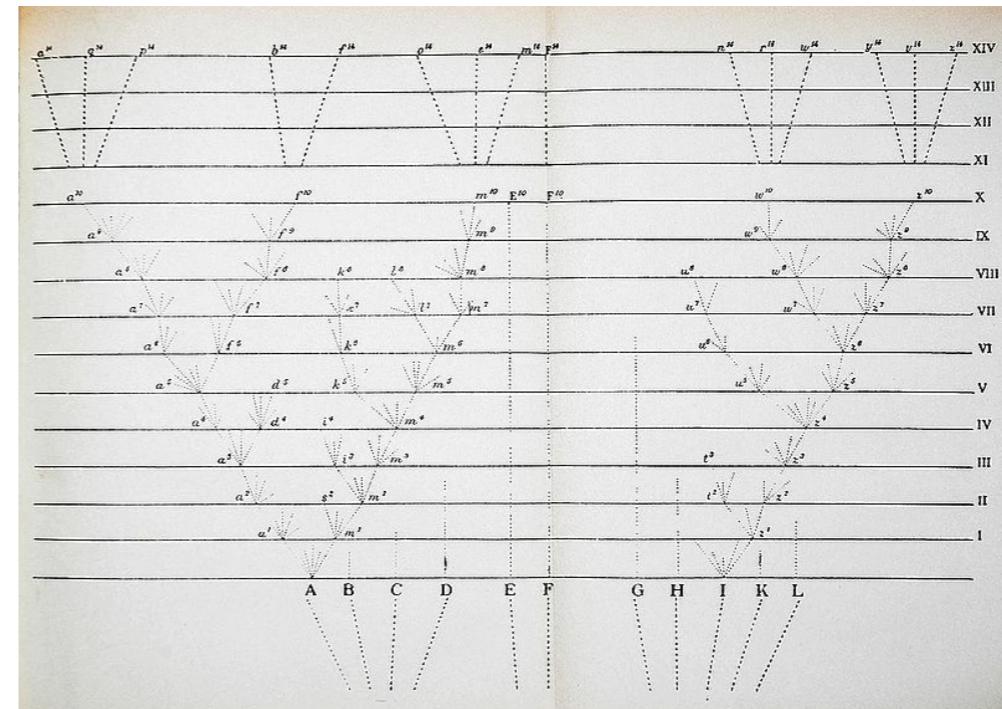
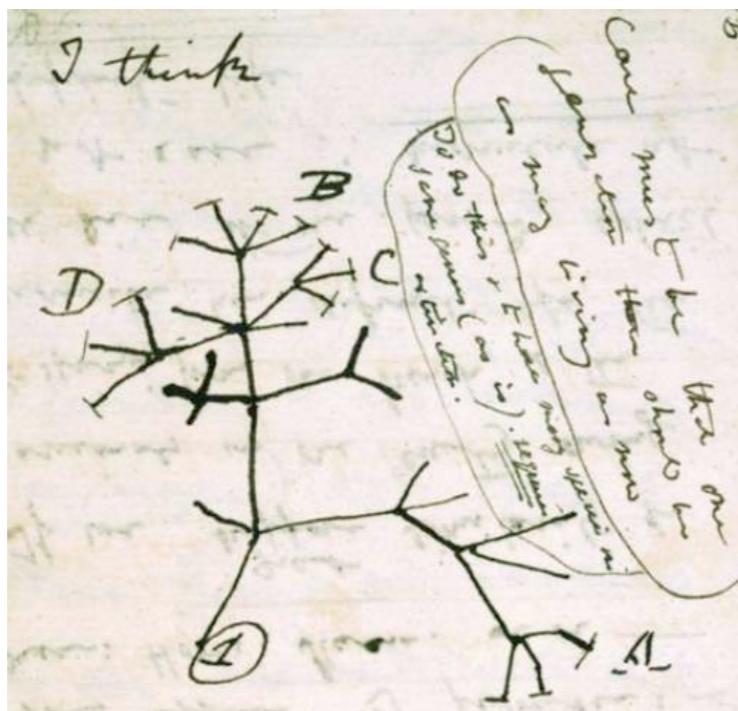


# Taming the BEAST

## Introduction to phylodynamic models

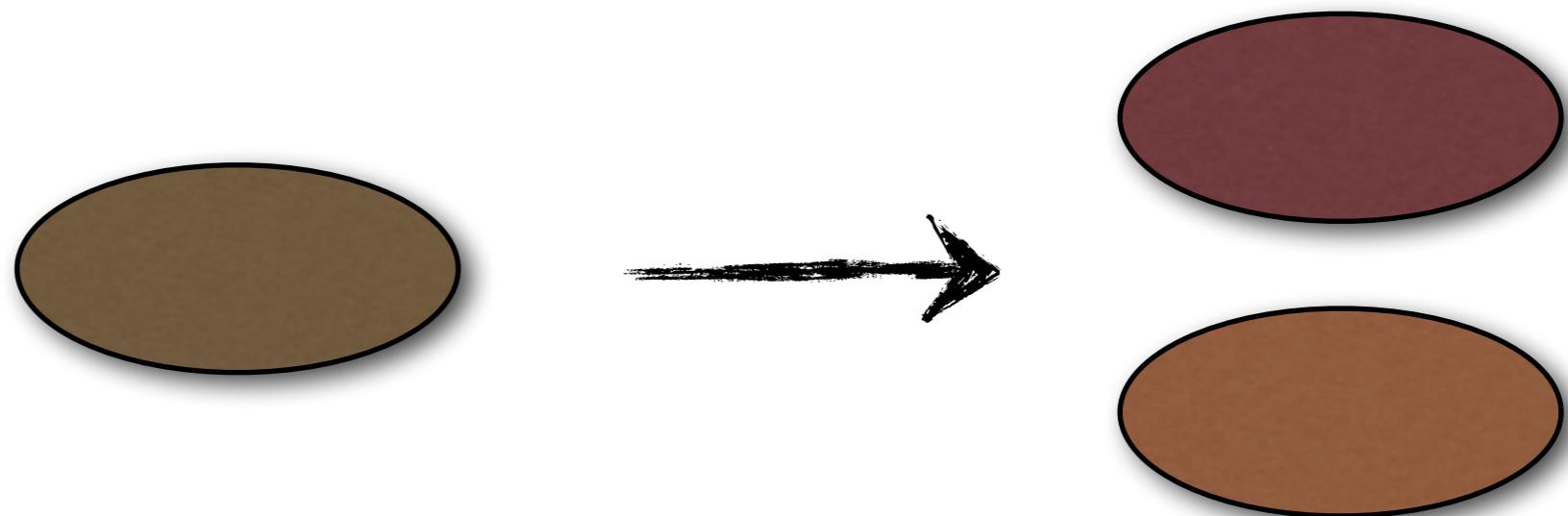


# Basis of biology: Reproduction & genetic change

---

## **Scale:**

DNA  
Virion  
Prokaryote  
Eukaryotic cell  
Eukaryote  
Species  
Infected host  
Immune cells

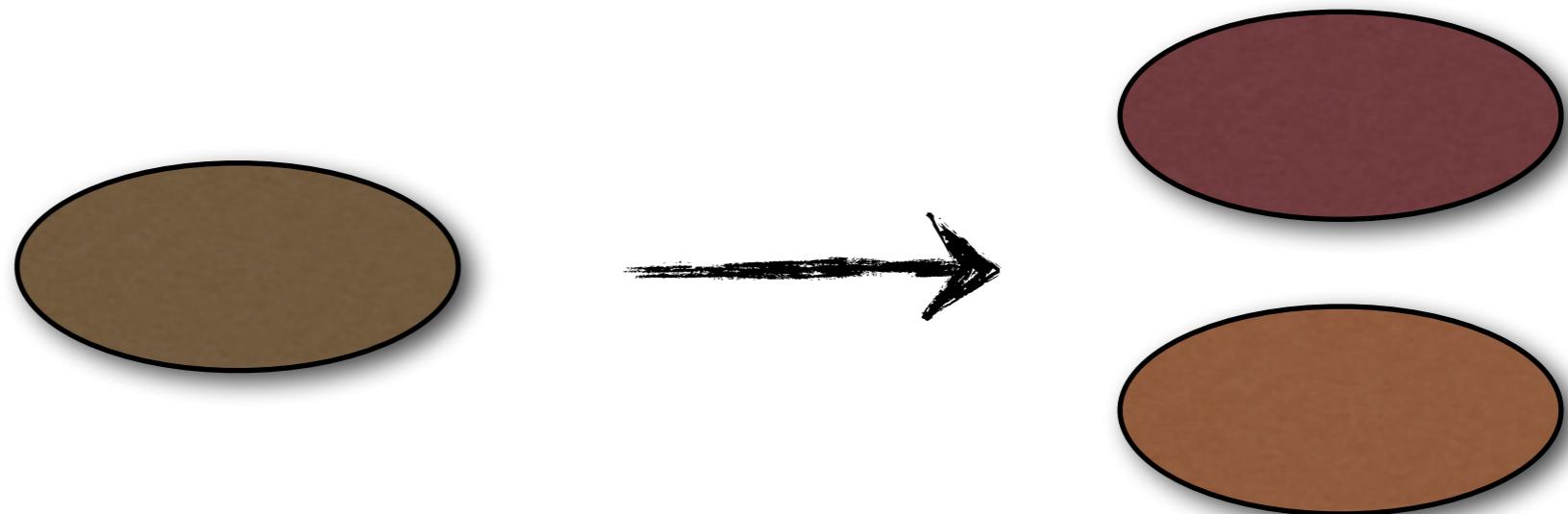


# Basis of biology: Reproduction & genetic change

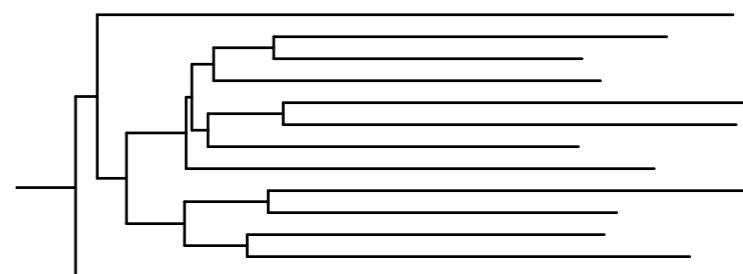
---

## Scale:

DNA  
Virion  
Prokaryote  
Eukaryotic cell  
Eukaryote  
Species  
Infected host  
Immune cells



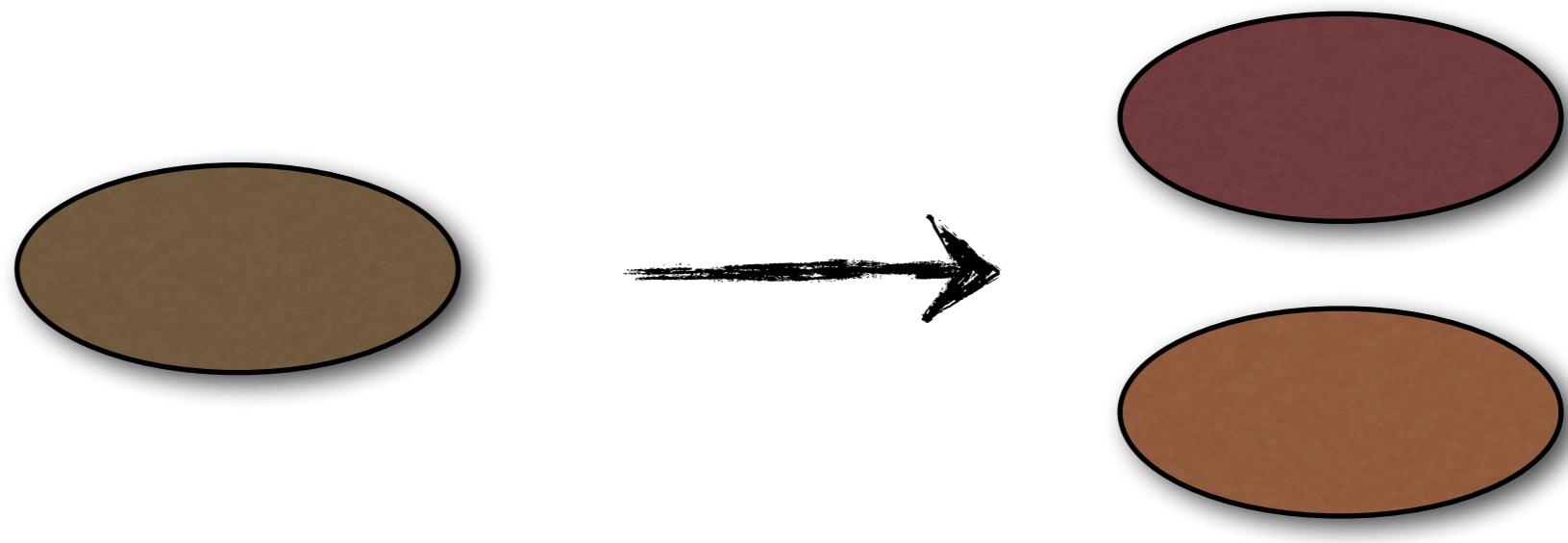
- ▶ Lineage tree (=phylogeny) represents **reproduction** process on any of these scales
- ▶ **Genetic** information changes along the tree



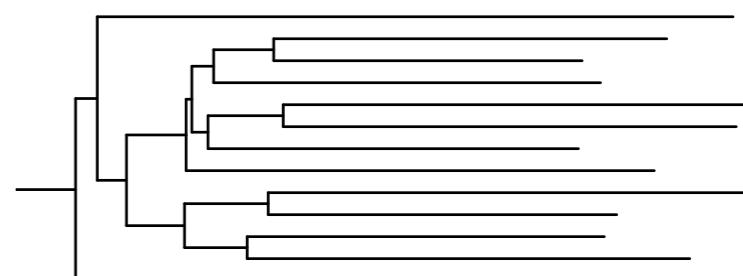
# Basis of biology: Reproduction & genetic change

## Scale:

DNA  
Virion  
Prokaryote  
Eukaryotic cell  
~~Eukaryote~~  
Species  
Infected host  
Immune cells



- ▶ Lineage tree (=phylogeny) represents **reproduction** process on any of these scales
- ▶ **Genetic** information changes along the tree



# Outline

---

- Why phylodynamics?
  - ▶ A macroevolutionary motivation
  - ▶ An epidemiological motivation
- A phylodynamic model and the math behind
  - ▶ Birth-death models
  - ▶ The coalescent
- Applications in Macroevolution
- Application in Epidemiology: SARS-CoV-2
- BEAST2 Tutorial

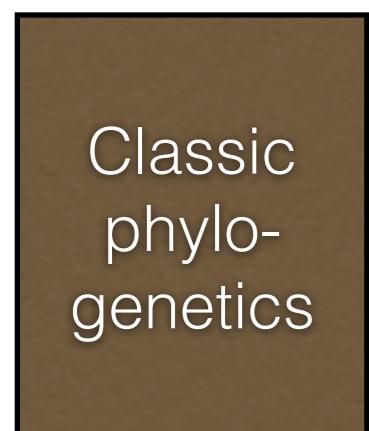
# Outline

---

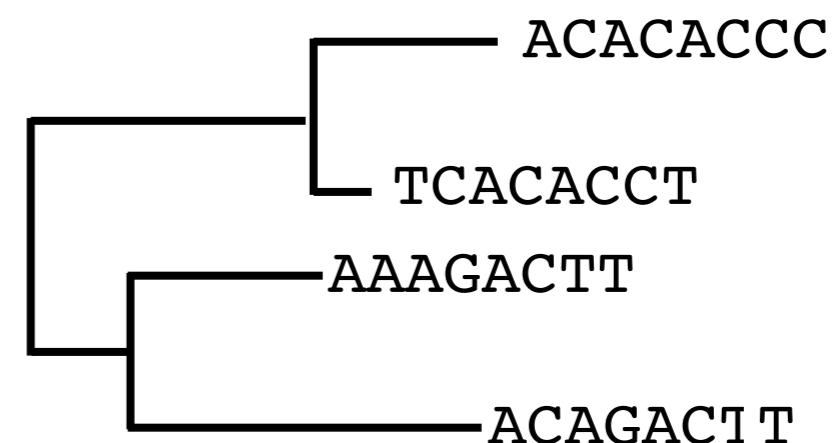
- Why phylodynamics?
  - ▶ A macroevolutionary motivation
  - ▶ An epidemiological motivation
- A phylodynamic model and the math behind
  - ▶ Birth-death models
  - ▶ The coalescent
- Applications in Macroevolution
- Application in Epidemiology: SARS-CoV-2
- BEAST2 Tutorial

# Phylogenetic tree inference

---



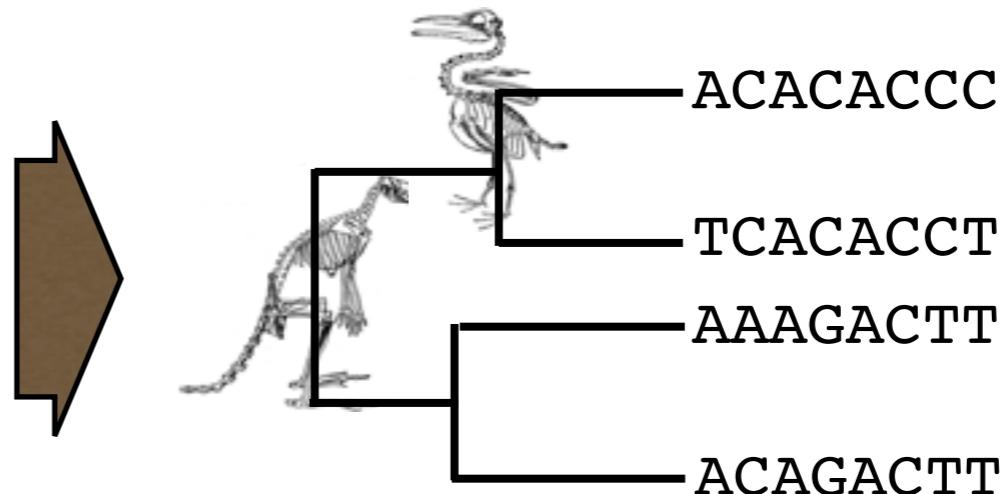
Species1 ACACACCC  
Species2 TCACACCT  
Species3 AAAGACTT  
Species4 ACAGACTT



# Phylogenetic tree inference

Classic phylogenetics

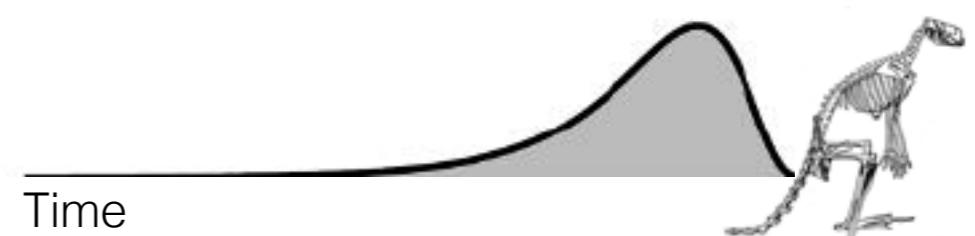
Species1 ACACACCC  
Species2 TCACACCT  
Species3 AAAGACTT  
Species4 ACAGACTT



Fossil calibration

Prior for each fossil calibration point (node dating)

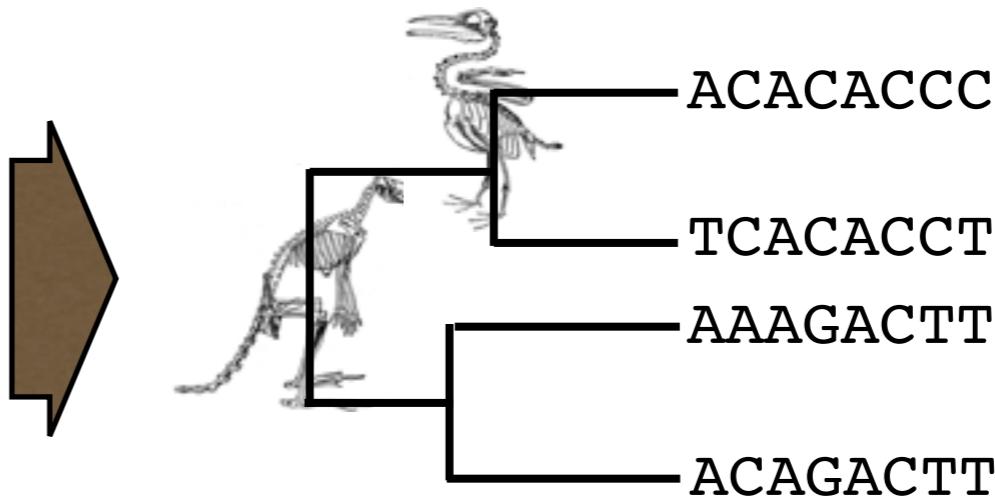
- ▶ Yang & Rannala (*MBE*, 2006)
- ▶ Heath (*Syst. Biol.*, 2010)



# Phylogenetic tree inference

Classic phylogenetics

Species1 ACACACCC  
Species2 TCACACCT  
Species3 AAAGACTT  
Species4 ACAGACTT



Fossil calibration

Prior for each fossil calibration point (node dating)

- ▶ Yang & Rannala (*MBE*, 2006)
- ▶ Heath (*Syst. Biol.*, 2010)



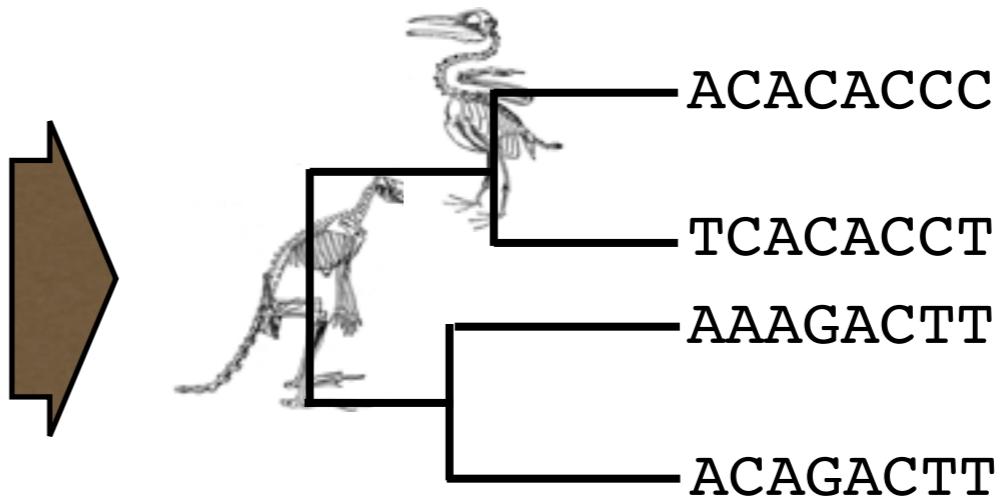
Limitation

- ▶ Prior for each node to define when **estimating divergence times** (ad hoc?)
- ▶ **Estimating speciation and extinction rates** cannot be done directly from trees with fossils

# Phylogenetic tree inference

Classic phylogenetics

Species1 ACACACCC  
Species2 TCACACCT  
Species3 AAAGACTT  
Species4 ACAGACTT



Fossil calibration

Prior for each fossil calibration point (node dating)

- ▶ Yang & Rannala (*MBE*, 2006)
- ▶ Heath (*Syst. Biol.*, 2010)



Limitation

- ▶ Prior for each node to date (ad hoc?)
- ▶ **Estimating speciation rates** directly from trees with fossils

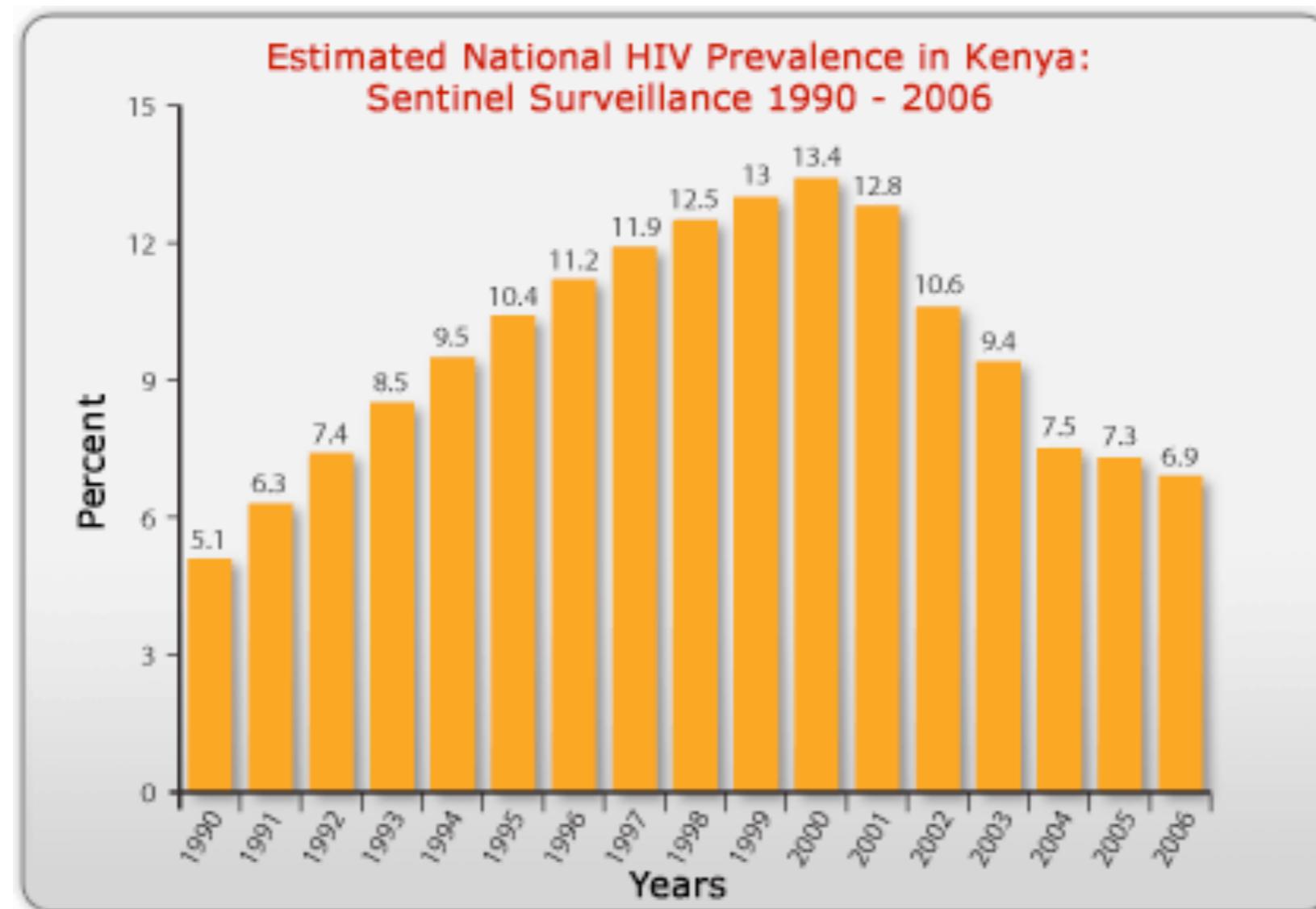
**Phylodynamic approach aims to overcome these limitations!**

# Outline

---

- Why phylodynamics?
  - ▶ A macroevolutionary motivation
  - ▶ An epidemiological motivation
- A phylodynamic model and the math behind
  - ▶ Birth-death models
  - ▶ The coalescent
- Applications in Macroevolution
- Application in Epidemiology: SARS-CoV-2
- BEAST2 Tutorial

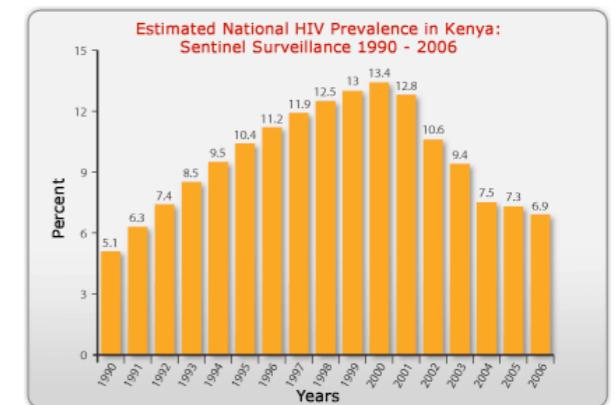
# Phylogenetics in epidemiology



# Phylogenetics in epidemiology

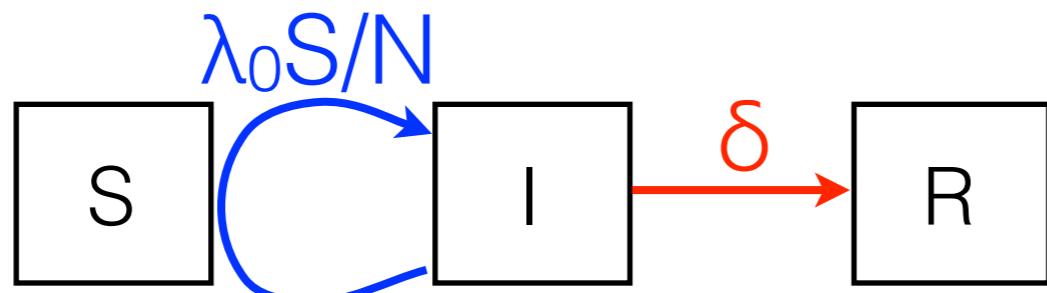
Data

# of infected individuals through time



Models

Population dynamics described by SIR models:

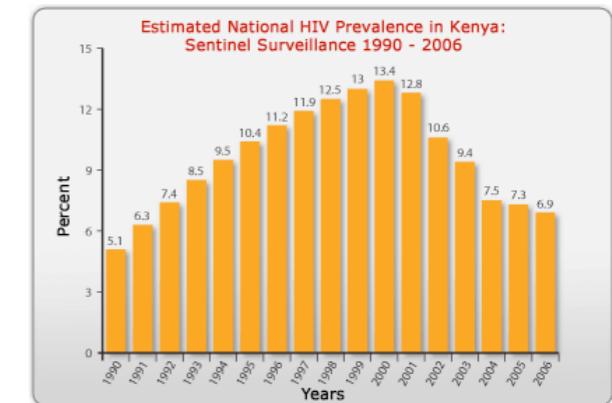


$$\begin{aligned}\frac{dS}{dt} &= -\lambda_0 IS/N \\ \frac{dI}{dt} &= \lambda_0 IS/N - \delta I \\ \frac{dR}{dt} &= \delta I\end{aligned}$$

# Phylogenetics in epidemiology

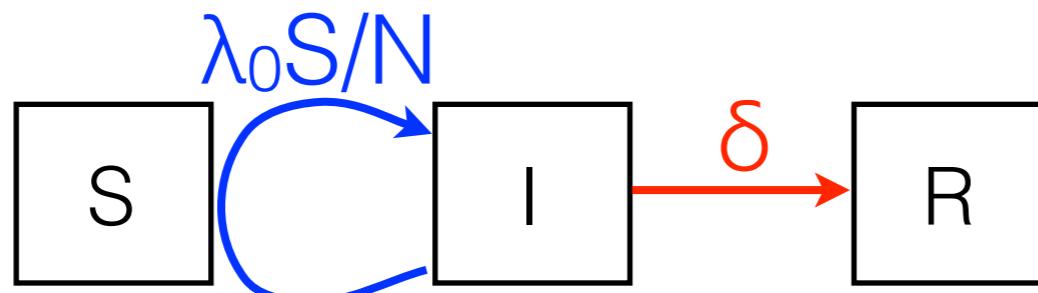
Data

# of infected individuals through time



Models

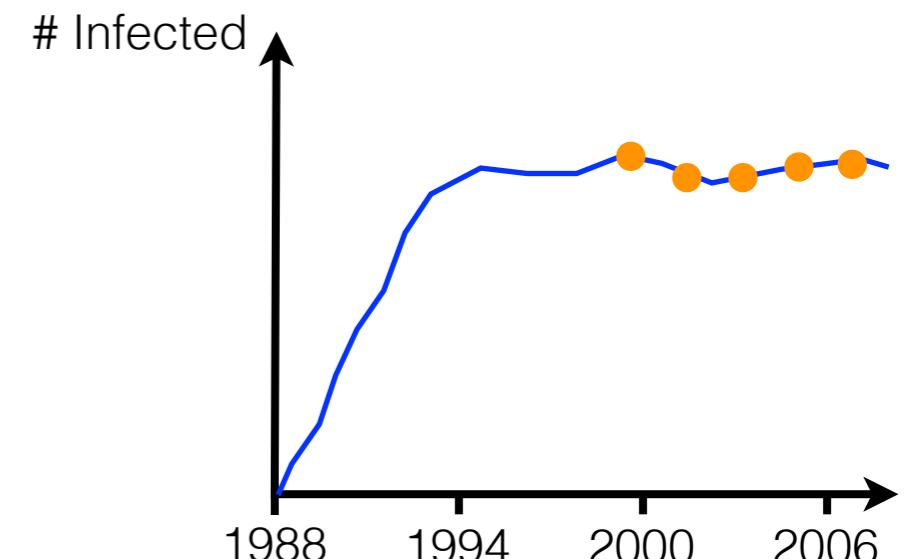
Population dynamics described by SIR models:



$$\begin{aligned}\frac{dS}{dt} &= -\lambda_0 IS/N \\ \frac{dI}{dt} &= \lambda_0 IS/N - \delta I \\ \frac{dR}{dt} &= \delta I\end{aligned}$$

Limitations

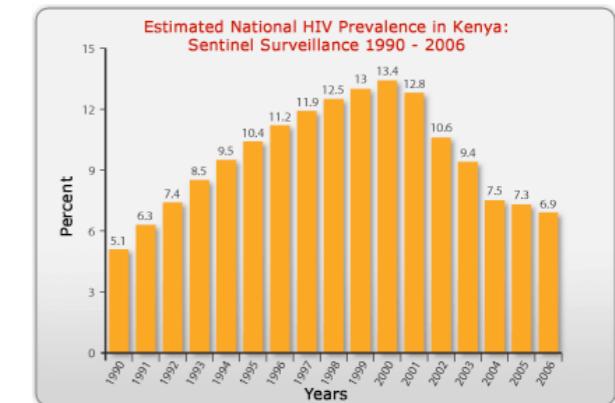
If sampling in early epidemic was missed:



# Phylogenetics in epidemiology

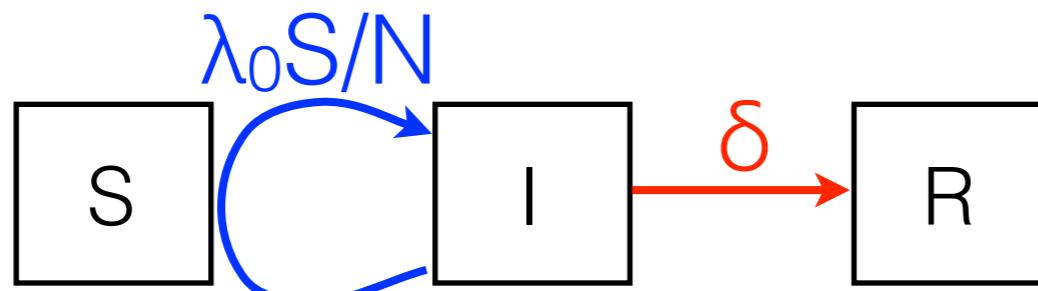
Data

# of infected individuals through time



Models

Population dynamics described by SIR models:

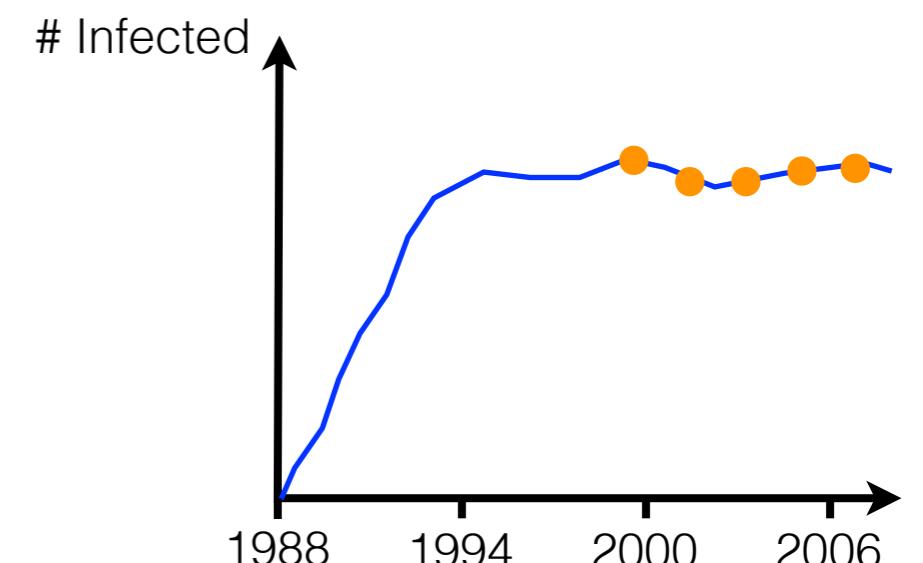


$$\begin{aligned}\frac{dS}{dt} &= -\lambda_0 IS/N \\ \frac{dI}{dt} &= \lambda_0 IS/N - \delta I \\ \frac{dR}{dt} &= \delta I\end{aligned}$$

Limitations

If sampling in early epidemic was missed:

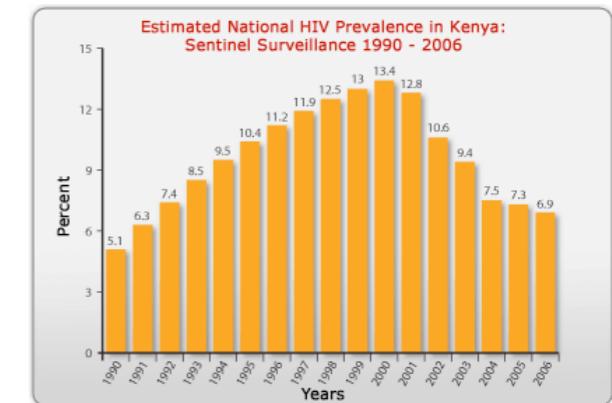
- ▶ **Time of epidemic outbreak?**
- ▶ **Basic reproductive number  $R_0$ ?**



# Phylogenetics in epidemiology

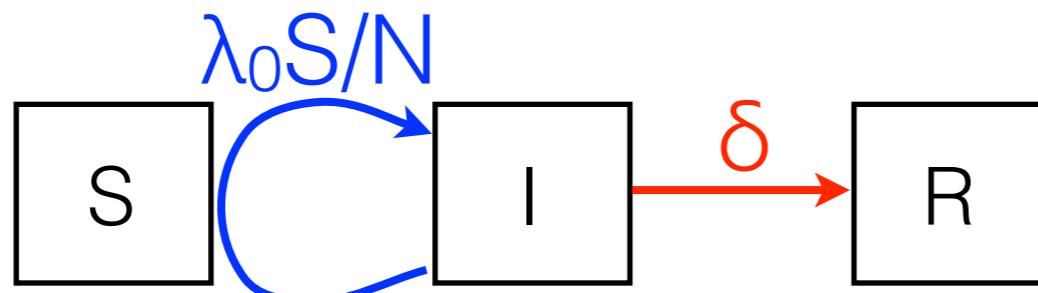
Data

# of infected individuals through time



Models

Population dynamics described by SIR models:



$$\begin{aligned}\frac{dS}{dt} &= -\lambda_0 IS/N \\ \frac{dI}{dt} &= \lambda_0 IS/N - \delta I \\ \frac{dR}{dt} &= \delta I\end{aligned}$$

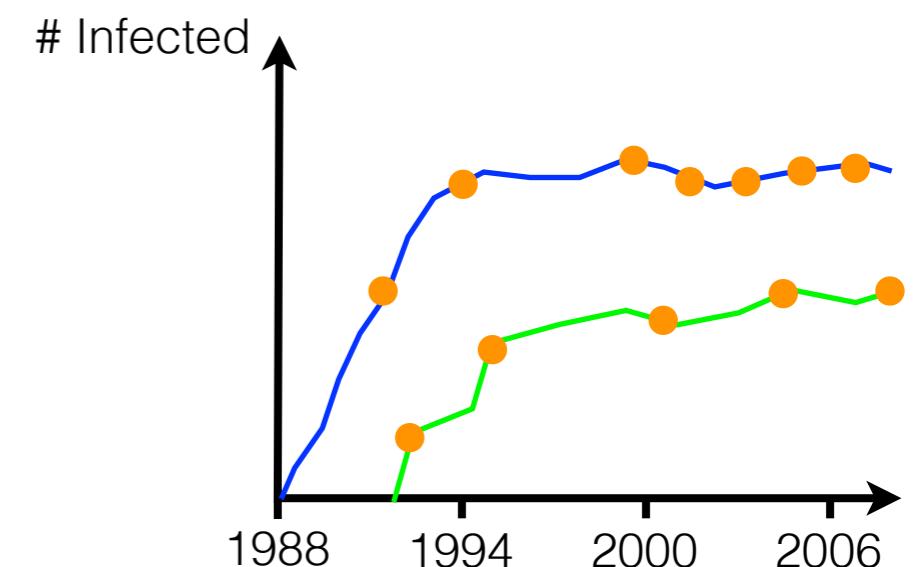
Limitations

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?**



# Genetic data entering epidemiology

---

Phylo-  
genetics

Inf1 ACACACCC  
Inf2 TCACACCT  
Inf3 AAAGACTT  
Inf4 ACAGACTT

# Genetic data entering epidemiology

---

Phylo-  
genetics

## Input:

Pathogen genetic data from different hosts

Inf1	<b>ACACACCC</b>
Inf2	<b>TCACACCT</b>
Inf3	<b>AAAGACTT</b>
Inf4	<b>ACAGACTT</b>

# Genetic data entering epidemiology

---

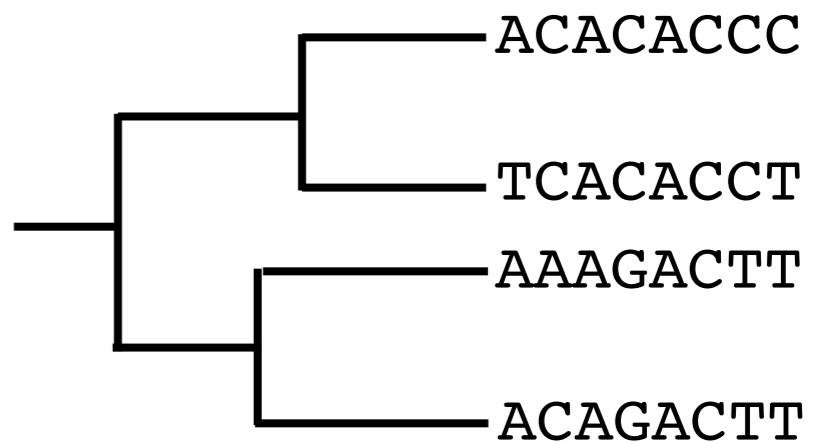
Phylogenetics

Input:

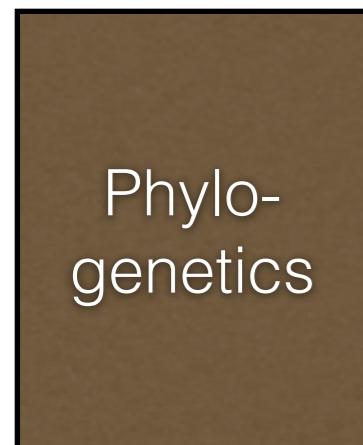
Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)



# Genetic data entering epidemiology

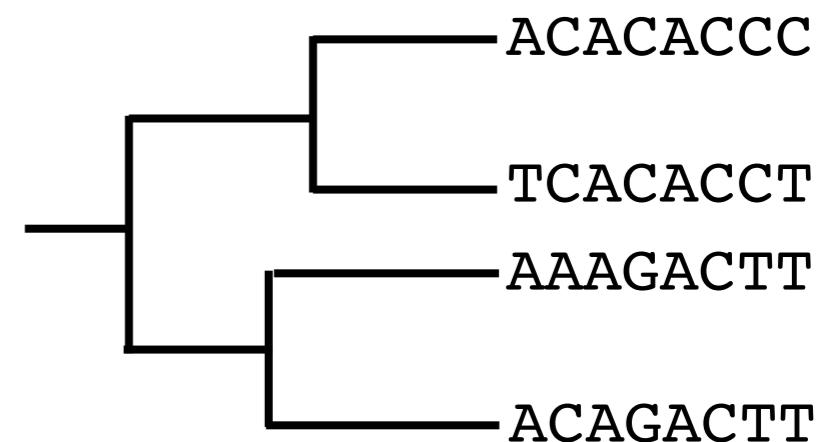


Input:

Pathogen genetic data from different hosts

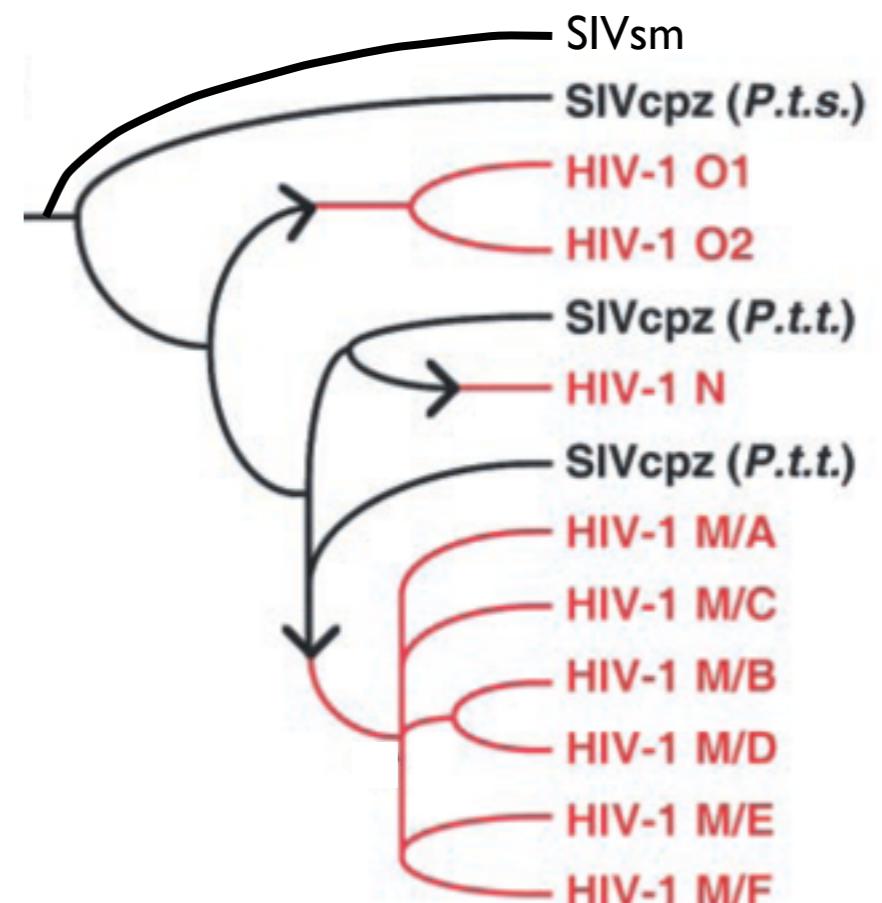
Output:

Transmission chain (who infected whom)



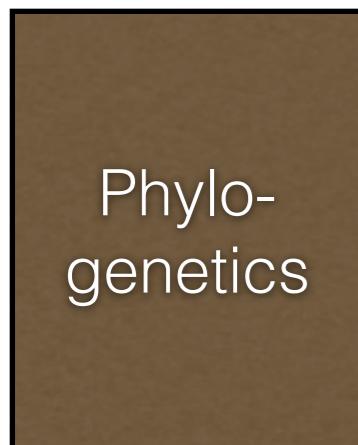
Emergence of HIV

Epidemic identified in 1980s



adopted from Hahn et al. (Science, 2000)

# Genetic data entering epidemiology

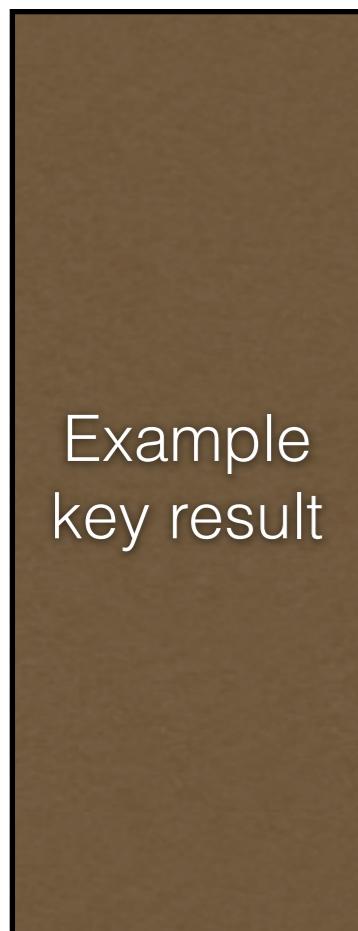
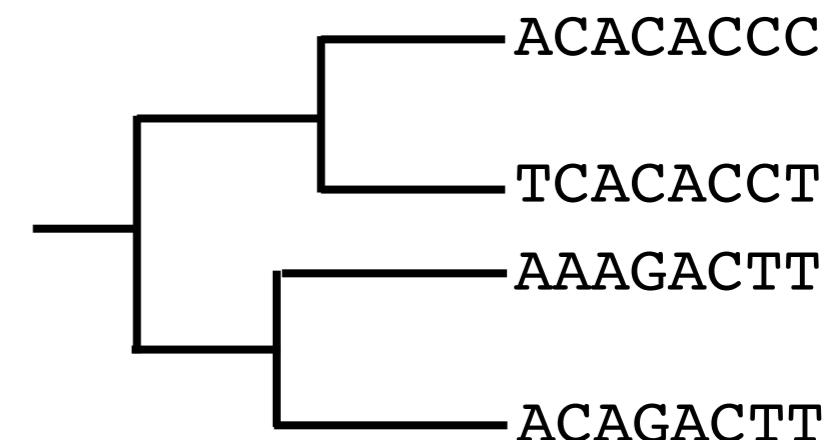


Input:

Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)



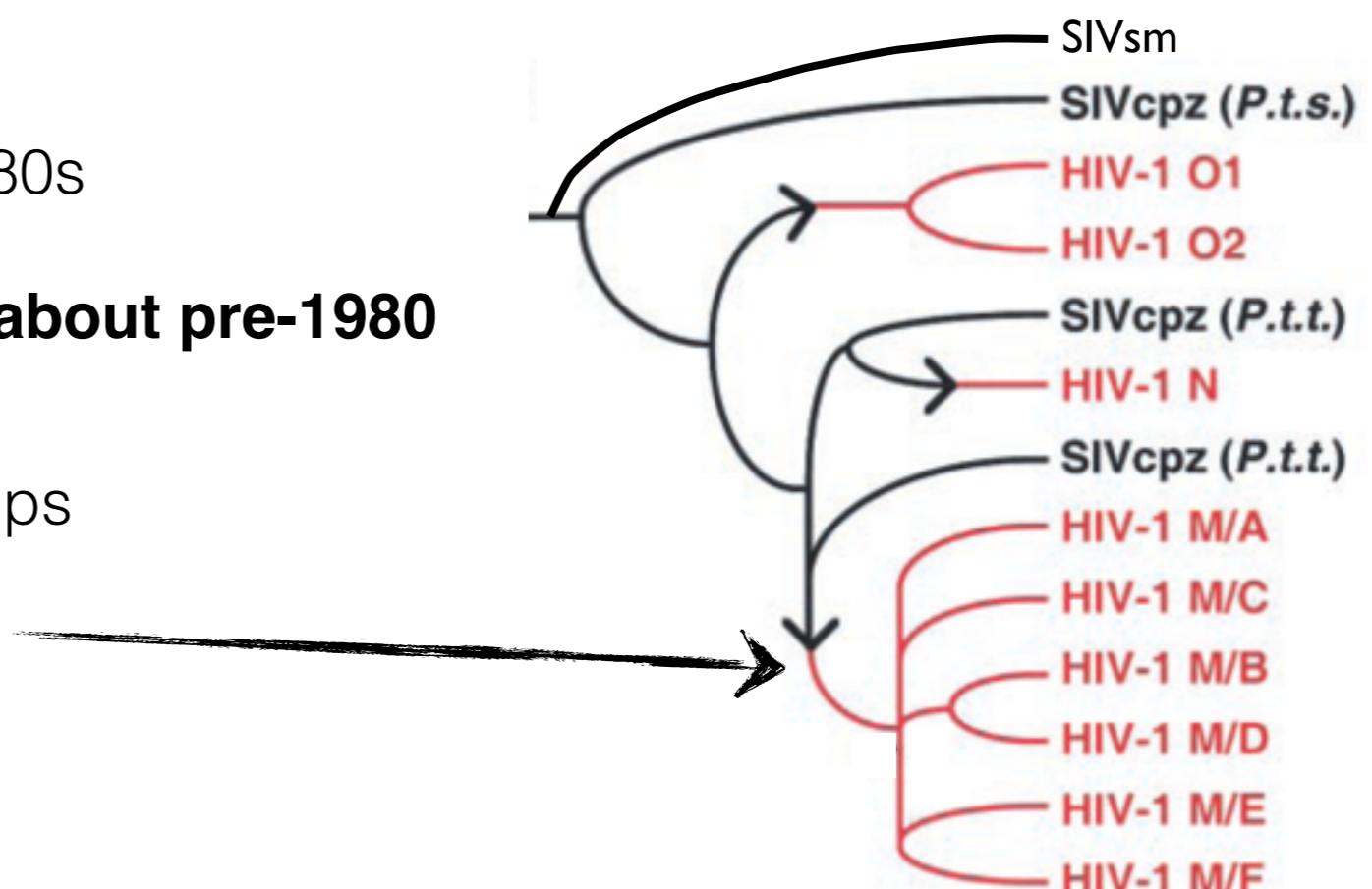
Emergence of HIV

Epidemic identified in 1980s

**Sequence data tells us about pre-1980**

Example  
key result

- ▶ Zoonosis from chimps
- ▶ In Africa
- ▶ Early 20th century



*adopted from Hahn et al. (Science, 2000)*

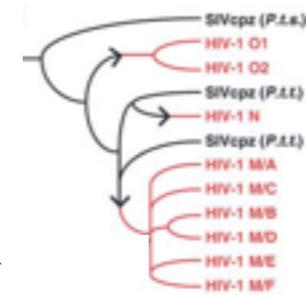
# Genetic data contains information about the epidemiological dynamics

---

Phylogenetics

State of process

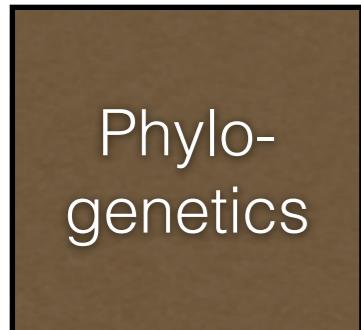
- ▶ Time and geographic location of outbreak



Emergence of a pathogen (here HIV)

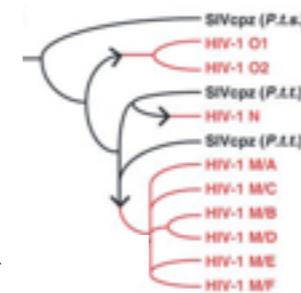
Hahn et al. (Science, 2000)

# Genetic data contains information about the epidemiological dynamics



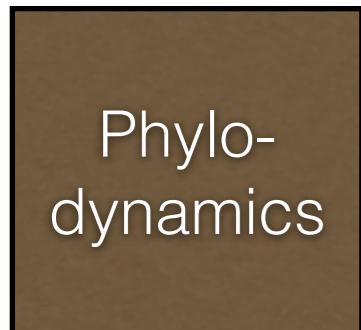
## State of process

- ▶ Time and geographic location of outbreak



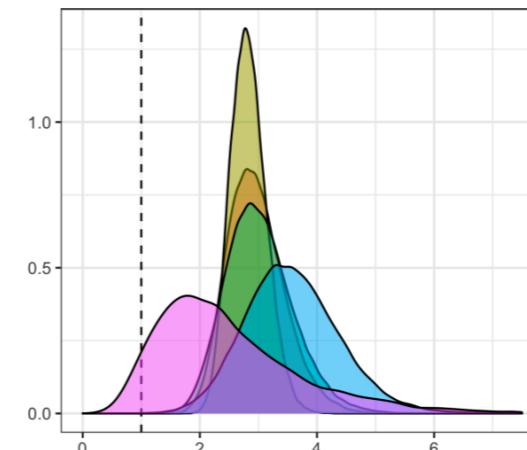
Emergence of a pathogen (here HIV)

Hahn et al. (Science, 2000)



## Dynamics of process

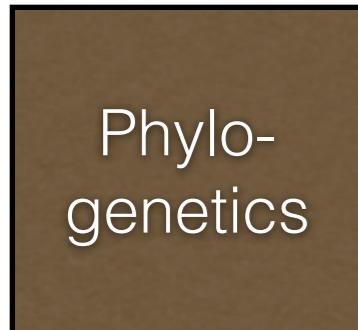
- ▶ Transmission process



Basic reproductive number of an emerging pathogen (here SARS-CoV-2)

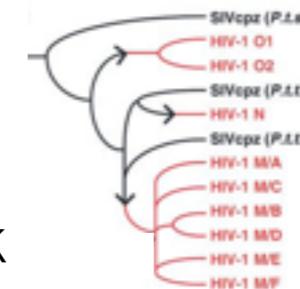
Vaughan et al. (submitted)

# Genetic data contains information about the epidemiological dynamics



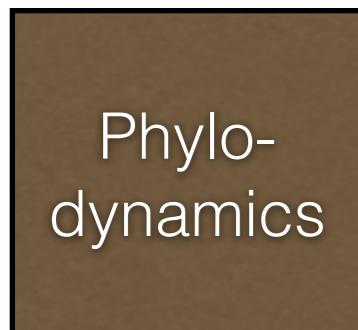
## State of process

- ▶ Time and geographic location of outbreak



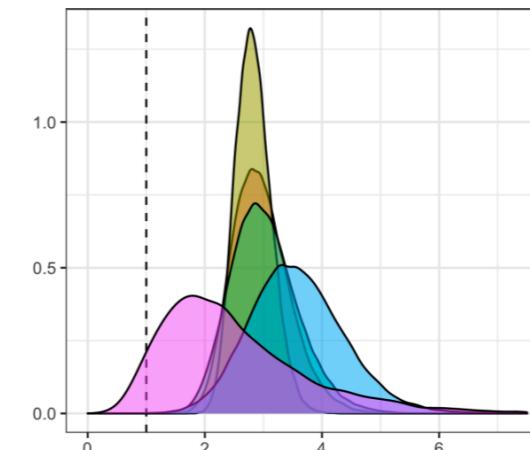
Emergence of a pathogen (here HIV)

Hahn et al. (Science, 2000)

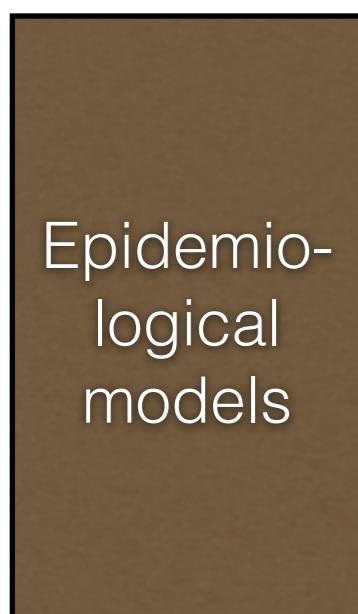


## Dynamics of process

- ▶ Transmission process



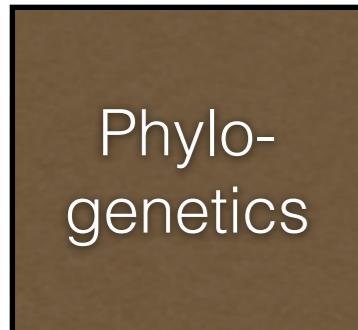
Basic reproductive number of an emerging pathogen (here SARS-CoV-2)  
Vaughan et al. (submitted)



Ideally: assume epidemiological model in phylogenetics (birth-death model)

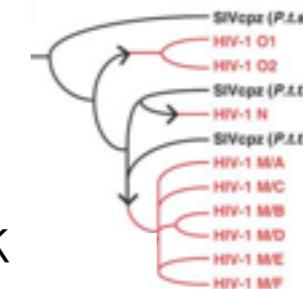
- ▶ Quantification of all model parameters!

# Genetic data contains information about the epidemiological dynamics



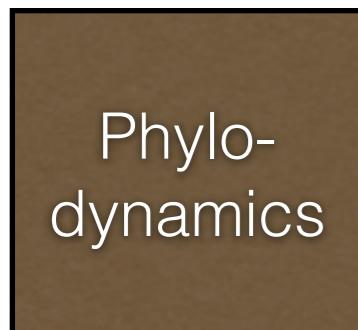
## State of process

- ▶ Time and geographic location of outbreak



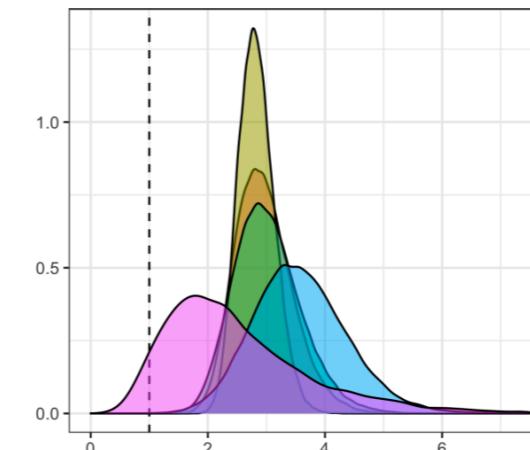
Emergence of a pathogen (here HIV)

Hahn et al. (Science, 2000)

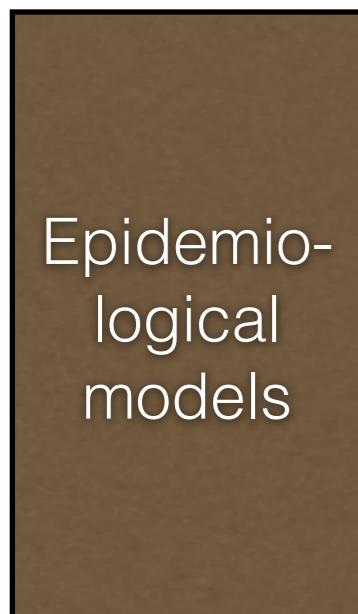


## Dynamics of process

- ▶ Transmission process



Basic reproductive number of an emerging pathogen (here SARS-CoV-2)  
Vaughan et al. (submitted)



Ideally: assume epidemiological model in phylogenetics (birth-death model)

- ▶ Quantification of all model parameters!

Approximate: by population genetic model (coalescent)

- ▶ Deterministic population size is parameterized!

# Outline

---

- Why phylodynamics?
  - ▶ A macroevolutionary motivation
  - ▶ An epidemiological motivation
- A phylodynamic model and the math behind
  - ▶ Birth-death models
  - ▶ The coalescent
- Applications in Macroevolution
- Application in Epidemiology: SARS-CoV-2
- BEAST2 Tutorial

# Phylodynamic model

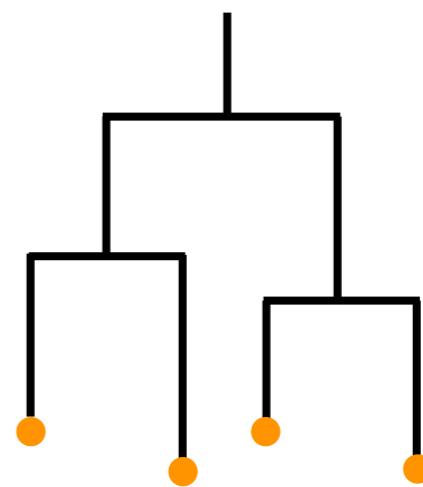
---

Population  
model:  
phylodyn.  
likelihood

Growth of tree

**(Population dynamics - replication)**

- parameters  $\eta$  ( $\rightarrow p[\text{tree} | \eta]$ )



# Phylodynamic model

---

Population model:  
phyodyn.  
likelihood

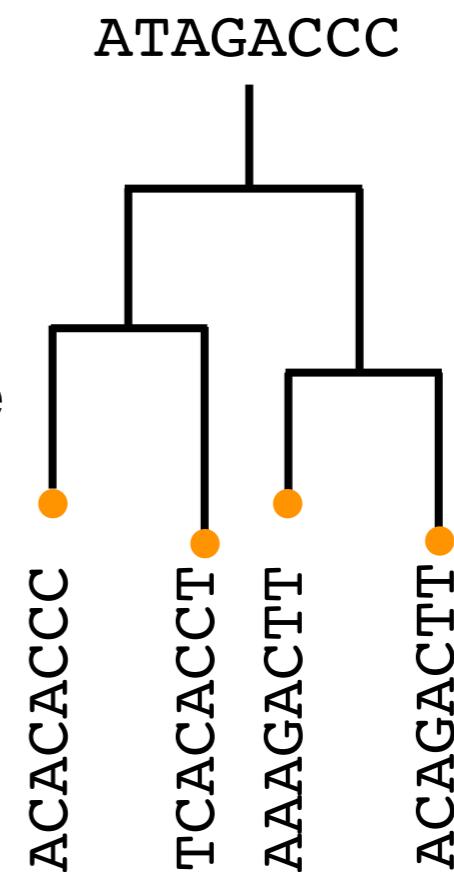
Growth of tree  
**(Population dynamics - replication)**

- parameters  $\eta$  ( $\rightarrow p[\text{tree} | \eta]$ )

Evolution model:  
phylogen.  
likelihood

Evolution of sequences/morphology along tree  
**(Evolutionary dynamics - genetic change)**

- parameters  $\theta$  ( $\rightarrow p[\text{seq} | \theta, \text{tree}]$ )



# Phylodynamic model

---

Population model:  
phyodyn.  
likelihood

Growth of tree  
**(Population dynamics - replication)**

- parameters  $\eta$  ( $\rightarrow p[\text{tree} | \eta]$ )

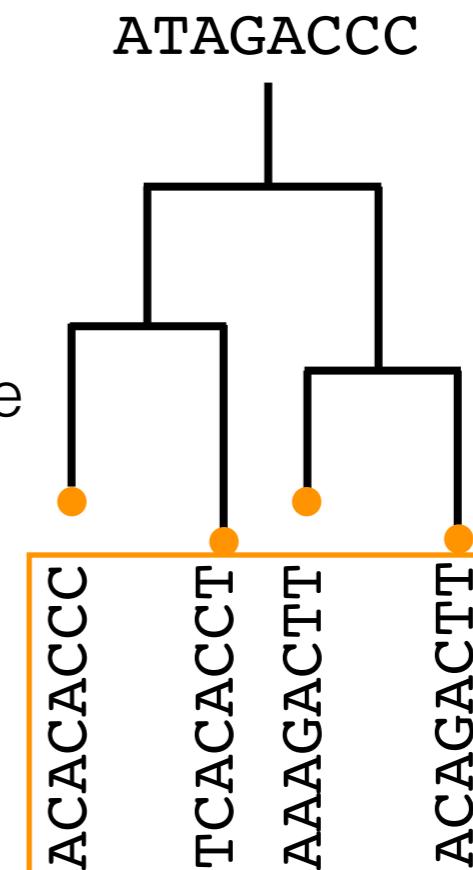
Evolution model:  
phylogen.  
likelihood

Evolution of sequences/morphology along tree  
**(Evolutionary dynamics - genetic change)**

- parameters  $\theta$  ( $\rightarrow p[\text{seq} | \theta, \text{tree}]$ )

Data

Sequence alignment



# Phylodynamic model

Population model:  
phyodyn.  
likelihood

Growth of tree

**(Population dynamics - replication)**

- parameters  $\eta$  ( $\rightarrow p[\text{tree} | \eta]$ )

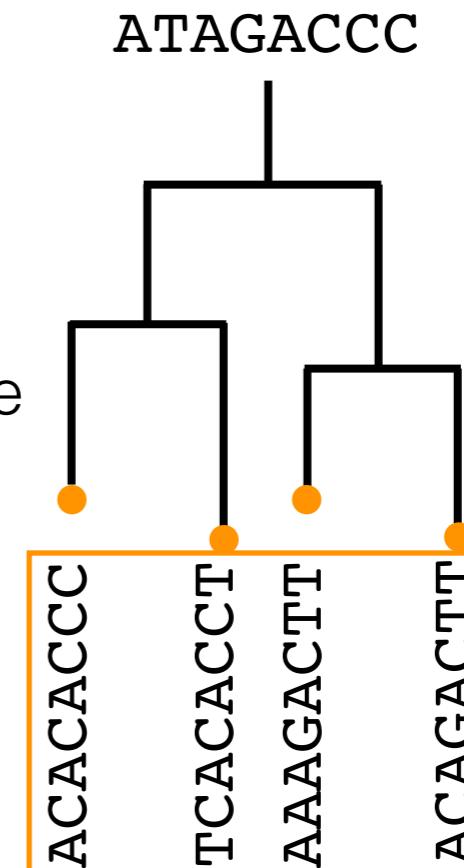
Evolution model:  
phylogen.  
likelihood

Evolution of sequences/morphology along tree

**(Evolutionary dynamics - genetic change)**

- parameters  $\theta$  ( $\rightarrow p[\text{seq} | \theta, \text{tree}]$ )

Sequence alignment



Data

**$p[\text{tree}, \eta, \theta | \text{seq}]$**

=

$$p[\text{seq} | \theta, \text{tree}] p[\text{tree} | \eta] p[\theta] p[\eta] / p[\text{seq}]$$

# Phylodynamic model

Population model:  
phyldyn.  
likelihood

Growth of tree

**(Population dynamics - replication)**

- parameters  $\eta$  ( $\rightarrow p[\text{tree} | \eta]$ )

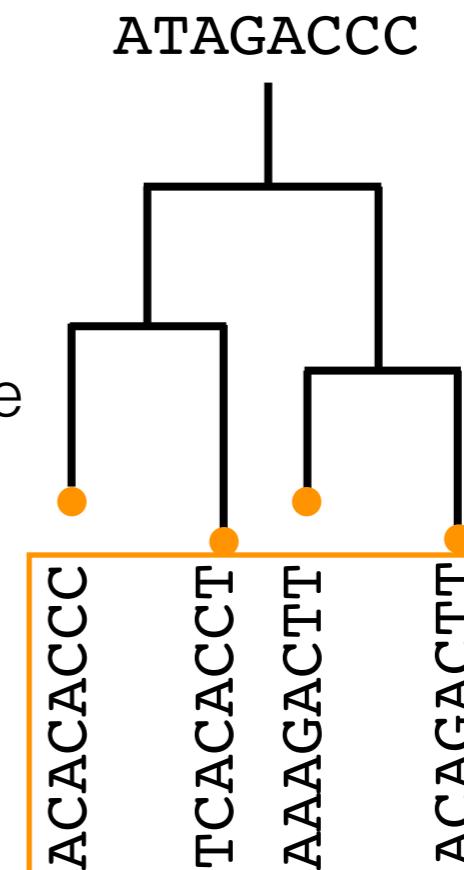
Evolution model:  
phylogen.  
likelihood

Evolution of sequences/morphology along tree

**(Evolutionary dynamics - genetic change)**

- parameters  $\theta$  ( $\rightarrow p[\text{seq} | \theta, \text{tree}]$ )

Sequence alignment

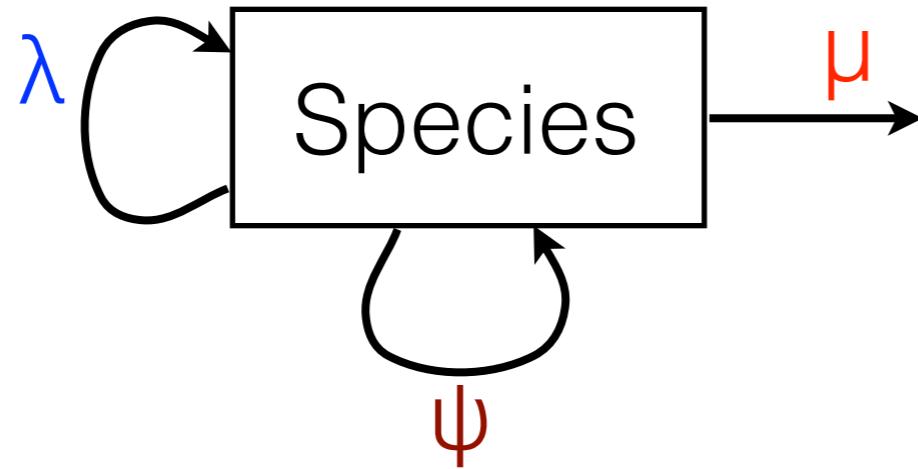


Data

**$p[\text{tree}, \eta, \theta | \text{seq}]$**

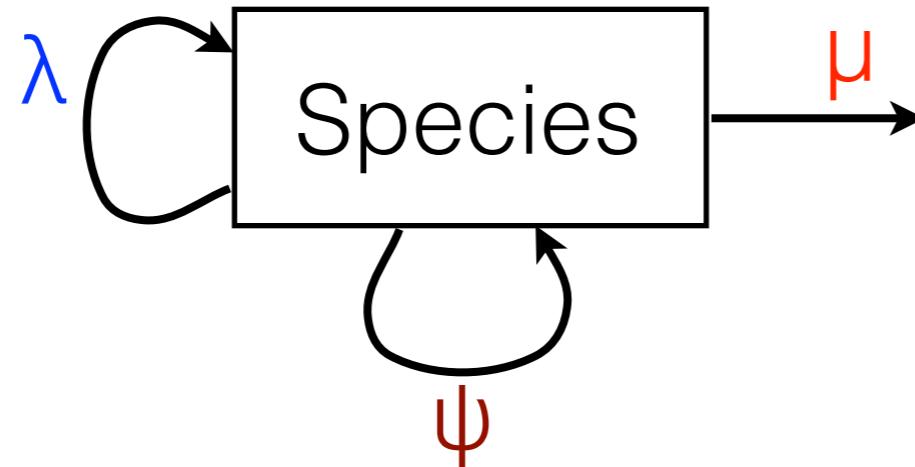
$$p[\text{seq} | \theta, \text{tree}] \cdot p[\text{tree} | \eta] \cdot p[\theta] \cdot p[\eta] / p[\text{seq}]$$

# Birth-death model for modelling macroevolutionary processes

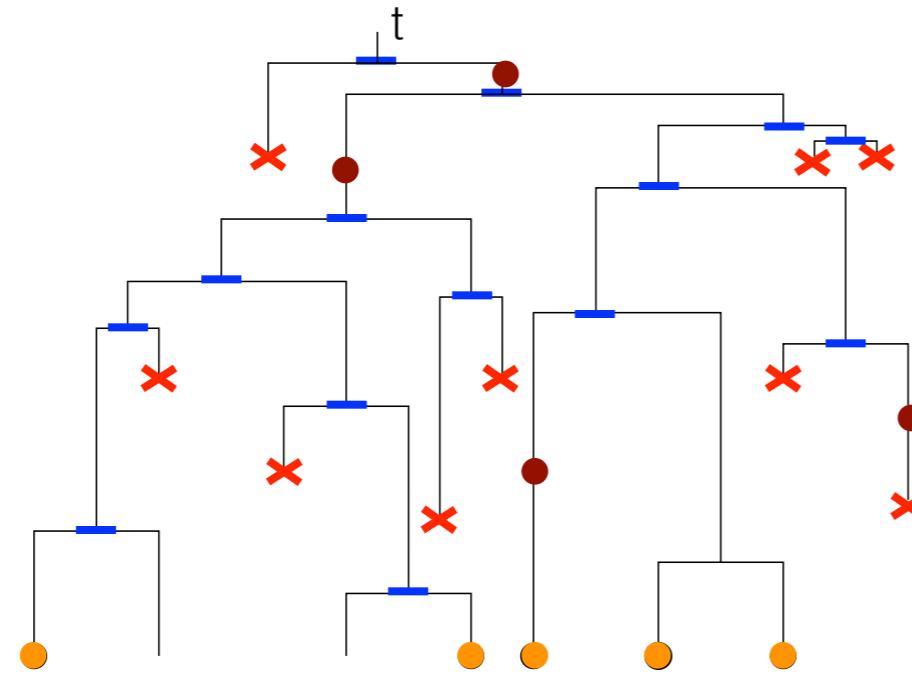
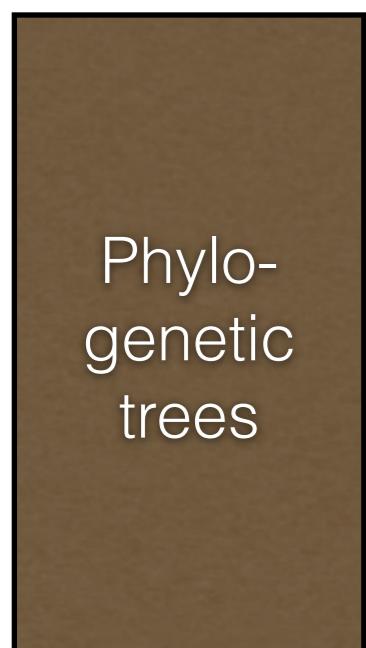


- stem age of a clade  $t_0$
- speciation rate  $\lambda$
- extinction rate  $\mu$
- fossil-observation rate  $\psi$
- sampling probability  $p$

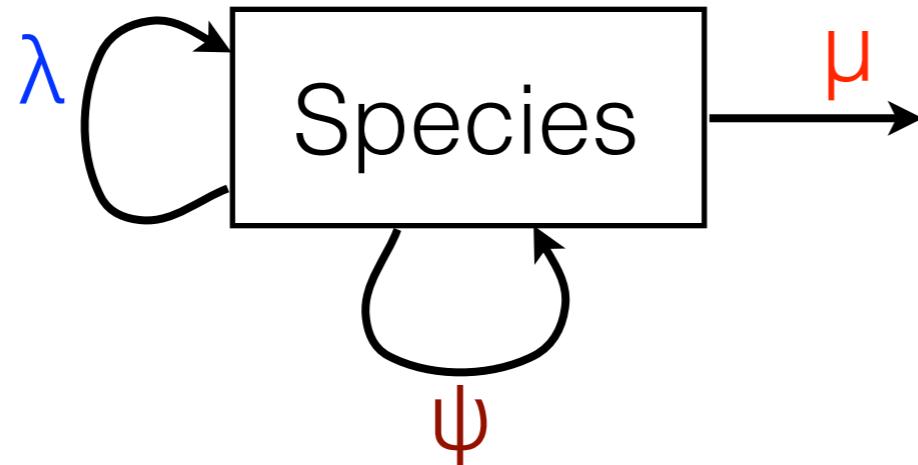
# Birth-death model for modelling macroevolutionary processes



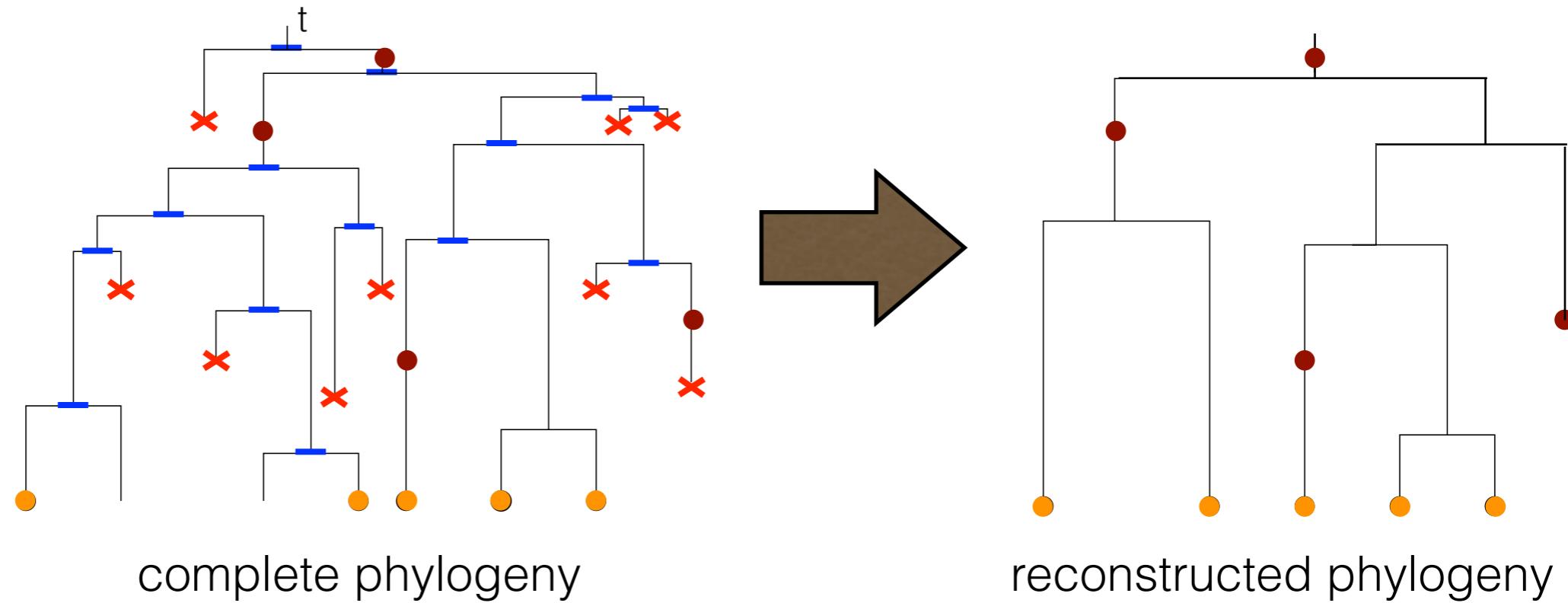
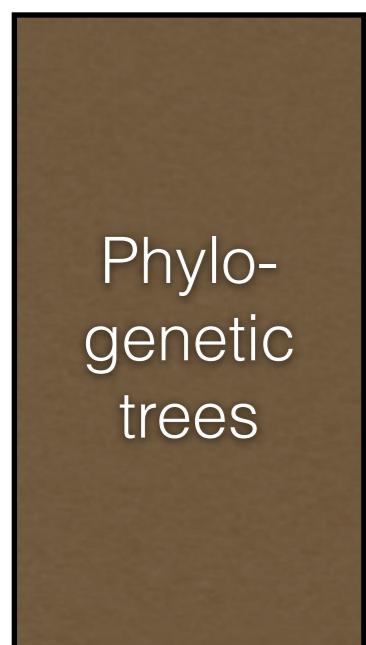
- stem age of a clade  $t_0$
- speciation rate  $\lambda$
- extinction rate  $\mu$
- fossil-observation rate  $\psi$
- sampling probability  $p$



# Birth-death model for modelling macroevolutionary processes

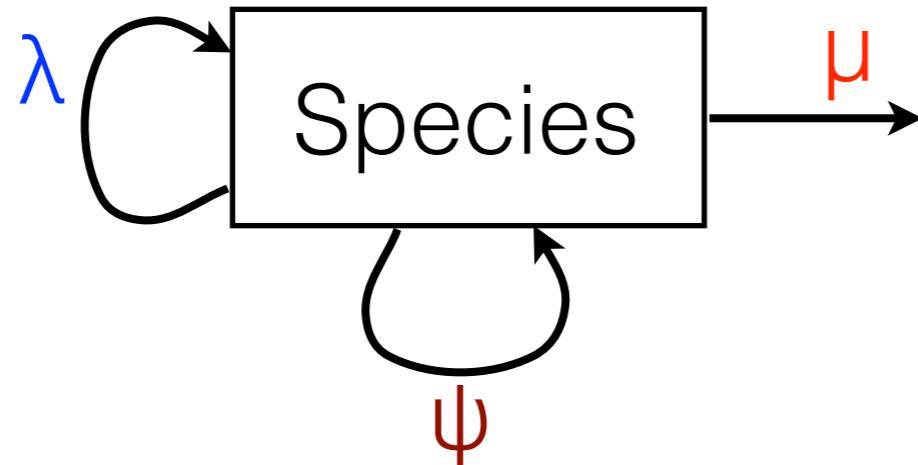


- stem age of a clade  $t_0$
- speciation rate  $\lambda$
- extinction rate  $\mu$
- fossil-observation rate  $\psi$
- sampling probability  $p$



# Birth-death model for modelling macroevolutionary processes

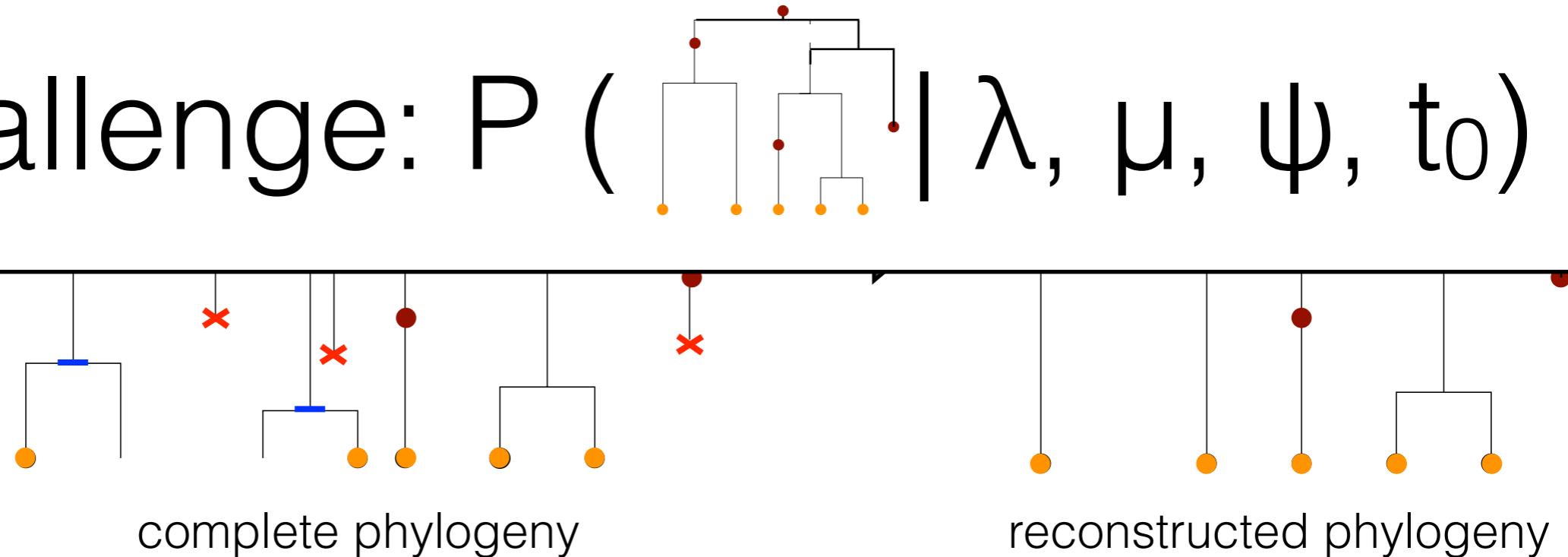
Birth-death model



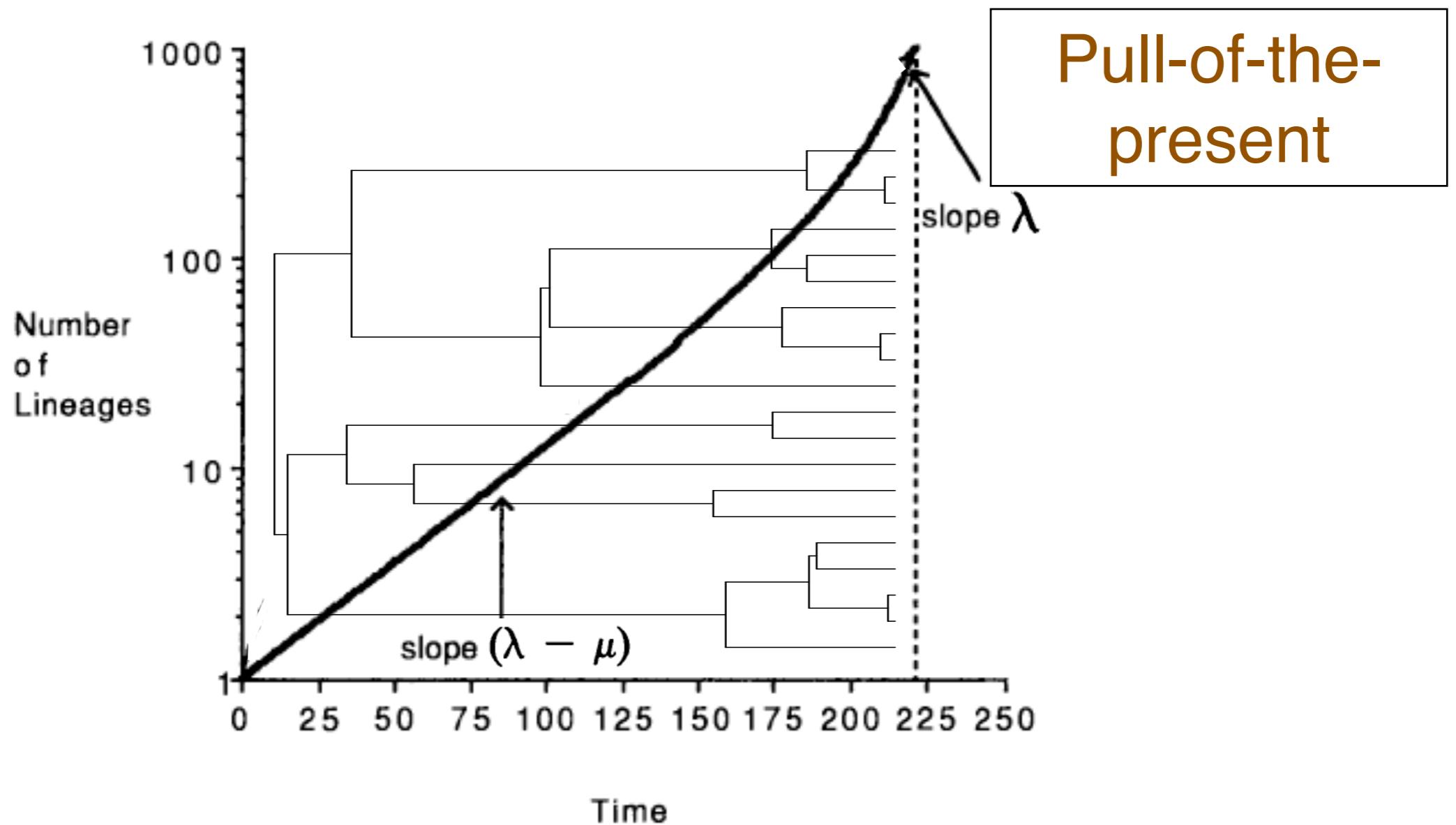
- stem age of a clade  $t_0$
- speciation rate  $\lambda$
- extinction rate  $\mu$
- fossil-observation rate  $\psi$
- sampling probability  $p$

Challenge:  $P(\text{ } | \lambda, \mu, \psi, t_0)$

genetic trees

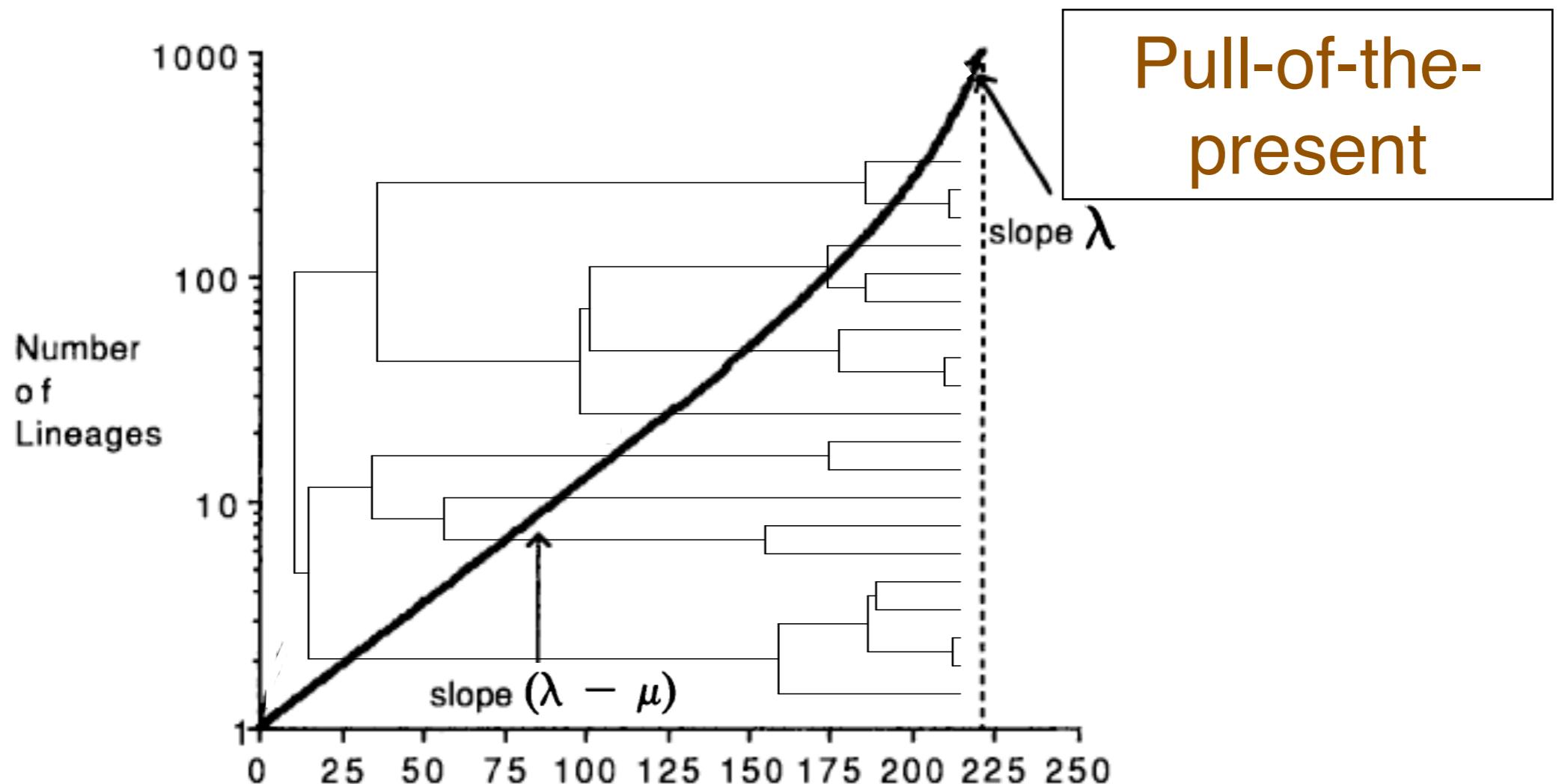


# Estimating speciation and extinction rates from phylogenies



Harvey et al (1994, *Evolution*); Nee et al. (1994, *Phil. Trans. B*)

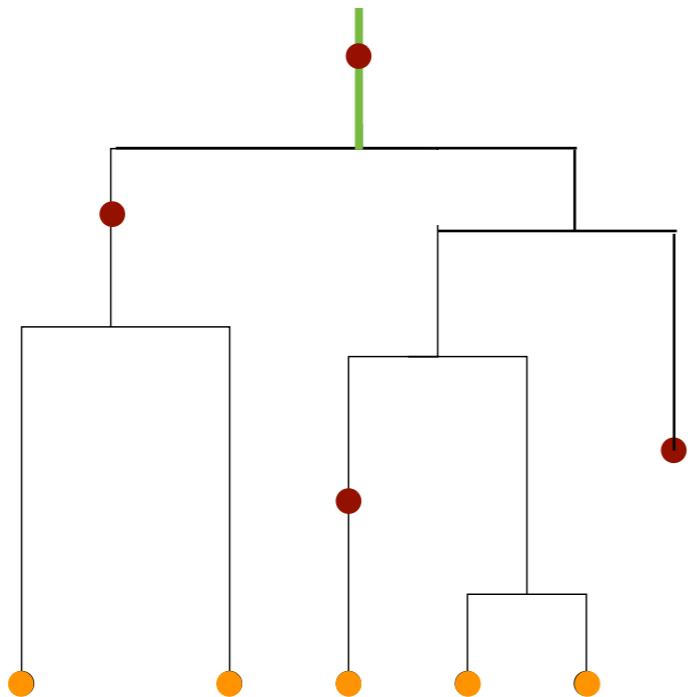
# Estimating speciation and extinction rates from phylogenies



Pull-of-the-present informs birth- and death rates

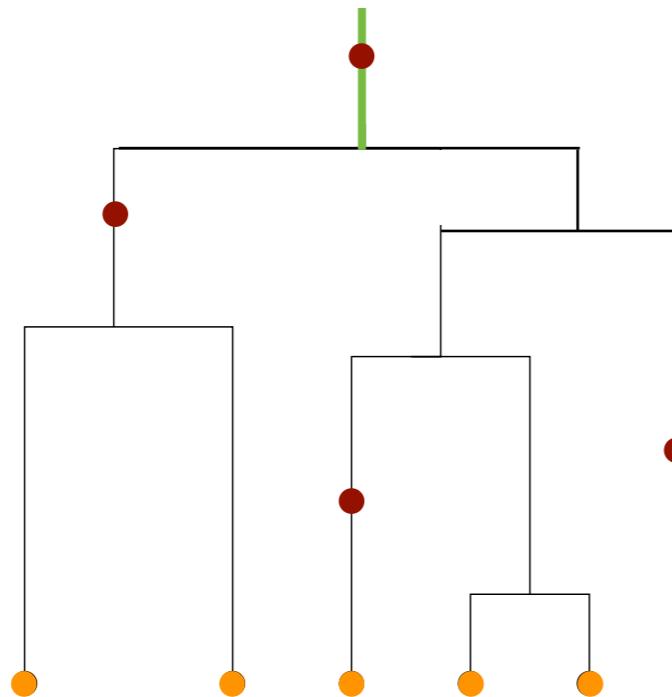
# Calculating the probability density: recursion

---

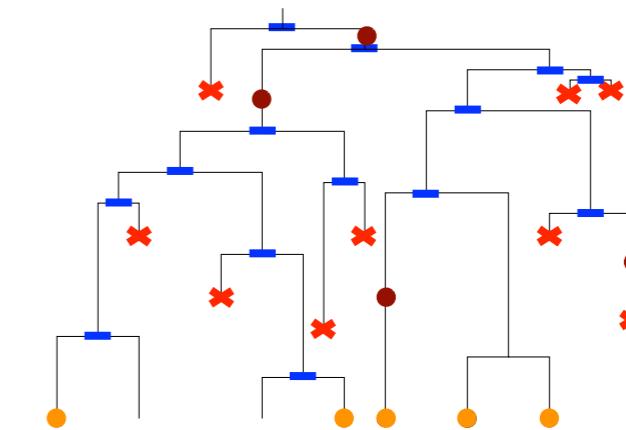


# Calculating the probability density: recursion

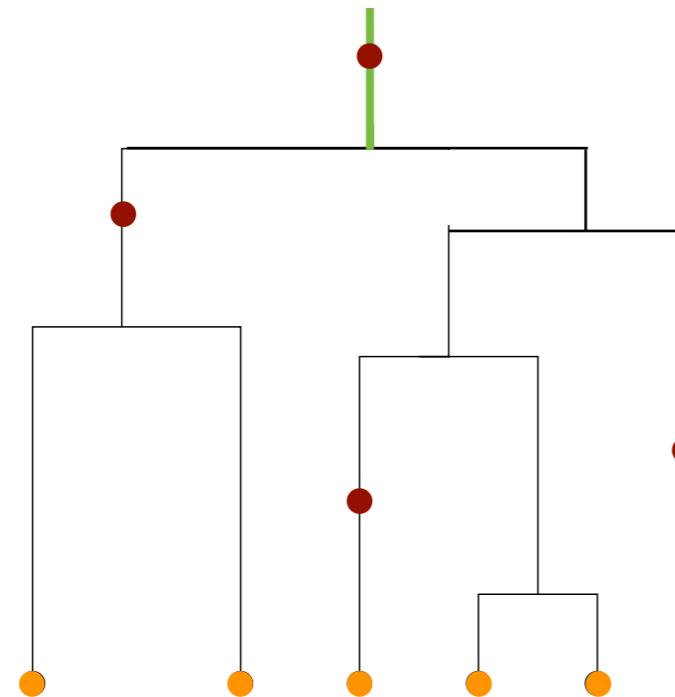
---



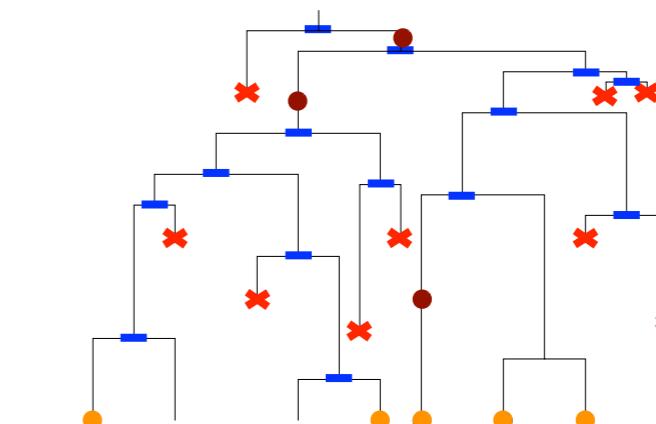
Integrate over all complete  
phylogenies:



# Calculating the probability density: recursion



Integrate over all complete phylogenies:



$$P(\text{ } \text{ } \text{ } ) = P(|) \lambda P(\text{ } \text{ } \text{ } ) P(\text{ } \text{ } \text{ } )$$

Differential equation

Recursive

# Combining molecular and fossil data: Probability density of the phylogeny

---

$$P \left( \begin{array}{c} \text{phylogenetic tree} \\ \text{with red dots at nodes} \end{array} \right) = \frac{1}{\lambda(1 - p_0(x_0))} \frac{4\lambda\rho}{q(x_0)} \prod_{i \in \mathcal{V}} \frac{4\lambda\rho}{q(x_i)} \prod_{f \in \mathcal{F}} \psi \left( 2\lambda \frac{p(y_f)q(y_f)}{q(z_f)} \right)^{\mathcal{I}_f}$$

$$p_0(t) = \frac{\lambda + \mu + \psi + c_1 \frac{e^{-c_1 t}(1 - c_2) - (1 + c_2)}{e^{-c_1 t}(1 - c_2) + (1 + c_2)}}{2\lambda}$$

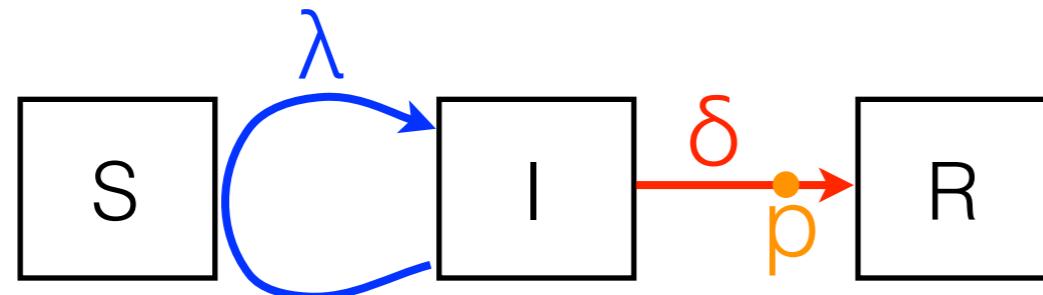
$$q(t) = 2(1 - c_2^2) + e^{-c_1 t}(1 - c_2)^2 + e^{c_1 t}(1 + c_2)^2$$

$$c_1 = \sqrt{(\lambda - \mu - \psi)^2 + 4\lambda\psi},$$

$$c_2 = -\frac{\lambda - \mu - 2\lambda\rho - \psi}{c_1}.$$

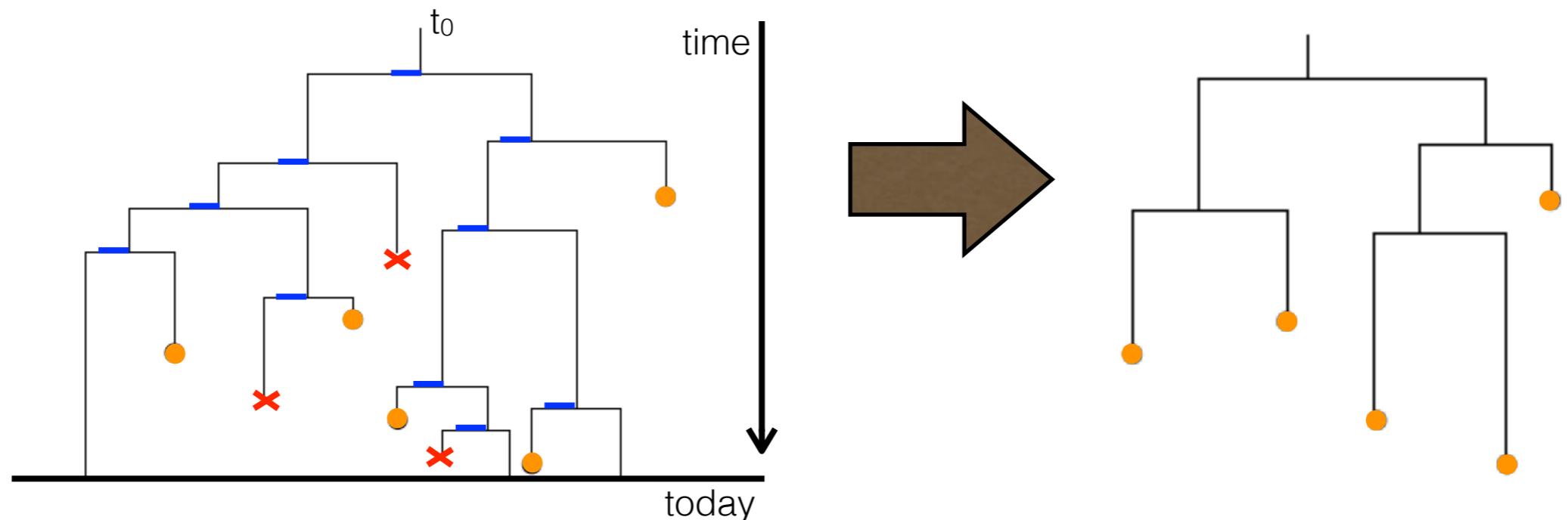
# Birth-death model for modelling epidemiological processes

Birth-death  
model for  
epidemic  
spread



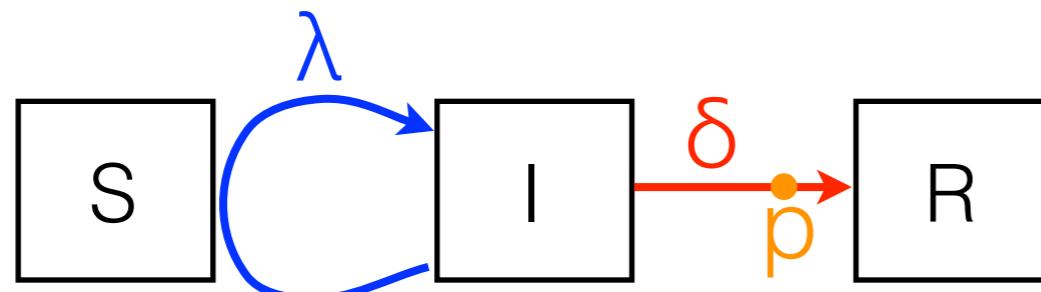
- start of outbreak  $t_0$
- transmission rate  $\lambda$
- “becoming-non-infectious” rate  $\delta$
- sampling probability  $p$

Phylo-  
genetic  
trees



# Birth-death model for modelling epidemiological processes

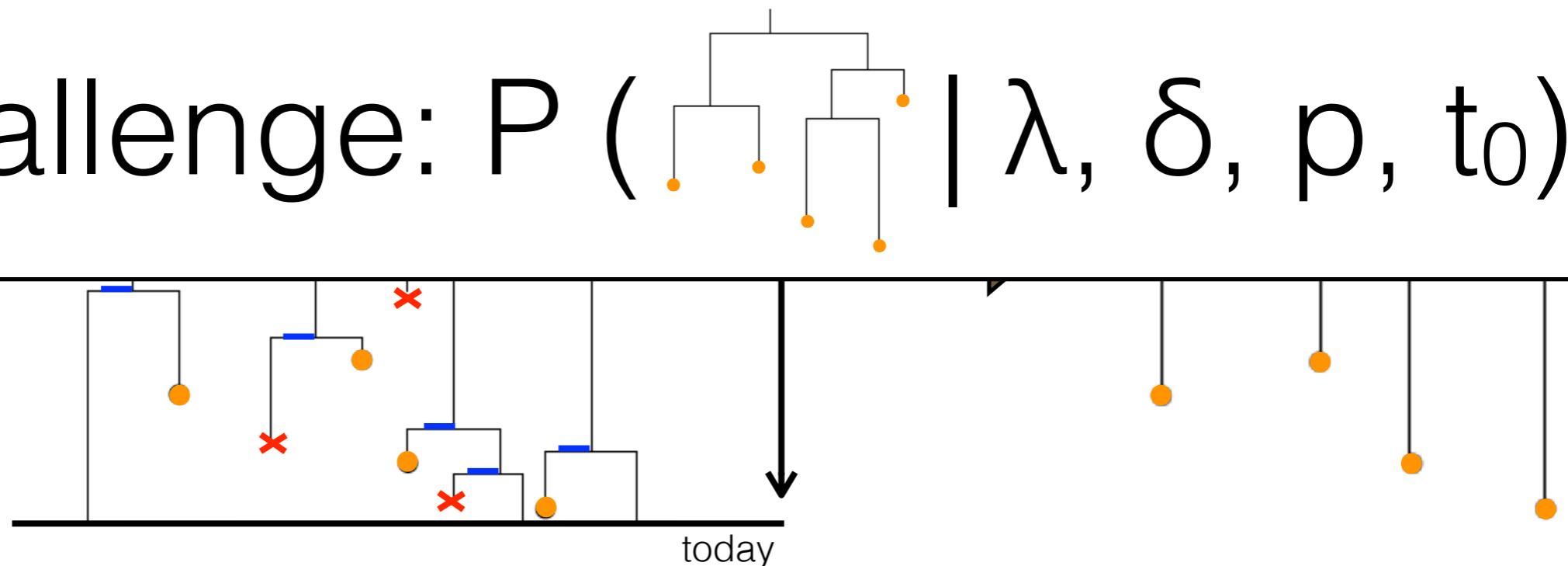
Birth-death  
model for  
epidemic  
spread



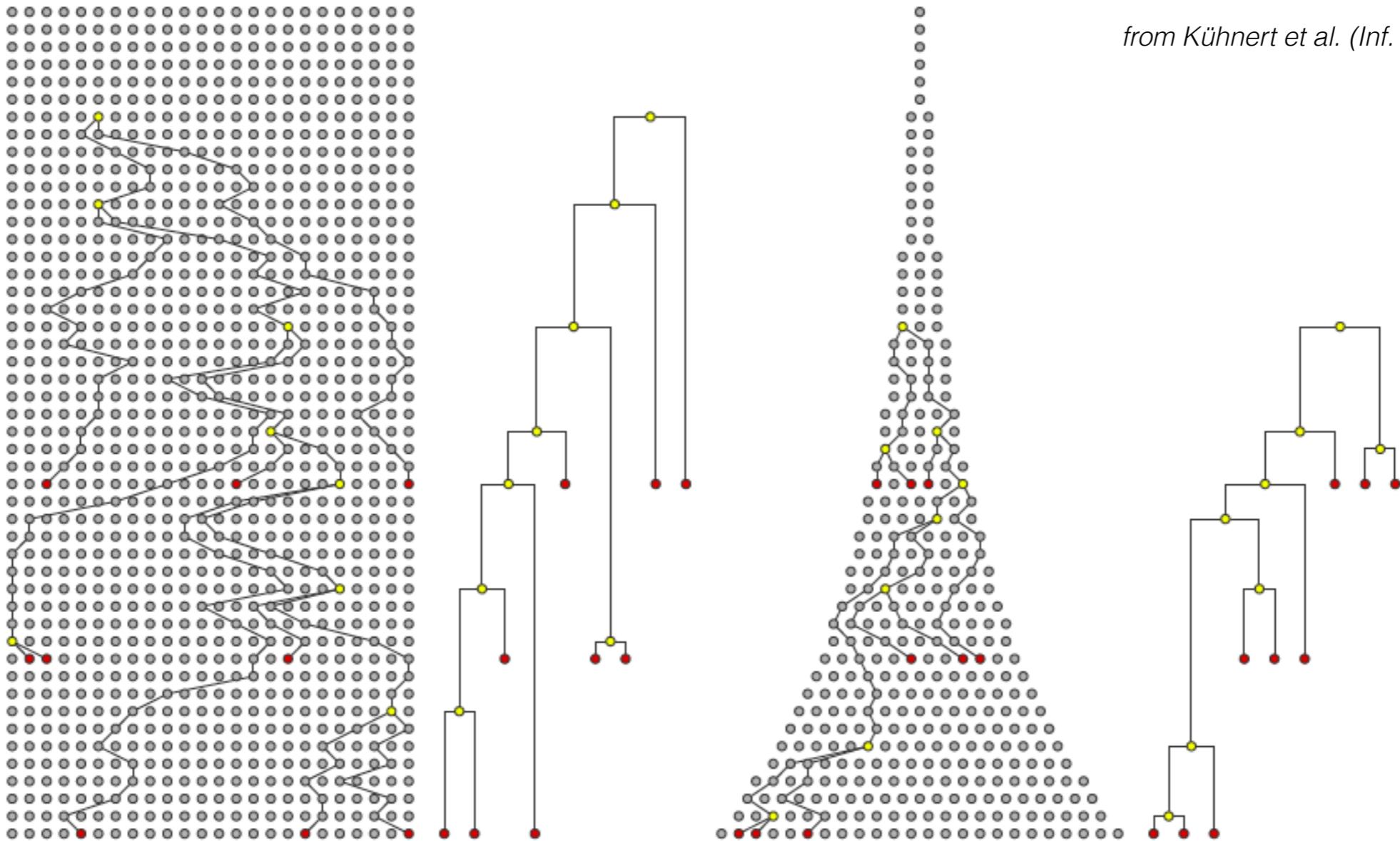
- start of outbreak  $t_0$
- transmission rate  $\lambda$
- “becoming-non-infectious” rate  $\delta$
- sampling probability  $p$

Challenge:  $P(\text{---} | \lambda, \delta, p, t_0)$

genetic  
trees

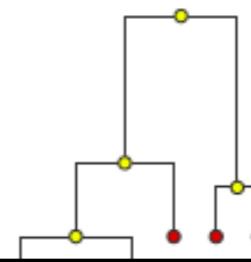
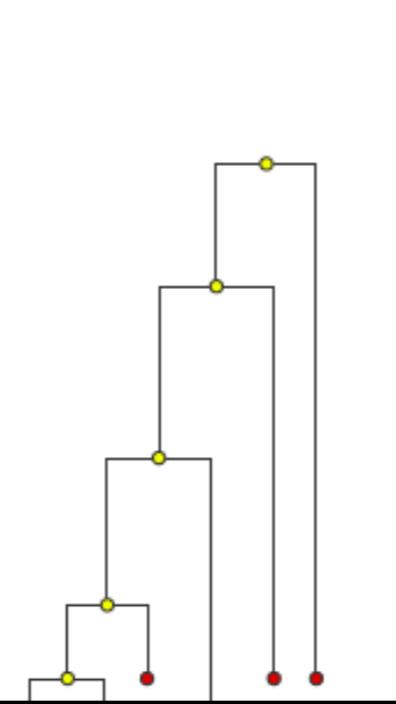
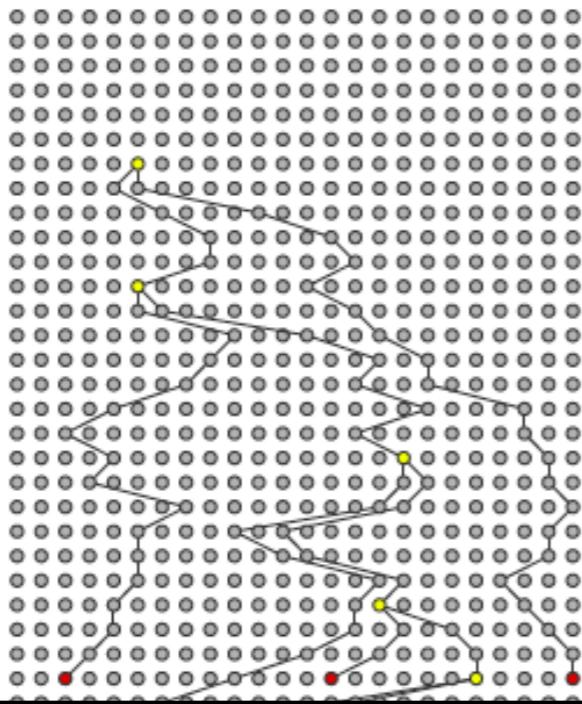


# Coalescent as an alternative model for epidemic spread



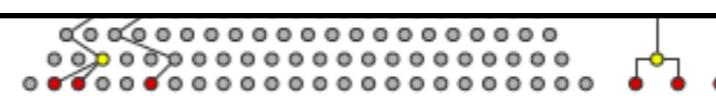
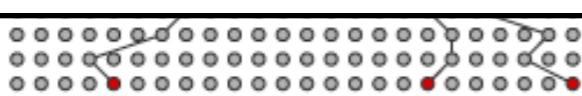
from Kühnert et al. (*Inf. Gen. Evol.*, 2011)

# Coalescent as an alternative model for epidemic spread

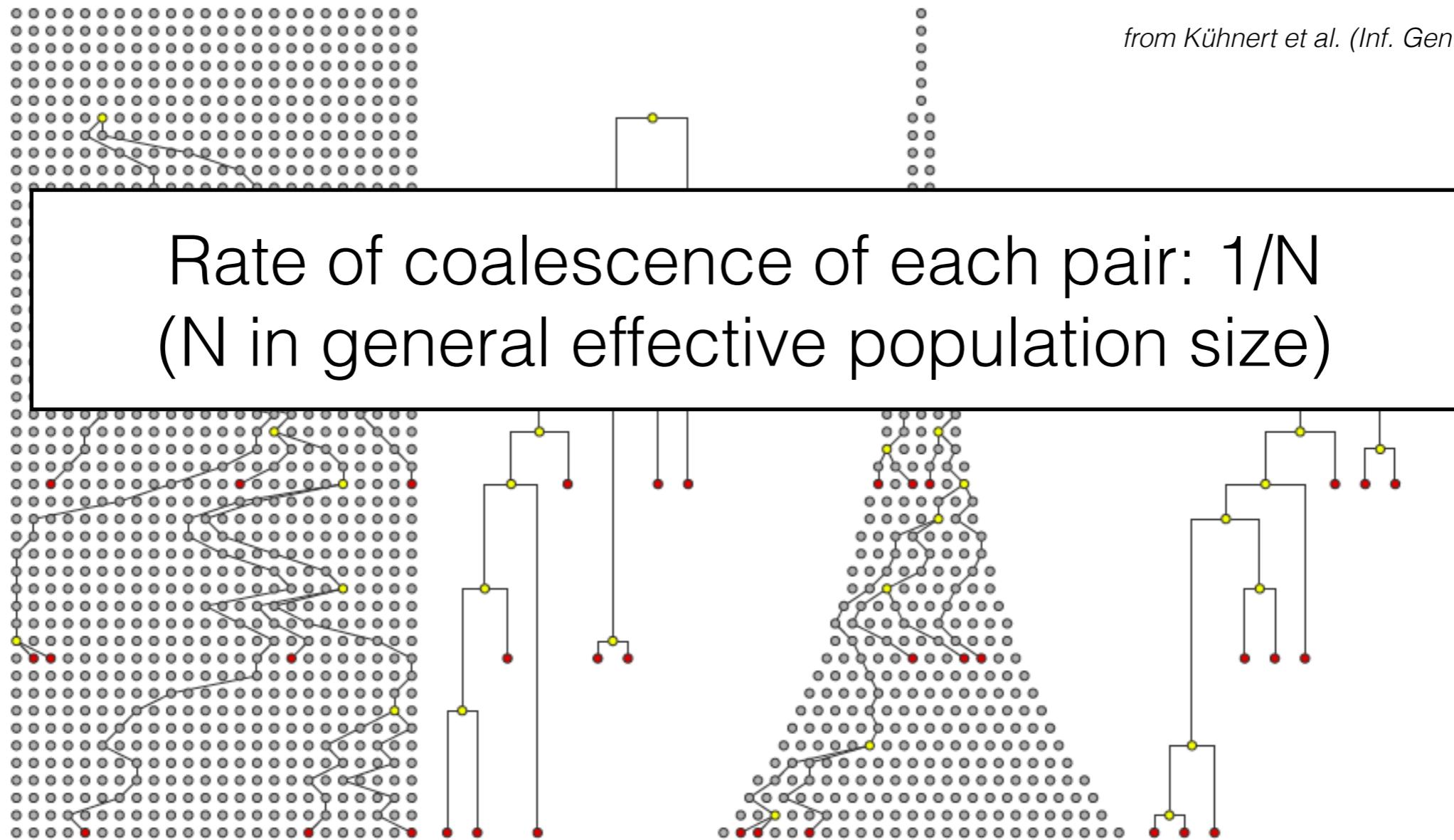


from Kühnert et al. (*Inf. Gen. Evol.*, 2011)

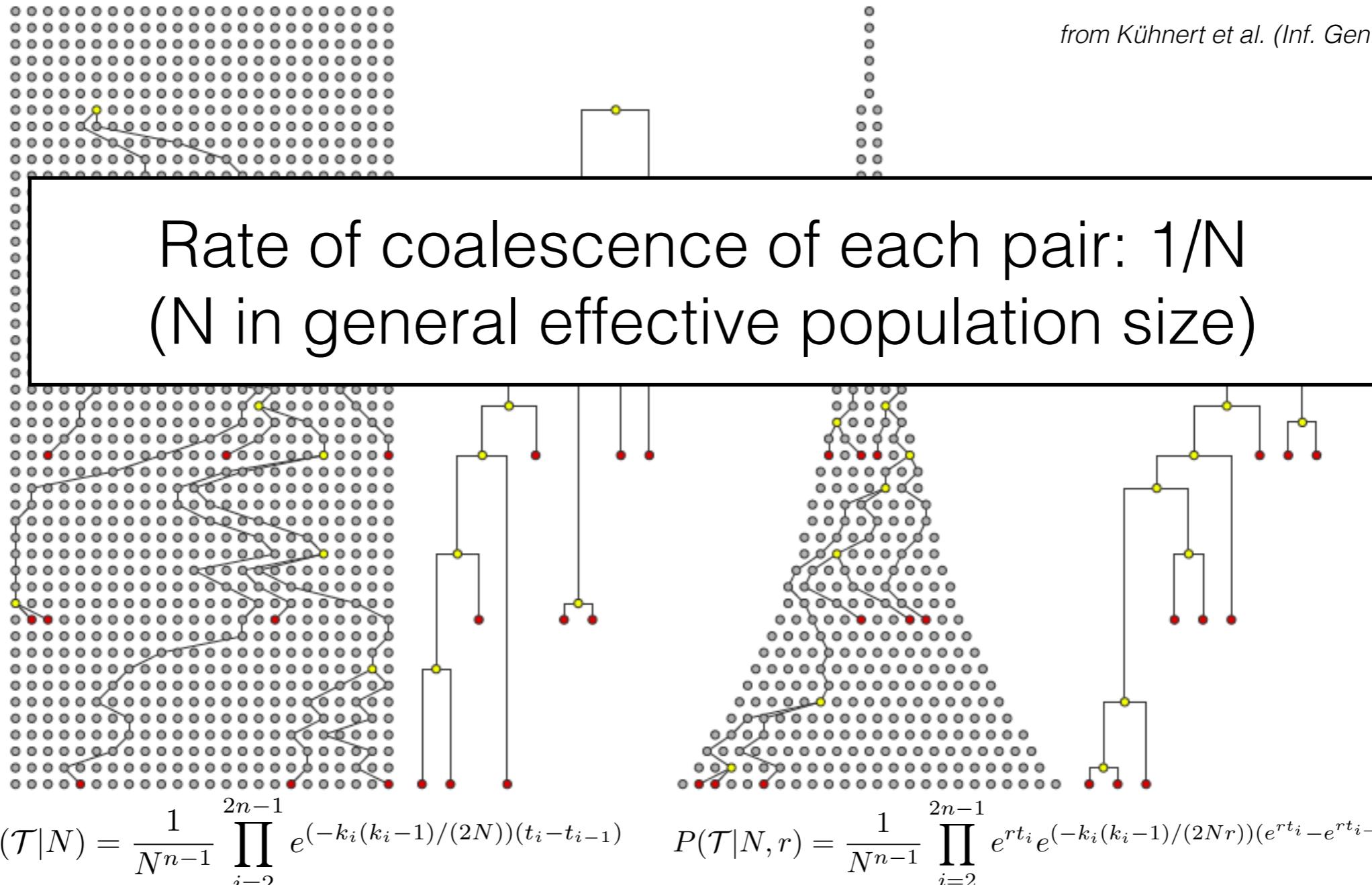
$$P \left( \text{ } \right) ?$$



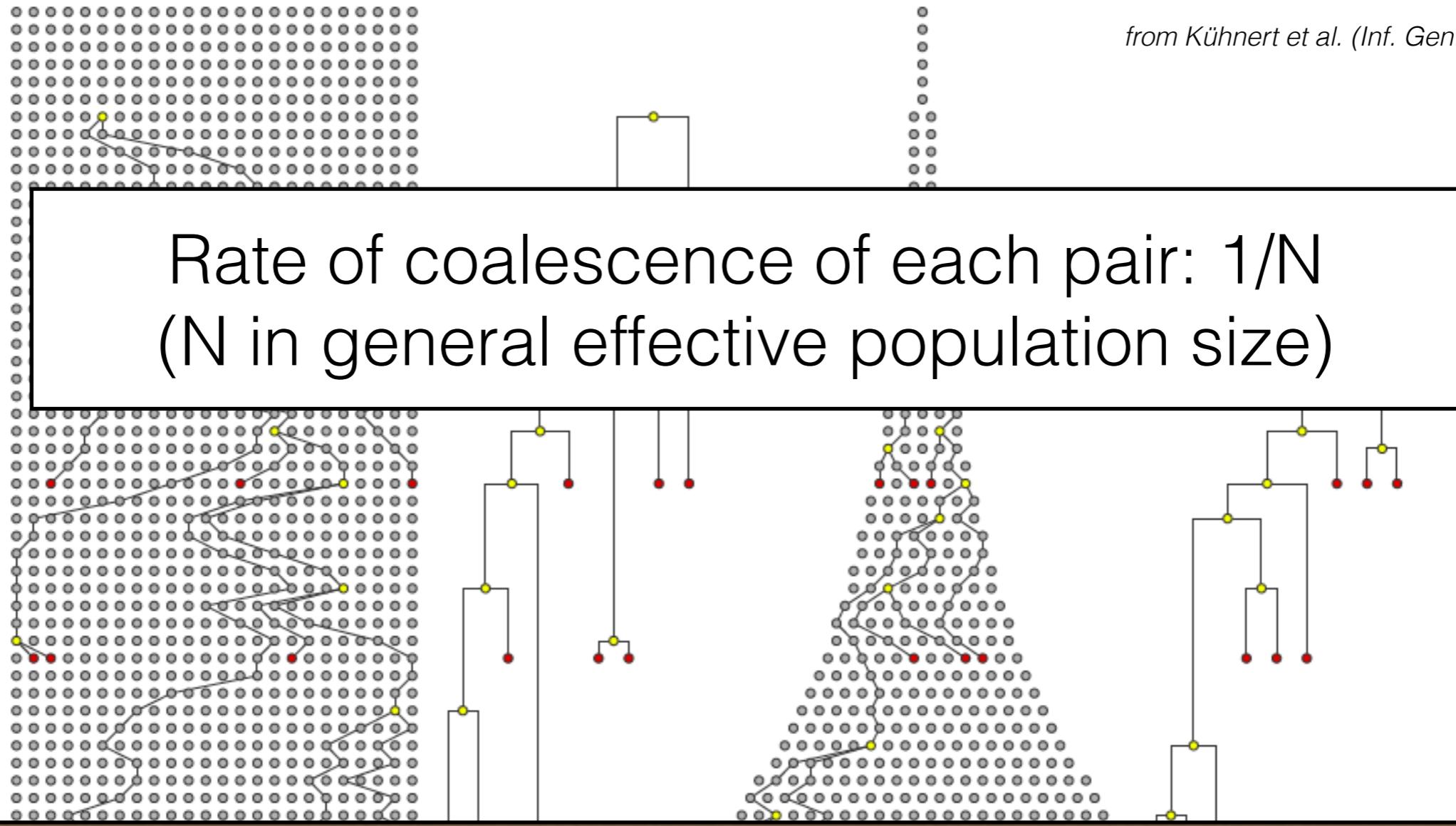
# Coalescent as an alternative model for epidemic spread



# Coalescent as an alternative model for epidemic spread



# Coalescent as an alternative model for epidemic spread



Many short branches indicate small population!

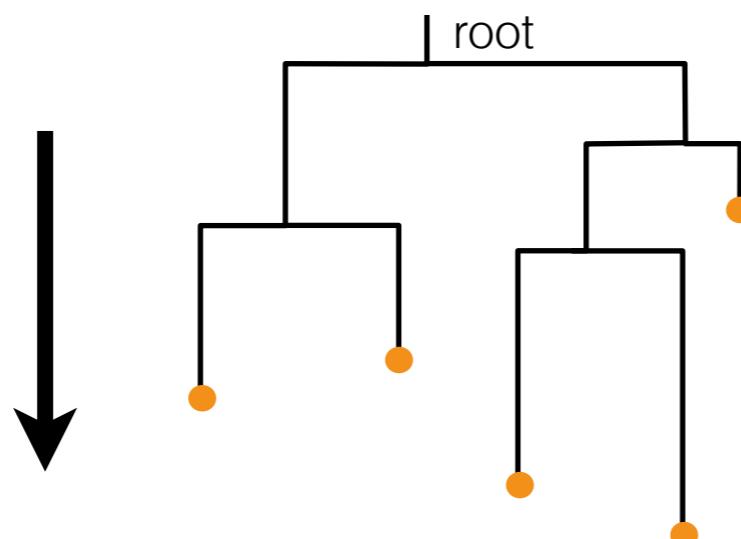
# How can we quantify population dynamics?

---

Birth-death model

Coalescent

- ▶ Process starts with initial ind.: *forward* in time!
- ▶ Birth, death & sampling describe the *stochastic* dynamics ( $\eta$ )
- ▶ root age is parameter
- ▶ tip ages are *random*
- ▶  $P(\text{tree, tips} \mid \text{root}; \eta)$



# How can we quantify population dynamics?

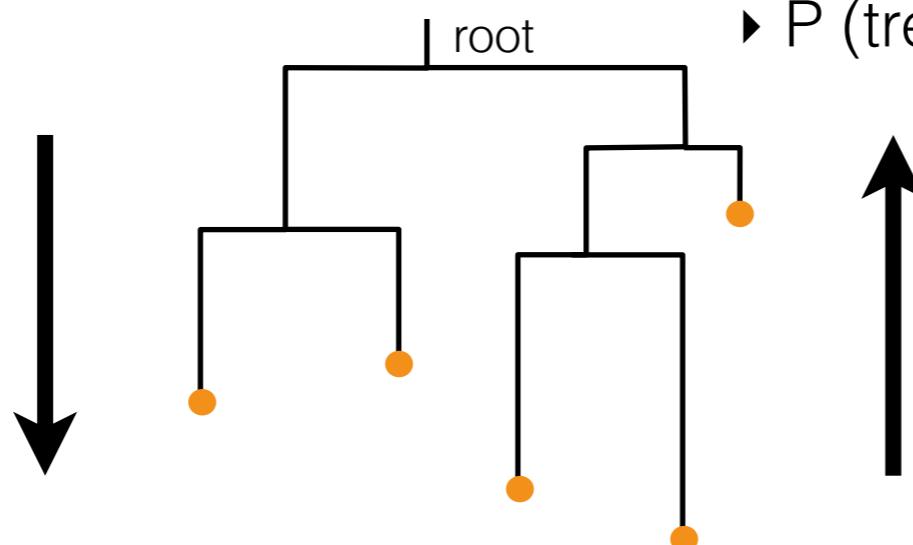
---

## Birth-death model

- ▶ Process starts with initial ind.: *forward* in time!
- ▶ Birth, death & sampling describe the *stochastic* dynamics ( $\eta$ )
- ▶ root age is parameter
- ▶ tip ages are *random*
- ▶  $P(\text{tree, tips} \mid \text{root}; \eta)$

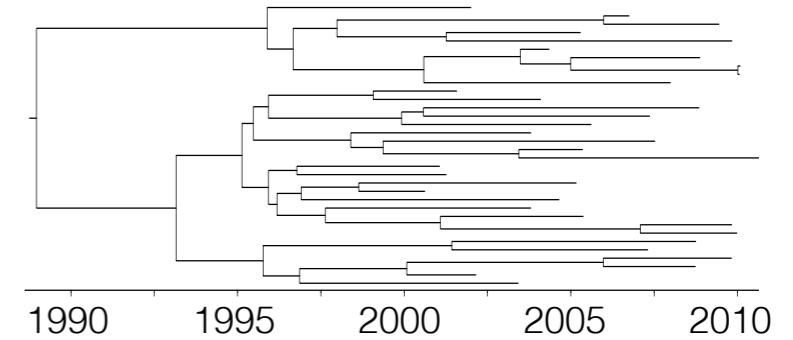
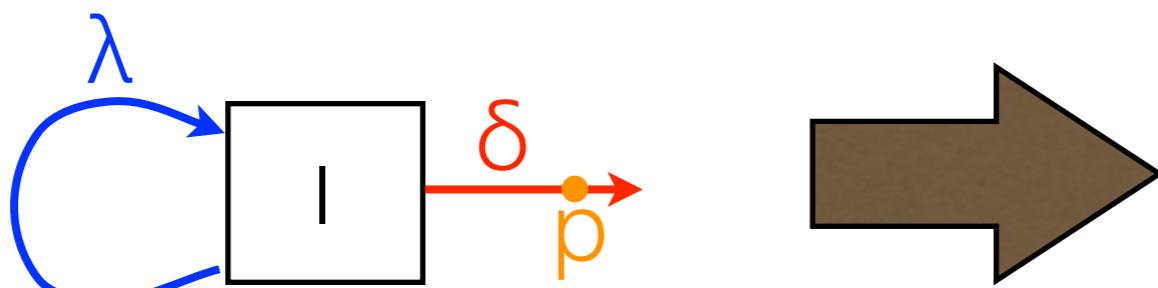
## Coalescent

- ▶ Process starts with  $n$  tips: *backward* in time!
- ▶ Population sizes (typically *deterministic*) describe the dynamics ( $\eta$ )
- ▶ root age is random
- ▶ tip ages are *parameters* (ie. conditioned on)
- ▶  $P(\text{tree, root} \mid \text{tips}; \eta)$



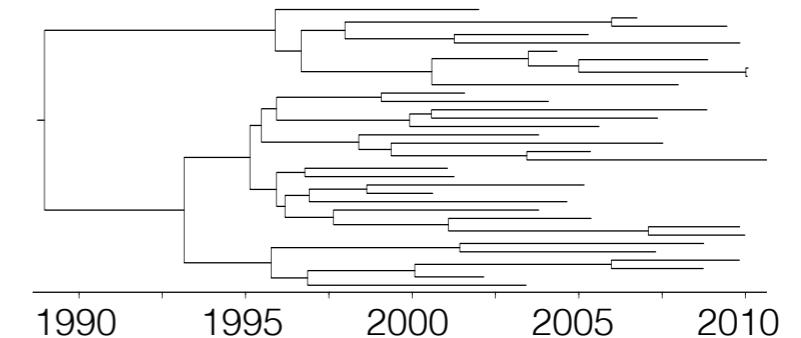
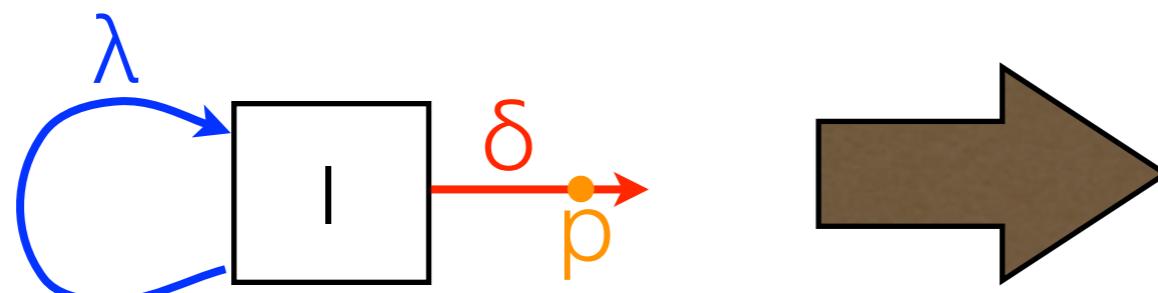
# Simulation study: Comparing birth-death model to coalescent

Simulating  
epidemic  
outbreak



# Simulation study: Comparing birth-death model to coalescent

Simulating  
epidemic  
outbreak



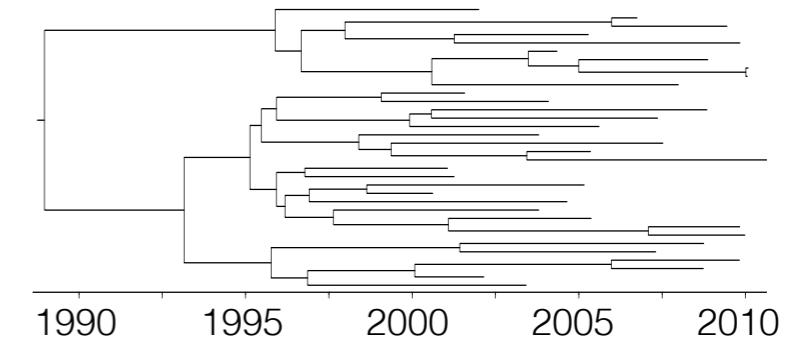
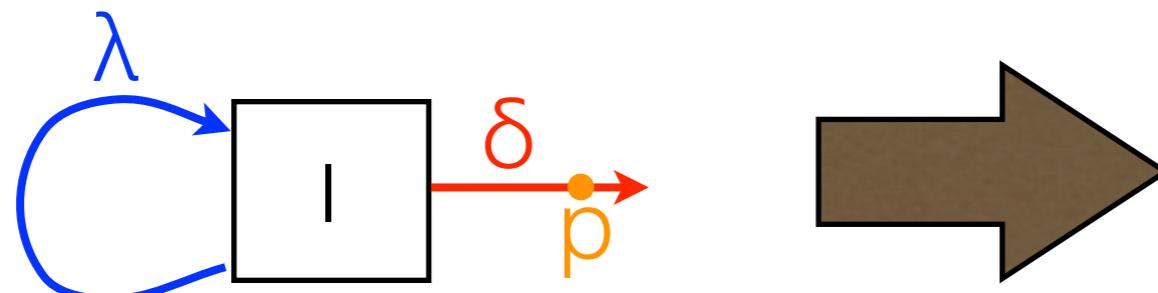
Exact  
method

$\lambda - \delta$  is population size growth rate

- ▶ 95% *HPD* interval contains true value in **97%** of the simulated trees

# Simulation study: Comparing birth-death model to coalescent

Simulating  
epidemic  
outbreak



Exact  
method

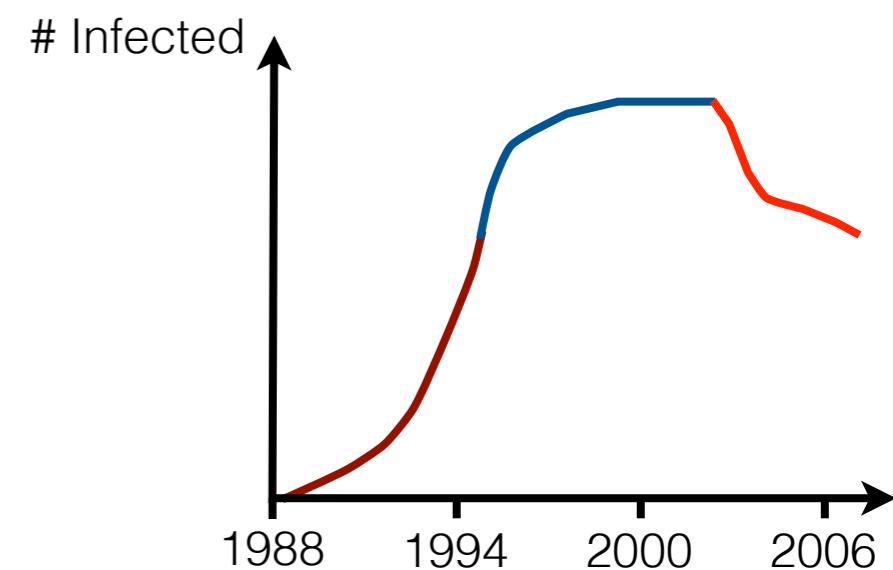
- $\lambda - \delta$  is population size growth rate
- ▶ 95% *HPD* interval contains true value in **97%** of the simulated trees

Coales-  
cent  
approxi-  
mation

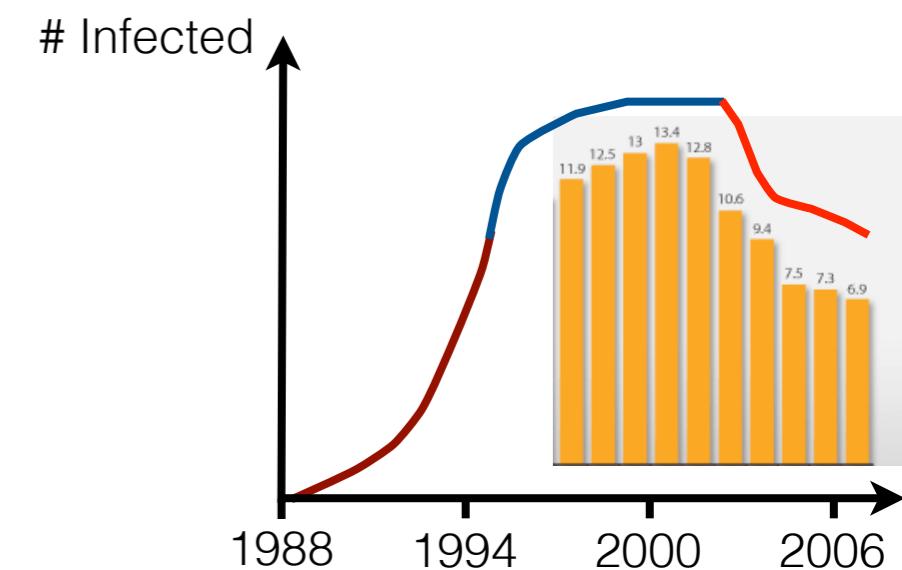
- Parameterizes deterministic population size growth
- ▶ 95% *HPD* interval contains true value in **55%** of the simulated trees

# Phylodynamic models for different dynamics

---



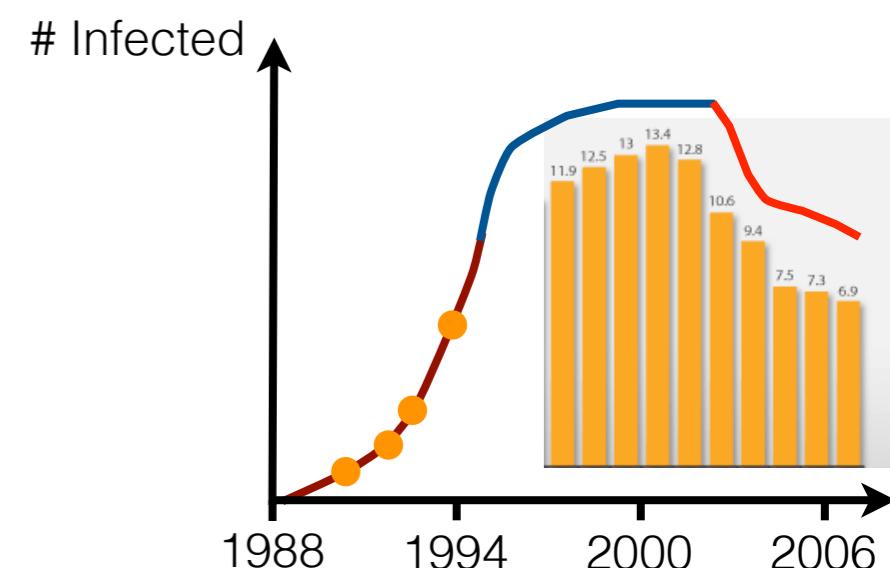
# Phylodynamic models for different dynamics



# Phylodynamic models for different dynamics

## 1. ...for analyzing outbreaks and adaptive radiations

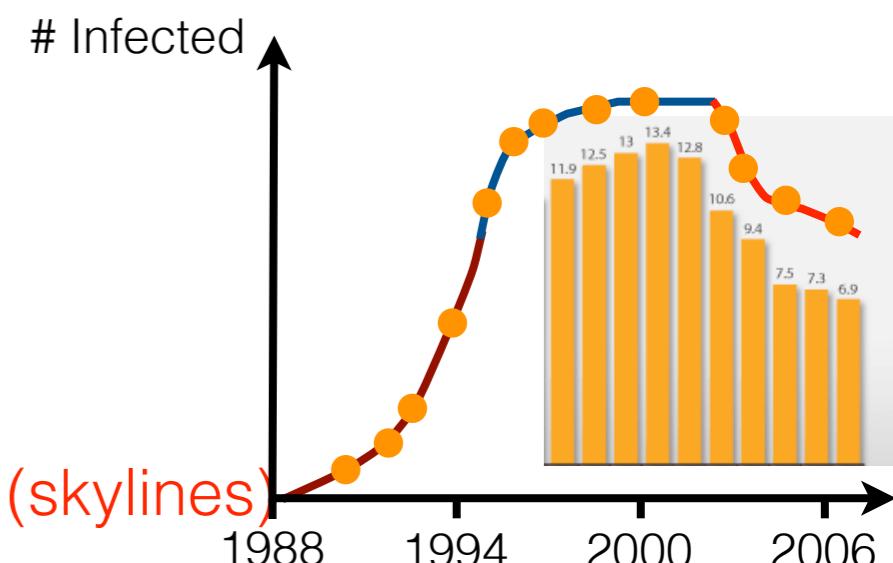
- ▶ C: Drummond et al. (*Genetics*, 2002)
- ▶ BD: Stadler et al. (*MBE*, 2012)



# Phylodynamic models for different dynamics

1. ...for analyzing outbreaks and adaptive radiations
  - ▶ C: Drummond et al. (*Genetics*, 2002)
  - ▶ BD: Stadler et al. (*MBE*, 2012)

2. ...for identifying population dynamic changes over time (skylines)
  - ▶ C: Drummond et al. (*MBE*, 2005)
  - ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)



# Phylodynamic models for different dynamics

1. ...for analyzing outbreaks and adaptive radiations

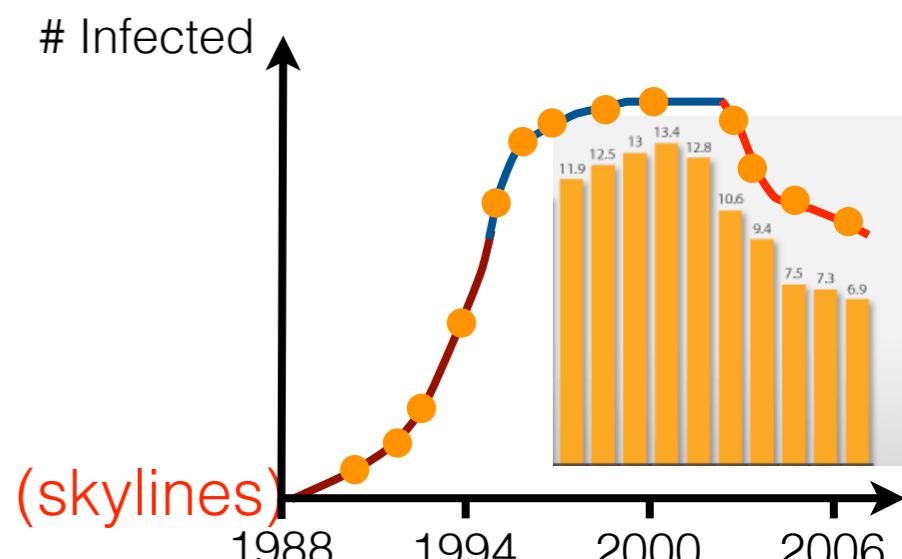
- ▶ C: Drummond et al. (*Genetics*, 2002)
- ▶ BD: Stadler et al. (*MBE*, 2012)

2. ...for identifying population dynamic changes over time (skylines)

- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

3. ...accounting for competition (SIR dynamics or diversity dependence)

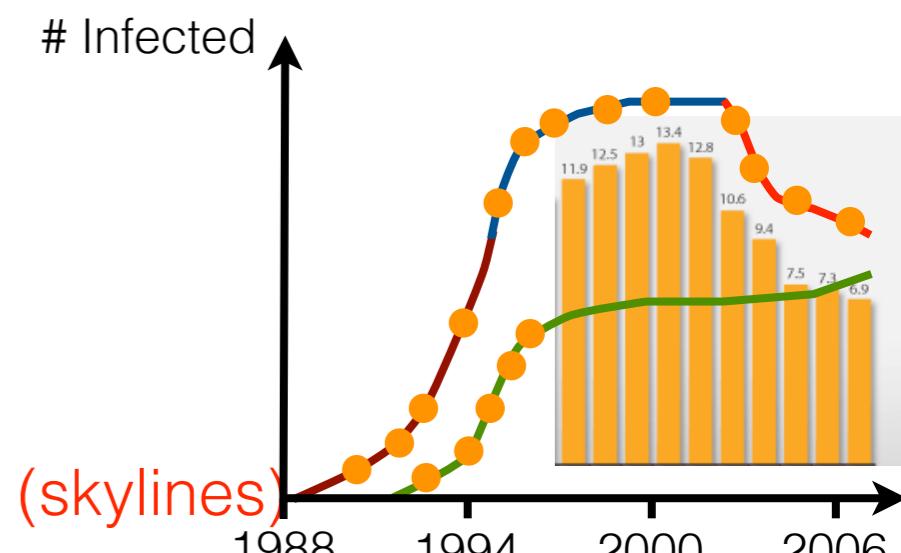
- ▶ C: Volz et al. (*Genetics*, 2009)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014), Leventhal et al. (*MBE*, 2014), , Vaughan et al. (*MBE*, 2019)



# Phylodynamic models for different dynamics

## 1. ...for analyzing outbreaks and adaptive radiations

- ▶ C: Drummond et al. (*Genetics*, 2002)
- ▶ BD: Stadler et al. (*MBE*, 2012)



## 2. ...for identifying population dynamic changes over time (skylines)

- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

## 3. ...accounting for competition (SIR dynamics or diversity dependence)

- ▶ C: Volz et al. (*Genetics*, 2009)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014), Leventhal et al. (*MBE*, 2014), , Vaughan et al. (*MBE*, 2019)

## 4. ...accounting for population structure

- ▶ C: Volz (*Genetics*, 2012), Vaughan et al. (*Bioinf.*, 2014), Müller et al. (*MBE*, 2017)
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013), Kühnert et al. (*MBE*, 2016)

# Phylodynamic models for different dynamics



1. ...for analyzing outbreaks and adaptive radiations

▶ C: Drummond et al. (*Genetics*, 2002)  
▶ BD: Stadler et al. (*MBE*, 2012)

2. ...analyzing population dynamic changes over time (skylines)

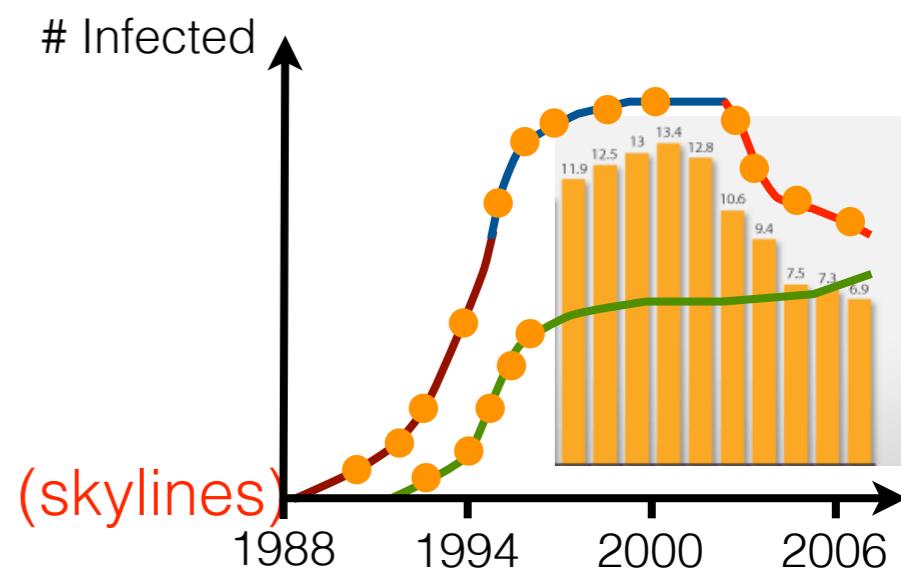
▶ C: Drummond et al. (*MBE*, 2005)  
▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

3. ...accounting for competition (SIR dynamics or diversity dependence)

▶ C: Volz et al. (*Genetics*, 2009)  
▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014), Leventhal et al. (*MBE*, 2014), , Vaughan et al. (*MBE*, 2019)

4. ...accounting for population structure

▶ C: Volz (*Genetics*, 2012), Vaughan et al. (*Bioinf.*, 2014), Müller et al. (*MBE*, 2017)  
▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013), Kühnert et al. (*MBE*, 2016)

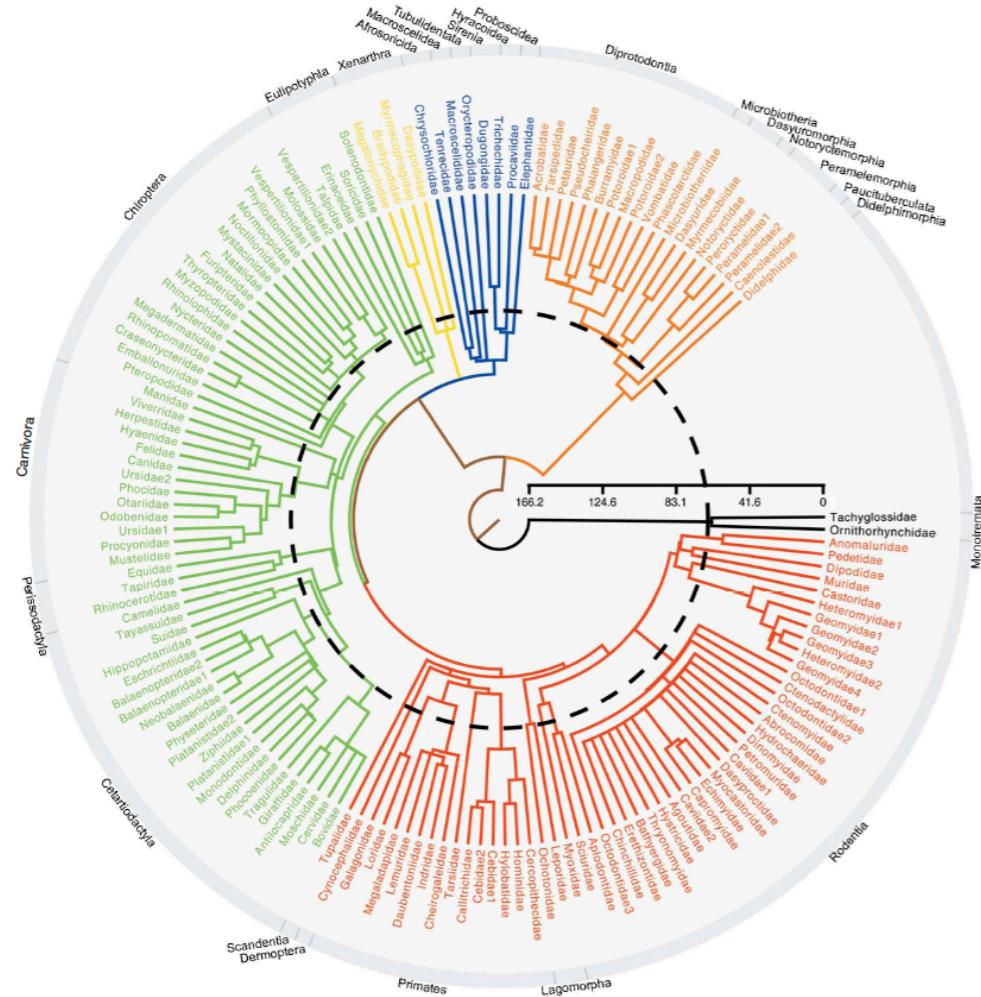


# Outline

---

- Why phylodynamics?
  - ▶ A macroevolutionary motivation
  - ▶ An epidemiological motivation
- A phylodynamic model and the math behind
  - ▶ Birth-death models
  - ▶ The coalescent
- Applications in Macroevolution
- Application in Epidemiology: SARS-CoV-2
- BEAST2 Tutorial

# Species phylogeny rejects increased mammalian diversification at KT-boundary

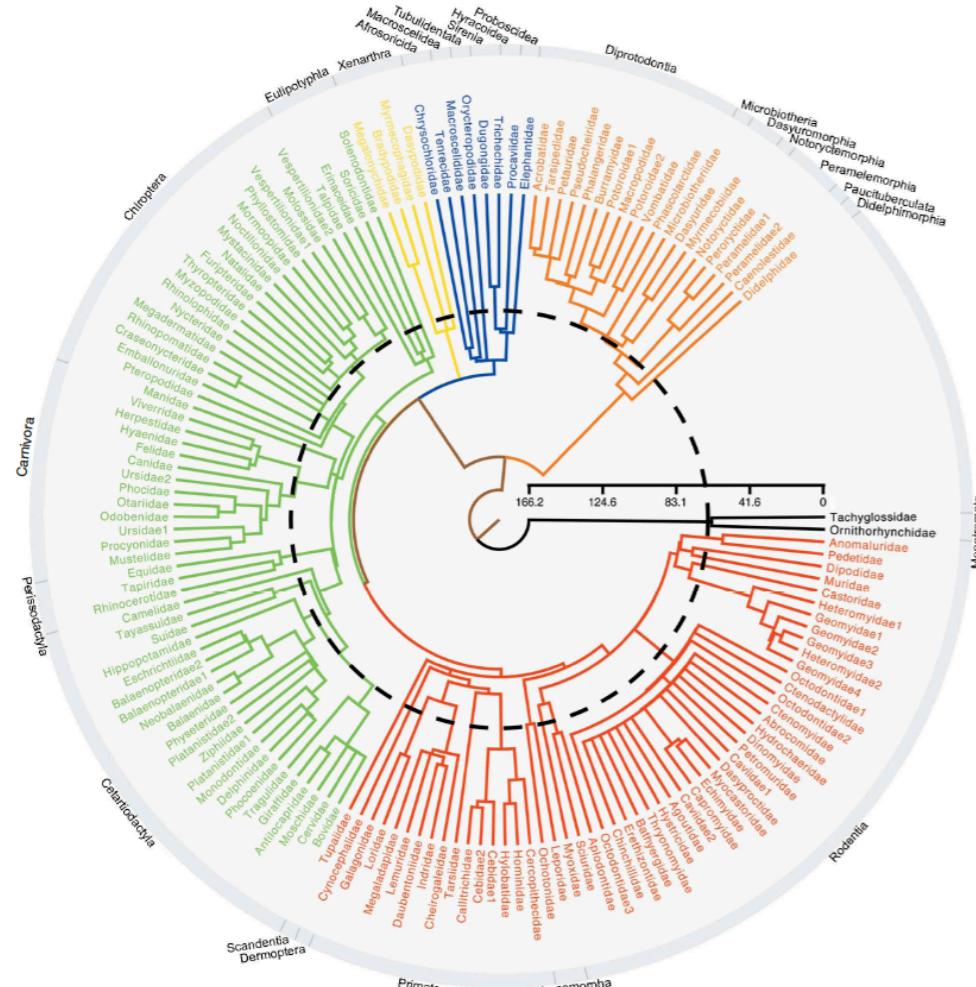


Mammalian phylogeny

(from Bininda-Emonds et al., Nature, 2007)

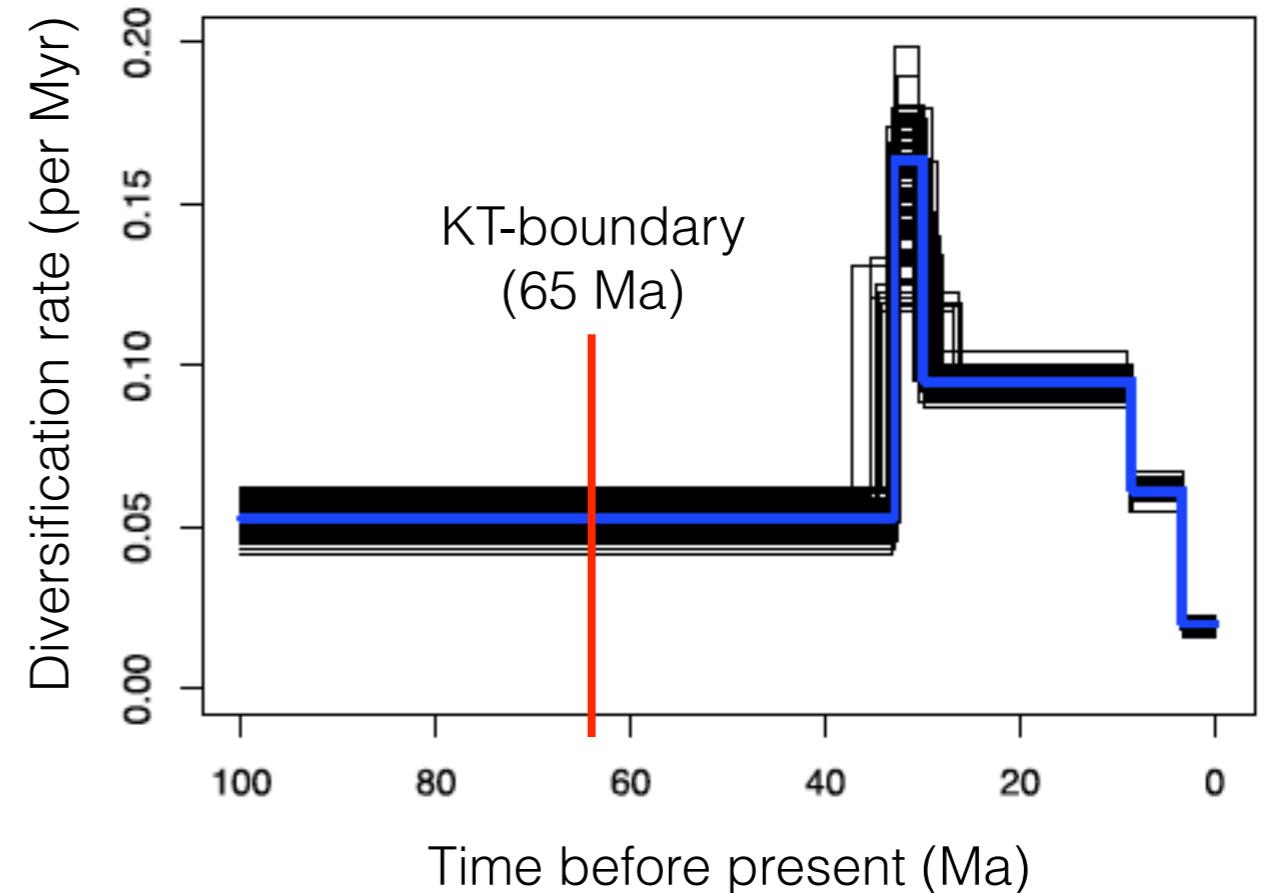
# Species phylogeny rejects increased mammalian diversification at KT-boundary

(diversification = speciation  $\lambda$  - extinction  $\mu$ )



Mammalian phylogeny

(from Bininda-Emonds et al., Nature, 2007)

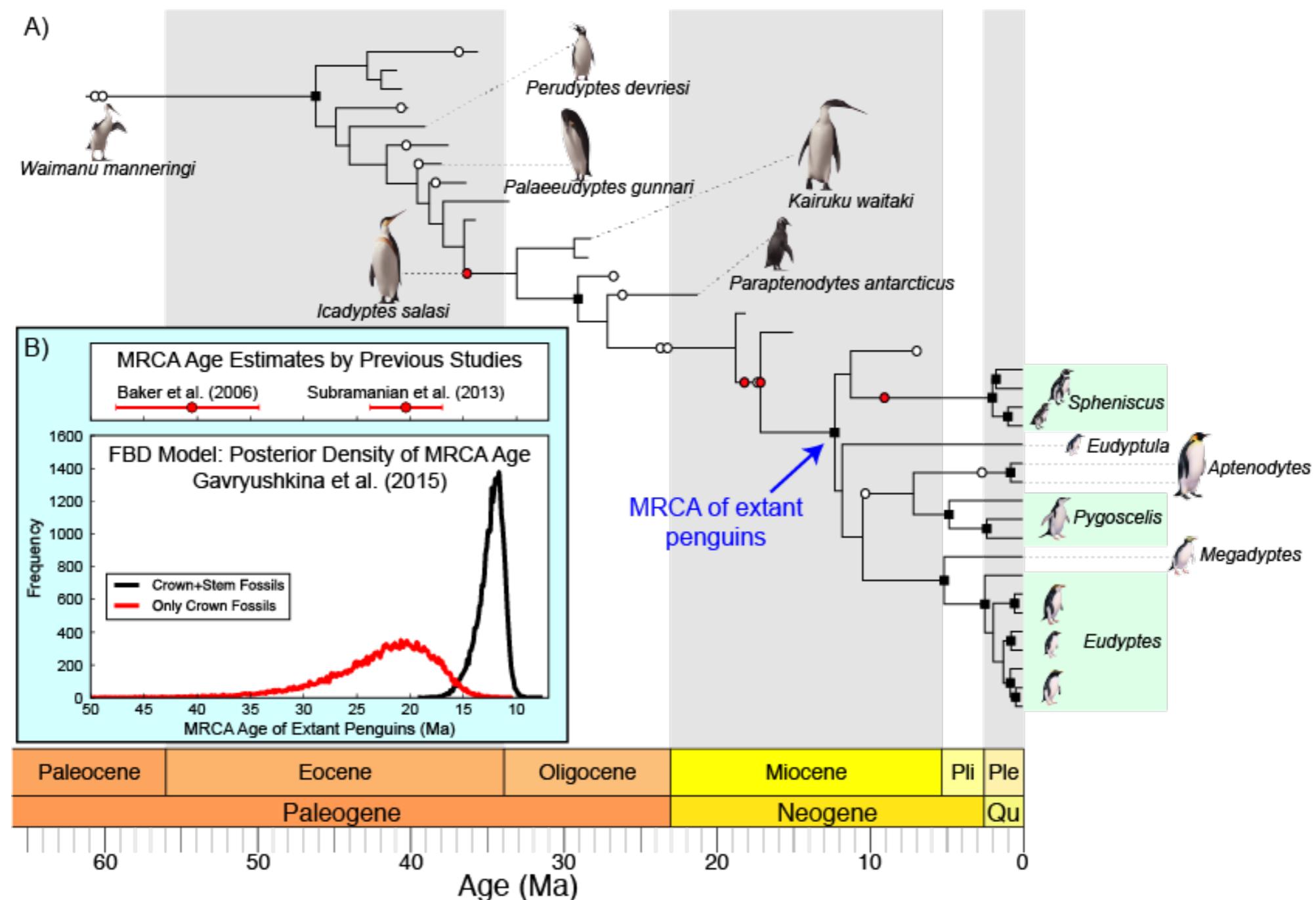


**Null model:** constant rates bad fit

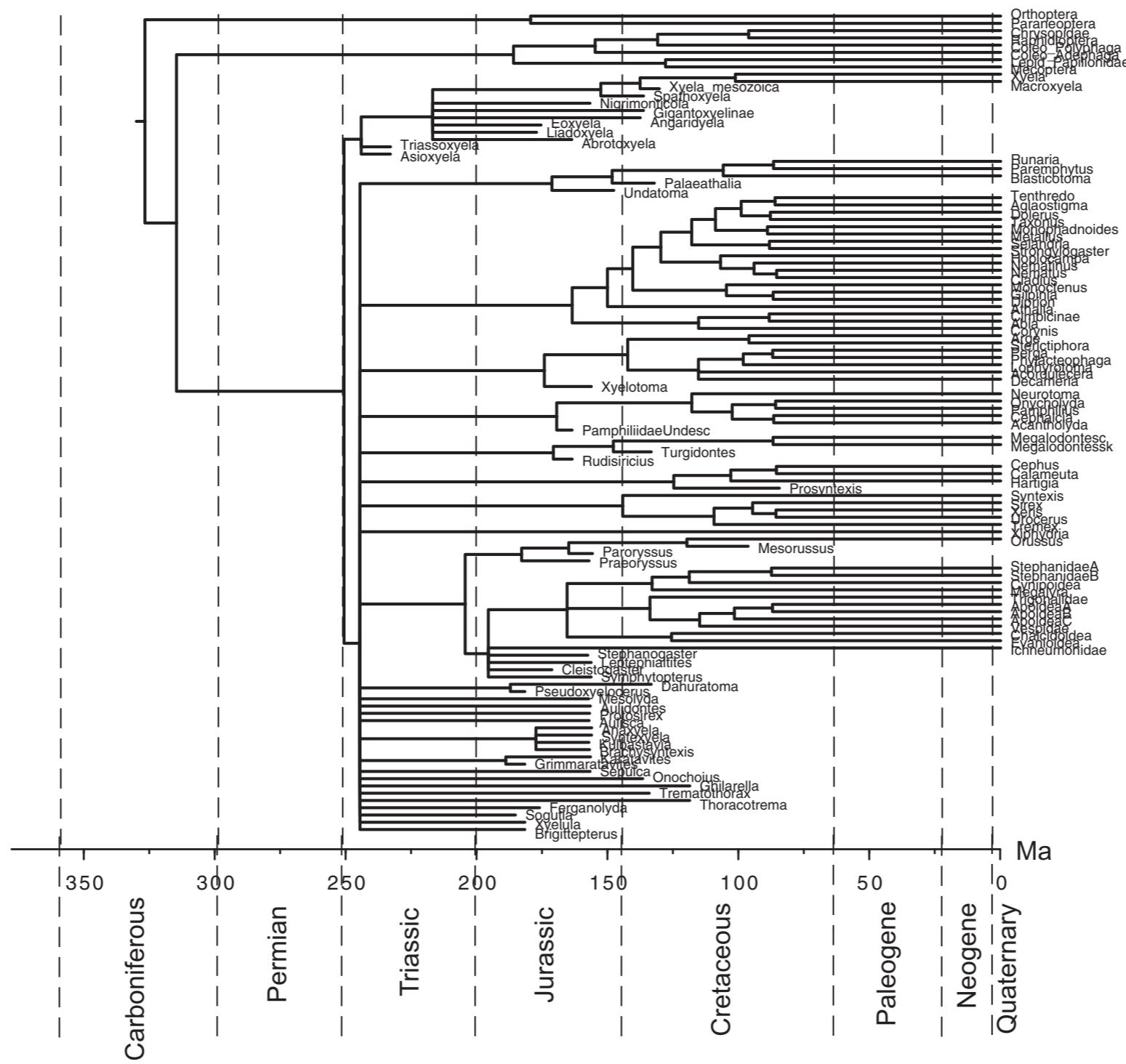
**Environmental model:** 4 shifts ( $p=1.0$ )

**Simulation results**

# Penguin phylogeny



# Hymenoptera phylogeny

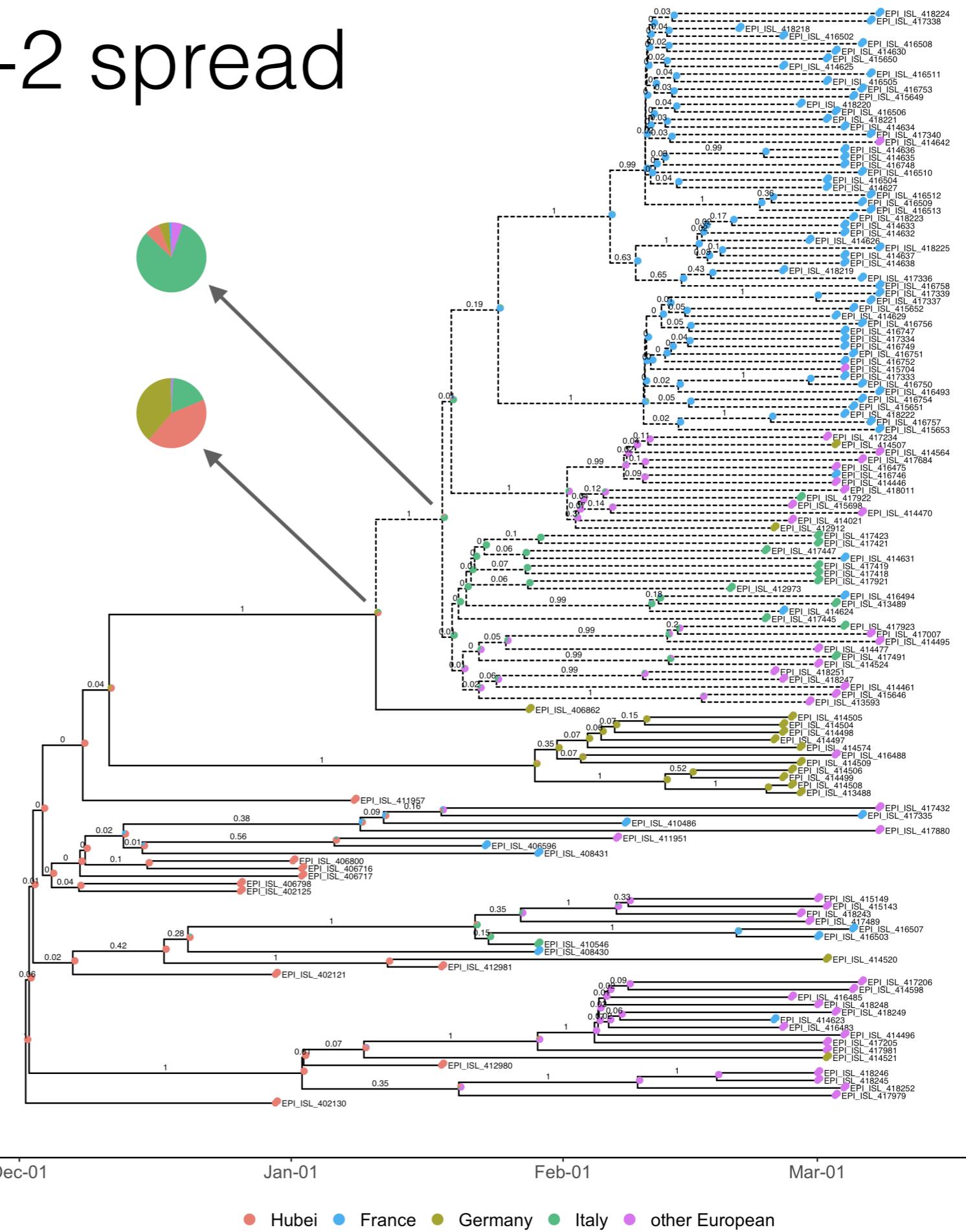


# Outline

---

- Why phylodynamics?
  - ▶ A macroevolutionary motivation
  - ▶ An epidemiological motivation
- A phylodynamic model and the math behind
  - ▶ Birth-death models
  - ▶ The coalescent
- Applications in Macroevolution
- Application in Epidemiology: SARS-CoV-2
- BEAST2 Tutorial

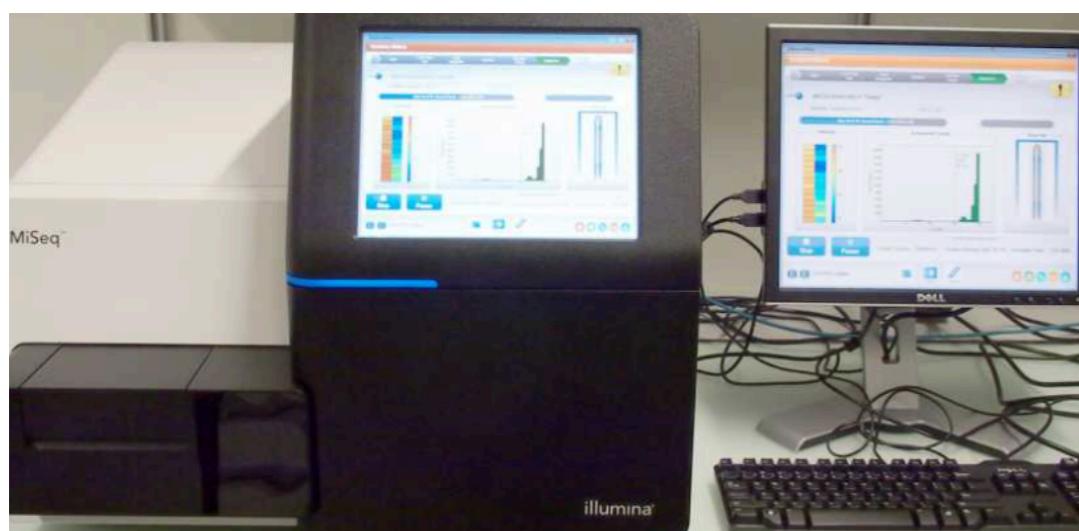
# Initial SARS-CoV-2 spread in Europe



# Swiss SARS-CoV-2 Sequencing Consortium (S3C)

Start in 03/2020

Meanwhile ~25000 Sequenzen published; almost 75% of all Swiss SARS-CoV-2 genomes.



## Members

**Genomic and epidemiological analysis & overall coordination:** Computational Evolution, D-BSSE, ETHZ. Chaoran Chen, Sarah Nadeau.

**Samples:** Violler AG. Christiane Beckmann, Maurice Redondo, Olivier Kobel, Christoph Noppen with initial ETH support by Sophie Seidel, Noemie Santamaria de Souza.

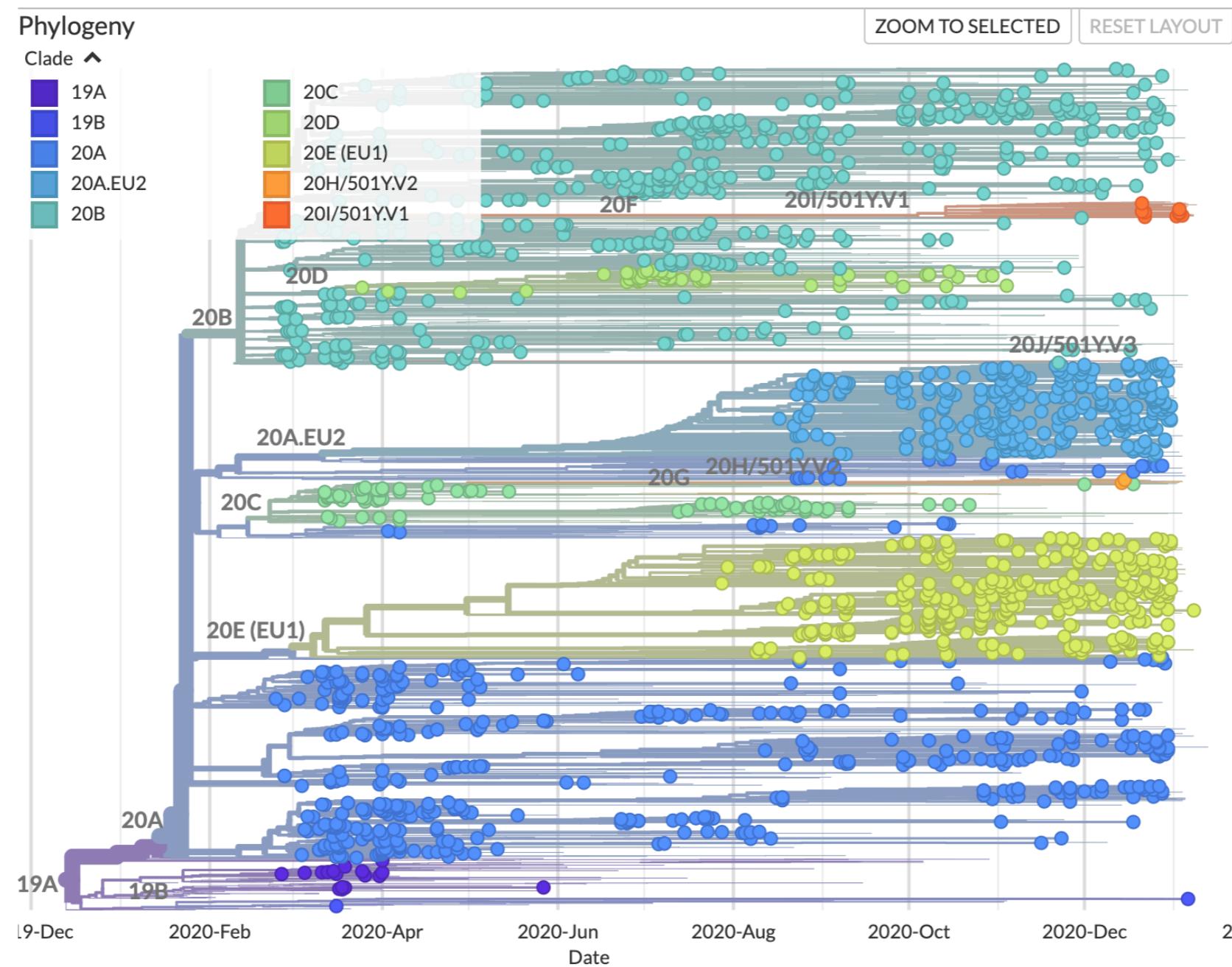
**Bioinformatics:** Computational Biology, D-BSSE, ETHZ . Niko Beerenwinkel, Ivan Topolsky, Philipp Jablonski, Lara Fuhrmann, David Dreifuss, Katharina Jahn, Pedro Ferreira, Susana Posada-Céspedes.

## Sequencing:

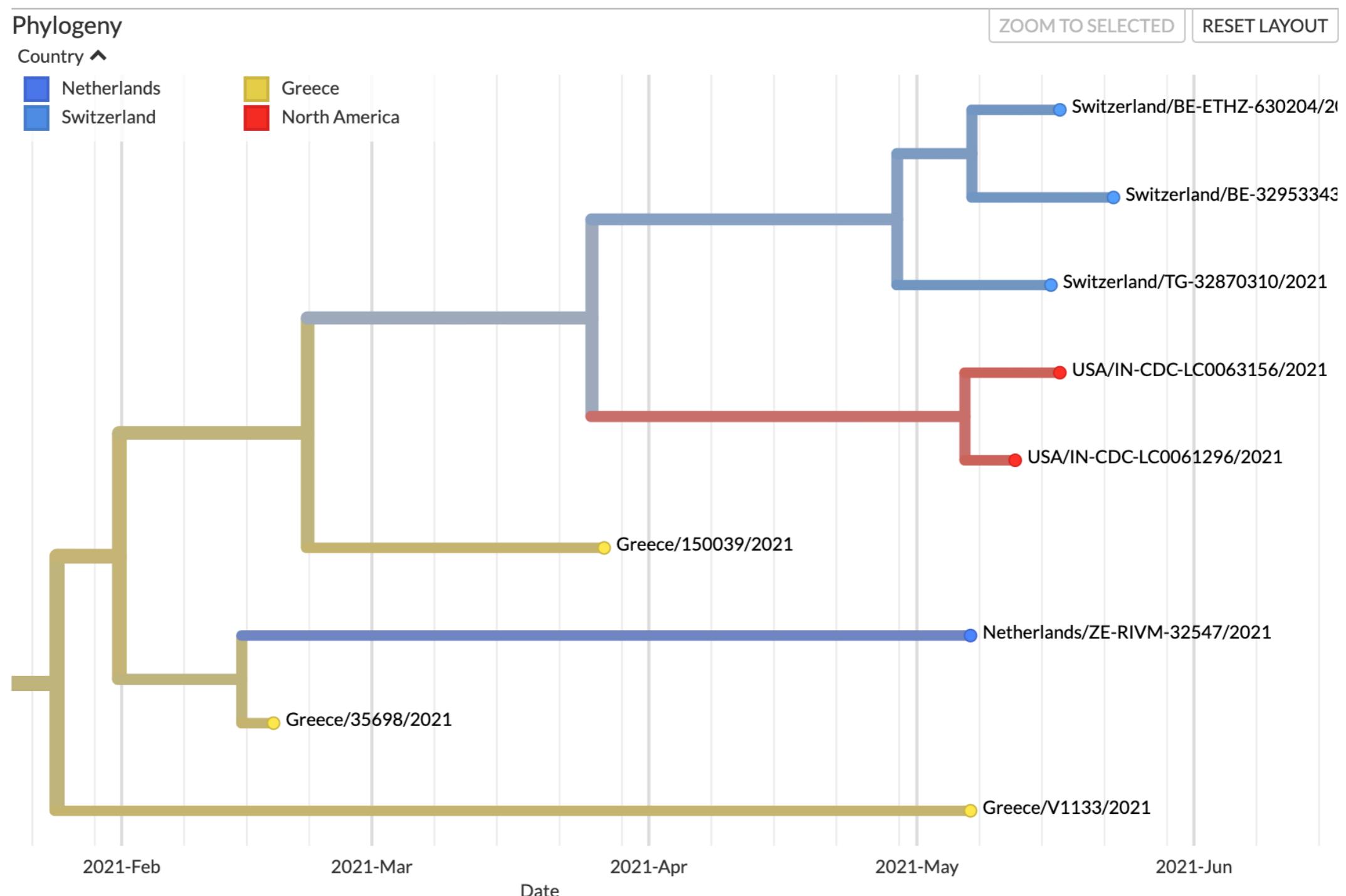
- **Genomic Facility Basel, ETHZ.** Christian Beisel, Rebecca Denes, Mirjam Feldkamp, Ina Nissen, Natascha Santacroce, Elodie Burcklen.
- **Functional Genomic Center Zurich, ETHZ.** Catharine Aquino, Andreia Cabral de Gouvea, Maria Domenica Moccia, Simon Grüter, Timothy Sykes, Lennart Opitz, Griffin White, Laura Neff, Doris Popovic, Andrea Patrignani, Jay Tracy, Ralph Schlapbach.
- **Health 2030 Genome Center, Geneva.** Emmanouil Dermitzakis, Keith Harshman, Ioannis Xenarios, Henri Pegeot, Lorenzo Cerutti, Deborah Penet.

**Lead:** Computational Evolution, D-BSSE, ETHZ. Tanja Stadler.

# SARS-CoV-2 phylogeny



# Zooming into the tree



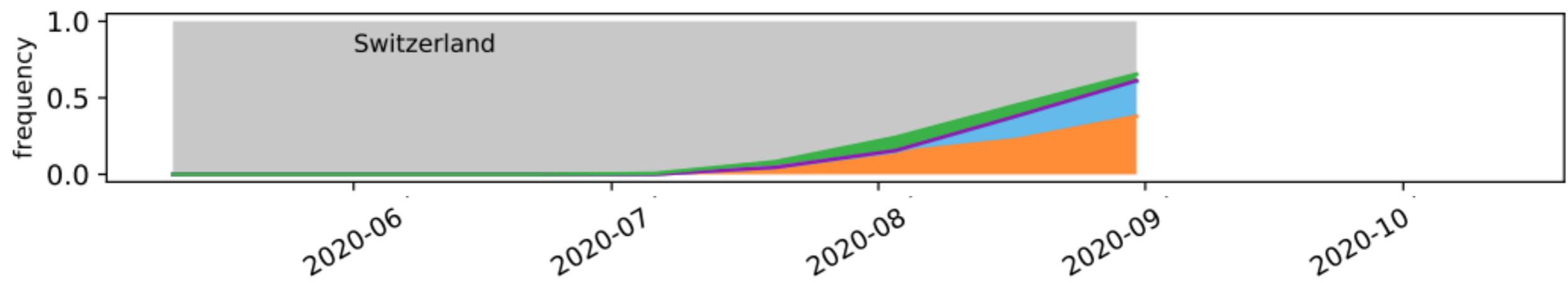
# SARS-CoV-2 Variants of Concern

---

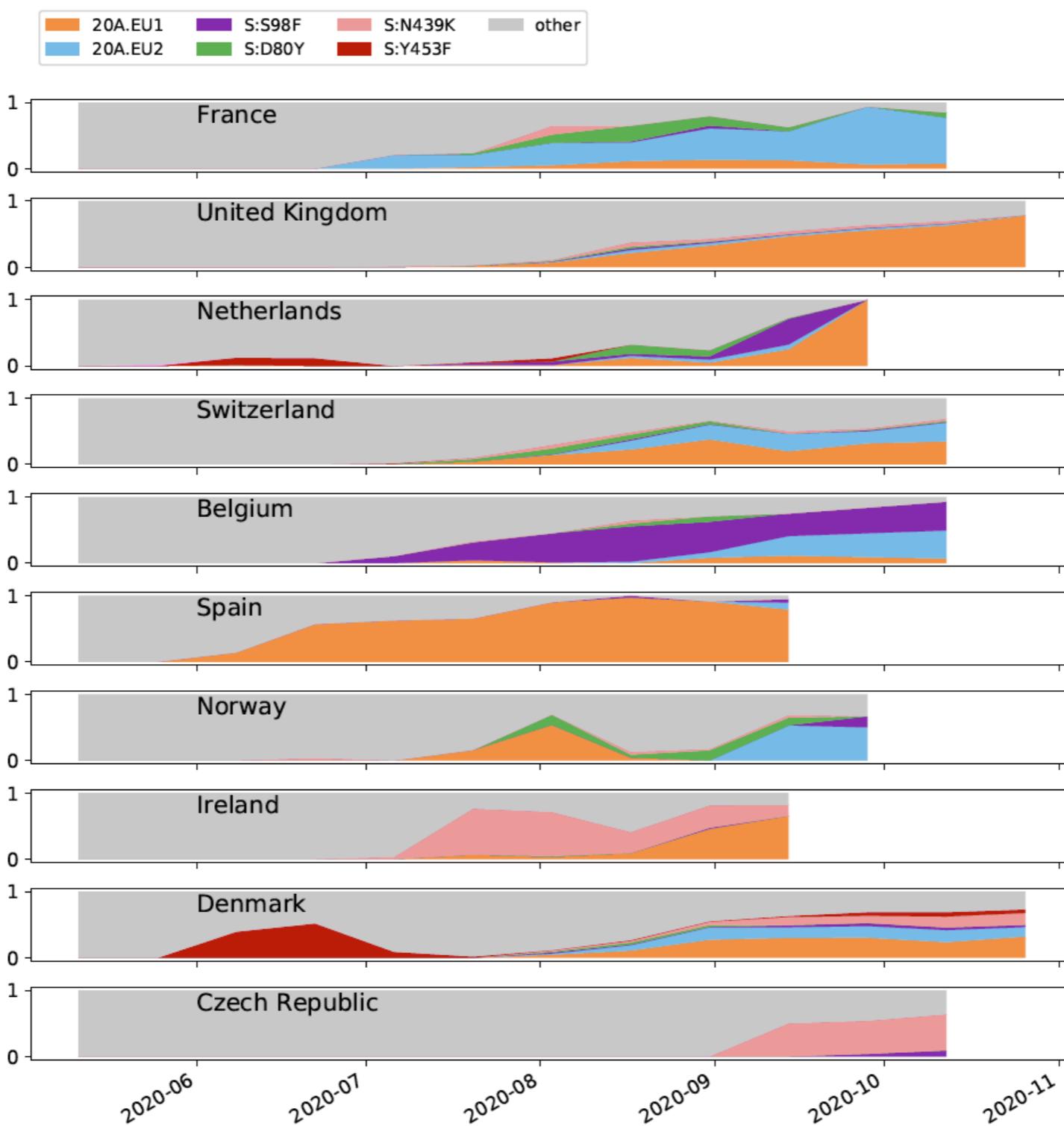
- Higher transmissibility or
- More severe disease outcome or
- Immune escape

# Did we find a VOC over summer 2020?

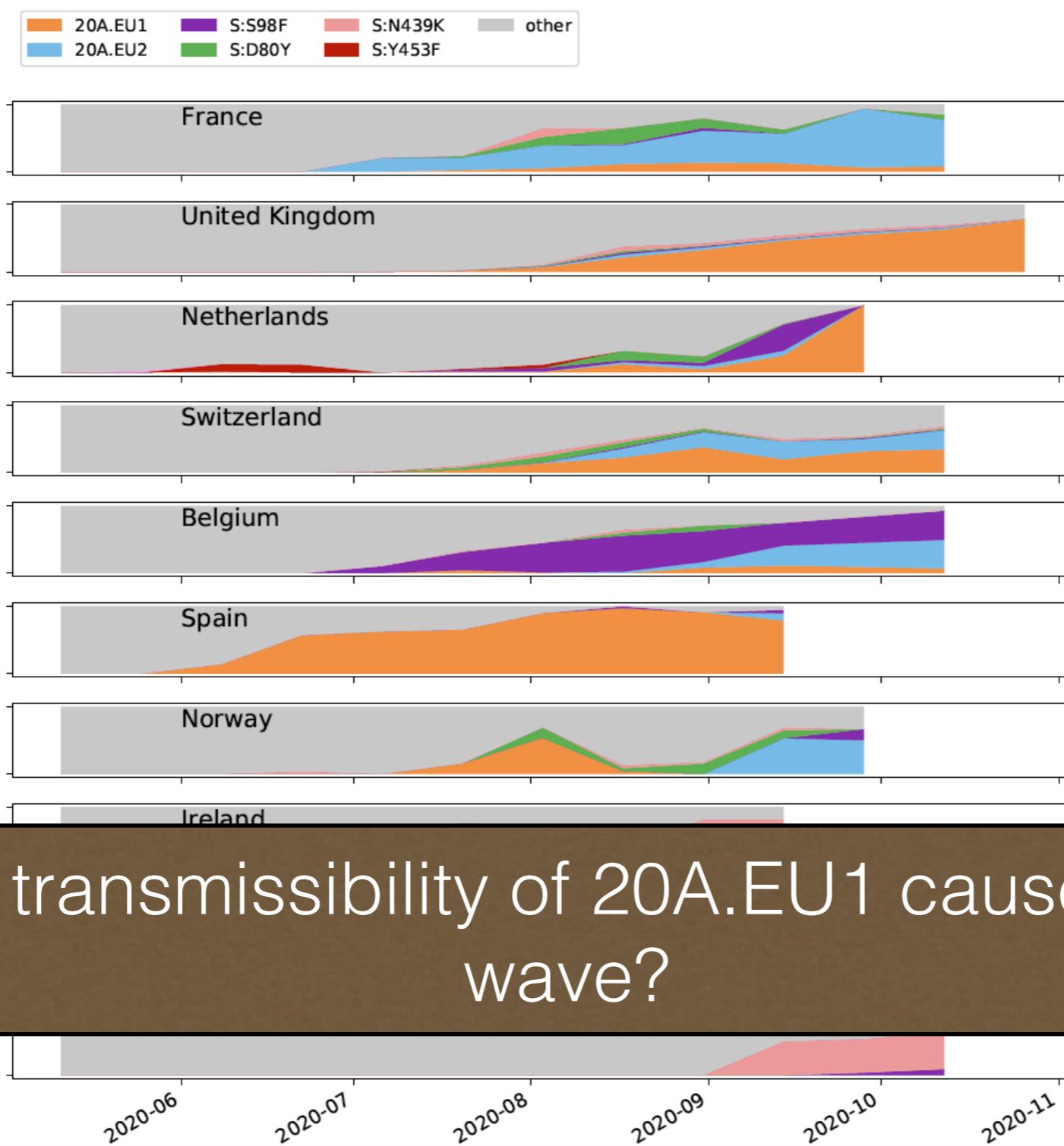
---



# Did we find a VOC over summer 2020?



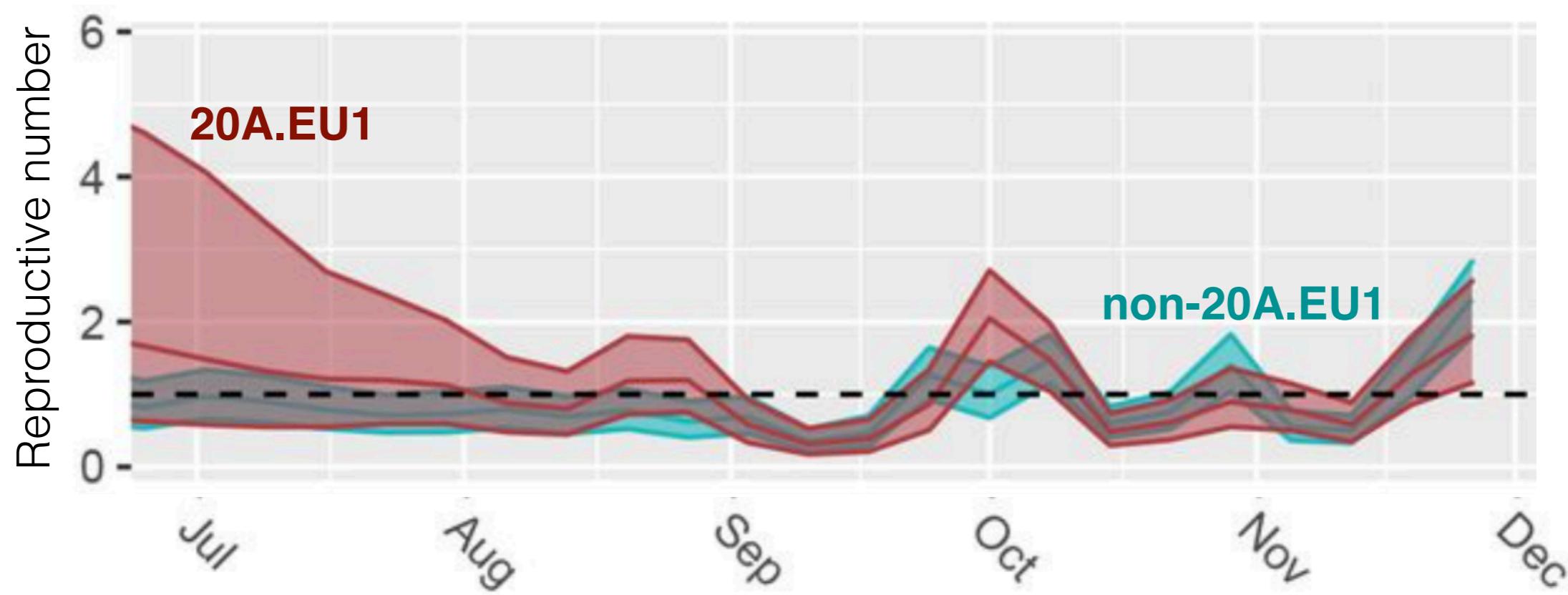
# Did we find a VOC over summer 2020?



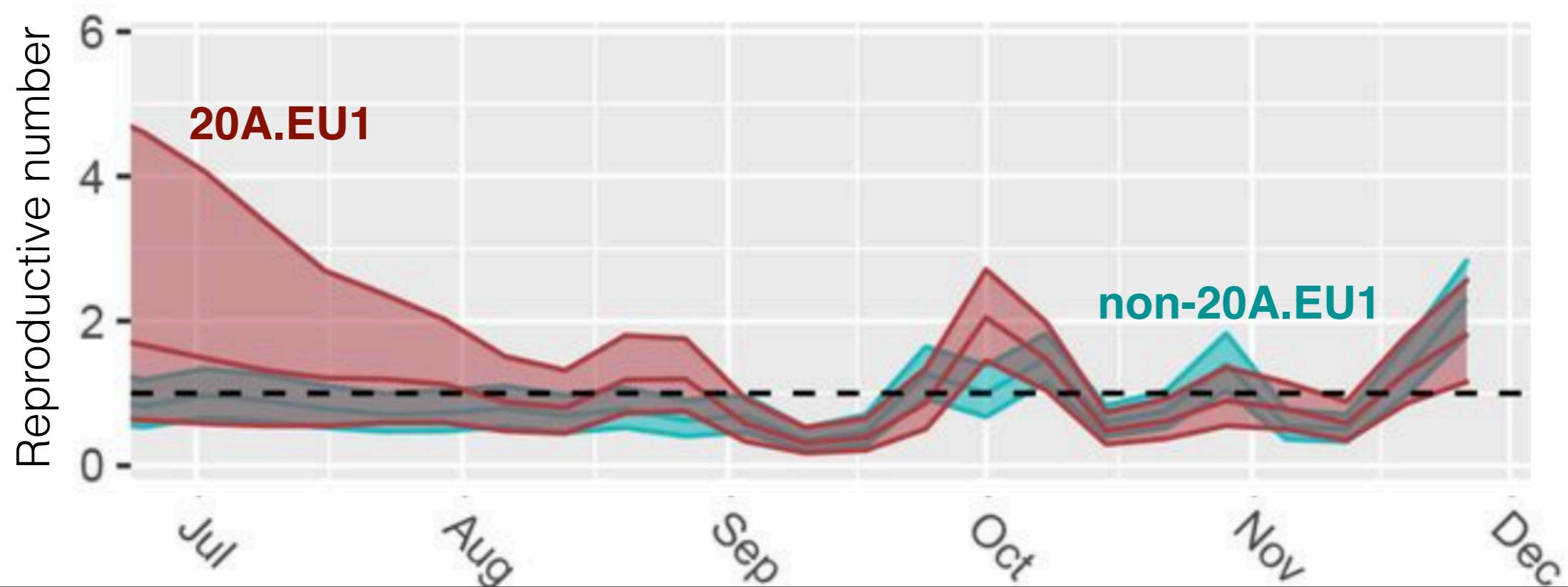
Did higher transmissibility of 20A.EU1 cause the second wave?

# Reproductive number for Switzerland

---

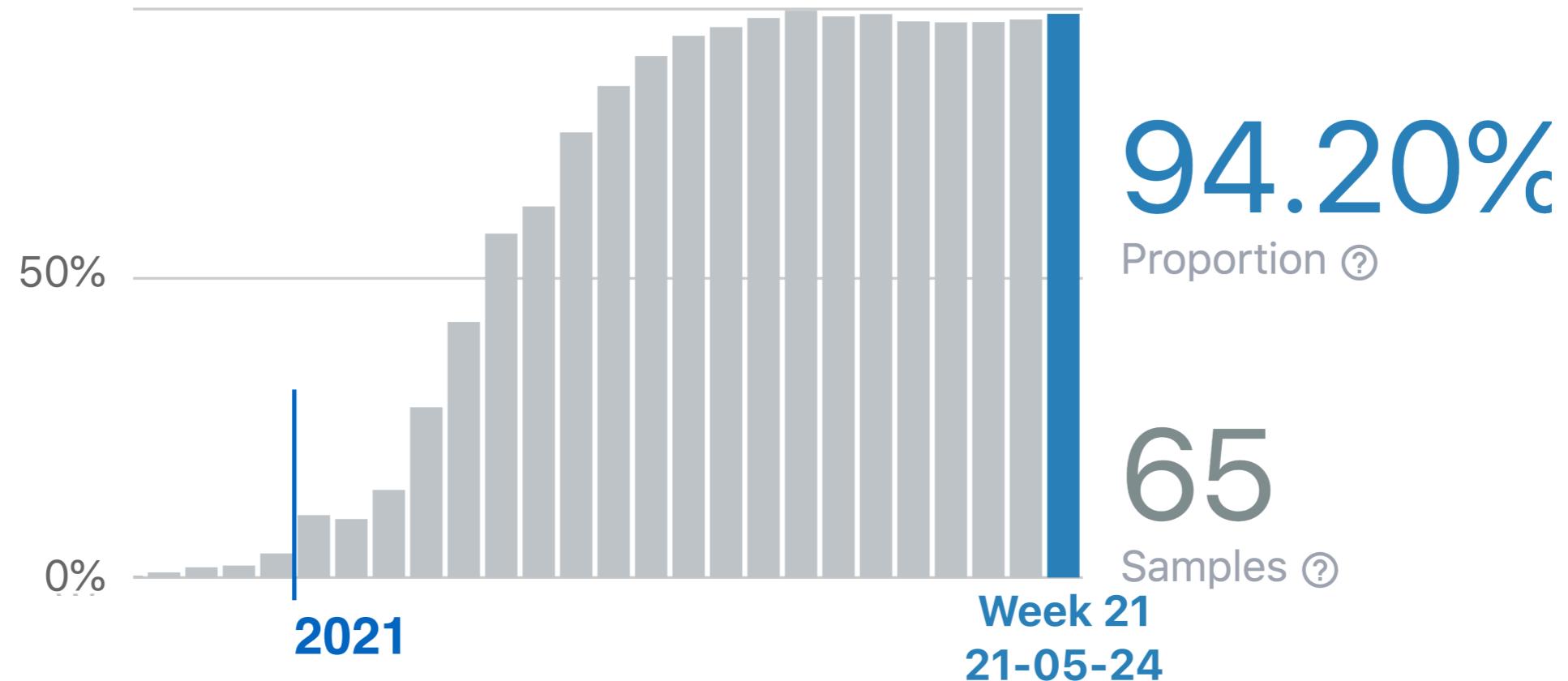


# Reproductive number for Switzerland



Fall and winter dynamics are not driven by the new variant!  
20A.EU1 is not a VOC.

# Variant B.1.1.7 has a transmission advantage and became dominant in CH



# Outline

---

- Why phylodynamics?
  - ▶ A macroevolutionary motivation
  - ▶ An epidemiological motivation
- A phylodynamic model and the math behind
  - ▶ Birth-death models
  - ▶ The coalescent
- Applications in Macroevolution
- Application in Epidemiology: SARS-CoV-2
- BEAST2 Tutorial

# Key tutorials in BEAST2 related to this talk

## Taming the BEAST



news workshops tutorials contribute 

### Skyline plots

Inference of past population dynamics using Bayesian Coalescent Skyline and Birth-Death Skyline plots.  
by Nicola F. Müller and Louis du Plessis

**Tutorial**  
 [Skyline-plots.pdf](#)  
 [Github repository](#)  
 [License](#)  
 [Statistics](#)

**Data**  
 [hcv.nexus](#)

**XML**

### Background

Population dynamics influence the shape of the tree and consequently, the shape of the tree contains some information about past population dynamics. The so-called Skyline methods allow to extract this information from phylogenetic trees in a non-parametric manner. It is non-parametric since there is no underlying system of differential equations governing the inference of these dynamics. In this tutorial we will look at two different methods to infer these dynamics from sequence data. The first one is the Coalescent Bayesian Skyline plot (Drummond, Rambaut, Shapiro, & Pybus, 2005), which is based on the coalescent model, and the second one is the Birth-Death Skyline plot (Stadler, Kuhnert, Bonhoeffer, &

## Taming the BEAST



news workshops tutorials contribute 

### Divergence Time Estimation using BEAST v2.X

Dating Species Divergences with the Fossilized Birth-Death Process  
by Tracy A. Heath

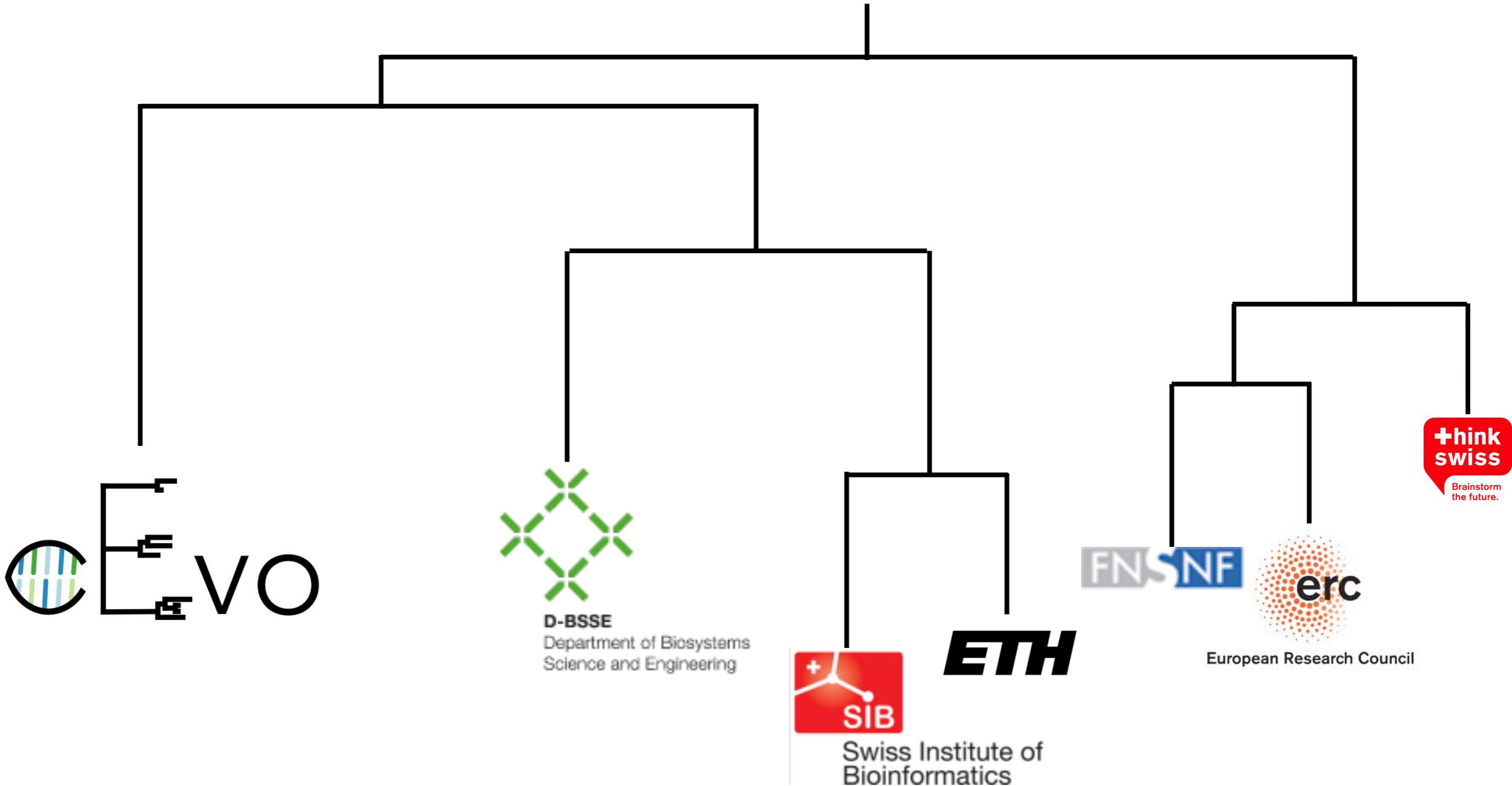
**Tutorial**  
 [FBD-tutorial.pdf](#)  
 [Github repository](#)  
 [License](#)  
 [Statistics](#)

**Data**  
 [bears\\_irbp\\_fossils.nex](#)  
 [bears\\_cytb\\_fossils.nex](#)

Central among the questions explored in biology are those that seek to understand the timing and rates of evolutionary processes. Accurate estimates of species divergence times are vital to understanding historical biogeography, estimating diversification rates, and identifying the causes of variation in rates of molecular evolution.

This tutorial will provide a general overview of divergence time estimation and fossil calibration using a stochastic branching process and relaxed-clock model in a Bayesian framework. The exercise will guide you through the steps necessary for estimating phylogenetic relationships and dating species divergences using the program BEAST v2.X.

# Phylogeny of Acknowledgements



Prof. Dr. Tanja Stadler  
Associate Professor for Computational Evolution  
Chair expert group Data & Modelling, Swiss National COVID-19 Science Task Force  
[tanja.stadler@bsse.ethz.ch](mailto:tanja.stadler@bsse.ethz.ch)

**ETH**  
Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich