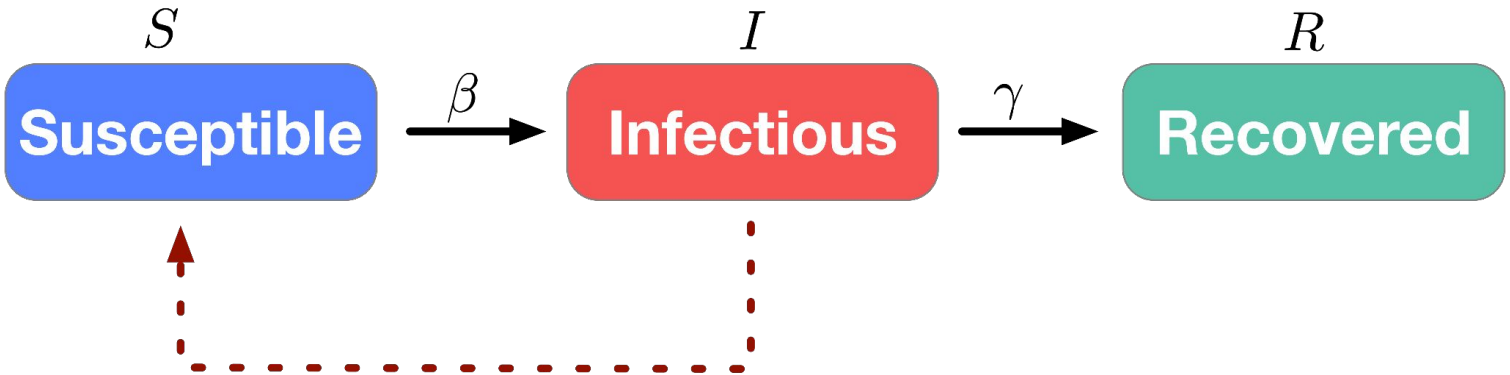
CSCI6802: Phylodynamics

Based on slides from Angela McLaughlin, Louis du Plessis material in the Taming the BEAST workshop, and Trevor Bedford's teaching materials

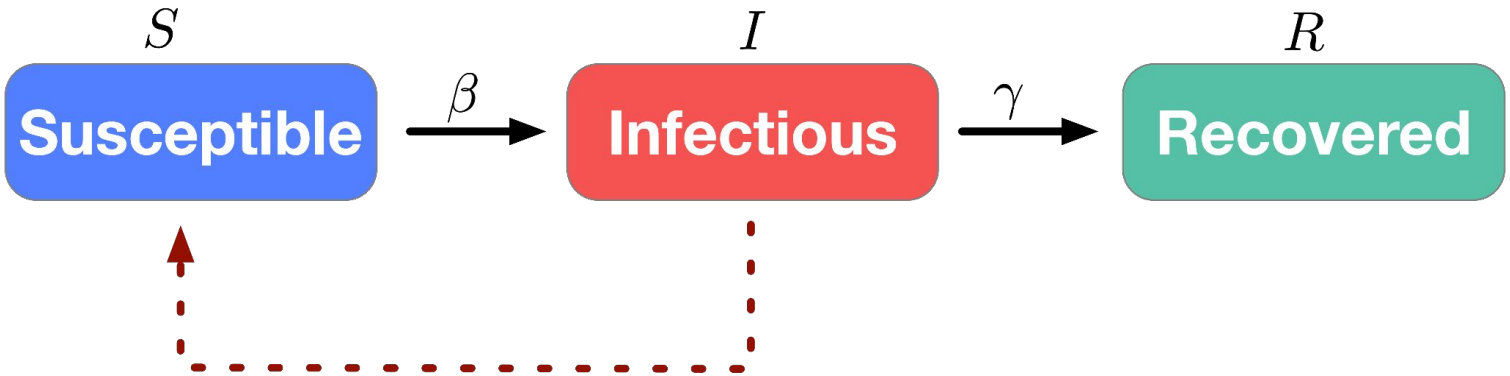
So, you want to understand the epidemiological dynamics of infectious diseases?

Compartmental models are used to model infections



https://covid19.uclaml.org/model.html

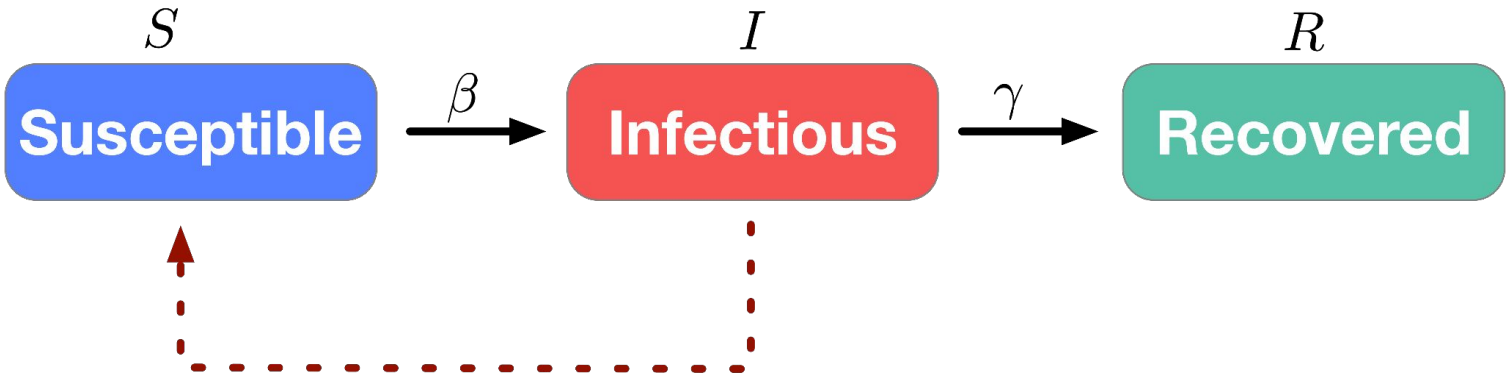
Compartmental models are used to model infections



Disclaimer: many more complex models!

https://covid19.uclaml.org/model.html

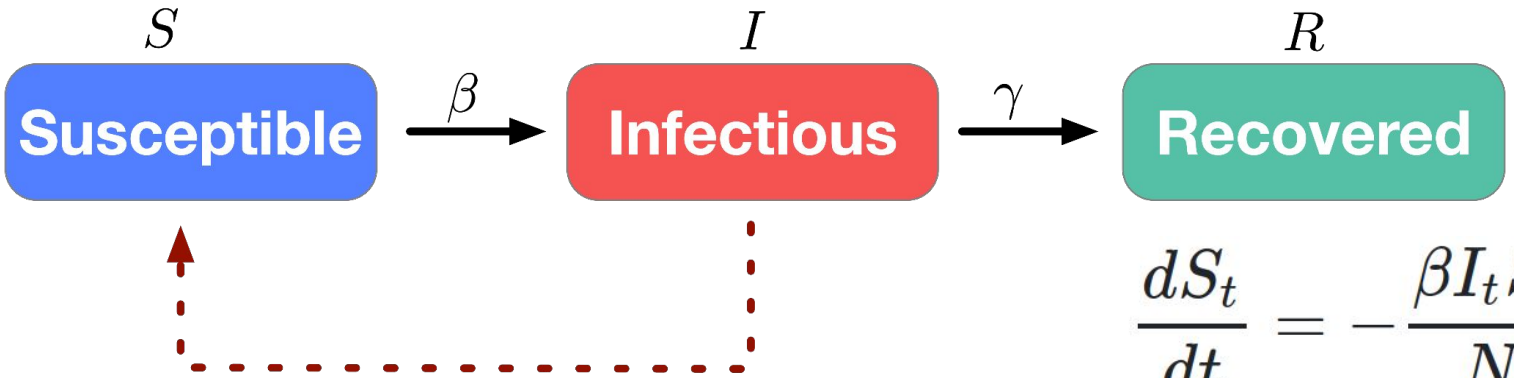
Compartmental models are used to model infections



- S_t: the number of susceptible individuals
- I_t: the number of infectious individuals
- R_t: the number of recovered/deceased/immune individuals

https://covid19.uclaml.org/model.html

Compartmental models are used to model infections



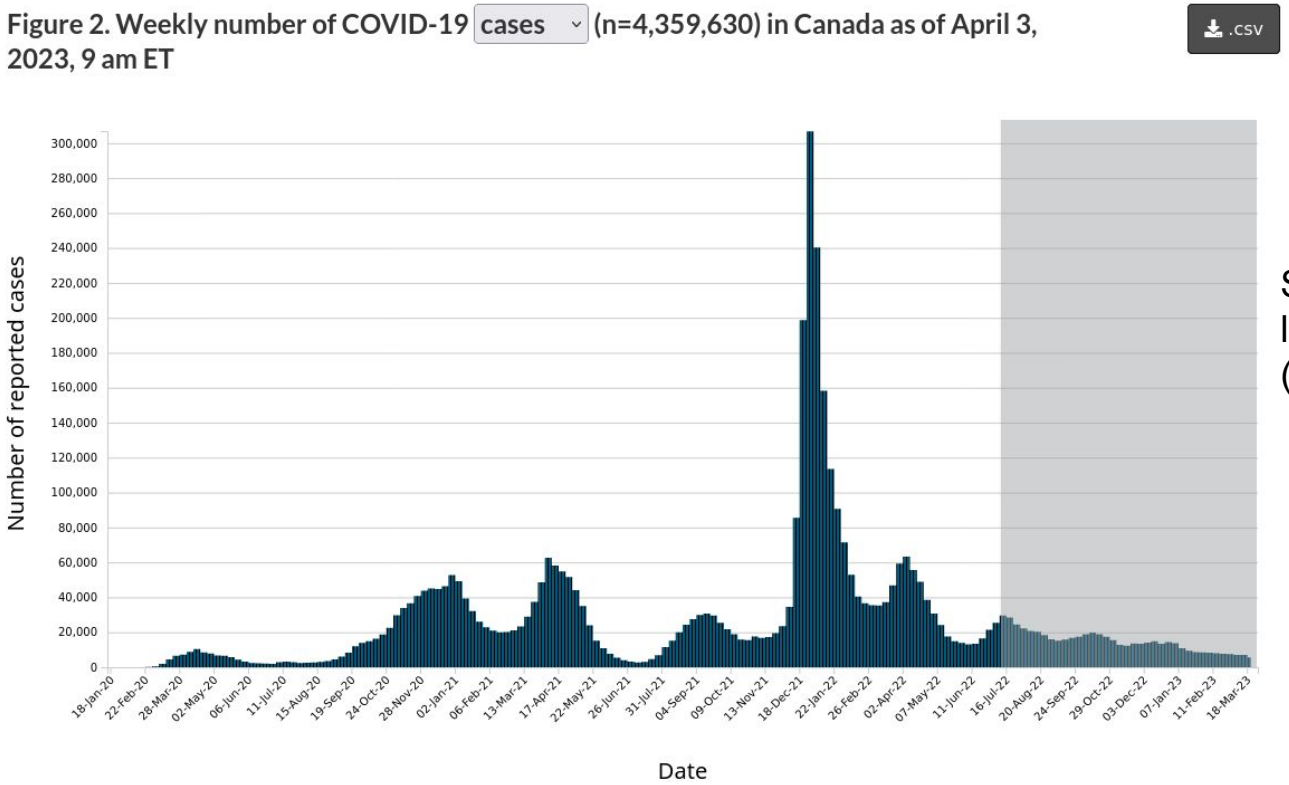
- S_t: the number of susceptible individuals
- I_t: the number of infectious individuals
- R_t: the number of recovered/deceased/immune individuals

$$\frac{dS_t}{dt} = -\frac{\beta I_t S_t}{N}$$
$$\frac{dI_t}{dt} = \frac{\beta I_t S_t}{N} - \gamma I_t$$
$$\frac{dR_t}{dt} = \gamma I_t$$

https://covid19.uclaml.org/model.html

Assuming N = fixed pop

Can calculate P(observed case counts | β=?,γ=?)



Same idea as Maximum likelihood Phylogenetics (just without any trees)

https://health-infobase.canada.ca/covid-19/current-situation.html

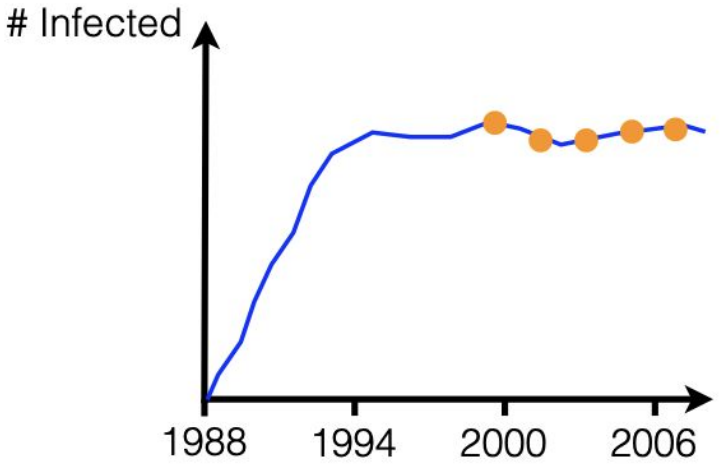
So, why do we need genomic data?

https://github.com/trvrb/phylodynamics-lecture

Genomics can be used to infer unobserved events

If sampling in early epidemic was missed:

- **Time of epidemic outbreak?**
- **Basic reproductive number R_0?**

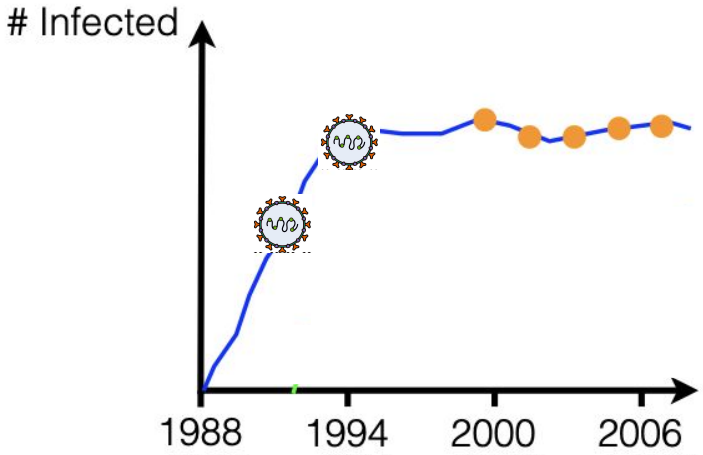


https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day2_Phylodynamics_-_Tanja_Stadler.pdf

Genomics can be used to infer unobserved events

If sampling in early epidemic was missed:

- **Time of epidemic outbreak?**
- **Basic reproductive number R_0?**



https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day2_Phylodynamics_-_Tanja_Stadler.pdf

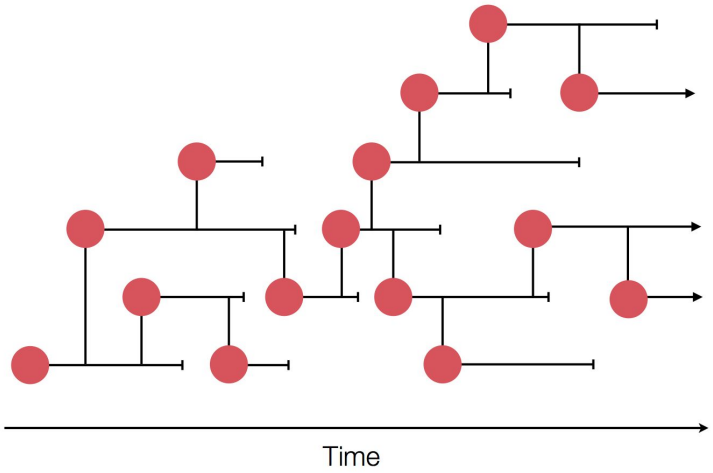
Genomics can be used to infer unobserved events

If sampling in early epidemic was missed:

- **Time of epidemic outbreak?**
- **Basic reproductive number R_0?**

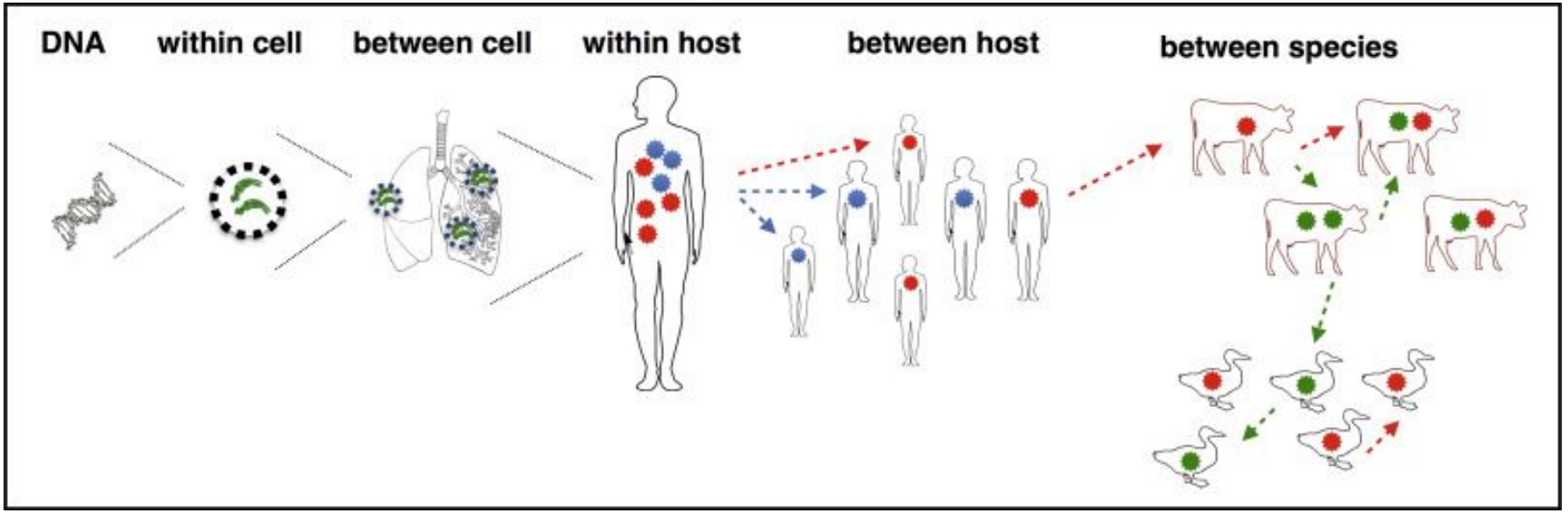
Data does not tell who infected whom:

- **Population structure?**



https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day2_Phylodynamics_-_Tanja_Stadler.pdf

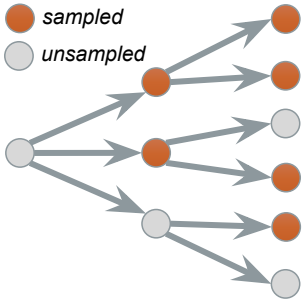
Cases don’t tell you (much) about pathogen evolution

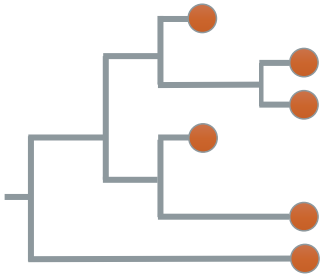


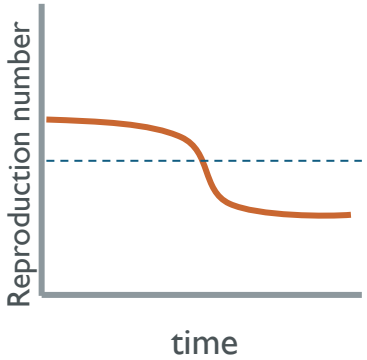
https://www.sciencedirect.com/science/article/pii/S1755436514000723

Many important genomic epidemiological questions

- Estimate lineages'/variants' divergence times and spatial origins
- Designate nomenclature track frequency
- Inform outbreak detection and support/refute epidemiological linkage
- Compare selective pressure across genomic sites and lineages
- Quantify groups' relative transmission rates and dispersal rates
- Evaluate impact of extrinsic forces, such as non-pharmaceutical interventions

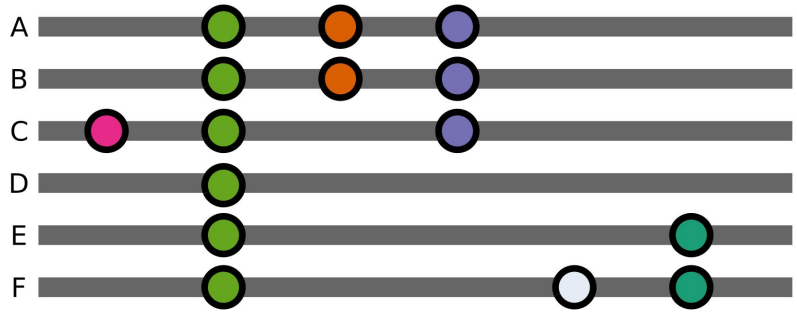




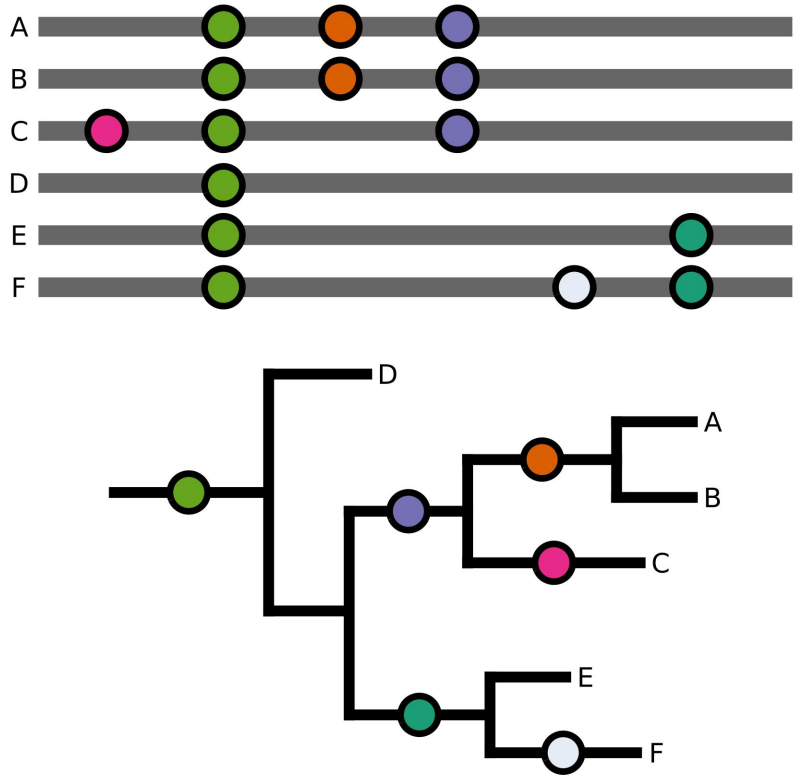


How do we actually link genomes and epidemiology?

Can infer a phylogeny from genomic data

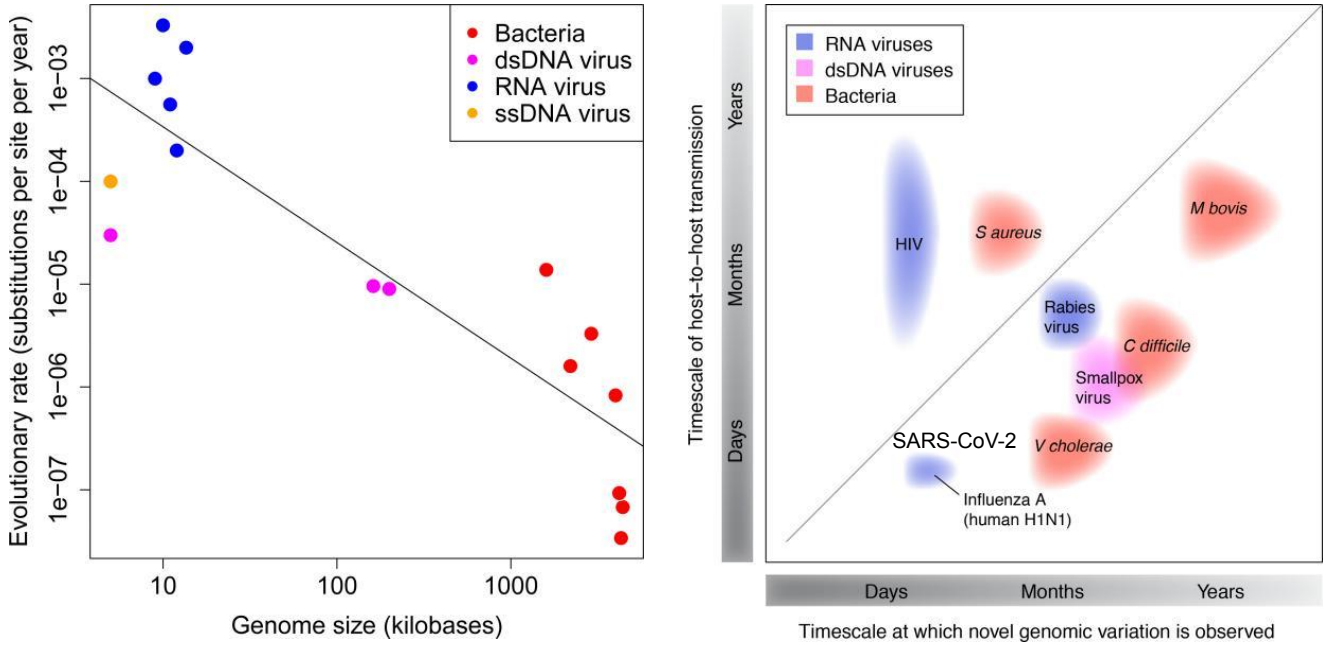


Can infer a phylogeny from genomic data



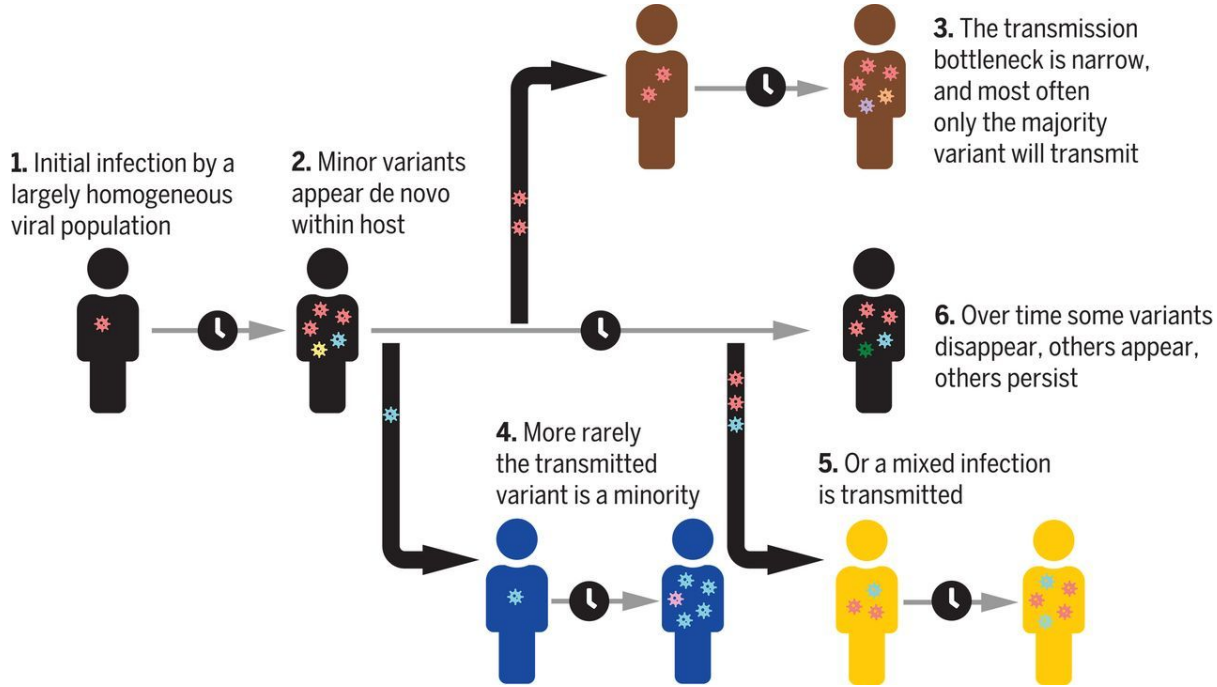
RNA viruses have measurably evolving populations

- RNA viruses have small genomes, large populations, and replicate frequently often with error-prone polymerases => rapid evolution!



Biek et al. 2016 *Trends Ecol Evol*

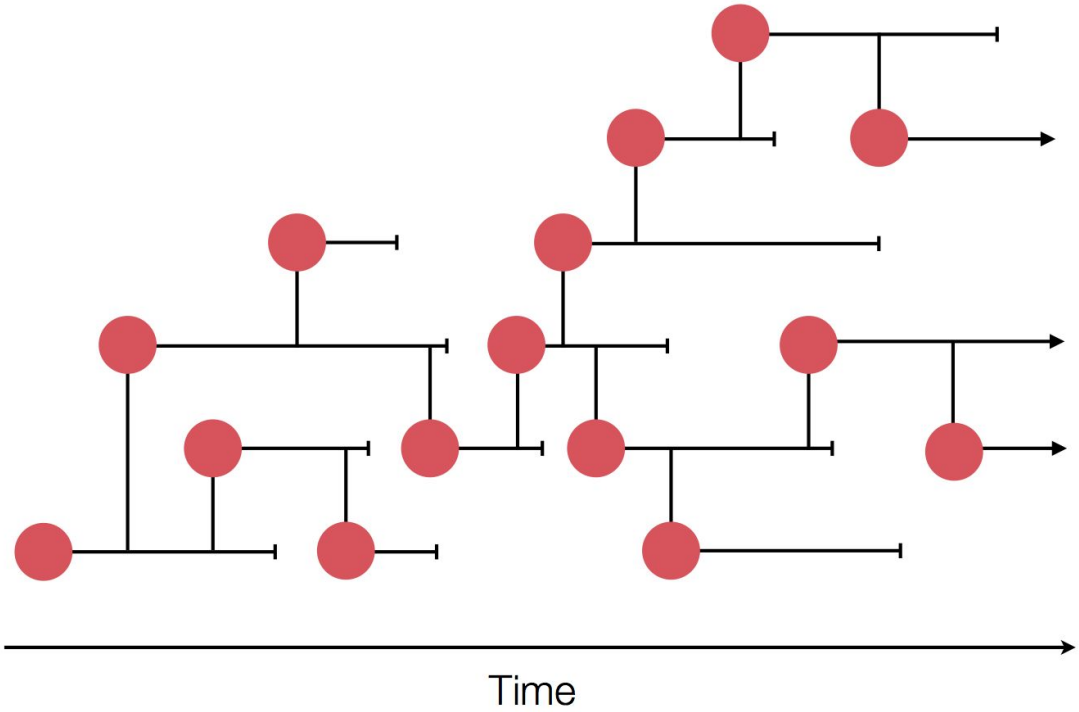
Complicated sampling of a (within-host) population of a (between host) population



https://www.science.org/doi/10.1126/science.abg0821

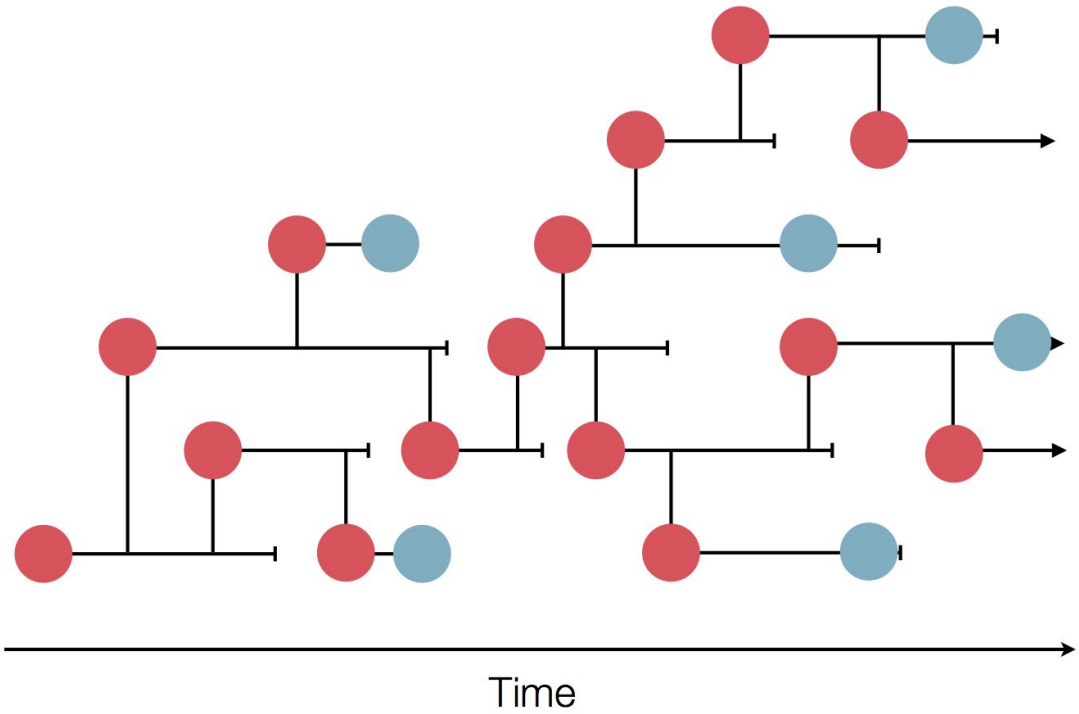
What does this tree actually represent?

Sampling from underlying process



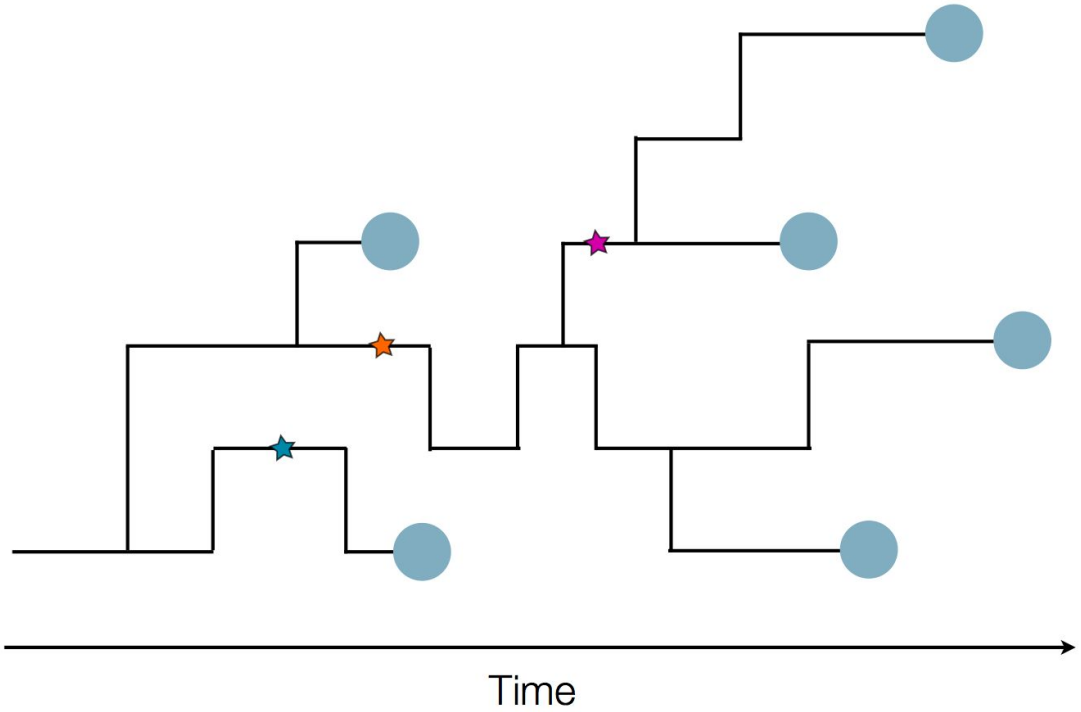
https://github.com/trvrb/phylodynamics-lecture

Sampling from underlying process



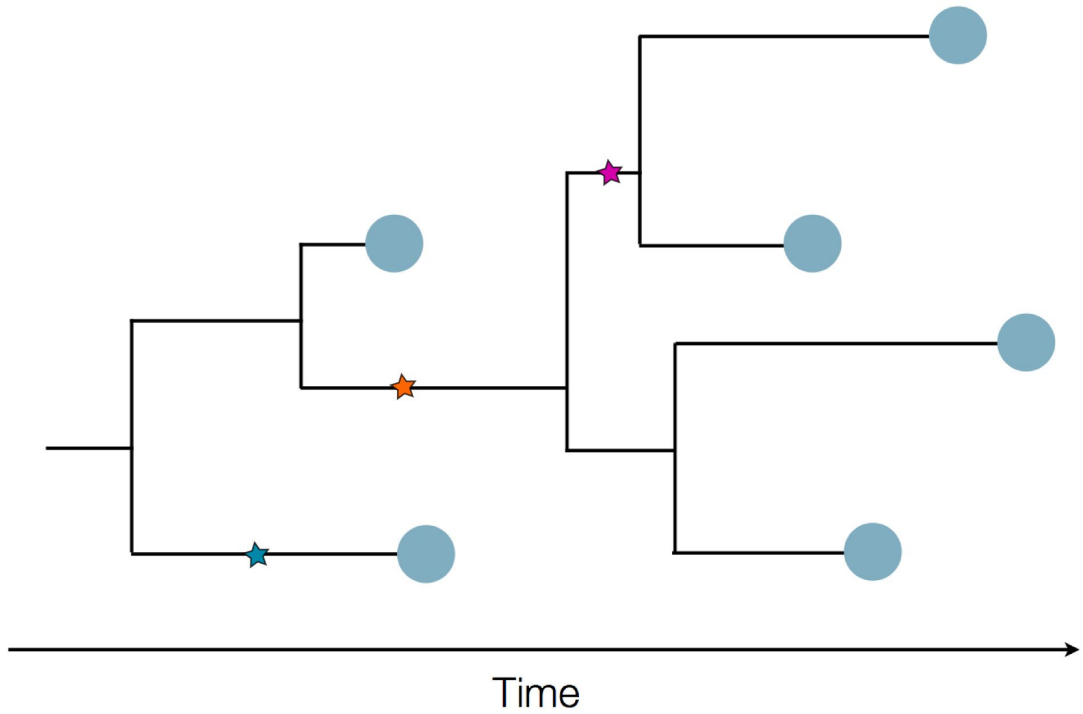
https://github.com/trvrb/phylodynamics-lecture

Sampling from underlying process



https://github.com/trvrb/phylodynamics-lecture

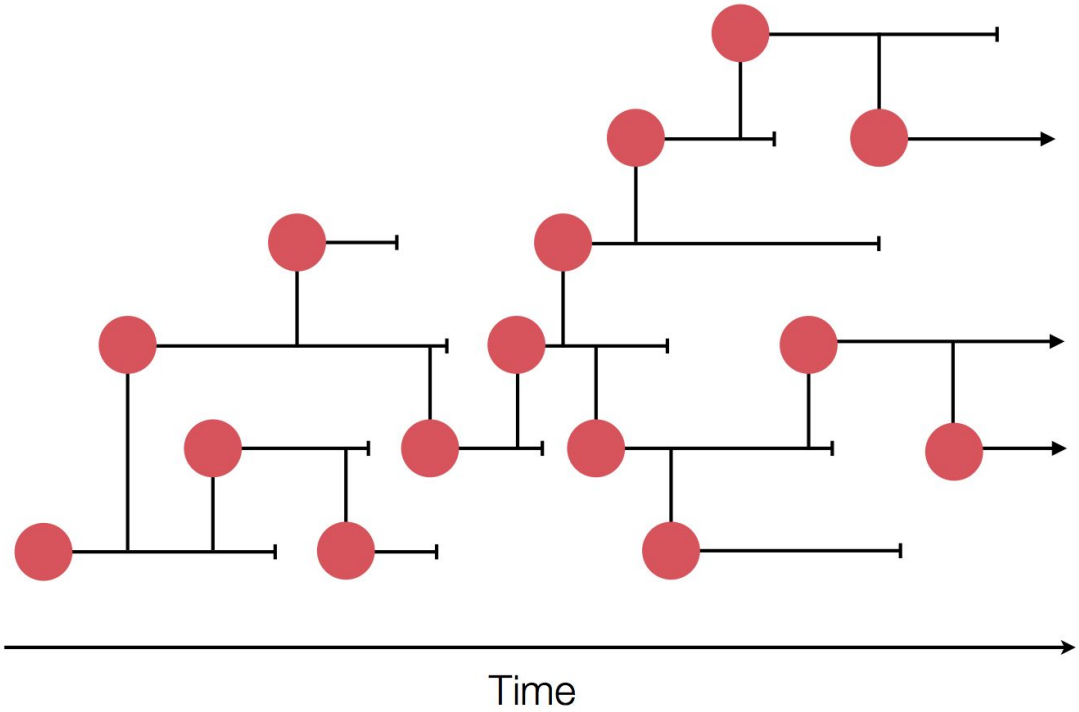
Sampling from underlying process



https://github.com/trvrb/phylodynamics-lecture

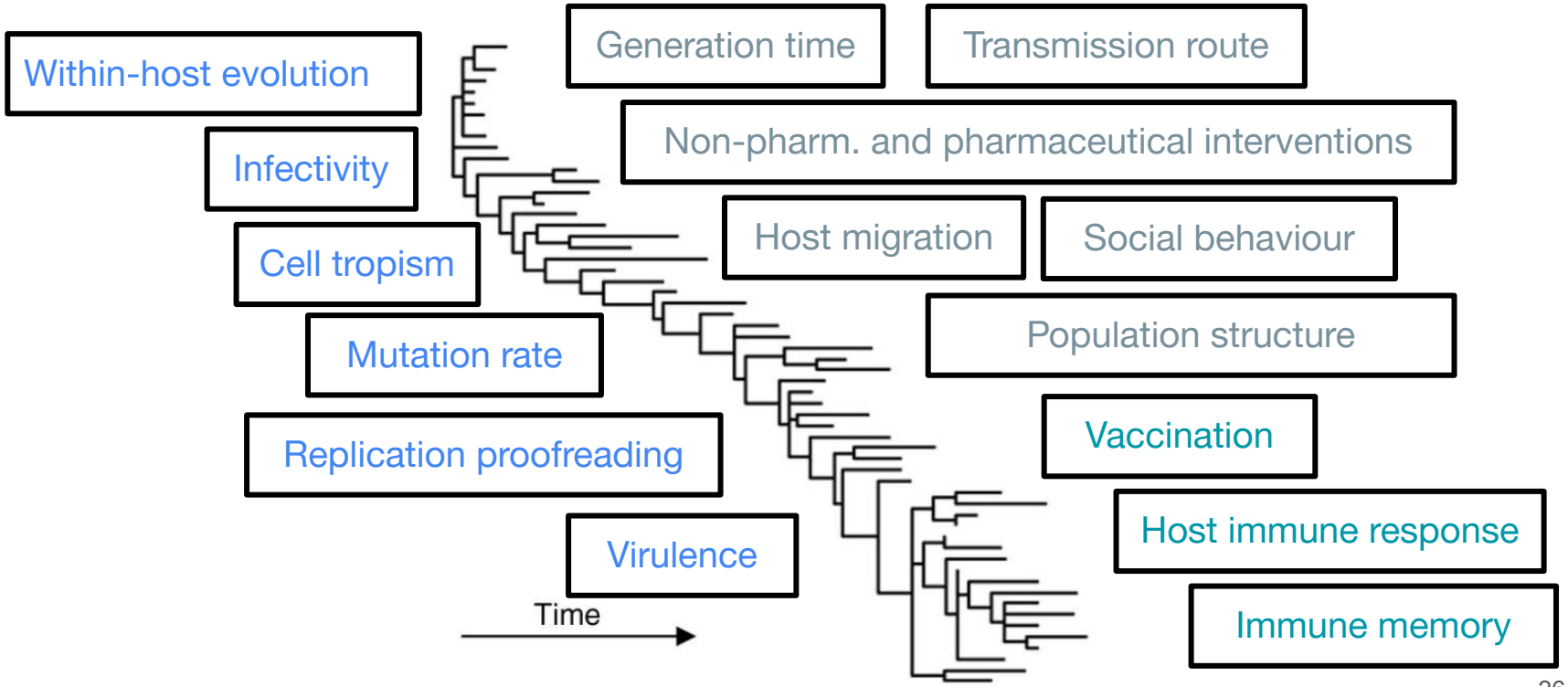
What determines underlying process?

Many forces shaping underlying process

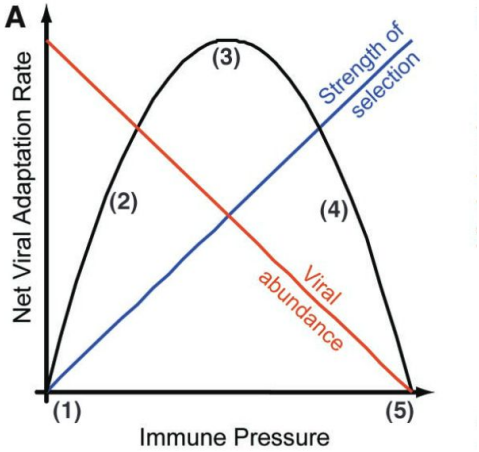


https://github.com/trvrb/phylodynamics-lecture

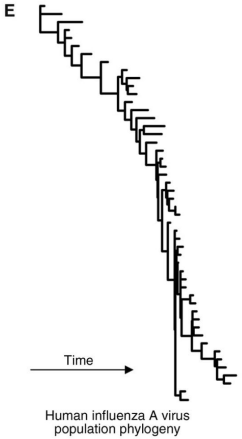
Evolutionary, epidemiological, and immunological forces shape phylogenies

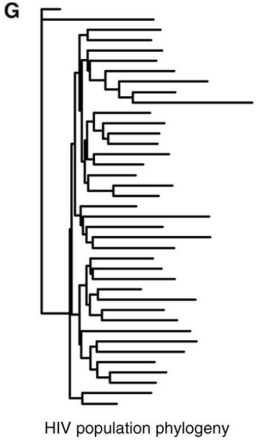


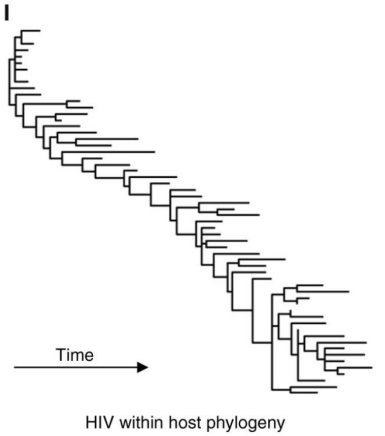
Different selection regimes shape phylogenies



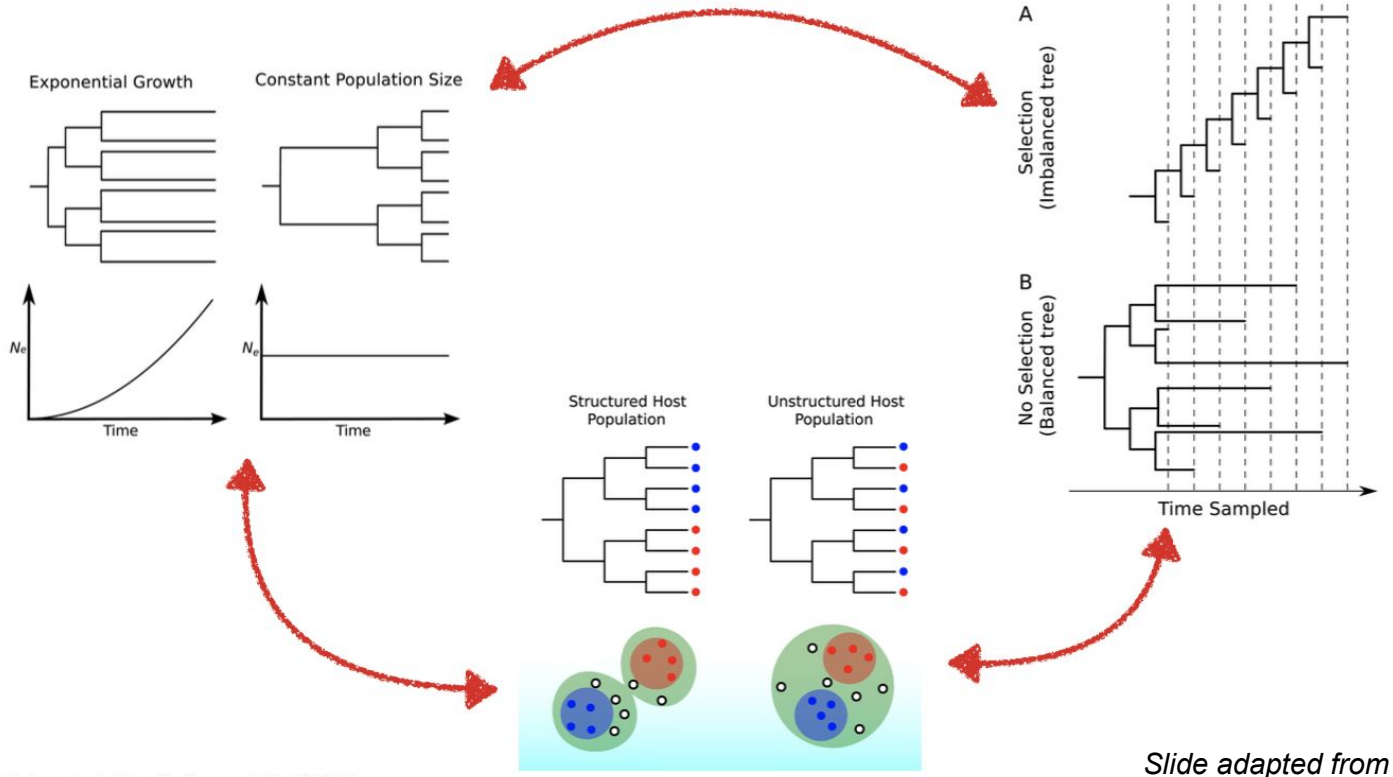
(Grenfell et al., 2004 *Science*)







Different population dynamics generate different tree shapes



Volz *et al.* **PLoS Comp Biol 2013**
Grenfell *et al.* **Science 2004**

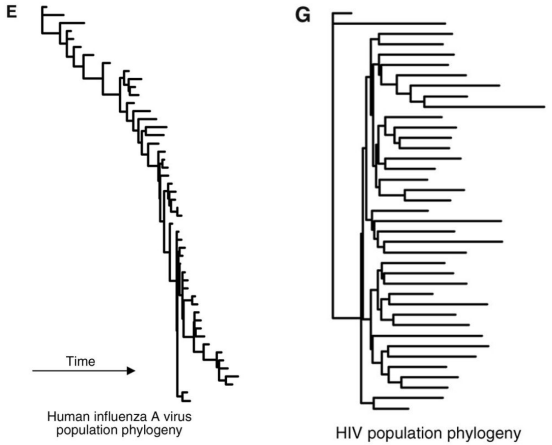
Slide adapted from Taming the BEAST 2019, Louis du Plessis

Spatial diffusion leaves also an imprint on the phylogeny

- **Phylogeography** = study of the principles and processes governing geographic distributions of genealogical lineages (Avise, 2000)

Are genealogies driven by dispersal or covariance? (Holmes, 2004)

- SARS-CoV-2: broad, rapid sweeps and global disassortativity. Acute infections, short generation time. Dispersal-dominant pattern
- HIV-1: co-existence of ~ geographically-structured subtypes and ongoing transmission within clusters. Persistent, chronic infections. More of a covariance distribution patterns



Phylodynamics is learning about this process from phylogeny (and vice versa!)

“Phylodynamics studies how ecological and evolutionary processes act and interact to shape a population’s phylogenetic history” (Grenfell et al., 2004 *Science*)

So, how do we do this?

Reminder: Bayes' Theorem

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

Prior → P(model)
- Original probability for the model parameters/components
- Belief in our hypothesis
- All parameters have priors, whether you specify them or not!

Likelihood → P(data | model)
- Probability of data given parameters (defined by model)

Posterior → P(model | data)
- Updated probability for the model parameters in light of the data

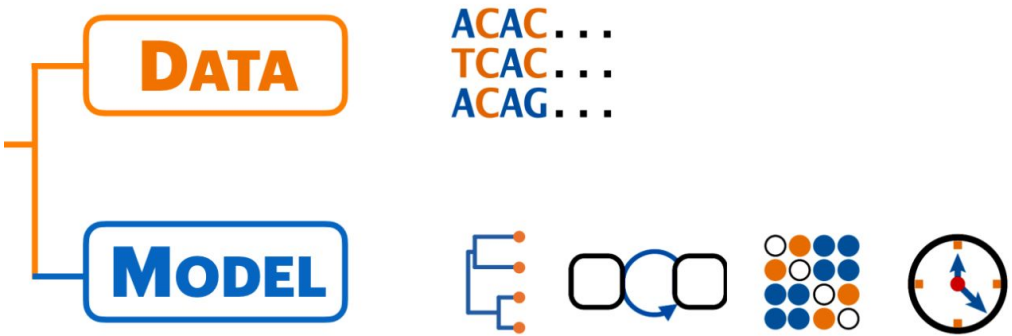
Model evidence → P(data) *Aka Marginal likelihood (hard to calculate)*
- Probability for data given model (any combination of parameters)
- Used for Bayesian model selection

Slide adapted from Taming the BEAST 2019 slides, Louis de Plessis

Bayesian inference is a key tool in phylodynamics

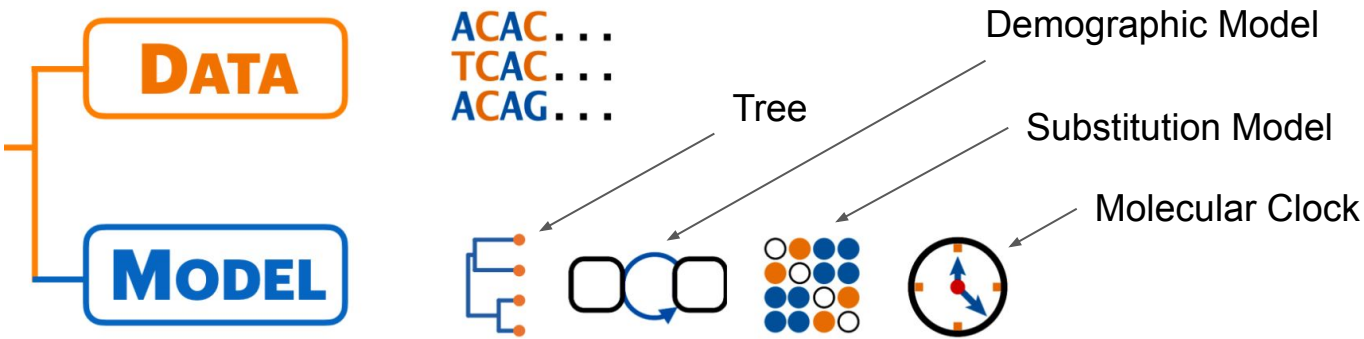
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/blob/master/Day1_First_Steps_-_Ceci_Valenzuela.pdf

Bayesian inference is a key tool in phylodynamics



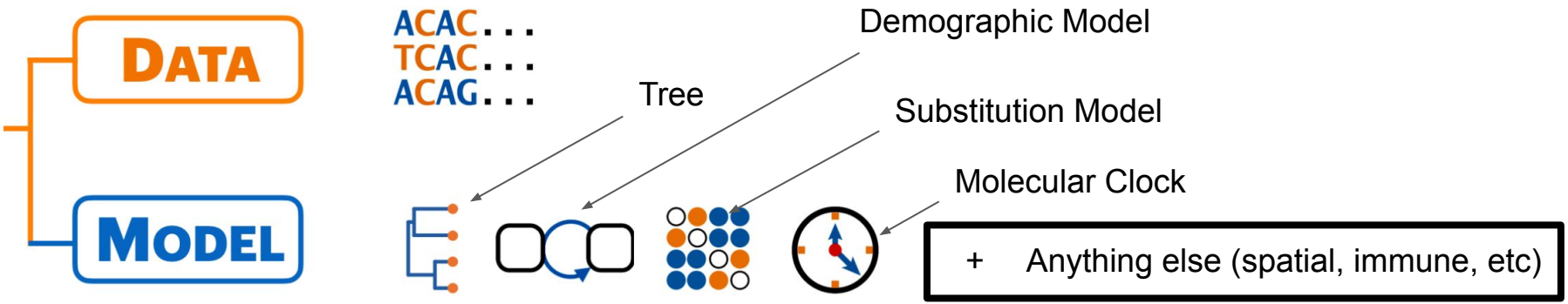
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/blob/master/Day1_First_Steps_-_Ceci_Valenzuela.pdf

Bayesian inference is a key tool in phylodynamics



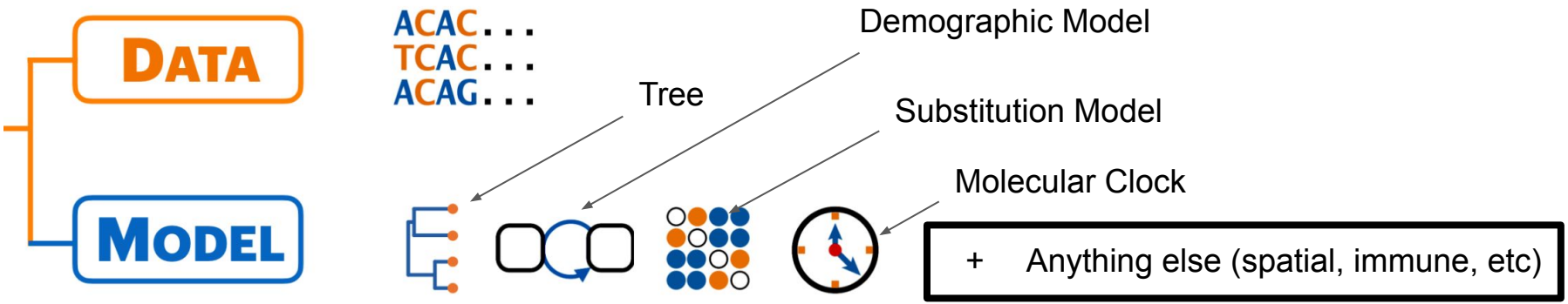
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/blob/master/Day1_First_Steps_-_Ceci_Valenzuela.pdf

Bayesian inference is a key tool in phylodynamics



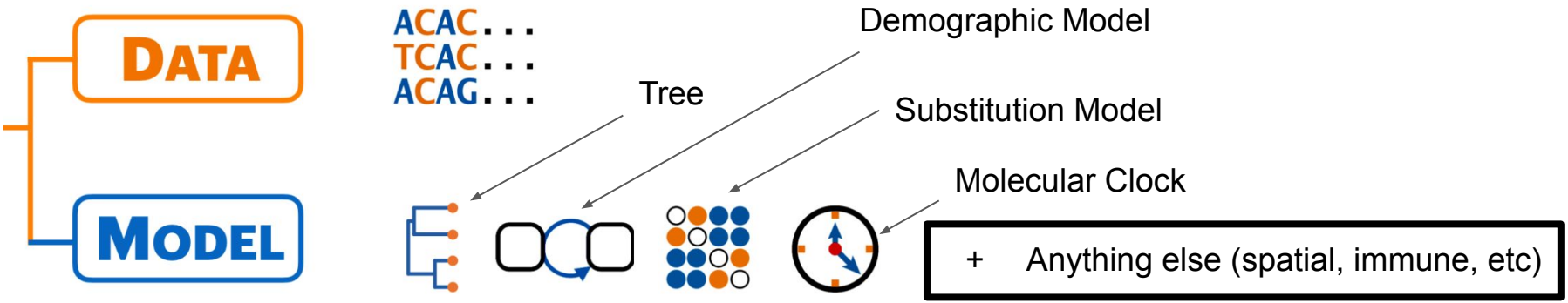
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/blob/master/Day1_First_Steps_-_Ceci_Valenzuela.pdf

Bayesian inference is a key tool in phylodynamics



https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/blob/master/Day1_First_Steps_-_Ceci_Valenzuela.pdf

Bayesian inference is a key tool in phylodynamics



$$P(\text{tree}, \text{model} \mid \text{ACAC}\ldots,\ \text{TCAC}\ldots,\ \text{ACAG}\ldots) = \frac{P(\text{ACAC}\ldots,\ \text{TCAC}\ldots,\ \text{ACAG}\ldots \mid \text{tree}, \text{model})\, P(\text{tree}, \text{model})}{P(\text{ACAC}\ldots,\ \text{TCAC}\ldots,\ \text{ACAG}\ldots)}$$

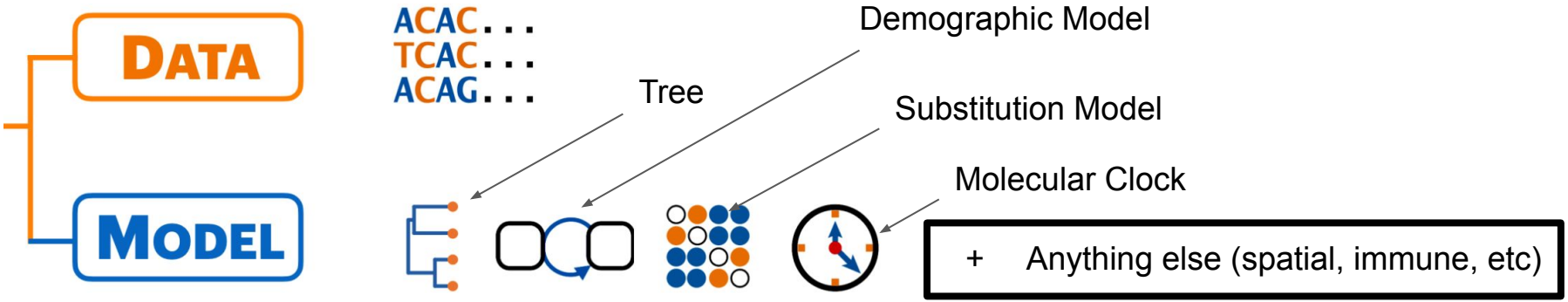
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Reminder: priors

$$P(\text{model})$$

- A distribution for some model parameter chosen based on your beliefs (from independent evidence/experiments) and with a particular analysis in mind
- Often a parametric distribution
  - e.g., uniform, normal, gamma, beta, lognormal, Laplace, …
- Sometimes a prior on a model component
  - e.g., substitution model (HKY, GTR, JC, …)
- Priors can have priors which can have priors (i.e., hyperpriors)
- Parameter bounds are part of the prior
  - e.g. normal distribution with lower bound

Bayesian inference is a key tool in phylodynamics

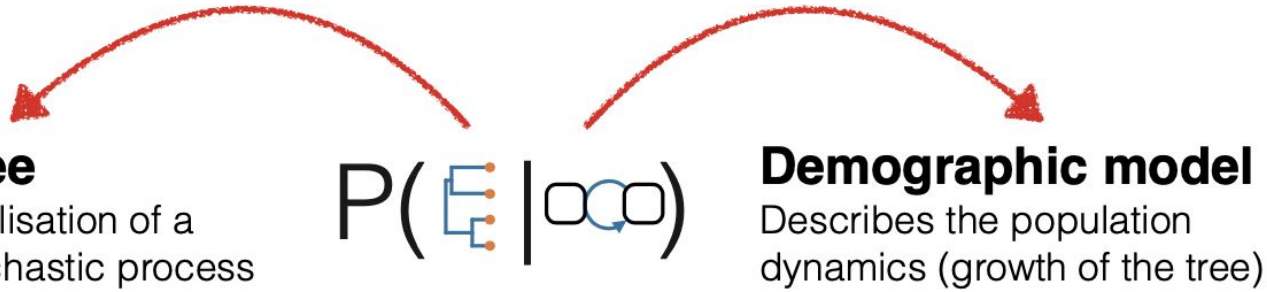


$$P(\text{tree}, \text{model} \mid \text{ACAC}\ldots,\ \text{TCAC}\ldots,\ \text{ACAG}\ldots) = \frac{P(\text{ACAC}\ldots,\ \text{TCAC}\ldots,\ \text{ACAG}\ldots \mid \text{tree}, \text{model})\, P(\text{tree}, \text{model})}{P(\text{ACAC}\ldots,\ \text{TCAC}\ldots,\ \text{ACAG}\ldots)}$$

https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Demographic model

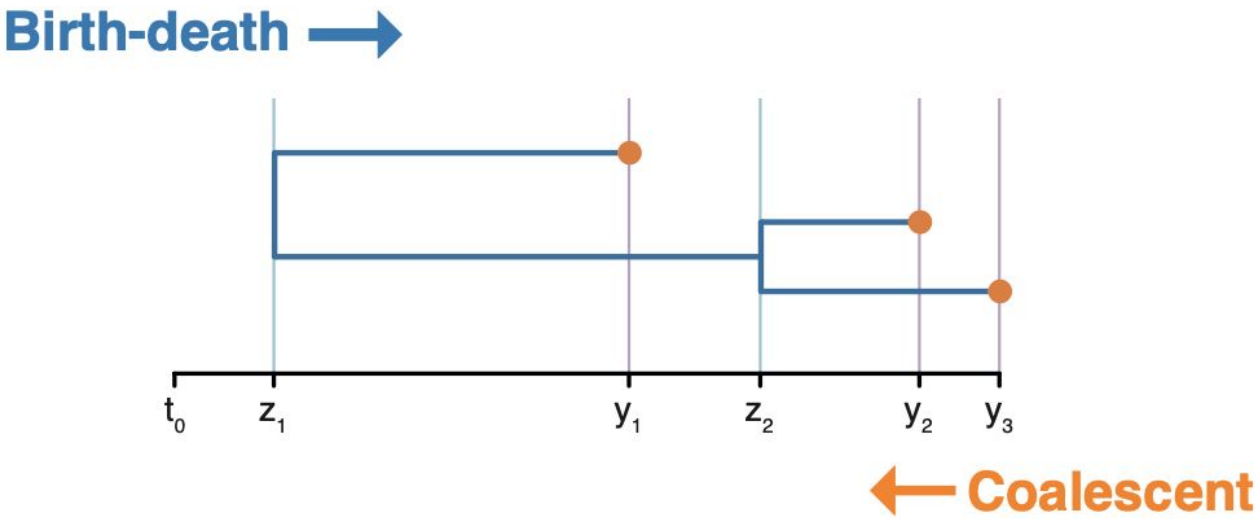
- Describes the population/speciation dynamics
- How does the population demographics / species diversity change over time?



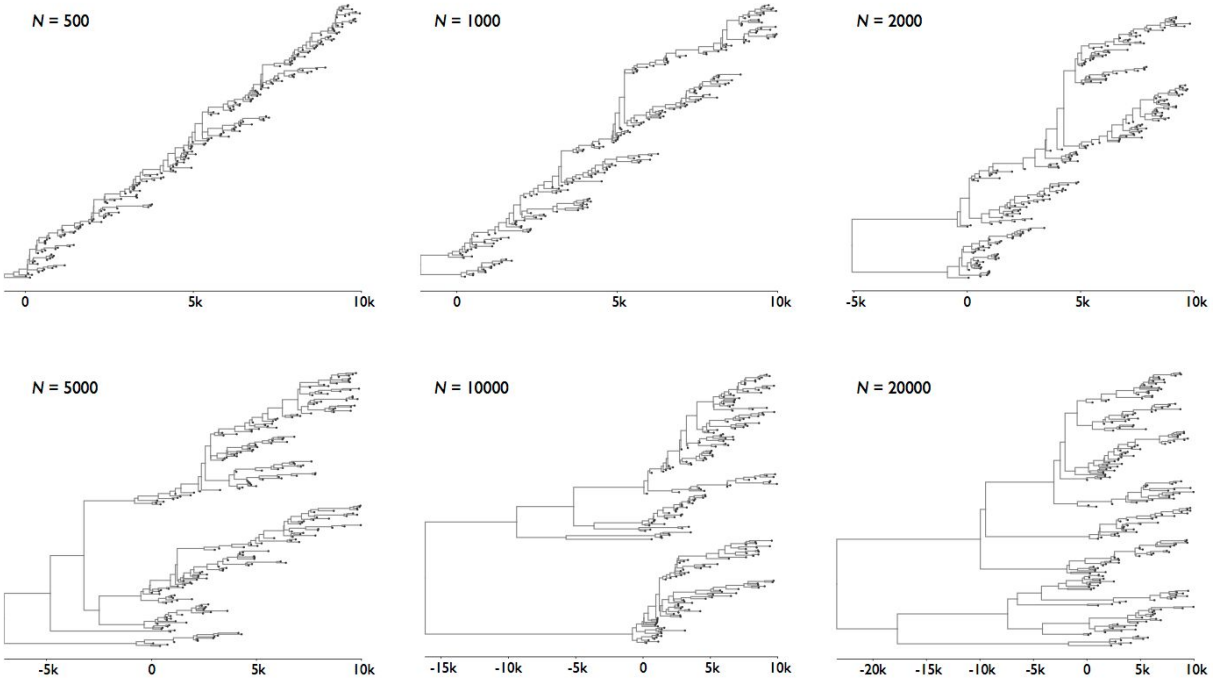
- How likely is the genealogy given a demographic model?
- Sometimes called a **tree prior**

Slide adapted from Taming the BEAST 2019 slides, Louis de Plessis

Demographic models are based on a **coalescent process** or a **birth-death (BD) process**

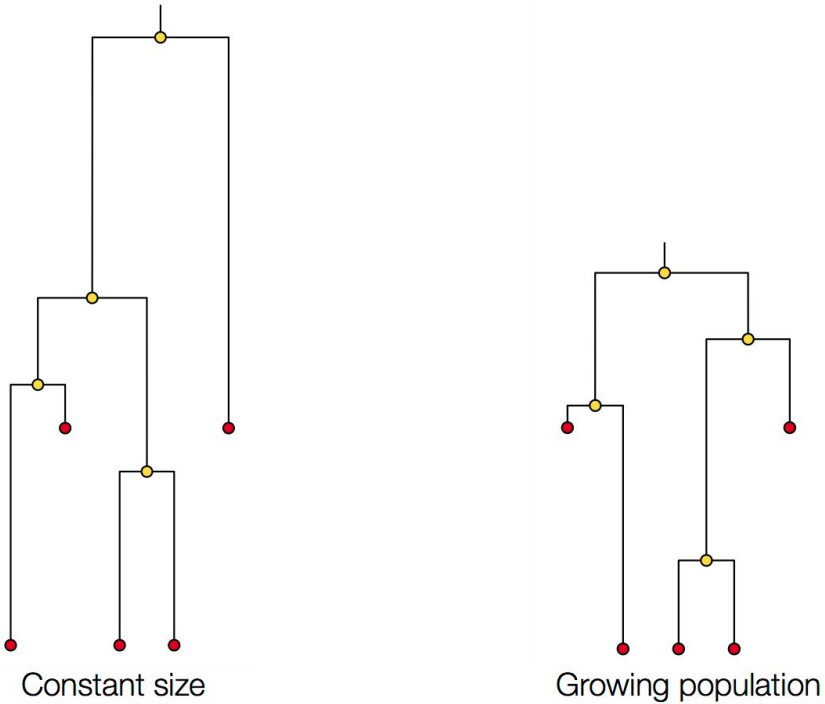


Shape of tree relates to population size (and structure)



https://github.com/trvrb/gs541-phylodynamics

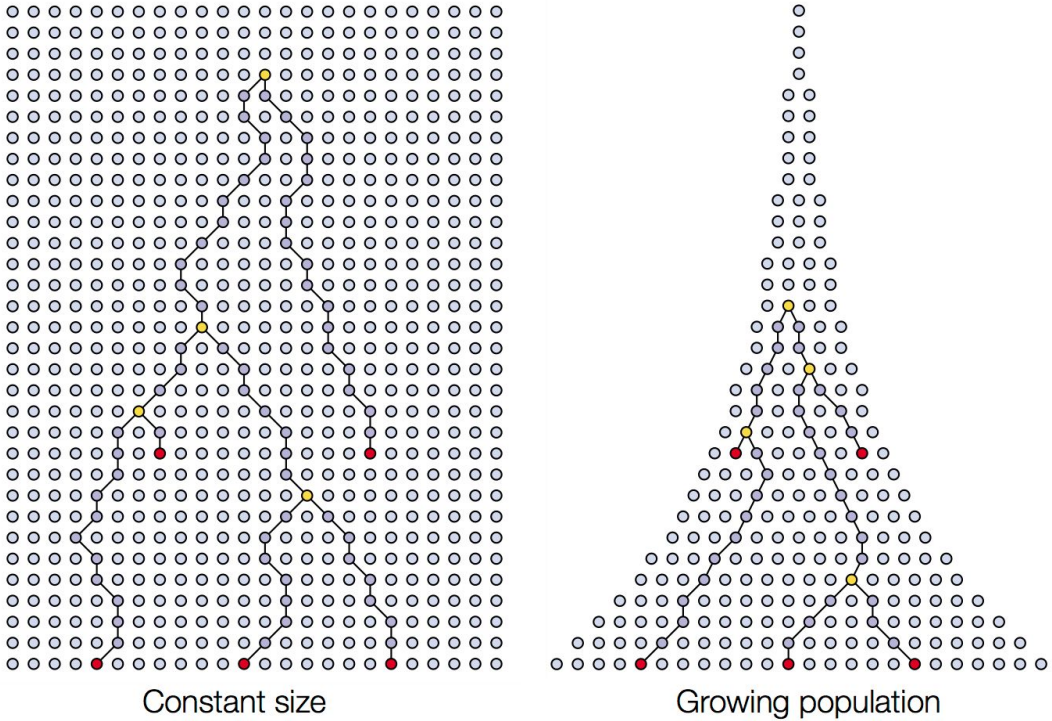
Coalescent processes let us quantify this relationship



https://github.com/trvrb/gs541-phylodynamics

Coalescent processes let us quantify this relationship

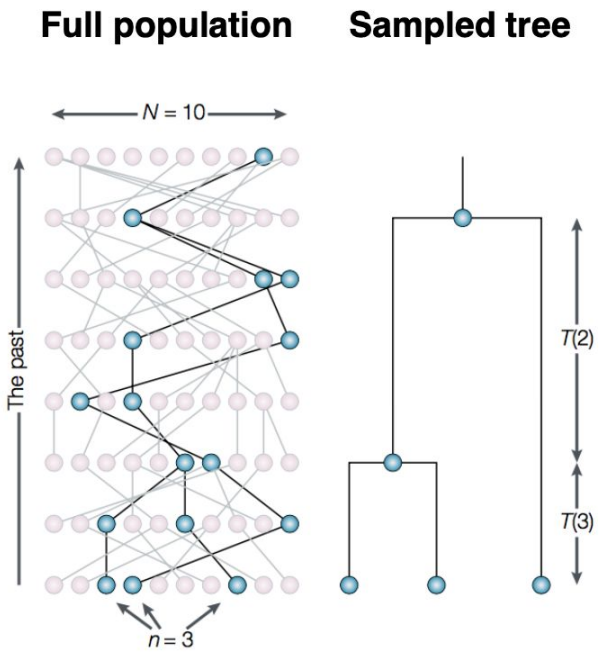
$P(coal|i=2) = 1/N$



https://github.com/trvrb/gs541-phylodynamics

The coalescent branching model

Describes the rate at which lineages coalesce into common ancestors backward-in-time, informing the rate of population growth (Kingman, 1982)



Rosenberg and Nordborg **Nature Reviews Genetics 2002**

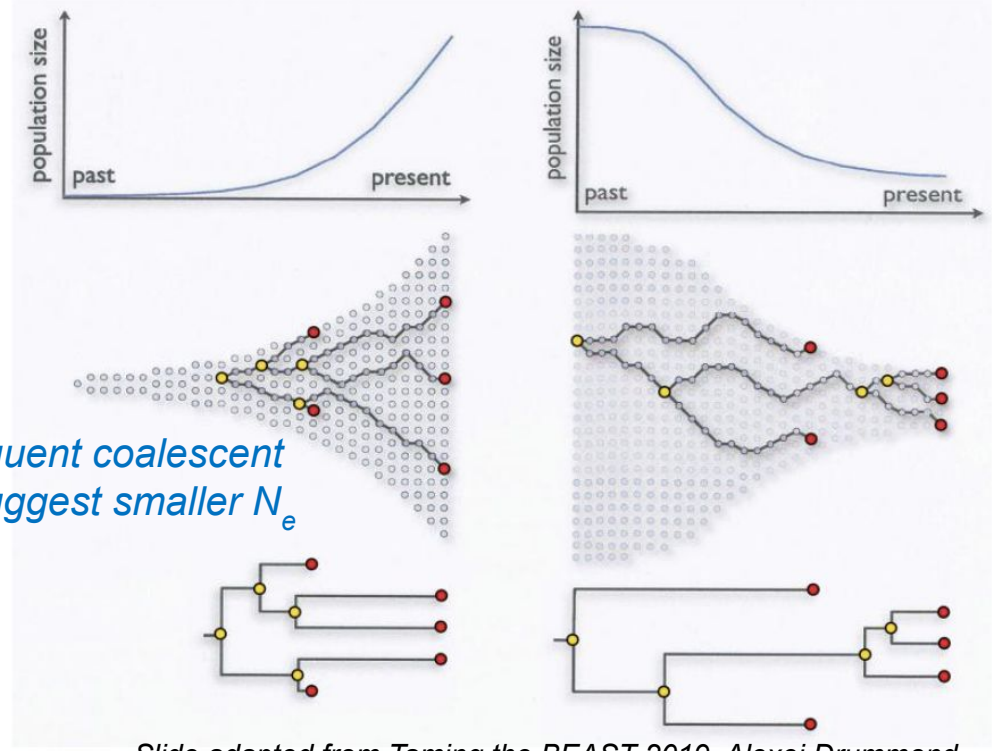
- Approximation to Wright-Fisher population dynamics (with large **N**)
- Trace ancestry of **n** samples in a population of size **N**
- Given **N** it is easy to calculate the probability for **2** nodes to coalesce in time **t**
- Calculate the probability of observing a given **tree** for a particular **N**
  → estimate **N** (**Ne** in practice)
- Easy to extend to time changing **N(t)**

Coalescent assumes:
mutations are neutral, no recombination or sampling bias, and the underlying susceptible population is described by the specified models

Prob. of coalescing for each pair at each generation = 1/N

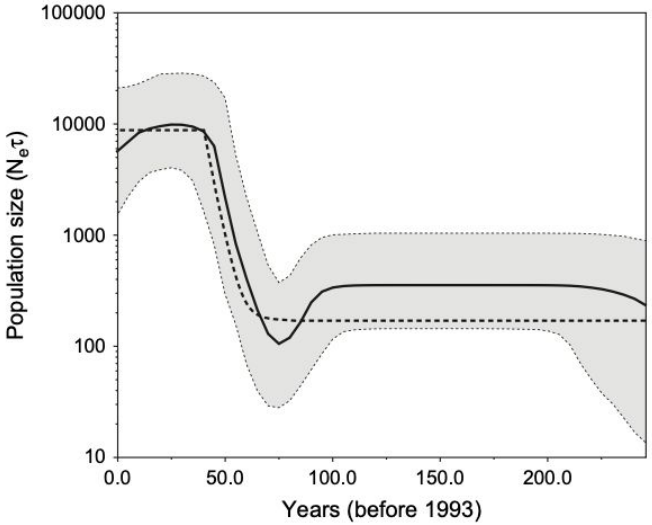
Slide adapted from Taming the BEAST 2019, Louis du Plessis

Coalescent applied to reconstruct the effective population size (N_e)



More frequent coalescent events suggest smaller N_e

Slide adapted from Taming the BEAST 2019, Alexei Drummond



Bayesian skyline plot derived from an alignment of Egyptian HCV sequences (Drummond et al., 2005)

Parametric and non-parametric population models

- Parametric changes in $N_{e,}$ constant or exponential (Pybus et al., 2001), or
- Non-parametric N_e (piecewise/epochal with/out smoothing)
  - Skyline model (Pybus et al., 2000)
  - Skyride model (Minin et al., 2008)
  - **Skygrid model + Gaussian Markov Random Field (Gill et al., 2013)**

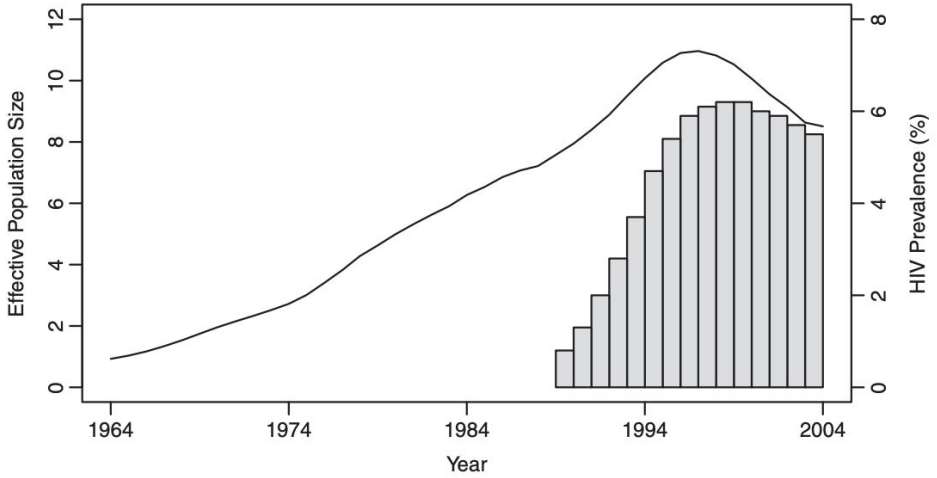
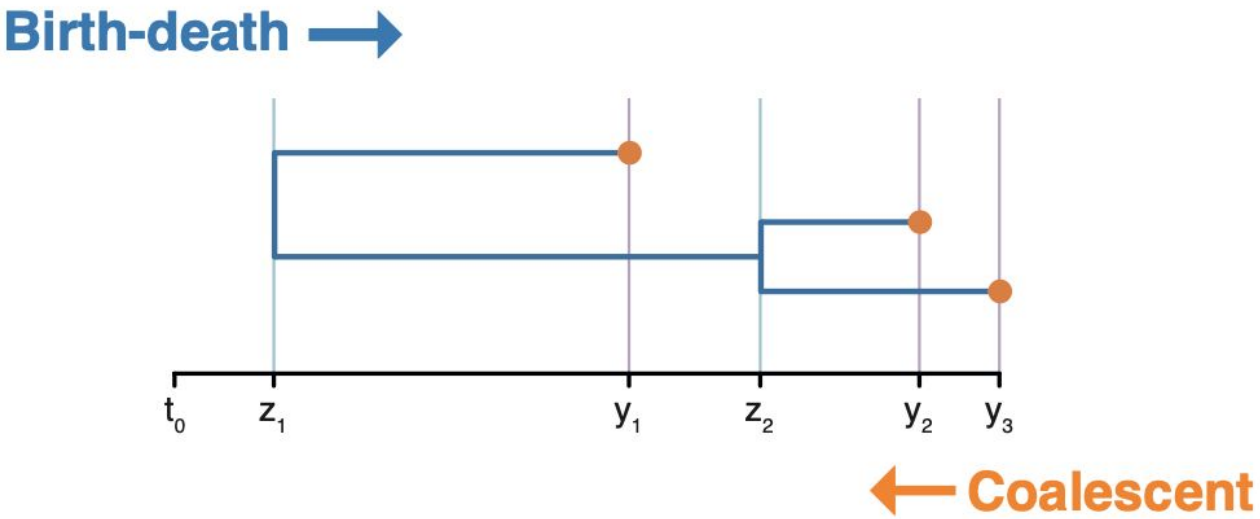
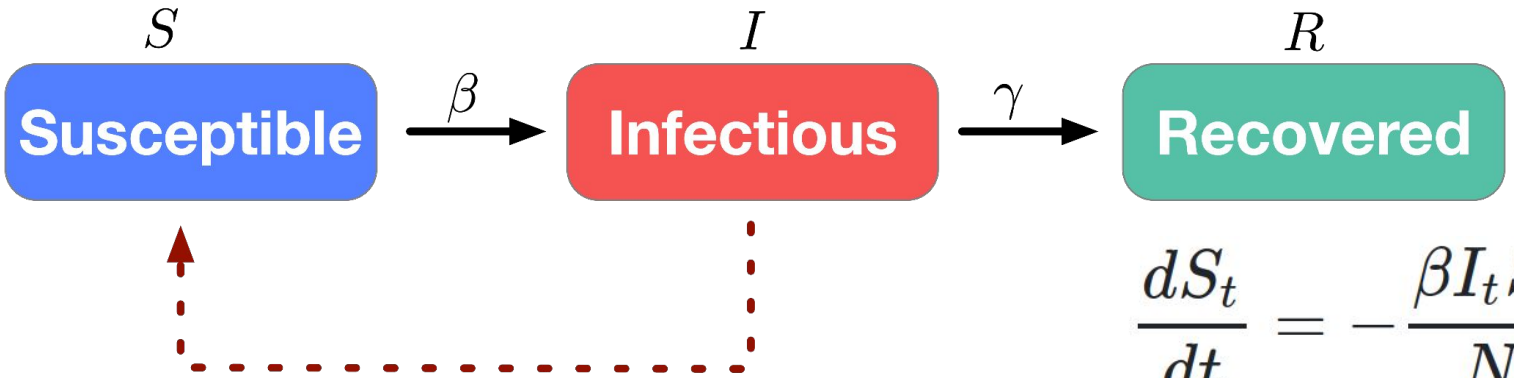


FIG. 5. Population history of HIV-1 CRF02_AG clade in Cameroon. The curve represents the estimated median log effective population size estimated from a multilocus alignment of 336 *gag*, *pol*, and *env* sequences sampled between 1996 and 2004. The bars represent estimated HIV prevalence counts in Cameroon.

Demographic models are based on a **coalescent process** or a **birth-death (BD) process**



Compartmental models are used to model infections



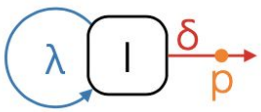
- S_t: the number of susceptible individuals
- I_t: the number of infectious individuals
- R_t: the number of recovered/deceased/immune individuals

$$\frac{dS_t}{dt} = -\frac{\beta I_t S_t}{N}$$
$$\frac{dI_t}{dt} = \frac{\beta I_t S_t}{N} - \gamma I_t$$
$$\frac{dR_t}{dt} = \gamma I_t$$

https://covid19.uclaml.org/model.html

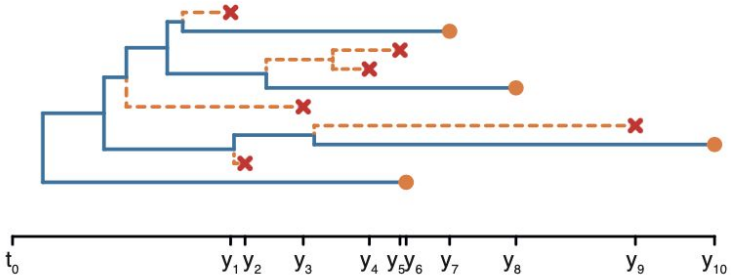
Assuming N = fixed pop

Birth-death-sampling (BDS) branching models



- λ — birth rate (lineages added to **full** tree)
- δ — death rate (lineages removed from the **full** tree)
- p — sampling probability (samples added to **sampled** tree)

- Forward-in-time branching process
- Events happen at different rates
  - infection/recovery
  - speciation/extinction
  - sampling/fossilization
  - …



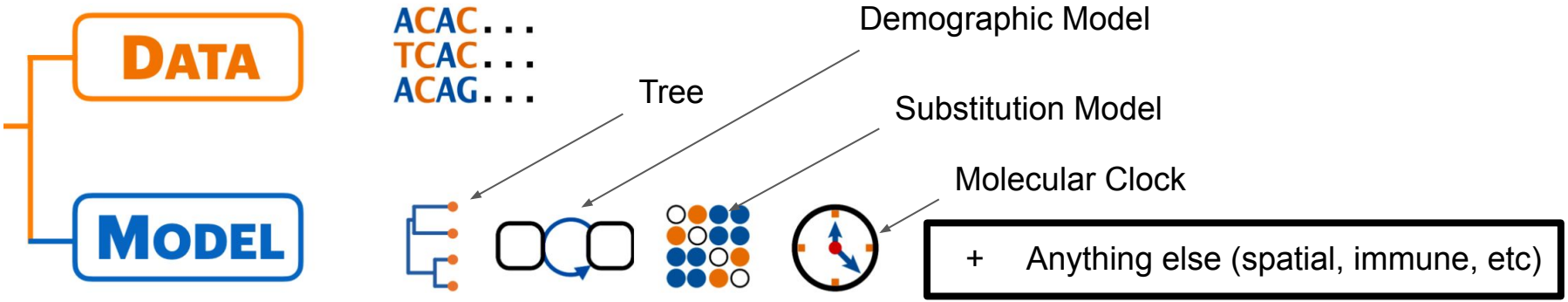
Slide adapted from Taming the BEAST 2019, Louis du Plessis

- Earliest BD model (Kendall, 1948)
- BD with binary characters (Maddison et al., 2007)
- BDS incomplete sampling (Stadler, 2009, 2010)
- Birth death skyline model: re-parameterized as effective reproductive number (Re) and become uninfectious rate (Stadler et al., 2013)

Assumptions

- *Vary by model re: sampling, the mass extinction events, whether sampling results in removal, and tree conditioning (MacPherson et al., 2021)*
- *All assume lineages are interchangeable, mutations are neutral, and random sampling*

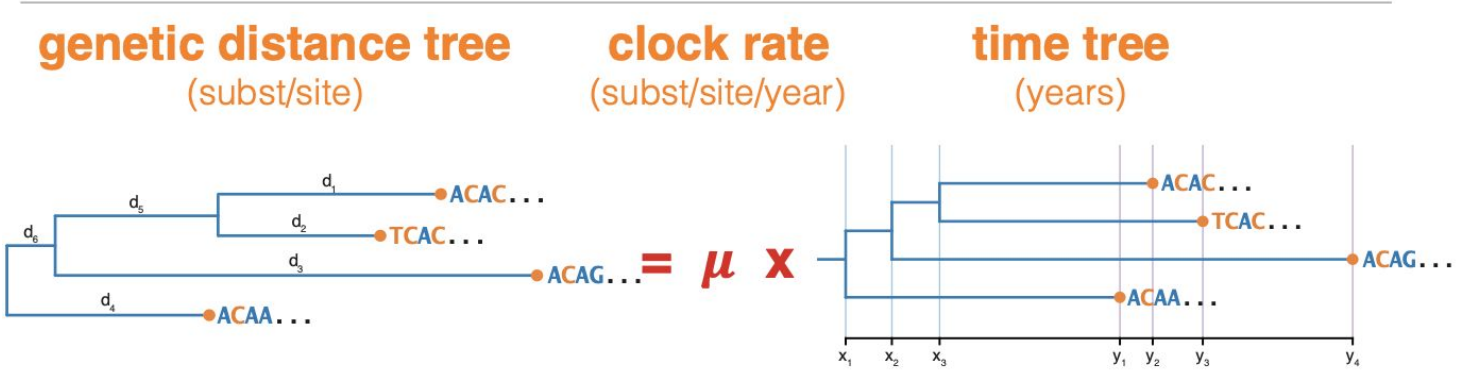
Bayesian inference is a key tool in phylodynamics



$$P(\text{tree}, \text{model} \mid \text{ACAC}\ldots,\ \text{TCAC}\ldots,\ \text{ACAG}\ldots) = \frac{P(\text{ACAC}\ldots,\ \text{TCAC}\ldots,\ \text{ACAG}\ldots \mid \text{tree}, \text{model})\, P(\text{tree}, \text{model})}{P(\text{ACAC}\ldots,\ \text{TCAC}\ldots,\ \text{ACAG}\ldots)}$$

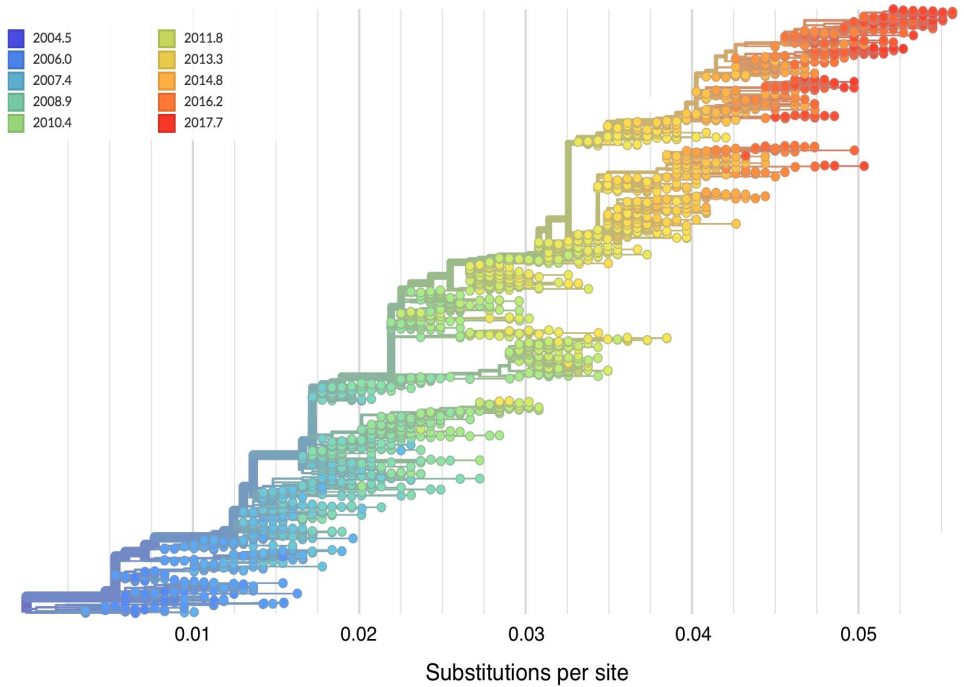
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Molecular clock model



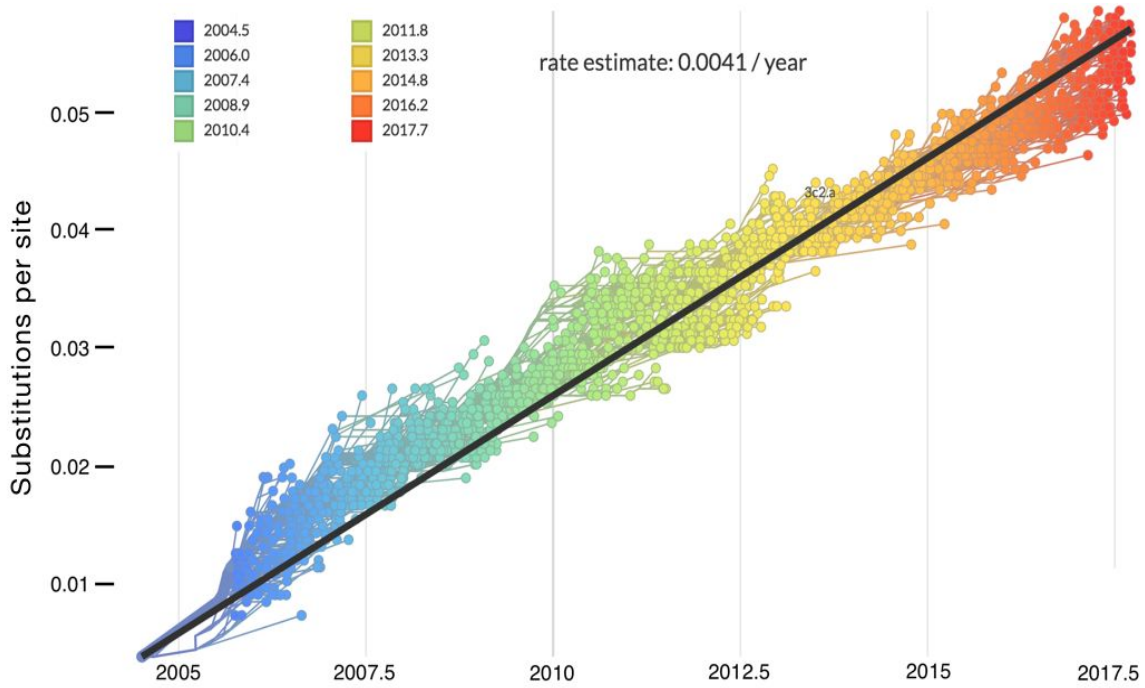
- Determines how quickly sequences are evolving along the tree
- **Genetic distance = Rate x Time**

Trees initially inferred as divergence from the root (substitutions/site)



https://github.com/trvrb/gs541-phylodynamics

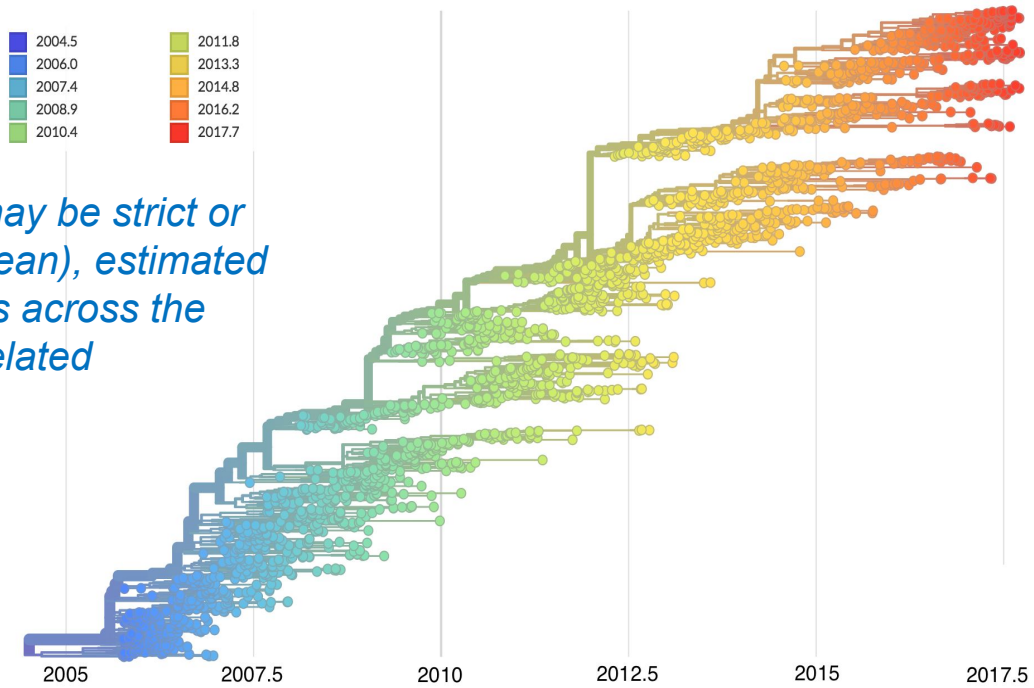
Root-to-tip linear regression used to estimate a strict molecular clock rate



https://github.com/trvrb/gs541-phylodynamics

Molecular clock assumptions are used to convert from divergence to time-scaled trees

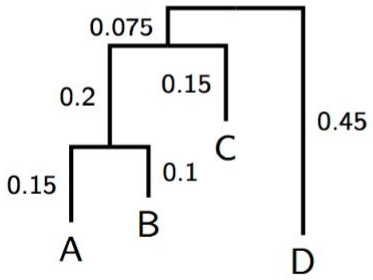
Molecular clock rate may be strict or relaxed (log-normal mean), estimated or fixed, homogeneous across the tree or local auto-correlated



https://github.com/trvrb/gs541-phylodynamics

Strict vs relaxed molecular clocks

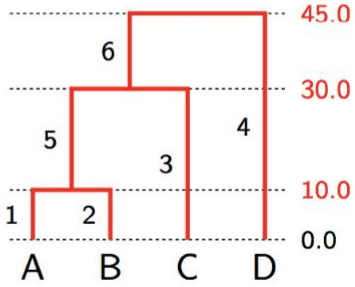
$$T = \vec{\mu} \star g$$



"substitution tree"

$$= \begin{pmatrix} 0.015 \\ 0.01 \\ 0.005 \\ 0.01 \\ 0.01 \\ 0.005 \end{pmatrix} \star$$

evolutionary rates
substitutions / site / unit time

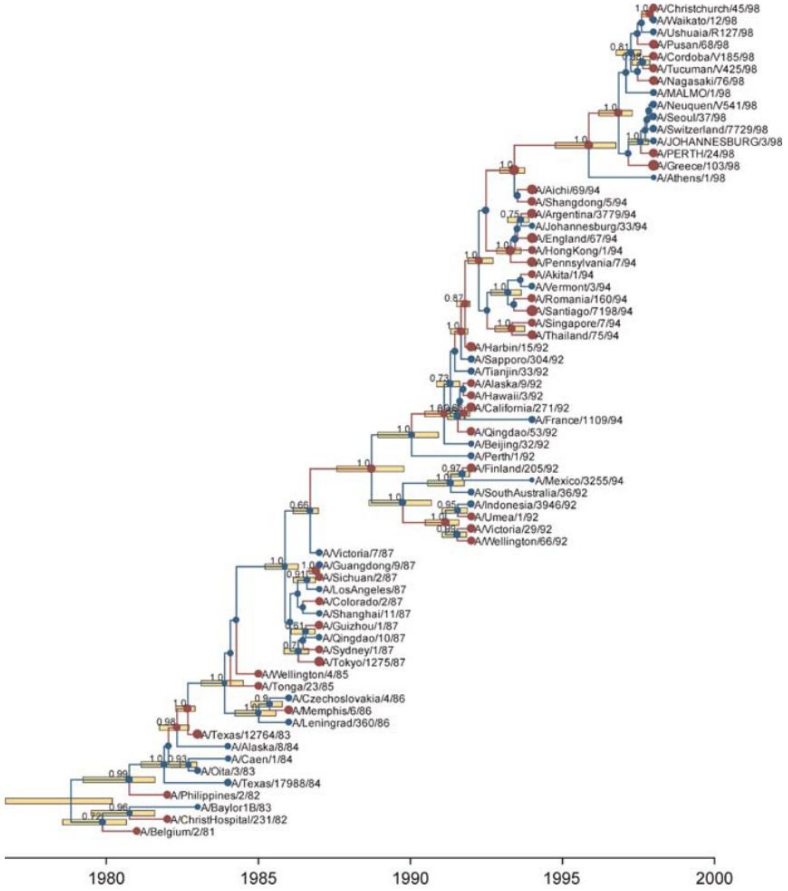


time tree

Molecular clock models

- Strict vs relaxed
- Lognormal
- Autocorrelated/local

Divergence dating



- Estimate the time of the most recent common ancestor (TMRCA) in time-scaled phylogenies
- Mean/median and 95% highest posterior density estimate (Bayesian), or 95% confidence interval (frequentist)

(Drummond 2006 PLoS Biology)

Figure 2. A Tree of 69 Influenza A Virus Sequences Drawn Randomly from the Posterior Distribution

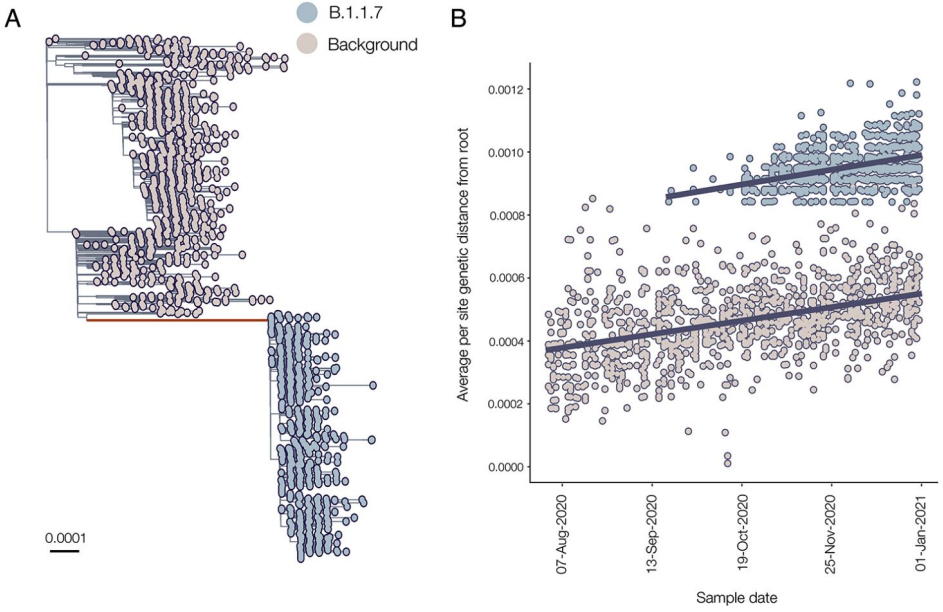
The divergence times correspond to the mean posterior estimate of their age in years. The yellow bars represent the 95% HPD interval for the divergence time estimates. Both the mean and 95% HPD of the divergence times were calculated conditional on the existence of the clade defined by the divergence. Each node in the tree that has a posterior probability greater than 0.5 is labeled with its posterior probability. The sampling times of the

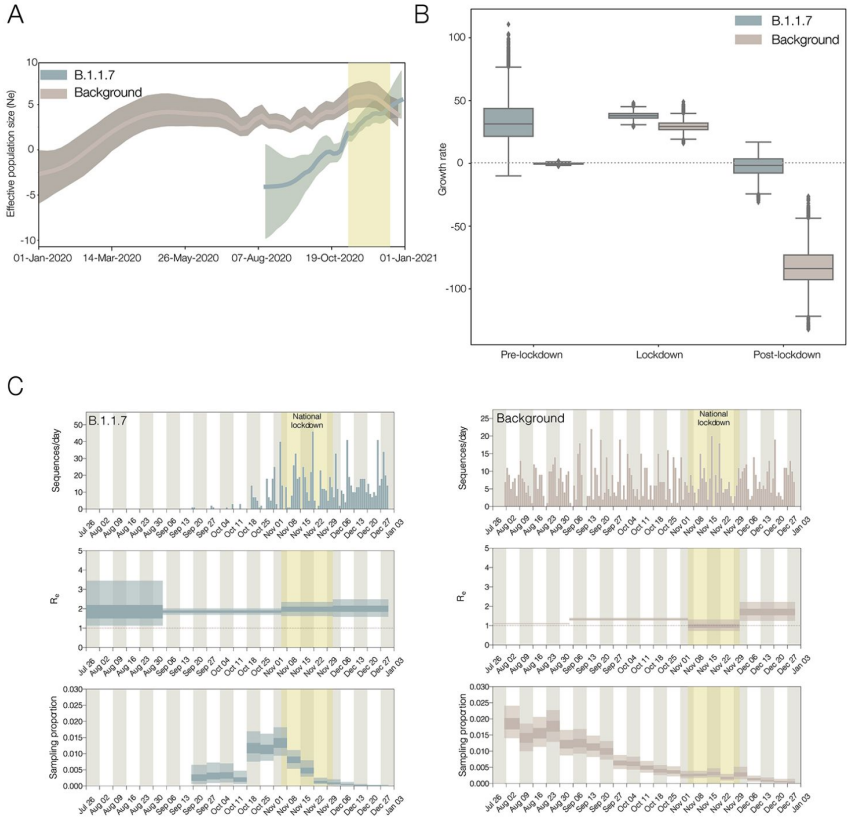
VIRUS EVOLUTION

Virus Evolution, 2022, **8(2)**, 1–13
DOI: https://doi.org/10.1093/ve/veac080
Advance access publication date: 26 August 2022
Research Article

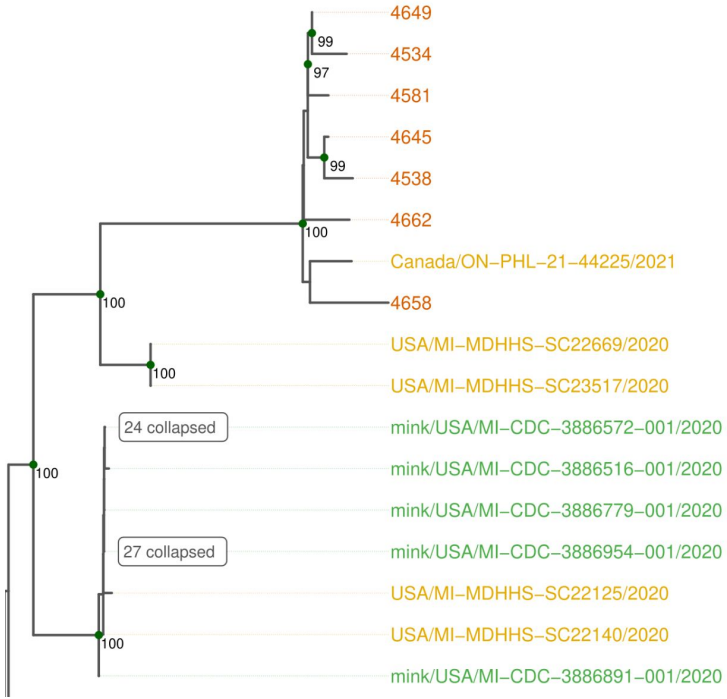
The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK

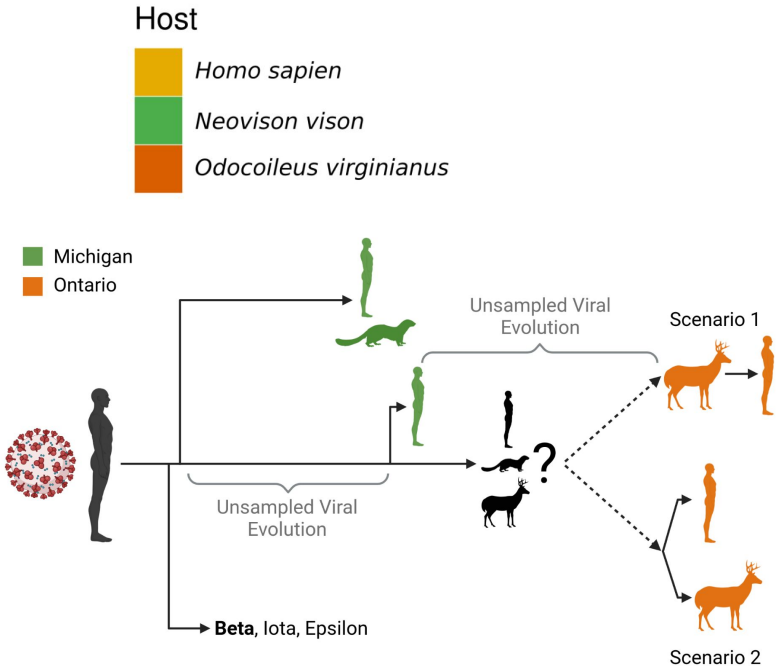
Verity Hill,[1,2,•,‡] Louis Du Plessis,[3,4] Thomas P. Peacock,[5] Dinesh Aggarwal,[6,7,8,9] Rachel Colquhoun,[1,§] Alesandro M. Carabelli,[8,••] Nicholas Ellaby,[7] Eileen Gallagher,[7] Natalie Groves,[7] Ben Jackson,[1,††] J. T. McCrone,[1,‡‡] Áine O'Toole,[1,§§] Anna Price,[10] Theo Sanderson,[6,11] Emily Scher,[1,••] Joel Southgate,[10] Erik Volz,[12,†††] The COVID-19 Genomics UK (COG-UK) Consortium,[†] Wendy S. Barclay,[5] Jeffrey C. Barrett,[6] Meera Chand,[7,13] Thomas Connor,[10,14,‡‡‡] Ian Goodfellow,[15] Ravindra K. Gupta,[8,16] Ewan M. Harrison,[6,8,17] Nicholas Loman,[18] Richard Myers,[7] David L. Robertson,[19,§§§] Oliver G. Pybus,[3,20,••••] and Andrew Rambaut[1,††††]



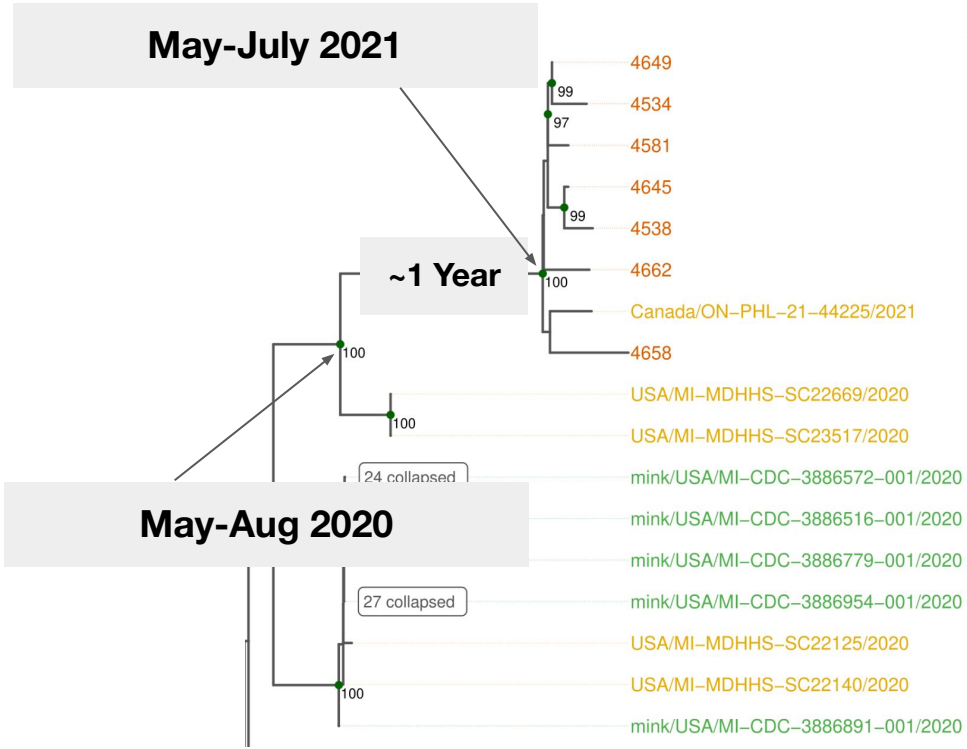


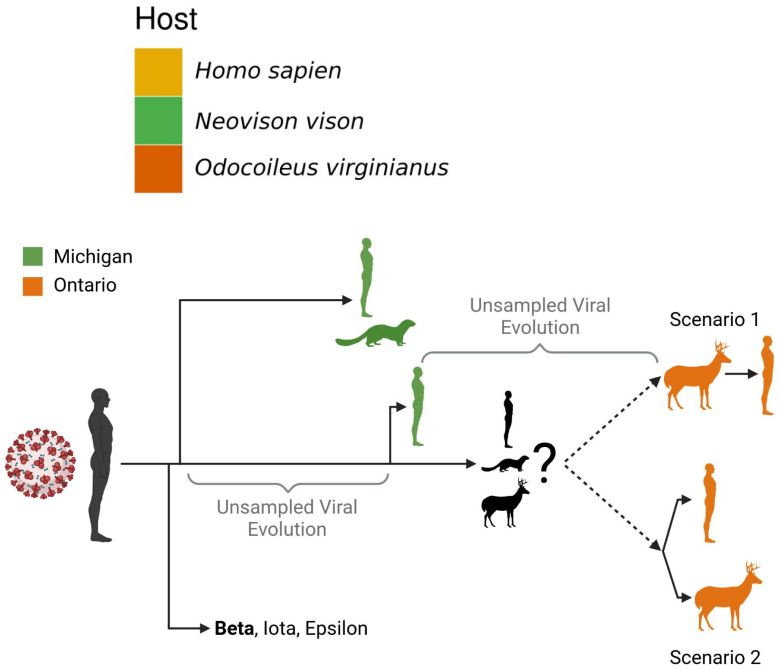
Time-trees let us estimate timing of unobserved events





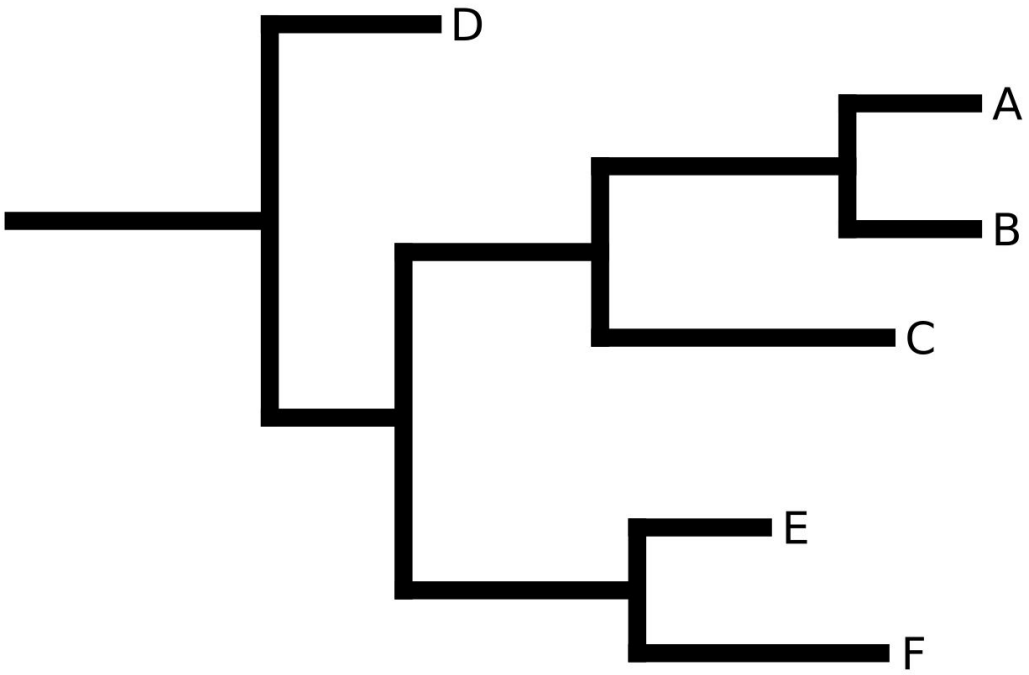
Time-trees let us estimate timing of unobserved events



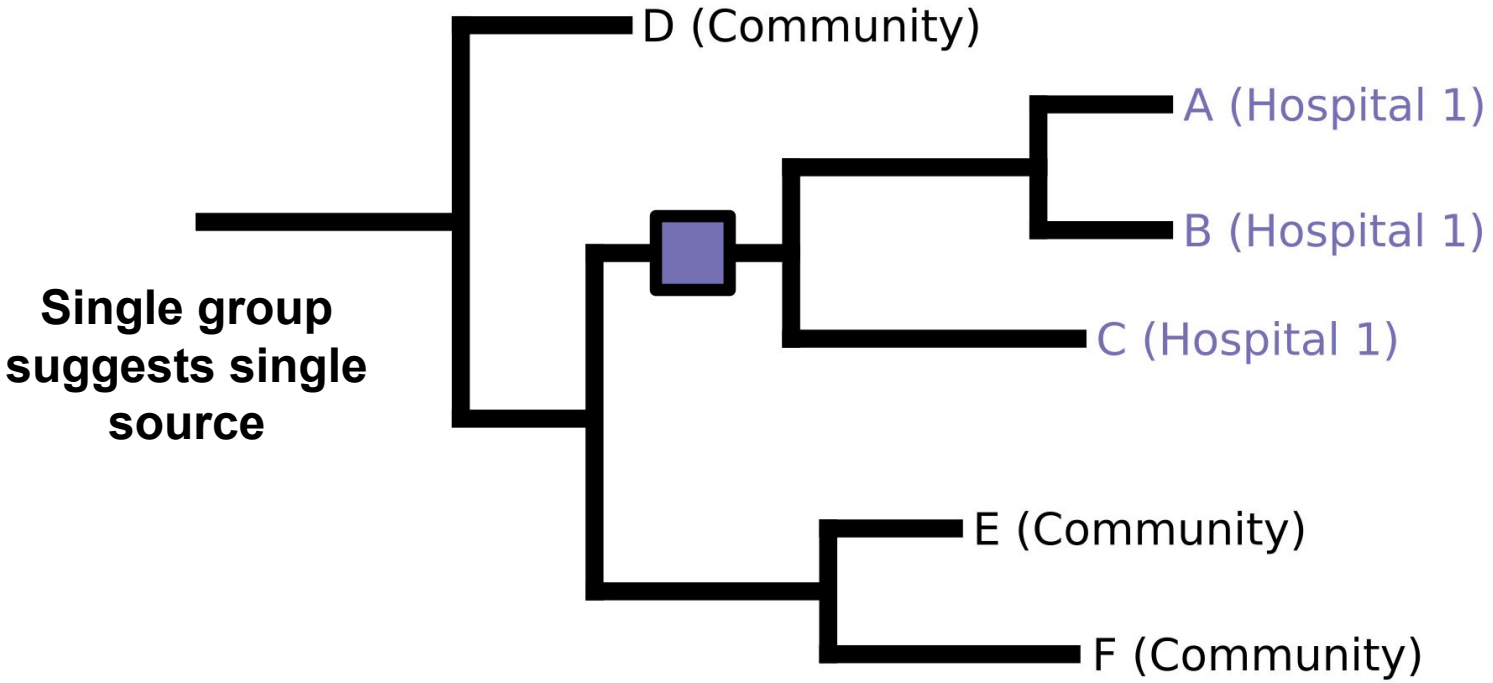


Inferring ancestral traits on our tree

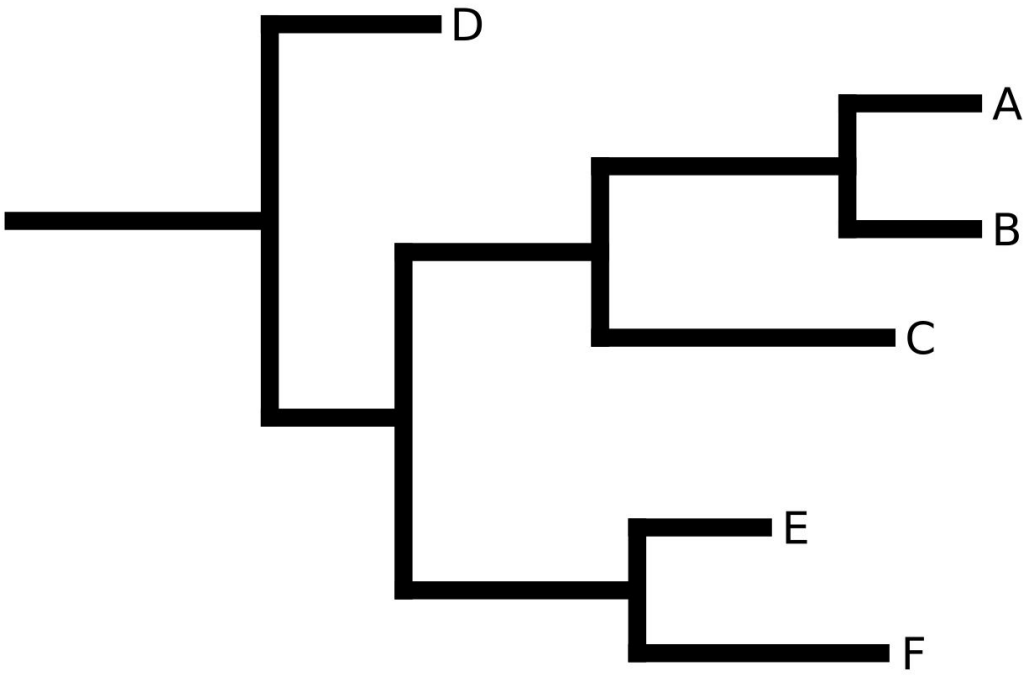
Trace sources of outbreaks



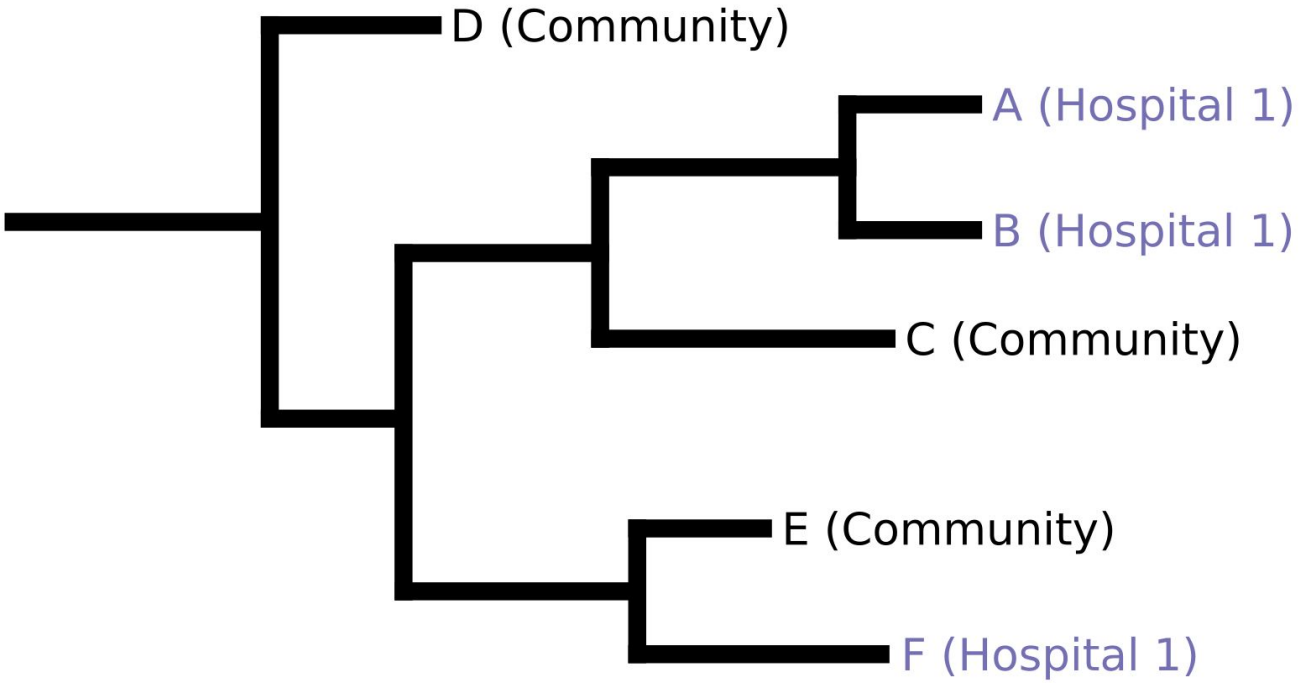
Trace sources of outbreaks



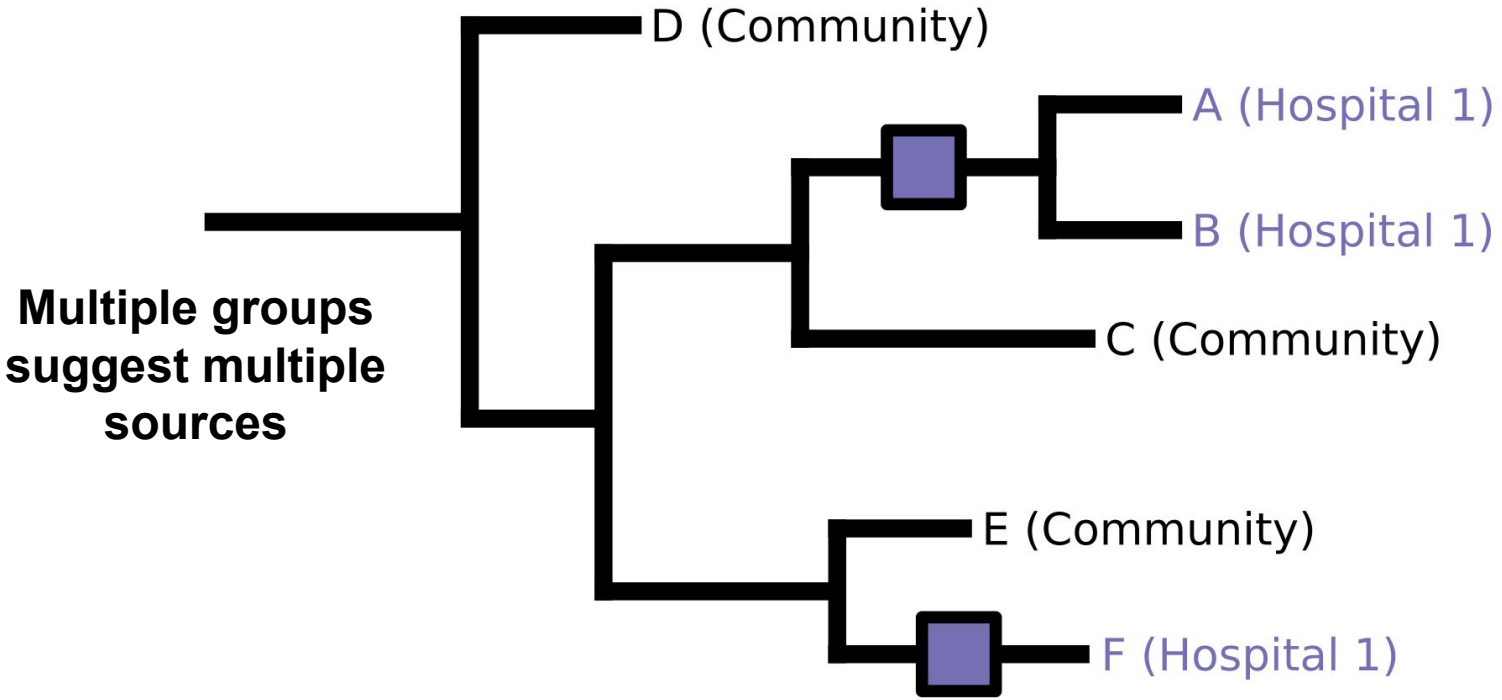
Trace sources of outbreaks



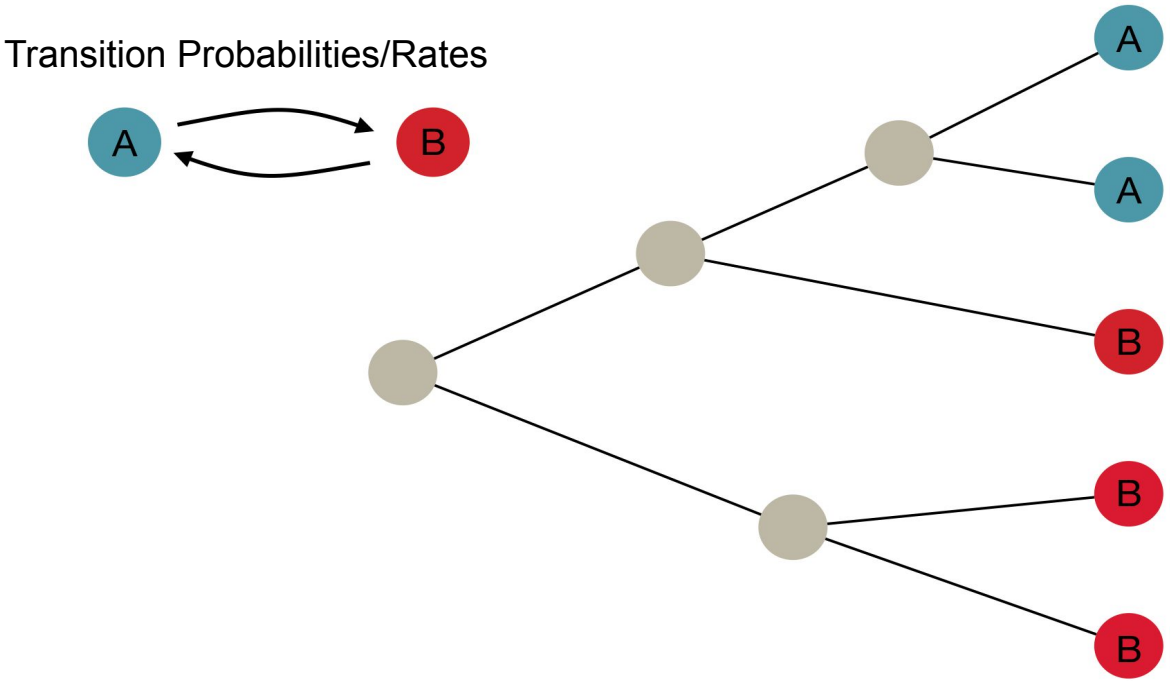
Trace sources of outbreaks



Trace sources of outbreaks

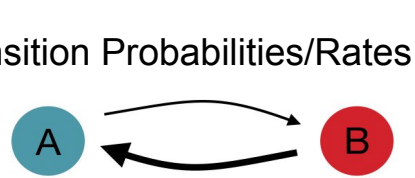


Inferring internal ancestral states from observed tips



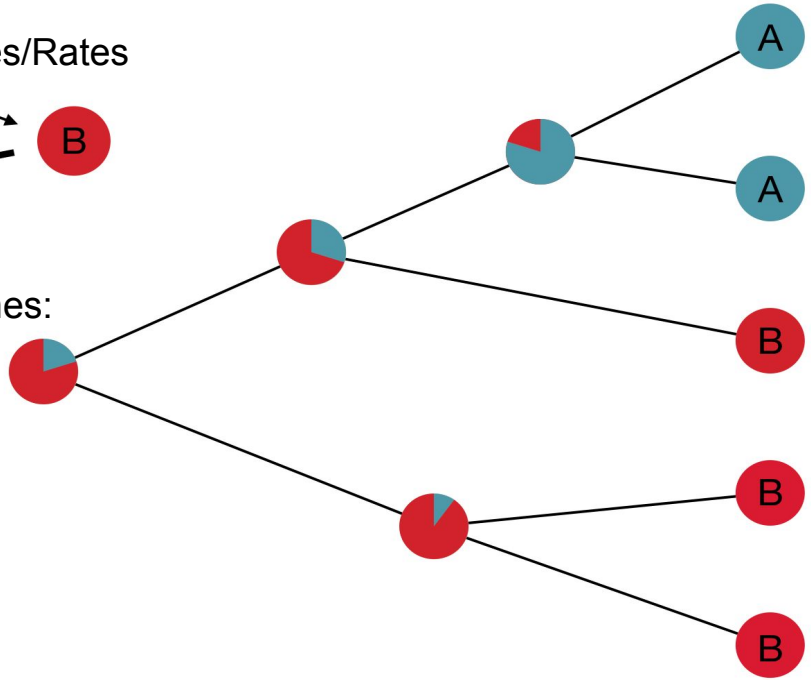
https://github.com/trvrb/gs541-phylodynamics

Inferring internal ancestral states from observed tips



Different statistical inference approaches:

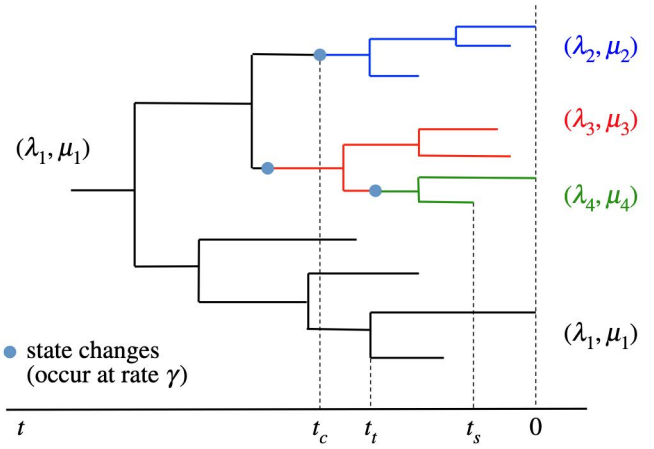
- Parsimony
- Maximum likelihood
- Bayesian



https://github.com/trvrb/gs541-phylodynamics

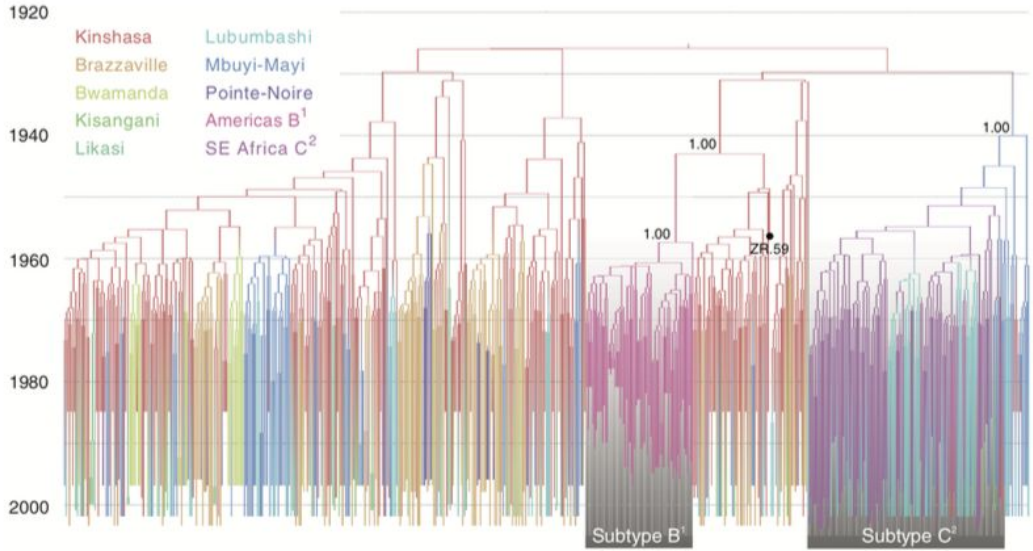
Bayesian phylogeography

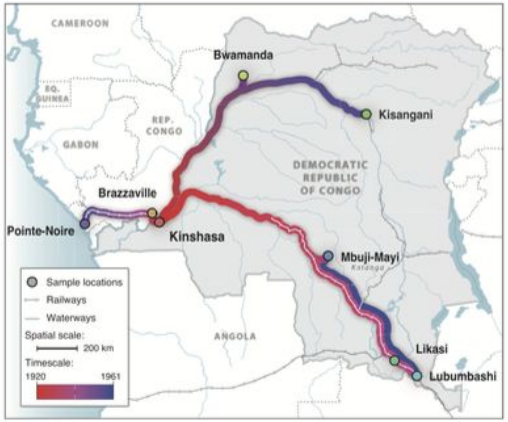
- Coalescent with discrete trait diffusion (Lemey et al., 2009)
  - The ’mugration’ model treats discrete traits as stochastically evolving characters
- Travel-aware coalescent with unsampled tips (Lemey et al., 2020)
- Structured coalescent approximation (Müller, et al 2018; de Maio et al., 2015)
- Structured birth-death (BD) with migration (Kuëhnert et al., 2016; Barido-Sottani et al., 2018)



Barido-Sottani et al., 2018

Bayesian phylogeography used to estimate the origins of the HIV-1 epidemic

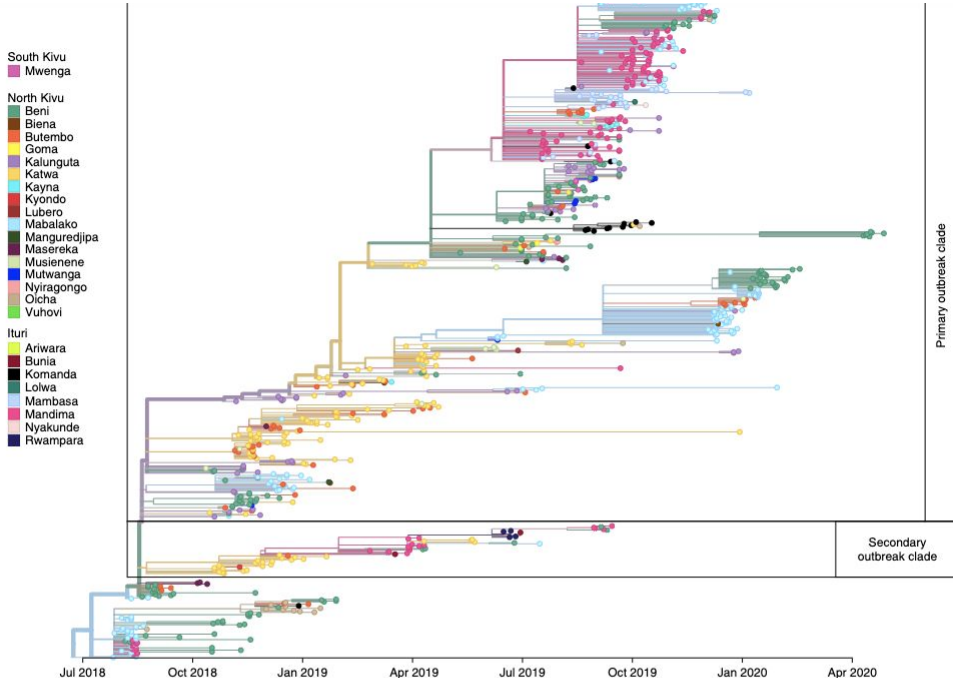


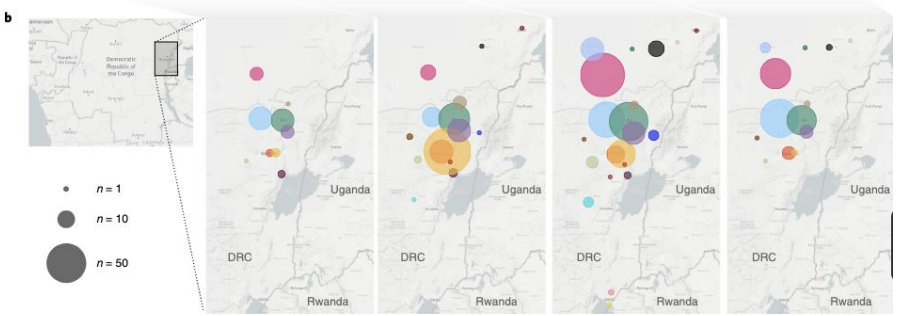


Faria et al. 2014 Science

Reconstruction of Ebola virus in DRC in 2018

(Kingada-Lusamaki et al., 2021, *Nature Medicine)*





Same approach works for traits like mutations or hosts

Or any evolving trait: geography, mutations, host species, drug resistance phenotypes, hospital outbreaks

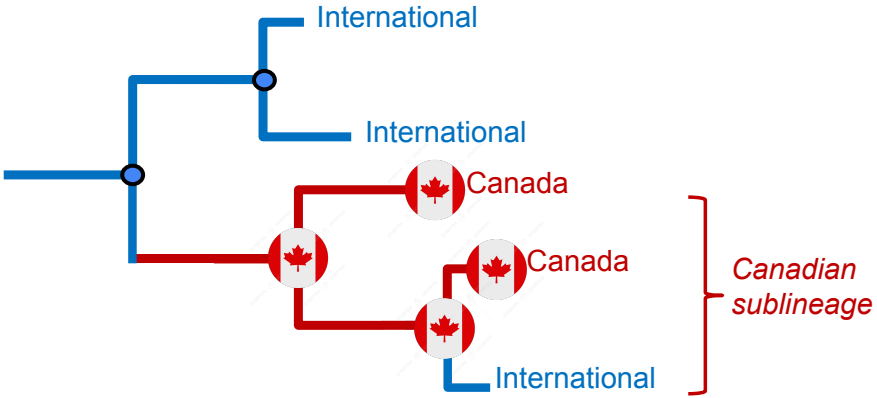
https://github.com/trvrb/gs541-phylodynamics

Phylogeography can be used to quantify viral importations and sublineages

Phylogeographic inference of viral introductions including sublineages & singletons

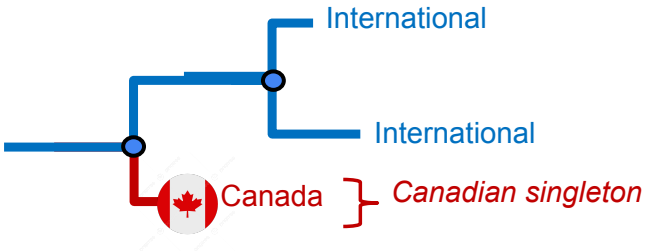
Sublineages

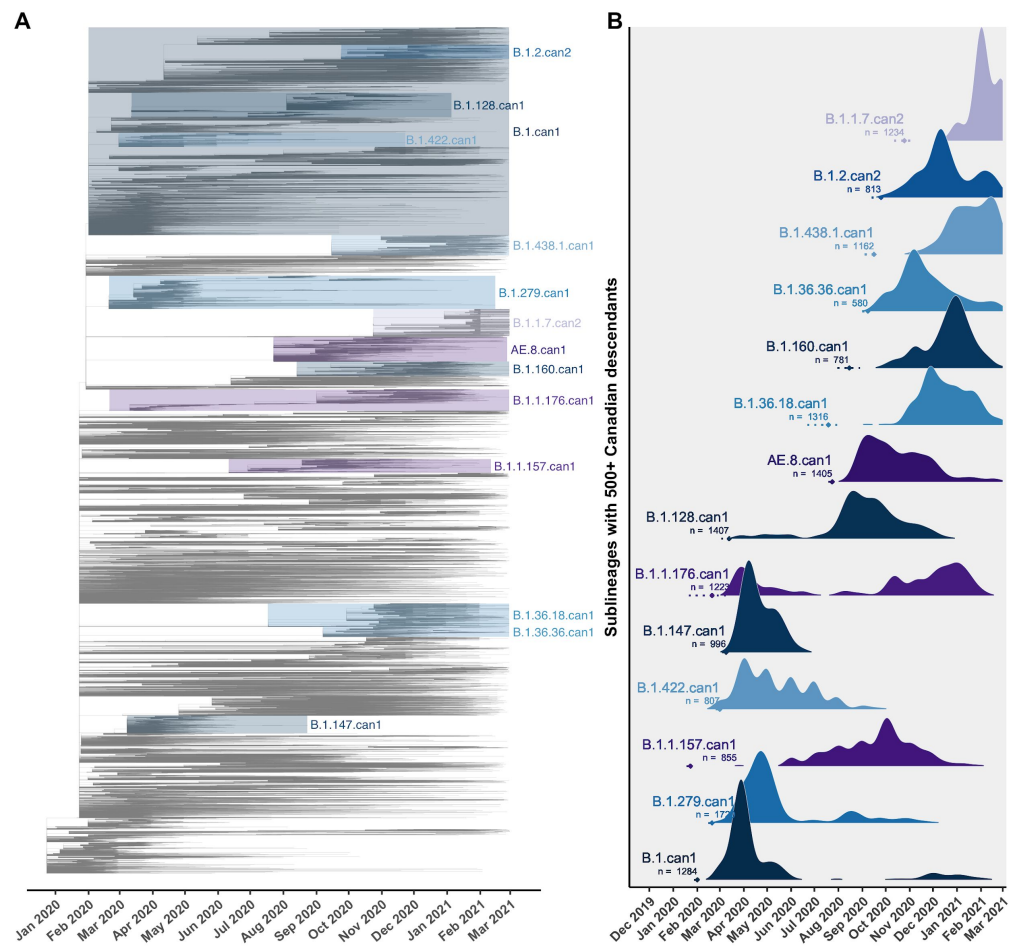
- International viral introductions resulting in sampled transmission



Singletons

- International viral introductions with no further sampled transmission



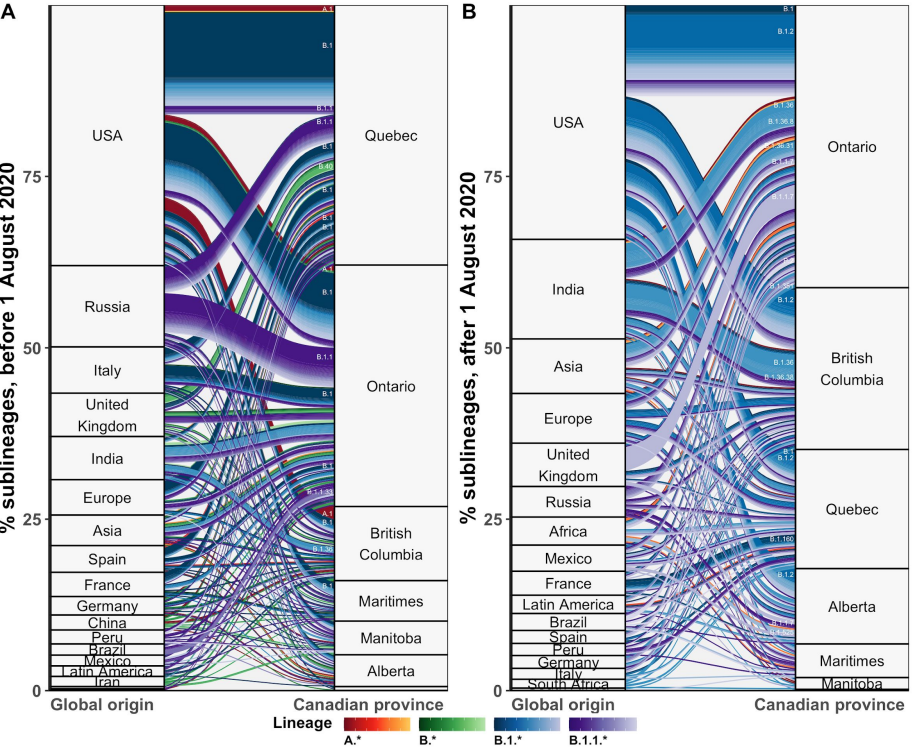


McLaughlin *et al.* eLife 2022;11:e73896.

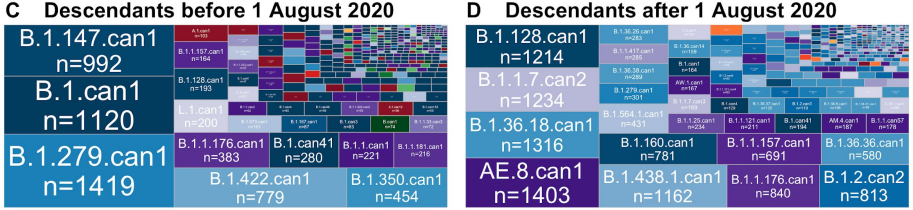
A few key sublineages dominated first waves of SARS-CoV-2 in Canada

- 680 (95% CI: 658-703) sublineages and 1582 (1501-1663) singletons up to 1 March 2021
- 0.7% of introductions led to 77% of sequences

Diverse origins of Canadian SARS-CoV-2 sublineages

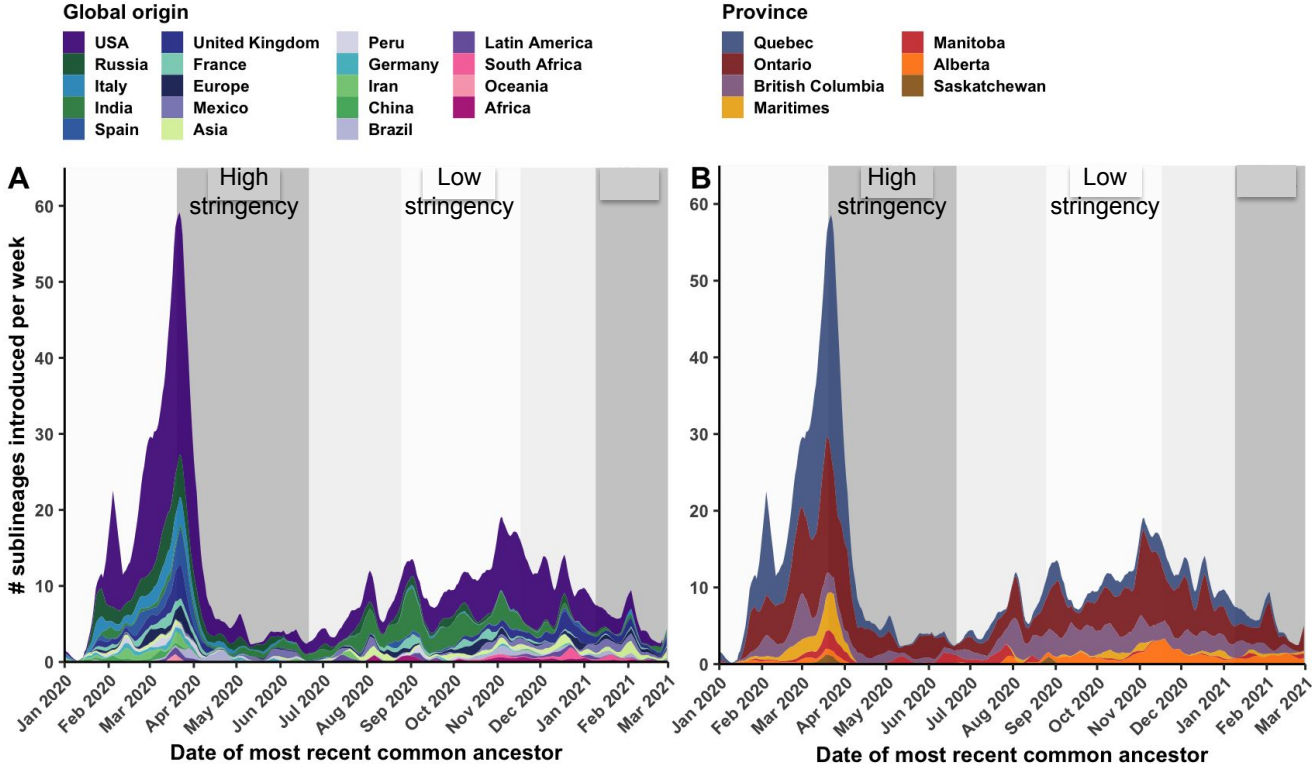


USA was the largest introduction source in first and second waves



McLaughlin *et al.* eLife 2022;11:e73896.

March 2020 restrictions reduced, but did not eliminate SARS-CoV-2 introductions into Canada



McLaughlin *et al.* eLife 2022;11:e73896.

- 10-fold decrease in sublineage introduction rate within 4 weeks

What about evolutionary forces like selection?

dN/dS is one way to detect selection

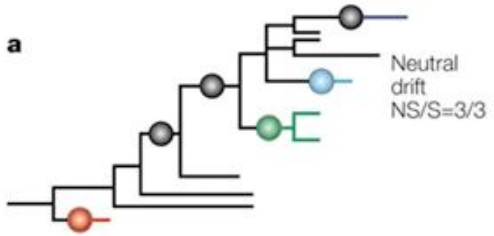
dN = non-synonymous mutations (normalised)

dS = synonymous mutations (normalised)

dN/dS is one way to detect selection

dN = non-synonymous mutations (normalised)

dS = synonymous mutations (normalised)



dN/dS ~ 1 : drift/neutral selection

Bush, R. Predicting adaptive evolution. *Nat Rev Genet* **2**, 387–392 (2001). https://doi.org/10.1038/35072023

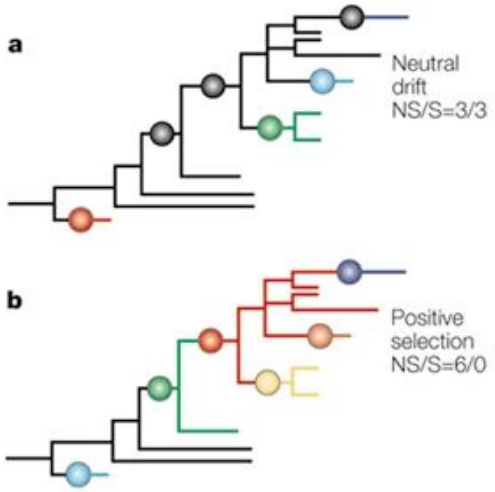
dN/dS is one way to detect selection

dN = non-synonymous mutations (normalised)

dS = synonymous mutations (normalised)

dN/dS > 1 : adaptive/positive selection

dN/dS ~ 1 : drift/neutral selection



Bush, R. Predicting adaptive evolution. *Nat Rev Genet* **2**, 387–392 (2001). https://doi.org/10.1038/35072023

dN/dS is one way to detect selection

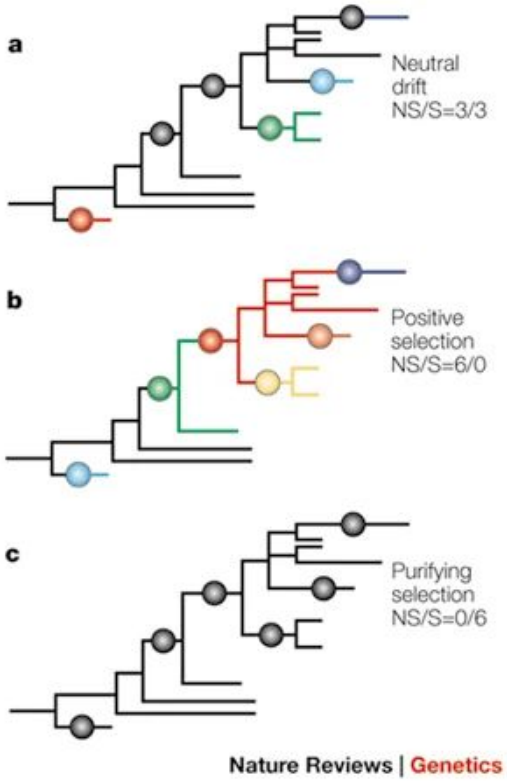
dN = non-synonymous mutations (normalised)

dS = synonymous mutations (normalised)

dN/dS > 1 : adaptive/positive selection

dN/dS ~ 1 : drift/neutral selection

dN/dS < 1: purifying/negative selection



Bush, R. Predicting adaptive evolution. *Nat Rev Genet* **2**, 387–392 (2001). https://doi.org/10.1038/35072023

dN/dS is one way to detect selection

dN = non-synonymous mutations (normalised)

dS = synonymous mutations (normalised)

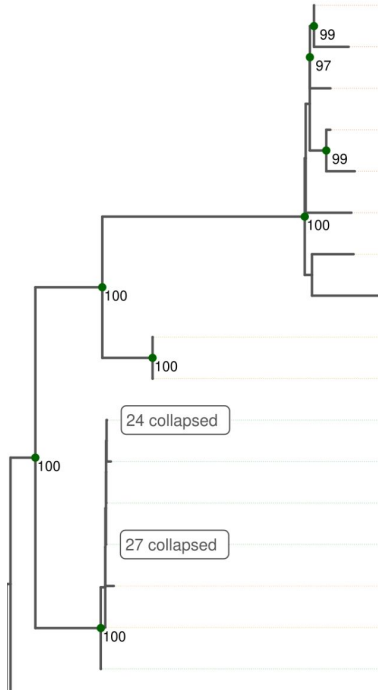
dN/dS > 1 : adaptive/positive selection

dN/dS ~ 1 : drift/neutral selection

dN/dS < 1: purifying/negative selection

Challenges:

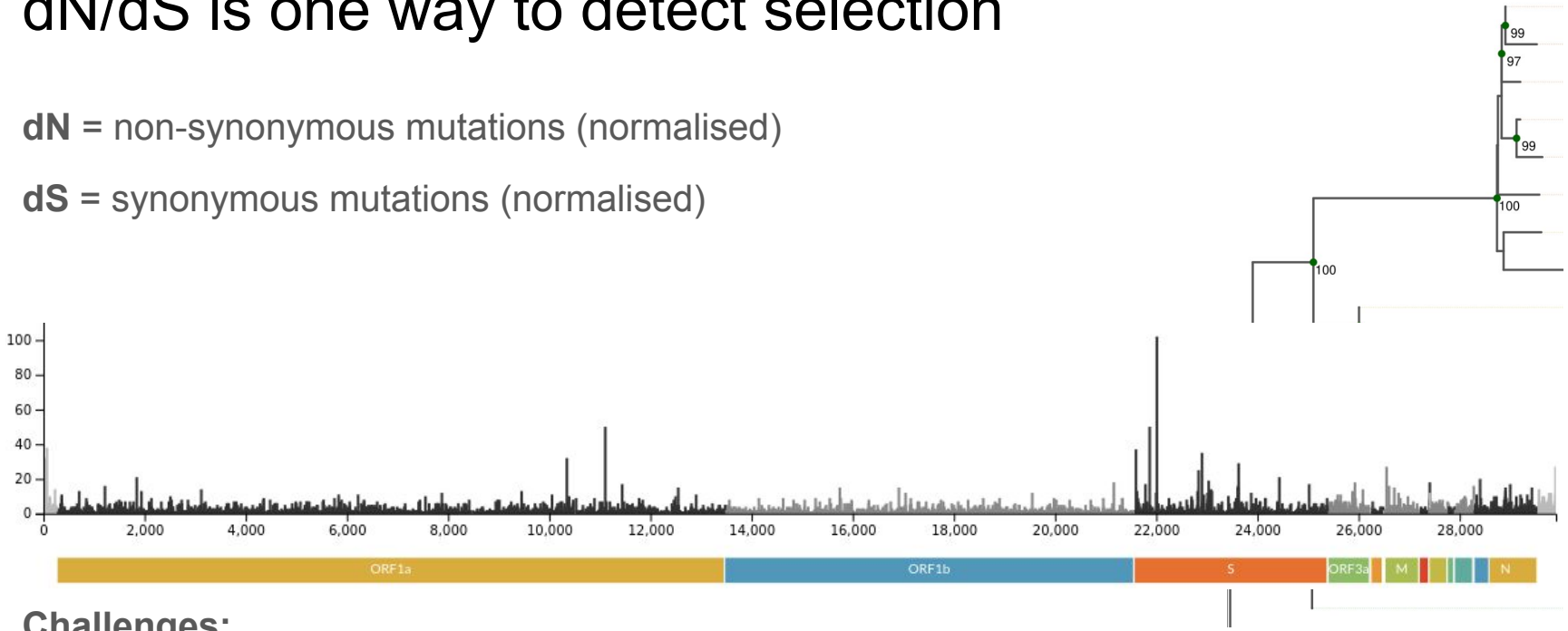
- Mutation rates vary over time/groups



dN/dS is one way to detect selection

dN = non-synonymous mutations (normalised)

dS = synonymous mutations (normalised)



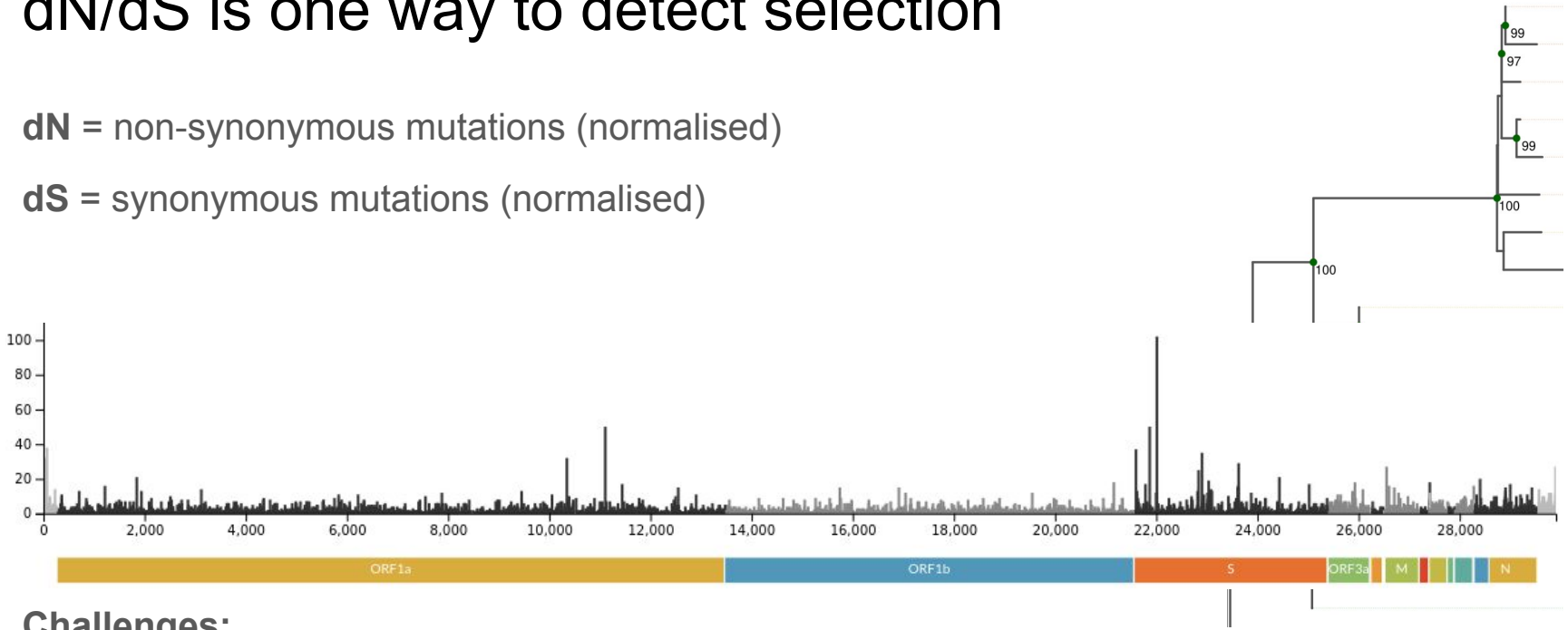
Challenges:

- Mutation rates vary over time/groups
- Mutation rates vary across genomes

dN/dS is one way to detect selection

dN = non-synonymous mutations (normalised)

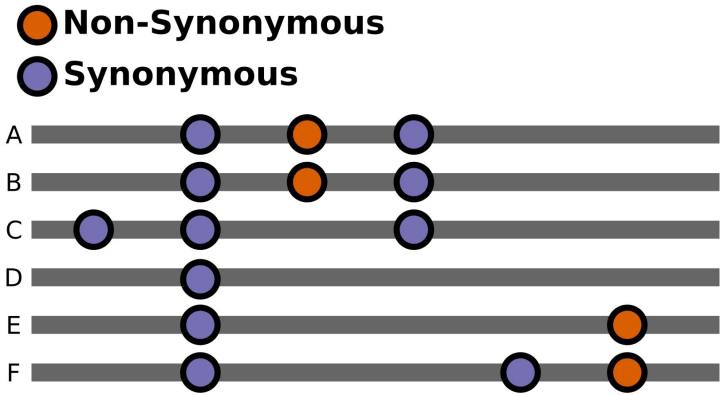
dS = synonymous mutations (normalised)



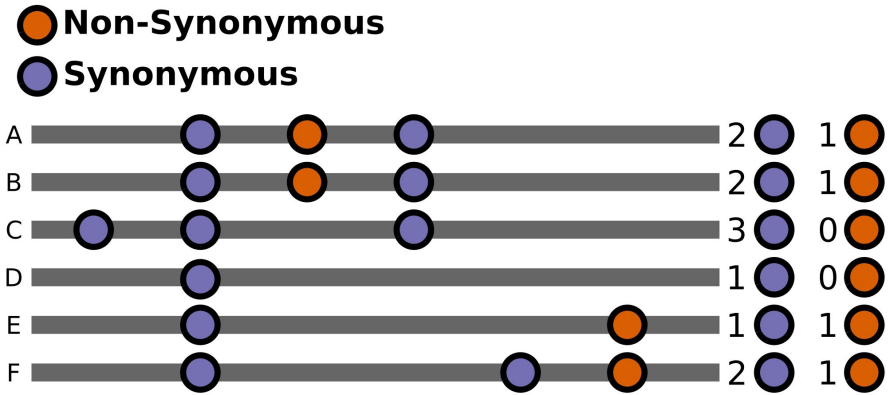
Challenges:

- Mutation rates vary over time/groups
- Mutation rates vary across genomes
- **Genomes are related** (mutations are non-independent)

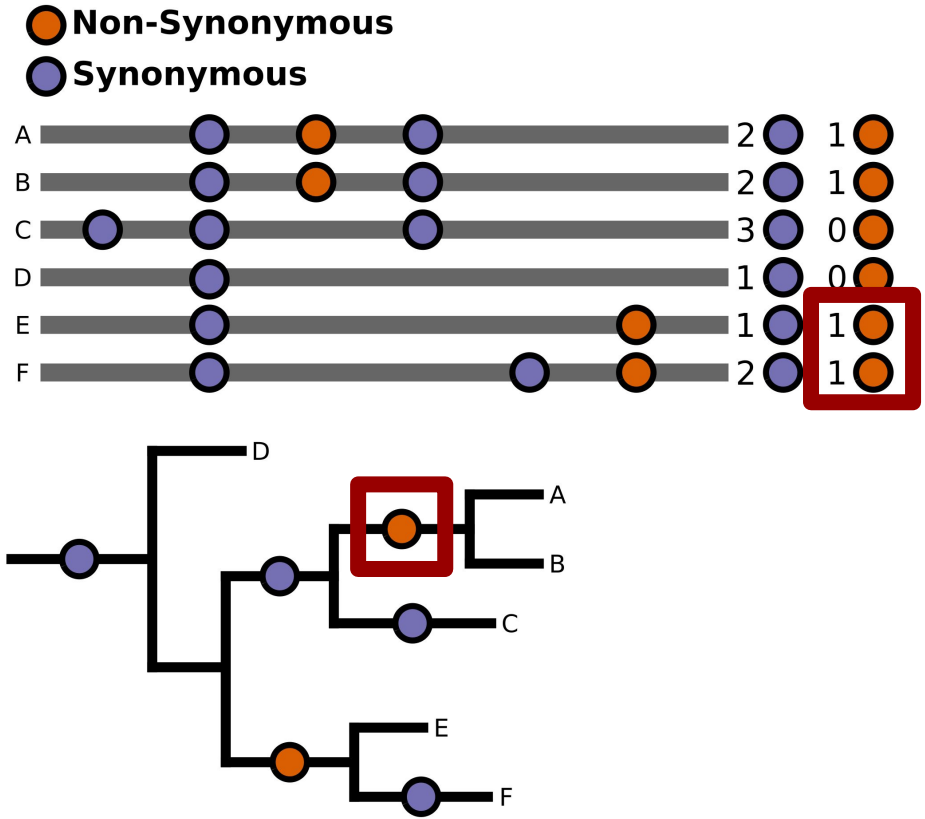
Non-independence of events in related genomes



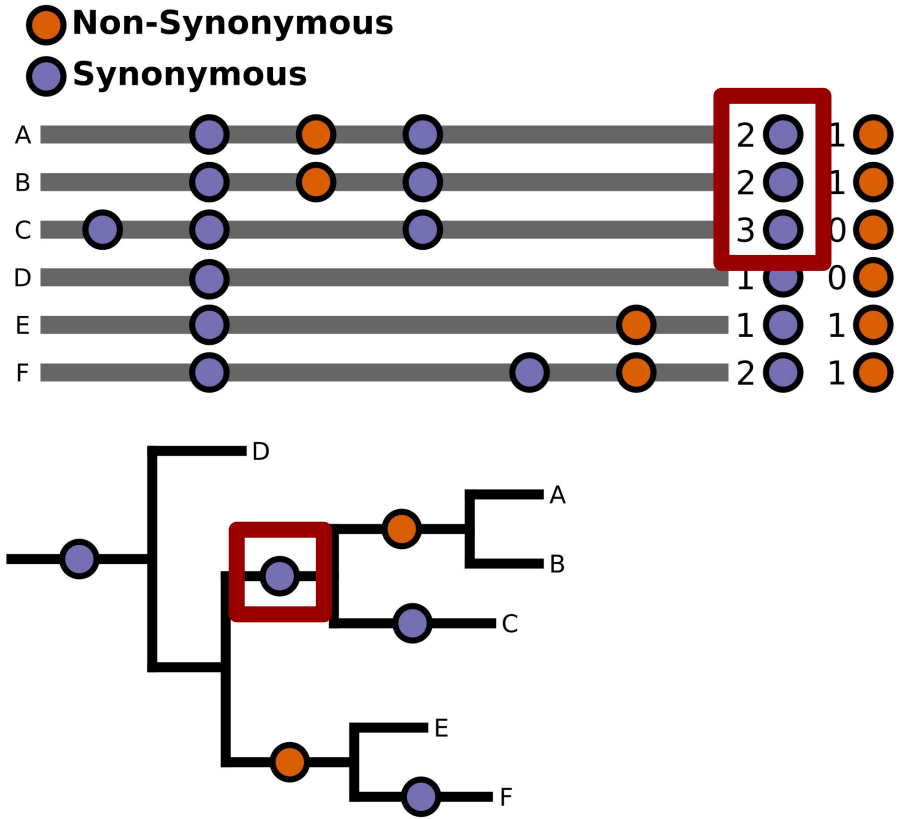
Non-independence of events in related genomes



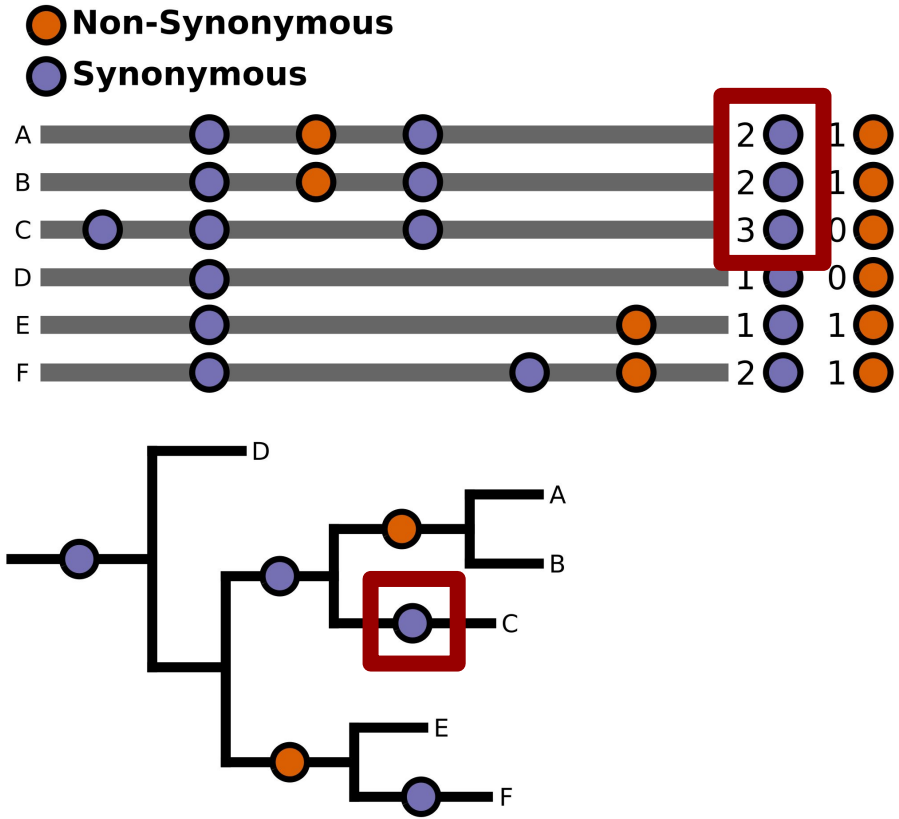
Non-independence of events in related genomes



Non-independence of events in related genomes

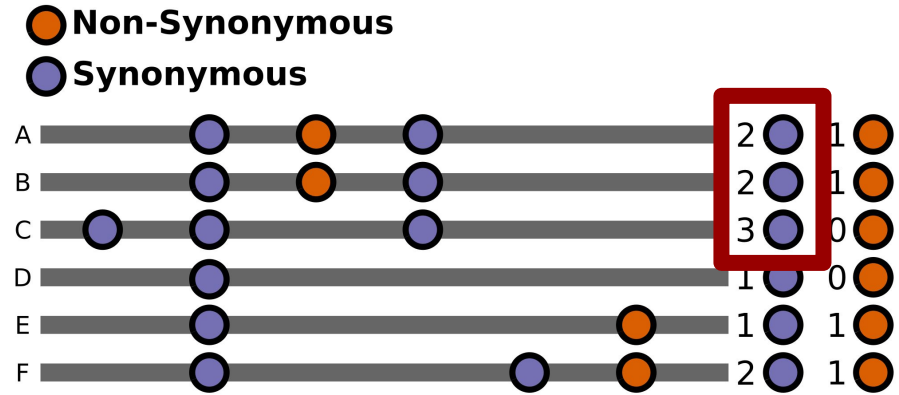


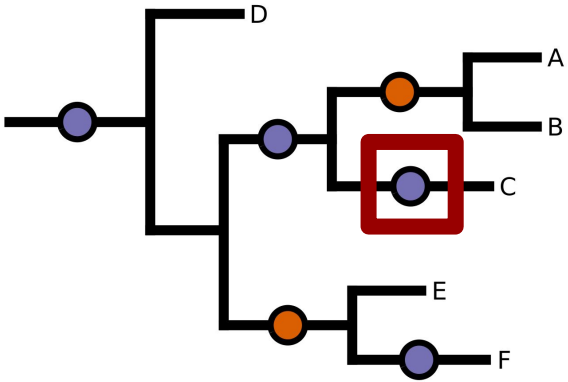
Non-independence of events in related genomes



- Phylogeny captures dependency structure of genomic data

Non-independence of events in related genomes



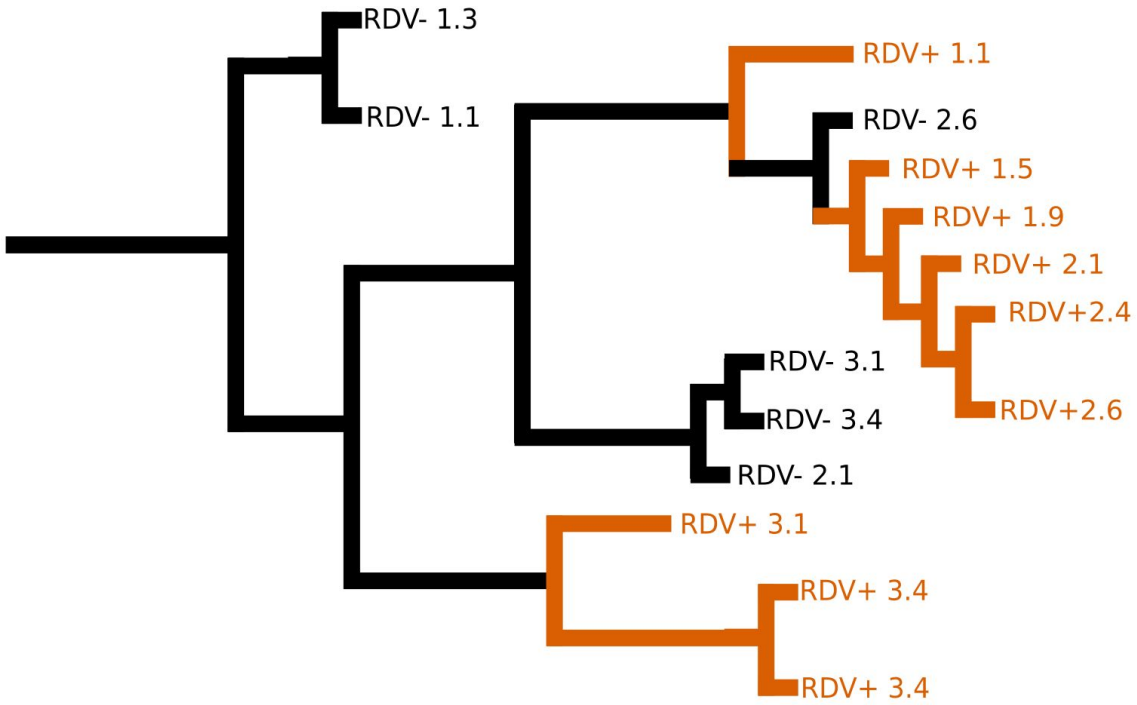


- Phylogeny captures dependency structure of genomic data
- Informs error term for models (e.g., regression)

- **a**daptive **B**ranch-**S**ite **R**andom **E**ffects **L**ikelihood: Is there a significant proportion of sites within selected branches with dN/dS > 1

Smith, MD et al. "Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection." Mol. Biol. Evol. 32, 1342–1353 (2015).

Testing for remdesevir resistance selection



Nirmalarajah, Kuganya, et al. "Use of whole genome sequencing to identify low-frequency mutations in SARS-CoV-2 patients treated with remdesivir." *Influenza and Other Respiratory Viruses* 17.9 (2023): e13179.

Limitations of phylodynamic analyses

- Biases due to non-random sampling: outbreaks, hospitalization, age, co-morbidity, changes in testing criteria
- Multiple sources of uncertainty: sequence inclusion, sequencing error, tree inference, polytomies/identical sequences, model assumptions
- Structuring sub-populations appropriately is challenging
- Misspecified priors: evolutionary clock rate, serial interval, reproduction number, dispersion, symmetrical migration rates

Summary

- Pathogen **evolution** and **epidemiology** are intrinsically linked
- Phylogenies are structured by **sampling**, **ecology, evolution, and epidemiology**
- Genomics provides insights into **evolution and unobserved events**
- Phylodynamics heavily uses **Bayesian phylogenetic models**
- Can use these approaches to do many things including:
  - Reconstruct **transmission**
  - Infer **timing/location** of outbreaks/events
  - Determine **epidemiological parameters**
  - Test for episodic **selection**