

Phylodynamics Workshop

Dr Samantha Lycett



Infection & Immunity
Roslin Institute
University of Edinburgh

June 2023



samantha.lycett@ed.ac.uk

Welcome to Phylodynamics !

- General introduction (talk)
- Molecular clock and Tempest (practical / demo)
- Time-scales and BEAST trees (talk)
- Phylodynamics and Phylogeography Practical (see handout)
 - Setting up BEAST1 using BEAUTI
 - Using Tracer, TreeAnnotator and FigTree to look at the results
 - Phylogeography: including traits in BEAUTI

What can pathogen sequences tell us ?

Introducing Phylodynamics and Phylogeography

How to use sequence data to infer transmission patterns



Pathogen genomes and Molecular Evolution

- Pathogen sequences (DNA / RNA, Protein) accumulate mutations over time
- Different types of mutations:
 - Mutations in nucleotide only
 - Mutations that change amino acid
 - Insertions and deletions
 - Recombination & Gene transfer

Nucleotides (A,C,G,T)

a	g	t	g	a	g	t	g	a	t	t	c	g	a	c	a	a	a	t	a	t	a	a	a	c	a	t	t	
a	g	t	g	a	g	t	g	a	t	t	c	a	a	c	a	a	a	c	t	a	t	a	a	a	c	a	t	t
a	g	t	g	g	g	t	g	a	t	t	c	a	a	c	a	a	a	t	a	t	a	a	a	c	a	t	t	
a	g	t	g	a	g	c	g	a	t	t	c	a	a	c	a	a	a	c	t	a	t	a	a	a	c	a	t	t
a	g	t	g	a	g	c	g	a	t	t	c	a	a	c	a	a	a	c	t	a	t	a	a	a	c	a	t	t

Amino Acids (Protein)

V	N	E	I	N	P	T	T	T	I	I	N	N	N	T	I	Q	N	N	F	T	N	I	T	N	I	I	L	N	K	D		
T	N	E	I	N	S	T	T	T	I	I	N	N	T	I	Y	N	T	I	Q	N	N	F	T	N	V	N	I	V	L	I	K	E
M	N	E	I	N	P	T	T	T															N	I	T	N	I	I	V	N	K	N
M	N	E	I	N	P	T	T	T															N	I	T	N	I	I	V	N	K	N
M	N	E	T	S	P	T	T	T	I	I	N	N	H	P	Q	N	N	F	T	N	I	T	N	I	I	V	T	K	N			
M	N	E	T	S	P	T	T	T	I	I	N	N	H	P	Q	N	N	F	T	N	I	T	N	I	I	V	T	K	N			

Sequence data: Each row is a sample
Each column is a genome position

Pathogen Genomes and Evolutionary Rates

	RNA Viruses	DNA Viruses	Bacteria
Replication & Evolution	Fast and error prone	Slower, more conserved	Very slow, but horizontal gene transfer
Genome size	8-30kb	20-200kb	4Mb
Mutations per year	10-100	1-20	0-1

HIV, Ebola
Foot-and-Mouth
Coronavirus

Segmented ssRNA

Avian influenza
Schmallenberg

African Swine
Fever

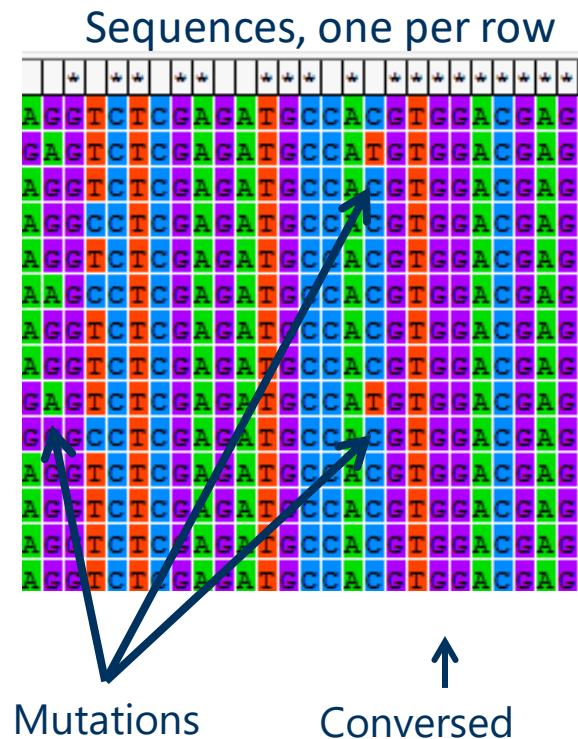
Segmented dsRNA

Blue Tongue
African Horse Sickness

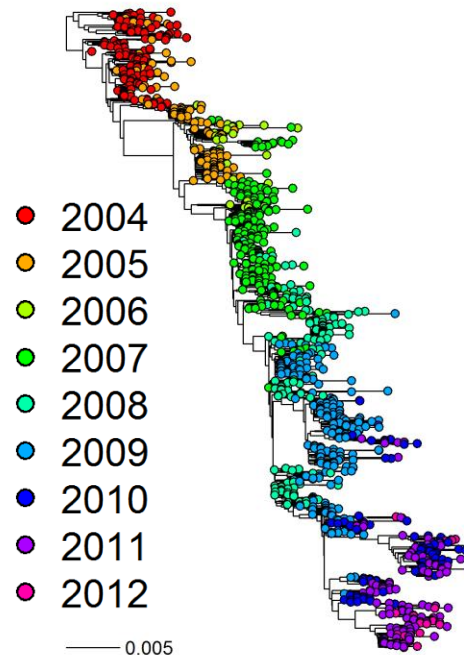
Bovine Tb

Pathogen Sequence Data

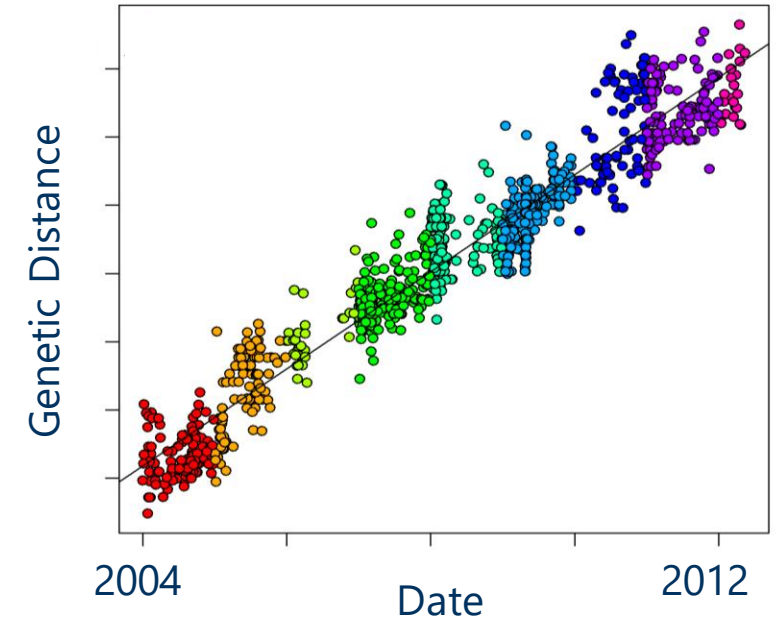
- Pathogen sequences provide richer information than strain type
- Sequences accumulate mutations over time



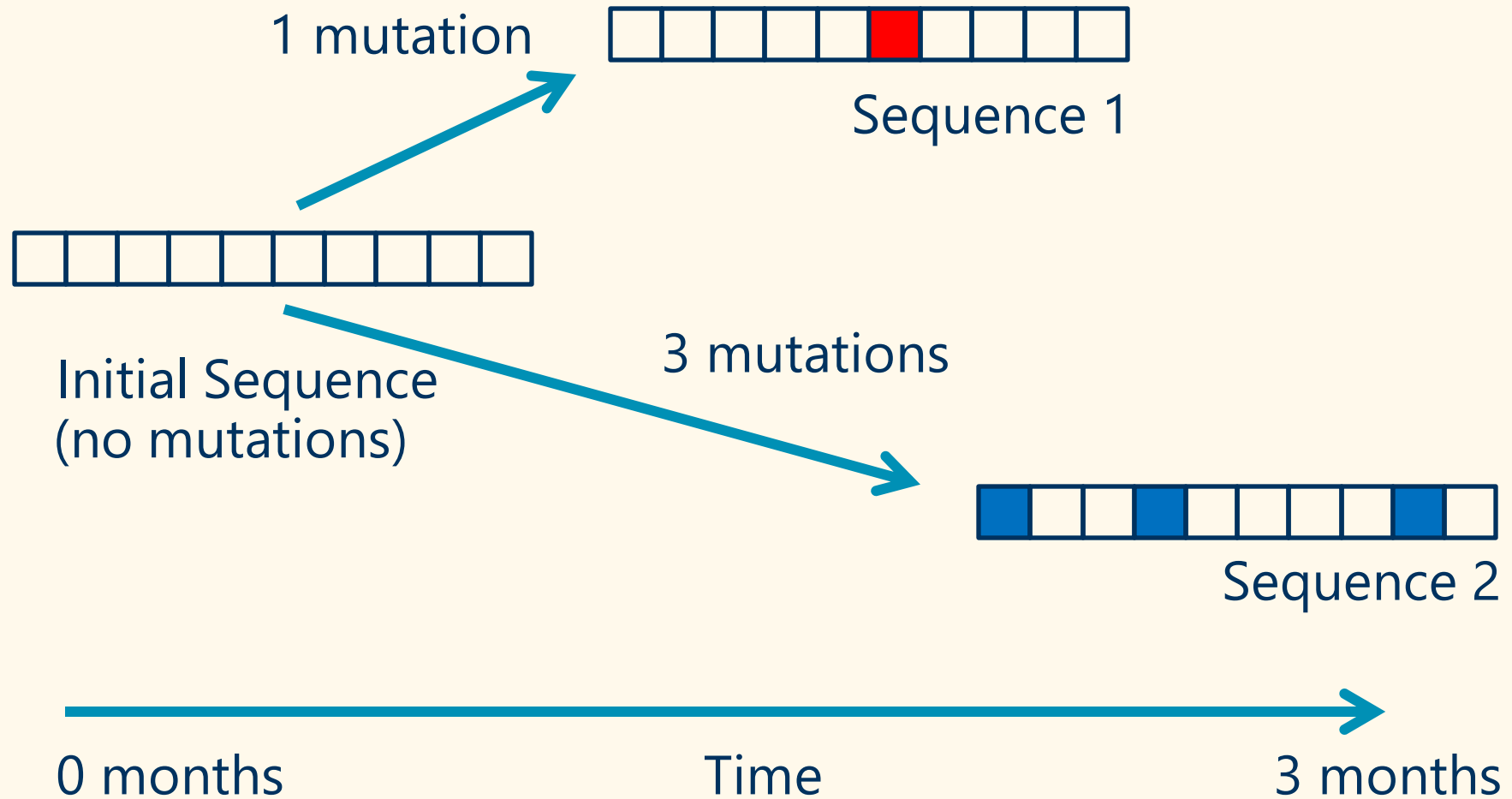
Tree of Human Influenza



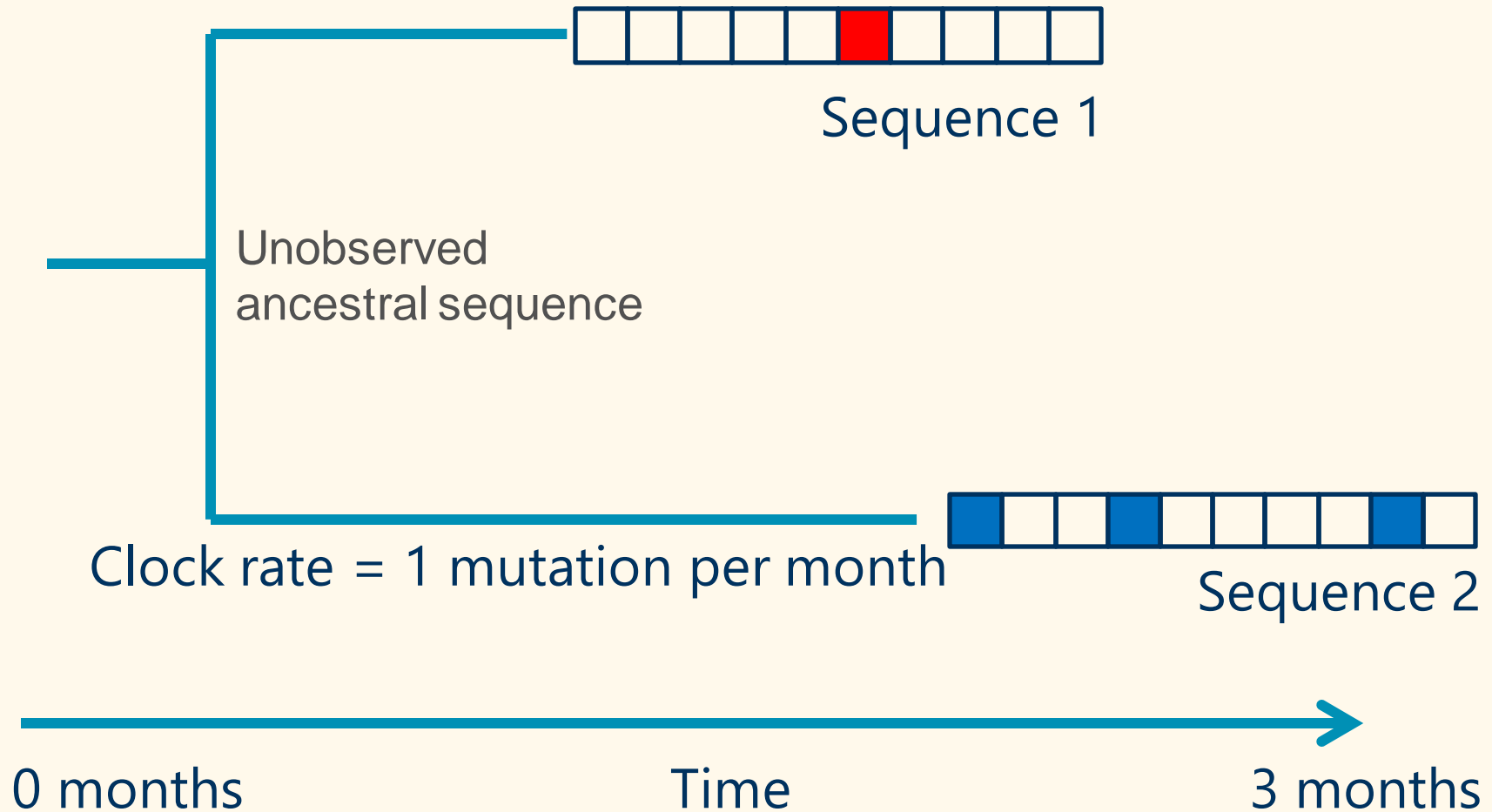
Genetic Distance from Root



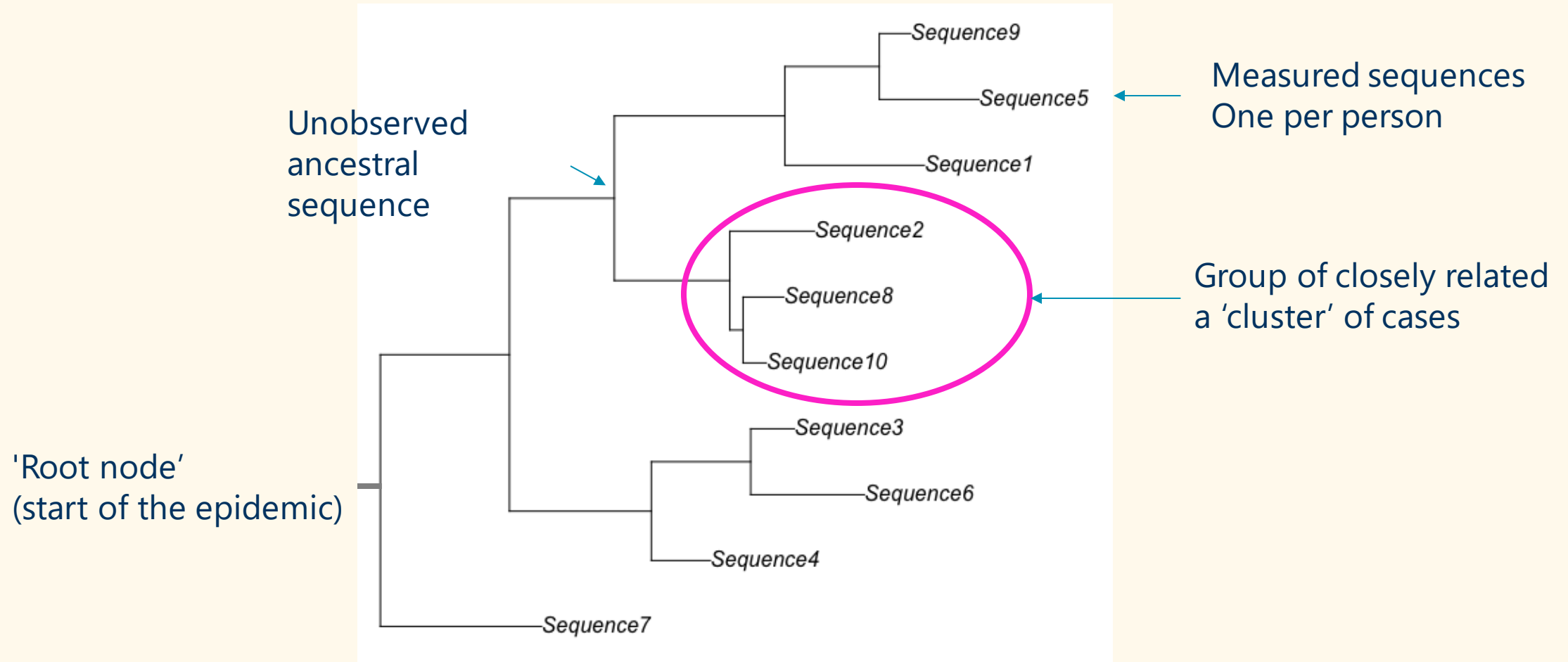
How are the sequences related ?



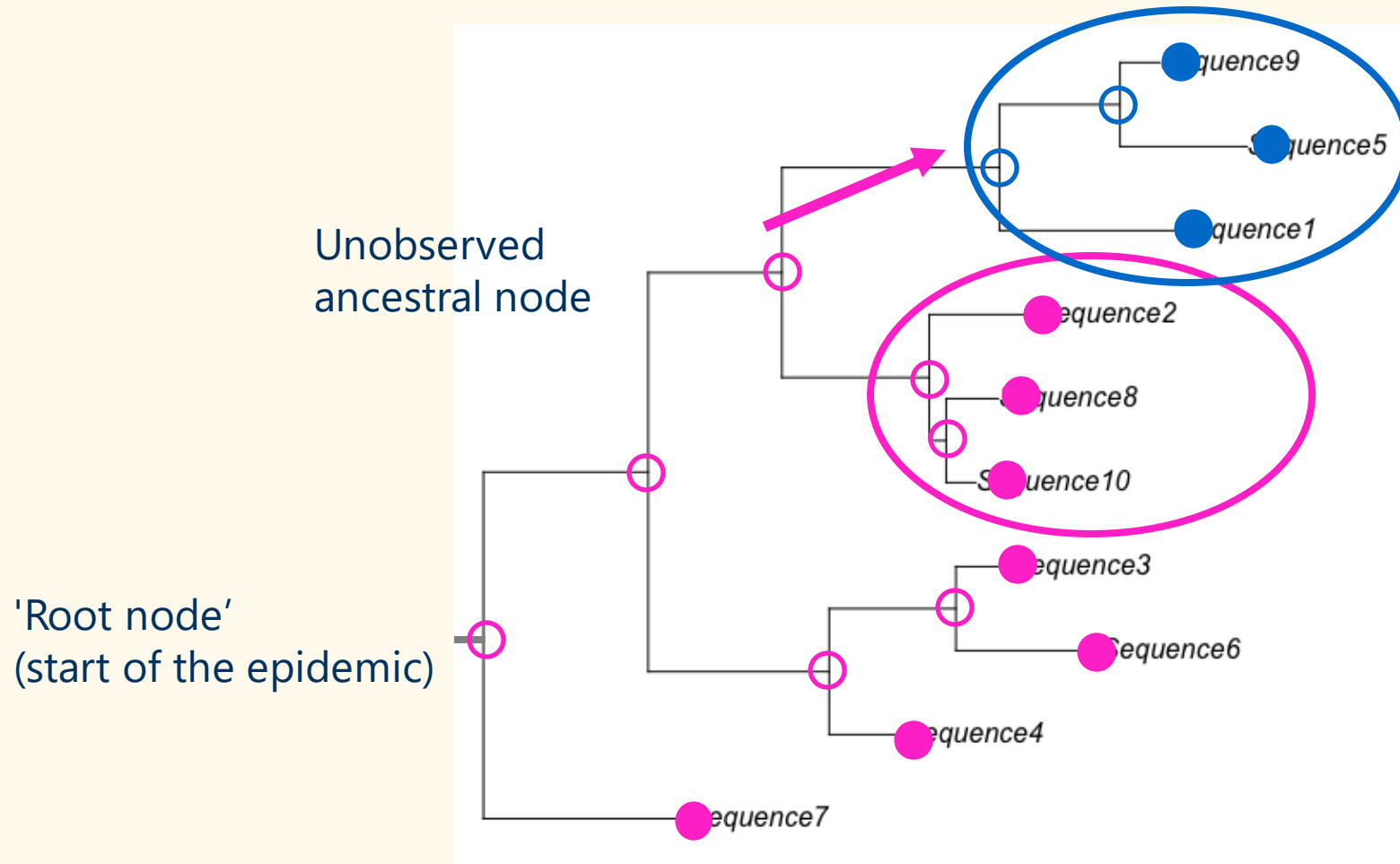
Adding a time scale and creating a Phylogenetic tree



Reading a phylogenetic tree



Reading a phylogenetic tree



Infer that:

Pink group (.. Place 1 ..)
infected

Blue group (..Place 2..)

Use a model to calculate
the probability of the
location of the ancestral
nodes

Phylodynamic Modelling Scenarios

Viral pathogen sequencing for disease tracking

Forensic Scenario

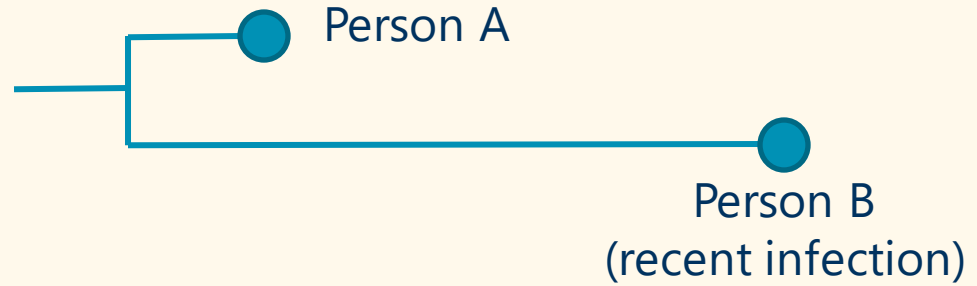
- For outbreak situations
- What strain is it ?
- When did it arrive ?
- Who infects whom ?
- Dense sampling required

Transmission Patterns

- For epidemics and pandemics
- Pattern of diversity
- Imports and Exports
- Quantify transmission between locations and hosts
- Surveillance sampling required

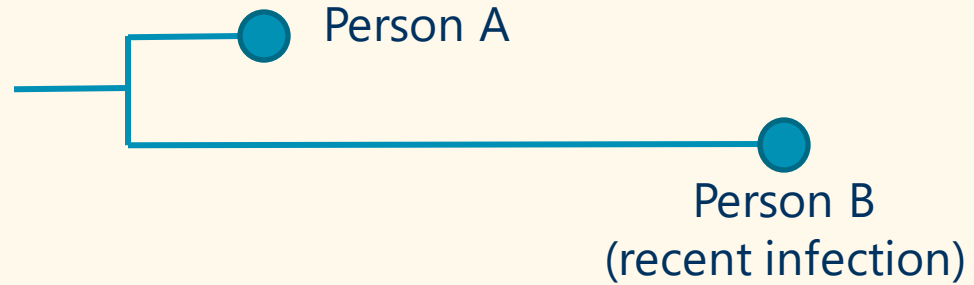
Detailed Forensic scenario - Can you tell who infected who ?

- Yes ?

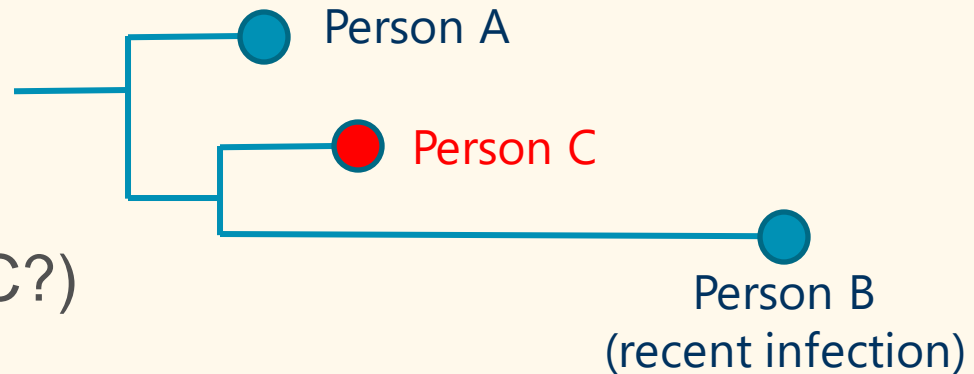


Detailed Forensic scenario - Can you tell who infected who ?

- Yes ?

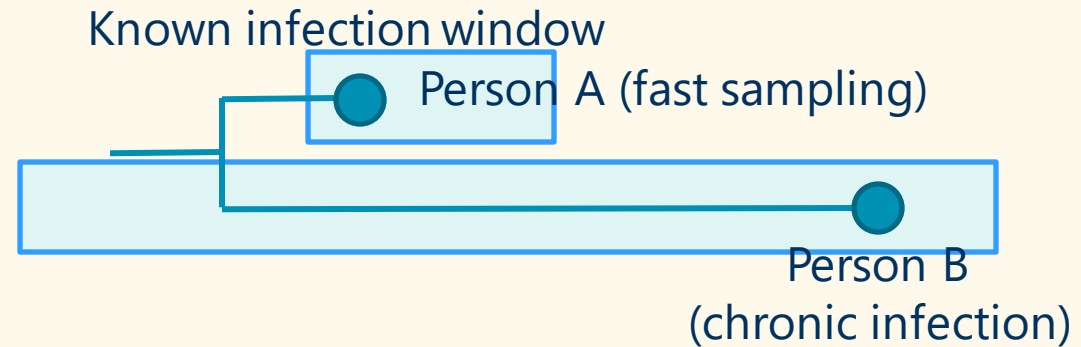


- No
- Needed to sample C
- (did you know about C?)



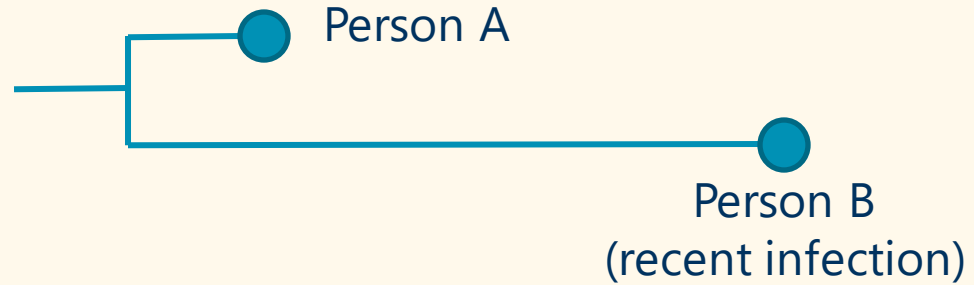
Detailed Forensic scenario - Can you tell who infected who ?

- No
- Need to know who was infected first, not who was sampled first

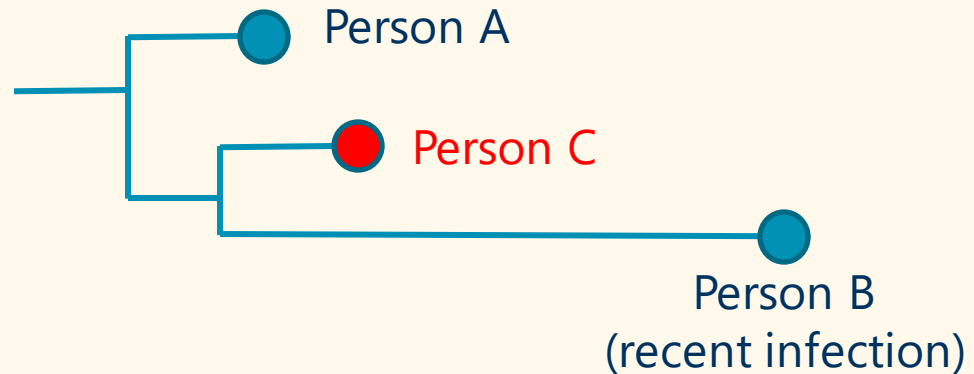


Detailed Forensic scenario - Can you tell who infected who ?

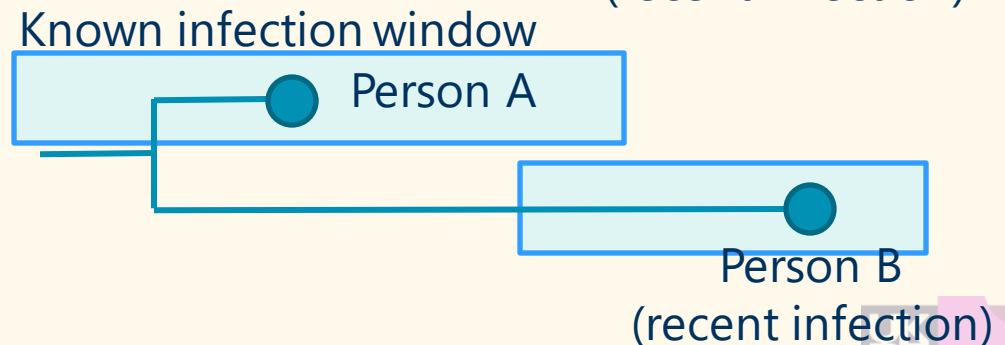
- Yes ?



- No

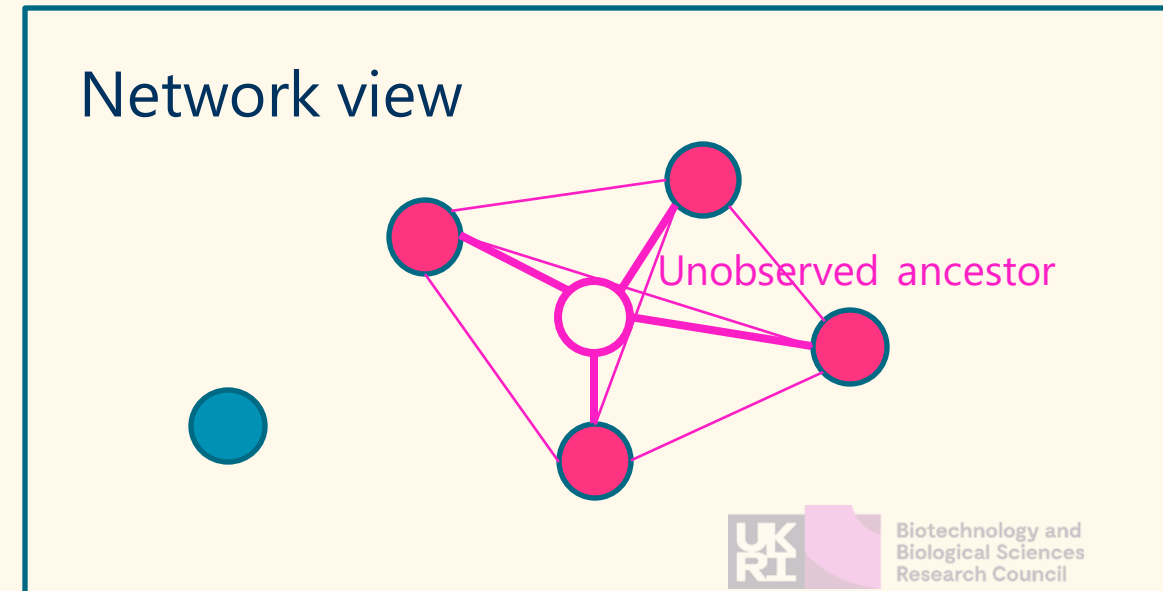
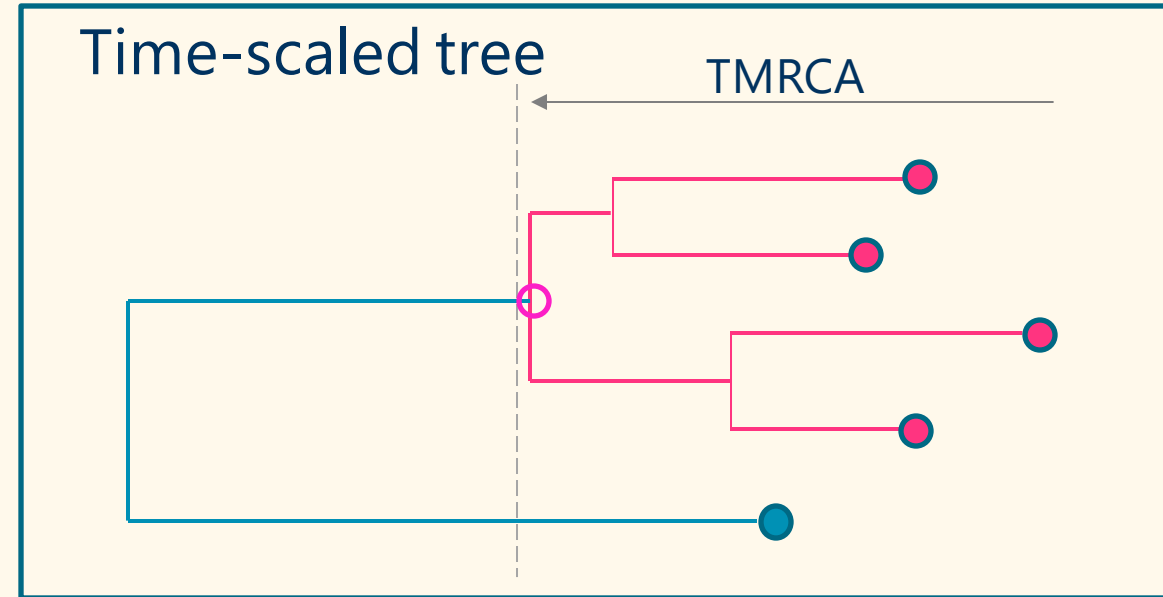


- Maybe ?



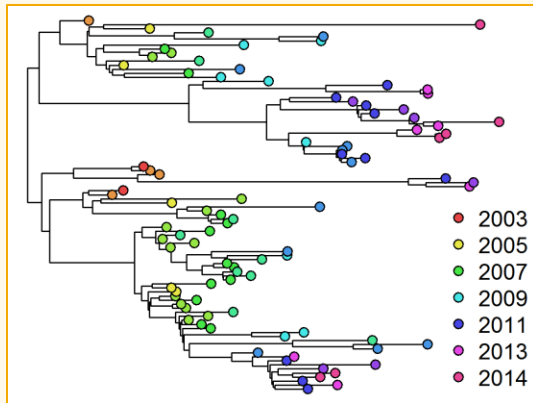
Genetic Clusters

- Exact determination of who infected whom is problematic; but can analyse collections of links between individuals
- **Clusters:** sequences from samples are **genetically similar**;
- (not had time to mutate much yet)
- Similar sequences have recent **time to most recent common ancestor (TMRCA)**
- Used in HIV analysis, Hospital infections and in Vet. Surveillance e.g. BVD

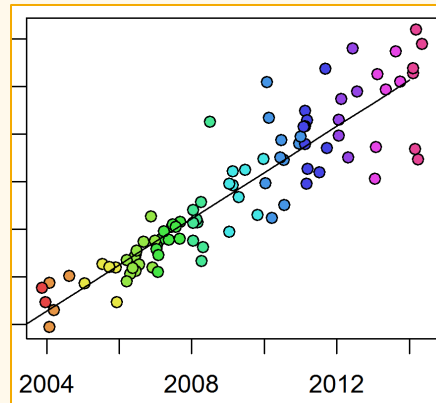


Pathogen Phylodynamics

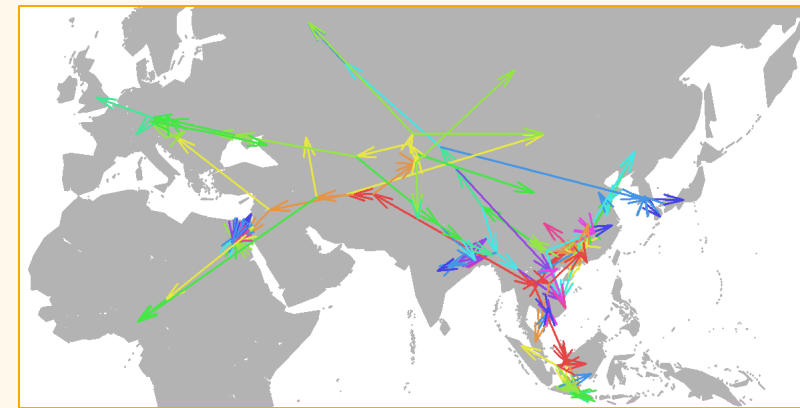
Harness pathogen sequence and disease surveillance data



Phylogenetic tree



Molecular Clock



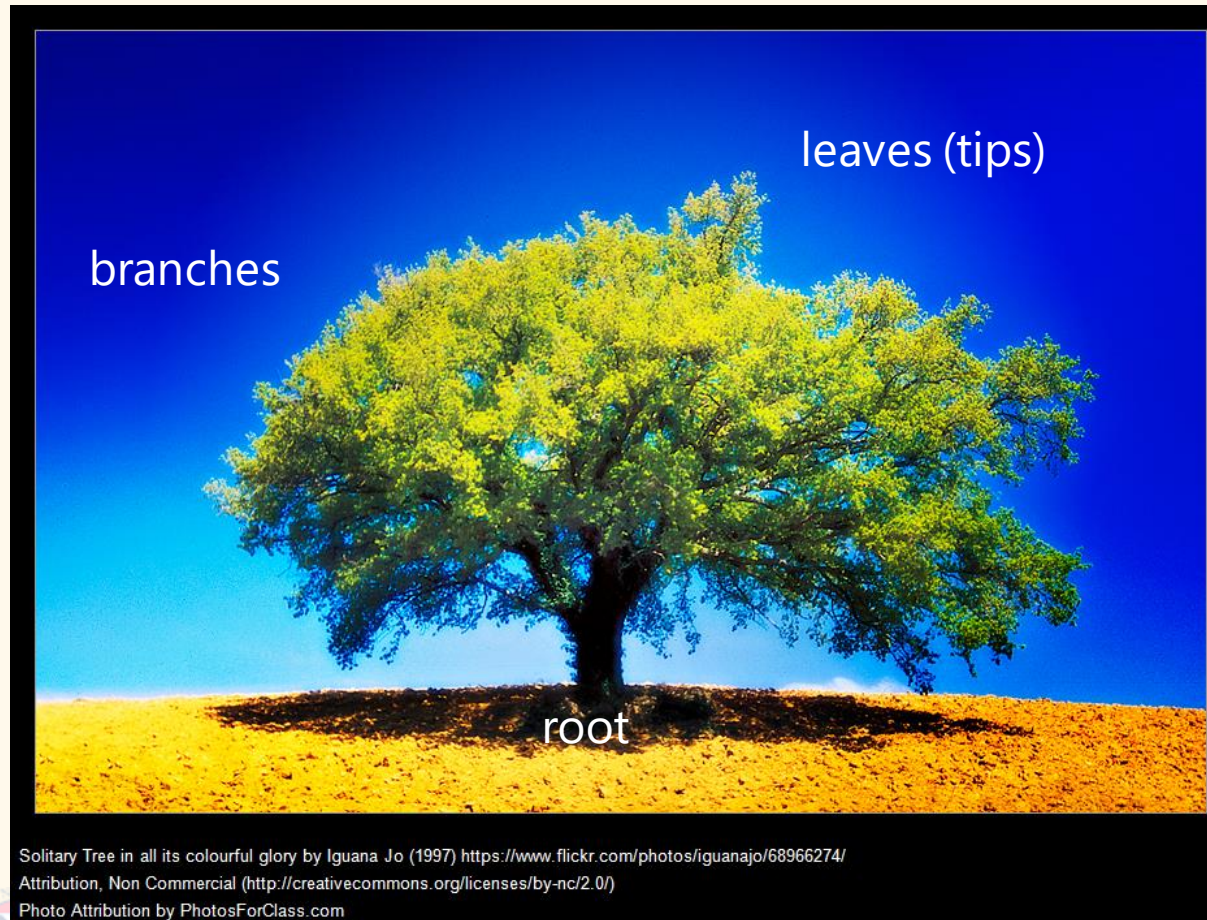
Tree + Spatial model

- Evolution and epidemiology of animal viruses and bacteria
- Cross species transmissions, host adaptations, epistatic interactions
- Combined evolutionary and spatial models
- Sources, routes and speed of spread
- Drivers of transmission patterns and evolution

Pathogens including: **Influenza**, **FMDV**, **BVDV**, **PRRSV**, **M.bovis**, **S.aureus** and **SARS-CoV-2**..

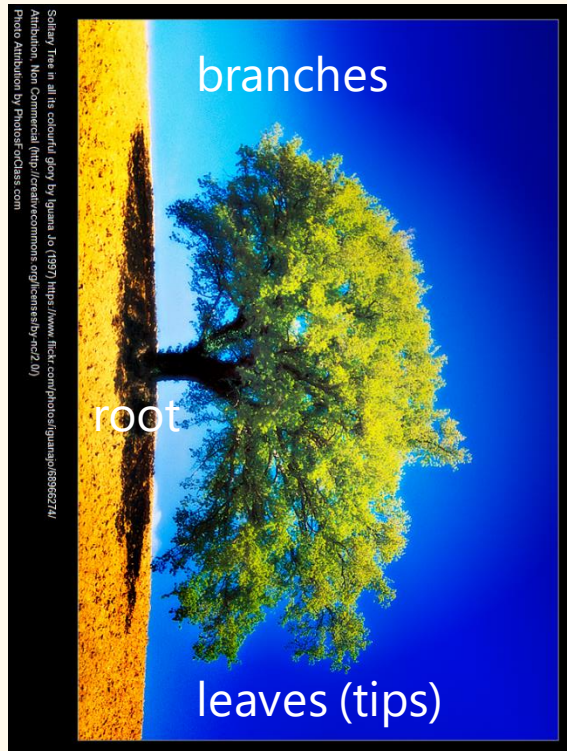
What can trees tell us ?

- A “plain” tree



What can trees tell us ?

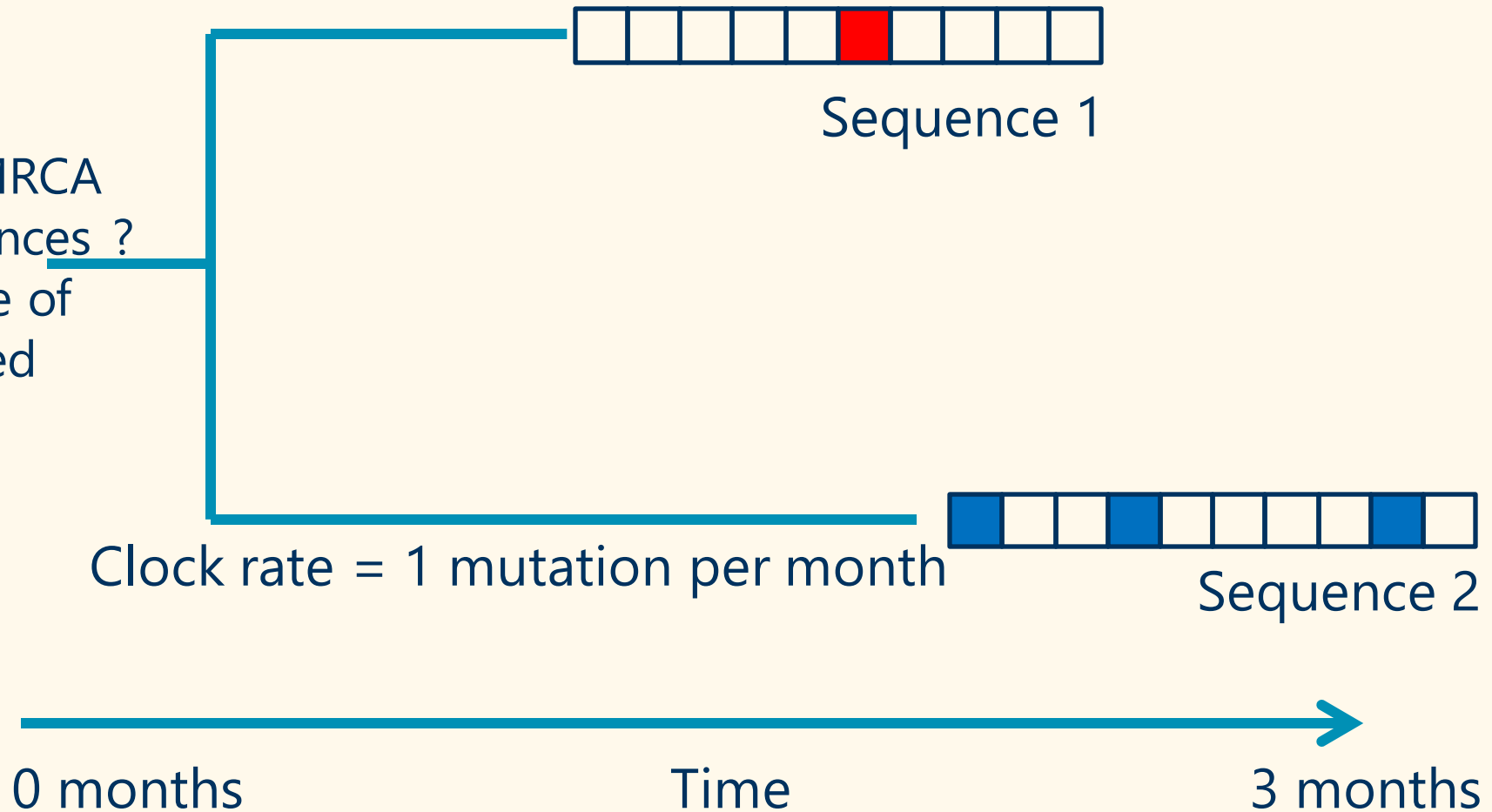
- A “plain” tree – timescale by molecular clock



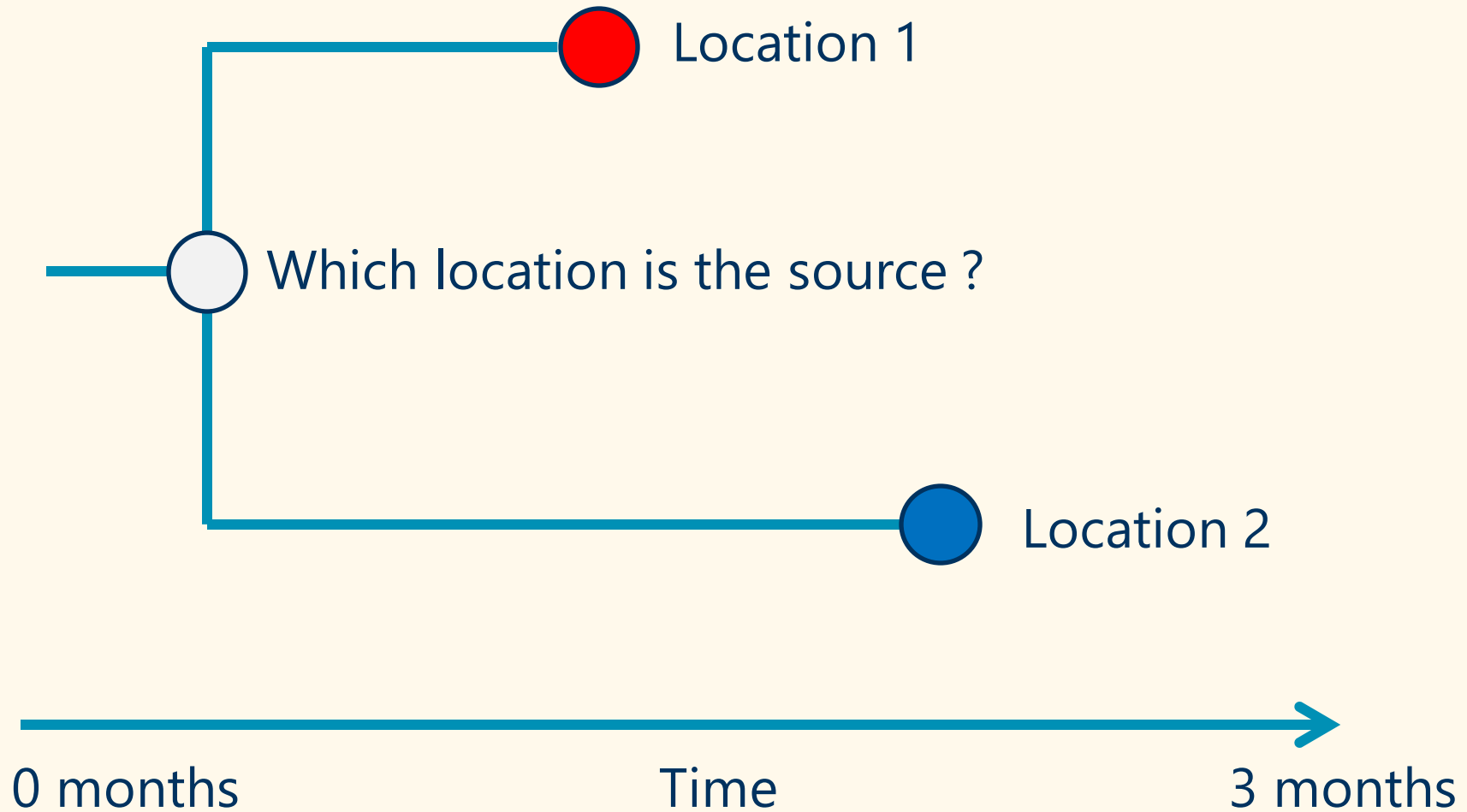
**Time to most recent common ancestor
(root age)**

Adding time scales to Trees (2)

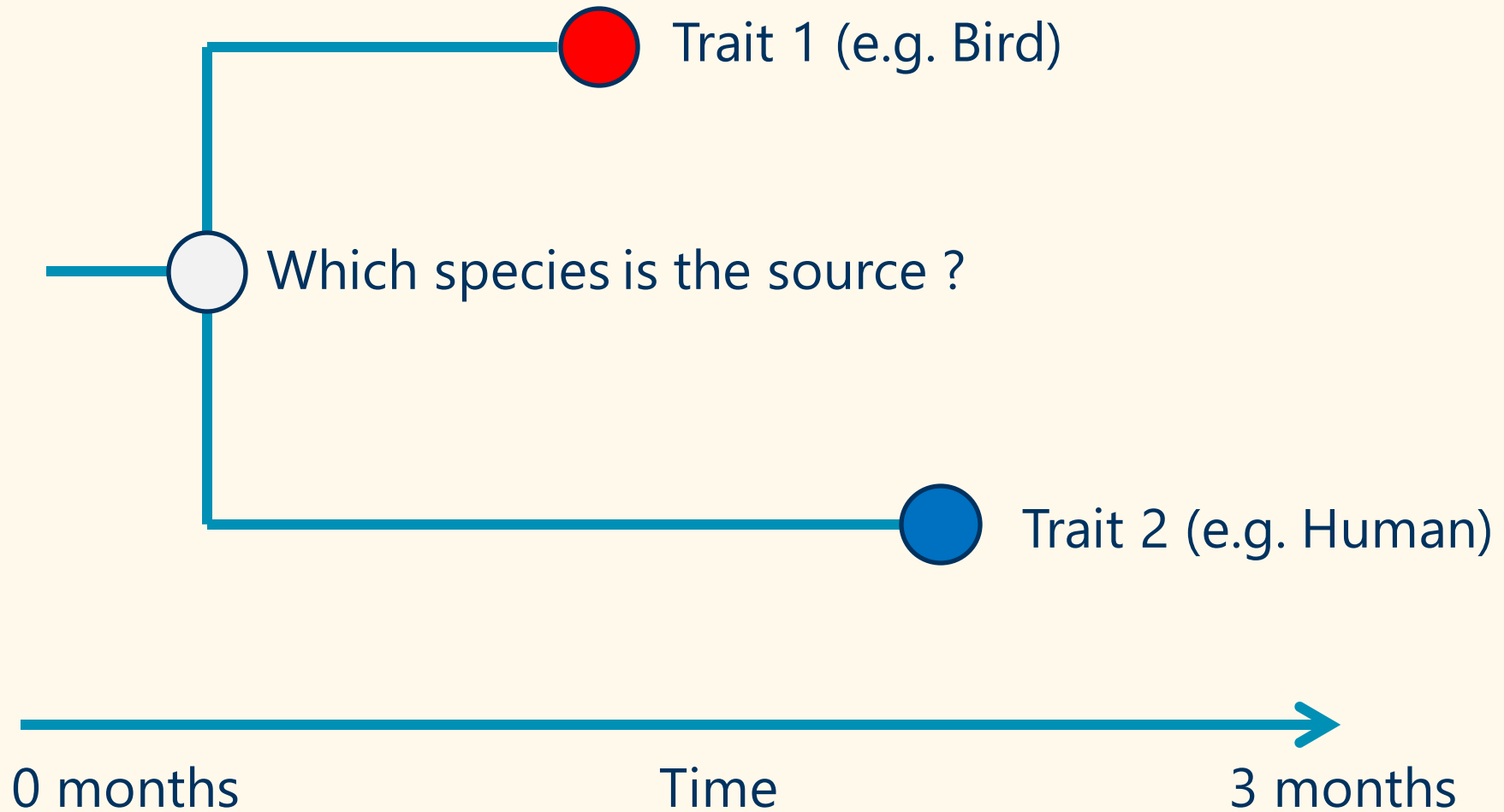
What is the TMRCA
of these sequences ?
(this is the time of
the un-observed
source)



Adding Locations to Trees



Adding other 'Traits' to trees



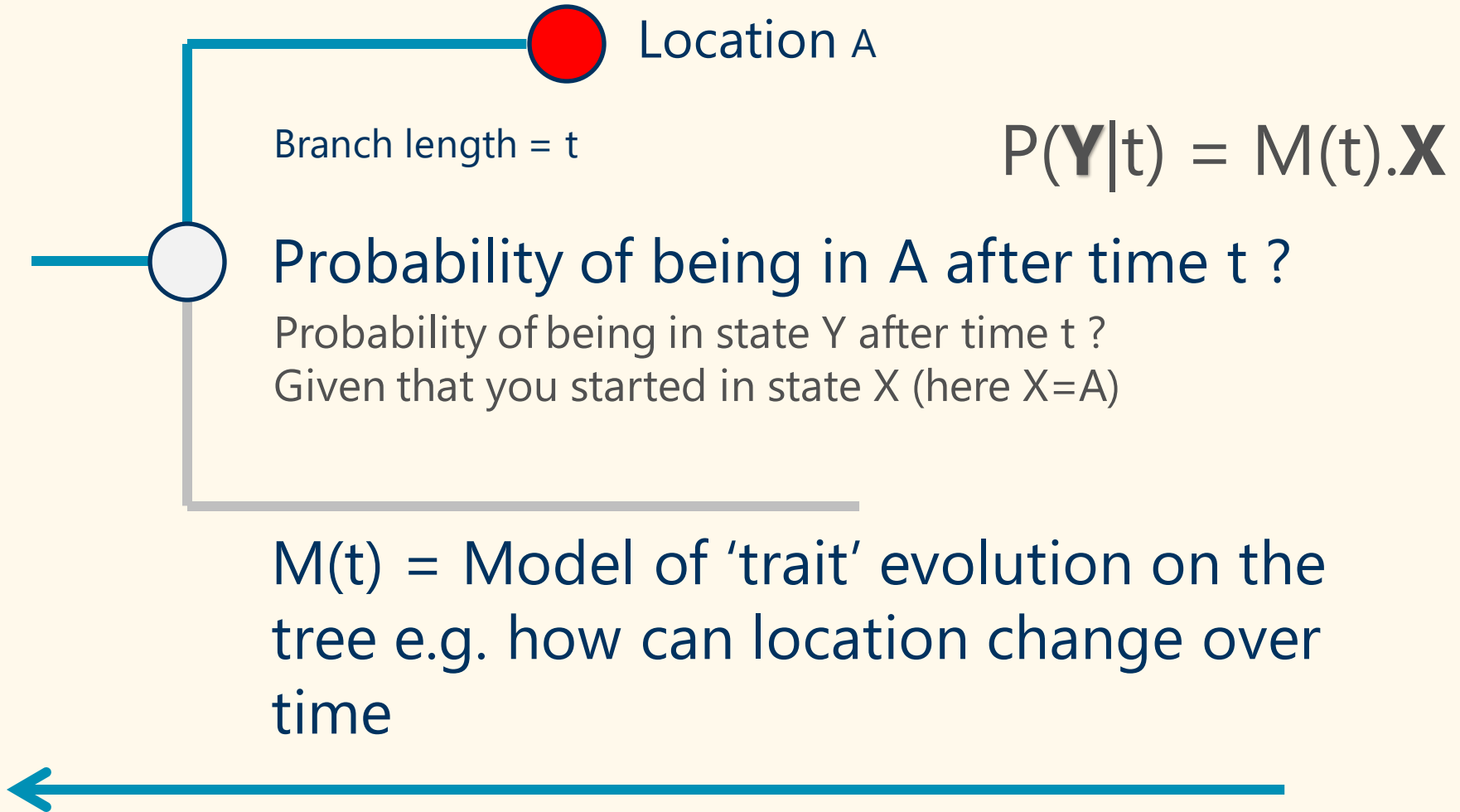
What can trees tell us ?

- Add traits to the tips;
- Infer ancestral states



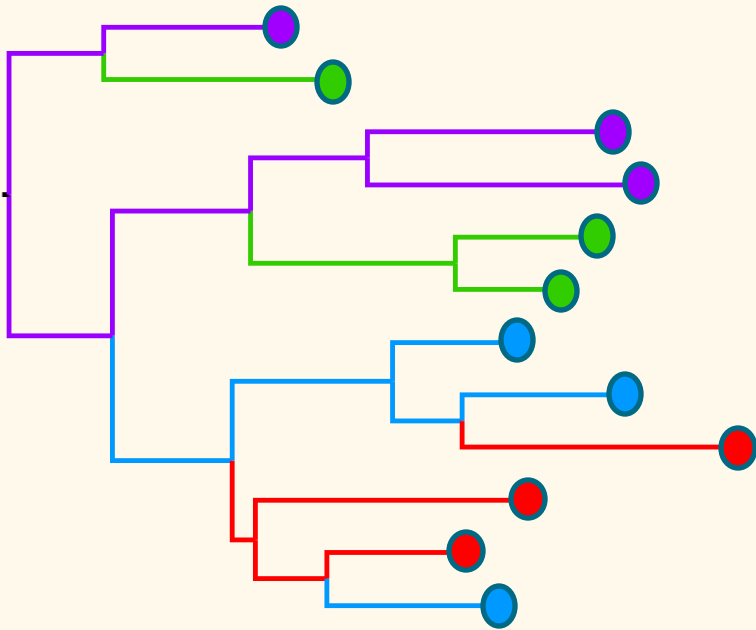
(decorated tree)

Focus on one branch



Discrete Trait Models

Tree with Location Traits



Transition Rate Matrix (M)

	A	B	C	D
A	-	A -> B	A -> C	A -> D
B	B -> A	-	B -> C	B -> D
C	C -> A	C -> B	-	C -> D
D	D -> A	D -> B	D -> C	-

Probability of Ancestral state (x'),
given branch length t and child state x :

$$p(x'|t) \sim e^{Mt}x$$

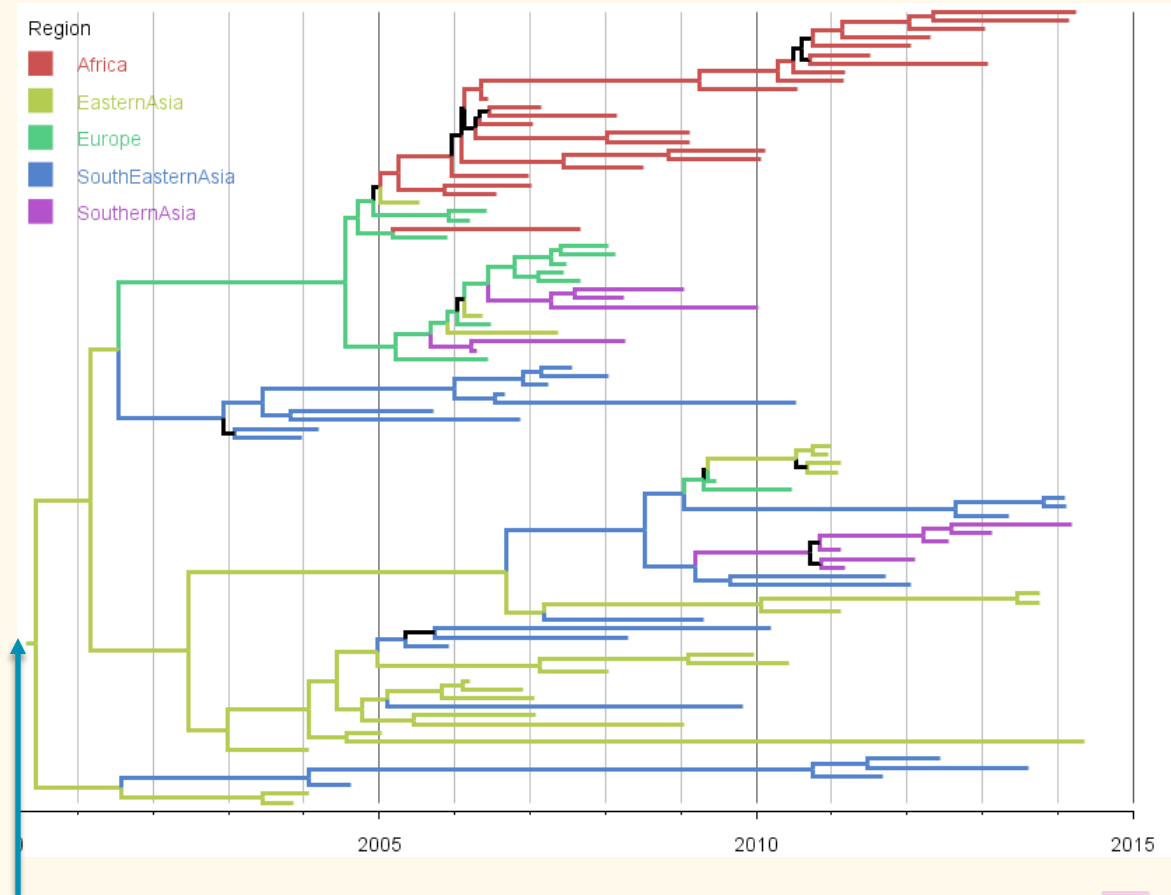


In a program,
e.g. BEAST !

- Add locations to phylogenetic tree
- Estimate transition rates between locations along branches
- Transmission pattern represented by rate matrix

What can trees tell us ?

- Add traits to the tips; infer ancestral states



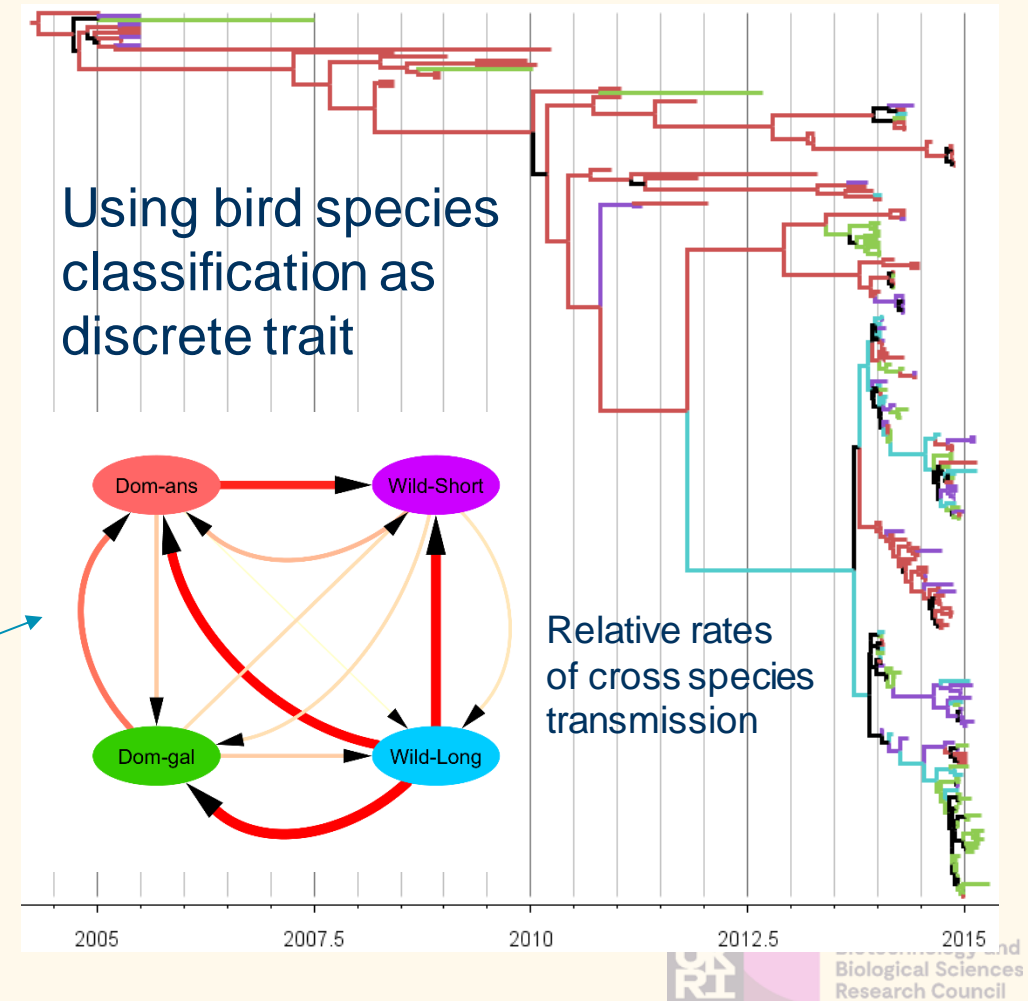
Epidemic started in Eastern Asia

What can trees tell us ?

- Add traits to the tips; infer ancestral states

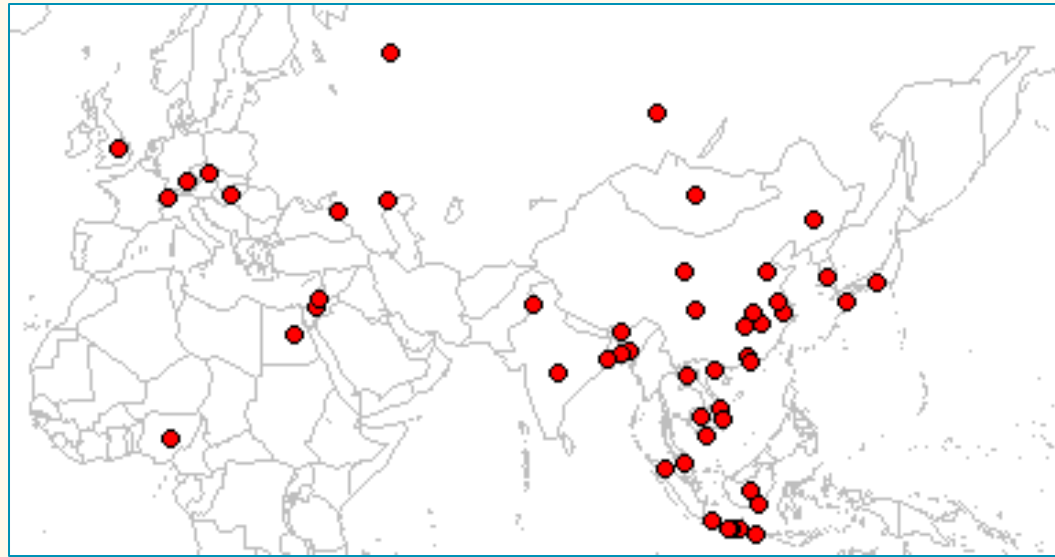


network figure version of the rate matrix of A->B etc
Arrow width = rate value
Arrow colour = importance measure of rate



Using Spatial Coordinates

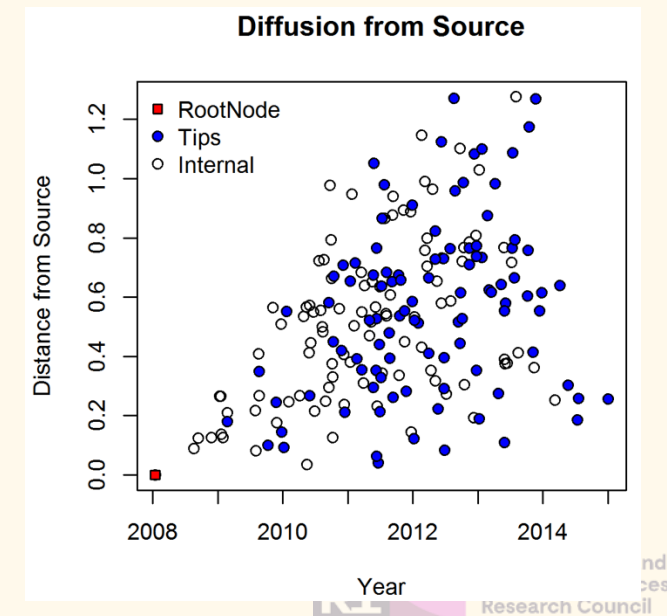
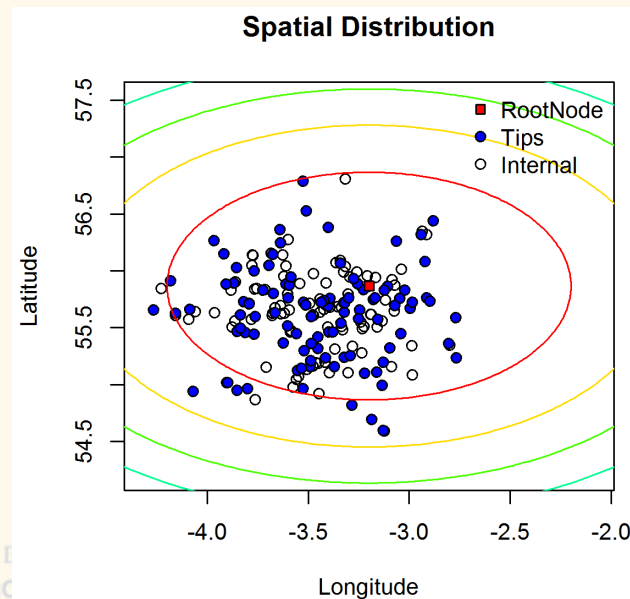
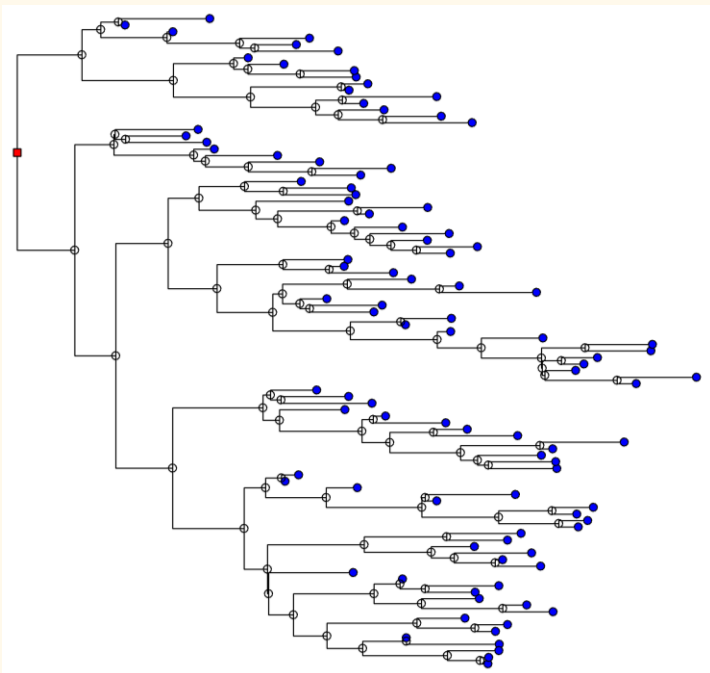
- Discrete traits models are good for when there are discrete populations
- However, the real distribution of samples may be more diverse



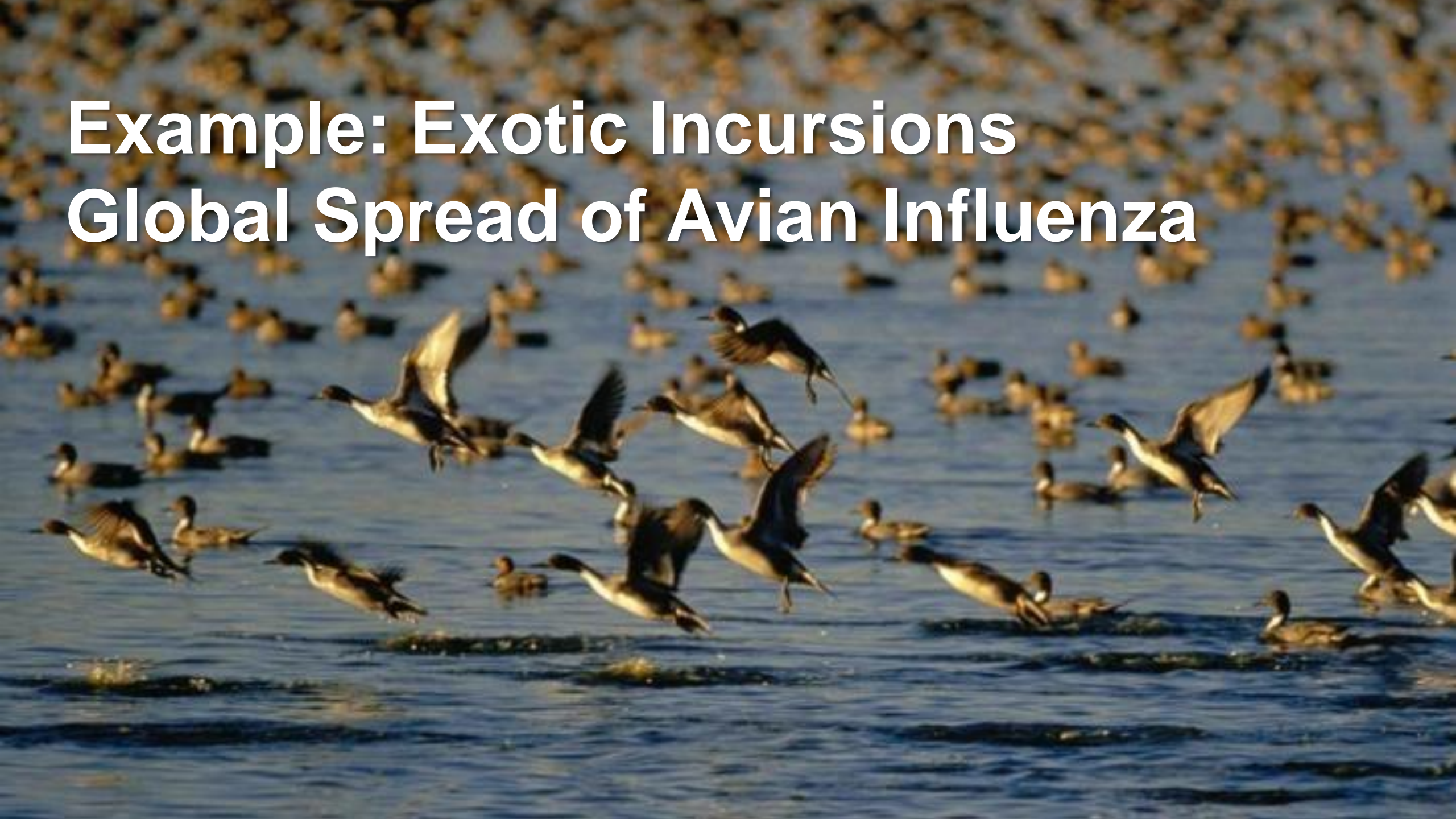
Actual distribution of H5N1 dataset

Spatial Diffusion

- Model the spatial coordinates as continuous traits on the tree
- Viral lineages “diffuse” from a point source
 - Distance of child node is expected diffusion distance from parent node assuming time t has elapsed
 - Uses Brownian motion (random walk) diffusion model and extensions

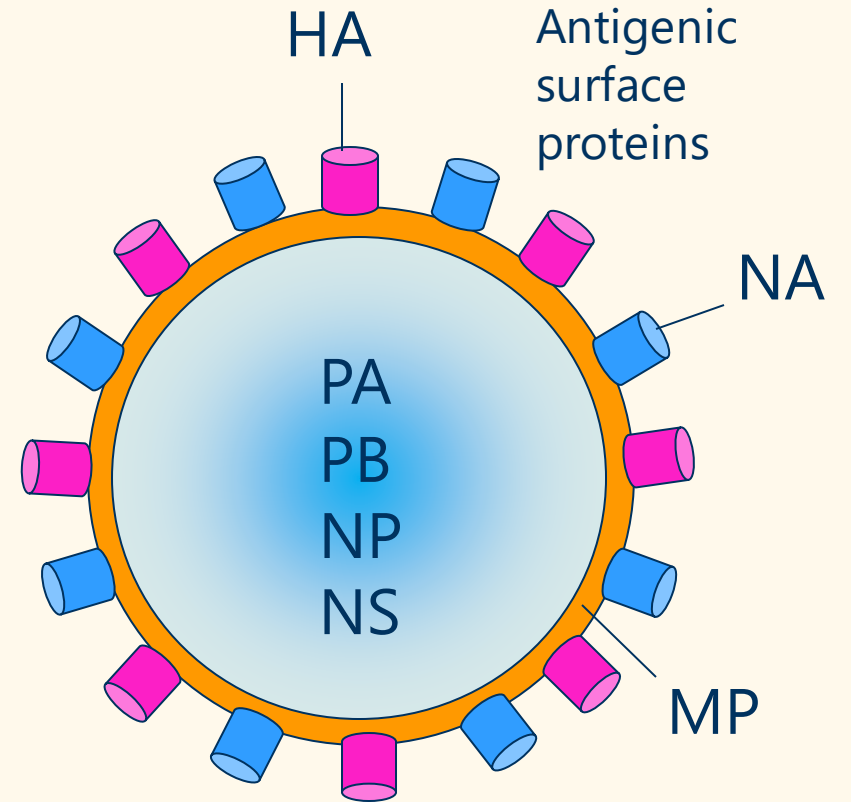


Example: Exotic Incursions Global Spread of Avian Influenza



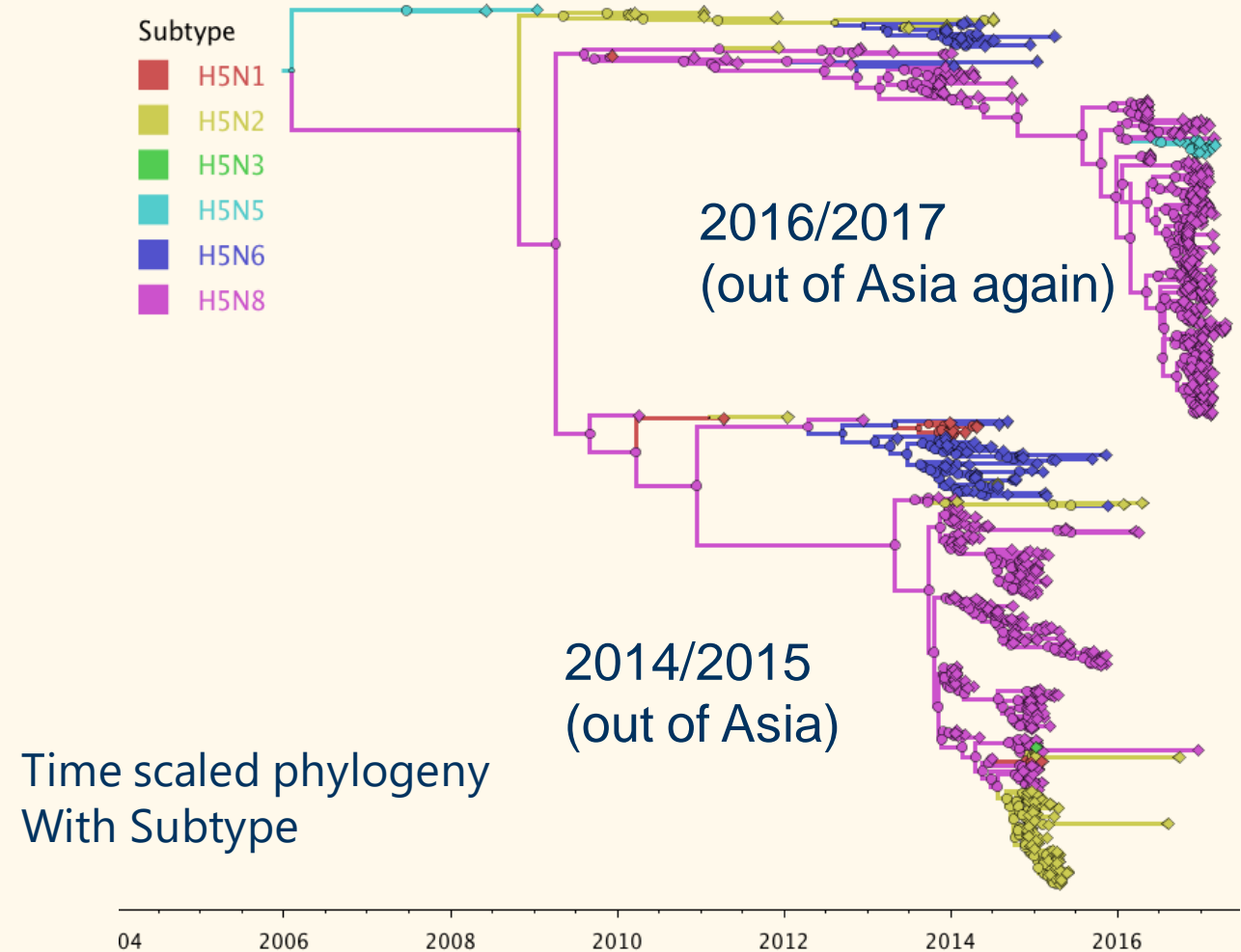
Influenza viruses

- 8 RNA Segments coding for 10+ proteins
- Virus subtype defined by surface proteins
 - Hemagglutinin (HA)
 - Neuraminidase (NA)
- Reassortments between all segments
- Substitution rate $\sim 5 \times 10^{-3}$ per site per year (10-70 nucleotides per year across genome)



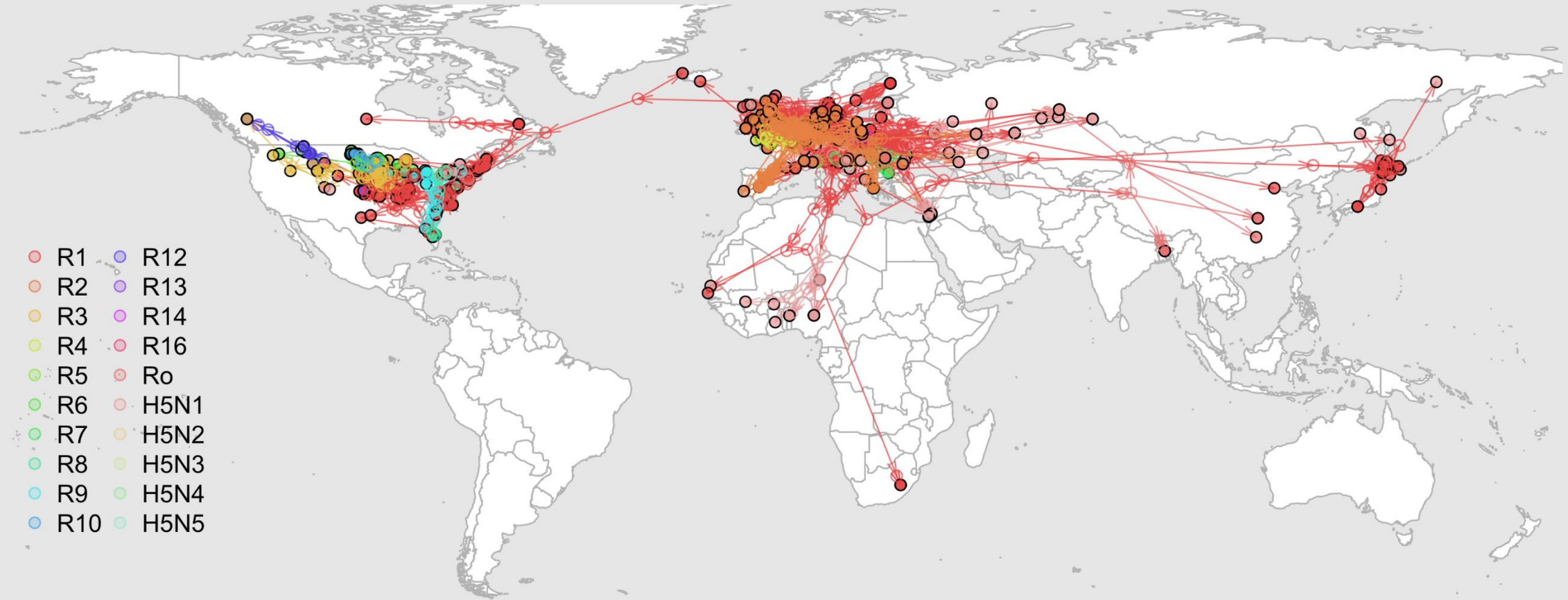
Avian Influenza Phylodynamics

- H5 + multiple subtypes (H5NX) since 2005
- Outbreaks in 2014/2015 in Asia, Europe, North America
- Outbreaks in 2016/2017 in Asia and Europe
- Data from 16 countries in Global Consortium
- Then Autumn Winter 20/21/22
- **AGAIN IN UK, EUROPE, AMERICAS NOW – 2022/2023**



Paper: “Role for migratory wild birds in the global spread of avian influenza H5N8”
Science 14 Oct 2016 “The Global Consortium for H5N8 and Related Influenza Viruses”:
 Samantha Lycett, Rogier Bodewes, (many authors), Mark Woolhouse, Thijs Kuiken.

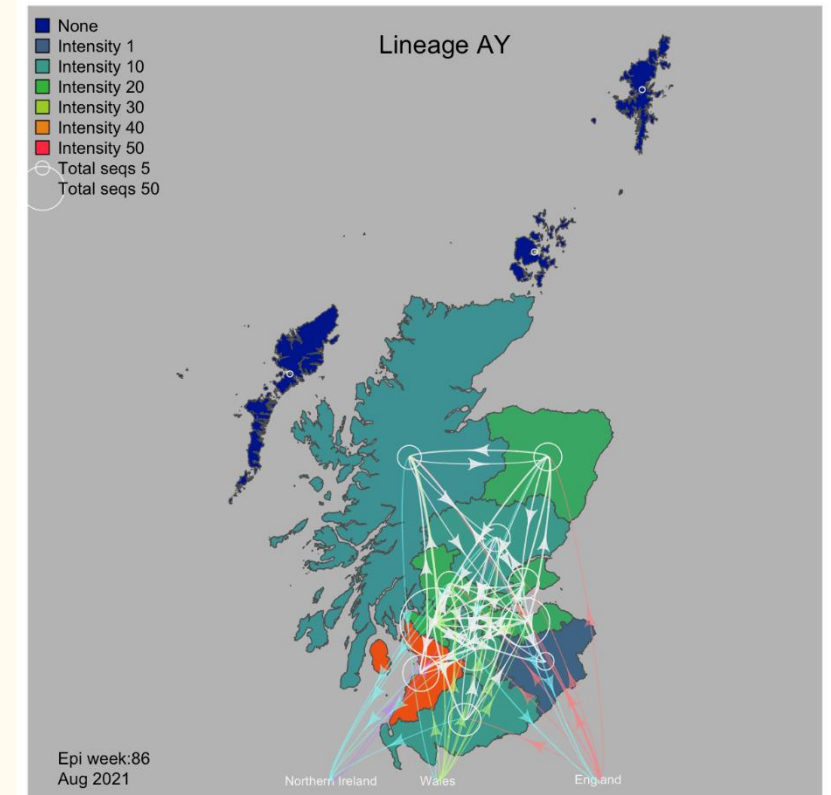
Reassortants on Map (Global) 2020-2022 data



- H5N1 data, but colours are by internal segment Reassortment code

Summary & Comments

- Tracking spread of infection using pathogen sequences and phylodynamics
 - Near real time surveillance sequencing
 - Global surveillance and data sharing important
 - Applicable to many measurably evolving systems: Influenza, SARS-CoV-2, FMDV..
- To make predictions or forecasts for viral spread in populations:
 - where are the current infections ? – surveillance & field
 - Imports and spreading patterns in the population ?
 - Fitness of (new) variants ? - Integrate experimental results with population scale growth rate estimates



Example of SARS-CoV-2 Delta AY lineages in Scotland calculated using whole genomes, time-scaled trees and discrete trait models

Arrow width: number of imports/exports in 7 days
Circle size: max sequences of AY Lineages in 7 days.

Background colour: "intensity" within healthboard transmissions

A flock of approximately 12 birds, likely swallows, are captured in flight against a clear, pale blue sky. They are arranged in a loose V-formation, with some birds leading and others following. The birds have dark wings and lighter-colored bodies. The text "Thank you !" is centered in the upper half of the image.

Thank you !

samantha.lycett@ed.ac.uk