



Dated Bayesian phylogenetic analysis to infer parameters of epidemiological interest

Sam Hong
Rega Institute of Medical Research
KU Leuven - Belgium
samuel.hong@kuleuven.be

Some applied questions addressed by viral genomic epidemiology



Tuesday What virus is causing the outbreak?

Wednesday What species are involved in virus transmission?

Thursday When did the outbreak begin?

Thursday How many introductions have there been?

Friday Where did the outbreak begin?

Friday How fast is the virus evolving?

Friday How rapidly is the virus transmitting?

Friday What factors drive an outbreak?

Workshop Materials
and Detailed Agenda:



Ask questions, participate, network and join the Workshop's WhatsApp groups ☺

Some applied questions addressed by viral genomic epidemiology



Tuesday

What virus is causing the outbreak?

Wednesday

What species are involved in virus transmission?

Thursday

When did the outbreak begin?

Thursday

How many introductions have there been?

Friday

Where did the outbreak begin?

Friday

How fast is the virus evolving?

Friday

How rapidly is the virus transmitting?

Friday

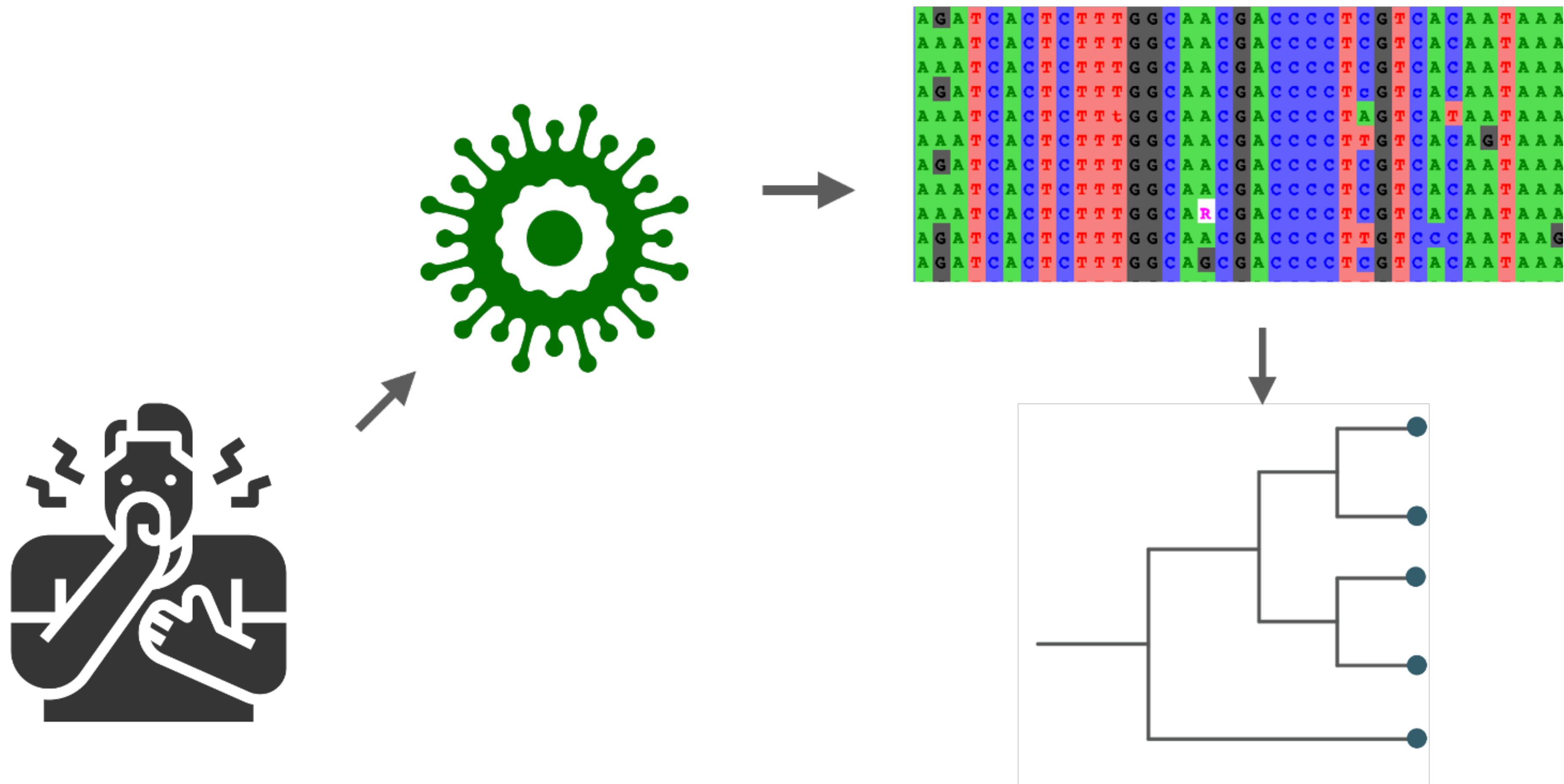
What factors drive an outbreak?

Workshop Materials
and Detailed Agenda:



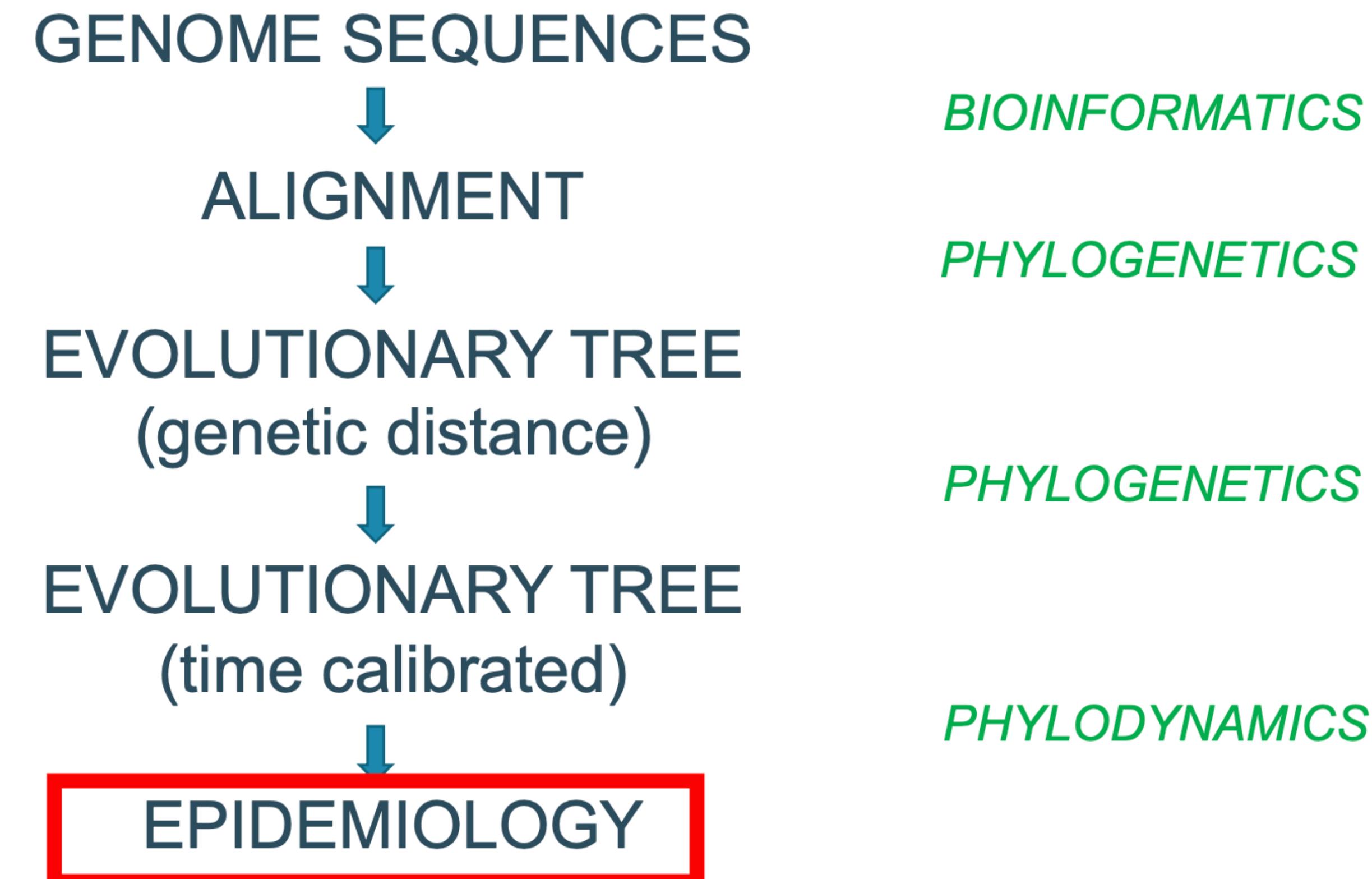
Ask questions, participate, network and join the Workshop's WhatsApp groups ☺

Genomic Epidemiology



From sequence to epidemiology

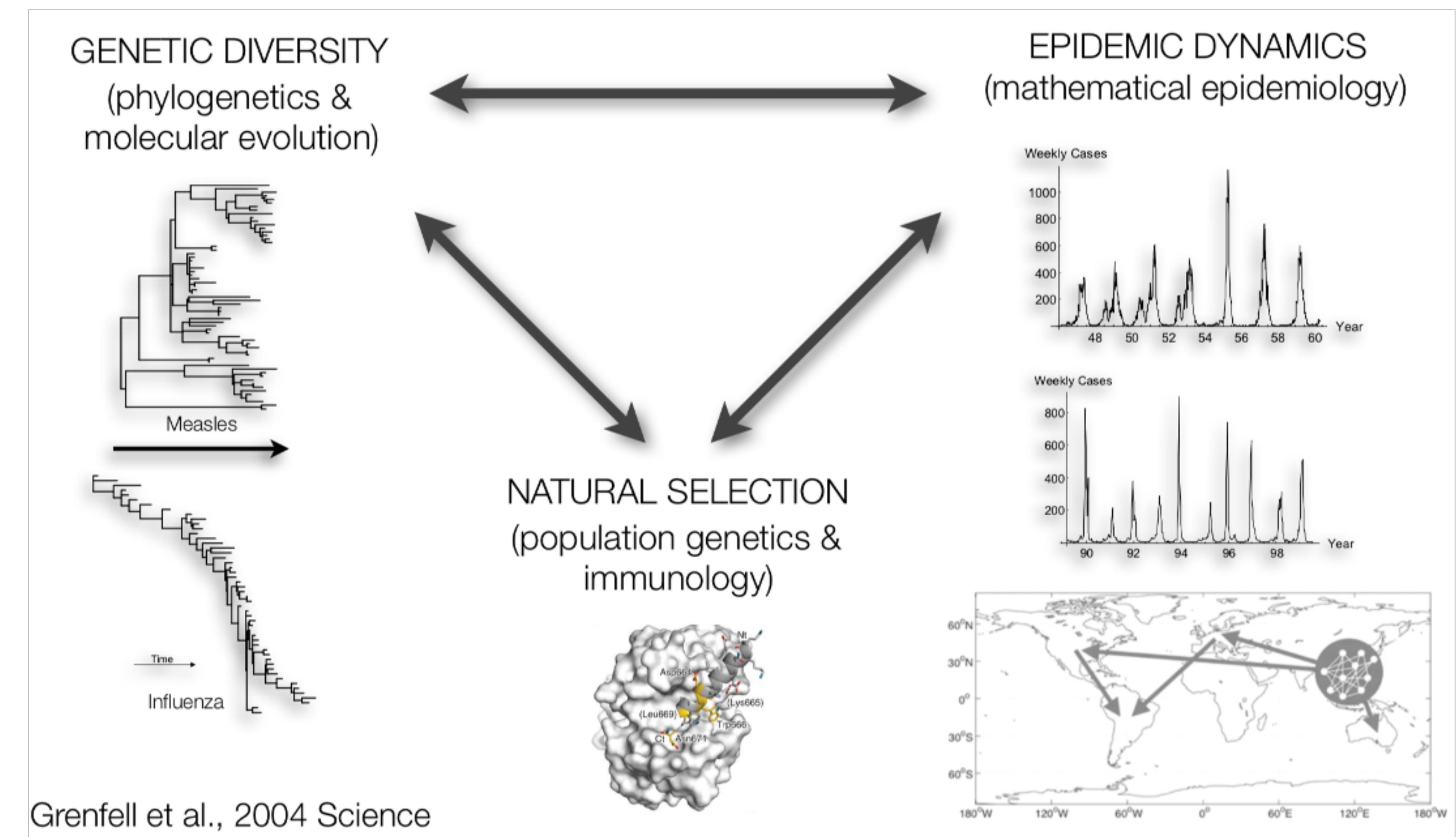
Alignment Methods
Sequence Evolution Models
Phylogenetic Reconstruction
Molecular Clock Models
Phylodynamic Models



Adapted from Philippe Lemey

Phyldynamics

In many infectious pathogens, evolutionary and epidemiological dynamics occur within the same time scale.

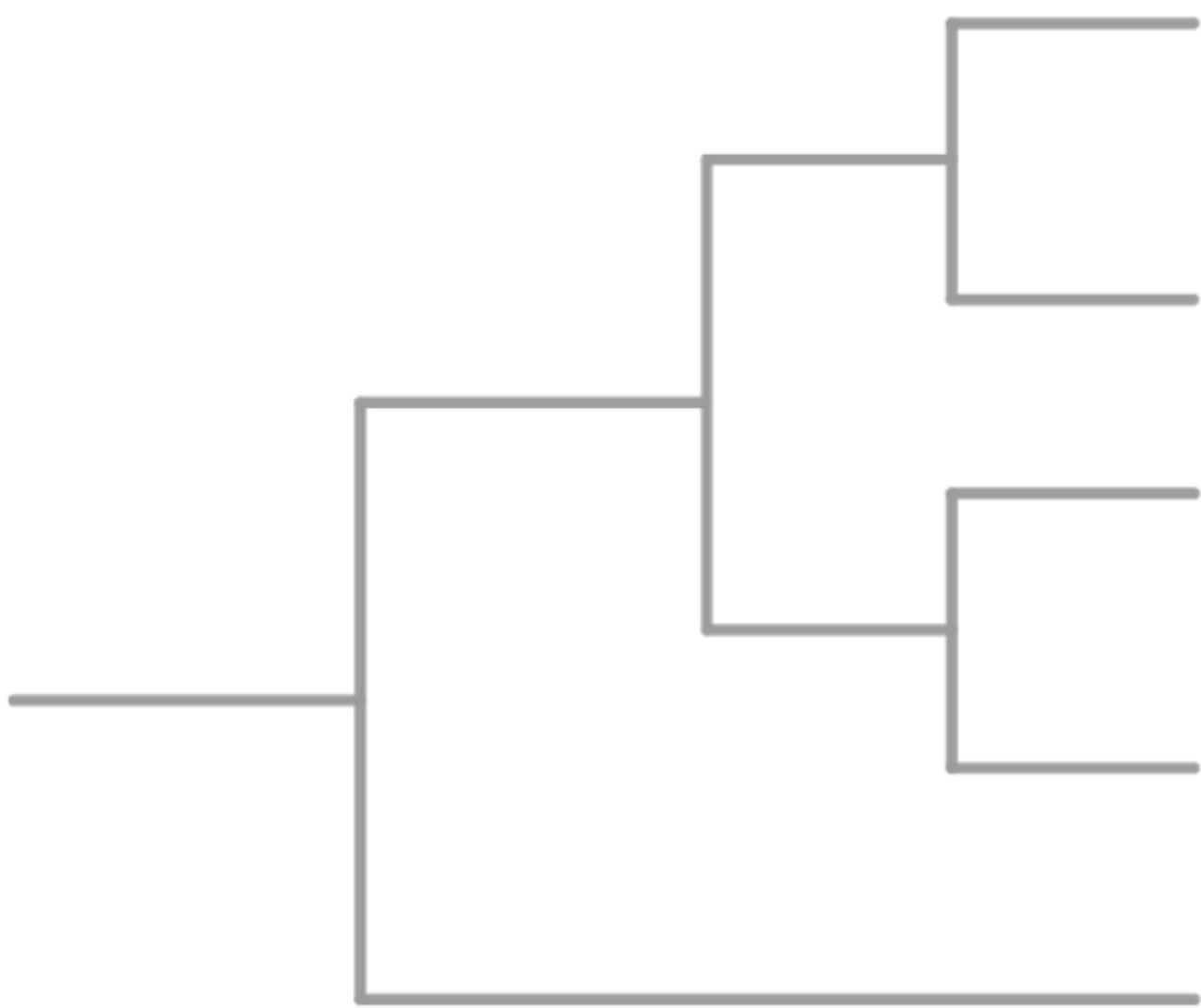


Bayesian Phylogenetics

Maximum Likelihood

$$\operatorname{argmax}_{\theta} L(\theta)$$

Likelihood: $L(\theta) \propto P(Data|\theta)$

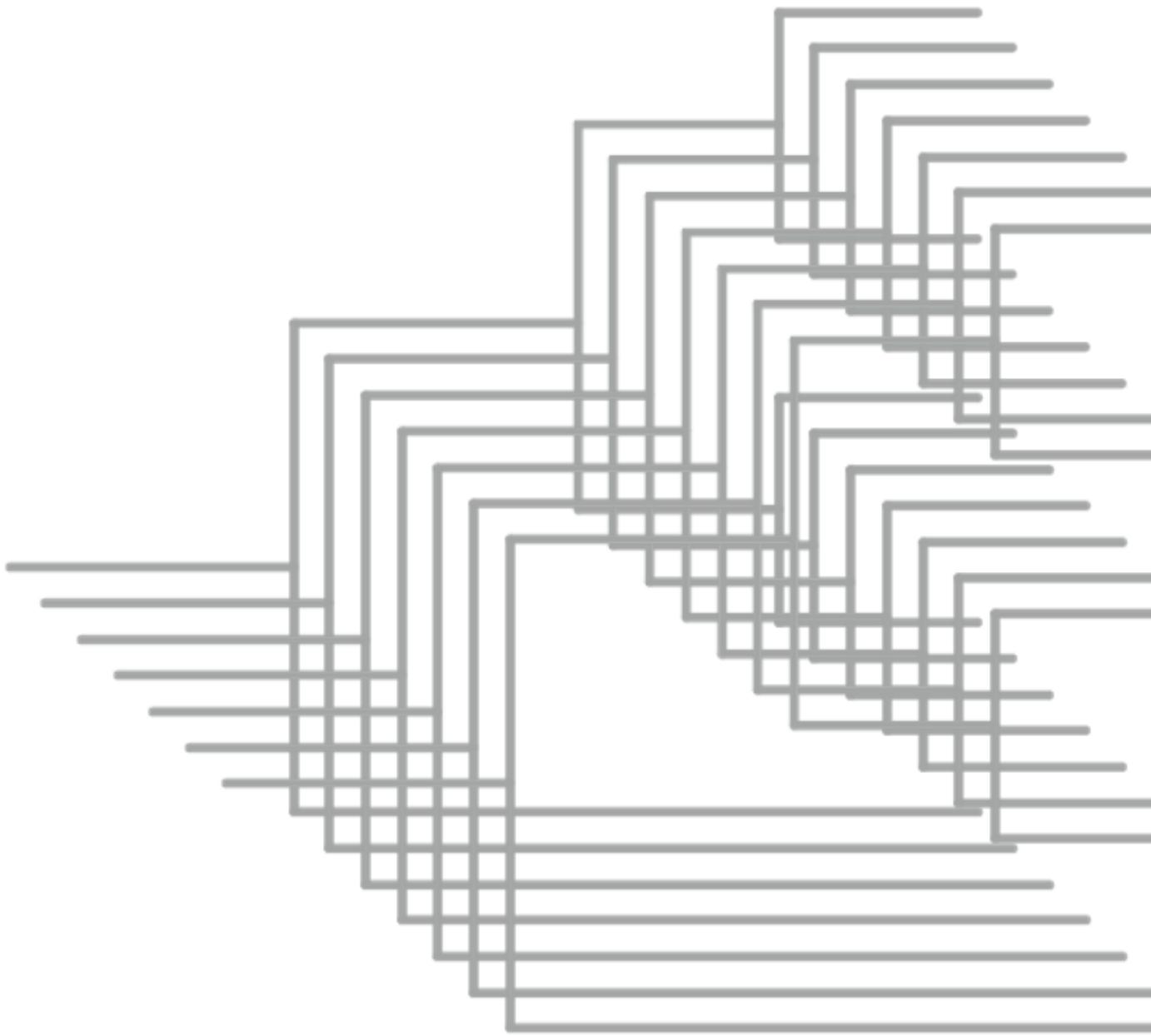


Posterior

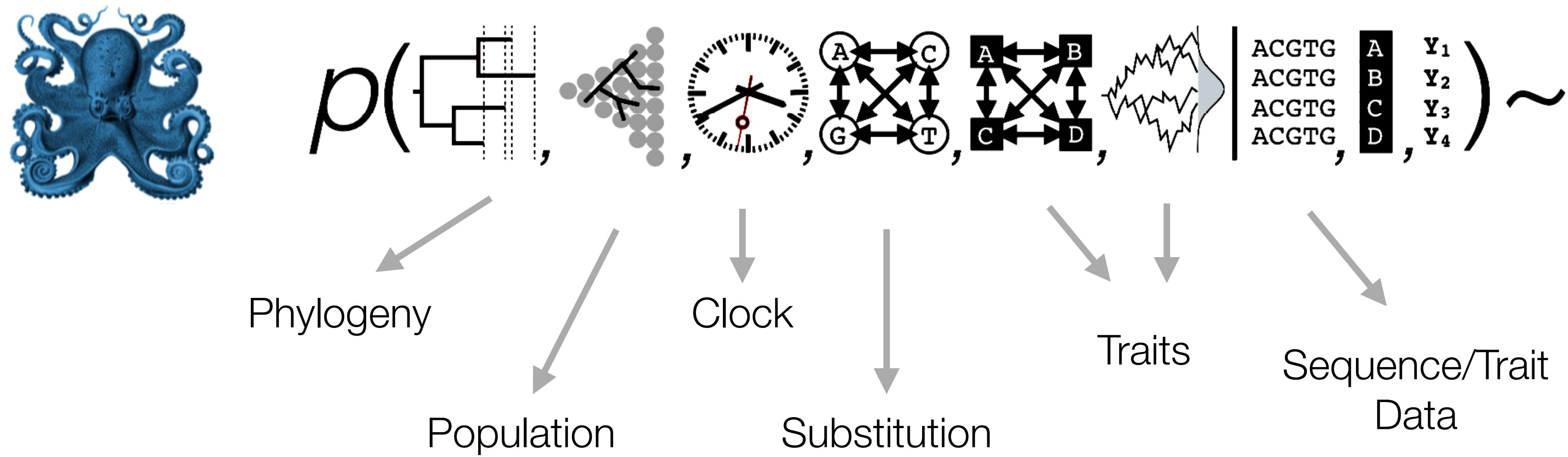
Bayesian

$$P(\theta|Data) = \frac{P(Data|\theta)P(\theta)}{P(Data)}$$

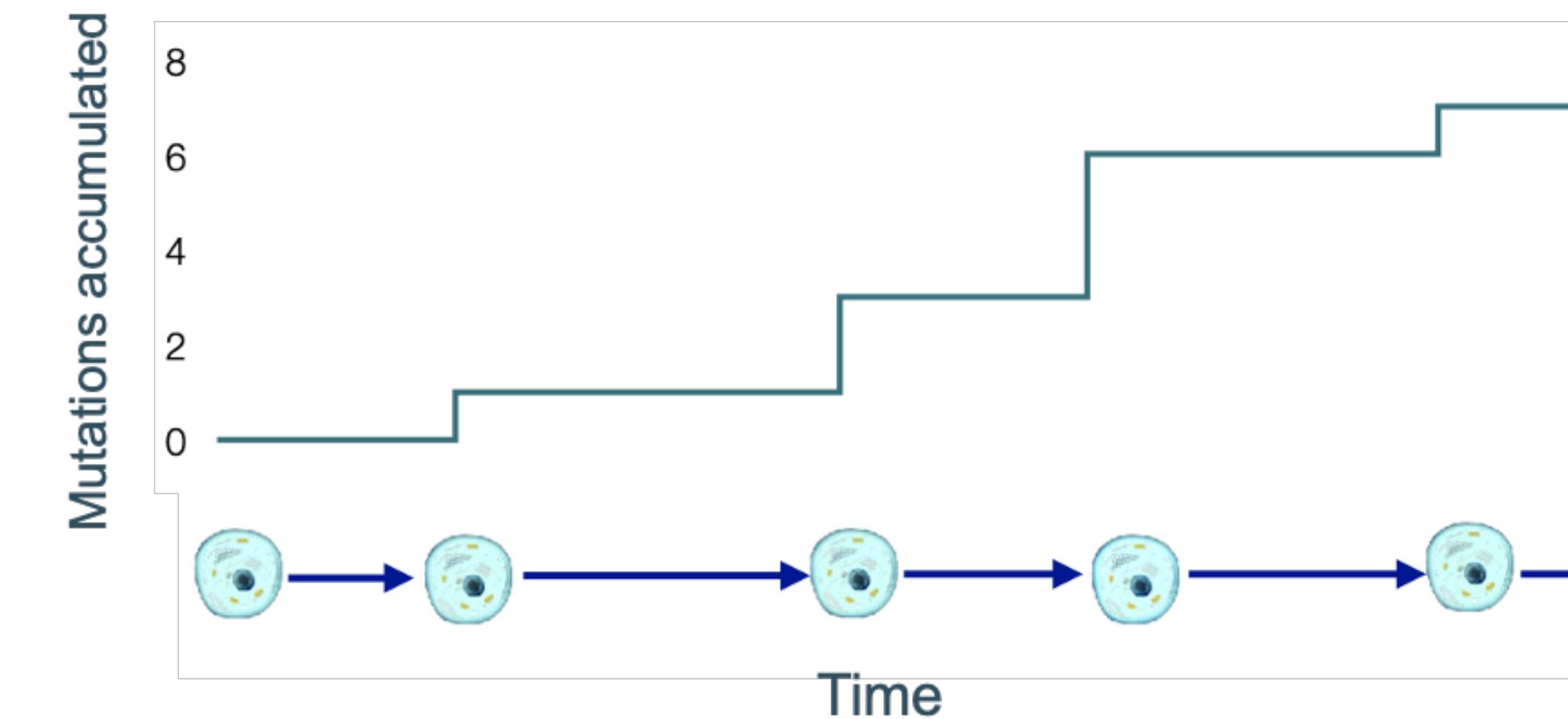
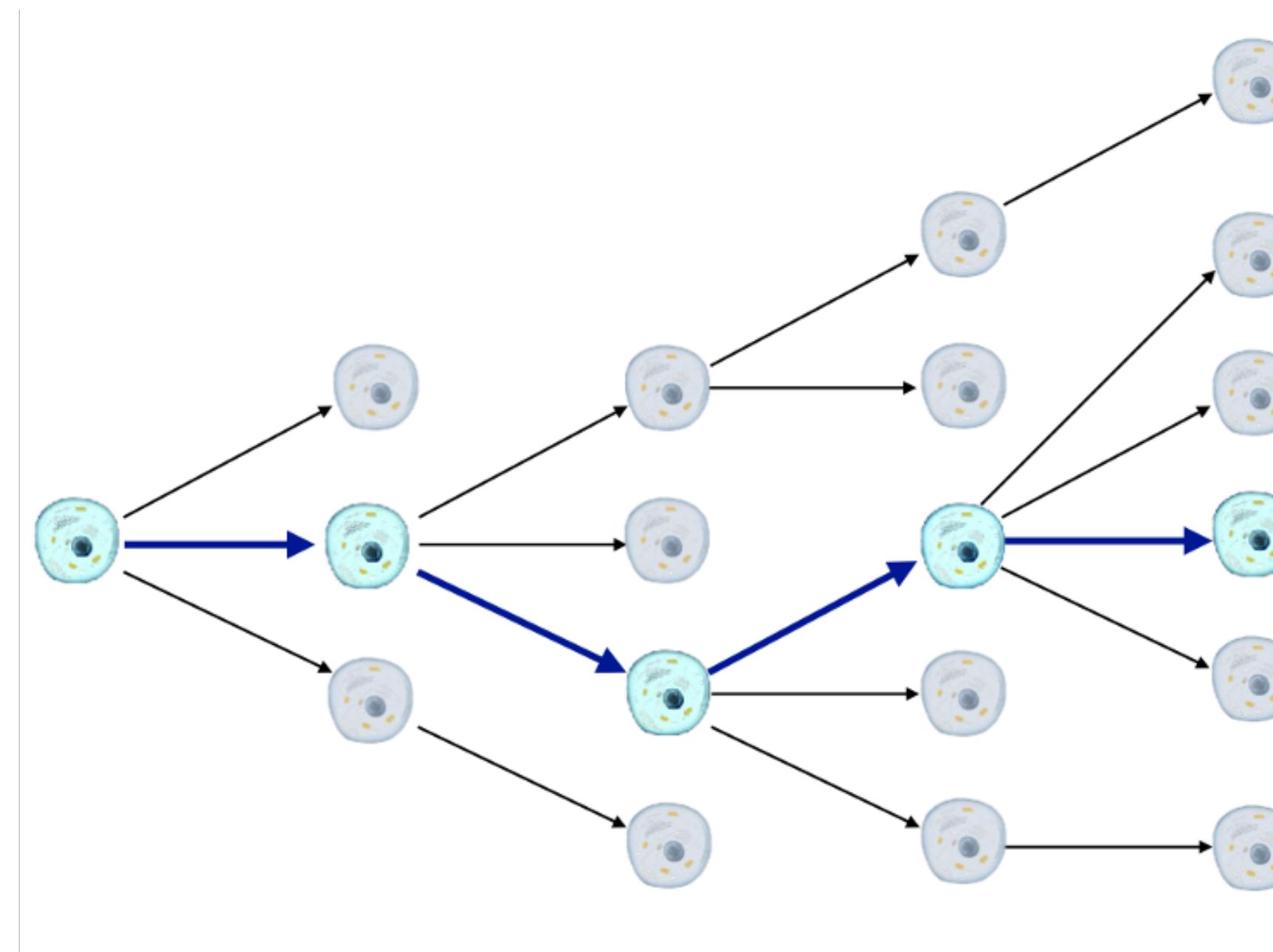
Prior



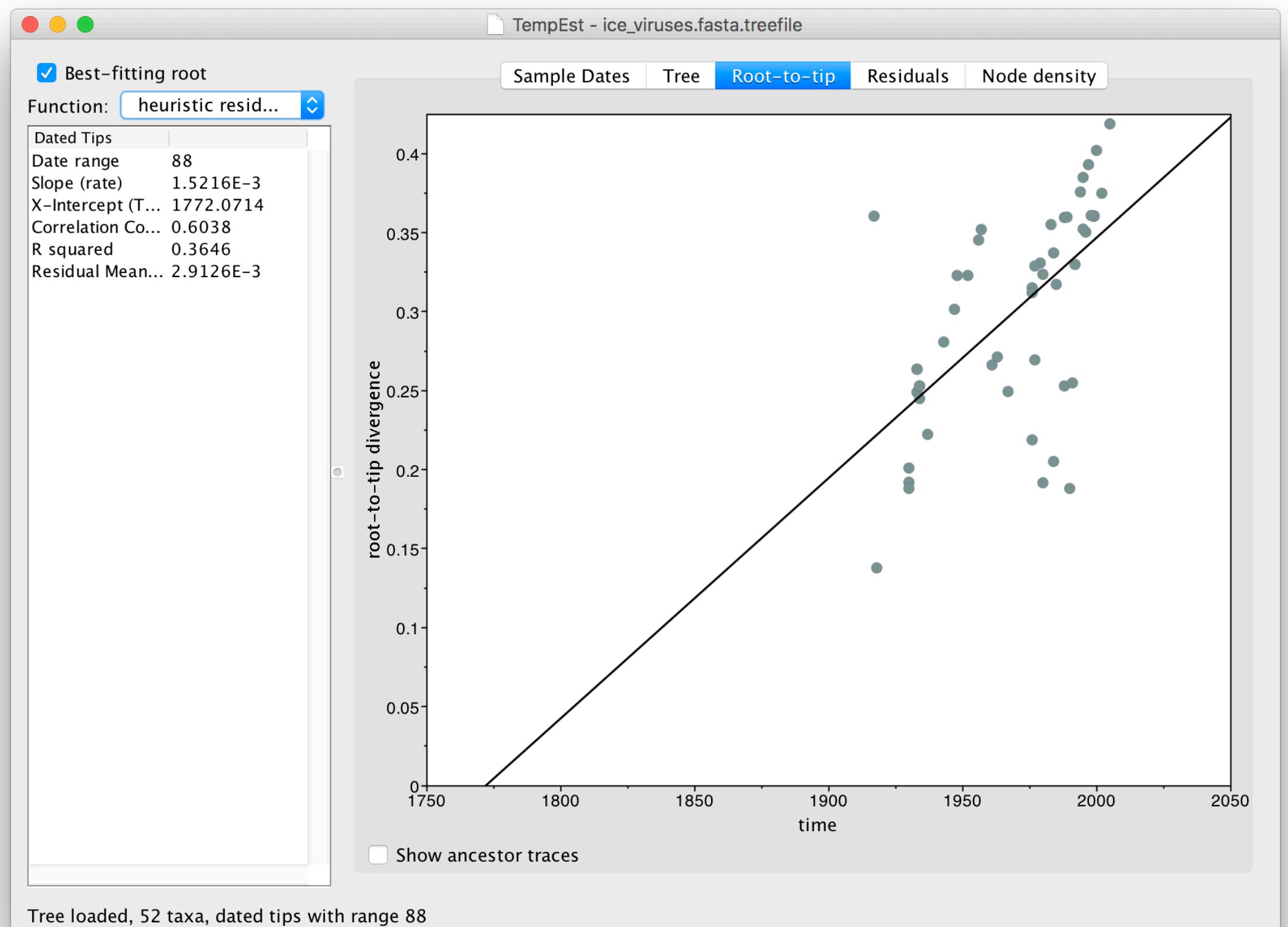
Bayesian phyloodynamics with BEAST



Learning from trees: molecular clocks

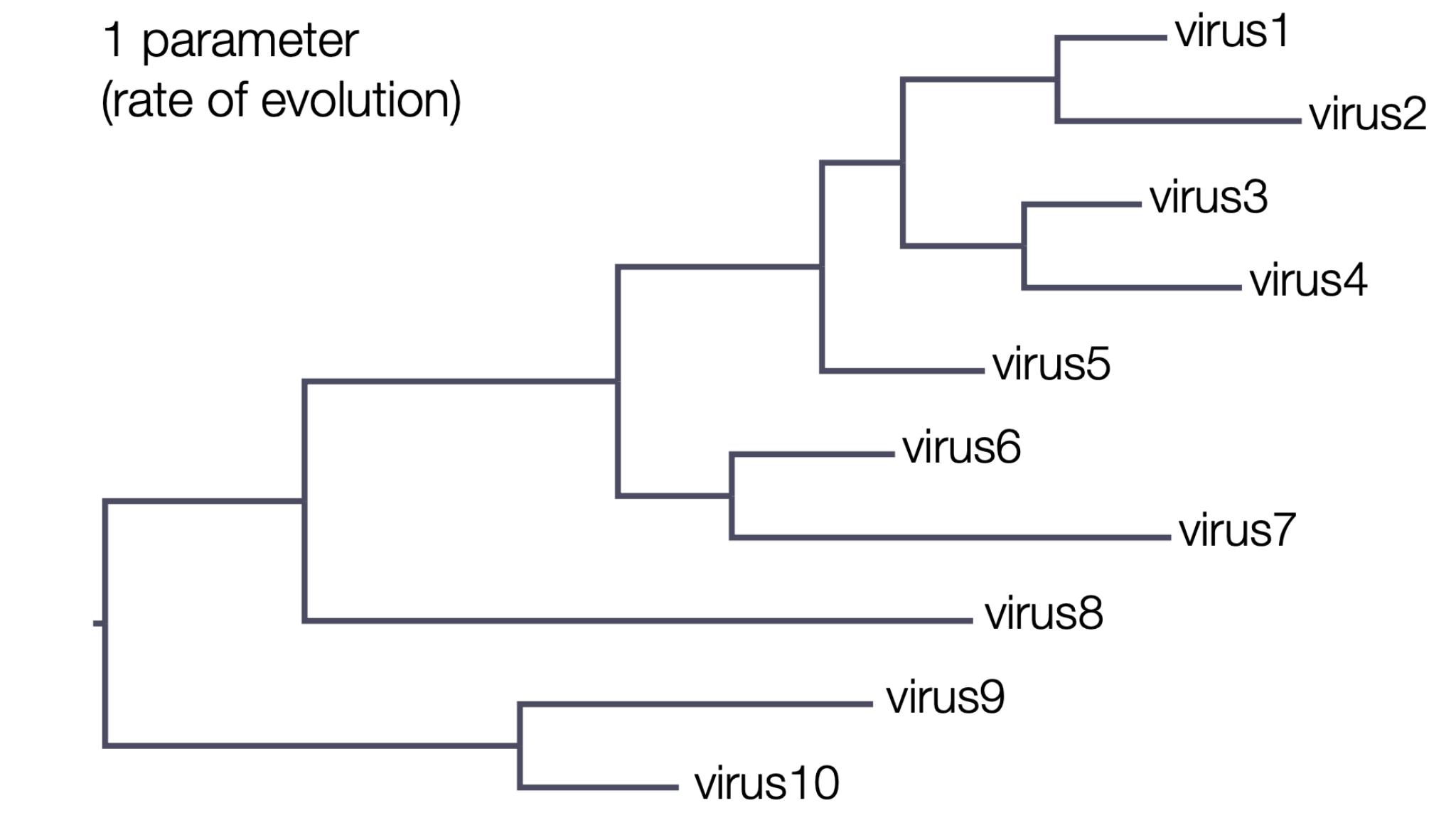


Clock models



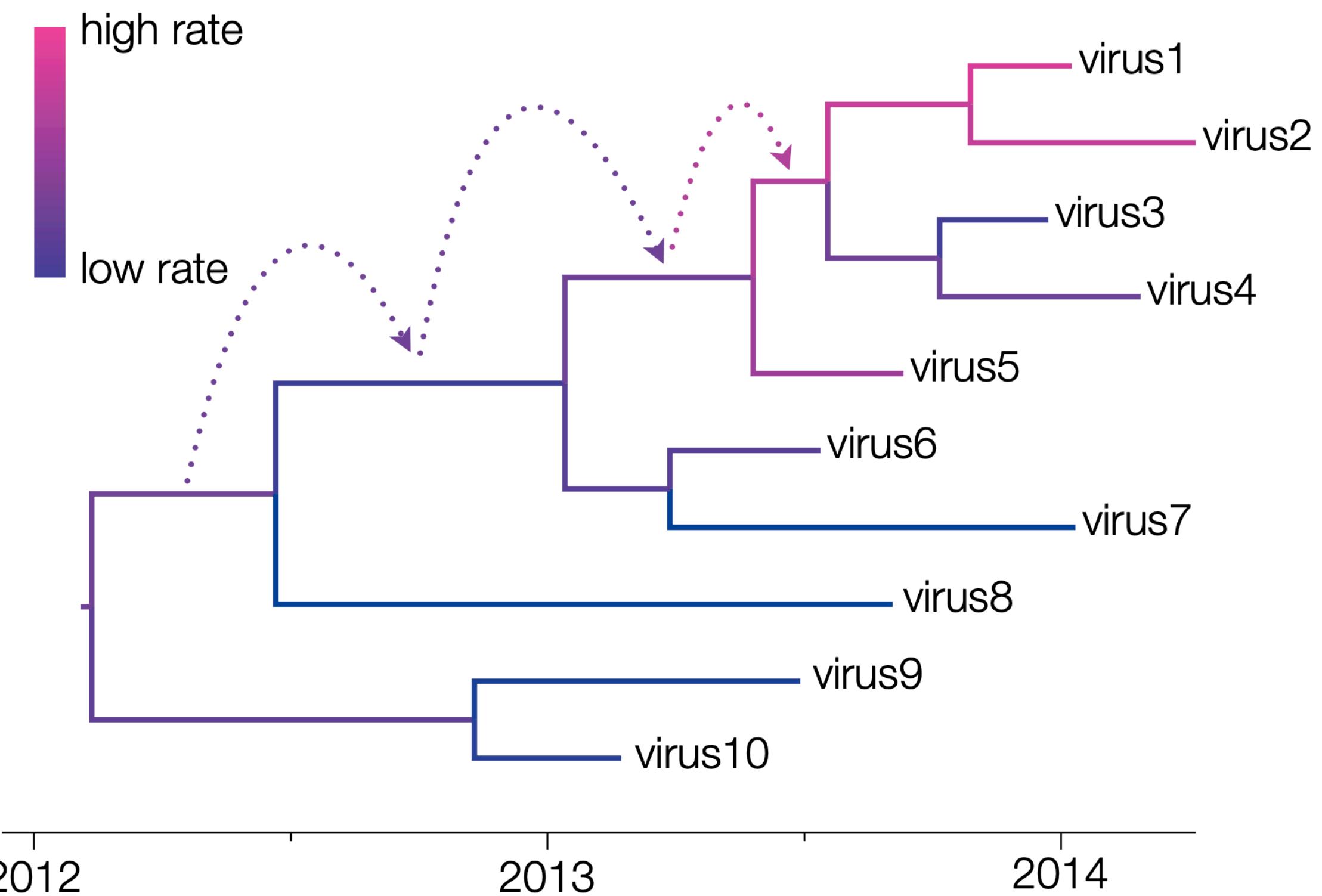
'strict' molecular clock

1 parameter
(rate of evolution)

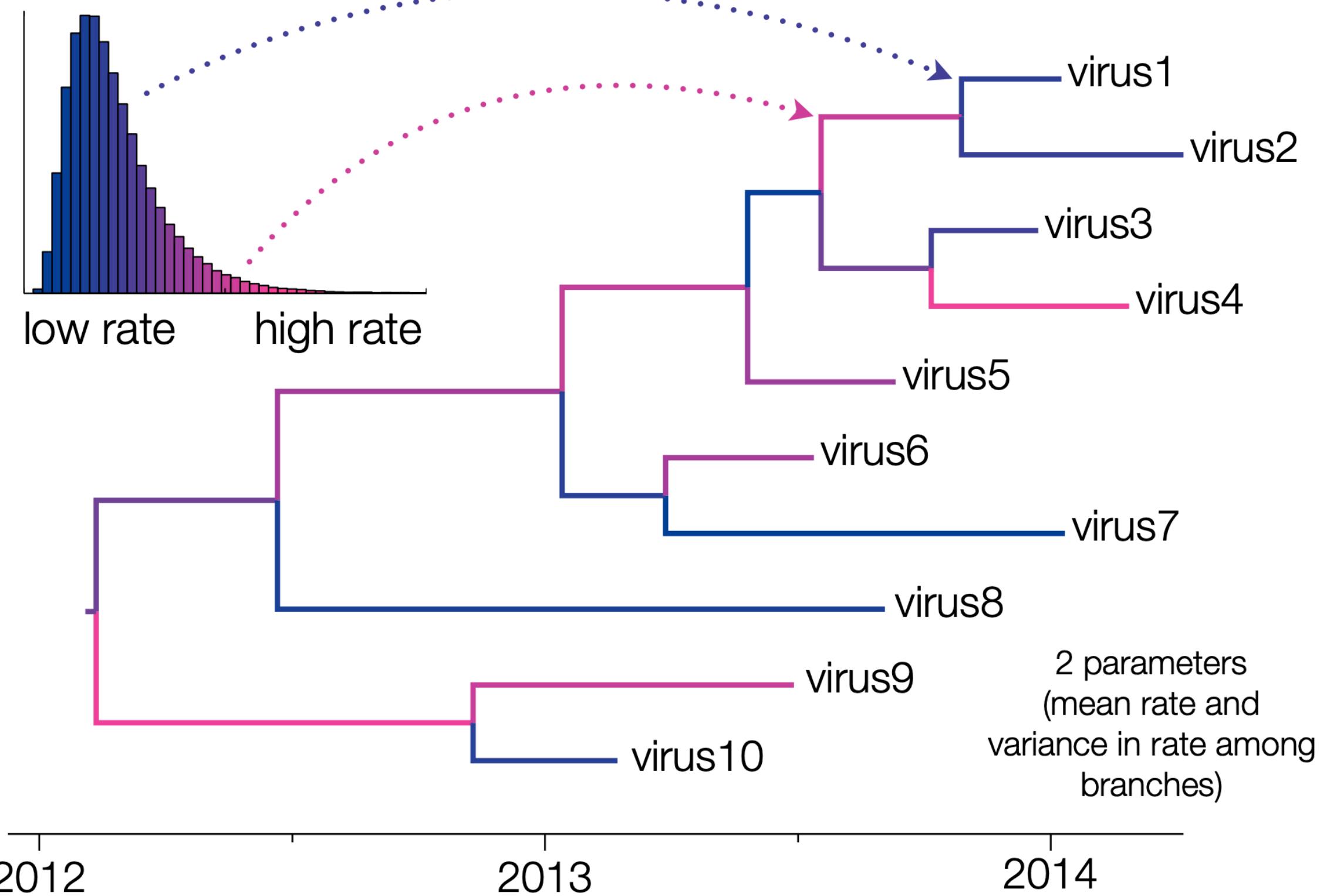


Relaxed Clocks

autocorrelated relaxed clock



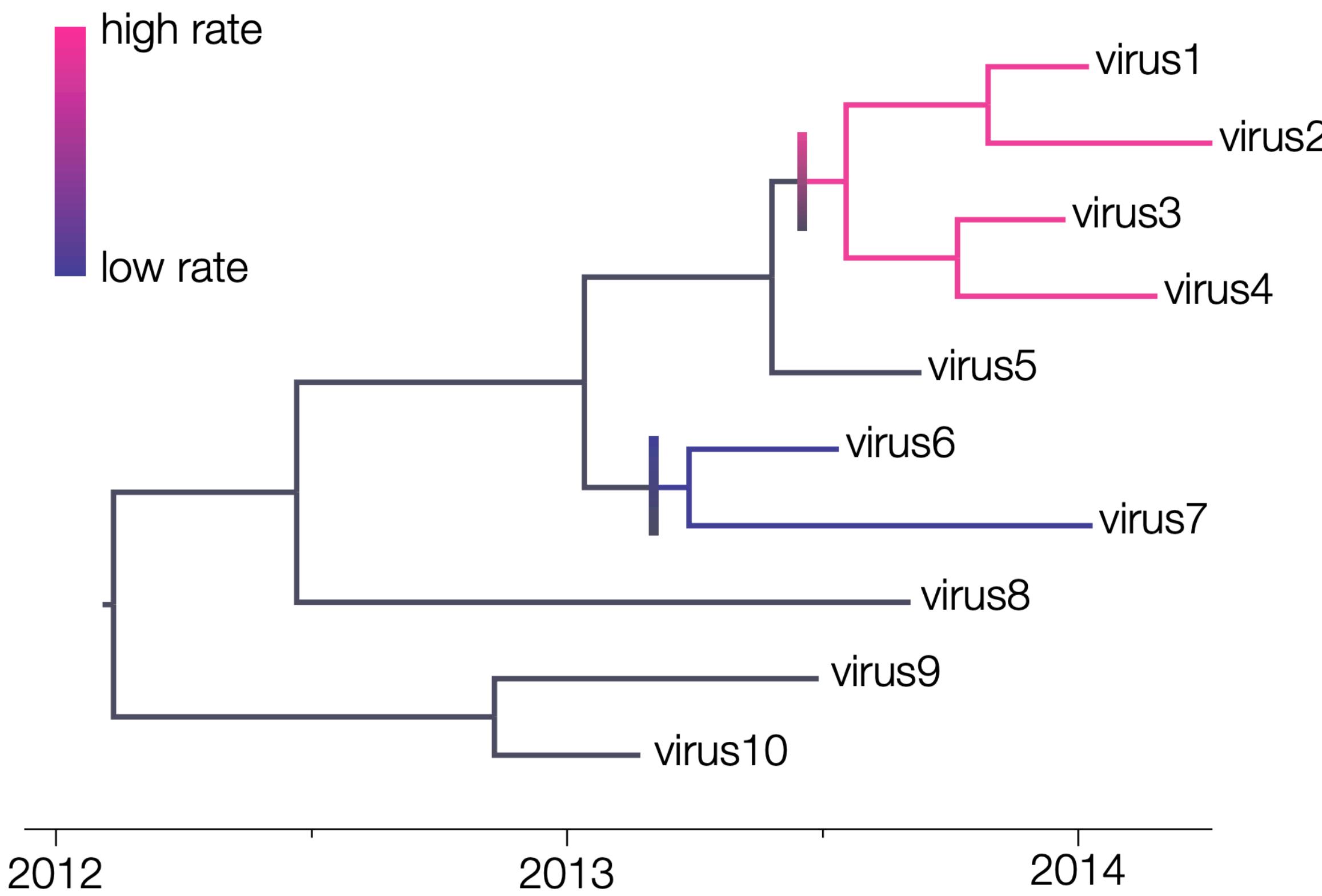
lognormal uncorrelated relaxed clock



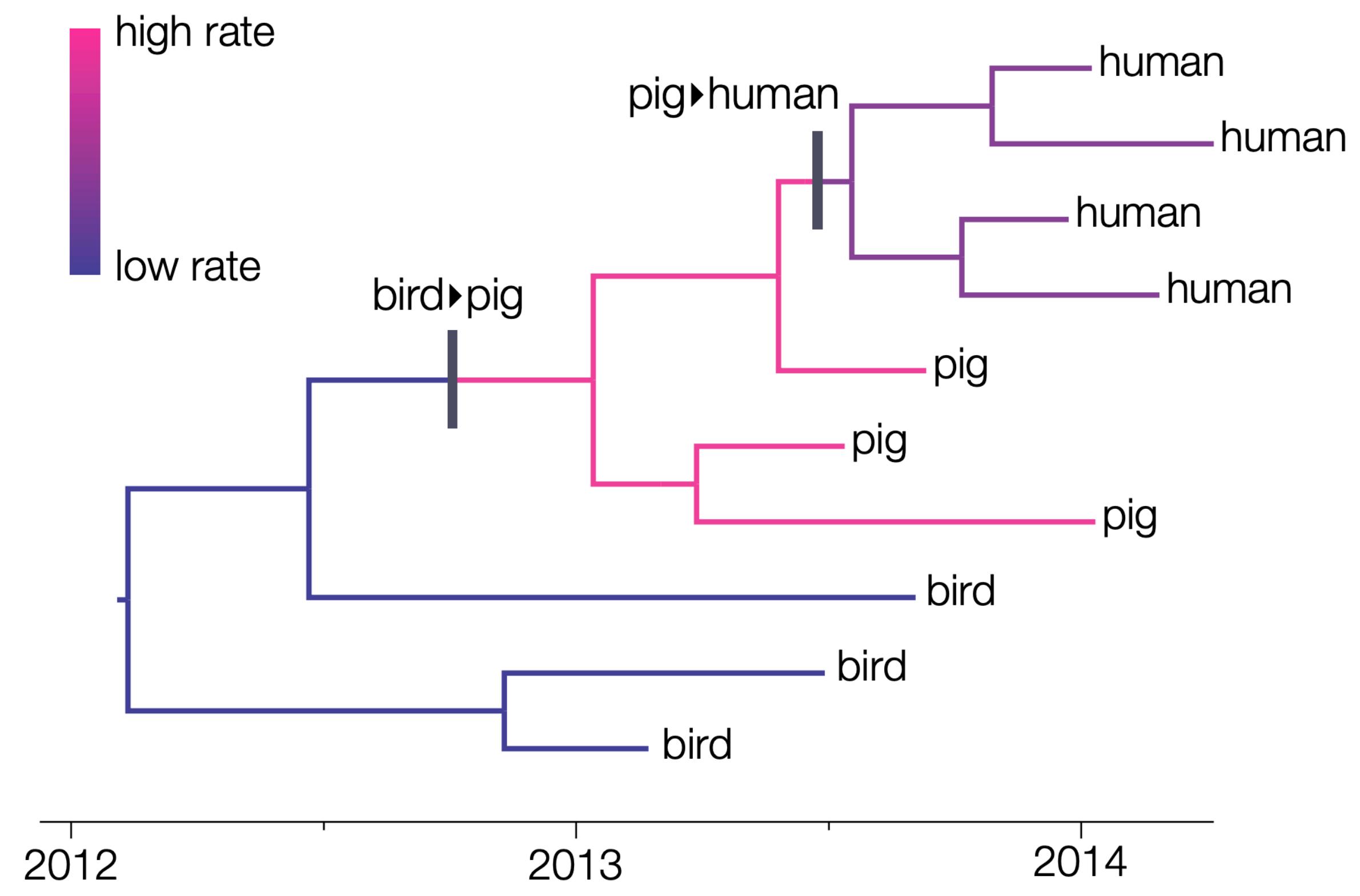
Adapted from Andrew Rambaut

Local clocks

'local' molecular clock



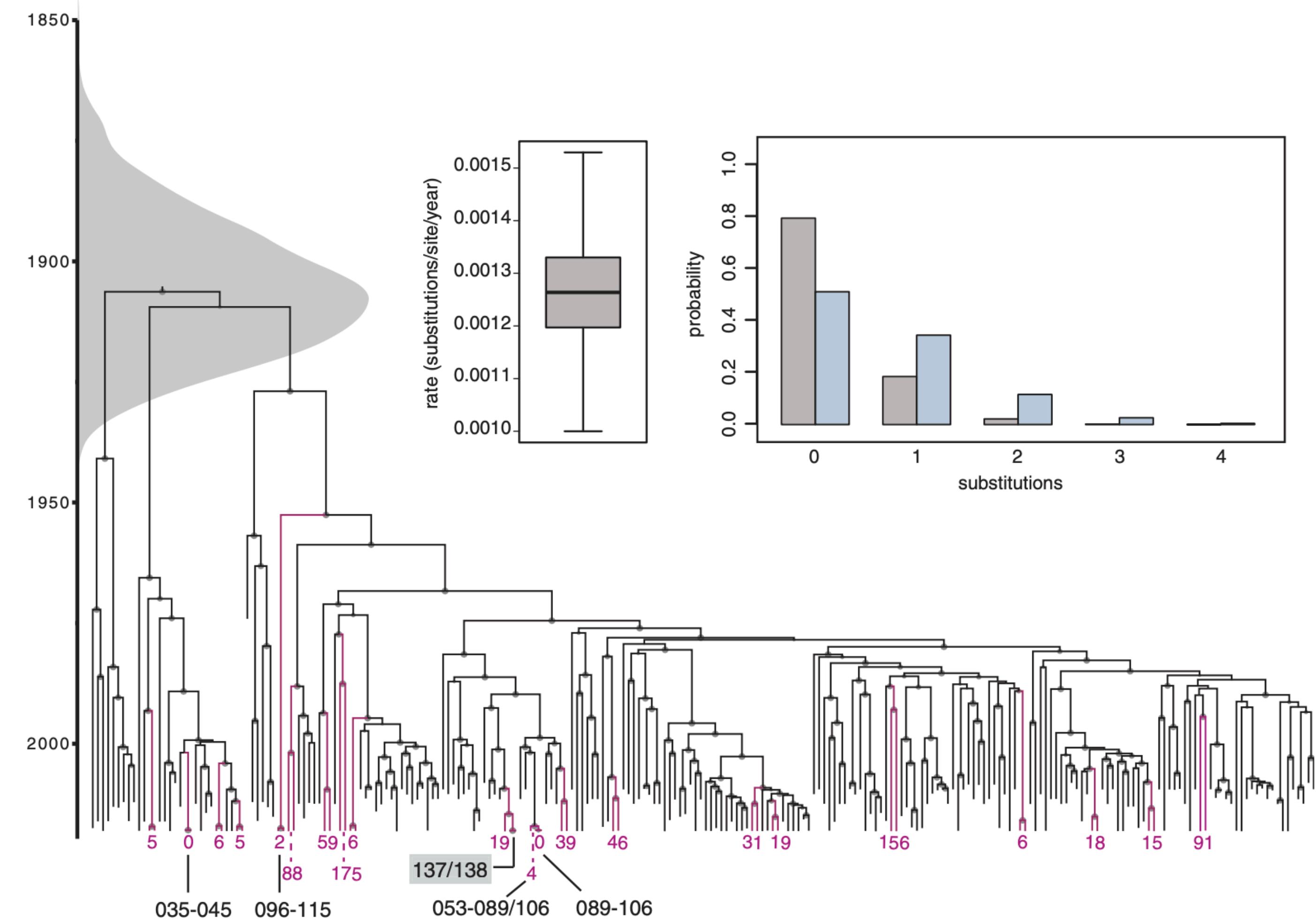
host specific local clock



Adapted from Andrew Rambaut

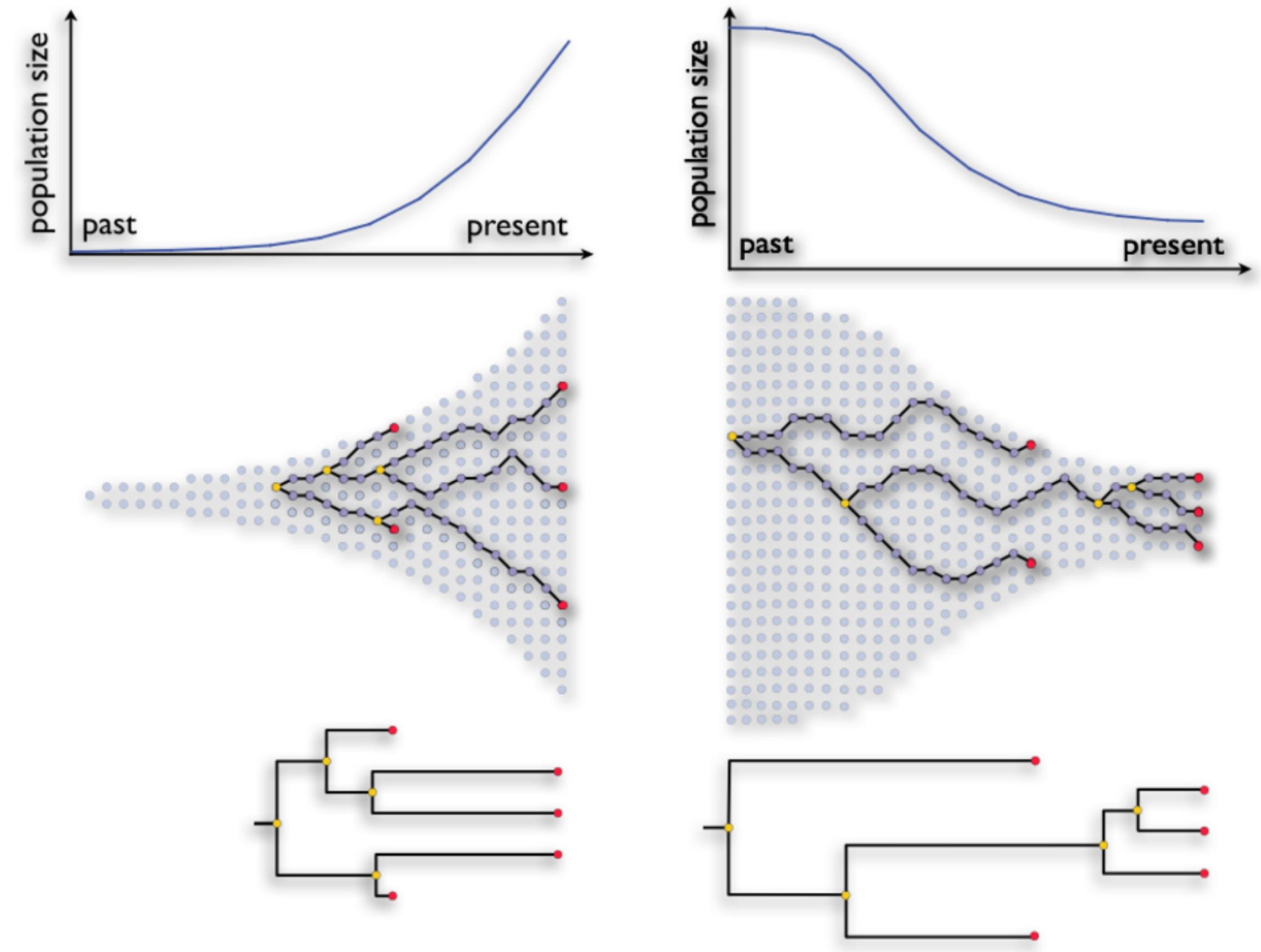
Time calibration applications

- Transforming branch lengths into units of time allow us to answer epidemiological questions such as :
 - When did the outbreak begin? (TMRCA)
 - How fast is the virus evolving? (Clock rate)
 - What is the mode of transmission?



Learning from trees: Coalescent

- Coalescent theory: Statistical model of the probability of two samples to merge (coalesce) after a given time
- We can learn about changes in population size from the shape of a tree:
 - The larger the population size, the longer it will take for two lineages to coalesce
 - Application: Calculate the exponential growth rate of an epidemic
 - How rapidly is the virus spreading? (Exponential growth)



Coalescent: Exponential Growth

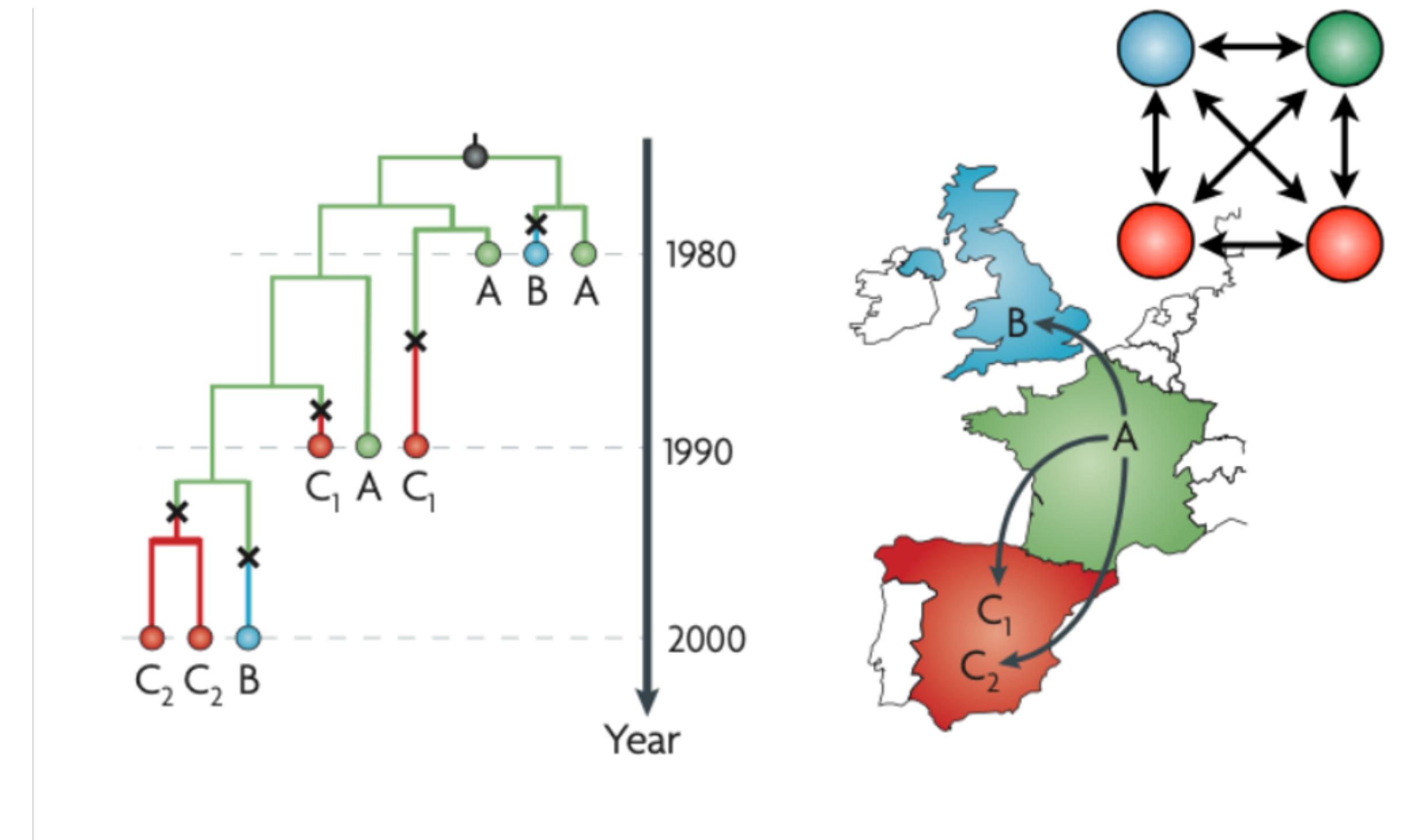
$$N_e(t) = N_0 e^{rt}$$

Effective population size Initial population size Growth rate

Doubling time: $N_e(t_{double}) = 2N_0 = N_0 e^{rt_{double}}$ $\xrightarrow{\text{Solve for t}}$ $t_{double} = \frac{\log(2)}{\text{Growth rate}}$

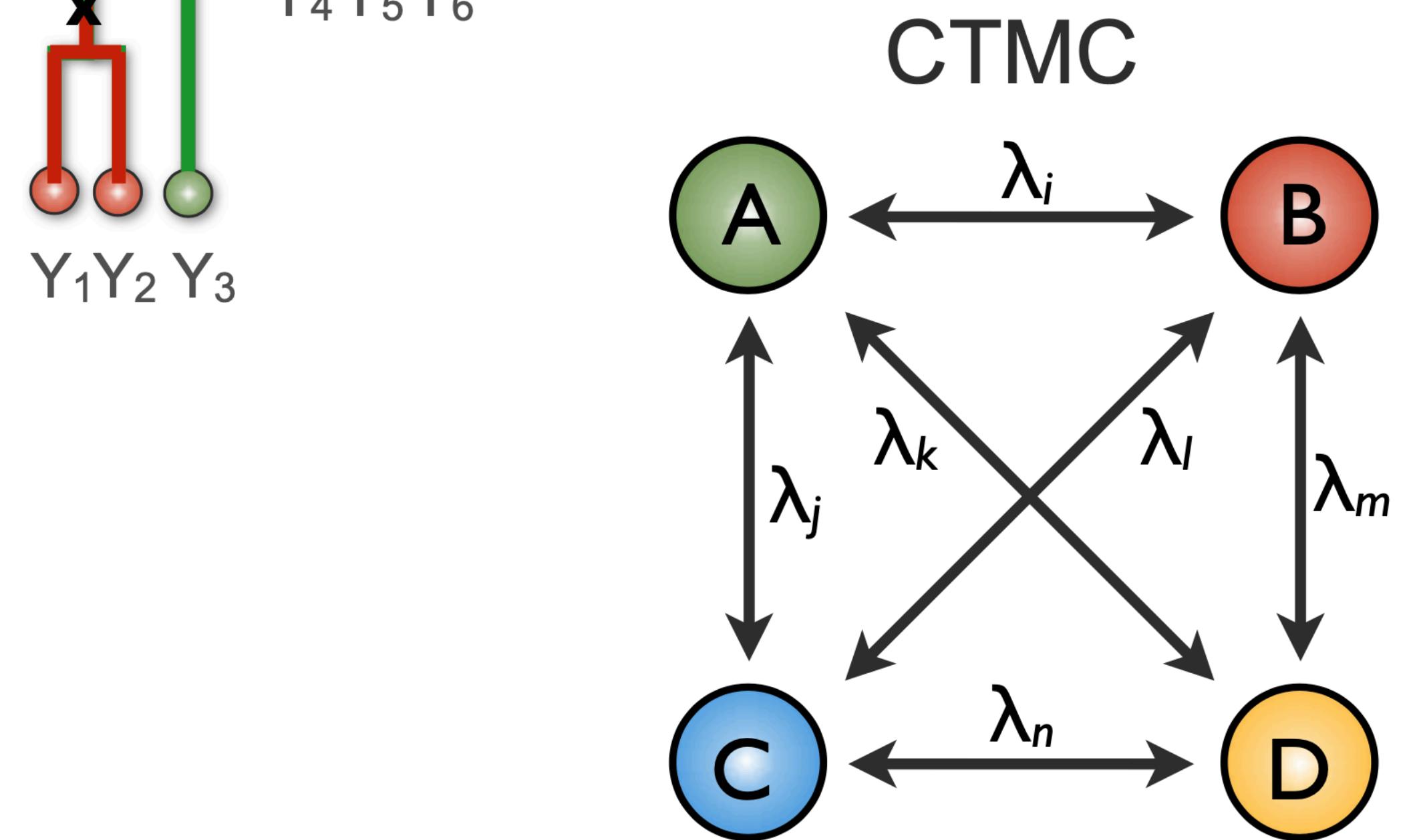
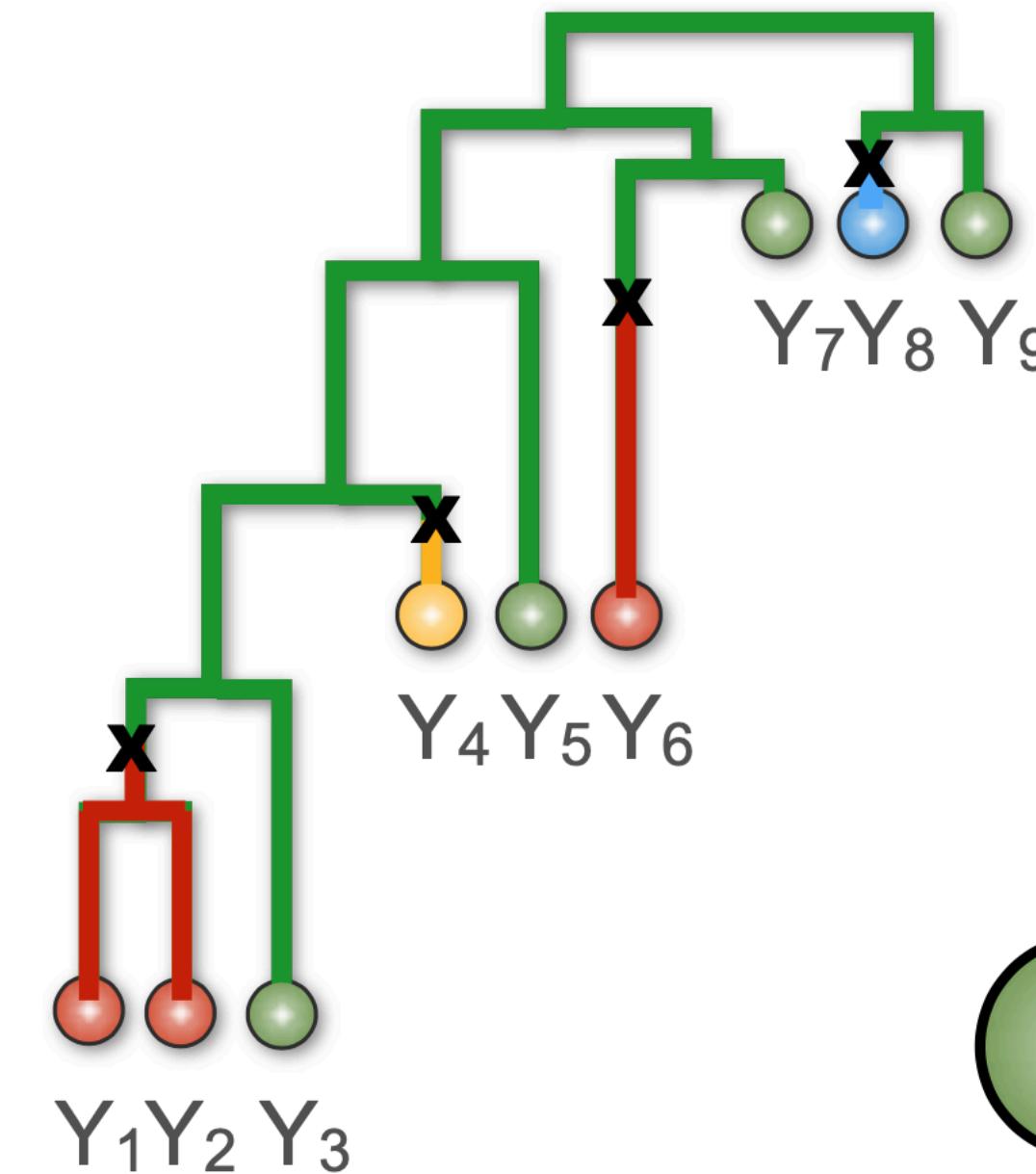
Learning from trees: Phylogeography

- Discrete trait analysis: model discrete locations as a trait coloring the tree
- Reconstruct probability of jumping from one location to another and expected number of jumps
- Answer questions such as:
 - How many introductions?
 - Where did the outbreak begin?
 - What factors drive spread?



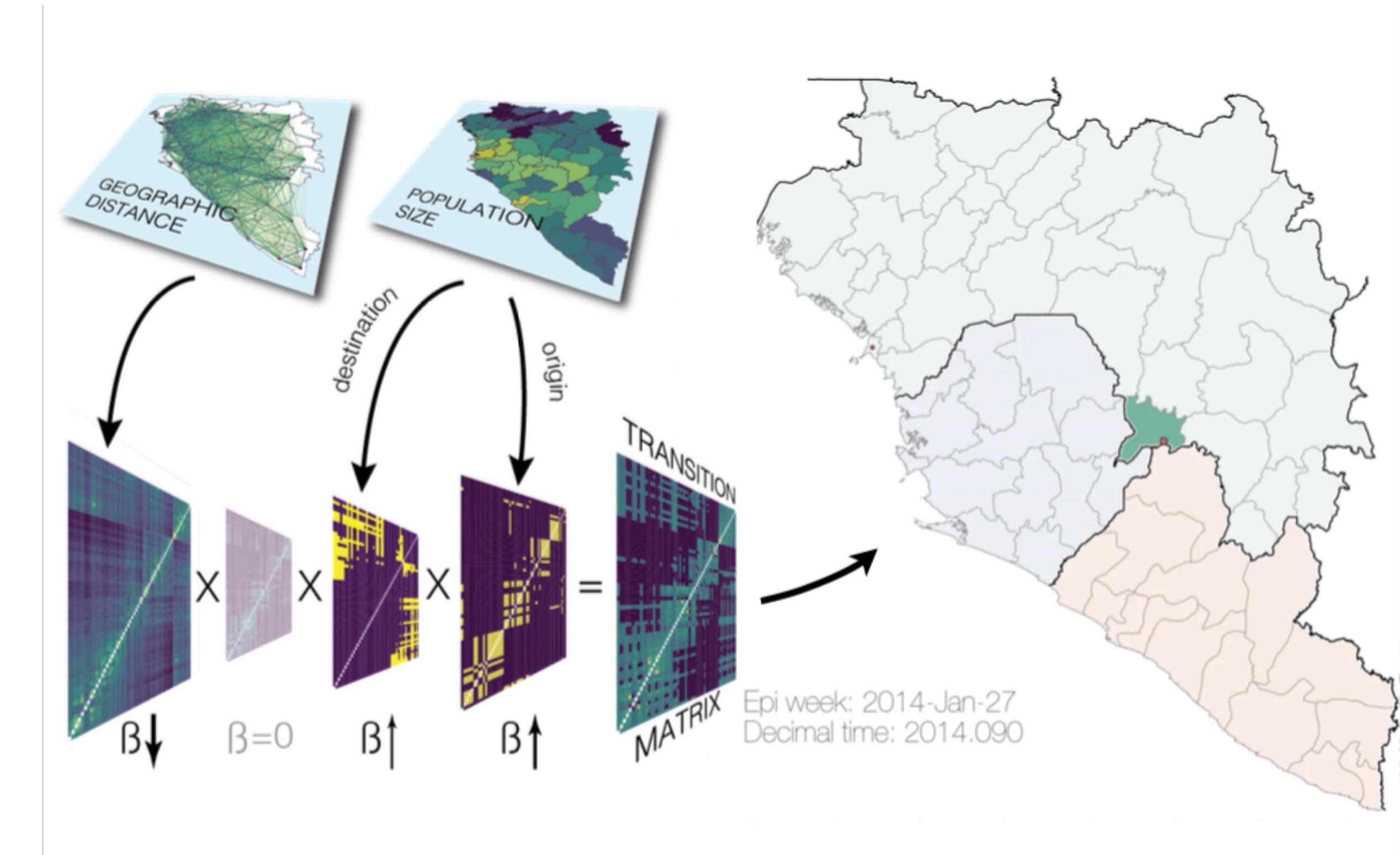
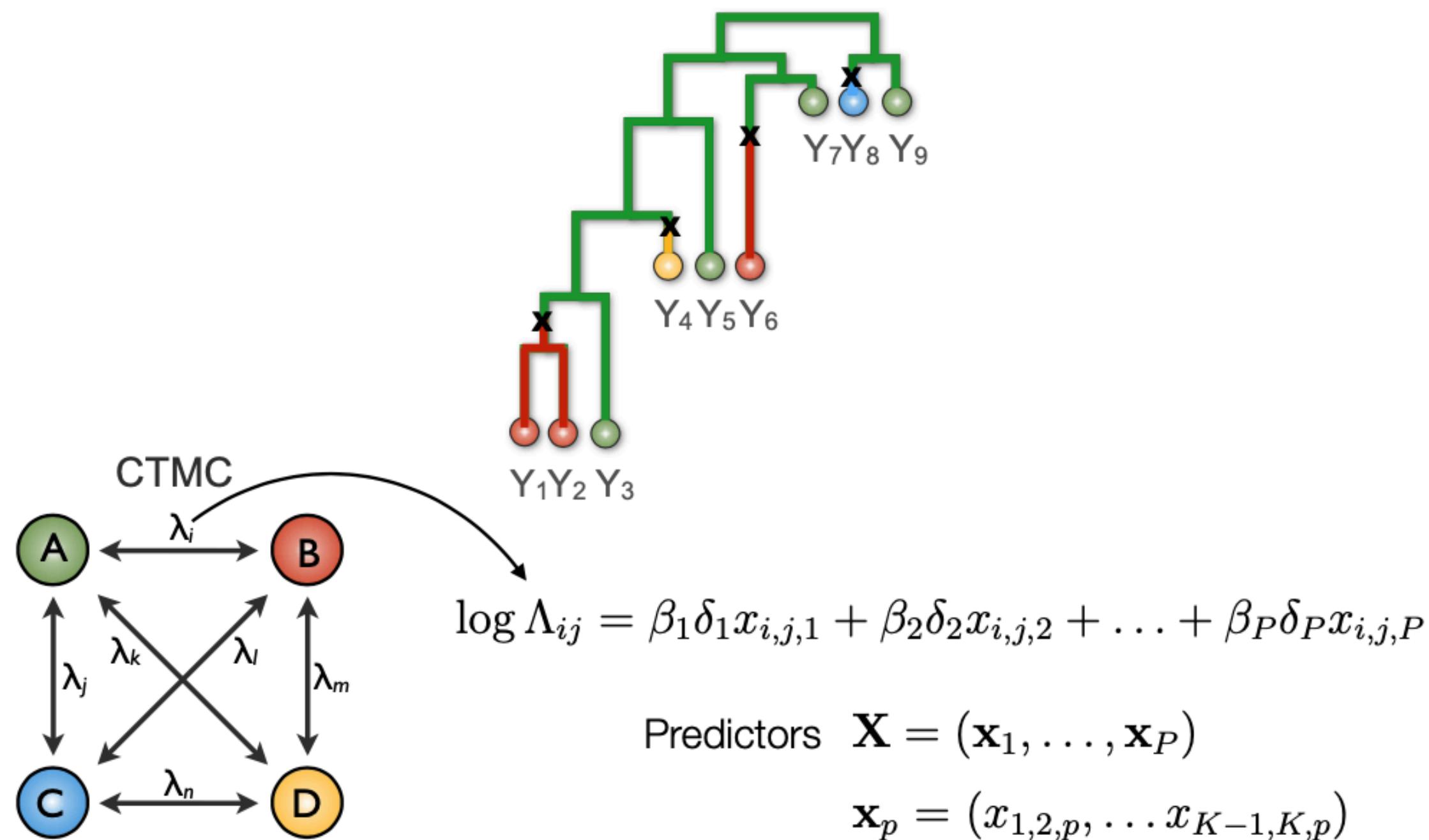
Discrete phylogeography

- Model spatial spread as a random process characterized by a rate matrix that describes the probability of transitioning to a location given the current location
 - Analogous to nucleotide substitution models
- Transition rates can be symmetric or asymmetric
- Use rates to reconstruct the location of internal nodes
 - How many introductions?
 - What is the origin?



Adapted from Philippe Lemey

Factors that drive spread?



Dudas et al. 2017

Adapted from Philippe Lemey

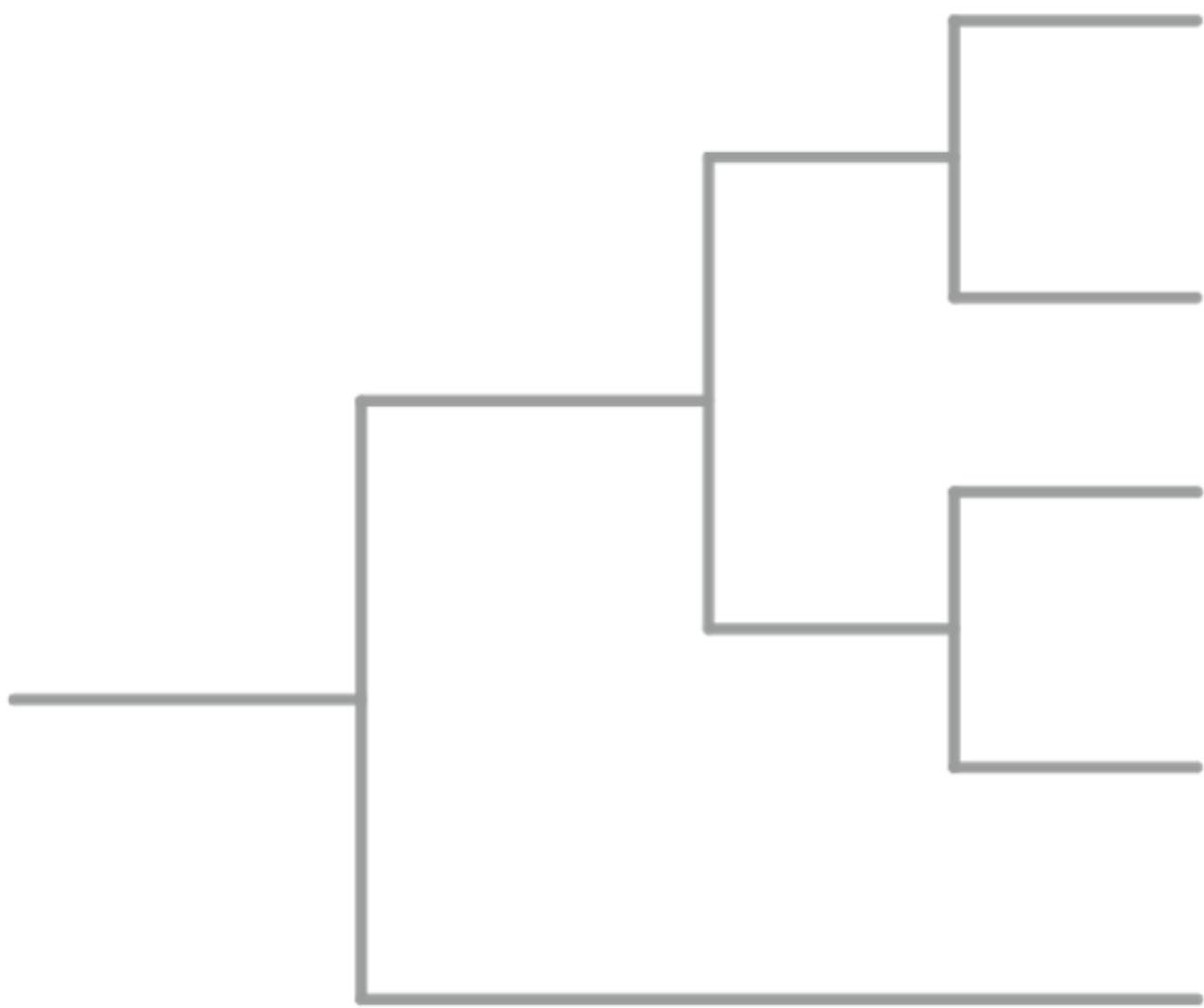
**How does Bayesian
Phylogenetic Inference work?**

Bayesian Phylogenetics

Maximum Likelihood

$$\operatorname{argmax}_{\theta} L(\theta)$$

Likelihood: $L(\theta) \propto P(Data|\theta)$

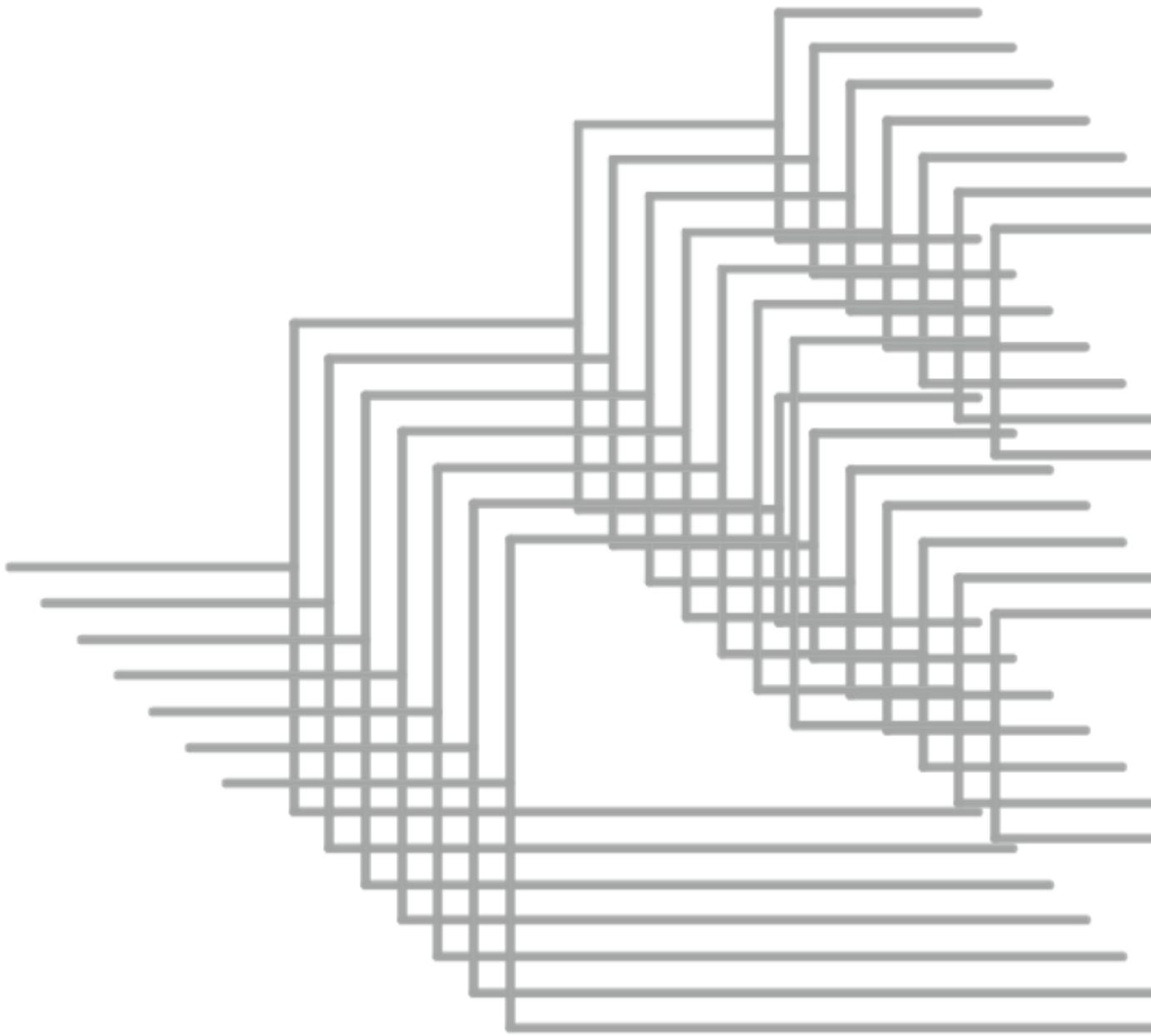


Posterior

Bayesian

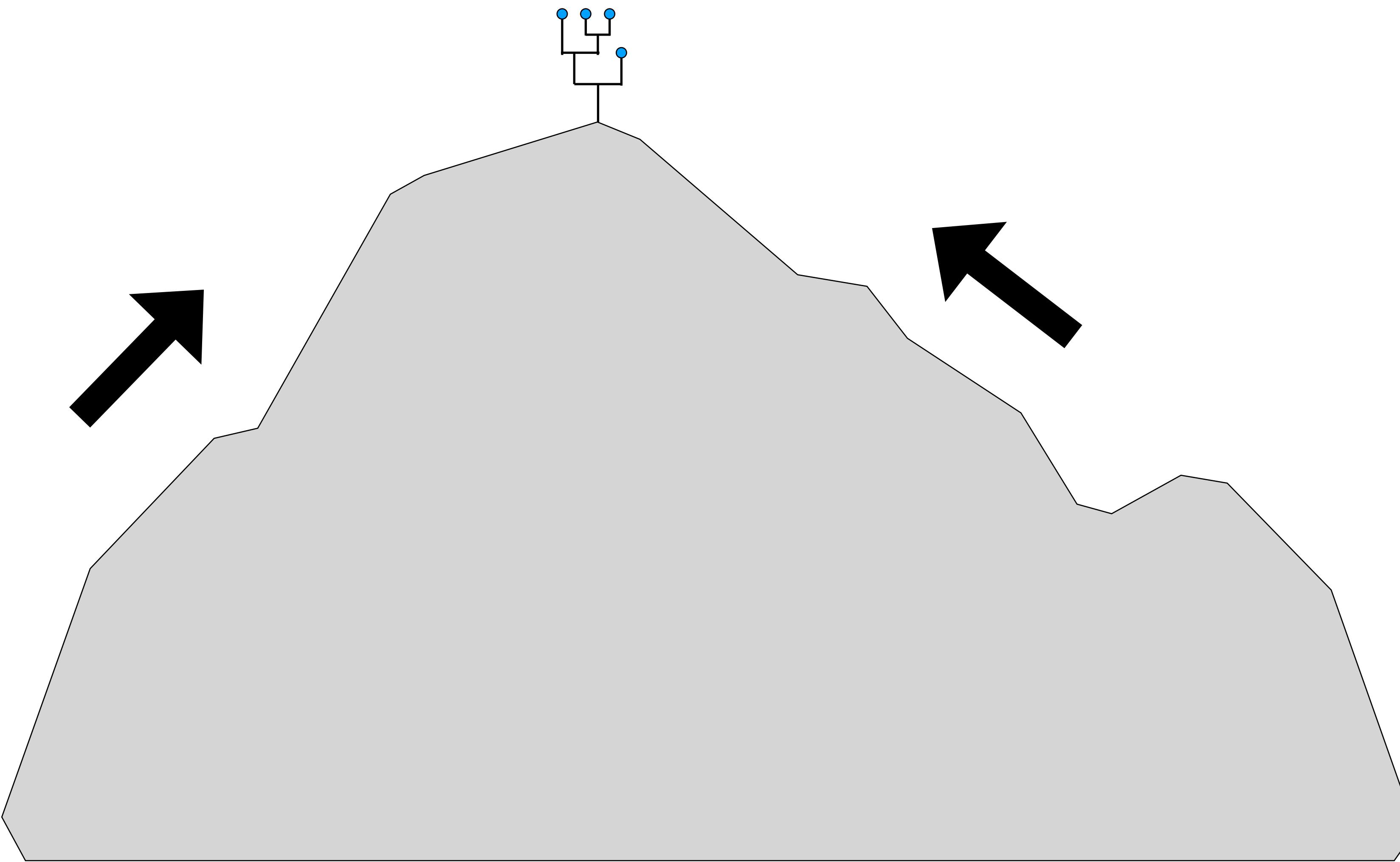
$$P(\theta|Data) = \frac{P(Data|\theta)P(\theta)}{P(Data)}$$

Prior





Maximum Likelihood



Bayesian

