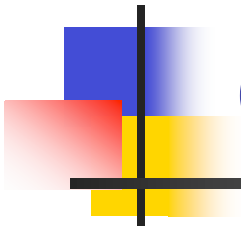
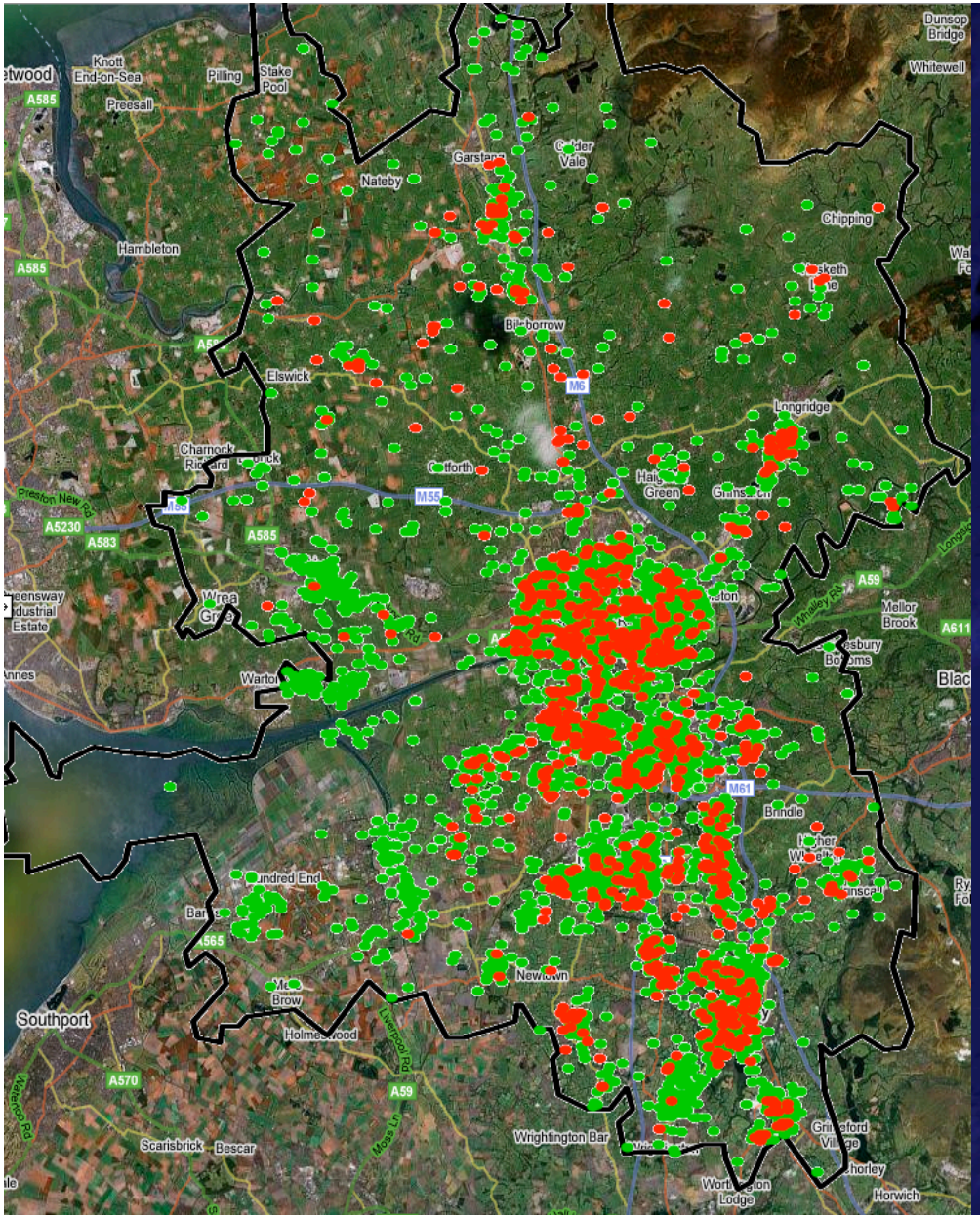


Campylobacter jejuni

Epidemiology and Evolution

Veterinary Training Research Initiative
Food-borne zoonotic pathogens: Transmission, pathogen evolution and control

- Lancaster University
  - Paul Fearnhead, Edith Gabriel, Peter Diggle
- University of Liverpool
  - Howard Leatherbarrow, Tony Hart, Malcolm Bennett
- Health Protection Agency
  - Andrew Fox, Steve Gee, Sam James
  - John Cheesbrough, Eric Bolton
- Funding: DEFRA
  - (Department for Environment, Food and Rural Affairs, UK Government)

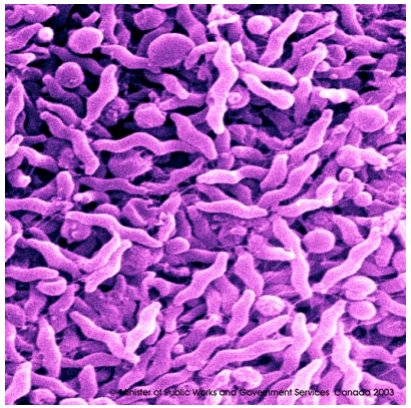


Characterisation of *C. jejuni* infection

- Epidemiology
- Evolutionary history
- Population genetics
- Source of human cases

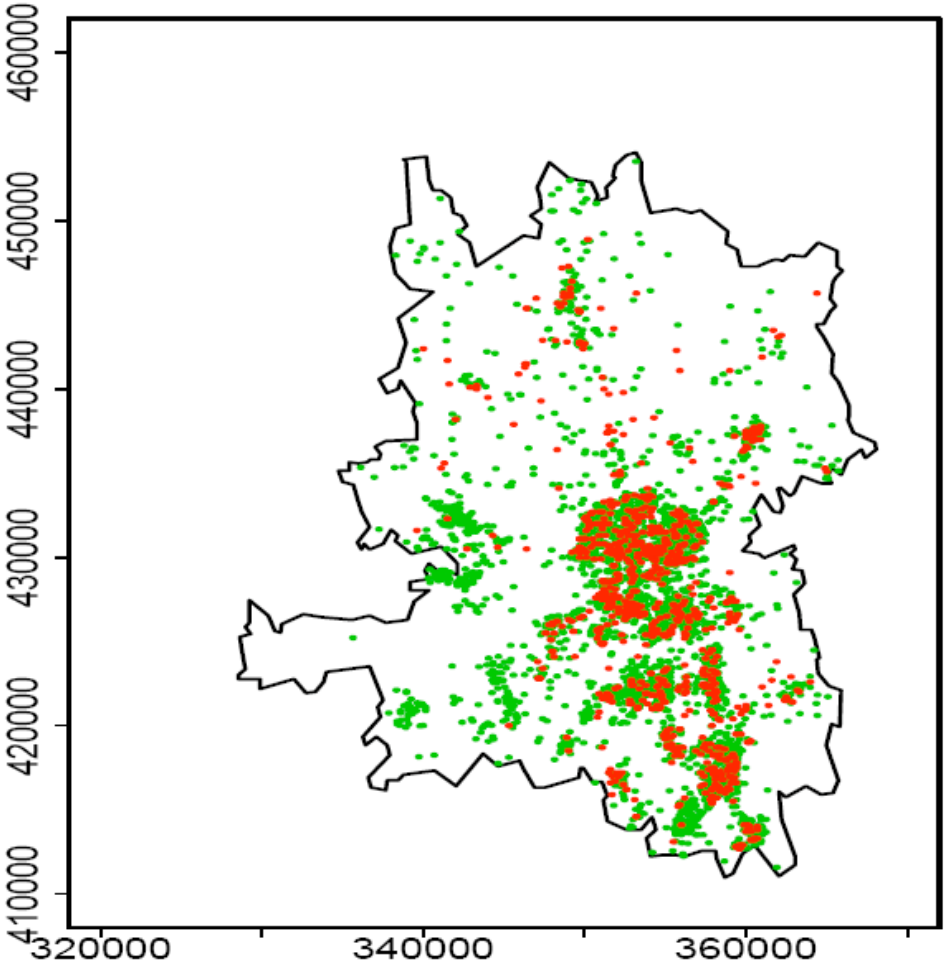
Foodborne illness in the UK
Food Standards Agency figures for 2000

Salmonella	16,987	20.9%
Campylobacter	**62,867**	**77.3%**
E.coli O157	1,147	1.4%
Clostridium perf.	166	0.2%
Listeria	113	0.1%
Total	81,280	

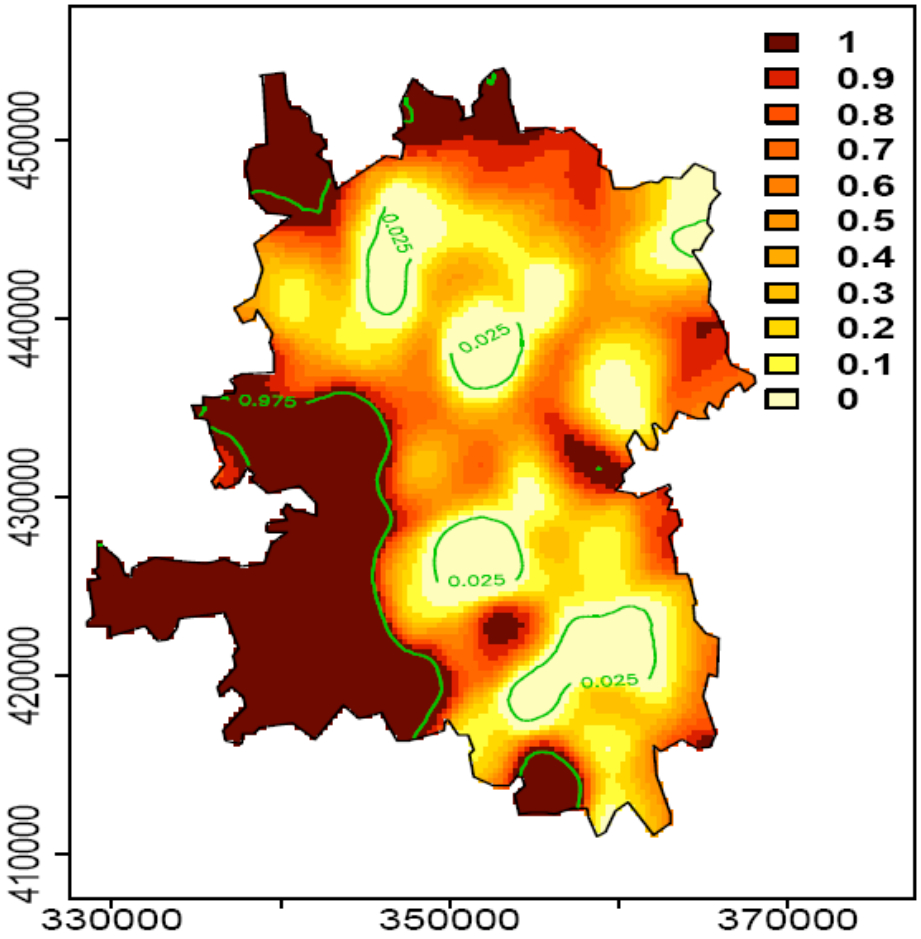


$8bn
Annual cost to US economy
Buzby et al. JID (1997)

Cases and controls



Significance



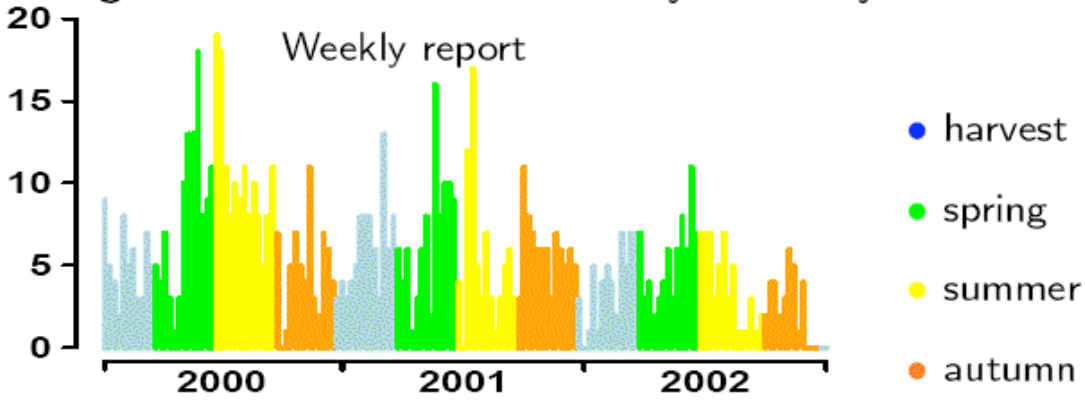
Poisson point process

- **Temporal variation**

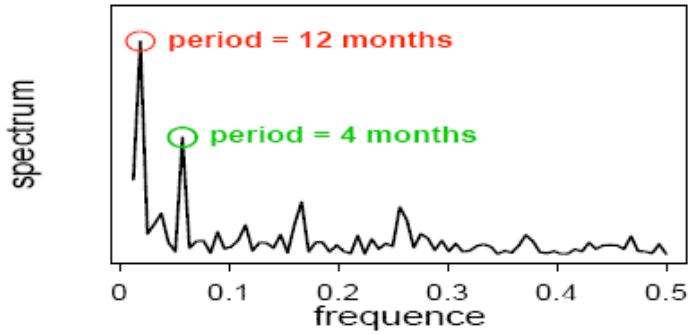
⋆ Day of week effect: less cases reported during the week-end.

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
151	192	171	174	153	29	11

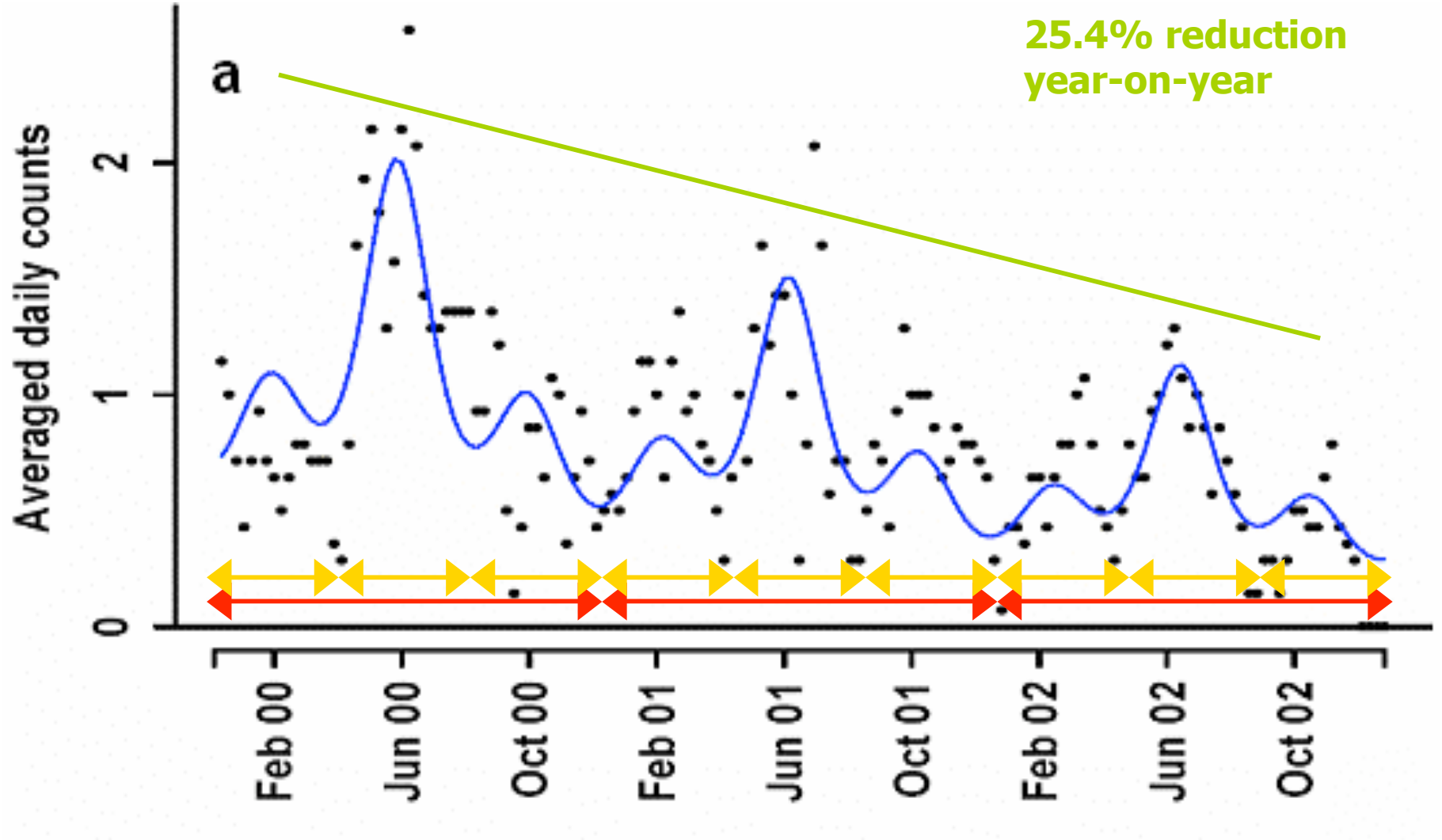
⋆ Decreasing trend: 25% reduction year-on-year.



⋆ Annual and 4-monthly periodicity.

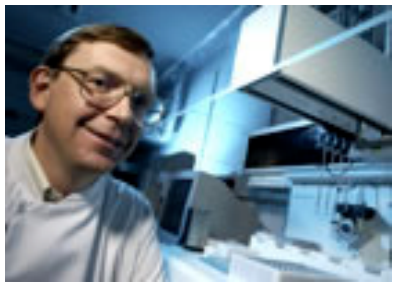


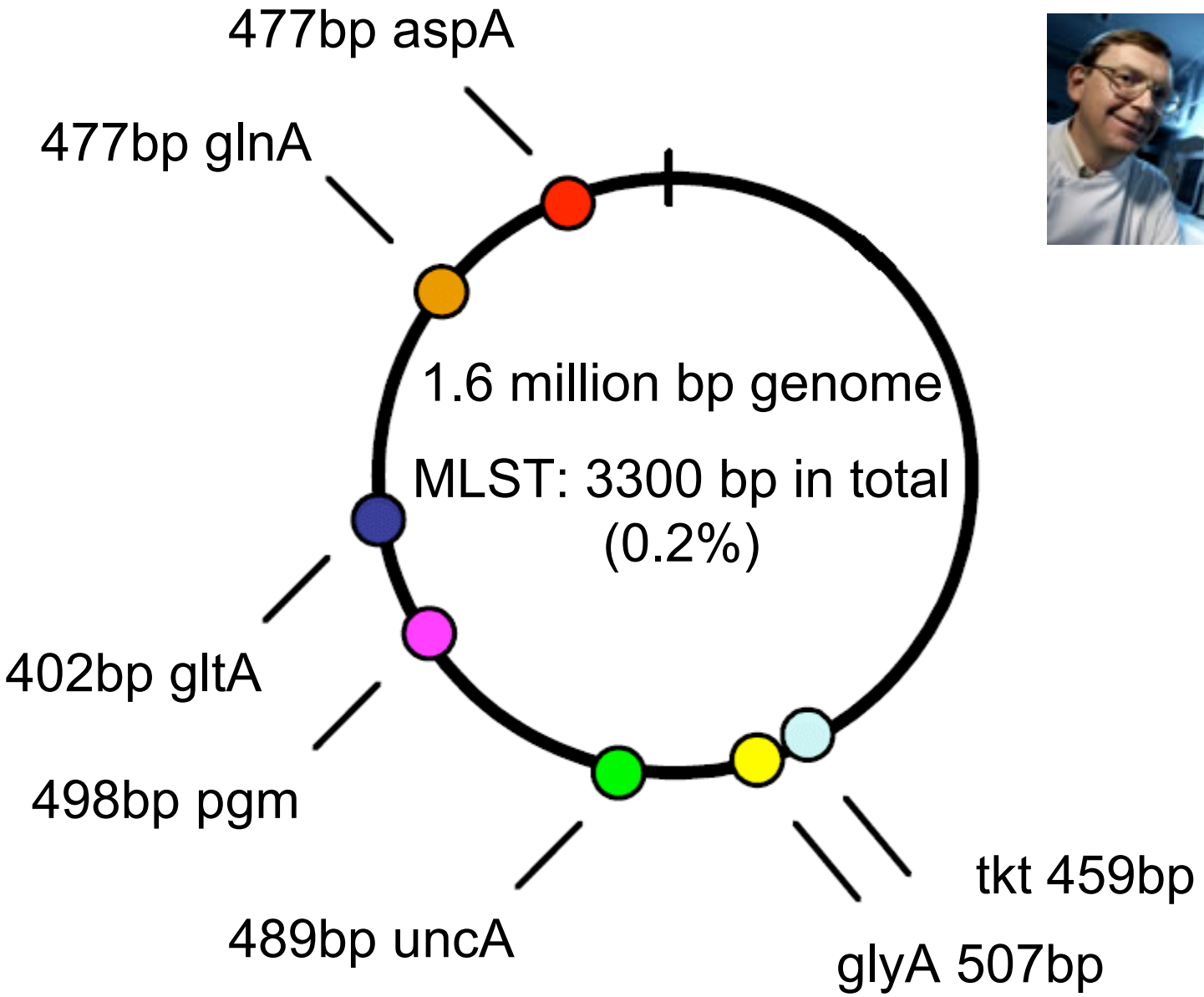
Seasonal patterns



Harmonic regression

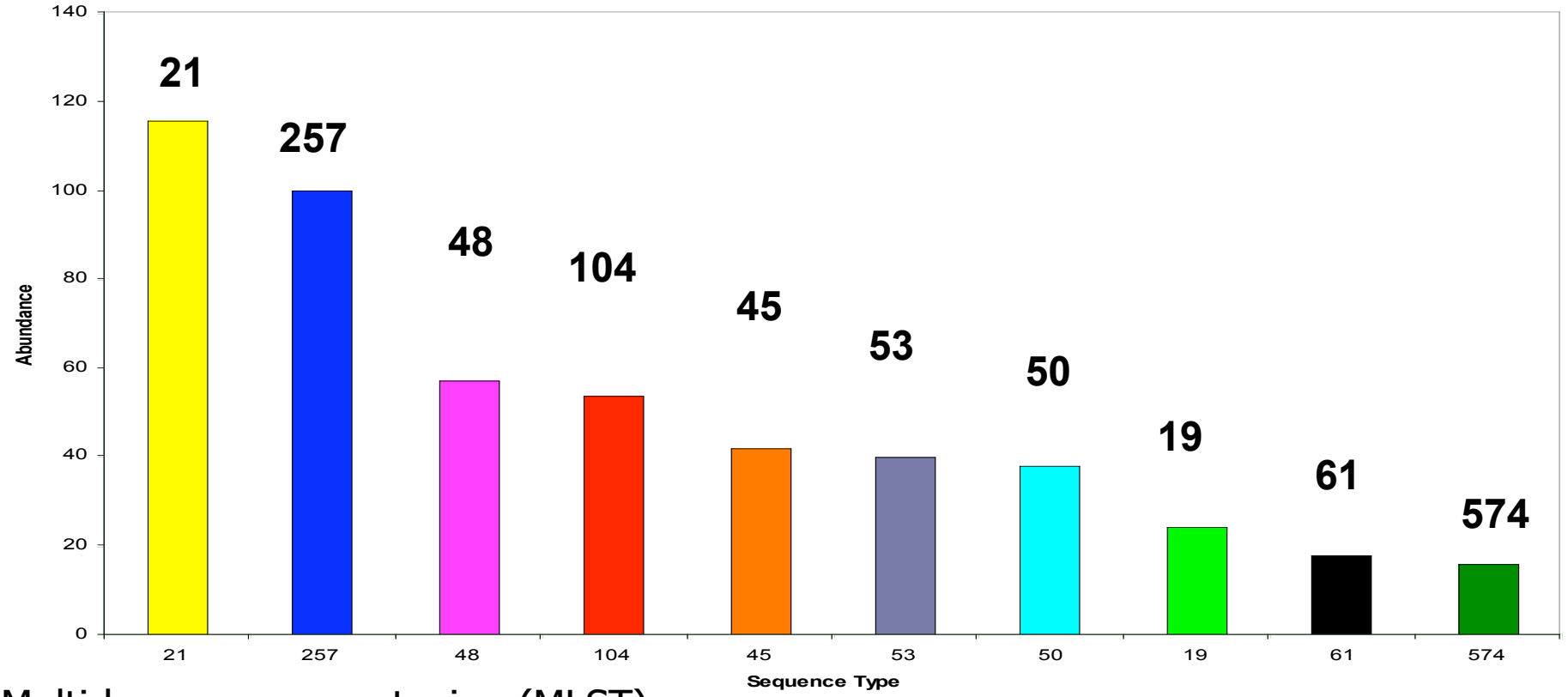
MLST: Multi-locus sequence typing



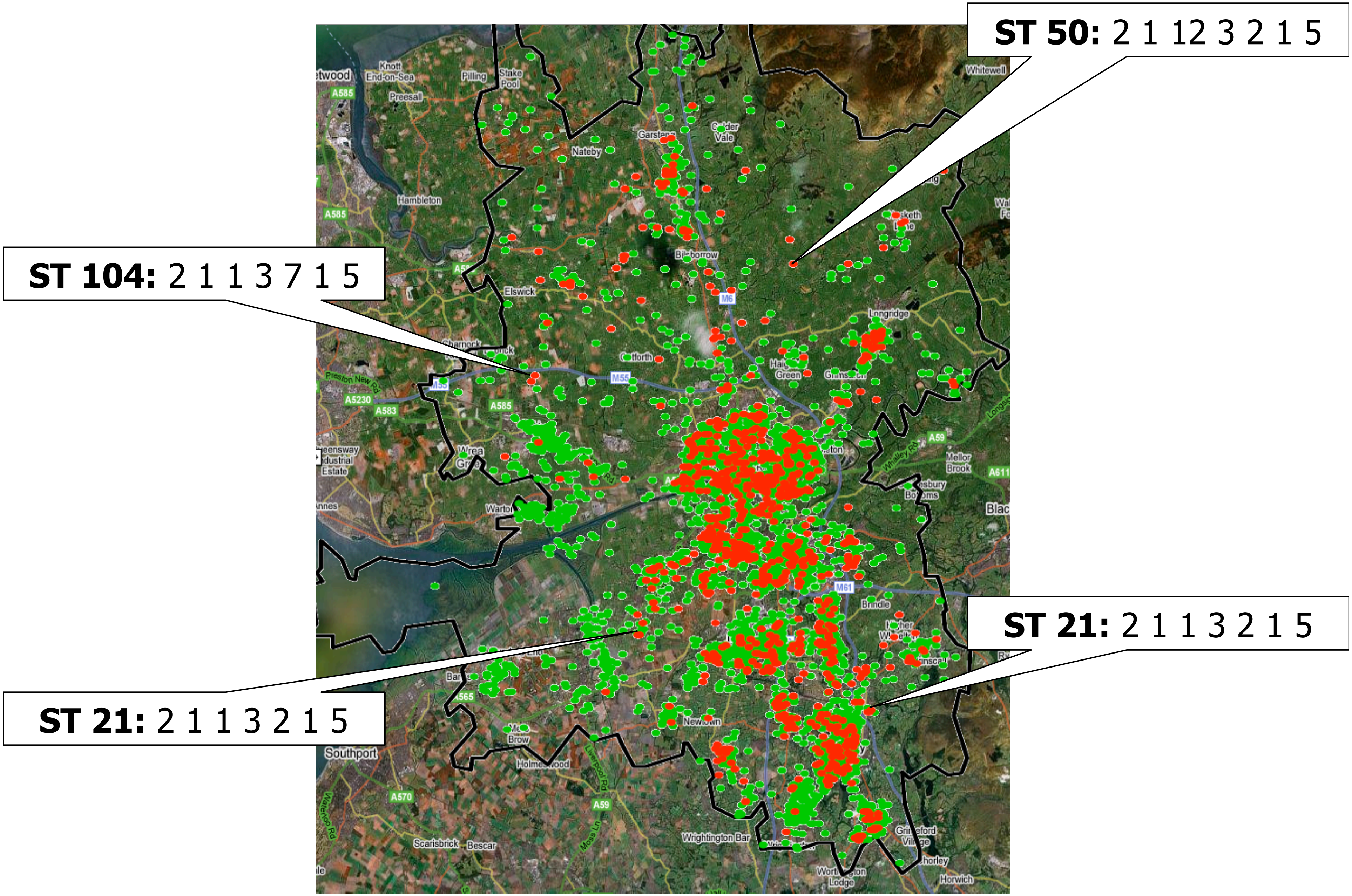


Multi-locus sequence typing (MLST)

st	aspA	glnA	gltA	glyA	pgm	tkt	uncA	freq
21	2	1	1	3	2	1	5	116
104	2	1	1	3	7	1	5	54
53	2	1	21	3	2	1	5	40
50	2	1	12	3	2	1	5	38
19	2	1	5	3	2	1	5	24



Multi-locus sequence typing (MLST)


ST 50: 2 1 12 3 2 1 5
ST 104: 2 1 1 3 7 1 5
ST 21: 2 1 1 3 2 1 5
ST 21: 2 1 1 3 2 1 5
Knott End-on-Sea
Pilling
Stake Pool
Preesall
Hambleton
Garstang
Calder Vale
Nateby
Bilsborrow
Whitewell
Elswick
M6
M55
Longridge
Goosnargh
Broughton
Grimsargh
A585
A5230
A583
Wrea Green
Warton
A59
Mellor Brook
M61
Brindle
Newtown
Mere Brow
Holmeswood
Southport
A570
A565
Scarisbrick
Bescar
Wrightington Bar
Worthington Lodge
Grimeford Village
Chorley
Horwich
A59
Liverpool Rd
Moss Ln

Genetic inhomogeneity in *Campylobacter jejuni*

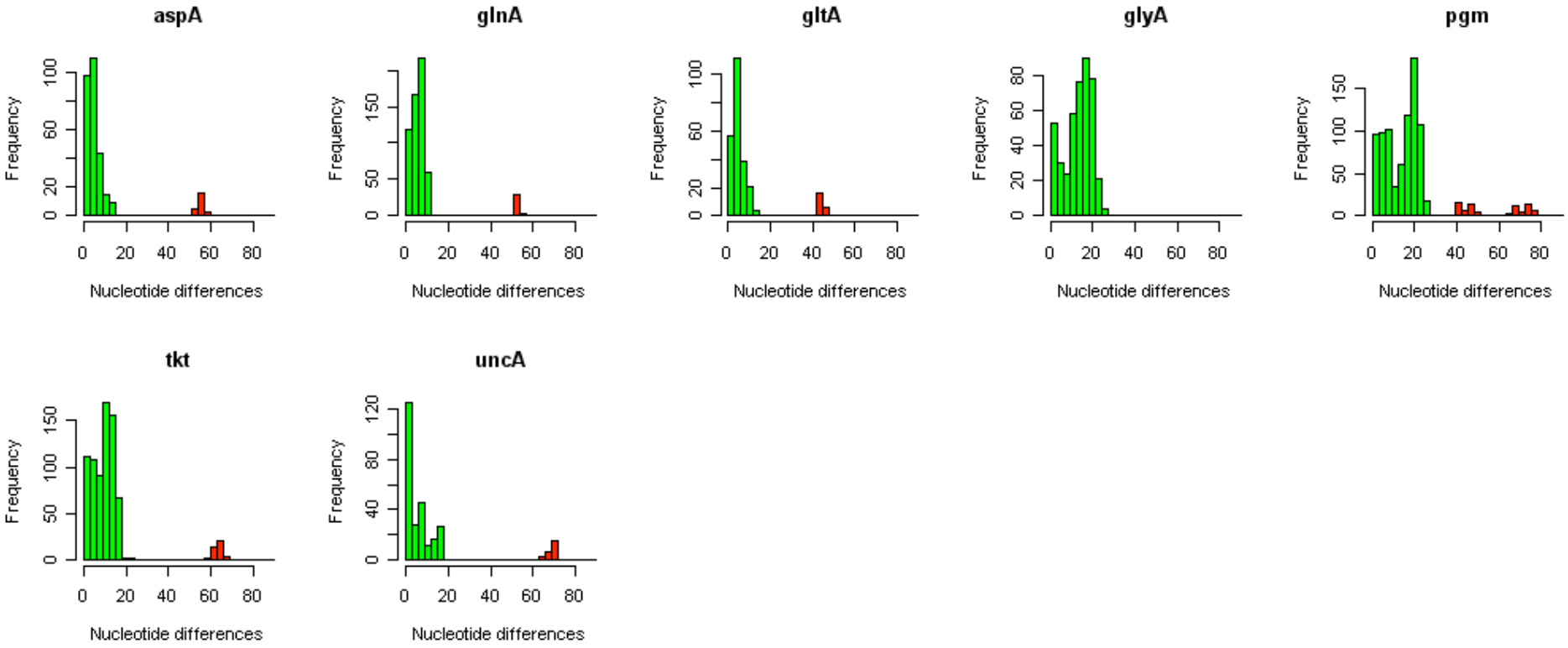
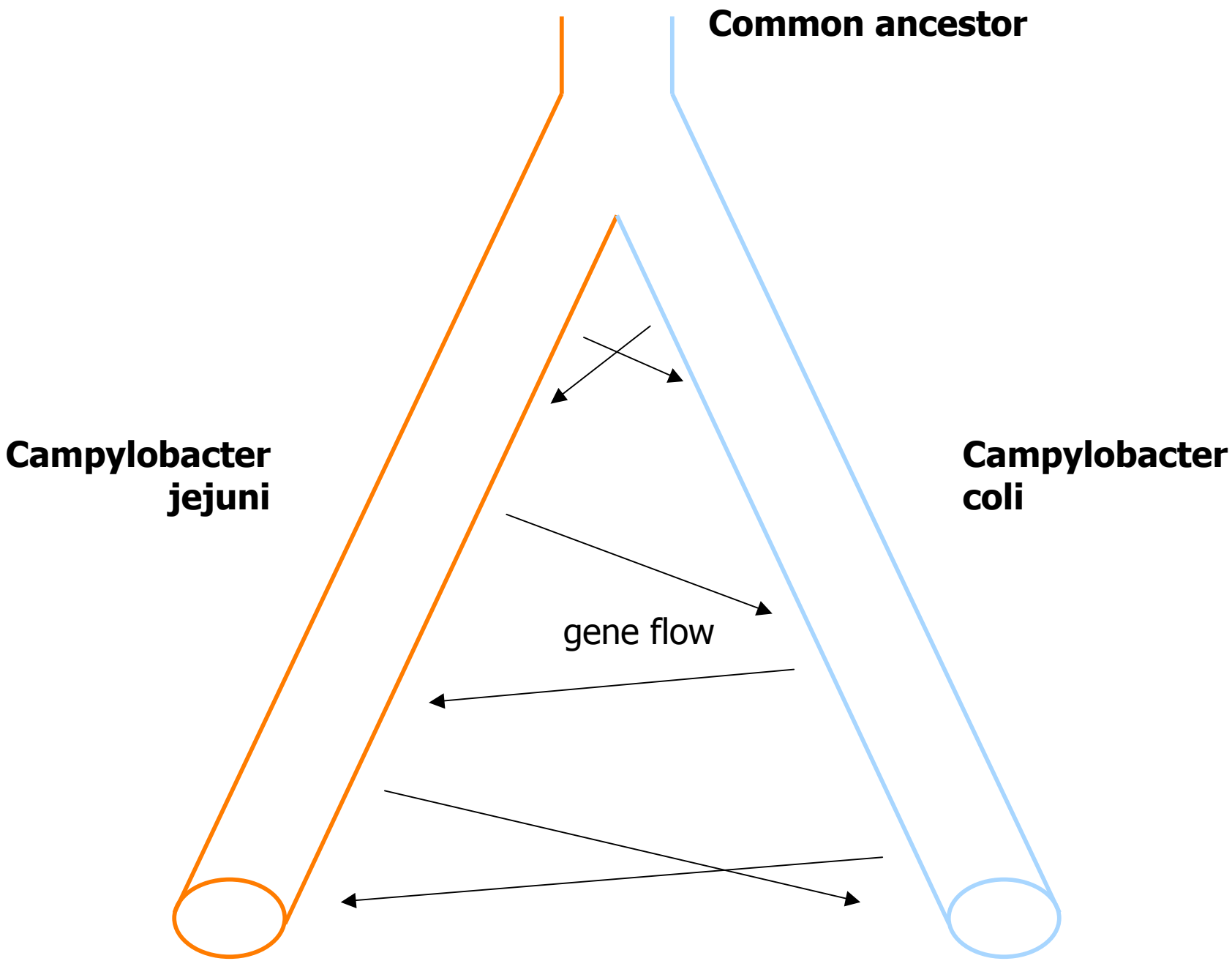


Figure 1. Number of nucleotide differences between each pairs of alleles at the seven loci. When *C. coli*-derived alleles are removed, the red portion of the histogram disappears.


Common ancestor
Campylobacter jejuni
Campylobacter coli
gene flow

Putatative *Campylobacter jejuni* isolates

ST	aspA	glnA	gltA	glyA	pgm	tkt	uncA
21	2	1	1	3	2	1	5
257	9	2	4	62	4	5	6
66	2	4	5	2	7	1	5
61	1	4	2	2	6	3	17

Putatative *Campylobacter coli* isolates

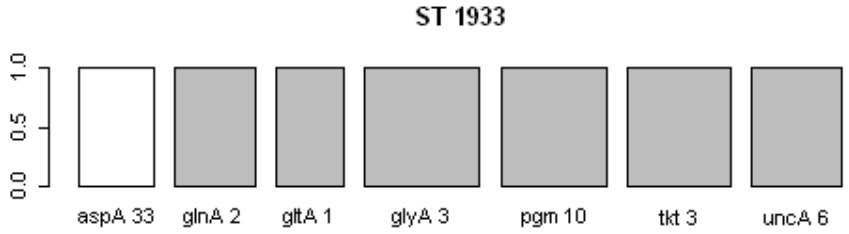
ST	aspA	glnA	gltA	glyA	pgm	tkt	uncA
825	33	39	30	82	113	47	17
826	33	39	30	114	104	35	17
832	33	39	30	79	113	43	17
868	81	104	81	113	143	119	67

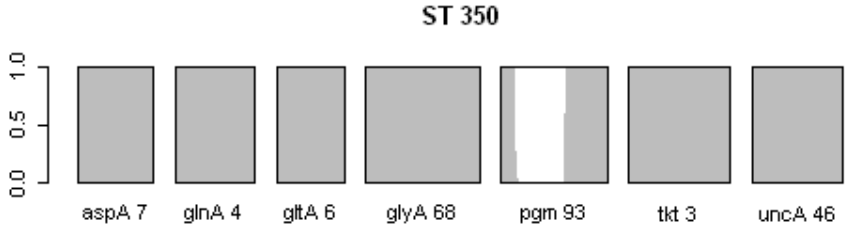
	ST	aspA	glnA	gltA	glyA	pgm	tkt	uncA
	61	1	4	2	2	6	3	17
Allele frequency								
C. jejuni		0.13	0.14	0.15	0.18	0.06	0.18	0.04
C. coli		0	0	0.01	0.01	0	0.01	0.48

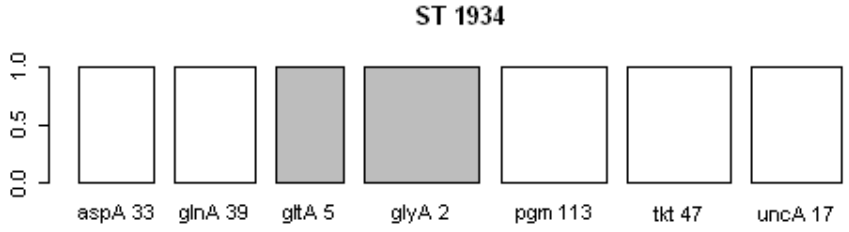
Population Structure: identifying hybrids

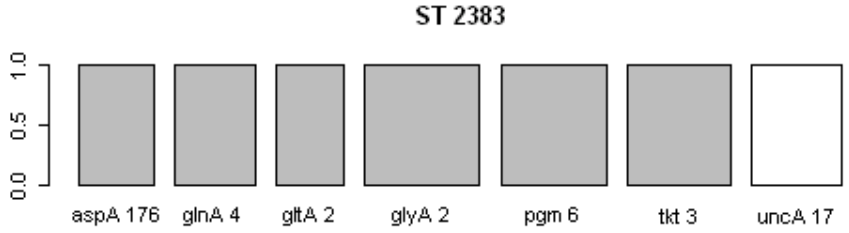
C. jejuni – C. coli hybrids

3 sequence types, 20/881 isolates (2.2%)









Population Structure: identifying hybrids

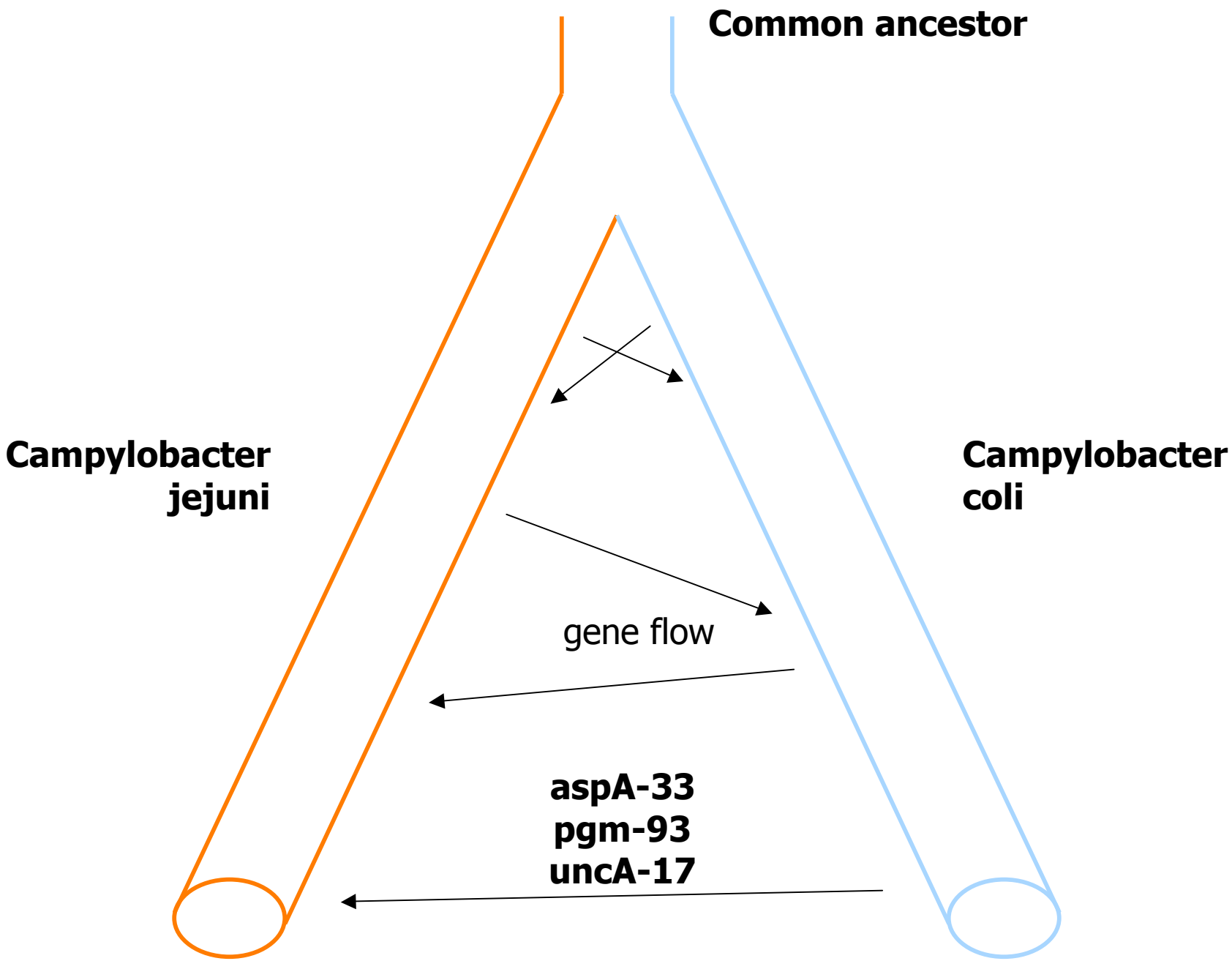
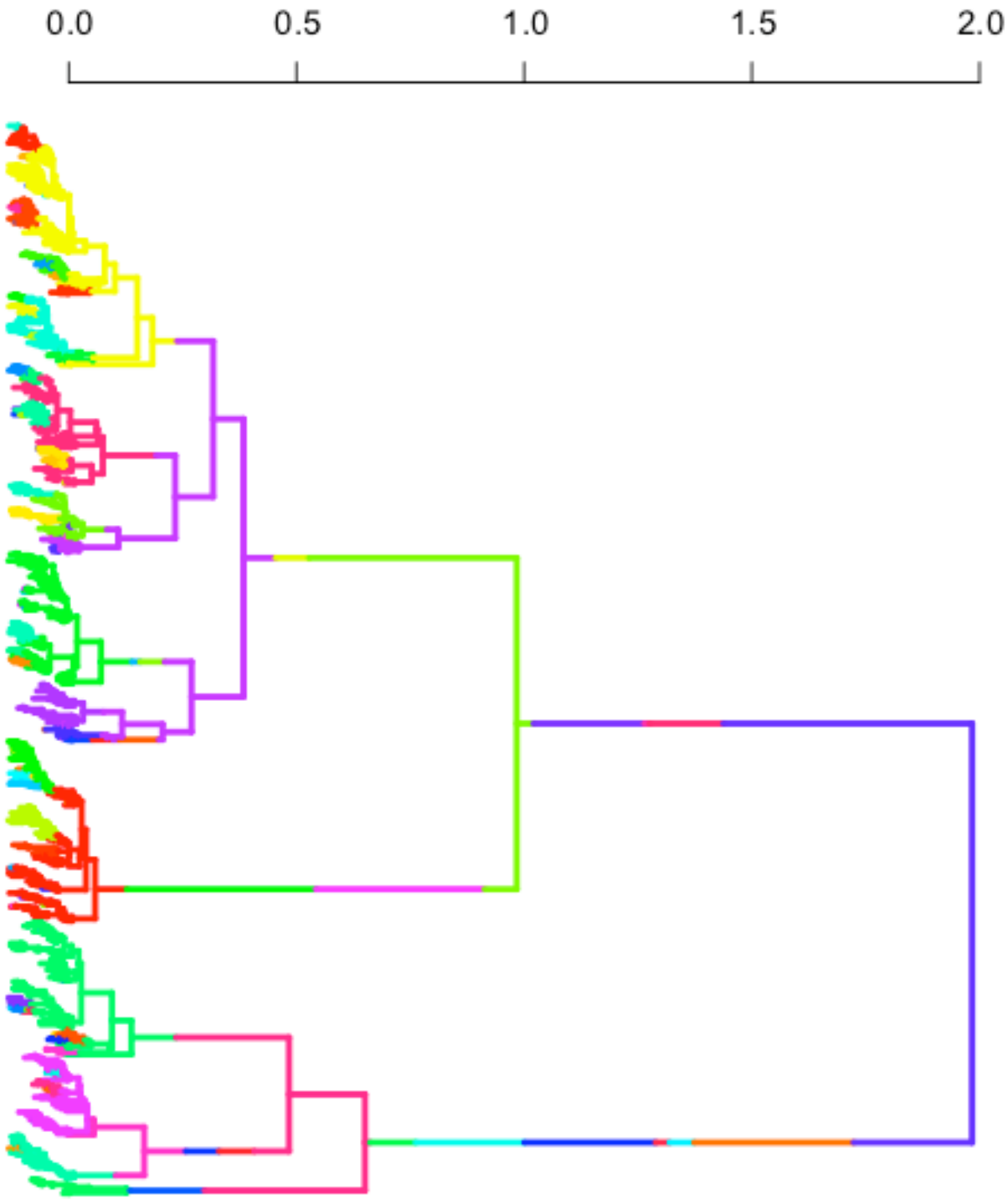

Common ancestor
Campylobacter jejuni
Campylobacter coli
gene flow
aspA-33
pgm-93
uncA-17

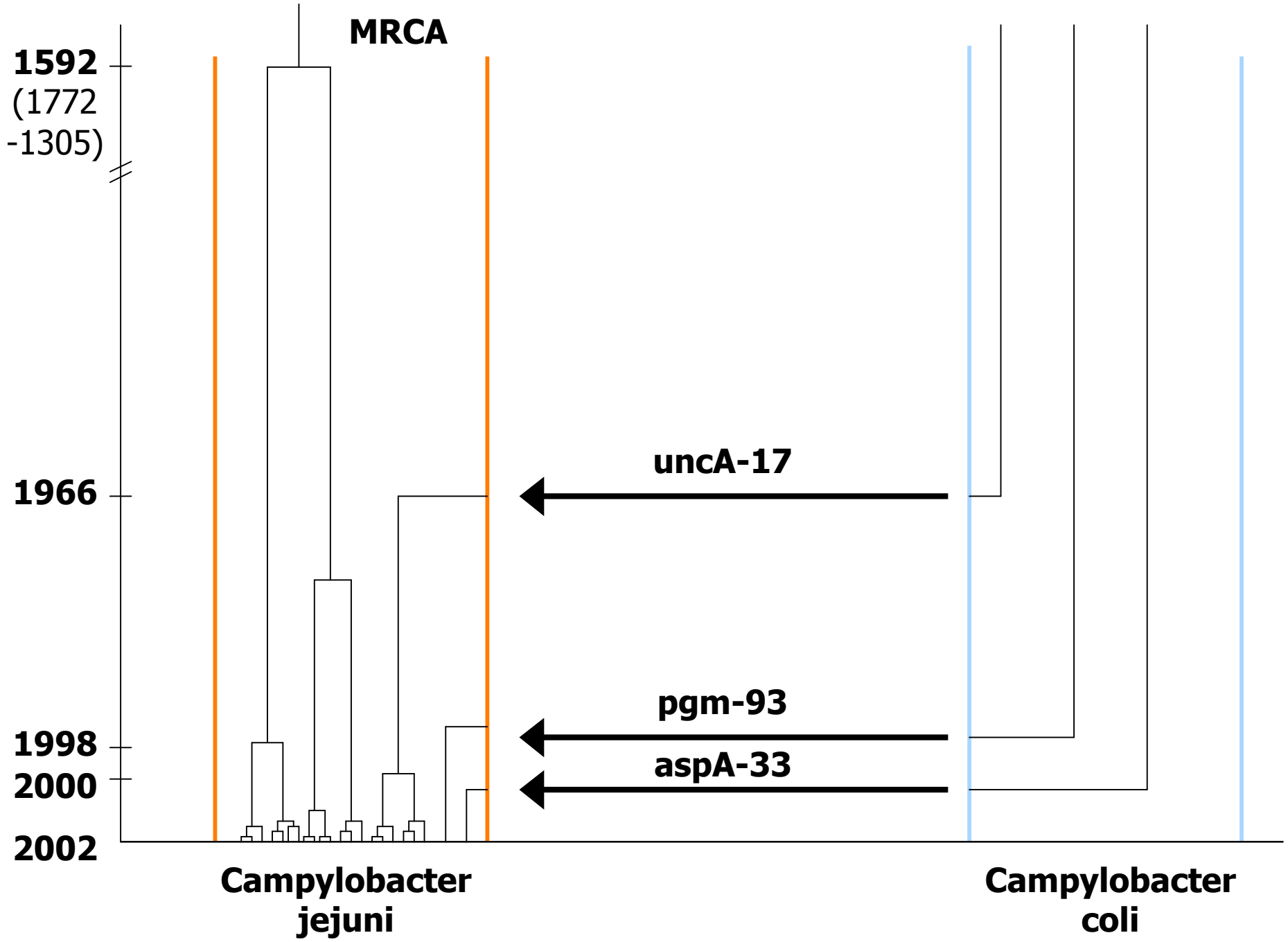
Table 1. Estimates of evolutionary parameters in *Campylobacter jejuni*.

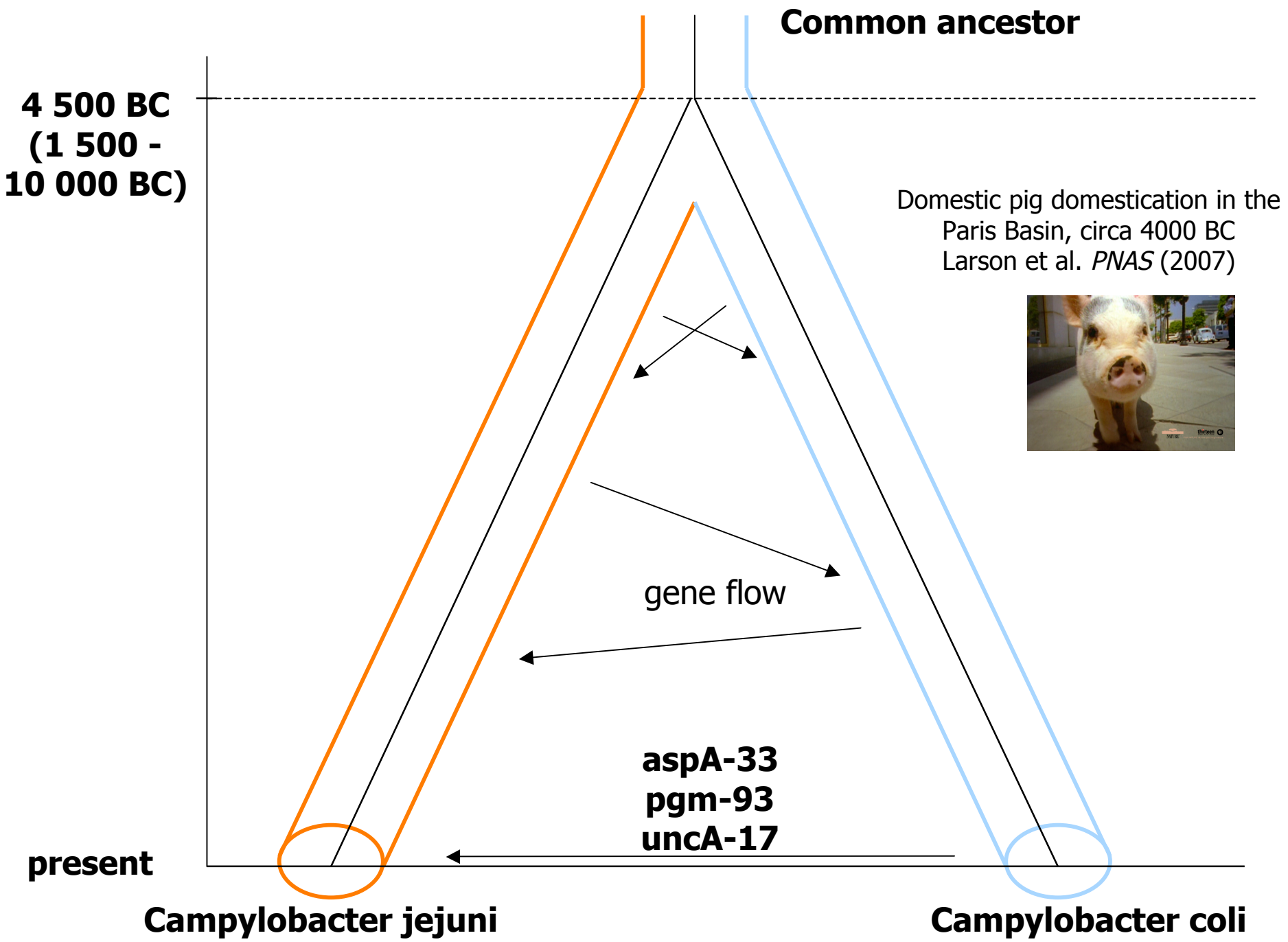
Parameter		Units	*A posteriori*		*A priori*
			Point estimate	95% C.I.	95% C.I.
	Sequence model				
θ	Total mutation rate	$kb^{-1}\ (2N_e g)^{-1}$			2.1 – 180
θ_S	Synonymous mutation rate	$kb^{-1}\ (2N_e g)^{-1}$			1.9 – 170
θ_N	Non-synonymous mutation rate	$kb^{-1}\ (2N_e g)^{-1}$			0.2 – 10
θ_0	Neutral mutation rate	$kb^{-1}\ (2N_e g)^{-1}$			5.8 – 510
κ	Transition-transversion ratio				3.3 – 180
ω	dN/dS ratio				0.0022 – 0.18
$\rho\tau$	Recombination rate between distant loci	$(2N_e g)^{-1}$			2×10^{-4} – 8×10^{5}
ρ	Recombination rate	$kb^{-1}\ (2N_e g)^{-1}$			0.0014 – 72
τ	Mean DNA import length	kb			0.015 – 6800

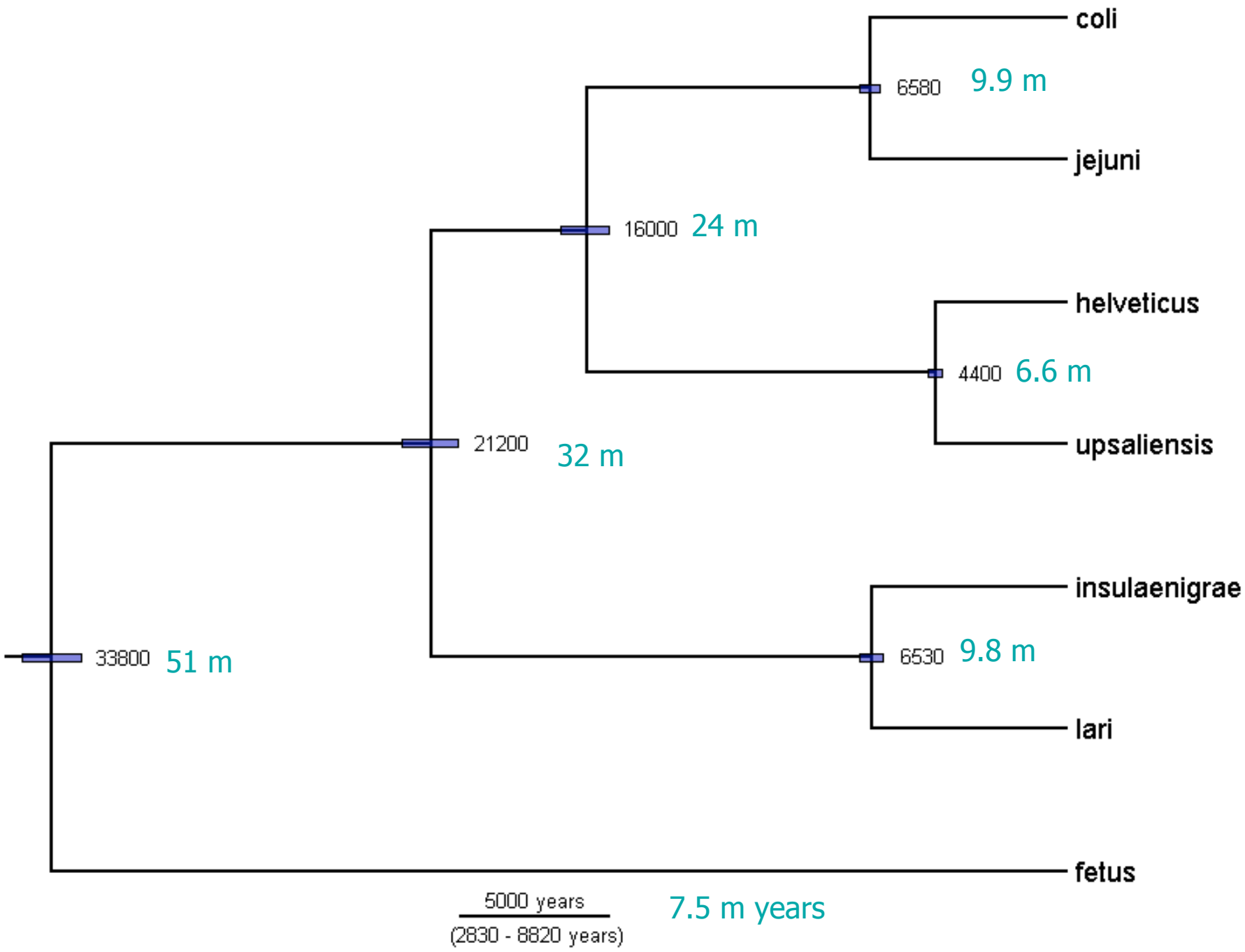
The geometric mean was used to obtain point estimates. The (2.5%, 97.5%) quantiles were used to calculate the 95% credible intervals (C.I.). All priors were uniform on the logarithmic scale.

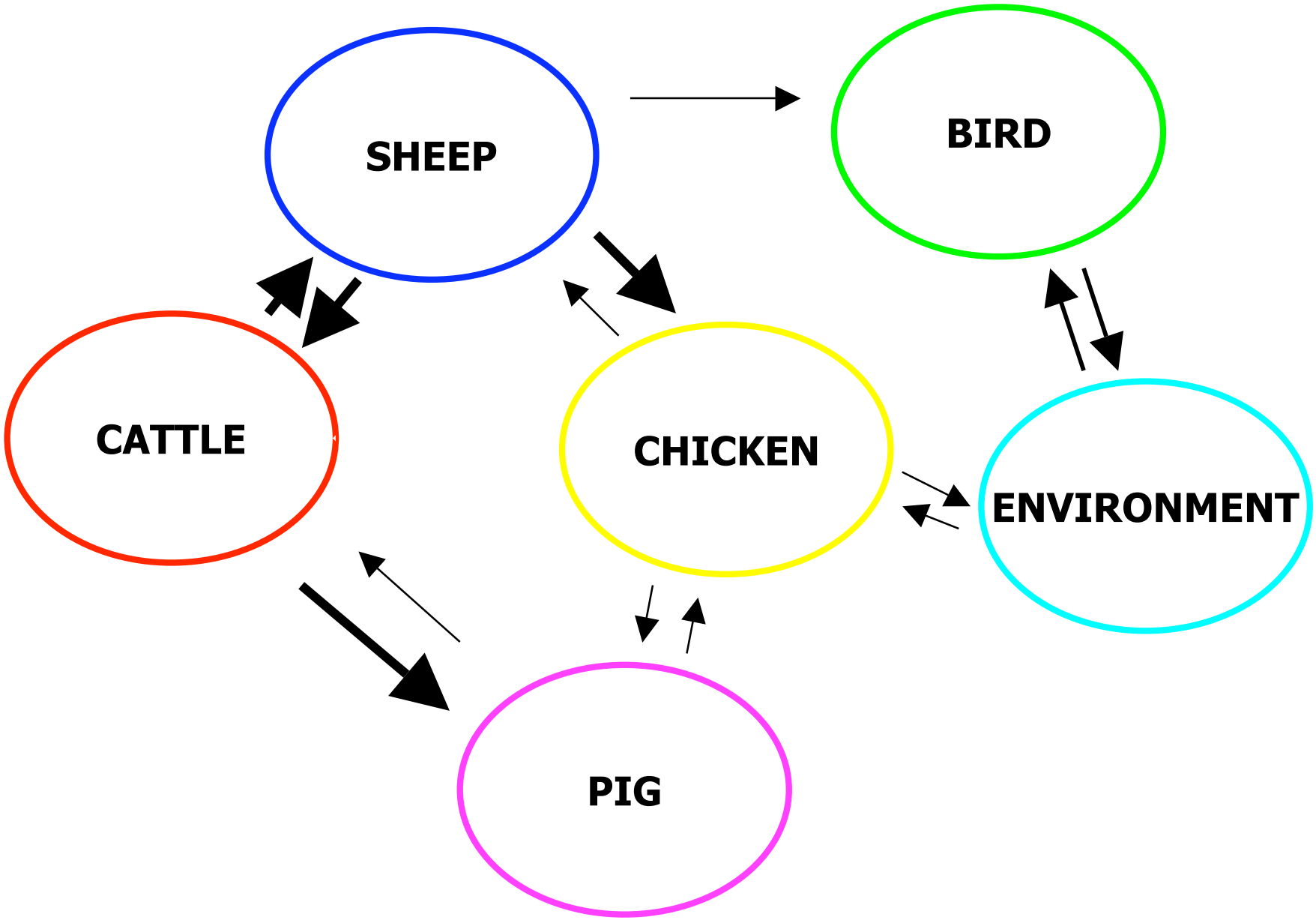
Coalescent modelling with ABC

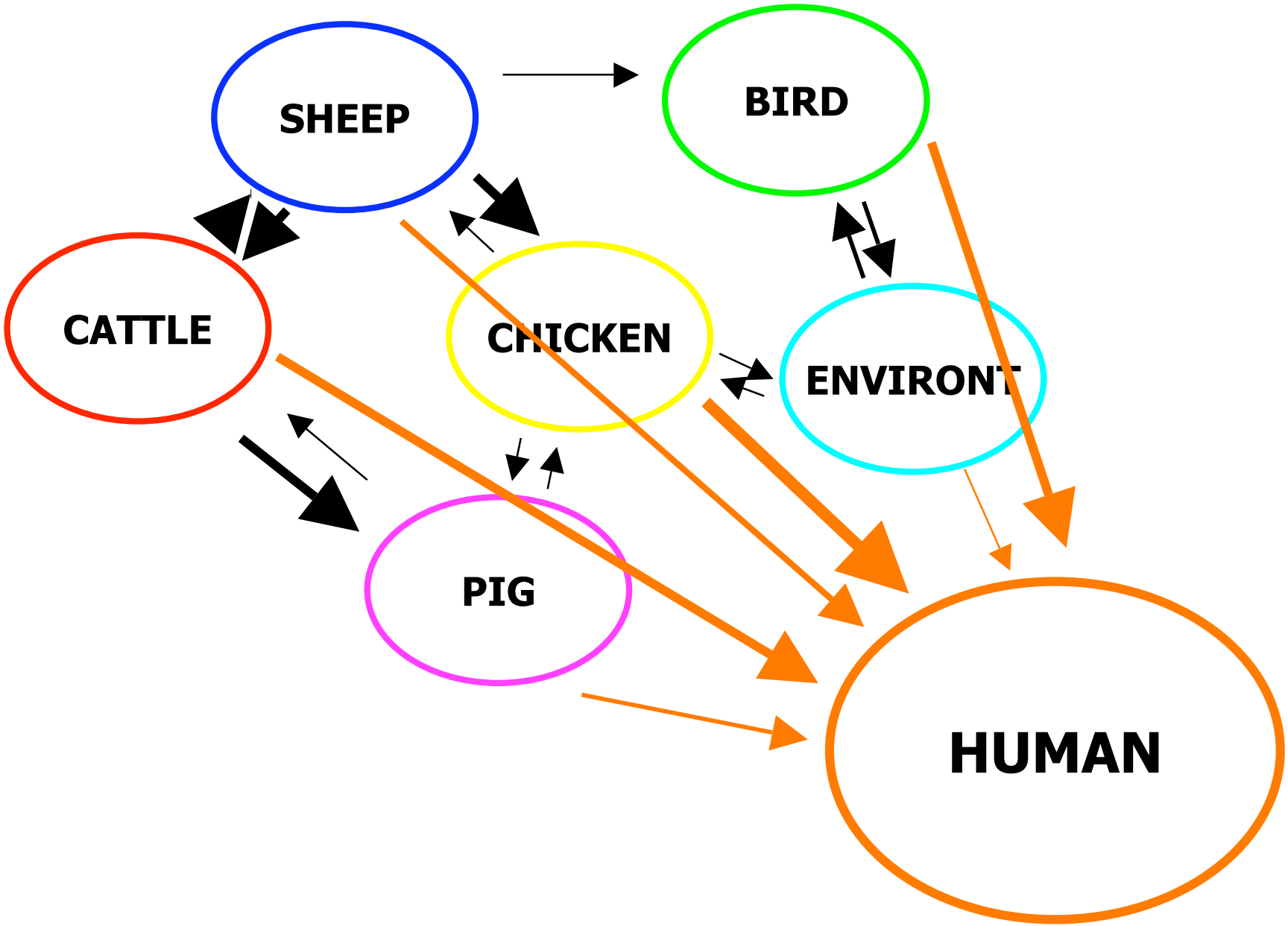

0.0
0.5
1.0
1.5
2.0


MRCA
1592
(1772
-1305)
1966
1998
2000
2002
uncA-17
pgm-93
aspA-33
Campylobacter
jejuni
Campylobacter
coli


Common ancestor
4 500 BC
(1 500 -
10 000 BC)
Domestic pig domestication in the
Paris Basin, circa 4000 BC
Larson et al. *PNAS* (2007)
gene flow
aspA-33
pgm-93
uncA-17
present
Campylobacter jejuni
Campylobacter coli


coli
6580
9.9 m
jejuni
16000
24 m
helveticus
4400
6.6 m
upsaliensis
21200
32 m
insulaenigrae
33800
51 m
6530
9.8 m
lari
fetus
5000 years
(2830 - 8820 years)
7.5 m years


SHEEP
BIRD
CATTLE
CHICKEN
ENVIRONMENT
PIG


SHEEP
BIRD
CATTLE
CHICKEN
ENVIRONT
PIG
HUMAN

Navigation

- **PubMLST**
MLST Home
Search / site map

- **Software**
Bio-Linux
Web tools
Software

- **Databases**
A. baumannii
A. fumigatus
B. cereus
Bordetella
B. burgdorferi
B. cepacia
C. fetus
C. helveticus
C. insulaenigrae
C. jejuni & C. coli
C. krusei
C. lari
C. upsaliensis
C. tropicalis
H. pylori
K. pneumoniae
L. monocytogenes
Neisseria
P. aeruginosa
P. gingivalis
S. agalactiae
S. uberis
Streptomyces
V. vulnificus
Wolbachia

+ **Mirrors**
+ **Developers**

Campylobacter jejuni and *Campylobacter coli* MLST Home Page

The Campylobacter MLST database has undergone re-organisation to split allelic profiles from isolate data. The original MLST database has become PubMLST and a new profiles database has been created. Further details about the database structure can be found here.

- Information
- Access main databases
  - Allelic Profile/ST Database
  - PubMLST Isolate Database
- Policy document
- Submission of data
- Submission history
- News and updates

- mlstdbNet software
- Other software
- Related links
- Recent publications using MLST in Campylobacter research

2006-11-01: Database submissions - We are now using an automated system for submitting new data to the database curators. Instructions | Log in

The use of this database is subject to the terms of the policy document and it should be acknowledged in all publications that make use of it. The preferred format for the acknowledgement can be found in the right-hand sidebar.

Website and database managed by Keith Jolley, curated by Alison Cody and Frances Colles.

The primary *Campylobacter jejuni* MLST website is hosted at The Peter Medawar Building for Pathogen Research, University of Oxford, UK. Initial development funded by the Wellcome Trust.

Citing the database

The preferred format for citing this website in publications is:

This publication made use of the *Campylobacter jejuni* Multi Locus Sequence Typing website (http://pubmlst.org/campylobacter/) developed by Keith Jolley and Man-Suen Chan and sited at the University of Oxford (Jolley et al. 2004, *BMC Bioinformatics*, **5**:86). The development of this site has been funded by the Wellcome Trust.

Related databases

C. fetus
C. helveticus
C. insulaenigrae
C. lari
C. upsaliensis

Status

Profile database
Profiles: 2914
Last updated: 2007-06-19

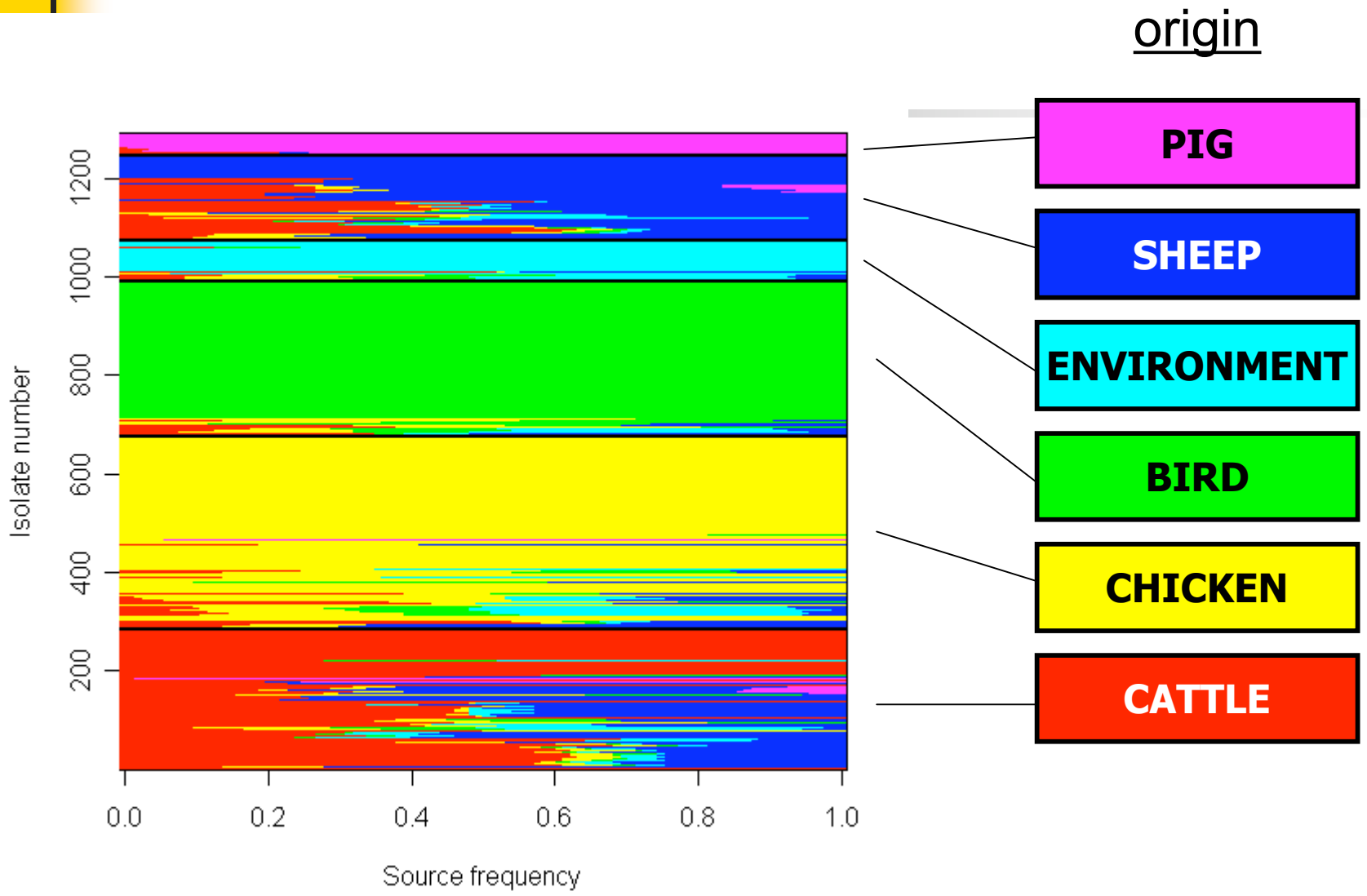
Isolate database
Isolates: 4453
Last updated: 2007-06-19

Devon's Racing Rams (photo: Rick Turner)

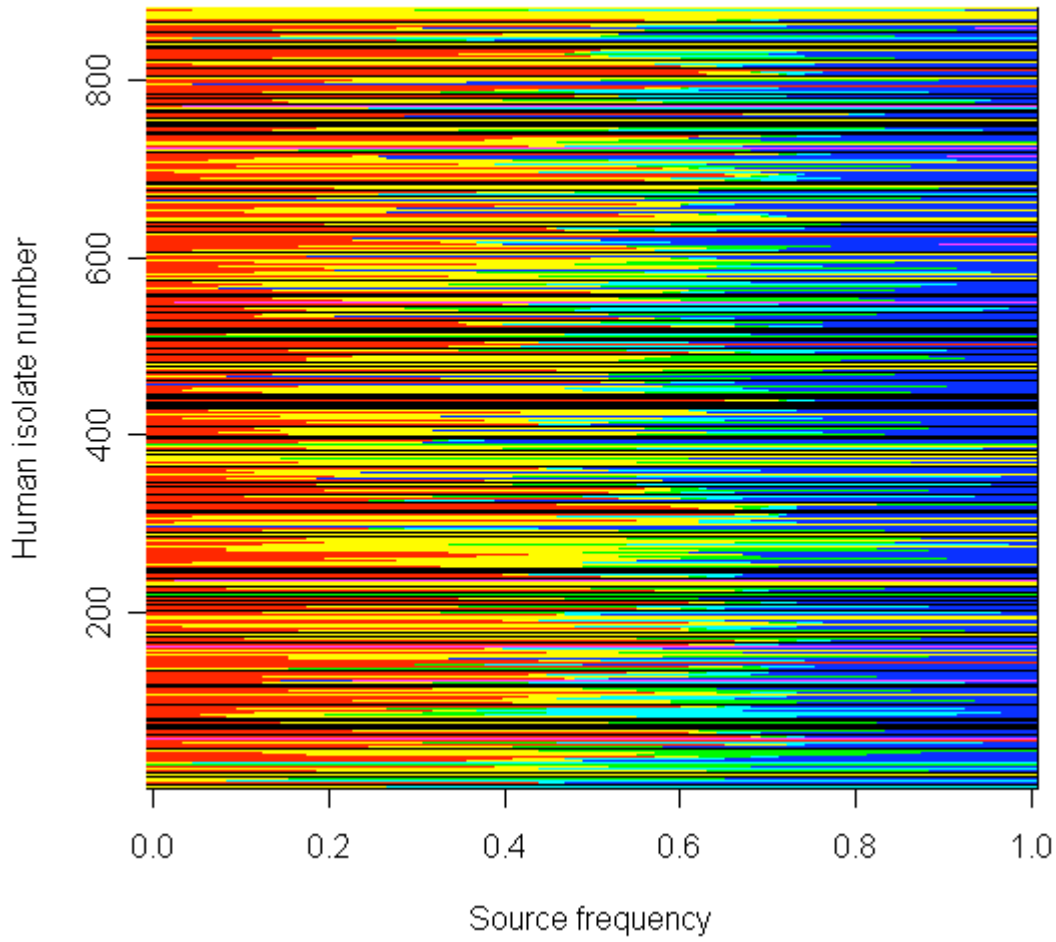
CATTLE	cattle	212	
	beef offal or meat	47	
	calf	12	
	cows milk	11	
	calf faeces	1	
	cattle faeces	1	284
CHICKEN	chicken	222	
	chicken offal or meat	153	
	chick	17	392
BIRD	wild bird	172	
	starling	71	
	goose faeces	25	
	turkey	22	
	goose	12	
	duck	7	
	starling faeces	4	313
ENVIRONMENT	sand (bathing beach)	52	
	environmental waters	28	
	soil	3	
	potable/drinking water	2	85
SHEEP	sheep	84	
	lamb offal or meat	74	
	lamb	10	
	sheep faeces	2	170
PIG	pig	35	
	pork offal or meat	10	
	piglet	1	46

oiler environment	17
	3
	5
y	10
	1
gira	1
goat	5
horse	1
human bloo	57
human stool	684
human unspecifie	
marmoset	
ostrich	
other animal	
rabbit	3
nspecified	144

Haplotype structure in sequences of known origin from pubMLST



Haplotype structure in human isolates



key

NOVEL

PIG

SHEEP

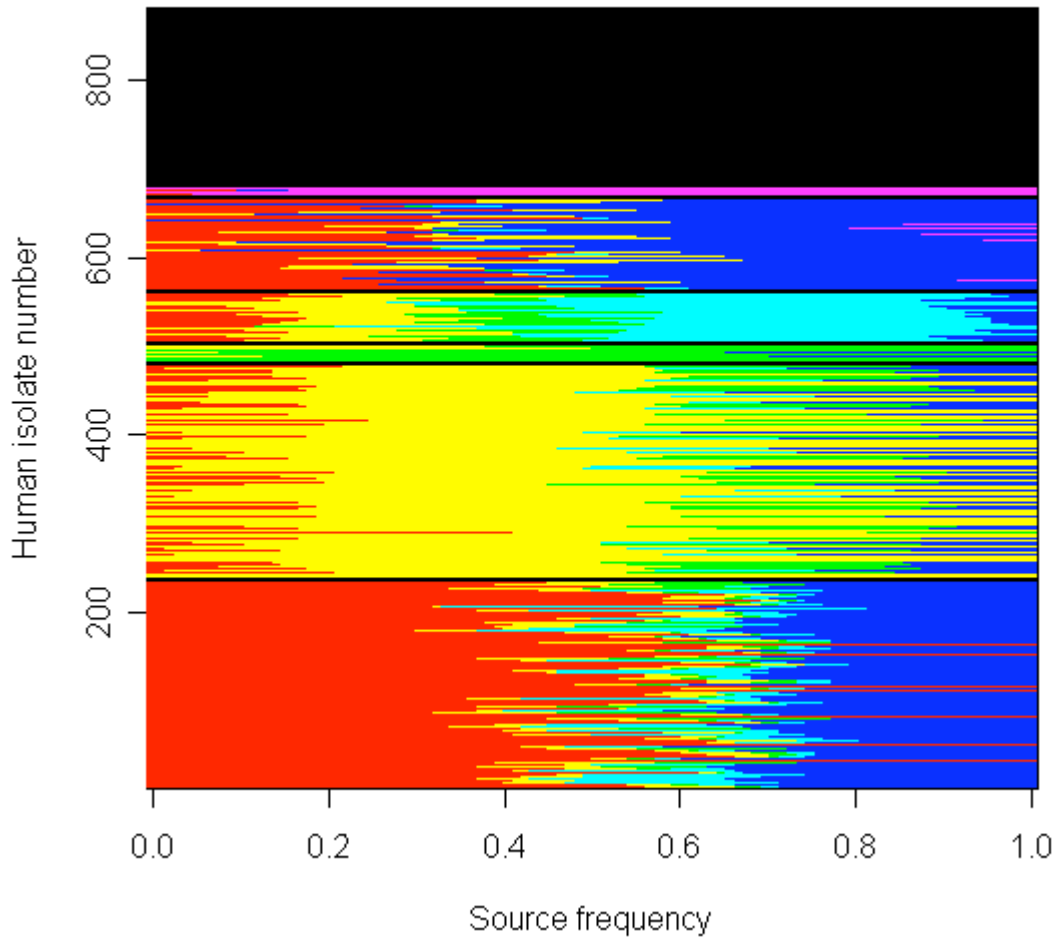
ENVIRONMENT

BIRD

CHICKEN

CATTLE

Haplotype structure in human isolates

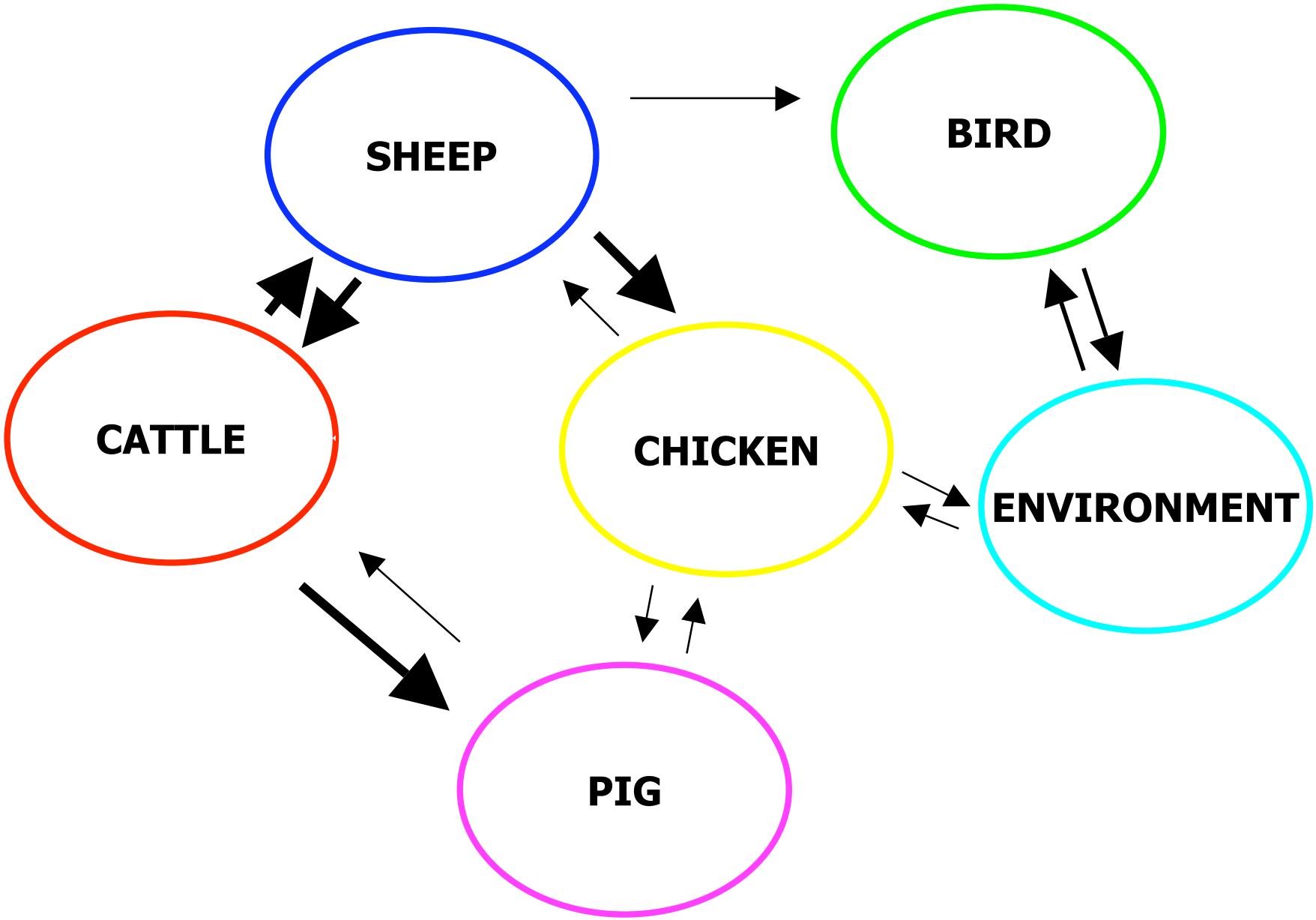


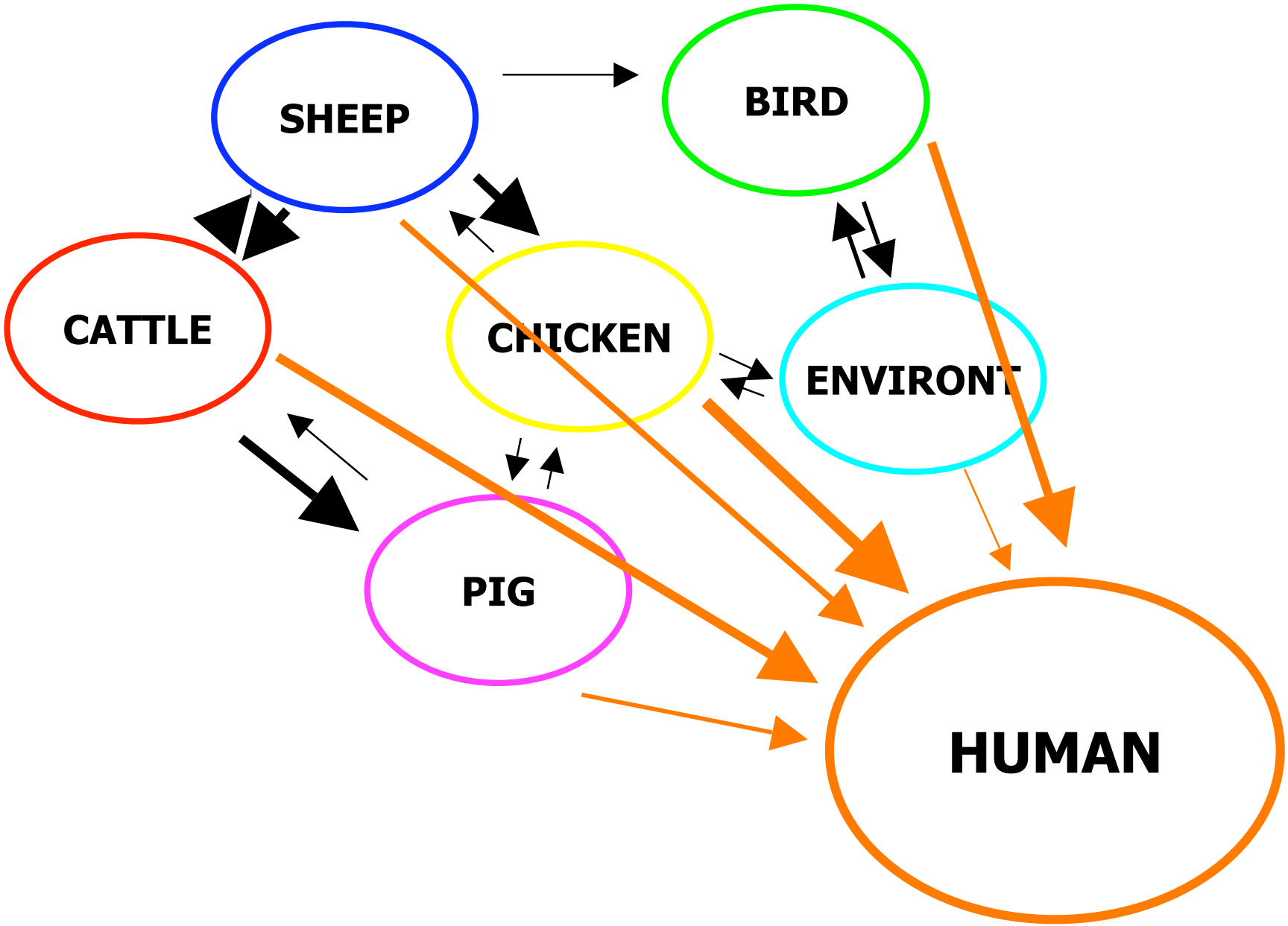
key

- NOVEL
- PIG
- SHEEP
- ENVIRONMENT
- BIRD
- CHICKEN
- CATTLE

Attributing novel genotypes

- **ST 574:** 7 53 2 10 11 3 3
- Human-specific, but similar to...
  - **ST 305:** 9 53 2 10 11 3 3
  - **ST 713:** 12 53 2 10 11 3 3
  - **ST 728:** 4 53 2 10 10 3 3
  - **ST 2585:** 7 2 3 10 11 3 3
- All these found in chicken, so the likely source is chicken


SHEEP
BIRD
CATTLE
CHICKEN
ENVIRONMENT
PIG


SHEEP
BIRD
CATTLE
CHICKEN
ENVIRONT
PIG
HUMAN

Model C: unlinked

$$l(G_i = k) = \prod_{l=1}^{7} \begin{cases} \mu_k & \text{if } H_i^{(l)} \text{ is unique} \\ (1-\mu_k) B^{(l)}_{H_i^{(l)} k} & \text{otherwise} \end{cases}$$

where μ_k is the population-specific probability of mutation, $B^{(l)}_{jk} = \sum_{m=1}^{n_g} M_{km} f^{(l)}_{jm}$ and M_{km} is the probability that a gene sampled from population k is a migrant from population m. $B^{(l)}_{jk}$ is like a weighted gene frequency that accounts for migration.

Model D: evolutionary model with linkage disequilibrium

Assuming that linked loci are unlinked can cause spurious results. Suppose that the ith human isolate is a copy of one of the animal/environmental isolates, except for differences caused by mutation and recombination events. Mutation generates novel alleles. Recombination generates (potentially) novel combinations of alleles.

$$l(G_i = k) = \sum_{c=1}^{N} \frac{M_{kX_c}}{N_{X_c}} \prod_{l=1}^{7} \begin{cases} \mu_k & \text{if } H_i^{(l)} \text{ is unique} \\ (1-\mu_k) R_k B^{(l)}_{H_i^{(l)} k} & \text{else if } H_i^{(l)} \neq Y_c^{(l)} \\ (1-\mu_k) R_k B^{(l)}_{H_i^{(l)} k} + (1-\mu_k)(1-R_k) & \text{else if } H_i^{(l)} = Y_c^{(l)} \end{cases}$$

Here the "template" isolate is c, which has known source population X_c and allelic profile $\mathbf{Y}_c = \left[Y_c^{(1)}, Y_c^{(2)}, \ldots, Y_c^{(7)}\right]$. N is the total number of isolates of known origin, N_k the number in population k, and R_k is the population-specific probability of recombination.

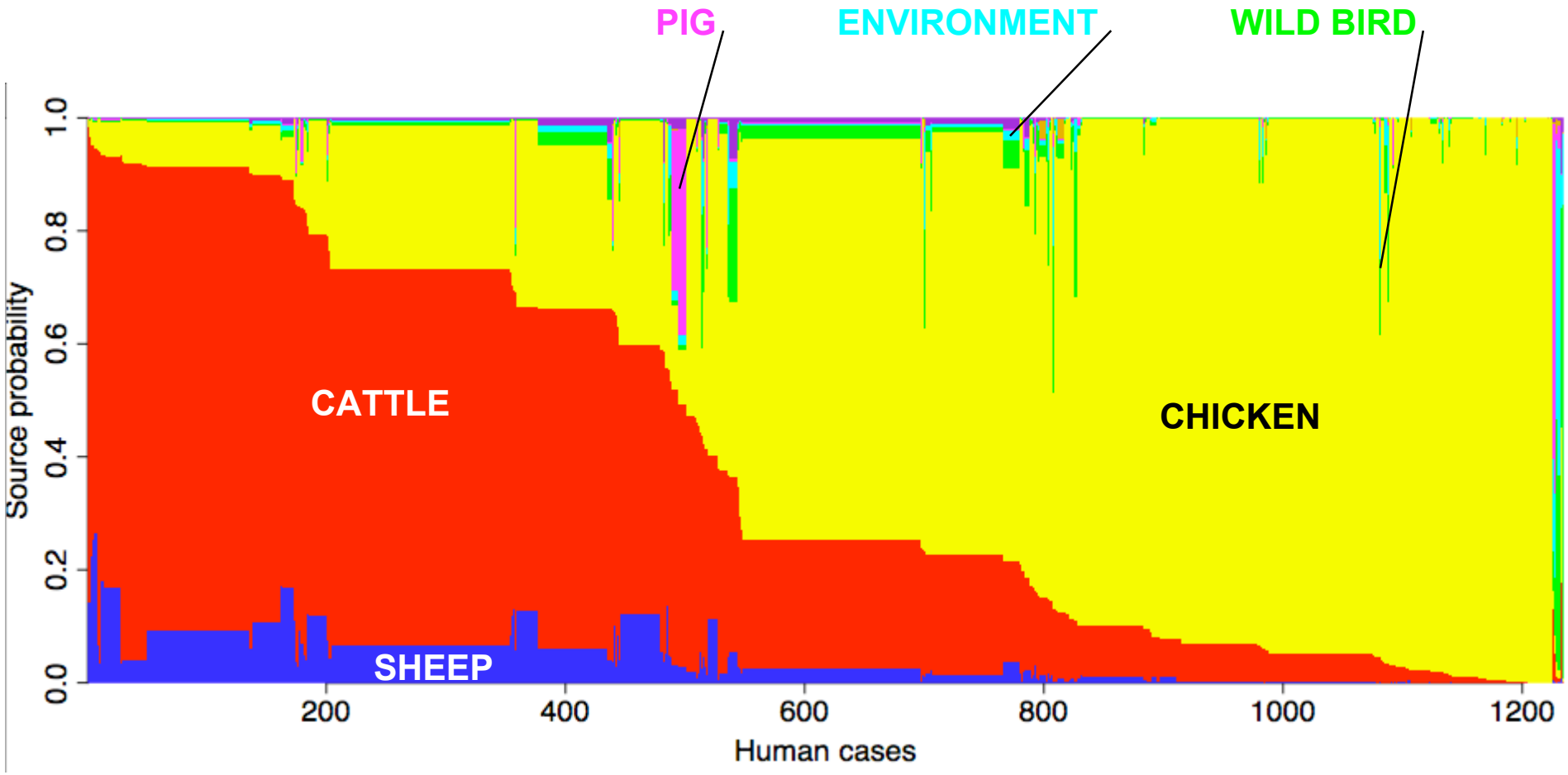
Does it work? Empirical cross-validation

- Split sequences of known origin into two groups. Treat one group as having unknown origin (pseudo-human cases)
- Infer the proportion of pseudo-human cases drawn from each source population
- Repeat 100 times to study the performance of the method

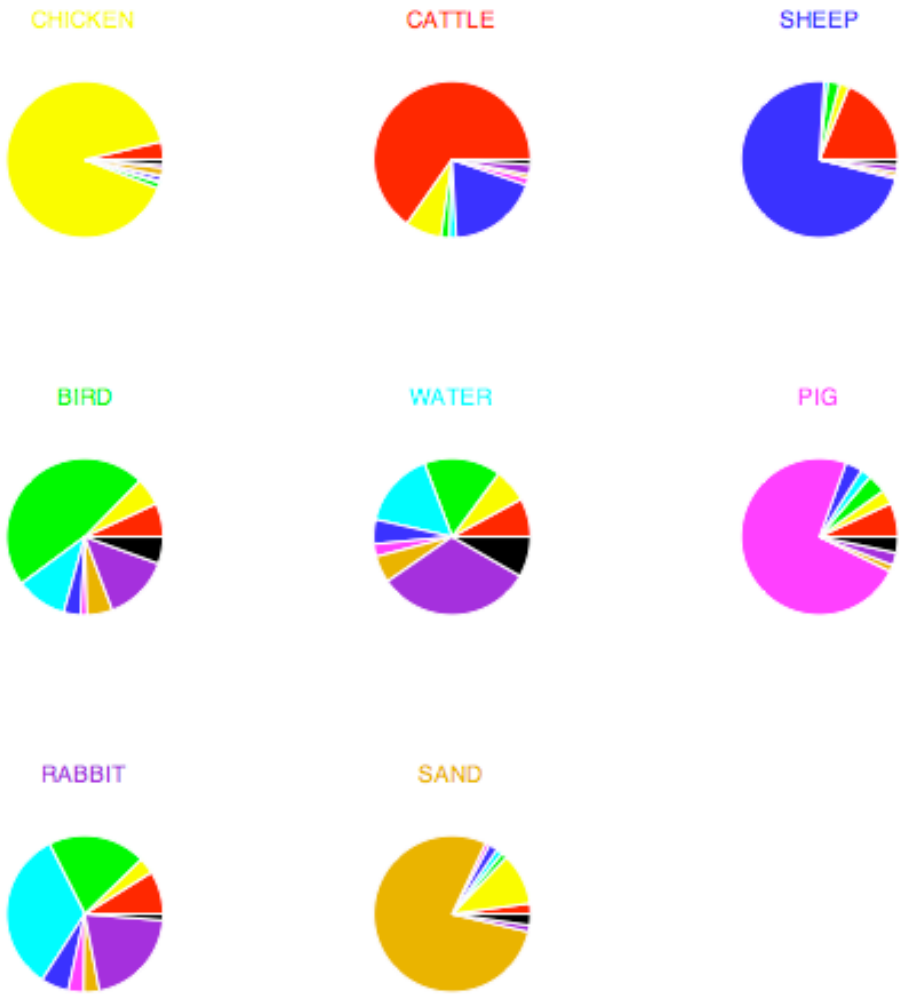
Simulation and empirical cross-validation Results: Linked model

		Unlinked model
Predicted Correct		0.86
Actual Correct		0.56
Coverage	CATTLE	22
	CHICKEN	85
	BIRD	81
	ENVIRONMENT	63
	SHEEP	16
	PIG	89

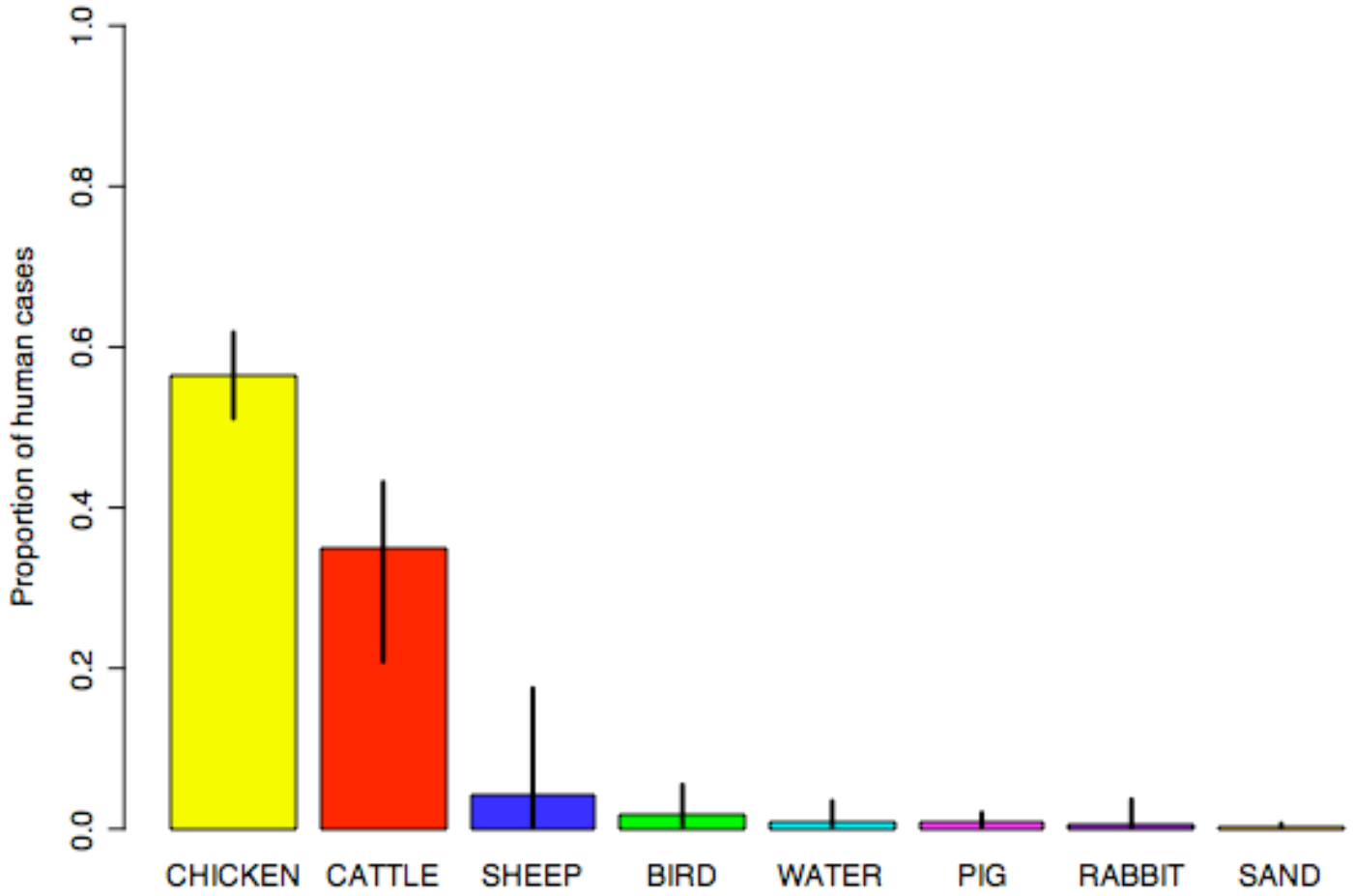
Case-by-case: probability of source



Gene flow between source populations



Tracing the source of infection: results



Tracing the host species of human food-borne infections

- Infected meat principal source (97%)
- No evidence for environment or wild birds as a major transmission route
- Prevention strategies:
  - Enhanced on-farm biosecurity
  - Interrupting transmission chain

Devon's Racing Rams (photo: Rick Turner)