Phylodynamics of measurably evolving pathogens
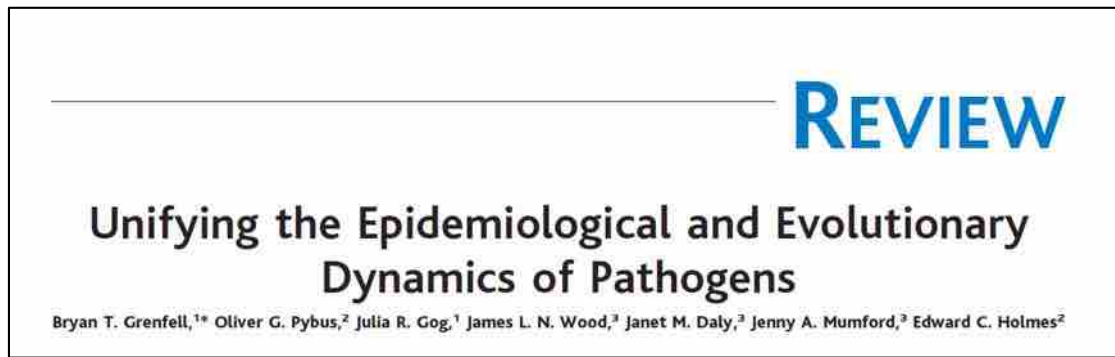
Roman Biek
(Roman.Biek@glasgow.ac.uk)

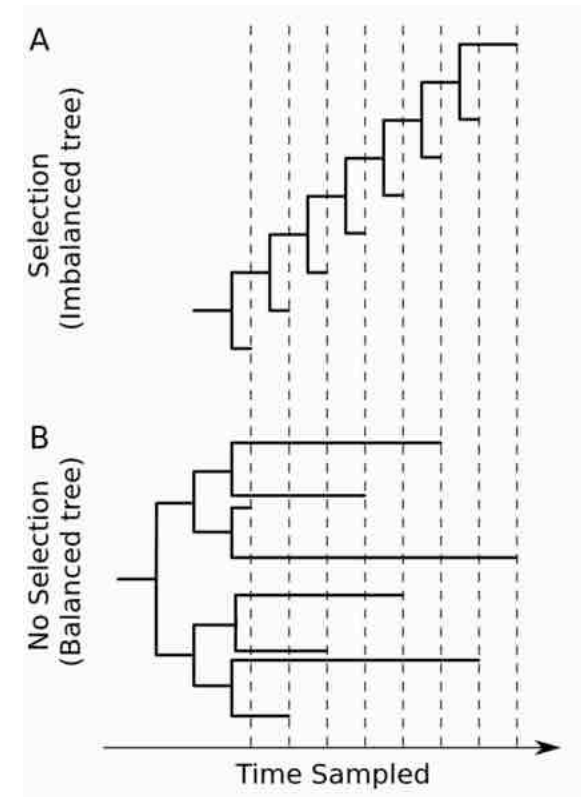Institute of Biodiversity, Animal Health &
Comparative Medicine

Genomics of Disease in Wildlife

University of Glasgow

VIA VERITAS VITA

From trees to *phylodynamics*

Grenfell et al. 2004, Science

What can <u>phylogenetic patterns</u> reveal about the underlying <u>population processes</u>?

Volz et al. 2012, PLoS Comp Biol

Phylodynamics

Traditionally, strongly linked to rapidly evolving pathogens such as RNA viruses: **transmission** and **mutational events** on similar time scales

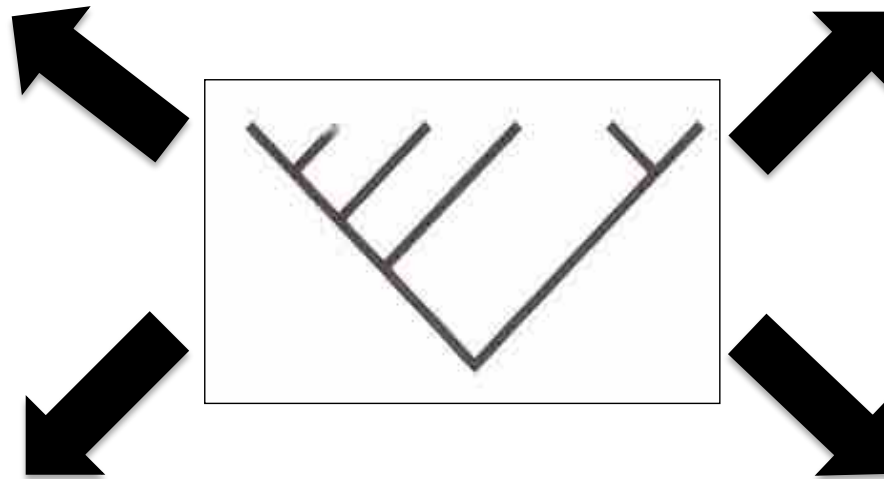Increasingly changing in the genomic era, as more organisms become "measurably evolving"

Phylogenies at the centre of phylodynamic questions

Molecular clocks
e.g. "*How long ago since two groups split?*"

Selection
e.g. "*Which genes or sites have undergone adaptive change*?"
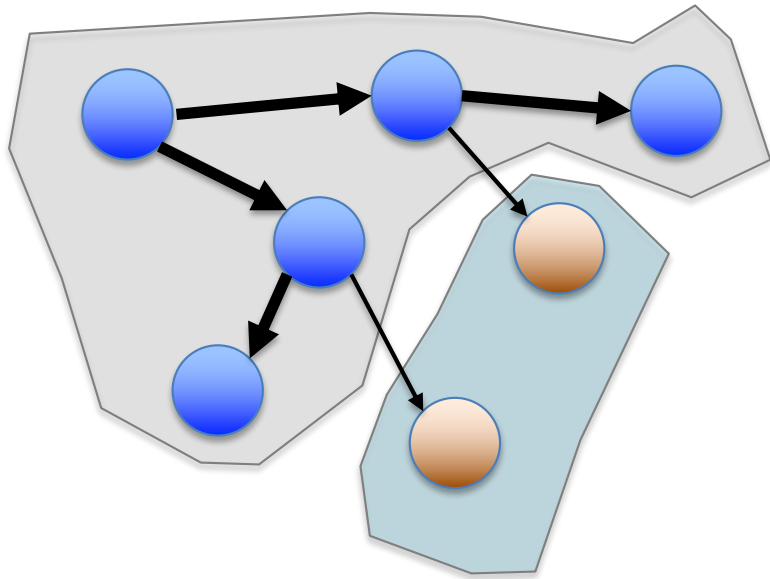
Ancestral state change
e.g. "*Movement rate between population A and B?*"

Demographic reconstruction
e.g. "*How has population size changed through time?*"
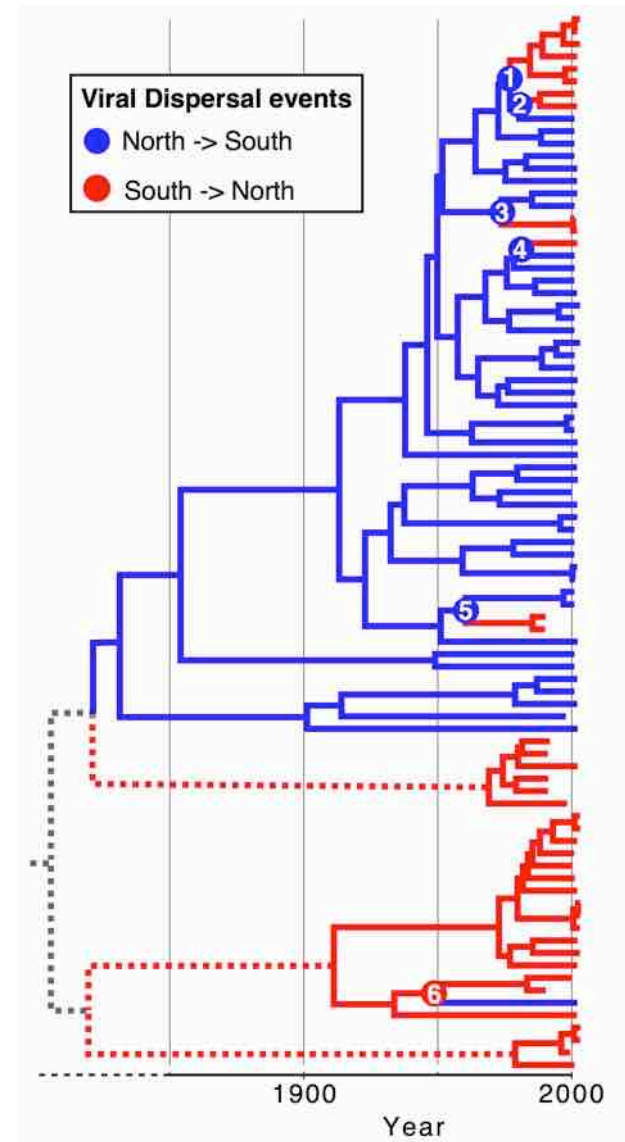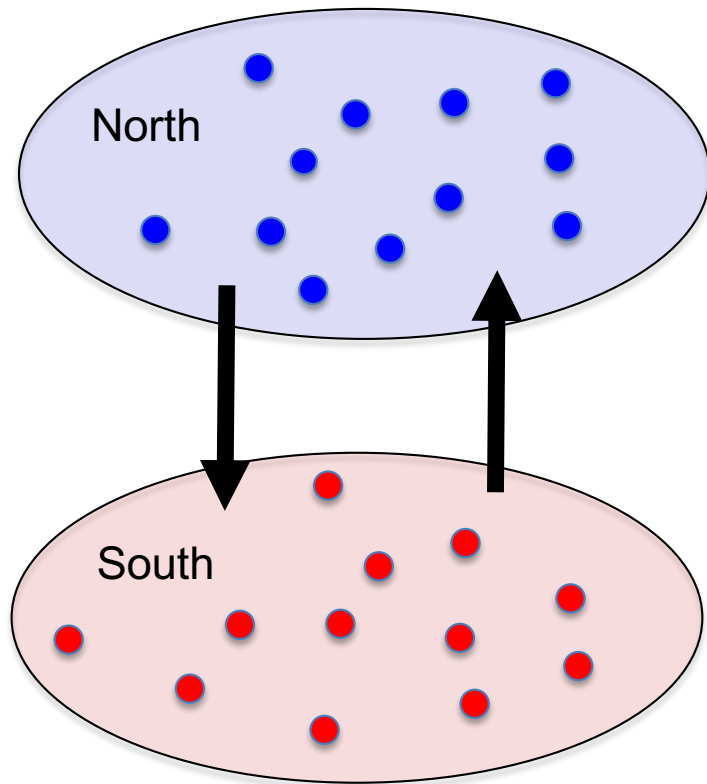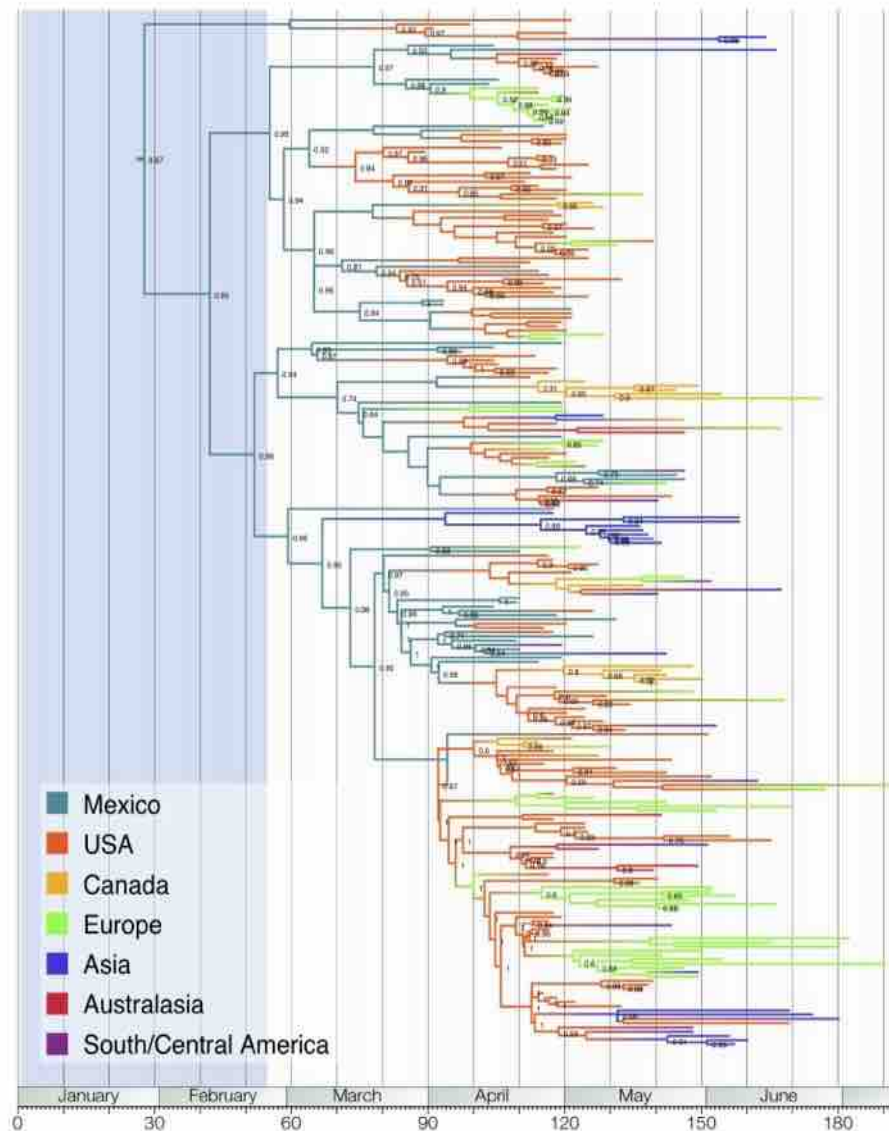
Structured host populations

Spatial
structure

Multiple
hosts

Estimating movement in discrete space: reconstructing ancestral states

Dispersal between two host populations

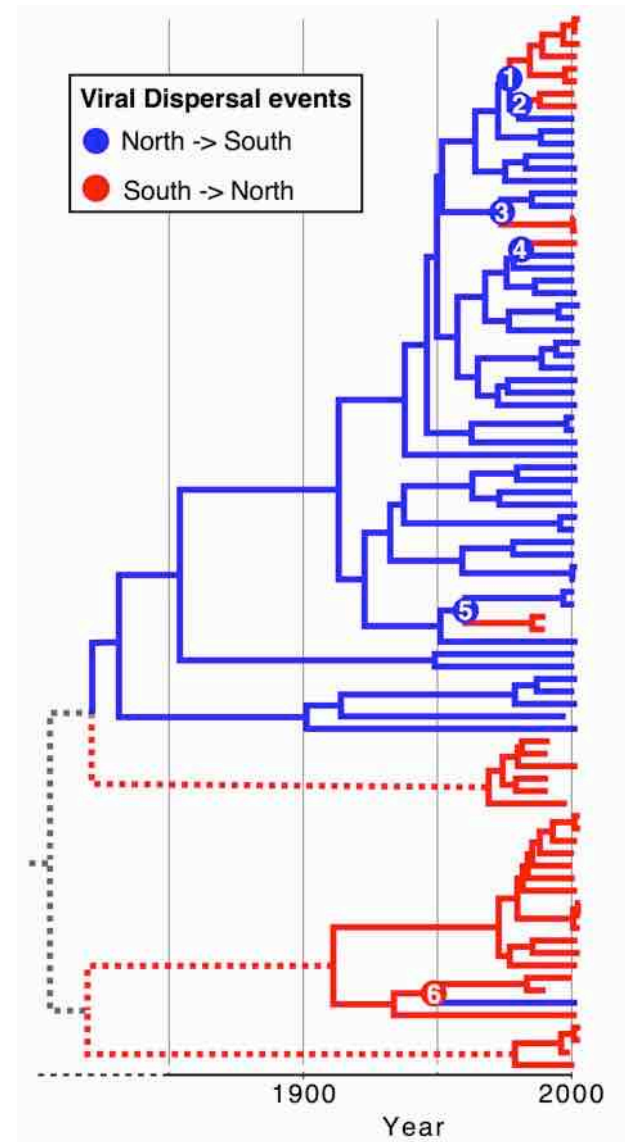# Global spread of H1N1 as inferred from phylogenetic data
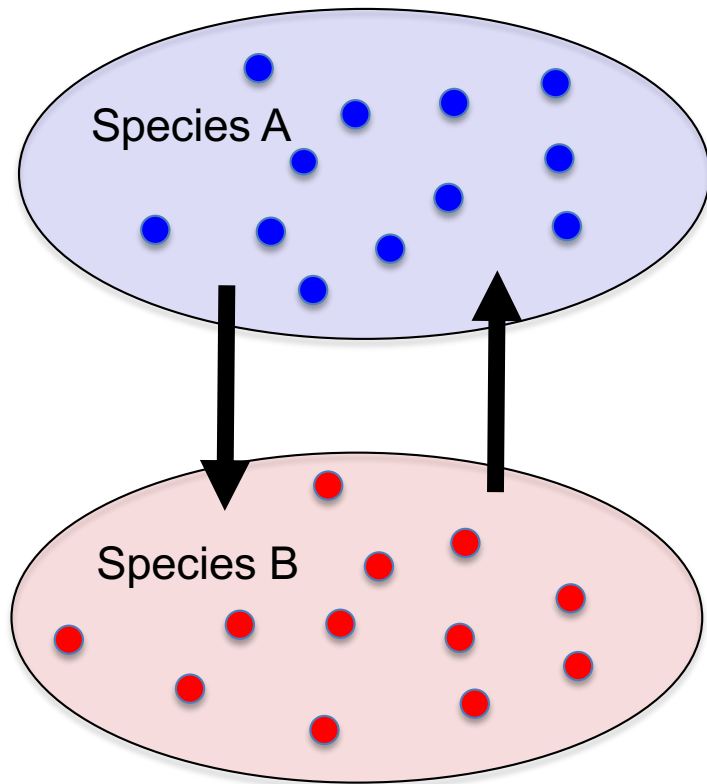

Lemey et al 2009,
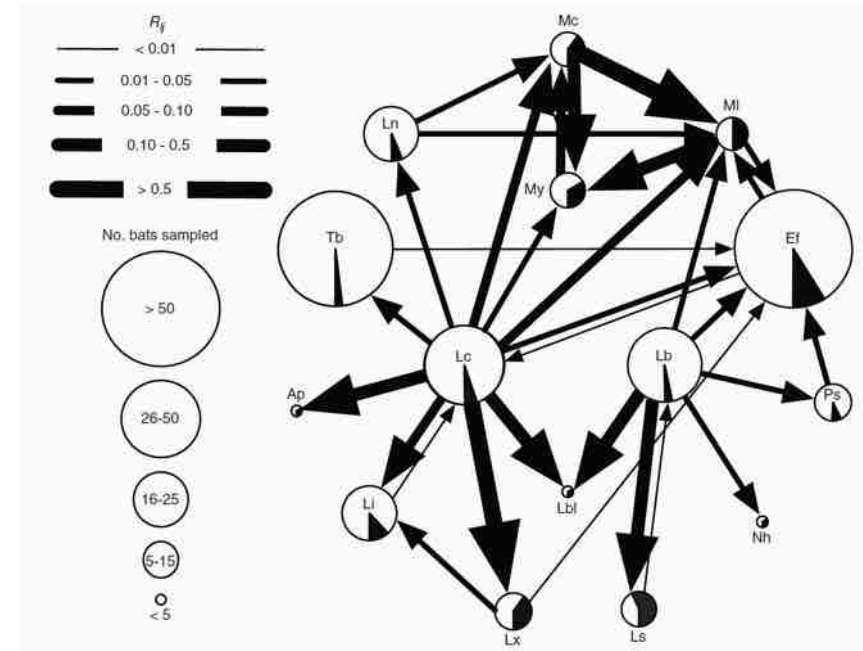*PLoS Current Influenza*
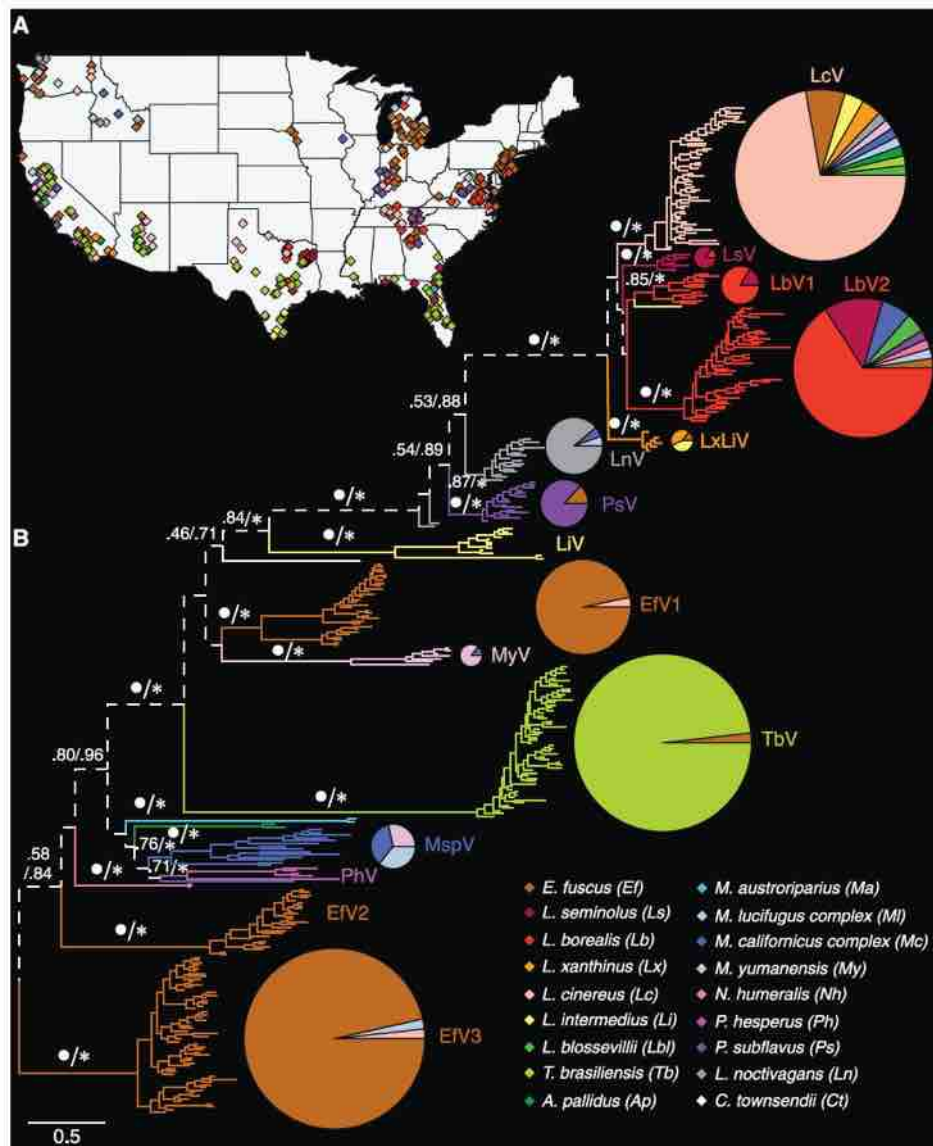
Global spread of H1N1

Discrete states don't need to be spatial

Movement between two
host ~~populations~~ species
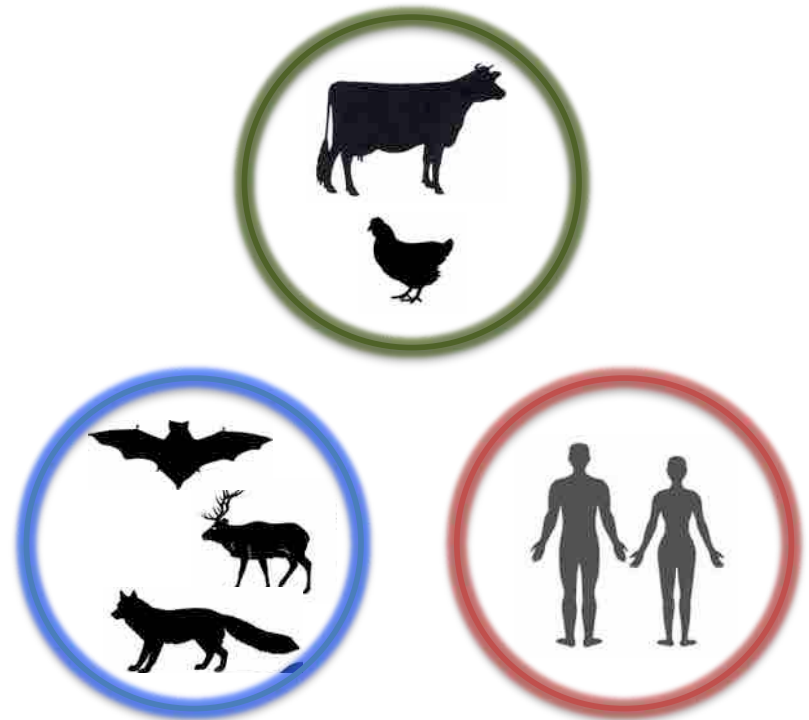
Cross-species transmission networks

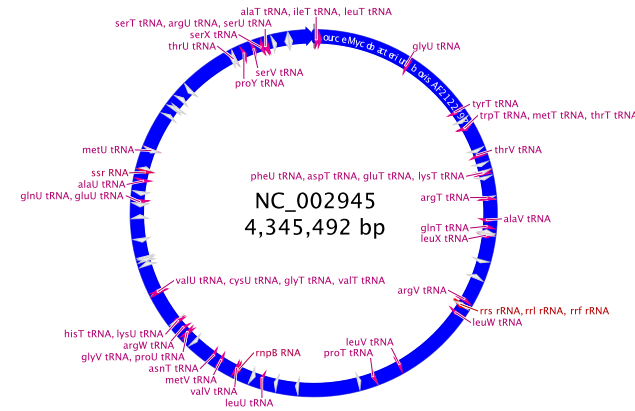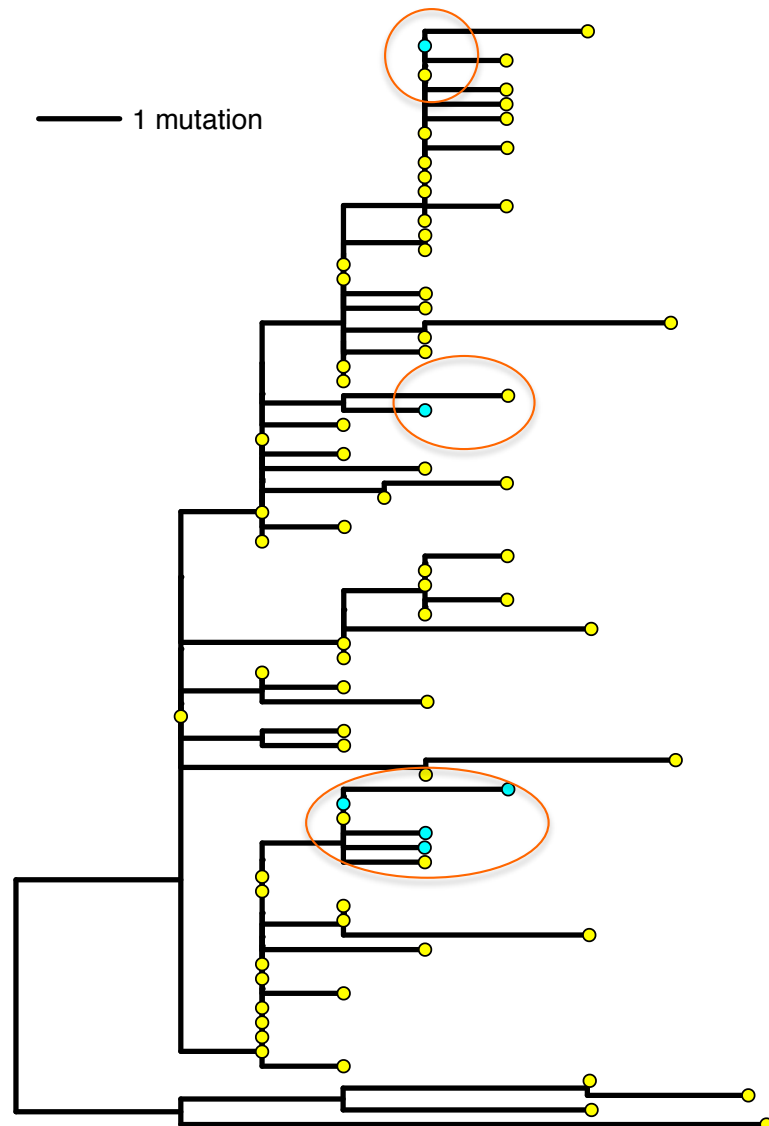*Streicker et al. 2010, Science*

Multi-host pathogens

Majority of emerging
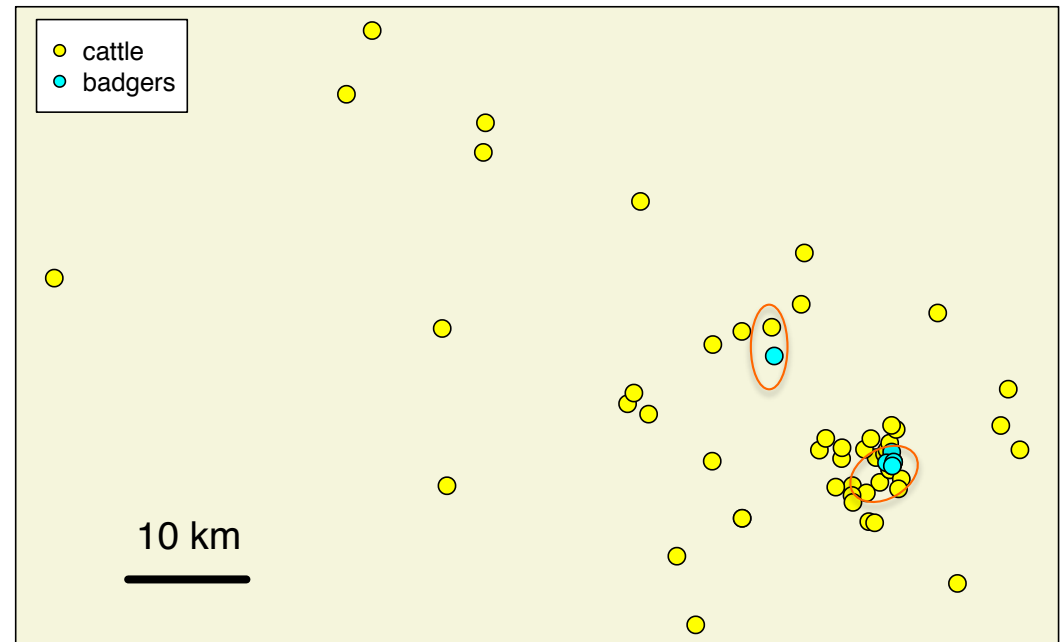infectious diseases are
caused by multi-host
pathogens

Complex transmission
dynamics that often involve
wildlife, livestock and
human populations

Bovine TB in badgers and cattle

1 mutation

M. bovis genome

NC_002945
4,345,492 bp

serT tRNA, argU tRNA, serU tRNA
serX tRNA
thrU tRNA
serV tRNA
proY tRNA
metU tRNA
ssr RNA
alaU tRNA
glnU tRNA, gluU tRNA
alaT tRNA, ileT tRNA, leuT tRNA
glyU tRNA
tyrT tRNA
trpT tRNA, metT tRNA, thrT tRNA
thrV tRNA
pheU tRNA, aspT tRNA, gluT tRNA, lysT tRNA
argT tRNA
alaV tRNA
glnT tRNA
leuX tRNA
argV tRNA
rrs rRNA, rrl rRNA, rrf rRNA
leuW tRNA
valU tRNA, cysU tRNA, glyT tRNA, valT tRNA
hisT tRNA, lysU tRNA
argW tRNA
glyV tRNA, proU tRNA
asnT tRNA
metV tRNA
valV tRNA
leuU tRNA
rnpB RNA
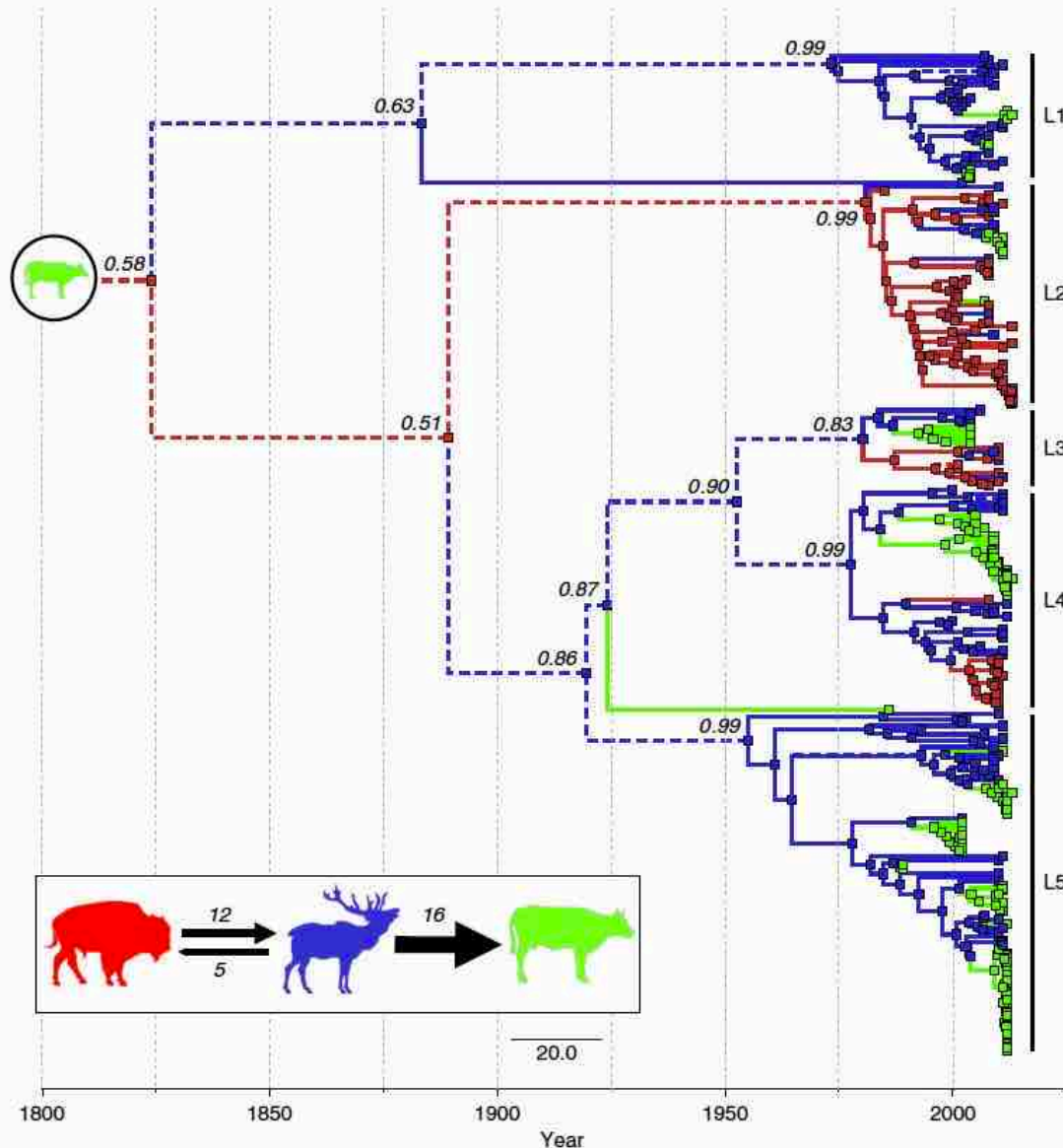leuV tRNA
proT tRNA
 source Myco bact erium b ovis AF212 2 297

- ○ cattle
- ● badgers

10 km

Identification of infection reservoirs

Brucella abortus
in the Greater
Yellowstone
Ecosystem
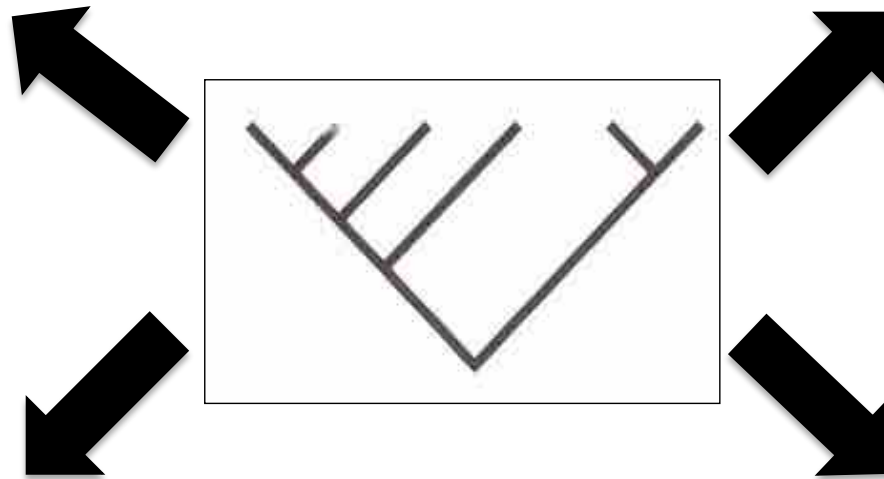
*Kamath et al. 2016,
Nature Comm*

Phylogenies at the centre of phylodynamic questions

Molecular clocks
e.g. "*How long ago since two groups split?*"

Selection
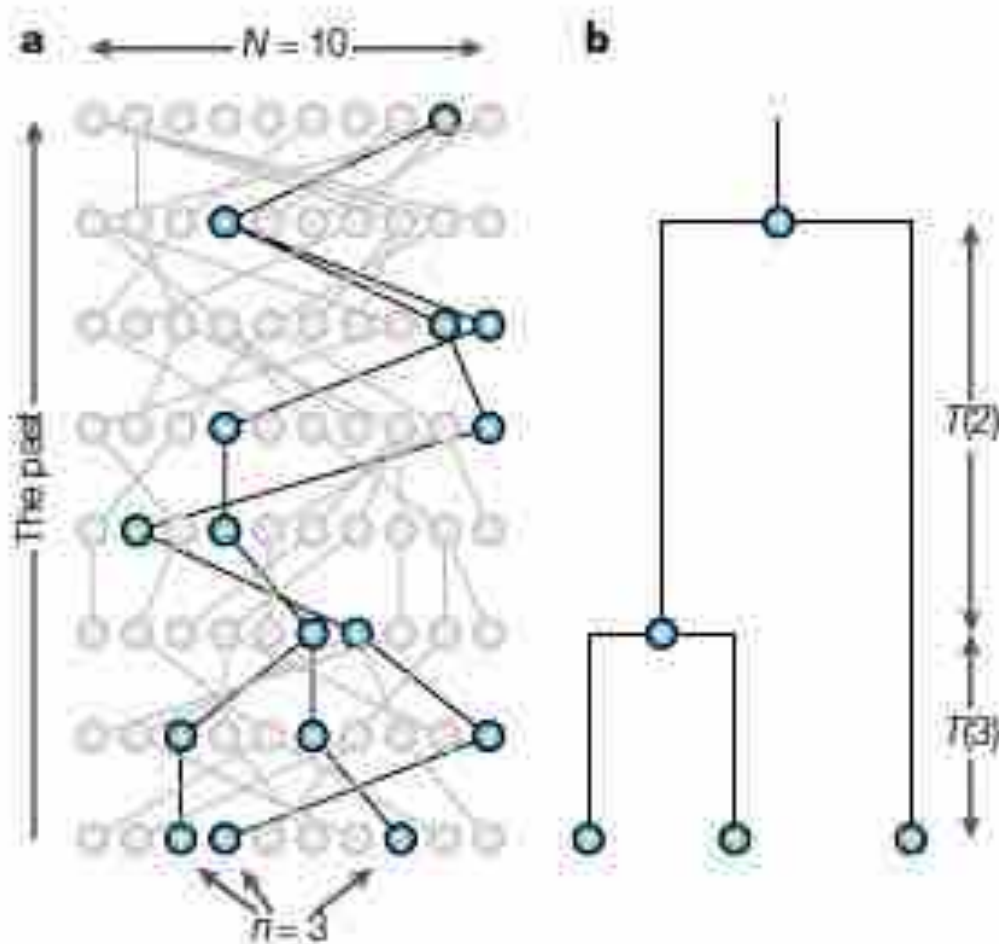e.g. "*Which genes or sites have undergone adaptive change*?"

Ancestral state change
e.g. "*Movement rate between population A and B?*"

Demographic reconstruction
e.g. "*How has population size changed through time?*"

Coalescent theory

a ← N = 10 → b

The past

T(2)

T(3)

n = 3

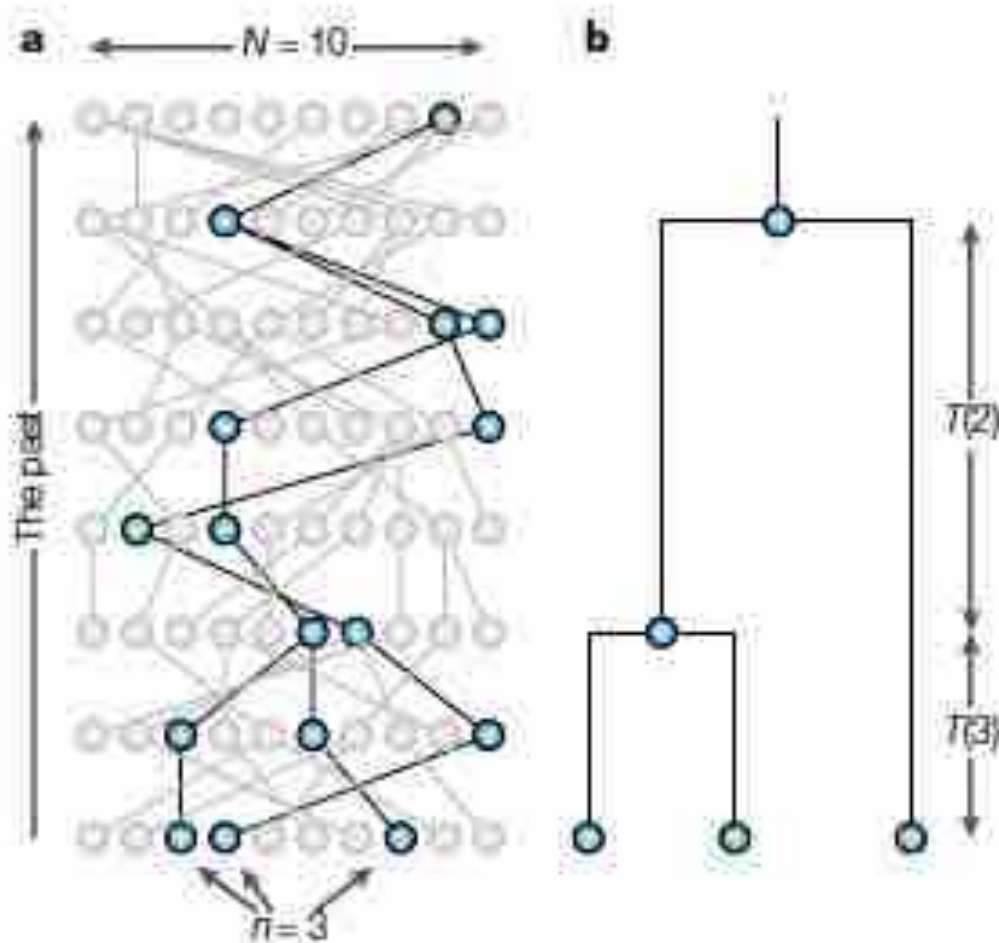*Rosenberg and Nordborg 2002,*
Nature Reviews Genetics

Going back in time, lineages randomly pick their parents

Coalescent event: when two lineages pick the same parent

Continues until all lineages coalesce at the most recent common ancestor

15

Coalescent theory

Rate of coalescence depends on

1. How many lineages are still left: more lineages = faster rate

2. The number of individuals in the population: fewer individuals means fewer parents to choose from = faster rate

Time between coalescent events depends on Pop size

16

Assumptions
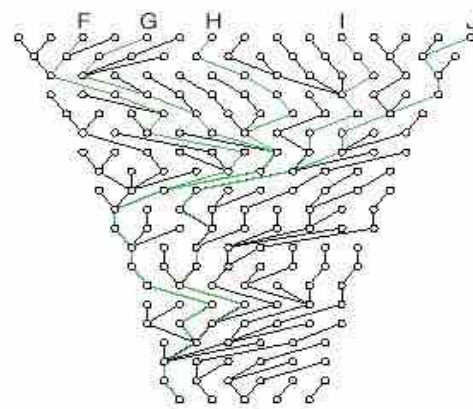
- N>>n – not more than one coalescent event in the same generation (diffusion approximation)
- No selection
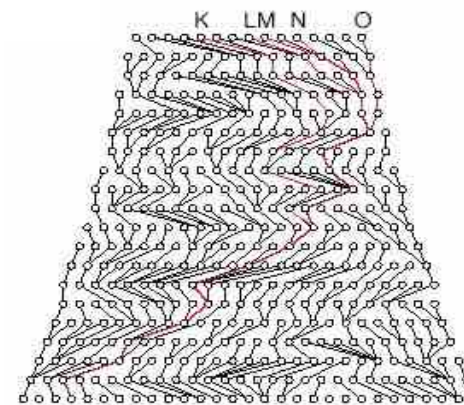- No recombination
- No population subdivision

Population size N

- Not the census population size but the effective population size (N_e)
- N_e = census size/ variance in the number of offspring
- For pathogen transmission, the estimate refers to the effective number of infections
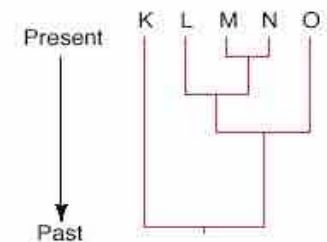
Coalescence in changing populations

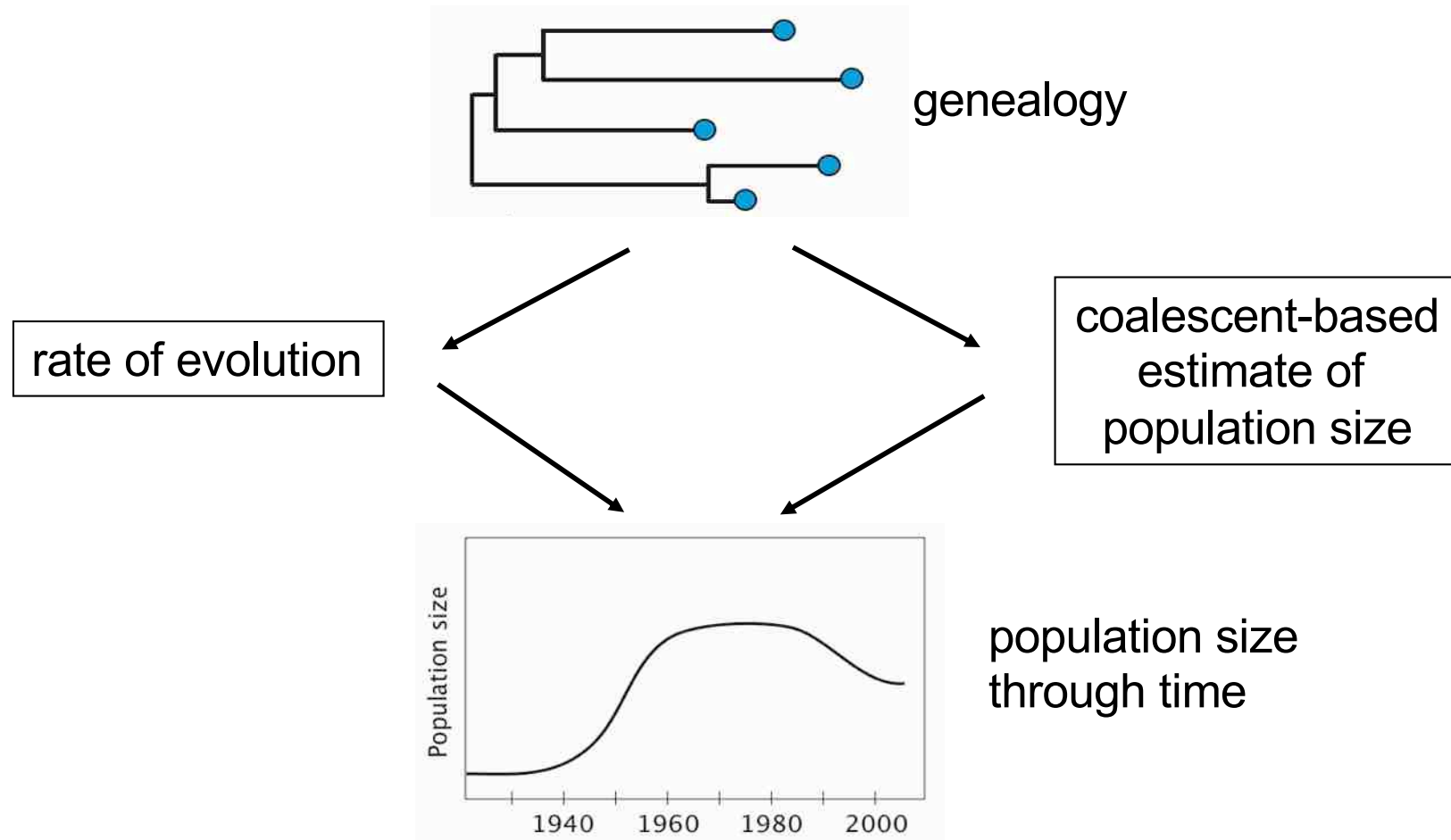Can be extended to changing population sizes by rescaling coalescent times

growing
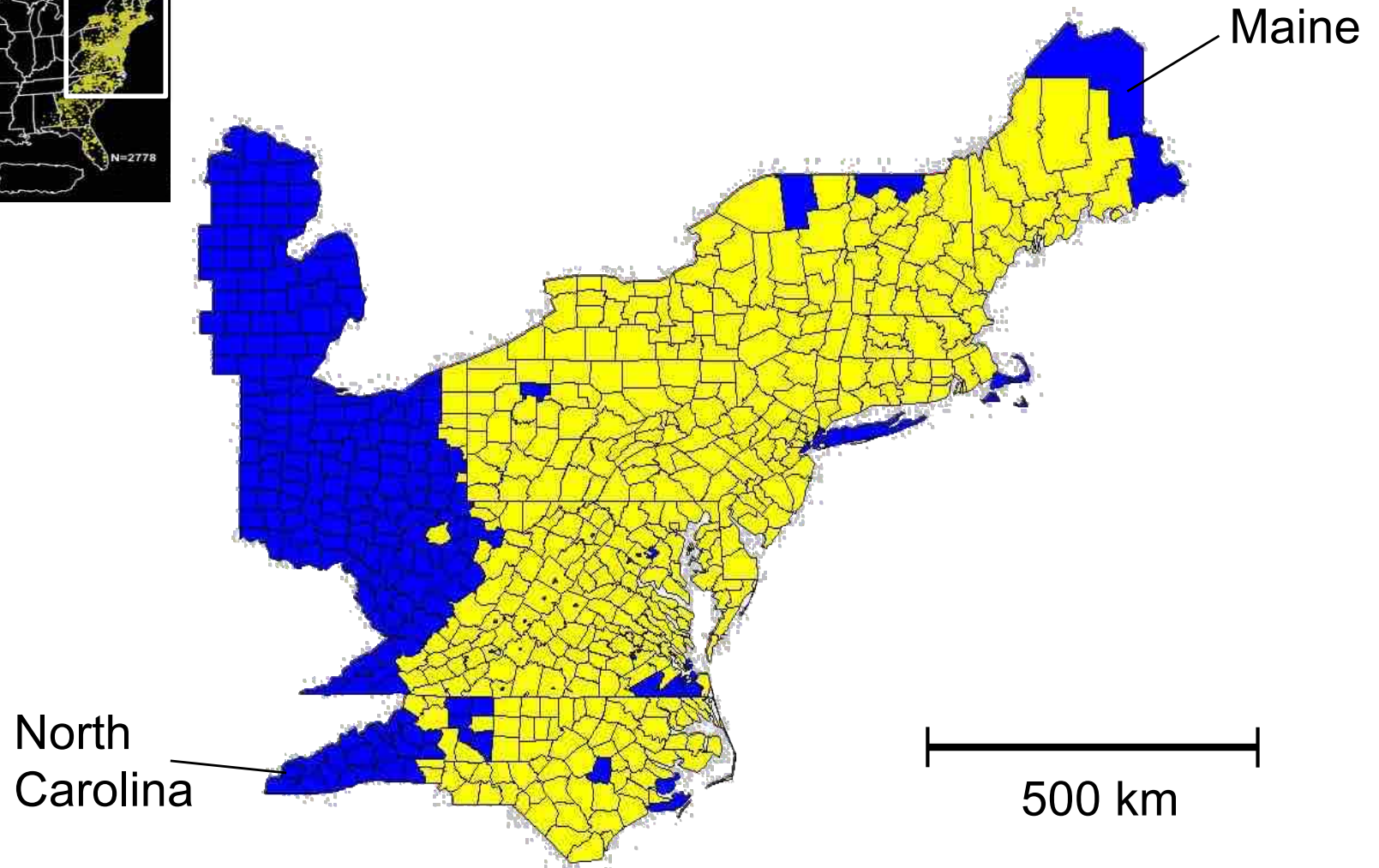
declining

Emerson 2001, Trends Ecol Evol

Combining clocks and demography

genealogy

rate of evolution

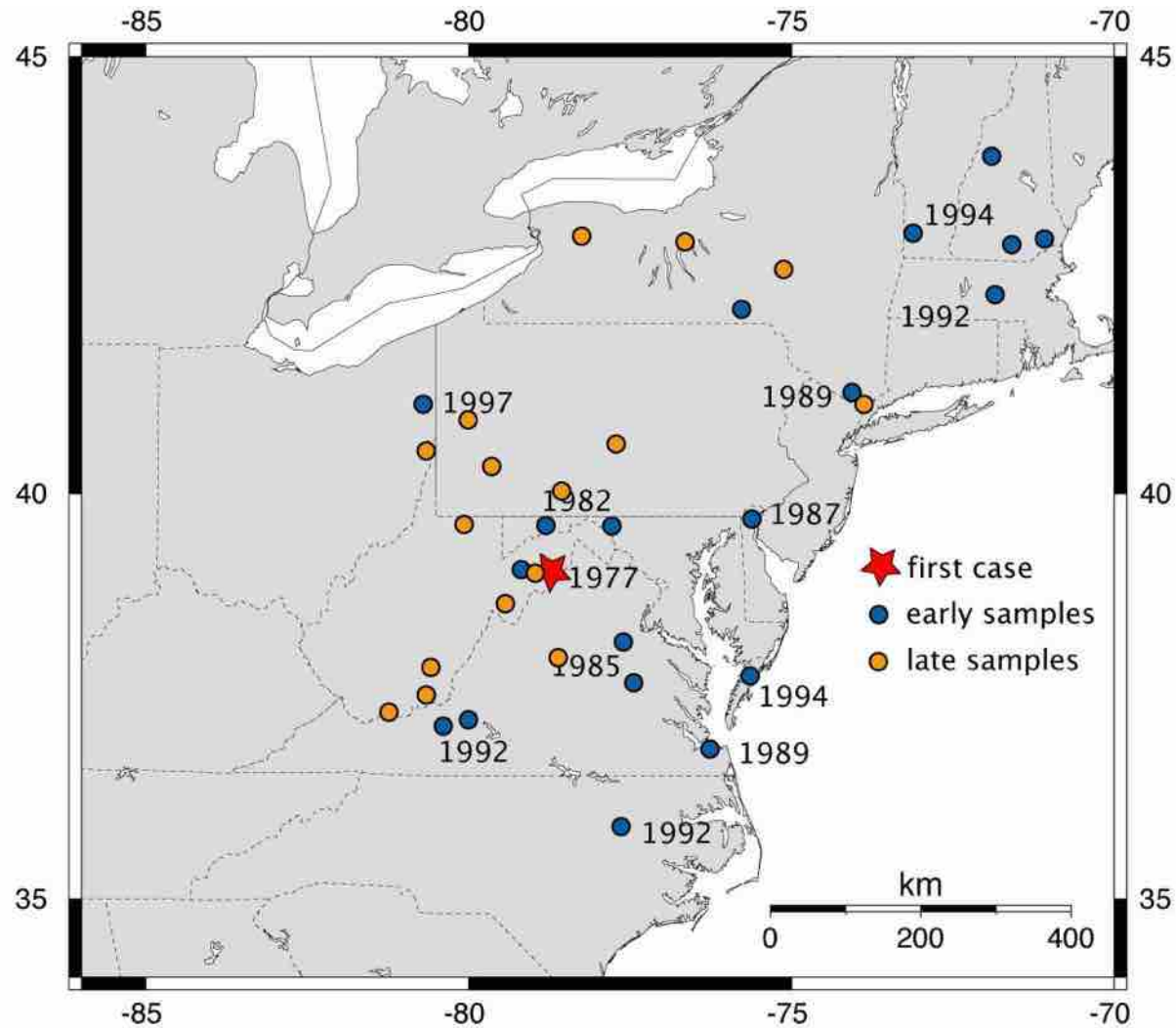coalescent-based
estimate of
population size

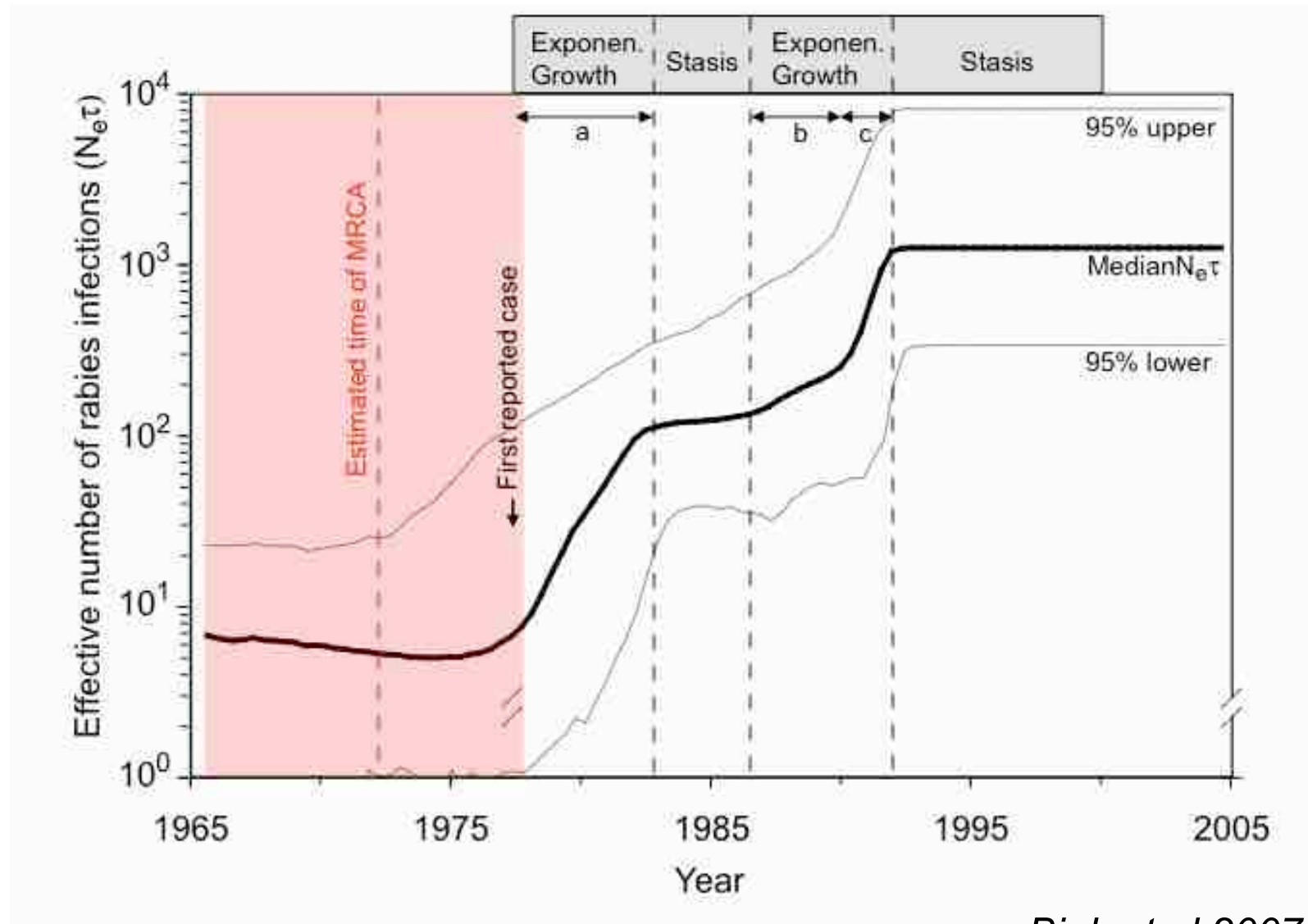population size
through time

Rabies invasion history 1977-1999
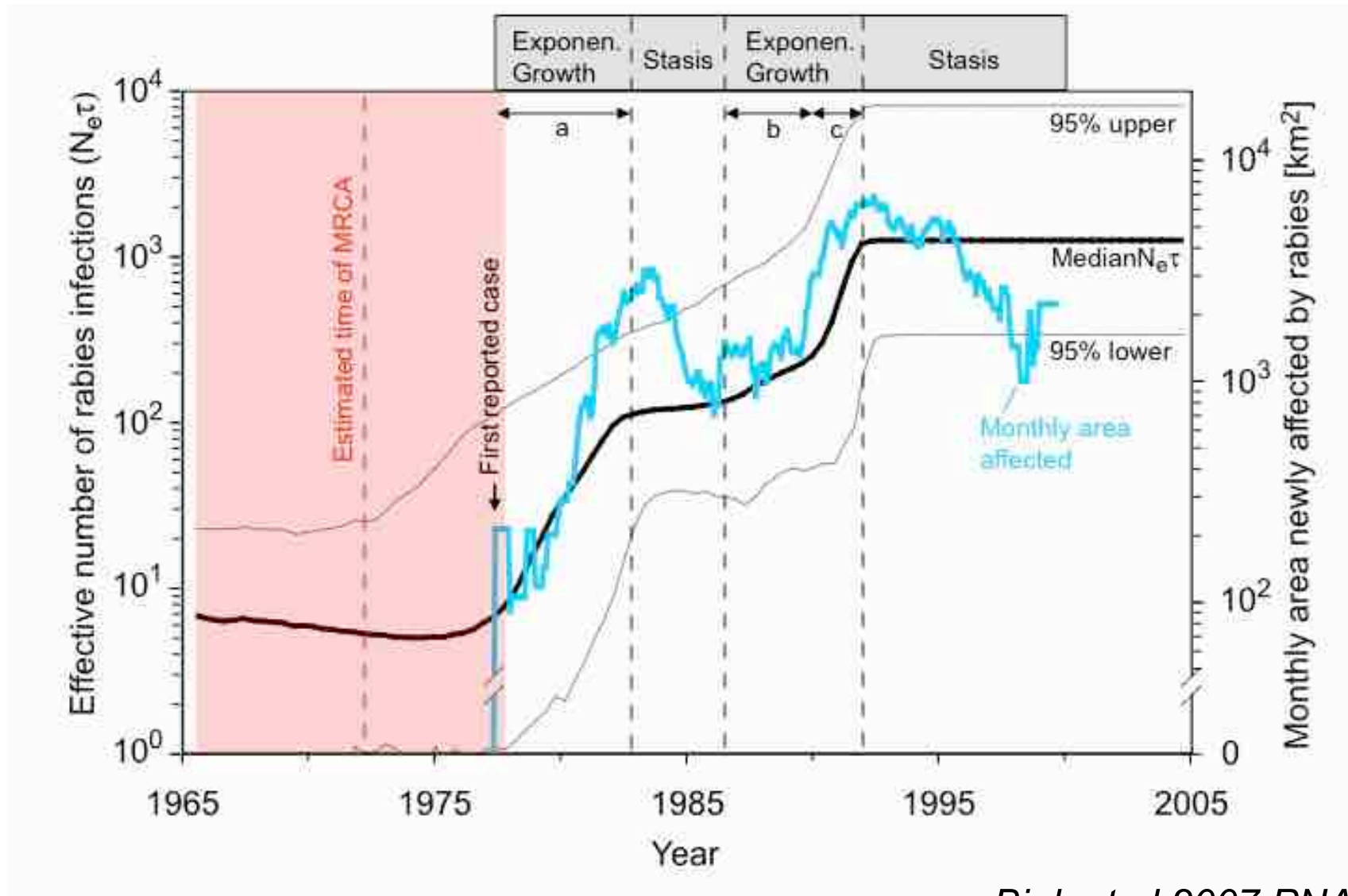
Sampling scheme

Change in the number of infected raccoons

Biek et al 2007,PNAS

Close correspondence to observed data

Biek et al 2007,PNAS

Ideal study design rarely achievable

- Typical problems:
    - Opportunistic sampling
    - Variable surveillance effort (e.g. alternative host species)
    - Variation in detectability (e.g. asymptomatic infections)
    - Appropriate scale often not known in advance

- Many of these will result in certain degree of sampling bias
    - How can bias be minimised?
   - How much does it matter for a given research question?

Dealing with sampling bias

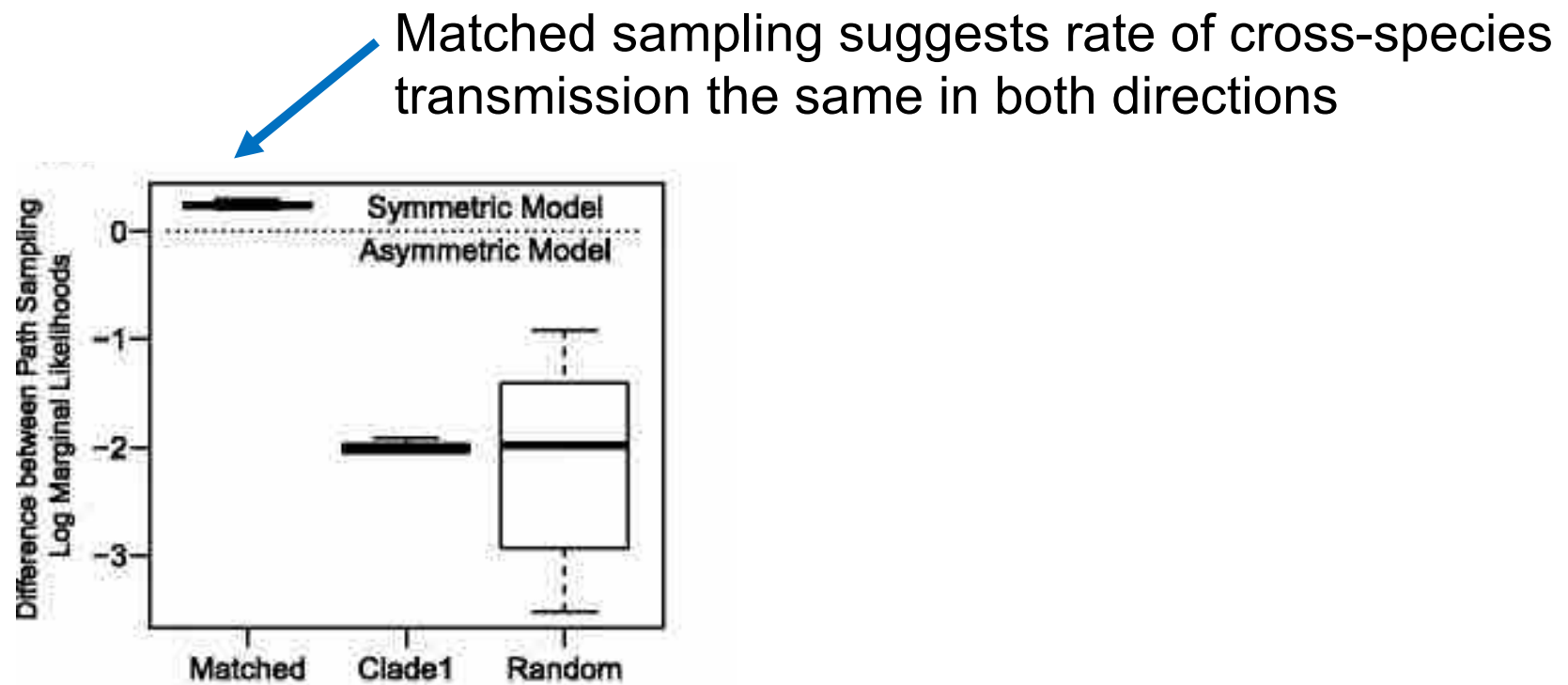| Type of inference	Potential for bias affecting results	Recommended practice
Molecular clock	Usually more robust	Sample evenly across time, maximise date range
Demographic inference (i.e. change in population size through time)	High	Define target population, sample evenly across space AND time
Ancestral states (e.g. cross species transmission	High	Sample evenly across space AND species AND time

Bovine TB in NZ cattle and wildlife

Crispell et al 2017,BMC Genomics

Bovine TB in NZ cattle and wildlife

Crispell et al 2017,BMC Genomics
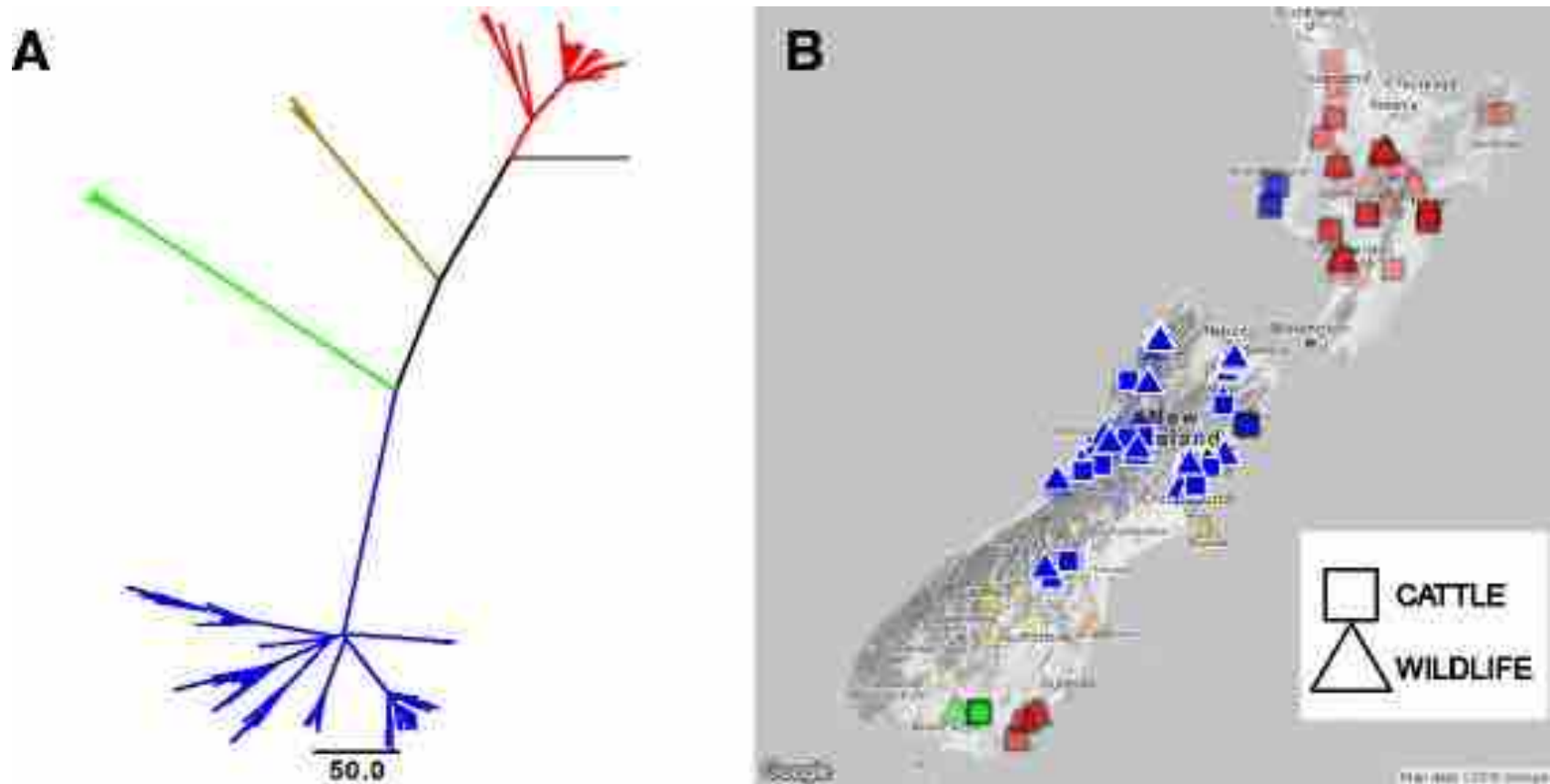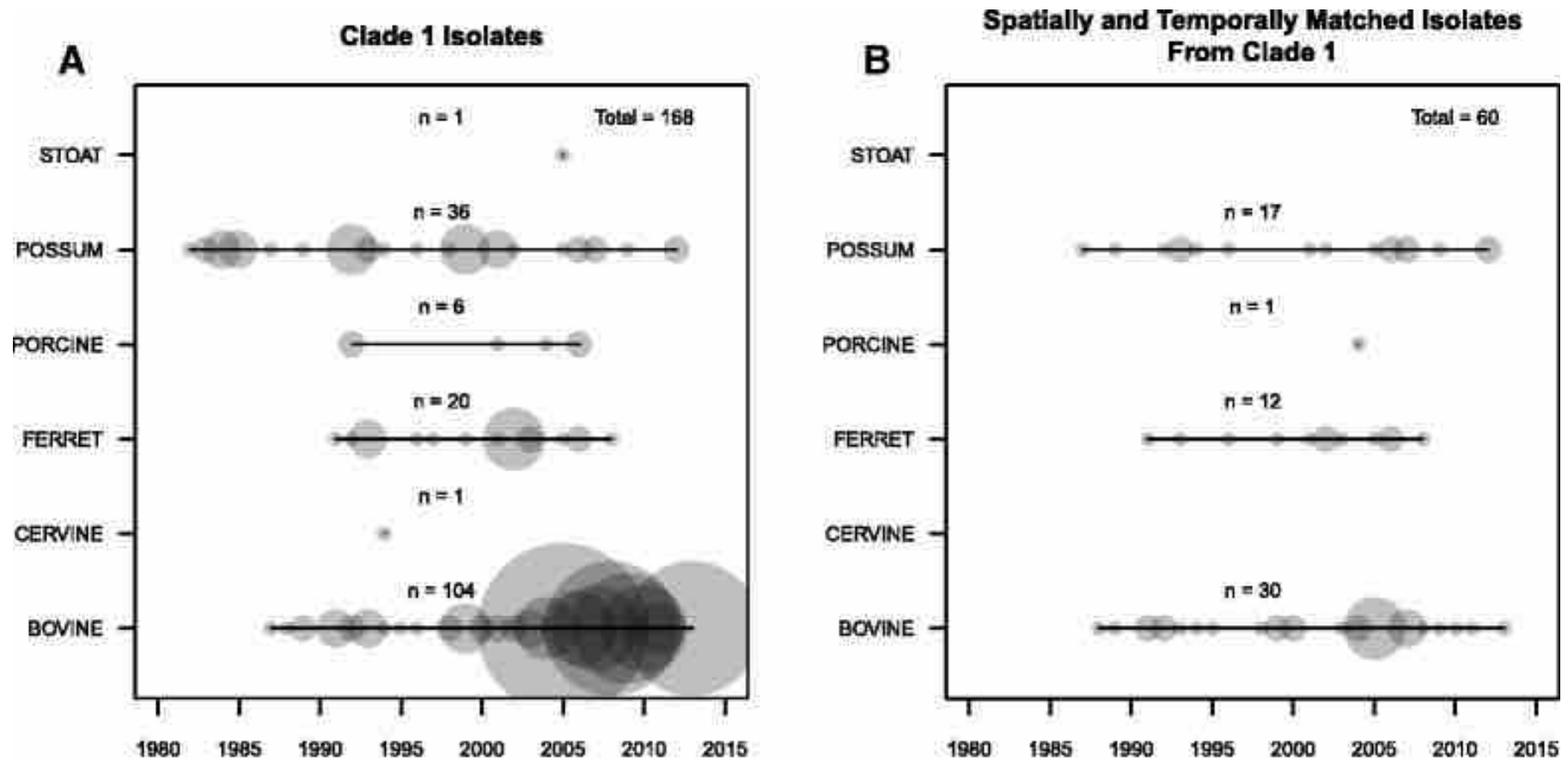
Sampling changes conclusions

Matched sampling suggests rate of cross-species transmission the same in both directions

Original (i.e. biased) sampling suggests more transmission from wildlife to cattle

Crispell et al 2017, BMC Genomics

Sampling biases

- In most cases inevitable
- Rarely explored systematically
- Molecular clock models more robust
- Well-documented effects on demographic and diffusion models