Helen Wimalarathna, Hertford College, University of Oxford.

The Molecular Epidemiology and Ecology of Campylobacter in the UK.

D.Phil in Zoology.

Trinity Term 2015

Contents

Acknowledgements

This thesis was based on data collected as part of a number of projects funded by the Health Protection Agency, PathoNgen Trace, The University of Oxford and the University of Reading.

I wish to thank Professor Martin Maiden for supervision, and for giving me the opportunity to prove myself. Many thanks to Professor Sam Sheppard for supervision, inspiration and enthusiasm, and Professor Noel McCarthy for extensive support and guidance.

I am grateful to Dr Judith Richardson, Dr Andy Lawson and colleagues at the Health Protection Agency Gastrointestinal Reference Unit in Colindale, for antimicrobial resistance phenotyping carried out in the reference laboratory; and I wish to thank the team of Biomedical Scientists at the John Radcliffe Hospital for their involvement in the study, specifically for the isolation of clinical *Campylobacter* from human stool specimens and resistance phenotyping.

Thank you to Dr Keith Jolley and Dr James Bray for their extensive work on the BIGSdb and the genome assembly pipeline.

Thank you to Dr Ali Cody for immeasurable help in the lab and beyond.

Thanks also to colleagues in the Maiden Group and Sheppard Group for help, encouragement and discussion of ideas, particularly Dr Fran Colles, Dr Carina Brehony, Melissa Jansen van Rensburg, Dorothea Hill, Kanny Diallo, Dr Guillaume Meric and Dr Ben Pascoe.

Above all I wish to thank my two daughters, Sachini and Amavi, for providing perspective and a reason to succeed.

Abstract

Campylobacter is a major Public Health problem in the UK, responsible for more than 50,000 recorded cases of gastroenteritis annually. This thesis examines the molecular epidemiology of clinical campylobacteriosis in Oxfordshire over a one-year period, considering the distribution of *Campylobacter jejuni* and *Campylobacter coli* clonal complexes, the relationship between clonal complex and ciprofloxacin resistance and the relationship between travel-associated campylobacteriosis and *Campylobacter* genotype.

The clinical picture is related back to the molecular epidemiology of *Campylobacter* isolated from retail poultry and the prevalence of resistance to antimicrobial substances. A widespread but non-random distribution of antibiotic resistance phenotypes among *Campylobacter* Sequence Types is indicative of the importance of horizontal transfer of resistant genes as well as clonal expansions of resistant genotypes.

Continuing backwards along the food production chain the molecular epidemiology of *Campylobacter* amongst livestock and farmyard commensals is considered. A study of *Campylobacter* in rats reveals considerable genetic diversity, and three distinct clades of 'rat-associated' *Campylobacter*. The use of whole genome sequence data here allows a higher resolution study of the population structure of *Campylobacter*, identifying a major split in the ST-45 clonal complex.

The potential to use next generation sequence data to improve the attribution of clinical campylobacteriosis cases to animal host source is investigated. Whole genome

sequencing reveals that there is insufficient differentiation between populations of *Campylobacter* isolated from farm animals.

Throughout this thesis data from conventional MLST and from next generation sequencing of whole genomes are compared and the relative benefits considered.

Table of Abbreviations

Abbreviation	Meaning
AIM	Asymmetric Island Model
AMOVA	Analysis of Molecular Variance
aspA	aspartase A
aST(1)	alternative Sequence Type (1)
aST(2)	alternative Sequence Type (2)
BIGSdb	Bacterial Isolate Genome Sequence Database
CC	Clonal Complex
cgMLST	core genome Multi Locus Sequence Typing
flaA	flagellar filament outer layer protein
flaB	flagellar filament core protein
glmM	phosphoglucosamine mutase
glnA	glutamine synthetase
gltA	citrate synthase
glyA	serine hydroxymethyltransferase
MLST	Multi Locus Sequence Typing
NCTC	National Collection of Type Cultures
NJ	Neighbor Joining
PCR	Polymerase Chain Reaction
pgm	phosphoglucomutase
porA	outer membrane porin A
PUBMLST	Public Multi Locus Sequence Typing Databases
RG 1-3	Rat Group 1-3
rMLST	ribosomal Multi Locus Sequence Typing
ST	Sequence Type
Thr_{86}	point mutation at position Thr-86 on the *gyrA*gene product
tkt	transketolase
U/K	Unknown (Sequence Type)
uncA	ATP synthase α subunit
UP	Uncharacterised Protein
wgMLST	whole genome Multi Locus Sequence Typing

Chapter One: Introduction.

1.1 The Campylobacter Genus

Campylobacter is a small gram negative rod, which displays a spiral shape under normal growth conditions. The organism is highly motile; a polar flagellum is present at one or both ends[1]. The optimal growth temperature varies between *Campylobacter* species. *C. jejuni* grows preferentially at 41^{o}C, while *C. coli* has an optimal growth temperature of 37^{o}C[2]. There is evidence of an association between primary host species and optimal growth temperature[2]. All *Campylobacter* species are microaerophilic, requiring an oxygen concentration of 3-15% and a carbon concentration of 3-5%[1].

Campylobacter isolates can be presumptively identified by their classic 'dew-drop' colonial morphology on selective agar, by positive reactions to catalase and oxidase; and by a negative reaction to urease.

C. jejuni lacks the metabolic pathways required for the use of carbohydrates such as glucose and galactose commonly used by bacterial pathogens. Instead they utilize a range of amino acids and intermediates of the citric acid cycle[3].

The *Campylobacter* genome is relatively small, at approximately 1600 – 1700 kb. The G+C content ranges between 30 – 35%[1].

Molecular characterization of *Campylobacter* isolates, using 23S rRNA places the genus in the epsilon subdivision of the Proteobacteria[4,5]. Closely related genera, falling within

the same subdivision are *Helicobacter* and *Arcobacter*, both of which may be associated with gastrointestinal disease[5].

1.2 The Importance of *Campylobacter* as a Public Health Problem in the UK

Campylobacter is a major Public Health problem in the UK, responsible for more than 50,000 recorded cases of gastroenteritis annually[6], and costing the NHS an estimated GBP 27.8 million each year[6,7]. Additionally there is an estimated economic cost of £265.10 per case on average to individual patients and to society in terms of medication charges and productive working days lost to illness[7]. Human campylobacteriosis is characterized by severe diarrhoea, usually accompanied by fever, abdominal pain, nausea and malaise[8].

While enteric campylobacteriosis is an acute, self-limiting disease; *Campylobacter* infection has been identified as an antecedent, or potential risk factor, for the chronic conditions Guillain-Barre Syndrome and reactive arthritis[9,10].

In the UK, in common with other industrialised nations, *Campylobacter jejuni* is responsible for approximately 90% of reported cases of human campylobacteriosis, with the remaining 10% of known cases being largely due to *Campylobacter coli*[11]. For this reason, *Campylobacter jejuni* will form the main focus of this thesis, with some consideration given also to *Campylobacter coli*.

1.3 Molecular Typing

Molecular typing of pathogenic bacteria has enhanced many epidemiological studies, including the identification of food-borne outbreaks of infection due to *E. coli* O157:

H7[12], *Salmonella enteritica*[13] and *Listeria monocytogenes*[14] and early identification of an outbreak source can in some cases enable effective disease containment[14]. However, in many species of bacteria it is currently impossible to predict the lineage from which a pathogenic phenotype will arise. For example, in *Bacillus cereus,* pathogenicity is associated with mobile elements which mediate spore-formation and toxin production[15]. These elements can be acquired by distantly related lineages and therefore, the likelihood that a strain will be pathogenic cannot be discerned from genotypes derived from non-plasmid DNA alone.

In species such as *Campylobacter*, particular genetically related groups often display similar disease associated phenotypes. The consistency of *Campylobacter* genotypes within sub-populations, and the variation between sub-populations can be exploited in order to determine the source of human infection by comparing clinical isolate genotype data with large reference sets isolated from known host-species.

Methods such as PFGE and serotyping have shown that far from being monomorphic pathogenic clones like *Yersinia pestis* or *Mycobacterium lepris*[16], *Campylobacter* populations are highly structured with complex associations among lineages at different levels of relatedness. The DNA sequence-based typing method of Multi Locus Sequence Typing (MLST) has provided considerable insight into population structure in recombining organisms such as *Campylobacter*. MLST is an unambiguous high-resolution genotyping method, exploiting genetic variation in fragments of seven separate housekeeping genes. Each locus is approximately 500bp in length, with a defined start and end point. Each unique sequence at a given locus is assigned an allele number, and a Sequence Type (ST) is identified by a unique series of seven numbers,

referring to the specific alleles present at each locus. Related STs can be grouped into Clonal Complexes, in which sequences are identical at four or more loci[17]. Molecular typing methods which have been used for the study of *Campylobacter* are summarised in table 1.

This high degree of genetic structuring is the result of a complex interplay of mutation, which leads to the gradual divergence of clonally related lineages, and Horizontal Gene Transfer (HGT), that can lead to the replacement of homologous DNA with sequence from another lineage or in extreme cases the introduction of new genes. While *Campylobacter* can be highly recombinogenic[18], mutation and HGT have not been sufficient to erase the clonal signal of descent from the genomes.

Table 1. Current and historical molecular typing methods used for the study of *Campylobacter*.

Method	Description	References
Serotyping	Agglutination-based methods, using a number of antisera to type *Campylobacter* isolates according to surface proteins, typically heat-stable antigens. Most widely used method Penner Serotyping, using 66 antisera. Problems include a lack of discriminatory power and lack of specificity, with agglutination to more than one antiserum common. Studies report approximately 20% of isolates being untypeable by serotyping. Serotyping has largely been displaced by other methods. It may be a useful tool for screening of large numbers of isolates for further investigation using higher resolution typing methods.	McKay, 2001[19] Frost, 1998[20]
Pulsed-Field Gel Electrophoresis	Restriction enzymes, typically *Kpn*1 or *Sma*1, are used to cut the chromosome into large fragments. Electrophoresis is used to separate the fragments according to size. The method was one of the first DNA-based methods for *Campylobacter* typing and was previously considered the gold-standard, although it has had mixed success in epidemiological studies.	Taboada, 2013[21] Ribot, 2001[22] Melero, 2012[23] Miller, 2010[24] Sahin, 2012[25] Champion, 2002[26]
Antigen typing	Short variable regions of the antigen genes, *flaA*, *flaB* or *porA* are used to distinguish between closely related isolates. The original method used RFLP, although this has been largely superseded by PCR	Taboada, 2013[21] Nachamkin, 1993[27] Meinersmann, 1997[28]

	amplification and DNA sequencing. While the method can provide good discrimination between closely related strains the hypervariability of the gene fragments, and the propensity to take up exogenous DNA lead to a lack of stability and to low species-specificity.	
Comparative Genome Fingerprinting	DNA-sequences at 40 loci, determined to be free of SNPs, and displaying variable carriage among isolates are arranged into eight groups of five. PCR sequencing is used to determine the presence or absence of the given locus within the isolate. Clusters are generated, using UPGMA methods, with a strict cut-off of 100% fingerprint identity for isolates within a cluster. This method has been shown to be highly effective in the identification of outbreak-associated clusters.	Clark, 2001[21]
Comparative Genome Hybridisation	A micro-array based method for ascertaining the presence and absence of genes across a reference and test isolates, by using differential patterns of fluorescence. Gained popularity as an accessible and efficient method for studies of comparative genomics before the advent of widely available NGS. Demonstrated the differential carriage of accessory genes.	Taboada, 2013[21] Hannon, 2009[29]

1.4 The Study of Campylobacter Genomics

The first fully annotated *Campylobacter* genome sequence, was published in 2000[30]. The NCTC 11168 isolate, belonging to ST-21, demonstrated a circular chromosome of 1,641,481 bp. A total of 1654 protein-coding sequences and 54 stable RNA sequences were identified. Insertion sequences and phage-associated sequences were absent, and there were very few repeat regions, however short hypervariable regions were present in high numbers.

The NCTC 11168 genome has subsequently been re-annotated, reducing the number of genes identified from 1654 to 1643 and identifying a number of functional differences from the original annotation[31].

The availability of *Campylobacter* reference genomes has facilitated the study of *Campylobacter* genomics, and numerous studies report *Campylobacter* genomes. A PubMed search on 10th March 2016, using the search terms 'Campylobacter' AND 'genome' returns 1210 entries. Next generation genome sequence data have been used to address contemporary public health questions, for example the study of an excess of Guillain-Barre cases in northern China[32]; and questions regarding the evolution and basic biology of the genus, for example a study of the mechanisms of genomic diversification in *Campylobacter*[33].

While a number of analytical methods have been developed and used, two underlying approaches have predominated for a number of years, that based on homologous sequence variation in genes uniformly present in isolates; and that based on non-

homologous sequence variation, or the presence and absence of genes or short sequences[34].

More recently there has been a move to identify the core genome and accessory genome and to combine the analysis of homologous and non-homologous variation in a pan-genomic or comparative genomics approach[34].

1.5 Contrasting Population Structure of *Campylobacter jejuni* and *Campylobacter coli*

A neighbour joining tree (Figure 1) shows that *C. jejuni* and *C. coli* display markedly different population structures. *C. jejuni* populations are highly structured into Clonal Complexes (Figure 2). When each locus is considered separately, there is evidence of considerable recombination within *C. jejuni*, with alleles from different lineages appearing within the same STs. In contrast, *C. coli* displays far greater genetic diversity, with three deep-branching clades, of which Clade 1 contains the vast majority of lineages described to date. The ST-828 clonal complex, part of clade 1, accounted for over 70% of the 5954 *C. coli* isolates submitted to the PubMLST database (http://pubmlst.org/campylobacter) before July 27th 2015, with most of the remainder sharing alleles with these, and therefore also being related. The second most common clonal complex (ST-1150 complex), also from clade 1, accounts for only 2% of isolates.

Figure 1. Contrasting population structure of *C. jejuni* and *C. coli* neighbor joining trees.

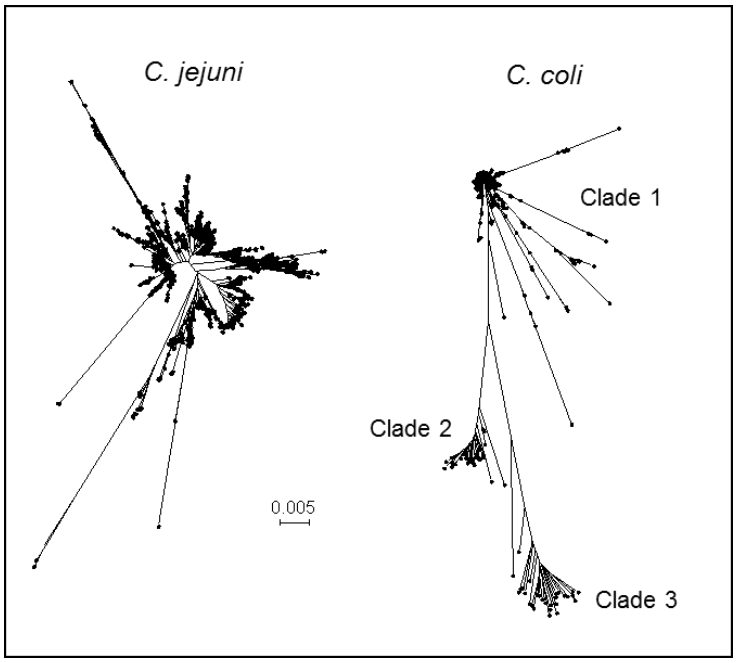


Figure 1. The genetic relatedness of 1341 *C. jejuni* and *C. coli* genotypes based on concatenated MLST alleles (3309bp) from published studies[35,36]. Contrasting tree topologies are visible on the neighbour-joining trees with three deep branching clades present among *C. coli* genotypes.

Figure 2. Contrasting population structure of *C. jejuni* and *C. coli* goeBurst diagrams

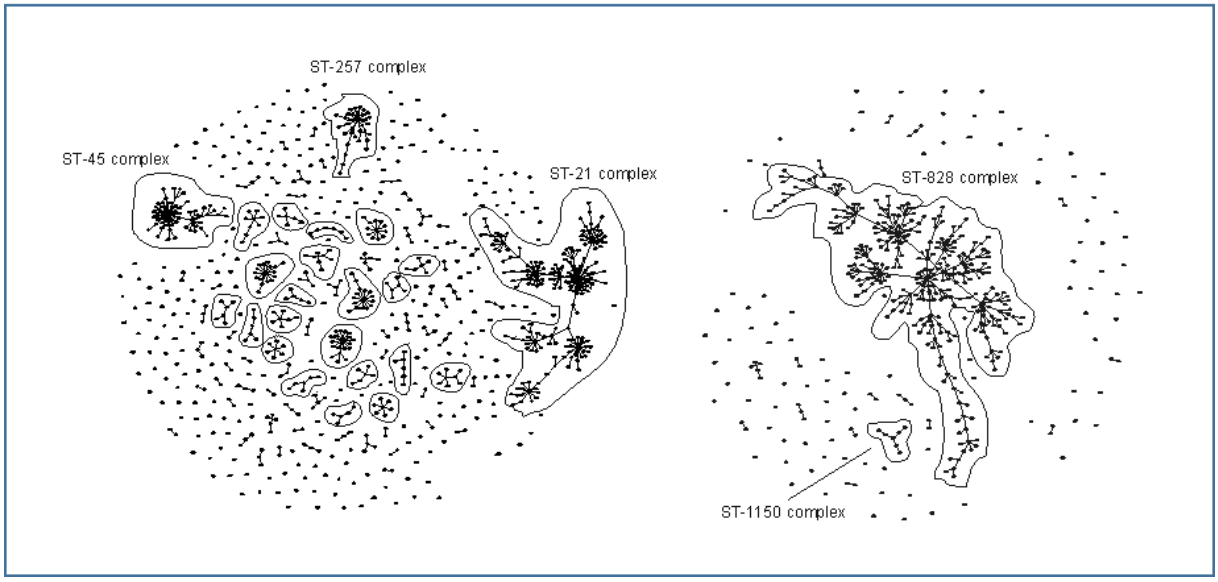


Figure 2. Comparison of the population structure of *C. jejuni* and *C. coli.* Different 7-locus genotypes are represented by points on a goeBURST diagram; strains differing at a single locus are joined by a lines that infer linkage by decent. Cluster size distribution is different for the two species with many more clonal complexes found among *C. jejuni* genotypes that within *C. coli* where most of the typed strains belong to the ST-828 complex.

1.6 Trefoil Structure in *Campylobacter coli*

The emergence and maintenance of the 3-clade structure in *C. coli* implies that three distinct bacterial gene pools exist, and that although recombination is evident within *C. coli* clades, there are or have been barriers to recombination between clades. Recombinational barriers can be considered in three broad categories: (i) adaptive, implying selection against hybrids; (ii) mechanistic, imposed by homology dependence of recombination or other factors promoting DNA specificity; (iii) ecological, a consequence of physical separation in different ecological niches[37]. In order to consider

the relative importance of each type of barrier in the evolution of a clade structure in *C. coli* it may be useful to look at both *C. coli* and *C. jejuni* in context.

Campylobacter jejuni and *C. coli* are approximately 12% divergent at the nucleotide level and are considered distinct microbial species, however there is strong evidence for a degree of hybridisation between the species through a process of horizontal gene transfer (HGT)[37,38]. Statistical model-based approaches have been used to investigate the sharing of both whole alleles and recombined elements or 'mosaic alleles' between *C. jejuni* and *C. coli*. While *C. coli* clade 1 remains distinct from clades 2 & 3, there is evidence of gene flow between *C. jejuni* and *C. coli* clade 1. Analysis of 1738 alleles from a total of 2953 Sequence Types identified 31 mosaic alleles, of which 25 occurred within *C. coli* isolates that had acquired *C. jejuni* DNA, and the remaining 6 had originated in *C. coli* and been acquired by *C. jejuni*. With the exception of a single mosaic allele, having originated in *C. coli* clade 3 and being acquired by *C. jejuni*, all genetic exchange events identified involved *C. coli* clade 1 as either donor or recipient.

The existence of hybrids and the maintenance of alleles of *C. jejuni* origin within the *C. coli* gene pool demonstrates that mechanistic barriers are not preventing interspecies gene flow. Furthermore, the resultant hybrid lineages are not sufficiently maladapted to prevent their proliferation (adaptive barrier). Ecological barriers to recombination are therefore likely to have been important in generating and maintaining the observed population structure in *C. coli* and *C. jejuni* species, clades and clonal complexes.

1.7 *Campylobacter* Epidemiology

Asymptomatic carriage of *Campylobacter* in humans has been reported in developing countries[39–42]. There is a dearth of reports of studies of asymptomatic *Campylobacter*

carriage in humans in industrialised countries, and prevalence of carriage in industrialised nations is believed to be low[43]. The apparent absence of asymptomatic carriage of *C. jejuni* and *C. coli* among individuals in the UK suggests that humans may not be a natural host for these organisms in high income countries, and may have undergone a relatively short history of co-evolution. For this reason, an appreciation of phylogenetic relationships, together with an examination of the ecology of *Campylobacter* can enhance understanding of the origin and causes of human campylobacteriosis and this forms the basis for much of the recent work to explain the epidemiology of these organisms[37,44–52] The majority of human campylobacteriosis cases are believed to occur sporadically, with an estimated rate of 9.3 cases per 1000 person-years in the community, although only an estimated 14% of cases present to a medical practitioner[6]. There is an apparent over-representation of UK *Campylobacter* outbreaks or epidemics in the academic literature. A PubMed literature search carried out on 17th March 2015, using the terms 'UK human Campylobacter sporadic' OR 'UK human Campylobacter endemic' yielded 28 relevant hits. On the same date a search using the terms 'UK Human Campylobacter outbreak' OR 'UK human Campylobacter epidemic' returned 67 relevant papers. This is likely to be the consequence of preferential follow-up and publication of outbreaks compared with sporadic disease, rather than a genuine reflection of epidemiology, although some recent research suggests that the actual burden of outbreak-related *Campylobacter* may be significantly underestimated[53], as a single outbreak may resemble dispersed and unrelated cases. A number of published outbreak studies using Multi-Locus Sequence Typing (MLST) have indicated that food preparation practices may result in a single 'point-source' of infection containing multiple *Campylobacter* Sequence Types (STs). This has been demonstrated particularly

in liver-related outbreaks, where chicken or duck liver dishes are prepared in bulk using homogenised under-cooked livers from a large number of individual birds, thereby creating a pool of a number of *Campylobacter* genotypes[54–56]. In the absence of a follow-up investigation using traditional epidemiological techniques, there would be no evidence to link the cases to a common source. Colonisation by multiple *Campylobacter* genotypes is also possible in an individual bird[57,58]. The occurrence of coinfection with multiple types of *Campylobacter jejuni* in human patients has been demonstrated[59,60] although the dynamics of competing strains and the ratio in which different strains are shed has not been extensively studied and is not well understood.

A study based in the Manawatu region of New Zealand used a Bayesian hierarchical model to identify potential *Campylobacter* outbreaks from a background of high sporadic disease incidence[61]. The model functioned independently of genotypic data, focusing solely on temporal and spatial data to identify anomalous clusters of increased incidence, using *Campylobacter* notification data for the region from the period between 2001 and 2007. Known outbreaks were correctly identified by the model, and a number of additional spatio-temporally localised clusters were identified for further investigation using traditional epidemiological methods such as questionnaires.

Similarly, a study in the north west of England found marked spatial and temporal clustering, consistent with a pattern of small scale point-source outbreaks, of *Campylobacter* cases that comprised various Sequence Types (STs) and were reported as sporadic events[62].

These studies highlight the benefits of studying *Campylobacter* epidemiology from a multi-disciplinary perspective, using more than one analytical approach to build up a picture of *Campylobacter* epidemiology.

1.8 The Ecology of *Campylobacter*

Multiple factors affect the epidemiology of campylobacteriosis in humans, including host characteristics and pathogen characteristics. An understanding of the ecology of *Campylobacter* in host reservoirs and the environment is essential in understanding the human behaviours which expose susceptible human hosts to infection by *Campylobacter jejuni* or *Campylobacter coli*.

The relationship between livestock and *Campylobacter* has long been documented[63–65]. Analysis of the genotypes of *Campylobacter* isolated from human disease cases has shown that the vast majority of campylobacteriosis cases are caused by *C. jejuni* and *C. coli* lineages also found in livestock, particularly cattle and chickens (Figure 3). Clinical isolates do not reflect the breadth of diversity seen in livestock, but are a non-random subset of these lineages. This is particularly marked in the case of *C. coli*, in which the average diversity per locus is 13 alleles in disease cases compared with 55 alleles in the general *C. coli* population[35].

Campylobacter infection has traditionally been considered asymptomatic in livestock[66] because of the absence of the recognisable gastrointestinal syndrome characteristic of human clinical cases, however more recent studies have indicated that *Campylobacter* infection in chickens is correlated with immune responses and pathologies which are suggestive of a non-commensal relationship between bacteria and host[67,68].

Campylobacter jejuni and *C. coli* are known to be transmitted via the faecal-oral route, which means that *Campylobacter* infection is the direct result of consumption of bacterial cells, which have been excreted by another host. An estimated 50-70% of human *Campylobacter* infection is caused by consumption of poultry meat contaminated with faecal material[63]. Slaughter methods used in large-scale food production create opportunities for the contamination of animal carcasses. A study carried out in Switzerland, where the agricultural system is based primarily on small mixed farms demonstrates that the majority of cross-contamination of *Campylobacter jejuni* types in chickens happens not on-farm but in the abattoir, and usually concerns *Campylobacter* types originating and being spread within the same flock, rather than being the result of insufficient decontamination processes between flocks, for example[58].

Figure 3. Comparative prevalence of clonal complexes in humans, chickens and cattle.

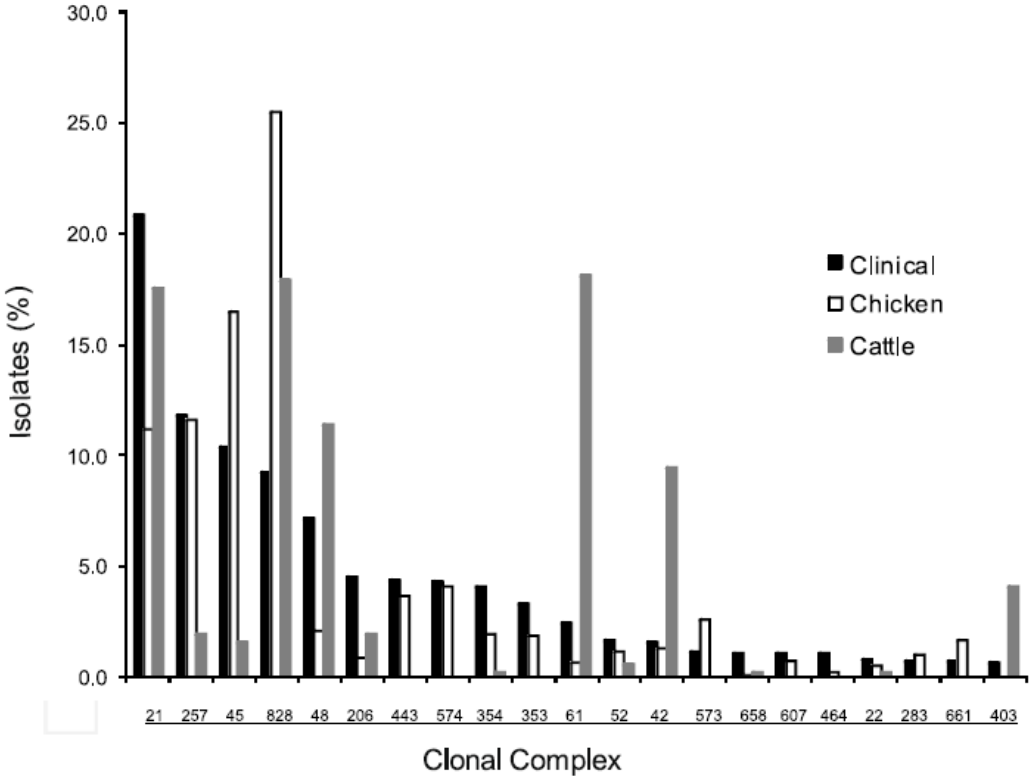


Figure 3. Human clinical isolates and those from chicken and cattle (faeces and meat). The relative abundance of clonal complexes (responsible for >1% of total UK disease) of isolates from human *C. jejuni* and *C. coli* infections and those from published chicken and cattle isolate collections[35,69]. All of the 21 most common disease causing clonal complexes are also found in cattle, chickens or both. Figure reproduced with author's permission.

1.9 Ecology and Host Association

There is evidence of association between clusters of related genotypes and the source or host from which the bacteria were isolated. At the species level, *C. jejuni* and *C. coli* have subtly different host ranges. Both species are found in a wide range of wild and

farm animals but *C. jejuni* dominate numerically in most sampled wild bird species[37,70] as well as chickens and cattle[69]. *C. coli* (clade 1) are also common in chicken and cattle, usually constituting around 10% of the *Campylobacter* population in these host animals (90% *C. jejuni*); but are more abundant than *C. jejuni* in pigs[71]. Within *C. coli,* isolates belonging to clades 2 & 3 are far less frequently identified and are usually isolated from environmental sources where they may be associated with waterfowl[72].

The host-genotype relationship goes further than this. In *C. coli* and *C. jejuni* there is a strong association between specific clonal complexes, STs, and alleles with host species[71,73]. This association is stronger than spatial or temporal signals and statistical assignment analyses show considerable consistency in grouping isolates from a range of host animal sources regardless of geographical source[36]. For example, a population of *C. jejuni* isolates from UK chickens is strikingly similar to a population of *C. jejuni* isolates from chickens in the US, mainland Europe or Senegal[36]. The equivalent is true of cattle, pigs and turkeys[36]. This host allelic signature between diverse lineages creates a pool of alleles common to a given source[73] and this signal of host association has been widely used to assign the origin reservoir of clinical isolates[69,74–76]. All of these studies identify farm-, especially chicken-, associated isolates as the main source of human infection. A summary of studies which have used MLST for the attribution of human campylobacteriosis can be found in table 2.

Table 2. Summary of studies using MLST for the attribution of human campylobacteriosis cases to animal host sources.

Description	Percentage of Clinical Cases Attributed to Each Source	Reference
Attribution of clinical campylobacteriosis cases to putative host source, using a reference set of 3419 isolates. Results of STRUCTURE and Asymmetric Island Model (AIM) compared.	STRUCTURE (*C. jejuni/ C. coli)*: Chicken: 58% / 40% Ruminants: 38% / 54% Wild Bird & Environmental: 4% / 0% Pigs: 0% / 6% AIM (*C. jejuni/ C. coli)*: Chicken: 78% / 56% Ruminants: 18% / 42% Wild Bird & Environmental: 4%	Sheppard *et al*, 2009[69].
Attribution of 3-year clinical dataset from Manawatu, New Zealand to putative host source using a reference set of 467 isolates. AIM used.	Poultry: 76% Ruminants: 18% Other: 4%	Mullner *et al*, 2009[74].
Attribution of clinical isolates from cases of campylobacteriosis in children aged under 5 years. Results from urban and rural north east Scotland compared.	Rural/ Urban: Cattle: 42% / 35% Birds (not chicken): 24% / 6% Chicken: 19% / 43% Sheep: 12% / 15%	Strachan *et al*, 2009[76].
Attribution of 1231 human disease cases in Lancashire, using 1145 reference isolates. AIM introduced and used.	Chicken: 56.59% Cattle: 35.10% Sheep: 4.35% Wild Birds: 1.67% Other Sources: 2.54%	Wilson *et al*, 2008[75].

As previously discussed, *C. jejuni* and *C. coli* have different host ranges and there is evidence that they exhibit colonisation and virulence factors differentially in response to different growth conditions, which may relate to host preferences[77]. For example, there is a wide range of carbon sources that *C. jejuni* utilize more effectively at 42°C rather than the lower temperature of 37°C[78]. The average core temperature of a chicken is 42°C, while a pig is 39°C so this could influence the ability of *C. jejuni* to colonize different hosts. Serine dehydratase, encoded by the *sdaA* gene has been demonstrated to be an essential colonisation factor in *C. jejuni*[79]. This gene is also expressed in *C. coli*, but the functionality of the enzyme is highly dependent on temperature[2]. In *C. coli* there is little or no serine dehydratase activity at 42°C, but at the lower temperature of 37°C activity is significantly increased, this could provide a partial explanation for the porcine host association with *C. coli* [2,78].

1.10 Why Do Farm Associated Isolates Cause Disease?

There are two possible explanations for the strong correlation between genotypes that cause human disease, and those that are associated with farm animals, especially chickens and ruminants. First, this could be the result of differential exposure. By definition, humans are more frequently exposed to domesticated food animals than to wild reservoirs of infection. The main risk factors for human campylobacteriosis include handling and consumption of raw or under-cooked poultry[11,80]; handling and consumption of barbequed meat[81]; contact with farm animals[11] and consumption of unpasteurised milk. These risk factors are all common behaviours which present opportunities for exposure to domesticated animals and animal products, whilst exposure to wild and environmental sources of *Campylobacter* may be less common. It

is therefore possible that all *Campylobacter* strains are equally infective and the dominance of farm associated genotypes in human disease is simply reflective of greater exposure to these strains.

An alternative explanation is that certain strains are more likely to cause acute infection than others. If it were the case that agricultural strains were more pathogenic to humans then they would be over represented in surveys of reported clinical cases. While this may be a less likely explanation than simple differential exposure, some genotypes do appear particularly well adapted to very specific ecological niches and in an evolutionary trade-off their ability to colonise diverse hosts may have been lost. There are numerous examples of host restricted STs among strains found only in specific wild bird species[82,83]. Genetic isolation could explain this but different colonization capacity could also be important. For example, *C. jejuni* strains (ST-3704) that are repeatedly found in the gut of bank voles are unable to colonise the chicken gut in laboratory experiments[84]. In a similar experiment, using a European Robin (*Erithacus rubecula*) infection model, *C. jejuni* from song thrushes (*Turdus philomelos*) successfully colonized but *C. jejuni* from human patients did not[85].

Colonisation and virulence factors in *Campylobacter* are not well understood, but evidence of differential abilities to invade the cells of different hosts[82–85] points to a possible explanation for the relationship between specific STs and human disease. Explanations based on differential exposure and colonization capacity are not mutually exclusive. It is plausible that those lineages that are found in a niche to which humans are routinely exposed have acquired the necessary colonisation factors to persist in this environment, and opportunistically to infect humans.

1.11 What is the Reservoir of Livestock *Campylobacter*?

While the reservoir and mode of transmission of *Campylobacter* to humans is well documented and understood the question of where livestock, and particularly chickens, acquire *Campylobacter* remains largely unanswered. The phrase 'farm to fork' has been coined to refer to the progress of food through the processing chain from production to consumption but it is necessary to consider what feeds into the farm environment, and the non-anthropogenic forces that could play a role in shaping the epidemiology of *Campylobacter* in the UK.

The poultry industry has instituted some major safeguards to try and reduce the prevalence of *Campylobacter* infection in chickens, and therefore to reduce the incidence of campylobacteriosis in humans. Fly screens were implemented in broiler houses in Denmark to reduce the influx of houseflies (*Musca domestica*) through air ducts[86,87]. This intervention was based on the observed correlation between the peak in housefly populations and the peak in *Campylobacter* prevalence on broiler carcasses at slaughter during the summer months. Accordingly the fly screens reduced the prevalence of *Campylobacter* positive flocks by an estimated 70%, based on extrapolation from prevalence estimates in broiler houses with and without the intervention[87]. Unfortunately there was no simultaneous molecular typing of *Campylobacter* isolates from chickens and fly carcasses, although a separate study in Denmark has used Pulsed-Field Gel Electrophoresis (PFGE) to demonstrate a correlation between a subset of *Campylobacter* types carried by flies and chickens from the same house[88]. This study, however, isolated *Campylobacter* from the dead carcasses of flies sampled from within the chicken houses and the possibility of *post mortem*

contamination with *Campylobacter* shed by chickens within the same space remains a possibility. Unfortunately there remains a lack of solid evidence to incriminate a particular source of *Campylobacter* in chickens or in livestock more generally, so many of the biosecurity methods implemented are based largely on conjecture, and the prevalence of *Campylobacter jejuni* and *C. coli* both in farmed livestock and in human disease remain high.

It is well known that brown rats (*Rattus norvegicus*) are common farmyard commensals, but to date there has been little research into the role they may play in the *Campylobacter* transmission cycle. A PUBMED search for the terms (Campylobacter wild Norway rat) or (Campylobacter wild brown rat) or (Campylobacter wild *Rattus norvegicus*) returned only eight English language articles and a single article in Spanish. Of these nine entries, five were rejected as irrelevant. The remaining four entries cover the period of time from 1977 to 2011. No entries more recent than 2011 were returned.

An early study of the isolation and nature of *Campylobacter* from laboratory and wild rodents[89] noted the occurrence of *Campylobacter* in three out of 29 laboratory-reared and kept Norway rats. Following presumptive identification by colonial morphology and growth patterns and Gram staining, traditional biochemical tests were carried out. The results suggested that the rat isolates were *Campylobacter coli*, and the researchers initially suggested a role for rats in the wild as reservoirs for porcine *Campylobacter coli*. Subsequent molecular testing carried out by comparing electrophoretograms of the acid plus phenol soluble (APS) proteins demonstrated small but marked differences between the three rat *Campylobacter* isolates and *Campylobacter coli* isolated from a symptomatic pig during an outbreak of porcine dysentery. The authors conclude that

the higher than expected prevalence of *Campylobacter* in rodent species may be of public health importance, with some cases of human campylobacteriosis potentially having arisen from contact with rodents rather than domestic animals. The authors report that this confirms the hypothesis that rats can be a reservoir of *Campylobacter*[89].

The presence of *Campylobacter* within the digestive tracts of laboratory rats, kept in isolation from other wild and farm animals, albeit with low prevalence, is interesting and may be suggestive of a commensal relationship between rats and *Campylobacter*, although the experimental methods reported do not definitively rule out the possibility of contamination from another source, for example a laboratory worker or a fomite entering the rats' living space. Further, the small size of the study and the poor resolution of the typing methods used do not really lend themselves to definitive conclusions regarding *Campylobacter* in rats.

The authors also noted that traditional biochemical tests to distinguish between *Campylobacter* species are too crude an instrument and that more sophisticated fine-typing methods using molecular markers are increasingly important.

The necessity for molecular tools for distinguishing between closely related strains of *Campylobacter* was reiterated by a later study from New Zealand on the relationship between intestinal *Campylobacter* species isolated from animals and humans as determined by Bacterial Restriction Endonuclease DNA analysis (BRENDA)[90]. A total of 18 *Campylobacter jejuni* isolates from Norway rats were examined, although it is not stated whether they were wild or captive-bred rats. The rat *Campylobacter* isolates were sub-divided into four separate BRENDA types, of which two were indistinguishable from

isolates from humans, chickens and a horse. The authors concluded that rats *may* play a role as carriers of infection for humans and other animals.

More recently a study of the presence of *Salmonella* and *Campylobacter* spp. in wild small mammals on organic farms in the Netherlands[91] identified a relatively low prevalence of *Campylobacter* spp. in house mice (9.64%) and Norway rats (12.5%), which were caught in live-traps on nine organic pig farms and one organic broiler farm during the summer and early autumn of 2004. *Campylobacter* isolates were speciated and genotyped using the Amplified Fragment Length Polymorphism method (AFLP). Rodent source isolates were compared with isolates from samples of pig manure collected at two of the study farms. AFLP did not identify any genotypes shared between the rodents and the pig manure samples. It is unclear why livestock faecal samples were collected only from selected pig farms, neglecting the chicken farm; equally the ratio of nine pig farms to a single chicken farm within the study is surprising and is not clearly justified within the text. The authors acknowledged that limited conclusions can be drawn from a study of this size, however they concluded that effective on-farm rodent management is required. Although this statement is non-controversial it is not well supported by the evidence from the study.

A study of the frequency of detection of *Escherichia coli, Salmonella* spp. and *Campylobacter* spp. in the faeces of wild rats in Trinidad and Tobago[92] also highlighted the importance of rodent control for public health. In this study 204 wild rats were caught predominantly in busy public spaces, such as markets and shopping areas, in the vicinity of fast food restaurants and in residential areas. The prevalence of *Campylobacter* isolated from rat faeces was relatively low at 3.4%. *Salmonella*

prevalence was also low at 2.0%, but *E. coli* showed high prevalence at 83.8%. Despite the low prevalence of *Campylobacter* amongst the rats, *Campylobacter* isolates displayed high resistance to antimicrobial substances, at 85.7%. Molecular typing was not carried out in this study to test for a correlation between 'rat types' and 'human disease types' of *Campylobacter* however the locations of sampling sites would in any case raise the question of whether rats may perhaps be incidental carriers of *Campylobacter* rather than a source, perhaps becoming infected by scavenging on the same food sources that may infect a human host. In the same way family pets have demonstrably been infected with the same *Campylobacter* types as their owners by eating from a common source[93]. These four studies all demonstrate the ability of Norway rats to become infected with, and to shed *Campylobacter*. There are significant gaps in the body of knowledge surrounding the epidemiology of *Campylobacter* in rats and their potential role as reservoirs or vectors of the lineages found in human disease.

1.12 Antimicrobial Resistance in *Campylobacter* Reservoirs

Mechanisms of Resistance

Fluoroquinolone resistance has been reported in a wide range of Gram positive and Gram negative organisms[94]. The most common mechanism of resistance to this class of drugs is the alteration of large enzymatic structures due to mutations in the DNA gyrase or DNA topoisomerase IV genes [95]. In both *C. jejuni* and *C. coli* a single point mutation at the *gyrA* locus causes amino acid substitutions in the quinolone resistance determining region on the surface of the *gyrA* subunit of the enzymatic quaternary structure, inhibiting the ability of fluoroquinolones to bind to the organism[95]. A mutation in the

parC locus in *Campylobacter* isolates has also been reported[96], although the presence of the *parC* gene in *Campylobacter* is questionable, as subsequent studies using whole genome sequencing have failed to identify the gene within the *Campylobacter* genome[31,95].

A variety of macrolide resistance mechanisms have been reported in different bacterial species[95]. In *Campylobacter* species the modification of the 23S rRNA target gene caused by a 2bp mutation in the peptidyl transferase region is the most common mechanism of macrolide resistance[95] although efflux pumps are also commonly reported[32,97] and may be observed in conjunction with the nucleotide mutations in the 23S rRNA gene [95].

Efflux pumps mediating resistance to tetracyclines are present in many bacterial species[95], however the predominant mechanism of tetracycline resistance in *Campylobacter* is the production of ribosomal protection proteins, interfering with the ability of the drug to bind to the organism[95]. Ribosomal protection proteins in *Campyobacter* are encoded by the *tet(O)* gene, which is commonly found on a self-transmissible plasmid, although increasingly has also been identified in the *Campylobacter* chromosome[95].

Resistance to aminoglycosides is the result of the production of the 3'-O-aminoglycoside phosphotransferase enzymes, which modify and inhibit the drugs[95]. The enzymes are widely distributed throughout Gram negative and Gram positive bacteria and the aminoglycoside-aminocyclitol resistance genes which encode them may be carried by a plasmid or incorporated in the chromosome. In the case of *Campylobacter* the *aphA-3* gene has been identified on plasmids[95].

DNA take-up and recombination

As a highly recombinogenic genus[18], *Campylobacter* has a natural propensity for the acquisition of DNA both from within the genus, and from distantly related organisms[95], therefore the acquisition of antimicrobial resistance genes and their rapid transfer between *Campylobacter* strains is problematic[95]. There are three main mechanisms by which bacteria acquire foreign DNA. Transformation and conjugation are two processes which involve the passage of DNA from one bacterium to another through bacterial membranes[98]. Transformation is the process of acquiring foreign DNA from the surrounding matrix of a bacterial organism. The *Campylobacter* genus is one of several bacterial genera which have been identified as 'competent' or capable of DNA uptake by transformation. In laboratory conditions, *Campylobacter* has been demonstrated to take up DNA originating from its own species in preference to *E. coli* chromosomal DNA[99] although the *tetM* gene, associated with tetracycline resistance, has demonstrably been acquired by *Campyobacter* from an *E. coli* plasmid[99].

Conjugation involves conjugative plasmids transferring themselves between bacteria[100]. This is one of the most widespread and efficient mechanisms for the spread of antimicrobial resistance genes among bacteria[100], and readily facilitates the transfer of DNA between very distantly related bacterial species. *Campylobacter* organisms have been demonstrated to acquire plasmids mediating tetracycline resistance in this way[101].

Transduction allows the transfer of plasmid or chromosomal DNA from one bacterial organism to another by a bacteriophage viral vector. This process does not require physical contact between the donor and recipient organisms, however the scope for

DNA transfer is very narrow, as a bacteriophage vector must be compatible with both donor and host organism[102]. Bacteriophages capable of vectoring DNA have been identified for both *C. jejuni* and *C. coli*[103]. In experimental conditions, species-specific bacteriophages have been used to transfer resistance to erythromycin between *C. coli* isolates[103].

The finding from Trinidad and Tobago that rat *Campylobacter* isolates display a high prevalence of resistance to antimicrobial substances[92] is consistent with evidence from the limited number of studies of antimicrobial resistance in populations of wild animals, discussed in a review paper on antibiotic resistance genes in natural environments[104]. Studies show that wild populations of baboons and apes living in contact with human populations display a higher prevalence of antibiotic resistant bacteria than those living in remote areas, not populated by humans[105]. Similar findings have been observed in populations of wild birds[106–108].

Antimicrobial resistance genes do occur naturally in populations not subject to antimicrobial use or consumption of contamination from pharmaceuticals, suggesting that mass dosing of artificially sustained populations of humans or animals is not a pre-requisite for the development of antimicrobial resistance. A study of the Murray Collection of 433 enterobacterial isolates collected from around the world between 1917 and 1952 demonstrated a 2.5% prevalence of resistance to either ampicillin or tetracycline before their discovery and use in the clinical setting[104]. Some putative explanations for the phenomenon include stochastic mutation events which are not deleterious, a response to the selective pressures of antimicrobials produced in natural ecosystems, and the horizontal transfer of transposons, integrons and plasmids[104].

Accepting that antibiotic resistance can occur in natural populations is not problematic, but random stochastic events or small-scale changes in response to selective pressures in low-density, remote populations do not explain the proliferation of antibiotic resistance in human clinical *Campylobacter*. A number of studies show that resistance to certain antimicrobial agents in *Campylobacter* isolated from symptomatic humans is at a level which causes concern[95,109]. The role of chickens bred for human consumption as effective incubators and multipliers of drug resistant *Campylobacter* has been suggested by multiple studies[110,111,112].

A longitudinal study of *Campylobacter* isolates from poultry products and from human stools in The Netherlands between 1982 and 1989 recorded increasing prevalence of quinolone resistance, from 0% in both humans and chicken at the beginning of the study to 11% and 14% respectively by the end of the study. The increase is correlated with an increase in use of quinolones in both veterinary and human medicine in The Netherlands during this time[111]. More recently the association between veterinary antimicrobial usage and resistance prevalence has been reiterated in a review paper on antibiotic resistance and antibiotic resistance mechanisms in *Campylobacter*[95]. European law now stringently controls the use of veterinary drugs. Licensed antibiotics are permitted for therapeutic use, but their use as growth promotors is expressly prohibited. Nevertheless the modern livestock farming environment has been identified as a niche which creates favourable conditions for the emergence and maintenance of populations of antibiotic resistant bacteria. In commercial farming antibiotics are administered through water at the flock or herd level, rather than the individual level, making accurate dosing difficult to achieve. The stock density of commercially farmed animals favours transmission of

pathogens within the flock or herd, allowing bacterial strains with a competitive advantage such as antimicrobial resistance to proliferate[109]. A Scottish study of antimicrobial resistance profiles in chicken, cattle and human *Campylobacter* isolates suggests that the similarity between resistance profiles and distributions of Minimum Inhibitory Concentrations (MICs) in chicken and human isolates, and their distinct differences from cattle isolates can be used to support the attribution of antibiotic resistant strains of *Campylobacter* from human disease cases to a chicken source[113]. It should be noted, however, that chicken isolates within this study were exclusively from chicken meat and cattle isolates were exclusively from cattle faeces. A more direct comparison between isolates from different species at the same point in the production process would provide more compelling evidence to support this conclusion.

1.12 Dating Lineage Divergence

Molecular methods can be applied to immediate public health questions regarding *Campylobacter* epidemiology, they also have an application in the study of longer-term population trends. Genotyping isolates from various sources can offer insight into the causes of the genetic structuring in *Campylobacter* populations. However, a more comprehensive understanding of the evolution of the genus can be obtained if the time scale for the divergence of lineages can be overlaid upon the tree of genetic relatedness. By cross-referencing estimated dates of divergence within the genus *Campylobacter* with ecological data it is possible to make inferences about the conditions which created the specific barriers which led to speciation, the formation of the lineage structure, and the gene-flow between certain clades and clonal complexes.

The traditional method for dating bacterial evolution is based on the rate of sequence divergence between *Escherichia coli* and *Salmonella typhimurium*, which is assumed to be 1% 16S rRNA divergence per 50 million years[114]. Applying this method to the *Campylobacter* genus estimates the *C. coli – C. jejuni* split to have occurred approximately 10 million years ago, and the divergence of 3 *C. coli* clades about 2.5 million years ago. An alternative dating method, using a molecular clock based on intra-specific diversity in *C. jejuni,* places these splitting events much more recently[18]. The speciation of *C. coli* and *C. jejuni* has been estimated to have occurred around 6,500 years ago, with *C. coli* clade divergence occurring 1,000-1,700 years ago[35]. While this large disparity between estimates is difficult to explain, there are reasons for favouring the more recent estimates for *Campylobacter* divergence. Methods that provide recent estimates are based on knowledge of genetic variation within the genus *Campylobacter* and not on the divergence of genera (*E. coli* and *S. typhimurium*) only distantly related to *Campylobacter*, and there is an increasing number of studies that use similar approaches and infer a more rapid rate of molecular evolution than in traditional models of bacterial evolution[18,115,116].

If the diversification leading to the population structure in extant *Campylobacter* populations is thought to have occurred in the last 6,500 years then it corresponds with important changes in human behaviour. For example, the development of agriculture, which began in the middle east about 10,000 years ago and became common in Europe about 5,000-3,000 BC or the establishment of the first cities and the rise of urbanization. Clearly this could have provided novel opportunities for *Campylobacter* to expand into

new host species and infect humans in a way that is, to some extent, mirrored in modern society and may have begun to shape the population structure that we observe today.

1.13 Scope of the Thesis and Specific Aims.

This thesis stems from a comprehensive 14-month study of human clinical campylobacteriosis in Oxfordshire, funded by the Health Protection Agency (HPA) and PathoNGen Trace, an EU-funded consortium. The study was carried out at a time of transition between traditional MLST techniques and next generation sequencing methods, allowing a direct comparison between methodologies.

Chapter one focuses on the descriptive molecular epidemiology of reported human *Campylobacter* during the study period from April 2010 to July 2014, using MLST data derived from traditional PCR techniques and from next generation sequencing. Genomic data have also been used to infer additional information of public health importance, specifically ciprofloxacin resistance phenotype.

The specific aims of chapter one are:

- To develop a protocol for the characterisation of *Campylobacter* isolates for use in sentinel surveillance and clinical public health settings.
- To describe the molecular epidemiology of human campylobacteriosis in Oxfordshire between April 2010 and July 2011.
- To compare the quality of MLST data obtained by PCR and Sanger sequencing with MLST data obtained by Illumina whole genome sequencing.
- To assess the prevalence of ciprofloxacin resistance in human *Campylobacter* isolates.
- To compare the epidemiology of domestically acquired campylobacteriosis cases with putative travel-related campylobacteriosis.

Chapter one provides the context for the thesis, describing the *Campylobacter* public health picture in Oxfordshire, as a representative model of human campylobacteriosis in the UK more generally. The subsequent chapters take a more applied approach, with a greater emphasis on analysis of data, within the public health context outlined by chapter one.

Chapter two reports a study of antimicrobial resistance in *Campylobacter* isolates from retail poultry in the UK. The study uses traditional PCR-derived MLST data to examine the patterns of antimicrobial resistance within and between lineages. This chapter has been published[117].

The specific aims of chapter two are:

- To quantify the prevalence of antimicrobial resistance in *Campylobacter* isolated from retail poultry.
- To investigate the relationship between lineage and antimicrobial resistance profile, and to consider the evolutionary and public health implications.

Chapter three reports a study of *Campylobacter jejuni* and *Campylobacter coli* in farm-dwelling Norway rats, which represent a potential bridging population spanning both the sylvatic and farmyard environments. The study uses whole genome data to provide much higher resolution than seven-locus MLST, and considers the evolutionary and ecological inferences that can be made using this additional data. In particular the sub-structuring of lineages within the ST-45 CC is explored.

The specific aims of chapter three are:

- To examine the role that wild, farm-dwelling Norway rats may play in the on-farm *Campylobacter* transmission cycle, using the gene-by-gene approach.
- To compare the information provided by seven-locus MLST and whole genome sequence data.

Chapter four returns to the Oxfordshire human clinical *Campylobacter* data and links the question of human disease with the prevalence of *Campylobacter* in animals – both farmed and wild. Whole genome data for both human clinical isolates and animal isolates are examined and the potential for improving the attribution of human disease cases to host sources is explored.

The specific aims of chapter four are:

- To evaluate the use of traditional seven-locus MLST for the attribution of human clinical campylobacteriosis cases to putative host sources.
- To explore the potential for using the gene-by-gene approach across the *Campylobacter* genome to develop an alternative attribution typing scheme.

Chapter Two: Analytical Methods and Approaches

2.1 Seven-Locus MLST

Multi-Locus Sequence Typing (MLST) is a robust molecular typing scheme which is easily transferable between laboratories, and therefore allows the easy comparison of any number of bacterial isolates regardless of isolate provenance or typing centre. The typing methodology was originally developed for typing *Neisseria meningitidis*[17] but has subsequently been expanded for use with a wide range of bacterial species, including *Campylobacter jejuni* and *Campylobacter coli*[118]. The *Campylobacter jejuni* and *Campylobacter coli* MLST scheme (henceforward referred to as *C. jejuni/ coli* MLST) uses fragments, between 402 bp and 507 bp in length, of the following seven housekeeping genes to identify and compare types: *aspA* (aspartase A), *glnA* (glutamine synthetase), *gltA* (citrate synthase), *glyA* (serine hydroxymethyltransferase), *glmM* (phosphoglucosamine mutase)[*], *tkt* (transketolase) and *uncA* (ATP synthase α subunit). Housekeeping genes are core genomic loci and are presumed to be under neutral or near-neutral selection pressures[119]. While the seven gene fragments of interest are not evenly distributed around the chromosome they are sufficiently distant from one another (a minimum of 70kb between each pair of loci) in the NCTC 11168 reference genome that it would be unlikely for any two or more of these genes to be affected by the same recombination event[118].

[*] The original *Campylobacter jejuni/ coli* MLST scheme lists the *glmM* locus as *pgm* (phosphoglucomutase), although all of the published primers amplify *glmM.* The mis-naming has been consistent and there are no compatibility issues between data obtained at various different times. The term *pgm* will continue to be used throughout this thesis.

A Sequence Type (ST) is defined by a string of seven allele numbers, corresponding to the seven MLST loci. Each ST is numbered arbitrarily and unchangeably, depending on when the ST was first identified. Sequence Types are grouped together into clonal complexes, defined by a sharing of at least four out of seven alleles with the central genotype[118].

Multi-Locus Sequence Typing does not take into account the number of nucleotide differences separating one allele from another at the same locus. A single point mutation, affecting only a single nucleotide is weighted the same as a recombination event, which can import a substantial quantity of DNA, thus causing multiple nucleotide changes in a single genetic event. This approach corrects for the occurrence of horizontal gene transfer[119].

2.2 The Gene-by-Gene Approach to Genomic Analysis

The gene-by-gene approach to the analysis of genomic data[120] is an extension of the concept of 7-locus MLST, which can be applied to any number of genes from across the genome. Like traditional MLST, the gene-by-gene approach to genomic analysis uses alleles, rather than single nucleotides, as the basic unit to measure differences between isolates. This has the advantages described above, and at the whole genome level is a computationally efficient and manageable way of analysing large amounts of data[119]. The approach can be applied to any number of loci, which could range in size from a short string of nucleotides to a whole gene. The highest level of resolution using this approach is whole-genome MLST (wgMLST), whereby all loci present in a given isolate are compared with all equivalent loci in other isolates. Whole genome MLST relies on a very high degree of genomic comparability, i.e. identical patterns of presence and

absence, between isolates. Another method of making high resolution comparisons between groups of similar isolates is core genome MLST (cgMLST) in which allelic variation across the shared core genome is analysed[120 121]. Ribosomal MLST (rMLST) is an alternative level of analysis which compares the 52 ribosomal protein loci present in *Campylobacter jejuni* and *C. coli*[122]. The number, functions, and genomic positions of the loci of interest will depend on the specific research question. The gene-by-gene approach provides a conceptual framework to support the analysis of genomic data at any level of resolution. The infrastructure which facilitates such analyses is discussed below.

2.3 Velvet Assembly

'Next Generation' sequencing technologies, including the Illumina and solexa genome sequencing platforms generate data in the form of short-read sequence files that require assembly and curation to render the data useable. Velvet is an algorithmic package designed specifically to assemble short-read sequencing data[123]. Velvet assembly, in contrast to many other genome assembly programs, is based on the construction of De Bruijn graphs made up of observed words or 'k-mers', making it ideal for use with short read data. Following construction of De Bruijn graphs potential errors are removed[123]. Currently Velvet is used by the in-house genome assembly pipeline to render the data compatible with the Bacterial Isolate Genome Sequence Database. The program's default parameters were used. A fuller description of Velvet and alternative genome assembly software can be found elsewhere[123–127].

2.4 The Bacterial Isolate Genome Sequence Database (BIGSdb)

The Bacterial Isolate Genome Sequence Database (BIGSdb)[128] was developed in Oxford University to address the need for a scalable, open-source database system to facilitate the storage, access and analysis of genomic data, in the form of assembled contigs, and linked phenotypic information. BIGSdb allows the comparison of any number of bacterial isolates at any number of loci for which nucleotide sequences are available. The user can organise relevant loci into units known as schemes, a single locus may be included in any number of schemes. The established MLST[118] and ribosomal MLST[122] schemes are included, and for a number of species including *Campylobacter jejuni* and *Campylobacter coli* there is progress towards establishing cgMLST schemes[121]. Such schemes facilitate efficient, replicable analyses but they are not restrictive and individual loci, including newly defined loci may be analysed separately or in conjunction with established schemes.

BIGSdb provides a number of analytical functions, which are discussed in depth elsewhere[128]. Throughout this thesis the BIGSdb Genome Comparator has been used to identify allelic differences between large sets of isolates. For each genome comparator run the parameters were set as follows:

- 70% minimum percentage identity for partial matches between alleles
- 50% minimum alignment length required between partially matching allele sequences
- BLASTN word size = 15
- Where available tagged allele sequences are extracted from the definitions database according to the allele designation.

Where incomplete or truncated loci were present they were ignored in pairwise comparison.

Data were output in the form of an excel spreadsheet, displaying the results for all loci, variable loci and loci which are incomplete or truncated in some isolates; a list of unique strains and a distance matrix showed the number of allelic differences between each pair of isolates. The distance matrix is also available in Nexus format, which is compatible with SplitsTree[129], enabling the relationships between isolates at the allelic level to be displayed on a Neighbour-Net diagram.

For comparisons between isolates at the nucleotide level the BIGSdb Export Sequences function was used. Where both tagged sequences and allele designations exist, tagged sequences are retrieved from the bin. Sequences with problems flagged and incomplete sequences were excluded from further analysis. Where alignments were required and the total number of isolates was fewer than 201, the MAFFT version 7 multiple alignment program[130] for amino acid or nucleotide sequences was used, through the BIGSdb user interface.

Where the total number of isolates requiring alignment exceeded 200, the 'produce alignments' option in BIGSdb was not selected, unaligned sequences were uploaded into Clustal Omega[131] for alignment.

Aligned nucleotide sequences were saved in FASTA format, and were uploaded into Mega[132,133] for analysis.

2.5 Neighbor Joining Trees

The Neighbor-Joining (NJ) algorithm[134] is a genetic distance-based method developed for constructing estimated phylogenetic trees, comprising nodes and branches of defined lengths, using evolutionary distance data and based on the principle of minimum evolution. In this thesis NJ trees are constructed using concatenated nucleotide sequences, as a means of representing inferred relationships between *Campylobacter* strains or lineages. Despite the accuracy implied by a fully resolved tree-like structure, NJ trees should be interpreted as estimated phylogenies, which do not take account of the recombination events, which are common in *Campylobacter*.

2.6 Neighbor-Net Diagrams

The Neighbor-Net method is another distance-based variation on the Neighbour Joining method, which allows for the graphical representation of the inherent ambiguities in the evolutionary histories of a recombining organism such as *Campylobacter*. Where the Neighbour Joining method produces neatly resolved phylogenetic trees with bifurcated branches, the Neighbor-Net algorithms produce a phylogenetic network. Neighbor-Net diagrams are the result of an agglomerative process, whereby triplets of linked nodes are replaced with pairs of linked nodes, reducing the distance matrix[135]. In this thesis the Neighbor-Net method has been used to construct networks based on matrices displaying the pairwise allelic differences between isolates.

2.7 ClonalFrame

ClonalFrame[136] is a package for the inference of bacterial microevolution using multi-locus sequence data. It provides an alternative approach to the problematic construction of phylogenetic histories by using Bayesian inference to estimate the clonal

relationships between isolates and simultaneously estimating the chromosomal position of recombination events that have disrupted the pattern of clonal descent. The ClonalFrame algorithm can define, and infer the relative age of bacterial lineages; infer the ancestral sequence and the recombination and mutation events that have occurred to give rise to each member of the sampled population. ClonalFrame was used in this thesis to reconstruct the relationships between *Campylobacter* isolates from retail poultry, using concatenated nucleotide sequences at the seven MLST loci.

2.8 Bayes' Theorem

Bayes' Theorem[137] is a tool for calculating the probability that a given hypothesis is true, based on the assumption that a new piece of evidence is true. Bayes Theorem can be expressed as follows:

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E|H) * P(H) + P(E|\neg H) * P(\neg H)}$$

Where H is the hypothesis being tested and E is the new piece of evidence which will support or undermine the hypothesis.

Bayesian theory underpins the model-based attribution of clinical disease cases to their putative host sources (chapter 6) by assigning a probability to the hypothesis that a given human isolate is attributable to each one of several putative host sources, based on the evidence of the observed distribution of genotypes among host sources.

2.9 STRUCTURE

STRUCTURE[138] is a model based method which uses Bayesian inference to assign individuals from unknown populations to putative source populations, defined by a set

of allele frequencies across a number of genetic loci. The model is 'trained' with genotypic data for a sample of each of the putative source populations. STRUCTURE assumes that each set of training data represents a random subset of the source population in question. Individuals of unknown origin are then probabilistically assigned to one or more source population. STRUCTURE allows for the possibility of admixture, whereby an individual may contain genetic material from more than one source population. This is appropriate in the case of recombining bacteria including *Campylobacter jejuni*.

2.10 Human Clinical *Campylobacter* Study Specific Methods

Sample Selection and Microbiology

During the 14 month period from April 2010 until July 2011 all laboratory confirmed cases of human campylobacteriosis from Oxford John Radcliffe Hospital were received by the Department of Zoology in Oxford University for MLST analysis, and patients were contacted to request participation in an extended epidemiological survey, covering recent exposures to potential sources of infection and relevant medical histories, such as prescription drug use.

Stool samples received by the John Radcliffe Hospital Clinical Microbiology Laboratory, from both community- and hospital-acquired cases of gastro-intestinal disease were routinely tested for the presence of *Campylobacter* by culturing on CCDA blood-free selective agar and incubating at 41°C for 48 hours in a microaerobic workstation (Don Whitley Scientific, Shipley, UK). During the period from April to the end of July 2010 primary cultures used for diagnostic purposes were provided for MLST analysis. Following incubation and reading, the positive primary cultures were stored at 5°C in

aerobic conditions until antimicrobial sensitivity testing had been completed. The average length of time between setting up primary cultures and providing them for MLST analysis was 4.91 days. Bacterial growth was harvested from each plate for the preparation of culture stocks and boilates (see section 2.11). Boilates were transported to Oxford University Department of Zoology on a weekly basis. Archived culture stocks were transported to Oxford University on one occasion at the end of the collection period.

As a result of ongoing monitoring the method of sample collection was changed with effect from the 29th of July 2010 for the remainder of the study. Positive *Campylobacter* cultures were swabbed after 48 hours' growth, by a Biomedical Scientist from within the hospital laboratory, during routine daily plate reading. Swabs were labelled with the laboratory number and were stored in Amies Charcoal Transport Medium at 5°C. Swabs were then transported to Oxford University on a weekly basis. Secondary cultures were grown on CCDA blood-free selective agar in the Department of Zoology research laboratory. Plates were incubated at 41°C for 48 hours in a microaerobic environment, using GENbox Microaerobic gas packs (Biomerieux). Immediately following incubation, bacterial growth was harvested for the preparation of culture stocks and boilates. A full isolate list is supplied in Appendix table 1.

PCR and Sanger Sequencing

All MLST of clinical isolates was carried out in the Department of Zoology, Oxford University. PCR and sequencing plates were made in advance, and were stored at -20°C until use. Each plate contained the PCR or sequencing mastermix for the 7 MLST loci and

for the three antigen loci, *flaA*, *flaB* and *porA*. A single plate could accommodate 7 isolates plus a negative control. Preparation of plates in advance and therefore reduction of batch size from 95 to 7 allowed for optimal use of reagents and rapid turn-around time in response to a variable case load.

PCR plates contained 23µl per well of mastermix, made up in the proportions previously described by Dingle[118]. Initially, all primer sequences were taken from earlier published protocols[139] were used for all loci.

As a result of ongoing monitoring, the original *asp* and *tkt* primers were subsequently replaced with *asp* S3 & S6 and *tkt* S5 & S4 primers[118], which were more reliable in obtaining results.

Forward and reverse sequencing plates contained 8µl per well of mastermix, made up in the proportions previously described by Dingle[118].

PCR plates were defrosted at room temperature and centrifuged at 500 rpm for < 1 minute. 15µl sterile MilliQ H_2O was dispensed into each of the 8 empty wells (A-H) in column 1. 15µl DNA was added to each well (A-G), and a further 15µl sterile MilliQ H_2O was added to the final well, as a negative control. The film seal was replaced, the plate vortexed briefly and then centrifuged at 500 rpm for < 1 minute, to ensure that the water and DNA template were both well mixed and in the bottom of the wells. 2µl of DNA + water was then transferred from column 1 into each column 2 – 11. The PCR plate was gently tapped on the bench before replacing the film lid; the plate was then vortexed briefly and centrifuged at 500 rpm for < 1 minute.

Polymerase Chain Reactions were carried out in a GeneAmp PCR 9700 System (Applied Biosystems) following the protocol described by Dingle[118].

Sequencing plates were defrosted at room temperature and centrifuged at 500 rpm for < 1 minute. 2µl re-suspended PCR product was dispensed into the corresponding well of the relevant forward and reverse sequencing plate. Plates were gently tapped on the bench top then the film seal replaced before vortexing briefly and centrifugation at 500 rpm for < 1 minute.

Sequencing reactions were carried out in a GeneAmp PCR 9700 System (Applied Biosystems, Massachusetts, USA) following the protocol reported by Dingle[118].

After the first round of sequencing, those isolates with fewer than 4 MLST alleles assigned were regrown from the glycerol stock, and the process repeated. For those isolates with 4 or more MLST alleles assigned, the PCR and sequencing process was repeated from the existing boilate. After the second round of sequencing those isolates with fewer than 3 MLST alleles assigned were eliminated; those isolates with 3 to 5 assigned alleles were regrown; for those isolates with 6 MLST loci assigned the PCR and sequencing process was repeated for the missing locus from the existing DNA boilate. Isolates yielding a novel ST were repeated from existing boilates in the first case. Confirmed new STs were then submitted to the PubMLST database for curation. Potential new STs which appeared to be mixed, or which gave very different results when repeated were regrown from stocks and the PCR and sequencing process repeated.

Whole Genome Sequencing

A subset of 729 *Campylobacter* isolates (Appendix table 2) was selected for Whole Genome Sequencing (WGS) by Genoscreen, using the HiSeq2500 Illumina sequencing platform. The sample was selected on an opportunistic basis.

Using a sterile plastic loop, CCDA plates were inoculated and spread with approximately 1-3 µl culture material from the relevant glycerol stock, without defrosting. Plates were incubated as previously described. Following 48 hours of incubation an individual colony from each plate was sub-cultured on Columbia Horseblood Agar (CBA), to ensure a pure culture for DNA extraction. After a further 48 hours growth, *Campylobacter* DNA was extracted, using a Promega DNA Wizard extraction kit, following the manufacturers' standard protocol.

The optical density (OD) and estimated DNA content of each extract was then measured using a spectrophotometer (nanodrop). Those falling within the range of 50 – 150 ng/µl of DNA were selected for WGS, those with a higher concentration of DNA were diluted to fall within this range, and those with a lower concentration were discarded.

Comparison of PCR-Derived and WGS-Derived MLST Data

MLST data derived from PCR sequencing of the individual gene fragments was compared with the MLST sequences extracted from WGS data. The proportion of isolates with identical MLST profiles, the proportion of isolates with MLST profile disagreements between the sequencing methods, and the proportion of isolates for which one method yielded better results than the other were established.

Evaluation of Ciprofloxacin Resistance Prevalence

For the 729 isolates with associated whole genome sequences, the *gyrA* locus was extracted and examined for the presence of the Thr_{86} mutation, known to be the primary mechanism to confer ciprofloxacin resistance in *Campylobacter* spp[140]. Chi-squared tests were carried out to test for an association between CC and ciprofloxacin resistance phenotype.

Identification of Putative 'Travel-Associated' Sequence Types and Ciprofloxacin Resistance Phenotypes

Data from the extended epidemiological survey were used to identify *Campylobacter* isolates from patients who had travelled abroad during the *Campylobacter* incubation period, and therefore who may have acquired the infection outside the UK. The region of travel was cross-referenced with the extracted *gyrA* sequence and the MLST data in order to discern ciprofloxacin resistance phenotype and ST. Tests for association were carried out.

2.11 Retail Poultry Antimicrobial Resistance Study Specific Methods

Retail poultry survey isolates

Campylobacter isolates (n = 1002) were obtained from the Health Protection Agency (HPA) Centre for Infections archive, comprising isolates from three UK retail chicken *Campylobacter* surveillance studies. Random, stratified samples of 214, 535 and 253 isolates were drawn from the National Retail Poultry Survey, April – June 2001; the Coordinated Local Authority Sentinel Surveillance (CLASSP) Study (2004-05); and Wales and Northern Ireland Surveillance Study (2001-06), respectively[139,141]. In total, 214

isolates from 2001 and 788 from 2004-05 were selected. A full isolate list is supplied in Appendix table 3. The isolates represented both independent butchers and large multiple outlet retail chains. 75% of all isolates in the current study were of *C. jejuni*, and the remainder were of *C. coli*, and the sample was stratified to ensure that 50% of isolates were collected in England, and the remaining 50% were divided evenly between Northern Ireland, Scotland and Wales.

Culture

All *Campylobacter* strains had been stored in the archive at - 80°C in Microbank cryovials (Prolab PL1605/G) prior to subculturing on Columbia Blood Agar (CBA). Plates were incubated for 48 hours in a MACS-VA500 Variable Atmosphere Workstation (Don Whitley Scientific Ltd) under microaerobic conditions (5% CO_2, 5% O_2, 3% H_2 and 87% N_2) at 37 °C. All microbiology procedures were performed according to the standards of the Clinical Pathology Accreditation (UK).

Determination of antimicrobial resistance

All isolates were screened for antimicrobial susceptibility (no growth) or resistance (growth) by the breakpoint screening method[142]. Isolates were grown on Columbia Blood Agar for 24 hours prior to suspension in distilled water, with a density of bacterial cells equal to a Macfarlands 0.5 standard for inoculation of antimicrobial test plates. Individual antimicrobial substances tested were incorporated into separate Iso-Sensitest Agar, enriched with 5% horse blood, in the following concentrations: chloramphenicol 8 µg/ml; gentamicin 4 µg/ml; kanamycin 16 µg/ml; neomycin 8 µg/ml; tetracycline 8 and 128 µg/ml; nalidixic acid 16 µg/ml; ciprofloxacin 1 µg/ml; and erythromycin 4 µg/ml. The

concentration of ampicillin tested changed from 32 µg/ml to 8 µg/ml during the course of the retail poultry surveys, thus ampicillin resistance was excluded from this study. A nationally or internationally recognised standard for the testing of antimoicrobial sensitivities in *Campylobacter* does not exist. The *Campylobacter* Reference Unit therefore developed and standardised a breakpoint method. While it differs from practices in some other laboratories it provides consistency within this dataset.

DNA boilate preparation

Boilates for use as template in PCR reactions were prepared as follows. A cell suspension of each culture was made in 125 µl phosphate buffered saline or in water (Sigma Aldrich, UK) in a 0.2 ml PCR tube. Suspensions were vortexed and transferred to a heat block at 100 °C for five minutes. This killed cell suspension was clarified by centrifugation at 13, 000 rpm for 10 min and stored at - 20 ºC.

PCR, Sequencing and bioinformatics

DNA template arrays were created in 96-well Thermo-fast®, polypropylene plates (Abgene, UK) and seven-locus MLST was carried out in Oxford by standard methods using published primers[143]. Each 25 µl PCR reaction comprised molecular grade water (Sigma-Aldrich, United Kingdom), 2.5 µl 10x PCR buffer (Qiagen Ltd, Hilden, Germany.), 0.25 µM each of forward and reverse primer, 0.2 mM dNTP mix (Invitrogen Ltd.), 0.025 units/µl (0.125 µl) taq polymerase (Qiagen Ltd.) and 2 µl of template DNA. The PCR thermal cycle began with a 15 min denaturation step at 95 °C, followed by 35 cycles of 94 °C for 30 seconds, 50 °C for 30 seconds and 72 °C for 1 minute, with a final extension at 72 °C for 5 minutes. 5 µl of PCR products were visualised with ultraviolet

transIllumination following electrophoresis at 200 V (10 min) on a 1 % (w / v) agarose gel in 1x TAE buffer (1 mM EDTA, 40mM Tris-acetate). The amplification products were purified by precipitation with 20% polyethylene glycol–2.5 M NaCl and stored at -20 °C. Nucleotide sequencing PCRs were performed in both directions with the same primers (forward or reverse), diluted in water. Reactions were carried out in 10 µl volumes containing 2 µl of PEG precipitated DNA resuspended in water, 1.0 µl 5x buffer, 0.02 µl BigDye Terminator v3.1 mix (Applied Biosystems) and 0.25 µM of either the forward or the reverse primer. Cycling parameters were as follows: 30 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 2 min. Unincorporated dye terminators were removed by precipitation of the termination products with 95% ethanol, and the reaction products were separated and detected with an ABI Prism 3730 automated DNA sequencer (Applied Biosyststems, UK). Forward and reverse sequences were assembled from the resultant chromatograms using the Staden suite of computer programs[144] from the Genetics Computer Group package. The consensus sequence was queried against the PubMLST *Campylobacter* sequence definitions database [http://pubmlst.org/perl/BIGSdb/BIGSdb.pl?db=pubmlst_campylobacter_seqdef] to give an allele number. The combination of alleles for the seven housekeeping genes gave the sequence type (ST). STs are assigned into genetically related clonal complexes, based on sharing four or more alleles with the central genotype.

A database was developed in the framework provided by the existing *Campylobacter* profiles database, [http://pubmlst.org/*Campylobacter*/] which covers the species *C. jejuni* and *C. coli* and is based on mlstdbNet software[145]. The molecular data on this database includes MLST and antigen sequence alleles.

Data Analysis

A phylogeny was estimated from the retail poultry study data using ClonalFrame[132]. This model-based approach to determine bacterial microevolution distinguishes point mutations from imported chromosomal recombination events – the source of the majority of allelic polymorphisms. This allows more accurate estimation of clonal relationships. A 75% consensus cut-off was imposed, meaning that only branches identified in 75% or more of the sampled trees were used in the final consensus trees. The trees shown are consensus trees of 6 ClonalFrame runs each with a 1,000 burn in and 10,000 iterations. The strict parameters used to generate the consensus trees ensured that cluster membership was robustly supported.

Binomial exact 95% Confidence Intervals were calculated for the percentage of *C. coli* and *C. jejuni* isolates resistant to each antimicrobial in the first and second phases of the study. To test for significant secular trends χ^2 tests were carried out, to test for homogeneity of resistance to each antimicrobial. The null hypothesis was that populations (species) are homogeneous in their resistance phenotypes**.** Permutation tests were then carried out for each antimicrobial to test the null hypothesis that there is no association between lineage and antimicrobial resistance phenotype within *C. jejuni*. Association between antimicrobial resistance and lineage in the observed data was summarised by an association score. This score was calculated by adding the absolute values for each lineage of the difference between the number of resistant and the number of susceptible isolates in that lineage. Resistance patterns were then randomised across the dataset and an association score estimated for this permuted

dataset. This process was repeated 10,000 times and the observed score compared with the range of scores obtained by permutation.

2.12 *Campylobacter* isolated from Farm-Caught *Rattus norvegicus* Study Specific Methods

Faecal pellets produced by Norway Rats were collected at seven non-contiguous livestock farms in Oxfordshire and Berkshire during the period from April – August 2012. Sampling took place in the early morning to ensure the freshness of faecal deposits and to maximise the chance of being able to culture any *Campylobacter* strains present.

Faecal pellets were transported individually in peptone water at room temperature. Upon arrival in the laboratory they were enriched in Bolton's broth at 37˚C for four hours, followed by incubation in the same medium at 42˚C for a further 48 hours. Following incubation the Bolton's broth suspensions were used to inoculate mCCDA plates which were incubated in microaerobic conditions at 42˚C for 36 – 48 hours, following standard protocols. A single colony was selected from each mCCDA plate for sub-culturing onto Columbia Blood Agar with horse blood, which was incubated at 42˚C for a further 36 – 48 hours. Gram staining was carried out on microaerophillic organisms. Curved or spiral Gram negative rods were presumptively identified as members of the *Campylobacter* genus. Immuno-agglutination assays were carried out following manufacturer's instructions to identify members of the *Campylobacter jejuni* and *Campylobacter coli* group, which together cause the vast majority of human campylobacteriosis.

All presumptive *Campylobacter jejuni* or *C. coli* isolates were then identified to species level using Matrix Assisted Laser Desorption/ Ionisation (MALDI) technology (Bruker). A

total of 306 *Campylobacter jejuni* and *C. coli* isolates were identified. All isolates were preserved in glycerol and stored at -80˚C for future reference.

A list of 144 unique random numbers was generated in Microsoft Excel, and the corresponding confirmed *C. jejuni* or *C. coli* isolates selected for whole genome sequencing. A full isolate list can be found in Appendix table 4. The selected isolates were grown from glycerol stocks, by inoculation of Columbia Blood Agar, followed by 48 hours' incubation at 42˚C in microaerobic conditions. A single colony was picked from each culture plate for sub-culturing onto fresh CBA before a further 48 hours' incubation as before.

Following incubation, genomic DNA was extracted using a DNA Wizard Extraction Kit (Promega), following manufacturer's instructions. The concentration of DNA in each sample was then measured using a liquid spectrophotometer (Nanodrop), and samples were diluted with DNA rehydration fluid (Promega) to fall within the range of 75 – 150 µg/ml of DNA. Genomes were sequenced at the Wellcome Trust Sanger Institute, using Illumina technology.

Genome sequence short reads were deposited in the European Bioinformatics Institute Short Read Archive (EBI SRA), and the assembled contigs at PUBMLST *Campylobacter jejuni/ coli* Isolate Database [http://pubmlst.org/campylobacter].

Curation and analyses were carried out using PUBMLST[146]. Known alleles across the genome were identified with a pre-existing allele number, while new alleles were checked to ensure a full coding sequence and were assigned the next available allele

number. Known and new STs were dealt with in a similar way, with new combinations of alleles receiving the next available ST number.

Sequences were examined and compared using the gene-by-gene approach, examining differences at the allelic level[120,147]; and using a nucleotide-based approach to identify and quantify single nucleotide differences across the genomes of smaller subsets of isolates. Pairwise differences at the allelic level were visualised using neighbournet graphs[148], produced by inputting Nexus format distance matrices, generated in the PUBMLST genome comparator[146]; while differences at the nucleotide level were visualised using neighbour joining trees[132], generated by inputting concatenated nucleotide sequences, in FASTA format, also generated using the PUBMLST genome comparator.

Human, chicken and cattle – derived *Campylobacter* isolates for comparison with *Campylobacter* from Norway Rats by generating lists of random numbers in excel and selecting isolates with the corresponding ID number. Wild bird and bank vole *Campylobacter* isolates with whole genome sequences were few in number, therefore all those available either in the PUBMLST *Campylobacter jejuni/ coli* Isolate Database, or in a private BIGS *Campylobacter* isolate database, were used for comparative analyses.

Loci at which exclusively rat alleles were present, and genes from NCTC 11168 which were absent from Norway Rat Campylobacter isolates were identified using the output from the PUBMLST genome comparator, exported to excel. The UniProt Knowledgebase[149] was used to identify the known or putative function of genes of interest within the Norway Rat *Campylobacter* isolate collection (table 3).

Table 3. Genes identified as absent from Rat Group 1 isolates.

Gene Name	Function
Cj0732	ABC transport system ATP-binding protein
uxaA'	Altronate dehydratase activity
uxaA'	Altronate dehydratase activity
panD	aspartate 1-decarboxylase activity
mloB	ATP binding
rloH	ATPase activity
Cj0139	ATPase activity; endonuclease activity
exbB1	Biopolymer transport
exbD1	Biopolymer transport
panC	catalytic activity
panB	catalytic activity
dapA	catalytic activity
Cj0729	catalytic activity
Cj0480c	DNA binding
hsdS	DNA binding
Cj0264c	electron carrier activity; oxidoreductase activity
cfrA	Ferric enterobactin uptake receptor
Cj0486	fucose transmembrane transporter activity
Cj0265c	heme-binding; electron carrier activity
Cj0241c	Iron ion binding
cfbpB	Iron-uptake ABC transporter permease
peb3	major antigenic peptide
cdtA	pathogenesis: cytolethal distending toxin
Cj0730	Putative ABC transport system permease
Cj0731	Putative ABC transport system permease
Cj0424	putative acidic periplasmic protein
ald'	Putative aldehyde dehydrogenase C-terminus
ald'	Putative aldehyde dehydrogenase C-terminus
Cj0487	Putative aminohydrolase
Cj0570	Putative ATP/GTP binding protein
Cj1138	Putative glycosyltransferase
Cj0733	Putative HAD-superfamily hydrolase
Cj0737	Putative haemagglutination activity domain containing protein
Cj1122c	Putative integral membrane protein- Surface polysaccharides, lipopolysaccharides and antigens
Cj0177	Putative iron transport protein
Cj0628	Putative lipoprotein

Cj0818	Putative lipoprotein
Cj1677	Putative lipoprotein
Cj1558	Putative membrane protein
Cj0484	Putative MFS transport protein
Cj1721c	Putative outer membrane protein
Cj0485	Putative oxidoreductase
Cj0425	putative periplasmic protein
Cj0735	Putative periplasmic protein
Cj0967	Putative periplasmic protein
Cj1376	Putative periplasmic protein
Cj1723c	Putative periplasmic protein
Cj0727	Putative periplasmic solute-binding protein
Cj0690c	Putative restriction/ modification enzyme
Cj1365c	Putative secreted serine protease
Cj1546	Putative transcriptional regulator
Cj1556	Putative transcriptional regulator
Cj1321	Putative transferase
hsdR	Putative type I resetriction enzyme R protein: ATP binding; Type I site-specific deoxyribonuclease activity
hsdM	Putative type I restriction enzyme M protein: DNA binding; site-specific DNA-methyltransferase activity
Cj0178	receptor activity, outer membrane put
Cj0819	Small hydrophobic protein
tonB1	TonB transport protein
tonB3	TonB transport protein
Cj0008	UP (Uncharacterised Protein)
Cj0055c	UP
Cj0056c	UP
Cj0488	UP
Cj0566	UP
Cj0567	UP
Cj0568	UP
Cj0569	UP
Cj0736	UP
Cj0738	UP
Cj0739	UP
Cj0740	UP
Cj0741	UP
Cj0747	UP
Cj0748	UP

Cj0970	UP
Cj0971	UP
Cj0972	UP
Cj1145c	UP
Cj1322	UP
Cj1323	UP
Cj1555c	UP
Cj1679	UP
Cj1722c	UP

2.13 Source Attribution Study Specific Methods

Selection of Isolates

The test set was a sample of 656 human clinical *Campylobacter jejuni* isolates collected at the John Radcliffe Hospital between April 2010 and July 2011 for which whole genome sequences were available. A full list of isolates is supplied in Appendix table 2. This represents approximately 55% of the laboratory confirmed *Campylobacter jejuni* cases identified at the John Radcliffe Hospital during the period of the study, and is broadly representative of the proportions of each clonal complex present in the complete MLST dataset (figure 4). There is no evident systematic bias affecting the likelihood of obtaining whole genome data for a given isolate. The clinical dataset is discussed in more detail in chapter 1.

Figure 4. Comparison of the composition of full dataset and subset selected for WGS.

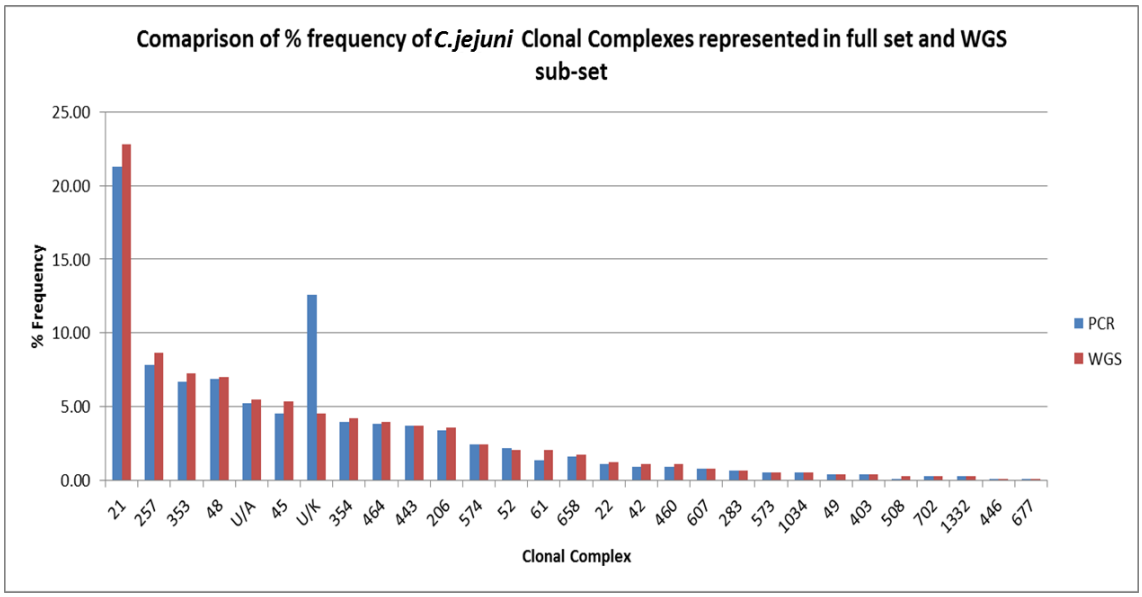


Figure 4. % Frequency of *C. jejuni* Clonal Complexes from PCR-derived MLST for full sample (n = 1187) and for sub-set selected for WGS (n = 656). U/A = Unassigned; U/K = Unknown, i.e. MLST profile incomplete.

Animal Source Reference Isolates

Sample Selection and Microbiology

A reference set of 448 *Campylobacter jejuni* isolates with associated genome sequences (Appendix table 5) was selected from a private database, hosted by BIGSdb, on the basis of species (*C. jejuni* only), host source (chicken, cattle, sheep, wild bird, Norway rat) and the availability of genome sequences. Isolates from studies which targeted particular genotypes, for example ST-21 were excluded. The reference set represents a broad cross-section of the existing library of *Campylobacter* isolated from a wide range of known hosts, which *may* be potential reservoirs for human infection, namely chicken (n=221), cattle (n=75), sheep (n=14), wild birds (n=12), and Norway rats (n=126) (figure 5). *Campylobacter* were isolated from the faeces, carcasses or caeca of each individual animal.

Chicken, cattle, sheep and wild bird specimens were originally plated onto Columbia Blood Agar, and incubated at 41˚C in micro-aerobic conditions for 48 hours. *Campylobacter* was presumptively identified based on colonial morphology and Gram-staining. Presumed *Campylobacter* isolates were sub-cultured onto mCCDA and incubated at 41˚C, as previously described. Norway rat faecal pellets were transported individually in peptone water at room temperature. Upon arrival in the laboratory they were enriched in Bolton's broth at 37˚C for four hours, followed by incubation in the same medium at 42˚C for a further 48 hours. Following incubation the Bolton's broth suspensions were used to inoculate mCCDA plates which were incubated in microaerobic conditions at 42˚C for 36 – 48 hours, following standard protocols. A single colony was selected from each mCCDA plate for sub-culturing onto Columbia Blood Agar with horse blood, which was incubated at 42˚C for a further 36 – 48 hours. Gram staining was carried out on microaerophillic organisms. Curved or spiral Gram negative rods were presumptively identified as members of the *Campylobacter* genus.

All presumptive *Campylobacter jejuni* or *C. coli* isolates were then identified to species level using Matrix Assisted Laser Desorption/ Ionisation (MALDI) technology (Bruker). Following incubation all isolates were stored at -80˚C in glycerol.

Whole Genome Sequencing

Isolates were subsequently re-grown from frozen glycerol stocks, by inoculating a labelled CCDA plate with 1µl of inoculum. Plates were again incubated micro-aerobically for 48 hours. Following incubation genomic DNA was extracted using a Promega DNA Wizard extraction kit, following the manufacturers' standard protocol. Isolates were

sent to GenoScreen (Lille, France) for whole genome sequencing, using the HiSeq 2500 sequencing platform (Illumina).

Figure 5. Distribution of Clonal Complexes and Sequence Types.

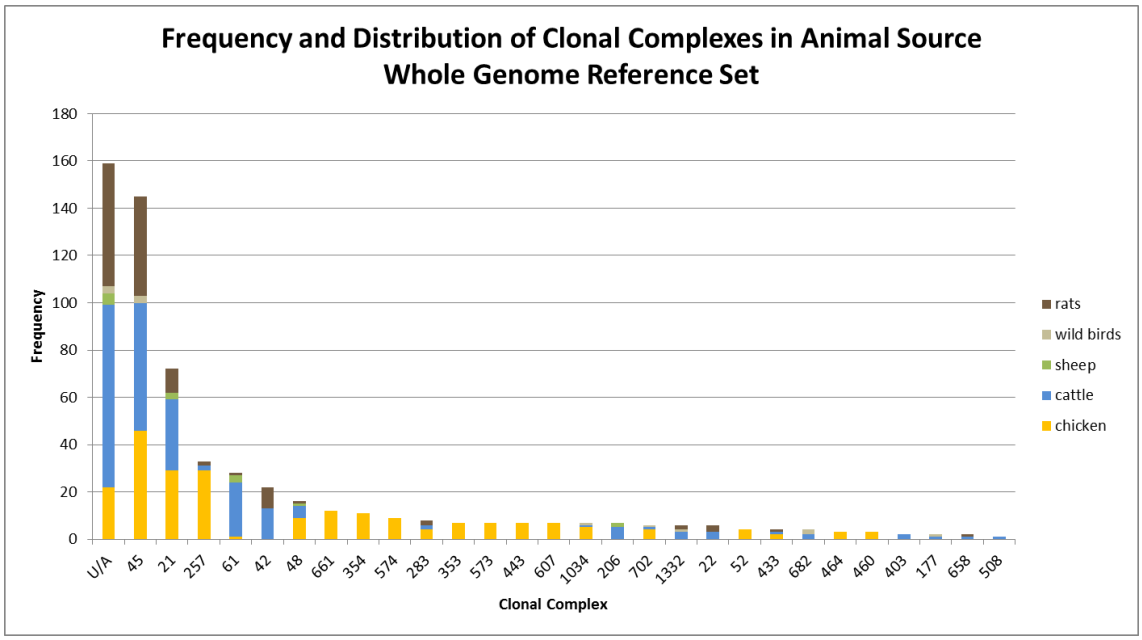
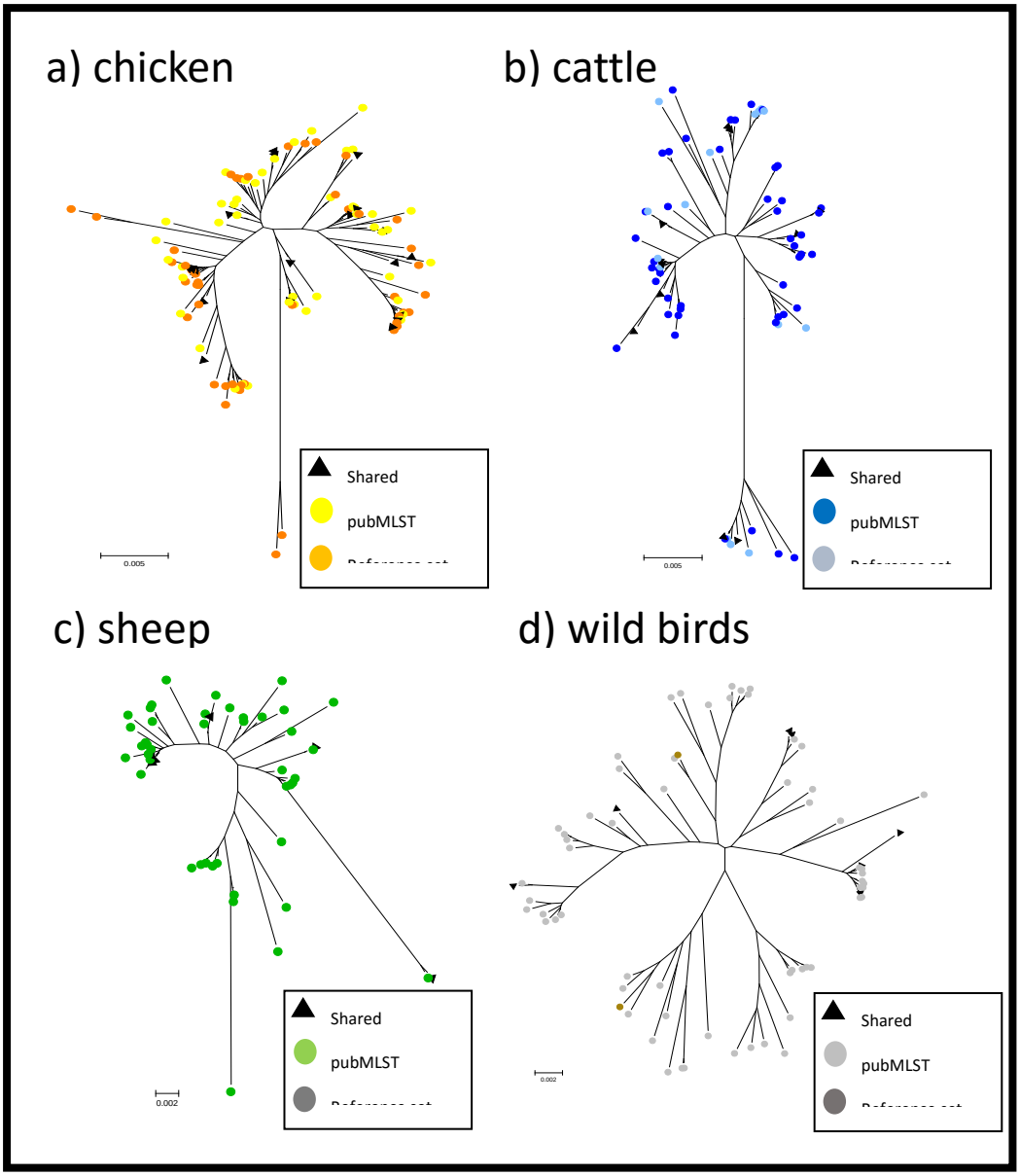


Figure 5. Clonal Complexes represented in animal reference set (n=448), coloured to represent the animal source of the isolate. U/A = unassigned.

To ensure that the reference set was broadly representative of the wider sampled populations from which the isolates were drawn, neighbour joining trees were constructed in Mega version 5.1[150], using the concatenated MLST allele sequences of the reference set for each of the agricultural and wild bird populations, together with a random selection of 100 isolates from the same source population from PubMLST, obtained by generating a list of random numbers using Microsoft Excel and selecting the corresponding isolate records.

Figure 6. Neighbor-Joining trees showing genome-sequenced reference populations in comparison with wider known populations.



The resulting NJ trees demonstrate that the reference set is sufficiently representative of the known diversity at the seven MLST loci for each population, with good coverage of the main background lineages and an absence of clusters or lineages which are not also seen in the background populations (figure 6). Norway Rats have not been extensively sampled, and the reference population for this species represents the

entirety of the *Campylobacter jejuni* diversity sampled in Norway rats to date, therefore no NJ tree was constructed.

Selection of Loci for Alternative Typing Scheme

Two separate methods were used for finding candidate loci for an alternative typing scheme, the first method based on an empirical study of diversity at each locus, and the second based on the predicted propensity for recombination at each locus.

Diversity-Based Alternative Typing Scheme [aST(1)]

The isolates constituting the reference populations were compared with each other at 1643 annotated loci[31] using BIGSdb genome comparator[146]. A total of 1623 variable loci were identified. All loci which appeared to be missing from at least 5% of isolates were discarded. Similarly, all loci which fell at the end of a contig and were therefore truncated in at least 5% of the isolates were discarded. The remaining 261 loci were retained for further investigation. A full list of variable loci is provided on Appendix p104.

Locus-by-Locus Analysis of Molecular Variance (AMOVA) was carried out using Arlequin version 3.5x[151], to quantify the relative inter- and intra-population variation at each of the variable loci. Source populations were defined as 'Chicken', 'Cattle', 'Sheep', 'Wild Bird' and 'Norway Rat'. Those loci with the highest inter-population variation scores were selected for testing as an alternative typing scheme. Unique patterns of alleles at each locus were arbitrarily numbered, in the same way as the original Multi Locus Sequence Typing scheme[17], and were given the working definition of 'alternative Sequence Types (1)' (aST (1)s). Standard AMOVA was then carried out to determine the relative within - and between – population diversity within the reference set, using aSTs.

Locus-by-Locus and Standard AMOVA were also performed using the MLST data for the reference set, to provide a comparison.

Recombination-Based Alternative Typing Scheme [aST(2)]

Each of the annotated loci from the NCTC 11168 genome[31] was ranked in order of predicted recombinogenicity, based on a comparison of matched pairs of isolates, and cross-referenced with a model-based approach using fineSTRUCTURE[152], carried out in Swansea University (Pascoe, manuscript in preparation). The top seven predicted recombinogenic loci were selected, discarding those loci that were contiguous (table 4).

Table 4. Predicted highly recombinogenic loci.

Gene Name	Molecular Function	Average Recombination Score
Cj0034c	Putative periplasmic protein	1.00
hypB	Nickel cation binding	1.00
Cj0036	Uncharacterised protein	0.94
hypD	Metal ion binding	0.93
pstA	Phosphate import	0.93
hypF	Double-stranded RNA binding; zinc ion binding	0.92
pstB	ATP binding; transmembrane activity	0.91
ilvC	Ketol-acid reductoisomerase activity	0.90
hypC	Hydrogenase isoenzymes formation	0.90
Cj0038c	Nucleotidyltransferase activity	0.90
Cj0619	Antiporter activity; drug transmembrane transporter activity	0.89

Table 4. Top eleven predicted recombinogenic loci, with molecular function (The Uniprot Consortium, 2015). Top seven non-contiguous predicted recombinogenic loci denoted by bold typeface.

Unique combinations of the seven selected loci were assigned an arbitrary identifier, in the same way as the original MLST definitions. Collectively, the new definitions were given the working term 'alternative Sequence Types (2)' or aST(2)s.

Evaluation of Shared and Unique Alleles and Types

Neighbour Joining Trees were constructed in Mega version 5.1[150] to estimate and visualise the genetic overlap between each of the source populations firstly at the seven MLST loci and then at the alternative aST(1) and aST(2) loci. The trees were compared visually to indicate the extent to which individual sources carry *Campylobacter jejuni* strains belonging to presumed separate lineages. The differentiation between the established MLST scheme and the experimental aST(1) and aST(2) schemes were compared.

Using MLST, aST(1) and aST(2) the number of alleles per locus per population, the number of population-specific and shared alleles and the number of population-specific and shared STs or aSTs was compared.

STRUCTURE Bayesian Attribution Model: Self – Attribution Test

The freely available STRUCTURE software was downloaded[138,153]. Input data were prepared in a space delimited text file, using the following columns: 'Label', which identifies the individual isolate; 'Popdata', an integer which defines the population from which the isolate was drawn, in this case 1 = Chicken; 2 = Cattle; 3 = Sheep; 4 = Wild Birds; 5 = Norway Rats; 6 = test data, 'PopFlag', which differentiates between reference and test data, with 1 denoting reference and 0 denoting test; 'Genotype', which spans seven individual columns stating allele number for each of the seven MLST or aST loci,

in alphabetical order by conventional abbreviation. The program was run using the No-Admixture Model with 10,000 burn-in iterations, followed by 100,000 experimental iterations for each test and the Structure-generated output graph was inspected to ensure convergence during the burn-in phase. Each dataset was run three times to check for consistency.

The reliability of attribution by STRUCTURE was tested by using isolates of known source as 'test' isolates against the reference set. The model was run using a stratified random selection of one third of each of the reference data host species. Lists of random numbers were generated in Microsoft Excel, and the corresponding isolate records selected, using a stratified approach to ensure that isolates from all host populations were subjected to testing.

This process was carried out using the same reference and dummy test MLST, aST(1) and aST(2) data. The probabilistic attribution was compared with the known source for each isolate and the reliability of the model, using each input type, measured by the degree of discrepancy between the actual and putative source for each isolate.

STRUCTURE: Attribution of Clinical Isolates to Putative Host Source

Following self-testing of STRUCTURE the putative sources of the human disease isolates were determined by re-running each of the models with the full animal-source reference data and human disease data entered as test data. The model was run separately for each of the typing schemes. All other parameters were unchanged.

Correlation Between Putative Source and Ciprofloxacin Resistance

The *gyrA* locus for each of the human disease isolates was extracted from the Whole Genome Sequence data. The isolates were identified as either resistant or susceptible to ciprofloxacin based on the presence or absence of the mutation, respectively. The proportion of ciprofloxacin resistant isolates from each putative source was compared and a chi-squared test for association carried out.

Chapter Three: The Molecular Epidemiology of Clinical Campylobacter in Oxfordshire: April 2010 – July 2011, Using Sanger Sequencing and Illumina Next Generation Whole Genome Sequencing.

3.1 Introduction

Human campylobacteriosis is a reportable disease in the UK, with each confirmed case being reported to the local Health Protection Unit, a sub-unit of the Health Protection Agency. National figures for campylobacteriosis incidence are compiled on an annual basis, but these figures provide little information on risk factors or epidemiological trends, and are therefore of limited public health use on their own.

Sentinel surveillance of human campylobacteriosis has been shown to be a useful method for strategy-focused surveillance, which feeds into research aimed at addressing questions pertinent to policy development[154]. The case of New Zealand provides a good example. Sentinel surveillance was carried out in the Manawatu region of the North Island of New Zealand, from 2005 – 2011[154]. The criteria for judging the success of sentinel surveillance in Manawatu included an assessment of the representativeness of the sub-population under study and an evaluation of the quality of data available. While the specifics of the demographic makeup of Manawatu differ from other regions in New Zealand it remains a valid site for strategy-focused sentinel surveillance due to the representation, to a greater or lesser extent, of the key demographic variables such as age, ethnicity and urban or rural status; and due to the availability of detailed and comprehensive epidemiological data, including

Campylobacter genomic data and epidemiological questionnaire responses[154]. Information gained from sentinel surveillance in Manawatu has been used to carry out attribution studies and identify specific campylobacteriosis risk factors and consequently to shape the development and implementation of public health policies which have successfully reduced the burden of campylobacteriosis by 54% in New Zealand[155].

In the UK the *Campylobacter* Sentinel Surveillance Scheme (CSSS) examined MLST data from reported cases of human campylobacteriosis in Southampton and Nottingham from 2000 – 2003[156]. During the three-year study period the distribution of *C. jejuni* clonal complexes in Southampton and Nottingham was almost identical, suggesting minimal regional variation within the UK despite obvious geographic and demographic differences. A subsequent 6-year longitudinal study of clinical *Campylobacter* in Oxfordshire[157] returned highly compatible results, identifying the same top four *C. jejuni* CCs as the Southampton and Nottingham data, in very similar proportions. Concentration of detailed data collection in a single location allows for targeted high-coverage extended epidemiological data collection, rather than spreading resources thinly across broader areas.

Advances in 'next generation' sequencing technology and accessibility have made vast amounts of genomic sequence data available, enabling the study of isolates, and the relationships between them, at a much higher level of resolution. The gene-by-gene approach to whole genome analysis[120] has the advantage of theoretical backwards compatibility with MLST. Various tools for the extraction of MLST data from short read WGS have been developed, to enable the rapid, objective identification of both known

and novel STs directly from Illumina output data, without going through an assembly step or manual assignment, both of which can introduce error and increase turnaround time[158,159]. A rapid flow from patient specimen to short read genomic sequence to confirmed ST, with minimal intervention by subjective observers, has great potential for streamlining both routine clinical diagnostic settings and public health surveillance settings. Both of these short read ST assignment tools have been developed and tested using genomic data from publicly available databases, including PUBMLST. For the majority of isolates data are derived only from whole genome sequencing, and a PCR and Sanger-sequencing record does not exist, therefore these tools demonstrate internal consistency, but they have not been tested for consistency with traditional Multi Locus Sequence Typing. There is a lack of formal comparisons between PCR-derived MLST data and MLST data extracted from whole genome data, with a PUBMED search on the terms 'whole genome sequence MLST comparison' OR 'WGS MLST comparison' returning no relevant records.

Here the findings from an extended epidemiological study of human campylobacteriosis in Oxfordshire are reported. Genotypic data were initially collected using conventional MLST methodology, and a large subset was later subjected to Illumina whole genome sequencing. This study considers both the epidemiology of human *Campylobacter* in Oxfordshire, and the quality and reliability of genotypic data obtained from traditional PCR and Sanger sequencing and from next generation whole genome sequencing.

3.2 Results

Distribution of Clonal Complexes and Sequence Types

Using PCR-derived MLST data for the complete HPA Oxfordshire dataset, 29 Clonal Complexes (CCs) were identified, representing 959 isolates (76.17%). A further 300 isolates (23.83%) were not assigned to any previously identified clonal complex. Of those isolates for which a CC was identified, 93.22% belonged to a *Campylobacter jejuni* CC. The remaining 6.78% of isolates assigned to a known CC belonged to *C. coli* ST-828 CC. No other *C. coli* CC was identified amongst the clinical isolates.

Twenty-seven of 29 Clonal Complexes were represented by multiple isolates, with the majority of isolates (56.16%) belonging to ST-21 CC; ST-257 CC; ST-353 CC; ST-48 CC and ST-828 CC (figure 7).

The thirteen most common STs, representing 51.25% of the sample were ST-21; ST-48; ST-5; ST-50; ST-257; ST-51; ST-464; ST-2030; ST354; ST-45; ST-574; ST-53; and ST-19 (figure 8).

Figure 7. Relative prevalence of clonal complexes.

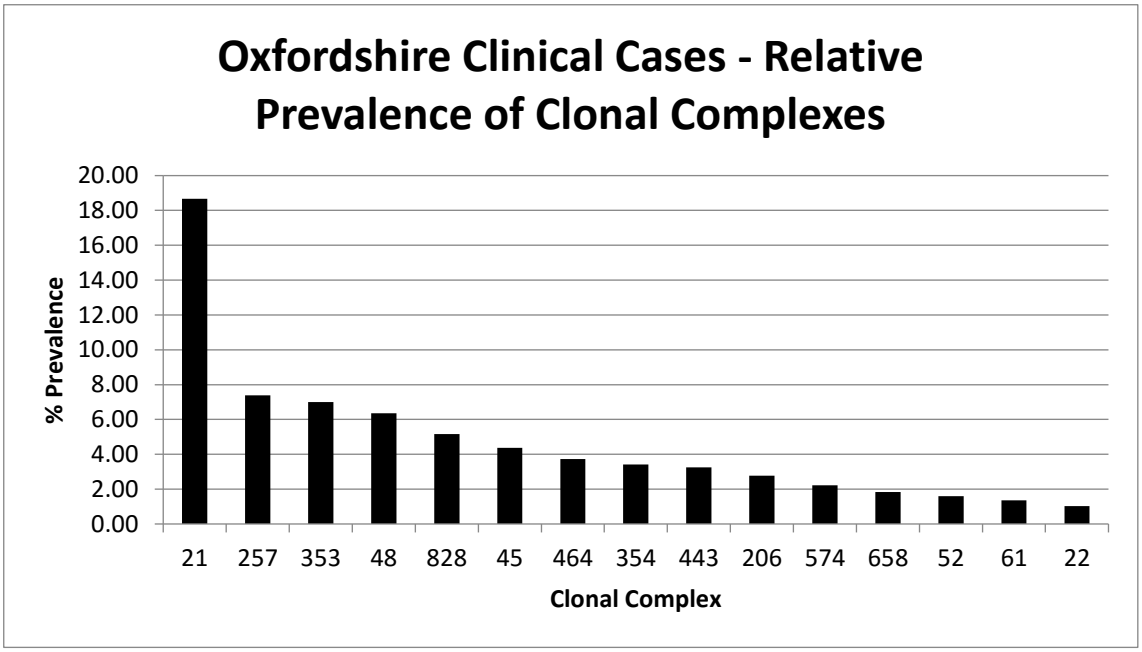


Figure 7. Bar chart displaying the percentage prevalence of clonal complexes represented in the Oxfordshire clinical dataset.

Figure 8. Relative prevalence of sequence types.

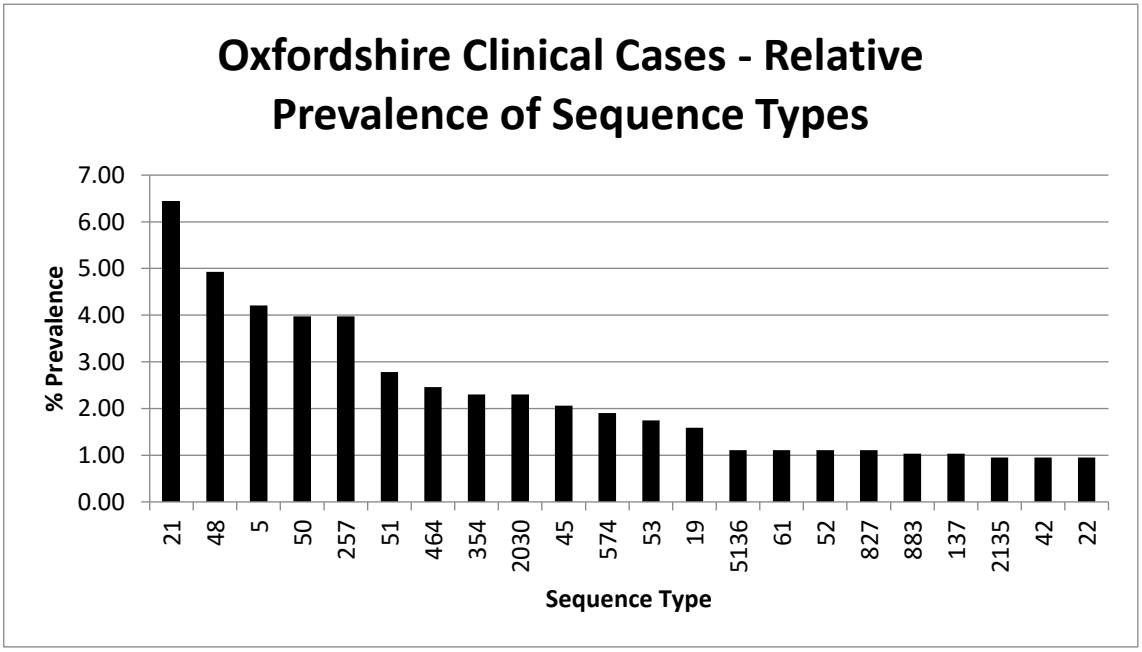


Figure 8. Bar chart displaying the percentage prevalence of sequence types represented in the Oxfordshire clinical dataset.

Comparison of Sequencing Methodologies

A total of 1259 *Campylobacter* isolates were received from the diagnostic bacteriology laboratory at the John Radcliffe Hospital between April 2010 and July 2011. Using both PCR and Sanger Sequencing methodology full MLST profiles were obtained for 1010 isolates (80.22%). Partial MLST profiles (i.e. between 1 and 6 MLST loci sequenced) were obtained for 217 isolates (17.24%). No MLST information at all was obtained for 32 isolates (2.54%). Of the 729 isolates selected for further analysis, PCR methods had yielded full MLST profiles for 628 isolates (86.15%). Partial MLST profiles had been obtained for 101 isolates (13.85%).

The 729 isolates selected for whole genome sequencing were broadly representative of the wider population from which they were drawn (figure 9). The reduced proportion of 'Unknown' (U/K) isolates in the sub-set compared with the full set of isolates is likely to be the result of failure to re-grow isolates which had initially been difficult to sequence.

Figure 9. Distribution of clonal complexes in full Oxfordshire dataset and in WGS dataset.

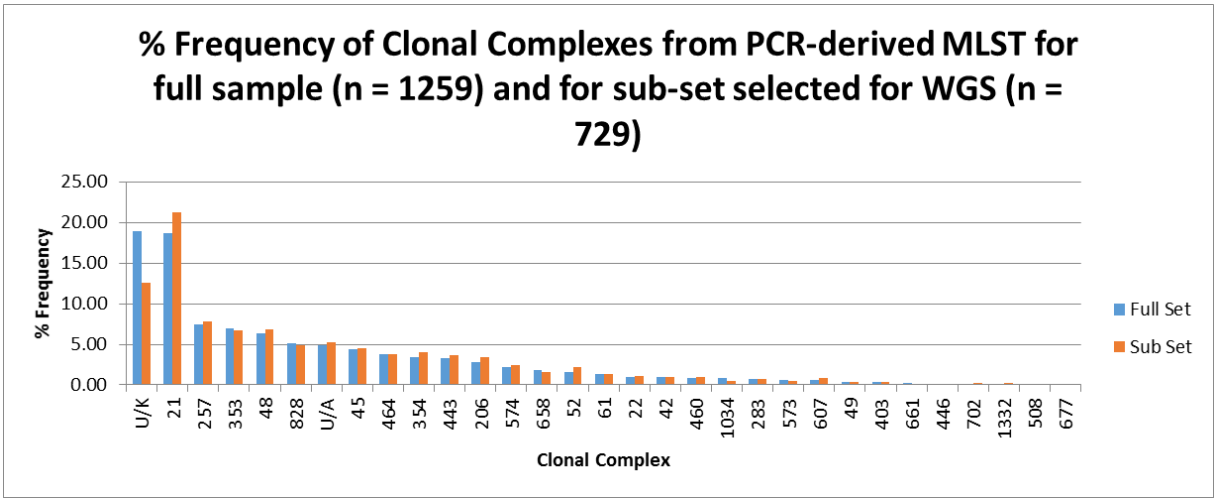


Figure 9. Bar chart displaying the percentage frequency of clonal complexes within the full sample and the sub-sample, demonstrating that the subset of isolates for which WGS exist are broadly representative of the wider sample.

Whole genome sequencing of 729 isolates yielded full MLST profiles for 696 isolates (95.47%). Partial MLST profiles were obtained for 28 isolates (3.84%) and no MLST data at all was available for 5 isolates (0.69%).

Of the 729 isolates for which both PCR and whole genome sequencing were carried out there was agreement on the ST between the two methods for 605 isolates (82.99%). For 78 isolates (10.70%) which had not been identified to ST level using the PCR method a full MLST profile was gained using WGS. In 13 cases (1.78%) a full MLST profile was obtained using PCR but not using WGS. There were disagreements between the two methods at the CC-level for 16 isolates (2.19%).

In an additional 17 cases (2.33%) there were disagreements at the ST-level between the two sequencing methods (table 5). Three of these differences represented a single nucleotide difference in only 1 out of 7 loci. A further four isolates differed by a number

of base pairs within a single locus. The remaining 10 disputed isolates differed at either two or three loci.

Table 5. Disagreements between PCR-derived and WGS-derived data. Legend printed over page.

	PCR data		WGS data	
id	ST	CC	ST	CC
5013	22	22	42	42
5172	441		825	828
5225	717	257	45	45
5231	48	48	45	45
5235	45, 109, 23	45	929	257
5242	5031	353	48	48
5247	375	48	48	48
5248	594	45	572	206
5251	5522	48	190	21
5393	233	45	50	21
5422	828	828	6601	48
5624	356	353	982	21
5660	464	464	2844	460
5677	5151		6132	828
5737	52	52	266	21
5833	52	52	6603	257
5037	42	42	7162	42
5058	286	257	257	257
5188	5237	48	5183	48
5209	5238	257	257	257
5226	5	353	400	353
5236	5628	45	45	45
5261	5241	206	273	206
5262	2926	354	354	354
5380	5135	574	2140	574
5381	5	353	7176	353
5484	5140	48	48	48
5616	1512	574	574	574
5623	2135	21	7172	21
5643	367	257	257	257
5688	4408	828	827	828
5742	3858	828	872	828
5884	273	206	5819	206

Table 5. Disagreements between PCR and WGS data. Data marked in red represent a conflict at the clonal complex level, while data marked in orange represent a conflict at the ST level only, with stated clonal complexes remaining unchanged.

Ciprofloxacin Resistance

Association between *gyrA* allele and resistance phenotype in clinical *Campylobacter*

For 716 clinical isolates both phenotypic and genomic data were available, allowing a direct comparison between the resistance phenotype and the *gyrA* allele. 135 unique *gyrA* alleles, were identified among these isolates. The alleles can be categorised into those that contain the Thr_{86} mutation and those that do not. 47.89% of the alleles contained the mutation and were identified as putative ciprofloxacin resistant alleles, with the remaining 52.11% not containing the mutation and being identified as putative ciprofloxacin susceptible alleles. A comparison of phenotypic and genomic data for each individual isolate revealed 97.76% consistency, with only 3 disagreements across the dataset, where a resistant phenotype has been recorded but the allelic data suggested ciprofloxacin sensitivity. Given the high rate of agreement, identification of the *gyrA* allele can thus be extrapolated to provide an estimate of the overall prevalence ciprofloxacin resistance across this dataset.

Prevalence of Ciprofloxacin Resistance and Association with Clonal Complex

Nucleotide sequences for the *gyrA* subunit were extracted from all 643 isolate records for which sequences were available, and the ciprofloxacin resistance phenotype inferred from the presence (resistant) or absence (susceptible) of the Thr_{86} mutation.

Overall, the inferred prevalence of ciprofloxacin resistance was 37.33%. The proportion of resistant isolates was unevenly distributed throughout the STs and CCs within the population.

A chi-squared test, using data from the 11 most common clonal complexes provided strong statistical support for an association between clonal complex and resistance phenotype ($p < 0.005$). ST-464, ST-354 and ST-574 clonal complexes displayed a significantly higher prevalence of ciprofloxacin resistance than would be expected by chance alone; while ST-45, ST-443, ST-48 and ST-257 CCs displayed a lower prevalence of resistance than would be expected due to chance (tables 6 & 7).

Table 6. Observed and expected frequencies of ciprofloxacin resistant and susceptible isolates.

CC	R Obs	R Exp	S Obs	S Exp	Total
21	68	60.06	86	93.94	154
257	10	23.40	50	36.6	60
828	21	19.11	28	29.89	49
353	18	17.94	28	28.06	46
48	6	15.99	35	25.01	41
45	3	12.48	29	19.52	32
354	26	11.31	3	17.69	29
206	10	9.75	15	15.25	25
443	3	9.36	21	14.64	24
464	22	9.36	2	14.64	24
574	8	6.24	8	9.76	16
	195		305		**500**

Table 6. Observed frequencies of ciprofloxacin resistant isolates within each clonal complex, and expected frequencies assuming a random distribution of resistance among the clonal complexes.

Table 7. Chi-squared test for association between CC and ciprofloxacin resistance phenotype.

CC	O	E	(O-E)	$((O-E)^2)/E$
21	68	60.06	7.94	1.05
257	10	23.40	-13.40	7.67
828	21	19.11	1.89	0.19
353	18	17.94	0.06	0.00
48	6	15.99	-9.99	6.24
45	3	12.48	-9.48	7.20
354	26	11.31	14.69	19.08
206	10	9.75	0.25	0.01
443	3	9.36	-6.36	4.32
464	22	9.36	12.64	17.07
574	8	6.24	1.76	0.50
			Total:	**63.33**
			d.f.	10

Table 7. Results of chi-square test for association between clonal complex and ciprofloxacin resistance phenotype for 11 most common clonal complexes within the Oxfordshire HPA dataset.

Neighbour joining trees constructed using concatenated MLST sequences (figure 10) demonstrated that despite the relationship between CC and ciprofloxacin resistance, ciprofloxacin resistance phenotypes were not clustered within clades but rather dispersed around the tree, with high and low prevalence resistance STs appearing within the same putative lineages.

Figure 10. Neighbor-joining tree displaying ciprofloxacin resistance of clinical isolates.

Figure 7. Neighbor-joining tree constructed using concatenated MLST sequences of all those STs representing at least 0.2% of the clinical dataset. Colour-coded to indicate percentage prevalence of ciprofloxacin resistance, with white indicating 0% recorded resistance and darkest red indicating 100% recorded resistance.

Association between *gyrA* alleles and Clonal Complex

A total of 135 unique *gyrA* alleles was identified from the WGS of the Oxfordshire clinical dataset. Each allele appeared between one and fifty-six times. The most frequently occurring alleles were *gyrA* 20 (8.93%), *gyrA* 1 (7.81%), *gyrA* 70 (7.81%) and *gyrA* 52 (6.86%) (figure 11). The *gyrA* 20 allele was identified as resistant, and appeared in both ST-21 CC and ST-354 CC isolates. The *gyrA* 1 and *gyrA* 70 alleles were both identified as ciprofloxacin susceptible and appeared predominantly in ST-21 CC and ST-257 CC, respectively. The *gyrA* 52 allele was also identified as susceptible and appeared in ST-21 CC, ST-48 CC and ST-206 CC.

Figure 11. Percentage frequency of *gyrA* alleles.

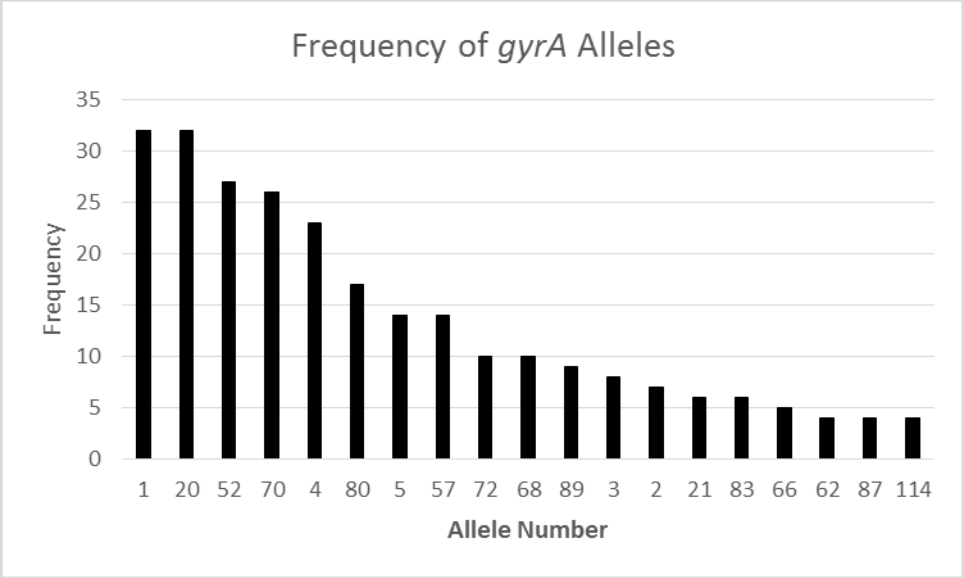


Figure 11. The most prevalent *gyrA* alleles within the WGS HPA Oxfordshire clinical dataset

For each of the clonal complexes represented here between 1 and 18 possible *gyrA* alleles were identified. In most cases there were both susceptible and resistant alleles, with the exception of ST-22, ST-42, ST-283, ST-403, ST-702 and ST-677 clonal complexes, for which there were only ciprofloxacin susceptible alleles and no resistant alleles. There were no examples of CCs with resistant alleles only.

A neighbour joining tree constructed using concatenated *gyrA* sequences shows that the Thr_{86} mutation was not confined to specific lineages, but was dispersed throughout the lineages, with genetically very similar alleles possessing either the mutation or the wildtype (figure 12).

Figure 12. Neighbor-joining displaying the distribution of *gyrA* alleles.

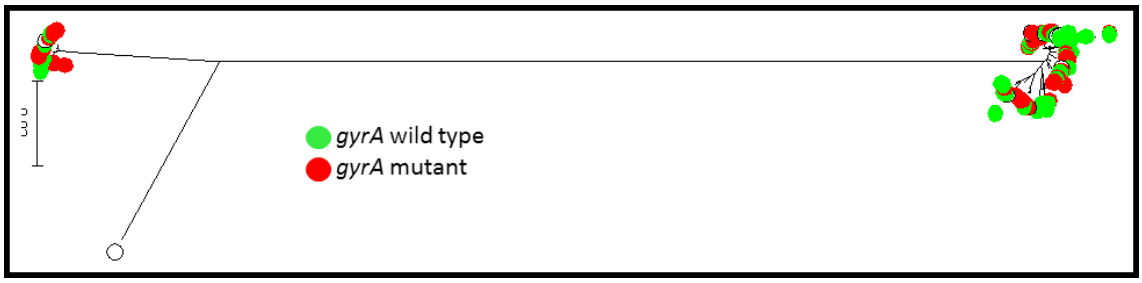


Figure 12. Neighbour Joining Tree constructed using concatenated sequences of each unique *gyrA* allele. Colour-coded green to denote ciprofloxacin sensitivity (wildtype) and red to denote ciprofloxacin resistance (mutant).

Association between *gyrA* alleles and Sequence Type

The association between ST and *gyrA* allele was even more marked. Using data from those STs with both genomic and phenotypic data, between 1 and 5 possible *gyrA* alleles were identified per ST. ST-354, ST-5136 and ST-572 displayed only resistant alleles within this subset, and displayed 100% resistance across the whole sample, while ST-53, ST-52, ST-827 and ST-273, which were 100% susceptible across the clinical dataset, displayed only susceptible alleles.

The majority of alleles appeared in one ST only, with a few exceptions. Resistant allele *gyrA* 20 appeared in ST-354, ST-21 and ST-50, and the resistant *gyrA* 57 appeared in ST-5136 and ST-464, both of which are members of the ST-464 CC. Susceptible allele *gyrA* 1 appeared in ST-21, ST-50, and ST-53, all of which are members of the ST-21 CC, susceptible *gyrA* 4 appeared in ST-50, ST-19 and ST-52, susceptible *gyrA* 70 appeared in ST-257, ST-2030 and ST-52 and susceptible *gyrA* 80 appeared in ST-443 and ST-574.

Travel – Associated STs and Ciprofloxacin Resistance Phenotypes

Of 999 isolates for which extended epidemiological data were available, a total of 120 individuals had travelled outside the UK within the 2 weeks prior to the onset of symptomatic campylobacteriosis. The most common travel destinations, by region, were Europe (n=44), Asia (n=37), and Africa (n=23) with fewer individuals having travelled to the Middle East, Central/ South America and the Caribbean, Oceania and North America (figure 13).

Figure 13. Frequency of patients having travelled outside UK during 2 weeks prior to onset of symptoms.

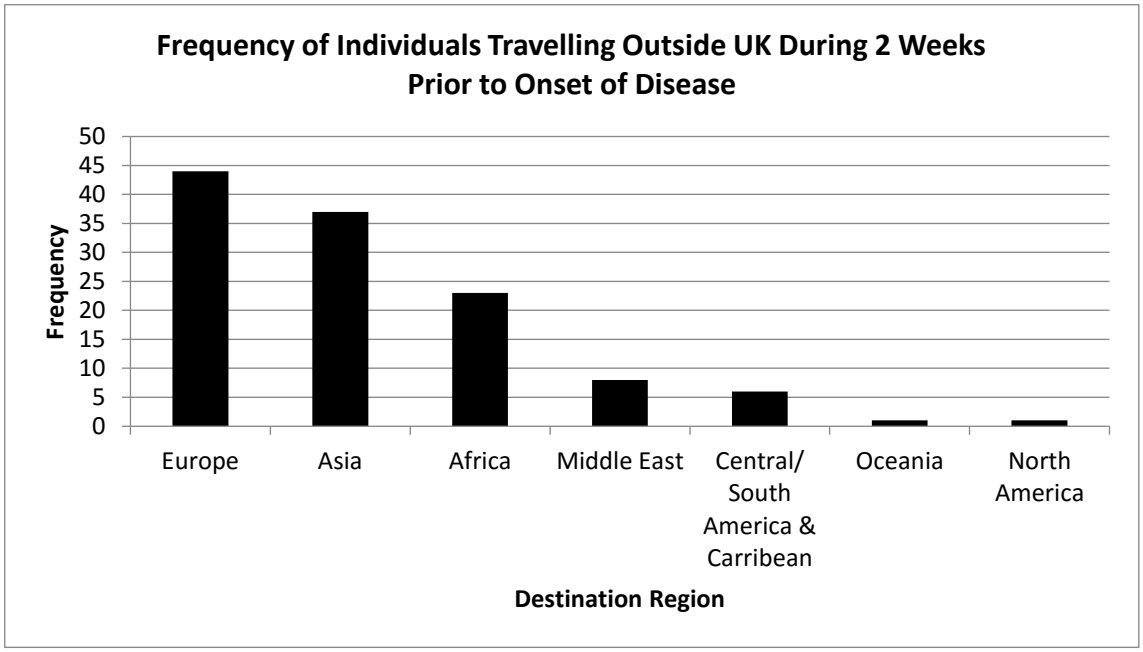


Figure 13. Bar chart showing the number of *Campylobacter* patients who had travelled outside the UK within the two weeks prior to onset of symptoms.

The most prevalent CCs amongst all returnee campylobacteriosis cases were ST-21 CC (20.35%); ST-353 CC (13.56%) and ST-828 CC (11.86%) (Figure 14). Eleven isolates

(9.32%) with full MLST profiles were unassigned to a clonal complex. These three most prevalent CCs amongst the 120 returning travellers were also within the five most prevalent clonal complexes identified in 1259 campylobacteriosis cases in Oxfordshire.

Figure 14. Frequency of clonal complexes in all patients with a recent history of travel.

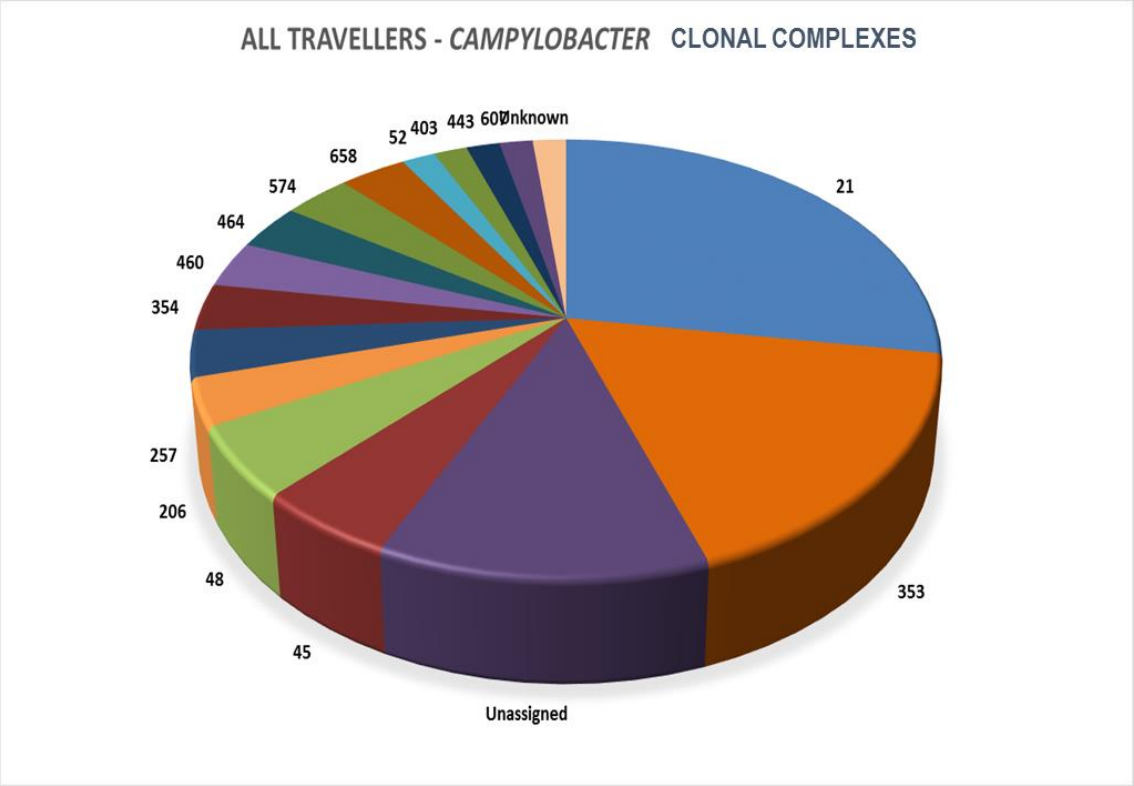


Figure 14. The relative frequency of *Campylobacter* clonal complexes amongst travellers to all regions.

Amongst returnees from Europe (figure 15) the most common CCs were ST-21 CC (25.58%) and ST-828 CC (9.30%). Three isolates (6.98%) with full MLST profiles were not assigned to any clonal complex. Amongst returnees from Asia (figure 16) the most common CCs were ST-353 CC (24.32%); ST-21 CC (13.51%) and ST-828 CC (10.81%). Six isolates (16.22%) from travellers to Asia had full MLST profiles but were not assigned to

any clonal complex. The top three most prevalent CCs amongst returnees from Africa (figure 17) were ST-828 CC (22.73%); ST-21 CC (18.18%) and ST-353 CC (9.09%).

Figure 15. Frequency of clonal complexes in patients with a recent history of travel to Europe.

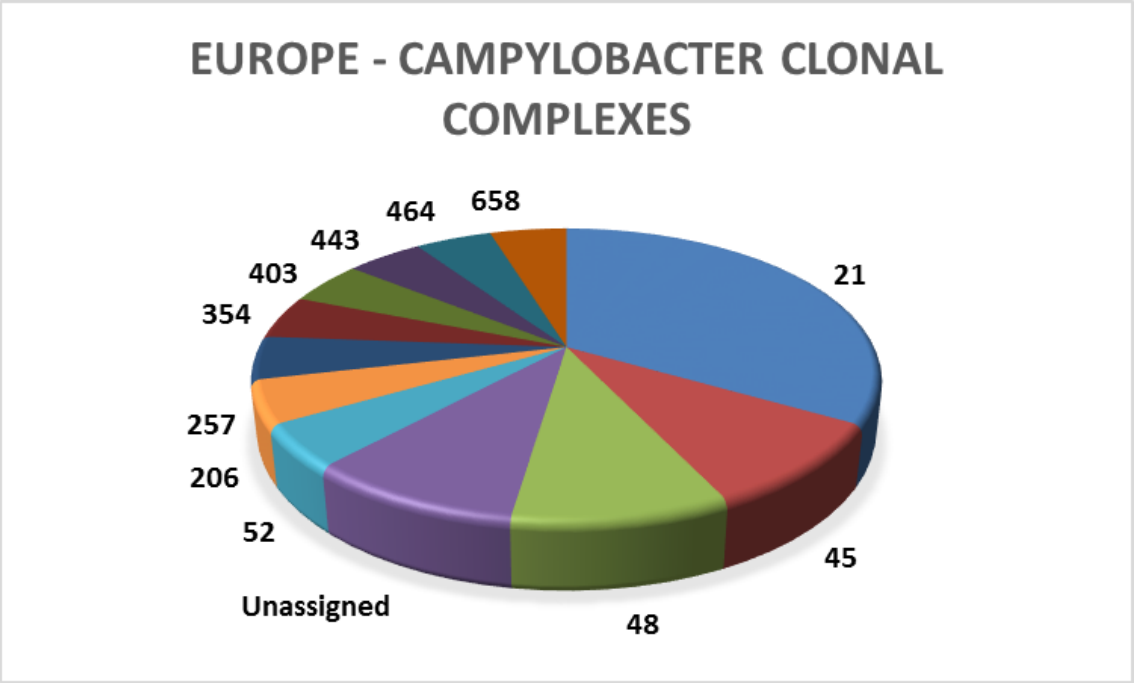


Figure 15. The relative frequency of *Campylobacter* clonal complexes amongst travellers returning from Europe.

Figure 16. Frequency of clonal complexes in patients with a recent history of travel to Asia.

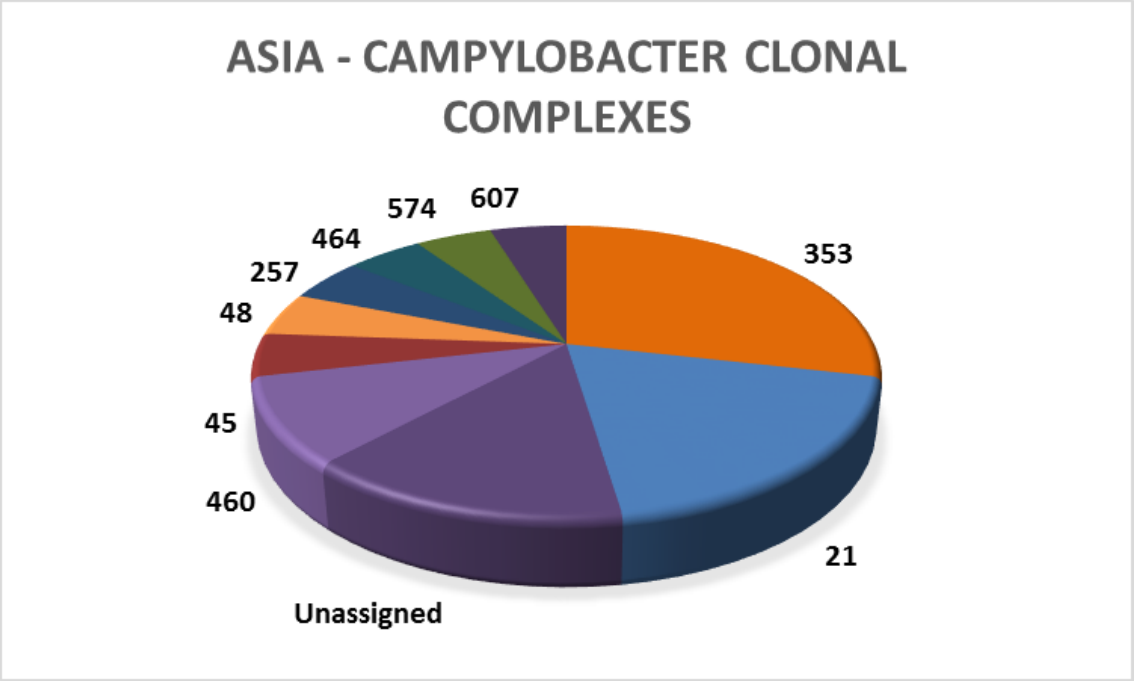


Figure 16. The relative frequency of *Campylobacter* clonal complexes amongst travellers returning from Asia.

Figure 17. Frequency of clonal complexes in patients with a recent history of travel to Africa.

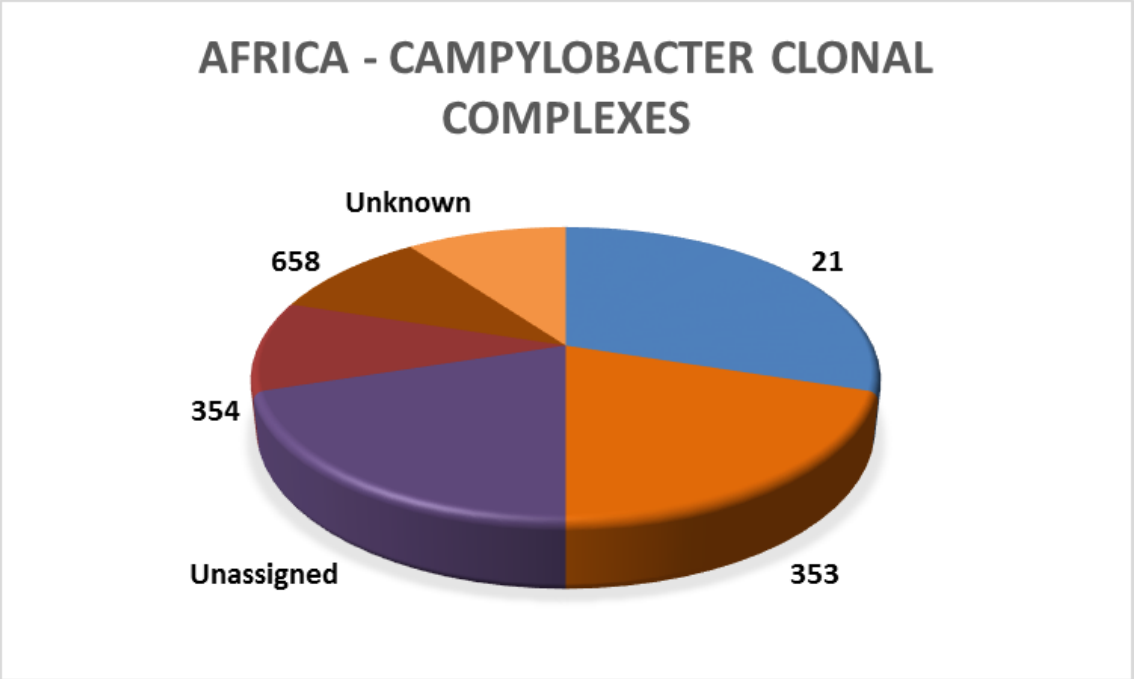


Figure 17. The relative frequency of *Campylobacter* clonal complexes amongst travellers returning from Africa.

All prevalent *Campylobacter* CCs amongst returning travellers were also prevalent amongst campylobacteriosis cases within the general population of Oxfordshire, not providing any support for an association between overseas travel and particular clonal complexes.

Whole genome sequences were available for 74 of the 120 cases of returning travellers' campylobacteriosis. In total 48 of these isolates (64.86%) had the Thr_{86} mutation in the *gyrA* allele, indicative of ciprofloxacin resistance, this is greater than would be expected by chance. A chi-squared test (tables 8 and 9) provided support for a statistically significant association between travel-associated campylobacteriosis and ciprofloxacin resistance ($p<0.005$).

Table 8. Observed and expected frequencies of ciprofloxacin resistance in travellers and non-travellers.

	R Obs	R Exp	S Obs	S Exp	Total
Non-Travellers	240	264.62	403	378.38	643
Travellers	74	49.38	46	70.62	120
	314		449		763

Table 9. Chi-squared test for an association between foreign travel and ciprofloxacin resistance.

	O	E	(O-E)	$((O-E)^2)/E$
Non-Travellers	240	264.62	-24.62	2.289911
Travellers	74	49.38	24.62	12.2701
			Total:	**14.56001**
			d.f.	1

Tables 8 and 9: results of chi-square test for association between travelling and ciprofloxacin resistance phenotype.

A chi-squared test (Tables 10 and 11) provided support for an association between the three most common clonal complexes within the returning travellers group and ciprofloxacin resistance phenotype ($p<0.005$). ST-21 CC in returning travellers was associated with sensitivity to ciprofloxacin, while ST-353 CC in returning travellers was significantly associated with resistance to ciprofloxacin, and ST-828 CC showed a significant but moderate association with ciprofloxacin resistance.

Table 10. Observed and expected frequencies of ciprofloxacin resistance among travellers for the most common clonal complexes.

CC	R Obs	R Exp	S Obs	S Exp	Total
21	6	11.24	10	4.76	16
353	11	7.73	0	3.27	11
828	9	7.03	1	2.97	10
	26		11		37

Table 11. Chi-squared test for association between CC and ciprofloxacin resistance among travellers.

CC	O	E	(O-E)	$((O-E)^2)/E$
21	6	11.24	-5.24	2.45
353	11	7.73	3.27	1.38
828	9	7.03	1.97	0.55
			Total:	**15.38**
			d.f.	2

Tables 10 and 11. Results of chi-square test for association between clonal complex and ciprofloxacin resistance phenotype.

3.3 Discussion & Conclusions

The John Radcliffe Hospital in Oxford serves a population of approximately 600,000 people; equivalent to approximately 1% of the population of the UK. During the 14 – month study period the incidence rate of reported campylobacteriosis cases in Oxfordshire was approximately equal to 180 cases per 100,000 person-years. This compares with an incidence rate in England and Wales of 113 cases per 100,000 person years in 2010. This study does not seek to provide an explanation for the excess of cases in Oxfordshire, however the fact that GP surgeries in the county were alerted to the existence of the HPA *Campylobacter* study may have led to an increased tendency to request stool specimens from patients presenting with gastrointestinal symptoms,

thereby increasing the number of diagnoses, rather than reflecting a truly elevated incidence of disease.

The distribution of CCs contributing to the overall burden of campylobacteriosis in Oxfordshire during the study period is broadly consistent with that reported by the Oxfordshire 2003 – 2009 longitudinal *Campylobacter* study[157]. In both studies ST-21 CC predominates, with ST-257; ST-828 and ST-48 CCs all appearing within the top 5 identified clonal complexes, each representing more than 5% of reported cases in each study.

Traditional PCR-derived MLST data and WGS-derived MLST data are highly consistent, with disagreements in less than 5% of cases. Disagreements at the ST level, especially where the two methods disagreed only at a single nucleotide within a single locus, can most probably be explained by misreads in either the original Sanger sequence or in the WGS. Improvements in assembly algorithms or in short read to ST software will reduce such errors going forward, but the low prevalence of such errors in this dataset (2.33%) suggests that it is not a major cause for concern.

Where there are disagreements at the CC level there is a high likelihood of mixed cultures. This is a problem characteristic of motile organisms such as *Campylobacter* which are prone to swarming on the culture plate, and a pure culture is hard to guarantee. Gene-by-gene analysis across the whole genome demonstrates that WGS has returned STs which are genuinely present in these cases, but this is not to suggest that the original PCR-derived STs were not also present. Again, the prevalence of these

disagreements is low (2.19%) and is unlikely to represent a major source of bias in the data.

These results effectively validate the use of WGS for the ascertainment of *Campylobacter jejuni* and *C. coli* STs, but also validate the historic results achieved by traditional Sanger sequencing. One criticism of the original MLST process was the way in which a single isolate was necessarily divided into seven separate portions for sequencing, before being re-assembled as a sequence type for analysis. The method developed for this study, using pre-frozen PCR and sequencing plates, accommodating the seven MLST loci and three antigen loci; and dealing with each isolate within a single row of the plate, rather than splitting them between seven separate locus-specific plates removed the need to divide isolates and provided an assurance that a string of seven allele numbers was indeed an ST relating to a single isolate.

The consistency demonstrated between PCR and WGS derived MLST supports comparisons between contemporary and historical datasets. There is an absence of direct comparisons of a single dataset sequenced by both traditional and next generation sequencing methods in the published literature. While this study uses a *Campylobacter* dataset the results can feasibly be extrapolated to other bacterial species.

Although there were 13 cases out of 729 where full MLST profiles could be obtained only by using PCR techniques and not by using whole genome sequencing overall the proportion of full MLST profiles achieved by WGS (95.47%) was greater than that achieved using PCR (86.15%). The process of PCR and Sanger sequencing is more time-

consuming, and therefore more expensive than the process of extracting DNA and whole genome sequencing. This study reports several repeats of the culture and PCR process to improve the coverage of MLST data in the earlier phase of the study, which also increases costs of both time and consumables, whereas the whole genome sequencing process was carried out only once, making it a more efficient choice. This is particularly pertinent to the clinical setting, where a patient's treatment could depend on the rapid turnaround of results; or the public health surveillance setting, where efficient processing and analysis of clinical isolates could lead to more efficient identification of outbreaks and rapid intervention.

The availability of whole genome sequences also opens up the possibility of investigating antibiotic resistance genotypes, as demonstrated in this study by an examination of ciprofloxacin resistance.

The estimated ciprofloxacin resistance prevalence of 37.33% in the Oxfordshire clinical isolates was at a level which would suggest ciprofloxacin may be of limited use for the treatment of human campylobacteriosis. The findings presented here are consistent with the published findings on the molecular epidemiology of clinical *Campylobacter* in Oxfordshire[157], using MLST data from the six years preceding the HPA study. The Oxfordshire longitudinal study found an association between ciprofloxacin sensitivity and nine clonal complexes (ST-22 CC; ST-45 CC; ST-48 CC; ST-61 CC; ST-257 CC; ST-283 CC; ST-403 CC; ST-658 CC and ST-677 CC). In the current study an association with ciprofloxacin sensitivity was found in four clonal complexes (ST-45 CC; ST-257 CC; ST-48 CC and ST-443 CC). Of these, three are consistent with the longitudinal study results. This study identified an association between ciprofloxacin resistance and three clonal

complexes (ST-464 CC; ST-354 CC and ST-574CC). Two of these clonal complexes (ST-464 CC and ST-354 CC) were also identified amongst highly ciprofloxacin resistant complexes by the Oxfordshire longitudinal study.

Previous studies using MLST to define *Campylobacter* strains isolated from various animal species in various geographical locations have generally concluded that there is a weak association, or no association between geographical providence and ST[36,160]. This study also provides no evidence to suggest a link between geography and ST. Isolates from *Campylobacter* infections which have putatively been acquired during travel outside the UK show a similar distribution of STs to those acquired within the UK, although the distribution of incubation times from infection to symptomatic campylobacteriosis is such that recent travel histories are not necessarily indicative of overseas acquisition of infection. Recent work using allelic data from recombinogenic variable loci across the genome has identified a number of putative geographically influenced loci, with the potential to develop an attribution scheme to attribute clinical cases reported in the UK to either a UK or overseas source (Pascoe, manuscript in preparation). An ability to distinguish domestically- and non-domestically acquired *Campylobacter* cases, without detailed follow-up of patients, could potentially be a useful tool for monitoring the UK *Campylobacter* burden and for targeting public health interventions within the UK and public health education for individuals travelling abroad.

From a public health perspective the international meat and livestock trade, and its potential to import *Campylobacter* to the UK is of greater importance than the movement of individuals with *Campylobacter*. Person-to-person transmission of the

infection is generally considered rare, and a returning traveller is likely to be a dead-end host, although there is evidence of limited person-to-person transmission of exotic, foreign-acquired *Campylobacter* strains within the UK[161]. The high prevalence of ciprofloxacin resistance in returning travellers' *Campylobacter* is of greater concern. The results of this study are consistent with the results of an earlier study carried out by the *Campylobacter* Sentinel Surveillance Scheme, which identified a prevalence of ciprofloxacin resistance of 55% in cases acquired abroad, compared with a resistance prevalence of 10% in cases acquired in the UK[162]. While the direct transmission of ciprofloxacin resistant *Campylobacter* strains to susceptible individuals may be a very low risk, there is a credible potential for the horizontal transfer of resistance genes from pathogenic organisms to commensal bacteria in the gastrointestinal tract[163]. The creation of endogenous reservoirs of antibiotic resistance genes could have important public health consequences and requires further investigation.

The increasing accessibility of whole genome sequencing technologies makes it possible to study aspects of the epidemiology of *Campylobacter*, and other organisms, including antibiotic resistance profiles based on genomic data. The validation of new methods of study is imperative. This study provides empirical evidence for the consistency and compatibility of WGS and PCR-derived data and also for the high level of reliability of genomic data in predicting the ciprofloxacin resistance phenotype of a *Campylobacter* isolate.

Chapter Four: Widespread acquisition of antimicrobial resistance among Campylobacter isolates from UK retail poultry and evidence for clonal expansion of resistant lineages.

4.1 Introduction

Campylobacteriosis is a major public health problem and is the most common bacterial cause of gastro-enteritis in the industrialised world. *Campylobacter* is a commensal constituent in the microflora of a wide range of animals, and has been isolated from numerous hosts including domestic and wild mammals, birds and reptiles[164,165]. In humans, however, *Campylobacter* is pathogenic, routinely causing acute diarrhoea and occasionally serious sequelae including Guillain-Barre Syndrome and reactive arthritis[166]. The majority of human campylobacteriosis is caused by *C. jejuni* and *C. coli*[162]. Most cases are self-limiting and do not require therapeutic intervention but persistent or complicated cases and those affecting immuno-compromised patients, require antimicrobial treatment. Ciprofloxacin, a second generation fluoroquinolone, is commonly prescribed for the treatment of diarrhoea, especially in returning travellers, while macrolides are recommended where treatment is required for laboratory confirmed *Campylobacter*.

Since the late 1980's there has been an increase in the incidence of resistance to antimicrobials, including fluoroquinolones and macrolides, in cases of human campylobacteriosis[167–170]. The development of resistance is often attributed to inappropriate or incomplete clinical usage of antimicrobials[171,172]. However, this explanation is insufficient in the case of *Campylobacter* because most human cases are

self-limiting and antimicrobial treatment is unusual. Furthermore person-to-person transmission is thought to be extremely rare in industrialised countries therefore there would be little opportunity for resistant lineages to proliferate[111]. With humans generally considered to be a dead-end host, there is a requirement to identify the most likely reservoirs for the acquisition of antimicrobial resistance in *Campylobacter*.

Contaminated chicken meat is among the major sources of *Campylobacter* associated with human disease. This has been demonstrated historically through risk assessment, case-control studies[173] and outbreak investigation[14], and through the 1999 'dioxin crisis' natural experiment in Belgium, where all domestically produced poultry meat was withdrawn from sale and the incidence of human campylobacteriosis was reduced by 40%[174]. More recently attribution studies, using MLST, have been used to compare genotypes of *Campylobacter* strains carried by wild and farmed host animals with those in human disease. This has shown a link between strains found on chickens, retail poultry and those causing disease in humans[35,69,74,76].

This study quantifies the occurrence of antimicrobial resistance and investigates temporal trends among *C. jejuni* and *C. coli* isolates from retail poultry. By considering this in the context of a phylogeny for *C. jejuni* and *C. coli*, this study was designed to investigate the extent to which increases in antimicrobial resistance are the result of (i) widespread acquisition of resistance among dispersed *Campylobacter* lineages or (ii) clonal expansion of resistant lineages. This provides evidence for the location and nature of increased antimicrobial resistance among clinical *Campylobacter* strains.

4.2 Results

Over the course of the study period a total of 194 STs, belonging to 27 clonal complexes (CCs), plus a further 82 STs not assigned to any recognised clonal complex were identified. Overall, the most abundant STs were ST 257 and ST 45, each representing 8.78% of the total sample, ST 827 (3.89%), ST 51 (3.19%), ST 21 (2.99%) and ST 573 (2.99%). There was no significant difference in the proportions of dominant STs between the two study periods.

Figure 18

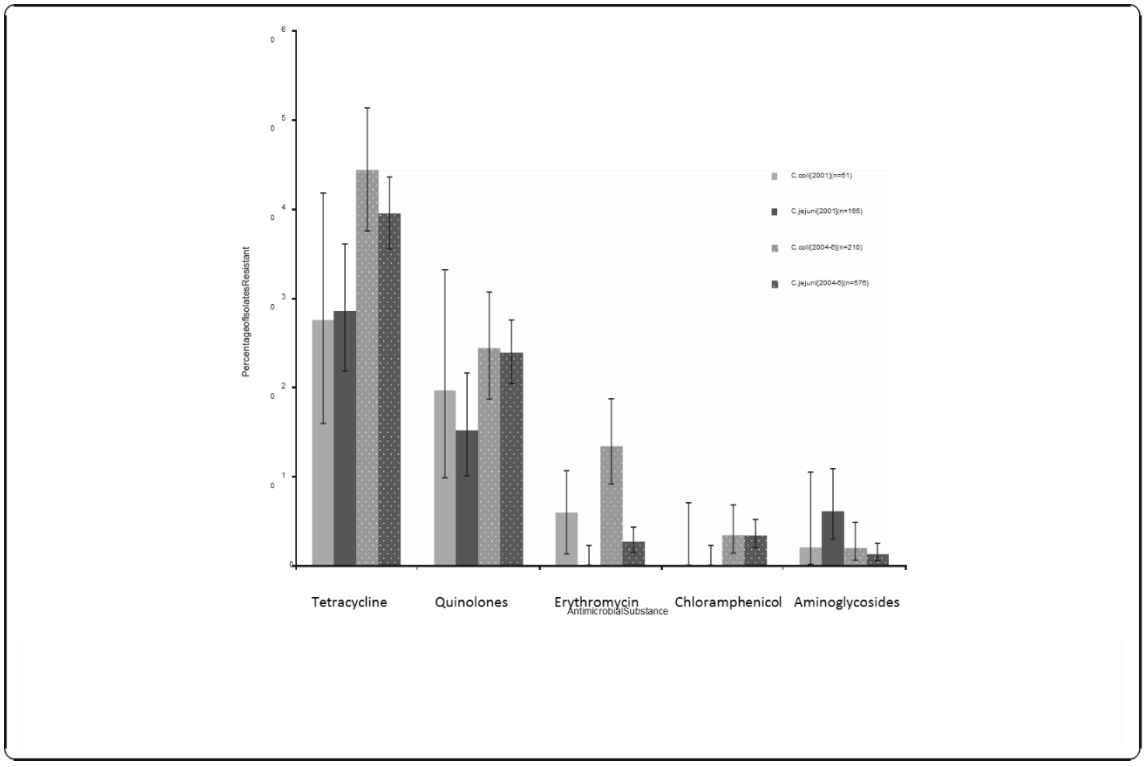


Figure 18. Proportion of resistant isolates for each antimicrobial. The percentages of resistant *C. coli* (light grey) and *C. jejuni* (dark grey) isolates are indicated for samples collected as part of the retail poultry survey in 2001 (solid colour) and 2004-5 (dotted colour). 95% confidence intervals based on a binomial distribution (resistant: susceptible) are given.

Figure 18 presents the data for the percentage of resistant isolates of both *C. jejuni* and *C. coli* between the first phase of the study in 2001 and the second phase, in 2004-5. While there appears to be an increase in resistance to all of the tested antimicrobials between the two phases it was not possible to detect a statistically significant secular trend with a sample of this size.

Overall, 38.02% (95% CI 35.01 – 41.02) *C. jejuni* and *C. coli* isolates combined were resistant to tetracycline, 22.26% (95% CI 19.68 – 24.84) were resistant to quinolones, 4.59% (95% CI 3.29 – 5.89) were resistant to erythromycin, and 2.59% (95% CI 1.29 – 3.11) resistant to chloramphenicol.

The genealogy estimated using CLONALFRAME, applied to MLST data, showed a high degree of genetic structuring among retail poultry isolates (Figure 19), with many of the lineages frequently identified from clinical samples being represented. Isolate clustering on the tree correlated with previously identified clonal complex designations (Table 12). For four (tetracycline, quinolones, chloramphenicol & erythromycin) out of the five antimicrobial substances tested in this study, resistance phenotypes were dispersed throughout clusters of related lineages (Table 12). Nearly all isolates tested were susceptible to aminoglycosides, therefore this class of antimicrobial agent was excluded from further analyses.

Table 12. Retail poultry isolates resistant to antimicrobial substances.

		Number and Percentage (%) of Tested Isolates Resistant to Antimicrobial Substance				
LINEAGE (n)	Dominant CC	Tetracycline	Quinolones[3]	Erythromycin	Chloramphenicol	Aminoglycosides
1 (209)	828	76 (36.4)	51 (24.40)	29 (13.88)	7 (3.35)	4 (1.91)
2 (187)	45	102 (54.55)	22 (11.76)	3 (1.60)	1 (0.53)	1 (0.53)
3 (131)	257	40 (30.53)	28 (21.37)	1 (0.76)	2 (1.53)	2 (1.53)
4 (44)	433	30 (68.18)	9 (20.45)	2 (4.55)	3 (6.82)	3 (6.82)
5 (21)	661	19 (90.48)	5 (23.81)	1 (4.76)	1 (4.76)	2 (9.52)
6 (16)	354	7 (43.75)	6 (37.50)	0	1 (6.25)	0
7 (7)	49	4 (57.14)	3 (42.86)	1 (14.29)	1 (14.29)	0
8 (5)	21	1 (20.00)	0	0	0	0
9 (35)	443	32 (91.43)	15 (42.86)	3 (8.57)	2 (8.57)	1 (2.86)
10 (5)	574	3 (60.00)	1 (20.00)	0	0	0
11 (8)	52	0	1 (12.50)	0	0	0
12 (3)	21	0	0	0	0	0
13 (11)	42	2 (18.18)	2 (18.18)	0	0	0
14 (12)	21	4 (33.33)	3 (25.00)	0	2 (16.67)	0
15 (21)	21	8 (38.10)	3 (14.29)	0	0	0
16 (3)	206	3 (100.00)	0	0	0	0

17 (4)	508	1 (25.00)	0	1 (25.00)	1 (25.00)	0
18 (10)	353	2 (20.00)	1 (10.00)	0	0	0
19 (10)	607	1 (10.00)	0	0	0	0
20 (7)	21	2 (28.57)	6 (85.71)	0	3 (42.86)	0
21 (4)	22	0	0	0	0	0
22 (7)	61	0	0	0	0	0
23 (10)		6 (60.00)	9 (90.00)	0	0	0
24 (3)		3 (100.00)	1 (33.33)	0	0	0
25 (2)		0	1 (50.00)	0	0	0

Table 12. Number and percentage of isolates from each lineage that were identified as resistant to each antimicrobial.

Figure 19. Estimated genealogies, showing distribution of antimicrobial resistance phenotypes throughout clades.

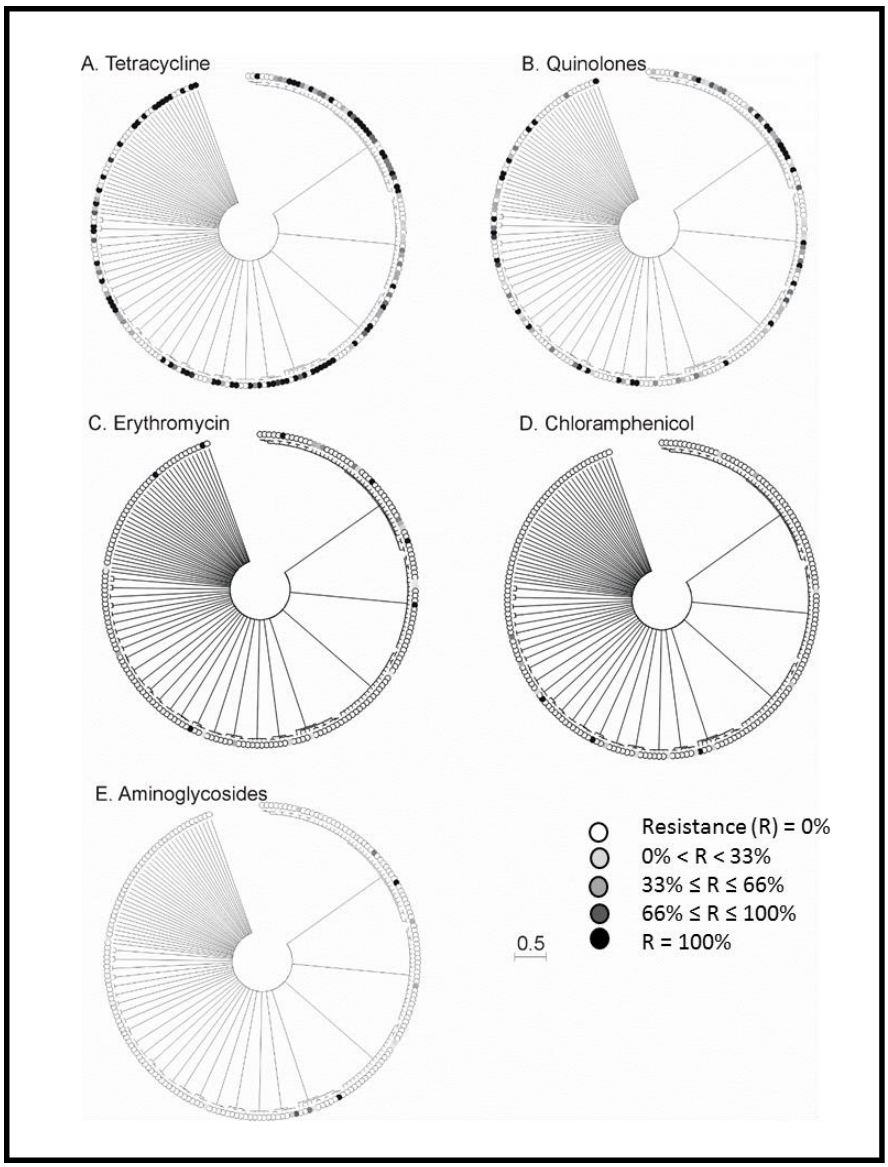


Figure 19. Clonal Frame genealogies of *Campylobacter* isolates from UK retail poultry surveys in 2001 and 2004-5. Greyscale shading indicates the percentage of isolates in each ST with resistance to each antimicrobial agent.

A total of 22 out of the 25 multi-ST lineages contained isolates resistant to one or more antimicrobial. Tetracycline resistant isolates were present in 20/25 clusters, with the percentage of resistant isolates per cluster ranging from 10% to 100%. Isolates resistant

to quinolone were present in 18/25 clusters and the proportion of resistant isolates ranged from 10% to 90%. Chloramphenicol and erythromycin resistant isolates were present in 11/25 and 8/25 clusters respectively and the proportion of resistant isolates per cluster did not exceed 42.9% in chloramphenicol or 25% in erythromycin.

For each antimicrobial, χ^2 tests for homogeneity were carried out to test the null hypothesis that populations (species) are homogeneous in their resistance phenotypes**.** In the case of tetracycline, quinolones and chloramphenicol, p values > 0.1 were obtained, providing no evidence to reject the null hypothesis. In the case of erythromycin ($p < 0.0005$) there was a significant difference in the incidence of resistance between *C. jejuni* and *C. coli,* with erythromycin resistance being associated with *C. coli* (OR 6.52).

Further, permutation tests were carried out for each antimicrobial, to test the null hypothesis that resistance was randomly distributed throughout the *C. jejuni* lineages. There was statistical support for some association between clade and probability of antimicrobial resistance for tetracycline and quinolones (naladixic acid and ciprofloxacin) in *C. jejuni*, although this is an incomplete explanation in itself. For erythromycin and chloramphenicol no statistical support for an association was identified (Figure 20).

Figure 20. Permutation test for a random association between lineage and resistance phenotype.

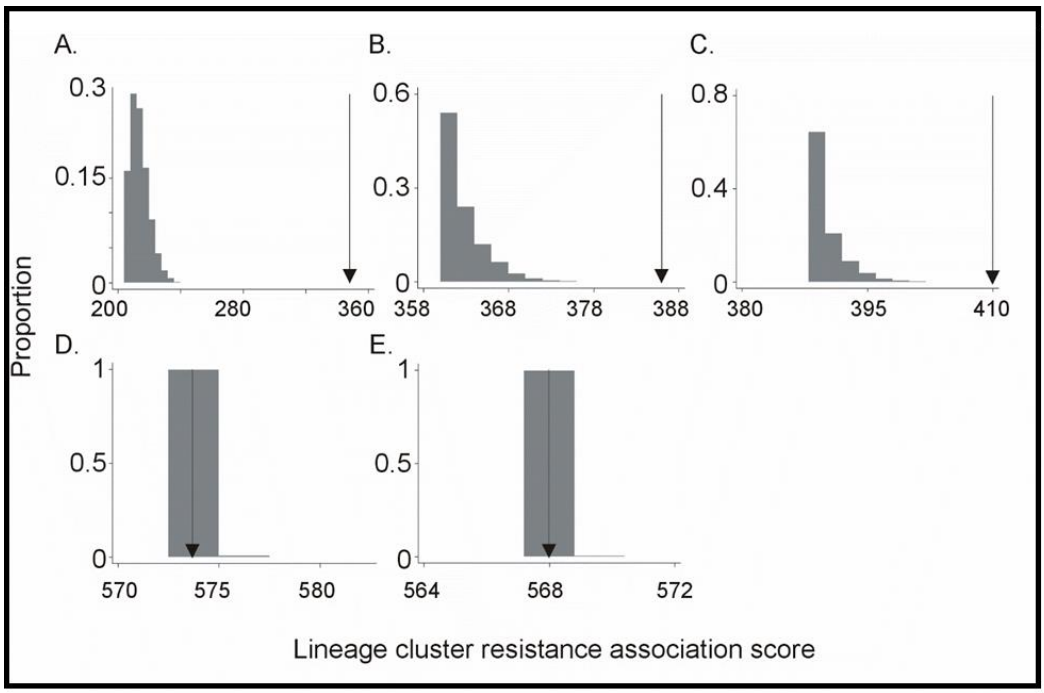


Figure 20. Permutation test results for the association of lineage with resistance phenotype for the tested antimicrobials. Comparison of a measure of association of resistant lineages with that expected by chance for (a) tetracycline, (b) naladixic acid, (c) ciprofloxacin, (d) erythromycin, (e) chloramphenicol. The arrows show the results from the data compared with frequency histograms of the scores from 10,000 permutations of the data which show the expected distribution of scores if no association exists. No comparison was made for aminoglycosides because too few isolates displayed resistance and so the test had no power.

4.3 Discussion and Conclusions

From the clinical perspective the observed prevalence of resistance of *C. jejuni* and *C. coli* isolates to antimicrobial agents is high throughout the study period. These findings are consistent with published data from clinical *Campylobacter* isolates which show high levels of antimicrobial resistance over a comparable time period[175] and with other studies that show that antimicrobial resistance patterns in clinical strains closely resemble those observed in chicken meat isolates[113].

The high incidence of resistance to tetracycline in both *C. jejuni* and *C. coli* indicates that this drug would be of little use for the treatment of campylobacteriosis. In the 2004-5 study period the combined prevalence of quinolone resistance in *C. jejuni* and *C. coli* isolates was 23.9% (188/176 samples) while the prevalence of erythromycin (currently recommended for the treatment of laboratory-confirmed campylobacteriosis) resistance in *C. coli* was 13.3% (28/ 210 samples). These levels of resistance are likely to represent an unacceptable frequency of therapeutic failure of the drugs indicated for the treatment of human campylobacteriosis.

The high levels of antimicrobial resistance cannot be accounted for exclusively by high numbers of a particular group of resistant genotypes. Rather, there is evidence for widespread acquisition of resistance among relatively distantly related lineages from retail poultry. This is consistent with a small-scale study of *C. jejuni* isolated from chicken meat in Senegal where quinolone resistant phenotypes were present in three out of four tested lineages, and also dispersed throughout singleton STs[176]. It is possible that mutations that confer antimicrobial resistance have occurred in multiple lineages. However, bacteria can acquire genetic material, including antimicrobial resistance genes, from relatively distantly related lineages through horizontal gene transfer[177,178]. Horizontal Gene Transfer (HGT) can involve recombination between lineages, or acquisition of plasmids, which has been demonstrated to be the main mechanism of tetracycline resistance in *Campylobacter*[179]. There is also evidence that plasmid acquisition may mediate resistance to chloramphenicol and aminoglycosides[180]. Resistance to macrolides is conferred by a 2 bp change in the putative erythromycin binding site[181]. Resistance to fluoroquinolones is most usually the result of a single

mutation in the *gyrA* region[181]. The widespread antimicrobial resistance in the *Campylobacter* populations, is likely to be the result of horizontal gene transfer as well as multiple independent mutation events.

When conditions are such that antimicrobial resistance confers a strong selective advantage, lineages that trace ancestry to resistant isolates will increase as a proportion of the population[182]. Under these circumstances a phylogenetic tree will show clusters of resistant lineages that have expanded clonally. Consistent with this, statistical analyses of the CLONALFRAME tree of retail poultry isolates indicated that resistant phenotypes were not randomly distributed but showed some clustering within lineages. At the highest level there was a species-specific association with erythromycin resistance correlated with *C. coli,* as in previous studies of *Campylobacter* in pigs, turkeys and chickens[52,58,183,184].

Resistance to tetracycline, quinolones and chloramphenicol showed no association with either *Campylobacter* species, but all were non-randomly distributed among *C. jejuni* lineages. This could indicate that antimicrobial resistance, having arisen in an ancestral lineage, is propagated clonally by vertical transmission. This finding is consistent with previous studies, in which associations were found between particular *C. jejuni* STs and serogroups, and a *gyrA* gene mutation which is a putative mechanism of resistance to quinolones [111].

For clonal expansion of resistant lineages to have occurred among isolates from retail poultry requires that strains had an opportunity to multiply. Mutation may occur stochastically but persistence is influenced by the fitness of organisms to compete in an

environment containing antimicrobials. Human campylobacteriosis is self-limiting and person-to-person spread is thought to be rare, therefore while the human gut may be an antimicrobial rich environment, strains that acquire resistance are not propagated and are lost from the population. Retail poultry meat itself is an unlikely environment in which antimicrobial resistant strains increase as a proportion of the population because *Campylobacter* are not thought to multiply outside of the host. Isolates from retail poultry essentially represent a subset of those found in chickens on the farm and therefore resistance among these strains is likely to reflect resistance patterns among isolates inhabiting chicken guts[36,45].

Antimicrobials have historically been used in livestock farming both for the treatment of infections and as growth promoters. The practice of administering growth promoters containing antimicrobials analogous to those used in human medicine was banned in EU countries in 2003, and in 2006 the use of all antimicrobial growth promoters was banned in the EU[185]. However, specific antimicrobials are licensed for therapeutic use in poultry. These include danofloxacin and difloxacin from the quinolone and fluoroquinolone family, several tetracyclines, several macrolides (including two varieties of erythromycin), and a number of aminoglycosides. Amphenicols are not licensed for use in poultry farming in the UK. Previous studies have speculated that where flocks testing positive for *Campylobacter* and other infections are treated *en masse* through the water supply accurate dosing is impossible and an individual bird may receive a dose too low to inhibit bacterial growth completely, thereby favouring antimicrobial resistant strains[186]. Chickens may be considered a possible reservoir in which antimicrobial resistant *Campylobacter* may emerge. This has been shown in experimental conditions

where resistance can be induced in *Campylobacter*-colonised chicken flocks, following treatment with fluoroquinolones[186].

The findings of this study suggest that antimicrobial resistance in *Campylobacter* isolated from chicken meat is widespread and may be increasing. Since retail poultry is considered to be one of the most important reservoirs of human *Campylobacter* infections, this pervasive resistance is likely to have far-reaching public health consequences. The diffuse pattern of resistance suggests that horizontal gene transfer has a role in the acquisition of resistance and evidence for the proliferation of resistant lineage clusters indicates that conditions occur that favour resistant strains – potentially on poultry farms. As pressure to meet the demand for poultry has increased there has been a requirement for greater intensification of farming practices. The consequences of this are not fully understood but the trend towards increasing levels of antimicrobial resistance among *Campylobacter* isolates from retail poultry has implications for containing outbreaks of drug resistant strains in humans.

Chapter Five: Genomic analysis of *Campylobacter* isolated from wild Norway Rats (Rattus norvegicus) on livestock farms, using a gene-by-gene approach.

5.1 Introduction

The overwhelming burden of human campylobacteriosis in the UK is attributable to commercially farmed meat[187–189]; however the source of infection in livestock is less clear. Identifying the reservoir or the vector responsible for *Campylobacter* colonisation or infection in chicken and other livestock species would facilitate the targeting of interventions to reduce the overall prevalence of *Campylobacter* in farmed animals, and thus to reduce the human campylobacteriosis burden. This approach has been successful in reducing the incidence of *Salmonella* in broiler chicken flocks, but the intervention methods used against *Salmonella* have had no impact on *Campylobacter* prevalence[190]. Logic might suggest that the introduction of *Campylobacter* to a closed livestock population may rely on vectors or fomites entering the population. Studies, primarily based on Pulsed-Field Gel Electrophoresis methodology which is described elsewhere[22], have provided a limited amount of evidence to support the roles of environmental water, farm workers, flies and other wildlife in the transmission of *Campylobacter* to farmed animals.

Laboratory analysis of drinking water in troughs and running and standing environmental waters in a dairy farm setting in the UK identified *Campylobacter spp* in 40.50% of all water samples, with *Campylobacter coli* being most prevalent overall[191]. Pulsed-Field Gel Electrophoresis Restriction Fragment Length Polymorphism (PFGE-

RFLP) analysis of *Campylobacter jejuni* isolates identified 17 different restriction patterns, of which three were shared between more than one source[191]. This study provides some evidence to support the possible role of water in circulating *Campylobacter* around the farm environment, essentially expanding the area over which faecal contamination occurs, but it does not suggest that environmental or trough waters are a source of *Campylobacter* in an un-colonised herd.

The process of partial depopulation or thinning has been reported as a risk factor for *Campylobacter* colonisation of broiler chicken flocks[192]. A study using PFGE to compare *Campylobacter* strains isolated from farm personnel, equipment, vehicles and the farm environment and from newly positive flocks within the week following thinning found a correlation[192]. There was also evidence to suggest the transmission of *Campylobacter* types between farms operated by the same company and using the same staff and equipment for the thinning process[193].

A study of broiler chicken sheds in Denmark described an abundance of flies (*Muscidae* and *Calliphoridae*), primarily drawn through the ventilation ducts. Nested PCR, carried out on a sample of 47 flies caught inside the chicken shed identified a *Campylobacter jejuni* prevalence of 56.4%. PFGE analysis of a sub sample of the captured flies was unable to distinguish the *Campylobacter* types carried by the flies from the types isolated from chickens within the same shed[88].

The reproducibility and specificity of PFGE has been questioned, and discussions can be found elsewhere[194,195].

Subsequent work using more sophisticated molecular techniques, including MLST, have demonstrated that while *Campylobacter jejuni* can be isolated from multiple non-livestock sources around the farm environment the prevalent sequence types differ between sources. A large survey of *Campylobacter* shedding in wild starlings (*Sturnus vulgaris*) caught alongside animal pens on a mixed livestock farm identified a *Campylobacter jejuni* shedding rate of 30.6%, but the most prevalent sequence types, belonging to the ST-177 and ST-682 clonal complexes, were distinct from those isolated from livestock[196].

MLST analysis of 327 *C. jejuni* isolates from cattle, wild birds, wild rabbits and environmental waters collected across a dairy farm and the surrounding environment described a mixed picture of some overlap between CCs isolated from different sources, as well as significant separation of other STs. A number of novel STs, not belonging to any known clonal complex, and isolated from only a single host source suggests possible host-associated types in less commonly sampled, i.e. non-livestock, populations[197].

Norway Rats (*Rattus norvegicus*) are commonly found in the close proximity of farmed livestock species in the UK and around the world[198]. Rats are a known vector of a number of pathogens of human public health importance, including hantavirus, *Leptospira* and *Salmonella*[199] but their propensity to act as a reservoir or vector in the livestock – human *Campylobacter* transmission cycle is not known. Biosecurity measures to exclude rats from animal houses are costly and their role in *Campylobacter* control has not been fully tested[200,201].

This study examines the genomic diversity of 144 whole genome sequenced *Campylobacter jejuni* and *C. coli* isolates from Norway rats, using the gene-by-gene approach, to examine the hypothesis that wild farm-dwelling Norway Rats play an important role in the transmission of *Campylobacter jejuni* and *C. coli* on-farm, contributing to the human campylobacteriosis burden.

5.2 Results

Faecal samples from 537 rats collected at 7 non-contiguous farms across Berkshire and Oxfordshire yielded 350 *Campylobacter* isolates. Speciation by Matrix-assisted laser desorption/ ionisation (MALDI) identified 306 as *Campylobacter jejuni* or *C. coli*. Whole Genome Sequencing of 144 randomly selected *C. jejuni/ C. coli* isolates was carried out. Initial analysis of the seven MLST loci demonstrated a diverse population of *Campylobacter jejuni* and *C. coli* isolated from Norway Rats, with a standard diversity index of 0.96 (0.95 – 0.97) at the allelic level, using 7-locus MLST. This is comparable with the diversity seen in equivalent-sized random samples of mixed *C. jejuni* and *C. coli* in chickens (0.99 ± 0.00), cattle (0.97 – 0.98) and humans (0.97 – 0.98). The rat *Campylobacter* sample, however, taps a source of diversity not previously identified. 59.59% of isolates sequenced were assigned to twelve known clonal complexes (CCs), and one unassigned CC (ST-586), spanning a range of Sequence Types (STs) known to occur in livestock, wild birds and human clinical cases, as well as a number of new but closely related STs. The remaining 40.41% of isolates were not assigned to any known CC, and were considerably different from known *Campylobacter jejuni* and *C. coli* strains.

Unassigned newly identified *Campylobacter jejuni* STs fell into three groups (figure 21) 'Rat Group 1', a distinct clade of *Campylobacter jejuni*, exclusively identified in Norway

Rats to date; 'Rat Group 2', a broad group of exclusively Rat *Campylobacter jejuni* types with a well-supported relationship with ST-45 CC; 'Rat Group 3', representing one or more clusters of *C. jejuni* STs exclusive to Norway Rats.

Figure 21. Neighbor-net diagram showing *Campylobacter* isolated from Norway rats in relation to known *Campylobacter* strains from livestock and human sources.

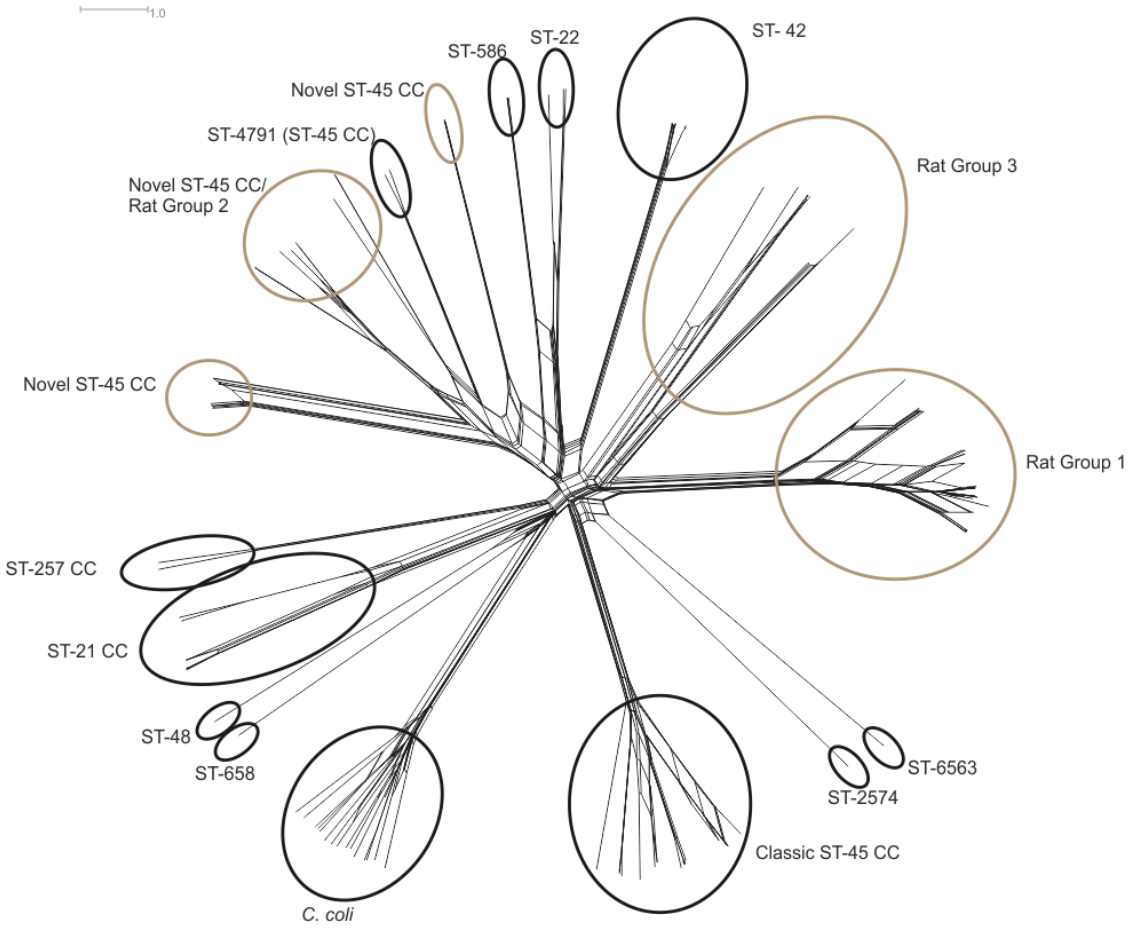


Figure 21. Whole Genome MLST Neighbor Net diagram displaying the diversity of *Campylobacter* isolated from Norway rats, with genomic data from human clinical and livestock *Campylobacter* isolates used for comparison.

Rat Group 1 (RG1) consists of 32 isolates (21.77%), representing all 7 farms, belonging to 6 complete STs; ST-5129, ST-6561, ST-6562, ST-6564, ST-7278, ST-7279 and one or

more incomplete ST. With the exception of ST-7279, which differs from the others at the *pgm* and *uncA* loci, all other STs differ only at the *pgm* locus, using MLST.

Gene-by-gene analysis of RG1 isolates using the loci identified in NCTC11168[31] identifies 933 variable loci across the genome, separating the clade into seven distinct clusters, and two individual outliers (figure 22). There was a strong correlation between farm site and WGMLST type, although there was no evidence for a correlation between livestock species on farm and *Campylobacter* genotype in rats from the corresponding farm site.

Figure 22. Neighbor-net diagram of Rat Group 1.

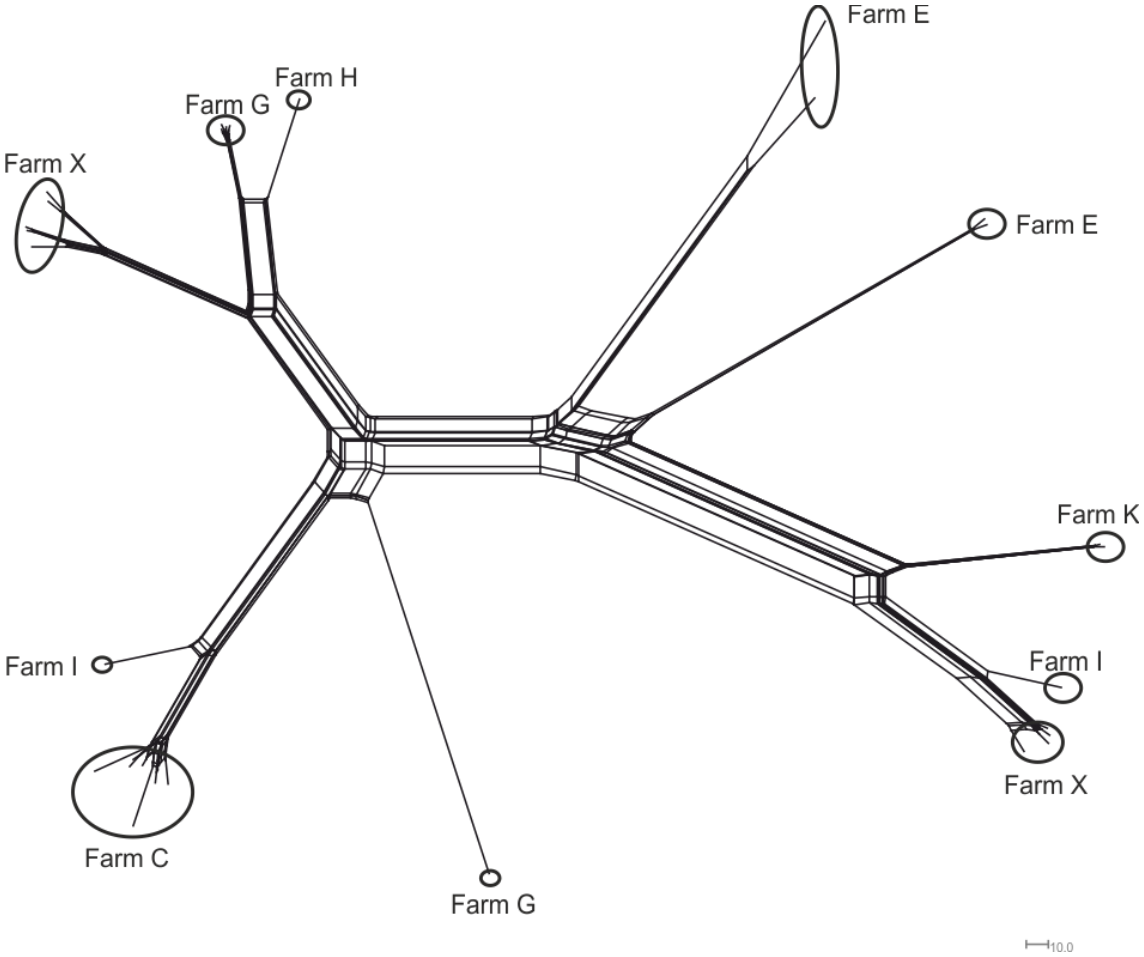


Figure 22. WGMLST Neighbor Net diagram showing Norway Rat isolates from RG1, and their provenance.

All isolates within RG1 were identical at 580 loci, with exclusively 'rat' alleles, i.e. alleles not previously identified in *Campylobacter* isolated from any other host, at 102 of the identical loci. A further 69 variable loci within RG1 demonstrated exclusively 'rat' alleles, with between 2 and 8 alleles identified at each locus. Analysis of differences between 'rat' alleles and the alleles present in NCTC11168 at these 171 loci demonstrates consistent, small to moderate differences between the rats and the reference strain

(figure 23). Rat Group 2 and 3 share a smaller proportion of loci with exclusively rat alleles, with 44 and 12 loci respectively. Within the sample of rat isolates belonging to known *C. jejuni* STs 70.30% of alleles at these 12 loci have been identified only in rats. ST-21 isolates from rats have fewer 'rat' alleles at these 12 loci compared with all other STs. There are 2 loci at which all rat *C. jejuni* isolates display exclusively new alleles; Cj0987c and Cj0988c which are a putative MFS transport protein and a hypothetical protein, respectively.

Figure 23. Neighbor-joining tree showing the relationship between Rat Group 1, NCTC 11168 and ST 3704.

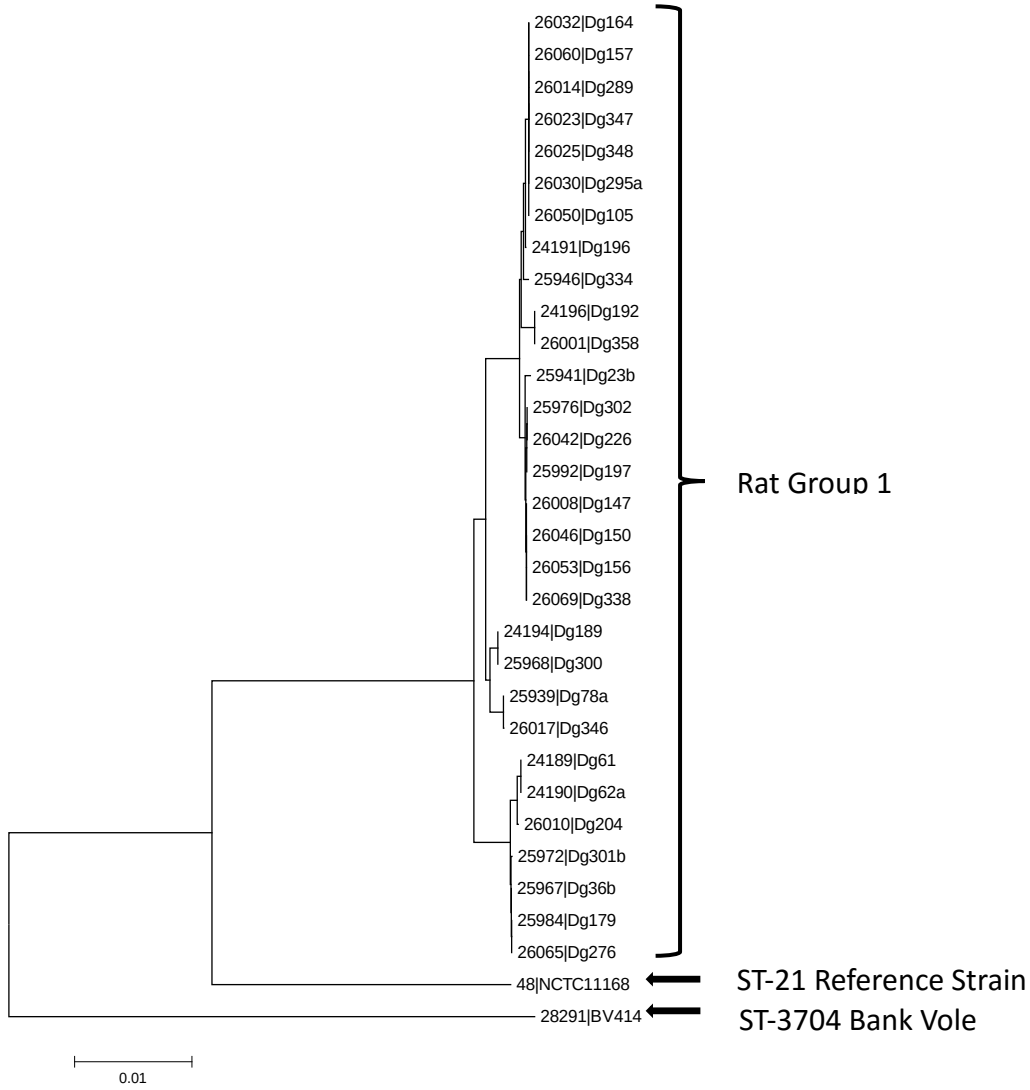


Figure 23. Neighbor-net diagram displaying the inferred relationships between Rat Group 1 isolates, NCTC 11168 and ST-3704, isolated from a bank vole at 171 loci identified as having exclusively rat alleles.

84 genes present in NCTC 11168 were routinely missing from all isolates within RG1. Absent genes can be categorised according to known or putative function (table 3). Six

absent genes (*cfrA, Cj0265c, Cj0241c, cfbpB, Cj0737* and Cj0177) are related to iron uptake and transport. The *cdtA and cdtC* genes, which encode the cytolethal distending toxin, which is involved in pathogenesis, were absent in all RG1 isolates.

Of these 33 genes were also missing from all RG2 and RG3 isolates, i.e. all rat-specific *C. jejuni* clades within this sample, although none of the missing genes from these groups is involved in either iron uptake and usage or pathogenesis. There was considerable variation in the pattern of presence and absence of these genes amongst the isolates of known ST. ST-22 and ST-42 rat isolates displayed patterns of presence and absence similar to those in the 3 *C. jejuni* rat clades, with 47 and 44 routinely missing genes, respectively. Analysis of ST-22 and ST-42 isolates from livestock and human sources identified 43 and 32 consistently missing genes, respectively. The *cfrA* and *tonB3* genes, absent in the rat ST-22 and ST-42 isolates were present in ST-22 and ST-42 isolates from livestock and humans.

ST-45 CC accounted for 28.38% of the Norway Rat *Campylobacter* sample. ST-45 CC isolates from rats fall into two distinct groups, those belonging to 'classic' STs found in livestock and human disease, and those belonging to 'novel' exclusively 'rat' STs, not previously recorded in PubMLST in any other animal, human or environmental source. Gene-by-gene comparison of the genomes of classic and novel ST-45 CC isolates identified in Norway Rats reveals 1431 variable loci. The alleles at 976 (69.86%) of those variable loci divided neatly into two pools, representing the classic and novel groups, with no over-lap.

At the whole genome level, using the gene-by-gene approach RG2 clustered with the novel ST-45 CC isolates, despite not sharing enough MLST alleles with the ST-45 central genotype, to be included within the ST-45 complex. Gene-by-gene comparison of RG2 genomes with novel ST-45 CC genomes identified 1411 variable loci, of which only 14.88% divided into discrete groups, corresponding with the RG2 and novel groups, at the remaining 85.12% of loci the alleles formed a single pool, shared between both RG2 and the novel ST-45 CC group.

The majority (61.90%) of rat *Campylobacter* isolates within this sample belong to a clonal complex also known to be present in populations of livestock. At the ST level 45.58% of rat *Campylobacter* isolates were indistinguishable from livestock-associated *Campylobacter jejuni* or *C. coli* types. Whole genome level analysis of ST-42, which is largely associated with cattle, provided higher resolution, demonstrating that rat isolates belonging to ST-42 formed a cluster distinct from members of the same ST isolated from either livestock or humans (figure 24). ST-21 isolates from Norway Rats behave differently from ST-42 isolates when compared with livestock and human ST-21 isolates using the gene-by-gene approach to construct a neighbour-net graph. Although the five rat ST-21 isolates appeared within the same clade as each other, they were interspersed with ST-21 isolates of human and chicken origin.

Figure 24. ST-42 isolated from Norway rats and from human sources.

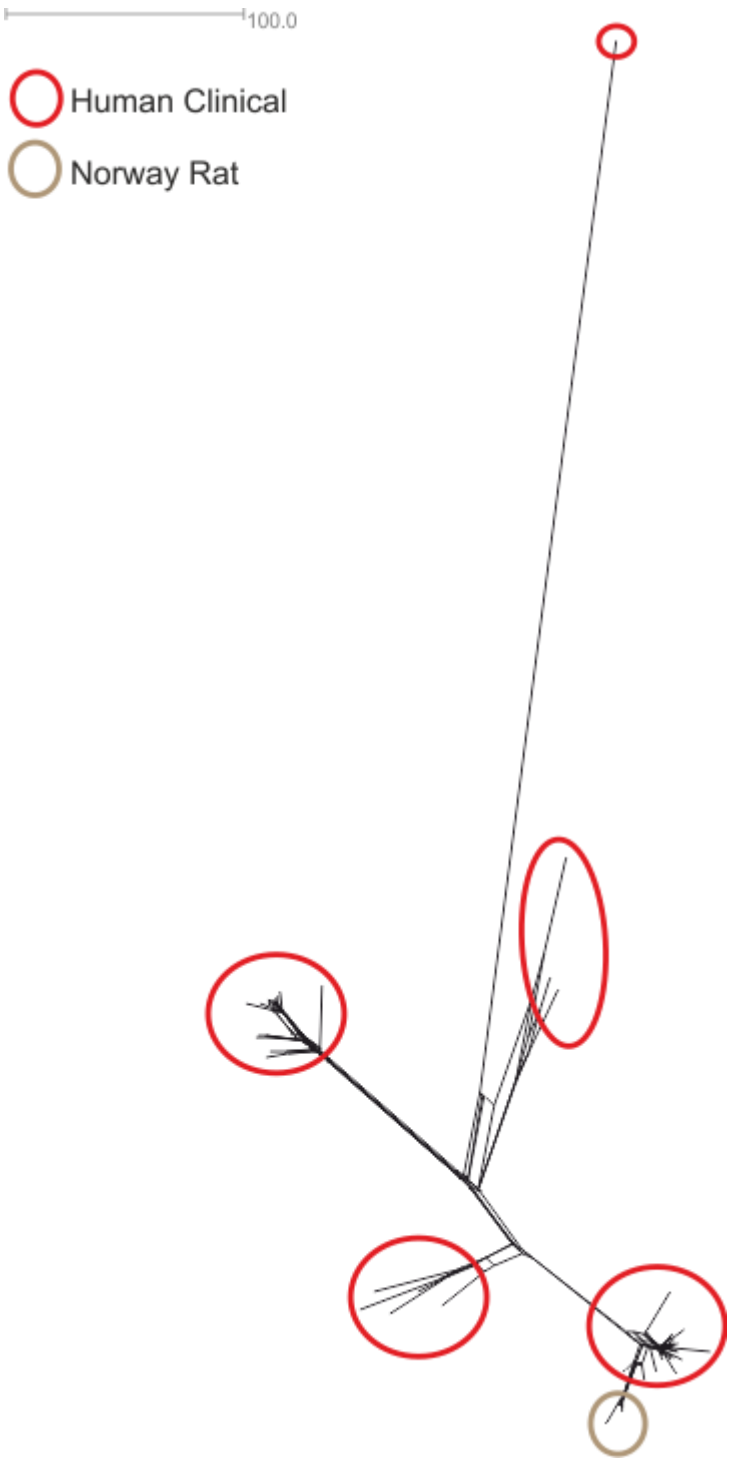


Figure 24. WGMLST Neighbor-Net diagram, showing the relationship between ST-42 isolates from Norway rats and human or livestock sources.

5.3 Discussion and Conclusions

The high genetic diversity observed overall in this sample of *Campylobacter jejuni* and *C. coli* isolated from Norway Rats, coupled with the very low diversity of types at farm level is consistent with the existence of relatively isolated island-like networks of *Campylobacter* transmission, with dominant strains sustained within a single colony, but less transmission between colonies. In the wild Norway Rats are gregarious; living in close contact with members of their own colony, but defending their territory violently against imposters. Unknown males entering a *Rattus norvegicus* colony are routinely attacked and usually die within hours of contact with the colony, often without external signs of injury[202]. In the presence of sufficient resources, Norway Rat colonies are relatively stable and largely isolated from one another, with ample transmission opportunities within colony but few transmission events between colonies; the migration of females, to avoid in-breeding[203] providing the main opportunity for between-colony transmission of *Campylobacter*.

In this study data were collected from each farm on one or more occasions during the study period. Longitudinal isolation of *Campylobacter* from rat faeces on specific farms over a sustained period would provide more evidence to elucidate the dynamics of the *Campylobacter* populations within rats over time.

The existence of a distinct rat *Campylobacter jejuni* clade (RG1) separated from livestock and disease-associated strains by an accumulation of small differences across the genome is suggestive of a long and slow process of divergence from a common ancestral type. This is consistent with a hypothesis of rat-adapted and livestock-adapted types which inhabit the same geographical location, but which exploit non-overlapping

ecological niches within the space. There is a possibility of 'livestock-type' *Campylobacter* and 'non-livestock-type' *Campylobacter* showing some genetic adaptations to their particular niches. There is a body of evidence to demonstrate association between common *Campylobacter* strains and host at the ST level[36,73] and there is some experimental evidence of host-specificity in certain strains of *Campylobacter*. ST-3704 shows a strong association with bank voles despite temporal and geographical boundaries[84,204]. Although the same ST has occasionally been isolated from other riparian and sylvatic sources, experimental infection *in vitro* has shown that it fails to colonise chicken cells[204]. The availability of whole genome data facilitates the identification of putative niche-specific genes which can be used to inform further empirical testing.

The lack of similarity between the genome of the bank vole-derived ST-3704 isolate and the Norway rat isolates coupled with the dramatic differences between the three main rat clades demonstrates that there is not a single rodent-associated *Campylobacter jejuni* lineage. The implication is that there is a far more complex evolutionary history shaping the structure of extant *Campylobacter jejuni* and *C. coli* populations.

The persistence of exclusively rat-specific alleles at 171 loci in RG1 is indicative of a bacterial population which is not readily recombining with livestock-associated *Campylobacter jejuni* strains. There is no compelling evidence to suggest that the rat-specific alleles have an adaptive advantage enabling their survival in rats, as only a small proportion of these alleles are seen in RG2 and RG3, and Norway Rats are also demonstrably capable of being infected with at least some livestock-associated *Campylobacter* strains, which do not possess rat-specific alleles. Given that

Campylobacter jejuni is a highly recombinogenic species[205] the evident lack of recombination between RG1 and other lineages, particularly livestock-associated lineages, is suggestive of separate populations which do not co-exist in the same host often enough or for long enough to facilitate recombination, and lends support to the hypothesis that ecological barriers to genetic exchange play an important role in the persistence of distinct *Campylobacter* populations.

The significance of the absence of 84 genes identified in NCTC11168 from all RG1 isolates is unclear. This study provides no evidence to suggest a loss of fitness, however the lack of the *cfrA, Cj0265c, Cj0241c, cfbpB, Cj0737* and *Cj0177* genes, which are related to iron uptake may represent a competitive disadvantage against ST-21 or other livestock-associated STs *in vivo*. Further experimentation would be required to test this hypothesis, but it is consistent with the apparent absence of RG1 STs in livestock, despite geographically congruent transmission cycles. The absence of the *cdtA* and *cdtC* genes from RG1 may provide a partial explanation for the absence of RG1 STs in human disease. The *cdtABC* gene cluster, which encodes the cytolethal distending toxin[206] has been implicated in pathogenesis in the human host[207], by gradually distending cells, resulting in cell death[1,208]. The full complement of *cdt* genes is routinely found in human clinical *Campylobacter jejuni* isolates[209]. There is some evidence in the literature to support the occurrence of asymptomatic *Campylobacter* carriage in humans[210]. The status of the *cdt* gene cluster in human isolates which are carried asymptomatically is not known.

The bifurcation of the ST-45 CC into 'classic' livestock-associated STs and 'novel' rat-associated STs, which are more closely related to isolates from the newly identified and

unassigned RG2 than to classic ST-45 CC isolates, suggests that the ST-45 Clonal Complex is a polyphyletic group. As a multi-host complex, ST-45 CC is epidemiologically different from the majority of known *Campylobacter jejuni* complexes[211]. While the ST-45 complex might be termed a 'generalist lineage' the apparent polyphyly within the clonal complex could suggest that it is not a single generalist lineage but a collection of highly adapted lineages displaying convergent evolution. This would then raise the question of the adaptive merit of the ST-45 complex phenotype. An alternative, and perhaps more satisfactory, explanation is that the ancestral ST-45 CC genotype has arisen in Norway Rats or another sylvatic species and has diversified, giving rise to the closely related Novel ST-45 CC and RG2, and the Classic ST-45 CC (figure 25).

Figure 25. Cladogram.

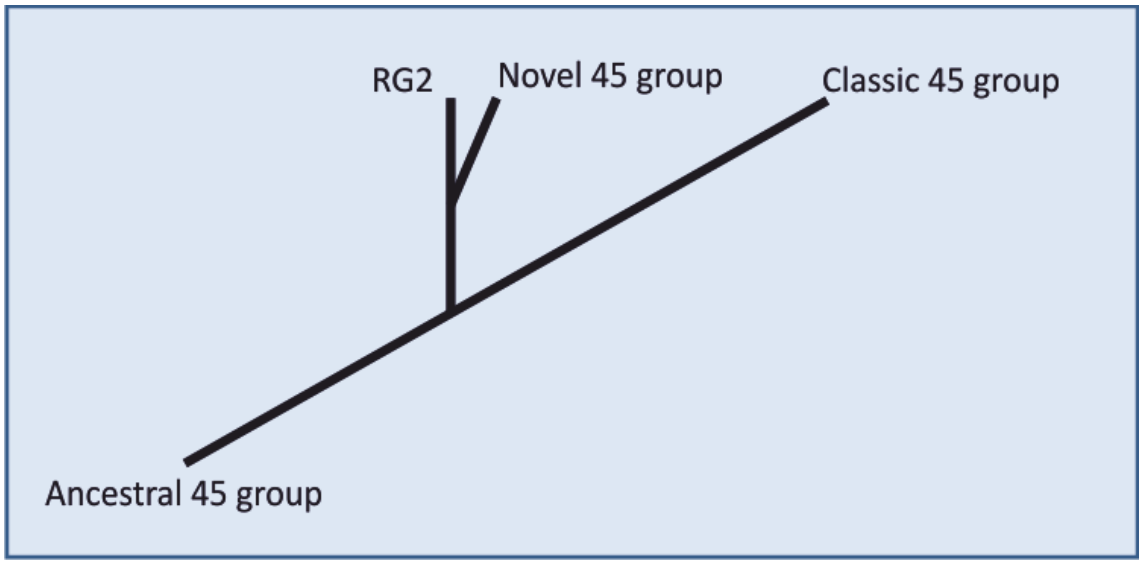


Figure 25. Cladogram depicting a plausible evolutionary history of ST-45 CC and related *Campylobacter jejuni* types.

The high prevalence of carriage of *Campylobacter* isolates belonging to CCs and STs also associated with livestock does not provide enough evidence to incriminate Norway Rats

as a reservoir or vector of public health importance in the livestock to human *Campylobacter* transmission cycle. Higher resolution analysis of the cattle-associated ST-42 using the gene-by-gene approach to identify similarities and differences across the genome demonstrates a separate clade of rat-associated *Campylobacter* types below the ST-level, the same sub-type has not been identified in livestock or disease (figure 24), although the current collection of whole genome data is not sufficient to test for a unique association between rats and the ST-42 sub-type. Of the nine rat isolates with the same ST-42 sub-type, eight were from a single mixed turkey and cattle farm and the ninth was from a separate cattle farm. The identification of the same sub-type at two separate farms could indicate that this is a specific genotype with a higher propensity to infect rats, although more extensive sampling would be required to address this question fully.

ST-21 CC, like the ST-45 CC, is considered a multi-host complex. Like rat-derived ST-42 isolates, all ST-21 isolates from rats are of a single sub-type. Apparently unlike ST-42, however, the ST-21 sub-type occurring in rats is also identified in livestock and human disease, but it is not clear whether this type has a specific property which facilitates colonisation of the rat. This study did not examine the concurrent *Campylobacter* types in livestock on the farms. It is possible that the ST-42 and ST-21 cases are examples of a livestock-associated *Campylobacter* type which infects rats temporarily. The current study does not provide conclusive evidence for or against this hypothesis and neither is there an existing body of data to address this question fully. The findings of a Danish study of *Campylobacter* carriage by wildlife in the immediate vicinity of livestock farms also suggest that transmission from livestock to wildlife may occur, although this is

largely speculation, based on the feeding patterns of wild bird species, rather than molecular evidence[212]. A small scale study using AFLP found no shared *Campylobacter* genotypes between a small sample of rats and pooled pig excreta samples[91]. A more thorough analysis would require long-term simultaneous testing of rats and livestock on the same farms, and the gene-by-gene approach to WGS should be used to provide sufficient resolution.

While the possibility of direct transmission of *Campylobacter* between Norway Rats and farmed livestock species cannot be ruled out in specific cases where an ST is shared, the overall picture emerging from this study is one of largely separate rat and livestock *Campylobacter* transmission cycles, with overlap only in a relatively small proportion of STs; this is suggested by the diversity of rat *Campylobacter* strains which have never been identified in livestock or human disease, and equally the diversity of livestock and human disease strains which are not reflected in the rat sample. Based on this observation, the benefit of on-farm biosecurity measures aimed specifically at the prevention of *Campylobacter* transmission between rats and livestock is questionable. The importance of good biosecurity and animal husbandry for the health and welfare of livestock is not disputed but it is suggested that the exclusion of rats from livestock houses is not a sufficient condition for the reduction of *Campylobacter* entering the human food chain, and neither has its necessity for this purpose been well supported to date.

Conventional 7-locus MLST provides an effective method for differentiating between relatively distantly related strains within a species, however the clonal complex and sequence type definitions would appear to be too blunt an instrument for a fuller

investigation of the evolution and ecology of *Campylobacter*. The availability of whole genome data and the ease of use of the gene-by-gene approach to handle genomic data efficiently and effectively facilitate the study of more complex relationships between bacteria and host, and between separate bacterial populations.

Chapter Six: An Exploration of the Potential to Use Genomic Data to Enable More Accurate Attribution of Human Clinical Campylobacteriosis to Putative Host Sources Using the STRUCTURE Bayesian Attribution Model.

6.1 Introduction

Campylobacter is a serious public health problem. An understanding of the reservoirs of infection and modes of transmission of *Campylobacter* is imperative for the implementation of successful public health interventions aimed at reducing the burden of campylobacteriosis in the UK and worldwide.

Since first being identified as an important human pathogen in the 1970s, farm animals, particularly chickens, have been implicated as the main reservoirs of *Campylobacter*.

Case control studies in the UK, Denmark and Sweden with comprehensive sampling of cases and well-matched controls have identified the consumption of raw or undercooked chicken as a major risk factor for sporadic campylobacteriosis[11,81,213].

An opportunistic study in Belgium, following the 'dioxin crisis' in 1999, whereby domestically raised poultry was contaminated with high levels of the chemical dioxin and was subsequently withdrawn from sale, saw a 40% reduction in the incidence of campylobacteriosis compared with projected figures based on seasonality and ongoing secular trends. *Campylobacter* incidence returned to normal after the four-week ban on poultry-sales was lifted[174].

Outbreak investigations have also contributed to an understanding of the role of poultry and waterfowl consumption in the transmission of *Campylobacter* by identifying the common source in a point-source outbreak, with epidemiological data supported by laboratory analyses[55,56,214].

Multi Locus Sequence Typing (MLST) is a robust framework for the standardised characterisation of *Campylobacter* isolates from any given source, which has provided an insight into the ecology of the genus. There is a marked association between Sequence Type (ST) and host species[36,71,73]. Populations of strains isolated from chickens show great diversity, and are highly structured into multiple STs and clonal complexes, but there is a high degree of consistency between chicken populations, regardless of geographical provenance, while there is far less similarity across the seven MLST loci between populations of *Campylobacter* isolated from different host species within the same geographical location[36].

Clinical *Campylobacter* isolates do not represent the full diversity of animal strains, but rather a non-random subset of predominantly 'agricultural' sequence types; this is particularly marked in the case of *C. coli* where broad sampling of humans and wild and domesticated animals identifies three separate clades, but disease cases and agricultural isolates are limited to clade 1[35].

A number of statistical models have used the molecular host signal, defined by MLST, as a means of attributing disease cases to their putative source. The *Campylobacter* Source Attribution Model (CAMSA) is an adaptation of the Hald Model[215] which compares the number of human campylobacteriosis cases caused by a given ST with the prevalence of

the same ST isolated from a range of food sources, and weighted by national figures for consumption, with the assumption that the annual incidence of domestically acquired sporadic campylobacteriosis attributed to a given food source is proportional to the amount of that food consumed that is contaminated with the same *Campylobacter* ST[216]. A significant proportion of clinical isolates cannot be attributed using this method, because there is no exact MLST match identified in target food sources, equally, STs which are common to several host species are problematic. The use of Bayesian statistical models, treating each locus individually within the context of a full allelic profile, facilitates a more nuanced analysis which allows for uncertainty and the possibility of mixed genetic background. Examples of Bayesian statistical models which have been used for disease attribution studies are the Asymmetric Island Model (AIM)[75] and STRUCTURE[138,217]. The Asymmetric Island Model is an adaptation of the Migration-Matrix Model[218] treating each *Campylobacter* reservoir as a discrete island. It is assumed that homogeneous mixing takes place within each island, and migration takes place to a greater or lesser extent between the islands[75]. Experimental attribution using AIM with MLST data was 95% accurate for chickens, 98% accurate for ruminants and 77% accurate for wild birds[69]. The STRUCTURE model characterises populations, in this case host reservoirs, by a set of allele frequencies at each locus. Clinical isolates are then probabilistically assigned to a population based on the combination of alleles at those loci. Experimental attribution of a set of isolates of known source using STRUCTURE and MLST data was 70% accurate in the case of chicken, 84% accurate in the case of ruminants, and 54% for wild birds[69]. The accuracy of all of the MLST-based models is impaired by the ubiquity of the ST-21 and ST-45 clonal complexes which, in contrast to

the other known clonal complexes, are abundant and widespread throughout farm animals and clinical cases, showing no obvious host associations[219,220].

Whole Genome Sequencing (WGS) of reference and disease strains provides high resolution, highly discriminatory data which demonstrates diversity within STs and enables the study of genealogical structuring below the ST level.

Here the use of seven-locus MLST with the STRUCTURE Bayesian attribution model[138] was examined, and the potential for a genome-wide scheme with greater resolution and the ability to distinguish closely related strains with different ecological requirements, investigated.

6.2 Results

Within and Between Population Diversity at MLST Loci

A Neighbour Joining Tree based on concatenated nucleotide sequences for the reference set isolates at the seven MLST loci (figure 26) demonstrated that while few individual STs were shared between multiple hosts there was no clear structure of putative lineages routinely correlating with individual hosts.

Figure 26. MLST neighbor-joining tree for animal reference isolates.

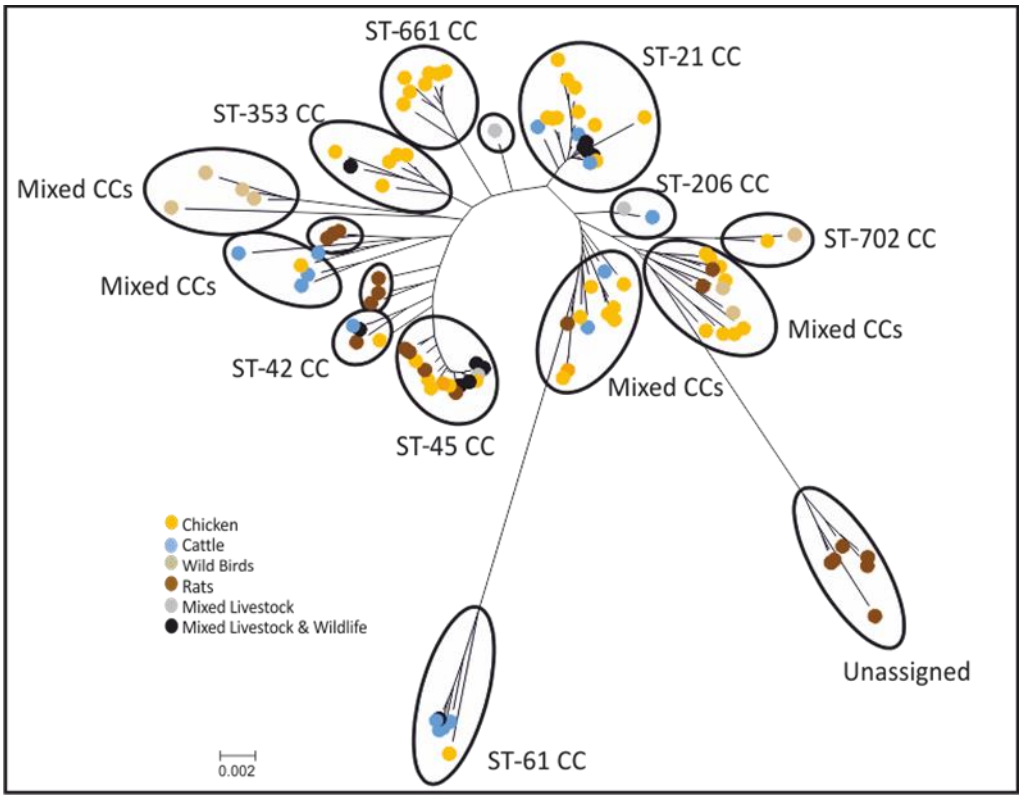


Figure 26. Neighbour-Joining Tree displaying estimated phylogenetic relationships between animal reference isolates (n = 448). Constructed using concatenated 7-locus MLST sequences.

The ST-661 and ST-353 CC clusters appeared to be associated with chicken, and there was evidence of three rat-associated clusters, one group of wild bird associated STs, which is likely to represent two distinct lineages, and a group of mixed CCs which was predominantly, though not exclusively, associated with cattle. The majority of isolates, however, appeared on branches of the tree which show no association with any given host from the reference set.

Using conventional MLST to define unique types, a total of 105 STs and 57 isolates with incomplete STs were identified across the entire reference dataset. Fifteen STs were

shared between more than one host population. Overall, 36.16% of all reference isolates belonged to shared STs, comprising 64.29% sheep isolates, 57.33% cattle isolates, 31.74% rat isolates, 33.33% of wild bird isolates and 29.86% chicken isolates. Sheep and cattle showed a considerable degree of sharing, with 64.29% of the sheep population having STs in common with 40.00% of the cattle population. Sequence types found in Norway Rats accounted for a significant proportion of each of the other populations; however a considerable degree of diversity not seen in any other population was also apparent within the Norway rat population. Based on absolute values the two most common sequence types, accounting for 14.29% of the overall sample were ST-45 and ST-21, both of which occurred in livestock and wild host sources (figure 27). The next most common sequence types were ST-257 and ST-60, each representing 4.46% of the overall sample, and identified exclusively in chickens. Given the variation in the sizes of the populations it is useful to look at each ST as a proportion of each host population (figure 28). In these terms the top 10 STs are ST-61, ST-45, ST-21, ST-206, ST-137, ST-7259, ST-48, ST-19, ST-42 and ST-53. With the single exception of ST-7259 all of these common STs were present in multiple hosts within the sample.

Figure 27. Common STs in animal reference set: absolute values.

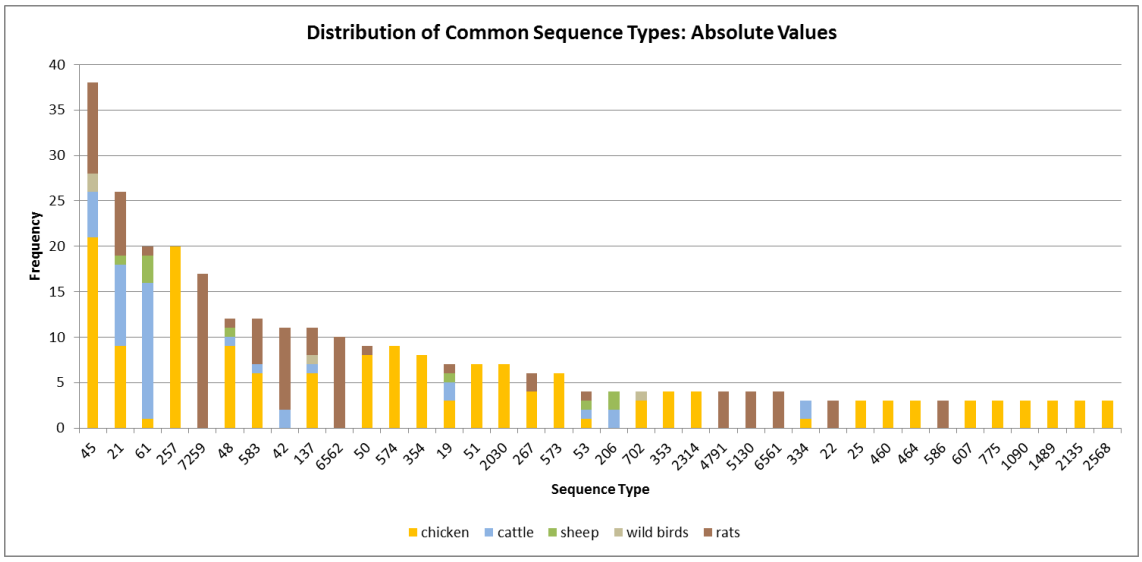


Figure 27. Stacked barchart showing the absolute frequency of STs isolated from each host source, for STs which are represented a minimum of 3 times in the reference dataset.

Figure 28. Common STs in animal reference set: relative values.

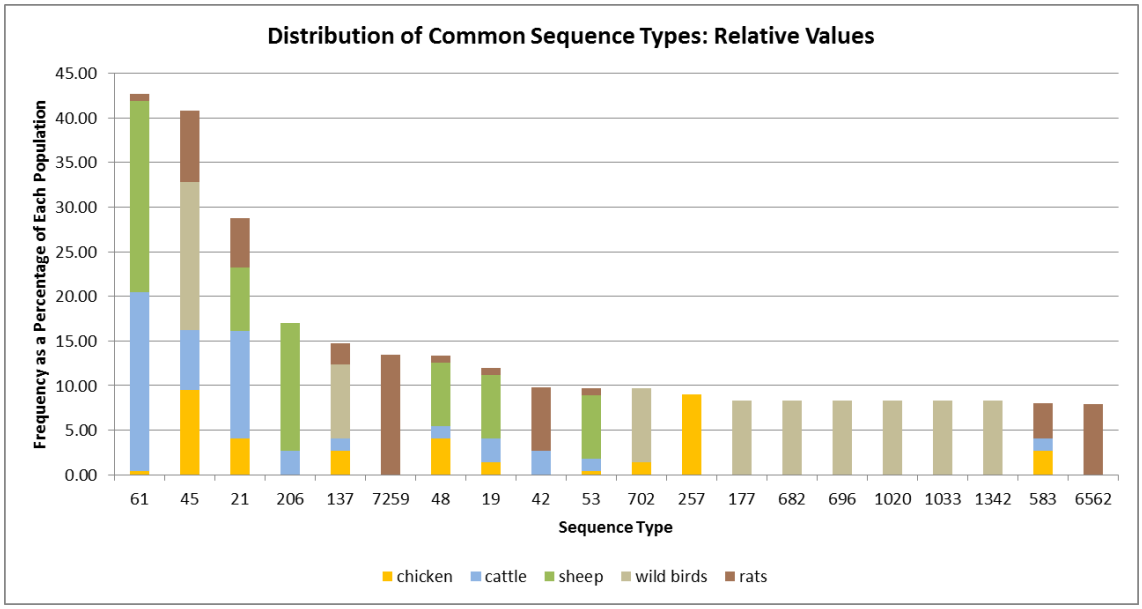


Figure 28. Stacked barchart showing the relative frequency, as a percentage of the respective populations, of STs which represent at least 5% of the stratified reference populations overall.

On average there were 30.14 alleles per locus, ranging from 20 at the *aspA* locus to 46 alleles at the *pgm* locus. Regardless of the allelic diversity of the locus across the complete reference set there was a high degree of sharing of alleles between populations (figure 29). The highest degree of sharing of alleles was at the *glt* locus (figure 29c), where 93.66% of isolates had non-source-specific alleles and the lowest was at the *pgm* locus (figure 29e), with 63.29% of isolates having alleles from a shared pool.

Figures 29a – 29g. Frequency of alleles at each locus.

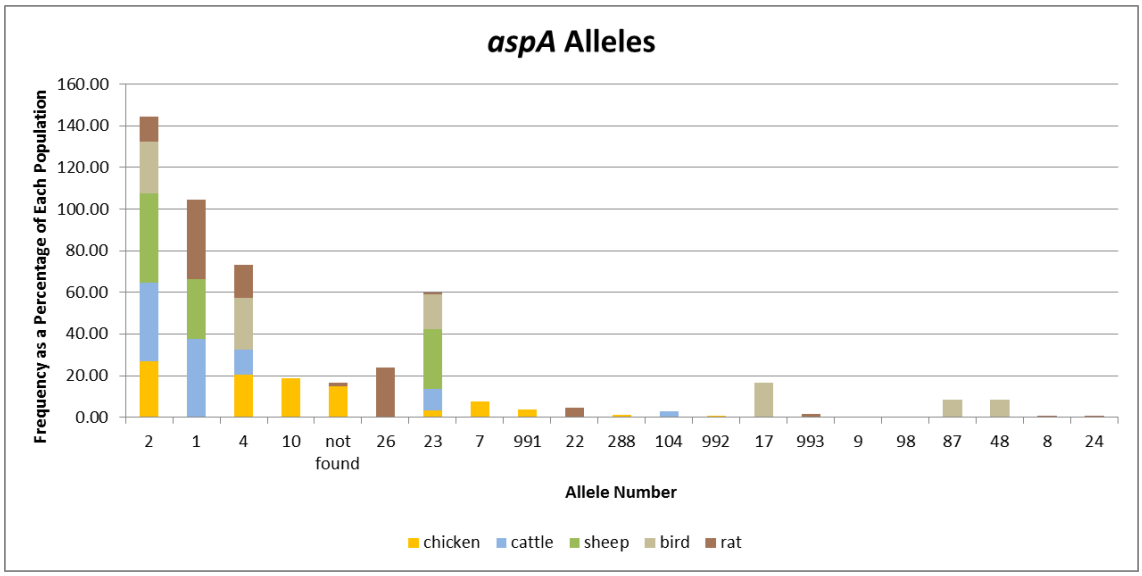


Figure 29a.

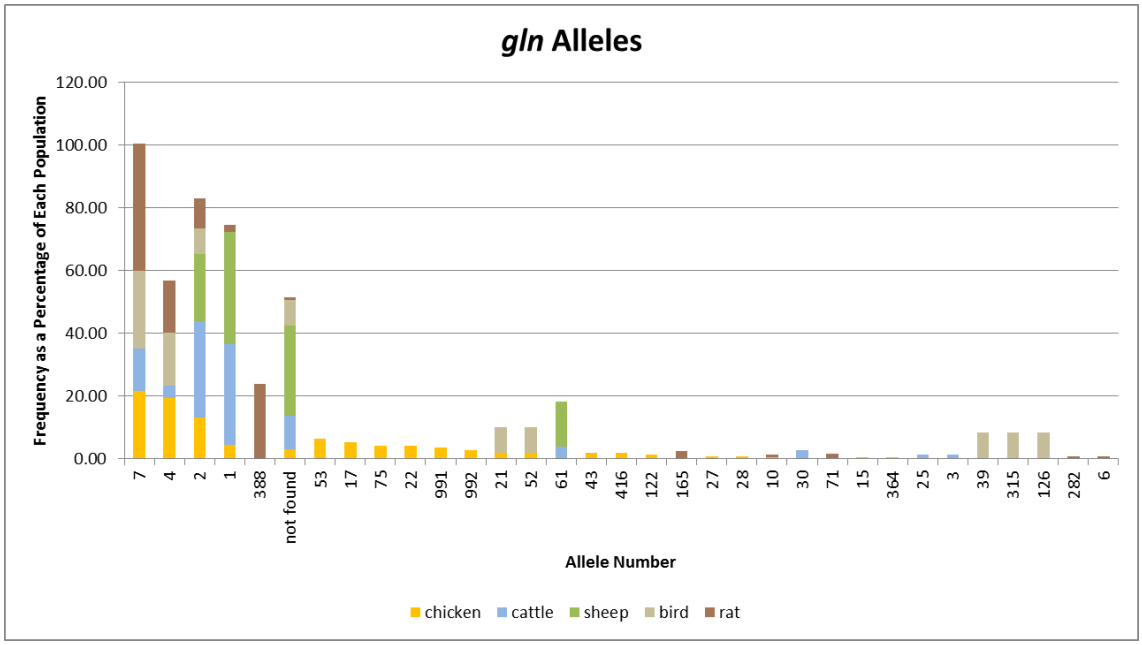


Figure 29b.

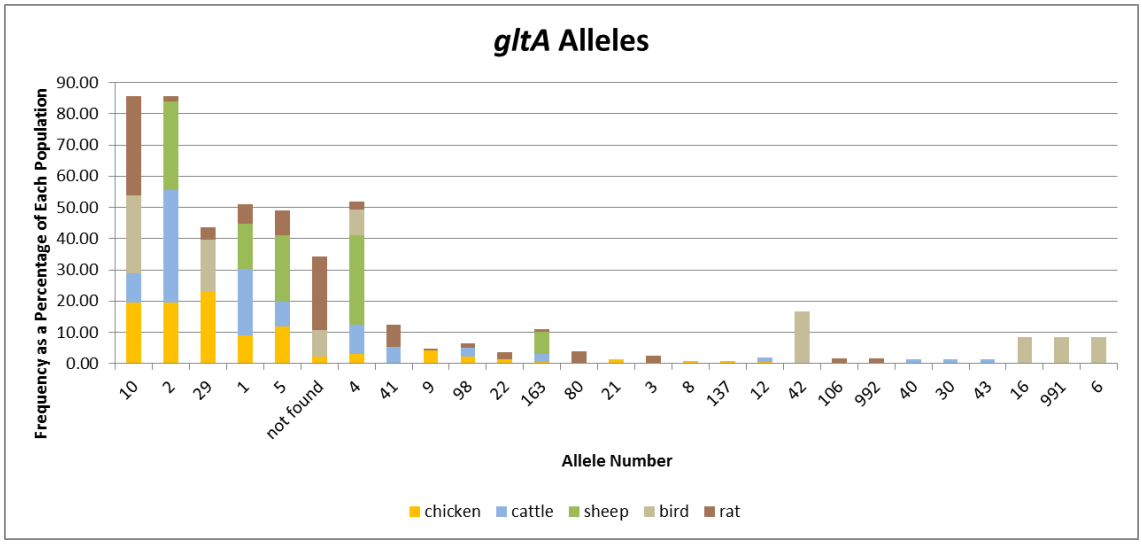


Figure 29c.

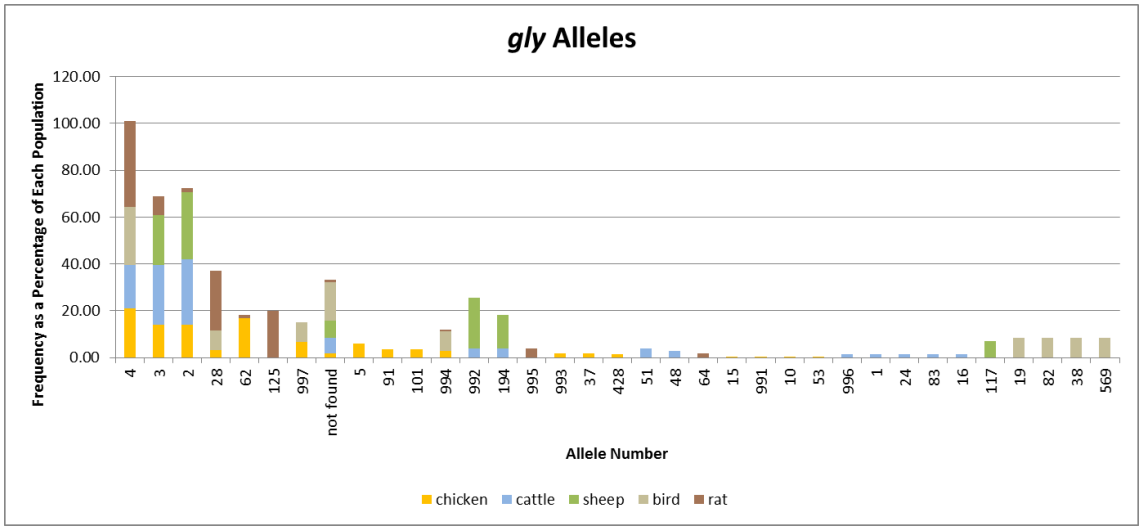


Figure 29d.

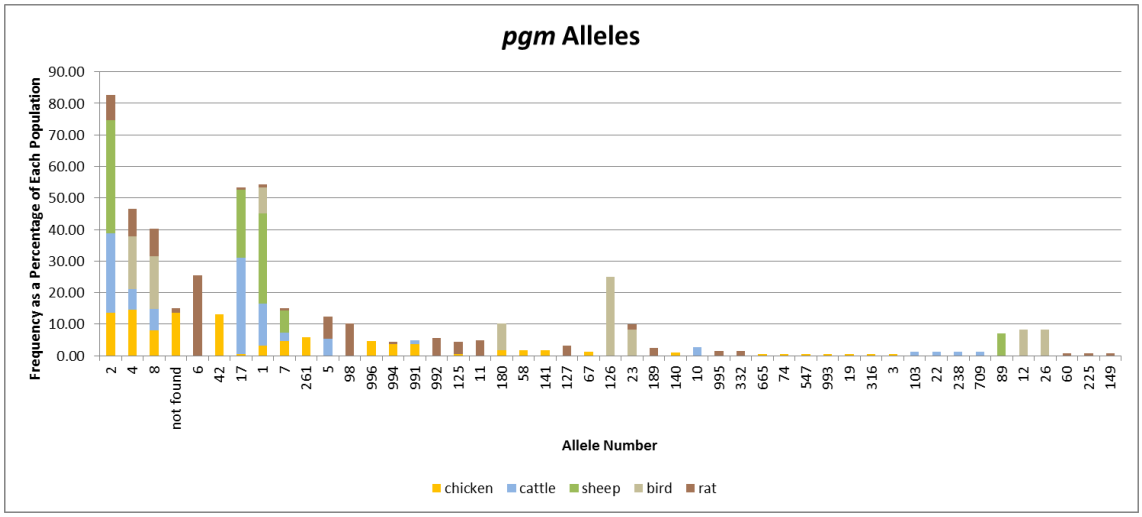


Figure 29e.

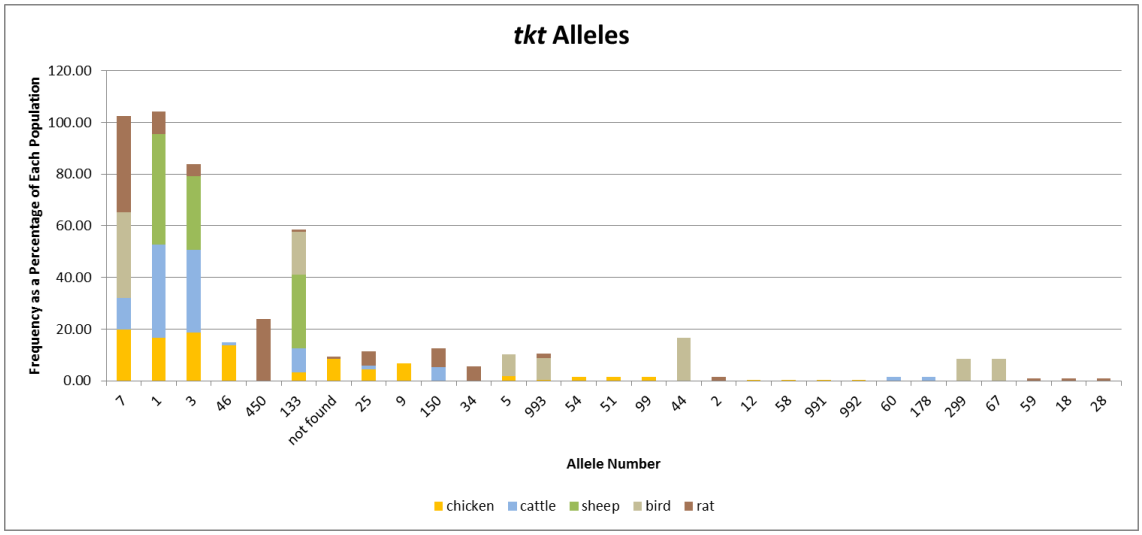


Figure 29f.

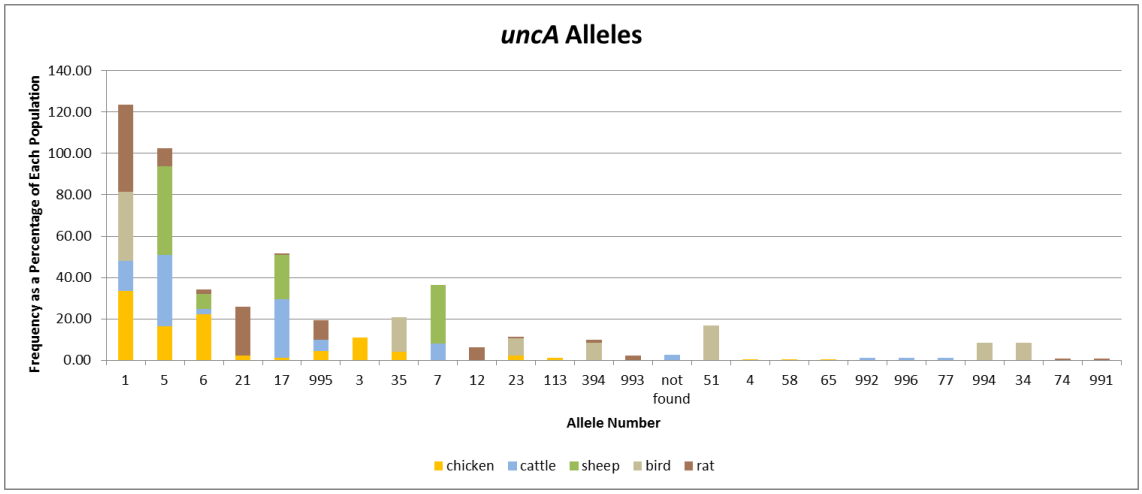


Figure 29g.

Figures 29a-g. Stacked barcharts showing the frequency of alleles at each of the MLST loci, as a percentage of each population.

Population genetic analyses performed in Arlequin[151] showed low discrimination between the host populations using MLST. Standard AMOVA, using allele numbers across the seven MLST loci demonstrated only 9.26% between-population variation and

90.74% within-population variation. The degree of within- and between-population diversity differed slightly at each locus when taken individually (table 13). Population pairwise F_{ST}s demonstrated statistically significant, but minimal distinctions between each pair of host populations, with the exception of cattle and sheep, which appeared to share a common pool of *Campylobacter jejuni* genotypes (table 14).

Table 13. Locus-by-locus AMOVA results for MLST loci.

	Between Populations				Within Populations					
Locus	SSD	d.f.	Va	% variation	SSD	d.f.	Vb	% variation	F_{ST}	P-value
asp	17.58	4	0.06	12.83	170.84	438	0.39	87.17	0.13	0.00
gln	12.32	4	0.04	8.37	182.91	442	0.41	91.63	0.08	0.00
glt	11.86	4	0.04	8.10	181.74	436	0.42	91.90	0.08	0.00
gly	12.06	4	0.04	8.28	182.49	438	0.42	91.72	0.08	0.00
pgm	12.80	4	0.04	8.68	184.62	420	0.44	91.32	0.09	0.00
tkt	12.87	4	0.04	8.91	180.37	440	0.41	91.09	0.09	0.00
unc	13.62	4	0.04	9.81	173.61	443	0.39	90.19	0.10	0.00

Table 13. Results of locus-by-locus AMOVA testing for the MLST scheme, showing the degree of between-population and within-population variation for each separate locus.

Table 14. Population pairwise F_{ST}s using the MLST scheme.

pop A	Population Pairwise F_{ST}s (p-values)				
Chicken	0.00				
Cattle	0.08 (0.00)	0.00			
Sheep	0.12 (0.00)	-0.01 (0.41)	0.00		
Wild Birds	0.03 (0.03)	0.14 (0.00)	0.20 (0.00)	0.00	
Rats	0.08 (0.00)	0.14 (0.00)	0.21 (0.00)	0.06 (0.01)	0.00
pop B	**Chicken**	**Cattle**	**Sheep**	**Wild Birds**	**Rats**

Table 14. Arlequin output showing F_{ST} values between each pair of host populations.

Within and Between Population Diversity beyond the Seven MLST Loci

Locus-by-Locus AMOVA reveals that there are significant differences in the degree of within- and between- population diversity across the genome.

Alternative Sequence Typing Scheme 1

The seven variable loci displaying the greatest percentage of between-population variation, discarding any contiguous loci, were selected for testing as alternative Sequence Typing scheme (1) (table 15).

Table 15. Candidate loci for alternative typing scheme (1).

Locus	Molecular Function	% between – population variation
cheY	metal ion binding	19.18
rpsR	rRNA binding	18.71
rpmH	structural constituent of ribosome	15.93
hupB	DNA binding	15.04
coaD	ATP binding	14.52
rplL	structural constituent of ribosome	15.84
Cj0711	RNA binding	14.41

Table 15. Seven non-contiguous loci showing the highest degree of between-population variation within the animal reference set.

A Neighbour Joining Tree (figure 30) based on concatenated nucleotide sequences for the reference set isolates at the seven aST(1) loci demonstrated some limited evidence of putative lineages being associated with a specific host, although the relationship was not unambiguous. The majority of chicken isolates fell within two large clusters. Cluster one also contained two isolates from rats while cluster 3 contained a single isolate from a wild bird and one small sub-cluster with a single aST(1) from chickens and one wild bird; and an aST(1), from a wild bird and two Norway rats within this reference sample. Cluster five was exclusive to wild birds, specifically starlings. Clusters eight and nine were both exclusive to Norway rats. Cluster two was a broad group, containing isolates from all sources, though more cattle isolates than any other single source. Cluster four was divided into two distinct sub-clusters, a and b. Cluster 4a was predominantly a cattle cluster, with one aST shared by cattle and sheep. Cluster 4b was a mix of cattle isolates, rat isolates and one aST shared by one bovid and nine rats. Clusters six and seven were

mixed, with aSTs isolated from both livestock and wild animals. Cluster ten was also mixed, but was a much deeper-branching structure than all other clusters.

Figure 30. Neighbor-joining tree constructed using concatenated sequences at aST(1) loci.

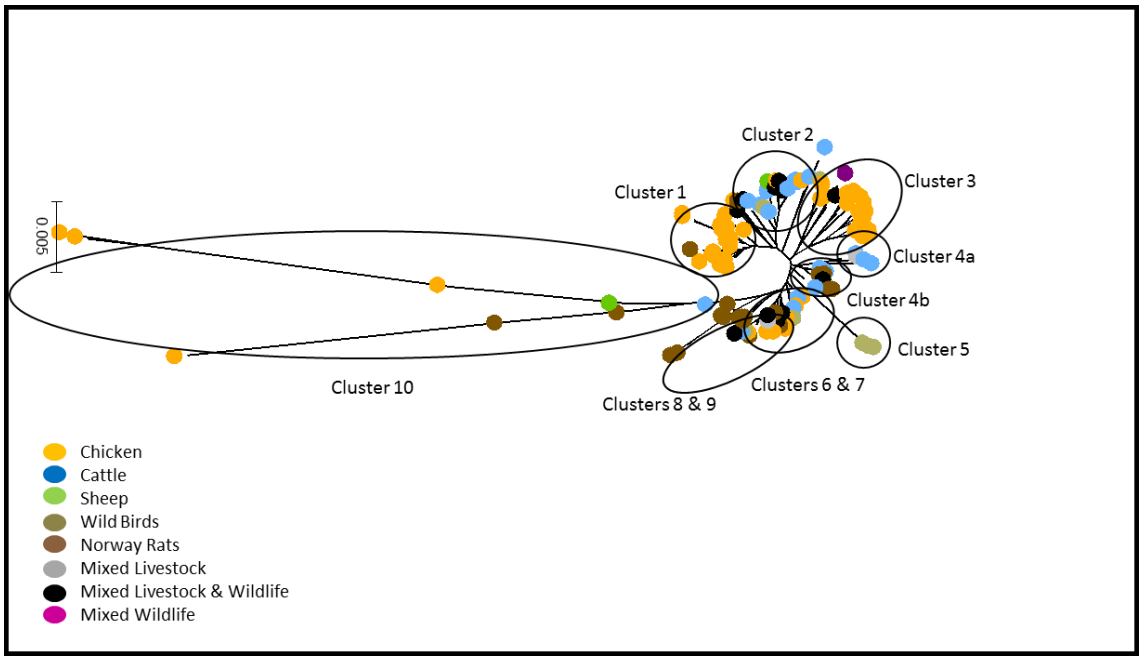


Figure 30. Neighbor Joining Tree constructed using concatenated nucleotide sequences for seven aST(1) loci, with isolate sources denoted by colour-coding.

Using alternative sequence typing scheme 1 to define types 111 unique aST(1)s, and 14 incomplete aST(1)s were identified across the reference set. Fifteen aST(1)s, representing a total of 173 isolates (38.62%) were shared between more than one host, of these two aST(1)s were shared between livestock sources only, one aST was shared between wild hosts only, two aST(1)s were shared between livestock and wild birds and 10 aSTs were shared between livestock and Norway rats (figure 31). In each case the proportion of chicken, wild birds or rats with a given shared aST was less than 10% of the respective host population. In relative terms the majority of the sharing was

between cattle and sheep, with 78.57% of sheep isolates sharing aST(1)s with 42.67% of the cattle population (figure 32).

Figure 31. Absolute frequency of alternative sequence types by source.

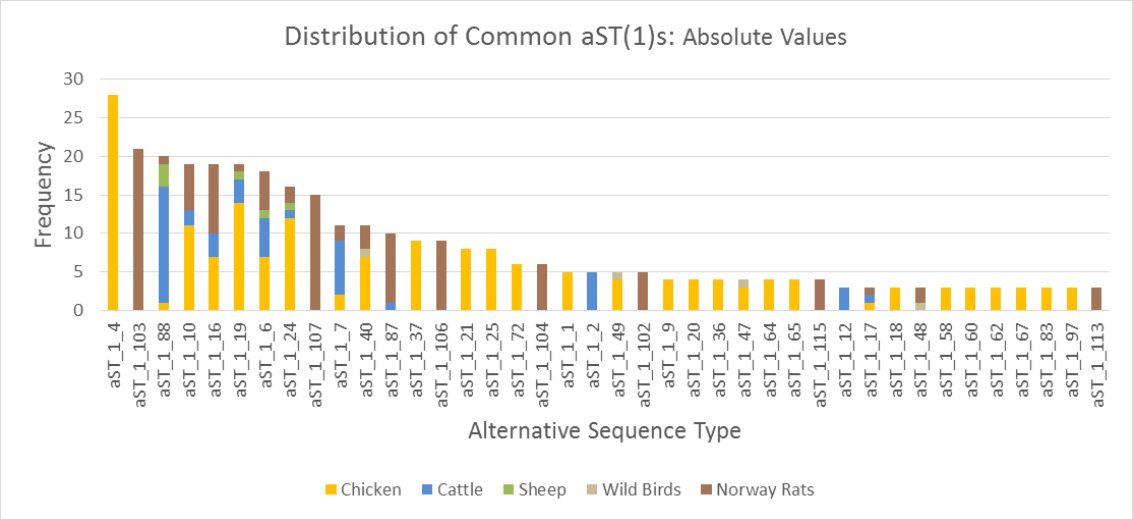


Figure 31. Stacked barchart showing the absolute frequency of aST(1)s isolated from each host source, for STs which are represented a minimum of 3 times in the reference dataset.

Figure 32. Relative frequencies of alternative sequence types by host source.

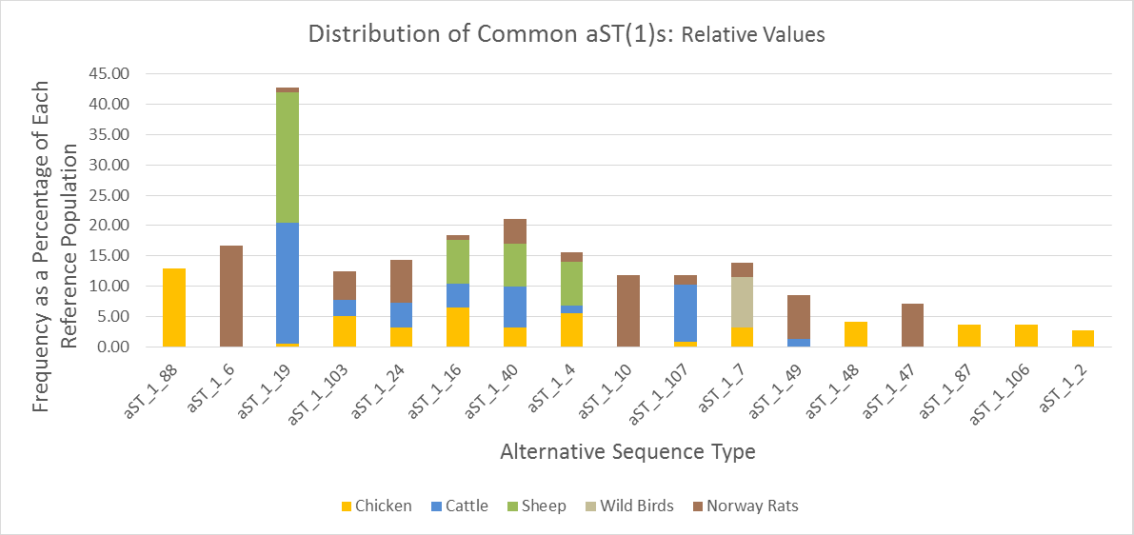


Figure 32. Stacked barchart showing the relative frequency, as a percentage of the respective populations, of aST(1)s which represent at least 5% of the stratified reference populations overall.

On average there were 20.29 alleles per aST(1) locus, ranging from 14 at the *cheY* and *rpmH* loci to 33 at the *coaD* locus. Figure 33 shows the percentage of alleles shared between each population at each locus. A total of 93.10% of isolates shared alleles from a common pool at the *rpmH* locus, while only 70.98% of isolates shared alleles at the *coaD* locus.

Figure 33a-g. Proportion of shared alleles at each of the aST(1) loci.

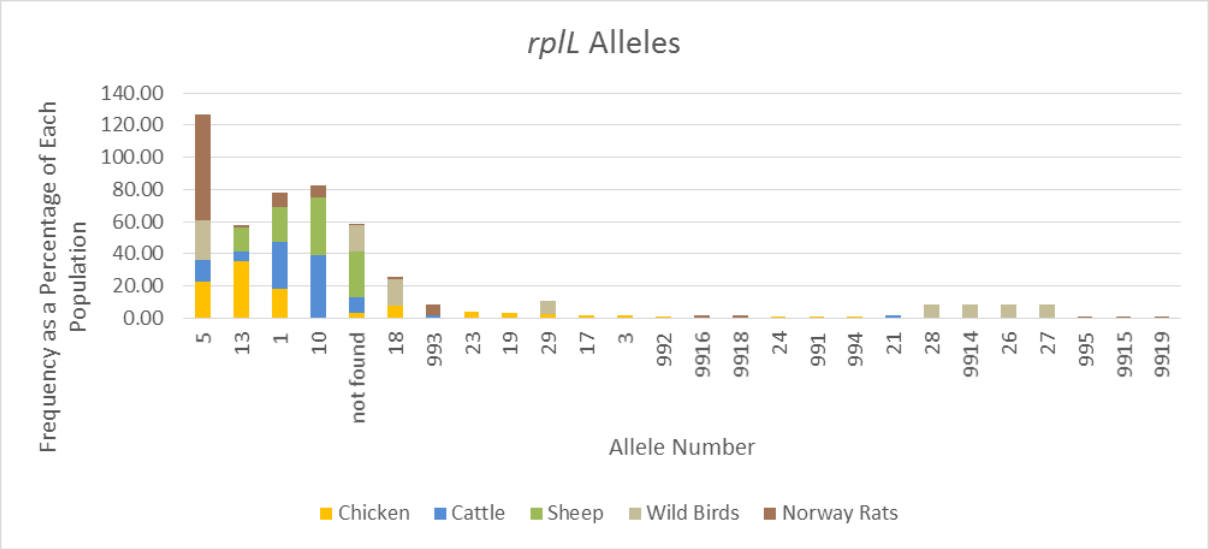


Figure 33a

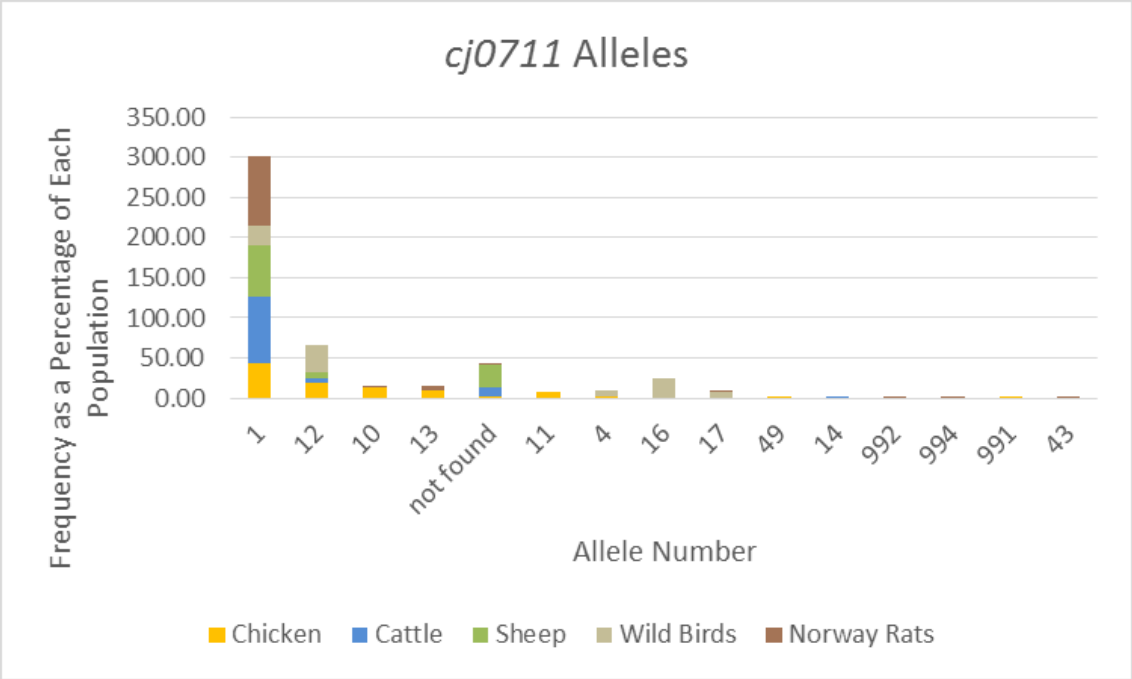


Figure 33b

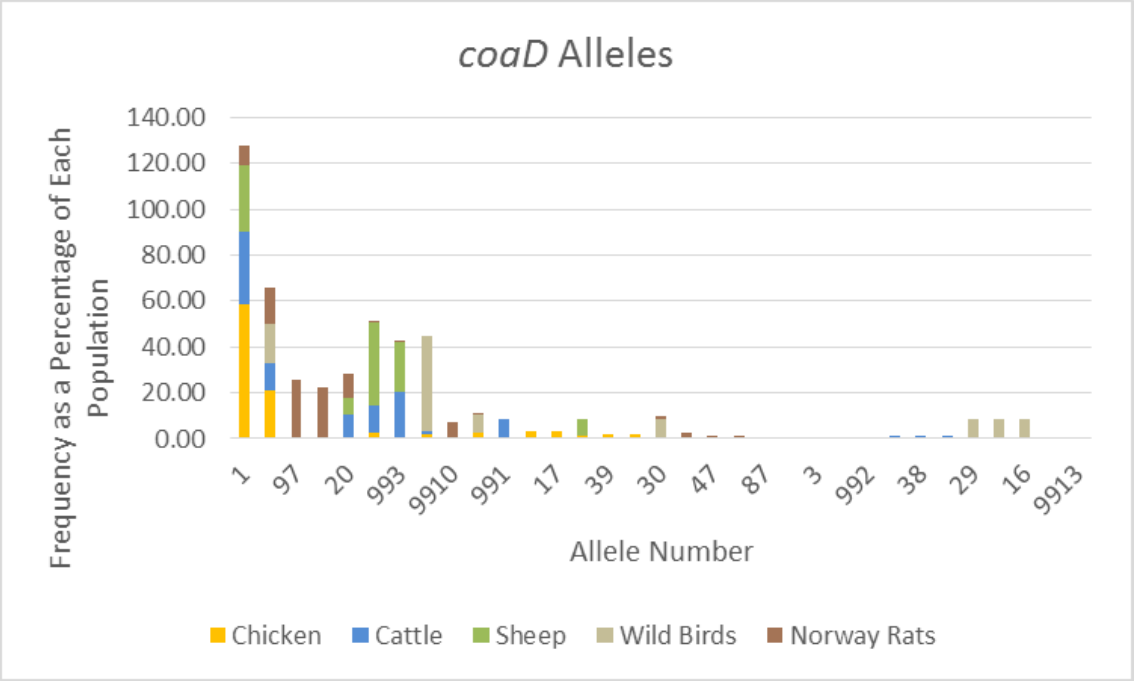


Figure 33c

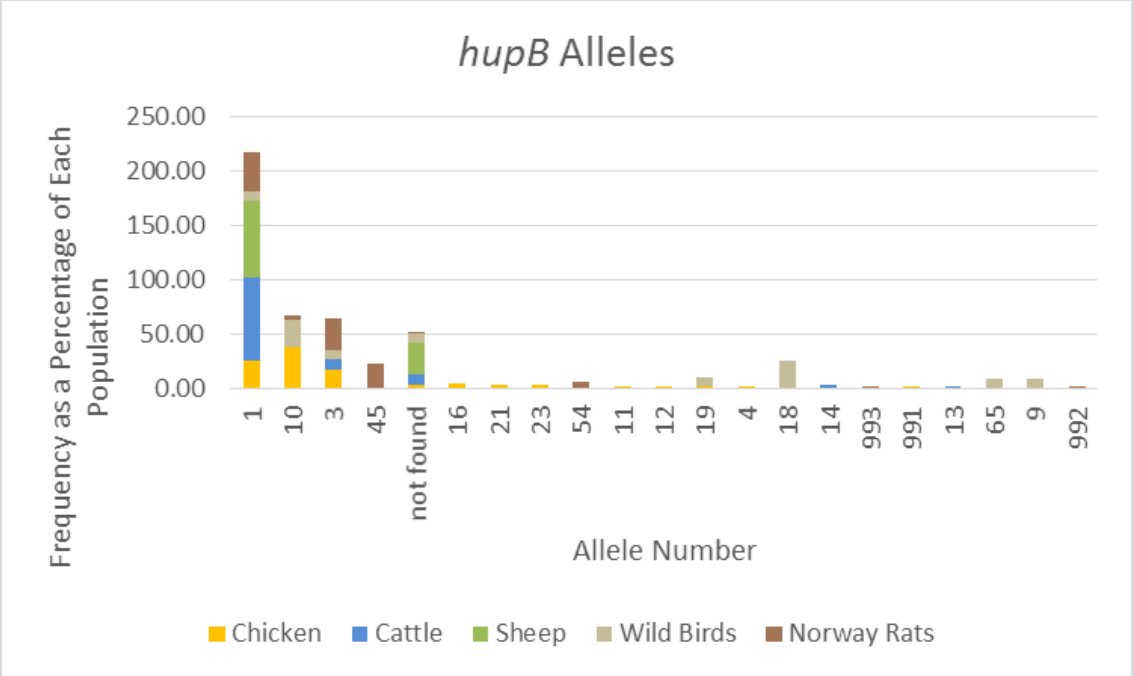


Figure 33d

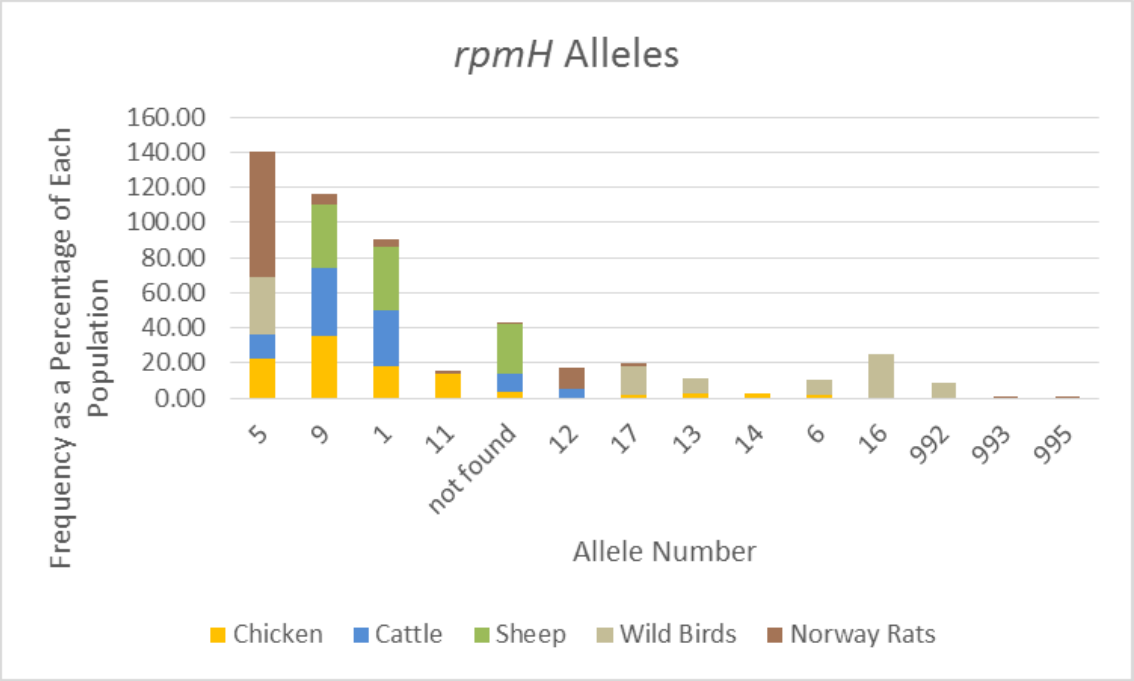


Figure 33e

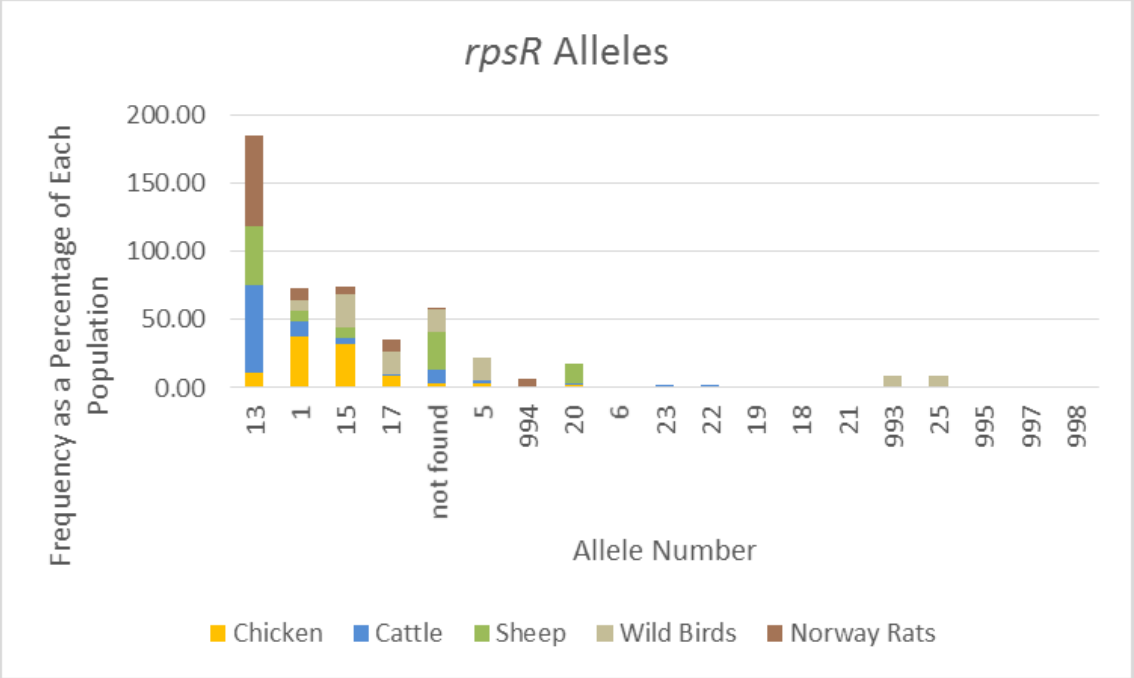


Figure 33f

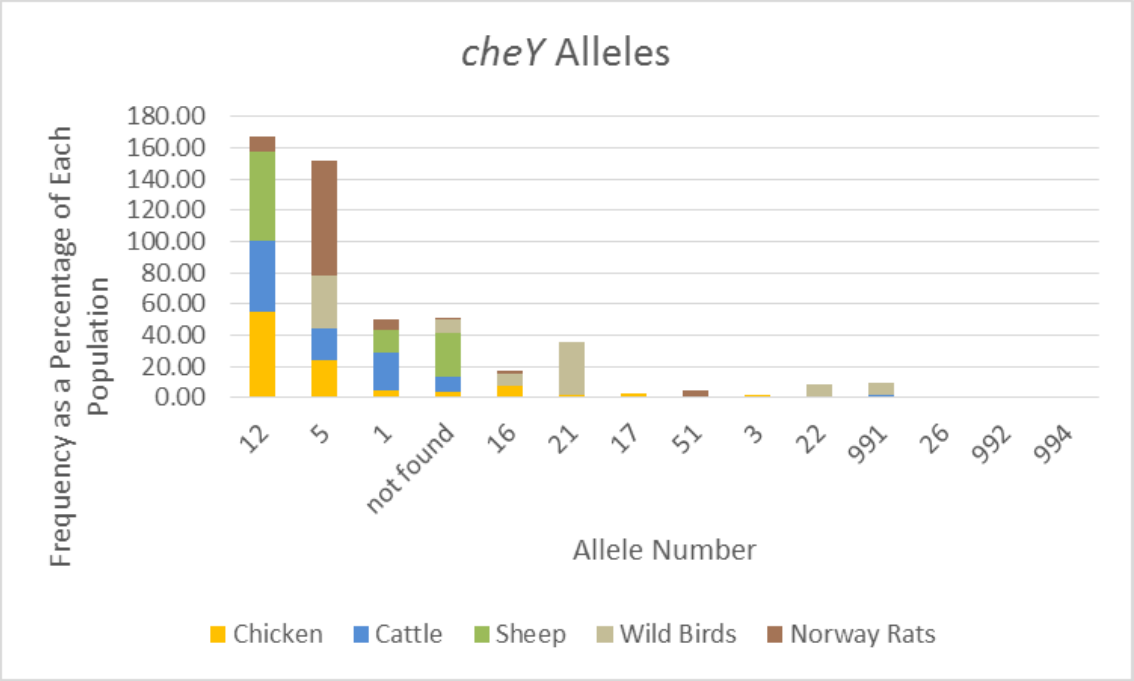


Figure 33g

Figures 33a-g. Stacked barcharts showing the frequency of alleles at each of the aST(1) loci, as a percentage of each population.

Alternative Sequence Typing Scheme 2

A Neighbour Joining Tree (figure 34) based on the concatenated nucleotide sequences for the reference set isolates at the seven aST(2) loci provided very little evidence for host-associated lineages, despite the evident relationships between particular individual aST(2) types and host sources. Of the nine main clusters identified by the NJ Tree only two appeared to be associated with a single host, chickens and Norway rats, respectively, within the reference set. All other clusters contained isolates from mixed hosts, both livestock and wildlife.

Figure 34. Neighbor-joining tree constructed using concatenated aST(2) sequences.

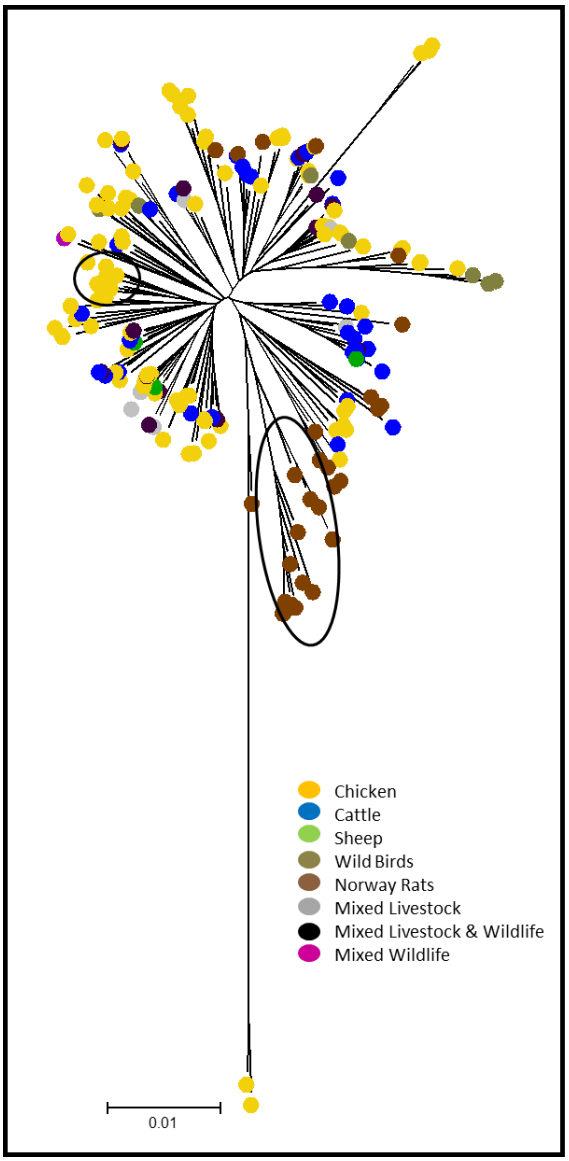


Figure 34. Neighbor Joining Tree constructed using concatenated nucleotide sequences at the seven aST(2) loci. Isolate sources denoted by colour scheme.

In total there were 186 unique aST(2)s and 28 incomplete aST(2)s. A total of 23 aST(2)s, representing 115 isolates (25.77%) were shared between two or more host sources (figure 35), of these six aST(2)s were shared between livestock sources only, 2 aSTs were shared between livestock and wild birds and 14 aSTs were shared between livestock and Norway rats. Only one aST(2) was present only in rats and wild birds. In each case of

sharing the proportion of chicken, cattle, wild birds or rats with a given shared aST(2) was less than 10% of the respective host population (figure 36), while shared aST(2)s were much more common in sheep, accounting for a total of 71.43% of sheep isolates overall, with a single shared aST(2) accounting for up to 28.57% of the sheep reference population.

Figure 35. Absolute frequency of alternative sequence type (2), by host source.

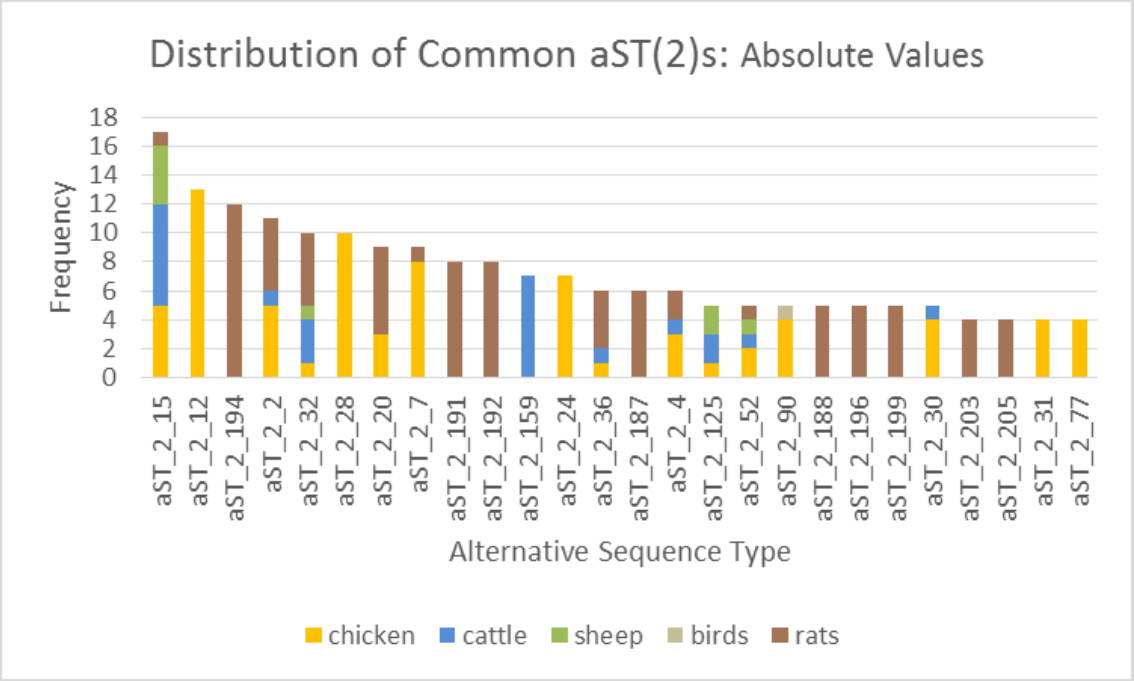


Figure 35. Bar chart showing the distribution of common aST(2) types.

Figure 36. Relative frequency of alternative sequence type (2), by host source.

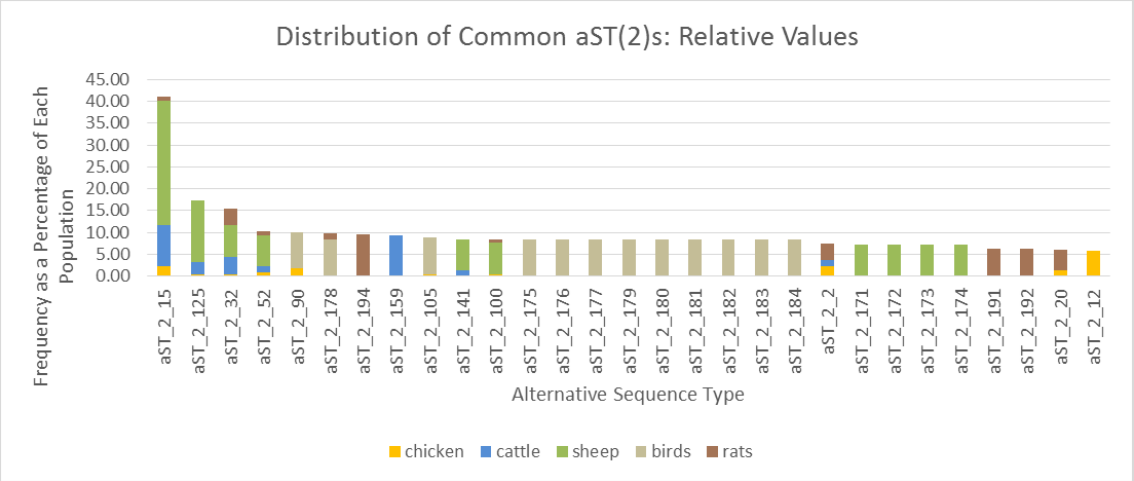


Figure 36. Bar chart showing the relative frequency of the common aST(2) types as a percentage of each population.

On average there were 81.14 alleles per aST(2) locus, ranging from 64 at the *cj0038c* locus to 97 at the *cj0036* locus. Figure 37 shows the percentage of alleles shared between each population at each locus. A total of 59.15% of isolates shared alleles from a common pool at the *ilvC* locus (figure 37g), while only 35.49% of isolates shared alleles at the *cj0038c* locus (figure 37c).

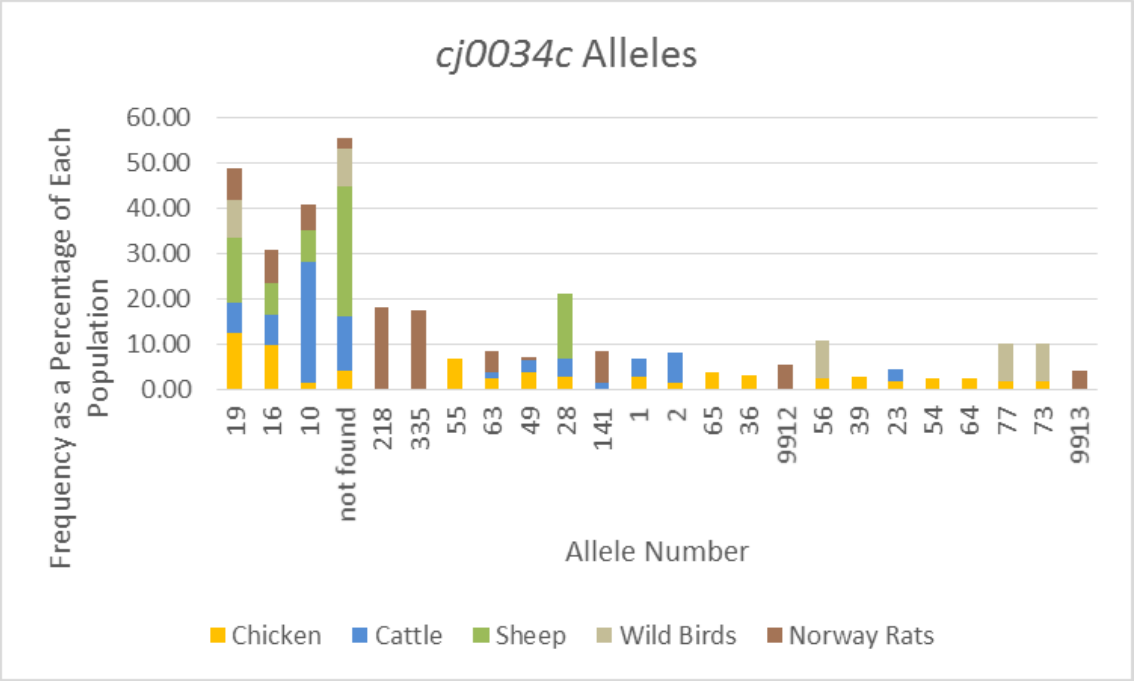


Figure 37a.

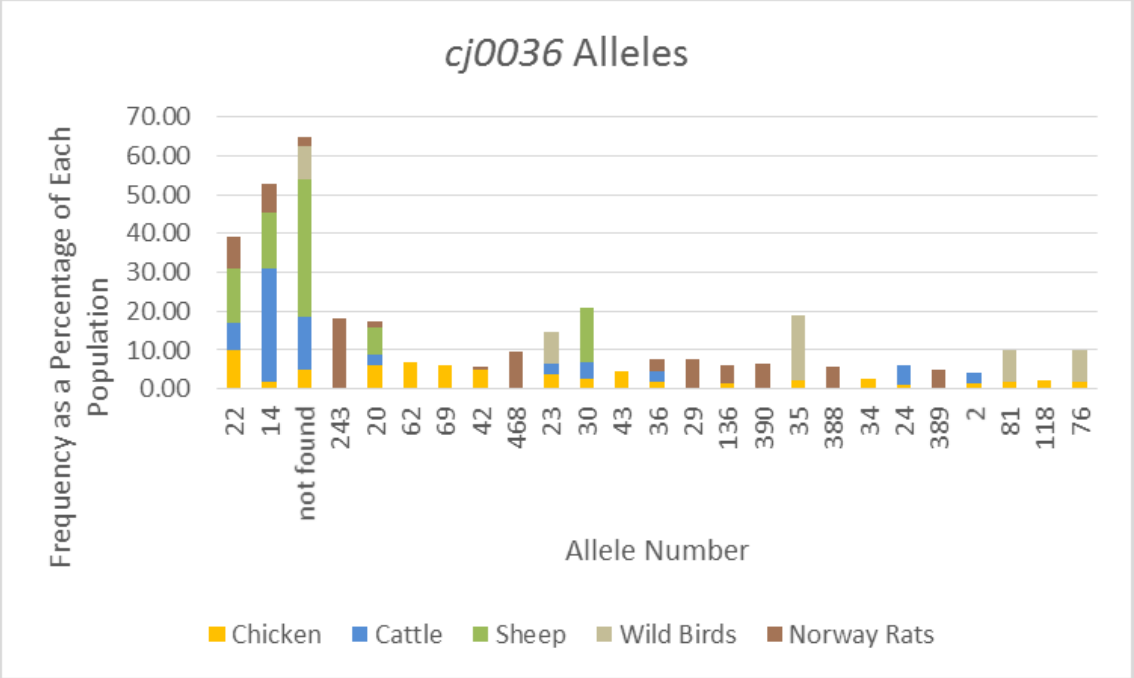


Figure 37b.

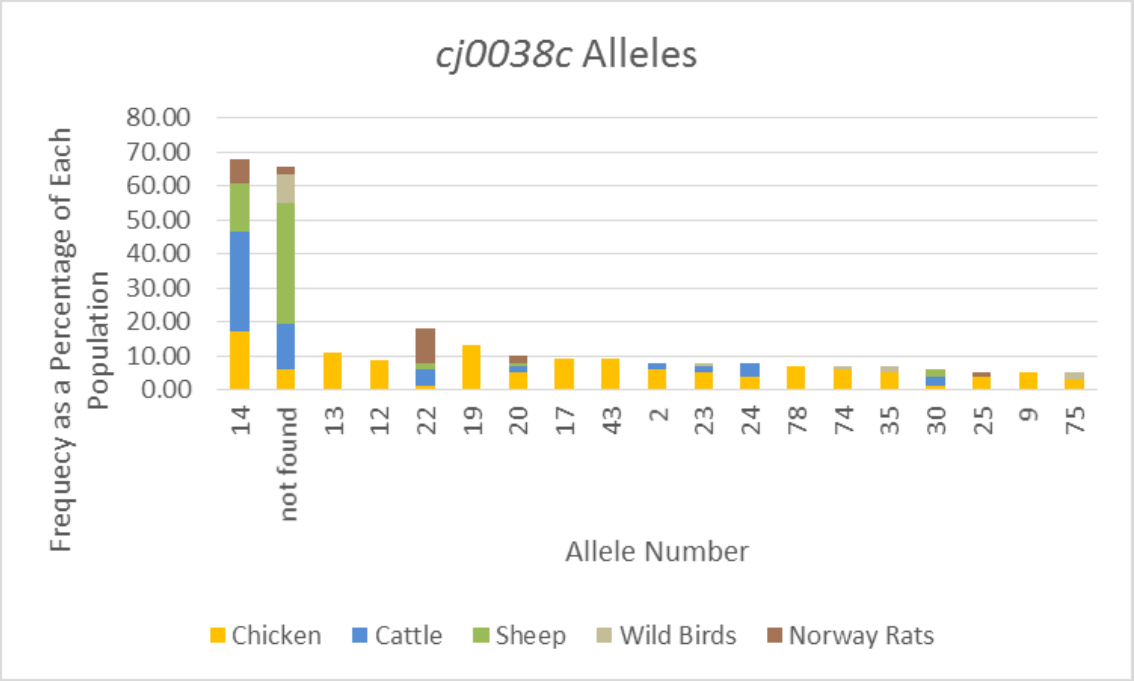


Figure 37c.

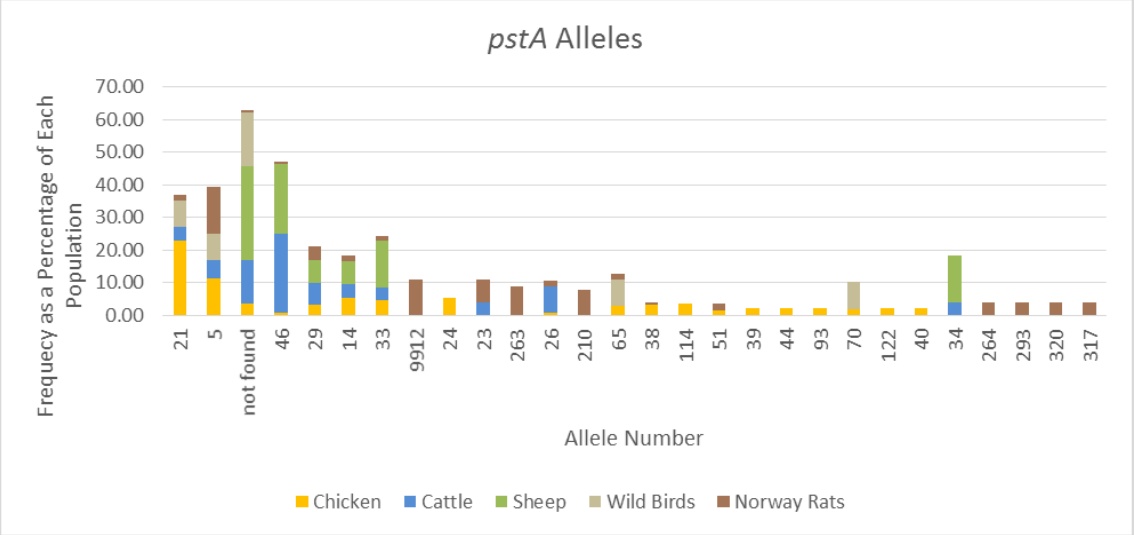


Figure 37d.

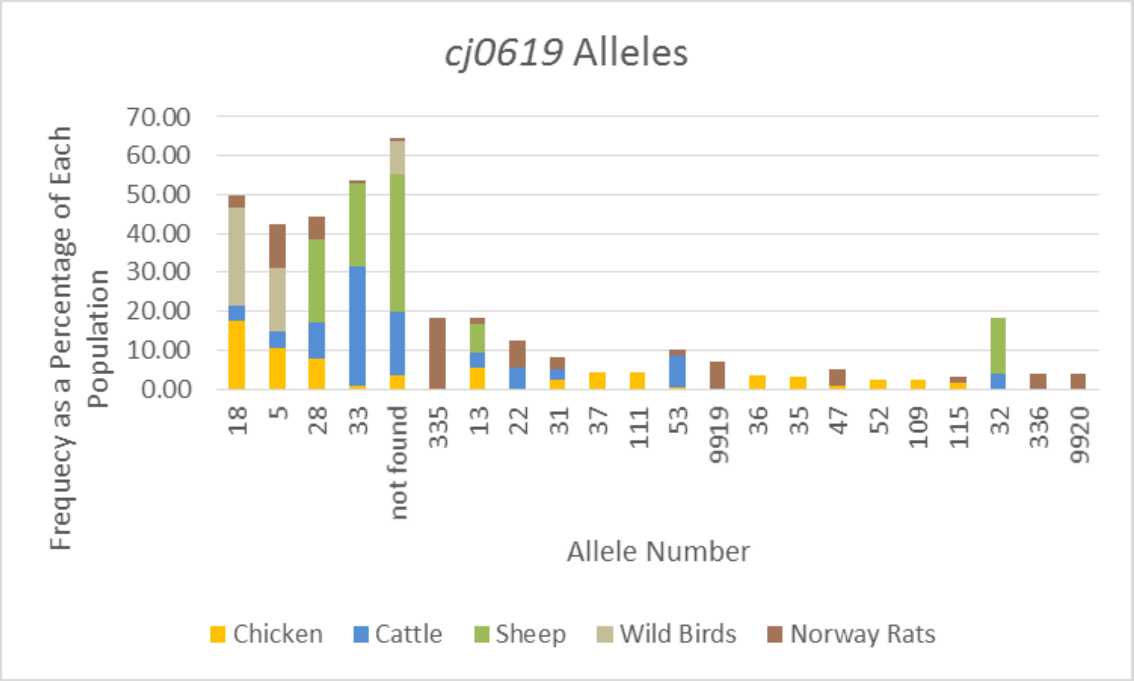


Figure 37e.

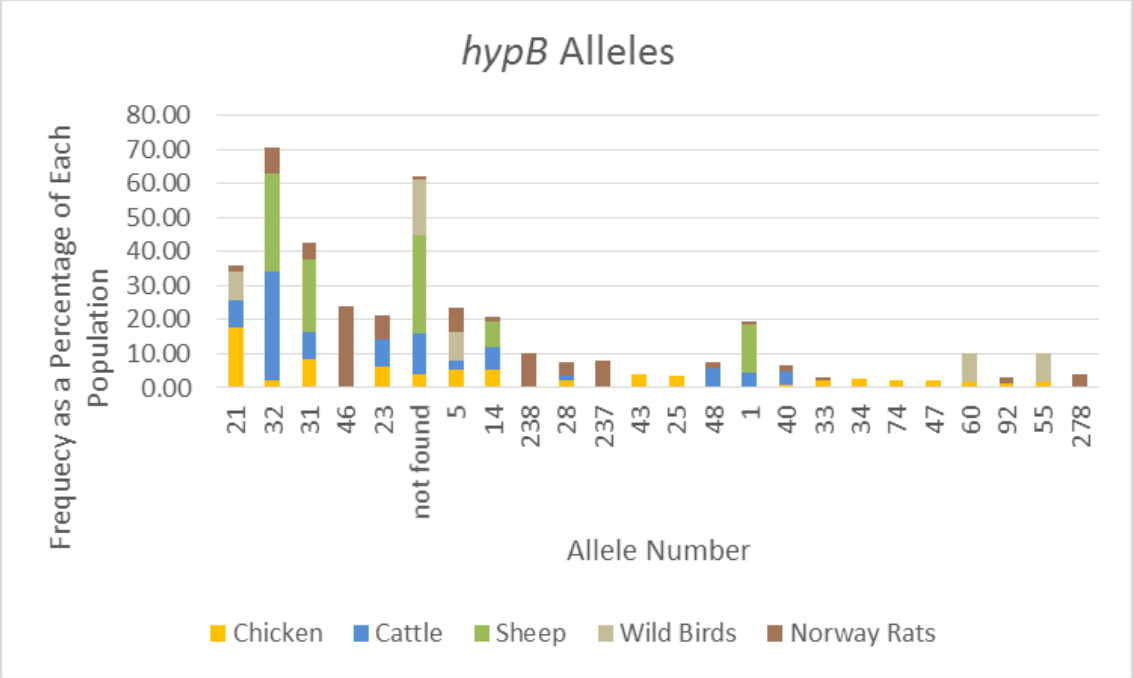


Figure 37f.

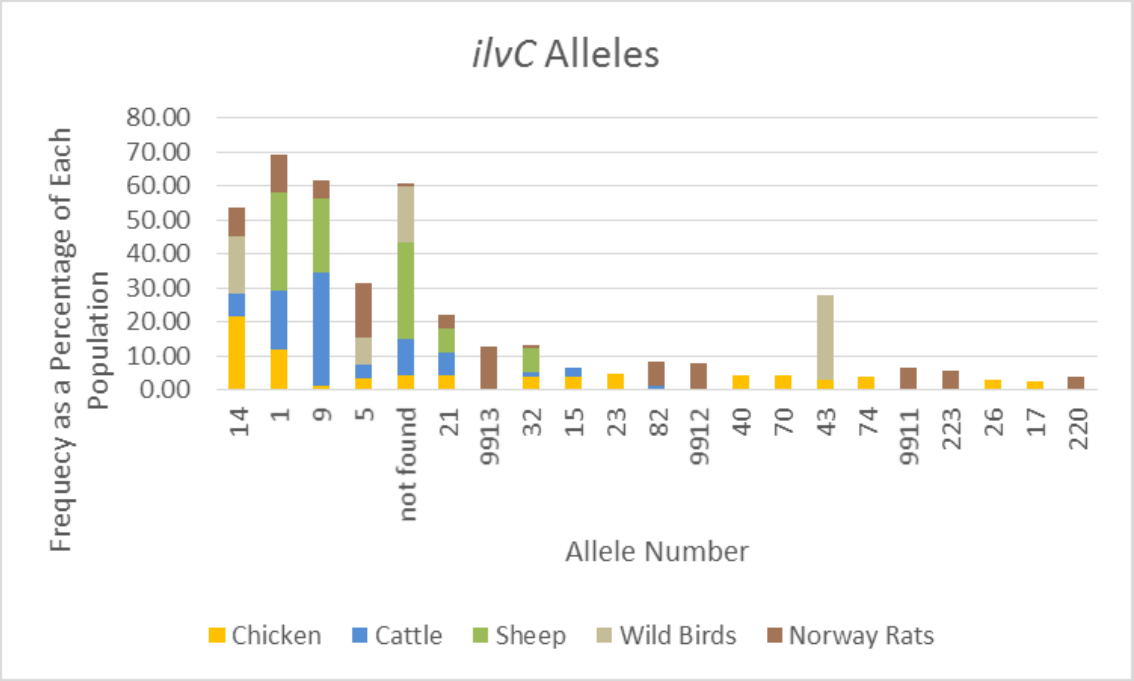


Figure 37g.

Figures 37a-g. Stacked bar charts showing the frequency of alleles at each of the aST(2) loci, as a percentage of each population.

STRUCTURE: MLST Self-Test Results

Self-testing of STRUCTURE using 147 'self-test' isolates of known source, against a reference set of 301 isolates was 62.56% correct overall. Self-attribution of chicken isolates was 53.42% correct, attribution of cattle isolates was 72.00% correct, 50.00% of wild bird isolates and 78.05% of rat isolates were correctly attributed. Attribution of isolates from sheep was the least successful, with only one correct attribution made (25.00%). 75% of sheep *Campylobacter* isolates were attributed to cattle.

On average, chicken, cattle and rat isolates had a higher probability of attribution to the correct source than an incorrect source, while isolates from sheep and wild birds were on average more likely to be attributed to an incorrect source (table 16).

Table 16. Average attribution probability of all isolates.

average attribution probability of all isolates						
		attributed source				
		chicken	**cattle**	**sheep**	**bird**	**rat**
actual source	**chicken**	0.56	0.22	0.04	0.04	0.15
	cattle	0.04	0.74	0.10	0.03	0.09
	sheep	0.02	0.72	0.03	0.02	0.01
	bird	0.12	0.08	0.17	0.39	0.25
	rat	0.05	0.17	0.02	0.03	0.73

Table 16. Results of attribution self-test, showing the average probability of attribution of isolates from each known source to each putative source, using MLST.

Of the 54 incorrectly attributed isolates, 16 (29.63%) belonged to ST-45 CC, and 12 isolates (22.22%) belonged to ST-21 CC, compared with 10.87% and 8.70% of correctly attributed isolates belonging to ST-45 CC and ST-21 CC respectively. Across all the source populations 10.76% of isolates were not attributed to a single host.

STRUCTURE: Alternative Sequence Typing Scheme 1 Self-Test Results

Attribution of isolates from known source, using the alternative sequence type (1) definitions was on average 59.86% correct overall. No sheep isolates were correctly attributed. 59.96% of chicken isolates were attributed to a chicken source, 60.00% of wild bird isolate were correctly attributed to a wild bird source, 72.00% of cattle isolates were correctly attributed, and 75.00% of rat isolates were correctly attributed. Overall, isolates from each host source apart from sheep had a higher probability of being

attributed to a correct, rather than an incorrect source (table 17). Across all the source populations 18.98% of isolates were not attributed to a single host.

Table 17. Average attribution probability of all isolates.

average attribution probability of all isolates						
		attributed source				
		chicken	**cattle**	**sheep**	**bird**	**rat**
actual source	**chicken**	0.58	0.19	0.02	0.03	0.18
	cattle	0.05	0.72	0.06	0.02	0.14
	sheep	0.28	0.48	0.57	0.05	0.05
	bird	0.09	0.04	0.16	0.57	0.14
	rat	0.04	0.18	0.03	0.04	0.72

Table 17. Results of attribution self-test, showing the average probability of attribution of isolates from each known source to each putative source, using aST(1).

STRUCTURE: Alternative Sequence Typing Scheme 2 Self-Test Results

Attribution of isolates from known source, using the aST(2) definitions was on average 63.27% correct overall. 64.38% of chicken isolates were correctly attributed to chicken, 56.00% of cattle isolates, 50.00% of wild bird isolates and 73.17% of rat isolates were attributed to the correct sources. No sheep isolates were correctly attributed.

On average, chicken, cattle and rat isolates showed a higher probability of being attributed to the correct rather than an incorrect source, while sheep and wild bird isolates were more likely to be attributed to an incorrect than a correct source (table 18). Across all the source populations 33.83% of isolates were not attributed to any single host population.

Table 18. Average attribution probability of all isolates.

average attribution probability of all isolates						
		attributed source				
		chicken	**cattle**	**sheep**	**bird**	**rat**
actual source	**chicken**	0.63	0.14	0.08	0.06	0.09
	cattle	0.13	0.61	0.11	0.04	0.11
	sheep	0.38	0.22	0.17	0.07	0.16
	bird	0.31	0.10	0.20	0.33	0.06
	rat	0.06	0.13	0.04	0.05	0.73

Table 18. Results of attribution self-test, showing the average probability of attribution of isolates from each known source to each putative source, using aST(2).

STRUCTURE: MLST Clinical Attribution Results

Attribution of 656 human clinical *C. jejuni* isolates to their host source using STRUCTURE, where any given host source with an attribution probability greater than 0.50 is accepted as the putative source of the corresponding human infection, estimates that 47.10% of clinical cases are attributable to chicken, 38.72% to cattle, 0.15% to sheep, 0.61% to wild birds and 9.60% attributable to Norway rats. Twenty-five (3.81%) of the human clinical isolates were unattributable. The average attribution probability of all isolates to chicken was 0.52, to cattle 0.35, to rats 0.08, to sheep 0.03 and to wild birds 0.01. Seventeen unattributed isolates belonged to eight clonal complexes, five unattributed isolates belonged to STs not assigned to any clonal complex and three unattributed isolates had incomplete MLST profiles and were not assigned to any ST.

Figure 38. Attribution probability of clinical isolates to host sources- MLST.

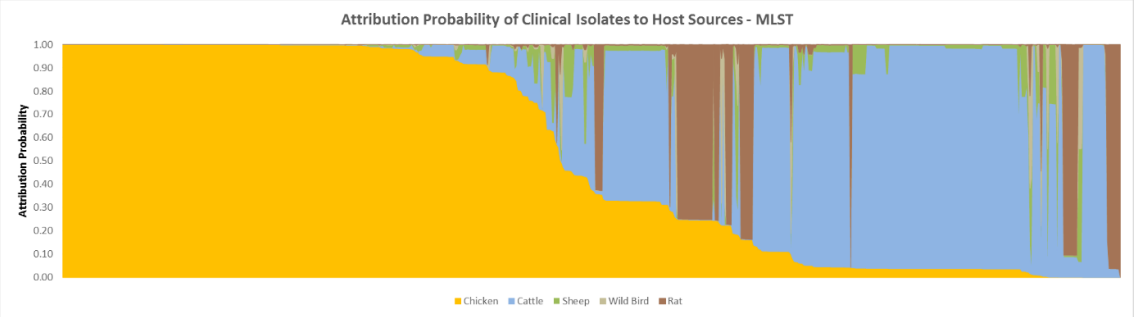


Figure 38. Attribution probability of clinical *Campylobacter* isolates to putative host source using MLST.

STRUCTURE: Alternative Sequence Typing Scheme 1 Clinical Attribution Results

Attribution of Clinical *C. jejuni* isolates using STRUCTURE with the newly defined aST(1) data in place of MLST data estimates that 56.44% of clinical cases are attributable to chicken, 31.29% attributable to cattle, 7.52% to Norway rats, 0.46% to sheep and 0.15% to wild birds. 4.14% of clinical isolates were unattributed, 15 out of 27 unattributed cases (55.56%) were split between two or more livestock hosts, and 11 unattributed cases (40.74%) split between one or more livestock sources and wild birds. A single isolate was attributed with equal probability to all five putative host sources. The overall average attribution probability of all isolates to each source closely correlated with the percentage of isolates attributed to the respective sources. The average attribution probability of all clinical cases to chickens was 0.57, to cattle was 0.32, to rats was 0.07, to sheep was 0.03, and to birds was 0.01.

Figure 39. Attribution probability of clinical isolates to host sources- aST(1).

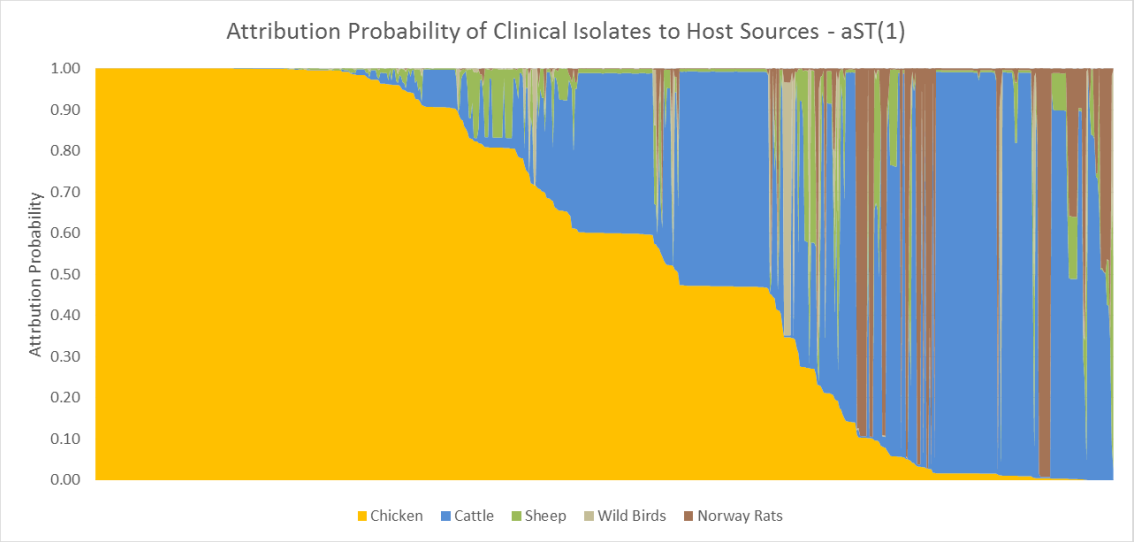


Figure 39. Attribution probability of clinical *Campylobacter* isolates to putative host source using aST(1).

STRUCTURE: Alternative Sequence Typing Scheme 2 Clinical Attribution Results

Attribution of Oxfordshire clinical *C. jejuni* isolates using aST(2) data attributes 58.69% of clinical cases to chicken, 10.37% to cattle, 2.44% to rats, 0.61% to wild birds and 0.30% to sheep (figure 40). 25.76% of isolates were not attributed to any single source, of the unattributed isolates 165 (97.63%) were split between livestock, wild birds and wild rats, and the remaining 4 isolates (2.37%) were split between livestock and rats.

Figure 40. Attribution probability of clinical isolates to host sources- aST(2).

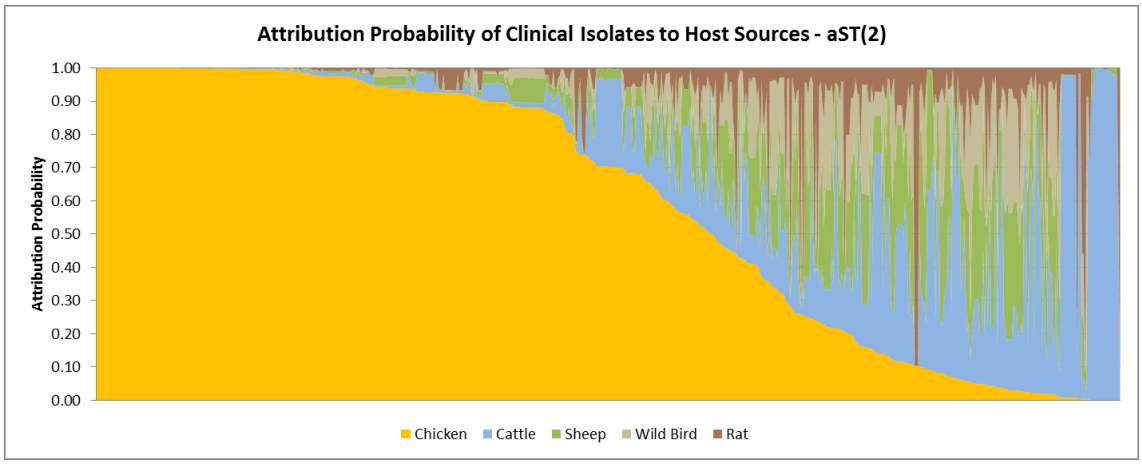


Figure 40. Attribution probability of clinical *Campylobacter* isolates to putative host source using aST(2).

Comparison Between MLST, aST(1) and aST(2) Attribution of Clinical Isolates

Figure 41. Comparison of attribution typing schemes.

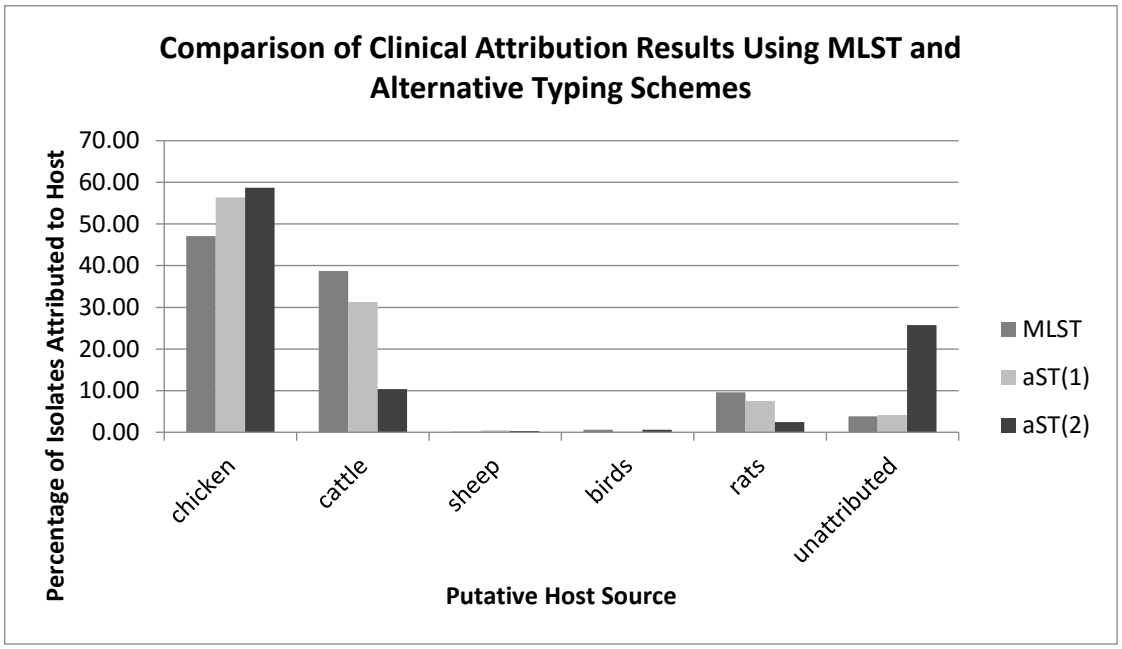


Figure 41. Comparison of the percentage of isolates attributed to each source by each typing method.

There was a general consistency between the three typing schemes in the rank order of the three most common putative sources of *Campylobacter jejuni* infection in humans. MLST attributed significantly more isolates to a cattle origin than did aST(2). Significantly more isolates were left unattributed by the aST(2) scheme than either MLST or aST(1). No other differences were significant.

Ciprofloxacin Resistance in Clinical Isolates

Of 656 human clinical *Campylobacter* isolates 36.68% are resistant to ciprofloxacin, as defined by the presence of the Thr_{86} mutation in the *gyrA* locus. The proportion of resistant isolates is unevenly distributed throughout the putative sources of infection.

MLST Attribution

Using conventional MLST to attribute disease cases to putative host sources 48.44% of chicken-attributed isolates, 30.93% of cattle-attributed isolates, 30.00% of unattributed isolates and 8.77% of rat-attributed isolates were ciprofloxacin resistant. Although percentage prevalence among sheep and wild bird-attributed isolates was high, at 100% and 50% respectively, the numbers of isolates attributed to these sources were so small that the resistance prevalence figures are somewhat meaningless.

The percentage prevalence of ciprofloxacin resistance by ST was examined for those STs with at least five isolates in the clinical test set (figure 42). Ciprofloxacin resistance ranged from 0% in ST-2030, ST-61, ST-22, ST-52, ST-2033, ST-267 and ST-273 to 100% in ST-464, ST-2135, ST-5136, ST-572 and ST-775. Ten STs had a ciprofloxacin resistance prevalence of greater than or equal to 40%, and of these all STs were attributed to livestock hosts (70% chicken, 30% cattle).

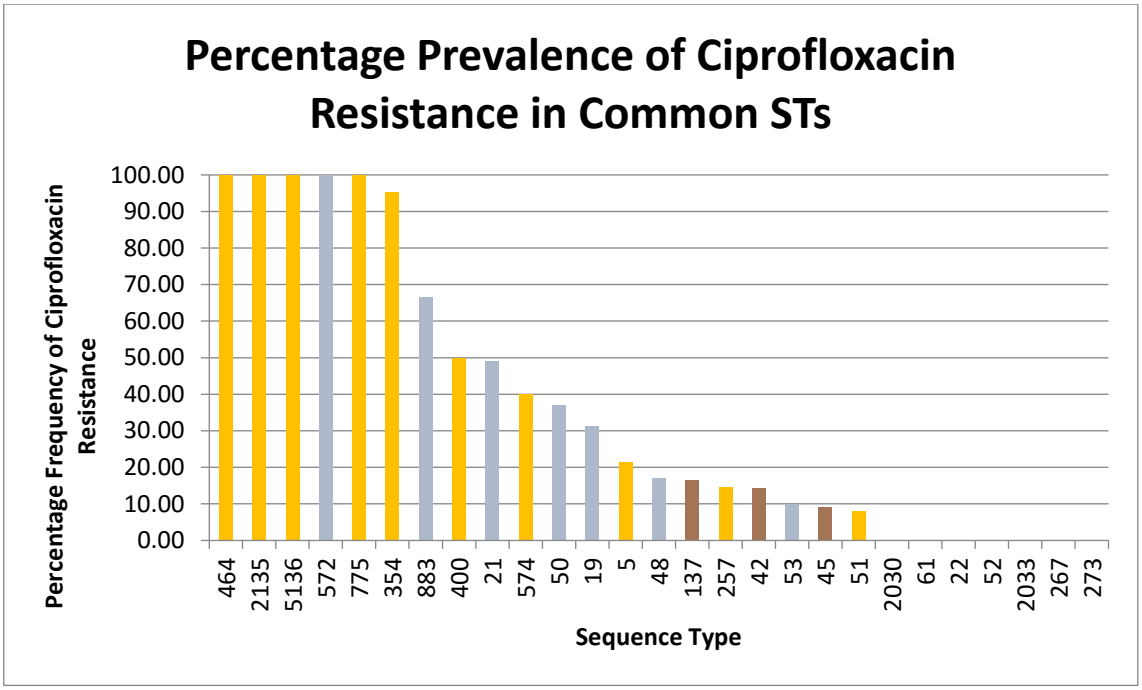


Figure 42. Percentage prevalence of ciprofloxacin resistant isolates, as defined by the presence of the Thr86 mutation, in common STs.

Alternative Sequence Typing Scheme 1

Using aST(1) definitions to attribute campylobacteriosis cases to putative host sources 42.52% of chicken-attributed isolates, 28.57% of cattle-attributed isolates and 13.16% of rat-attributed isolates were resistant to ciprofloxacin. The prevalence of ciprofloxacin resistance amongst unattributed isolates was 39.13%. Sheep-attributed and wild bird-attributed isolates displayed a prevalence of 33.33% and 0.00% ciprofloxacin resistance, respectively.

Alternative Sequence Typing Scheme 2

Using aST(2) to define and attribute human disease cases to putative host sources 35.36% of chicken-attributed isolates and 12.12% of cattle-attributed isolates were ciprofloxacin resistant. The ciprofloxacin resistance mutation was not present in any of

the Norway rat – attributed isolates. Ciprofloxacin resistance prevalence in the sheep and wild bird-attributed isolates was high at 50.00% and 25.00% respectively, but with extremely large confidence intervals as a result of the small number of isolates attributed to these sources. The prevalence of ciprofloxacin resistance in unattributed isolates was 50.34%.

Ciprofloxacin Resistance in Reference Isolates

The prevalence of ciprofloxacin resistance in the reference population differed by source (figure 43), with chickens displaying a significantly higher prevalence of resistance than all other sources, with the exception of sheep. The prevalence of resistance in wild birds and in Norway rats was 0.

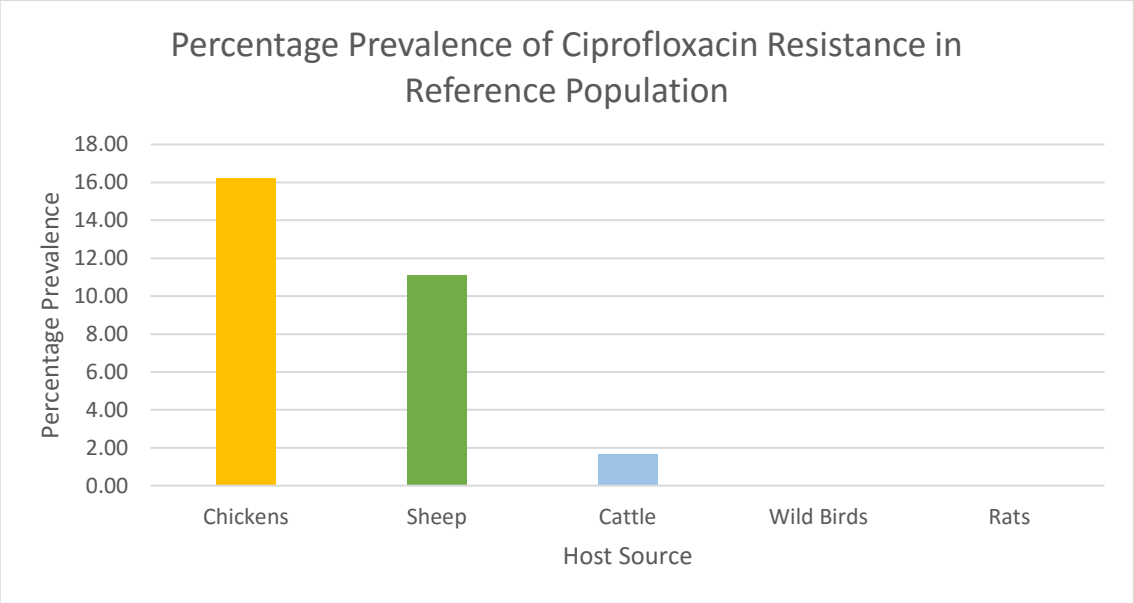


Figure 43. Percentage of ciprofloxacin resistant isolates in each of the reference populations.

6.3 Discussion and Conclusions

The selection of isolates for the reference set was dictated by the availability of whole genomes. The selection of isolates from a relatively small number of studies raises the question of how representative the isolates are. The availability of more data would doubtless improve the validity of the whole genome reference set but neighbour joining tree comparison of the reference chicken, cattle and sheep isolates with background MLST data demonstrates that the reference sets were not unrepresentative of the corresponding wider sampled populations (figure 5). The Norway rat isolates represented the entire collection of Norway rat *Campylobacter jejuni* sampled to date. The Norway rat collection was diverse and represented a number of clades never encountered before in the sampling of animals or human disease, the existence of greater unsampled rat *Campylobacter* diversity remains a possibility but if this hypothetical diversity represented one or more new clades it would not have an effect on the attribution of human clinical cases. The low availability of wild bird genomes means that 'wild birds' was taken as a generic term to cover wild ducks (n = 5); starlings (n = 4); geese (n = 5); assorted wild birds (n =2). This is not ideal but there are currently too few whole genome sequences from wild bird isolates to treat them separately.

Human clinical isolates were taken from a coherent sample of hospital and community isolates representing the known campylobacteriosis burden from across Oxfordshire. The representativeness of Oxfordshire *Campylobacter* isolates of the disease across the rest of the UK has previously been established[156,157].

Neighbour-Joining trees, estimating the relatedness between individual isolates, as defined by nucleotide sequences at the seven MLST loci (figure 26) demonstrate that

while only a minority of STs can be considered genuine 'multi-host' strains, and the majority of STs within this dataset is unique to a given host there is little separation of sources according to lineage. Separate host-affiliated STs appear side-by-side on the same branch of the estimated phylogeny, and STs isolated from a single host are distributed throughout the putative lineages. Similarly, Analysis of Molecular Variance reveals that only a small proportion of the molecular diversity within the MLST loci of the reference population is attributable to host source, with high within-source diversity. Although a large proportion of STs is unique to each of the given sources in the reference set there is a high degree of sharing of alleles between sources, with unique STs usually being a new combination of known alleles.

This pattern of high within-population diversity and extensive sharing of alleles between populations is an obstacle to the use of MLST data for accurate attribution and raises the question of what drives host-ST affiliation.

STs based on seven housekeeping loci represent only a small fraction of the genome. Using whole genome sequence data provides higher resolution than MLST, separating STs into smaller clusters. Variation across the genome, outwith the MLST loci could provide an explanation for host association in *Campylobacter;* however wgMLST clustering did not correlate well with host source. Looking at each of the variable loci individually demonstrated that there was a varying degree of plasticity and between-population diversity across the genome but there were no loci at which alleles separate well by host source. For this reason STRUCTURE attribution using aSTs did not significantly improve the reliability of attribution compared with conventional MLST-based attribution.

In the absence of evidence for host-adapted lineages the relationship between sequence type and host source is remarkable. Small but statistically significant F_{ST} values between each pair of the host populations, with the exception of cattle and sheep, point to some degree of genetic isolation between populations with a substantial but not infinite level of genetic exchange between the disparate host species. Furthermore, when the reference set was divided into two groups of 'Agricultural' and 'Non-Agricultural' isolates F_{ST} values were higher than when all populations were considered as a single group. This result would indicate that there is a higher degree of gene flow within the Agricultural group than between the Agricultural and Non-Agricultural groups. The Migration Matrix Model[218] and the Asymmetric Islands Model[75] provide a good description for this scenario. Coherent populations and a statistical imbalance of sequence types are maintained by certain obstacles which prevent panmixia; however these obstacles are not insuperable and genetic mixing of *Campylobacter* takes place to a greater or lesser extent between each pair of host sources, leading to distinctive but overlapping samples from the same pool of *Campylobacter* types, which are not well differentiated using a gene-by-gene approach. Recent work using a Genome-Wide Association Study (GWAS) approach has identified genomic elements or words, 30 base pairs in length, which are almost universally present in *Campylobacter* isolated from a bovine source, but which are also present in a minority of chicken isolates[221]. The GWAS approach, looking at allelic variation above the level of single nucleotides but below the level of whole genes would appear to be a promising approach to improve the accuracy of host source attribution.

Self-attribution tests yielded similar results for MLST and each of the trial alternative typing schemes. The newly defined aST(2) was marginally, but not significantly, more accurate on average with 63.27% correct attribution overall, compared with 62.56% accuracy of MLST and 59.86% accuracy with aST(1). The rank order of correct attributions by host varied slightly between the three schemes. Using MLST rat isolates were correctly attributed (78.05%) more than any other source, followed by cattle isolates (72.00%), chicken isolates (53.42%), wild bird isolates (50.00%) and sheep (25.00%). Using aST(1) rat isolates were also the most successfully attributed (75.00%), followed by cattle isolates (72.00%), wild birds (60.00%) and chickens (59.96%). No sheep isolates were correctly attributed using aST(1). Using aST(2) rat isolates were again the most accurately attributed (73.17%), followed by chicken isolates (64.38%), cattle isolates (56.00%), and wild bird isolates (50.00%). No sheep isolates were correctly attributed using aST(2).

The successful self-attribution of the majority of Norway rat isolates is easily explained by the distinctive nature of the population. While there was a number of shared STs between rats and other sources the high proportion of very distinctive strains (figures 26, 30, 34) allowed easy differentiation by the attribution model. It is predicted that the self-attribution of wild bird isolates would be improved by using a larger and more coherent reference set. 'Wild birds' was the only reference population that represented more than one host species, and the isolates were few in number. For the purpose of this study the available genomes from wild bird *Campylobacter jejuni* isolates were used as a method of highlighting potentially host-adapted genes through large-scale population genetic analyses. The poor self-attribution of sheep isolates was

symptomatic of their similarity with bovine isolates, as demonstrated by F_{ST} values, this effect was exacerbated by the small sheep *Campylobacter* population sample.

Self-attribution of chicken isolates was poor across all three typing schemes, with up to 46.48% of chicken isolates being attributed to a non-chicken source. This is suggestive of a chicken *Campylobacter jejuni* population which partially resembles other populations. A possible hypothesis based on this finding is that a super-population of *Campylobacter jejuni* exists in chickens, with other livestock species most commonly colonised by sub-samples of this super-population. If this were the case attribution only of human clinical isolates to putative host source would represent a major underestimate of the contribution of chicken to the *Campylobacter* public health problem.

The proportion of unattributable isolates was 10.76% for MLST, 18.98% for aST(1) and 33.83% for aST(2). The inability of aST(2) to attribute such a large proportion of isolates of known origin should not be misinterpreted as a failure of the typing scheme, rather in the context of the other evidence from this study it is suggestive of poor differentiation between *Campylobacter* populations from different sources, particularly from different livestock sources. Recent research focusing on ST-21 using whole genome sequencing and PCR sequencing of target genes from *C. jejuni* isolates from various host sources demonstrates high genomic plasticity but little or no correlation between genomic traits and host source[219].

The results of self-attribution tests should inform the interpretation of the results of attribution of clinical isolates to putative source. Based on the assumption of

approximately 62% accuracy of attribution to source on average across the three typing schemes, the results of the clinical attribution using both MLST and the alternative typing schemes are plausible, attributing the majority of cases to chicken, with a significant minority attributed to cattle and a far lesser proportion attributed to sheep, Norway rats and wild birds. Recent research based on MLST attribution has demonstrated that wild birds may be an important source of *Campylobacter* amongst specific at-risk groups[222]. The proportion of putative wild bird-type cases in this study is in line with previous findings. Norway rats as a reservoir or vector of *Campylobacter* have not previously been studied but the MLST attribution of 9.60% of human clinical cases from the year 2010 – 2011 seems unlikely, although Norway rats can live alongside human habitation it would seem implausible that human – rat interaction in Oxfordshire is sufficient to cause almost 1 in 10 campylobacteriosis cases across the county. Inspection of the isolates attributed by MLST to rats reveals that 37 isolates (58.37%) belonged to the multi-host ST-45 CC. The remainder of rat-attributed human isolates belonged to STs generally associated with livestock, which appear to occur sporadically in Norway rats. Similarly there is a question mark over the extent of the role of cattle in *Campylobacter* transmission. MLST, aST(1) and aST(2)-based attribution linked 38.72%, 31.29% and 10.37%, respectively, of human cases with cattle. While these results are in line with previous findings[36,74] they are somewhat at odds with the results of studies of the prevalence of *Campylobacter* on retail beef[223,224]. It is possible that a percentage of cattle-attributed human *Campylobacter* infections result from imperfect pasteurisation of milk[225] rather than consumption of beef. Amongst the 254 disease isolates attributed by MLST to cattle, 145 isolates belong to the multi-host ST-21 CC. Where *Campylobacter* types are shared by multiple sources it is difficult to discern the actual source of

infection. ST-61, for example, has been isolated from chicken, cattle, sheep and a Norway rat; therefore it is conceivable that contact with the faeces of any of these hosts may result in an infection with ST 61. If it is assumed that on a particular occasion an ST-61 campylobacteriosis case is known to be the direct result of contact with Norway rat faeces, but that carriage of ST-61 in the rat was transient, and was itself the result of contact with cattle faeces, from a public health perspective culling of rats as an occasional transient host of ST-61 would be unlikely to reduce the burden of ST-61 disease when the reservoir, cattle, still harboured and shed bacteria. In order to assess the utility of Bayesian model-based attribution using either MLST or an alternative typing scheme the purpose of attribution should be better defined. Although campylobacteriosis is not considered to be a vector-borne disease the notion of transient hosts as vectors may be a useful one to distinguish a short term host, for example the Norway rat with ST-61, from the natural reservoir, which would appear to be cattle in the case of ST-61.

The prevalence of ciprofloxacin resistance, as defined by the presence of the Thr_{86} mutation in the *gyrA* locus, is 34.66% in the human clinical isolates. Resistance at this level is likely to represent therapeutic failure of ciprofloxacin.

At 44.40%, the prevalence of ciprofloxacin resistance in isolates attributed to chickens is significantly higher than that of any other host population. This is consistent with previous studies linking ciprofloxacin resistance in human infection with antibiotic use in chickens, however considerable caution should be exercised in drawing conclusions from this. Several previous studies[117,157] have demonstrated a correlation between clonal complex and prevalence of antimicrobial resistance. In this study members of the

ST-353 CC, ST-354 CC and ST-464 CC have been attributed to chickens. On-farm studies have identified these complexes as being highly prevalent in chickens and they have also been linked with resistance to ciprofloxacin[157]. What is not currently understood is whether ciprofloxacin resistance develops as an inevitable consequence of long-term chicken gut colonisation, or whether there are different factors leading separately to high prevalence in chickens and a tendency to develop ciprofloxacin resistance within these clonal complexes.

Chapter Seven: Overall Summary and Conclusions

Summary of Findings

The specific theme of this thesis is the molecular epidemiology and ecology of *Campylobacter* in the UK. Each chapter has addressed a distinct subtopic within this overall theme, covering aspects of *Campylobacter* epidemiology and ecology from the clinical perspective and working back towards the farm environment. The important issue of antimicrobial resistance has been visited repeatedly throughout this thesis, considering the epidemiology of resistance in the clinical setting, in retail poultry and in *Campylobacter* reservoirs and the natural environment. Underpinning each chapter is the question of how best to use the available data, and the most appropriate level of resolution to answer each of the specific research questions.

Chapter three demonstrated a high level of consistency between PCR-derived and WGS-derived MLST data, effectively validating both the increasing use of WGS MLST data in clinical and public health settings, and comparisons between historic and contemporary datasets. Ciprofloxacin resistance was used as an example of a phenotypic trait which can be inferred from WGS data. The same principle can be applied to other traits of public health importance, with a known genomic determiner. The sequencing of whole genomes, rather than selected loci allows for forward compatibility as genes of interest are identified through laboratory experiments.

Chapter four reports a study of antimicrobial resistance using MLST data and phenotypic resistance data. Here MLST data provided sufficient resolution to examine the relationships between antimicrobial resistance profile and putative lineage,

demonstrating that resistance phenotypes were not confined to specific lineages. Horizontal gene transfer was implicated in the spread of antimicrobial resistance genes, although stochastic mutations, which persist in an antibiotic-rich environment, were not ruled out. The high prevalence of retail poultry isolates of multiple different lineages with phenotypic resistance to a number of antibiotics is compatible with the popular hypothesis that indiscriminate antibiotic usage in poultry is responsible for high levels of resistance in clinical isolates. The findings of the attribution study, reported in chapter six, whereby chicken-attributed clinical isolates displayed higher levels of ciprofloxacin resistance than isolates attributed to any other source were also largely compatible with this hypothesis, although the generally low prevalence of ciprofloxacin resistance in the highly chicken-associated ST-257 CC suggested that the relationship is not a simple one. This thesis does not attempt an in-depth study of the causal relationship between the use of antibiotics in treating livestock and antibiotic resistance in clinical isolates, however it is suggested that whole genome sequence data would facilitate such a study, both by improving estimates of population genetic structuring and by allowing highly accurate presumptive identification of resistant isolates based on genotype rather than phenotype.

Empirical testing of antimicrobial resistance in the laboratory, as carried out in the retail poultry survey study, is time consuming and the problem of inconsistent protocols was demonstrated by the changing ampicillin breakpoints used throughout the study period. Genomic data, if it were available, would help to mitigate against the effects of changing protocols.

Chapter five reports a study of *Campylobacter* in wild Norway rats. MLST data alone provided an interesting insight into the genetic diversity of rat *Campylobacter* and suggested a pattern of a degree of 'sharing' of STs prominent in livestock and clinical samples and a high degree of uniqueness, with novel STs. A high proportion of novel STs were assigned to the ST-45 CC. Only with whole genome sequence data did it become apparent that there was a highly distinctive clade of rat-associated *Campylobacter jejuni* types which formed a subset of the ST-45 CC. This finding demonstrated that MLST data alone may not be sufficient for the study of more complex evolutionary relationships below the clonal complex level.

The distinctive rat *Campylobacter jejuni* genome collection was used alongside wild bird *C. jejuni* genomes in chapter six to provide an outgroup for comparison with genomes from livestock *Campylobacter* isolates to identify genes which may be host associated, in order to develop a typing scheme which would enable more accurate attribution of human clinical isolates to their putative host species. Genomic data, analysed using the gene-by-gene approach, provided higher resolution to differentiate closely related isolates but they did not improve the accuracy of attribution. This finding was somewhat counterintuitive, it would be expected that more data would lead to better information.

Population genetic analyses using genomic data revealed extremely low differentiation between isolates from agricultural sources, suggesting that *Campylobacter jejuni* in chickens, cattle and sheep were essentially alternative subsamples of the same population of livestock strains. WGS did not improve attribution but increased understanding of ecology of *Campylobacter jejuni*.

The findings of the attribution study lent support to the hypothesis that ecological barriers are one of the primary forces shaping the population genetic structure of *Campylobacter jejuni*. A common pool of *C. jejuni* genotypes spanned the farmed ruminant and chicken populations studied here, despite evident differences in host biology and physiology. Conversely the physiological similarities between farmed chickens and wild birds did not result in substantial overlap between *Campylobacter* STs. It is suggested that the farm environment creates an artificial ecological niche which sustains a population of livestock-associated *Campylobacter*.

The findings relating to antimicrobial resistance reported in chapters three, four and six were suggestive of a correlation between the occurrence of *Campylobacter* in livestock and the development or proliferation of resistance. Surveillance of human clinical cases in Oxfordshire revealed that three 'chicken-associated' STs (ST-464, 353 and 574) displayed a higher than expected prevalence of resistance to ciprofloxacin. This was consistent with recent published findings on a longitudinal study in Oxfordshire[157].

Chapter four, looking specifically at retail poultry identified a high overall prevalence of resistance to antimicrobial substances at 38%, suggesting that poultry meat could be a major source of drug-resistant *Campylobacter* entering the food chain.

Similarly, chapter six identified an excess of ciprofloxacin resistant genotypes among clinical isolates attributed to livestock, particularly chicken. Reference isolates from wildlife sources displayed 100% sensitivity to ciprofloxacin, while the highest prevalence of resistance was among chicken reference isolates (16.22%).

These findings indicate the importance of monitoring the prevalence of both *Campylobacter* carriage and of antimicrobial resistance in livestock populations.

Limitations

The studies reported in this thesis were carried out at a time of transition between Sanger sequencing of fragments of the *Campylobacter* genome and large-scale whole genome sequencing. Inevitably this lead to certain deficiencies with the datasets. Most notably the availability of *Campylobacter* isolates from animal sources was not ideal, representing two layers of bias. Unlike the Oxfordshire clinical campylobacteriosis study which sampled isolates from all laboratory-confirmed cases of *Campylobacter* in humans within the catchment area; studies of *Campylobacter* in livestock and wild animals are necessarily selective, furthermore the subsequent genome-sequencing of isolates from animal sources is largely based on specific interests or research questions, and is not an unbiased sample of the diversity of *Campylobacter* within animal populations. Nevertheless, the available data have been used with caution to develop and test new methods for the attribution of disease cases to host source, and to consider the ecology of *C. jejuni* and *C. coli*.

Coinfection with, or concurrent carriage of, multiple *Campylobacter* strains has not been considered in great depth within this thesis. The conventional microbiological methods used within these studies and in the routine diagnostic environment essentially isolate from each specimen the dominant or most highly shed *Campylobacter* strain, which thrives in laboratory conditions. Advances in genomic technologies in the future could enable the identification and relative quantification of competing bacterial strains within a single specimen.

Future Directions

As the field of bacterial genomics continues to develop, the study of the concepts within this thesis will continue to evolve. A study of antimicrobial resistance in *Campylobacter* isolates from farm to fork, using a large and well-sampled genomic dataset is a priority for future work in terms of public health. While various studies focus on resistance at different points along the food chain[117,226,227], there is a need for a systematic and joined-up investigation, aimed at identifying the specific risk factors for the acquisition and transfer of resistance genes and the proliferation of resistant strains at each stage.

While rats have not been implicated as important *Campylobacter* reservoirs or vectors, their relationship with the *Campylobacter* genus is interesting from the evolutionary and ecological perspective. A more in-depth study of *Campylobacter* isolated from rats, using a comparative genomics approach, for example using gene-finding software such as PROKKA[228] to identify genes which are not present in livestock *Campylobacter*, might provide a valuable insight into the host-parasite relationship in the *Campylobacter* genus.

Ongoing surveillance of *Campylobacter*, and particularly of the prevalence of antimicrobial resistance is imperative. The ability of *Campylobacter* to acquire resistance genes from mobile elements and to incorporate them into the chromosome[93] demonstrates the importance of genomic surveillance of resistance, rather than simply phenotypic surveillance, making possible the study of acquisition and transmission of resistance genes.

References

1. van Vliet, A. H. & Ketley, J. M. Pathogenesis of enteric Campylobacter infection. *Symp. Ser. Soc. Appl. Microbiol.* 45S–56S (2001).
2. Line, J., Hiett, K., Guard, J. & Seal, B. Temperature affects sole carbon utilization patterns of Campylobacter coli 49941. *Curr. Microbiol.* **62,** 821–5 (2011).
3. Stahl, M., Butcher, J. & Stintzi, A. Nutrient acquisition and metabolism by Campylobacter jejuni. *Front. Cell. Infect. Microbiol.* **2,** 5 (2012).
4. Trust, T. J. *et al.* Phylogenetic and molecular characterization of a 23S rRNA gene positions the genus Campylobacter in the epsilon subdivision of the Proteobacteria and shows that the presence of transcribed spacers is common in Campylobacter spp. *J. Bacteriol.* **176,** 4597–609 (1994).
5. On, S. L. Taxonomy of Campylobacter, Arcobacter, Helicobacter and related bacteria: current status, future prospects and immediate concerns. *Symp. Ser. Soc. Appl. Microbiol.* 1S–15S (2001).
6. Tam, C. C. *et al.* Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut* **61,** 69–77 (2012).
7. Roberts, J. A. *et al.* The study of infectious intestinal disease in England: socio-economic impact. *Epidemiol. Infect.* **130,** 1–11 (2003).
8. Acheson, D. & Allos, B. M. Campylobacter jejuni Infections: Update on Emerging Issues and Trends. *Clin. Infect. Dis.* **32,** 1201–1206 (2001).
9. Kuroki, S. *et al.* Guillain-Barré syndrome associated with Campylobacter infection. *Pediatr. Infect. Dis. J.* **10,** 149–51 (1991).
10. Nachamkin, I., Allos, B. M. & Ho, T. Campylobacter species and Guillain-Barré syndrome. *Clin. Microbiol. Rev.* **11,** 555–67 (1998).
11. Friedman, C. R. *et al.* Risk factors for sporadic Campylobacter infection in the United States: A case-control study in FoodNet sites. *Clin. Infect. Dis.* **38 Suppl 3,** S285-96 (2004).
12. Bender, J. B. *et al.* Surveillance by molecular subtype for Escherichia coli O157:H7 infections in Minnesota by molecular subtyping. *N. Engl. J. Med.* **337,** 388–94 (1997).
13. Bender, J. B. *et al.* Use of molecular subtyping in surveillance for Salmonella enterica serotype typhimurium. *N. Engl. J. Med.* **344,** 189–95 (2001).
14. Olsen, S. J. *et al.* Multistate outbreak of Listeria monocytogenes infection linked to delicatessen turkey meat. *Clin. Infect. Dis.* **40,** 962–7 (2005).
15. Raymond, B., Wyres, K. L., Sheppard, S. K., Ellis, R. J. & Bonsall, M. B. Environmental factors determining the epidemiology and population genetic structure of the Bacillus cereus group in the field. *PLoS Pathog.* **6,** e1000905 (2010).
16. Achtman, M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu. Rev. Microbiol.* **62,** 53–70 (2008).
17. Maiden, M. C. *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 3140–5 (1998).
18. Wilson, D. J. *et al.* Rapid evolution and the importance of recombination to the gastroenteric pathogen Campylobacter jejuni. *Mol. Biol. Evol.* **26,** 385–97 (2009).
19. McKay, D., Fletcher, J., Cooper, P. & Thomson-Carter, F. M. Comparison of two methods for serotyping Campylobacter spp. *J. Clin. Microbiol.* **39,** 1917–21 (2001).
20. Frost, J. A., Oza, A. N., Thwaites, R. T. & Rowe, B. Serotyping scheme for Campylobacter jejuni and Campylobacter coli based on direct agglutination of heat-stable antigens. *J. Clin. Microbiol.* **36,** 335–9 (1998).

21. Taboada, E. N., Clark, C. G., Sproston, E. L. & Carrillo, C. D. Current methods for molecular typing of Campylobacter species. *J. Microbiol. Methods* **95,** 24–31 (2013).

22. Ribot, E. M., Fitzgerald, C., Kubota, K., Swaminathan, B. & Barrett, T. J. Rapid pulsed-field gel electrophoresis protocol for subtyping of Campylobacter jejuni. *J. Clin. Microbiol.* **39,** 1889–94 (2001).

23. Melero, B., Juntunen, P., Hänninen, M.-L., Jaime, I. & Rovira, J. Tracing Campylobacter jejuni strains along the poultry meat production chain from farm to retail by pulsed-field gel electrophoresis, and the antimicrobial resistance of isolates. *Food Microbiol.* **32,** 124–8 (2012).

24. Miller, R. S. *et al.* DNA identification and characterization of Campylobacter jejuni and Campylobacter coli isolated from caecal samples of chickens in Grenada. *J. Appl. Microbiol.* **108,** 1041–9 (2010).

25. Sahin, O. *et al.* Molecular evidence for zoonotic transmission of an emergent, highly pathogenic Campylobacter jejuni clone in the United States. *J. Clin. Microbiol.* **50,** 680–7 (2012).

26. Champion, O. L., Best, E. L. & Frost, J. A. Comparison of pulsed-field gel electrophoresis and amplified fragment length polymorphism techniques for investigating outbreaks of enteritis due to campylobacters. *J. Clin. Microbiol.* **40,** 2263–5 (2002).

27. Nachamkin, I., Bohachick, K. & Patton, C. M. Flagellin gene typing of Campylobacter jejuni by restriction fragment length polymorphism analysis. *J. Clin. Microbiol.* **31,** 1531–6 (1993).

28. Meinersmann, R. J., Hiett, K. L. & Tarplay, A. Cloning of an outer membrane protein gene from Campylobacter jejuni. *Curr. Microbiol.* **34,** 360–6 (1997).

29. Hannon, S. J. *et al.* Genomics-based molecular epidemiology of Campylobacter jejuni isolates from feedlot cattle and from people in Alberta, Canada. *J. Clin. Microbiol.* **47,** 410–20 (2009).

30. Parkhill, J. *et al.* The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. *Nature* **403,** 665–8 (2000).

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

32. Zhang, M. *et al.* Molecular typing and antimicrobial susceptibility profiles of Campylobacter jejuni isolates from north China. *J. Med. Microbiol.* **59,** 1171–7 (2010).

33. Zhou, Y. *et al.* Comprehensive genomic characterization of campylobacter genus reveals some underlying mechanisms for its genomic diversification. *PLoS One* **8,** e70241 (2013).

34. Méric, G. *et al.* A reference pan-genome approach to comparative bacterial genomics: identification of novel epidemiological markers in pathogenic Campylobacter. *PLoS One* **9,** e92798 (2014).

35. Sheppard, S. K. *et al.* Evolution of an agriculture-associated disease causing Campylobacter coli clade: evidence from national surveillance data in Scotland. *PLoS One* **5,** e15708 (2010).

36. Sheppard, S. K. *et al.* Host association of Campylobacter genotypes transcends geographic variation. *Appl. Environ. Microbiol.* **76,** 5269–77 (2010).

37. Sheppard, S. K. *et al.* Niche segregation and genetic structure of Campylobacter jejuni populations from wild and agricultural host species. *Mol. Ecol.* **20,** 3484–90 (2011).

38. Sheppard, S. K., McCarthy, N. D., Falush, D. & Maiden, M. C. J. Convergence of Campylobacter species: implications for bacterial evolution. *Science* **320,** 237–9 (2008).

39. Pazzaglia, G. *et al.* Campylobacter diarrhoea and an association of recent disease with asymptomatic shedding in Egyptian children. *Epidemiol. Infect.* **106,** 77 (2009).

40. Billingham, J. D. Campylobacter enteritis in The Gambia. *Trans. R. Soc. Trop. Med. Hyg.* **75,** 641–644 (1981).

41. Glass, R. I. *et al.* Epidemiologic and Clinical Features of Endemic Campylobacter jejuni

Infection in Bangladesh. *J. Infect. Dis.* **148,** 292–296 (1983).

42. Rajan, D. P. & Mathan, V. I. Prevalence of Campylobacter fetus subsp. jejuni in healthy populations in southern India. *J. Clin. Microbiol.* **15,** 749–51 (1982).

43. Hellard, M. E., Sinclair, M. I., Hogg, G. G. & Fairley, C. K. Prevalence of enteric pathogens among community based asymptomatic individuals. *J. Gastroenterol. Hepatol.* **15,** 290–3 (2000).

44. de Haan, C. P. A., Kivistö, R., Hakkinen, M., Rautelin, H. & Hänninen, M. L. Decreasing trend of overlapping multilocus sequence types between human and chicken Campylobacter jejuni isolates over a decade in Finland. *Appl. Environ. Microbiol.* **76,** 5228–36 (2010).

45. Hastings, R., Colles, F. M., McCarthy, N. D., Maiden, M. C. J. & Sheppard, S. K. Campylobacter genotypes from poultry transportation crates indicate a source of contamination and transmission. *J. Appl. Microbiol.* **110,** 266–76 (2011).

46. Jorgensen, F. *et al.* Influence of season and geography on Campylobacter jejuni and C. coli subtypes in housed broiler flocks reared in Great Britain. *Appl. Environ. Microbiol.* **77,** 3741–8 (2011).

47. Kittl, S., Kuhnert, P., Hächler, H. & Korczak, B. M. Comparison of genotypes and antibiotic resistance of Campylobacter jejuni isolated from humans and slaughtered chickens in Switzerland. *J. Appl. Microbiol.* **110,** 513–520 (2011).

48. Lang, P. *et al.* Expanded multilocus sequence typing and comparative genomic hybridization of Campylobacter coli isolates from multiple hosts. *Appl. Environ. Microbiol.* **76,** 1913–25 (2010).

49. Müllner, P. *et al.* Molecular epidemiology of Campylobacter jejuni in a geographically isolated country with a uniquely structured poultry industry. *Appl. Environ. Microbiol.* **76,** 2145–54 (2010).

50. Sproston, E. L. *et al.* Temporal variation and host association in the Campylobacter population in a longitudinal ruminant farm study. *Appl. Environ. Microbiol.* **77,** 6579–86 (2011).

51. Sproston, E. L. *et al.* Multi-locus sequence types of Campylobacter carried by flies and slugs acquired from local ruminant faeces. *J. Appl. Microbiol.* **109,** 829–38 (2010).

52. Thakur, S. & Gebreyes, W. A. Phenotypic and genotypic heterogeneity of Campylobacter coli within individual pigs at farm and slaughter in the US. *Zoonoses Public Health* **57 Suppl 1,** 100–6 (2010).

53. Gillespie, I. A. *et al.* Point source outbreaks of Campylobacter jejuni infection--are they more common than we think and what might cause them? *Epidemiol. Infect.* **130,** 367–75 (2003).

54. Young, N. J. *et al.* Campylobacter infection associated with consumption of duck liver pâté: a retrospective cohort study in the setting of near universal exposure. *Epidemiol. Infect.* **142,** 1269–76 (2014).

55. Abid, M. *et al.* Duck liver-associated outbreak of Campylobacteriosis among humans, United Kingdom, 2011. *Emerg. Infect. Dis.* **19,** 1310–3 (2013).

56. Forbes, K. J. *et al.* Campylobacter immunity and coinfection following a large outbreak in a farming community. *J. Clin. Microbiol.* **47,** 111–6 (2009).

57. Colles, F. M., McCarthy, N. D., Bliss, C. M., Layton, R. & Maiden, M. C. J. The long-term dynamics of Campylobacter colonizing a free-range broiler breeder flock: an observational study. *Environ. Microbiol.* **17,** 938–46 (2015).

58. Wirz, S. E., Overesch, G., Kuhnert, P. & Korczak, B. M. Genotype and antibiotic resistance analyses of Campylobacter isolates from ceca and carcasses of slaughtered broiler flocks. *Appl. Environ. Microbiol.* **76,** 6377–86 (2010).

59. Nichols, G. L., Richardson, J. F., Sheppard, S. K., Lane, C. & Sarran, C. Campylobacter epidemiology: a descriptive study reviewing 1 million cases in England and Wales between 1989 and 2011. *BMJ Open* **2,** (2012).

60. Richardson, J. F. *et al.* Coinfection with Campylobacter species: an epidemiological problem? *J. Appl. Microbiol.* **91,** 206–211 (2001).

61. Spencer, S. E. F., Marshall, J., Pirie, R., Campbell, D. & French, N. P. The detection of spatially localised outbreaks in campylobacteriosis notification data. *Spat. Spatiotemporal. Epidemiol.* **2,** 173–83 (2011).

62. Gabriel, E. *et al.* Spatio-temporal epidemiology of Campylobacter jejuni enteritis, in an area of Northwest England, 2000-2002. *Epidemiol. Infect.* **138,** 1384–90 (2010).

63. Adak, G. K., Cowden, J. M., Nicholas, S. & Evans, H. S. The Public Health Laboratory Service national case-control study of primary indigenous sporadic cases of campylobacter infection. *Epidemiol. Infect.* **115,** 15–22 (1995).

64. Altekruse, S. F., Hunt, J. M., Tollefson, L. K. & Madden, J. M. Food and animal sources of human Campylobacter jejuni infection. *J. Am. Vet. Med. Assoc.* **204,** 57–61 (1994).

65. Shane, S. M. & Montrose, M. S. The occurrence and significance of Campylobacter jejuni in man and animals. *Vet. Res. Commun.* **9,** 167–98 (1985).

66. Hood, A. M., Pearson, A. D. & Shahamat, M. The extent of surface contamination of retailed chickens with Campylobacter jejuni serogroups. *Epidemiol. Infect.* **100,** 17–25 (1988).

67. Humphrey, S. *et al.* Campylobacter jejuni is not merely a commensal in commercial broiler chickens and affects bird welfare. *MBio* **5,** e01364-14 (2014).

68. Meade, K. G. *et al.* Comparative in vivo infection models yield insights on early host immune response to Campylobacter in chickens. *Immunogenetics* **61,** 101–10 (2009).

69. Sheppard, S. K. *et al.* Campylobacter genotyping to determine the source of human infection. *Clin. Infect. Dis.* **48,** 1072–8 (2009).

70. Colles, F. M. *et al.* Campylobacter infection of broiler chickens in a free-range environment. *Environ. Microbiol.* **10,** 2042–50 (2008).

71. Miller, W. G. *et al.* Identification of host-associated alleles by multilocus sequence typing of Campylobacter coli strains from food animals. *Microbiology* **152,** 245–55 (2006).

72. Sheppard, H. M. L. W. and S. K. *Current Topics in Phylogenetics and Phylogeography of Terrestrial and Aquatic Systems*. (InTech, 2012). doi:10.5772/1947

73. McCarthy, N. D. *et al.* Host-associated genetic import in Campylobacter jejuni. *Emerg. Infect. Dis.* **13,** 267–72 (2007).

74. Mullner, P. *et al.* Assigning the source of human campylobacteriosis in New Zealand: a comparative genetic and epidemiological approach. *Infect. Genet. Evol.* **9,** 1311–9 (2009).

75. Wilson, D. J. *et al.* Tracing the source of campylobacteriosis. *PLoS Genet.* **4,** e1000203 (2008).

76. Strachan, N. J. C. *et al.* Attribution of Campylobacter infections in northeast Scotland to specific sources by use of multilocus sequence typing. *J. Infect. Dis.* **199,** 1205–8 (2009).

77. Leach, S., Harvey, P. & Wali, R. Changes with growth rate in the membrane lipid composition of and amino acid utilization by continuous cultures of Campylobacter jejuni. *J. Appl. Microbiol.* **82,** 631–40 (1997).

78. Line, J. E., Hiett, K. L., Guard-Bouldin, J. & Seal, B. S. Differential carbon source utilization by Campylobacter jejuni 11168 in response to growth temperature variation. *J. Microbiol. Methods* **80,** 198–202 (2010).

79. Velayudhan, J., Jones, M. A., Barrow, P. A. & Kelly, D. J. L-Serine Catabolism via an Oxygen-Labile L-Serine Dehydratase Is Essential for Colonization of the Avian Gut by Campylobacter jejuni. *Infect. Immun.* **72,** 260–268 (2003).

80. Kapperud, G., Skjerve, E., Bean, N. H., Ostroff, S. M. & Lassen, J. Risk factors for sporadic Campylobacter infections: results of a case-control study in southeastern Norway. *J. Clin. Microbiol.* **30,** 3117–21 (1992).

81. Studahl, A. & Andersson, Y. Risk factors for indigenous campylobacter infection: a Swedish case-control study. *Epidemiol. Infect.* **125,** 269–75 (2000).

82. Waldenström, J., On, S. L. W., Ottvall, R., Hasselquist, D. & Olsen, B. Species diversity of campylobacteria in a wild bird community in Sweden. *J. Appl. Microbiol.* **102,** 424–32 (2007).

83. Colles, F. M., Dingle, K. E., Cody, A. J. & Maiden, M. C. J. Comparison of Campylobacter populations in wild geese with those in starlings and free-range poultry on the same farm. *Appl. Environ. Microbiol.* **74,** 3583–90 (2008).

84. Williams, N. J. *et al.* Isolation of a novel Campylobacter jejuni clone associated with the bank vole, Myodes glareolus. *Appl. Environ. Microbiol.* **76,** 7318–21 (2010).

85. Waldenström, J. *et al.* Campylobacter jejuni colonization in wild birds: results from an infection experiment. *PLoS One* **5,** e9082 (2010).

86. Hald, B., Skovgård, H., Pedersen, K. & Bunkenborg, H. Influxed insects as vectors for Campylobacter jejuni and Campylobacter coli in Danish broiler houses. *Poult. Sci.* **87,** 1428–34 (2008).

87. Hald, B., Sommer, H. M. & Skovgård, H. Use of fly screens to reduce Campylobacter spp. introduction in broiler houses. *Emerg. Infect. Dis.* **13,** 1951–3 (2007).

88. Hald, B. *et al.* Flies and Campylobacter infection of broiler flocks. *Emerg. Infect. Dis.* **10,** 1490–2 (2004).

89. Fernie, D. S. & Park, R. W. The isolation and nature of campylobacters (microaerophilic vibrios) from laboratory and wild rodents. *J. Med. Microbiol.* **10,** 325–9 (1977).

90. Kakoyiannis, C. K., Winter, P. J. & Marshall, R. B. The relationship between intestinal Campylobacter species isolated from animals and humans as determined by BRENDA. *Epidemiol. Infect.* **100,** 379–87 (1988).

91. Meerburg, B. G., Jacobs-Reitsma, W. F., Wagenaar, J. A. & Kijlstra, A. Presence of Salmonella and Campylobacter spp. in wild small mammals on organic farms. *Appl. Environ. Microbiol.* **72,** 960–2 (2006).

92. Nkogwe, C., Raletobana, J., Stewart-Johnson, A., Suepaul, S. & Adesiyun, A. Frequency of Detection of Escherichia coli, Salmonella spp., and Campylobacter spp. in the Faeces of Wild Rats (Rattus spp.) in Trinidad and Tobago. *Vet. Med. Int.* **2011,** 686923 (2011).

93. Peel, R. N. & McIntosh, A. W. The dog it was that died. *Lancet (London, England)* **2,** 1212 (1978).

94. Piddock, L. J. Fluoroquinolone resistance: overuse of fluoroquinolones in human and veterinary medicine can breed resistance. *BMJ* **317,** 1029–30 (1998).

95. Alfredson, D. A. & Korolik, V. Antibiotic resistance and resistance mechanisms in Campylobacter jejuni and Campylobacter coli. *FEMS Microbiol. Lett.* **277,** 123–32 (2007).

96. Gibreel, A., Sjögren, E., Kaijser, B., Wretlind, B. & Sköld, O. Rapid emergence of high-level resistance to quinolones in Campylobacter jejuni associated with mutational changes in gyrA and parC. *Antimicrob. Agents Chemother.* **42,** 3276–8 (1998).

97. Zhou, J. *et al.* A seventeen-year observation of the antimicrobial susceptibility of clinical Campylobacter jejuni and the molecular mechanisms of erythromycin-resistant isolates in Beijing, China. *Int. J. Infect. Dis.* **42,** 28–33 (2016).

98. Chen, I., Christie, P. J. & Dubnau, D. The ins and outs of DNA transfer in bacteria. *Science* **310,** 1456–60 (2005).

99. Wang, Y. & Taylor, D. E. Natural transformation in Campylobacter species. *J. Bacteriol.* **172,** 949–55 (1990).

100. Llosa, M., Gomis-Rüth, F. X., Coll, M. & de la Cruz Fd, F. Bacterial conjugation: a two-step mechanism for DNA transport. *Mol. Microbiol.* **45,** 1–8 (2002).

101. Taylor, D. E., De Grandis, S. A., Karmali, M. A. & Fleming, P. C. Transmissible plasmids from Campylobacter jejuni. *Antimicrob. Agents Chemother.* **19,** 831–835 (1981).

102. Mazodier, P. & Davies, J. Gene Transfer Between Distantly Related Bacteria - annurev.ge.25.120191.001051. *Annual review of microbiology* (1991). Available at: http://www.annualreviews.org/doi/pdf/10.1146/annurev.ge.25.120191.001051.

(Accessed: 30th April 2016)

103. Taylor, D. E. Genetics of Campylobacter and Helicobacter - annurev.mi.46.100192.000343. *Annual review of microbiology* 35–64 (1992). Available at: http://www.annualreviews.org/doi/pdf/10.1146/annurev.mi.46.100192.000343. (Accessed: 30th April 2016)

104. Allen, H. K. *et al.* Call of the wild: antibiotic resistance genes in natural environments. *Nat. Rev. Microbiol.* **8,** 251–9 (2010).

105. Rolland, R. M., Hausfater, G., Marshall, B. & Levy, S. B. Antibiotic-resistant bacteria in wild primates: increased prevalence in baboons feeding on human refuse. *Appl. Environ. Microbiol.* **49,** 791–4 (1985).

106. Cole, D. *et al.* Free-living Canada geese and antimicrobial resistance. *Emerg. Infect. Dis.* **11,** 935–8 (2005).

107. Sjölund, M. *et al.* Dissemination of multidrug-resistant bacteria into the Arctic. *Emerg. Infect. Dis.* **14,** 70–2 (2008).

108. Sjölund, M. *et al.* Antimicrobial susceptibility in Escherichia coli of human and avian origin--a comparison of wild-type distributions. *Clin. Microbiol. Infect.* **15,** 461–5 (2009).

109. Aarestrup, F. M. & Wegener, H. C. The effects of antibiotic usage in food animals on the development of antimicrobial resistance of importance for humans in Campylobacter and Escherichia coli. *Microbes Infect.* **1,** 639–44 (1999).

110. Aarestrup, F. M. & Engberg, J. Antimicrobial resistance of thermophilic Campylobacter. *Vet. Res.* **32,** 311–21

111. Endtz, H. P. *et al.* Quinolone resistance in campylobacter isolated from man and poultry following the introduction of fluoroquinolones in veterinary medicine. *J. Antimicrob. Chemother.* **27,** 199–208 (1991).

112. Karp, B. E. & Engberg, J. Comment on: Does the use of antibiotics in food animals pose a risk to human health? A critical review of published data. *J. Antimicrob. Chemother.* **54,** 273-4-8 (2004).

113. Gormley, F. J. *et al.* Antimicrobial resistance profiles of Campylobacter from humans, retail chicken meat, and cattle feces. *Foodborne Pathog. Dis.* **7,** 1129–31 (2010).

114. Ochman, H. & Wilson, A. C. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* **26,** 74–86 (1987).

115. Falush, D., Almqvist, E. W., Brinkmann, R. R., Iwasa, Y. & Hayden, M. R. Measurement of mutational flow implies both a high new-mutation rate for Huntington disease and substantial underascertainment of late-onset cases. *Am. J. Hum. Genet.* **68,** 373–85 (2001).

116. Feng, L. *et al.* A recalibrated molecular clock and independent origins for the cholera pandemic clones. *PLoS One* **3,** e4053 (2008).

117. Wimalarathna, H. M. L. *et al.* Widespread acquisition of antimicrobial resistance among Campylobacter isolates from UK retail poultry and evidence for clonal expansion of resistant lineages. *BMC Microbiol.* **13,** 160 (2013).

118. Dingle, K. E. *et al.* Multilocus sequence typing system for Campylobacter jejuni. *J. Clin. Microbiol.* **39,** 14–23 (2001).

119. Jolley, K. A. & Maiden, M. C. J. Using multilocus sequence typing to study bacterial variation: prospects in the genomic era. *Future Microbiol.* **9,** 623–30 (2014).

120. Sheppard, S. K., Jolley, K. A. & Maiden, M. C. J. A Gene-By-Gene Approach to Bacterial Population Genomics: Whole Genome MLST of Campylobacter. *Genes (Basel).* **3,** 261–77 (2012).

121. van Tonder, A. J. *et al.* Defining the estimated core genome of bacterial populations using a Bayesian decision model. *PLoS Comput. Biol.* **10,** e1003788 (2014).

122. Jolley, K. A. *et al.* Ribosomal multilocus sequence typing: Universal characterization of bacteria from domain to strain. *Microbiology* **158,** 1005–1015 (2012).

123. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18,** 821–9 (2008).

124. Zerbino, D. R. Using the Velvet de novo assembler for short-read sequencing technologies. *Curr. Protoc. Bioinformatics* **Chapter 11,** Unit 11.5 (2010).

125. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–77 (2012).

126. Miller, J. R., Koren, S. & Sutton, G. Assembly algorithms for next-generation sequencing data. *Genomics* **95,** 315–27 (2010).

127. Ekblom, R. & Wolf, J. B. W. A field guide to whole-genome sequencing, assembly and annotation. *Evol. Appl.* **7,** n/a-n/a (2014).

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

129. Huson, D. H. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14,** 68–73 (1998).

130. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–80 (2013).

131. McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41,** W597-600 (2013).

132. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28,** 2731–9 (2011).

133. Kumar, S., Tamura, K. & Nei, M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput. Appl. Biosci.* **10,** 189–91 (1994).

134. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–25 (1987).

135. Bryant, D. & Moulton, V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21,** 255–65 (2004).

136. Didelot, X. & Falush, D. Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175,** 1251–66 (2007).

137. London, R. S. *Philosophical Transactions of the Royal Society of London: Giving Some Accounts of the Present Undertakings, Studies, and Labours, of the Ingenious, in Many Considerable Parts of the World, Volume 53*. (1764).

138. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–59 (2000).

139. Miller, W. G. *et al.* Extended multilocus sequence typing system for Campylobacter coli, C. lari, C. upsaliensis, and C. helveticus. *J. Clin. Microbiol.* **43,** 2315–29 (2005).

140. Gaunt, P. N. & Piddock, L. J. Ciprofloxacin resistant Campylobacter spp. in humans: an epidemiological and laboratory study. *J. Antimicrob. Chemother.* **37,** 747–57 (1996).

141. Thwaites, R. T. & Frost, J. A. Drug resistance in Campylobacter jejuni, C coli, and C lari isolated from humans in north west England and Wales, 1997. *J. Clin. Pathol.* **52,** 812–4 (1999).

142. Dingle, K. E., Colles, F. M., Falush, D. & Maiden, M. C. J. Sequence typing and comparison of population biology of Campylobacter coli and Campylobacter jejuni. *J. Clin. Microbiol.* **43,** 340–7 (2005).

143. Embley, T. M. The linear PCR reaction: a simple and robust method for sequencing amplified rRNA genes. *Lett. Appl. Microbiol.* **13,** 171–4 (1991).

144. Staden, R. The Staden sequence analysis package. *Mol. Biotechnol.* **5,** 233–41 (1996).

145. Jolley, K. A. *et al.* mlstdbNet – distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics* **5,** 86 (2004).

146. Jolley, K. A. Internet-based sequence-typing databases for bacterial molecular

epidemiology. *Methods Mol. Biol.* **551,** 305–12 (2009).

147. Maiden, M. C. J. *et al.* MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* **11,** 728–36 (2013).

148. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23,** 254–67 (2006).

149. Magrane, M. & Consortium, U. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford).* **2011,** bar009 (2011).

150. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24,** 1596–9 (2007).

151. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10,** 564–7 (2010).

152. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8,** e1002453 (2012).

153. Hubisz, M. J., Falush, D., Stephens, M. & Pritchard, J. K. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **9,** 1322–32 (2009).

154. Bolwell, C. F., Gilpin, B. J., Campbell, D. & French, N. P. Evaluation of the representativeness of a sentinel surveillance site for campylobacteriosis. *Epidemiol. Infect.* **143,** 1990–2002 (2015).

155. Sears, A. *et al.* Marked campylobacteriosis decline after interventions aimed at poultry, New Zealand. *Emerg. Infect. Dis.* **17,** 1007–15 (2011).

156. McCarthy, N. D. *et al.* Molecular epidemiology of human Campylobacter jejuni shows association between seasonal and international patterns of disease. *Epidemiol. Infect.* **140,** 2247–55 (2012).

157. Cody, A. J. *et al.* A longitudinal 6-year study of the molecular epidemiology of clinical campylobacter isolates in Oxfordshire, United kingdom. *J. Clin. Microbiol.* **50,** 3193–201 (2012).

158. Inouye, M., Conway, T. C., Zobel, J. & Holt, K. E. Short read sequence typing (SRST): multi-locus sequence types from short reads. *BMC Genomics* **13,** 338 (2012).

159. Larsen, M. V *et al.* Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* **50,** 1355–61 (2012).

160. Griekspoor, P. *et al.* Marked host specificity and lack of phylogeographic population structure of Campylobacter jejuni in wild birds. *Mol. Ecol.* **22,** 1463–72 (2013).

161. Mughini-Gras, L. *et al.* Campylobacteriosis in returning travellers and potential secondary transmission of exotic strains. *Epidemiol. Infect.* **142,** 1277–88 (2014).

162. Gillespie, I. A. *et al.* A case-case comparison of Campylobacter coli and Campylobacter jejuni infection: a tool for generating hypotheses. *Emerg. Infect. Dis.* **8,** 937–42 (2002).

163. Stecher, B. *et al.* Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 1269–74 (2012).

164. Broman, T. *et al.* Campylobacter jejuni in black-headed gulls (Larus ridibundus): prevalence, genotypes, and influence on C. jejuni epidemiology. *J. Clin. Microbiol.* **40,** 4594–602 (2002).

165. Tu, Z.-C. *et al.* Campylobacter fetus of reptile origin as a human pathogen. *J. Clin. Microbiol.* **42,** 4405–7 (2004).

166. Zia, S. *et al.* Health problems following Campylobacter jejuni enteritis in a Lancashire population. *Rheumatology (Oxford).* **42,** 1083–8 (2003).

167. Cody, A. J., Clarke, L., Bowler, I. C. J. W. & Dingle, K. E. Ciprofloxacin-resistant campylobacteriosis in the UK. *Lancet (London, England)* **376,** 1987 (2010).

168. Reina, J., Borrell, N. & Serra, A. Emergence of resistance to erythromycin and

fluoroquinolones in thermotolerant Campylobacter strains isolated from feces 1987-1991. *Eur. J. Clin. Microbiol. Infect. Dis.* **11,** 1163–6 (1992).

169. Lee, M. D. *et al.* Class 1 integron-associated tobramycin-gentamicin resistance in Campylobacter jejuni isolated from the broiler chicken house environment. *Antimicrob. Agents Chemother.* **46,** 3660–4 (2002).

170. Talsma, E., Goettsch, W. G., Nieste, H. L., Schrijnemakers, P. M. & Sprenger, M. J. Resistance in Campylobacter species: increased resistance to fluoroquinolones and seasonal variation. *Clin. Infect. Dis.* **29,** 845–8 (1999).

171. Levy, S. B. Factors impacting on the problem of antibiotic resistance. *J. Antimicrob. Chemother.* **49,** 25–30 (2002).

172. Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **74,** 417–33 (2010).

173. Neimann, J., Engberg, J., Mølbak, K. & Wegener, H. C. A case-control study of risk factors for sporadic campylobacter infections in Denmark. *Epidemiol. Infect.* **130,** 353–66 (2003).

174. Vellinga, A. & Van Loock, F. The dioxin crisis as experiment to determine poultry-related campylobacter enteritis. *Emerg. Infect. Dis.* **8,** 19–22 (2002).

175. Strachan, N. J. C. & Forbes, K. J. The growing UK epidemic of human campylobacteriosis. *Lancet (London, England)* **376,** 665–7 (2010).

176. Kinana, A. D. *et al.* Genetic diversity and quinolone resistance in Campylobacter jejuni isolates from poultry in Senegal. *Appl. Environ. Microbiol.* **72,** 3309–13 (2006).

177. Spratt, B. G. Hybrid penicillin-binding proteins in penicillin-resistant strains of Neisseria gonorrhoeae. *Nature* **332,** 173–6 (1988).

178. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405,** 299–304 (2000).

179. Taylor, D. E., Hiratsuka, K., Ray, H. & Manavathu, E. K. Characterization and expression of a cloned tetracycline resistance determinant from Campylobacter jejuni plasmid pUA466. *J. Bacteriol.* **169,** 2984–9 (1987).

180. Courvalin, P. M., Shaw, W. V & Jacob, A. E. Plasmid-mediated mechanisms of resistance to aminoglycoside-aminocyclitol antibiotics and to chloramphenicol in group D streptococci. *Antimicrob. Agents Chemother.* **13,** 716–25 (1978).

181. Engberg, J., Aarestrup, F. M., Taylor, D. E., Gerner-Smidt, P. & Nachamkin, I. Quinolone and macrolide resistance in Campylobacter jejuni and C. coli: resistance mechanisms and trends in human isolates. *Emerg. Infect. Dis.* **7,** 24–34

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

183. D'lima, C. B. *et al.* Clonal population structure and specific genotypes of multidrug-resistant Campylobacter coli from Turkeys. *Appl. Environ. Microbiol.* **73,** 2156–64 (2007).

184. Bywater, R. J. Veterinary use of antimicrobials and emergence of resistance in zoonotic and sentinel bacteria in the EU. *J. Vet. Med. B. Infect. Dis. Vet. Public Health* **51,** 361–3

185. Castanon, J. I. R. History of the use of antibiotic as growth promoters in European poultry feeds. *Poult. Sci.* **86,** 2466–71 (2007).

186. McDermott, P. F. *et al.* Ciprofloxacin resistance in Campylobacter jejuni evolves rapidly in chickens treated with fluoroquinolones. *J. Infect. Dis.* **185,** 837–40 (2002).

187. Harrison, W. A., Griffith, C. J., Tennant, D. & Peters, A. C. Incidence of Campylobacter and Salmonella isolated from retail chicken and associated packaging in South Wales. *Lett. Appl. Microbiol.* **33,** 450–4 (2001).

188. Whyte, P. *et al.* Occurrence of Campylobacter in retail foods in Ireland. *Int. J. Food Microbiol.* **95,** 111–8 (2004).

189. Taylor, E. V *et al.* Common source outbreaks of Campylobacter infection in the USA, 1997-2008. *Epidemiol. Infect.* **141,** 987–96 (2013).

190. Newell, D. G. & Fearnley, C. Sources of Campylobacter colonization in broiler chickens. *Appl. Environ. Microbiol.* **69,** 4343–51 (2003).

191. Kemp, R. *et al.* Prevalence and genetic diversity of Campylobacter spp. in environmental water samples from a 100-square-kilometer predominantly dairy farming area. *Appl. Environ. Microbiol.* **71,** 1876–82 (2005).

192. Allen, V. M. *et al.* Sources and spread of thermophilic Campylobacter spp. during partial depopulation of broiler chicken flocks. *J. Food Prot.* **71,** 264–70 (2008).

193. Allen, V. M. *et al.* Evaluation of the performance of different cleaning treatments in reducing microbial contamination of poultry transport crates. *Br. Poult. Sci.* **49,** 233–40 (2008).

194. Davis, M. A., Hancock, D. D., Besser, T. E. & Call, D. R. Evaluation of pulsed-field gel electrophoresis as a tool for determining the degree of genetic relatedness between strains of Escherichia coli O157:H7. *J. Clin. Microbiol.* **41,** 1843–9 (2003).

195. Zou, W. *et al.* Evaluation of pulsed-field gel electrophoresis profiles for identification of Salmonella serotypes. *J. Clin. Microbiol.* **48,** 3122–6 (2010).

196. Colles, F. M. *et al.* Dynamics of Campylobacter colonization of a natural host, Sturnus vulgaris (European starling). *Environ. Microbiol.* **11,** 258–67 (2009).

197. Kwan, P. S. L. *et al.* Molecular epidemiology of Campylobacter jejuni populations in dairy cattle, wildlife, and the environment in a farmland area. *Appl. Environ. Microbiol.* **74,** 5130–8 (2008).

198. Brown, P. R. & Khamphoukeo, K. Farmers' knowledge, attitudes, and practices with respect to rodent management in the upland and lowland farming systems of the Lao People's Democratic Republic. *Integr. Zool.* **2,** 165–73 (2007).

199. Webster, J. P. & Macdonald, D. W. Parasites of wild brown rats (Rattus norvegicus) on UK farms. *Parasitology* **111 (Pt 3,** 247–55 (1995).

200. Backhans, A. & Fellström, C. Rodents on pig and chicken farms - a potential threat to human and animal health. *Infect. Ecol. Epidemiol.* **2,** (2012).

201. Newell, D. G. *et al.* Biosecurity-based interventions and strategies to reduce Campylobacter spp. on poultry farms. *Appl. Environ. Microbiol.* **77,** 8605–14 (2011).

202. BARNETT, S. A. Physiological effects of social stress in wild rats. I. The adrenal cortex. *J. Psychosom. Res.* **3,** 1–11 (1958).

203. Gardner-Santana, L. C. *et al.* Commensal ecology, urban landscapes, and their influence on the genetic characteristics of city-dwelling Norway rats (Rattus norvegicus). *Mol. Ecol.* **18,** 2766–78 (2009).

204. Hepworth, P. J. *et al.* Genomic variations define divergence of water/wildlife-associated Campylobacter jejuni niche specialists from common clonal complexes. *Environ. Microbiol.* **13,** 1549–60 (2011).

205. Taboada, E. N. *et al.* Comparative genomic assessment of Multi-Locus Sequence Typing: rapid accumulation of genomic heterogeneity among clonal isolates of Campylobacter jejuni. *BMC Evol. Biol.* **8,** 229 (2008).

206. Pickett, C. L. *et al.* Prevalence of cytolethal distending toxin production in Campylobacter jejuni and relatedness of Campylobacter sp. cdtB gene. *Infect. Immun.* **64,** 2070–8 (1996).

207. Abuoun, M. *et al.* Cytolethal distending toxin (CDT)-negative Campylobacter jejuni strains and anti-CDT neutralizing antibodies are induced during human infection but not during colonization in chickens. *Infect. Immun.* **73,** 3053–62 (2005).

208. Zilbauer, M., Dorrell, N., Wren, B. W. & Bajaj-Elliott, M. Campylobacter jejuni-mediated disease pathogenesis: an update. *Trans. R. Soc. Trop. Med. Hyg.* **102,** 123–9 (2008).

209. Rozynek, E. *et al.* Prevalence of potential virulence markers in Polish Campylobacter jejuni and Campylobacter coli isolates obtained from hospitalized children and from chicken carcasses. *J. Med. Microbiol.* **54,** 615–9 (2005).

210. Blaser, M. J. *et al.* Isolation of Campylobacter fetus subsp. jejuni from Bangladeshi

children. *J. Clin. Microbiol.* **12,** 744–7 (1980).

211. Sheppard, S. K. *et al.* Progressive genome-wide introgression in agricultural Campylobacter coli. *Mol. Ecol.* **22,** 1051–64 (2013).

212. Hald, B. *et al.* Campylobacter jejuni and Campylobacter coli in wild birds on Danish livestock farms. *Acta Vet. Scand.* **58,** 11 (2016).

213. Rodrigues, L. C. *et al.* The study of infectious intestinal disease in England: risk factors for cases of infectious intestinal disease with Campylobacter jejuni infection. *Epidemiol. Infect.* **127,** 185–93 (2001).

214. Pires, S. M., Vigre, H., Makela, P. & Hald, T. Using outbreak data for source attribution of human salmonellosis and campylobacteriosis in Europe. *Foodborne Pathog. Dis.* **7,** 1351–61 (2010).

215. Hald, T., Vose, D., Wegener, H. C. & Koupeev, T. A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. *Risk Anal.* **24,** 255–69 (2004).

216. Boysen, L. Campylobacter in Denmark: control, human risk and source attribution. (2012).

217. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164,** 1567–87 (2003).

218. Bodmer, W. F. & Cavalli-Sforza, L. L. A migration matrix model for the study of random genetic drift. *Genetics* **59,** 565–92 (1968).

219. Gripp, E. *et al.* Closely related Campylobacter jejuni strains from different sources reveal a generalist rather than a specialist lifestyle. *BMC Genomics* **12,** 584 (2011).

220. Dearlove, B. L. *et al.* Rapid host switching in generalist Campylobacter strains erodes the signal for tracing human infections. *ISME J.* (2015). doi:10.1038/ismej.2015.149

221. Sheppard, S. K. *et al.* Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 11923–7 (2013).

222. Cody, A. J. *et al.* Wild bird associated Campylobacter jejuni isolates are a consistent source of human disease, in Oxfordshire, United Kingdom. *Environ. Microbiol. Rep.* n/a-n/a (2015). doi:10.1111/1758-2229.12314

223. Noormohamed, A. & Fakhr, M. K. A higher prevalence rate of Campylobacter in retail beef livers compared to other beef and pork meat cuts. *Int. J. Environ. Res. Public Health* **10,** 2058–68 (2013).

224. Osano, O. & Arimi, S. M. Retail poultry and beef as sources of Campylobacter jejuni. *East Afr. Med. J.* **76,** 141–3 (1999).

225. Fernandes, A. M. *et al.* Partial Failure of Milk Pasteurization as a Risk for the Transmission of Campylobacter From Cattle to Humans. *Clin. Infect. Dis.* **61,** 903–9 (2015).

226. Piddock, L. J. V. Activity of antibiotics used in human medicine for Campylobacter jejuni isolated from farm animals and their environment in Lancashire, UK. *J. Antimicrob. Chemother.* **46,** 303–306 (2000).

227. Thakur, S. & Gebreyes, W. A. Prevalence and Antimicrobial Resistance of Campylobacter in Antimicrobial-Free and Conventional Pig Production Systems.

228. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–9 (2014).