bioinformatics.ca

Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io

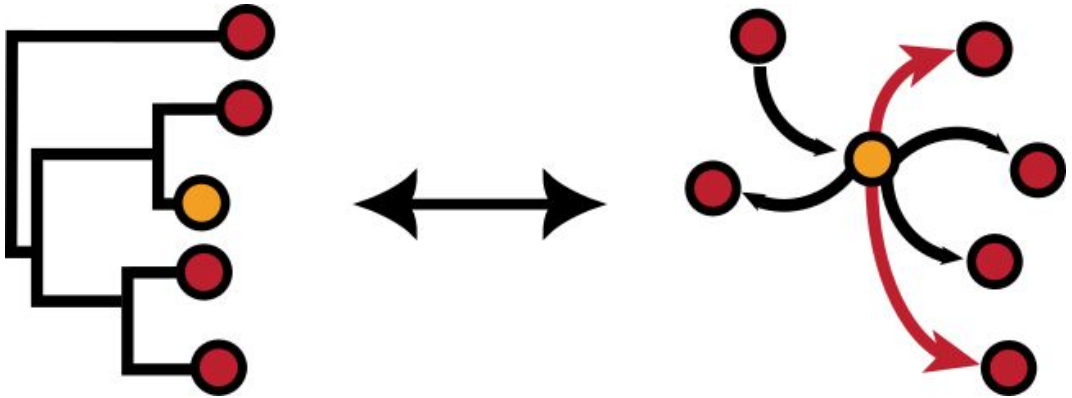
Module 7: Phylodynamics

Finlay Maguire

Genomic Epidemiology of Infectious Disease

April, 18-21, 2023

bioinformatics.ca



Finlay Maguire (finlay.maguire@dal.ca)

Computer Science and Community Health & Epidemiology, Dalhousie University, Halifax

Shared Hospital Laboratory, Toronto

Sunnybrook Research Institute, Toronto

Institute of Comparative Genomics, Dalhousie University, Halifax

Overview

Module 7: Phylodynamics (Finlay Maguire)

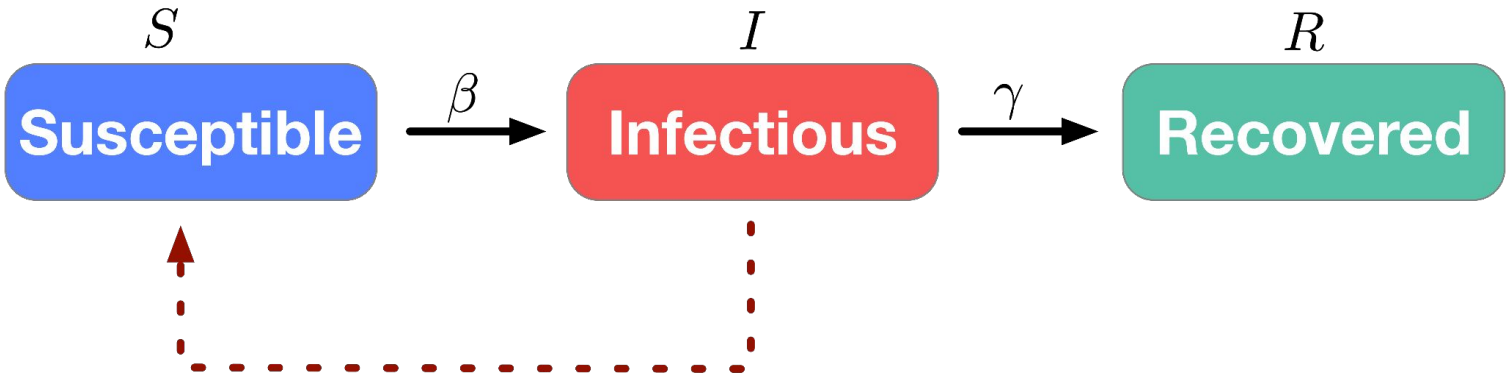
- Overview of phylodynamics
- Bayesian modelling
- Temporal inference
- Spatial/trait inference
- Epidemiological parameter estimation
- Inference of selection

Lab Practical:

- Likelihood-based phylodynamic analyses of SARS-CoV-2 genomes
    a. Time estimation
    b. Ancestral state reconstruction
    c. Testing for selection

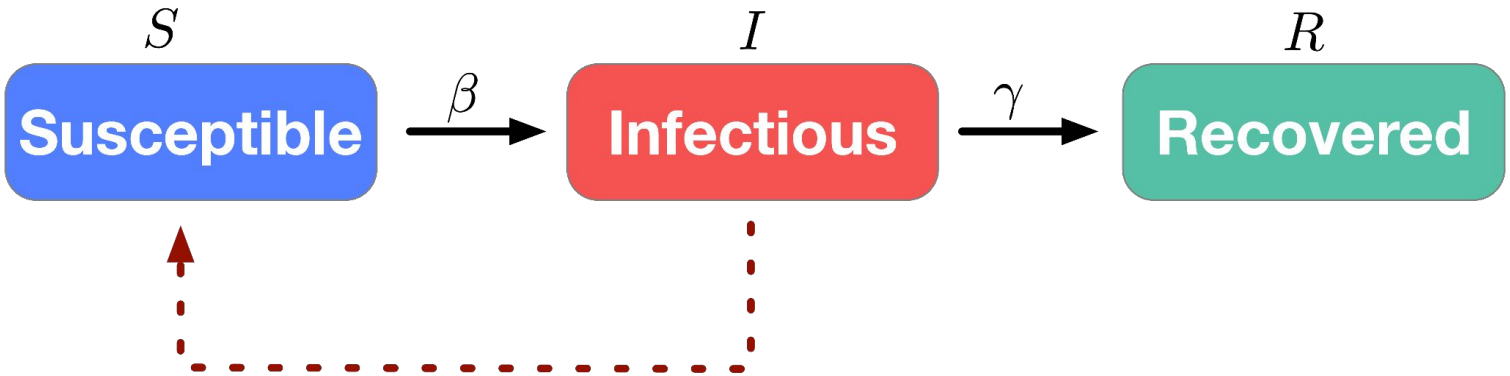
So, you want to understand the epidemiology of infectious diseases?

Compartmental models are used to model infections



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

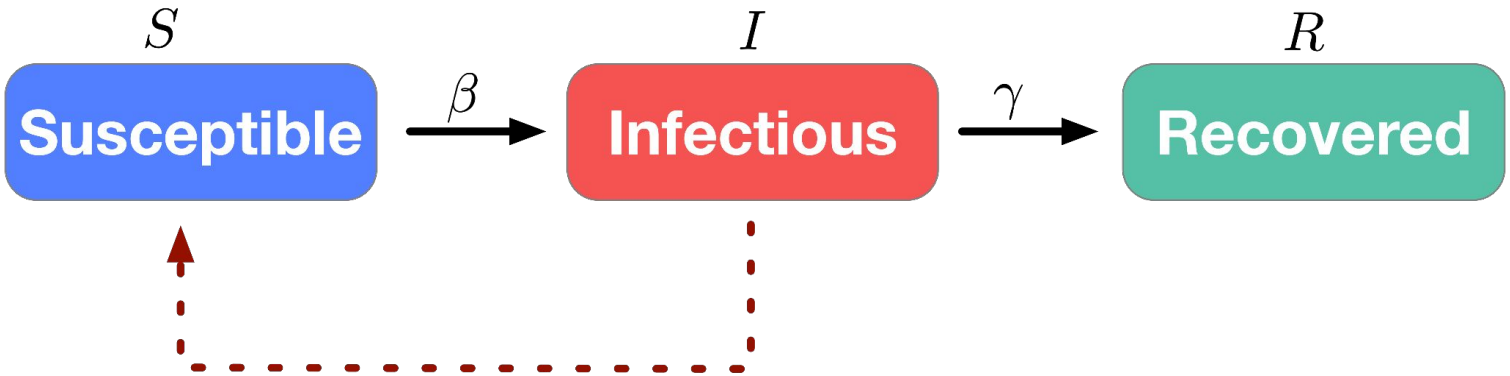
Compartmental models are used to model infections



Disclaimer: many more complex models!

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

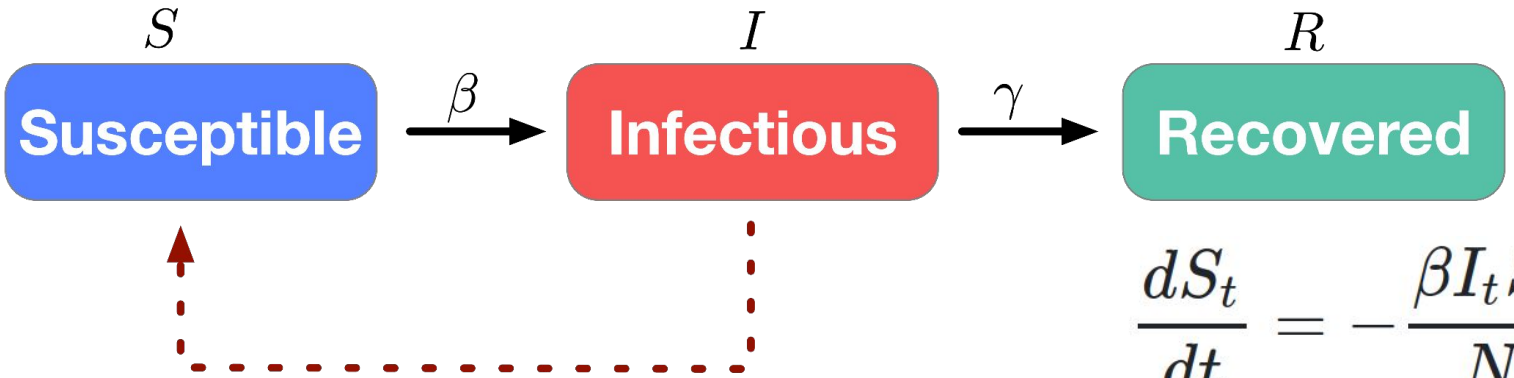
Compartmental models are used to model infections



- 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



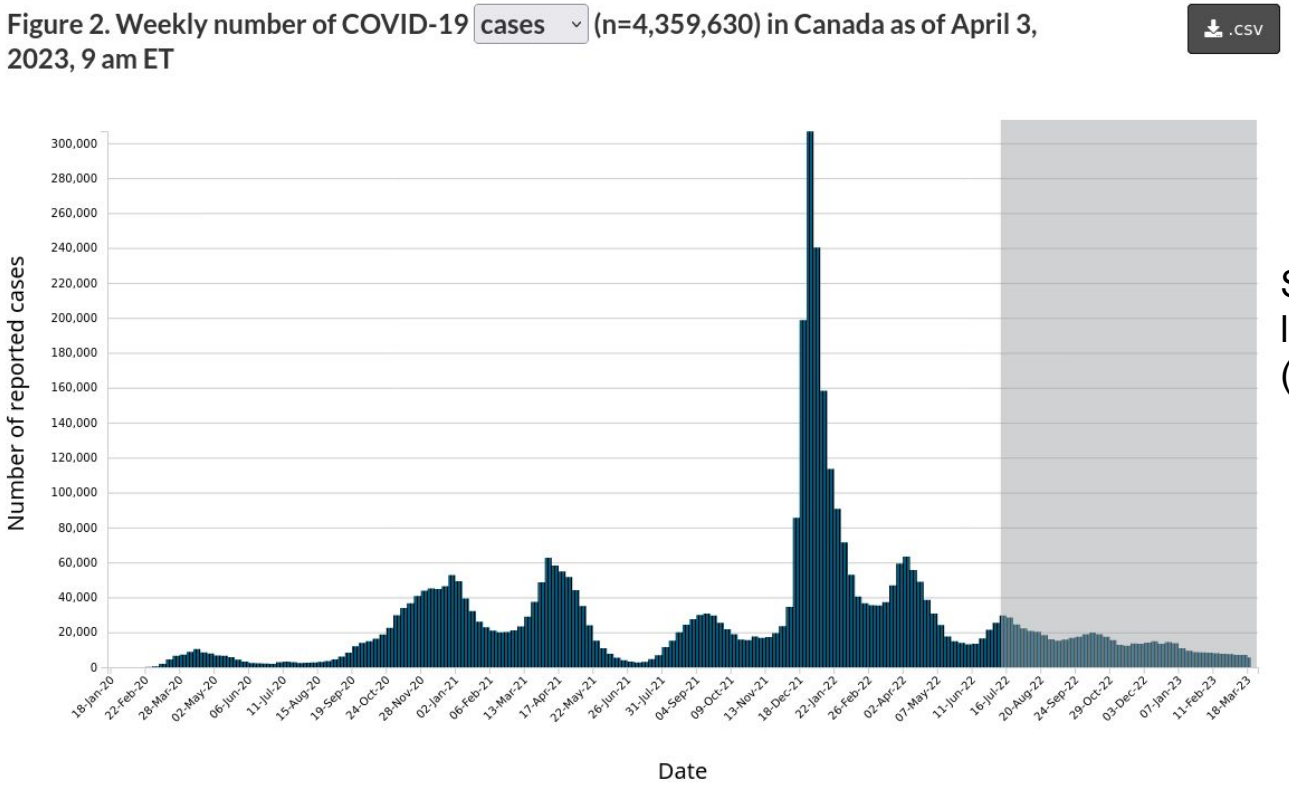
- S_t: the number of susceptible individuals
- I_t: the number of infectious individuals
- R_t: the number of recovered/deceased/immune individuals

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

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

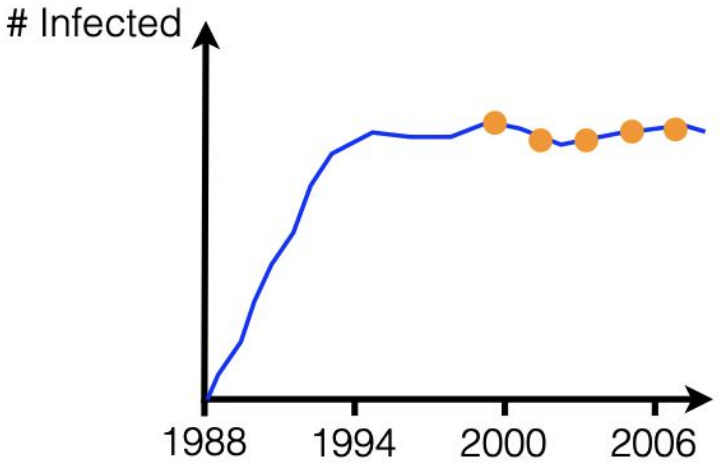
So, why do we need genomic data?

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?**

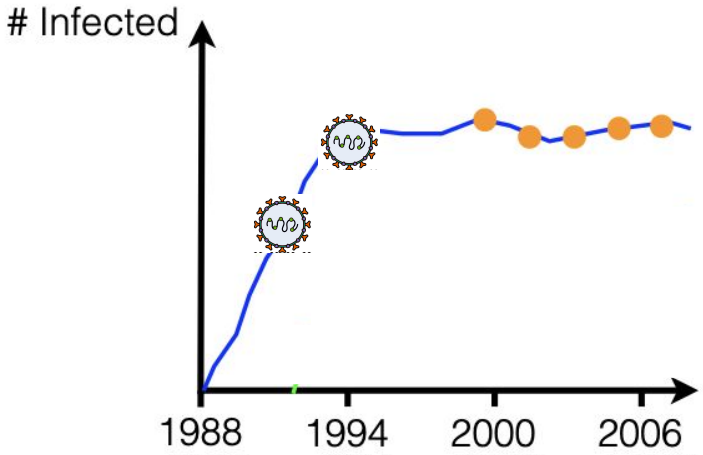


https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day2_Phylodynamics_-_Tanja_Stadler.pdf

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

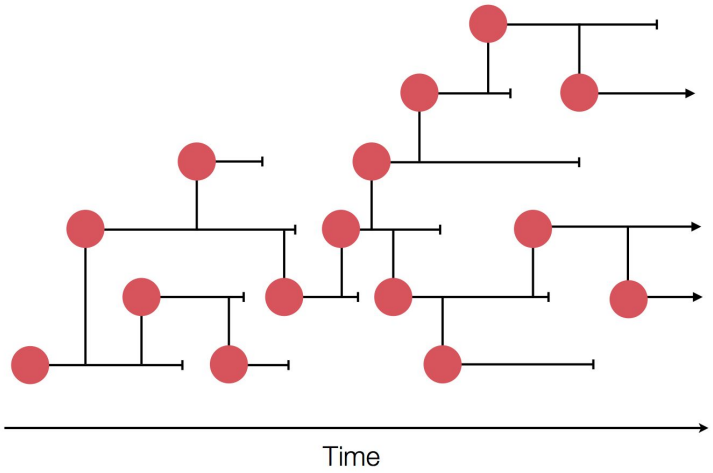
Genomics can be used to infer unobserved events

If sampling in early epidemic was missed:

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

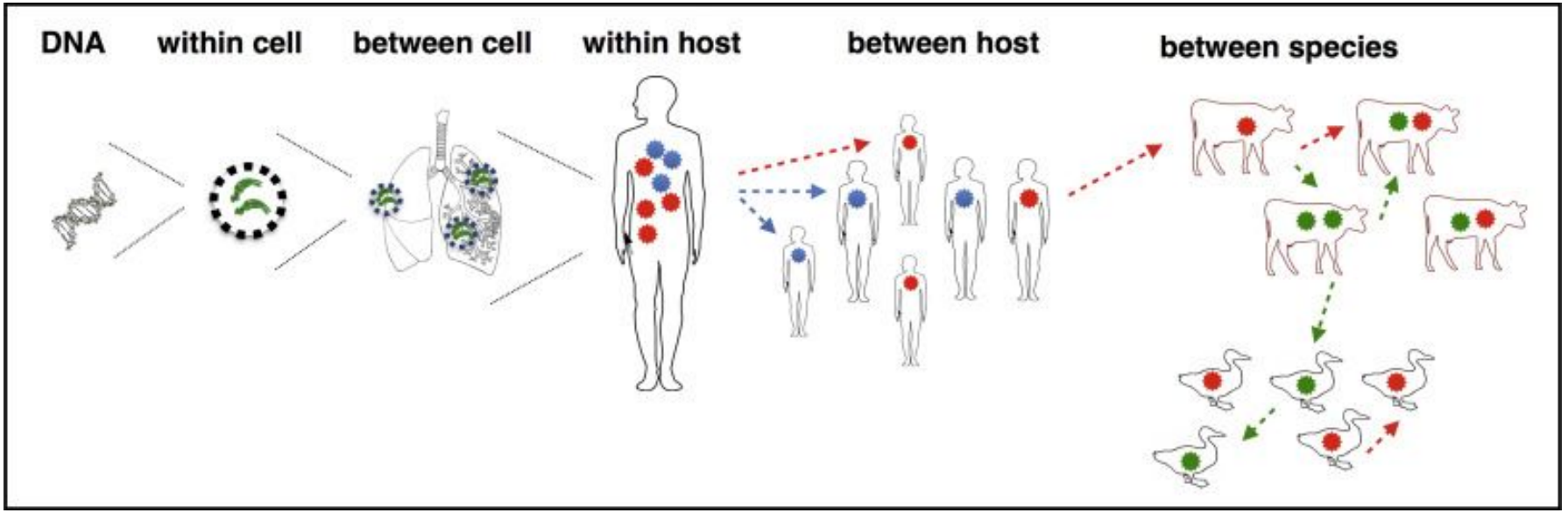
Data does not tell who infected whom:

- **Population structure?**



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

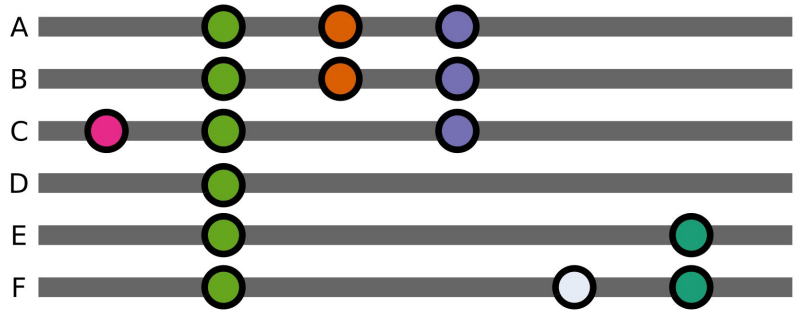
Cases don’t tell you (much) about pathogen evolution



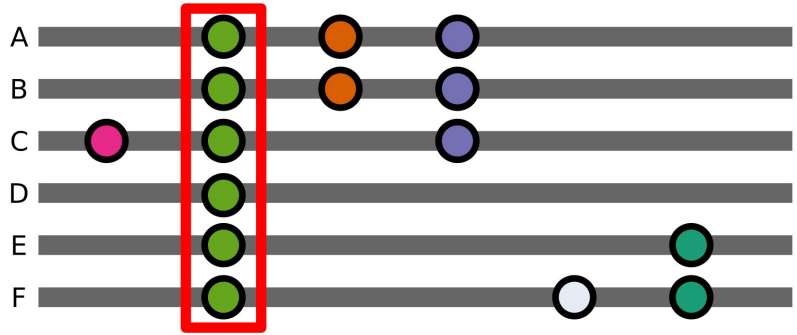
https://www.sciencedirect.com/science/article/pii/S1755436514000723

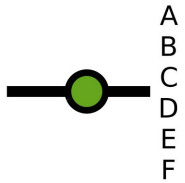
How do we actually link genomes and epidemiology?

Can infer a phylogeny from genomic data

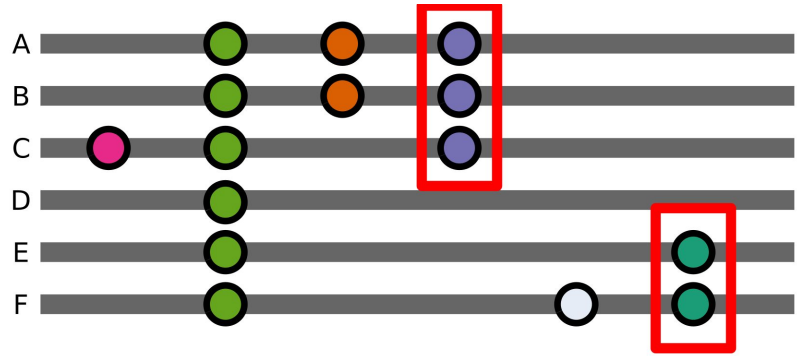


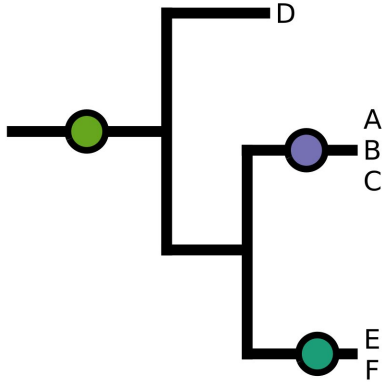
Can infer a phylogeny from genomic data



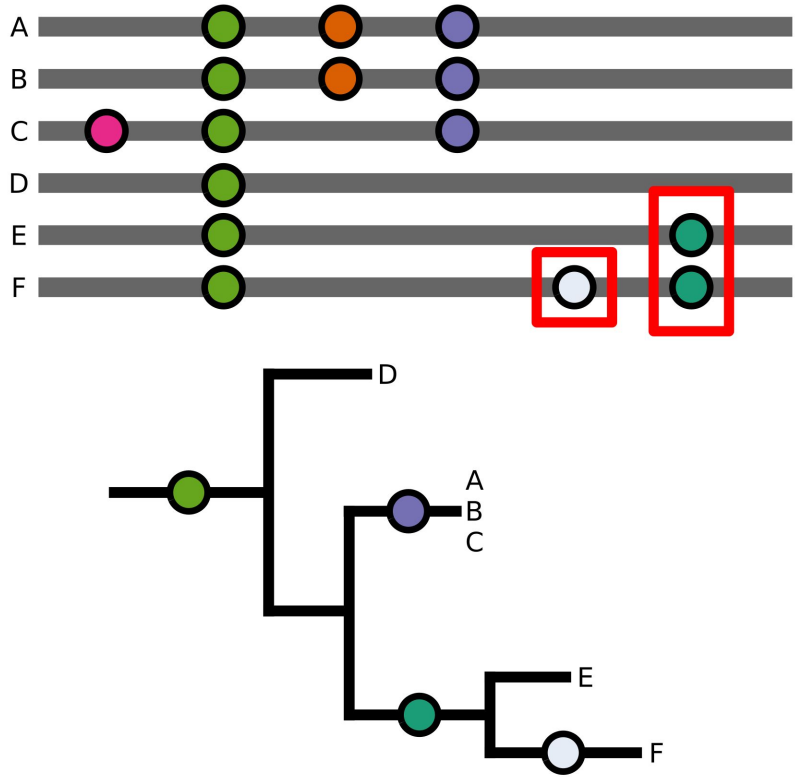


Can infer a phylogeny from genomic data

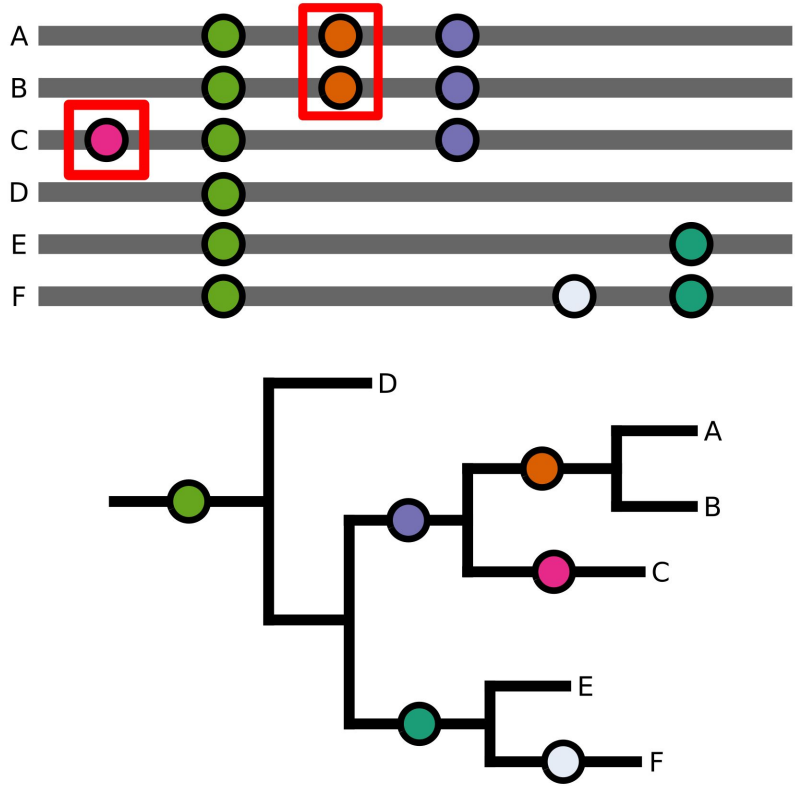




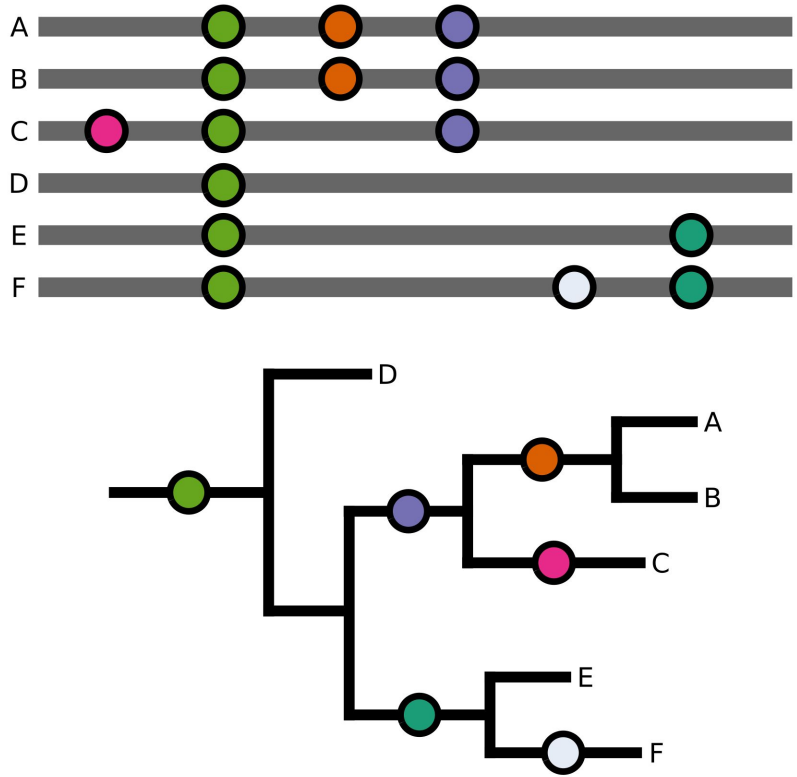
Can infer a phylogeny from genomic data



Can infer a phylogeny from genomic data

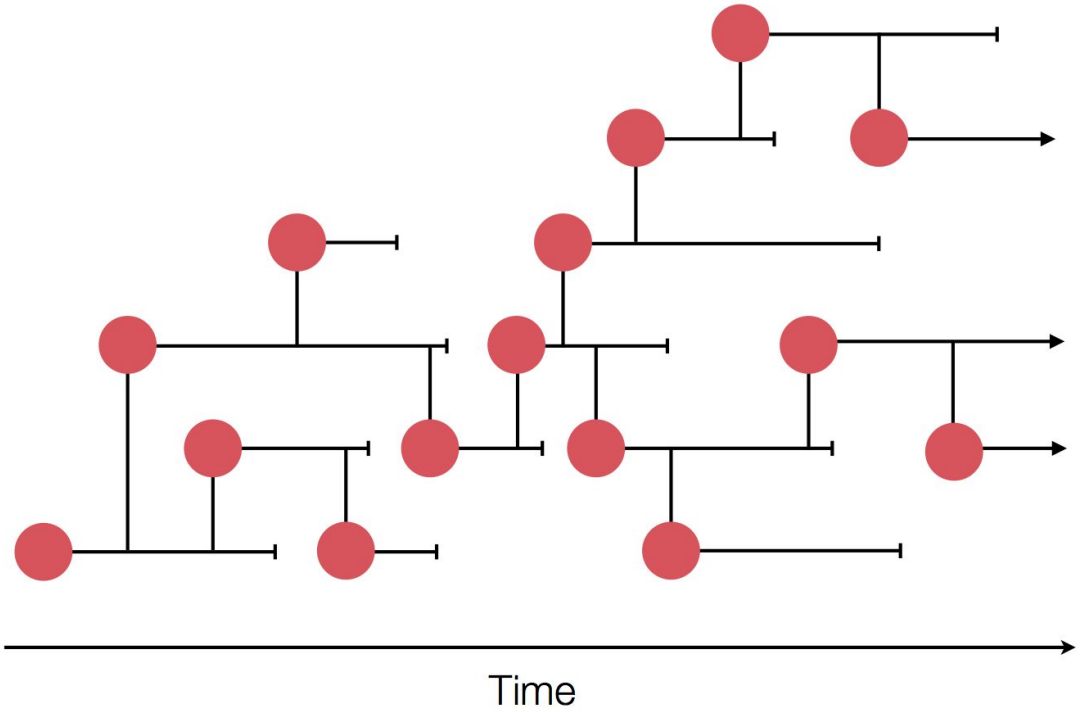


Can infer a phylogeny from genomic data



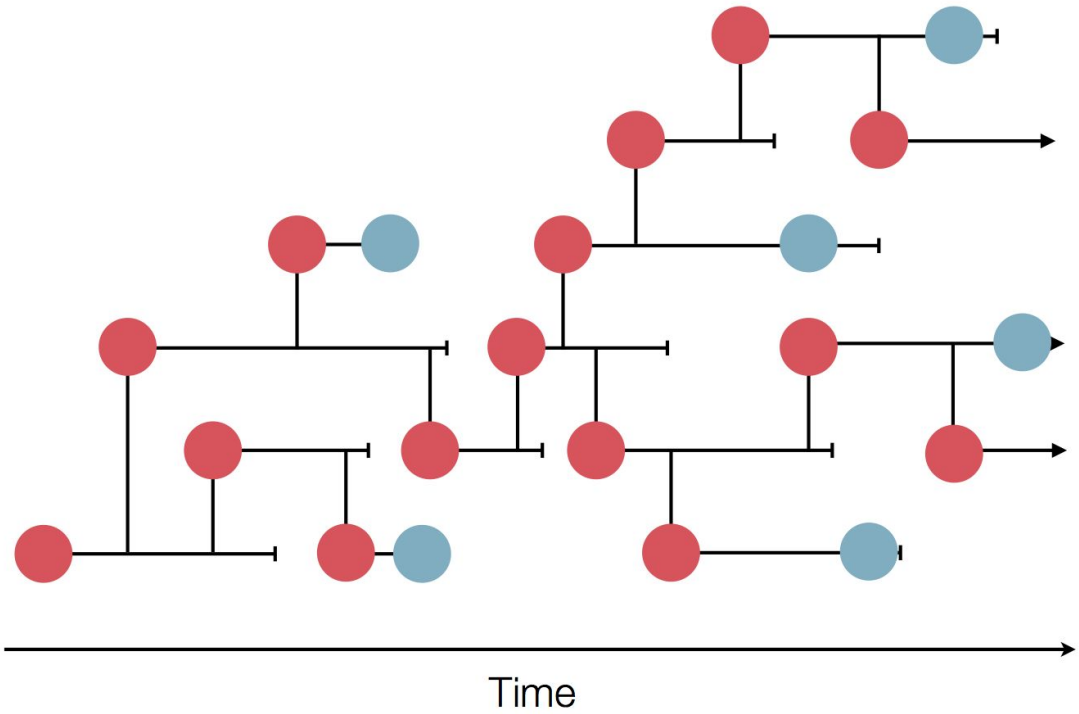
What does this tree actually represent?

Sampling from underlying process



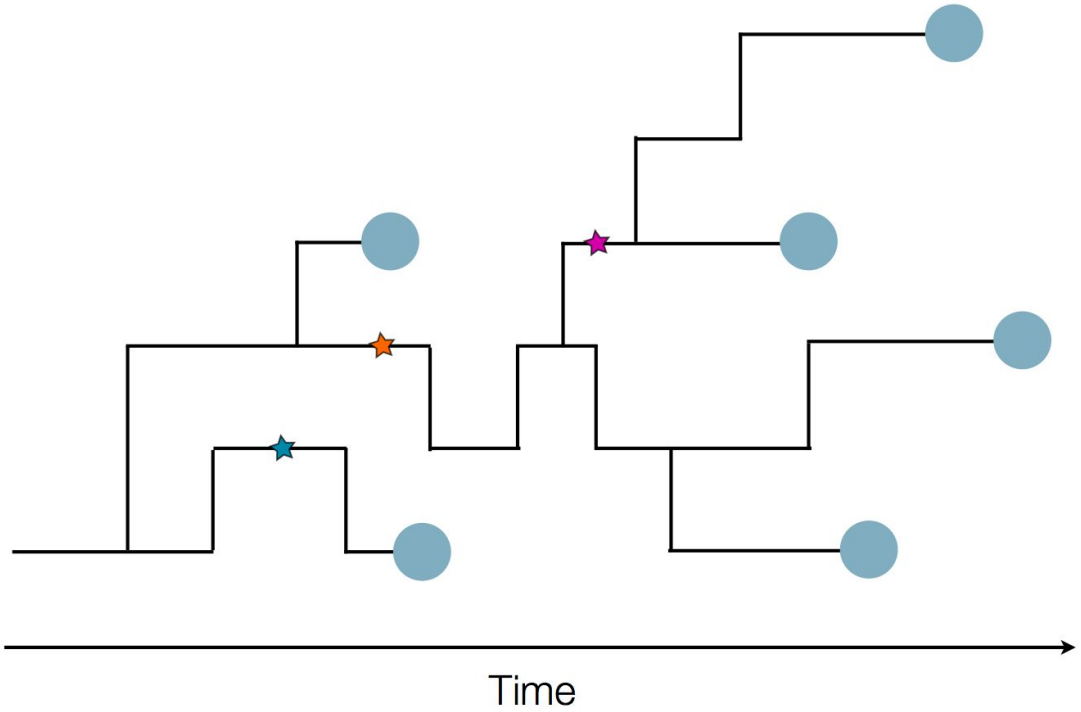
https://github.com/trvrb/phylodynamics-lecture

Sampling from underlying process



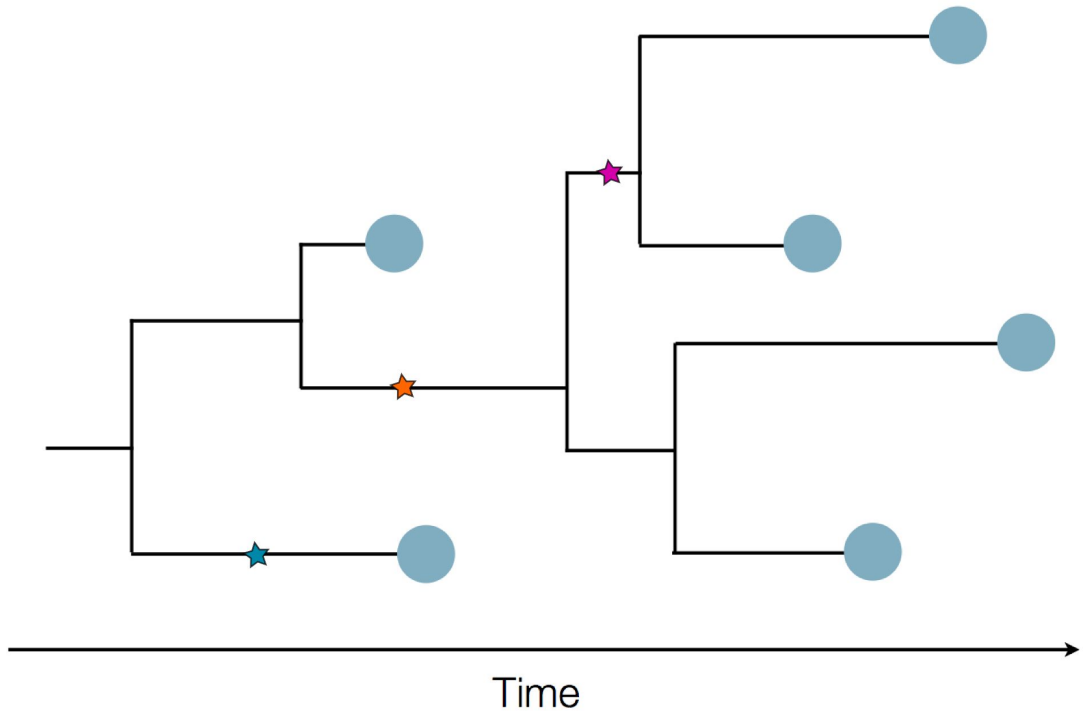
https://github.com/trvrb/phylodynamics-lecture

Sampling from underlying process



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

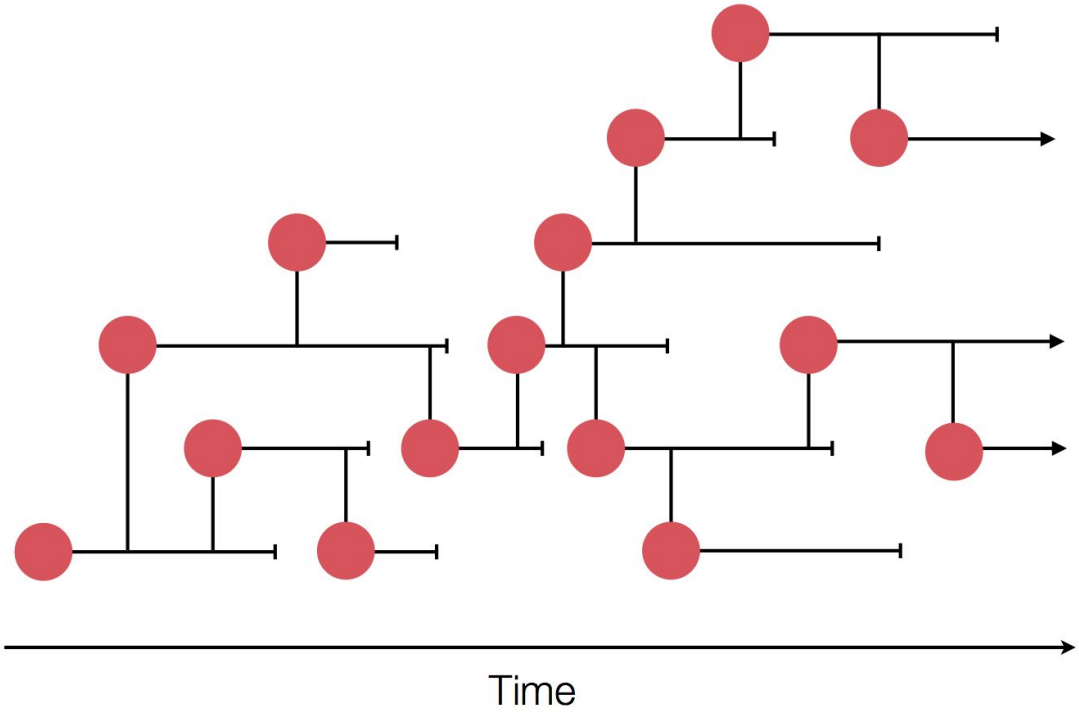
Sampling from underlying process



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

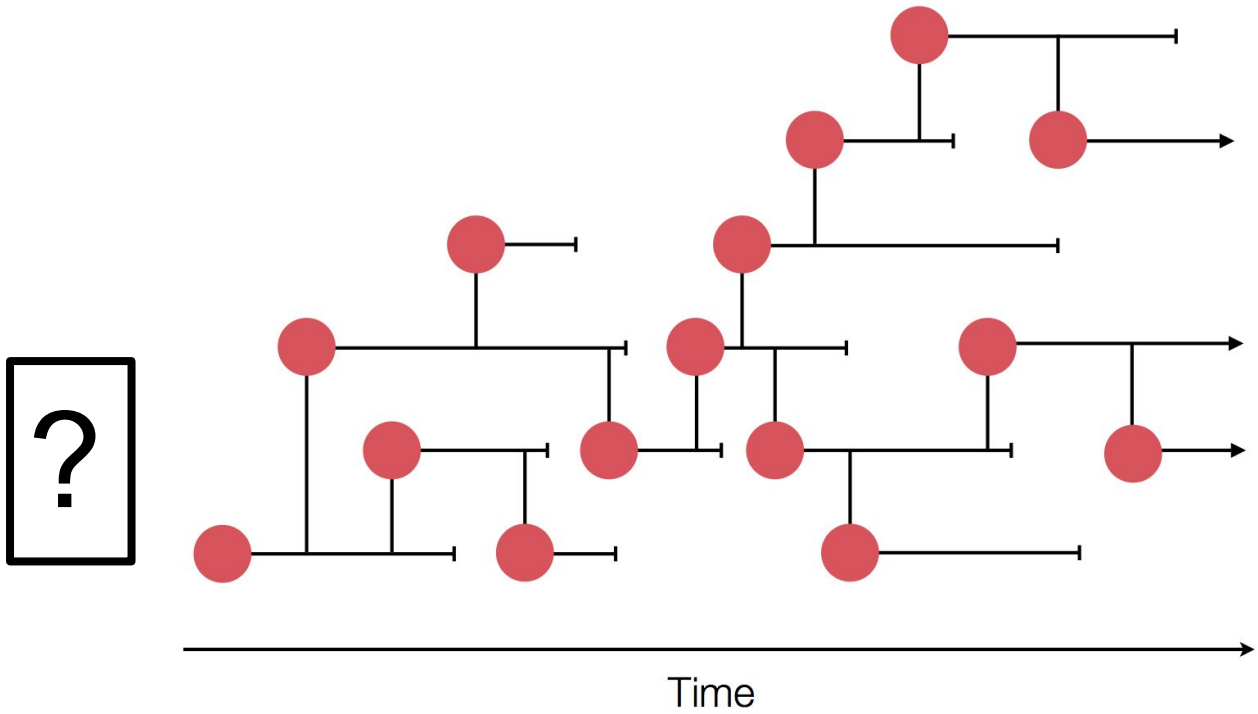
What determines underlying process?

Many forces shaping underlying process



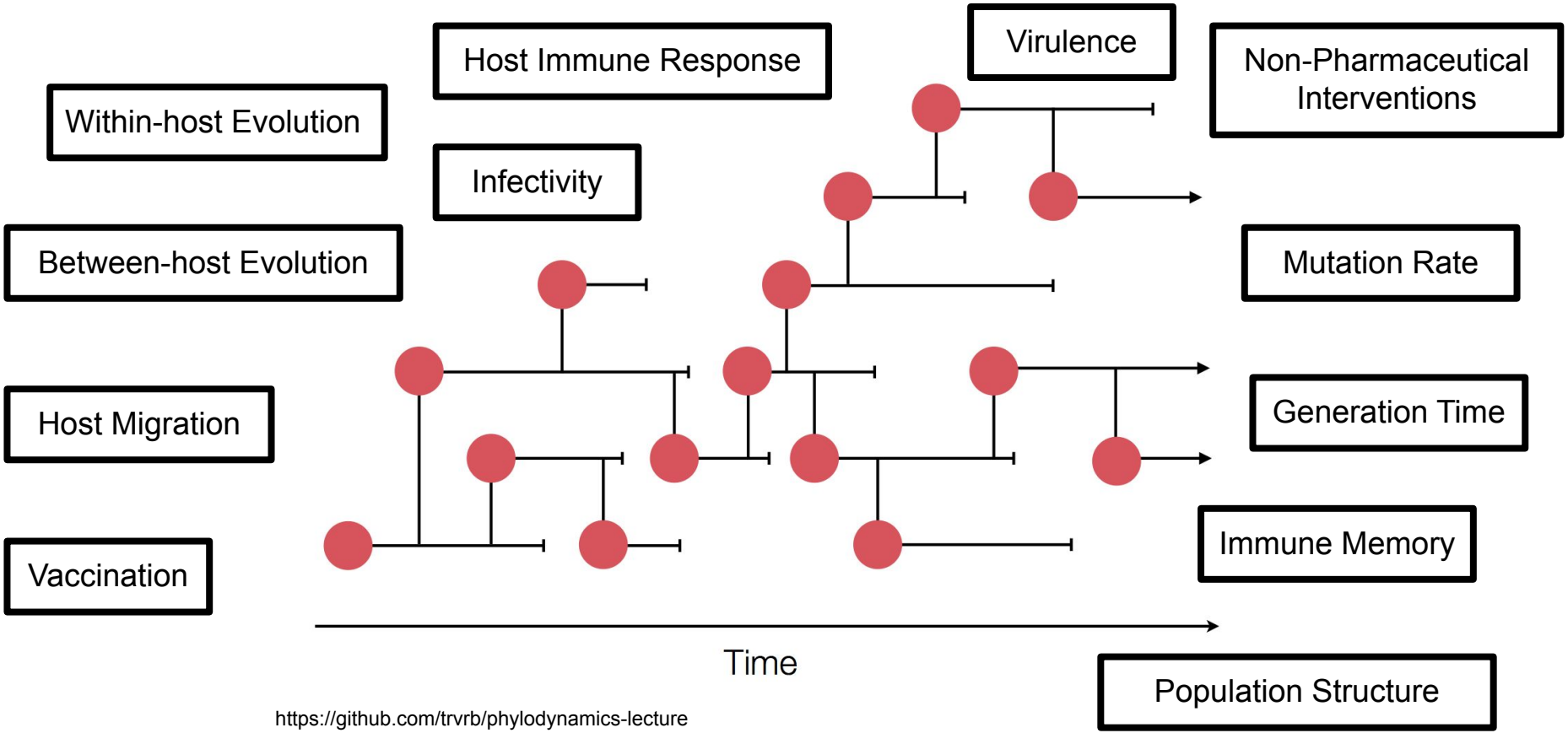
https://github.com/trvrb/phylodynamics-lecture

Many forces shaping underlying process



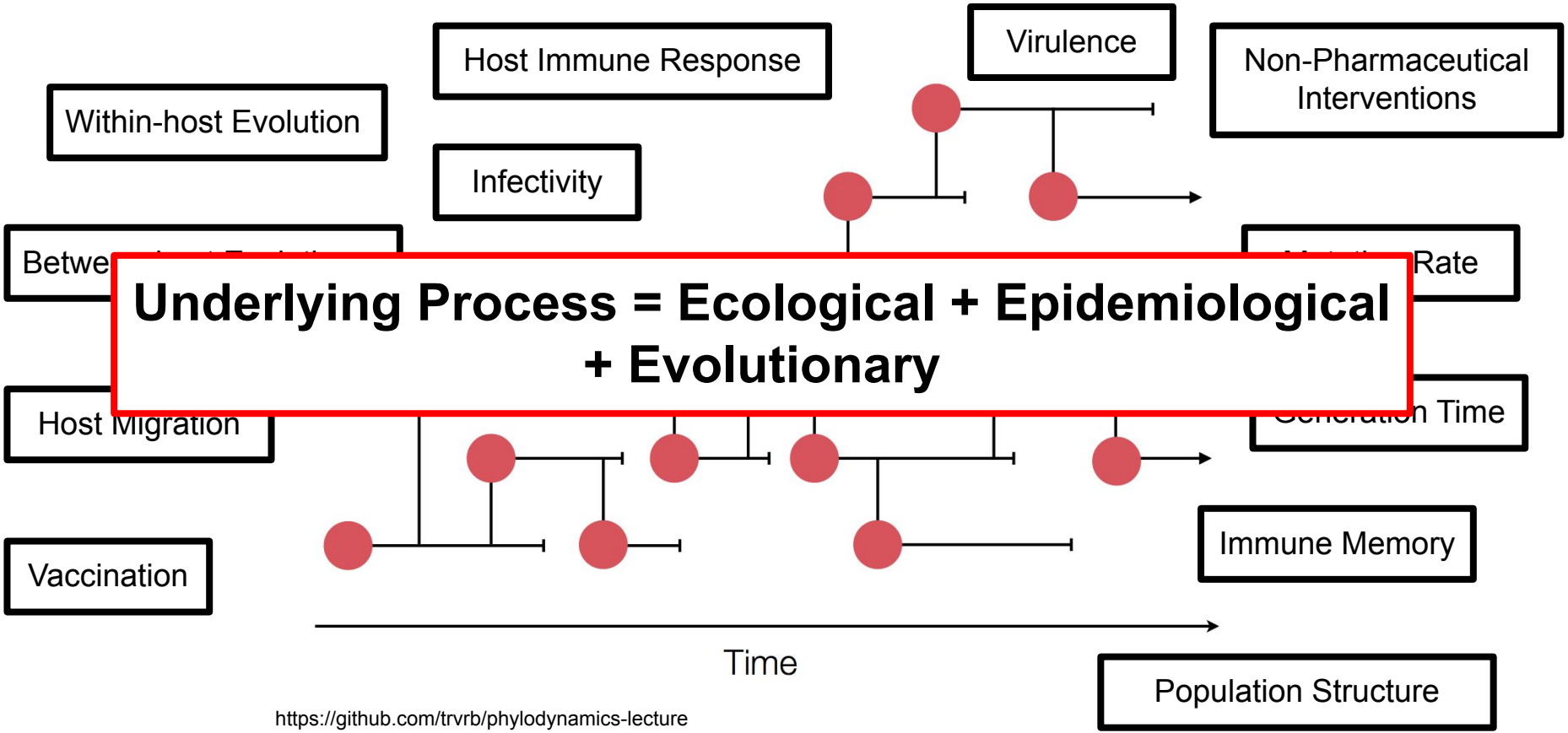
https://github.com/trvrb/phylodynamics-lecture

Many forces shaping underlying process



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

Many forces shaping underlying process

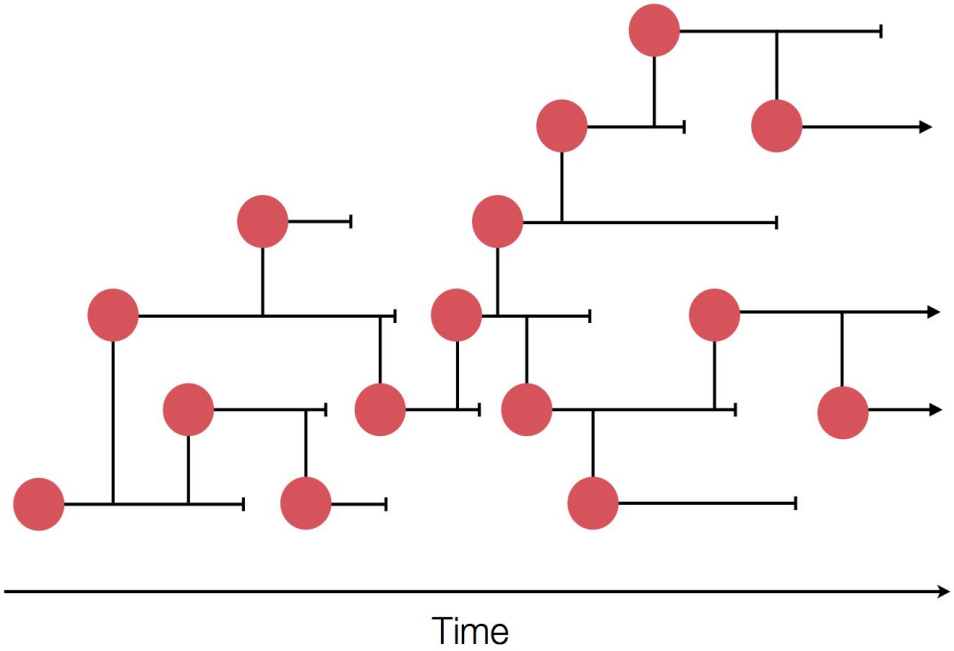


https://github.com/trvrb/phylodynamics-lecture

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

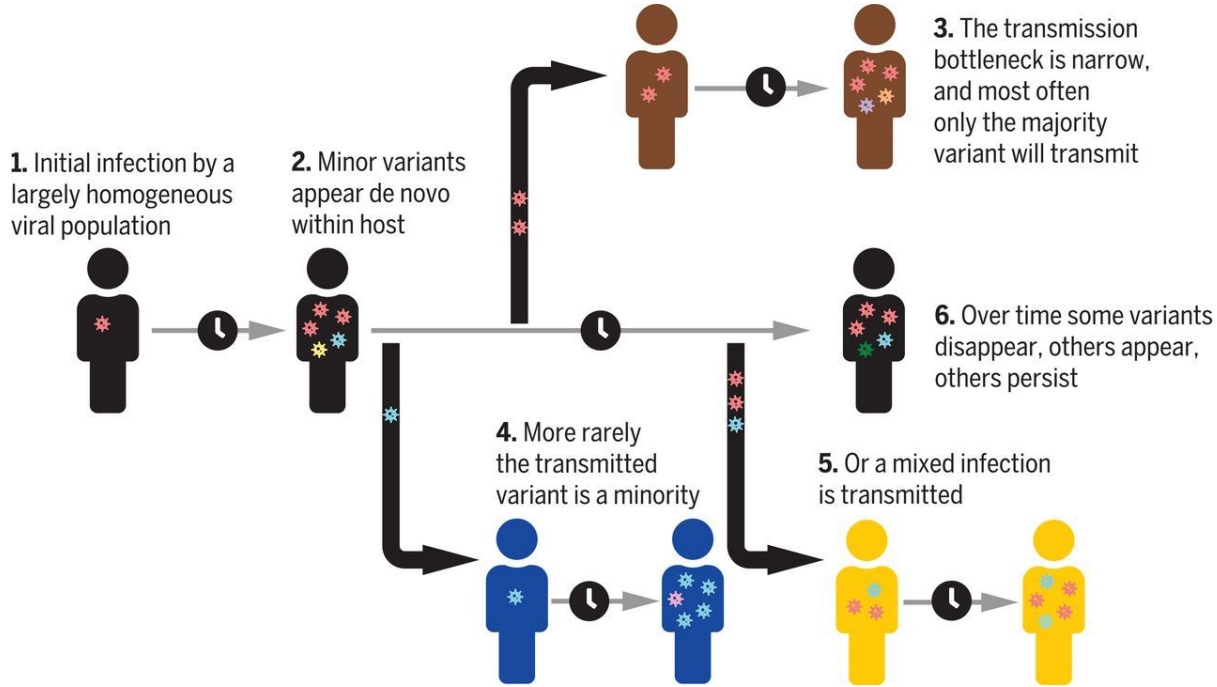
Let’s start with the “simple” reconstruction of the transmission network

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



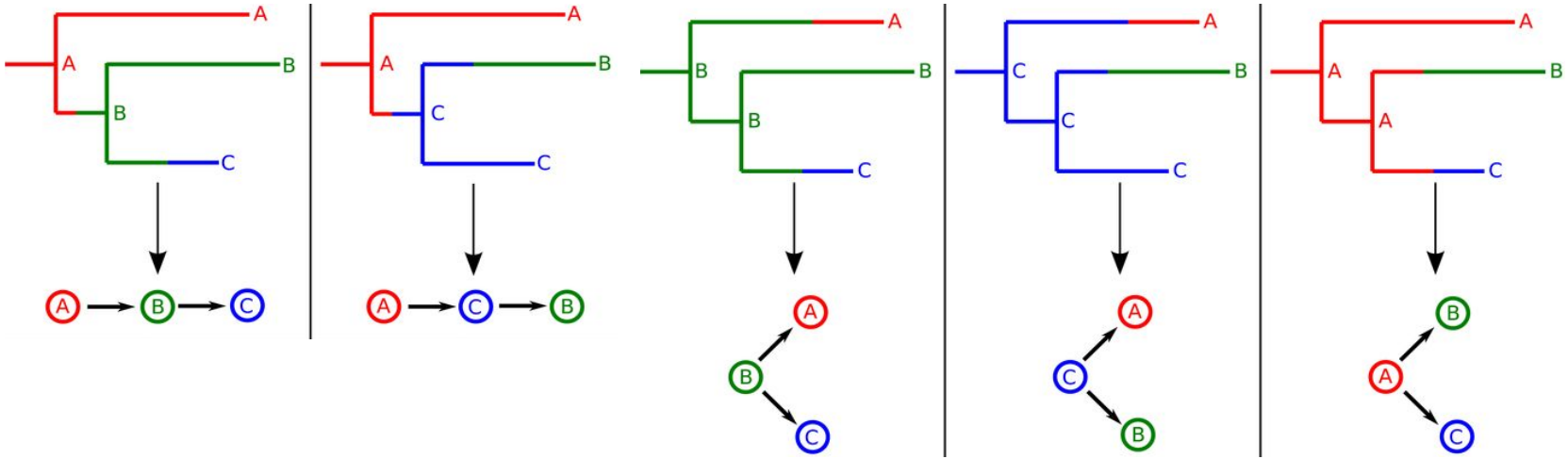
https://github.com/trvrb/phylodynamics-lecture

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



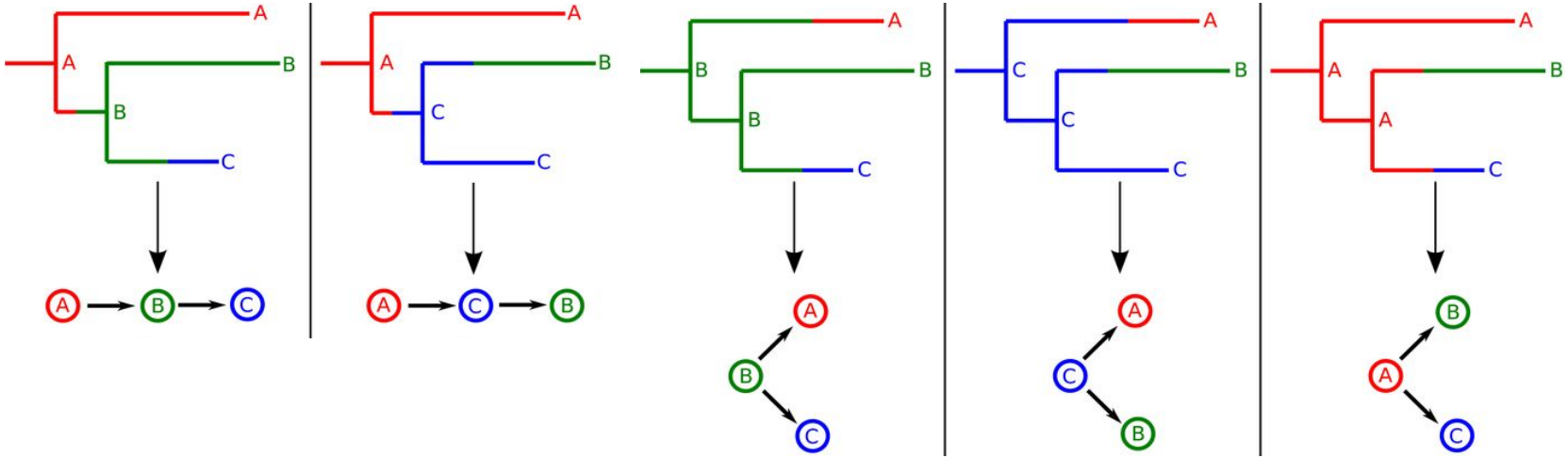
https://www.science.org/doi/10.1126/science.abg0821

Same tree can be consistent with different scenarios



10.1371/journal.pcbi.1004613

Same tree can be consistent with different scenarios



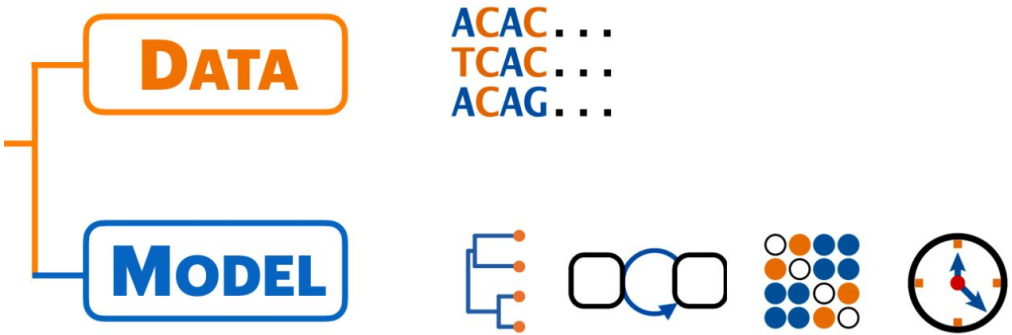
10.1371/journal.pcbi.1004613

=> Probabilistic inference!

Bayesian inference is a key tool in phylodynamics

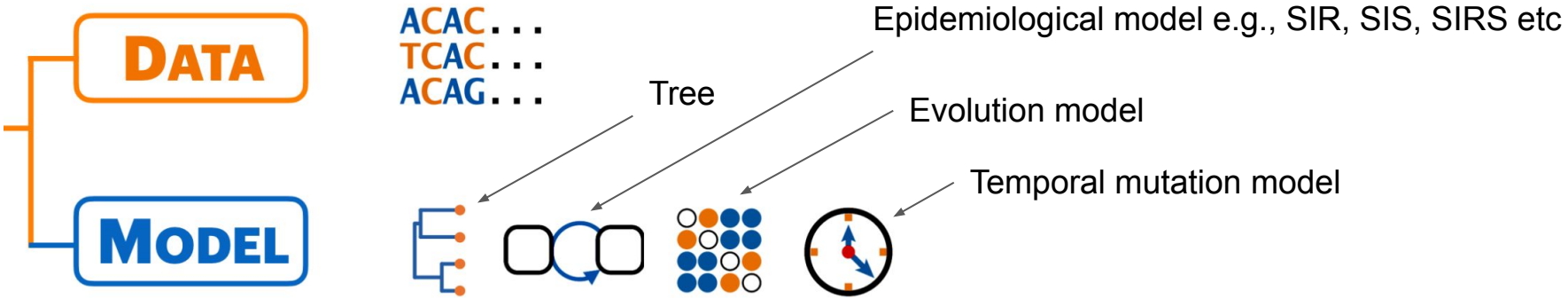
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Bayesian inference is a key tool in phylodynamics



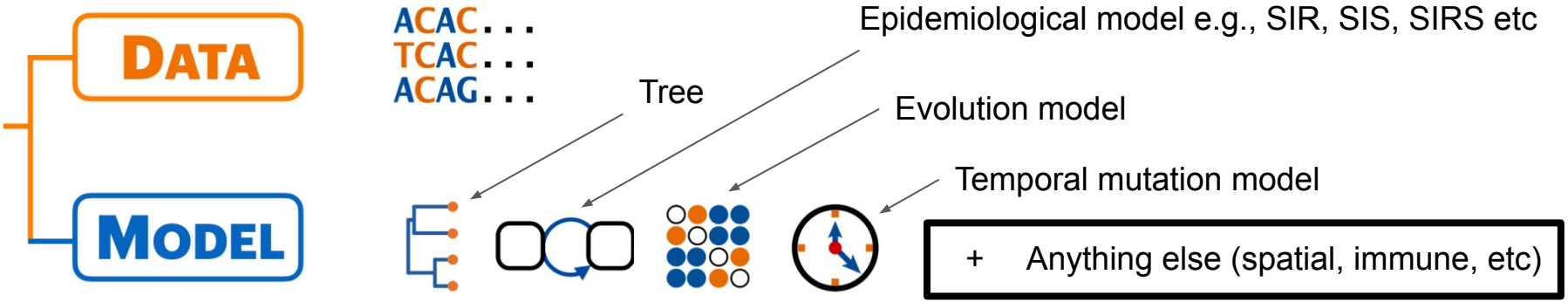
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Bayesian inference is a key tool in phylodynamics



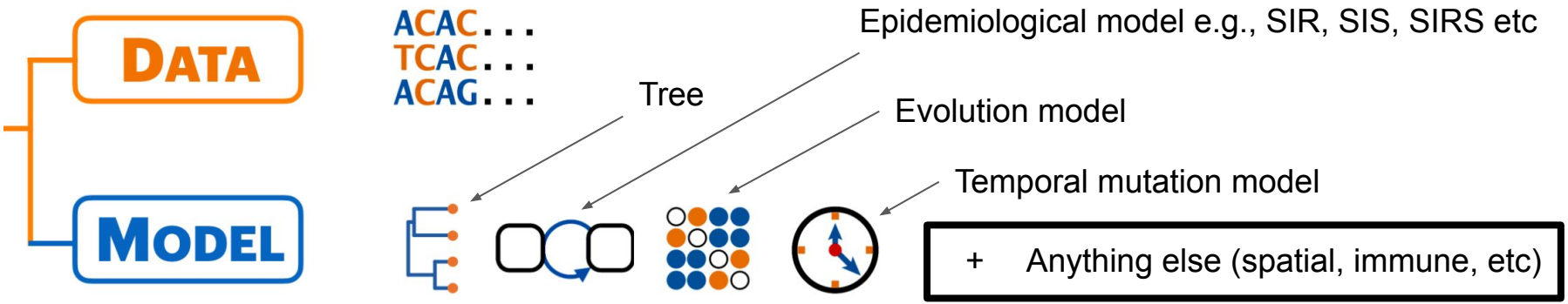
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Bayesian inference is a key tool in phylodynamics



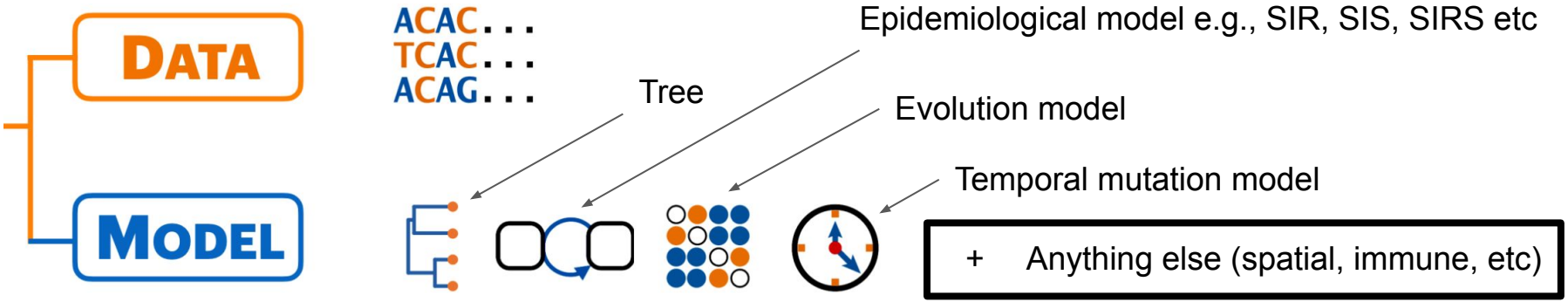
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Bayesian inference is a key tool in phylodynamics



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

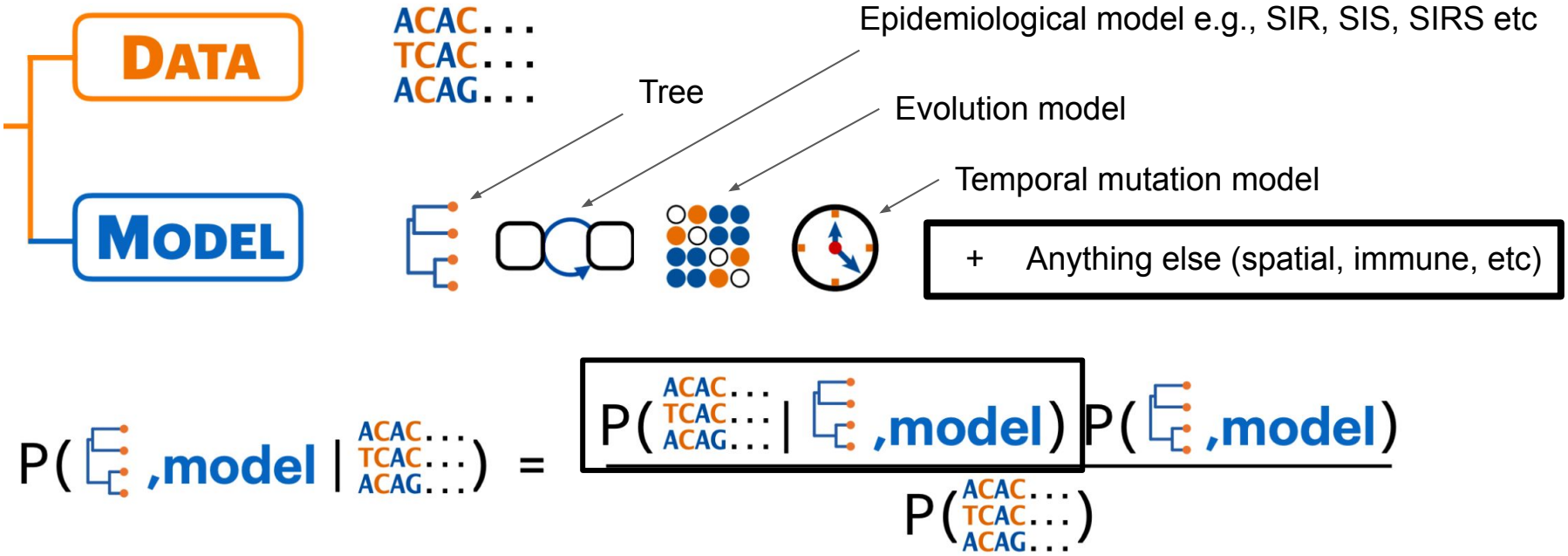
Bayesian inference is a key tool in phylodynamics



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

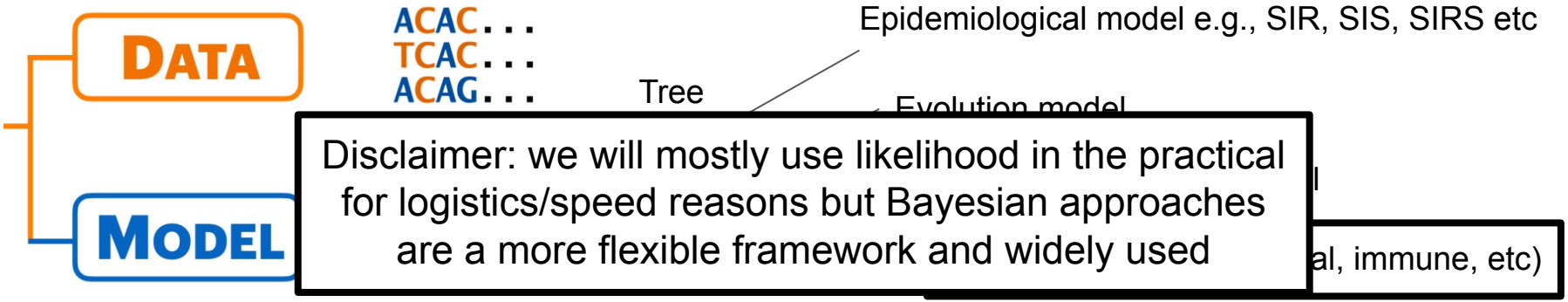
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Bayesian inference is a key tool in phylodynamics



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

Bayesian inference is a key tool in phylodynamics

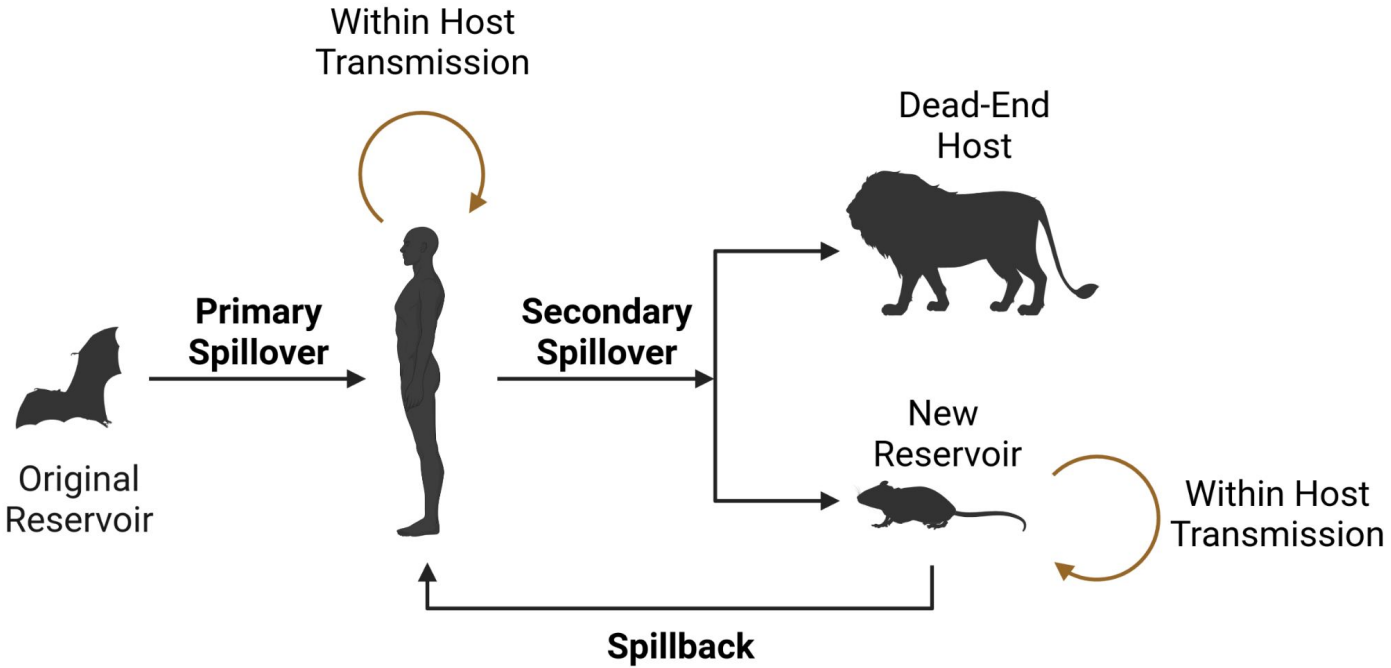


$$P(\text{tree}, \text{model} \mid \text{ACAC...}, \text{TCAC...}, \text{ACAG...}) = \frac{P(\text{ACAC...}, \text{TCAC...}, \text{ACAG...} \mid \text{tree}, \text{model})\, P(\text{tree}, \text{model})}{P(\text{ACAC...}, \text{TCAC...}, \text{ACAG...})}$$

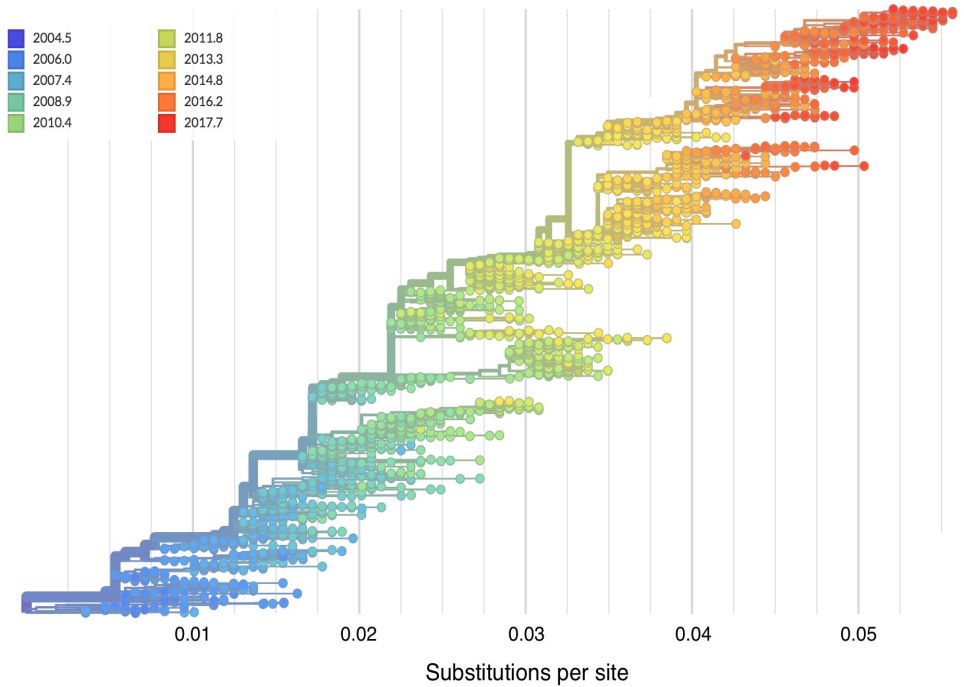
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day4_Structured_Phylodynamics_-_Denise_Kuehnert.pdf

Now we’ve got the general concept let’s look at some specific analyses

Knowing when zoonoses happen is key to reducing them

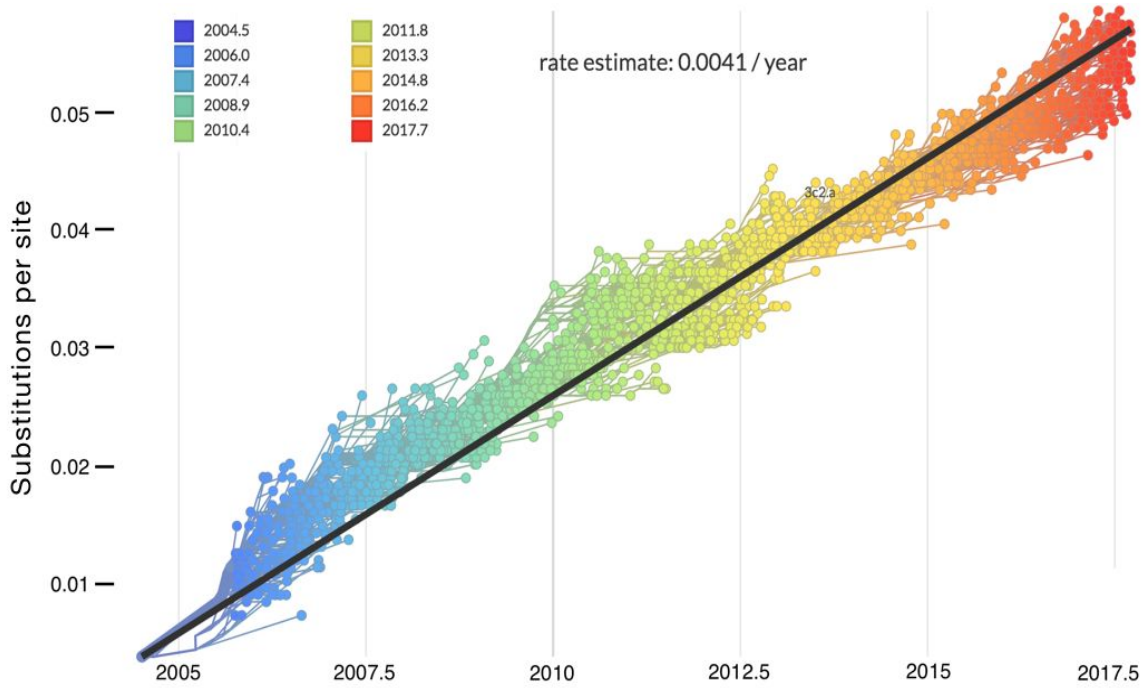


Need mutation rate to convert tree from distance to time



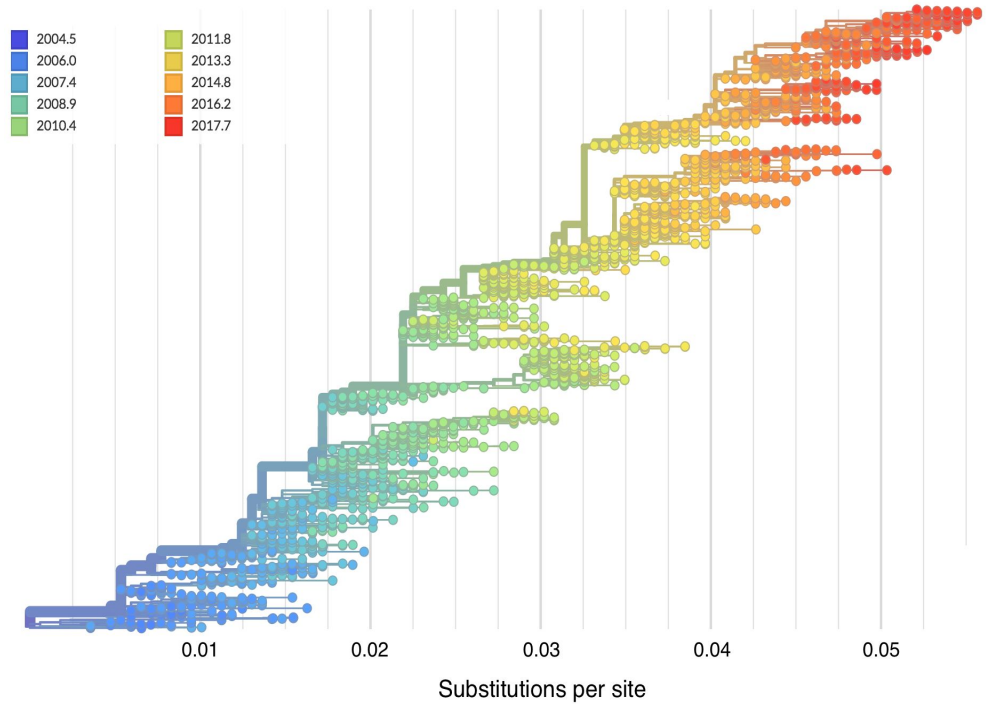
https://github.com/trvrb/gs541-phylodynamics

Root-to-Tip Regression can estimate mutation rate



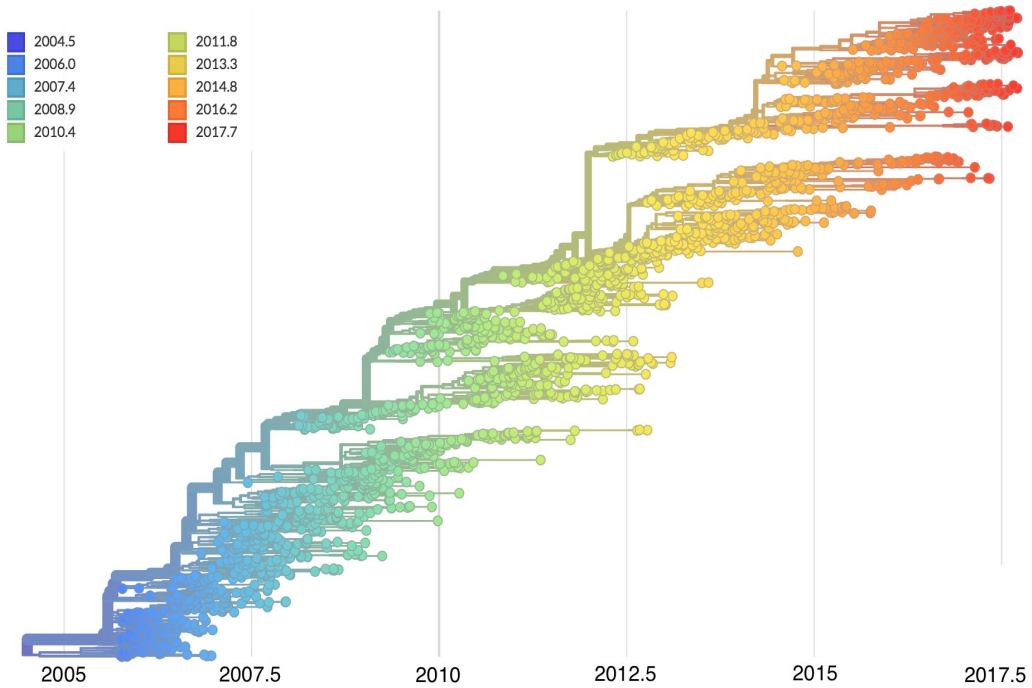
https://github.com/trvrb/gs541-phylodynamics

Estimate timing with fixed points, lengths & mutation rates



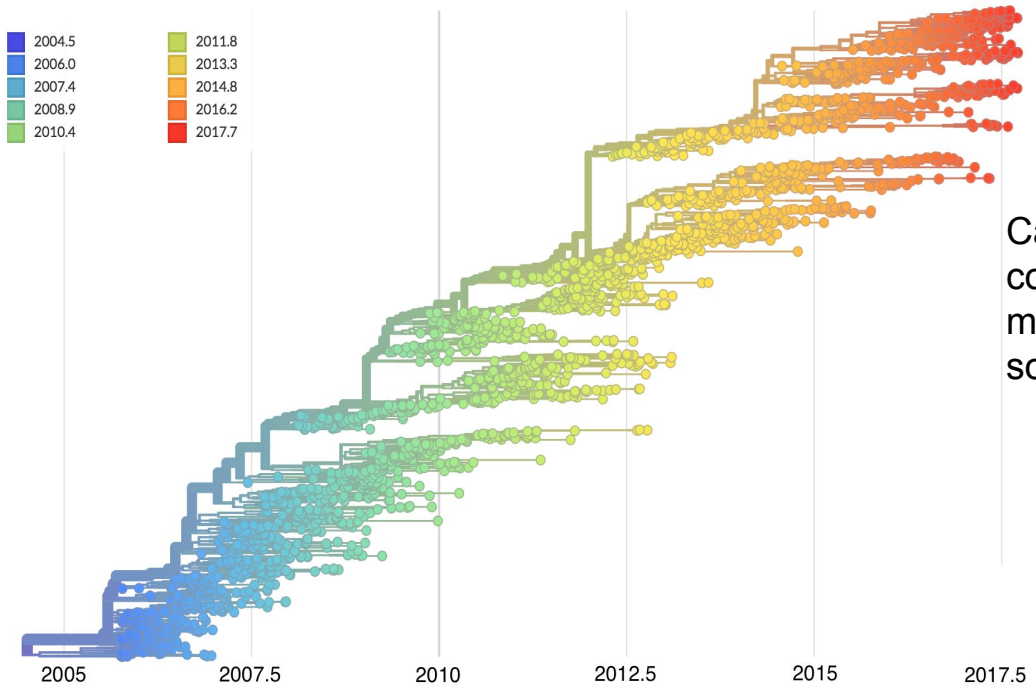
https://github.com/trvrb/gs541-phylodynamics

Estimate timing with fixed points, lengths & mutation rates



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

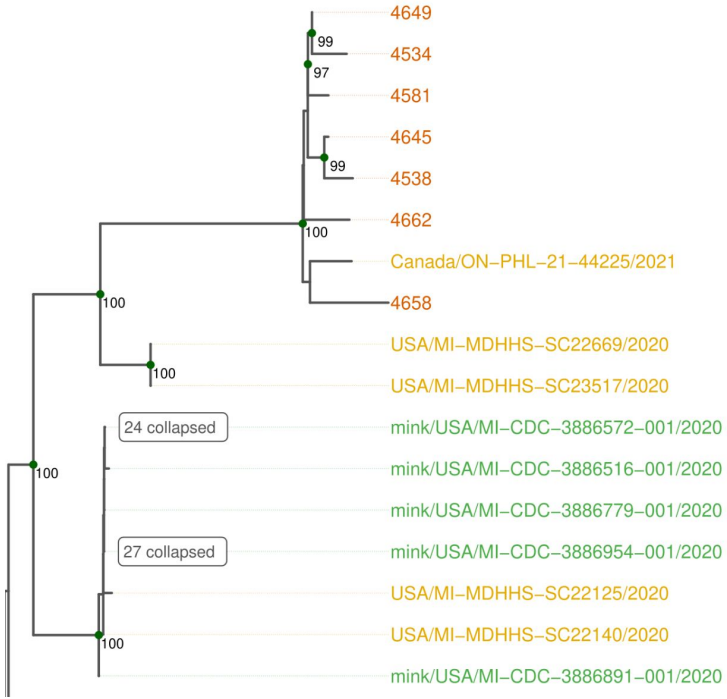
Estimate timing with fixed points, lengths & mutation rates

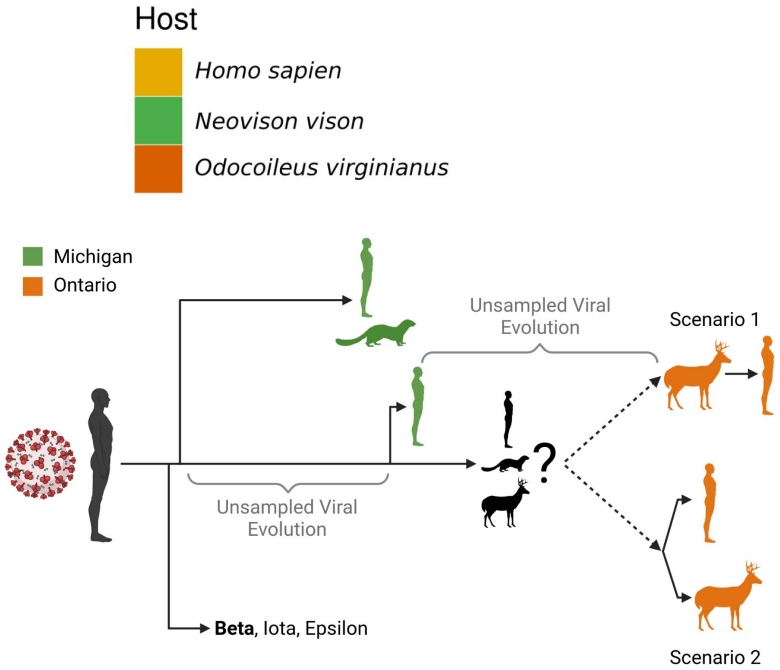


Can infer much more complex and varying mutation rates using more sophisticated models

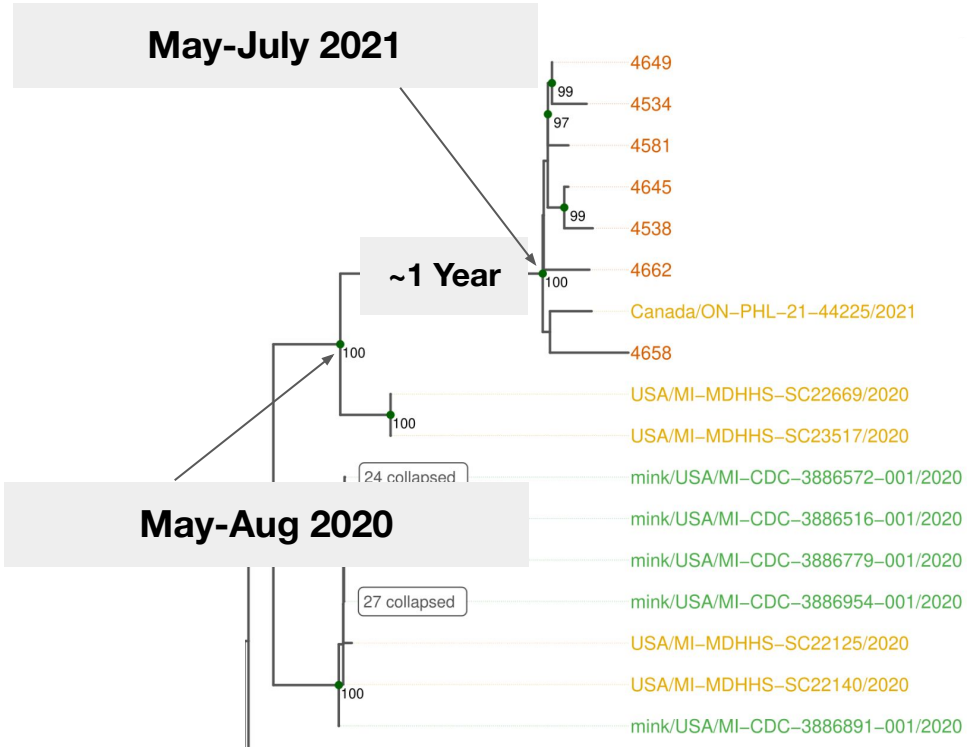
https://github.com/trvrb/gs541-phylodynamics

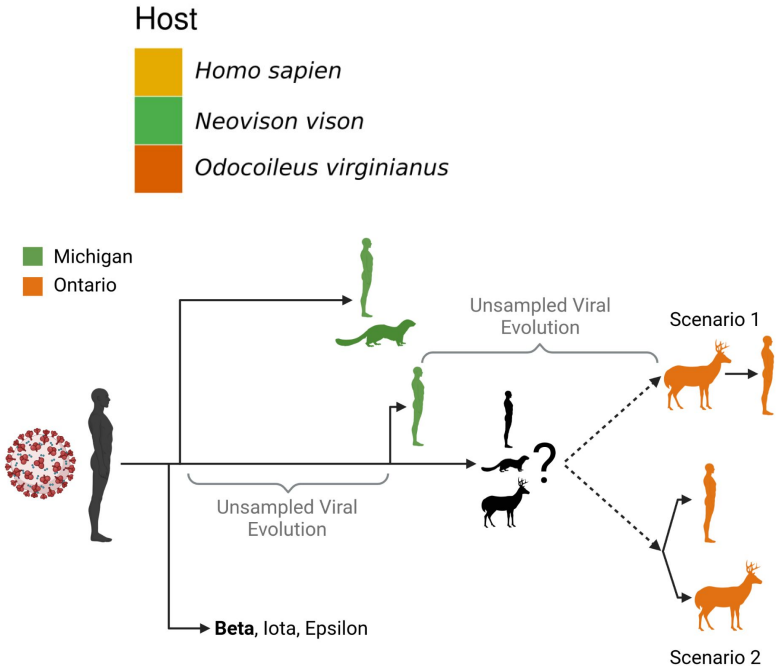
Time-trees let us estimate timing of unobserved events





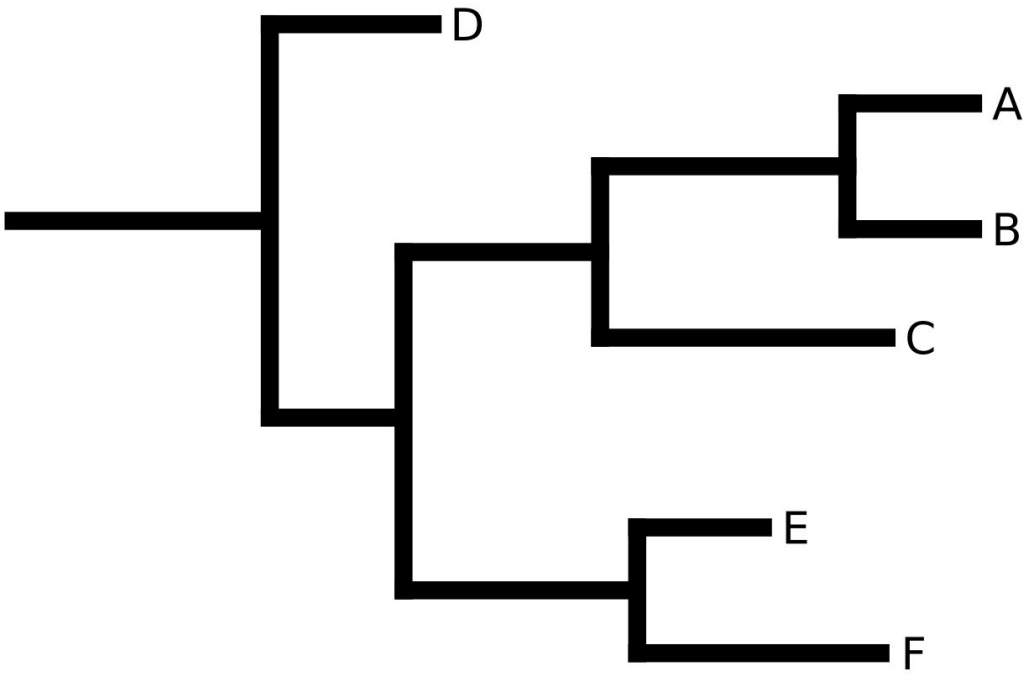
Time-trees let us estimate timing of unobserved events



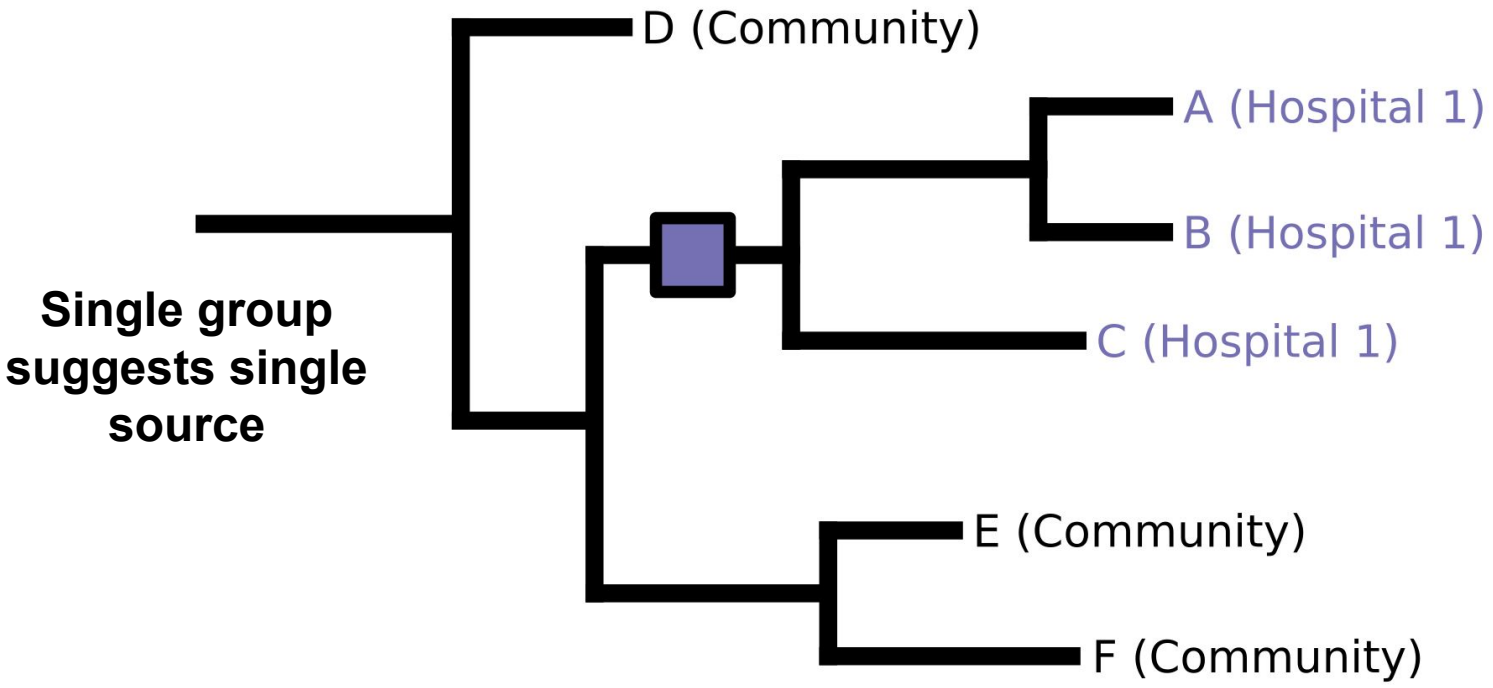


We also often want to know WHERE something happened

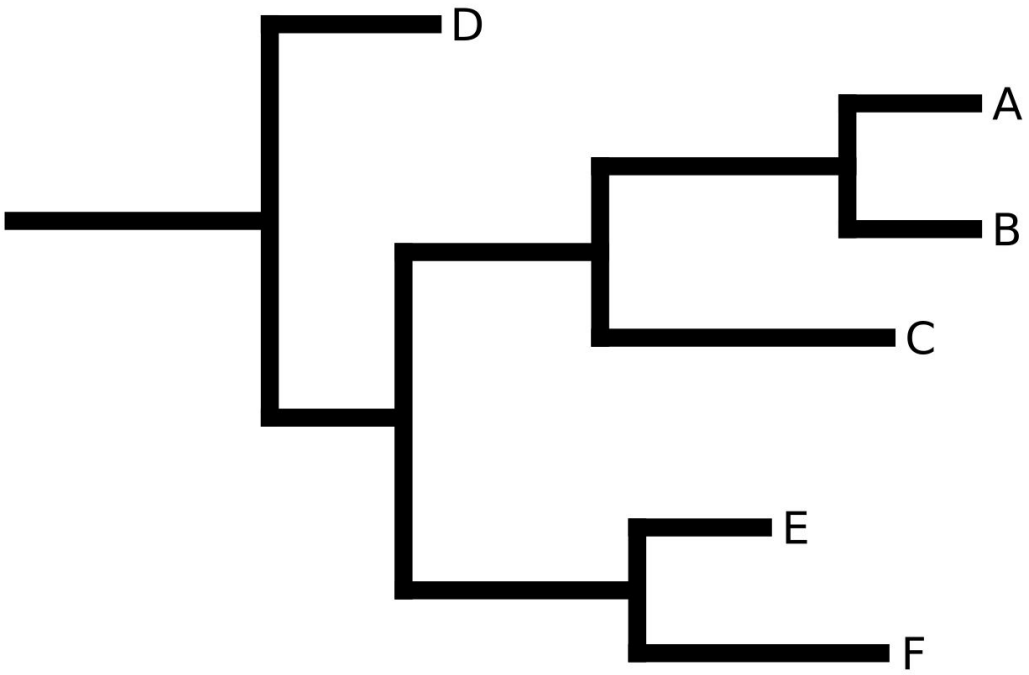
Trace sources of outbreaks



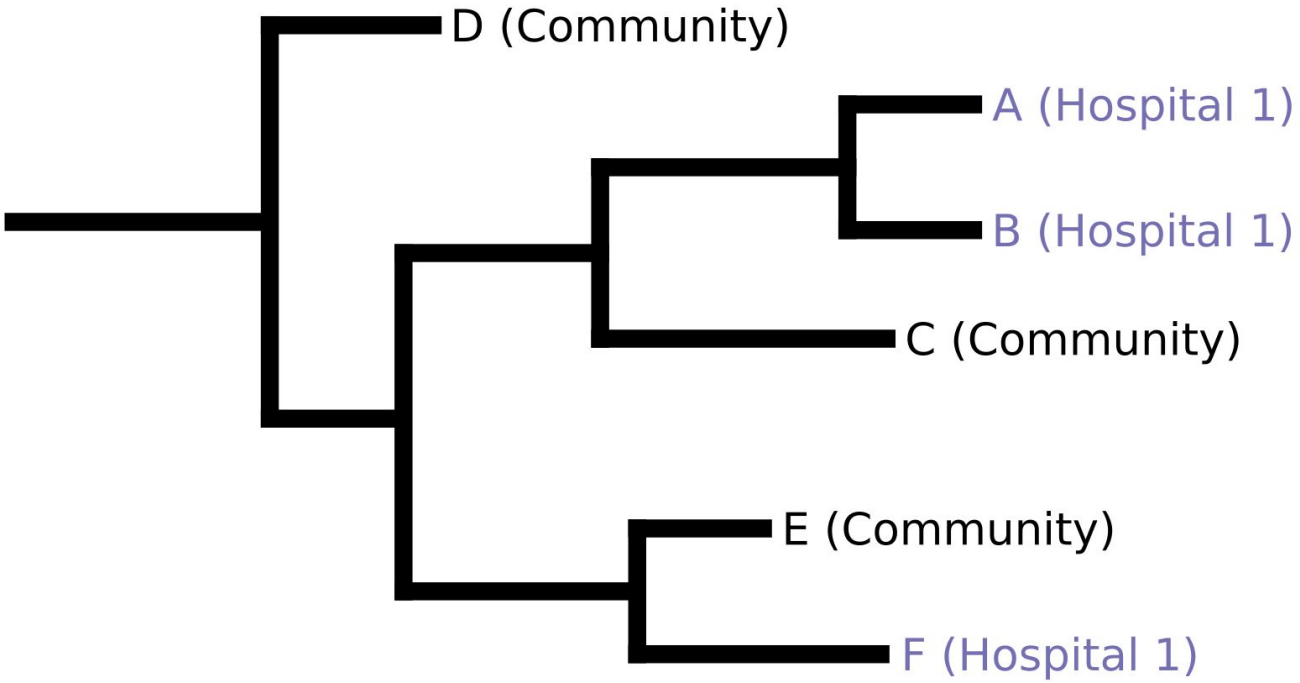
Trace sources of outbreaks



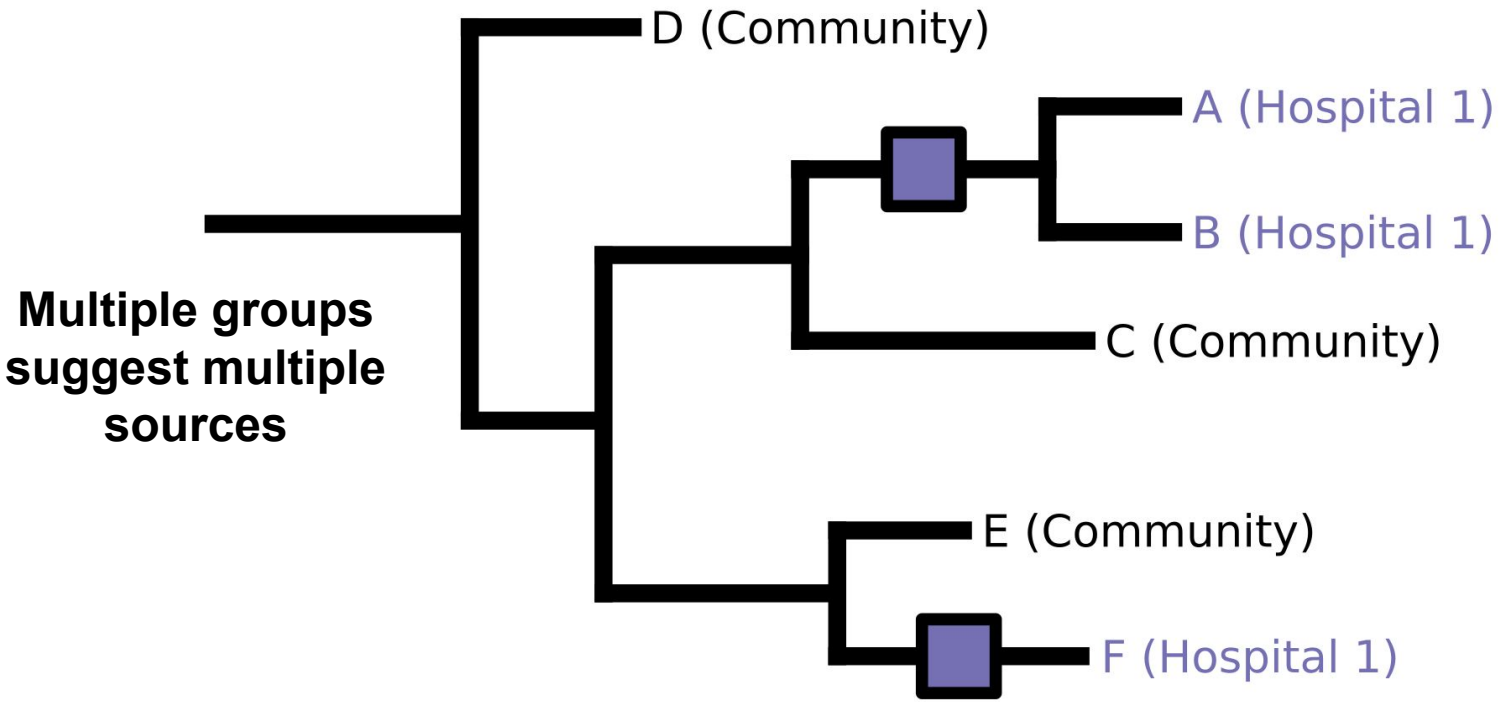
Trace sources of outbreaks



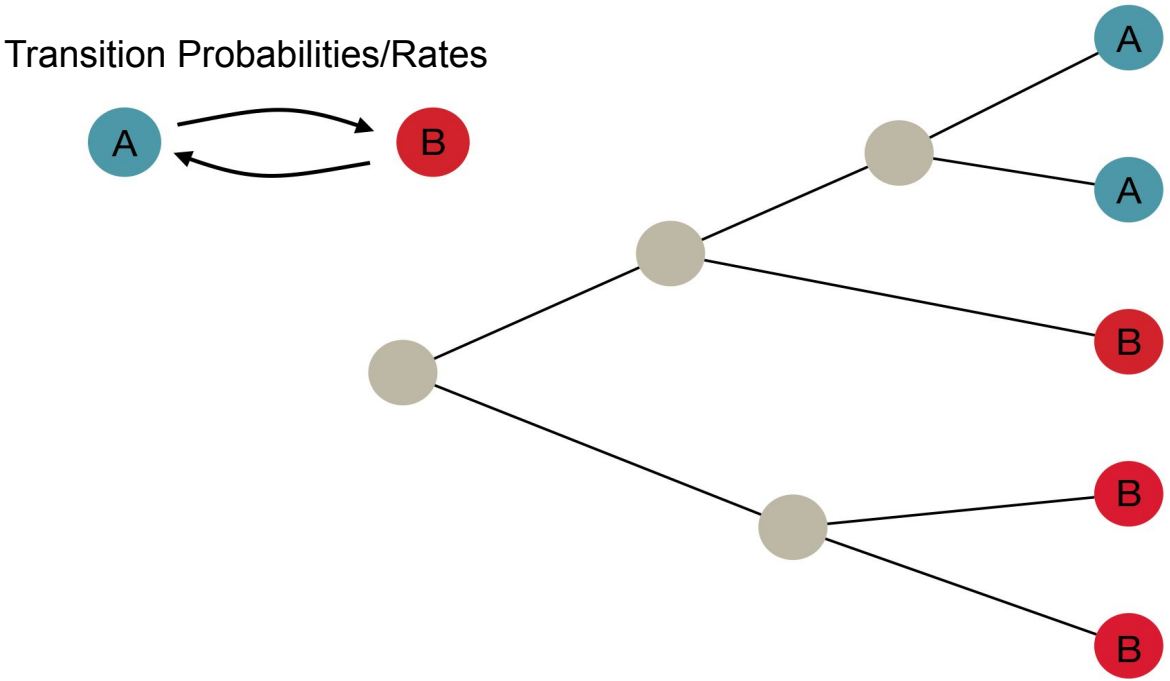
Trace sources of outbreaks



Trace sources of outbreaks

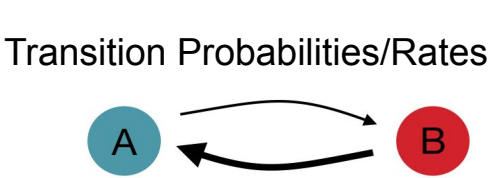


Inferring internal ancestral states from observed tips



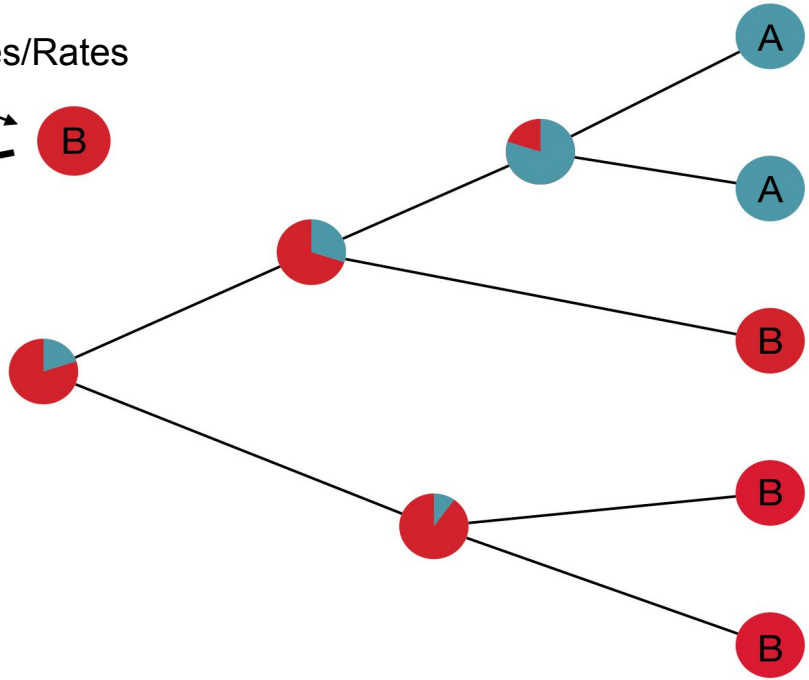
https://github.com/trvrb/gs541-phylodynamics

Inferring internal ancestral states from observed tips



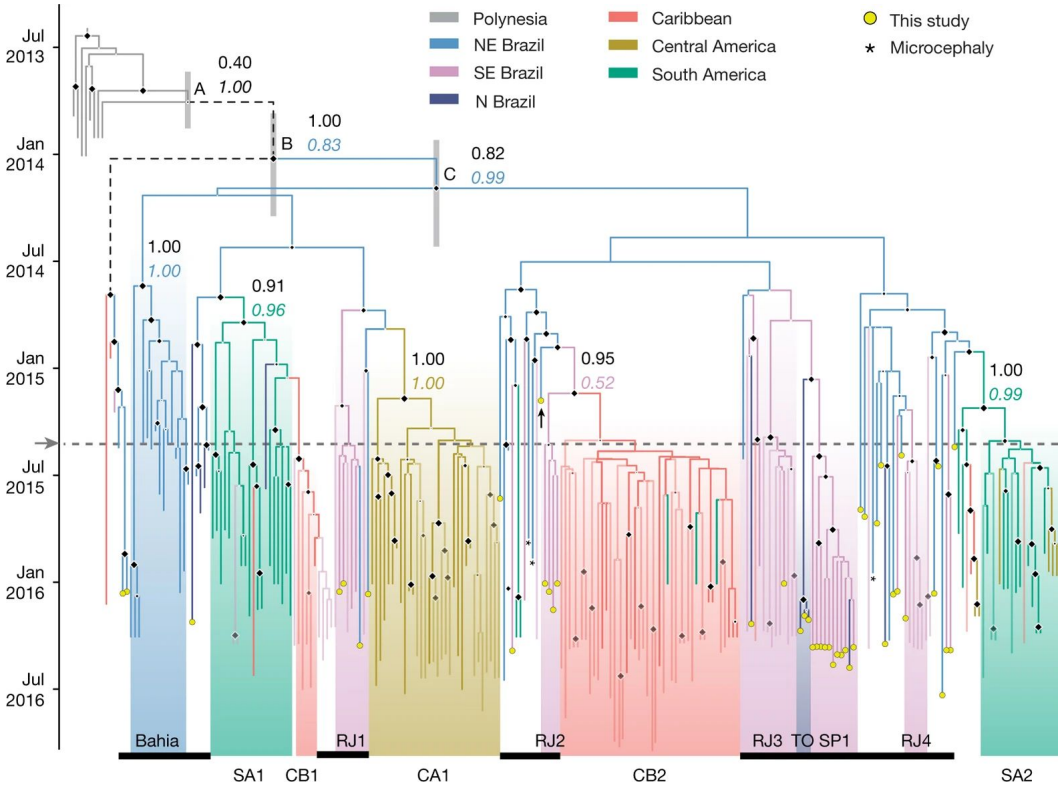
Many different inference approaches:

- Parsimony (e.g., Dollo)
- Maximum likelihood on tree
- **Bayesian modelling**



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

Tracing origin location of Zika



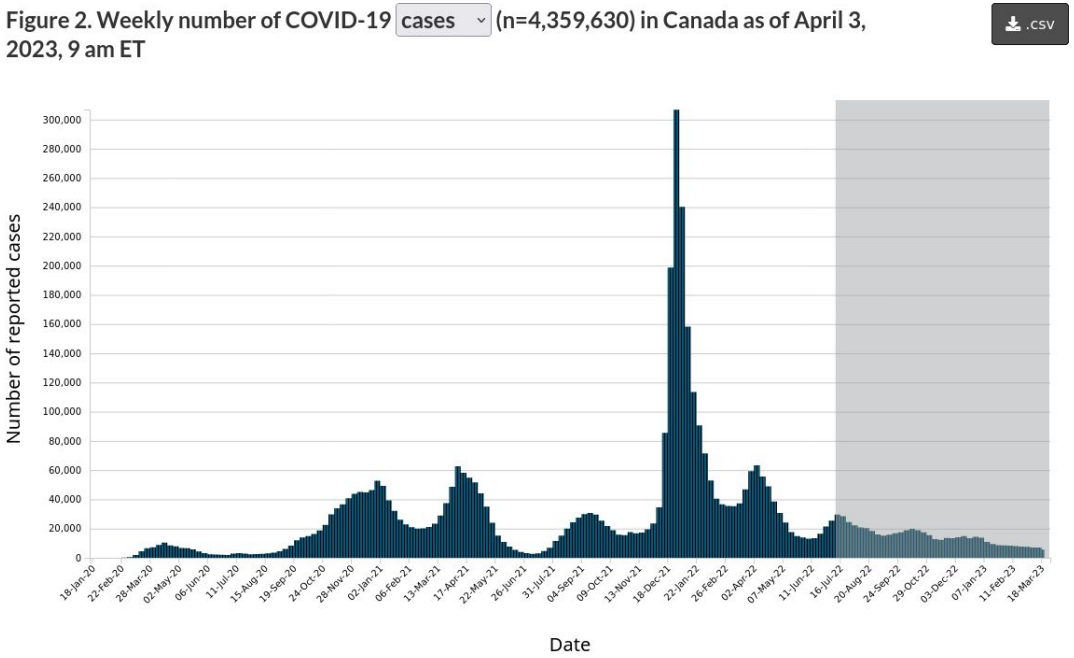
10.1038/nature22401
https://github.com/trvrb/gs541-phylodynamics

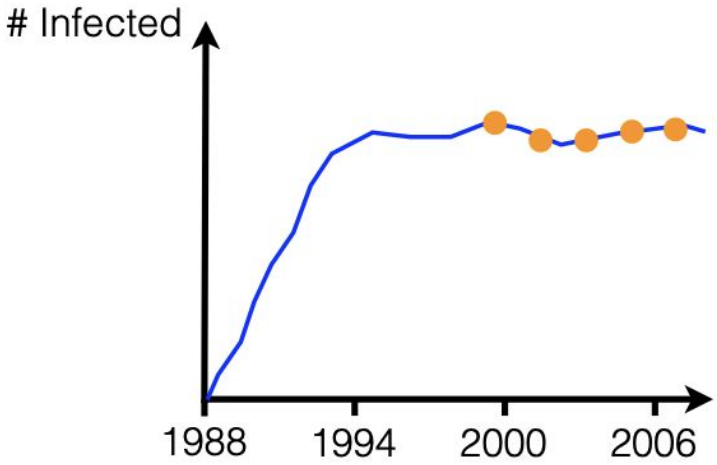
Same approach works for traits like mutations or hosts

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

What about epidemiological parameters?

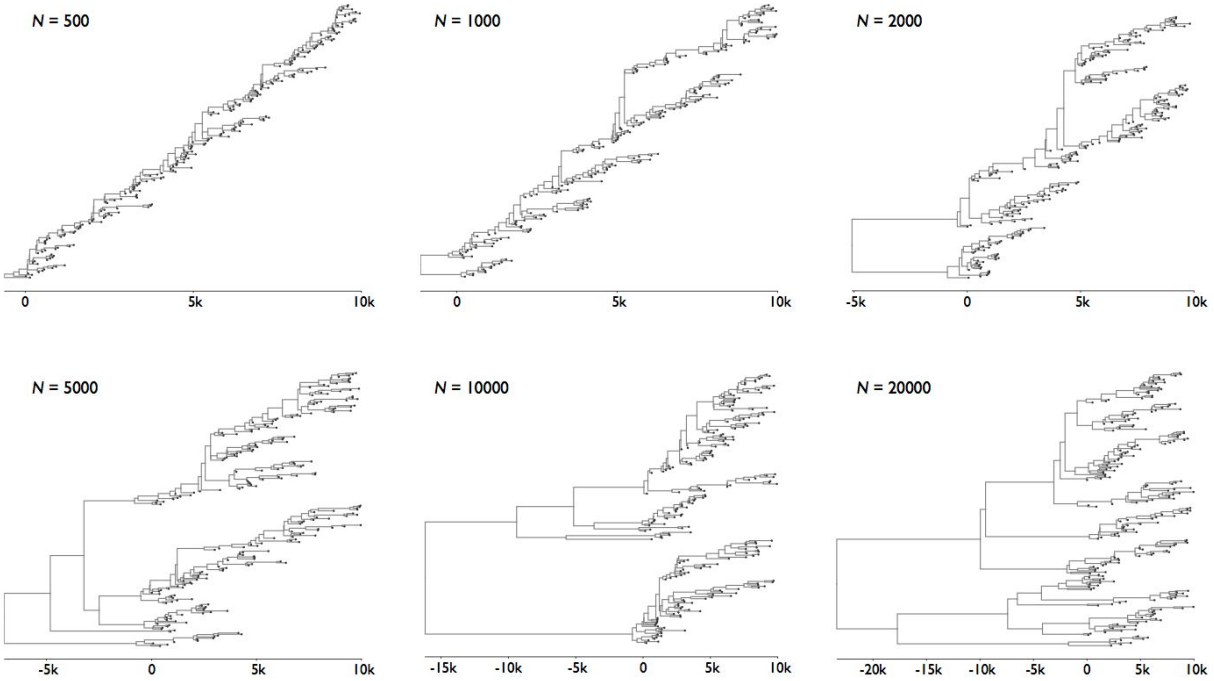
Estimating pathogen population size/structure





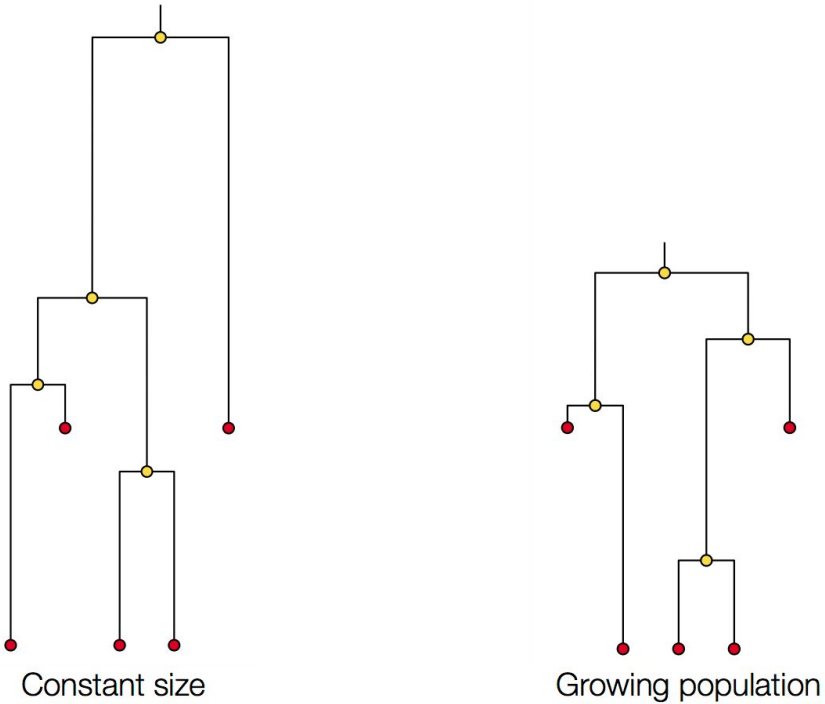
https://github.com/Taming-the-BEAST/Taming-the-BEAST-2021-Online-Lectures/raw/master/Day2_Phylodynamics_-_Tanja_Stadler.pdf
https://covid19.uclaml.org/model.html

Shape of tree relates to population size (and structure)



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

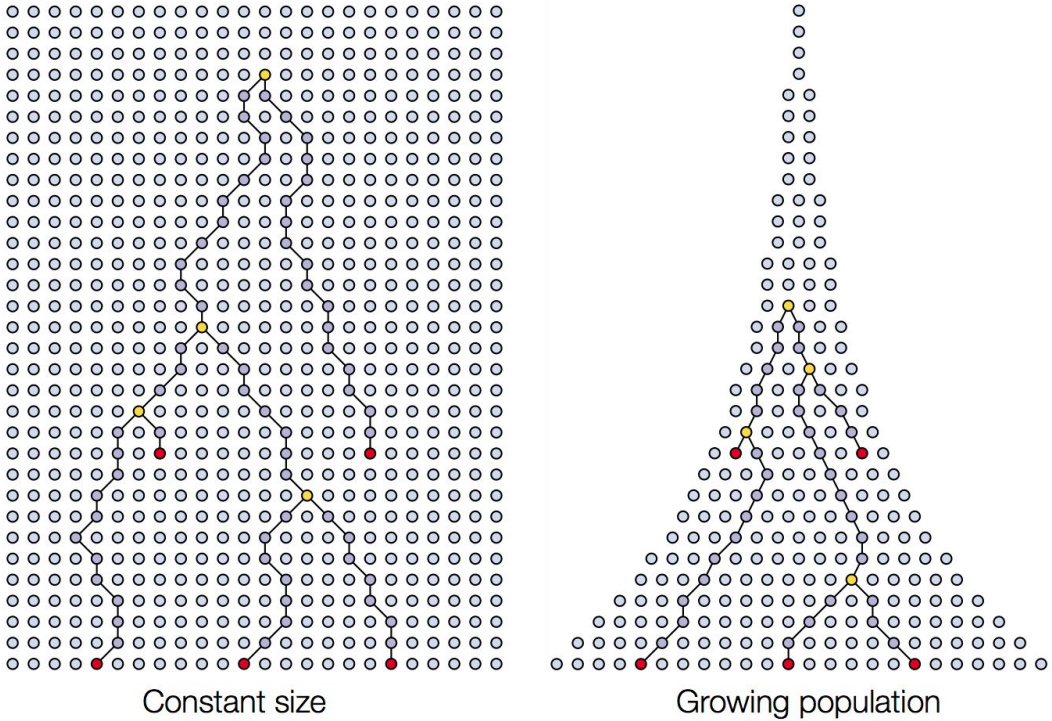
Coalescent processes let us quantify this relationship



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

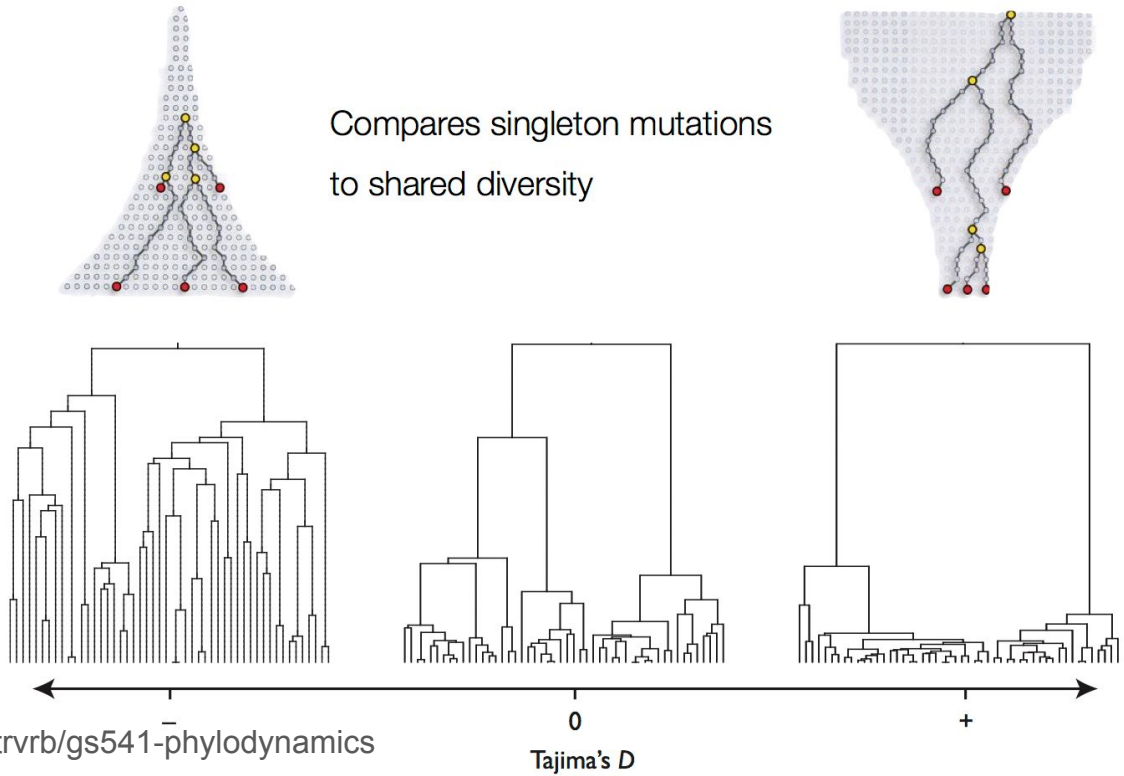
Coalescent processes let us quantify this relationship

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



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

Tajima's D captures deviation from neutral change



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

What about evolutionary forces like selection?

dN/dS is one way to detect selection

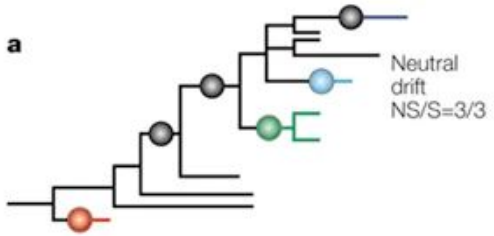
dN = non-synonymous mutations (normalised)

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

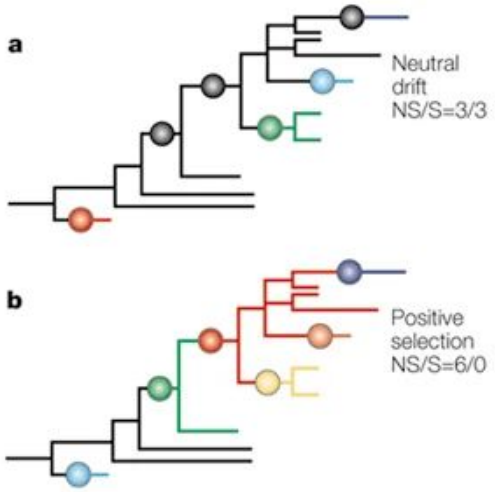
dN/dS is one way to detect selection

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

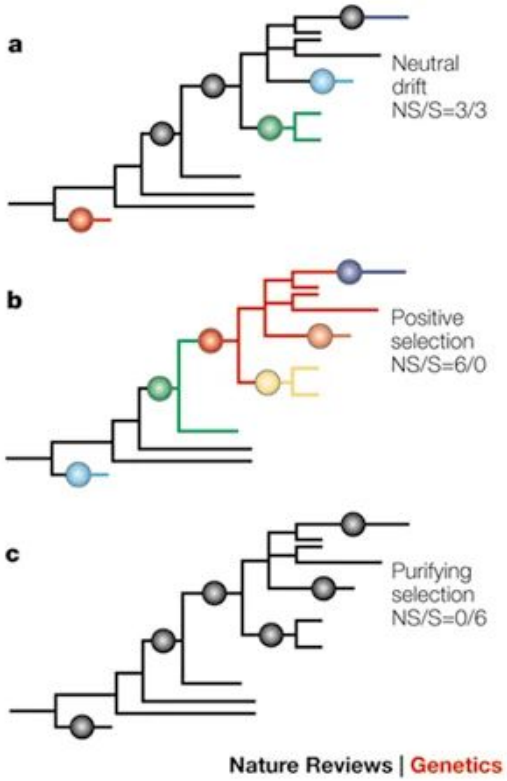
dN = non-synonymous mutations (normalised)

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)

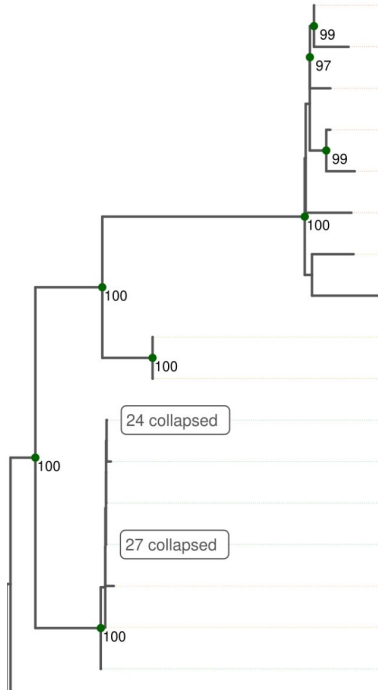
dN/dS > 1 : adaptive/positive selection

dN/dS ~ 1 : drift/neutral selection

dN/dS < 1: purifying/negative selection

Challenges:

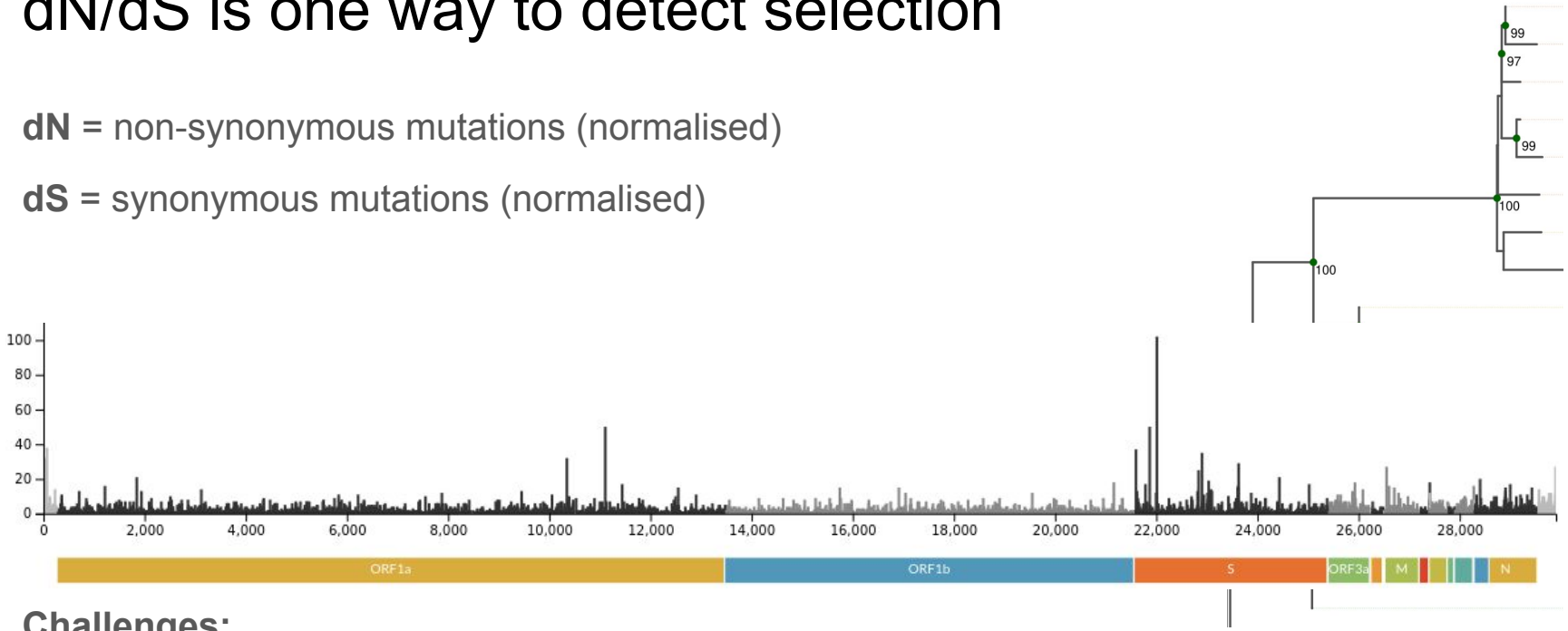
- Mutation rates vary over time/groups



dN/dS is one way to detect selection

dN = non-synonymous mutations (normalised)

dS = synonymous mutations (normalised)



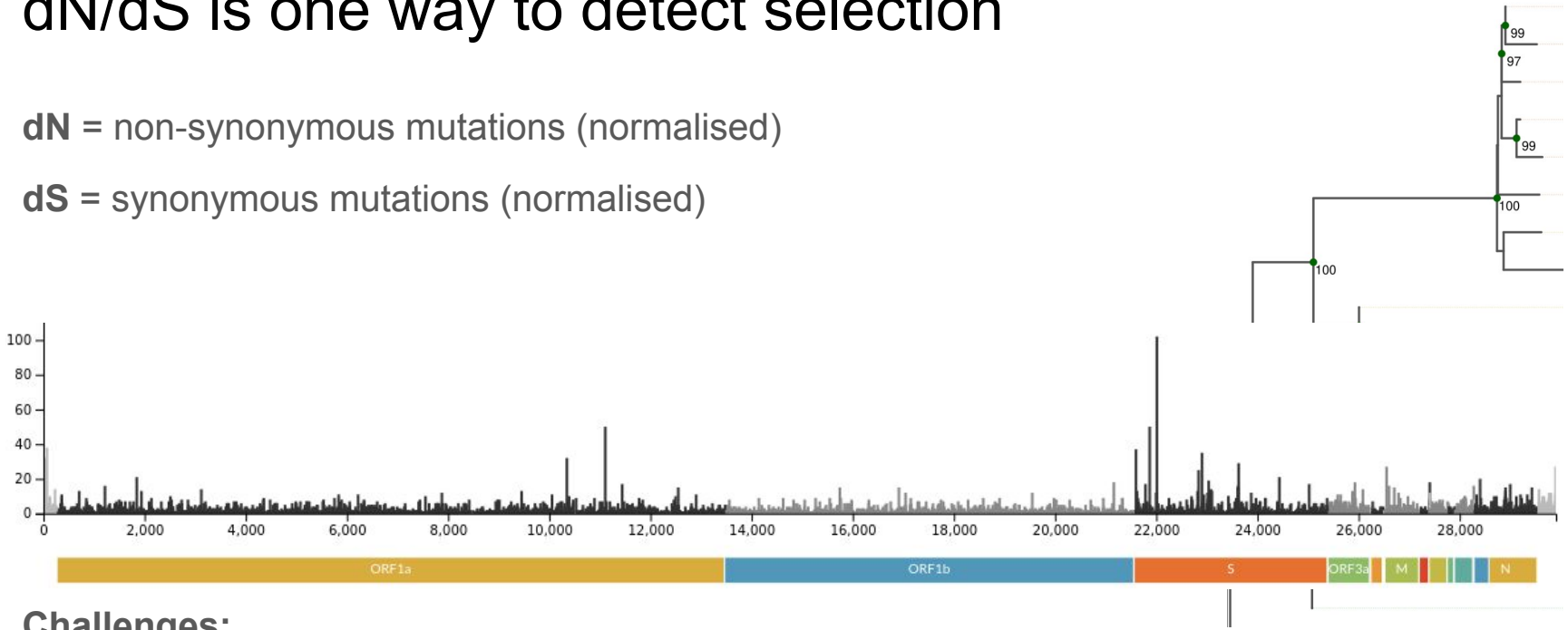
Challenges:

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

dN/dS is one way to detect selection

dN = non-synonymous mutations (normalised)

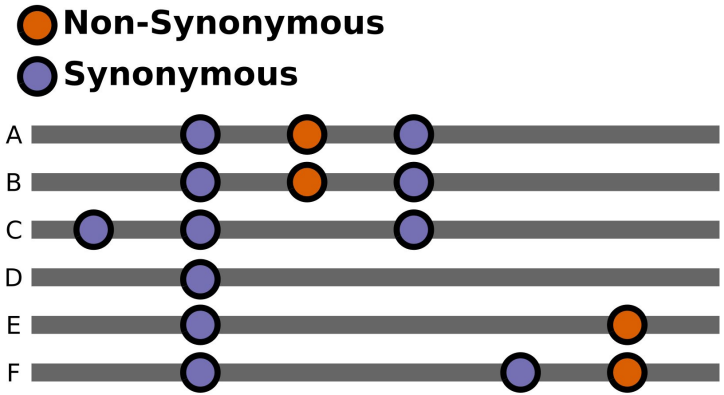
dS = synonymous mutations (normalised)



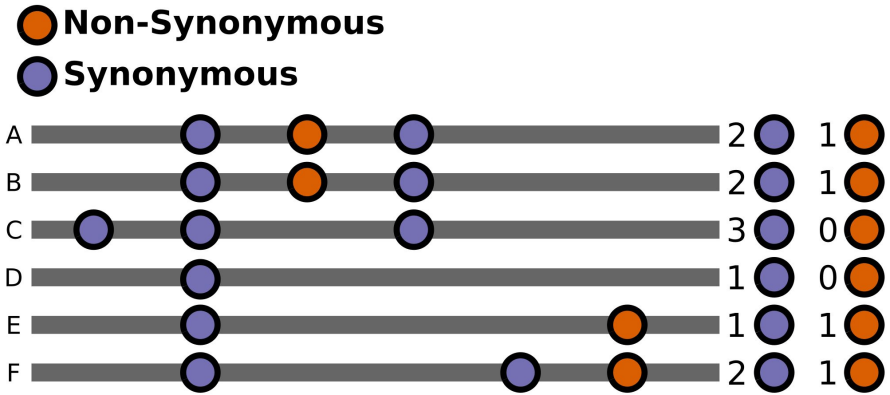
Challenges:

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

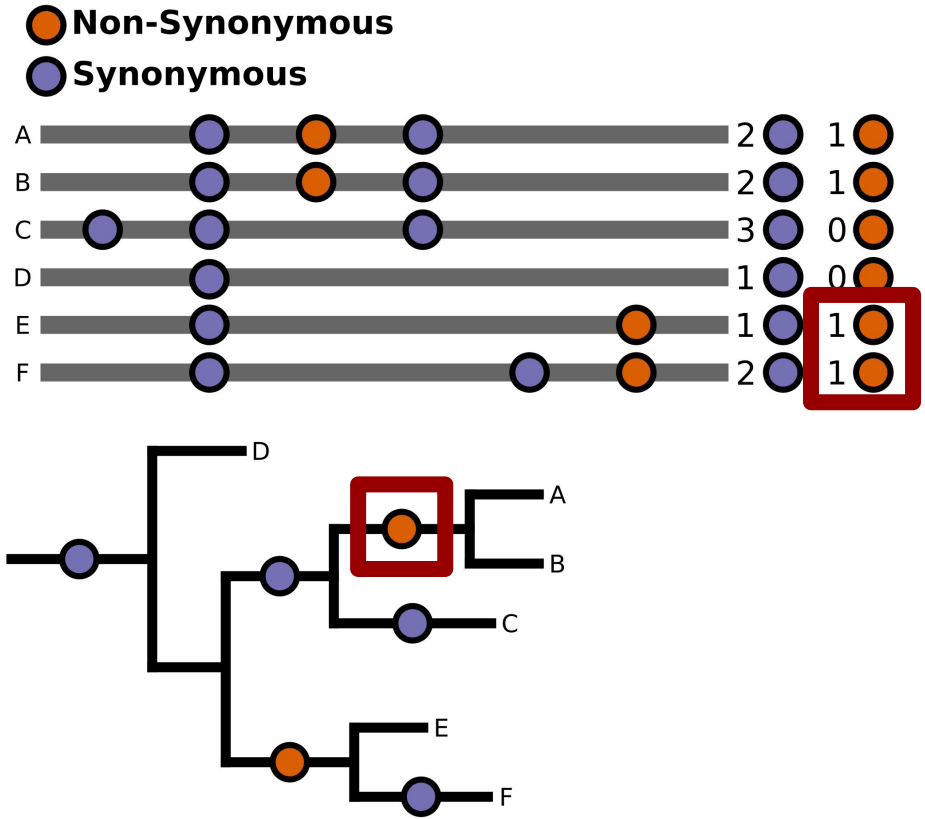
Non-independence of events in related genomes



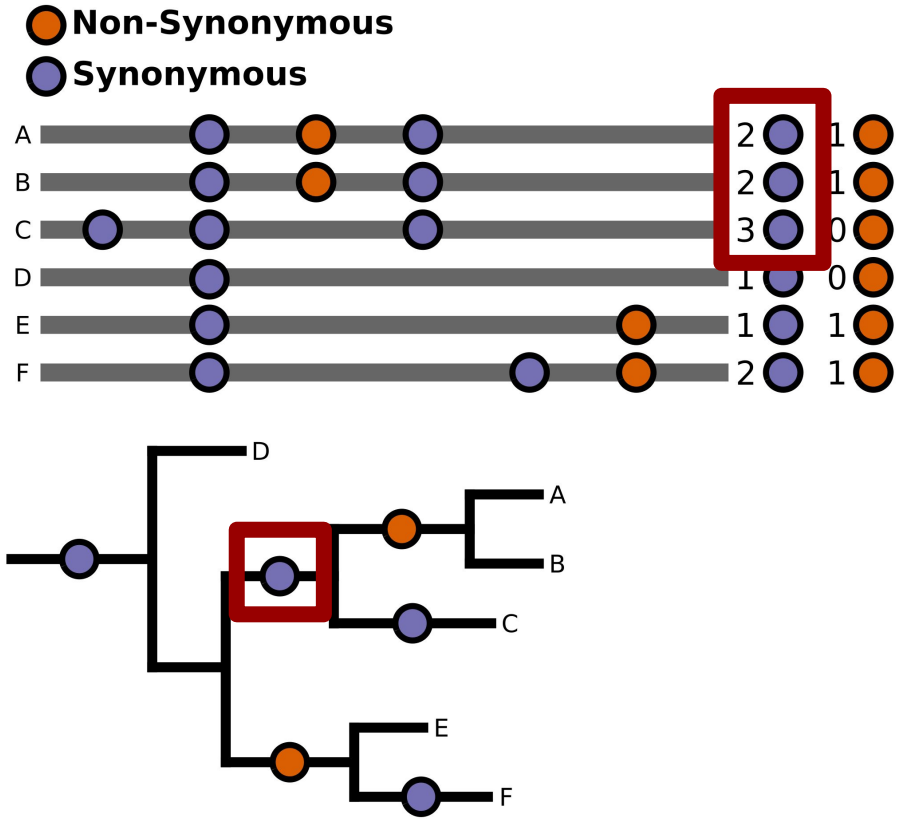
Non-independence of events in related genomes



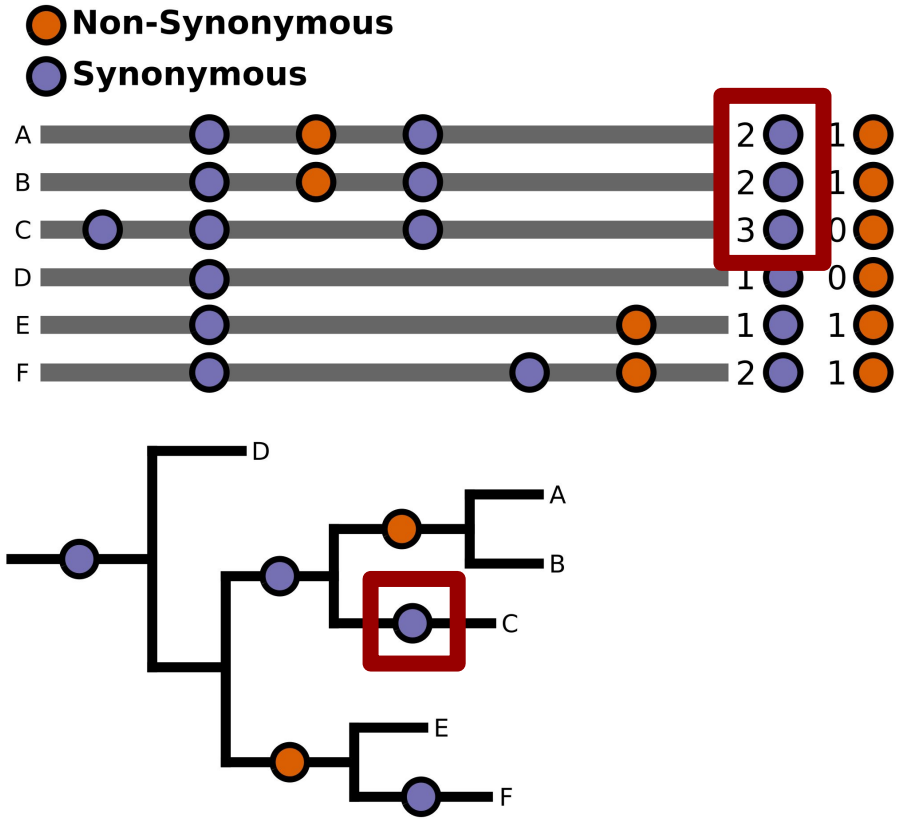
Non-independence of events in related genomes



Non-independence of events in related genomes

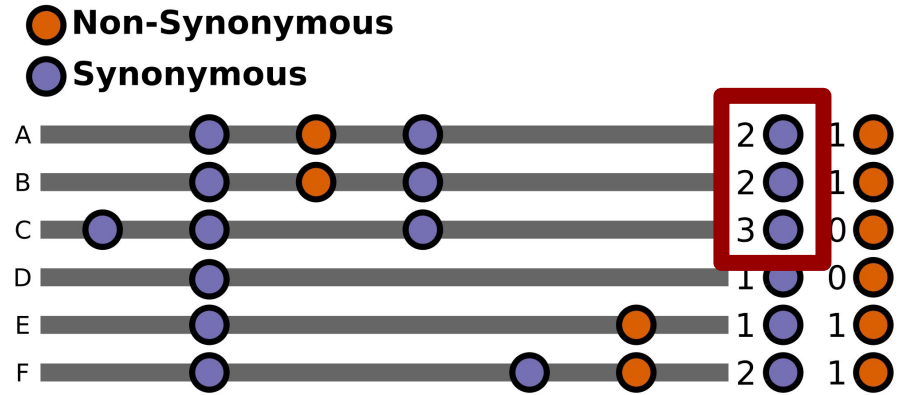


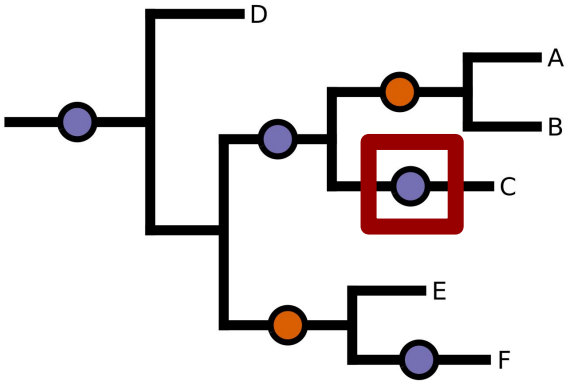
Non-independence of events in related genomes



- Phylogeny captures dependency structure of genomic data

Non-independence of events in related genomes



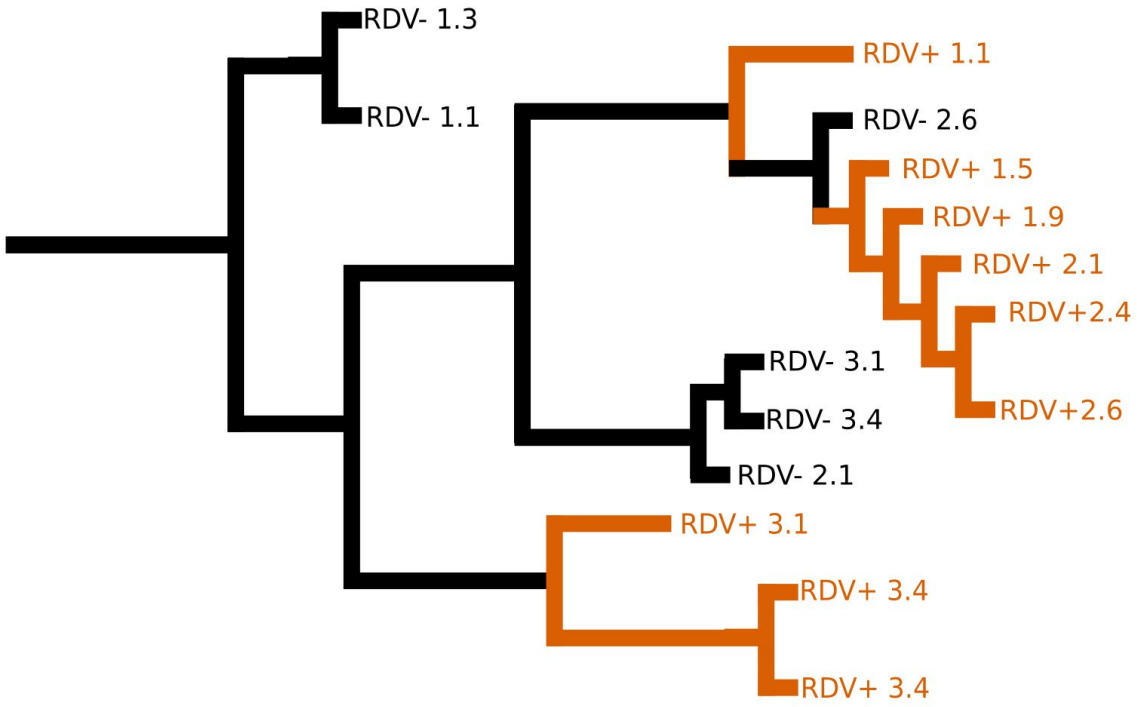


- Phylogeny captures dependency structure of genomic data
- Informs error term for models (e.g., regression)

- **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



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**

We are on a Coffee Break & Networking Session

Workshop Sponsors: