Application of sequence typing to the epidemiology of *Campylobacter jejuni*

Noel McCarthy

St Peter's College
University of Oxford

Michaelmas Term 2008

A thesis submitted for the degree of Doctor of Philosophy

RADCLIFFE SCIENCE LIBRARY OXFORD
DEPOSITED
16 FEB 2010
THESIS

Application of sequence typing to the epidemiology of *Campylobacter jejuni*

ABSTRACT

The bacterium *Campylobacter jejuni* is responsible for a substantial burden of human disease. It has a wide host and environmental distribution. The ecology in non-human hosts has been studied, in particular for farm animals, but the relationship between different host and environmental niches, including the extent to which host and geography are associated with genetic differentiation, is uncertain. Open epidemiological questions include the largely unexplained summer peaks in many temperate countries, apparent paucity of outbreaks despite a high incidence of foodborne infection, and uncertainty in the quantitative contributions of different sources to human infection. Bacterial subtyping has not made a substantial contribution to the ecology or epidemiology of this species. This thesis applied a housekeeping gene multilocus sequence typing scheme, complemented with more variable antigen genes, in epidemiological and ecological studies of *C. jejuni*.

These studies identified a large contribution of one clonal group to the high level of human infection during summer along with a similar seasonal rise in relative prevalence for this clade in poultry meat, and the association of other phenotypic characteristics with the clade. A lack of geographical variation between distant populations within England contrasted with substantial differences at international level. Genetic differentiation by host species exceeded geographic and temporal effects and showed the potential of using multilocus genotype to attribute human infection to animal host sources. Recombination played a major role in the generation of this genetic differentiation, which finding informed the use of alleles rather than summary measures of genotype in population assignment of *C. jejuni*. Using fine typing with *porA* and *flaA* gene fragments to produce higher discrimination than previously applied allowed the identification of genotypic clusters, demonstrated emergent clades within the poultry industry and human disease, and confirmed empirically the predictions from theoretical work that even more highly discriminatory methods will be needed to reliably identify outbreaks and their specific source using genotype.

This work has thus exploited the of integration multilocus sequencing in studies of the epidemiology and ecology of *C. jejuni*, developed analytical approaches for this application, and identified some limitations and the extent to which these may be tractable with more extensive sequence data and further development of analytical approaches.

Preface

This thesis is based on collaborative projects, mostly involving typing sub-studies within larger pre-existing projects. The work presented in the thesis is restricted to those areas where I was the main contributor to the overall study or to the conception and conduct of the specific sub-studies.

Specifically: In the *Campylobacter* Sentinel Surveillance Scheme (CSSS) component of this thesis, a typing sub-study of a large multicentre enhanced surveillance project, I

1. Formulated the hypotheses to be tested in the typing sub-study
2. Wrote the funding application with guidance from Martin Maiden, my thesis supervisor, and Sarah O'Brien, the principal investigator for the CSSS
3. Defined the subset of isolates to be included with Iain Gillespie who ran the epidemiological component of the overall CSSS
4. Completed the sequence typing of the isolates presented in the thesis with the exception of 190 isolates where this work was shared with Alison Cody and
5. Analysed the results.

The epidemiology data included in this analysis was supplied by Ian Gillespie from the main study. Preparation of DNA extracts was undertaken by staff at the Health Protection Agency Centre for Infections led by Andy Lawson.

In the host association studies I formulated the questions, established the dataset from published isolates and collaborators, and analysed the data with supervision in the application of analytical techniques from Daniel Falush as well as overall guidance from my thesis supervisor.

In the retail poultry survey I wrote the application for funding, coordinated the project and performed the analyses presented. Others funded by the project undertook the laboratory work. The sample information was provided from several individuals and institutions involved in the original studies.

For the case-control study I led the MLST substudy within a larger ongoing case-control study led by Sarah O'Brien. This included writing the funding application for the MLST substudy in collaboration with Robert Owen, coordination of this study and analysis of the data as presented. Frances Colles performed most of the laboratory work (approximately 85%). I completed the laboratory work (repeats of failing isolates and loci).

I planned and performed the analyses for the assessment of discrimination in relation to outbreak investigation.

Acknowledgements
Many others have directly and indirectly supported me in the work presented in this thesis. Especially large contributions require particular acknowledgement.

Martin Maiden for flexible supervision and support in this thesis as well as personally, and for having developed the foundation and context that made this work possible.

Frances Colles and Alison Cody, who supported me in learning sequence typing and contributed to several of the projects described in this thesis. Keith Jolley for bioinformatics support. Daniel Falush for guidance in population genetic analysis. Kate Dingle for access to her unpublished work on the application of *porA* sequencing.

Robert Owen, Andy Lawson, Judith Richardson, Ian Gillespie and Sarah O' Brien (now University of Manchester) at the Centre for Infections have been extremely positive collaborators on studies presented here through which we have jointly exploited the HPA held sample archives and added sequence typing components to ongoing projects.

Eamonn O'Moore for his generosity in supporting my delay in taking up a joint clinical post with him, despite the attendant additional work that resulted for him during that time.

Last, and most of all, my family who have put up with the impact of this work on family life, and for rather longer than was foreseen or agreed.

Funding: This work was funded by a clinical training fellowship from the Wellcome Trust (Grant No. WT072782MA). The work on retail poultry was funded by a project grant from the FSA (Grant No. B15011).

Table of Contents

Abstract.....i
Preface.....ii
Acknowledgements.....iii
Table of Contents.....iv

1. Introduction.....1
1.1 Campylobacteriosis.....1
1.1.1 Usual symptoms in high income countries.....1
1.1.2 Complications.....2
1.1.3 Asymptomatic infection.....2
1.1.4 Clinical picture in low income countries.....3
1.2 Descriptive epidemiology of *Campylobacter jejuni*.....3
1.2.1 Extent and pattern of human disease.....3
1.2.2 Demographic pattern.....4
1.2.3 Small scale geographic effects.....6
1.2.4 Summer peak and seasonal variation.....7
1.3 Attributing human disease to source.....14
1.3.1 Risk assessment.....15
1.3.2 Outbreak investigation.....16
1.3.3 Analytical epidemiology of incident cases.....18
1.3.4 Natural experiments.....31
1.3.5 Microbial subtype.....32
1.3.6 Summary of risk factors.....33
1.3.7 Integrating genetic sequencing into source attribution.....34
1.4 *Campylobacter jejuni* biology.....35
1.4.1 Phenotype.....35
1.4.2 Taxonomy and genome.....36
1.4.3 Ecology of *Campylobacter jejuni*.....36
1.5 Typing of *Campylobacter jejuni*.....40
1.5.1 Phenotyping.....41
1.5.2 Genotyping other than sequence typing.....43
1.5.3 Sequence typing.....49
1.6 Using published literature to guide the work of the thesis.....64
2. Project Aims.....65
2.1 Subtypes of *Campylobacter jejuni* in a human population (Chapter 4).....65
2.2 Source attribution by genotype (Chapter 5).....66
2.3 The *Campylobacter jejuni* population in chicken (Chapter 6).....66
2.4 Outbreak detection (Chapter 7).....67
2.5 Genotype, source of infection, and population biology (Chapter 8).....68
3. Methods.....69
3.1 Laboratory methods.....69
3.1.1 Preparation of heat-killed cell suspensions.....69
3.1.2 Sample handling and PCR amplification.....70
3.1.3 Sequencing.....72
3.1.4 Primers.....74
3.1.5 Sequence assembly and bioinformatics.....76
3.2 Study samples and datasets.....77
3.2.1 *Campylobacter* Sentinel Surveillance Scheme (CSSS) dataset.....77
3.2.2 Retail poultry survey (RPS) dataset.....79

3.2.3 Case control study (CaCo) dataset 82
3.2.4 Cattle sheep and chicken (CSP) dataset 83
3.2.5 Published human disease (PHD) dataset 85
3.2.6 International published animal and environmental datasets 86
3.2.7 Attribution reference dataset 87
3.3 Analysis 88
3.4 Methods for descriptive molecular epidemiology of CSSS data 88
3.4.1 Geographical and demographic analyses 88
3.4.2 Change in subtypes over time 89
3.4.3 Clonal complex and symptoms 89
3.4.4 Description of population structure 90
3.5 Source attribution by genotype 91
3.6 Analysis of the *Campylobacter jejuni* population in chicken 93
3.6.1 Population structure 93
3.6.2 Statistical analysis 93
3.6.3 Population assignment 94
3.6.4 Discrimination 94
3.7 Simulation of sequencing approaches for outbreak investigation 95
3.7.1 Population 95
3.7.2 Simulation of typing 96
3.7.3 Scaling simulation to produce correct results 97
3.7.4 Selection of simulation results presented 97
3.8 Genotype and source of human infection 98
3.8.1 Seasonality in human and chicken samples 98
3.8.2 Comparing genotypic attribution with risk factor information 99
3.8.3 Use of discriminatory typing to investigate outbreaks 101
4. Subtypes of *Campylobacter jejuni* in human populations 102
4.1 Geographical distribution of sequence types 103
4.1.1 Two geographically separate populations in England 103
4.1.2 Comparison with international datasets 104
4.2 Temporal patterns among subtypes 107
4.2.1 Temporal distribution by clonal complex 107
4.2.2 Temporal distribution by sequence type 112
4.3 Association of demographic characteristics with genotype 112
4.3.1 Age 112
4.3.2 Ethnic group 116
4.3.3 Gender 118
4.4 Clonal complex and symptoms 118
4.5 Population structure of *C. jejuni* in the CSSS and published samples 120
4.5.1 Clonal population structure 120
4.5.2 Population structure and geography 123
4.6 Discussion 125
5. Source attribution using typing 130
5.1 Dataset description and gene flow 131
5.2 Population assignment 132
5.2.1 Results of assignment using different data forms 132
5.2.2 Tests for confounding 134
5.2.3 Biology of the observed host association 135
6. *Campylobacter jejuni* contamination of retail poultry 143
6.1 Description of population sampled 143
6.1.1 Sample characteristics 143

6.1.2 Sequence types and population structure 144
6.2 Relationship between subtype and sample characteristics 145
6.2.1 Association of time of sample and subtype 145
6.2.2 Geography and subtype 148
6.2.3 Production and handling of chicken and subtype 150
6.3 Joint analysis with published data from other animal hosts 151
6.4 Antigen gene results 153
6.4.1 Discriminatory index and epidemiology using 9 locus genotype 153
6.4.2 Recent changes in each gene as a contribution to discrimination 156
6.5 Discussion 157
7. Typing approaches to identify outbreaks 161
7.1 Results 164
7.2 Discussion 169
8. Genotype, source of infection, & population biology 173
8.1 Seasonality and clonal group in human and chicken samples 176
8.2 Genotype and risk factors in the sentinel surveillance scheme 179
8.3 Discriminatory typing to isolates of human and poultry origin 185
8.4 Discussion 189
8.4.1 Seasonality 189
8.4.2 Attributing human infection using genotype 191
8.4.3 Analysis of discriminatory typing of human and poultry isolates 194
9. References 199
10. Annexes 219
10.1 Annex 1 Summary of risk factors 219
10.1.1 Host factors 219
10.1.2 Food types 220
10.1.3 Routes of exposure to food 221
10.1.4 Water 222
10.1.5 Environment and animal exposures 222
10.1.6 Occupational risk 223
10.1.7 Person to person spread 223
10.2 Annex 2. Questionnaire 224
10.3 Annex 3 – Additional tables from chapter 4. 229
10.3.1 Allele distributions by locus for isolates with complete and incomplete types. 229
10.3.2 Clonal complexes by county in CSSS 236
10.4 Annex 4. CC and ST distribution in datasets from 4 different countries 237
10.5 Annex 5. Tabulation of ST distribution in peak and non-peak periods. 243
10.6 Annex 6 – Between country population comparisons using eBURST 247
10.7 Annex 7 – Detailed data tables of clonal complex and ST distribution for retail poultry survey 253
10.8 Annex 8. Temporal pattern among human and chicken isolates. (Figures for chapter 8) 259
10.9 Annex 9. ST distribution form human disease and animal hosts 262

1. Introduction

This thesis describes studies of the human zoonotic pathogen *Campylobacter jejuni.* It has the substantive aim of increasing understanding of the epidemiology of *C. jejuni* infection as a basis for disease control, and the methodological aim of further developing the application of sequence typing at the interface of epidemiology, ecology and evolution[1]. Clarification of the population biology and ecology of this species supports the primary epidemiological motivation but is also important of itself since *C. jejuni* can act as a model to address general questions regarding the biology and ecology of multi-host bacterial species.

In order to provide an overall context for the work and to identify questions amenable to investigation this introduction first reviews the disease and the general epidemiological features in humans including the evidence for the sources of human infection with a focus on identifying those gaps in current knowledge that may be susceptible to the integration of sequence typing into other investigative approaches. It then describes the biology and ecology of *C. jejuni*. Lastly it reviews the approaches to microbial sub-typing that have been applied and assesses the characteristics of these methods and their contribution to understanding of the population biology and epidemiology of *C. jejuni*. This review of typing includes consideration of the potential of sequence typing to address identified gaps in understanding and to support disease control.

1.1 Campylobacteriosis

1.1.1 Usual symptoms in high income countries

C. jejuni is primarily a cause of gastroenteritis with onset 2-5 days following infection[2]. Laboratory confirmed cases in one large UK study typically reported diarrhoea (95%), abdominal pain (85%) and fever (78%) and less commonly vomiting (35%) and bloody

diarrhoea (27%), with factors such as young age and large infecting dose associated with more severe disease[3]. Ten percent of these cases were admitted to hospital. This disease spectrum was based on laboratory confirmed cases in a high-income country. Symptoms and certainly hospitalisation rates of non-laboratory confirmed clinical cases are likely to be less severe. The substantial evidence for asymptomatic infection is described below.

1.1.2 Complications

Extra-intestinal infection is rare[4,5] but non-infectious extra-intestinal complications may occur in the weeks following infection notably reactive arthritis and the neurological condition of Guillain-Barré syndrome, including the Miller Fisher variant. Reactive arthritis incidence post gastroenteritis has been estimated at between one and five percent[6] and *C. jejuni* infection may be the most common preceding infectious cause[7] of reactive arthritis. It is usually self-limiting within six months[6,7]. Neurological complications are more severe but much less common with an incidence of approximately 1 in 3,000 for Guillain-Barré syndrome following *C. jejuni* gastroenteritis[8,9]. Since Guillain-Barré syndrome produces a flaccid paralysis which in some cases affects the muscles of respiration and the autonomic nervous system it is an important complication and although rare contributes substantially to the total health impact and economic burden of this infection[10].

1.1.3 Asymptomatic infection

Asymptomatic infection is common. Firstly, of 89 experimental infections in healthy adults only 29 were associated with symptoms[11]. Secondly, the observed point prevalence of *C. jejuni* infection has been estimated at 0.1%-1%[12-19]. Given that duration of excretion after infection is usually less than a few weeks[20 ,21-23], and symptomatic infections have an incidence below 1% in high income countries[14,17,24-26] the expected population prevalence if all infections were symptomatic (with 1% annual infection and mean duration of 3 weeks) would be less than 0.06%, substantially lower that the observed prevalence. Thirdly,

population seroepidemiology has shown over 20% seropositivity in young adults[27] although antibody levels generally decline substantially within two to three years of an infection[28,29]. These three lines of evidence suggest that the majority, perhaps a large majority, of infections are asymptomatic.

1.1.4 Clinical picture in low income countries

Clinical disease in low-income countries appears to be milder on average than in high-income countries[30], but with the same range of severity and types of complications. Infection is very frequent in young children, over 100 times higher than in high income countries[2,22] . There is a decreasing probability of disease following infection with increasing age among children in these settings[31,32]. The prevalence of *C. jejuni* is often over 10% among asymptomatic children over 1 year of age and both serological[33,34] and serial culture[31,35-37] studies have shown that the large majority of children have been infected by the age of two, many several times.

1.2 Descriptive epidemiology of *Campylobacter jejuni*

This summary of the epidemiology of *C. jejuni* focuses on high-income countries and in particular the UK where the work of this thesis is based. It is intended to summarise this area and to identify questions where reliable definition of sub-types may contribute to increased understanding and so guide the work of this thesis.

1.2.1 Extent and pattern of human disease

There have been between 44,294 (2004) and 57,674 (2000) reports of laboratory confirmed human *Campylobacter* infection per year in England and Wales (Figure 1.1) between 1997 and 2006. These are not exclusively *C. jejuni*, and accurate speciation is not always performed. Where speciation has been performed it suggests that approximately 93% of human isolates in England and Wales are likely to be *C. jejuni*[38].

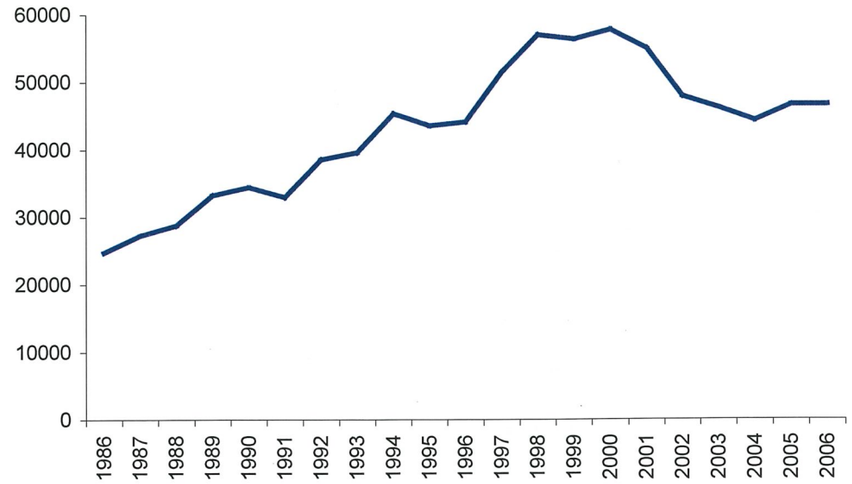


Figure 1.1. Laboratory reports of faecal isolation of *Campylobacter* in England and Wales, 1986-2006.[39]

The best estimate for the relationship between the number of incident symptomatic *Campylobacter* infections in the community and the number reported nationally is that 1 in 7.6 is diagnosed and reported[24]. These adjustments give an estimated 300,000 to 400,000 cases per year in England and Wales over the past decade, an incidence rate in the range of 6 to 8 cases per 1,000 population per year. Differences between healthcare systems make direct comparison of laboratory confirmed cases between countries difficult but many other high-income countries with comparable surveillance systems do nonetheless show broadly similar rates of reported laboratory confirmed cases[40] with the exception of New Zealand where rates are approximately five fold higher[41].

1.2.2 Demographic pattern

1.2.2.1 Age and sex

Incidence rates of laboratory confirmed infection in the UK are highest in young children (age under 5). Older children (aged 5 to 14) have the lowest rate. Young and middle-aged adults also have a relatively high incidence of reported infection, with a lower incidence of

reported infection in the oldest age groups (Figure 1.2). A similar pattern of high incidence in young children in Manitoba, Canada is confined to rural children[42]. These strong patterns support comparison of subtypes between age groups within this thesis.

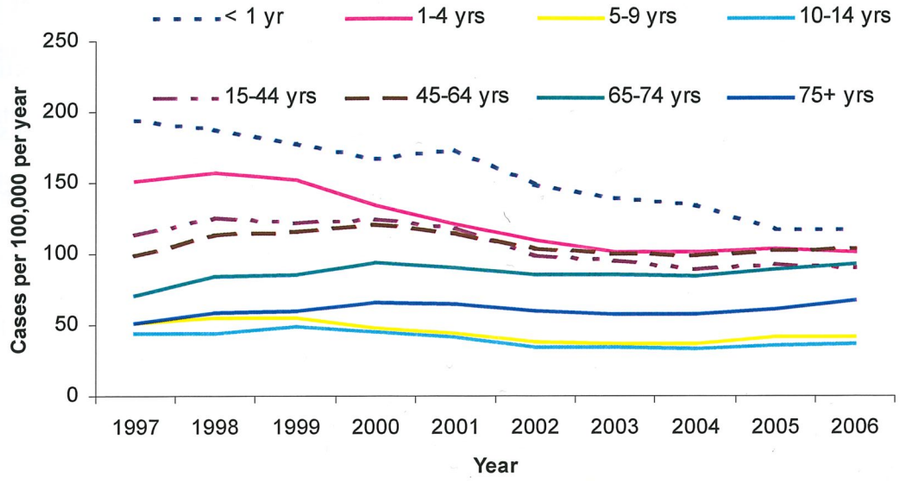


Figure 1.2. Laboratory reports of *Campylobacter spp.* from faecal isolates reported to the Health Protection Agency Centre for Infections. Cases per 100,000 population by age. England & Wales, 1990-2006.[43]

A higher incidence of *C. jejuni* infection in males than in females is usual, including in large studies[44,45] although the effect is fairly small. Some reports suggest that this male dominance varies with age but results of these more detailed analyses are less consistent. In one UK study substantial male excess in childhood led to the overall effect despite higher levels in females in adolescence and young adulthood (15-24), with older adults showing similar incidences in males and females[46]. Other studies in England[47], Norway[48], Canada[42,49] and Scotland[50] report a stronger male bias among young children, but not a reversal in other age groups while one study reports a dip in incidence among older girls and young adult females (age 5-19) compared to younger girls which dip is absent among boys of this age[51]. A review of US and international data supports a male excess of reported cases extending up to ages in the range of 30-40 but not above[25], which may be due, at least in

part, to demographics. Variation in incidence by gender is thus a weak effect that is inconsistent across different studies and populations.

1.2.2.2 Ethnic group

The Campylobacter Sentinel Surveillance Scheme (CSSS) in the UK reported that the Pakistani population had relatively high rates of laboratory confirmed *C. jejuni* infection, Indian and black ethnic groups relatively low rates and the white and other populations an intermediate level[52]. The age distribution varied by ethnic group and the excess incidence among Pakistani relative to white populations was entirely driven by very high rates among children while adult rates were less than those in the white population. A similar finding of more childhood cases in those of Asian (mainly Indian subcontinent) ethnicity and lower rates among adults was reported from Birmingham in a study not distinguishing among different Asian ethnicities and where the overall rates of reported disease in the white and Asian populations were similar[53]. The adult Asian population at the time of the study were mainly first generation immigrants. The basis for these apparently marked differences in the age distribution of laboratory confirmed *C. jejuni* in different ethnic groups living in England is uncertain. Findings of higher incidence in immigrant children have also been reported from Norway, where travel associated infection may be a major contributing factor[54].

1.2.3 Small scale geographic effects

Many studies have reported incidence differences between rural and urban populations. The earliest report to address this showed higher incidence in childhood in rural areas and in adults in urban areas in Scotland[50], with raw milk exposure causing disease and subsequent immunity suggested as a possible explanation. This study covered the period before pasteurisation of milk was made compulsory. A Canadian study from the same period identified elevated rates among farm dwelling but not non farm-dwelling rural residents.

Analysis of risk factors suggested that the excess of cases in this group appeared to be mostly explained by drinking raw milk[49]. An English comparison of rates in the catchment areas of different laboratories showed an excess incidence in laboratories with a mainly rural population across all ages[47]. More recent studies showing an excess of cases in rural populations in Denmark[55] and Canada (Manitoba)[42] showed that the excess was mainly confined to children. In the Danish study this made only a small contribution to the overall population incidence.

In contrast, higher rates have been found in urban populations in Norway[48] where the excess was mainly associated with increased incidence of travel related infections and the Netherlands[44] where travel related infection was also a contributor to the urban excess. In New Zealand excess reported disease in urban settings was mainly due to a more pronounced summer peak with similar incidence during the rest of the year[56]. A separate analysis of New Zealand data showed that a small overall urban excess could be broken down into a more marked urban excess for adults and a higher childhood incidence in rural areas[41]. Reported urban - rural differences have therefore generally been modest but may indicate some particular risks to children in rural areas, particularly in the past when consumption of unpasteurised milk may have been more important.

1.2.4 Summer peak and seasonal variation

A seasonal peak during summer was an early finding in the study of campylobacteriosis[49,54] and has been described widely since then[57,58] with a sudden very sharp increase, often doubling the weekly incidence of reported infection over the course of 2 weeks at the same time each year[57]. More complex patterns of seasonal incidence have also been described. Since the seasonality of *Campylobacter* infection has been a striking yet poorly explained phenomenon, and one that may be susceptible to the inclusion of a reliable subtyping

approach within epidemiological studies as proposed in this thesis, a rather detailed review of the existing literature describing seasonal patterns and possible mediators of these patterns is given here.

1.2.4.1 Summer peak

A distinct seasonal peak has been described in many temperate countries including Canada[49,58], the Czech Republic[58], Denmark[57,58], England[58,59,60], Finland[57,61], Norway[48,54], Iceland[62], the Netherlands[58,63], New Zealand[56,57,58], Scotland[57,58,64], Sweden[57], Switzerland[58] and Wales[57]. The peaks described in most Northern European countries have been sufficiently pronounced that at least 10% of cases occur in the fortnight spanning the peak[57] with some exceptions such as Estonia[58]. Summer peaks were less reliably identified in warmer countries in one international comparative study[58] consistent with single country studies in Israel and Portugal[65,66] where levels may in fact be higher in winter and Italy[67] and Spain[68] where no obvious seasonality was seen. Exceptions include summer peaks reported in Pakistan[69] and Curacao[70].

The timing of the peak can be relatively early such as in Wales at around week 22 (late May / Early June) or later with a sharp peak in Finland occurring around week 30 (August). In general there have been later peaks in colder countries[57,58] with Scandinavian countries peaks around 8-11 weeks, and Scotland 4-7 weeks after the peak in Wales. However the reported timing has been less consistent in some temperate countries (e.g. Germany[57]) and the peak less distinct in many[57].

Varying seasonality within a country has been observed in New Zealand with the seasonal peak in the rural parts of North Island being lower and less persistent than on South

Island[56,58]. Urban areas had a definite seasonal peak on both islands[56]. Variation in seasonality was less evident between regions in England[45].

1.2.4.2 More complex seasonal patterns

The shape of the summer peak has varied between countries, with differences in shape between countries generally much more striking than year-to-year variation for individual countries. The initial sharp peak has fallen off relatively quickly in Finland[57], and Wales[57,58], while the observed decline has been more gradual in Scotland[57,58], and Denmark[57,58]. In other countries an initial sharp rise has been followed by a further increase from this higher summer baseline such that the highest incidence has been late in the summer with the Netherlands[58,63], Switzerland[58], and the Czech Republic[58] offering clear examples of this pattern. The more complex seasonal patterns suggested by these comparisons have been analysed further in Scotland and the Netherlands. The Scottish analysis suggested that alongside the main June/July summer peak there was a smaller preceding April peak and subsequent October peak. When the national data was broken into regions clear differences were evident with very little seasonal variation in Greater Glasgow compared to other areas such as Lothian (which includes Edinburgh) while the reduction following the summer peak was slower and the autumn peak higher in Lothian than elsewhere. These differences were not formally tested[64]. An earlier study in Scotland considering campylobacteriosis reports from 1978 – 1982 noted a more erratic seasonal pattern overall and marked differences between urban and rural populations, with a first quarter peak in rural populations and a third quarter peak in urban. The rural first quarter peak was particularly prominent among younger children. However, at least three milk-borne outbreaks had a substantial effect on the winter (first quarter) figures at this time when raw milk consumption was common and these findings may not be relevant to current epidemiology[50]. In the Netherlands peaks were identified in January, early July and

August/September, with this last peak representing the highest incidence. While the first two peaks were consistent across the country the third showed distinct differences between Dutch regions. Although it occurred approximately 1 week after the school holidays in each region, with school start dates varying be region, it affected both adults and children[63].

1.2.4.3 Seasonal effects in different demographic groups

A study based in four laboratories in England noted that the summer peak was more prominent in young men and speculated that servicemen participating in outdoor exercises may have contributed to the effect[47] since the catchment populations included military bases. A more recent and larger analysis identified that the summer peak is earlier and stronger in children aged 1 –4 years of age, with the adult peak an average of 2 weeks later. While children aged 1-4 had a fairly discrete early summer peak followed by a decline, infants (aged under 1) showed further smaller peaks later in the summer[45]. This analysis also showed some trend for the peak in men to be slightly earlier than for women. The laboratory specimen date was used in this analysis with no data allowing to assess whether time from symptom onset to sampling differed by age or gender.

1.2.4.4 Mediation of seasonality

The reason for the seasonal pattern of human *C. jejuni* disease is not understood. Published studies can be divided into those that focus on climate and others focussed on a range of more specific factors.

Climate

Many other gastrointestinal pathogens show seasonal variation in incidence[71]. The most similar pattern is that of *Salmonella* infection with a summer peak that closely mirrors ambient temperature[71,72]. However, although the impact of temperature related growth of *Salmonella* on food might be important this is not believed to occur with *C. jejuni*[73,74] for which species the peak often occurs earlier than peak ambient temperatures in any case[57].

Several studies have described more complex associations between ambient temperature and the incidence of *C. jejuni* infection, although the relationships described differ. These include findings that countries with colder winters have later peaks[58], that average temperature 4 weeks earlier is the best predictive climatic factor (with some prediction by amount of sunlight and none for humidity or precipitation)[75]. Two studies explicitly model season and still find a residual association with temperature[76,77]. In one of these the positive association is observed up to 14°C but not at higher temperatures[77]. Although climate may thus be an important background factor none of the findings to date is very useful in determining the mechanism by which the seasonal peak occurs.

Specific factors associated with seasonally varying risk

In theory seasonal factors affecting host immunity could produce such patterns. There is no evidence for such host factor variation and it is not considered further here. In broad terms the rise in disease could be due to (i) new seasonally restricted sources of infection, (ii) an increase in a subset of the usual year-round sources, or (iii) an increase originating across the range of usual sources. In each case increased human infection could be due to increased populations of *C. jejuni* in, or improved transmission from, these sources to humans. The evidence for a seasonal, or expanded seasonal, role for the range of sources of human infection evaluated to date is summarised below. Seasonality in particular niches and host species is described in the section on the ecology of *C. jejuni*.

Isolate characteristics. Two studies from the Netherlands describe lower resistance to quinolone antibiotics during the summer peak, which may indicate different origins of infection at this time. In one study this was a fairly minor effect[44] but in the other it was more marked with 22% resistance during summer months compared to 33% during the rest

of the year[51]. These findings thus support a hypothesis that at least the relative contributions of different sources may vary seasonally, and that subtyping may provide insight into this.

Travel. The seasonal peak in Norway was more pronounced when travel associated cases were excluded[48], the peak in UK countries is early relative to peak holiday travel and the marked peak in young children at this time[45,57] are examples of evidence to show that the peak is not driven by travel although travel related infections may contribute to observed seasonal patterns in some countries[58].

Environment. Direct infection from the environment during summer has been suggested with outdoor activities reported by far more cases in summer than at other times in a UK study[46]. Deliberate consumption of, or recreational contact with, untreated water has been identified as a risk factor for *C. jejuni* infection[78,79]. However, the effects described are not sufficient to explain the summer peak by increased contact with water and moreover, contamination levels in surface waters decrease during warm summer months[59,80]. Direct contact with animals may increase and contribute to the summer peak. This has not been evaluated.

Wild birds. Bird pecked milk has been identified as a risk factor for *C. jejuni*[81]. A marked increase of exposure to bird pecked milk has been reported among cases in the UK in May and June with the additional finding that those reporting this exposure tended to have infection with a particular serotype (HS6)[46] suggesting that this may contribute to the summer peak. An earlier small case-control study also identified bird pecked milk as a risk factor in late spring[82], This exposure does not have the capacity to explain the peak fully or across countries with a wide range of milk distribution systems. Seasonality of infection in

wild birds is not well described other than an autumn peak in black headed gulls in Sweden, which occurred later than the human peak[83].

Poultry. There is evidence for increased contamination of poultry in the summer months. However the peak in human infection has been reported to precede the peak in chicken meat[57,84,85], so that the role of increased poultry contamination in seasonality is uncertain.

Raw milk and untreated water. Outbreaks from milk and waterborne sources in the US have been identified more in spring and autumn than during the summer peak period[25] and drinking raw milk was identified as a risk factor for apparently sporadic *C. jejuni* infection outside but not during the seasonal peak in the Netherlands[86]. This may therefore indicate an even greater unexplained proportion of cases during the summer peak if some sources contributing cases at other times of year are reduced during the peak.

Food preparation. Barbecuing has been identified as being associated with *C. jejuni* infection only during the peak period in one Dutch study[86] and may therefore contribute to the summer peak both by an increasing rate of barbecuing and increased risk per exposure event at this time. However this comes with a caveat that case and control interviews had a time lag which may have biased the results, in particular regarding factors showing variation over time such as barbecuing.

1.2.4.5 Conclusion on seasonality and questions arising for this study

Seasonal variation varies between different countries and sometimes within countries. In those places where seasonal effects are strong they appear stable over the timescales of several years for which they have been studied. Although several factors such as increased

contamination of poultry, increased human exposure to milk from bird pecked containers and increased direct contact with the outdoor environment may contribute to the seasonal peak the aetiology of this peak is not clear and the similarity or difference of sources for infections during peak and trough periods has not been established. Several studies have suggested that the subtype distribution of incident infections may vary seasonally using serotyping and antimicrobial susceptibility as an index of type. Integration of sequence based subtyping may help to explore the basis for seasonal variation in human *C. jejuni* infection.

1.3 Attributing human disease to source

The relative contributions of different sources and transmission routes can guide disease control programmes for foodborne zoonoses that inhabit more than one animal host or environmental reservoir[87]. The attribution of *Campylobacter* using typing data has proved difficult despite the application of a wide range of approaches[86,88,89]. This contrasts with considerable success in microbial source tracking for human *Salmonella* infection[90,91]. This section summarises what is known regarding the source of human *Campylobacter* infection based on a range of approaches to determining source and what potential these approaches have to further clarify the origin of human infection. A particular focus is the role of subtyping in attribution and the potential and difficulties for sequence typing to make a contribution.

The main approaches to identifying the sources of human infection are

1. Risk assessment
2. Analytical epidemiology
    a. Outbreak investigations
    b. Investigation of incident cases
3. Natural experiments and other sources of information.
4. Microbial typing

1.3.1 Risk assessment

Quantification of exposure of humans to particular sources of infection, by for example estimating the extent of contamination and frequency and intensity of human interaction with particular sources, is the essence of a risk assessment approach to predicting the importance of these potential sources of infection. Risk assessment models require either extensive data or assumptions for each source of interest for all of the following: degree of contamination with the pathogen of interest; frequency of human contact with the source; risk of transmission per contact; and likely virulence of the transmitted organisms. One published report describes a project that attempted a quantitative overall prediction of source for human campylobacteriosis (based on a consideration of 31 possible sources). This study assembled the available information on the hazard of various potential sources, the routes by which humans can be exposed and the risk of infection and disease due to each exposure. It concluded that the data gaps are too great to allow reliable conclusions on several counts[89]. For example, the dose-response model was judged inadequate for general use in risk assessment, there was substantial uncertainty regarding the level of contamination of many sources and the routes of transmission from contaminated sources to humans are not well characterised.

Thus although risk assessment models considering a single source of a pathogen can be useful in guiding and monitoring control interventions from a particular source[92] and there is extensive work for *Campylobacter* on the risk from poultry meat and factors determining this[93] and the potentially important impact of cross contamination in the kitchen[94] the risk assessment approach has not contributed to overall source attribution for human campylobacteriosis. Quantification of differing virulence between identifiable subtypes is an area in which subtyping may contribute to future risk assessment for this pathogen.

1.3.2 Outbreak investigation

Outbreak investigations have been used to define the major causes of an infection, based on the argument that the better quality of evidence of source arising from outbreak investigations than for investigations of unrelated single sporadic cases allows more reliable ascertainment of source in this setting. However this assumes firstly that sporadic and outbreak cases share the same source, secondly that sufficient data is available from investigated outbreaks to allow reliable estimation, and thirdly that outbreaks from different sources are similarly likely to be identified. There is some evidence against the first assumption since *C. jejuni* outbreaks in the USA, which were mainly driven by water and milk associated cases, occurred more in spring and autumn with an intervening peak of cases not associated with outbreaks in the summer[25] and some evidence for a small October peak of outbreaks associated with commercial food premises in the UK[95,96]. The second assumption has been highlighted as a particular weakness in relation to *Campylobacter*[97] where outbreaks are relatively rare compared to the overall burden of disease. For example in the UK 21 reported outbreaks between 1992 and 1994 comprised only 0.2% of reported cases[95] and between 1995 and 1999 the 50 reported outbreaks accounted for just 0.4% of cases[96]. Whether the third assumption holds is speculative, but this assumption is further vulnerable if reliant on published reports since a substantial publication bias appears to operate with published outbreaks not generally representative of incident outbreaks[98].

Accepting these caveats, which mean that outbreaks are unlikely to be a reliable quantitative guide to the source of human infection, they may still allow to identify some recurrent sources of human infection. The source of infection in the UK review of 21 outbreaks between 1992 and 1994 identified water as the origin in six. In all cases supplies were private and in some issues regarding poor maintenance, inadequate chlorination and identified contamination events were identified[95]. Water was identified as a source in four

of the 50 outbreaks in the later review (1995-9), all involving small water supply systems in rural locations[96]. In a review of 111 outbreaks between 1978 and 1996 from the USA[25] 12 were waterborne. These included eight larger (community) water supply associated outbreaks causing an estimated 5,068 cases in total. In these larger outbreaks disruption of chlorination or contamination of unchlorinated groundwater supplies were implicated. In the USA waterborne origin was more common in earlier years with only three in the second half of the study. This appears to be supported by the relatively greater importance of waterborne origin in the earlier UK survey, although numbers are too small to draw strong conclusions.

In the USA study a further 30 outbreaks (27%) were milk associated, generally due to consumption of unpasteurised milk. Of these 25 occurred in the first half of the period studied. In the 1992-4 UK review five outbreaks (24%) are milk associated. Three were due to unpasteurised milk, one to inadequate pasteurisation and one to contamination of pasteurised milk by birds. In the later UK review just two (4%) outbreaks were associated with milk or milk products.

A wide range of foodstuffs was reported in other outbreaks where a source was identified. The most common was poultry identified in four outbreaks in the early UK study (including chicken liver mousse in one), in 14 in the later UK study (13 chicken and one duck) and three in the US series (two chicken and one turkey). Restaurants and hotels were identified as the setting for 16 and ten outbreaks respectively among the 50 in the second UK review, which combine to give 52% of *C. jejuni* outbreaks associated with hotels or restaurants compared to 15% of reported outbreaks from other pathogens over the same period. Identification of food handling faults in 26 food associated outbreaks included faults likely to cause cross contamination in 18[96].

Outbreaks reported from other industrialised countries have also mainly resulted from contamination of water, milk or chicken and with evidence of cross contamination in some investigations[25]. Outbreak evidence therefore indicates that consumption of water from vulnerable supplies and raw milk have been important, though probably decreasing sources of *Campylobacter* infection, that poultry meat is the most commonly implicated food and that identified outbreaks are often linked to hotels and restaurants where specific procedural failings may contribute to risk. All of these factors may also be important contributors to cases not belonging to identified outbreaks.

Regarding the ambitions of this thesis, the lack of identified outbreaks for such a common infection is striking and subtyping may have a role in improved outbreak detection if this lack is substantially due to failed identification rather than non-occurrence of outbreaks. If subtypes are predictive of source then comparison of subtypes in outbreaks with those from apparently sporadic cases will allow to identify whether cases in outbreaks represent the same sources of infection, a subset, or substantially different sources to non-outbreak cases. Clarification of whether typing information can predict source would therefore support the establishment and analysis of sample collections from temporospatially matched sporadic and outbreak associated cases to identify the extent to which these cases share similar sources.

1.3.3 Analytical epidemiology of incident cases

Analytic epidemiology has been the main approach contributing to the published literature for source attribution in *Campylobacter*. The case-control design is the most commonly used one with exposure histories from cases compared to histories from control individuals who have been selected to represent the population from which the cases arose. In theory

analytical epidemiology allows the quantitative attribution of foodborne and other disease to different sources. The potential for bias resulting from inaccurate classification of exposures, selection of controls that do not represent the underlying study population or imprecise definition of cases of disease is an important limitation of the method. Most case-control studies have not distinguished between *C. jejuni* and *C. coli* so that the results mainly refer to the more dominant *C. jejuni* but may be affected by associations with *C. coli.*

The following section summarises a review of analytical studies published until 1998[86] and then reviews similar studies published since that time (up to December 2007). Studies restricted to children are considered separately given the high incidence of *Campylobacter* infection in young children. The aim of these summaries is to identify what analytical epidemiology infers regarding source of human infection, to assess methodological issues affecting the quality of evidence to date, to assess the scope for future contributions from this approach alone and if modified by the integration of sequence typing information.

1.3.3.1 Previously reviewed epidemiological studies

Eleven out of the 16 published case-control studies conducted between 1979 and 1998 and reviewed by Neimann *et al.*[86] identified a positive association between Campylobacteriosis and exposure to chicken meat (Table 1.1), making it the most frequently identified risk factor. In seven studies a variable indicating general or domestic consumption of chicken was associated with disease whereas in others only particular categories of chicken exposure showed significant association, such as eating under-cooked chicken, restaurant chicken, or chicken liver. Given the wide and incompletely described range of chicken variables investigated in each study the results are difficult to interpret quantitatively.

Other risk factors identified in more than one study are also listed in Table 1.1. Like chicken, although pets are identified often the particular pet exposure implicated varies between studies. There were two reports of disease associated with puppies and only single reports of other particular pet exposures. Similarly for exposures to farms or animals and to water, although many studies identified an association with these classes of exposure the precise exposure varies between studies. In contrast, drinking either raw milk (5 studies) or bird pecked milk (3 studies) are frequently identified risk factors that appear to represent the same association independently identified in different studies.

Table 1.1. Summary results of a review of *Campylobacter* case-control study results from 16 studies. (summarised from Neimann *et al.* 2003.[86])

Exposure	Studies*	Comment
Chicken	11	11 have some form of chicken exposure with 3 identifying in particular under-cooked chicken, 2 commercial/restaurant consumption, 1 barbecued chicken and 1 chicken liver.
Raw milk	5	
Pets	5	3 dogs of which 2 specify puppies, 1 cats and 1 any pet with diarrhoea
Water	4	2 water from environmental source, 1 rainwater used at home and 1 small domestic supplier.
Bird pecked milk	3	
Farm / farm animals	3	1 each of living on farm, contact with cattle, contact with chicken.
Pork	2	
Barbecued food	2	

*** Number of studies reporting this association.**

1.3.3.2 Studies since the last published review

Eleven more recent studies (ten case control studies and one environmental factor analytical study) reported between 1999 and December 2007 are reviewed here and summarised in Table 1.2. Seven were conducted in EU countries, two were from the USA and one each based in Australia and Curacao. One of the case control studies used registry data only from 22,036 cases[55] while the other nine, involving active data acquisition, included a total of

3,193 cases. Only these nine studies reported on food consumption and individual animal and environmental exposures with the other two limited to population level factors covering, for example, water systems and environmental measures.

Six of the nine studies assessing individual level exposures reported an association between Campylobacteriosis and one or more chicken meat exposures, with domestic consumption associated in three, undercooked chicken in three and consumption at a restaurant or takeaway in three. Contact with farm animals was associated with disease in five studies with two specifying chicken contact and one cattle contact. The two studies reporting an association with companion animal exposures both report an association with puppy ownership. Water associated exposures include consumption of outdoor water in 2 and having a dug well as a domestic supply in 1. Positive associations each reported in two studies were drinking unpasteurised milk, eating at a barbecue (eating red meat in one of these), eating pork (undercooked in one), and eating meat (other than chicken) at a restaurant. Single reports included eating offal, avoiding offal, eating salad vegetables, eating grapes, eating raw seafood, drinking bottled water, and swimming in natural outdoor water. The highest population attributable risks for any factors estimated were 24% each for consumption of chicken at a restaurant[99] in one study and avoiding consumption of offal in another[86]. At an ecological level campylobacteriosis is reported as being positively associated with length of water supply pipe and living in an area with high ruminant density, although such findings are particularly prone to uncontrolled confounding.

Table 1.2. Review of case control and environmental analytical studies not included in review by Neimann *et al.*[86]

Study, location, period, size (cases/controls)	Design and particular features	Risk factors	Measure of effect * OR [95% CI] PAF%
Rodrigues *et al.*[100] 2001 UK Aug 1993 – Jan 1995 229/229	Population based. Age, sex and GP practice matched control. 22 day median delay between case and control interviews. Time covered not specified but is for 10 days.	Chicken at restaurant Boiled rice at home Pulses Fruit (skin on, none, dry) Salad at home Pasteurised foods Raw eggs in home desserts	1.8 0.5 0.3 0.2 – 0.4 0.3 0.5 0.3
Neimann *et al.*[86] 2003 9 of 16 Dutch counties May 1996-May 1997 217/236	Age, sex and county matched controls. Postal questionnaire. Posted questionnaire then checked by phone. Exposures in 2 weeks before illness (case) and 2 weeks preceding questionnaire (control). Control interviews 118 days later (mean) than case interviews.	Undercooked poultry Pork chops Barbecue red meat Grapes Unpasteurised milk Avoid eating offal Eat pig offal Thaw poultry Scald cutting boards Organic milk Unpeeled apples and pears	8.2 [1.1-63.1] 5% 4.4 [1.4 – 13.7] 6% 4.1 [1.5-10.9] 13% 2.8 [1.4-5.9] 12% 11.8 [2.0-70.3] 4% 2.6 [1.2-6.0] 24% 0.2 [0.1-0.8] 0.2 [0.0-0.9] 0.3 [0.1-1.2] 0.4 [0.2-0.8] 0.2 [0.1-0.4]
Friedman *et al.*[99] 2004 US Foodnet sites (7 states) 1998-1999 1152/ 1152	Population based matched 1:1 on age band and telephone exchange by systematic dialling protocol. Analysis restricted to age 2+ and non-travel cases. No *coli* – *jejuni* differentiation. Case / parent telephone interview within 21 days of specimen regarding 7 days before onset of diarrhoea. Control within 7 days of case and for same time period.	Chicken at restaurant Undercooked chicken Chicken at home Fried chicken Turkey at restaurant Turkey at home Other meat at restaurant Other meat at home Ate raw seafood Raw milk Water from river/lake Pet puppy Farm animal contact (2-11 yo) Farm animal contact (12+ yo) Contact animal faeces Fresh berries	2.2 [1.7-2.9] 24% 2.1 [1.2-3.4] 0.7 [0.6-0.9] 0.5 [0.3-0.6] 2.5 [1.3-4.7] 4% 0.5 [0.4-0.8] 1.7 [1.3-2.2] 21% 0.7 [0.5-0.9] 1.9 [1.1-3.4] 3% 4.3 [1.3-14.2] 1.5% 3.3 [1.5-7.5] 3% 3.4 [1.8-6.5] 5% 21.0 [2.0-178.0] 2% 2.0 [1.2-3.6] 4% 1.4 [1.0-1.9] 6% 0.6 [0.5-0.9]

Study, location, period, size (cases/controls)	Design and particular features	Risk factors	Measure of effect * OR [95% CI] PAF%
Nygard *et al.*[101] 2004 Sweden 1998-2000 7007 cases / registry data	Ecological analysis at municipality level. Routine datasets combined in GIS	Ruminant density per 10 ruminants per km sq Per 10% increase in public water supply Per 10m increase in water pipe length Mean temperature	1.12 [1.09-1.15] 0.93 [0.90-0.95] 1.13 [1.09-1.17] 1.02 [1.01-1.04]
Endtz *et al.*[70] 2003 Curacao Dec1999-March 2000 106/108	Catchment area (population) based. Controls selected with lab negative faecal specimens. Controls selected after cases identified. Timing not clear.	No associations including animals, foods, well water, food preparation.	
Kapperud *et al.*[102] 2003 Norway 3 counties with 1,061,800 pop. Jul 1999- Jun 2000 212/422	Matched case control study. Postal questionnaire checked by phone. Exposures in 2 weeks before illness (case) and 2 immediate preceding 2 weeks (control). Control interview median 41 days (mean 47) after onset date. Case med 27 days after). Note that over 50% of cases drank untreated water.	Drink untreated water Eat at barbecue Eat poultry bought raw Contact with animals (occupational) Eat undercooked pork Eat mutton Eat fruits and berries (per time) Swim is lake/sea/pool	2.5 [1.2-5.4] 4.1 [1.7-9.9] 1.4 [1.0-2.0] 19.3 [1.7-222.3] 37.0 [1.6-831] 0.4 [0.2-0.8] 0.9 [0.9-1] 0.7 [0.5-1.0]
Ethelberg *et al.*[55] 2005 Denmark 1991-2001 22,066/ 318958	Population registry study. Excluded cases identified as travel related. 15 age matched controls per case. No *coli* – *jejuni* differentiation. Restricted to registry data – no patient recall. Only registry information available. Rural defined as most rural with 5.2% of population or most rural 20%. Differences mainly due to those in children.	Rural area Private vs public water	OR 1.3 [1.2-1.4] 1.0 overall but varied by county.
Potter *et al.*[103] 2003 Michigan Oct 2000-Oct 2001 83/122	RURAL communities. Matched 1:2 by age, sex and county using telephone exchange matched controls by systematic dialling protocol. Fortnight before symptoms (cases) and questionnaire (controls). Cases reported up to 30 days after symptoms. Postal questionnaires.	Poultry husbandry	6.9 [1.4-33.0] 18%.

Study, location, period, size (cases/controls)	Design and particular features	Risk factors	Measure of effect * OR [95% CI] PAF%
Evans *et al.*[104] Wales Jan – Dec 2001 213/ 1,144	Catchment area population case control. Controls were negative sample patients. Questionnaire sent on specimen receipt before diagnosis. Equivalent case-control timing. Bias possibilities from other diseases. Very different demographics etc. Questions 7 days before onset except antibiotics (1 month).	Chicken Fried chicken outlet Bottled water Cattle contact Salad vegetables (tomato, cucumber)	1.8 [1.2-2.7] 1.9 [1.1-3.4] 1.4 [1.0-2.0] 5.1 [1.3-19.7] 1.5 [1.1-2.2]
Stafford *et al.*[105] 2007 Australia Sep 2001- Sep 2002 881/833	Population lab catchment in 5 of 8 states. Systematic sample within labs. 1:1 broad age band match with controls from previous national gastro survey volunteers. Only age 5 or older. Cases telephone interviewed within 30 days of illness regarding 7 days prior to illness and controls within 30 days of cases regarding 7 days prior interview.	Ate cooked chicken Undercooked chicken Offal (any species) Own puppy Domestic chickens Vegetable (5+ types) Ate fresh fish Foods with raw egg Organic fruit and veg. Home grown fruit	1.4 [1.0-1.9] 4.7 [2.6-8.4] 2.0 [1.0-4.0] 3.0 [1.1-4.2] 12.4 [2.6-59.3] 0.2 [0.1-0.5] 0.7 [0.5-0.9] 0.5 [0.3-0.8] 0.6 [0.4-1.0] 0.4 [0.3-0.6]
Schonberg-norio *et al.*[106] 2004 Finland July-Sept 2002 100/137	Multicentre (3 lab catchments). Matched. 99 of 100 were *jejuni*. Only univariate results given. Note that the three positive associations shown persist in multivariate but not indicated for protective factors. Cases answered within 37 days of symptoms. Controls identified after and approached 2 at a time if non-response. Period interviewed not reported.	Ate/tasted raw meat Water from dug well Swimming in natural water Ate berries Ate black/redcurrants Yoghurt Pasteurised milk Carrots Cooked fish Beef liver Water from large plant Ate at friend's house	12. 0 [1.5-93.8]] 3.2 [1.6-6.5] 2.3 [1.2-4.2] 0.4 [0.2-0.9] 0.2 [0.1-0.4] 0.4 [0.2-0.9] 0.4 [0.2-0.9] 0.4 [0.2-0.8] 0.4 [0.2-0.9] 0.2 [0-0.9] 0.5 [0.3-1.0] 0.4 [0.1-1.0]

* Effects given are Odds ratios [95% confidence intervals] and percentage population attributable fractions (PAF) where available.

Exposures reported to have a negative association with disease are fully listed in Table 1.2. These included eating fruit (5 studies reporting fruit generally or in particular peeled, home grown, currants or fruit and berries), vegetables (4 studies reporting vegetables generally or salad, organic or carrot as the particular vegetable exposure), berries (fruit and berries in 1 study and berries in 2), and offal (2 reports one pig offal and one beef liver). Although these quantitative associations with apparently protective factors were almost as numerous as the risk factors their interpretation is much more difficult. They may be the result of confounding as markers for particular dietary habits, the result of chance given the large numbers of exposures assessed, or have a true causal relationship. Only relatively strong protective effects would be expected to be detectable in case control studies. It is difficult to see which biological mechanisms could be operating to produce such strong effects for so many factors and therefore unlikely that many of them represent true causal relationships. If these apparent protective effects were identified to be associated with infection with different subtypes of *C. jejuni* it would reduce the probability that the effects were due to chance and may allow to investigate whether these fit with a plausible biological mechanism or are more likely due to confounding. This may be an argument for integrating subtype into case control studies, if subtype can be used to predict source of infection.

1.3.3.3 Epidemiological studies in children

The results of case control studies among children are summarised in Table 1.3. The association of chicken consumption was less evident than in adult studies being reported in just one study, and then only in an analysis restricted to children aged one or older. In contrast the importance of water consumption in both the US and Swedish

studies (though highlighted as a negative result in one Australian study[107]) and the association with farm environment and pet animals are greater than seen in adult studies. Farm animal contact has also been highlighted as more important among children in a large general population study in the US[99]. Specific association with live chickens in the home in the two Australian studies was supported by univariate analysis results in the Swedish study with the multivariate result not given but reported to be non-significant. These findings regarding animal exposures come with the caveat that the positive associations varied between studies. Some reported an association with dogs of any age, some only puppies, and some dogs with diarrhoea while null results for these same exposures were reported from other studies. Direct exposure to raw meat was identified as a risk factor in the US in the retail setting and in Spain in the home. The US, Swedish and very small Spanish studies analysed subgroups, which analyses suggested changing risk factors with age but sample sizes were too small to formally test variation by age group.

A negative association with breast-feeding was reported in two studies from analyses restricted to those aged less than one. A third study in those aged one to five reported no effect. Eating sausage (Sweden) and Salami (Australia) was associated with reduced risk of disease. Swedish sausage is not usually eaten raw[79] and so these appear to be separate negative associations rather than independent observations of an effect.

Table 1.3. *Campylobacter* case-control studies in children in industrialised countries.

Country	Australia	Australia	Sweden	Spain	USA
Period	1996	2000	2001/2	1998-00	2002-4
Cases/controls	81/144	172/173	126/270	28/84	123/928
Age	0-3	1-5	0-6	0-3	0-1
Setting	Popn.	Popn.	Popn.	Hospital	Popn.
Reference	107	108	79	109	110
Risk factors	Odds ratios (population attributable fraction)				
Animal / environmental contact					
Dog	5.2*		3.8 (30%)		
Pet with diarrhoea					5.3 (6%)
Own puppy	16.6				
Chickens at home	11.8	4.5	5.4 †		
Living /visiting farm			6.6**		4.1 (16%)
Contact with meat					
In shopping trolley near meat					2.2 (7%)
Chicken or mincemeat in kitchen	4.1				
Protective factors					
Breast feeding	0.1‡	No effect			0.3
Sausage			0.05		
Salami		0.5			
Cheese		0.33			
Food exposures					
Grilled meat			2.1* (27%)		
Home prepared fruit and veg					2.0 (28%)
Porridge		2.2			
Butter		1.9			
Mayonnaise	4.1				
Chicken				8.8¶	
Water consumption					
Water					
Well water	- ve		2.6 (19%)		2.6 (14%)
Lake/river water			6.0 (19%)		
Raw milk			3.7* (7%)		

*$0.05 < p < 0.1$
† on univariate analysis only
‡ In analysis of those under 1 year old
¶ Aged 1 or more
** Univariate analysis only. Text describes living on a farm with animals as risk. The tables give odds ratios for contact with ruminants, which is the odds ratio given here.

1.3.3.4 Summary of findings and issues from analytical epidemiological studies

Overall there has been strong evidence from analytical epidemiology for chicken, direct animal contact and consumption of contaminated water as risk factors in both older and more recent analytical studies. Raw milk is also a risk for *Campylobacter* but less commonly identified in more recent work, while bird pecked milk, identified

in three separate earlier studies has not been reported in later studies. Evidence for other exposures is more variable. There is some evidence that risk factors among children are not identical to those in adults and may indeed vary with among different childhood age groups.

The reviewed analytical epidemiology literature does not give a clear picture of the proportion of disease arising from different sources. Firstly, although some studies presented results in quantitative attribution terms and some others gave sufficient information to calculate this but many did not. Secondly, differences in study design including exposure ascertainment make combination of information across studies difficult. Thirdly, although the observed variation in results where studies did report on similar exposures may be affected by true variation in risk factors at different times and in different places, the marked methodological differences and limitations of studies reported, some of which limitations are inherent to the case-control may also be important. Table 1.2 summarises features of the design and conduct of those analytical studies that I reviewed individually, i.e. those published between 1999 and 2007. Important methodological issues that could affect the validity of reported studies and their fitness to act as a basis for quantitative attribution include differential timing of exposures from cases and controls, imperfect representation of the population giving rise to cases by controls, uncontrolled confounding, and recall bias as summarised below.

Period of exposure history

Cases were generally interviewed concerning the period before disease while control interviews covered a variety of time periods as early as the time period before the case became ill and as late as the period immediately before the interview or questionnaire.

In most studies the controls were interviewed much later than the cases, with the mean difference being 118 days in the most extreme case[86] among the studies that reported this information clearly. Any seasonality in exposures or risk due to particular exposures will be strongly affected by this bias. The only questionnaire study to avoid this bias completely was one where test patients with negative faecal samples were analysed as controls for *Campylobacter* infection[104]. This source of controls was also used in another study where differential delay was not avoided[70].

Representativeness

The two studies using population registries compare cases and controls arising from the same population with controls fully representative of the population from which cases arose[55,101]. In contrast the two studies using those submitting *Campylobacter* negative samples as controls[70,104] did not try to fully represent the underlying population and have biased control groups relative to this population. In the other studies using individual level data the sampling frames used aimed to represent the population, but bias may have resulted from non-response among controls. The percentage and pattern of non-response is often not reported fully, limiting the capacity to assess how well the control groups represented the populations giving rise to cases.

Control of confounding

Although most case-control studies have been matched on a range of criteria including age the nature of this matching varied substantially from controls with birth dates within a week of cases[86] to quite broad age band matching in others[99] and more complex matching with tight matching among younger participants and weaker matching with increasing age[100,111]. No attempts are described in attempting to control

for selection bias. Most studies used multivariable analysis to control confounding due to measured exposures but the approaches used and the extent to which these analyses were reported varied between studies.

Recall bias

Non-differential recall bias is a particular limitation of many of these studies given the long gap between exposure and data capture, and the recording of exposures over lengthy periods relative to the at risk time that would be estimated from the incubation time of *C. jejuni*. Although the incubation time for *C. jejuni* is usually 2-5 days the shortest exposure histories were for 7 days preceding onset of symptoms in three studies[99,104,105] with periods of 10 days[100] and 14 days[86,102,103] also used and some studies not clearly specifying the period covered[70,106]. Such delayed histories covering overlong exposure periods will likely dilute the association between causal exposures and disease.

1.3.3.5 Conclusion from analytical epidemiology

Published analytical epidemiology studies have given a guide to some of the main exposures but are not able to provide a basis for quantitative attribution to source or to evaluate the consistency of risk factors over time or between populations. Many of the practical and methodological limitations are difficult to overcome and this approach seems unlikely to allow reliable quantitative attribution on it's own. The role of subtyping within these studies depends on the extent to which subtype varies between different sources of infection. If this effect is strong then studies on risk factors for particular subtypes may allow a more reliable picture with far less noise than results from studies seeking to identify the source of all *C. jejuni* or a combination of *C. jejuni* and *C. coli.* In addition, identification of the relationship between apparently

protective exposures and subtypes may clarify whether these are true effects or artefacts.

1.3.4 Natural experiments

Two "natural experiments" have occurred to allow estimation of the likely importance of poultry as a source of *Campylobacter* infection at a population level.

1.3.4.1 Belgium

In June 1999 Belgian poultry meat and eggs were withdrawn from the retail market in Belgium. Belgian produced poultry held a 59% market share at the time. A 40% reduction in the incidence of Campylobacteriosis was observed compared to expectations derived from a Fourier time series model of data from the previous 5 years. An extrapolation from a 40% reduction in *Campylobacter* infection due to a 59% reduction in poultry on sale would predict that 68% of disease is due to poultry. This linear extrapolation depends on assumptions that sales of imported poultry were unaffected (they could have either decreased or risen) and that *Campylobacter* contamination in locally produced and imported chicken is similar. The true share of human *Campylobacter* infection due to poultry in Belgium can therefore be stated confidently at more than 40% with a best, but more guarded, estimate of 68%.

1.3.4.2 Iceland

Iceland is self sufficient for poultry production[112]. Levels of laboratory confirmed *Campylobacter* infection were low until the late 1990s[62,112] when the rate per 100,000 population rose from 13.8 in 1997 to 52 in 1998 and 116 in 1999. Poultry consumption had risen markedly over this time. In addition poultry was increasingly retailed fresh rather than frozen and monitoring programmes at slaughter showed substantial infection of flocks with *Campylobacter*. Poultry was therefore seen as a

major contributor to the increase[112]. A range of interventions included increased on farm biosecurity, altered slaughtering practices including freezing of positive flocks and public education regarding safe poultry handling. The rate dropped to 33 per 100,000 in 2000. A small rise in travel associated *Campylobacter* infection at this time suggests that reduced ascertainment is not the reason for this change and most of the 71% reduction may represent reduced risk from poultry, although which interventions were most important is uncertain[112].

1.3.4.3 Conclusion from natural experiments

These two natural experiments indicate that contaminated poultry meat contributed to a substantial proportion of human infection in these countries at these times. A testable hypothesis, if type is informative of source, is that the types infecting humans during the phases of reduced infection would comprise more non-chicken types relative to the preceding time periods. This highlights the importance of developing subtyping methods to predict origin of infection to allow for fuller exploitation of such events in determining source of infection.

1.3.5 Microbial subtype

The use of microbial subtype is the other main approach to attribution of infectious gastroenteritis[87,92]. A feature of this approach, which can be a limitation, is that it identifies the original (typically zoonotic) source and does not usually help in assessment of the route of infection, such as cross contamination during processing. Within this restriction the approach is well developed for *Salmonella*. Initially this was dependent on some subtypes occurring almost exclusively in particular sources and this assumption continues to be used in the application of this approach in the Netherlands[92]. A Bayesian framework has been used in Denmark to allow attribution with the less restrictive assumption that different sources harbour different

distributions of *Salmonella* subtypes, and allows modelling of, for example, differential virulence and growth rates in the food chain[90]. In combination with the availability of substantial data on the distribution of subtypes in different sources and incidence by subtype in humans it has been possible to estimate the source of human infections probabilistically with regard to food animal species and whether domestically produced or imported[90]. The results in the Netherlands and Denmark are broadly similar[92]. Addition of antimicrobial resistance data has allowed more detailed analysis of imported versus domestic origin[113]. The existence of robust, easily applied and internationally standardised phenotyping methods which can identify clones with particular host association underpins the success of this approach.

No similar success has been reported with *C. jejuni*. Phenotyping is complex, more difficult to standardise and no robust host species associations have been described. There is no published evidence for reliable host association of phenotype. Some host association of genotype has been identified that indicates that subtype based attribution may be feasible[114] but no robust methods of indexing associations have been described and no attempt made or method developed to use the observed host association to attribute human disease to source. This approach has not therefore contributed effectively to the identification of sources of infection with *C. jejuni* to date and addressing this gap using multilocus sequence typing is a main focus of this thesis.

1.3.6 Summary of risk factors

A detailed overview of risk factors for *Campylobacter* infection in humans, based on exposures identified by the formal methods outlined above, and on any other quantitative evidence of association reported in the peer-reviewed literature is given

in Annex 1 considering host factors, food types, routes of exposure to foodborne infection, water, environmental and animal exposures, occupational risk and person to person spread. It is not included here since it does not add new questions that are addressed in the thesis.

1.3.7 Integrating genetic sequencing into source attribution

Reviewing existing approaches to determining the source of human infection highlights that the cornerstone of using sequence based subtyping to supplement and support identification of the sources of human infection is the identification of a reliable approach to attributing human infection to source based on subtype. This would allow direct use of typing data to attribute to source and also support:

1. The impact of outbreak investigation by clarifying the extent to which outbreaks and sporadic cases represent common sources of infection and guiding source identification during outbreak investigation.
2. The integration of subtype information in analytical studies of incident *C. jejuni* infection to allow more targeted determination of subtype specific risk factors and to evaluate apparently protective factors.
3. Opportunistic identification sources of infection through exploitation of phenomena producing sudden and substantial changes exposure of people to particular potential sources of infection.
4. Risk assessment approaches by reducing the information gaps that restrict the ability to develop useful overall risk assessment models.

The ability to predict source by subtype required to provide these benefits depends firstly on the biology of *C. jejuni* and secondly on available typing technologies, which are reviewed in the two remaining sections of this introduction.

1.4 *Campylobacter jejuni* biology

This section provides a general overview of the biology of *C. jejuni* with particular reference to those aspects that impact on the epidemiology of human disease and on the use of microbial subtype to support the investigation of this epidemiology.

1.4.1 Phenotype

The Gram-negative, motile, spiral-shaped bacterium *Campylobacter jejuni* has width and length varying from 0.2–0.9 μm and 0.5–5 μm, respectively. The motility of *C. jejuni* is mediated by flagella, usually a single polar flagellum which, together with the spiral cell body imparts a fast rotation around the axis. The bacteria are microaerophilic, do not metabolise sugars, and use intermediates of the tricarboxylic acid cycle as sources for energy. Isolates will usually grow readily under microaerophilic conditions (5% oxygen) at 37°C and at 42°C[2,115].

Fastidious growth requirements and lack of robustness to environmental stress[73,74] result in poor growth on foodstuffs in contrast to, for example, *Salmonella.* This limits the likelihood of outbreaks occurring due to bacterial multiplication near the point of consumption. Nonetheless the organism can persist on foodstuffs[73] resulting in high levels of contamination of some widely distributed foods[85,116].

1.4.2 Taxonomy and genome

C. jejuni belongs to the genus *Campylobacter* in the epsilon group of proteobacteria[117]. Many species within the genus are associated with diarrhoeal disease in humans and other animals[117]. The *C. jejuni* genome is approximately 1.7 megabases in size[118-120]. It is naturally competent and recombination is important in generating and maintaining diversity[121,122]. This high degree of recombination needs to be considered in designing and interpreting typing approaches.

1.4.3 Ecology of *Campylobacter jejuni*

1.4.3.1 Hosts and environmental niches

Most *Campylobacter* species have established relationships with a particular range of hosts[123-126]. Among human pathogens a broad host range is particularly characteristic of emerging diseases[127] and *C. jejuni* is striking for the broad range of hosts from which it has been isolated. These include many species of wild birds[126,128,129], farm animals[129-134], and companion animals[23,135] all of which have been relatively extensively studied, but also more exotic animal species, although these have not been studied to the same extent. A zoo based case series over one year included isolates from non-human primates, ungulates, felids, a red panda, birds, and a reptile[136] and other isolations from wild animals such as muskrats[137], rabbits[138] and hares[139] have been reported. More recently isolation has been demonstrated from amoebae[140,141]. Although *C. jejuni* can infect this very wide host range, high prevalence of infection and persistent carriage is less widespread so that true reservoirs for the species may be more restricted. Carriage prevalence is usually low and carriage duration probably short in humans as reviewed in the description of human asymptomatic infection (Section 1.1.3). Point prevalence was also low among for companion animals[135], although prevalence may be higher in young cats[23]. Higher prevalence of carriage has

been reported for a wide range of wild birds[126,142] and farmed ungulates[130,143-148] and higher levels still in gallinaceous (e.g. chicken)[132,133,149], anseriform (e.g. duck)[128,150-152] and ratite (e.g. ostrich)[134] birds where carriage rates exceeding 50% have been described. Within farm animal niches, which are the most fully studied, findings have included early infection of immature animals including cattle[145,146] and chicken with a decrease in infection with increasing age[149]. Co-infection with multiple lineages of *C. jejuni* appears common[153-155]. The intensity of infection also varies substantially by host species. Infection has been shown to spread rapidly within chickens flocks[156] and infected birds to excrete high concentrations of *Campylobacter*[149] whereas levels of excretion have been lower in infected herds of ungulates[130,146,157]. This high level of carriage in the various types of fowl, combined with extensive contamination of carcases occurring at slaughter[158] has been reflected in contamination of meat at retail sale with levels of contamination of the order of 50% for retail poultry[85,116] whereas prevalence of contamination of red meat has generally been 10 fold lower[116,159-161]. In addition to the prevalence of infection, a high intensity of contamination, as estimated by quantitative microbiology, has been identified repeatedly on retail poultry[89,149,162].

C. jejuni has been isolated from many environmental sources including fresh water[163] and the sand of bathing beaches[164]. Survival within free-living aquatic amoebae[140] has been suggested as a possible environmental reservoir facilitating infection of chickens[141]. The role of water in human infection is also clearly identified[86,95,96]. However there is no evidence for replication in water and it is not clear whether this or any other environmental niche is important as a reservoir or as a vehicle for transmission between species in the ecology of the *C. jejuni*. Contamination of flies[165,166] and beetles[167,168] has been described and this includes visceral contamination[165,168]. One intervention study reported a reduction in the incidence of

infection in broiler houses following interventions to reduce entry of flies[169]. Although it is possible that other effects of the intervention studied led to the reduction rather than an impact via reduction in fly exposure the combination of some evidence from an intervention study and repeated demonstration of survival of *C. jejuni* in flies highlights insects as potential vectors for transmission. Sustained replication in insects has not been reported and there is no evidence for or against insects acting as a reservoir.

1.4.3.2 Seasonality in non-human hosts

Seasonal patterns have been described in many animal and environmental niches. Spring and autumn peaks in prevalence of faecal excretion of *C.jejuni*, with an intervening summer dip have been described in the UK dairy cattle but not in beef herds.[130,146]. A small study in dairy cattle in New Zealand showed a lower rate of carriage in winter than in summer or autumn[170]. Consistently high rates without seasonal variation have been described in sheep[157]. No obvious seasonality was evident in red meat at retail[161]. Most studies of chicken and retail poultry have found a marked increase in prevalence of infection during the summer months. A peak in the prevalence of faecal carriage at slaughter between June and September was reported from a large study in the Netherlands[132], an approximate doubling of the proportion of flocks testing positive for *Campylobacter* during July and August has been described in two studies in Denmark[75,171] while August was the month of highest prevalence in a Norwegian study of broiler flocks[172]. Although one UK study of chickens[131] and one small survey of retail poultry[173] did not find any seasonality of *Campylobacter* carriage other UK studies found a rise in early summer in chickens[174] and in retail poultry where a three year study showed a rise in summer in each of three years with

an additional December peak in one year[85]. In the US contamination levels in turkey at slaughter in two plants were more reliably high in winter months, although relatively high levels were sometimes seen in July. The majority of samples in this study were *C. jejuni* but seasonality is assessed using combined *C. jejuni* and *C. coli* data[175]. A study of black headed gulls peak in Sweden showed the prevalence of *C. jejuni* to be highest during autumn[83].

C. jejuni was most frequently cultured from samples of river water taken during winter[59] and least during summer[80] in North West England. Some other studies have shown autumn peaks but generally environmental water contamination and survival decreases with increasing heat and sun[176]. This is consistent with experimental evidence for decreased survival of *Campylobacter* in the presence of ultraviolet light[59]. Just one study, from New Zealand, has identified a summer (as well as a winter) peak in the proportion of river water samples to test positive for *Campylobacter*[177]. Contamination of sewage appears to be most common in summer in studies from the UK[178] and Italy[179]. Further investigation in the UK suggested that this peak was driven by a rise in *Campylobacter* contamination of the effluent from an abattoir rather than changes in human sewage[180].

Seasonal variation is thus well studied in chicken where it shows a similar pattern to human seasonality. This is not true of all other niches with, for example, environmental water showing an opposite trend and little seasonal variation in red meat contamination with *C. jejuni* at point of sale. Many species and niches are less studied and there is a lot of residual uncertainty regarding seasonality in many species including most wild birds.

1.4.3.3 Host association within *C. jejuni*

Given the wide and distinctive host and environmental range of *C. jejuni* there is substantial scope for ecological structuring of this bacterial species across niches, including between host species. Little is known regarding circulation of *C. jejuni* between host species and in the environment[163], and by extension about barriers to gene flow across these boundaries. Greater clarity regarding the extent of host species association, whether genetically related hosts harbour more similar *C. jejuni* populations, whether animals in similar ecological niches share similar types and the importance of geographic structuring and temporal change relative to these other characteristics would provide a more robust foundation for understanding the ecology of this species and supporting monitoring and control. Such identification of population structure within a species requires the ability to index subtypes of the species and this aspect of *C. jejuni* ecology is therefore described in the next section, which reviews the approaches to and applications of subtyping.

1.5 Typing of *Campylobacter jejuni*

This section summarises the literature on subtyping *C. jejuni*, and the applications of this subtyping in order to provide the final component of background to the main focus of the thesis on integrating genetic sequence information and population biology approaches into the investigation of the epidemiology and ecology of this species. A wide range of approaches to subtyping *C. jejuni* have been developed and applied as outlined below. Identical types have been used as evidence for a common recent origin in attempts to detect outbreaks, to define which cases belong to outbreaks detected by other means, and to assess whether different clonal groups dominate particular niches. Source prediction has been undertaken on both local and

larger scales and has involved attempts to predict source of human and chicken infections. Although there has been a clear progression from phenotyping to non-sequence genotyping to sequence typing all methods continue to be used. Particular attention is given in this section to the utility of different typing approaches in contributing to epidemiological and ecological questions and to a thorough treatment of multilocus sequence typing, which approach is used in the work of the thesis.

1.5.1 Phenotyping

1.5.1.1 Laboratory methods of phenotyping

Serotyping[181,182] and phage typing[183-185] are the main phenotypic methods applied separately or in combination[185]. As well as being technically demanding, poor and variable typeability (the proportion of isolates for which a type can be determined) is inherent to the serotyping methods applied to *C. jejuni*[186]. Reproducibility has also been poor and although some modifications have reduced the contribution of laboratory limitations to this problem[187] serotype is itself relatively unstable and can change during the course of a single host infection[188]. Serotype is not strongly linked with more global assessments of phylogenetic relationships[189] suggesting that it does not act as a reliable indicator of clonal relatedness in *C. jejuni*. Lack of discrimination, with just 10 serotypes including 53% of isolates in one large study[187] is a further major limitation for discriminatory purposes. This may in part be due to the geographical and temporal restriction of the sample used to define the origin panel of serotyped isolates for the most widely applied Penner scheme[182], which limits the diversity indexed among these reference serotypes. Increasing the number of phages used in phage typing schemes has improved the typeability and discrimination offered by this approach[184,185]. Combination of phage typing with serotyping improves discrimination and this combination became the UK reference laboratory approach[185].

1.5.1.2 Applications of phenotyping

Despite limitations to phenotyping in this species it has been widely applied and has produced some substantive contributions. In studies considering pathogen factors and outcome of infection several serotypes with increased risk of GBS complicating infection have been reported[190,191], indicating a possible role for serotype itself or another factor with similar population genetic structuring in the aetiology of this complication. An investigation into infecting type and demographic characteristics showed non-random patterning of *C. jejuni* serotype by geography and age. However this analysis did not differentiate between fixed geographical differences or heterogeneity due to links between clustered cases and used statistical approaches that did not account for such plausible sources of variance[78], making it difficult to interpret the origin of the effect as being due to clustering or another form of population structuring. Several studies have assessed ecology in terms of host association, in some cases including human isolates. A comparison of serotypes from wildlife isolates (hedgehogs, squirrels, other mammals and birds, n=47), broiler chickens (n=29) and humans (n=44) showed lesser differences between human and broiler isolates than between either of these species and the wildlife collection. Some serotypes were present in all three groups and further differentiation by genotypic analyses supported these isolates as being clonally related[192]. An early assessment of serotype among both *C. jejuni* and *C. coli* in Norway, using a different scheme, also suggested that human isolates were closer to those from poultry than those from wild birds, flies or pigs, but with relatively minor differences[129]. Studies in New Zealand have shown similar serotype distributions in humans at different times and identified the same serotypes frequently in farm animals, particularly ruminants, but less commonly in wild ducks[193,194]. A UK study on serotypes in humans noted an

increased rate of reporting drinking bird pecked milk (12%) in those with Penner serotype HS6, but this was the only positive association from many hypotheses tested and test statistics did not consider multiple comparisons[46]. Use of serotyping to determine the origin of infections in a large study in chickens (3,304 isolates) was limited by lack of variation with 3 serotype complexes accounting for 58% of isolates and 10 for 85%[195].

1.5.1.3 Summary of phenotyping and its application

Phenotyping has thus provided evidence that at least one complication may vary with subtype, that there is partial differentiation between *C. jejuni* in different host species, that human types resemble those isolated from chicken and other farm animals more than they do isolated from wild animals, and that there may be overlapping populations between different host groups. No phenotyping study has reported quantitative attribution to source and the reported overlap in apparently clonally related isolates from different species would prove a complication to attribution approaches. In part due to lack of comparability, with different or modified schemes being used as well as difficulties in between laboratory standardisation, no joint analysis has been made of different studies. The limited capacity for joint analysis is a barrier to this approach making a more substantial contribution.

1.5.2 Genotyping other than sequence typing

A wide range of methods assaying variation across the genome involve the generation of DNA fragments using restriction enzymes (e.g. Pulse Field Gel Electrophoresis (PFGE)[186,196]) or amplification of DNA (e.g. randomly amplified polymorphic DNA (RAPD)[186]) or a combination of both such as amplified fragment length polymorphism (AFLP)[197]. These approaches generate band patterns representing the genome wide distribution of the binding sites of the enzymes or primers used[186].

Among methods focussed on a more restricted part of the genome several distinct approaches to polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) typing of the *flaA* gene[186] or other genes[198] have been developed giving much more detailed information but from just a single short section of DNA. More recently, microarray methods have been developed based on hybridisation with short gene fragments[199].

1.5.2.1 Pulse field gel electrophoresis

Standardisation and reproducibility have not been achieved effectively for the genome wide approaches[186] with the partial exception of PFGE where it has been maintained in the USA through PulseNet[200,201]. Although there are concerns that PFGE type is quite unstable and can change substantially due to genome rearrangement in a single infection[188,202] and *Campylobacter* phylogenies inferred from PFGE data may be unreliable (with, for example, isolates from different species clustering together[203]) it is by far the most widely used of these approaches and is considered the gold standard for discriminatory typing[201,204].

1.5.2.2 Application of pulse field gel electrophoresis

Outbreak investigation

In two outbreaks, which had been identified by other means, PFGE was reported as useful in identifying whether cases belonged to the outbreak[205,206]. Three studies performed to evaluate PFGE have confirmed this capacity to confirm cases as belonging to an outbreak by comparing a range of sporadic isolates and isolates from several outbreaks[204,207,208]. The most extensive use of PFGE to support outbreak investigation has been within the USA PulseNet, where more than 2,500 profiles have been identified using standardised methods[201]. A consensus view from this experience is that it is a useful approach to support the investigation of detected outbreaks but not

useful as a means to detect *C. jejuni* outbreaks[201,209], in contrast to the PulseNet findings for other bacteria where PFGE has been used to detect outbreaks[201]. While PulseNet collaborators have used the SmaI restriction enzyme others have proposed that KpnI or a combination is more discriminatory[194,210]. A New Zealand study[194] and a Canadian[211] each studied 183 cases. The New Zealand study was entirely retrospective using available data. It identified clusters by genotyping some of which genotypic clusters also showed temporal clustering (not formally analysed) and some genotypes may also have clustered by area of residence and reported occupation. In this retrospective study further investigation was not done to assess if these cases had true epidemiological links. The authors conclude that similar PFGE patterns following the more discriminatory KpnI digestion may identify epidemiologically meaningful outbreaks arising from a common source. The earlier Canadian report included both a retrospective and prospective component. Outbreaks identified by other means were confirmed as being of identical type. However clusters identified by typing did not yield evidence for a common source on further investigation. Since most clusters were small (2 to 3 cases) identification of source was difficult and the authors conclude that following such small genotypic clusters is not justified[211].

Source attribution

Groups of PGFE types in human cases appear to vary over time and according to patient characteristics[212], which may indicate evidence for different sources. Host association of subtypes has been examined in several studies. In Sweden isolates from wild birds, broilers and humans were analysed jointly by a UPGMA cluster analysis. Two main clusters were produced. Sixty of 73 gull isolates were in one genotypic cluster, mostly in an almost exclusively gull-origin sub-cluster, while human and chicken isolates were more widely scattered and generally closer to each other than to

gull isolates[83]. A comparison of 49 human isolates with 87 from a range of wild birds identified only two isolates in wild birds that were similar to those from humans, one each from a starling and a blackbird[213]. A Finnish study of isolates from chickens and humans during the seasonal peaks between 1996 and 1998 showed a wider range of types in chickens that mainly persisted across the three years. Fewer types were isolated from humans with more variation from year to year. However most human isolates represented types also identified in chickens[214]. The Danish phenotyping study with a wide range of wild animal as well as human and broiler isolates also conducted PFGE and identified fourteen clusters with similar PFGE patterns. Seven of these 14 subtype clusters included human isolates[192]. Four of the subtypes with human isolates also included broiler isolates. The only wild animal or bird isolates (hedgehog, fox, deer, squirrel, duck, buzzard, gull, sparrow hawk, hare, seal, pheasant, rook, raven, magpie or seagull) to share a subtype with human isolates was hedgehog. There was some sharing of subtype between broiler and wild animal and between different wild animal species with four wild animal species (fox, squirrel, duck, buzzard) identified as carrying the most widely distributed type. These results are consistent with the overall results of phenotyping in this sample. A New Zealand comparison of isolates from humans, water, chicken and veterinary origins sampled during winter (August) and summer (February) found clear differences between the two time periods. Human isolates mostly shared subtypes with either chicken or water subtypes. No attribution analysis was performed and the tabulated data suggests that little accurate attribution would be possible. A study in Curacao combined case control epidemiology with PFGE of isolates. Generally there was no association between type and risk factors with the possible exception of those owning a deep domestic well being more likely to have infection with one PFGE pattern[70]. A large

Canadian broiler study (93 lots of chickens on 57 farms) showed marked diversity including diversity on individual farms but with some identical types at different periods and on different farms. Approximately 20% of human genotypes were identical to clones identified among poultry[215]. No other host samples were available.

1.5.2.3 Summary of PFGE and its application

Like phenotyping PFGE indicates that there is partial differentiation of *C. jejuni* populations by host species, and that human isolates are closer to those from broilers than of wild animal origin. Also like phenotyping, combined analysis of the many PFGE studies described above is not possible given lack of standardisation. Although outbreak identification has not been well supported by PFGE, the method appears useful in including or excluding cases from outbreaks already under investigation. This poor performance in outbreak detection suggests either that a very highly discriminatory methods may be needed to support outbreak identification for this pathogen or that outbreaks may really be very rare events.

1.5.2.4 Amplified fragment length polymorphism

Five AFLP studies have assessed host association and possible sources of human infection. A Danish study considering possible sources of human infection reported that most clones infecting humans were also isolated from several non-human animal sources making it difficult to infer source based on type[216]. A UK study found relatively marked differences between pig and poultry *C. coli* but no clear clonal separation of *C. jejuni* in samples from poultry, cattle, swine and humans although relative prevalences differed among some AFLP defined clones[189]. A Swiss study identified quite marked separation between poultry, pets, cattle and humans and some evidence for cattle associated clones identified on multiple farms. Most (47 of 64) human isolates in this study were in human only clones. Of two Dutch studies, one

including only humans and chicken isolates identified 3 main clusters, each of which included isolates of both human and chicken origin with no cluster restricted to either host[197], while the second, including human, chicken and cattle isolates found no convincing evidence for host association with AFLP[88]. The information on type provided by AFLP has not therefore been shown to be useful in predicting source of human infection. The approach has not been evaluated for other epidemiological applications.

1.5.2.5 Single gene restriction fragment length polymorphism

There are at least seven *flaA* RFLP schemes in use[186]. Relatively frequent genetic events can substantially alter *flaA* genes[207,217] and, as a consequence, these genes have minimal relationship to the clonal history of an isolate and even show evidence for substantial cross-species recombination[218]. As a result, although allowing a rough prediction of whether or not isolates are identical, *flaA* RFLP are unlikely to index other epidemiologically or ecologically important information.

1.5.2.6 Microarrays

More recently microarray approaches have allowed the capture of more detailed genome wide information, such as gene content, by assaying whether an isolate has very similar genes to those from a fully sequenced isolate on a gene by gene basis[199]. This approach uses hybridisation with a cut point determining whether the isolate under study is classified as having a gene that is sufficiently similar to the probe on the microarray. Early arrays are being superseded as increasing numbers of completed genomes are available allowing a better index of the *C. jejuni* pan-genome. Integrating data produced from different generations of arrays will provide a further challenge. In common with the genome wide approaches that produce band patterns, there is a lack of good models describing the genetic processes producing the

observed patterns of results, in this case gene gain and loss. This limits the quality of inference available from microarray data[219]. One study has reported that this approach identified a livestock and non-livestock clade with most human cases fitting into the livestock clade[199]. However, this study lacked an adequate sample for non-human isolates, used an array based on a single genome study, and did not take account of the population structure already identified by sequence typing methods[114,220] the importance of which unbiased population structure has been shown for other bacterial species[219]. To date microarray analysis has thus not contributed usefully to the understanding of *C. jejuni* population structure or epidemiology.

1.5.2.7 Summary of non-sequence genotyping and it's application

Among non-sequence based genotyping PFGE has been the most useful approach, and has supported case definition in outbreak studies as well as giving relatively consistent results in studies of source attribution suggesting a large role for farm animals, and poultry in particular, in the origin of human infection. All methods have suffered from lack of comparability between studies and none appears likely to allow efficient combined analysis between studies, outbreak detection, quantitative source attribution, or substantial insight into the processes leading to the observed bacterial population structure and biology.

1.5.3 Sequence typing

This section reviews the application of sequence typing to *C. jejuni* with particular detail on multi locus sequence typing and the findings from this approach, which is applied in the later work of this thesis.

1.5.3.1 Sequencing single highly variable gene fragments

Sequencing of a variable portion of the *flaA* gene is the most common single gene typing approach to date[217,218,221], although other genes have also been used[222]. Two factors limit the inference that can be drawn from *flaA* sequencing. Firstly, identical alleles of the *flaA* gene fragment have been found in *C. jejuni* and the related species *C. coli*[218]. Secondly, substantial change can occur in a single recombination so that closely related isolates can have very different *flaA* sequences[217]. Sequence based approaches have been useful in demonstrating explicitly these extreme limitations of any single gene system in a frequently recombining species. However the implication is that sequencing of the *flaA* gene without a wider genotypic context has no substantial advantage over non-sequence typing of this gene. In other bacterial species systems using sequences from two or more variable genes[223] or one or more variable genes in the context provided by more stable typing approaches such as multilocus sequence typing (MLST)[224] have been applied and are more reliable in both discriminating between closely related isolates and avoiding the false identification of apparently close relatedness between distant isolates sharing a common sequence at a single locus. Fragments of the *flaA* and *porA* genes are used in this project alongside a *C. jejuni* MLST scheme (Figure 1.3) to support discrimination.

1.5.3.2 Whole genome sequencing

The first full genome sequence of a *C. jejuni* strain was reported in 2000[119]. Since then other full sequences have become available[118] and the increasing availability of high-throughput sequencing approaches means that whole genome sequences may soon be feasible even for large sample collections[225]. During the work of this thesis this was not feasible on a large scale and this approach is not reviewed other than to

acknowledge that it may be available soon and should be considered when assessing useful extensions to the approaches presented in this thesis.

1.5.3.3 Multilocus sequence typing (MLST) of *C. jejuni*

MLST of bacteria involves sequencing the forward and reverse strands of several, typically seven, gene fragments[220,226,227]. Gene fragments are chosen to allow (1) typing of all, or virtually all, members of the target group of bacteria and (2) identification of genetic relatedness between isolates using analysis based on genetic processes involved in evolution. The genes chosen code for essential metabolic functions. This means that they are likely to be present in all isolates, assuring the aim of typeability. A consequence is that the genes are under stabilising selection which limits the diversity available from each gene fragment and therefore lowers the efficiency of identification of differences between isolates. The use of seven genes provides sufficient information to allow isolates to be grouped according to a moderately detailed genealogy. Closely related isolates, sharing identical alleles at several of the loci, are thus identified as belonging to a particular clonal complex (CC) or even having an identical sequence type (ST) across all 7 loci. Isolates identified as belonging to the same clonal complex are very likely to share a more recent common ancestor with each other than with isolates which are not members of the clonal complex[226,228].

Analysis of MLST data can take advantage of data from genes at multiple separate loci on the genome to identify overall population structure, the place of an isolate in that structure, and to estimate the contribution of different effects such as mutation, recombination (swapping of segments of DNA between lineages of bacteria), selection and neutral processes such as genetic drift to the generation of genetic

diversity, maintenance of species genetic characteristics and the shaping of population structure[229,230]. This capacity for analysis using models based on the understanding of basic genetic processes contrasts with typing schemes using empirical typing data where no clear model of the processes producing change is available.

Three MLST schemes for *C. jejuni* have been published[121,220,231]. One of these has become widely applied[220] while the others have not been used in subsequently published work to date. The large scale and increasing application of a single scheme[61,88,114,123,138,158,204,207,218,231-255] since first described in 2001[220] allows direct comparison between, and combined analysis of data from, different studies using this standard approach. This is further supported by an accessible public database[256]. The position on the genome of the seven gene fragments used in this scheme, which has also been applied to *C. coli*[218], are shown in black writing in Figure 1.3[218,220]. The size and the detailed position of each gene fragment on the *Campylobacter* genome is indicated in Table 1.4.

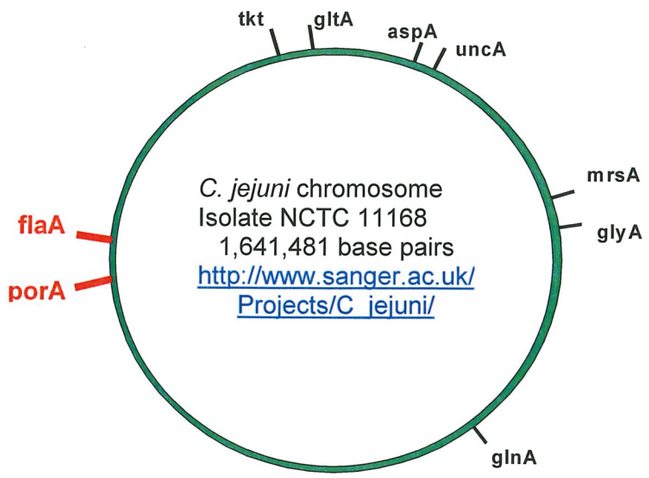


Figure 1.3. Position of multi locus sequence typing loci[220] (in balck) and antigen genes[257] (in red) on the NCTC11168 genome.

Table 1.4. Location and size of MLST[220] loci on NCTC11168 *C. jejuni* genome

MLST gene	Identifier and position on NCTC11168 genome		Base pairs
aspA	Cj0087	96692 to 97168	477
*uncA**	Cj0105	112163 to 112651	489
mrsA / pgm†	Cj0360	327773 to 328270	498
glyA	Cj0402	367573 to 368079	507
glnA	Cj0699c	657609 to 658085**	477
tkt	Cj1645	1569415 to 1569873	459
gltA	Cj1682c	1604529 to 1604930**	402

* *uncA* renamed to *atpA* and now identified as such on the NCTC1118 genome.
† *"pgm"* most similar to *E. coli mrsA.* Currently not named on the NCTC1118 genome having been previously identified as both *pgm* and *mrsA*.
** *glnA* and *gltA* are coded on the negative sense DNA strand

Several papers have described single nucleotide polymorphism[236,253,258] or other[241,246] methods to shortcut sequencing requirements in obtaining groupings similar to MLST defined population structure. These have generally produced a rapid approach capable of moderate accuracy. However they are not capable of indexing further diversity, which is inherent to standard application of MLST though new allele discovery and the identification of new allele combinations. There are no published reports of the practical application for scientific or public health purposes of these short cut approaches and they are not considered further in this thesis.

The reported applications of MLST to the population biology, ecology, and epidemiology of *C. jejuni* are reviewed below. Throughout the rest of this thesis MLST refers to the most widely applied scheme[220] unless otherwise specified.

1.5.3.4 Application of MLST to the population genetics and biology of *C. jejuni*

Population genetic parameters

Analysis using MLST has shown that *C. jejuni* is weakly clonal. This was reported using index of association analyses in the first report of the widely used MLST

scheme[220], with largely consistent results from analyses in the other two schemes that have been developed[121,231]. Initial estimates for the basic population genetic parameters suggested that recombination events were eight-fold more common than mutations with an average fragment length estimated at 3.3 kb[88]. Later model based population genetic analysis of data from a structured sample of a local area and it's animals[138] estimated the recombination rate as being similar to the mutation rate in *C. jejuni* and the mean recombination fragment size as in the order of 500-750 bp long and similar in each host species studied[259]. Although conclusions about the lack of evidence for large recombination fragments in this study used an erroneous gene order this is unlikely to have made a major contribution to this estimate so that current combined evidence supports mainly relatively short recombination fragments, on average in the order of hundreds of base pairs.

Population structure and diversity within *C. jejuni*

A comparison of MLST and MLEE among a diverse set of *C. jejuni* isolates confirmed the similarity of grouping obtained by each method[235] as expected from the design of MLST[220,226]. The large collections of isolates which have been MLST typed have demonstrated substantial diversity with 3360 unique STs identified in the combined *C. jejuni* / *C. coli* database (at 21 February 2008) and 43 defined clonal complexes. The most diverse clonal complexes within *C. jejuni* are ST-21 complex including 332 STs and ST-45 complex with 184 (at 21 February 2008). MLST has also been used to describe the phylogenetic relationship of the subspecies *C. jejuni* subsp. *doylei* showing that a geographically diverse set of *C. jejuni* subsp. *doylei* isolates formed a phylogenetically distinct cluster within *C. jejuni*[252] consistent with *C. jejuni* subsp. *doylei* isolates sharing a more recent common ancestor with each other than with other *C. jejuni* isolates.

Gene flow between species

The relationship between *C. jejuni* and *C. coli* has been analysed using the common MLST scheme[218]. This has shown the presence of hybrid forms, that these result from recombination between *C. jejuni* and *C. coli*[88,218,232], that such recombination may be particularly frequent among isolates obtained from ruminants[232], and that it can be found in isolates on different continents[245]. Recombination with other species appears to be less common[232] but does occur with evidence for lateral gene transfer between *C. jejuni* and *C. lari*[123] and unidentified cross species imports have been reported among Danish *C. coli*, as well as imports from *C. jejuni*[248].

1.5.3.5 MLST and the distribution of genotypic and phenotypic characteristics

Population structure indexed by MLST has been used along with other genotype information to describe the population structuring of *porA* antigen gene variants[251], to assess the stability of the flagellar antigen genes *flaA* and *flaB*[207], and to identify that gene content differences mainly map to clonal population structure rather than host of origin[250] in a subtractive hybridisation study using isolates from an earlier study[138]. The complication of Guillain-Barré syndrome has been identified in association with genetically diverse isolates in several MLST defined clonal complexes[260], consistent with earlier findings using MLEE[261]. Most studies considering antibiotic resistance in the light of MLST defined population structure have focussed on *C. coli* where they have shown patterns such as a clonal distribution of erythromycin resistance due to a 23S rRNA mutation[262] and susceptibility to other antibiotics[242] among isolates from turkeys in the USA and contrasting findings of a lack of strong clonal patterning of resistance to a range of antibiotics among *C. coli* from swine in the USA even when the resistance had the same genetic basis[263] or of quinolone resistance in isolates from

chicken in Senegal[245]. The associations in these papers were observational rather than based on formal analysis of resistance in relation to population structure.

1.5.3.6 MLST applied to the ecology of *C. jejuni* subtypes

The seven published studies describing the distribution of isolates by clonal complex across multiple hosts, and one restricted to chicken isolates are summarised in Table 1.5. The table only includes the most commonly identified clonal complexes. ST-21 complex and ST-45 complex appear to be relatively widely distributed across many host species as well as abundant. Ruminants and chicken are the most usual sources, with ruminant slightly more dominant in ST-21 complex and chicken in ST-45 complex. However these farm animals are the most sampled sources and isolation has also been reported from wild birds, other animals and environmental samples for both clonal complexes. ST-206 complex isolates have been reported frequently from both chicken and ruminant sources but not from wild animals or environmental samples. The importance of ruminant hosts for ST-61 complex, ST-42 complex and to an extent ST-48 complex is well supported across studies in different European countries, while ST-257 and ST-354 complexes are somewhat more common in isolates of chicken origin. ST-177 and ST-179 have been mostly reported from sand isolates, although three ST-177 isolates from starlings and one from a pet have also been identified. One study restricted to chicken isolates is included as it is substantially separated geographically from other isolates[240]. These Senegalese chicken isolates mostly fit into the chicken associated complexes (257 and 354) or the multi-host ST-21 complex, with one isolate in ST-42 complex which is usually more ruminant associated. At a quite different geographical level, a study within a 10 km square grid showed that isolates that were sampled near to each other were more likely to be identical than distant samples. Inference on whether this only operated at

less than 1km or extended longer was dependent on the analytical model chosen limiting the precision of the inference that can be drawn[138]. Published work is therefore consistent with some geographical structuring of *C. jejuni* populations at a local level, but suggests that host association may be a much stronger effect than geography, even over large distances. The direct comparability of MLST results between studies is central to enabling these inferences that rely on data from several independent projects.

Table 1.5. Host species of isolation of major clonal complexes of *C. jejuni* in published reports, 2001-2007.

Paper	Country	Size	Year	21	42	45	48	61	177/179	257	206	354	403	Other
Colles 2003[234]	UK,	n=112	1990s	**Rm St** Ch	Rm	**Ch Tu** Rm St	Rm	Rm	St	Tu Rm St	Rm			
Dingle 2002[114]	UK, NL mainly	n=313	Up to 2001	**Rm Ch** Sa	Rm	**Ch** Rm Sa	**Rm Sa** Ch	**Rm** Sa	Sa	Ch	Ch Rm	Ch	Rm	
Dingle 2001[220]	UK, NL	n=115	1990s	**Rm Ch** Sa		**Ch** Sa			Sa					
French[138]	UK	n=172	2000	**Rm** Bi Wa Ra	**Rm** Ra Wa Bi	**Rm Bi Wa** Ra So Ba		**Rm** Ra Wa						
Karenlampi[61]	Finland	n= 53	2003	**Rm Ch**		**Ch** Rm	Rm	Rm						ST-677 Ch
Kinana[240]	Senegal	n= 46	2000-2	Ch	Ch					Ch		Ch		
Manning 2003[231]	UK mainly	n=215		**Rm Ch** Sw Pe	Rm	**Ch** Rm Pe	**Rm Pe** Ch	**Rm** Sw Gi	Pe	Ch Rm	**Rm** Ch	Rm	**Sw** Rm	ST-573 Ch
Schouls 2003[88]	NL mainly	n=90	1990s	**Rm Ch**	Rm	Rm Ch	Rm Ch	Rm		Rm Ch				ST-46 Ch Rm

Ba Badger; Bi Wild bird – unspecified; Ch Chicken; Gi Giraffe; Pe Pet animal; Ra Rabbit; Rm Ruminant; Sa Sand; So soil; St Starling; Sw Swine; Tu Turkey; Wa Water.
Dominant hosts identified in bold type. Order of listing host gives a further estimate of most frequent source of isolation. Given variable sample numbers from each source in each study these are very approximate estimates of apparent host association.

1.5.3.7 MLST and the dynamics of infection in agricultural animals and settings

MLST has been used to identify population dynamics of infection among successive chicken flocks in a study assessing the effect of phage susceptibility on broiler infection showing that phage resistance was mediated by incursion of phage resistant types rather than *de novo* development of phage resistance[264]. A study describing the diversity and dynamics of ten housed broiler flocks in the UK used MLST to demonstrate the limited number of infecting types at any one time, successive infection by different subtypes over time, contamination of water and the environment with similar types and on occasion isolation of the types infecting broilers from the breeder flocks or environment prior to the detection of infection in the broiler flock[265]. The dynamics of chicken meat contamination during chicken processing used MLST to compare infecting types and identified evidence for differences in types isolated from caecal culture and carcass culture as well as for carry over of infecting type onto successive flocks at the same processing plant[158]. Isolation from header tank water of *C. jejuni* that in some cases matched the contaminating type among chickens has been interpreted as suggesting a possible route of entry of infection[266], although both the lack of discrimination afforded by MLST alone and the lack of a clear direction of causality limits the strength of these conclusions[266]. The uptake of MLST as an approach to trace infection in agricultural settings indicates that development of more reliable fine typing to complement the MLST type may be useful to support studies exploring the dynamics of *C. jejuni* infection on farms and during processing.

1.5.3.8 MLST applied to epidemiology

The main epidemiological functions for which MLST has been evaluated or applied are determination of host association as described above, outbreak investigation,

comparison of trends in infecting subtypes over time or between populations, and integration in a case control study to guide subgroup analysis.

Outbreak investigation

Isolates that were part of a Canadian waterborne outbreak were identical. However, this type was also identified in non-outbreak isolates showing that is did not effectively discriminate outbreak isolates from unrelated isolates[237]. Validation studies each including several outbreaks have also been completed[204,207] giving similar results, with MLST inadequately discriminating to allow it's application in identifying which cases are outbreak related and which are not. A combination of MLST and highly variable gene typing should increase discrimination to allow this functution, as it would provide discrimination at least equivalent to PFGE[204,207]. It is less certain whether an MLST plus antigen gene fragments scheme will be effective in outbreak detection, an important potential task for typing schemes in relation to *C. jejuni* where outbreak detection is uncommon.

Trends between populations and over time

Collections of human isolates typed by MLST have been published from the UK[114,244] and the Netherlands[114], Curacao[249], the USA[238], Finland[61], Australia[243], and New Zealand[267] allowing direct comparison. The first large scale study of human isolates (n=501) was of an international collection but with many isolates from the UK (n=370) and the Netherlands (n=79), the large majority of which were isolated between 1990 and 1999 including 213 in 1991. This showed that human infections included diverse isolates from 15 identified clonal complexes with ST-21, ST-45, ST-61, ST-48, and ST-257 complexes the most common in that order[114]. A study of 234 isolates[249] which had been collected during a 1999/2000 case control study[70] in

Curacao found many isolates in the same clonal complexes identified in Europe[114] with 7 of the 10 most frequent complexes from Curacao being ones already identified along with the novel ST-41, 508 and 657 complex isolates which complexes were first identified in this study and between them comprised 61 isolates (26% of the total). Additionally, 41 isolates belonged to ST-403 complex, which is relatively rare in humans in Europe[114]. These ST-403 complex isolates were restricted to the November-February period of the Curacao study, consistent with either a seasonal or transient source.

An Australian case control study conducted between 1999 and 2001 provided 153 isolates which were subsequently typed using MLST[243]. Of 40 STs, 19 were new to this study, but 16 of these new STs were within 11 known clonal complexes (257, 354, 21, 52, 353, 658, 45, 607, 61, 22 and 460). The most abundant newly described ST was ST-28. Identified 18 times, ST-28 is a single locus variant of ST-354, the central genotype of ST-354[243]. A study from Michigan, USA of just 30 human and chicken isolates found 9 new STs, 5 of which were within established clonal complexes[238]. A Finnish study sequence typed 298 human isolates, from 1996, 2002 and 2003. Although 28 new STs were identified only 9 of these (each represented by just 1 isolate) were not assigned to existing clonal complexes so that in total 264 of the 298 isolates were previously reported STs[61], 25 were among the 19 new STs within known clonal complexes and just 9 were new STs and not members of known clonal complexes. A later UK study reported only clonal complexes of 326 isolates over a one-year period (2003/4). The main clonal complexes identified appear typical of earlier studies with no comment made on individual STs and no comparison with other datasets in the report[244]. This study also noted a seasonal or transient peak in the

summer of ST-45 complex, with the rise more marked in rural than urban areas. A sample from New Zealand in winter 2006, when rates of disease were unusually high identified just 2 new *C. jejuni* STs (4 isolates) in a sample of 107 isolates. Some groups of isolates within this study were unusual variants of common sequence types e.g. ST-474 and ST-190 from clonal complexes ST-48 and 21 respectively. A common source local outbreak is suggested as the likely source[267].

These results indicate geographical variation, but more strikingly that the same main clonal complexes dominate in human disease in much of the world, as evidenced by later publications identifying STs that are already known or are minor variants of widely distributed STs identified in earlier studies in geographically distant areas. The precise ST profile present in each population[243,249] and relative proportions of each ST[267] varied somewhat between studies and suggests that there is some geographic structuring. Because several studies were of short duration the extent to which such locally dominant STs represent stable geographical structuring or a transient expansion, perhaps due to a common source is not always clear[249,267], but this uncertainty would be amenable to longitudinal sampling and typing or to the investigation of possible sources of transient expansions. Other expansions are clearly time limited such as the ST-403 complex isolates in Curacao and ST-45 complex in North West England[244], which are consistent with either seasonal or temporary increases. As well as indexing short term variation these published studies form a geographically widespread baseline against which future studies using the same typing scheme can be compared to determine the longer term dynamics of population structure. The published work to date also shows the potential for more structured

datasets to clarify questions of temporospatial variation in subtypes infecting humans and to use this to explore the origin of infection.

Subset analysis

A Finnish study used MLST to determine the sequence type of 46 isolates from an earlier case control study[106] and analysed risk factors by sequence type defined subgroups[61]. This identified some apparent associations e.g. unassigned STs associated with swimming in natural bodies of water, eating raw mince with ST-48 complex, contact with a dog with ST-45 complex and drinking water from natural sources or not treated with chlorine with ST-677 complex. The sample sizes were small and analyses did not account for multiple testing statistically, which limits the certainty regarding the associations observed. Nonetheless the findings are biologically plausible with, for example, the consumption of raw mince associated with a type most associated with ruminants and raw water consumption associated with a type typical of wild birds.

1.5.3.9 Summary of findings from the application of MLST

The direct comparison of studies using a standard MLST scheme allowed substantially more extensive and robust inference than other typing methods, perhaps most clearly illustrated in reviewing host association and the geographical and temporal structure of human disease isolates. However questions remain in these areas and the value of more structured longitudinal samples, for example to allow separation of transient and seasonal effects, is evident. The limitations of MLST alone in supporting outbreak investigation highlight the importance of identifying effective complementary fine typing to develop the functions of supporting outbreak case definition and ideally to also allow outbreak detection. Such discriminatory fine

typing should also support extension of the application of tracing infection through the food chain, where MLST is already widely used. The results identified when MLST data was used in a small analytical epidemiology study support the integrated epidemiological analysis of classical risk factor information along with MLST type[61]. Alongside these more applied questions the demonstration of extensive recombination as contributing substantially to the patterns of MLST genotypes highlights that use of this data should consider this genetic process in any model based analysis and inference from typing data.

1.6 Using published literature to guide the work of the thesis

This introduction reviewed the descriptive epidemiology, outbreak investigation, attribution of human disease to sources, population biology, and subtyping methods of *C. jejuni.* This has identified open questions, with particular emphasis on those susceptible to the application of sequence typing alone or combined with other methods. These questions are set out, where possible in the form of testable hypotheses in the next chapter describing the aims of the work presented in this thesis.

2. Project Aims

The central premises of this work were that firstly, it would be possible to use sequence typing data to both improve understanding of the population biology of *C. jejuni* and address epidemiological questions and secondly, increased understanding of the population biology could enhance epidemiological analysis. The main questions investigated and their specific motivations are set out below. The sections 2.1 to 2.5 correspond to results chapters 4 to 8 respectively.

2.1 Subtypes of *Campylobacter jejuni* in a human population (Chapter 4)

A descriptive epidemiological study of *C. jejuni* in combination with MLST subtype was undertaken using a structured sample of cases from two geographically separate human populations in England over three years. This was to provide a framework and to raise further hypotheses for later work in the thesis, as well as addressing some questions identified in the introductory literature review. For example, the degree of secular change over several years determines the extent to which contemporaneous non-human origin samples are needed to assess likely sources of infection and the extent of spatial structuring of subtypes causing human disease dictates how geography must be considered, guiding later work in the thesis. The assessment of the sub-types contributing to the summer peak of human disease may help to clarify the source of this phenomenon as a more direct output from this first analysis. The main questions addressed in this descriptive study were:

Do the subtypes of *C. jejuni* causing human disease change over time?

In particular

i) Do the subtypes change systematically within 3 years?

ii) Are particular subtypes responsible for the seasonal peak?

Do infections in geographically separate populations in the UK come from one or distinguishable populations of *C. jejuni*?

How does a large UK sample of *C. jejuni* from human infections compare with human isolates from distant countries?

Do different subtypes of *C. jejuni* infect different demographic groups?

Is symptom complex associated with subtype?

2.2 Source attribution by genotype (Chapter 5)

Differences between *C. jejuni* in different animal host populations, but with substantial overlap, were described in the introduction. This apparent host association has never been quantified, the population biology of the association has not been explored and no attempt has been made to assess the potential accuracy of assignment to host of origin based on *C. jejuni* subtype information. The questions addressed were:

Can MLST data be used to describe the relationship between different host species and to support attribution of human disease burden?

In particular

i) Is there a reliable association between genotype and host?

ii) What is the population biology of host association in *C. jejuni*?

iii) How is genotypic data best used to describe host association?

2.3 The *Campylobacter jejuni* population in chicken (Chapter 6)

The attribution of human infections to source using MLST data requires the availability of appropriate reference populations. Chicken is probably the most important source of human infection. Characterisation of the population of *C. jejuni*

contaminating chicken meat was undertaken to contribute a foundation to this attribution work as well as describing this population in its own right. The questions addressed were:

What is the population of *C. jejuni* from chicken to which humans are exposed?

In particular

i) What is the distribution of genotypes on retail poultry meat?

ii) How do these vary over time and in relation to place of origin?

iii) Do food production and handling methods alter subtype distribution?

2.4 Outbreak detection (Chapter 7)

The lack of identified outbreaks in *C. jejuni* may reflect the lack of epidemiological linkage between cases but may also, at least in part, be due to ineffective detection of outbreaks. An assessment of the sensitivity of typing methods used to date for outbreak detection, the required sensitivity, and the performance possible with currently available sequence typing approaches is the focus of this section of the thesis. The questions addressed were:

How can molecular typing contribute to outbreak investigation in *C. jejuni*?

In particular

i) How sensitive are currently used methods to detect small or diffuse outbreaks?

ii) What typing system characteristics would be needed to detect of outbreaks?

iii) What typing characteristics would allow source identification for outbreaks?

iv) What contribution do additional genes make to outbreak detection capacity?

v) How useful is a combination of MLST, *flaA* and *porA* in outbreak detection?

2.5 Genotype, source of infection, and population biology (Chapter 8)

The last results chapter combined the findings and datasets from earlier chapters to extend the questions addressed and findings from these chapters. Since identifying the source of infection is central to disease control, one main focus on this integration of the different components of the work was to support identification of the source of infection in humans. The other was to use the available data to better understand the population biology of *C. jejuni* and the application of this to epidemiological questions. Because the specific aims of this chapter arose from work in the earlier chapters they are specified at the start of chapter 8.

3. Methods

This methods section describes laboratory methods, sampling and analysis. The laboratory methods (section 3.1) were applied to the three isolate collections subtyped for the work of this thesis. The sampling approach used to select the isolate collections typed in the study and also to construct further datasets from published work is described in section 3.2. These datasets were analysed individually, in combined analyses, and together with external datasets from published work. Section 3.3 describes the analytical approaches.

3.1 Laboratory methods

3.1.1 Preparation of heat-killed cell suspensions

Isolates typed within this study were supplied either as the supernatant from boiled cell suspensions of archived isolates by the Health Protection Agency Centre for Infections (*Campylobacter* Sentinel Surveillance Scheme (CSSS), retail poultry surveys and case control study) prepared according to the protocol described below, or as aliquots from DNA extractions undertaken for genome array studies at the London School of Hygiene and Tropical Medicine (Nottinghamshire subset of case control study isolates) using their protocols[199].

Protocol for heat killed cell suspension preparation: One bead was removed from the -80°C archive freezer isolate sample* and streaked onto blood agar†. This was incubated for 48 hours in microaerophilic conditions. A subculture was then plated

*For the CSSS component of this thesis 3 frozen beads were boiled directly without subculture but this was associated with poor MLST results and altered after 150 isolates to the culturing approach described above.
†Columbia blood agar is composed of special peptone, starch, sodium chloride and agar, after is has been autoclaved and cooled 5% sterile defibrinated blood is added (information from Oxoid).

onto CCDA‡ agar and incubated for 24 hours in microaerophilic conditions. A 1µl loopful of fresh bacterial growth (the equivalent of a large colony of fresh bacterial growth) was taken from each of these plates, and placed into 125µl of sterile PBS and boiled for 10 minutes. The supernatant was transferred into a fresh tube following centrifugation and transported to Oxford in person or using standard post.

3.1.2 Sample handling and PCR amplification

Samples were stored in individual tubes in Oxford with 50µl from each sample arrayed in 96 well plates for PCR including negative control wells. Tubes and plates were stored in a -20°C freezer. MLST[220] and *flaA* sequencing used standard methods from published primers and protocols[123,207,220,221] and others from the *Campylobacter* MLST website http://pubmlst.org/ as detailed below. Sequencing of *porA* used similar methods developed by colleagues as detailed below and available in more detail at http://pubmlst.org/campylobacter/momp/.

3.1.2.1 PCR amplification

Amplification of each gene product by PCR was performed separately for the 7 MLST loci and, in isolates from the case control and retail poultry studies, for *flaA* and *porA* genes in 96 well microtitre plates. Double volume reactions were used in the early part of the project and when repeating failed reactions. 25µl PCR reactions were performed using 2µl of template, 8.875µl of milliQ H_2O, 2.5µl of 10X buffer, 0.5 µl of 10µM stock of dNTP mix, 0.125µl of Qiagen *Taq* polymerase, 0.5µl each of 10µM stock of forward and reverse primers.

‡ *Campylobacter* selective blood free agar (CCDA) is composed of LabM (LAB 112) and contains cefoperazone (16mg per vial) and amphotericin B (5mg per vial) antibiotics (information from Media, Centre for Infections, Colindale).

Incubation conditions for PCR were based on those published[220] for the MLST and *flaA* loci as described below. The number of cycles used was either 35 or 40 and the storage temperature following completion either 4°C or15°C at different periods of the study.

Incubation for MLST loci and *flaA*

Step 1	95°C	3 minutes
35 or 40 cycles of steps 2-4		
Step 2	95°C	30 seconds
Step 3	55-60°C	30 seconds
Step 4	72° C	1 minute
Step 5	72°C	10 minutes
Step 6	4°C /15 °C	Until plate retrieved (up to 16 hours)

Incubation conditions for *porA* amplification were

Step 1	95 °C for 15 minutes	
35 cycles of steps 2-4		
Step 2	94 °C for 30 seconds	
Step 3	50 °C for 30 seconds	
Step 4	72 °C for 1 minute 30 seconds	
Step 5	72 °C for 5 minutes	
Step 6	4°C /15 °C	Until plate retrieved (up to 16 hours)

At least eight samples from each plate (5 µl of each sample with 1 µl 6x loading buffer) including negative controls were checked for PCR product using a 1% agarose gel containing ethidium bromide 0.5 mg / ml by electrophoresis at 140 V for 15 min and visualisation on a U.V. trans-illuminator. When repeating failed reactions all samples were tested to allow individual dilution.

3.1.2.2 Precipitation and dilution of PCR products

After PCR, the amplicons were precipitated using the following protocol. Polyethylene glycol (20% PEG_{8000}, Sigma) / 2.5 M NaCl in milliQ H_2O was added to each well (30µl for 25µl reactions and 60µl for 50µl reactions). The plate was re-sealed, mixed using the vortex mixer, briefly centrifuged at 500g to ensure the whole mix was at the bottom of the wells and incubated at room temperature for 60 minutes or 4°C overnight. The PCR products were formed into a pellet by centrifugation for one hour at 2750g (at 4°C) and the supernatant discarded. The DNA pellet was washed twice with the addition of 150 ml of cold 70% EtOH (from -20°C freezer) to each well, centrifugation at 2750g for 10 min of the resealed plate and discarding of the supernatant by a brief 500g centrifugation of the inverted plate. The precipitated PCR products were diluted in 20 µl milliQ H_2O. Where PCR was repeated on failed samples dilution used between 5 µl and 40 µl depending on band strength assessed by UV visualisation of electrophoresed samples from each reaction on ethidium bromide containing agarose gels.

3.1.3 Sequencing

Arrayed PCR products and negative controls were transferred in duplicate to 96 well plates to allow forward and reverse sequencing reactions.

3.1.3.1 Nucleotide sequence extension reactions

Nucleotide sequence extension reactions were as previously published[220] based on manufacturers protocols using BigDye™ reaction mix and 5X buffer but performed at 1/32 of manufacturers recommended quantities. Each 10µl sequencing reaction contained 2µl template DNA, 4µl 0.67µM forward or reverse primer (final concentration of 0.27µM), 0.25µl Big Dye (Applied Biosystems), 1.875µl 5x reaction

buffer (Applied Biosystems or equivalent laboratory stock), 1.875µl molecular biology grade water (milliQ).

Incubation conditions were:

30 cycles of steps 1-3

Step 1	96 °C for 10 sec
Step 2	50 °C for 5 sec
Step 3	60 °C for 2 min
Step 4	Hold: 4°C

3.1.3.2 Precipitation of products

The extension products were precipitated by addition of 60µl of a mixture of 100% ethanol, sodium acetate (3M NaOAc at pH 5.2, Anachem Ltd, Luton, UK) and milliQ H_2O in proportions to allow addition of 50µl ethanol, and 2µl sodium acetate and 8µl H_2O to each well. Plates were then sealed, vortexed gently and allowed to incubate for 45 minutes at room temperature. The incubated plates were centrifuged for one hour at 2750g (at 4°C). The supernatant was then immediately discarded by spinning the inverted plates at 500g for one minute. The pellet was washed with the addition of 150 ml of cold 70% EtOH (from -20°C freezer) to each well, centrifugation at 2750g for 10 min of the resealed plate and discarding of the supernatant by a brief 500g centrifugation of the inverted plate.

3.1.3.3 Sequencing extension reaction products

The nucleotide extension reaction products were detected on an ABI Prism 3730 automated DNA analyser, a core facility in the Zoology Department at the University of Oxford.

3.1.4 Primers

3.1.4.1 Antigen gene primers

Primers developed by Meinersmann[221] were used to sequence the 321 base pair short variable region of *flaA*. The following primers were used for amplification and sequencing:

fla4F (forward)	5′- GGA TTT CGT ATT AAC ACA AAT GGT GC -3′
fla625R (reverse)	5′- CAA G(AT)C CTG TTC C(AT)A CTG AAG -3

In some sequencing reactions the alternative forward primer fla242F was used

fla242F	5′- CTA TGG ATG AGC AAT T(AT)A AAA T-3′.

Primers developed by Kate Dingle (unpublished) were used for amplification and sequencing of the *porA* gene fragment:

MOMP-1 (forward)	5'-GAT GGT TTA ACT CTA GCT GC-3'
MOMP-2 (reverse)	5'-TGA GAA GTT AAG TTT TGG AGA G-3'.

3.1.4.2 MLST primers

The primers used for amplification and sequencing are either from the original scheme[220], those developed subsequently by Kate Dingle and available from the pubmlst website [http://pubmlst.org/campylobacter/mlst-info/cjejuni/primers.shtml], which are all designated by A"integer" or S"integer" in the Tables 3.1 and 3.2 respectively, or from joint *C. jejuni* and *C. coli* primers[123] which are designated as AF and AR for forward and reverse primers for each locus. The preferred primers for amplification and sequencing are listed first in Tables 3.1 and 3.2 followed by the full range of alternatives used when initial and subsequent repeat reactions failed to produce initial PCR product or sequencing products.

Table 3.1. Amplification primers used in sequence typing within this work. For each locus the first two listed were the default forward and reverse primers with others used in some early work or for failed reactions listed below.

Locus	Primer	Sequence (5'-3')
aspA	F aspA9	AGT ACT AAT GAT GCT TAT CC
	R aspS6	TTC ATT TGC GGT AAT ACC ATC
	F aspAF	GAG AGA AAA GCW GAA GAA TTT AAA GAT
	R aspAR	TTT TTT CAT TWG CRS TAA TAC CAT C
	R aspA10	ATT TCA TCA ATT TGT TCT TTG C
glnA	F glnS1	GCT CAA TTC ATG GAT GGC
	R glnS6	TTC CAT AAG CTC ATA TGA AC
	F glnA1	TAG GAA CTT GGC ATC ATA TTA CC
	R glnA2	TTG GAC GAG CTT CTA CTG GC
	F glnAF	TGA TAG GMA CTT GGC AYC ATA TYA C
	R glnAR	ARR CTC ATA TGM ACA TGC ATA CCA
gltA	F gltS1	GTG GCT ATC CTA TAG AGT GGC
	R gltS6	CCA AAG CGC ACC AAT ACC TG
	F gltA1	GGG CTT GAC TTC TAC AGC TAC TTG
	R gltA2	CCA AAT AAA GTT GTC TTG GAC GG
	F gltS3	CTT ATA TTG ATG GAG AAA ATG G
	F gltAF	GAR TGG CTT GCK GAA AAY AAR CTT T
	R gltAR	TAT AAA CCC TAT GYC CAA AGC CCA T
glyA	F glyAF	ATT CAG GTT CTC AAG CTA ATC AAG G
	R glyAR	GCT AAA TCY GCA TCT TTK CCR CTA AA
	F glyA1	GAG TTA GAG CGT CAA TGT GAA GG
	R glyA2	AAA CCT CTG GCA GTA AGG GC
	F glyS3	AGC TAA TCA AGG TGT TTA TGC GG
	F glyS5	GCT AAT CAA GGT GTT TAT AT
	F glyS7	AGC CTA ATT CAG GTT CTC AA
	R glyS4	AGG TGA TTA TCC GTT CCATCGC
pgm	F pgmAF	CAT TGC GTG TDG TTT TAG ATG TVG C
	R pgmAR	AAT TTT CHG TBC CAG AAT AGC GAA A
	F pgmA7	TAC TAA TAA TAT CTT AGT AGG
	R pgmA8	CAC AAC ATT TTT CAT TTC TTT TTC
	F pgmS5	GGT TTT AGA TGT GGC TCA TG
	R pgmS2	TCC AGA ATA GCG AAA TAA GG
tkt	F tktS5	GCT TAG CAG ATA TTT TAA GTG
	R tktS4	ACT TCT TCA CCC AAA GGT GCG
	F tktA3	GCA AAC TCA GGA CAC CCA GG
	R tktA6	AAA GCA TTG TTA ATG GCT GC
	R tktS6	AAG CCT GCT TGT TCT TTG GC
	F tktAF	GCA AAY TCA GGM CAY CCA GGT GC
	R tktAR	TTT TAA THA VHT CTT CRC CCA AAG GT
uncA	F uncAF	GWC AAG GDG TTA TYT GTA TWT ATG TTG C
	R uncAR	TTT AAD AVY TCA ACC ATT CTTTGTCC
	FuncA7	ATG GAC TTA AGA ATA TTA TGGC
	R uncA8	ATA AAT TCC ATC TTC AAA TTCC
	F uncS3	AAA GTA CAG TGG CAC AAG TGG
	F uncS5	TGT TGC AAT TGG TCA AAA GC
	R uncS4	TGC CTC ATC TAA ATC ACT AGC

3.1.4 Primers

3.1.4.1 Antigen gene primers

Primers developed by Meinersmann[221] were used to sequence the 321 base pair short variable region of *flaA*. The following primers were used for amplification and sequencing:

fla4F (forward)	5′- GGA TTT CGT ATT AAC ACA AAT GGT GC -3′
fla625R (reverse)	5′- CAA G(AT)C CTG TTC C(AT)A CTG AAG -3

In some sequencing reactions the alternative forward primer fla242F was used

fla242F	5′- CTA TGG ATG AGC AAT T(AT)A AAA T-3′.

Primers developed by Kate Dingle (unpublished) were used for amplification and sequencing of the *porA* gene fragment:

MOMP-1 (forward)	5'-GAT GGT TTA ACT CTA GCT GC-3'
MOMP-2 (reverse)	5'-TGA GAA GTT AAG TTT TGG AGA G-3'.

3.1.4.2 MLST primers

The primers used for amplification and sequencing are either from the original scheme[220], those developed subsequently by Kate Dingle and available from the pubmlst website [http://pubmlst.org/campylobacter/mlst-info/cjejuni/primers.shtml], which are all designated by A"integer" or S"integer" in the Tables 3.1 and 3.2 respectively, or from joint *C. jejuni* and *C. coli* primers[123] which are designated as AF and AR for forward and reverse primers for each locus. The preferred primers for amplification and sequencing are listed first in Tables 3.1 and 3.2 followed by the full range of alternatives used when initial and subsequent repeat reactions failed to produce initial PCR product or sequencing products.

Table 3.1. Amplification primers used in sequence typing within this work. For each locus the first two listed were the default forward and reverse primers with others used in some early work or for failed reactions listed below.

Locus	Primer	Sequence (5'-3')
aspA	F aspA9	AGT ACT AAT GAT GCT TAT CC
	R aspS6	TTC ATT TGC GGT AAT ACC ATC
	F aspAF	GAG AGA AAA GCW GAA GAA TTT AAA GAT
	R aspAR	TTT TTT CAT TWG CRS TAA TAC CAT C
	R aspA10	ATT TCA TCA ATT TGT TCT TTG C
glnA	F glnS1	GCT CAA TTC ATG GAT GGC
	R glnS6	TTC CAT AAG CTC ATA TGA AC
	F glnA1	TAG GAA CTT GGC ATC ATA TTA CC
	R glnA2	TTG GAC GAG CTT CTA CTG GC
	F glnAF	TGA TAG GMA CTT GGC AYC ATA TYA C
	R glnAR	ARR CTC ATA TGM ACA TGC ATA CCA
gltA	F gltS1	GTG GCT ATC CTA TAG AGT GGC
	R gltS6	CCA AAG CGC ACC AAT ACC TG
	F gltA1	GGG CTT GAC TTC TAC AGC TAC TTG
	R gltA2	CCA AAT AAA GTT GTC TTG GAC GG
	F gltS3	CTT ATA TTG ATG GAG AAA ATG G
	F gltAF	GAR TGG CTT GCK GAA AAY AAR CTT T
	R gltAR	TAT AAA CCC TAT GYC CAA AGC CCA T
glyA	F glyAF	ATT CAG GTT CTC AAG CTA ATC AAG G
	R glyAR	GCT AAA TCY GCA TCT TTK CCR CTA AA
	F glyA1	GAG TTA GAG CGT CAA TGT GAA GG
	R glyA2	AAA CCT CTG GCA GTA AGG GC
	F glyS3	AGC TAA TCA AGG TGT TTA TGC GG
	F glyS5	GCT AAT CAA GGT GTT TAT AT
	F glyS7	AGC CTA ATT CAG GTT CTC AA
	R glyS4	AGG TGA TTA TCC GTT CCATCGC
pgm	F pgmAF	CAT TGC GTG TDG TTT TAG ATG TVG C
	R pgmAR	AAT TTT CHG TBC CAG AAT AGC GAA A
	F pgmA7	TAC TAA TAA TAT CTT AGT AGG
	R pgmA8	CAC AAC ATT TTT CAT TTC TTT TTC
	F pgmS5	GGT TTT AGA TGT GGC TCA TG
	R pgmS2	TCC AGA ATA GCG AAA TAA GG
tkt	F tktS5	GCT TAG CAG ATA TTT TAA GTG
	R tktS4	ACT TCT TCA CCC AAA GGT GCG
	F tktA3	GCA AAC TCA GGA CAC CCA GG
	R tktA6	AAA GCA TTG TTA ATG GCT GC
	R tktS6	AAG CCT GCT TGT TCT TTG GC
	F tktAF	GCA AAY TCA GGM CAY CCA GGT GC
	R tktAR	TTT TAA THA VHT CTT CRC CCA AAG GT
uncA	F uncAF	GWC AAG GDG TTA TYT GTA TWT ATG TTG C
	R uncAR	TTT AAD AVY TCA ACC ATT CTTTGTCC
	FuncA7	ATG GAC TTA AGA ATA TTA TGGC
	R uncA8	ATA AAT TCC ATC TTC AAA TTCC
	F uncS3	AAA GTA CAG TGG CAC AAG TGG
	F uncS5	TGT TGC AAT TGG TCA AAA GC
	R uncS4	TGC CTC ATC TAA ATC ACT AGC

Table 3.2. Sequencing primers used in sequence typing within this work. For each locus the first two listed were the default forward and reverse primers with others used in some early work or for failed reactions listed below.

Locus	Primer	Sequence (5'-3')
aspA	F aspS3	CCA ACT GCA AGA TGC TGT ACC
	R aspS6	TTC ATT TGC GGT AAT ACC ATC
	F aspAF	GAG AGA AAA GCW GAA GAA TTT AAA GAT
	R aspAR	TTT TTT CAT TWG CRS TAA TAC CAT C
glnA	F glnS1	GCT CAA TTC ATG GAT GGC
	R glnS6	TTC CAT AAG CTC ATA TGA AC
	F glnS3	CAT GCA ATC AAT GAA GAA AC
	F glnAF	TGA TAG GMA CTT GGC AYC ATA TYA C
	R glnAR	ARR CTC ATA TGM ACA TGC ATA CCA
gltA	F gltS1	GTG GCT ATC CTA TAG AGT GGC
	R gltS6	CCA AAG CGC ACC AAT ACC TG
	F gltS3	CTT ATA TTG ATG GAG AAA ATGG
	F gltAF	GAR TGG CTT GCK GAA AAY AAR CTT T
	R gltAR	TAT AAA CCC TAT GYC CAA AGC CCA T
glyA	F glyS3	AGC TAA TCA AGG TGT TTA TGCG G
	R glyS4	AGG TGA TTA TCC GTT CCA TCG C
	F glyAF	ATT CAG GTT CTC AAG CTA ATC AAG G
	R glyAR	GCT AAA TCY GCA TCT TTK CCR CTA AA
	F glyS5	GCT AAT CAA GGT GTT TAT AT
	F glyS7	AGC CTA ATT CAG GTT CTC AA
pgm	F pgmAF	CAT TGC GTG TDG TTT TAG ATG TVG C
	R pgmAR	AAT TTT CHG TBC CAG AAT AGC GAA A
	F pgmS5	GGT TTT AGA TGT GGC TCA TG
	R pgmS2	TCC AGA ATA GCG AAA TAA GG
tkt	F tktS5	GCT TAG CAG ATA TTT TAA GTG
	R tktS4	ACT TCT TCA CCC AAA GGT GCG
	R tktS6	AAG CCT GCT TGT TCT TTG GC
	F tktAF	GCA AAY TCA GGM CAY CCA GGT GC
	R tktAR	TTT TAA THA VHT CTT CRC CCA AAG GT
uncA	F uncS5	TGT TGC AAT TGG TCA AAA GC
	R uncS4	TGC CTC ATC TAA ATC ACT AGC
	F uncS3	AAA GTA CAG TGG CAC AAG TGG
	F uncAF	GWC AAG GDG TTA TYT GTA TWT ATG TTG C
	R uncAR	TTT AAD AVY TCA ACC ATT CTT TGT CC

3.1.5 Sequence assembly and bioinformatics

Sequence traces were assembled using methods Sequence Typing Analysis and Retrieval System software developed by Drs Man-Suen Chan and Nicki Ventress http://neelix.molbiol.ox.ac.uk:8080/userweb/mchan/stars/ (Copyright University of

Oxford 2001). Failed sequences were reviewed manually and assembled where all base pairs could be clearly identified for the gene fragments specified in Table 1.4.

The consensus sequence was queried against the allele sequences maintained on the *Campylobacter* database http://pubmlst.org/campylobacter/ to give an allele number Original traces for new alleles were submitted for curation in line with database policy. The combination of alleles for the seven housekeeping genes gave the sequence type (ST). Antigen gene sequences were similarly assigned allele numbers using the *flaA* and *porA* databases and curation at the University of Oxford.

3.2 Study samples and datasets

3.2.1 *Campylobacter* Sentinel Surveillance Scheme (CSSS) dataset

The CSSS was undertaken between May 2000 and April 2003. Selected health authorities and laboratories throughout England and Wales (Figure 3.1) collaborated with the *Campylobacter* Reference Unit of the Public Health Laboratory Service (CRU). Laboratories in these areas referred isolates of *Campylobacter* to CRU for serotyping and phage typing, while epidemiological information was collected using a standard questionnaire (Annex 2). The questionnaire was administered to each patient by the health or local authority as part of the routine investigation of foodborne infection. This captured demographic and clinical data as well as an exposure history during the 2 weeks before the onset of illness to ask patients about possible risk factors. This data was collated centrally[268].

The scheme was intended to allow application of the phenotyping approaches developed at the CRU and specific aims of the scheme included the:

• production of a representative sample of *Campylobacter* isolates throughout England and Wales, with which to evaluate typing data;

• provision of a statistically valid sample from which trends can be identified;

• provision of data to support detailed studies on specific food, animal, environmental, and cultural issues that may play a part in the epidemiology of *Campylobacter* infection[268].

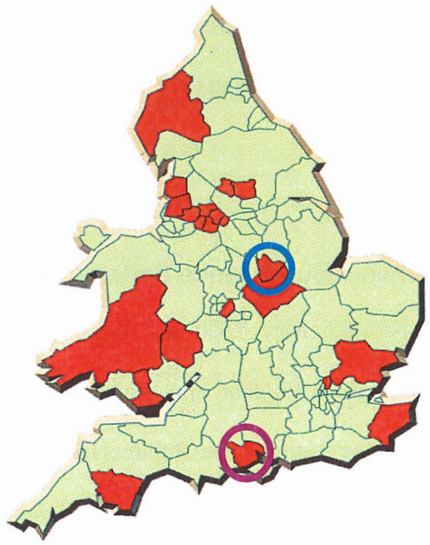


Figure 3.1. Areas included in the CSSS, highlighted in red including the two sites in Nottinghamshire (blue circle) and Hampshire (purple circle) types in this study [map from Gillespie *et al.* 2002[38]]

Table 3.3 shows that a minority of isolates in the CSSS are suitable for inclusion in the current study which was restricted to cases with a questionnaire, a linked sample of *C. jejuni* and with no history of travel in the two weeks before onset of illness. For 29% of those who would have been suitable for entry in the current study no isolate was available and for 10% postcode was incomplete.

Table 3.3. Proportion of CSSS participants available to the current study with linked isolates of *C. jejuni*. [data from Ian Gillespie, Health Protection Agency]

Data	n	Percent of total	Percent loss
All questionnaires*	20482	100%	2%
Non travel	15975	78%	22%
Non-duplicates	15851	77%	1%
Linked data	11199	55%	29%
C. jejuni	9552	47%	15%
Postcodes complete	8619	42%	10%

The part of the dataset and isolate archive used in the work of this thesis was from two populations (Nottingham and Southampton and parts of their surrounding counties, Figure 3.1). These populations were selected because:

1. Recruitment had been successful and relatively stable in these areas over the full three-year duration of the scheme;

2. They were relatively distant geographically (separated by 270km);

3. The sample sizes in the two areas were similar;

4. No other researchers had declared an interest in sequence typing isolates from these components of the study.

Only those cases with isolates linked to a questionnaire with complete postcode and where there was no history of recent international travel were included. These criteria produced 1596 isolates, 885 from Nottinghamshire and 711 from Hampshire.

3.2.2 Retail poultry survey (RPS) dataset

Isolates were derived from three retail poultry surveys in which fresh and frozen chicken was sampled from retail outlets across the UK: the 2001 retail poultry survey; the Coordinated Local Authority Sentinel Surveillance of Pathogens (CLASSP); and the Wales and Northern Ireland survey with samples from the latter two surveys coming from 2004-2005. A stratified random sample was selected from the 2001

survey as described below while samples from the later surveys were taken as available until the total for the project was reached. Due to limited availability of samples in the early part of the CLASSP survey this led to over-representation of the isolates from the Wales and Northern Ireland study. In the latter part of the study sample numbers within CLASSP increased substantially.

3.2.2.1 The 2001 retail poultry survey

This was a national survey undertaken by the FSA to determine the level of *Campylobacter* and *Salmonella* in retail chicken between April and June 2001. A total of 4866 samples of fresh, frozen, whole and portioned chicken purchased from over 1500 retail outlets throughout the UK were tested. Chicken samples were purchased from a representative cross section of retail outlets, including major retailers, butchers, grocers, market and farm stalls according to market share. Samplers were permitted to take a maximum of five samples from any one store, and the samples were required to be different types of chicken product. Chicken samples were sent to one of three laboratories and tested for the presence or absence of *Salmonella*, and the presence or absence and numbers of *Campylobacter*. Overall 50% of these were positive for *Campylobacter*, though this varied with Scotland and Northern Ireland having a substantially higher rate of isolation than samples from England and Wales. The laboratory processing samples from Scotland and Northern Ireland did not process English and Welsh samples. All *Salmonella* isolates and a proportion of the *Campylobacter* isolates were sent to reference laboratories for serotyping, phage typing, screening for antimicrobial resistance and archiving. Full details of this study have been published[269].

For the current study, to ensure that 200 typed isolates would be available given the possibility of non-viable or incorrectly speciated isolates, 240 were selected randomly within the following strata and constraints:

1. Restricted to samples with a speciated isolate in the HPA archive.
2. No more than 1 isolate from any sample was included.
3. 50% of isolates from England with 17% from each of Northern Ireland, Scotland and Wales.
4. 75% *C. jejuni* and 25% *C. coli*, approximately in line with the proportional levels of contamination.

Only the *C. jejuni* isolates are analysed in the work presented in this thesis.

3.2.2.2 CLASSP

CLASSP is a collaboration between local authorities, the Local Authorities Coordinators of Regulatory Services (LACORS) and the Health Protection Agency. It commenced in November 2004 to collect standardised data on the prevalence of *Campylobacter* spp. and *Salmonella* spp. on raw whole chicken samples, either fresh or frozen. *Campylobacter* isolates originating from these samples were provided for this study. Participating local authorities were asked to obtain 30% of samples from butchers and 70% from large multiple retail chains. Additional information regarding the sample was obtained using a standardised pro-forma. Laboratories were asked to send a sample of isolates to the HPA Centre for Infections for confirmation, speciation and archiving. All isolates were archived as part of a project funded by the FSA. All available isolates were included in the current typing study until it was completed.

3.2.2.3 Wales and Northern Ireland survey

The Wales and N. Ireland retail chicken survey was a collaboration between The Food Standards Agency, and the Welsh and Northern Irish local authorities. The survey aimed to monitor the prevalence and type of *Campylobacter* and *Salmonella* on the surface of raw whole chicken on sale in Wales and Northern Ireland. This was a rolling survey that started with sampling of whole chickens in Wales in 2001, 2002 and 2003. It was extended to Northern Ireland in 2004, continuing in 2005 and 2006 when the standardised methods for CLASSP were also adopted for use by the Wales and Northern Ireland Survey. Raw whole retail chickens, both fresh and frozen, were sampled from a range of retail outlets across Northern Ireland and Wales including supermarkets and butchers. Samples were tested for presence of *Salmonella* and *Campylobacter* and 25% of *Campylobacter*s isolated from chicken were sent to the HPA for typing and archiving[85,270] from which available isolates were included in this project for typing.

3.2.2.4 Typed isolates in the RPS dataset

In total 742 *C. jejuni* isolates were typed in this project: 165 from the 2001 survey, 408 from the CLASSP and 169 from the Wales and Northern Ireland surveys.

3.2.3 Case control study (CaCo) dataset

The FSA funded "Case-control Study of Risk Factors for *Campylobacter jejuni* infectious intestinal disease in England and Wales" was a multicentre case control study conducted between April 2005 and June 2006. The MLST and antigen gene typed subset of isolates comprised two sets of isolates that were typed as a sub-study. The first set, 120 isolates from Nottinghamshire in the preparatory stages of the case-control study (2003-2004) had already been typed using a microarray. Comparative

analysis of population structure defined by MLST and phylogenomics was a requirement of funding of the MLST sub-study thus necessitating that these isolates be included in the sequence typing sub-study. The second set comprised the 246 available isolates from Lancashire, Cheshire and Merseyside, in the case control study proper (2005 – 2006). These were selected because of the relatively good availability of questionnaire data for cases and controls in this region of this multicentre study. No other criteria were applied to this selection.

Table 3.4. Timing and geographical origin of case control study isolates included in the sequence typing sub-study.

Year	Nottinghamshire	Lancashire, Cheshire and Merseyside
2003	58	0
2004	58	0
2005	0	174
2006	0	67
Not recorded	4	5
	120	246

3.2.4 Cattle sheep and chicken (CSP) dataset

A reference dataset to represent cattle, sheep and chicken isolates of *C. jejuni* was established by including isolates using the following criteria:

1. All isolates on the pubmlst database (www.pubmlst.org) with date-stamp before 1 August 2004 and which had been published in peer reviewed literature or for which permission was obtained from those who had submitted the data. All but 10 of the isolates on pubmlst at this time were available for inclusion by these criteria.

2. During the contact process to seek permission to include unpublished isolates from the pubmlst database additional typed isolates (n=27) made available by contacted researchers and were also included.

3. It has been shown that *C. jejuni* recombines with *C. coli*[218]. Those isolates with at least 4/7 alleles typical of *C. jejuni* are included.

In total 713 isolates were available by these criteria. These comprised 245 cattle isolates, 159 sheep isolates and 309 chicken isolates. Their distribution by year and country is summarised in Tables 3.5 and 3.6 showing that it is UK dominated but covers a fairly large geographical and temporal spread.

Table 3.5. Isolates by year of isolation and host species.

Year	Chicken	Cattle	Sheep	Total
1981	0	4	0	4
1982	2	1	2	5
1983	0	3	0	3
1984	0	1	0	1
1986	0	2	0	2
1988	2	18	0	20
1989	0	1	0	1
1990	54	1	0	55
1991	30	6	0	36
1992	1	3	0	4
1993	8	6	1	15
1994	6	1	0	7
1995	12	1	0	13
1996	35	0	0	35
1997	2	0	0	2
1998	40	41	68	149
1999	10	38	38	86
2000	15	6	0	21
2001	45	83	5	133
2002	0	0	2	2
2003	13	0	0	13
Unspecified	34	29	43	106
Total	309	245	159	713

Table 3.6. Isolates by country and host species.

Country	Chicken	Cattle	Sheep	Total
Canada	0	5	0	5
Czech Republic	8	0	0	8
Denmark	6	1	0	7
Netherlands	53	4	0	57
New Zealand	5	1	0	6
Northern Ireland	1	2	0	3
UK	217	218	158	593
USA	17	13	1	31
Unknown	2	1	0	3
Total	309	245	159	713

This dataset is based on submissions to the pubmlst database and is not a structured sample. It includes several small local collections and includes unusual STs out of proportion to their population frequency since the first identified isolate from each ST is generally present on this database but many isolates from frequently isolated STs are not submitted. These caveats were considered in the use of this dataset and analytical approaches taken to allow robust inference not susceptible to these sampling issues.

3.2.5 Published human disease (PHD) dataset

This dataset was assembled using all fully published structured human disease datasets from defined time and place. It comprised isolates from parts of Australia[243], New Zealand[267] and Finland[61] as summarised in Table 3.7.

Table 3.7. *C. jejuni* isolates by country and year assembled from published literature.

	Country		
Year	Australia	Finland	New Zealand
1999-2001	153		
1996		92	
2002		109	
2003		97	
2006			107

3.2.6 International published animal and environmental datasets

A collection of isolates from animal and environmental origin from the published literature was extracted for comparison with data in this thesis. All isolates were taken from some studies[138,240] which gave a line listing allowing identification of all isolates sequence type and host of origin. Only non-human source isolates were included in studies comprising a mix of human and non-human origin isolates[61,114]. Where the published paper did not allow reliable linkage of ST with origin[231] only those isolates where this information could be recovered from the pubmlst database were included.

Tables 3.8, 3.9 and 3.10 summarise this collection of isolates overall by study, country of isolation, and host or environmental niche. Subsets of this dataset are used in analyses as indicated in the relevant analysis and results sections.

Table 3.8. Origin of isolates by country of isolation and published study.

	Study				
Country	Dingle[114]	Karenlampi[61]	Kinana[240]	Manning[231]	French[138]
Czech Republic	2			8	
Denmark	2			5	
Finland		53			
Netherlands	54			3	
New Zealand	6				
Senegal			46		
UK	248			135	172
USA	1				
Unknown	2				
	315	53	46	151	172

Table 3.9. Origin of isolates by host species and published study.

	Study					
Host species	Dingle[114]	Karenlampi[61]	Kinana[240]	Manning[231]	French[138]	Total
Chicken	132	33	46	42		253
Cattle	54	20		52	82	208
Sheep	72			35	5	112
Sand (bathing beach)	50					50
Bird					35	35
Water				1	23	24
Rabbit					20	20
Pig	5			11		16
Broiler environment	2			8		10
Badger					5	5
Soil					2	2
Giraffe				1		1
Pet				1		1
	315	53	46	151	172	737

Table 3.10. Origin of isolates by host species and country.

	Country									
Host	UK	NL	Finland	Senegal	CZ	DK	NZ	USA	Unknown	Total
Chicken	99	52	33	46	10	6	5	1	1	253
Cattle	180	5	20			1	1		1	208
Sheep	112									112
Sand	50									50
Bird	35									35
Water	24									24
Rabbit	20									20
Pig	16									16
Broiler environment	10									10
Badger	5									5
Soil	2									2
Giraffe	1									1
Pet	1									1
	555	57	53	46	10	7	6	1	2	737

3.2.7 Attribution reference dataset

As a reference dataset for attribution of human cases (chapter 8) the RPS dataset was combined with ruminants from the CSP sampled in the UK between 1998 and 2002 (n=273) and isolates from starlings (n= 285) and wild geese (n=166) from a published study based on a free-range farm in Oxfordshire[271] collected between 2002-2005 and 2002-2003 respectively.

3.3 Analysis

The overall analytical approach to each of the questions identified under project aims in chapter 2 is given here with additional detail in the relevant results chapters. The numbering corresponds to the results chapters 4 to 8 (methods outlined in section 3.4 for chapter 4, section 3.5 for chapter 5 etc). Statistical analyses were undertaken using Stata Version 7. Other software used for particular analyses is indicated in the relevant sections.

3.4 Methods for descriptive molecular epidemiology of CSSS data

All of these analyses used the CSSS dataset to perform descriptive epidemiology of *C. jejuni* with the additional information provided by MLST defined sub-type. Some geographical comparisons additionally use the PHD dataset.

3.4.1 Geographical and demographic analyses

Comparisons between different countries, between the two main study sites in the CSSS sub-study and between different demographic groups (e.g. gender and age groups) were made graphically, by Chi square test for clonal complex distribution and by F_{ST}. For Chi square testing minor clonal complexes were collapsed into a "minor" category to avoid severe violations of Chi square assumptions and in some analyses further merging of categories is undertaken as indicated when numbers were small. Two-by two comparisons used Fisher exact tests. F_{ST} testing used both concatenated sequences and allelic profiles and was performed using ARLEQUIN software[272] with statistical significance assessed by permutation tests using 3,000 iterations. Diversity is compared using Simpson's index of diversity.

3.4.2 Change in subtypes over time

The association of particular subtypes with a seasonal peak was tested by

(1) Defining the seasonal peak within the dataset, dividing isolates into those within and outwith the seasonal peak period and testing differences in clonal complex distribution with a Chi square test. The seasonal peak was defined *a priori* as the 50% of isolates with onset dates closest to the peak day identified from the data. This peak day was identified using a Poisson regression of daily count of cases against day of the year modelled using sine and cosine terms and date as a continuous variable to account for secular trend[71]:

$$\text{Count} = \beta_1*\text{sine}(2\pi*(\text{day of year})/365.25) + \beta_2*\text{cosine}(2\pi*(\text{day of year})/365.25) + \beta_3*(\text{date}).$$

(2) A logistic regression model including coefficients to model season and secular trends as above was used to explore the seasonal patterns of individual clonal complexes compared to all other isolates in more detail. For example:

$$\text{Logit}\ ((\text{ST-45 complex})/(\text{not ST-45 complex})) = \beta_0 + \beta_1*\text{sine}(2\pi*(\text{day of year})/365.25) + \beta_2*\text{cosine}(2\pi*(\text{day of year})/365.25) + \beta_3*(\text{year}).$$

An initial global test used clonal complex as a multinomial outcome variable. Year was modelled as both a continuous and categorical variable to assess any non linear effects of year.

3.4.3 Clonal complex and symptoms

The association between duration of symptoms and clonal complex is assessed using a linear regression model with clonal complex as a categorical variable. The association of other symptoms with clonal complex used univariate and multivariate logistic regression.

3.4.4 Description of population structure

Population structures were described and compared using

1. Clonal complex designation defined from the MLST website[220].
2. The eBURST (Based Upon Related Sequence Types) algorithm[273]

The eBURST algorithm identifies mutually exclusive groups of related genotypes in a set of multilocus sequence types based on sharing of alleles, and attempts to identify the founding genotype (sequence type or ST) of each group. The rules to identify the more central (putatively ancestral) genotypes are based on the isolate with the largest number of single locus variants in the dataset. Where a genotype could be linked to several other isolates as a single locus variant the algorithm preferentially assigns linkages to those STs that have the largest number of single locus variants. The algorithm then predicts the descent from the predicted founding genotype to the other genotypes in the group, displaying the output as a radial diagram, centred on the predicted founding genotype. It was implemented in the standard (most restricted) setting of only identifying relationships where STs share six alleles and differ at one.

3. CLONALFRAME[230]

This is a model based analytical algorithm designed to infer phylogeny from multilocus sequence data in bacteria such as *Campylobacter*. The algorithm assumes that data represent genotypes generated by a clonal genealogy with recombination. Given a dataset of multilocus sequence types the model infers the genealogy, the sequences at each of the internal nodes in this genealogy, the position of the recombined regions in each of the branches of the genealogy, and model parameters for the rate of recombination, rate of mutation, mean length of recombination fragments and the rate of nucleotide differences in recombined stretches. This inference is based on the posterior distributions sampled from a Markov chain Monte

Carlo process. The model generating the distributions has simplifying assumptions including that distant sequence blocks are not affected by a single recombination event and that recombination fragment length follows an exponential distribution. The former is likely to hold for *C. jejuni*[259] and the inaccuracy in the latter should not affect the inferred clonal phylogeny substantially. The assumption that recombination is from outside the dataset studied can lead to incorrect inference of relatedness in deeper branches[230] which needs to be considered in the evaluation of this part of any inferred phylogeny.

CLONALFRAME analyses were performed with 50,000 burn in and 50,000 subsequent iterations sampling each 100th iteration. Three runs were completed, two with default settings and one with an initial UPGMA tree. All trees reached similar values for estimates of recombination rates, fragment lengths and mutation rates. The 50% consensus trees from the three runs were also similar and combined into a joint 65% consensus tree. All CLONALFRAME outputs were redrawn and manipulated for presentation using MEGA software (version 3) [274].

3.5 Source attribution by genotype

This analysis used the 713 isolate CSP dataset comprising chicken, cattle and sheep isolates as per section 3.2.4.

Genetic differentiation between the three populations was assessed using F_{ST} pairwise comparisons as implemented in ARLEQUIN software[275]. This was conducted using the 3309 base pair concatenated sequence as input information and the seven alleles using standard MLST allelic profile number coding. Analysis was conducted on the full

dataset. It was repeated using each ST only once (or once per host if identified in more than one host species) to assess robustness to sampling variation in isolates contributing to this dataset. Statistics for F_{ST} values were estimated using permutation tests within ARLEQUIN.

Population assignment was performed on the data in the forms of ST, Clonal complex and allelic profile. Assignment used STRUCTURE software, a model-based clustering method designed to infer population structure and assign individuals to populations using multilocus genotype data[276]. Differences in genotype frequency between populations allow probabilistic assignment of isolates to populations, even if there is some sharing of genotypes between those populations. The source of the isolates to be assigned was predicted based on a training set that consisted of other relevant isolates by using the USEPOPINFO option, which allows the population of origin to be known for some strains (in this case the training set) while for other strains (the isolates to be assigned) it is assumed unknown.

STRUCTURE estimates the genotype frequencies in each host species based on all of the isolates, as well as estimating the population of origin of isolates of unknown origin, taking into account uncertainty due to sample size. To allow maximum use of the data, some analyses employed a leave-one-out strategy in which a single isolate was assigned using the remaining strains as the training dataset, with the procedure being repeated in turn for each isolate.

The parameters used for all STRUCTURE simulations were a no admixture model with lambda = 1, and gene frequencies uncorrelated between populations. One thousand burn-in cycles and 10000 further repetitions were performed for each analysis.

Empirical assignment accuracy was measured as the average probability p_{k*} with which each isolate was assigned to the correct host source k*. Predicted assignment accuracy is estimated as the average of $\sum_{k=1..K} p_k^2$, where each individual is assigned to one of K different sources.

In supplementary analyses a permutation test was used to assess statistical significance of the association of discordant alleles among non ST-21 isolates of ST-21 complex with host of isolation. The permutation test was performed by randomly permuting the actual host species amongst the predictions obtained from STRUCTURE repeated 10,000 times. The percentage correctly assigned to host was counted from the observed data and compared to the expected distribution under the null hypothesis of no host association generated by the counts from these 10,000 permuted datasets.

3.6 Analysis of the *Campylobacter jejuni* population in chicken

3.6.1 Population structure

The population structure was inferred by CLONALFRAME (more fully described in 3.4.4) and compared with clonal complex designations.

3.6.2 Statistical analysis

Evaluation of variation over time and by source (e.g. country of origin) used Chi square testing. Logistic regression models were used to assess seasonal and secular trends and to assess association of type with origin adjusted for time of sampling. The logistic regression model for temporal effects was of the form:

logit (clonal_complex X/not CC X) = $\beta_0 + \beta_1 \text{sine}(2\pi*(\text{day of year})/365.25) + \beta_2 \text{cosine}(2\pi*(\text{day of year})/365.25) + \beta_3[\text{year} = 2004] + \beta_4[\text{year} = 2005]$

Likelihood ratio tests were performed following removal of either the year terms or seasonal terms from this model against the full model. The six most abundant clonal complexes were modelled (at least 30 isolates in each) as well as a collection of unassigned and rare clonal complexes (less than 12 isolates in the dataset). To assess possible confounding of the association between country of origin and clonal complex by time of sampling the model was extended to include country of origin and the association with type reviewed in the full model and with year terms removed. The comparison between fresh or frozen status and point of sale and clonal complex is similarly tested in unadjusted models and adjusted for year and country of origin in a logistic regression model.

3.6.3 Population assignment

Comparison of the retail poultry populations with cattle, sheep and poultry isolates (the CSP dataset) and other between host comparisons (using isolates from collections summarised in "International published animal and environmental datasets", section 3.2.6) used allele based population assignment with STRUCTURE software as described in 3.5. Reference populations were separate from assigned populations in each case.

3.6.4 Discrimination

Estimation of the degree of discrimination afforded by increasing the number and diversity of sequenced genes was indexed using Simpson's index of diversity.

Identification of the proportion of recent change in each gene fragment was assayed by making two-way comparisons between nine locus sequence types (MLST, *flaA* and *porA*). Comparisons where eight of the nine alleles were shared were counted and in each of these cases the locus with non-identical alleles identified and added to the count for that gene. The process was repeated counting those sequence types identical

at seven of the nine loci and then further repeated using isolates rather than entering each sequence type only once. This was implemented in a Stata programme script.

3.7 Simulation of sequencing approaches for outbreak investigation

In this study an arbitrary model population of isolates, some epidemiologically linked and some unrelated, was assumed and the impact of microbial subtyping of isolates from this population assessed. The main aim for subtyping was to identify outbreaks (epidemiologically linked cases) by picking them out as having an identical type which was not shared by cases that were not part of the outbreak. Subtyping was simulated assuming a range of parameter values for the characteristics of the typing scheme, in particular for discriminatory index.

3.7.1 Population

The model population used comprised 500 cases, approximately the annual number of detected cases expected in a UK population of 500,000 people, with the assumption that half are truly sporadic (unique), that small clusters of 2 to 3 identical isolates are quite common and comprise a further 20% of the total and that larger outbreaks are rare and inversely related to outbreak size. The model population of surveillance cases based on these criteria is summarised in Table 3.11 and is the "true" population of cases and outbreaks considered in all analyses.

Table 3.11. Model population of cases

Epidemiological cluster size	Occurrence	Isolates
1 (sporadic cases)	250	250
2	35	70
3	10	30
4	7	28
5	4	20
6	2	12
7	2	14
8	1	8
9	1	9
10	1	10
13	1	13
16	1	16
20	1	20
Total		500

3.7.2 Simulation of typing

The results from typing were estimated for levels of DI of 90%, 95%, 99%, 99.5% and 99.9% with 100% reproducibility and for DI 99.5% with reproducibility of 95%. This was implemented for each level of discrimination by the following process:

1. The isolates, identified by an identity and a cluster number indicating their epidemiological cluster identity, were permuted into a random order.

2. The first isolate in the randomised order was assigned to typing group number one. All isolates of the same cluster were then assigned the same number subject to reproducibility criteria i.e. to simulate 100% reproducibility all were assigned, to simulate 95% reproducibility it was assigned probabilistically to each with 95% probability. A proportion of the remaining non-identical isolates were then assigned to typing group 1 using a probability that would produce the required discriminatory index. This probability was identified iteratively as described below. For isolates belonging to true clusters the probability was scaled so that the probability of the cluster being assigned to typing group number 1 was equal to the probability for

sporadic isolates. If one isolate in a cluster was assigned to typing group 1 all other isolates in the cluster were then probabilistically assigned the same number with probability based on reproducibility.

3. The process was repeated by finding the next isolate in the dataset without an assigned type, numbering it with a new typing group number and repeating the above process to include other isolates into this type group.

4. Further repetitions continued until the all isolates had a typing group number.

3.7.3 Scaling simulation to produce correct results

The simulation was tested empirically to ensure that the correct DI was achieved. This was done by calculating DI using one randomly selected isolate from each true genotypic cluster for each simulation. The simulation process was repeated 100 times to assess whether the process produced the correct DI (a distribution of DIs centred on the correct value). Where the correct result was not obtained probability parameters were altered and the full process repeated until the planned levels of DI were obtained. Reproducibility was also confirmed empirically by simulation for reproducibility of 95% meaning that isolates in a cluster had a 95% probability of being typed into the typing group that contained the majority of isolates from that cluster.

3.7.4 Selection of simulation results presented

250 simulations were then carried out for each level of DI and reproducibility and results stored. Results are presented across the 250 datasets for each level of DI and some more detailed results using one typical dataset per parameter set. These datasets were selected to be typical in each case by the criteria that the DI was at the median of the group to 4 decimal places and the size of the typing group in which the true cluster of 20 isolates fell was within 1 of the median for the 250 datasets.

3.8 Genotype and source of human infection

This chapter jointly analyses data from different parts of the thesis to address questions of the use of sequence typing to identify source of human infection.

3.8.1 Seasonality in human and chicken samples

The main clonal complexes showing temporal trends among human isolates (Chapter 4) were graphed using data from human isolates in the CSSS and retail poultry isolates in the RPS to show the monthly proportion of isolates belonging to these clonal complexes along with 95% binomial exact confidence intervals for each time point.

Joint analysis for the combined ST-45 and ST-283 clonal complexes was performed using a logistic regression model predicting membership of one of these complexes based on host, year (continuous variable) and season (modelled as in section 3.4.2):

Logit ((ST-45 complex or ST-283 complex)/(not one of these complexes)) = β_0 + β_1*sine(2π*(day of year)/365.25) + β_2*cosine(2π*(day of year)/365.25) + β_3*(year) + β_4*(host)

Differences in seasonality between humans and retail poultry were assessed by addition of interaction terms as below and performing a likelihood ratio test on the model with and without these terms:

…+β_5*(host)*sine(2π*(day of year)/365.25) + β_6*(host)*cosine(2π*(day of year)/365.25)

3.8.2 Comparing genotypic attribution with risk factor information

3.8.2.1 Attribution

Isolates from the CSSS dataset were assigned to probable host of origin using the STRUCTURE algorithm as described above with default settings run for 50,000 iterations after a burn-in of 50,000 iterations and the attribution reference dataset (section 3.2.7).

Isolates were assigned to one of the four options of chicken, ruminant, starling and geese when the predicted probability from the algorithm was greater than 0.75 for one population. The results were reviewed qualitatively given the limitations of the algorithm and of the reference datasets and a restricted subset of the assignment predictions used to compare with the questionnaire data from the CSSS. This is described more fully in the results of this analysis (Chapter 8.3).

3.8.2.2 Comparison of genotypic with exposure history

Predicted origin based on genotype was compared with exposure history from patient questionnaires. The questionnaire completed by cases comprised 85 exposures across a wide range of food consumption, water consumption, recreational water activity, animal contact, and contact with other cases (Annex 2). All exposures were tested in all four sets of analyses (chicken vs ruminant and starling, ruminant vs chicken and starling, starling vs chicken and ruminant, ST 45 or 283 complex vs non ST-45 or 283 complex during the peak period) by univariate logistic regression. Food exposures classified as consumption "never", "once" or "more often" in the two weeks before illness were classified as 0,1, and 2 respectively with the assumption of risk increasing similarly at each increased level of exposure on a multiplicative scale. Similarly for

water consumption (glasses per day) and the quantitative measure of handling raw meat (steps 0, 1, 2-5, 6-10, 11 or more). For household size (for each of children, adults and total) analysis assumed a similar grading of risk associated with household size up to size 10 but no increase above this size.

Given the large number of possible associations the primary outcomes were defined as follows:

1. Association of chicken genotype with chicken consumption (the subject of just one exposure question).

2. Association of ruminant genotype with consumption of lamb or unspecified meat (the four exposure categories Halal meat, Lamb, Meat pies and organic meat are included) or unpasteurised milk.

3. Association of starling (wild bird) genotypes with outdoor exposures (four recreational water exposures and the consumption of river water) or consumption of bird-pecked milk).

Apart from these predefined exposures – subtype hypotheses other observed associations must be considered in the light of the 340 potential analyses and consequently high risk of detecting significant associations by chance.

Positive associations (at $p < 0.05$) were evaluated further by detailed tabulation and multivariate analyses as appropriate, which is described in the results section.

3.8.3 Use of discriminatory typing to investigate outbreaks

The human case control (CaCo) dataset was used to identify genotypic clusters of infection defined as isolates with an identical sequence type at the seven MLST loci, the short variable region of *flaA* and the sequenced variable region of *porA*.

Tabulation and graphing was used to describe spatio-temporal clustering of identical isolates.

Temporal clustering was evaluated formally by

1) Measuring the numbers of days between all pairs of isolates.
2) Grouping according to whether the comparison was within a cluster (i.e. time between two identical isolates) or between clusters (i.e. non-identical isolates).
3) Comparison of the distributions from 2) above by rank-sum test.

Genotypic clusters defined by nine-locus identical genotype in the CaCo dataset were compared with the RPS dataset to assess the possibility of chicken as a source and to describe the dynamics of these highly resolved subtypes.

4. Subtypes of *Campylobacter jejuni* in human populations

This chapter integrates MLST data in the descriptive epidemiology of *C. jejuni* from the structured sample of isolates available in the CSSS dataset. Of the 1596 cases identified for inclusion the laboratory sample was either missing from the archive or contaminated such that it was not possible to recover the *C. jejuni* strain for 36 (2.3%) (Table 4.1). Six isolates proved to be *C. coli* on MLST and seven appeared to represent mixed cultures that could not be resolved. Of the remaining 1547 *C. jejuni* isolates available to the study 1505 were fully typed by MLST while profiles were not completed for 42. Incomplete profiles were those where template was exhausted before completion of typing.

Table 4.1. Outcome of laboratory typing for 1596 cases of *C. jejuni* identified for inclusion in the study.

Outcome of MLST		
Completed *C. jejuni*	1505	94.3%
C. coli	6	0.4%
Unresolved mixes	7	0.4%
Missing	30	1.9%
Contaminated	6	0.4%
Incomplete	42	2.6%
	1596	100.0%

The demographic characteristics of cases with a complete MLST profile were similar to those for whom a complete profile was not available with a complete profile for 94% of Nottingham cases and 94.5% of Hampshire cases, for 94% of both male and female cases and with similar age distributions among cases with a complete profile (mean 42.5 years, standard deviation 20.9) and without (mean 44.9 years, standard deviation 19.5). The allele distributions of cases with complete and partial MLST profiles were broadly similar [Annex 3]. Given the lack of evidence to suggest that any missing data is likely to be biased in terms of ST or allele distribution missing data is not considered in analyses.

Although aiming to recruit all cases over the full three-year period recruitment was not always complete and may have varied over time. Table 4.2 shows the numbers of isolates from the three years of the study, with a probable fall off in recruitment in year 3 (the drop exceeds the drop in incidence estimated from national data at this time).

Table 4.2. Typed isolates by location and study year.

Year	Nottingham	Hampshire	Total
May 2000-April 2001	316	276	592
May 2001-April 2002	292	262	554
May 2002-April 2003	188	112	300
Total	796	650	1446*

* Only those cases reporting onset of illness. Of the 59 cases not recording onset date an estimated date based on laboratory data suggested onset of 17, 30 and 12 cases in year 1, 2 and 3 respectively.

4.1 Geographical distribution of sequence types

Spatial distribution of subtype is assessed across regional and global distances using the two geographically separated English populations from the CSSS and by comparison of the full CSSS dataset with those published studies that report full detail on their population MLST types.

4.1.1 Two geographically separate populations in England

The distribution of clonal complex in the two study populations is almost identical as shown in Figure 4.1. Full data is given in Annex 3. There is no statistical support for variation between these two populations assessed at the clonal complex level (χ^2 p=0.79 for the data in Figure 4.1) or at the nucleotide (F_{ST} value of 0, p = 0.4) or allele (F_{ST} value of 0, p = 0.6) levels. Diversity is similar in the two population samples, with Simpson's Index of Diversity for sequence type 0.947 in the isolates from Nottinghamshire and 0.952 in those from Hampshire.

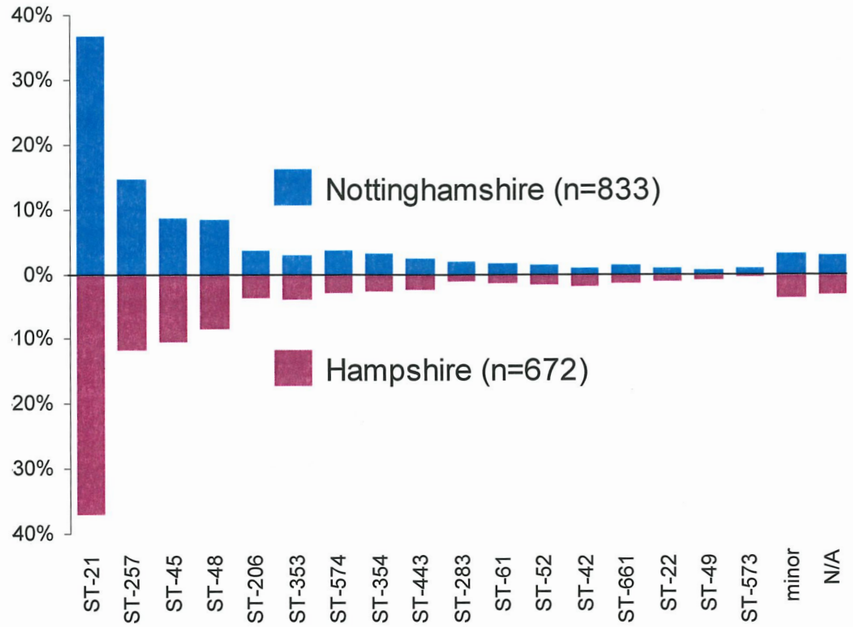


Figure 4.1. Clonal complex distribution* of 1505 isolates in two geographically distinct populations, May 2000 – April 2003 (likelihood ratio χ^2 test p = 0.79).
*Clonal complexes with less than 10 isolates in the dataset are combined grouped as "minor" clonal complexes and isolates not assigned to a clonal complex as N/A.

4.1.2 Comparison with international datasets

The data from this study is compared with population samples of 107 isolates from New Zealand (winter 2006), 153 isolates from Australia (1999-2001) and 298 isolates from Finland (1996, 2002 and 2003). These are summarised in Figure 4.2 with full detail in Annex 4. Calculation of F_{ST} values between countries supported substantial genetic differentiation (Table 4.3) and showed the Finnish sample to be most distant from the others, in particular from the Australian and New Zealand samples.

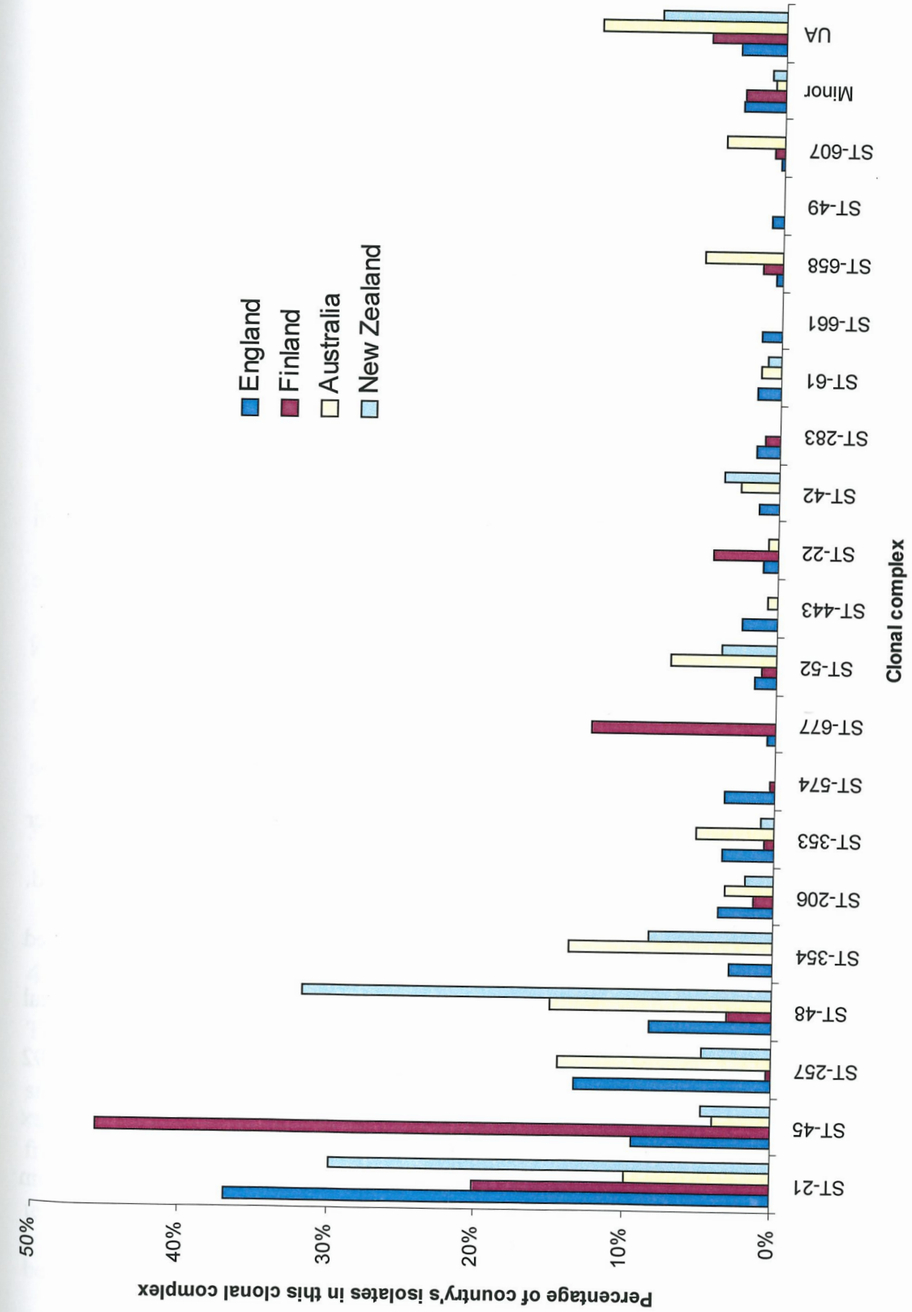


Figure 4.2. International comparisons of clonal complexes of *C. jejuni* in population based samples.

Table 4.3. F_{ST} values for *C. jejuni* isolates from human infections in four countries, allele based above diagonal and sequence based below. All differences are statistically significant with $p < 0.0001$.

Country	England	Finland	Australia	New Zealand
England		0.116	0.053	0.028
Finland	0.154		0.173	0.177
Australia	0.035	0.214		0.073
New Zealand	0.032	0.239	0.033	

A Chi square test on the data in Figure 4.2 with all clonal complexes represented by less than 40 isolates collapsed into a combined "minor clonal complex" category confirms the strong statistical support for variation between countries (Pearson $\chi^2 = 754$ with 33 degrees of freedom, $p < 10^{-136}$).

The most frequently seen clonal complex overall, ST-21 complex was geographically widespread. It was the most common complex in England, the second most common in Finland and New Zealand and the fourth most common in Australia. ST-48 complex was the most abundant complex in both Australia and New Zealand, the fourth most abundant in England and the fifth in Finland.

Forty six percent of isolates in Finland were of ST-45 complex, which is a fivefold higher relative abundance than seen in England and ten fold higher than Australia and New Zealand. Overall ST-45 comprised 62% of isolates in this complex and ST-137 12%, with no marked differences between countries in the ST distribution within the complex. The other unusual feature of the Finnish data is the prevalence of ST-677 complex, which accounted for 16 of 92 isolates in 1996 and 21 of 206 in 2003-3. Eighty percent of these were ST-677. This complex was otherwise rare with just 8 ST-677 complex isolates in the English dataset and none from Australia or New Zealand.

ST-257 complex is contrastingly rare in Finland (just 1 isolate, ST-824, a three locus variant which is usually assigned to ST-257 complex but fits similarly well in ST-354 complex) but very common in England and Australia, with 13% and 14% of isolates respectively, and comprising 5% of isolates in New Zealand.

Other marked variations between countries involved ST-574 complex and ST-354 complex. There were 50 ST-574 isolates in England with this ST absent from other countries, and just one ST-574 complex isolate seen elsewhere, a single locus variant in Finland. ST-354 complex was absent from the Finnish dataset, but particularly common in the Australian and New Zealand data. More minor variations by country include the relatively high prevalences of ST-52 complex in Australia and New Zealand and ST-22 in Finland.

4.2 Temporal patterns among subtypes

Given the marked and largely unexplained seasonal pattern of human *C. jejuni* infection this section reviews the temporal distribution with the added dimension of MLST using the three year CSSS sample which allows separation of seasonal from transient or secular changes.

4.2.1 Temporal distribution by clonal complex

The distribution of isolates by month across the study is shown in Figure 4.3, with peaks each summer in line with expectations. This figure also shows the complexity of contributions from different subtypes, even when aggregated to the clonal complex level. A single global test of variation in contribution by different subtypes to the summer peak was preformed first before considering clonal complexes individually or more complex seasonal patterns. Isolates were divided into two approximately equal groups by estimating a peak day from a Poisson regression of daily counts (day 191) and considering the 50% of isolates occurring closest to

this day as peak and others as non-peak. This showed clear evidence for differences in the distribution of isolates by clonal complex in the peak and non-peak periods (χ^2 test $p < 0.00001$). F_{ST} values between peak and non-peak populations are small but also highly significant when calculated using nucleotides (0.011, $p<0.0001$) or alleles (0.007, $p<0.0001$). Given good statistical support for variation in clonal complex distribution overall, individual clonal complexes were assessed (Table 4.4). ST-45 complex and ST-283 complex isolates form a substantially greater proportion of the total during the peak period compared to the non-peak period. ST-257 complex shows the opposite pattern, forming a larger proportion of isolates in the non-peak months. Although less extreme, this is well supported statistically given the large numbers of ST-257 complex isolates overall. Variation away from 50% is seen among other clonal complexes but with less robust statistical support due to smaller numbers of isolates in these complexes.

Table 4.4. Tabulation of clonal complexes and occurrence in the peak period or outside it (global χ^2 square test $p < 0.00001$, individual p values < 0.05 in bold type in table).

Clonal complex	Non-peak	Peak	Percent	Total	Fisher exact p
ST-21 complex	274	283	51%	557	0.63
ST-257 complex	121	80	40%	201	**0.002**
ST-45 complex	43	99	70%	142	**$<10^{-6}$**
ST-48 complex	61	64	51%	125	0.78
ST-206 complex	32	24	43%	56	0.34
ST-353 complex	31	21	40%	52	0.20
ST-574 complex	30	20	40%	50	0.20
ST-354 complex	25	19	43%	44	0.45
ST-443 complex	24	12	33%	36	0.06
ST-283 complex	4	19	83%	23	**0.001**
ST-61 complex	9	14	61%	23	0.30
ST-52 complex	15	7	32%	22	0.13
ST-42 complex	13	8	38%	21	0.38
ST-661 complex	9	12	57%	21	0.52
ST-22 complex	8	7	47%	15	1.00
ST-49 complex	4	8	67%	12	0.26
ST-573 complex	6	4	40%	10	0.75
Other	23	27	54%	50	0.57
N/A	23	22	49%	45	1.00
Total	755	750	50%	1505	

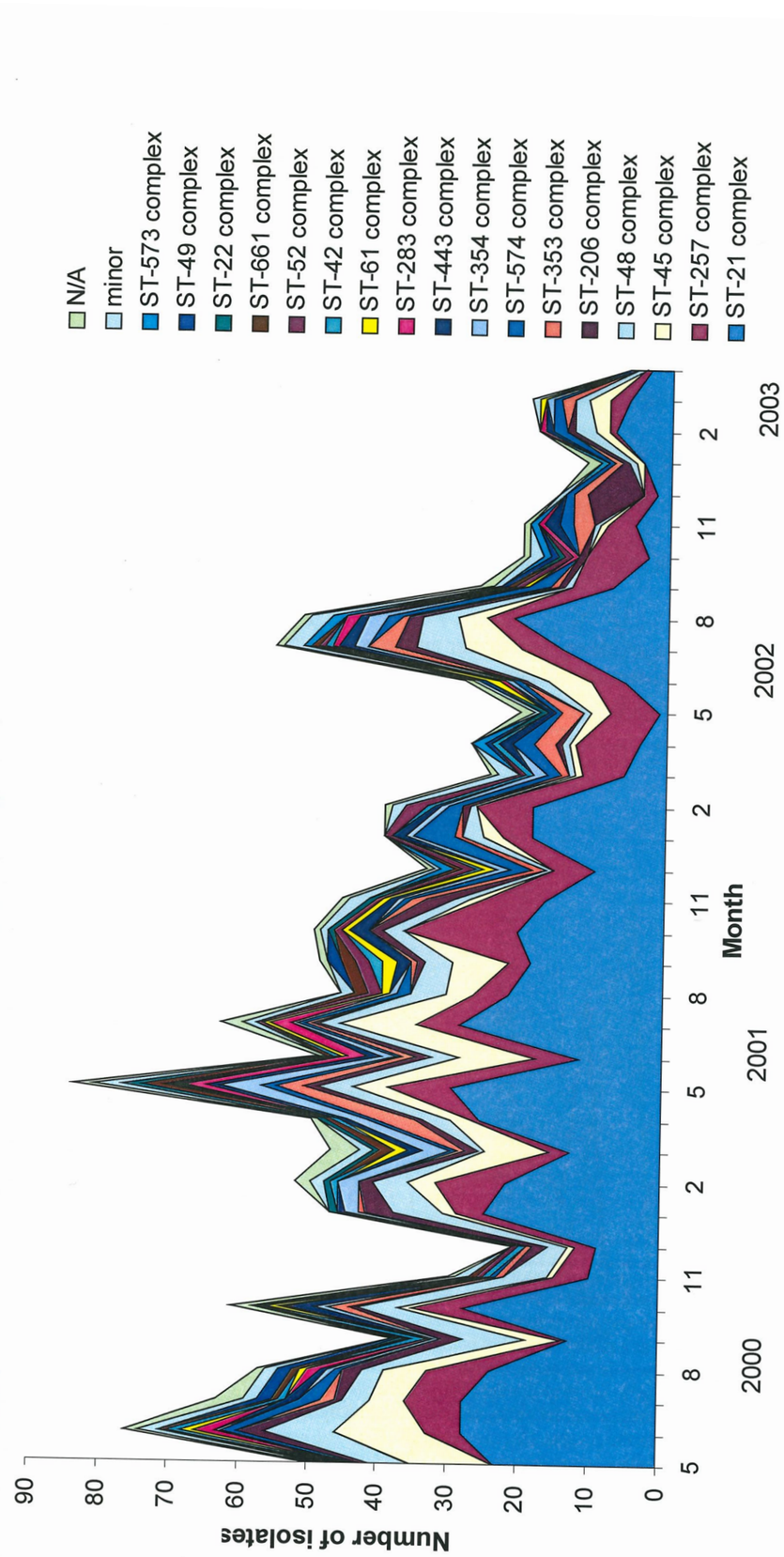


Figure 4.3. Distribution of isolates by clonal complex and time for which patient reported date of onset is available (n=1446).

A more detailed regression model analysis of secular and seasonal trends is summarised in Table 4.5. There is again strong statistical evidence for both seasonal and secular change in the relative prevalence of clonal complexes. The main long-term trends in abundant clonal complexes are of ST-21 complex decreasing, and ST-257 and ST-574 complex increasing (Table 4.5 and Figure 4.4 Panel A). Among the less abundant complexes ST-353 complex also increases over the study period (Figure 4.4, Panel B). ST-573 complex appears to emerge during the study with 1 isolate in the first half and 9 across the second half, although total numbers are small.

Table 4.5. Statistical evidence for secular and seasonal variation of clonal complexes in human *C. jejuni* infection in Hampshire and Nottinghamshire, May 2000-April 2003, n = 1505 (p values < 0.05 highlighted in bold type).

Clonal complex	Number (in CC)	Seasonal†	Year‡	Linear effect (year)¶
Overall *		**<0.0001**	**<0.0001**	**0.0009**
ST-21	557	0.56	**0.001**	**0.8 [0.70-0.92]**
ST-257	201	**0.02**	**0.004**	**1.29 [1.06-1.55]**
ST-45	142	**<0.0001**	0.40	
ST-48	125	0.98	**0.02**	0.80 [0.63-1.02]
ST-206	56	0.07	0.99	
ST-353	52	0.52	0.11	**1.48 [1.04-2.10]**
ST-574	50	0.29	**0.003**	**1.73 [1.20-2.47]**
ST-354	44	**0.01**	0.12	
ST-443	36	0.92	0.78	
ST-283	23	**0.002**	0.69	
ST-61	23	0.55	0.45	
ST-52**	22	**0.04****	0.11	
ST-42	21	0.61	0.33	
ST-661**	21	0.56	0.11	
ST-22**	15	0.23	0.31	
ST-49**	12	**0.04****	0.23	
ST-573**	10	0.87	**0.02****	**2.38 [1.05-5.42]**
Unassigned	45	0.83	0.97	
Minor	50	0.78	0.12	

* Multinomial regression using year and sine and cosine transforms of day of year.
† Likelihood ratio test of a model to predict being a member of the clonal complex under study using season modelled by sine and cosine terms and year as a categorical variable.
‡ Likelihood ratio test of a model to predict being a member of the clonal complex under study using year as a categorical variable and including season modelled by sine and cosine terms.
¶ Likelihood ratio test of a model to predict being a member of the clonal complex under study using year as a continuous variable and including season modelled by sine and cosine terms. Inclusion of day instead of year gives similar results.
** No isolates in these clonal complexes present in 2003 sample reducing sample size to 1446 for analyses with year as a categorical variable. Analysis of seasonal effect with year as a continuous variable to include the full dataset gave similar statistical support for the seasonal variation of ST-52 complex and ST-49 complex.

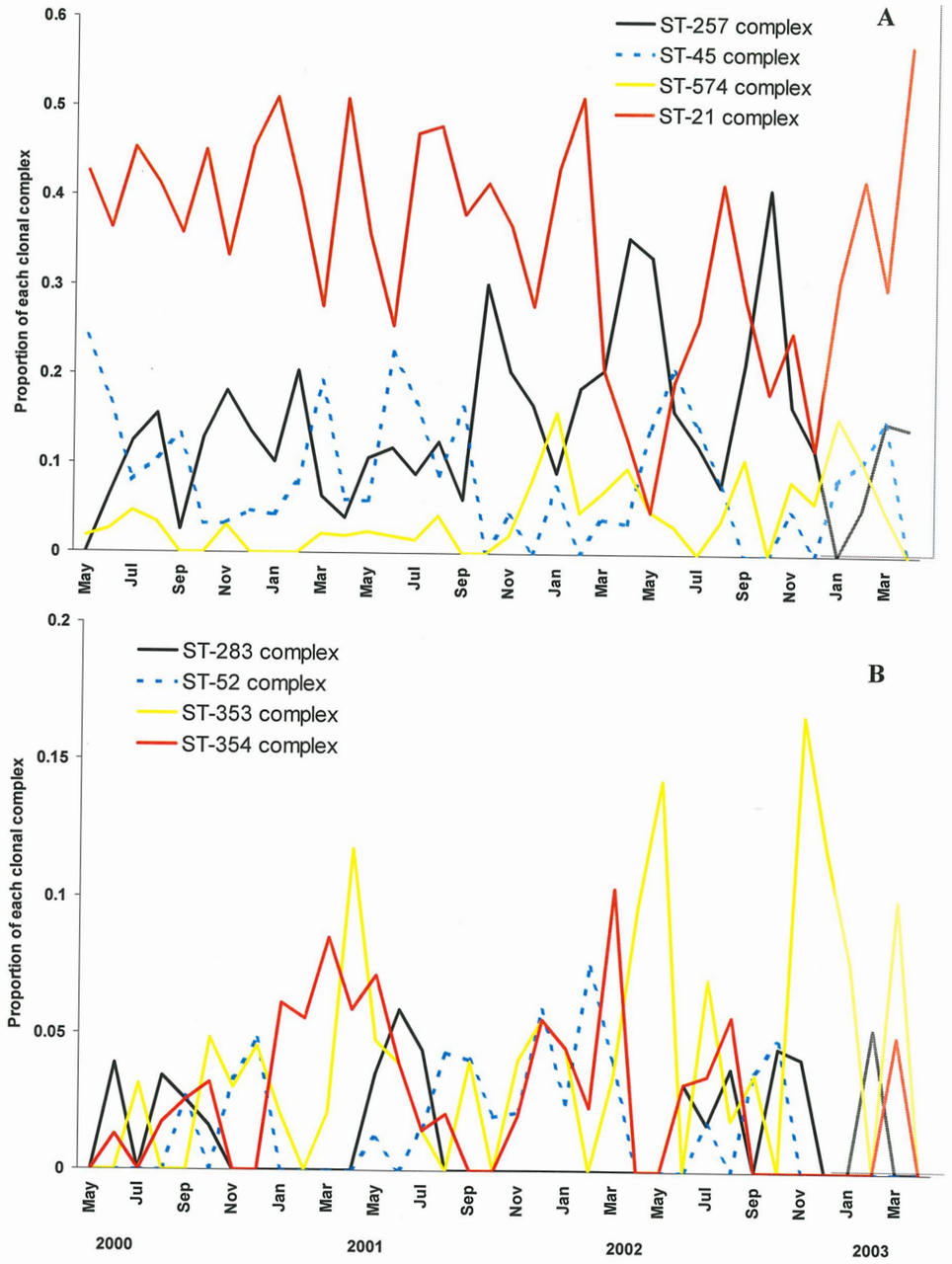


Figure 4.4. Distribution of clonal complexes with evidence for secular or seasonal trends in time, showing the proportion of monthly isolates in each month due to each clonal complex. **A.** Major complexes **B.** Minor complexes.
Fewer than 20 isolates per month in the last 4 months of the study makes estimates unreliable in this time period.

Seasonal patterns were reviewed by graphing both model fitted values and raw data of clonal complexes for which there was good statistical support for secular or seasonal trends. The raw data for these is graphed in Figure 4.4. The strongest seasonal effects among the major clonal complexes were for ST-45 complex and ST-283 complex, which showed a seasonal peak in relative incidence at the same time as the seasonal peak in overall disease incidence is observed and for ST-257 which show an opposite, but less marked seasonal pattern, in line with the peak - non-peak tabulation. Among less abundant clonal complexes ST-354 peaks in the first half of the year while ST-52 complex and ST-49 complex (not shown) are relatively more common in the later part of the year.

4.2.2 Temporal distribution by sequence type

In comparing the peak to non-peak distribution of other individual STs the most marked differences are in ST-45, ST-257 and ST-267 each the most common ST in ST-45, ST-257 and ST-283 clonal complexes respectively. Analysis of other individual STs does not show any notable findings (tabulation in Annex 5).

4.3 Association of demographic characteristics with genotype

4.3.1 Age

The distribution of cases by age is summarised in Figure 4.5. Consistent with other studies there is a higher rate of diagnosis in young children, followed by a drop in later childhood before rising again in adolescence and young adulthood. The further peak in the 50 to 59 year old age group has also been identified across the fuller dataset from this sentinel surveillance scheme. Factors such as age specific population size, varying probability of diagnosis with age, varying risk of disease following exposure and differential exposures by age may contribute to such patterns.

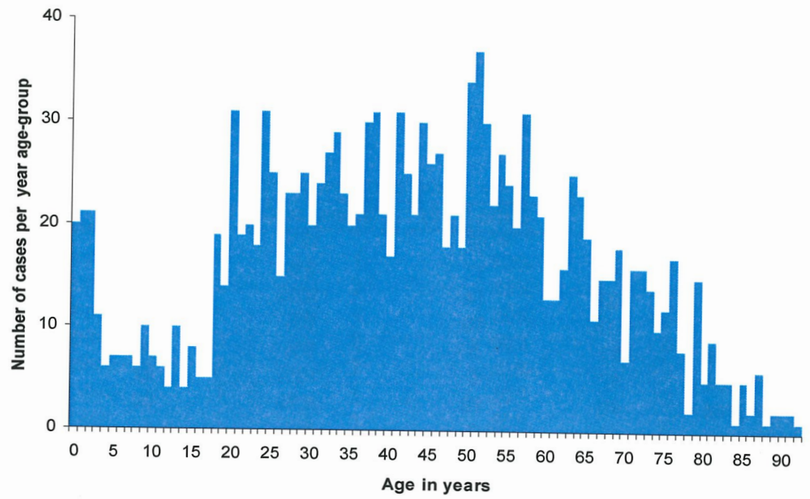


Figure 4.5. Number of cases in the study according to age of participants (in years).

A report that older people have more diverse types[277] suggests that age related subtype distribution may contribute to the observed demographic pattern. This guided analysis of diversity of subtypes by age group as well as the association of age and subtype per se.

4.3.1.1 Comparing those aged 0-17 with those aged 18 and over

There was some variation in distribution of the main clonal complexes between those under 18 and aged 18 and over (χ^2 test p = 0.007), although differences are not very substantial for most clonal complexes. F_{ST} analysis shows minor differentiation with F_{ST} 0.003 (p =0.05) for alleles and 0.004 (p=0.04) for nucleotide sequence. The types which appear to differ most were ST-283 complex seen almost 3 times as often as expected among those under age 18 (Fisher exact p = 0.001) and ST-257 complex seen about half as often as expected (Fisher exact p = 0.01) (Table 4.6). For other clonal complexes differences were relatively small and may well reflect chance alone. The diversity of STs is similar among children (Simpson's Index of Diversity 0.95) or

adults (0.95) in contrast to the increased diversity reported among adults in analyses of serotypes[277].

Table 4.6. Distribution of isolates by clonal complex in children and adults.

Clonal complex	Adult	Child (<18)	Percent under 18	Total	Fisher p
ST-21 complex	490	67	12%	557	>0.05
ST-257 complex	190	11	5%	201	**0.01**
ST-45 complex	128	14	10%	142	>0.05
ST-48 complex	115	10	8%	125	>0.05
ST-206 complex	51	5	9%	56	>0.05
ST-353 complex	47	5	10%	52	>0.05
ST-574 complex	40	10	20%	50	>0.05
ST-354 complex	42	2	5%	44	>0.05
ST-443 complex	30	6	17%	36	>0.05
ST-283 complex	16	7	30%	23	**0.001**
ST-61 complex	18	5	22%	23	>0.05
ST-52 complex	19	3	14%	22	>0.05
ST-42 complex	19	2	10%	21	>0.05
ST-661 complex	20	1	5%	21	>0.05
ST-22 complex	14	1	7%	15	>0.05
ST-49 complex	11	1	8%	12	>0.05
ST-573 complex	10	0	0%	10	>0.05
minor	39	11	22%	50	**0.02**
N/A	41	4	9%	45	>0.05
Total	1340	165	**11%**	1505	

4.3.1.2 Children age 0-3 compared to children aged 4-17

Younger children had no strong overall difference in infecting clonal complexes compared to older children (χ^2 test $p = 0.07$). At the allele and concatenated sequence level differences were also small (0.012 for alleles and 0.017 for sequence) and marginally statistically significant ($p < 0.04$ & 0.02 respectively). The types with strong seasonal patterns, ST-45 complex and ST-283 complex appear more prominent in younger children than older children, while ST-574 complex, which appears to be relatively recent in this dataset is more common in older children than expected by chance (Table 4.7). However the marginal statistical support for these in the context of multiple post-hoc testing means that these associations are speculative. The diversity of ST in younger (Simpson's Index of Diversity (0.96) and older children (0.95) is similar.

Table 4.7. Distribution of isolates by clonal complex and age in children.

Clonal complex	Under 4	4 to 17	Percent under 4	Total	Fisher p
ST-21 complex	28	39	42%	67	>0.05
ST-45 complex	10	4	71%	14	**0.047**
ST-257 complex	4	7	36%	11	>0.05
ST-48 complex	2	8	20%	10	>0.05
ST-574 complex	1	9	10%	10	**0.044**
ST-283 complex	6	1	86%	7	**0.045**
ST-443 complex	5	1	83%	6	>0.05
ST-206 complex	2	3	40%	5	>0.05
ST-353 complex	1	4	20%	5	>0.05
ST-61 complex	3	2	60%	5	>0.05
ST-52 complex	1	2	33%	3	>0.05
ST-354 complex	1	1	50%	2	>0.05
ST-42 complex	1	1	50%	2	>0.05
ST-22 complex	1	0	100%	1	>0.05
ST-49 complex	0	1	0%	1	>0.05
ST-661 complex	0	1	0%	1	>0.05
minor	6	5	55%	11	>0.05
N/A	1	3	25%	4	>0.05
Total	73	92	**44%**	165	

4.3.1.3 Older adults (60 or over) and younger adults and children (under 30)

There is statistical support for some variation in identified types of *C. jejuni* with age when comparing older individuals (age 60 or more) with those under 30 (χ^2 test $p < 0.0001$, Table 4.8). Similarly comparison of alleles or nucleotide sequences shows $F_{ST} = 0.012$ for alleles and 0.017 for sequence with $p < 0.00001$ in each case. Only ST-257 complex ($p < 0.001$) and ST-574 complex ($p = 0.002$) have statistical evidence for being more and less frequent among older cases respectively than would be expected by chance. Although showing opposite patterns of age distribution both of these complexes increased in relative prevalence during the course of the three-year study (Table 4.5 and Figure 4.4 Panel A). Neither Simpson's index of diversity (0.94 for both groups) nor the proportion of infections due to isolates from uncommon clonal complexes or not assigned to any clonal complex (Table 4.8) is supportive of the older population being mainly infected with unusual types.

Table 4.8. Distribution of isolates by clonal complex and age, comparing those 60 or older with those under 30 (Chi square $p < 0.0001$).

Clonal complex	Under 30	60 or older	Percent over 59	Total
ST-21 complex	173	99	36%	272
ST-257 complex	34	64	65%*	98
ST-45 complex	32	35	52%	67
ST-48 complex	37	26	41%	63
ST-574 complex	25	5	17%*	30
ST-206 complex	12	17	59%	29
ST-353 complex	10	13	57%	23
ST-354 complex	10	10	50%	20
ST-443 complex	15	5	25%	20
ST-283 complex	14	5	26%	19
ST-52 complex	10	5	33%	15
ST-61 complex	5	9	64%	14
ST-661 complex	8	1	11%	9
ST-49 complex	4	4	50%	8
ST-22 complex	4	2	33%	6
ST-42 complex	3	3	50%	6
ST-573 complex	2	4	67%	6
Minor	16	13	45%	29
N/A	14	9	39%	23
Total	428	329	**43%**	757

* ($p<0.01$)

4.3.2 Ethnic group

Most isolates were from individuals with white or unknown ethnic groups (Table 4.9)

Table 4.9. Self reported ethnic groups of participants.

Ethnic group (Office for National Statistics classification)	Freq.
Asian British Bangladeshi	1
Asian British Indian	2
Asian British / Other Asian	10
Asian British Pakistani	2
Black British Black African	2
Black British Black Caribbean	5
Black British / Other Black	3
Mixed	2
Mixed White and Asian	2
Mixed White and Black Caribbean	1
Other ethnic group	1
Other ethnic group - Chinese	3
Unknown	308
White British	1147
White Irish	3
White Other White	13
Total	1505

Comparison of the clonal complex distribution among different ethnic groupings does not show any strong ethnic differences. When combining groups to form an Asian, a

Black, a White, a mixed and an "Other and unknown" ethnic group those clonal complexes appearing 2 or more times (maximum 4) in Black or Asian or mixed ethnicity cases also appeared at least 50 times in the dataset as a whole (Table 4.10). The distribution in the White and the "Other and unknown" groups were similar. No evidence is therefore available for association of type and ethnicity in this study.

Table 4.10. Distribution of isolates by clonal complex and ethnic groups of participants.

Clonal complex	Asian	Black	Mixed	Unknown or other	White	Total
ST-21	4	1	3	130	419	557
ST-22	0	0	0	4	11	15
ST-42	0	0	0	4	17	21
ST-45	3	1	0	33	105	142
ST-48	0	0	1	32	92	125
ST-49	0	0	0	4	8	12
ST-52	1	0	0	2	19	22
ST-61	0	0	0	6	17	23
ST-177	0	0	0	0	1	1
ST-206	0	0	0	8	48	56
ST-257	1	4	1	36	159	201
ST-283	1	1	0	4	17	23
ST-353	0	3	0	5	44	52
ST-354	0	0	0	7	37	44
ST-403	0	0	0	0	5	5
ST-433	0	0	0	1	1	2
ST-443	1	0	0	3	32	36
ST-446	0	0	0	2	0	2
ST-460	0	0	0	2	7	9
ST-508	0	0	0	2	6	8
ST-573	0	0	0	3	7	10
ST-574	2	0	0	14	34	50
ST-607	0	0	0	1	3	4
ST-658	0	0	0	0	7	7
ST-661	0	0	0	2	19	21
ST-677	0	0	0	1	7	8
ST-682	0	0	0	0	1	1
ST-692	0	0	0	0	1	1
ST-1275	0	0	0	0	1	1
ST-1332	0	0	0	0	1	1
Unassigned	0	0	0	8	37	45
Total	13	10	5	314	1163	1505

4.3.3 Gender

There is no evidence for differences in the subtypes infecting men and women (Figure 4.6). Diversity of subtypes is also similar (Simpson's index 0.95 for both genders).

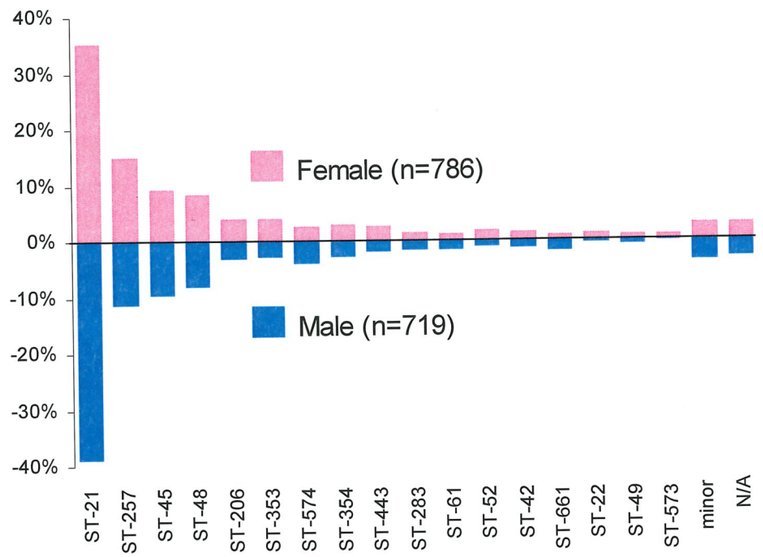


Figure 4.6. Gender distribution of the major clonal complexes among male and female cases, (Chi square p value = 0.41).

4.4 Clonal complex and symptoms

There was no overall evidence for variation of duration of illness with clonal complex ($p = 0.11$ for linear regression using major clonal complexes as categorical variables). Comparison by clonal complex of median duration of illness shows only 1 (ST-49 complex) with a duration illness of more than 5 days different from the overall median, which is based on a group of just 11 patients (Table 4.11). Results are therefore inconsistent with very great differences in median duration of illness among those diagnosed with infection due to different clonal complexes.

Table 4.11. Average duration of illness by clonal complex of *C. jejuni* isolates.

Clonal complex	Mean days ill	Median days ill	Number with data
ST-21 complex	15.9	15	540
ST-257 complex	15.1	15	191
ST-45 complex	16.5	17	137
ST-48 complex	15.4	15	122
ST-206 complex	12.9	12	56
ST-353 complex	17.2	18	49
ST-574 complex	18.6	19	47
ST-354 complex	15.7	16	41
ST-443 complex	14.2	13	35
ST-283 complex	15.7	14	23
ST-61 complex	15.2	13.5	22
ST-42 complex	17.9	20	21
ST-52 complex	12.8	10	21
ST-661 complex	17.1	18.5	18
ST-22 complex	17.8	18.5	14
ST-49 complex	10.4	4	11
ST-573 complex	16.9	16.5	10
Minor	15.7	14	45
N/A	15.1	17	43
Total	15.7	15	1446

There is some evidence for association of clonal complex with reported bloody diarrhoea on univariate analysis (Table 4.12), but not for reported vomiting, abdominal pain or fever. Age is strongly associated with reporting bloody stool, which symptom is relatively common in age groups up to age 29 and decreases gradually after this age. Logistic regression analysis with adjustment for age (using 10 year age groups) reduces the estimate for the association between clonal complex and likelihood of reporting blood and renders it non-significant ($p=0.09$). Clonal complex does not thus appear to dictate symptom complex although it is not possible to rule out moderate effects, particularly for an association with bloody diarrhoea. For bloody diarrhoea the clonal complexes with the strongest evidence for an increased tendency to cause bloody diarrhoea are ST-353 complex or ST-52 complex. ST-206 complex may be associated with lower incidence of bloody diarrhoea.

Table 4.12. Percentage of individuals reporting symptoms by clonal complex.

	Blood (p = 0.02)*		Vomiting (p = 0.25)*		Abdominal pain (p = 0.92)*		Fever (p = 0.62)*	
Clonal complex	Percent	N†	Percent	N†	Percent	N†	Percent	N†
ST-21 complex	37%	464	36%	515	94%	537	90%	533
ST-257 complex	29%	168	37%	193	92%	199	86%	192
ST-45 complex	30%	120	37%	126	94%	138	89%	132
ST-48 complex	38%	103	42%	113	93%	124	89%	120
ST-206 complex	20%	49	21%	53	96%	54	88%	51
ST-353 complex	48%	48	40%	50	90%	49	86%	49
ST-574 complex	42%	43	37%	49	94%	47	88%	48
ST-354 complex	21%	39	21%	43	95%	44	91%	43
ST-443 complex	24%	33	47%	34	97%	35	77%	35
ST-61 complex	35%	17	23%	22	91%	22	77%	22
ST-283 complex	37%	19	37%	19	86%	21	80%	20
ST-661 complex	42%	19	33%	21	100%	21	90%	21
ST-42 complex	37%	19	35%	20	95%	20	95%	20
ST-52 complex	67%	18	35%	20	95%	20	90%	20
ST-22 complex	25%	12	36%	14	93%	15	80%	15
ST-49 complex	20%	10	18%	11	82%	11	82%	11
ST-573 complex	13%	8	10%	10	88%	8	75%	8
Minor	30%	44	36%	45	91%	45	80%	49
N/A	33%	39	35%	40	98%	43	86%	43
Total	34%	1272	35%	1398	93%	1453	88%	1432

* Estimated p values from univariate logistic regression with major clonal complex as a categorical variable. Multivariate result reported in the text for blood in faeces.
† Reporting of symptoms was not complete so that the numbers vary in each analysis as tabulated.

4.5 Population structure of *C. jejuni* in the CSSS and published samples

Although coming with the caveat that human disease isolates are not a true single population of *C. jejuni* but rather are likely to represent differential sampling from several source populations, this large collection does allow a description of the structure of the *C. jejuni* population infecting humans in England during a three year period.

4.5.1 Clonal population structure

A clonal phylogeny estimated by the ClonalFrame algorithm is shown in Figure 4.7 with the same data presented as a radial tree in Figure 4.8. Clonal groups inferred by ClonalFrame are similar to those estimated by clonal complex designations, although some such as ST-257, ST-206 and ST-353 complexes break down in this

analysis. The tree suggests a closer relationship between ST-661 complex and the large, poorly resolved ST-21 complex than with other clonal groups, but does not otherwise particularly close or distant relationships between major clonal complexes. The minor ST-283 clonal complex appears to be part of a single clonal group with ST-45 complex, with which complex it shared similar characteristics regarding seasonality and age distribution in other analyses in this chapter. Among other minor clonal complexes, three represented by just one ST each, ST 677 (8 isolates), ST-682 (1 isolate), and ST 177 (1 isolate) appear to share a more recent ancestor with each other than with other STs in the dataset. These three clonal complexes appear to have been mainly identified from wild rather than food or human sources in the UK, with the isolates on the pubmlst database from sand on bathing beaches (ST-177), starlings (ST-682) a range of samples from outdoor water, a rabbit, a hare, a wild bird, cow faeces and one from a human blood culture, although ST-682 was common among Finnish human isolates (Annex 3).

Table 4.12. Percentage of individuals reporting symptoms by clonal complex.

	Blood (p = 0.02)*		Vomiting (p = 0.25)*		Abdominal pain (p = 0.92)*		Fever (p = 0.62)*	
Clonal complex	Percent	N†	Percent	N†	Percent	N†	Percent	N†
ST-21 complex	37%	464	36%	515	94%	537	90%	533
ST-257 complex	29%	168	37%	193	92%	199	86%	192
ST-45 complex	30%	120	37%	126	94%	138	89%	132
ST-48 complex	38%	103	42%	113	93%	124	89%	120
ST-206 complex	20%	49	21%	53	96%	54	88%	51
ST-353 complex	48%	48	40%	50	90%	49	86%	49
ST-574 complex	42%	43	37%	49	94%	47	88%	48
ST-354 complex	21%	39	21%	43	95%	44	91%	43
ST-443 complex	24%	33	47%	34	97%	35	77%	35
ST-61 complex	35%	17	23%	22	91%	22	77%	22
ST-283 complex	37%	19	37%	19	86%	21	80%	20
ST-661 complex	42%	19	33%	21	100%	21	90%	21
ST-42 complex	37%	19	35%	20	95%	20	95%	20
ST-52 complex	67%	18	35%	20	95%	20	90%	20
ST-22 complex	25%	12	36%	14	93%	15	80%	15
ST-49 complex	20%	10	18%	11	82%	11	82%	11
ST-573 complex	13%	8	10%	10	88%	8	75%	8
Minor	30%	44	36%	45	91%	45	80%	49
N/A	33%	39	35%	40	98%	43	86%	43
Total	34%	1272	35%	1398	93%	1453	88%	1432

* Estimated p values from univariate logistic regression with major clonal complex as a categorical variable. Multivariate result reported in the text for blood in faeces.
† Reporting of symptoms was not complete so that the numbers vary in each analysis as tabulated.

4.5 Population structure of *C. jejuni* in the CSSS and published samples

Although coming with the caveat that human disease isolates are not a true single population of *C. jejuni* but rather are likely to represent differential sampling from several source populations, this large collection does allow a description of the structure of the *C. jejuni* population infecting humans in England during a three year period.

4.5.1 Clonal population structure

A clonal phylogeny estimated by the CLONALFRAME algorithm is shown in Figure 4.7 with the same data presented as a radial tree in Figure 4.8. Clonal groups inferred by CLONALFRAME are similar to those estimated by clonal complex designations, although some such as ST-257, ST-206 and ST-353 complexes break down in this

analysis. The tree suggests a closer relationship between ST-661 complex and the large, poorly resolved ST-21 complex than with other clonal groups, but does not otherwise particularly close or distant relationships between major clonal complexes. The minor ST-283 clonal complex appears to be part of a single clonal group with ST-45 complex, with which complex it shared similar characteristics regarding seasonality and age distribution in other analyses in this chapter. Among other minor clonal complexes, three represented by just one ST each, ST 677 (8 isolates), ST-682 (1 isolate), and ST 177 (1 isolate) appear to share a more recent ancestor with each other than with other STs in the dataset. These three clonal complexes appear to have been mainly identified from wild rather than food or human sources in the UK, with the isolates on the pubmlst database from sand on bathing beaches (ST-177), starlings (ST-682) a range of samples from outdoor water, a rabbit, a hare, a wild bird, cow faeces and one from a human blood culture, although ST-682 was common among Finnish human isolates (Annex 3).

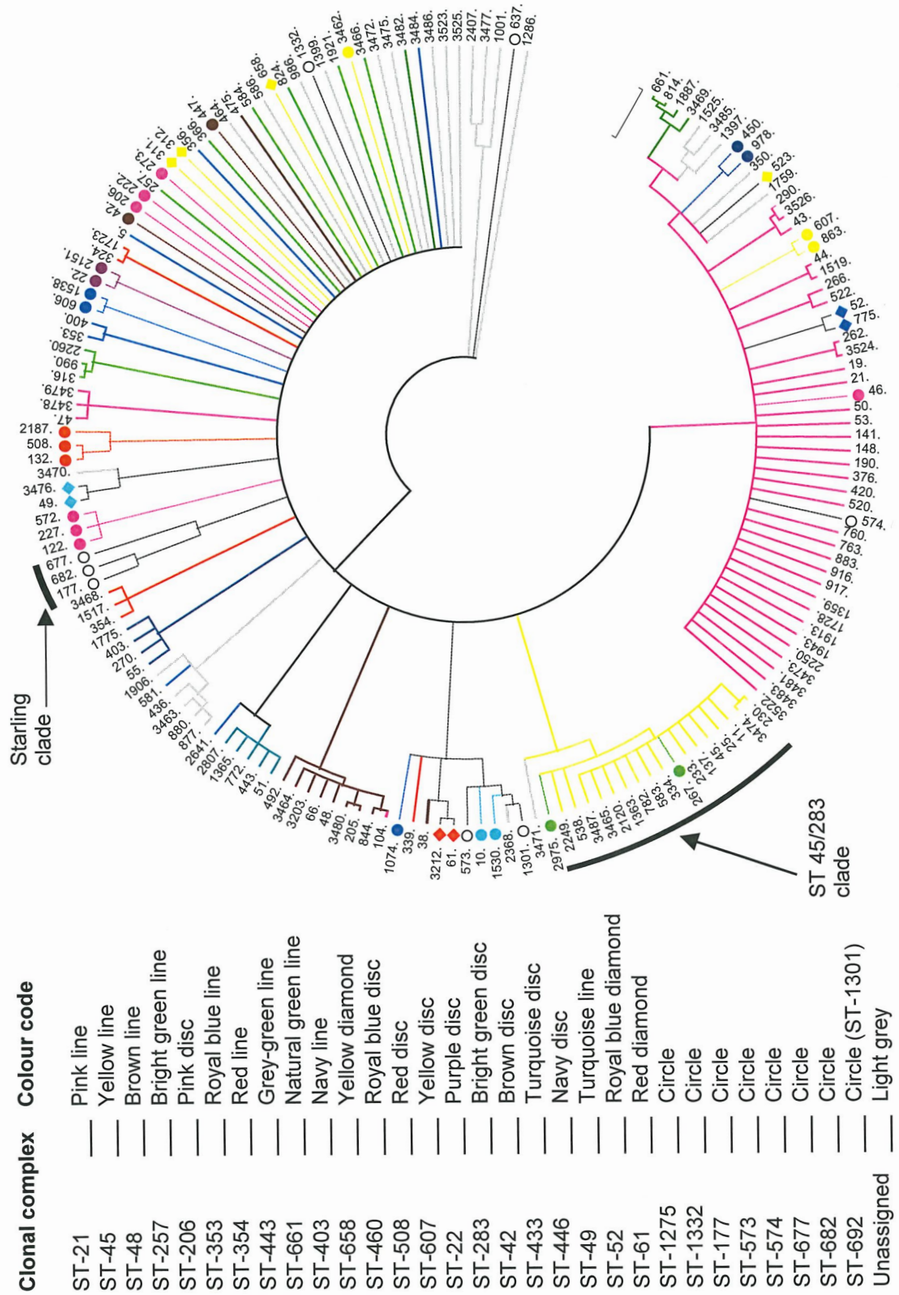


Figure 4.7. Clonal phylogeny estimated by CLONALFRAME algorithm coloured to indicate clonal complex designations and annotated with ST.

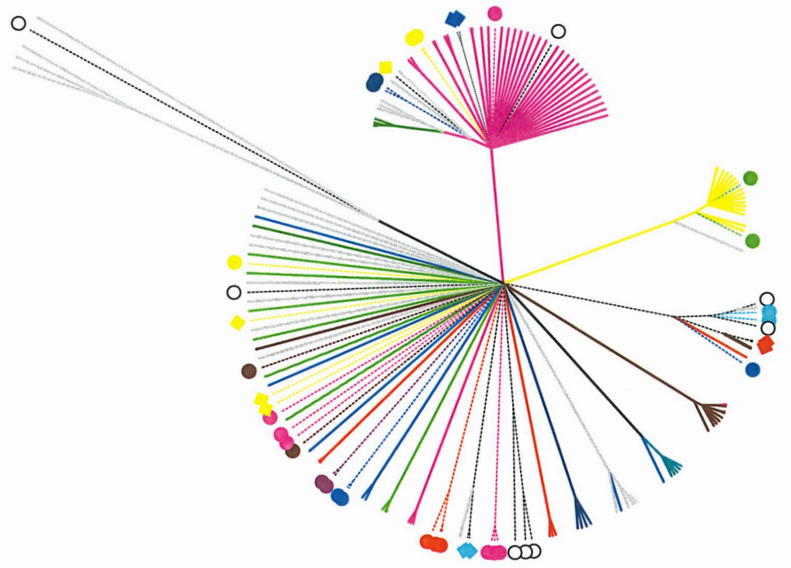


Figure 4.8. Radial tree presentation of phylogeny from CLONALFRAME algorithm coloured to indicate clonal complex designations as in the Figure 4.7 legend.

4.5.2 Population structure and geography

The variation of population structure with geography was assessed descriptively using the eBURST algorithm and the international datasets from Figure 4.2. Datasets from each of the four countries are compared in two-way comparisons with figures in Annex 6 to show which STs were present in both and which unique to one country in each case.

The large and diverse ST-21 complex was geographically widespread, with ST-21 and a single locus variant ST-50 identified in all 4 countries and ST-53 and ST-190 in all except Finland and present over the full time period of the CSSS study. Although some diversity was present across countries the dominant STs differ. In England ST-21 and ST-53 comprised 31% and 22% of the 557 isolations of this complex in the

CSSS respectively but these STs are less common in the other countries with just 4 and 6 observations in total respectively (Annex 4). In Finland 50 of the 60 ST-21 complex isolates were ST-50, which also accounted for 7 of the 15 ST-21 complex isolates in Australia, 8 of the 32 in New Zealand and 80 in the UK. In New Zealand the majority (18) of ST-21 complex isolates were ST-190, another single locus variant of ST-21.

ST-48 was the commonest ST of ST-48 complex in all countries except for New Zealand and accounted for 65% of all ST-48 complex isolates. In New Zealand ST-474, a single locus variant comprised 32 of 34 ST-48 complex isolates and 30% of all isolates in that study. It was not observed in any of the other countries.

Table 4.13. Distribution of ST-257 complex STs by country.

	Country					Divergent alleles	
ST	England	Finland	Australia	New Zealand	Total	Number	Allele
197	0	0	2	0	2	1	*aspA* 75
257	181	0	18	5	204	0	
316	1	0	0	0	1	1	*tkt* 1
366	1	0	0	0	1	1	*aspA* 4
532	0	0	2	0	2	1	*pgm* 105
584	9	0	0	0	9	3	*aspA* 8 *gltA* 5 *uncA* 1
824	3	1	0	0	4	3	*gltA* 2 *glyA* 2 *pgm* 11
990	1	0	0	0	1	1	*pgm* 133
1921	1	0	0	0	1	1	*uncA* 8
2260	2	0	0	0	2	1	*tkt* 3
3466	2	0	0	0	2	1	*pgm* 264
ST-257C	201	1	22	5	229		
Full dataset	1505	298	153	107	2063		

Two abundant clonal complexes, ST-574 and ST-257, showed limited diversity. ST-574 complex, increasing over time in the CSSS dataset and comprising 50 isolates, has no single locus variants and is absent outside England. ST-257 complex shows an intermediate pattern. Geographically widespread and comprising 14%, 13% and 5% of isolates in Australia, England and New Zealand respectively it is dominated by the

ST-257 sequence type, which accounts for 89% of ST-257 complex overall and also makes ST-257 the most common ST in the dataset as a whole as well as commonest in England and joint second in Australia. The relatively rare variants of this most abundant genotype are each restricted to just one country with the exception of ST-824 identified in both England and Finland (Table 4.13).

4.6 Discussion

This study integrated MLST information into the descriptive epidemiology of *C. jejuni* infection in a three-year sample from two English populations and compared this data with external datasets. There was no evidence for differentiation between the two geographically separate human populations in England. This is not consistent with local sources being important contributors to *C. jejuni* infections in the areas around Nottingham or Southampton, unless such local sources harbour identical genotypes. Although geographically separate both populations are similar in that each comprises a city and the surrounding areas. The lack of effect does not therefore rule out geographical differences between, for example, urban and rural populations. However since such a large proportion of the population are urban or peri-urban the public health implications of potentially shared sources across large geographical areas may be important. Moreover this suggests that sentinel sampling of human cases and possible source populations can be an efficient index of *C. jejuni* infection in England.

The larger scale geographical assessment using data from 3 other countries showed clear differences between countries but also similarities. Differentiation was not related to linear distance with the English sample being more similar to the samples

from Australia and New Zealand than to the Finnish one, which sample was from the Helsinki area. Although clear differences could be identified there was also substantial overlap for many of the major clonal groupings across the countries analysed. In some cases there was geographical patterning within the major complexes with different sub-lineages dominating in different countries.

Clear seasonal effects were observed. Some clonal groups (e.g. ST-45 and ST-283 complexes) seem to be important contributors to the peak, others (e.g. ST-21 complex and ST-48 complex) maintain a similar proportion over the year and still others contribute a greater proportion outside the seasonal peak, e.g. ST-257 complex and do not appear to rise with the seasonal peak. Other more complex seasonal patterns are also suggested e.g. for ST-354 complex and ST-52 complex which have their peaks away from the overall one. These varying seasonal patterns could be due to a range of factors including exposure of humans to different sources across the year, different dominant types of infection in these sources over the year, or seasonal variation of survival in the food chain by type. Although statistically well supported the seasonal variation is marginal in terms of absolute numbers and the main clonal groups remain the main groups across both peak and non-peak periods in the CSSS.

ST-45 complex stands out as having both a marked summer peak and strong differentiation between countries. The domination of the Finnish data by this clonal complex which peaks in summer fits with observations that the seasonal peak is much more dominant in this country than in others, whereas this complex is much less common in New Zealand which has far less seasonal variation[57]. Although designated as separate a clonal complex, ST-283 complex isolates group with ST-45

complex in the CLONALFRAME analysis to the population structure, consistent with their similar seasonal patterns. A further similarity of these isolates is their greater preponderance in young cases. This set of findings suggests some ecological or phenotypic characteristics of an ST-45 complex / ST-283 complex clade.

In contrast isolates of ST-257, ST-354 and ST-49 complexes, which have lower relative prevalence during the seasonal peak are not seen in Finland but are quite common in England, Australia and New Zealand. Similarly ST-52 complex, although present, is uncommon in Finland but not in the other countries and also tends to peak outside the overall peak period. The coherence of seasonal and geographical patterns may reflect a single underlying biological phenomenon causing temporal and geographical variation in dominant clonal complexes.

Overall the temporal trend shows substantial stability of dominant clones over the three-year study period. This relatively slow rate of secular change in the dataset is less than the between country differences and assures that the time differences between the samples in each country are not likely to explain the differences observed. Against this stability there is clearly some change in the CSSS dataset with a decrease in ST-21 complex and a rise in other complexes including the increasingly common ST-257 complex and ST-574 complex. Other less abundant complexes may also have been emerging over this time such as ST-573 complex and ST-353 complex. The wide diversity of the prevalent but decreasing ST-21 complex contrasts with the narrow clonal structure of ST-574 complex and ST-257 complex also consistent with these being emerging clones. The demographic pattern of these clonal complexes with ST-574 complex common in younger individuals and ST-257

complex in older people does not support a general model of the elderly being generally susceptible to new clones.

Apart from the age associations described above there was no evidence for substantial demographic patterning of subtypes. In particular there was no evidence for increased diversity with age. Although this contrasts with a previous report, that report assessed association with sero-types[277], which may be a more suitable measure when considering an immune-driven effect. It is possible that variation in antigenic properties within clonal complex is quite large compared to between complex differences, which would produce a null result in the current study even if sero-type were important but not tightly linked to clonal group.

There was no strong evidence for virulence differences among the major clonal groupings in this study. Differences are certainly less than described for other human pathogens such as *Neisseria meningitidis*[278] where this has been well studied. However this study would miss isolates that are completely avirulent and therefore never cause symptomatic infections. A comparison of types infecting humans with types in sources is needed to assess whether this is the case. In addition, the rare occurrence of very severe (bacteraemic) or complicated (e.g. by Guillain Barré syndrome) infection was not assessed here due to the lack of such severe or complicated cases. Published work suggests that some serotypes may be more strongly associated with Guillain-Barré syndrome[190,279] but not that the risk of this complication varies strongly with clonal group[260].

This study has thus identified areas where subtype is clearly important in the population biology of *C. jejuni*, with implications for epidemiology and other areas where clonally defined subtypes do not appear to show major differences. The findings regarding geographical and temporal patterning, which existed but did not show rapid change suggest that temporospatial effects need to be considered in source attribution but that if host association is strong then some flexibility of temporospatial matching may be possible.

5. Source attribution using typing

Identification of the relative contributions of pathogen sources and transmission routes is necessary to underpin evidence based disease control programmes[87]. As outlined in the introduction, one approach to address it, microbial source tracking, is the application of microbial typing to isolates from human cases and possible sources in the food chain to allow attribution of disease to food sources at individual case and population levels[90,91]. Evidence-based control programs based on this information have worked well with *Salmonella* at a population level in Denmark[91]. Source tracking depends on accurate estimation of the frequency of different subtypes in each host reservoir. In *Salmonella*, particular serotypes and phage subtypes are stably found in the same host[90]. The biology underlying this success is firstly that specific clones are well adapted to specific hosts and secondly that the serophage type is a stable and reliable indicator of a specific clone. For other organisms it can be much harder to find reliable host associated markers and *C. jejuni* is a case of this. Genetic methods of discrimination show enormous diversity, with studies typically reporting that more than half of the strains have unique genotypes[88,114,129,138,189,216,220,231]. Many common genotypes are broadly distributed, while for rare genotypes it is not possible to estimate the relative frequency of genotypes in different host reservoirs accurately. These difficulties have meant that although host associations have been identified for particular genotypes, no generally useable approach has been developed.

Although some host association of clonal complex with host species has been described[114] no general method to predict host from MLST data has been developed. This chapter describes an evaluation of the overall capacity of MLST data to predict host of origin. Using isolates from known sources, namely cattle, sheep and poultry from the CSP dataset (section 3.2.4), it was possible to compare predictions with the

true origin of each strain. Analysis of MLST data in the different forms of sequence type, clonal complex, and allelic profile made it possible to allow for effects of recombination as well as clonal relationship and to explore the population biology of host association underlying observed signals.

5.1 Dataset description and gene flow

There were 330 MLST sequence types among 713 isolates. Two isolates (ST-284 and ST-327) had 4 alleles typical of *C. jejuni* and 3 typical of *C. coli*. All others had 5 or more typical *C. jejuni* alleles. The diversity in isolates of chicken origin was substantially greater (219 sequence types for 309 isolates) than either cattle or sheep isolates (Table 5.1), which may represent sampling or true differences in diversity. There were no fixed nucleotide differences between the three host populations.

Table 5.1. Barriers to gene flow between host populations estimated by F_{ST} pairwise comparisons as implemented in ARELQUIN software[275].

	Isolates (frequency within sample used)			
	DNA sequence F_{ST} [p value]		**Allele data F_{ST} [p value]**	
	Chicken (n=309)	Cattle (n=245)	Chicken (n=309)	Cattle (n=245)
Cattle (n=245)	0.053 [<0.001]		0.05 [<0.001]	
Sheep (n=157)	0.055 [<0.001]	0.006 [0.07]	0.06 [<0.001]	0.001 [0.009]
	Each ST only once per host			
	DNA sequence F_{ST} [p value]		**Allele data F_{ST} [p value]**	
	Chicken (n=219)	Cattle (n=93)	Chicken (n=219)	Cattle (n=93)
Cattle (n=93)	0.029 [<0.001]		0.026 [<0.001]	
Sheep (n=57)	0.028 [<0.001]	0.024 [0.015]	0.034 [<0.001]	0.011 [0.04]

Table 5.1 reports F_{ST} values between the three host populations including the observed frequency of each haplotype (upper panel) and when counting each haplotype only once per host (lower panel) to allow for potential sampling effects.

This showed differentiation between the chicken and both ruminant populations and more marginal differentiation between the cattle and sheep populations of *C. jejuni*. Analysis with each ST used only once per host showed weaker differentiation but the same general pattern with highly significant differences between chicken and the other two host species.

5.2 Population assignment

Results are presented for initial planned population assignment using different forms of MLST data as input and then for follow up analyses based on these results.

5.2.1 Results of assignment using different data forms

Table 5.2 shows assignment accuracy, from population assignment using the STRUCTURE algorithm, when using the whole dataset and a leave-one-out strategy to assign strains to three host species; cow, sheep and chicken. These analyses were performed using the seven alleles, clonal complex, ST, and combinations thereof as indicated in the table. Since random guessing would be correct one third of the time, comparison of how much improvement genotype information made above random assignment is more informative than the percentage correct, i.e. what proportion of the gap between 33% correct expected using random assignment and 100% correct with perfect prediction has been closed. Assignment using the 7 alleles closed 37% of this gap compared to 10% for sequence type and 13% for clonal complex. Prediction did not improve substantially when ST or clonal complex information was added to allele information. These overall results emphasised the limits to using sequence type or clonal complex as a summary of multilocus sequencing typing when predicting host of origin.

Table 5.2. Capacity of alleles, overall sequence type and clonal complex information to predict host species for *C. jejuni* isolates from cattle, sheep and chickens.

Genotype information used	Percent correct	Percent of uncertainty removed*
Alleles	58	37
ST	40	10
Clonal complex (1) †	42	13
Clonal complex (2) †	42	13
Alleles plus ST	60	40
Alleles plus clonal complex†	58	37

* Random selection would be expected to predict correctly 33% of the time. The proportion of the remaining uncertainty (67%) that is resolved is given here.

† Clonal complex (1) substituted ST for clonal complex where no clonal complex is assigned and Clonal complex (2) substituted a missing value code. Clonal complex (1) was used in addition to alleles to assess "Alleles plus clonal complex".

Prediction of host of origin based on alleles is shown in more detail in Table 5.3. The method performed much better in distinguishing strains from chicken and those from cow or sheep than in distinguishing between strains from the two ruminant species.

Table 5.3. Predicted host species compared to actual species of host among *C.jejuni* from chicken, cattle and sheep.

True host	Sample size	Predicted host species		
		Chicken	Cow	Sheep
Chicken	309	**66%**	14%	19%
Cow	245	12%	**50%**	38%
Sheep	159	10%	36%	**54%**

Analysis restricted to cattle and sheep isolates achieved an assignment accuracy of 58% between these two species compared to 50% expected by chance, and thus explained only 16% of remaining uncertainty showing little detectable host association between these two closely related host species. Comparison of chicken isolates with the combined cattle and sheep populations gave improved resolution and allowed correct prediction 80% of the time (60% of uncertainty removed) indicating substantial host association.

5.2.2 Tests for confounding

Given the nature of the dataset possible confounding factors such as an association of species with time or location of sampling might have led either to completely spurious associations or to overestimates of their magnitude. Evidence for temporal and spatial effects was tested by splitting the data according to country and time of isolation (Table 5.4). In a comparison of UK chicken isolates from 1997 or earlier and 1998 or later (Table 5.4) 66% were assigned to the population of the correct period based on allelic profile. In a comparison of UK and Dutch chicken isolates, 69% were assigned to the correct country.

Table 5.4. Subpopulations for comparisons considering time and geography.

Description of source	Number
Early* UK chickens	114
Late† UK chickens	78
All UK chickens	217
Dutch chickens	53
Non UK chickens	92
Late† UK cattle and sheep	273

*1990 – 1997 = early
†1998 – 2003 = late

To test whether these temporospatial effects could have produced the apparent host association the following additional analyses tested association when host was negatively associated with time and/or space using the populations in Table 5.4. Late UK chicken isolates (1998-2003) were assigned using early UK chicken (1997 or earlier) and late UK ruminant isolates (1998-2003) as training sets, giving 77% assignment to chicken. UK chicken isolates were assigned using non-UK chicken and UK ruminant isolates as training sets, giving 64% assignment to chicken. These analyses showed that host effect was stronger than that of time or space and that the

host association observed was not simply the result of confounding due to these factors.

5.2.3 Biology of the observed host association

There were 252 ST-21 complex isolates in this study. This clonal complex comprises a substantial proportion of isolates and is highly diverse[114,120,220,280]. Of the 252 ST-21 isolates 188 were not ST-21 but differed at between 1 and 3 alleles from the central genotype. Further analysis of host association within ST-21 complex explored the mechanism underlying the better performance observed for allele based assignment, and by extension, the biological processes producing this host signature in the bacterial genome using these divergent alleles from the central genotype. The 188 isolates within the complex that were not ST-21 were assigned to chicken or ruminant host based only on the alleles at which they differed from ST-21, using all non ST-21 complex isolates as the training set. Isolates were assigned to the correct host with 66% accuracy. This analysis suggests that ST-21 complex isolates are picking up alleles that are characteristic of the host population. In order to demonstrate that this deviation from 50% is not a sampling artefact or chance effect, a repeat analysis was restricted to the 88 unique ST-host combinations, to avoid the influence of clonal expansion within host, and a permutation test was performed to assess the possible role of chance. 67% of these combinations were correctly assigned, which was a higher proportion than observed in any of 10,000 iterations in the permutation test (Figure 5.1.A).

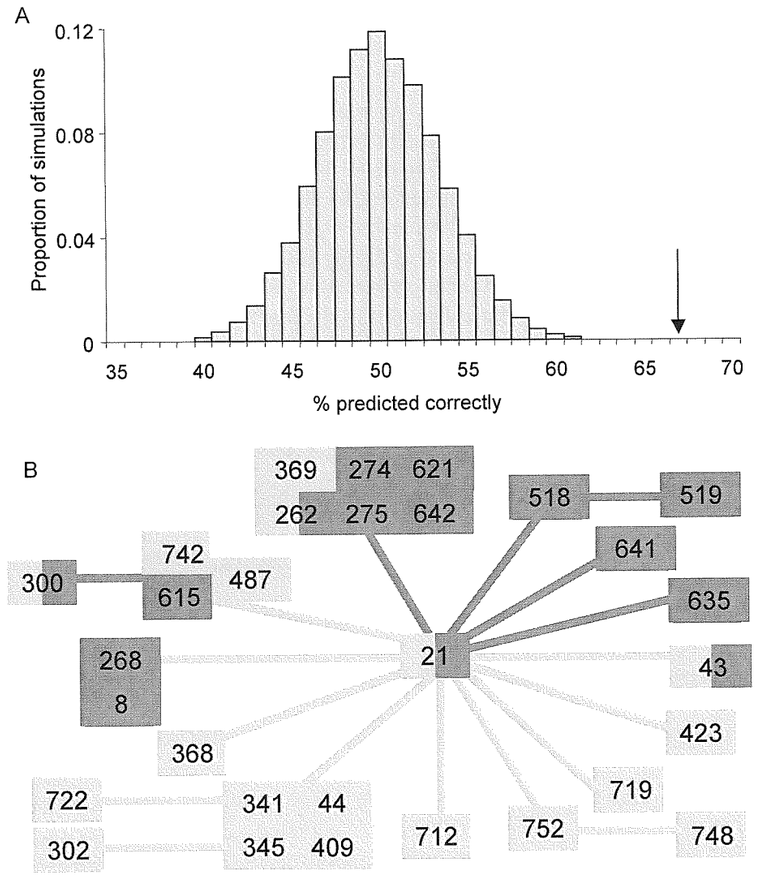


Figure 5.1. Prediction of source of origin within ST-21 complex. A. Observed accuracy of prediction compared with distribution of values obtained by permuting host labels so that the alleles varying from central genotype are not informative on host of origin. B. Prediction of origin using only alleles for which substantial reference information is available. Light grey lines indicate presence of an allele different from ST-21 present mainly in chickens in the reference population (i.e. an allele that would predict chicken origin) and dark grey those present mainly in ruminants (i.e. predicts ruminant origin). Light grey boxes are STs found only in chickens and dark grey only in ruminants. Mixed boxes indicate STs found in ruminants and chicken.

The overall accuracy of host assignment based on imported alleles is hampered by the fact that many of these alleles are individually too rare for their frequency in particular host gene pools to be estimated accurately. Imported alleles that are frequently observed in the source population should allow more reliable host prediction than those where for which there is less data to estimate population

prevalence. Figure 5.1.B shows the results of using only those alleles that are both found in at least 10 different ST-host combinations in the non ST-21 complex isolates and are also substantially differentiated between the chicken and ruminant populations (based on a 65% cut-off) are used as predictors. All 4 isolates having 2 alleles both of which are suggestive of either chicken or ruminant origin are indeed from the predicted source. In one case two potentially informative alleles gave conflicting information, one suggesting bovid origin and one chicken. Isolates with this ST came from both sources. Of the 24 STs with only 1 informative allele available, 18 are correctly assigned and only 4 incorrectly with 2 STs isolated from both chicken and ruminant sources.

Discussion

These analyses confirm a partial association of *C. jejuni* genotypes with host species and demonstrate a clear distinction between isolates obtained from chickens and those obtained from ruminants, when alleles are considered independently in statistical analysis. This finding was robust to sampling differences in time and place, suggesting that host effects were stronger than geographic and temporal effects, an important consideration if these associations are to be employed in epidemiological investigations. Moreover, populations of *C. jejuni* in farm animals such as ruminants and chickens may be rather similar compared to those from other hosts[114,231] so that the approach may be even more accurate when considering *C. jejuni* from a more diverse host range. The distinction between cow and sheep isolates is much weaker. Differentiation between these species might be demonstrable using more genetic information. However even the minor differences observed may be a sampling artifact with these species sharing a common gene pool.

The allele-based method used here goes a long way towards solving the problem of high diversity and rare subtypes in *C. jejuni*. Many alleles show differences in frequency between hosts. These alleles provide useful information on source for STs that are too rare to allow estimates of their frequency in different hosts, for example because they are entirely absent from training sets.

There are some limitations to this implementation that must be considered in any more extensive application. The current estimate of 80% accuracy in distinguishing chicken isolates from ruminant ones may be somewhat over optimistic if sampling effects are important. Sampling effects would include the nature of the sample (feces, meat etc.) as well as time and place. For example, the dominant *Campylobacter* types found on processed carcasses have been shown to differ from those found on the live poultry entering the processing plant[281]. Nonetheless, the easily identifiable sampling effects are overwhelmed by the host effect. Moreover, the analysis within ST-21 complex (Figure 5.1) is robust to both identified and unidentifiable sampling effects and suggests that such effects are not a major problem.

A further limitation of this allele-based application of STRUCTURE is that it assumes allelic independence, which is clearly violated for the dataset analysed. There are two different ways of estimating assignment accuracy. The first, used throughout this report, is a holdout procedure, whereby source of origin of strains for which the actual origin is known is predicted using the rest of the sample as a training set, providing an unbiased empirical measure of accuracy. For purposes of prediction of isolates where the source is unknown, this procedure is not possible and the second approach to

estimating accuracy would need to be employed, using the estimates of accuracy that the algorithm itself provides. Because STRUCTURE assumes that each allele is independent, it is overconfident regarding the accuracy with which it estimates the origin of multilocus genotypes where alleles are not independent. For example, in differentiating chicken isolates from those originating in cattle and sheep, STRUCTURE predicts 91% accuracy for itself, but empirically it only achieves 80% on average. A better estimate of uncertainty would be necessary for predictive purposes if applied to isolates of unknown origin. More sophisticated genetic models that reflect the dependence amongst the loci should achieve more accurate assignment as well as better estimates of statistical uncertainty.

Accepting these limitations this approach nonetheless demonstrates the ability to assign isolates probabilistically to populations. When broad reference populations from the full range of possible sources are available it will be possible to apportion groups of isolates, such as those affecting a human population over a period of time, to their sources, although precision in the attribution of *C. jejuni* may be less than with other pathogens such as *Salmonella* where host species appear to harbour more differentiated populations[90]. Any application to human isolates before such comprehensive reference databases are in place comes with strong caveats. The accuracy in the current setting was in attributing isolates that came from chicken or ruminants back to the correct source. Isolates from humans may come from a wide range of sources. If the correct source is not represented in the reference datasets such isolates would of necessity be attributed to those putative sources that are present. Prediction is possible with individual isolates, sometimes to a single source, although in some cases prediction will suggest a range of populations rather than one. For

example, two of the fully sequenced *C. jejuni* genomes are from known sources, one being isolated from poultry (isolate RM-1221)[118] and the other from a human involved in a cow's milk outbreak (isolate 81-176)[120,280]. Assigning these based on the reference datasets used in this study predicted their origin as chicken (99% probability) and cattle/sheep (97% probability) respectively.

The broad host range of *C. jejuni* spanning a range of mammalian, avian and other species make it a good model to study features that may be informative of the ecology of multi-host pathogens. *C. jejuni* imports fragments from other members of the species, which have been estimated to be typically a few hundred base pairs in length[259]. The analysis within ST-21 complex demonstrates that isolates in this complex have imported genetic material prevalent in the population of *Campylobacter* carried by their host species (Figure 5.1). This observation implies firstly that there is persistent differentiation in allele frequencies between different host species and secondly that many of the ST-21 isolates represent lineages that have persisted within the same host species long enough to import a substantial number of alleles. The current study surveyed 7 loci and found on average 0.32 host-specific alleles in the 81 STs other than ST-21 that were members of the ST-21 complex, i.e. approximately 5% of the alleles in this analysis. The acquired genes were widely distributed between these types and across loci, with 6 of the 7 loci having one or more imports. Since the MLST loci themselves were chosen because they represent core metabolic functions of *C. jejuni*[220] and are not obvious candidates for host adaptation we are unlikely to be observing a rate above that produced by the neutral level of genetic import. Extrapolating linearly from the seven loci in this work to the 1654 gene coding sequences in the *C. jejuni* genome[119] gives an estimate of 86 genes with alleles typical

of a particular host species within each ST-21 complex ST. This is obviously a rough estimate since it is based on fairly limited data and because recombination and selection at other genes may behave differently. However this approximate estimate demonstrates the potential for substantial adaptation to the most recent host by homologous recombination. Indeed, homologous recombination may be an important factor in allowing a single bacterial species to stably colonize a wide range of host species while adapting to some extent to each.

The weak overall host association of clonal complexes and the greater power of attribution using the signal generated by recombination does not mean that host-associated, and host-adapted, sub lineages do not occur. A more detailed population structure, probably requiring substantially more genetic information than afforded by the 7 MLST gene fragment sequences for resolution, may well identify host-associated lineages within clonal complexes in addition to the known host species-clonal complex associations.

In conclusion, a population genetic approach using multi-locus data has allowed host assignment in *C. jejuni* where single markers to identify host associated lineages are unavailable but host species populations are differentiated by allele frequency at a range of loci. Homologous recombination generates a host signature in the *C. jejuni* genome and analyses taking advantage of this signal have improved accuracy of host prediction. The inherent standardization and portability of sequence typing in combination with the availability of such improved assignment techniques support the application of this approach to clarify aspects of *C. jejuni* epidemiology on a global scale and application to the study of other suitable microbes. Further developments of assignment approaches should allow more efficient use of currently available data.

Host association appears stronger than temporal and geographical effects. However this study was severely constrained by the nature of the sample investigated. Structured sampling of *C. jejuni* from defined hosts, places and times, and where frequency of different types is representative of the population rather than an artefact of sampling, is needed. This would allow better understanding of the relative importance of these factors on population biology and underpin more reliable epidemiological application.

6. *Campylobacter jejuni* contamination of retail poultry

The potential for host attribution using multi-locus sequence data was demonstrated in the previous chapter and the importance of chicken as a source reviewed in the introduction. Attribution using subtype requires samples from possible source populations. The type of sample required depends on the population biology of *C. jejuni* in host species. This chapter describes a large sample of *C. jejuni* from retail poultry and explores variation in type in relation to factors including time of sampling and country of origin. This is to assess the need to match temporal, spatial, and other factors for an attribution reference dataset. It also describes the *C. jejuni* population in poultry, with some caveats regarding how well contamination of poultry meat reflects the infecting population during life. Discriminatory genotyping using 9-locus genotype, with the addition of the *flaA* and *porA* antigen genes was also undertaken and analysed in relation to sample characteristics to explore the scope for more detailed source identification by subtyping, such as during outbreak investigations.

6.1 Description of population sampled

6.1.1 Sample characteristics

This study typed 742 isolates with completed MLST, 165 from the 2001 UK retail poultry survey, 408 from CLASSP (in England and Scotland) and 169 from the 2004/5 period of the Wales and Northern Ireland survey. The distribution of the typed isolates by time and place is given in Tables 6.1 and 6.2.

Table 6.1. Distribution of retail poultry sample by year and month

Month	2001	2004	2005	Total
January	0	12	40	52
February	0	13	54	67
March	0	11	52	63
April	59	8	23	90
May	105	12	46	163
June	1	5	44	50
July	0	8	43	51
August	0	12	30	42
September	0	9	30	39
October	0	11	27	38
November	0	46	16	62
December	0	25	0	25
Total	165	172	405	742

These studies, their sampling approach, and the selection of isolates from them for this study are described in the Methods chapter, section 3.2.2.

Table 6.2. Distribution of retail poultry sample by year and country of sale.

Year	2001	2004	2005	Total
England	78	55	294	427
Northern Ireland	30	39	18	87
Scotland	30	1	23	54
Wales	27	76	36	139
Not recorded	0	1	34	35
Total	165	172	405	742

6.1.2 Sequence types and population structure

Figure 6.1 compares the population structure inferred by CLONALFRAME with the clonal complex definition of genetic group for *C. jejuni*. Sequence types are clustered into similar but not identical groups. Clonal complex groups are used in this chapter since they are widely used and broadly similar to those from the phylogeny produced by CLONALFRAME. ST-267, comprising 10 of the 11 ST-283 complex appears closely related to ST-45 complex as identified in CLONALFRAME analysis of human isolates.

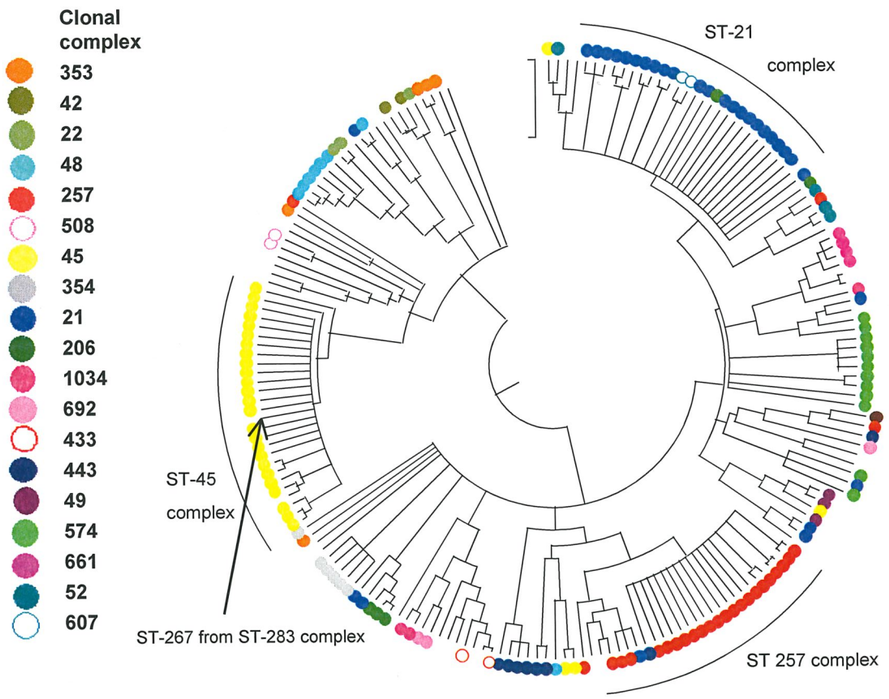


Figure 6.1. Comparison of clonal complexes to population structure from a 75% consensus CLONALFRAME tree. Full clonal complex and ST data is given in Annex 7.

6.2 Relationship between subtype and sample characteristics

6.2.1 Association of time of sample and subtype

The separation of seasonal and secular trends is complicated by the differential sampling in different years, in particular the restriction of the 2001 survey to between April and June. In line with this, the data is divided into April – June 2001 (n=165), April – June 2004/5 (n=138) and other months 2004/5 (n=439) for graphical assessment of association of subtype with time (Figure 6.2). The clonal complex distribution of isolates sampled at these different times was broadly similar, with for example the same three being most abundant, albeit in a different order with ST-21 complex most prevalent in the 2001 survey and ST-45 complex in the 2004/5 studies. Among other clonal complexes ST-353 complex was more prevalent in the 2001 and ST-574, ST-443, and ST-573 complexes in the later surveys. The data thus shows evidence for variation in relative importance of different clonal complexes over a time scale of years (χ^2 $p < 0.001$). There is no evidence for substantially different types in April-June and other times although this ignores other possible seasonal effects.

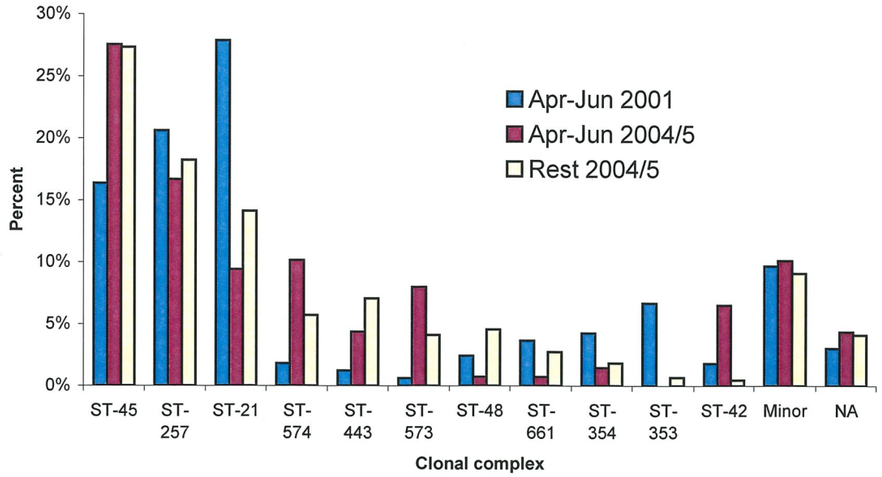


Figure 6.2. Comparison of percent of *C. jejuni* isolates in major clonal complexes for, April-June 2001 (n=165), April-June 2004/2005 (n=138), rest of 2004/2005 (n= 439).
*Minor clonal complexes include 11 or fewer isolates. NA - not assigned to a clonal complex.

To allow more flexible analysis of temporal effects including season a logistic regression model was used to predict membership of individual clonal complexes (for those with n≥12 in the study) by the explanatory factors of season and year. Results are summarised in Table 6.3. There is strong evidence for a decline in ST-21 complex in the later surveys compared to 2001 and a rise in ST-45 and ST-574. There is also evidence for seasonal variation. A summer peak was supported for ST-45 complex while ST-443 complex appeared to peak more in winter (Table 6.3 and Figure 6.3).

Table 6.3. Seasonal and secular changes in prevalent *C. jejuni* clonal complexes on retail poultry from a logistic regression model (n=742).

Clonal complex	n	Year effect* Highest	p	Seasonal effect† Peak months	p
45	185	2004/5	0.003	June/July	0.005
257	137	None	0.44	Jan/Feb	0.07
21	121	2001	<0.001	None	0.83
574	42	2004/5	0.02	March	0.13
443	39	2005	0.02	October	0.005
573	30	2005	<0.001	April	0.10
Infrequent (n< 12)	99	None	0.91	October / November	0.95

* For "Year effect" the p values for year are from likelihood ratio tests for models with and without year as a categorical variable and thus tests for variation associated with year without assuming a monotonic form of relationship.
† The p values for seasonality are from likelihood ratio tests for models with and without sine and cosine terms for time. The description "Peak months" is from the graphic display of fitted values for these coefficients with examples given in Figure 6.3.

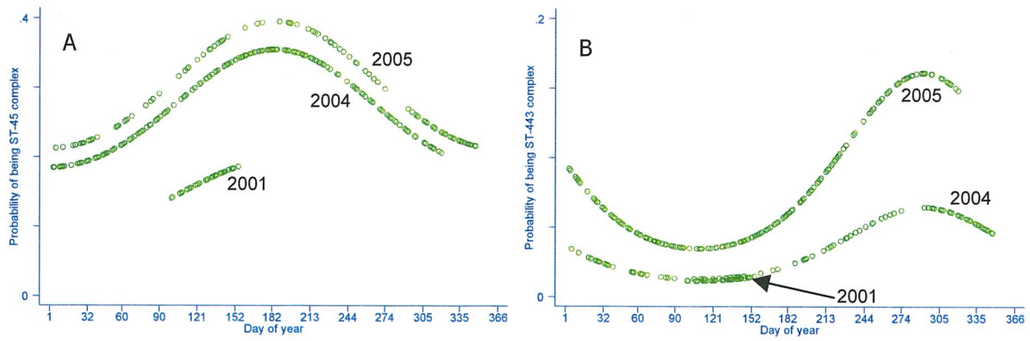


Figure 6.3. Fitted values from logistic regression model predicting the probability of isolated belonging to ST-45 clonal complex (panel A) or ST-443 clonal complex (panel B) dependent on year and day of year of sample.

Statistical support for other seasonally patterned variation was weaker. Although ST-283 complex is not common in this dataset (n=11) it was tested in a similar way in the light of the CSSS results. Isolates were confined to the March – Sept period. There was no strong statistical evidence for seasonality, but given only 11 isolates there was little power to confirm seasonality statistically and only an extreme effect could have produced a significant result. There was no statistical support for interaction between year and seasonal effect for ST-45 complex (p=0.16 from a likelihood ratio test for interaction between cyclical effect and year). The seasonal effect in ST-443 complex is less stable with year (p < 0.001 for interaction) in this dataset. Figure 6.4 shows the proportion of isolates each month from each of these two clonal complexes. This graph shows that there may be more complicated seasonal patterns for ST-45 complex than modelled in the statistical analysis, with a small early peak followed by a larger late summer peak.

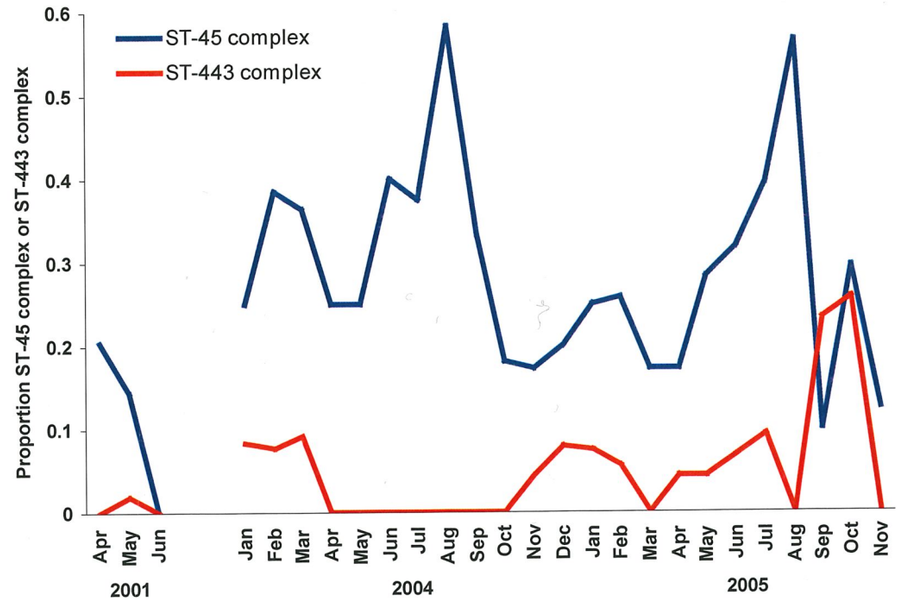


Figure 6.4. Proportion of *C. jejuni* isolates from retail poultry that belong to ST-45 complex and ST-443 complex by month.

6.2.2 Geography and subtype

6.2.2.1 Country of sale

Although broadly similar between countries, subtype of *C. jejuni* does vary with country of sampling (Figure 6.5, χ^2 test p = 0.003 for a global test of association). Clear differences on inspection are the relatively low number of ST-21 complex isolates in the English sample, low ST-257 complex in both the English and Northern Irish samples and the apparently high level of ST-574 complex in Scotland.

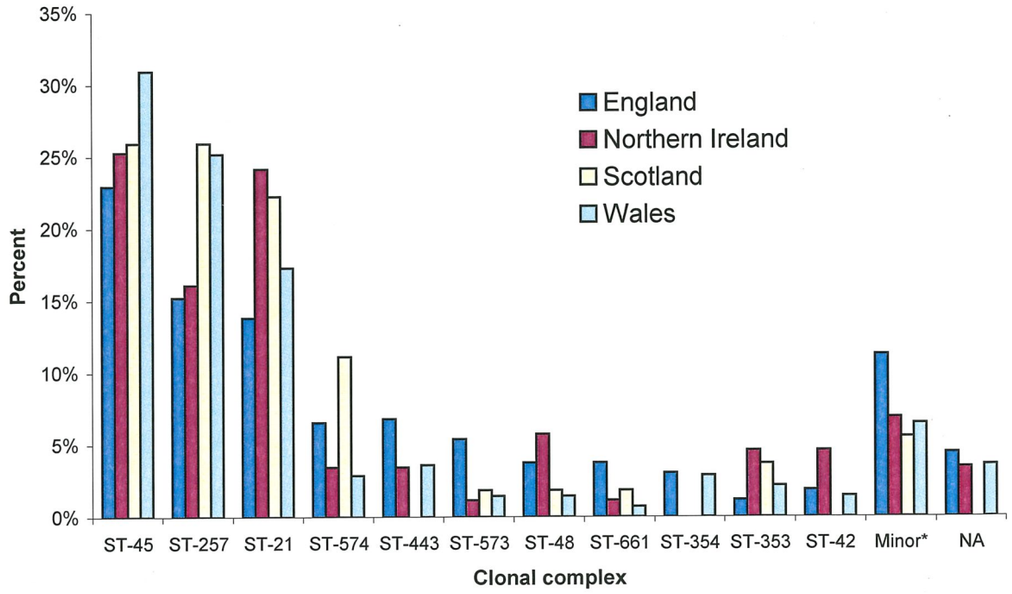


Figure 6.5. Comparison of percent of *C. jejuni* isolates in each major clonal complex by country of sale (n=707 total, England =427, N. Ire. = 87, Scotland = 54 and Wales = 139). *Minor clonal complexes = 11 or fewer isolates and NA = not assigned to a clonal complex.

The ST-21 relationship with country decreased when adjusted for year and day of sampling using the same model as used to assess seasonal and secular trends (p = 0.29 association between probability of being ST-21 complex and country of sale) and may have been due to differential sampling across time in each country. Adjustment for sampling time had only a minor effect on the association between country and ST-257 (p=0.02 for adjusted association between type and country) while for ST-574 this

adjustment enhances the positive association with Scotland and negative association with Wales and Northern Ireland (p=0.01 for adjusted association between type and country).

6.2.2.2 Country of origin

Most sampled poultry was of UK origin limiting the scope to test variation in this. Non-UK produced poultry was mainly from other countries within the European Union. There was no overall evidence for different clonal complex distribution between samples produced in the UK or elsewhere (Figure 6.6 and (χ^2 p value = 0.6). The apparent lack of some low frequency clonal complexes in the non-UK data may have been due to sample size of just 41 for non-UK origin *C. jejuni* isolates. A higher level of ST-45 complex in non-UK chicken was not significant and was further reduced when adjusted for whether the sample was from fresh or frozen poultry, with more non-UK poultry being frozen and frozen poultry tending to have a higher proportion of ST-45 complex isolates as described in the next section.

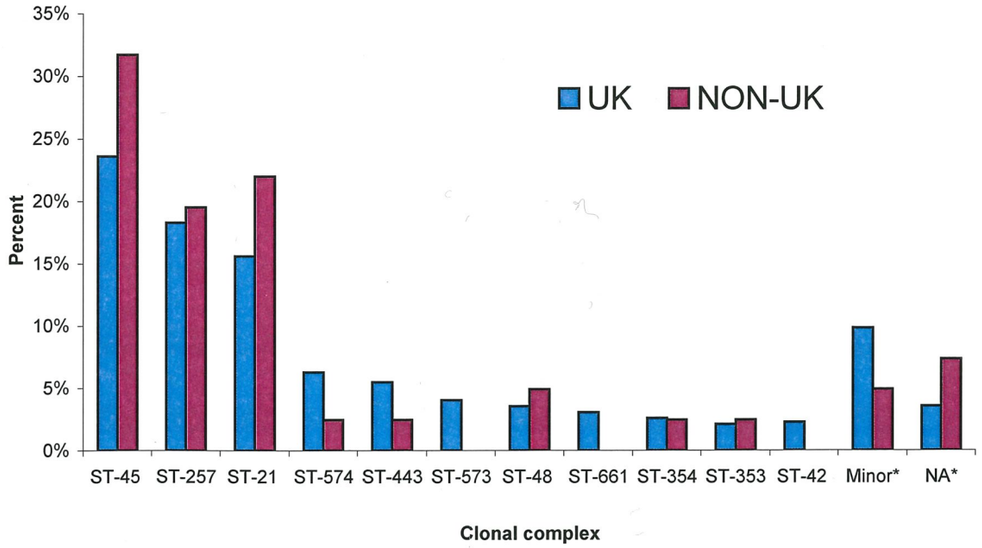


Figure 6.6. Distribution of 664 *C. jejuni* isolates from retail chicken by clonal complex and whether produced in the UK (n=623) or abroad (n=41).
*Minor clonal complexes = 11 or fewer isolates and NA = not assigned to a clonal complex.

6.2.3 Production and handling of chicken and subtype

6.2.3.1 Production method

Table 6.4 shows the percentage of each clonal complex that originated from a free range or organic rather than broiler production. There was no evidence for unusual types (minor clonal complexes or unassigned) being more common in extensive than broiler production or indeed of a systematic difference in the proportions of different clonal complexes by production method (χ^2 p value = 0.16). However, there are too few *C. jejuni* isolates from free range or organic flocks to have reasonable power to assess differences in clonal complex distribution by production type unless such differences were very large.

Table 6.4. Percentage of isolates from extensive production by clonal complex.

Clonal complex	Broiler	Free-range / organic	Percent free-range / organic
ST-45 complex	145	9	6%
ST-257 complex	116	1	1%
ST-21 complex	102	3	3%
ST-574 complex	35	3	8%
ST-443 complex	30	3	9%
ST-573 complex	25	1	4%
ST-48 complex	21	3	13%
ST-661 complex	17	1	6%
ST-354 complex	16	0	0%
ST-353 complex	10	1	9%
ST-42 complex	11	0	0%
Minor	61	1	2%
Unassigned	26	0	0%
Total	615	26	4%

6.2.3.2 Fresh or frozen at point of sale

C. jejuni isolates from fresh (n=583) and frozen (n=108) chickens showed similar clonal complex distributions (Figure 6.7). Overall the variation is consistent with that expected by chance alone (χ^2 p value = 0.2). Within the overall lack of variation the greater dominance of the ST-45 complex does appear marked, and if compared to all other isolates gives an odds ratio of 2 (Fisher exact test p = 0.003). Although this is

post-hoc and multiple comparisons must therefore be considered this may represent a real effect. In analysis of the 659 isolates with complete data for country of origin the association between ST-45 complex and isolation from frozen chicken (Odds ratio 1.98 [95% CI 1.27-3.09]) is robust to adjustment for country of origin (1.93 [1.23-3.03]) and year and season of isolation (1.79 [1.13-2.85]).

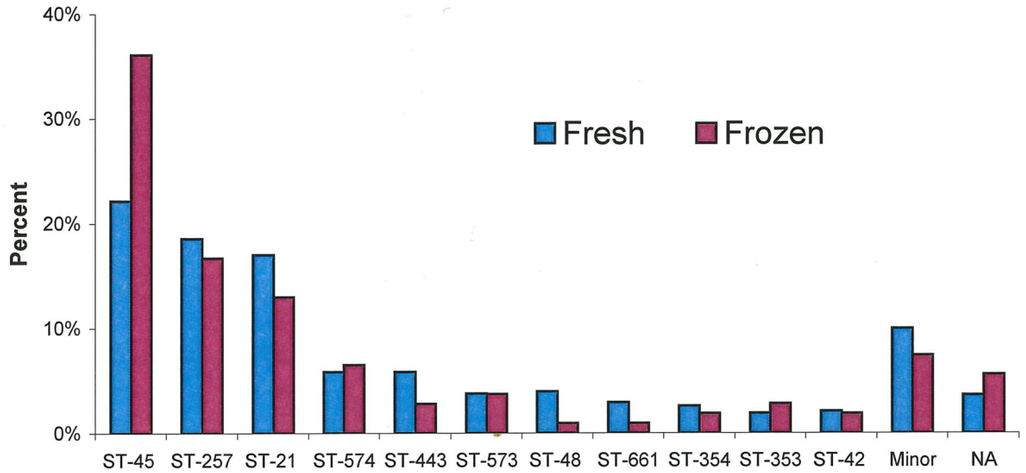


Figure 6.7. Clonal complex distribution of 664 *C. jejuni* isolates from retail chicken by clonal complex and whether fresh (n=583) or frozen (n=108) at the point of sale.
*Minor clonal complexes are those with 11 or fewer isolates. NA = not assigned to a clonal complex.

6.2.3.3 Type of retailer

There was no evidence for type of retailer (supermarket (n=495), butcher (n=144), other (n=53) being associated with subtype (χ^2 test p value = 0.6).

6.3 Joint analysis with published data from other animal hosts

Retail poultry isolates show a more similar clonal complex distribution to other isolates from chicken than to those from ruminant sources using the CSP dataset as an external reference population (Figure 6.8). For example, ST-45 and ST-257

complexes are more typical of chicken and ST-61 more typical of ruminants, but with considerable overlap between host species. Assignment to host using allele frequency and the STRUCTURE algorithm as described for the analysis of the CSP dataset, assigns the retail poultry 68% to chicken and 32% to ruminant populations.

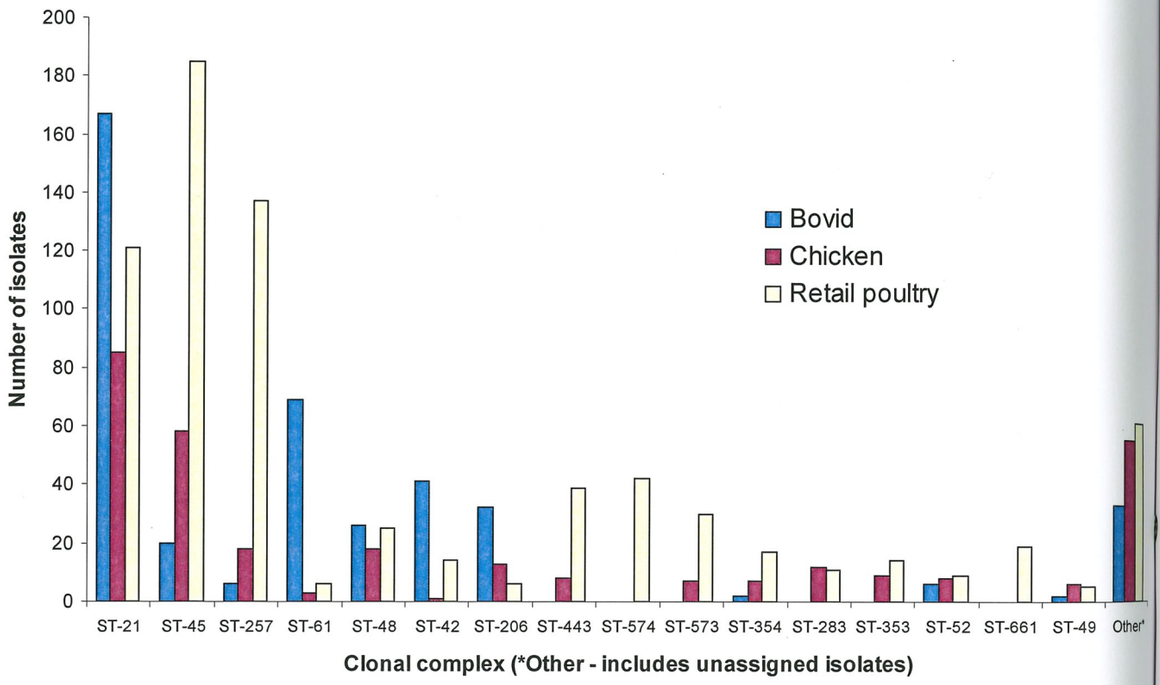


Figure 6.8. Comparison of clonal complex distribution of 742 retail poultry isolates of *C. jejuni* with chicken (n=309) and bovid (n=404) isolates from the CSP dataset.

Wider comparisons are shown in Table 6.5. Firstly isolates from the RPS of UK origin (n=623) and collated published isolates from cattle sampled in the UK from 1997 to 2001 were used as a reference dataset to assign isolates from these same host animals in other countries which were identified from published isolate collections from Finland (chicken and cattle) and Senegal (chicken). These datasets are described in section 3.2.6 and the original publications referenced in footnotes to Table 6.5. Isolates are assigned correctly with between 71% and 83% accuracy (Table 6.5)

compared to 50% expected by chance. The process was reversed using the small Finnish dataset[282] of 20 cattle and 33 chicken isolates as the reference population. This produced very erratic results for assignment.

Table 6.5. Assignment of international isolates from published data and the retail poultry survey.

Reference populations	Assigned population	Percent correct
UK cattle & UK poultry	Senegalese poultry	71
UK cattle & UK poultry	Finnish poultry	78
UK cattle & UK poultry	Finnish cattle	83
Finnish cattle & Finnish poultry	Senegalese poultry	9
Finnish cattle & Finnish poultry	UK poultry	30
Finnish cattle & Finnish poultry	UK cattle	76

UK poultry - 623 isolates of UK origin from the UK retail poultry surveys 2001 and 2004-5.
UK cattle – 153 cattle isolates from UK cattle 1997-2001 from three published studies (n=42[114], n=29[231] and n=82[138]).
Finnish cattle (n= 20) and poultry (n=33) all from a 2003 study[282].
Senegalese poultry from a single study 2000-2002[240].

6.4 Antigen gene results

Highly discriminatory methods are used in outbreak detection and the identification of specific sources of infection. As a basis for evaluating this approach using variable gene sequencing in *C. jejuni* fragments of two highly variable genes, *flaA* and *porA*, were sequenced in this project to allow evaluation of the degree of discrimination that this would afford for isolates from a chicken source. Further work, on the theoretical and empirical assessment of the role of highly discriminatory typing in outbreak detection and source tracing, is presented in the next two chapters

6.4.1 Discriminatory index and epidemiology using 9 locus genotype

The discriminatory index of a typing scheme measures the proportion of non-identical isolates that the scheme identifies as having a different type (it is assumed that isolates not sharing a very recent common ancestor, and by extension ecological origin, have some genetic differentiation which would allow discrimination between them). The MLST genotype offers 96.0% discrimination among the *C. jejuni* isolates in this study

(n=741 as genotyping of variable genes incomplete) with the combined set of 7 MLST genes producing marginally more discrimination than either antigen gene alone but less than the combination of two antigen genes. Use of all nine genes raises the DI from to 99%, removing three quarters of the gap in discrimination of MLST alone (Table 6.6). These estimates are based on the assumption that all isolates did not share a common origin such that discrimination should be both possible between all isolates. If some isolates did share a common origin the estimated indices will have been conservative and the true DI values higher.

Table 6.6. Discriminatory index available from MLST, *flaA* and *porA* genes and combinations of these genes among 741 *C. jejuni* isolates from retail poultry surveys.

Loci	Discriminatory index*
MLST	0.961
flaA	0.937
porA	0.957
MLST & *flaA*	0.981
MLST & *porA*	0.986
flaA & *porA*	0.984
MLST & *flaA porA*	0.990

* Assuming no isolates share a recent common origin.

Using the 9-locus genotype 301 (41%) of the isolates were unique in this dataset. There are nonetheless clusters of genotypically identical isolates with 9 clusters of size 9 or larger (Table 6.7). Isolates sharing a 9-locus genotype did not all share an identical origin. Table 6.8 summarises the diversity of origin of the 9 clusters of genotypically identical isolates, which comprise 9 or more isolates. Within each cluster there is evidence for production in several farms showing that identity across these 9 gene fragments is not, on it's own, strong evidence for a common proximate origin. Although all 58 isolates in the largest cluster (ST 257, *flaA* allele 16, *porA* allele 1) appear to be from standard broiler production they originate from several producers and across the years 2001 (n=19), 2004 (n=16), and 2005 (n=23). Two of

the isolates in 2001 are of non-UK origin while all others with known origin are from the UK, although isolates in 2005 were of both Scottish and English origin.

Table 6.7. Clusters of genotypically identical isolates in *C. jejuni*. (n=741*)

Isolates in cluster	Sequence type	*flaA*	MOMP
58	257	16	1
17	51	21	11
14	45	15	49
12	25	22	180
12	583	239	73
10	45	21	49
10	267	239	73
9	45	21	44
9	257	16	137
8	334	239	73
8	607	14	168
7	21	8	7
7	45	8	44
7	48	32	14
7	50	36	6
7	104	36	14
7	262	37	8
7	573	314	122
6	5	14	15
6	51	21	10
6	573	32	171
6	574	8	297
5	17	16	1
5	51	34	10
5	61	42	36
5	354	18	57
5	573	67	4
5	2030	16	1
4	19	36	7
4	45	5	44
4	45	70	53
4	53	32	27
4	53	52	27
4	354	18	58
4	447	189	27
4	661	36	2
3	22 clusters of 3 identical isolates		
2	36 pairs of identical isolates		
1	301 unique isolates		
741	Isolates in analysis		

* In one *C. jejuni* isolate *flaA* and MOMP typing was not complete and it is not possible to identify whether it is unique or part of a cluster.

Table 6.8. Diversity of sample characteristics within genotypically identical clusters.

ST, *flaA*, *porA*	n	Origin	2001	2004	2005	Healthmarks*	Production†
257, 16, 1	58	UK/non UK	+	+	+	13	B
51, 21, 11	17	UK/non UK			+	7	B,F
45, 15, 49	14	UK	+	+	+	9	B,F
25, 22, 180	12	UK/non UK	+	+	+	5	B
583, 239, 73	12	UK/non UK		+	+	7	B,F
45, 21, 49	10	UK		+	+	4	B
267, 239, 73	10	UK		+	+	3	B, C
45, 21, 44	9	UK	+	+	+	5	B
257, 16, 137	9	UK		+	+	5	B

* Number of different health marks associated this type. Healthmarks are linked to production sites. Recording not always complete and reliable. Numbers given here are conservative not counting any unclear marks.
† Production - B (standard broiler production), F (free range) and C (corn fed).

6.4.2 Recent changes in each gene as a contribution to discrimination

The greater extant diversity in antigen genes should be mirrored in more rapid rates of recent change. This is assessed below considering which alleles differed between isolates (or genotypes) which shared 8 or 7 identical alleles across the nine loci (MLST, *flaA* and *porA*). The analysis uses 738 isolates (392 different nine locus genotypes) with typing data for all nine loci (*porA* sequence was not completed for four isolates). The method is described in the methods chapter (section 3.6.4). The presence of identical alleles at seven or eight loci is evidence for a recent ancestor with the variant alleles therefore likely to represent recent change in one of the isolates since that time. The MLST loci are rarely the locus at which the most recent change has occurred. When analysed using all isolates the MLST genes together contributed 28% of the most recent changes (i.e. single allele differences) with *uncA* contributing most at 8% compared to 21% for *flaA* and 51% for *porA*. Analysis using each genotype only once produced the same pattern but to an even more marked extent with over 88% of single allele differences involving either *flaA* (27.5%) or *porA* (61%) and less than 12% an MLST locus (Table 6.9). Analysis considering

differences at 2 loci gave a similar dominance of *flaA* and *porA* over MLST genes as the site of recent change, although the differentiation between *flaA* and *porA* was less marked as would be expected when two changes were present in each comparison.

Table 6.9. Distribution of recent evolution in isolates across loci in comparisons of isolates and of genotypes at 9 loci (738 isolates, 392 genotypes).

Locus	Proportion of single or two allele differences per locus			
	Isolate comparisons		Genotype comparisons	
	1 allele differs*	2 alleles differ*†	1 allele differs‡	2 alleles differ‡†
aspA	0.055	0.033	0.027	0.040
glnA	0.017	0.016	0.011	0.024
gltA	0.036	0.032	0.004	0.030
glyA	0.061	0.050	0.016	0.032
pgm	0.001	0.013	0.002	0.012
tkt	0.030	0.021	0.031	0.020
uncA	0.080	0.069	0.024	0.057
flaA	0.214	0.346	0.275	0.355
porA	0.506	0.420	0.610	0.429

* n = 5008 and 7757 isolate pairs where isolates differ from each other by 1 and 2 alleles respectively
† As two loci are involved in each of these events results are divided by 2 to facilitate comparison between columns.
‡ Analysis of genotypes to avoid inflation of the importance of isolates which differ from very common 9 locus genotypes by one or two alleles (451 one allele and 1849 two allele difference pairs of genotypes were identified).

6.5 Discussion

This study described the population of *C. jejuni* contaminating UK retail poultry meat. There was substantial stability of subtype distribution over the course of several years but with some clear change in relative prevalences. The ST-573 and ST-574 clonal complexes appeared to emerge over the period studied (2001, 2004, and 2005) from initially low levels while ST-21 complex decreased as a proportion of the total. These findings mirrored those described from humans in the CSSS between 2000 and 2003. The lack of change in relative prevalence for ST-257 complex and the rise in ST-45 complex differed from the CSSS finding. Some clonal complexes appeared to show distinct seasonal patterns while for others there is no clear evidence of seasonal variation, with ST-45 complex again standing out, as it did in the CSSS human isolate dataset. The basis for seasonal variation in ST-45 complex in retail poultry meat is not

clear. It could reflect variation in infection rates during production or differential survival post slaughter. Analysis of seasonal pattern at different points in the poultry production chain would allow clarification of how infection rates or differential survival contribute to these findings. This clonal complex has been isolated from quite a broad host range in past studies including chicken, ruminants, wild birds (especially starlings) and sand in two or more studies and from turkey, pet animals, a badger, a rabbit, water and soil each in one study (Table 1.5) but without any study of temporal variation in these niches. The combination of seasonal variation at point of sale, the greater survival of ST-45 complex isolates on frozen poultry and isolation from relatively diverse niches suggests that ST-45 complex may have a distinct phenotype. The clonal relationship between ST-283 complex and ST-45 complex was again supported by estimation of a clonal phylogeny within this dataset. Although formal analysis does not give statistical support for the small number of ST-283 complex isolates having a seasonal pattern they were all isolated across the spring to autumn period. The true clonal group with any such distinct phenotype may therefore include isolates classified as ST-283 complex. The seasonality of ST-45 in retail poultry meat indicates that seasonal variation in human types may not always be due to seasonally varying sources of infection but could involve changing subtype exposure from the same source or sources.

The weakness of geographical effect for country of origin in the RPS data set, and the apparently strong host association when assigning Finnish and even Senegalese isolates using UK reference populations supports the hypothesis that there is strong host association overriding substantial geographical distance. The assignment results using the Finnish study as a reference population did not show reliable host

association. This study had just 20 cattle and 33 poultry isolates, which was likely to have provided an inadequate estimation of gene frequencies, and so these discordant results are not good evidence to question the host association supported by other analyses in this and the previous chapter. Although overcome by host association geographical effects were not absent. In addition, the association of country of sampling with subtype showed some population structure at this level. This is consistent with the supply of poultry to retailers not being homogenous across the UK with, for example, different producers dominating in different areas. Healthmarks, which indicate producer, were not completely or consistently recorded and this association not formally assessed.

The limited temporal and spatial patterning described in this study means that tight geographical and temporal matching of farm animal reference data to human isolates may not be necessary and that use of larger external datasets should be considered where sufficient local and contemporaneous data are not available. Evaluation of geographical structuring in other hosts is important in this regard. If sufficient host data becomes available it may be possible to use both local and wider data, with appropriately estimated weights accounting for temporal and geographical distance, to allow accurate attribution of human infection to source.

This study demonstrated, as expected, that highly diverse genes encoding surface exposed proteins undergo more frequent change than MLST loci and may be efficient markers to distinguish sub-lineages within sequence types. The 9-locus genotype applied in this work was highly discriminatory by usual standards[204,207]. However substantial stability over time and a widespread geographical origin of identical 9-

locus types show that this discrimination is not sufficient to reliably implicate a particular source in the absence of complementary evidence, as well as giving insight into the population of *C. jejuni*. Considering *C. jejuni* population biology the identification of such widespread genotypic clusters, identical across 9 gene fragments including two that are rather diverse, provides evidence for rapid spread of relatively recently evolved clones and may provide a population structure for more detailed study. For epidemiology this finding implies that identical genotypes in humans, even at apparently high levels of discrimination by usual standards, may provide only weak evidence of close epidemiological relationship and raises the question of just how discriminatory a typing approach needs to be to allow reliable outbreak detection and source tracing. These are dependent on the questions being asked. In the case of source of infection, even within poultry there is a potentially complex ecology. In 2006 there were 14,823 premises on the GB Poultry Register registered as commercial chicken producers[283]. If considering the ideal situation in which each farm produced chicken meat which, when contaminated, always had the same stable genotype and that these genotypes were all non-identical then a scheme with a DI of 99% would on average be able to narrow the possible farm down to 148 out of the 14,823. Increasing discrimination to very high levels will increase the weight that can be given to the finding of isolates of a common type in cases and a suspect source. The levels of discrimination required to support epidemiological needs, typically outbreak detection, case definition within outbreaks, and source identification should guide the development of epidemiological typing schemes. The following chapter presents a theoretical approach to identifying the characteristics that are needed to support outbreak detection.

7. Typing approaches to identify outbreaks

This chapter evaluates the performance of existing typing methods in outbreak detection and considers what characteristics should guide the design of highly discriminatory typing approaches for epidemiological purposes. Identified outbreaks, which have offered important insights supporting disease control in many other diseases[284,285], have been few and mostly small relative to the total burden of *C. jejuni* infection[95,96,204]. The biology of the organism contributes to the lack of identifiable outbreaks[73,74] as discussed in the main introduction (section 1.4.1). However it is possible that many incident outbreaks are not detected[209,286]. Typing has inherent limitations in *C. jejuni* outbreak identification since contamination of a poultry flock, water supply or other source may involve several different lineages of *C. jejuni.* However, effective highly discriminatory subtyping should be able to identify the reasonably large proportion of outbreaks involving a single or dominant lineage[96,204].

Since heavy contamination of flocks of poultry at slaughter is usual[158], flocks of poultry often harbour a dominant strain of *C. jejuni* which is also dominant after slaughter[158] and survival of *C. jejuni* to the point of sale well demonstrated[85,287], it follows that clusters of cases originating from the same poultry production system, involving infection with the same subtype, are likely. Such linked cases may be more diffusely spread in the population and offer only weak clustering in time and space as a signal to prompt outbreak identification[288]. Identified *C. jejuni* outbreaks have not tended to be of this diffuse form that might be expected to be common. They have more usually been clustered according to classical epidemiological links and have often been due to raw milk, local water supplies and food contamination involving social groupings or events[95,96,204]. Typing might therefore identify entirely different

types of outbreak extending further up the food chain than those with a proximate epidemiological link identified currently[288] (Figure 7.1).

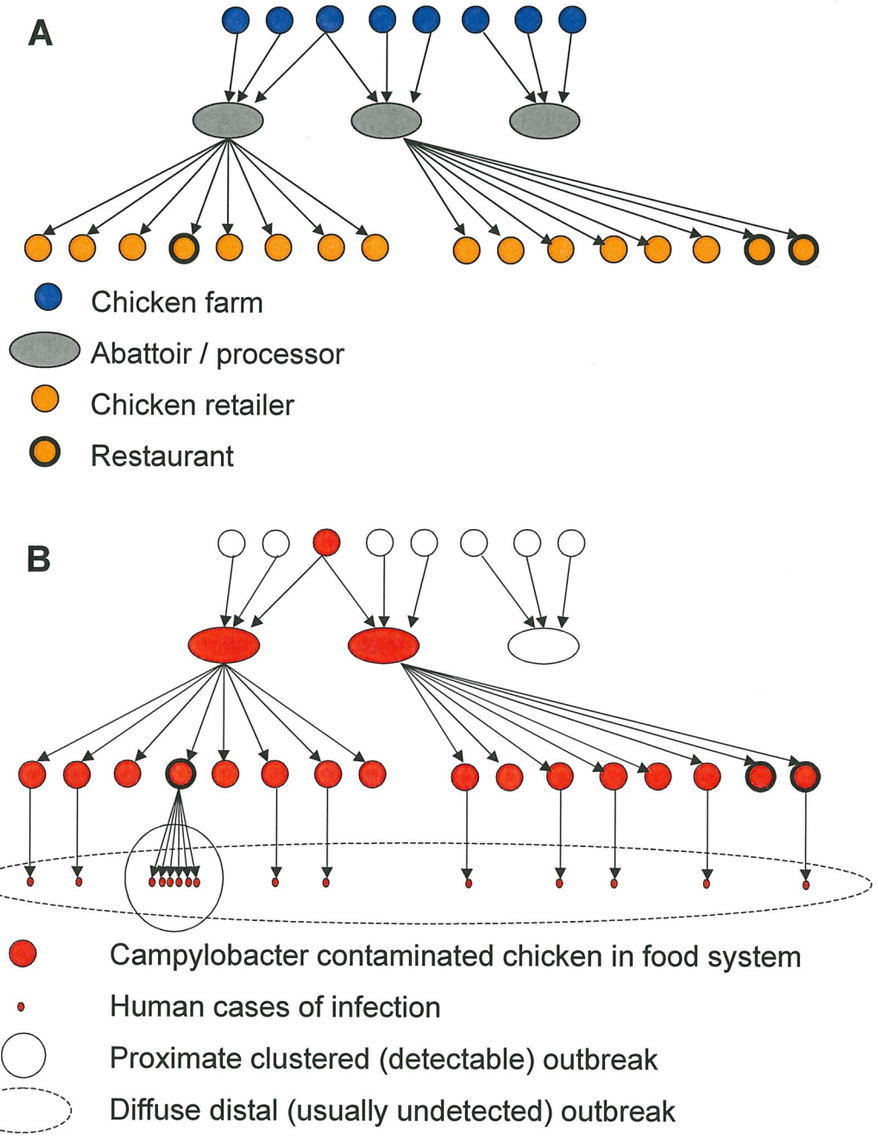


Figure 7.1. Diagrammatic representation of (A) part of the poultry meat supply chain and (B) contamination at a single farm leading to a diffuse outbreak which is unlikely to be detected and traced as well a smaller more proximately linked outbreak, where local factors (e.g. hygiene deficits) produced several more identifiably clustered cases.

As discussed in the introduction many schemes using approaches considered highly discriminatory have been applied to *Campylobacter jejuni* and evaluated in terms of their capacity to link outbreak cases and distinguish them from other cases [88,186,204,209,216,289]. Of these, no currently published method surpasses PFGE for discriminatory power[204,289]. The discrimination index[290] (DI) of PFGE has been estimated as approximately 0.94[207] and 0.97[289]. PFGE has also been more widely applied in outbreak detection than other methods, mainly though PulseNet, a PFGE based outbreak detection scheme for foodborne disease in the US[201,288]. However, although PFGE has been found useful in the further investigation of an outbreak of *C. jejuni* detected by other means [209,288], the lack of PulseNet detected outbreaks for *C. jejuni* stands in stark relief to successes with other organisms[201,288].

This lack of detected outbreaks from PFGE surveillance intended to detect the type of diffuse outbreak predicted in Figure 7.1 for *C. jejuni* may indicate that such outbreaks are not present or that different approaches are needed. This chapter considers how diffuse *C. jejuni* outbreaks might be distributed in the population of surveillance-detected cases and what would result from the application of typing methodologies with different DI and reproducibility characteristics to such a population. This evaluation of performance is based on how far a typing scheme is from the ideal where it linked all outbreak strains into a single genetic subtype, with all non-outbreak isolates excluded from the cluster. The DI of a typing scheme for this purpose is defined as "... the probability that two unrelated strains sampled from the test population will be placed into different typing groups."[290] where "unrelated" refers to not having an epidemiological relationship. The impact of different DI and reproducibility[291] on the usefulness of identified typing groups was evaluated by

simulating the application of typing schemes with different levels of these characteristics to a model population. The model population and simulation method are described in the Methods chapter, section 3.7.

7.1 Results

The number of additional isolates grouped along with the 20-member true cluster (Table 3.11) is shown in Figure 7.2. At a DI of 0.9 and 0.95 the cluster of 20 isolates is usually typed in a group comprising many other isolates, with a median of 51 and 28 extra isolates respectively across the 250 simulations (Figure 7.2. A). At a DI of 0.99 the median number of extra cases is 5 and there are 12 or fewer on 90% of simulations. At a DI of 0.995 or 0.999 (Figure 7.2.B) the outbreak cluster of 20 isolates is identified with few additional incorrectly linked cases in most cases. The distribution of additional cases associated with smaller clusters was similar, such that the ratio of non-cluster to cluster cases was correspondingly higher.

The relationship between true epidemiological clusters of five or more isolates (Table 3.11) and groups of isolates identified as indistinguishable by typing is shown in Figure 7.3 from one typical simulated dataset each for the DI levels of 0.95. 0.99, 0.995 and 0.999, all assuming 100% reproducibility. Isolates that belong to an indistinguishable type were termed a typing group. The typing groups shown on Figure 7.3 are those with a size greater than the smaller of 1) the group size at which all clusters of five or more will be included and 2) a group of size ten. At a DI of 0.95 large typing groups (of up to 32 undifferentiated isolates) were present which contained no true cluster of size five or larger. A single large typing group contained four clusters of isolates, related to each other within each cluster but with no epidemiological relationship between clusters, and a further 40 isolates not belonging

to any of the four clusters (Figure 7.3. Panel A). In the example with DI 0.99 (Figure 7.3. Panel B) two large groups each contained two true clusters in addition to isolates that belonged to neither group. In order to identify all true clusters of five typing groups of size 6 or larger need to be considered, which also included 201 isolates that did not belong in a cluster of size five or more. With DI 0.995 some clusters are not discriminated from each other but most are found in typing groups with only a few additional incorrectly grouped isolates (Figure 7.3. Panel C). Clusters were effectively separated into different groups with very few additional isolates by typing with a DI of 0.999 (Figure 7.3. Panel D). The three typing groups of 5 not containing a cluster of size 5 contain clusters of size 4 and 1 additional isolate, so that no true cluster has more than two unrelated isolates incorrectly grouped with it.

Figure 7.4 shows the impact of reproducibility. Results from 250 simulations are presented for the clusters of size 8 and size 20 generated with a DI 99.5% and reproducibility 95%. Imperfect reproducibility had predictably minor effects with over 90% of simulations identifying seven or more members of the eight-member cluster and 18 or more of the 20-member cluster in a single typing group. For the cluster of size eight the presence of four or more unrelated isolates in the typing group (i.e. at least 33% of the observed typing group not being part of the main cluster) in 66 of the 250 simulations, which is a function of imperfect DI, is of greater relative importance than the effect of imperfect reproducibility.

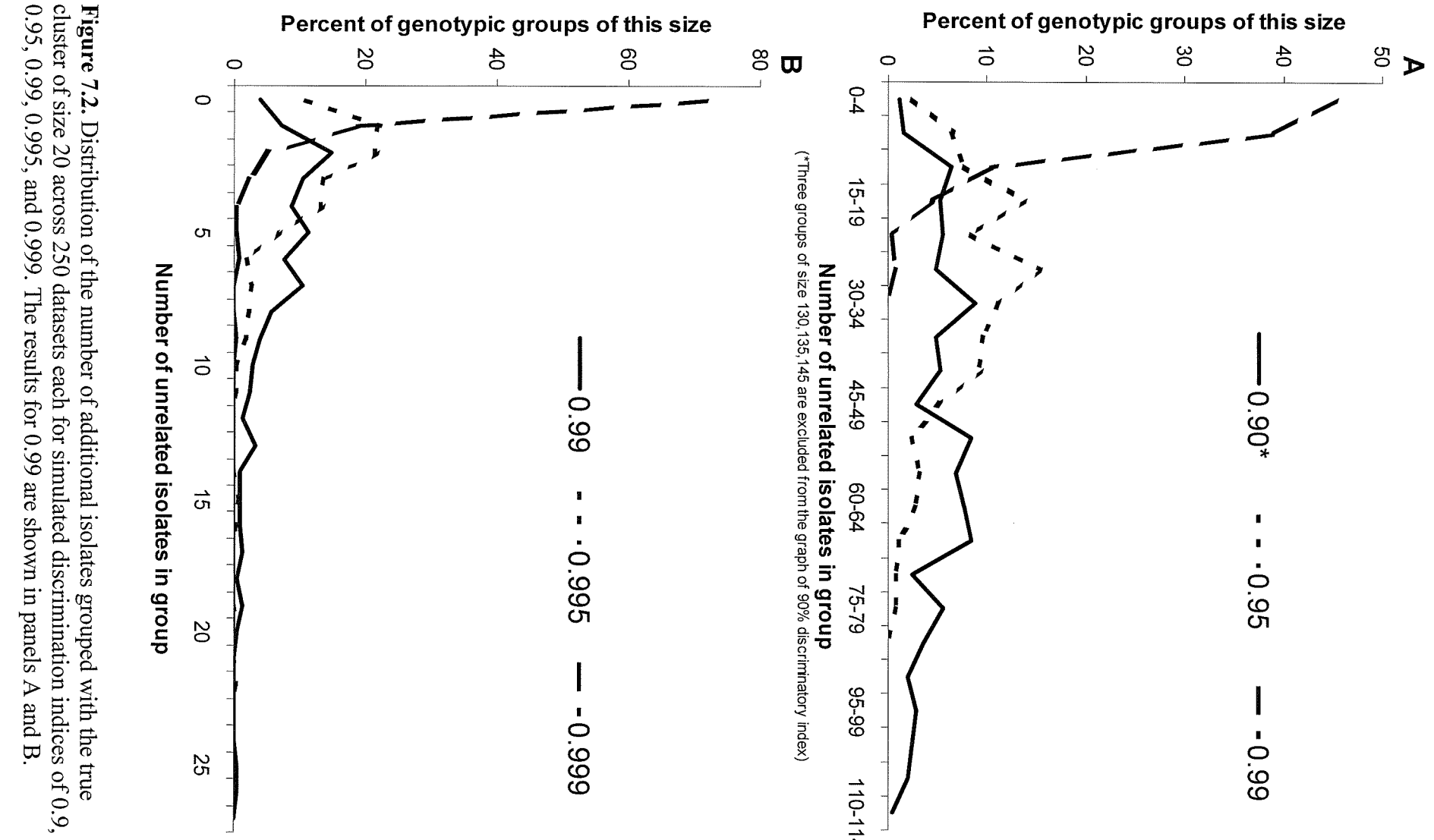


Figure 7.2. Distribution of the number of additional isolates grouped with the true cluster of size 20 across 250 datasets each for simulated discrimination indices of 0.9, 0.95, 0.99, 0.995, and 0.999. The results for 0.99 are shown in panels A and B.

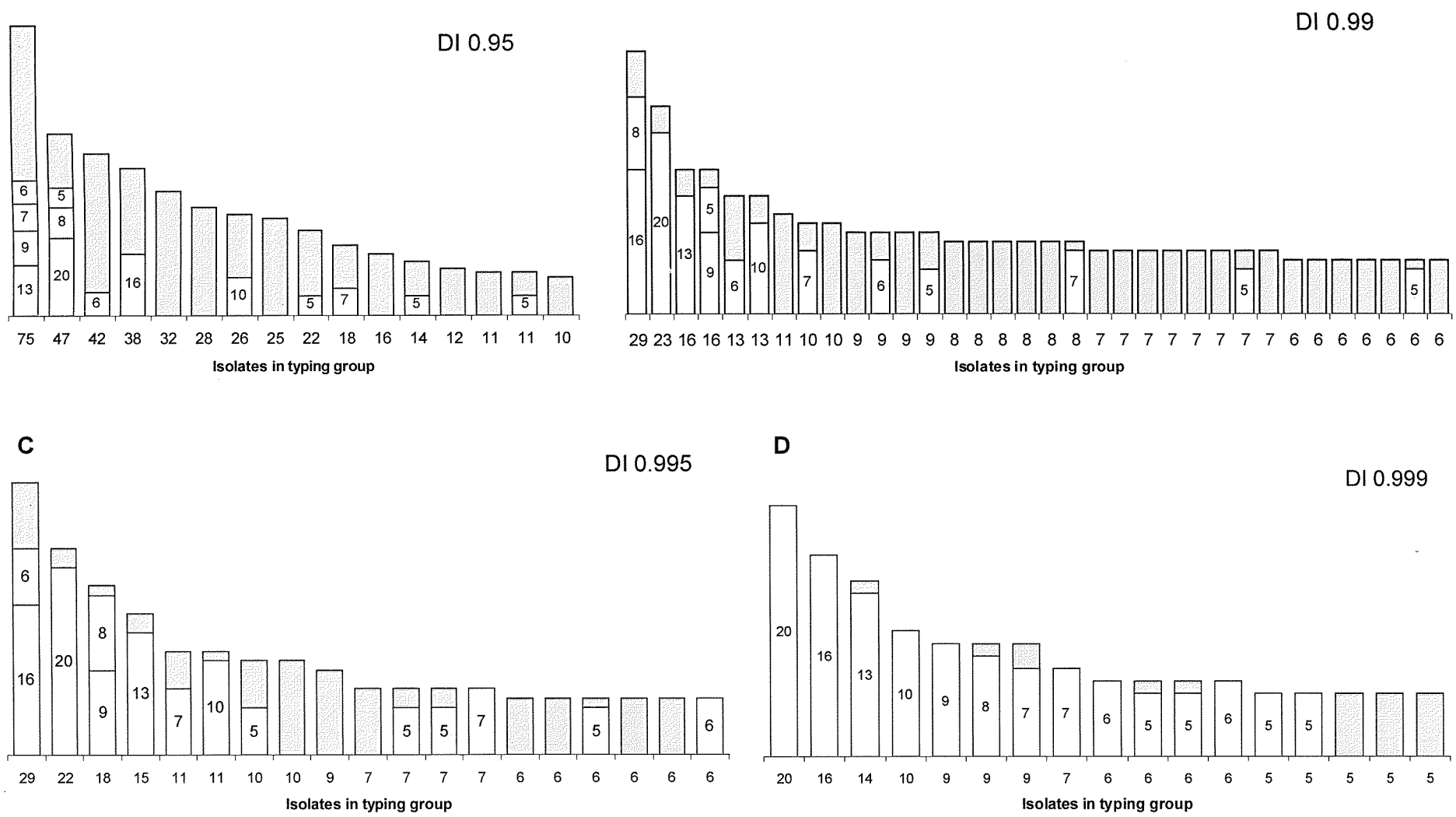


Figure 7.3. Identification of clustered and other isolates within typing groups produced by simulation with discrimination indices of 0.95 (Panel A), 0.99 (Panel B), 0.995 (Panel D), 0.999 (Panel D). Each vertical column represents a group of isolates typed as being identical to each other. True epidemiological clusters are indicated as white areas surrounded by a black line and containing the number of isolates in the cluster. Isolates not belonging to a cluster of 5 or more are shaded grey.

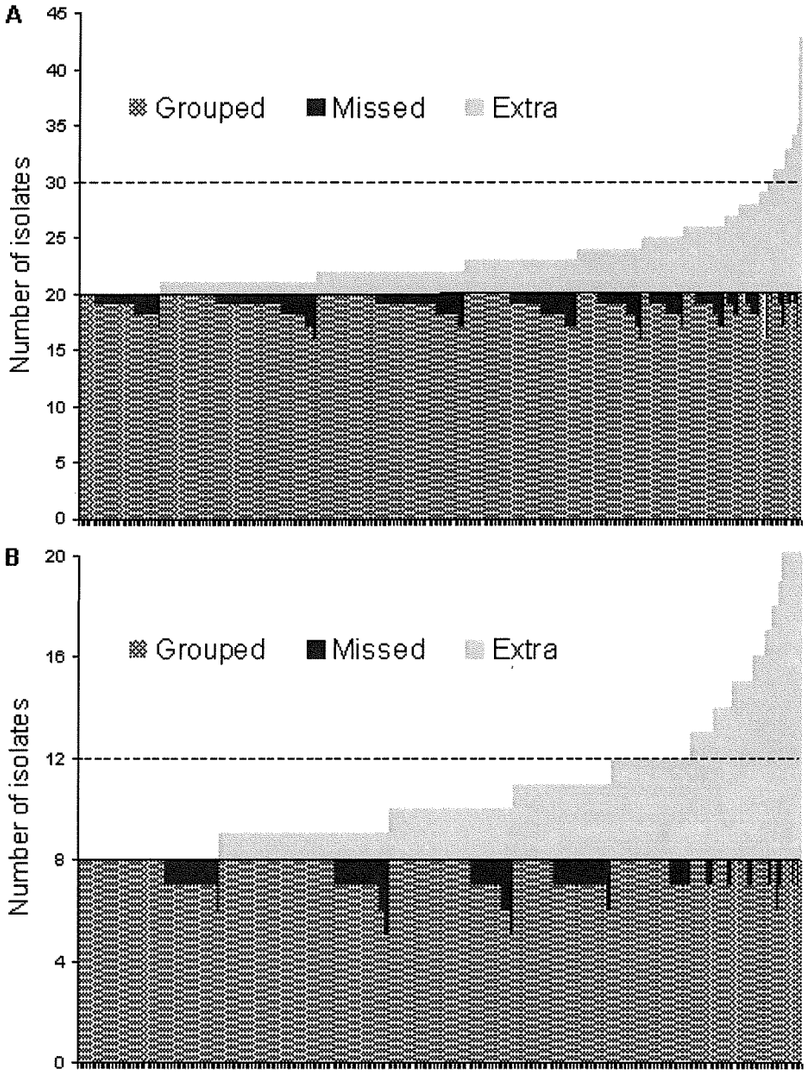


Figure 7.4. Isolates in epidemiological clusters that were correctly identified as identical to each other by the typing scheme (patterned), those that were missed (black) and isolates incorrectly identified as identical to cluster members (grey). Each panel summarises results from 250 simulations with one vertical line per simulation. Simulations were run to ensure that reproducibility was 95% and discriminatory index 99.5%. Panel A shows a cluster of 20 isolates and Panel B a cluster of 8. Dotted lines indicate point at which at least 33% of grouped isolates appearing to be of the cluster type do not belong to the cluster.

7.2 Discussion

This study shows that very highly discriminatory methods are likely to be required to detect outbreaks of *C. jejuni* by typing with the results presented in Figure 7.3 showing that levels of discrimination that have been considered high are at substantial risk of combining more than 1 outbreak and several sporadic cases into a single typing group and that levels of resolution to produce a DI in excess of 99% would be needed. The lesser impact of poor reproducibility, even when as low as 95% is shown in Figure 7.4, where inclusion of external isolates (a feature of DI below 100%) is a greater problem than failure to include some true cluster members in the typing group.

Although often termed highly discriminatory, methods applied to date have not reached these levels of discrimination[204,207,289]. For a population similar to the one in this study current methods would not detect outbreaks with sufficient specificity to be useful. Attempts to identify outbreaks by follow up of such typing groups, which groups comprise true outbreaks contaminated by unrelated cases that are either sporadic or from other clusters is very time consuming[292].

The findings of this study are based on a model population. The model population is likely to be sufficiently close to the real biology and epidemiology of *C. jejuni* to provide lessons for the investigation of this infection. Many cases are sporadic and studies which type isolates typically find a large proportion and sometimes even a majority of cases to be unique within their study population[88,209]. A small number of outbreaks are detected, and these are rarely large unless waterborne[95,206,209,286]. Population based surveillance studies are consistent with the occurrence of many very small clusters[209,286]. This literature suggests that the model population assumed for

this simulation is an acceptable estimate of reality. The marked increase in useful information relative to noise as DI moves from 99% towards 100%, and the very high noise to useful signal ratio with lower levels of DI in this study may therefore explain the inability of previously applied schemes to detect outbreaks[209,288]. The main conclusions of this study will apply to real populations of cases that have even a moderate resemblance to the model population used, and may indeed be generalisable to some other infectious agents. The results are not generalisable to pathogens with very different diversity and a tendency to occur as outbreaks, or to very different populations, such as if applied to a much larger multi-centre population since the different underlying epidemiology could have important effects[293].

As identified in the previous chapter the same issues regarding discrimination arise if using typing to implicate sources of infection. Our preliminary results of applying a typing method with a DI of at least 0.99 to *C. jejuni* isolates from retail poultry suggested that at this level of discrimination apparently identical isolates have originated from different producers, in different countries, and in two sampling periods separated by 4 years. This may be a combination of imperfect discrimination and the possibility for rapid expansion and widespread circulation of some sub types within the poultry production system.

Given the level of discrimination required based on these simulations it is important to assess whether it is achievable in practice. A related underlying question is how best to think about levels of discrimination. The real questions are epidemiological "Were the cases caused by isolates from a common source?" which can also be considered as an applied evolutionary question such as "Did the isolates from these

patients share a common ancestor within the past six months?" if postulating a shared source such as a recent cycle of poultry from a particular producer. Translating this into typing is not trivial, but a worthwhile goal. Although DI is the most widely used metric to evaluate discriminatory typing schemes it was developed and initially applied when typing schemes were mainly phenotypic[291,294,295] and before the current increasing interplay between microbial population genetics and infectious disease epidemiology[1]. A move to identifying the closeness of relatedness rather than whether isolates are the same or different is a useful direction for typing to ensure that applications can benefit from the increasing concordance of many disciplines. This view suggests that typing targets should assay genetic data that is the result of analysable processes. Examples are MLST and variable nucleotide tandem repeats (VNTR). The parameters of mutation and recombination driving change can be estimated from MLST data[259] and support interpretation of different MLST types as degrees of relatedness. Parallels with microsatellite loci in higher organisms for which extensive parametric models have been developed and can index degrees of relatedness[296]serve as a guide for what may be feasible using VNTR typing in microbes. In the interim there are also immediate benefits to using a phylogenetic framework to select typing targets. MLST data which identifies the broad population structure and where an isolate fits in this structure is complemented by addition of genes under strong diversifying selection since these will be likely to have undergone recent change and allow identification of structure at the tips of the phylogenetic tree as shown in the previous chapter.

A further benefit of sequence types compared to older approaches is extreme reproducibility which is critical when concluding that a level of discrimination

equivalent to a DI of 0.999 may be of benefit in identifying epidemiological clusters by typing. The trade-off between increased discrimination and loss of reproducibility has been identified and has been a problem in developing highly discriminatory typing approaches[291]. Since dideoxynucleotide termination sequencing has error rates below 1/10,000[297] it is possible to increase discrimination to a very level while maintaining acceptable reproducibility by appropriate selection of gene targets. In this work on *C. jejuni* combinations of genes used in current typing such as MLST, *flaA* and *porA*[204,207,289] have been chosen. Other targets such as VNTRs may be indexed by sequencing and other methods to offer discrimination levels in this range[298] in other organisms although in vivo stability of these loci[299] may pose problems as well as some issues with between laboratory reproducibility[300] if non sequence approaches are used to assay them. The particular approach and typing targets chosen to deliver the required discrimination may also be influenced by other goals such as ability to predict host in multi-host zoonoses if these are of importance[247,301]. Since sequence types can be stored digitally they also support the building up of a library of reference types. The adoption of such common approaches allows past work to act as a foundation for new applications of such methods, which is particularly important to allow adequate reference data for source attribution[247] as well as facilitating the use of typing data to address additional public health and scientific questions[302].

This study has highlighted the need and possibility to revisit difficult epidemiological problems such as the detection of *C. jejuni* outbreaks, the importance of estimating the required performance characteristics for a typing scheme to address these, and the need to have an ambitious and long-term approach to our use of the array of developing technologies to support this work.

8. Genotype, source of infection, & population biology

This thesis has studied the application of sequence typing to *C. jejuni* in theory and practice in a series of related studies. This final chapter reviews findings from these studies and reports further analyses to extend these findings regarding observed patterns of seasonality, attribution of human infection, the emergence of identifiable clones, and the application of fine typing using increased amounts of sequence data.

Differences in seasonality and geographical distribution were identified between subtypes of *C. jejuni*. The most marked seasonal changes were for ST-45 complex and ST-283 complex which had a substantially higher relative prevalence during the period of peak incidence each summer than during the rest of the year in the UK. This was reflected in a similar seasonality among retail poultry and a related geographical pattern whereby this clonal group comprised a larger proportion of isolates in those countries where the summer seasonal peak is prominent. Although classified as two clonal complexes, analysis to estimate clonal relatedness supported sequence types from these clonal complexes as being more closely related to each other than to isolates from other complexes. The other phenotypic characteristics of ST-45 complex were an apparently wide host range identified in the introduction (Table 1.5), a greater resistance to freezing inferred from relative prevalence on fresh and frozen poultry in retail poultry survey samples, and some evidence for higher prevalence in young children. These findings combined to identify a clonal grouping with particular phenotypic characteristics including a distinctive ecology. In this chapter the comparable peaks between human and retail poultry isolates were examined more fully through a joint analysis and the likely origin of human infection with isolates

from this clonal group assessed by both genotype based attribution and using case-case comparisons [303] of risk factors reported by cases.

Chapter five showed that analysis of MLST data for *C. jejuni* allowed prediction of host, reopening the possibility of using microbial genotype data to predict the source of human *C. jejuni* infections [247]. However substantial limitations to attribution using currently available multilocus data and analytical techniques remain. For example, the very abundant ST-21 complex inhabits many hosts [114,138] and prediction of source in this complex was only successful for sequence types which have imported alleles that have substantial and well estimated differences in frequency between host species [247]. Moreover, identifying the origin of isolates from two known populations (chicken or ruminant) is less challenging than identifying the source of disease isolates from human infection. In attributing human infection the range of possible sources is wide, the sources may not all be known, and reference data is not available for all possible sources. Exploitation of multi-locus data from human cases as a tool for source attribution has not been validated. The further analysis reported on isolates from human cases in this chapter attempted to predict their source by genotype, reviewed these predictions and then compared the self reported exposures between cases from different predicted sources to evaluate the usefulness of the approach for public health application in a more realistic setting.

The observed pattern of a decrease in the prevalent, geographically widespread and very genetically diverse ST-21 complex in humans over the three years of the CSSS, with a similar decline in samples from retail poultry contrasted with evidence for the emergence of other clonal groups characterised by marked increases in prevalence in

humans and UK retail poultry, which groups show limited genetic diversity and more restricted geographic spread. At a finer level of resolution within samples from UK retail poultry, there was some evidence for substantial expansion of narrow clones with identical sequence types across all 7 MLST loci and two antigen loci. Such large and possibly rapid expansion of clones, sharing substantial identity across the genome, within the poultry industry offers an insight into the population biology of *C. jejuni* in this host species and niche. It is also important for epidemiological work in identifying outbreaks and in the specific identification of source for human disease. This extended nine-locus sequence typing was not undertaken in the CSSS dataset and the very limited temporal overlap between this and the RPS dataset may limit the value of comparing isolates using such fine typing between these two data sources. A better temporally matched set of isolates for the retail poultry dataset was available from a national multicentre human case control study of *Campylobacter*. This was used to consider fine typing jointly between human and retail poultry isolates and further explore the identified clonal expansion from population biology and epidemiological perspectives.

This chapter reports the extended analysis to explore the above themes more fully. The specific aims of these three areas of analysis in this chapter were to consider:

What final conclusions can be drawn in this thesis on seasonally patterned subtypes, especially ST-45/ST-283 complex, and the role of this in human infection?

In particular:

i) Is the rise in poultry responsible for the rise in human disease during summer?

ii) Are particular exposures associated with infection due to these subtypes?

What can be inferred regarding the origin of human cases using MLST data?

In particular

i) What proportions of human cases would genotype based assignment ascribe to different sources?

ii) How reliable is this genotype based attribution?

ii) How does genotypic assigned origin fit with "classic epidemiology"?

What can be learnt of the emergence of genetically homogenous clones and what is the significance of this?

In particular

i) Does MLST, *flaA* and *porA* typing allow specific source identification?

ii) Can MLST, *flaA* and *porA* typing index the emergence and spread of clones in the human and retail poultry populations?

8.1 Seasonality and clonal group in human and chicken samples

Temporal patterns of ST-283 complex and ST-45 complex isolates from humans and retail poultry are compared in Figure 8.1 using the CSSS and RPS datasets. The May-August period is highlighted to allow ease of identification of the summer season. Inspection of Figure 8.1 shows a consistent summer peak for the proportion of isolates belonging to the closely related ST-45 / ST-283 complexes. Equivalent graphs are given in Annex 8 for ST-21 complex, ST-257 complex and ST-574 complex. There was a trend for a relatively low proportion of ST-574 complex and ST-257 complex isolates during the summer peak but this was weaker and less consistent and there was no clear seasonal pattern with ST-21 complex (Figure A.8.1, Annex 8).

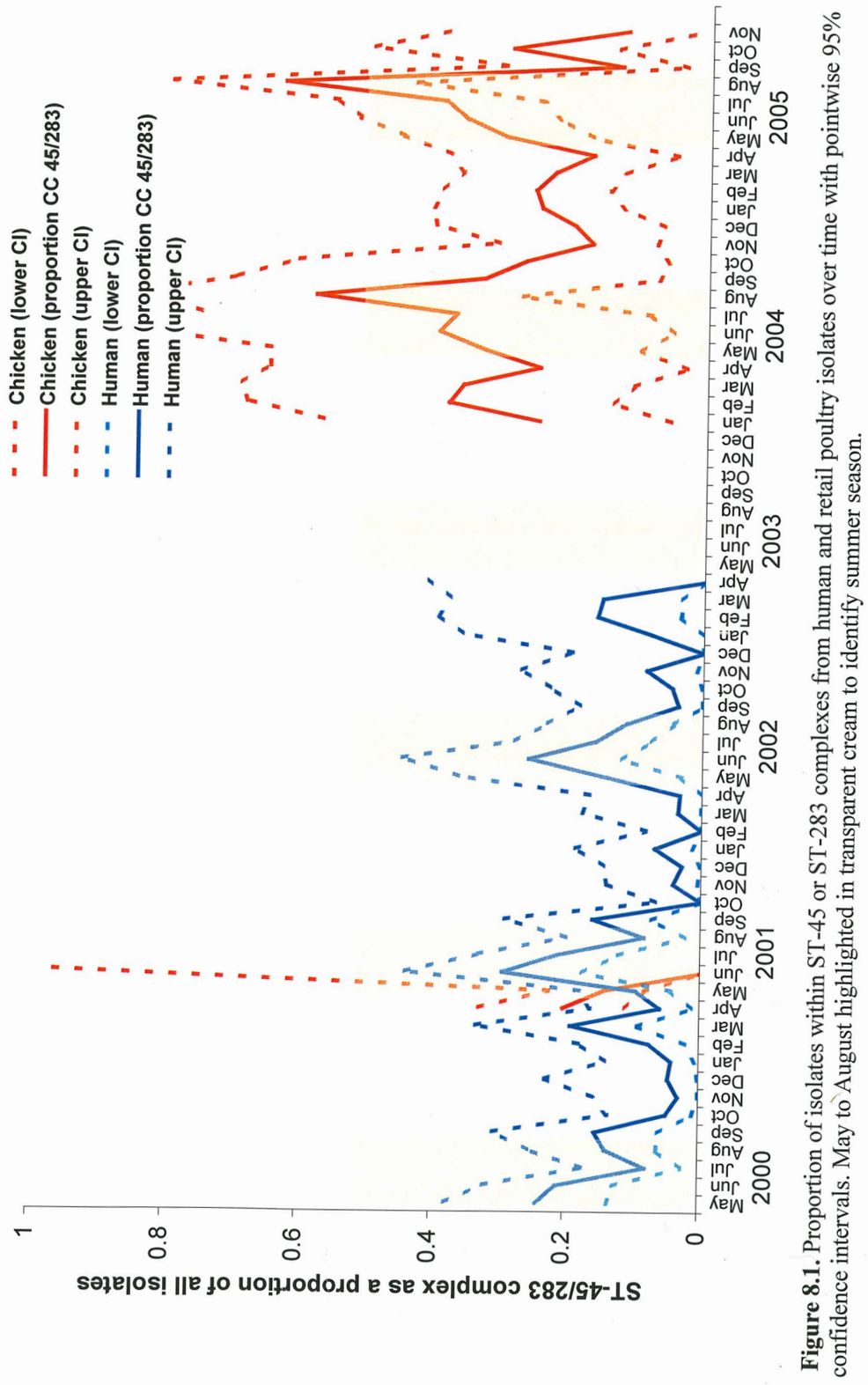


Figure 8.1. Proportion of isolates within ST-45 or ST-283 complexes from human and retail poultry isolates over time with pointwise 95% confidence intervals. May to August highlighted in transparent cream to identify summer season.

The independent associations of host, season and year with membership of ST-45 complex or ST-283 complex were assessed in a joint analysis of the CSSS and RPS using a regression model and showed that:

1. These complexes form a lesser proportion of isolates from human cases (odds ratio 0.5 [95% confidence interval 0.4-0.8], adjusted for long-term trend and seasonal effects) than from retail poultry. This is an attenuation of the unadjusted association, with 26% of the retail poultry isolates belonging to one of these complexes compared to 11% of the sentinel surveillance scheme human isolates.

2. Although both share a similar and marked summer peak the human cases have a summer peak earlier (day 171 for estimated or reported date of onset) than the retail poultry isolates (day 195 for sample date). Interaction testing of the host and seasonality terms indicates that this difference is larger than expected by chance alone ($p = 0.007$). The retail poultry sample dates used in this analysis are the recorded date of sampling from a retail premises. The human sample dates are the recorded or inferred clinical sample date. The timing of human infection leading to these clinical samples would have been earlier requiring the passage of an incubation period for *C. jejuni* and the amount of time from onset of symptoms for a patient to present to healthcare services and have a sample taken, so that the extent to which the peak in human infection preceded the peak in retail poultry contamination is likely to be greater than identified in this analysis.

The limited overlap in year of study for chicken and human isolates limits inference, in particular regarding long-term secular trends in these humans and chicken meat or

relative proportions of particular types in each host. These results should be considered with this strong caveat.

8.2 Genotype and risk factors in the sentinel surveillance scheme

Probabilistic attribution of the 1505 human isolates in the CSSS used the STRUCTURE algorithm as applied in chapter five to isolates of known provenance, but here applied to the human isolates of unknown origin and an attribution reference dataset comprising isolates from ruminants, chicken, starlings and geese (Section 3.2.7). This assigned 51% to ruminants, 44% to chicken and just under 3% each to starlings and geese. The same percentages applied to the 1493 isolates for which the algorithm assigned to a single source species population with at least 0.75 probability. The results for common human STs (those isolated at least 10 times) are summarised in Table 8.1, with full data Annex 9. STs occurring 10 or more times among the 1505 human isolates in the CSSS dataset accounted for 81% of the human isolates, 63% of chicken isolates and 60% of ruminants isolates but just 4% each for starlings and geese, suggesting a substantially shared gene pool for much human and farm animal infection but a differentiated *C. jejuni* population from the two wild bird species in line with the results of assignment based on alleles.

Some sequence types assigned to chicken (e.g. ST-257, ST-573 and ST-574) or to ruminants (e.g. ST-61 and ST-206) based on allele frequencies were also supported by the presence of these STs almost exclusively in that species. However STRUCTURE assigned all ST-21 complex isolates with high probability to a ruminant source. Although overall there was evidence for ruminants as important hosts for this complex (42% of ruminant isolates in the assignment reference dataset were of ST-21

complex compared to 16% of chicken isolates) the substantial prevalence of ST-21 complex among isolates from chicken means that this assignment of human cases with ST-21 complex almost exclusively to ruminant origin is almost certainly inaccurate. The assignment to ruminant origin of some STs that are particularly common among humans and show no obvious differentiation between chicken and ruminant populations further questions the accuracy of this overall allele based assignment approach within ST-21 complex. For example ST-53 occurred 125 times in the CSSS, 12 times in chicken (1.6% of chicken isolates) and 4 times in ruminants (1.5% of ruminant isolates), but all ST-53 isolates were assigned to ruminant origin in the STRUCTURE analysis. If considered with a prior belief that human exposure to *C. jejuni* generally is much more from chicken than ruminant sources then the majority of human ST-21 complex infection could be due to chicken sources.

Other attributions may also be unreliable. ST-403 isolates were attributed to goose origin with high confidence (confirmed on a repeat run of the algorithm) despite no similar isolates in the goose population and low prevalence of alleles in ST-403 in isolates from geese (*aspA* allele 10 once, *pgm* allele 10 three times and *uncA* allele 7 once and no appearance of other ST-403 alleles in the 166 goose isolates). This ST has previously been reported mainly from pigs[231]. A combination of a reference population lacking the true source population and the properties of the STRUCTURE algorithm therefore appear to have led to a high-probability prediction that is unlikely to be correct. The quantitative assignment estimates from STRUCTURE analysis are therefore presented with caveats regarding the accuracy of individual ST assignments, and with evidence, at least for ST-21 complex isolates, that there may be an overall bias producing excessive assignment to ruminant origin over chicken origin.

Table 8.1. Common STs in the CSSS human disease dataset (n ≥10), their distribution by host, and predicted origin using allele information in the STRUCTURE algorithm.

Clonal complex	ST	Human isolates	Predicted origin	Non-human				
				Chicken	Goose	Starling	Ruminant	Total
21	21	174	Ruminant	30			32	62
	53	125	Ruminant	12			4	16
	50	80	Ruminant	18			3	21
	104	54	Ruminant	8				8
	19	42	Ruminant	8			17	25
	262	20	Ruminant	10			22	32
	47	11	Ruminant					0
257	257	181	Chicken	88		2	3	93
45	45	88	Chicken	88		5	10	103
	137	17	Chicken	17	7		1	25
	334	11	Chicken	13		1		14
48	48	95	Ruminant	16			4	20
	475	11	Ruminant	2				2
574	574	50	Chicken	29		1		30
354	354	25	Chicken	12			2	14
	324	13	Chicken					0
353	5	22	Chicken	9				9
	400	13	Chicken	2				2
206	572	20	Ruminant	1				1
	206	10	Ruminant	2			15	17
443	51	27	Chicken	32				32
61	61	22	Ruminant	6			33	39
283	267	22	Chicken	10		1		11
52	52	20	Chicken	4			1	5
42	42	16	Starling	9		1	12	22
22	22	14	Chicken	3			2	5
661	661	12	Geese	10	9			9
49	49	11	Chicken	3			2	5
573	573	10	Chicken	30				30
Total (common STs)		1216		471	7	11	163	652
Overall total		1505		742	166	285	273	1466

In order to identify populations to assess whether origin estimated by genotype correlates with reported exposures it was important to have reliable genotypic assignment. The dataset for this comparison was restricted to where STRUCTURE assignment and ST relative prevalence in the attribution dataset both supported a particular host of origin. Predicted origins from the STRUCTURE algorithm are listed for each common ST in Tables 8.1 and for all STs in Table A.8.1. (Annex 9) along with the distribution of that ST across the four reference populations. Isolates from STs meeting criteria for reliable genotypic estimate of source are summarized in Table 8.2. This table also includes a listing of the ST-45 complex and ST-283 complex isolates. These isolates are listed since their unusual phenotype raised the question of whether infection with this clonal group is associated with particular risk factors. Assessment of association between membership of ST-45/ST-283 complex and exposure history during the summer peak period was made without an assumption of source or prior hypotheses.

The univariate associations between host-specific type and exposures are summarised in Table 8.3. The numbers involved in each analysis are given since some of the total (468) did not have complete questionnaire information for particular questions. Questions on pet ownership were only posed to those saying yes to animal contact and questions on type of restaurant to those who reported eating at a restaurant. This limits the sample size in these analyses as well affecting the meaning of any association observed.

Table 8.2. Isolates where coherent prediction is possible from genotype or belonging to ST-45 complex or ST-283 complex.

Chicken		
CC	ST	Isolates
257	257	181
	316	1
	366	1
	584	9
	824	3
	990	1
	1921	1
	2260	2
	3466	2
353	5	22
	353	3
	356	8
	400	13
	581	4
	2641	1
	3484	1
354	324	13
	339	1
	354	25
	1517	3
	1723	1
	3468	1
433	10	1
	1530	1
443	51	27
	443	1
	772	1
	1365	5
	2807	2
573	573	10
574	574	50
607	607	2
	863	1
	3462	1
	Total	399

Ruminant		
CC.	ST	Isolates
61	61	22
	3212	1
206	46	6
	122	9
	206	10
	222	1
	227	7
	273	2
	290	1
	Total	59

Starling		
CC.	ST	Isolates
177	177	1
677	677	8
682	682	1
	Total	10

ST-45 and ST-283 complex		
CC	ST	Isolates
45	11	2
	25	1
	45	88
	137	17
	230	5
	233	1
	334	11
	538	1
	583	9
	2120	1
	2249	1
	3465	1
	3474	1
	3487	1
283	267	22
	2975	1
	Total	163

Infection with a chicken rather than a ruminant or starling type was positively associated with consumption of chicken or ownership of pets and negatively associated with consumption of "offal or tripe" or eating at a restaurant classified as "miscellaneous". The association with chicken consumption was the primary hypothesis and shows complementary evidence for chicken as the source of infection

in this group of cases reporting more chicken consumption than others on the one hand and infected with a genotype typical of chicken on the other. The negative association with a particular classification of restaurants must be interpreted in the light of no overall association of infection with a chicken type and eating at a restaurant or with the frequency of restaurant use. Reviewing the free-text responses of those who were classified as eating at restaurants of classed as miscellaneous their responses were either "miscellaneous" (n= 8) or "several different restaurants (n = 5). Similarly the association with pet ownership is that among those reporting contact with animals the ownership of one or more pets was associated with infection with a chicken sub-type. Among the primary hypotheses for ruminant types there is no association with meat consumption and a strong negative association with consumption of unpasteurised milk. The offal and pet exposure associations are the complement of those observed for chicken consumption.

Table 8.3. Univariate associations between exposures and infection with host specific subtypes of *C. jejuni*.

		Chicken (n=399)		Ruminant (n=59)		Starling† (n=10)	
Exposure	n	Odds Ratio	p	Odds Ratio	p	Odds Ratio	p
Chicken	455	1.6 [1.1-2.3]	0.03	0.7 [0.4-1.0]	0.05		
Offal and tripe	425	0.4 [0.2-0.9]	0.02	2.7 [1.3-5.7]	0.01		
Unpast. milk	441			0.1 [0.02-1.0]	0.05		
Restaurants							
Miscellan.	186	0.2 [0.1-0.6]	0.01	5.4 [1.6-18.2]	0.01		
Swimming	461					4.4 [1.2-16.2]	0.02
Windsurf	441					47.8 [2.8-825]	0.01
Rec. water*	463					3.9 [1.1-14.2]	0.04
Pets	294	2.1 [1.1-4.2]	0.03	0.4 [0.2-0.8]	0.01		

* Rec. water is a composite from the four listed exposures (fishing, sailing, swimming and windsurfing) and marked as positive for those reporting any of the four.
† On logistic regression starling subtype was associated with consumption of babyfood based on standard approximate statistical tests from regression analysis but not supported (p = 0.115) on exact statistical testing and additionally was lost in an adjusted analysis where age group under 5 or 5 years and older was introduced to test for confounding. Similarly an association with drinking river water (odds ratio 16.1 [1.6-166]) was not supported on exact statistical testing (p = 0.09).

Strong positive associations were identified for starling genotypes with recreational water (freshwater or sea) exposures. Four out of ten cases with a starling sub-type reported any of the recreational water exposures (40%) compared to 15% among the 453 individuals answering these exposure questions but infected with chicken or ruminant types. Among the four individuals reporting recreational water exposure all four reported swimming and 1 windsurfing but none fishing or sailing.

Infection with either ST-45 complex or ST-283 complex isolates during the summer was associated with having mains water (n = 738, odds ratio 2.3 [1.1-4.9], p = 0.03) and consumption of food at a restaurant at school rather than at another type of restaurant in analysis among those reporting eating at a restaurant (n= 305, odds ratio 7.1 [1.4-36.4], p = 0.02). This association persists when adjusted for age as a continuous variable or age groups (0-4, 5-16, 17 and over). Univariate association with babyfood consumption disappeared with adjustment for age group (age four or less or five and over). A positive association with number of pets among those having pets was lost when those without pets were included (coded as zero).

8.3 Discriminatory typing of isolates of human and poultry origin

The same MLST-*flaA-porA* typing scheme used in the RPS survey was applied to isolates from the human case control (CaCo) dataset. The isolates available for subtyping were from two locations and time periods in this multi-centre study, Nottinghamshire during 2003-4 and Lancashire, Cheshire and Merseyside during 2005-6 and thus allowed substantial overlap with the sampling period of the RPS dataset which was mainly sampled during 2004 and 2005. The MLST-*flaA-porA* genotype identified 220 sequence types among the 366 isolates in the CaCo dataset, 168 of which were represented by just 1 isolate with the others in genotypic clusters

of 2 to 27 identical isolates. These are summarised in Table 8.4. showing the sequence type of the largest clusters: six comprising five isolates each; two with nine; and one each with 27, 12 and eight. Sequence type and location is not given for unique isolates or the 41 clusters of size four or less. Eight out of the 11 groups of five or more identical isolates involved both study locations and time periods. The distribution of these clusters in time is shown in Figure 8.2.

Table 8.4. Genotypically identical clusters of *C. jejuni* in the case control study.

Cluster size	Isolates by place		Total	Sequence type		
	LCM*	Notts*		ST	*flaA*	*porA*
27	19	8	27	257	16	1
12	12	0	12	257	16	137
9	8	1	9	21	78	7
9	9	0	9	354	18	58
8	8	0	8	2030	16	1
5	3	2	5	5	14	15
5	3	2	5	19	36	7
5	4	1	5	48	32	14
5	2	3	5	51	21	11
5	3	2	5	53	52	27
5	3	2	5	262	37	8
Summary data for smaller clusters (two to four isolates) and unique isolates						
4			20	5 groups of 4 identical isolates		
3			33	11 groups of 3 identical isolates		
2			50	25 groups of 2 identical isolates		
1			168	168 singletons		
	246	120	366			

* LCM isolates are from Lancashire, Cheshire and Merseyside during 2005-6 and Notts isolates from Nottinghamshire during 2003-4 (see Table 3.4 and accompanying text for fuller description of dataset).

Analysis of temporal clustering gave strong overall statistical evidence (rank sum test $p < 0.0001$) for shorter average times between two isolates of the same genotype (median = 166 days) than between isolates of different genotypic clusters (median = 280 days) albeit that the average time between identical isolates is quite long and no sequence type is very tightly clustered in time. Assessed individually only the three clusters restricted to the second phase of the study (genotypes 257-16-137, 2030-16-1 and 354-18-58) show evidence of temporal clustering relative to isolates from other clusters, although statistical power is low given generally small cluster sizes.

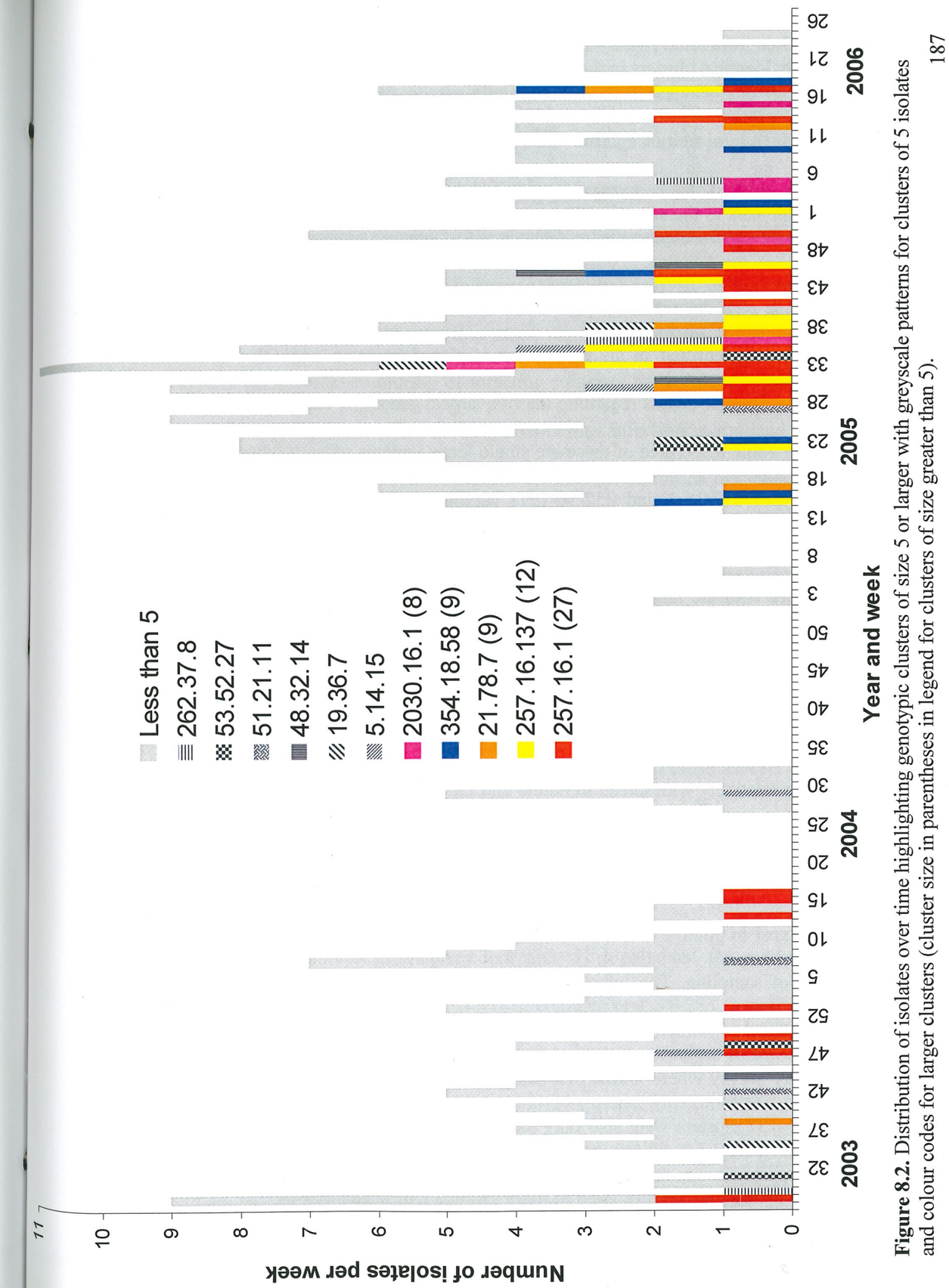


Figure 8.2. Distribution of isolates over time highlighting genotypic clusters of size 5 or larger with greyscale patterns for clusters of 5 isolates and colour codes for larger clusters (cluster size in parentheses in legend for clusters of size greater than 5).

Table 8.5. Data from retail poultry surveys for the genotypes representing the five largest genotypically identical clusters in the case control study.

ST	*flaA*	*porA*	CaCo (n)	RPS (n)	Dates	Producers*	UK / non-UK
257	16	1	27	58	Apr 01 – Nov 05	17	Y/Y
257	16	137	12	9	Nov 04 – Sep 05	5	Y/N
21	78	7	9	1	April 01	1	Y/N
354	18	58	9	3	Aug 04 – Sep 05	1	Y/N
2030	16	1	8	5	Jul 05 – Nov 05	2	Y/N

* Using health mark as the index of different production facilities. This data is incomplete and only clearly different marks were identified as different so that these are conservative estimates.

Reviewing the RPS dataset in the light of these clusters in human cases identifies that the 9 locus genotypes identified in all 11 clusters of size 5 or more were also identified from chicken meat. Further details regarding the five largest genotypic clusters in the CaCo dataset are given in Table 8.5. Three of these are single locus variants of each other, with the largest two (genotypes 257-16-1 and 257-16-137) varying only at the *porA* locus and ST-2030 varying from ST-257 at the *uncA* locus with the nine locus genotype 2030-16-1 sharing antigen genes with the largest 257-16-1 cluster. The allelic variation in *porA* between *porA* allele 1 and *porA* allele 137 is substantial with 5 separate insertion sites and several distant substitutions. ST-2030 differs from ST-257 by having *uncA* allele 12 in place of allele 6 with allele 12 having cytosine residues at positions 165 and 390 in place of thymine in allele 6. Allele 12 is relatively widespread in nature and part of the central genotype for both ST-443 complex and ST-573 complex. In both cases recombination events are the likely origin of variation between these three lineages. The largest human cluster (genotypes 257-16-1) was also the largest in poultry where it was identified 58 times across a wide range of producers and times of sampling. The latest of the five large clusters to appear in the human data, genotype 2030-16-1 which appeared at week 33 in 2005 was also last to be identified in the retail poultry survey where it was first sampled on during week 30 in 2005. The 354-18-58 and 257-16-137 clusters were also only present in the second half of the CaCo data (Figure 8.2) and appeared late in the retail poultry survey (Table 8.5).

8.4 Discussion

This concluding chapter has built on the findings of earlier chapters and on the shared sequence typing information available from each study through further joint analyses of data from different components of the work and additional sources. Some findings from earlier chapters have not been explored further. This was not an assessment of relative importance but was either because they stand alone following the initial analysis or because it has not been possible to explore them further at this stage. An example of the former is the strong evidence against geographic structuring of infecting subtypes between two separate English populations. This has implications for the origin of human infection in England and for the type of reference populations needed to explore disease attribution as discussed in chapter four that do not require further analysis or elaboration. In contrast, the lack of an association between the clonal structure of *C. jejuni* and virulence was based on an analysis restricted to isolates from cases of gastroenteritis. Further comparison with isolates from cases of severe infection such as from bacteraemic cases and with isolates from asymptomatic infection are needed to explore this more fully but represent work that was not feasible within the limits and priorities of this thesis. The rest of this discussion focuses on those aspects of the work that were reviewed and extended in this chapter.

8.4.1 Seasonality

Analysis of *C. jejuni* subtype in chapters four and six highlighted the marked seasonal variation in proportion of infections due to ST-45 complex and ST-283 complex as well as indicating the relatedness of these complexes and the apparent resistance to freezing for ST-45 complex, which is known to have a wide natural host range. This motivated more formal comparison of infection with these subtypes in humans and chicken meat. Although common in both humans and chicken and with a summer peak in both species, joint analysis suggested

some differences. The higher relative prevalence of these isolates in chicken than in humans was substantial and together with evidence for a higher relative prevalence in chicken meat during summer showed that changing levels of contamination in a known source could contribute to the human increase of this subtype rather than needing to identify different sources of infection to explain the subtype increase. However the apparently earlier summer peak in humans does not fit well with a summer ST-45/ST-283 clonal group peak in chicken producing the parallel peak in humans. There are caveats to these findings. As evident from the graphs there is minimal overlap in the years covered by the datasets and it is not possible to be certain that apparent differences between hosts are not the effect of secular changes in relative prevalence of this clonal group and time shift of the seasonal peak from year to year. Although the findings persisted in a multivariate model the adjustment may not have been adequate given the lack of overlapping data on which to model the effect and further data is needed to explore whether these differences between humans and chickens are true.

Estimation of the association of exposures reported by cases with infection due to this clonal group showed associations with both the use of mains water and eating at a school restaurant (among those reporting eating at any restaurant). These do not offer any insight into the seasonal pattern of this clonal group and must be considered in the light of this being a speculative analysis with 85 associations tested and the consequent risk of false positive associations. In particular this analysis does not clarify the role of contaminated poultry meat in the increased relative prevalence observed in humans during the summer peak. Further clarification of the origin of human ST-45 complex and ST-283 complex infection during summer will require further research. Well-matched large sample collections from humans, poultry and other potential sources will be particularly important in this regard. The application of fine typing to such isolates may also be helpful in both identifying the origin of

lineages infecting humans more reliably and in mapping the spread of lineages between and expansion within other hosts and niches. Countries such as Finland with a pronounced summer peak [57] and where ST-45 complex is an important contributor to human infection [61] may offer ideal populations to study these questions.

8.4.2 Attributing human infection using genotype

The analysis in this chapter allowed an evaluation of whether genotype can identify source of human infection, following on from the description of the biology of host association and the demonstration of both the feasibility and limitations of model based attribution using allele data in chapter five. The analysis exploited datasets from published work[247,271], the characterisation of the retail poultry population of *C. jejuni* described in chapter six and the availability of a large sample of sequence typed isolates from patients who had reported their exposure histories. The data strongly supported two of the primary hypotheses tested with consumption of chicken being associated with infection due to a genotype typical of chicken and outdoor water exposures associated with infection due to genotypes typical of a starling population. There was no evidence for association of ruminant types with meat consumption and a negative association with unpasteurised milk.

Despite the small numbers of starling (wild bird) attributable isolates in the CSSS the associations for infection from this putative source with the predicted outdoor environment exposures and infection were highly statistically significant. Although coming from three separate clonal complexes, mainly ST-677 complex, review of the estimated clonal phylogeny from the CSSS dataset (Figure 4.7) shows that these three clonal complexes share a closer clonal relationship to each other than to other isolates. While all of these complexes were rare in the CSSS dataset (8 of 1505 isolates), ST-677 complex was more common in Finland where it comprised 37 of 298 human isolates (28 ST-677 and 9 ST-794) [61]. In a subset of that

study with available exposure questionnaires (45 cases) 4 of 11 cases infected with ST-677 reported drinking water from natural sources or drinking non-chlorinated water from a small water treatment plant compared to 3 and 2 cases reporting these exposures respectively among the 34 cases infected with different clonal complexes. The more marked genetic differentiation of wild bird *C. jejuni* from farm animal types and apparently strong association of these types with particular transmission routes may mean that genetic attribution will be effective in those cases which are due to environmental exposures as such reference datasets build up. Outdoor water exposure has been identified as both a risk factor [86,99,102] and a protective factor [102] in analytical epidemiological studies not considering subtype (Table 1.2). This motivates the further investigation of the potential of MLST to clarify the role of such exposures including the development of reference datasets from wild bird and environmental isolates of *C. jejuni*.

Although less striking than the recreational water findings for starling genotypes the positive association of chicken genotypes with chicken consumption was statistically significant if considered as the main hypothesis under test. Moreover, since this is a common exposure, a weaker level of association can contribute substantially to overall diseased risk, consistent with the evidence that chicken may be the dominant source of human infection (Tables 1.1. and 1.2). There was no association with consumption of meat other than chicken and infection with a ruminant type, which is not very surprising given the lack of evidence for red meat consumption as an important source of infection. A more striking result was the strong negative association of drinking unpasteurised milk with a ruminant subtype. The prior expectation would have been for a positive association of this exposure with infection due to ruminant types. As described in the thesis introduction contaminated unpasteurised milk can pose a major risk of *C. jejuni* gastroenteritis and has been responsible for many reported

outbreaks of infection[25,95,96]. In the CSSS study the question asked to define exposure was for consumption during the two weeks before illness (Annex 2). If consumption during these two weeks was a good marker for chronic consumption then it is possible that immunity due to recurrent exposure could produce a protective effect. Although there is evidence that chronic raw milk exposure may protect well against *C. jejuni* gastroenteritis following consumption of raw milk[304] the argument that this produced the findings reported here would additionally require that immunity acquired by exposure to ruminant subtypes of *C. jejuni* will preferentially protect against disease due to ruminant subtypes, which is speculative.

The positive association of infection with a ruminant type and the consumption of tripe and offal and complementary negative association with chicken consumption are hard to interpret since the species of offal was not requested in the questionnaire. In case control studies offal has also been identified as both a risk factor[105] and protective factor[86] (Table 1.2). The current findings do little to reduce the uncertainty regarding this exposure! Given the recurrent identification of this exposure it may be worth specifying it in more detail in future studies to clarify the species origin of offal consumed. The association with ownership of pet animals (contrasted with exposure to other non-pet animals) with chicken subtypes and corresponding negative association for ruminant subtypes is similarly difficult to interpret since pet ownership was a supplementary question to those reporting that they did have some form of animal exposure. A true tendency for pets to become infected with subtypes similar to chicken could produce this finding but so could a risk for ruminant subtypes occurring in those who had contact in animals other than pets. There is also the possibility that these associations observed outside the primary hypotheses analysed represent chance findings in the context of multiple testing.

Overall this analysis has identified further potential utility but also limitations for the use of MLST and other sequence typing data in the attribution of cases of human *C. jejuni* infection to possible sources. It provides a strong basis for further extending available reference datasets to wild animal and environmental niches, where genetic differentiation may be stronger and particularly informative and offers some guidance on important exposure information to gather from cases in order to maximise the capacity to interpret subtype effectively in source identification.

8.4.3 Analysis of discriminatory typing of human and poultry isolates

Building on the evaluation of discriminatory typing in chapters six and seven a combined analysis of data from human isolates with that from poultry identified genotypic clusters of identical isolates that were present both datasets. All eleven clusters of size five or larger in the human data involved nine locus genotypes also present among the poultry isolates. Some of these nine locus genotype clusters were widespread in space and time as well as in both human and chicken isolates. This further supports the evidence from the retail poultry analysis in chapter six that identical types should be treated cautiously as evidence for a particular shared point source, even at this level of discriminatory typing.

The ST-257 complex was striking for the extremely large proportion of isolates that belong to the central genotype in datasets analysed in this thesis, which finding was most obvious in the large CSSS and RPS datasets (90% of the 201 ST-257 complex isolates in the CSSS and 64% of the 137 in the RPS), but also supported in the smaller Australian[243] and New Zealand[267] datasets where it was the most common ST in the complex. It is not possible to be certain whether the increased diversity within the complex in the RPS compared to the CSSS was due to these isolates being later or due to differences in chicken and human populations. In any

case, the marked increase in contribution to human disease over the course of the CSSS and the relative lack of diversity are consistent with the expansion of ST-257 complex being a relatively recent and rapid event. The wide geographical range contrasts with the apparently narrow host range and limited diversity of ST-257 complex, although there is some geographical restriction with no evidence for ST-257 complex in 298 cases from Finland apart from the single isolate of ST-854[61] which, although ascribed to ST-257 complex, is a three locus variant of ST-257, a three locus variant of ST-354 and a single locus variant of STs in both ST-257 complex and ST-354 complex making it's clonal origin difficult to determine .Similarly the only ST-257 complex contribution to the Senegalese chicken dataset (n=46)[240] is a single isolate of ST-854.

This picture of the emergence of ST-257 complex is mirrored in the analysis using more highly resolved typing described in this chapter. Three of the five largest clusters in humans were within ST-257 complex and these lineages had evidence for even stronger clonal relatedness to each other, with the two nine-locus sequence types identified later in both human and retail poultry samples each a single locus variant of the MLST-*flaA-porA* genotype 257-16-1. The lack of typed isolates from other sources limits the extent to which it can be inferred that these clones were specific to chicken and humans. However as was evident in the retail poultry study and further shown in the attribution work in the current chapter ST-257 complex was typical of chicken and has rarely been found outside this host, other than in humans, making the most plausible explanation that the three ST-257 complex clusters identified in humans came from chicken and represent lineages within a particular clonal expansion in this host species. The evidence for recent emergence, most likely within poultry, is most compelling for the ST-2030-16-1 (MLST-*flaA-porA*) genotype. The two records of ST-2030 on the pubMLST website (http://pubmlst.org/) at 9 August 2008 were

human disease cases from 2006, one for Northwest England and the other from Grampian in Scotland. ST-2030 is absent from all datasets studied in this thesis until July 2005 including the 1505 isolate CSSS dataset in England and the PHD dataset (from Finland, Australia and New Zealand) but is present repeatedly in the human and retail poultry samples from that time onwards. It is a single locus variant of the widespread ST-257-16-1 and reviewing all ST-2030 isolates in the thesis datasets all eight in the case control dataset and five of the six in the RPS dataset have this identical antigen type at both *porA* and *flaA* loci. All available data are therefore supportive for this clone originating during the summer of 2005 within retail poultry arising from the widespread ST-257-16-1 clone. The 354-18-58 and 257-16-137 clusters were also only present in the second half of the CaCo data (Figure 8.2) and appeared late in the retail poultry survey (Table 8.5), although here wider dataset comparison is not possible since ST-354 and ST-257 are common and the *porA* and *flaA* sequences not available to allow assessment of whether these lineages were absent in the CSSS and other earlier datasets.

Although there are caveats in interpreting identity at these 9-locus types as indicating a very close epidemiological or phylogenetic link, the further analysis of this chapter does appear to allow tracking of emerging clones consistent with the more diffuse "outbreaks" hypothesised in chapter seven with evidence for clonal lineages expanding rapidly, contributing to the substantial ST-257 complex increase in recent years, and causing human cases over a wide geography and timescale within the UK. For at least one lineage (2030-16-1) an approximate time of emergence can be identified more accurately. The detail added by antigen genes therefore further supports the picture of expansion of non-diverse lineages within ST-257 complex that was suggested by analysis of MLST data. Moreover it suggests that such expansions within the retail poultry industry may be substantial with a single rather homogenous clone comprising a substantial proportion of the total RPS sample, and two

closely related lineages which are likely to have arisen from this clone also identified on repeated occasions. These findings offer both insight into the population dynamics of *C. jejuni* in the poultry industry niche and an opportunity for further study by identifying and mapping out the behaviour of these clonal groups. For example, the ST-2030-16-1 lineage with an estimated time of emergence and several identified isolates provides an opportunity to assess the rate of change across the genome in the retail poultry niche.

The other main emergent clonal complex in chicken and humans over this time was ST-574 complex. As discussed in chapter four it showed more geographical restriction than ST-257 complex in the international comparison of human cases, a substantial increase over time from a low initial level in the CSSS dataset (Figure 4.4), and very strong dominance of the central genotype. However, perhaps due to lower overall numbers, there was less evidence for 9-locus clusters and sharing of identical 9 locus genotypes between human and chicken datasets was more limited. Of the 14 isolates in the case control dataset seven were part of clusters of size 4 and 3. However neither of these genotypes was present in the retail poultry survey samples. Two singletons in the case control sample did match isolates from the RPS, ST-574-8-2 appearing three times in the RPS and ST-574-8-285 once. If this clonal complex continues to expand it may provide future opportunities to observe microevolution within *C. jejuni*.

Overall this thesis has developed approaches to integrate sequence typing into overlapping questions in the epidemiology and population biology of *C. jejuni*. It has demonstrated important positive and negative findings regarding the ecological structure and epidemiological distribution of *C. jejuni* sub populations and investigated the related questions of outbreak investigation and micro-evolution of this species, as well as considering

the picture demonstrated in broader brush-stroke analysis regarding attribution and overall sources for human infection. The great added value of the addition of common sequence type data to analysis within and between studies has thus allowed increased knowledge and understanding of the epidemiology and biology of *C. jejuni* in these areas as well as making explicit the limitations of this additional information. Many of these limitations are tractable with additional data, which is rapidly accruing in this field and likely to be scientifically productive and of benefit to public health interventions in the coming years.

9. References

1. Levin BR, Lipsitch M, Bonhoeffer S. Population biology, evolution, and infectious disease: convergence and synthesis. *Science*. 1999;283:806-9.
2. Karmali MA, Fleming PC. *Campylobacter* enteritis. *Can Med Assoc J.* 1979;120:1525-32.
3. Gillespie IA, O'Brien S J, Frost JA, Tam C, Tompkins D, Neal KR, Syed Q, Farthing MJ. Investigating vomiting and/or bloody diarrhoea in *Campylobacter jejuni* infection. *J Med Microbiol*. 2006;55:741-6.
4. Skirrow MB, Jones DM, Sutcliffe E, Benjamin J. *Campylobacter* bacteraemia in England and Wales, 1981-91. *Epidemiol Infect*. 1993;110:567-73.
5. Skirrow MB, Blaser MJ. Clinical aspects of *Campylobacter* infection. In: Nachamkin I, Blaser MJ, eds. *Campylobacter*. Washington D.C.: ASM; 2000.
6. Pope JE, Krizova A, Garg AX, Thiessen-Philbrook H, Ouimet JM. *Campylobacter* reactive arthritis: a systematic review. *Semin Arthritis Rheum*. 2007;37:48-55.
7. Soderlin MK, Kautiainen H, Puolakkainen M, Hedman K, Soderlund-Venermo M, Skogh T, Leirisalo-Repo M. Infections preceding early arthritis in southern Sweden: a prospective population-based study. *J Rheumatol*. 2003;30:459-64.
8. McCarthy N, Giesecke J. Incidence of Guillain-Barré syndrome following infection with *Campylobacter jejuni*. *Am J Epidemiol*. 2001;153:610-4.
9. Tam CC, Rodrigues LC, Petersen I, Islam A, Hayward A, O'Brien SJ. Incidence of Guillain-Barré syndrome among patients with *Campylobacter* infection: a general practice research database study. *J Infect Dis*. 2006;194:95-7.
10. Buzby JC, Allos BM, Roberts T. The economic burden of *Campylobacter*-associated Guillain-Barré syndrome. *J Infect Dis*. 1997;176 Suppl 2:S192-7.
11. Black RE, Levine MM, Clements ML, Hughes TP, Blaser MJ. Experimental *Campylobacter jejuni* infection in humans. *J Infect Dis*. 1988;157:472-9.
12. Tompkins DS, Hudson MJ, Smith HR, Eglin RP, Wheeler JG, Brett MM, Owen RJ, Brazier JS, Cumberland P, King V, Cook PE. A study of infectious intestinal disease in England: microbiological findings in cases and controls. *Commun Dis Public Health*. 1999;2:108-13.
13. Amar CF, East CL, Gray J, Iturriza-Gomara M, Maclure EA, McLauchlin J. Detection by PCR of eight groups of enteric pathogens in 4,627 faecal samples: re-examination of the English case-control Infectious Intestinal Disease Study (1993-1996). *Eur J Clin Microbiol Infect Dis*. 2007;26:311-23.
14. Olesen B, Neimann J, Bottiger B, Ethelberg S, Schiellerup P, Jensen C, Helms M, Scheutz F, Olsen KE, Krogfelt K, Petersen E, Molbak K, Gerner-Smidt P. Etiology of diarrhea in young children in Denmark: a case-control study. *J Clin Microbiol*. 2005;43:3636-41.
15. Svenungsson B, Lagergren A, Ekwall E, Evengard B, Hedlund KO, Karnell A, Lofdahl S, Svensson L, Weintraub A. Enteropathogens in adult patients with diarrhea and healthy control subjects: a 1-year prospective study in a Swedish clinic for infectious diseases. *Clin Infect Dis*. 2000;30:770-8.
16. Hellard ME, Sinclair MI, Hogg GG, Fairley CK. Prevalence of enteric pathogens among community based asymptomatic individuals. *J Gastroenterol Hepatol*. 2000;15:290-3.
17. Rees JH, Soudain SE, Gregson NA, Hughes RA. *Campylobacter jejuni* infection and Guillain-Barré syndrome. *N Engl J Med*. 1995;333:1374-9.

18. Skirrow MB, Butzler JP. Foreword. In: Nachamkin I, Blaser MJ, eds. *Campylobacter*. Washington D.C.: ASM; 2000:xviii-xxiii.
19. Blaser MJ, Taylor DN, Feldman RA. Epidemiology of *Campylobacter jejuni* infections. *Epidemiol Rev*. 1983;5:157-76.
20. Karmali MA, Fleming PC. *Campylobacter* enteritis in children. *J Pediatr*. 1979;94:527-33.
21. Svedhem A, Kaijser B. *Campylobacter fetus* subspecies *jejuni*: a common cause of diarrhea in Sweden. *J Infect Dis*. 1980;142:353-9.
22. Oberhelman RA, Taylor DN. *Campylobacter* infections in developing countries. In: Nachamkin I, Blaser MJ, eds. *Campylobacter*. Washington D.C.: ASM; 2000.
23. Blaser MJ, LaForce FM, Wilson NA, Wang WL. Reservoirs for human campylobacteriosis. *J Infect Dis*. 1980;141:665-9.
24. Wheeler JG, Sethi D, Cowden JM, Wall PG, Rodrigues LC, Tompkins DS, Hudson MJ, Roderick PJ. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. *BMJ*. 1999;318:1046-50.
25. Friedman CR, Neimann J, Wegener HC, Tauxe RV. Epidemiology of *Campylobacter jejuni* infection in the United States and other industrialized nations. In: Nachamkin I, Blaser MJ, eds. *Campylobacter*. 2ND ed. Washington: ASM Press,; 2000:121-138.
26. Butzler JP, Dekeyser P, Detrain M, Dehaen F. Related vibrio in stools. *J Pediatr*. 1973;82:493-5.
27. Linneberg A, Ostergaard C, Tvede M, Andersen LP, Nielsen NH, Madsen F, Frolund L, Dirksen A, Jorgensen T. IgG antibodies against microorganisms and atopic disease in Danish adults: the Copenhagen Allergy Study. *J Allergy Clin Immunol*. 2003;111:847-53.
28. Strid MA, Engberg J, Larsen LB, Begtrup K, Molbak K, Krogfelt KA. Antibody responses to *Campylobacter* infections determined by an enzyme-linked immunosorbent assay: 2-year follow-up study of 210 patients. *Clin Diagn Lab Immunol*. 2001;8:314-9.
29. Cawthraw SA, Feldman RA, Sayers AR, Newell DG. Long-term antibody responses following human infection with *Campylobacter jejuni*. *Clin Exp Immunol*. 2002;130:101-6.
30. Coker AO, Isokpehi RD, Thomas BN, Amisu KO, Obi CL. Human campylobacteriosis in developing countries. *Emerg Infect Dis*. 2002;8:237-44.
31. Calva JJ, Ruiz-Palacios GM, Lopez-Vidal AB, Ramos A, Bojalil R. Cohort study of intestinal infection with *Campylobacter* in Mexican children. *Lancet*. 1988;1:503-6.
32. Taylor DN, Perlman DM, Echeverria PD, Lexomboon U, Blaser MJ. *Campylobacter* immunity and quantitative excretion rates in Thai children. *J Infect Dis*. 1993;168:754-8.
33. Ani EA, Takahashi T, Shonekan RA. *Campylobacter jejuni* antibodies in Nigerian children. *J Clin Microbiol*. 1988;26:605-6.
34. Figueroa G, Galeno H, Troncoso M, Toledo S, Soto V. Prospective study of *Campylobacter jejuni* infection in Chilean infants evaluated by culture and serology. *J Clin Microbiol*. 1989;27:1040-4.
35. Georges-Courbot MC, Beraud-Cassel AM, Gouandjika I, Georges AJ. Prospective study of enteric *Campylobacter* infections in children from birth to 6 months in the Central African Republic. *J Clin Microbiol*. 1987;25:836-9.
36. Pazzaglia G, Bourgeois AL, el Diwany K, Nour N, Badran N, Hablas R. *Campylobacter* diarrhoea and an association of recent disease with asymptomatic shedding in Egyptian children. *Epidemiol Infect*. 1991;106:77-82.

37. Richardson NJ, Koornhof HJ, Bokkenheuser VD, Mayet Z, Rosen EU. Age related susceptibility to *Campylobacter jejuni* infection in a high prevalance population. *Arch Dis Child*. 1983;58:616-9.
38. Gillespie IA, O'Brien SJ, Frost JA, Adak GK, Horby P, Swan AV, Painter MJ, Neal KR. A case-case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: a tool for generating hypotheses. *Emerg Infect Dis*. 2002;8:937-42.
39. HPA. *Campylobacter* spp. Laboratory reports of faecal isolates reported to the Health Protection Agency 1986-2006. In; http://www.hpa.org.uk/infections/topics_az/campy/data_ew.htm Accessed 10 December 2007.
40. Humphrey T, O'Brien S, Madsen M. *Campylobacters* as zoonotic pathogens: a food production perspective. *Int J Food Microbiol*. 2007;117:237-57.
41. Baker MG, Sneyd E, Wilson NA. Is the major increase in notified campylobacteriosis in New Zealand real? *Epidemiol Infect*. 2007;135:163-70.
42. Green CG, Krause D, Wylie J. Spatial analysis of *Campylobacter* infection in the Canadian province of Manitoba. *Int J Health Geogr*. 2006;5:2.
43. HPA. *Campylobacter* spp. Laboratory reports from faecal isolates reported to the Health Protection Agency Centre for Infections. England & Wales, 1990-2006. In; http://www.hpa.org.uk/infections/topics_az/campy/data_ew_age_rates.htm Acessed 12 December 2007.
44. van Hees BC, Veldman-Ariesen MJ, de Jongh BM, Tersmette M, van Pelt W. Regional and seasonal differences in incidence and antibiotic resistance of *Campylobacter* from a nationwide surveillance study in The Netherlands: an overview of 2000-2004. *Clin Microbiol Infect*. 2007;13:305-10.
45. Louis VR, Gillespie IA, O'Brien SJ, Russek-Cohen E, Pearson AD, Colwell RR. Temperature-driven *Campylobacter* seasonality in England and Wales. *Appl Environ Microbiol*. 2005;71:85-92.
46. Sopwith W, Ashton M, Frost JA, Tocque K, O'Brien S, Regan M, Syed Q. Enhanced surveillance of campylobacter infection in the North West of England 1997-1999. *J Infect*. 2003;46:35-45.
47. Skirrow MB. A demographic survey of *Campylobacter*, *Salmonella* and *Shigella* infections in England. A Public Health Laboratory Service Survey. *Epidemiol Infect*. 1987;99:647-57.
48. Kapperud G, Aasen S. Descriptive epidemiology of infections due to thermotolerant *Campylobacter* spp. in Norway, 1979-1988. *APMIS*. 1992;100:883-90.
49. Thompson JS, Cahoon FE, Hodge DS. Rate of *Campylobacter* spp. isolation in three regions of Ontario, Canada, from 1978 to 1985. *J Clin Microbiol*. 1986;24:876-8.
50. Sibbald CJ, Sharp JC. *Campylobacter* infection in urban and rural populations in Scotland. *J Hyg (Lond)*. 1985;95:87-93.
51. Talsma E, Goettsch WG, Nieste HL, Schrijnemakers PM, Sprenger MJ. Resistance in *Campylobacter* species: increased resistance to fluoroquinolones and seasonal variation. *Clin Infect Dis*. 1999;29:845-8.
52. CSSS. Ethnicity and *Campylobacter* infection: a population-based questionnaire survey. *J Infect*. 2003;47:210-6.
53. Manaseki S, Hawker J, Ali S. Ethnic inequalities in *Campylobacter* infection in Birmingham, UK: descriptive study of notified cases. *J Epidemiol Community Health*. 2004;58:278-9.
54. Lassen J, Kapperud G. Epidemiological aspects of enteritis due to *Campylobacter* spp. in Norway. *J Clin Microbiol*. 1984;19:153-6.

55. Ethelberg S, Simonsen J, Gerner-Smidt P, Olsen KE, Molbak K. Spatial distribution and registry-based case-control analysis of *Campylobacter* infections in Denmark, 1991-2001. *Am J Epidemiol*. 2005;162:1008-15.
56. Hearnden M, Skelly C, Eyles R, Weinstein P. The regionality of campylobacteriosis seasonality in New Zealand. *Int J Environ Health Res*. 2003;13:337-48.
57. Nylen G, Dunstan F, Palmer SR, Andersson Y, Bager F, Cowden J, Feierl G, Galloway Y, Kapperud G, Megraud F, Molbak K, Petersen LR, Ruutu P. The seasonal distribution of *Campylobacter* infection in nine European countries and New Zealand. *Epidemiol Infect*. 2002;128:383-90.
58. Kovats RS, Edwards SJ, Charron D, Cowden J, D'Souza RM, Ebi KL, Gauci C, Gerner-Smidt P, Hajat S, Hales S, Hernandez Pezzi G, Kriz B, Kutsar K, McKeown P, Mellou K, Menne B, O'Brien S, van Pelt W, Schmid H. Climate variability and *Campylobacter* infection: an international study. *Int J Biometeorol*. 2005;49:207-14.
59. Jones K, Betaieb M, Telford DR. Thermophilic *Campylobacters* in surface waters around Lancaster, UK: negative correlation with *Campylobacter* infections in the community. *J Appl Bacteriol*. 1990;69:758-64.
60. Elliott TS, Payne GM, Lewis MJ. *Campylobacter* enteritis in Nottingham. *Ecol Dis*. 1983;2:291-4.
61. Karenlampi R, Rautelin H, Schonberg-Norio D, Paulin L, Hanninen ML. Longitudinal Study of Finnish *Campylobacter jejuni* and *C. coli* Isolates from Humans, Using Multilocus Sequence Typing, Including Comparison with Epidemiological Data and Isolates from Poultry and Cattle. *Appl Environ Microbiol*. 2007;73:148-55.
62. Steingrimsson O, Thorsteinsson SB, Hjalmarsdottir M, Jonasdottir E, Kolbeinsson A. *Campylobacter* ssp. infections in Iceland during a 24-month period in 1980-1982. Clinical and epidemiological characteristics. *Scand J Infect Dis*. 1985;17:285-90.
63. van Pelt W, van der Heijden SJ, van Duynhoven Y. Similarities and differences in seasonality of *Campylobacter* in broilers and humans, 1998-2006, the Netherlands. In: *Campylobacter Helicobacter and Related Organisms*. Rotterdam: Blackwell; 2007:51-52.
64. Miller G, Dunn GM, Smith-Palmer A, Ogden ID, Strachan NJ. Human campylobacteriosis in Scotland: seasonality, regional trends and bursts of infection. *Epidemiol Infect*. 2004;132:585-93.
65. Shmilovitz M, Kretzer B, Rotman N. *Campylobacter jejuni* as an etiological agent of diarrheal diseases in Israel. *Isr J Med Sci*. 1982;18:935-40.
66. Cabrita J, Pires I, Vlaes L, Coignau H, Levy J, Goossens H, Goncalves AP, de Mol P, Butzler JP. *Campylobacter* enteritis in Portugal: epidemiological features and biological markers. *Eur J Epidemiol*. 1992;8:22-6.
67. Varoli O, Gatti M, Piscolla F. A one-year study on thermophilic *Campylobacters* isolated from fecal specimens. *Microbiologica*. 1989;12:263-5.
68. Velasco AC, Mateos ML, Mas G, Pedraza A, Diez M, Gutierrez A. Three-year prospective study of intestinal pathogens in Madrid, Spain. *J Clin Microbiol*. 1984;20:290-2.
69. Alam M, Akhtar YN, Ali SS, Ahmed M, Atiq M, Ansari A, Chaudhry FA, Bashir H, Bangash MA, Awais A, Safdar A, Hasnain SF, Zafar A. Seasonal variation in bacterial pathogens isolated from stool samples in Karachi, Pakistan. *J Pak Med Assoc*. 2003;53:125-9.
70. Endtz HP, van West H, Godschalk PC, de Haan L, Halabi Y, van den Braak N, Kesztyus BI, Leyde E, Ott A, Verkooyen R, Price LJ, Woodward DL, Rodgers FG, Ang CW, van Koningsveld R, van Belkum A, Gerstenbluth I. Risk factors associated with *Campylobacter jejuni* infections in Curacao, Netherlands Antilles. *J Clin Microbiol*. 2003;41:5588-92.
71. Naumova EN, Jagai JS, Matyas B, DeMaria A, Jr., MacNeill IB, Griffiths JK. Seasonality in six enterically transmitted diseases and ambient temperature. *Epidemiol Infect*. 2007;135:281-92.
72. Kovats RS, Edwards SJ, Hajat S, Armstrong BG, Ebi KL, Menne B. The effect of temperature on food poisoning: a time-series analysis of salmonellosis in ten European countries. *Epidemiol Infect*. 2004;132:443-53.
73. Alter T, Scherer K. Stress response of *Campylobacter* spp. and its role in food processing. *J Vet Med B Infect Dis Vet Public Health*. 2006;53:351-7.
74. Park SF. The physiology of *Campylobacter* species and its relevance to their role as foodborne pathogens. *Int J Food Microbiol*. 2002;74:177-88.
75. Patrick ME, Christiansen LE, Waino M, Ethelberg S, Madsen H, Wegener HC. Effects of climate on incidence of *Campylobacter* spp. in humans and prevalence in broiler flocks in Denmark. *Appl Environ Microbiol*. 2004;70:7474-80.
76. Fleury M, Charron DF, Holt JD, Allen OB, Maarouf AR. A time series analysis of the relationship of ambient temperature and common bacterial enteric infections in two Canadian provinces. *Int J Biometeorol*. 2006;50:385-91.
77. Tam CC, Rodrigues LC, O'Brien SJ, Hajat S. Temperature dependence of reported *Campylobacter* infection in England, 1989-1999. *Epidemiol Infect*. 2006;134:119-25.
78. Schonberg-Norio D, Sarna S, Hanninen ML, Katila ML, Kaukoranta SS, Rautelin H. Strain and host characteristics of *Campylobacter jejuni* infections in Finland. *Clin Microbiol Infect*. 2006;12:754-60.
79. Carrique-Mas J, Andersson Y, Hjertqvist M, Svensson A, Torner A, Giesecke J. Risk factors for domestic sporadic campylobacteriosis among young children in Sweden. *Scand J Infect Dis*. 2005;37:101-10.
80. Obiri-Danso K, Jones K. Distribution and seasonality of microbial indicators and thermophilic *Campylobacters* in two freshwater bathing sites on the River Lune in northwest England. *J Appl Microbiol*. 1999;87:822-32.
81. Neal KR, Slack RC. Diabetes mellitus, anti-secretory drugs and other risk factors for campylobacter gastro-enteritis in adults: a case-control study. *Epidemiol Infect*. 1997;119:307-11.
82. Lighton LL, Kaczmarski EB, Jones DM. A study of risk factors for *Campylobacter* infection in late spring. *Public Health*. 1991;105:199-203.
83. Broman T, Palmgren H, Bergstrom S, Sellin M, Waldenstrom J, Danielsson-Tham ML, Olsen B. *Campylobacter jejuni* in black-headed gulls (*Larus ridibundus*): prevalence, genotypes, and influence on *C. jejuni* epidemiology. *J Clin Microbiol*. 2002;40:4594-602.
84. Meldrum RJ, Griffiths JK, Smith RM, Evans MR. The seasonality of human campylobacter infection and *Campylobacter* isolates from fresh, retail chicken in Wales. *Epidemiol Infect*. 2005;133:49-52.
85. Meldrum RJ, Smith RM, Wilson IG. Three-year surveillance program examining the prevalence of *Campylobacter* and *Salmonella* in whole retail raw chicken. *J Food Prot*. 2006;69:928-31.
86. Neimann J, Engberg J, Molbak K, Wegener HC. A case-control study of risk factors for sporadic *Campylobacter* infections in Denmark. *Epidemiol Infect*. 2003;130:353-66.
87. Batz MB, Doyle MP, Morris G, Jr., Painter J, Singh R, Tauxe RV, Taylor MR, Wong DM. Attributing illness to food. *Emerg Infect Dis*. 2005;11:993-9.

88. Schouls LM, Reulen S, Duim B, Wagenaar JA, Willems RJ, Dingle KE, Colles FM, Van Embden JD. Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and recombination. *J Clin Microbiol*. 2003;41:15-26.
89. Evers EG, Van der Fels-Klerx HJ, Nauta MJ, Schijven JF, Havelaar AH. *Campylobacter* source attribution by exposure assessment. *International Journal of Risk Assessment and Management*. 2007;8:174-190.
90. Hald T, Vose D, Wegener HC, Koupeev T. A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. *Risk Anal*. 2004;24:255-69.
91. Wegener HC, Hald T, Lo Fo Wong D, Madsen M, Korsgaard H, Bager F, Gerner-Smidt P, Molbak K. *Salmonella* control programs in Denmark. *Emerg Infect Dis*. 2003;9:774-80.
92. Havelaar AH, Braunig J, Christiansen K, Cornu M, Hald T, Mangen MJ, Molbak K, Pielaat A, Snary E, Van Pelt W, Velthuis A, Wahlstrom H. Towards an integrated approach in supporting microbiological food safety decisions. *Zoonoses Public Health*. 2007;54:103-17.
93. Nauta MJ, Jacobs-Reitsma WF, Havelaar AH. A risk assessment model for *Campylobacter* in broiler meat. *Risk Anal*. 2007;27:845-61.
94. Mylius SD, Nauta MJ, Havelaar AH. Cross-contamination during food preparation: a mechanistic model applied to chicken-borne *Campylobacter*. *Risk Anal*. 2007;27:803-13.
95. Pebody RG, Ryan MJ, Wall PG. Outbreaks of *Campylobacter* infection: rare events for a common pathogen. *Commun Dis Rep CDR Rev*. 1997;7:R33-7.
96. Frost JA, Gillespie IA, O'Brien SJ. Public health implications of *Campylobacter* outbreaks in England and Wales, 1995-9: epidemiological and microbiological investigations. *Epidemiol Infect*. 2002;128:111-8.
97. Adak GK, Meakins SM, Yip H, Lopman BA, O'Brien SJ. Disease risks from foods, England and Wales, 1996-2000. *Emerg Infect Dis*. 2005;11:365-72.
98. O'Brien SJ, Gillespie IA, Sivanesan MA, Elson R, Hughes C, Adak GK. Publication bias in foodborne outbreaks of infectious intestinal disease and its implications for evidence-based food policy. England and Wales 1992-2003. *Epidemiol Infect*. 2006;134:667-74.
99. Friedman CR, Hoekstra RM, Samuel M, Marcus R, Bender J, Shiferaw B, Reddy S, Ahuja SD, Helfrick DL, Hardnett F, Carter M, Anderson B, Tauxe RV. Risk factors for sporadic *Campylobacter* infection in the United States: A case-control study in FoodNet sites. *Clin Infect Dis*. 2004;38 Suppl 3:S285-96.
100. Rodrigues LC, Cowden JM, Wheeler JG, Sethi D, Wall PG, Cumberland P, Tompkins DS, Hudson MJ, Roberts JA, Roderick PJ. The study of infectious intestinal disease in England: risk factors for cases of infectious intestinal disease with *Campylobacter jejuni* infection. *Epidemiol Infect*. 2001;127:185-93.
101. Nygard K, Andersson Y, Rottingen JA, Svensson A, Lindback J, Kistemann T, Giesecke J. Association between environmental risk factors and *Campylobacter* infections in Sweden. *Epidemiol Infect*. 2004;132:317-25.
102. Kapperud G, Espeland G, Wahl E, Walde A, Herikstad H, Gustavsen S, Tveit I, Natas O, Bevanger L, Digranes A. Factors associated with increased and decreased risk of *Campylobacter* infection: a prospective case-control study in Norway. *Am J Epidemiol*. 2003;158:234-42.

103. Potter RC, Kaneene JB, Hall WN. Risk factors for sporadic *Campylobacter jejuni* infections in rural michigan: a prospective case-control study. *Am J Public Health*. 2003;93:2118-23.
104. Evans MR, Ribeiro CD, Salmon RL. Hazards of healthy living: bottled water and salad vegetables as risk factors for *Campylobacter* infection. *Emerg Infect Dis*. 2003;9:1219-25.
105. Stafford RJ, Schluter P, Kirk M, Wilson A, Unicomb L, Ashbolt R, Gregory J. A multi-centre prospective case-control study of *Campylobacter* infection in persons aged 5 years and older in Australia. *Epidemiol Infect*. 2007;135:978-988.
106. Schonberg-Norio D, Takkinen J, Hanninen ML, Katila ML, Kaukoranta SS, Mattila L, Rautelin H. Swimming and *Campylobacter* infections. *Emerg Infect Dis*. 2004;10:1474-7.
107. Tenkate TD, Stafford RJ. Risk factors for *Campylobacter* infection in infants and young children: a matched case-control study. *Epidemiol Infect*. 2001;127:399-404.
108. Cameron S, Ried K, Worsley A, Topping D. Consumption of foods by young children with diagnosed *Campylobacter* infection - a pilot case-control study. *Public Health Nutr*. 2004;7:85-9.
109. Bellido-Blasco JB, Gonzalez-Cano JM, Galiano-Arlandis JV, Herrero-Carot C, Tirado-Balaguer MD, Arnedo-Pena A, Safont-Adsuara L, Romeu-Garcia MA. [Risk factors for the occurrence of sporadic *Campylobacter*, *Salmonella* and rotavirus diarrhea in preschool children]. *An Pediatr (Barc)*. 2007;66:367-74.
110. Fullerton KE, Ingram LA, Jones TF, Anderson BJ, McCarthy PV, Hurd S, Shiferaw B, Vugia D, Haubert N, Hayes T, Wedel S, Scallan E, Henao O, Angulo FJ. Sporadic *Campylobacter* infection in infants: a population-based surveillance case-control study. *Pediatr Infect Dis J*. 2007;26:19-24.
111. Sethi D, Wheeler JG, Cowden JM, Rodrigues LC, Sockett PN, Roberts JA, Cumberland P, Tompkins DS, Wall PG, Hudson MJ, Roderick PJ. A study of infectious intestinal disease in England: plan and methods of data collection. *Commun Dis Public Health*. 1999;2:101-7.
112. Stern NJ, Hiett KL, Alfredsson GA, Kristinsson KG, Reiersen J, Hardardottir H, Briem H, Gunnarsson E, Georgsson F, Lowman R, Berndtson E, Lammerding AM, Paoli GM, Musgrove MT. *Campylobacter* spp. in Icelandic poultry operations and human disease. *Epidemiol Infect*. 2003;130:23-32.
113. Hald T, Lo Fo Wong DM, Aarestrup FM. The attribution of human infections with antimicrobial resistant *Salmonella* bacteria in Denmark to sources of animal origin. *Foodborne Pathog Dis*. 2007;4:313-26.
114. Dingle KE, Colles FM, Ure R, Wagenaar JA, Duim B, Bolton FJ, Fox AJ, Wareing DR, Maiden MC. Molecular characterization of *Campylobacter jejuni* clones: a basis for epidemiologic investigation. *Emerg Infect Dis*. 2002;8:949-55.
115. Nachamkin I, Engberg J, Møller Aarestrup F. Diagnosis and Antimicrobial Susceptibility of *Campylobacter* spp. In: Nachamkin I, Blaser MJ, eds. *Campylobacter*. Washington D.C.: ASM; 2000.
116. Whyte P, McGill K, Cowley D, Madden RH, Moran L, Scates P, Carroll C, O'Leary A, Fanning S, Collins JD, McNamara E, Moore JE, Cormican M. Occurrence of *Campylobacter* in retail foods in Ireland. *Int J Food Microbiol*. 2004;95:111-8.
117. On SL. Taxonomy of *Campylobacter*, *Arcobacter*, *Helicobacter* and related bacteria: current status, future prospects and immediate concerns. *J Appl Microbiol*. 2001;90:1S-15S.
118. Fouts DE, Mongodin EF, Mandrell RE, Miller WG, Rasko DA, Ravel J, Brinkac LM, DeBoy RT, Parker CT, Daugherty SC, Dodson RJ, Durkin AS, Madupu R, Sullivan

SA, Shetty JU, Ayodeji MA, Shvartsbeyn A, Schatz MC, Badger JH, Fraser CM, Nelson KE. Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. *PLoS Biol*. 2005;3:e15.
119. Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D, Chillingworth T, Davies RM, Feltwell T, Holroyd S, Jagels K, Karlyshev AV, Moule S, Pallen MJ, Penn CW, Quail MA, Rajandream MA, Rutherford KM, van Vliet AH, Whitehead S, Barrell BG. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature*. 2000;403:665-8.
120. Hofreuter D, Tsai J, Watson RO, Novik V, Altman B, Benitez M, Clark C, Perbost C, Jarvie T, Du L, Galan JE. Unique features of a highly pathogenic *Campylobacter jejuni* strain. *Infect Immun*. 2006;74:4694-707.
121. Suerbaum S, Lohrengel M, Sonnevend A, Ruberg F, Kist M. Allelic diversity and recombination in *Campylobacter jejuni*. *J Bacteriol*. 2001;183:2553-9.
122. Young KT, Davis LM, Dirita VJ. *Campylobacter jejuni*: molecular biology and pathogenesis. *Nat Rev Microbiol*. 2007;5:665-79.
123. Miller WG, On SL, Wang G, Fontanoz S, Lastovica AJ, Mandrell RE. Extended multilocus sequence typing system for *Campylobacter coli*, *C. lari*, *C. upsaliensis*, and *C. helveticus*. *J Clin Microbiol*. 2005;43:2315-29.
124. Stanley J, Burnens AP, Linton D, On SL, Costas M, Owen RJ. *Campylobacter helveticus* sp. nov., a new thermophilic species from domestic animals: characterization, and cloning of a species-specific DNA probe. *J Gen Microbiol*. 1992;138:2293-303.
125. Endtz HP, Vliegenthart JS, Vandamme P, Weverink HW, van den Braak NP, Verbrugh HA, van Belkum A. Genotypic diversity of *Campylobacter lari* isolated from mussels and oysters in The Netherlands. *Int J Food Microbiol*. 1997;34:79-88.
126. Waldenstrom J, Broman T, Carlsson I, Hasselquist D, Achterberg RP, Wagenaar JA, Olsen B. Prevalence of *Campylobacter jejuni*, *Campylobacter lari*, and *Campylobacter coli* in different ecological guilds and taxa of migrating birds. *Appl Environ Microbiol*. 2002;68:5911-7.
127. Cleaveland S, Laurenson MK, Taylor LH. Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philos Trans R Soc Lond B Biol Sci*. 2001;356:991-9.
128. Yogasundram K, Shane SM, Harrington KS. Prevalence of *Campylobacter jejuni* in selected domestic and wild birds in Louisiana. *Avian Dis*. 1989;33:664-7.
129. Rosef O, Kapperud G, Lauwers S, Gondrosen B. Serotyping of *Campylobacter jejuni*, *Campylobacter coli*, and *Campylobacter laridis* from domestic and wild animals. *Appl Environ Microbiol*. 1985;49:1507-10.
130. Stanley K, Jones K. Cattle and sheep farms as reservoirs of *Campylobacter*. *J Appl Microbiol*. 2003;94 Suppl:104S-113S.
131. Humphrey TJ, Henley A, Lanning DG. The colonization of broiler chickens with *Campylobacter jejuni*: some epidemiological investigations. *Epidemiol Infect*. 1993;110:601-7.
132. Jacobs-Reitsma WF, Bolder NM, Mulder RW. Cecal carriage of *Campylobacter* and *Salmonella* in Dutch broiler flocks at slaughter: a one-year study. *Poult Sci*. 1994;73:1260-6.
133. Bang DD, Nielsen EM, Knudsen K, Madsen M. A one-year study of *Campylobacter* carriage by individual Danish broiler chickens as the basis for selection of *Campylobacter* spp. strains for a chicken infection model. *Epidemiol Infect*. 2003;130:323-33.

134. Cuomo A, Dipineto L, Santaniello A, Matteoli G, Sarli T, Vecchia DD, Fioretti A, Menna LF. Detection of thermotolerant *Campylobacter* in ostriches (*Struthio camelus*) in Italy. *Vet J*. 2007;174:439-41.
135. Hald B, Pedersen K, Waino M, Jorgensen JC, Madsen M. Longitudinal study of the excretion patterns of thermophilic *Campylobacter* spp. in young pet dogs in Denmark. *J Clin Microbiol*. 2004;42:2003-12.
136. Luechtefeld NW, Cambre RC, Wang WL. Isolation of *Campylobacter fetus* subsp *jejuni* from zoo animals. *J Am Vet Med Assoc*. 1981;179:1119-22.
137. Pacha RE, Clark GW, Williams EA. Occurrence of *Campylobacter jejuni* and *Giardia* species in muskrat (*Ondatra zibethica*). *Appl Environ Microbiol*. 1985;50:177-8.
138. French N, Barrigas M, Brown P, Ribiero P, Williams N, Leatherbarrow H, Birtles R, Bolton E, Fearnhead P, Fox A. Spatial epidemiology and natural population structure of *Campylobacter jejuni* colonizing a farmland ecosystem. *Environ Microbiol*. 2005;7:1116-26.
139. Rosef O, Gondrosen B, Kapperud G, Underdal B. Isolation and characterization of *Campylobacter jejuni* and *Campylobacter coli* from domestic and wild mammals in Norway. *Appl Environ Microbiol*. 1983;46:855-9.
140. Axelsson-Olsson D, Waldenstrom J, Broman T, Olsen B, Holmberg M. Protozoan *Acanthamoeba polyphaga* as a potential reservoir for *Campylobacter jejuni*. *Appl Environ Microbiol*. 2005;71:987-92.
141. Snelling WJ, Stern NJ, Lowery CJ, Moore JE, Gibbons E, Baker C, Dooley JS. Colonization of broilers by *Campylobacter jejuni* internalized within *Acanthamoeba castellanii*. *Arch Microbiol*. 2007.
142. Kapperud G, Rosef O. Avian wildlife reservoir of *Campylobacter fetus* subsp. *jejuni*, *Yersinia* spp., and *Salmonella* spp. in Norway. *Appl Environ Microbiol*. 1983;45:375-80.
143. Oporto B, Esteban JI, Aduriz G, Juste RA, Hurtado A. Prevalence and strain diversity of thermophilic *Campylobacters* in cattle, sheep and swine farms. *J Appl Microbiol*. 2007;103:977-84.
144. Englen MD, Hill AE, Dargatz DA, Ladely SR, Fedorka-Cray PJ. Prevalence and antimicrobial resistance of *Campylobacter* in US dairy cattle. *J Appl Microbiol*. 2007;102:1570-7.
145. Nielsen EM. Occurrence and strain diversity of thermophilic *Campylobacters* in cattle of different age groups in dairy herds. *Lett Appl Microbiol*. 2002;35:85-9.
146. Stanley KN, Wallace JS, Currie JE, Diggle PJ, Jones K. The seasonal variation of thermophilic *Campylobacters* in beef cattle, dairy cattle and calves. *J Appl Microbiol*. 1998;85:472-80.
147. Acik MN, Cetinkaya B. Heterogeneity of *Campylobacter jejuni* and *Campylobacter coli* strains from healthy sheep. *Vet Microbiol*. 2006;115:370-5.
148. Ertas HB, Ozbey G, Kilic A, Muz A. Isolation of *Campylobacter jejuni* and *Campylobacter coli* from the gall bladder samples of sheep and identification by polymerase chain reaction. *J Vet Med B Infect Dis Vet Public Health*. 2003;50:294-7.
149. Corry JE, Atabay HI. Poultry as a source of *Campylobacter* and related organisms. *J Appl Microbiol*. 2001;90:96S-114S.
150. Boonmar S, Yingsakmongkon S, Songserm T, Hanhaboon P, Passadurak W. Detection of *Campylobacter* in duck using standard culture method and multiplex polymerase chain reaction. *Southeast Asian J Trop Med Public Health*. 2007;38:728-31.
151. Prescott JF, Bruin-Mosch CW. Carriage of *Campylobacter jejuni* in healthy and diarrheic animals. *Am J Vet Res*. 1981;42:164-5.

152. Luechtefeld NA, Blaser MJ, Reller LB, Wang WL. Isolation of *Campylobacter fetus* subsp. *jejuni* from migratory waterfowl. *J Clin Microbiol*. 1980;12:406-8.
153. Rivoal K, Denis M, Salvat G, Colin P, Ermel G. Molecular characterization of the diversity of *Campylobacter* spp. isolates collected from a poultry slaughterhouse: analysis of cross-contamination. *Lett Appl Microbiol*. 1999;29:370-4.
154. Hiett KL, Stern NJ, Fedorka-Cray P, Cox NA, Musgrove MT, Ladely S. Molecular subtype analyses of *Campylobacter* spp. from Arkansas and California poultry operations. *Appl Environ Microbiol*. 2002;68:6220-36.
155. Moore J, Caldwell P, Millar B. Molecular detection of *Campylobacter* spp. in drinking, recreational and environmental water supplies. *Int J Hyg Environ Health*. 2001;204:185-9.
156. Evans SJ, Sayers AR. A longitudinal study of *Campylobacter* infection of broiler flocks in Great Britain. *Prev Vet Med*. 2000;46:209-23.
157. Stanley KN, Wallace JS, Currie JE, Diggle PJ, Jones K. Seasonal variation of thermophilic *Campylobacters* in lambs at slaughter. *J Appl Microbiol*. 1998;84:1111-6.
158. Allen VM, Bull SA, Corry JE, Domingue G, Jorgensen F, Frost JA, Whyte R, Gonzalez A, Elviss N, Humphrey TJ. *Campylobacter* spp. contamination of chicken carcasses during processing in relation to flock colonisation. *Int J Food Microbiol*. 2007;113:54-61.
159. Wong TL, Hollis L, Cornelius A, Nicol C, Cook R, Hudson JA. Prevalence, numbers, and subtypes of *Campylobacter jejuni* and *Campylobacter coli* in uncooked retail meat samples. *J Food Prot*. 2007;70:566-73.
160. Zhao C, Ge B, De Villena J, Sudler R, Yeh E, Zhao S, White DG, Wagner D, Meng J. Prevalence of *Campylobacter* spp., *Escherichia coli*, and *Salmonella* serovars in retail chicken, turkey, pork, and beef from the Greater Washington, D.C., area. *Appl Environ Microbiol*. 2001;67:5431-6.
161. Turnbull PCB, Rose P. *Campylobacter jejuni* and *Salmonella* in Raw Red Meats. A Public Health Laboratory Service Survey. *The Journal of Hygiene*. 1982;88:29-37.
162. Evers EG, Van der Fels-Klerx HJ, Nauta MJ, Schijven JF, Havelaar AH. Het relatieve belang van *Campylobacter* transmissieroutes op basis van blootstellingsschatting. In. Bilthoven: RIVM; 2004.
163. Skelly C, Weinstein P. Pathogen survival trajectories: an eco-environmental approach to the modeling of human campylobacteriosis ecology. *Environ Health Perspect*. 2003;111:19-28.
164. Bolton FJ, Surman SB, Martin K, Wareing DR, Humphrey TJ. Presence of *Campylobacter* and *Salmonella* in sand from bathing beaches. *Epidemiol Infect*. 1999;122:7-13.
165. Shane SM, Montrose MS, Harrington KS. Transmission of *Campylobacter jejuni* by the housefly (*Musca domestica*). *Avian Dis*. 1985;29:384-91.
166. Hald B, Skovgard H, Bang DD, Pedersen K, Dybdahl J, Jespersen JB, Madsen M. Flies and *Campylobacter* infection of broiler flocks. *Emerg Infect Dis*. 2004;10:1490-2.
167. Aarts HJ, van Lith LA, Jacobs-Reitsma WF. Discrepancy between Penner serotyping and polymerase chain reaction fingerprinting of *Campylobacter* isolated from poultry and other animal sources. *Lett Appl Microbiol*. 1995;20:371-4.
168. Strother KO, Steelman CD, Gbur EE. Reservoir competence of lesser mealworm (Coleoptera: *Tenebrionidae*) for *Campylobacter jejuni* (Campylobacterales: *Campylobacteraceae*). *J Med Entomol*. 2005;42:42-7.

169. Hald B, Sommer HM, Skovgard H. Use of fly screens to reduce *Campylobacter* spp. introduction in broiler houses. *Emerg Infect Dis*. 2007;13:1951-3.
170. Meanger JD, Marshall RB. Seasonal prevalence of thermophilic *Campylobacter* infections in dairy cattle and a study of infection of sheep. *N Z Vet J*. 1989;37:18-20.
171. Wedderkopp A, Gradel KO, Jorgensen JC, Madsen M. Pre-harvest surveillance of *Campylobacter* and *Salmonella* in Danish broiler flocks: a 2-year study. *Int J Food Microbiol*. 2001;68:53-9.
172. Kapperud G, Skjerve E, Vik L, Hauge K, Lysaker A, Aalmen I, Ostroff SM, Potter M. Epidemiological investigation of risk factors for *Campylobacter* colonization in Norwegian broiler flocks. *Epidemiol Infect*. 1993;111:245-55.
173. Wilson IG. *Salmonella* and *Campylobacter* contamination of raw retail chickens from different producers: a six year survey. *Epidemiol Infect*. 2002;129:635-45.
174. Wallace JS, Stanley KN, Currie JE, Diggle PJ, Jones K. Seasonality of thermophilic *Campylobacter* populations in chickens. *J Appl Microbiol*. 1997;82:219-24.
175. Logue CM, Sherwood JS, Elijah LM, Olah PA, Dockter MR. The incidence of *Campylobacter* spp. on processed turkey from processing plants in the midwestern United States. *J Appl Microbiol*. 2003;95:234-41.
176. Abulreesh HH, Paget TA, Goulder R. *Campylobacter* in waterfowl and aquatic environments: incidence and methods of detection. *Environ Sci Technol*. 2006;40:7122-31.
177. Eyles R, Niyogi D, Townsend C, Benwell G, Weinstein P. Spatial and temporal patterns of *Campylobacter* contamination underlying public health risk in the Taieri River, New Zealand. *J Environ Qual*. 2003;32:1820-8.
178. Jones K, Betaieb M, Telford DR. Seasonal variation of thermophilic *Campylobacters* in sewage sludge. *J Appl Bacteriol*. 1990;69:185-9.
179. Stampi S, De Luca G, Varoli O, Zanetti F. Occurrence, removal and seasonal variation of thermophilic *Campylobacters* and *Arcobacter* in sewage sludge. *Zentralbl Hyg Umweltmed*. 1999;202:19-27.
180. Jones K, Betaieb M, Telford DR. Correlation between environmental monitoring of thermophilic *Campylobacters* in sewage effluent and the incidence of *Campylobacter* infection in the community. *J Appl Bacteriol*. 1990;69:235-40.
181. Lior H, Woodward DL, Edgar JA, Laroche LJ, Gill P. Serotyping of *Campylobacter jejuni* by slide agglutination based on heat-labile antigenic factors. *J Clin Microbiol*. 1982;15:761-8.
182. Penner JL, Hennessy JN. Passive hemagglutination technique for serotyping *Campylobacter fetus* subsp. *jejuni* on the basis of soluble heat-stable antigens. *J Clin Microbiol*. 1980;12:732-7.
183. Grajewski BA, Kusek JW, Gelfand HM. Development of a bacteriophage typing system for *Campylobacter jejuni* and *Campylobacter coli*. *J Clin Microbiol*. 1985;22:13-8.
184. Patton CM, Wachsmuth IK, Evins GM, Kiehlbauch JA, Plikaytis BD, Troup N, Tompkins L, Lior H. Evaluation of 10 methods to distinguish epidemic-associated *Campylobacter* strains. *J Clin Microbiol*. 1991;29:680-8.
185. Frost JA, Kramer JM, Gillanders SA. Phage typing of *Campylobacter jejuni* and *Campylobacter coli* and its use as an adjunct to serotyping. *Epidemiol Infect*. 1999;123:47-55.
186. Wassenaar TM, Newell DG. Genotyping of *Campylobacter* spp. *Appl Environ Microbiol*. 2000;66:1-9.

187. Frost JA, Oza AN, Thwaites RT, Rowe B. Serotyping scheme for *Campylobacter jejuni* and *Campylobacter coli* based on direct agglutination of heat-stable antigens. *J Clin Microbiol*. 1998;36:335-9.
188. Hanninen ML, Hakkinen M, Rautelin H. Stability of related human and chicken *Campylobacter jejuni* genotypes after passage through chick intestine studied by pulsed-field gel electrophoresis. *Appl Environ Microbiol*. 1999;65:2272-5.
189. Hopkins KL, Desai M, Frost JA, Stanley J, Logan JM. Fluorescent amplified fragment length polymorphism genotyping of *Campylobacter jejuni* and *Campylobacter coli* strains and its relationship with host specificity, serotyping, and phage typing. *J Clin Microbiol*. 2004;42:229-35.
190. Lastovica AJ, Goddard EA, Argent AC. Guillain-Barré syndrome in South Africa associated with *Campylobacter jejuni* O:41 strains. *J Infect Dis*. 1997;176 Suppl 2:S139-43.
191. Mishu B, Blaser MJ. Role of infection due to *Campylobacter jejuni* in the initiation of Guillain-Barré syndrome. *Clin Infect Dis*. 1993;17:104-8.
192. Petersen L, Nielsen EM, Engberg J, On SL, Dietz HH. Comparison of genotypes and serotypes of *Campylobacter jejuni* isolated from Danish wild mammals and birds and from broiler flocks and humans. *Appl Environ Microbiol*. 2001;67:3115-21.
193. Devane ML, Nicol C, Ball A, Klena JD, Scholes P, Hudson JA, Baker MG, Gilpin BJ, Garrett N, Savill MG. The occurrence of *Campylobacter* subtypes in environmental reservoirs and potential transmission routes. *J Appl Microbiol*. 2005;98:980-90.
194. Garrett N, Devane ML, Hudson JA, Nicol C, Ball A, Klena JD, Scholes P, Baker MG, Gilpin BJ, Savill MG. Statistical comparison of *Campylobacter jejuni* subtypes from human cases and environmental sources. *J Appl Microbiol*. 2007;103:2113-21.
195. Pearson AD, Greenwood MH, Feltham RK, Healing TD, Donaldson J, Jones DM, Colwell RR. Microbial ecology of *Campylobacter jejuni* in a United Kingdom chicken supply chain: intermittent common source, vertical transmission, and amplification by flock propagation. *Appl Environ Microbiol*. 1996;62:4614-20.
196. Yan W, Chang N, Taylor DE. Pulsed-field gel electrophoresis of *Campylobacter jejuni* and *Campylobacter coli* genomic DNA and its epidemiologic application. *J Infect Dis*. 1991;163:1068-72.
197. Duim B, Wassenaar TM, Rigter A, Wagenaar J. High-resolution genotyping of *Campylobacter* strains isolated from poultry and humans with amplified fragment length polymorphism fingerprinting. *Appl Environ Microbiol*. 1999;65:2369-75.
198. Ragimbeau C, Salvat G, Colin P, Ermel G. Development of a multiplex PCR gene fingerprinting method using *gyrA* and *pflA* polymorphisms to identify genotypic relatedness within *Campylobacter jejuni* species. *J Appl Microbiol*. 1998;85:829-38.
199. Champion OL, Gaunt MW, Gundogdu O, Elmi A, Witney AA, Hinds J, Dorrell N, Wren BW. Comparative phylogenomics of the food-borne pathogen *Campylobacter jejuni* reveals genetic markers predictive of infection source. *Proc Natl Acad Sci U S A*. 2005;102:16043-8.
200. Ribot EM, Fitzgerald C, Kubota K, Swaminathan B, Barrett TJ. Rapid pulsed-field gel electrophoresis protocol for subtyping of *Campylobacter jejuni*. *J Clin Microbiol*. 2001;39:1889-94.
201. Gerner-Smidt P, Hise K, Kincaid J, Hunter S, Rolando S, Hyytia-Trees E, Ribot EM, Swaminathan B. PulseNet USA: a five-year update. *Foodborne Pathog Dis*. 2006;3:9-19.
202. Scott AE, Timms AR, Connerton PL, Loc Carrillo C, Adzfa Radzum K, Connerton IF. Genome dynamics of *Campylobacter jejuni* in response to bacteriophage predation. *PLoS Pathog*. 2007;3:e119.

203. Keller J, Wieland B, Wittwer M, Stephan R, Perreten V. Distribution and genetic variability among *Campylobacter* spp. isolates from different animal species and humans in Switzerland. *Zoonoses Public Health*. 2007;54:2-7.
204. Sails AD, Swaminathan B, Fields PI. Utility of multilocus sequence typing as an epidemiological tool for investigation of outbreaks of gastroenteritis caused by *Campylobacter jejuni*. *J Clin Microbiol*. 2003;41:4733-9.
205. Fitzgerald C, Helsel LO, Nicholson MA, Olsen SJ, Swerdlow DL, Flahart R, Sexton J, Fields PI. Evaluation of methods for subtyping *Campylobacter jejuni* during an outbreak involving a food handler. *J Clin Microbiol*. 2001;39:2386-90.
206. Kuusi M, Nuorti JP, Hanninen ML, Koskela M, Jussila V, Kela E, Miettinen I, Ruutu P. A large outbreak of campylobacteriosis associated with a municipal water supply in Finland. *Epidemiol Infect*. 2005;133:593-601.
207. Mellmann A, Mosters J, Bartelt E, Roggentin P, Ammon A, Friedrich AW, Karch H, Harmsen D. Sequence-Based Typing of *flaB* Is a More Stable Screening Tool Than Typing of *flaA* for Monitoring of *Campylobacter* Populations. *J Clin Microbiol*. 2004;42:4840-2.
208. Suzuki Y, Ishihara M, Saito M, Ishikawa N, Yokochi T. Discrimination by means of pulsed-field gel electrophoresis between strains of *Campylobacter jejuni* Lior type 4 derived from sporadic cases and from outbreaks of infection. *J Infect*. 1994;29:183-7.
209. Hedberg CW, Smith KE, Besser JM, Boxrud DJ, Hennessy TW, Bender JB, Anderson FA, Osterholm MT. Limitations of pulsed-field gel electrophoresis for the routine surveillance of *Campylobacter* infections. *J Infect Dis*. 2001;184:242-4.
210. Michaud S, Menard S, Gaudreau C, Arbeit RD. Comparison of SmaI-defined genotypes of *Campylobacter jejuni* examined by KpnI: a population-based study. *J Med Microbiol*. 2001;50:1075-81.
211. Michaud S, Menard S, Arbeit RD. Role of real-time molecular typing in the surveillance of *Campylobacter* enteritis and comparison of pulsed-field gel electrophoresis profiles from chicken and human isolates. *J Clin Microbiol*. 2005;43:1105-11.
212. Hook H, Ekegren MB, Ericsson H, Vagsholm I, Danielsson-Tham ML. Genetic and epidemiological relationships among *Campylobacter* isolates from humans. *Scand J Infect Dis*. 2004;36:435-42.
213. Broman T, Waldenstrom J, Dahlgren D, Carlsson I, Eliasson I, Olsen B. Diversities and similarities in PFGE profiles of *Campylobacter jejuni* isolated from migrating birds and humans. *J Appl Microbiol*. 2004;96:834-43.
214. Hanninen ML, Perko-Makela P, Pitkala A, Rautelin H. A three-year study of *Campylobacter jejuni* genotypes in humans with domestically acquired infections and in chicken samples from the Helsinki area. *J Clin Microbiol*. 2000;38:1998-2000.
215. Nadeau E, Messier S, Quessy S. Prevalence and comparison of genetic profiles of *Campylobacter* strains isolated from poultry and sporadic cases of campylobacteriosis in humans. *J Food Prot*. 2002;65:73-8.
216. Siemer BL, Harrington CS, Nielsen EM, Borck B, Nielsen NL, Engberg J, On SL. Genetic relatedness among *Campylobacter jejuni* serotyped isolates of diverse origin as determined by numerical analysis of amplified fragment length polymorphism (AFLP) profiles. *J Appl Microbiol*. 2004;96:795-802.
217. Harrington CS, Thomson-Carter FM, Carter PE. Evidence for recombination in the flagellin locus of *Campylobacter jejuni*: implications for the flagellin gene typing scheme. *J Clin Microbiol*. 1997;35:2386-92.

218. Dingle KE, Colles FM, Falush D, Maiden MC. Sequence Typing and Comparison of Population Biology of *Campylobacter coli* and *Campylobacter jejuni*. *J Clin Microbiol*. 2005;43:340-7.
219. Gressmann H, Linz B, Ghai R, Pleissner KP, Schlapbach R, Yamaoka Y, Kraft C, Suerbaum S, Meyer TF, Achtman M. Gain and Loss of Multiple Genes During the Evolution of *Helicobacter pylori*. *PLoS Genet*. 2005;1:e43.
220. Dingle KE, Colles FM, Wareing DR, Ure R, Fox AJ, Bolton FE, Bootsma HJ, Willems RJ, Urwin R, Maiden MC. Multilocus sequence typing system for *Campylobacter jejuni*. *J Clin Microbiol*. 2001;39:14-23.
221. Meinersmann RJ, Helsel LO, Fields PI, Hiett KL. Discrimination of *Campylobacter jejuni* isolates by fla gene sequencing. *J Clin Microbiol*. 1997;35:2810-4.
222. Zhang Q, Meitzler JC, Huang S, Morishita T. Sequence polymorphism, predicted secondary structures, and surface-exposed conformational epitopes of *Campylobacter* major outer membrane protein. *Infect Immun*. 2000;68:5679-89.
223. Martin IM, Ison CA, Aanensen DM, Fenton KA, Spratt BG. Rapid sequence-based identification of gonococcal transmission clusters in a large metropolitan area. *J Infect Dis*. 2004;189:1497-505.
224. Feavers IM, Gray SJ, Urwin R, Russell JE, Bygraves JA, Kaczmarski EB, Maiden MC. Multilocus sequence typing and antigen gene sequencing in the investigation of a meningococcal disease outbreak. *J Clin Microbiol*. 1999;37:3883-7.
225. Meyer M, Stenzel U, Hofreiter M. Parallel tagged sequencing on the 454 platform. *Nat Protoc*. 2008;3:267-78.
226. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A*. 1998;95:3140-5.
227. Maiden MC. Multilocus sequence typing of bacteria. *Annu Rev Microbiol*. 2006;60:561-88.
228. Smith JM, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? *Proc Natl Acad Sci U S A*. 1993;90:4384-8.
229. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 2003;164:1567-87.
230. Didelot X, Falush D. Inference of bacterial microevolution using multilocus sequence data. *Genetics*. 2007;175:1251-66.
231. Manning G, Dowson CG, Bagnall MC, Ahmed IH, West M, Newell DG. Multilocus sequence typing for comparison of veterinary and human isolates of *Campylobacter jejuni*. *Appl Environ Microbiol*. 2003;69:6370-9.
232. Meinersmann RJ, Dingle KE, Maiden MC. Genetic Exchange among *Campylobacter* Species. *Genome Letters*. 2003;2:48 - 52.
233. Wareing DR, Ure R, Colles FM, Bolton FJ, Fox AJ, Maiden MC, Dingle KE. Reference isolates for the clonal complexes of *Campylobacter jejuni*. *Lett Appl Microbiol*. 2003;36:106-10.
234. Colles FM, Jones K, Harding RM, Maiden MC. Genetic diversity of *Campylobacter jejuni* isolates from farm animals and the farm environment. *Appl Environ Microbiol*. 2003;69:7409-13.
235. Sails AD, Swaminathan B, Fields PI. Clonal complexes of *Campylobacter jejuni* identified by multilocus sequence typing correlate with strain associations identified by multilocus enzyme electrophoresis. *J Clin Microbiol*. 2003;41:4058-67.

236. Best EL, Fox AJ, Frost JA, Bolton FJ. Identification of *Campylobacter jejuni* multilocus sequence type ST-21 clonal complex by single-nucleotide polymorphism analysis. *J Clin Microbiol*. 2004;42:2836-9.
237. Clark CG, Bryden L, Cuff WR, Johnson PL, Jamieson F, Ciebin B, Wang G. Use of the oxford multilocus sequence typing protocol and sequencing of the flagellin short variable region to characterize isolates from a large outbreak of waterborne *Campylobacter* sp. strains in Walkerton, Ontario, Canada. *J Clin Microbiol*. 2005;43:2080-91.
238. Fitch BR, Sachen KL, Wilder SR, Burg MA, Lacher DW, Khalife WT, Whittam TS, Young VB. Genetic diversity of *Campylobacter* sp. isolates from retail chicken products and humans with gastroenteritis in Central Michigan. *J Clin Microbiol*. 2005;43:4221-4.
239. Miller WG, Englen MD, Kathariou S, Wesley IV, Wang G, Pittenger-Alley L, Siletz RM, Muraoka W, Fedorka-Cray PJ, Mandrell RE. Identification of host-associated alleles by multilocus sequence typing of *Campylobacter coli* strains from food animals. *Microbiology*. 2006;152:245-55.
240. Kinana AD, Cardinale E, Tall F, Bahsoun I, Sire JM, Garin B, Breurec S, Boye CS, Perrier-Gros-Claude JD. Genetic diversity and quinolone resistance in *Campylobacter jejuni* isolates from poultry in Senegal. *Appl Environ Microbiol*. 2006;72:3309-13.
241. Djordjevic SP, Unicomb LE, Adamson PJ, Mickan L, Rios R. Clonal complexes of *Campylobacter jejuni* identified by multilocus sequence typing are reliably predicted by restriction fragment length polymorphism analyses of the *flaA* gene. *J Clin Microbiol*. 2007;45:102-8.
242. D'Lima C B, Miller WG, Mandrell RE, Wright SL, Siletzky RM, Carver DK, Kathariou S. Clonal Population Structure and Specific Genotypes of Multidrug Resistant *Campylobacter coli* from Turkeys. *Appl Environ Microbiol*. 2007.
243. Mickan L, Doyle R, Valcanis M, Dingle KE, Unicomb L, Lanser J. Multilocus sequence typing of *Campylobacter jejuni* isolates from New South Wales, Australia. *J Appl Microbiol*. 2007;102:144-52.
244. Sopwith W, Birtles A, Matthews M, Fox A, Gee S, Painter M, Regan M, Syed Q, Bolton E. *Campylobacter jejuni* multilocus sequence types in humans, northwest England, 2003-2004. *Emerg Infect Dis*. 2006;12:1500-7.
245. Kinana AD, Cardinale E, Bahsoun I, Tall F, Sire JM, Breurec S, Garin B, Saad-Bouh Boye C, Perrier-Gros-Claude JD. Campylobacter coli isolates derived from chickens in Senegal: Diversity, genetic exchange with *Campylobacter jejuni* and quinolone resistance. *Res Microbiol*. 2007.
246. O'Reilly LC, Inglis TJ, Unicomb L. Australian multicentre comparison of subtyping methods for the investigation of *Campylobacter infection*. *Epidemiol Infect*. 2006;134:768-79.
247. McCarthy ND, Colles FM, Dingle KE, Bagnall MC, Manning G, Maiden MC, Falush D. Host-associated genetic import in *Campylobacter jejuni*. *Emerg Infect Dis*. 2007;13:267-72.
248. Litrup E, Torpdahl M, Nielsen EM. Multilocus sequence typing performed on *Campylobacter coli* isolates from humans, broilers, pigs and cattle originating in Denmark. *J Appl Microbiol*. 2007;103:210-8.
249. Duim B, Godschalk PC, van den Braak N, Dingle KE, Dijkstra JR, Leyde E, van der Plas J, Colles FM, Endtz HP, Wagenaar JA, Maiden MC, van Belkum A. Molecular evidence for dissemination of unique *Campylobacter jejuni* clones in Curacao, Netherlands Antilles. *J Clin Microbiol*. 2003;41:5593-7.

250. Hepworth PJ, Leatherbarrow H, Hart CA, Winstanley C. Use of suppression subtractive hybridisation to extend our knowledge of genome diversity in *Campylobacter jejuni*. *BMC Genomics*. 2007;8:110.
251. Clark CG, Beeston A, Bryden L, Wang G, Barton C, Cuff W, Gilmour MW, Ng LK. Phylogenetic relationships of *Campylobacter jejuni* based on *porA* sequences. *Can J Microbiol*. 2007;53:27-38.
252. Parker CT, Miller WG, Horn ST, Lastovica AJ. Common genomic features of *Campylobacter jejuni* subsp. *doylei* strains distinguish them from *C. jejuni* subsp. *jejuni*. *BMC Microbiol*. 2007;7:50.
253. Price EP, Thiruvenkataswamy V, Mickan L, Unicomb L, Rios RE, Huygens F, Giffard PM. Genotyping of *Campylobacter jejuni* using seven single-nucleotide polymorphisms in combination with *flaA* short variable region sequencing. *J Med Microbiol*. 2006;55:1061-70.
254. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol*. 2004;186:1518-30.
255. Thakur S, Morrow WE, Funk JA, Bahnson PB, Gebreyes WA. Molecular epidemiologic investigation of *Campylobacter coli* in swine production systems, using multilocus sequence typing. *Appl Environ Microbiol*. 2006;72:5666-9.
256. Jolley KA, Chan MS, Maiden MC. mlstdbNet - distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics*. 2004;5:86.
257. Dingle KE, McCarthy ND, Cody AJ, Peto TE, Maiden MC. Extended sequence typing of *Campylobacter* spp., United Kingdom. *Emerg Infect Dis*. 2008;14:1620-2.
258. Best EL, Fox AJ, Frost JA, Bolton FJ. Real-time single-nucleotide polymorphism profiling using Taqman technology for rapid recognition of *Campylobacter jejuni* clonal complexes. *J Med Microbiol*. 2005;54:919-25.
259. Fearnhead P, Smith NG, Barrigas M, Fox A, French N. Analysis of Recombination in *Campylobacter jejuni* from MLST Population Data. *J Mol Evol*. 2005;61:333-40.
260. Dingle KE, Van Den Braak N, Colles FM, Price LJ, Woodward DL, Rodgers FG, Endtz HP, Van Belkum A, Maiden MC. Sequence typing confirms that *Campylobacter jejuni* strains associated with Guillain-Barre and Miller-Fisher syndromes are of diverse genetic lineage, serotype, and flagella type. *J Clin Microbiol*. 2001;39:3346-9.
261. Engberg J, Nachamkin I, Fussing V, McKhann GM, Griffin JW, Piffaretti JC, Nielsen EM, Gerner-Smidt P. Absence of clonality of *Campylobacter jejuni* in serotypes other than HS:19 associated with Guillain-Barré syndrome and gastroenteritis. *J Infect Dis*. 2001;184:215-20.
262. Chan K, Miller WG, Mandrell RE, Kathariou S. The absence of intervening sequences in 23S rRNA genes of *Campylobacter coli* isolates from Turkeys is a unique attribute of a cluster of related strains which also lack resistance to erythromycin. *Appl Environ Microbiol*. 2007;73:1208-14.
263. Thakur S, Gebreyes WA. *Campylobacter coli* in swine production: antimicrobial resistance mechanisms and molecular epidemiology. *J Clin Microbiol*. 2005;43:5705-14.
264. Connerton PL, Loc Carrillo CM, Swift C, Dillon E, Scott A, Rees CE, Dodd CE, Frost J, Connerton IF. Longitudinal study of *Campylobacter jejuni* bacteriophages and their hosts from broiler chickens. *Appl Environ Microbiol*. 2004;70:3877-83.
265. Bull SA, Allen VM, Domingue G, Jorgensen F, Frost JA, Ure R, Whyte R, Tinker D, Corry JE, Gillard-King J, Humphrey TJ. Sources of *Campylobacter* spp. colonizing housed broiler flocks during rearing. *Appl Environ Microbiol*. 2006;72:645-52.

266. Ogden ID, MacRae M, Johnston M, Strachan NJ, Cody AJ, Dingle KE, Newell DG. Use of multilocus sequence typing to investigate the association between the presence of *Campylobacter* spp. in broiler drinking water and *Campylobacter* colonization in broilers. *Appl Environ Microbiol*. 2007;73:5125-9.
267. McTavish SM, Pope CE, Nicol C, Sexton K, French N, Carter PE. Wide geographical distribution of internationally rare *Campylobacter* clones within New Zealand. *Epidemiol Infect*. 2007:1-9.
268. Anon. Sentinel surveillance of *campylobacter* in England and Wales. *Commun Dis Rep CDR Wkly*. 2000;10:169, 172.
269. Food Standards Agency. UK-wide Survey of *Salmonella* and *Campylobacter* Contamination of Fresh and Frozen Chicken on Retail Sale. In. London: Food Standards Agency; 2003.
270. Meldrum RJ. A Survey of *Campylobacter* and *Salmonella* in Raw Retail Chicken Available to Consumers in Wales and Northern Ireland. In: National Public Health Service for Wales; 2006:1-26.
271. Colles FM, Dingle KE, Cody AJ, Maiden MC. Comparison of *Campylobacter* populations in wild geese with those in starlings and free-range poultry on the same farm. *Appl Environ Microbiol*. 2008;74:3583-90.
272. Schneider S, Roessli D, Excoffier L. Arlequin: A software for population genetics data analysis. In. Ver 2.000. ed. Geneva: Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.; 2000.
273. Feil EJ. eBURST. In; 2004:1-14.
274. Kumar S, Tamura K, Nei M. MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Briefings in Bioinformatics*. 2004;5:150-163.
275. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*. 1992;131:479-91.
276. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945-59.
277. Miller G, Dunn GM, Reid TM, Ogden ID, Strachan NJ. Does age acquired immunity confer selective protection to common serotypes of *Campylobacter jejuni*? *BMC Infect Dis*. 2005;5:66.
278. Yazdankhah SP, Kriz P, Tzanakaki G, Kremastinou J, Kalmusova J, Musilek M, Alvestad T, Jolley KA, Wilson DJ, McCarthy ND, Caugant DA, Maiden MC. Distribution of serogroups and genotypes among disease-associated and carried isolates of *Neisseria meningitidis* from the Czech Republic, Greece, and Norway. *J Clin Microbiol*. 2004;42:5146-53.
279. Kuroki S, Saida T, Nukina M, Haruta T, Yoshioka M, Kobayashi Y, Nakanishi H. *Campylobacter jejuni* strains from patients with Guillain-Barré syndrome belong mostly to Penner serogroup 19 and contain beta-N-acetylglucosamine residues. *Ann Neurol*. 1993;33:243-7.
280. Korlath JA, Osterholm MT, Judy LA, Forfang JC, Robinson RA. A point-source outbreak of campylobacteriosis associated with consumption of raw milk. *J Infect Dis*. 1985;152:592-6.
281. Slader J, Domingue G, Jorgensen F, McAlpine K, Owen RJ, Bolton FJ, Humphrey TJ. Impact of transport crate reuse and of catching and processing on *Campylobacter* and *Salmonella* contamination of broiler chickens. *Appl Environ Microbiol*. 2002;68:713-9.

282. Karenlampi R, Rautelin H, Hakkinen M, Hanninen ML. Temporal and geographical distribution and overlap of Penner heat-stable serotypes and pulsed-field gel electrophoresis genotypes of *Campylobacter jejuni* isolates collected from humans and chickens in Finland during a seasonal peak. *J Clin Microbiol*. 2003;41:4870-2.
283. Surveillance ZaEIDS, Department for the Environment, Food and Rural Affairs (Defra). Density of premises and poultry species registered on the GB Poultry Register, for level 2 statistical areas of Great Britain. In. London: Department for the Environment, Food and Rural Affairs (Defra); 2006.
284. Todd EC. Epidemiology of foodborne diseases: a worldwide review. *World Health Stat Q*. 1997;50:30-50.
285. Tauxe RV. Emerging foodborne diseases: an evolving public health challenge. *Emerg Infect Dis*. 1997;3:425-34.
286. Gillespie IA, O'Brien SJ, Adak GK, Tam CC, Frost JA, Bolton FJ, Tompkins DS. Point source outbreaks of *Campylobacter jejuni* infection--are they more common than we think and what might cause them? *Epidemiol Infect*. 2003;130:367-75.
287. Luber P, Bartelt E. Enumeration of *Campylobacter* spp. on the surface and within chicken breast fillets. *J Appl Microbiol*. 2007;102:313-8.
288. Tauxe RV. Molecular subtyping and the transformation of public health. *Foodborne Pathog Dis*. 2006;3:4-8.
289. Huang S, Luangtongkum T, Morishita TY, Zhang Q. Molecular typing of *Campylobacter* strains using the cmp gene encoding the major outer membrane protein. *Foodborne Pathog Dis*. 2005;2:12-23.
290. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol*. 1988;26:2465-6.
291. Hunter PR. Reproducibility and indices of discriminatory power of microbial typing methods. *J Clin Microbiol*. 1990;28:1903-5.
292. Gupta A, Hunter SB, Bidol SA, Dietrich S, Kincaid J, Salehi E, Nicholson L, Genese CA, Todd-Weinstein S, Marengo L, Kimura AC, Brooks JT. *Escherichia coli* O157 cluster evaluation. *Emerg Infect Dis*. 2004;10:1856-8.
293. Hedberg CW, Besser JM. Commentary: cluster evaluation, PulseNet, and public health practice. *Foodborne Pathog Dis*. 2006;3:32-5.
294. Hunter PR, Fraser CA. Application of a numerical index of discriminatory power to a comparison of four physiochemical typing methods for *Candida albicans*. *J Clin Microbiol*. 1989;27:2156-60.
295. Hunter PR. A critical review of typing methods for *Candida albicans* and their applications. *Crit Rev Microbiol*. 1991;17:417-34.
296. Weir BS, Anderson AD, Hepler AB. Genetic relatedness analysis: modern data and new challenges. *Nat Rev Genet*. 2006;7:771-80.
297. Hill F, Gemund C, Benes V, Ansorge W, Gibson TJ. An estimate of large-scale sequencing accuracy. *EMBO Rep*. 2000;1:29-31.
298. Sawires YS, Songer JG. Multiple-locus variable-number tandem repeat analysis for strain typing of *Clostridium perfringens*. *Anaerobe*. 2005;11:262-72.
299. Boxrud D, Pederson-Gulrud K, Wotton J, Medus C, Lyszkowicz E, Besser J, Bartkus JM. Comparison of multiple-locus variable-number tandem repeat analysis, pulsed-field gel electrophoresis, and phage typing for subtype analysis of *Salmonella enterica* serotype Enteritidis. *J Clin Microbiol*. 2007;45:536-43.
300. Pasqualotto AC, Denning DW, Anderson MJ. A cautionary tale: Lack of consistency in allele sizes between two laboratories for a published multilocus microsatellite typing system. *J Clin Microbiol*. 2007;45:522-8.

301. Top J, Schouls LM, Bonten MJ, Willems RJ. Multiple-locus variable-number tandem repeat analysis, a novel typing scheme to study the genetic relatedness and epidemiology of Enterococcus faecium isolates. *J Clin Microbiol*. 2004;42:4503-11.
302. Urwin R, Maiden MC. Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol*. 2003;11:479-87.
303. McCarthy N, Giesecke J. Case-case comparisons to study causation of common infectious diseases. *Int J Epidemiol*. 1999;28:764-8.
304. Blaser MJ, Sazie E, Williams LP, Jr. The influence of immunity on raw milk--associated *Campylobacter* infection. *JAMA*. 1987;257:43-6.
305. Garcia Rodriguez LA, Ruigomez A. Gastric acid, acid-suppressing drugs, and bacterial gastroenteritis: how much of a risk? *Epidemiology*. 1997;8:571-4.
306. Garcia Rodriguez LA, Ruigomez A, Panes J. Use of acid-suppressing drugs and the risk of bacterial gastroenteritis. *Clin Gastroenterol Hepatol*. 2007;5:1418-23.
307. Effler P, Ieong MC, Kimura A, Nakata M, Burr R, Cremer E, Slutsker L. Sporadic *Campylobacter jejuni* infections in Hawaii: associations with prior antibiotic use and commercially prepared chicken. *J Infect Dis*. 2001;183:1152-5.
308. Neal KR, Scott HM, Slack RC, Logan RF. Omeprazole as a risk factor for campylobacter gastroenteritis: case-control study. *BMJ*. 1996;312:414-5.
309. Sorvillo FJ, Lieb LE, Waterman SH. Incidence of campylobacteriosis among patients with AIDS in Los Angeles County. *J Acquir Immune Defic Syndr*. 1991;4:598-602.
310. Pulik M, Genet P, Leturdu F, Lionnet F, Louvel D, Touahri T. Rifabutin prophylaxis against *Mycobacterium avium* complex infections in HIV-infected patients: impact on the incidence of campylobacteriosis. *AIDS Patient Care STDS*. 1999;13:467-72.
311. Tee W, Mijch A. *Campylobacter jejuni* bacteremia in human immunodeficiency virus (HIV)-infected and non-HIV-infected patients: comparison of clinical features and review. *Clin Infect Dis*. 1998;26:91-6.
312. Tee W, Mijch A, Wright E, Yung A. Emergence of multidrug resistance in *Campylobacter jejuni* isolates from three patients infected with human immunodeficiency virus. *Clin Infect Dis*. 1995;21:634-8.
313. Vellinga A, Van Loock F. The dioxin crisis as experiment to determine poultry-related campylobacter enteritis. *Emerg Infect Dis*. 2002;8:19-22.
314. Studahl A, Andersson Y. Risk factors for indigenous *Campylobacter* infection: a Swedish case-control study. *Epidemiol Infect*. 2000;125:269-75.
315. Christenson B, Ringner A, Blucher C, Billaudelle H, Gundtoft KN, Eriksson G, Bottiger M. An outbreak of *campylobacter* enteritis among the staff of a poultry abattoir in Sweden. *Scand J Infect Dis*. 1983;15:167-72.
316. Cawthraw SA, Lind L, Kaijser B, Newell DG. Antibodies, directed towards *Campylobacter jejuni* antigens, in sera from poultry abattoir workers. *Clin Exp Immunol*. 2000;122:55-60.
317. Adak GK, Cowden JM, Nicholas S, Evans HS. The Public Health Laboratory Service national case-control study of primary indigenous sporadic cases of *Campylobacter* infection. *Epidemiol Infect*. 1995;115:15-22.
318. Stuart J, Sufi F, McNulty C, Park P. Outbreak of campylobacter enteritis in a residential school associated with bird pecked bottle tops. *Commun Dis Rep CDR Rev*. 1997;7:R38-40.
319. Roels TH, Wickus B, Bostrom HH, Kazmierczak JJ, Nicholson MA, Kurzynski TA, Davis JP. A foodborne outbreak of *Campylobacter jejuni* (O:33) infection associated with tuna salad: a rare strain in an unusual vehicle. *Epidemiol Infect*. 1998;121:281-7.

320. Gaudreau C, Michaud S. Cluster of erythromycin- and ciprofloxacin-resistant *Campylobacter jejuni* subsp. jejuni from 1999 to 2001 in men who have sex with men, Quebec, Canada. *Clin Infect Dis*. 2003;37:131-6.
321. Quinn TC, Goodell SE, Fennell C, Wang SP, Schuffler MD, Holmes KK, Stamm WE. Infections with *Campylobacter jejuni* and *Campylobacter*-like organisms in homosexual men. *Ann Intern Med*. 1984;101:187-92.
322. Olsen SJ, Hansen GR, Bartlett L, Fitzgerald C, Sonder A, Manjrekar R, Riggs T, Kim J, Flahart R, Pezzino G, Swerdlow DL. An outbreak of *Campylobacter jejuni* infections associated with food handler contamination: the use of pulsed-field gel electrophoresis. *J Infect Dis*. 2001;183:164-7.
323. Cohen DI, Rouach TM, Rogol M. A *Campylobacter* enteritis outbreak in a military base in Israel. *Isr J Med Sci*. 1984;20:216-8.

10. Annexes

10.1 Annex 1 Summary of risk factors

The following overview of risk factors for Campylobacteriosis in humans is based on exposures identified by the formal methods outlined above and any other association reported in the peer-reviewed literature for which quantitative evidence is described not fitting into the criteria for inclusion of the preceding sections.

10.1.1 Host factors

The age and sex associations with reported *C. jejuni* infection are described above. Other host factors have also been identified. Diabetes mellitus was associated with an odds ratio of 4.1 in the only published analytical study to address this risk factor[81]. Gastric acid suppressing medicines also appear to increase risk of *Campylobacter* gastroenteritis[81,305-308] (as well as Salmonella gastroenteritis[306]). The evidence is stronger for proton pump antagonists where it has been demonstrated in independent studies[81,306] and where there is evidence of a dose response relationship[306] than for H2 histamine receptor antagonists where the initial positive findings[81] have not been supported in later studies[305,306]. Just one study has noted a risk associated with recent use of antibiotics[307] . Very high rates of Campylobacteriosis have been reported in AIDS patients[309,310] and severe and persistent disease is also well recognised in patients with AIDS[311,312]. Three are no good recent studies to indicate the impact of current antiviral regimens on this and whether well-treated HIV infection is a risk factor is uncertain. A reported association with chronic gastrointestinal conditions could represent a real association but is particularly prone to a diagnostic bias with increased testing in those with underlying conditions[105].

10.1.2 Food types

10.1.2.1 Chicken

The "natural experiment" evidence[112,313] provides strong evidence for chicken meat as the dominant origin of infection, perhaps in the regions of 70%, in two populations. Additionally it is the commonest food exposure identified in outbreaks and the commonest single food exposure in case control studies, although the strength of association in these studies was both weaker and more variable than might be expected form the natural experiment based estimates. Consumption of undercooked chicken or chicken in a restaurant were often the chicken exposures associated with disease rather than, or as well as, a general association with consumption of chicken. Contact with chicken meat was also identified as a risk in some studies.

10.1.2.2 Other meats

Other meat consumption and contact are variously identified in case control studies, with pork, undercooked meat and consumption of meat at a barbecue often reported. However these are reported far less than poultry. Offal is reported in several studies associated with both increased and decreased odds of infection.

10.1.2.3 Raw and contaminated milk

Unpasteurised milk is a recurrent association in both outbreak and case control studies. Although it appears that this exposure is decreasing the association has been identified in two relatively large and recent studies[86,99]. Contamination following pasteurisation has been identified in outbreak studies but appears uncommon with the exception of bird-pecked milk

in the UK. Unpasteurised and contaminated milk thus appears to be a strong risk factor when it occurs that is decreasing in importance due to decreased exposure.

10.1.2.4 Other foods

There is no recurrent or reliable evidence for other foods although many foods have been associated with disease in some studies.

10.1.3 Routes of exposure to food

10.1.3.1 Restaurants

The frequent association with consumption of chicken in a restaurant in case control studies, the over-representation of hotels and restaurants as the sources of outbreaks and the frequent identification of cross contamination faults in implicated catering premises suggest that commercial catering may be an important transmission route for *C. jejuni* from contaminated food to people.

10.1.3.2 Barbecue

Attendance at a barbecue per se and consumption of meats at a barbecue have both been repeatedly associated with infection. Risks of cross-contamination, direct contact with meat and undercooking may each contribute to the apparently increased risk from this setting.

10.1.3.3 Direct contact and cross contamination

Direct contact with poultry and meat in shops and at home have been associated with infection. Domestic cross contamination and direct contact may also be important routes of infection. Questions assessing the structure and processes of kitchen hygiene have produced mixed results (Table 1.2).

10.1.4 Water

Water has been a prominent cause in outbreaks where small local supplies or larger supplies with deficits in maintenance or chlorination appear to pose a higher risk. The relative importance of water supplies appears to be declining over the years. Consumption of outdoor or otherwise untreated water appears to be an increased risk

10.1.5 Environment and animal exposures

10.1.5.1 Farm animal

Contact with a range of farm animals has been identified in many studies, with some evidence that it may be more important among children than animals. The frequency of this exposure is low at a population level limiting the overall impact of this exposure with the exception of one study in Michigan where poultry husbandry was associated with up to 18% of incident disease[103].

10.1.5.2 Companion animals

Many studies report an association between disease and pet ownership or contact. Although the precise nature of contact varies ownership of puppies is frequently identified as a risk. Like farm animal contact the population attributable risk for this exposure is generally small even when a strong association is identified. Since pet dogs do not appear to sustain *C. jejuni* infection they may act as a vehicle for infection originating in other sources[135].

10.1.5.3 Environment

Recreational contact with water has been found to have both positive and negative associations and may present a risk (e.g. Table 2). Although living on a farm[314] or in areas of high ruminant density[101] have been associated with increased incidence of infection there is no reliable association with other specific forms of direct environmental contact apart from direct contact with animals or consumption of untreated water.

10.1.6 Occupational risk

Several population based case control studies have identified poultry husbandry as a risk factor for *Campylobacter* infection[102,103,314] as well as other occupational animal exposures[102]. In an outbreak of among staff at a poultry abattoir disease incidence was lower in longstanding than new abattoir workers[315] while 7 of 78 abattoir workers in another study had asymptomatic carriage and longstanding abattoir workers also had higher levels of antibody to *Campylobacter*[316] indicating occupational exposure to *Campylobacter* in this setting. Other occupational contact with raw meat has also been associated with disease[317].

10.1.7 Person to person spread

Person to person spread of *C. jejuni* is generally reported as uncommon and humans clearly do not sustain transmission in industrialised countries. However the increasing understanding of the frequency of asymptomatic infection as described under the clinical features of this infection means that moderate amounts of secondary infection may occur without obvious clusters of linked clinical cases and many studies seeking evidence for secondary spread have found good evidence for limited person to person transmission such as in a milk related outbreak in a residential home for those with learning difficulties[318], in an abattoir outbreak[315] and among adolescents at a summer camp[319]. Studies in families of cases have identified that 27% of cases reported similar symptoms in family members[314] while another study screened 200 asymptomatic household contacts and identified infection in 20. In these cases it is difficult to separate secondary infection from joint primary infection. There is evidence for sexual transmission among homosexual men[320,321] and one well characterised outbreak in this population[320]. Only two published reports describe outbreaks with foodborne spread from infected food handlers as the likely source of human infection[322,323].

10.2 Annex 2. Questionnaire

CAMPYLOBACTER INVESTIGATION FORM.

Date completed ...../...../..... NHS no. ..... Lab. code .....

Personal details

Family name _____ First name_____

Date of Birth ___/___/___ (day/month/year) Age ____ years _____ months (if <1yr)

Sex ☐ Male ☐ Female Telephone number _____ Postcode _____

Occupation (please include retired, student, pre-school child etc.) _____

Are you a commercial food handler? Yes ☐ No ☐

Describe your ethnic origin: _____

2. Symptoms.

On what date did your illness start ___/___/___ (day/month/year)

How long did your illness last ____ (days)

Have you experienced any of the following symptoms due to your illness:

Diarrhoea (four or more loose/runny stools in 24 hours)	☐ Yes	☐ No	☐ Unsure
Bloody stools	☐ Yes	☐ No	☐ Unsure
Vomiting	☐ Yes	☐ No	☐ Unsure
Abdominal pain (cramps)	☐ Yes	☐ No	☐ Unsure
Fever (feeling hot and cold)	☐ Yes	☐ No	☐ Unsure

Are you still ill? ☐ Yes ☐ No

Did you seek advice from a doctor ☐ Yes ☐ No

Were you admitted into hospital ☐ Yes ☐ No

If yes, for how many days were you admitted _____ (days)

For how many days were you prevented from going to work or from undertaking normal daily activities because of your illness _____ (days)

3. Travel

Did you travel OUTSIDE THE UK in the two weeks before your illness started

☐ Yes ☐ No

If yes, where did you go

Country _____ Town/Resort _____

Hotel _____ Date of arrival ___/___/___ (day/month/year)

Date of departure ___/___/___ (day/month/year)

Did you travel IN THE UK in the two weeks before your illness started ☐ Yes☐ No

If yes, where did you go

Town/Resort _____ Hotel _____

Date of arrival ___/___/___ Date of departure ___/___/___ (day/month/year)

4. Food consumption in the two weeks before your illness

Food type	How often eaten		
Baby food	☐ Never	☐ Once	☐ More often
Barbecued food	☐ Never	☐ Once	☐ More often
Beef (inc roast, mince, steak)	☐ Never	☐ Once	☐ More often
Cheese	☐ Never	☐ Once	☐ More often
Chicken	☐ Never	☐ Once	☐ More often
Cold meats (pre-cooked)	☐ Never	☐ Once	☐ More often
Fish/shellfish	☐ Never	☐ Once	☐ More often
Halal meat	☐ Never	☐ Once	☐ More often
Lamb	☐ Never	☐ Once	☐ More often
Meat pies	☐ Never	☐ Once	☐ More often
Offal or tripe	☐ Never	☐ Once	☐ More often
Other poultry	☐ Never	☐ Once	☐ More often
Organic meat	☐ Never	☐ Once	☐ More often
Organic vegetables	☐ Never	☐ Once	☐ More often
Pate	☐ Never	☐ Once	☐ More often
Pork, ham or bacon	☐ Never	☐ Once	☐ More often
Pre-prepared sandwiches	☐ Never	☐ Once	☐ More often
Salads	☐ Never	☐ Once	☐ More often
Sausages	☐ Never	☐ Once	☐ More often
Vegetarian food	☐ Never	☐ Once	☐ More often

Were any of the above foods cooked 'rare' (i.e. was pink inside) ☐ Yes ☐ No ☐ Unsure

Have you eaten in a restaurant in the last two weeks ☐ Yes ☐ No ☐ Unsure

If yes please give details _____

Please estimate how often you handled raw meat in the two weeks before your illness

☐ 0 ☐ 1 ☐ 2 - 5 ☐ 6 - 10 ☐ 11plus

Did you drink (or have in your cereal) unpasteurised (e.g. farm fresh; green top) milk in the two weeks before illness

☐ Yes☐ No ☐ Unsure

Did you drink (or take in cereal) bird-pecked milk in the two weeks before illness

☐ Yes☐ No ☐ Unsure

Is most of your milk: ☐ Doorstep-delivered ☐ Supermarket-bought

5. Water consumption

Did you drink cold, unboiled water in the two weeks before illness from:

Mains water supply ☐ Yes ☐ No

Private water supply ☐ Yes ☐ No *If yes, was it treated* ☐ Yes ☐ No ☐ Unsure

Bottled water ☐ Yes ☐ No

River, stream, spring ☐ Yes ☐ No

A filter jug ☐ Yes ☐ No

Had you noticed any engineering work or supply problem (e.g. discoloured or tainted water)

☐ Yes ☐ No

*If yes, please specify*_____

How many glasses of unboiled water do you drink a day (i.e. straight from the tap or in squash) _____

6. Recreational water activity

Did you take part in any of the following water (fresh or sea) sports or activities in the two weeks before illness:

Swimming/paddling ☐ Yes ☐ No
Sailing ☐ Yes ☐ No
Windsurfing ☐ Yes ☐ No
Fishing ☐ Yes ☐ No

If yes, please give details

_____

7. Contact with animals

Have you had any contact with animals in the two weeks prior to illness? ☐ Yes ☐ No

If yes

Do you have any pets ☐ Yes ☐ No

*If yes, please specify type*_____

Have any of your pets had diarrhoea in the two weeks before illness ☐ Yes ☐ No

If yes, please list _____

Did you clean up the faeces (mess) ☐ Yes ☐ No

Have you visited a farm in the two weeks before illness ☐ Yes ☐ No

If yes, please give details

_____

8. Household members

How many people live in your household _____ Adults _____ Children

Has anyone else in your household been ill with similar symptoms recently (whether they ate the same food or not) ☐ Yes ☐ No

If yes, please give details

_____

_____

_____

Has anyone else outside the household had a similar illness recently ☐ Yes ☐ No

If yes, please give their name and address

_____

_____

_____

10.3 Annex 3 – Additional tables from chapter 4.

10.3.1 Allele distributions by locus for isolates with complete and incomplete types.

aspA	Complete		Incomplete	
	n	Percent	n	Percent
1	79	5.2%	4	21.1%
2	734	48.7%	11	57.9%
3	12	0.8%	0	0.0%
4	162	10.7%	1	5.3%
6	9	0.6%	0	0.0%
7	139	9.2%	0	0.0%
8	65	4.3%	0	0.0%
9	217	14.4%	1	5.3%
10	14	0.9%	0	0.0%
14	9	0.6%	0	0.0%
17	1	0.1%	0	0.0%
19	1	0.1%	0	0.0%
22	2	0.1%	0	0.0%
24	16	1.1%	0	0.0%
26	1	0.1%	0	0.0%
33	3	0.2%	0	0.0%
47	2	0.1%	0	0.0%
48	2	0.1%	0	0.0%
55	1	0.1%	0	0.0%
57	13	0.9%	0	0.0%
62	20	1.3%	0	0.0%
64	3	0.2%	1	5.3%
79	0	0.0%	1	5.3%
91	1	0.1%	0	0.0%
95	1	0.1%	0	0.0%
186	1	0.1%	0	0.0%
Total	1508		19	

glnA	Complete		Incomplete	
	n	Percent	n	Percent
1	561	37.2%	17	51.5%
2	272	18.0%	6	18.2%
3	15	1.0%	1	3.0%
4	192	12.7%	4	12.1%
6	8	0.5%	0	0.0%
7	166	11.0%	4	12.1%
10	29	1.9%	0	0.0%
15	4	0.3%	0	0.0%
17	77	5.1%	0	0.0%
21	42	2.8%	0	0.0%
22	1	0.1%	0	0.0%
23	1	0.1%	0	0.0%
25	22	1.5%	0	0.0%
27	4	0.3%	0	0.0%
28	12	0.8%	0	0.0%
29	1	0.1%	0	0.0%
30	1	0.1%	0	0.0%
33	2	0.1%	0	0.0%
38	1	0.1%	0	0.0%
39	1	0.1%	0	0.0%
45	1	0.1%	0	0.0%
53	50	3.3%	0	0.0%
55	2	0.1%	0	0.0%
59	2	0.1%	0	0.0%
70	0	0.0%	1	3.0%
75	21	1.4%	0	0.0%
81	8	0.5%	0	0.0%
105	1	0.1%	0	0.0%
112	8	0.5%	0	0.0%
115	1	0.1%	0	0.0%
124	1	0.1%	0	0.0%
172	1	0.1%	0	0.0%
Total	1508		33	

gltA	Complete		Incomplete	
	n	Percent	n	Percent
1	389	25.8%	12	34.3%
2	207	13.7%	3	8.6%
3	16	1.1%	0	0.0%
4	217	14.4%	2	5.7%
5	201	13.3%	5	14.3%
6	22	1.5%	1	2.9%
8	1	0.1%	0	0.0%
9	1	0.1%	0	0.0%
10	129	8.6%	3	8.6%
12	87	5.8%	3	8.6%
16	3	0.2%	0	0.0%
21	138	9.2%	2	5.7%
22	5	0.3%	0	0.0%
29	1	0.1%	0	0.0%
30	3	0.2%	1	2.9%
40	34	2.3%	1	2.9%
41	6	0.4%	0	0.0%
42	19	1.3%	2	5.7%
43	1	0.1%	0	0.0%
50	8	0.5%	0	0.0%
52	4	0.3%	0	0.0%
57	1	0.1%	0	0.0%
59	1	0.1%	0	0.0%
60	4	0.3%	0	0.0%
75	1	0.1%	0	0.0%
77	1	0.1%	0	0.0%
84	1	0.1%	0	0.0%
85	1	0.1%	0	0.0%
94	1	0.1%	0	0.0%
149	3	0.2%	0	0.0%
218	1	0.1%	0	0.0%
227	1	0.1%	0	0.0%
Total	1508		35	

glyA	Complete		Incomplete	
	n	Percent	n	Percent
1	1	0.1%	0	0.0%
2	278	18.4%	2	9.1%
3	544	36.1%	13	59.1%
4	228	15.1%	1	4.5%
5	14	0.9%	0	0.0%
9	2	0.1%	0	0.0%
10	78	5.2%	0	0.0%
15	36	2.4%	0	0.0%
17	12	0.8%	0	0.0%
19	5	0.3%	0	0.0%
24	7	0.5%	0	0.0%
26	1	0.1%	0	0.0%
27	1	0.1%	0	0.0%
28	12	0.8%	0	0.0%
37	12	0.8%	0	0.0%
38	2	0.1%	0	0.0%
48	21	1.4%	0	0.0%
49	2	0.1%	0	0.0%
51	1	0.1%	0	0.0%
53	3	0.2%	0	0.0%
62	219	14.5%	2	9.1%
68	1	0.1%	0	0.0%
77	1	0.1%	0	0.0%
79	0	0.0%	1	4.5%
82	2	0.1%	0	0.0%
88	4	0.3%	2	9.1%
93	1	0.1%	0	0.0%
97	0	0.0%	1	4.5%
99	8	0.5%	0	0.0%
105	1	0.1%	0	0.0%
111	1	0.1%	0	0.0%
127	1	0.1%	0	0.0%
139	1	0.1%	0	0.0%
155	1	0.1%	0	0.0%
194	4	0.3%	0	0.0%
195	1	0.1%	0	0.0%
292	1	0.1%	0	0.0%
293	1	0.1%	0	0.0%
Total	1508		22	

pgm	Complete		Incomplete	
	n	Percent	n	Percent
1	98	6.5%	0	0.0%
2	550	36.5%	14	66.7%
3	21	1.4%	1	4.8%
4	205	13.6%	1	4.8%
5	16	1.1%	0	0.0%
6	28	1.9%	2	9.5%
7	164	10.9%	1	4.8%
8	2	0.1%	0	0.0%
10	43	2.9%	0	0.0%
11	131	8.7%	0	0.0%
12	8	0.5%	0	0.0%
13	3	0.2%	0	0.0%
17	12	0.8%	0	0.0%
19	15	1.0%	0	0.0%
22	21	1.4%	0	0.0%
23	41	2.7%	0	0.0%
42	68	4.5%	0	0.0%
43	1	0.1%	0	0.0%
58	1	0.1%	0	0.0%
59	1	0.1%	0	0.0%
60	2	0.1%	0	0.0%
67	8	0.5%	0	0.0%
89	9	0.6%	0	0.0%
93	1	0.1%	0	0.0%
94	1	0.1%	0	0.0%
98	9	0.6%	1	4.8%
103	18	1.2%	0	0.0%
104	0	0.0%	1	4.8%
113	1	0.1%	0	0.0%
114	1	0.1%	0	0.0%
118	1	0.1%	0	0.0%
120	8	0.5%	0	0.0%
122	1	0.1%	0	0.0%
127	1	0.1%	0	0.0%
128	1	0.1%	0	0.0%
132	1	0.1%	0	0.0%
134	1	0.1%	0	0.0%
141	1	0.1%	0	0.0%
169	1	0.1%	0	0.0%
172	1	0.1%	0	0.0%
189	1	0.1%	0	0.0%
236	1	0.1%	0	0.0%
261	2	0.1%	0	0.0%
264	2	0.1%	0	0.0%
341	1	0.1%	0	0.0%
342	1	0.1%	0	0.0%
363	1	0.1%	0	0.0%
366	2	0.1%	0	0.0%
371	1	0.1%	0	0.0%
Total	1508		21	

tkt	Complete		Incomplete	
	n	Percent	n	Percent
1	705	46.8%	18	75.0%
2	1	0.1%	0	0.0%
3	211	14.0%	2	8.3%
5	201	13.3%	2	8.3%
7	130	8.6%	1	4.2%
9	18	1.2%	0	0.0%
11	12	0.8%	1	4.2%
12	45	3.0%	0	0.0%
18	2	0.1%	0	0.0%
20	1	0.1%	0	0.0%
25	3	0.2%	0	0.0%
28	8	0.5%	0	0.0%
29	1	0.1%	0	0.0%
34	31	2.1%	0	0.0%
35	1	0.1%	0	0.0%
40	5	0.3%	0	0.0%
46	1	0.1%	0	0.0%
47	2	0.1%	0	0.0%
48	4	0.3%	0	0.0%
51	40	2.7%	0	0.0%
54	1	0.1%	0	0.0%
57	2	0.1%	0	0.0%
58	7	0.5%	0	0.0%
59	21	1.4%	0	0.0%
61	8	0.5%	0	0.0%
62	11	0.7%	0	0.0%
67	8	0.5%	0	0.0%
76	8	0.5%	0	0.0%
82	1	0.1%	0	0.0%
83	1	0.1%	0	0.0%
99	1	0.1%	0	0.0%
100	4	0.3%	0	0.0%
103	1	0.1%	0	0.0%
120	1	0.1%	0	0.0%
123	1	0.1%	0	0.0%
133	1	0.1%	0	0.0%
137	1	0.1%	0	0.0%
144	1	0.1%	0	0.0%
160	1	0.1%	0	0.0%
174	1	0.1%	0	0.0%
226	1	0.1%	0	0.0%
291	1	0.1%	0	0.0%
297	1	0.1%	0	0.0%
298	1	0.1%	0	0.0%
303	1	0.1%	0	0.0%
Total	1508		24	

uncA	Complete		Incomplete	
	n	Percent	n	Percent
1	215	14.3%	0	0.0%
3	108	7.2%	0	0.0%
4	1	0.1%	0	0.0%
5	710	47.1%	12	80.0%
6	349	23.1%	1	6.7%
7	5	0.3%	0	0.0%
8	4	0.3%	1	6.7%
12	47	3.1%	0	0.0%
16	3	0.2%	0	0.0%
17	25	1.7%	0	0.0%
21	1	0.1%	0	0.0%
23	2	0.1%	0	0.0%
31	1	0.1%	0	0.0%
34	9	0.6%	1	6.7%
35	3	0.2%	0	0.0%
44	8	0.5%	0	0.0%
46	1	0.1%	0	0.0%
51	1	0.1%	0	0.0%
52	8	0.5%	0	0.0%
57	1	0.1%	0	0.0%
58	1	0.1%	0	0.0%
74	1	0.1%	0	0.0%
114	1	0.1%	0	0.0%
144	1	0.1%	0	0.0%
145	1	0.1%	0	0.0%
195	1	0.1%	0	0.0%
Total	1508		15	

10.3.2 Clonal complexes by county in CSSS

Table 4.3. Distribution of clonal complexes between Hampshire and Nottinghamshire in the *Campylobacter* Sentinel Surveillance Scheme.

Clonal	Nottinghamshire		Hampshire		Total
complex	Number	Percent	Number	Percent	
ST-21	307	37%	250	37%	557
ST-257	122	15%	79	12%	201
ST-45	72	9%	70	10%	142
ST-48	69	8%	56	8%	125
ST-206	31	4%	25	4%	56
ST-353	25	3%	27	4%	52
ST-574	30	4%	20	3%	50
ST-354	27	3%	17	3%	44
ST-443	20	2%	16	2%	36
ST-61	14	2%	9	1%	23
ST-283	16	2%	7	1%	23
ST-52	11	1%	11	2%	22
ST-42	8	1%	13	2%	21
ST-661	11	1%	10	1%	21
ST-22	8	1%	7	1%	15
ST-49	6	1%	6	1%	12
ST-573	7	1%	3	0%	10
ST-460	3	0%	6	1%	9
ST-508	4	0%	4	1%	8
ST-677	4	0%	4	1%	8
ST-658	5	1%	2	0%	7
ST-403	3	0%	2	0%	5
ST-607	3	0%	1	0%	4
ST-433	1	0%	1	0%	2
ST-446	1	0%	1	0%	2
ST-177	1	0%	0	0%	1
ST-682	1	0%	0	0%	1
ST-692	0	0%	1	0%	1
ST-1275	0	0%	1	0%	1
ST-1332	0	0%	1	0%	1
Unassigned	23	3%	22	3%	45
Total	833		672		1505

10.4 Annex 4. CC and ST distribution in datasets from 4 different countries

Table. Distribution of clonal complexes by country

Clonal	Percentage of country total per clonal complex				
complex	England	Finland	Australia	New Zealand	Percent
ST-21	37%	20%	10%	30%	32%
ST-45	9%	46%	4%	5%	14%
ST-257	13%	0%	14%	5%	11%
ST-48	8%	3%	15%	32%	9%
ST-354	3%	0%	14%	8%	4%
ST-206	4%	1%	3%	2%	3%
ST-353	3%	1%	5%	1%	3%
ST-574	3%	0%	0%	0%	2%
ST-677	1%	12%	0%	0%	2%
ST-52	1%	1%	7%	4%	2%
ST-443	2%	0%	1%	0%	2%
ST-22	1%	4%	1%	0%	1%
ST-42	1%	0%	3%	4%	1%
ST-283	2%	1%	0%	0%	1%
ST-61	2%	0%	1%	1%	1%
ST-661	1%	0%	0%	0%	1%
ST-658	0%	1%	5%	0%	1%
ST-49	1%	0%	0%	0%	1%
ST-607	0%	1%	4%	0%	1%
Minor	3%	3%	1%	1%	3%
Unassigned	3%	5%	12%	8%	4%
Total	1505	298	153	107	2066

Table . Distribution of sequence type by country

	Country				
ST	England	Finland	Australia	New Zealand	Total
5	22	0	3	1	26
10	1	0	0	0	1
11	2	5	0	0	7
19	42	0	0	0	42
21	174	1	2	1	178
22	14	10	0	0	24
25	1	0	2	0	3
38	3	0	0	0	3
42	16	0	4	4	24
43	1	0	0	0	1
44	4	0	0	0	4
45	88	85	1	4	178
46	6	1	0	0	7
47	11	0	0	0	11
48	95	5	23	1	124
49	11	0	0	0	11
50	80	50	7	8	145
51	27	0	1	0	28
52	20	3	5	4	32
53	125	0	1	5	131
55	1	0	0	0	1
61	22	0	1	1	24
66	2	0	0	0	2
70	0	0	1	0	1
104	54	0	0	0	54
122	9	1	0	0	10
132	3	0	0	0	3
137	17	17	0	0	34
141	4	0	0	0	4
148	1	0	0	0	1
161	0	0	4	0	4
177	1	0	0	0	1
190	2	0	1	18	21
197	0	0	2	0	2
205	9	0	0	0	9
206	10	0	0	0	10
222	1	0	0	0	1
227	7	1	5	0	13
230	5	12	0	0	17
233	1	0	1	0	2
257	181	0	18	5	204
262	20	0	0	0	20
266	1	0	0	0	1
267	22	2	0	0	24
270	1	0	0	0	1
273	2	0	0	0	2
290	1	0	0	0	1
305	0	1	0	0	1
311	2	0	0	0	2
312	1	0	1	0	2
316	1	0	0	0	1
324	13	0	0	0	13
334	11	1	0	0	12
339	1	0	0	0	1
350	1	0	0	0	1
353	3	1	0	0	4
354	25	0	3	9	37
356	8	0	0	0	8
366	1	0	0	0	1
376	1	0	0	0	1
383	0	1	0	0	1
400	13	0	0	0	13
403	2	0	0	0	2
420	1	0	0	0	1
436	8	0	0	3	11
443	1	0	0	0	1
447	5	0	0	0	5
449	0	0	1	0	1
450	1	1	0	0	2
451	0	1	2	0	3
464	2	0	0	0	2
474	0	0	0	32	32
475	11	3	0	0	14
492	1	0	0	0	1
508	4	0	0	0	4
520	4	0	0	0	4
522	2	0	0	0	2
523	1	0	7	0	8
524	0	0	2	0	2
525	0	0	6	0	6
526	0	0	1	0	1
527	0	0	2	0	2
528	0	0	18	0	18
529	0	0	1	0	1
530	0	0	7	1	8
531	0	0	10	0	10
532	0	0	2	0	2
533	0	0	1	0	1
535	0	0	1	0	1
536	0	0	1	0	1
537	0	0	1	0	1
538	1	4	1	0	6
567	0	0	1	0	1
568	0	0	1	0	1
569	0	0	1	0	1
572	20	1	0	0	21
573	10	0	0	0	10
574	50	0	0	0	50
581	4	0	0	0	4
583	9	1	0	1	11
584	9	0	0	0	9
586	7	1	0	0	8
606	4	0	0	0	4
607	2	0	0	0	2
637	1	0	0	0	1

658	3	1	0	0	4
661	12	0	0	0	12
677	8	28	0	0	36
682	1	0	0	0	1
760	4	0	0	0	4
763	1	0	0	0	1
772	1	0	0	0	1
775	2	0	0	0	2
782	1	0	0	0	1
794	0	9	0	0	9
814	1	0	0	0	1
824	3	1	0	0	4
825	1	0	0	0	1
844	1	0	0	0	1
863	1	0	0	0	1
877	4	0	0	0	4
880	1	0	0	0	1
883	7	2	0	0	9
916	1	0	0	0	1
917	1	0	0	0	1
918	0	1	0	0	1
945	0	3	0	0	3
978	1	0	0	0	1
986	1	0	0	0	1
990	1	0	0	0	1
1001	1	0	0	0	1
1003	0	2	0	0	2
1074	1	0	0	0	1
1080	0	1	0	0	1
1286	1	0	0	0	1
1301	1	0	0	0	1
1326	0	2	0	0	2
1332	1	1	0	0	2
1359	2	0	0	0	2
1363	1	0	0	0	1
1365	5	0	0	0	5
1367	0	2	0	0	2
1395	0	1	0	0	1
1397	1	0	0	0	1
1399	2	0	0	0	2
1446	1	0	0	0	1
1517	3	0	0	0	3
1519	1	0	0	0	1
1525	1	1	0	0	2
1530	1	0	0	0	1
1538	4	0	0	0	4
1597	1	0	0	0	1
1716	0	1	0	0	1
1723	1	0	0	0	1
1728	2	0	0	0	2
1759	2	2	0	0	4
1775	1	0	0	0	1
1887	6	0	0	0	6
1906	1	0	0	0	1

1913	1	0	0	0	1
1915	0	1	0	0	1
1921	1	0	0	0	1
1943	1	1	0	0	2
1944	0	3	0	0	3
1945	0	1	0	0	1
1947	0	2	0	0	2
1948	0	1	0	0	1
1949	0	1	0	0	1
1950	0	1	0	0	1
1951	0	1	0	0	1
1952	0	1	0	0	1
1953	0	1	0	0	1
1954	0	1	0	0	1
1955	0	1	0	0	1
1956	0	1	0	0	1
1958	0	1	0	0	1
1959	0	1	0	0	1
1960	0	1	0	0	1
1961	0	1	0	0	1
1962	0	1	0	0	1
1963	0	1	0	0	1
1964	0	1	0	0	1
1966	0	1	0	0	1
1967	0	2	0	0	2
1969	0	2	0	0	2
1973	0	1	0	0	1
1974	0	1	0	0	1
2026	0	0	0	1	1
2120	1	0	0	0	1
2151	1	0	0	0	1
2187	1	0	0	0	1
2249	1	0	0	0	1
2250	1	0	0	0	1
2260	2	0	0	0	2
2343	0	0	0	1	1
2345	0	0	0	2	2
2347	0	0	0	1	1
2368	1	0	0	0	1
2398	0	0	0	1	1
2407	1	0	0	0	1
2535	0	0	0	3	3
2641	1	0	0	0	1
2807	2	0	0	0	2
2975	1	0	0	0	1
3203	1	0	0	0	1
3212	1	0	0	0	1
3462	1	0	0	0	1
3463	1	0	0	0	1
3464	1	0	0	0	1
3465	1	0	0	0	1
3466	2	0	0	0	2
3468	1	0	0	0	1
3469	1	0	0	0	1

3470	1	0	0	0	1
3471	1	0	0	0	1
3472	1	0	0	0	1
3473	1	0	0	0	1
3474	1	0	0	0	1
3475	1	0	0	0	1
3476	1	0	0	0	1
3477	1	0	0	0	1
3478	1	0	0	0	1
3479	1	0	0	0	1
3480	1	0	0	0	1
3481	1	0	0	0	1
3482	1	0	0	0	1
3483	1	0	0	0	1
3484	1	0	0	0	1
3485	1	0	0	0	1
3486	1	0	0	0	1
3487	1	0	0	0	1
3522	1	0	0	0	1
3523	1	0	0	0	1
3524	1	0	0	0	1
3525	1	0	0	0	1
3526	1	0	0	0	1
Total	1508	298	153	107	2066

10.5 Annex 5. Tabulation of ST distribution in peak and non-peak periods.

ST	Non peak	Peak	Total
5	12	10	22
10	0	1	1
11	0	2	2
19	21	21	42
21	86	88	174
22	7	7	14
25	0	1	1
38	1	2	3
42	10	6	16
43	0	1	1
44	3	1	4
45	26	62	88
46	5	1	6
47	9	2	11
48	50	45	95
49	4	7	11
50	44	36	80
51	17	10	27
52	14	6	20
53	58	67	125
55	1	0	1
61	8	14	22
66	1	1	2
104	25	29	54
122	9	0	9
132	1	2	3
137	4	13	17
141	1	3	4
148	0	1	1
177	0	1	1
190	0	2	2
205	4	5	9
206	3	7	10
222	1	0	1
227	1	6	7
230	0	5	5
233	1	0	1
257	107	74	181
262	6	14	20
266	0	1	1
267	4	18	22
270	1	0	1
273	1	1	2
290	0	1	1
311	2	0	2
312	1	0	1
316	1	0	1
324	10	3	13
334	4	7	11

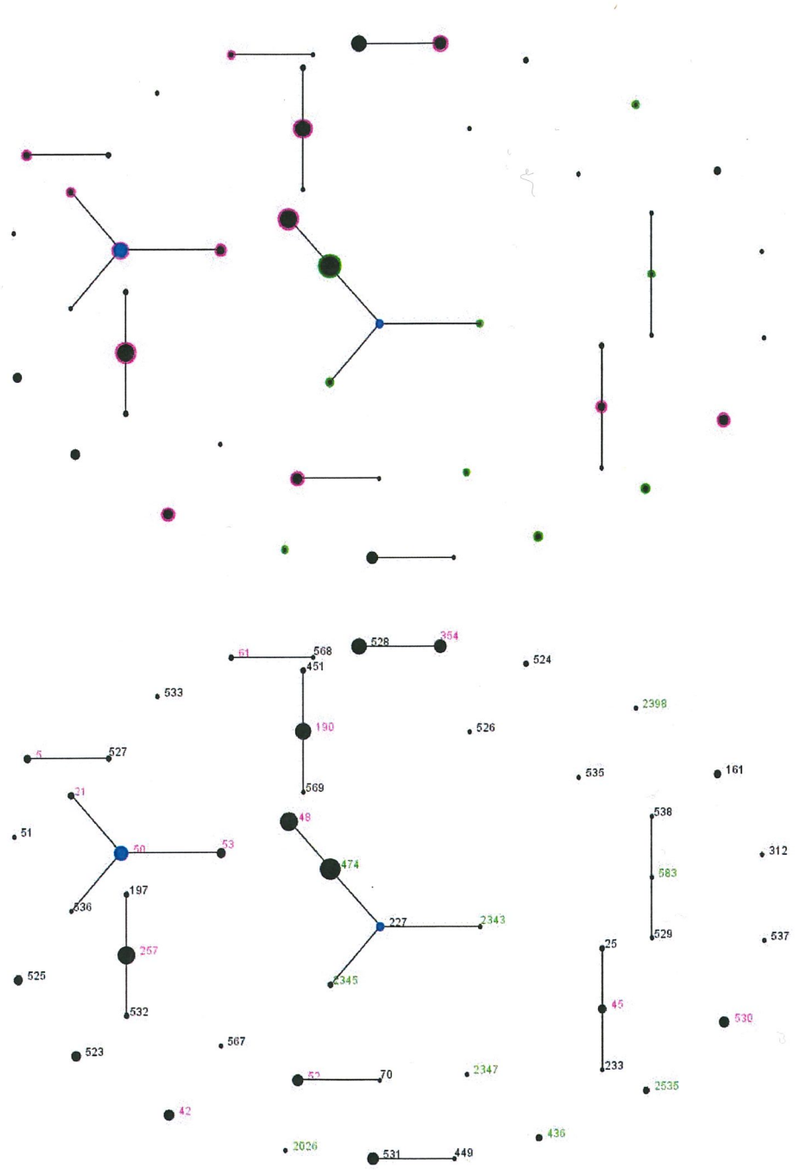
339	0	1	1
350	0	1	1
353	0	3	3
354	13	12	25
356	6	2	8
366	1	0	1
376	1	0	1
400	8	5	13
403	2	0	2
420	1	0	1
436	4	4	8
443	1	0	1
447	3	2	5
450	0	1	1
464	1	1	2
475	4	7	11
492	0	1	1
508	2	2	4
520	0	4	4
522	1	1	2
523	1	0	1
538	0	1	1
572	12	8	20
573	6	4	10
574	30	20	50
581	3	1	4
583	5	4	9
584	5	4	9
586	3	4	7
606	3	1	4
607	2	0	2
637	0	1	1
658	2	1	3
661	3	9	12
677	0	8	8
682	0	1	1
760	3	1	4
763	0	1	1
772	0	1	1
775	1	1	2
782	0	1	1
814	1	0	1
824	2	1	3
825	0	1	1
844	0	1	1
863	0	1	1
877	3	1	4
880	0	1	1
883	5	2	7
916	1	0	1
917	1	0	1
978	0	1	1

986	1	0	1
990	0	1	1
1001	1	0	1
1074	0	1	1
1286	1	0	1
1301	0	1	1
1332	0	1	1
1359	1	1	2
1363	0	1	1
1365	5	0	5
1397	0	1	1
1399	2	0	2
1446	0	1	1
1517	1	2	3
1519	0	1	1
1525	1	0	1
1530	0	1	1
1538	3	1	4
1597	0	1	1
1723	0	1	1
1728	1	1	2
1759	0	2	2
1775	1	0	1
1887	4	2	6
1906	0	1	1
1913	1	0	1
1921	1	0	1
1943	1	0	1
2120	1	0	1
2151	1	0	1
2187	1	0	1
2249	1	0	1
2250	0	1	1
2260	2	0	2
2368	1	0	1
2407	0	1	1
2641	1	0	1
2807	1	1	2
2975	0	1	1
3203	1	0	1
3212	1	0	1
3462	0	1	1
3463	0	1	1
3464	0	1	1
3465	0	1	1
3466	2	0	2
3468	1	0	1
3469	1	0	1
3470	1	0	1
3471	0	1	1
3472	0	1	1
3473	0	1	1

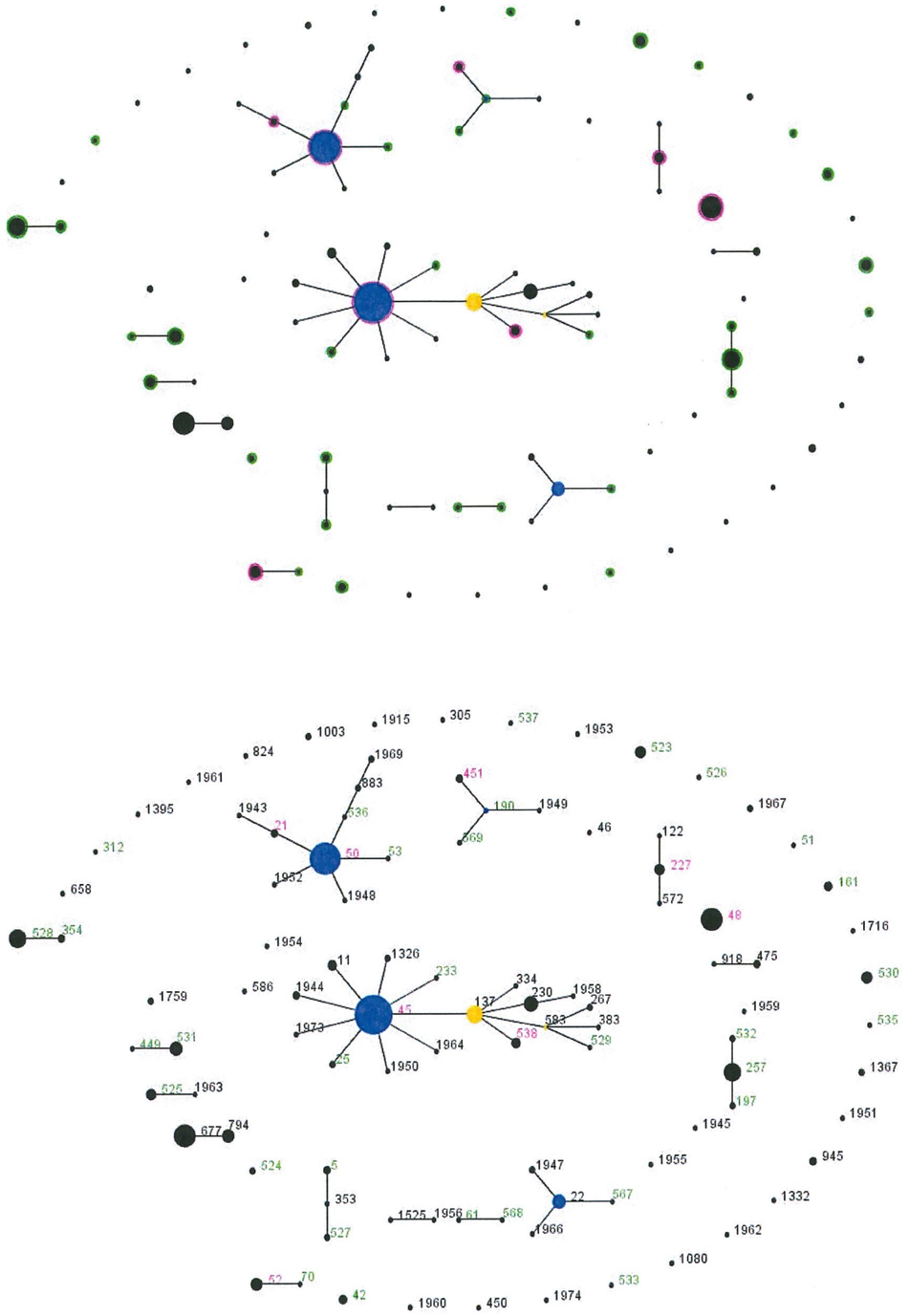
3474	0	1	1
3475	0	1	1
3476	0	1	1
3477	1	0	1
3478	1	0	1
3479	1	0	1
3480	0	1	1
3481	0	1	1
3482	0	1	1
3483	1	0	1
3484	1	0	1
3485	1	0	1
3486	1	0	1
3487	1	0	1
3522	0	1	1
3523	0	1	1
3524	1	0	1
3525	1	0	1
3526	0	1	1
Total	755	753	1508

10.6 Annex 6 – Between country population comparisons using eBURST

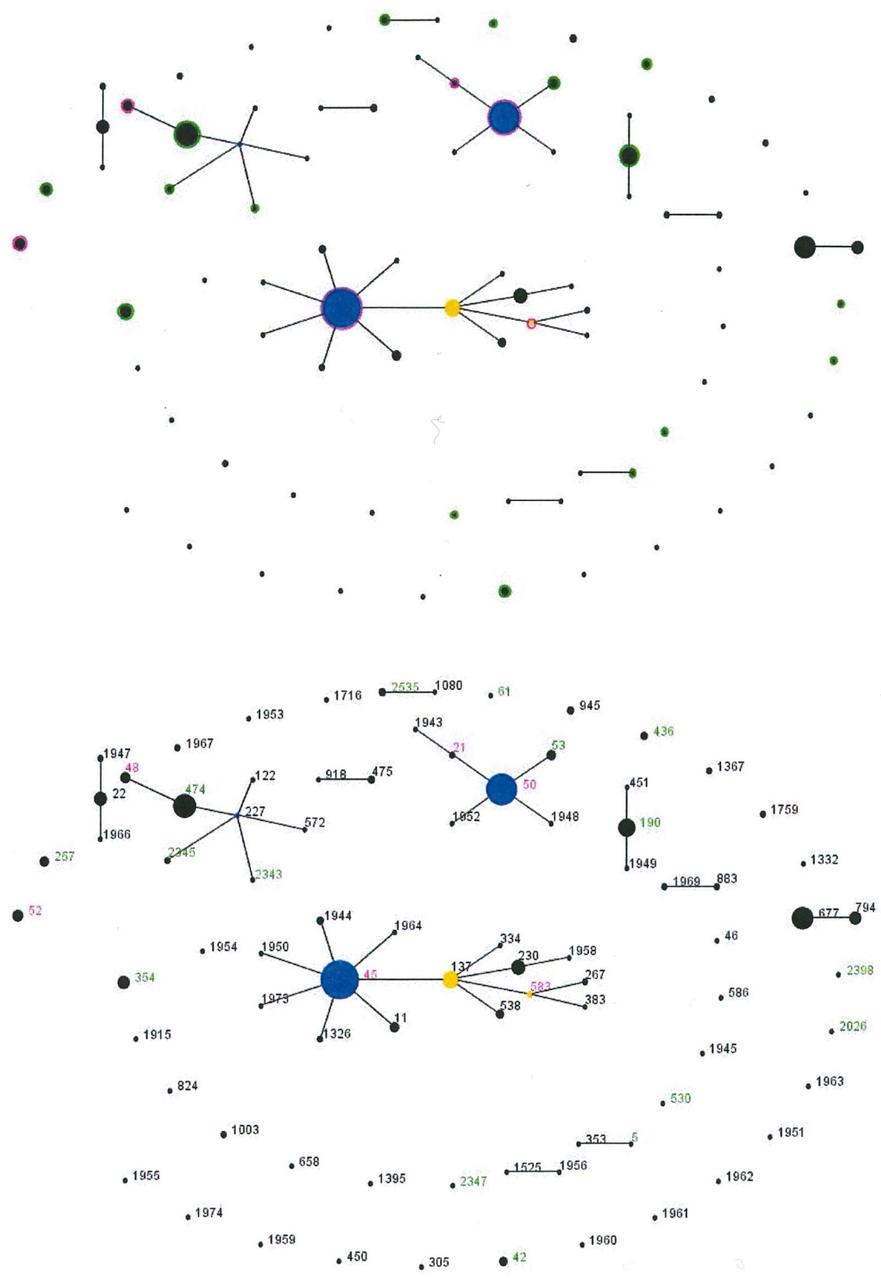
E-BURST datasets comparing collections of isolates from England (CSSS dataset n=1508), Finland (n=298), Australia (n=153) and New Zealand (n=107). Each panel gives a two country comparison with an unlabelled diagram for an overall impression followed by an ST labelled diagram.



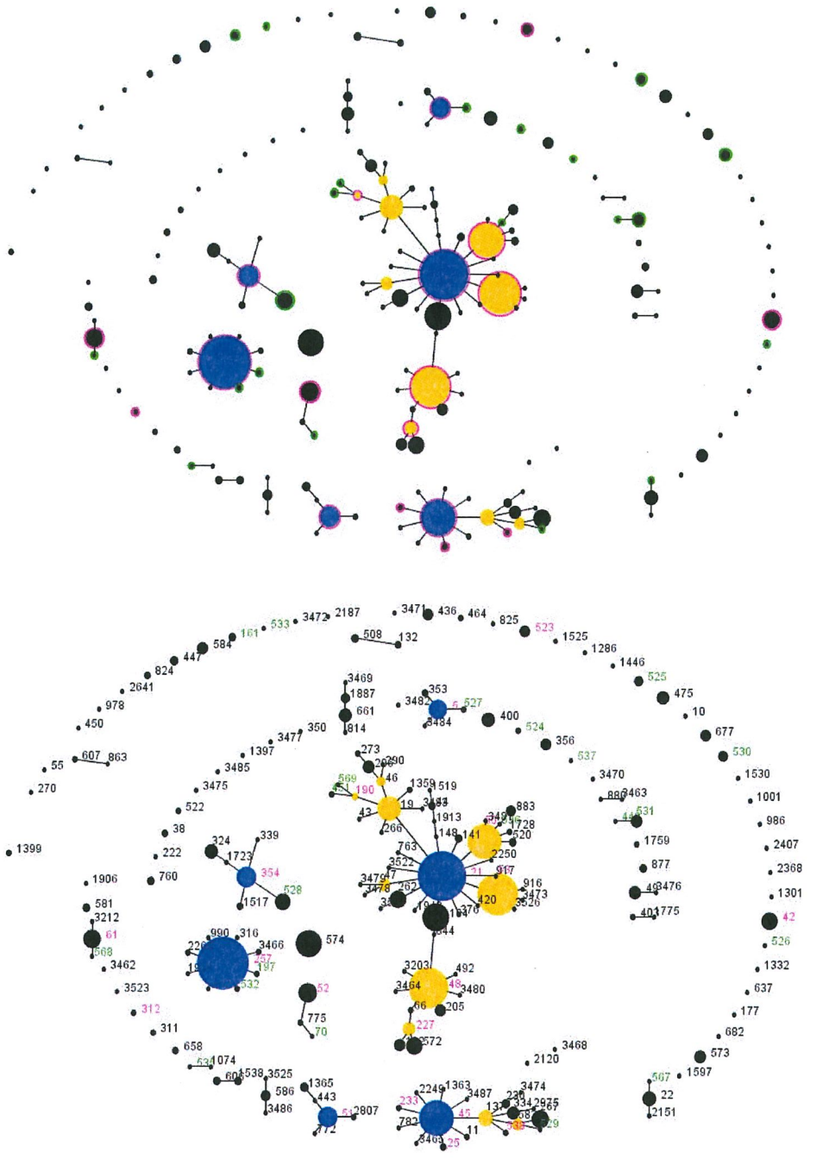
Annex 6. Figure A5.1. E-Burst diagram comparing isolates of *C. jejuni* from humans in Australia (no halo), New Zealand (green halo) and both (pink halo).



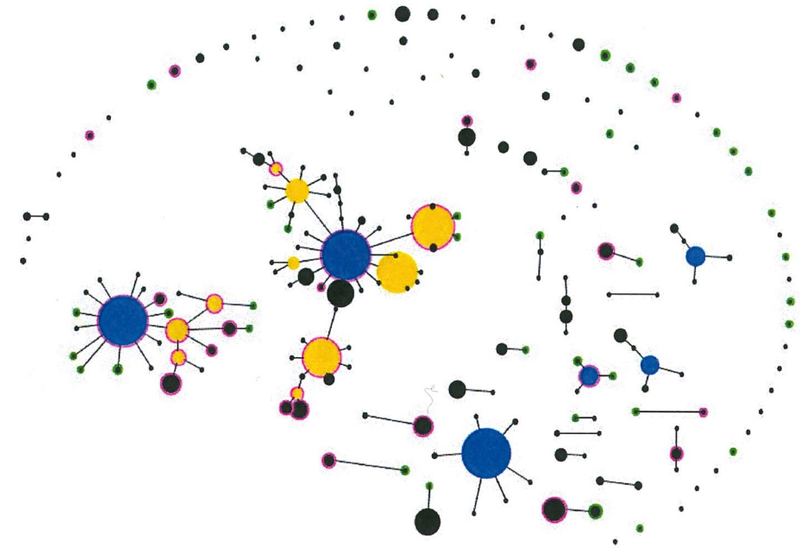
Annex 6. Figure A5.2. E-Burst diagram comparing isolates of *C. jejuni* from humans in Finland (no halo), Australia (green halo) and both (pink halo).



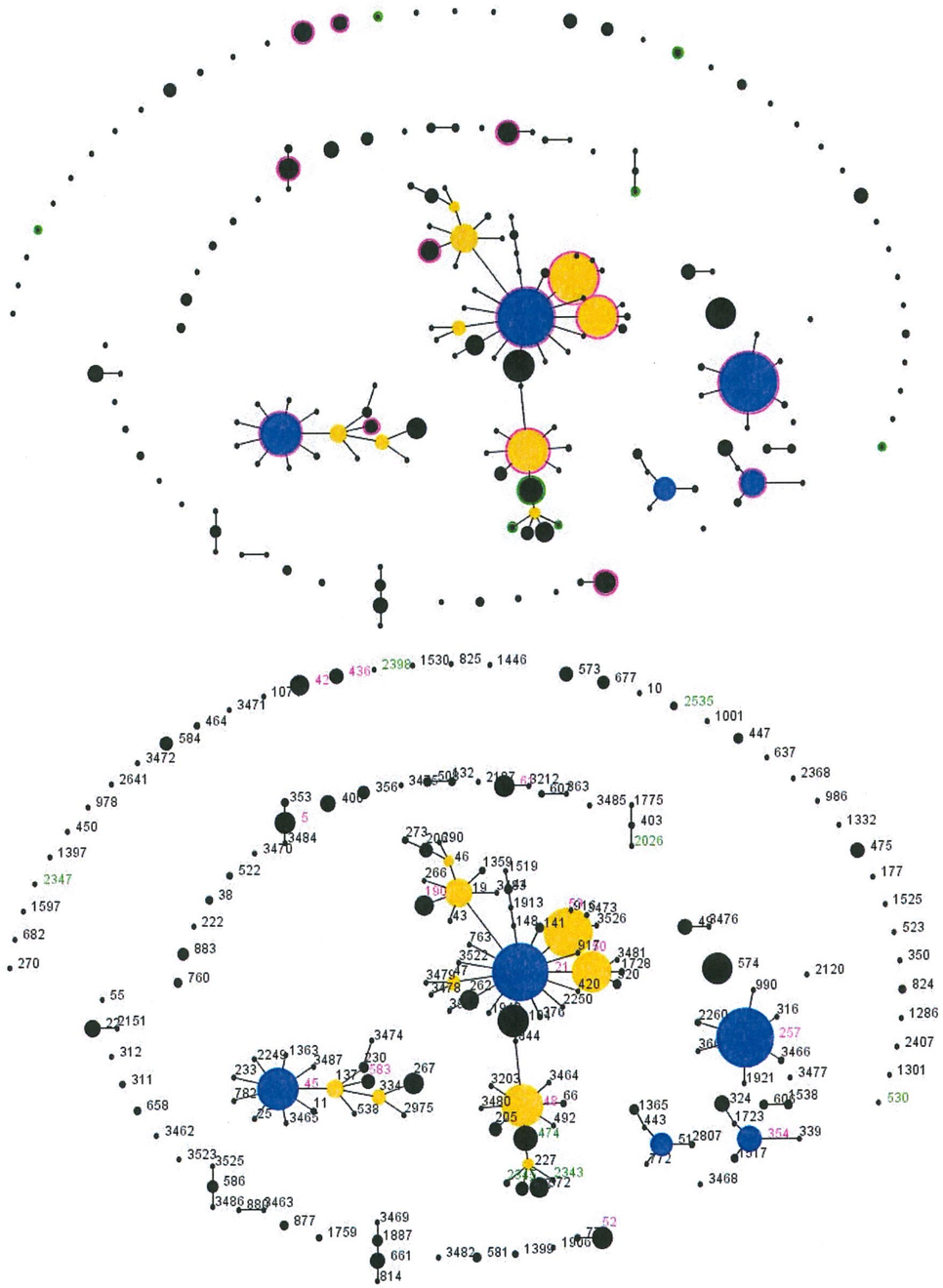
Annex 6. Figure A5.3. E-Burst diagram comparing isolates of *C. jejuni* from humans in Finland (no halo), New Zealand (green halo) and both (pink halo).



Annex 6. Figure A5.4. E-Burst diagram comparing isolates of *C. jejuni* from humans in England (CSSS) (no halo), Australia (green halo) and both (pink halo).



Annex 6. Figure A5.5. E-Burst diagram comparing isolates of *C. jejuni* from humans in England (CSSS) (no halo), Finland (green halo) and both (pink halo).



Annex 6. Figure A5.6. E-Burst diagram comparing isolates of *C. jejuni* from humans in England (CSSS) (no halo), New Zealand (green halo) and both (pink halo).

10.7 Annex 7 – Detailed data tables of clonal complex and ST distribution for retail poultry survey.

Table A6.1. Distribution of *Campylobacter jejuni* by clonal complex in retail poultry survey isolates.

Clonal complex	Number of isolates
ST-45 complex	185
ST-257 complex	137
ST-21 complex	121
ST-574 complex	42
ST-443 complex	39
ST-573 complex	30
ST-48 complex	25
ST-661 complex	19
ST-354 complex	17
ST-353 complex	14
ST-42 complex	14
ST-283 complex	11
ST-607 complex	10
ST-52 complex	9
ST-206 complex	6
ST-61 complex	6
ST-22 complex	5
ST-49 complex	5
ST-1034 complex	4
ST-508 complex	4
ST-433 complex	3
ST-692 complex	3
ST-1275 complex	1
ST-1287 complex	1
ST-446 complex	1
ST-677 complex	1
Unassigned	29
Total	742

Table A.6.2.B. Sequence types of *Campylobacter jejuni* in retail poultry survey isolates.

Sequence type	Number of isolates
45	88
257	88
51	32
21	30
573	30
574	29
50	18
137	17
48	16
583	14
25	13
334	13
53	12
354	12
262	10
267	10
5	9
17	9
42	9
661	9
19	8
104	8
607	8
814	8
44	6
61	6
233	6
1712	6
2030	6
447	5
584	5
52	4
22	3
49	3
132	3
586	3
845	3
2120	3
2223	3
2411	3

Sequence type	Number of isolates
63	2
206	2
266	2
372	2
400	2
475	2
775	2
782	2
844	2
883	2
1530	2
1700	2
1701	2
1709	2
1915	2
1993	2
2110	2
2254	2
2413	2
2414	2
2436	2
2437	2
8	1
11	1
46	1
188	1
190	1
201	1
227	1
286	1
294	1
359	1
366	1
441	1
443	1
464	1
489	1
560	1
571	1
572	1

Sequence type	Number of isolates
640	1
659	1
677	1
704	1
717	1
761	1
877	1
916	1
945	1
958	1
977	1
1231	1
1399	1
1400	1
1408	1
1509	1
1517	1
1519	1
1673	1
1702	1
1703	1
1792	1
1922	1
1924	1
1954	1
2033	1
2109	1
2121	1
2187	1
2211	1
2227	1
2231	1
2233	1
2249	1
2412	1
2415	1
2416	1
2417	1
2418	1
2419	1

Sequence type	Number of isolates
2420	1
2421	1
2422	1
2423	1
2424	1
2425	1
2426	1
2427	1
2428	1
2429	1
2430	1
2431	1
2432	1
2433	1
2434	1
2435	1
2438	1
2439	1
2440	1
2441	1
2442	1
2443	1
2444	1
2445	1
2446	1
2447	1
2448	1
2449	1
2450	1
2451	1
2452	1
2453	1
2454	1
2455	1
2456	1
2457	1
2458	1
2459	1
2467	1
2489	1

Sequence type	Number of isolates
2491	1
2492	1
2582	1
2583	1
2584	1
2585	1
2586	1
2589	1
2590	1
2591	1
2592	1
2593	1
2594	1
2595	1
2596	1
2597	1
2599	1
2600	1
2601	1
2602	1
2603	1
2604	1
2605	1
2606	1
2607	1
2609	1
2610	1
2611	1
2612	1
2613	1
2615	1
2622	1
Total	742

10.8 Annex 8. Temporal pattern among human and chicken isolates. (Figures for chapter 8)

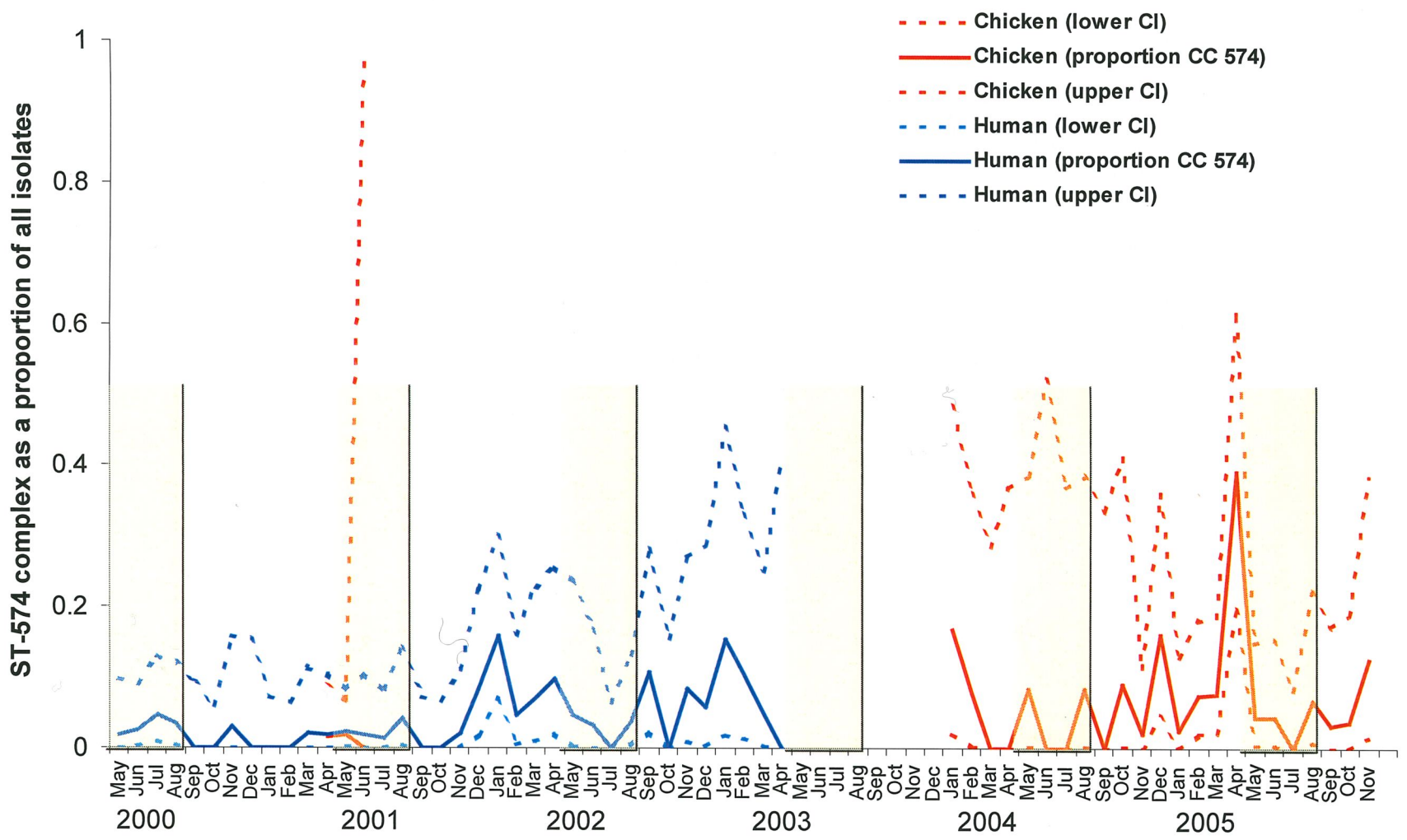


Figure A.8.1A. Proportion of isolates within ST-574 complex from human and retail poultry isolates over time with pointwise 95% confidence intervals. May to August highlighted in transparent cream to identify summer season.

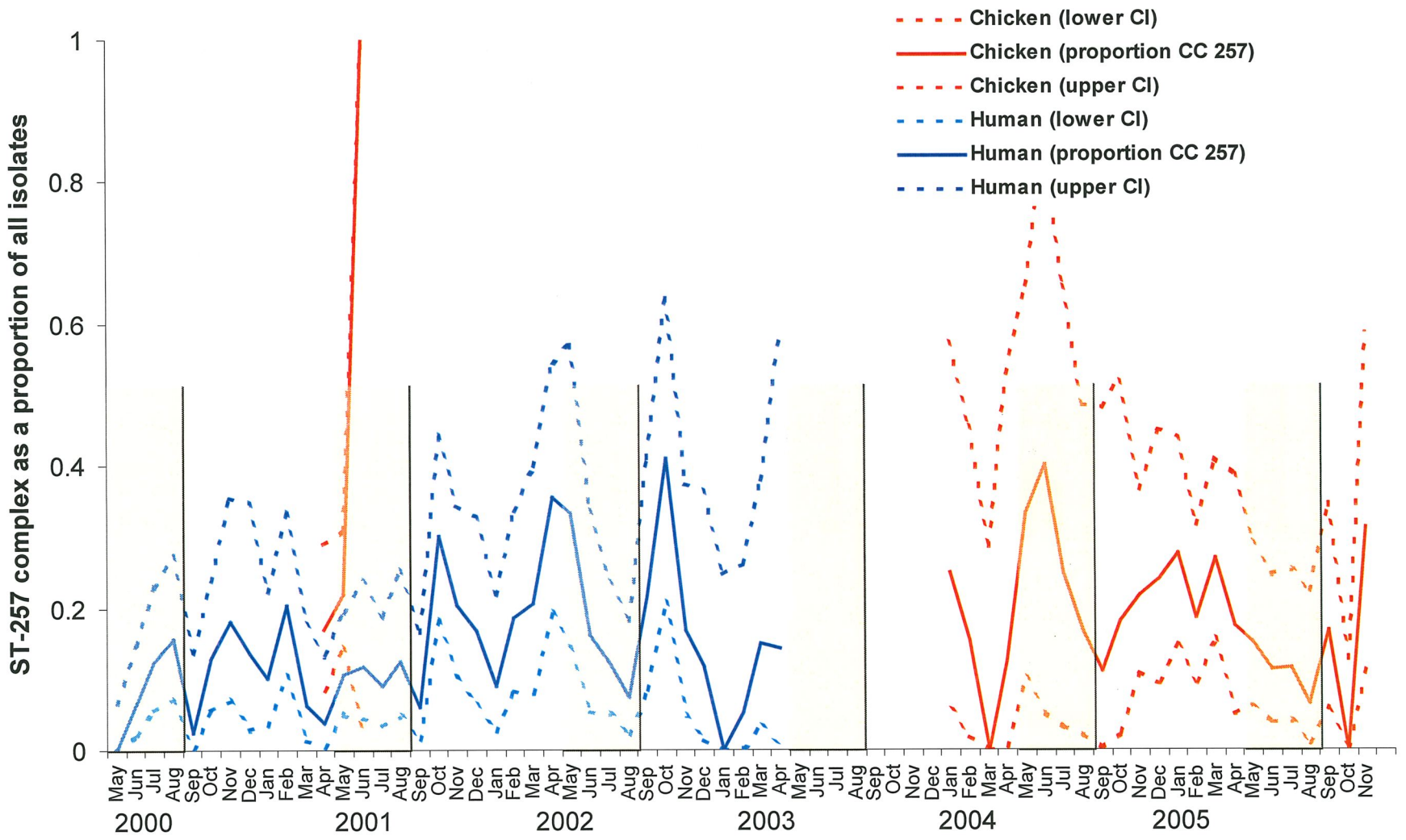


Figure A.8.1B. Proportion of isolates within ST-257 complex from human and retail poultry isolates over time with pointwise 95% confidence intervals. May to August highlighted in transparent cream to identify summer season.

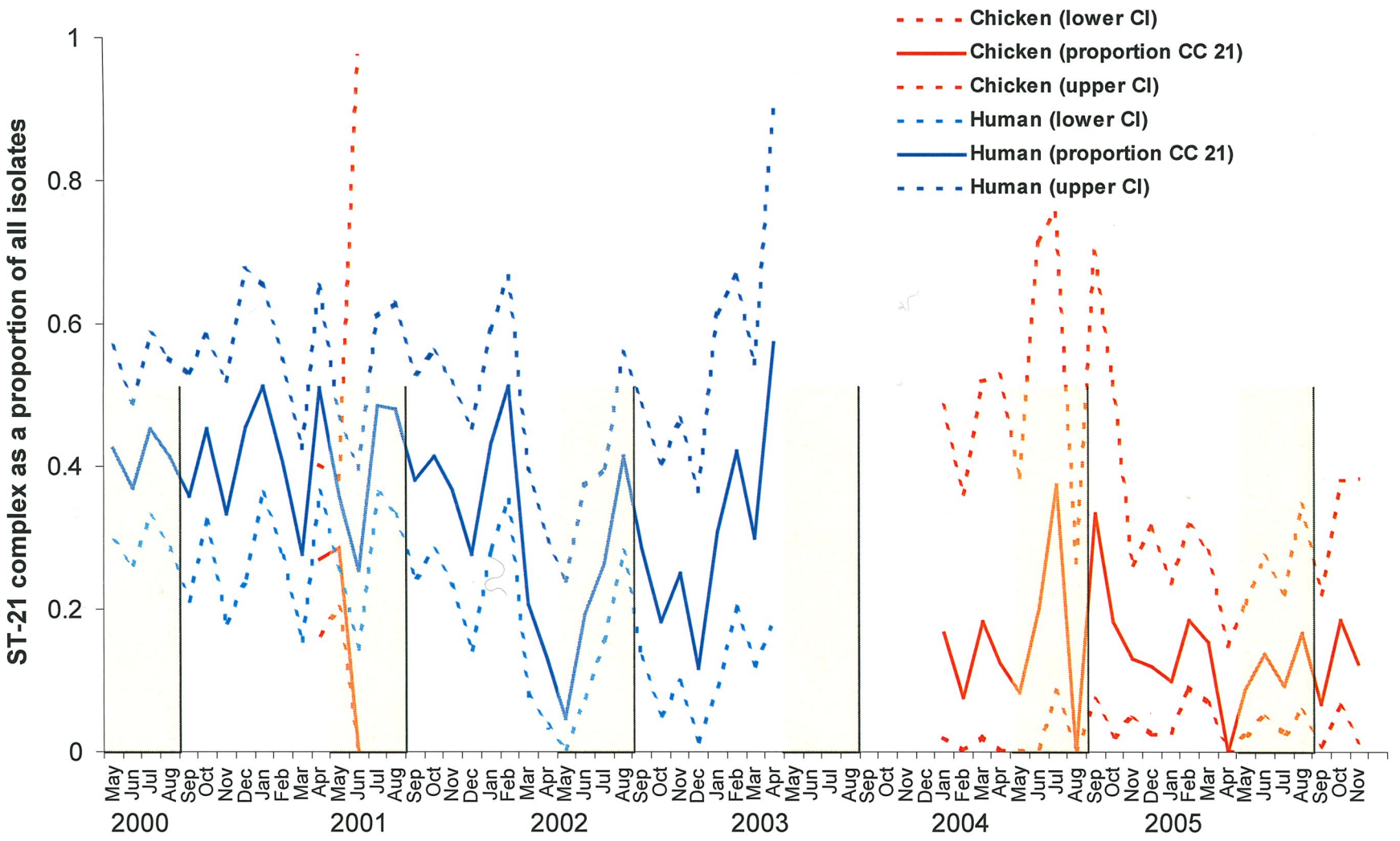


Figure A.8.1C. Proportion of isolates within ST-21 complex from human and retail poultry isolates over time with pointwise 95% confidence intervals. May to August highlighted in transparent cream to identify summer season.

10.9 Annex 9. ST distribution form human disease and animal hosts

Table A8.1. Number of cases of each ST in the CSSS dataset, predicted source of these STs using allele information in the STRUCTURE algorithm and distribution among animal hosts.

Clonal complex	ST	Human	Non-human reference populations					**Assigned** population
			Chicken	Goose	Starling	Ruminant	Total	
21	8		1			1	2	
	19	42	8			17	25	Ruminant
	21	174	30			32	62	Ruminant
	43	1					0	Ruminant
	44	4	6				6	Ruminant
	47	11					0	Ruminant
	50	80	18			3	21	Ruminant
	53	125	12			4	16	Ruminant
	104	54	8				8	Ruminant
	141	4					0	Ruminant
	148	1					0	Ruminant
	188		1				1	
	190	2	1				1	Ruminant
	193			1			1	
	201		1				1	
	260					1	1	
	262	20	10			22	32	Ruminant
	264					1	1	
	266	1	2			8	10	Ruminant
	268					1	1	
	272					1	1	
	274					1	1	
	275					1	1	
	280					1	1	
	282					1	1	
	300					2	2	
	376	1					0	Ruminant
	420	1					0	Ruminant
	489		1				1	
	520	4					0	Ruminant
	522	2				8	8	Ruminant
	559					1	1	
	561					1	1	
	615					1	1	
	616					1	1	
	621					1	1	
	635					1	1	
	640		1			1	2	
	641					1	1	
	642					1	1	
	760	4					0	Ruminant
	761		1				1	
	763	1					0	Ruminant
	797					1	1	
	806					1	1	

CC	ST	Human	Chicken	Goose	Starling	Ruminant	Total	Assigned
21	883	7	2				2	Ruminant
contd	916	1	1				1	Ruminant
	917	1					0	Ruminant
	1359	2					0	Ruminant
	1383					1	1	
	1519	1	1				1	Ruminant
	1728	2					0	Ruminant
	1913	1					0	Ruminant
	1943	1					0	Ruminant
	2110		2				2	
	2121		1				1	
	2233		1				1	
	2250	1					0	Ruminant
	2413		2				2	
	2414		2				2	
	2415		1				1	
	2417		1				1	
	2418		1				1	
	2419		1				1	
	2420		1				1	
	2422		1				1	
	2424		1				1	
	2457		1				1	
	3473	1					0	Ruminant
	3478	1					0	Ruminant
	3479	1					0	Ruminant
	3481	1					0	Ruminant
	3483	1					0	Ruminant
	3522	1					0	Ruminant
	3524	1					0	Ruminant
	3526	1					0	Ruminant
22	22	14	3			2	5	Chicken
	617					1	1	
	2151	1					0	Chicken
	2421		1				1	
	2615		1				1	
42	42	16	9		1	12	22	Starling
	269					1	1	
	279					1	1	
	447	5	5			1	6	UA
	575					1	1	
	590					1	1	
	629					1	1	
	632					1	1	
	634					1	1	
	638					1	1	
	643					1	1	

CC	ST	Human	Chicken	Goose	Starling	Ruminant	Total	Assigned
45	11	2	1			1	2	Chicken
	25	1	13				13	Chicken
	45	88	88		5	10	103	Chicken
	137	17	17	7		1	25	Chicken
	230	5				1	1	Chicken
	233	1	6				6	Chicken
	334	11	13		1		14	Chicken
	538	1					0	Chicken
	583	9	14			2	16	Chicken
	633					1	1	
	659		1				1	
	706			1			1	
	782	1	2				2	Chicken
	845		3				3	
	998				1		1	
	1025				1		1	
	1363	1					0	Chicken
	1509		1				1	
	1701		2				2	
	2109		1				1	
	2120	1	3				3	Chicken
	2223		3				3	
	2249	1	1				1	Chicken
	2411		3				3	
	2412		1				1	
	2423		1				1	
	2437		2				2	
	2454		1				1	
	2589		1				1	
	2590		1				1	
	2595		1				1	
	2599		1				1	
	2600		1				1	
	2604		1				1	
	2612		1				1	
	2613		1				1	
	3465	1					0	Chicken
	3474	1					0	Chicken
	3487	1					0	Chicken
48	38	3			2	6	8	Ruminant
	48	95	16			4	20	Ruminant
	66	2					0	Ruminant
	205	9					0	Ruminant
	281					1	1	
	475	11	2				2	Ruminant
	492	1					0	Ruminant
	541					1	1	
	798					1	1	
	804					1	1	

CC	ST	Human	Chicken	Goose	Starling	Ruminant	Total	Assigned
48	844	1	2				2	Ruminant
contd	2416		1				1	
	2427		1				1	
	2442		1				1	
	2446		1				1	
	2583		1				1	
	3203	1					0	Ruminant
	3464	1					0	Ruminant
	3480	1					0	Ruminant
49	49	11	3			2	5	Chicken
	2448		1				1	
	2582		1				1	
	3476	1					0	Chicken
52	52	20	4			1	5	Chicken
	63		2				2	
	261					1	1	
	775	2	2				2	Chicken
	796					1	1	
	2602		1				1	
61	61	22	6			33	39	Ruminant
	219					1	1	
	263					1	1	
	352					2	2	
	618					3	3	
	620					1	1	
	622					1	1	
	628					1	1	
	636					1	1	
	820					1	1	
	848					1	1	
	3212	1					0	Ruminant
177	144				3		3	
	177	1			46		46	Starling
	563				2		2	
	685				1		1	
	1004				3		3	
	1014				1		1	
	1381				2		2	
	1382				2		2	
	1388				2		2	
	1394				3		3	
	1482				1		1	
	1485				1		1	
	1500				1		1	
	1506				1		1	
	1533				1		1	

CC	ST	Human	Chicken	Goose	Starling	Ruminant	Total	Assigned
177 contd	1535				1		1	
179	220				3		3	
206	46	6	1				1	Ruminant
	122	9					0	Ruminant
	206	10	2			15	17	Ruminant
	222	1					0	Ruminant
	227	7	1				1	Ruminant
	271					2	2	
	273	2				3	3	Ruminant
	276					1	1	
	277					1	1	
	290	1					0	Ruminant
	471					1	1	
	562					1	1	
	572	20	1				1	Ruminant
	795					1	1	
	2610		1				1	
257	17		9				9	
	257	181	88		2	3	93	Chicken
	286		1				1	
	294		1				1	
	316	1					0	Chicken
	366	1	1				1	Chicken
	372		2				2	
	560		1			1	2	
	584	9	5				5	Chicken
	717		1				1	
	824	3					0	Chicken
	990	1					0	Chicken
	1712		6				6	
	1792		1				1	
	1921	1					0	Chicken
	1922		1				1	
	1924		1				1	
	2030		6				6	
	2254		2				2	
	2260	2					0	Chicken
	2425		1				1	
	2428		1				1	
	2447		1				1	
	2450		1				1	
	2451		1				1	
	2455		1				1	
	2456		1				1	
	2591		1				1	
	2594		1				1	
	2605		1				1	

CC	ST	Human	Chicken	Goose	Starling	Ruminant	Total	Assigned
257	2607		1				1	
contd	3466	2					0	Chicken
283	267	22	10		1		11	Chicken
	2444		1				1	
	2975	1					0	Chicken
353	5	22	9				9	Chicken
	353	3					0	Chicken
	356	8					0	Chicken
	400	13	2				2	Chicken
	581	4					0	Chicken
	2231		1				1	
	2429		1				1	
	2586		1				1	
	2641	1					0	Chicken
	3484	1					0	Chicken
354	324	13					0	Chicken
	339	1					0	Chicken
	354	25	12			2	14	Chicken
	359		1				1	
	1517	3	1				1	Chicken
	1723	1					0	Chicken
	2033		1				1	
	2438		1				1	
	2606		1				1	
	3468	1					0	Chicken
403	55	1				3	3	UA
	270	1				5	5	UA
	285					1	1	
	403	2					0	Geese
	1775	1					0	Geese
433	10	1					0	Chicken
	1408		1				1	
	1530	1	2				2	Chicken
443	51	27	32				32	Chicken
	443	1	1				1	Chicken
	772	1					0	Chicken
	1365	5					0	Chicken
	1703		1				1	
	2458		1				1	
	2459		1				1	
	2592		1				1	
	2593		1				1	
	2609		1				1	
	2807	2					0	Chicken

CC	ST	Human	Chicken	Goose	Starling	Ruminant	Total	Assigned
446	450	1					0	Chicken
	978	1					0	Chicken
	2489		1				1	
460	606	4					0	Chicken
	1074	1					0	Chicken
	1538	4					0	Chicken
508	132	3	3				3	Geese
	508	4					0	Geese
	2187	1	1				1	Geese
573	573	10	30				30	Chicken
574	571		1				1	
	574	50	29		1		30	Chicken
	1993		2				2	
	2433		1				1	
	2435		1				1	
	2436		2				2	
	2439		1				1	
	2441		1				1	
	2449		1				1	
	2585		1				1	
	2596		1				1	
	2601		1				1	
607	607	2	8				8	Chicken
	863	1					0	Chicken
	1915		2				2	
	3462	1					0	Chicken
658	311	2					0	Chicken
	312	1					0	Chicken
	523	1					0	
	658	3					0	Chicken
661	661	12	9				9	Geese
	814	1	8				8	Geese
	958		1				1	
	1400		1				1	
	1887	6					0	Geese
	3469	1					0	Geese
	3482	1					0	Chicken
677	677	8	1		2	1	4	Starling
	1024				1		1	
	1534				1		1	
682	681				1		1	

CC	ST	Human	Chicken	Goose	Starling	Ruminant	Total	Assigned
682	682	1			9		9	Starling
contd	686				15		15	
	687				1		1	
	818				11		11	
	1019				1		1	
	1020				63		63	
	1021				4		4	
	1022				13		13	
	1027				13		13	
	1385				1		1	
	1386				1		1	
	1387				1		1	
	1390				1		1	
	1391				1		1	
	1392				1		1	
	1503				2		2	
	1505				1		1	
	1507				1		1	
	1542				2		2	
692	692			4			4	
	699			7			7	
	707			6			6	
	1301	1					0	Geese
	1954		1				1	
	2584		1				1	
	2622		1				1	
702	702			18			18	
	703			1			1	
	705			1			1	
1034	694			2			2	
	697			1			1	
	698			1			1	
	711			2			2	
	788			1			1	
	977		1	1			2	
	1029			1			1	
	1033			16			16	
	1608			1			1	
	1709		2				2	
	2611		1				1	
1275	637	1					0	UA
	1231		1				1	
1287	945		1				1	
1332	696			9		1	10	

CC	ST	Human	Chicken	Goose	Starling	Ruminant	Total	Assigned
1332 cntd	1332	1					0	Geese
-	265					2	2	
-	284					1	1	
-	327					1	1	
-	350	1					0	UA
-	436	8			2		2	Chicken
-	441		1				1	
-	464	2	1				1	Chicken
-	586	7	3			1	4	Starling
-	619					1	1	
-	639					1	1	
-	683				14		14	
-	684				1		1	
-	691			5			5	
-	693			9			9	
-	695			5			5	
-	700			8			8	
-	701			8			8	
-	704		1	3			4	
-	708			1			1	
-	709			6			6	
-	710			13			13	
-	784					1	1	
-	785			1			1	
-	786			1			1	
-	787					1	1	
-	789			1			1	
-	877	4	1				1	Chicken
-	880	1					0	Chicken
-	986	1					0	Starling
-	993			1			1	
-	997				2		2	
-	999				1		1	
-	1000				1		1	
-	1001	1					0	Starling
-	1002				1		1	
-	1023				9		9	
-	1026				1		1	
-	1028			3			3	
-	1030			4			4	
-	1031			10			10	
-	1032			2			2	
-	1286	1					0	UA
-	1384				1		1	
-	1389				1		1	
-	1397	1					0	Geese
-	1399	2	1				1	Chicken
-	1481				1		1	
-	1483				1		1	

CC	ST	Human	Chicken	Goose	Starling	Ruminant	Total	Assigned
-	1484				1		1	
-	1486				1		1	
-	1501				1		1	
-	1502				1		1	
-	1504				1		1	
-	1508				1		1	
-	1525	1					0	Geese
-	1536				1		1	
-	1537				1		1	
-	1539				1		1	
-	1606			3			3	
-	1607			1			1	
-	1609				1		1	
-	1612				1		1	
-	1613				1		1	
-	1673		1				1	
-	1700		2				2	
-	1702		1				1	
-	1759	2					0	Chicken
-	1906	1					0	Chicken
-	2211		1				1	
-	2227		1				1	
-	2368	1					0	Geese
-	2407	1					0	Starling
-	2426		1				1	
-	2430		1				1	
-	2431		1				1	
-	2432		1				1	
-	2434		1				1	
-	2440		1				1	
-	2443		1				1	
-	2445		1				1	
-	2452		1				1	
-	2453		1				1	
-	2467		1				1	
-	2491		1				1	
-	2492		1				1	
-	2597		1				1	
-	2603		1				1	
-	3463	1					0	Chicken
-	3470	1					0	
-	3471	1					0	Chicken
-	3472	1					0	Chicken
-	3475	1					0	Geese
-	3477	1					0	Starling
-	3485	1					0	Geese
-	3486	1					0	Starling
-	3523	1					0	Chicken
-	3525	1					0	Starling
Total		1505	742	166	285	273	1466	