maxbin_505_002	k141_8083	*erm(C)*_13	72.11	100	No	Gram-positive bacteria (C,P)
metagenome_6916509	maxbin_509_006	k141_4732	*erm(B)*_18	100	99.34	No	Gram-positive bacteria (C,P)
	maxbin_509_006	k141_5602	*aac(6')-Iaa*_1	100	98.63	Yes	
metagenome_6916513	maxbin_513_007	k141_24986	$bla_{\text{CTX-M-15}}$_1	100	100	No	Enterobacterales (C,P)
	maxbin_513_007	k141_25055	$bla_{\text{CTX-M-27}}$_1	100	100	No	Enterobacterales (C,P)
	maxbin_513_007	k141_25222	*aac(6')-Iaa*_1	100	100	Yes	
	maxbin_513_007	k141_6097	*qnrS1*_1	100	100	Yes	Enterobacterales (C,P), Vibrionales (P)
	maxbin_513_007	k141_8401	*erm(X)*_1	99.88	98.01	No	Gram-positive bacteria (C,P), *Gardnerella* (C)
metagenome_6916514	maxbin_514_008	k141_7549	*erm(X)*_2	90.81	97.4	No	Gram-positive bacteria (C,P), *Gardnerella* (C)
	maxbin_514_008	k141_9075	*aac(6')-Iaa*_1	100	100	Yes	
metagenome_6916518	maxbin_518_006	k141_14134	*aac(6')-Iaa*_1	100	98.63	Yes	
	maxbin_518_006	k141_1957	*aph(3')-Ia*_3	99.88	99.75	No	Enterobacterales (C,P)
	maxbin_518_006	k141_5827	*lnu(F)*_1	100	100	No	Enterobacterales (C,P)
	maxbin_518_006	k141_5827	*ant(3'')-Ia*_1	99.59	97.43	No	
	maxbin_518_006	k141_8656	*mef(B)*_1	90.24	100	No	Enterobacterales (C,P), Pasteurellales (C)
	maxbin_518_006	k141_9771	*dfrA12*_8	100	100	No	Enterobacterales (C,P), Pseudomonadales (C)
	maxbin_518_006	k141_9911	$bla_{\text{CTX-M-55}}$_1	100	100	No	Enterobacterales (C,P), Vibrionales (P)
metagenome_6916521	maxbin_521_014	k141_24086	*aac(6')-Iaa*_1	100	100	Yes	
	maxbin_521_014	k141_4434	*erm(X)*_4	100	97.19	No	Gram-positive bacteria (C,P), *Gardnerella* (C)

6.4.5 Consideration of the wider sample microbiome in the clinical context

To investigate differences in gut microbial profiles associated with *Salmonella* infection, the relative abundance of Enterobacteriaceae in the case and control samples were compared. This was generally higher in the case metagenomes (0.219-96.6%, median = 40.2%) compared to controls (0.0246-19.4%, median = 1.47%); this difference was statistically significant (p = <0.001). However, there was an overlap between cases and controls (Figure 6.1), as previously observed when comparing *S. enterica* relative abundance. Enterobacteriaceae relative abundances over 5% were observed in 21 case and six control metagenomes. In most cases, the Enterobacteriaceae reads were largely represented by *E. coli* (Figure 6.5). However, in three case metagenomes and one control metagenome, the Enterobacteriaceae levels observed were related to *K. pneumoniae* or *K*. sp. KP20-425-1 (metagenome_6916508, metagenome_6916514, metagenome_6916521 and ERR9904459), whereas in two control metagenomes the Enterobacteriaceae reads were predominated by a combination of *Klebsiella* and *Enterobacter* species (ERR9904457 and ERR9904462). In two case metagenomes, the Enterobacteriaceae reads were almost entirely represented by *S. enterica* itself (metagenome_6916507 and metagenome_6916509; Appendix 10).

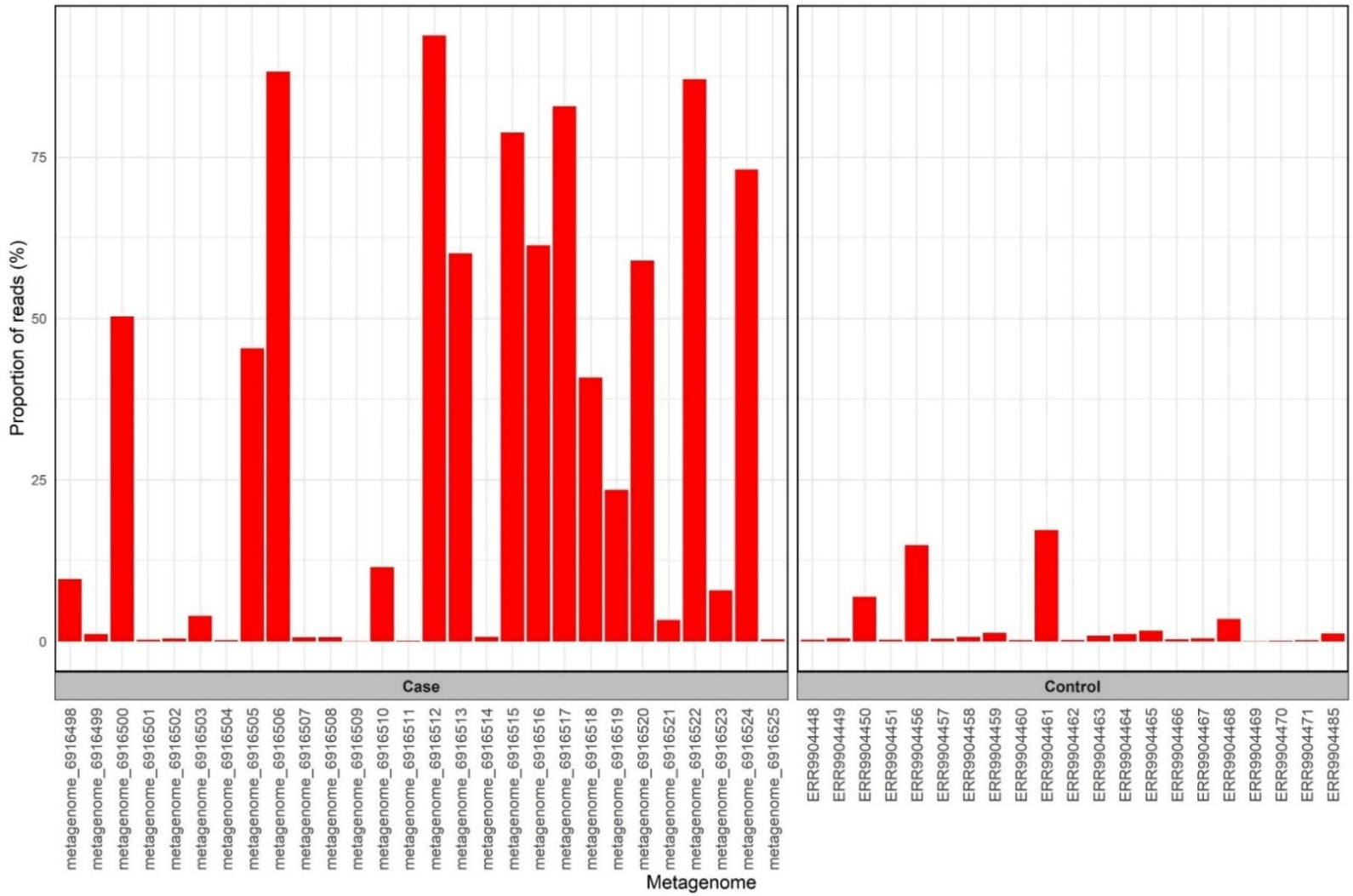

Figure 6.5: Relative abundances of *Escherichia coli* across the case and control metagenomes, determined with Bracken

A number of potential diarrheagenic bacterial species were identified amongst the top 20 most abundant species in the case metagenomes (Appendix 10). *S. enterica* was among the 20 most abundant species in 20 (71.4%) of the case metagenomes and none of the control metagenomes. In 12 case metagenomes, other potential diarrhoeagenic bacterial species (*Shigella flexneri, Shigella boydii, Shigella dysenteriae, Shigella sonnei, Staphylococcus aureus, Clostridium perfringens* and *Clostridioides difficile*) were identified amongst the 20 most abundant species. *S. flexneri* was identified in 10 case metagenomes with the relative abundance ranging between 0.0599-5.16%, *S. boydii* in eight (0.0198-0.694%), *S. dysenteriae* in five (0.0251-0.485%), *S. sonnei* in four (0.0295-0.132%), *S. aureus* in four (0.701-19.7%), *C. perfringens* in one (0.0720%) and *C. difficile* in two (0.193-0.265%). Considering the top 20 most abundant species, at least one of the potential diarrhoeagenic species was more abundant than *S*. *enterica* in seven samples; specifically, *S. aureus* was more abundant than *S. enterica* in one metagenome, and at least one *Shigella* species was more abundant than *S. enterica* in six metagenomes. *C. difficile* was also identified amongst the 20 most abundant species in one control (0.687%) metagenome.

The number of unique *Salmonella* associated AMR genes (i.e. those identified in the associated genome) identified in the metagenome ranged between 0-11 (Figure 6.6A). The relative abundances of *Salmonella* AMR genes were calculated and compared to the overall AMR gene abundance in each sample (Figure 6.6B). *Salmonella* AMR genes represented 0.00-67.1% of the overall sample resistome.

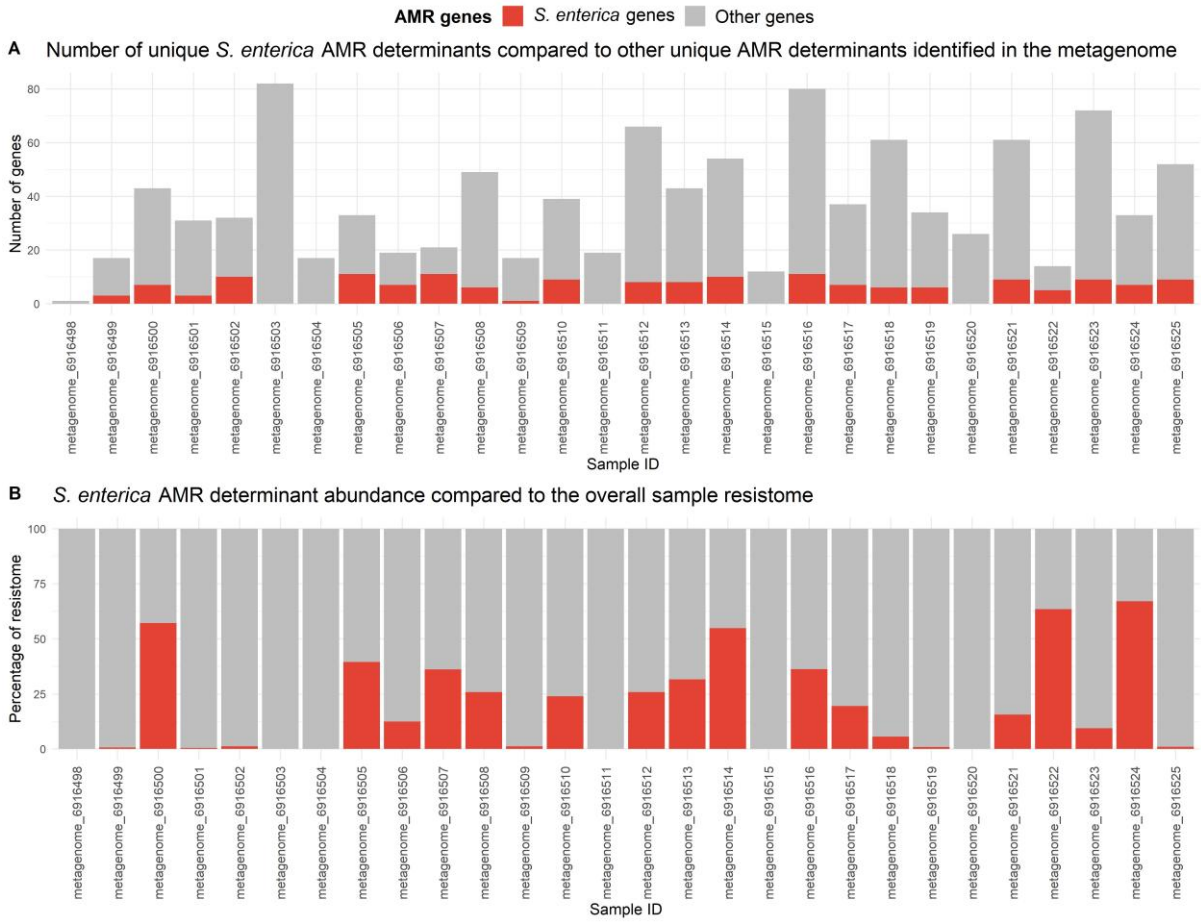

Figure 6.6: Comparison of the number of unique *Salmonella enterica* associated antimicrobial resistance (AMR) genes and other genes present in the associated faecal metagenomes according to KMA (A) and the relative abundance of unique *S. enterica* AMR genes compared to the overall AMR gene repertoire in the metagenomes (B)

6.5 Discussion

Metagenomic identification of organisms considered to be contaminating rather than colonising food samples can be difficult due to their low abundance, as observed with *Campylobacter* on retail chicken (Chapter 5). However, clinical samples may contain a higher abundance of the pathogen potentially responsible for the observed infection due to pathogen replication, making metagenomic identification more likely. Pathogen colonisation may also be associated with observable changes in the microbiome. This work aimed to assess the use of metagenomics for the identification of *S. enterica, S. enterica* serovars and associated AMR genes as well as infection related microbiome changes in faecal metagenomes from culture-confirmed salmonellosis cases. The resistomes of the case metagenomes were also compared to healthy controls, to determine whether or not infection cases can be distinguished from symptom-free controls for treatment guidance.

6.5.1 Metagenomics for the identification of *Salmonella enterica* and the serovar isolated from salmonellosis cases

The work described in the current chapter provided evidence to suggest that metagenomics can be used to detect the diarrhoeagenic pathogen, as evidenced by the identification of *S. enterica* in all 28 case metagenomes. However, the relative abundance between case metagenomes varied, and the pathogen was also detected in up to 100% of healthy control metagenomes depending on classification method. A similar issue was observed in the previous chapter with an overlap in *Campylobacter* abundance between culture positive and negative samples. Although in this chapter the relative abundance of *S. enterica* in the control metagenomes (n = 21) was significantly lower than in cases, some overlap between the sample groups was evident, thus making diagnostic cutoffs ambiguous. One potential explanation for this overlap is pathogen carriage in the healthy control cohort. The individuals from whom the control samples were derived did not have episodes of diarrhoea in the six months leading up to sampling, and the samples were screened for gastrointestinal pathogens including *Salmonella* using a nucleic acid test kit (Luminex xTag) (Pereira-Dias *et al.*, 2021). However, such kits are subject to detection limits, thus it is possible that pathogens of interest may not always be identified. This kit has previously displayed 83.3-92.3% sensitivity in detection of *Salmonella* spp. from clinical samples (Navidad *et al.*, 2013; Deng *et al.*, 2015; Duong *et al.*, 2016), leading to speculation that the control samples that displayed higher *S. enterica* relative abundances could represent potential asymptomatic carriers. Although this is generally considered to be rare (Gal-Mor, 2018), such asymptomatic carriage of *Salmonella* has been previously reported in Vietnam (Thompson *et al.*, 2015; The *et al.*, 2018).

The range in the *S. enterica* relative abundance in the case metagenomes and overlap with control metagenomes could represent biological differences observed at different stages of *S. enterica* infection. Low *S. enterica* relative abundances in symptomatic case samples are not unusual and have been observed previously (Singh *et al.*, 2015; The *et al.*, 2018). During salmonellosis, the pathogen must overcome infection barriers that prevent colonisation, including challenge from the innate gut microbiome (Sibinelli-Sousa *et al.*, 2022). An important step in *Salmonella* infection that allows the pathogen to establish itself in the gut is the induction of an inflammatory state, which is facilitated by the expression of *Salmonella* pathogenicity islands and production of effectors. One of the outcomes of this is an increase in oxygen levels in the gut lumen, allowing *Salmonella* to utilise aerobic respiration, which results in expansion. It takes time for these processes to occur, thus the abundance of the pathogen can vary depending on when the samples were taken (Sibinelli-Sousa *et al.*, 2022).

The aim of the current work was to also test whether the pathogen can be further characterised with metagenomics in clinical cases, as this was not viable in most chicken meat metagenomes containing *Campylobacter* (Chapter 5). Here, characterisation included identification of the *Salmonella* serovar present. A total of 11 different *S. enterica* serovars were identified from the genomic data in this study. Although the isolated serovar was identifiable in up to 78.6% of the associated case metagenomes, serovars were also identified in the healthy control samples. There was again an overlap in the relative abundance of the isolated *S. enterica* serovar in the case metagenomes and the most abundant serovar in the control metagenomes based on Centrifuge analysis. Alignment of reads to reference *Salmonella* genomes, local assembly and realignment displayed similar limitations, whereby the proportion of the reference genome covered for assemblies from some case samples was very low, resembling proportions observed in the healthy control group. SISTR cgMLST analysis yielded predictions for all samples, including controls, accompanied by warnings regarding missing cgMLST loci, thus affecting reliability of the predictions. The alternative approach of *de novo* assembly followed by MAG assembly resulted in the identification of the isolated serovar in a small proportion (17.9-21.4%) of the case metagenomes. This all indicates that although the metagenomic characterisation efficacy is better for infection cases than food samples, the reliability of serovar characterisation with these methods may still be limited in absence of culture and genomic data.

The findings also highlight important limitations and considerations regarding the analysis approach used in metagenomics studies, and previous studies have also highlighted that the selection of tools and databases can have large implications on findings (Zhou *et al.*, 2016; Doster *et al.*, 2019). Some of these, including the risk of false positive classifications with taxonomic profiling tools, were highlighted in the previous chapter, but also apply here. This may imply that low-level detection of *S. enterica* and *S. enterica* serovars in the control samples can be a result of false positive hits. Additionally, the differences between Centrifuge and Bracken results observed for *S. enterica* relative abundances can be a result of differences in tool algorithms, parameters, databases and the way in which relative abundance is inferred between the two approaches. Database limitations can also explain the lack of identification of particular serovars, as evidenced here by the lack of *S.* Meleagridis (n = 1) identification. A reference genome for this serovar was not available in NCBI, and therefore this serovar was not present in the database for alignment. Although a draft genome could be used instead, these are often fragmented; alignment of a highly fragmented assembly of *S. enterica* reads from the metagenomes to another fragmented sequence would likely result in low coverage for this serovar, making identification difficult.

Additional limitations that may affect the ability to detect and characterise the pathogen at the serovar level pertain to the sample preparation and metagenomic sequencing approaches.

Although there was no significant correlation between the relative abundance of *S. enterica* and overall number of reads in the case metagenomes, there was a large range in the read number (30,576-94,890,832) compared to the control metagenomes (10,821,636-15,177,561). Influx of immune cells into the gastrointestinal tract during infection, shedding of the gut lining, or presence of blood in the stool (Peterson *et al.*, 2022), can potentially result in a high proportion of host DNA in samples, affecting the coverage of taxa of interest and thus varying relative abundance of *S. enterica* and other taxa (Broz, Ohlson and Monack, 2012; Pereira-Marques *et al.*, 2019). Use of improved host DNA depletion prior to sequencing could help improve pathogen coverage.

6.5.2 Metagenomics for the identification of *Salmonella enterica* associated antimicrobial resistance genes and comparison with healthy controls

This chapter also aimed to explore the efficacy of metagenomics in resistome comparisons between cases and controls for potential application in treatment guidance, which may be particularly important in paediatric cases that are more at risk of severe infection.

The current chapter found that overall community resistome profiling may not be particularly useful for treatment guidance. Many of the AMR genes identified in the genomes and case metagenomes were also present in the control metagenomes (Figure 6.4), and there was no significant difference in the overall number of unique AMR genes or the number of antimicrobial classes associated with the genes between the cases and controls. Comparisons of the proportion of samples containing resistance genes relevant to antimicrobial classes used for salmonellosis treatment, specifically beta-lactams, macrolides and fluoroquinolones (Wen, Best and Nourse, 2017), also revealed no significant difference between cases and controls. This indicates that the ability to accurately profile the AMR genes carried by the pathogen specifically may be important in this context, in order to establish metagenomics as an efficient method for pathogen detection and treatment guidance.

The number of unique AMR genes identified in the *S. enterica* genomes analysed in this chapter ranged from 1-18 (median = 9) and MDR genotypes were observed in 82.1% of genomes. Between 0.00-100% of the AMR genes in the genomes were identified in the associated metagenomes and there was a significant association between this and the *S. enterica* relative abundance in the associated metagenomes, indicating that the coverage of *S. enterica* in the metagenome is linked with the identification of *S. enterica* associated AMR genes. *S. enterica* associated AMR genes also represented high proportions of the resistome in several cases (up to 67.1%). It is important to note that *Salmonella* associated AMR genes were defined as AMR genes

identified in the *Salmonella* isolated from the stool samples. Many of these genes may be carried by multiple different organisms, and attribution of specific AMR genes to *S. enterica* specifically may not be possible using mapping approaches applied to unassembled metagenome reads. The *aac(6')-Iaa* gene identified in all of the genomes is endogenous to *Salmonella* (Magnet, Courvalin and Lambert, 1999; Salipante and Hall, 2003) and can thus be used to determine the efficacy of detection of *Salmonella* specific genes. The identification of this gene in the case metagenomes was generally poor, indicating that AMR genes carried specifically by the pathogen of interest may be difficult to identify directly from stool metagenomes.

Finding AMR genes carried by *S. enterica* specifically may in theory be possible with MAG assembly, although the two MAG assembly tools used in the current chapter yielded different results. The Metabat2 MAGs contained either no AMR genes or one AMR gene and the Maxbin2 MAGs contained up to 10 AMR genes. *S. enterica* AMR genes may be carried on plasmids (Emond-Rheault *et al.*, 2020), and plasmid sequences are often excluded from MAGs (Maguire *et al.*, 2020), which may explain the low AMR gene identification on the Metabat2 MAGs. However, when the Maxbin2 MAG contigs containing AMR genes were screened against publicly available sequences, most mapped either to distantly related bacteria, or chromosomes or plasmids of Enterobacterales or Vibrionales, indicating possibility of both plasmid and chromosomal sequences being incorporated into the MAGs. There were differences in the AMR genes identified in the *Salmonella* genomes associated with the stool samples and the Maxbin2 MAGs generated and although this could potentially indicate presence of multiple *Salmonella* strains in the samples with different AMR genotypes, it is likely that these MAGs were contaminated with sequences from other bacteria, particularly as such errors have been previously documented (Meziti *et al.*, 2021). MAGs with low completeness are also subject to underestimation of contamination (Parks *et al.*, 2015), which further affects the reliability of MAGs for characterisation of pathogens within stool samples.

6.5.3 Differences in the microbiome between the salmonellosis cases and controls

In the previous chapter, *Campylobacter* contamination was not associated with significant changes in the chicken meat sample microbiome. However, in the current chapter, the case metagenomes (Figure 6.1) were marked with a significant difference in the relative abundance of Enterobacteriaceae between the cases and controls, which was mostly associated with *E. coli* and *Klebsiella* species. Previous research has also described an increase in Enterobacteriaceae abundance in *S. enterica* infection and an overall reduction in bacterial diversity (Aljahdali *et al.*, 2020). This is likely linked to the inflammatory state induced by *Salmonella* to progress infection, as other closely related Enterobacteriaceae are also able to thrive in these conditions (Sibinelli-

Sousa *et al.*, 2022). However, high Enterobacteriaceae relative abundance was not observed in all of the case metagenomes. This could again be potentially related to an earlier stage of infection (Sibinelli-Sousa *et al.*, 2022). Previous research also indicates that host factors including age may play a significant role in microbial shifts (The *et al.*, 2018), which may also be relevant here. The innate gut microbiome, particularly during development in early life, can affect the rate of *Salmonella* clearance and disease progression (Gal-Mor, 2018; Aljahdali *et al.*, 2020) and therefore result in different taxonomic profiles. The participants in this study from whom samples were obtained were under 5 years of age. In early life, the gut microbiome is subject to changes as certain bacterial groups increase in abundance and others decrease (Arrieta *et al.*, 2014). Indeed, the individual differences in the microbial profiles amongst the control group described in this chapter can be partly explained by differences in age between the paediatric participants, as noted by the authors (Pereira-Dias *et al.*, 2021). In LMICs, gut microbiome development can also be confounded by more frequent exposure to pathogens through the environment and food, which can introduce further complexity between individuals (Robertson *et al.*, 2019). This could also partially explain the observation of Enterobacteriaceae relative abundances in some control metagenomes resembling those in the case metagenomes. This highlights that microbiome profiling can be a useful indicator of gastrointestinal disease, but further work is required to understand the microbiome dynamics during infection whilst accounting for these potential confounding variables.

At lower taxonomic levels, the detection of other potential diarrheagenic bacterial species (*Shigella*, *C. perfringens* and *C. difficile*) among the 20 most abundant species in 12 case metagenomes could be indicative of potential coinfections, which have been previously documented in paediatric cases in Southeast Asia (Deng *et al.*, 2015). It is also possible that *S. enterica* is present, but other pathogens may be causing the infection symptoms. However, for taxa like *Shigella*, definitive identification in metagenome data is difficult as *Shigella* is a specialised sub-clone of *E. coli* (Chaudhuri and Henderson, 2012), and it is difficult to discern commensal and pathogenic variants of these organisms, thus these results should be interpreted with caution. *S. aureus* was identified amongst the 20 most abundant species in four case metagenomes, and has been previously associated with food poisoning. However, this requires production of enterotoxins (Argudín, Mendoza and Rodicio, 2010), the presence of which was not evaluated in this study. *C. difficile* was also identified among the 20 most abundant species in two cases and one healthy control metagenome at similar abundance. *C. difficile* is often carried asymptomatically (Poxton, McCoubrey and Blair, 2001), and pathogenesis is associated with toxin production (Di Bella *et al.*, 2016). Toxin presence was not evaluated, thus the presence of these species alone provides insufficient evidence for coinfection. Nonetheless, the ability to identify other potential diarrhoeagenic pathogens highlights the potential of metagenomics-based

applications in clinical settings, and may be another potential reason for differing *S. enterica* relative abundances in the case metagenomes.

6.6 Conclusions

Metagenomics provides a snapshot of the microorganisms present in a sample. For many salmonellosis faecal samples, NTS made up a relatively large proportion of the microorganisms present, and the pathogen could be detected in all case samples using metagenomics. This highlights that pathogen colonisation in clinical cases can result in detectable levels of the pathogen in metagenome samples, which is not often seen when the pathogen is merely a contaminant, as discussed in the previous chapter. Characterisation to the serovar level was possible in a proportion of samples, though improvements in sample processing, sequencing and analysis approaches are required to improve characterisation efficacy. Variation in the proportion of the metagenome made up of NTS between case samples could be potentially attributed to the stage of infection at the time of sampling, individual host microbiome differences during development, and possible coinfections, resulting in some cases that had similar NTS metagenome proportions to controls. The case metagenomes were marked with significant differences in Enterobacteriaceae relative abundances, highlighting that pathogen-associated microbiome changes can be identified in scenarios where the pathogen colonises the host, but less so when it is merely contaminating the sample, as observed in Chapter 5. Many AMR genes identified in salmonellosis cases were also identified in healthy control stools, thus indicating that the resistomes of children with diarrhoea and healthy children are similar. *Salmonella*-specific AMR genes, such as *aac(6')-Iaa*, were seldom identified in the case samples, indicating that reliable AMR genotyping of pathogens with metagenomics alone may not be currently possible in this context. Taken together, the current chapter showcases the potential of metagenomics for pathogen detection and characterisation, but highlights the requirement for optimised host DNA depletion, sequencing and analysis methods to facilitate routine use.

7. Chapter 7: General discussion

Campylobacter spp. are the leading cause of gastroenteritis worldwide (Kaakoush *et al.*, 2015), with chicken meat reported as the leading source of infection (Skarp, Hänninen and Rautelin, 2016; European Food Safety Authority and European Centre for Disease Prevention and Control, 2019). However, the epidemiology of *Campylobacter* is incompletely understood, largely due to case underreporting (de Wit *et al.*, 2001; Wagenaar, French and Havelaar, 2013), as well as difficulties with detection and limited isolate sampling and characterisation. While culturing is the current gold standard method of detection (Harrison *et al.*, 2022), the existence of different culturing standards (Center for Food Safety and Applied Nutrition, 2000; Rosenquist, Bengtsson and Hansen, 2007; Health Canada, 2014; Standards Australia, 2015; International Organization for Standardization, 2017a, 2017b; United States Department of Agriculture, Food Safety and Inspection Service, 2022), and their recognised limitations (Ugarte-Ruiz *et al.*, 2012; Seliwiorstow *et al.*, 2016; Jo *et al.*, 2017) suggest that method choice may affect reported contamination rates. *Campylobacter* populations can be diverse, with evidence of different STs colonising chickens, and possible further cross-contamination during meat processing (Colles *et al.*, 2010; Corry *et al.*, 2017; Würfel *et al.*, 2019; Inglis *et al.*, 2021; Faverjon, Cameron and De Nardi, 2022). Difficulties with culture alongside limited isolate sampling (Food Standards Agency, 2016; Ugarte-Ruiz *et al.*, 2018; Hull *et al.*, 2021; McDermott, 2021; Royden *et al.*, 2021; Habib *et al.*, 2023), mean that the diversity of *Campylobacter* on retail chicken in the UK has not been fully elucidated. Diverse *Campylobacter* populations on retail chicken meat can have significant impacts on surveillance, source attribution and outbreak tracking.

Culture-independent techniques like shotgun metagenomics are emerging as tools for pathogen detection and characterisation (Loman *et al.*, 2013; Zhou *et al.*, 2016; Jesser *et al.*, 2023; Royer *et al.*, 2024). This could be a useful surveillance strategy that enables characterisation of *Campylobacter* from retail chicken, overcoming difficulties with culturing and potentially high within-sample diversity. As an unbiased approach, shotgun metagenomics can additionally facilitate the identification of organisms associated with *Campylobacter* persistence on retail meat, providing avenues for novel intervention strategies to reduce *Campylobacter* prevalence and improve food safety.

This project aimed to address the knowledge gaps regarding *Campylobacter* recovery and diversity on retail meat and determine whether or not metagenomics can be reliably used for pathogen detection and characterisation. The following approaches were utilised:

1. A combination of culture conditions was used to isolate *Campylobacter* from retail chicken, and the diversity of chromosomes and plasmids assessed using WGS.

2. Retail chicken samples were subjected to shotgun metagenomic sequencing to determine if *Campylobacter* can be detected and characterised, and to identify microbial signatures associated with *Campylobacter* presence and absence.
3. In light of the difficulties associated with the identification of low abundance pathogens from food, metagenomic detection of *Salmonella* from stool samples in culture-confirmed salmonellosis cases was assessed, to showcase the potential and limitations of metagenomics for pathogen detection and characterisation.

7.1 Key findings and contributions to the field

7.1.1 Investigation of *Campylobacter* diversity on retail chicken

Utilisation of a combination of culture methods revealed that the recovery of *Campylobacter* in individual samples can be dependent on the culture method used, indicating that the use of different national and international standards can affect *Campylobacter* recovery in surveillance, source attribution and outbreak investigation studies. In this work, individual chicken samples contained up to two different *Campylobacter* species and up to eight different STs, highlighting the importance of characterising multiple isolates per sample. This is particularly important in outbreak investigation scenarios, as STs associated with outbreaks can be missed. This was demonstrated through simulation studies that revealed a requirement of up to 26 isolates to reach 95% average probability of recovering a theoretical outbreak ST based on the ST distribution from the *Campylobacter* culture positive retail chicken samples. Individual ST groups within samples can display further diversity at the SNP and AMR genotype level, suggesting that the resolution of bacterial identification beyond the level of ST may require even more intensive sampling. These findings may have significant resource implications for public health laboratories.

7.1.2 Investigation of *Campylobacter* plasmid diversity

Extending the diversity investigation from the chromosome to plasmids, there was evidence of plasmid sharing between *C. jejuni* and *C. coli* and between different STs within the species. This highlights the value of long read sequencing that can facilitate the study of MGEs, and further highlights the diversity of *Campylobacter* on retail chicken, particularly as individual isolates can carry up to three plasmids. Some *Campylobacter* plasmids can harbour AMR genes conferring tetracycline resistance, and genes that may enhance chicken colonisation and subsequent presence on meat, highlighting the previously overlooked but potential public health significance of *Campylobacter* plasmids. The study of specific plasmids in epidemiologically important STs could help to elucidate the reasons for the prevalence of individual lineages in human infection.

Focusing on a clinically relevant ST (ST-6175) recovered from 11 of the first 45 retail chicken samples in this project, all of the isolates were found to carry the same plasmid, but core unique features associated with the plasmid cluster were not identified; this suggests that the combination of different genes also present in other plasmid clusters could be driving persistence, or that persistence is related to chromosomal features. This work contributes to the limited knowledge of *Campylobacter* plasmid ecology in an epidemiologically important niche.

7.1.3 Metagenomic identification of *Campylobacter* and organisms associated with *Campylobacter* presence and absence on retail chicken

The deep isolate sampling required to capture the *Campylobacter* diversity on retail chicken may not be feasible in public health laboratories due to high costs and time constraints. Time-sensitive situations like outbreak investigations could benefit from culture-independent approaches such as shotgun metagenomics, particularly as sequencing costs continue to decrease. The current work found that the identification of *Campylobacter* directly from retail chicken with metagenomics is difficult due to its low abundance on the meat. Characterisation may only be possible in cases where abundance is higher, as evident in one retail chicken sample in which the isolated ST could be identified. It was also not possible to reliably identify microbial signatures associated with *Campylobacter* presence and absence, which could indicate that there are no true differences, that any differences may be difficult to identify with current approaches, or that current limitations of databases prevent such inferences from being possible. Although the recent advances in metagenomics have increased interest in replacing traditional culturing methods with culture-independent solutions, this work highlighted the need for improvements in the tools and databases for application in food safety studies. Given the low abundance and large diversity of *Campylobacter*, culture methods are still very important.

7.1.4 The potential of metagenomics for pathogen detection and characterisation: *Salmonella* and culture-confirmed salmonellosis cases

In an infection scenario, the pathogen may be more abundant and may modulate the microbiome. This work explored the dynamics of foodborne pathogen infection by looking at *S. enterica* in stool samples, which revealed significant changes in microbial composition at the family level in salmonellosis paediatric cases compared to controls. The pathogen could be detected in all case metagenomes, and the serovar and AMR genes associated with the isolated *S. enterica* could be characterised in a proportion of samples; this highlighted that while the reliable metagenomic detection and characterisation of non-growing pathogens on food is difficult, it can

be a viable approach in clinical cases where the pathogen is colonising and causing an infection. However, the detection of the pathogen, albeit at a low abundance, in healthy control samples may confound the applicability of metagenomics for diagnostic purposes, and more work is needed to evaluate diagnostic cutoffs. Improvements in sample processing, particularly host DNA depletion, as well as sequencing and analysis methods are needed to further increase the utility of metagenomics for clinical application.

7.2 Future directions

This project has provided a number of important insights into the diversity of *Campylobacter* on retail chicken meat and the related public health implications, finding multiple STs, novel STs and diversity within individual ST groups. Future work could apply cgMLST or wgMLST methods to further type the lineages (Cody *et al.*, 2013, 2017); the development of tools enabling rapid and comprehensive strain characterisation, akin to those available for other pathogens like *Salmonella* (Yoshida *et al.*, 2016), could also simplify typing. This would be highly beneficial for public health applications, given the importance of *Campylobacter* as a leading bacterial cause of diarrhoea.

A combination of different culture conditions was used to recover *Campylobacter* from retail chicken in this project (Chapter 3). However, these are still mostly selective for *C. jejuni* and *C. coli*, and thus the diversity of other species could not be reliably inferred. Future work could explore additional methods, for example by changing the gas compositions, incubation time or plating method to aid recovery of other species that could be present on retail chicken (Lynch *et al.*, 2011).

The exploration of plasmid diversity (Chapter 4) was limited to a small dataset (n = 46) covering only 12 retail chicken samples, some of which were represented by only one ST, thus reducing the ability to infer within-sample plasmid sharing between STs. Future work should employ long read sequencing on a larger number of isolates from a larger number of samples to fully elucidate *Campylobacter* plasmid diversity. The mobility of the plasmids should also be evaluated further, though this may require improvement in current databases that lack *Campylobacter* plasmid sequences for typing.

The identification of *Campylobacter* from retail chicken samples with metagenomics was limited by the low abundance of the pathogen (Chapter 5). Recent improvements in metagenomics, such as adaptive sequencing during which reads of the organism of interest are preferentially sequenced (Martin *et al.*, 2022), have been suggested as a viable solution for strain-level detection of pathogens from food in potential outbreak scenarios (Buytaers *et al.*, 2024). However, given the large diversity of *Campylobacter* reported in this project, a large number of

reference genomes used during sequencing would be needed to capture this variation for accurate detection, indicating the need for optimisation. Adaptive sequencing can also be used to exclude host reads, which can make up a large proportion of metagenomes representing human stool samples during active infection (Chapter 6), thus potentially increasing efficacy of pathogen typing as the coverage of microbes is improved (Marquet *et al.*, 2022).

This project has highlighted the limitations of many databases and their lack of representation of microorganisms present on food samples. This underlines the need for further sampling from different niches, including retail food, to populate databases and enable investigations of microbial community associations in order to reduce pathogen burden. Improvements in bioinformatics software are also needed. The analysis of short read genome data in this project predominantly utilised standard pipelines that are commonly applied in genomics, but metagenome analysis as well as long read genome analysis required the use of multiple approaches. This was largely because long read and metagenome analysis methods are less standardised, thus multiple assemblers, read classifiers and MAG algorithms had to be used in attempt to avoid making erroneous conclusions based on the limitations of any one tool. The application of metagenomics for pathogen detection and characterisation may also only be possible in high income settings; despite sequencing costs reducing with time, the cost of culture-independent sequencing may still be prohibitive for routine use in LMICs.

7.3 Final remarks

Overall, this work has significantly contributed to the understanding of *Campylobacter* diversity on retail chicken, which is a leading infection source. In light of the high sampling effort required due to high within-sample diversity, metagenomics was explored as a method for pathogen detection and characterisation, and to identify organisms associated with *Campylobacter* presence and absence for potential intervention. This identified a number of areas that require refinement before metagenomics can be used for pathogen tracking, especially for application in food safety. Application of metagenomics to identify *Salmonella* from paediatric diarrhoeal infection cases demonstrated improved efficacy, but highlighted that further refinement is also needed for clinical applications.

8. References

Abraham, S. *et al.* (2020) 'Emergence of fluoroquinolone-resistant *Campylobacter jejuni* and *Campylobacter coli* among Australian chickens in the absence of fluoroquinolone use', *Applied and Environmental Microbiology*, 86(8), pp. e02765-19. Available at: https://doi.org/10.1128/AEM.02765-19.

Acik, M.N. *et al.* (2013) 'The first isolation of *Campylobacter lanienae* from chickens', *Revue de Medecine Veterinaire*, p. 6.

Adewale, B.A. (2020) 'Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years?', *African Journal of Laboratory Medicine*, 9(1), p. 1340. Available at: https://doi.org/10.4102/ajlm.v9i1.1340.

Adkin, A. *et al.* (2006) 'Use of a systematic review to assist the development of *Campylobacter* control strategies in broilers', *Journal of Applied Microbiology*, 100(2), pp. 306–315. Available at: https://doi.org/10.1111/j.1365-2672.2005.02781.x.

Afgan, E. *et al.* (2018) 'The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update', *Nucleic Acids Research*, 46(W1), pp. W537–W544. Available at: https://doi.org/10.1093/nar/gky379.

Alcock, B.P. *et al.* (2023) 'CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database', *Nucleic Acids Research*, 51(D1), pp. D690–D699. Available at: https://doi.org/10.1093/nar/gkac920.

Aljahdali, N.H. *et al.* (2020) 'Current knowledge and perspectives of potential impacts of *Salmonella enterica* on the profile of the gut microbiota', *BMC Microbiology*, 20(1), p. 353. Available at: https://doi.org/10.1186/s12866-020-02008-x.

Allos, B.M. (1997) 'Association between *Campylobacter* infection and Guillain-Barré syndrome', *The Journal of Infectious Diseases*, 176 Suppl 2, pp. S125-128. Available at: https://doi.org/10.1086/513783.

Altschul, S.F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403–410. Available at: https://doi.org/10.1016/S0022-2836(05)80360-2.

Altschul, S.F. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Research*, 25(17), pp. 3389–3402. Available at: https://doi.org/10.1093/nar/25.17.3389.

An, J.-U. *et al.* (2018) 'Dairy cattle, a potential reservoir of human campylobacteriosis: epidemiological and molecular characterization of *Campylobacter jejuni* from cattle farms', *Frontiers in Microbiology*, 9. Available at: https://doi.org/10.3389/fmicb.2018.03136.

Andersen, S.C. *et al.* (2017) 'Towards diagnostic metagenomics of *Campylobacter* in fecal samples.', *BMC Microbiology*, 17(1), p. 133. Available at: https://doi.org/10.1186/s12866-017-1041-3.

Anisimova, M. *et al.* (2011) 'Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes', *Systematic Biology*, 60(5), pp. 685–699. Available at: https://doi.org/10.1093/sysbio/syr041.

Argudín, M.Á., Mendoza, M.C. and Rodicio, M.R. (2010) 'Food poisoning and *Staphylococcus aureus* enterotoxins', *Toxins*, 2(7), pp. 1751–1773. Available at: https://doi.org/10.3390/toxins2071751.

Arrieta, M.-C. *et al.* (2014) 'The intestinal microbiome in early life: health and disease', *Frontiers in Immunology*, 5. Available at: https://www.frontiersin.org/articles/10.3389/fimmu.2014.00427.

Ayrapetyan, M. and Oliver, J.D. (2016) 'The viable but non-culturable state and its relevance in food safety', *Current Opinion in Food Science*, 8, pp. 127–133. Available at: https://doi.org/10.1016/j.cofs.2016.04.010.

Bacon, D.J. *et al.* (2000) 'Involvement of a plasmid in virulence of *Campylobacter jejuni* 81-176', *Infection and Immunity*, 68(8), pp. 4384–4390. Available at: https://doi.org/10.1128/iai.68.8.4384-4390.2000.

Baffoni, L. *et al.* (2012) 'A *Bifidobacterium*-based synbiotic product to reduce the transmission of *C. jejuni* along the poultry food chain', *International Journal of Food Microbiology*, 157(2), pp. 156–161. Available at: https://doi.org/10.1016/j.ijfoodmicro.2012.04.024.

Barrett, E. *et al.* (2018) 'Post-infectious sequelae after *Campylobacter* enteric infection: a pilot study in Maricopa County, Arizona, USA', *Pilot and Feasibility Studies*, 4(1), p. 142. Available at: https://doi.org/10.1186/s40814-018-0335-z.

Baylis, C. l. *et al.* (2000) 'Comparison of three enrichment media for the isolation of *Campylobacter* spp. from foods', *Journal of Applied Microbiology*, 89(5), pp. 884–891. Available at: https://doi.org/10.1046/j.1365-2672.2000.01203.x.

Beghini, F. *et al.* (2021) 'Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3', *eLife*. Edited by P. Turnbaugh, E. Franco, and C.T. Brown, 10, p. e65088. Available at: https://doi.org/10.7554/eLife.65088.

Bharat, A. *et al.* (2022) 'Correlation between phenotypic and *in silico* detection of antimicrobial resistance in *Salmonella enterica* in Canada using Staramr', *Microorganisms*, 10(2), p. 292. Available at: https://doi.org/10.3390/microorganisms10020292.

Billington, C., Kingsbury, J.M. and Rivas, L. (2022) 'Metagenomics approaches for improving food safety: a review', *Journal of Food Protection*, 85(3), pp. 448–464. Available at: https://doi.org/10.4315/JFP-21-301.

Blaser, M.J. and Engberg, J. (2008) 'Clinical aspects of *Campylobacter jejuni* and *Campylobacter coli* infections', in *Campylobacter*. John Wiley & Sons, Ltd, pp. 97–121. Available at: https://doi.org/10.1128/9781555815554.ch6.

Bloomfield, S.J. *et al.* (2018) 'Long-term colonization by *Campylobacter jejuni* within a human host: evolution, antimicrobial resistance, and adaptation', *The Journal of Infectious Diseases*, 217(1), pp. 103–111. Available at: https://doi.org/10.1093/infdis/jix561.

Bloomfield, S.J. *et al.* (2021) 'Genomic adaptations of *Campylobacter jejuni* to long-term human colonization', *Gut Pathogens*, 13(1), p. 72. Available at: https://doi.org/10.1186/s13099-021-00469-7.

Bloomfield, S.J. *et al.* (2023) 'Determination and quantification of microbial communities and antimicrobial resistance on food through host DNA-depleted metagenomics', *Food Microbiology*, 110, p. 104162. Available at: https://doi.org/10.1016/j.fm.2022.104162.

de Boer, P. *et al.* (2002) 'Generation of *Campylobacter jejuni* genetic diversity *in vivo*', *Molecular Microbiology*, 44(2), pp. 351–359. Available at: https://doi.org/10.1046/j.1365-2958.2002.02930.x.

de Boer, R.F. *et al.* (2013) 'Detection of *Campylobacter* species and *Arcobacter butzleri* in stool samples by use of real-time multiplex PCR', *Journal of Clinical Microbiology*, 51(1), pp. 253–259. Available at: https://doi.org/10.1128/JCM.01716-12.

Bolton, F.J. and Coates, D. (1983) 'A study of the oxygen and carbon dioxide requirements of thermophilic campylobacters', *Journal of Clinical Pathology*, 36(7), pp. 829–834. Available at: https://doi.org/10.1136/jcp.36.7.829.

Bolton, F.J., Coates, D. and Hutchinson, D.N. (1984) 'The ability of *Campylobacter* media supplements to neutralize photochemically induced toxicity and hydrogen peroxide', *The Journal of Applied Bacteriology*, 56(1), pp. 151–157. Available at: https://doi.org/10.1111/j.1365-2672.1984.tb04707.x.

Brenner, F.W. *et al.* (2000) '*Salmonella* Nomenclature', *Journal of Clinical Microbiology*, 38(7), pp. 2465–2467. Available at: https://doi.org/10.1128/jcm.38.7.2465-2467.2000.

Brooks, M.E. *et al.* (2017) 'glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling', *The R Journal*, 9(2), pp. 378–400.

Brown, H.L. *et al.* (2014) 'Chicken juice enhances surface attachment and biofilm formation of *Campylobacter jejuni*', *Applied and Environmental Microbiology*, 80(22), pp. 7053–7060. Available at: https://doi.org/10.1128/AEM.02614-14.

Broz, P., Ohlson, M.B. and Monack, D.M. (2012) 'Innate immune response to *Salmonella* Typhimurium, a model enteric pathogen', *Gut Microbes*, 3(2), pp. 62–70. Available at: https://doi.org/10.4161/gmic.19141.

Buermans, H.P.J. and den Dunnen, J.T. (2014) 'Next generation sequencing technology: advances and applications', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842(10), pp. 1932–1941. Available at: https://doi.org/10.1016/j.bbadis.2014.06.015.

Bull, S.A. *et al.* (2006) 'Sources of *Campylobacter* spp. colonizing housed broiler flocks during rearing', *Applied and Environmental Microbiology*, 72(1), pp. 645–652. Available at: https://doi.org/10.1128/AEM.72.1.645-652.2006.

Burdock, G.A. and Carabin, I.G. (2004) 'Generally recognized as safe (GRAS): history and description', *Toxicology Letters*, 150(1), pp. 3–18. Available at: https://doi.org/10.1016/j.toxlet.2003.07.004.

Burfoot, D. *et al.* (2015) 'Investigation of the efficacy, practicality and cost effectiveness of modified atmosphere packaging on *Campylobacter* numbers on raw chicken intended for retail'. Available at: https://www.food.gov.uk/sites/default/files/media/document/fs101038mapfinalreport.pdf.

Burnham, P.M. and Hendrixson, D.R. (2018) '*Campylobacter jejuni*: collective components promoting a successful enteric lifestyle', *Nature Reviews Microbiology*, 16(9), pp. 551–565. Available at: https://doi.org/10.1038/s41579-018-0037-9.

Buytaers, F.E. *et al.* (2020) 'A practical method to implement strain-level metagenomics-based foodborne outbreak investigation and source tracking in routine', *Microorganisms*, 8(8), p. 1191. Available at: https://doi.org/10.3390/microorganisms8081191.

Buytaers, F.E. *et al.* (2024) 'Strain-level characterization of foodborne pathogens without culture enrichment for outbreak investigation using shotgun metagenomics facilitated with Nanopore adaptive sampling', *Frontiers in Microbiology*, 15. Available at: https://doi.org/10.3389/fmicb.2024.1330814.

Callahan, B.J., McMurdie, P.J. and Holmes, S.P. (2017) 'Exact sequence variants should replace operational taxonomic units in marker-gene data analysis', *The ISME Journal*, 11(12), pp. 2639–2643. Available at: https://doi.org/10.1038/ismej.2017.119.

Campitelli, E. (2022) 'ggnewscale: multiple fill and colour scales in "ggplot2". R package version 0.4.8.' Available at: https://CRAN.R-project.org/package=ggnewscale.

Cappellato, M., Baruzzo, G. and Di Camillo, B. (2022) 'Investigating differential abundance methods in microbiome data: a benchmark study', *PLOS Computational Biology*, 18(9), p. e1010467. Available at: https://doi.org/10.1371/journal.pcbi.1010467.

Carrique-Mas, J.J. *et al.* (2020) 'An estimation of total antimicrobial usage in humans and animals in Vietnam', *Antimicrobial Resistance and Infection Control*, 9, p. 16. Available at: https://doi.org/10.1186/s13756-019-0671-7.

Carrique-Mas, J.J. *et al.* (2023) 'Restrictions on antimicrobial use in aquaculture and livestock, Viet Nam', *Bulletin of the World Health Organization*, 101(3), pp. 223–225. Available at: https://doi.org/10.2471/BLT.22.289187.

Cayrou, C. *et al.* (2021) 'Phase variation during host colonization and invasion by *Campylobacter jejuni* and other *Campylobacter* species', *Frontiers in Microbiology*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.705139.

Center for Food Safety and Applied Nutrition (2000) 'BAM Chapter 7: *Campylobacter*'. FDA. Available at: https://www.fda.gov/food/laboratory-methods-food/bam-chapter-7-campylobacter (Accessed: 9 May 2024).

Chaloner, G. *et al.* (2014) 'Dynamics of dual infection with *Campylobacter jejuni* strains in chickens reveals distinct strain-to-strain variation in infection ecology', *Applied and Environmental Microbiology*, 80(20), pp. 6366–6372. Available at: https://doi.org/10.1128/AEM.01901-14.

Charalampous, T. *et al.* (2019) 'Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection', *Nature Biotechnology*, 37(7), pp. 783–792. Available at: https://doi.org/10.1038/s41587-019-0156-5.

Charalampous, T. *et al.* (2024) 'Routine metagenomics service for ICU patients with respiratory infection', *American Journal of Respiratory and Critical Care Medicine*, 209(2), pp. 164–174. Available at: https://doi.org/10.1164/rccm.202305-0901OC.

Chaudhuri, R.R. and Henderson, I.R. (2012) 'The evolution of the *Escherichia coli* phylogeny', *Infection, Genetics and Evolution*, 12(2), pp. 214–226. Available at: https://doi.org/10.1016/j.meegid.2012.01.005.

Chaveerach, P. *et al.* (2002) '*in vitro* study on the effect of organic acids on *Campylobacter jejuni/coli* populations in mixtures of water and feed', *Poultry Science*, 81(5), pp. 621–628. Available at: https://doi.org/10.1093/ps/81.5.621.

Chen, S. *et al.* (2018) 'fastp: an ultra-fast all-in-one FASTQ preprocessor', *Bioinformatics*, 34(17), pp. i884–i890. Available at: https://doi.org/10.1093/bioinformatics/bty560.

Chomsky, E. and Lifshitz, A. (2024) ‘tgutil: simple utility functions for Tanay lab code’. Available at: https://github.com/tanaylab/tgutil.

Chon, J.-W. *et al.* (2014) ‘Supplementation of Bolton broth with triclosan improves detection of *Campylobacter jejuni* and *Campylobacter coli* in chicken carcass rinse’, *International Journal of Food Microbiology*, 181, pp. 37–39. Available at: https://doi.org/10.1016/j.ijfoodmicro.2014.04.006.

Chon, J.-W. *et al.* (2018) ‘Improvement of Bolton broth by supplementation with tazobactam for the isolation of *Campylobacter* from chicken rinses’, *Poultry Science*, 97(1), pp. 289–293. Available at: https://doi.org/10.3382/ps/pex169.

Chu, J. *et al.* (2014) ‘BioBloom tools: fast, accurate and memory-efficient host species sequence screening using bloom filters’, *Bioinformatics*, 30(23), pp. 3402–3404. Available at: https://doi.org/10.1093/bioinformatics/btu558.

Clausen, P.T.L.C., Aarestrup, F.M. and Lund, O. (2018) ‘Rapid and precise alignment of raw reads against redundant databases with KMA’, *BMC Bioinformatics*, 19(1), p. 307. Available at: https://doi.org/10.1186/s12859-018-2336-6.

Cody, A.J. *et al.* (2013) ‘Real-time genomic epidemiological evaluation of human *Campylobacter* isolates by use of whole-genome multilocus sequence typing’, *Journal of Clinical Microbiology*, 51(8), pp. 2526–2534. Available at: https://doi.org/10.1128/JCM.00066-13.

Cody, A.J. *et al.* (2017) ‘Core genome multilocus sequence typing scheme for stable, comparative analyses of *Campylobacter jejuni* and *C. coli* human disease isolates’, *Journal of Clinical Microbiology*, 55(7), pp. 2086–2097. Available at: https://doi.org/10.1128/JCM.00080-17.

Cody, A.J. *et al.* (2019) ‘A systematic review of source attribution of human campylobacteriosis using multilocus sequence typing’, *Eurosurveillance*, 24(43). Available at: https://doi.org/10.2807/1560-7917.ES.2019.24.43.1800696.

Colles, F.M. *et al.* (2008) ‘Comparison of *Campylobacter* populations in wild geese with those in starlings and free-range poultry on the same farm’, *Applied and Environmental Microbiology*, 74(11), pp. 3583–3590. Available at: https://doi.org/10.1128/AEM.02491-07.

Colles, F.M. *et al.* (2010) ‘Comparison of *Campylobacter* populations isolated from a free-range broiler flock before and after slaughter’, *International Journal of Food Microbiology*, 137(2), pp. 259–264. Available at: https://doi.org/10.1016/j.ijfoodmicro.2009.12.021.

Colles, F.M. and Maiden, M.C.J. (2012) ‘*Campylobacter* sequence typing databases: applications and future prospects’, *Microbiology*, 158(11), pp. 2695–2709. Available at: https://doi.org/10.1099/mic.0.062000-0.

Connerton, P.L. *et al.* (2018) ‘The effect of the timing of exposure to *Campylobacter jejuni* on the gut microbiome and inflammatory responses of broiler chickens’, *Microbiome*, 6, p. 88. Available at: https://doi.org/10.1186/s40168-018-0477-5.

Connor, T.R. *et al.* (2016) ‘CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community’, *Microbial Genomics*, 2(9), p. e000086. Available at: https://doi.org/10.1099/mgen.0.000086.

Constantinides, B., Hunt, M. and Crook, D.W. (2023) ‘Hostile: accurate decontamination of microbial host sequences’, *Bioinformatics*, 39(12), p. btad728. Available at: https://doi.org/10.1093/bioinformatics/btad728.

Corry, J. *et al.* (2017) *Reducing* Campylobacter *cross-contamination during poultry processing*. FS990010 (M01039). University of Bristol. Available at: https://www.food.gov.uk/research/research-projects/reducing-campylobacter-cross-contamination-during-poultry-processing (Accessed: 6 April 2021).

Corry, J.E.L. *et al.* (1995) 'Culture media for the isolation of campylobacters', *International Journal of Food Microbiology*, 26(1), pp. 43–76. Available at: https://doi.org/10.1016/0168-1605(95)00044-K.

Costa, D. and Iraola, G. (2019) 'Pathogenomics of emerging *Campylobacter* species', *Clinical Microbiology Reviews*, 32(4), pp. e00072-18. Available at: https://doi.org/10.1128/CMR.00072-18.

Couto, N. *et al.* (2018) 'Critical steps in clinical shotgun metagenomics for the concomitant detection and typing of microbial pathogens', *Scientific Reports*, 8(1), p. 13767. Available at: https://doi.org/10.1038/s41598-018-31873-w.

Croucher, N.J. *et al.* (2015) 'Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins', *Nucleic Acids Research*, 43(3), p. e15. Available at: https://doi.org/10.1093/nar/gku1196.

Crowther, J.R. (2008) *ELISA: Theory and practice*. Springer Science & Business Media. Available at: https://doi.org/10.1385/0-89603-279-5:1.

Danecek, P. *et al.* (2021) 'Twelve years of SAMtools and BCFtools', *GigaScience*, 10(2), p. giab008. Available at: https://doi.org/10.1093/gigascience/giab008.

De Cesare, A. *et al.* (2019) 'Effect of a low protein diet on chicken ceca microbiome and productive performances', *Poultry Science*, 98(9), pp. 3963–3976. Available at: https://doi.org/10.3382/ps/pez132.

Dearlove, B.L. *et al.* (2016) 'Rapid host switching in generalist *Campylobacter* strains erodes the signal for tracing human infections', *The ISME Journal*, 10(3), pp. 721–729. Available at: https://doi.org/10.1038/ismej.2015.149.

Deng, J. *et al.* (2015) 'A comparison of Luminex xTAG® Gastrointestinal Pathogen Panel (xTAG GPP) and routine tests for the detection of enteropathogens circulating in Southern China', *Diagnostic Microbiology and Infectious Disease*, 83(3), pp. 325–330. Available at: https://doi.org/10.1016/j.diagmicrobio.2015.07.024.

Di Bella, S. *et al.* (2016) '*Clostridium difficile* toxins A and B: insights into pathogenic properties and extraintestinal effects', *Toxins*, 8(5), p. 134. Available at: https://doi.org/10.3390/toxins8050134.

Dingle, K.E. *et al.* (2001) 'Multilocus sequence typing system for *Campylobacter jejuni*', *Journal of Clinical Microbiology*, 39(1), pp. 14–23. Available at: https://doi.org/10.1128/JCM.39.1.14-23.2001.

Dingle, K.E. *et al.* (2005) 'Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*', *Journal of Clinical Microbiology*, 43(1), pp. 340–347. Available at: https://doi.org/10.1128/JCM.43.1.340-347.2005.

Djeghout, B. *et al.* (2022) 'Comparative genomics of *Campylobacter jejuni* from clinical campylobacteriosis stool specimens', *Gut Pathogens*, 14(1), p. 45. Available at: https://doi.org/10.1186/s13099-022-00520-1.

Doorduyn, Y. *et al.* (2010) 'Risk factors for indigenous *Campylobacter jejuni* and *Campylobacter coli* infections in The Netherlands: a case-control study', *Epidemiology & Infection*, 138(10), pp. 1391–1404. Available at: https://doi.org/10.1017/S095026881000052X.

Doster, E. *et al.* (2019) 'A cautionary report for pathogen identification using shotgun metagenomics; a comparison to aerobic culture and polymerase chain reaction for *Salmonella enterica* identification', *Frontiers in Microbiology*, 10. Available at: https://doi.org/10.3389/fmicb.2019.02499.

Dowle, M. and Srinivasan, A. (2023) 'data.table: extension of `data.frame`_. R package version 1.14.8'. Available at: https://CRAN.R-project.org/package=data.table.

Duffy, G., Cagney, C. and Lynch, Ó. (2007) *Recovery and identification of emerging Campylobacteraceae from food*. 85. Dublin: Ashtown Food Research Centre. Available at: https://core.ac.uk/download/pdf/45656643.pdf (Accessed: 10 May 2024).

Duong, V.T. *et al.* (2016) 'Evaluation of Luminex xTAG Gastrointestinal Pathogen Panel assay for detection of multiple diarrheal pathogens in fecal samples in Vietnam', *Journal of Clinical Microbiology*, 54(4), pp. 1094–1100. Available at: https://doi.org/10.1128/JCM.03321-15.

Duong, V.T. *et al.* (2020) 'Genomic serotyping, clinical manifestations, and antimicrobial resistance of nontyphoidal *Salmonella* gastroenteritis in hospitalized children in Ho Chi Minh City, Vietnam', *Journal of Clinical Microbiology*, 58(12), pp. e01465-20. Available at: https://doi.org/10.1128/JCM.01465-20.

Durazzi, F. *et al.* (2021) 'Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota', *Scientific Reports*, 11(1), p. 3030. Available at: https://doi.org/10.1038/s41598-021-82726-y.

Dymond, J.S. (2013) 'Explanatory chapter: quantitative PCR', in J. Lorsch (ed.) *Methods in Enzymology*. Academic Press (Laboratory Methods in Enzymology: DNA), pp. 279–289. Available at: https://doi.org/10.1016/B978-0-12-418687-3.00023-9.

Dziegiel, A.H. *et al.* (2024) 'High *Campylobacter* diversity in retail chicken: epidemiologically important strains may be missed with current sampling methods', *Epidemiology & Infection*, 152, p. e101. Available at: https://doi.org/10.1017/S0950268824000906.

Edgar, R.C. (2004) 'MUSCLE: multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*, 32(5), pp. 1792–1797. Available at: https://doi.org/10.1093/nar/gkh340.

EFSA Panel on Biological Hazards (BIOHAZ) (2010) 'Scientific opinion on quantification of the risk posed by broiler meat to human campylobacteriosis in the EU', *EFSA Journal*, 8(1), p. 1437. Available at: https://doi.org/10.2903/j.efsa.2010.1437.

Ehuwa, O., Jaiswal, A.K. and Jaiswal, S. (2021) '*Salmonella*, food safety and food handling practices', *Foods*, 10(5), p. 907. Available at: https://doi.org/10.3390/foods10050907.

El Baaboua, A. *et al.* (2021) 'Prevalence and antimicrobial profiling of *Campylobacter* spp. isolated from meats, animal, and human feces in Northern of Morocco.', *International Journal of Food Microbiology*, 349, p. 109202. Available at: https://doi.org/10.1016/j.ijfoodmicro.2021.109202.

El-Adawy, H. *et al.* (2023) 'Genomic insight into *Campylobacter jejuni* isolated from commercial turkey flocks in Germany using whole-genome sequencing analysis', *Frontiers in Veterinary Science*, 10, p. 1092179. Available at: https://doi.org/10.3389/fvets.2023.1092179.

Elnifro, E.M. *et al.* (2000) 'Multiplex PCR: optimization and application in diagnostic virology', *Clinical Microbiology Reviews*, 13(4), pp. 559–570. Available at: https://doi.org/10.1128/cmr.13.4.559-570.2000.

Emond-Rheault, J.-G. *et al.* (2020) 'The *Salmonella enterica* plasmidome as a reservoir of antibiotic resistance', *Microorganisms*, 8(7), p. 1016. Available at: https://doi.org/10.3390/microorganisms8071016.

Epping, L. *et al.* (2021) 'Genome-wide insights into population structure and host specificity of *Campylobacter jejuni*', *Scientific Reports*, 11(1), p. 10358. Available at: https://doi.org/10.1038/s41598-021-89683-6.

Epping, L., Antão, E.-M. and Semmler, T. (2021) 'Population biology and comparative genomics of *Campylobacter* species', *Current Topics in Microbiology and Immunology*, 431, pp. 59–78. Available at: https://doi.org/10.1007/978-3-030-65481-8_3.

Ernst, F. *et al.* (2022) 'mia: microbiome analysis'. Available at: https://github.com/microbiome/mia.

Espinosa, E. *et al.* (2024) 'Advancements in long-read genome sequencing technologies and algorithms', *Genomics*, 116(3), p. 110842. Available at: https://doi.org/10.1016/j.ygeno.2024.110842.

European Centre for Disease Prevention and Control (2024) *Campylobacteriosis: annual epidemiological report for 2022*. Stockholm: ECDC. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/CAMP_AER_2022_final.pdf (Accessed: 25 April 2024).

European Commission (2021) *Guidance document on multiplex real-time PCR methods.* LU: Publications Office. Available at: https://data.europa.eu/doi/10.2760/243914 (Accessed: 17 May 2024).

European Food Safety Authority (2007) 'Introduction of a Qualified Presumption of Safety (QPS) approach for assessment of selected microorganisms referred to EFSA - Opinion of the Scientific Committee', *EFSA Journal*, 5(12), p. 587. Available at: https://doi.org/10.2903/j.efsa.2007.587.

European Food Safety Authority (2010) 'Analysis of the baseline survey on the prevalence of *Campylobacter* in broiler batches and of *Campylobacter* and *Salmonella* on broiler carcasses in the EU, 2008 - Part A: *Campylobacter* and *Salmonella* prevalence estimates', *EFSA Journal*, 8(3), p. 1503. Available at: https://doi.org/10.2903/j.efsa.2010.1503.

European Food Safety Authority and European Centre for Disease Prevention and Control (2019) *The European Union one health 2018 zoonoses report*. Available at: https://efsa.onlinelibrary.wiley.com/doi/10.2903/j.efsa.2019.5926 (Accessed: 6 April 2021).

Fakruddin, Md., Mannan, K.S.B. and Andrews, S. (2013) 'Viable but nonculturable bacteria: food safety and public health perspective', *ISRN Microbiology*, 2013, pp. 1–6. Available at: https://doi.org/10.1155/2013/703813.

Faverjon, C., Cameron, A. and De Nardi, M. (2022) *Modelling framework to quantify the risk of AMR exposure via food products - example of chicken and lettuce*. Food Standards Agency. Available at: https://doi.org/10.46756/sci.fsa.qum110.

FC, M. and Davis, T. (2022) 'ggpattern: "ggplot2" pattern geoms_. R package version 1.0.1'. Available at: https://CRAN.R-project.org/package=ggpattern.

Fernandes, A.D. *et al.* (2013) 'ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-seq', *PLOS ONE*, 8(7), p. e67019. Available at: https://doi.org/10.1371/journal.pone.0067019.

Fernandes, A.D. *et al.* (2014) 'Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis', *Microbiome*, 2(1), p. 15. Available at: https://doi.org/10.1186/2049-2618-2-15.

Figura, N. *et al.* (1993) 'Two cases of *Campylobacter mucosalis* enteritis in children', *Journal of Clinical Microbiology*, 31(3), pp. 727–728. Available at: https://doi.org/10.1128/jcm.31.3.727-728.1993.

Florensa, A.F. *et al.* (2022) 'ResFinder – an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes', *Microbial Genomics*, 8(1), p. 000748. Available at: https://doi.org/10.1099/mgen.0.000748.

Flynn, O.M.J., Blair, I.S. and Mcdowell, D.A. (1994) 'Prevalence of *Campylobacter* species on fresh retail chicken wings in Northern Ireland', *Journal of Food Protection*, 57(4), pp. 334–336. Available at: https://doi.org/10.4315/0362-028X-57.4.334.

Food Standards Agency (2016) *Microbiological survey of* Campylobacter *contamination in fresh whole chilled chickens: protocol*. Available at: https://www.food.gov.uk/sites/default/files/media/document/retail-survey-protocol-year-4_0.pdf (Accessed: 13 June 2023).

Fornefett, J. *et al.* (2021) 'Bacterial gastroenteritis caused by the putative zoonotic pathogen *Campylobacter lanienae*: first reported case in Germany', *Access Microbiology*, 3(3), p. 000199. Available at: https://doi.org/10.1099/acmi.0.000199.

Fourgeaud, J. *et al.* (2024) 'Performance of clinical metagenomics in France: a prospective observational study', *The Lancet Microbe*, 5(1), pp. e52–e61. Available at: https://doi.org/10.1016/S2666-5247(23)00244-6.

Fuks, G. *et al.* (2018) 'Combining 16S rRNA gene variable regions enables high-resolution microbial community profiling', *Microbiome*, 6(1), p. 17. Available at: https://doi.org/10.1186/s40168-017-0396-x.

Gaasbeek, E.J. *et al.* (2009) 'Functional characterization of excision repair and RecA-dependent recombinational DNA repair in *Campylobacter jejuni*', *Journal of Bacteriology*, 191(12), pp. 3785–3793. Available at: https://doi.org/10.1128/JB.01817-08.

Gal-Mor, O. (2018) 'Persistent infection and long-term carriage of typhoidal and nontyphoidal Salmonellae', *Clinical Microbiology Reviews*, 32(1), pp. e00088-18. Available at: https://doi.org/10.1128/CMR.00088-18.

Gal-Mor, O., Boyle, E.C. and Grassl, G.A. (2014) 'Same species, different diseases: how and why typhoidal and non-typhoidal *Salmonella enterica* serovars differ', *Frontiers in Microbiology*, 5, p. 391. Available at: https://doi.org/10.3389/fmicb.2014.00391.

Gao, B. *et al.* (2017) 'Metabolic and fitness determinants for *in vitro* growth and intestinal colonization of the bacterial pathogen *Campylobacter jejuni*', *PLOS Biology*, 15(5), p. e2001390. Available at: https://doi.org/10.1371/journal.pbio.2001390.

Garcia-Fernandez, A. *et al.* (2024) 'Antibiotic resistance, plasmids, and virulence-associated markers in human strains of *Campylobacter jejuni* and *Campylobacter coli* isolated in Italy', *Frontiers in Microbiology*, 14. Available at: https://doi.org/10.3389/fmicb.2023.1293666.

García-Sánchez, L. *et al.* (2017) '*Campylobacter jejuni* survival in a poultry processing plant environment', *Food Microbiology*, 65, pp. 185–192. Available at: https://doi.org/10.1016/j.fm.2017.02.009.

Garrity, G., Bell, J. and Lilburn, T. (2005) 'Campylobacterales ord. nov', in *Bergey's Manual of Systematic Bacteriology*. 2nd edn. New York: Springer, p. 1145.

Gee, B. *et al.* (2002) 'Effect of incubation temperature on the isolation of thermophilic species of *Campylobacter* from faeces', *Communicable Disease and Public Health*, 5(4), pp. 282–284.

Ghielmetti, G. *et al.* (2023) 'Whole-genome-based characterization of *Campylobacter jejuni* from human patients with gastroenteritis collected over an 18 year period reveals increasing prevalence of antimicrobial resistance', *Microbial Genomics*, 9(2), p. mgen000941. Available at: https://doi.org/10.1099/mgen.0.000941.

Gilligan, P.H. (2013) 'Identification of pathogens by classical clinical tests', in E. Rosenberg et al. (eds) *The prokaryotes: human microbiology*. Berlin, Heidelberg: Springer, pp. 57–89. Available at: https://doi.org/10.1007/978-3-642-30144-5_90.

Gilroy, R. *et al.* (2024) 'The immunobiology of persistent intestinal infection by *Campylobacter jejuni* in the chicken'. bioRxiv, p. 2024.08.23.609314. Available at: https://doi.org/10.1101/2024.08.23.609314.

Golz, J.C. *et al.* (2020) 'Whole genome sequencing reveals extended natural transformation in *Campylobacter* impacting diagnostics and the pathogens adaptive potential', *Scientific Reports*, 10(1), p. 3686. Available at: https://doi.org/10.1038/s41598-020-60320-y.

Golz, J.C. and Stingl, K. (2021) 'Natural competence and horizontal gene transfer in *Campylobacter*', in S. Backert (ed.) *Fighting* Campylobacter *Infections: Towards a One Health Approach*. Cham: Springer International Publishing, pp. 265–292. Available at: https://doi.org/10.1007/978-3-030-65481-8_10.

González-Fandos, E., Maya, N. and Pérez-Arnedo, I. (2015) 'Effect of propionic acid on *Campylobacter jejuni* attached to chicken skin during refrigerated storage', *International Microbiology: The Official Journal of the Spanish Society for Microbiology*, 18(3), pp. 171–175. Available at: https://doi.org/10.2436/20.1501.01.247.

Gouy, M. *et al.* (2021) 'Seaview version 5: a multiplatform software for multiple sequence alignment, molecular phylogenetic analyses, and tree reconciliation', in K. Katoh (ed.) *Multiple Sequence Alignment: Methods and Protocols*. New York, NY: Springer US (Methods in Molecular Biology), pp. 241–260. Available at: https://doi.org/10.1007/978-1-0716-1036-7_15.

Govender, K.N. and Eyre, D.W. (2022) 'Benchmarking taxonomic classifiers with Illumina and Nanopore sequence data for clinical metagenomic diagnostic applications', *Microbial Genomics*, 8(10), p. mgen000886. Available at: https://doi.org/10.1099/mgen.0.000886.

van der Graaf-van Bloois, L., Wagenaar, J.A. and Zomer, A.L. (2021) 'RFPlasmid: predicting plasmid sequences from short-read assembly data using machine learning', *Microbial Genomics*, 7(11). Available at: https://doi.org/10.1099/mgen.0.000683.

Grant, K.A. and Park, S.F. (1995) 'Molecular characterization of katA from *Campylobacter jejuni* and generation of a catalase-deficient mutant of *Campylobacter coli* by interspecific allelic exchange', *Microbiology*, 141(6), pp. 1369–1376. Available at: https://doi.org/10.1099/13500872-141-6-1369.

Grützke, J. *et al.* (2019) 'Fishing in the soup – pathogen detection in food safety using metabarcoding and metagenomic sequencing', *Frontiers in Microbiology*, 10. Available at: https://doi.org/10.3389/fmicb.2019.01805.

Guo, B. *et al.* (2008) 'CmeR functions as a pleiotropic regulator and is required for optimal colonization of *Campylobacter jejuni in vivo*', *Journal of Bacteriology*, 190(6), pp. 1879–1890. Available at: https://doi.org/10.1128/JB.01796-07.

Gurevich, A. *et al.* (2013) 'QUAST: quality assessment tool for genome assemblies', *Bioinformatics*, 29(8), pp. 1072–1075. Available at: https://doi.org/10.1093/bioinformatics/btt086.

Guyard-Nicodème, M. *et al.* (2015) 'Prevalence and characterization of *Campylobacter jejuni* from chicken meat sold in French retail outlets', *International Journal of Food Microbiology*, 203, pp. 8–14. Available at: https://doi.org/10.1016/j.ijfoodmicro.2015.02.013.

Habib, I. *et al.* (2023) 'Genomic characterization of molecular markers associated with antimicrobial resistance and virulence of the prevalent *Campylobacter coli* isolated from retail chicken meat in the United Arab Emirates', *Current Research in Food Science*, 6, p. 100434. Available at: https://doi.org/10.1016/j.crfs.2023.100434.

Hardy Diagnostics (2020) *Hunt broth*, *Hardy Diagnostics*. Available at: https://hardydiagnostics.com/media/assets/product/documents/HuntBroth.pdf (Accessed: 13 May 2024).

Harrington, C.S., Thomson-Carter, F.M. and Carter, P.E. (1997) 'Evidence for recombination in the flagellin locus of *Campylobacter jejuni*: implications for the flagellin gene typing scheme', *Journal of Clinical Microbiology*, 35(9), pp. 2386–2392. Available at: https://doi.org/10.1128/jcm.35.9.2386-2392.1997.

Harrison, L.M. *et al.* (2022) 'Current methodologies and future direction of *Campylobacter* isolation and detection from food matrices, clinical samples, and the agricultural environment', *Journal of Microbiological Methods*, 201, p. 106562. Available at: https://doi.org/10.1016/j.mimet.2022.106562.

Hasman, H. *et al.* (2014) 'Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples', *Journal of Clinical Microbiology*, 52(1), pp. 139–146. Available at: https://doi.org/10.1128/JCM.02452-13.

Haverkamp, T.H.A. *et al.* (2021) 'Detection of *Campylobacter* in air samples from poultry houses using shot-gun metagenomics – a pilot study'. bioRxiv, p. 2021.05.17.444449. Available at: https://doi.org/10.1101/2021.05.17.444449.

Hawinkel, S. *et al.* (2019) 'A broken promise: microbiome differential abundance methods do not control the false discovery rate', *Briefings in Bioinformatics*, 20(1), pp. 210–221. Available at: https://doi.org/10.1093/bib/bbx104.

Health Canada (2014) 'MFLP-46: Isolation of thermophilic *Campylobacter* from food'. Ottawa. Available at: https://www.canada.ca/en/health-canada/services/food-nutrition/research-programs-analytical-methods/analytical-methods/compendium-methods/laboratory-procedures-microbiological-analysis-foods-compendium-analytical-methods.html (Accessed: 10 May 2024).

Heather, J.M. and Chain, B. (2016) 'The sequence of sequencers: the history of sequencing DNA', *Genomics*, 107(1), pp. 1–8. Available at: https://doi.org/10.1016/j.ygeno.2015.11.003.

Hendrixson, D.R. (2006) 'A phase-variable mechanism controlling the *Campylobacter jejuni* FlgR response regulator influences commensalism', *Molecular Microbiology*, 61(6), pp. 1646–1659. Available at: https://doi.org/10.1111/j.1365-2958.2006.05336.x.

Hendrixson, D.R. and DiRita, V.J. (2004) 'Identification of *Campylobacter jejuni* genes involved in commensal colonization of the chick gastrointestinal tract', *Molecular Microbiology*, 52(2), pp. 471–484. Available at: https://doi.org/10.1111/j.1365-2958.2004.03988.x.

Hertogs, K. *et al.* (2021) 'Contamination sources and transmission routes for *Campylobacter* on (mixed) broiler farms in Belgium, and comparison of the gut microbiota of flocks colonized and uncolonized with *Campylobacter*', *Pathogens*, 10(1), p. 66. Available at: https://doi.org/10.3390/pathogens10010066.

Hetman, B.M. *et al.* (2020) '"These aren't the strains you're looking for": recovery bias of common *Campylobacter jejuni* subtypes in mixed cultures', *Frontiers in Microbiology*, 11, p. 541. Available at: https://doi.org/10.3389/fmicb.2020.00541.

Hoang, D.T. *et al.* (2018) 'UFBoot2: improving the ultrafast bootstrap approximation', *Molecular Biology and Evolution*, 35(2), pp. 518–522. Available at: https://doi.org/10.1093/molbev/msx281.

Hodges, L.M. *et al.* (2021) 'Systematic evaluation of whole-genome sequencing based prediction of antimicrobial resistance in *Campylobacter jejuni* and *C. coli*', *Frontiers in Microbiology*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.776967.

Hofreuter, D. (2014) 'Defining the metabolic requirements for the growth and colonization capacity of *Campylobacter jejuni*', *Frontiers in Cellular and Infection Microbiology*, 4, p. 137. Available at: https://doi.org/10.3389/fcimb.2014.00137.

Huang, P. *et al.* (2018) 'The chicken gut metagenome and the modulatory effects of plant-derived benzylisoquinoline alkaloids', *Microbiome*, 6(1), p. 211. Available at: https://doi.org/10.1186/s40168-018-0590-5.

Hull, D.M. *et al.* (2021) 'Antimicrobial resistance and interspecies gene transfer in *Campylobacter coli* and *Campylobacter jejuni* isolated from food animals, poultry processing, and retail meat in North Carolina, 2018-2019.', *PLOS ONE*, 16(2), p. e0246571. Available at: https://doi.org/10.1371/journal.pone.0246571.

Hull, D.M. *et al.* (2023) 'Detection of resistance and virulence plasmids in *Campylobacter coli* and *Campylobacter jejuni* isolated from North Carolina food animal production, 2018–2019', *Food Microbiology*, 116, p. 104348. Available at: https://doi.org/10.1016/j.fm.2023.104348.

d'Humières, C. *et al.* (2021) 'The potential role of clinical metagenomics in infectious diseases: therapeutic perspectives', *Drugs*, 81(13), pp. 1453–1466. Available at: https://doi.org/10.1007/s40265-021-01572-4.

Humphrey, S. *et al.* (2014) '*Campylobacter jejuni* is not merely a commensal in commercial broiler chickens and affects bird welfare', *mBio*, 5(4), p. 10.1128/mbio.01364-14. Available at: https://doi.org/10.1128/mbio.01364-14.

Hunt, M. *et al.* (2017) 'ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads', *Microbial Genomics,* 3(10), p. e000131. Available at: https://doi.org/10.1099/mgen.0.000131.

Hussein, N.H. *et al.* (2021) 'Mobilized colistin resistance (mcr) genes from 1 to 10: a comprehensive review', *Molecular Biology Reports*, 48(3), pp. 2897–2907. Available at: https://doi.org/10.1007/s11033-021-06307-y.

Ica, T. *et al.* (2012) 'Characterization of mono- and mixed-culture *Campylobacter jejuni* biofilms', *Applied and Environmental Microbiology*, 78(4), pp. 1033–1038. Available at: https://doi.org/10.1128/AEM.07364-11.

Illumina (2024) *Cost of NGS | Comparisons and budget guidance*, *Illumina*. Available at: https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-cost.html (Accessed: 12 February 2024).

Indikova, I., Humphrey, T.J. and Hilbert, F. (2015) 'Survival with a helping hand: *Campylobacter* and microbiota', *Frontiers in Microbiology*, 6, p. 1266. Available at: https://doi.org/10.3389/fmicb.2015.01266.

Inglis, G.D. *et al.* (2021) 'Analysis of *Campylobacter jejuni* subtype distribution in the chicken broiler production continuum: a longitudinal examination to identify primary contamination points', *Applied and Environmental Microbiology*, 87(3), pp. e02001-20. Available at: https://doi.org/10.1128/AEM.02001-20.

International Organization for Standardization (2017a) 'EN ISO 10272-1:2017: Microbiology of the food chain — horizontal method for detection and enumeration of *Campylobacter* spp. — Part 1: Detection method'. Geneva.

International Organization for Standardization (2017b) 'EN ISO 10272-2:2017: Microbiology of the food chain — horizontal method for detection and enumeration of *Campylobacter* spp. — Part 2: Colony-count technique'. Geneva.

Iovine, N.M. (2013) 'Resistance mechanisms in *Campylobacter jejuni*', *Virulence*, 4(3), pp. 230–240. Available at: https://doi.org/10.4161/viru.23753.

Jacob, P., Mdegela, R.H. and Nonga, H.E. (2011) 'Comparison of Cape Town and Skirrow's *Campylobacter* isolation protocols in humans and broilers in Morogoro, Tanzania', *Tropical Animal Health and Production*, 43(5), pp. 1007–1013. Available at: https://doi.org/10.1007/s11250-011-9799-z.

Jacoby, G.A. (2005) 'Mechanisms of resistance to quinolones', *Clinical Infectious Diseases*, 41(Supplement_2), pp. S120–S126. Available at: https://doi.org/10.1086/428052.

Jasson, V. *et al.* (2009) 'Characterization of *Escherichia coli* from raw poultry in Belgium and impact on the detection of *Campylobacter jejuni* using Bolton broth', *International Journal of Food Microbiology*, 135(3), pp. 248–253. Available at: https://doi.org/10.1016/j.ijfoodmicro.2009.09.007.

Jesser, K.J. *et al.* (2023) 'Why are so many enteric pathogen infections asymptomatic? Pathogen and gut microbiome characteristics associated with diarrhea symptoms and carriage of diarrheagenic *E. coli* in northern Ecuador', *Gut Microbes*, 15(2), 2281010. Available at: https://www.tandfonline.com/doi/full/10.1080/19490976.2023.2281010.

Jo, Y. *et al.* (2017) 'Enrichment broth for the detection of *Campylobacter jejuni* and *Campylobacter coli* in fresh produce and poultry', *Journal of Food Protection*, 80(11), pp. 1842–1850. Available at: https://doi.org/10.4315/0362-028X.JFP-16-450.

Joensen, K.G. *et al.* (2020) 'Whole-genome sequencing to detect numerous *Campylobacter jejuni* outbreaks and match patient isolates to sources, Denmark, 2015–2017', *Emerging infectious diseases*, 26(3), pp. 523–532. Available at: https://doi.org/10.3201/eid2603.190947.

Joensen, K.G. *et al.* (2021) 'Whole genome sequencing data used for surveillance of *Campylobacter* infections: detection of a large continuous outbreak, Denmark, 2019', *Eurosurveillance*, 26(22), p. 2001396. Available at: https://doi.org/10.2807/1560-7917.ES.2021.26.22.2001396.

Johnson, J.S. *et al.* (2019) 'Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis', *Nature Communications*, 10(1), p. 5029. Available at: https://doi.org/10.1038/s41467-019-13036-1.

Jolley, K.A., Bray, J.E. and Maiden, M.C.J. (2018) 'Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications', *Wellcome Open Research*, 3, p. 124. Available at: https://doi.org/10.12688/wellcomeopenres.14826.1.

Jorgensen, F. *et al.* (2019) *A microbiological survey of* Campylobacter *contamination in fresh whole UK-produced chilled chickens at retail sale*. FSA Project FS102121, p. 37. Available at: https://www.food.gov.uk/sites/default/files/media/document/campylobacter-contamination-uk-chickens-year-4-report.pdf.

Jorgensen, F. *et al.* (2021) *A survey of the levels of* Campylobacter *spp. contamination and prevalence of selected antimicrobial resistance determinants in fresh whole UK-produced chilled chickens at retail sale (non-major retailers)*. Year 5 and 6 (2018 to 2020) Final Report. Food Standards Agency. Available at: https://doi.org/10.46756/sci.fsa.xxz973.

Jorgensen, F. *et al.* (2022) *Levels and trends of antimicrobial resistance in* Campylobacter *spp. from chicken in the UK*. Food Standards Agency. Available at: https://doi.org/10.46756/sci.fsa.dud728.

Josefsen, M.H. *et al.* (2004) 'Towards an international standard for PCR-based detection of foodborne thermotolerant campylobacters: interaction of enrichment media and pre-PCR treatment on carcass rinse samples', *Journal of Microbiological Methods*, 58(1), pp. 39–48. Available at: https://doi.org/10.1016/j.mimet.2004.03.001.

Josefsen, M.H. *et al.* (2015) 'Microbial food safety: potential of DNA extraction methods for use in diagnostic metagenomics', *Journal of Microbiological Methods*, 114, pp. 30–34. Available at: https://doi.org/10.1016/j.mimet.2015.04.016.

Jribi, H. *et al.* (2017) 'Isolation and identification of *Campylobacter* spp. from poultry and poultry by-products in Tunisia by conventional culture method and multiplex real-time PCR', *Journal of Food Protection*, 80(10), pp. 1623–1627. Available at: https://doi.org/10.4315/0362-028X.JFP-16-321.

Juraschek, K. *et al.* (2021) 'Outcome of different sequencing and assembly approaches on the detection of plasmids and localization of antimicrobial resistance genes in commensal *Escherichia coli*', *Microorganisms*, 9(3), p. 598. Available at: https://doi.org/10.3390/microorganisms9030598.

Jurinović, L. *et al.* (2023) 'Genetic diversity and resistome analysis of *Campylobacter lari* Isolated from gulls in Croatia', *Antibiotics*, 12(8), p. 1310. Available at: https://doi.org/10.3390/antibiotics12081310.

Kaakoush, N.O. *et al.* (2007) 'Oxygen requirement and tolerance of *Campylobacter jejuni*', *Research in Microbiology*, 158(8), pp. 644–650. Available at: https://doi.org/10.1016/j.resmic.2007.07.009.

Kaakoush, N.O. *et al.* (2014) 'The interplay between *Campylobacter* and *Helicobacter* species and other gastrointestinal microbiota of commercial broiler chickens', *Gut Pathogens*, 6, p. 18. Available at: https://doi.org/10.1186/1757-4749-6-18.

Kaakoush, N.O. *et al.* (2015) 'Global epidemiology of *Campylobacter* infection', *Clinical Microbiology Reviews*, 28(3), pp. 687–720. Available at: https://doi.org/10.1128/CMR.00006-15.

Kang, D.D. *et al.* (2019) 'MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies', *PeerJ*, 7, p. e7359. Available at: https://doi.org/10.7717/peerj.7359.

Kantar (2024) *Grocery market share: Great Britain (accurate as of 14 April 2024)*, *Kantar*. Available at: https://www.kantarworldpanel.com/global/grocery-market-share/great-britain/snapshot/22.04.18/ (Accessed: 2 May 2024).

Karikari, A.B. *et al.* (2017) 'Antibiotic resistance of *Campylobacter* recovered from faeces and carcasses of healthy livestock', *BioMed Research International*, 2017, p. 4091856. Available at: https://doi.org/10.1155/2017/4091856.

Karst, S.M. *et al.* (2021) 'High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing', *Nature Methods*, 18(2), pp. 165–169. Available at: https://doi.org/10.1038/s41592-020-01041-y.

Kassambara, A. (2023) 'ggpubr: "ggplot2" based publication ready plots'. Available at: https://CRAN.R-project.org/package=ggpubr.

Katz, L.S. *et al.* (2019) 'Mashtree: a rapid comparison of whole genome sequence files', *Journal of open source software*, 4(44), p. 10.21105/joss.01762. Available at: https://doi.org/10.21105/joss.01762.

Kchouk, M., Gibrat, J.F. and Elloumi, M. (2017) 'Generations of sequencing technologies: from first to next generation', *Biology and Medicine*, 09(03). Available at: https://doi.org/10.4172/0974-8369.1000395.

Kempf, I. *et al.* (2017) '*Campylobacter coli* in organic and conventional pig production in France and Sweden: prevalence and antimicrobial resistance', *Frontiers in Microbiology*, 8, p. 955. Available at: https://doi.org/10.3389/fmicb.2017.00955.

Khachatryan, L. *et al.* (2020) 'Taxonomic classification and abundance estimation using 16S and WGS—A comparison using controlled reference samples', *Forensic Science International: Genetics*, 46, p. 102257. Available at: https://doi.org/10.1016/j.fsigen.2020.102257.

Kim, C., Pongpanich, M. and Porntaveetus, T. (2024) 'Unraveling metagenomics through long-read sequencing: a comprehensive review', *Journal of Translational Medicine*, 22(1), p. 111. Available at: https://doi.org/10.1186/s12967-024-04917-1.

Kim, D. *et al.* (2016) 'Centrifuge: rapid and sensitive classification of metagenomic sequences', *Genome Research*, 26(12), pp. 1721–1729. Available at: https://doi.org/10.1101/gr.210641.116.

Kim, J. *et al.* (2016) 'An improved culture method for selective isolation of *Campylobacter jejuni* from wastewater', *Frontiers in Microbiology*, 7, p. 1345. Available at: https://doi.org/10.3389/fmicb.2016.01345.

Kim, Jinshil *et al.* (2019) 'Microbiota analysis for the optimization of *Campylobacter* isolation from chicken carcasses using selective media', *Frontiers in Microbiology*, 10. Available at: https://www.frontiersin.org/article/10.3389/fmicb.2019.01381.

Kim, J.S. *et al.* (2010) 'Multiplexed magnetic microsphere immunoassays for detection of pathogens in foods', *Sensing and Instrumentation for Food Quality and Safety*, 4(2), pp. 73–81. Available at: https://doi.org/10.1007/s11694-010-9097-x.

Kim, Junhyung *et al.* (2019) 'Metagenomic analysis of isolation methods of a targeted microbe, *Campylobacter jejuni*, from chicken feces with high microbial contamination', *Microbiome*, 7(1), p. 67. Available at: https://doi.org/10.1186/s40168-019-0680-z.

Kim, S.-H. *et al.* (2021) 'Review on stress tolerance in *Campylobacter jejuni*', *Frontiers in Cellular and Infection Microbiology*, 10. Available at: https://doi.org/10.3389/fcimb.2020.596570.

Ko, K.K.K., Chng, K.R. and Nagarajan, N. (2022) 'Metagenomics-enabled microbial surveillance', *Nature Microbiology*, 7(4), pp. 486–496. Available at: https://doi.org/10.1038/s41564-022-01089-w.

Kolmogorov, M. *et al.* (2019) 'Assembly of long, error-prone reads using repeat graphs', *Nature Biotechnology*, 37(5), pp. 540–546. Available at: https://doi.org/10.1038/s41587-019-0072-8.

Koren, S. *et al.* (2017) 'Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation', *Genome Research*, 27(5), pp. 722–736. Available at: https://doi.org/10.1101/gr.215087.116.

Kovanen, S.M. *et al.* (2014) 'Multilocus sequence typing (MLST) and whole-genome MLST of *Campylobacter jejuni* isolates from human infections in three districts during a seasonal peak in Finland', *Journal of Clinical Microbiology*, 52(12), pp. 4147–4154. Available at: https://doi.org/10.1128/JCM.01959-14.

Kramer, J.M. *et al.* (2000) '*Campylobacter* contamination of raw meat and poultry at retail sale: identification of multiple types and comparison with isolates from human infection', *Journal of Food Protection*, 63(12), pp. 1654–1659. Available at: https://doi.org/10.4315/0362-028X-63.12.1654.

Krause, R. *et al.* (2002) 'Recurrent septicemia due to *Campylobacter fetus* and *Campylobacter lari* in an immunocompetent patient', *Infection*, 30(3), pp. 171–174. Available at: https://doi.org/10.1007/s15010-002-2115-0.

Kuana, S.L. *et al.* (2008) 'Occurrence and characterization of *Campylobacter* in the Brazilian production and processing of broilers', *Avian Diseases*, 52(4), pp. 680–684. Available at: https://doi.org/10.1637/8296-032608-Reg.1.

Kumar, H. *et al.* (2020) 'Whole metagenome sequencing of cecum microbiomes in Ethiopian indigenous chickens from two different altitudes reveals antibiotic resistance genes', *Genomics*, 112(2), pp. 1988–1999. Available at: https://doi.org/10.1016/j.ygeno.2019.11.011.

Kumar, P.S. *et al.* (2011) 'Target region selection is a critical determinant of community fingerprints generated by 16S pyrosequencing', *PLOS ONE*, 6(6), p. e20956. Available at: https://doi.org/10.1371/journal.pone.0020956.

Kwon, Y.-K. *et al.* (2017) 'Prevalence of *Campylobacter* species in wild birds of South Korea', *Avian Pathology: Journal of the W.V.P.A*, 46(5), pp. 474–480. Available at: https://doi.org/10.1080/03079457.2017.1315048.

Lacharme-Lora, L. *et al.* (2017) 'B lymphocytes play a limited role in clearance of *Campylobacter jejuni* from the chicken intestinal tract', *Scientific Reports*, 7(1), p. 45090. Available at: https://doi.org/10.1038/srep45090.

Ladely, S.R. *et al.* (2017) 'Isolation method (direct plating or enrichment) does not affect antimicrobial susceptibility of *Campylobacter* from chicken carcasses', *Journal of Food Safety*, 37(1), p. e12279. Available at: https://doi.org/10.1111/jfs.12279.

Lahti, L. and Shetty, S. (2022) 'Tools for microbiome analysis in R.' Available at: http://microbiome.github.com/microbiome.

Lan, N.P.H. *et al.* (2016) 'Invasive non-typhoidal *Salmonella* infections in Asia: clinical observations, disease outcome and dominant serovars from an infectious disease hospital in Vietnam', *PLOS Neglected Tropical Diseases*, 10(8), p. e0004857. Available at: https://doi.org/10.1371/journal.pntd.0004857.

Langmead, B. and Salzberg, S.L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature Methods*, 9(4), pp. 357–359. Available at: https://doi.org/10.1038/nmeth.1923.

Lanzl, M.I. *et al.* (2022) 'Combining enrichment with multiplex real-time PCR leads to faster detection and identification of *Campylobacter* spp. in food compared to ISO 10272–1:2017', *Food Microbiology*, 108, p. 104117. Available at: https://doi.org/10.1016/j.fm.2022.104117.

Lastovica, A.J. (2006) 'Emerging *Campylobacter* spp.: the tip of the iceberg', *Clinical Microbiology Newsletter*, 28(7), pp. 49–56. Available at: https://doi.org/10.1016/j.clinmicnews.2006.03.004.

Leinonen, R. *et al.* (2011) 'The Sequence Read Archive', *Nucleic Acids Research*, 39(suppl_1), pp. D19–D21. Available at: https://doi.org/10.1093/nar/gkq1019.

Lenth, R. (2023) 'emmeans: estimated marginal means, aka least-squares means. R package version 1.8.5'. Available at: https://CRAN.R-project.org/package=emmeans (Accessed: 27 July 2023).

Li, D. *et al.* (2015) 'MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph', *Bioinformatics*, 31(10), pp. 1674–1676. Available at: https://doi.org/10.1093/bioinformatics/btv033.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows–Wheeler transform', *Bioinformatics*, 25(14), pp. 1754–1760. Available at: https://doi.org/10.1093/bioinformatics/btp324.

Li, S. *et al.* (2020) 'Microbiome-informed food safety and quality: longitudinal consistency and cross-sectional distinctiveness of retail chicken breast microbiomes', *mSystems*, 5(5), pp. e00589-20. Available at: https://doi.org/10.1128/mSystems.00589-20.

Li, Y. *et al.* (2014) 'Nontyphoidal *Salmonella* infection in children with acute gastroenteritis: prevalence, serotypes, and antimicrobial resistance in Shanghai, China', *Foodborne Pathogens and Disease*, 11(3), pp. 200–206. Available at: https://doi.org/10.1089/fpd.2013.1629.

Liaw, J. *et al.* (2019) 'The *Campylobacter jejuni* type VI secretion system enhances the oxidative stress response and host colonization', *Frontiers in Microbiology*, 10, p. 2864. Available at: https://doi.org/10.3389/fmicb.2019.02864.

Lin, H. and Peddada, S.D. (2020) 'Analysis of compositions of microbiomes with bias correction', *Nature Communications*, 11(1), p. 3514. Available at: https://doi.org/10.1038/s41467-020-17041-7.

Lin, H. and Peddada, S.D. (2024) 'Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures', *Nature Methods*, 21(1), pp. 83–91. Available at: https://doi.org/10.1038/s41592-023-02092-7.

Liu, B. *et al.* (2021) 'VFDB 2022: a general classification scheme for bacterial virulence factors', *Nucleic Acids Research*, 50(D1), pp. D912–D917. Available at: https://doi.org/10.1093/nar/gkab1107.

Liu, F. *et al.* (2022) 'Global epidemiology of campylobacteriosis and the impact of COVID-19', *Frontiers in Cellular and Infection Microbiology*, 12, p. 979055. Available at: https://doi.org/10.3389/fcimb.2022.979055.

Liu, Q. *et al.* (2023) 'Advances in the application of molecular diagnostic techniques for the detection of infectious disease pathogens (Review)', *Molecular Medicine Reports*, 27(5), p. 104. Available at: https://doi.org/10.3892/mmr.2023.12991.

Liu, Y.-X. *et al.* (2021) 'A practical guide to amplicon and metagenomic analysis of microbiome data', *Protein & Cell*, 12(5), pp. 315–330. Available at: https://doi.org/10.1007/s13238-020-00724-8.

Llarena, A.-K. *et al.* (2016) 'Monomorphic genotypes within a generalist lineage of *Campylobacter jejuni* show signs of global dispersion', *Microbial Genomics*, 2(10), p. e000088. Available at: https://doi.org/10.1099/mgen.0.000088.

Loman, N.J. *et al.* (2013) 'A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of shiga-toxigenic *Escherichia coli* O104:H4', *JAMA*, 309(14), pp. 1502–1510. Available at: https://doi.org/10.1001/jama.2013.3231.

Louis, V.R. *et al.* (2005) 'Temperature-driven *Campylobacter* seasonality in England and Wales', *Applied and Environmental Microbiology*, 71(1), pp. 85–92. Available at: https://doi.org/10.1128/AEM.71.1.85-92.2005.

Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550. Available at: https://doi.org/10.1186/s13059-014-0550-8.

Lu, J. *et al.* (2017) 'Bracken: estimating species abundance in metagenomics data', *PeerJ Computer Science*, 3, p. e104. Available at: https://doi.org/10.7717/peerj-cs.104.

Luethy, P.M. *et al.* (2017) 'Microbiota-derived short-chain fatty acids modulate expression of *Campylobacter jejuni* determinants required for commensalism and virulence', *mBio*, 8(3), pp. e00407-17. Available at: https://doi.org/10.1128/mBio.00407-17.

Lynch, Ó.A. *et al.* (2011) 'Occurrence of fastidious *Campylobacter* spp. in fresh meat and poultry using an adapted cultural protocol', *International Journal of Food Microbiology*, 150(2), pp. 171–177. Available at: https://doi.org/10.1016/j.ijfoodmicro.2011.07.037.

Ma, L. *et al.* (2016) 'Metagenomic assembly reveals hosts of antibiotic resistance genes and the shared resistome in pig, chicken, and human feces', *Environmental Science & Technology*, 50(1), pp. 420–427. Available at: https://doi.org/10.1021/acs.est.5b03522.

Maechler, M. *et al.* (2022) 'cluster: cluster analysis basics and extensions. R package version 2.1.4.' Available at: https://CRAN.R-project.org/package=cluster.

Magiorakos, A.-P. *et al.* (2012) 'Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance', *Clinical Microbiology and Infection*, 18(3), pp. 268–281. Available at: https://doi.org/10.1111/j.1469-0691.2011.03570.x.

Magnet, S., Courvalin, P. and Lambert, T. (1999) 'Activation of the cryptic *aac(6')-Iy* aminoglycoside resistance gene of *Salmonella* by a chromosomal deletion generating a transcriptional fusion', *Journal of Bacteriology*, 181(21), pp. 6650–6655. Available at: https://doi.org/10.1128/JB.181.21.6650-6655.1999.

Maguire, F. *et al.* (2020) 'Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic islands', *Microbial Genomics*, 6(10), p. e000436. Available at: https://doi.org/10.1099/mgen.0.000436.

Manson, A.L. *et al.* (2019) 'Chicken meat-associated Enterococci: influence of agricultural antibiotic use and connection to the clinic', *Applied and Environmental Microbiology*, 85(22), pp. e01559-19. Available at: https://doi.org/10.1128/AEM.01559-19.

Marasini, D. *et al.* (2018) 'Phylogenetic relatedness among plasmids harbored by *Campylobacter jejuni* and *Campylobacter coli* isolated from retail meats', *Frontiers in Microbiology*, 9, p. 2167. Available at: https://doi.org/10.3389/fmicb.2018.02167.

Marasini, D. and Fakhr, M.K. (2014) 'Exploring PFGE for detecting large plasmids in *Campylobacter jejuni* and *Campylobacter coli* isolated from various retail meats', *Pathogens*, 3(4), pp. 833–844. Available at: https://doi.org/10.3390/pathogens3040833.

Marasini, D. and Fakhr, M.K. (2016) 'Whole-genome sequencing of a *Campylobacter jejuni* strain isolated from retail chicken meat reveals the presence of a megaplasmid with Mu-like prophage and multidrug resistance genes', *Genome Announcements*, 4(3), p. 10.1128/genomea.00460-16. Available at: https://doi.org/10.1128/genomea.00460-16.

Marçais, G. *et al.* (2018) 'MUMmer4: a fast and versatile genome alignment system', *PLOS Computational Biology*, 14(1), p. e1005944. Available at: https://doi.org/10.1371/journal.pcbi.1005944.

Mardis, E.R. (2017) 'DNA sequencing technologies: 2006-2016', *Nature Protocols*, 12(2), pp. 213–218. Available at: https://doi.org/10.1038/nprot.2016.182.

Marmion, M. *et al.* (2021) 'The changing microbiome of poultry meat; from farm to fridge', *Food Microbiology*, 99, p. 103823. Available at: https://doi.org/10.1016/j.fm.2021.103823.

Marotta, F. *et al.* (2019) 'Antimicrobial resistance genotypes and phenotypes of *Campylobacter jejuni* isolated in Italy from humans, birds from wild and urban habitats, and poultry', *PLOS ONE*, 14(10), p. e0223804. Available at: https://doi.org/10.1371/journal.pone.0223804.

Marquet, M. *et al.* (2022) 'Evaluation of microbiome enrichment and host DNA depletion in human vaginal samples using Oxford Nanopore's adaptive sequencing', *Scientific Reports*, 12(1), p. 4000. Available at: https://doi.org/10.1038/s41598-022-08003-8.

Martin, S. *et al.* (2022) 'Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples', *Genome Biology*, 23(1), p. 11. Available at: https://doi.org/10.1186/s13059-021-02582-x.

Martinez-Porchas, M. *et al.* (2017) 'How conserved are the conserved 16S-rRNA regions?', *PeerJ*, 5, p. e3036. Available at: https://doi.org/10.7717/peerj.3036.

Masanta, W.O. *et al.* (2013) 'Modification of intestinal microbiota and its consequences for innate immune response in the pathogenesis of campylobacteriosis', *Journal of Immunology Research*, 2013(1), p. 526860. Available at: https://doi.org/10.1155/2013/526860.

McCarthy, N. *et al.* (2021) *Enhanced molecular-based (MLST/whole genome) surveillance and source attribution of* Campylobacter *infections in the UK*. Food Standards Agency. Available at: https://doi.org/10.46756/sci.fsa.ksj135.

McDermott, S. (2021) *NARMS Retail Meat Surveillance Laboratory Protocol*. FDA, Center for Veterinary Medicine. Available at: https://www.fda.gov/media/93332/download#:~:text=Notes%3A%20For%20Campylobacter%2C%20only%20set,Do%20not%20stack%20containers.

McMurdie, P.J. and Holmes, S. (2013) 'phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data', *PLOS ONE*, 8(4), p. e61217. Available at: https://doi.org/10.1371/journal.pone.0061217.

McMurdie, P.J. and Holmes, S. (2014) 'Waste not, want not: why rarefying microbiome data is inadmissible', *PLOS Computational Biology*, 10(4), p. e1003531. Available at: https://doi.org/10.1371/journal.pcbi.1003531.

von Meijenfeldt, F.A.B. *et al.* (2019) 'Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT', *Genome Biology*, 20(1), p. 217. Available at: https://doi.org/10.1186/s13059-019-1817-x.

Meziti, A. *et al.* (2021) 'The reliability of metagenome-assembled genomes (MAGs) in representing natural populations: insights from comparing MAGs against isolate genomes derived from the same fecal sample', *Applied and Environmental Microbiology*, 87(6), pp. e02593-20. Available at: https://doi.org/10.1128/AEM.02593-20.

Miller, W.G. *et al.* (2005) 'Extended multilocus sequence typing system for *Campylobacter coli*, *C. lari*, *C. upsaliensis*, and *C. helveticus*', *Journal of Clinical Microbiology*, 43(5), pp. 2315–2329. Available at: https://doi.org/10.1128/JCM.43.5.2315-2329.2005.

Miller, W.G. *et al.* (2024) '*Campylobacter devanensis* sp. nov., *Campylobacter porcelli* sp. nov., and *Campylobacter vicugnae* sp. nov., three novel *Campylobacter lanienae*-like species recovered from swine, small ruminants, and camelids', *International Journal of Systematic and Evolutionary Microbiology*, 74(6), p. 006405. Available at: https://doi.org/10.1099/ijsem.0.006405.

Minh, B.Q., Nguyen, M.A.T. and von Haeseler, A. (2013) 'Ultrafast approximation for phylogenetic bootstrap', *Molecular Biology and Evolution*, 30(5), pp. 1188–1195. Available at: https://doi.org/10.1093/molbev/mst024.

Mohamed, H.A.A., Williams, L.K. and van Klink, E. (2021) 'The diversity of *Campylobacter* spp. throughout the poultry processing plant', *Zoonoses and Public Health*, 68(7), pp. 769–780. Available at: https://doi.org/10.1111/zph.12852.

Molina-Bolívar, J.A. and Galisteo-González, F. (2005) 'Latex immunoagglutination assays', *Journal of Macromolecular Science, Part C*, 45(1), pp. 59–98. Available at: https://doi.org/10.1081/MC-200045819.

Moran, L. *et al.* (2011) 'Restoring the selectivity of Bolton broth during enrichment for *Campylobacter* spp. from raw chicken', *Letters in Applied Microbiology*, 52(6), pp. 614–618. Available at: https://doi.org/10.1111/j.1472-765X.2011.03046.x.

Morgulis, A. *et al.* (2008) 'Database indexing for production MegaBLAST searches', *Bioinformatics*, 24(16), pp. 1757–1764. Available at: https://doi.org/10.1093/bioinformatics/btn322.

Morisse, P. *et al.* (2021) 'Scalable long read self-correction and assembly polishing with multiple sequence alignment', *Scientific Reports*, 11(1), p. 761. Available at: https://doi.org/10.1038/s41598-020-80757-5.

Morita, D. *et al.* (2023) 'Whole-genome and plasmid comparative analysis of *Campylobacter jejuni* from human patients in Toyama, Japan, from 2015 to 2019', *Microbiology Spectrum*, 11(1), pp. e02659-22. Available at: https://doi.org/10.1128/spectrum.02659-22.

Mourkas, E. *et al.* (2019) 'Gene pool transmission of multidrug resistance among *Campylobacter* from livestock, sewage and human disease', *Environmental Microbiology*, 21(12), pp. 4597–4613. Available at: https://doi.org/10.1111/1462-2920.14760.

Muhamad Rizal, N.S. *et al.* (2020) 'Advantages and limitations of 16S rRNA next-generation sequencing for pathogen identification in the diagnostic microbiology laboratory: perspectives from a middle-income country', *Diagnostics*, 10(10), p. 816. Available at: https://doi.org/10.3390/diagnostics10100816.

Mukherjee, S. *et al.* (2015) 'Large-scale contamination of microbial isolate genomes by Illumina PhiX control', *Standards in Genomic Sciences*, 10(1), p. 18. Available at: https://doi.org/10.1186/1944-3277-10-18.

Müller, K. and Wickham, H. (2023) 'tibble: simple data frames_. R package version 3.2.0'. Available at: https://CRAN.R-project.org/package=tibble.

Munk, P. *et al.* (2018) 'Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries', *Nature Microbiology*, 3(8), pp. 898–908. Available at: https://doi.org/10.1038/s41564-018-0192-9.

Murphy, C., Carroll, C. and Jordan, K. (2006) 'Environmental survival mechanisms of the foodborne pathogen *Campylobacter jejuni*', *Journal of Applied Microbiology*, 100(4), pp. 623–632. Available at: https://doi.org/10.1111/j.1365-2672.2006.02903.x.

Navidad, J.F. *et al.* (2013) 'Evaluation of Luminex xTAG gastrointestinal pathogen analyte-specific reagents for high-throughput, simultaneous detection of bacteria, viruses, and parasites of clinical and public health importance', *Journal of Clinical Microbiology*, 51(9), pp. 3018–3024. Available at: https://doi.org/10.1128/JCM.00896-13.

Nearing, J.T. *et al.* (2022) 'Microbiome differential abundance methods produce different results across 38 datasets', *Nature Communications*, 13(1), p. 342. Available at: https://doi.org/10.1038/s41467-022-28034-z.

Neoh, H. *et al.* (2019) 'Pulsed-field gel electrophoresis (PFGE): a review of the "gold standard" for bacteria typing and current alternatives', *Infection, Genetics and Evolution*, 74, p. 103935. Available at: https://doi.org/10.1016/j.meegid.2019.103935.

Ng, L.K., Stiles, M.E. and Taylor, D.E. (1985) 'Inhibition of *Campylobacter coli* and *Campylobacter jejuni* by antibiotics used in selective growth media', *Journal of Clinical Microbiology*, 22(4), pp. 510–514.

Nguyen, L.-T. *et al.* (2015) 'IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies', *Molecular Biology and Evolution*, 32(1), pp. 268–274. Available at: https://doi.org/10.1093/molbev/msu300.

Nhung, N.T. *et al.* (2024) 'A review and meta-analysis of non-typhoidal *Salmonella* in Vietnam: challenges to the control and antimicrobial resistance traits of a neglected zoonotic pathogen', *One Health*, 18, p. 100698. Available at: https://doi.org/10.1016/j.onehlt.2024.100698.

Nichols, G.L. (2005) 'Fly transmission of *Campylobacter*', *Emerging Infectious Diseases*, 11(3), pp. 361–364. Available at: https://doi.org/10.3201/eid1103.040460.

Nichols, G.L. *et al.* (2012) '*Campylobacter* epidemiology: a descriptive study reviewing 1 million cases in England and Wales between 1989 and 2011', *BMJ Open*, 2(4), p. e001179. Available at: https://doi.org/10.1136/bmjopen-2012-001179.

Nielsen, H.L. *et al.* (2012) 'Short-term and medium-term clinical outcomes of *Campylobacter concisus* infection', *Clinical Microbiology and Infection*, 18(11), pp. E459–E465. Available at: https://doi.org/10.1111/j.1469-0691.2012.03990.x.

Oakley, B.B. *et al.* (2013) 'The poultry-associated microbiome: network analysis and farm-to-fork characterizations', *PLOS ONE*, 8(2), p. e57190. Available at: https://doi.org/10.1371/journal.pone.0057190.

Oksanen, J. *et al.* (2022) 'vegan: community ecology package'. Available at: https://CRAN.R-project.org/package=vegan.

Ondov, B.D. *et al.* (2016) 'Mash: fast genome and metagenome distance estimation using MinHash', *Genome Biology*, 17(1), p. 132. Available at: https://doi.org/10.1186/s13059-016-0997-x.

Ondov, B.D. *et al.* (2019) 'Mash Screen: high-throughput sequence containment estimation for genome discovery', *Genome Biology*, 20(1), p. 232. Available at: https://doi.org/10.1186/s13059-019-1841-x.

Oren, A. and Garrity, G.M. (2021) 'Valid publication of the names of forty-two phyla of prokaryotes', *International Journal of Systematic and Evolutionary Microbiology*, 71(10). Available at: https://doi.org/10.1099/ijsem.0.005056.

Overdevest, I. *et al.* (2011) 'Extended-spectrum β-lactamase genes of *Escherichia coli* in chicken meat and humans, the Netherlands', *Emerging Infectious Diseases*, 17(7), pp. 1216–1222. Available at: https://doi.org/10.3201/eid1707.110209.

Oyarzabal, O. and Battie, C. (2012) 'Immunological methods for the detection of *Campylobacter* spp. - current applications and potential use in biosensors', in *Trends in Immunolabelled and Related Techniques*, pp. 203–226. Available at: 10.5772/35692 (Accessed: 19 May 2024).

Page, A.J. *et al.* (2015) 'Roary: rapid large-scale prokaryote pan genome analysis', *Bioinformatics*, 31(22), pp. 3691–3693. Available at: https://doi.org/10.1093/bioinformatics/btv421.

Page, A.J., Ainsworth, E.V. and Langridge, G.C. (2020) 'socru: typing of genome-level order and orientation around ribosomal operons in bacteria', *Microbial Genomics*, 6(7), p. e000396. Available at: https://doi.org/10.1099/mgen.0.000396.

Parisi, A. *et al.* (2020) 'The role of animals as a source of antimicrobial resistant nontyphoidal *Salmonella* causing invasive and non-invasive human disease in Vietnam', *Infection, Genetics and Evolution*, 85, p. 104534. Available at: https://doi.org/10.1016/j.meegid.2020.104534.

Park, S.H. *et al.* (2014) 'Current and emerging technologies for rapid detection and characterization of *Salmonella* in poultry and poultry products', *Food Microbiology*, 38, pp. 250–262. Available at: https://doi.org/10.1016/j.fm.2013.10.002.

Parker, C.T. *et al.* (2007) 'Common genomic features of *Campylobacter jejuni* subsp. *doylei* strains distinguish them from *C. jejuni* subsp. *jejuni*', *BMC Microbiology*, 7(1), p. 50. Available at: https://doi.org/10.1186/1471-2180-7-50.

Parks, D.H. *et al.* (2015) 'CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes', *Genome Research*, 25(7), pp. 1043–1055. Available at: https://doi.org/10.1101/gr.186072.114.

Parsons, C.M., Potter, L.M. and Brown, R.D. (1983) 'Effects of dietary carbohydrate and of intestinal microflora on excretion of endogenous amino acids by poultry', *Poultry Science*, 62(3), pp. 483–489. Available at: https://doi.org/10.3382/ps.0620483.

Parte, A.C. *et al.* (2020) 'List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ', *International Journal of Systematic and Evolutionary Microbiology*, 70(11), pp. 5607–5612. Available at: https://doi.org/10.1099/ijsem.0.004332.

Pascoe, B. *et al.* (2020) 'Genomic epidemiology of *Campylobacter jejuni* associated with asymptomatic pediatric infection in the Peruvian Amazon', *PLOS Neglected Tropical Diseases*, 14(8), p. e0008533. Available at: https://doi.org/10.1371/journal.pntd.0008533.

Patel, S. *et al.* (2021) 'Differential analysis of longitudinal methicillin-resistant *Staphylococcus aureus* colonization in relation to microbial shifts in the nasal microbiome of neonatal piglets', *mSystems*, 6(4), pp. e00152-21. Available at: https://doi.org/10.1128/mSystems.00152-21.

Pazzaglia, G. *et al.* (1991) '*Campylobacter* diarrhoea and an association of recent disease with asymptomatic shedding in Egyptian children', *Epidemiology & Infection*, 106(1), pp. 77–82.

Peabody, M.A. *et al.* (2015) 'Evaluation of shotgun metagenomics sequence classification methods using *in silico* and *in vitro* simulated communities', *BMC Bioinformatics*, 16(1), p. 362. Available at: https://doi.org/10.1186/s12859-015-0788-5.

Pedersen, T.L. (2022) 'patchwork: the composer of plots. R package version 1.1.2.' Available at: https://CRAN.R-project.org/package=patchwork.

Pereira-Dias, J. *et al.* (2021) 'The gut microbiome of healthy Vietnamese adults and children is a major reservoir for resistance genes against critical antimicrobials', *The Journal of Infectious Diseases*, 224(12 Suppl 2), pp. S840–S847. Available at: https://doi.org/10.1093/infdis/jiab398.

Pereira-Marques, J. *et al.* (2019) 'Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis', *Frontiers in Microbiology*, 10, p. 1277. Available at: https://doi.org/10.3389/fmicb.2019.01277.

Persson, S. and Olsen, K.E. (2005) 'Multiplex PCR for identification of *Campylobacter coli* and *Campylobacter jejuni* from pure cultures and directly on stool samples', *Journal of Medical Microbiology*, 54(11), pp. 1043–1047. Available at: https://doi.org/10.1099/jmm.0.46203-0.

Peterson, C.-L. *et al.* (2022) 'Clinical metagenomics is increasingly accurate and affordable to detect enteric bacterial pathogens in stool', *Microorganisms*, 10(2), p. 441. Available at: https://doi.org/10.3390/microorganisms10020441.

Picot, J. *et al.* (2012) 'Flow cytometry: retrospective, fundamentals and recent instrumentation', *Cytotechnology*, 64(2), pp. 109–130. Available at: https://doi.org/10.1007/s10616-011-9415-0.

Piddock, L.J.V. *et al.* (2003) 'Fluoroquinolone resistance in *Campylobacter* species from man and animals: detection of mutations in topoisomerase genes.', *The Journal of Antimicrobial Chemotherapy*, 51(1), pp. 19–26. Available at: https://doi.org/10.1093/jac/dkg033.

Pires, S.M. and Devleesschauwer, B. (2021) 'Chapter 1 - Estimates of global disease burden associated with foodborne pathogens', in J.G. Morris and D.J. Vugia (eds) *Foodborne Infections and Intoxications (Fifth Edition)*. Academic Press, pp. 3–17. Available at: https://doi.org/10.1016/B978-0-12-819519-2.00020-7.

Platts-Mills, J.A. *et al.* (2014) 'Detection of *Campylobacter* in stool and determination of significance by culture, enzyme immunoassay, and PCR in developing countries', *Journal of Clinical Microbiology*, 52(4), pp. 1074–1080. Available at: https://doi.org/10.1128/JCM.02935-13.

Polansky, O. *et al.* (2016) 'Important metabolic pathways and biological processes expressed by chicken cecal microbiota', *Applied and Environmental Microbiology*, 82(5), pp. 1569–1576. Available at: https://doi.org/10.1128/AEM.03473-15.

Poly, F. *et al.* (2018) 'Update on *Campylobacter* vaccine development', *Human Vaccines & Immunotherapeutics*, 15(6), pp. 1389–1400. Available at: https://doi.org/10.1080/21645515.2018.1528410.

Pölzler, T., Stüger, H. and Lassnig, H. (2018) 'Prevalence of most common human pathogenic *Campylobacter* spp. in dogs and cats in Styria, Austria', *Veterinary Medicine and Science*, 4(2), pp. 115–125. Available at: https://doi.org/10.1002/vms3.93.

Poudel, S. *et al.* (2022) 'Prevalence, antimicrobial resistance, and molecular characterization of *Campylobacter* isolated from broilers and broiler meat raised without antibiotics', *Microbiology Spectrum*, 10(3), pp. e00251-22. Available at: https://doi.org/10.1128/spectrum.00251-22.

Poxton, I.R., McCoubrey, J. and Blair, G. (2001) 'The pathogenicity of *Clostridium difficile*', *Clinical Microbiology and Infection*, 7(8), pp. 421–427. Available at: https://doi.org/10.1046/j.1198-743x.2001.00287.x.

Prendergast, D.M. *et al.* (2022) 'Genomic diversity, virulence and source of *Campylobacter jejuni* contamination in Irish poultry slaughterhouses by whole genome sequencing', *Journal of Applied Microbiology*, 133(5), pp. 3150–3160. Available at: https://doi.org/10.1111/jam.15753.

Prjibelski, A. *et al.* (2020) 'Using SPAdes *de novo* assembler', *Current Protocols in Bioinformatics*, 70(1), p. e102. Available at: https://doi.org/10.1002/cpbi.102.

Public Health England (2020) 'Detection and enumeration of *Campylobacter* species: National Infection Service Food Water and Environmental Microbiology Standard Method'. London: Public Health England.

Quainoo, S. *et al.* (2017) 'Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis', *Clinical Microbiology Reviews*, 30(4), pp. 1015–1063. Available at: https://doi.org/10.1128/CMR.00016-17.

Quince, C. *et al.* (2017) 'Shotgun metagenomics, from sampling to analysis', *Nature Biotechnology*, 35(9), pp. 833–844. Available at: https://doi.org/10.1038/nbt.3935.

R Core Team (2023) 'R: a language and environment for statistical computing'. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Rantsiou, K. *et al.* (2018) 'Next generation microbiological risk assessment: opportunities of whole genome sequencing (WGS) for foodborne pathogen surveillance, source tracking and risk assessment', *International Journal of Food Microbiology*, 287, pp. 3–9. Available at: https://doi.org/10.1016/j.ijfoodmicro.2017.11.007.

Rao, M.R. *et al.* (2001) 'Pathogenicity and convalescent excretion of *Campylobacter* in rural Egyptian children', *American Journal of Epidemiology*, 154(2), pp. 166–173. Available at: https://doi.org/10.1093/aje/154.2.166.

Rasschaert, G. *et al.* (2020) '*Campylobacter* contamination of broilers: the role of transport and slaughterhouse', *International Journal of Food Microbiology*, 322, p. 108564. Available at: https://doi.org/10.1016/j.ijfoodmicro.2020.108564.

Rawson, T., Dawkins, M.S. and Bonsall, M.B. (2019) 'A mathematical model of *Campylobacter* dynamics within a broiler flock', *Frontiers in Microbiology*, 10, p. 1940. Available at: https://doi.org/10.3389/fmicb.2019.01940.

Revell, L.J. (2012) 'phytools: an R package for phylogenetic comparative biology (and other things)', *Methods in Ecology and Evolution*, 3(2), pp. 217–223. Available at: https://doi.org/10.1111/j.2041-210X.2011.00169.x.

Ricke, S.C. *et al.* (2019) 'Developments in rapid detection methods for the detection of foodborne *Campylobacter* in the United States', *Frontiers in Microbiology*, 9. Available at: https://doi.org/10.3389/fmicb.2018.03280.

Robertson, J. *et al.* (2020) 'Universal whole-sequence-based plasmid typing and its utility to prediction of host range and epidemiological surveillance', *Microbial Genomics*, 6(10), p. e000435. Available at: https://doi.org/10.1099/mgen.0.000435.

Robertson, R.C. *et al.* (2019) 'The human microbiome and child growth – first 1000 days and beyond', *Trends in Microbiology*, 27(2), pp. 131–147. Available at: https://doi.org/10.1016/j.tim.2018.09.008.

Rodríguez, A. *et al.* (2015) 'Design of primers and probes for quantitative real-time PCR methods', in C. Basu (ed.) *PCR Primer Design*. New York, NY: Springer, pp. 31–56. Available at: https://doi.org/10.1007/978-1-4939-2365-6_3.

Rodríguez-Valera, F. (2004) 'Environmental genomics, the big picture?', *FEMS Microbiology Letters*, 231(2), pp. 153–158. Available at: https://doi.org/10.1016/S0378-1097(04)00006-0.

Rosanova, M.T. *et al.* (2002) 'Risk factors for mortality caused by nontyphoidal *Salmonella* sp. in children', *International Journal of Infectious Diseases*, 6(3), pp. 187–190. Available at: https://doi.org/10.1016/s1201-9712(02)90109-8.

Rosenquist, H., Bengtsson, A. and Hansen, T.B. (2007) 'A collaborative study on a Nordic standard protocol for detection and enumeration of thermotolerant *Campylobacter* in food (NMKL 119, 3. Ed., 2007)', *International Journal of Food Microbiology*, 118(2), pp. 201–213. Available at: https://doi.org/10.1016/j.ijfoodmicro.2007.07.037.

Rossi, M. *et al.* (2009) '*Campylobacter avium* sp. nov., a hippurate-positive species isolated from poultry', *International Journal of Systematic and Evolutionary Microbiology*, 59(9), pp. 2364–2369. Available at: https://doi.org/10.1099/ijs.0.007419-0.

Royden, A. *et al.* (2021) '*Campylobacter* contamination at retail of halal chicken produced in the United Kingdom.', *Journal of Food Protection*, 84(8), pp. 1433–1445. Available at: https://doi.org/10.4315/JFP-20-428.

Royer, C. *et al.* (2024) 'Comparison of metagenomic and traditional methods for diagnosis of *E. coli* enteric infections', *mBio*, 15(4), pp. e03422-23. Available at: https://doi.org/10.1128/mbio.03422-23.

RStudio Team (2022) 'RStudio: integrated development environment for R'. Boston, MA: RStudio, Inc. Available at: http://www.rstudio.com/.

Sakaridis, I. *et al.* (2018) 'Investigating the association between the caecal microbiomes of broilers and *Campylobacter* burden', *Frontiers in Microbiology*, 9, p. 927. Available at: https://doi.org/10.3389/fmicb.2018.00927.

Salipante, S.J. and Hall, B.G. (2003) 'Determining the limits of the evolutionary potential of an antibiotic resistance gene', *Molecular Biology and Evolution*, 20(4), pp. 653–659. Available at: https://doi.org/10.1093/molbev/msg074.

Samarth, D.P. and Kwon, Y.M. (2020) 'Horizontal genetic exchange of chromosomally encoded markers between *Campylobacter jejuni* cells', *PLOS ONE*, 15(10), p. e0241058. Available at: https://doi.org/10.1371/journal.pone.0241058.

San Millan, A. and MacLean, R.C. (2017) 'Fitness costs of plasmids: a limit to plasmid transmission', *Microbiology Spectrum*, 5(5), p. 10.1128/microbiolspec.mtbp-0016–2017. Available at: https://doi.org/10.1128/microbiolspec.mtbp-0016-2017.

Santini, C. *et al.* (2010) 'Characterization of probiotic strains: an application as feed additives in poultry against *Campylobacter jejuni*', *International Journal of Food Microbiology*, 141 Suppl 1, pp. S98-108. Available at: https://doi.org/10.1016/j.ijfoodmicro.2010.03.039.

Satam, H. *et al.* (2023) 'Next-generation sequencing technology: current trends and advancements', *Biology*, 12(7), p. 997. Available at: https://doi.org/10.3390/biology12070997.

Schwengers, O. *et al.* (2021) 'Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification', *Microbial Genomics*, 7(11), p. 000685. Available at: https://doi.org/10.1099/mgen.0.000685.

Seemann, T. (2014) 'Prokka: rapid prokaryotic genome annotation', *Bioinformatics*, 30(14), pp. 2068–2069. Available at: https://doi.org/10.1093/bioinformatics/btu153.

Seliwiorstow, T. *et al.* (2015) 'Comparison of sample types and analytical methods for the detection of highly *Campylobacter*-colonized broiler flocks at different stages in the poultry meat production chain', *Foodborne Pathogens and Disease*, 12(5), pp. 399–405. Available at: https://doi.org/10.1089/fpd.2014.1894.

Seliwiorstow, T. *et al.* (2016) 'Comparative performance of isolation methods using Preston broth, Bolton broth and their modifications for the detection of *Campylobacter* spp. from naturally contaminated fresh and frozen raw poultry meat', *International Journal of Food Microbiology*, 234, pp. 60–64. Available at: https://doi.org/10.1016/j.ijfoodmicro.2016.06.040.

Shen, C.-H. (2019) 'Chapter 9 - Amplification of nucleic acids', in C.-H. Shen (ed.) *Diagnostic Molecular Biology*. Academic Press, pp. 215–247. Available at: https://doi.org/10.1016/B978-0-12-802823-0.00009-2.

Shen, Z. *et al.* (2016) 'Identification and functional analysis of two toxin–antitoxin systems in *Campylobacter jejuni*', *Molecular Microbiology*, 101(6), pp. 909–923. Available at: https://doi.org/10.1111/mmi.13431.

Sheppard, S.K., Dallas, J.F., *et al.* (2010) 'Evolution of an agriculture-associated disease causing *Campylobacter coli* clade: evidence from national surveillance data in Scotland', *PLOS ONE*, 5(12), p. e15708. Available at: https://doi.org/10.1371/journal.pone.0015708.

Sheppard, S.K., Colles, F., *et al.* (2010) 'Host association of *Campylobacter* genotypes transcends geographic variation', *Applied and Environmental Microbiology*, 76(15), pp. 5269–5277. Available at: https://doi.org/10.1128/AEM.00124-10.

Sheppard, S.K. *et al.* (2011) 'Introgression in the genus *Campylobacter*: generation and spread of mosaic alleles', *Microbiology*, 157(Pt 4), pp. 1066–1074. Available at: https://doi.org/10.1099/mic.0.045153-0.

Sheppard, S.K. *et al.* (2013) 'Progressive genome-wide introgression in agricultural *Campylobacter coli*', *Molecular Ecology*, 22(4), pp. 1051–1064. Available at: https://doi.org/10.1111/mec.12162.

Shi, Y. *et al.* (2022) 'Metagenomic sequencing for microbial DNA in human samples: emerging technological advances', *International Journal of Molecular Sciences*, 23(4), p. 2181. Available at: https://doi.org/10.3390/ijms23042181.

Sibinelli-Sousa, S. *et al.* (2022) 'Revisiting the steps of *Salmonella* gut infection with a focus on antagonistic interbacterial interactions', *The FEBS Journal*, 289(14), pp. 4192–4211. Available at: https://doi.org/10.1111/febs.16211.

Šikić Pogačar, M. *et al.* (2020) 'Effect of *Lactobacillus* spp. on adhesion, invasion, and translocation of *Campylobacter jejuni* in chicken and pig small-intestinal epithelial cell lines', *BMC veterinary research*, 16(1), p. 34. Available at: https://doi.org/10.1186/s12917-020-2238-5.

Singh, P. *et al.* (2015) 'Intestinal microbial communities associated with acute enteric infections and disease recovery', *Microbiome*, 3(1), p. 45. Available at: https://doi.org/10.1186/s40168-015-0109-2.

Sinulingga, T.S. *et al.* (2020) 'Occurrence of *Campylobacter* species from broiler chickens and chicken meat in Malaysia', *Tropical Animal Health and Production*, 52(1), pp. 151–157. Available at: https://doi.org/10.1007/s11250-019-01995-y.

Skarp, C.P.A., Hänninen, M.-L. and Rautelin, H.I.K. (2016) 'Campylobacteriosis: the role of poultry meat', *Clinical Microbiology and Infection*, 22(2), pp. 103–109. Available at: https://doi.org/10.1016/j.cmi.2015.11.019.

Slatko, B.E., Gardner, A.F. and Ausubel, F.M. (2018) 'Overview of next generation sequencing technologies', *Current Protocols in Molecular Biology*, 122(1), p. e59. Available at: https://doi.org/10.1002/cpmb.59.

Sofka, D. *et al.* (2015) 'Changes within the intestinal flora of broilers by colonisation with *Campylobacter jejuni*', *Berliner Und Munchener Tierarztliche Wochenschrift*, 128(3–4), pp. 104–110.

Speegle, L. *et al.* (2009) 'Use of cellulose filters to isolate *Campylobacter* spp. from naturally contaminated retail broiler meat', *Journal of Food Protection*, 72(12), pp. 2592–2596. Available at: https://doi.org/10.4315/0362-028x-72.12.2592.

Stahl, M. *et al.* (2011) 'l-fucose utilization provides *Campylobacter jejuni* with a competitive advantage', *Proceedings of the National Academy of Sciences of the United States of America*, 108(17), pp. 7194–7199. Available at: https://doi.org/10.1073/pnas.1014125108.

Stahl, M., Butcher, J. and Stintzi, A. (2012) 'Nutrient acquisition and metabolism by *Campylobacter jejuni*', *Frontiers in Cellular and Infection Microbiology*, 2, p. 5. Available at: https://doi.org/10.3389/fcimb.2012.00005.

Standards Australia (2015) 'AS 5013.6:2015: Method 6: Examination for specific organisms—*Campylobacter*'. Sydney.

Stanley, D. *et al.* (2015) 'Comparison of fecal and cecal microbiotas reveals qualitative similarities but quantitative differences', *BMC Microbiology*, 15(1), p. 51. Available at: https://doi.org/10.1186/s12866-015-0388-6.

Stern, N.J. *et al.* (1995) '*Campylobacter* spp. in broilers on the farm and after transport', *Poultry Science*, 74(6), pp. 937–941. Available at: https://doi.org/10.3382/ps.0740937.

Stoakes, E. *et al.* (2022) 'Substrate utilisation and energy metabolism in non-growing *Campylobacter jejuni* M1cam', *Microorganisms*, 10(7), p. 1355. Available at: https://doi.org/10.3390/microorganisms10071355.

Summers, J. *et al.* (2024) '*Campylobacter*', in G.W. Smithers (ed.) *Encyclopedia of Food Safety (Second Edition)*. Oxford: Academic Press, pp. 95–110. Available at: https://doi.org/10.1016/B978-0-12-822521-9.00035-6.

Taboada, E.N. *et al.* (2013) 'Current methods for molecular typing of *Campylobacter* species', *Journal of Microbiological Methods*, 95(1), pp. 24–31. Available at: https://doi.org/10.1016/j.mimet.2013.07.007.

Tam, C.C. *et al.* (2012) 'Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice', *Gut*, 61(1), pp. 69–77. Available at: https://doi.org/10.1136/gut.2011.238386.

Tam, C.C. and O'Brien, S.J. (2016) 'Economic cost of *Campylobacter*, norovirus and rotavirus disease in the United Kingdom', *PLOS ONE*, 11(2), p. e0138526. Available at: https://doi.org/10.1371/journal.pone.0138526.

Tauxe, R.V. *et al.* (1985) 'Illness associated with *Campylobacter laridis*, a newly recognized *Campylobacter* species', *Journal of Clinical Microbiology*, 21(2), pp. 222–225. Available at: https://doi.org/10.1128/jcm.21.2.222-225.1985.

Taylor, T.L. *et al.* (2019) 'Rapid, multiplexed, whole genome and plasmid sequencing of foodborne pathogens using long-read nanopore technology', *Scientific Reports*, 9(1), p. 16350. Available at: https://doi.org/10.1038/s41598-019-52424-x.

Tedersoo, T. *et al.* (2022) 'The prevalence, counts, and MLST genotypes of *Campylobacter* in poultry meat and genomic comparison with clinical isolates', *Poultry Science*, 101(4), p. 101703. Available at: https://doi.org/10.1016/j.psj.2022.101703.

Teh, K.H., Flint, S. and French, N. (2010) 'Biofilm formation by *Campylobacter jejuni* in controlled mixed-microbial populations', *International Journal of Food Microbiology*, 143(3), pp. 118–124. Available at: https://doi.org/10.1016/j.ijfoodmicro.2010.07.037.

Tejera, N. *et al.* (2020) 'Genome-scale metabolic model driven design of a defined medium for *Campylobacter jejuni* M1cam', *Frontiers in Microbiology*, 11. Available at: https://doi.org/10.3389/fmicb.2020.01072.

Teunisse, G. (2022) 'Fantaxtic - nested bar plots for phyloseq data'. Available at: https://github.com/gmteunisse/Fantaxtic.

The, H.C. *et al.* (2018) 'Assessing gut microbiota perturbations during the early phase of infectious diarrhea in Vietnamese children', *Gut Microbes*, 9(1), pp. 38–54. Available at: https://doi.org/10.1080/19490976.2017.1361093.

The, H.C. and Le, S.-N.H. (2022) 'Dynamic of the human gut microbiome under infectious diarrhea', *Current Opinion in Microbiology*, 66, pp. 79–85. Available at: https://doi.org/10.1016/j.mib.2022.01.006.

Thibodeau, A. *et al.* (2015) 'Chicken caecal microbiome modifications induced by *Campylobacter jejuni* colonization and by a non-antibiotic feed additive', *PLOS ONE*, 10(7), p. e0131978. Available at: https://doi.org/10.1371/journal.pone.0131978.

Thompson, C.N. *et al.* (2013) 'Epidemiological features and risk factors of *Salmonella* gastroenteritis in children resident in Ho Chi Minh City, Vietnam', *Epidemiology & Infection*, 141(8), pp. 1604–1613. Available at: https://doi.org/10.1017/S0950268812002014.

Thompson, C.N. *et al.* (2015) 'A prospective multi-center observational study of children hospitalized with diarrhea in Ho Chi Minh City, Vietnam', *The American Journal of Tropical Medicine and Hygiene*, 92(5), pp. 1045–1052. Available at: https://doi.org/10.4269/ajtmh.14-0655.

Tribble, D.R. *et al.* (2010) 'Assessment of the duration of protection in *Campylobacter jejuni* experimental infection in humans', *Infection and Immunity*, 78(4), pp. 1750–1759. Available at: https://doi.org/10.1128/IAI.01021-09.

Trifinopoulos, J. *et al.* (2016) 'W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis', *Nucleic Acids Research*, 44(W1), pp. W232-235. Available at: https://doi.org/10.1093/nar/gkw256.

Trigodet, F. *et al.* (2022) 'High molecular weight DNA extraction strategies for long-read sequencing of complex metagenomes', *Molecular Ecology Resources*, 22(5), pp. 1786–1802. Available at: https://doi.org/10.1111/1755-0998.13588.

Uelze, L. *et al.* (2020) 'Typing methods based on whole genome sequencing data', *One Health Outlook*, 2(1), p. 3. Available at: https://doi.org/10.1186/s42522-020-0010-1.

Ugarte-Ruiz, M. *et al.* (2012) 'Evaluation of four protocols for the detection and isolation of thermophilic *Campylobacter* from different matrices', *Journal of Applied Microbiology*, 113(1), pp. 200–208. Available at: https://doi.org/10.1111/j.1365-2672.2012.05323.x.

Ugarte-Ruiz, M. *et al.* (2018) 'Exploring the oxidative, antimicrobial and genomic properties of *Campylobacter jejuni* strains isolated from poultry', *Research in Veterinary Science*, 119, pp. 170–175. Available at: https://doi.org/10.1016/j.rvsc.2018.06.016.

Ugboko, H.U. *et al.* (2020) 'Childhood diarrhoeal diseases in developing countries', *Heliyon*, 6(4), p. e03690. Available at: https://doi.org/10.1016/j.heliyon.2020.e03690.

UK Health Security Agency (2024) Campylobacter *data 2013 to 2022*, *GOV.UK*. Available at: https://www.gov.uk/government/publications/campylobacter-infection-annual-data/campylobacter-data-2013-to-2022 (Accessed: 3 May 2024).

United States Department of Agriculture, Food Safety and Inspection Service (2022) 'MLG 41.07: Isolating and identifying *Campylobacter jejuni/coli/lari* from poultry rinsate, sponge and raw product samples'. Athens. Available at: https://www.fsis.usda.gov/sites/default/files/media_file/2022-03/MLG_41.07.pdf (Accessed: 10 May 2024).

Van Kinh, N. (2010) *Situation Analysis: Antibiotic use and resistance in Vietnam*. CDDEP. Available at: https://onehealthtrust.org/wp-content/uploads/2017/06/vn_report_web_1_8.pdf (Accessed: 19 April 2023).

Van, T.T.H. *et al.* (2012) 'The antibiotic resistance characteristics of non-typhoidal *Salmonella enterica* isolated from food-producing animals, retail meat and humans in South East Asia', *International Journal of Food Microbiology*, 154(3), pp. 98–106. Available at: https://doi.org/10.1016/j.ijfoodmicro.2011.12.032.

Van, T.T.H. *et al.* (2016) '*Campylobacter hepaticus* sp. nov., isolated from chickens with spotty liver disease', *International Journal of Systematic and Evolutionary Microbiology*, 66(11), pp. 4518–4524. Available at: https://doi.org/10.1099/ijsem.0.001383.

Van, T.T.H. *et al.* (2023) '*Campylobacter bilis,* the second novel *Campylobacter* species isolated from chickens with spotty liver disease, can cause the disease', *Veterinary Microbiology*, 276, p. 109603. Available at: https://doi.org/10.1016/j.vetmic.2022.109603.

Vandamme, P. *et al.* (2010) 'Reclassification of *Bacteroides ureolyticus* as *Campylobacter ureolyticus* comb. nov., and emended description of the genus *Campylobacter*', *International Journal of Systematic and Evolutionary Microbiology*, 60(9), pp. 2016–2022. Available at: https://doi.org/10.1099/ijs.0.017152-0.

Vegge, C.S. *et al.* (2016) 'Glucose metabolism via the Entner-Doudoroff pathway in *Campylobacter*: a rare trait that enhances survival and promotes biofilm formation in some isolates', *Frontiers in Microbiology*, 7, p. 1877. Available at: https://doi.org/10.3389/fmicb.2016.01877.

Vermeulen, S. (2019) *Fast and easy metagenomic data analysis with FastDeMe*. Available at: https://github.com/aldertzomer/FastDeMe/blob/master/Report%20Sander%20Vermeulen.pdf (Accessed: 8 February 2024).

Vorwerk, H. *et al.* (2014) 'Utilization of host-derived cysteine-containing peptides overcomes the restricted sulphur metabolism of *Campylobacter jejuni*', *Molecular Microbiology*, 93(6), pp. 1224–1245. Available at: https://doi.org/10.1111/mmi.12732.

Wadie, B. *et al.* (2021) '*in silico* characterization of toxin-antitoxin systems in *Campylobacter* isolates recovered from food sources and sporadic human illness', *Genes*, 12(1), pp. 1–16. Available at: https://doi.org/10.3390/genes12010072.

Wagenaar, J.A. *et al.* (2014) '*Campylobacter fetus* infections in humans: exposure and disease', *Clinical Infectious Diseases*, 58(11), pp. 1579–1586. Available at: https://doi.org/10.1093/cid/ciu085.

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), pp. 1600–1606. Available at: https://doi.org/10.1093/cid/cit555.

Walker, B.J. *et al.* (2014) 'Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement', *PLOS ONE*, 9(11), p. e112963. Available at: https://doi.org/10.1371/journal.pone.0112963.

Wallace, R.L. *et al.* (2021) '*Campylobacter jejuni* ST50, a pathogen of global importance: A comparative genomic analysis of isolates from Australia, Europe and North America', *Zoonoses and Public Health*, 68(6), pp. 638–649. Available at: https://doi.org/10.1111/zph.12853.

Wallen, Z.D. (2021) 'Comparison study of differential abundance testing methods using two large Parkinson disease gut microbiome datasets derived from 16S amplicon sequencing', *BMC Bioinformatics*, 22(1), p. 265. Available at: https://doi.org/10.1186/s12859-021-04193-6.

Walsh, C.J. *et al.* (2024) 'GROND: a quality-checked and publicly available database of full-length 16S-ITS-23S rRNA operon sequences', *Microbial Genomics*, 10(6), p. 001255. Available at: https://doi.org/10.1099/mgen.0.001255.

Wang, L.-G. *et al.* (2020) 'Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data', *Molecular Biology and Evolution*, 37(2), pp. 599–603. Available at: https://doi.org/10.1093/molbev/msz240.

Weerakoon, D.R. *et al.* (2009) 'The role of respiratory donor enzymes in *Campylobacter jejuni* host colonization and physiology', *Microbial Pathogenesis*, 47(1), pp. 8–15. Available at: https://doi.org/10.1016/j.micpath.2009.04.009.

Weiss, S. *et al.* (2017) 'Normalization and microbial differential abundance strategies depend upon data characteristics', *Microbiome*, 5(1), p. 27. Available at: https://doi.org/10.1186/s40168-017-0237-y.

Wen, S.C., Best, E. and Nourse, C. (2017) 'Non-typhoidal *Salmonella* infections in children: review of literature and recommendations for management', *Journal of Paediatrics and Child Health*, 53(10), pp. 936–941. Available at: https://doi.org/10.1111/jpc.13585.

Wenger, A.M. *et al.* (2019) 'Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome', *Nature Biotechnology*, 37(10), pp. 1155–1162. Available at: https://doi.org/10.1038/s41587-019-0217-9.

Wesolowska-Andersen, A. *et al.* (2014) 'Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis', *Microbiome*, 2(1), p. 19. Available at: https://doi.org/10.1186/2049-2618-2-19.

Wick, R.R. *et al.* (2017) 'Unicycler: resolving bacterial genome assemblies from short and long sequencing reads', *PLOS Computational Biology*, 13(6), p. e1005595. Available at: https://doi.org/10.1371/journal.pcbi.1005595.

Wickham, H. (2007) 'Reshaping data with the reshape package', *Journal of Statistical Software*, 21, pp. 1–20. Available at: https://doi.org/10.18637/jss.v021.i12.

Wickham, H. (2016) ‘ggplot2: elegant graphics for data analysis’. Springer-verlag New York.

Wickham, H. *et al.* (2019) ‘Welcome to the Tidyverse’, *Journal of Open Source Software*, 4(43), p. 1686. Available at: https://doi.org/10.21105/joss.01686.

Wickham, H. and Bryan, J. (2023) ‘readxl: read Excel files. R package version 1.4.2’. Available at: https://CRAN.R-project.org/package=readxl.

Wilke, C. (2022) ‘ggridges: ridgeline plots in ’ggplot2’_. R package version 0.5.4’. Available at: https://CRAN.R-project.org/package=ggridges.

Williams, L.K. *et al.* (2012) ‘Enrichment culture can bias the isolation of *Campylobacter* subtypes’, *Epidemiology & Infection*, 140(7), pp. 1227–1235. Available at: https://doi.org/10.1017/S0950268811001877.

Williams, M.S., Ebel, E.D. and Nyirabahizi, E. (2021) ‘Comparative history of *Campylobacter* contamination on chicken meat and campylobacteriosis cases in the United States: 1994–2018’, *International Journal of Food Microbiology*, 342, p. 109075. Available at: https://doi.org/10.1016/j.ijfoodmicro.2021.109075.

Wimalarathna, H.M. *et al.* (2013) ‘Widespread acquisition of antimicrobial resistance among *Campylobacter* isolates from UK retail poultry and evidence for clonal expansion of resistant lineages’, *BMC Microbiology*, 13(1), p. 160. Available at: https://doi.org/10.1186/1471-2180-13-160.

de Wit, M. a. S. *et al.* (2001) ‘Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology’, *American Journal of Epidemiology*, 154(7), pp. 666–674. Available at: https://doi.org/10.1093/aje/154.7.666.

Wood, D.E., Lu, J. and Langmead, B. (2019) ‘Improved metagenomic analysis with Kraken 2’, *Genome Biology*, 20(1), p. 257. Available at: https://doi.org/10.1186/s13059-019-1891-0.

Wu, Y.-W., Simmons, B.A. and Singer, S.W. (2016) ‘MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets’, *Bioinformatics*, 32(4), pp. 605–607. Available at: https://doi.org/10.1093/bioinformatics/btv638.

Würfel, S.F.R. *et al.* (2019) ‘Genetic diversity of *Campylobacter jejuni* and *Campylobacter coli* isolated from poultry meat products sold on the retail market in Southern Brazil’, *Poultry Science*, 98(2), pp. 932–939. Available at: https://doi.org/10.3382/ps/pey365.

Xu, S., Dai, Z., *et al.* (2021) ‘ggtreeExtra: compact visualization of richly annotated phylogenetic data’, *Molecular Biology and Evolution*, 38(9), pp. 4039–4042. Available at: https://doi.org/10.1093/molbev/msab166.

Xu, S., Chen, M., *et al.* (2021) ‘Use ggbreak to effectively utilize plotting space to deal with large datasets and outliers’, *Frontiers in Genetics*, 12. Available at: https://doi.org/10.3389/fgene.2021.774846.

Xu, S. (2022) ‘ggstar: multiple geometric shape point layer for “ggplot2”. R package version 1.0.4.’ Available at: https://CRAN.R-project.org/package=ggstar.

Yadav, R. *et al.* (2018) ‘Typing of *Campylobacter jejuni* isolated from poultry on the basis of flaA-RFLP by various restriction enzymes’, *Veterinary and Animal Science*, 6, pp. 1–5. Available at: https://doi.org/10.1016/j.vas.2018.06.003.

Yang, L. and Chen, J. (2022) ‘A comprehensive evaluation of microbial differential abundance analysis methods: current status and potential solutions’, *Microbiome*, 10(1), p. 130. Available at: https://doi.org/10.1186/s40168-022-01320-0.

Yap, M. *et al.* (2022) ‘Next-generation food research: use of meta-omic approaches for characterizing microbial communities along the food chain’, *Annual Review of Food Science and Technology*, 13(Volume 13, 2022), pp. 361–384. Available at: https://doi.org/10.1146/annurev-food-052720-010751.

Ye, S.H. *et al.* (2019) ‘Benchmarking metagenomics tools for taxonomic classification’, *Cell*, 178(4), pp. 779–794. Available at: https://doi.org/10.1016/j.cell.2019.07.010.

Yorki, S. *et al.* (2023) ‘Comparison of long- and short-read metagenomic assembly for low-abundance species and resistance genes’, *Briefings in Bioinformatics*, 24(2), p. bbad050. Available at: https://doi.org/10.1093/bib/bbad050.

Yoshida, C.E. *et al.* (2016) ‘The *Salmonella in silico* Typing Resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies’, *PLOS ONE*, 11(1), p. e0147101. Available at: https://doi.org/10.1371/journal.pone.0147101.

Yu, G. *et al.* (2017) ‘ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data’, *Methods in Ecology and Evolution*, 8(1), pp. 28–36. Available at: https://doi.org/10.1111/2041-210X.12628.

Zamudio, R. *et al.* (2024) ‘Global transmission of extended-spectrum cephalosporin resistance in *Escherichia coli* driven by epidemic plasmids’, *eBioMedicine*, 103, p. 105097. Available at: https://doi.org/10.1016/j.ebiom.2024.105097.

Zankari, E. *et al.* (2017) ‘PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens’, *Journal of Antimicrobial Chemotherapy*, 72(10), pp. 2764–2768. Available at: https://doi.org/10.1093/jac/dkx217.

Zhang, L. *et al.* (2020) ‘Molecular characterization and antibiotic resistant profiles of *Campylobacter* species isolated from poultry and diarrheal patients in southeastern China 2017-2019.’, *Frontiers in Microbiology*, 11, p. 1244. Available at: https://doi.org/10.3389/fmicb.2020.01244.

Zhang, Z. *et al.* (2000) ‘A greedy algorithm for aligning DNA sequences’, *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 7(1–2), pp. 203–214. Available at: https://doi.org/10.1089/10665270050081478.

Zhou, Y. *et al.* (2016) ‘Metagenomic approach for identification of the pathogens associated with diarrhea in stool specimens’, *Journal of Clinical Microbiology*, 54(2), pp. 368–375. Available at: https://doi.org/10.1128/jcm.01965-15.

Zilhao, R., Papadopoulou, B. and Courvalin, P. (1988) ‘Occurrence of the *Campylobacter* resistance gene *tetO* in *Enterococcus* and *Streptococcus* spp.’, *Antimicrobial Agents and Chemotherapy*, 32(12), pp. 1793–1796.

Zolfo, M. *et al.* (2017) ‘MetaMLST: multi-locus strain-level bacterial typing from metagenomic samples’, *Nucleic Acids Research*, 45(2), p. e7. Available at: https://doi.org/10.1093/nar/gkw837.

Appendices

Appendix 1: Retail chicken samples collected, and the chapters they are referenced in

Sample	Date Collected	Collection City	Commodity	Sample cut	Country of Origin	Chapter(s)
CH-0312	14/03/2021	Norwich	Chicken	Chicken drumstick (skin on)	United Kingdom	3, 4, 5
CH-0313	14/03/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0314	14/03/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0315	14/03/2021	Norwich	Chicken	Chicken thigh (skinless)	United Kingdom	3, 5
CH-0316	14/03/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 5
CH-0317	10/04/2021	Norwich	Chicken	Chicken thigh (skinless, boneless)	United Kingdom	3, 4, 5
CH-0318	10/04/2021	Norwich	Chicken	Chicken drumstick (skin on)	United Kingdom	3, 4, 5
CH-0319	11/04/2021	Norwich	Chicken	Chicken thigh (skinless, boneless)	United Kingdom	3, 4, 5
CH-0320	11/04/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 5
CH-0321	11/04/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0322	08/05/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 5
CH-0323	08/05/2021	Norwich	Chicken	Chicken thigh (skinless, boneless)	United Kingdom	3, 5
CH-0324	08/05/2021	Norwich	Chicken	Chicken drumstick (skin on)	United Kingdom	3, 5
CH-0325	08/05/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0326	08/05/2021	Norwich	Chicken	Chicken thigh (skin on)	United Kingdom	3, 4, 5
CH-0327	05/06/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0328	05/06/2021	Norwich	Chicken	Chicken thigh (skinless, boneless)	United Kingdom	3, 4, 5
CH-0329	05/06/2021	Norwich	Chicken	Chicken thigh (skinless, boneless)	United Kingdom	3, 4, 5
CH-0330	05/06/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0331	05/06/2021	Norwich	Chicken	Chicken drumstick (skin on)	United Kingdom	3, 4, 5
CH-0332	19/06/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0333	19/06/2021	Norwich	Chicken	Chicken drumstick (skin on)	United Kingdom	3, 4, 5
CH-0334	19/06/2021	Norwich	Chicken	Chicken thigh (skin on)	United Kingdom	3, 4, 5
CH-0335	19/06/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0336	19/06/2021	Norwich	Chicken	Chicken drumstick (skin on)	United Kingdom	3, 4, 5
CH-0337	13/08/2021	Norwich	Chicken	Chicken thigh (skin on)	United Kingdom	3, 4, 5
CH-0338	13/08/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0339	14/08/2021	Norwich	Chicken	Chicken thigh (skinless, boneless)	United Kingdom	3, 4, 5
CH-0340	14/08/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0341	14/08/2021	Norwich	Chicken	Chicken drumstick (skin on)	United Kingdom	3, 4, 5

Appendix 1: Retail chicken samples collected, and the chapters they are referenced in

Sample	Date Collected	Collection City	Commodity	Sample cut	Country of Origin	Chapter(s)
CH-0347	14/10/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 5
CH-0348	14/10/2021	Norwich	Chicken	Chicken drumstick (skin on)	United Kingdom	3, 5
CH-0349	14/10/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 5
CH-0350	15/10/2021	Norwich	Chicken	Chicken breast (boneless, skin on)	United Kingdom	3, 4, 5
CH-0351	15/10/2021	Norwich	Chicken	Chicken thigh (skinless, boneless)	United Kingdom	3, 4, 5
CH-0352	04/11/2021	Norwich	Chicken	Chicken thigh (skinless, boneless)	United Kingdom	3, 5
CH-0353	04/11/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 5
CH-0354	04/11/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 5
CH-0355	04/11/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0356	05/11/2021	Norwich	Chicken	Chicken thigh (skin on)	Poland	3, 5
CH-0357	26/11/2021	Norwich	Chicken	Chicken thigh (skin on)	Poland	3, 4, 5
CH-0358	26/11/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	Poland	3, 4, 5
CH-0359	26/11/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0360	26/11/2021	Norwich	Chicken	Chicken drumstick (skin on)	United Kingdom	3, 5
CH-0361	26/11/2021	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	3, 4, 5
CH-0362	17/02/2022	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	5
CH-0363	17/02/2022	Norwich	Chicken	Chicken thigh (skinless, boneless)	United Kingdom	5
CH-0364	17/02/2022	Norwich	Chicken	Chicken leg (skin on, bone in)	United Kingdom	5
CH-0365	18/02/2022	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	5
CH-0366	18/02/2022	Norwich	Chicken	Chicken thigh (skin on, bone in)	United Kingdom	5
CH-0367	18/02/2022	Norwich	Chicken	Chicken wings (skin on, bone in)	United Kingdom	5
CH-0368	18/02/2022	Norwich	Chicken	Chicken (whole)	United Kingdom	5
CH-0369	18/02/2022	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	5
CH-0370	18/02/2022	Norwich	Chicken	Chicken thigh & drumstick (skin on, bone in)*	United Kingdom	5
CH-0371	18/02/2022	Norwich	Chicken	Chicken thigh (skinless, boneless)	United Kingdom	5
CH-0372	18/02/2022	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	5
CH-0373	10/03/2022	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	5
CH-0374	10/03/2022	Norwich	Chicken	Chicken thigh & drumstick (skin on, bone in)**	United Kingdom	5
CH-0375	10/03/2022	Norwich	Chicken	Chicken (whole)	United Kingdom	5
CH-0376	10/03/2022	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	5
CH-0377	10/03/2022	Norwich	Chicken	Chicken (whole)	United Kingdom	5
CH-0378	11/03/2022	Norwich	Chicken	Chicken leg (skin on, bone in)	United Kingdom	5

Appendix 1: Retail chicken samples collected, and the chapters they are referenced in

Sample	Date Collected	Collection City	Commodity	Sample cut	Country of Origin	Chapter(s)
CH-0379	11/03/2022	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	5
CH-0380	11/03/2022	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	5
CH-0381	11/03/2022	Norwich	Chicken	Chicken thigh & drumstick (skin on, bone in)**	United Kingdom	5
CH-0382	11/03/2022	Norwich	Chicken	Chicken breast (boneless, skinless)	United Kingdom	5
CH-0383	11/03/2022	Norwich	Chicken	Chicken thigh & drumstick (skin on, bone in)*	United Kingdom	5

*sampled drumstick only; **sampled thigh only

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0312	CA21CH-0312-6-1	*C. jejuni*	45	42	CAT	mCCDA
CH-0313	CA19CH-0313-3-1	*C. jejuni*	257	37	CAT	u-mCCDA
CH-0313	CA21CH-0313-3-2	*C. jejuni*	5136	37	CAT	u-mCCDA
CH-0313	CA21CH-0313-3-3	*C. jejuni*	257	37	CAT	u-mCCDA
CH-0313	CA21CH-0313-3-4	*C. jejuni*	257	37	CAT	u-mCCDA
CH-0313	CA21CH-0313-3-5	*C. jejuni*	257	37	CAT	u-mCCDA
CH-0313	CA21CH-0313-3-6	*C. jejuni*	257	37	CAT	u-mCCDA
CH-0313	CA21CH-0313-6-1	*C. jejuni*	5136	42	CAT	mCCDA
CH-0313	CA21CH-0313-6-2	*C. jejuni*	257	42	CAT	mCCDA
CH-0313	CA21CH-0313-6-3	*C. jejuni*	5136	42	CAT	u-mCCDA
CH-0313	CA21CH-0313-6-4	*C. jejuni*	5136	42	CAT	u-mCCDA
CH-0313	CA21CH-0313-6-5	*C. jejuni*	257	42	CAT	u-mCCDA
CH-0313	CA21CH-0313-6-6	*C. jejuni*	5136	42	CAT	u-mCCDA
CH-0314	CA21CH-0314-3-2	*C. jejuni*	464	37	CAT	mCCDA
CH-0314	CA21CH-0314-3-3	*C. jejuni*	6175	37	CAT	mCCDA
CH-0314	CA21CH-0314-3-4	*C. jejuni*	6175	37	CAT	mCCDA
CH-0314	CA21CH-0314-6-1R2	*C. jejuni*	464	42	CAT	u-mCCDA
CH-0314	CA21CH-0314-6-2R2	*C. jejuni*	464	42	CAT	u-mCCDA
CH-0314	CA21CH-0314-6-3R2	*C. jejuni*	464	42	CAT	u-mCCDA
CH-0315	CA21CH-0315-1-1	*C. jejuni*	cj unknown6	37	None (Direct plating)	mCCDA
CH-0315	CA21CH-0315-1-2-R2	*C. jejuni*	cj unknown6	37	None (Direct plating)	mCCDA
CH-0317	CA21CH-0317-1-1	*C. jejuni*	400	37	None (Direct plating)	mCCDA
CH-0317	CA21CH-0317-1-2	*C. jejuni*	400	37	None (Direct plating)	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0317	CA21CH-0317-1-3	*C. jejuni*	400	37	None (Direct plating)	mCCDA
CH-0317	CA21CH-0317-1-4	*C. jejuni*	400	37	None (Direct plating)	mCCDA
CH-0317	CA21CH-0317-1-5	*C. jejuni*	400	37	None (Direct plating)	u-mCCDA
CH-0317	CA21CH-0317-1-6	*C. jejuni*	400	37	None (Direct plating)	u-mCCDA
CH-0317	CA21CH-0317-1-7	*C. jejuni*	400	37	None (Direct plating)	u-mCCDA
CH-0317	CA21CH-0317-2-1-R	*C. jejuni*	400	37	Bolton	mCCDA
CH-0317	CA21CH-0317-2-2-R	*C. jejuni*	400	37	Bolton	mCCDA
CH-0317	CA21CH-0317-2-3	*C. jejuni*	400	37	Bolton	mCCDA
CH-0317	CA21CH-0317-2-4	*C. jejuni*	400	37	Bolton	mCCDA
CH-0317	CA21CH-0317-2-5	*C. jejuni*	400	37	Bolton	u-mCCDA
CH-0317	CA21CH-0317-2-6-R	*C. jejuni*	400	37	Bolton	u-mCCDA
CH-0317	CA21CH-0317-2-7-R	*C. jejuni*	400	37	Bolton	u-mCCDA
CH-0317	CA21CH-0317-2-8	*C. jejuni*	400	37	Bolton	u-mCCDA
CH-0317	CA21CH-0317-3-1	*C. jejuni*	400	37	CAT	mCCDA
CH-0317	CA21CH-0317-3-2	*C. jejuni*	400	37	CAT	mCCDA
CH-0317	CA21CH-0317-3-3	*C. jejuni*	400	37	CAT	mCCDA
CH-0317	CA21CH-0317-3-4	*C. jejuni*	400	37	CAT	mCCDA
CH-0317	CA21CH-0317-3-5	*C. jejuni*	400	37	CAT	u-mCCDA
CH-0317	CA21CH-0317-3-6	*C. jejuni*	400	37	CAT	u-mCCDA
CH-0317	CA21CH-0317-3-7	*C. jejuni*	400	37	CAT	u-mCCDA
CH-0317	CA21CH-0317-3-8	*C. jejuni*	400	37	CAT	u-mCCDA
CH-0317	CA21CH-0317-4-1	*C. jejuni*	400	42	None (Direct plating)	mCCDA
CH-0317	CA21CH-0317-4-2	*C. jejuni*	400	42	None (Direct plating)	mCCDA
CH-0317	CA21CH-0317-4-3	*C. jejuni*	400	42	None (Direct plating)	mCCDA
CH-0317	CA21CH-0317-4-4	*C. jejuni*	400	42	None (Direct plating)	mCCDA
CH-0317	CA21CH-0317-4-5	*C. jejuni*	400	42	None (Direct plating)	u-mCCDA
CH-0317	CA21CH-0317-4-6	*C. jejuni*	400	42	None (Direct plating)	u-mCCDA
CH-0317	CA21CH-0317-5-1	*C. jejuni*	400	42	Bolton	mCCDA
CH-0317	CA21CH-0317-5-2	*C. jejuni*	400	42	Bolton	mCCDA
CH-0317	CA21CH-0317-5-3	*C. jejuni*	400	42	Bolton	mCCDA
CH-0317	CA21CH-0317-5-4	*C. jejuni*	400	42	Bolton	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0317	CA21CH-0317-5-5	*C. jejuni*	400	42	Bolton	u-mCCDA
CH-0317	CA21CH-0317-5-6	*C. jejuni*	400	42	Bolton	u-mCCDA
CH-0317	CA21CH-0317-5-7	*C. jejuni*	400	42	Bolton	u-mCCDA
CH-0317	CA21CH-0317-5-8	*C. jejuni*	400	42	Bolton	u-mCCDA
CH-0317	CA21CH-0317-6-1	*C. jejuni*	400	42	CAT	mCCDA
CH-0317	CA21CH-0317-6-2	*C. jejuni*	400	42	CAT	mCCDA
CH-0317	CA21CH-0317-6-3	*C. jejuni*	400	42	CAT	mCCDA
CH-0317	CA21CH-0317-6-4	*C. jejuni*	400	42	CAT	mCCDA
CH-0317	CA21CH-0317-6-5	*C. jejuni*	400	42	CAT	u-mCCDA
CH-0317	CA21CH-0317-6-6	*C. jejuni*	400	42	CAT	u-mCCDA
CH-0317	CA21CH-0317-6-7	*C. jejuni*	400	42	CAT	u-mCCDA
CH-0317	CA21CH-0317-6-8	*C. jejuni*	400	42	CAT	u-mCCDA
CH-0318	CA21CH-0318-3-1	*C. jejuni*	5136	37	CAT	mCCDA
CH-0318	CA21CH-0318-3-2	*C. jejuni*	5136	37	CAT	mCCDA
CH-0318	CA21CH-0318-3-3	*C. jejuni*	5136	37	CAT	mCCDA
CH-0318	CA21CH-0318-3-4	*C. jejuni*	5136	37	CAT	mCCDA
CH-0318	CA21CH-0318-4-1	*C. jejuni*	5136	42	None (Direct plating)	mCCDA
CH-0318	CA21CH-0318-4-2	*C. jejuni*	5136	42	None (Direct plating)	mCCDA
CH-0318	CA21CH-0318-4-3	*C. jejuni*	5136	42	None (Direct plating)	mCCDA
CH-0318	CA21CH-0318-4-4	*C. jejuni*	5136	42	None (Direct plating)	mCCDA
CH-0319	CA21CH-0319-3-1	*C. jejuni*	6175	37	CAT	mCCDA
CH-0319	CA21CH-0319-3-2	*C. jejuni*	6175	37	CAT	mCCDA
CH-0319	CA21CH-0319-3-3	*C. jejuni*	6175	37	CAT	mCCDA
CH-0319	CA21CH-0319-3-4	*C. jejuni*	6175	37	CAT	mCCDA
CH-0319	CA21CH-0319-3-5	*C. jejuni*	6175	37	CAT	u-mCCDA
CH-0319	CA21CH-0319-3-6-R	*C. jejuni*	6175	37	CAT	u-mCCDA
CH-0319	CA21CH-0319-3-7-R	*C. jejuni*	6175	37	CAT	u-mCCDA
CH-0319	CA21CH-0319-3-8	*C. jejuni*	6175	37	CAT	u-mCCDA
CH-0319	CA21CH-0319-6-1	*C. jejuni*	6175	42	CAT	mCCDA
CH-0319	CA21CH-0319-6-2	*C. jejuni*	122	42	CAT	mCCDA
CH-0319	CA21CH-0319-6-3	*C. jejuni*	6175	42	CAT	mCCDA
CH-0319	CA21CH-0319-6-4	*C. jejuni*	6175	42	CAT	mCCDA
CH-0319	CA21CH-0319-6-5	*C. jejuni*	6175	42	CAT	u-mCCDA
CH-0319	CA21CH-0319-6-6	*C. jejuni*	6175	42	CAT	u-mCCDA
CH-0319	CA21CH-0319-6-7	*C. jejuni*	6175	42	CAT	u-mCCDA
CH-0319	CA21CH-0319-6-8	*C. jejuni*	6175	42	CAT	u-mCCDA
CH-0320	CA21CH-0320-2-1-R	*C. lari*	27	37	Bolton	mCCDA
CH-0320	CA21CH-0320-2-2-R	*C. lari*	27	37	Bolton	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0320	CA21CH-0320-2-3-R	*C. lari*	27	37	Bolton	mCCDA
CH-0320	CA21CH-0320-2-4-R	*C. lari*	27	37	Bolton	mCCDA
CH-0320	CA21CH-0320-2-5-R	*C. lari*	27	37	Bolton	u-mCCDA
CH-0320	CA21CH-0320-2-6-R	*C. lari*	27	37	Bolton	u-mCCDA
CH-0320	CA21CH-0320-2-7-R	*C. lari*	27	37	Bolton	u-mCCDA
CH-0320	CA21CH-0320-2-8-R	*C. lari*	27	37	Bolton	u-mCCDA
CH-0320	CA21CH-0320-3-1	*C. lari*	27	37	CAT	mCCDA
CH-0320	CA21CH-0320-3-2	*C. lari*	27	37	CAT	mCCDA
CH-0320	CA21CH-0320-3-3	*C. lari*	27	37	CAT	mCCDA
CH-0320	CA21CH-0320-3-4	*C. lari*	27	37	CAT	mCCDA
CH-0320	CA21CH-0320-3-5	*C. lari*	27	37	CAT	u-mCCDA
CH-0320	CA21CH-0320-3-6	*C. lari*	27	37	CAT	u-mCCDA
CH-0320	CA21CH-0320-3-7	*C. lari*	27	37	CAT	u-mCCDA
CH-0320	CA21CH-0320-3-8	*C. lari*	27	37	CAT	u-mCCDA
CH-0320	CA21CH-0320-6-1	*C. jejuni*	464	42	CAT	mCCDA
CH-0320	CA21CH-0320-6-2	*C. jejuni*	464	42	CAT	mCCDA
CH-0320	CA21CH-0320-6-3	*C. jejuni*	464	42	CAT	mCCDA
CH-0320	CA21CH-0320-6-4	*C. jejuni*	464	42	CAT	mCCDA
CH-0320	CA21CH-0320-6-5	*C. jejuni*	464	42	CAT	u-mCCDA
CH-0320	CA21CH-0320-6-6	*C. jejuni*	464	42	CAT	u-mCCDA
CH-0320	CA21CH-0320-6-7	*C. jejuni*	464	42	CAT	u-mCCDA
CH-0320	CA21CH-0320-6-8-R	*C. jejuni*	464	42	CAT	u-mCCDA
CH-0321	CA21CH-0321-3-1	*C. jejuni*	5136	37	CAT	mCCDA
CH-0321	CA21CH-0321-3-2	*C. jejuni*	5136	37	CAT	mCCDA
CH-0321	CA21CH-0321-3-3	*C. jejuni*	5136	37	CAT	mCCDA
CH-0321	CA21CH-0321-3-4	*C. jejuni*	5136	37	CAT	mCCDA
CH-0321	CA21CH-0321-4-1	*C. jejuni*	5136	42	None (Direct plating)	mCCDA
CH-0321	CA21CH-0321-4-2	*C. jejuni*	5136	42	None (Direct plating)	mCCDA
CH-0321	CA21CH-0321-6-1	*C. jejuni*	6175	42	CAT	mCCDA
CH-0321	CA21CH-0321-6-2	*C. jejuni*	6175	42	CAT	mCCDA
CH-0321	CA21CH-0321-6-3	*C. jejuni*	6175	42	CAT	mCCDA
CH-0321	CA21CH-0321-6-4	*C. jejuni*	5136	42	CAT	mCCDA
CH-0321	CA21CH-0321-6-5	*C. jejuni*	6175	42	CAT	u-mCCDA
CH-0321	CA21CH-0321-6-6	*C. jejuni*	5136	42	CAT	u-mCCDA
CH-0321	CA21CH-0321-6-7	*C. jejuni*	5136	42	CAT	u-mCCDA
CH-0321	CA21CH-0321-6-8	*C. jejuni*	6175	42	CAT	u-mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0323	CA21CH-0323-6-1	*C. jejuni*	cj unknown7	42	CAT	mCCDA
CH-0323	CA21CH-0323-6-2	*C. jejuni*	cj unknown7	42	CAT	mCCDA
CH-0323	CA21CH-0323-6-3	*C. jejuni*	cj unknown7	42	CAT	mCCDA
CH-0323	CA21CH-0323-6-4	*C. jejuni*	cj unknown7	42	CAT	u-mCCDA
CH-0323	CA21CH-0323-6-5-R	*C. jejuni*	cj unknown7	42	CAT	u-mCCDA
CH-0323	CA21CH-0323-6-6	*C. jejuni*	cj unknown7	42	CAT	u-mCCDA
CH-0323	CA21CH-0323-6-7	*C. jejuni*	cj unknown7	42	CAT	u-mCCDA
CH-0325	CA21CH-0325-2-1	*C. jejuni*	21	37	Bolton	mCCDA
CH-0325	CA21CH-0325-3-1	*C. coli*	cc unknown4	37	CAT	mCCDA
CH-0325	CA21CH-0325-3-2	*C. jejuni*	21	37	CAT	mCCDA
CH-0325	CA21CH-0325-3-3	*C. jejuni*	21	37	CAT	mCCDA
CH-0325	CA21CH-0325-3-4	*C. jejuni*	21	37	CAT	mCCDA
CH-0325	CA21CH-0325-3-5-R	*C. jejuni*	21	37	CAT	u-mCCDA
CH-0325	CA21CH-0325-3-6	*C. jejuni*	21	37	CAT	u-mCCDA
CH-0325	CA21CH-0325-3-7	*C. jejuni*	21	37	CAT	u-mCCDA
CH-0325	CA21CH-0325-3-8	*C. jejuni*	21	37	CAT	u-mCCDA
CH-0325	CA21CH-0325-6-1	*C. coli*	1096	42	CAT	mCCDA
CH-0325	CA21CH-0325-6-2	*C. coli*	1096	42	CAT	u-mCCDA
CH-0325	CA21CH-0325-6-3	*C. jejuni*	21	42	CAT	u-mCCDA
CH-0325	CA21CH-0325-6-4	*C. jejuni*	21	42	CAT	u-mCCDA
CH-0325	CA21CH-0325-6-5	*C. jejuni*	21	42	CAT	u-mCCDA
CH-0325	CA21CH-0325-6-6	*C. jejuni*	21	42	CAT	u-mCCDA
CH-0325	CA21CH-0325-6-7	*C. coli*	1096	42	CAT	u-mCCDA
CH-0326	CA21CH-0326-3-1-R	*C. coli*	1595	37	CAT	mCCDA
CH-0326	CA21CH-0326-3-2	*C. coli*	1595	37	CAT	mCCDA
CH-0326	CA21CH-0326-3-3	*C. coli*	1595	37	CAT	mCCDA
CH-0326	CA21CH-0326-3-4	*C. coli*	1595	37	CAT	mCCDA
CH-0326	CA21CH-0326-3-5	*C. coli*	1595	37	CAT	mCCDA
CH-0326	CA21CH-0326-3-6	*C. coli*	1595	37	CAT	mCCDA
CH-0326	CA21CH-0326-3-7	*C. coli*	1595	37	CAT	mCCDA
CH-0326	CA21CH-0326-3-8	*C. coli*	1595	37	CAT	mCCDA
CH-0326	CA21CH-0326-5-1-R	*C. coli*	1595	42	Bolton	mCCDA
CH-0326	CA21CH-0326-5-2	*C. coli*	1595	42	Bolton	mCCDA
CH-0326	CA21CH-0326-5-3	*C. coli*	1595	42	Bolton	u-mCCDA
CH-0326	CA21CH-0326-5-4	*C. coli*	1595	42	Bolton	u-mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0326	CA21CH-0326-5-5	*C. coli*	1595	42	Bolton	u-mCCDA
CH-0326	CA21CH-0326-5-6	*C. coli*	1595	42	Bolton	u-mCCDA
CH-0326	CA21CH-0326-6-1	*C. coli*	1595	42	CAT	mCCDA
CH-0327	CA21CH-0327-5-1	*C. coli*	1096	42	Bolton	mCCDA
CH-0327	CA21CH-0327-5-2	*C. coli*	1096	42	Bolton	mCCDA
CH-0327	CA21CH-0327-5-3	*C. coli*	1096	42	Bolton	mCCDA
CH-0327	CA21CH-0327-5-4	*C. coli*	1096	42	Bolton	mCCDA
CH-0327	CA21CH-0327-5-5	*C. coli*	1096	42	Bolton	u-mCCDA
CH-0327	CA21CH-0327-5-6	*C. coli*	1096	42	Bolton	u-mCCDA
CH-0327	CA21CH-0327-5-7	*C. coli*	1096	42	Bolton	u-mCCDA
CH-0327	CA21CH-0327-5-8	*C. coli*	1096	42	Bolton	u-mCCDA
CH-0327	CA21CH-0327-6-1	*C. jejuni*	2282	42	CAT	mCCDA
CH-0327	CA21CH-0327-6-2	*C. jejuni*	2282	42	CAT	mCCDA
CH-0327	CA21CH-0327-6-3	*C. jejuni*	2282	42	CAT	mCCDA
CH-0327	CA21CH-0327-6-4	*C. jejuni*	2282	42	CAT	u-mCCDA
CH-0327	CA21CH-0327-6-5	*C. jejuni*	2282	42	CAT	u-mCCDA
CH-0327	CA21CH-0327-6-6	*C. jejuni*	2282	42	CAT	u-mCCDA
CH-0327	CA21CH-0327-6-7	*C. jejuni*	2282	42	CAT	u-mCCDA
CH-0328	CA21CH-0328-3-1	*C. jejuni*	cj unknown4	37	CAT	mCCDA
CH-0328	CA21CH-0328-3-2	*C. jejuni*	cj unknown4	37	CAT	mCCDA
CH-0328	CA21CH-0328-3-3	*C. jejuni*	cj unknown4	37	CAT	mCCDA
CH-0328	CA21CH-0328-3-4	*C. jejuni*	9401	37	CAT	mCCDA
CH-0328	CA21CH-0328-3-5	*C. jejuni*	cj unknown4	37	CAT	mCCDA
CH-0328	CA21CH-0328-3-6	*C. jejuni*	cj unknown4	37	CAT	mCCDA
CH-0328	CA21CH-0328-3-7	*C. jejuni*	cj unknown4	37	CAT	mCCDA
CH-0328	CA21CH-0328-3-8	*C. jejuni*	814	37	CAT	mCCDA
CH-0328	CA21CH-0328-5-1	*C. jejuni*	814	42	Bolton	mCCDA
CH-0328	CA21CH-0328-5-2	*C. jejuni*	814	42	Bolton	mCCDA
CH-0328	CA21CH-0328-5-3	*C. jejuni*	814	42	Bolton	mCCDA
CH-0328	CA21CH-0328-5-4	*C. jejuni*	814	42	Bolton	mCCDA
CH-0328	CA21CH-0328-5-5	*C. jejuni*	814	42	Bolton	u-mCCDA
CH-0328	CA21CH-0328-5-6	*C. jejuni*	814	42	Bolton	u-mCCDA
CH-0328	CA21CH-0328-5-7	*C. jejuni*	814	42	Bolton	u-mCCDA
CH-0328	CA21CH-0328-5-8	*C. jejuni*	814	42	Bolton	u-mCCDA
CH-0328	CA21CH-0328-6-1	*C. jejuni*	814	42	CAT	mCCDA
CH-0329	CA21CH-0329-2-1	*C. coli*	829	37	Bolton	mCCDA
CH-0329	CA21CH-0329-2-2	*C. coli*	829	37	Bolton	mCCDA
CH-0329	CA21CH-0329-2-3	*C. coli*	829	37	Bolton	mCCDA
CH-0329	CA21CH-0329-2-4	*C. coli*	829	37	Bolton	mCCDA
CH-0329	CA21CH-0329-2-5	*C. coli*	829	37	Bolton	u-mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0329	CA21CH-0329-2-6	*C. coli*	829	37	Bolton	u-mCCDA
CH-0329	CA21CH-0329-2-7	*C. coli*	829	37	Bolton	u-mCCDA
CH-0329	CA21CH-0329-2-8	*C. coli*	829	37	Bolton	u-mCCDA
CH-0329	CA21CH-0329-3-1	*C. coli*	829	37	CAT	mCCDA
CH-0329	CA21CH-0329-3-2	*C. coli*	829	37	CAT	mCCDA
CH-0329	CA21CH-0329-3-3	*C. coli*	829	37	CAT	mCCDA
CH-0329	CA21CH-0329-3-4	*C. coli*	828	37	CAT	mCCDA
CH-0329	CA21CH-0329-3-5	*C. coli*	829	37	CAT	u-mCCDA
CH-0329	CA21CH-0329-3-6	*C. coli*	829	37	CAT	u-mCCDA
CH-0329	CA21CH-0329-3-7	*C. coli*	829	37	CAT	u-mCCDA
CH-0329	CA21CH-0329-3-8	*C. coli*	829	37	CAT	u-mCCDA
CH-0329	CA21CH-0329-5-1	*C. coli*	829	42	Bolton	mCCDA
CH-0329	CA21CH-0329-5-2	*C. coli*	829	42	Bolton	mCCDA
CH-0329	CA21CH-0329-5-3	*C. coli*	829	42	Bolton	mCCDA
CH-0329	CA21CH-0329-5-4	*C. coli*	829	42	Bolton	mCCDA
CH-0329	CA21CH-0329-5-5	*C. coli*	829	42	Bolton	u-mCCDA
CH-0329	CA21CH-0329-5-6	*C. coli*	829	42	Bolton	u-mCCDA
CH-0329	CA21CH-0329-5-7	*C. coli*	829	42	Bolton	u-mCCDA
CH-0329	CA21CH-0329-5-8	*C. coli*	829	42	Bolton	u-mCCDA
CH-0329	CA21CH-0329-6-1	*C. coli*	829	42	CAT	mCCDA
CH-0329	CA21CH-0329-6-2	*C. coli*	829	42	CAT	mCCDA
CH-0329	CA21CH-0329-6-3	*C. coli*	829	42	CAT	mCCDA
CH-0329	CA21CH-0329-6-4	*C. coli*	829	42	CAT	mCCDA
CH-0329	CA21CH-0329-6-5	*C. coli*	829	42	CAT	u-mCCDA
CH-0329	CA21CH-0329-6-6	*C. coli*	829	42	CAT	u-mCCDA
CH-0329	CA21CH-0329-6-7	*C. coli*	829	42	CAT	u-mCCDA
CH-0329	CA21CH-0329-6-8	*C. coli*	829	42	CAT	u-mCCDA
CH-0330	CA21CH-0330-2-1	*C. coli*	1191	37	Bolton	mCCDA
CH-0330	CA21CH-0330-2-2	*C. coli*	1191	37	Bolton	mCCDA
CH-0330	CA21CH-0330-2-3	*C. coli*	1191	37	Bolton	mCCDA
CH-0330	CA21CH-0330-2-4	*C. coli*	1191	37	Bolton	mCCDA
CH-0330	CA21CH-0330-2-5	*C. coli*	1191	37	Bolton	u-mCCDA
CH-0330	CA21CH-0330-2-6	*C. coli*	1191	37	Bolton	u-mCCDA
CH-0330	CA21CH-0330-2-7	*C. coli*	1191	37	Bolton	u-mCCDA
CH-0330	CA21CH-0330-2-8	*C. coli*	1191	37	Bolton	u-mCCDA
CH-0330	CA21CH-0330-5-1	*C. coli*	1191	42	Bolton	mCCDA
CH-0330	CA21CH-0330-5-2	*C. coli*	1191	42	Bolton	mCCDA
CH-0330	CA21CH-0330-5-3	*C. coli*	1191	42	Bolton	mCCDA
CH-0330	CA21CH-0330-5-4	*C. coli*	1191	42	Bolton	mCCDA
CH-0330	CA21CH-0330-5-5	*C. coli*	1191	42	Bolton	u-mCCDA
CH-0330	CA21CH-0330-5-6	*C. coli*	1191	42	Bolton	u-mCCDA
CH-0330	CA21CH-0330-5-7	*C. coli*	1191	42	Bolton	u-mCCDA
CH-0330	CA21CH-0330-5-8	*C. coli*	1191	42	Bolton	u-mCCDA
CH-0330	CA21CH-0330-6-1	*C. jejuni*	441	42	CAT	mCCDA
CH-0330	CA21CH-0330-6-2	*C. coli*	1191	42	CAT	mCCDA
CH-0330	CA21CH-0330-6-3	*C. jejuni*	441	42	CAT	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0330	CA21CH-0330-6-4	*C. jejuni*	441	42	CAT	mCCDA
CH-0330	CA21CH-0330-6-5	*C. coli*	830	42	CAT	u-mCCDA
CH-0330	CA21CH-0330-6-6	*C. coli*	830	42	CAT	u-mCCDA
CH-0330	CA21CH-0330-6-7	*C. coli*	830	42	CAT	u-mCCDA
CH-0330	CA21CH-0330-6-8	*C. jejuni*	441	42	CAT	u-mCCDA
CH-0331	CA21CH-0331-2-1	*C. jejuni*	5136	37	Bolton	mCCDA
CH-0331	CA21CH-0331-2-2	*C. jejuni*	5136	37	Bolton	mCCDA
CH-0331	CA21CH-0331-2-3	*C. jejuni*	5136	37	Bolton	mCCDA
CH-0331	CA21CH-0331-2-4	*C. jejuni*	5136	37	Bolton	mCCDA
CH-0331	CA21CH-0331-2-5	*C. jejuni*	5136	37	Bolton	u-mCCDA
CH-0331	CA21CH-0331-2-6	*C. jejuni*	5136	37	Bolton	u-mCCDA
CH-0331	CA21CH-0331-2-7	*C. jejuni*	5136	37	Bolton	u-mCCDA
CH-0331	CA21CH-0331-2-8	*C. jejuni*	5136	37	Bolton	u-mCCDA
CH-0331	CA21CH-0331-5-1	*C. jejuni*	6175	42	Bolton	mCCDA
CH-0331	CA21CH-0331-5-2	*C. jejuni*	6175	42	Bolton	mCCDA
CH-0331	CA21CH-0331-5-3	*C. jejuni*	6175	42	Bolton	mCCDA
CH-0331	CA21CH-0331-5-4	*C. jejuni*	6175	42	Bolton	mCCDA
CH-0331	CA21CH-0331-5-5	*C. jejuni*	6175	42	Bolton	u-mCCDA
CH-0331	CA21CH-0331-5-6	*C. jejuni*	6175	42	Bolton	u-mCCDA
CH-0331	CA21CH-0331-5-7	*C. jejuni*	6175	42	Bolton	u-mCCDA
CH-0331	CA21CH-0331-5-8	*C. jejuni*	6175	42	Bolton	u-mCCDA
CH-0331	CA21CH-0331-6-1	*C. jejuni*	2282	42	CAT	mCCDA
CH-0331	CA21CH-0331-6-2	*C. jejuni*	2282	42	CAT	mCCDA
CH-0331	CA21CH-0331-6-3	*C. jejuni*	2282	42	CAT	mCCDA
CH-0331	CA21CH-0331-6-4	*C. jejuni*	2282	42	CAT	mCCDA
CH-0331	CA21CH-0331-6-5	*C. jejuni*	5136	42	CAT	u-mCCDA
CH-0331	CA21CH-0331-6-6	*C. jejuni*	2282	42	CAT	u-mCCDA
CH-0331	CA21CH-0331-6-7	*C. jejuni*	2282	42	CAT	u-mCCDA
CH-0331	CA21CH-0331-6-8	*C. jejuni*	2282	42	CAT	u-mCCDA
CH-0332	CA21CH-0332-3-1	*C. jejuni*	51	37	CAT	mCCDA
CH-0332	CA21CH-0332-3-2	*C. jejuni*	51	37	CAT	mCCDA
CH-0332	CA21CH-0332-3-3	*C. jejuni*	51	37	CAT	mCCDA
CH-0332	CA21CH-0332-3-4	*C. jejuni*	51	37	CAT	mCCDA
CH-0332	CA21CH-0332-3-5	*C. jejuni*	51	37	CAT	u-mCCDA
CH-0332	CA21CH-0332-3-6	*C. jejuni*	51	37	CAT	u-mCCDA
CH-0332	CA21CH-0332-3-7	*C. jejuni*	51	37	CAT	u-mCCDA
CH-0332	CA21CH-0332-3-8	*C. jejuni*	51	37	CAT	u-mCCDA
CH-0332	CA21CH-0332-6-1	*C. jejuni*	61	42	CAT	mCCDA
CH-0332	CA21CH-0332-6-2	*C. jejuni*	61	42	CAT	mCCDA
CH-0332	CA21CH-0332-6-3	*C. jejuni*	61	42	CAT	mCCDA
CH-0332	CA21CH-0332-6-4	*C. jejuni*	61	42	CAT	mCCDA
CH-0332	CA21CH-0332-6-5	*C. jejuni*	61	42	CAT	u-mCCDA
CH-0332	CA21CH-0332-6-6	*C. jejuni*	61	42	CAT	u-mCCDA
CH-0332	CA21CH-0332-6-7	*C. jejuni*	61	42	CAT	u-mCCDA
CH-0332	CA21CH-0332-6-8	*C. jejuni*	61	42	CAT	u-mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0333	CA21CH-0333-1-1	*C. jejuni*	cj unknown1	37	None (Direct plating)	mCCDA
CH-0333	CA21CH-0333-1-2	*C. jejuni*	48	37	None (Direct plating)	mCCDA
CH-0333	CA21CH-0333-1-3	*C. jejuni*	cj unknown2	37	None (Direct plating)	mCCDA
CH-0333	CA21CH-0333-1-4	*C. jejuni*	53	37	None (Direct plating)	mCCDA
CH-0333	CA21CH-0333-1-5	*C. jejuni*	cj unknown1	37	None (Direct plating)	u-mCCDA
CH-0333	CA21CH-0333-1-6	*C. jejuni*	cj unknown2	37	None (Direct plating)	u-mCCDA
CH-0333	CA21CH-0333-2-1	*C. jejuni*	267	37	Bolton	mCCDA
CH-0333	CA21CH-0333-2-2	*C. jejuni*	267	37	Bolton	mCCDA
CH-0333	CA21CH-0333-2-3	*C. jejuni*	267	37	Bolton	mCCDA
CH-0333	CA21CH-0333-2-4	*C. jejuni*	267	37	Bolton	mCCDA
CH-0333	CA21CH-0333-2-5	*C. jejuni*	267	37	Bolton	u-mCCDA
CH-0333	CA21CH-0333-2-6	*C. jejuni*	267	37	Bolton	u-mCCDA
CH-0333	CA21CH-0333-2-7	*C. jejuni*	267	37	Bolton	u-mCCDA
CH-0333	CA21CH-0333-2-8	*C. jejuni*	267	37	Bolton	u-mCCDA
CH-0333	CA21CH-0333-4-1	*C. jejuni*	cj unknown3	42	None (Direct plating)	mCCDA
CH-0333	CA21CH-0333-4-2	*C. jejuni*	5136	42	None (Direct plating)	mCCDA
CH-0333	CA21CH-0333-5-1	*C. jejuni*	267	42	Bolton	mCCDA
CH-0333	CA21CH-0333-5-2	*C. jejuni*	267	42	Bolton	mCCDA
CH-0333	CA21CH-0333-5-3	*C. jejuni*	267	42	Bolton	mCCDA
CH-0333	CA21CH-0333-5-4	*C. jejuni*	267	42	Bolton	mCCDA
CH-0333	CA21CH-0333-5-5	*C. coli*	9012	42	Bolton	u-mCCDA
CH-0333	CA21CH-0333-5-6	*C. jejuni*	267	42	Bolton	u-mCCDA
CH-0333	CA21CH-0333-6-1	*C. coli*	9012	42	CAT	mCCDA
CH-0333	CA21CH-0333-6-2	*C. jejuni*	267	42	CAT	u-mCCDA
CH-0333	CA21CH-0333-6-3	*C. jejuni*	267	42	CAT	u-mCCDA
CH-0334	CA21CH-0334-2-1	*C. jejuni*	53	37	Bolton	mCCDA
CH-0334	CA21CH-0334-2-2	*C. jejuni*	53	37	Bolton	mCCDA
CH-0334	CA21CH-0334-2-3	*C. jejuni*	53	37	Bolton	mCCDA
CH-0334	CA21CH-0334-2-4	*C. jejuni*	53	37	Bolton	mCCDA
CH-0334	CA21CH-0334-2-5	*C. jejuni*	53	37	Bolton	u-mCCDA
CH-0334	CA21CH-0334-2-6	*C. jejuni*	53	37	Bolton	u-mCCDA
CH-0334	CA21CH-0334-2-7	*C. jejuni*	53	37	Bolton	u-mCCDA
CH-0334	CA21CH-0334-2-8	*C. jejuni*	53	37	Bolton	u-mCCDA
CH-0334	CA21CH-0334-4-1	*C. jejuni*	cj unknown2	42	None (Direct plating)	mCCDA
CH-0334	CA21CH-0334-4-2	*C. jejuni*	53	42	None (Direct plating)	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0334	CA21CH-0334-4-3	*C. jejuni*	257	42	None (Direct plating)	mCCDA
CH-0334	CA21CH-0334-5-1	*C. jejuni*	53	42	Bolton	mCCDA
CH-0334	CA21CH-0334-5-2	*C. jejuni*	230	42	Bolton	mCCDA
CH-0334	CA21CH-0334-5-3	*C. jejuni*	230	42	Bolton	mCCDA
CH-0334	CA21CH-0334-5-4	*C. jejuni*	53	42	Bolton	mCCDA
CH-0334	CA21CH-0334-5-5	*C. jejuni*	230	42	Bolton	u-mCCDA
CH-0334	CA21CH-0334-5-6	*C. jejuni*	230	42	Bolton	u-mCCDA
CH-0334	CA21CH-0334-5-7	*C. jejuni*	53	42	Bolton	u-mCCDA
CH-0334	CA21CH-0334-5-8	*C. jejuni*	53	42	Bolton	u-mCCDA
CH-0334	CA21CH-0334-6-1	*C. jejuni*	53	42	CAT	mCCDA
CH-0334	CA21CH-0334-6-2	*C. jejuni*	53	42	CAT	mCCDA
CH-0334	CA21CH-0334-6-3	*C. jejuni*	53	42	CAT	mCCDA
CH-0334	CA21CH-0334-6-4	*C. jejuni*	53	42	CAT	mCCDA
CH-0334	CA21CH-0334-6-5	*C. jejuni*	449	42	CAT	u-mCCDA
CH-0335	CA21CH-0335-1-1	*C. jejuni*	21	37	None (Direct plating)	mCCDA
CH-0335	CA21CH-0335-1-2	*C. jejuni*	6175	37	None (Direct plating)	mCCDA
CH-0335	CA21CH-0335-2-1	*C. coli*	cc unknown1	37	Bolton	mCCDA
CH-0335	CA21CH-0335-2-2	*C. coli*	cc unknown1	37	Bolton	mCCDA
CH-0335	CA21CH-0335-2-3	*C. coli*	cc unknown1	37	Bolton	mCCDA
CH-0335	CA21CH-0335-2-4	*C. jejuni*	227	37	Bolton	mCCDA
CH-0335	CA21CH-0335-2-5	*C. coli*	cc unknown1	37	Bolton	u-mCCDA
CH-0335	CA21CH-0335-2-6	*C. coli*	cc unknown1	37	Bolton	u-mCCDA
CH-0335	CA21CH-0335-2-7	*C. coli*	cc unknown1	37	Bolton	u-mCCDA
CH-0335	CA21CH-0335-2-8	*C. coli*	cc unknown1	37	Bolton	u-mCCDA
CH-0335	CA21CH-0335-3-1	*C. coli*	cc unknown1	37	CAT	mCCDA
CH-0335	CA21CH-0335-3-2	*C. jejuni*	227	37	CAT	mCCDA
CH-0335	CA21CH-0335-3-3	*C. coli*	cc unknown1	37	CAT	mCCDA
CH-0335	CA21CH-0335-3-4	*C. coli*	cc unknown1	37	CAT	mCCDA
CH-0335	CA21CH-0335-3-5	*C. coli*	cc unknown1	37	CAT	u-mCCDA
CH-0335	CA21CH-0335-3-6	*C. coli*	cc unknown1	37	CAT	u-mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0335	CA21CH-0335-3-7	*C. coli*	cc unknown1	37	CAT	u-mCCDA
CH-0335	CA21CH-0335-3-8	*C. coli*	cc unknown1	37	CAT	u-mCCDA
CH-0335	CA21CH-0335-4-1	*C. jejuni*	2254	42	None (Direct plating)	mCCDA
CH-0335	CA21CH-0335-5-1	*C. coli*	cc unknown1	42	Bolton	mCCDA
CH-0335	CA21CH-0335-5-2	*C. coli*	cc unknown1	42	Bolton	mCCDA
CH-0335	CA21CH-0335-5-3	*C. coli*	cc unknown1	42	Bolton	mCCDA
CH-0335	CA21CH-0335-5-4	*C. coli*	cc unknown1	42	Bolton	mCCDA
CH-0335	CA21CH-0335-5-5	*C. coli*	cc unknown1	42	Bolton	u-mCCDA
CH-0335	CA21CH-0335-5-6	*C. coli*	cc unknown1	42	Bolton	u-mCCDA
CH-0335	CA21CH-0335-5-7	*C. coli*	cc unknown1	42	Bolton	u-mCCDA
CH-0335	CA21CH-0335-5-8	*C. coli*	cc unknown1	42	Bolton	u-mCCDA
CH-0335	CA21CH-0335-6-1	*C. coli*	cc unknown1	42	CAT	mCCDA
CH-0335	CA21CH-0335-6-2	*C. coli*	cc unknown1	42	CAT	mCCDA
CH-0335	CA21CH-0335-6-3	*C. coli*	cc unknown1	42	CAT	mCCDA
CH-0335	CA21CH-0335-6-4	*C. coli*	cc unknown1	42	CAT	mCCDA
CH-0335	CA21CH-0335-6-5	*C. coli*	cc unknown1	42	CAT	u-mCCDA
CH-0335	CA21CH-0335-6-6	*C. coli*	cc unknown1	42	CAT	u-mCCDA
CH-0335	CA21CH-0335-6-7	*C. coli*	cc unknown1	42	CAT	u-mCCDA
CH-0335	CA21CH-0335-6-8	*C. jejuni*	227	42	CAT	u-mCCDA
CH-0336	CA21CH-0336-2-1	*C. jejuni*	7743	37	Bolton	mCCDA
CH-0336	CA21CH-0336-2-2	*C. jejuni*	6175	37	Bolton	mCCDA
CH-0336	CA21CH-0336-2-3	*C. jejuni*	7743	37	Bolton	mCCDA
CH-0336	CA21CH-0336-2-4	*C. jejuni*	6175	37	Bolton	mCCDA
CH-0336	CA21CH-0336-2-5	*C. jejuni*	7743	37	Bolton	u-mCCDA
CH-0336	CA21CH-0336-2-6	*C. jejuni*	6175	37	Bolton	u-mCCDA
CH-0336	CA21CH-0336-2-7	*C. jejuni*	7743	37	Bolton	u-mCCDA
CH-0336	CA21CH-0336-2-8	*C. jejuni*	7743	37	Bolton	u-mCCDA
CH-0336	CA21CH-0336-3-1	*C. jejuni*	5136	37	CAT	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0336	CA21CH-0336-3-2	*C. jejuni*	6175	37	CAT	mCCDA
CH-0336	CA21CH-0336-3-3	*C. jejuni*	6175	37	CAT	mCCDA
CH-0336	CA21CH-0336-3-4	*C. jejuni*	6175	37	CAT	mCCDA
CH-0336	CA21CH-0336-3-5	*C. jejuni*	6175	37	CAT	u-mCCDA
CH-0336	CA21CH-0336-3-6	*C. jejuni*	6175	37	CAT	u-mCCDA
CH-0336	CA21CH-0336-3-7	*C. jejuni*	6175	37	CAT	u-mCCDA
CH-0336	CA21CH-0336-3-8	*C. jejuni*	6175	37	CAT	u-mCCDA
CH-0336	CA21CH-0336-5-1	*C. jejuni*	230	42	Bolton	mCCDA
CH-0336	CA21CH-0336-5-2	*C. jejuni*	230	42	Bolton	mCCDA
CH-0336	CA21CH-0336-5-3	*C. jejuni*	230	42	Bolton	mCCDA
CH-0336	CA21CH-0336-5-4	*C. jejuni*	230	42	Bolton	mCCDA
CH-0336	CA21CH-0336-5-5	*C. jejuni*	230	42	Bolton	u-mCCDA
CH-0336	CA21CH-0336-5-6	*C. jejuni*	230	42	Bolton	u-mCCDA
CH-0336	CA21CH-0336-5-7	*C. jejuni*	230	42	Bolton	u-mCCDA
CH-0336	CA21CH-0336-6-1	*C. jejuni*	6175	42	CAT	mCCDA
CH-0336	CA21CH-0336-6-2	*C. jejuni*	400	42	CAT	mCCDA
CH-0336	CA21CH-0336-6-3	*C. jejuni*	400	42	CAT	mCCDA
CH-0336	CA21CH-0336-6-4	*C. jejuni*	400	42	CAT	mCCDA
CH-0336	CA21CH-0336-6-5	*C. jejuni*	5136	42	CAT	u-mCCDA
CH-0336	CA21CH-0336-6-6	*C. jejuni*	400	42	CAT	u-mCCDA
CH-0336	CA21CH-0336-6-7	*C. jejuni*	400	42	CAT	u-mCCDA
CH-0336	CA21CH-0336-6-8	*C. jejuni*	5136	42	CAT	u-mCCDA
CH-0337	CA21CH-0337-2-1	*C. jejuni*	2066	37	Bolton	mCCDA
CH-0337	CA21CH-0337-2-2	*C. jejuni*	2211	37	Bolton	mCCDA
CH-0337	CA21CH-0337-2-3	*C. jejuni*	2066	37	Bolton	mCCDA
CH-0337	CA21CH-0337-2-4	*C. jejuni*	2066	37	Bolton	mCCDA
CH-0337	CA21CH-0337-2-5	*C. jejuni*	2066	37	Bolton	u-mCCDA
CH-0337	CA21CH-0337-2-6	*C. jejuni*	2066	37	Bolton	u-mCCDA
CH-0337	CA21CH-0337-2-7	*C. jejuni*	2211	37	Bolton	u-mCCDA
CH-0337	CA21CH-0337-2-8	*C. jejuni*	2211	37	Bolton	u-mCCDA
CH-0337	CA21CH-0337-4-1	*C. jejuni*	2066	42	None (Direct plating)	u-mCCDA
CH-0337	CA21CH-0337-5-1	*C. jejuni*	51	42	Bolton	mCCDA
CH-0337	CA21CH-0337-5-2	*C. jejuni*	2066	42	Bolton	mCCDA
CH-0337	CA21CH-0337-5-3	*C. jejuni*	400	42	Bolton	mCCDA
CH-0337	CA21CH-0337-5-4	*C. jejuni*	2066	42	Bolton	mCCDA
CH-0337	CA21CH-0337-5-5	*C. jejuni*	2066	42	Bolton	u-mCCDA
CH-0337	CA21CH-0337-5-6	*C. jejuni*	2066	42	Bolton	u-mCCDA
CH-0337	CA21CH-0337-5-7	*C. jejuni*	2066	42	Bolton	u-mCCDA
CH-0337	CA21CH-0337-5-8	*C. jejuni*	2066	42	Bolton	u-mCCDA
CH-0337	CA21CH-0337-6-1	*C. jejuni*	2066	42	CAT	mCCDA
CH-0337	CA21CH-0337-6-2	*C. jejuni*	2066	42	CAT	u-mCCDA
CH-0337	CA21CH-0337-6-3	*C. jejuni*	2066	42	CAT	u-mCCDA
CH-0337	CA21CH-0337-6-4	*C. jejuni*	2066	42	CAT	u-mCCDA
CH-0338	CA21CH-0338-2-1	*C. jejuni*	6876	37	Bolton	mCCDA
CH-0338	CA21CH-0338-2-2	*C. jejuni*	6876	37	Bolton	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0338	CA21CH-0338-2-3	*C. jejuni*	50	37	Bolton	mCCDA
CH-0338	CA21CH-0338-2-4	*C. coli*	cc unknown3	37	Bolton	mCCDA
CH-0338	CA21CH-0338-2-5	*C. jejuni*	6876	37	Bolton	u-mCCDA
CH-0338	CA21CH-0338-2-6	*C. jejuni*	6876	37	Bolton	u-mCCDA
CH-0338	CA21CH-0338-2-7	*C. jejuni*	6876	37	Bolton	u-mCCDA
CH-0338	CA21CH-0338-2-8	*C. jejuni*	6876	37	Bolton	u-mCCDA
CH-0338	CA21CH-0338-3-1	*C. coli*	828	37	CAT	mCCDA
CH-0338	CA21CH-0338-3-2	*C. coli*	cc unknown3	37	CAT	mCCDA
CH-0338	CA21CH-0338-3-3	*C. coli*	828	37	CAT	mCCDA
CH-0338	CA21CH-0338-3-4	*C. jejuni*	6175	37	CAT	mCCDA
CH-0338	CA21CH-0338-3-5	*C. coli*	cc unknown3	37	CAT	u-mCCDA
CH-0338	CA21CH-0338-3-6	*C. jejuni*	6175	37	CAT	u-mCCDA
CH-0338	CA21CH-0338-3-7	*C. coli*	828	37	CAT	u-mCCDA
CH-0338	CA21CH-0338-3-8	*C. coli*	828	37	CAT	u-mCCDA
CH-0338	CA21CH-0338-5-1	*C. coli*	1541	42	Bolton	mCCDA
CH-0338	CA21CH-0338-5-2	*C. coli*	1541	42	Bolton	mCCDA
CH-0338	CA21CH-0338-5-3	*C. coli*	1541	42	Bolton	mCCDA
CH-0338	CA21CH-0338-5-4	*C. coli*	cc unknown3	42	Bolton	mCCDA
CH-0338	CA21CH-0338-5-5	*C. coli*	1541	42	Bolton	u-mCCDA
CH-0338	CA21CH-0338-5-6	*C. coli*	1541	42	Bolton	u-mCCDA
CH-0338	CA21CH-0338-5-7	*C. coli*	cc unknown3	42	Bolton	u-mCCDA
CH-0338	CA21CH-0338-5-8	*C. coli*	1541	42	Bolton	u-mCCDA
CH-0338	CA21CH-0338-6-1	*C. coli*	cc unknown3	42	CAT	mCCDA
CH-0338	CA21CH-0338-6-2	*C. coli*	cc unknown3	42	CAT	mCCDA
CH-0338	CA21CH-0338-6-3	*C. coli*	cc unknown3	42	CAT	mCCDA
CH-0338	CA21CH-0338-6-4	*C. coli*	cc unknown3	42	CAT	mCCDA
CH-0338	CA21CH-0338-6-5	*C. coli*	1541	42	CAT	u-mCCDA
CH-0338	CA21CH-0338-6-6	*C. coli*	cc unknown3	42	CAT	u-mCCDA
CH-0338	CA21CH-0338-6-7	*C. coli*	cc unknown3	42	CAT	u-mCCDA
CH-0338	CA21CH-0338-6-8	*C. coli*	cc unknown3	42	CAT	u-mCCDA
CH-0339	CA21CH-0339-2-1	*C. jejuni*	cj unknown5	37	Bolton	mCCDA
CH-0339	CA21CH-0339-2-2	*C. jejuni*	cj unknown5	37	Bolton	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0339	CA21CH-0339-2-3	*C. jejuni*	cj unknown5	37	Bolton	mCCDA
CH-0339	CA21CH-0339-2-4	*C. jejuni*	cj unknown5	37	Bolton	mCCDA
CH-0339	CA21CH-0339-2-5	*C. jejuni*	cj unknown5	37	Bolton	u-mCCDA
CH-0339	CA21CH-0339-2-6	*C. jejuni*	cj unknown5	37	Bolton	u-mCCDA
CH-0339	CA21CH-0339-2-7	*C. jejuni*	cj unknown5	37	Bolton	u-mCCDA
CH-0339	CA21CH-0339-2-8	*C. jejuni*	cj unknown5	37	Bolton	u-mCCDA
CH-0339	CA21CH-0339-3-1	*C. jejuni*	51	37	CAT	mCCDA
CH-0339	CA21CH-0339-3-2	*C. jejuni*	51	37	CAT	mCCDA
CH-0339	CA21CH-0339-3-3	*C. jejuni*	51	37	CAT	mCCDA
CH-0339	CA21CH-0339-3-4	*C. jejuni*	51	37	CAT	mCCDA
CH-0339	CA21CH-0339-3-5	*C. jejuni*	51	37	CAT	u-mCCDA
CH-0339	CA21CH-0339-3-6	*C. jejuni*	51	37	CAT	u-mCCDA
CH-0339	CA21CH-0339-3-7	*C. jejuni*	51	37	CAT	u-mCCDA
CH-0339	CA21CH-0339-3-8	*C. jejuni*	51	37	CAT	u-mCCDA
CH-0339	CA21CH-0339-5-1	*C. jejuni*	51	42	Bolton	mCCDA
CH-0339	CA21CH-0339-5-2	*C. jejuni*	51	42	Bolton	mCCDA
CH-0339	CA21CH-0339-5-3	*C. jejuni*	51	42	Bolton	mCCDA
CH-0339	CA21CH-0339-5-4	*C. jejuni*	51	42	Bolton	mCCDA
CH-0339	CA21CH-0339-5-5	*C. jejuni*	51	42	Bolton	u-mCCDA
CH-0339	CA21CH-0339-5-6	*C. jejuni*	51	42	Bolton	u-mCCDA
CH-0339	CA21CH-0339-5-7	*C. jejuni*	51	42	Bolton	u-mCCDA
CH-0339	CA21CH-0339-5-8	*C. jejuni*	51	42	Bolton	u-mCCDA
CH-0339	CA21CH-0339-6-1	*C. jejuni*	2258	42	CAT	mCCDA
CH-0339	CA21CH-0339-6-2	*C. jejuni*	2258	42	CAT	mCCDA
CH-0339	CA21CH-0339-6-3	*C. coli*	6795	42	CAT	mCCDA
CH-0339	CA21CH-0339-6-4	*C. coli*	6795	42	CAT	mCCDA
CH-0339	CA21CH-0339-6-5	*C. coli*	6795	42	CAT	u-mCCDA
CH-0339	CA21CH-0339-6-6	*C. coli*	6795	42	CAT	u-mCCDA
CH-0339	CA21CH-0339-6-7	*C. jejuni*	2258	42	CAT	u-mCCDA
CH-0340	CA21CH-0340-1-1	*C. jejuni*	257	37	None (Direct plating)	u-mCCDA
CH-0340	CA21CH-0340-2-1	*C. jejuni*	441	37	Bolton	mCCDA
CH-0340	CA21CH-0340-2-2	*C. jejuni*	441	37	Bolton	mCCDA
CH-0340	CA21CH-0340-2-3	*C. jejuni*	441	37	Bolton	mCCDA
CH-0340	CA21CH-0340-2-4	*C. jejuni*	441	37	Bolton	mCCDA
CH-0340	CA21CH-0340-2-5	*C. jejuni*	441	37	Bolton	u-mCCDA
CH-0340	CA21CH-0340-2-6	*C. jejuni*	441	37	Bolton	u-mCCDA
CH-0340	CA21CH-0340-2-7	*C. jejuni*	257	37	Bolton	u-mCCDA
CH-0340	CA21CH-0340-2-8	*C. jejuni*	441	37	Bolton	u-mCCDA
CH-0340	CA21CH-0340-3-1	*C. jejuni*	447	37	CAT	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0340	CA21CH-0340-3-2	*C. jejuni*	6175	37	CAT	mCCDA
CH-0340	CA21CH-0340-3-3	*C. jejuni*	6175	37	CAT	mCCDA
CH-0340	CA21CH-0340-3-4	*C. jejuni*	447	37	CAT	mCCDA
CH-0340	CA21CH-0340-5-1	*C. jejuni*	441	42	Bolton	mCCDA
CH-0340	CA21CH-0340-5-2	*C. jejuni*	441	42	Bolton	mCCDA
CH-0340	CA21CH-0340-5-3	*C. jejuni*	441	42	Bolton	mCCDA
CH-0340	CA21CH-0340-5-4	*C. jejuni*	441	42	Bolton	mCCDA
CH-0340	CA21CH-0340-5-5	*C. coli*	827	42	Bolton	u-mCCDA
CH-0340	CA21CH-0340-5-6	*C. coli*	827	42	Bolton	u-mCCDA
CH-0340	CA21CH-0340-5-7	*C. coli*	cc unknown3	42	Bolton	u-mCCDA
CH-0340	CA21CH-0340-5-8	*C. jejuni*	441	42	Bolton	u-mCCDA
CH-0340	CA21CH-0340-6-1	*C. coli*	cc unknown3	42	CAT	mCCDA
CH-0340	CA21CH-0340-6-2	*C. coli*	cc unknown3	42	CAT	mCCDA
CH-0340	CA21CH-0340-6-3	*C. jejuni*	441	42	CAT	mCCDA
CH-0340	CA21CH-0340-6-4	*C. coli*	cc unknown3	42	CAT	u-mCCDA
CH-0340	CA21CH-0340-6-5	*C. coli*	cc unknown3	42	CAT	u-mCCDA
CH-0340	CA21CH-0340-6-6	*C. coli*	cc unknown3	42	CAT	u-mCCDA
CH-0340	CA21CH-0340-6-7	*C. coli*	cc unknown3	42	CAT	u-mCCDA
CH-0341	CA21CH-0341-2-1	*C. coli*	cc unknown3	37	Bolton	mCCDA
CH-0341	CA21CH-0341-2-2	*C. coli*	cc unknown3	37	Bolton	mCCDA
CH-0341	CA21CH-0341-2-3	*C. coli*	1595	37	Bolton	mCCDA
CH-0341	CA21CH-0341-2-4	*C. coli*	1595	37	Bolton	mCCDA
CH-0341	CA21CH-0341-2-5	*C. jejuni*	447	37	Bolton	u-mCCDA
CH-0341	CA21CH-0341-2-6	*C. jejuni*	6175	37	Bolton	u-mCCDA
CH-0341	CA21CH-0341-2-7	*C. coli*	cc unknown3	37	Bolton	u-mCCDA
CH-0341	CA21CH-0341-2-8	*C. jejuni*	447	37	Bolton	u-mCCDA
CH-0341	CA21CH-0341-3-1	*C. jejuni*	21	37	CAT	mCCDA
CH-0341	CA21CH-0341-3-2	*C. jejuni*	21	37	CAT	mCCDA
CH-0341	CA21CH-0341-3-3	*C. jejuni*	21	37	CAT	mCCDA
CH-0341	CA21CH-0341-3-4	*C. jejuni*	21	37	CAT	mCCDA
CH-0341	CA21CH-0341-3-5	*C. jejuni*	21	37	CAT	u-mCCDA
CH-0341	CA21CH-0341-3-6	*C. jejuni*	21	37	CAT	u-mCCDA
CH-0341	CA21CH-0341-3-7	*C. jejuni*	21	37	CAT	u-mCCDA
CH-0341	CA21CH-0341-3-8	*C. jejuni*	21	37	CAT	u-mCCDA
CH-0341	CA21CH-0341-5-1	*C. coli*	827	42	Bolton	mCCDA
CH-0341	CA21CH-0341-5-2	*C. coli*	827	42	Bolton	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0341	CA21CH-0341-5-3	*C. coli*	827	42	Bolton	mCCDA
CH-0341	CA21CH-0341-5-4	*C. coli*	827	42	Bolton	mCCDA
CH-0341	CA21CH-0341-5-5	*C. coli*	827	42	Bolton	u-mCCDA
CH-0341	CA21CH-0341-5-6	*C. coli*	827	42	Bolton	u-mCCDA
CH-0341	CA21CH-0341-5-7	*C. coli*	827	42	Bolton	u-mCCDA
CH-0341	CA21CH-0341-5-8	*C. coli*	827	42	Bolton	u-mCCDA
CH-0341	CA21CH-0341-6-1	*C. coli*	827	42	CAT	mCCDA
CH-0341	CA21CH-0341-6-2	*C. coli*	827	42	CAT	mCCDA
CH-0341	CA21CH-0341-6-3	*C. coli*	827	42	CAT	mCCDA
CH-0341	CA21CH-0341-6-4	*C. coli*	827	42	CAT	mCCDA
CH-0341	CA21CH-0341-6-5	*C. coli*	827	42	CAT	u-mCCDA
CH-0341	CA21CH-0341-6-6	*C. coli*	827	42	CAT	u-mCCDA
CH-0341	CA21CH-0341-6-7	*C. jejuni*	cj unknown8	42	CAT	u-mCCDA
CH-0341	CA21CH-0341-6-8	*C. coli*	827	42	CAT	u-mCCDA
CH-0347	CA21CH-0347-3-1	*C. jejuni*	262	37	CAT	mCCDA
CH-0347	CA21CH-0347-3-2	*C. jejuni*	262	37	CAT	mCCDA
CH-0347	CA21CH-0347-3-3	*C. jejuni*	262	37	CAT	mCCDA
CH-0347	CA21CH-0347-5-1	*C. jejuni*	262	42	Bolton	mCCDA
CH-0347	CA21CH-0347-5-2	*C. jejuni*	262	42	Bolton	mCCDA
CH-0347	CA21CH-0347-5-3	*C. jejuni*	262	42	Bolton	mCCDA
CH-0347	CA21CH-0347-5-4	*C. jejuni*	262	42	Bolton	u-mCCDA
CH-0347	CA21CH-0347-5-5	*C. jejuni*	262	42	Bolton	u-mCCDA
CH-0347	CA21CH-0347-5-6	*C. jejuni*	262	42	Bolton	u-mCCDA
CH-0347	CA21CH-0347-6-1	*C. jejuni*	262	42	CAT	mCCDA
CH-0347	CA21CH-0347-6-2	*C. jejuni*	262	42	CAT	mCCDA
CH-0347	CA21CH-0347-6-3	*C. jejuni*	262	42	CAT	mCCDA
CH-0347	CA21CH-0347-6-4	*C. jejuni*	262	42	CAT	u-mCCDA
CH-0347	CA21CH-0347-6-5	*C. jejuni*	262	42	CAT	u-mCCDA
CH-0347	CA21CH-0347-6-6	*C. jejuni*	262	42	CAT	u-mCCDA
CH-0348	CA21CH-0348-6-1	*C. coli*	827	42	CAT	mCCDA
CH-0348	CA21CH-0348-6-2	*C. coli*	827	42	CAT	mCCDA
CH-0348	CA21CH-0348-6-3	*C. coli*	827	42	CAT	mCCDA
CH-0348	CA21CH-0348-6-4	*C. coli*	827	42	CAT	u-mCCDA
CH-0348	CA21CH-0348-6-5	*C. coli*	827	42	CAT	u-mCCDA
CH-0348	CA21CH-0348-6-6	*C. coli*	827	42	CAT	u-mCCDA
CH-0349	CA21CH-0349-2-1	*C. coli*	cc unknown2	37	Bolton	mCCDA
CH-0349	CA21CH-0349-2-2	*C. coli*	cc unknown2	37	Bolton	mCCDA
CH-0349	CA21CH-0349-2-3	*C. coli*	cc unknown2	37	Bolton	mCCDA
CH-0349	CA21CH-0349-2-4	*C. coli*	cc unknown2	37	Bolton	u-mCCDA
CH-0349	CA21CH-0349-2-5	*C. coli*	cc unknown2	37	Bolton	u-mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0349	CA21CH-0349-2-6	*C. coli*	cc unknown2	37	Bolton	u-mCCDA
CH-0349	CA21CH-0349-5-1	*C. jejuni*	814	42	Bolton	mCCDA
CH-0349	CA21CH-0349-5-2	*C. jejuni*	814	42	Bolton	mCCDA
CH-0349	CA21CH-0349-5-3	*C. jejuni*	814	42	Bolton	mCCDA
CH-0349	CA21CH-0349-5-4	*C. jejuni*	814	42	Bolton	u-mCCDA
CH-0349	CA21CH-0349-5-5	*C. jejuni*	814	42	Bolton	u-mCCDA
CH-0349	CA21CH-0349-5-6	*C. jejuni*	814	42	Bolton	u-mCCDA
CH-0349	CA21CH-0349-6-1	*C. coli*	cc unknown2	42	CAT	mCCDA
CH-0349	CA21CH-0349-6-2	*C. coli*	cc unknown2	42	CAT	mCCDA
CH-0349	CA21CH-0349-6-3	*C. coli*	cc unknown2	42	CAT	mCCDA
CH-0349	CA21CH-0349-6-4	*C. coli*	cc unknown2	42	CAT	u-mCCDA
CH-0349	CA21CH-0349-6-5	*C. coli*	cc unknown2	42	CAT	u-mCCDA
CH-0349	CA21CH-0349-6-6	*C. coli*	cc unknown2	42	CAT	u-mCCDA
CH-0350	CA21CH-0350-2-1	*C. jejuni*	19	37	Bolton	mCCDA
CH-0350	CA21CH-0350-2-2	*C. jejuni*	19	37	Bolton	mCCDA
CH-0350	CA21CH-0350-2-3	*C. jejuni*	48	37	Bolton	mCCDA
CH-0350	CA21CH-0350-2-4	*C. jejuni*	48	37	Bolton	u-mCCDA
CH-0350	CA21CH-0350-2-5	*C. jejuni*	48	37	Bolton	u-mCCDA
CH-0350	CA21CH-0350-2-6	*C. jejuni*	19	37	Bolton	u-mCCDA
CH-0350	CA21CH-0350-3-1	*C. jejuni*	2066	37	CAT	mCCDA
CH-0350	CA21CH-0350-3-2	*C. jejuni*	2066	37	CAT	mCCDA
CH-0350	CA21CH-0350-3-3	*C. jejuni*	50	37	CAT	mCCDA
CH-0350	CA21CH-0350-5-1	*C. jejuni*	19	42	Bolton	mCCDA
CH-0350	CA21CH-0350-5-2	*C. jejuni*	19	42	Bolton	mCCDA
CH-0350	CA21CH-0350-5-3	*C. jejuni*	19	42	Bolton	mCCDA
CH-0350	CA21CH-0350-5-4	*C. jejuni*	19	42	Bolton	u-mCCDA
CH-0350	CA21CH-0350-5-5	*C. jejuni*	19	42	Bolton	u-mCCDA
CH-0350	CA21CH-0350-5-6	*C. jejuni*	19	42	Bolton	u-mCCDA
CH-0350	CA21CH-0350-6-1	*C. jejuni*	50	42	CAT	mCCDA
CH-0350	CA21CH-0350-6-2	*C. jejuni*	50	42	CAT	mCCDA
CH-0350	CA21CH-0350-6-3	*C. jejuni*	50	42	CAT	mCCDA
CH-0350	CA21CH-0350-6-4	*C. jejuni*	8334	42	CAT	u-mCCDA
CH-0350	CA21CH-0350-6-5	*C. jejuni*	2066	42	CAT	u-mCCDA
CH-0350	CA21CH-0350-6-6	*C. jejuni*	2066	42	CAT	u-mCCDA
CH-0351	CA21CH-0351-1-1	*C. coli*	825	37	None (Direct plating)	mCCDA
CH-0351	CA21CH-0351-2-1	*C. jejuni*	cj unknown8	37	Bolton	mCCDA
CH-0351	CA21CH-0351-2-2	*C. coli*	829	37	Bolton	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0351	CA21CH-0351-2-3	*C. jejuni*	48	37	Bolton	mCCDA
CH-0351	CA21CH-0351-2-4	*C. jejuni*	cj unknown8	37	Bolton	u-mCCDA
CH-0351	CA21CH-0351-2-5	*C. jejuni*	48	37	Bolton	u-mCCDA
CH-0351	CA21CH-0351-2-6	*C. jejuni*	cj unknown8	37	Bolton	u-mCCDA
CH-0351	CA21CH-0351-3-1	*C. coli*	829	37	CAT	mCCDA
CH-0351	CA21CH-0351-5-1	*C. coli*	825	42	Bolton	mCCDA
CH-0351	CA21CH-0351-5-2	*C. coli*	825	42	Bolton	mCCDA
CH-0351	CA21CH-0351-5-3	*C. coli*	825	42	Bolton	mCCDA
CH-0351	CA21CH-0351-5-4	*C. coli*	825	42	Bolton	u-mCCDA
CH-0351	CA21CH-0351-5-5	*C. coli*	825	42	Bolton	u-mCCDA
CH-0351	CA21CH-0351-5-6	*C. coli*	825	42	Bolton	u-mCCDA
CH-0351	CA21CH-0351-6-1	*C. coli*	825	42	CAT	mCCDA
CH-0351	CA21CH-0351-6-2	*C. coli*	825	42	CAT	mCCDA
CH-0351	CA21CH-0351-6-3	*C. coli*	825	42	CAT	mCCDA
CH-0353	CA21CH-0353-2-1	*C. jejuni*	814	37	Bolton	mCCDA
CH-0353	CA21CH-0353-2-2	*C. jejuni*	814	37	Bolton	mCCDA
CH-0353	CA21CH-0353-2-3	*C. jejuni*	814	37	Bolton	mCCDA
CH-0353	CA21CH-0353-2-4	*C. jejuni*	814	37	Bolton	u-mCCDA
CH-0353	CA21CH-0353-2-5	*C. jejuni*	814	37	Bolton	u-mCCDA
CH-0353	CA21CH-0353-2-6	*C. jejuni*	814	37	Bolton	u-mCCDA
CH-0353	CA21CH-0353-3-1	*C. jejuni*	814	37	CAT	mCCDA
CH-0353	CA21CH-0353-3-2	*C. jejuni*	814	37	CAT	mCCDA
CH-0353	CA21CH-0353-3-3	*C. jejuni*	814	37	CAT	mCCDA
CH-0353	CA21CH-0353-3-4	*C. jejuni*	814	37	CAT	u-mCCDA
CH-0353	CA21CH-0353-3-5	*C. jejuni*	814	37	CAT	u-mCCDA
CH-0353	CA21CH-0353-3-6	*C. jejuni*	814	37	CAT	u-mCCDA
CH-0353	CA21CH-0353-5-1	*C. coli*	827	42	Bolton	mCCDA
CH-0353	CA21CH-0353-5-2	*C. coli*	827	42	Bolton	mCCDA
CH-0353	CA21CH-0353-5-3	*C. coli*	827	42	Bolton	mCCDA
CH-0353	CA21CH-0353-5-4	*C. coli*	827	42	Bolton	u-mCCDA
CH-0353	CA21CH-0353-5-5	*C. coli*	827	42	Bolton	u-mCCDA
CH-0353	CA21CH-0353-5-6	*C. coli*	827	42	Bolton	u-mCCDA
CH-0353	CA21CH-0353-6-1	*C. jejuni*	814	42	CAT	mCCDA
CH-0353	CA21CH-0353-6-2	*C. jejuni*	814	42	CAT	mCCDA
CH-0353	CA21CH-0353-6-3	*C. jejuni*	814	42	CAT	mCCDA
CH-0353	CA21CH-0353-6-4	*C. jejuni*	814	42	CAT	u-mCCDA
CH-0353	CA21CH-0353-6-5	*C. jejuni*	814	42	CAT	u-mCCDA
CH-0353	CA21CH-0353-6-6	*C. jejuni*	814	42	CAT	u-mCCDA
CH-0355	CA21CH-0355-3-1	*C. jejuni*	9401	37	CAT	mCCDA
CH-0355	CA21CH-0355-3-2	*C. jejuni*	9401	37	CAT	mCCDA
CH-0355	CA21CH-0355-3-3	*C. jejuni*	9401	37	CAT	mCCDA
CH-0355	CA21CH-0355-5-1	*C. jejuni*	9401	42	Bolton	mCCDA
CH-0355	CA21CH-0355-5-2	*C. jejuni*	9401	42	Bolton	mCCDA
CH-0355	CA21CH-0355-5-3	*C. jejuni*	9401	42	Bolton	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0355	CA21CH-0355-5-4	*C. jejuni*	9401	42	Bolton	u-mCCDA
CH-0355	CA21CH-0355-5-5	*C. jejuni*	9401	42	Bolton	u-mCCDA
CH-0355	CA21CH-0355-5-6	*C. jejuni*	9401	42	Bolton	u-mCCDA
CH-0356	CA21CH-0356-6-1	*C. jejuni*	305	42	CAT	mCCDA
CH-0356	CA21CH-0356-6-2	*C. jejuni*	305	42	CAT	mCCDA
CH-0356	CA21CH-0356-6-3	*C. jejuni*	305	42	CAT	mCCDA
CH-0356	CA21CH-0356-6-4	*C. jejuni*	305	42	CAT	u-mCCDA
CH-0356	CA21CH-0356-6-5	*C. jejuni*	305	42	CAT	u-mCCDA
CH-0356	CA21CH-0356-6-6	*C. jejuni*	305	42	CAT	u-mCCDA
CH-0357	CA21CH-0357-2-1	*C. jejuni*	cj unknown9	37	Bolton	mCCDA
CH-0357	CA21CH-0357-2-2	*C. jejuni*	cj unknown9	37	Bolton	mCCDA
CH-0357	CA21CH-0357-2-3	*C. jejuni*	cj unknown9	37	Bolton	mCCDA
CH-0357	CA21CH-0357-2-4	*C. jejuni*	cj unknown9	37	Bolton	u-mCCDA
CH-0357	CA21CH-0357-2-5	*C. jejuni*	cj unknown9	37	Bolton	u-mCCDA
CH-0357	CA21CH-0357-2-6	*C. jejuni*	cj unknown9	37	Bolton	u-mCCDA
CH-0357	CA21CH-0357-3-1	*C. jejuni*	918	37	CAT	mCCDA
CH-0357	CA21CH-0357-3-2	*C. jejuni*	918	37	CAT	mCCDA
CH-0357	CA21CH-0357-3-3	*C. jejuni*	918	37	CAT	mCCDA
CH-0357	CA21CH-0357-5-1	*C. jejuni*	2036	42	Bolton	mCCDA
CH-0357	CA21CH-0357-5-2	*C. jejuni*	2036	42	Bolton	mCCDA
CH-0357	CA21CH-0357-5-3	*C. jejuni*	2036	42	Bolton	mCCDA
CH-0357	CA21CH-0357-5-4	*C. jejuni*	2036	42	Bolton	u-mCCDA
CH-0357	CA21CH-0357-5-5	*C. jejuni*	2036	42	Bolton	u-mCCDA
CH-0357	CA21CH-0357-5-6	*C. jejuni*	2036	42	Bolton	u-mCCDA
CH-0357	CA21CH-0357-6-1	*C. jejuni*	918	42	CAT	mCCDA
CH-0357	CA21CH-0357-6-2	*C. jejuni*	918	42	CAT	mCCDA
CH-0357	CA21CH-0357-6-3	*C. jejuni*	918	42	CAT	mCCDA
CH-0357	CA21CH-0357-6-4	*C. jejuni*	918	42	CAT	u-mCCDA
CH-0357	CA21CH-0357-6-5	*C. jejuni*	918	42	CAT	u-mCCDA
CH-0357	CA21CH-0357-6-6	*C. jejuni*	918	42	CAT	u-mCCDA
CH-0358	CA21CH-0358-5-1	*C. jejuni*	6175	42	Bolton	mCCDA
CH-0358	CA21CH-0358-5-2	*C. jejuni*	6175	42	Bolton	mCCDA
CH-0358	CA21CH-0358-5-3	*C. jejuni*	6175	42	Bolton	mCCDA
CH-0358	CA21CH-0358-5-4	*C. jejuni*	6175	42	Bolton	u-mCCDA
CH-0358	CA21CH-0358-5-5	*C. jejuni*	6175	42	Bolton	u-mCCDA
CH-0358	CA21CH-0358-5-6	*C. jejuni*	6175	42	Bolton	u-mCCDA
CH-0359	CA21CH-0359-2-1	*C. coli*	962	37	Bolton	mCCDA
CH-0359	CA21CH-0359-2-2	*C. jejuni*	6175	37	Bolton	mCCDA
CH-0359	CA21CH-0359-2-3	*C. coli*	962	37	Bolton	mCCDA
CH-0359	CA21CH-0359-2-4	*C. jejuni*	6175	37	Bolton	u-mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0359	CA21CH-0359-2-5	*C. coli*	962	37	Bolton	u-mCCDA
CH-0359	CA21CH-0359-2-6	*C. coli*	962	37	Bolton	u-mCCDA
CH-0359	CA21CH-0359-3-1	*C. coli*	962	37	CAT	mCCDA
CH-0359	CA21CH-0359-3-2	*C. jejuni*	6175	37	CAT	mCCDA
CH-0359	CA21CH-0359-3-3	*C. coli*	962	37	CAT	mCCDA
CH-0359	CA21CH-0359-3-4	*C. coli*	962	37	CAT	u-mCCDA
CH-0359	CA21CH-0359-3-5	*C. coli*	962	37	CAT	u-mCCDA
CH-0359	CA21CH-0359-3-6	*C. coli*	962	37	CAT	u-mCCDA
CH-0359	CA21CH-0359-5-1	*C. coli*	962	42	Bolton	mCCDA
CH-0359	CA21CH-0359-5-2	*C. coli*	962	42	Bolton	mCCDA
CH-0359	CA21CH-0359-5-3	*C. coli*	962	42	Bolton	mCCDA
CH-0359	CA21CH-0359-5-4	*C. coli*	962	42	Bolton	u-mCCDA
CH-0359	CA21CH-0359-5-5	*C. coli*	962	42	Bolton	u-mCCDA
CH-0359	CA21CH-0359-5-6	*C. coli*	962	42	Bolton	u-mCCDA
CH-0359	CA21CH-0359-6-1	*C. coli*	962	42	CAT	mCCDA
CH-0359	CA21CH-0359-6-2	*C. coli*	962	42	CAT	mCCDA
CH-0359	CA21CH-0359-6-3	*C. coli*	962	42	CAT	u-mCCDA
CH-0359	CA21CH-0359-6-4	*C. coli*	962	42	CAT	u-mCCDA
CH-0359	CA21CH-0359-6-5	*C. coli*	962	42	CAT	u-mCCDA
CH-0361	CA21CH-0361-2-1	*C. jejuni*	cj unknown10	37	Bolton	mCCDA
CH-0361	CA21CH-0361-2-2	*C. jejuni*	cj unknown10	37	Bolton	mCCDA
CH-0361	CA21CH-0361-2-3	*C. jejuni*	cj unknown10	37	Bolton	mCCDA
CH-0361	CA21CH-0361-2-4	*C. jejuni*	cj unknown10	37	Bolton	u-mCCDA
CH-0361	CA21CH-0361-2-5	*C. jejuni*	cj unknown10	37	Bolton	u-mCCDA
CH-0361	CA21CH-0361-2-6	*C. jejuni*	cj unknown10	37	Bolton	u-mCCDA
CH-0361	CA21CH-0361-3-1	*C. coli*	1541	37	CAT	mCCDA
CH-0361	CA21CH-0361-3-2	*C. coli*	1541	37	CAT	mCCDA
CH-0361	CA21CH-0361-3-3	*C. coli*	1541	37	CAT	mCCDA
CH-0361	CA21CH-0361-3-4	*C. coli*	1541	37	CAT	u-mCCDA
CH-0361	CA21CH-0361-3-5	*C. coli*	1541	37	CAT	u-mCCDA
CH-0361	CA21CH-0361-3-6	*C. coli*	1541	37	CAT	u-mCCDA
CH-0361	CA21CH-0361-5-1	*C. jejuni*	574	42	Bolton	mCCDA
CH-0361	CA21CH-0361-5-2	*C. jejuni*	574	42	Bolton	mCCDA
CH-0361	CA21CH-0361-5-3	*C. jejuni*	574	42	Bolton	mCCDA
CH-0361	CA21CH-0361-5-4	*C. jejuni*	574	42	Bolton	u-mCCDA
CH-0361	CA21CH-0361-5-5	*C. jejuni*	574	42	Bolton	u-mCCDA
CH-0361	CA21CH-0361-5-6	*C. jejuni*	574	42	Bolton	u-mCCDA
CH-0361	CA21CH-0361-6-1	*C. jejuni*	574	42	CAT	mCCDA
CH-0361	CA21CH-0361-6-2	*C. jejuni*	574	42	CAT	mCCDA
CH-0361	CA21CH-0361-6-3	*C. jejuni*	574	42	CAT	mCCDA

Appendix 2: Summary of the *Campylobacter* genomes obtained from the first 45 chicken samples, the samples they were obtained from, the culture method they were isolated with and their sequence type (ST)

Sample	Genome	Species	ST	Temperature (°C)	Broth	Plate
CH-0361	CA21CH-0361-6-4	*C. jejuni*	574	42	CAT	u-mCCDA
CH-0361	CA21CH-0361-6-5	*C. jejuni*	574	42	CAT	u-mCCDA
CH-0361	CA21CH-0361-6-6	*C. jejuni*	51	42	CAT	u-mCCDA

CAT=cefoperazone, amphotericin B, teicoplanin; mCCDA=modified charcoal-cefoperazone-deoxycholate agar; u-mCCDA=modified charcoal-cefoperazone-deoxycholate agar without supplements

Appendix 3: Allelic profiles of the novel sequence types (STs) identified

ST	Alleles						
	aspA	***glnA***	***gltA***	***glyA***	***pgm***	***tkt***	***uncA***
C. coli unknown1	*aspA*(33)	*glnA*(39)	*gltA*(240)	*glyA*(82)	*pgm*(104)	*tkt*(56)	*uncA*(17)
C. coli unknown2	*aspA*(124)	*glnA*(39)	*gltA*(30)	*glyA*(139)	*pgm*(104)	*tkt*(47)	*uncA*(17)
C. coli unknown3	*aspA*(~430)	*glnA*(39)	*gltA*(30)	*glyA*(82)	*pgm*(104)	*tkt*(44)	*uncA*(17)
C. coli unknown4	*aspA*(53)	*glnA*(38)	*gltA*(30)	*glyA*(81)	*pgm*(118)	*tkt*(47)	*uncA*(36)
C. jejuni unknown1	*aspA*(2)	*glnA*(115)	*gltA*(5)	*glyA*(26)	*pgm*(127)	*tkt*(29)	*uncA*(1)
C. jejuni unknown2	*aspA*(2)	*glnA*(2)	*gltA*(4)	*glyA*(64)	*pgm*(332)	*tkt*(7)	*uncA*(23)
C. jejuni unknown3	*aspA*(22)	*glnA*(28)	*gltA*(5)	*glyA*(17)	*pgm*(363)	*tkt*(3)	*uncA*(6)
C. jejuni unknown4	*aspA*(2)	*glnA*(75)	*gltA*(4)	*glyA*(48)	*pgm*(141)	*tkt*(67)	*uncA*(1)
C. jejuni unknown5	*aspA*(2)	*glnA*(231)	*gltA*(4)	*glyA*(48)	*pgm*(11)	*tkt*(67)	*uncA*(1)
C. jejuni unknown6	*aspA*(7)	*glnA*(75)	*gltA*(4)	*glyA*(15)	*pgm(*141)	*tkt*(34)	*uncA*(1)
C. jejuni unknown7	*aspA*(4)	*glnA*(7)	*gltA*(40)	*glyA*(85)	*pgm*(42)	*tkt*(51)	*uncA*(1)
C. jejuni unknown8	*aspA*(2)	*glnA*(21)	*gltA*(12)	*glyA*(62)	*pgm*(11)	*tkt*(67)	*uncA*(6)
C. jejuni unknown9	*aspA*(9)	*glnA*(21)	*gltA*(4)	*glyA*(62)	*pgm*(363)	*tkt*(3)	*uncA*(35)
C. jejuni unknown10	*aspA*(2)	*glnA*(231)	*gltA*(4)	*glyA*(48)	*pgm*(~886)	*tkt*(3)	*uncA*(1)

Appendix 4: Pairwise non-recombinogenic single nucleotide polymorphism (SNP) distances for *Campylobacter* sequence types (STs) identified in the study within samples and overall

ST	Species	Sample	Number of isolates	Pairwise SNPs	Median
257	*C. jejuni*	CH-0313	7	23-42	32
		CH-0334	1	N/A	N/A
		CH-0340	2	53	N/A
		All	10	23-92	53
45	*C. jejuni*	CH-0312	1	N/A	N/A
		All	1	N/A	N/A
5136	*C. jejuni*	CH-0313	5	27-66	52
		CH-0318	8	37-94	65
		CH-0321	9	29-70	53
		CH-0331	9	24-70	45
		CH-0333	1	N/A	N/A
		CH-0336	3	52-62	58
		All	35	24-115	66
464	*C. jejuni*	CH-0314	4	28-60	45
		CH-0320	8	26-84	49.5
		All	12	20-88	48.5
6175	*C. jejuni*	CH-0314	2	34	N/A
		CH-0319	15	15-52	34
		CH-0321	5	23-50	32
		CH-0331	8	15-56	35
		CH-0335	1	N/A	N/A
		CH-0336	11	14-46	24
		CH-0338	2	22	N/A
		CH-0340	2	29	N/A
		CH-0341	1	N/A	N/A
		CH-0358	6	14-31	24
		CH-0359	3	28-34	33
		All	56	9-72	34
cj unknown6	*C. jejuni*	CH-0315	2	53	N/A
		All	2	53	N/A
400	*C. jejuni*	CH-0317	45	13-74	41
		CH-0336	5	23-59	41.5
		CH-0337	1	N/A	N/A
		All	51	11-74	41
122	*C. jejuni*	CH-0319	1	N/A	N/A
		All	1	N/A	N/A
27	*C. lari*	CH-0320	16	14-65	39
		All	16	14-65	39
cj unknown7	*C. jejuni*	CH-0323	7	63-123	89
		All	7	63-123	89
21	*C. jejuni*	CH-0325	12	5-35	15
		CH-0335	1	N/A	N/A
		CH-0341	8	2-17	11
		All	21	2-89	33
cc unknown4	*C. coli*	CH-0325	1	N/A	N/A
		All	1	N/A	N/A
1096	*C. coli*	CH-0325	3	42-48	44
		CH-0327	8	8-53	29.5
		All	11	8-2413	44
1595	*C. coli*	CH-0326	15	23-111	56
		CH-0341	2	24	N/A
		All	17	23-128	72.5
2282	*C. jejuni*	CH-0327	7	49-88	60
		CH-0331	7	27-78	54
		All	14	19-90	59
cj unknown4	*C. jejuni*	CH-0328	6	43-99	68
		All	6	43-99	68
9401	*C. jejuni*	CH-0328	1	N/A	N/A
		CH-0355	9	11-42	28

Appendix 4: Pairwise non-recombinogenic single nucleotide polymorphism (SNP) distances for *Campylobacter* sequence types (STs) identified in the study within samples and overall

ST	Species	Sample	Number of isolates	Pairwise SNPs	Median
		All	10	11-100	30
814	*C. jejuni*	CH-0328	10	43-124	71
		CH-0349	6	13-37	28
		CH-0353	18	17-68	33
		All	34	13-125	49
829	*C. coli*	CH-0329	31	19-135	67
		CH-0351	2	17	N/A
		All	33	17-1955	70
828	*C. coli*	CH-0329	1	N/A	N/A
		CH-0338	4	27-41	32.5
		All	5	27-285	38
1191	*C. coli*	CH-0330	17	21-163	51.5
		All	17	21-163	51.5
441	*C. jejuni*	CH-0330	4	21-31	27
		CH-0340	13	14-35	23
		All	17	14-54	28
830	*C. coli*	CH-0330	3	32-46	41
		All	3	32-46	41
51	*C. jejuni*	CH-0332	8	15-46	29
		CH-0337	1	N/A	N/A
		CH-0339	16	11-244	30
		CH-0361	1	N/A	N/A
		All	26	11-244	63
61	*C. jejuni*	CH-0332	8	25-100	46
		All	8	25-100	46
cj unknown1	*C. jejuni*	CH-0333	2	108	N/A
		All	2	108	N/A
48	*C. jejuni*	CH-0333	1	N/A	N/A
		CH-0350	3	16-33	31
		CH-0351	2	49	N/A
		All	6	16-132	42
cj unknown2	*C. jejuni*	CH-0333	2	36	N/A
		CH-0334	1	N/A	N/A
		All	3	36-64	53
53	*C. jejuni*	CH-0333	1	N/A	N/A
		CH-0334	17	0-22	10
		All	18	0-22	10
267	*C. jejuni*	CH-0333	15	38-129	78
		All	15	38-129	78
cj unknown3	*C. jejuni*	CH-0333	1	N/A	N/A
		All	1	N/A	N/A
9012	*C. coli*	CH-0333	2	74	N/A
		All	2	74	N/A
230	*C. jejuni*	CH-0334	4	32-113	90.5
		CH-0336	7	44-125	67
		All	11	32-185	103
449	*C. jejuni*	CH-0334	1	N/A	N/A
		All	1	N/A	N/A
cc unknown1	*C. coli*	CH-0335	29	20-121	41
		All	29	20-121	41
227	*C. jejuni*	CH-0335	3	22-28	24
		All	3	22-28	24
2254	*C. jejuni*	CH-0335	1	N/A	N/A
		All	1	N/A	N/A
7743	*C. jejuni*	CH-0336	5	47-83	67
		All	5	47-83	67
2066	*C. jejuni*	CH-0337	16	9-43	28
		CH-0350	4	20-30	27.5
		All	20	9-43	28

Appendix 4: Pairwise non-recombinogenic single nucleotide polymorphism (SNP) distances for *Campylobacter* sequence types (STs) identified in the study within samples and overall

ST	Species	Sample	Number of isolates	Pairwise SNPs	Median
2211	*C. jejuni*	CH-0337	3	62-134	120
		All	3	62-134	120
6876	*C. jejuni*	CH-0338	6	34-73	43
		All	6	34-73	43
50	*C. jejuni*	CH-0338	1	N/A	N/A
		CH-0350	4	13-21	16.5
		All	5	13-177	20
cc unknown3	*C. coli*	CH-0338	12	20-63	40
		CH-0340	7	19-45	36
		CH-0341	3	28-41	36
		All	22	19-64	40
1541	*C. coli*	CH-0338	7	34-71	60
		CH-0361	6	19-54	37
		All	13	19-241	210.5
cj unknown5	*C. jejuni*	CH-0339	8	31-69	48
		All	8	31-69	48
2258	*C. jejuni*	CH-0339	3	30-47	33
		All	3	30-47	33
6795	*C. coli*	CH-0339	4	32-58	48.5
		All	4	32-58	48.5
447	*C. jejuni*	CH-0340	2	25	N/A
		CH-0341	2	29	N/A
		All	4	25-86	70
827	*C. coli*	CH-0340	2	32	N/A
		CH-0341	15	14-42	29
		CH-0348	6	18-40	36
		CH-0353	6	29-53	43
		All	29	14-110	77
cj unknown8	*C. jejuni*	CH-0341	1	N/A	N/A
		CH-0351	3	46-72	62
		All	4	46-76	59.5
262	*C. jejuni*	CH-0347	15	1-26	11
		All	15	1-26	11
cc unknown2	*C. coli*	CH-0349	12	17-70	33
		All	12	17-70	33
19	*C. jejuni*	CH-0350	9	5-27	16
		All	9	5-27	16
8334	*C. jejuni*	CH-0350	1	N/A	N/A
		All	1	N/A	N/A
825	*C. coli*	CH-0351	10	17-64	48
		All	10	17-64	48
305	*C. jejuni*	CH-0356	6	24-43	35
		All	6	24-43	35
cj unknown9	*C. jejuni*	CH-0357	6	14-86	44
		All	6	14-86	44
918	*C. jejuni*	CH-0357	9	19-62	43
		All	9	19-62	43
2036	*C. jejuni*	CH-0357	6	38-85	53
		All	6	38-85	53
962	*C. coli*	CH-0359	20	17-85	39
		All	20	17-85	39
cj unknown10	*C. jejuni*	CH-0361	6	41-87	69
		All	6	41-87	69
574	*C. jejuni*	CH-0361	11	23-67	41
		All	11	23-67	41

Appendix 5: Antimicrobial resistance genotypes of *Campylobacter* sequence type (ST) groups within samples

Sample	Species	ST	Number of isolates	AMR determinants	% isolates positive for determinant
CH-0312	*C. jejuni*	45	1	*bla*$_{OXA-447}$	100
CH-0313	*C. jejuni*	257	7	*bla*$_{OXA-193}$	100
		5136	5	*bla*$_{OXA-193}$	100
				tet(O/32/O)	100
				gyrA mutation (T86I)	100
CH-0314	*C. jejuni*	464	4	*bla*$_{OXA-193}$	100
				tet(O/32/O)	100
				gyrA mutation (T86I)	100
		6175	2	*bla*$_{OXA-193}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
CH-0315	*C. jejuni*	cj unknown6	2	*bla*$_{OXA-184}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
CH-0317	*C. jejuni*	400	45	*tet(O/32/O)*	100
				gyrA mutation (T86I)	100
CH-0318	*C. jejuni*	5136	8	*bla*$_{OXA-193}$	100
				tet(O/32/O)	100
				gyrA mutation (T86I)	100
CH-0319	*C. jejuni*	122	1	*bla*$_{OXA-193}$	100
		6175	15	*bla*$_{OXA-193}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
CH-0320	*C. lari*	27	16	*bla*$_{OXA-493}$	100
	C. jejuni	464	8	*bla*$_{OXA-193}$	100
				tet(O/32/O)	100
				gyrA mutation (T86I)	100
CH-0321	*C. jejuni*	5136	9	*bla*$_{OXA-193}$	100
				tet(O/32/O)	100
				gyrA mutation (T86I)	100
		6175	5	*bla*$_{OXA-193}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
CH-0323	*C. jejuni*	cj unknown7	7	*bla*$_{OXA-193}$	100
CH-0325	*C. jejuni*	21	12	*bla*$_{OXA-193}$	100
				gyrA mutation (T86I)	100
	C. coli	1096	3	*aadE-Cc*	100
		cc unknown4	1	*aadE-Cc*	100
CH-0326	*C. coli*	1595	15	*bla*$_{OXA-193}$	100
				tet(O)	100
CH-0327	*C. jejuni*	2282	7	*bla*$_{OXA-193}$	85.7
	C. coli	1096	8	NA	NA
CH-0328	*C. jejuni*	814	10	*bla*$_{OXA-184}$	100
				tet(O)	100
				gyrA mutation (T86I)	100

Appendix 5: Antimicrobial resistance genotypes of *Campylobacter* sequence type (ST) groups within samples

Sample	Species	ST	Number of isolates	AMR determinants	% isolates positive for determinant
		9401	1	*gyrA* mutation (T86I)	100
		cj unknown4	6	*bla*$_{OXA-184}$	100
				tet(O)	100
CH-0329	*C. coli*	828	1	*tet(O)*	100
				gyrA mutation (T86I)	100
		829	31	NA	NA
CH-0330	*C. jejuni*	441	4	*bla*$_{OXA-193}$	100
				gyrA mutation (T86I, P104S)	100
	C. coli	830	3	*bla*$_{OXA-193}$	100
				tet(O/32/O)	100
				gyrA mutation (T86I)	100
		1191	17	*bla*$_{OXA-193}$	94.1
CH-0331	*C. jejuni*	2282	7	*bla*$_{OXA-193}$	100
		5136	9	*bla*$_{OXA-193}$	100
				tet(O/32/O)	100
				gyrA mutation (T86I)	100
		6175	8	*bla*$_{OXA-193}$	87.5
				tet(O)	100
				gyrA mutation (T86I)	100
CH-0332	*C. jejuni*	51	8	*bla*$_{OXA-193}$	100
				tet(O)	100
		61	8	*bla*$_{OXA-193}$	100
				gyrA mutation (T86I)	100
CH-0333	*C. jejuni*	48	1	*bla*$_{OXA-61}$	100
		53	1	*bla*$_{OXA-193}$	100
		267	15	*bla*$_{OXA-193}$	93.3
		5136	1	*bla*$_{OXA-193}$	100
				tet(O/32/O)	100
				gyrA mutation (T86I)	100
		cj unknown1	2	*bla*$_{OXA-184}$	100
				tet(O)	100
		cj unknown2	2	*bla*$_{OXA-465}$	100
				tet(O)	100
		cj unknown3	1	*bla*$_{OXA-184}$	100
				tet(O)	100
	C. coli	9012	2	NA	NA
CH-0334	*C. jejuni*	53	17	*bla*$_{OXA-193}$	94.1
		230	4	*bla*$_{OXA-193}$	100
		257	1	*bla*$_{OXA-193}$	100
		449	1	*bla*$_{OXA-184}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
		cj unknown2	1	*bla*$_{OXA-465}$	100
				tet(O)	100

Appendix 5: Antimicrobial resistance genotypes of *Campylobacter* sequence type (ST) groups within samples

Sample	Species	ST	Number of isolates	AMR determinants	% isolates positive for determinant
CH-0335	*C. jejuni*	21	1	*bla*$_{OXA-193}$	100
		227	3	*bla*$_{OXA-193}$	100
		2254	1	*bla*$_{OXA-193}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
		6175	1	*bla*$_{OXA-193}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
	C. coli	cc unknown1	29	*aadE-Cc*	100
				bla$_{OXA-489}$	96.6
CH-0336	*C. jejuni*	230	7	*bla*$_{OXA-193}$	100
		400	5	*tet(O/32/O)*	100
				gyrA mutation (T86I)	100
		5136	3	*bla*$_{OXA-193}$	100
				tet(O/32/O)	100
				gyrA mutation (T86I)	100
		6175	11	*bla*$_{OXA-193}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
		7743	5	*bla*$_{OXA-184}$	20.0
				bla$_{OXA-185}$	80.0
				gyrA mutation (T86I)	100
CH-0337	*C. jejuni*	51	1	*bla*$_{OXA-184}$	100
		400	1	*tet(O/32/O)*	100
				gyrA mutation (T86I)	100
		2066	16	*tet(O)*	100
				gyrA mutation (T86I)	100
		2211	3	*bla*$_{OXA-184}$	100
				tet(O)	100
CH-0338	*C. jejuni*	50	1	*bla*$_{OXA-193}$	100
				tet(O/32/O)	100
				gyrA mutation (T86I)	100
		6175	2	*bla*$_{OXA-193}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
		6876	6	*ant(6)-Ia*	100
				bla$_{OXA-465}$	100
				gyrA mutation (T86I)	100
				tet(O)	100
	C. coli	828	4	*tet(O)*	100
				gyrA mutation (T86I)	100
		1541	7	NA	NA
		cc unknown3	12	*bla*$_{OXA-453}$	100
				tet(O)	100
				gyrA mutation (T86I)	100

Appendix 5: Antimicrobial resistance genotypes of *Campylobacter* sequence type (ST) groups within samples

Sample	Species	ST	Number of isolates	AMR determinants	% isolates positive for determinant
CH-0339	*C. jejuni*	51	16	*bla*$_{OXA-193}$	87.5
				bla$_{OXA-184}$	12.5
				tet(O)	87.5
		2258	3	*bla*$_{OXA-193}$	100
		cj unknown5	8	*bla*$_{OXA-185}$	100
				gyrA mutation (T86I)	100
	C. coli	6795	4	*gyrA* mutation (T86I)	100
CH-0340	*C. jejuni*	257	2	*bla*$_{OXA-193}$	100
				tet(O)	100
		441	13	*bla*$_{OXA-193}$	100
				gyrA mutation (T86I, P104S)	100
		447	2	*tet(O)*	100
		6175	2	*bla*$_{OXA-193}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
	C. coli	827	2	*bla*$_{OXA-489}$	100
		cc unknown3	7	*bla*$_{OXA-453}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
CH-0341	*C. jejuni*	21	8	*bla*$_{OXA-193}$	100
				gyrA mutation (T86I)	100
		447	2	NA	NA
		6175	1	*bla*$_{OXA-193}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
		cj unknown8	1	*bla*$_{OXA-193}$	100
				tet(O/32/O)	100
				gyrA mutation (T86I)	100
	C. coli	827	15	*bla*$_{OXA-489}$	100
		1595	2	*bla*$_{OXA-193}$	100
				tet(O)	100
		cc unknown3	3	*bla*$_{OXA-453}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
CH-0347	*C. jejuni*	262	15	*bla*$_{OXA-193}$	100
				tet(O)	100
CH-0348	*C. coli*	827	6	*bla*$_{OXA-489}$	100
CH-0349	*C. jejuni*	814	6	*bla*$_{OXA-184}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
	C. coli	cc unknown2	12	*bla*$_{OXA-452}$	100
				tet(O)	100
CH-0350	*C. jejuni*	19	9	*bla*$_{OXA-193}$	100
				gyrA mutation (T86I)	100
		48	3	*bla*$_{OXA-61}$	100
				tet(O)	100

Appendix 5: Antimicrobial resistance genotypes of *Campylobacter* sequence type (ST) groups within samples

Sample	Species	ST	Number of isolates	AMR determinants	% isolates positive for determinant
		50	4	*bla*$_{OXA-193}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
		2066	4	*tet(O)*	100
				gyrA mutation (T86I)	100
		8334	1	*tet(O)*	100
				gyrA mutation (T86I)	100
CH-0351	*C. jejuni*	48	2	*bla*$_{OXA-61}$	100
				tet(O)	100
		cj unknown8	3	*bla*$_{OXA-193}$	100
				gyrA mutation (T86I)	100
				tet(O/32/O)	100
	C. coli	825	10	*gyrA* mutation (T86I)	100
		829	2	NA	NA
CH-0353	*C. jejuni*	814	18	*bla*$_{OXA-184}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
	C. coli	827	6	*bla*$_{OXA-489}$	100
CH-0355	*C. jejuni*	9401	9	*gyrA* mutation (T86I)	100
CH-0356	*C. jejuni*	305	6	*bla*$_{OXA-193}$	100
				gyrA mutation (T86I)	100
CH-0357	*C. jejuni*	918	9	*bla*$_{OXA-193}$	100
				gyrA mutation (T86I)	100
		2036	6	*bla*$_{OXA-193}$	100
				tet(O/32/O)	100
				gyrA mutation (T86I)	100
		cj unknown9	6	*bla*$_{OXA-184}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
CH-0358	*C. jejuni*	6175	6	*bla*$_{OXA-184}$	100
				gyrA mutation (T86I)	100
				tet(O)	100
CH-0359	*C. jejuni*	6175	3	*bla*$_{OXA-193}$	100
				tet(O)	100
				gyrA mutation (T86I)	100
	C. coli	962	20	*tet(O)*	100
CH-0361	*C. jejuni*	51	1	*bla*$_{OXA-193}$	100
				tet(O)	100
		574	11	*bla*$_{OXA-193}$	100
				tet(O)	100
		cj unknown10	6	*bla*$_{OXA-185}$	100
				gyrA mutation (T86I)	100
	C. coli	1541	6	NA	NA

Appendix 6: Plate counts and colony forming units (CFU)/mL values for comparisons of growth of four *Campylobacter* isolates representing four STs initially identified through one condition only (Bolton broth/cefoperazone, amphotericin B, teicoplanin (CAT) broth)

Isolate	Condition	plate	Plate counts	Dilution	CFU/mL	Notes
314-6-1	Direct plating 37°C	mCCDA	26, 14	10^{-6}	2·00E+08	
314-6-1	Direct plating 42°C	mCCDA	31, 37	10^{-6}	3·40E+08	some colony swarming was observed, though enumeration was possible as the colonies were isolated
314-6-1	Bolton broth 37°C	mCCDA	133, 118	10^{-6}	1·26E+09	
314-6-1	Bolton broth 42°C	mCCDA	60, 77	10^{-6}	6·85E+08	
314-6-1	CAT broth 37°C	mCCDA	54, 74	10^{-4}	6·40E+06	
314-6-1	CAT broth 42°C	mCCDA	55, 33	10^{-6}	4·40E+08	
314-6-1	Direct plating 37°C	CBA	50, 40	10^{-6}	4·50E+08	
314-6-1	Direct plating 42°C	CBA	41, 38	10^{-6}	3·95E+08	some colony swarming was observed, though enumeration was possible as the colonies were isolated
314-6-1	Bolton broth 37°C	CBA	159, 167	10^{-6}	1·63E+09	
314-6-1	Bolton broth 42°C	CBA	75, 76	10^{-6}	7·55E+08	
314-6-1	CAT broth 37°C	CBA	72, 109	10^{-6}	9·05E+08	
314-6-1	CAT broth 42°C	CBA	98, 122	10^{-6}	1·10E+09	
330-6-7	Direct plating 37°C	mCCDA	43, 38	10^{-6}	4·05E+08	
330-6-7	Direct plating 42°C	mCCDA	30, 24	10^{-6}	2·70E+08	some colony swarming was observed, though enumeration was possible as the colonies were isolated
330-6-7	Bolton broth 37°C	mCCDA	7, 14	10^{-6}	1·05E+08	
330-6-7	Bolton broth 42°C	mCCDA	166, 172	10^{-6}	1·69E+09	
330-6-7	CAT broth 37°C	mCCDA	171, 190	10^{-6}	1·81E+09	
330-6-7	CAT broth 42°C	mCCDA	178, 180	10^{-6}	1·79E+09	
330-6-7	Direct plating 37°C	CBA	63, 61	10^{-6}	6·20E+08	

Appendix 6: Plate counts and colony forming units (CFU)/mL values for comparisons of growth of four *Campylobacter* isolates representing four STs initially identified through one condition only (Bolton broth/cefoperazone, amphotericin B, teicoplanin (CAT) broth)

Isolate	Condition	plate	Plate counts	Dilution	CFU/mL	Notes
330-6-7	Direct plating 42°C	CBA	48, 52	10^{-6}	5·00E+08	some colony swarming was observed, though enumeration was possible as the colonies were isolated
330-6-7	Bolton broth 37°C	CBA	24, 23	10^{-6}	2·35E+08	
330-6-7	Bolton broth 42°C	CBA	215, 175	10^{-6}	1·95E+09	
330-6-7	CAT broth 37°C	CBA	271, 307	10^{-6}	2·89E+09	
330-6-7	CAT broth 42°C	CBA	272, 249	10^{-6}	2·61E+09	
334-5-2	Direct plating 37°C	mCCDA	8, 16	10^{-6}	1·20E+08	
334-5-2	Direct plating 42°C	mCCDA	26	10^{-6}	2·60E+08	only one plate count was available for calculation of CFU/mL due to swarming on the other plate, thus the results should be treated with caution
334-5-2	Bolton broth 37°C	mCCDA	247, 236	10^{-6}	2·42E+09	
334-5-2	Bolton broth 42°C	mCCDA	24, 25	10^{-6}	2·45E+08	
334-5-2	CAT broth 37°C	mCCDA	268, 260	10^{-6}	2·64E+09	
334-5-2	CAT broth 42°C	mCCDA	95, 97	10^{-6}	9·60E+08	
334-5-2	Direct plating 37°C	CBA	36, 39	10^{-6}	3·75E+08	
334-5-2	Direct plating 42°C	CBA	42, 45	10^{-6}	4·35E+08	
334-5-2	Bolton broth 37°C	CBA	245, 234	10^{-6}	2·40E+09	
334-5-2	Bolton broth 42°C	CBA	177, 169	10^{-6}	1·73E+09	
334-5-2	CAT broth 37°C	CBA	311, 258	10^{-6}	2·85E+09	
334-5-2	CAT broth 42°C	CBA	156, 156	10^{-6}	1·56E+09	
350-2-1	Direct plating 37°C	mCCDA	142, 101*	10^{-4}	4·86E+07	*too numerous to count - one quarter of plate counted and count multiplied to obtain CFU/mL
350-2-1	Direct plating 42°C	mCCDA	15, 22	10^{-6}	1·85E+08	
350-2-1	Bolton broth 37°C	mCCDA	147, 154*	10^{-4}	6·02E+07	*too numerous to count - one quarter of plate counted and count multiplied to obtain CFU/mL

Appendix 6: Plate counts and colony forming units (CFU)/mL values for comparisons of growth of four *Campylobacter* isolates representing four STs initially identified through one condition only (Bolton broth/cefoperazone, amphotericin B, teicoplanin (CAT) broth)

Isolate	Condition	plate	Plate counts	Dilution	CFU/mL	Notes
350-2-1	Bolton broth 42°C	mCCDA	29, 17	10^{-6}	2·30E+08	
350-2-1	CAT broth 37°C	mCCDA	75, 90	10^{-6}	8·25E+08	
350-2-1	CAT broth 42°C	mCCDA	65, 60	10^{-6}	6·25E+08	
350-2-1	Direct plating 37°C	CBA	42, 34	10^{-6}	3·80E+08	
350-2-1	Direct plating 42°C	CBA	38, 35	10^{-6}	3·65E+08	
350-2-1	Bolton broth 37°C	CBA	74, 75*	10^{-6}	2·98E+09	*too numerous to count - one quarter of plate counted and count multiplied to obtain CFU/mL
350-2-1	Bolton broth 42°C	CBA	50, 44	10^{-6}	4·70E+08	
350-2-1	CAT broth 37°C	CBA	292, 283	10^{-6}	2·88E+09	
350-2-1	CAT broth 42°C	CBA	106, 118	10^{-6}	1·12E+09	

CBA=Columbia blood agar; mCCDA=modified charcoal-cefoperazone-deoxycholate agar

Appendix 7: Genes within each cluster of orthologous groups (COG) category identified in different plasmid clusters based on eggNOG analysis, summarised by preferred name and available description

Preferred name	COG category	Description	Cluster
-	-	-*	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
-	D	ATPase MipZ	2, 6
-	D	CobQ/CobB/MinD/ParA nucleotide binding domain	2, 6
-	D	Involved in chromosome partitioning	1, 3, 4, 7, 8, 9, 12
-	D	VirC1 protein	2
-	E	AAA domain, putative AbiEii toxin, Type IV TA system	8
-	G	4-amino-4-deoxy-alpha-L-arabinopyranosyl undecaprenyl phosphate biosynthetic process	3, 4
-	G	Phage lysozyme	1
-	G	TraG-like protein, N-terminal region	1, 3, 4, 8
-	H	PFAM C-5 cytosine-specific DNA methylase	8
-	I	Lipase (class 3)	1
-	I	Pfam Lipase (class 3)	1, 3, 4, 8
-	I	Protein of unknown function (DUF2974)	3, 8
-	J	elongation factor	7
-	J	elongation factor G	8
-	K	BRO family, N-terminal domain	2
-	K	nucleotide-binding Protein	12
-	L	AAA domain	7, 8, 9
-	L	Bacterial DNA topoisomeraes I ATP-binding domain	7, 8, 9, 12
-	L	COG0827 Adenine-specific DNA methylase	7, 8, 9, 12
-	L	COG1943 Transposase and inactivated derivatives	9
-	L	Domain of unknown function (DUF1738)	2, 6
-	L	Initiator Replication protein	11, 13
-	L	nucleotide-binding Protein	12
-	L	Phage integrase family	1, 3, 4, 8
-	L	Putative transposase DNA-binding domain	9
-	L	R.Pab1 restriction endonuclease	6
-	L	RecT family	1, 3, 4, 8
-	L	Resolvase, N terminal domain	7, 8, 9, 12
-	L	Single-strand binding protein family	2, 6, 7, 8, 9, 12
-	M	LysM domain	3, 4
-	M	Transglycosylase SLT domain	2, 6
-	N	COG0630 Type IV secretory pathway, VirB11 components, and related ATPases involved in archaeal flagella biosynthesis	7, 8, 9, 12
-	N	HicA toxin of bacterial toxin-antitoxin	1, 3, 4, 7, 8, 9
-	N	Type IV secretory pathway, VirB3-like protein	2, 6
-	O	Highly conserved protein containing a thioredoxin domain	3, 4
-	O	Signal peptidase, peptidase S26	2, 6
-	P	AAA domain, putative AbiEii toxin, Type IV TA system	8

Appendix 7: Genes within each cluster of orthologous groups (COG) category identified in different plasmid clusters based on eggNOG analysis, summarised by preferred name and available description

Preferred name	COG category	Description	Cluster
-	Q	Calcium- and calmodulin-responsive adenylate cyclase activity	8
-	Q	COG2931, RTX toxins and related Ca2 -binding proteins	8
-	S	Annotation was generated automatically without manual curation	8, 9
-	S	Bacterial mobilisation protein (MobC)	10
-	S	Cag pathogenicity island protein Cag12	7, 8, 9, 12
-	S	DNA/RNA non-specific endonuclease	2
-	S	Domain of unknown function (DUF4145)	1, 3, 4
-	S	Fic/DOC family	3, 4, 8
-	S	HicB_like antitoxin of bacterial toxin-antitoxin system	1, 3, 4, 7, 8, 9, 12
-	S	NYN domain	1, 3, 4, 7, 8
-	S	Peptidase C39 family	8, 9, 10
-	S	PFAM Uncharacterised protein family UPF0150	3, 4, 8
-	S	Phage regulatory protein Rha (Phage_pRha)	2, 6, 7, 8, 9, 12
-	S	PIN domain	2, 6
-	S	Plasmid stabilization system	3, 4
-	S	Protein conserved in bacteria	1, 3, 4, 8
-	S	Psort location Cytoplasmic, score	8
-	S	Relaxase/Mobilisation nuclease domain	10
-	S	RelE StbE family addiction module toxin	3, 4, 8
-	S	TIR domain	8, 9
-	S	TOPRIM	1, 3, 4, 7, 8, 9, 12
-	S	TraT complement resistance protein	3, 4
-	S	TrbC/VIRB2 family	7, 8, 9, 12
-	S	TrbM	2, 6, 7, 8, 9, 12
-	S	Type-1V conjugative transfer system mating pair stabilisation	3, 4
-	S	Type-F conjugative transfer system pilin assembly protein	3, 4
-	S	Virulence-associated protein D	8, 9, 12
-	S	WGR domain	1, 3, 4, 8
-	T	Conserved protein contains FHA domain	1, 3, 8
-	T	Nucleotide-binding Protein	12
-	U	ATPases associated with a variety of cellular activities	2, 6
-	U	Bacterial conjugation TrbI-like protein	2, 6, 7, 8, 9, 12
-	U	COG0630 Type IV secretory pathway, VirB11 components, and related ATPases involved in archaeal flagella biosynthesis	7, 8, 9, 12
-	U	COG3505 Type IV secretory pathway, VirD4 components	7, 8, 9, 12
-	U	Conjugal transfer protein	2, 6, 7, 8, 9, 12
-	U	Conjugation	2, 6
-	U	Relaxase	7, 8, 9, 12
-	U	Relaxase/Mobilisation nuclease domain	2, 6, 8

Appendix 7: Genes within each cluster of orthologous groups (COG) category identified in different plasmid clusters based on eggNOG analysis, summarised by preferred name and available description

Preferred name	COG category	Description	Cluster
-	U	Signal peptidase, peptidase S26	2, 6
-	U	TraM recognition site of TraD and TraG	2, 6
-	U	TrbC/VIRB2 family	2, 6
-	U	TrbL/VirB6 plasmid conjugal transfer protein	2, 6, 7, 8, 9, 12
-	U	Type II/IV secretion system protein	2, 6
-	U	Type IV secretion system proteins	7, 8, 9, 12
-	U	Type IV secretory pathway, VirB3-like protein	2, 6
-	U	VirB8 protein	2, 6, 7, 8, 9, 12
cagE	U	COG3451 Type IV secretory pathway, VirB4 components	7, 8, 9, 12
dcm	H	C-5 cytosine-specific DNA methylase	8
dcm	L	C-5 cytosine-specific DNA methylase	8
dinP	L	impB/mucB/samB family C-terminal domain	1, 3, 4, 8
dnaG	L	Participates in initiation and elongation during chromosome replication	1, 3, 4, 8, 12
fic	D	Fic/DOC family	2, 6
mob	-	-	13
tetM	J	Elongation factor G, domain IV	8
traG	U	Type IV secretory system Conjugative DNA transfer	2
tssA	S	Type VI secretion, EvfE, EvfF, ImpA, BimE, VC_A0119, VasJ	1, 3, 8
tssB	S	Type VI secretion	1, 3, 8
tssC	S	Type VI secretion protein	1, 3, 8
tssD	S	Type VI secretion system effector, Hcp	1, 3, 8
tssE	-	-	1, 3, 8
tssG	S	Type VI secretion, TssG	1, 3, 4, 8
tssJ	S	Type VI secretion	1, 3, 8
tssK	S	Type VI secretion	1, 3, 8
tssL	S	Type VI secretion system protein DotU	1, 3, 8
tssM	S	blastn search against coding sequences failed to identify a	1, 3, 8
vasA	S	Type VI secretion	1, 3, 4, 8
yafQ	S	Endonuclease activity	8

*the table displays summarised data, thus multiple genes without available preferred name or description are summarised here

Appendix 8: DESeq2 results showing differentially abundant taxa at the 0.05 p-value threshold (Benjamini-Hochberg corrected)

Base Mean	log_2Fold Change (ML)	lfcSE (ML)	stat	pvalue	padj	Species
2984.589	-5.807	0.640	-9.071	1.18E-19	5.36E-16	*Latilactobacillus sakei*
255.860	-5.081	0.820	-6.199	5.68E-10	1.29E-06	*Gordonia otitidis*
261.610	5.283	0.938	5.636	1.74E-08	2.64E-05	*Lawsonia intracellularis*
21.417	-3.051	0.560	-5.450	5.04E-08	5.72E-05	*Mycolicibacterium gilvum*
55.767	-3.167	0.593	-5.344	9.09E-08	8.25E-05	*Gordonia polyisoprenivorans*
417.219	-3.076	0.587	-5.237	1.63E-07	0.000113	*Stenotrophomonas sp. 169*
107.512	-2.439	0.467	-5.225	1.74E-07	0.000113	*Ralstonia pickettii*
32816.074	4.148	0.815	5.089	3.60E-07	0.00019	*Hafnia alvei*
33.494	-2.380	0.468	-5.081	3.76E-07	0.00019	*Ralstonia* sp. 56D2
575.685	4.450	0.891	4.993	5.94E-07	0.000269	*Hafnia* sp. CBA7124
47.604	-3.125	0.632	-4.948	7.52E-07	0.00031	*Nocardioides* sp. CF8
28.320	-2.084	0.426	-4.894	9.86E-07	0.000344	*Bosea* sp. Tri-49
139.498	-4.288	0.875	-4.900	9.61E-07	0.000344	*Plantibacter* sp. PA-3-X8
1331.991	3.553	0.734	4.839	1.30E-06	0.000422	*Photobacterium damselae*
22.305	-3.061	0.635	-4.819	1.44E-06	0.000436	*Agrococcus carbonis*
35.106	-2.656	0.564	-4.713	2.44E-06	0.000693	*Aureimonas altamirensis*
121.663	-3.032	0.649	-4.669	3.03E-06	0.000808	*Microbacterium aurum*
87.905	-4.036	0.868	-4.651	3.31E-06	0.000817	*Plantibacter* sp. M259
23.918	-2.883	0.621	-4.644	3.42E-06	0.000817	*Gordonia insulae*
27.963	-2.549	0.552	-4.622	3.80E-06	0.000863	*Bradyrhizobium* sp. SK17
77.392	4.146	0.903	4.590	4.42E-06	0.000956	*Citrobacter* sp. RHBSTW-01044
30.260	-2.877	0.635	-4.530	5.90E-06	0.001216	*Cupriavidus gilardii*
2821.661	2.719	0.604	4.505	6.63E-06	0.001301	*Providencia heimbachae*
31.108	-3.088	0.687	-4.497	6.88E-06	0.001301	*Nocardioides* sp. S5
69.064	-2.953	0.660	-4.477	7.58E-06	0.001376	*Aureimonas* sp. OT7
255.047	-3.738	0.839	-4.454	8.43E-06	0.001471	*Sanguibacter keddieii*
37.511	-2.399	0.542	-4.430	9.44E-06	0.001586	*Bosea* sp. RAC05
16.042	-2.607	0.594	-4.389	1.14E-05	0.001845	*Rhodococcus* sp. PBTS 1
110.332	-2.758	0.635	-4.341	1.42E-05	0.002217	*Schlegelella thermodepolymerans*
17.504	-2.179	0.503	-4.330	1.49E-05	0.002258	*Rhizobium* sp. WL3
79.171	3.100	0.717	4.323	1.54E-05	0.002258	*Myroides* sp. A21
74.809	-3.242	0.754	-4.302	1.69E-05	0.002324	*Brevibacterium* sp. CS2
265.827	-3.497	0.812	-4.306	1.67E-05	0.002324	*Rhodococcus* sp. P-2
64183.503	3.479	0.813	4.278	1.89E-05	0.002449	*Aeromonas salmonicida*
13.692	-2.314	0.540	-4.281	1.86E-05	0.002449	*Rhodococcus rhodochrous*
197.799	3.285	0.774	4.244	2.19E-05	0.00269	*Helicobacter pullorum*
96.300	2.658	0.625	4.249	2.15E-05	0.00269	*Myroides profundi*
187.219	10.902	2.577	4.231	2.33E-05	0.002776	*Macrococcus* sp. 19Msa1099
96.050	5.911	1.412	4.185	2.85E-05	0.003311	*Citrobacter* sp. RHBSTW-00017
35.773	-1.838	0.441	-4.167	3.09E-05	0.003506	*Aquabacterium olei*
41.278	4.497	1.086	4.140	3.48E-05	0.003849	*Chryseobacterium* sp. JV274
2844.528	2.111	0.514	4.105	4.05E-05	0.004372	*Providencia rettgeri*
136.680	-3.320	0.814	-4.077	4.56E-05	0.004812	*Gordonia bronchialis*

Appendix 8: DESeq2 results showing differentially abundant taxa at the 0.05 p-value threshold (Benjamini-Hochberg corrected)

Base Mean	log_2Fold Change (ML)	lfcSE (ML)	stat	pvalue	padj	Species
43.551	2.444	0.609	4.010	6.07E-05	0.006256	*Acetivibrio thermocellus*
132.433	2.843	0.711	4.001	6.32E-05	0.006368	*Avibacterium volantium*
73.807	5.641	1.419	3.976	7.01E-05	0.006913	*Weissella ceti*
529.688	-3.183	0.802	-3.968	7.23E-05	0.006982	*Brevundimonas naejangsanensis*
391.089	2.392	0.604	3.958	7.56E-05	0.007	*Serratia* sp. JSRIV006
595.367	2.918	0.736	3.962	7.43E-05	0.007	*Photobacterium profundum*
2027.753	2.836	0.719	3.947	7.92E-05	0.007187	*Gallibacterium anatis*
155.995	2.582	0.657	3.928	8.58E-05	0.007633	*Avibacterium paragallinarum*
3255.953	2.904	0.742	3.911	9.19E-05	0.007993	*Aeromonas* sp. CA23
86.455	-3.198	0.818	-3.907	9.34E-05	0.007993	*Janibacter indicus*
9459.438	1.948	0.501	3.891	9.98E-05	0.008381	*Serratia fonticola*
13.668	-2.006	0.522	-3.839	0.000124	0.010196	*Bradyrhizobium quebecense*
185.068	2.684	0.701	3.829	0.000129	0.010419	*Photobacterium gaetbulicola*
232.333	2.066	0.541	3.816	0.000136	0.01063	*Photorhabdus thracensis*
142.849	1.883	0.493	3.818	0.000134	0.01063	*Myroides phaeus*
582.553	2.011	0.532	3.778	0.000158	0.01214	*Yersinia aldovae*
335.396	1.981	0.526	3.768	0.000164	0.012336	*Providencia sneebia*
155.031	-2.647	0.703	-3.764	0.000167	0.012336	*Brevundimonas* sp. M20
266.519	1.878	0.499	3.762	0.000169	0.012336	*Empedobacter falsenii*
303.518	1.982	0.528	3.753	0.000175	0.012337	*Proteus hauseri*
48.716	-1.894	0.505	-3.750	0.000177	0.012337	*Schlegelella brevitalea*
25.026	-1.657	0.442	-3.753	0.000175	0.012337	*Mesorhizobium* sp. Pch-S
45.280	-2.961	0.794	-3.730	0.000192	0.013034	*Plantibacter flavus*
18.685	-2.371	0.636	-3.729	0.000192	0.013034	*Brachybacterium avium*
20139.612	3.318	0.892	3.721	0.000199	0.013099	*Yersinia entomophaga*
71.501	2.330	0.627	3.716	0.000202	0.013099	*Proteus columbae*
283.420	-2.278	0.612	-3.719	0.0002	0.013099	*Acidihalobacter ferrooxydans*
35.711	-1.875	0.505	-3.712	0.000205	0.013129	*Bosea* sp. ANAM02
1882.674	2.002	0.541	3.697	0.000218	0.013367	*Moellerella wisconsensis*
13.468	-1.948	0.527	-3.699	0.000217	0.013367	*Nocardioides* sp. MC1495
60.028	2.970	0.803	3.698	0.000218	0.013367	*Myroides* sp. ZB35
15069.421	2.307	0.627	3.679	0.000234	0.013706	*Morganella morganii*
16.897	-2.135	0.580	-3.678	0.000235	0.013706	*Metabacillus* sp. B2-18
756.779	3.219	0.873	3.686	0.000228	0.013706	*Rothia nasimurium*
26.727	-2.530	0.688	-3.677	0.000236	0.013706	*Microbacterium* sp. PM5
2031.312	2.181	0.594	3.668	0.000244	0.014012	*Yersinia pestis*
961.904	1.779	0.488	3.648	0.000264	0.014998	*Yersinia pseudotuberculosis*
24.459	-2.272	0.627	-3.622	0.000292	0.015965	*Agrococcus* sp. SCSIO52902
66.316	-2.472	0.682	-3.626	0.000287	0.015965	*Rhodococcus* sp. P1Y
83.855	-2.645	0.730	-3.624	0.00029	0.015965	*Rhodococcus* sp. B7740
243.707	1.775	0.493	3.597	0.000322	0.01724	*Escherichia marmotae*
159.417	2.510	0.698	3.596	0.000323	0.01724	*Ornithobacterium rhinotracheale*
332.137	1.578	0.440	3.583	0.000339	0.017717	*Yersinia mollaretii*

Appendix 8: DESeq2 results showing differentially abundant taxa at the 0.05 p-value threshold (Benjamini-Hochberg corrected)

Base Mean	log_2Fold Change (ML)	lfcSE (ML)	stat	pvalue	padj	Species
81.308	-2.707	0.756	-3.583	0.00034	0.017717	*Brevundimonas* sp. AJA228-03
24.129	-2.191	0.617	-3.552	0.000382	0.019718	*Blastococcus saxobsidens*
18.734	-1.967	0.555	-3.543	0.000396	0.020189	*Gordonia amarae*
198.247	1.652	0.469	3.520	0.000431	0.021632	*Yersinia similis*
77.714	-2.065	0.587	-3.519	0.000434	0.021632	*Ralstonia mannitolilytica*
114.695	-2.697	0.772	-3.495	0.000473	0.023345	*Brevundimonas* sp. LM2
18.433	3.753	1.075	3.490	0.000482	0.023523	*Citrobacter* sp. RHBSTW-00127
1096.273	2.184	0.627	3.485	0.000492	0.023759	*Kluyvera intermedia*
147.610	2.814	0.810	3.473	0.000515	0.023841	*Proteus* sp. NMG38-2
303.890	-2.797	0.805	-3.473	0.000515	0.023841	*Brevundimonas mediterranea*
123.141	-2.670	0.768	-3.478	0.000505	0.023841	*Brevundimonas goettingensis*
41.703	4.105	1.181	3.475	0.000511	0.023841	*Geobacillus thermoleovorans*
21.952	-1.588	0.458	-3.464	0.000532	0.023889	*Bradyrhizobium* sp. BTAi1
9149.135	-2.841	0.819	-3.467	0.000526	0.023889	*Bartonella krasnovii*
21.678	-1.997	0.576	-3.465	0.000529	0.023889	*Blastococcus* sp. PRF04-17
205.923	1.947	0.564	3.454	0.000553	0.024005	*Yersinia* sp. KBS0713
1188.852	-1.317	0.381	-3.459	0.000541	0.024005	*Xanthomonas citri*
50.320	-1.283	0.372	-3.445	0.000571	0.024005	*Massilia* sp. YMA4
57.400	-2.540	0.737	-3.445	0.000571	0.024005	*Massilia* sp. MB5
378.064	-2.725	0.789	-3.452	0.000557	0.024005	*Brevundimonas diminuta*
83.230	-2.727	0.790	-3.454	0.000552	0.024005	*Brevundimonas vitisensis*
170.758	2.346	0.680	3.449	0.000562	0.024005	*Clostridium isatidis*
20.984	-2.096	0.611	-3.431	0.000601	0.025027	*Microbacterium protaetiae*
364.310	1.619	0.472	3.428	0.000608	0.025091	*Escherichia albertii*
19.476	-2.169	0.633	-3.425	0.000616	0.025163	*Brachybacterium ginsengisoli*
17.115	-2.229	0.652	-3.420	0.000627	0.025407	*Cupriavidus* sp. USMAHM13
34.922	-1.719	0.504	-3.411	0.000648	0.026011	*Thermomonas* sp. XSG
82.330	-2.564	0.753	-3.407	0.000657	0.026141	*Brevundimonas subvibrioides*
17.638	-1.726	0.508	-3.401	0.000672	0.026506	*Bosea* sp. PAMC 26642
63.599	3.134	0.924	3.391	0.000695	0.026786	*Serratia* sp. HRI
98.908	1.535	0.454	3.379	0.000726	0.026786	*Pectobacterium odoriferum*
80.683	1.897	0.560	3.390	0.0007	0.026786	*Vibrio navarrensis*
54.147	-2.224	0.658	-3.383	0.000718	0.026786	*Sulfuritortus calidifontis*
20.877	-1.484	0.439	-3.380	0.000724	0.026786	*Bosea vaviloviae*
29.878	-2.642	0.780	-3.388	0.000704	0.026786	*Rhodococcus* sp. R79
76.023	1.582	0.466	3.393	0.000692	0.026786	*Myroides odoratus*
1103.363	2.217	0.655	3.384	0.000713	0.026786	*Bacteroides fragilis*
94.893	1.345	0.400	3.365	0.000767	0.0279	*Erwinia [Pantoea] beijingensis*
83.723	2.173	0.646	3.364	0.000769	0.0279	*Aeromonas* sp. FDAARGOS 1410
220.491	1.642	0.490	3.350	0.000809	0.029134	*Jinshanibacter zhutongyuii*
90.047	2.154	0.644	3.347	0.000816	0.02915	*Vibrio aphrogenes*
84.192	2.212	0.664	3.330	0.000868	0.030609	*Thaumasiovibrio subtropicus*

Appendix 8: DESeq2 results showing differentially abundant taxa at the 0.05 p-value threshold (Benjamini-Hochberg corrected)

Base Mean	log_2Fold Change (ML)	lfcSE (ML)	stat	pvalue	padj	Species
21.320	-2.158	0.648	-3.329	0.00087	0.030609	*Gordonia* sp. KTR9
166.369	2.048	0.616	3.327	0.000879	0.030665	*Bacteroides thetaiotaomicron*
17.724	2.961	0.893	3.317	0.00091	0.031512	*Citrobacter* sp. 172116965
109.035	-2.072	0.627	-3.303	0.000956	0.03287	*Luteimonas* sp. YGD11-2
29.201	-2.210	0.670	-3.299	0.000972	0.033153	*Cutibacterium avidum*
17.636	-1.667	0.506	-3.294	0.000986	0.033179	*Bradyrhizobium* sp. CCBAU 53421
528.028	2.804	0.851	3.294	0.000987	0.033179	*Caldibacillus thermoamylovorans*
242.613	1.490	0.453	3.290	0.001003	0.033214	*Salmonella bongori*
32.772	1.895	0.576	3.291	0.001	0.033214	*Cellulophaga algicola*
32164.115	2.023	0.615	3.287	0.001013	0.033313	*Yersinia intermedia*
21.380	-1.853	0.564	-3.284	0.001023	0.033404	*Methylobacterium brachiatum*
32.334	-2.104	0.642	-3.276	0.001052	0.034093	*Microbacterium* sp. A18JL241
74.985	2.468	0.754	3.271	0.001071	0.034364	*Pseudoalteromonas* sp. SM9913
82.942	-1.875	0.573	-3.270	0.001076	0.034364	*Pseudoxanthomonas suwonensis*
74.734	-2.415	0.739	-3.267	0.001087	0.034483	*Microbacterium* sp. SSW1-36
25.221	1.447	0.444	3.262	0.001107	0.034892	*Capnocytophaga canimorsus*
125.413	-2.830	0.870	-3.252	0.001148	0.035907	*Brevundimonas vancanneytii*
154.584	-2.830	0.873	-3.241	0.00119	0.036476	*Paracoccus sanguinis*
12.538	-1.661	0.512	-3.242	0.001185	0.036476	*Mycolicibacterium fluoranthenivorans*
30.657	-2.261	0.697	-3.245	0.001175	0.036476	*Gordonia terrae*
452.403	2.376	0.737	3.223	0.001267	0.03833	*Citrobacter freundii* complex sp. CFNIH3
155.742	2.440	0.757	3.221	0.001277	0.03833	*Aeromonas* sp. FDAARGOS 1414
170.608	1.767	0.551	3.207	0.001343	0.03833	*Vibrio taketomensis*
12.361	-1.526	0.474	-3.221	0.001279	0.03833	*Rhizobium rhizoryzae*
15.956	-1.501	0.467	-3.216	0.001301	0.03833	*Bosea* sp. AS-1
195.378	-2.442	0.760	-3.211	0.001323	0.03833	*Brevundimonas* sp. CS1
31.297	-2.119	0.660	-3.212	0.001317	0.03833	*Cutibacterium granulosum*
29.725	1.508	0.470	3.208	0.001338	0.03833	*Formosa* sp. L2A11
20.026	3.227	1.005	3.210	0.001329	0.03833	*Bacteroides* sp. CACC 737
74.166	2.314	0.720	3.216	0.001301	0.03833	*Bacteroides stercoris*
500.549	2.585	0.804	3.214	0.00131	0.03833	*Alistipes onderdonkii*
19.914	-1.791	0.559	-3.201	0.001368	0.03868	*Rhodococcus* sp. MTM3W5.2
15.507	1.779	0.556	3.200	0.001373	0.03868	*Dokdonia* sp. Dokd-P16
657.596	1.460	0.457	3.193	0.001408	0.039422	*Enterobacter roggenkampii*
138.540	2.211	0.695	3.179	0.001477	0.040372	*Pseudoalteromonas agarivorans*
49.082	-1.224	0.385	-3.181	0.001465	0.040372	*Ralstonia insidiosa*
36.299	-2.120	0.667	-3.179	0.001477	0.040372	*Microbacterium* sp. 1S1

Appendix 8: DESeq2 results showing differentially abundant taxa at the 0.05 p-value threshold (Benjamini-Hochberg corrected)

Base Mean	log_2Fold Change (ML)	lfcSE (ML)	stat	pvalue	padj	Species
22.189	-1.981	0.622	-3.184	0.001451	0.040372	*Microbacterium cremeum*
545.078	2.038	0.642	3.175	0.001499	0.040719	*Serratia* sp. JSRIV001
29.768	-2.170	0.684	-3.172	0.001516	0.04095	*Miniimonas* sp. S16
114.070	1.366	0.431	3.169	0.001529	0.041043	*Xenorhabdus poinarii*
23.067	-1.732	0.548	-3.162	0.001567	0.041815	*Rhodococcus globerulus*
104.467	7.861	2.496	3.149	0.001637	0.043251	*Serratia* sp. AS13
29.473	-1.745	0.554	-3.149	0.00164	0.043251	*Yimella* sp. cx-51
87.239	1.755	0.558	3.146	0.001657	0.043451	*Vibrio alfacsensis*
259.364	1.933	0.615	3.144	0.001668	0.043496	*Aliivibrio fischeri*
228.815	1.517	0.484	3.135	0.001718	0.044387	*Enterobacter* sp. SA187
22.562	2.658	0.849	3.133	0.001732	0.044387	*Pseudoalteromonas donghaensis*
98.049	1.924	0.614	3.134	0.001725	0.044387	*Bacteroides eggerthii*
25.437	-1.860	0.594	-3.130	0.001751	0.04454	*Microbacterium lemovicicum*
88.082	-1.524	0.487	-3.128	0.001757	0.04454	*Leptospira santarosai*
55.645	1.811	0.579	3.126	0.001772	0.044666	*Staphylococcus argenteus*
297.254	2.022	0.649	3.115	0.001838	0.045295	*Aeromonas* sp. ASNIH4
15.612	-1.355	0.435	-3.114	0.001847	0.045295	*Caulobacter* sp. FWC26
20.145	1.960	0.629	3.118	0.001821	0.045295	*Weissella soli*
24.397	-1.921	0.616	-3.116	0.001831	0.045295	*Microbacterium* sp. 10M-3C3
19.339	-1.969	0.632	-3.115	0.001842	0.045295	*Ornithinimicrobium* sp. HY006
4813.847	1.352	0.435	3.106	0.001898	0.045312	*Enterobacter hormaechei*
8497.636	2.237	0.719	3.111	0.001863	0.045312	*Hafnia paralvei*
116.171	1.773	0.570	3.109	0.001878	0.045312	*Iodobacter ciconiae*
16.316	3.427	1.103	3.108	0.001884	0.045312	*Lysinibacillus varians*
128.908	1.484	0.478	3.106	0.001897	0.045312	*Empedobacter brevis*
14.599	-1.652	0.532	-3.104	0.001909	0.045354	*Mycolicibacterium aubagnense*
42.453	-2.373	0.767	-3.095	0.001966	0.045575	*Pseudoxanthomonas daejeonensis*
30.356	2.319	0.749	3.095	0.001969	0.045575	*Salinivibrio* sp. YCSC6
26.983	-2.029	0.655	-3.097	0.001956	0.045575	*Microbacterium resistens*
13.519	-1.829	0.590	-3.099	0.001942	0.045575	*Ornithinimicrobium avium*
25.700	-1.965	0.634	-3.099	0.001945	0.045575	*Deinococcus proteolyticus*
20.748	-1.782	0.576	-3.093	0.001979	0.045577	*Gordonia pseudoamarae*
9216.295	2.518	0.815	3.088	0.002014	0.046143	*Aeromonas media*
208.545	1.496	0.485	3.084	0.002043	0.046574	*Vibrio mediterranei*
149.557	2.005	0.651	3.079	0.002074	0.047043	*Aliivibrio salmonicida*
122.087	1.417	0.461	3.075	0.002106	0.04754	*Xenorhabdus budapestensis*
50.208	-2.044	0.665	-3.072	0.002123	0.04759	*Microbacterium hominis*
174.290	1.690	0.550	3.072	0.002129	0.04759	*Flavobacterium* sp. CECT 9288
718.290	1.681	0.548	3.066	0.00217	0.048271	*Serratia* sp. JSRIV002

ML=maximum likelihood

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711050	Stanley	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-
ERR2711051	Typhimurium	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-
		aph(6)-Id_1_M28829	Aminoglycoside	Yes	0.242640518	0.242640518	
		aph(3'')-Ib_5_AF321551	Aminoglycoside	different variant (aph(3'')-Ib_3_AF321550)	0.177414572	0.177414572	
		aac(3)-IIa_4_L22613	Aminoglycoside	No			
		blaTEM-1B_1_AY458016	Beta-lactam	No			
		mcr-3.1_1_KY924928	Polymyxin	No			
		mph(A)_1_D16251	Macrolide	No			
		floR_2_AF118107	Phenicol	No			
		qnrS1_1_AB187515	Fluoroquinolone	No			
		sul2_2_AY034138	Sulphonamide	Yes	0.20124638	0.20124638	
		sul2_6_FN995456	Sulphonamide	No			
		tet(A)_6_AF534183	Tetracycline	No			
		tet(B)_2_AF326777	Tetracycline	No			
ERR2711052	I 4,[5],12:i:-	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-
		aph(6)-Id_1_M28829	Aminoglycoside	Yes	1.007362658	10.35654569	
		aph(3')-Ia_3_EF015636	Aminoglycoside	No			

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711052	I 4,[5],12:i:-	aph(3'')-Ib_5_ AF321551	Aminoglycoside	Yes	10.71231254	10.71401591	-
		blaTEM-1B_1_ AY458016	Beta-lactam	Yes	0.446738976	0.446738976	
		mcr-3.20_1_ NG055493	Polymyxin	No			
		mph(A)_1_D16251	Macrolide	Yes	6.69517145	8.285606395	
		floR_2_AF118107	Phenicol	No			
		qnrS1_1_ AB187515	Fluoroquinolone	Yes	0.036754212	0.036754212	
		ARR-2_1_ HQ141279	Rifamycin	No			
		ARR-3_4_ FM207631	Rifamycin	No			
		sul2_3_HQ840942	Sulphonamide	different variants (sul2_2_AY034138, sul2_5_AY524415, sul2_9_FJ197818, sul2_12_AF497970, sul2_17_U57647, sul2_19_AJ319822)		10.70872236	
		tet(A)_6_AF534183	Tetracycline	Yes	14.61973114	16.5856848	
		tet(B)_2_AF326777	Tetracycline	No			
		dfrA14_5_ DQ388123	Diaminopyrimidine	No			
		dfrA14_1_ KF921535	Diaminopyrimidine	No			

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711053	Meleagridis	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-
		aadA2_1_NC_010870	Aminoglycoside	Yes	0.067102505	0.067102505	
		aadA1_3_JQ414041	Aminoglycoside	No			
		blaTEM-1B_1_AY458016	Beta-lactam	No			
		fosA7_1_LAPJ01000014	Phosphonic acid	No			
		cmlA1_1_M64556	Phenicol	No			
		floR_2_AF118107	Phenicol	No			
		qnrS1_1_AB187515	Fluoroquinolone	No			
		sul2_2_AY034138	Sulphonamide	Yes	0.152893619	0.152893619	
		sul3_2_AJ459418	Sulphonamide	No			
		tet(A)_6_AF534183	Tetracycline	Yes	0.243177575	0.243177575	
		dfrA12_8_AM040708	Diaminopyrimidine	No			
ERR2711054	Saintpaul	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			*gyrA* (S83Y)
		aph(6)-Id_1_M28829	Aminoglycoside	Yes	0.101796245	0.101796245	
		aph(3')-Ia_3_EF015636	Aminoglycoside	Yes	0.01482189	0.01482189	
		aph(3'')-Ib_5_AF321551	Aminoglycoside	different variant (aph(3'')-Ib_2_AF024602)	0.111910479	0.111910479	
		aac(3)-IIa_4_L22613	Aminoglycoside	No			

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711054	Saintpaul	blaTEM-1B_1_AY458016	Beta-lactam	Yes	0.212320133	0.212320133	*gyrA* (S83Y)
		lnu(F)_1_EU118119	Lincosamide	Yes	0.041384488	0.041384488	
		mph(A)_1_D16251	Macrolide	Yes	0.019676438	0.019676438	
		floR_2_AF118107	Phenicol	Yes	0.056852584	0.056852584	
		qnrS1_1_AB187515	Fluoroquinolone	Yes	0.113241392	0.113241392	
		ARR-2_1_HQ141279	Rifamycin	No			
		ARR-3_4_FM207631	Rifamycin	No			
		sul2_2_AY034138	Sulphonamide	Yes	0.134524759	0.134524759	
		tet(A)_6_AF534183	Tetracycline	Yes	0.366016091	0.366016091	
		dfrA14_5_DQ388123	Diaminopyrimidine	No			
		dfrA14_1_KF921535	Diaminopyrimidine	No			
ERR2711055	Weltevreden	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-
ERR2711056	Enteritidis	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			*gyrA* (D87N)
		aph(6)-Id_4_CP000971	Aminoglycoside	No			
		aph(3'')-Ib_5_AF321551	Aminoglycoside	No			
		blaTEM-1B_1_AY458016	Beta-lactam	No			
		sul2_2_AY034138	Sulphonamide	No			
		tet(A)_6_AF534183	Tetracycline	No			

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711057	Typhimurium	aac(6')-Iaa_1_NC_003197	Aminoglycoside	Yes	0.453846078	0.453846078	-
		aadA2_1_NC_010870	Aminoglycoside	Yes	3.183680231	3.183680231	
		aadA1_3_JQ414041	Aminoglycoside	Yes	0.917500452	0.917500452	
		blaTEM-1B_1_AY458016	Beta-lactam	Yes	4.710105962	4.710105962	
		cmlA1_1_M64556	Phenicol	Yes	4.735785163	4.735785163	
		floR_2_AF118107	Phenicol	Yes	4.863795011	4.863795011	
		qnrS1_1_AB187515	Fluoroquinolone	Yes	1.357309855	1.357309855	
		sul2_2_AY034138	Sulphonamide	Yes	6.758919998	6.758919998	
		sul3_2_AJ459418	Sulphonamide	Yes	2.298848355	2.298848355	
		tet(A)_6_AF534183	Tetracycline	Yes	8.616047492	8.616047492	
		dfrA12_8_AM040708	Diaminopyrimidine	Yes	1.583306126	1.583306126	
ERR2711058	Typhimurium	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-
		ant(3'')-Ia_1_X02340	Aminoglycoside	Yes	1.391013952	1.391013952	
		aadA1_3_JQ414041	Aminoglycoside	No			
		aadA2_2_JQ364967	Aminoglycoside	Yes	0.639419271	0.639419271	
		blaTEM-1B_1_AY458016	Beta-lactam	Yes	8.013713028	8.013713028	
		cmlA1_1_M64556	Phenicol	Yes	0.568167474	0.568167474	

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711058	Typhimurium	qnrS1_1_ AB187515	Fluoroquinolone	No			-
		sul3_2_AJ459418	Sulphonamide	Yes	0.352856789	0.352856789	
		tet(A)_4_AJ517790	Tetracycline	Yes	1.182669677	1.182669677	
		tet(A)_6_AF534183	Tetracycline	No			
		dfrA12_8_ AM040708	Diaminopyrimidine	Yes	0.258178411	0.258178411	
ERR2711059	I 4,[5],12:i:-	aac(6')-Iaa_1_ NC_003197	Aminoglycoside	Yes	0.961872926	0.961872926	-
		aph(6)-Id_1_ M28829	Aminoglycoside	Yes	4.005033604	4.005033604	
		aph(3'')-Ib_5_ AF321551	Aminoglycoside	Yes	4.422173607	4.422173607	
		aac(3)-IIa_4_ L22613	Aminoglycoside	No			
		blaTEM-1B_1_ AY458016	Beta-lactam	Yes	4.024465686	4.024465686	
		blaCTX-M-55_1_ DQ810789	Beta-lactam	No			
		mcr-3.1_1_ KY924928	Polymyxin	Yes	4.193219563	4.193219563	
		floR_2_AF118107	Phenicol	Yes	2.533574738	2.533574738	
		catA2_1_X53796	Phenicol	Yes	1.56508342	1.56508342	
		qnrS1_1_AB187515	Fluoroquinolone	Yes	2.839290037	2.839290037	
		sul2_3_HQ840942	Sulphonamide	different variant (sul2_2_AY034138)	4.455295786	4.455295786	
		sul2_9_FJ197818	Sulphonamide				
		tet(A)_6_AF534183	Tetracycline	Yes	4.05867582	4.05867582	
		tet(B)_2_AF326777	Tetracycline	Yes	3.01290949	3.01290949	

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711060	Newport	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-
		aph(6)-Id_1_M28829	Aminoglycoside	Yes	0.299886083	0.299886083	
		aph(3')-Ia_3_EF015636	Aminoglycoside	different variant (aph(3')-Ia_1_V00359)	0.001136258	0.001136258	
		aadA22_1_AM261837	Aminoglycoside	No			
		aac(3)-IIa_4_L22613	Aminoglycoside	Yes	12.79994127	12.80293856	
		blaTEM-1B_1_AY458016	Beta-lactam	No			
		blaCTX-M-55_1_DQ810789	Beta-lactam	No			
		lnu(F)_1_EU118119	Lincosamide	No			
		mph(A)_1_D16251	Macrolide	No			
		floR_2_AF118107	Phenicol	Yes	0.240109712	0.240109712	
		qnrS1_1_AB187515	Fluoroquinolone	Yes	11.95631681	11.95631681	
		ARR-2_1_HQ141279	Rifamycin	No			
		ARR-3_4_FM207631	Rifamycin	Yes	0.083789211	0.481303135	
		sul3_2_AJ459418	Sulphonamide	No			
		tet(A)_6_AF534183	Tetracycline	No			
		dfrA14_5_DQ388123	Diaminopyrimidine	No			
		dfrA14_1_KF921535	Diaminopyrimidine	No			

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711061	Paratyphi B var. Java monophasic	aac(6')-Iaa_1_NC_003197	Aminoglycoside	Yes	1.166501022	1.166501022	-
ERR2711062	I 4,[5],12:i:-	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-
		aph(6)-Id_1_M28829	Aminoglycoside	Yes	0.175346728	0.175346728	
		aph(3')-Ia_3_EF015636	Aminoglycoside	No			
		aadA22_1_AM261837	Aminoglycoside	No			
		aph(3'')-Ib_5_AF321551	Aminoglycoside	Yes	0.181137959	0.181137959	
		aac(3)-IIa_4_L22613	Aminoglycoside	Yes	3.588696963	3.657143931	
		blaTEM-1B_1_AY458016	Beta-lactam	Yes	9.190267767	9.190267767	
		blaCTX-M-55_1_DQ810789	Beta-lactam	Yes	7.346959198	7.346959198	
		lnu(F)_1_EU118119	Lincosamide	No			
		mph(A)_1_D16251	Macrolide	Yes	2.413405601	2.870860512	
		qnrS1_1_AB187515	Fluoroquinolone	No			
		ARR-2_1_HQ141279	Rifamycin	No			
		ARR-3_4_FM207631	Rifamycin	No			
		sul2_3_HQ840942	Sulphonamide	different variants (sul2_6_FN995456, sul2_9_FJ197818)		0.155537299	

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711062	I 4,[5],12:i:-	sul3_2_AJ459418	Sulphonamide	No			-
		tet(A)_6_AF534183	Tetracycline	Yes	0.308022595	0.308022595	
		tet(B)_2_AF326777	Tetracycline	No			
		dfrA14_5_ DQ388123	Diaminopyrimidine	No			
		dfrA14_1_ KF921535	Diaminopyrimidine	Yes	0.054437566	0.054437566	
ERR2711063	Hvittingfoss	aac(6')-Iaa_1_ NC_003197	Aminoglycoside	No			-
ERR2711064	Typhimurium	aac(6')-Iaa_1_ NC_003197	Aminoglycoside	No			-
		aadA2_1_ NC_010870	Aminoglycoside	Yes	0.004046167	0.004046167	
		ant(3'')-Ia_1_ X02340	Aminoglycoside	Yes	1.123347158	1.123347158	
		aadA1_3_ JQ414041	Aminoglycoside	No			
		blaTEM-1B_1_ AY458016	Beta-lactam	Yes	6.324976233	6.324976233	
		cmlA1_1_M64556	Phenicol	No			
		floR_2_AF118107	Phenicol	Yes	0.068615416	0.068615416	
		qnrS1_1_ AB187515	Fluoroquinolone	Yes	0.003077939	0.003077939	
		sul2_2_AY034138	Sulphonamide	Yes	6.920569488	7.151305257	
		sul3_2_AJ459418	Sulphonamide	No			
		tet(A)_6_AF534183	Tetracycline	Yes	9.834767619	11.11173522	
		dfrA12_8_ AM040708	Diaminopyrimidine	different variant (dfrA12_4_EU650399)	0.002842078	0.002842078	

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711065	I 4,[5],12:i:-	aac(6')-Iaa_1_NC_003197	Aminoglycoside	Yes	0.12404847	0.12404847	-
		aph(6)-Id_1_M28829	Aminoglycoside	Yes	0.960082441	7.12970683	
		aph(3'')-Ib_5_AF321551	Aminoglycoside	Yes	6.816103635	6.816103635	
		blaTEM-1B_1_AY458016	Beta-lactam	Yes	1.753853707	1.753853707	
		blaCTX-M-55_1_DQ810789	Beta-lactam	Yes	0.163077748	0.163077748	
		qnrS1_1_AB187515	Fluoroquinolone	Yes	0.198751005	0.198751005	
		sul2_2_AY034138	Sulphonamide	Yes	6.695068943	6.695068943	
		tet(A)_6_AF534183	Tetracycline	Yes	8.74238291	8.74238291	
ERR2711066	I 4,[5],12:i:-	aac(6')-Iaa_1_NC_003197	Aminoglycoside	Yes	0.283873586	0.283873586	-
		aph(6)-Id_4_CP000971	Aminoglycoside	Yes	0.203845023	13.07805691	
		aph(3'')-Ib_5_AF321551	Aminoglycoside	Yes	13.40677378	13.40829515	
		blaCTX-M-14_1_AF252622	Beta-lactam	Yes	1.714382567	1.714382567	
		blaTEM-1B_1_AY458016	Beta-lactam	Yes	0.796534721	0.796534721	
		floR_2_AF118107	Phenicol	Yes	1.212080236	1.212080236	
		qnrS1_1_AB187515	Fluoroquinolone	Yes	0.648296632	0.648296632	
		sul2_2_AY034138	Sulphonamide	Yes	0.215784418	13.64905462	
		tet(A)_6_AF534183	Tetracycline	Yes	8.296964891	9.242776054	
		tet(B)_2_AF326777	Tetracycline	Yes	0.858994851	0.858994851	

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711067	Stanley	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-
ERR2711068	I 4,[5],12:i:-	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-
		aadA2_1_NC_010870	Aminoglycoside	No			
		aph(6)-Id_1_M28829	Aminoglycoside	Yes	6.343500802	6.417087442	
		aadA1_3_JQ414041	Aminoglycoside	No			
		aph(3'')-Ib_5_AF321551	Aminoglycoside	Yes	6.442363522	6.447583623	
		blaTEM-1B_1_AY458016	Beta-lactam	Yes	0.08271993	0.08271993	
		mcr-1.1_1_KP347127	Polymyxin	Yes	0.006231927	0.006231927	
		cmlA1_1_M64556	Phenicol	Yes	0.005910325	0.005910325	
		floR_2_AF118107	Phenicol	Yes	9.236108798	9.236108798	
		qnrS1_1_AB187515	Fluoroquinolone	Yes	0.04942175	0.04942175	
		sul2_2_AY034138	Sulphonamide	Yes	6.163852994	6.444070034	
		sul2_6_FN995456	Sulphonamide	Yes	0.010596026		
		sul3_2_AJ459418	Sulphonamide	Yes	7.457827792	7.457827792	
		tet(B)_2_AF326777	Tetracycline	Yes	0.06236526	0.06236526	
		dfrA12_8_AM040708	Diaminopyrimidine	different variant (dfrA12_7_AB196348)	0.001111018	0.001111018	
ERR2711069	I 4,[5],12:i:-	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711069	I 4,[5],12:i:-	aph(6)-Id_1_M28829	Aminoglycoside	different variant (aph(6)-Id_4_CP000971)	2.962765841	2.962765841	-
		aph(3'')-Ib_5_AF321551	Aminoglycoside	Yes	4.380016112	4.380016112	
		blaTEM-1B_1_AY458016	Beta-lactam	Yes	0.60166608	0.60166608	
		floR_2_AF118107	Phenicol	Yes	0.485166276	0.485166276	
		qnrS1_1_AB187515	Fluoroquinolone	Yes	1.5530134	1.5530134	
		sul2_3_HQ840942	Sulphonamide	different variant (sul2_2_AY034138)	2.970381326	2.970381326	
		tet(A)_6_AF534183	Tetracycline	Yes	6.586806044	6.586806044	
		dfrA14_5_DQ388123	Diaminopyrimidine	No			
		dfrA14_1_KF921535	Diaminopyrimidine	No			
ERR2711070	Albany	aac(6')-Iaa_1_NC_003197	Aminoglycoside	Yes	0.010199037	0.010199037	*gyrA* (S83F)
		blaCARB-2_1_M69058	Beta-lactam	Yes	0.047139535	0.047139535	
		floR_2_AF118107	Phenicol	Yes	5.292150243	5.292150243	
		sul1_2_U12338	Sulphonamide	Yes	0.141239546	0.168679571	
		tet(G)_2_AF133140	Tetracycline	Yes	0.049503109	0.049503109	
		dfrA1_8_X00926	Diaminopyrimidine	Yes	0.061589329	0.061589329	
ERR2711071	Typhimurium	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-
		aph(6)-Id_1_M28829	Aminoglycoside	Yes	0.154680583	0.154680583	

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711071	Typhimurium	aph(3'')-Ib_5_ AF321551	Aminoglycoside	Yes	0.147934649	0.147934649	-
		blaTEM-1B_1_ AY458016	Beta-lactam	Yes	0.269699664	0.269699664	
		floR_2_AF118107	Phenicol	No			
		sul2_3_HQ840942	Sulphonamide	different variant (sul2_9_FJ197818)			
		sul2_9_FJ197818	Sulphonamide	Yes	0.072723222	0.072723222	
		tet(A)_6_AF534183	Tetracycline	Yes	0.202770577	0.202770577	
		tet(B)_2_AF326777	Tetracycline	Yes	0.05049848	0.05049848	
ERR2711072	Enteritidis	aac(6')-Iaa_1_ NC_003197	Aminoglycoside	No			*gyrA* (D87N)
		aph(6)-Id_4_ CP000971	Aminoglycoside	No			
		aph(3'')-Ib_5_ AF321551	Aminoglycoside	No			
		blaTEM-1B_1_ AY458016	Beta-lactam	No			
		sul2_2_AY034138	Sulphonamide	No			
		tet(A)_4_AJ517790	Tetracycline	No			
		tet(A)_6_AF534183	Tetracycline	No			
ERR2711073	I 4,[5],12:i:-	aac(6')-Iaa_1_ NC_003197	Aminoglycoside	Yes	0.078790867	0.078790867	-
		aph(6)-Id_1_ M28829	Aminoglycoside	Yes	1.918235938	1.923238448	
		aph(3'')-Ib_5_ AF321551	Aminoglycoside	Yes	1.976303727	1.976303727	
		blaTEM-1B_1_ AY458016	Beta-lactam	Yes	0.677523481	0.677523481	

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711073	I 4,[5],12:i:-	floR_2_AF118107	Phenicol	Yes	2.150443144	2.150443144	-
		qnrS1_1_ AB187515	Fluoroquinolone	Yes	0.968150533	0.968150533	
		sul2_3_HQ840942	Sulphonamide	different variants (sul2_2_AY034138, sul2_5_AY524415, sul2_6_FN995456, sul2_9_FJ197818)		2.104828903	
		tet(A)_4_AJ517790	Tetracycline	Yes	0.234335761	4.696171722	
		tet(A)_6_AF534183	Tetracycline	Yes	4.461835961		
		dfrA14_5_ DQ388123	Diaminopyrimidine	Yes	0.14842424	1.021178347	
		dfrA14_1_ KF921535	Diaminopyrimidine	Yes	0.872754108		
ERR2711074	I 4,[5],12:i:-	aac(6')-Iaa_1_ NC_003197	Aminoglycoside	No			-
		aph(6)-Id_4_ CP000971	Aminoglycoside	Yes	10.34702779	11.46385391	
		aph(3'')-Ib_5_ AF321551	Aminoglycoside	Yes	10.90863154	10.90863154	
		blaTEM-1B_1_ AY458016	Beta-lactam	Yes	13.00683877	13.00683877	
		floR_2_AF118107	Phenicol	No			
		qnrS1_1_ AB187515	Fluoroquinolone	No			
		sul2_2_AY034138	Sulphonamide	Yes	10.9479816	11.51037108	
		tet(A)_6_AF534183	Tetracycline	Yes	14.56899962	16.63524953	
		tet(B)_2_AF326777	Tetracycline	No			

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711075	I 4,[5],12:i:-	aac(6')-Iaa_1_NC_003197	Aminoglycoside	Yes	0.00280451	0.00280451	-
		aph(6)-Id_4_CP000971	Aminoglycoside	Yes	0.010491886	1.164780107	
		aph(3'')-Ib_5_AF321551	Aminoglycoside	Yes	1.149503473	1.166349291	
		blaTEM-1B_1_AY458016	Beta-lactam	Yes	2.5709875	2.5709875	
		floR_2_AF118107	Phenicol	Yes	0.048467483	0.048467483	
		qnrS1_1_AB187515	Fluoroquinolone	Yes	0.097288843	0.097288843	
		sul2_2_AY034138	Sulphonamide	Yes	3.104659002	3.333789029	
		tet(A)_6_AF534183	Tetracycline	Yes	0.123530373	0.123530373	
		dfrA14_5_DQ388123	Diaminopyrimidine	Yes	0.123575104	0.831959599	
		dfrA14_1_KF921535	Diaminopyrimidine	Yes	0.708384495		
ERR2711076	I 4,[5],12:i:-	aac(6')-Iaa_1_NC_003197	Aminoglycoside	No			-
		aph(6)-Id_1_M28829	Aminoglycoside	Yes	1.553299818	14.99226722	
		aph(3'')-Ib_5_AF321551	Aminoglycoside	Yes	14.88344911	14.88344911	
		blaTEM-1B_1_AY458016	Beta-lactam	Yes	0.010529683	0.010529683	
		floR_2_AF118107	Phenicol	No			
		qnrS1_1_AB187515	Fluoroquinolone	Yes	0.011248783	0.011248783	
		sul2_2_AY034138	Sulphonamide	Yes	13.88880019	14.46062717	

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR

Genome	Serovar	KMA					StarAMR
		Resistance determinant ID	Class summary	Identified in metagenome (KMA)	Proportion of sample resistome	Proportion of sample resistome (all variants combined)	Point mutation
ERR2711076	I 4,[5],12:i:-	tet(A)_6_AF534183	Tetracycline	Yes	20.03210788	22.69311764	-
		dfrA14_5_ DQ388123	Diaminopyrimidine	different variant (dfrA14_2_Z50805)	0.007268332	0.007268332	
		dfrA14_1_ KF921535	Diaminopyrimidine	different variant (dfrA14_2_Z50805)	0.007268332		
ERR2711077	Saintpaul	aac(6')-Iaa_1_ NC_003197	Aminoglycoside	No			*gyrA* (S83Y)
		aph(6)-Id_1_ M28829	Aminoglycoside	Yes	0.066014293	0.066014293	
		aph(3')-Ia_3_ EF015636	Aminoglycoside	different variant (aph(3')-Ia_1_V00359)	0.025283508	0.025283508	
		aph(3'')-Ib_5_ AF321551	Aminoglycoside	Yes	0.062656679	0.062656679	
		aac(3)-IIa_4_ L22613	Aminoglycoside	Yes	0.034438639	0.034438639	
		blaTEM-1B_1_ AY458016	Beta-lactam	Yes	0.540185418	0.540185418	
		lnu(F)_1_ EU118119	Lincosamide	No			
		mph(A)_1_D16251	Macrolide	Yes	0.043043933	0.043043933	
		floR_2_AF118107	Phenicol	Yes	0.0433504	0.0433504	
		qnrS1_1_ AB187515	Fluoroquinolone	No			
		ARR-2_1_ HQ141279	Rifamycin	No			
		ARR-3_4_ FM207631	Rifamycin	No			
		sul2_2_AY034138	Sulphonamide	Yes	0.039687446	0.039687446	
		tet(A)_6_AF534183	Tetracycline	Yes	0.079729244	0.079729244	

Appendix 9: Antimicrobial resistance determinants identified in the 28 *Salmonella enterica* genomes with KMA and StarAMR							
Genome	**Serovar**	**KMA**					**StarAMR**
		Resistance determinant ID	**Class summary**	**Identified in metagenome (KMA)**	**Proportion of sample resistome**	**Proportion of sample resistome (all variants combined)**	**Point mutation**
ERR2711077	Saintpaul	dfrA14_5_ DQ388123	Diaminopyrimidine	No			*gyrA* (S83Y)
		dfrA14_1_ KF921535	Diaminopyrimidine	No			

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
metagenome_6916498	Case	*Homo sapiens*	40.7
metagenome_6916498	Case	Human gammaherpesvirus 4	20.7
metagenome_6916498	Case	*Escherichia coli*	9.61
metagenome_6916498	Case	*Phocaeicola dorei*	3.51
metagenome_6916498	Case	*Klebsiella pneumoniae*	2.41
metagenome_6916498	Case	*Streptococcus salivarius*	1.93
metagenome_6916498	Case	*Megamonas funiformis*	1.73
metagenome_6916498	Case	*Bacteroides fragilis*	1.48
metagenome_6916498	Case	*Bifidobacterium longum*	1.44
metagenome_6916498	Case	*Bifidobacterium breve*	1.30
metagenome_6916498	Case	*[Ruminococcus] gnavus*	0.981
metagenome_6916498	Case	*Salmonella enterica*	0.942
metagenome_6916498	Case	*Bacteroides ovatus*	0.638
metagenome_6916498	Case	*Bacteroides zhangwenhongii*	0.491
metagenome_6916498	Case	*Bacteroides caccae*	0.428
metagenome_6916498	Case	*Plasmodium falciparum*	0.399
metagenome_6916498	Case	*Lactobacillus acidophilus*	0.360
metagenome_6916498	Case	*Pan troglodytes*	0.350
metagenome_6916498	Case	*Faecalibacterium prausnitzii*	0.347
metagenome_6916498	Case	*Streptococcus thermophilus*	0.281
metagenome_6916499	Case	*Enterococcus faecium*	85.1
metagenome_6916499	Case	*Bifidobacterium longum*	3.68
metagenome_6916499	Case	*Bacteroides thetaiotaomicron*	2.14
metagenome_6916499	Case	*Homo sapiens*	1.96
metagenome_6916499	Case	*Bifidobacterium breve*	1.94
metagenome_6916499	Case	*Escherichia coli*	1.10
metagenome_6916499	Case	*Bacteroides fragilis*	0.895
metagenome_6916499	Case	*Klebsiella pneumoniae*	0.840
metagenome_6916499	Case	*Candida albicans*	0.252
metagenome_6916499	Case	*Bacteroides ovatus*	0.172
metagenome_6916499	Case	*[Ruminococcus] gnavus*	0.0862
metagenome_6916499	Case	*Salmonella enterica*	0.0849
metagenome_6916499	Case	*Streptococcus salivarius*	0.0834
metagenome_6916499	Case	*Bacteroides caccae*	0.0757
metagenome_6916499	Case	*Enterococcus lactis*	0.0431
metagenome_6916499	Case	*Bacteroides xylanisolvens*	0.0410
metagenome_6916499	Case	*Rothia mucilaginosa*	0.0357
metagenome_6916499	Case	*Prevotella oris*	0.0355
metagenome_6916499	Case	*Phocaeicola vulgatus*	0.0336
metagenome_6916499	Case	Bifidobacteriaceae bacterium	0.0328
metagenome_6916500	Case	*Escherichia coli*	50.3
metagenome_6916500	Case	*Enterococcus faecalis*	23.7
metagenome_6916500	Case	*Staphylococcus aureus*	19.7
metagenome_6916500	Case	*Enterobacter hormaechei*	3.05
metagenome_6916500	Case	*Salmonella enterica*	0.598
metagenome_6916500	Case	*Klebsiella pneumoniae*	0.328
metagenome_6916500	Case	*Homo sapiens*	0.319

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
metagenome_6916500	Case	*Enterobacter cloacae*	0.0862
metagenome_6916500	Case	*Shigella flexneri*	0.0811
metagenome_6916500	Case	*Clostridium perfringens*	0.0720
metagenome_6916500	Case	*Streptococcus salivarius*	0.0605
metagenome_6916500	Case	*Enterobacter* sp. BIDMC100	0.0535
metagenome_6916500	Case	*Enterobacter* sp. CRENT-193	0.0499
metagenome_6916500	Case	*Siphoviridae* sp.	0.0440
metagenome_6916500	Case	*Schaalia odontolytica*	0.0428
metagenome_6916500	Case	*Actinomyces oris*	0.0324
metagenome_6916500	Case	uncultured *Clostridium* sp.	0.0310
metagenome_6916500	Case	*Streptococcus mitis*	0.0301
metagenome_6916500	Case	*Actinomyces* sp. HMT 175	0.0255
metagenome_6916500	Case	*Shigella boydii*	0.0254
metagenome_6916501	Case	*Bifidobacterium pseudocatenulatum*	24.0
metagenome_6916501	Case	*Bifidobacterium bifidum*	16.7
metagenome_6916501	Case	*Streptococcus pasteurianus*	14.6
metagenome_6916501	Case	*Veillonella atypica*	13.8
metagenome_6916501	Case	*Limosilactobacillus vaginalis*	6.10
metagenome_6916501	Case	*Streptococcus salivarius*	4.61
metagenome_6916501	Case	*Bifidobacterium catenulatum*	2.24
metagenome_6916501	Case	*Streptococcus gallolyticus*	1.64
metagenome_6916501	Case	*Lactobacillus paragasseri*	1.60
metagenome_6916501	Case	*Streptococcus lutetiensis*	1.54
metagenome_6916501	Case	*Lactobacillus gasseri*	0.97
metagenome_6916501	Case	*Bifidobacterium adolescentis*	0.706
metagenome_6916501	Case	*Bifidobacterium longum*	0.651
metagenome_6916501	Case	*Homo sapiens*	0.472
metagenome_6916501	Case	*Streptococcus* sp. FDAARGOS_192	0.377
metagenome_6916501	Case	*Limosilactobacillus fermentum*	0.343
metagenome_6916501	Case	*Veillonella nakazawae*	0.341
metagenome_6916501	Case	*Enterobacter hormaechei*	0.324
metagenome_6916501	Case	*Streptococcus* sp. LPB0220	0.322
metagenome_6916501	Case	*Bifidobacterium animalis*	0.310
metagenome_6916502	Case	*Phocaeicola vulgatus*	86.9
metagenome_6916502	Case	*Bacteroides fragilis*	1.46
metagenome_6916502	Case	*Bacteroides uniformis*	1.13
metagenome_6916502	Case	*Bacteroides eggerthii*	1.10
metagenome_6916502	Case	*Phocaeicola dorei*	1.03
metagenome_6916502	Case	*Parabacteroides merdae*	0.486
metagenome_6916502	Case	*Siphoviridae* sp. ctF6o6	0.483
metagenome_6916502	Case	uncultured organism	0.418
metagenome_6916502	Case	*Escherichia coli*	0.408
metagenome_6916502	Case	*Candida albicans*	0.405
metagenome_6916502	Case	*Bacteroides thetaiotaomicron*	0.302
metagenome_6916502	Case	*Bacteroides* sp. CACC 737	0.293
metagenome_6916502	Case	uncultured bacterium	0.263
metagenome_6916502	Case	*Bacteroides cellulosilyticus*	0.151

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
metagenome_6916502	Case	*Parabacteroides distasonis*	0.150
metagenome_6916502	Case	*Bacteroides xylanisolvens*	0.149
metagenome_6916502	Case	*Bacteroides ovatus*	0.109
metagenome_6916502	Case	*Siphoviridae* sp.	0.0912
metagenome_6916502	Case	*Streptococcus mutans*	0.0864
metagenome_6916502	Case	*Homo sapiens*	0.0665
metagenome_6916503	Case	*Phocaeicola dorei*	25.2
metagenome_6916503	Case	*Megamonas funiformis*	11.6
metagenome_6916503	Case	*[Ruminococcus] gnavus*	6.93
metagenome_6916503	Case	*Bacteroides ovatus*	5.25
metagenome_6916503	Case	*Escherichia coli*	3.94
metagenome_6916503	Case	*Bacteroides zhangwenhongii*	3.44
metagenome_6916503	Case	*Streptococcus salivarius*	3.18
metagenome_6916503	Case	*Bacteroides fragilis*	2.58
metagenome_6916503	Case	*Faecalibacterium prausnitzii*	2.56
metagenome_6916503	Case	*Bacteroides caccae*	2.29
metagenome_6916503	Case	*Bacteroides uniformis*	1.16
metagenome_6916503	Case	*Haemophilus parainfluenzae*	0.944
metagenome_6916503	Case	*Blautia wexlerae*	0.844
metagenome_6916503	Case	*Phocaeicola vulgatus*	0.832
metagenome_6916503	Case	*Veillonella* sp. S12025-13	0.820
metagenome_6916503	Case	*Streptococcus pasteurianus*	0.803
metagenome_6916503	Case	*Siphoviridae* sp.	0.764
metagenome_6916503	Case	*Bacteroides* sp. ZJ-18	0.690
metagenome_6916503	Case	*Enterococcus faecalis*	0.671
metagenome_6916503	Case	*Veillonella nakazawae*	0.619
metagenome_6916504	Case	*Limosilactobacillus fermentum*	60.2
metagenome_6916504	Case	*Veillonella atypica*	8.23
metagenome_6916504	Case	*Streptococcus salivarius*	6.16
metagenome_6916504	Case	*Lacticaseibacillus paracasei*	5.95
metagenome_6916504	Case	*Homo sapiens*	5.50
metagenome_6916504	Case	*Streptococcus* sp. LPB0220	0.785
metagenome_6916504	Case	*Phocaeicola dorei*	0.763
metagenome_6916504	Case	*Streptococcus thermophilus*	0.618
metagenome_6916504	Case	*Streptococcus* sp. FDAARGOS_192	0.506
metagenome_6916504	Case	*Streptococcus parasanguinis*	0.489
metagenome_6916504	Case	*Siphoviridae* sp.	0.364
metagenome_6916504	Case	*Megamonas funiformis*	0.361
metagenome_6916504	Case	*Veillonella nakazawae*	0.290
metagenome_6916504	Case	*Prevotella melaninogenica*	0.262
metagenome_6916504	Case	*Veillonella* sp. OK1	0.246
metagenome_6916504	Case	*Rothia mucilaginosa*	0.241
metagenome_6916504	Case	*Veillonella dispar*	0.227
metagenome_6916504	Case	*Streptococcus* sp. ZB199	0.221
metagenome_6916504	Case	*Streptococcus pasteurianus*	0.214
metagenome_6916504	Case	*[Ruminococcus] gnavus*	0.201
metagenome_6916505	Case	*Escherichia coli*	45.3

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
metagenome_6916505	Case	*Salmonella enterica*	27.7
metagenome_6916505	Case	*Staphylococcus haemolyticus*	10.7
metagenome_6916505	Case	*Homo sapiens*	10.2
metagenome_6916505	Case	*Staphylococcus aureus*	0.973
metagenome_6916505	Case	*Klebsiella pneumoniae*	0.835
metagenome_6916505	Case	*Shigella flexneri*	0.428
metagenome_6916505	Case	*Saccharomyces cerevisiae*	0.428
metagenome_6916505	Case	*Bifidobacterium breve*	0.254
metagenome_6916505	Case	*Shigella boydii*	0.215
metagenome_6916505	Case	*Bifidobacterium longum*	0.134
metagenome_6916505	Case	*Bacteroides fragilis*	0.134
metagenome_6916505	Case	*Corynebacterium tuberculostearicum*	0.129
metagenome_6916505	Case	*Shigella dysenteriae*	0.0796
metagenome_6916505	Case	*Phocaeicola dorei*	0.0607
metagenome_6916505	Case	*Prevotella buccalis*	0.0593
metagenome_6916505	Case	*Spirometra erinaceieuropaei*	0.0441
metagenome_6916505	Case	*Megamonas funiformis*	0.0436
metagenome_6916505	Case	*Pan troglodytes*	0.0423
metagenome_6916505	Case	*Gari tellinella*	0.0409
metagenome_6916506	Case	*Escherichia coli*	88.1
metagenome_6916506	Case	*Homo sapiens*	3.66
metagenome_6916506	Case	*Shigella flexneri*	1.12
metagenome_6916506	Case	Bacteriophage sp.	1.04
metagenome_6916506	Case	*Salmonella enterica*	0.791
metagenome_6916506	Case	*Prevotella buccalis*	0.406
metagenome_6916506	Case	*Bifidobacterium pseudocatenulatum*	0.373
metagenome_6916506	Case	*Veillonella parvula*	0.319
metagenome_6916506	Case	*Bifidobacterium breve*	0.292
metagenome_6916506	Case	*Salmonella* phage D10	0.270
metagenome_6916506	Case	*Shigella boydii*	0.252
metagenome_6916506	Case	*Siphoviridae* sp.	0.246
metagenome_6916506	Case	*Klebsiella pneumoniae*	0.179
metagenome_6916506	Case	*Streptococcus pasteurianus*	0.172
metagenome_6916506	Case	*Shigella dysenteriae*	0.106
metagenome_6916506	Case	*Skatevirus skate*	0.0921
metagenome_6916506	Case	*Shigella sonnei*	0.0642
metagenome_6916506	Case	*Skatevirus KFSSE2*	0.0564
metagenome_6916506	Case	*Segzyvirus segz1*	0.0549
metagenome_6916506	Case	*Gryllus bimaculatus*	0.0435
metagenome_6916507	Case	*Bifidobacterium longum*	15.8
metagenome_6916507	Case	*Salmonella enterica*	12.8
metagenome_6916507	Case	*Phocaeicola vulgatus*	10.0
metagenome_6916507	Case	*Homo sapiens*	7.17
metagenome_6916507	Case	*Bacteroides uniformis*	6.69
metagenome_6916507	Case	*Herelleviridae* sp.	2.34
metagenome_6916507	Case	*Bacteroides fragilis*	1.65
metagenome_6916507	Case	uncultured bacterium	1.57

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
metagenome_6916507	Case	*Faecalibacterium prausnitzii*	1.03
metagenome_6916507	Case	*Siphoviridae* sp.	0.680
metagenome_6916507	Case	*Bifidobacterium breve*	0.647
metagenome_6916507	Case	*Escherichia coli*	0.612
metagenome_6916507	Case	*Fusobacterium mortiferum*	0.504
metagenome_6916507	Case	*[Clostridium] innocuum*	0.410
metagenome_6916507	Case	*Enterobacter hormaechei*	0.400
metagenome_6916507	Case	*Clostridioides difficile*	0.265
metagenome_6916507	Case	*Bacteroides* sp. HF-162	0.237
metagenome_6916507	Case	*Phocaeicola dorei*	0.217
metagenome_6916507	Case	*Veillonella dispar*	0.177
metagenome_6916507	Case	*Bacteroides humanifaecis*	0.175
metagenome_6916508	Case	*Klebsiella* sp. KP20-425-1	52.5
metagenome_6916508	Case	*Klebsiella pneumoniae*	30.1
metagenome_6916508	Case	*Acinetobacter baumannii*	7.62
metagenome_6916508	Case	*Citrobacter* sp. RHBSTW-00053	7.05
metagenome_6916508	Case	*Escherichia coli*	0.656
metagenome_6916508	Case	*Candida albicans*	0.388
metagenome_6916508	Case	*Candida tropicalis*	0.292
metagenome_6916508	Case	*Klebsiella variicola*	0.101
metagenome_6916508	Case	*Enterococcus faecalis*	0.0804
metagenome_6916508	Case	*Streptococcus salivarius*	0.0782
metagenome_6916508	Case	*Salmonella enterica*	0.0778
metagenome_6916508	Case	*Enterobacter hormaechei*	0.0776
metagenome_6916508	Case	*Rothia mucilaginosa*	0.0697
metagenome_6916508	Case	*Klebsiella quasipneumoniae*	0.0591
metagenome_6916508	Case	*Streptococcus macedonicus*	0.0578
metagenome_6916508	Case	*Homo sapiens*	0.0516
metagenome_6916508	Case	*Enterobacter kobei*	0.0495
metagenome_6916508	Case	*Streptococcus gallolyticus*	0.0491
metagenome_6916508	Case	*Streptococcus thermophilus*	0.0477
metagenome_6916508	Case	*Limosilactobacillus fermentum*	0.0421
metagenome_6916509	Case	*Bifidobacterium longum*	55.9
metagenome_6916509	Case	*Salmonella enterica*	22.3
metagenome_6916509	Case	*Limosilactobacillus fermentum*	4.45
metagenome_6916509	Case	*Streptococcus pasteurianus*	3.81
metagenome_6916509	Case	*Bifidobacterium breve*	2.24
metagenome_6916509	Case	*Veillonella parvula*	2.12
metagenome_6916509	Case	*Limosilactobacillus oris*	1.61
metagenome_6916509	Case	*Streptococcus salivarius*	1.19
metagenome_6916509	Case	*Staphylococcus aureus*	1.03
metagenome_6916509	Case	*Streptococcus gallolyticus*	0.626
metagenome_6916509	Case	*Siphoviridae* sp.	0.585
metagenome_6916509	Case	*Streptococcus lutetiensis*	0.475
metagenome_6916509	Case	*Homo sapiens*	0.314
metagenome_6916509	Case	*Streptococcus vestibularis*	0.198
metagenome_6916509	Case	*Streptococcus thermophilus*	0.184

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
metagenome_6916509	Case	*Limosilactobacillus vaginalis*	0.118
metagenome_6916509	Case	*Bifidobacterium bifidum*	0.0995
metagenome_6916509	Case	*Streptococcus equinus*	0.0955
metagenome_6916509	Case	*Streptococcus* sp. FDAARGOS_192	0.0825
metagenome_6916509	Case	*Bifidobacterium catenulatum*	0.0776
metagenome_6916510	Case	*Bifidobacterium longum*	46.5
metagenome_6916510	Case	*Bacteroides fragilis*	23.8
metagenome_6916510	Case	*Escherichia coli*	11.4
metagenome_6916510	Case	*Megamonas funiformis*	3.94
metagenome_6916510	Case	*Bifidobacterium breve*	1.60
metagenome_6916510	Case	Bifidobacteriaceae bacterium	0.876
metagenome_6916510	Case	*Streptococcus salivarius*	0.854
metagenome_6916510	Case	*Akkermansia muciniphila*	0.652
metagenome_6916510	Case	*Acidaminococcus intestini*	0.515
metagenome_6916510	Case	*Enterococcus raffinosus*	0.492
metagenome_6916510	Case	*Streptococcus lactarius*	0.448
metagenome_6916510	Case	*Erysipelatoclostridium ramosum*	0.390
metagenome_6916510	Case	*Shigella flexneri*	0.273
metagenome_6916510	Case	*Megamonas hypermegale*	0.235
metagenome_6916510	Case	*Streptococcus thermophilus*	0.207
metagenome_6916510	Case	*Siphoviridae* sp.	0.203
metagenome_6916510	Case	*Bifidobacterium bifidum*	0.135
metagenome_6916510	Case	*Myoviridae* sp.	0.131
metagenome_6916510	Case	*[Clostridium] symbiosum*	0.125
metagenome_6916510	Case	*Streptococcus* sp. FDAARGOS_192	0.108
metagenome_6916511	Case	*Staphylococcus haemolyticus*	84.8
metagenome_6916511	Case	*Enterococcus faecium*	2.38
metagenome_6916511	Case	*Homo sapiens*	2.11
metagenome_6916511	Case	*Salmonella enterica*	1.82
metagenome_6916511	Case	*Streptococcus* sp. LPB0220	1.09
metagenome_6916511	Case	*Streptococcus parasanguinis*	0.750
metagenome_6916511	Case	*Staphylococcus aureus*	0.701
metagenome_6916511	Case	*Haemophilus* sp. oral taxon 036	0.695
metagenome_6916511	Case	*Haemophilus seminalis*	0.687
metagenome_6916511	Case	*Haemophilus haemolyticus*	0.635
metagenome_6916511	Case	*Streptococcus oralis*	0.404
metagenome_6916511	Case	*Streptococcus* sp. ZB199	0.313
metagenome_6916511	Case	*Bifidobacterium breve*	0.287
metagenome_6916511	Case	*Haemophilus influenzae*	0.211
metagenome_6916511	Case	*Corynebacterium tuberculostearicum*	0.190
metagenome_6916511	Case	*Streptococcus australis*	0.129
metagenome_6916511	Case	*Streptococcus mitis*	0.121
metagenome_6916511	Case	*Staphylococcus hominis*	0.113
metagenome_6916511	Case	*Escherichia coli*	0.103
metagenome_6916511	Case	*Streptococcus lactarius*	0.0943
metagenome_6916512	Case	*Escherichia coli*	93.8
metagenome_6916512	Case	*Streptococcus pasteurianus*	1.32

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
metagenome_6916512	Case	*Bacteroides fragilis*	1.23
metagenome_6916512	Case	*Klebsiella pneumoniae*	0.803
metagenome_6916512	Case	*Streptococcus gallolyticus*	0.261
metagenome_6916512	Case	*Enterococcus faecalis*	0.157
metagenome_6916512	Case	*Fusobacterium mortiferum*	0.144
metagenome_6916512	Case	*Finegoldia magna*	0.141
metagenome_6916512	Case	*Shigella flexneri*	0.105
metagenome_6916512	Case	*Salmonella enterica*	0.089
metagenome_6916512	Case	*Veillonella atypica*	0.0769
metagenome_6916512	Case	*Anaerococcus obesiensis*	0.0653
metagenome_6916512	Case	*Anaerococcus vaginalis*	0.0572
metagenome_6916512	Case	*Prevotella buccalis*	0.0517
metagenome_6916512	Case	*Klebsiella quasipneumoniae*	0.0506
metagenome_6916512	Case	*Prevotella melaninogenica*	0.0458
metagenome_6916512	Case	*Escherichia albertii*	0.0456
metagenome_6916512	Case	*Homo sapiens*	0.0452
metagenome_6916512	Case	*Shigella boydii*	0.0389
metagenome_6916512	Case	*Enterococcus avium*	0.0364
metagenome_6916513	Case	*Escherichia coli*	60.1
metagenome_6916513	Case	*Streptococcus salivarius*	10.4
metagenome_6916513	Case	*Haemophilus parainfluenzae*	7.89
metagenome_6916513	Case	*Streptococcus* sp. LPB0220	3.07
metagenome_6916513	Case	*Salmonella enterica*	2.82
metagenome_6916513	Case	*Streptococcus parasanguinis*	2.25
metagenome_6916513	Case	*Fusobacterium mortiferum*	1.50
metagenome_6916513	Case	*Streptococcus* sp. ZB199	0.962
metagenome_6916513	Case	*Streptococcus* sp. FDAARGOS_192	0.855
metagenome_6916513	Case	*Acinetobacter baumannii*	0.728
metagenome_6916513	Case	*Rothia mucilaginosa*	0.725
metagenome_6916513	Case	*Homo sapiens*	0.511
metagenome_6916513	Case	*Bifidobacterium longum*	0.497
metagenome_6916513	Case	*Veillonella atypica*	0.478
metagenome_6916513	Case	*Neisseria subflava*	0.462
metagenome_6916513	Case	*Enterococcus faecium*	0.299
metagenome_6916513	Case	*Siphoviridae* sp.	0.289
metagenome_6916513	Case	*Streptococcus oralis*	0.273
metagenome_6916513	Case	*Streptococcus lactarius*	0.227
metagenome_6916513	Case	*Veillonella parvula*	0.213
metagenome_6916514	Case	*Klebsiella pneumoniae*	74.8
metagenome_6916514	Case	*Salmonella enterica*	16.5
metagenome_6916514	Case	*Bifidobacterium longum*	5.33
metagenome_6916514	Case	*Enterococcus faecalis*	1.00
metagenome_6916514	Case	*Escherichia coli*	0.677
metagenome_6916514	Case	*Bifidobacterium breve*	0.361
metagenome_6916514	Case	*Acinetobacter baumannii*	0.175
metagenome_6916514	Case	*Schaalia odontolytica*	0.127
metagenome_6916514	Case	*Salmonella bongori*	0.119

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
metagenome_6916514	Case	unidentified plasmid	0.112
metagenome_6916514	Case	*Klebsiella variicola*	0.0666
metagenome_6916514	Case	*Klebsiella quasipneumoniae*	0.0632
metagenome_6916514	Case	*Salmonella* sp.	0.0599
metagenome_6916514	Case	*Parabacteroides distasonis*	0.0588
metagenome_6916514	Case	*Siphoviridae* sp.	0.0375
metagenome_6916514	Case	*Streptococcus salivarius*	0.0315
metagenome_6916514	Case	Bifidobacteriaceae bacterium	0.0256
metagenome_6916514	Case	*Bifidobacterium catenulatum*	0.0228
metagenome_6916514	Case	*Yersinia pestis*	0.0130
metagenome_6916514	Case	*Enterobacter hormaechei*	0.0128
metagenome_6916515	Case	*Escherichia coli*	78.8
metagenome_6916515	Case	*Homo sapiens*	16.3
metagenome_6916515	Case	*Bifidobacterium breve*	0.335
metagenome_6916515	Case	*Bacteroides fragilis*	0.247
metagenome_6916515	Case	*Klebsiella pneumoniae*	0.186
metagenome_6916515	Case	*Bifidobacterium longum*	0.178
metagenome_6916515	Case	*Salmonella enterica*	0.172
metagenome_6916515	Case	*Phocaeicola dorei*	0.141
metagenome_6916515	Case	*Spirometra erinaceieuropaei*	0.140
metagenome_6916515	Case	*[Ruminococcus] gnavus*	0.112
metagenome_6916515	Case	*Staphylococcus epidermidis*	0.105
metagenome_6916515	Case	*Megamonas funiformis*	0.0907
metagenome_6916515	Case	*Pan troglodytes*	0.0851
metagenome_6916515	Case	*Prevotella buccalis*	0.0772
metagenome_6916515	Case	*Lactobacillus acidophilus*	0.0638
metagenome_6916515	Case	*Finegoldia magna*	0.0627
metagenome_6916515	Case	*Pan paniscus*	0.0604
metagenome_6916515	Case	*Enterococcus faecium*	0.0515
metagenome_6916515	Case	*Streptococcus salivarius*	0.0492
metagenome_6916515	Case	*Anaerococcus obesiensis*	0.0481
metagenome_6916516	Case	*Escherichia coli*	61.3
metagenome_6916516	Case	*Salmonella enterica*	14.3
metagenome_6916516	Case	*Bacteroides fragilis*	3.88
metagenome_6916516	Case	*Streptococcus salivarius*	1.29
metagenome_6916516	Case	*Bifidobacterium catenulatum*	1.12
metagenome_6916516	Case	*Phocaeicola vulgatus*	1.04
metagenome_6916516	Case	*Bacteroides humanifaecis*	0.837
metagenome_6916516	Case	*Klebsiella pneumoniae*	0.809
metagenome_6916516	Case	*Bifidobacterium breve*	0.774
metagenome_6916516	Case	*Megamonas funiformis*	0.718
metagenome_6916516	Case	*Bifidobacterium longum*	0.661
metagenome_6916516	Case	*Veillonella* sp. S12025-13	0.563
metagenome_6916516	Case	*Faecalibacterium prausnitzii*	0.526
metagenome_6916516	Case	*Bacteroides* sp. CACC 737	0.521
metagenome_6916516	Case	*Shigella flexneri*	0.500
metagenome_6916516	Case	*Bacteroides* sp. A1C1	0.481

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
metagenome_6916516	Case	*Bacteroides* sp. HF-162	0.427
metagenome_6916516	Case	*Veillonella nakazawae*	0.420
metagenome_6916516	Case	*Schaalia odontolytica*	0.389
metagenome_6916516	Case	*Bifidobacterium pseudocatenulatum*	0.335
metagenome_6916517	Case	*Escherichia coli*	82.8
metagenome_6916517	Case	*Homo sapiens*	5.97
metagenome_6916517	Case	*Enterococcus faecalis*	3.85
metagenome_6916517	Case	*Salmonella enterica*	2.56
metagenome_6916517	Case	*Acinetobacter baumannii*	2.10
metagenome_6916517	Case	*Bifidobacterium longum*	0.473
metagenome_6916517	Case	*Klebsiella pneumoniae*	0.455
metagenome_6916517	Case	*Shigella flexneri*	0.252
metagenome_6916517	Case	*Escherichia fergusonii*	0.155
metagenome_6916517	Case	*Bifidobacterium breve*	0.105
metagenome_6916517	Case	*Acinetobacter nosocomialis*	0.0871
metagenome_6916517	Case	*Enterococcus faecium*	0.0744
metagenome_6916517	Case	*Shigella boydii*	0.0580
metagenome_6916517	Case	*Bacteroides fragilis*	0.0570
metagenome_6916517	Case	*Spirometra erinaceieuropaei*	0.0385
metagenome_6916517	Case	*Phocaeicola dorei*	0.0325
metagenome_6916517	Case	*Pan troglodytes*	0.0308
metagenome_6916517	Case	*Shigella sonnei*	0.0295
metagenome_6916517	Case	*Megamonas funiformis*	0.0281
metagenome_6916517	Case	*Shigella dysenteriae*	0.0251
metagenome_6916518	Case	*Escherichia coli*	40.8
metagenome_6916518	Case	*Lactobacillus acidophilus*	31.5
metagenome_6916518	Case	*Klebsiella pneumoniae*	19.4
metagenome_6916518	Case	*Streptococcus salivarius*	2.18
metagenome_6916518	Case	*Shigella flexneri*	0.819
metagenome_6916518	Case	*Salmonella enterica*	0.678
metagenome_6916518	Case	*Klebsiella quasipneumoniae*	0.557
metagenome_6916518	Case	*Streptococcus thermophilus*	0.352
metagenome_6916518	Case	*Streptococcus* sp. FDAARGOS_192	0.349
metagenome_6916518	Case	*Shigella boydii*	0.232
metagenome_6916518	Case	*Streptococcus vestibularis*	0.202
metagenome_6916518	Case	*Streptococcus equinus*	0.176
metagenome_6916518	Case	*Shigella dysenteriae*	0.176
metagenome_6916518	Case	*Enterobacter hormaechei*	0.162
metagenome_6916518	Case	*Enterobacter cloacae*	0.141
metagenome_6916518	Case	*Punavirus P1*	0.105
metagenome_6916518	Case	*Shigella sonnei*	0.101
metagenome_6916518	Case	*Yersinia pestis*	0.0737
metagenome_6916518	Case	*Streptococcus lactarius*	0.0715
metagenome_6916518	Case	*Enterobacter roggenkampii*	0.0698
metagenome_6916519	Case	*Bacteroides fragilis*	27.8
metagenome_6916519	Case	*Escherichia coli*	23.5
metagenome_6916519	Case	*Fusobacterium mortiferum*	17.1

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
metagenome_6916519	Case	*Streptococcus pasteurianus*	5.33
metagenome_6916519	Case	*Phocaeicola vulgatus*	5.31
metagenome_6916519	Case	*Bacteroides* sp. ZJ-18	4.80
metagenome_6916519	Case	*Bifidobacterium longum*	1.95
metagenome_6916519	Case	*Bifidobacterium pseudocatenulatum*	0.987
metagenome_6916519	Case	*Sutterella wadsworthensis*	0.949
metagenome_6916519	Case	*Streptococcus gallolyticus*	0.475
metagenome_6916519	Case	*Streptococcus lutetiensis*	0.459
metagenome_6916519	Case	*Siphoviridae* sp.	0.352
metagenome_6916519	Case	uncultured bacterium	0.348
metagenome_6916519	Case	*Phocaeicola dorei*	0.335
metagenome_6916519	Case	*[Ruminococcus] gnavus*	0.261
metagenome_6916519	Case	*Homo sapiens*	0.246
metagenome_6916519	Case	*Bifidobacterium breve*	0.240
metagenome_6916519	Case	*Bacteroides ovatus*	0.191
metagenome_6916519	Case	*Bacteroides* sp. HF-162	0.145
metagenome_6916519	Case	*Herelleviridae* sp.	0.134
metagenome_6916520	Case	*Escherichia coli*	58.9
metagenome_6916520	Case	*Megamonas funiformis*	34.0
metagenome_6916520	Case	*Megamonas hypermegale*	2.08
metagenome_6916520	Case	*Erysipelatoclostridium ramosum*	1.28
metagenome_6916520	Case	*Streptococcus salivarius*	0.485
metagenome_6916520	Case	*[Ruminococcus] gnavus*	0.301
metagenome_6916520	Case	*Myoviridae* sp.	0.295
metagenome_6916520	Case	*Siphoviridae* sp.	0.142
metagenome_6916520	Case	*Streptococcus thermophilus*	0.138
metagenome_6916520	Case	*Faecalibacterium prausnitzii*	0.136
metagenome_6916520	Case	*Bifidobacterium animalis*	0.117
metagenome_6916520	Case	*Homo sapiens*	0.116
metagenome_6916520	Case	Bacteriophage sp.	0.0973
metagenome_6916520	Case	*[Clostridium] innocuum*	0.0825
metagenome_6916520	Case	*Shigella flexneri*	0.0599
metagenome_6916520	Case	*Streptococcus* sp. FDAARGOS_192	0.0428
metagenome_6916520	Case	uncultured bacterium	0.0301
metagenome_6916520	Case	*Anopheles gambiae*	0.0271
metagenome_6916520	Case	*Shigella boydii*	0.0198
metagenome_6916520	Case	*Klebsiella pneumoniae*	0.0187
metagenome_6916521	Case	*Bifidobacterium breve*	56.9
metagenome_6916521	Case	*Klebsiella pneumoniae*	18.9
metagenome_6916521	Case	*Escherichia coli*	3.30
metagenome_6916521	Case	*Streptococcus salivarius*	2.49
metagenome_6916521	Case	*[Ruminococcus] gnavus*	2.34
metagenome_6916521	Case	*Salmonella enterica*	2.18
metagenome_6916521	Case	*Enterococcus faecium*	1.85
metagenome_6916521	Case	*Bifidobacterium longum*	1.74
metagenome_6916521	Case	*Bacteroides fragilis*	1.38
metagenome_6916521	Case	*Prevotella oris*	1.02

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
metagenome_6916521	Case	*Goslarvirus goslar*	0.704
metagenome_6916521	Case	*Siphoviridae* sp.	0.508
metagenome_6916521	Case	*Veillonella atypica*	0.500
metagenome_6916521	Case	*Streptococcus thermophilus*	0.322
metagenome_6916521	Case	*Prevotella veroralis*	0.289
metagenome_6916521	Case	*Eggerthella lenta*	0.221
metagenome_6916521	Case	*Myoviridae* sp.	0.219
metagenome_6916521	Case	*Streptococcus* sp. FDAARGOS_192	0.212
metagenome_6916521	Case	*Streptococcus equinus*	0.209
metagenome_6916521	Case	*Streptococcus vestibularis*	0.203
metagenome_6916522	Case	*Escherichia coli*	87.0
metagenome_6916522	Case	*Shigella flexneri*	5.16
metagenome_6916522	Case	*Salmonella enterica*	1.93
metagenome_6916522	Case	*Homo sapiens*	1.82
metagenome_6916522	Case	*Klebsiella pneumoniae*	1.13
metagenome_6916522	Case	*Shigella boydii*	0.694
metagenome_6916522	Case	*Shigella dysenteriae*	0.485
metagenome_6916522	Case	*Bifidobacterium longum*	0.388
metagenome_6916522	Case	*Streptococcus pasteurianus*	0.167
metagenome_6916522	Case	*Shigella sonnei*	0.132
metagenome_6916522	Case	*Saccharomyces cerevisiae*	0.132
metagenome_6916522	Case	*Staphylococcus hominis*	0.112
metagenome_6916522	Case	*Bifidobacterium breve*	0.0728
metagenome_6916522	Case	*Corynebacterium tuberculostearicum*	0.0477
metagenome_6916522	Case	*Staphylococcus haemolyticus*	0.0317
metagenome_6916522	Case	*Escherichia marmotae*	0.0271
metagenome_6916522	Case	*Siphoviridae* sp.	0.0254
metagenome_6916522	Case	*Bacteroides fragilis*	0.0218
metagenome_6916522	Case	*Anopheles gambiae*	0.0213
metagenome_6916522	Case	synthetic construct	0.0193
metagenome_6916523	Case	*Bacteroides fragilis*	59.9
metagenome_6916523	Case	*Escherichia coli*	7.89
metagenome_6916523	Case	*Haemophilus parainfluenzae*	5.81
metagenome_6916523	Case	*Bacteroides thetaiotaomicron*	3.23
metagenome_6916523	Case	*Klebsiella pneumoniae*	2.98
metagenome_6916523	Case	*Morganella morganii*	2.83
metagenome_6916523	Case	*Salmonella enterica*	1.53
metagenome_6916523	Case	*Klebsiella* phage VLCpiP4a	0.520
metagenome_6916523	Case	*Klebsiella quasipneumoniae*	0.441
metagenome_6916523	Case	uncultured bacterium	0.432
metagenome_6916523	Case	*Podoviridae* sp.	0.430
metagenome_6916523	Case	*Sutterella megalosphaeroides*	0.334
metagenome_6916523	Case	*Veillonella parvula*	0.271
metagenome_6916523	Case	*Klebsiella* phage VLCpiP4b	0.270
metagenome_6916523	Case	*Enterobacter roggenkampii*	0.214
metagenome_6916523	Case	*Parabacteroides distasonis*	0.209
metagenome_6916523	Case	*Citrobacter* sp. MGH105	0.189

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
metagenome_6916523	Case	*Anopheles gambiae*	0.178
metagenome_6916523	Case	*Homo sapiens*	0.167
metagenome_6916523	Case	*Enterococcus avium*	0.162
metagenome_6916524	Case	*Escherichia coli*	73.0
metagenome_6916524	Case	*Salmonella enterica*	11.7
metagenome_6916524	Case	*Phocaeicola vulgatus*	6.15
metagenome_6916524	Case	*Klebsiella pneumoniae*	1.59
metagenome_6916524	Case	*Erysipelatoclostridium ramosum*	1.20
metagenome_6916524	Case	*Bacteroides uniformis*	0.90
metagenome_6916524	Case	Bacteriophage sp.	0.428
metagenome_6916524	Case	*Veillonella atypica*	0.413
metagenome_6916524	Case	*Haemophilus parainfluenzae*	0.286
metagenome_6916524	Case	*Bifidobacterium longum*	0.272
metagenome_6916524	Case	*Hungatella hathewayi*	0.239
metagenome_6916524	Case	*Clostridioides difficile*	0.193
metagenome_6916524	Case	*Parabacteroides merdae*	0.185
metagenome_6916524	Case	*Citrobacter portucalensis*	0.165
metagenome_6916524	Case	*Bacteroides caccae*	0.158
metagenome_6916524	Case	*Streptococcus salivarius*	0.150
metagenome_6916524	Case	*Fusobacterium pseudoperiodonticum*	0.147
metagenome_6916524	Case	*Veillonella dispar*	0.144
metagenome_6916524	Case	*Siphoviridae* sp.	0.141
metagenome_6916524	Case	*Phocaeicola dorei*	0.130
metagenome_6916525	Case	*Enterococcus faecium*	41.2
metagenome_6916525	Case	*Streptococcus thermophilus*	18.5
metagenome_6916525	Case	*Enterococcus lactis*	12.1
metagenome_6916525	Case	*Gordonia otitidis*	2.21
metagenome_6916525	Case	*Streptococcus salivarius*	2.20
metagenome_6916525	Case	*Bifidobacterium animalis*	2.16
metagenome_6916525	Case	*Schaalia odontolytica*	1.32
metagenome_6916525	Case	*Enterococcus* sp. DA9	0.836
metagenome_6916525	Case	*Streptococcus mitis*	0.677
metagenome_6916525	Case	*Streptococcus* sp. LPB0220	0.665
metagenome_6916525	Case	*Rothia mucilaginosa*	0.510
metagenome_6916525	Case	*Bifidobacterium* sp. KRGSERBCFTRI	0.504
metagenome_6916525	Case	*Enterococcus durans*	0.467
metagenome_6916525	Case	*Enterococcus faecalis*	0.463
metagenome_6916525	Case	*Alkalihalobacillus clausii*	0.419
metagenome_6916525	Case	*Streptococcus parasanguinis*	0.375
metagenome_6916525	Case	*[Ruminococcus] gnavus*	0.363
metagenome_6916525	Case	*Bifidobacterium longum*	0.336
metagenome_6916525	Case	*Escherichia coli*	0.283
metagenome_6916525	Case	*Rothia kristinae*	0.269
ERR9904448	Control	*Bifidobacterium longum*	17.3
ERR9904448	Control	*Phocaeicola dorei*	9.29
ERR9904448	Control	*Anaerostipes hadrus*	7.75
ERR9904448	Control	*Bifidobacterium catenulatum*	7.55

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
ERR9904448	Control	*Bacteroides uniformis*	6.94
ERR9904448	Control	*Blautia wexlerae*	6.08
ERR9904448	Control	*[Ruminococcus] gnavus*	4.85
ERR9904448	Control	*Bifidobacterium pseudocatenulatum*	3.56
ERR9904448	Control	*Bacteroides thetaiotaomicron*	2.90
ERR9904448	Control	*Collinsella aerofaciens*	2.80
ERR9904448	Control	*Faecalibacterium prausnitzii*	1.82
ERR9904448	Control	*Megasphaera elsdenii*	1.77
ERR9904448	Control	*[Ruminococcus] torques*	1.18
ERR9904448	Control	*Sellimonas intestinalis*	1.06
ERR9904448	Control	*Bifidobacterium breve*	0.786
ERR9904448	Control	*Siphoviridae* sp.	0.655
ERR9904448	Control	*Roseburia intestinalis*	0.641
ERR9904448	Control	*Flavonifractor plautii*	0.577
ERR9904448	Control	*Eggerthella lenta*	0.454
ERR9904448	Control	uncultured bacterium	0.450
ERR9904449	Control	*Bifidobacterium longum*	20.0
ERR9904449	Control	*Phocaeicola vulgatus*	7.72
ERR9904449	Control	*Bacteroides humanifaecis*	6.41
ERR9904449	Control	*Bifidobacterium catenulatum*	6.14
ERR9904449	Control	*Bifidobacterium bifidum*	6.03
ERR9904449	Control	*Anaerostipes hadrus*	4.87
ERR9904449	Control	*Bacteroides stercoris*	3.60
ERR9904449	Control	*Bifidobacterium pseudocatenulatum*	3.38
ERR9904449	Control	*Blautia wexlerae*	2.30
ERR9904449	Control	*[Ruminococcus] torques*	1.95
ERR9904449	Control	*Bifidobacterium breve*	1.62
ERR9904449	Control	*Bacteroides thetaiotaomicron*	1.59
ERR9904449	Control	*Megamonas funiformis*	1.54
ERR9904449	Control	*Megasphaera elsdenii*	1.41
ERR9904449	Control	*[Ruminococcus] gnavus*	1.31
ERR9904449	Control	Oscillospiraceae bacterium	1.16
ERR9904449	Control	*Sellimonas intestinalis*	1.14
ERR9904449	Control	*Bacteroides fragilis*	1.09
ERR9904449	Control	*Faecalibacterium prausnitzii*	1.02
ERR9904449	Control	*Bacteroides* sp. HF-162	0.775
ERR9904450	Control	*Bifidobacterium bifidum*	22.6
ERR9904450	Control	*Bifidobacterium pseudocatenulatum*	10.2
ERR9904450	Control	*Bifidobacterium longum*	9.60
ERR9904450	Control	Oscillospiraceae bacterium	9.04
ERR9904450	Control	*Escherichia coli*	6.88
ERR9904450	Control	*Streptococcus lutetiensis*	5.18
ERR9904450	Control	*Collinsella aerofaciens*	4.82
ERR9904450	Control	*[Ruminococcus] torques*	4.77
ERR9904450	Control	*Streptococcus pasteurianus*	4.20
ERR9904450	Control	*Blautia wexlerae*	1.70
ERR9904450	Control	*Enterococcus faecalis*	1.55

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
ERR9904450	Control	*Veillonella parvula*	1.43
ERR9904450	Control	*Bifidobacterium breve*	1.14
ERR9904450	Control	Clostridiaceae bacterium	0.899
ERR9904450	Control	*Dorea formicigenerans*	0.813
ERR9904450	Control	*Streptococcus gallolyticus*	0.593
ERR9904450	Control	Lachnospiraceae bacterium	0.586
ERR9904450	Control	*Siphoviridae* sp.	0.578
ERR9904450	Control	*Streptococcus thermophilus*	0.539
ERR9904450	Control	*Erysipelatoclostridium ramosum*	0.359
ERR9904451	Control	*Bacteroides uniformis*	9.99
ERR9904451	Control	*Blautia wexlerae*	8.58
ERR9904451	Control	*Bacteroides fragilis*	4.61
ERR9904451	Control	*Alistipes onderdonkii*	3.99
ERR9904451	Control	*Streptococcus salivarius*	2.59
ERR9904451	Control	*Anaerobutyricum hallii*	2.57
ERR9904451	Control	*Parabacteroides distasonis*	2.55
ERR9904451	Control	*Bifidobacterium pseudocatenulatum*	2.13
ERR9904451	Control	*Phocaeicola vulgatus*	1.81
ERR9904451	Control	*Siphoviridae* sp.	1.66
ERR9904451	Control	*Roseburia intestinalis*	1.58
ERR9904451	Control	*Akkermansia muciniphila*	1.56
ERR9904451	Control	*Streptococcus thermophilus*	1.46
ERR9904451	Control	*Collinsella aerofaciens*	1.36
ERR9904451	Control	*[Ruminococcus] gnavus*	1.17
ERR9904451	Control	*Anaerostipes hadrus*	0.995
ERR9904451	Control	uncultured bacterium	0.948
ERR9904451	Control	*[Ruminococcus] torques*	0.829
ERR9904451	Control	*Dorea longicatena*	0.809
ERR9904451	Control	*Bacteroides xylanisolvens*	0.772
ERR9904456	Control	*Bifidobacterium breve*	15.4
ERR9904456	Control	*[Ruminococcus] gnavus*	14.9
ERR9904456	Control	*Escherichia coli*	14.9
ERR9904456	Control	*Erysipelatoclostridium ramosum*	6.38
ERR9904456	Control	*Bacteroides fragilis*	3.96
ERR9904456	Control	*Bacteroides xylanisolvens*	2.40
ERR9904456	Control	Lachnospiraceae bacterium	1.66
ERR9904456	Control	*Bacteroides* sp. ZJ-18	1.45
ERR9904456	Control	*Blautia argi*	1.09
ERR9904456	Control	*Bacteroides* sp. PHL 2737	0.976
ERR9904456	Control	*[Clostridium] innocuum*	0.867
ERR9904456	Control	*Akkermansia muciniphila*	0.842
ERR9904456	Control	*Bifidobacterium longum*	0.799
ERR9904456	Control	*Enterocloster bolteae*	0.775
ERR9904456	Control	*Streptococcus salivarius*	0.746
ERR9904456	Control	*Clostridioides difficile*	0.687
ERR9904456	Control	uncultured bacterium	0.627
ERR9904456	Control	*Siphoviridae* sp.	0.583

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
ERR9904456	Control	*Bacteroides ovatus*	0.472
ERR9904456	Control	*Eggerthella lenta*	0.461
ERR9904457	Control	*[Ruminococcus] gnavus*	8.00
ERR9904457	Control	*Klebsiella pneumoniae*	5.61
ERR9904457	Control	*Streptococcus pasteurianus*	5.17
ERR9904457	Control	*Enterobacter hormaechei*	4.65
ERR9904457	Control	*Enterobacter* sp. BIDMC100	4.39
ERR9904457	Control	*Anaerostipes hadrus*	2.83
ERR9904457	Control	*Bacteroides fragilis*	2.06
ERR9904457	Control	*Bifidobacterium longum*	1.79
ERR9904457	Control	unidentified plasmid	1.65
ERR9904457	Control	*Phocaeicola vulgatus*	1.53
ERR9904457	Control	*Siphoviridae* sp.	1.49
ERR9904457	Control	*Bacteroides thetaiotaomicron*	1.48
ERR9904457	Control	*Faecalibacterium prausnitzii*	1.47
ERR9904457	Control	*Collinsella aerofaciens*	1.37
ERR9904457	Control	*Enterobacter* sp. JBIWA005	1.23
ERR9904457	Control	*Parabacteroides distasonis*	1.21
ERR9904457	Control	*[Clostridium] innocuum*	1.07
ERR9904457	Control	*Bacteroides uniformis*	0.790
ERR9904457	Control	*[Ruminococcus] torques*	0.789
ERR9904457	Control	*Dorea formicigenerans*	0.781
ERR9904458	Control	*[Ruminococcus] gnavus*	12.8
ERR9904458	Control	*Megamonas funiformis*	11.5
ERR9904458	Control	*Phocaeicola vulgatus*	9.77
ERR9904458	Control	*Faecalibacterium prausnitzii*	6.83
ERR9904458	Control	*Bifidobacterium longum*	6.15
ERR9904458	Control	*Bacteroides uniformis*	3.88
ERR9904458	Control	*Streptococcus salivarius*	3.82
ERR9904458	Control	*Bacteroides thetaiotaomicron*	2.63
ERR9904458	Control	*Bacteroides xylanisolvens*	2.16
ERR9904458	Control	*Veillonella atypica*	1.84
ERR9904458	Control	*Bacteroides caccae*	1.48
ERR9904458	Control	*Parabacteroides merdae*	1.41
ERR9904458	Control	*Siphoviridae* sp.	1.10
ERR9904458	Control	uncultured bacterium	1.07
ERR9904458	Control	*Anaerostipes hadrus*	0.899
ERR9904458	Control	*Megamonas hypermegale*	0.836
ERR9904458	Control	*Sellimonas intestinalis*	0.745
ERR9904458	Control	*Escherichia coli*	0.696
ERR9904458	Control	*Lachnospira eligens*	0.599
ERR9904458	Control	*Veillonella dispar*	0.555
ERR9904459	Control	*Bacteroides fragilis*	60.2
ERR9904459	Control	*[Ruminococcus] gnavus*	15.8
ERR9904459	Control	*Klebsiella pneumoniae*	5.63
ERR9904459	Control	*Veillonella atypica*	2.06
ERR9904459	Control	*Enterocloster bolteae*	1.85

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
ERR9904459	Control	*Escherichia coli*	1.30
ERR9904459	Control	*Siphoviridae* sp.	0.757
ERR9904459	Control	*Bacteroides ovatus*	0.377
ERR9904459	Control	*Haemophilus parainfluenzae*	0.364
ERR9904459	Control	*Klebsiella quasipneumoniae*	0.255
ERR9904459	Control	*Parabacteroides distasonis*	0.252
ERR9904459	Control	*Myoviridae* sp.	0.219
ERR9904459	Control	*Prevotella intermedia*	0.217
ERR9904459	Control	*Blautia argi*	0.203
ERR9904459	Control	*Roseburia intestinalis*	0.194
ERR9904459	Control	Bacteriophage sp.	0.193
ERR9904459	Control	*Enterobacter hormaechei*	0.182
ERR9904459	Control	uncultured bacterium	0.172
ERR9904459	Control	*Streptococcus thermophilus*	0.165
ERR9904459	Control	*Enterobacter cloacae*	0.147
ERR9904460	Control	*Blautia wexlerae*	9.16
ERR9904460	Control	*Anaerostipes hadrus*	8.86
ERR9904460	Control	Oscillospiraceae bacterium	8.40
ERR9904460	Control	*Phocaeicola vulgatus*	6.99
ERR9904460	Control	*Bacteroides ovatus*	5.82
ERR9904460	Control	*Bifidobacterium longum*	4.58
ERR9904460	Control	*Bifidobacterium pseudocatenulatum*	3.03
ERR9904460	Control	*Collinsella aerofaciens*	2.52
ERR9904460	Control	*Parabacteroides merdae*	1.96
ERR9904460	Control	*Megamonas funiformis*	1.84
ERR9904460	Control	*[Ruminococcus] torques*	1.74
ERR9904460	Control	*Faecalibacterium prausnitzii*	1.67
ERR9904460	Control	*Bacteroides caccae*	1.39
ERR9904460	Control	*Bifidobacterium bifidum*	1.27
ERR9904460	Control	uncultured bacterium	1.25
ERR9904460	Control	Bacteriophage sp.	1.19
ERR9904460	Control	*Bacteroides* sp. A1C1	1.18
ERR9904460	Control	*Bacteroides thetaiotaomicron*	1.05
ERR9904460	Control	*Siphoviridae* sp.	0.954
ERR9904460	Control	*Parabacteroides distasonis*	0.845
ERR9904461	Control	*Megamonas funiformis*	18.1
ERR9904461	Control	*Escherichia coli*	17.2
ERR9904461	Control	*Bifidobacterium pseudocatenulatum*	7.99
ERR9904461	Control	*Blautia wexlerae*	5.06
ERR9904461	Control	*Phocaeicola vulgatus*	4.58
ERR9904461	Control	*Bifidobacterium bifidum*	3.77
ERR9904461	Control	*Bifidobacterium breve*	3.12
ERR9904461	Control	*Collinsella aerofaciens*	2.65
ERR9904461	Control	*Megasphaera elsdenii*	1.89
ERR9904461	Control	*Acidaminococcus intestini*	1.62
ERR9904461	Control	*Siphoviridae* sp.	1.26
ERR9904461	Control	*Bifidobacterium catenulatum*	1.23

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
ERR9904461	Control	*Bacteroides uniformis*	1.14
ERR9904461	Control	uncultured bacterium	1.13
ERR9904461	Control	*[Ruminococcus] gnavus*	1.13
ERR9904461	Control	*Parabacteroides distasonis*	0.959
ERR9904461	Control	*Megamonas hypermegale*	0.851
ERR9904461	Control	Bacteriophage sp.	0.717
ERR9904461	Control	*Streptococcus pasteurianus*	0.602
ERR9904461	Control	*Bifidobacterium longum*	0.449
ERR9904462	Control	*Bifidobacterium longum*	38.4
ERR9904462	Control	*Megamonas funiformis*	7.71
ERR9904462	Control	*Klebsiella pneumoniae*	7.60
ERR9904462	Control	*Bifidobacterium breve*	6.71
ERR9904462	Control	*Enterobacter roggenkampii*	5.33
ERR9904462	Control	*Bacteroides fragilis*	4.21
ERR9904462	Control	*Collinsella aerofaciens*	2.28
ERR9904462	Control	*[Ruminococcus] torques*	2.13
ERR9904462	Control	*Bacteroides stercoris*	1.98
ERR9904462	Control	*Bacteroides uniformis*	1.49
ERR9904462	Control	*Blautia wexlerae*	1.26
ERR9904462	Control	Oscillospiraceae bacterium	1.24
ERR9904462	Control	*[Ruminococcus] gnavus*	1.23
ERR9904462	Control	*Sutterella wadsworthensis*	0.965
ERR9904462	Control	*Bacteroides ovatus*	0.711
ERR9904462	Control	*Bacteroides xylanisolvens*	0.554
ERR9904462	Control	*Megamonas hypermegale*	0.437
ERR9904462	Control	*Enterobacter hormaechei*	0.395
ERR9904462	Control	*Erysipelatoclostridium ramosum*	0.382
ERR9904462	Control	*Siphoviridae* sp.	0.379
ERR9904463	Control	*Bacteroides fragilis*	43.1
ERR9904463	Control	*Bifidobacterium longum*	14.5
ERR9904463	Control	*Bacteroides* sp. ZJ-18	5.25
ERR9904463	Control	*Bifidobacterium breve*	3.94
ERR9904463	Control	*Faecalibacterium prausnitzii*	3.47
ERR9904463	Control	*Erysipelatoclostridium ramosum*	3.12
ERR9904463	Control	*Blautia* sp. SC05B48	2.74
ERR9904463	Control	*Veillonella parvula*	2.71
ERR9904463	Control	*Siphoviridae* sp.	1.65
ERR9904463	Control	*Sellimonas intestinalis*	1.40
ERR9904463	Control	*[Clostridium] innocuum*	0.976
ERR9904463	Control	*Escherichia coli*	0.875
ERR9904463	Control	*[Ruminococcus] gnavus*	0.862
ERR9904463	Control	*Enterocloster bolteae*	0.526
ERR9904463	Control	*Intestinibacter bartlettii*	0.447
ERR9904463	Control	*Streptococcus pasteurianus*	0.421
ERR9904463	Control	*Enterococcus raffinosus*	0.402
ERR9904463	Control	*Veillonella atypica*	0.393
ERR9904463	Control	*[Clostridium] spiroforme*	0.359

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
ERR9904463	Control	*Clostridium neonatale*	0.356
ERR9904464	Control	*[Ruminococcus] gnavus*	12.2
ERR9904464	Control	*Megamonas funiformis*	8.66
ERR9904464	Control	*Faecalibacterium prausnitzii*	5.18
ERR9904464	Control	*Blautia wexlerae*	4.50
ERR9904464	Control	*Collinsella aerofaciens*	4.04
ERR9904464	Control	*Streptococcus salivarius*	3.43
ERR9904464	Control	*Phocaeicola vulgatus*	3.14
ERR9904464	Control	*Prevotella copri*	2.14
ERR9904464	Control	*Bacteroides caccae*	2.04
ERR9904464	Control	*Bifidobacterium pseudocatenulatum*	1.95
ERR9904464	Control	Bacteriophage sp.	1.77
ERR9904464	Control	*Bifidobacterium longum*	1.40
ERR9904464	Control	*Anaerobutyricum hallii*	1.39
ERR9904464	Control	*Siphoviridae* sp.	1.29
ERR9904464	Control	*Roseburia* sp. NSJ-69	1.17
ERR9904464	Control	*Parabacteroides distasonis*	1.15
ERR9904464	Control	*Escherichia coli*	1.12
ERR9904464	Control	*Streptococcus pasteurianus*	1.11
ERR9904464	Control	*Anaerostipes hadrus*	1.05
ERR9904464	Control	*Intestinibacter bartlettii*	1.04
ERR9904465	Control	*Blautia wexlerae*	13.21
ERR9904465	Control	*Phocaeicola vulgatus*	5.85
ERR9904465	Control	*[Ruminococcus] gnavus*	5.12
ERR9904465	Control	*Lachnospiraceae bacterium*	4.93
ERR9904465	Control	*[Ruminococcus] torques*	2.73
ERR9904465	Control	*Bifidobacterium catenulatum*	2.72
ERR9904465	Control	*Collinsella aerofaciens*	2.61
ERR9904465	Control	*Anaerostipes hadrus*	2.49
ERR9904465	Control	*Bacteroides xylanisolvens*	2.10
ERR9904465	Control	*Streptococcus thermophilus*	1.73
ERR9904465	Control	*Escherichia coli*	1.65
ERR9904465	Control	*Bifidobacterium longum*	1.45
ERR9904465	Control	*Sellimonas intestinalis*	1.37
ERR9904465	Control	*Streptococcus salivarius*	1.34
ERR9904465	Control	*Bacteroides thetaiotaomicron*	1.25
ERR9904465	Control	*Siphoviridae* sp.	1.22
ERR9904465	Control	*Faecalibacterium prausnitzii*	1.20
ERR9904465	Control	*Ruthenibacterium lactatiformans*	1.16
ERR9904465	Control	*Parabacteroides distasonis*	1.14
ERR9904465	Control	*Bacteroides fragilis*	0.898
ERR9904466	Control	*Bacteroides uniformis*	8.46
ERR9904466	Control	*Bacteroides intestinalis*	5.92
ERR9904466	Control	*Faecalibacterium prausnitzii*	3.83
ERR9904466	Control	*Bacteroides thetaiotaomicron*	3.19
ERR9904466	Control	*Parabacteroides merdae*	2.68
ERR9904466	Control	Oscillospiraceae bacterium	2.64

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
ERR9904466	Control	*Phascolarctobacterium faecium*	2.52
ERR9904466	Control	*Prevotella copri*	2.21
ERR9904466	Control	*Megamonas funiformis*	1.60
ERR9904466	Control	*Siphoviridae* sp.	1.55
ERR9904466	Control	uncultured bacterium	1.40
ERR9904466	Control	*Anaerobutyricum hallii*	1.20
ERR9904466	Control	*Bifidobacterium pseudocatenulatum*	1.17
ERR9904466	Control	*[Ruminococcus] torques*	1.13
ERR9904466	Control	*Catenibacterium mitsuokai*	1.08
ERR9904466	Control	*[Eubacterium] rectale*	1.08
ERR9904466	Control	*Anaerostipes hadrus*	0.809
ERR9904466	Control	*Parabacteroides distasonis*	0.793
ERR9904466	Control	Clostridiaceae bacterium	0.764
ERR9904466	Control	Bacteriophage sp.	0.743
ERR9904467	Control	*Prevotella copri*	12.67
ERR9904467	Control	*Megamonas funiformis*	5.78
ERR9904467	Control	*Blautia wexlerae*	5.07
ERR9904467	Control	Oscillospiraceae bacterium	4.58
ERR9904467	Control	*Faecalibacterium prausnitzii*	3.90
ERR9904467	Control	*Parabacteroides distasonis*	3.67
ERR9904467	Control	*Bifidobacterium longum*	2.41
ERR9904467	Control	*Phocaeicola vulgatus*	2.35
ERR9904467	Control	*Bifidobacterium catenulatum*	2.20
ERR9904467	Control	uncultured bacterium	2.15
ERR9904467	Control	*Bifidobacterium bifidum*	1.81
ERR9904467	Control	*Siphoviridae* sp.	1.70
ERR9904467	Control	*Anaerobutyricum hallii*	1.66
ERR9904467	Control	Bacteriophage sp.	1.60
ERR9904467	Control	*Bacteroides uniformis*	1.34
ERR9904467	Control	*Bacteroides caccae*	1.31
ERR9904467	Control	*Anaerostipes hadrus*	1.13
ERR9904467	Control	*Phocaeicola coprophilus*	1.11
ERR9904467	Control	uncultured organism	1.10
ERR9904467	Control	*Blautia obeum*	1.08
ERR9904468	Control	*Phocaeicola dorei*	7.61
ERR9904468	Control	*Bacteroides xylanisolvens*	6.07
ERR9904468	Control	*Bifidobacterium pseudocatenulatum*	5.95
ERR9904468	Control	*Bacteroides uniformis*	5.57
ERR9904468	Control	*Bifidobacterium longum*	4.92
ERR9904468	Control	*Bacteroides eggerthii*	4.40
ERR9904468	Control	*Akkermansia muciniphila*	3.92
ERR9904468	Control	*Bacteroides thetaiotaomicron*	3.46
ERR9904468	Control	*Escherichia coli*	3.43
ERR9904468	Control	*Bifidobacterium breve*	2.94
ERR9904468	Control	*Bifidobacterium bifidum*	2.71
ERR9904468	Control	*Collinsella aerofaciens*	2.71
ERR9904468	Control	*Blautia wexlerae*	2.46

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
ERR9904468	Control	*Anaerostipes hadrus*	2.13
ERR9904468	Control	*Sellimonas intestinalis*	1.95
ERR9904468	Control	*[Ruminococcus] gnavus*	1.79
ERR9904468	Control	*Bacteroides fragilis*	1.53
ERR9904468	Control	*Streptococcus pasteurianus*	1.37
ERR9904468	Control	*Parabacteroides distasonis*	1.27
ERR9904468	Control	*Faecalibacterium prausnitzii*	0.840
ERR9904469	Control	*Blautia wexlerae*	12.25
ERR9904469	Control	*Roseburia hominis*	6.11
ERR9904469	Control	*Parabacteroides distasonis*	5.33
ERR9904469	Control	*Parabacteroides merdae*	4.94
ERR9904469	Control	*Megamonas funiformis*	4.34
ERR9904469	Control	*Faecalibacterium prausnitzii*	4.21
ERR9904469	Control	*Prevotella copri*	3.71
ERR9904469	Control	*Bacteroides caccae*	3.52
ERR9904469	Control	*Bacteroides xylanisolvens*	3.04
ERR9904469	Control	*Bacteroides thetaiotaomicron*	2.73
ERR9904469	Control	Bacteriophage sp.	2.47
ERR9904469	Control	*Phocaeicola vulgatus*	2.31
ERR9904469	Control	*Bacteroides humanifaecis*	2.14
ERR9904469	Control	*[Ruminococcus] gnavus*	1.67
ERR9904469	Control	*Bifidobacterium longum*	1.55
ERR9904469	Control	*Bacteroides intestinalis*	1.46
ERR9904469	Control	*Bifidobacterium bifidum*	1.24
ERR9904469	Control	*Bacteroides* sp. CACC 737	1.15
ERR9904469	Control	*Anaerostipes hadrus*	1.07
ERR9904469	Control	*Lachnospira eligens*	1.05
ERR9904470	Control	*Blautia wexlerae*	26.0
ERR9904470	Control	*[Clostridium] innocuum*	10.4
ERR9904470	Control	*Anaerostipes hadrus*	5.46
ERR9904470	Control	*Phocaeicola dorei*	4.37
ERR9904470	Control	Bacteriophage sp.	4.14
ERR9904470	Control	*[Ruminococcus] lactaris*	2.60
ERR9904470	Control	*Faecalibacterium prausnitzii*	2.09
ERR9904470	Control	*Bacteroides ovatus*	1.84
ERR9904470	Control	uncultured human fecal virus	1.62
ERR9904470	Control	*Bacteroides xylanisolvens*	1.52
ERR9904470	Control	Lachnospiraceae bacterium	1.42
ERR9904470	Control	*Bacteroides thetaiotaomicron*	1.17
ERR9904470	Control	*Bacteroides caccae*	1.15
ERR9904470	Control	*Phascolarctobacterium faecium*	1.00
ERR9904470	Control	*Bifidobacterium pseudocatenulatum*	0.914
ERR9904470	Control	*Parabacteroides distasonis*	0.900
ERR9904470	Control	*Bifidobacterium longum*	0.891
ERR9904470	Control	uncultured bacterium	0.753
ERR9904470	Control	*Blautia obeum*	0.712
ERR9904470	Control	*[Ruminococcus] gnavus*	0.668

Appendix 10: Relative abundance of the top 20 most abundant species in the case and control metagenomes, determined with Kraken2 and Bracken, with *Salmonella enterica* highlighted in yellow and other potential diarrheagenic pathogens highlighted in orange

Sample	Type	Species	Relative abundance (%)
ERR9904471	Control	*Prevotella copri*	12.57
ERR9904471	Control	*Blautia wexlerae*	8.72
ERR9904471	Control	*Phocaeicola vulgatus*	4.16
ERR9904471	Control	*Faecalibacterium prausnitzii*	3.51
ERR9904471	Control	*Phascolarctobacterium faecium*	3.47
ERR9904471	Control	*Bifidobacterium catenulatum*	2.74
ERR9904471	Control	*Parabacteroides merdae*	2.58
ERR9904471	Control	*Megamonas funiformis*	2.53
ERR9904471	Control	*Parabacteroides distasonis*	2.36
ERR9904471	Control	*Bifidobacterium longum*	2.04
ERR9904471	Control	*Siphoviridae* sp.	1.71
ERR9904471	Control	*Streptococcus salivarius*	1.71
ERR9904471	Control	*Bacteroides thetaiotaomicron*	1.56
ERR9904471	Control	*Roseburia hominis*	1.44
ERR9904471	Control	uncultured bacterium	1.44
ERR9904471	Control	*Collinsella aerofaciens*	1.23
ERR9904471	Control	Oscillospiraceae bacterium	1.15
ERR9904471	Control	*Sellimonas intestinalis*	1.07
ERR9904471	Control	Bacteriophage sp.	1.06
ERR9904471	Control	*[Ruminococcus] gnavus*	1.01
ERR9904485	Control	*Phocaeicola vulgatus*	20.4
ERR9904485	Control	*Bifidobacterium breve*	12.8
ERR9904485	Control	*Bifidobacterium pseudocatenulatum*	7.75
ERR9904485	Control	*Bacteroides uniformis*	5.21
ERR9904485	Control	*Bifidobacterium longum*	3.90
ERR9904485	Control	*Bacteroides thetaiotaomicron*	3.67
ERR9904485	Control	*Bacteroides xylanisolvens*	2.53
ERR9904485	Control	*Bacteroides fragilis*	2.02
ERR9904485	Control	*Phocaeicola dorei*	1.44
ERR9904485	Control	*Escherichia coli*	1.23
ERR9904485	Control	*Siphoviridae* sp.	1.21
ERR9904485	Control	Oscillospiraceae bacterium	1.16
ERR9904485	Control	*Limosilactobacillus fermentum*	1.13
ERR9904485	Control	*Enterococcus avium*	1.05
ERR9904485	Control	*Collinsella aerofaciens*	1.04
ERR9904485	Control	*[Ruminococcus] torques*	1.04
ERR9904485	Control	*Blautia wexlerae*	0.955
ERR9904485	Control	*Streptococcus salivarius*	0.787
ERR9904485	Control	*Faecalibacterium prausnitzii*	0.774
ERR9904485	Control	Bacteriophage sp.	0.752

Appendix 11: Proportion of *Salmonella enterica* and *Salmonella enterica* serovars determined through genome analysis identified in the associated metagenomes; values highlighted in green indicate that the proportion represents the most likely serovar present

Metagenome	Serovar of associated genome	Total number of reads in sample	Classification tool				Alignment and assembly	
			Centrifuge			Bracken	Nucmer+BBsplit	SISTR cgMLST
			Serovar ID	Serovar proportion (%)	Species proportion (%)	Species proportion (%)	Proportion of serovar reference covered	Serovar predicted
metagenome_6916498	Stanley	30576		0.00	0.474	0.942	0.00243	Choleraesuis
metagenome_6916499	Typhimurium	796619	*Salmonella enterica* subsp. *enterica* serovar Typhimurium	0.000628	0.0678	0.085	0.00949	Choleraesuis
metagenome_6916500	I 4,[5],12:i:-	8300300	*Salmonella enterica* subsp. *enterica* serovar 4,[5],12:i:-	0.0000482	0.156	0.598	0.0200	Enteritidis
metagenome_6916501	Meleagridis	4716550		0.00	0.0113	0.0200	Not in database	Typhimurium
metagenome_6916502	Saintpaul	30670355	*Salmonella enterica* subsp. *enterica* serovar Saintpaul	0.00000326	0.00432	0.00978	0.0207	Typhimurium
metagenome_6916503	Weltevreden	70506774	*Salmonella enterica* subsp. *enterica* serovar Weltevreden	0.00000425	0.00557	0.00259	0.0155	Enteritidis
metagenome_6916504	Enteritidis	3312164	*Salmonella enterica* subsp. *enterica* serovar Enteritidis	0.0000604	0.00537	0.0144	0.00485	Choleraesuis
metagenome_6916505	Typhimurium	222420	*Salmonella enterica* subsp. *enterica* serovar Typhimurium	0.190	18.3	27.7	0.621	Typhimurium
metagenome_6916506	Typhimurium	579680	*Salmonella enterica* subsp. *enterica* serovar Typhimurium	0.0361	0.465	0.791	0.0182	Rissen
metagenome_6916507	I 4,[5],12:i:-	403987		0.00	9.25	12.8	0.734	I 4,[5],12:i:-
metagenome_6916508	Newport	21808691	*Salmonella enterica* subsp. *enterica* serovar Newport	0.000381	0.0476	0.0778	0.00953	Enteritidis
metagenome_6916509	Paratyphi B var. Java	4431232	*Salmonella enterica* subsp. *enterica* serovar Java	1.82	16.0	22.3	0.605	Paratyphi B var. Java monophasic

Appendix 11: Proportion of *Salmonella enterica* and *Salmonella enterica* serovars determined through genome analysis identified in the associated metagenomes; values highlighted in green indicate that the proportion represents the most likely serovar present

Metagenome	Serovar of associated genome	Total number of reads in sample	Classification tool				Alignment and assembly	
			Centrifuge			Bracken	Nucmer+BBsplit	SISTR cgMLST
			Serovar ID	Serovar proportion (%)	Species proportion (%)	Species proportion (%)	Proportion of serovar reference covered	Serovar predicted
metagenome_6916510	I 4,[5],12:i:-	48036427	*Salmonella enterica* subsp. *enterica* serovar 4,[5],12:i:-	0.00000208	0.0766	0.0225	0.00840	Bredeney
metagenome_6916511	Hvittingfoss	957107	*Salmonella enterica* subsp. *enterica* serovar Hvittingfoss	0.238	1.30	1.82	0.202	Hvittingfoss
metagenome_6916512	Typhimurium	26414703	*Salmonella enterica* subsp. *enterica* serovar Typhimurium	0.0481	0.257	0.0891	0.0784	Enteritidis
metagenome_6916513	I 4,[5],12:i:-	3369247	*Salmonella enterica* subsp. *enterica* serovar 4,[5],12:i:-	0.0000594	2.17	2.82	0.869	Enteritidis
metagenome_6916514	I 4,[5],12:i:-	51944841	*Salmonella enterica* subsp. *enterica* serovar 4,[5],12:i:-	0.0000886	12.7	16.5	0.928	I 4,[5],12:i:-
metagenome_6916515	Stanley	89351		0.00	0.172	0.172	0.0270	Enteritidis
metagenome_6916516	I 4,[5],12:i:-	53522335	*Salmonella enterica* subsp. *enterica* serovar 4,[5],12:i:-	0.0000206	0.173	14.3	0.0519	Bredeney
metagenome_6916517	I 4,[5],12:i:-	298453		0.0000	1.97	2.56	0.0412	Enteritidis
metagenome_6916518	Albany	60890445	*Salmonella enterica* subsp. *enterica* serovar Albany	0.0689	0.359	0.678	0.804	Albany
metagenome_6916519	Typhimurium	8559665	*Salmonella enterica* subsp. *enterica* serovar Typhimurium	0.000619	0.0498	0.0527	0.0211	Enteritidis
metagenome_6916520	Enteritidis	23433178	*Salmonella enterica* subsp. *enterica* serovar Enteritidis	0.000486	0.462	0.0149	0.00685	Enteritidis
metagenome_6916521	I 4,[5],12:i:-	94890832	*Salmonella enterica* subsp. *enterica* serovar 4,[5],12:i:-	0.0000148	1.62	2.18	0.904	I 4,[5],12:i:-

Appendix 11: Proportion of *Salmonella enterica* and *Salmonella enterica* serovars determined through genome analysis identified in the associated metagenomes; values highlighted in green indicate that the proportion represents the most likely serovar present

Metagenome	Serovar of associated genome	Total number of reads in sample	Classification tool				Alignment and assembly	
			Centrifuge			Bracken	Nucmer+BBsplit	SISTR cgMLST
			Serovar ID	Serovar proportion (%)	Species proportion (%)	Species proportion (%)	Proportion of serovar reference covered	Serovar predicted
metagenome_6916522	I 4,[5],12:i:-	1785951		0.00	0.124	1.93	0.0151	Bredeney
metagenome_6916523	I 4,[5],12:i:-	42489614	*Salmonella enterica* subsp. *enterica* serovar 4,[5],12:i:-	0.00000235	0.105	1.53	0.430	Enteritidis
metagenome_6916524	I 4,[5],12:i:-	33217278	*Salmonella enterica* subsp. *enterica* serovar 4,[5],12:i:-	0.00000602	0.0581	11.7	0.0593	Enteritidis
metagenome_6916525	Saintpaul	26916352	*Salmonella enterica* subsp. *enterica* serovar Saintpaul	0.0000297	0.0223	0.0289	0.0652	Enteritidis

cgMLST=core genome multilocus sequence typing

Appendix 12: Proportion of *Salmonella enterica* identified in the healthy control metagenomes and the top serovar result

Metagenome	Total number of reads - sample	Classification tool				Alignment and assembly		
		Centrifuge			Bracken	Nucmer+BBsplit		SISTR cgMLST
		Serovar ID - top result (*S. enterica*)	Serovar proportion (%) - top result (*S. enterica*)	Species proportion (%)	Species proportion (%)	Serovar - top result	Proportion	Serovar predicted
ERR9904448	11058806	*Salmonella enterica* subsp. *salamae* serovar 57:z29:z42	0.000172	0.00401	0.00	Stanley	0.0183	Choleraesuis
ERR9904449	11347650	*Salmonella enterica* subsp. *enterica* serovar Enteritidis	0.000291	0.00485	0.00292	Senftenberg	0.0363	Rissen
ERR9904450	11972059	*Salmonella enterica* subsp. *enterica* serovar Typhimurium	0.000735	0.00941	0.0744	Indiana	0.0240	Enteritidis
ERR9904451	11649380	*Salmonella enterica* subsp. *enterica* serovar Matopeni	0.000558	0.00469	0.00	Indiana	0.0111	Bredeney
ERR9904456	12348497	*Salmonella enterica* subsp. *enterica* serovar Anatum	0.00263	0.0546	0.230	Indiana	0.0174	Enteritidis
ERR9904457	12807184	*Salmonella enterica* subsp. *enterica* serovar Typhi	0.00365	0.0489	0.0262	Senftenberg	0.0449	Enteritidis
ERR9904458	13036869	*Salmonella enterica* subsp. *enterica* serovar Newport	0.000552	0.0045	0.00	Reading	0.0347	Enteritidis
ERR9904459	15177561	*Salmonella enterica* subsp. *enterica* serovar Typhi	0.00257	0.0129	0.0624	Senftenberg	0.0687	Mikawasima
ERR9904460	12182543	*Salmonella enterica* subsp. *enterica* serovar Krefeld	0.000337	0.00609	0.00204	Senftenberg	0.0192	Enteritidis
ERR9904461	12904277	*Salmonella enterica* subsp. *enterica* serovar Typhimurium	0.00105	0.0171	0.0087	Napoli	0.0161	Bredeney
ERR9904462	11283567	*Salmonella enterica* subsp. *enterica* serovar Typhi	0.00806	0.0454	0.0412	Senftenberg	0.0435	Enteritidis
ERR9904463	12383599	*Salmonella enterica* subsp. *enterica* serovar Typhi	0.000420	0.00381	0.0152	Indiana	0.0262	Bredeney

Appendix 12: Proportion of *Salmonella enterica* identified in the healthy control metagenomes and the top serovar result

Metagenome	Total number of reads - sample	Classification tool				Alignment and assembly		
		Centrifuge			Bracken	Nucmer+BBsplit		SISTR cgMLST
		Serovar ID - top result (*S. enterica*)	Serovar proportion (%) - top result (*S. enterica*)	Species proportion (%)	Species proportion (%)	Serovar - top result	Proportion	Serovar predicted
ERR9904464	13483755	*Salmonella enterica* subsp. *enterica* serovar Typhimurium	0.000356	0.00731	0.00488	California	0.0226	Enteritidis
ERR9904465	10821636	*Salmonella enterica* subsp. *enterica* serovar Typhimurium	0.000296	0.00941	0.0141	Indiana	0.0253	Enteritidis
ERR9904466	11395229	*Salmonella enterica* subsp. *salamae* serovar 55:k:z39	0.000281	0.00774	0.00183	Senftenberg	0.0284	Typhimurium
ERR9904467	11396301	*Salmonella enterica* subsp. *enterica* serovar Newport	0.000325	0.00683	0.00223	Napoli	0.0364	Rissen
ERR9904468	11438979	*Salmonella enterica* subsp. *enterica* serovar Typhimurium	0.000245	0.00719	0.0110	Indiana	0.0222	Typhimurium
ERR9904469	12397759	*Salmonella enterica* subsp. *enterica* serovar India	0.000210	0.00401	0.00	Brancaster	0.000396	Choleraesuis
ERR9904470	11449445	*Salmonella enterica* subsp. *enterica* serovar Macclesfield	0.00193	0.00720	0.00	Cerro	0.00715	Rissen
ERR9904471	11286775	*Salmonella enterica* subsp. *salamae* serovar 55:k:z39	0.000301	0.00714	0.00330	Schwarzengrund	0.0204	Enteritidis
ERR9904485	11082554	*Salmonella enterica* subsp. *enterica* serovar Typhimurium	0.00031	0.00429	0.00134	Hadar	0.0160	Typhimurium

cgMLST=core genome multilocus sequence typing

Appendix 13: *Salmonella* metagenome assembled genomes (MAGs) identified in the case metagenomes

Metagenome	*S. enterica* MAGs											
	Metabat2						Maxbin2					
	No. of MAGs	Comp. (%)	Cont. (%)	GC (%)	Size	Serovar classification	No. of MAGs	Comp. (%)	Cont. (%)	GC (%)	Size	Serovar classification
metagenome_6916498	-	-	-	-	-	-	-	-	-	-	-	-
metagenome_6916499	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916500	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916501	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916502	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916503	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916504	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916505	0	-	-	-	-	-	1	14.7	1.72	48.7	1819582	Typhimurium
metagenome_6916506	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916507	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916508	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916509	2	12.5-75.9	0.00-0.0400	51.2-52.6	1143091-3399705	Java	1	98.8	0.390	52.2	4633441	Java
metagenome_6916510	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916511	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916512	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916513	1	14.7	0.00	53.4	1264848	Typhimurium	1	57.1	0.00	52.1	3343026	Typhimurium
metagenome_6916514	3	12.5-37.9	0.00-0.03	51.0-52.5	949825-2151674	I 4,[5],12:i:-, Typhimurium	1	67.6	0.00	52.2	4663338	Typhimurium
metagenome_6916515	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916516	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916517	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916518	1	33.5	1.75	52.2	2541824	Albany	1	93.6	3.03	52.0	4769587	Albany

Appendix 13: *Salmonella* metagenome assembled genomes (MAGs) identified in the case metagenomes

Metagenome	*S. enterica* MAGs											
	Metabat2						Maxbin2					
	No. of MAGs	Comp. (%)	Cont. (%)	GC (%)	Size	Serovar classification	No. of MAGs	Comp. (%)	Cont. (%)	GC (%)	Size	Serovar classification
metagenome_6916519	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916520	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916521	2	19.8-39.7	0.00-1.72	52.1-52.5	714422-1994898	I 4,[5],12:i:-	1	99.2	0.23	52.1	4989203	Typhimurium
metagenome_6916522	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916523	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916524	0	-	-	-	-	-	0	-	-	-	-	-
metagenome_6916525	0	-	-	-	-	-	0	-	-	-	-	-

No. = number; comp = completeness; cont = contamination