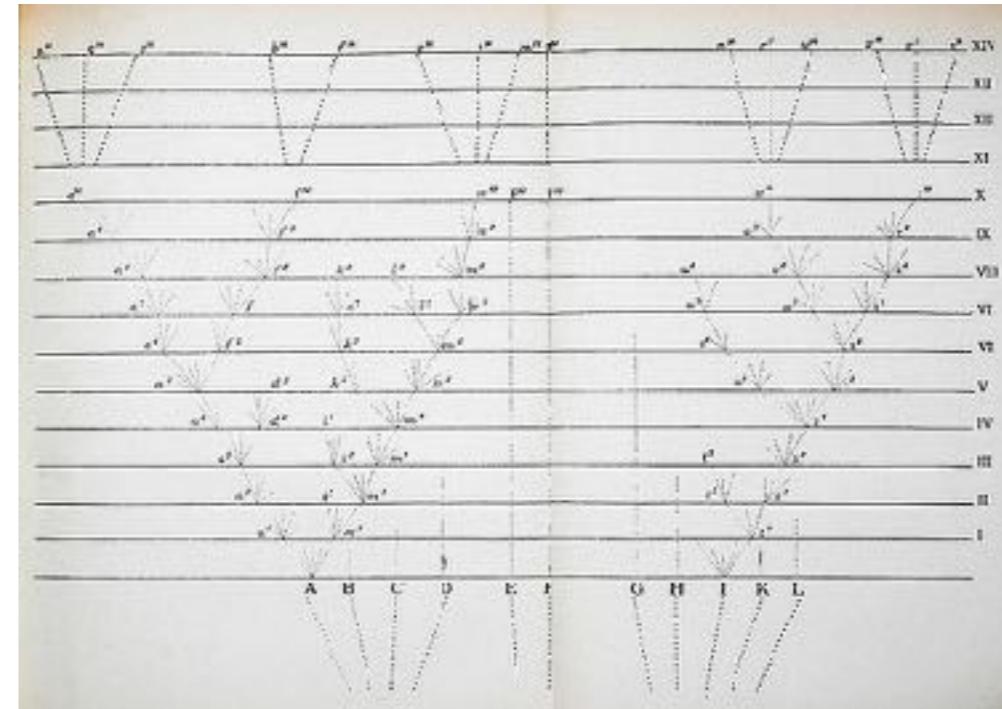
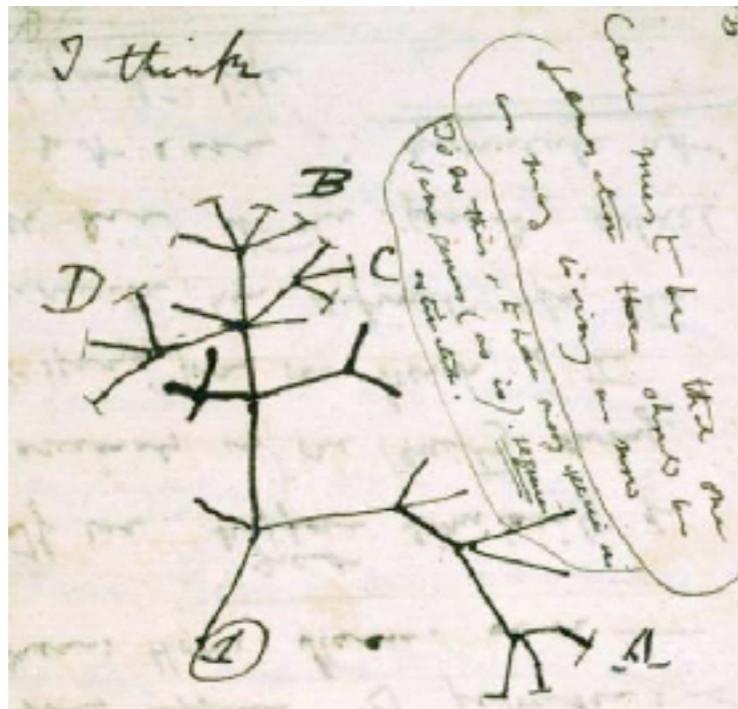


# Phylogenetics

Tree priors based on  
the birth-death model and the coalescent

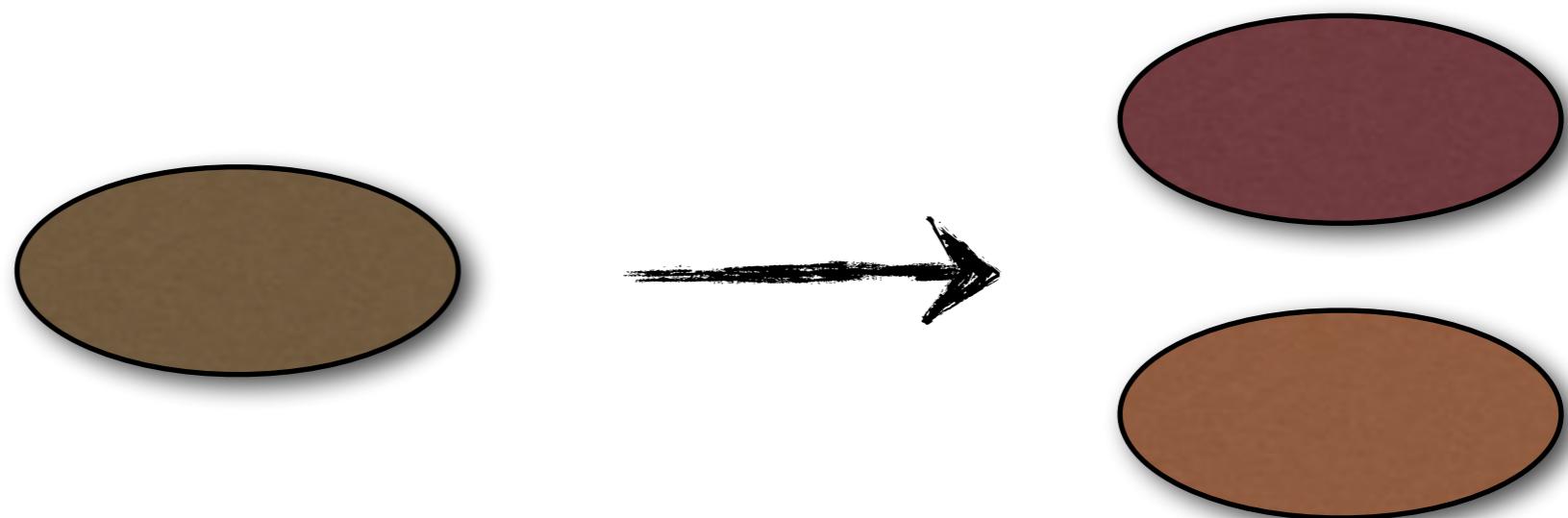


# Basis of biology: Reproduction & genetic change

---

## **Scale:**

DNA  
Virion  
Prokaryote  
Eukaryotic cell  
Eukaryote  
Species  
Infected host  
Immune cells

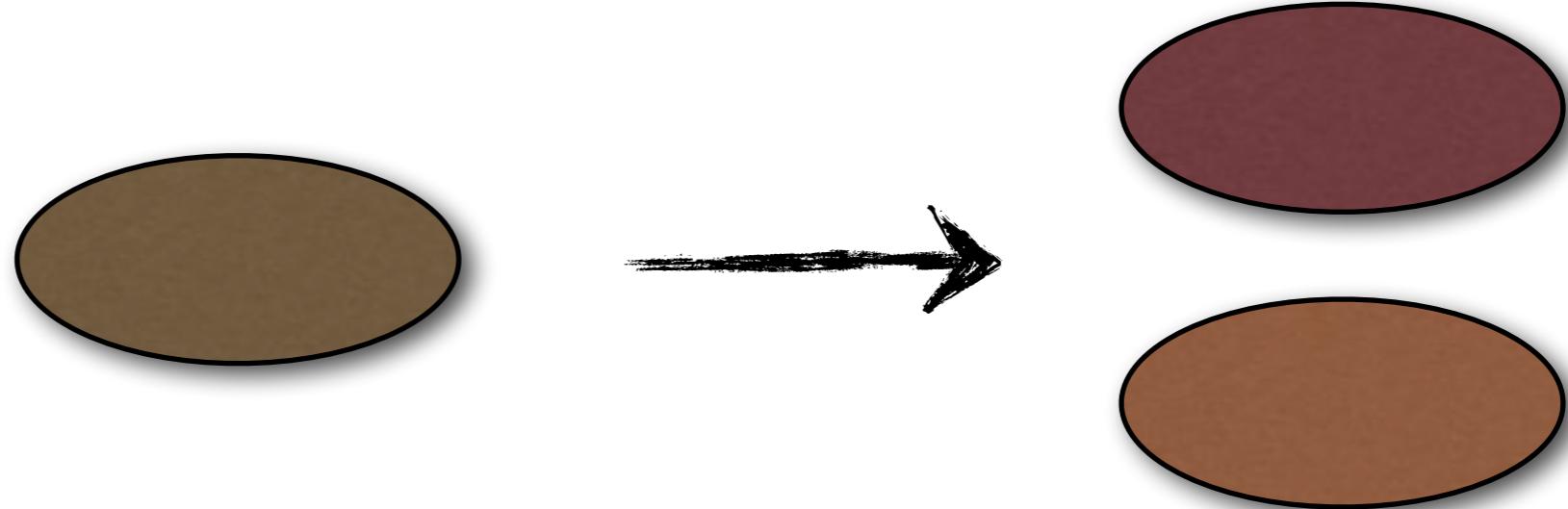


# Basis of biology: Reproduction & genetic change

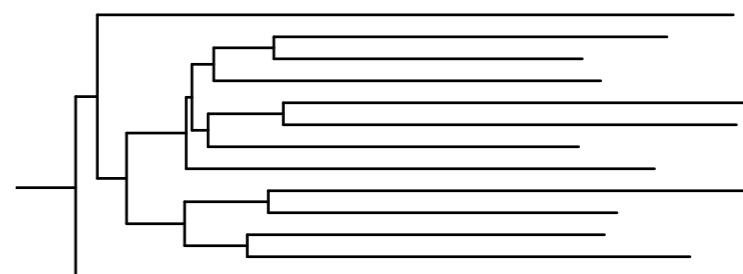
---

## Scale:

DNA  
Virion  
Prokaryote  
Eukaryotic cell  
Eukaryote  
Species  
Infected host  
Immune cells



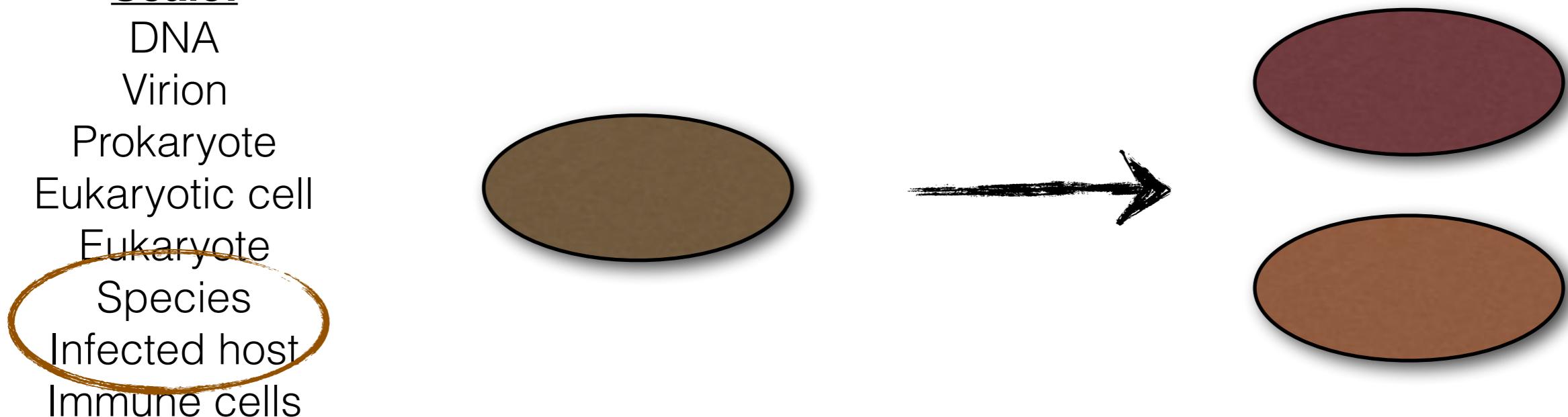
- ▶ Lineage tree (=phylogeny) represents **reproduction** process on any of these scales
- ▶ **Genetic** information changes along the tree



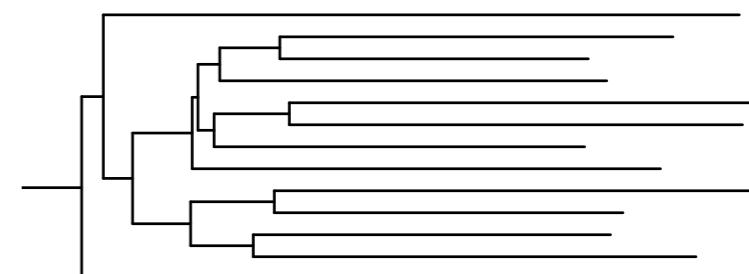
# Basis of biology: Reproduction & genetic change

## 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



# Phylogenetic model

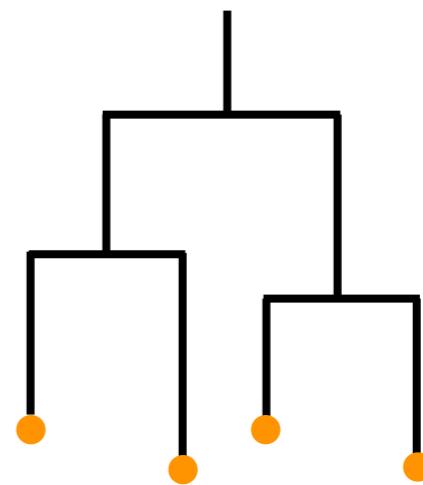
---

Population  
model

Growth of tree

**(Population dynamics - replication)**

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



# Phylogenetic model

---

Population  
model

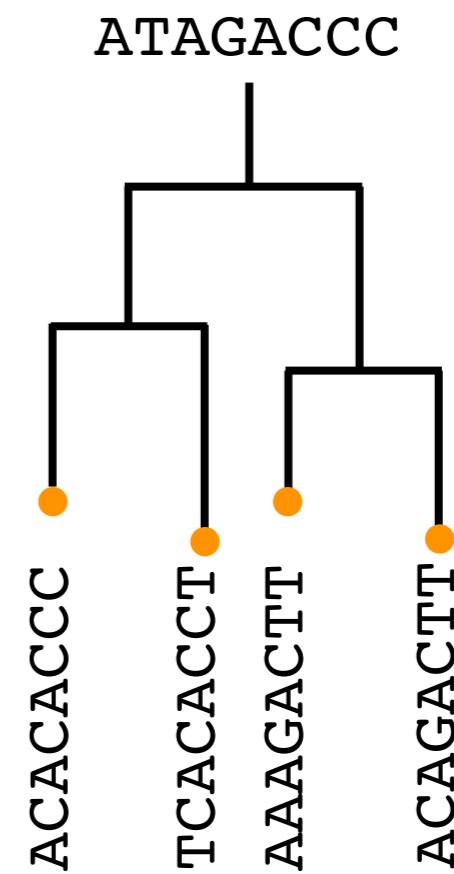
Growth of tree  
**(Population dynamics - replication)**

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

Evolution  
model

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

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



# Phylogenetic model

---

Population  
model

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

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

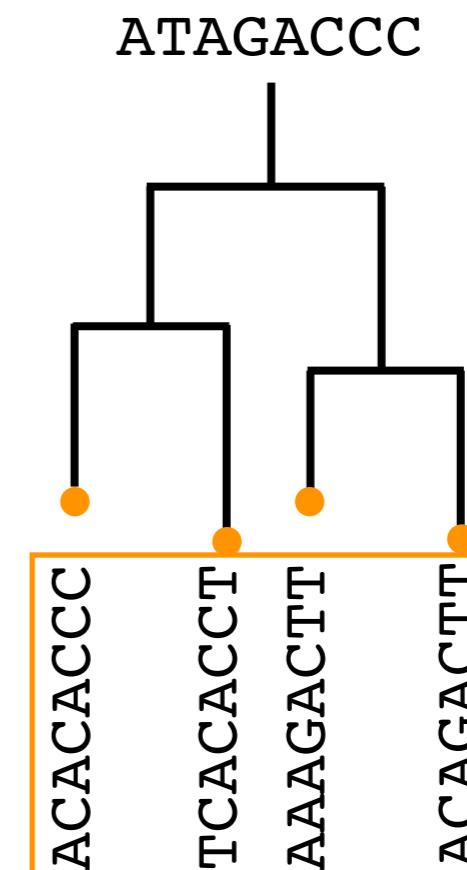
Evolution  
model

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

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

Data

Sequence alignment



# Phylogenetic model

Population  
model

Growth of tree

**(Population dynamics - replication)**

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

Evolution  
model

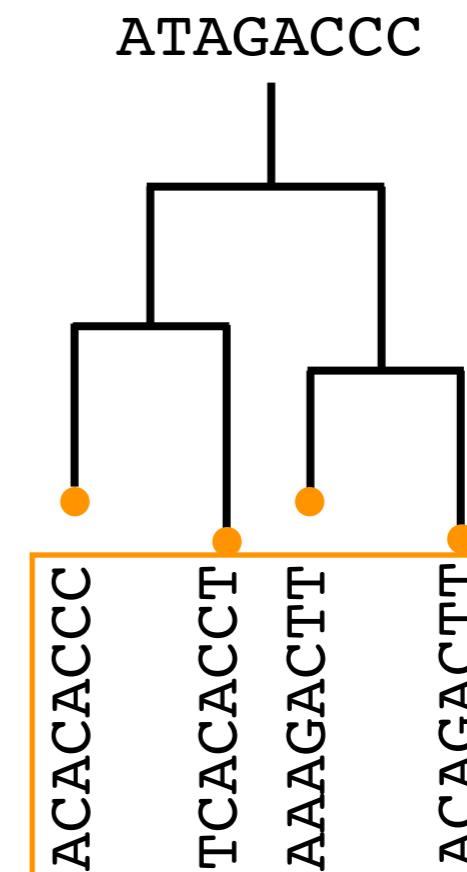
Evolution of sequences along tree

**(Evolutionary dynamics - genetic change)**

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

Data

Sequence alignment



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

=

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

# Phylogenetic model

Population  
model

Growth of tree

**(Population dynamics - replication)**

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

Evolution  
model

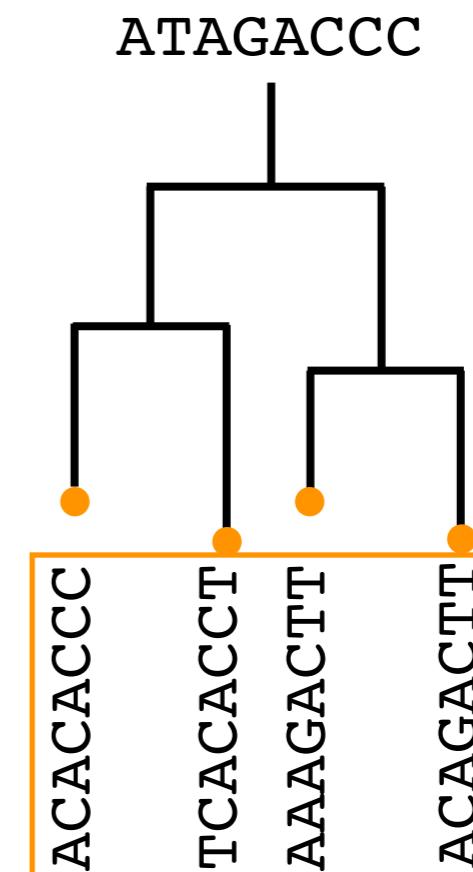
Evolution of sequences along tree

**(Evolutionary dynamics - genetic change)**

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

Data

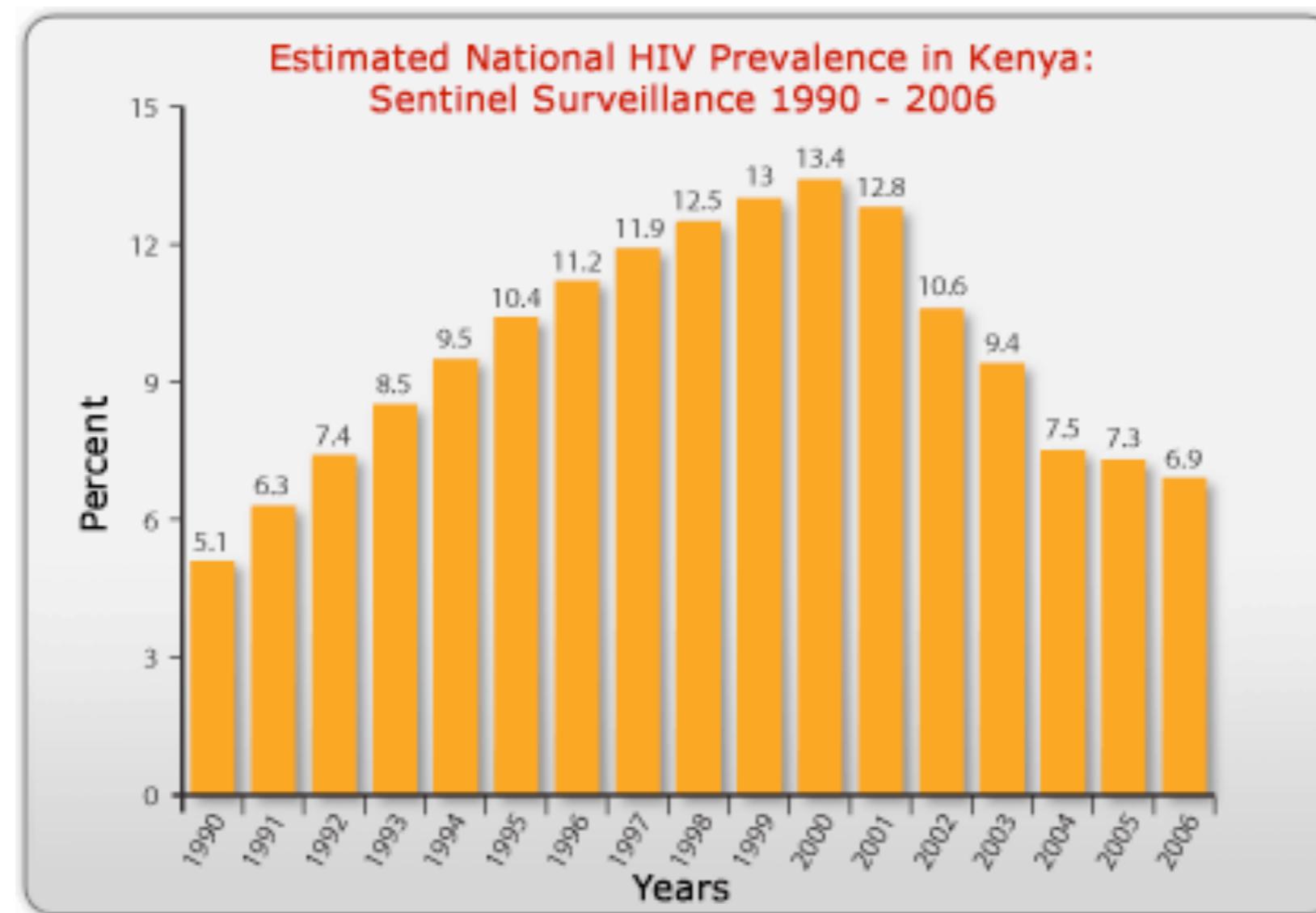
Sequence alignment



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

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

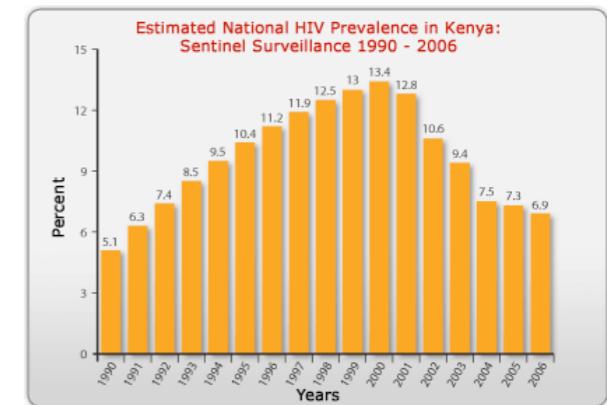
# Epidemiology



# Epidemiology

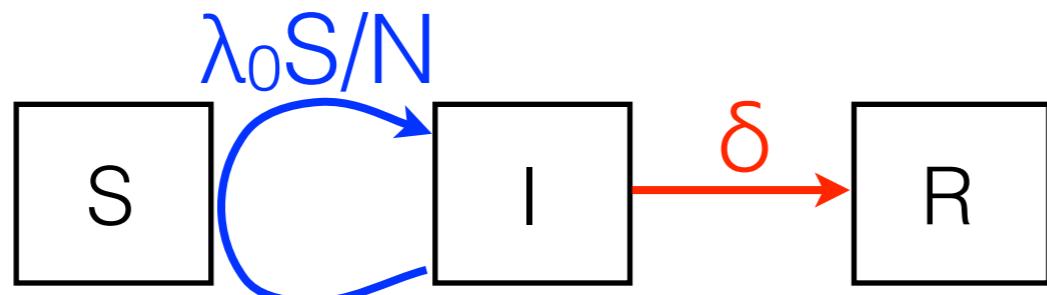
Data

# of infected individuals through time



Models

Population dynamics described by SIR models:

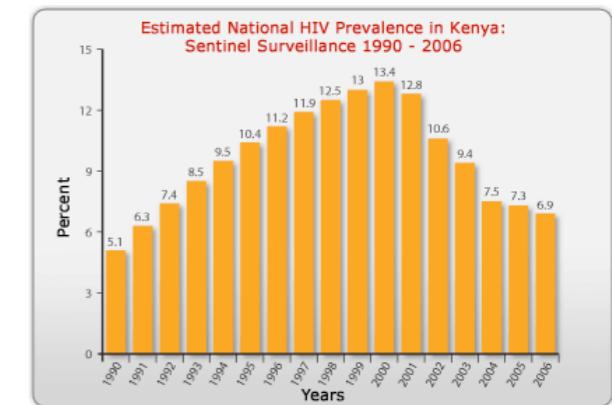


$$\begin{aligned}\frac{dS}{dt} &= -\lambda_0 IS/N \\ \frac{dI}{dt} &= \lambda_0 IS/N - \delta I \\ \frac{dR}{dt} &= \delta I\end{aligned}$$

# Epidemiology

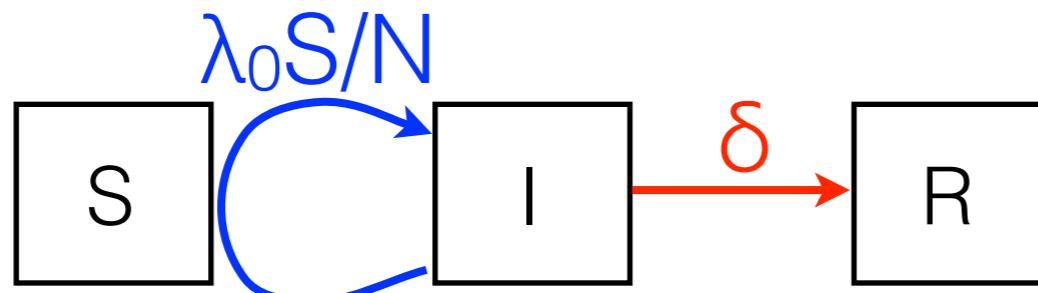
Data

# of infected individuals through time



Models

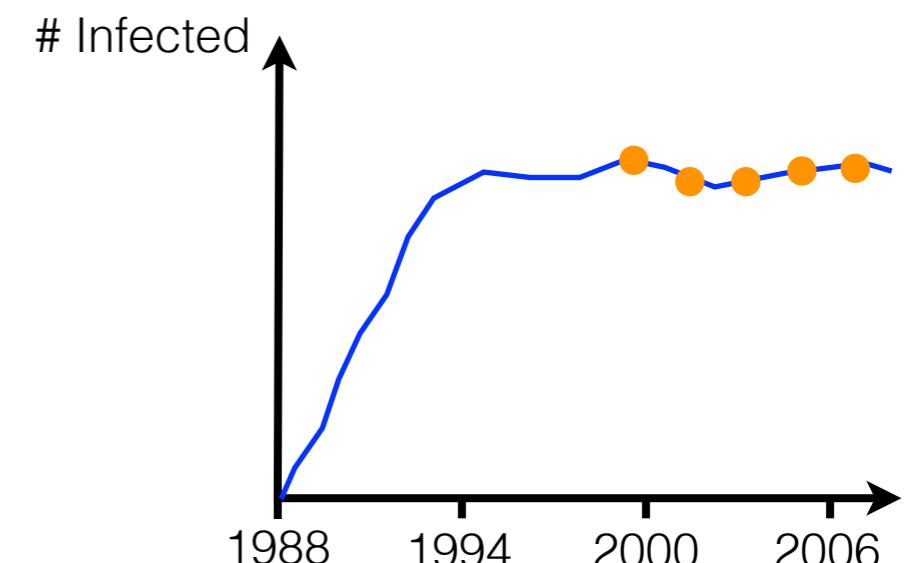
Population dynamics described by SIR models:



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

Limitations

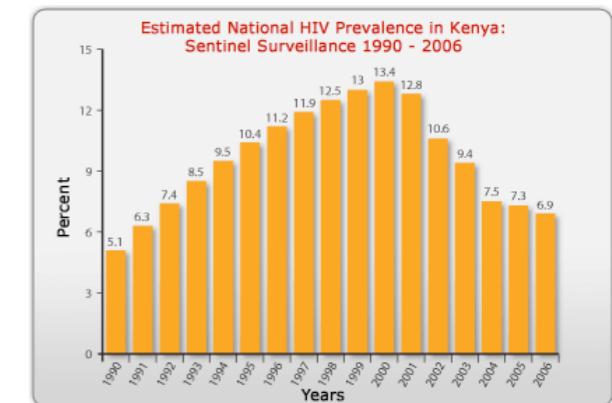
If sampling in early epidemic was missed:



# Epidemiology

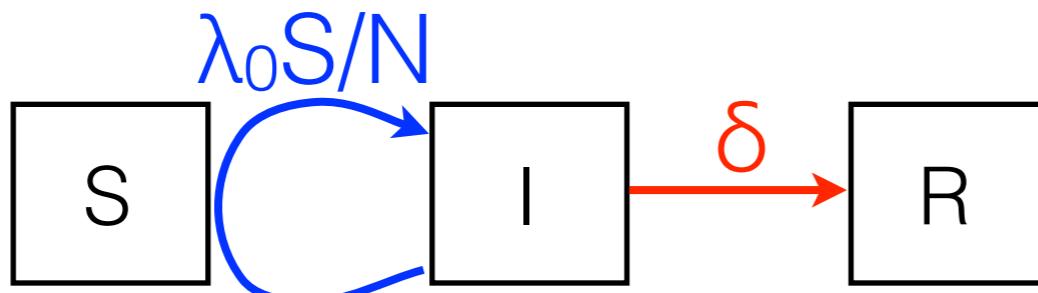
Data

# of infected individuals through time



Models

Population dynamics described by SIR models:

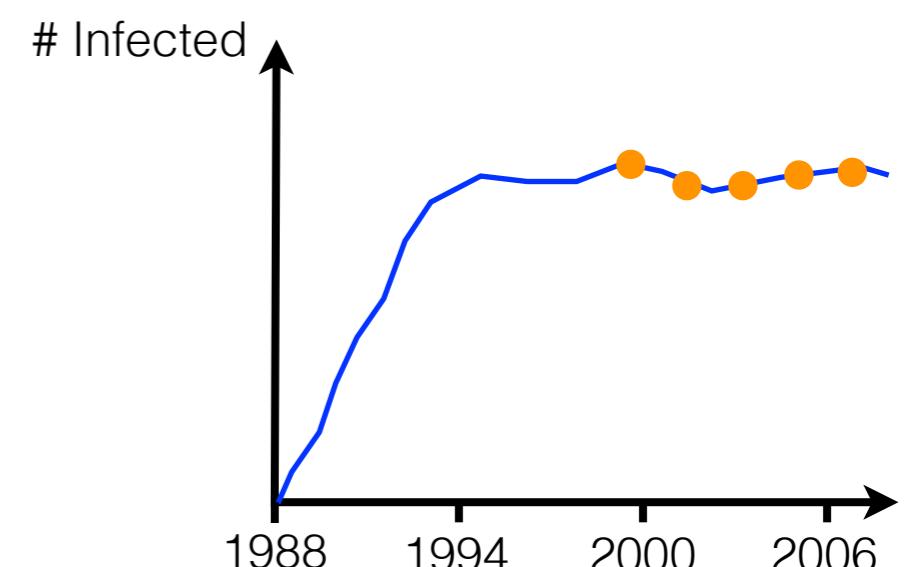


$$\begin{aligned}\frac{dS}{dt} &= -\lambda_0 IS/N \\ \frac{dI}{dt} &= \lambda_0 IS/N - \delta I \\ \frac{dR}{dt} &= \delta I\end{aligned}$$

Limitations

If sampling in early epidemic was missed:

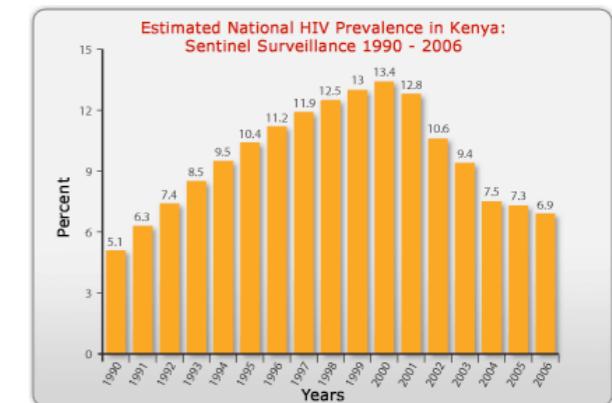
- ▶ **Time of epidemic outbreak?**
- ▶ **Basic reproductive number  $R_0$ ?**



# Epidemiology

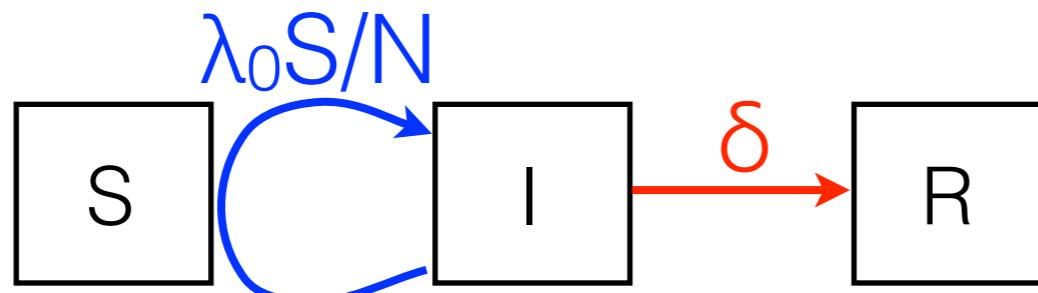
Data

# of infected individuals through time



Models

Population dynamics described by SIR models:



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

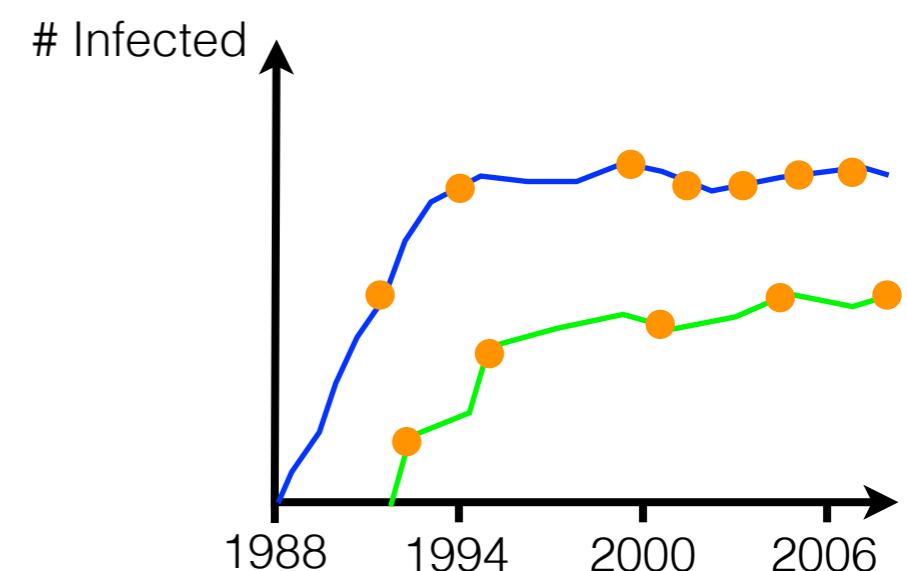
Limitations

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

---

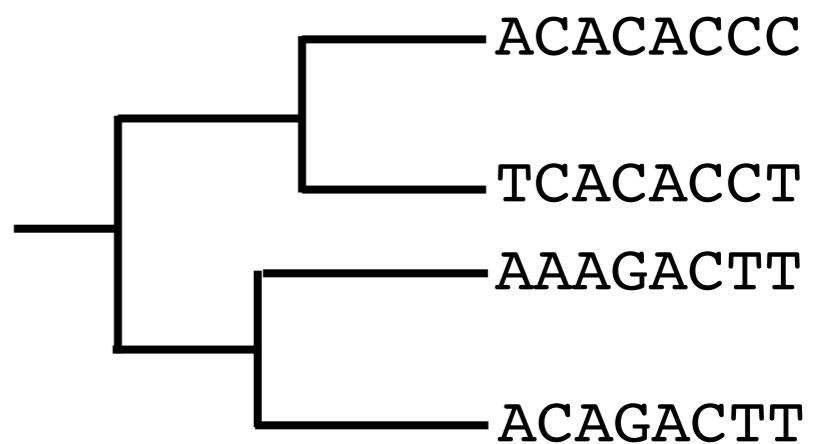
Phylogenetics

Input:

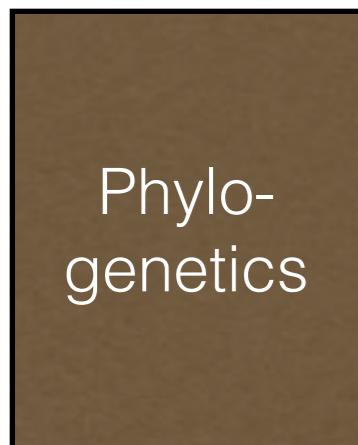
Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)



# Genetic data entering epidemiology

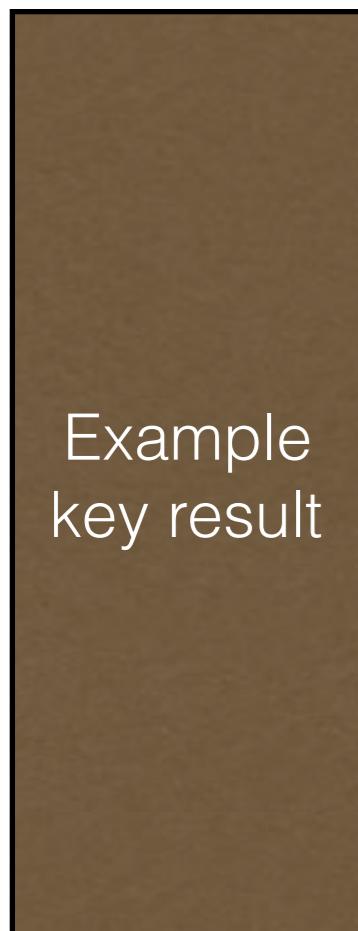
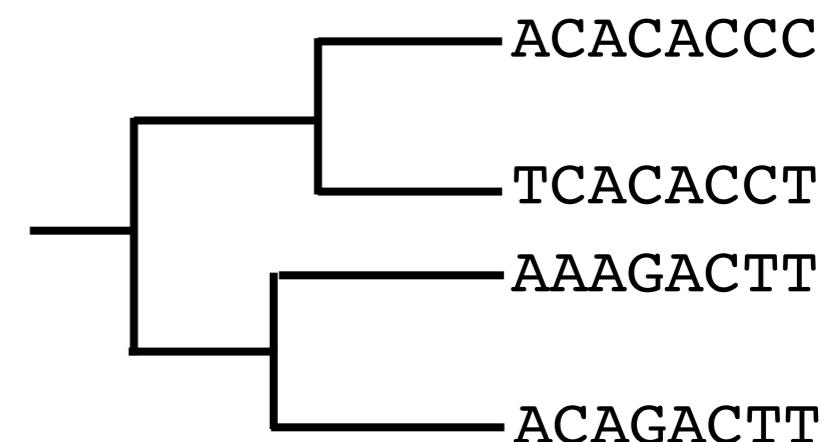


Input:

Pathogen genetic data from different hosts

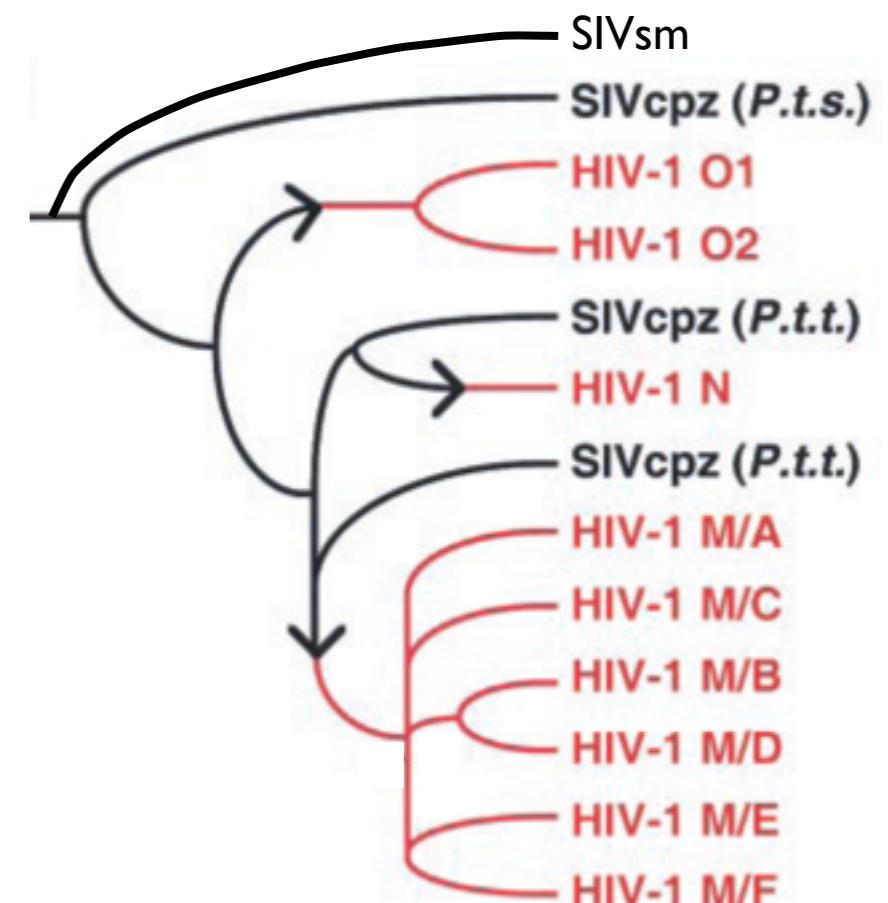
Output:

Transmission chain (who infected whom)



Emergence of HIV

Epidemic identified in 1980s



adopted from Hahn et al. (Science, 2000)

# Genetic data entering epidemiology

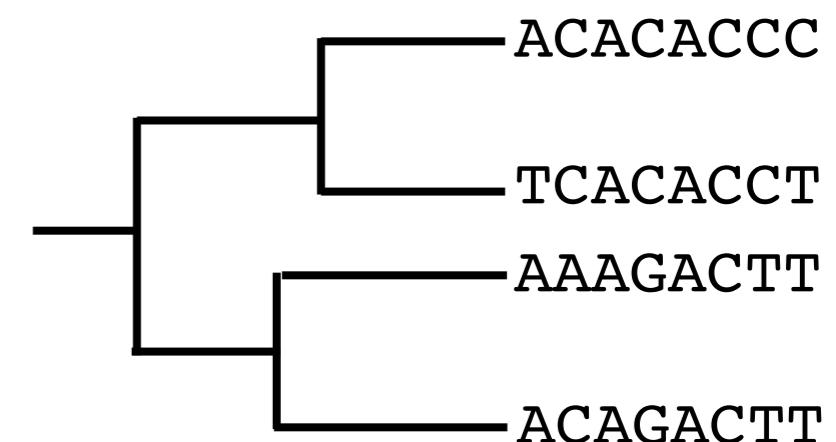
Phylogenetics

Input:

Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)



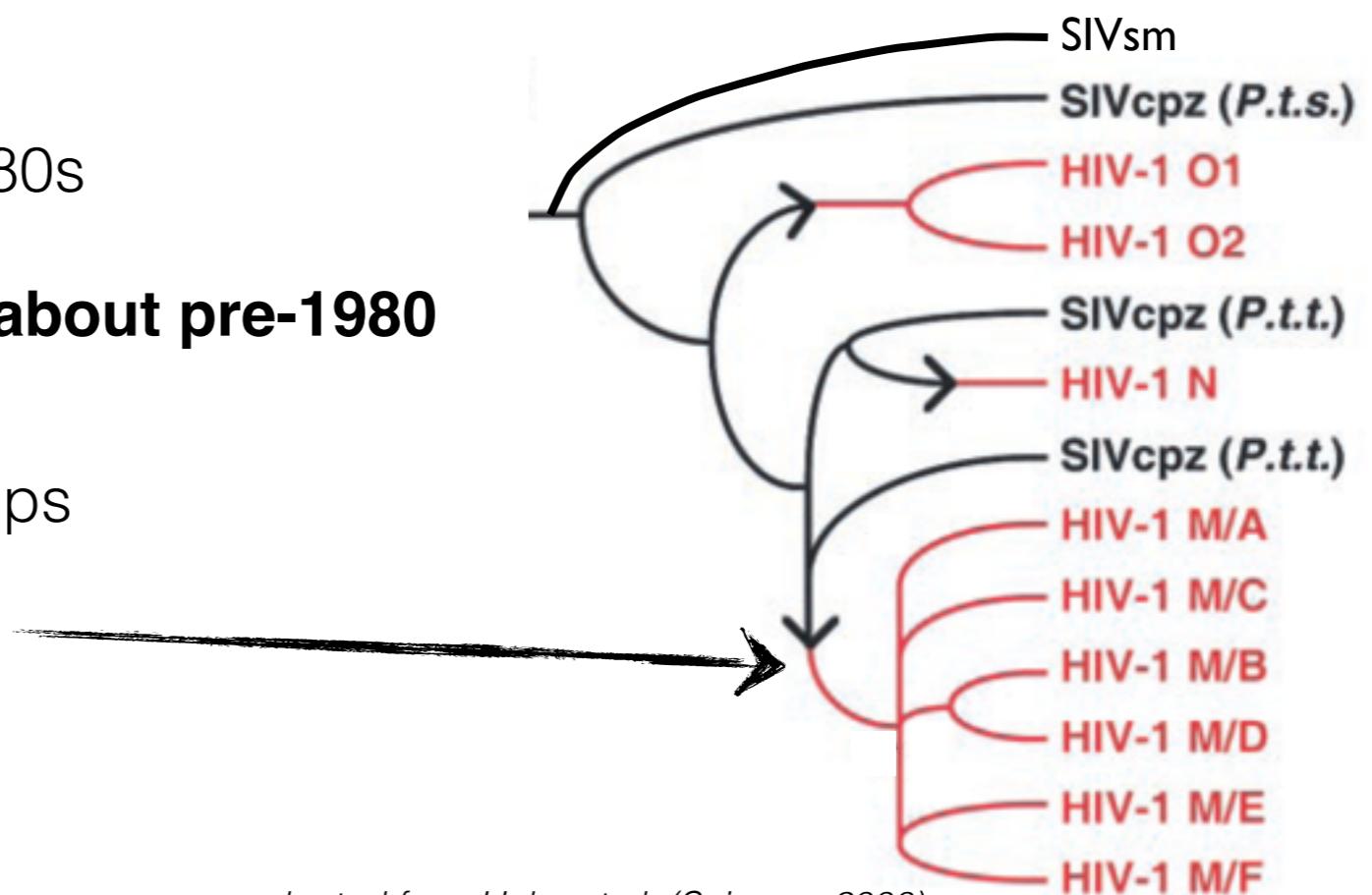
Example key result

Emergence of HIV

Epidemic identified in 1980s

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

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



adopted from Hahn et al. (Science, 2000)

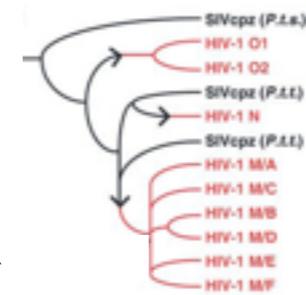
# Genetic data contains information about the epidemiological dynamics

---

Phylogenetics

State of process

- ▶ Time and geographic location of outbreak



Emergence of a pathogen (here HIV)

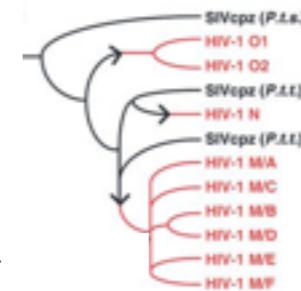
Hahn et al. (Science, 2000)

# Genetic data contains information about the epidemiological dynamics

Phylogenetics

State of process

- ▶ Time and geographic location of outbreak



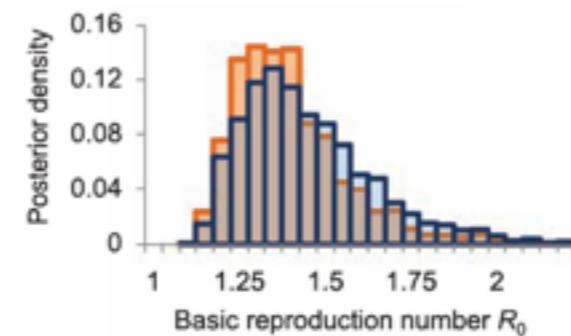
Emergence of a pathogen (here HIV)

Hahn et al. (Science, 2000)

Phylogenetics

Dynamics of process

- ▶ Transmission process



Basic reproductive number of an emerging pathogen (here H1N1)

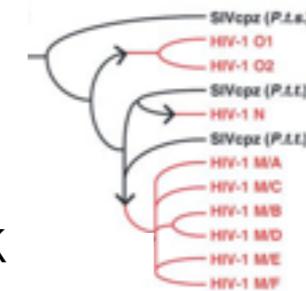
Fraser et al. (Science, 2009)

# Genetic data contains information about the epidemiological dynamics

Phylogenetics

State of process

- ▶ Time and geographic location of outbreak



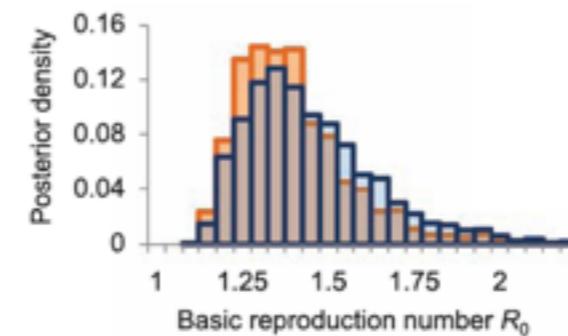
Emergence of a pathogen (here HIV)

Hahn et al. (Science, 2000)

Phylogenetics

Dynamics of process

- ▶ Transmission process



Basic reproductive number of an emerging pathogen (here H1N1)

Fraser et al. (Science, 2009)

Epidemiological models

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

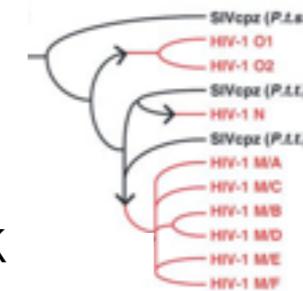
- ▶ Quantification of all model parameters!

# Genetic data contains information about the epidemiological dynamics

Phylogenetics

State of process

- ▶ Time and geographic location of outbreak



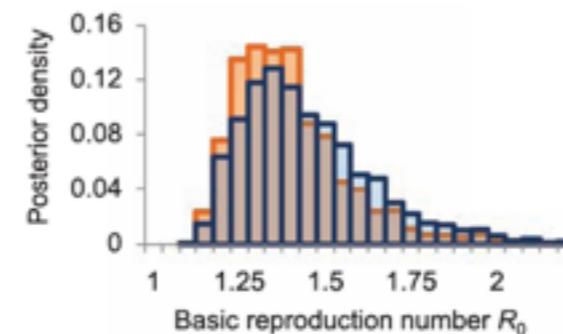
Emergence of a pathogen (here HIV)

Hahn et al. (Science, 2000)

Phylogenetics

Dynamics of process

- ▶ Transmission process



Basic reproductive number of an emerging pathogen (here H1N1)

Fraser et al. (Science, 2009)

Epidemiological models

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

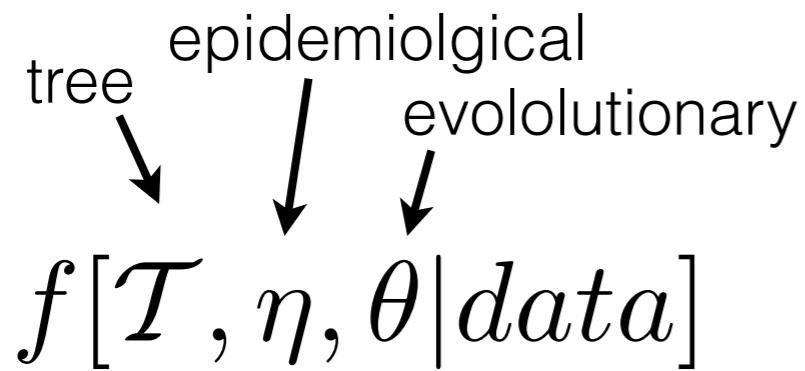
- ▶ Quantification of all model parameters!

Approximate: by population genetic model (coalescent)

- ▶ Deterministic population size is parameterized!

# Bayesian approach for estimating epidemiological parameters

---



# Bayesian approach for estimating epidemiological parameters

---

$$f[\mathcal{T}, \eta, \theta | data] = \frac{f[data | \mathcal{T}, \theta] f[\mathcal{T} | \eta] f[\eta, \theta]}{f[data]}$$

tree      epidemiological      evolutionary

# Bayesian approach for estimating epidemiological parameters

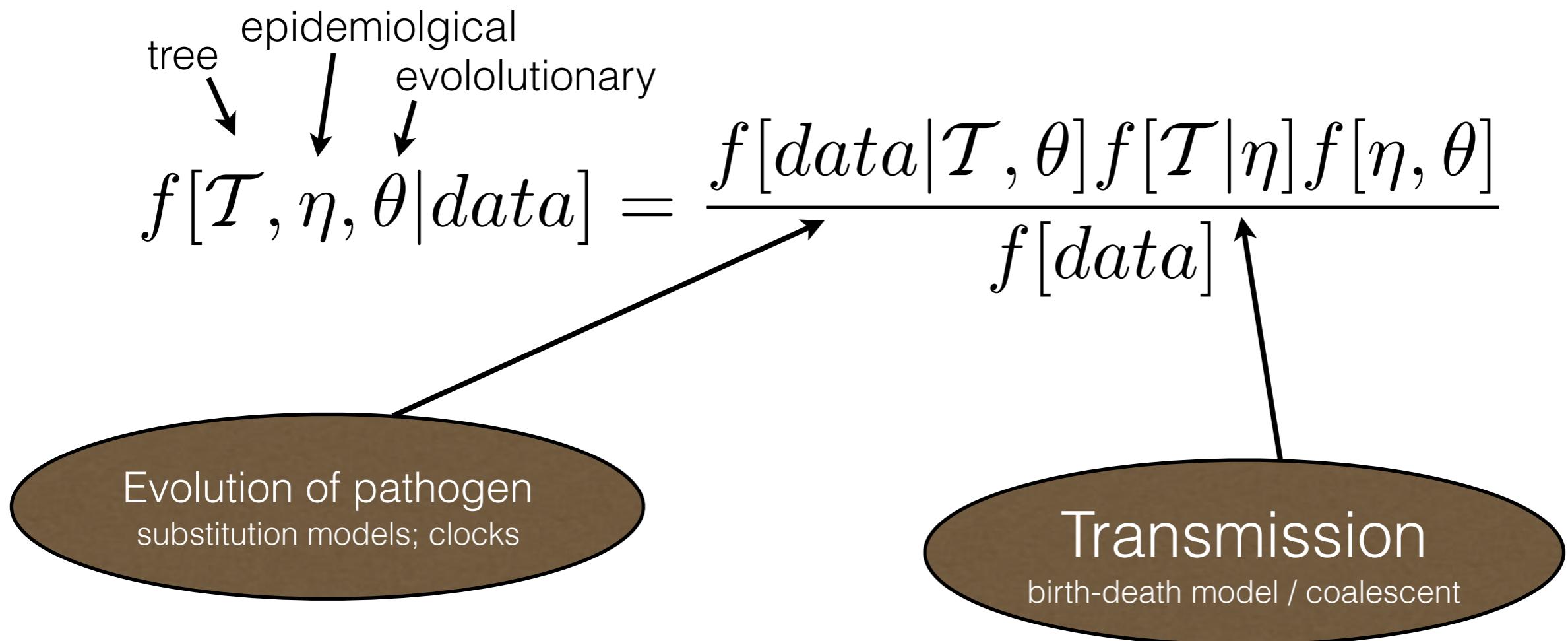
---

$$f[\mathcal{T}, \eta, \theta | data] = \frac{f[data | \mathcal{T}, \theta] f[\mathcal{T} | \eta] f[\eta, \theta]}{f[data]}$$



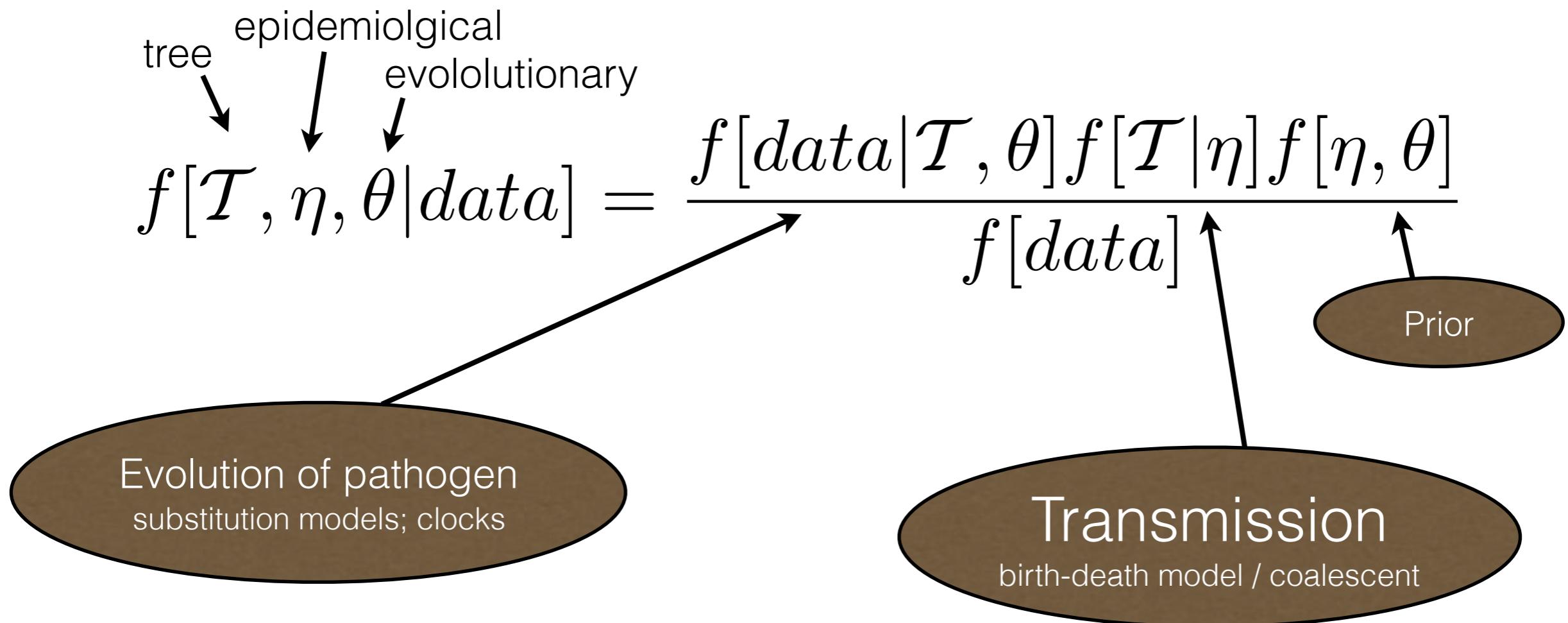
# Bayesian approach for estimating epidemiological parameters

---

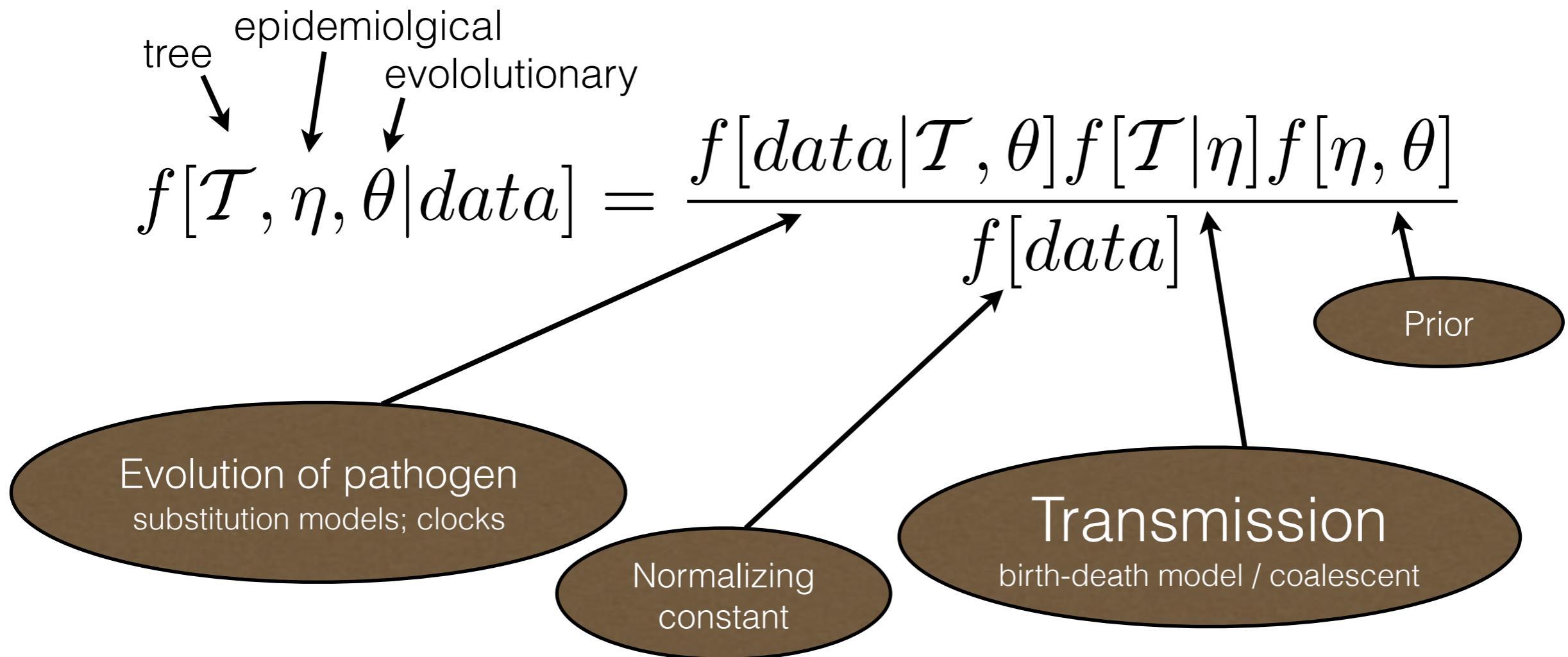


# Bayesian approach for estimating epidemiological parameters

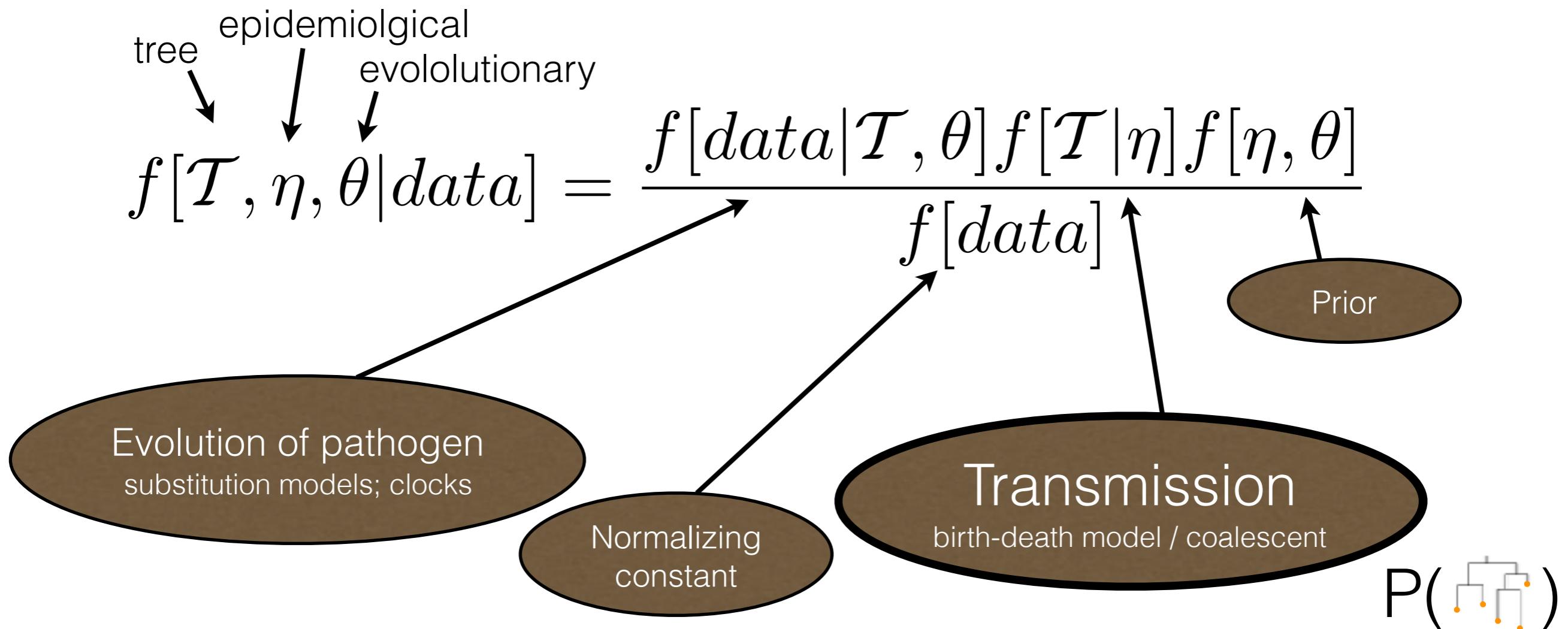
---



# Bayesian approach for estimating epidemiological parameters

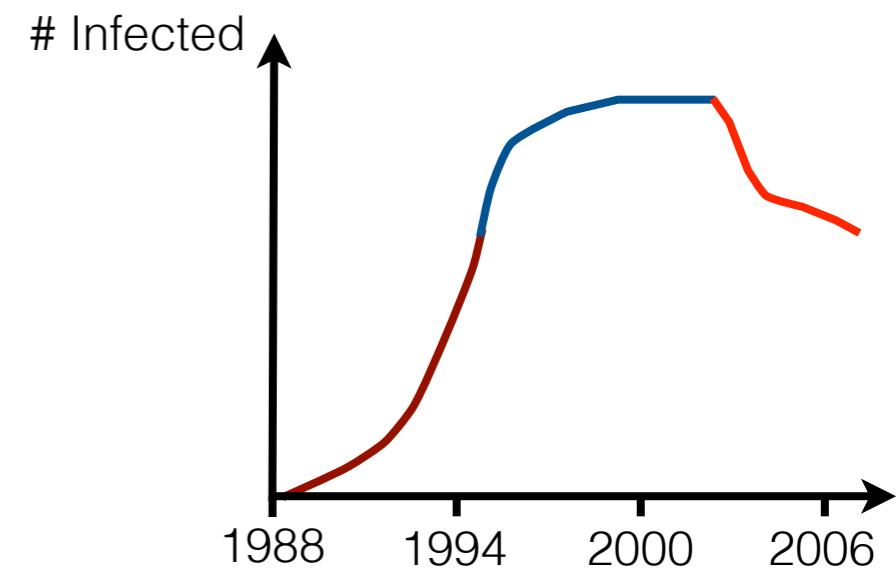


# Bayesian approach for estimating epidemiological parameters



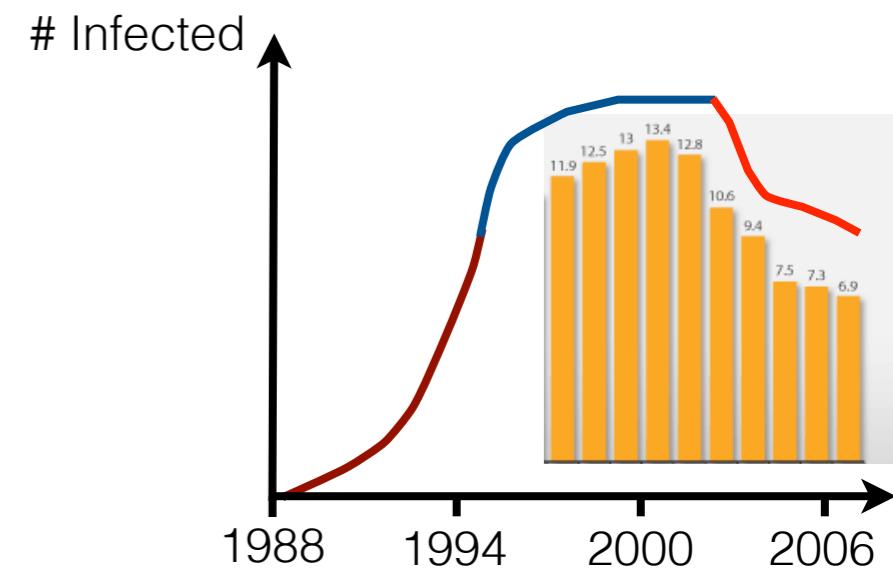
# Phylogenetic methods for different epidemic dynamics

---



# Phylogenetic methods for different epidemic dynamics

---

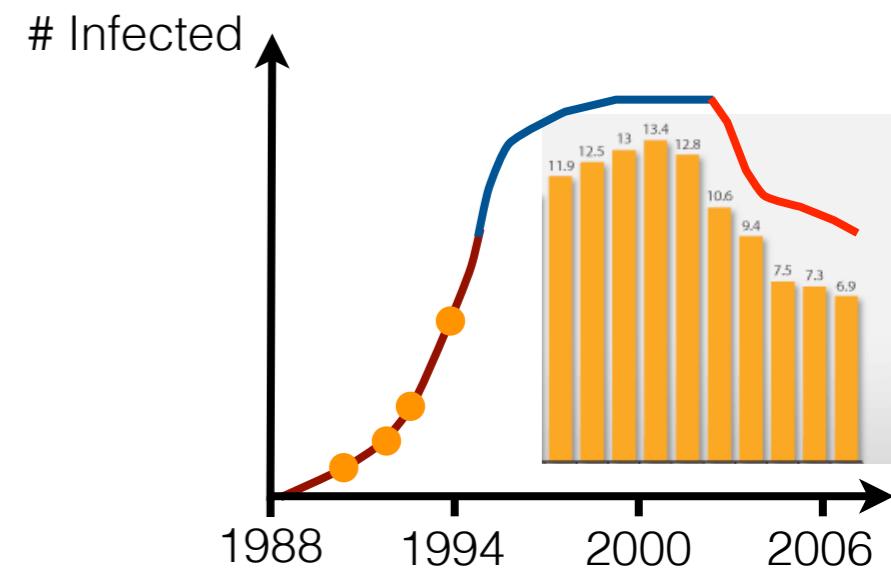


# Phylogenetic methods for different epidemic dynamics

---

## 1. ...for analyzing epidemic outbreaks

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



# Phylogenetic methods for different epidemic dynamics

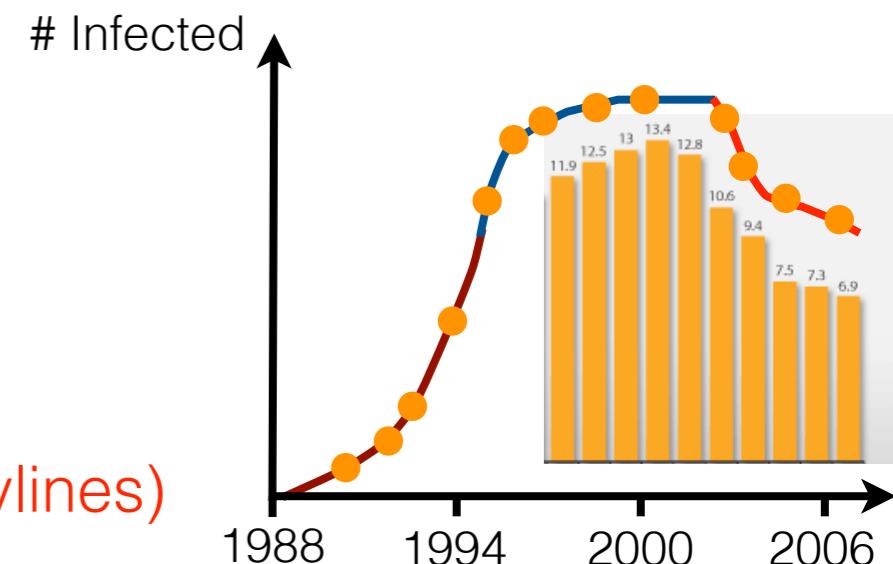
---

## 1. ...for analyzing epidemic outbreaks

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

## 2. ...for identifying epidemiological changes over time (skylines)

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



# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

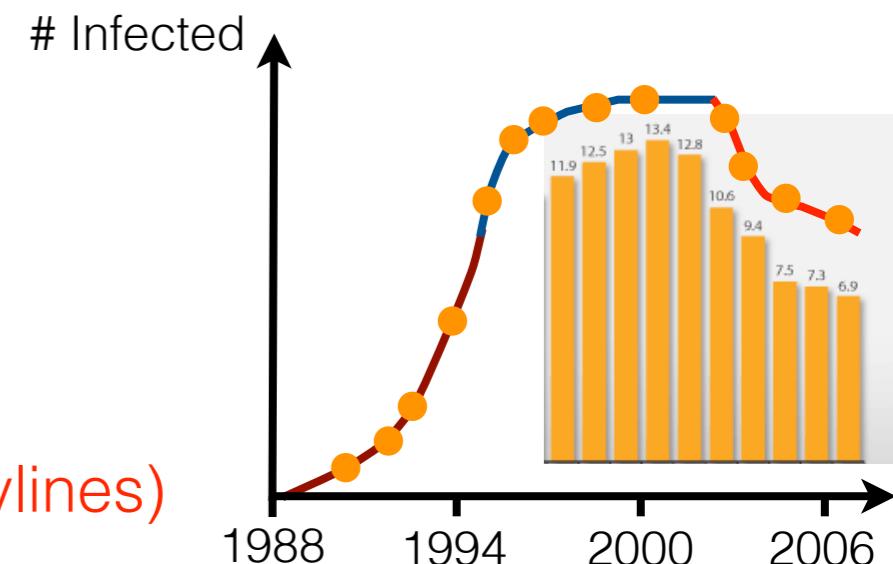
- ▶ C: Drummond et al. (*Genetics*, 2002)
- ▶ BD: Stadler et al. (*MBE*, 2012) - **BDSKY add-on**

## 2. ...for identifying epidemiological changes over time (skylines)

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

## 3. ...accounting for SIR model dynamics

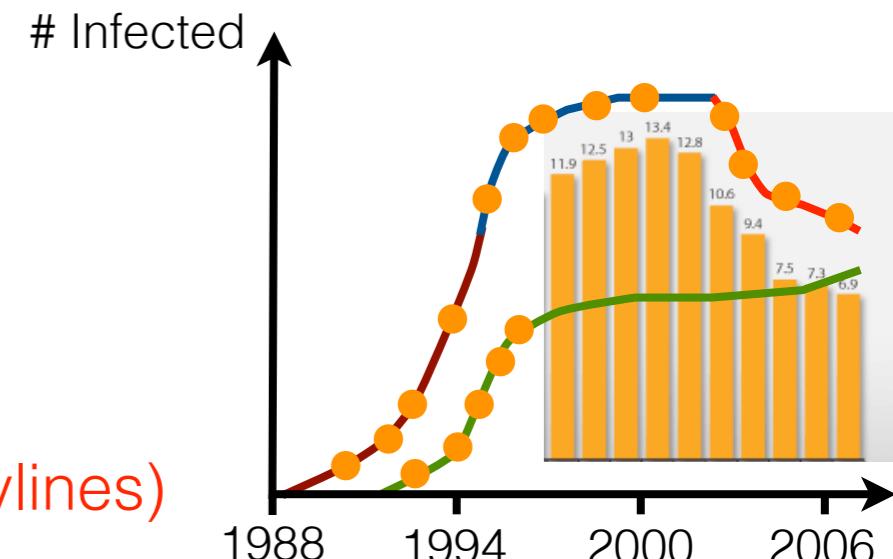
- ▶ C: Volz et al. (*Genetics*, 2009)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014) - **BDSIR add-on**



# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

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



## 2. ...for identifying epidemiological changes over time (skylines)

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

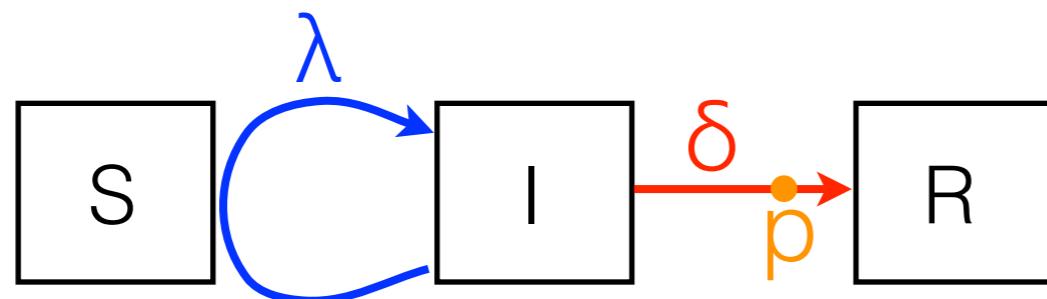
## 3. ...accounting for SIR model dynamics

- ▶ C: Volz et al. (*Genetics*, 2009)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014) - **BDSIR add-on**

## 4. ...accounting for population structure

- ▶ C: Volz (*Genetics*, 2012), Vaughan et al. (*Bioinf.*, 2014), de Maio et al. (*PLoS Genetics*, 2015), Mueller et al. (*MBE*, 2017) - **Mascot add-on**
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013), Kühnert et al. (*MBE*, 2016) - **bdmm add-on**

# Birth-death model as a model for epidemic spread

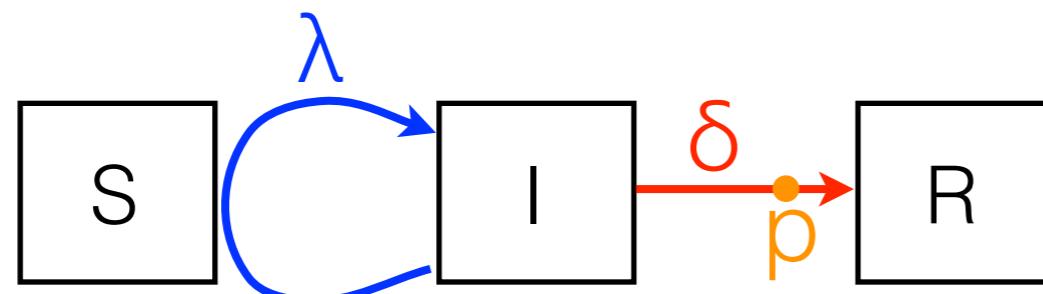


- transmission rate  $\lambda$
- “becoming-non-infectious” rate  $\delta$
- sampling probability  $p$

**Parameters may depend on:**

- 1) constant (epidemic outbreak)
- 2) # of susceptibles (SIR model)
- 3) time (environmental effect)
- 4) type of infected individual

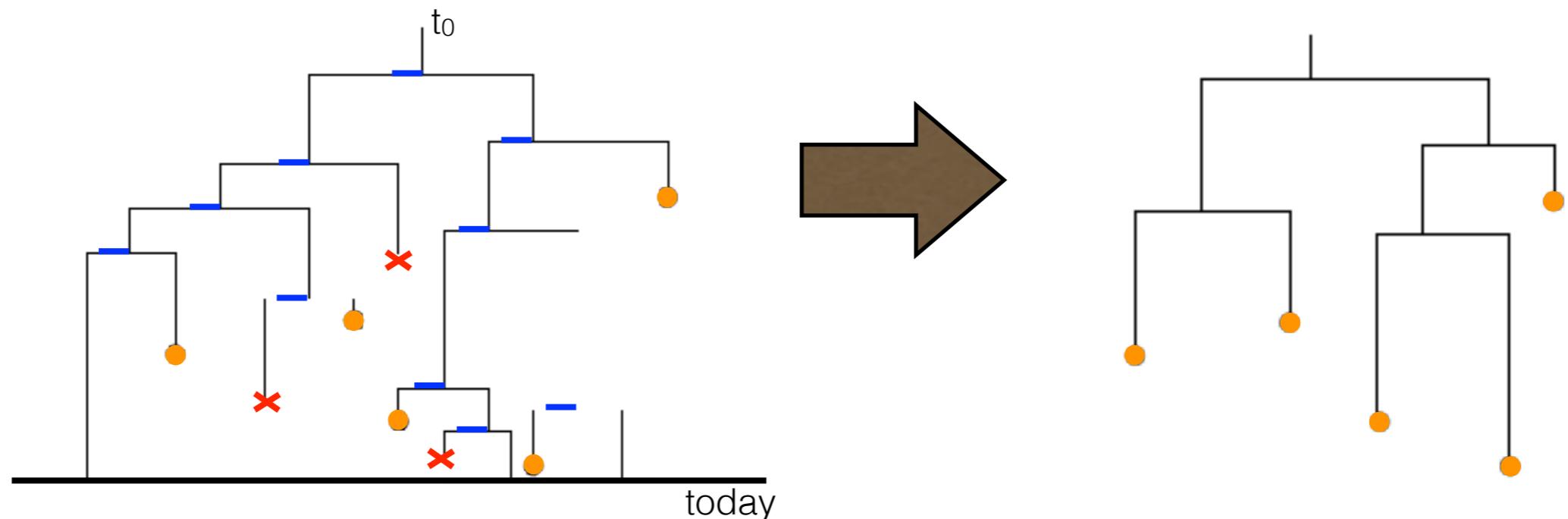
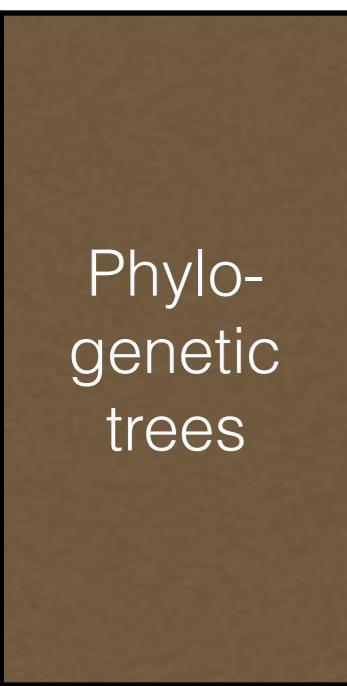
# Birth-death model as a model for epidemic spread



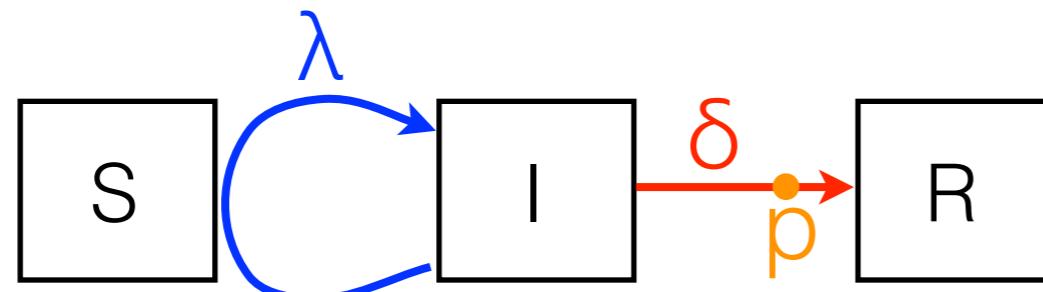
- transmission rate  $\lambda$
- “becoming-non-infectious” rate  $\delta$
- sampling probability  $p$

## Parameters may depend on:

- 1) constant (epidemic outbreak)
- 2) # of susceptibles (SIR model)
- 3) time (environmental effect)
- 4) type of infected individual



# Birth-death model as a model for epidemic spread

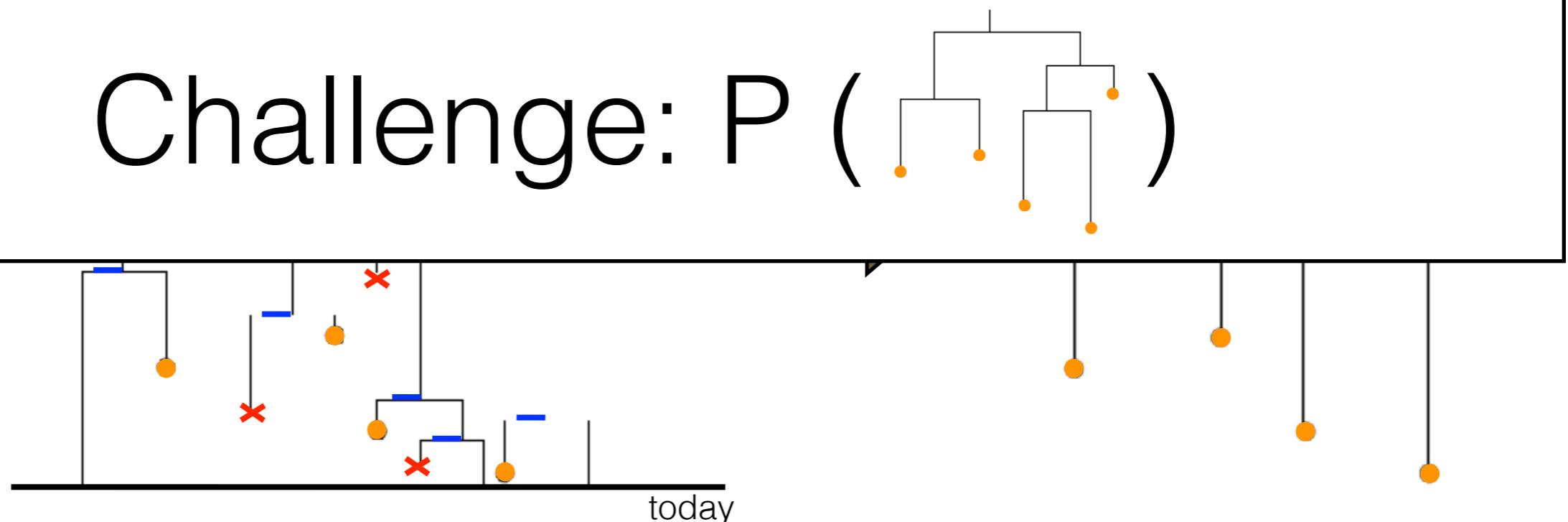


- transmission rate  $\lambda$
- “becoming-non-infectious” rate  $\delta$
- sampling probability  $p$

## Parameters may depend on:

- 1) constant (epidemic outbreak)
- 2) # of susceptibles (SIR model)
- 3) time (environmental effect)
- 4) type of infected individual

Challenge:  $P( \text{ } )$



# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

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

## 2. ...for identifying epidemiological changes over time (skylines)

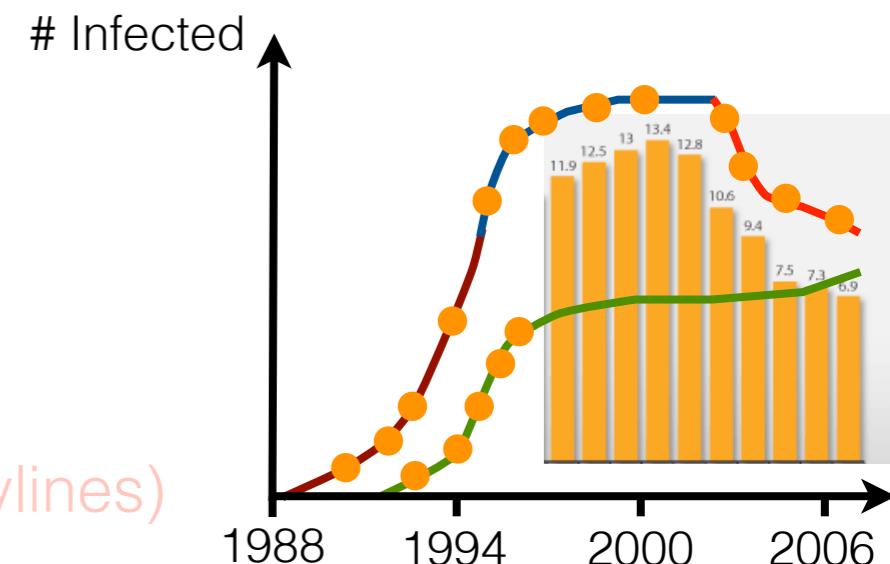
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013) - **BDSKY add-on**

## 3. ...accounting for SIR model dynamics

- ▶ C: Volz et al. (*Genetics*, 2009)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014) - **BDSIR add-on**

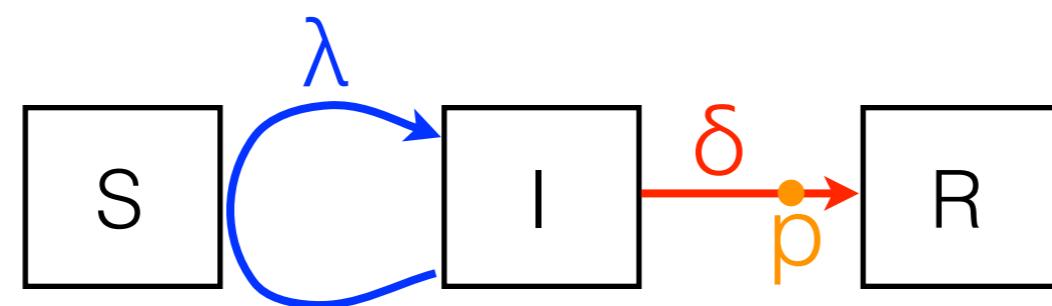
## 4. ...accounting for population structure

- ▶ C: Volz (*Genetics*, 2012), Vaughan et al. (*Bioinf.*, 2014), de Maio et al. (*PLoS Genetics*, 2015), Mueller et al. (*MBE*, 2017) - **Mascot add-on**
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013), Kühnert et al. (*MBE*, 2016) - **bdmm add-on**



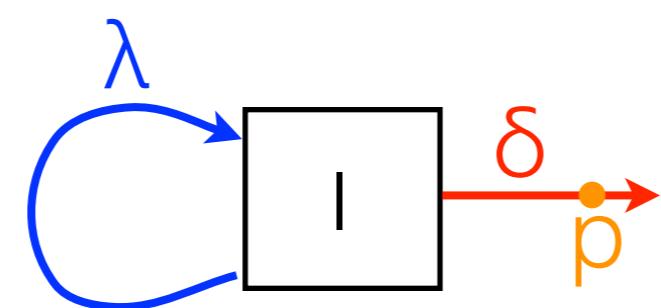
# 1) Epidemic outbreaks

---

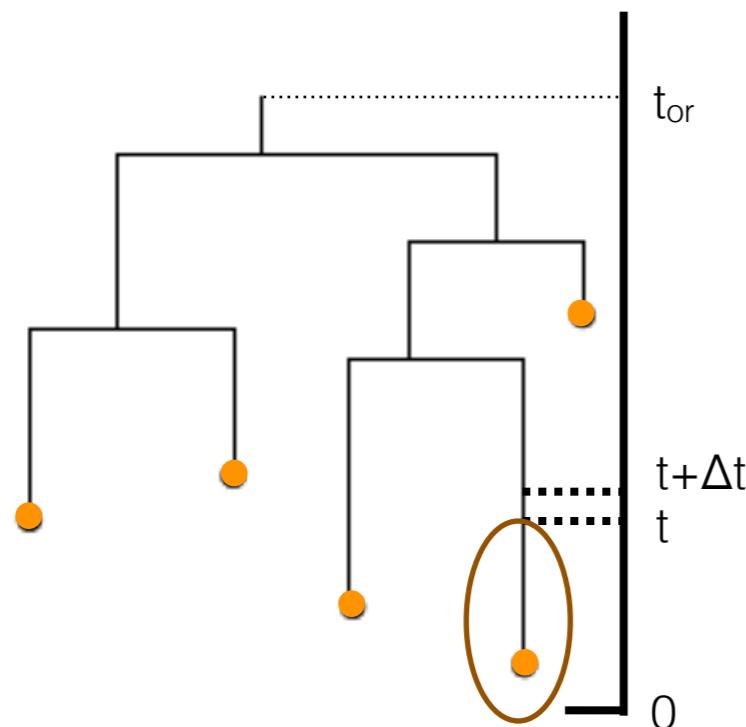
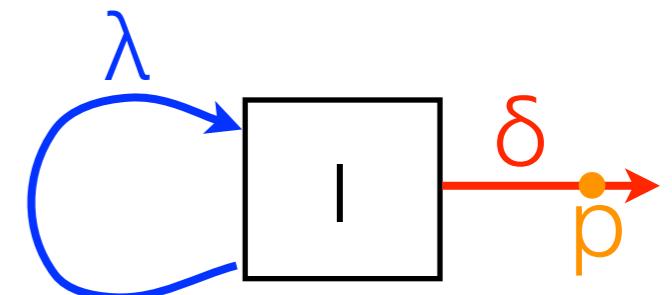


# 1) Epidemic outbreaks

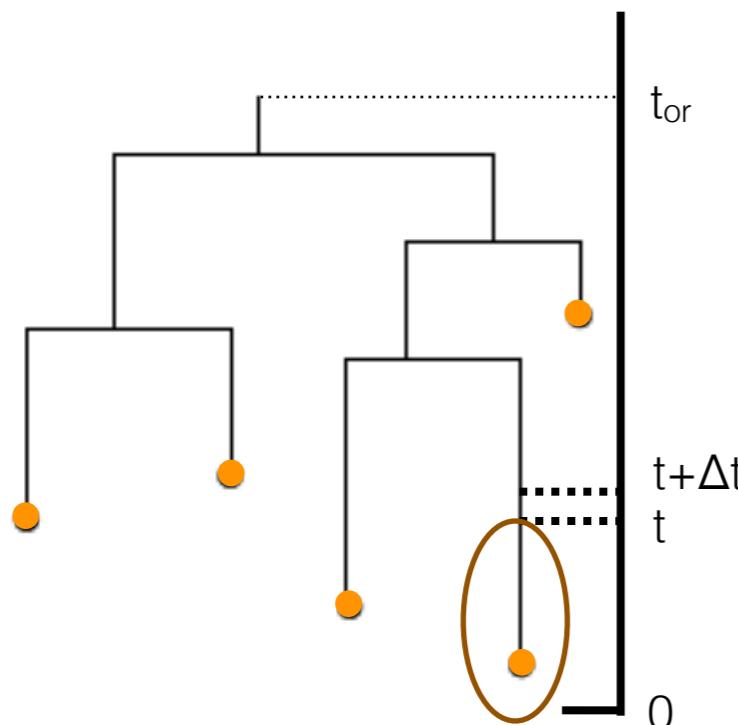
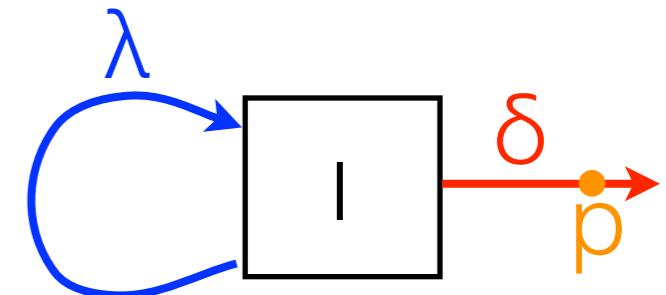
---



# 1) Epidemic outbreaks

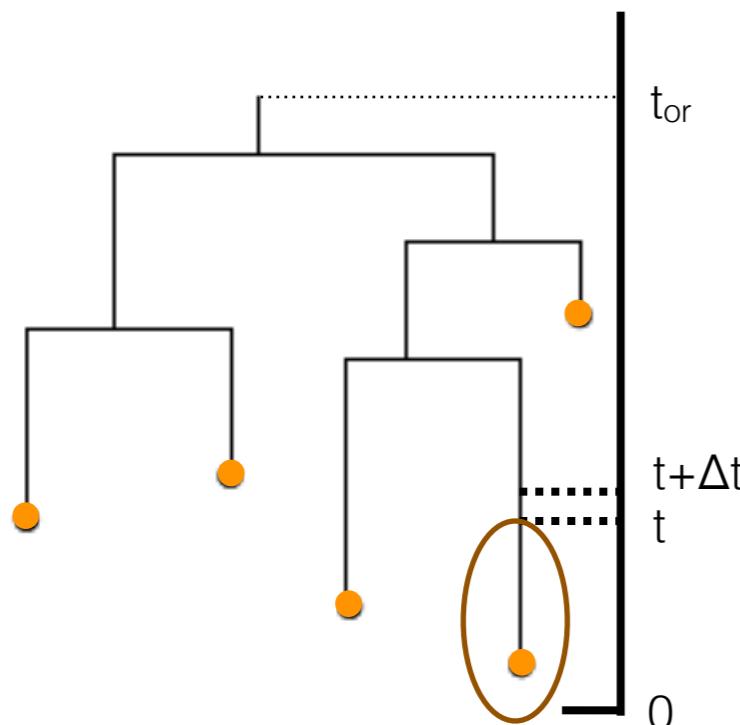
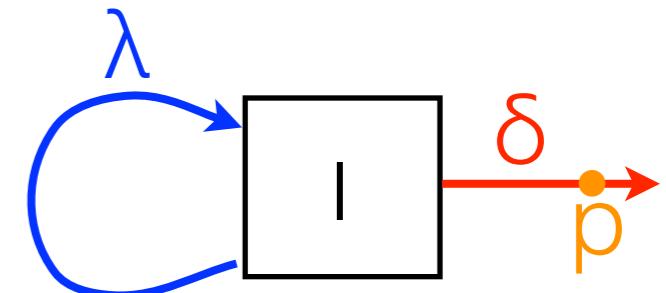


# 1) Epidemic outbreaks



$p(t)$ : probability density that a given individual at time  $t$  produces an descending tree as observed.

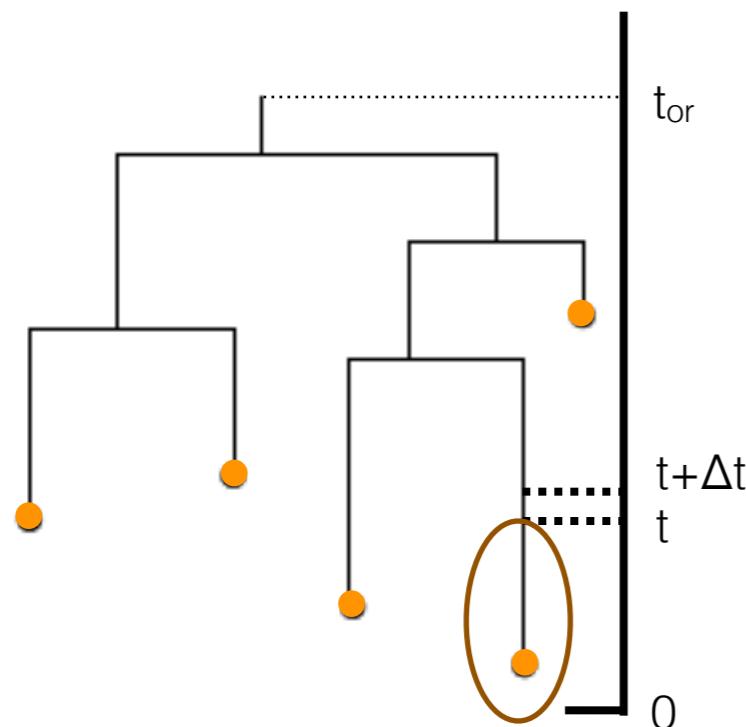
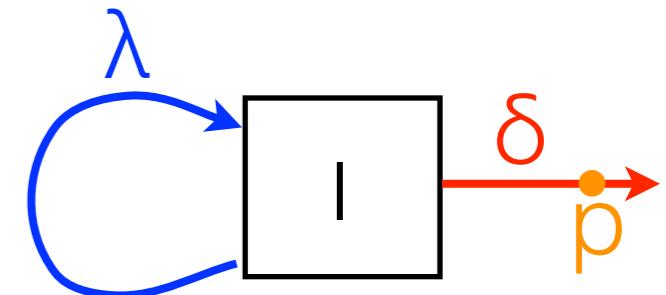
# 1) Epidemic outbreaks



$p(t)$ : probability density that a given individual at time  $t$  produces an descending tree as observed.

$p(t_{\text{or}}) = P(\text{tree})$ : probability density of sampled tree with origin at time  $t_{\text{or}}$ .

# 1) Epidemic outbreaks

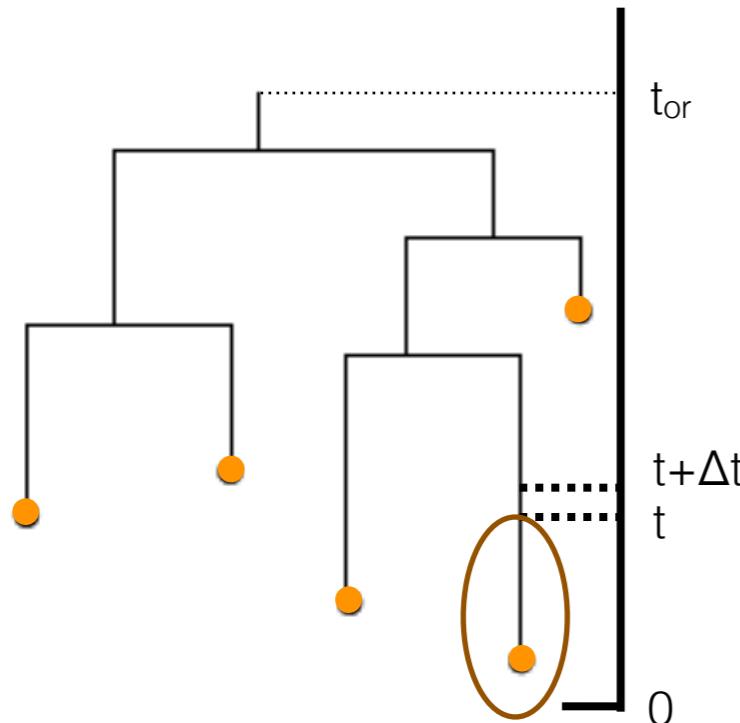
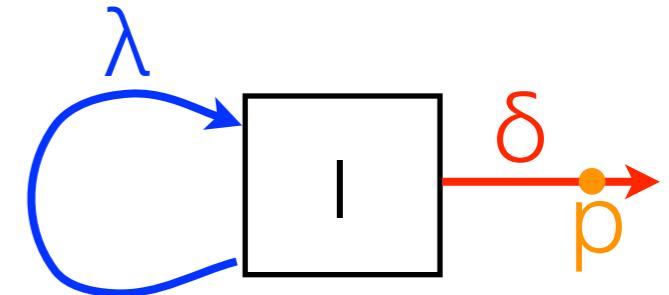


$p(t)$ : probability density that a given individual at time  $t$  produces an descending tree as observed.

$p(t_{\text{or}}) = P(\text{tree})$ : probability density of sampled tree with origin at time  $t_{\text{or}}$ .

$p_0(t)$ : probability density that a given individual at time  $t$  has no sampled descendants.

# 1) Epidemic outbreaks



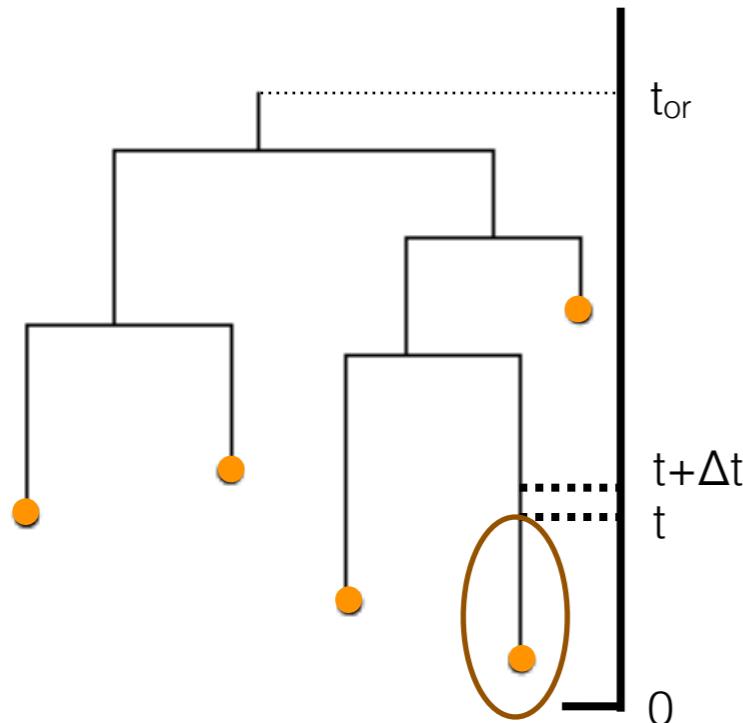
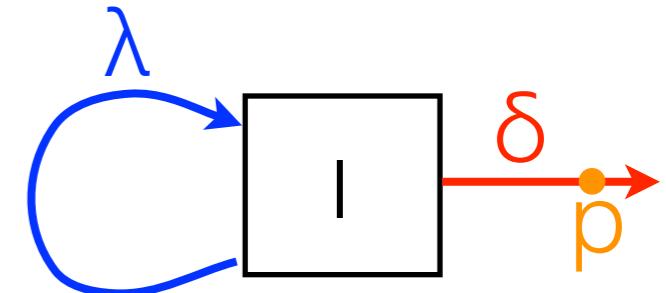
$p(t)$ : probability density that a given individual at time  $t$  produces an descending tree as observed.

$p(t_{\text{or}}) = P(\text{tree})$ : probability density of sampled tree with origin at time  $t_{\text{or}}$ .

$p_0(t)$ : probability density that a given individual at time  $t$  has no sampled descendants.

$$p(t + \Delta t) = (1 - (\lambda + \delta)\Delta t - O(\Delta t^2))p(t) + \lambda\Delta t p_0(t)p(t) + O(\Delta t^2)$$

# 1) Epidemic outbreaks



$p(t)$ : probability density that a given individual at time  $t$  produces an descending tree as observed.

$p(t_{\text{or}}) = P(\text{tree})$ : probability density of sampled tree with origin at time  $t_{\text{or}}$ .

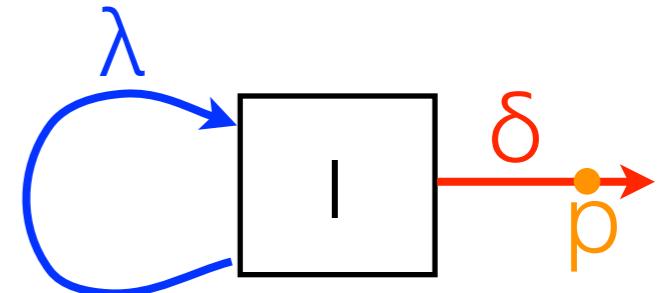
$p_0(t)$ : probability density that a given individual at time  $t$  has no sampled descendants.

$$p(t + \Delta t) = (1 - (\lambda + \delta)\Delta t - O(\Delta t^2))p(t) + \lambda\Delta t p_0(t)p(t) + O(\Delta t^2)$$

$\Delta t \rightarrow 0$

$$\frac{d}{dt}p(t) = -(\lambda + \delta)p(t) + 2\lambda p_0(t)p(t)$$

# Tree probability



$$P \left( \text{tree diagram} \right) = \lambda^{m-1} \prod_{i=0}^{m-1} \frac{1}{q(x_i)} \prod_{i=1}^m p \delta q(y_i)$$

where

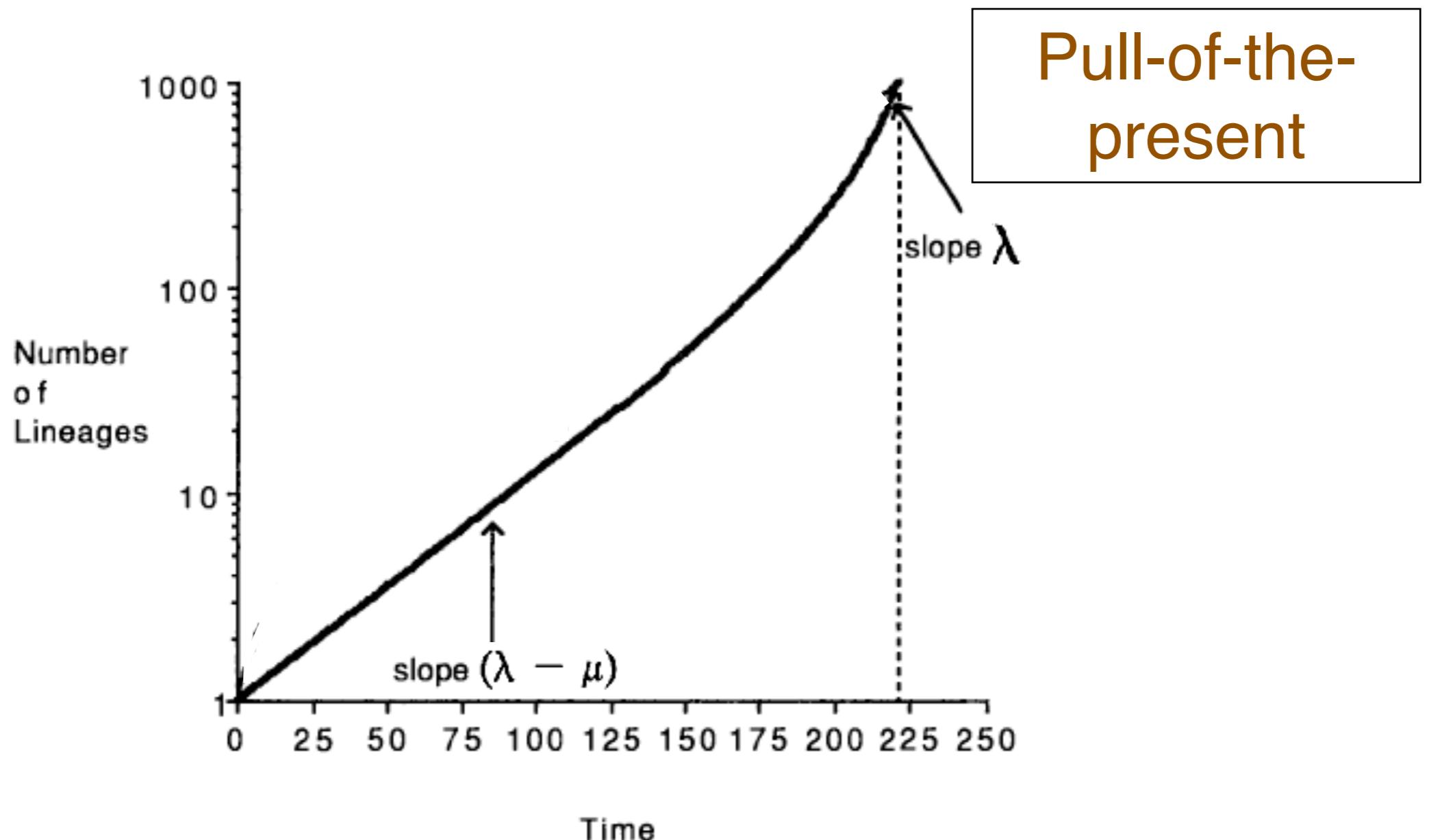
$$q(t) = 2(1 - c_2^2) + e^{-c_1 t}(1 - c_2)^2 + e^{c_1 t}(1 + c_2)^2$$

$$c_1 = |\sqrt{(\lambda - \delta)^2} + 4p\lambda\delta|$$

$$c_2 = -\frac{\lambda - \delta}{c_1}$$

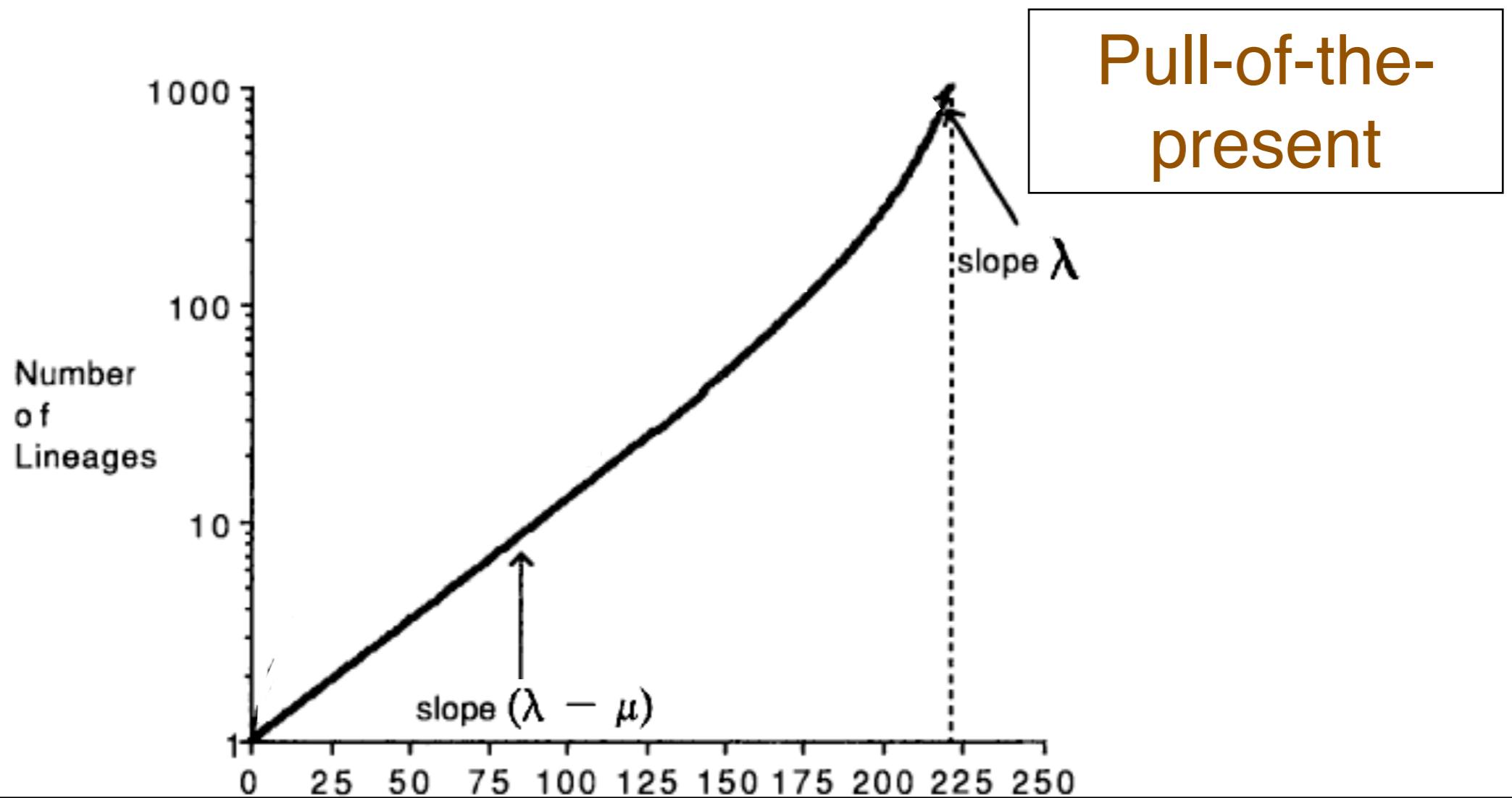
and  $x_i$  branching times,  $y_i$  sampling times.

# Fingerprint of the birth-and death rates in phylogenies



*Harvey et al (1994, Evolution)*

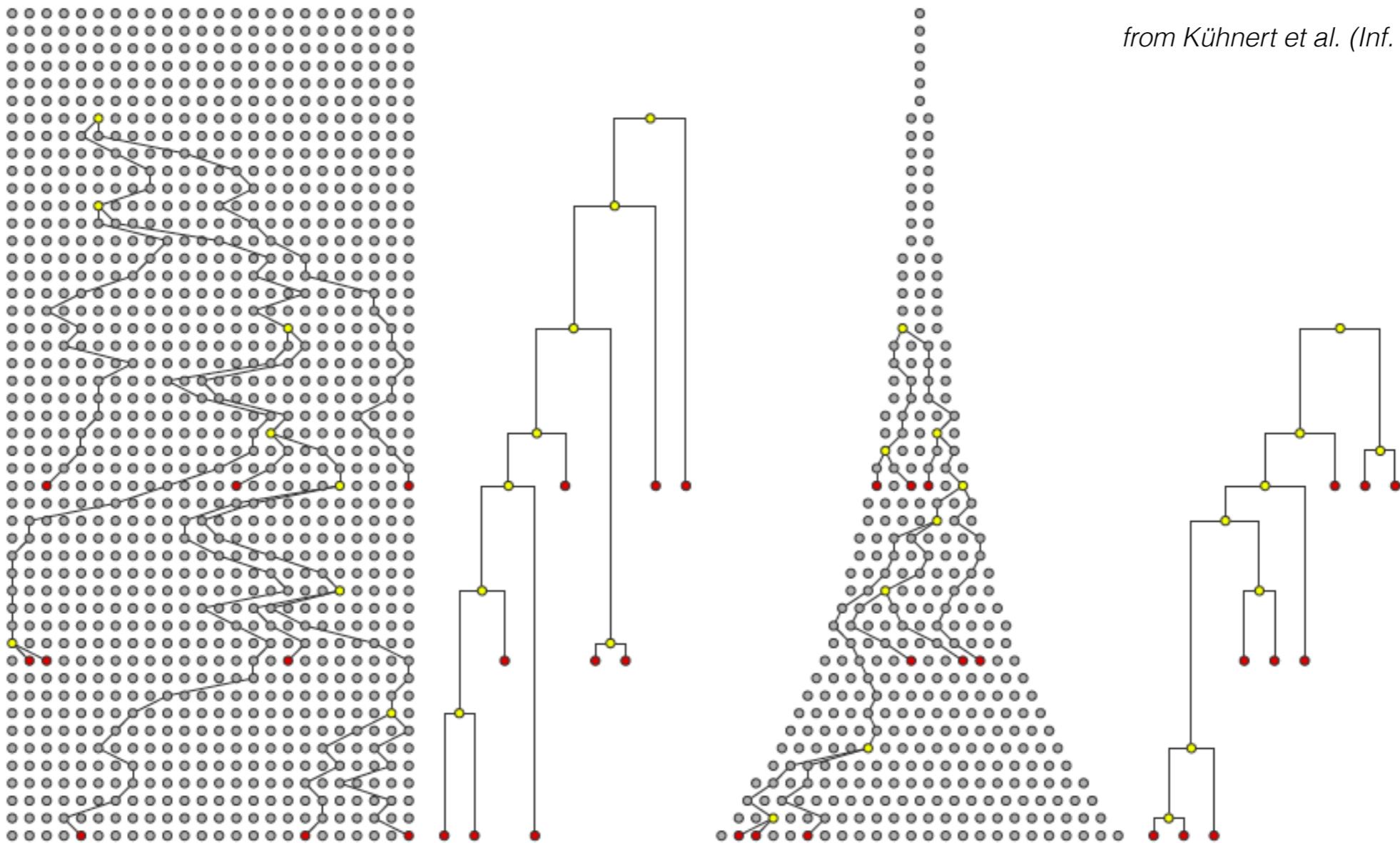
# Fingerprint of the birth-and death rates in phylogenies



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

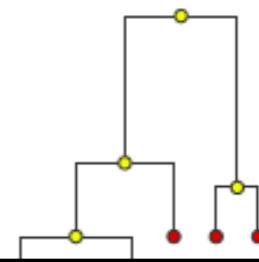
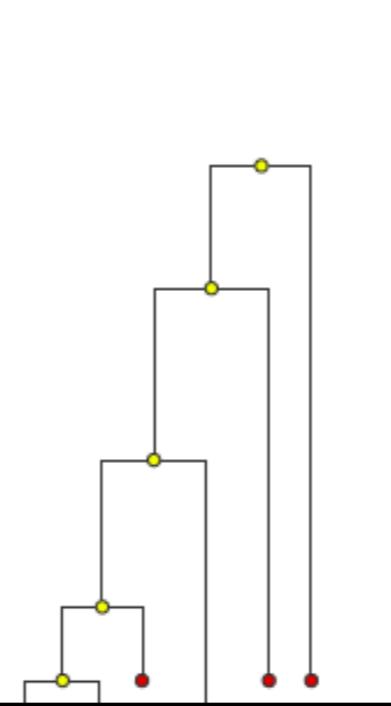
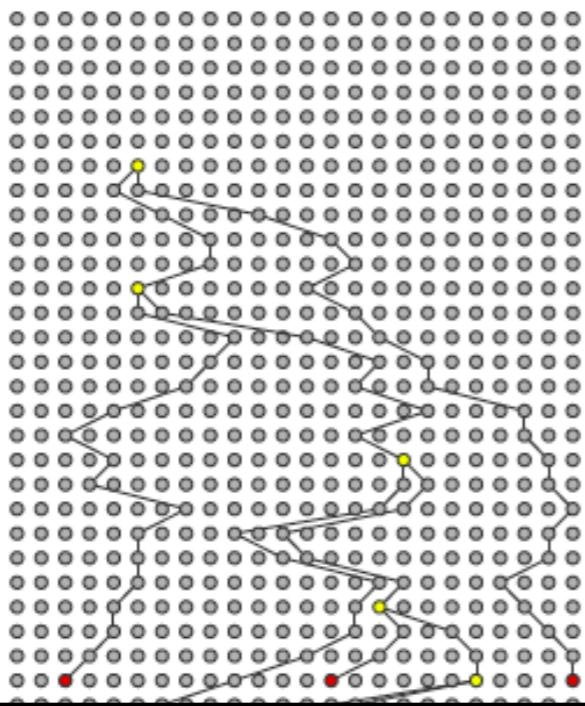
# Coalescent as an alternative model for epidemic spread

---



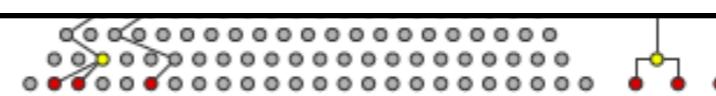
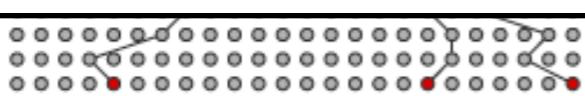
from Kühnert et al. (*Inf. Gen. Evol.*, 2011)

# Coalescent as an alternative model for epidemic spread

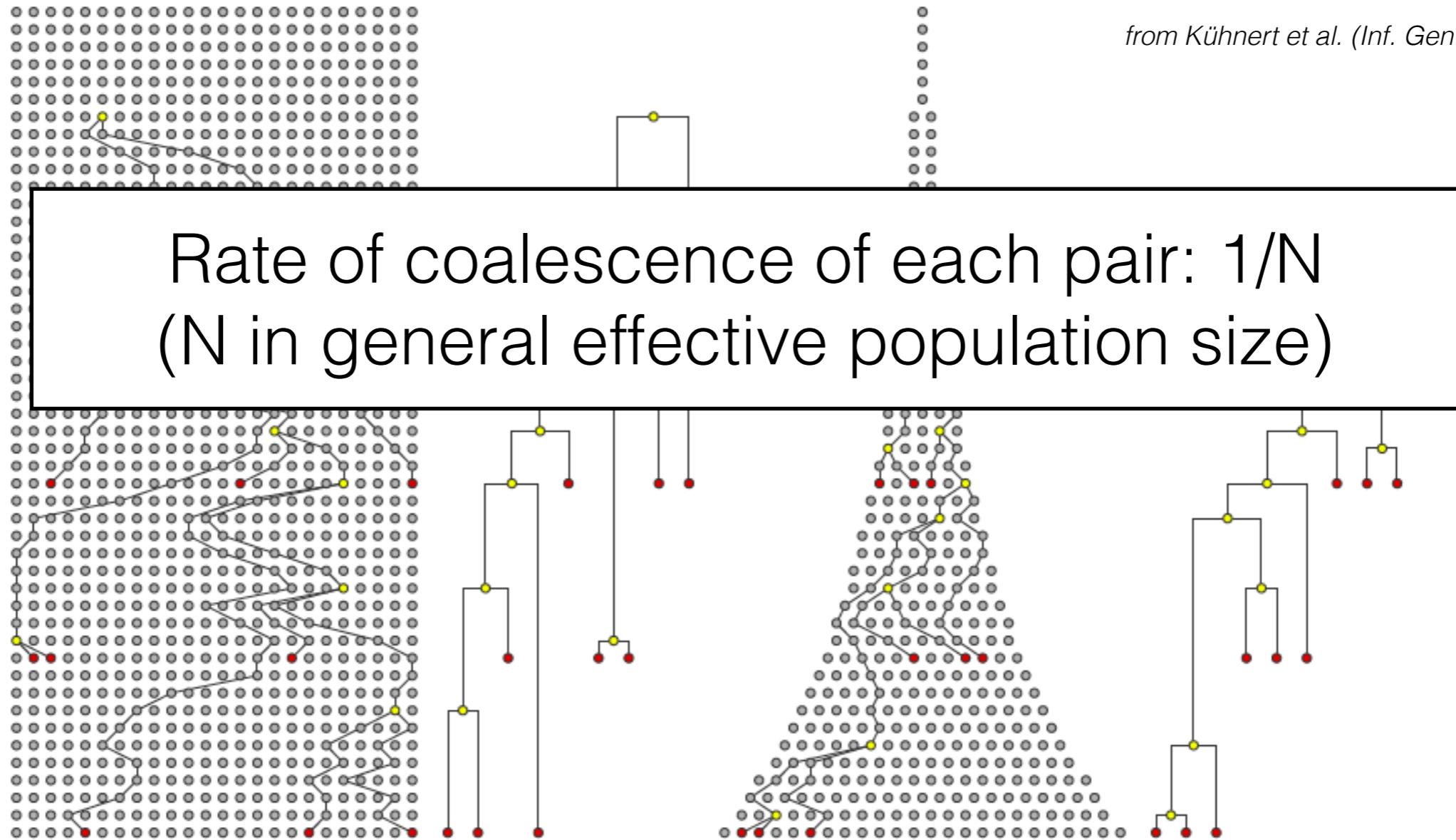


from Kühnert et al. (*Inf. Gen. Evol.*, 2011)

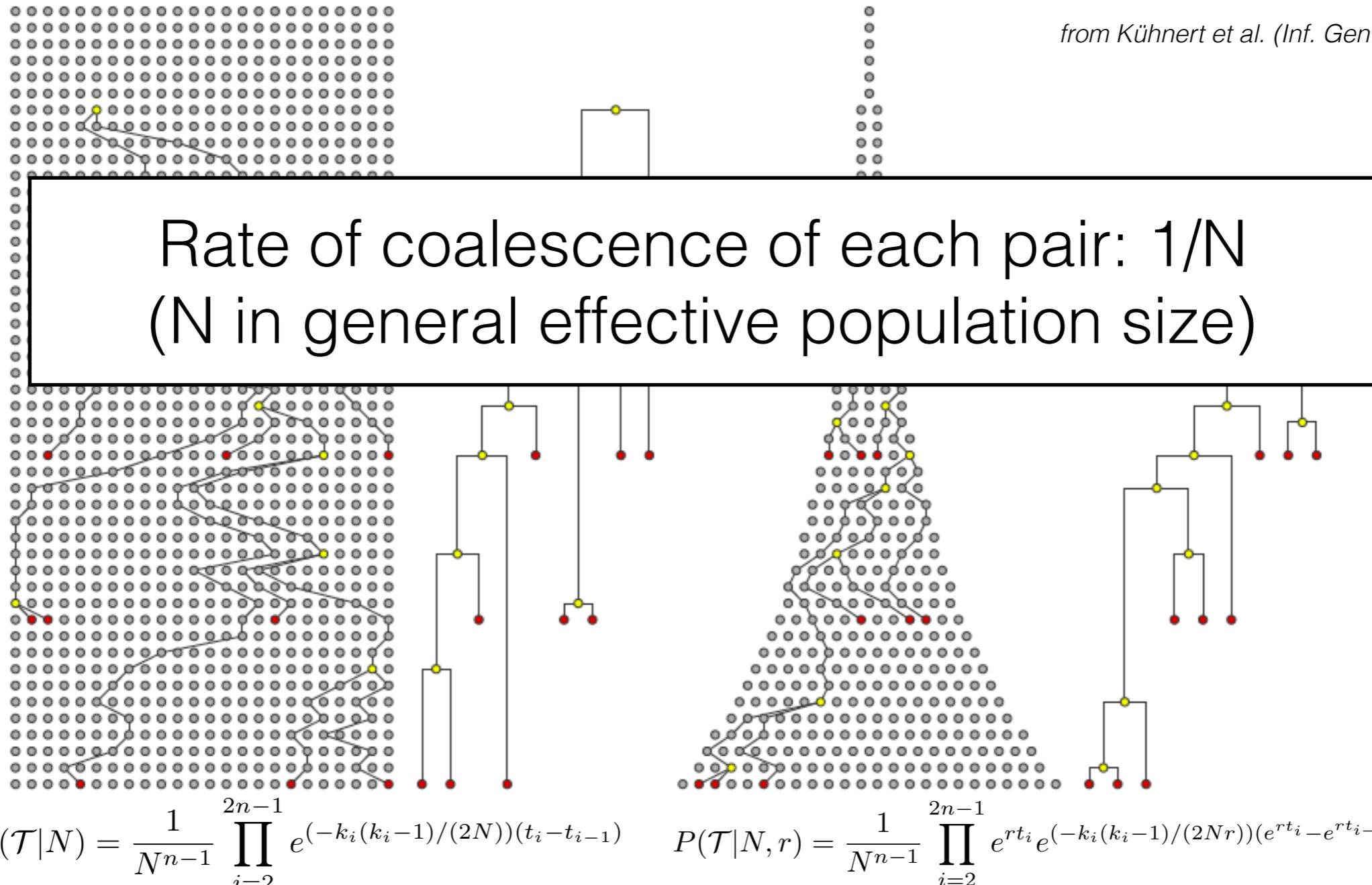
$$P \left( \text{ } \right) ?$$



# Coalescent as an alternative model for epidemic spread

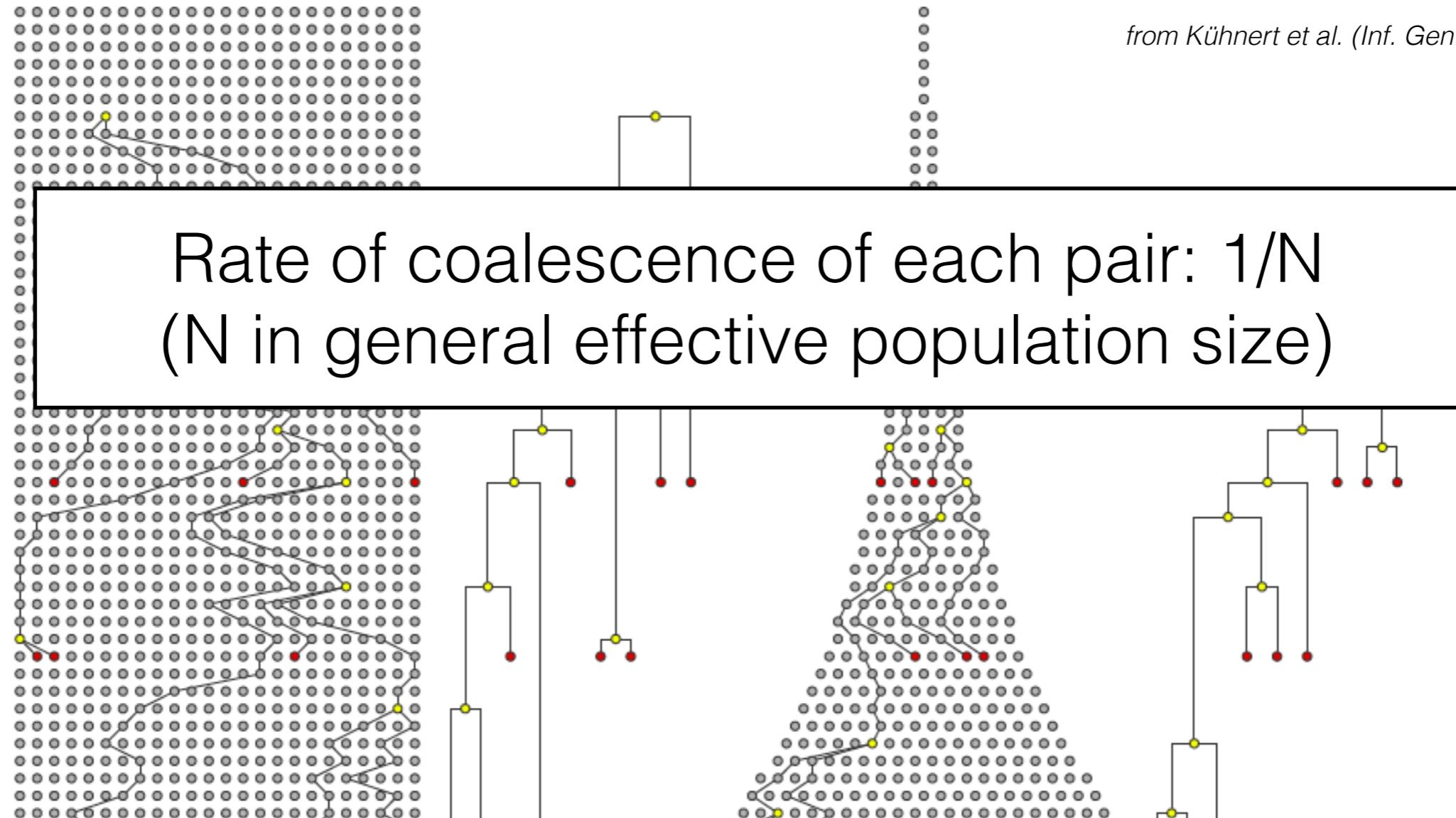


# Coalescent as an alternative model for epidemic spread



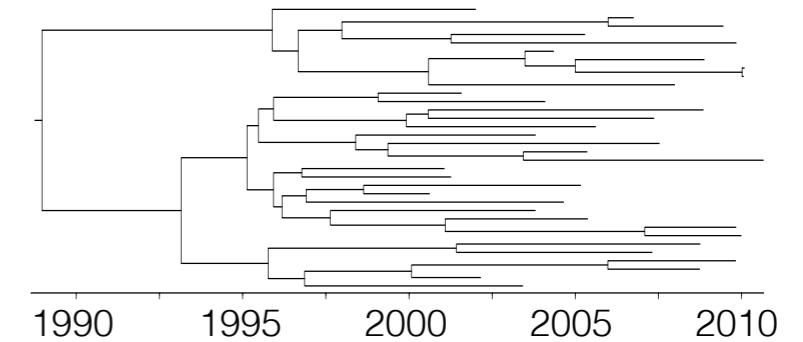
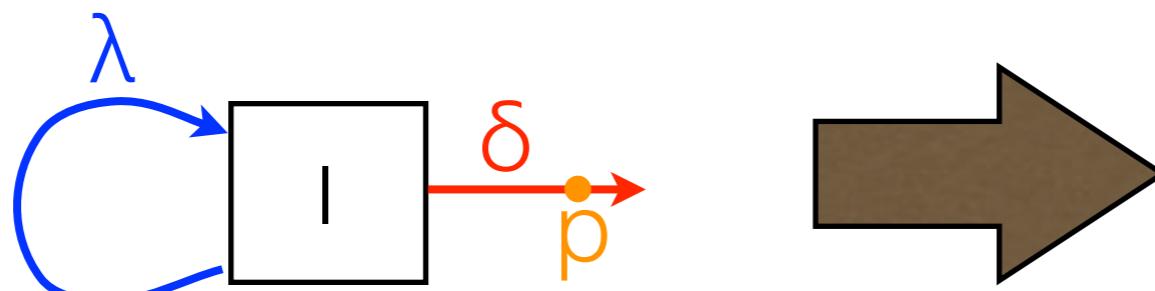
# Coalescent as an alternative model for epidemic spread

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



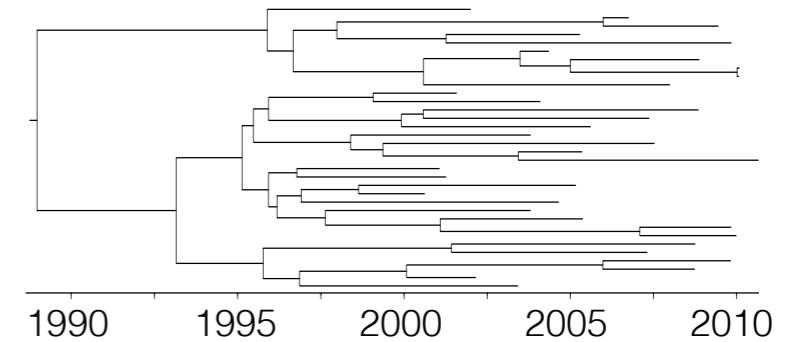
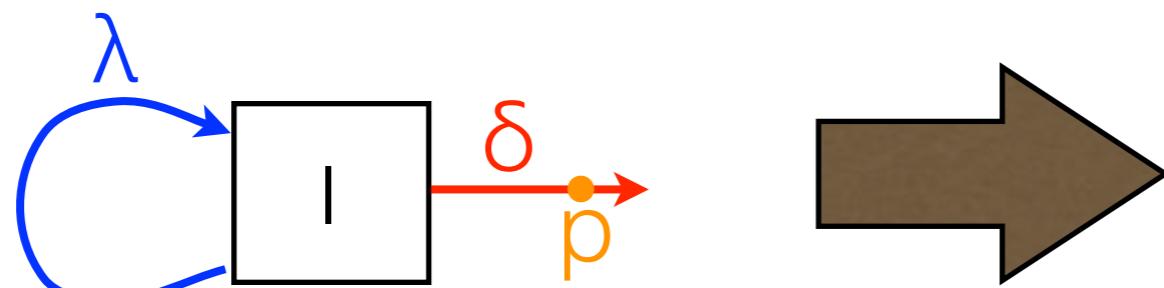
# Simulation study: Comparing exact method to coalescent

Simulating  
epidemic  
outbreak



# Simulation study: Comparing exact method to coalescent

Simulating  
epidemic  
outbreak



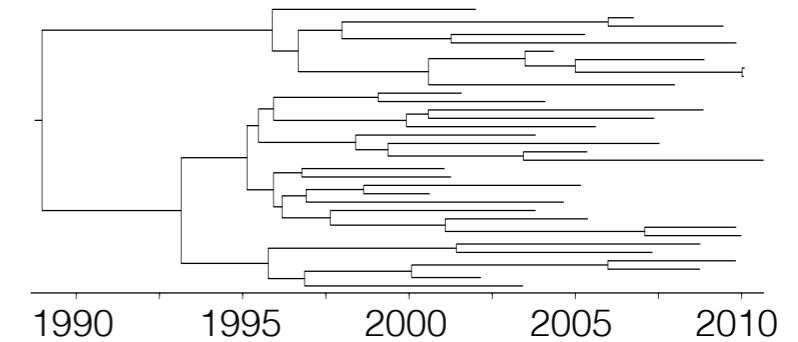
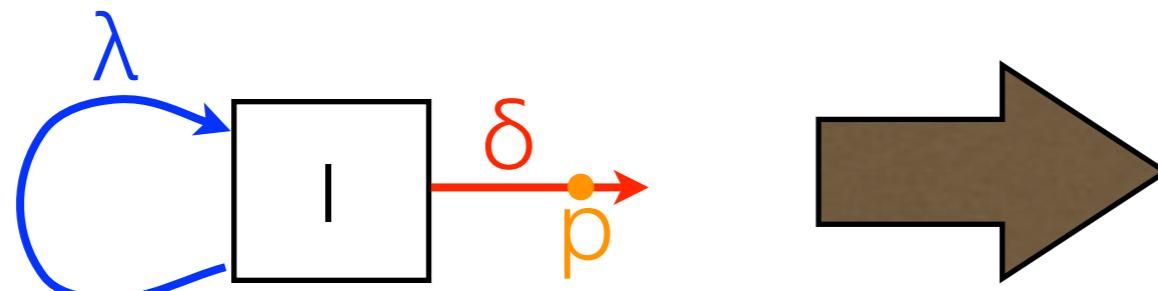
Exact  
method

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

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

# Simulation study: Comparing exact method 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

# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

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

## 2. ...for identifying epidemiological changes over time (skylines)

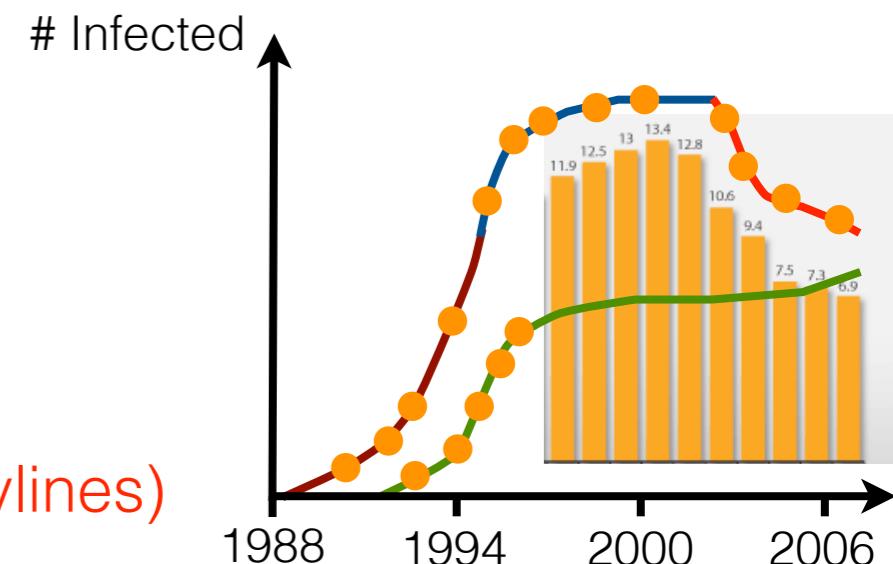
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014) - **BDSIR add-on**

## 3. ...accounting for SIR model dynamics

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

## 4. ...accounting for population structure

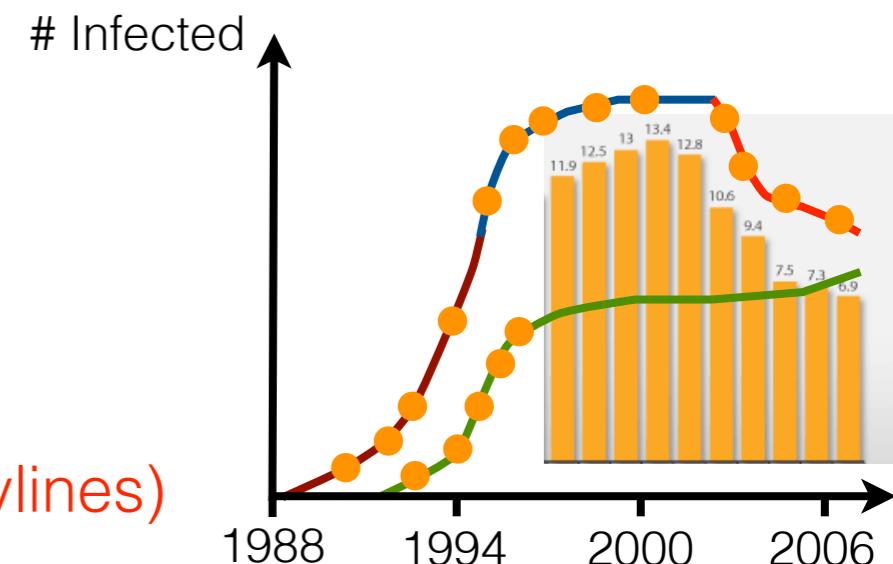
- ▶ C: Volz (*Genetics*, 2012), Vaughan et al. (*Bioinf.*, 2014), de Maio et al. (*PLoS Genetics*, 2015), Mueller et al. (*MBE*, 2017) - **Mascot add-on**
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013), Kühnert et al. (*MBE*, 2016) - **bdmm add-on**



# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

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



## 2. ...for identifying epidemiological changes over time (skylines)

- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014) - **BDSIR add-on**

## 3. ...accounting for SIR model dynamics

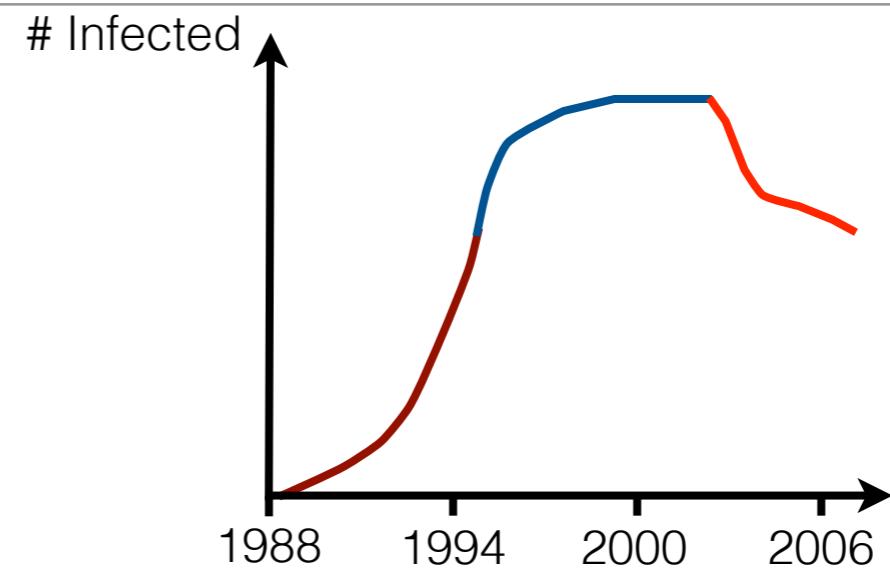
- ▶ C: Volz et al. (*Genetics*, 2009)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014), Leventhal et al. (*MBE*, 2014)

## 4. ...accounting for population structure

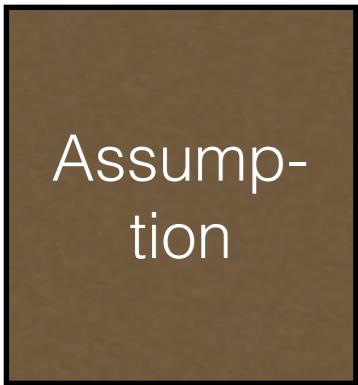
- ▶ C: Volz (*Genetics*, 2012), Vaughan et al. (*Bioinf.*, 2014), de Maio et al. (*PLoS Genetics*, 2015), Mueller et al. (*MBE*, 2017) - **Mascot add-on**
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013), Kühnert et al. (*MBE*, 2016) - **bdmm add-on**

# Bayesian skyline plot relaxes constant growth rate

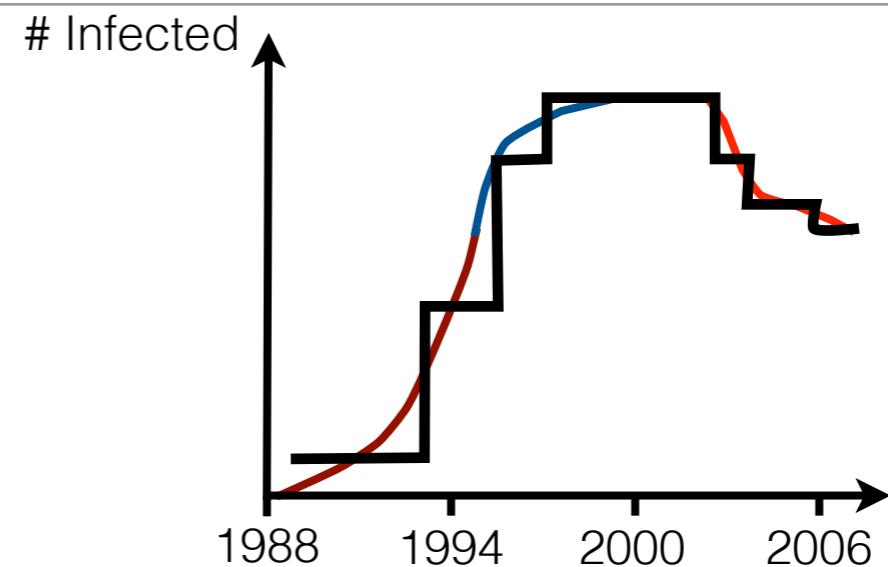
---



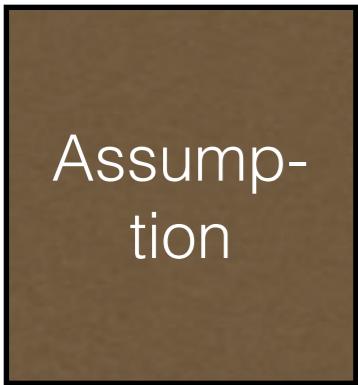
# Bayesian skyline plot relaxes constant growth rate



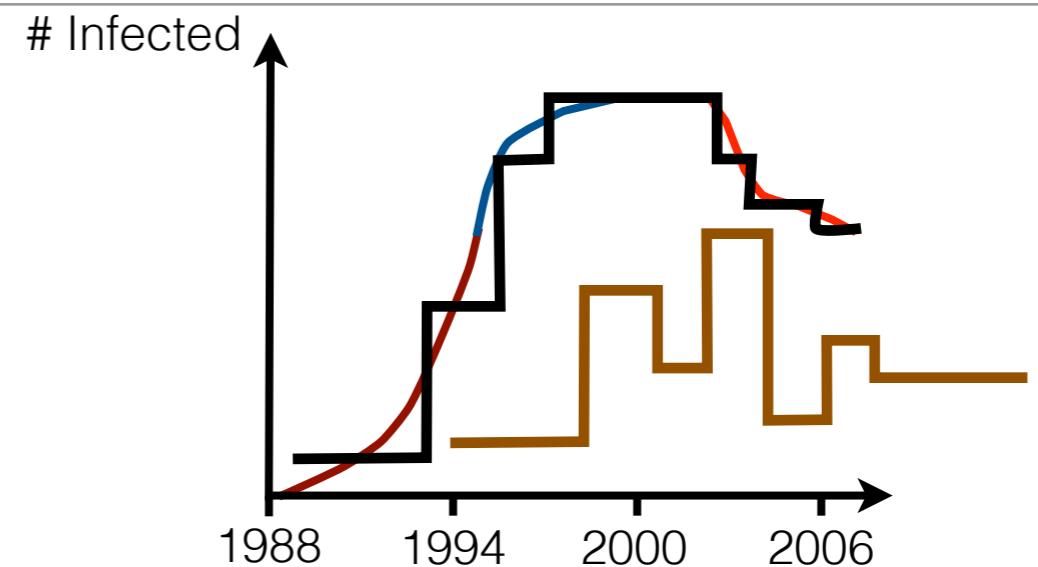
Piecewise constant population size



# Bayesian skyline plot relaxes constant growth rate



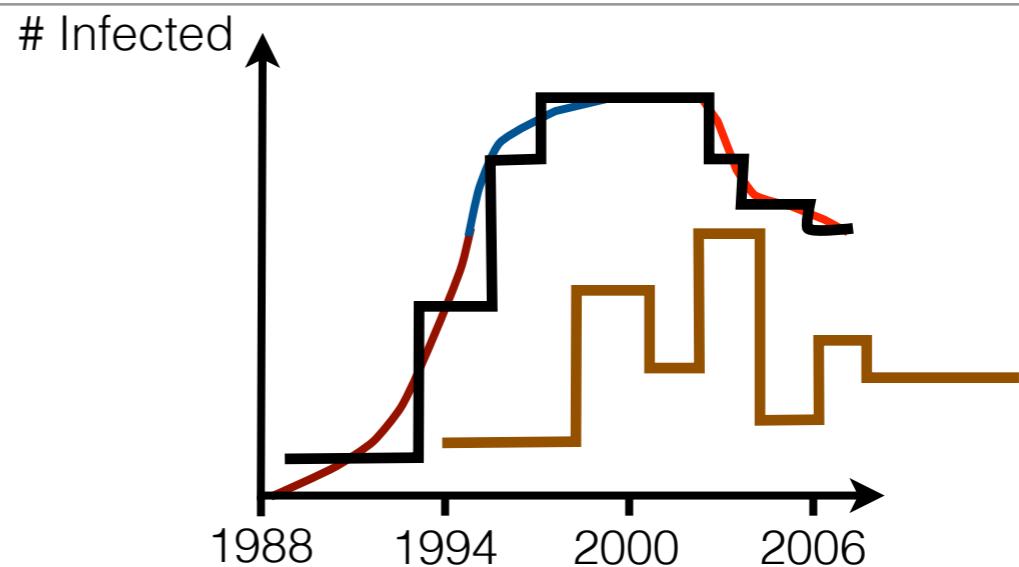
Piecewise constant population size



# Bayesian skyline plot relaxes constant growth rate

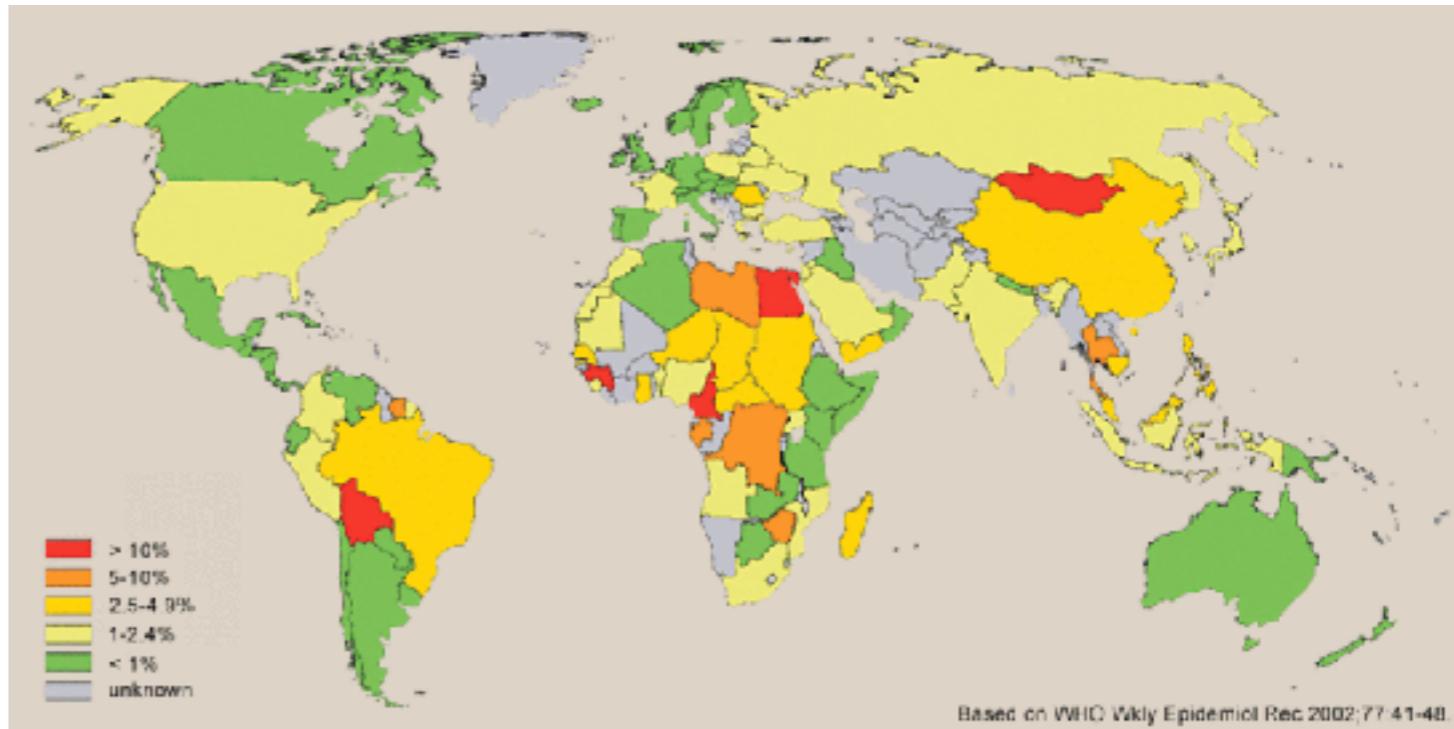
Assump-  
tion

Piecewise constant  
population size



HCV in  
Egypt

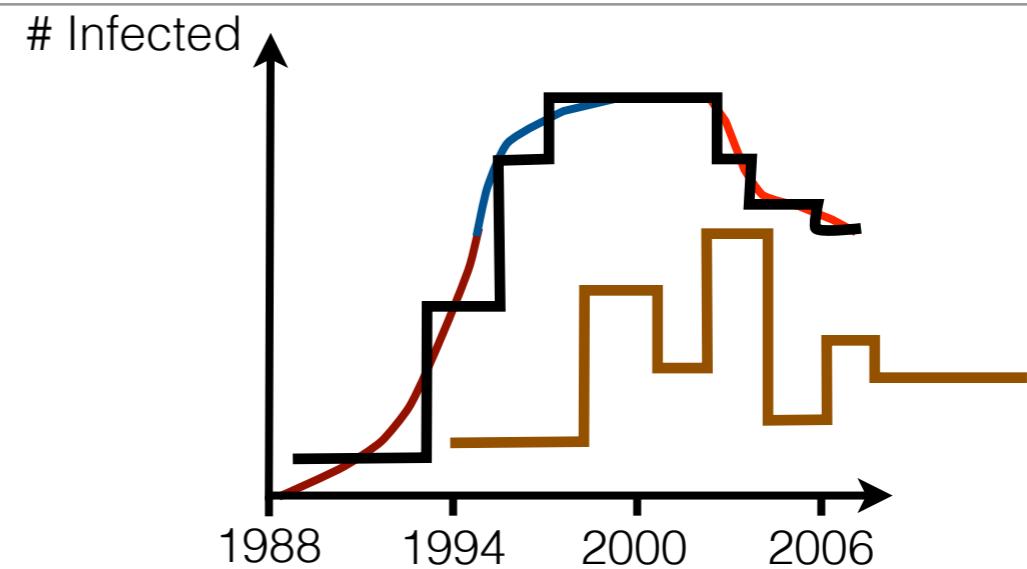
- ▶ Egypt highest HCV prevalence worldwide (about 13 %)
- ▶ Neighboring countries low prevalence
- ▶ When did the high prevalence emerge?
- ▶ Why?



# Bayesian skyline plot relaxes constant growth rate

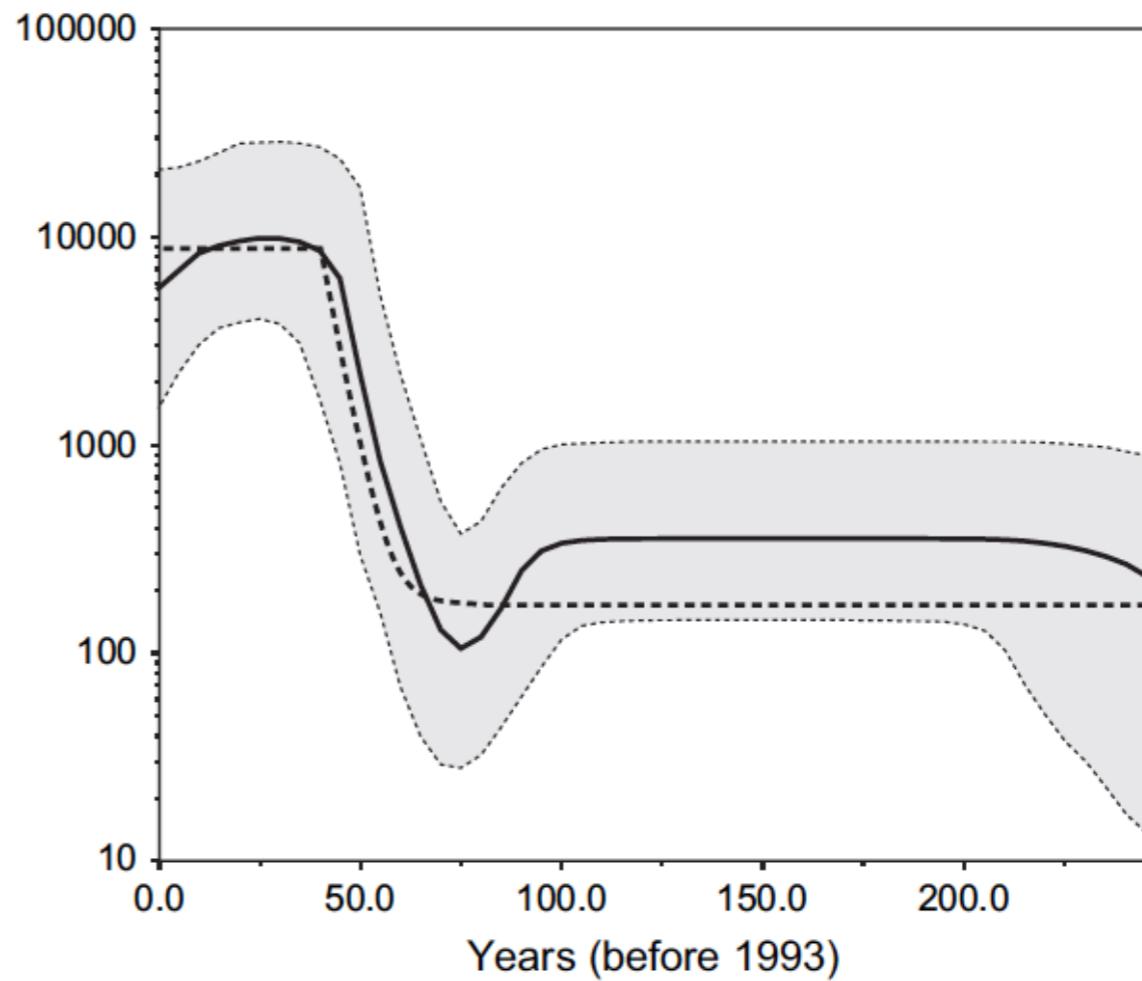
Assump-  
tion

Piecewise constant  
population size



HCV in  
Egypt

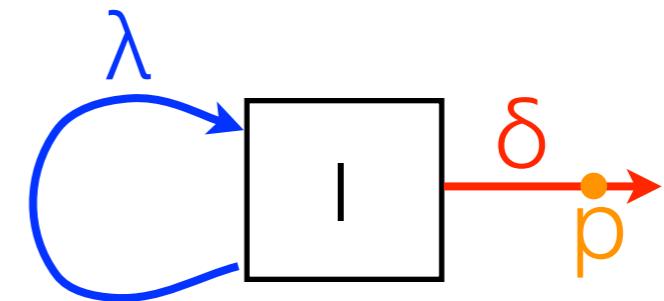
- ▶ Egypt highest HCV prevalence worldwide (about 13 %)
- ▶ Neighboring countries low prevalence
- ▶ When did the high prevalence emerge?
- ▶ Why?



# Birth-death-skyline plot

---

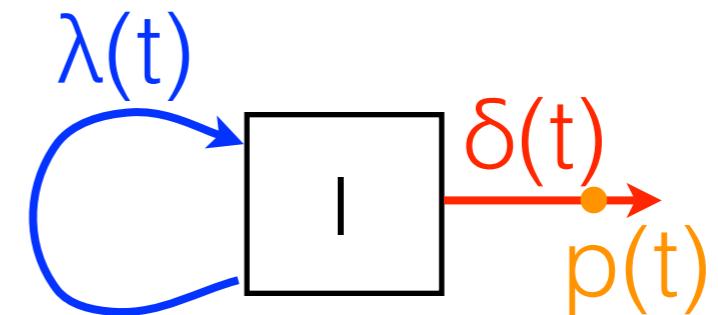
Epidemiological rates may change through time:



# Birth-death-skyline plot

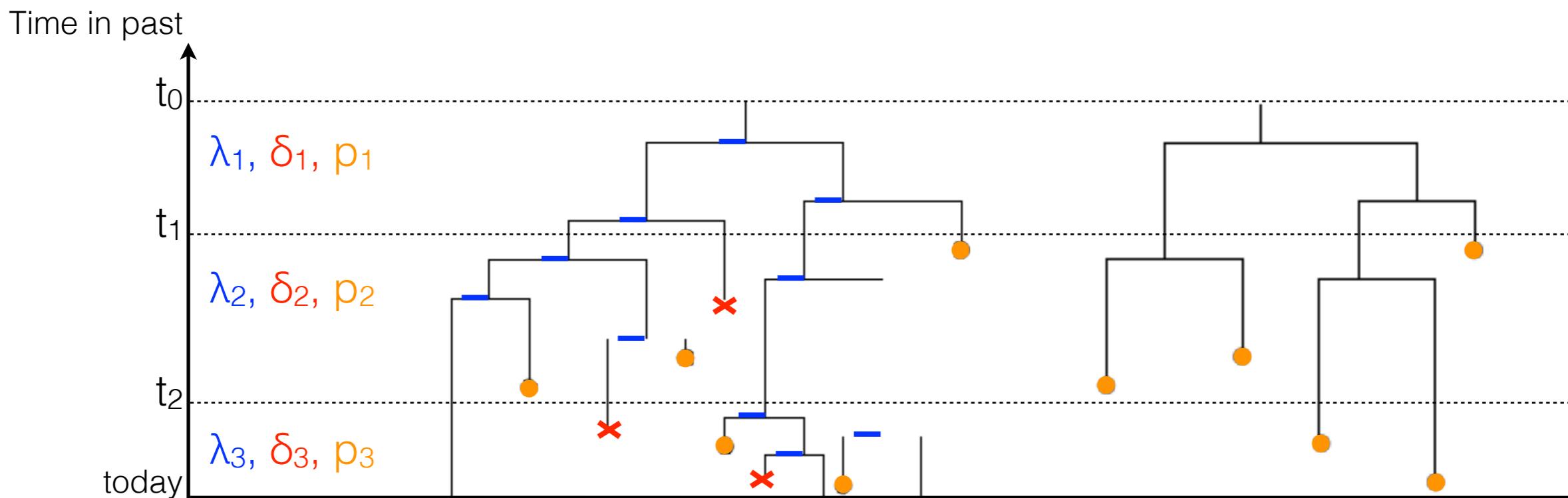
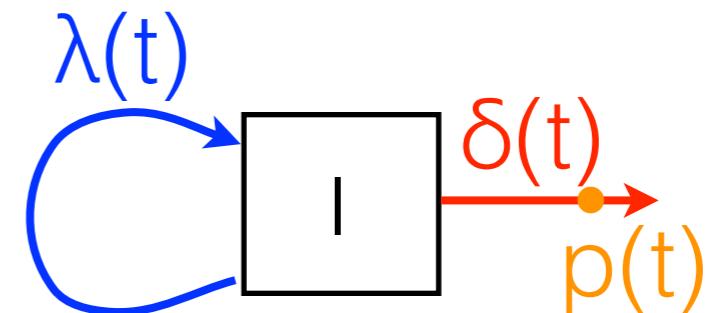
---

Epidemiological rates may change through time:



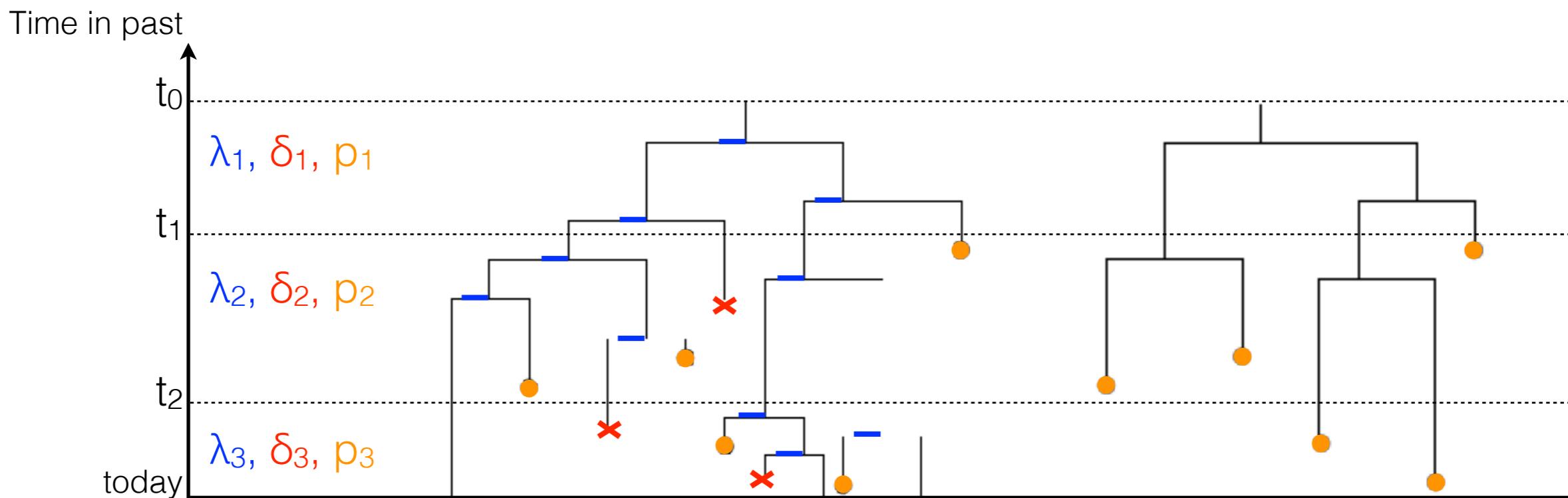
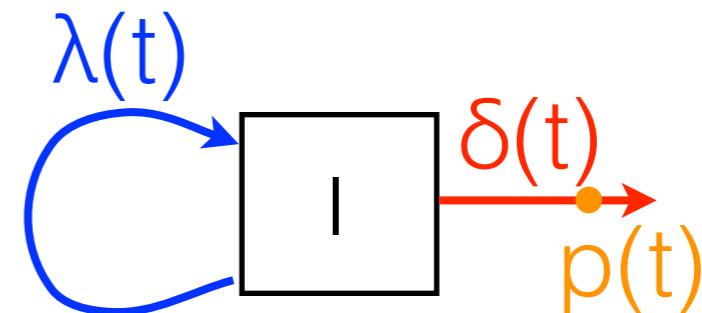
# Birth-death-skyline plot

Epidemiological rates may change through time:



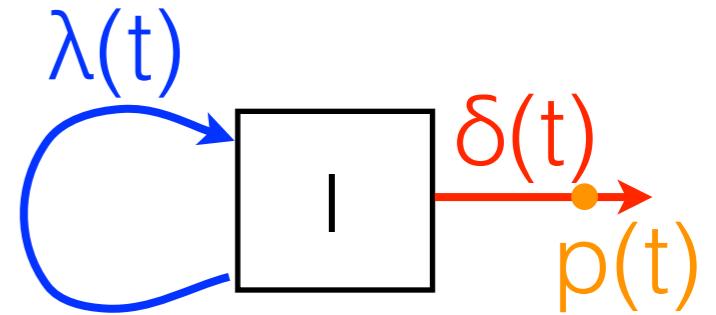
# Birth-death-skyline plot

Epidemiological rates may change through time:



**BEAST add-on BDSKY**

# Hepatitis C virus in Egypt

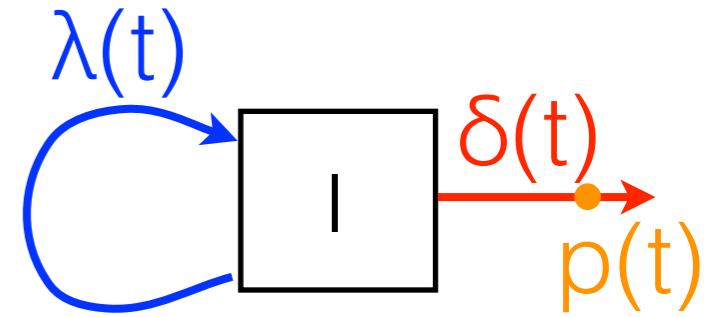


Effective reproductive number  $R_e(t)$

generalizes

Basic reproductive number  $R_0 = R_e(0)$

# Hepatitis C virus in Egypt



Effective reproductive number  $R_e(t)$   
generalizes

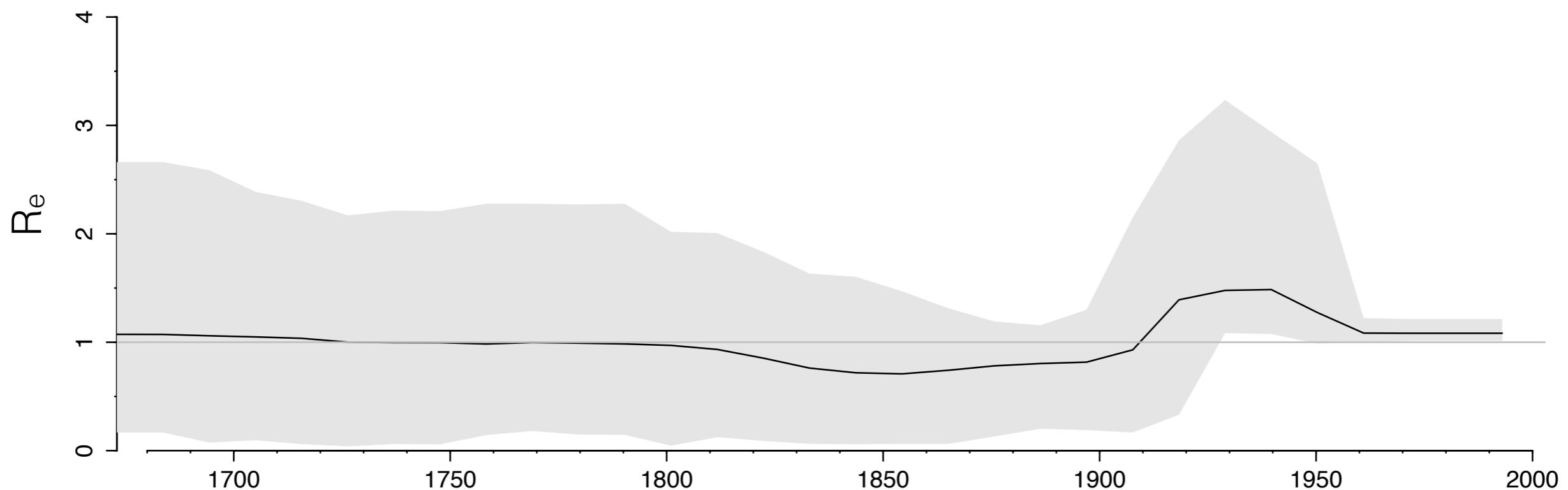
Basic reproductive number  $R_0 = R_e(0)$

Analysis of 63 seq:

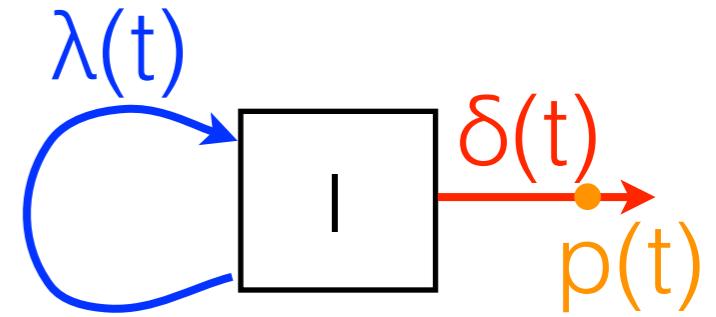
$$R_e(t) = \lambda(t)/\delta(t)$$

median

95% HPD interval



# Hepatitis C virus in Egypt



Effective reproductive number  $R_e(t)$   
generalizes

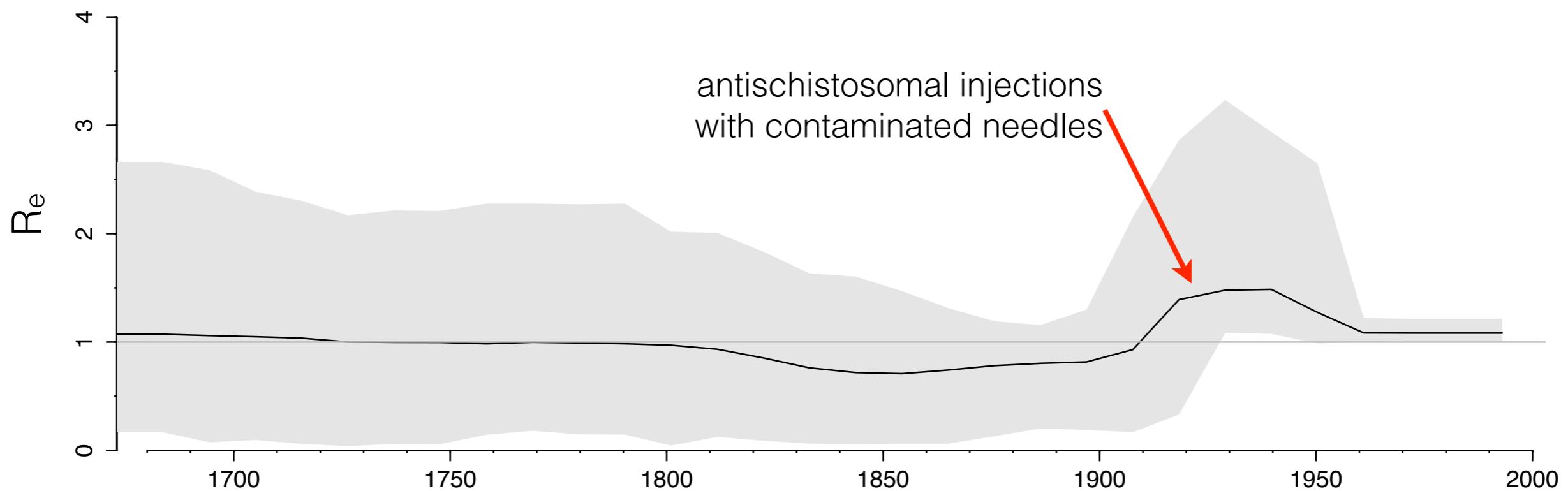
Basic reproductive number  $R_0 = R_e(0)$

Analysis of 63 seq:

$$R_e(t) = \lambda(t)/\delta(t)$$

median

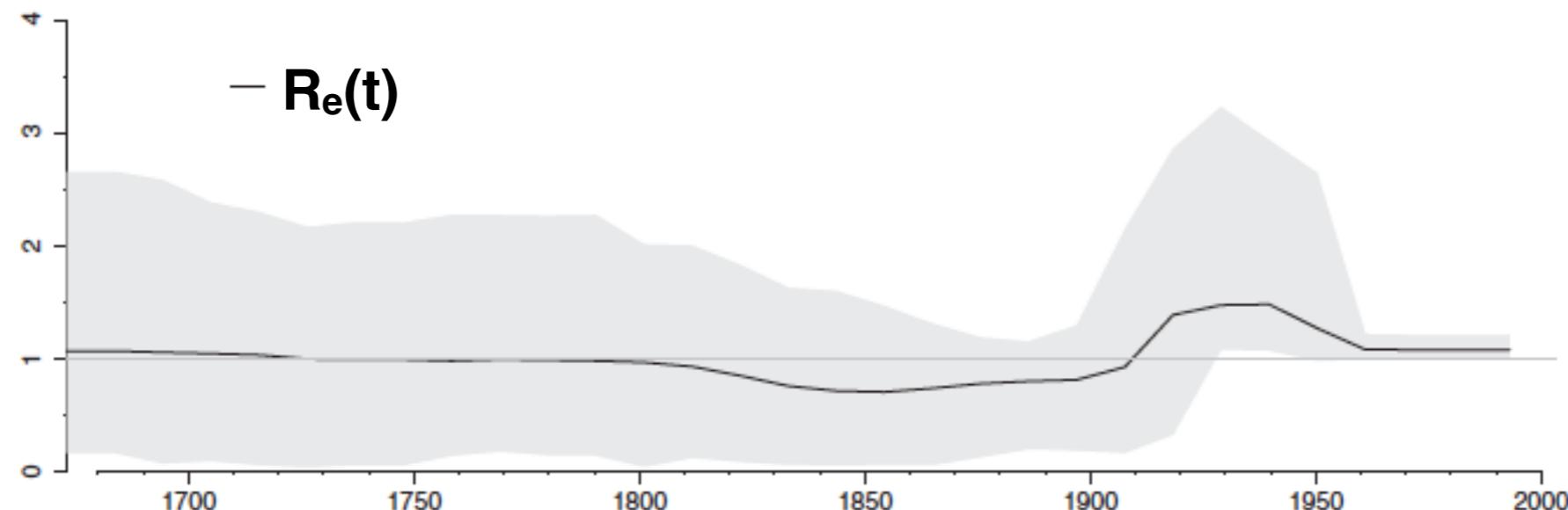
95% HPD interval



# The two skyline models

---

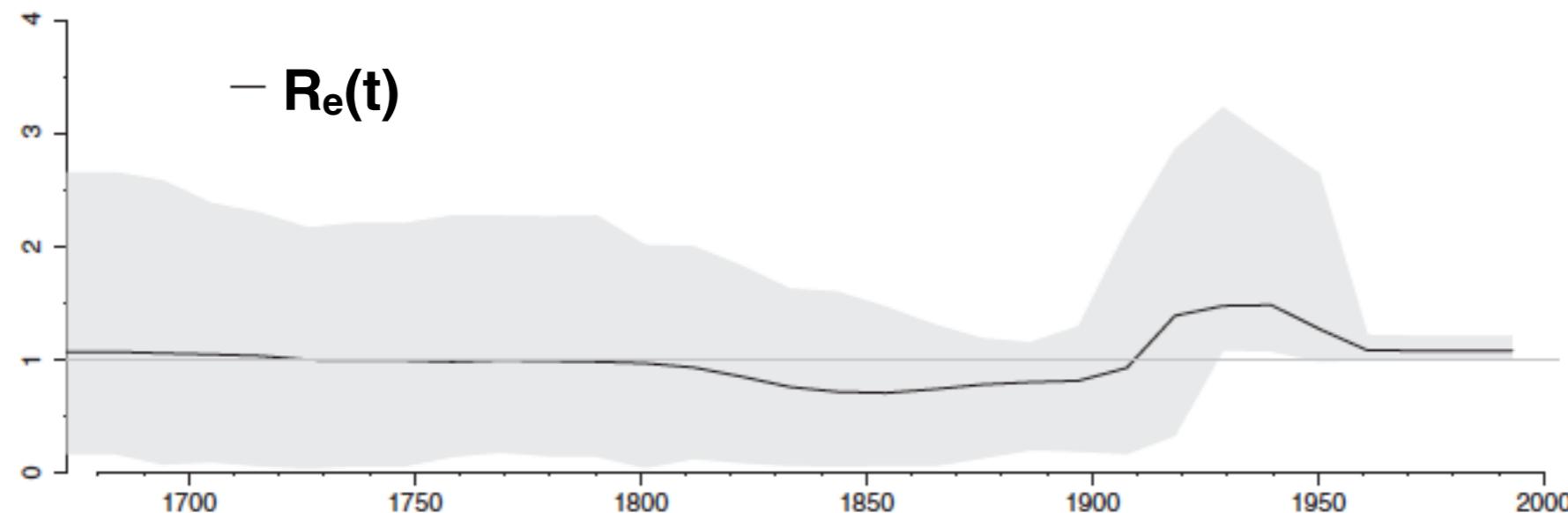
Birth-death skyline plot: effective reproductive number



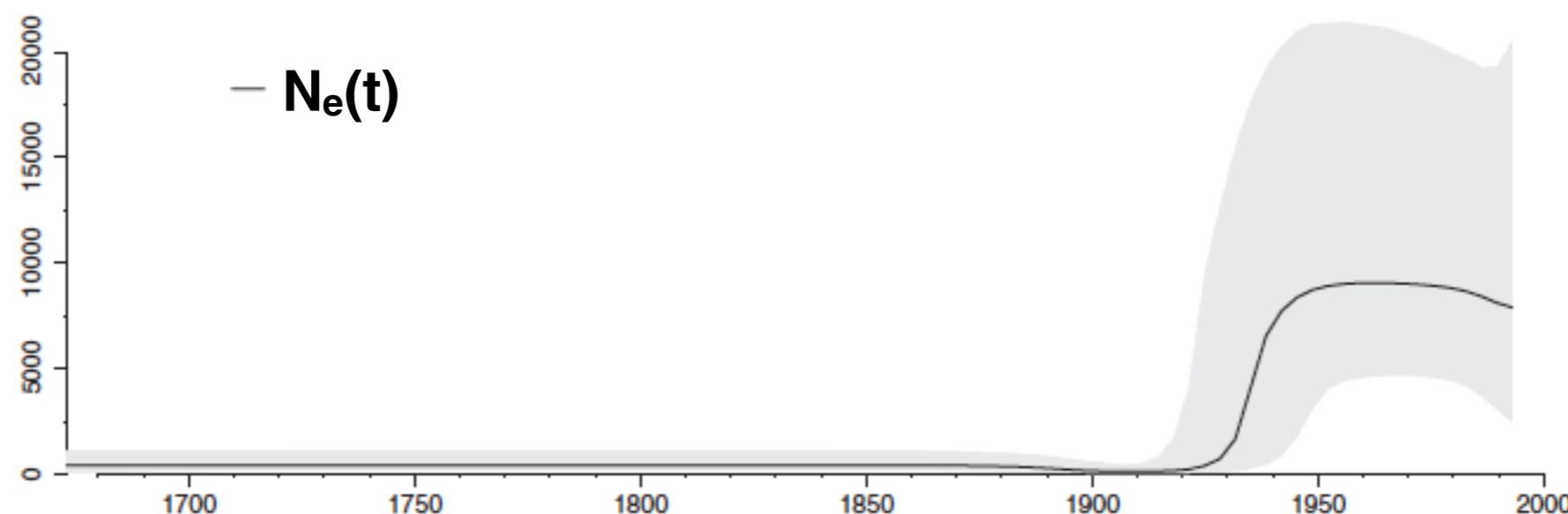
# The two skyline models

---

Birth-death skyline plot: effective reproductive number

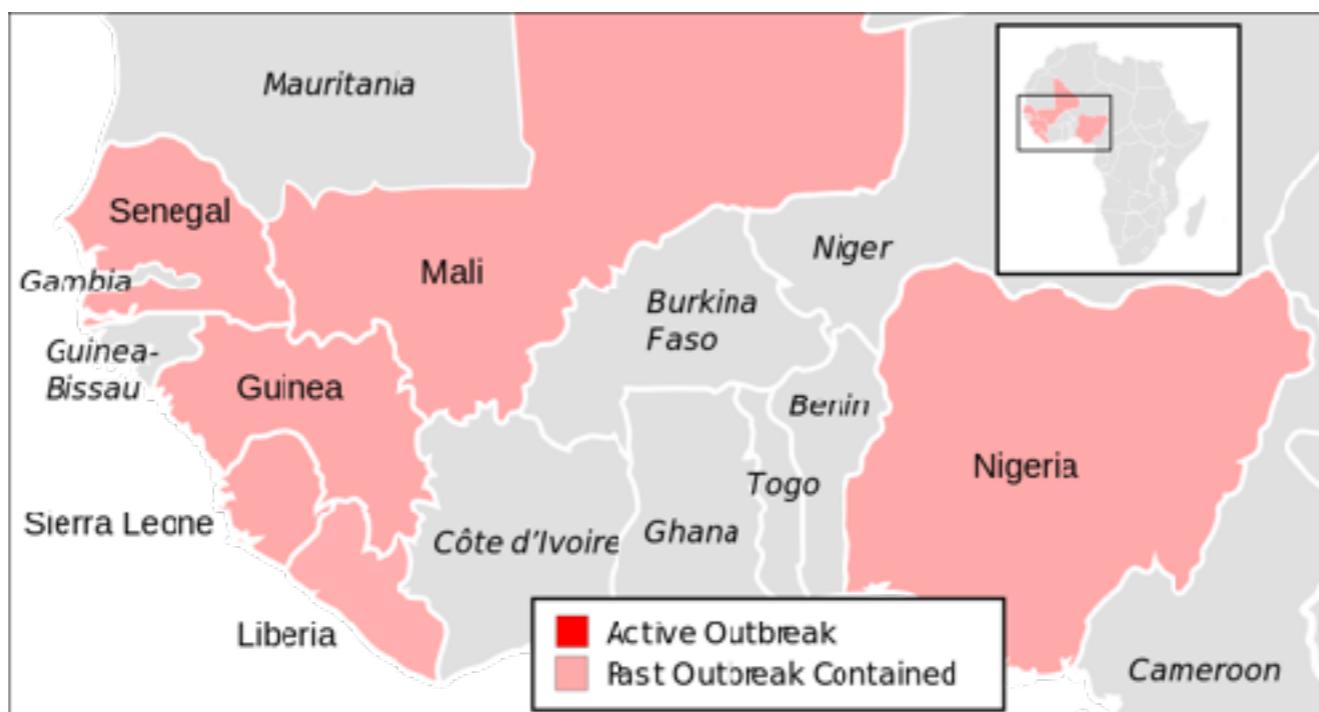


Coalescent skyline plot: effective population size



# Ebola in West Africa

---



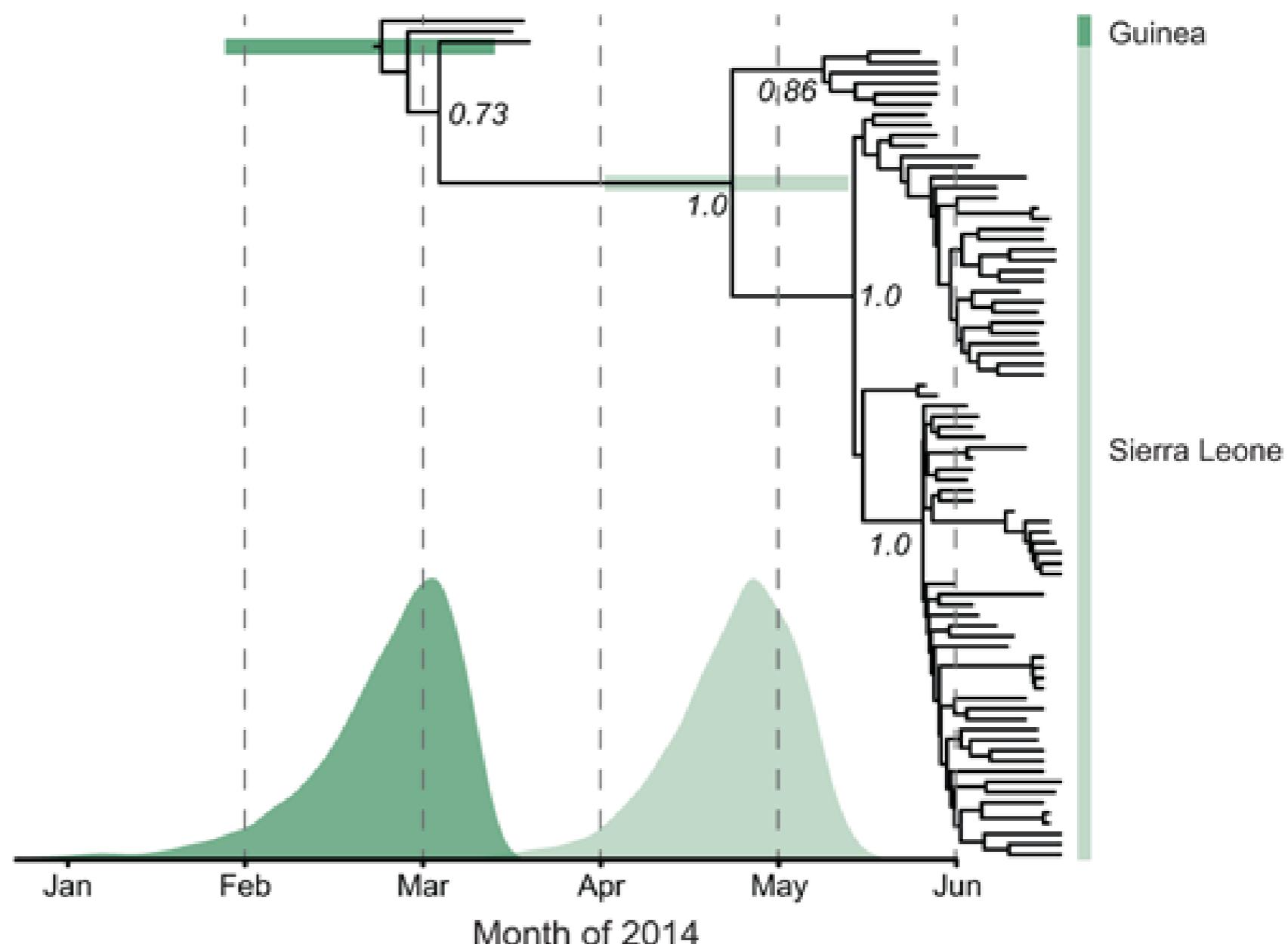
28657 suspected cases, 11325 deaths (probably an underestimate)

# Single zoonosis in December caused the current epidemic

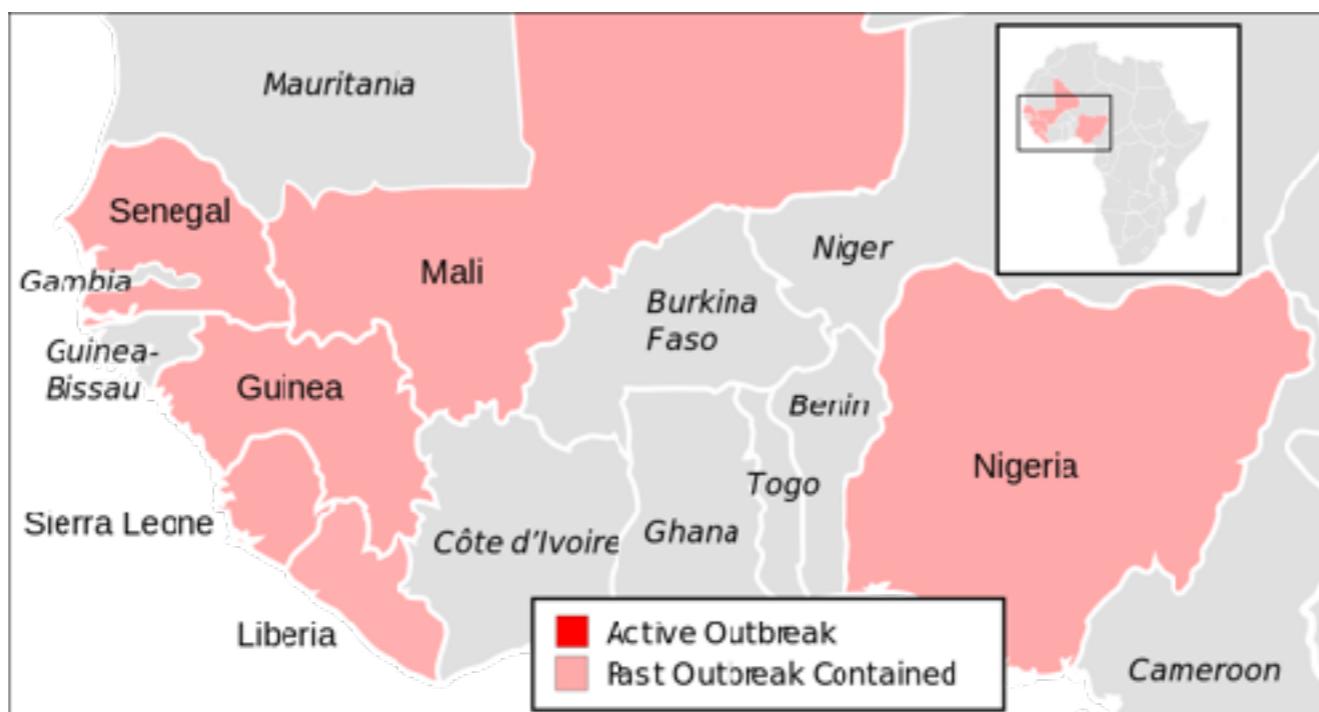


## Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak

Stephen K. Gire,<sup>1,2\*</sup> Augustine Goba,<sup>3\*†</sup> Kristian G. Andersen,<sup>1,2\*†</sup> Rachel S. G. Sealton,<sup>2,4\*‡</sup> Daniel J. Park,<sup>2\*</sup> Lansana Kanneh,<sup>3</sup> Simbirie Jalloh,<sup>3</sup> Mambu Momoh,<sup>3,5</sup> Mohamed Fullah,<sup>3,5‡</sup> Gytis Dudas,<sup>6</sup> Shirlee Wohl,<sup>1,2,7</sup> Lina M. Moses,<sup>8</sup> Nathan L. Yozwiak,<sup>1,2</sup> Sarah Winnicki,<sup>1,2</sup> Christian B. Matranga,<sup>2</sup> Christine M. Malboeuf,<sup>2</sup> James Qu,<sup>2</sup> Adrienne D. Gladden,<sup>2</sup> Stephen F. Schaffner,<sup>1,2</sup> Xiao Yang,<sup>2</sup> Pan-Pan Jiang,<sup>1,2</sup> Mahan Nekoui,<sup>1,2</sup> Andres Colubri,<sup>1</sup> Moinya Ruth Coomber,<sup>3</sup> Mbalu Fonnie,<sup>3‡</sup> Alex Moigboi,<sup>3‡</sup> Michael Gbakie,<sup>3</sup> Fatima K. Kamara,<sup>3</sup> Veronica Tucker,<sup>3</sup> Edwin Konuwa,<sup>3</sup> Sidiki Saffa,<sup>3</sup> Josephine Sellu,<sup>3</sup> Abdul Azziz Jalloh,<sup>3</sup> Alice Kovoma,<sup>3‡</sup> James Koninga,<sup>3</sup> Ibrahim Mustapha,<sup>3</sup> Kandeh Kargbo,<sup>3</sup> Momoh Foday,<sup>3</sup> Mohamed Yillah,<sup>3</sup> Franklyn Kanneh,<sup>3</sup> Willie Robert,<sup>3</sup> James L. B. Massally,<sup>3</sup> Sinéad B. Chapman,<sup>2</sup> James Bochicchio,<sup>2</sup> Cheryl Murphy,<sup>2</sup> Chad Nusbaum,<sup>2</sup> Sarah Young,<sup>2</sup> Bruce W. Birren,<sup>2</sup> Donald S. Grant,<sup>3</sup> John S. Scheiffelin,<sup>8</sup> Eric S. Lander,<sup>2,7,9</sup> Christian Happi,<sup>10</sup> Sahr M. Gevao,<sup>11</sup> Andreas Gnarke,<sup>2§</sup> Andrew Rambaut,<sup>6,12,13§</sup> Robert F. Garry,<sup>8§</sup> S. Humarr Khan,<sup>3‡§</sup> Pardis C. Sabeti<sup>1,2†§</sup>

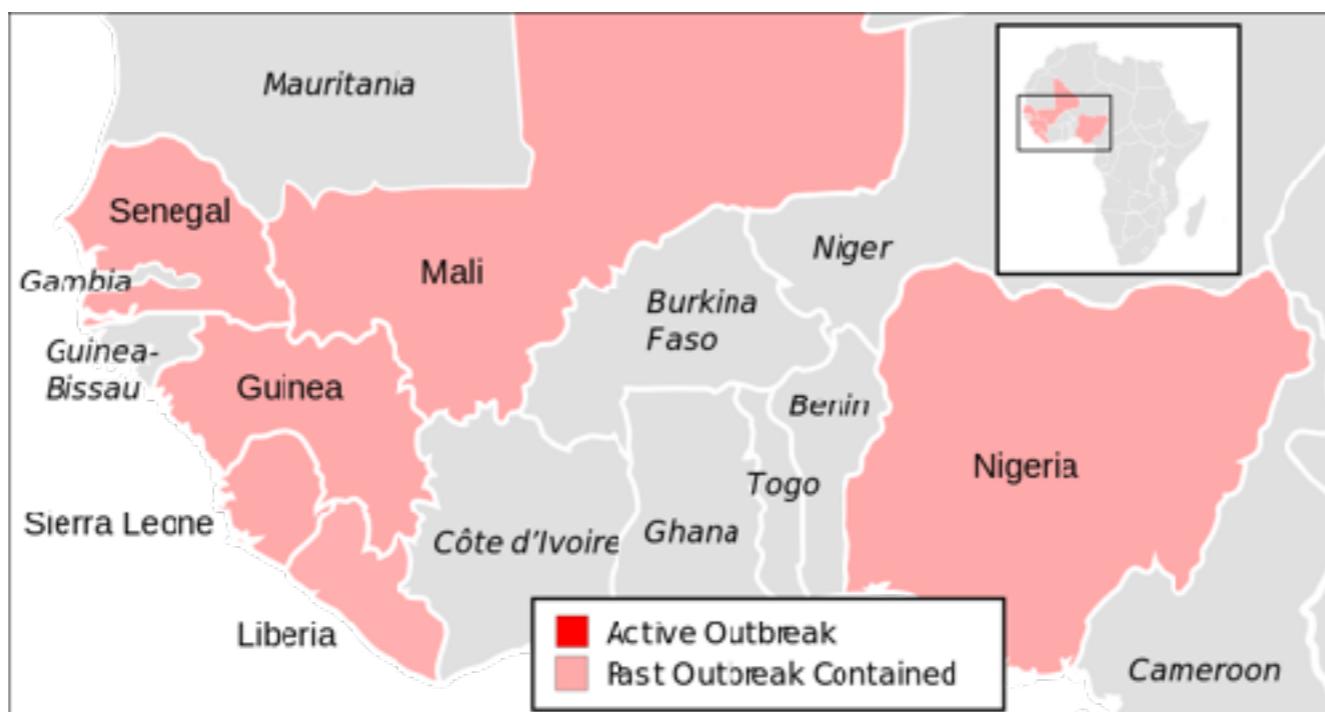


# Ebola in West Africa



28657 suspected cases, 11325 deaths (probably an underestimate)

# Ebola in West Africa



28657 suspected cases, 11325 deaths (probably an underestimate)

$R_e(0) = \lambda/\mu$ : basic reproductive number

$R_0$  around 2 from genetic sequences

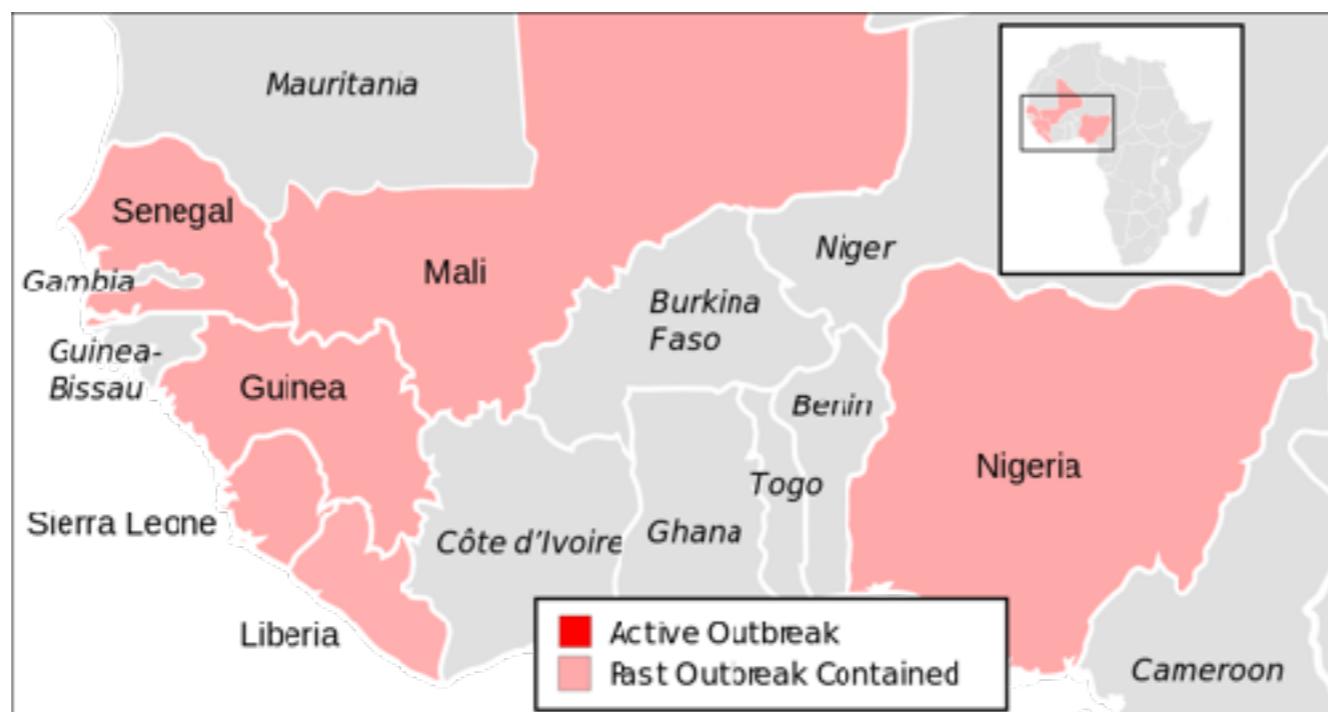
(Stadler et al.; PLoS Curr. Outbreaks, 2014)

TagesAnzeiger

Okt. 08, 2014

**ETH-Forscher errechnen das wahre Ausmass der Ebola-Epidemie**

# Ebola in West Africa



28657 suspected cases, 11325 deaths (probably an underestimate)

$R_e(0) = \lambda/\mu$ : basic reproductive number

$R_0$  around 2 from genetic sequences

(Stadler et al.; PLoS Curr. Outbreaks, 2014)

$R_0$  around 2 from incidences [= new cases through time]

(WHO Ebola Response Team; New Engl. Jour. Med., 2014)

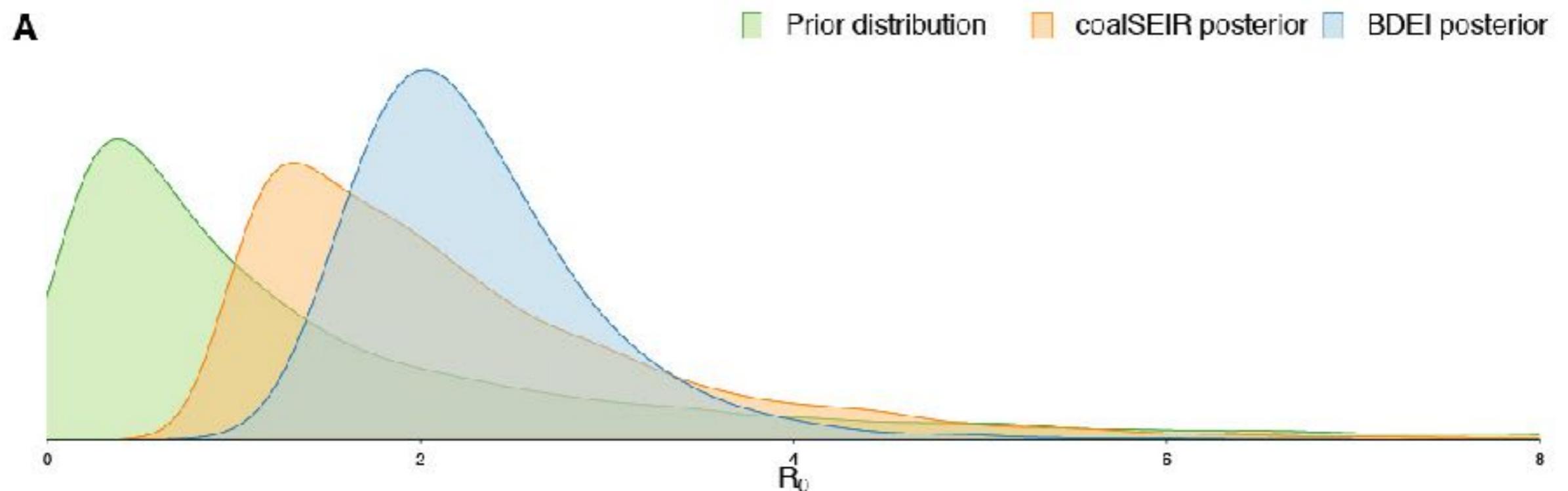
TagesAnzeiger

Okt. 08, 2014

**ETH-Forscher errechnen das wahre Ausmass der Ebola-Epidemie**

# Estimation of $R_e(0)$ for Ebola in Sierra Leone (72 genomes)

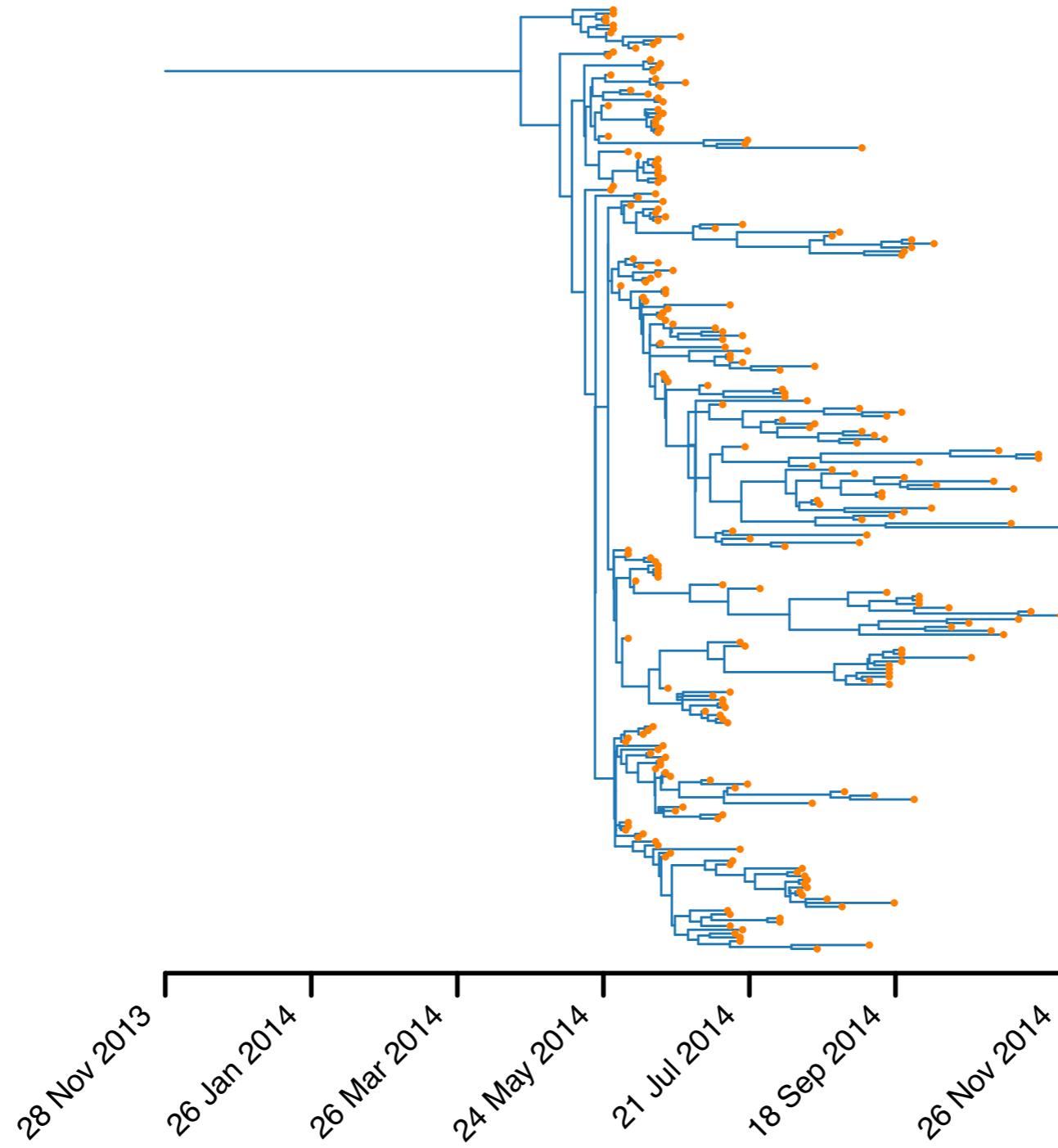
$$f[\mathcal{T}, \eta, \theta | data] = \frac{f[data | \mathcal{T}, \theta] f[\mathcal{T} | \eta] f[\eta, \theta]}{f[data]}$$





# Results in Southeast Sierra Leone

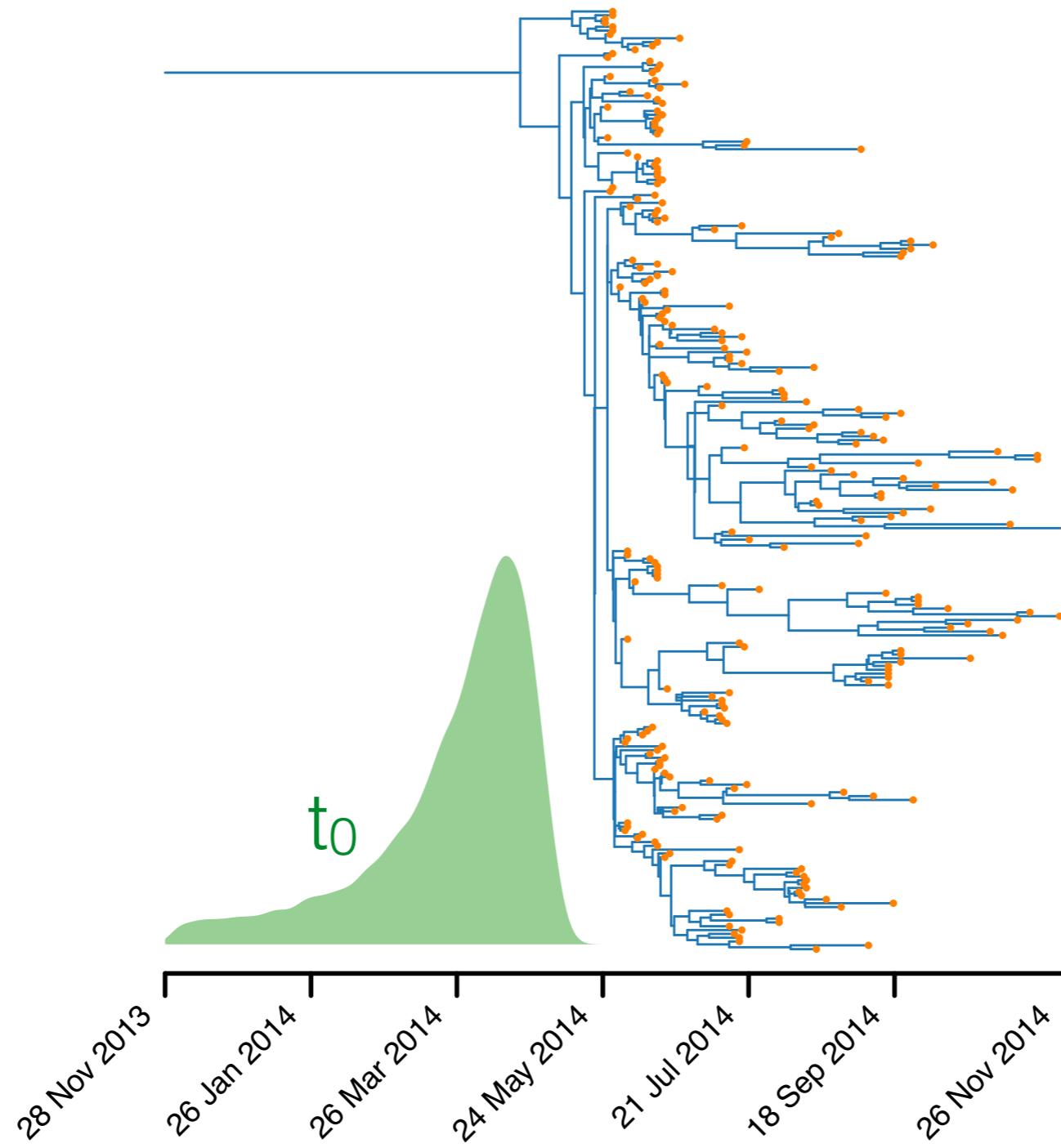
(Based on **246** genomes sampled between **25 May** and **26 Nov 2014**)





# Results in Southeast Sierra Leone

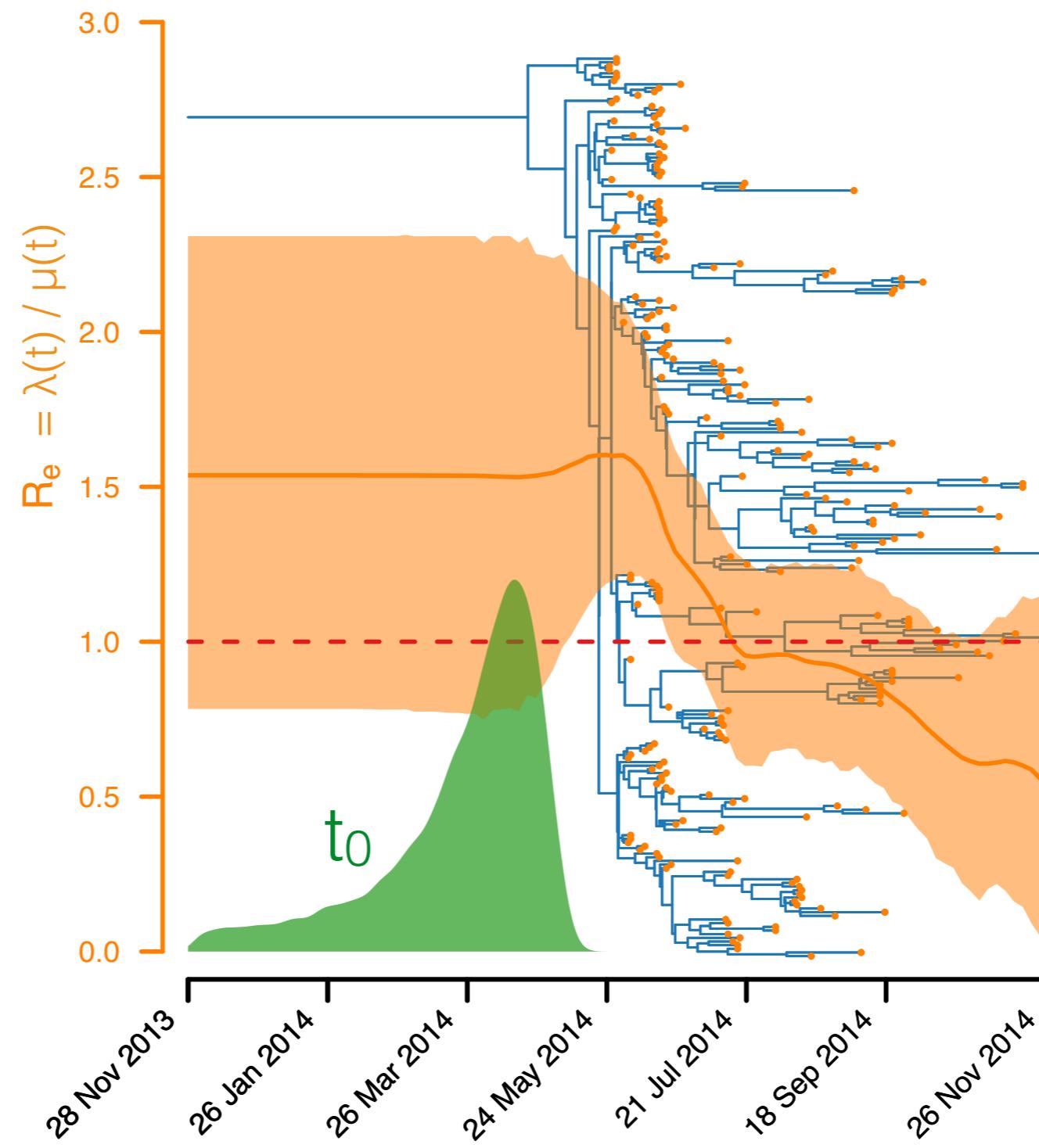
(Based on **246** genomes sampled between **25 May** and **26 Nov 2014**)





# Results in Southeast Sierra Leone

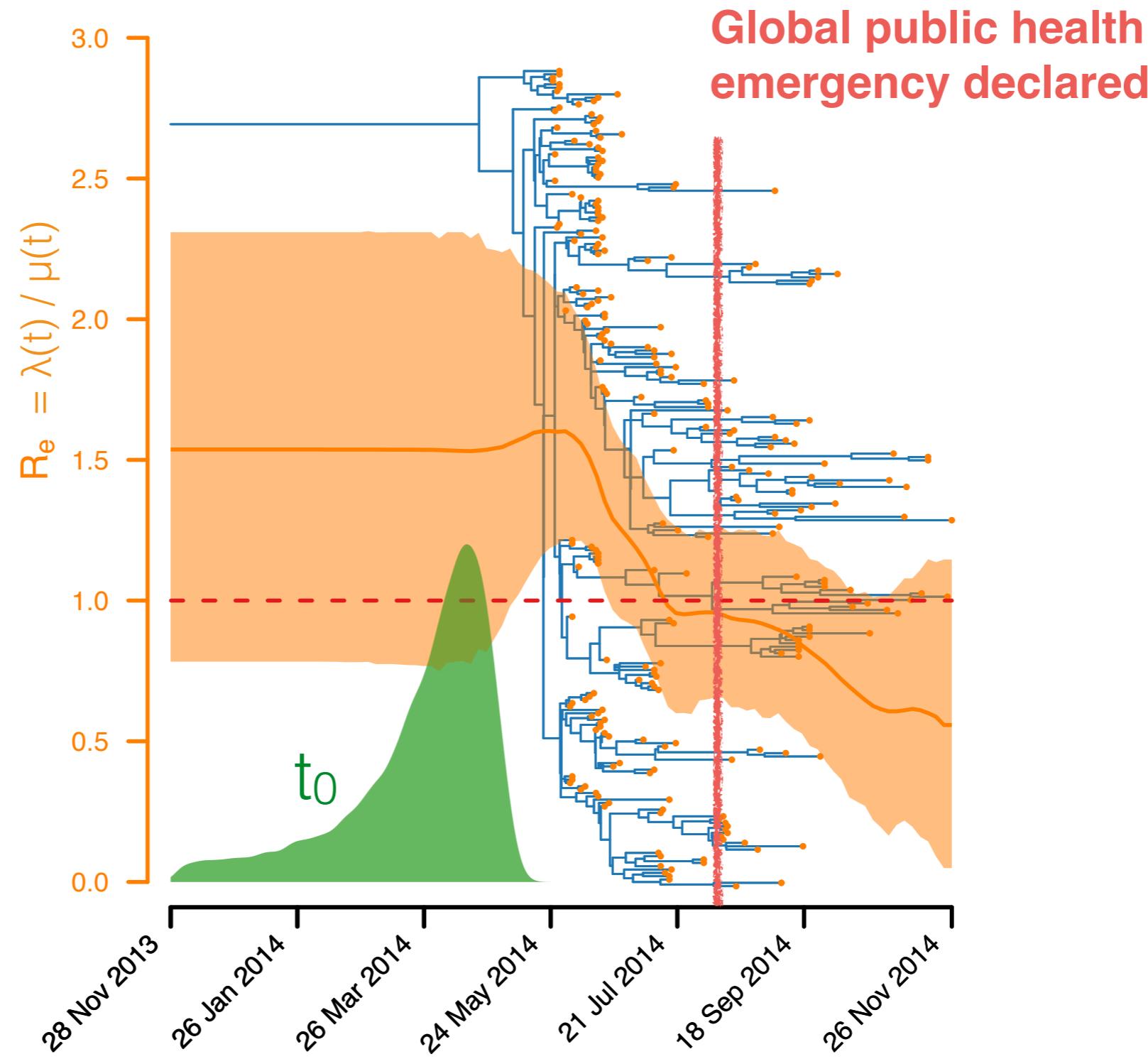
(Based on **246** genomes sampled between **25 May** and **26 Nov 2014**)





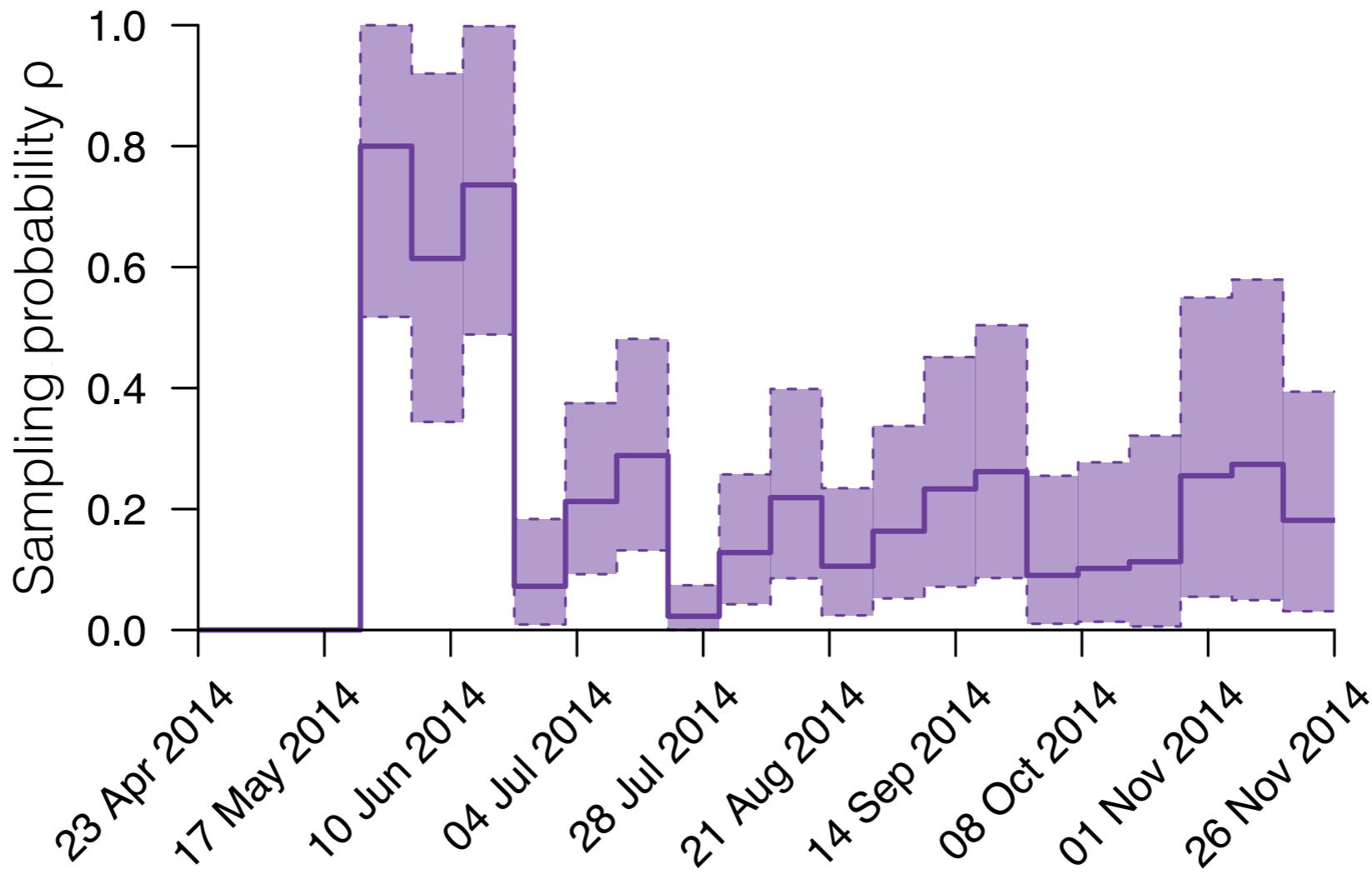
# Results in Southeast Sierra Leone

(Based on **246** genomes sampled between **25 May** and **26 Nov 2014**)

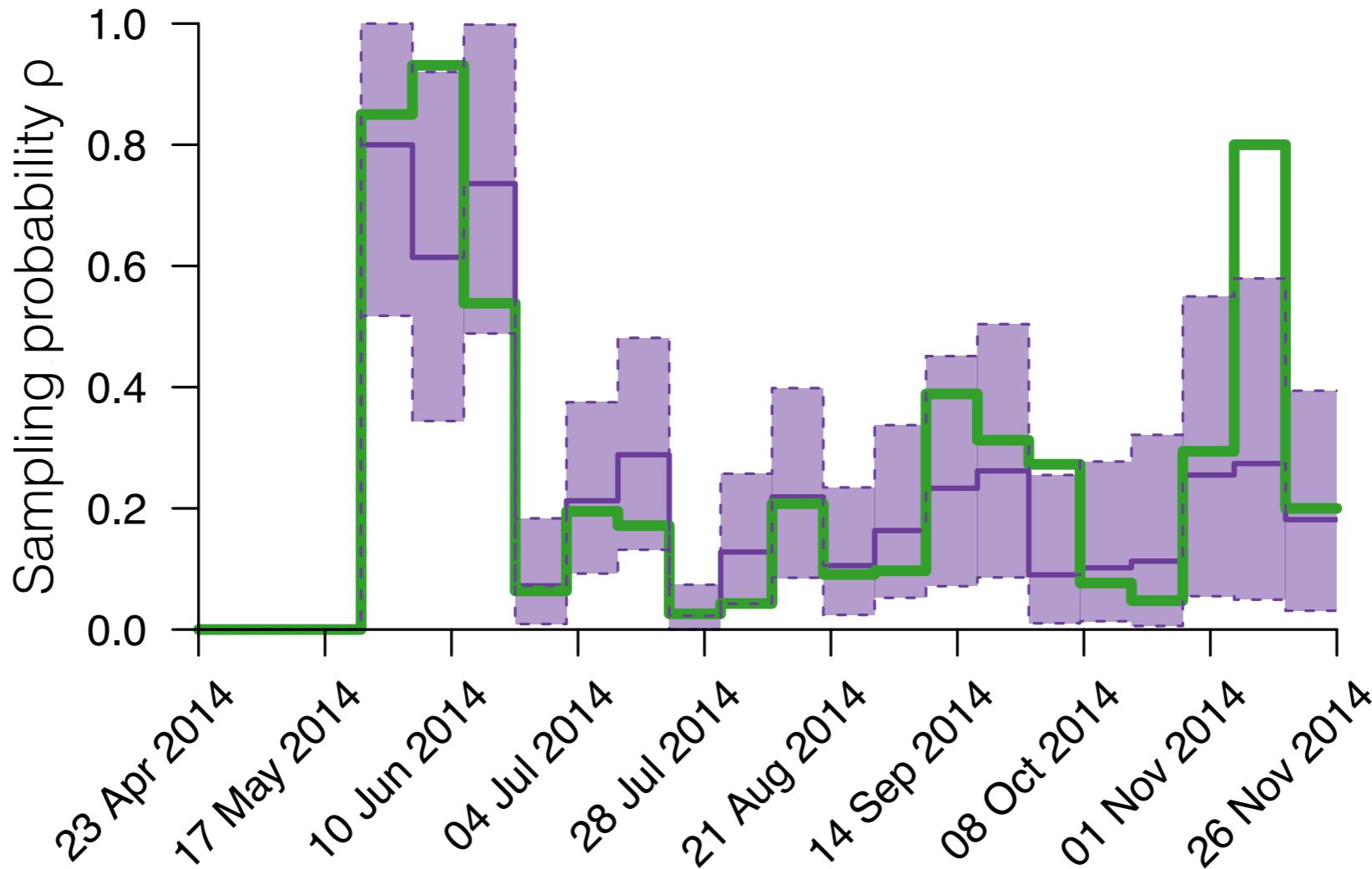


# Sampling probability in Southeast Sierra Leone

---

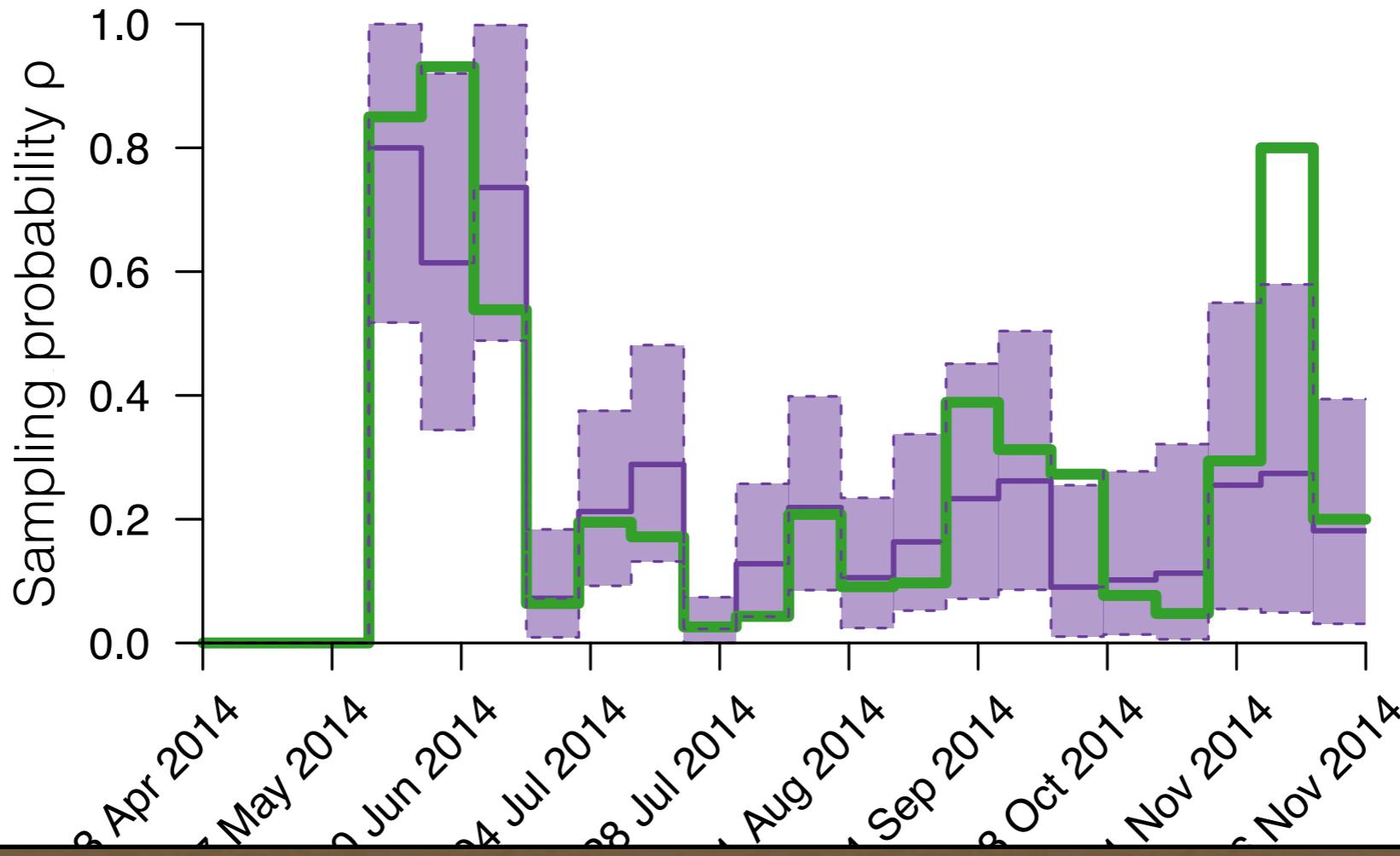


# Sampling probability in Southeast Sierra Leone



— WHO Sampling proportion  
= #sequenced / [Incidence (i.e. new cases) reported by WHO]

# Sampling probability in Southeast Sierra Leone



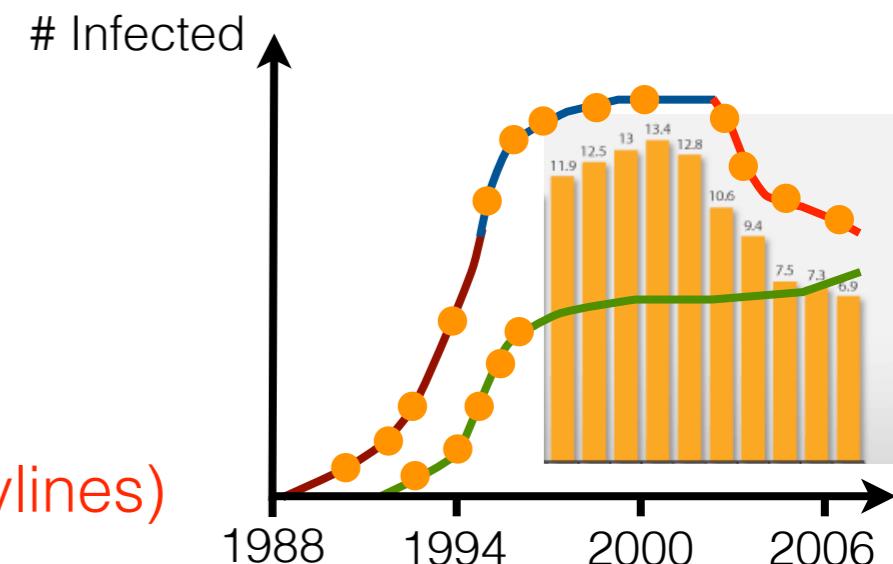
Phylogenetics is alternative to classic epidemiological approaches for estimating

**Incidence, Time of outbreak, Reproductive number**

# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

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



## 2. ...for identifying epidemiological changes over time (skylines)

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

## 3. ...accounting for SIR model dynamics

- ▶ C: Volz et al. (*Genetics*, 2009)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014) - **BDSIR add-on**

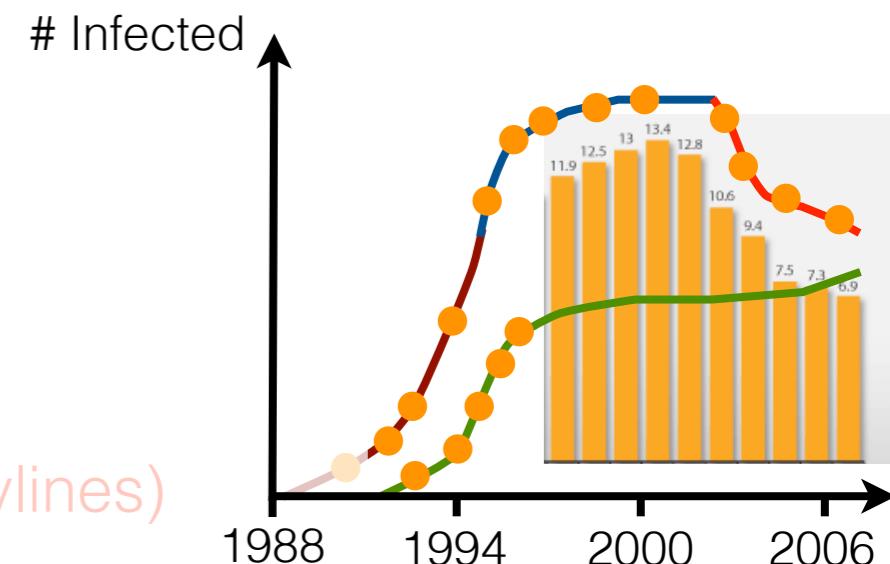
## 4. ...accounting for population structure

- ▶ C: Volz (*Genetics*, 2012), Vaughan et al. (*Bioinf.*, 2014), de Maio et al. (*PLoS Genetics*, 2015), Mueller et al. (*MBE*, 2017) - **Mascot add-on**
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013), Kühnert et al. (*MBE*, 2016) - **bdmm add-on**

# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

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



## 2. ...for identifying epidemiological changes over time (skylines)

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

## 3. ...accounting for SIR model dynamics

- ▶ C: Volz et al. (*Genetics*, 2009)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014) - **BDSIR add-on**

## 4. ...accounting for population structure

- ▶ C: Volz (*Genetics*, 2012), Vaughan et al. (*Bioinf.*, 2014), de Maio et al. (*PLoS Genetics*, 2015), Mueller et al. (*MBE*, 2017) - **Mascot add-on**
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013), Kühnert et al. (*MBE*, 2016) - **bdmm add-on**

# Host population structure leaves pattern in the phylogenetic tree

---

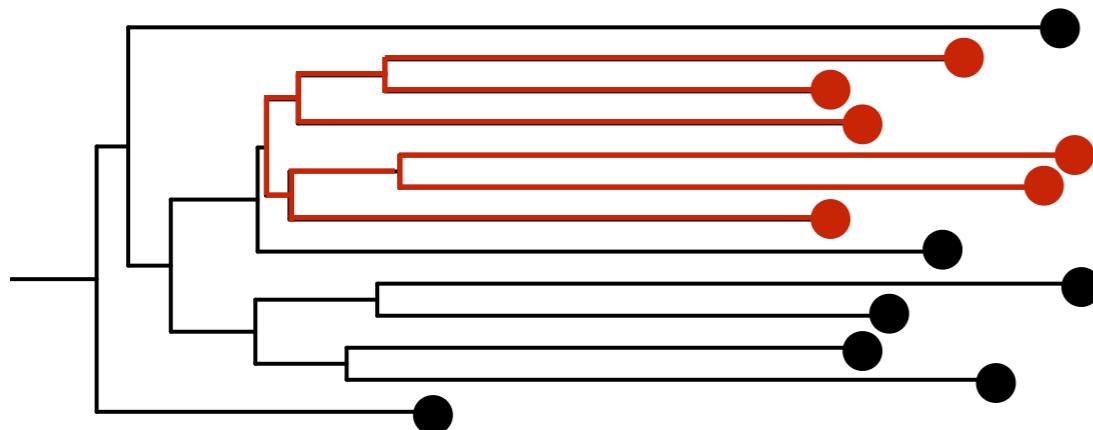
Are new infections explained through ongoing transmission or migration from another area?

-> identifying source demes!

# Host population structure leaves pattern in the phylogenetic tree

Are new infections explained through ongoing transmission or migration from another area?  
-> identifying source demes!

Transmission in red deme:

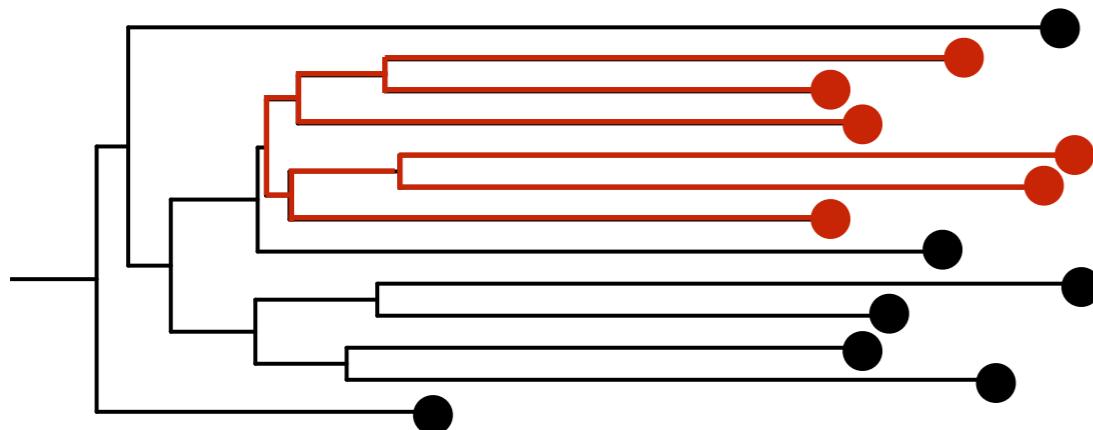


- Red deme
- Black deme

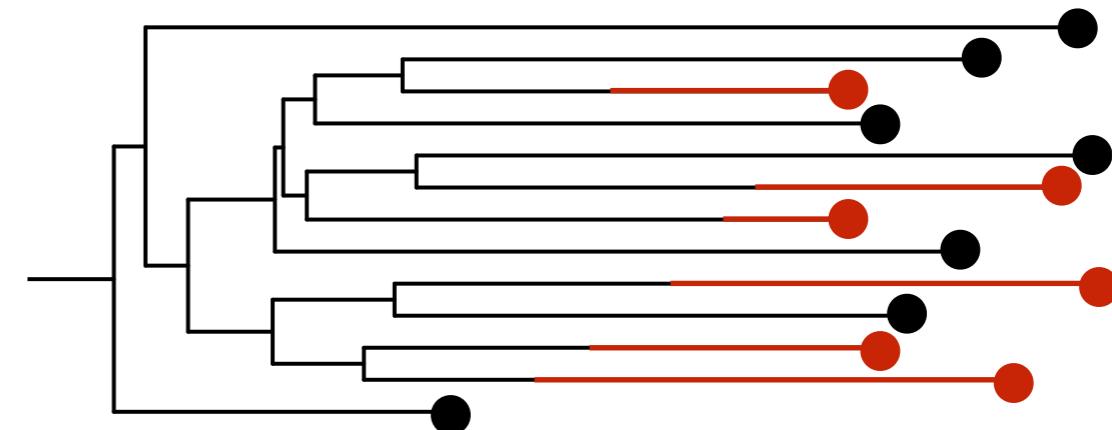
# Host population structure leaves pattern in the phylogenetic tree

Are new infections explained through ongoing transmission or migration from another area?  
-> identifying source demes!

Transmission in red deme:



Migration into red deme:

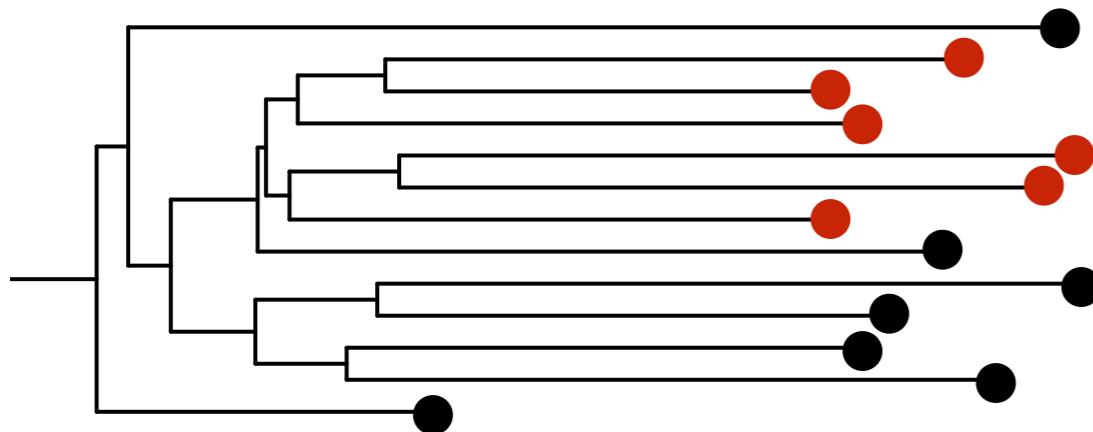


- Red deme
- Black deme

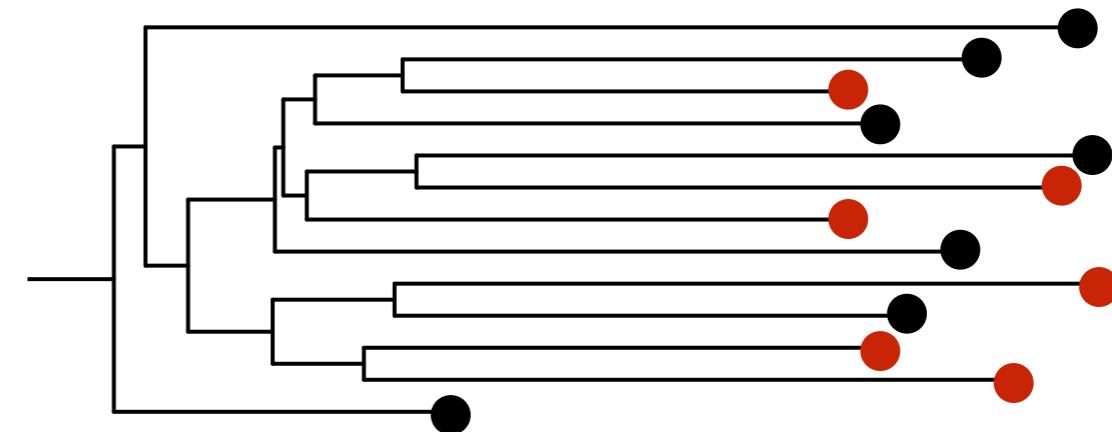
# Host population structure leaves pattern in the phylogenetic tree

Are new infections explained through ongoing transmission or migration from another area?  
-> identifying source demes!

Transmission in red deme:



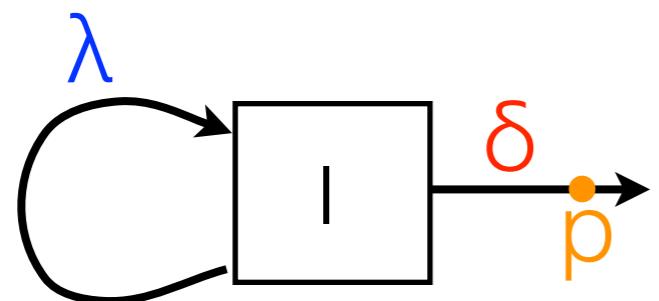
Migration into red deme:



- Red deme
- Black deme

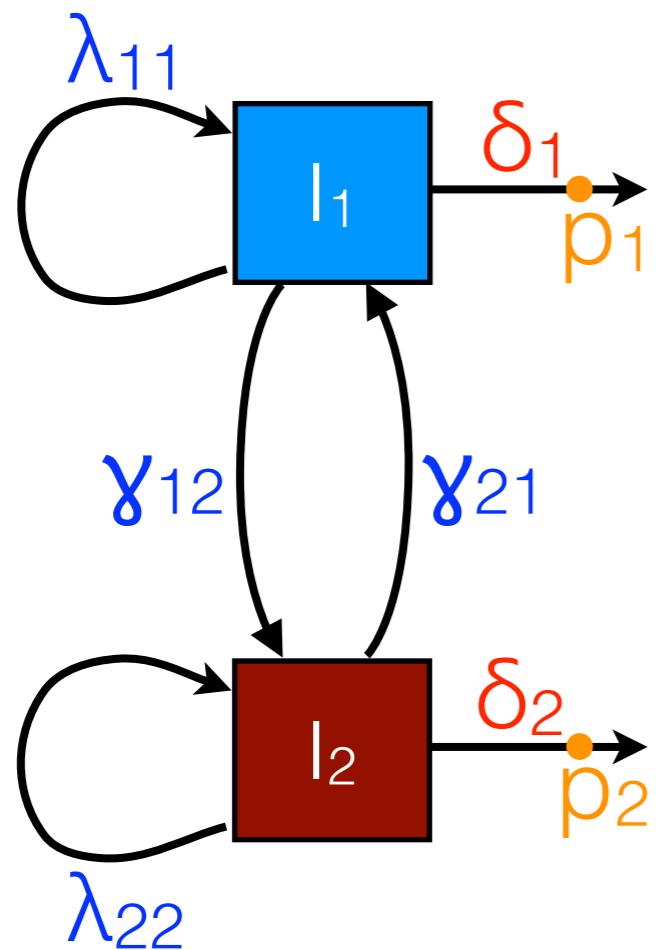
# Population structure

---



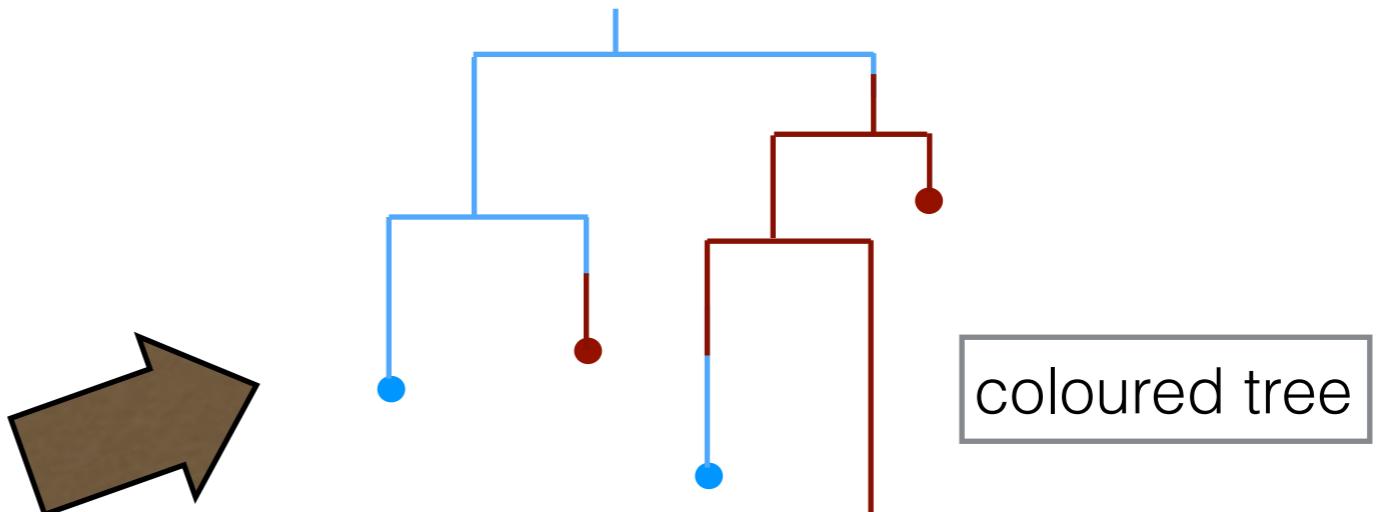
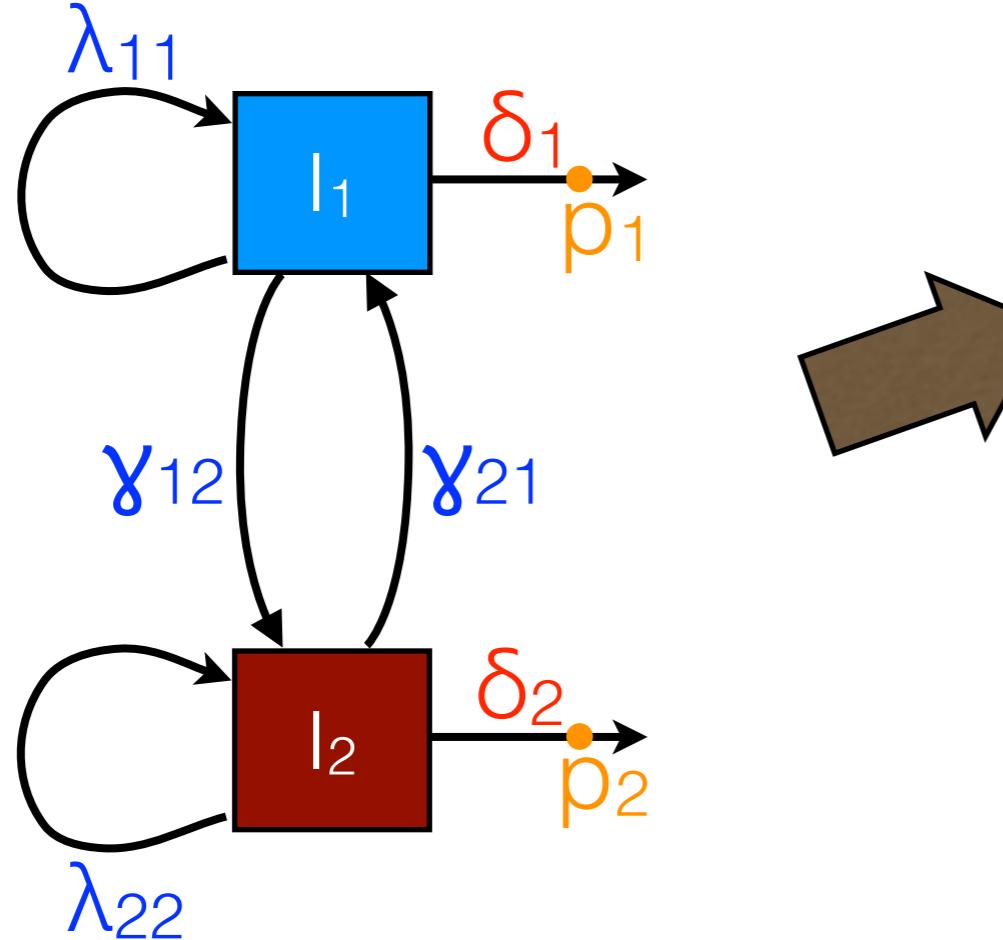
# Population structure

---

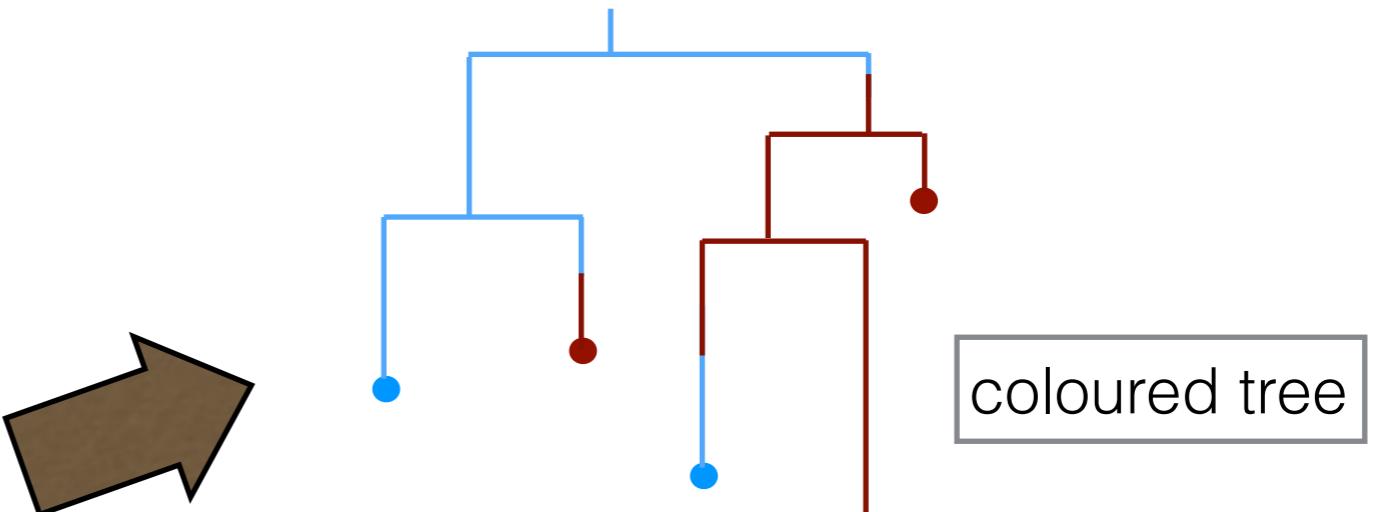
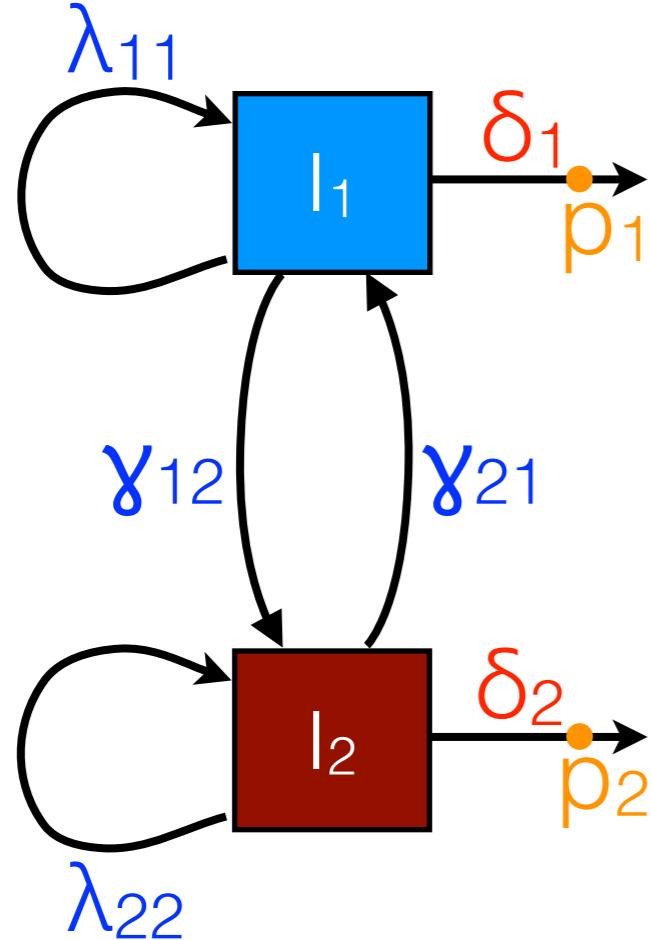


# Population structure

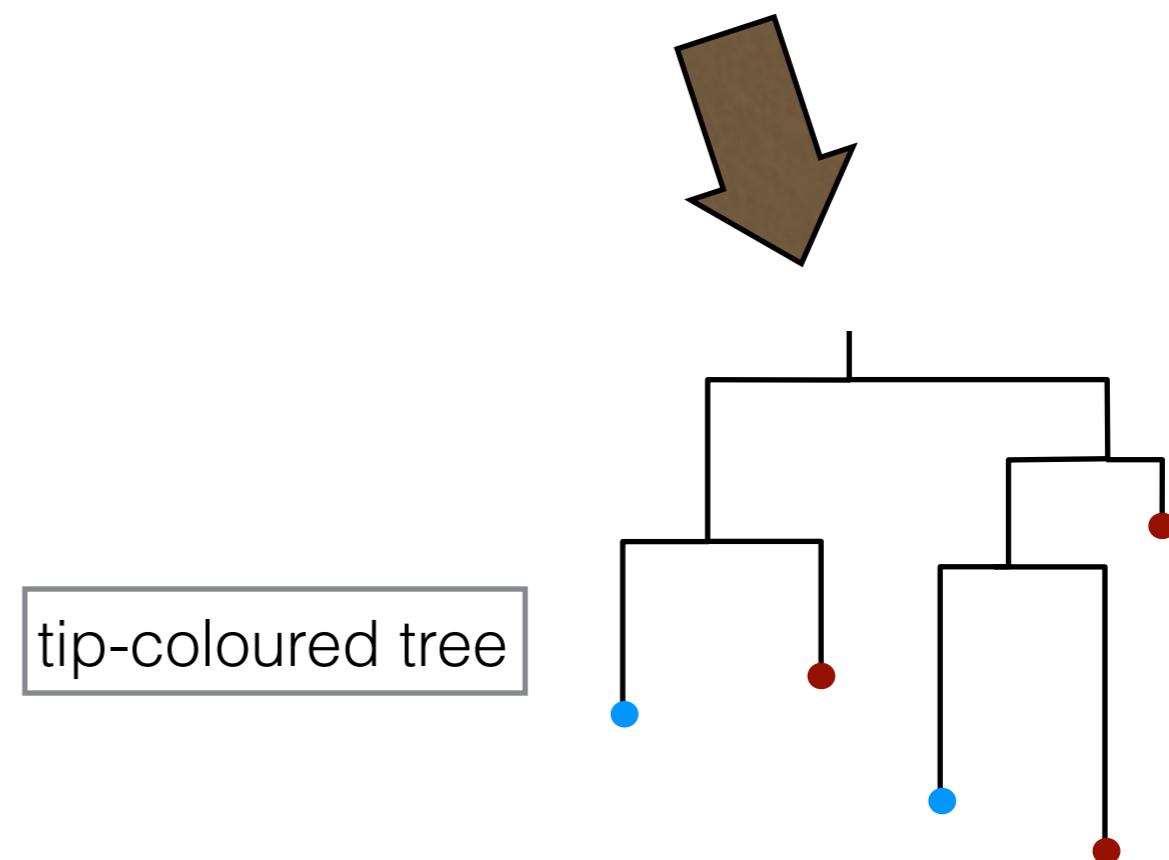
---



# Population structure



tip-coloured tree

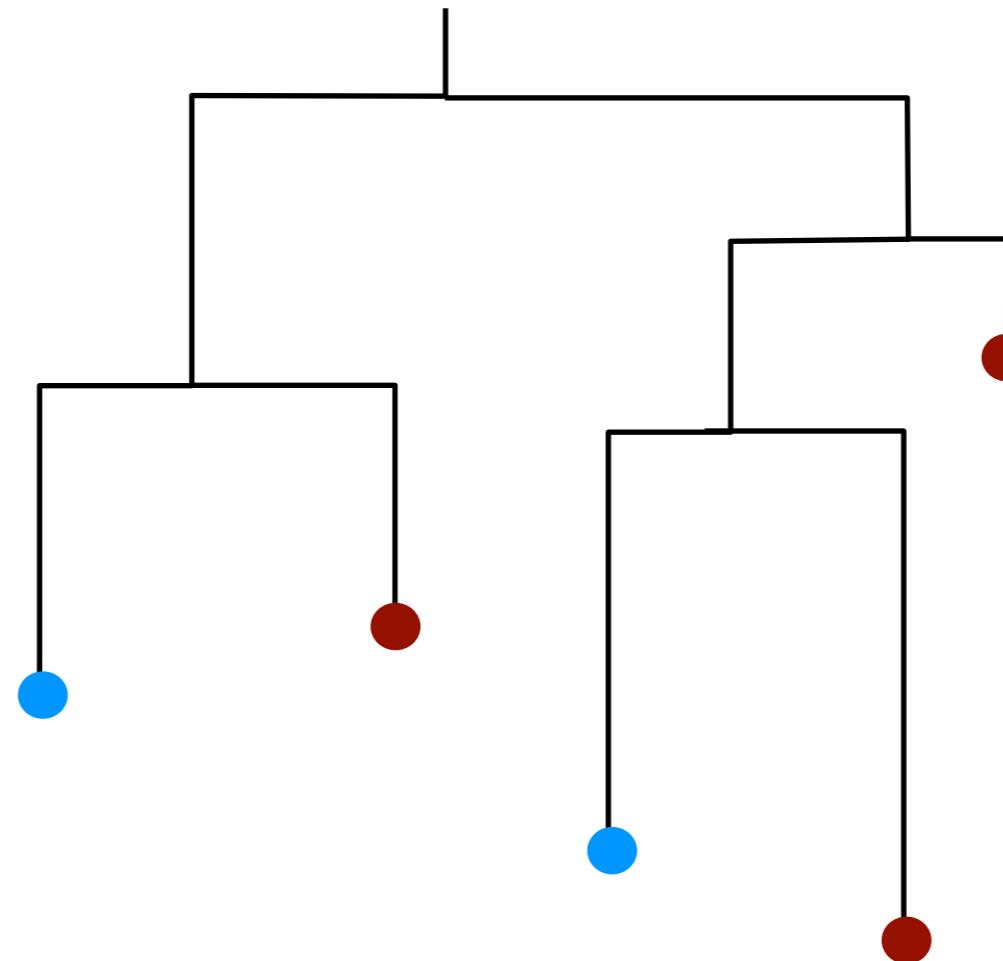


# How can we quantify transmission rates and migration rates?

---

Multi-type birth-death model

Structured coalescent

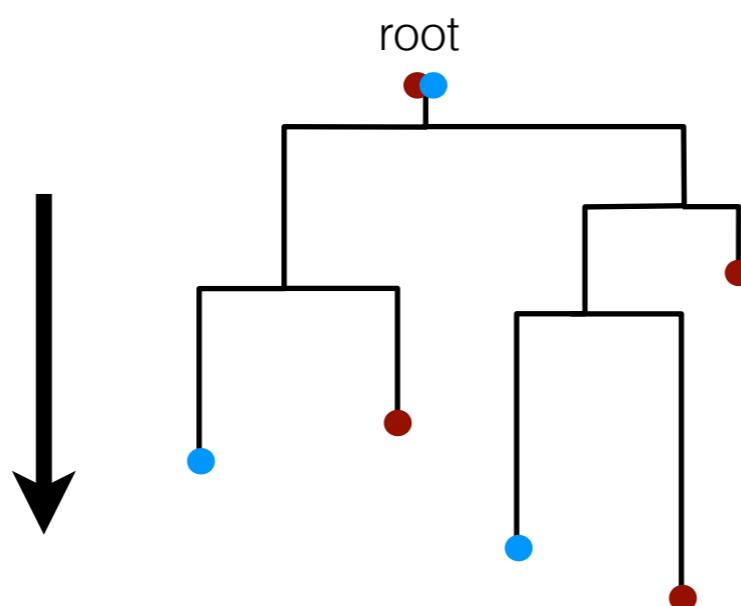


# How can we quantify transmission rates and migration rates?

Multi-type birth-death model

Structured coalescent

- ▶ Process starts with coloured ind.: forward in time!
- ▶ Transmission, migration, recovery & sampling describe the dynamics ( $\eta$ )
- ▶ root (age and state) is parameter
- ▶ tips (age and state) is random
- ▶  $P(\text{tree, tips} \mid \text{root}; \eta)$



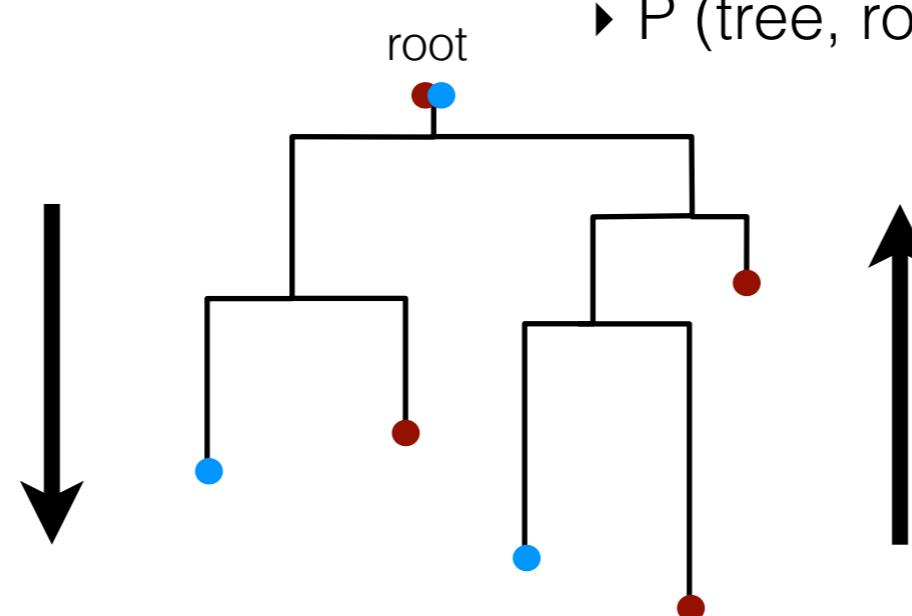
# How can we quantify transmission rates and migration rates?

## Multi-type birth-death model

- ▶ Process starts with coloured ind.: forward in time!
- ▶ Transmission, migration, recovery & sampling describe the dynamics ( $\eta$ )
- ▶ root (age and state) is parameter
- ▶ tips (age and state) is random
- ▶  $P(\text{tree}, \text{tips} | \text{root}; \eta)$

## Structured coalescent

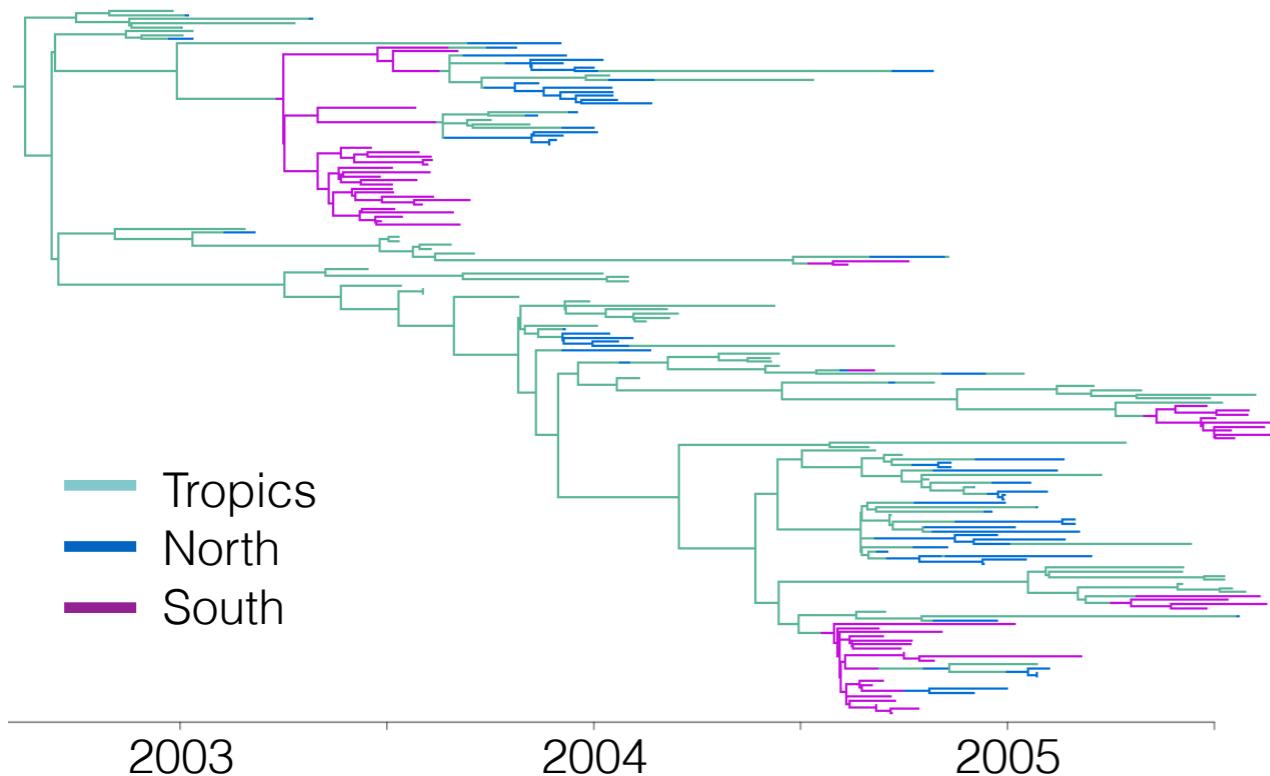
- ▶ Process starts with  $n$  coloured tips: backward in time!
- ▶ Population sizes and migrations describe the dynamics ( $\eta$ )
- ▶ root (age and state) is random
- ▶ tips (age and state) is parameter
- ▶  $P(\text{tree}, \text{root} | \text{tips}; \eta)$



# Host population structure

## Global spread of seasonal influenza

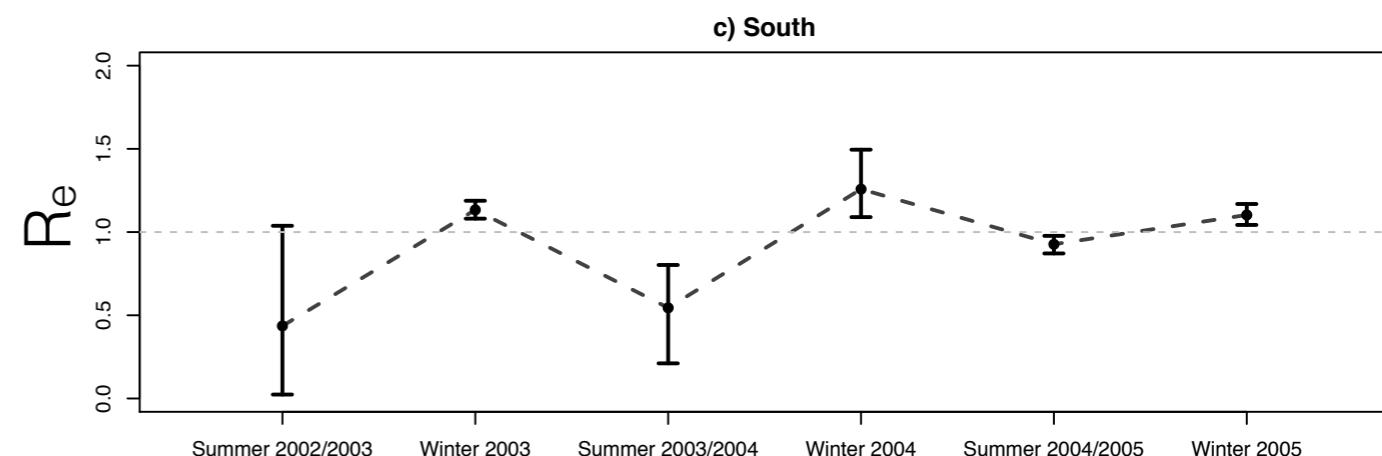
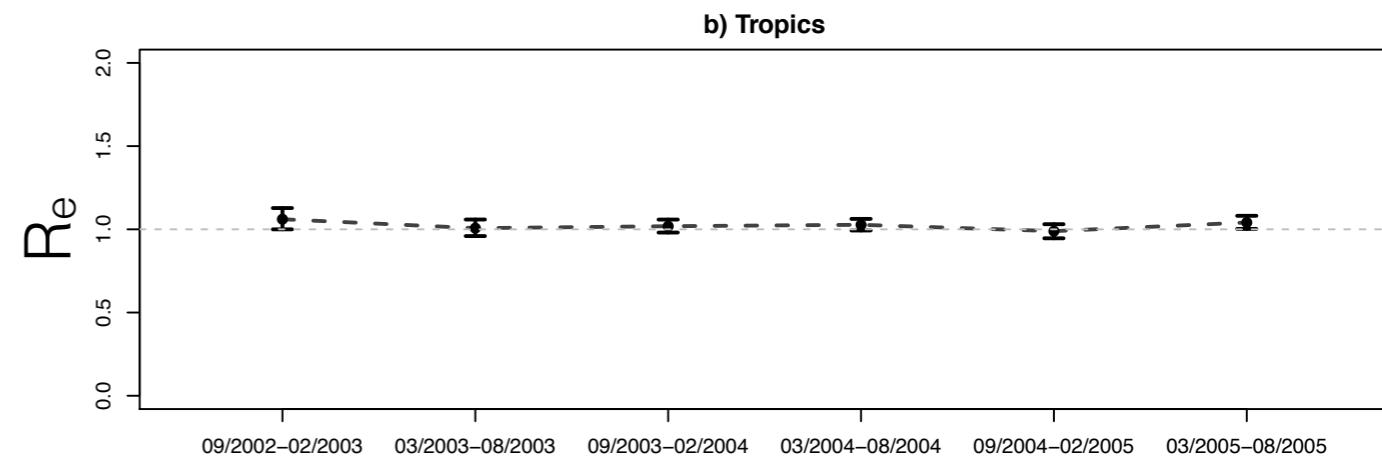
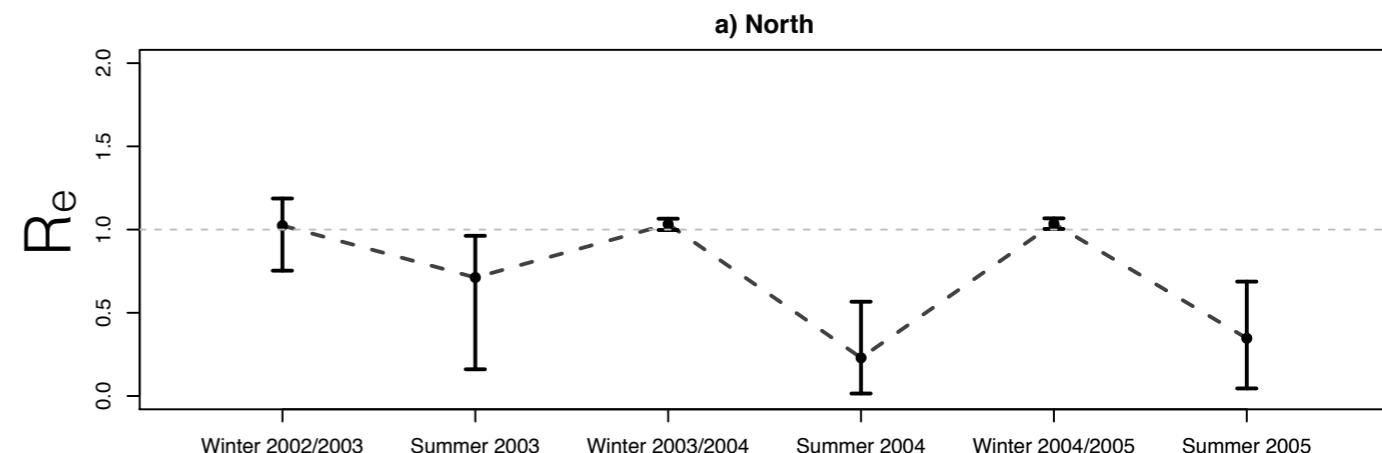
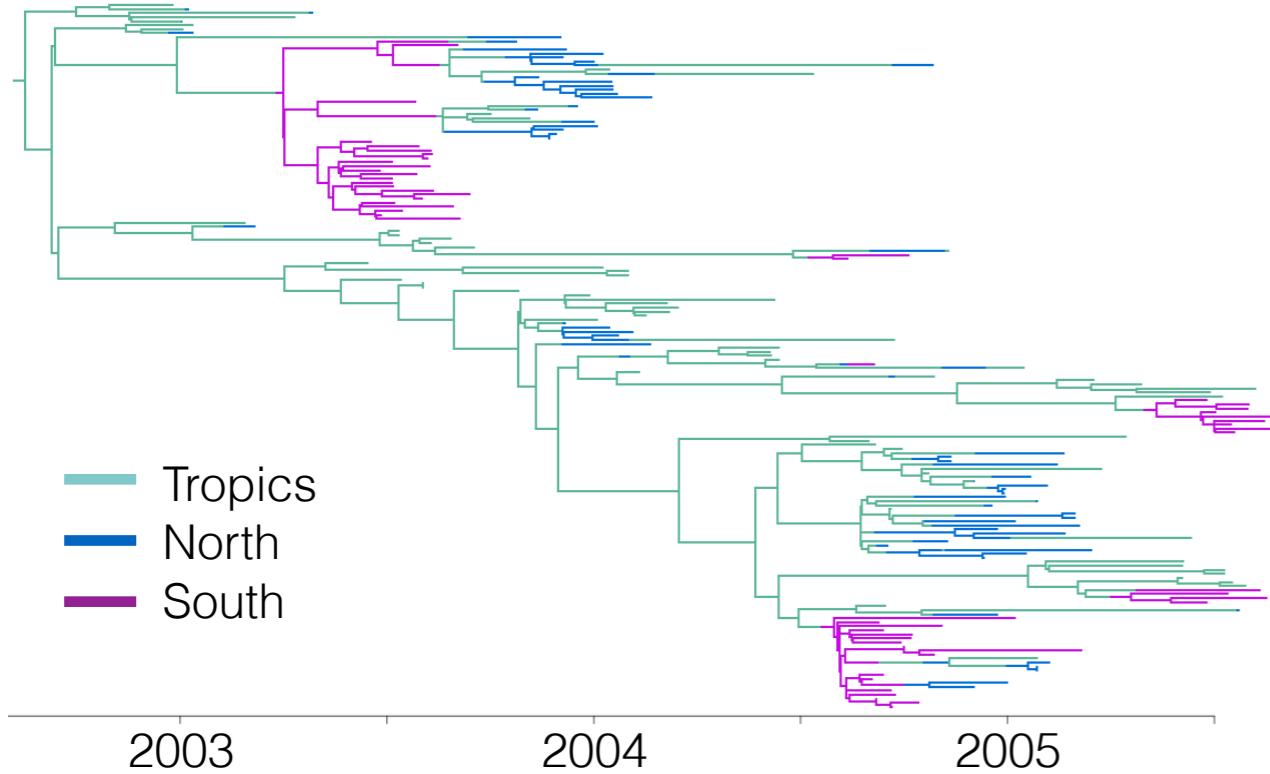
Inference of  
coloured tree



# Host population structure

## Global spread of seasonal influenza

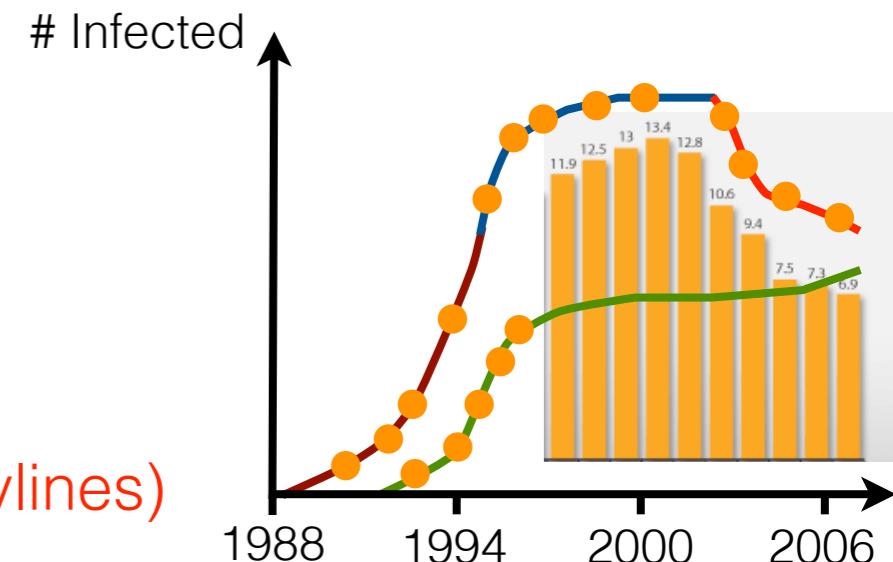
Inference of  
coloured tree



# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

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



## 2. ...for identifying epidemiological changes over time (skylines)

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

## 3. ...accounting for SIR model dynamics

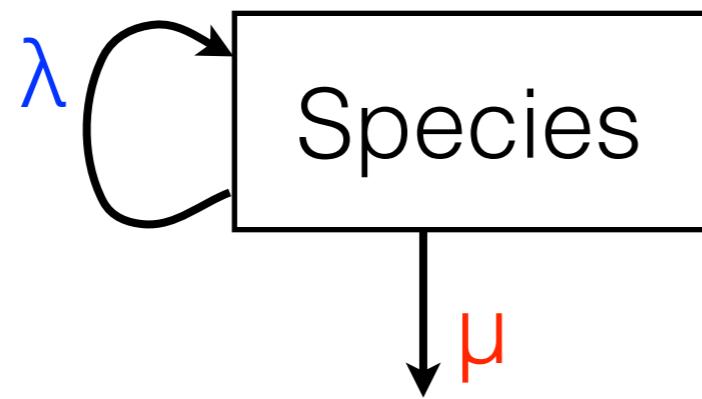
- ▶ C: Volz et al. (*Genetics*, 2009)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014) - **BDSIR add-on**

## 4. ...accounting for population structure

- ▶ C: Volz (*Genetics*, 2012), Vaughan et al. (*Bioinf.*, 2014), de Maio et al. (*PLoS Genetics*, 2015), Mueller et al. (*MBE*, 2017) - **Mascot add-on**
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013), Kühnert et al. (*MBE*, 2016) - **bdmm add-on**

# In afternoon: The birth-death model as a model for speciation and extinction

---

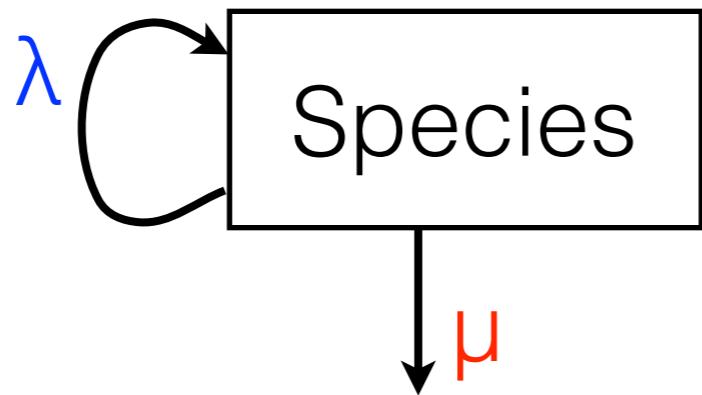


- stem age of a clade  $t$
- speciation rate  $\lambda$
- extinction rate  $\mu$
- sampling probability  $p$

## Rates may depend on:

- 1) time (environmental-dependence)
- 2) # of species (density-dependence)

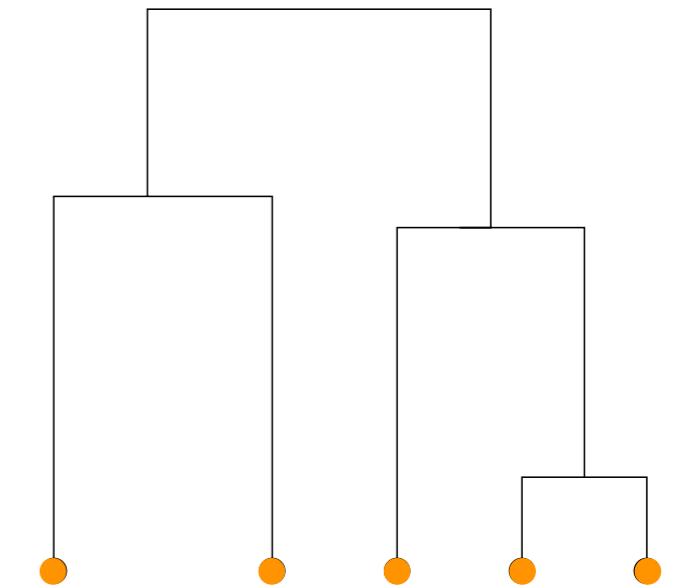
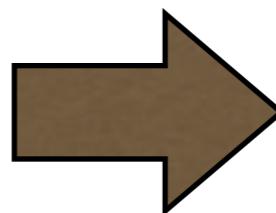
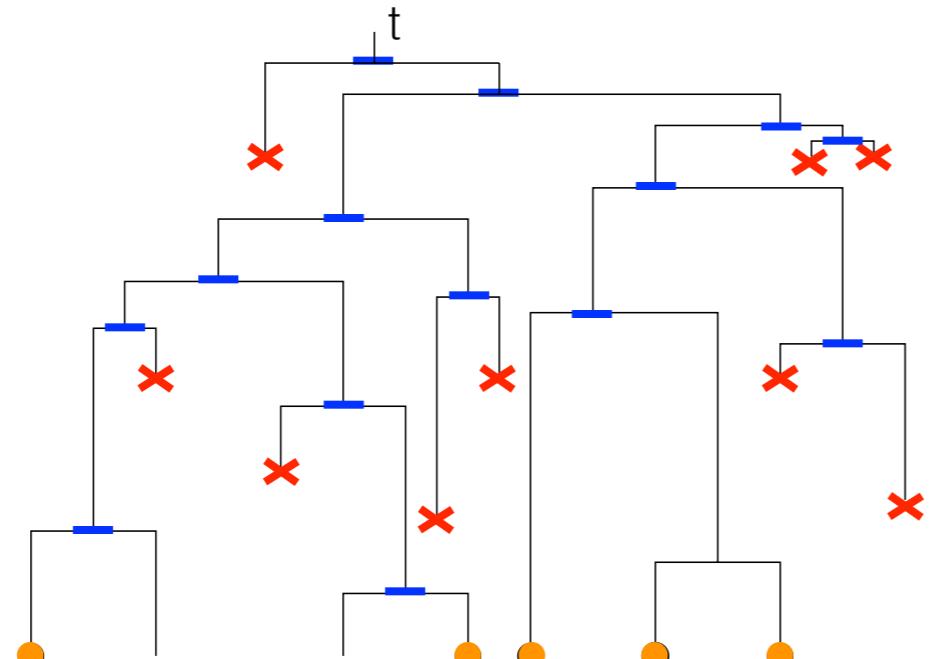
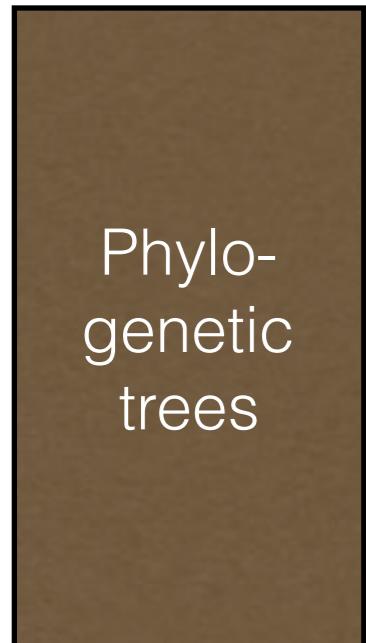
# In afternoon: The birth-death model as a model for speciation and extinction



- stem age of a clade  $t$
- speciation rate  $\lambda$
- extinction rate  $\mu$
- sampling probability  $p$

## Rates may depend on:

- 1) time (environmental-dependence)
- 2) # of species (density-dependence)



# Phylogenetic methods for different population dynamics

---

## 1. ...for analyzing exponential growth

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

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

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

## 3. ...accounting for diversity-dependent dynamics

- ▶ *C: Volz et al. (Genetics, 2009)*
- ▶ *BD: Kühnert et al. (Roy. Soc. Int., 2014) - **BDSIR add-on***

## 4. ...accounting for population structure

- ▶ *C: Volz (Genetics, 2012), Vaughan et al. (Bioinf., 2014), de Maio et al. (PLoS Genetics, 2015), Mueller et al. (MBE, 2017) - **Mascot add-on***
- ▶ *BD: Stadler & Bonhoeffer (Phil. Trans. Roy. Soc. B, 2013), Kühnert et al. (MBE, 2016) - **bdmm add-on***