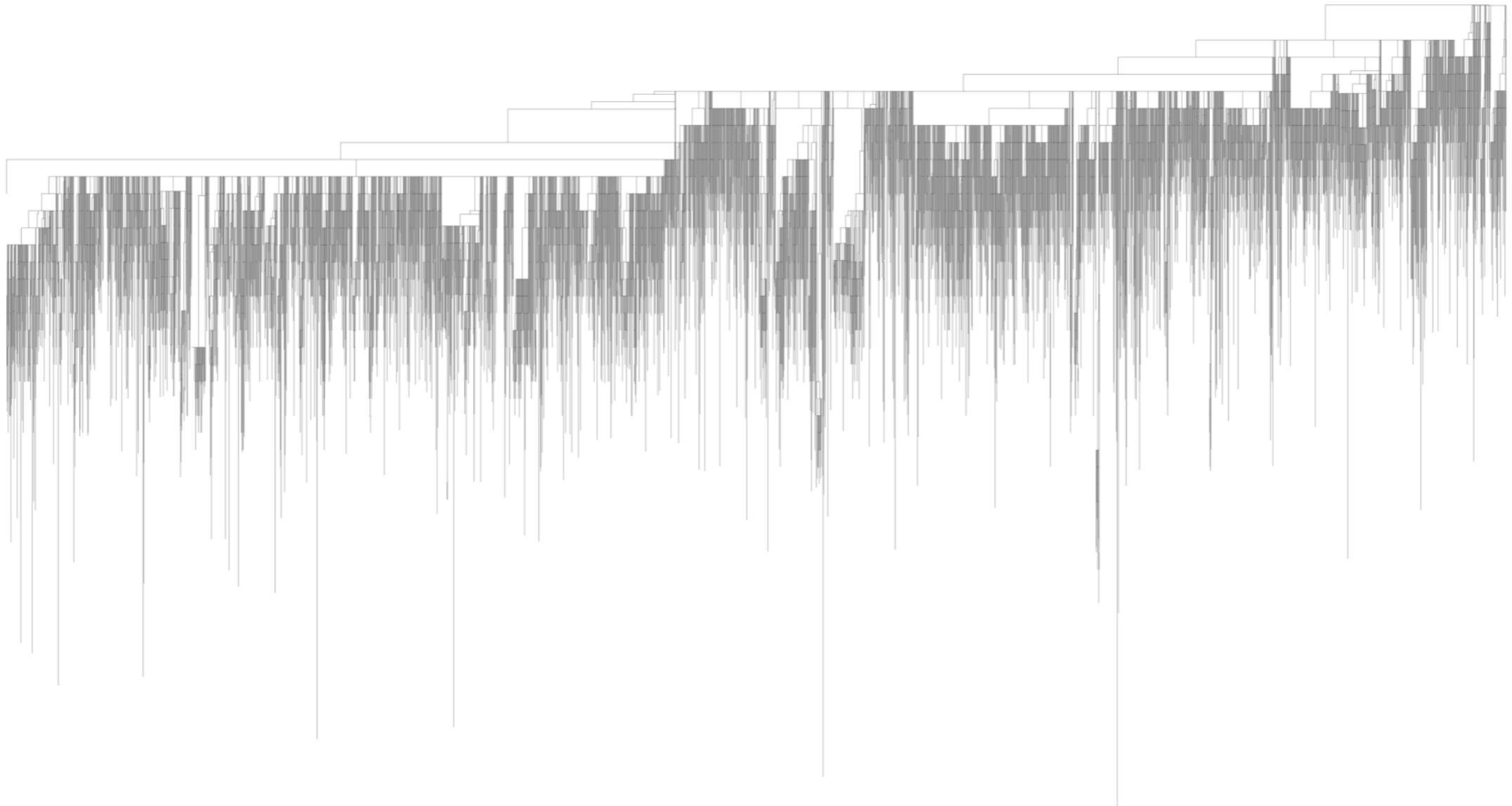


ESTABLISHMENT & LINEAGE DYNAMICS OF THE SARS-COV-2 EPIDEMIC IN THE UK

LOUIS DU PLESSIS



~88,000 SARS-CoV-2 genomes

<https://microreact.org/project/cogconsortium-2020-09-24>

Everything
I'm presenting
about

Phylogenetics
pipelines

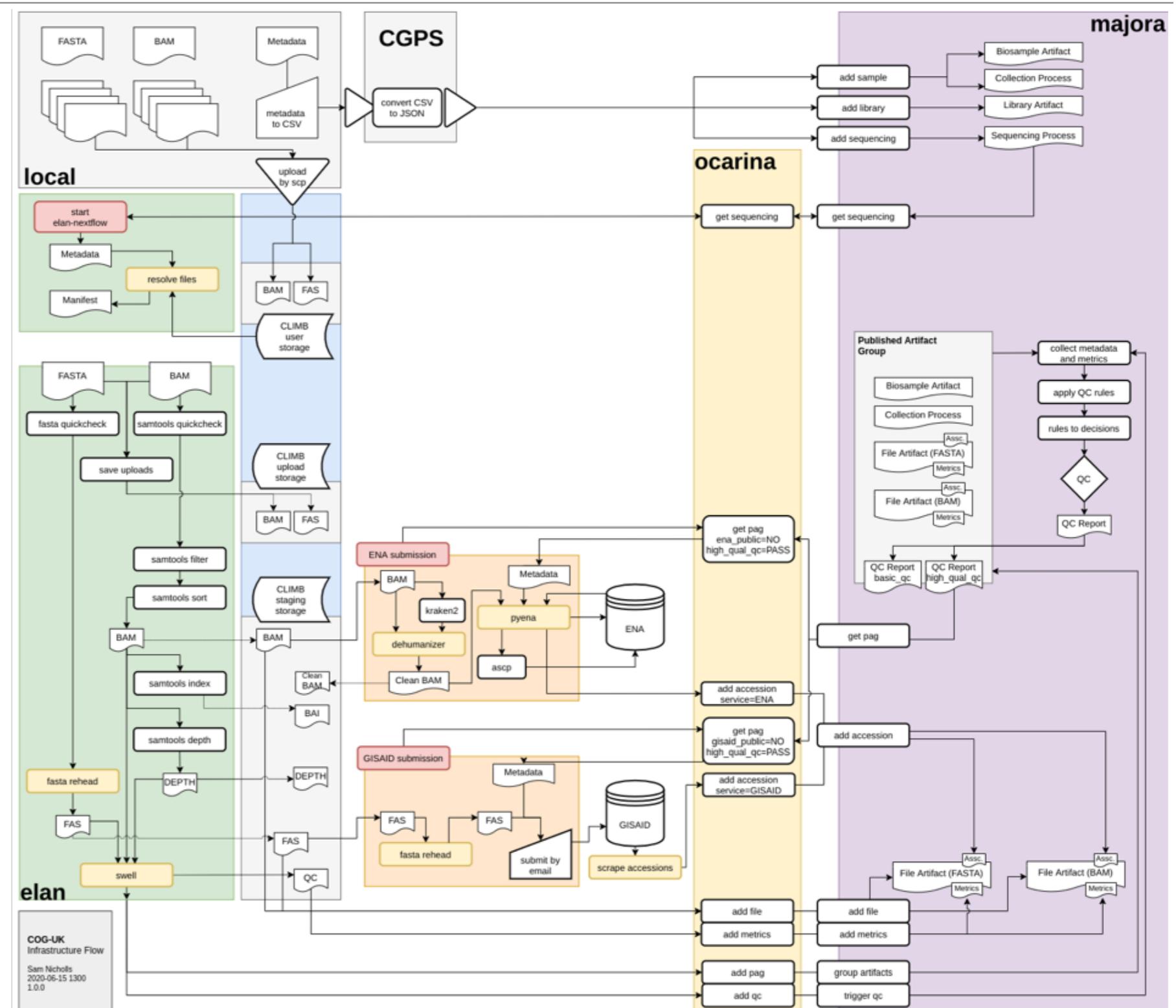
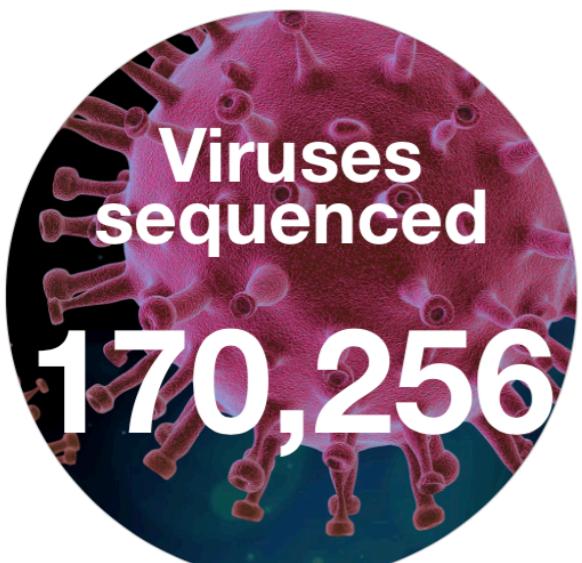
Sequencing
pipelines

Lighthouse
Labs



COVID-19 GENOMICS UK CONSORTIUM

- Multiple sequencing teams nationwide
- Unified informatics for genomes & metadata
- Data uploaded to GISAID and ENA



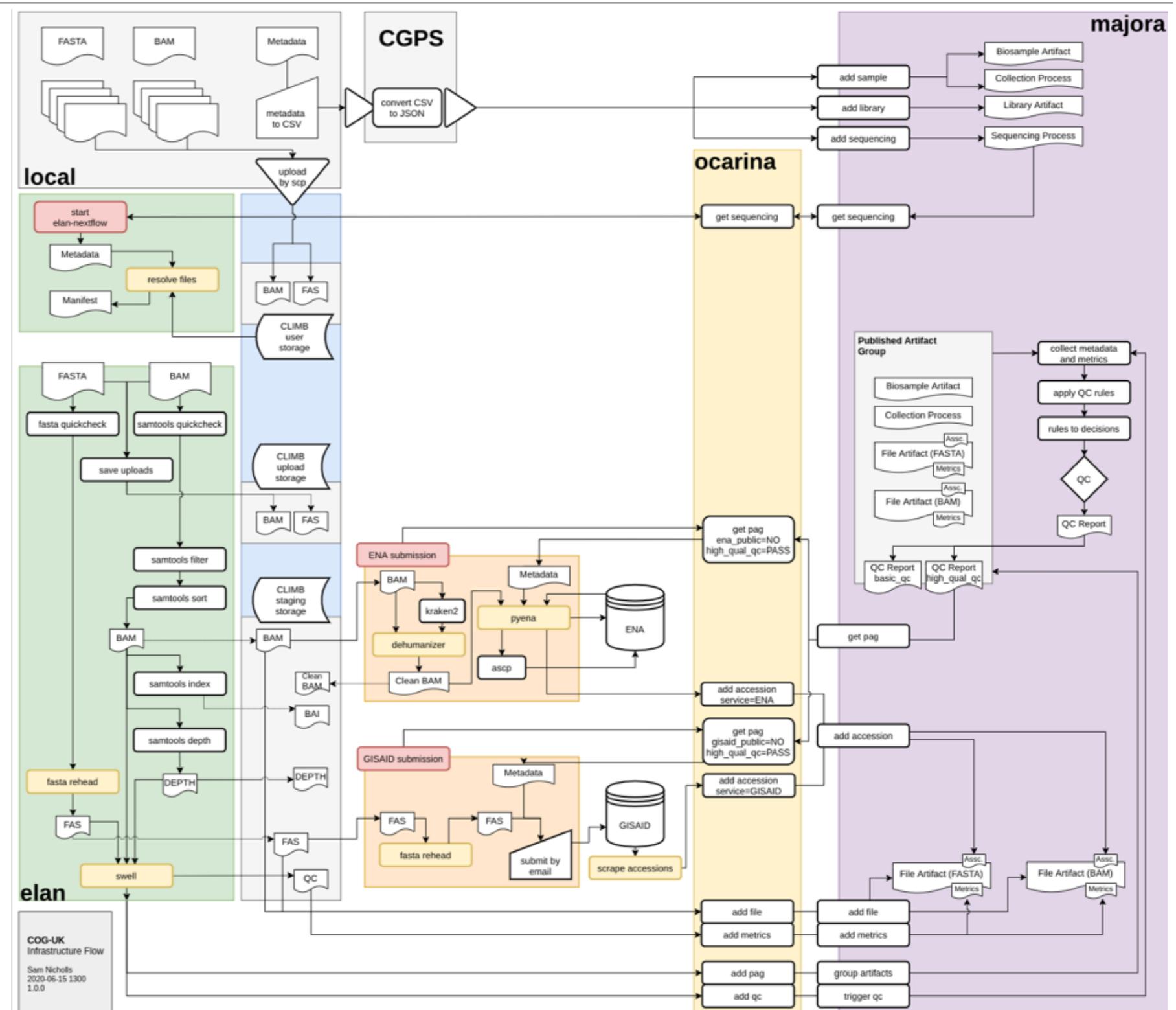


COVID-19 GENOMICS UK CONSORTIUM

- Multiple sequencing teams nationwide
- Unified informatics for genomes & metadata
- Data uploaded to GISAID and ENA

Viruses sequenced

469,041





COVID-19 GENOMICS UK CONSORTIUM

- Multiple sequencing teams nationwide
- Unified information for genomes & metadata
- Data uploaded to GISAID and ENA

Viruses sequenced

469,041

Lots of hard work from:

University of Birmingham

Sam Nicholls

Nick Loman

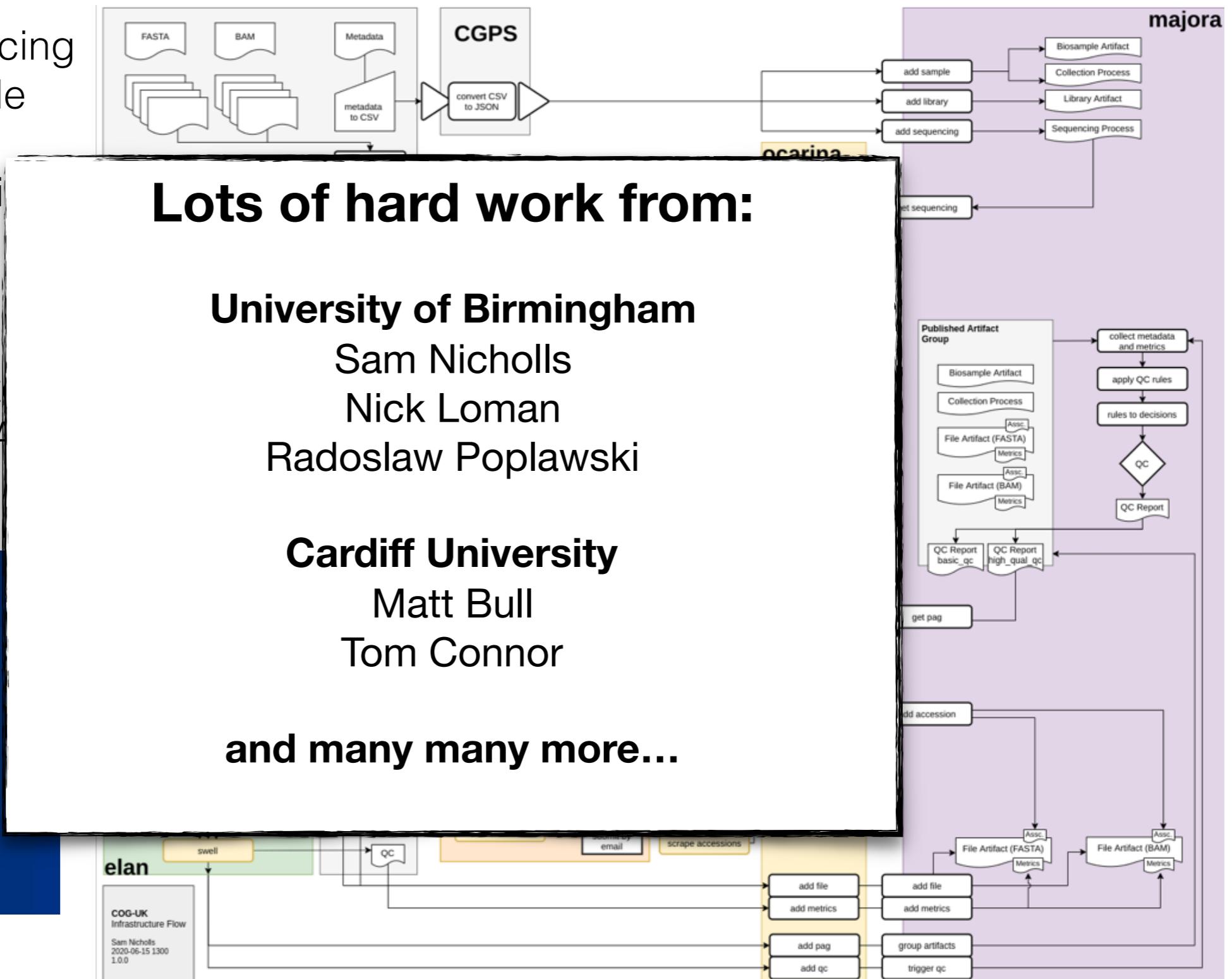
Radoslaw Poplawski

Cardiff University

Matt Bull

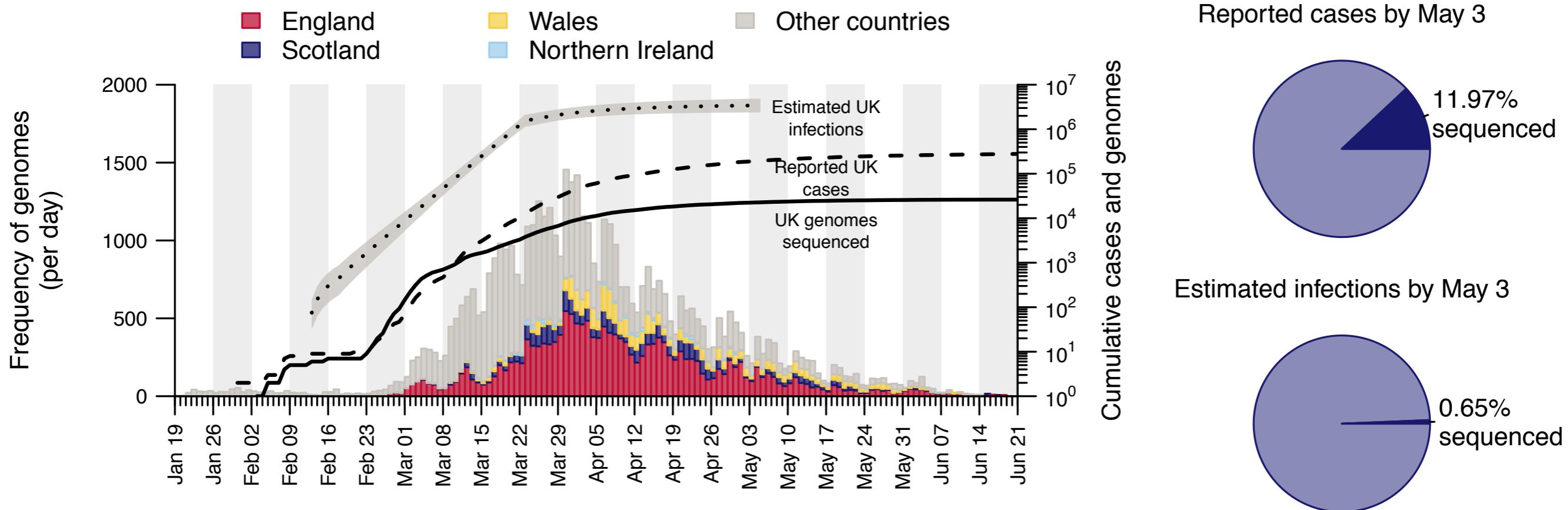
Tom Connor

and many many more...

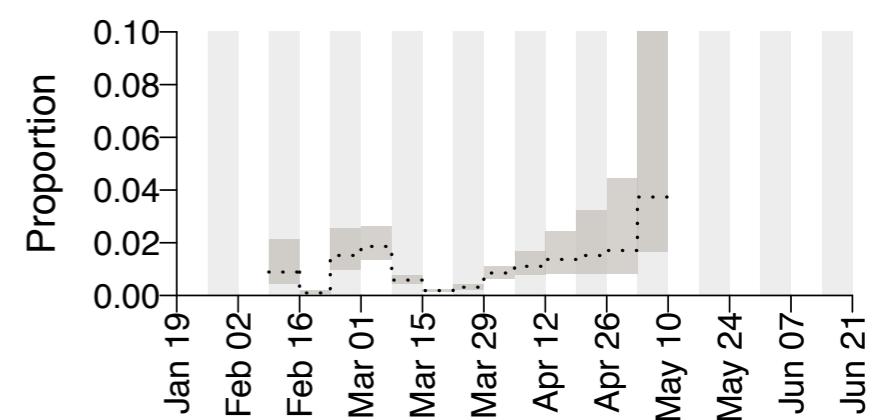


Sampling of SARS-CoV-2 genomes

First 6 months of the pandemic



- **50,887** SARS-CoV-2 genomes
- **26,181** from the UK (51.5%)



Seasonal Influenza (2010)

OPEN  ACCESS Freely available online

PLOS 

Global Migration Dynamics Underlie Evolution and Persistence of Human Influenza A (H3N2)

Trevor Bedford^{1,2*}, Sarah Cobey^{1,2}, Peter Beerli³, Mercedes Pascual^{1,2}

1 Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, United States of America, **2** Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan, United States of America, **3** Department of Scientific Computing, Florida State University, Tallahassee, Florida, United States of America

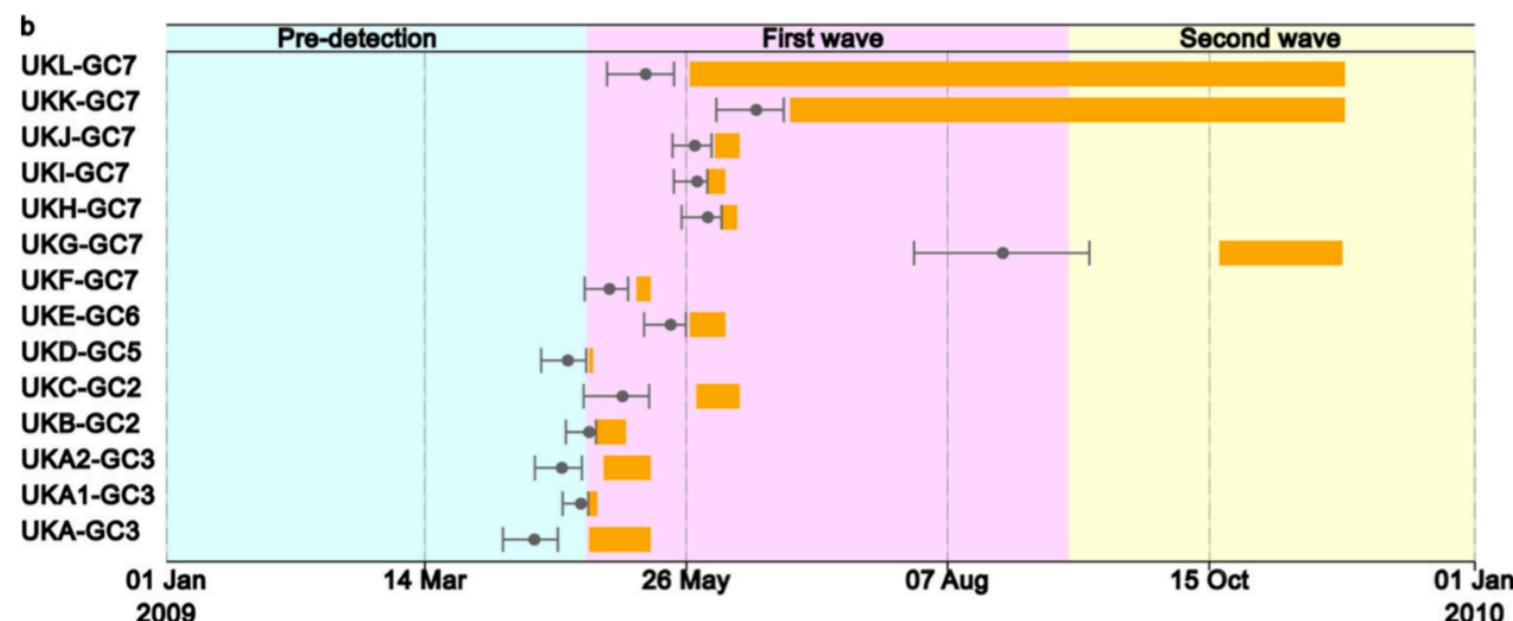
"Examining the influenza genealogy, it is apparent that regional outbreaks often result from very few immigration events, consistent with previous results. For example, the 2003 epidemic in Oceania appears almost completely monophyletic and can trace its history to a single migration event (or perhaps multiple migration events of identical strains) in early 2003."

Pandemic H1N1 (2012)

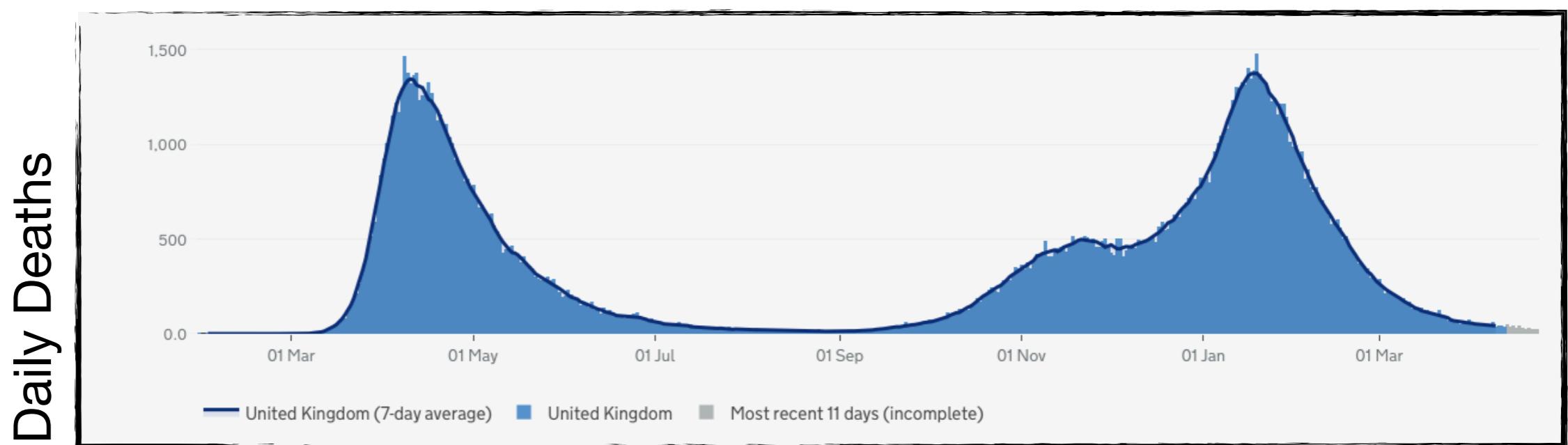
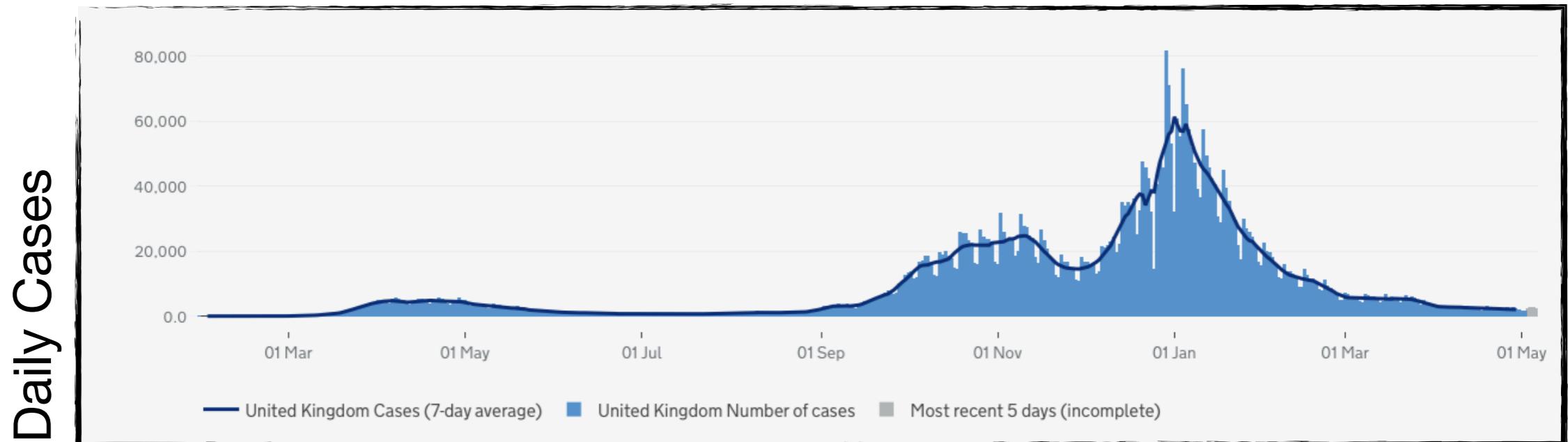


Evolutionary Dynamics of Local Pandemic H1N1/2009 Influenza Virus Lineages Revealed by Whole-Genome Analysis

Gregory J. Baillie,^a Monica Galiano,^b Paul-Michael Agapow,^b Richard Myers,^b Rachael Chiam,^a Astrid Gall,^a Anne L. Palser,^a Simon J. Watson,^a Jessica Hedge,^c Anthony Underwood,^b Steven Platt,^b Estelle McLean,^b Richard G. Pebody,^b Andrew Rambaut,^{c,d} Jonathan Green,^b Rod Daniels,^e Oliver G. Pybus,^f Paul Kellam,^a and Maria Zambon^b



Epidemic time-series (United Kingdom)



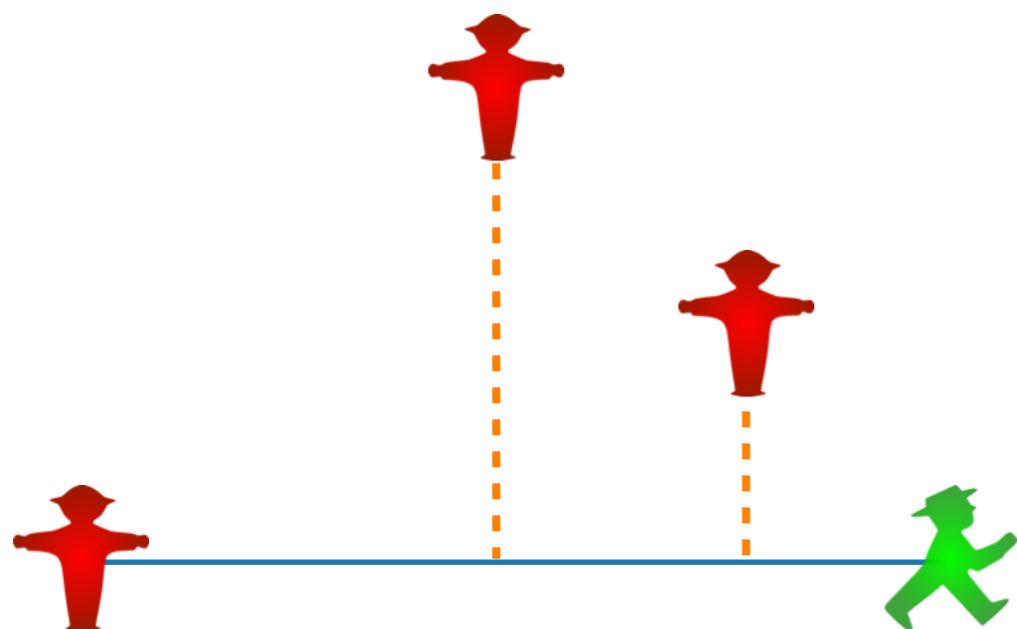
Definition: Transmission tree



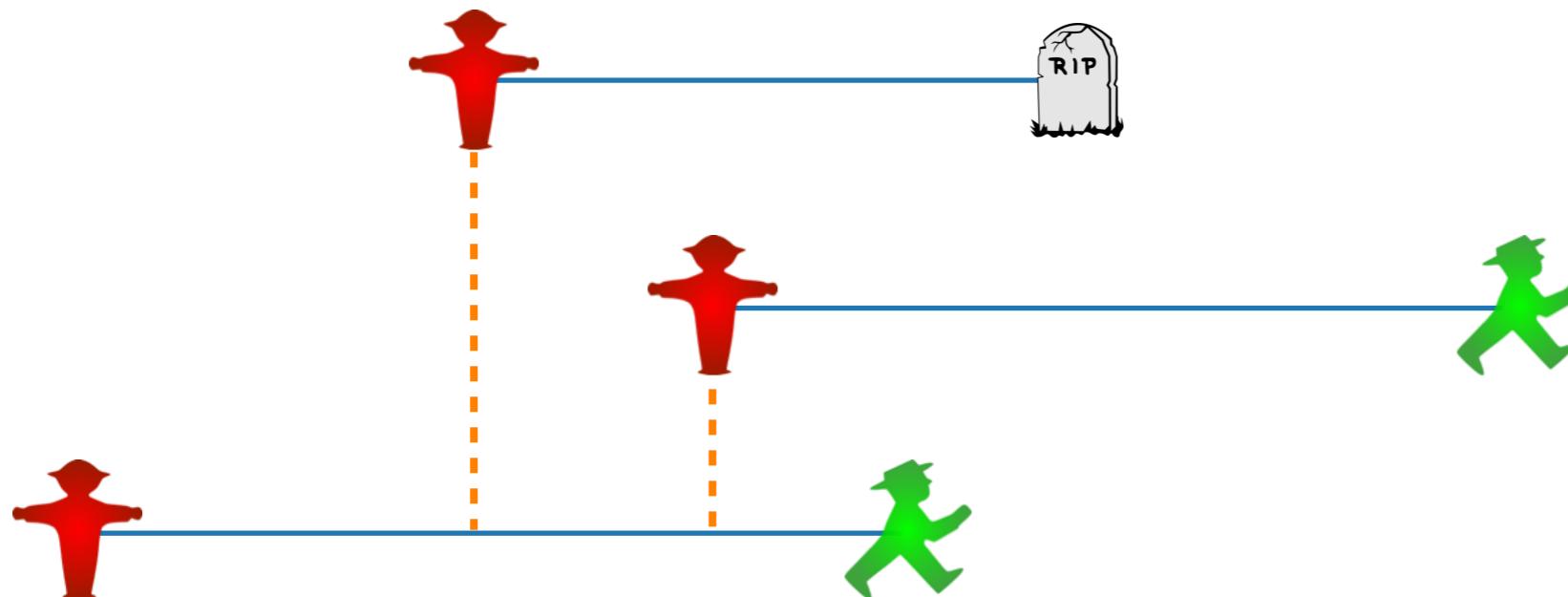
Definition: Transmission tree



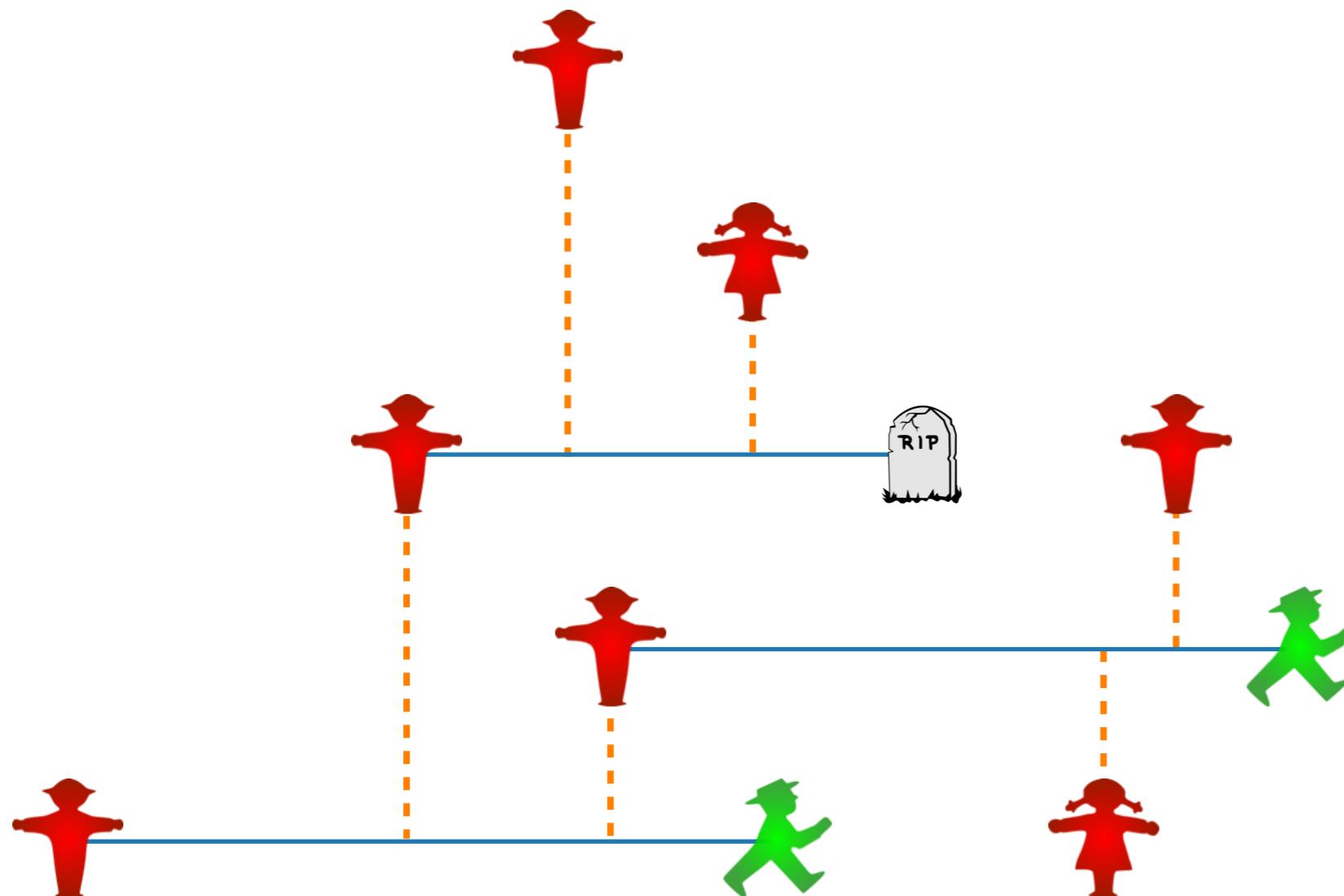
Definition: Transmission tree



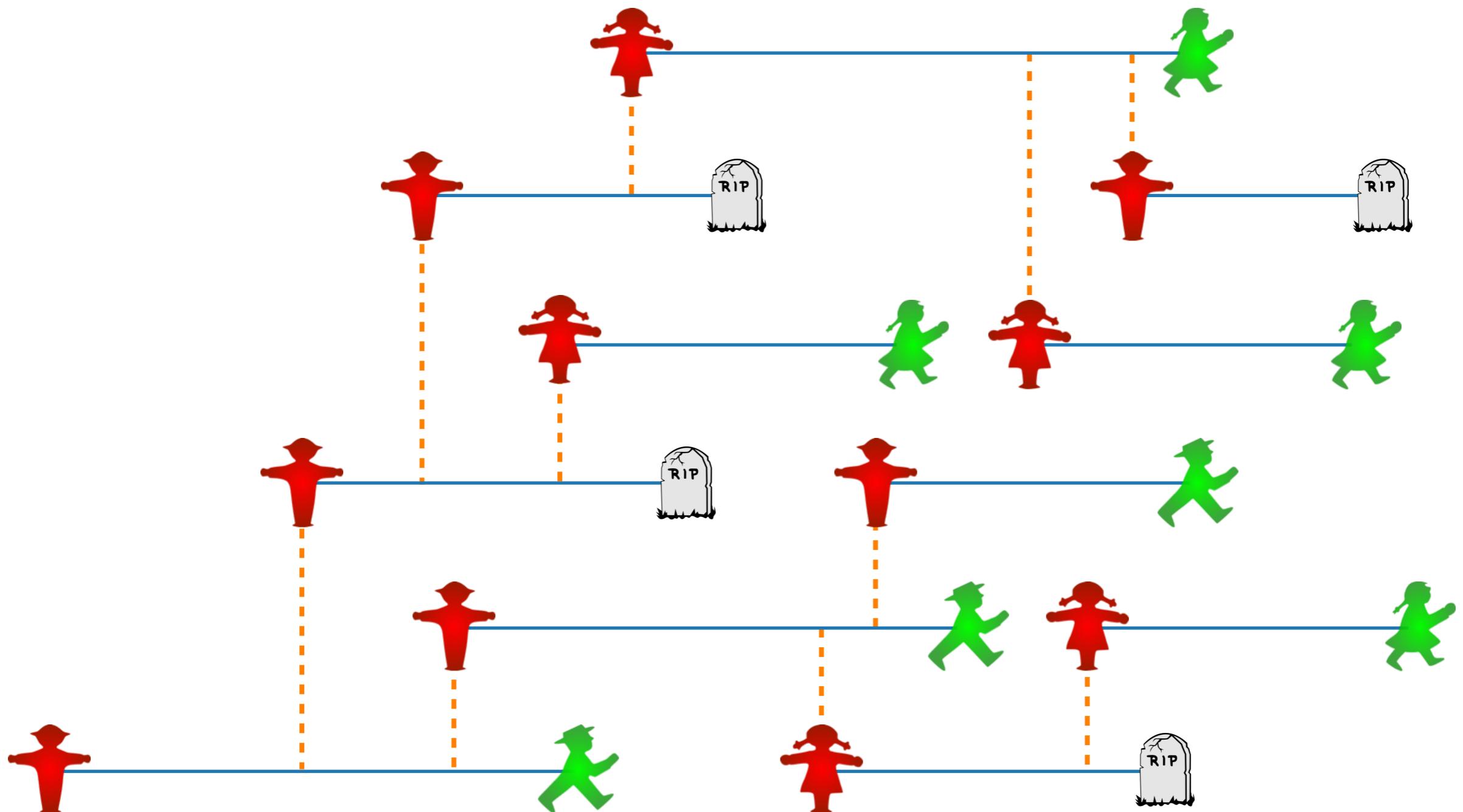
Definition: Transmission tree



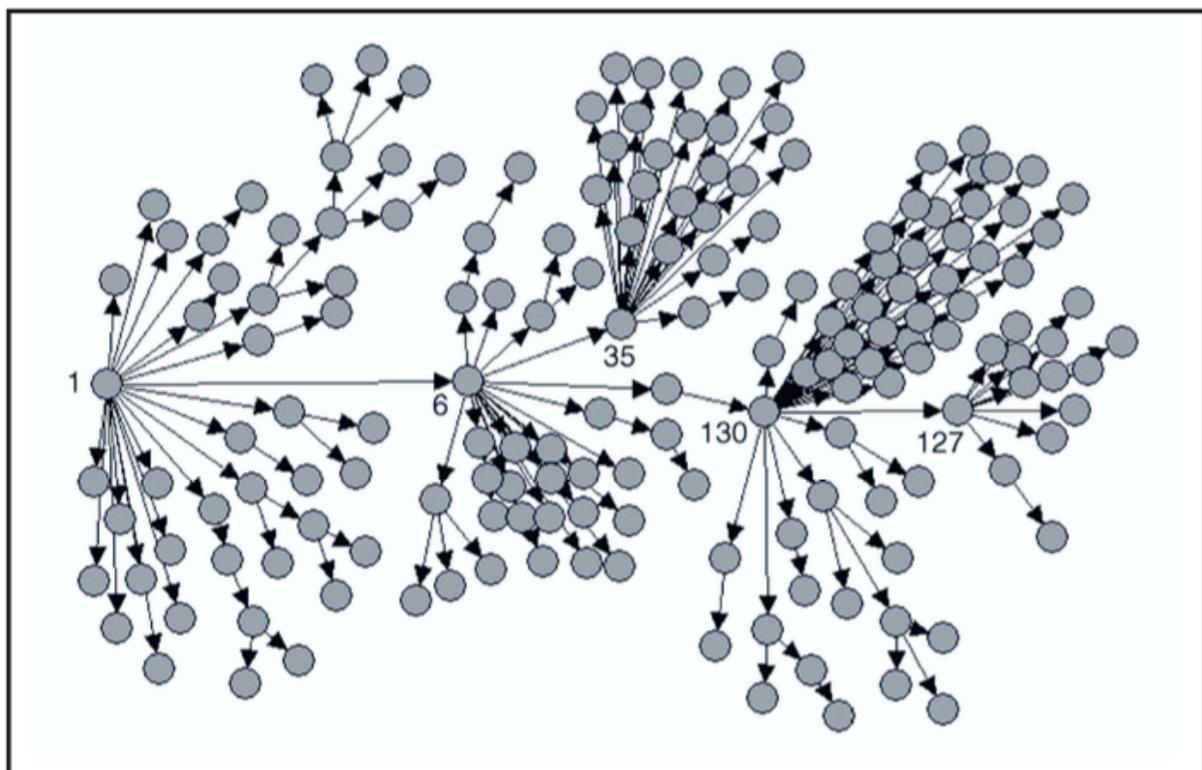
Definition: Transmission tree



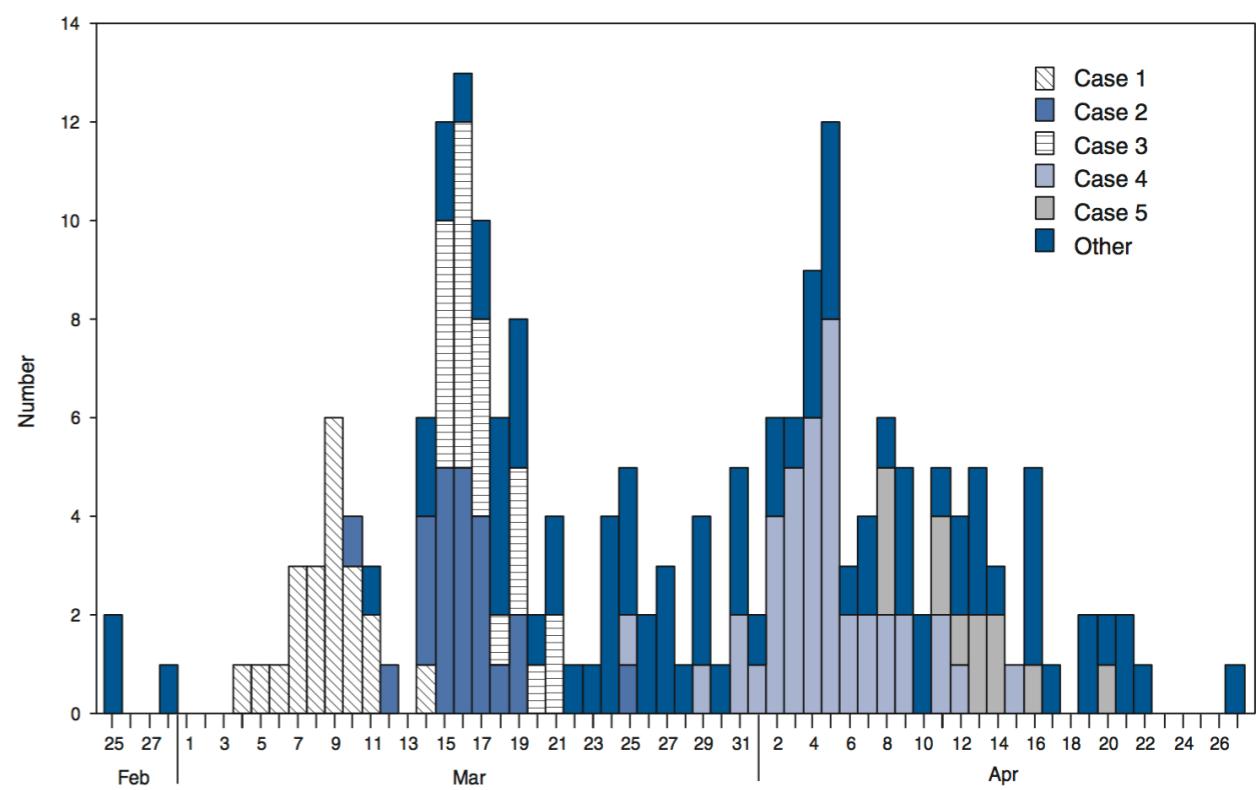
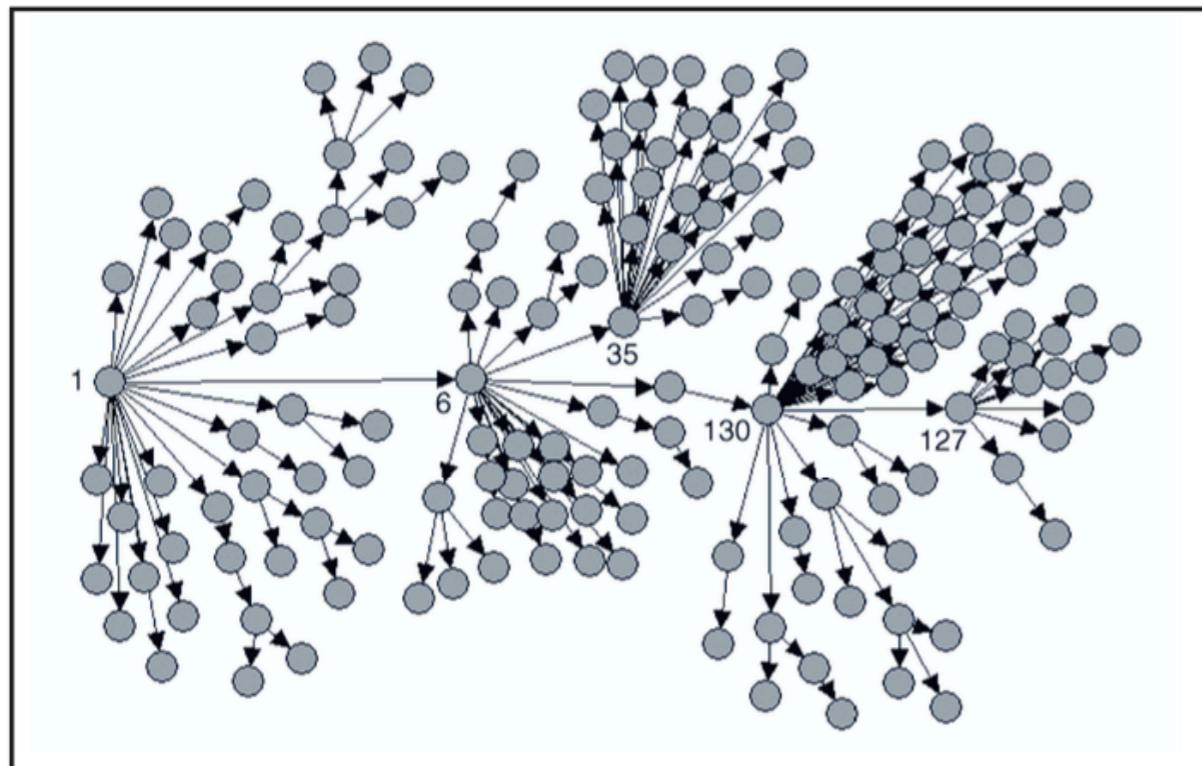
Definition: Transmission tree



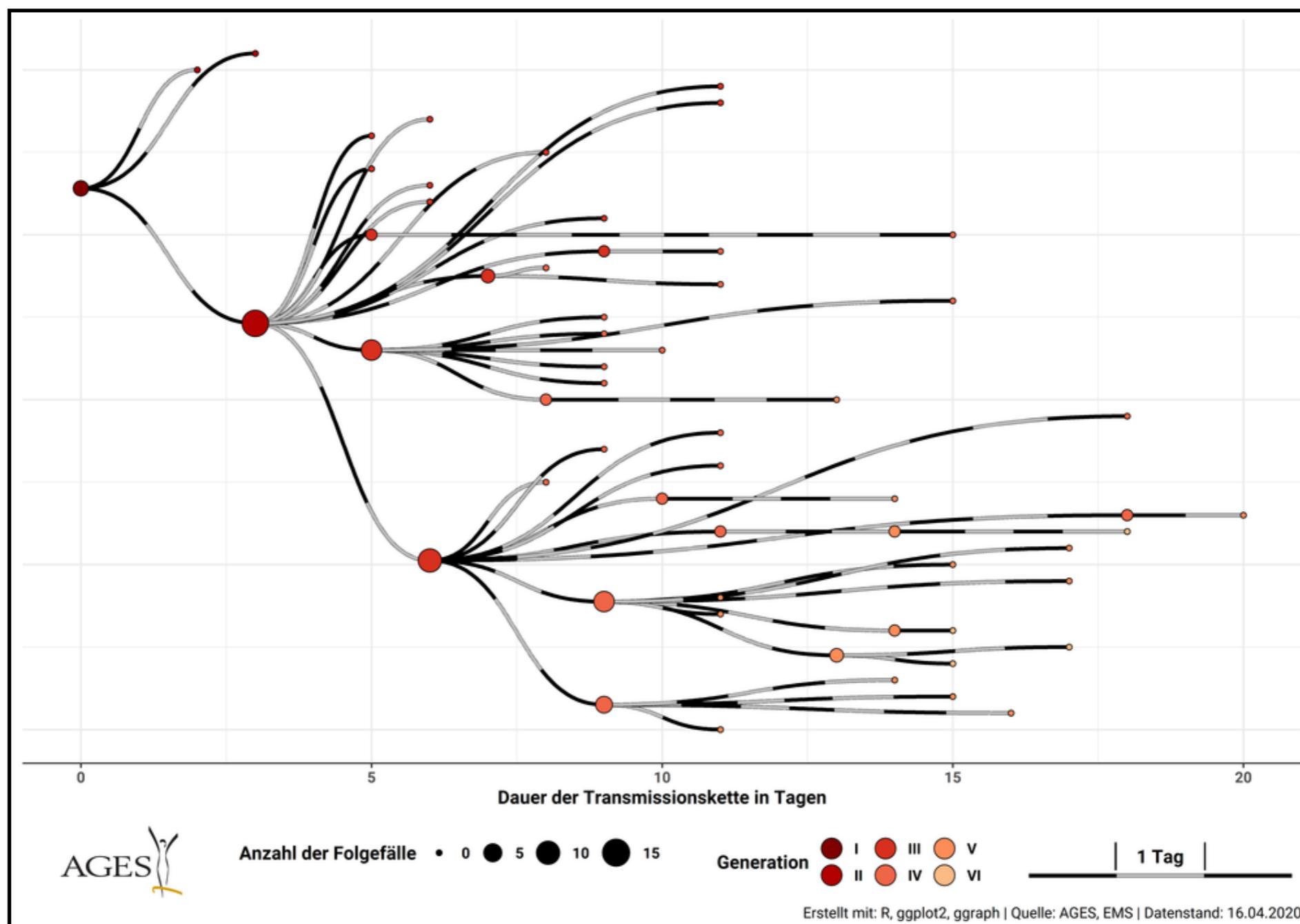
SARS-1 in Singapore (Feb-Apr 2003)



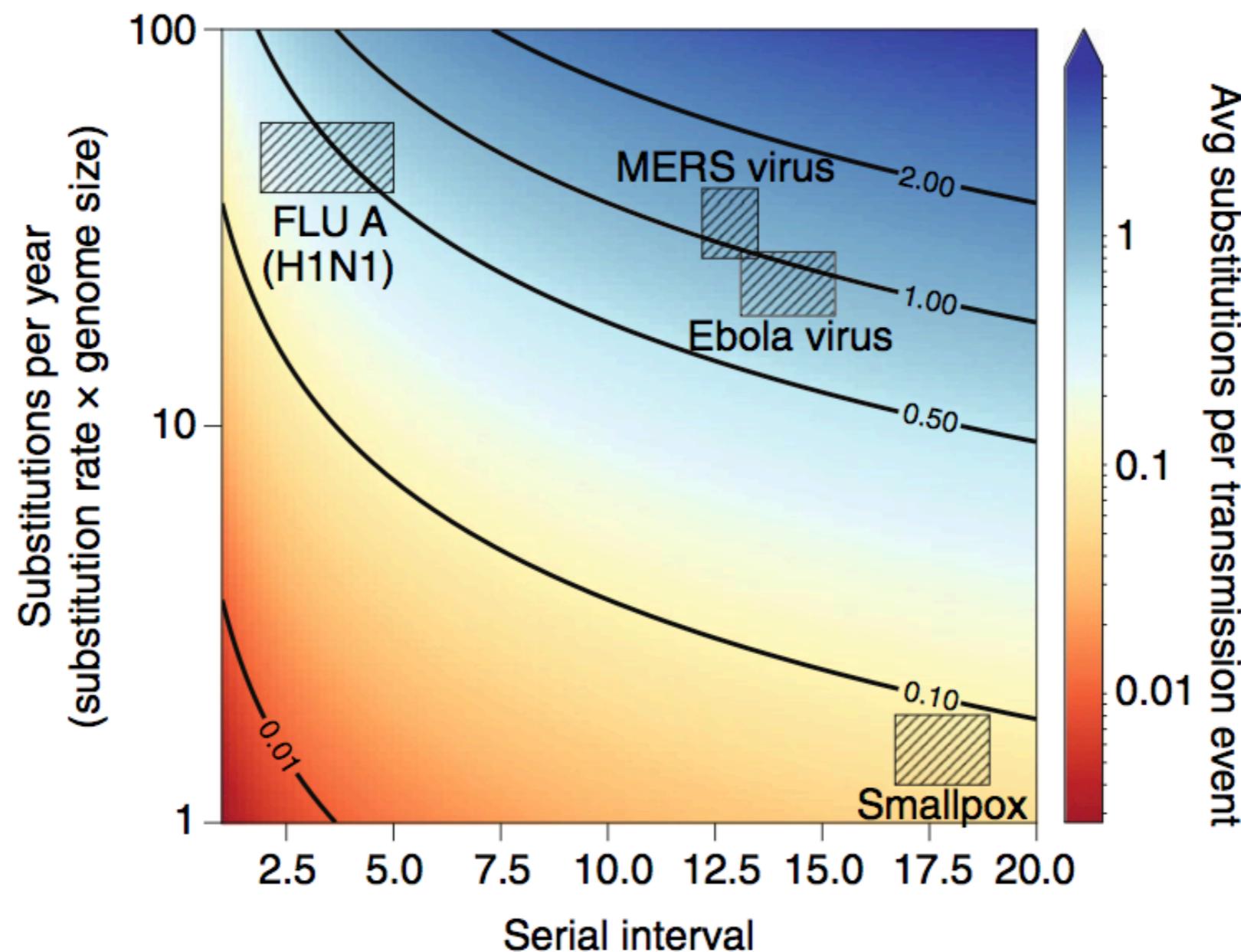
SARS-1 in Singapore (Feb-Apr 2003)



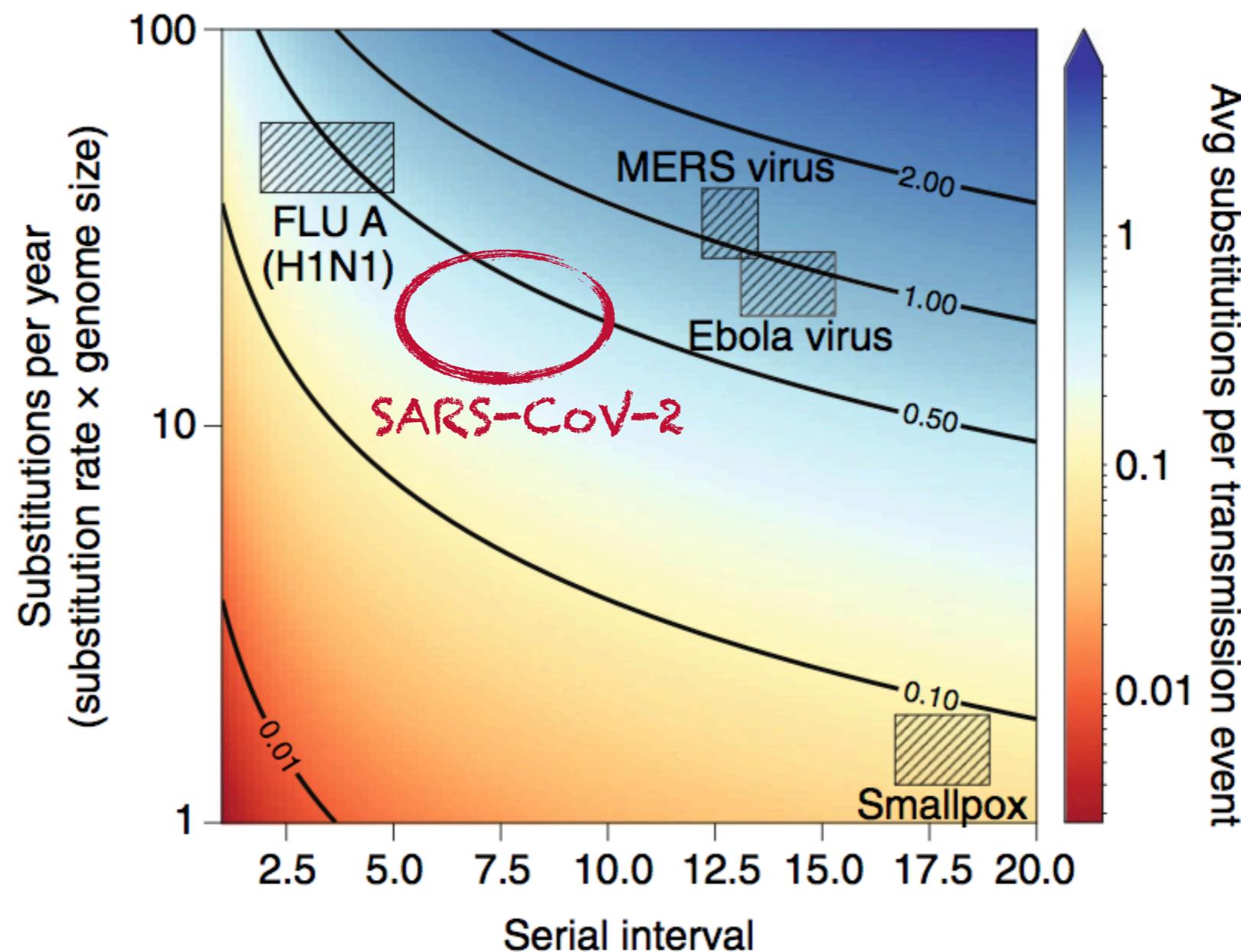
SARS-CoV-2 in Austria (Apr 2020)



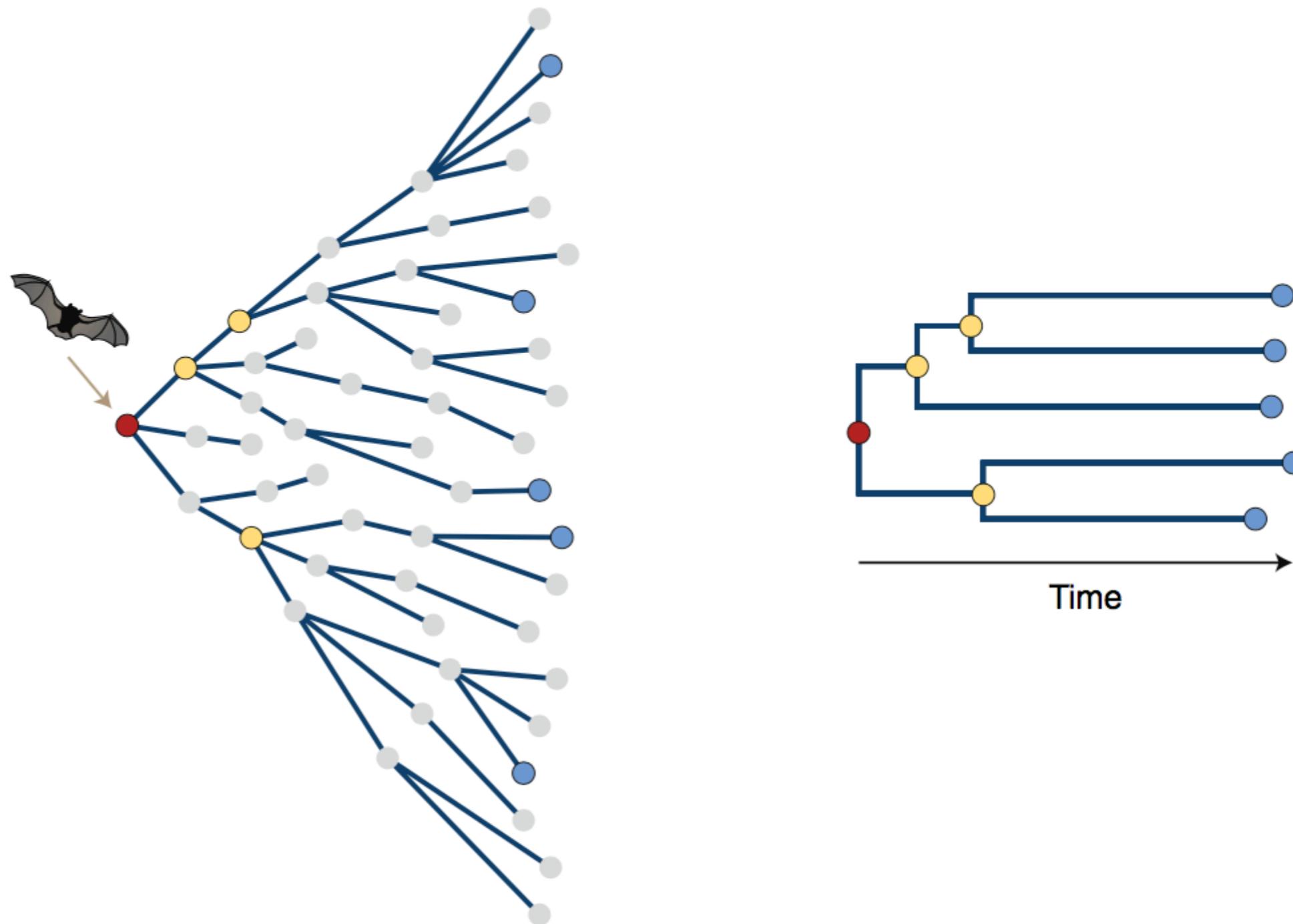
Measurably evolving populations and molecular clocks



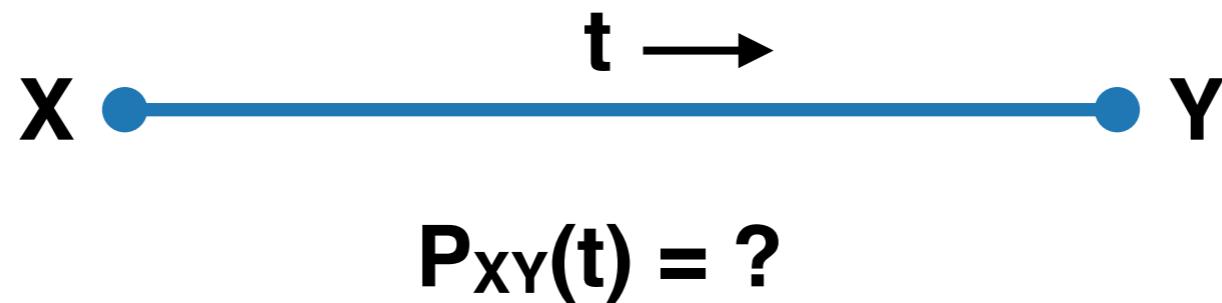
Measurably evolving populations and molecular clocks



Definition: Partial transmission tree (phylogeny)



How do we reconstruct the tree?



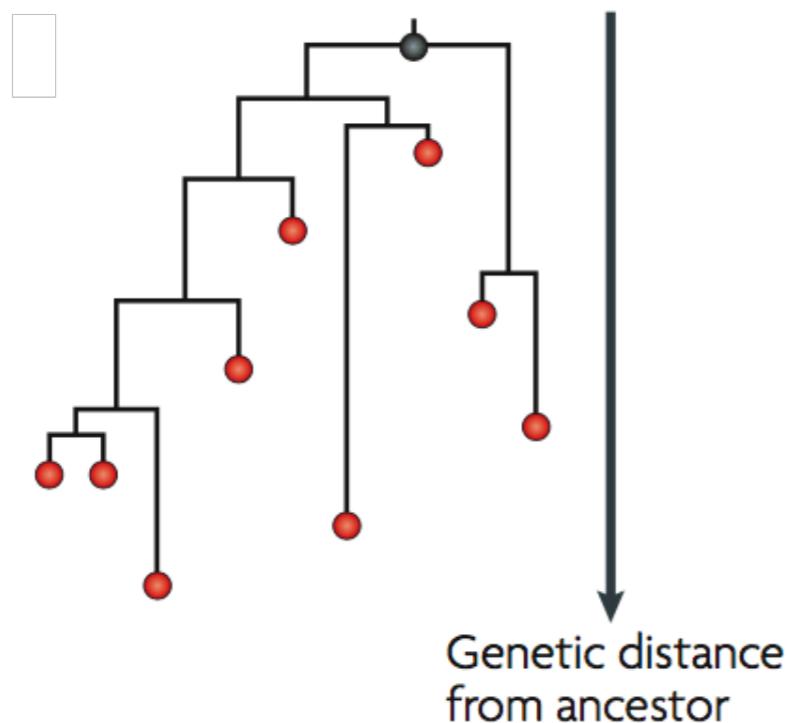
- What is the probability of observing **Y** at the end of the branch?
- Assume a site model that describes probabilities of different types of mutations over time (usually a CTMC)
- Write down the likelihood:

$$P(\begin{array}{c} \text{ACAC...} \\ \text{TCAC...} \\ \text{ACAG...} \end{array} \mid \text{tree})$$

- Optimise the tree and site model parameters using maximum-likelihood or MCMC

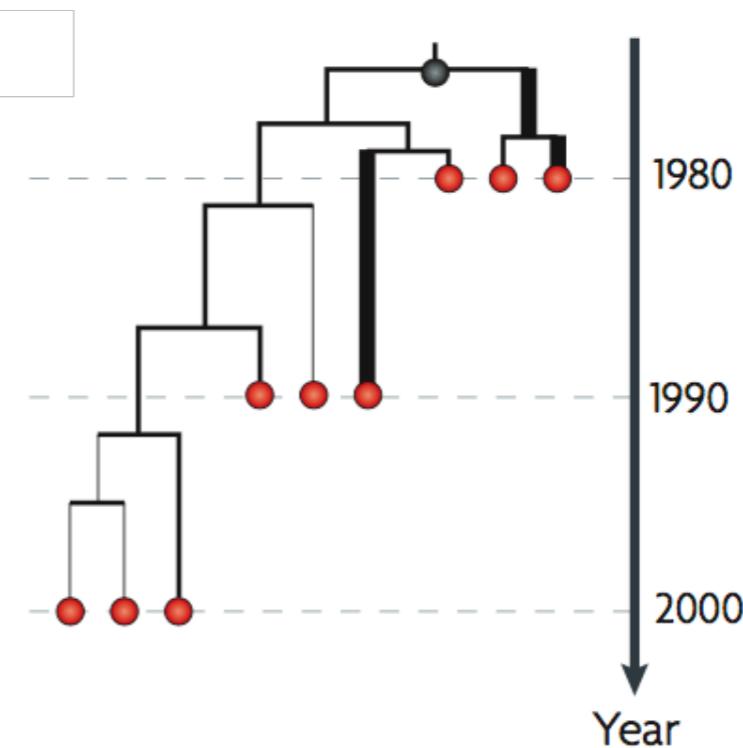
Getting to a time-calibrated tree

Genetic distance tree
(subst/site)



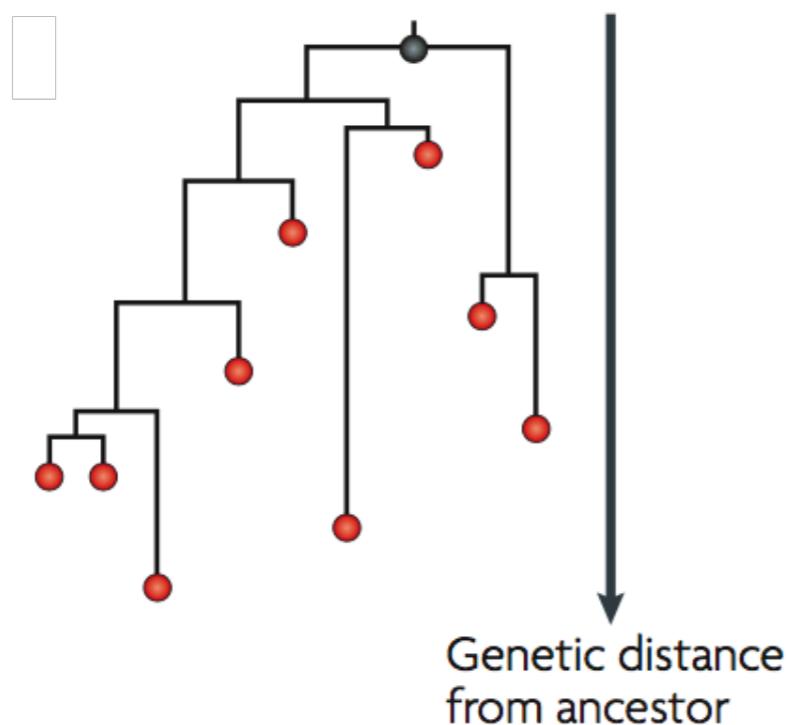
Clock rate
(subst/site/year)
 $= \mu \times$

Time-calibrated tree
(year)



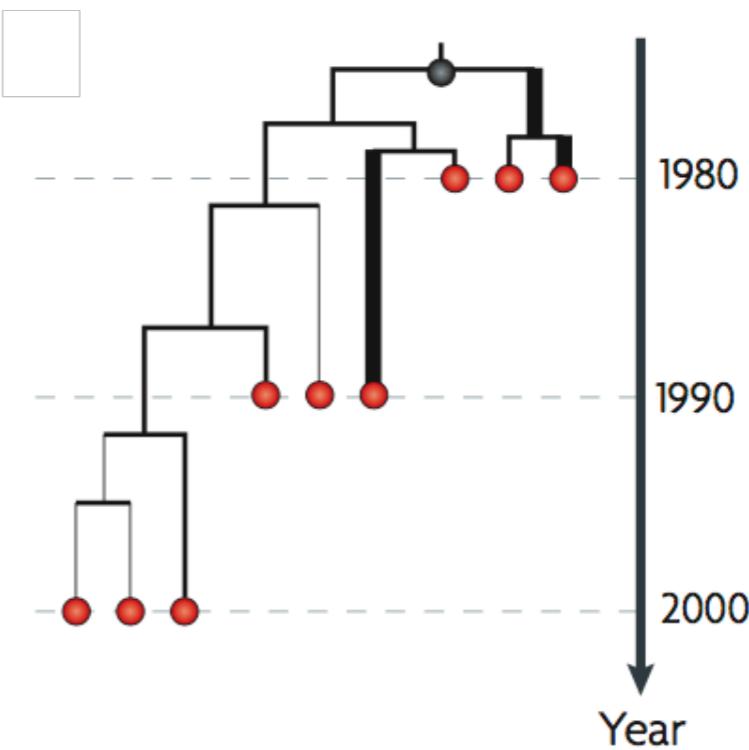
Getting to a time-calibrated tree

Genetic distance tree (subst/site)



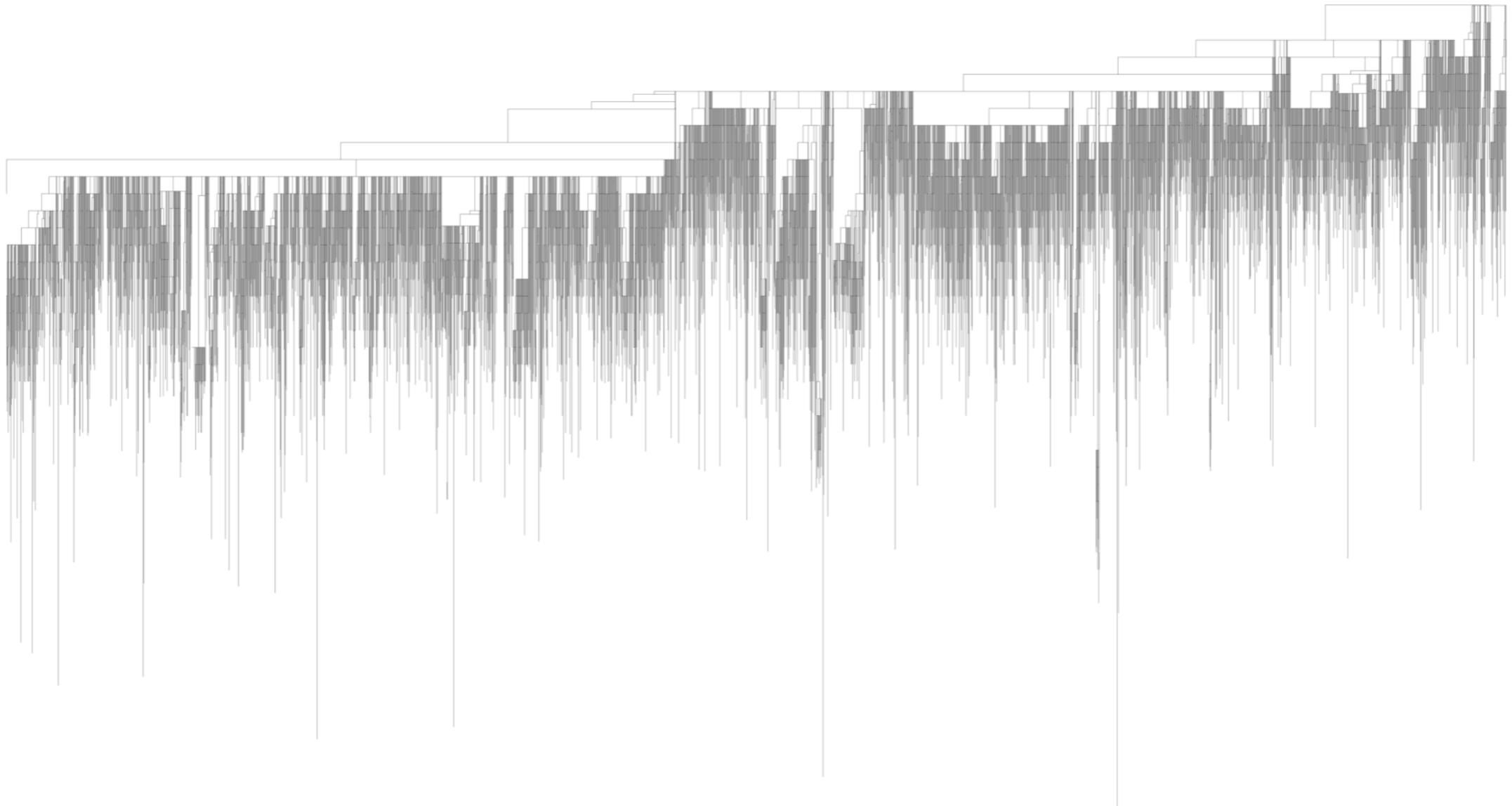
Clock rate
(subst/site/year)
 $= \mu \times$

Time-calibrated tree (year)



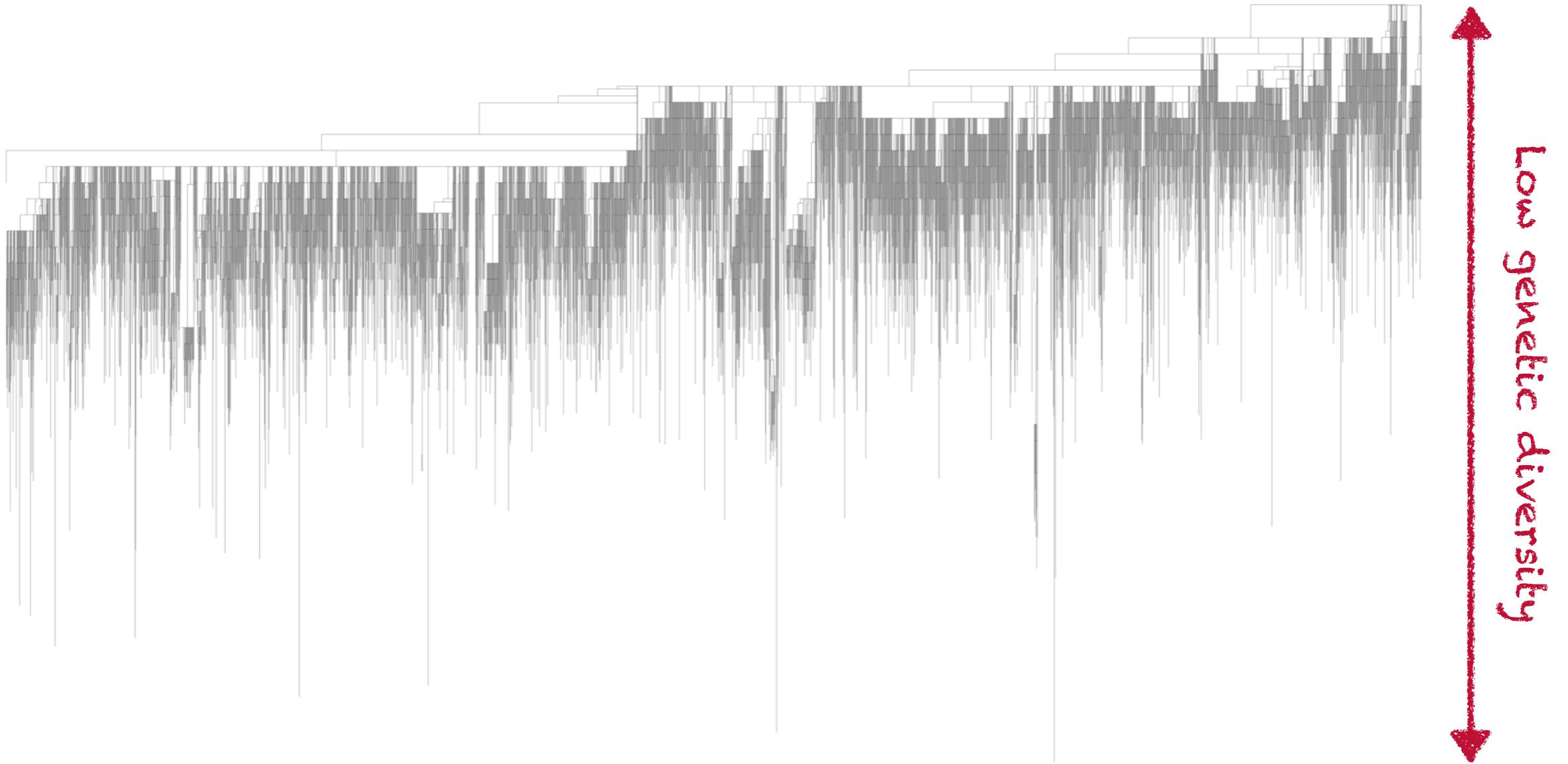
$$P(\text{ACAC...} | \text{TCAC...}, \text{ACAG...})$$

The equation shows a sequence alignment with three rows of DNA bases: ACAC..., TCAC..., and ACAG.... To the right of the alignment is a phylogenetic tree with blue branches. Next is a grid of colored circles (orange and blue) representing mutation states. Finally, there is a clock icon with a red dot and a blue arrow pointing clockwise, symbolizing the clock rate.



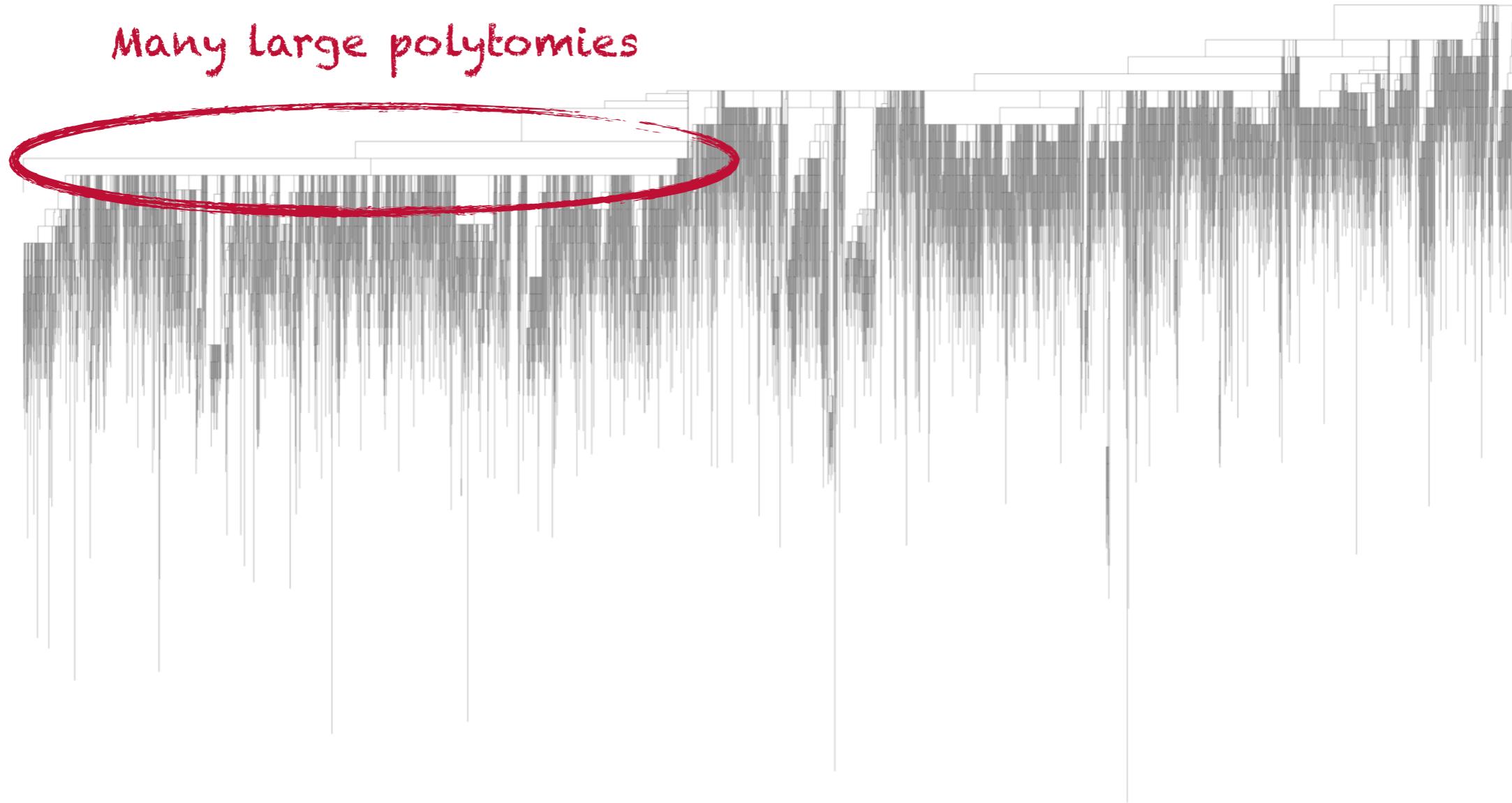
~88,000 SARS-CoV-2 genomes

<https://microreact.org/project/cogconsortium-2020-09-24>



~88,000 SARS-CoV-2 genomes

<https://microreact.org/project/cogconsortium-2020-09-24>



~88,000 SARS-CoV-2 genomes

<https://microreact.org/project/cogconsortium-2020-09-24>

Bayesian phylogenetics


genetic
sequences


time-calibrated
phylogeny


demographic
model


site model


molecular clock
model

$$P(\text{model} \mid \text{data}) = \frac{P(\text{data} \mid \text{model})P(\text{model})}{P(\text{data})}$$

Bayesian phylogenetics

genetic
sequences

time-calibrated
phylogeny

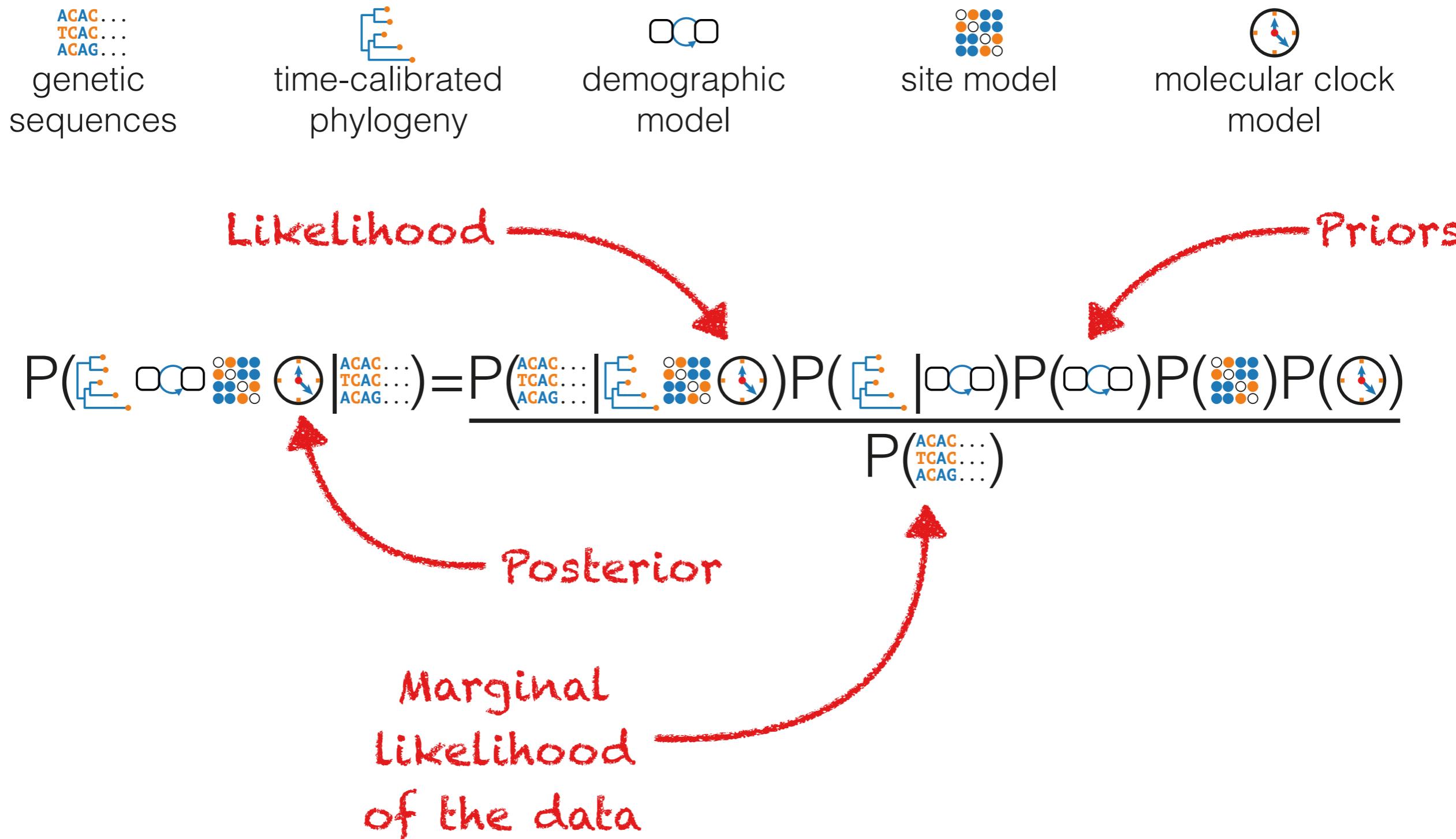
demographic
model

site model

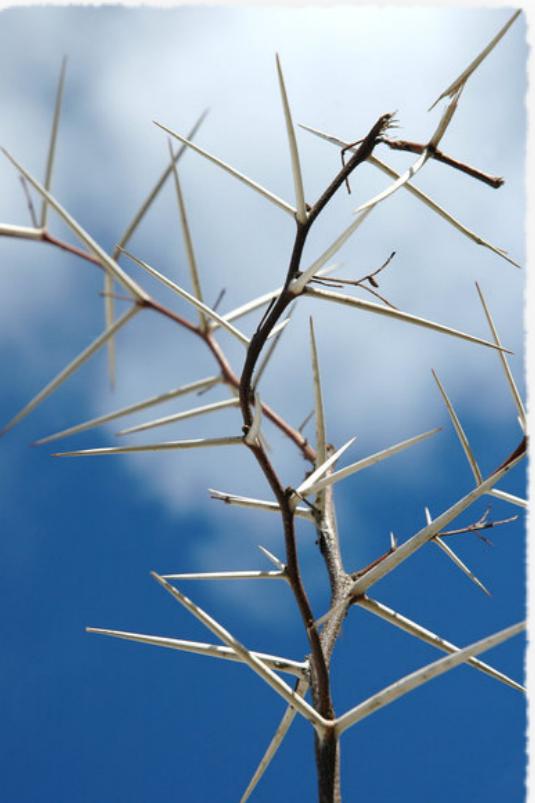
molecular clock
model

$$P(\text{E} \mid \text{ACAC...}, \text{TCAC...}, \text{ACAG...}) = \frac{P(\text{ACAC...} \mid \text{E}) P(\text{TCAC...} \mid \text{E}) P(\text{ACAG...} \mid \text{E})}{P(\text{ACAC...})}$$

Bayesian phylogenetics



Efficiently analysing huge datasets in BEAST



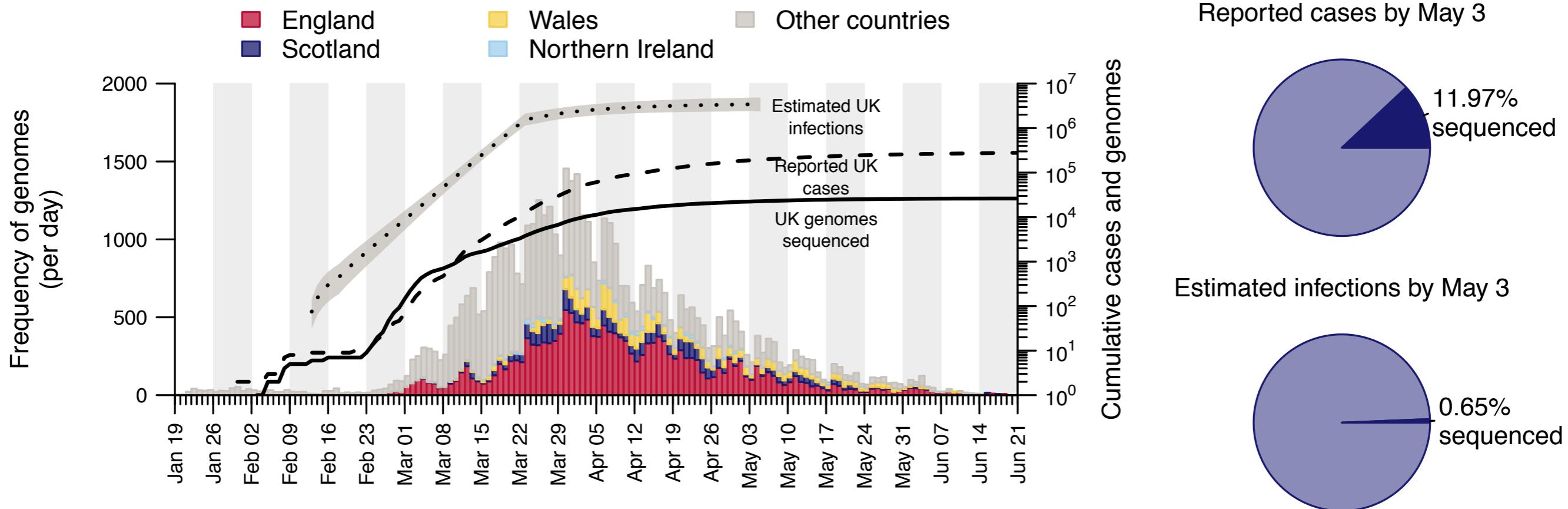
Thorney BEAST (by JT McCrone & Andrew Rambaut)

- Scale branches on a guide tree
- Only integrate over polytomy resolutions
- More efficient data structures

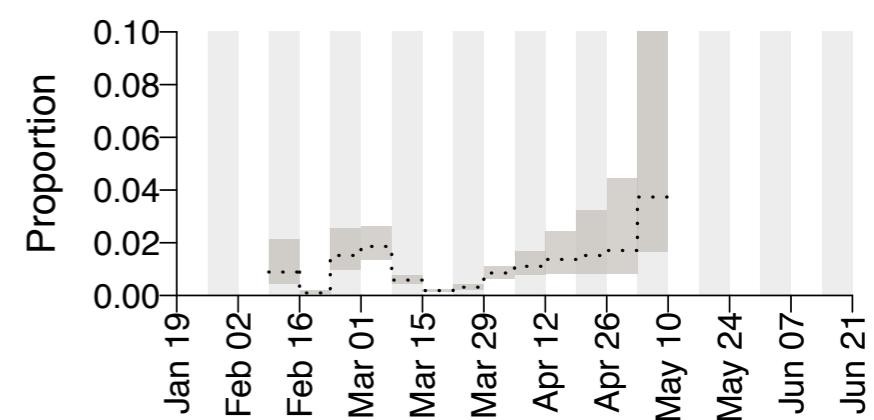
https://beast.community/thorney_beast

Sampling of SARS-CoV-2 genomes

First 6 months of the pandemic



- **50,887** SARS-CoV-2 genomes
- **26,181** from the UK (51.5%)



Definition: UK transmission lineages

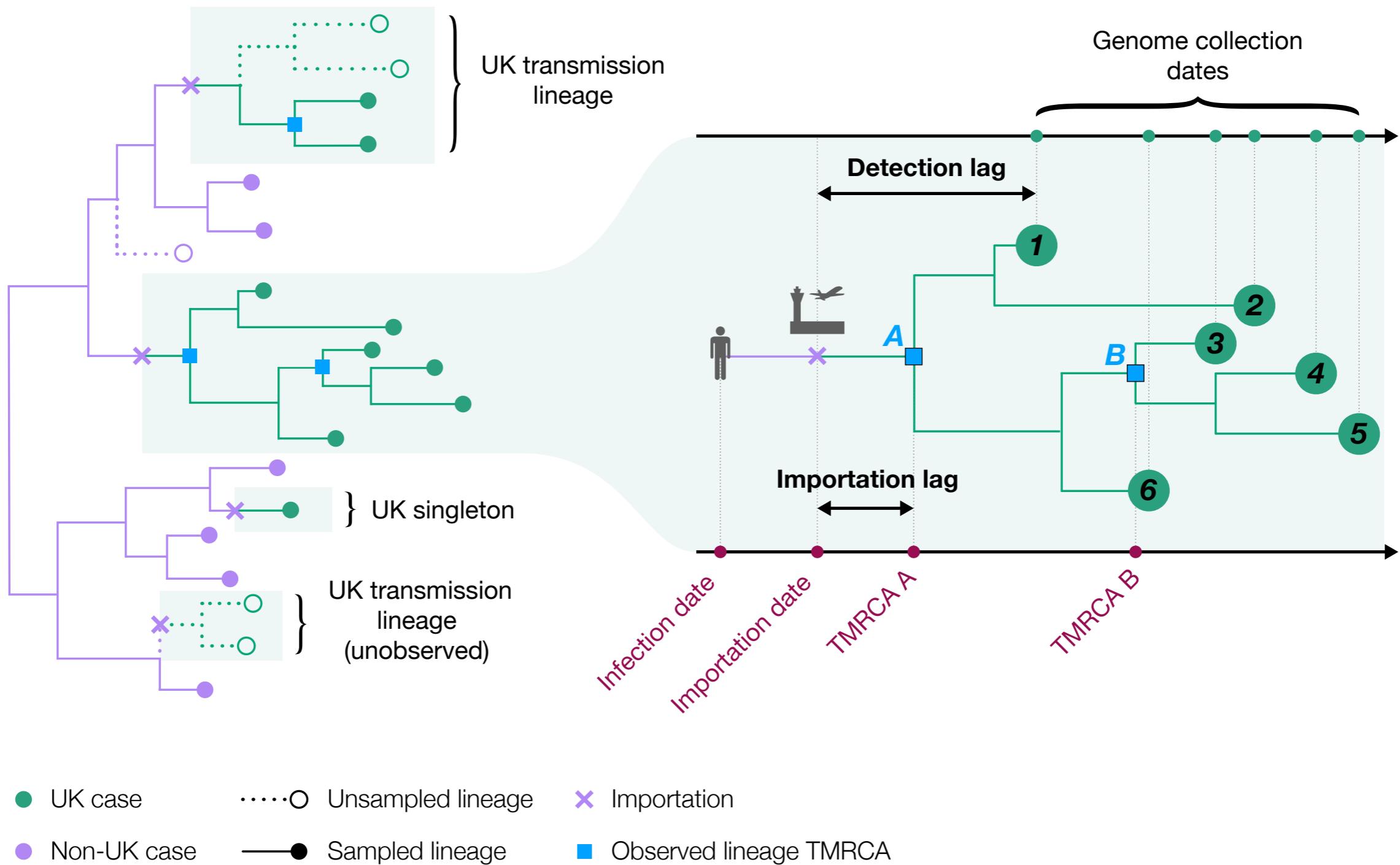


figure by Jayna Raghwani

Definition: UK transmission lineages

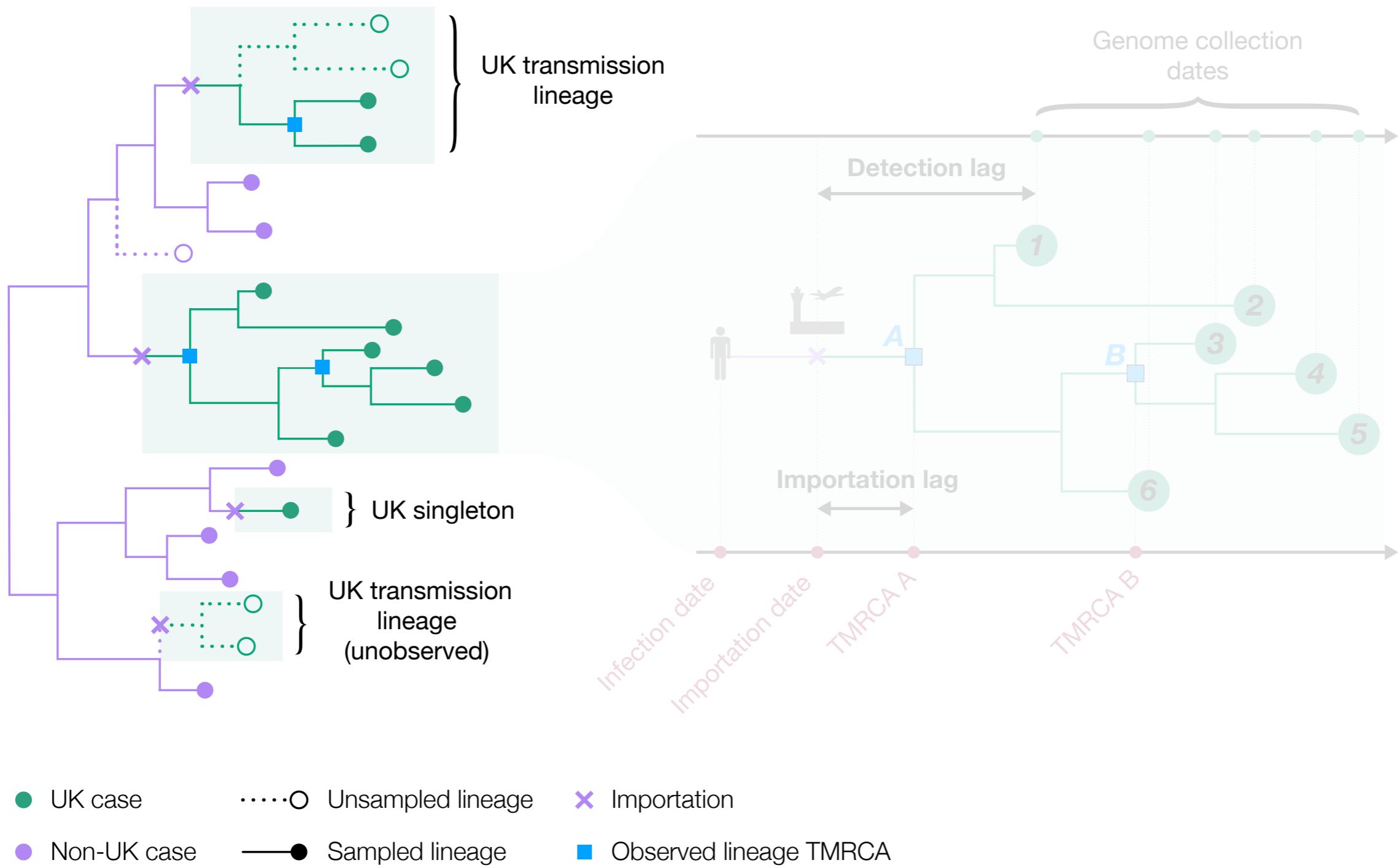
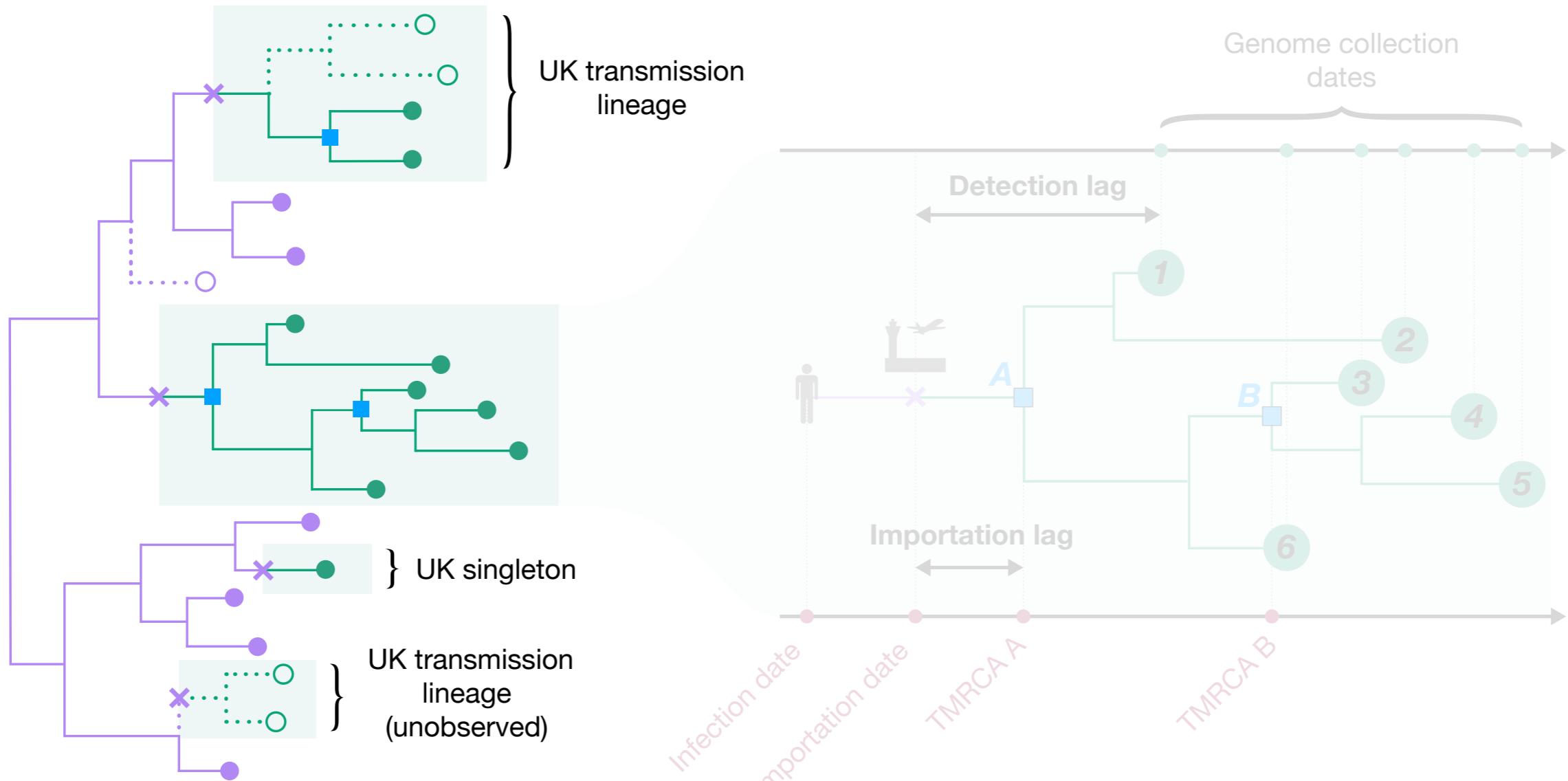


figure by Jayna Raghwani

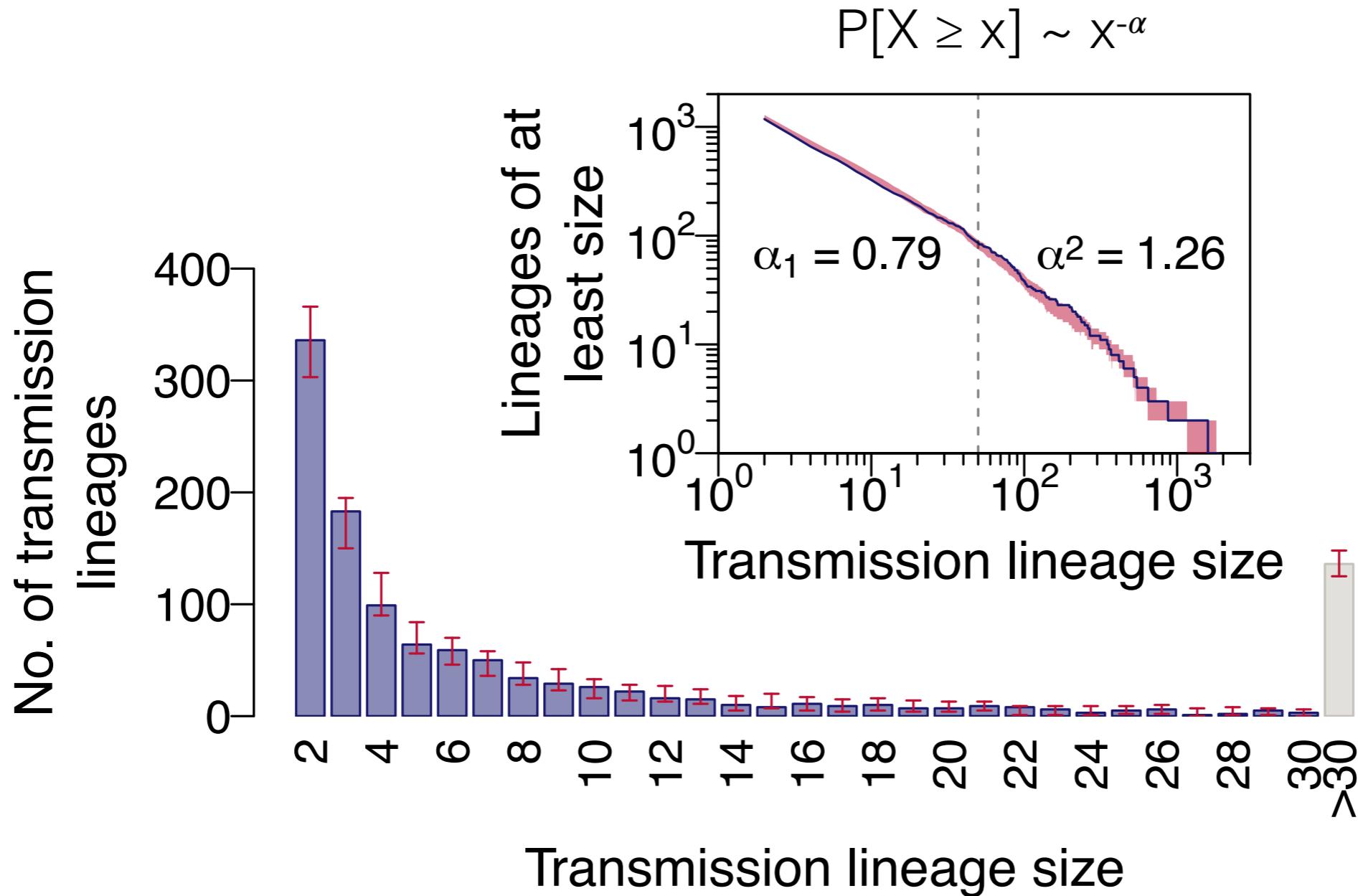
Definition: UK transmission lineages



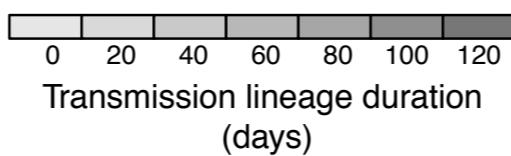
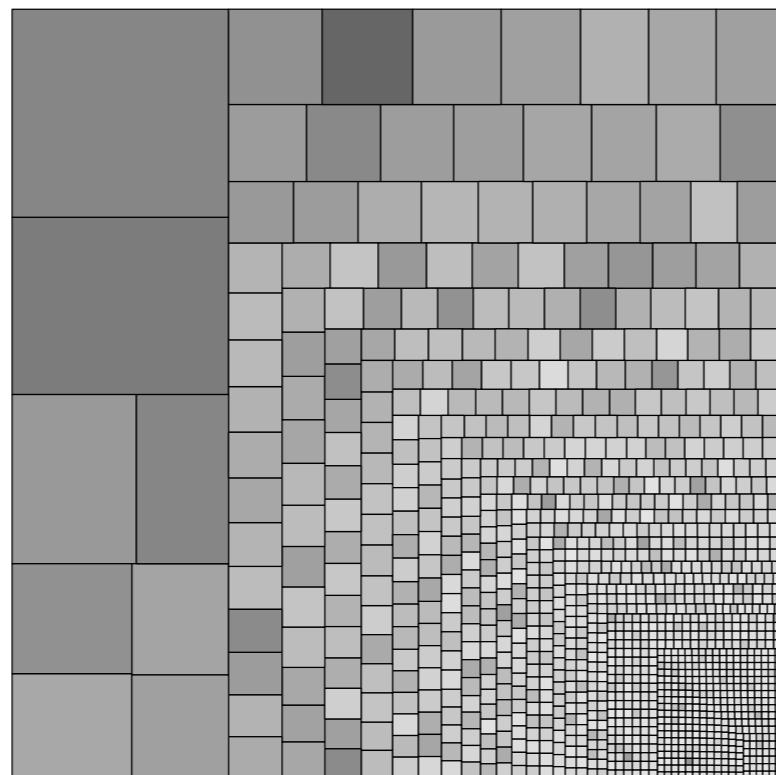
UK transmission lineages ≠ Pango lineages

Sizes of UK transmission lineages

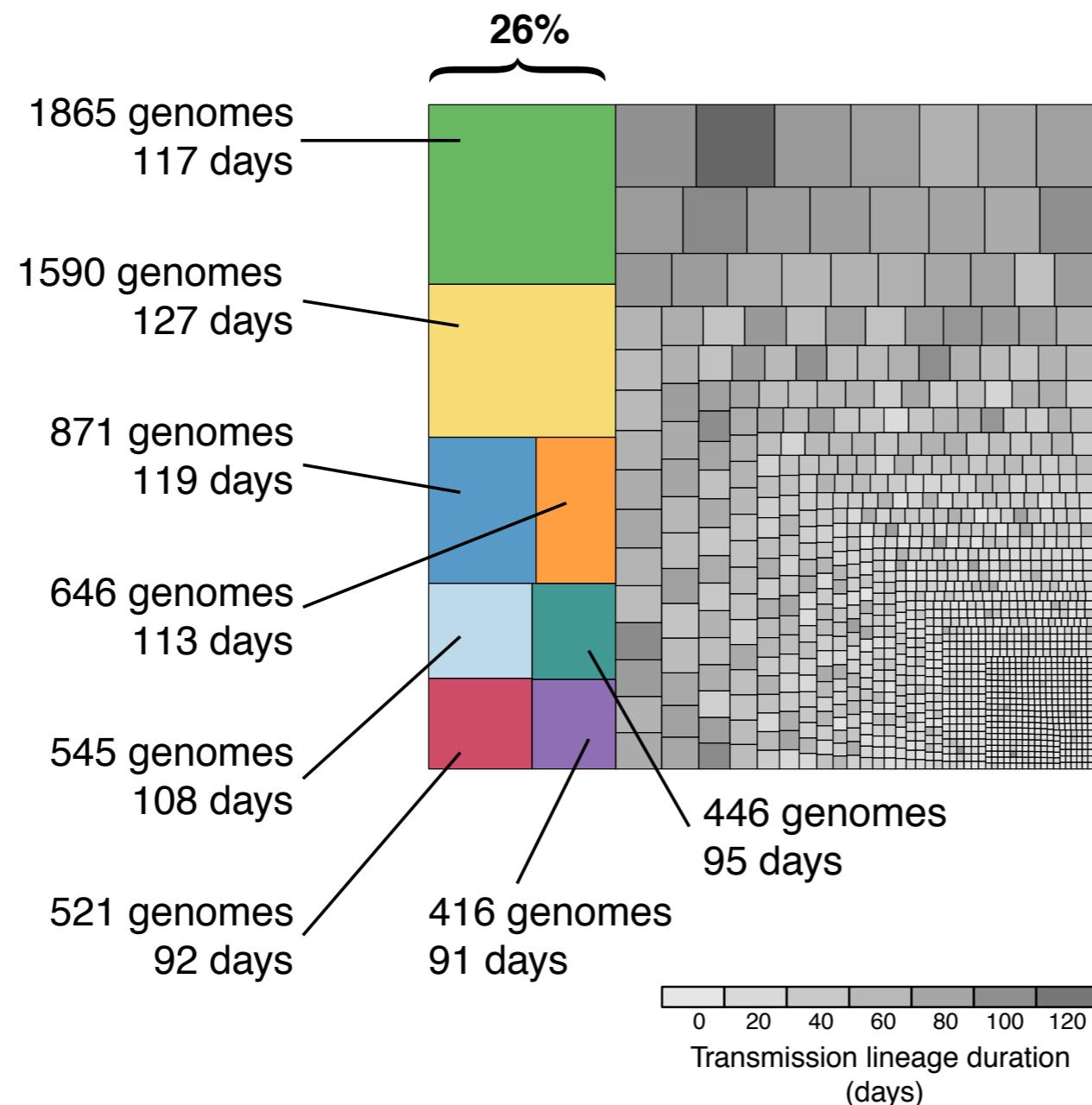
26,181 UK genomes into **1179** lineages and **1650** singletons



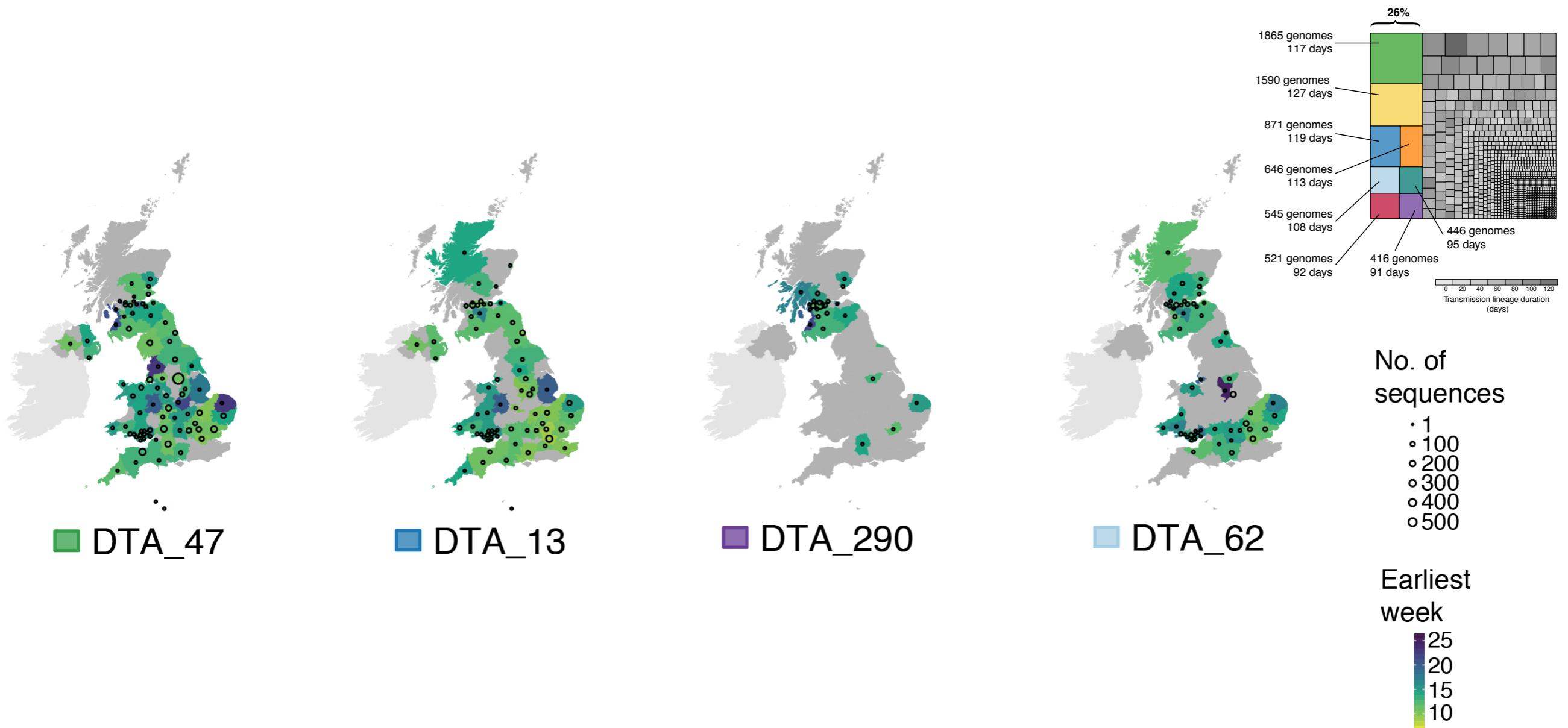
UK transmission lineage breakdown



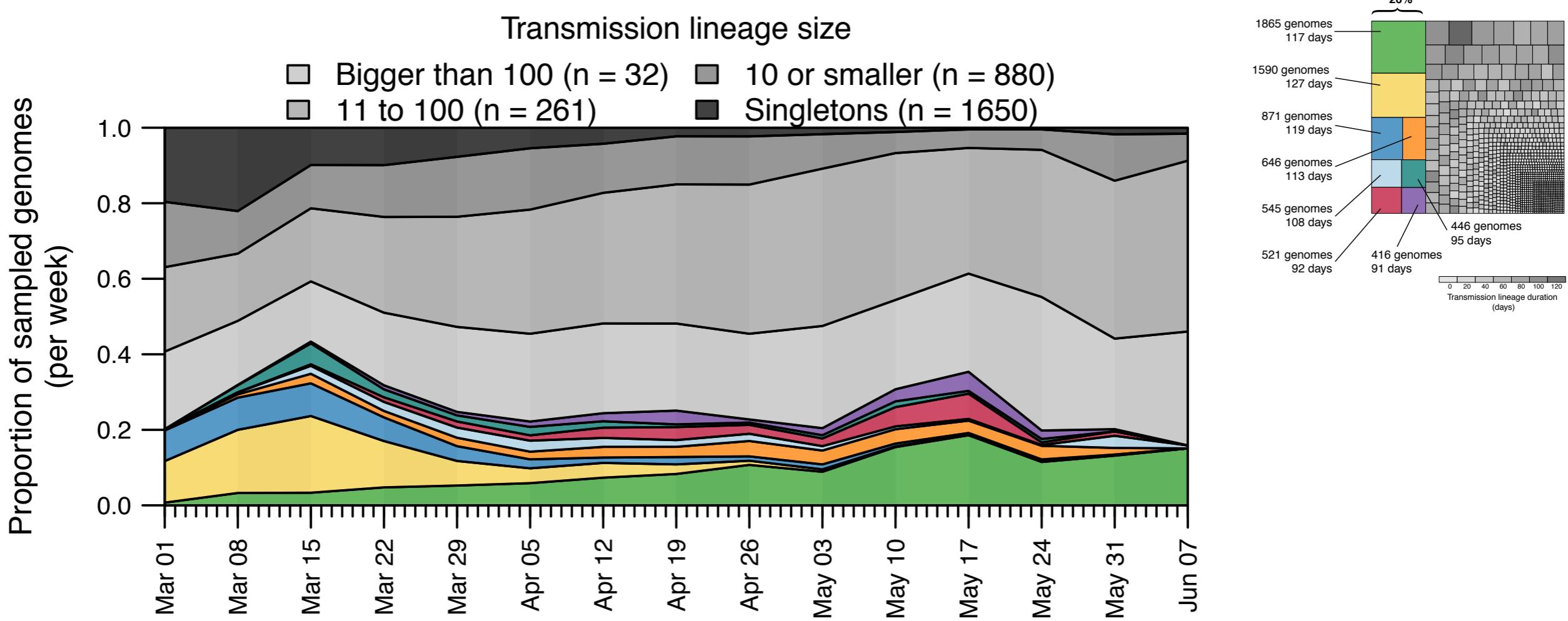
UK transmission lineage breakdown



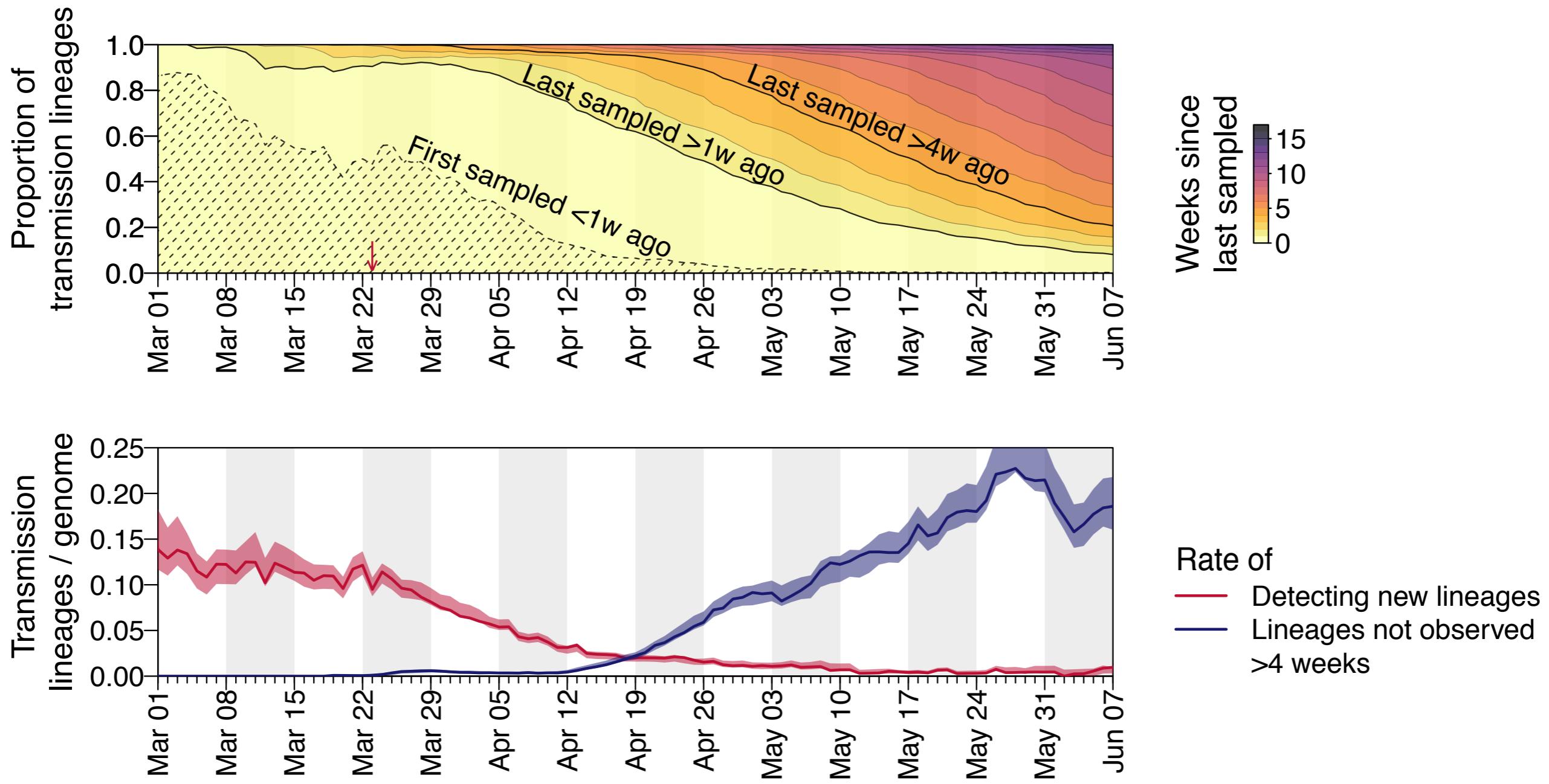
Geographic spread of transmission lineages



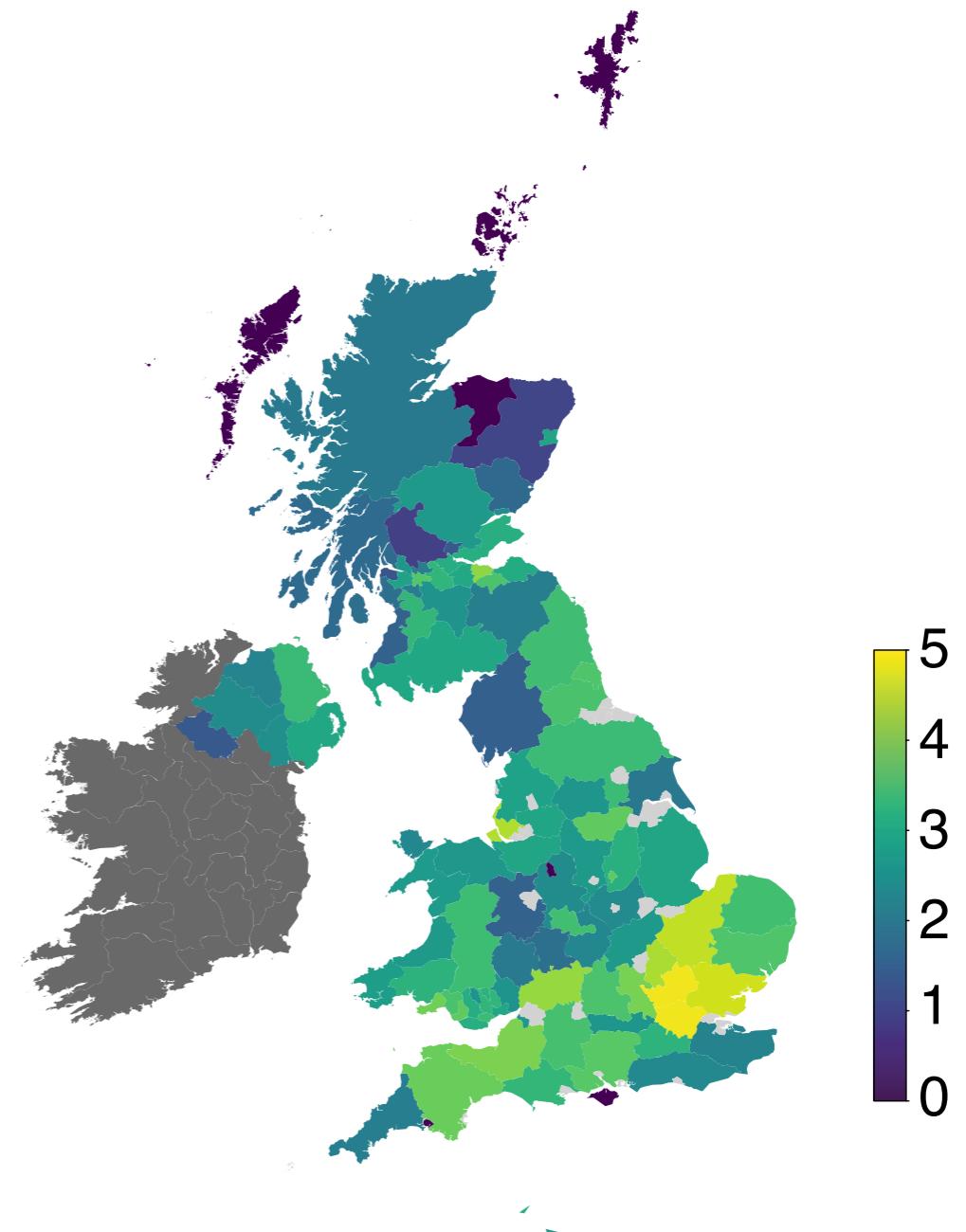
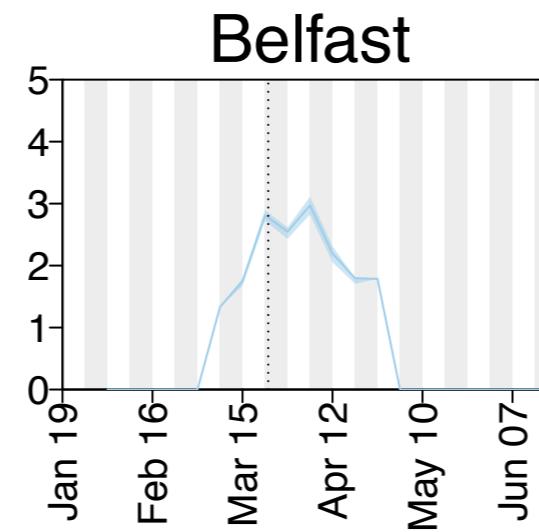
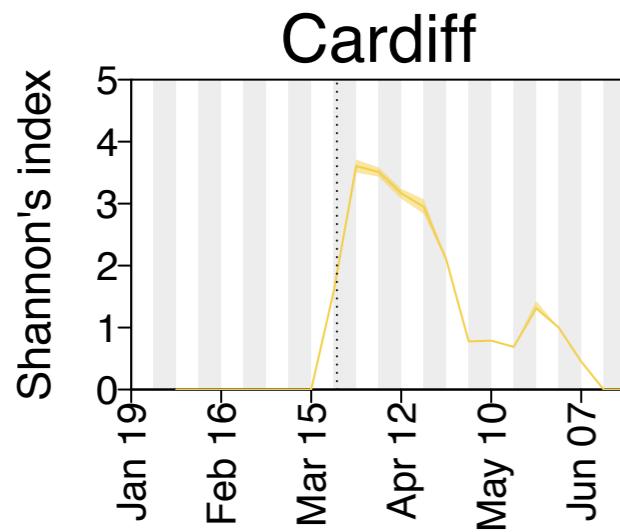
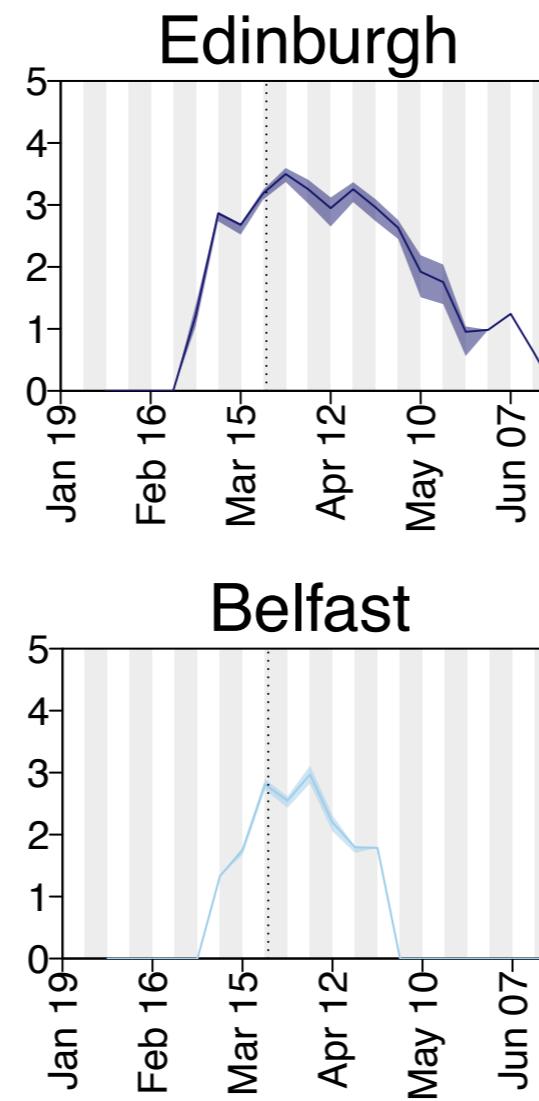
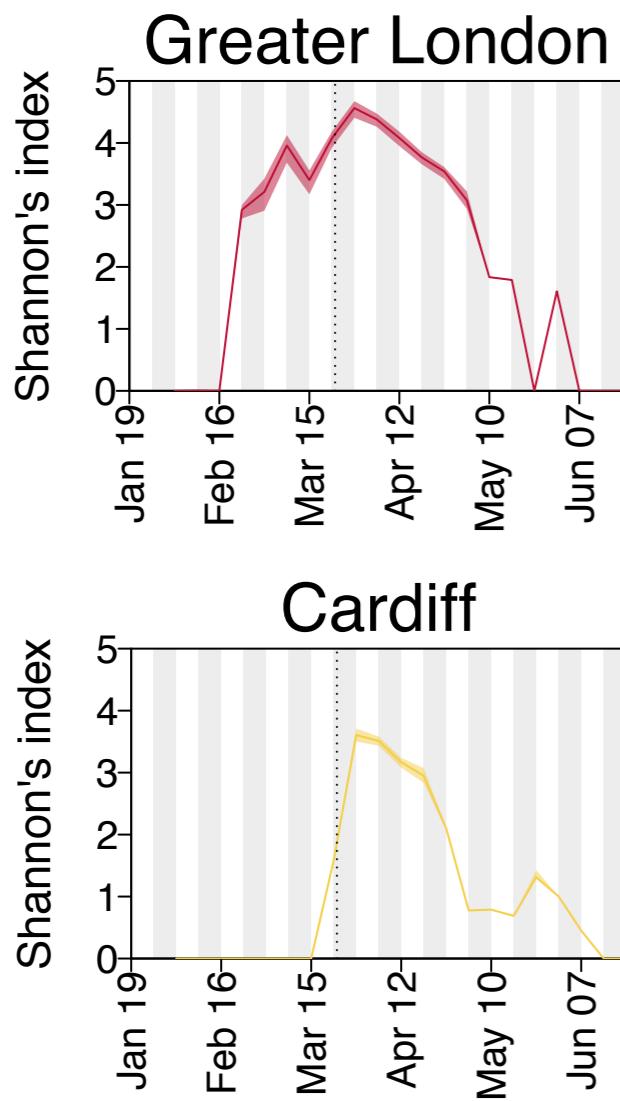
Sample breakdown by UK transmission lineage



Detection of UK transmission lineages



Shannon's index for lineage diversity



Definition: TMRCA and importation lag

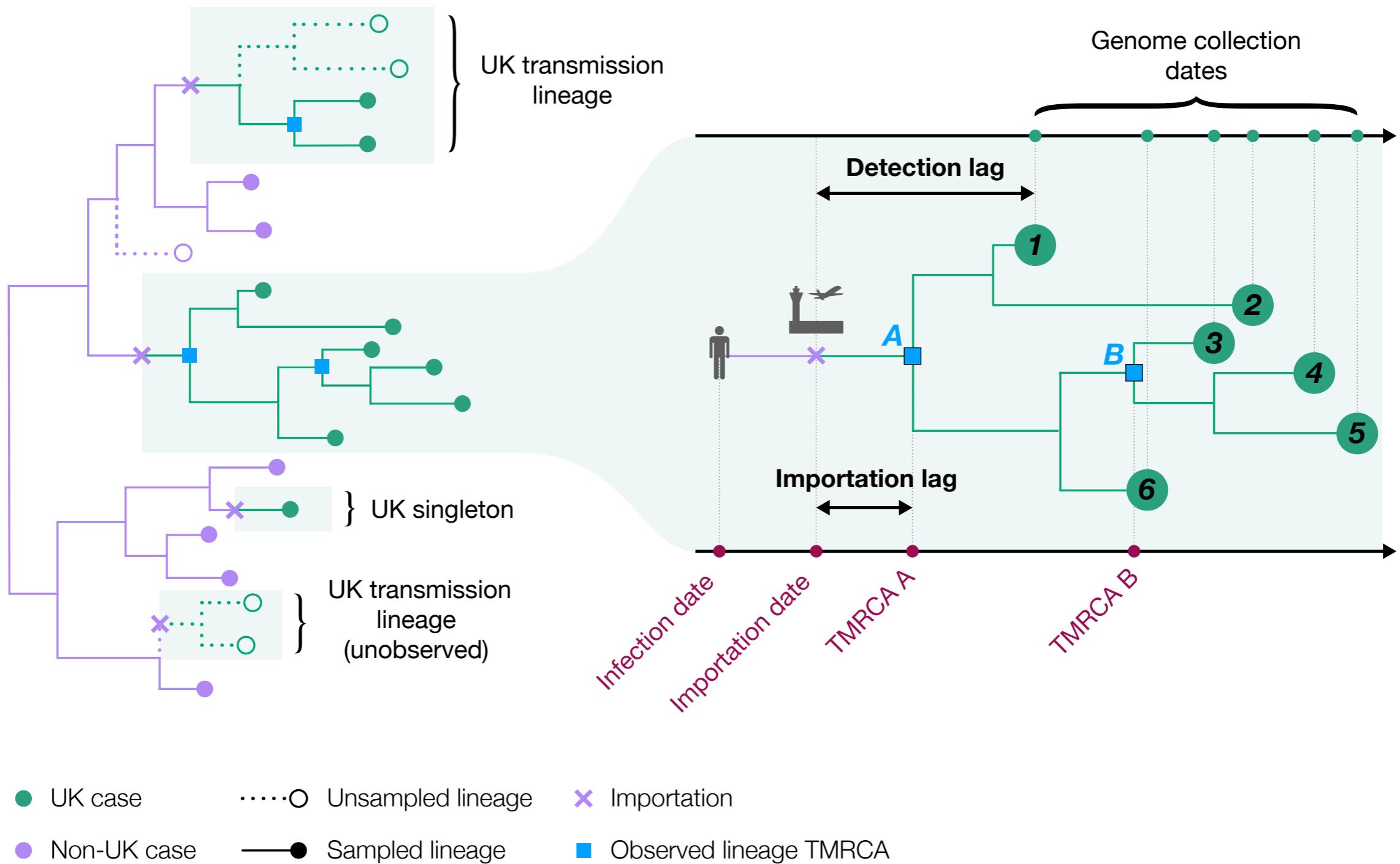


figure by Jayna Raghwani

Definition: TMRCA and importation lag

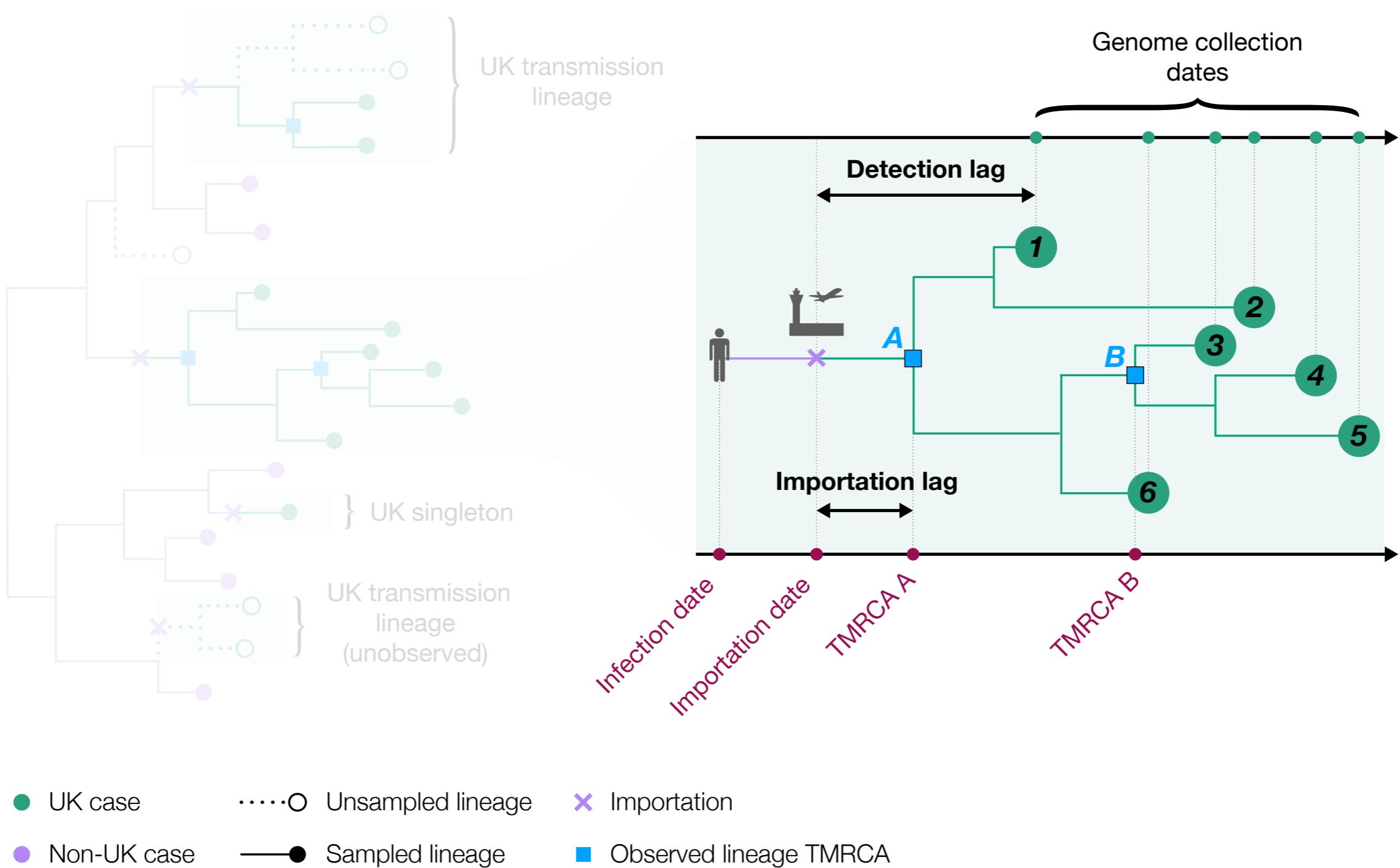
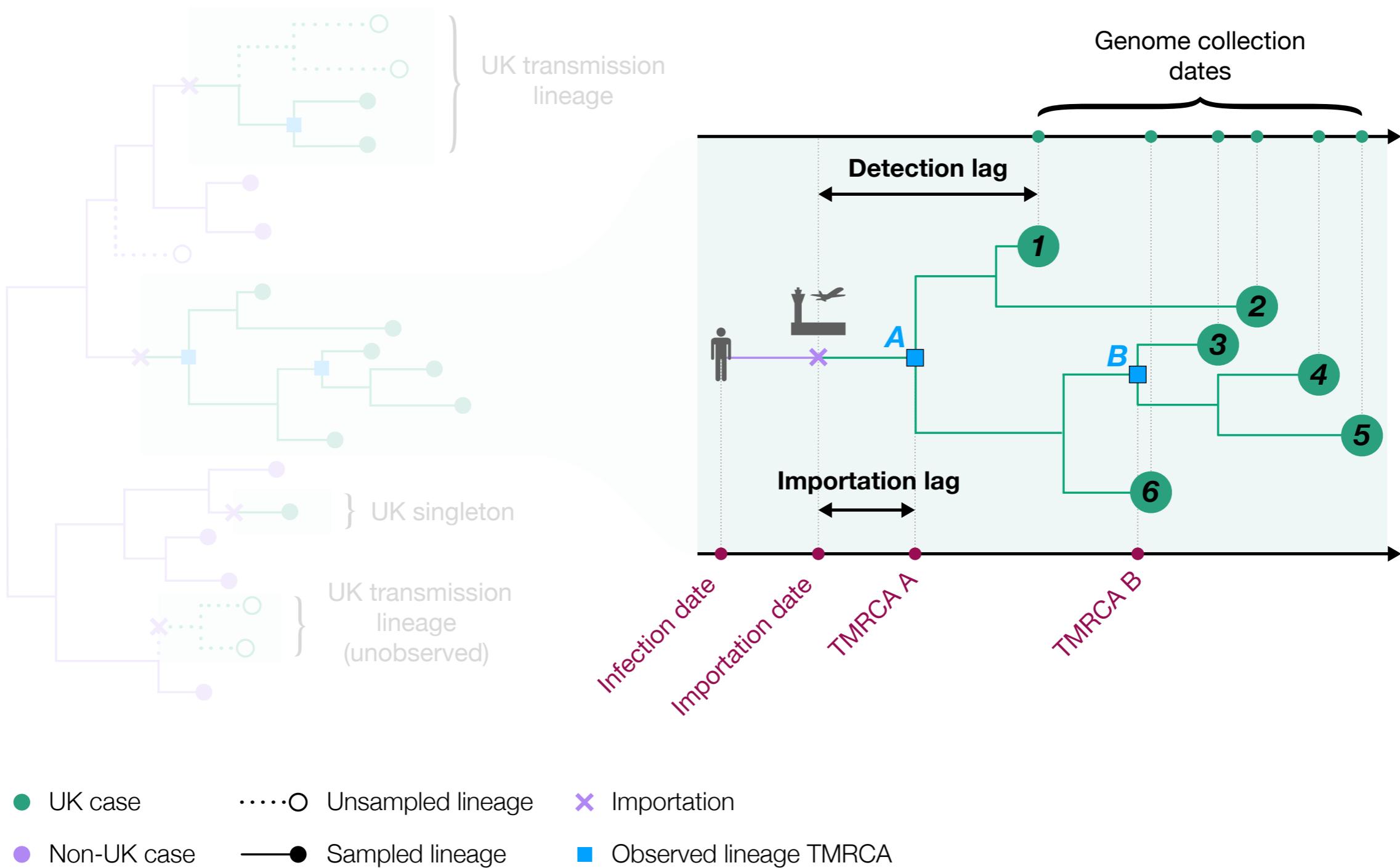


figure by Jayna Raghwani

time to the most recent
common ancestor

Definition: TMRCA and importation lag



Definition: TMRCA and importation lag

time to the most recent common ancestor

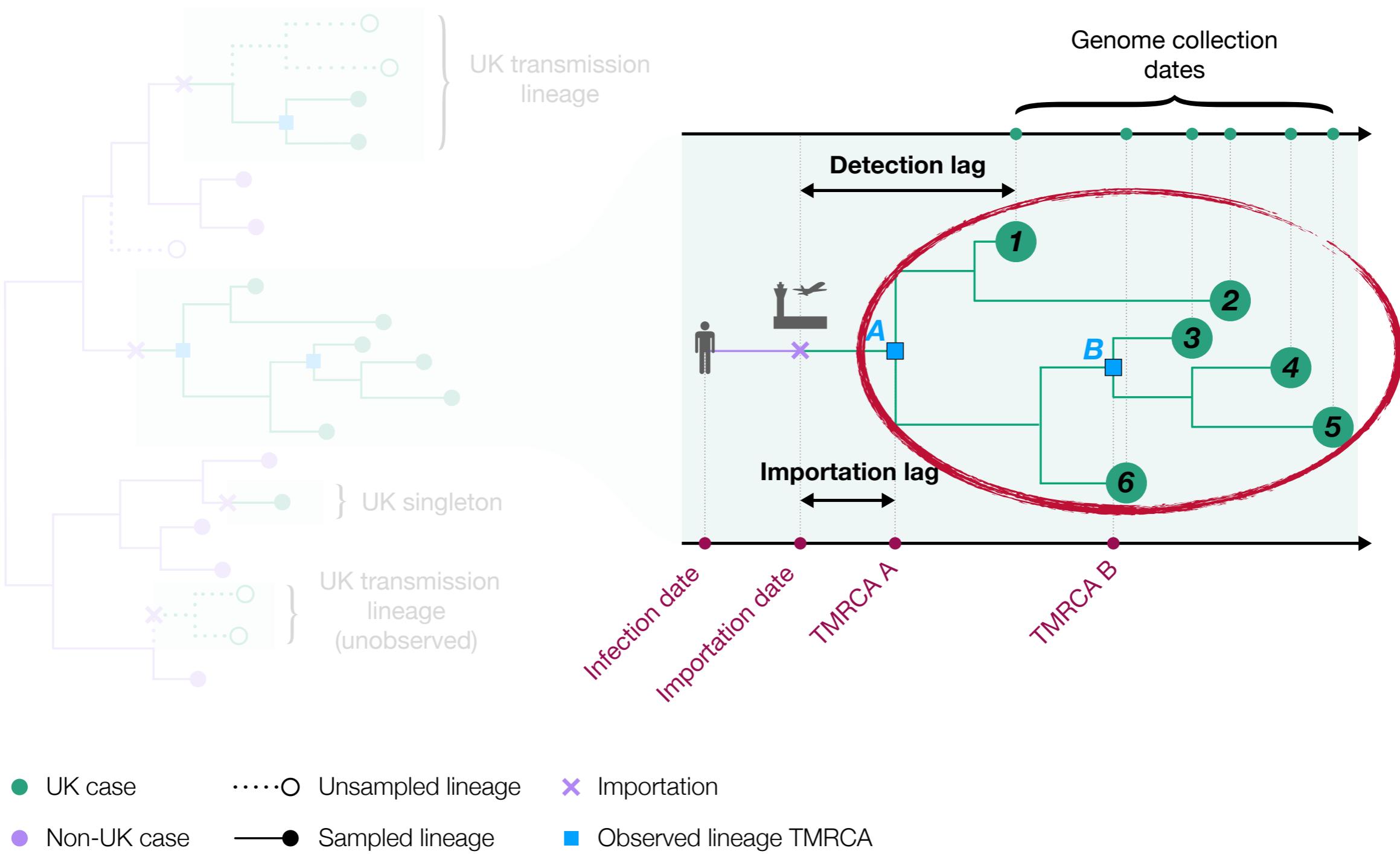


figure by Jayna Raghwani

time to the most recent
common ancestor

Definition: TMRCA and importation lag

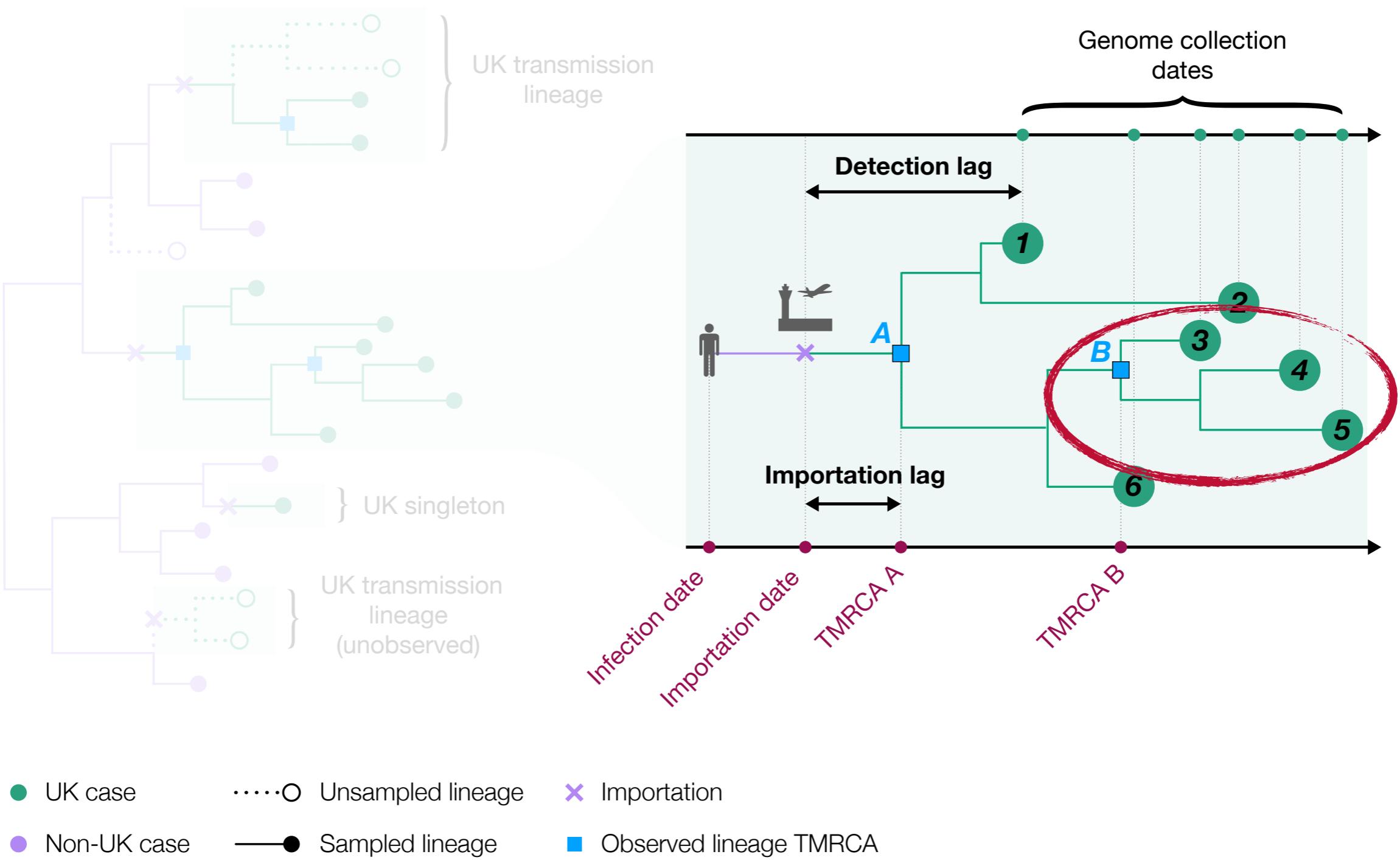


figure by Jayna Raghwani

time to the most recent
common ancestor

Definition: TMRCA and importation lag

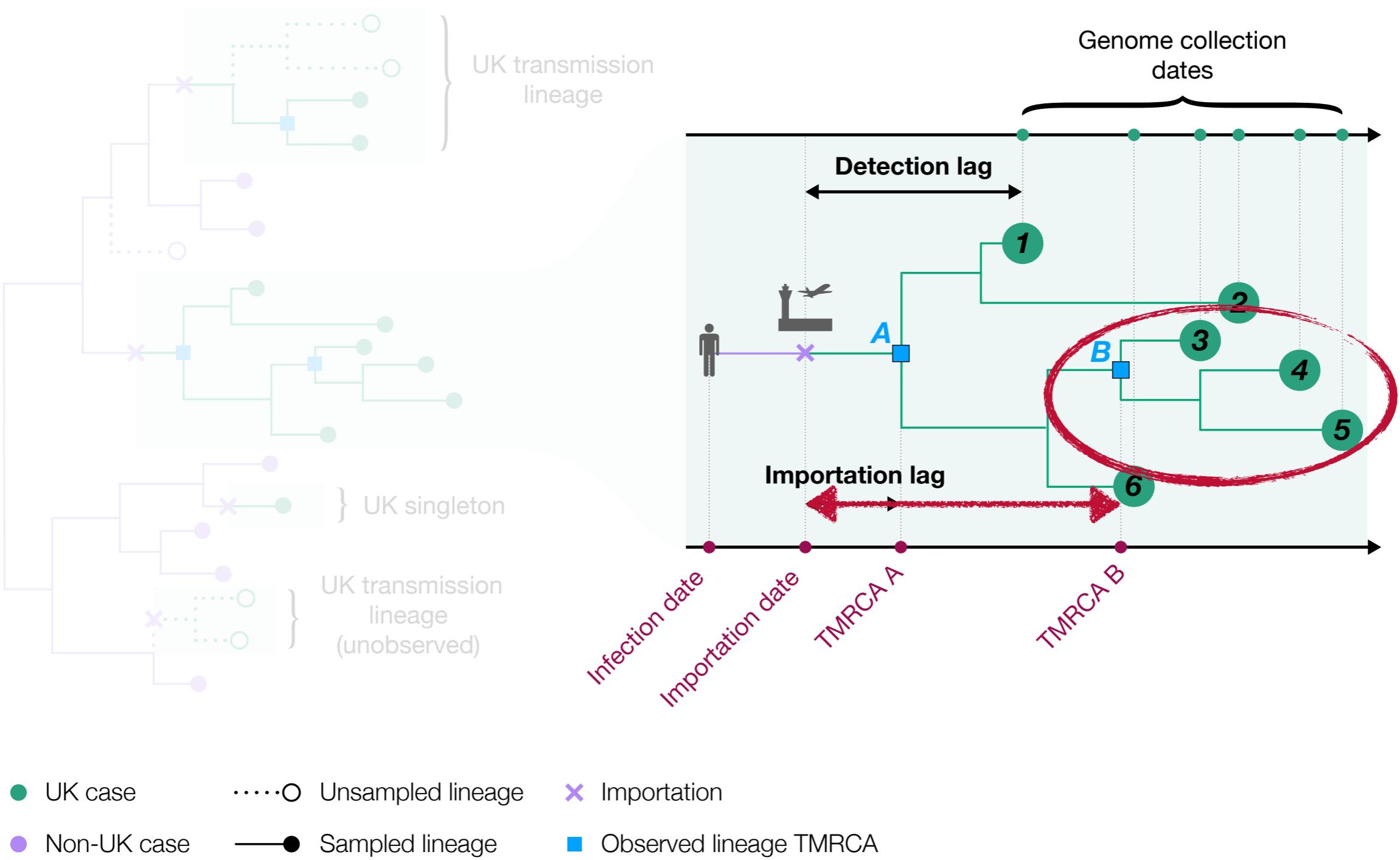
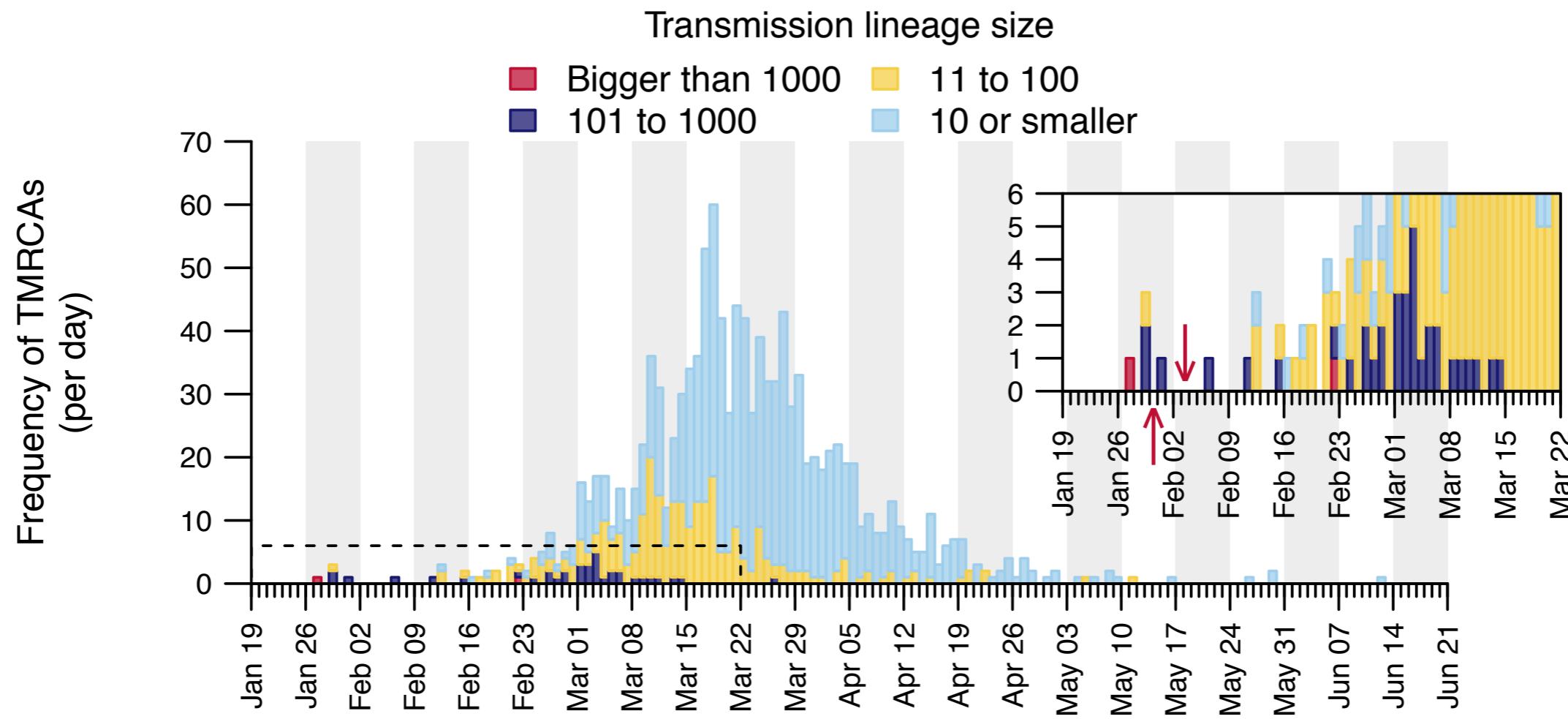


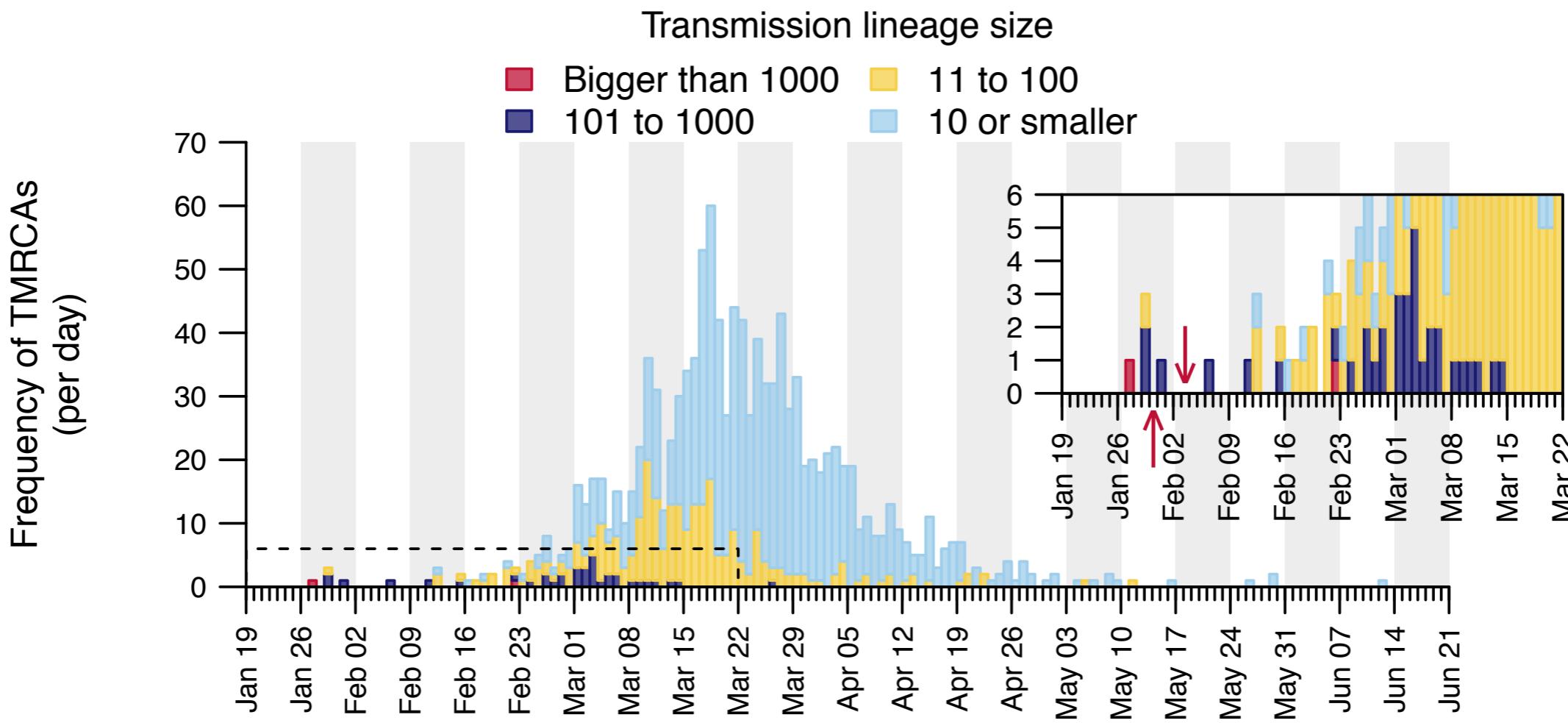
figure by Jayna Raghwani

Estimated TMRCAs of UK transmission lineages



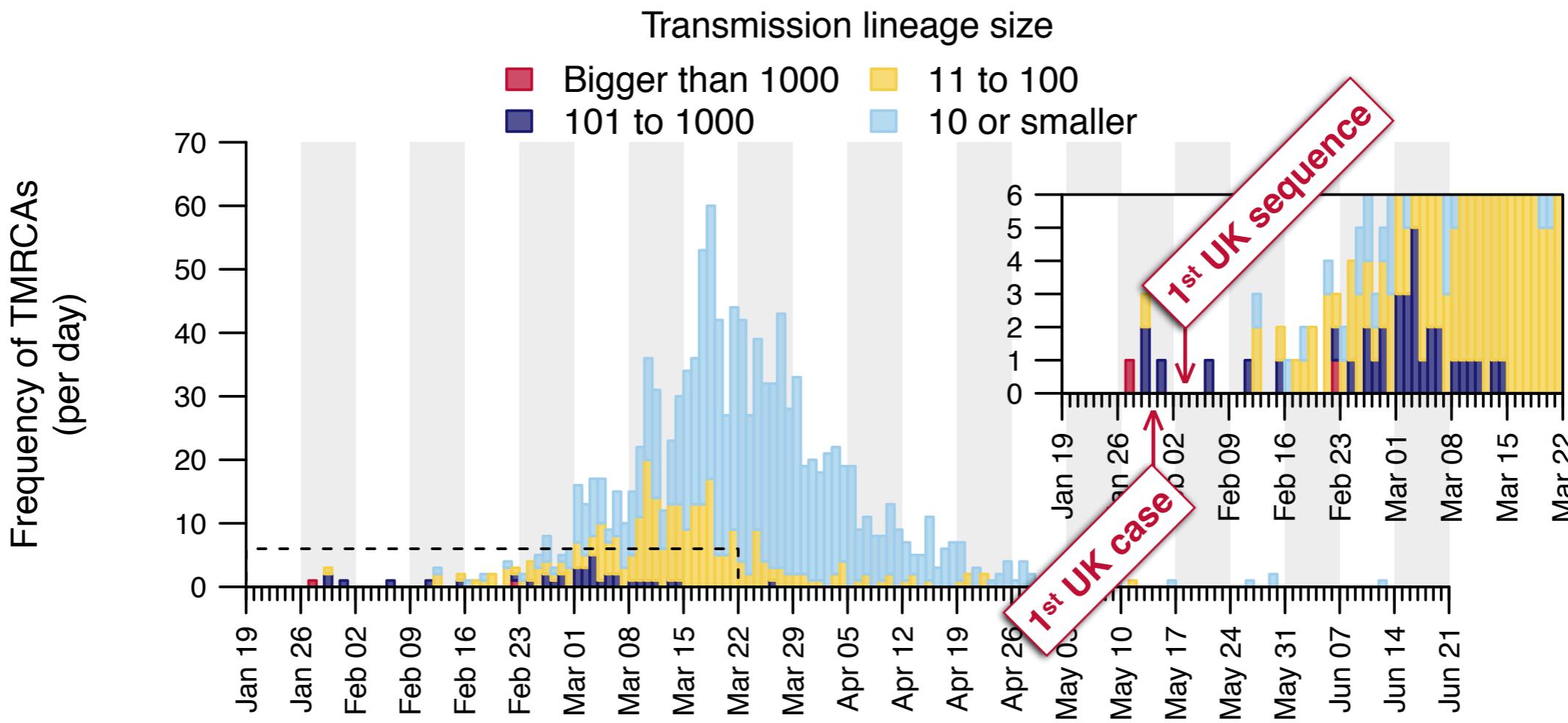
time to the most recent
common ancestor

Estimated TMRCAs of UK transmission lineages

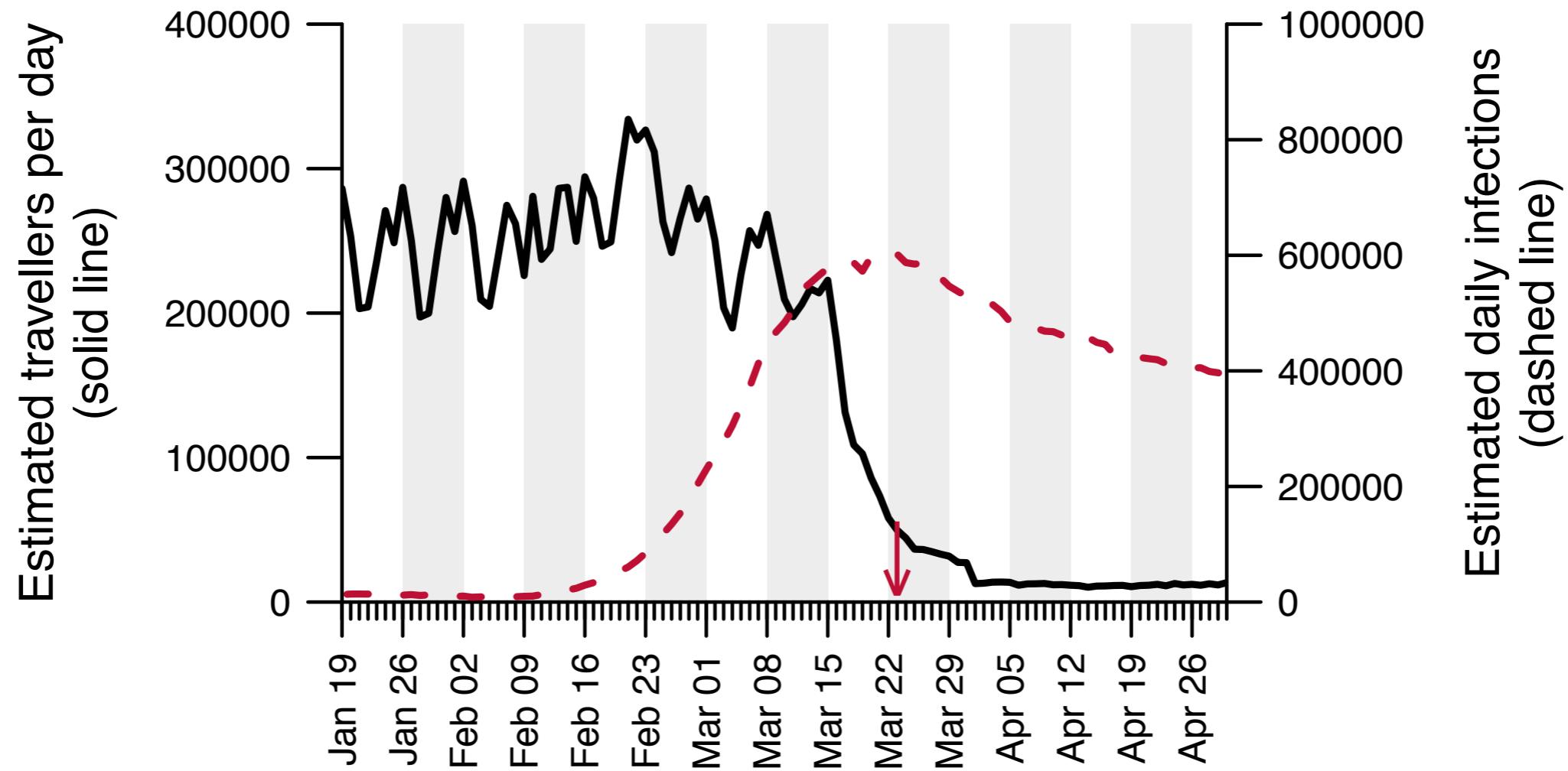


time to the most recent
common ancestor

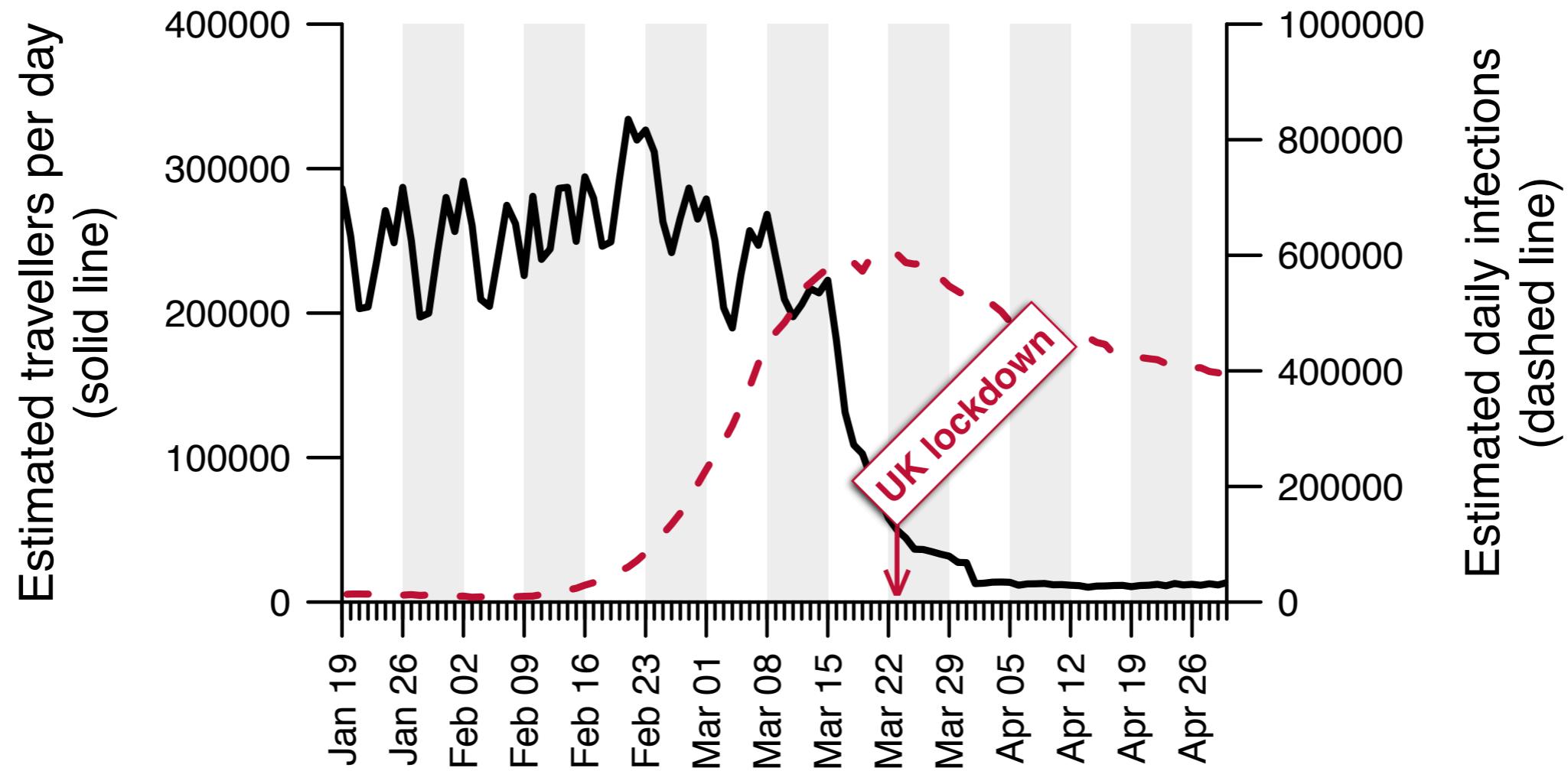
Estimated TMRCAs of UK transmission lineages



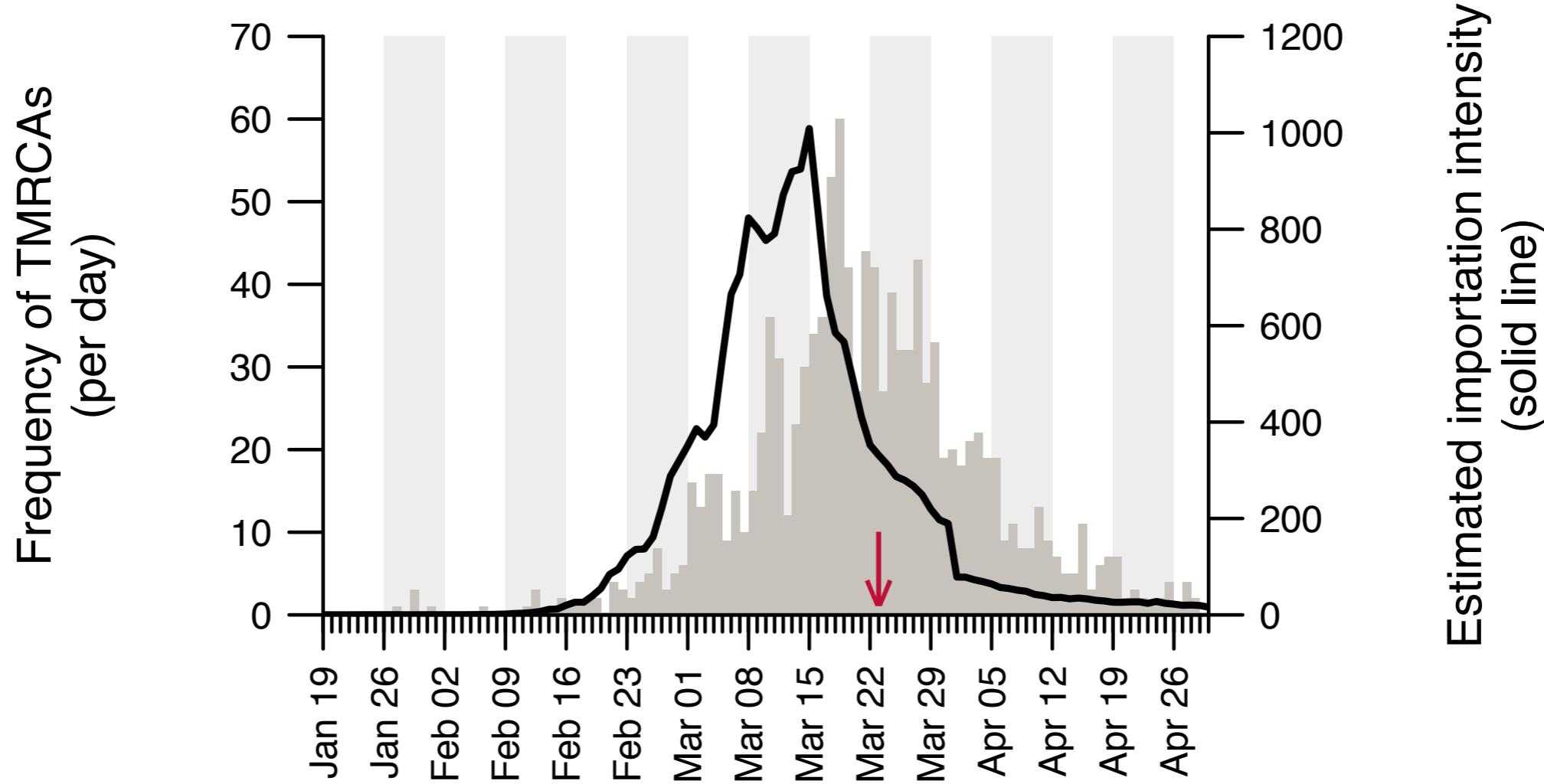
Trends in inbound travel and COVID-19 cases



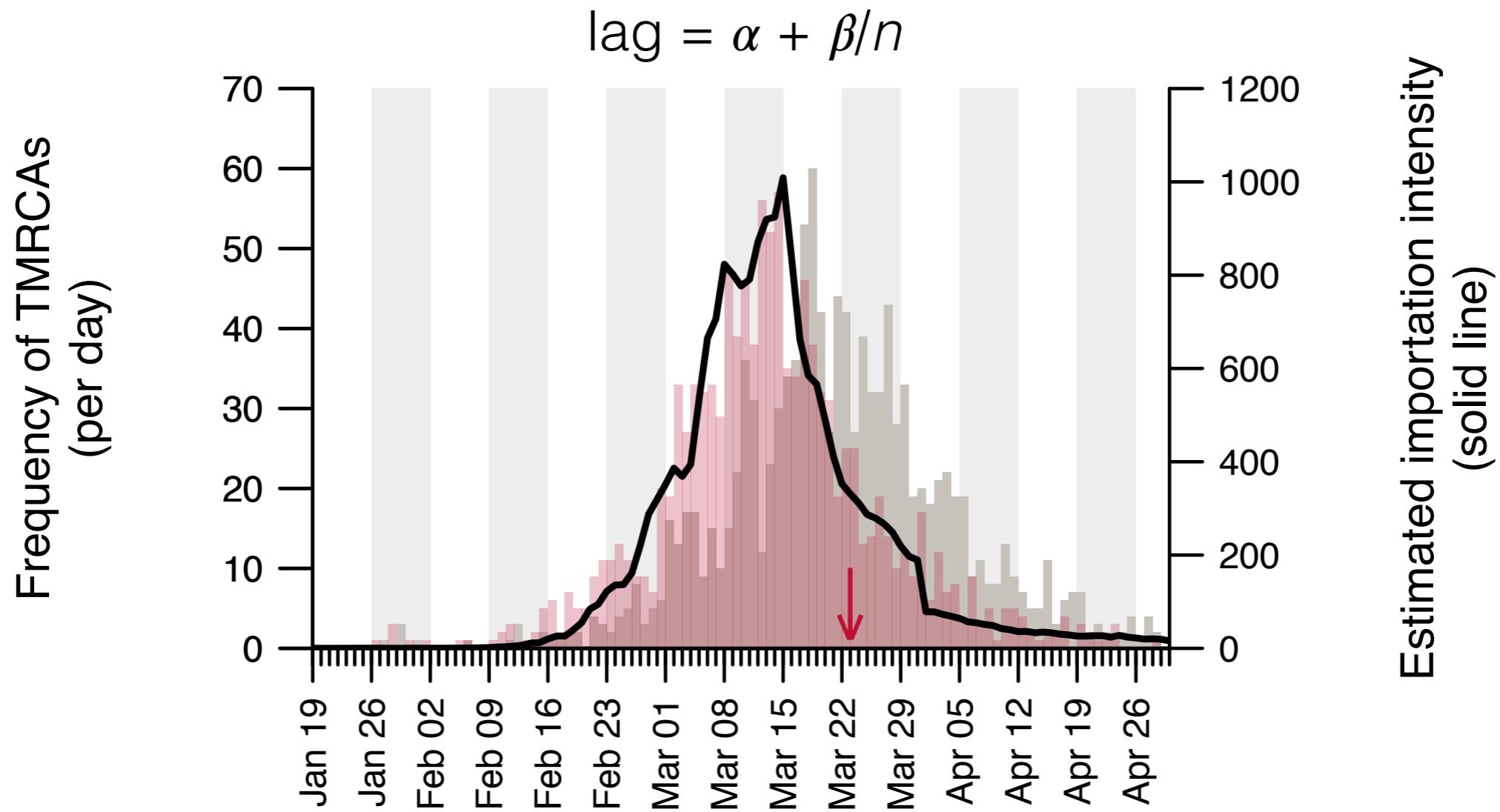
Trends in inbound travel and COVID-19 cases



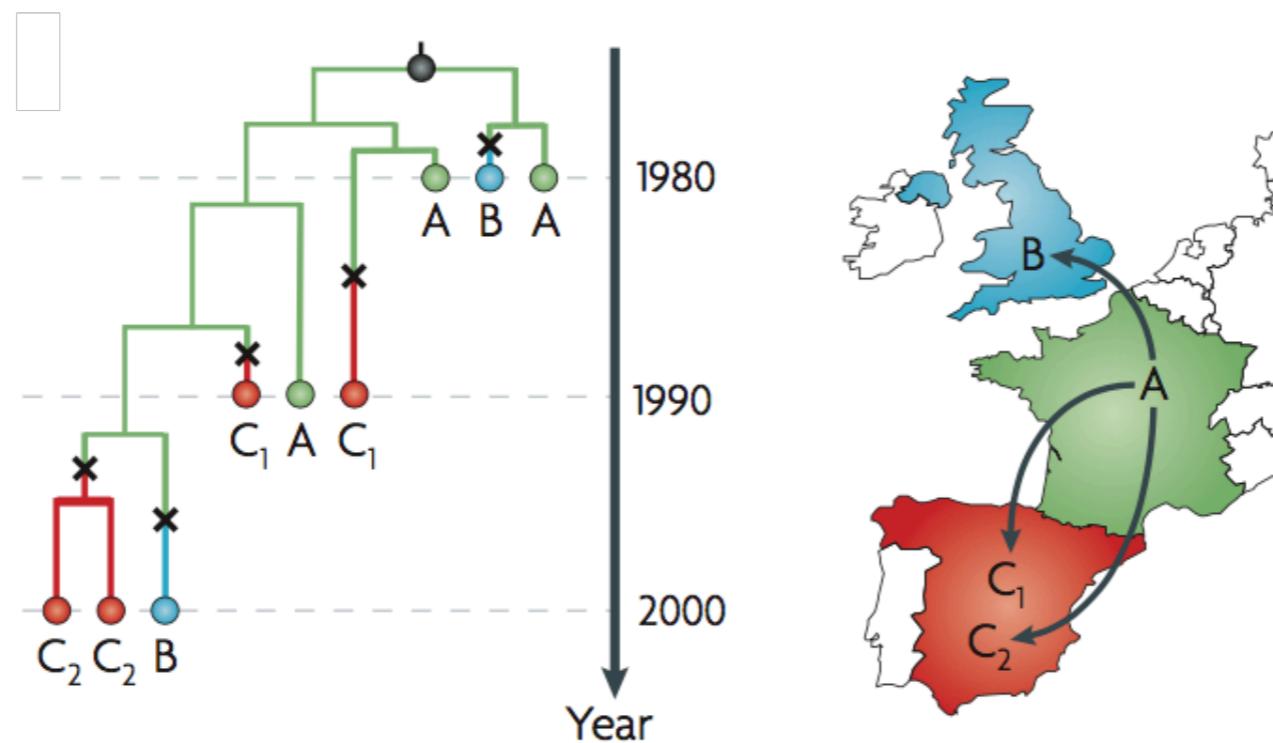
Estimated Importation Intensity



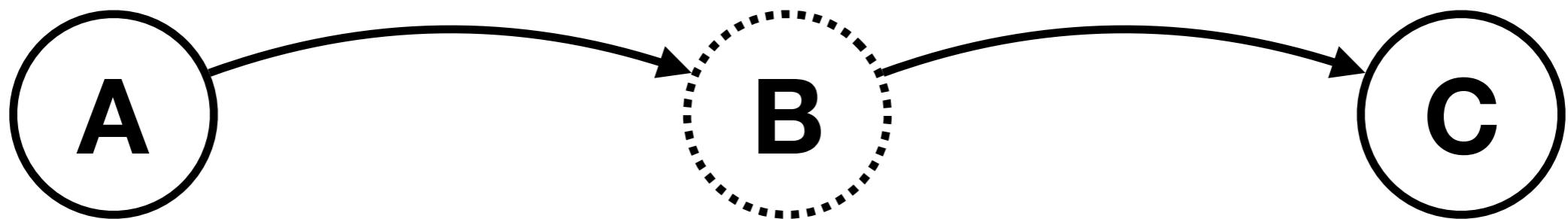
Estimated Importation Intensity



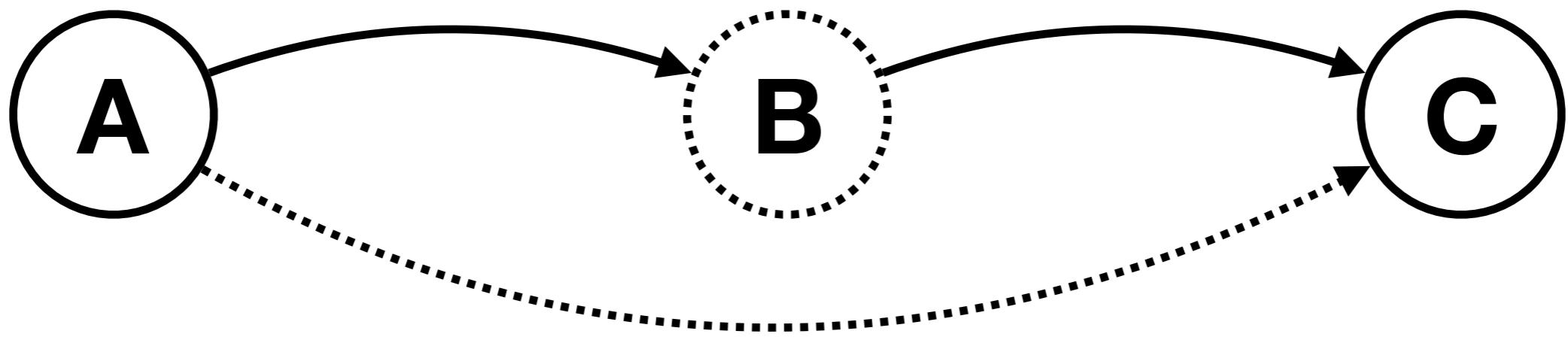
Aside: Phylogeography



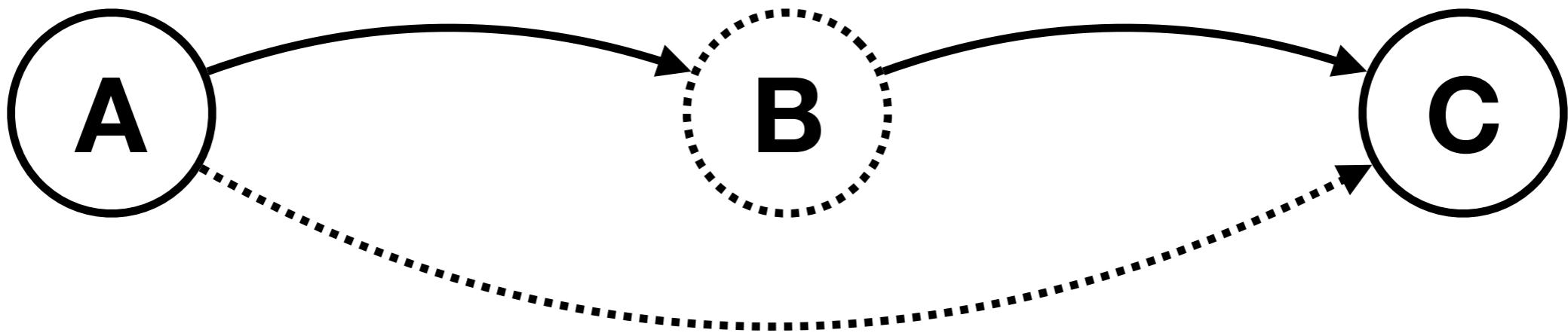
3-body problem



3-body problem



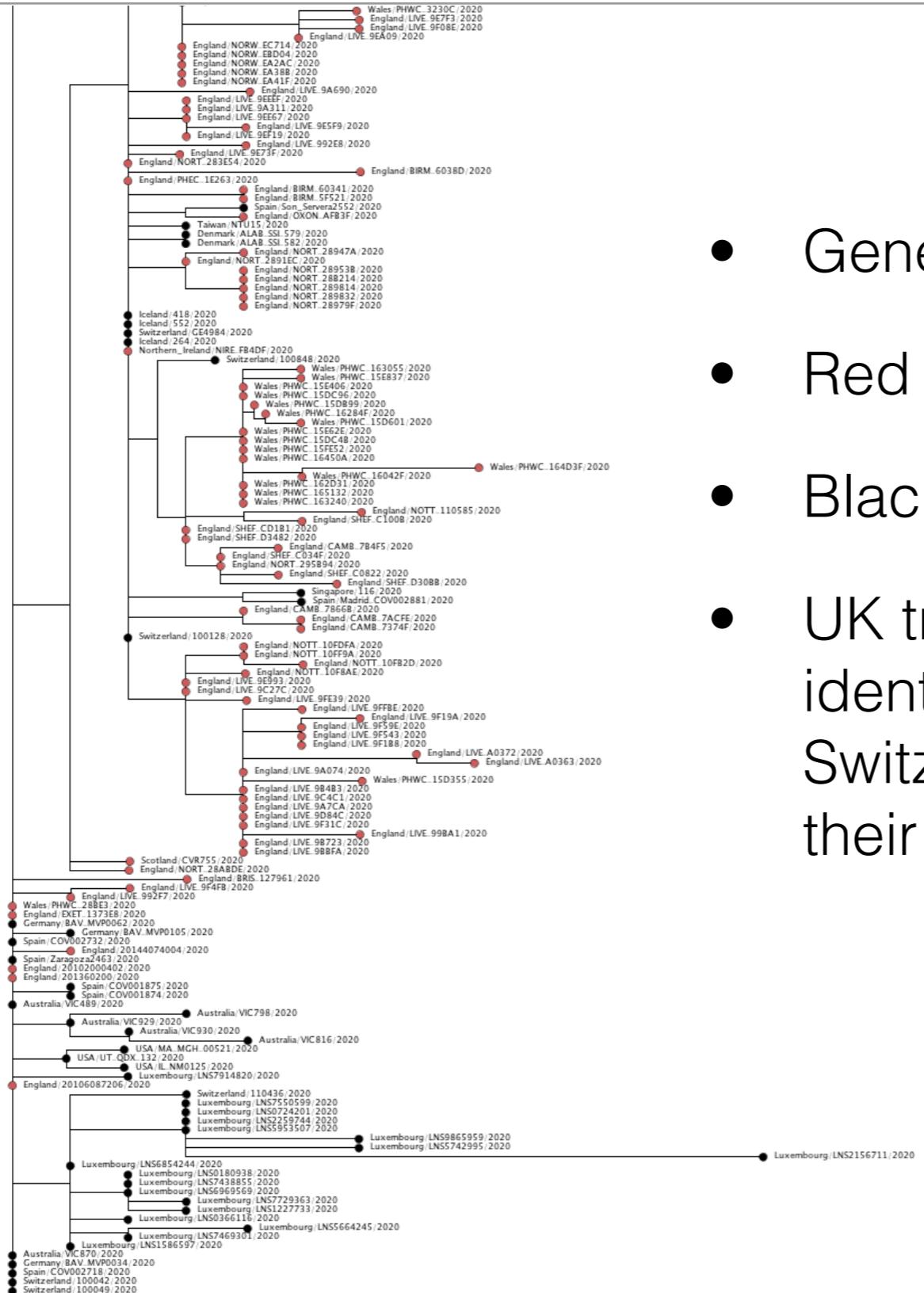
3-body problem



Cases, sequences and genomes until 5 May 2020

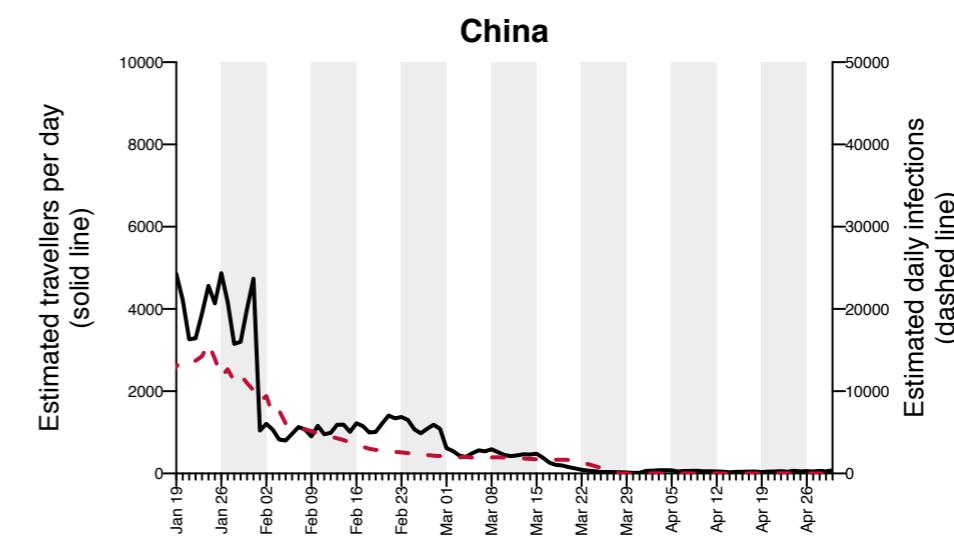
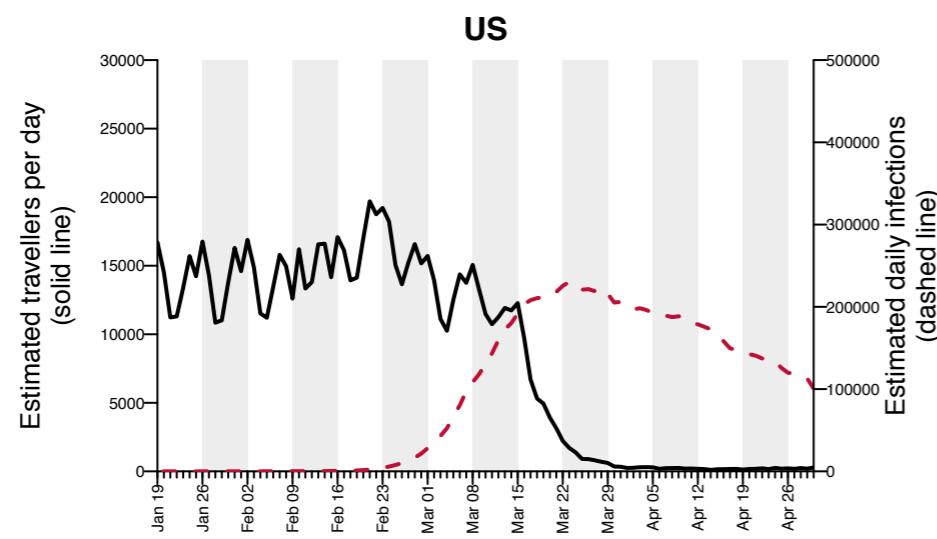
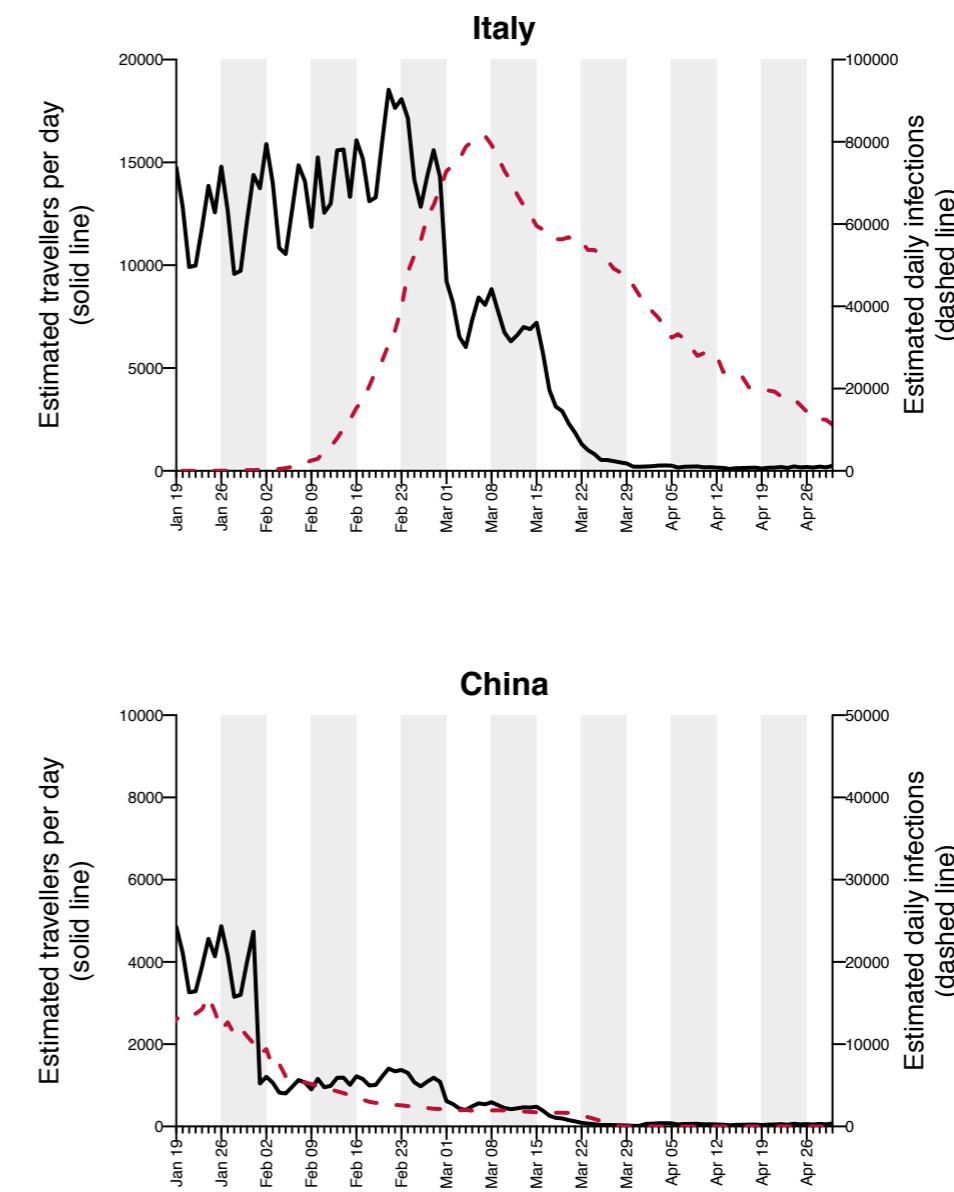
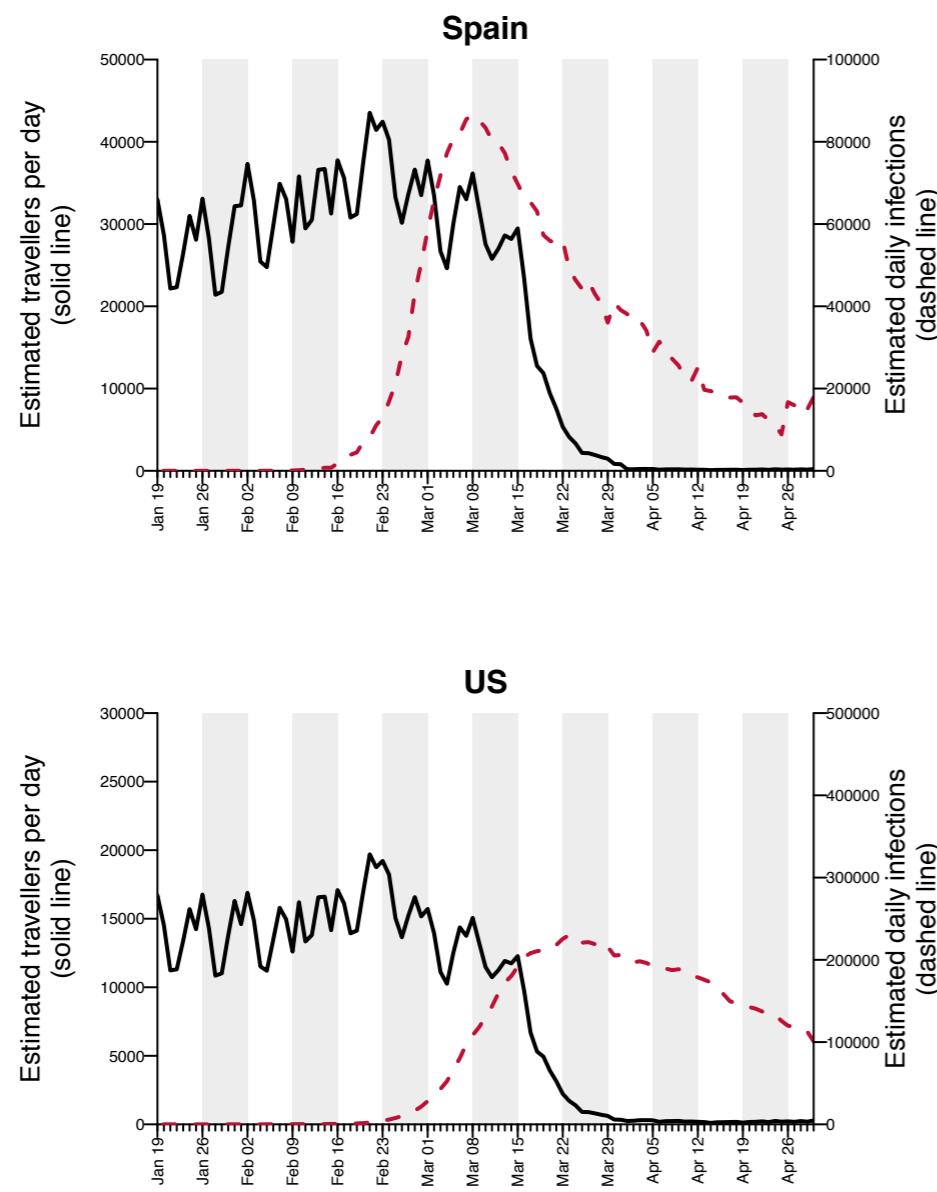
Country	Reported cases	Estimated infections	Genomes sequenced	Estimated sampling
United Kingdom	190,576	3,431,442	22,749	0.66%
Spain	218,010	2,592,585	1,612	0.062%
France	131,857	2,215,432	295	0.013%
Italy	211,938	2,764,563	136	0.0049%

Genetic diversity problem

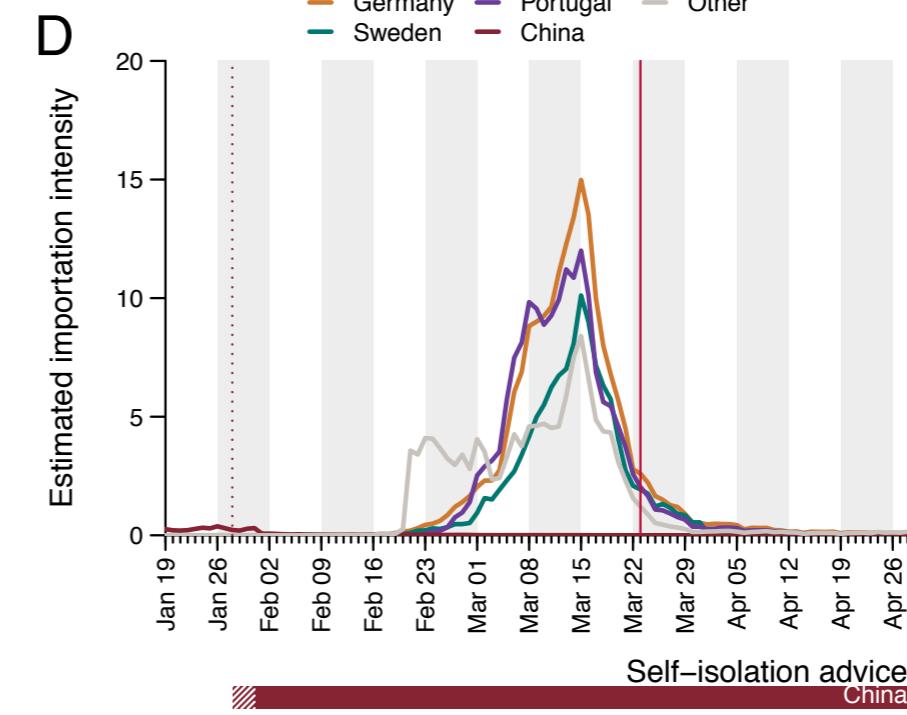
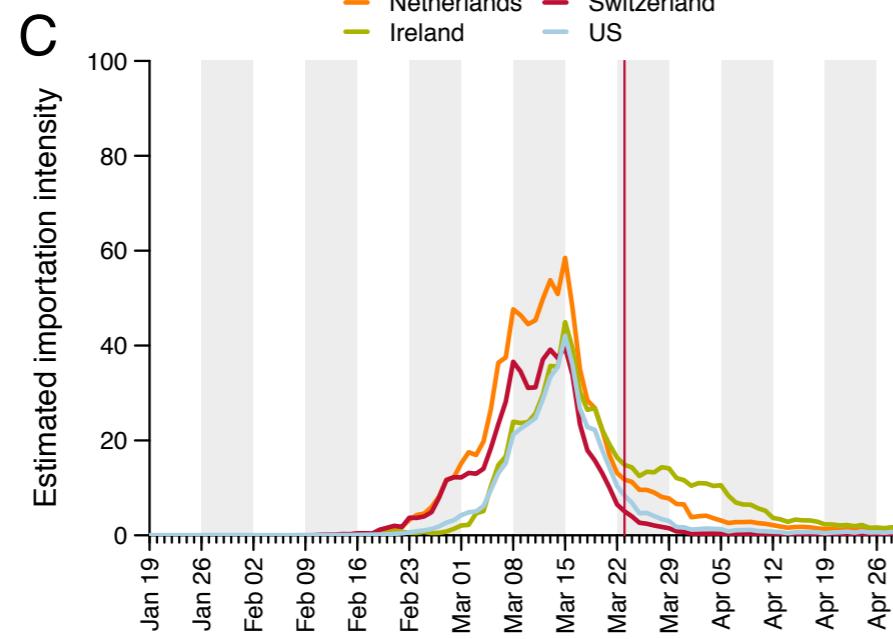
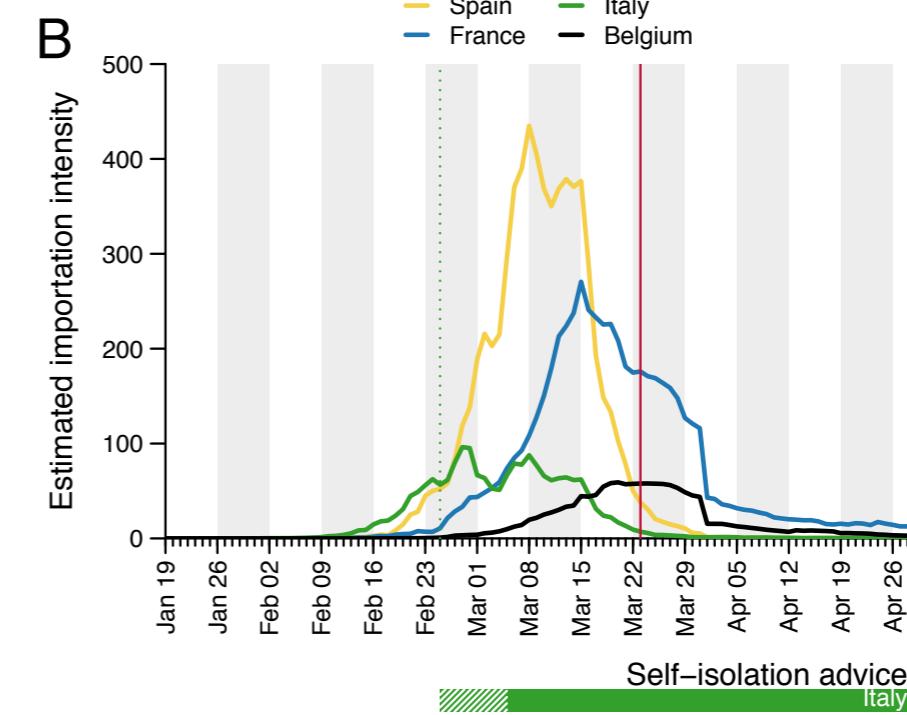
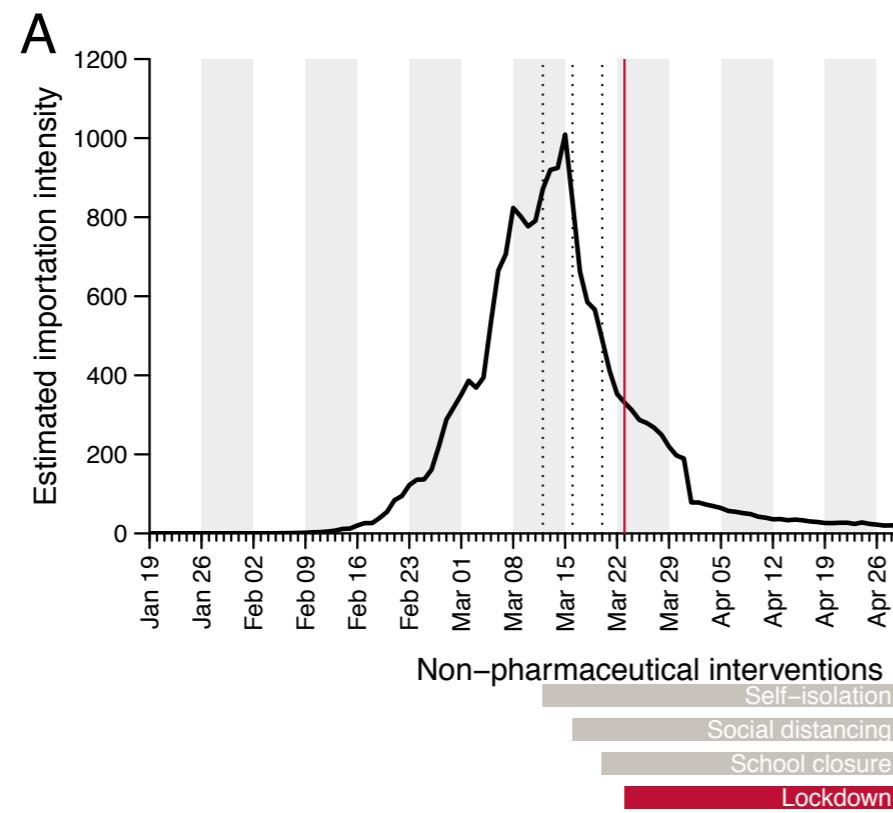


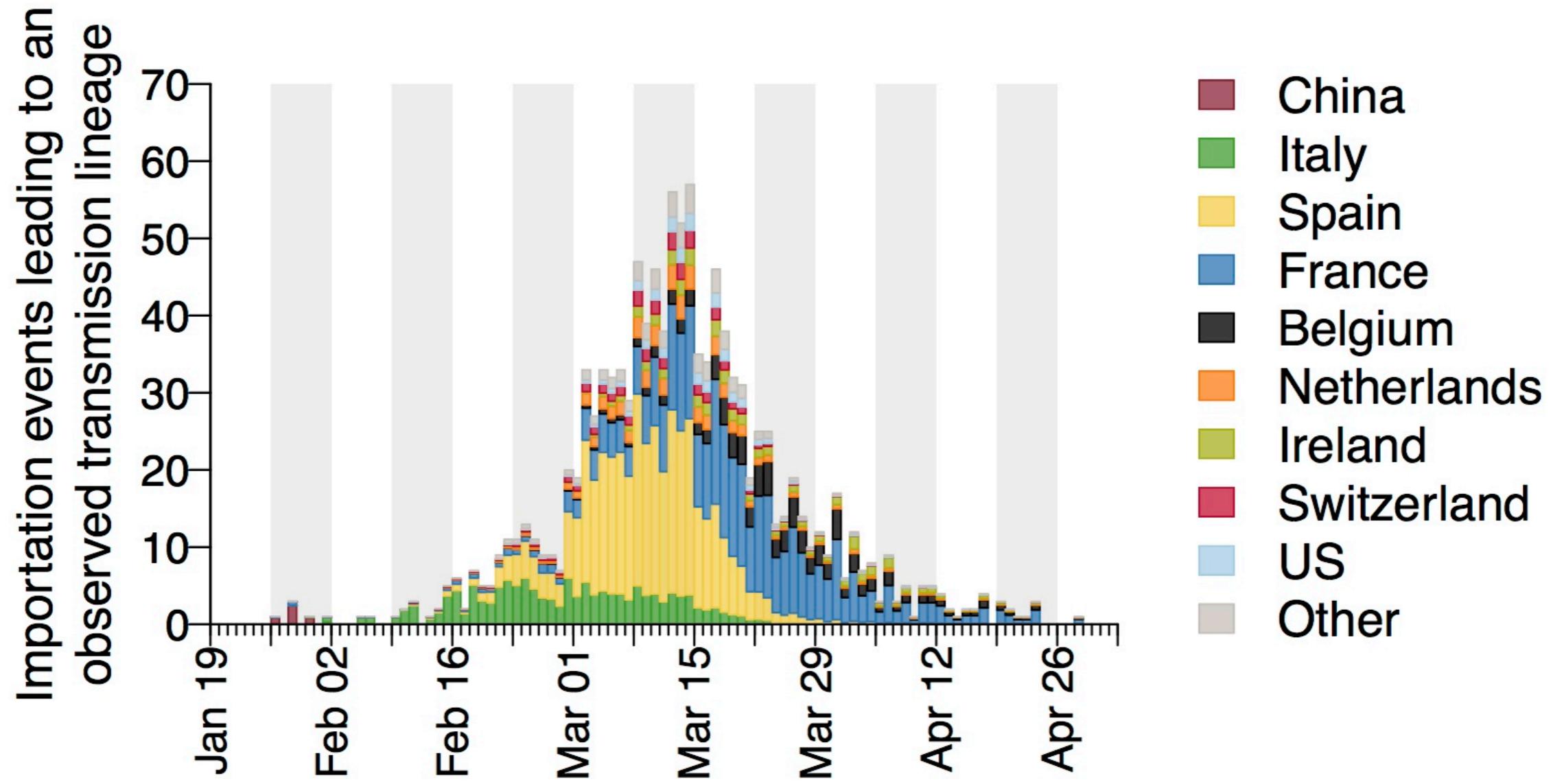
- Genetic distance tree
- Red nodes are in the UK
- Black nodes are outside of the UK
- UK transmission lineages have identical genomes from Iceland, Switzerland, Spain, Germany etc. at their roots

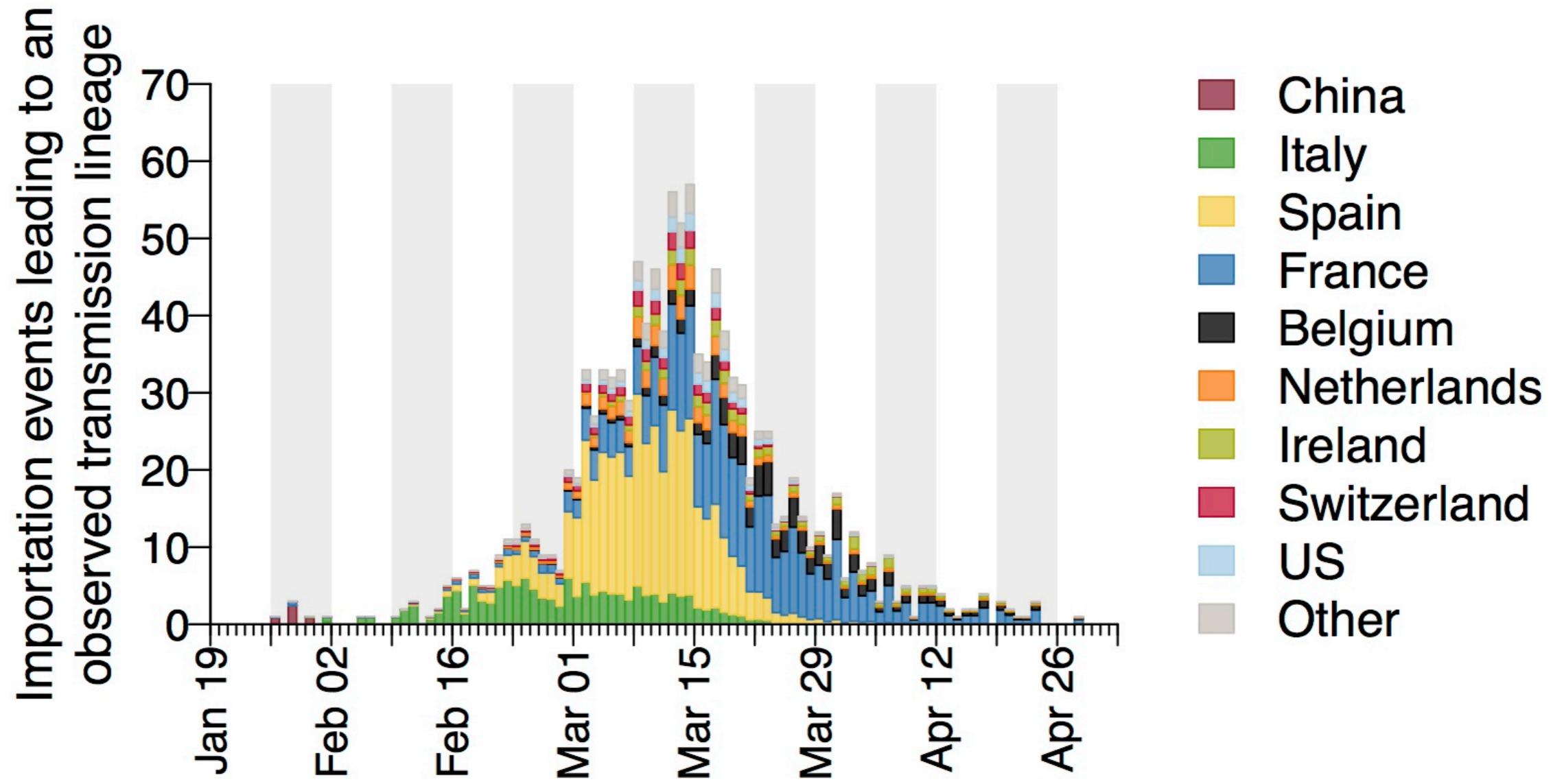
Country-specific travel and COVID-19 cases



Country-specific importation intensity

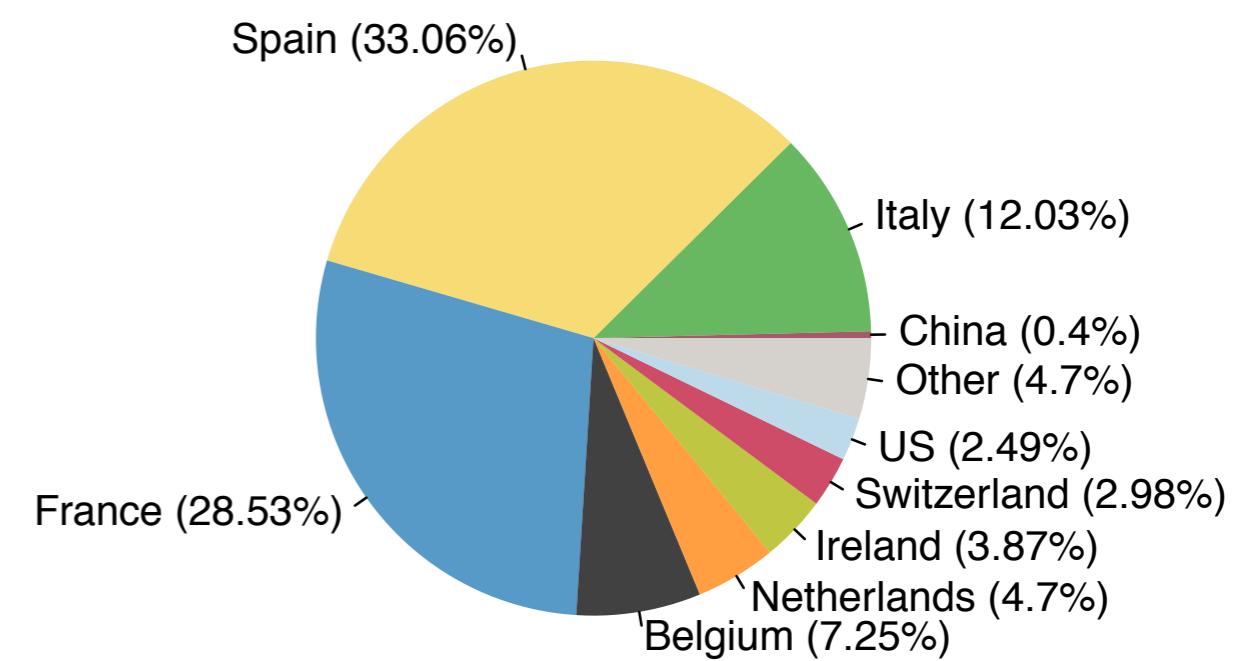
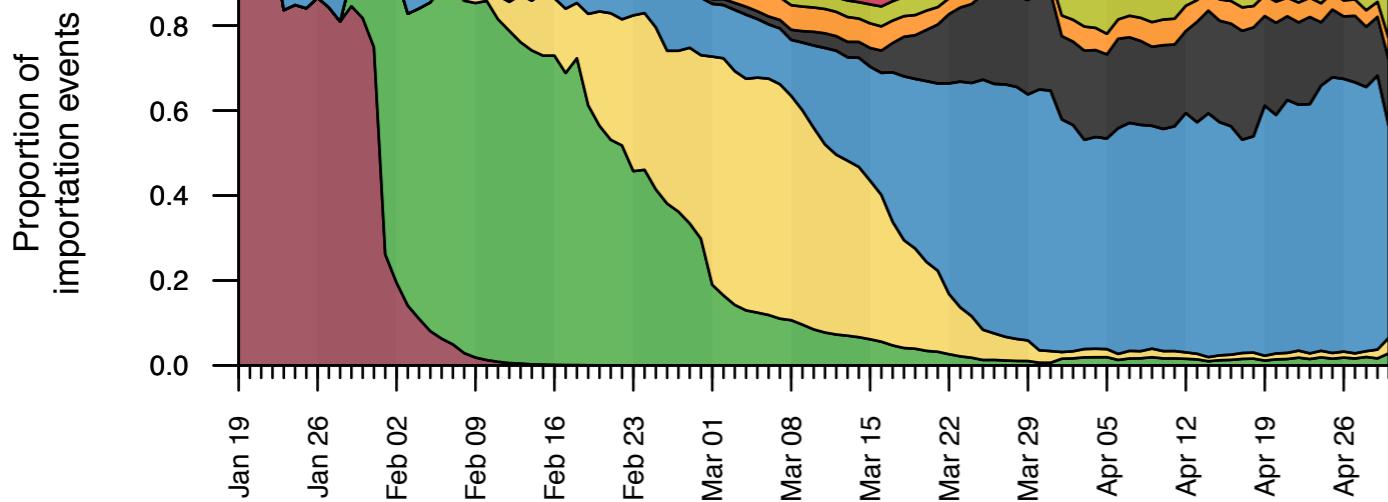






Legend:

- China
- Italy
- Spain
- France
- Belgium
- Netherlands
- Ireland
- Switzerland
- US
- Other



Acknowledgments

University of Oxford

Oliver Pybus
Alex Zarebski
Moritz Kraemer
Bernardo Gutierrez
Jayna Raghwani
Nuno Faria
Tetyana Vasylyeva
David Aanensen

University of Edinburgh

Andrew Rambaut
JT McCrone
Verity Hill
Jordan Ashworth
Rachel Colquhoun
Ben Jackson
Áine O'Toole
Emily Scher

**and everyone who sampled, sequenced
and shared SARS-CoV-2 genomes!**

University of Birmingham

Nicholas Loman
Samuel Nicholls

University of Cambridge

Chris Ruis

University of Toronto / BlueDot

Alexander Watts
Isaac Bogoch
Kamran Khan

Imperial College

Kris Parag
Erik Volz

Public Health Wales

Tom Connor



**COVID-19
GENOMICS
UK CONSORTIUM**

<https://www.cogconsortium.uk>



RESEARCH

CORONAVIRUS

Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK

Louis du Plessis^{1*}, John T. McCrone^{2*}, Alexander E. Zarebski^{1*}, Verity Hill^{2*}, Christopher Ruis^{3,4*}, Bernardo Gutierrez^{1,5}, Jayna Raghwani¹, Jordan Ashworth², Rachel Colquhoun², Thomas R. Connor^{6,7}, Nuno R. Faria^{1,8}, Ben Jackson², Nicholas J. Loman⁹, Áine O'Toole², Samuel M. Nicholls⁹, Kris V. Parag⁸, Emily Scher², Tetyana I. Vasylyeva¹, Erik M. Volz⁸, Alexander Watts^{10,11}, Isaac I. Bogoch^{12,13}, Kamran Khan^{10,11,12}, COVID-19 Genomics UK (COG-UK) Consortium†, David M. Aanensen^{14,15}, Moritz U. G. Kraemer^{1‡}, Andrew Rambaut^{2‡§}, Oliver G. Pybus^{1,16‡§}