i-CaMPS
impact of interventions - *Campylobacter* MLST Project in Scotland

Division of Applied Medicine
School of Medicine and Dentistry
University of Aberdeen
Foresterhill
Aberdeen AB25 2ZD
United Kingdom

Tel: +44 (0)1224 437023
Email: CaMPS@abdn.ac.uk

Food Standards Agency- Scotland
Contract S14054

"Employing Source Attribution and Molecular Epidemiology to measure the impact of interventions on human campylobacteriosis in Scotland"

Final Report for April 2011 – March 2012

Contents

Figures

Tables

1. Introduction

1.1 Background

Campylobacter are Gram-negative bacteria that live commensally in the gastrointestinal tracts of a wide range of animals and birds, including farmed species and companion animals. Some *Campylobacter* species are also zoonotic human pathogens. A typical human infection consists of a self-limiting bout of diarrhoea, abdominal cramps and fever lasting about five days. *Campylobacter* infection was implicated in causing human enteritis in the late 1970s (18), and has since become recognised as the commonest known cause of bacterial infectious intestinal disease (IID) worldwide. According to WHO estimates, *Campylobacter*-related illness affects around 1% of populations in developed countries every year.

Campylobacter infection causes almost half of all IID cases in the UK, with *Campylobacter jejuni* causing around 90% of cases and the closely-related *Campylobacter coli* causing almost all the rest. In 2012 a total of 6321 isolates of *Campylobacter* were reported in Scotland, which was a slight decrease compared to the 6363 reports in 2011. In recent years numbers had peaked in 2010 at 6597 up from the 2004 low of 4365 (Figure 1). Because there is substantial under-reporting, the actual number of cases is likely to be closer to 500,000(20). Further, about 10% of reported cases are hospitalised. In Scotland the overall rate of *Campylobacter* infection in 2012 was 120.9 per 100,000. Among the mainland NHS boards the lowest rate of 77.3 per 100,000 was in Fife, which is historically low. The highest rate of 160.4 per 100,000 was observed in Tayside, with Grampian having a rate of 143.1 per 100,000 (Figure 2).

Most cases of *Campylobacter* are apparently sporadic with few identified outbreaks. There was one outbreak of *Campylobacter* in 2012 reported to ObSurv (the surveillance system for all general outbreaks of IID in Scotland) which is typical; in the previous 17 years since the start of ObSurv there had been 34 reported general outbreaks of *Campylobacter*.

High rates of *Campylobacter* incidence translate into substantial annual economic costs, estimated at £503M in the UK (all likely cases) (9), EUR9M in the Netherlands (reported cases in 1999) (22), and $4.3bn in the USA (all likely cases) (1). *Campylobacter* infection can also lead to serious longer-term illness. Approximately one case for every 1000 reported cases leads to Guillain-Barré syndrome: a serious condition of reversible or permanent loss of limb motor function that is the commonest cause of acute flaccid paralysis. *Campylobacter* infection is also associated with the non-paralytic version of GBS, Miller-Fisher syndrome, and with reactive arthritis.

The i-CaMPS 2010-11 annual report established a number of methodological parameters:

- Human campylobacteriosis incidence in Scotland is modelled well by Grampian Region.
- The five variant molecular attribution models used gave broadly the same source attribution results, with Asymmetric Island showing a higher proportion of attribution to chicken. Accordingly this study has focused on the STRUCTURE with alleles and Asymmetric Island models.

- The reference host source isolate datasets will continue to be extended by the addition of contemporaneous isolates.
- Attribution in Grampian mirrors that of Scotland, and so can be used as a proxy to interpret Scottish case data.

The main source of human campylobacteriosis in Scotland(4,17) and elsewhere in the developed world is retail chicken with a significant proportion of the remainder attributable to ruminants(2,12,13,19,21). Both the UK and Scottish governments have a responsibility to promote health and minimise logistic burden on the health care sector, and therefore want the incidence of human *Campylobacter* infection substantially reduced. Human *Campylobacter* infection is viewed as having a significant food-borne component based on the best available evidence(3), and therefore food safety regulation bodies and organisations in the food production sector are best-placed to identify and implement effective interventions. The 'Joint Working Group on *Campylobacter*' was established in August 2009 as a joint industry and government group (www.food.gov.uk/safereating/microbiology/campylobacterevidenceprogramme/wgcampy). It aims to identify interventions that would reduce *Campylobacter* in chicken. The membership includes the British Poultry Council (BPC), the National Farmers' Union (NFU) the British Retail Consortium (BRC), the FSA and Defra. Their aim is to identify and put in place interventions that will reduce *Campylobacter* through a Joint Action Plan. The key activities of the action plan relate to on-farm, transport, processing, retail, consumer and catering sector trials and interventions, as well as surveillance and monitoring. The present study will address aspects of surveillance and monitoring by seeking to clarify the sources of human campylobacteriosis in Scotland during 2011-12 which in 2005-07(4) were determined to be principally retail chicken with a significant proportion of the remainder attributable to ruminants. It will continue base line data started in April 2010(5) of campylobacteriosis and the molecular attribution of source of these clinical isolates which can be used to monitor the success of the other elements of the Joint Action Plan.

Figure 1. Annual incidence of campylobacteriosis in UK.

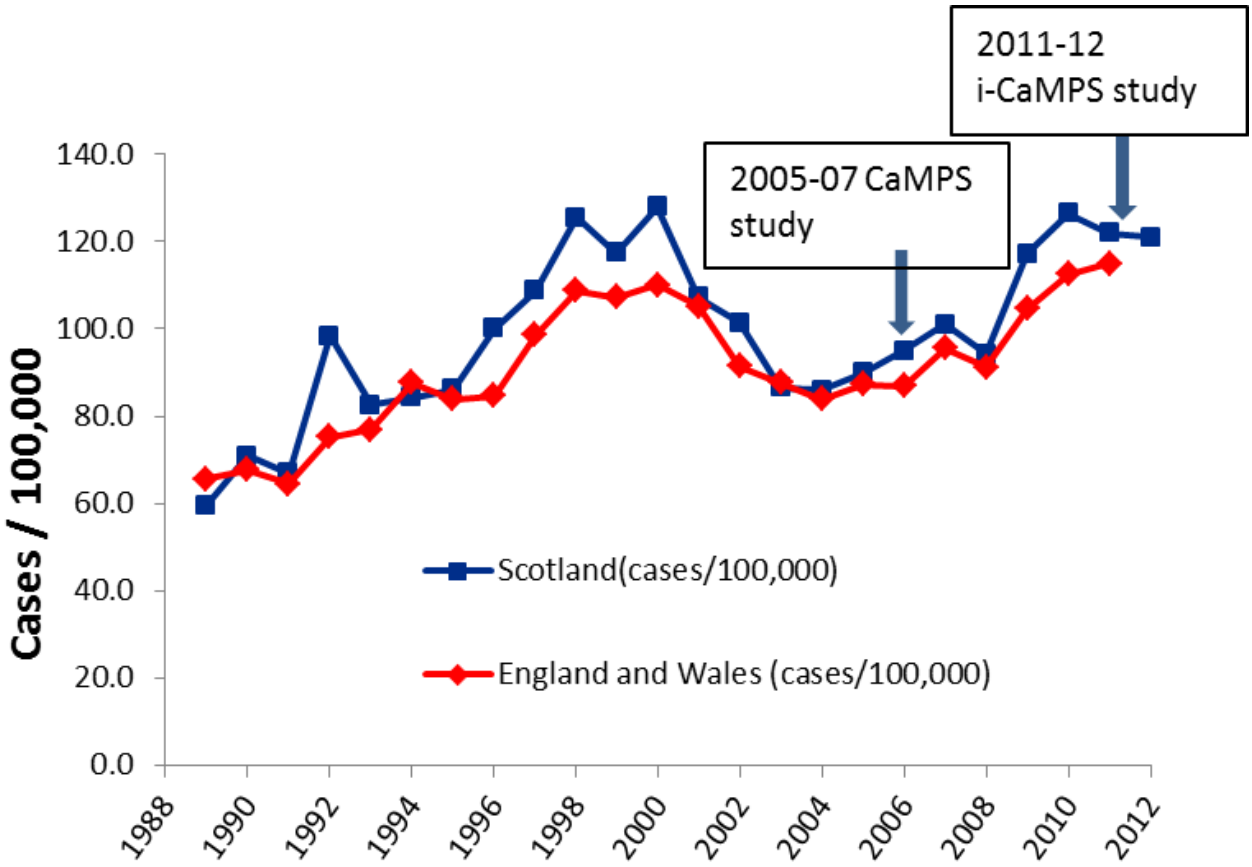
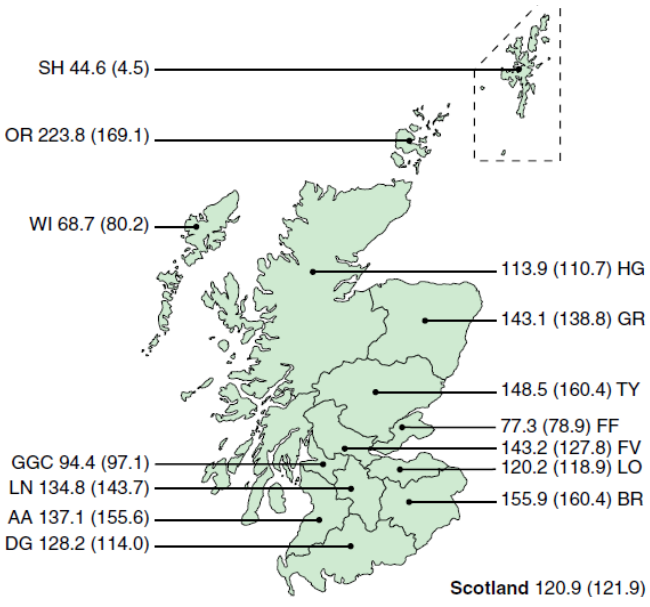


Figure 2. Incidence per 100,000 population of reports of *Campylobacter* infection 2012 (2011).



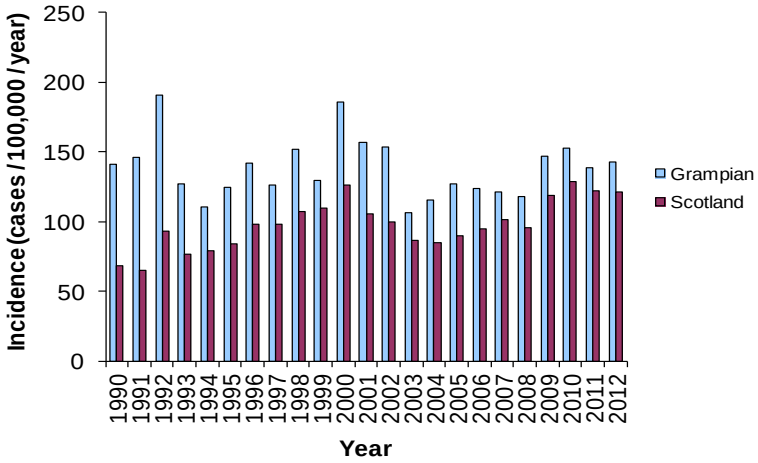
Data from Health Protection Scotland

1.2 Campylobacteriosis in Scotland and Grampian.

The trends in overall incidence of Scottish campylobacteriosis continue to be broadly mirrored by those in Grampian (Figure 3).

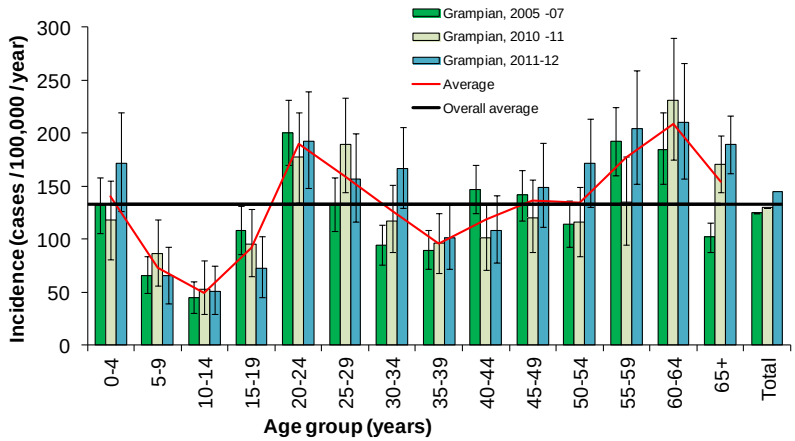
Within Grampian (Figure 4), age stratification generally showed no change in incidence within the age groups (exceptions: 30-34 where there is an increase in 2011-12 compared with 2005-07; 65+ where there is an increase in 2010-11 &2011-12 compared with 2005-07). There continues to be four age groups (5-9, 10-14, 15-19, 35-39) which have lower incidence than the overall average (133/100,000) of the three periods (P<0.0001) and two age groups (20-24, 60-64) which have higher incidence than the overall average (P<0.0001)

Figure 3. Incidence of campylobacteriosis in Scotland and Grampian from 1990 to 2012.



Data from Health Protection Scotland.

Figure 4. Age structured incidence of campylobacteriosis in Grampian for 2005-07 and 2010-12.



Data from Health Protection Scotland. Error bars are 95% CI.

1.3 Aims

This Research Requirement seeks to estimate the proportion of clinical *Campylobacter* isolates that are attributable to retail chicken sources and to compare this with the previous CaMPS study of 2005-7 and the i-CaMPS 2010-11 study. This attribution is dependent on appropriate source isolates typed by MLST. The CaMPS study(4,17) identified that those from chicken, cattle and sheep were of greatest relevance. This Research Requirement will establish baseline data against which the success of future interventions, over a number of years, at many points along the 'farm to fork' pathway to chicken consumption will be measured. It is therefore important that this baseline dataset includes contemporaneous chicken, cattle and sheep isolates. These will all be sourced predominantly from Grampian which we have shown previously to be typical for cattle and sheep strains(16) when compared with other Scottish regions, whilst retail chicken is both sourced and distributed all around the UK (our survey of abattoir locations displayed on retail chicken products in Grampian shows these to be sourced from across the UK).

As in the previous studies molecular source attribution will be performed using 7-locus MLST. An innovative approach has been adopted in the present study by employing a next-generation, whole genome sequencing strategy with a bioinformatics pipeline. This allowed the extraction of the 'classical' 7-locus MLST data whilst also making available the nearly complete genomes of the 1000+ isolates studied and thus an invaluable future data resource.

2. Materials and Methods

2.1 Isolate Collections

All available clinical isolates in Grampian for the 12 month period 1 April 2011 – 31 March 2012 (n=783) were collected (Table 1).

Contemporaneous *Campylobacter* isolates from the principal source hosts were also collected (Table 1). Retail chicken was sourced from shops around Aberdeen. Cattle and sheep faecal samples were collected by FSA Operations staff at Portleithen abattoir on a regular basis and couriered to our labs; the sources of the originating animals was selected to be predominantly from NE Scotland.

Isolation and culture was carried out as described previously(4).

Table 1. Number of specimens collected, number of presumptive *Campylobacter* spp. isolated, number of MLST-confirmed *Campylobacter* spp, number of MLST 7 locus isolates.

	Specimens collected	*Campylobacter* positive specimens[a]	*Campylobacter* prevalence[a]	Isolates with 7-locus MLST[b]
Human	**684**	**-**	**-**	**634**
Cattle	301	169	56%	91
Sheep	209	142	68%	95
Chicken	243	230	95%	176

a: Confirmed *Campylobacter* spp. by latex agglutination test
b: Not all isolates were MLST typed

2.2 MLST of isolates

The isolation of genomic DNA which is suitable for WGS used the Promega Wizard Genomic DNA Purification Kit (Catalogue # A1125).

For Genome Sequencing, DNA extracts were sequenced using an Illumina HiSeq sequencer using 100 nt paired-end sequencing in a 96-plex format using bar-coded tags for each sample. The paired read files were de novo assembled using the Velvet assembler in an established pipeline at University of Oxford. The sequences were imported into BIGSdb (Bacterial Isolate Genome Sequence Database(11)), a bioinformatic pipeline developed at University of Oxford.

BIGSdb is software designed to store and analyse sequence data for bacterial isolates. Any number of sequences can be linked to isolate records - these can be small contigs assembled from dideoxy sequencing through to whole genomes (complete or multiple contigs generated from parallel sequencing technologies such as 454 or Illumina Solexa). All the functionality of mlstdbnet and agdbnet has been incorporated into BIGSdb and this software will be used to eventually host all the databases on the PubMLST.org site. BIGSdb extends the principle of MLST to genomic data, where large numbers of loci can be defined, with alleles assigned by reference to sequence definition databases (which can also be set up with BIGSdb). Loci can also be grouped into schemes so that types can be defined by combinations of allelic profiles, a concept analogous to MLST.

The whole genome sequences of the isolates were used to classify them into strain types using 7-locus MLST(10). Allele numbers and sequence types (ST) will be assigned using the public *Campylobacter* PubMLST database http://publmst.org/campylobacter/.

Multi-locus Sequence Typing was carried out on all isolates and this is summarised in Table 1. Not all presumptive isolates were confirmed to be *Campylobacter jejuni/ coli* by MLST and this was most probably due to the difficulty of achieving this by visual inspection of colonies and latex sero-agglutination testing.

2.3 Host reservoir isolate datasets

The poultry, cattle and sheep data were compared with that obtained in CaMPS 2005/6(4) using Nei's genetic distance(8) and rarefaction to establish whether the species data can be combined from the two different years. This will also provide evidence of the stability or otherwise of sources over time.

To maximise the use of available source datasets, typed isolates from the 2005-06 Scottish study; clinical isolates from the overlapping 27 month period July 2005 -Oct 2007 (n=1452) from Grampian and isolates from the i-CaMPS 2010-11 period, in addition to the isolates from the current period were used in the molecular attribution analyses (Table 2).

Table 2. Isolate datasets.

Host dataset 1. 2005-06 Scottish-wide hosts

Host		*C.jejuni*	*C.coli*	TOTAL
Cattle	2005 -06	336	25	361
Sheep	2005 -06	91	56	147
Chicken	2005 -06	255	47	302
Wild Birds	2005 -06	176	12	188
Pigs	2005 -06	7	33	40

Host dataset 2. 2010-11 Grampian-wide hosts

Host		Total
Cattle	2010 -11	77
Sheep	2010 -11	100
Chicken	2010 -11	181
Wild Birds	2005 -06	as above
Pigs	2005 -06	as above

Host dataset 3. 2011-12 Grampian-wide hosts

Host		Total
Cattle	2011 -12	91
Sheep	2011 -12	95
Chicken	2011 -12	176
Wild Birds	2005 -06	as above
Pigs	2005 -06	as above

Host dataset 1+2+3. Combined 2005-07 Scottish-wide plus 2010-12 Grampian-wide hosts

Host		Total
Cattle	2010 -12 & 2005 -06	529
Sheep	2010 -12 & 2005 -06	342
Chicken	2010 -12 & 2005 -06	659
Wild Birds	2005 -06	188
Pigs	2005 -06	40

Clinical isolate datasets.

Period	Region	Total
2005 -06	Scotland	5674
2005 -07	Grampian	1452
2010 -11	Grampian	697
2011 -12	Grampian	600

2.4 Molecular attribution methods

Attribution by microbial sub-typing is a relatively new area of research. The term "source attribution" has been defined(14) as: "…the partitioning of the human disease burden of one or more foodborne infections to specific source, where the term *source* includes animal reservoirs and vehicles (e.g. foods)."

Further, the microbial subtyping methodology uses the distribution of subtypes in each of the sources and compares this with that found in humans. This can be done in terms of simple proportions (e.g. the Dutch model) or using Bayesian stochastic methods (e.g. STRUCTURE). Currently, there are 5 main techniques for attributing disease on a population level using microbial sub-typing(2). Three of these methods will be used in the current study (Table 3) and are detailed below.

The Dutch Model (6) is a straight forward way to estimate the attribution of a particular genotype (e.g. ST) to a reservoir, when the frequency distribution of each type is known for each reservoir. If p_{ij} represents the frequency of type *i* (eg ST 19) in source *j* (e.g. poultry) then the proportion of attribution of type *i* in source *j* is given by

$$\lambda_{ij} = \frac{p_{ij}}{\sum_{j} p_{ij}}$$

where the summation by *j* considers all the reservoirs where data exist (e.g. cattle, sheep, wild birds, poultry etc.).

When applied at ST level this model does not guarantee that all STs will be attributed to sources. This is because human types that are not found in the animal reservoir cannot be attributed. However, if genetic information exists at multiple loci as in this study, then the Dutch Model can make use of the frequency of each individual allele at each individual locus, and estimate attribution even for STs that are not present in the animal reservoirs. In particular, at allele level the frequencies $p_{a_{ijk}}$ can be calculated for each allele a_{ijk} of all isolates from the animal reservoirs. Where i is subtype, j source and k the loci number.

The attribution score of bacterial subtype *i* in source *j* is

$$\lambda_{ij} = \frac{\prod_{k=1}^{7} p_{a_{ijk}}}{\sum_{j}\left(\prod_{k=1}^{7} p_{a_{ijk}}\right)}$$

where $p_{a_{ijk}} = BetaInv(0.5, 0+1, N_{isolates}+1)$ if its frequency is zero (*BetaInv* fn in Excel). This assumes that we have no prior knowledge of $p_{a_{ijk}}$ and so is maximally noncommittal or conservative.

The Dutch Model does not take into account the uncertainty in the frequency distribution of genotypes. It does not consider any information about the exposure of humans to sources or the viability/virulence of pathogens once they are ingested by humans.

STRUCTURE (15) is a Bayesian clustering model designed to infer population structure and to attribute individuals to population groups. The program can use MLST genotyping data. Each isolate is attributed on the basis of a training dataset consisting of isolates from known populations (i.e. set USEPOPINFO to 1). The algorithm calculates the frequency of each particular sequence type in each population taking into account the uncertainty due to the sample size. Based on these frequencies the probability to belong to a population group/reservoir is calculated, following multiple iterative steps (Markov chain Monte Carlo - MCMC) for the estimation of frequencies. The programme has the option to consider the allele independent (no-admixture model – independent alleles) and starts with equal frequencies for each isolate type. Following an initial number of MCMC burn-in steps (e.g. 1000) further iterations (e.g. 10000) are used for estimation of the probabilities that an isolate belongs to each particular population being considered (eg cattle, sheep, poultry etc.). To enable the largest reference dataset to be used (often datasets are small due to the cost of typing many isolates) only one ST is selected at a time from the unknown dataset by using the jacknife method. This process is repeated to enable multiple estimations of the same sequence type so that uncertainty in the attribution scores can be determined.

STRUCTURE can be used at ST or allele level, it incorporates uncertainty and takes account of sample size. Hence, in principal it gives a more realistic estimation of the attribution to a specific reservoir than the Dutch Model. Also, like the Dutch Model at allele level it can assign human cases that have STs that are not found in the animal reservoirs. However it is highly time consuming and does not consider any exposure to risk factors or the viability of pathogens.

The Asymmetric Island (AI) Model (23) incorporates a Bayesian approach and uses the allelic profile of the sequence subtypes to reconstruct the genealogical history of the isolates. The host populations are considered to exist on separate "islands" (e.g. the sheep island). Mutations and recombination occur on each island. Migrations from between each reservoir (island) and into the human population are used to estimate the degree of attribution to each source. This model has been applied to MLST data from England(23), Scotland(17) and New Zealand where 56%, 78% and 75% of human cases were attributed to poultry respectively.

The Asymmetric Island model incorporates recombination and mutation, uses MLST data at the allele level and achieves relatively high values for self-attribution. However, the model appears to be complicated and the current explanations of its operation difficult to comprehend. The Asymmetric Island model assigns each human case to the potential source populations on the basis of DNA sequence similarity. By comparing human isolates to a panel of reference sequences of known source (e.g. cattle, sheep, chickens, pigs, wild birds and the environment), each human case can be assigned a probability of originating in each source population. The source attribution probabilities are calculated using a statistical model of the way the DNA sequences evolve in the populations of bacteria. In the statistical model, there are parameters representing the processes of mutation, DNA exchange between bacteria (recombination or horizontal gene transfer) and zoonotic transmission between populations. These processes lead to differences in gene frequencies between the source populations, facilitating source attribution. The model can be trained, by estimating the parameters exclusively from the sequences of known source, before using it to calculate source attribution probabilities for human sequences.

Table 3. Molecular attribution models used.

Model	Genetic unit of assessment	
	ST	Allele
Dutch proportional	√	√
STRUCTURE	√	√
Asymmetric Island		√

3. Results and Discussion

3.1 Has the prevalence of *Campylobacter* in food and animal reservoirs changed over time?

The prevalence of *Campylobacter* in the different reservoirs sampled and in retail chicken shows a continuing increase in prevalence for all three. This increase might be due to subtle changes in laboratory protocols or to differing staff, or to real increases over the eight year period. If true, it might provide an explanation for the increase in clinical cases over the last few years.

Table 4. *Campylobacter* prevalence in 2005 -06, 2010 -11 and 2011 -12 in cattle, sheep and retail chicken.

Reservoir	2005 -06 +ve/total (%)	2010 -11 +ve/total (%)	[a] 2011 -12 +ve/total (%)
Cattle	104/474 (22)	47/142 (33)	301/169 (56)
Sheep	97/292 (33)	88/167 (53)	209/142 (68)
Chicken	142/222 (64)	215/238 (90)	243/230 (95)

[a]Taken from Table 1

3.2 Do strain types change over time?

The extent to which the isolates from sources represented the maximum hypothetical diversity was characterised using rarefaction. Rarefaction is a data re-sampling technique that indicates whether diversity has reached a plateau or is still rising at the total sample size, i.e., at the end of collection. A rarefaction curve that has reached a plateau indicates that all diversity (i.e. all MLST genotypes) has been sampled whereas an increasing slope indicates that some diversity remains unsampled (i.e. there are likely to be MLST types in the reservoir that have not yet been sampled). This method assumes that the dataset represents a random sample taken from a closed system characterised by a constant, stable spectrum of types. As in the large 2005 -06 study(4) and the 2010-11 study(5), the rarefaction curves for all clinical, environmental and food sources were still rising, even at the maximum sample sizes (Figure 5 a, b). This is because the system being studied is open to immigration (e.g. for human clinical strains there will be immigration by foreign travel) and also that the sampling size is not sufficiently large to be comprehensive. For all isolates over the seven year period comparison of 2005-07 with 200-12 (Supplementary Figure 1) suggests that there is a significant reduction ($p<0.020$) in diversity of cattle isolates over the period (Supplementary Figure 1,b) in contrast to no significant difference for sheep ($p=0.174$) and chicken ($p<0.084$) and a just significant difference for clinical isolates ($p<0.043$). The reasons for these differences are not clear, but may be due to sampling biases. Both the clinical and chicken strains have similar levels of diversity. However, the cattle and sheep strains exhibited less diversity (for both the 2005-06 and 2010-11 studies) than those from retail chicken and human clinical strains. Collectively, over the all study periods (Figure 5 c) the diversity of clinical isolates is quite similar to the aggregated diversity of isolates from all sources ($p=0.191$), however it should be remembered that a proportion of the food and environmental isolates are likely to be non-pathogenic to humans.

It was apparent visually that the proportions of animal and clinical strains were changing with time (Figure 6). Comparing the 27 month period to October 2007 with the twelve month periods to March 2011 and to March 2012 (Supplementary Table 1) there were statistically significant changes in clinical strain abundance. Eight ST (ST5, ST21, ST22, ST42, ST50, ST464, ST1044, ST5136) increased statistically significantly in abundance over the seven year period with two (ST1044, ST5136) first appearing during this period. Seven ST (ST48, ST137, ST257, ST354, SR475, ST574, ST2030) decreased significantly in abundance over the seven year period.

These descriptive changes in the *Campylobacter* population which were apparent at a strain level were also examined by calculating, Nei's, genetic distance between isolates from each source from the two study periods (Table 6). Nei's genetic distance is a measure of the overlap in the genetic content of populations and this was measured at both strain level (a single measure of similarity using ST number) and at allele level (similarity measured across the seven MLST loci). Again significant differences were observed both between hosts and between the study periods. The clinical isolates were always most similar (the genetic distance is the smallest) to the chicken isolates. The cattle and sheep isolates also tended to be more similar to each other than to chicken or clinical isolates.

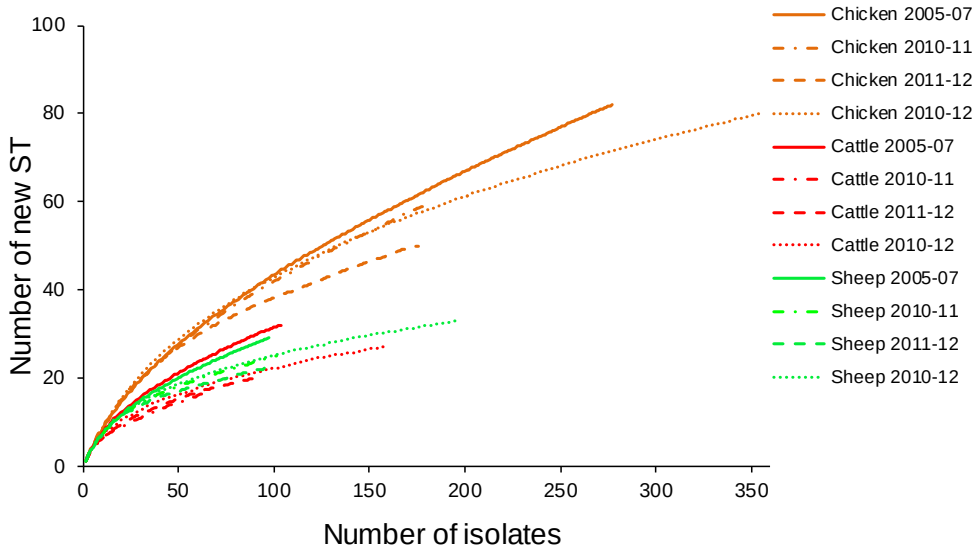
The most parsimonious explanation of this strain diversity is that *Campylobacter* present in clinical, environmental and food sources in Scotland represents an

extremely large pool of strains that is continually being augmented: internally by mutation and recombination and externally by strain input from human travel and migrating wildlife.

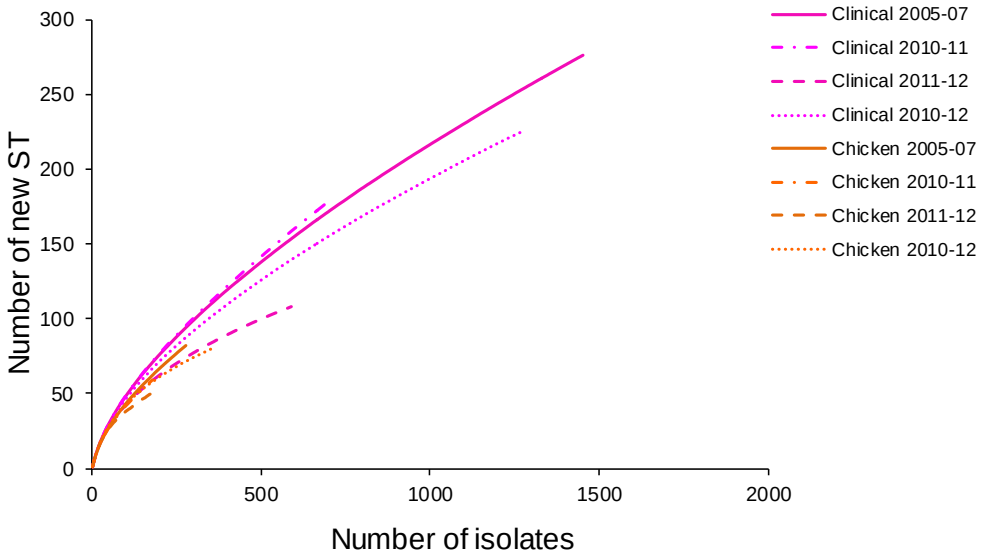
Since the Nei genetic distance findings imply that the host datasets are only somewhat genetically similar, then combining datasets for a particular host from several periods may be problematical, however small sample size (cf rarefaction) will have contributed to this. Accordingly and as previously, attribution analyses have used all available host datasets as indicated in Table 2 and in the analyses following. The genetic distances were always smaller when determined using the allele rather than the ST datasets suggesting that attribution analyses using allele level data should be more refined.

Figure 5. Rarefaction (saturation) analysis.

(a)



(b)



(c)

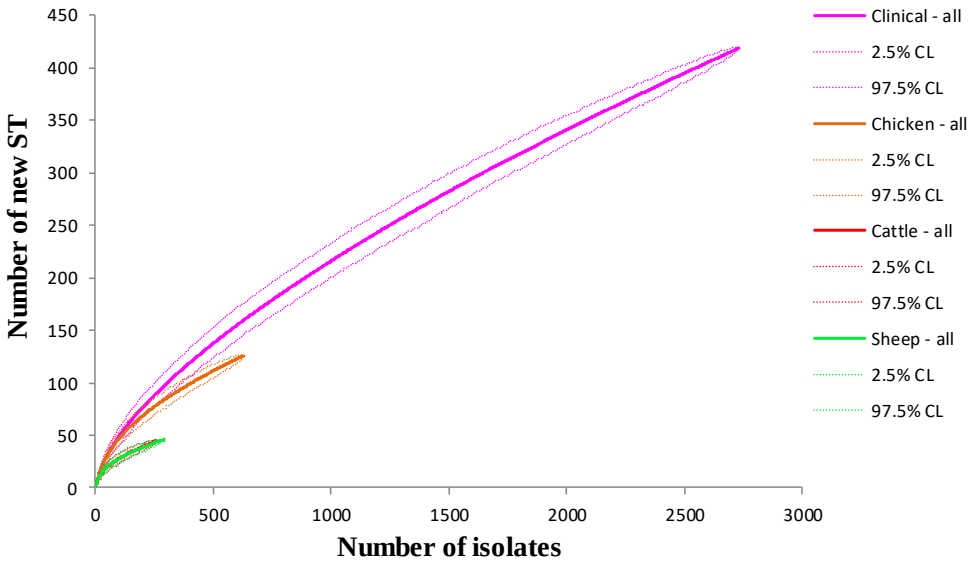


Figure 6. Observed changes over time in abundance of Sequence Types by Reservoir.

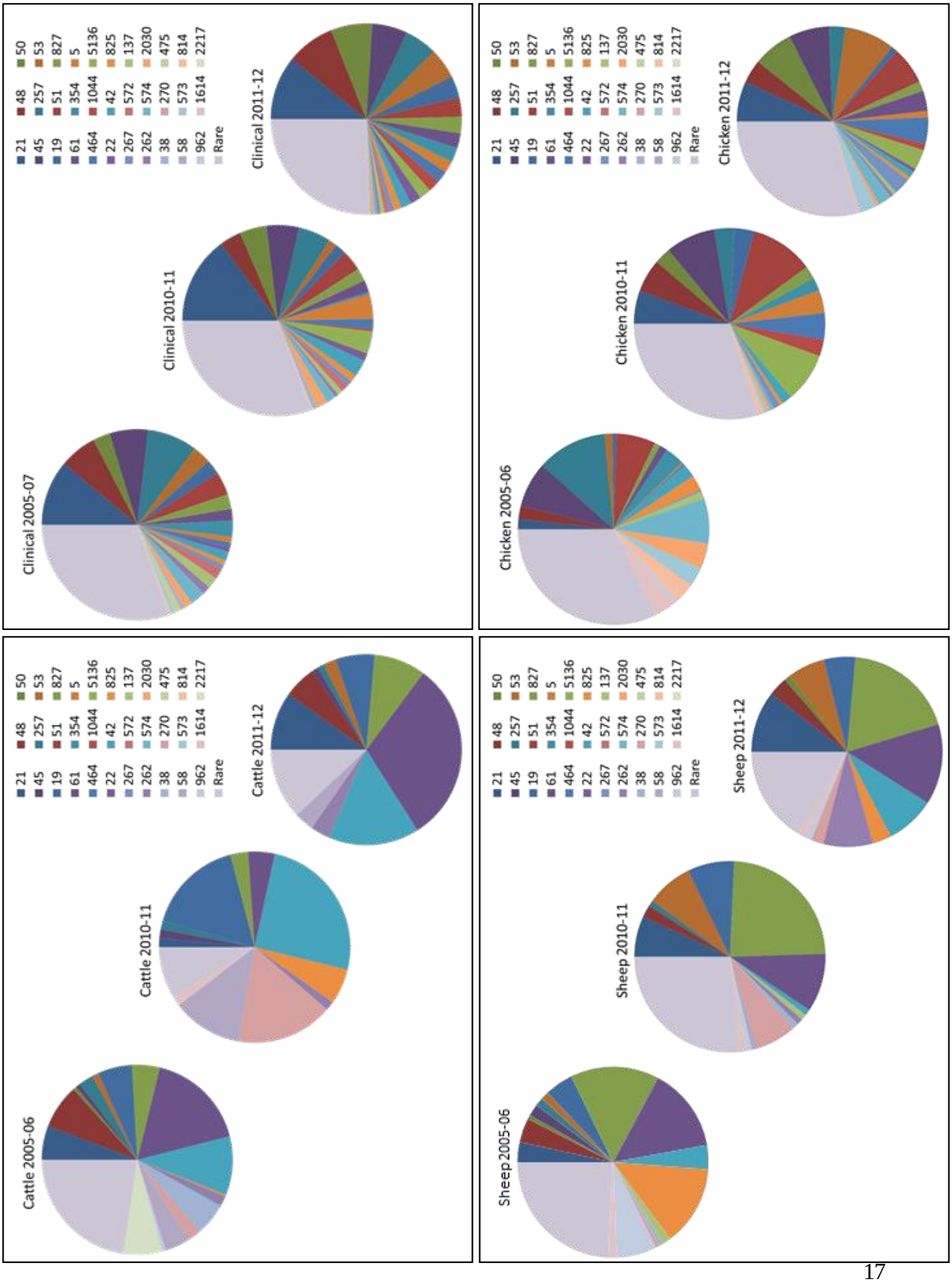


Figure 7. Observed abundance of Sequence Types by Reservoir, 2005-12.

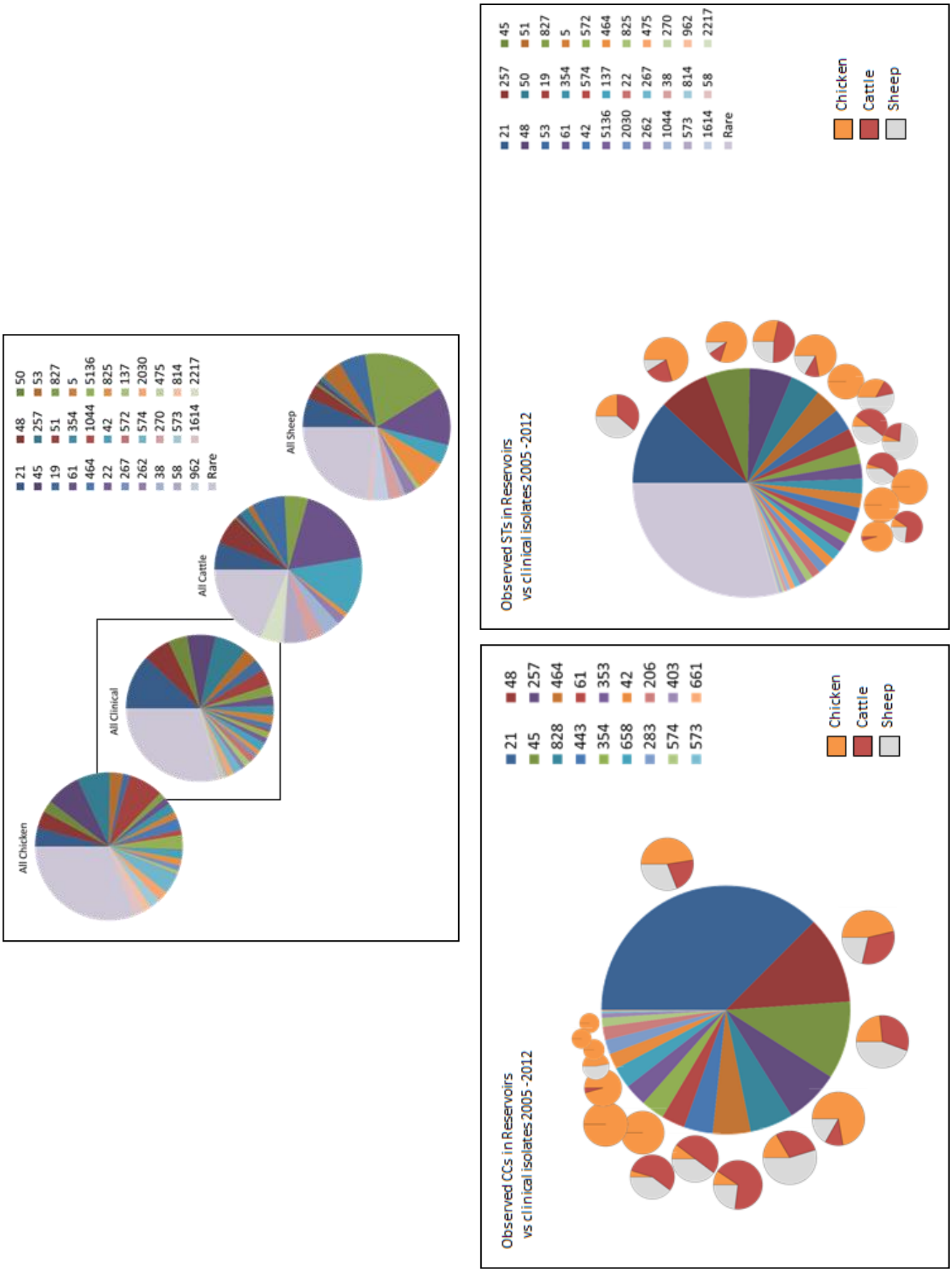


Table 6. Genetic distances (Nei) between isolates in the three study periods for all sources.

(a) Within hosts

ST level

Group	Genetic Distance (p-value)
Cattle 2010 -11 vs. 2011 -12	0.6074 (<0.0001)
Cattle 2005 -07 vs. 2011 -12	0.3932 (0.0328)
Cattle 2005 -07 vs. 2010 -11	0.5655 (0.0001)
Cattle 2005 -07 vs. 2010 -12	0.3492 (0.1073)
Sheep 2010 -11 vs. 2011 -12	0.3753 (0.0187)
Sheep 2005 -07 vs. 2011 -12	0.3460 (0.1272)
Sheep 2005 -07 vs. 2010 -11	0.4519 (0.0007)
Sheep 2005 -07 vs. 2010 -12	0.3994 (0.0223)
Chicken 2010 -11 vs. 2011 -12	0.4741 (<0.0001)
Chicken 2005 -07 vs. 2011 -12	0.6278 (<0.0001)
Chicken 2005 -07 vs. 2010 -11	0.6221 (<0.0001)
Chicken 2005 -07 vs. 2010 -12	0.5896 (<0.0001)
Clinical 2010 -11 vs. 2011 -12	0.3573 (<0.0001)
Clinical 2005 -07 vs. 2011 -12	0.3613 (<0.0001)
Clinical 2005 -07 vs. 2010 -11	0.4096 (<0.0001)
Clinical 2005 -07 vs. 2010 -12	0.3351 (<0.0001)

Allele level

Group	Genetic Distance (p-value)
Cattle 2010 -11 vs. 2011 -12	0.3817 (0.0002)
Cattle 2005 -07 vs. 2011 -12	0.1298 (0.7765)
Cattle 2005 -07 vs. 2010 -11	0.3735 (0.0004)
Cattle 2005 -07 vs. 2010 -12	0.1862 (0.1162)
Sheep 2010 -11 vs. 2011 -12	0.2124 (0.0538)
Sheep 2005 -07 vs. 2011 -12	0.2162 (0.0421)
Sheep 2005 -07 vs. 2010 -11	0.2378 (0.0247)
Sheep 2005 -07 vs. 2010 -12	0.2184 (0.0510)
Chicken 2010 -11 vs. 2011 -12	0.2242 (0.0027)
Chicken 2005 -07 vs. 2011 -12	0.3775 (<0.0001)
Chicken 2005 -07 vs. 2010 -11	0.3089 (<0.0001)
Chicken 2005 -07 vs. 2010 -12	0.3242 (<0.0001)
Clinical 2010 -11 vs. 2011 -12	0.1269 (0.0044)
Clinical 2005 -07 vs. 2011 -12	0.1221 (0.0148)
Clinical 2005 -07 vs. 2010 -11	0.1288 (0.0028)
Clinical 2005 -07 vs. 2010 -12	0.1016 (<0.0001)

(b) Between hosts

2011-12

ST level

Group	Genetic Distance (p-value)
Cattle 2011 -12 vs. Sheep 2011 -12	0.4000 (0.0025)
Cattle 2011 -12 vs. Chicken 2011 -12	0.7698 (<0.0001)
Sheep 2011 -12 vs. Chicken 2011 -12	0.7333 (<0.0001)
Cattle 2011 -12 vs. Clinical 2011 -12	0.6910 (<0.0001)
Sheep 2011 -12 vs. Clinical 2011 -12	0.6689 (<0.0001)
Chicken 2011 -12 vs. Clinical 2011 -12	0.4671 (<0.0001)

Allele level

Group/Group	Genetic Distance (p-value)
Cattle 2011 -12 vs. Sheep 2011 -12	0.3100 (0.0005)
Cattle 2011 -12 vs. Chicken 2011 -12	0.4265 (<0.0001)
Sheep 2011 -12 vs. Chicken 2011 -12	0.4805 (<0.0001)
Cattle 2011 -12 vs. Clinical 2011 -12	0.4176 (<0.0001)
Sheep 2011 -12 vs. Clinical 2011 -12	0.4179 (<0.0001)
Chicken 2011 -12 vs. Clinical 2011 -12	0.2376 (0.0008)

2010-11

ST level

Group	Genetic Distance (p-value)
Cattle 2010 -11 vs. Sheep 2010 -11	0.6875 (<0.0001)
Cattle 2010 -11 vs. Chicken 2010 -11	0.8605 (<0.0001)
Sheep 2010 -11 vs. Chicken 2010 -11	0.8232 (<0.0001)
Cattle 2010 -11 vs. Clinical 2010 -11	0.8428 (<0.0001)
Sheep 2010 -11 vs. Clinical 2010 -11	0.7651 (<0.0001)
Chicken 2010 -11 vs. Clinical 2010 -11	0.5355 (<0.0001)

Allele level

Group/Group	Genetic Distance (p-value)
Cattle 2010 -11 vs. Sheep 2010 -11	0.4983 (<0.0001)
Cattle 2010 -11 vs. Chicken 2010 -11	0.5939 (<0.0001)
Sheep 2010 -11 vs. Chicken 2010 -11	0.5955 (<0.0001)
Cattle 2010 -11 vs. Clinical 2010 -11	0.5383 (<0.0001)
Sheep 2010 -11 vs. Clinical 2010 -11	0.4713 (<0.0001)
Chicken 2010 -11 vs. Clinical 2010 -11	0.2752 (<0.0001)

2005-07

ST level

Group	Genetic Distance (p-value)
Cattle 2005 -07 vs. Sheep 2005 -07	0.5273 (0.0002)
Cattle 2005 -07 vs. Chicken 2005 -07	0.8722 (<0.0001)
Sheep 2005 -07 vs. Chicken 2005 -07	0.8351 (<0.0001)
Cattle 2005 -07 vs. Clinical 2005 -07	0.7663 (<0.0001)
Sheep 2005 -07 vs. Clinical 2005 -07	0.7481 (<0.0001)
Chicken 2005 -07 vs. Clinical 2005 -07	0.5785 (<0.0001)

Allele level

Group	Genetic Distance (p-value)
Cattle 2005 -07 vs. Sheep 2005 -07	0.3227 (0.0001)
Cattle 2005 -07 vs. Chicken 2005 -07	0.6226 (<0.0001)
Sheep 2005 -07 vs. Chicken 2005 -07	0.6460 (<0.0001)
Cattle 2005 -07 vs. Clinical 2005 -07	0.4324 (<0.0001)
Sheep 2005 -07 vs. Clinical 2005 -07	0.4918 (<0.0001)
Chicken 2005 -07 vs. Clinical 2005 -07	0.3984 (<0.0001)

2005-12

ST level

Group	Genetic Distance (p-value)
Cattle all vs. Sheep all	0.4599 (<0.0001)
Cattle all vs. Chicken all	0.8058 (<0.0001)
Sheep all vs. Chicken all	0.7709 (<0.0001)
Cattle all vs. Clinical all	0.7358 (<0.0001)
Sheep all vs. Clinical all	0.6681 (<0.0001)
Chicken all vs. Clinical all	0.4474 (<0.0001)

Allele level

Group	Genetic Distance (p-value)
Cattle all vs. Sheep all	0.3174 (<0.0001)
Cattle all vs. Chicken all	0.5123 (<0.0001)
Sheep all vs. Chicken all	0.5665 (<0.0001)
Cattle all vs. Clinical all	0.4179 (<0.0001)
Sheep all vs. Clinical all	0.4340 (<0.0001)
Chicken all vs. Clinical all	0.2772 (<0.0001)

Nei's genetic distance take a value of 0.0 where the genetic distance between the two populations is completely overlapping, and 1.0 when the two populations are completely genetically distinct.

3.3 Self-Attribution tests of models

Self-attribution is a key performance measure for these models. This is the average percentage accuracy that any given isolate from a reservoir can be correctly attributed back to its own reservoir. This can be performed in a number of ways. However, one simple approach is to use a jacknife method to predict the source of an isolate that was unknown to the model and known to the user. This is then repeated for all the source isolates a number of times (e.g. up to 10,000) so that an average, and confidence intervals, can be calculated. Self-attribution ranges are reported as between 62-97% for between 5-7 hosts for the Asymmetric Island model(17,23) and 38-70% for STRUCTURE(17). Note that by chance you would expect a correct self-attribution of 20% and 14% for 5 and 7 sources respectively. The poorest self-attribution in these methods is environment, which is likely to contain isolates from a number of hosts. These data demonstrate that there are differences in the frequencies of MLST types between hosts and that this information can be used for source attribution.

Overall the correct attribution percentages varied between approximately 30% to 99% (Figure 8). The average correct attribution percentages for each model (Table 7) shows that Structure-alleles has the highest (~66%) average correct attribution score. Dutch alleles and AI had the poorest score of predicting sheep (30.6%). Dutch alleles had the highest score of predicting pigs (99.8%). As in past reports, most analyses will continue to be reported for the Structure-alleles and the AI models; the former as it gives output typical for the other tests, the latter as its underlying assumptions and methodology are rather different – and so gives different outcomes. Future work will seek to clarify these differences, for instance by seeking validation of the molecular attributions by correlation with cases' reported exposures in the questionnaires.

Figure 8. Self-attribution (correct attribution) of animal isolates by Dutch, Structure and AI models.

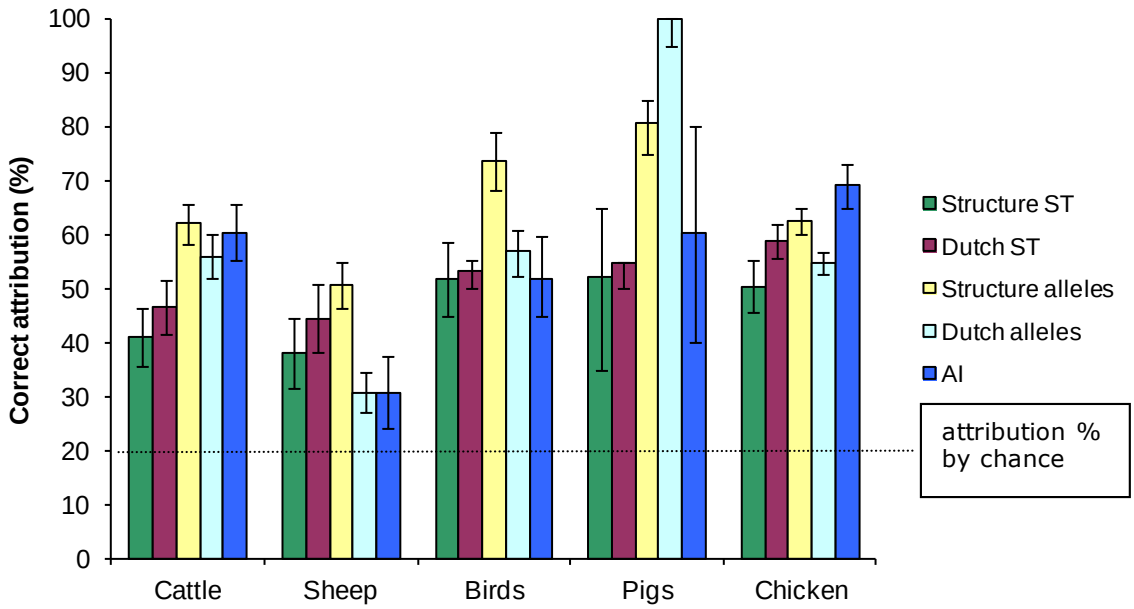


Table 7. Average correct self-attribution of animal strains by Dutch, Structure and AI models.

Model	Average correct self-attribution (%)	
	ST	Allele
Dutch proportional	51.5	59.5
STRUCTURE	46.5	65.7
Asymmetric Island	-	54.4

3.4 The sources of human campylobacteriosis in Grampian

The above analyses have provided evidence to support the following claims:

- Grampian is representative of Scotland for the understanding of the sources of campylobacteriosis
- Larger host datasets are more informative for attribution than smaller datasets
- STRUCTURE with alleles and the Asymmetric Island models were the most appropriate for attribution analyses.

Overview

In Grampian 2011-12 (Figure 9,a), neither pigs nor wild birds contributed significantly to the burden of campylobacteriosis together contributing less than $^1/_{5^{th}}$ of cases. Cattle and sheep attributed cases comprised just under ½ of all cases (47% Structure alleles model) and chicken attributed cases comprised the balance at slightly less (44% Structure alleles model). The Asymmetric Island model attributed cases more predominantly to chicken (3% pigs /wild birds; 16% cattle /sheep; 81% chicken). These proportions are not dissimilar to those for all study periods combined (Figure 9,b).

The excess (37%; 222 cases out of 600 cases) of clinical isolates attributed to chicken by the AI model compared to the Structure-Alleles model is largely due to a bias in attribution towards chicken for those strains belonging to three clonal complexes: CC21 (111 isolates), CC48 (49 isolates) and CC828 (26 isolates). These strains are predominantly isolated from ruminants (CC21 and CC48) and pigs and sheep (CC828) (Figure 7). The reason for this bias is unknown, but might be due to undetected reservoir sampling biases or to an aspect of the AI model. The future use of whole genome sequences will enable broader genotypic characterisation of strain types in the modelling processes, while the expanding animal and food reservoir isolate collections will enable better estimates to be made of reservoir compositions.

Age stratified analysis

Stratifying attribution source by case age (Figure 10; Supplementary Figure 2) indicated that for pigs and wild birds the burden is constant with age. In the case of ruminant and retail chicken sources there is an age dependent increase in attribution to retail chicken sources at the expense of ruminant sources. Since there is a trend of increasing number of cases in the elderly population in recent times(7) this could be explained by poultry sources. From the study periods 2005 -12 it is difficult to confirm whether this is actually the case, however the continuance of this study may clarify this hypothesis.

Date stratified analysis

Little change in the relative importance of the sources was seen over the study periods (Figures 11, 12).

Urban-rural stratified analysis

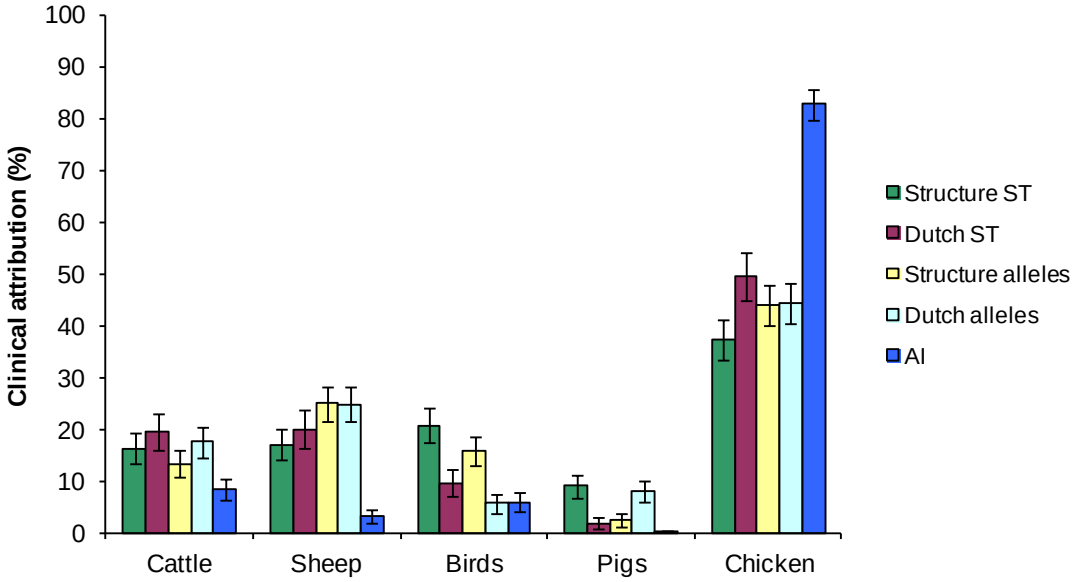
Figure 13 illustrates the partitioning of cases by three criteria: "chicken" vs "non-chicken" attributed; urban vs rural residence; age. For all ages, in both "chicken attributed" and "non-chicken attributed" cases the incidence is higher in the rural population. In urban population there is no difference in incidence between

"chicken attributed" and "non-chicken attributed" cases at the level of each individual age group and overall; young children and school pupils have lower incidences than the adults. In the rural-"chicken" group only the school pupils have a lower incidence than the other age groups. In the rural-"non-chicken" group young children have the highest incidence (10.8 cases/100,000/month) and school pupils have the lowest incidence.

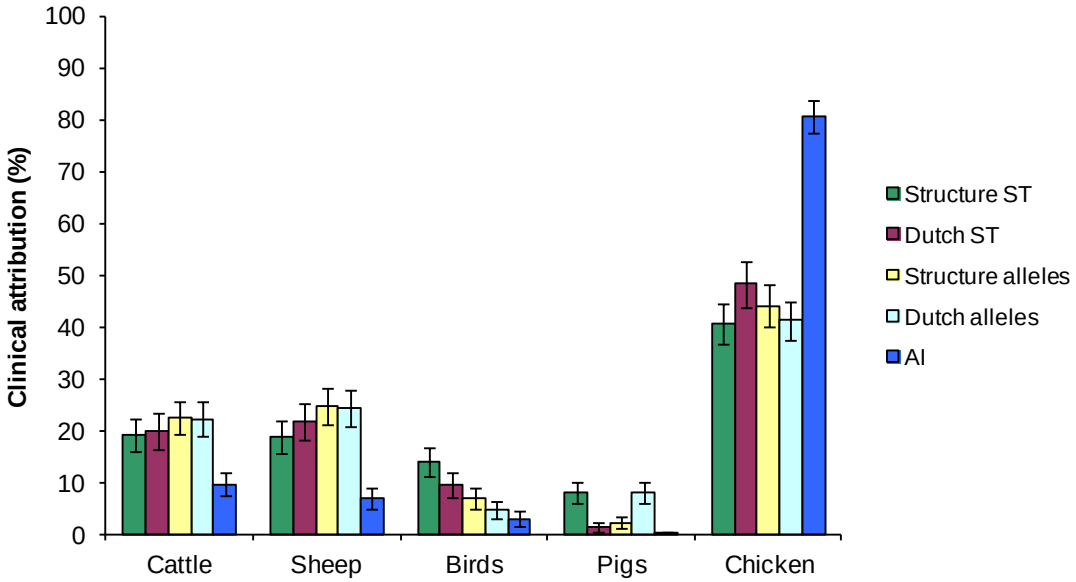
This dramatically higher incidence in rural, young children with non-chicken attributed cases strongly implies that there is an important reservoir for infection, that is associated with the countryside and with non-chicken sources. Further analysis of this cohort is likely to shed light on our understanding of the relative importance, not just of the size of the different reservoirs, which is quite well established, but on the differential exposures to these sources.

Figure 9. Source attribution of Grampian clinical isolates using (2011 -12) (a) the 2011 -12 host dataset or (b) the combined 2005-12 host dataset.

(a)



(b)



Source attribution of Grampian clinical isolates using Host Dataset 1+2+3. 95% CI.

Figure 10. Attributed host sources of clinical isolates from Grampian (2005-07 and 2010 -12) and Scotland (2005-06) partitioned by patient age.

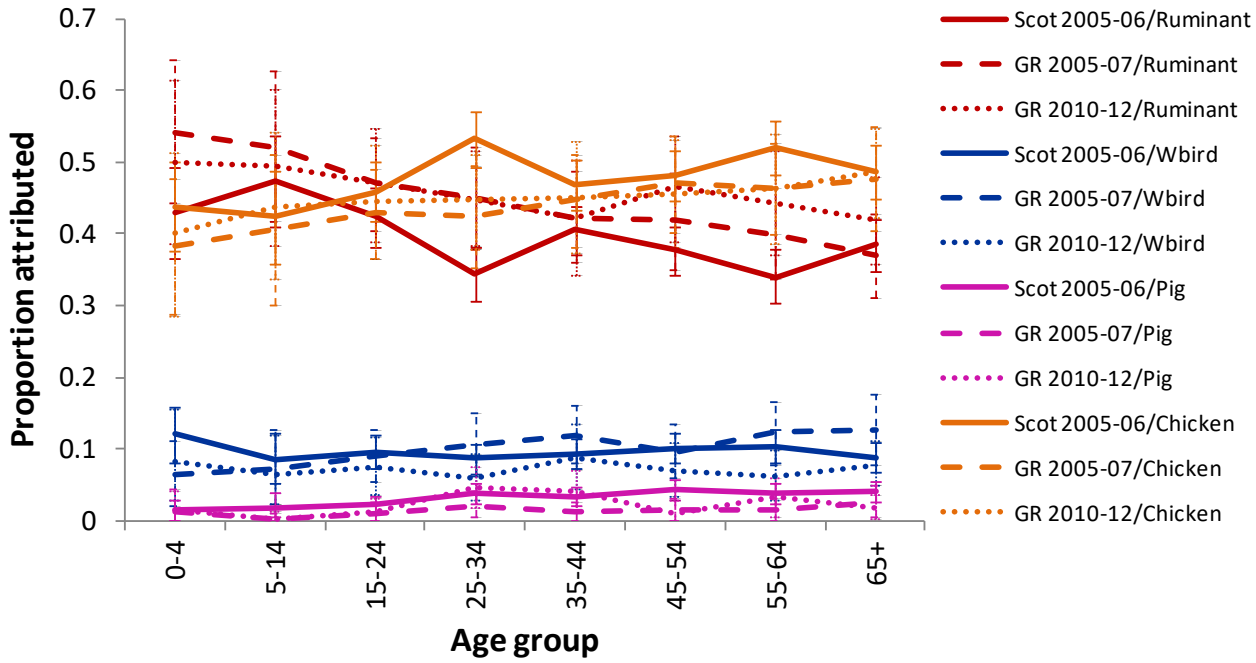
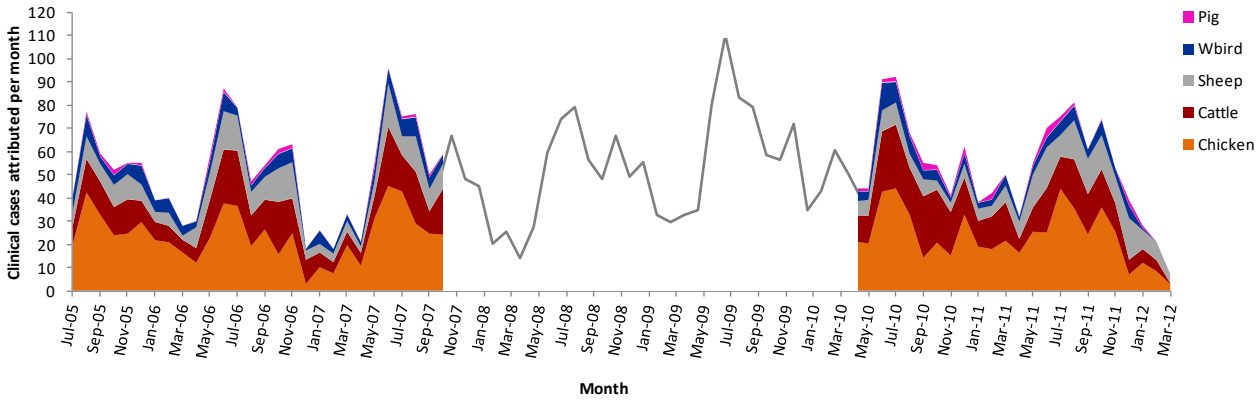
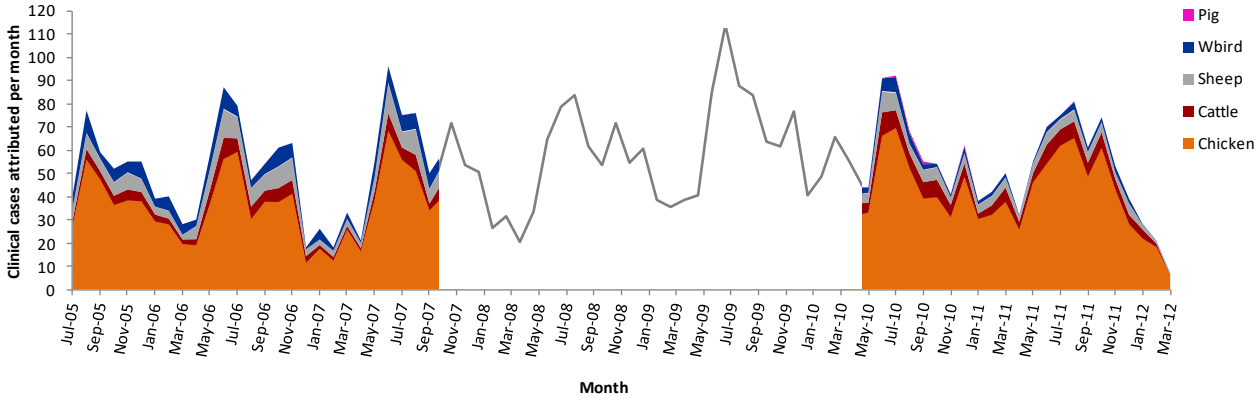


Figure 11. Attribution to five potential host reservoirs of clinical *Campylobacter* cases in Grampian per month by (a) STRUCTURE with alleles Model, (b) Asymmetric Island Model.

(a)



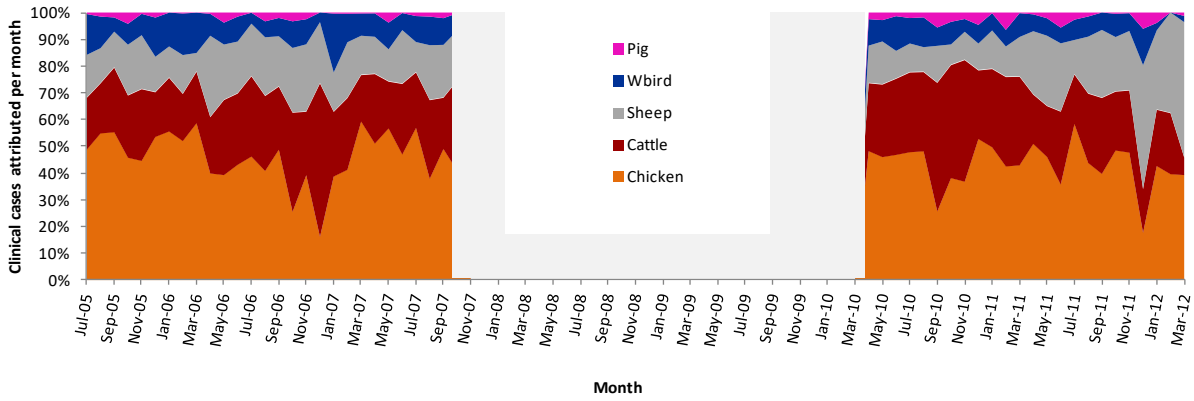
(b)



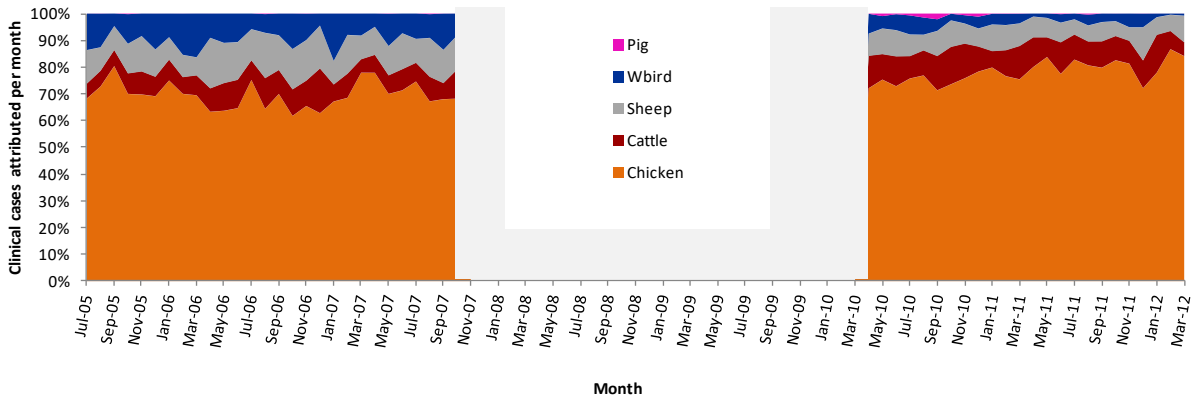
Graph stacked to total number of clinical cases. Attribution based on host datasets available up to each period.

Figure 12. Attribution to five potential host reservoirs of clinical *Campylobacter* cases in Grampian per month by (a) STRUCTURE with alleles Model, (b) Asymmetric Island Model.

(a)

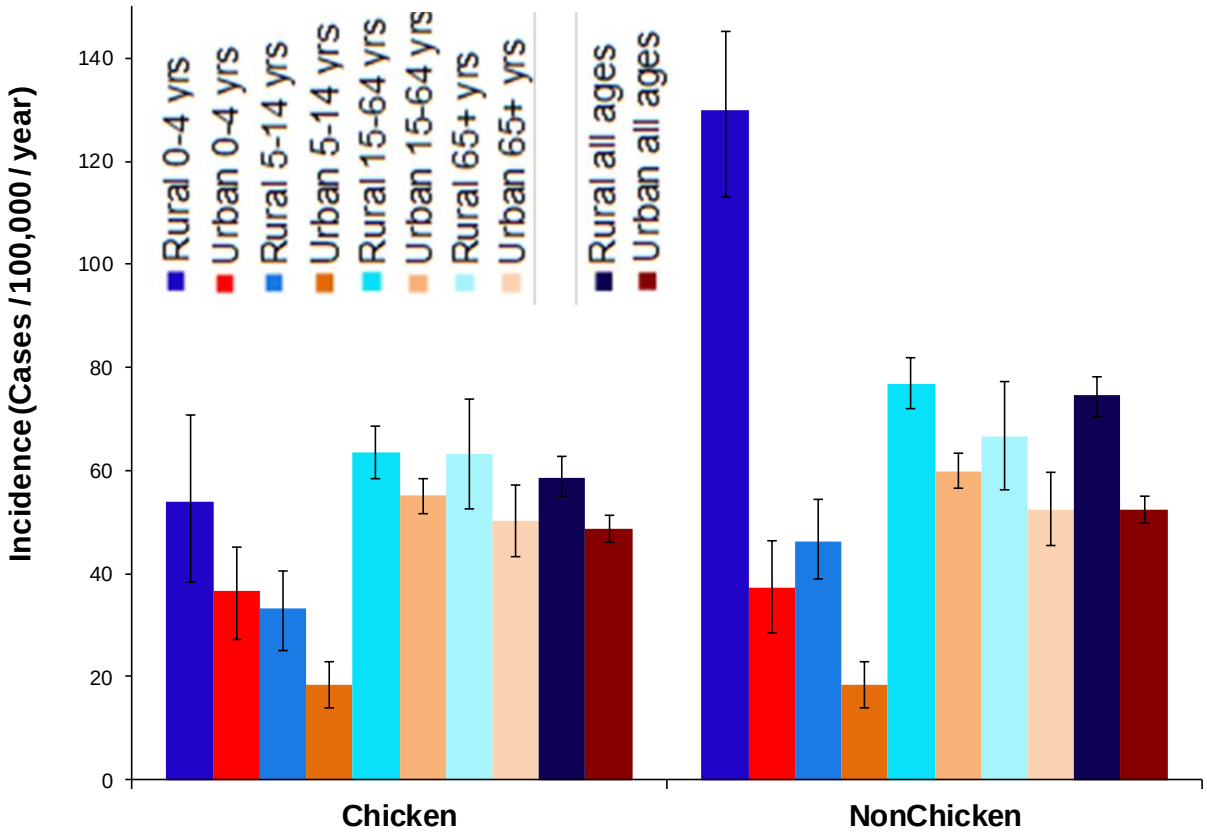


(b)



Graph stacked to 100% of clinical cases. Attribution based on host datasets available up to each period.

Figure 13. Chicken- NonChicken and Rural-Urban variation.



All isolates from 2005 -12 analysed by the STRUCTURE alleles model.

3.5 Conclusions

The prevalence of *Campylobacter* in the different reservoirs sampled and in retail chicken shows a continuing increase in prevalence for all three. This increase might be due to subtle changes in laboratory protocols or to differing staff, or to real increases over eight year period. If true, it might provide an explanation for the increase in clinical cases over the last few years.

There continues to be extensive population diversity of *Campylobacter* strains in farm animals, in retail chicken and in human isolates. The relative abundance of the strain types found in these reservoirs continues to be quite dynamic with even the relative abundance of commoner strains changing significantly between the 2005 -07 study and the 2011 -12 study, indeed even over periods as short as one year. Notwithstanding this, the strain profiles in each species remain characteristic, and thus the basis of molecular attribution modelling continues to hold.

The attribution models and associated datasets were validated by self-attribution testing. The average correct self-attribution percentages for each model showed that Structure-alleles had the highest (~66%) average correct attribution score, and this was therefore used in most analyses. Asymmetric Island gave lower (~54%) average correct self-attribution scores, and has been used for comparative purposes as it attributes a higher proportion to chicken.

Host attribution modelling of putative sources of human infection suggests that there continues to be broadly the same proportional attribution over all the study periods with retail chicken making the largest contribution.

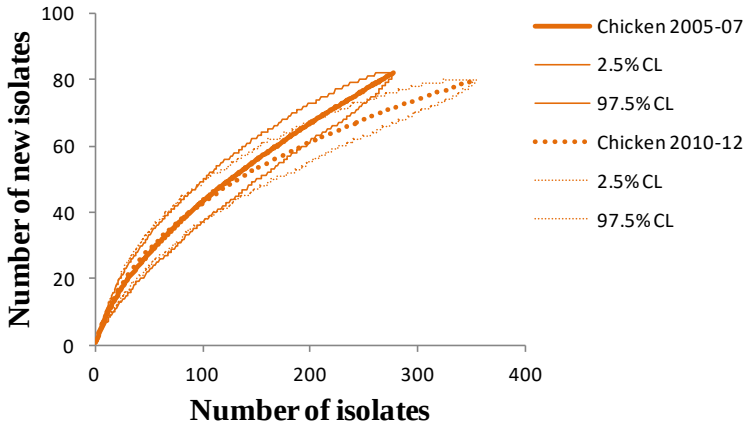
The impact, in Scotland and the UK, of forthcoming intervention strategies to reduce human campylobacteriosis originating from the poultry food chain should be observable by a decrease in human cases and confirmed by a subsequent decrease in the proportion of clinical isolates associated with chicken.

The current study has highlighted the dynamic nature of *Campylobacter* and the requirement to monitor prevalence, counts and strain types.

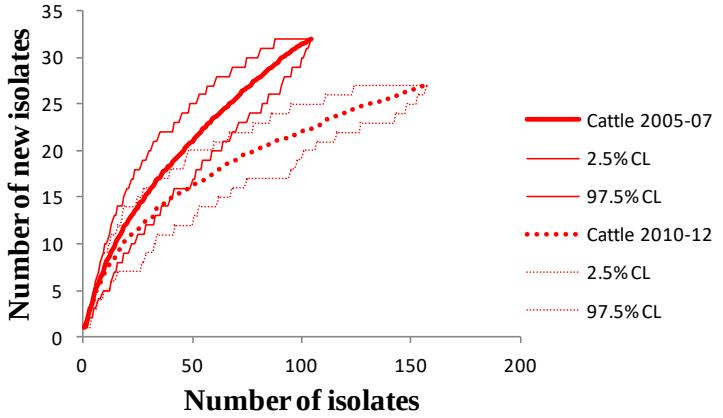
4. Supplementary Data

Supplementary Figure 1. Rarefaction (saturation) analysis.

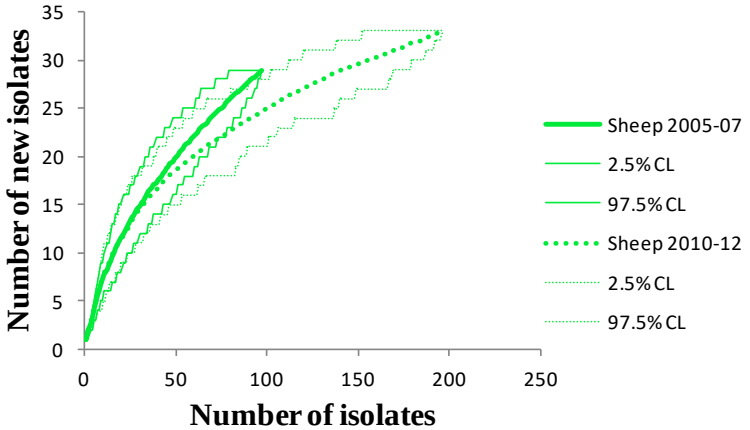
(a)



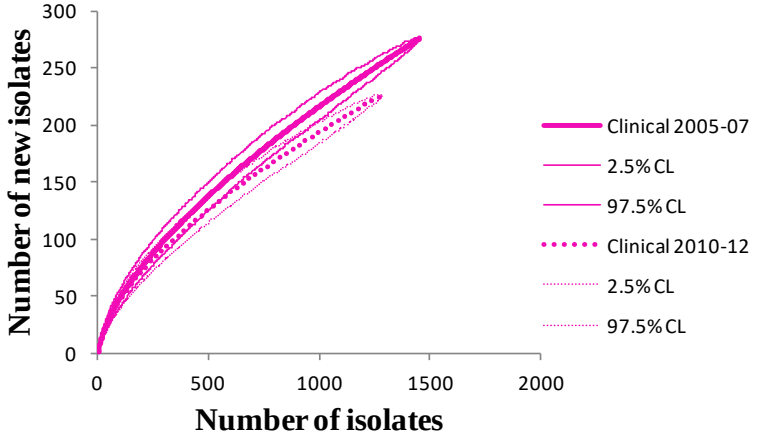
(b)



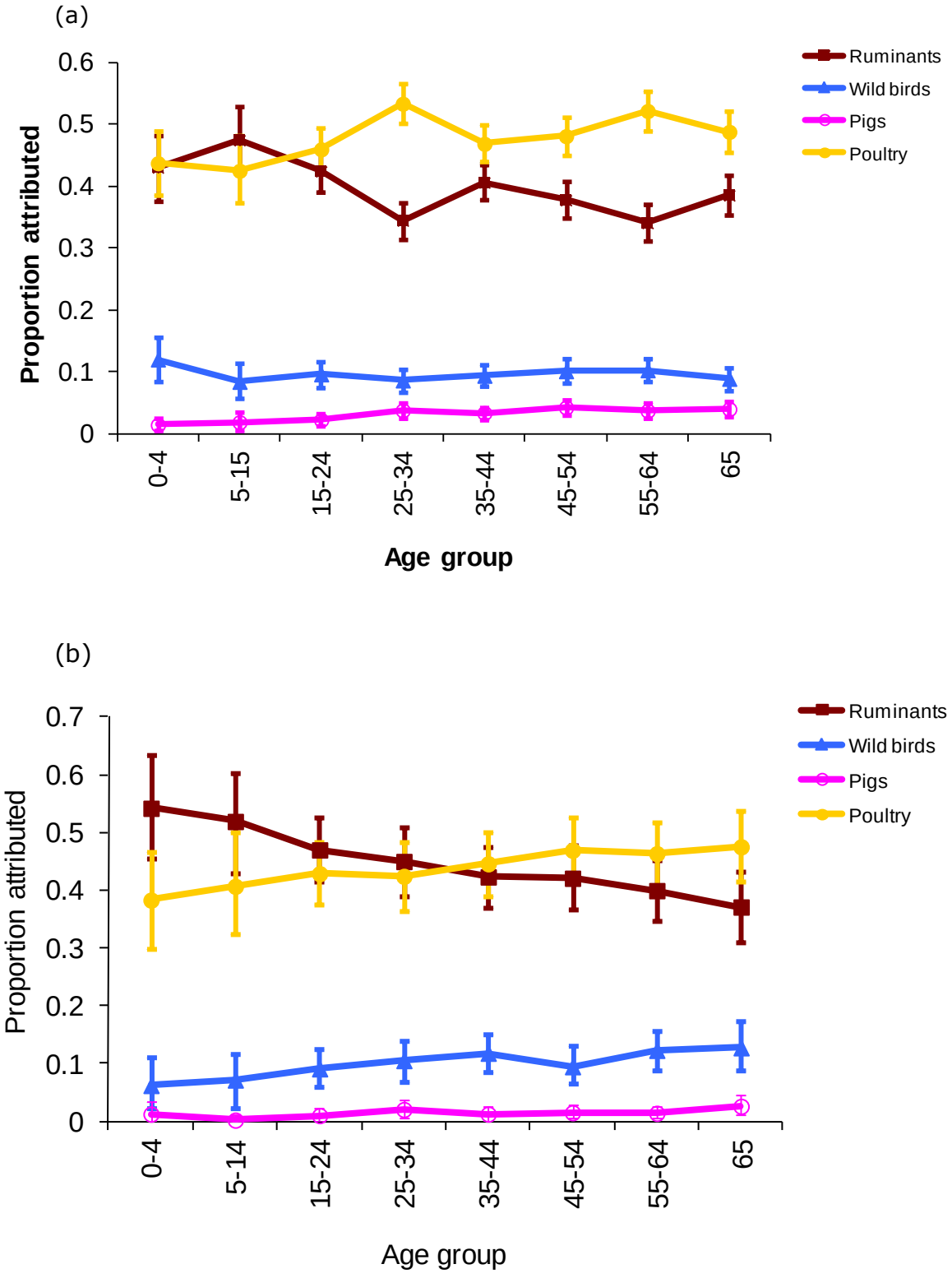
(c)



(d)

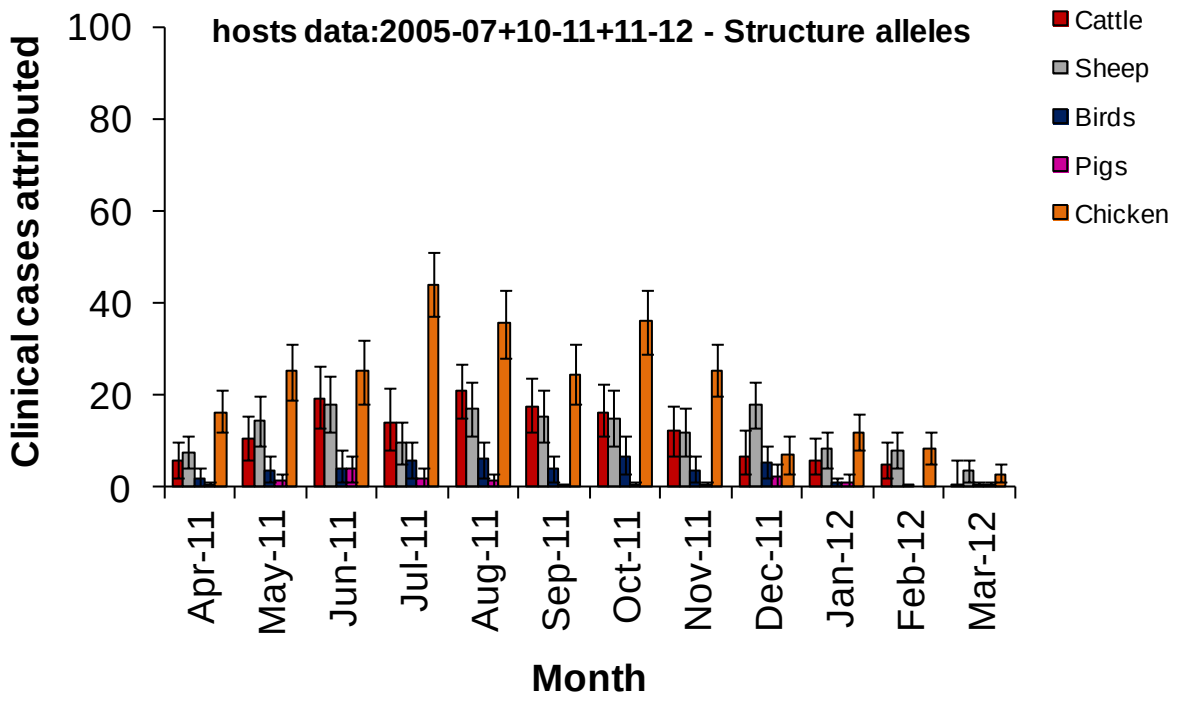


Supplementary Figure 2. Patient age vs attributed host source of (a) Scottish clinical isolates (2005 -06) or (b) Grampian clinical isolates (2005 -07).



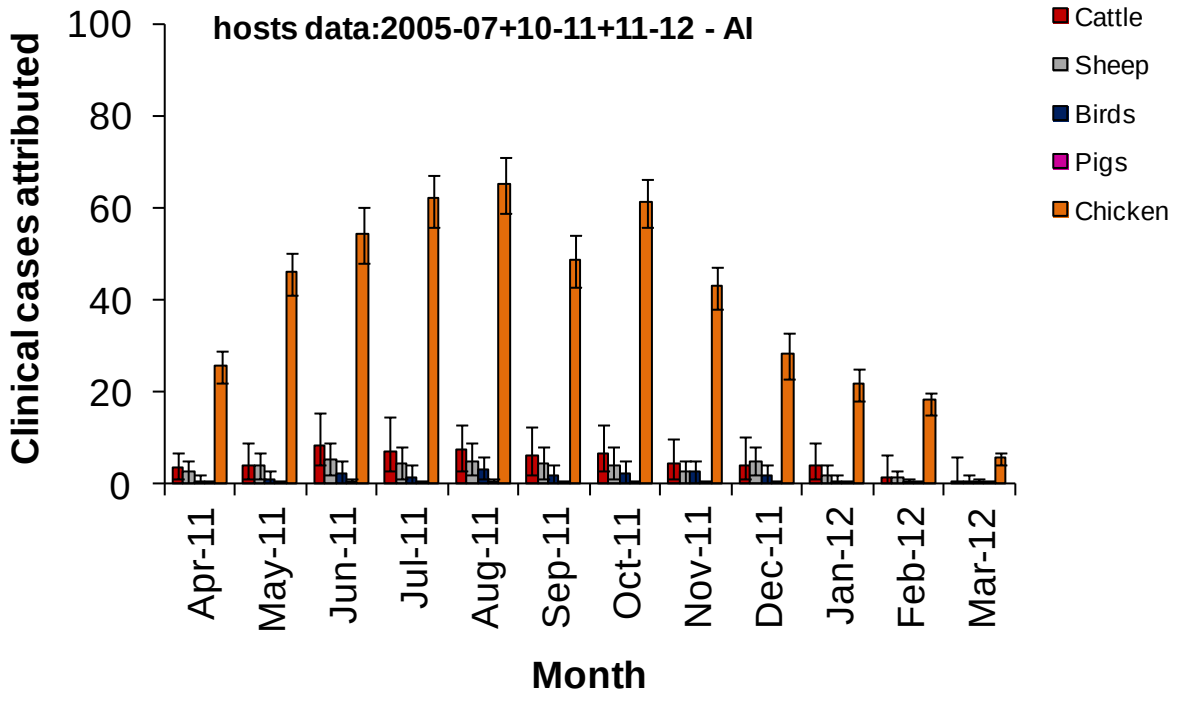
Attribution based on Host Dataset 1. 90% CI.

Supplementary Figure 3. Molecular attribution of clinical *Campylobacter* by STRUCTURE with alleles.



Attribution based on Host Dataset 3. 90% CI.

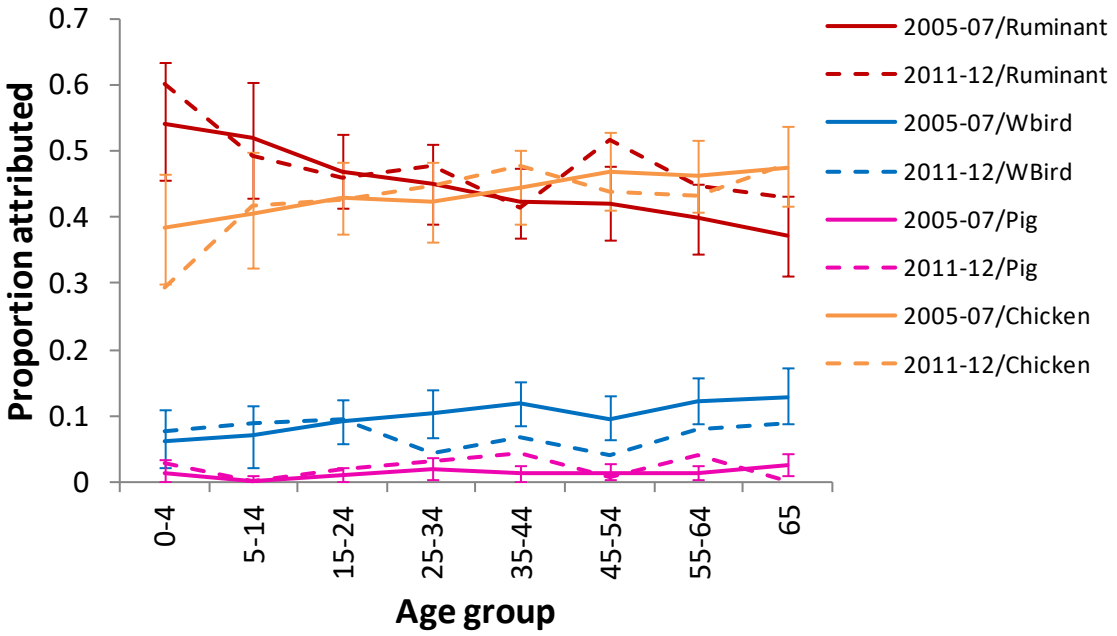
Supplementary Figure 4. Molecular attribution of clinical *Campylobacter* by Asymmetric Island.



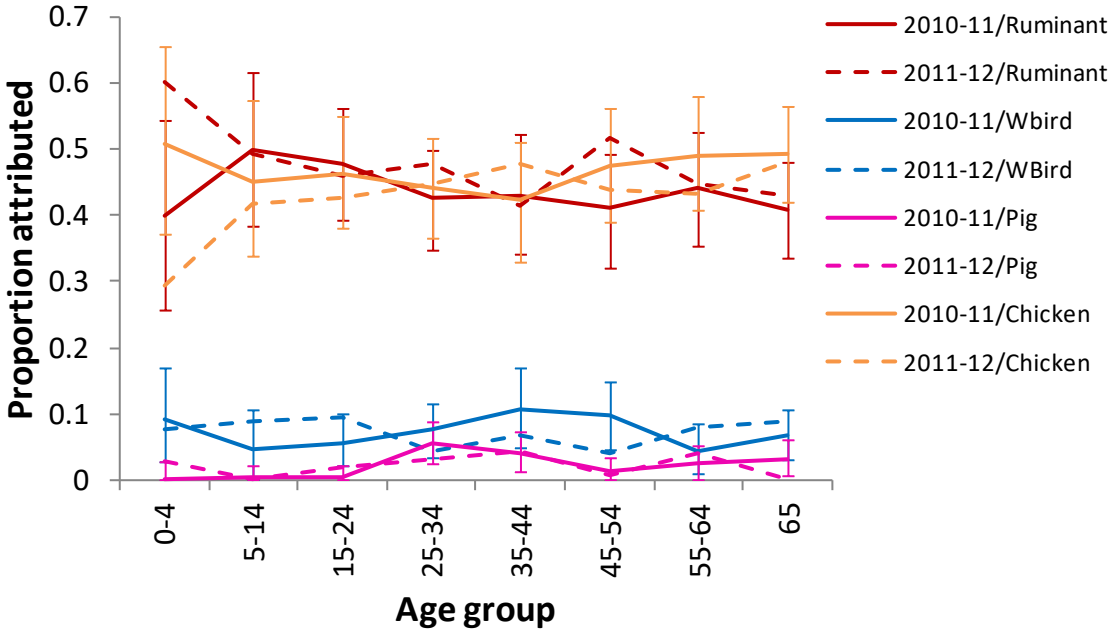
Attribution based on Host Dataset 3. 90% CI.

Supplementary Figure 5. Attributed host sources of clinical isolates from Grampian in: (a) 2005-07 and from 2011 -12 partitioned by patient age; (b) 2010-11 and from 2011 -12

(a)



(b)



Attribution based on Host Dataset 3. 90% CI.

Supplementary Table 1. Comparing the relative abundance of the most prevalent clinical MLST types over time.

Period	ST	OR[a]	P-value[b]
Jul2005-Oct2007	5	1.00	-
Apr 2010- Mar2011	5	**4.49**	**0.0000**
Apr 2011- Mar2012	5	2.10	0.0797
Jul2005-Oct2007	19	1.00	-
Apr 2010- Mar2011	19	0.75	0.4418
Apr 2011- Mar2012	19	1.50	0.1447
Jul2005-Oct2007	21	1.00	-
Apr 2010- Mar2011	21	**1.40**	**0.0139**
Apr 2011- Mar2012	21	0.97	0.9382
Jul2005-Oct2007	22	1.00	-
Apr 2010- Mar2011	22	1.58	0.3485
Apr 2011- Mar2012	22	**2.24**	**0.0633**
Jul2005-Oct2007	38	1.00	-
Apr 2010- Mar2011	38	1.75	0.3478
Apr 2011- Mar2012	38	0.40	0.6810
Jul2005-Oct2007	42	1.00	-
Apr 2010- Mar2011	42	**1.93**	**0.0440**
Apr 2011- Mar2012	42	1.21	0.5689
Jul2005-Oct2007	45	1.00	-
Apr 2010- Mar2011	45	0.83	0.3891
Apr 2011- Mar2012	45	0.91	0.6921
Jul2005-Oct2007	48	1.00	-
Apr 2010- Mar2011	48	**0.56**	**0.0106**
Apr 2011- Mar2012	48	1.33	0.1242
Jul2005-Oct2007	50	1.00	-
Apr 2010- Mar2011	50	**1.63**	**0.0434**
Apr 2011- Mar2012	50	**2.53**	**0.0000**
Jul2005-Oct2007	51	1.00	-
Apr 2010- Mar2011	51	0.88	0.7164
Apr 2011- Mar2012	51	0.80	0.4434
Jul2005-Oct2007	53	1.00	-
Apr 2010- Mar2011	53	**0.46**	**0.0274**
Apr 2011- Mar2012	53	**1.70**	**0.0287**
Jul2005-Oct2007	58	-	-
Apr 2010- Mar2011	58	-	-
Apr 2011- Mar2012	58	-	-
Jul2005-Oct2007	61	1.00	-
Apr 2010- Mar2011	61	1.01	1.0000
Apr 2011- Mar2012	61	1.26	0.5035

Jul2005-Oct2007	137	1.00	-
Apr 2010- Mar2011	137	0.44	0.0941
Apr 2011- Mar2012	137	**0.34**	**0.0478**
Jul2005-Oct2007	257	1.00	-
Apr 2010- Mar2011	257	**0.63**	**0.0141**
Apr 2011- Mar2012	257	**0.63**	**0.0219**
Jul2005-Oct2007	262	1.00	-
Apr 2010- Mar2011	262	0.42	0.1248
Apr 2011- Mar2012	262	0.36	0.1066
Jul2005-Oct2007	267	1.00	-
Apr 2010- Mar2011	267	1.05	1.0000
Apr 2011- Mar2012	267	1.01	1.0000
Jul2005-Oct2007	270	1.00	-
Apr 2010- Mar2011	270	0.42	0.6708
Apr 2011- Mar2012	270	0.48	0.6780
Jul2005-Oct2007	354	1.00	-
Apr 2010- Mar2011	354	**0.16**	**0.0002**
Apr 2011- Mar2012	354	0.93	0.8804
Jul2005-Oct2007	464	1.00	-
Apr 2010- Mar2011	464	**2.31**	**0.0456**
Apr 2011- Mar2012	464	**2.67**	**0.0205**
Jul2005-Oct2007	475	1.00	-
Apr 2010- Mar2011	475	**0.12**	**0.0113**
Apr 2011- Mar2012	475	**0.13**	**0.0204**
Jul2005-Oct2007	572	1.00	-
Apr 2010- Mar2011	572	0.72	0.4694
Apr 2011- Mar2012	572	0.46	0.1153
Jul2005-Oct2007	573	1.00	-
Apr 2010- Mar2011	573	0.00	0.1042
Apr 2011- Mar2012	573	0.00	0.1140
Jul2005-Oct2007	574	1.00	-
Apr 2010- Mar2011	574	0.62	0.2109
Apr 2011- Mar2012	574	**0.19**	**0.0013**
Jul2005-Oct2007	814	1.00	-
Apr 2010- Mar2011	814	0.52	1.0000
Apr 2011- Mar2012	814	0.00	0.3283
Jul2005-Oct2007	825	1.00	-
Apr 2010- Mar2011	825	1.40	0.4756
Apr 2011- Mar2012	825	1.62	0.3238
Jul2005-Oct2007	827	1.00	-
Apr 2010- Mar2011	827	0.85	0.6546
Apr 2011- Mar2012	827	1.12	0.7619
Jul2005-Oct2007	962	1.00	-
Apr 2010- Mar2011	962	0.00	0.5556
Apr 2011- Mar2012	962	0.00	0.5604
Jul2005-Oct2007	1044	-	-
Apr 2010- Mar2011	1044	**∞**	**0.0335**
Apr 2011- Mar2012	1044	**∞**	**0.0000**

Jul2005-Oct2007	1614	1.00	-
Apr 2010- Mar2011	1614	1.05	1.0000
Apr 2011- Mar2012	1614	0.00	1.0000
Jul2005-Oct2007	2030	1.00	-
Apr 2010- Mar2011	2030	1.37	0.4528
Apr 2011- Mar2012	2030	**0.24**	**0.0349**
Jul2005-Oct2007	2217	-	-
Apr 2010- Mar2011	2217	-	-
Apr 2011- Mar2012	2217	-	-
Jul2005-Oct2007	5136	-	-
Apr 2010- Mar2011	5136	**∞**	**0.0000**
Apr 2011- Mar2012	5136	**∞**	**0.0000**
Jul2005-Oct2007	Other	1.00	-
Apr 2010- Mar2011	Other	1.02	0.8407
Apr 2011- Mar2012	Other	**0.79**	**0.0366**

ORs are given relative to 2005-07 period. Significance is highlighted in bold.
[a] Odds ratio (if >1.0 indicates an increase with time)
[b] Calculated by Fisher's exact test

Supplementary Table 2. Recent outputs from FSAS projects and enabled resources.

Publications

1. Bessell, P., O. Rotariu, G. T. Innocent, A. Smith-Palmer, N. J. C. Strachan, K. J. Forbes, J. M. Cowden, S. W. J. Reid, and L. Matthews. 2012. Using sequence data to identify alternative routes and risk of infection: A case-study of *Campylobacter* in Scotland. BMC Infect.Dis. 12:80.

2. Read, D. S., D. J. Woodcock, N. J. Strachan, K. J. Forbes, F. M. Colles, M. C. Maiden, F. Clifton-Hadley, A. Ridley, A. Vidal, J. Rodgers, A. S. Whiteley, and S. K. Sheppard. 2012. Evidence for phenotypic plasticity amongst multi-host Campylobacter jejuni and C. coli lineages using ribosomal MLST and Raman spectroscopy. Appl.Environ.Microbiol.

3. Strachan, N. J. C., M. Macrae, A. Thomson, O. Rotariu, I. D. Ogden, and K. J. Forbes. 2012. Source attribution, prevalence and enumeration of *Campylobacter* spp. from retail liver. Int.J.Food Microbiol. 153:234-236.

4. Sheppard, S. K., F. M. Colles, N. D. McCarthy, N. J. C. Strachan, I. D. Ogden, K. J. Forbes, J. F. Dallas, and M. C. J. Maiden. 2011. Niche segregation and genetic structure of Campylobacter jejuni populations from wild and agricultural host species. Mol Ecol 20:3484-3490.

5. Sproston, E. L., I. D. Ogden, M. Macrae, J. F. Dallas, S. K. Sheppard, A. J. Cody, M. Colles, M. J. Wilson, K. J. Forbes, and N. J. C. Strachan. 2011. Temporal Variation and Host Association in the Campylobacter Population in a Longitudinal Ruminant Farm Study. Appl.Environ.Microbiol. 77:6579-6586.

Presentations

1. Strachan, NJC "BBC Radio 4, Face the Facts, Campylobacter - the silent epidemic" January 2013.
2. Forbes KJ, Strachan NJC "Campylobacter: when will it ever stop? Human campylobacteriosis in Scotland" NHS Grampian Public Health, Aberdeen, Nov 2012.
3. Forbes KJ "Measuring the impact of interventions on *Campylobacter*." FSA - defra - BBSRC Workshop. London, January 2012.
4. Strachan NJC, Rotariu O, Smith-Palmer A., Cowden J, Sheppard SK, O'Brien SJ, Maiden M, Macrae M, Bessell PR, Matthews L, Reid S, Innocent G, Ogden ID and Forbes KJ. "Elucidating the seasonality of human *Campylobacter* infections". CampylobacterUK, London, Jan 2012.
5. Sproston EL, Ogden ID, MacRae M, Dallas JF, Sheppard SK, Cody AJ, Colles F, Wilson MJ, Forbes, KJ, Strachan NJC. "The temporal and host variation in the *Campylobacter* population of ruminants – a longitudinal farm study." CampylobacterUK, London, Jan 2012.
6. Forbes KJ, Colles F, Rotariu O, Thomson A, Strachan NJC "Can source attribution explain the dramatic rise in *Campylobacter* infections in the UK?" CampylobacterUK, London, Jan 2012.

7. Strachan NJC, Rotariu O, Smith-Palmer A., Cowden J, Sheppard SK, O'Brien SJ, Maiden M, Macrae M, Bessell PR, Matthews L, Reid S, Innocent G, Ogden ID and Forbes KJ. "Elucidating the seasonality of human *Campylobacter* infections". 16th Int Workshop on CHRO, Vancouver, Canada, Sept 2011.
8. Sproston EL, Ogden ID, MacRae M, Dallas JF, Sheppard SK, Cody AJ, Colles F, Wilson MJ, Forbes, KJ, Strachan NJC. "The temporal and host variation in the *Campylobacter* population of ruminants – a longitudinal farm study." 16th Int Workshop on CHRO, Vancouver, Canada, Sept 2011.
9. Forbes KJ, Colles F, Rotariu O, Thomson A, Strachan NJC "Can source attribution explain the dramatic rise in *Campylobacter* infections in the UK?" 16th Int Workshop on CHRO, Vancouver, Canada, Sept 2011.
10. Sheppard SK, Forbes KJ, Maiden M. "Hybrid Speciation in agricultural *Campylobacter coli*." 16th Int Workshop on CHRO, Vancouver, Canada, Sept 2011.
11. "Molecular epidemiological research on zoonoses at Aberdeen." Presented to Tim Smith, Chief Executive of the Food Standards Agency. Aberdeen, August 2011.
12. "*Campylobacter*. Using phylogeny to understand the biology of a human zoonotic gastro-intestinal bacterial pathogen." St Andrews, June 2011.

5. References

1. Buzby, J. C. and T. Roberts. 1997. Economic costs and trade impacts of microbial foodborne illness. World Health Statistics Quarterly - Rapport Trimestriel de Statistiques Sanitaires Mondiales 50:57-66.

2. EFSA. 2010. Scientific Opinion on Quantification of the risk posed by broiler meat to human campylobacteriosis in the EU. EFSA Journal 8:1-89.

3. Food Standards Agency, BBSRC, and DEFRA. 2010. UK research and innovation strategy for campylobacter in the food chain.

4. Forbes, K. J. 2009. The Molecular Epidemiology of Scottish *Campylobacter* Isolates from Human Cases of Infection using Multilocus Sequence Typing (MLST). CaMPS -*Campylobacter* MLST Project in ScotlandFood Standards Agency.

5. Forbes, K. J. and i-CaMPS team. 2011. i-CaMPS impact of interventions - *Campylobacter* MLST Project in Scotland. 2010- 1011.

6. French, N. and Molecular Epidemiology and Veterinary Public Health Group. 2008. Enhancing Surveillance of Potentially Foodborne Enteric Diseases in New Zealand: Human Campylobacteriosis in the Manawatu, p. 1-56. Massey University, New Zealand.

7. Gillespie, I. A., S. J. O'Brien, and F. E. Bolton. 2009. Age Patterns of Persons with Campylobacteriosis, England and Wales, 1990–2007. Emerg.Infect.Dis. 15:2046-2048.

8. Gormley, F. J., M. Macrae, K. J. Forbes, I. D. Ogden, J. F. Dallas, and N. J. C. Strachan. 2008. Has retail chicken played a role in the decline of human campylobacteriosis? Appl.Environ.Microbiol. 74:383-390.

9. Humphrey, T. J., S. O'Brien, and M. Madsen. 2007. *Campylobacter* as zoonotic pathogens: a food production perspective. Int.J.Food Microbiol. 117:237-257.

10. Jolley, K. A., C. M. Bliss, J. S. Bennett, H. B. Bratcher, C. Brehony, F. Colles, H. Wimalarathna, O. B. Harrison, S. K. Sheppard, A. J. Cody, and M. C. J. Maiden. 2012. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology 158:1005-1015.

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

12. Little, C. L., F. J. Gormley, N. Rawal, and J. F. Richardson. 2010. A recipe for disaster: outbreaks of campylobacteriosis associated with poultry liver pâté in England and Wales. Epidemiol.Infect.

13. NELSON, W. A. R. R. 2010. Campylobacteriosis in New Zealand. Epidemiol.Infect. 138:1762-1764.

14. Pires, S. M., E. G. Evers, W. van Pelt, T. Ayers, E. Scallan, F. J. Angulo, A. Havelaar, and T. Hald. 2009. Attributing the Human Disease Burden of Foodborne Infections to Specific Sources. Foodborne Path.Dis. 6:417-424.

15. Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

16. Rotariu, O., J. F. Dallas, I. D. Ogden, M. Macrae, S. K. Sheppard, M. C. J. Maiden, F. J. Gormley, K. J. Forbes, and N. J. C. Strachan. 2009. Spatiotemporal homogeneity of *Campylobacter* subtypes from cattle and sheep across NE and SW Scotland. Appl.Environ.Microbiol. 75:6275-6281.

17. Sheppard, S. K., J. F. Dallas, N. J. C. Strachan, M. Macrae, N. D. McCarthy, D. Falush, I. D. Ogden, M. C. J. Maiden, and K. J. Forbes. 2009. *Campylobacter* Genotyping to Determine the Source of Human Infection. Clin.Infect.Dis. 48:1072-1078.

18. Skirrow, M. B. 1977. Campylobacter enteritis: a "new" disease. BMJ 2:9-11.

19. Strachan, N. J. C. and K. J. Forbes. 2010. The growing UK epidemic of human campylobacteriosis. Lancet 376:665-667.

20. Tam, C. C., L. C. Rodrigues, L. Viviani, J. P. Dodds, M. R. Evans, P. R. Hunter, J. J. Gray, L. H. Letley, G. Rait, D. S. Tompkins, and S. J. O'Brien. 2011. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. Gut .

21. Tustin, J., K. Laberge, P. Michel, J. Reiersen, S. Dadadóttir, H. Briem, H. Hardardottir, K. Kristinsson, E. Gunnarsson, V. Fridriksdottir, and F. Georgsson. 2011. A National Epidemic of Campylobacteriosis in Iceland, Lessons Learned. Zoonoses and Public Health 58:440-447.

22. van den Brandhof, W. E., G. A. De Wit, M. A. S. de Wit, and Y. Van Duynhoven. 2004. Costs of gastroenteritis in The Netherlands. Epidemiol.Infect. 132:211-221.

23. Wilson, D. J., E. Gabriel, A. J. H. Leatherbarrow, J. Cheesbrough, S. Gee, E. Bolton, A. J. Fox, P. Fearnhead, C. A. Hart, and P. J. Diggle. 2008. Tracing the Source of Campylobacteriosis. PLoS Genet 4:e1000203.