Determining the source of human campylobacteriosis cases through time

Jonathan Marshall

27 November 2013

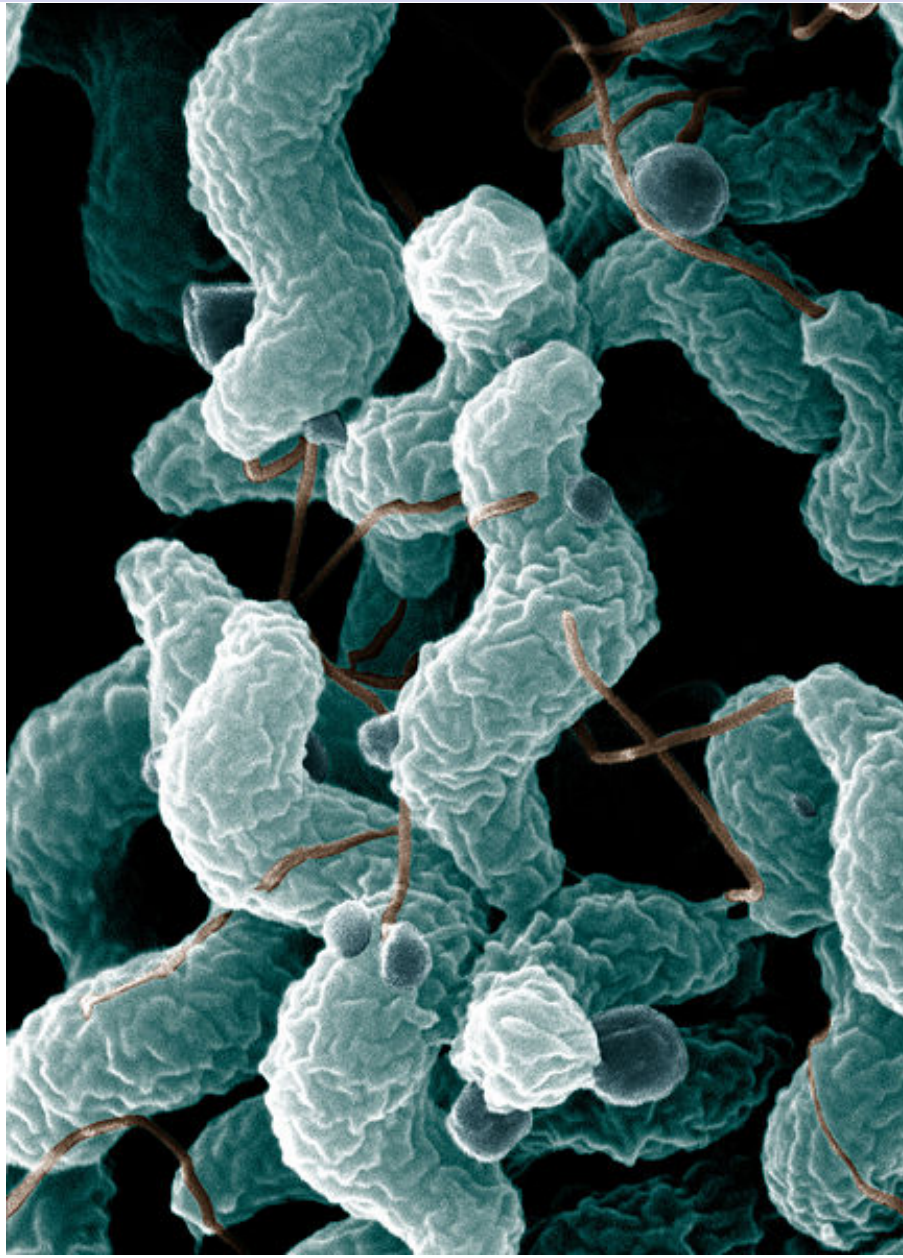
Zoonoses in New Zealand

- Five of the six most notified diseases in New Zealand are zoonoses.
- In 2012, there were over 10,000 notified cases of zoonoses.

Disease	Number	Rate	
Campylobacteriosis	7031	158.6	
Pertussis	5902	133.1	
Giardiasis	1719	38.8	
Salmonellosis	1085	24.5	
Cryptosporidiosis	877	19.8	
Yersiniosis	517	11.7	

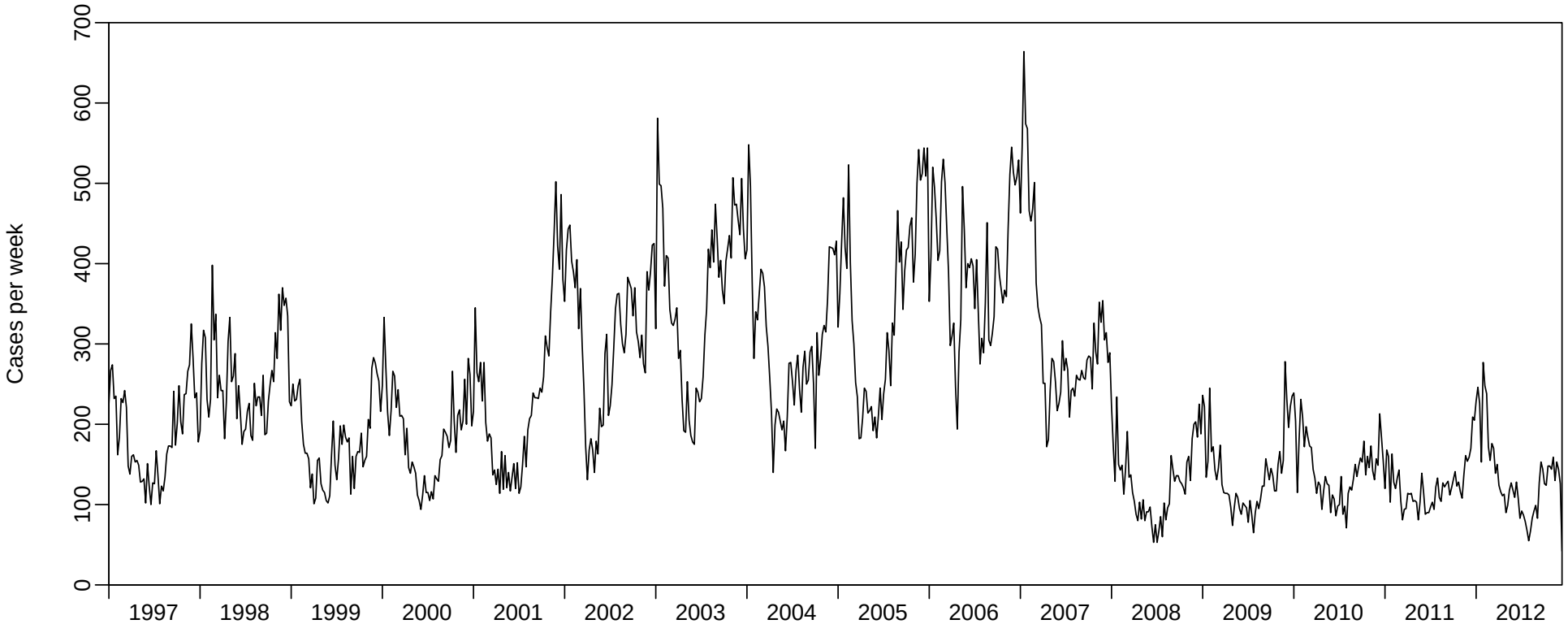
Campylobacter

- *Campylobacter* "twisted bacteria".
- Predominantly a food-borne illness.
- For every case notified, there are 7.6 community cases of campylobacteriosis.
  (Wheeler et. al., BMJ 1999)
- Each case is estimated to cost around $600.
  (Lake et. al., Risk Anal., 2009)



Goal 1: Determine the source of infection for human cases.

New Zealand campylobacteriosis cases



Goal 2: Does the proportion of human cases attributed to each source change seasonally?

Goal 3: Is the intervention in the poultry industry the related to the drop in the number of cases after 2007?

Data from the Manawatu

- Human isolates were collected from cases within the MidCentral DHB from 2005-2012.
- During 2005-2008, isolates were collected from chicken, beef and lamb from supermarkets in the Manawatu.
- In addition, water and environmental samples were collected from common swimming spots and neighbouring farms around the Manawatu.

Humans	Poultry	Sheep	Cattle	Water/Env
978	556	117	169	127

How can we determine the origin of a human case?

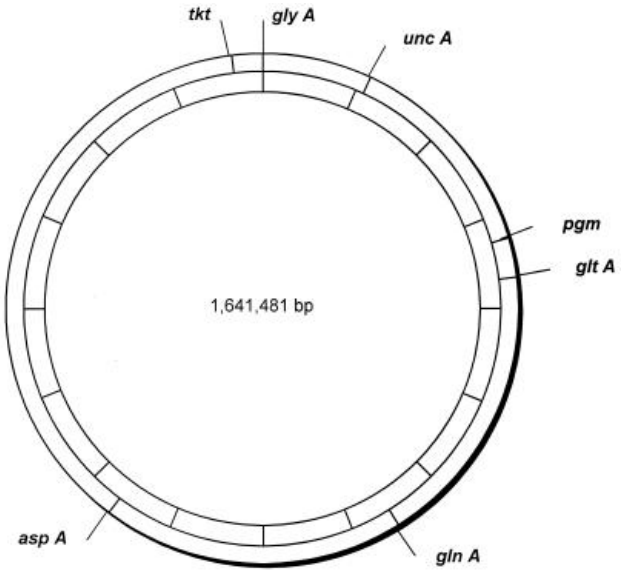
- The majority of cases are sporadic rather than being part of outbreaks.
- Epidemiological information associated with a case may be minimal.
- We often have no information on the exposure for sporadic cases.
- Often the only thing we have is genotype information for each case.

MLST for *Campylobacter*

Seven housekeeping genes (loci).

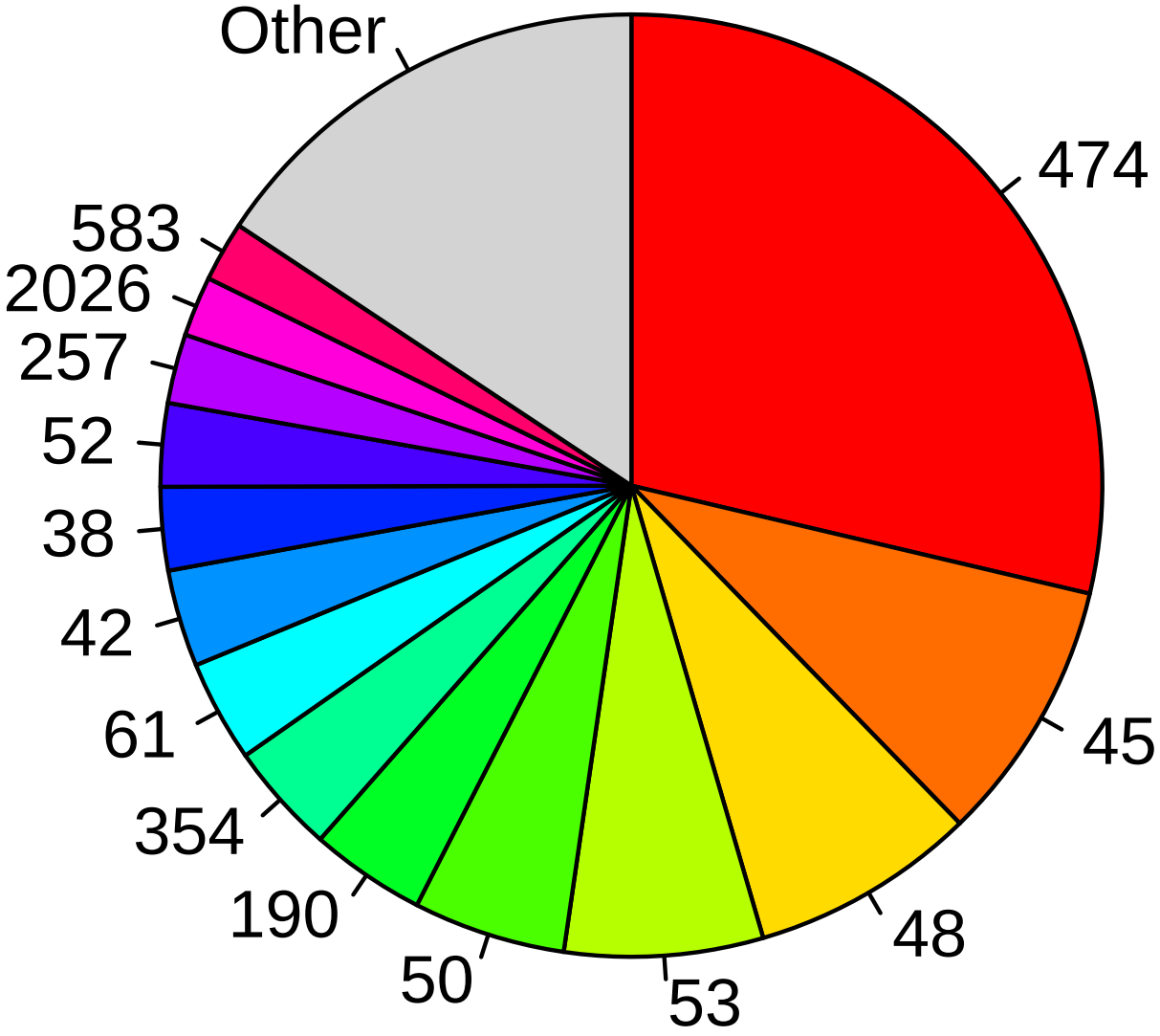
Unique genes are assigned different numbers (alleles).

The combination of alleles at the seven loci gives the **multilocus sequence type**.



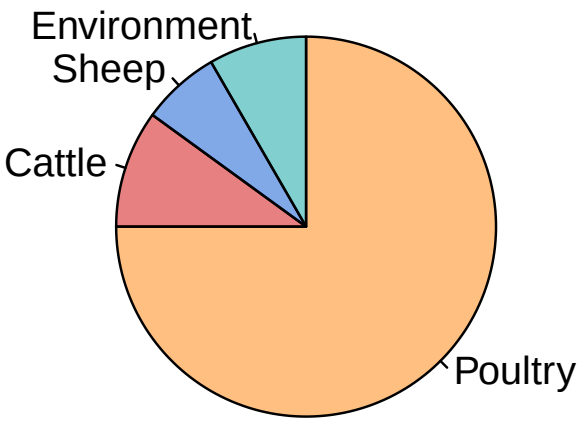
ST	aspA	glnA	gltA	glyA	pgm	tkt	uncA
474	2	4	1	2	2	1	5
61	1	4	2	2	6	3	17
190	2	1	5	3	2	3	5
2381	175	251	216	282	359	293	102
48	2	4	1	2	7	1	5

Human MLST types

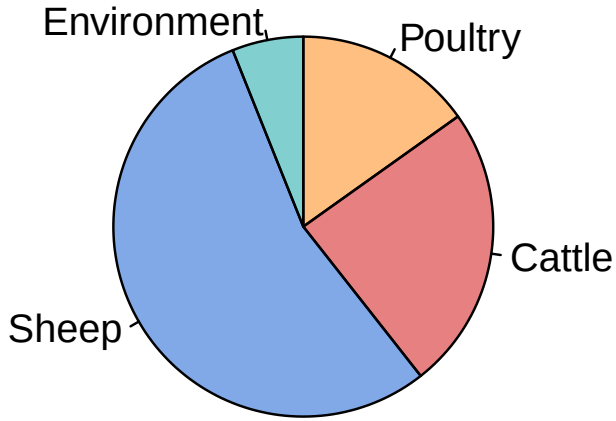


Source specific types

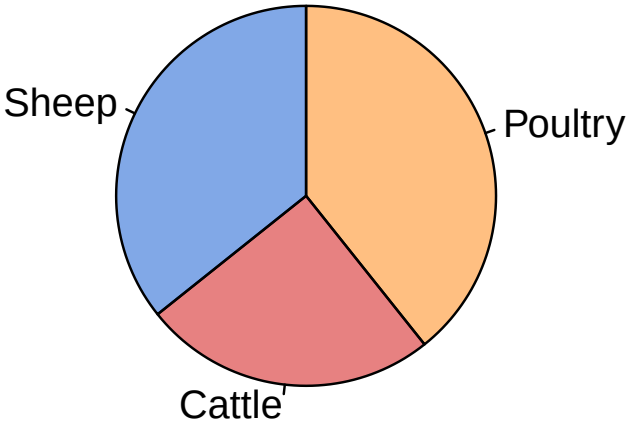
ST-474: $n = 60$



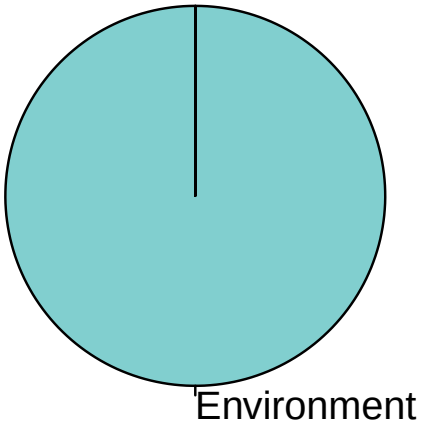
ST-61: $n = 33$



ST-190: $n = 28$



ST-2381: $n = 22$



Attribution using MLST

Assign each human ST to the most likely source, given the distribution of STs on the source.

Use **Bayes' Theorem**

$$P(\text{source} = k|\text{ST} = i) = \frac{P(\text{ST} = i|\text{source} = k)P(\text{source} = k)}{\sum_k P(\text{ST} = i|\text{source} = k)P(\text{source} = k)}$$

where

- $P(\text{ST} = i|\text{source} = k)$ is the distribution of STs on each source.
- $P(\text{source} = k)$ is the prior probability that an isolate picked at random is from source k.

Island model (D. Wilson, 2008)

Model the distribution of allelic profiles $P(\text{ST}|\text{source})$ on each source by assuming that the observed sequences arise due to:

- **Mutation**, where an allele at a locus is novel.
- **Migration** between sources, where the allelic profile has been observed before in one of the sources, including the current one.
- **Recombination**, where the allele at a given loci has been observed before but not in this allelic profile.

Island model

ST	aspA	glnA	gltA	glyA	pgm	tkt	uncA
474	2	4	1	2	2	1	5
?	2	4	1	2	29	1	5

We have a novel allele at the pgm locus. We assume this genotype has arisen through **mutation**.

Island model

ST	aspA	glnA	gltA	glyA	pgm	tkt	uncA
474	2	4	1	2	2	1	5
?	2	4	1	2	1	1	5

The pgm allele looks familiar...

ST	aspA	glnA	gltA	glyA	pgm	tkt	uncA
45	4	7	10	4	1	7	1
3718	2	4	1	4	1	1	5

But we haven't seen this genotype before. We assume it arose through **recombination**.

Island model

ST	aspA	glnA	gltA	glyA	pgm	tkt	uncA
474	2	4	1	2	2	1	5
?	2	4	1	2	2	1	5

This is just 474 - we've seen this before, but possibly not in this source. We assume it arose through **migration**.

Island model

The probability of observing $y = (y^1, y^2, \ldots, y^7)$ in source k is

$$\phi(y|k,X) = \sum_{c \in X} \frac{M_{S_c k}}{N_{S_c}} \prod_{l=1}^{7} \begin{cases} \mu_k & \text{if } y^l \text{ is novel,} \\ (1-\mu_k) R_k \sum_{j=1}^{K} M_{jk} f^l_{y^l j} & \text{if } y^l \neq c^l \\ (1-\mu_k) \left[1 - R_k (1 - \sum_{j=1}^{K} M_{jk} f^l_{y^l j}) \right] & \text{if } y^l = c^l \end{cases}$$

where

- X are the previously observed sequences.
- c comes from source S_c.
- μ_k be the probability of a novel mutant allele.
- R_k be the probability that a type has undergone recombination.
- M_{jk} be the probability of an allele migrating from source j to k.
- f^l_{aj} be the frequency with which allele a has been observed at locus l in those genotypes sampled from source j.

Estimating parameters

The likelihood for observing source types Y given μ, M, and R may be approximated using a leave-one-out approach.

$$p(Y|\mu, M, R) \sim \prod_{y \in Y} \phi(y|S_y, Y - \{y\}).$$

We fit this using MCMC with switching and lognormal Metropolis-Hastings proposals for μ, M, and R.

Attribution of human isolates to sources

Let F_k be the proportion of human isolates attributed to source k.

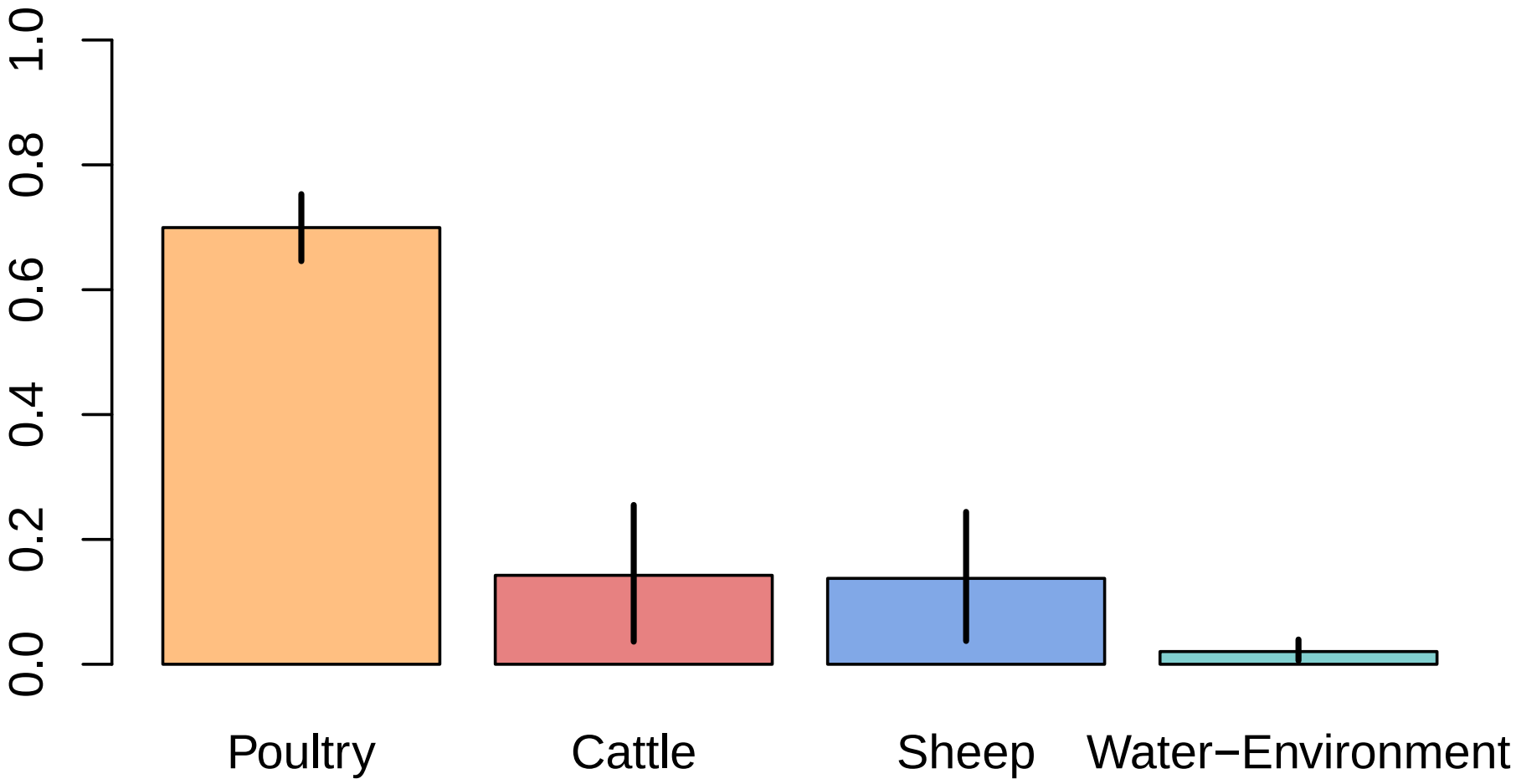
Then the posterior distribution for F_k given human isolates H and source isolates Y is

$$p(F_k|H, Y) \propto \prod_{h \in H} \sum_k F_k \phi(h|k, Y) p(F_k)$$

where $p(F_k)$ is the prior where we assume each source is equally likely.

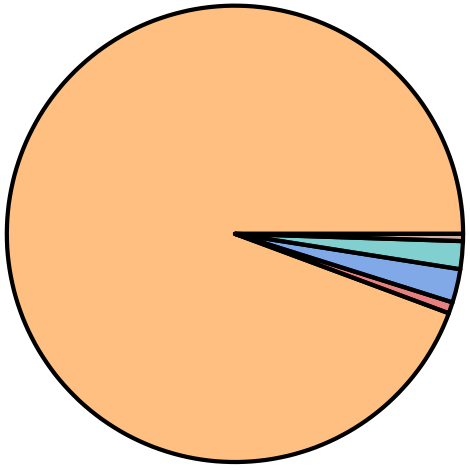
This depends on μ, M and R, thus for each posterior sample from their MCMC chain we run a side chain to estimate F_k.

Proportion of human cases attributed to each source

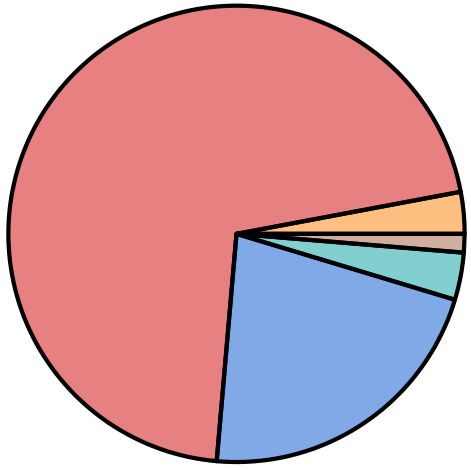


Migration, mutation, and recombination

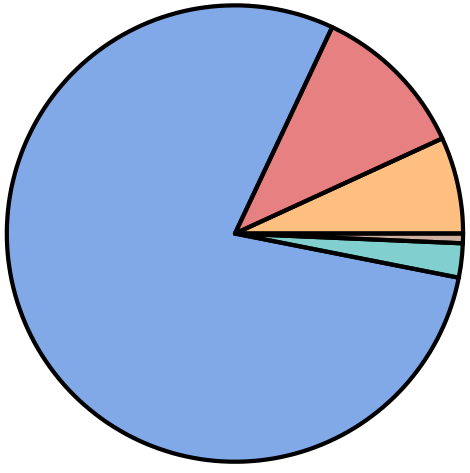
Poultry



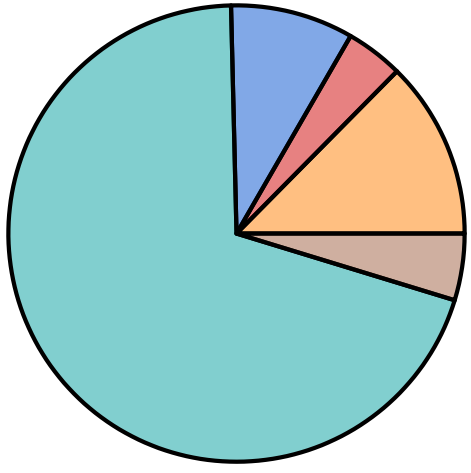
Cattle



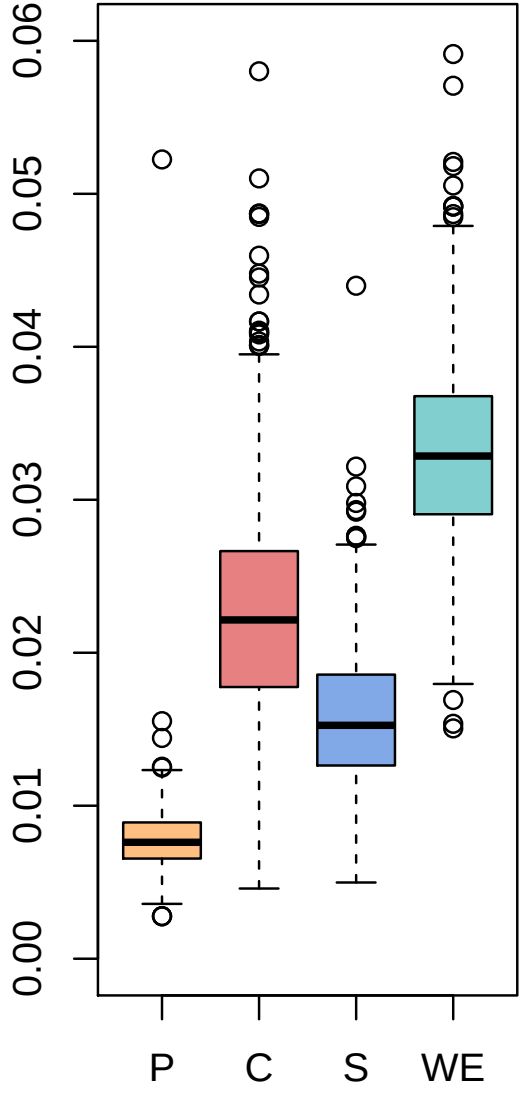
Sheep



Water-Environment



Recombination



Seasonal model for F_k

Suppose F_k can change through time, and let

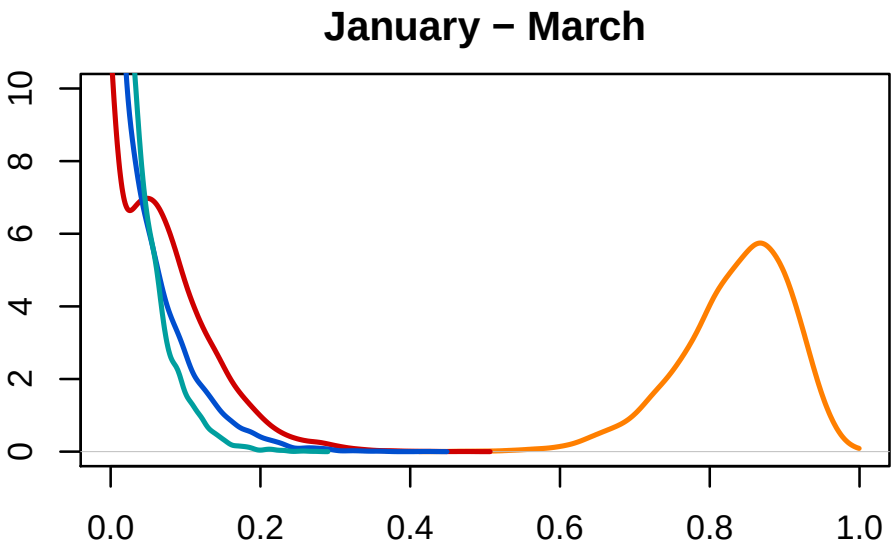
$$F_{kt} = \begin{cases} \dfrac{e^{-f_{kt}}}{1 + \sum_{k=1}^{K-1} e^{-f_{kt}}} & k = 1 \ldots K-1, \\ \dfrac{1}{1 + \sum_{k=1}^{K-1} e^{-f_{kt}}} & k = K. \end{cases}$$

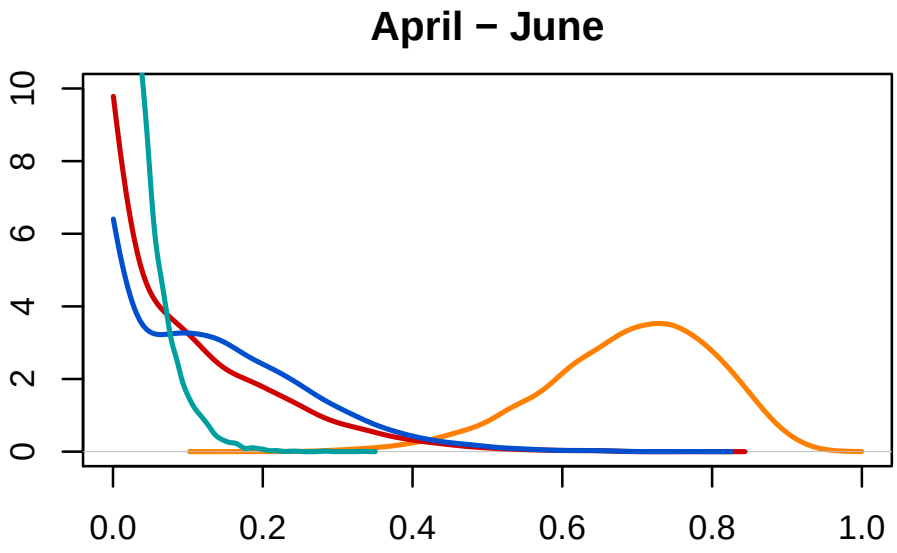
where

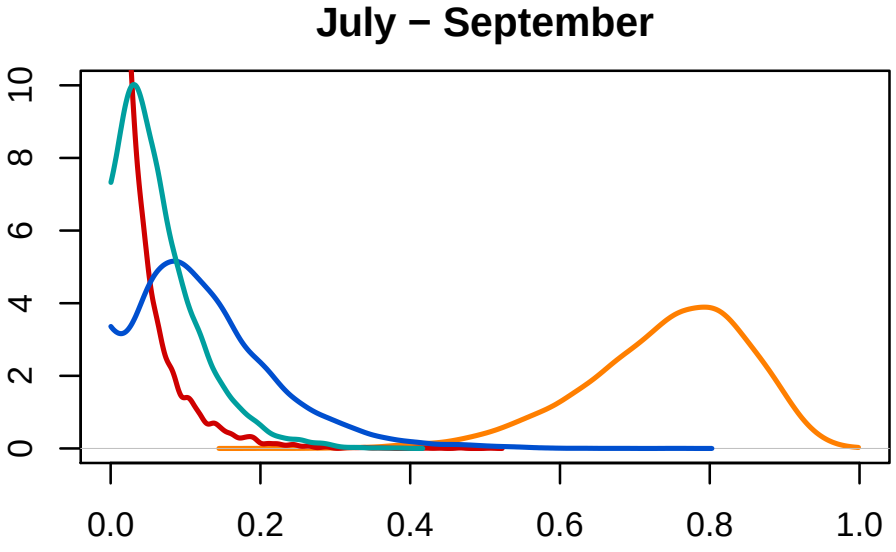
$$f_{kt} = \sum_{i=1}^{4} \alpha_{ik} \mathbf{1}_{t \in \text{Season}_i} + \beta_k \mathbf{1}_{t \geq 2007} + \epsilon_{kt},$$

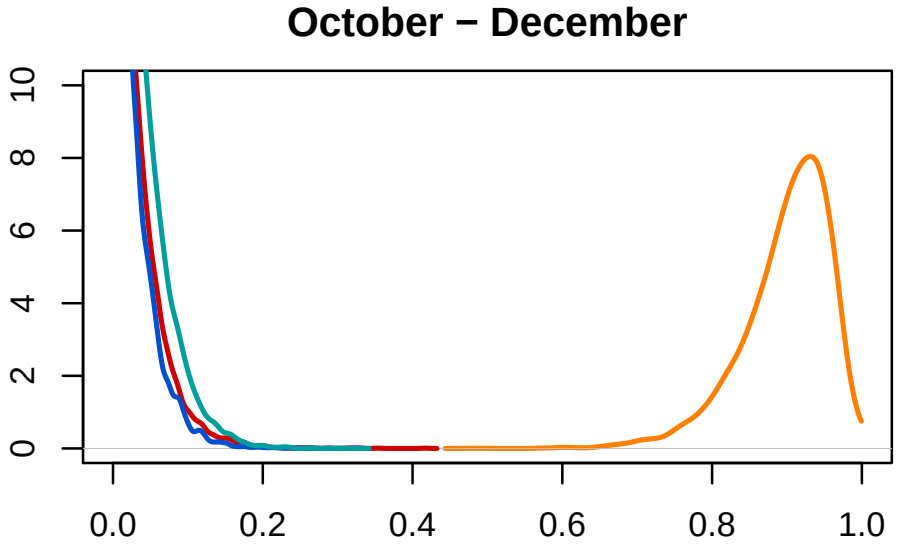
$$\epsilon_{kt} \sim \text{Normal}(\rho_k \epsilon_{k(t-1)}, \sigma_k^2).$$

Proportions: pre-2007

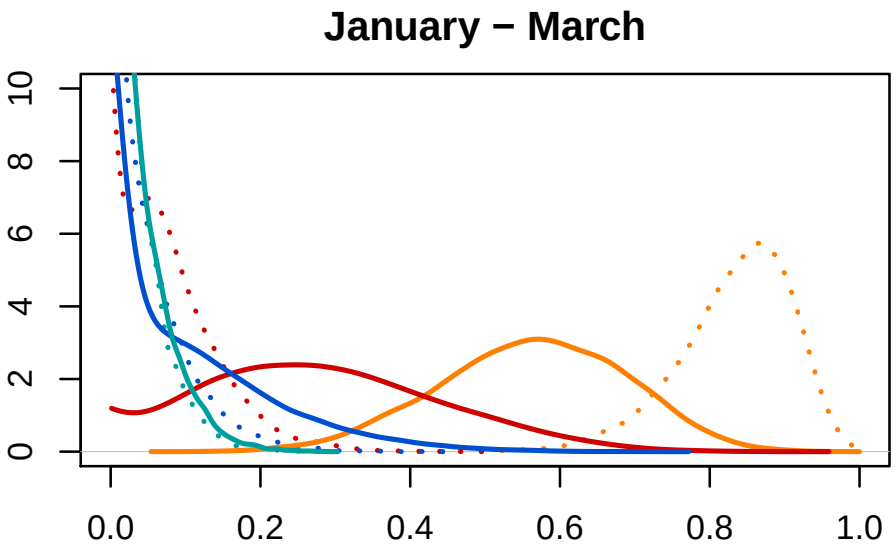


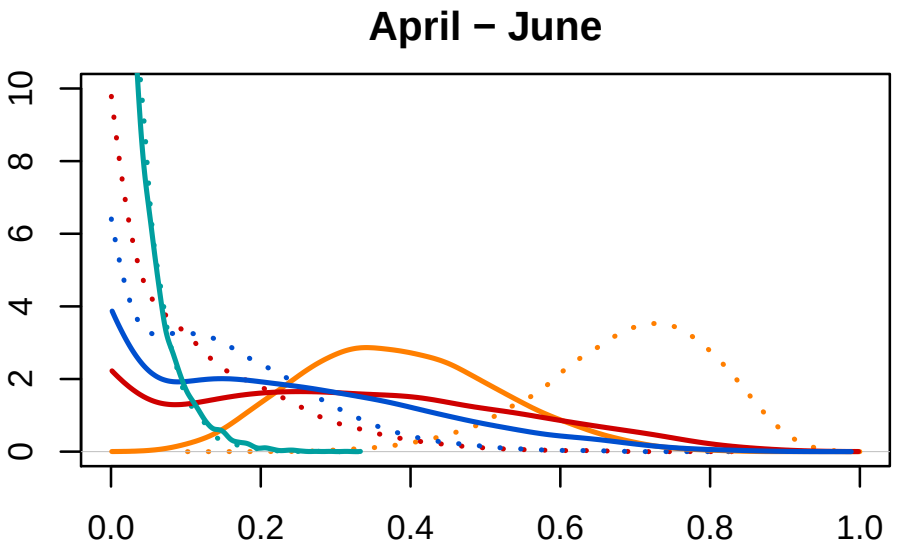


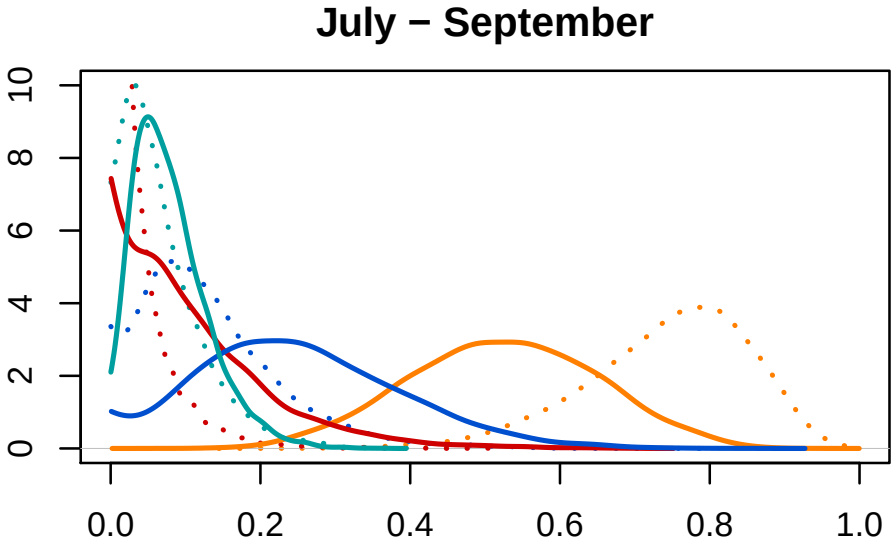


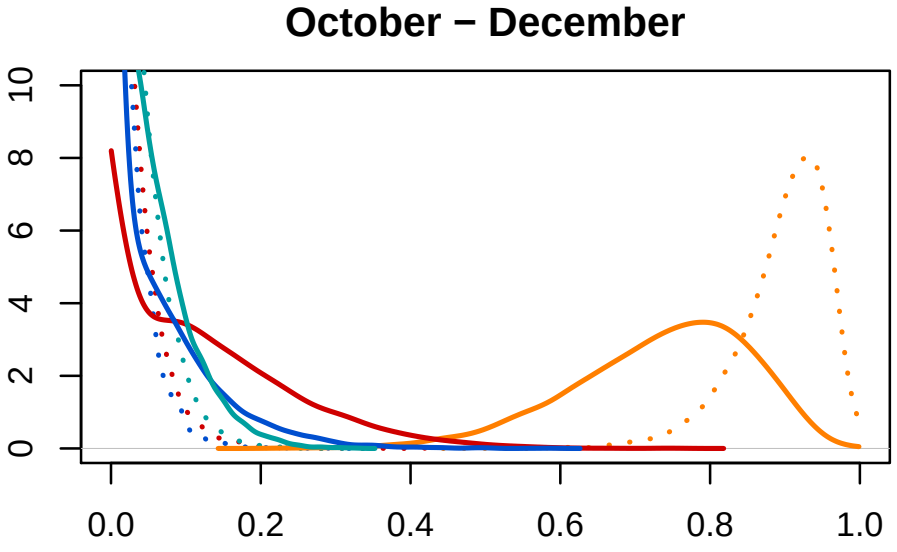


Proportions: post-2007

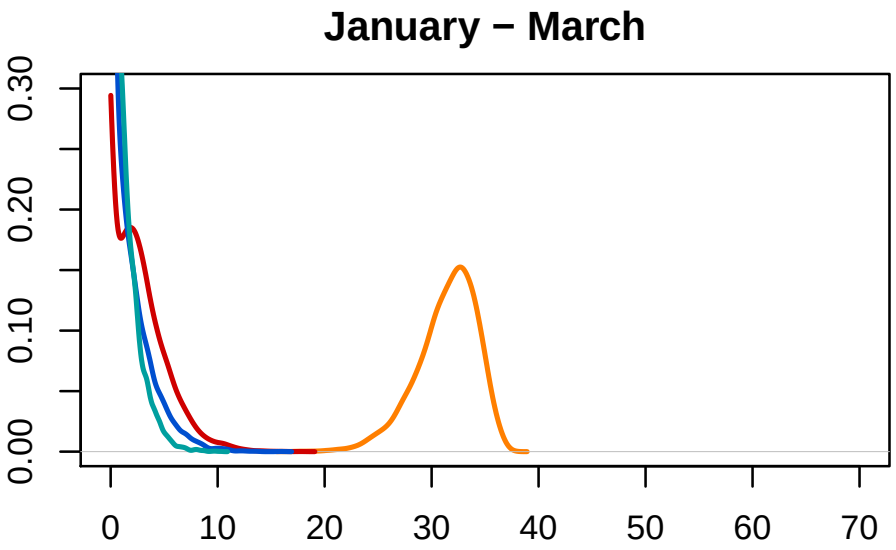


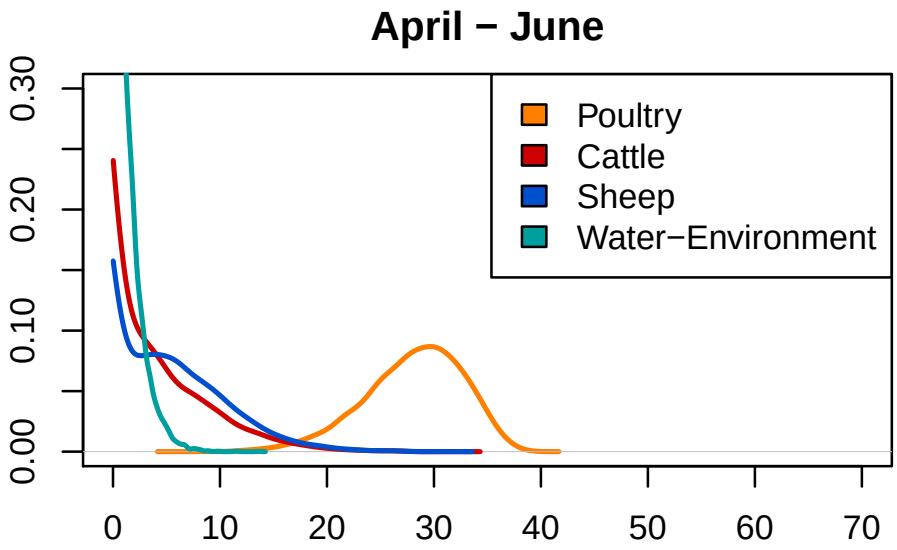


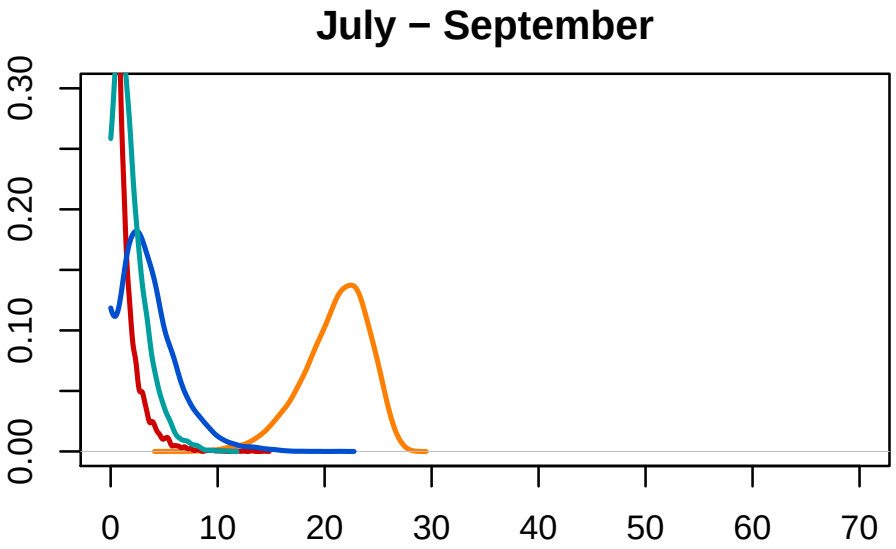


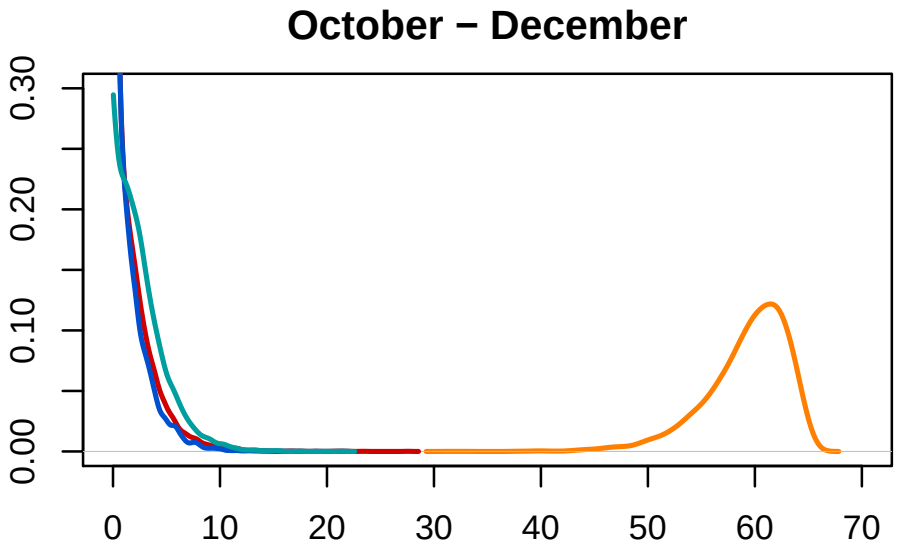


Totals: pre-2007

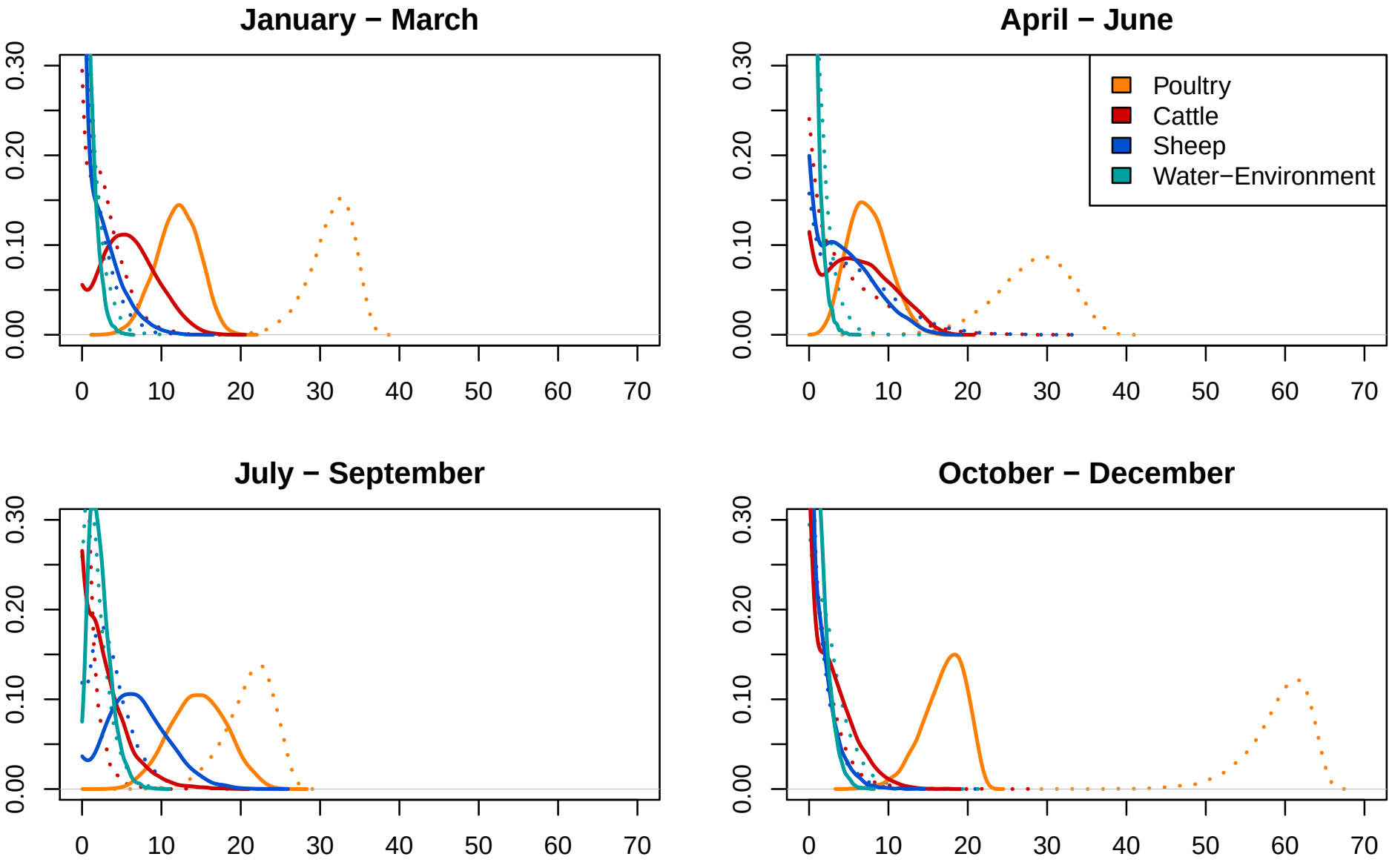




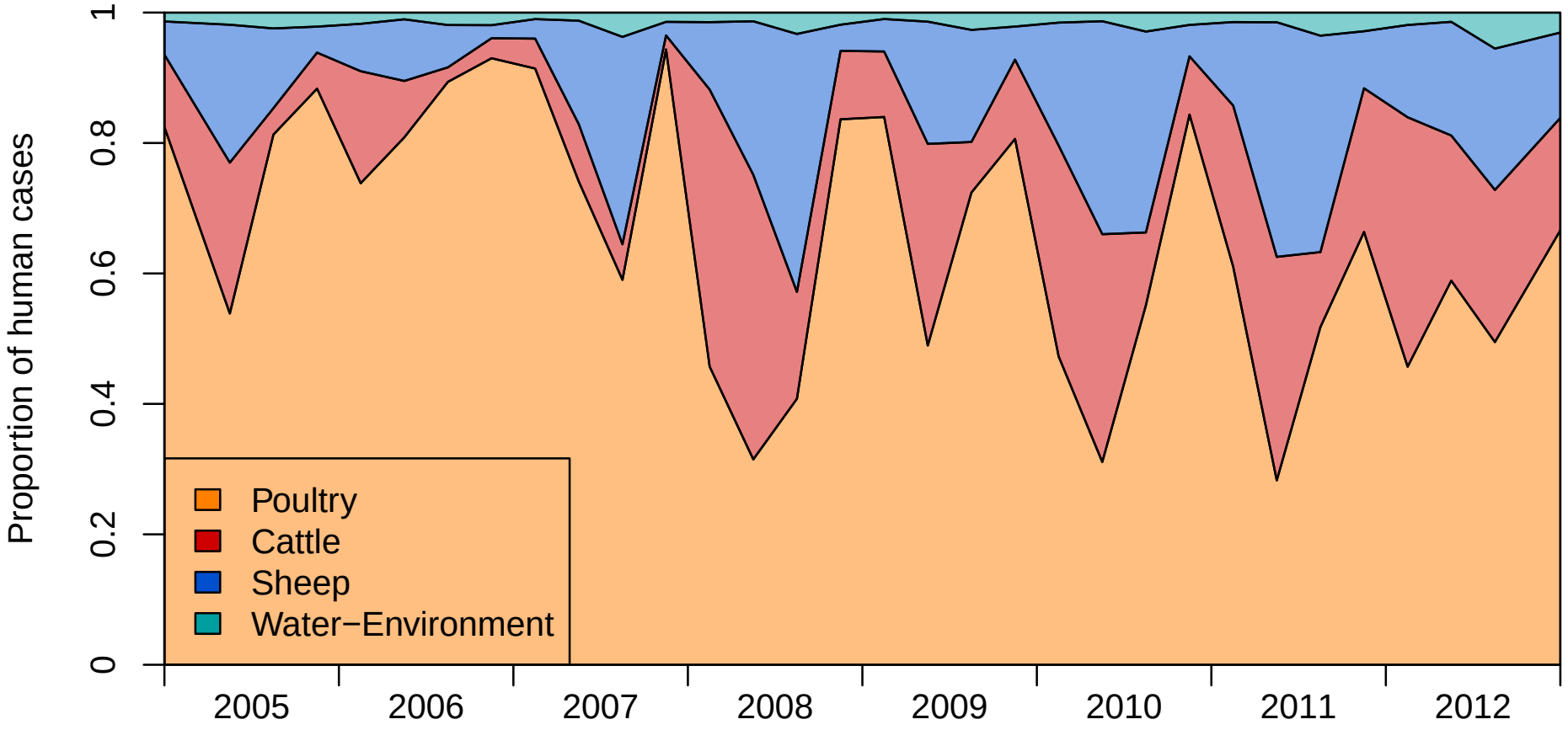




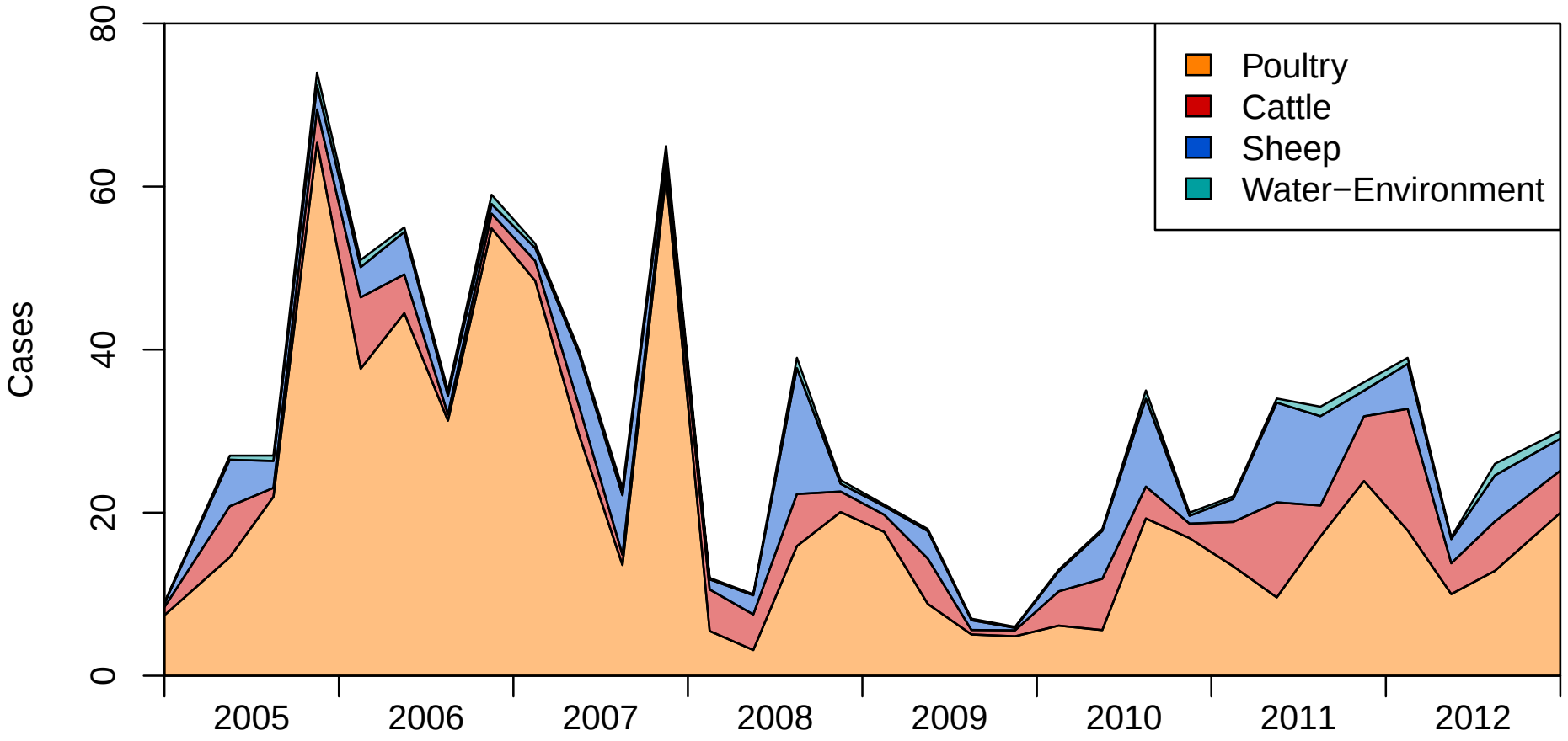
Totals: post-2007



Campylobacter in the Manawatu



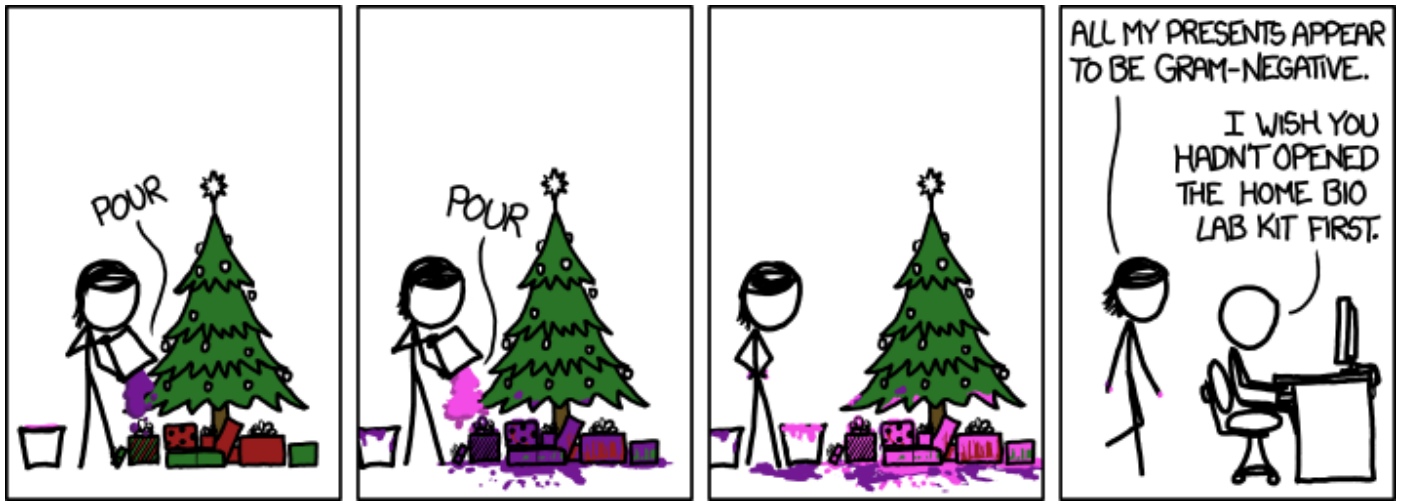
Campylobacter in the Manawatu



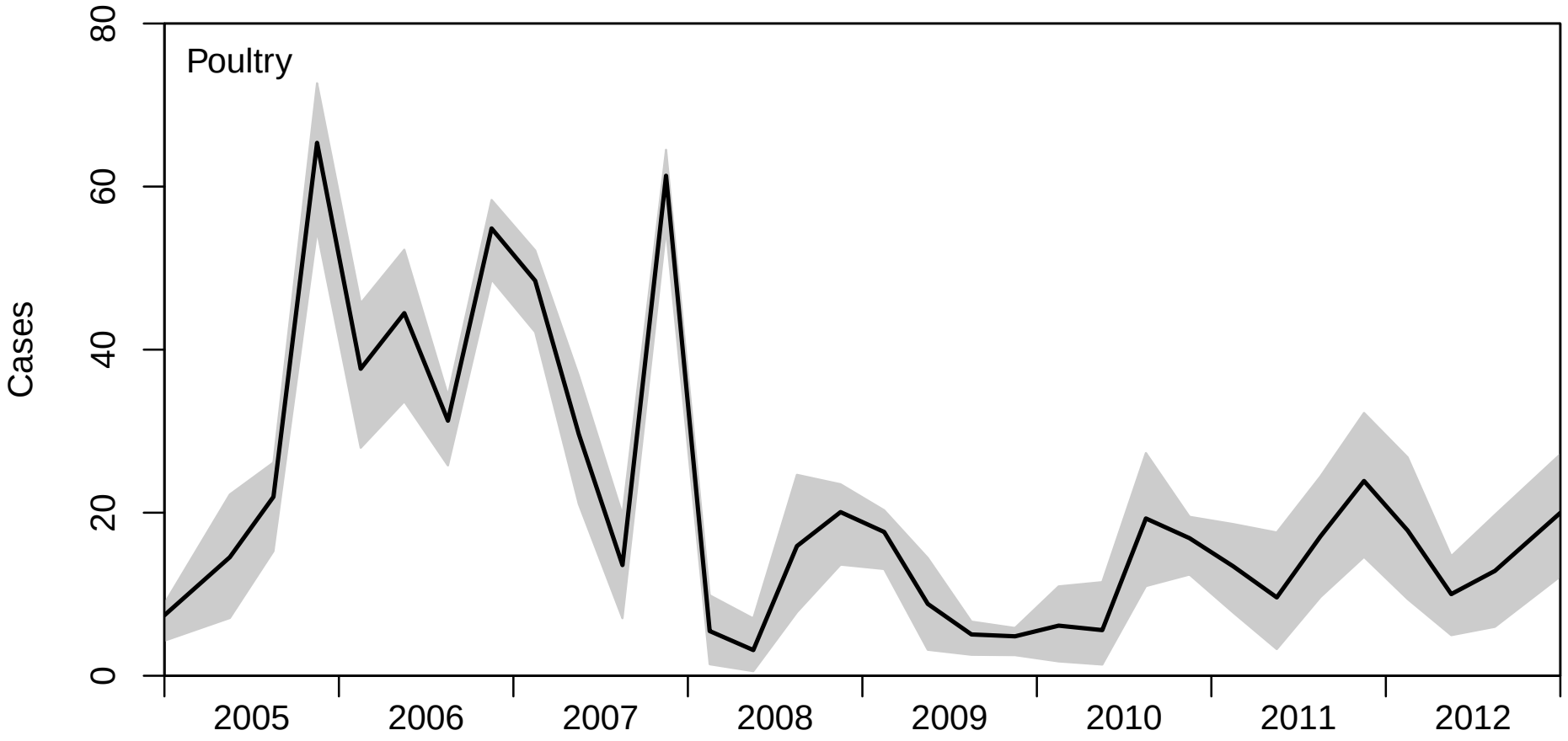
Summary

- Genotyping information allows us to assign human cases to their most likely source.
- The majority of cases are attributed to poultry.
- There are interesting seasonal trends, with more poultry cases in summer and more ruminant cases in winter.
- The water/environmental strains don't seem to be associated with human illness, but do show higher probabilities of mutation and recombination.
- The poultry intervention in 2007 appears to be the main reason for the decrease in total human cases.

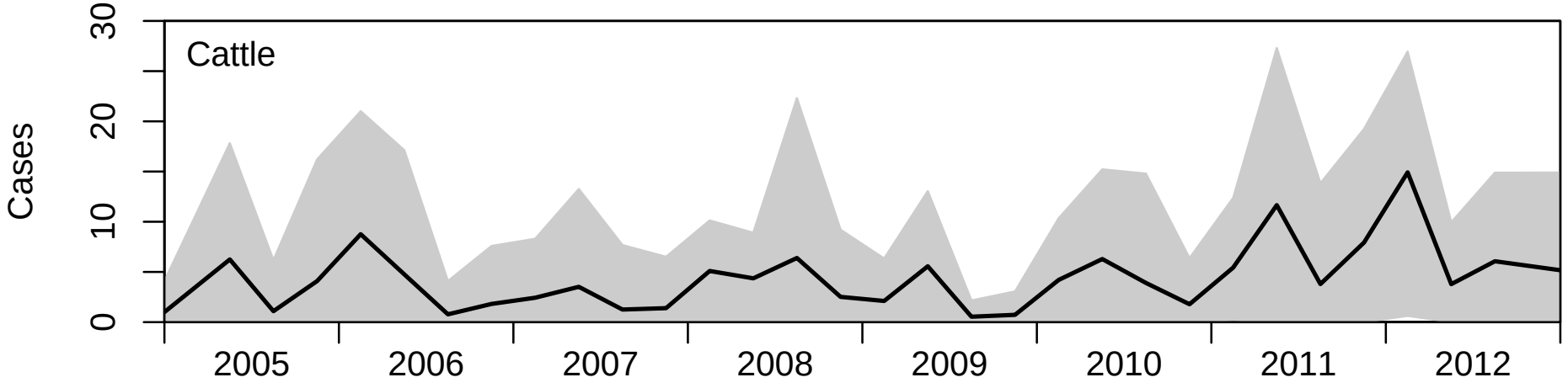
Thanks for listening

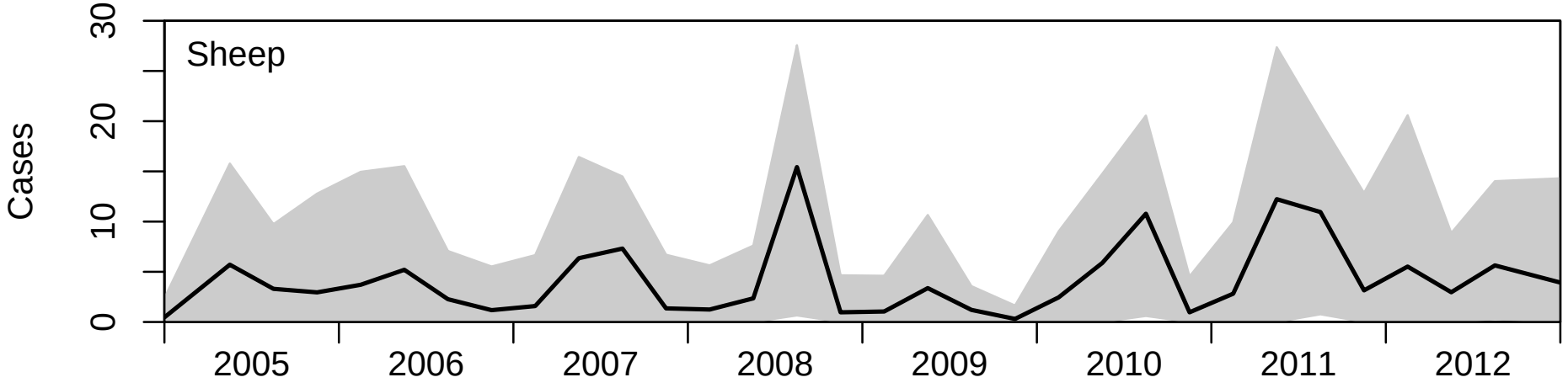


Campylobacter in the Manawatu

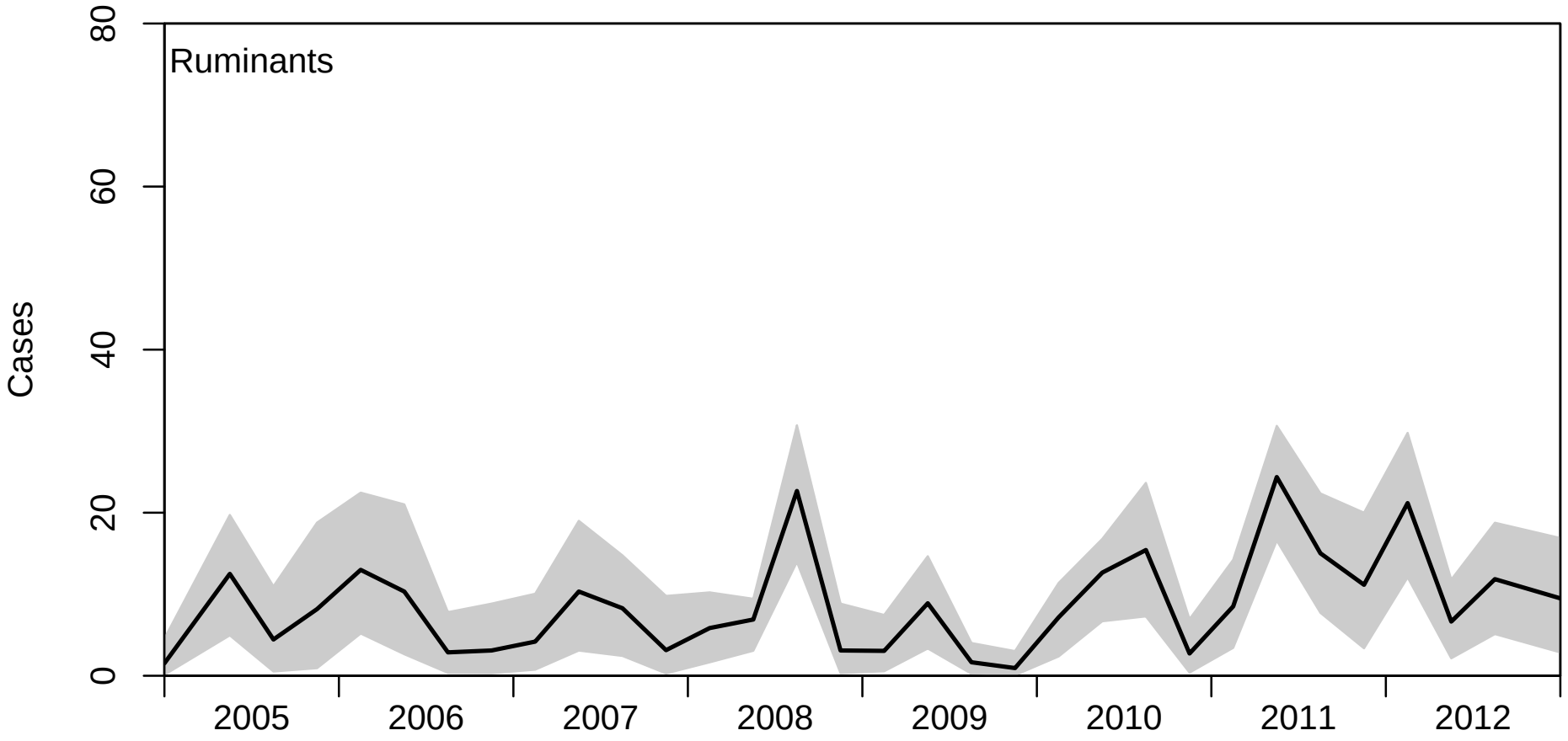


Campylobacter in the Manawatu

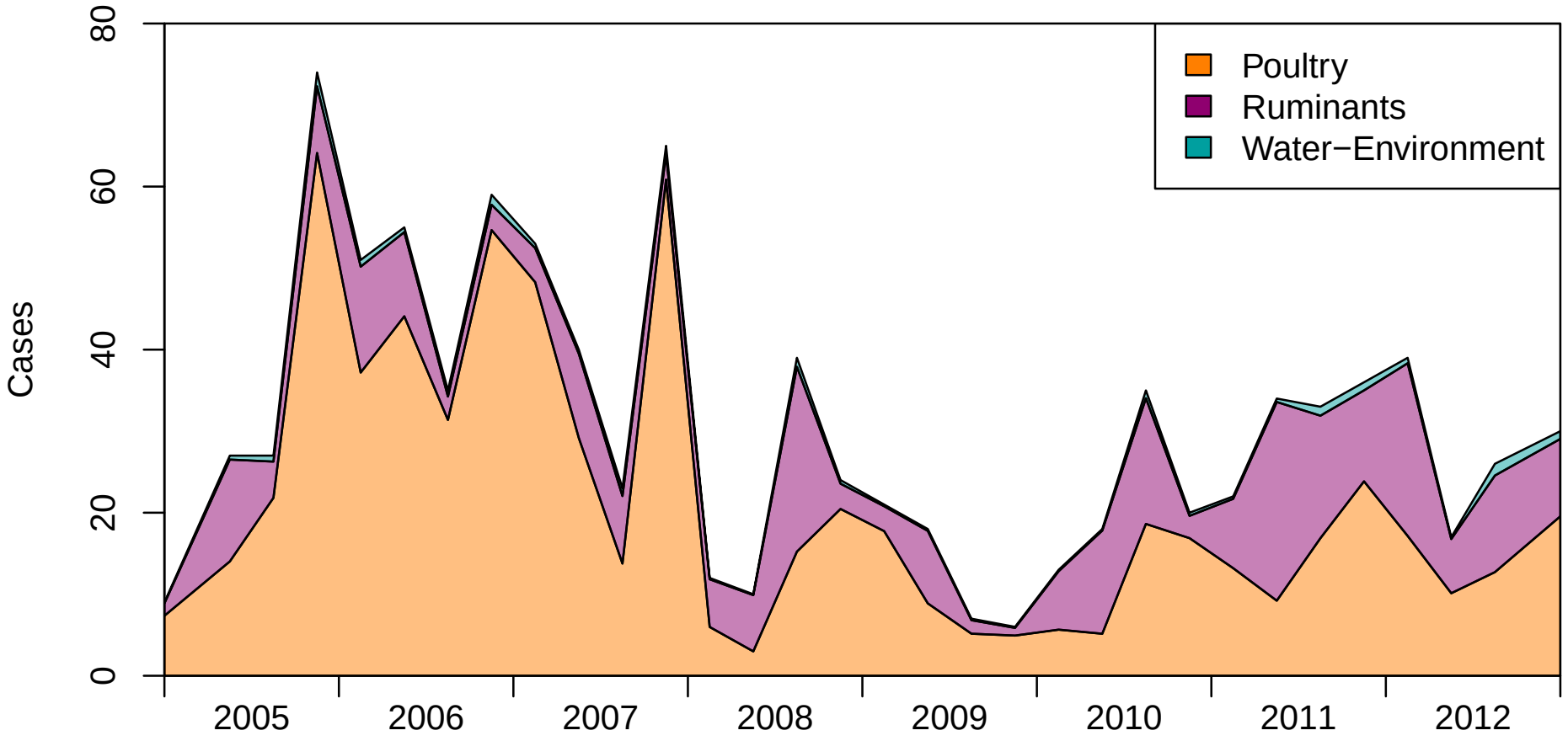




Campylobacter in the Manawatu



Campylobacter in the Manawatu



Campylobacter in the Manawatu