

Phylodynamics Workshop

Dr Samantha Lycett



Infection & Immunity
Roslin Institute
University of Edinburgh

Feb 2024



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	Y	N	T	I	Q	N	N	F	T	N	V	I	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	

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, HCV
Foot-and-Mouth
Coronavirus

Segmented ssRNA

Influenza



THE UNIVERSITY of EDINBURGH
Royal (Dick) School of
Veterinary Studies

Mpox,
African Swine Fever

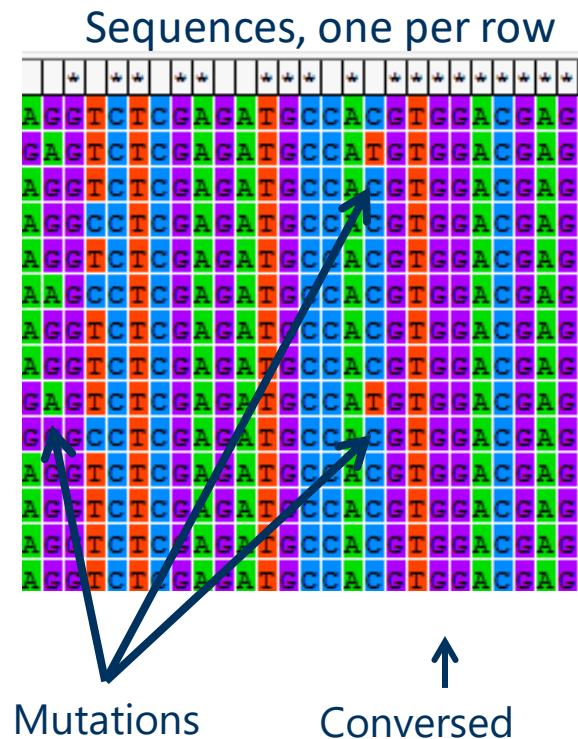
Segmented dsRNA

Blue Tongue

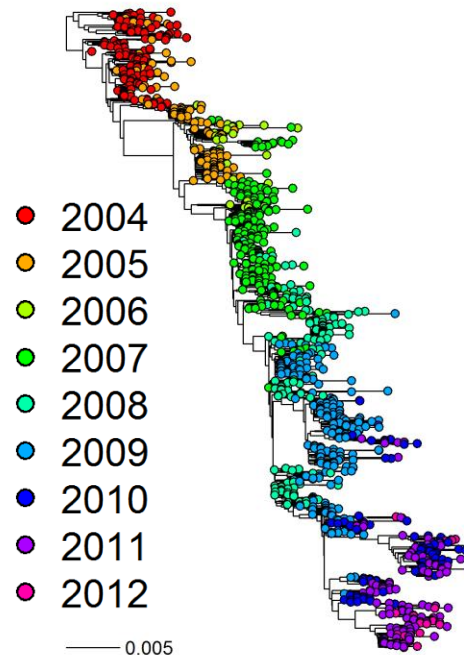
Bovine Tb

Pathogen Sequence Data

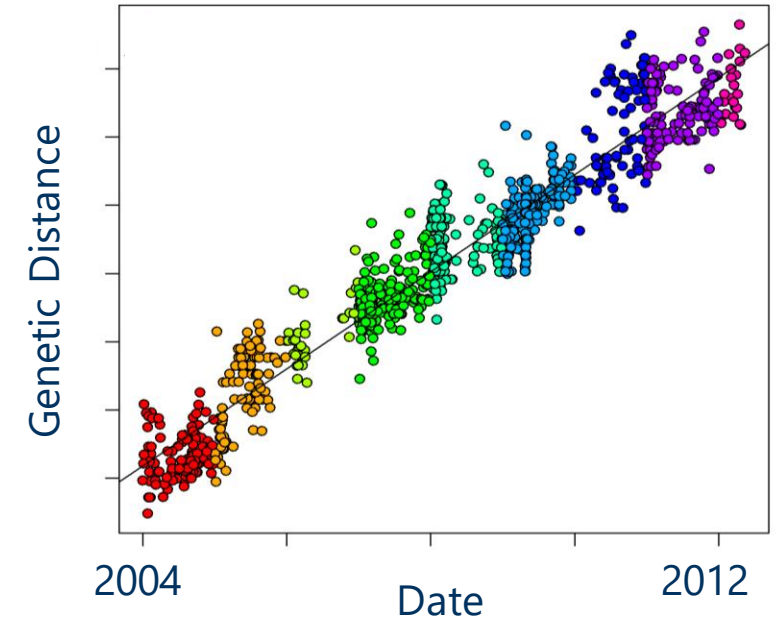
- Pathogen sequences provide richer information than strain type
- Sequences accumulate mutations over time



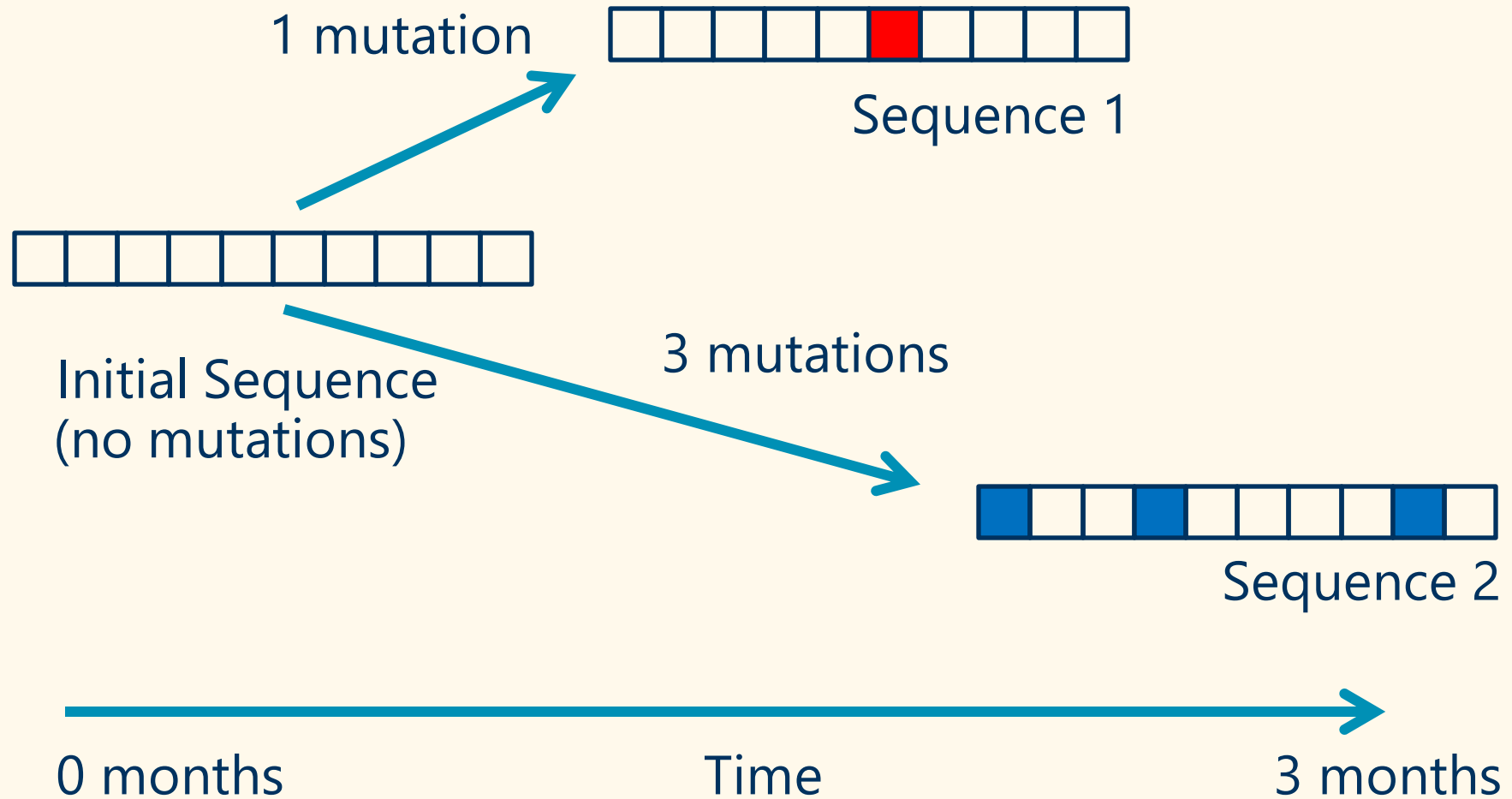
Tree of Human Influenza



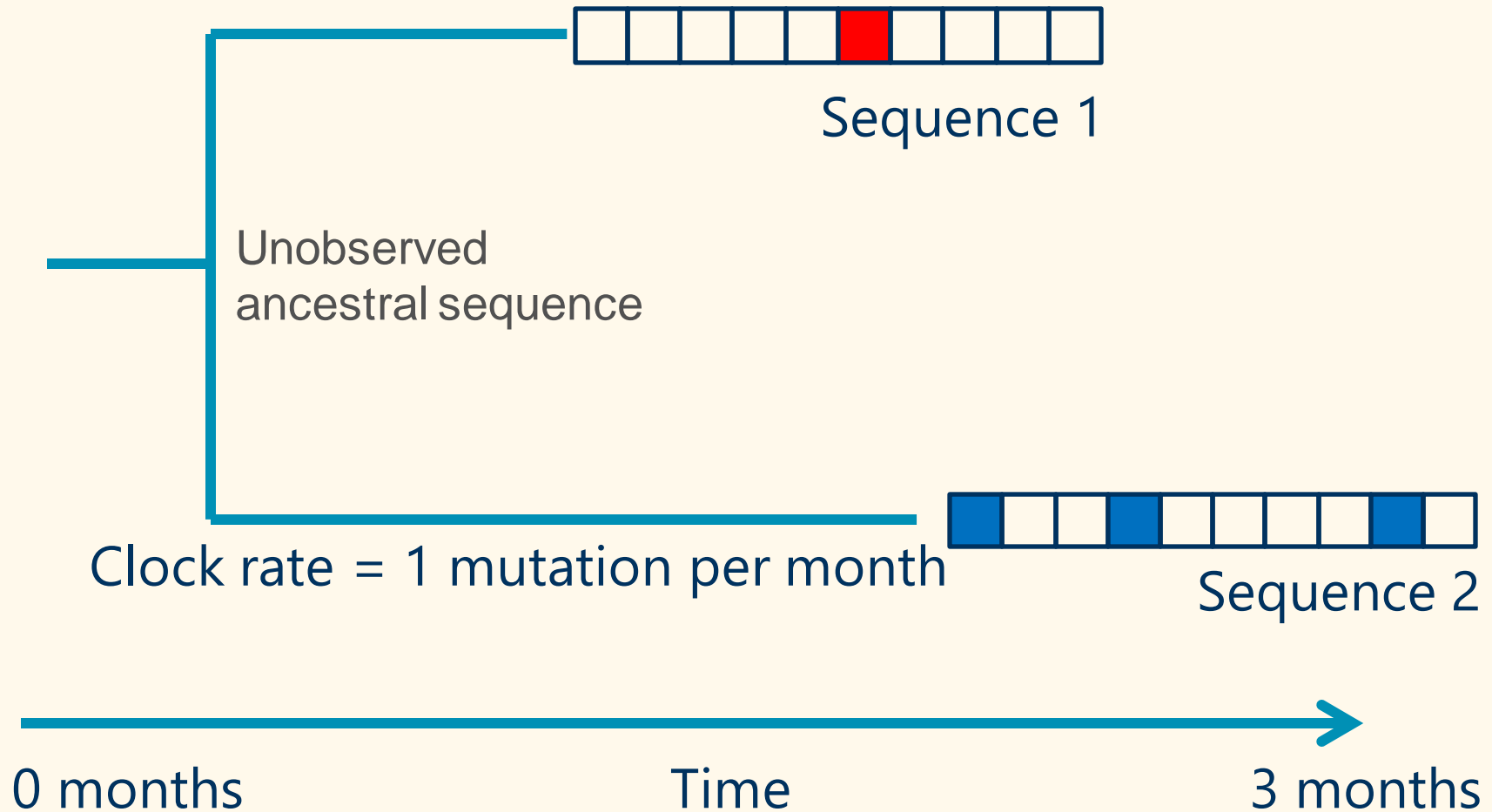
Genetic Distance from Root



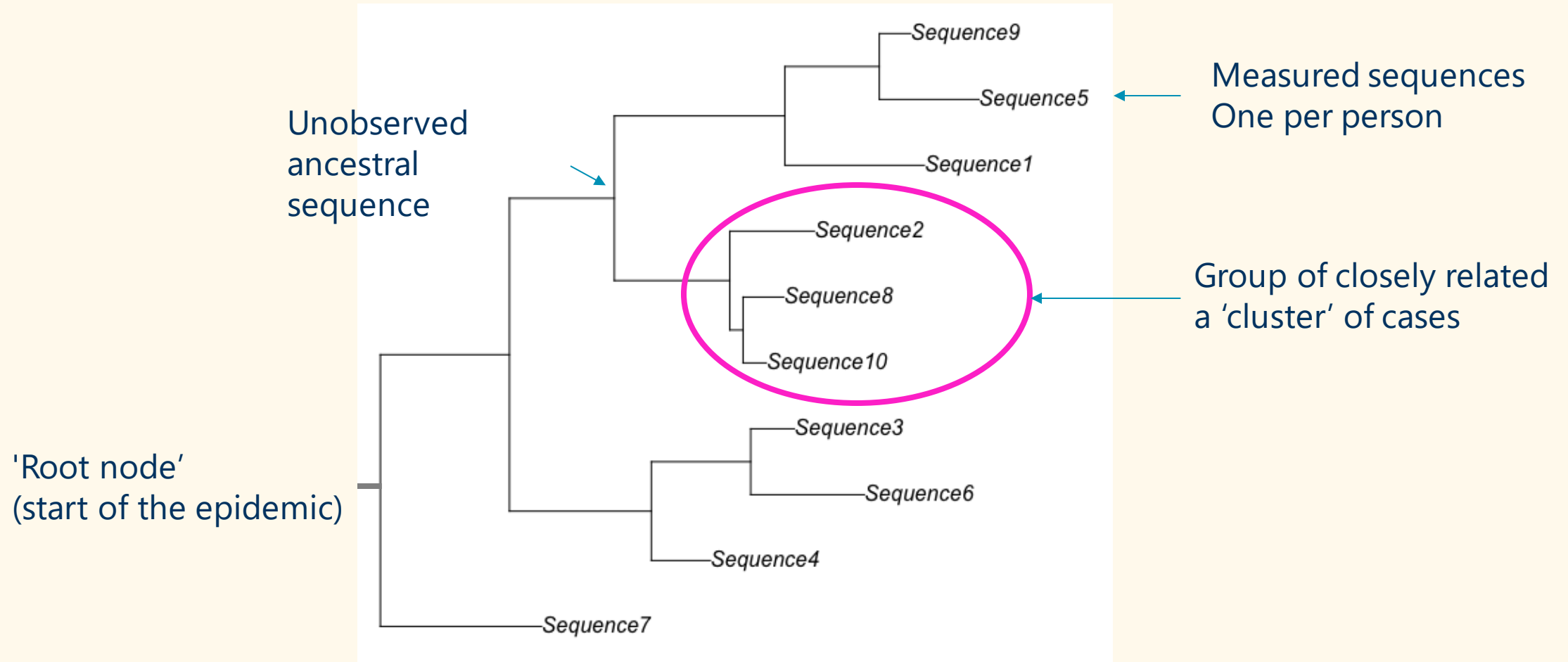
How are the sequences related ?



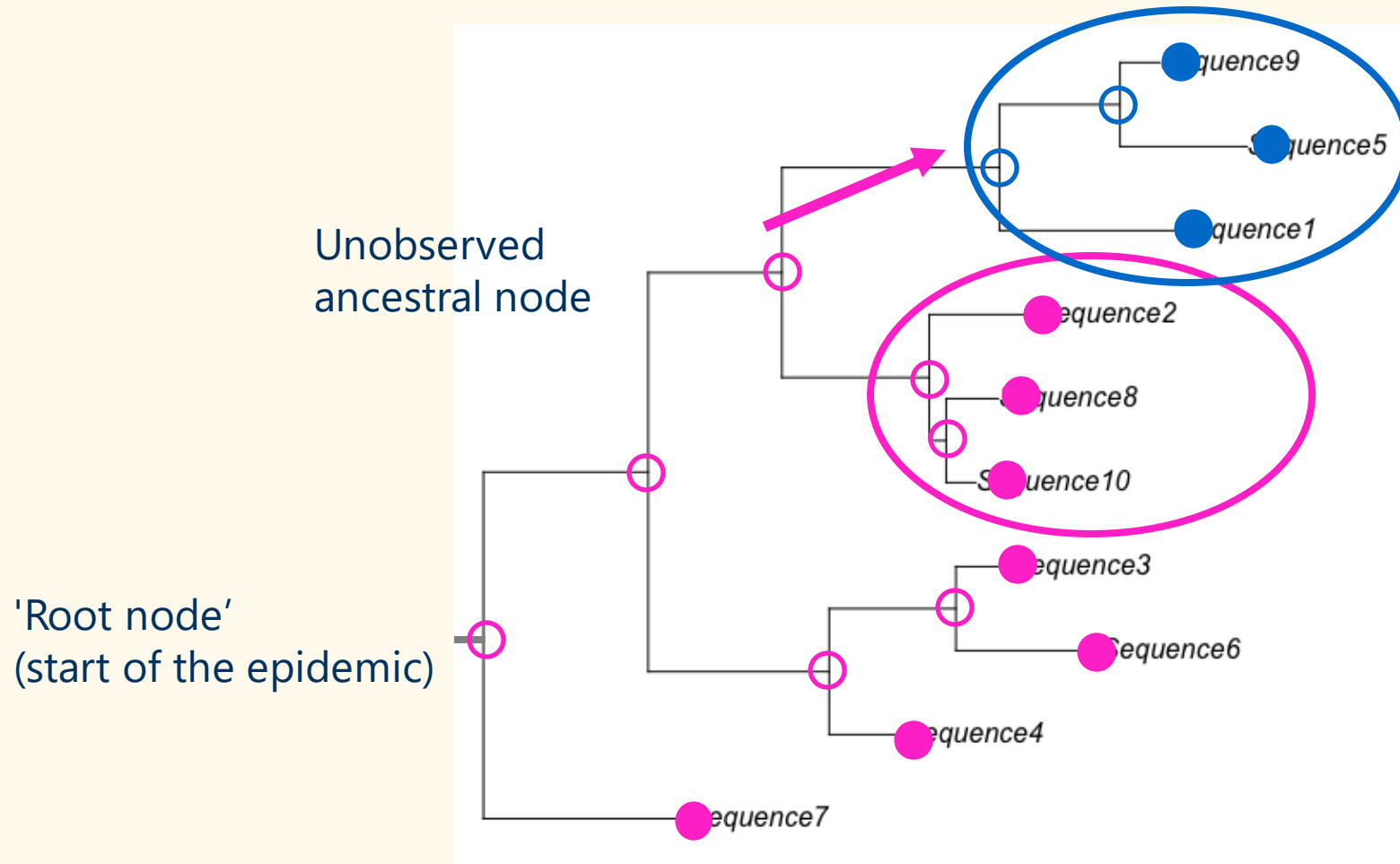
Adding a time scale and creating a Phylogenetic tree



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

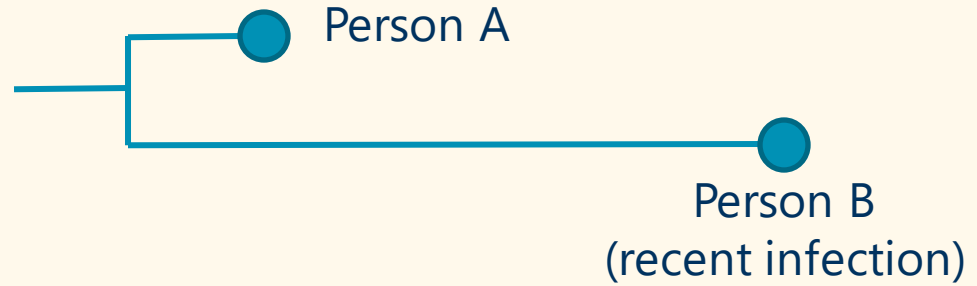
- For outbreak situations
- What strain is it ?
- When did it arrive ?
- Who infects whom ?
- Dense sampling required

Transmission Patterns

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

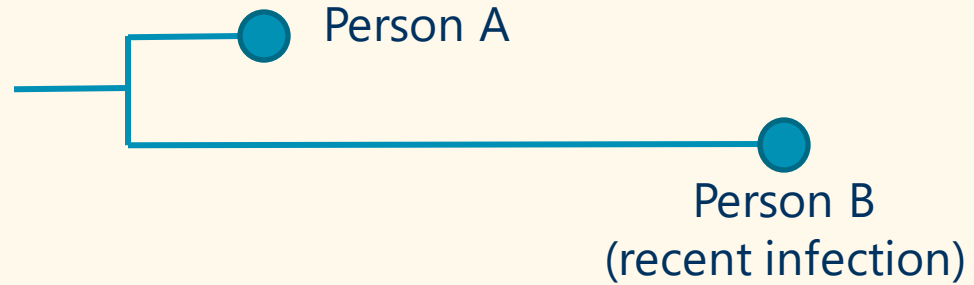
Detailed Forensic scenario - Can you tell who infected who ?

- Yes ?

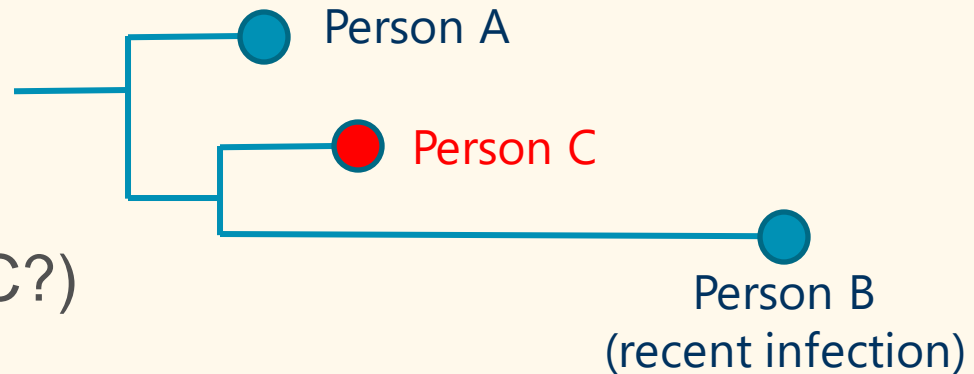


Detailed Forensic scenario - Can you tell who infected who ?

- Yes ?

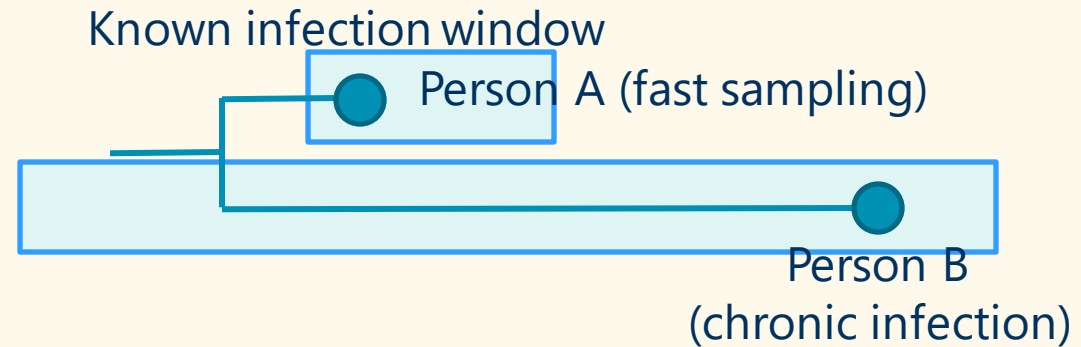


- No
- Needed to sample C
- (did you know about C?)



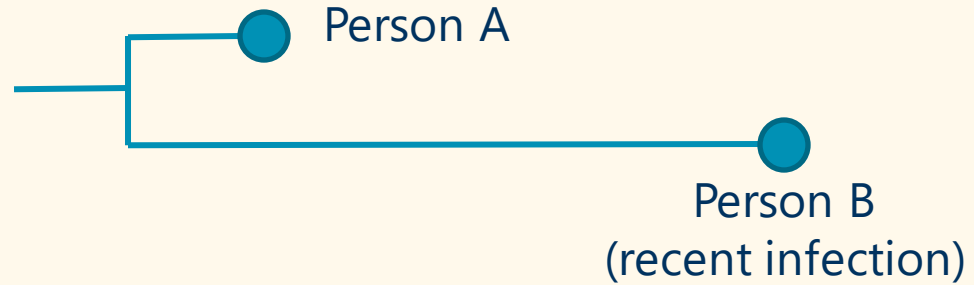
Detailed Forensic scenario - Can you tell who infected who ?

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

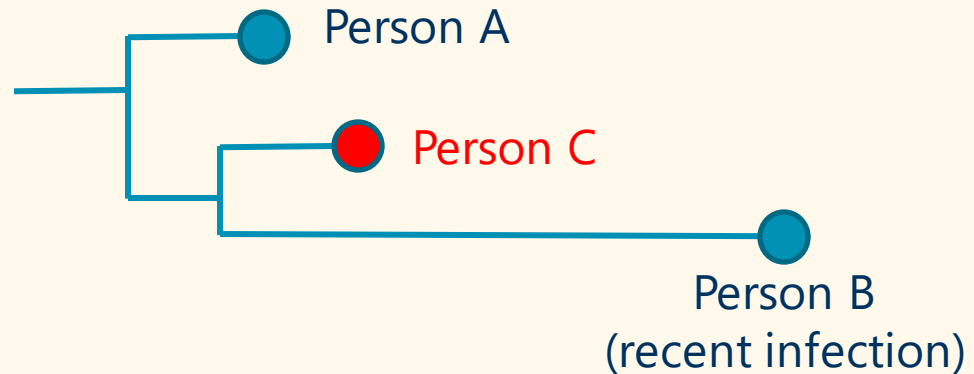


Detailed Forensic scenario - Can you tell who infected who ?

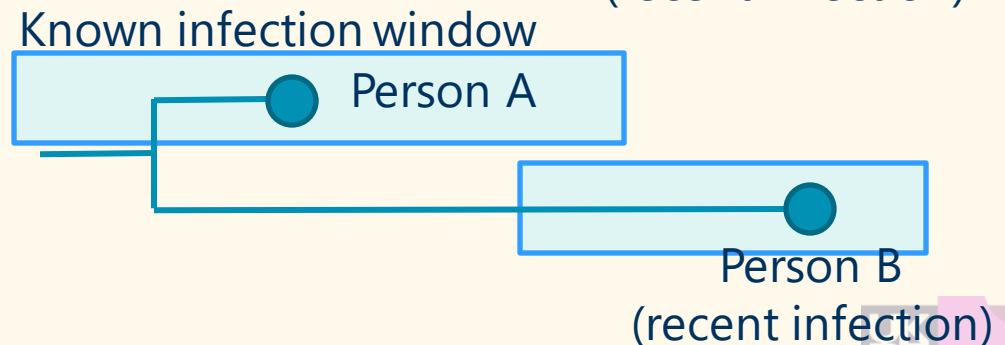
- Yes ?



- No

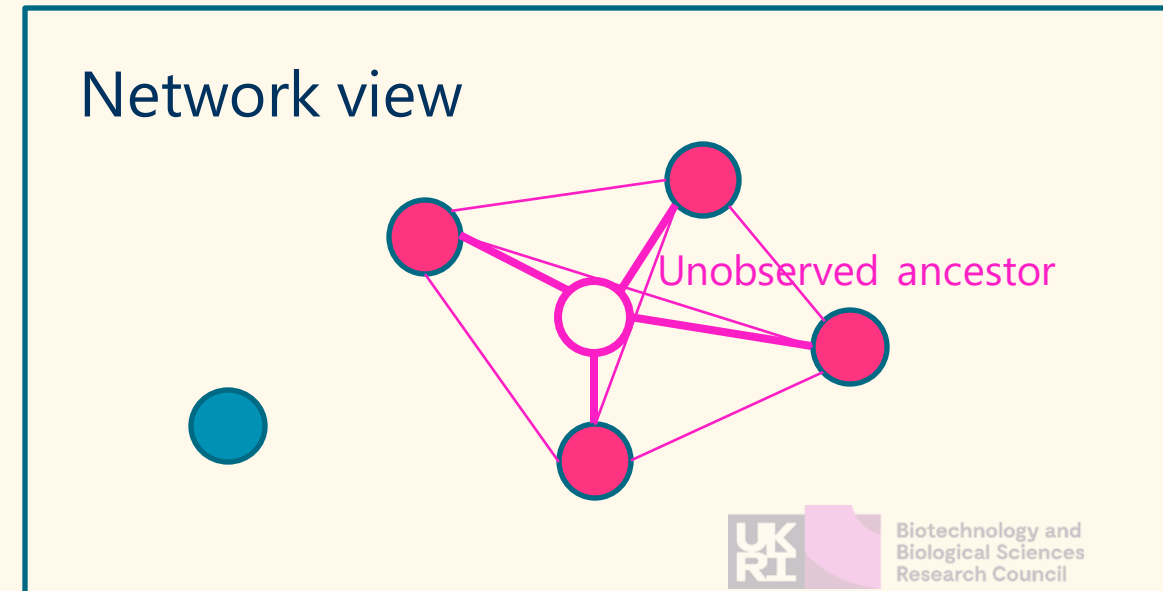
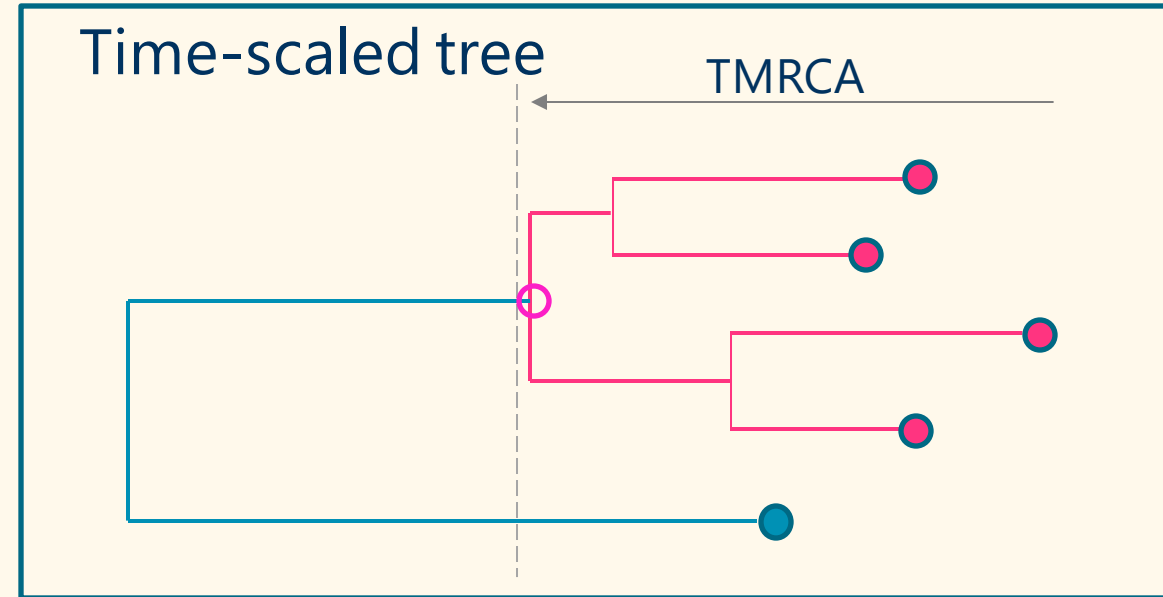


- Maybe ?



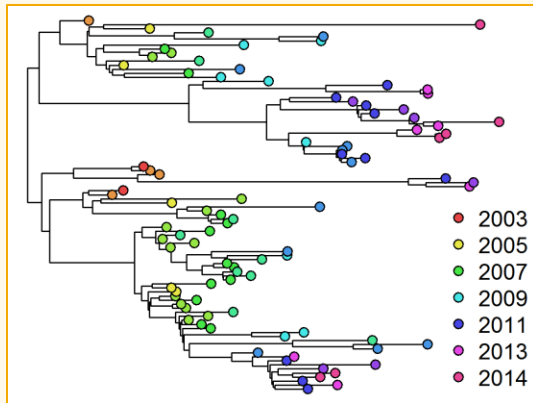
Genetic Clusters

- 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

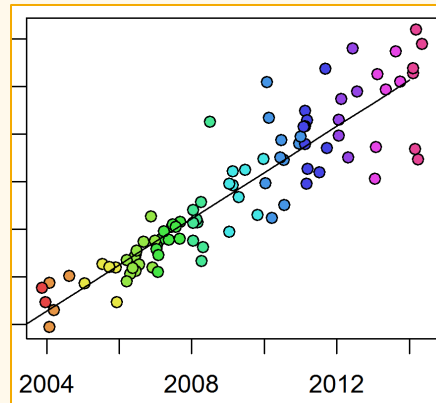


Pathogen Phylodynamics

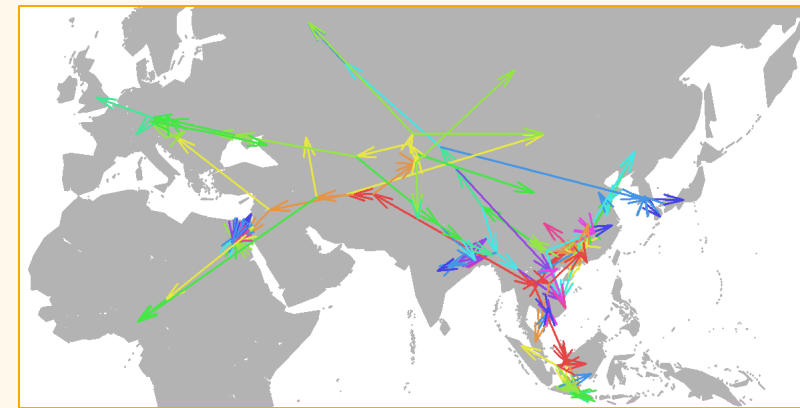
Harness pathogen sequence and disease surveillance data



Phylogenetic tree



Molecular Clock



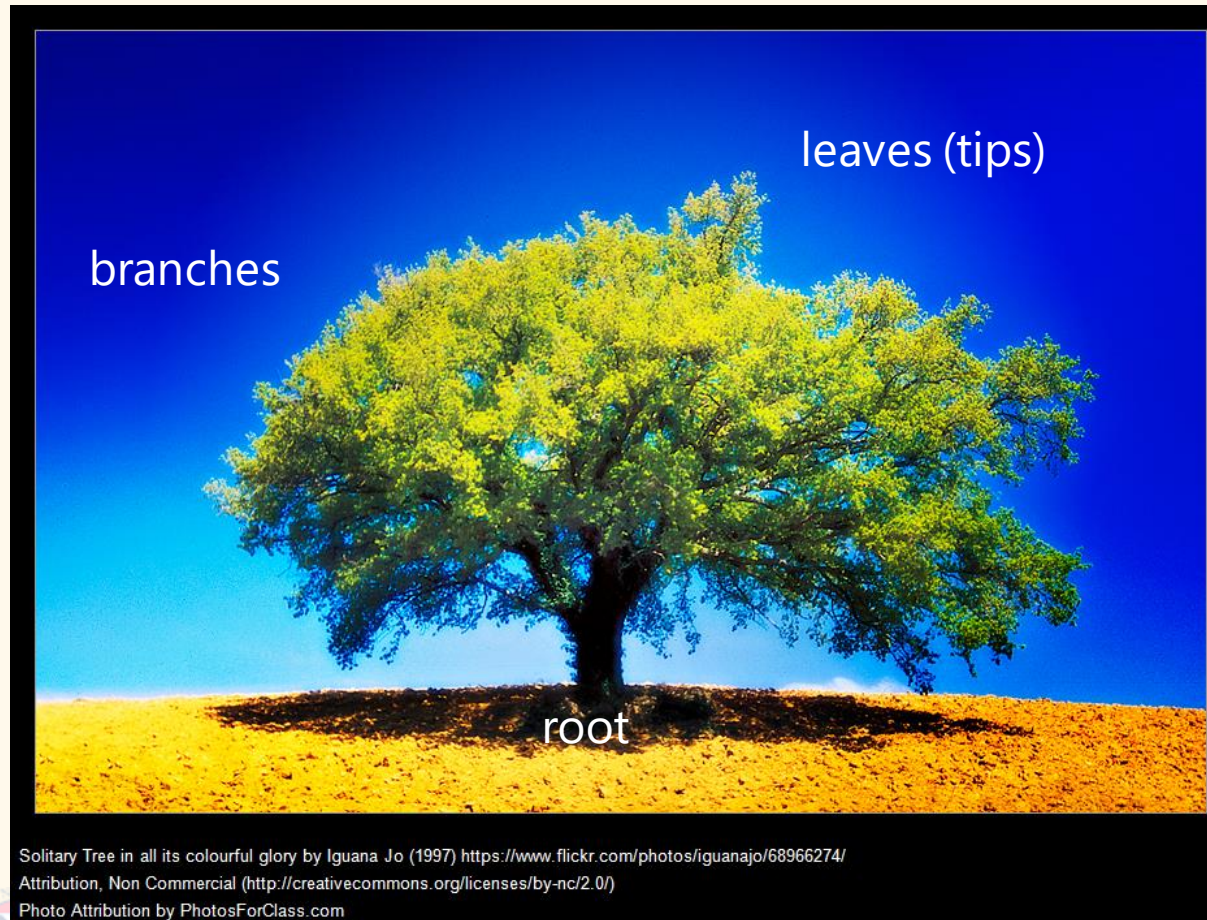
Tree + Spatial model

- 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..**

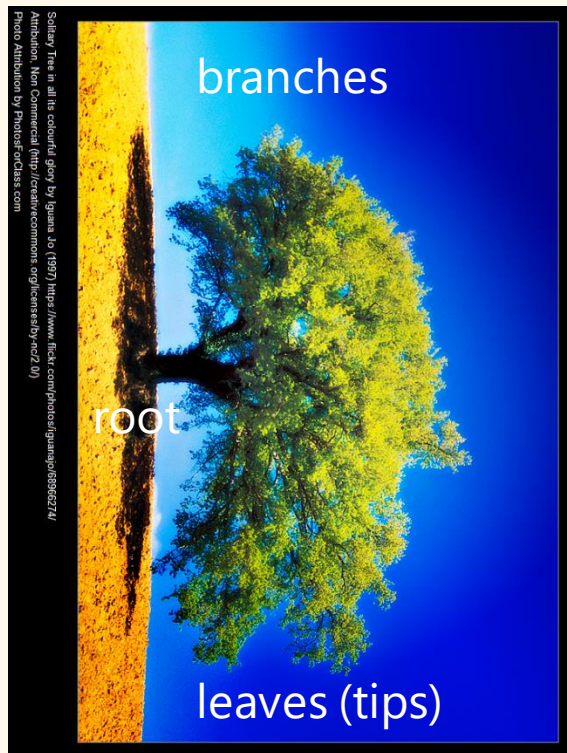
What can trees tell us ?

- A “plain” tree



What can trees tell us ?

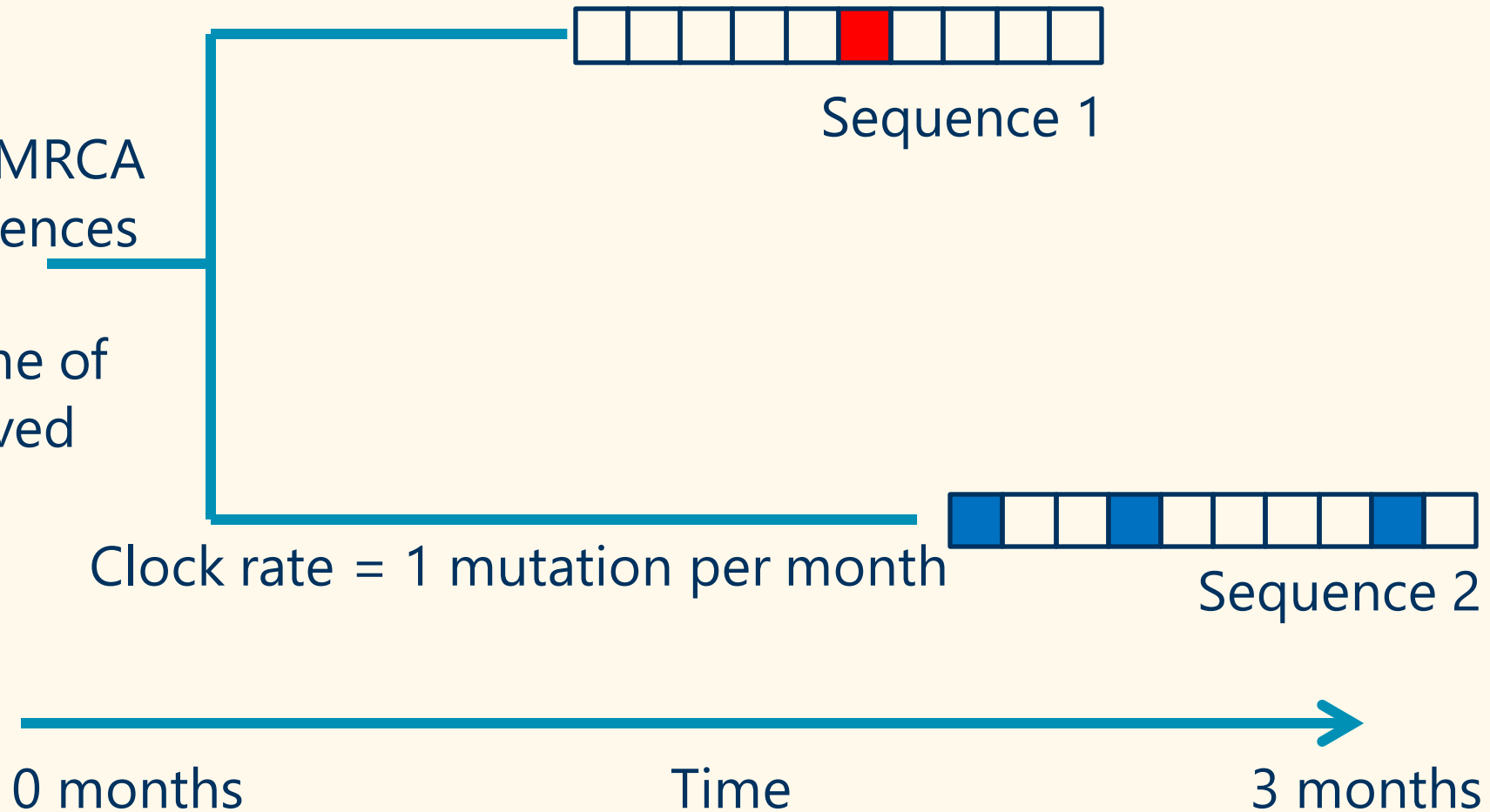
- A “plain” tree – timescale by molecular clock



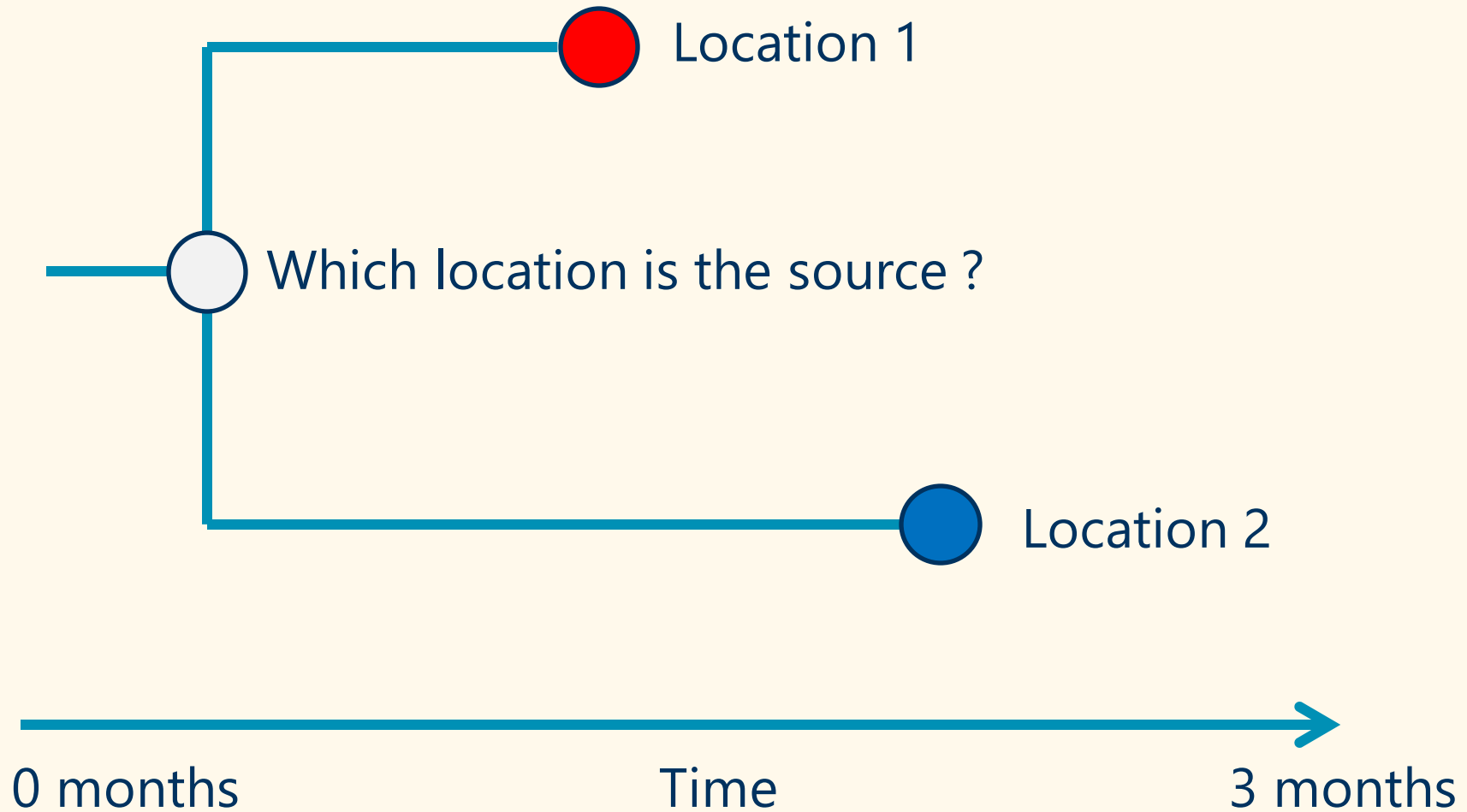
**Time to most recent common ancestor
(root age)**

Adding time scales to Trees (2)

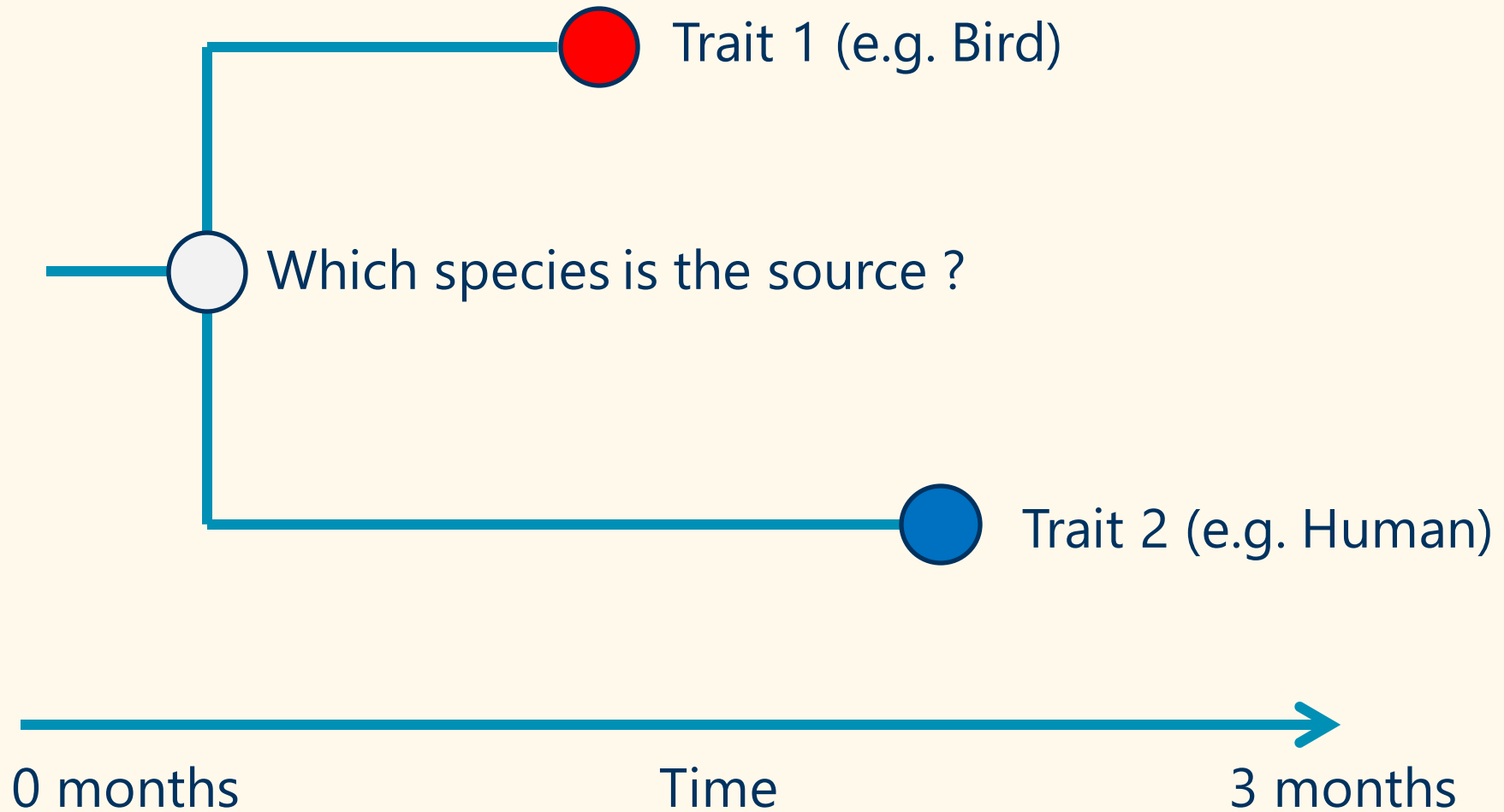
What is the TMRCA
of these sequences
?
(this is the time of
the un-observed
source)



Adding Locations to Trees



Adding other 'Traits' to trees



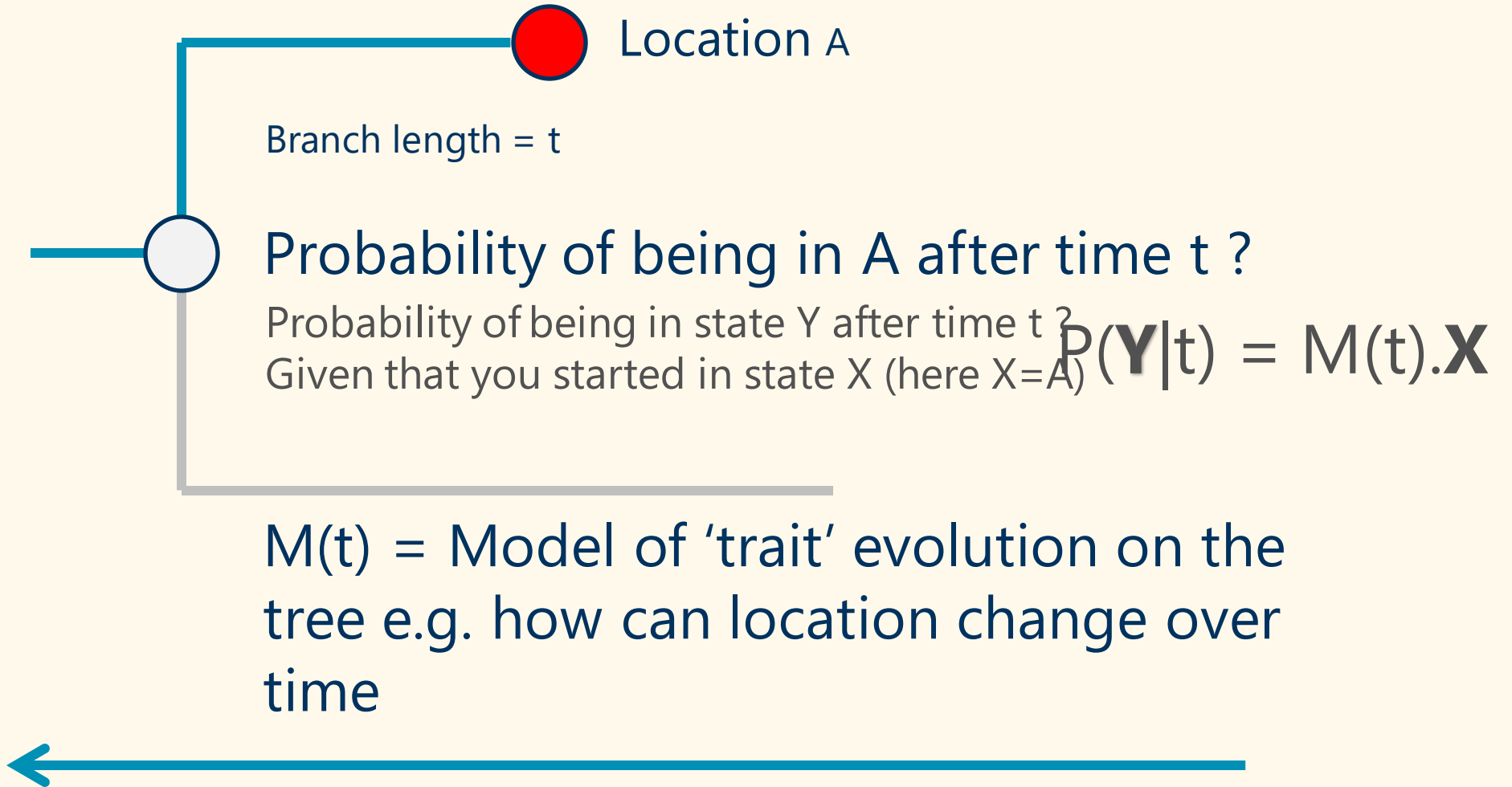
What can trees tell us ?

- Add traits to the tips;
- Infer ancestral states



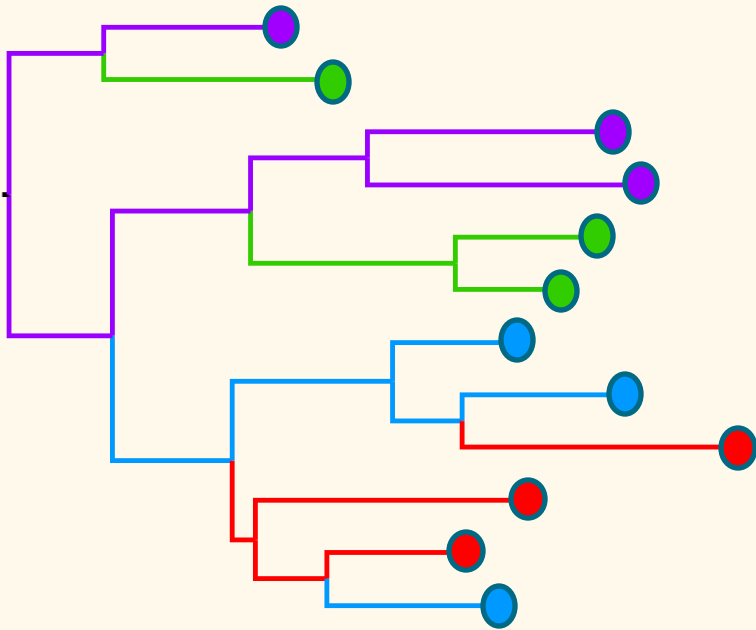
(decorated tree)

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 :

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

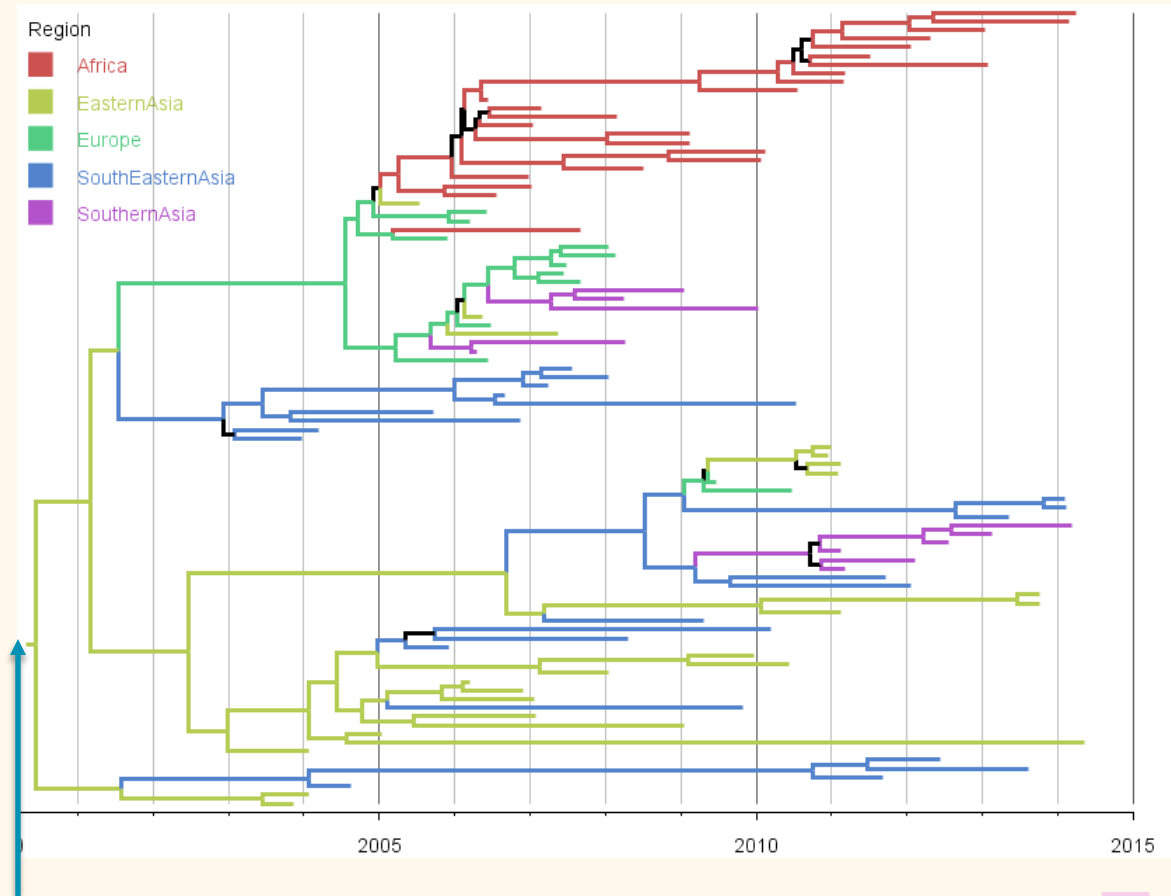


In a program,
e.g. BEAST !

- 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



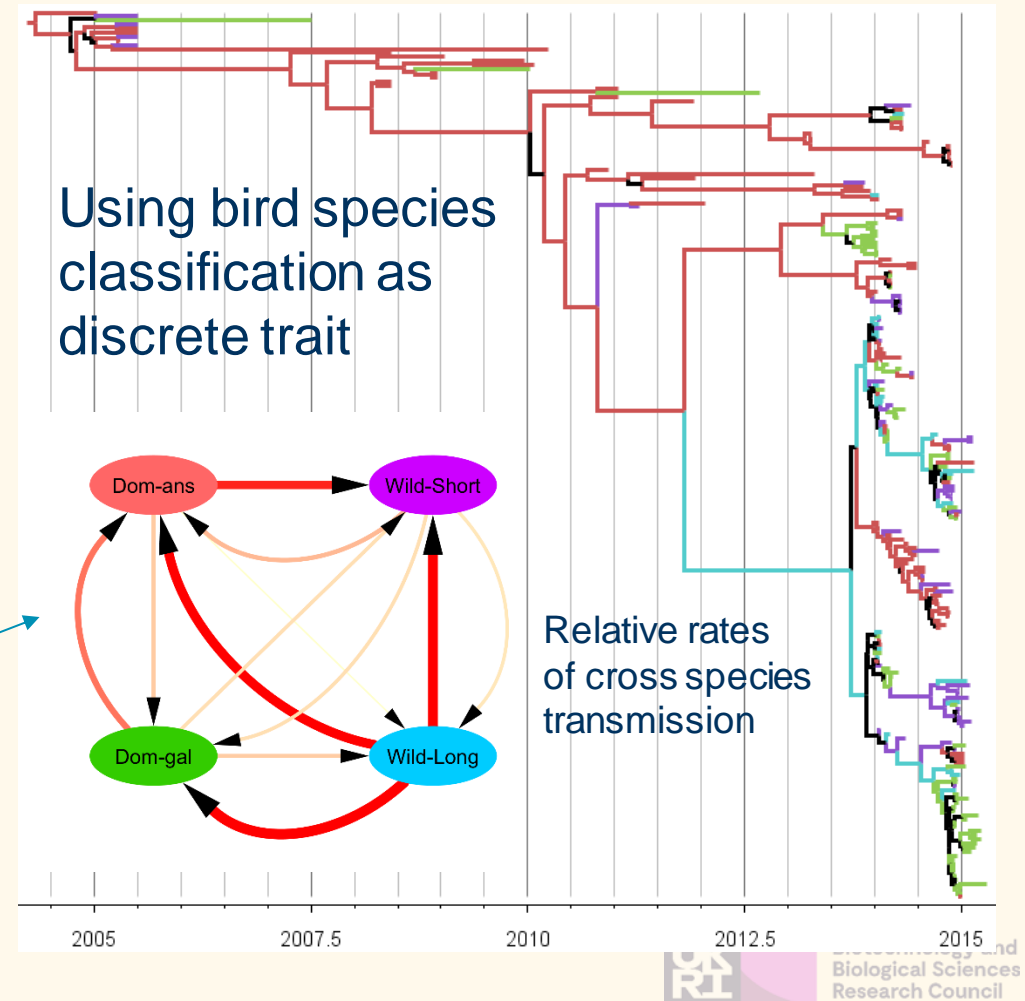
Epidemic started in Eastern Asia

What can trees tell us ?

- Add traits to the tips; infer ancestral states

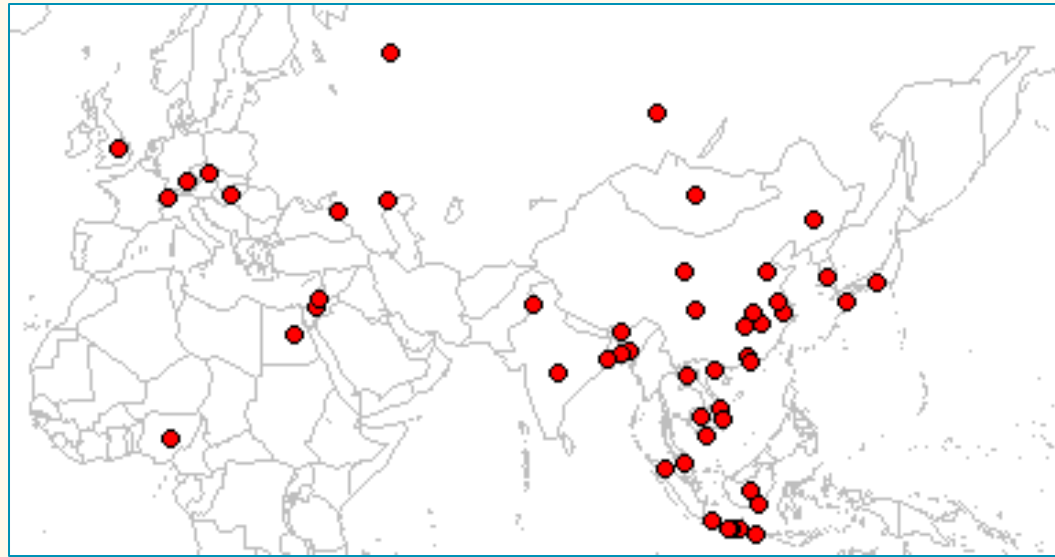


network figure version of the rate matrix of A- \rightarrow B etc
Arrow width = rate value
Arrow colour = importance measure of rate



Using Spatial Coordinates

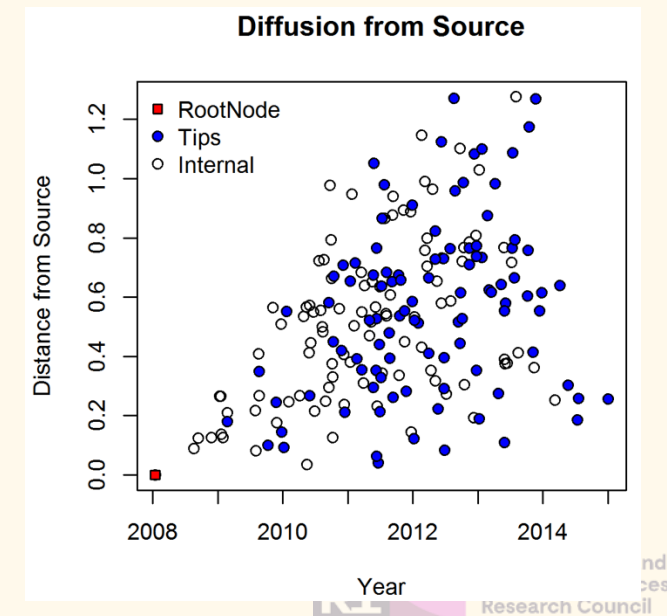
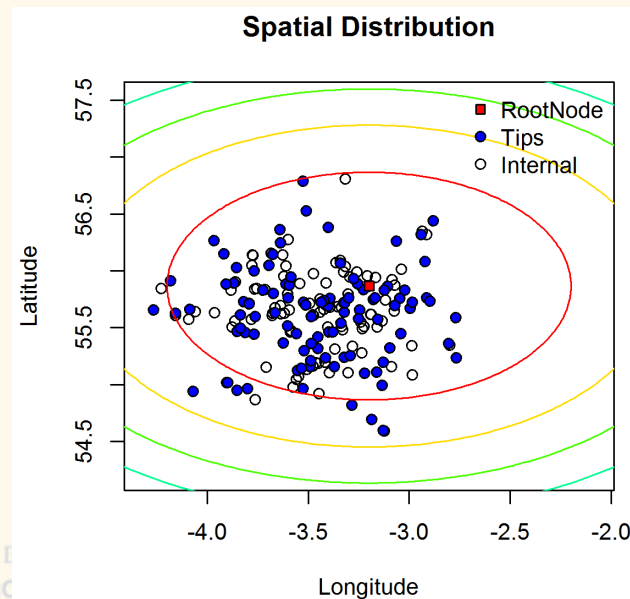
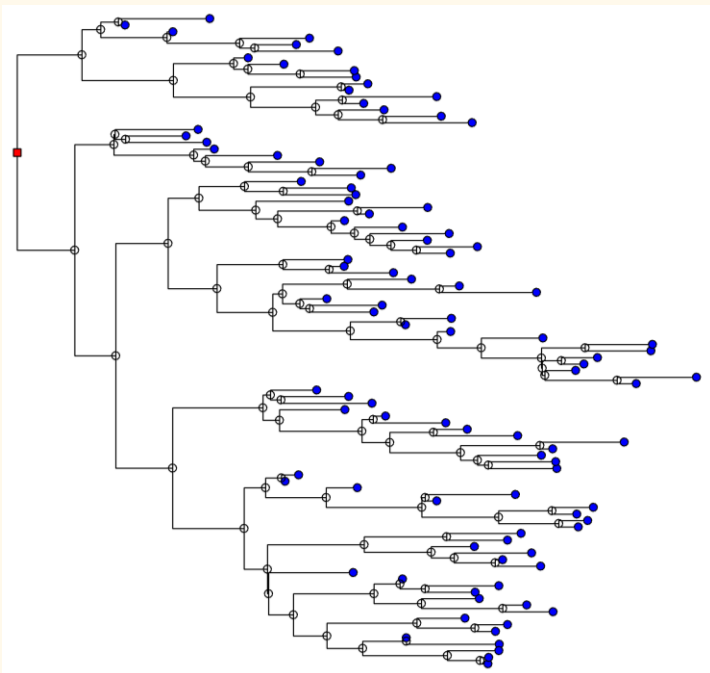
- Discrete traits models are good for when there are discrete populations
- However, the real distribution of samples may be more diverse



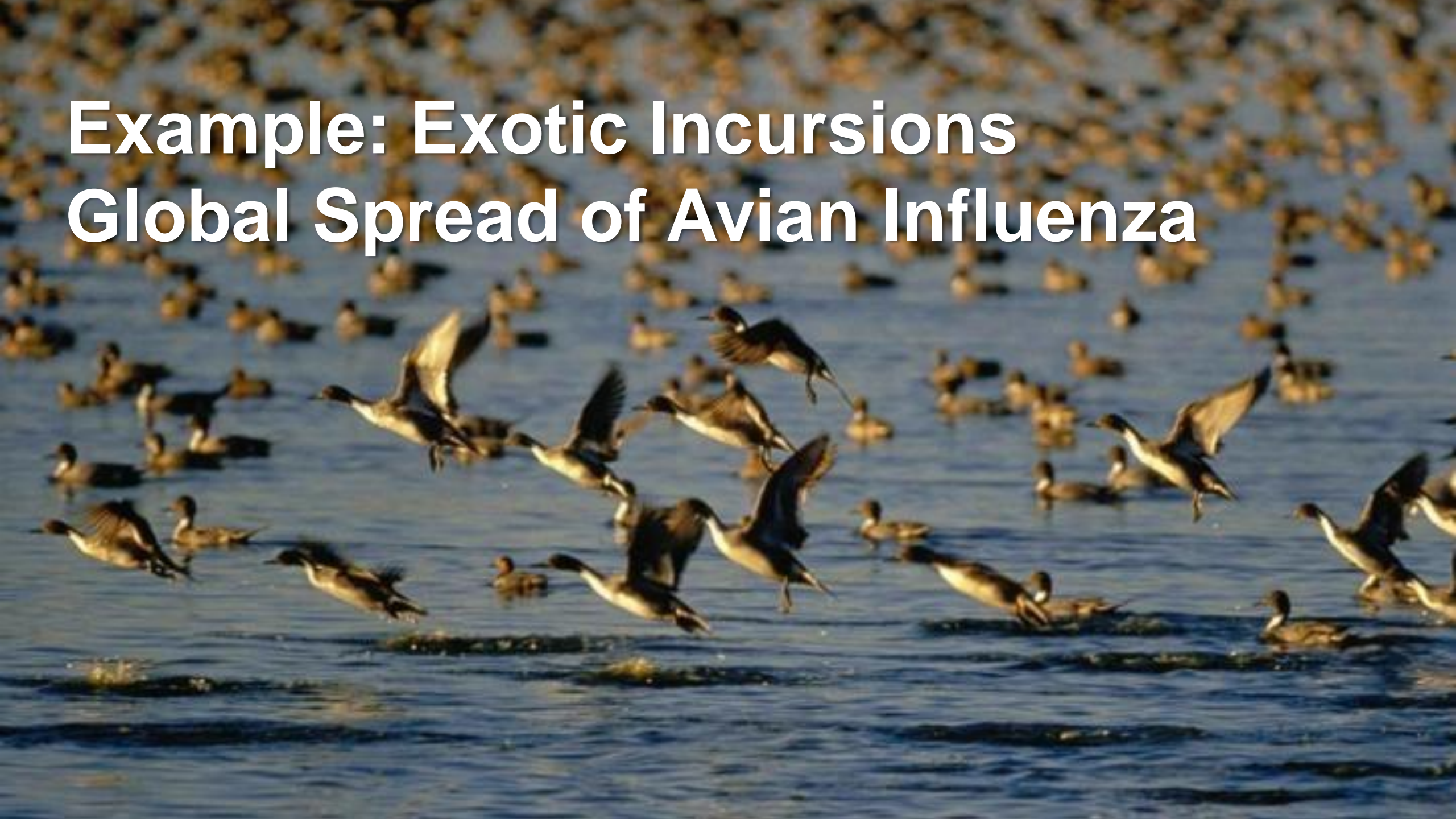
Actual distribution of H5N1 dataset

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

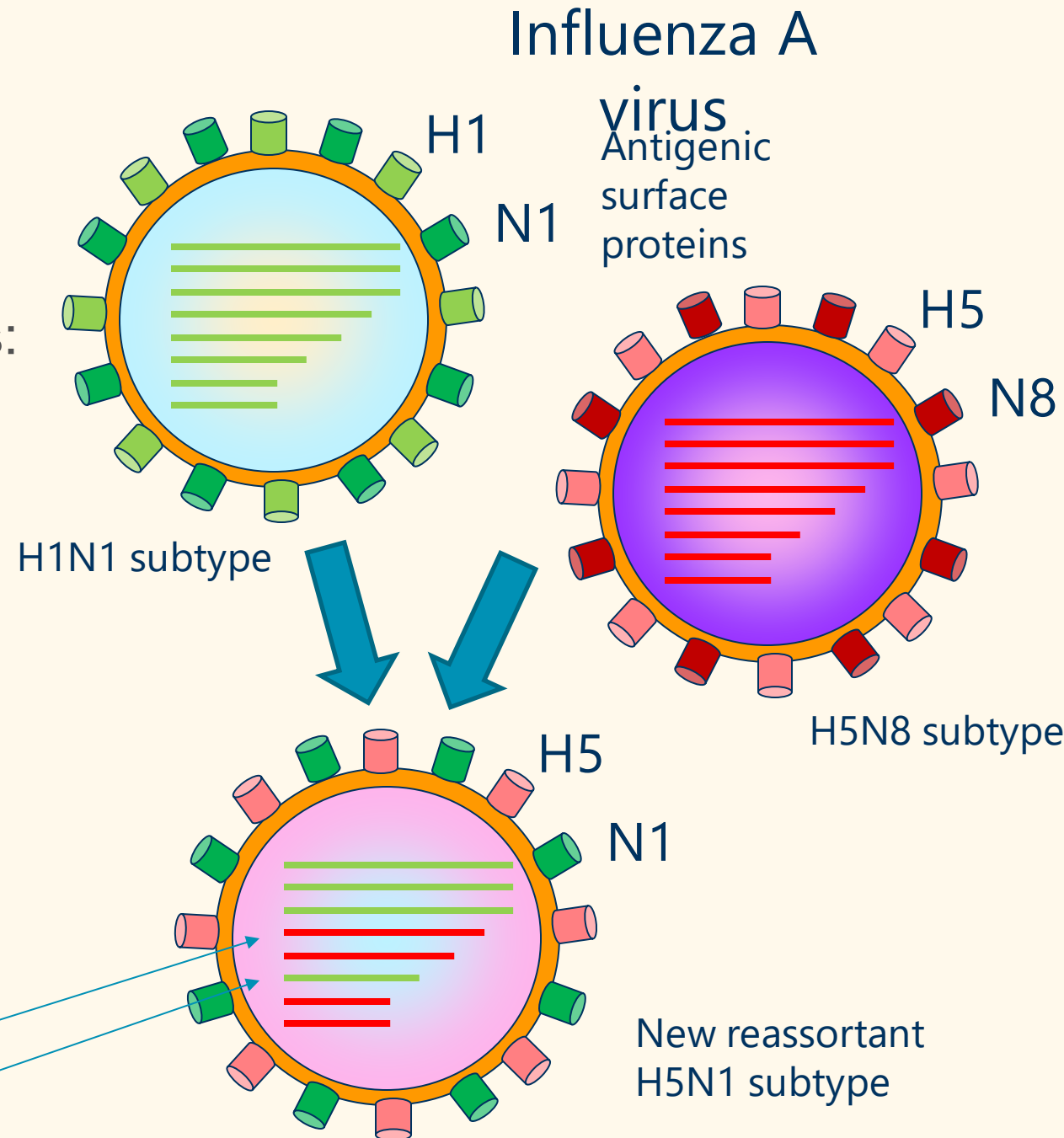


Influenza A Viruses

- 8 RNA Segments coding for 10+ proteins
- Virus subtype defined by surface proteins:
 - Hemagglutinin (HA) and
 - Neuraminidase (NA)
- New variants generated by:
 - reassortment (mixing) of segments
 - Fast mutation rate
 - Approx $5e-3$ per site per year
 - 10-70 nucleotides per year

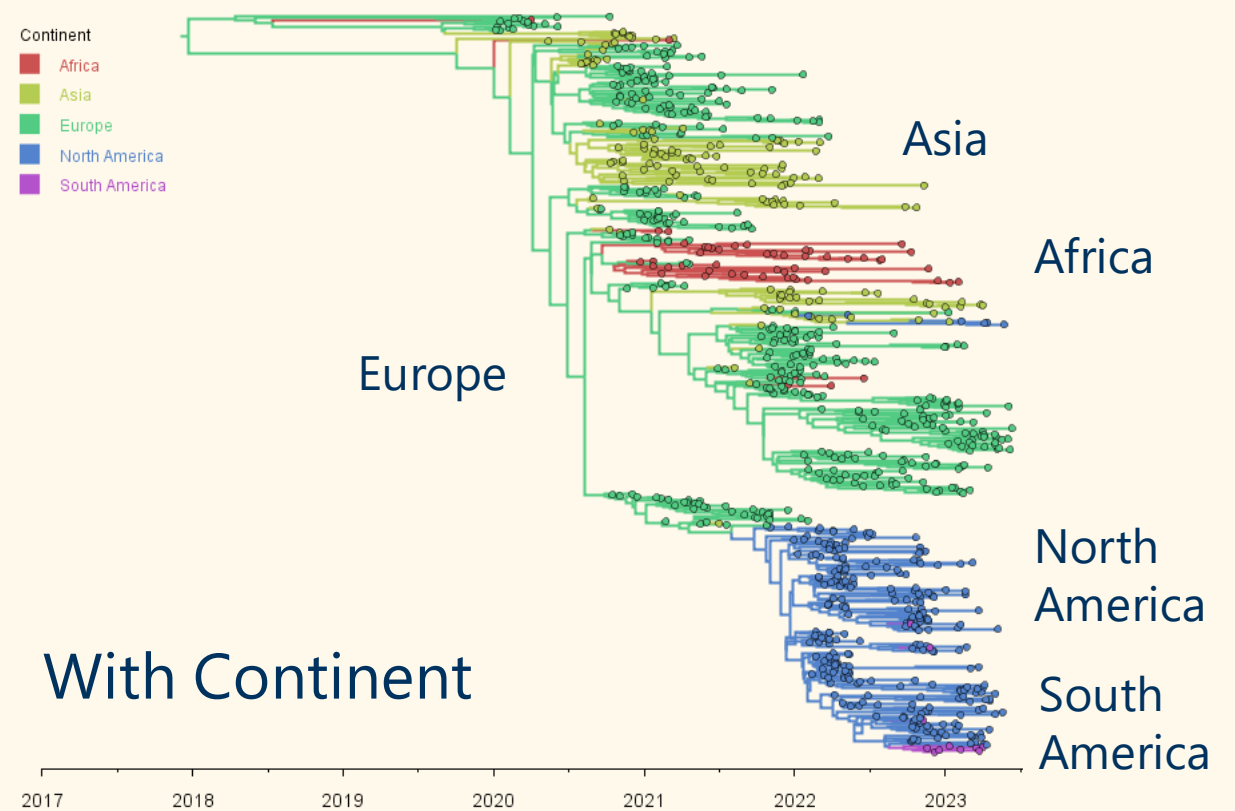
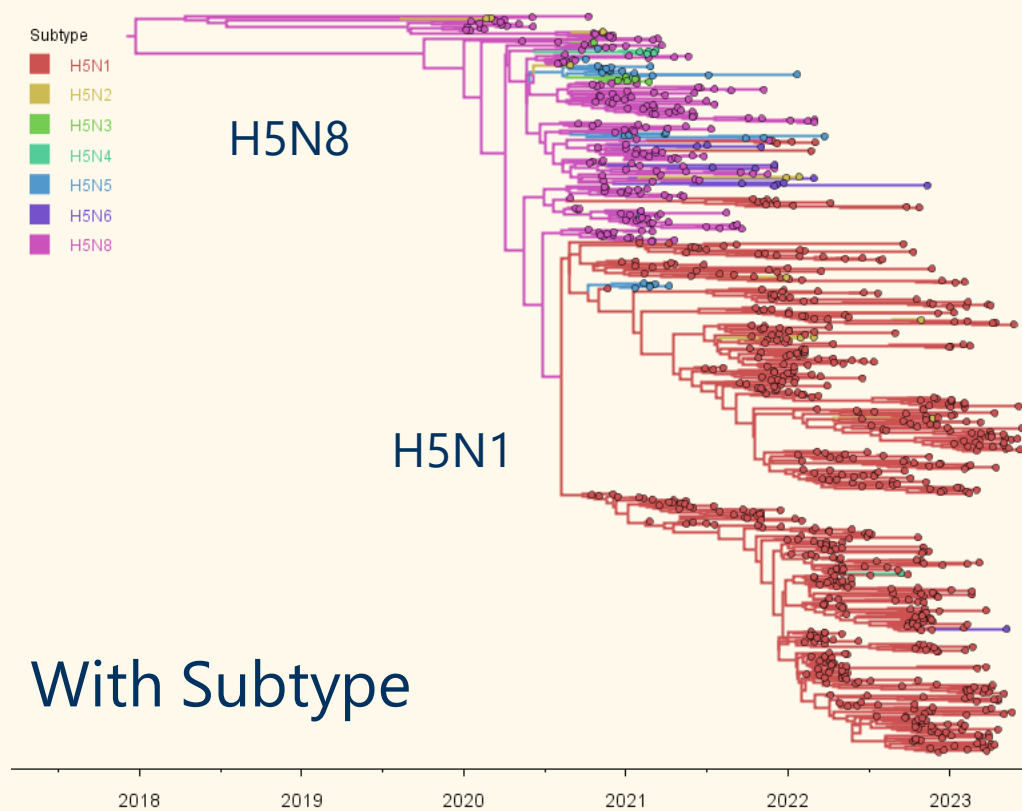
A segment = virus 'chromosome' of RNA
8 RNA segments in influenza genomes

Number 4 = HA
Number 6 = NA



Avian Influenza 2020-2023 Global Sequence Data

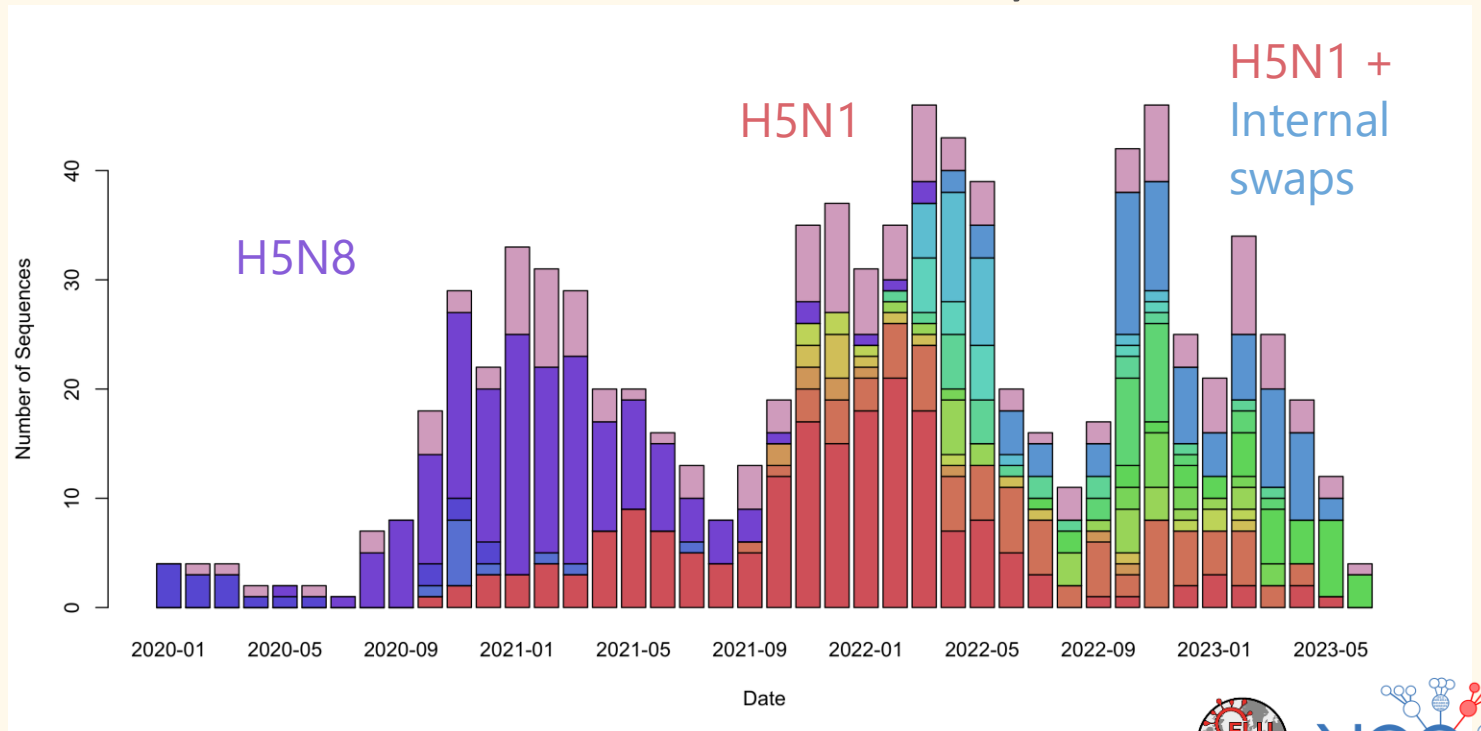
- 5000+ whole genomes from 2020-2023 including H5N8, H5N1 (2.3.4.4b)
- Use genome sequences to track transmission in time, space and species



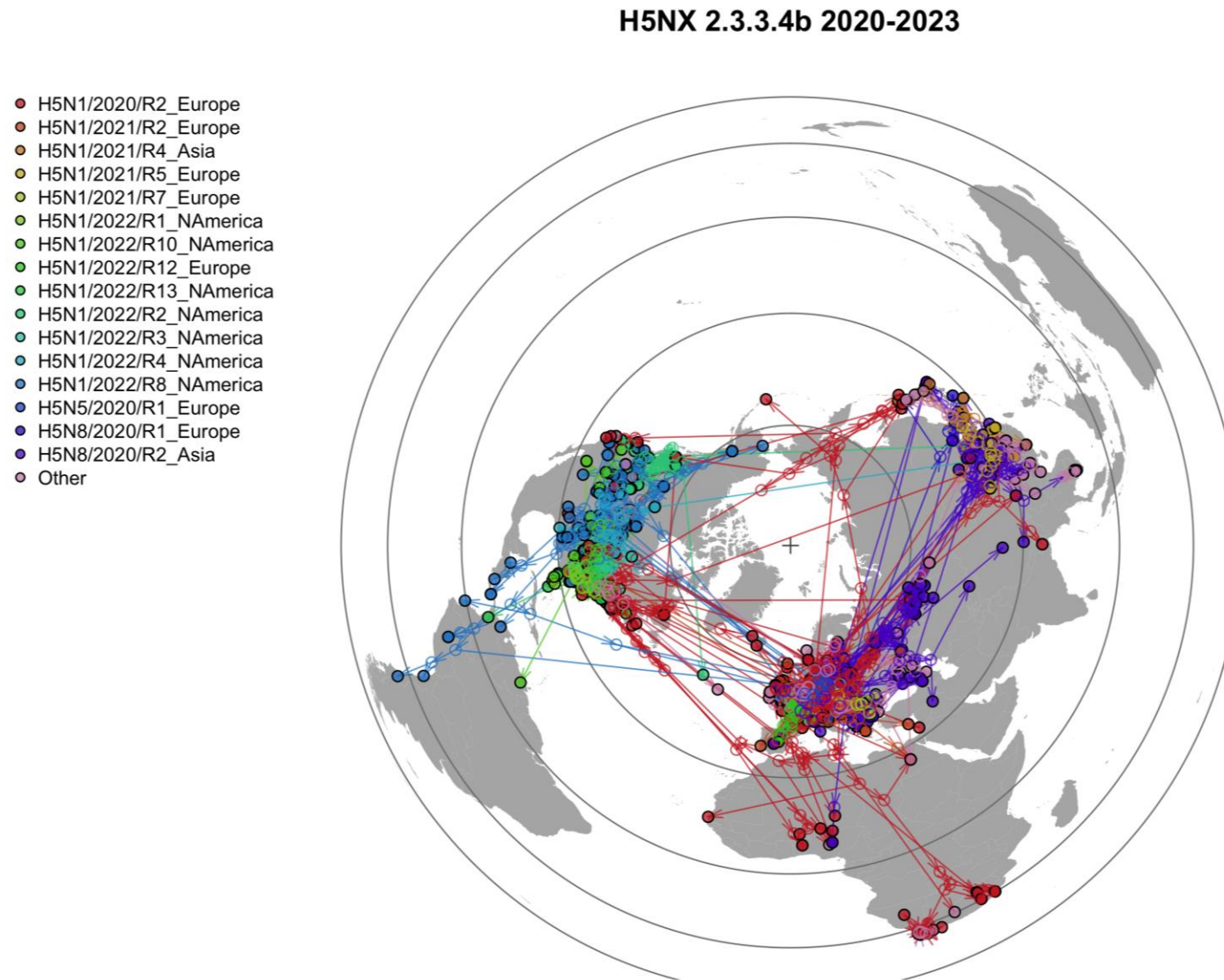
Avian Influenza 2020-2023 Reassortants in Detail

- 5000+ H5NX ("2.3.3.4b") whole genomes from 2020-2023 including H5N8, H5N1
- Other internal segments acquired from circulating low pathogenic AI
- Each segment split into clusters of 1-2% genetic distance and given segment code
- 8 Segment codes combined to Reassortment code to describe diversity

● H5N1/2020/R2_Europe	1	1	1	1	1	1	1	1
● H5N1/2021/R2_Europe	2	1	2	1	1	1	1	1
● H5N1/2021/R4_Asia	1	1	9	1	1	4	1	1
● H5N1/2021/R5_Europe	2	1	1	1	1	1	1	1
● H5N1/2021/R7_Europe	10	1	1	1	1	1	1	7
● H5N1/2022/R1_NAmerica	7	4	1	1	2	1	1	1
● H5N1/2022/R10_NAmerica	4	1	1	1	2	1	1	8
● H5N1/2022/R12_Europe	2	1	4	1	3	1	1	3
● H5N1/2022/R13_NAmerica	4	4	8	1	2	1	1	1
● H5N1/2022/R2_NAmerica	8	1	1	1	2	1	1	1
● H5N1/2022/R3_NAmerica	7	4	6	1	2	1	1	1
● H5N1/2022/R4_NAmerica	4	1	1	1	2	1	1	1
● H5N1/2022/R8_NAmerica	5	3	1	1	2	1	1	4
● H5N5/2020/R1_Europe	3	2	2	1	1	7	1	2
● H5N8/2020/R1_Europe	6	1	5	2	1	3	1	2
● H5N8/2020/R2_Asia	3	2	3	1	1	2	1	2
	PB2	PB1	PA	HA	NP	NA	MP	NS



Phylogeography: Global Spreading Pattern of Reassortants from North Pole



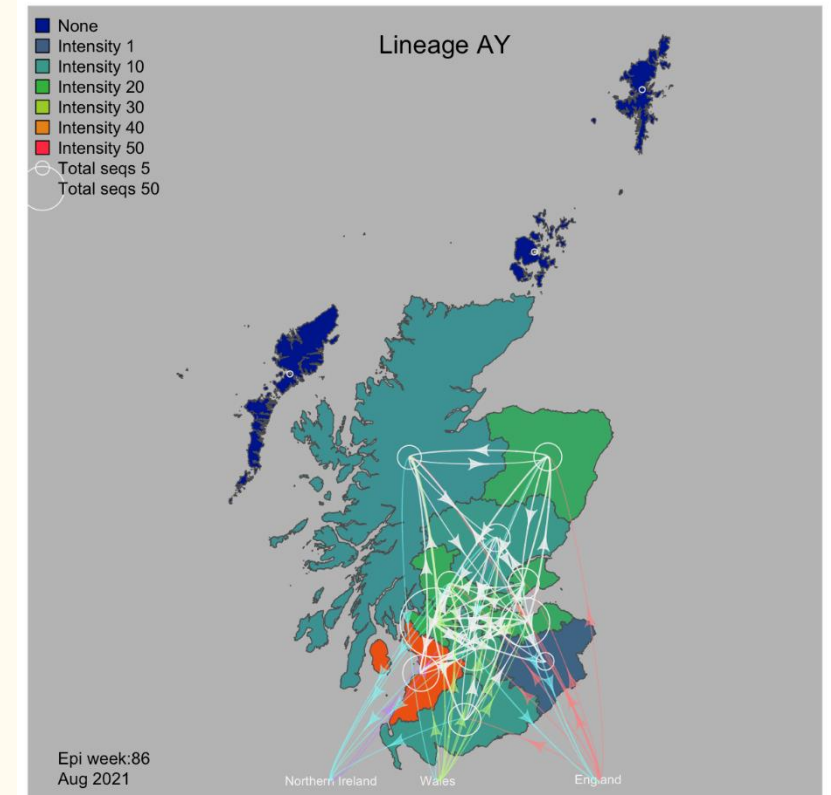
Looking down from
North Pole

Using virus sequence
data to track
infection in time and
space

Coloured by
reassortant type

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, is captured in flight against a clear, pale blue sky. The birds are arranged in a loose V-formation, with some leading and others following. Their wings are spread wide, showing dark feathers on the upper surfaces and lighter feathers on the undersides. The overall composition is clean and minimalist, emphasizing the natural beauty of the birds in flight.

Thank you !

samantha.lycett@ed.ac.uk