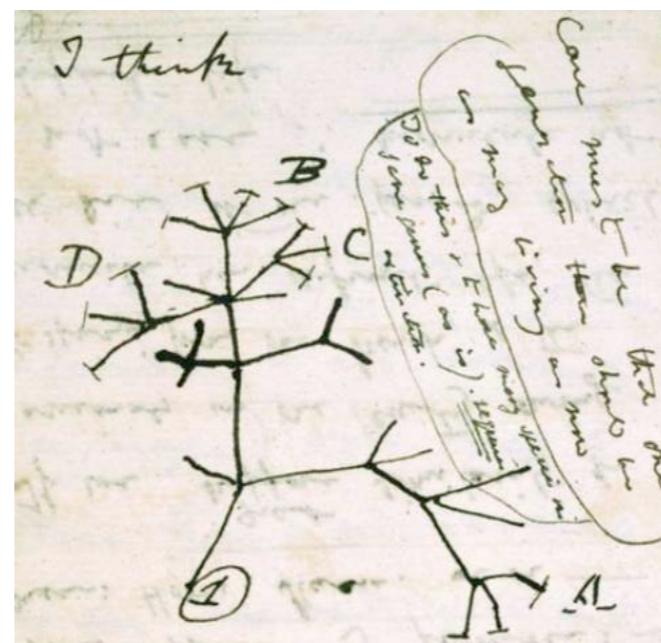


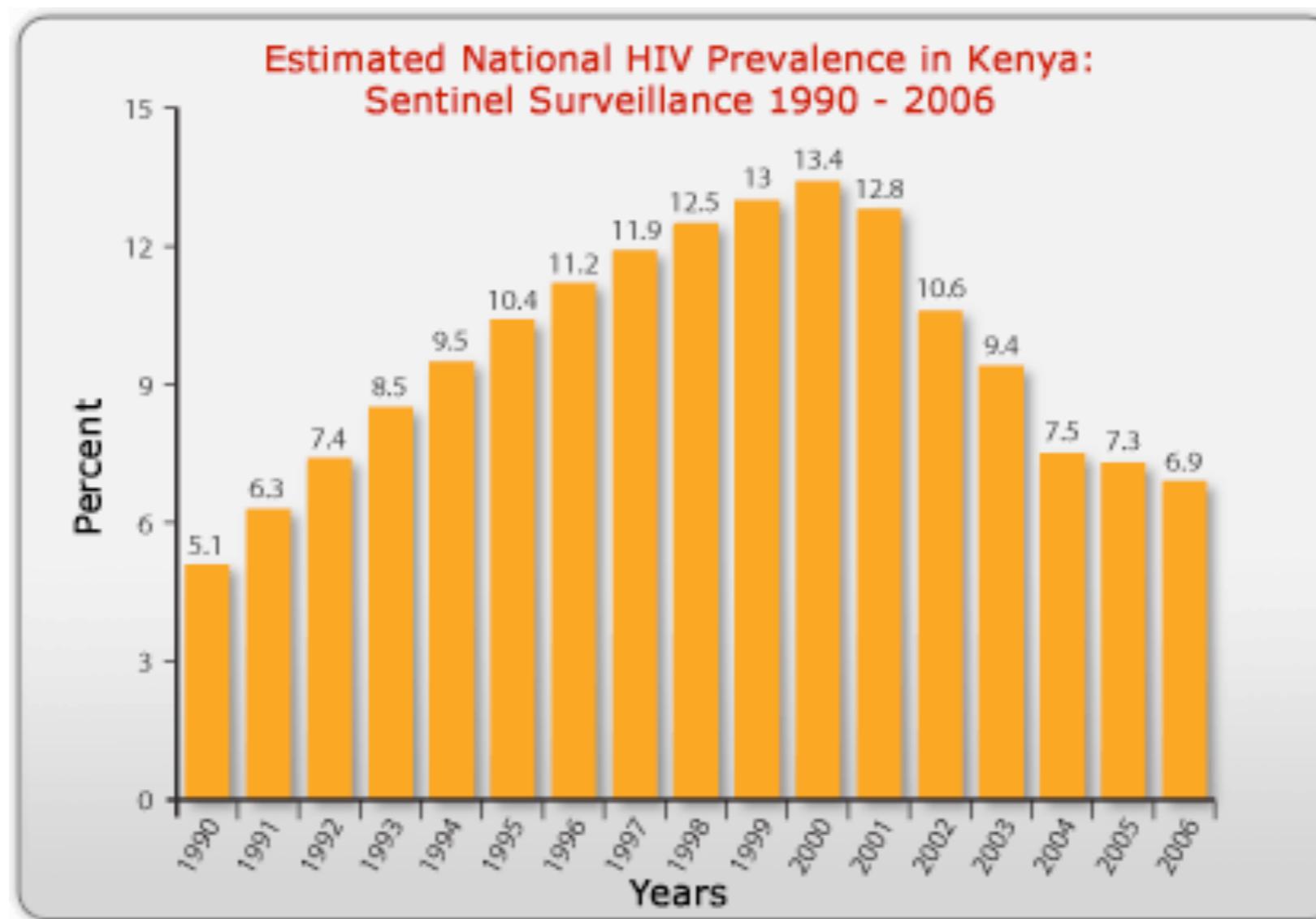
# Phylogenetics

Uncovering  
epidemiological dynamics  
based on genetic sequence data



# Epidemiology

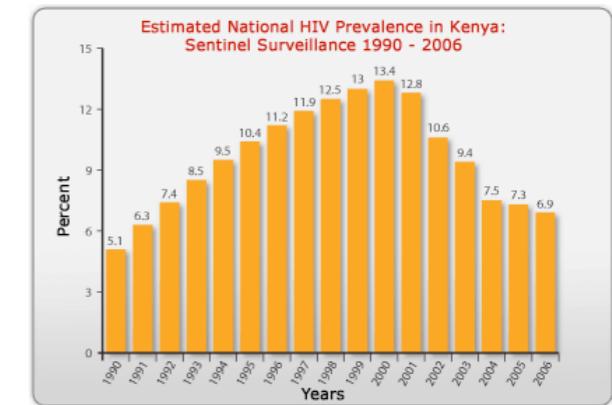
---



# Epidemiology

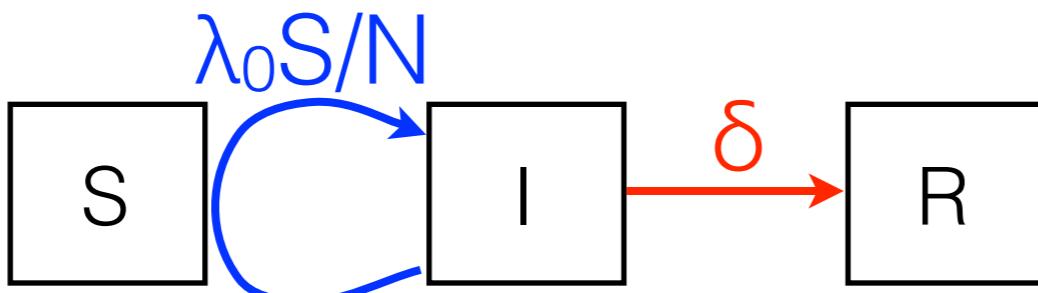
Data

# of infected individuals through time



Models

Population dynamics described by SIR models:

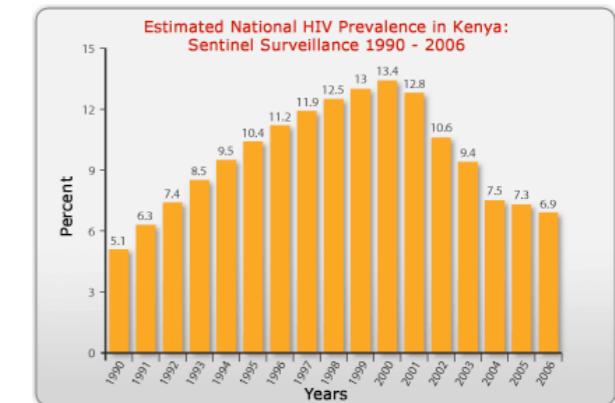


$$\begin{aligned}\frac{dS}{dt} &= -\lambda_0 IS/N \\ \frac{dI}{dt} &= \lambda_0 IS/N - \delta I \\ \frac{dR}{dt} &= \delta I\end{aligned}$$

# Epidemiology

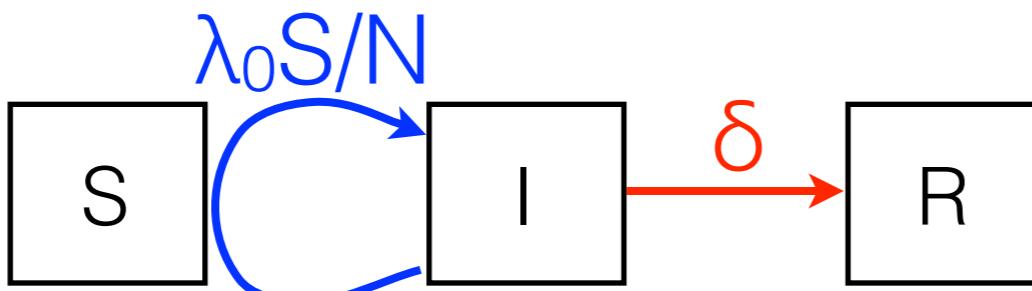
Data

# of infected individuals through time



Models

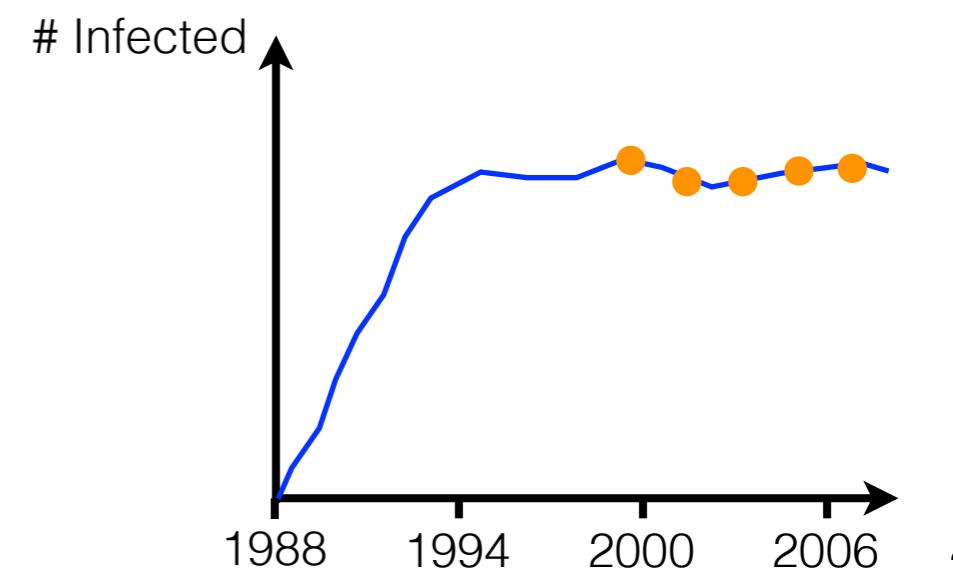
Population dynamics described by SIR models:



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

Limitations

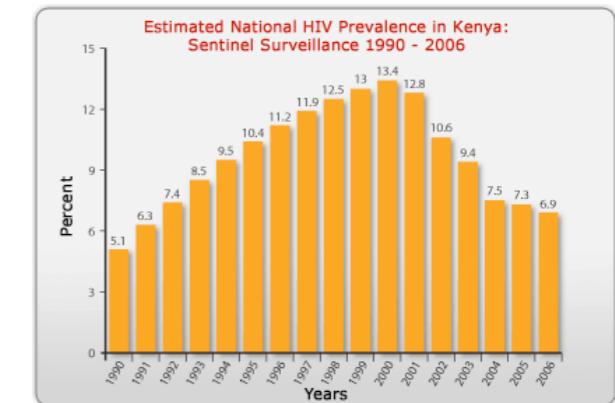
If sampling in early epidemic was missed:



# Epidemiology

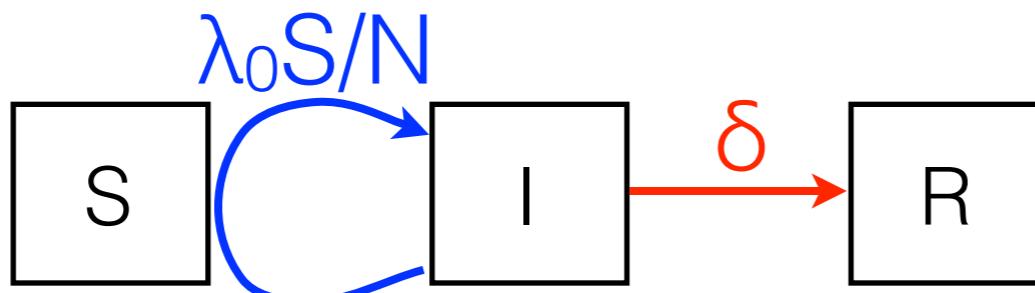
Data

# of infected individuals through time



Models

Population dynamics described by SIR models:

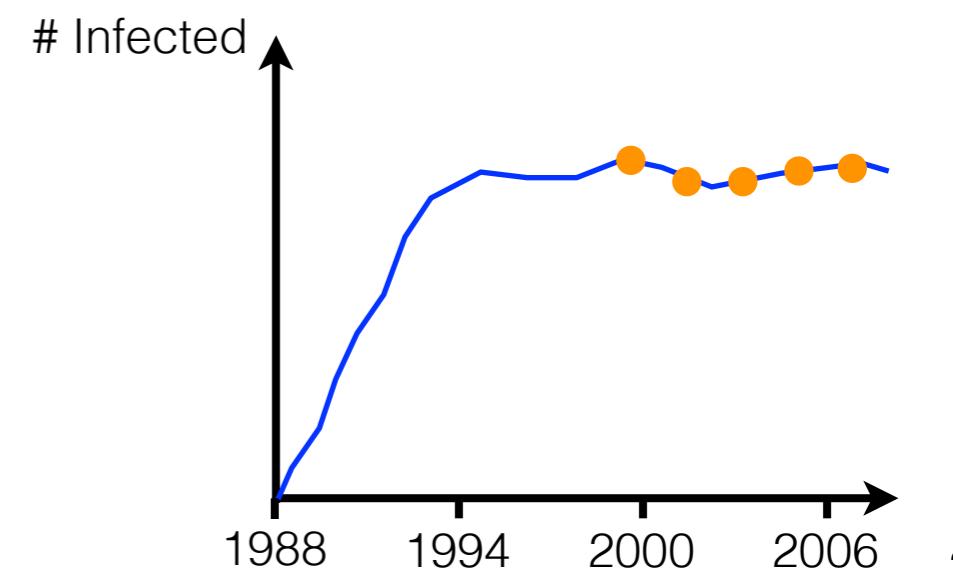


$$\begin{aligned}\frac{dS}{dt} &= -\lambda_0 IS/N \\ \frac{dI}{dt} &= \lambda_0 IS/N - \delta I \\ \frac{dR}{dt} &= \delta I\end{aligned}$$

Limitations

If sampling in early epidemic was missed:

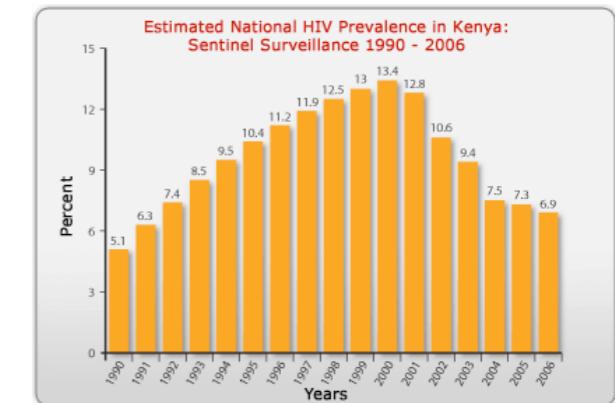
- ▶ **Time of epidemic outbreak?**
- ▶ **Basic reproductive number  $R_0$ ?**



# Epidemiology

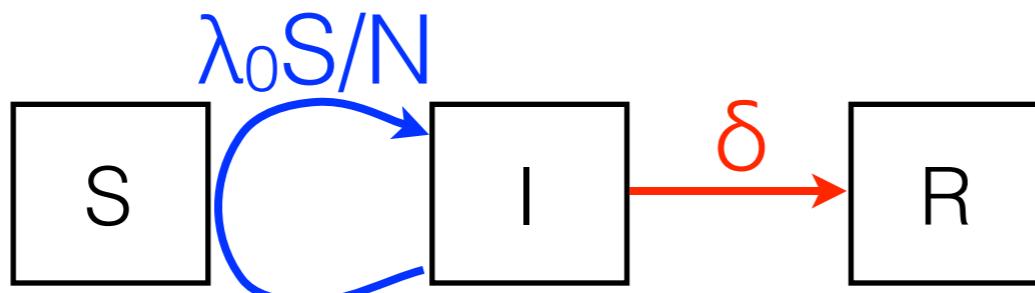
Data

# of infected individuals through time



Models

Population dynamics described by SIR models:



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

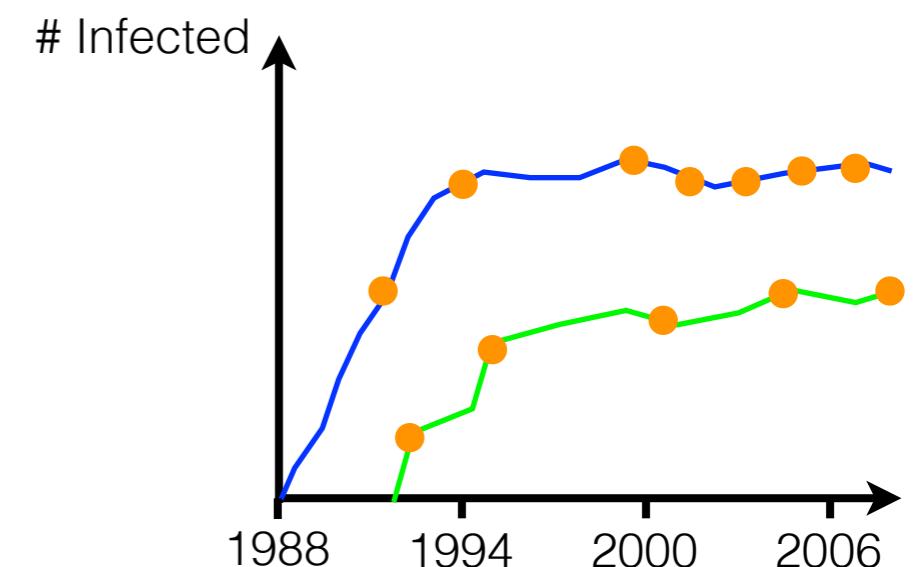
Limitations

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

---

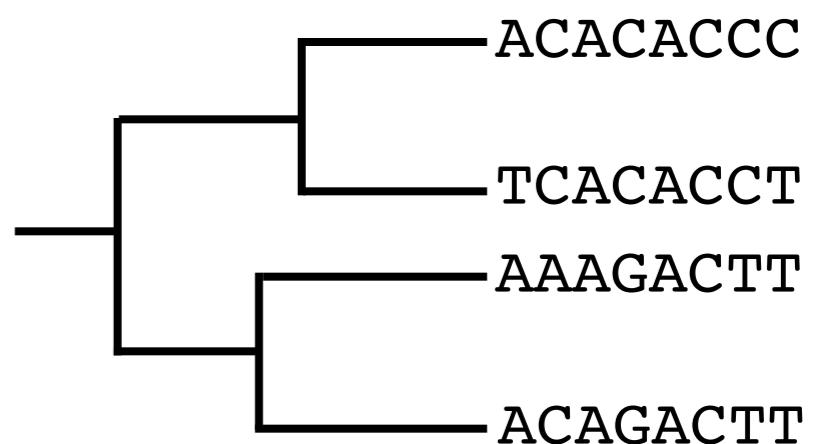
Phylogenetics

Input:

Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)



# Genetic data entering epidemiology

---

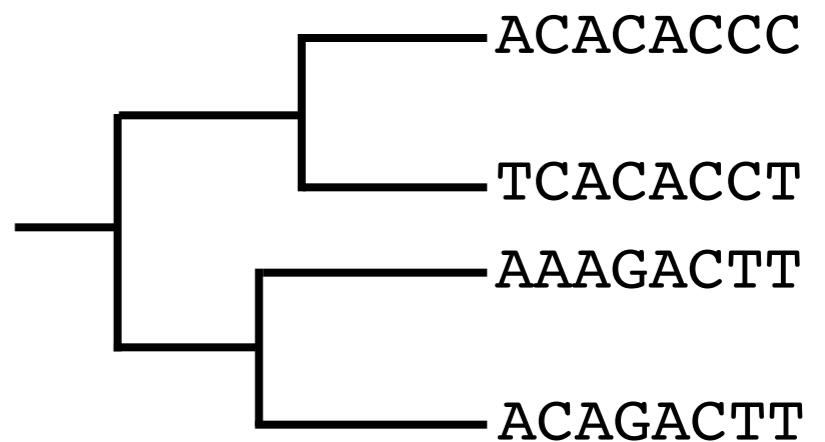
Phylogenetics

Input:

Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)

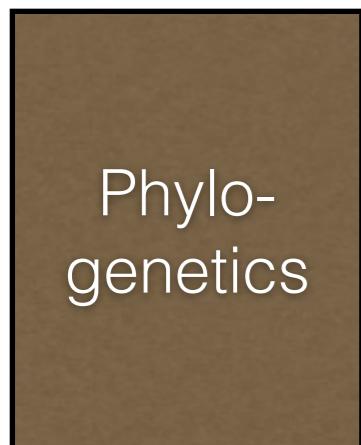


Example  
key result

Emergence of HIV

Epidemic identified in 1980s

# Genetic data entering epidemiology

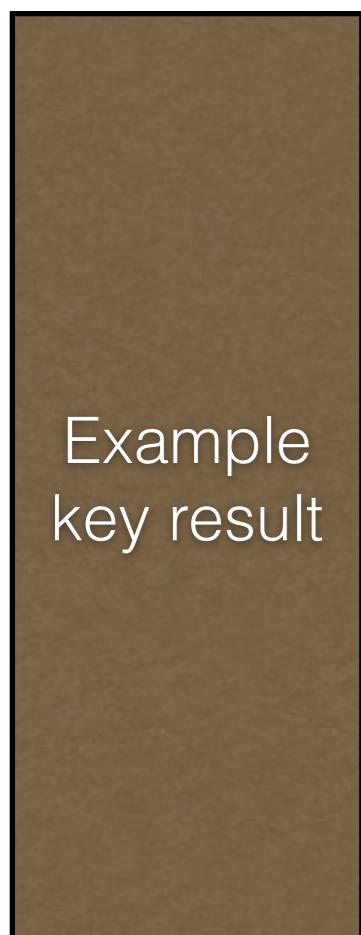
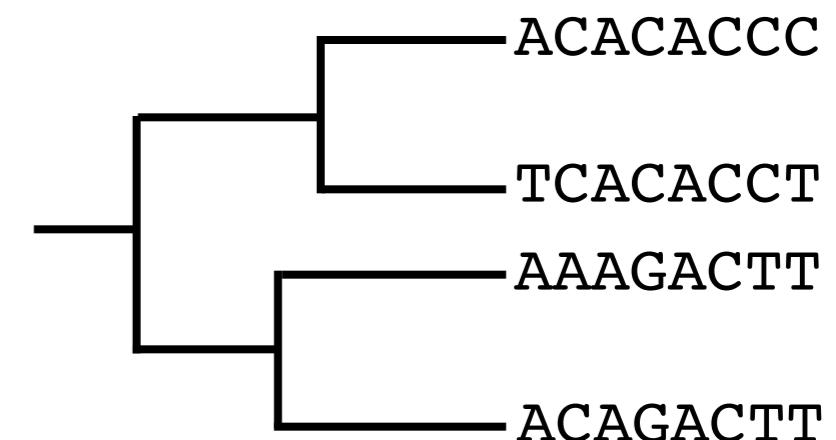


Input:

Pathogen genetic data from different hosts

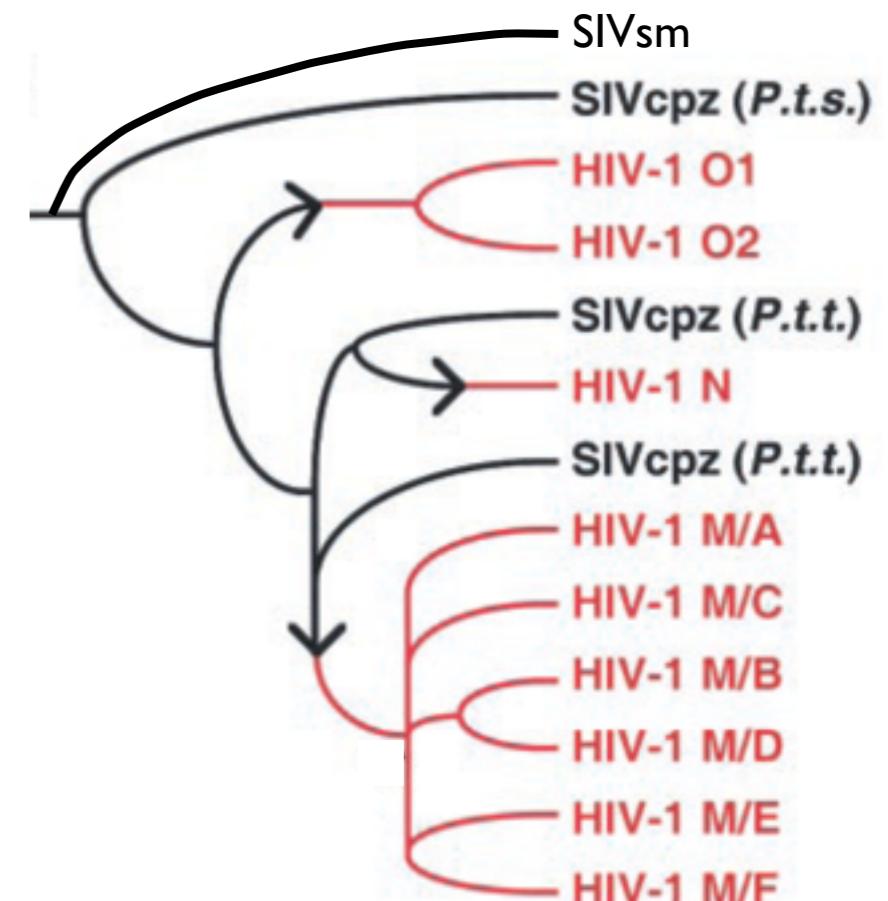
Output:

Transmission chain (who infected whom)



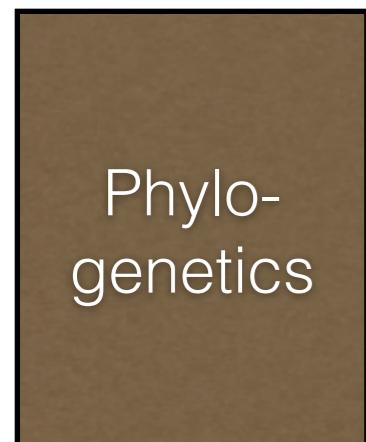
Emergence of HIV

Epidemic identified in 1980s



adopted from Hahn et al. (Science, 2000)

# Genetic data entering epidemiology

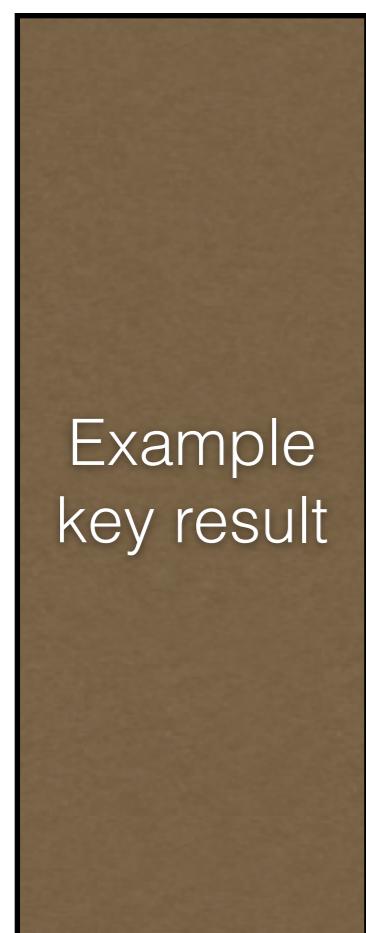
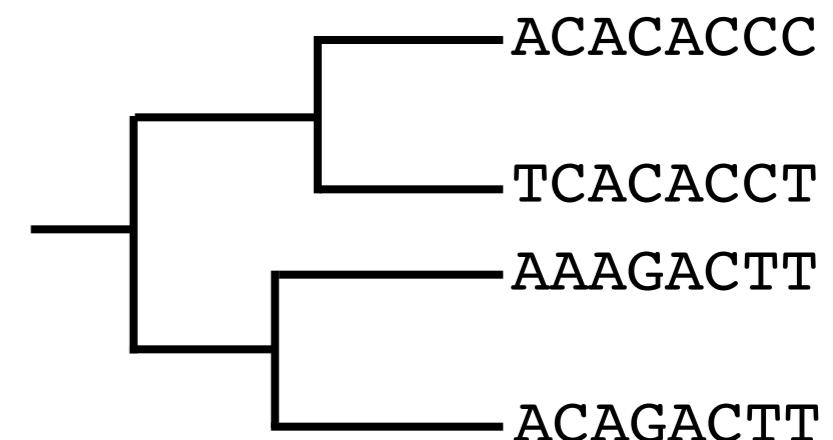


Input:

Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)

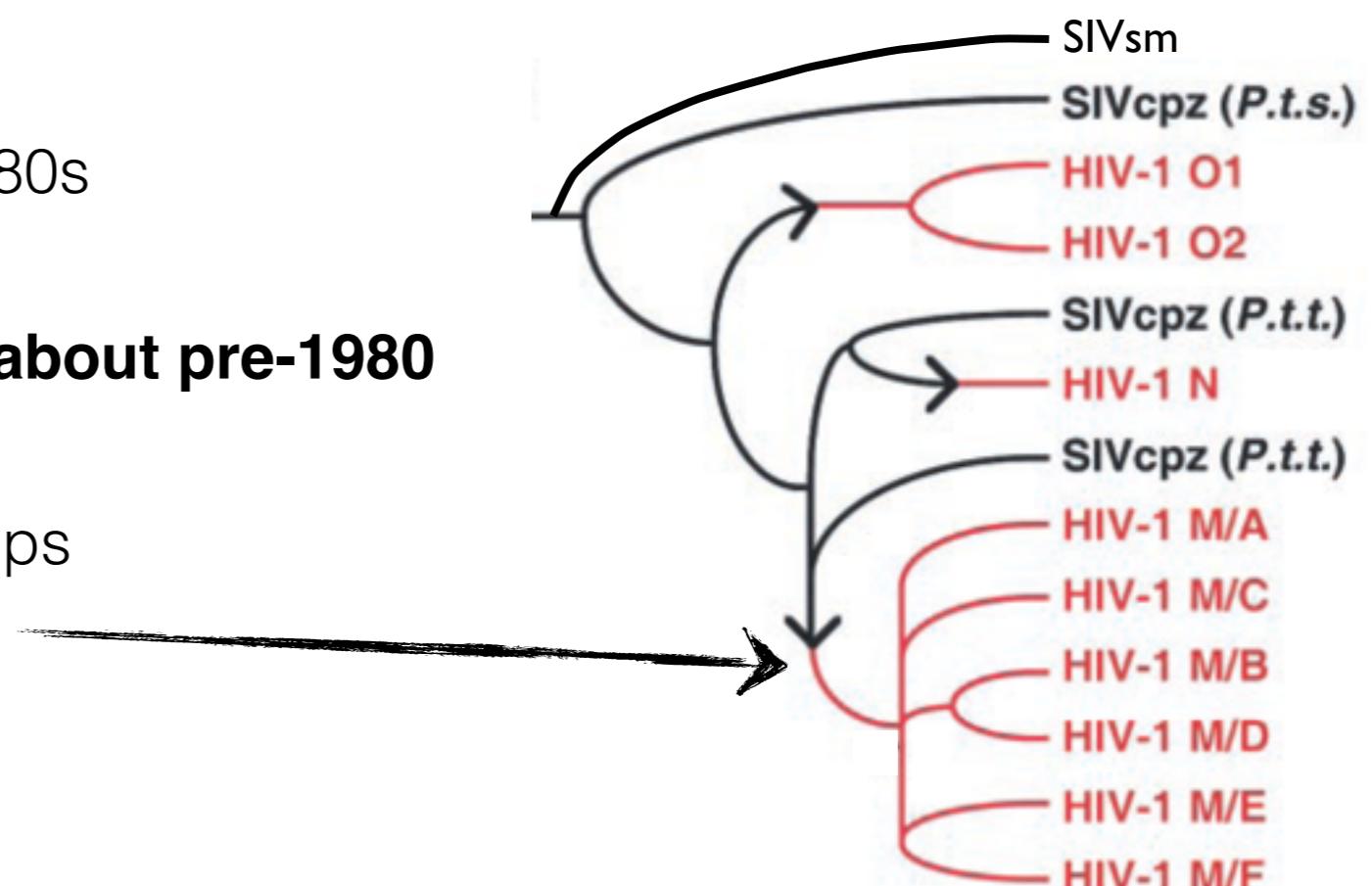


Emergence of HIV

Epidemic identified in 1980s

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

- ▶ Zoonosis from chimps



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

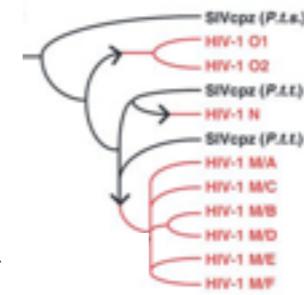
# Genetic data contains information about the epidemiological dynamics

---

Phylogenetics

State of process

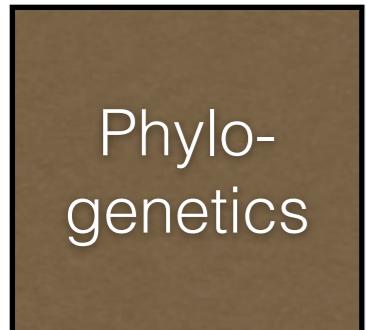
- ▶ Time and geographic location of outbreak



Emergence of a pathogen (here HIV)

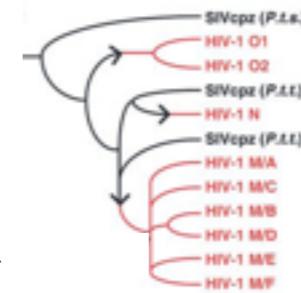
Hahn et al. (Science, 2000)

# Genetic data contains information about the epidemiological dynamics



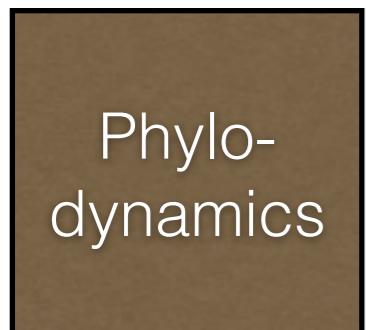
## State of process

- ▶ Time and geographic location of outbreak



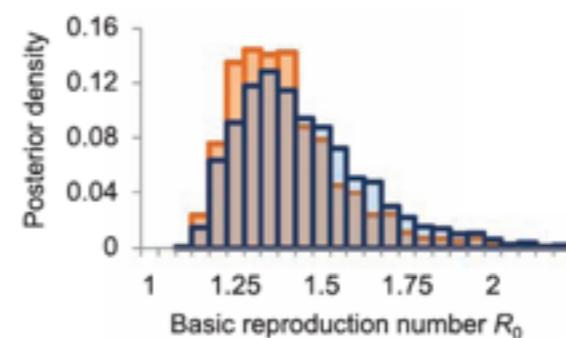
Emergence of a pathogen (here HIV)

Hahn et al. (Science, 2000)



## Dynamics of process

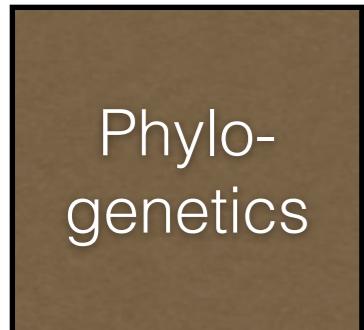
- ▶ Transmission process



Basic reproductive number of an emerging pathogen (here H1N1)

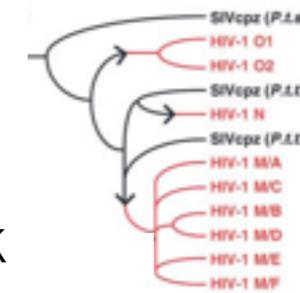
Fraser et al. (Science, 2009)

# Genetic data contains information about the epidemiological dynamics



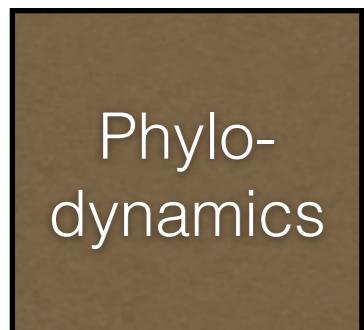
## State of process

- ▶ Time and geographic location of outbreak



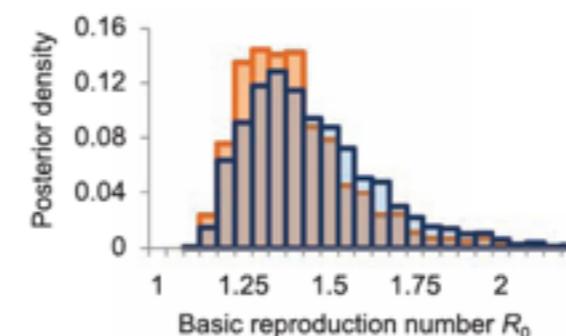
Emergence of a pathogen (here HIV)

Hahn et al. (Science, 2000)



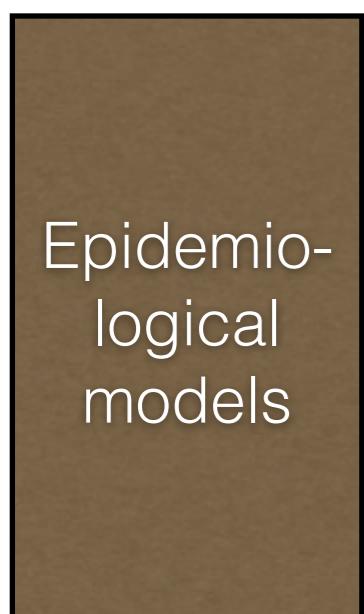
## Dynamics of process

- ▶ Transmission process



Basic reproductive number of an emerging pathogen (here H1N1)

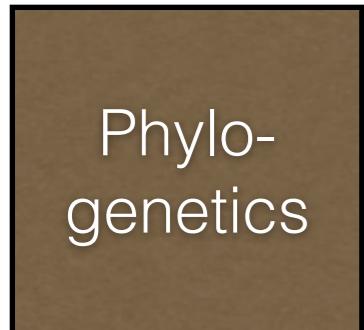
Fraser et al. (Science, 2009)



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

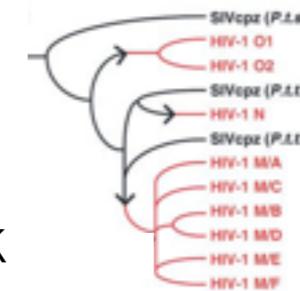
- ▶ Quantification of all model parameters!

# Genetic data contains information about the epidemiological dynamics



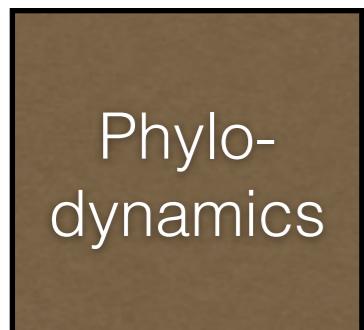
## State of process

- ▶ Time and geographic location of outbreak



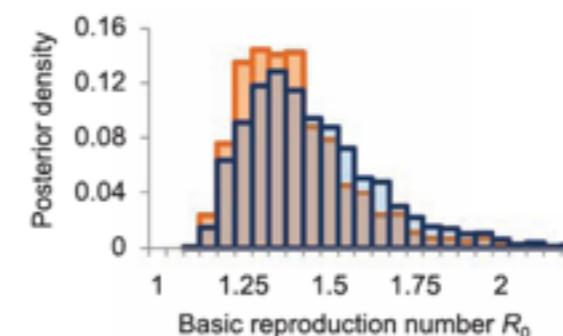
Emergence of a pathogen (here HIV)

Hahn et al. (Science, 2000)



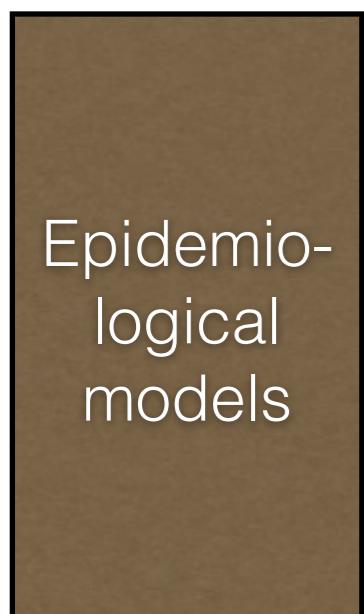
## Dynamics of process

- ▶ Transmission process



Basic reproductive number of an emerging pathogen (here H1N1)

Fraser et al. (Science, 2009)



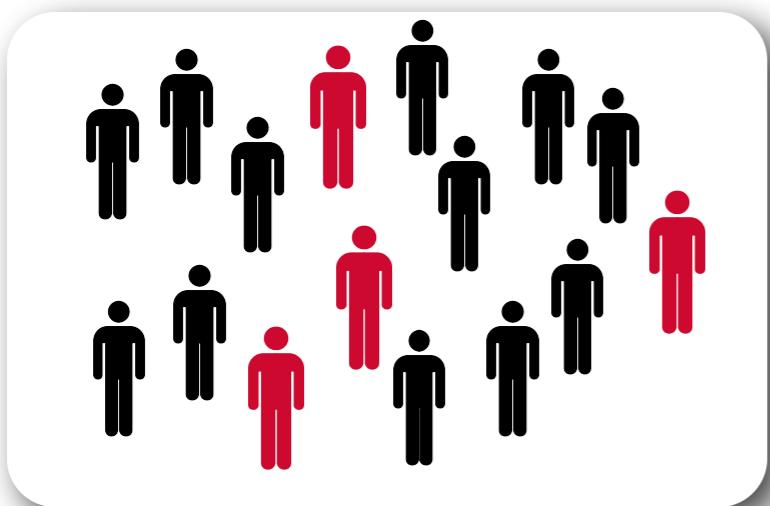
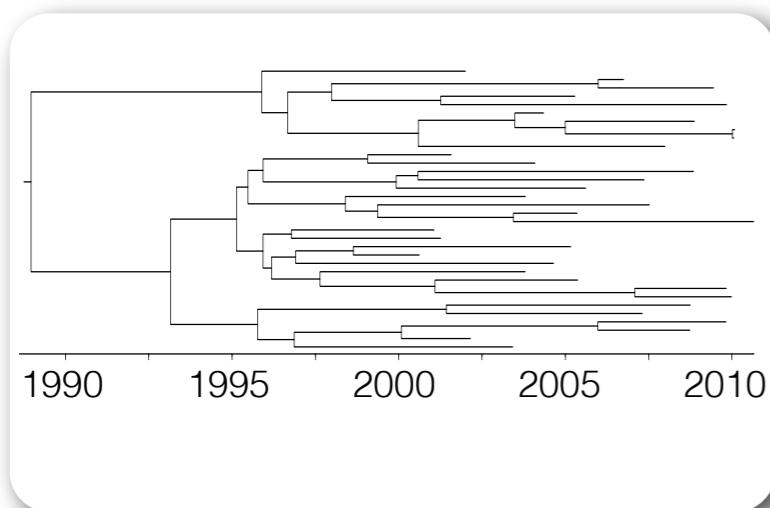
Ideally: assume epidemiological model in phylogenetics (birth-death model)

- ▶ Quantification of all model parameters!

Approximate: by population genetic model (coalescent)

# Merge epidemiology and phylogenetics

---



## **Methodology to:**

- 1) Reconstruct phylogenetic trees
- 2) Fit epidemiological models to phylogenetic trees

**Quantification of spread of infectious diseases?**

**Factors governing transmission dynamics?**

(e.g. population structure, environment, vaccines)

# General modeling framework

---

Epidemiological model

Growth of tree  
► parameters  $\eta$

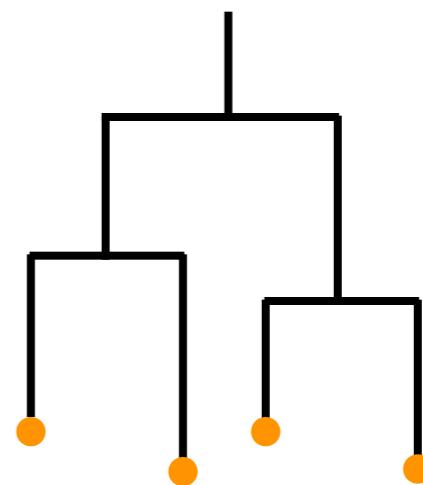


# General modeling framework

---

Epidemiological model

- Growth of tree
  - parameters  $\eta$



# General modeling framework

---

Epidemiological model

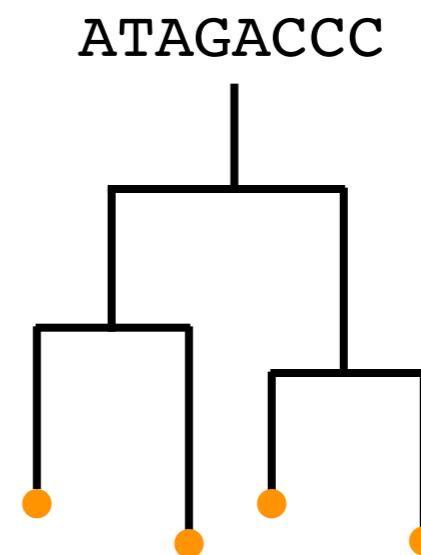
Growth of tree

- ▶ parameters  $\eta$

Evolutionary model

Evolution of sequences along tree

- ▶ parameters  $\theta$



# General modeling framework

---

Epidemiological model

Growth of tree

- ▶ parameters  $\eta$

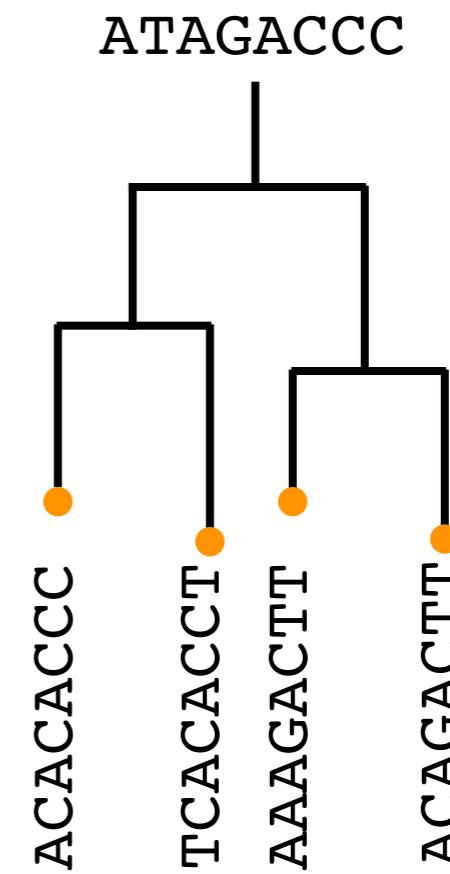
Evolutionary model

Evolution of sequences along tree

- ▶ parameters  $\theta$

Tree distribution

Defined through  $\eta$  and  $\theta$



# General modeling framework

---

Epidemiological model

Growth of tree  
▶ parameters  $\eta$

Evolutionary model

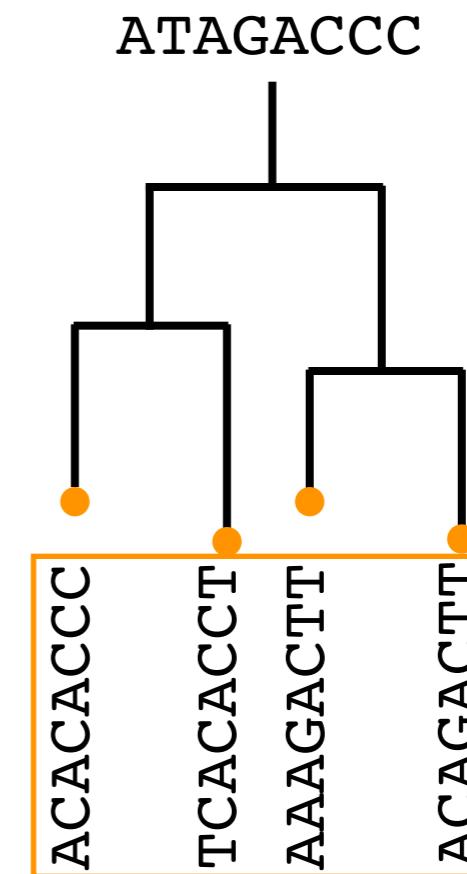
Evolution of sequences along tree  
▶ parameters  $\theta$

Tree distribution

Defined through  $\eta$  and  $\theta$

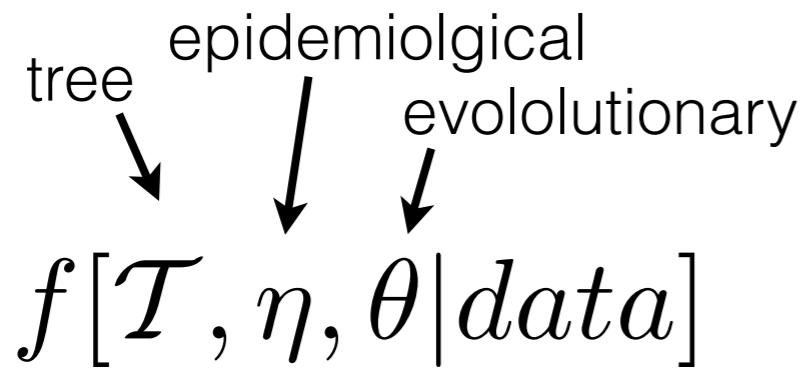
Data

Sequence alignment



# Bayesian approach for estimating epidemiological parameters

---



# Bayesian approach for estimating epidemiological parameters

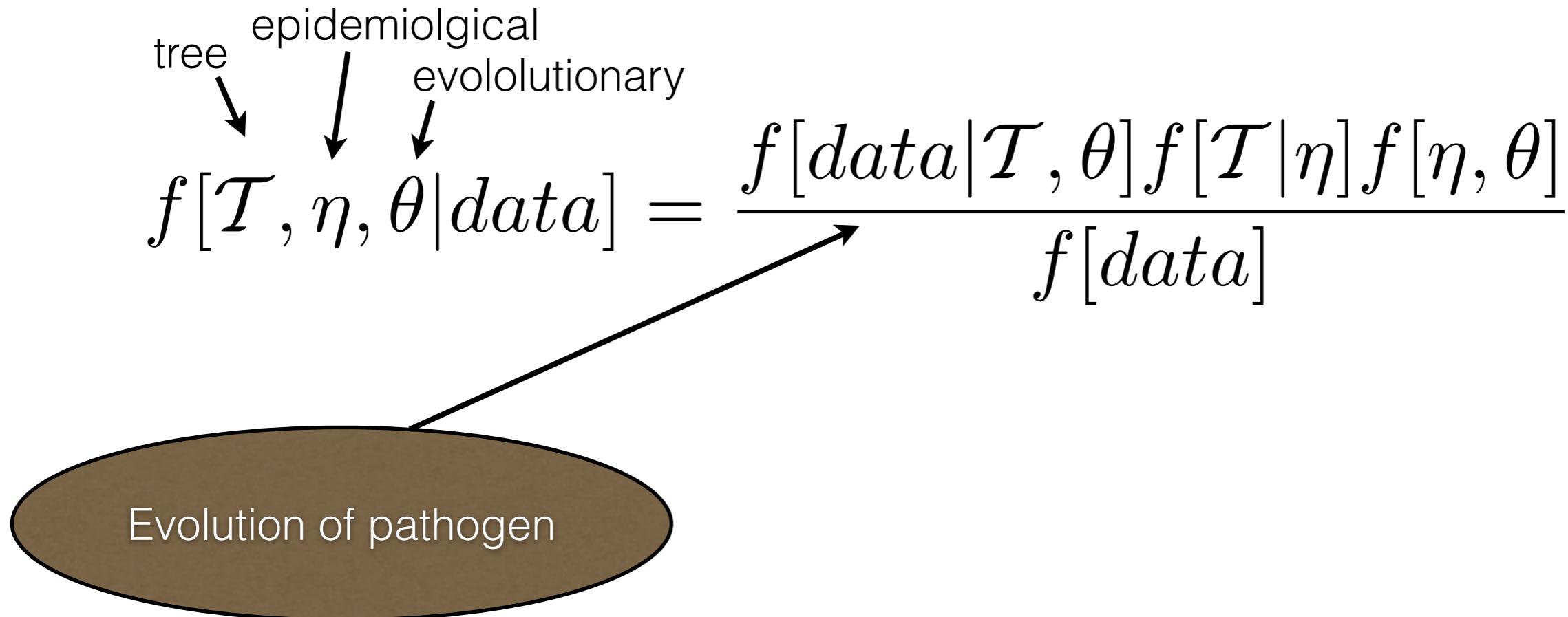
---

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

tree      epidemiological  
             ↓           ↓  
             evolutionary

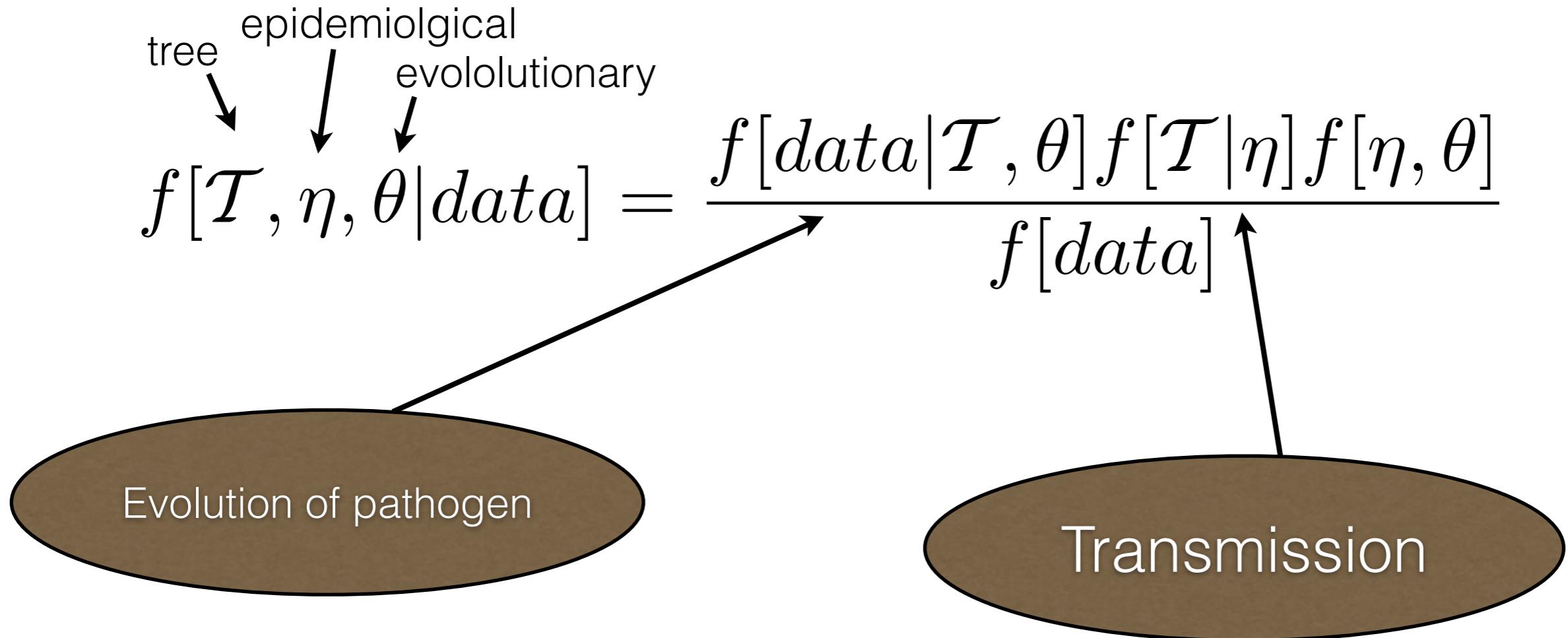
# Bayesian approach for estimating epidemiological parameters

---

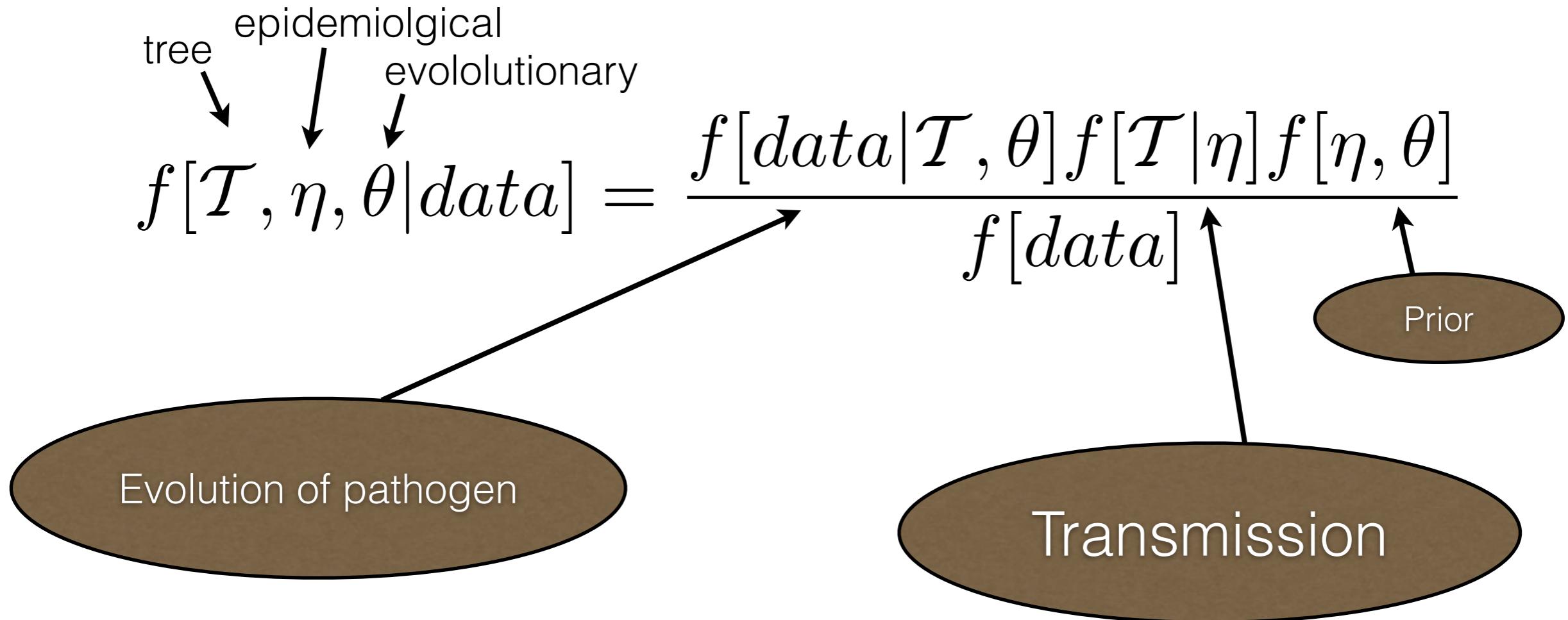


# Bayesian approach for estimating epidemiological parameters

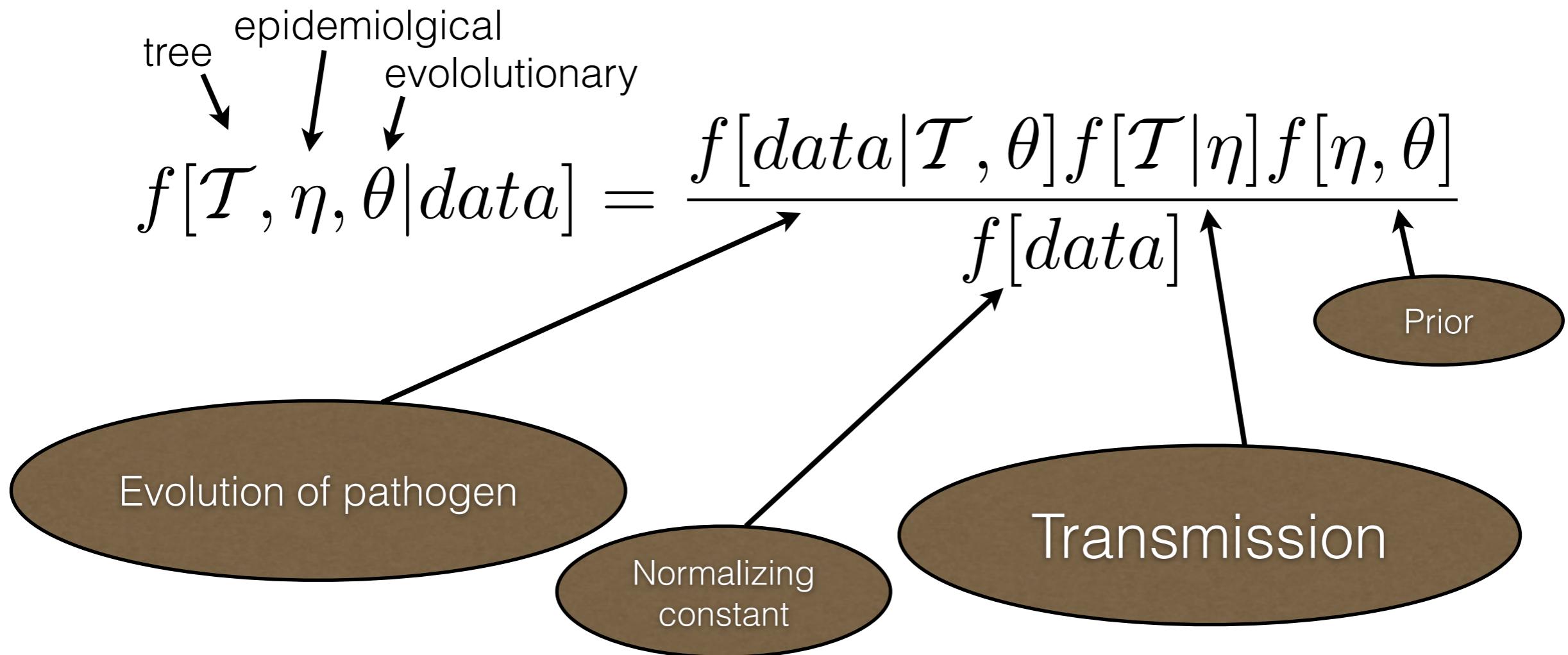
---



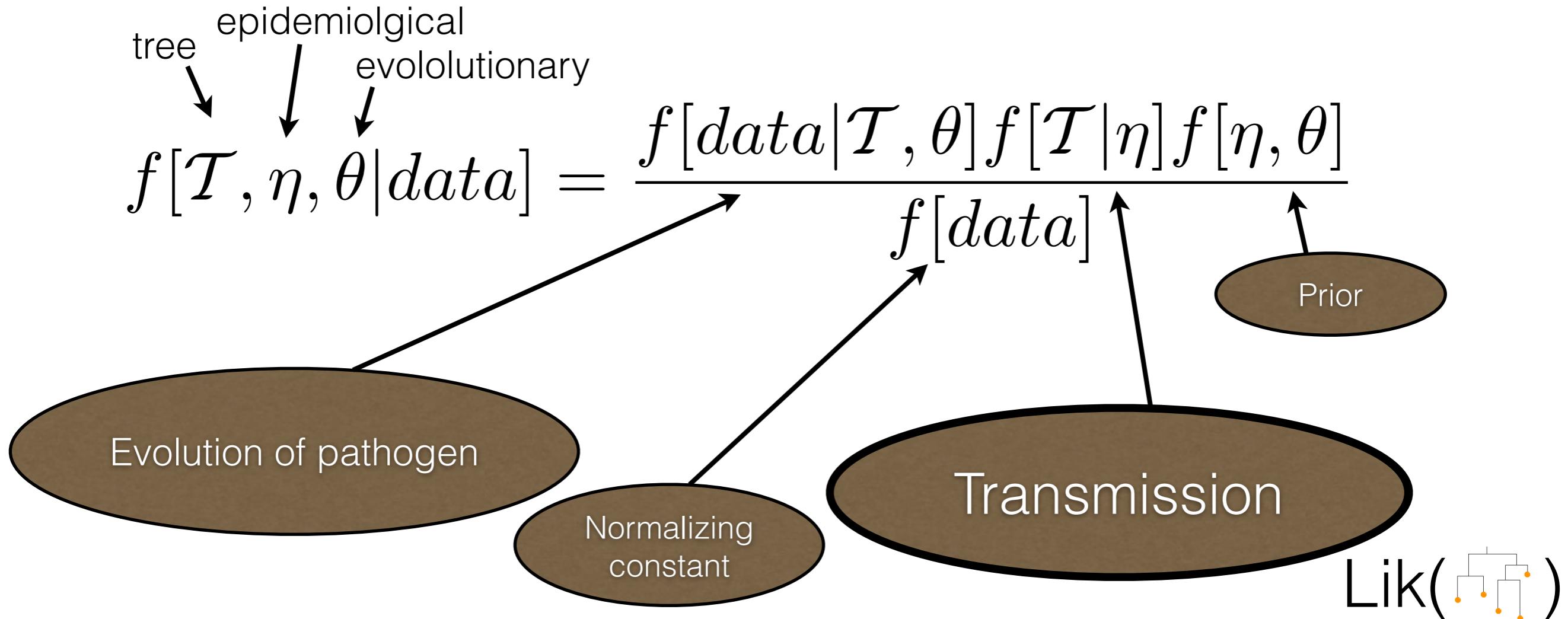
# Bayesian approach for estimating epidemiological parameters



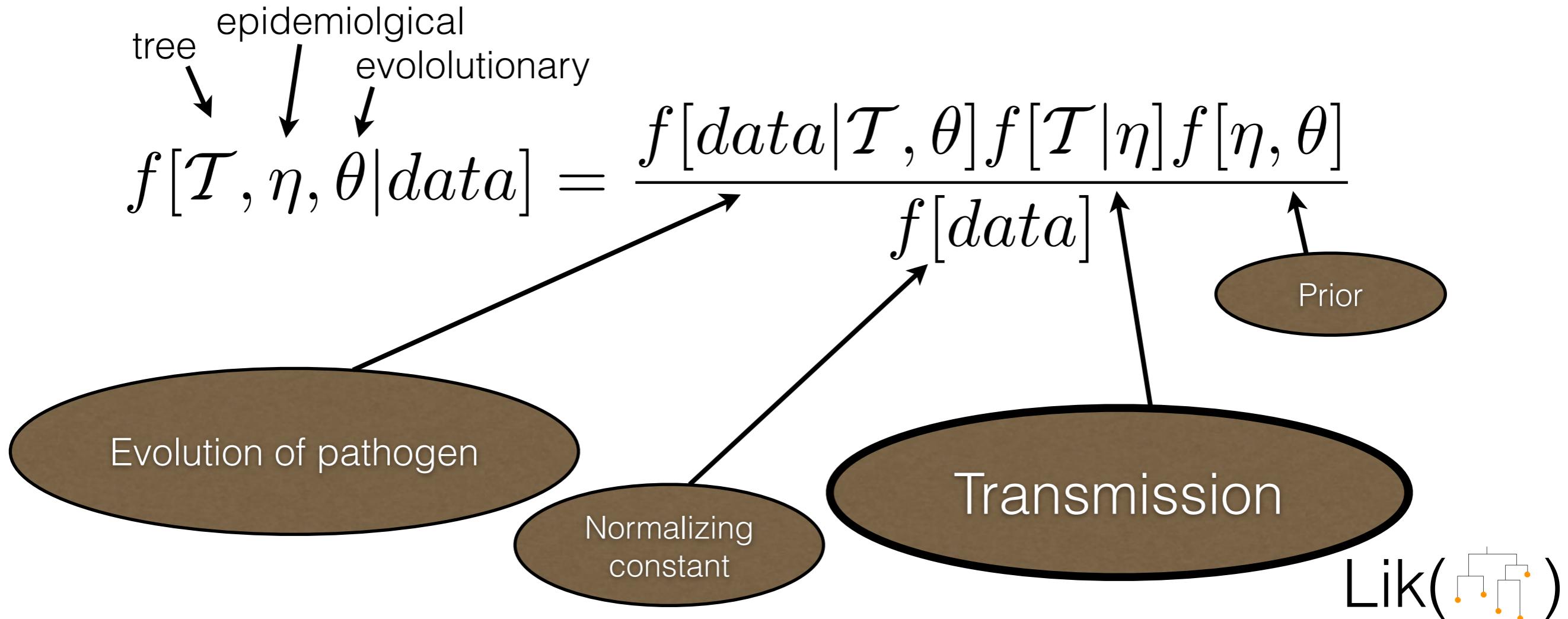
# Bayesian approach for estimating epidemiological parameters



# Bayesian approach for estimating epidemiological parameters



# Bayesian approach for estimating epidemiological parameters

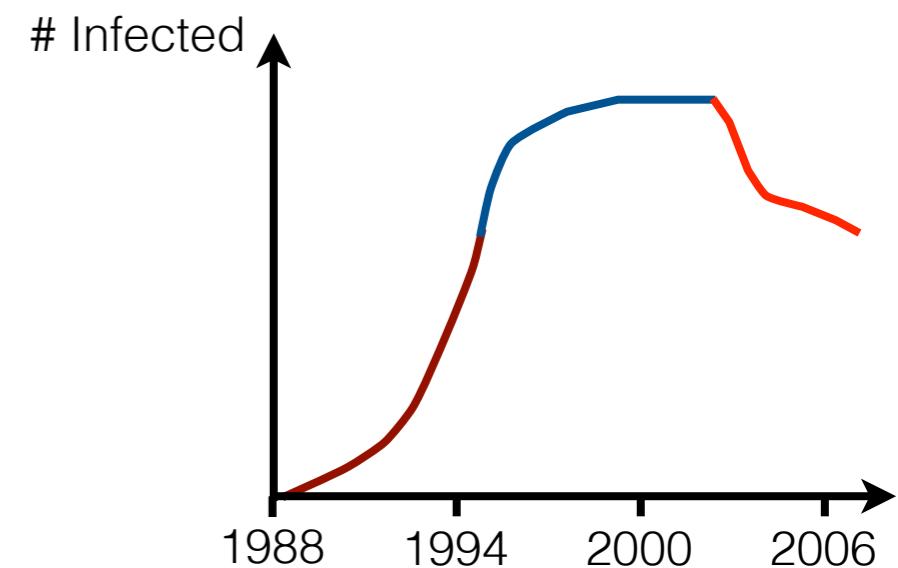


Implemented in Beast

Transmission model: coalescent or birth-death model

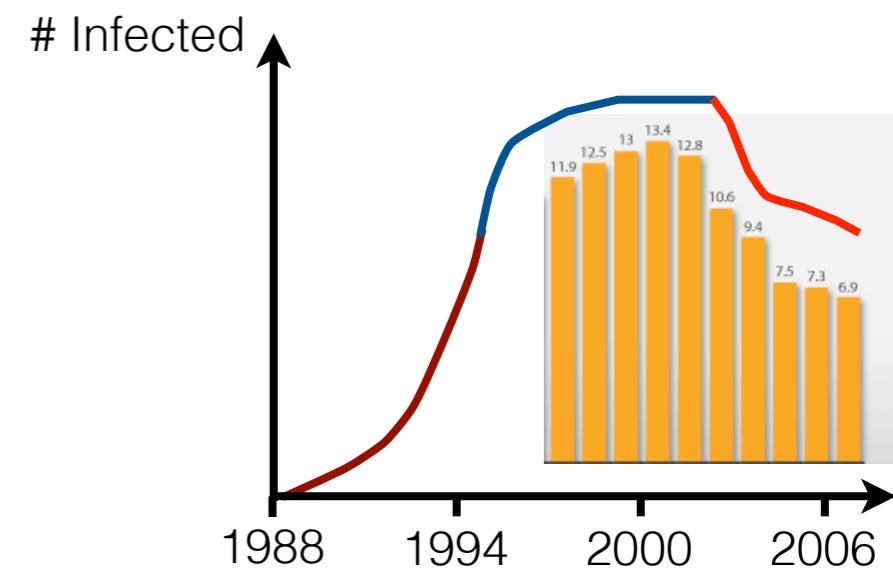
# Phylogenetic methods for different epidemic dynamics

---



# Phylogenetic methods for different epidemic dynamics

---

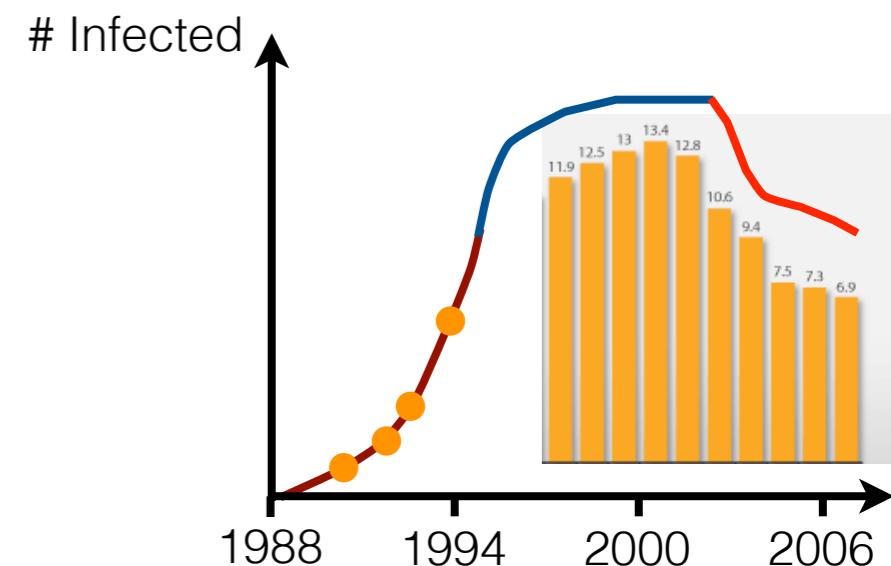


# Phylogenetic methods for different epidemic dynamics

---

## 1. ...for analyzing epidemic outbreaks

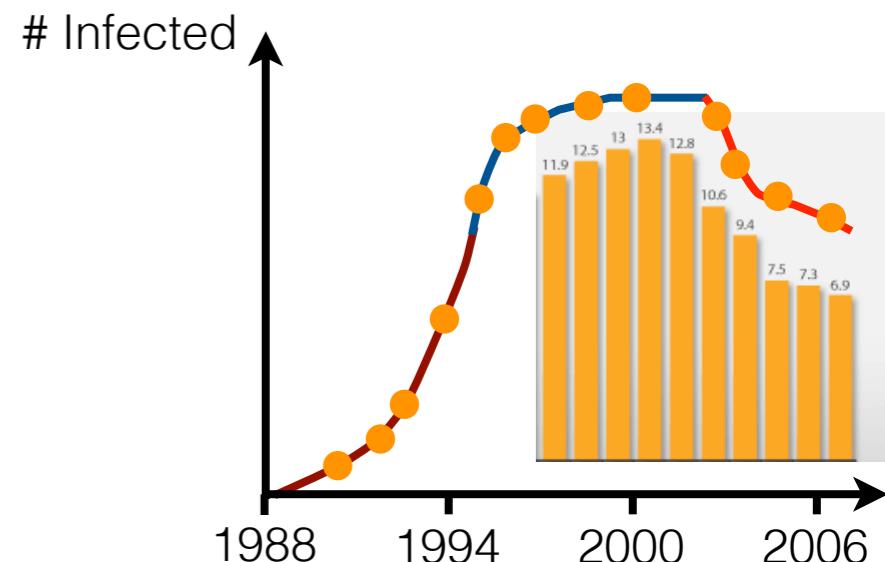
- ▶ C: Drummond et al. (*Genetics*, 2002)
- ▶ BD: Stadler et al. (*MBE*, 2012)



# Phylogenetic methods for different epidemic dynamics

---

1. ...for analyzing epidemic outbreaks
  - ▶ C: Drummond et al. (*Genetics*, 2002)
  - ▶ BD: Stadler et al. (*MBE*, 2012)
  
2. ...for identifying epidemiological changes over time
  - ▶ C: Drummond et al. (*MBE*, 2005)
  - ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)



# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

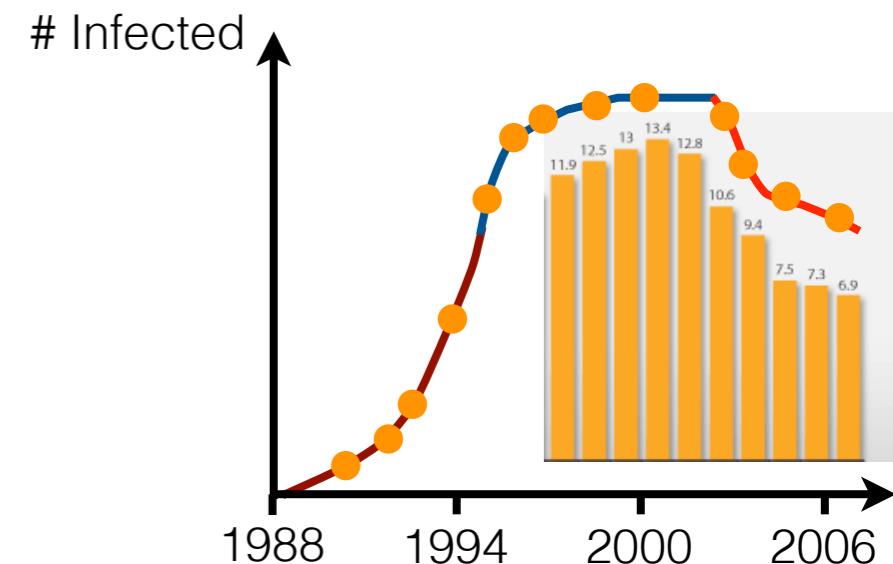
- ▶ C: Drummond et al. (*Genetics*, 2002)
- ▶ BD: Stadler et al. (*MBE*, 2012)

## 2. ...for identifying epidemiological changes over time

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

## 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)



# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

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

## 2. ...for identifying epidemiological changes over time

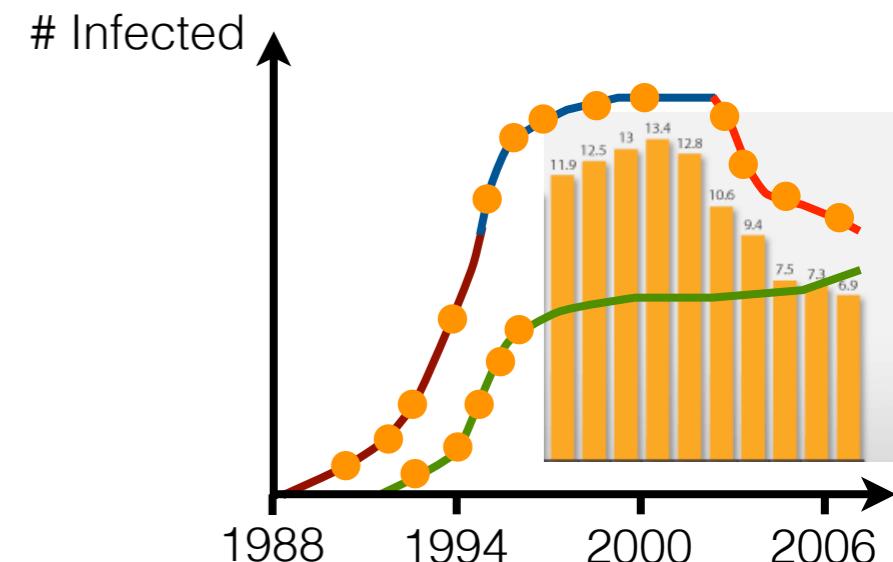
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

## 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)

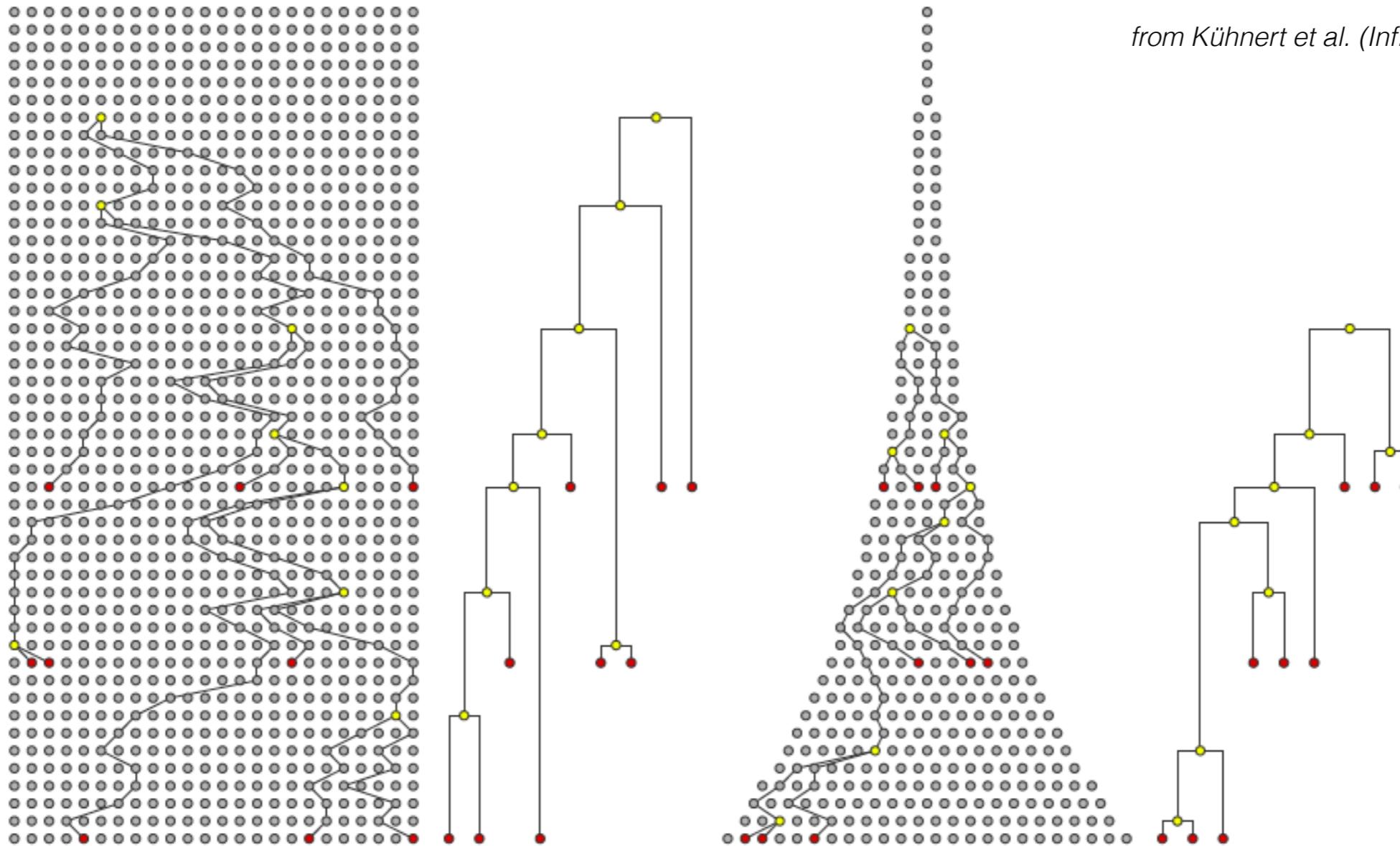
## 4. ...accounting for host population structure

- ▶ C: Volz (*Genetics*, 2012)
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013)



# Part I

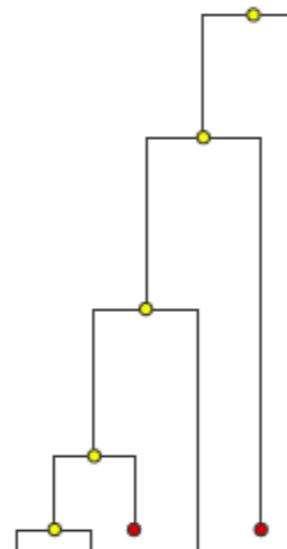
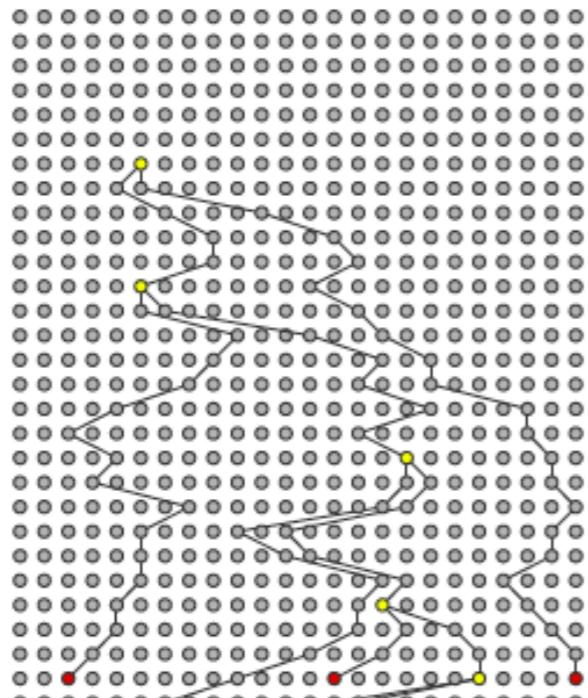
## Coalescent as a model for transmission



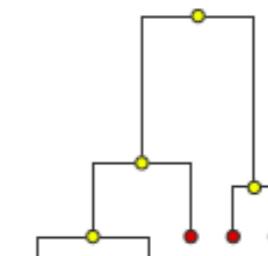
from Kühnert et al. (*Inf. Gen. Evol.*, 2011)

# Part I

## Coalescent as a model for transmission



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

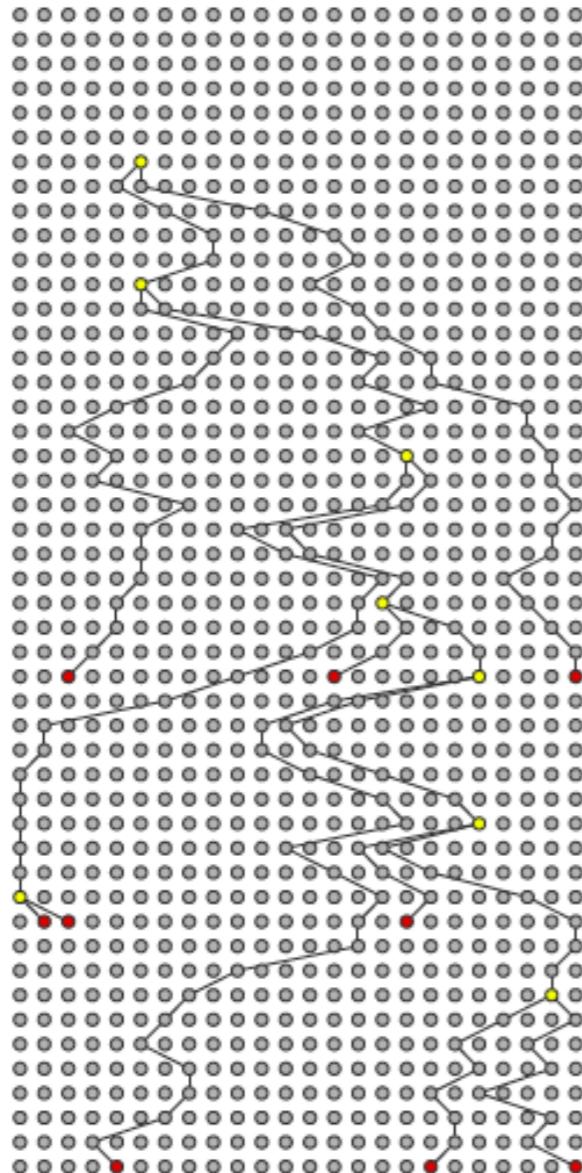


Lik ( ) ?

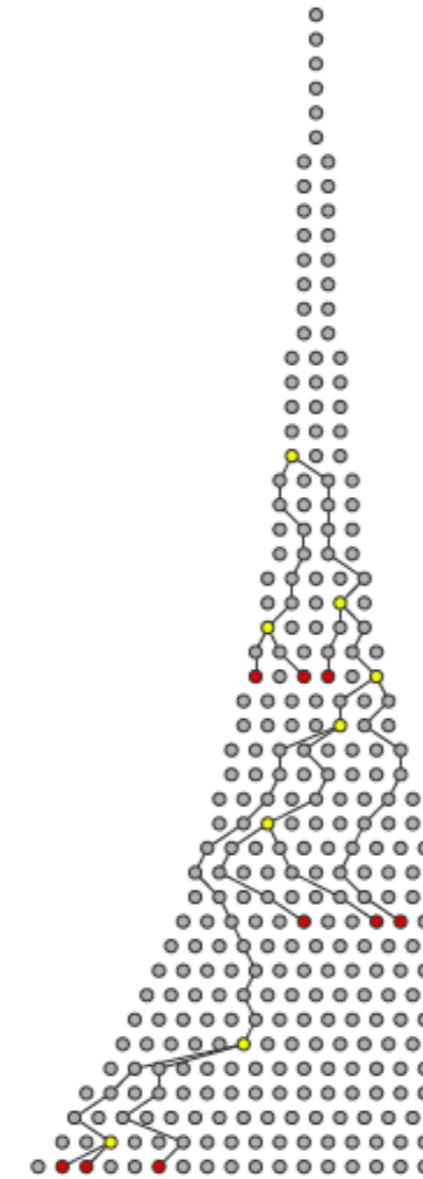
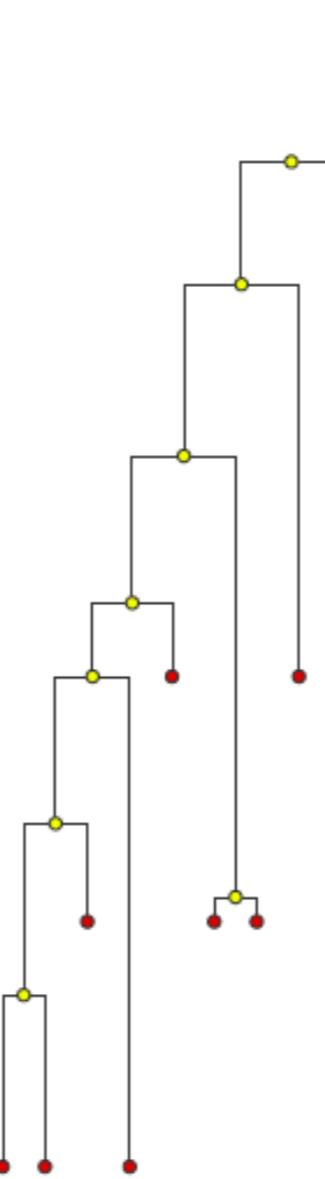


# Part I

## Coalescent as a model for transmission



$$f_G(g|\theta) = \frac{1}{\theta^{n-1}} \cdot \prod_{i=2}^{2n-1} e^{(-k_i(k_i-1)/2\theta)(t_i - t_{i-1})}$$



$$f_G(g|\theta, r) = \frac{1}{\theta^{n-1}} \cdot \prod_{i=2}^{2n-1} e^{rt_i} e^{(-k_i(k_i-1)/2\theta r)(rt_i - rt_{i-1})}$$

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

from Drummond et al. (*Genetics*, 2002)

# Estimating $R_0$ from sequencing data

---

Hepatitis C

Little known before first isolation, ie. pre-1989.

How can we determine basic reproductive number  $R_0$ ?

# Estimating $R_0$ from sequencing data

---

Hepatitis C

Little known before first isolation, ie. pre-1989.

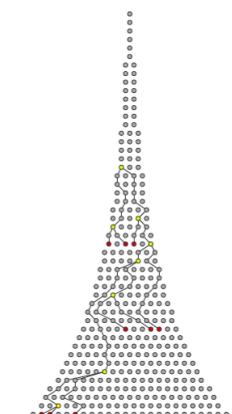
How can we determine basic reproductive number  $R_0$ ?

Inference

40-100 sequences per subtype

Fit exponential-growth coalescent to sequences and obtain growth rate  $r$  (i.e.  $\eta=r$ )

$R_0 = rD+1$  (where  $D$  is expected time of infectiousness)



# Estimating $R_0$ from sequencing data

Hepatitis C

Little known before first isolation, ie. pre-1989.

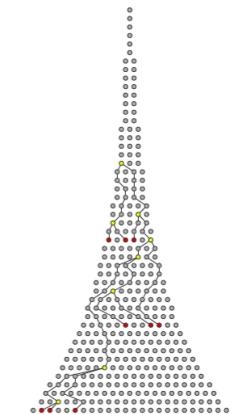
How can we determine basic reproductive number  $R_0$ ?

Inference

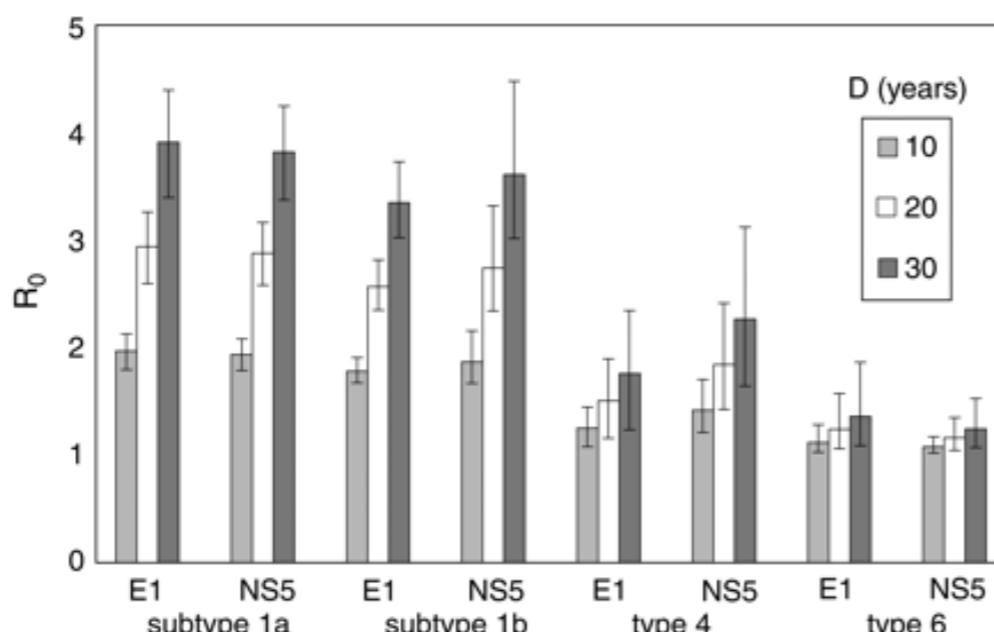
40-100 sequences per subtype

Fit exponential-growth coalescent to sequences and obtain growth rate  $r$  (i.e.  $\eta=r$ )

$R_0 = rD+1$  (where  $D$  is expected time of infectiousness)

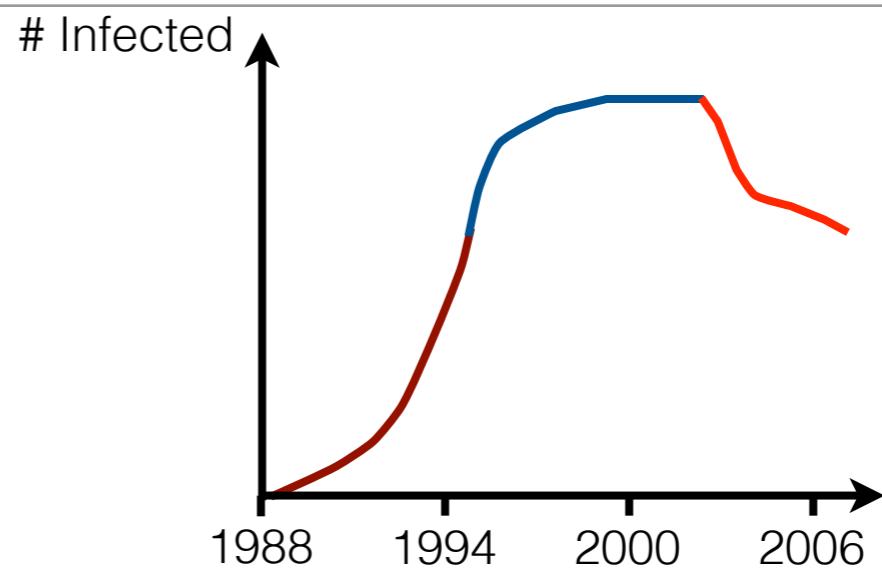


Results

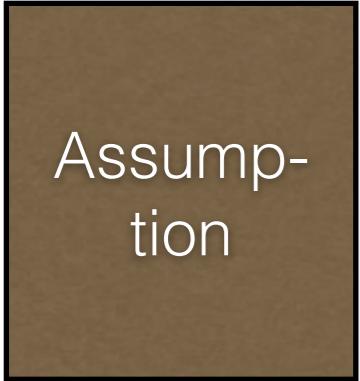


# Bayesian skyline plot relaxes constant growth rate

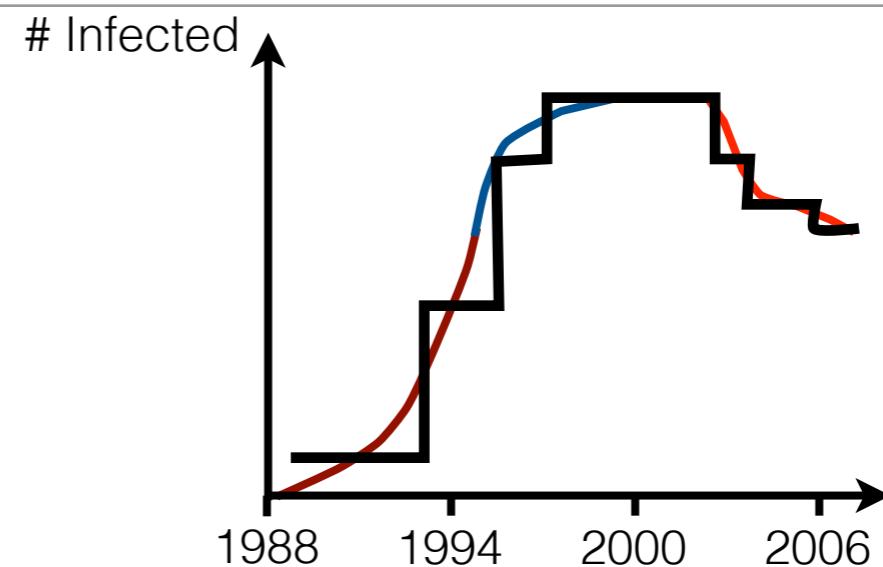
---



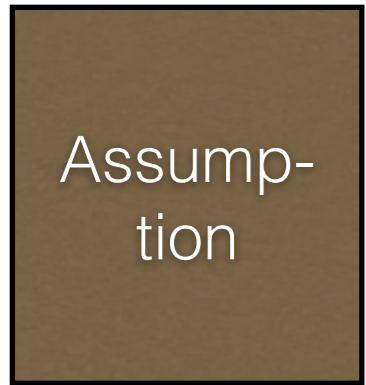
# Bayesian skyline plot relaxes constant growth rate



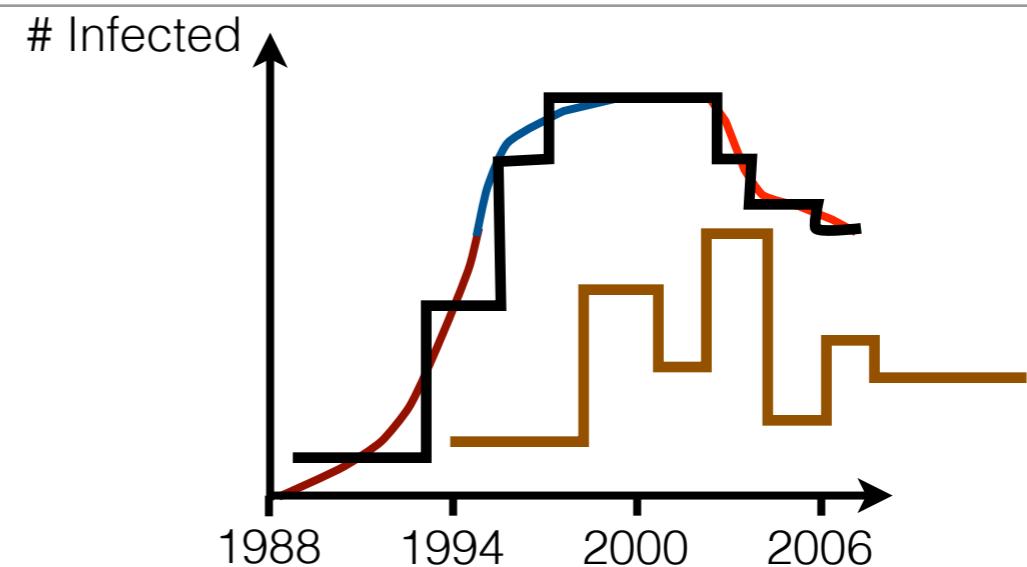
Piecewise constant population size



# Bayesian skyline plot relaxes constant growth rate



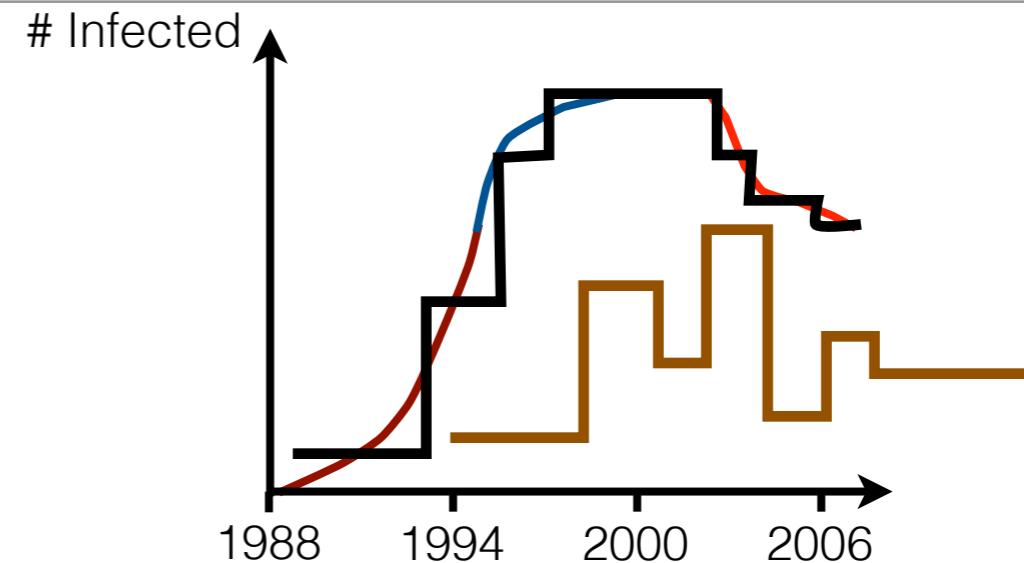
Piecewise constant population size



# Bayesian skyline plot relaxes constant growth rate

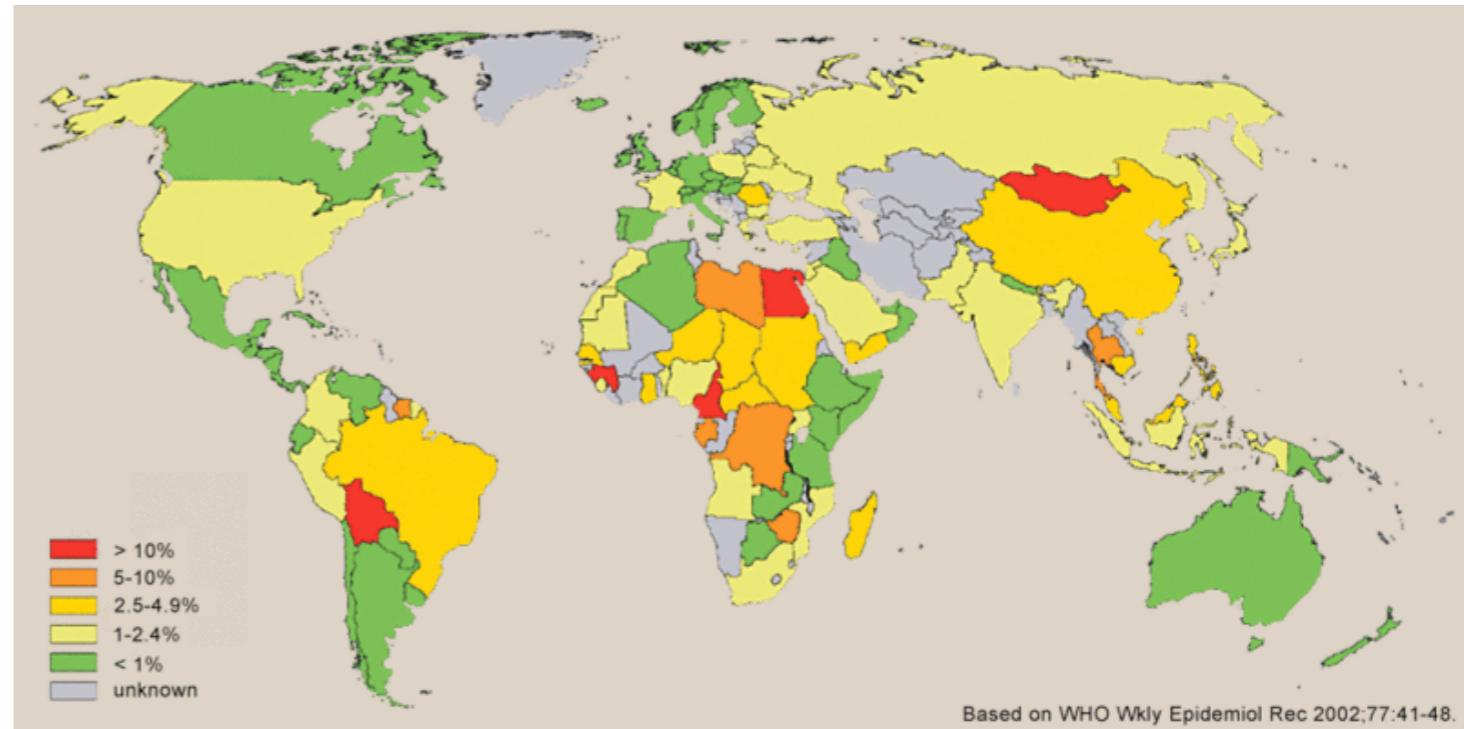
Assump-  
tion

Piecewise constant  
population size



HCV in  
Egypt

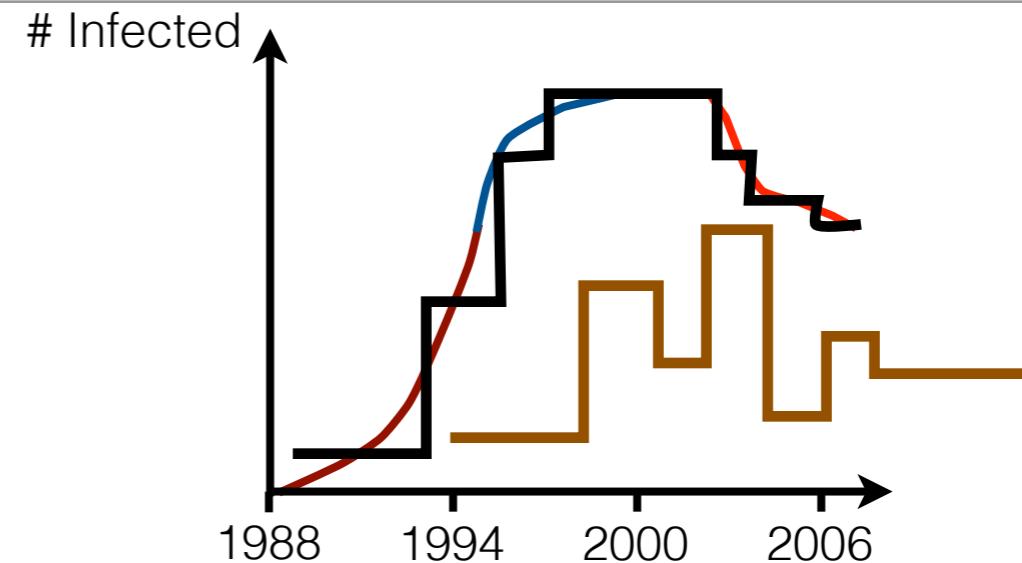
- ▶ Egypt highest HCV prevalence worldwide (about 13 %)
- ▶ Neighboring countries low prevalence
- ▶ When did the high prevalence emerge?
- ▶ Why?



# Bayesian skyline plot relaxes constant growth rate

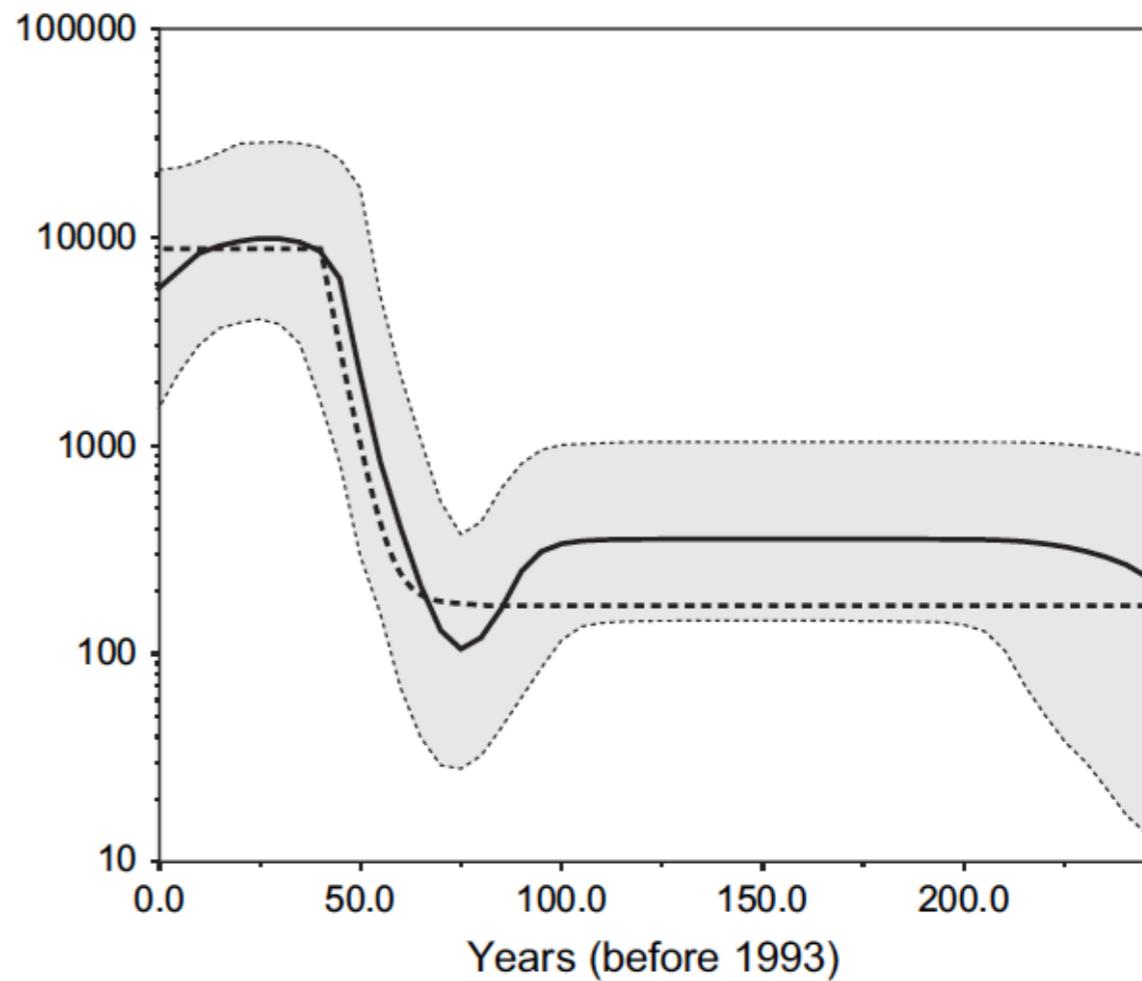
Assump-  
tion

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?



# Limitations of classic coalescent

---

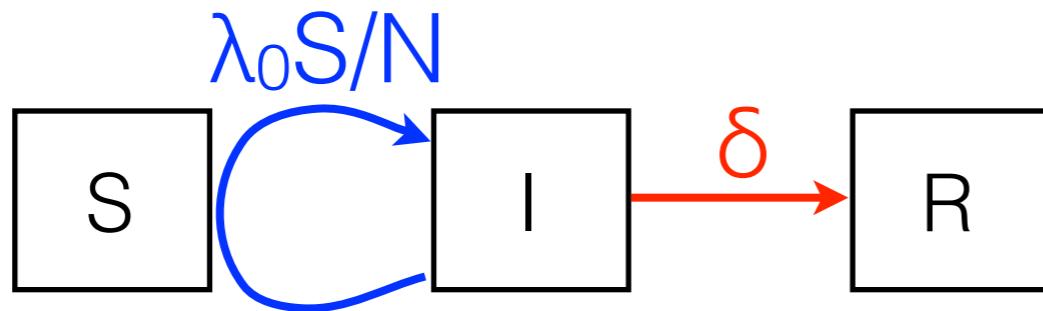
Coales-  
cent

Coalescent rates only depend on # infected, i.e.  $\lambda_0 l!$

# Limitations of classic coalescent

Coalescent

Coalescent rates only depend on # infected, i.e.  $\lambda_0 I$ !



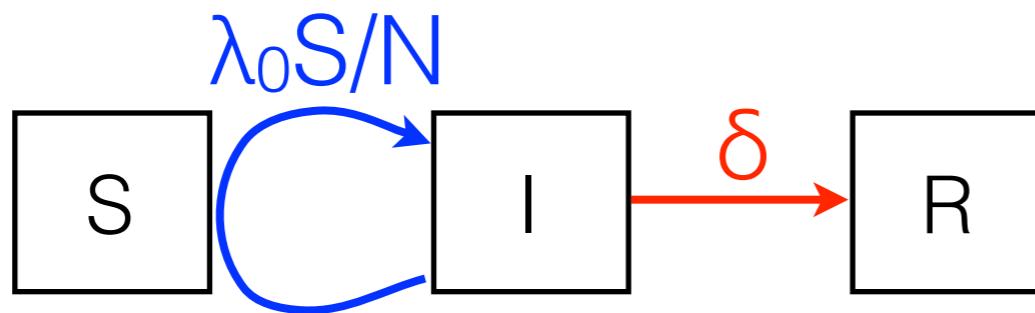
SIR  
extension

# Limitations of classic coalescent

Coalescent

Coalescent rates only depend on # infected, i.e.  $\lambda_0 I$ !

SIR  
extension

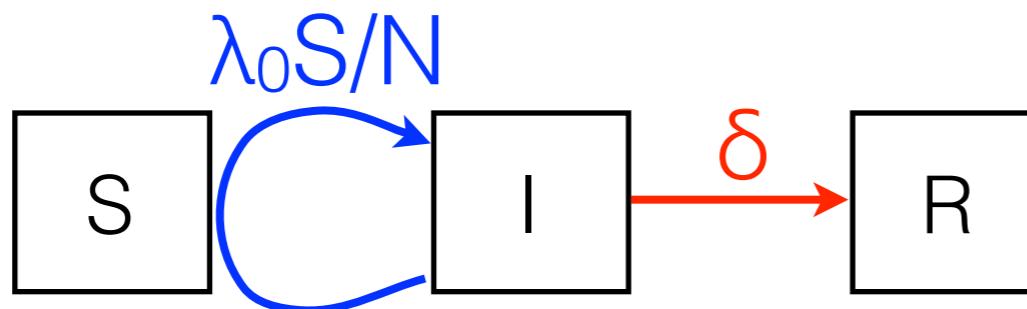


overall coalescent rate:  $\lambda_0 IS/N$

# Limitations of classic coalescent

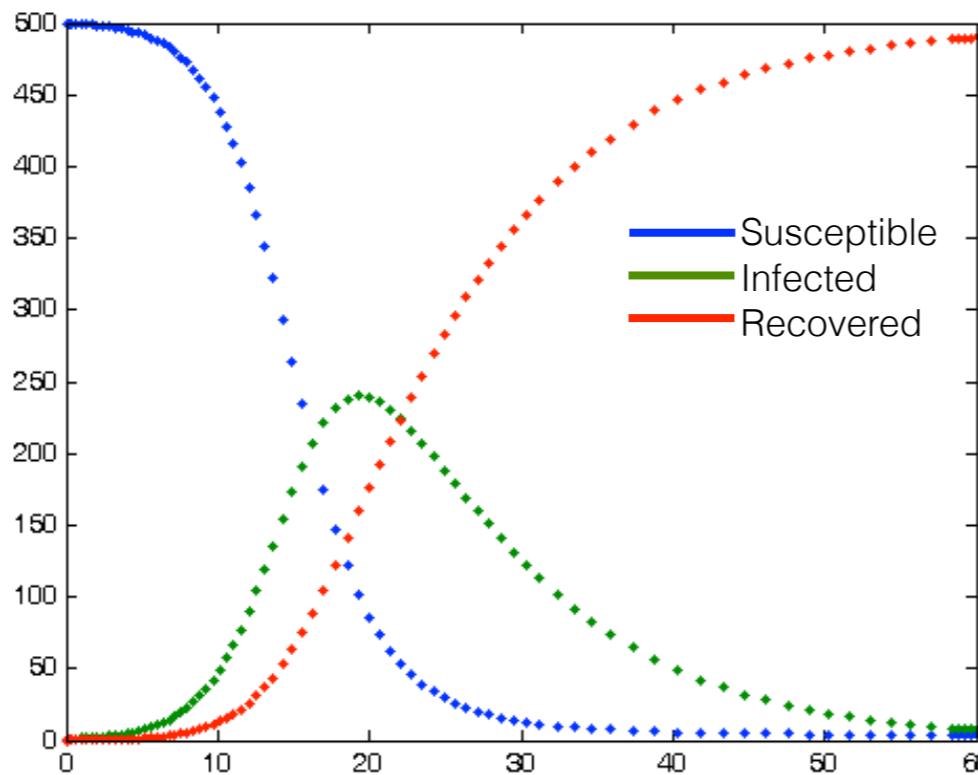
Coalescent

Coalescent rates only depend on # infected, i.e.  $\lambda_0 I$ !



overall coalescent rate:  $\lambda_0 I S/N$

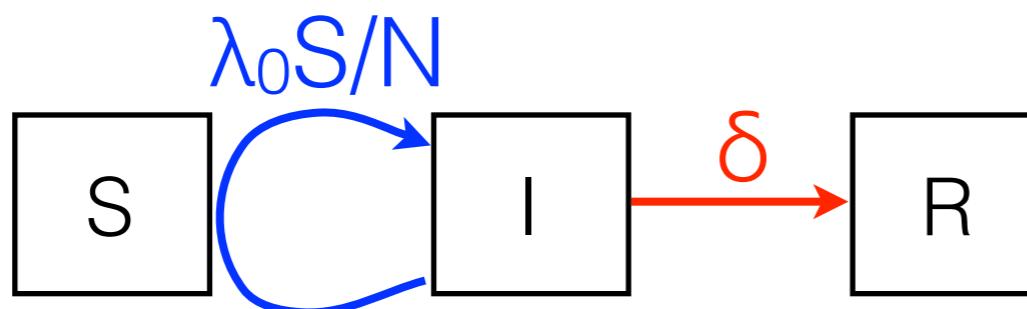
SIR  
extension



# Limitations of classic coalescent

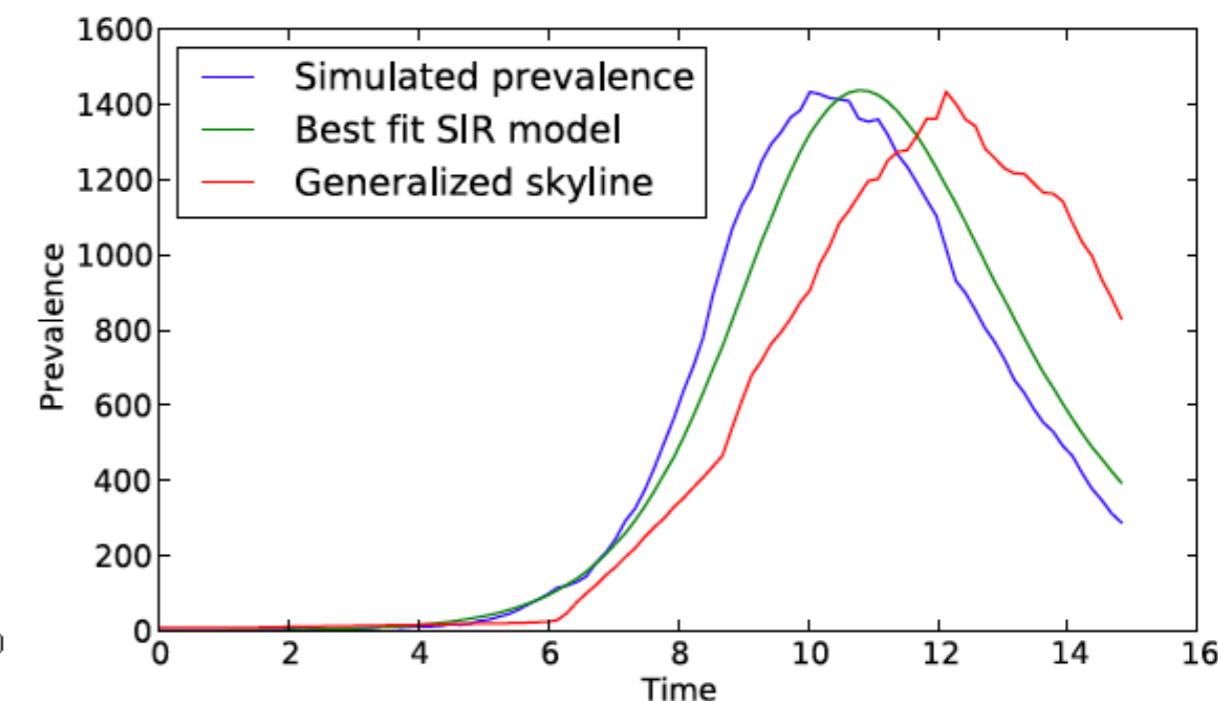
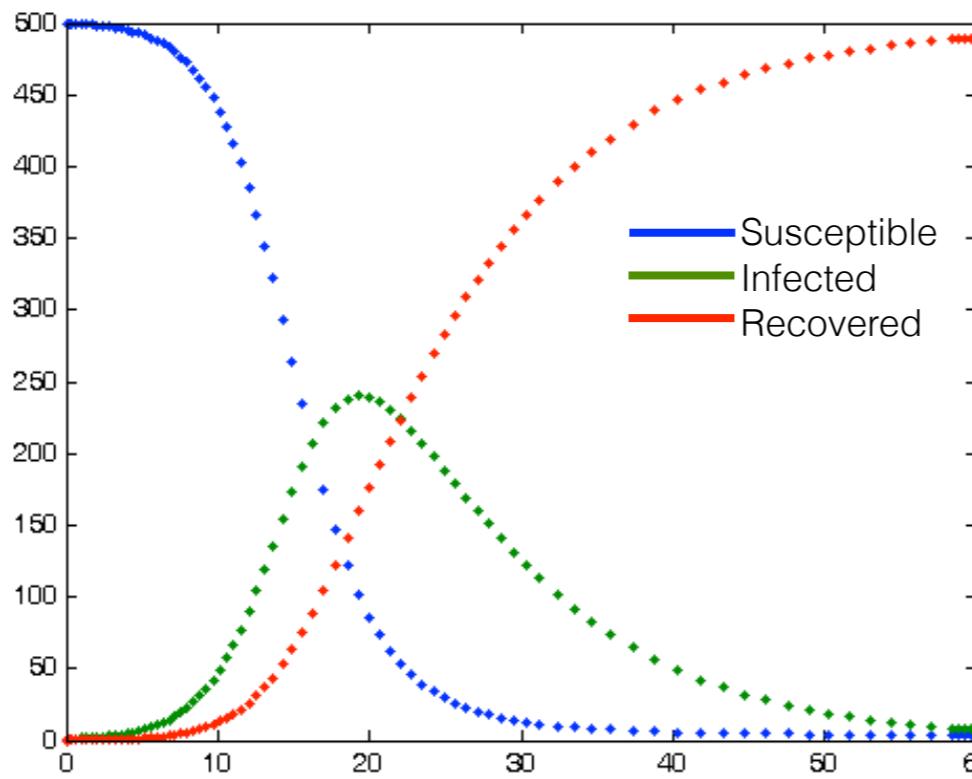
Coalescent

Coalescent rates only depend on # infected, i.e.  $\lambda_0 I$ !



overall coalescent rate:  $\lambda_0 IS/N$

SIR  
extension



# Phylogenetic methods for different epidemic dynamics

---

1. ...for analyzing epidemic outbreaks

► *C: Drummond et al. (Genetics, 2002)*

2. ...for identifying epidemiological changes over time

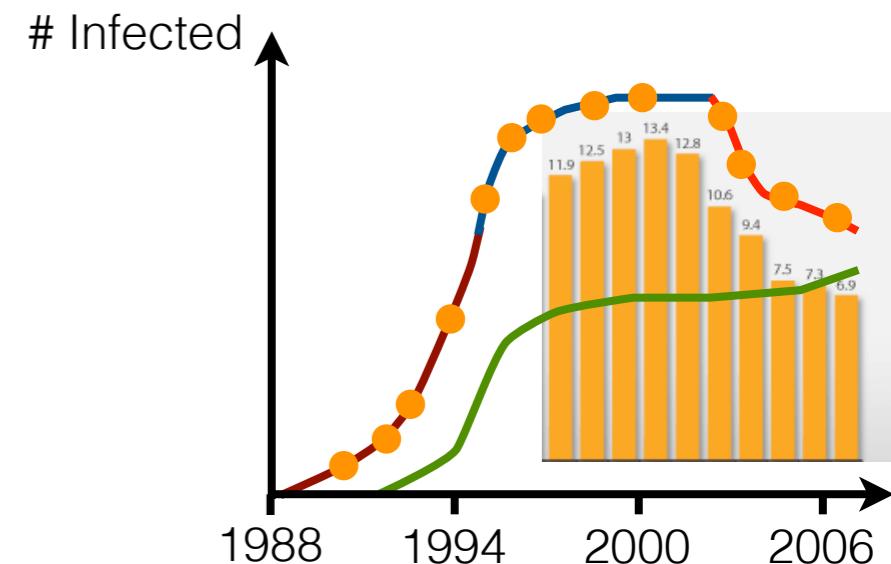
► *C: Drummond et al. (MBE, 2005)*

3. ...accounting for SIR model dynamics

► *C: Volz et al. (Genetics, 2009)*

4. ...accounting for host population structure

► *C: Volz (Genetics, 2012)*



# Phylogenetic methods for different epidemic dynamics

1. ...for analyzing epidemic outbreaks

► *C: Drummond et al. (Genetics, 2002)*

2. ...for identifying epidemiological changes over time

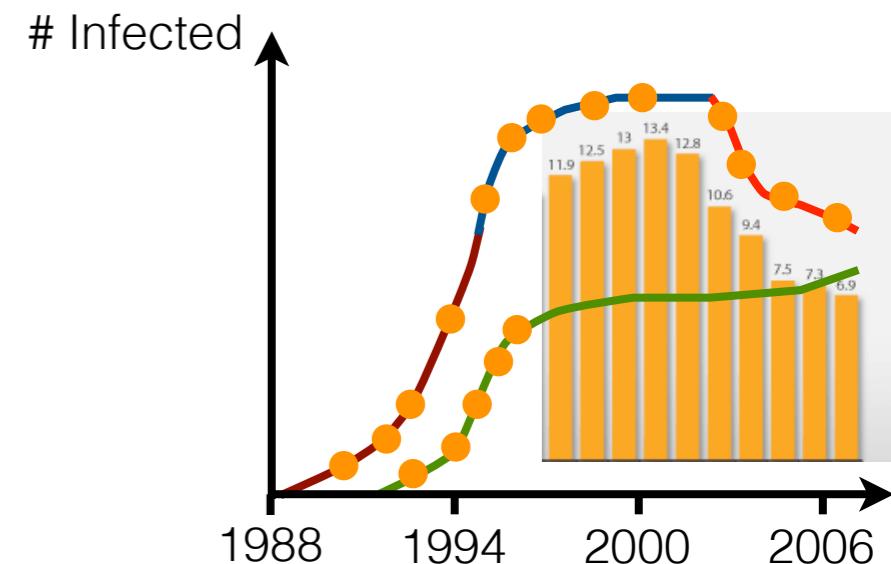
► *C: Drummond et al. (MBE, 2005)*

3. ...accounting for SIR model dynamics

► *C: Volz et al. (Genetics, 2009)*

4. ...accounting for host population structure

► *C: Volz (Genetics, 2012)*



# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

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

## 2. ...for identifying epidemiological changes over time

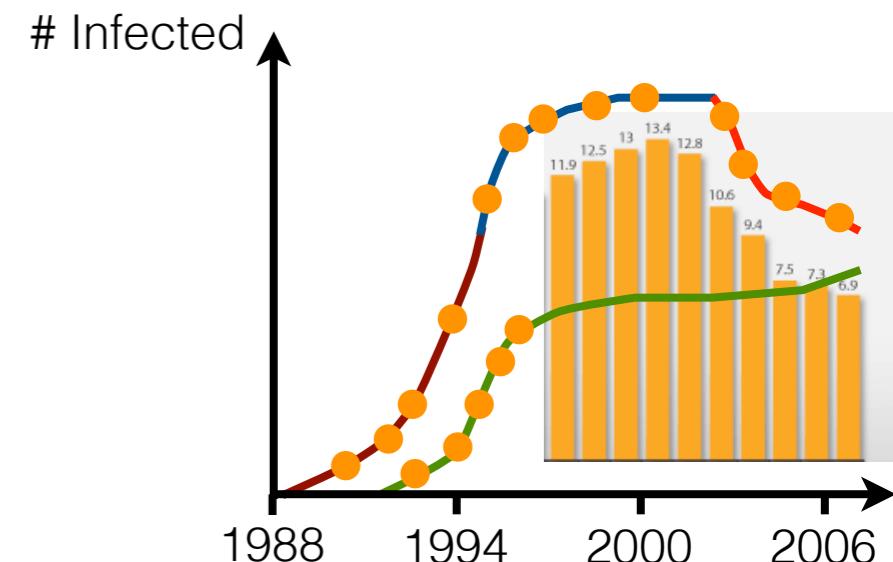
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

## 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 host population structure

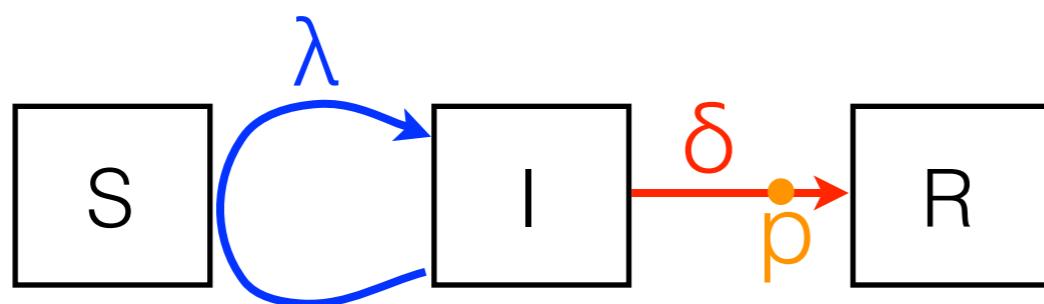
- ▶ C: Volz (*Genetics*, 2012)
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013)



# Part II

## Birth-death model for transmission

Birth-death  
model



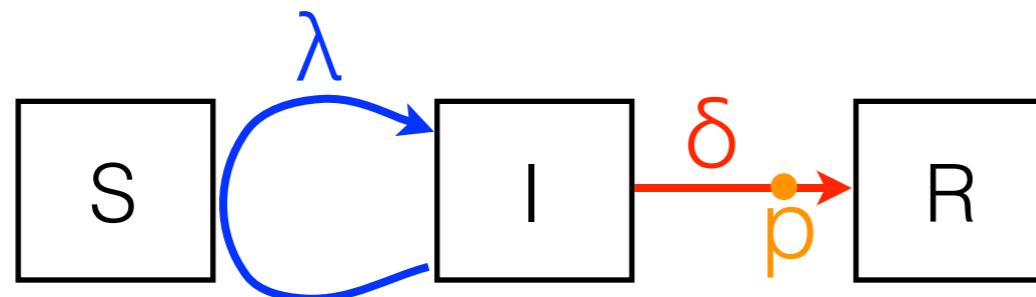
- transmission rate  $\lambda$
- “becoming-non-infectious” rate  $\delta$
- sampling probability  $p$

**Parameters may depend on:**

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

# Part II

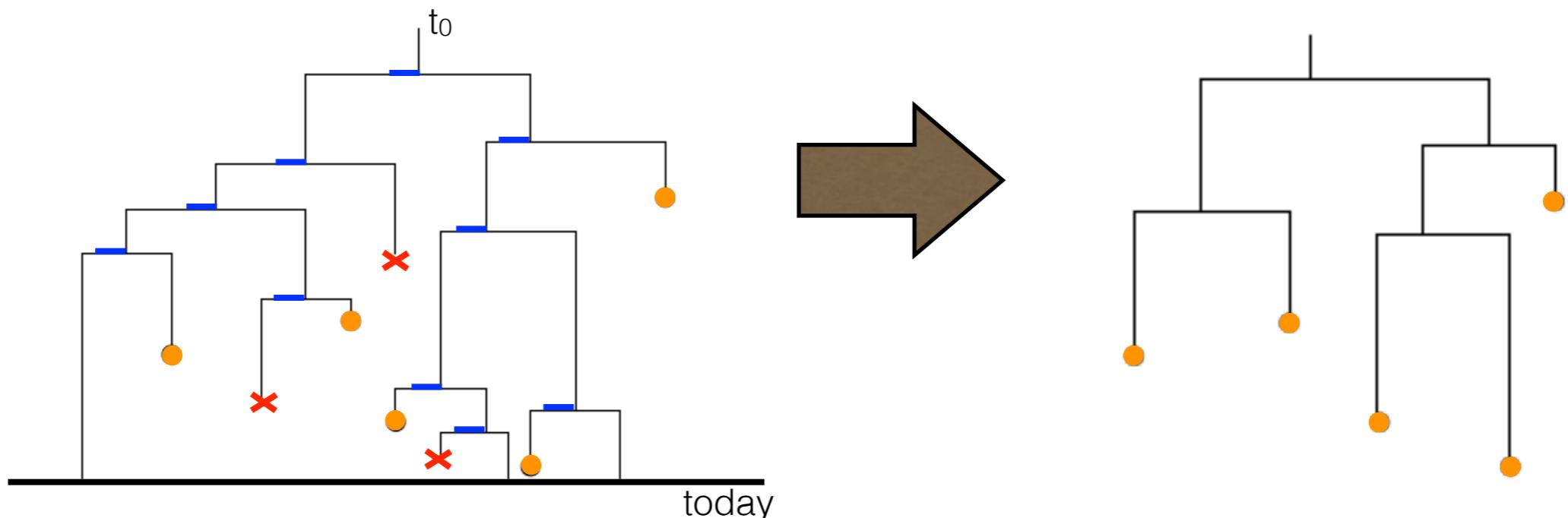
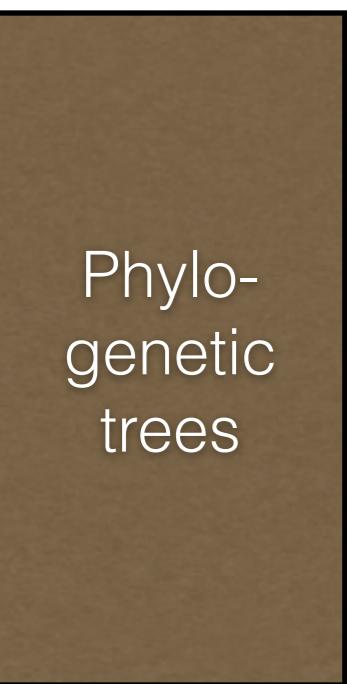
## Birth-death model for transmission



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

### Parameters may depend on:

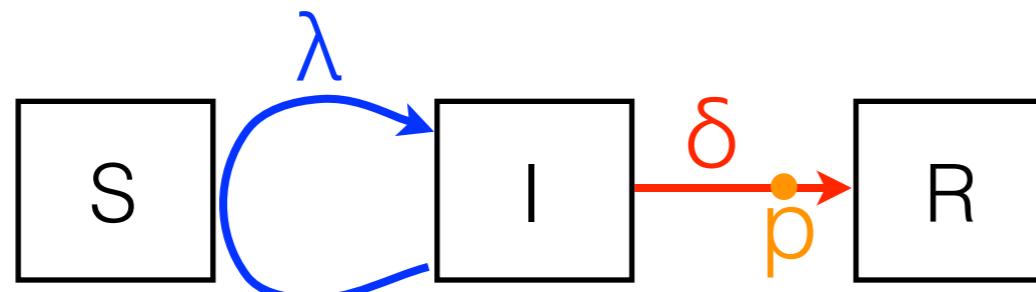
- 1) constant (epidemic outbreak)
- 2) time (environmental effect)
- 3) # of susceptibles (SIR model)
- 4) type of infected individual



# Part II

## Birth-death model for transmission

Birth-death  
model



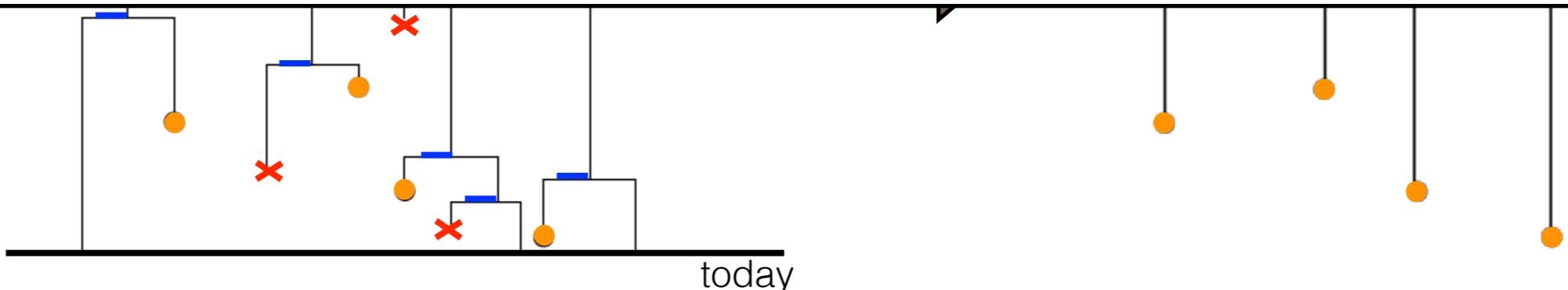
- transmission rate  $\lambda$
- “becoming-non-infectious” rate  $\delta$
- sampling probability  $p$

### Parameters may depend on:

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

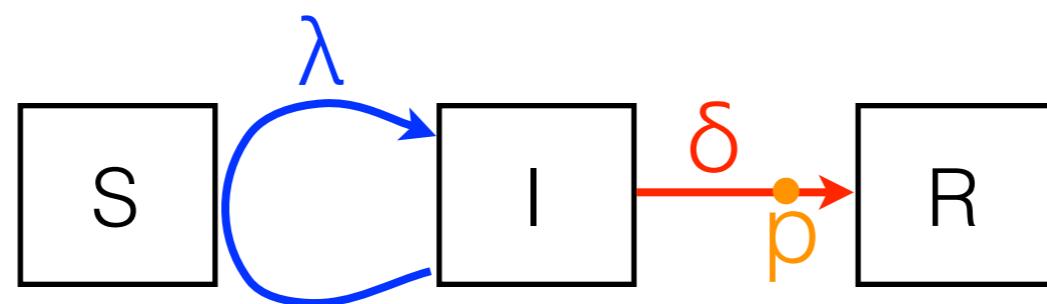
Lik ( ) ?

genetic  
trees



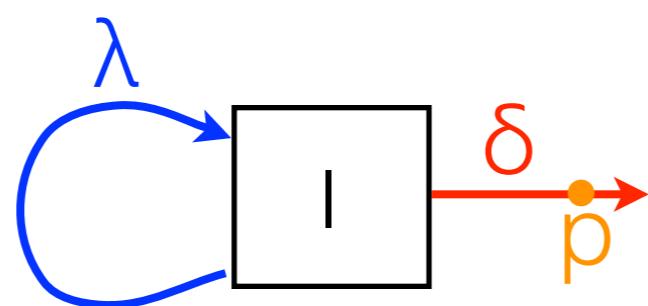
# 1) Epidemic outbreaks

---

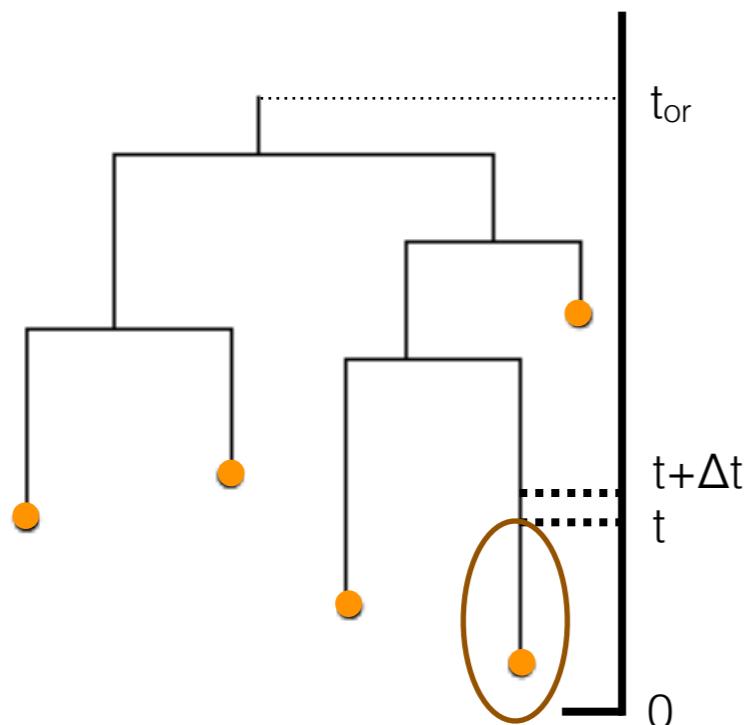
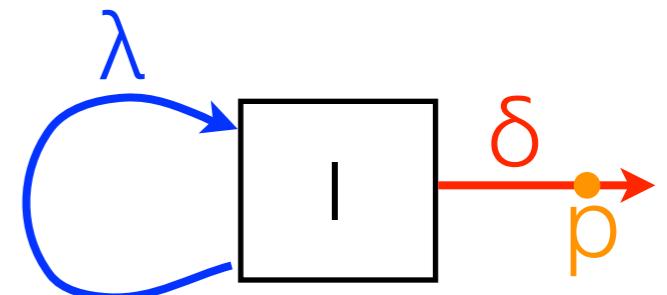


# 1) Epidemic outbreaks

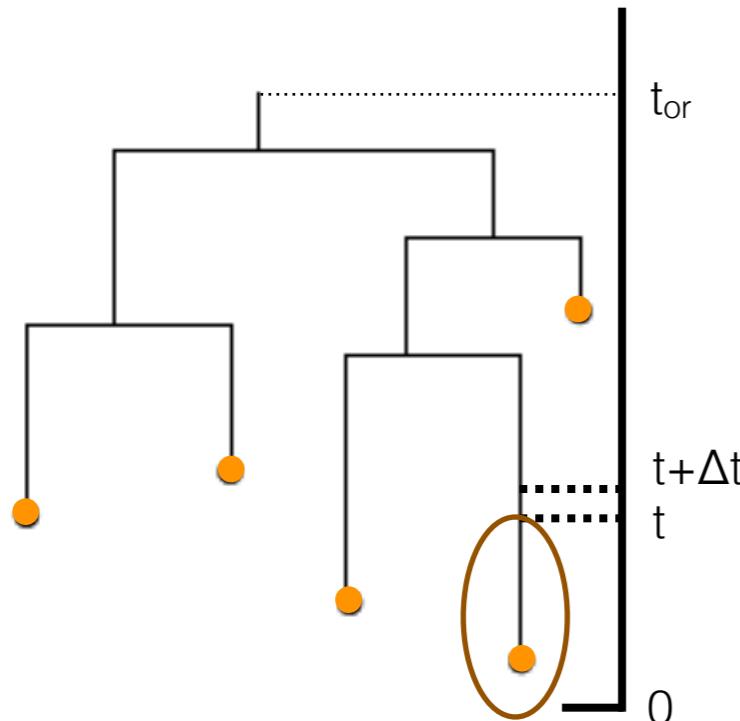
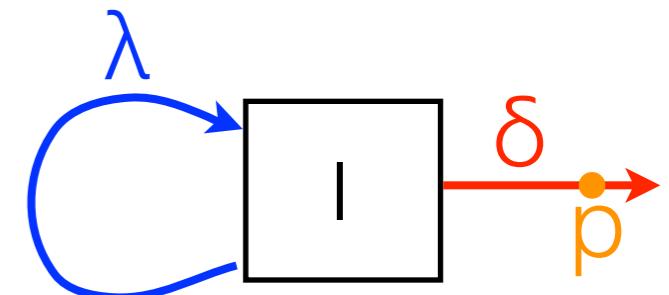
---



# 1) Epidemic outbreaks

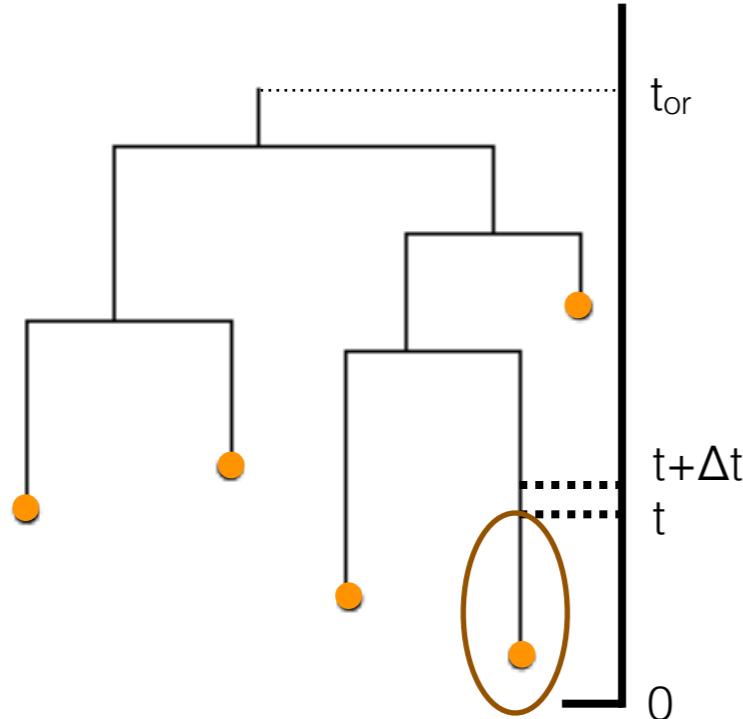
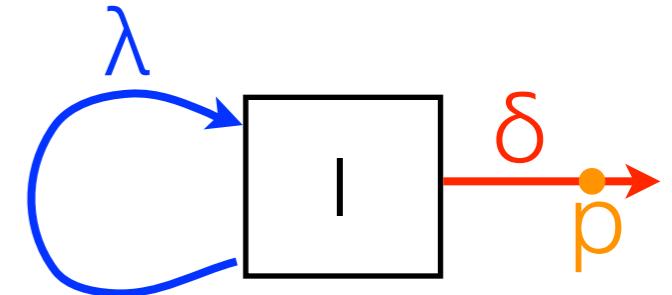


# 1) Epidemic outbreaks



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

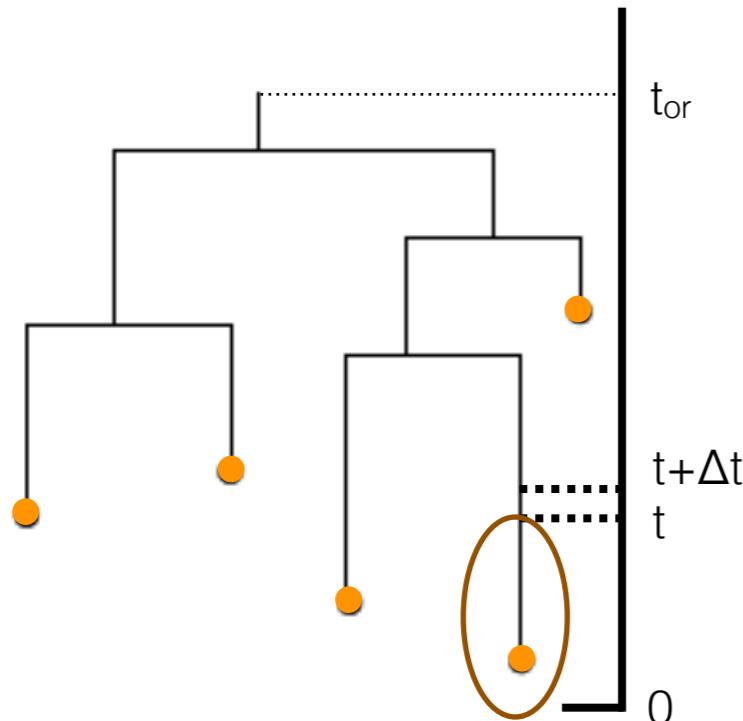
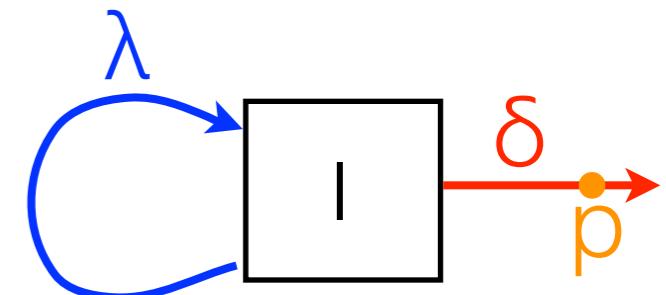
# 1) Epidemic outbreaks



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

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

# 1) Epidemic outbreaks

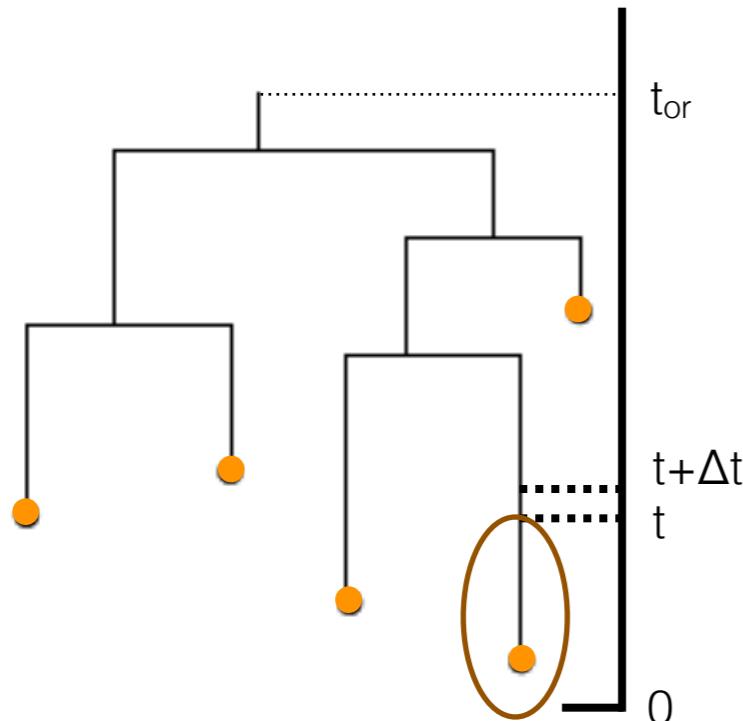
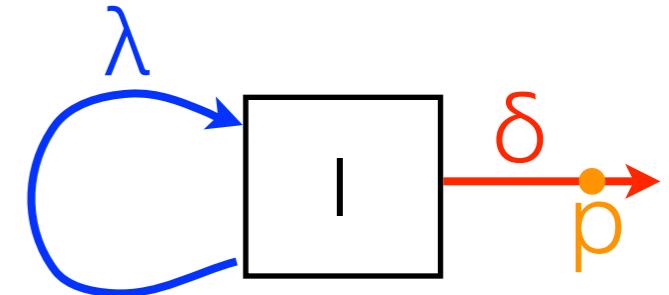


$p(t)$ : probability density that a given individual at time  $t$  produces an descending tree as observed.

$p(t_{\text{or}}) = \text{Lik}(\text{tree})$ : 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



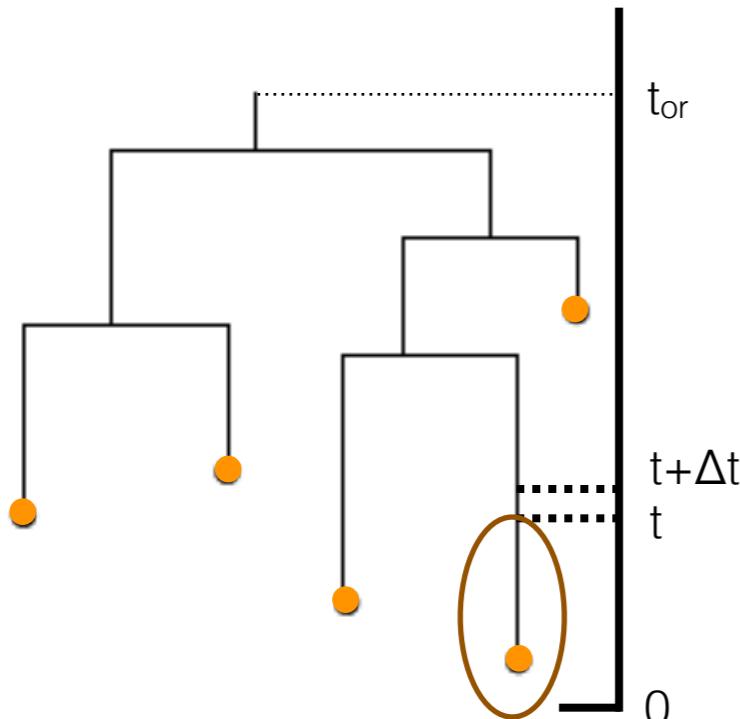
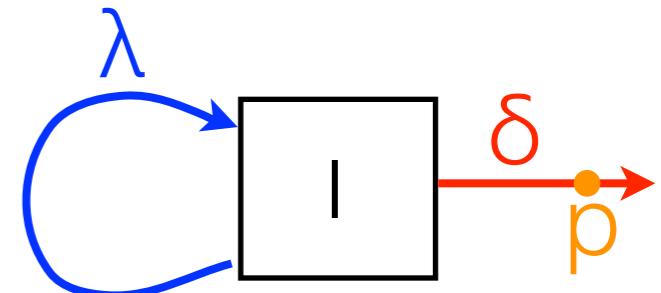
$p(t)$ : probability density that a given individual at time  $t$  produces an descending tree as observed.

$p(t_{\text{or}}) = \text{Lik}(\text{tree})$ : 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.

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

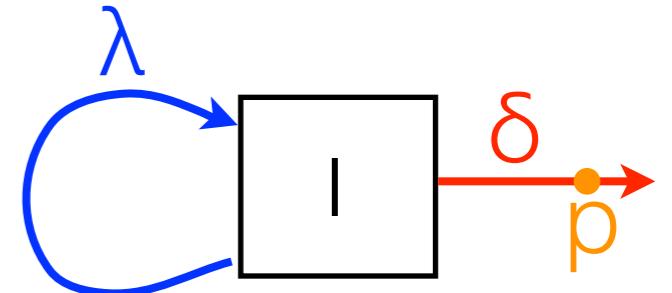
$p_0(t)$ : probability density that a given individual at time  $t$  has no sampled descendants.

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

$\Delta t \rightarrow 0$

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

# Tree likelihood



$$\text{Lik} \left( \begin{array}{c} \text{tree diagram} \\ \text{with orange dots at nodes} \end{array} \right) = \lambda^{m-1} \prod_{i=0}^{m-1} \frac{1}{q(x_i)} \prod_{i=1}^m p \delta q(y_i)$$

where

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

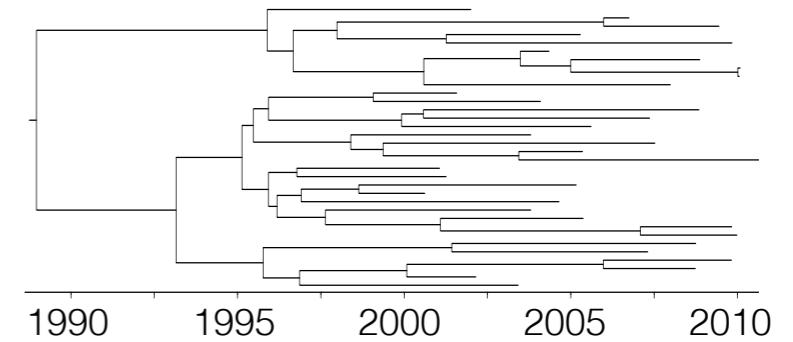
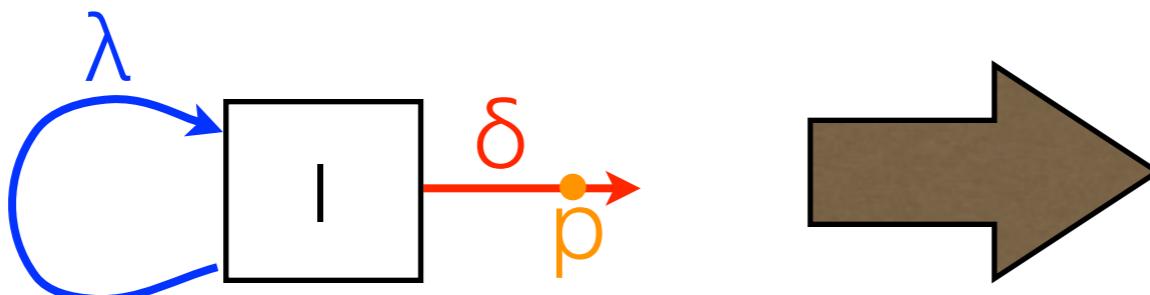
$$c_1 = |\sqrt{(\lambda - \delta)^2} + 4p\lambda\delta|$$

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

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

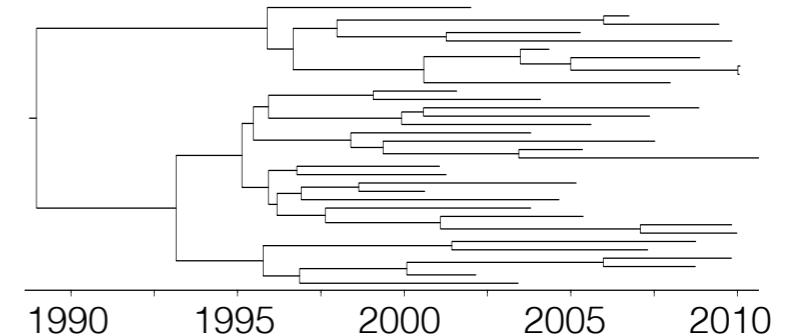
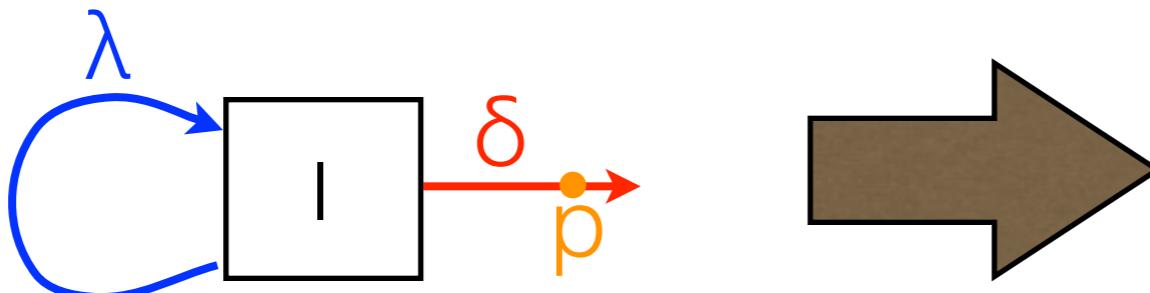
# Simulation study: Comparing exact method to coalescent

Simulating  
epidemic  
outbreak



# Simulation study: Comparing exact method to coalescent

Simulating  
epidemic  
outbreak



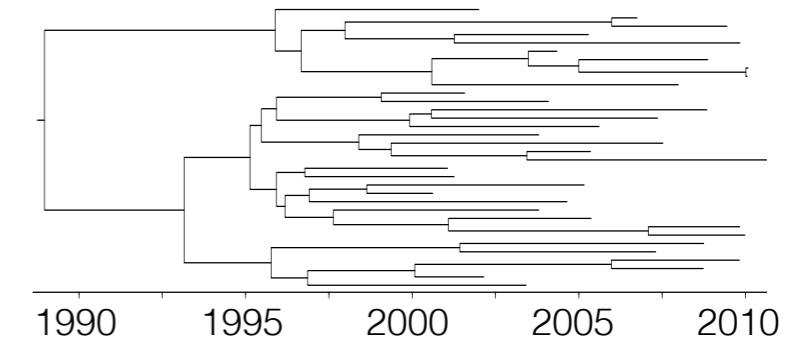
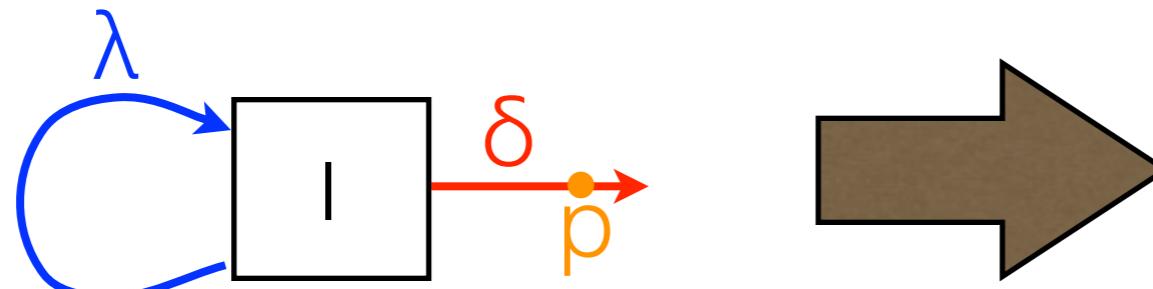
Exact  
method

$\lambda - \delta$  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...

## 1. ...for analyzing epidemic outbreaks

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

## 2. ...for identifying epidemiological changes over time

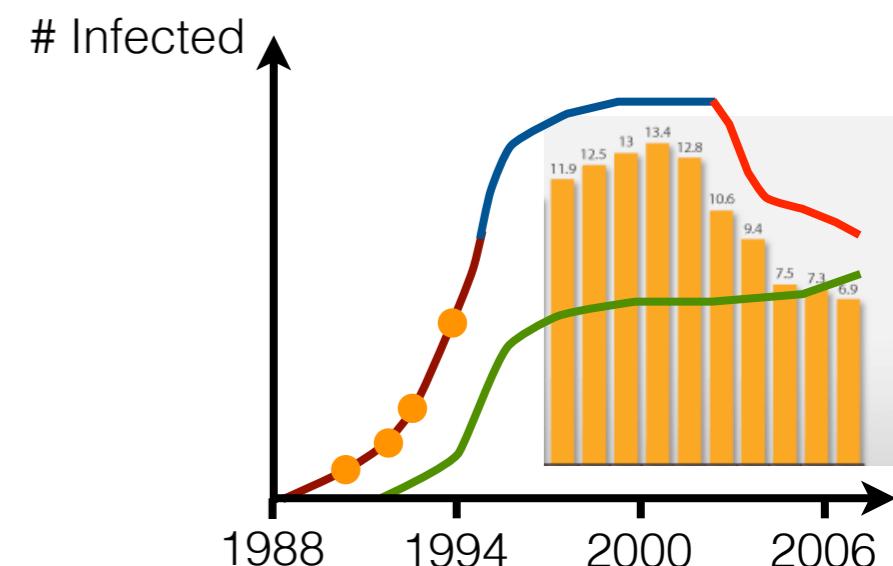
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

## 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 host population structure

- ▶ C: Volz (*Genetics*, 2012)
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013)



# Phylogenetic methods...

## 1. ...for analyzing epidemic outbreaks

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

## 2. ...for identifying epidemiological changes over time

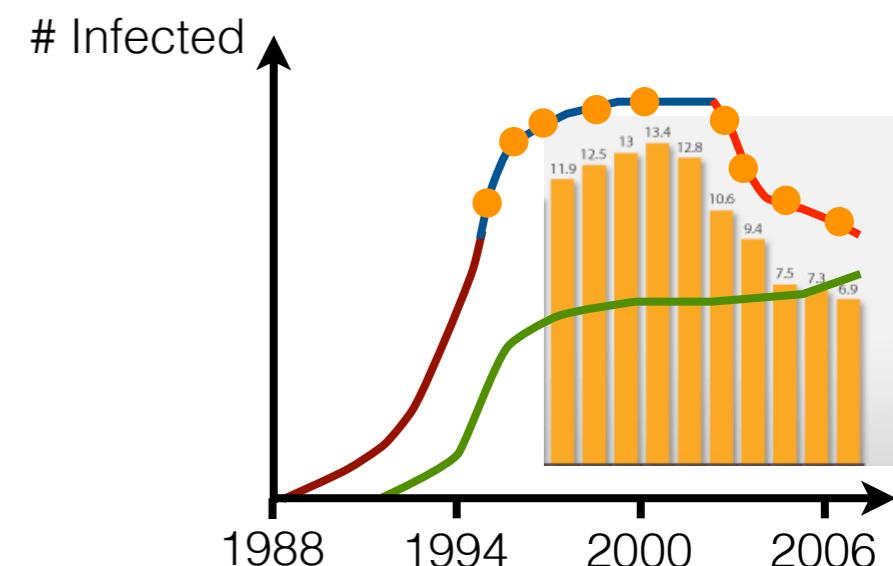
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

## 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 host population structure

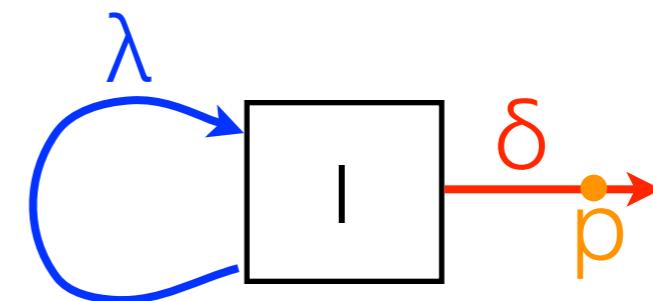
- ▶ C: Volz (*Genetics*, 2012)
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013)



# Birth-death-skyline plot

---

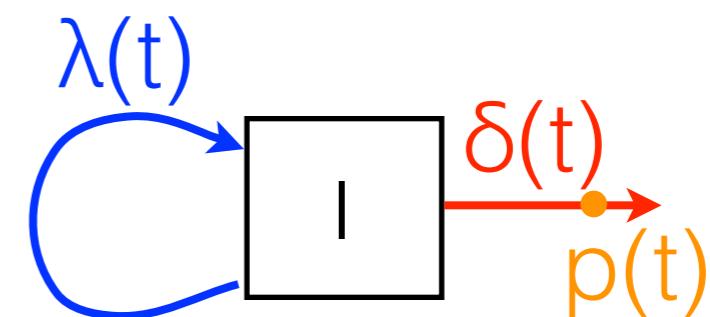
Epidemiological rates may change through time:



# Birth-death-skyline plot

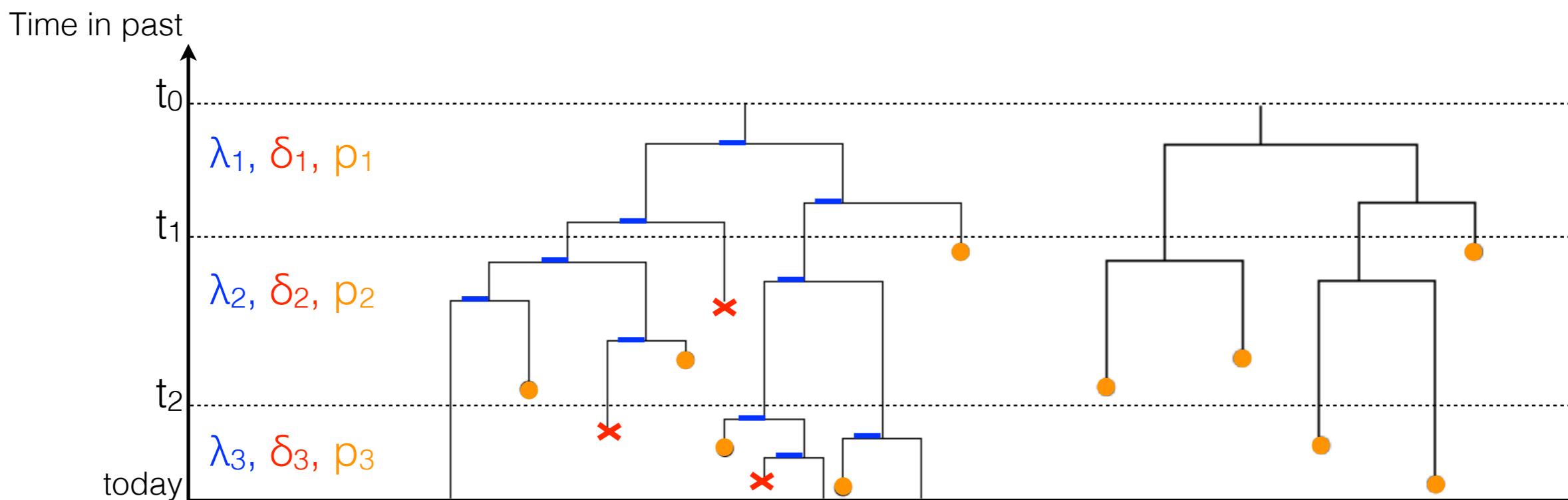
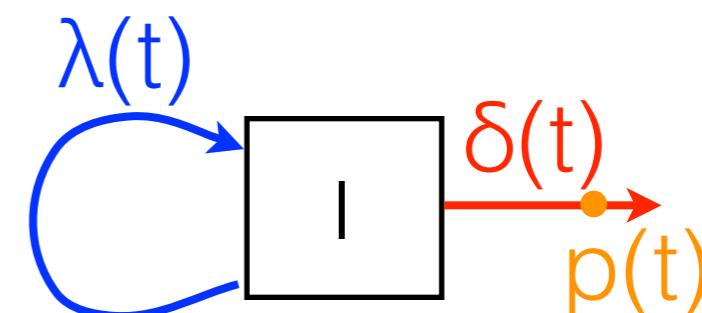
---

Epidemiological rates may change through time:

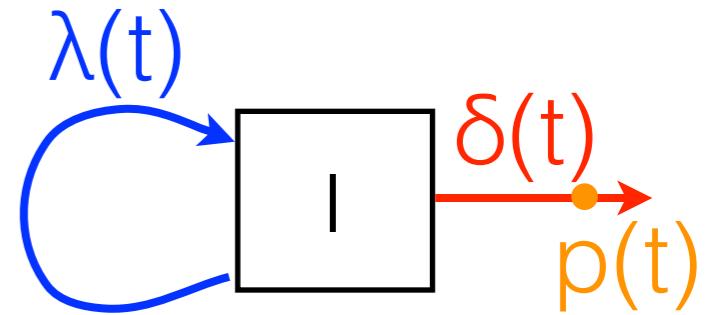


# Birth-death-skyline plot

Epidemiological rates may change through time:



# Hepatitis C virus in Egypt

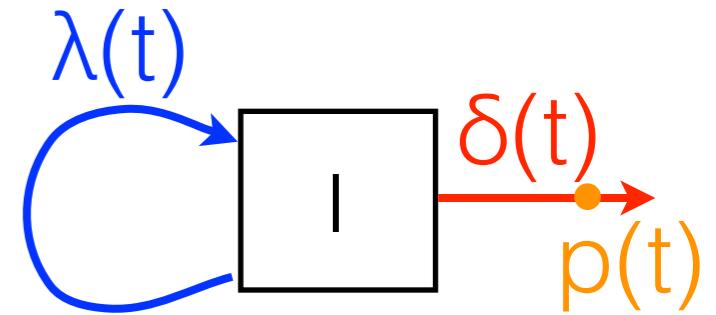


Effective reproductive number  $R_e(t)$

generalizes

Basic reproductive number  $R_0 = R_e(0)$

# Hepatitis C virus in Egypt



Effective reproductive number  $R_e(t)$   
generalizes

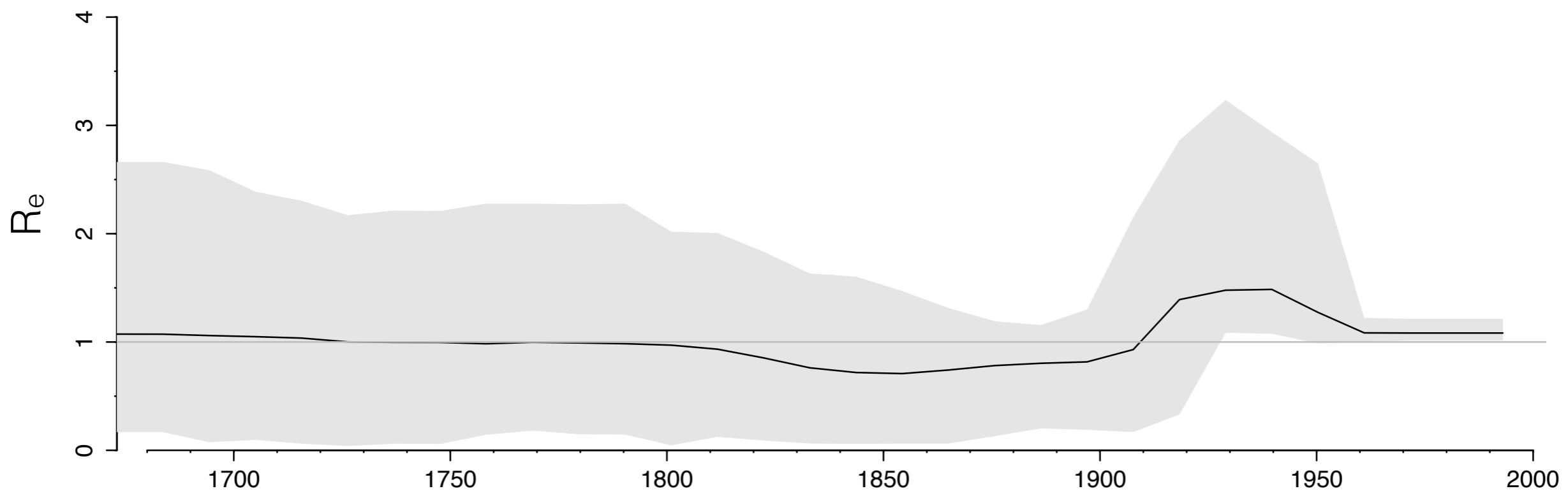
Basic reproductive number  $R_0 = R_e(0)$

Analysis of 63 seq:

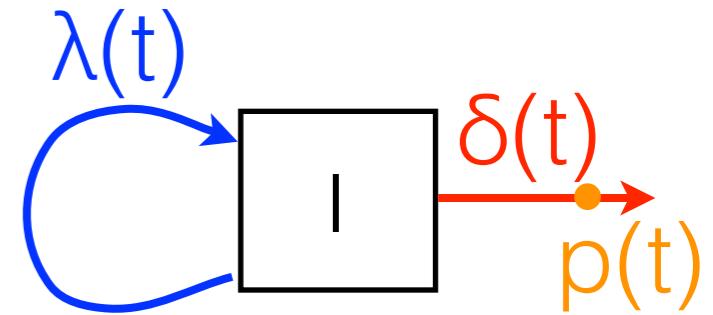
$$R_e(t) = \lambda(t)/\delta(t)$$

median

95% HPD interval



# Hepatitis C virus in Egypt



Effective reproductive number  $R_e(t)$   
generalizes

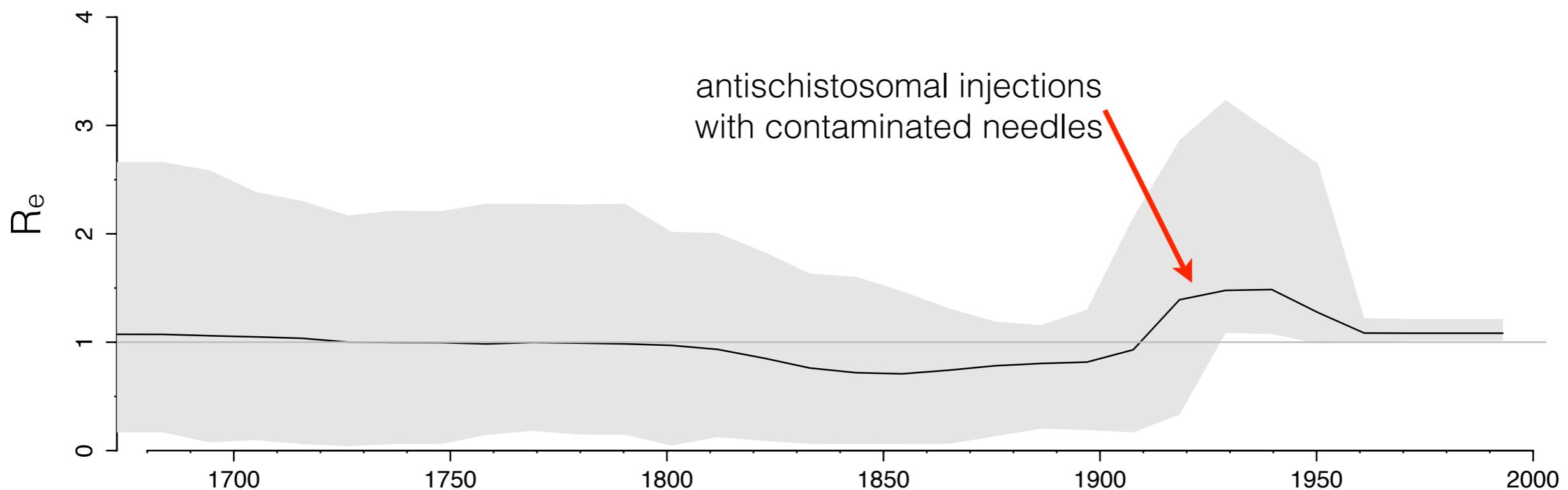
Basic reproductive number  $R_0 = R_e(0)$

Analysis of 63 seq:

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

median

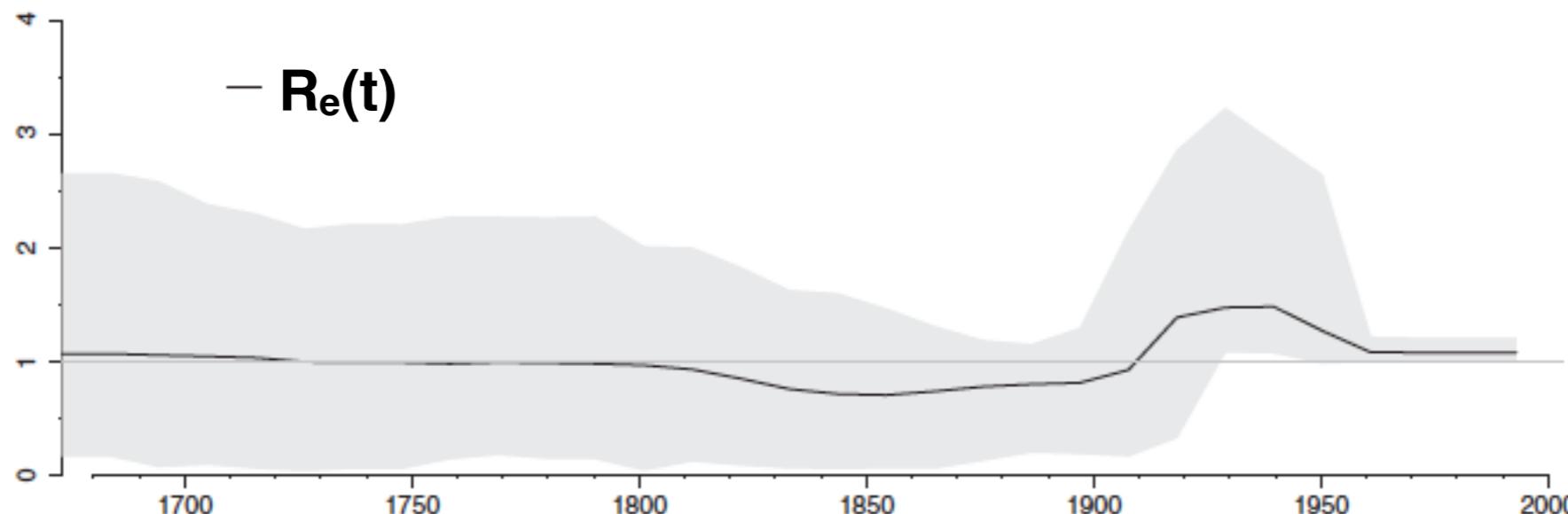
95% HPD interval



# The two skyline models

---

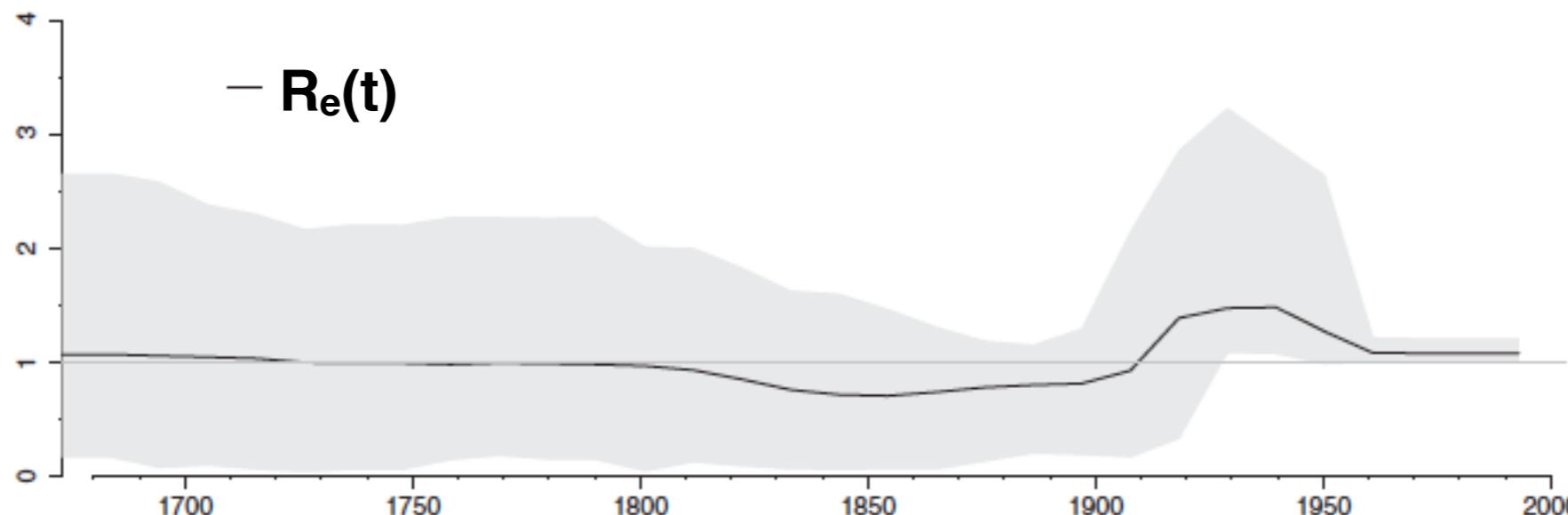
Birth-death skyline plot: effective reproductive number



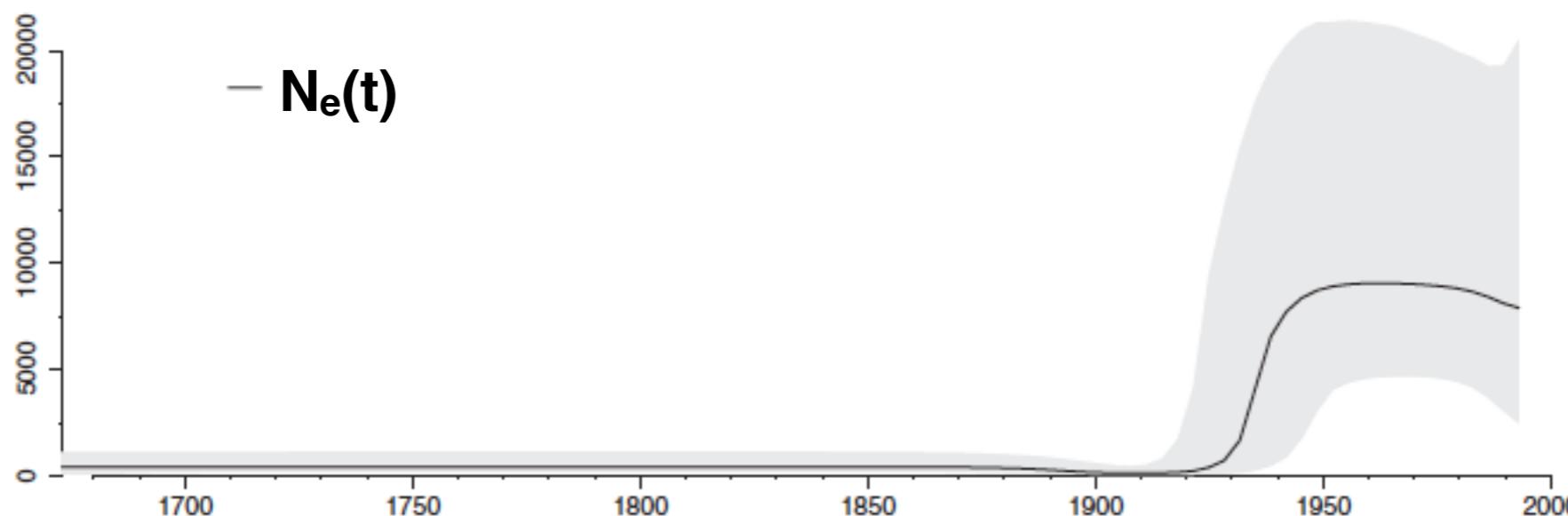
# The two skyline models

---

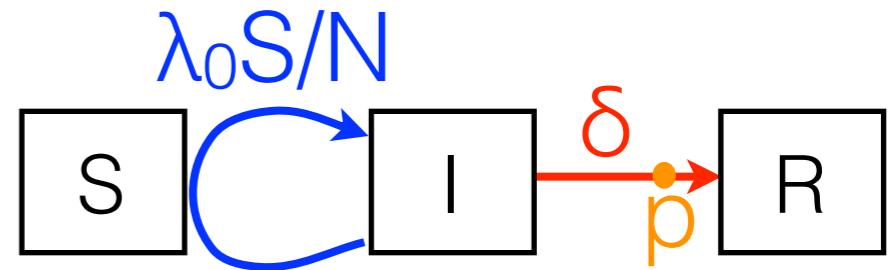
Birth-death skyline plot: effective reproductive number



Coalescent skyline plot: effective population size



### 3) SIR models



#### SIR simulation (trees with 200 tips sampled):

population size:  $N=5000$ , start with one infected,  $N=S+I+R$

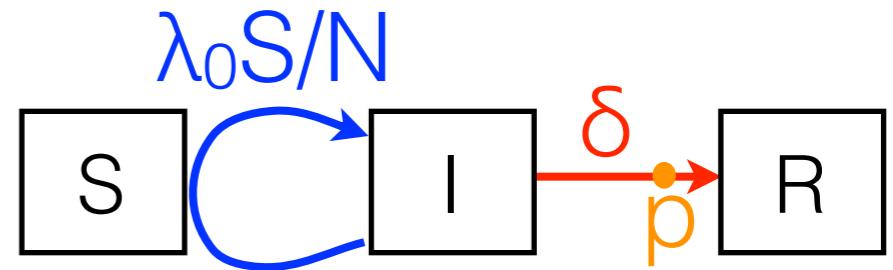
$$\lambda(S) = \lambda_0 S/N$$

$$R_0 = \lambda_0/\delta = 4$$

host recovery rate:  $\delta = 0.5$

sampling probability:  $p = 0.05$

### 3) SIR models



#### SIR simulation (trees with 200 tips sampled):

population size:  $N=5000$ , start with one infected,  $N=S+I+R$

$$\lambda(S) = \lambda_0 S/N$$

$$R_0 = \lambda_0 / \delta = 4$$

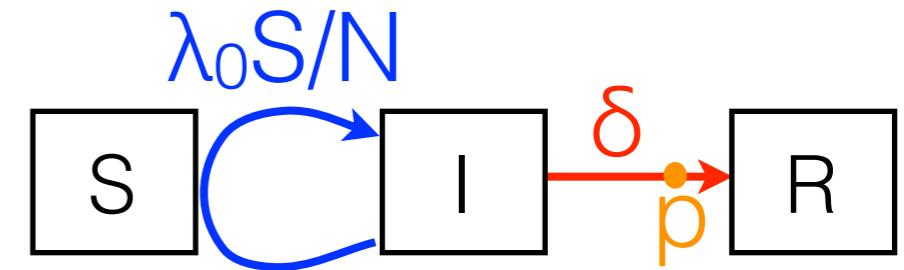
host recovery rate:  $\delta = 0.5$

sampling probability:  $p = 0.05$

**Inference:** 10 intervals for  $\lambda$ ,  $\delta$  and  $p$  constant across interval

meaning  $\lambda(S)=\lambda_0 S/N$  is approximated by  $\lambda(t)$

### 3) SIR models



#### SIR simulation (trees with 200 tips sampled):

population size:  $N=5000$ , start with one infected,  $N=S+I+R$

$$\lambda(S) = \lambda_0 S/N$$

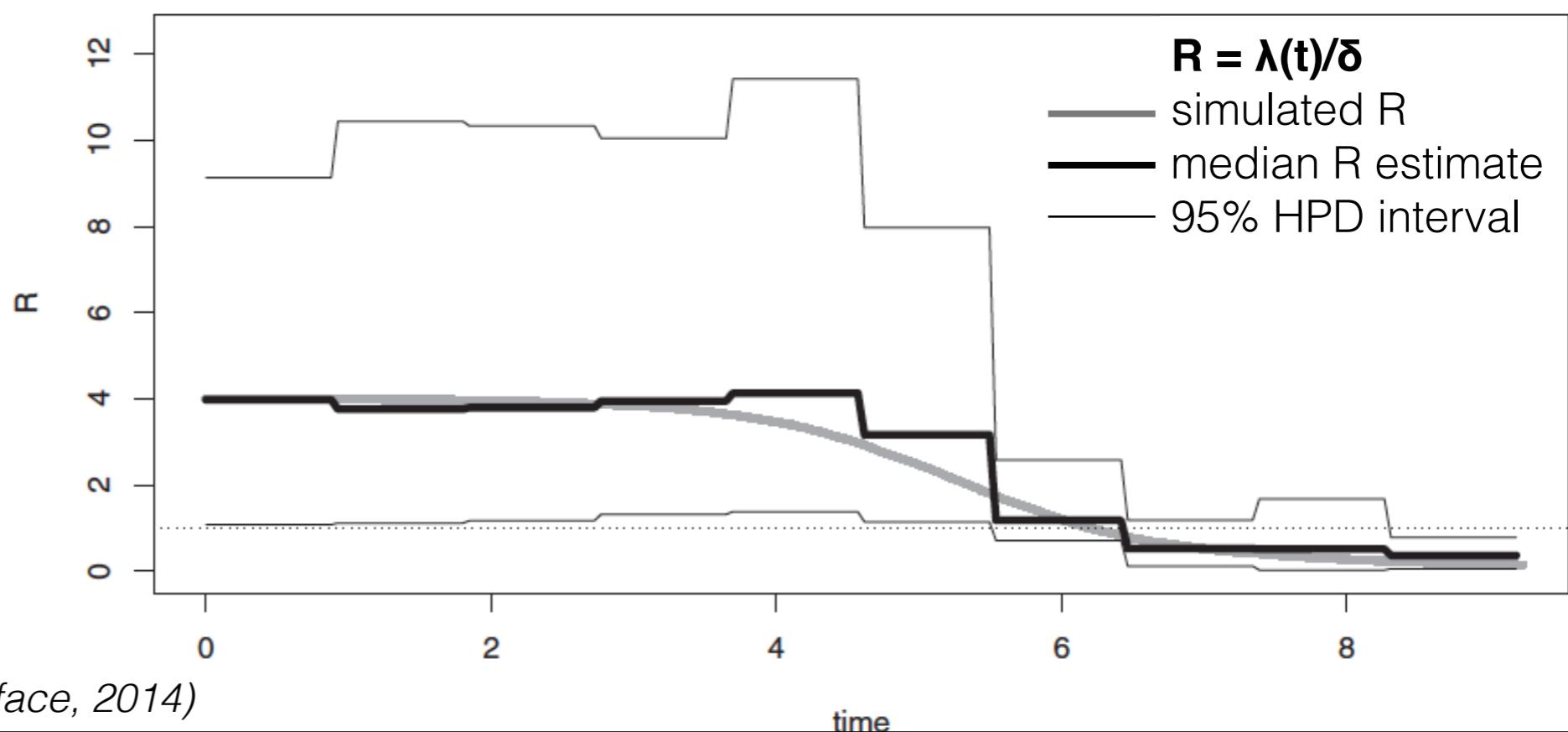
$$R_0 = \lambda_0/\delta = 4$$

host recovery rate:  $\delta = 0.5$

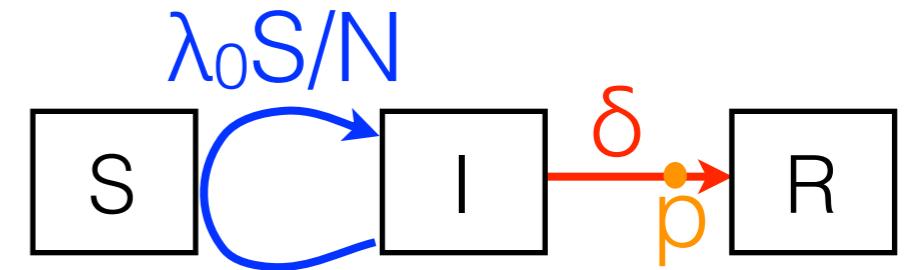
sampling probability:  $p = 0.05$

**Inference:** 10 intervals for  $\lambda$ ,  $\delta$  and  $p$  constant across interval

meaning  $\lambda(S)=\lambda_0 S/N$  is approximated by  $\lambda(t)$



### 3) SIR models



#### SIR simulation (trees with 200 tips sampled):

population size:  $N=5000$ , start with one infected,  $N=S+I+R$

$$\lambda(S) = \lambda_0 S/N$$

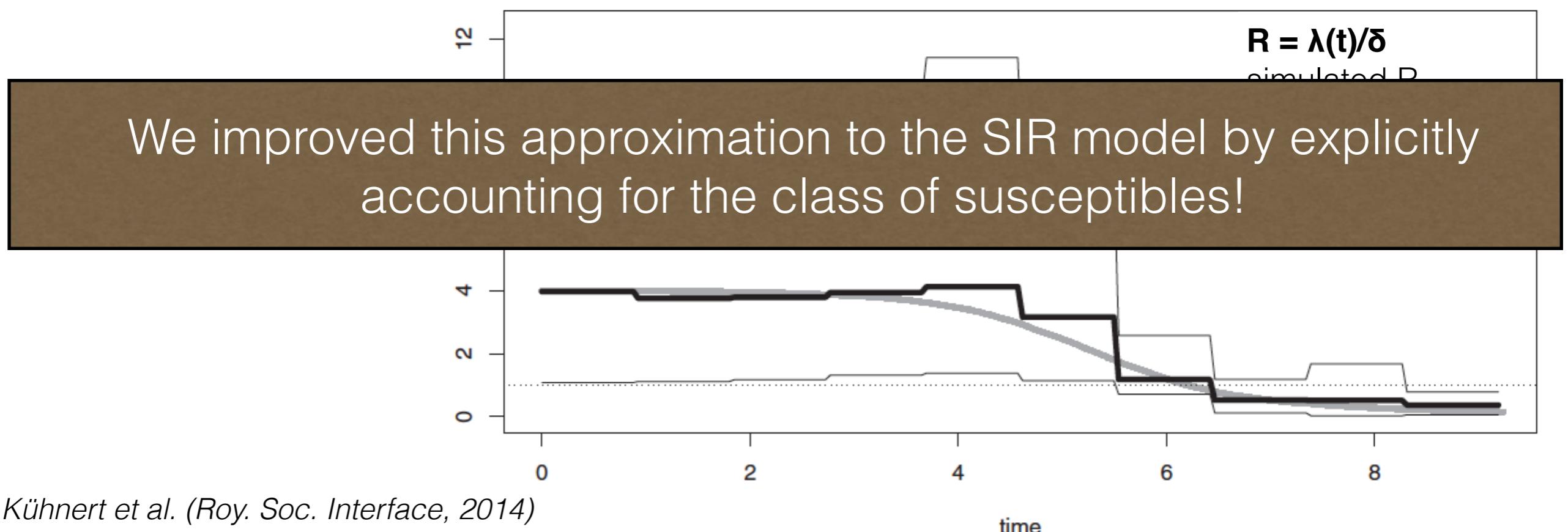
$$R_0 = \lambda_0/\delta = 4$$

host recovery rate:  $\delta = 0.5$

sampling probability:  $p = 0.05$

**Inference:** 10 intervals for  $\lambda$ ,  $\delta$  and  $p$  constant across interval

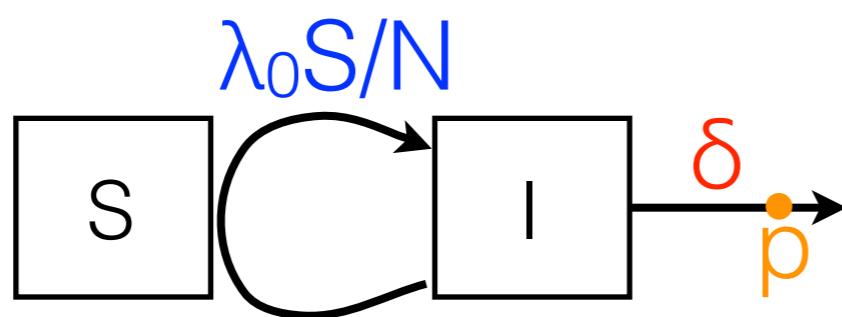
meaning  $\lambda(S)=\lambda_0 S/N$  is approximated by  $\lambda(t)$



# SI dynamics - exact method

---

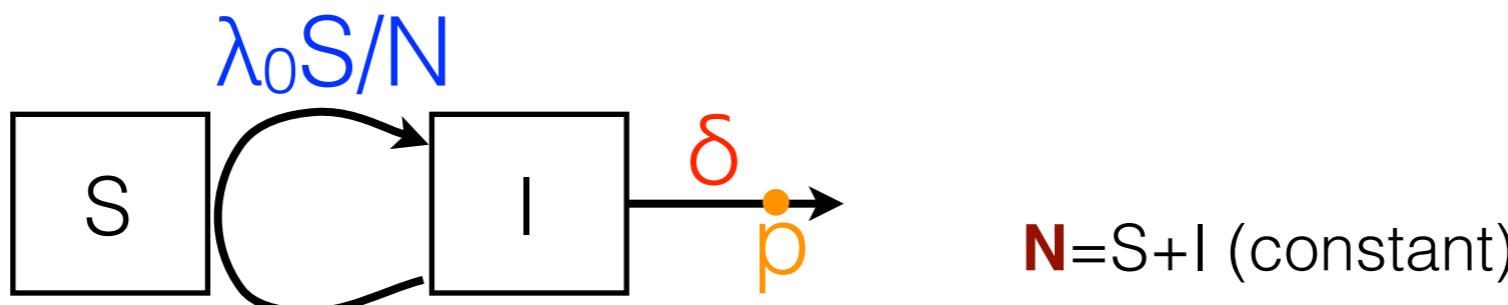
Para-  
meters



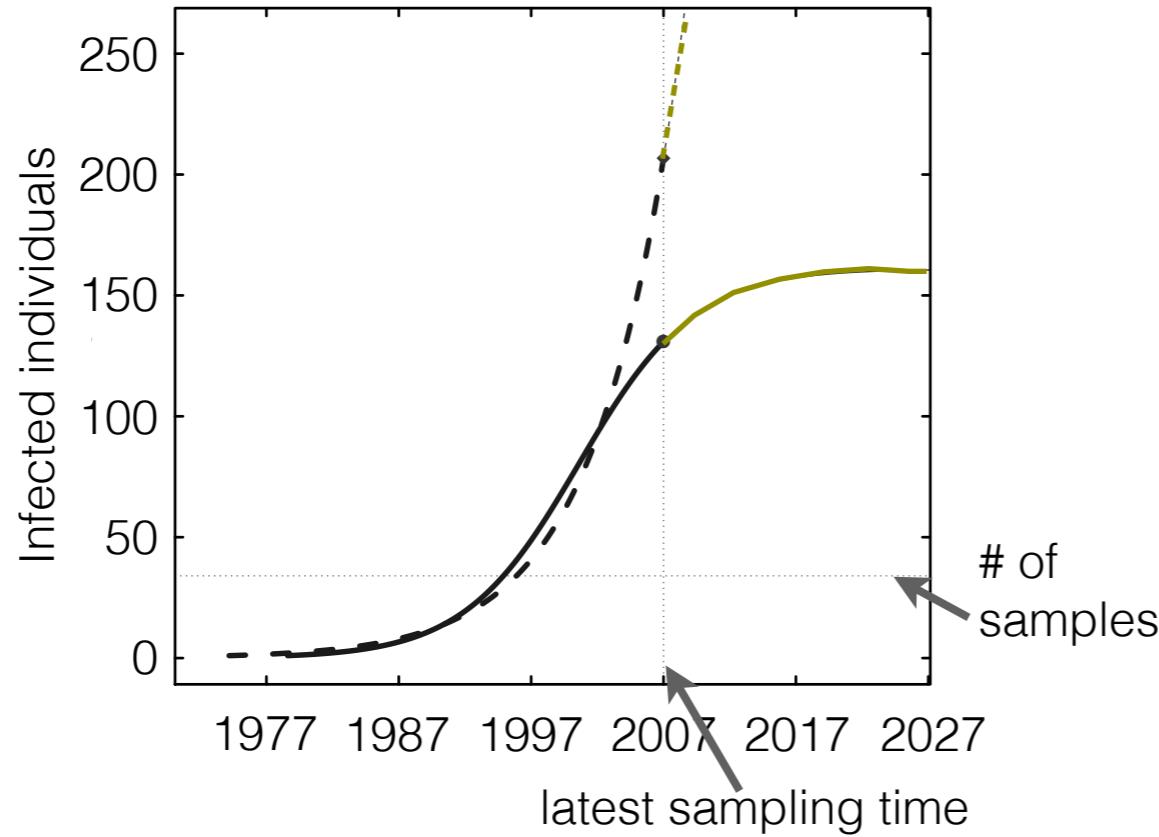
$$\mathbf{N} = S + I \text{ (constant)}$$

# SI dynamics - exact method

Para-  
meters

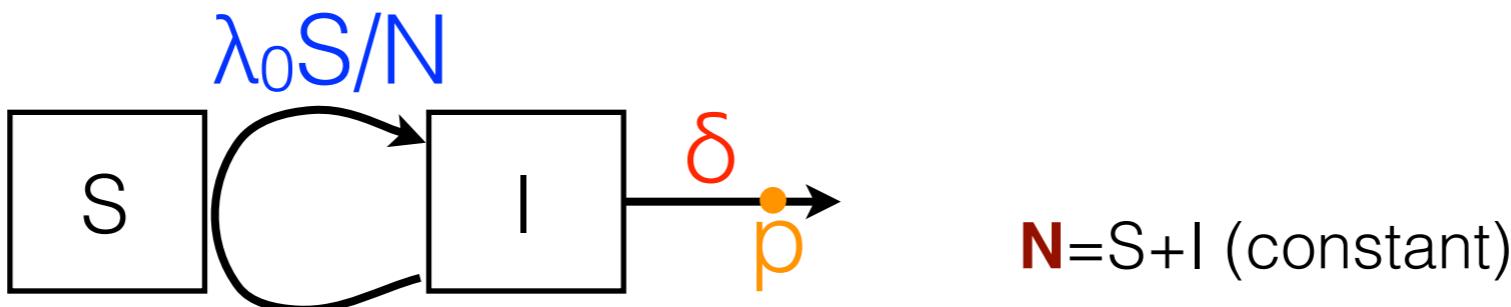


HIV  
trans-  
mission  
cluster  
in  
Switzer-  
land

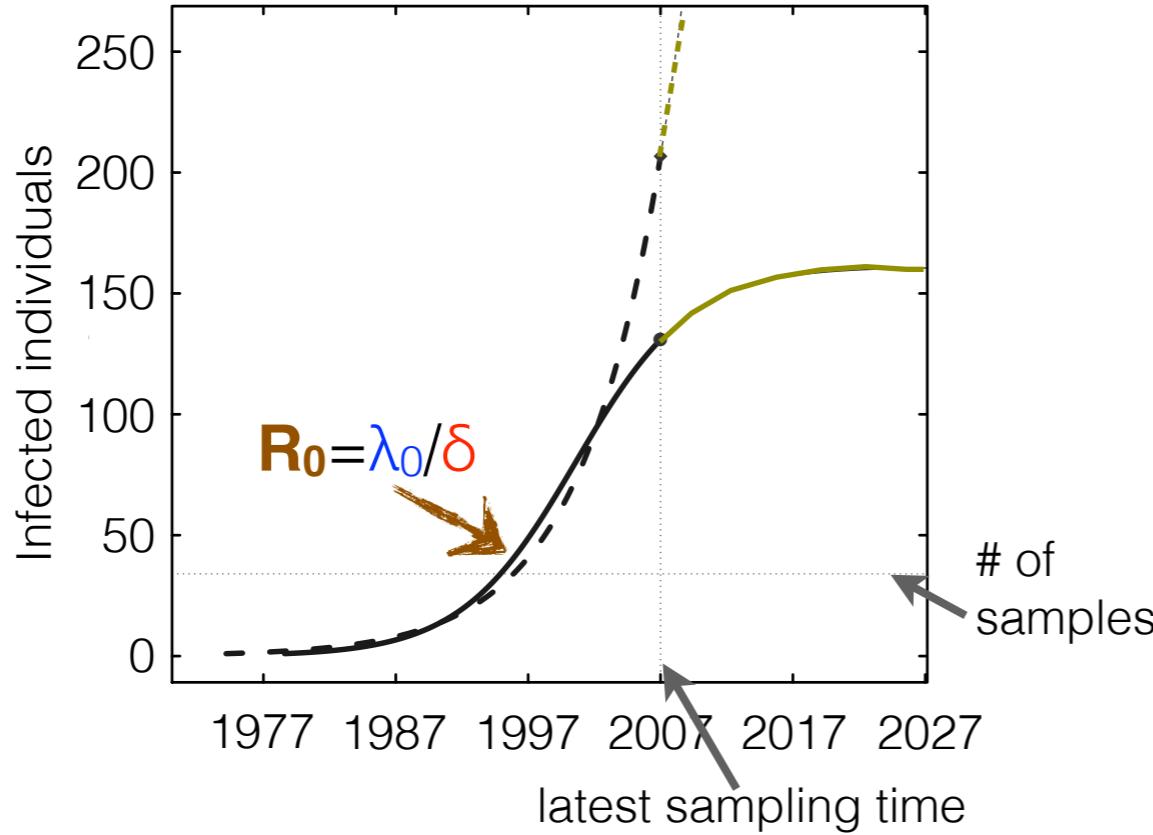


# SI dynamics - exact method

Para-meters



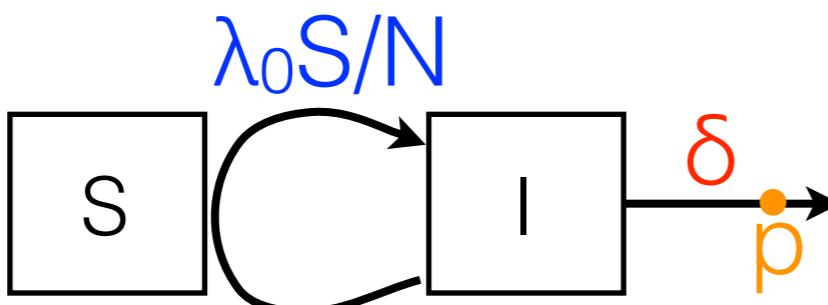
HIV transmission cluster in Switzerland



— SI model  
- - - epidemic outbreak model

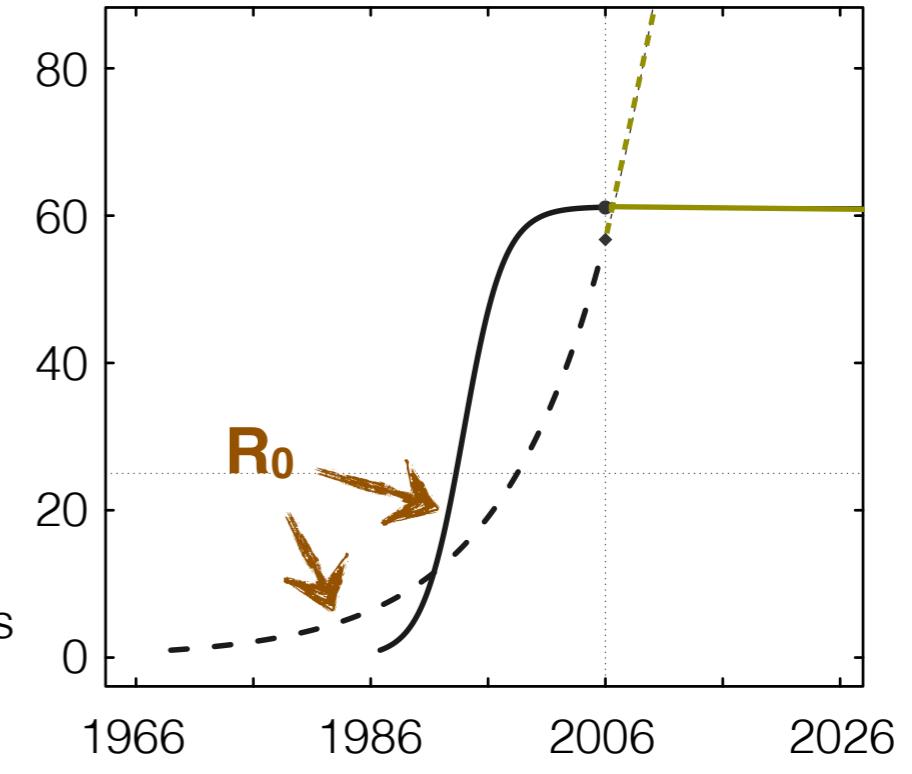
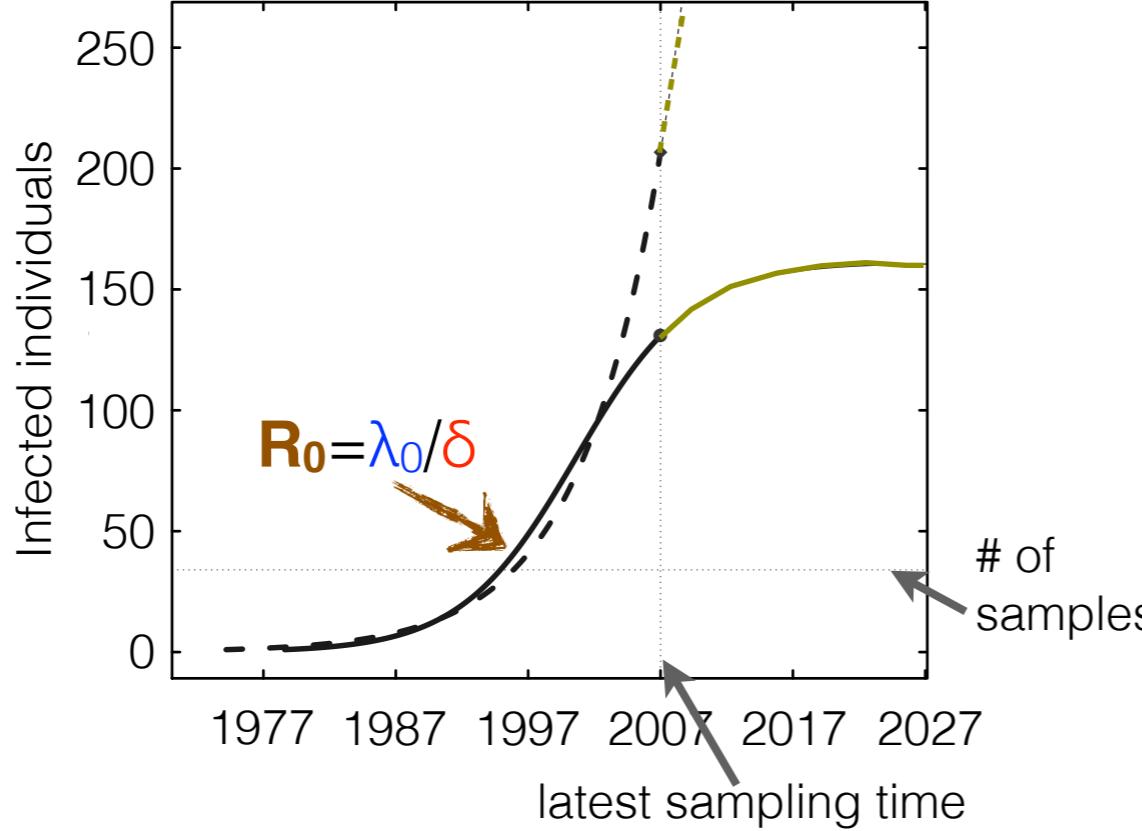
# SI dynamics - exact method

Para-meters



$$N = S + I \text{ (constant)}$$

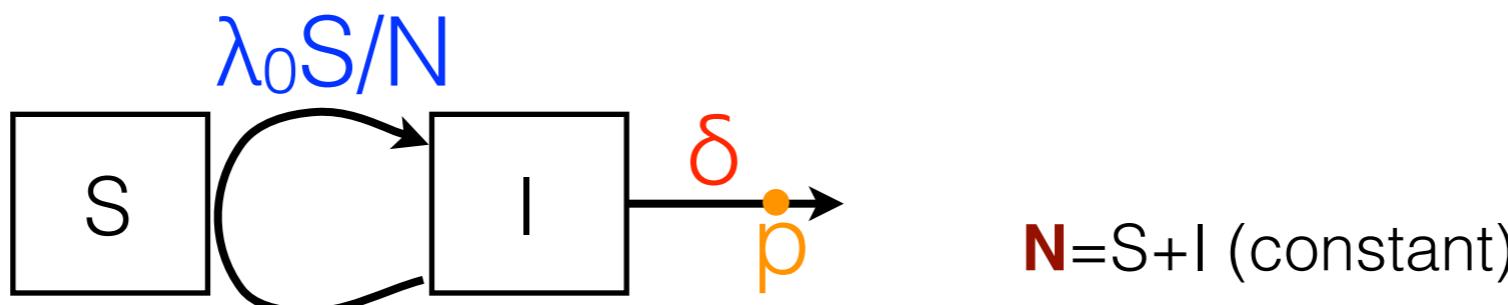
HIV transmission cluster in Switzerland



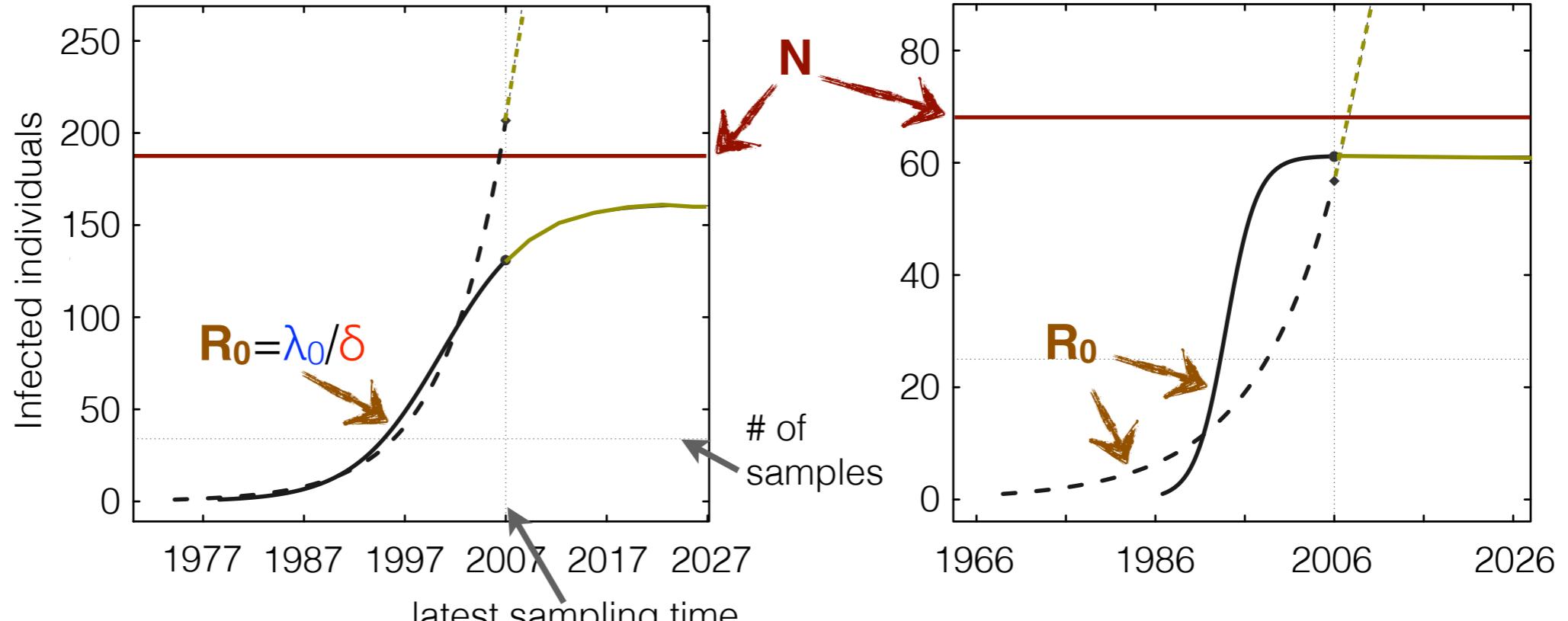
— SI model  
- - - epidemic outbreak model

# SI dynamics - exact method

Para-meters

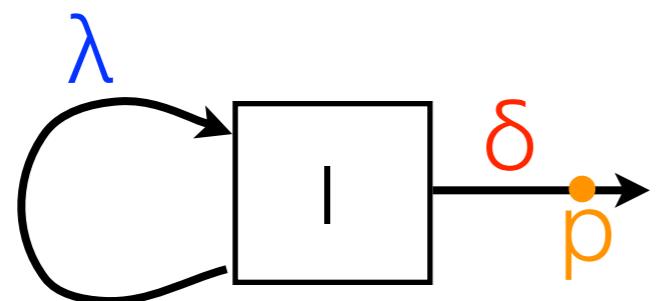


HIV transmission cluster in Switzerland



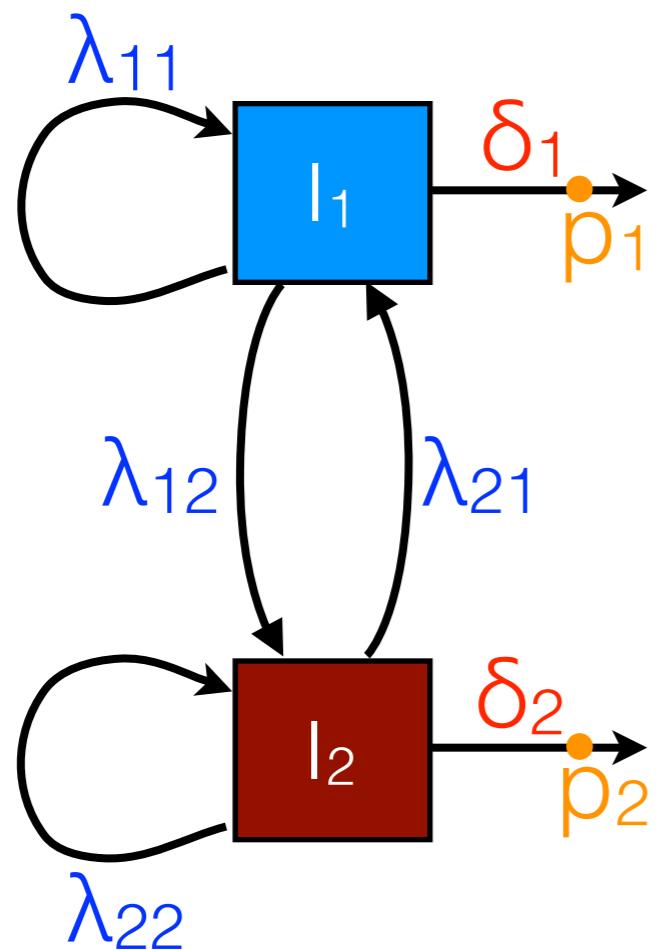
## 4) Host population structure

---



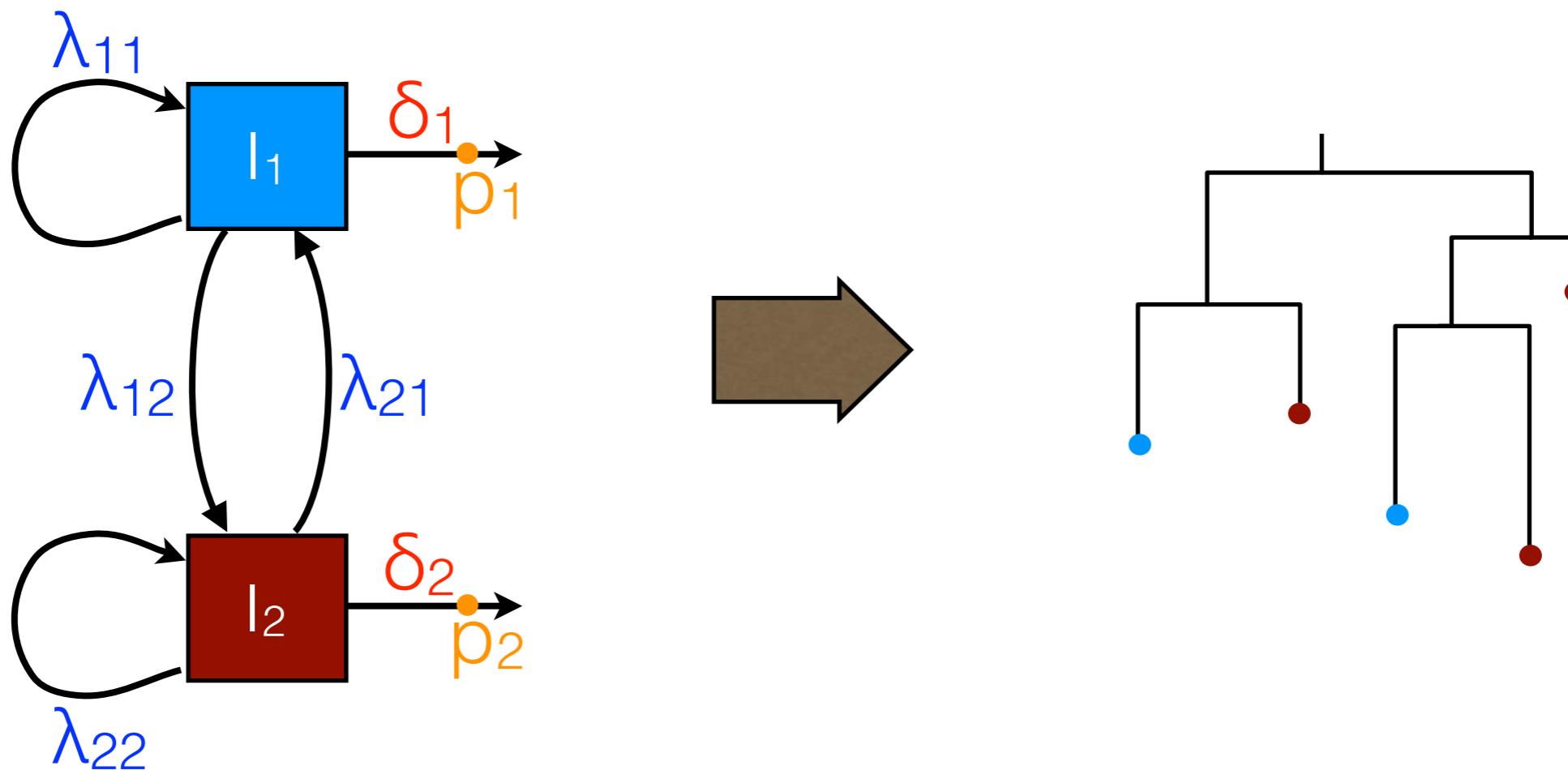
## 4) Host population structure

---



## 4) Host population structure

---



# Quantification of risk-group dependent spread of HIV

---

Two states

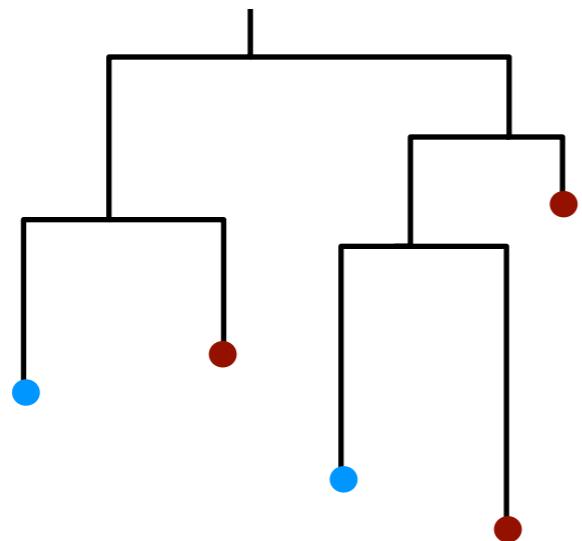
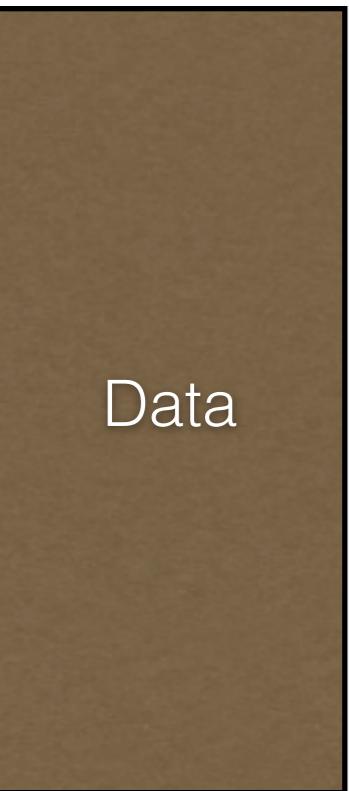
Heterosexual

Intravenous drug user

# Quantification of risk-group dependent spread of HIV



Heterosexual  
Intravenous drug user

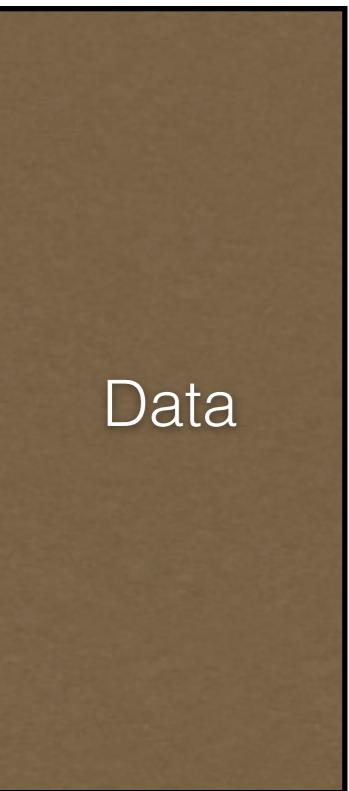


220 Latvian sequences from  
Balode et al. (2012)

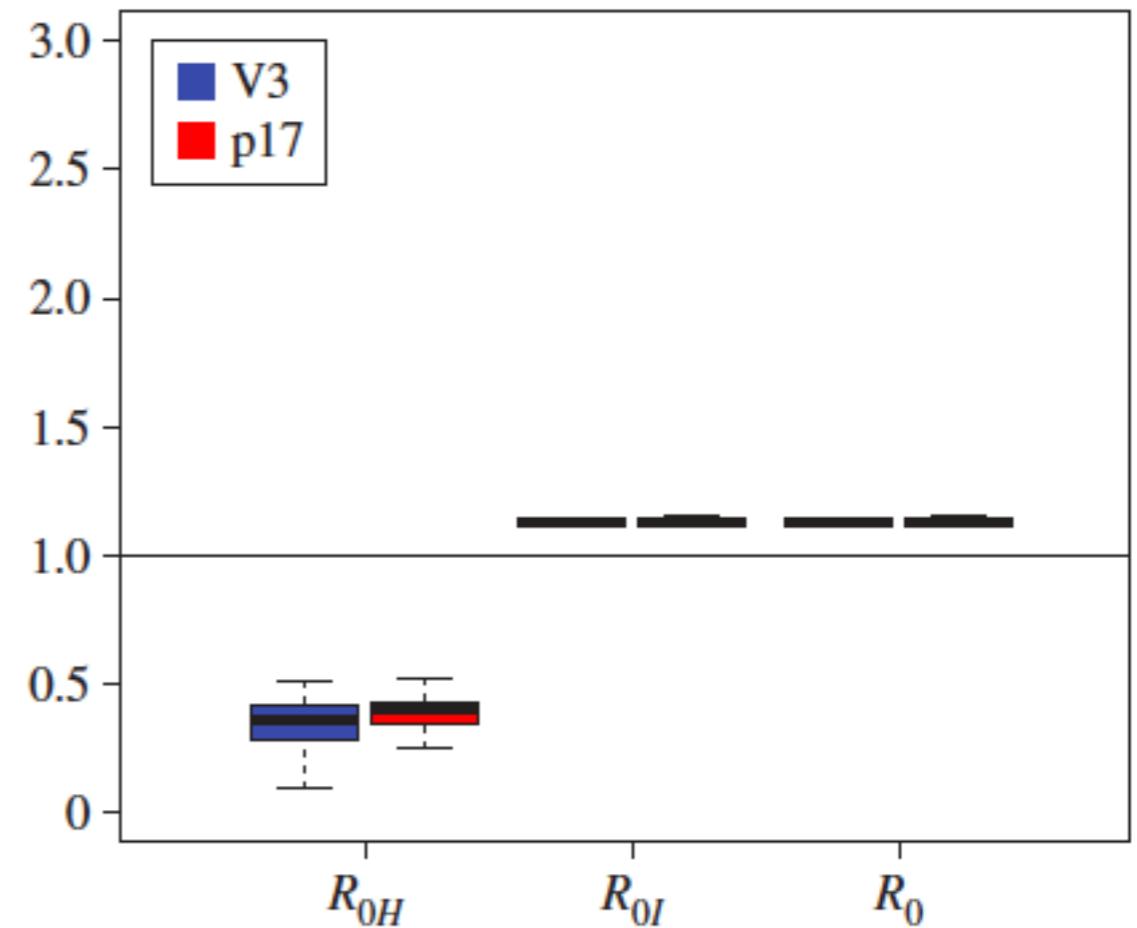
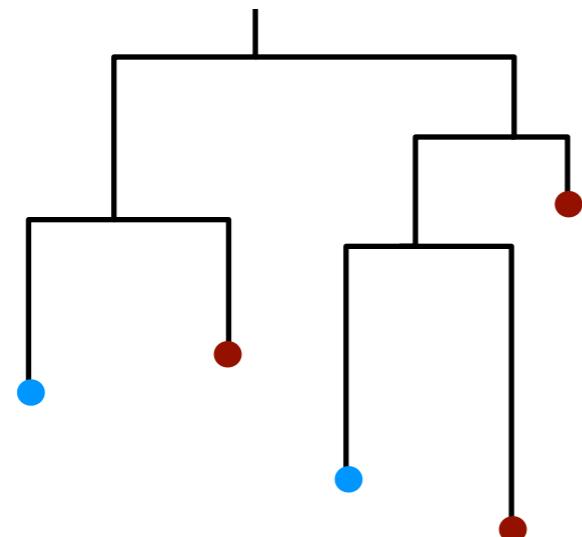
# Quantification of risk-group dependent spread of HIV



Heterosexual  
Intravenous drug user



220 Latvian sequences from  
Balode et al. (2012)



# Quantification of super-spreading in HIV

---



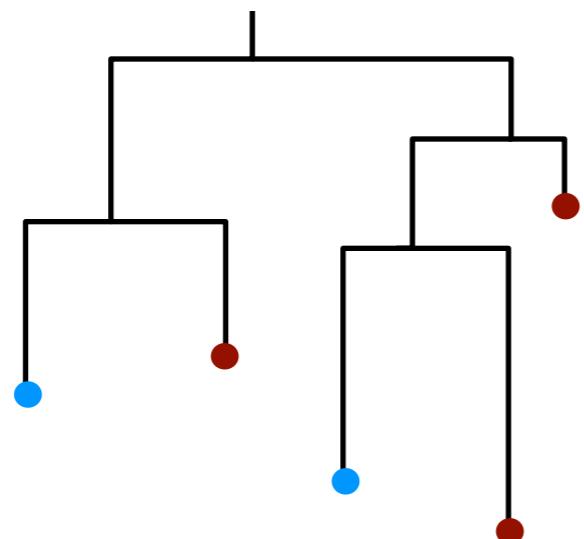
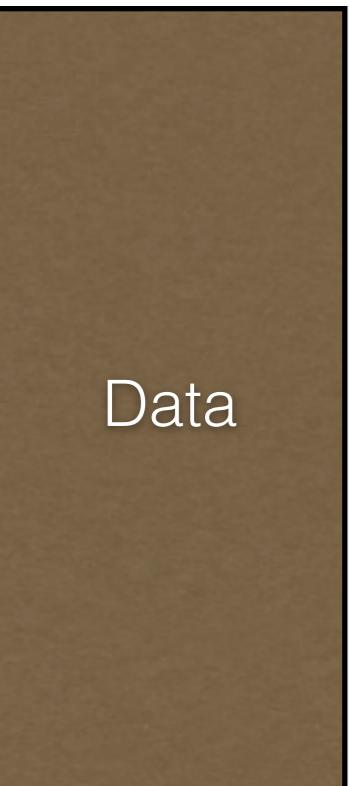
**Super-spreader**

**Normal-spreader**

# Quantification of super-spreading in HIV



**Super-spreader**  
**Normal-spreader**

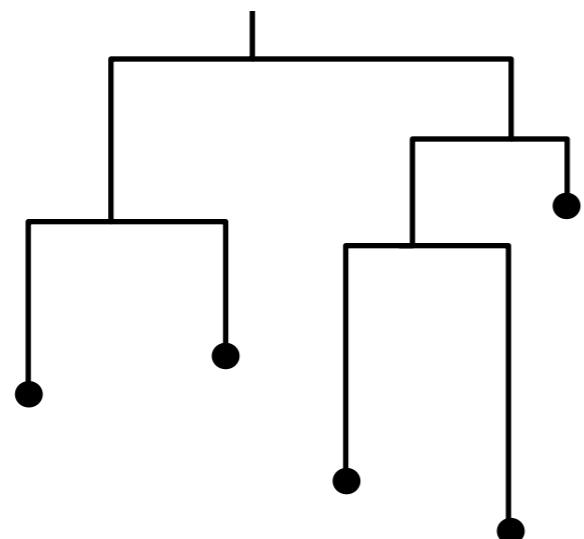
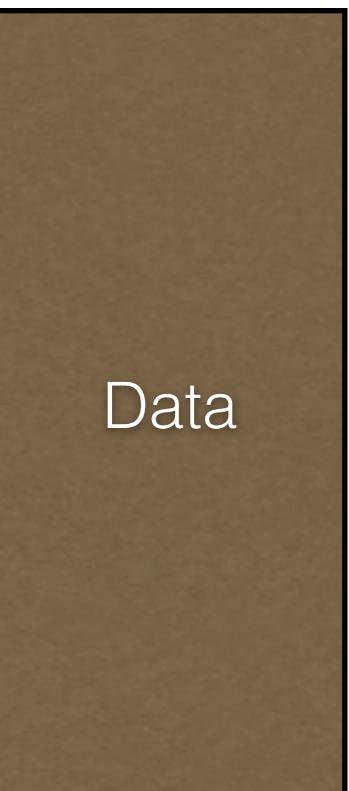


40 Latvian sequences from  
Balode et al. (2012)

# Quantification of super-spreading in HIV



**Super-spreader**  
**Normal-spreader**



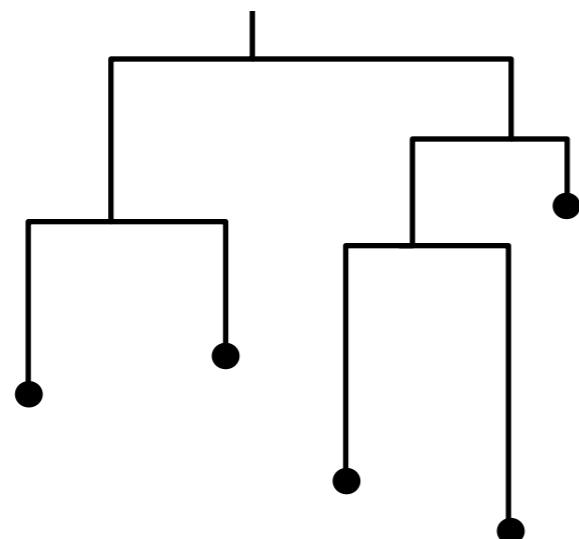
40 Latvian sequences from  
Balode et al. (2012)

# Quantification of super-spreading in HIV

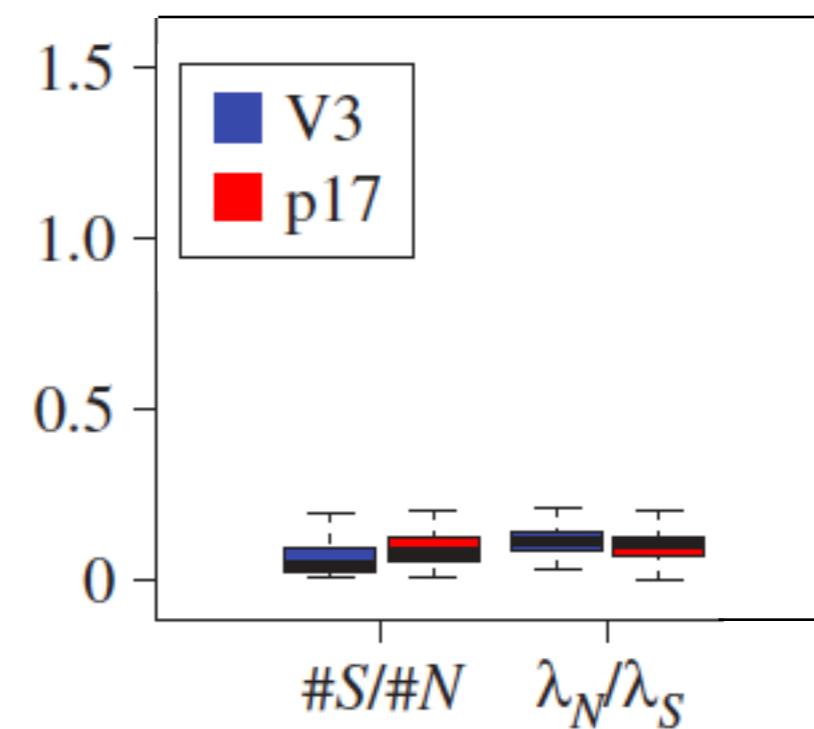
Two states

Super-spreader  
Normal-spreader

Data

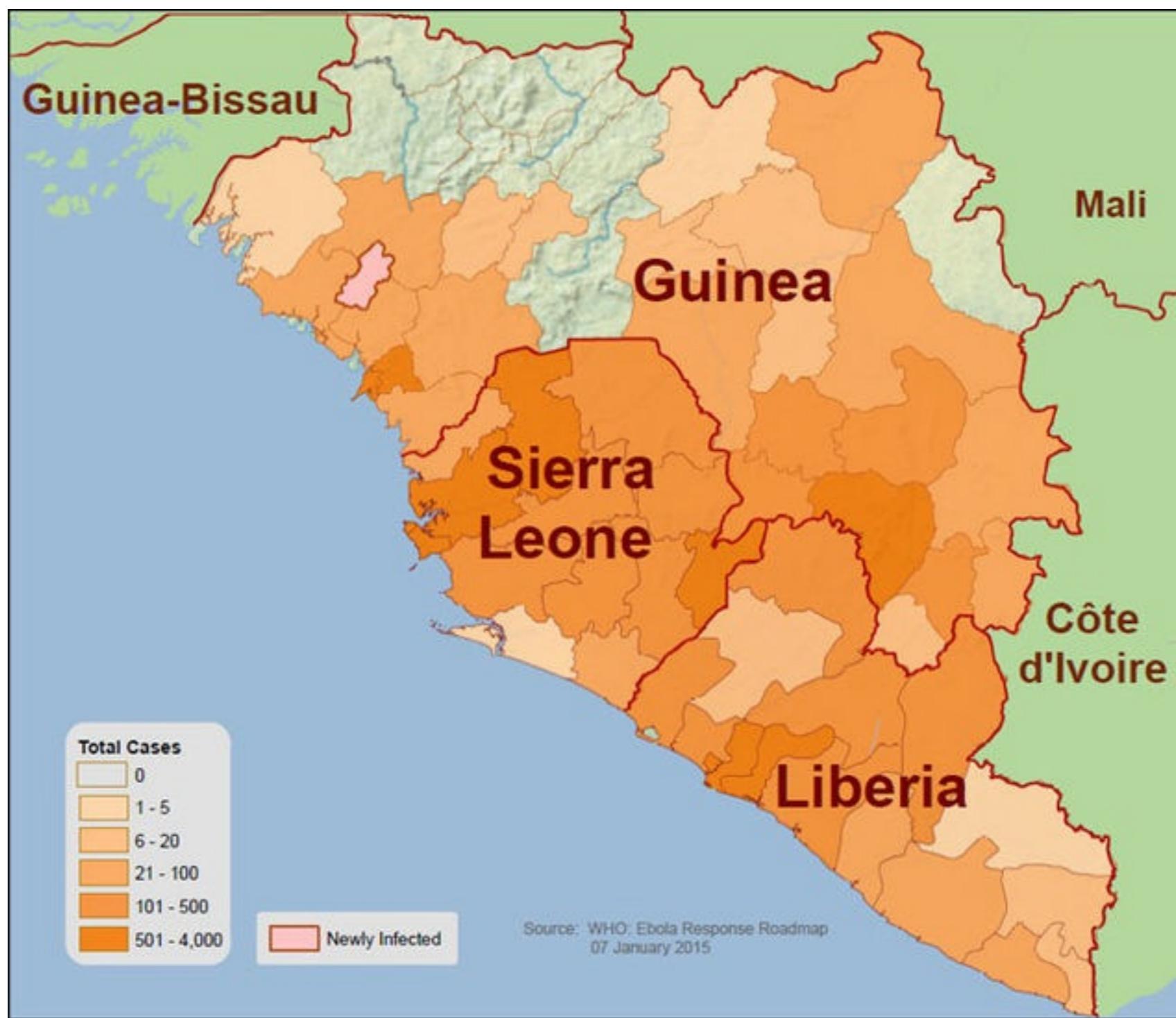


40 Latvian sequences from  
Balode et al. (2012)



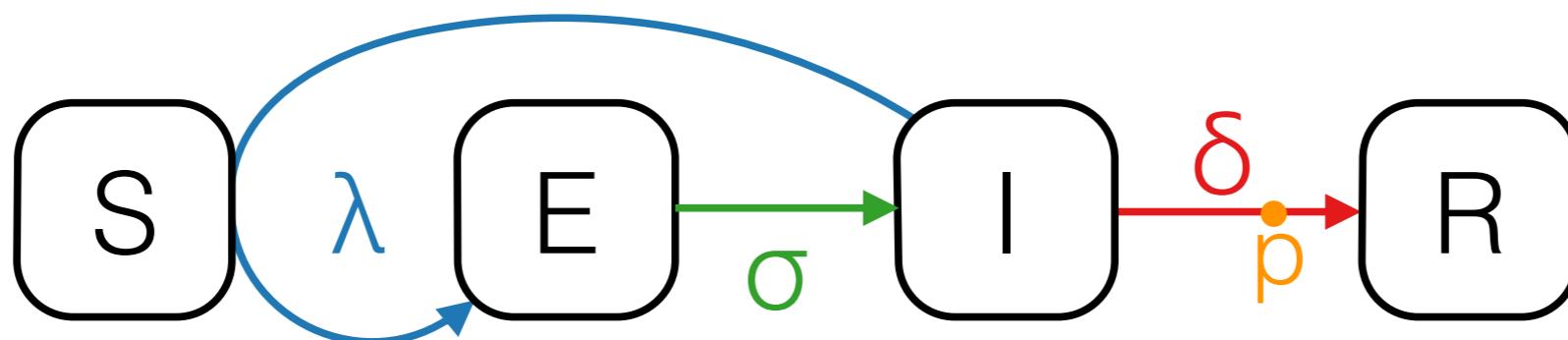
# Ebola in West Africa

---



# Epidemiological model for Ebola

---

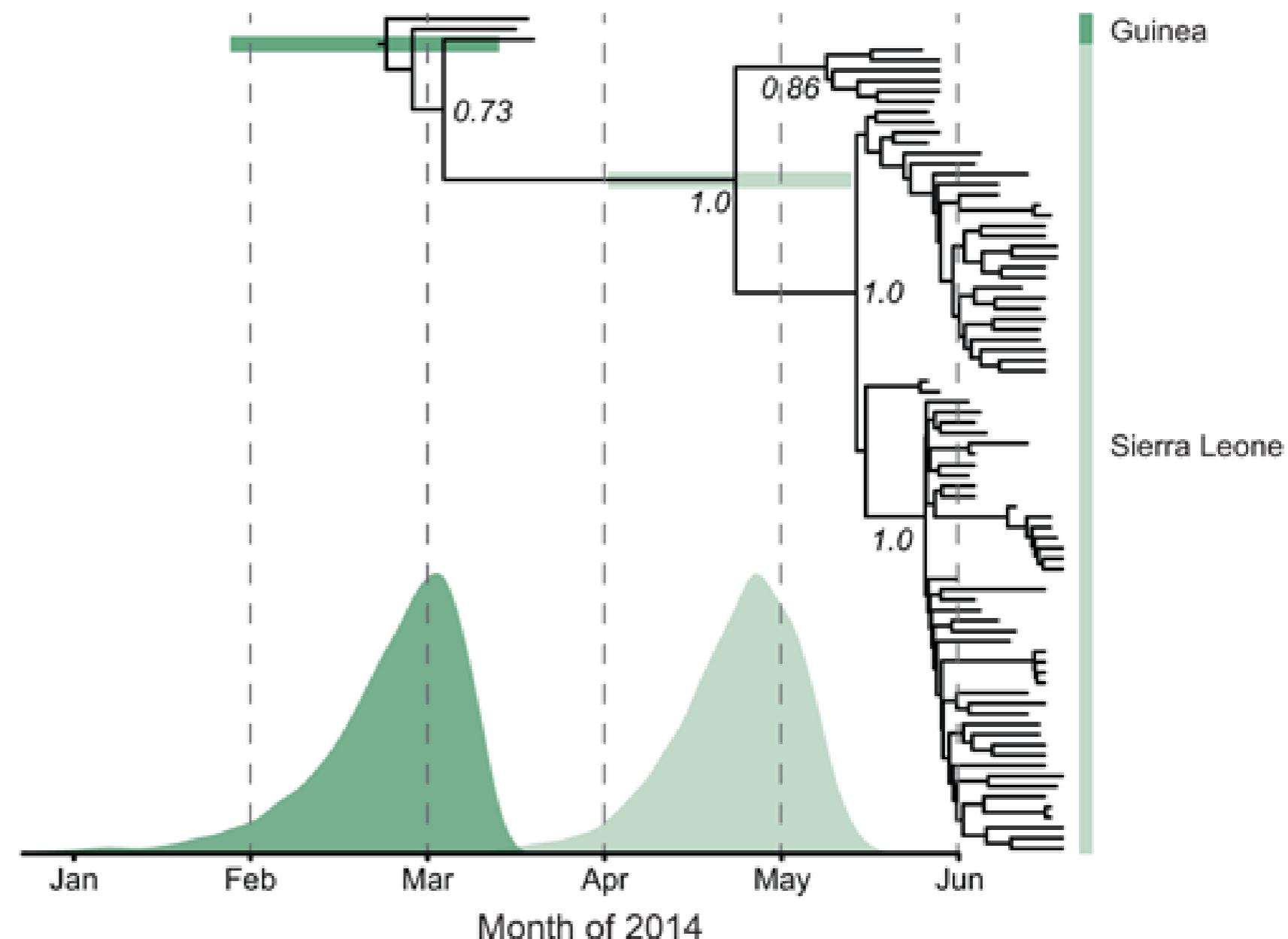


# Single zoonosis in December caused the current epidemic



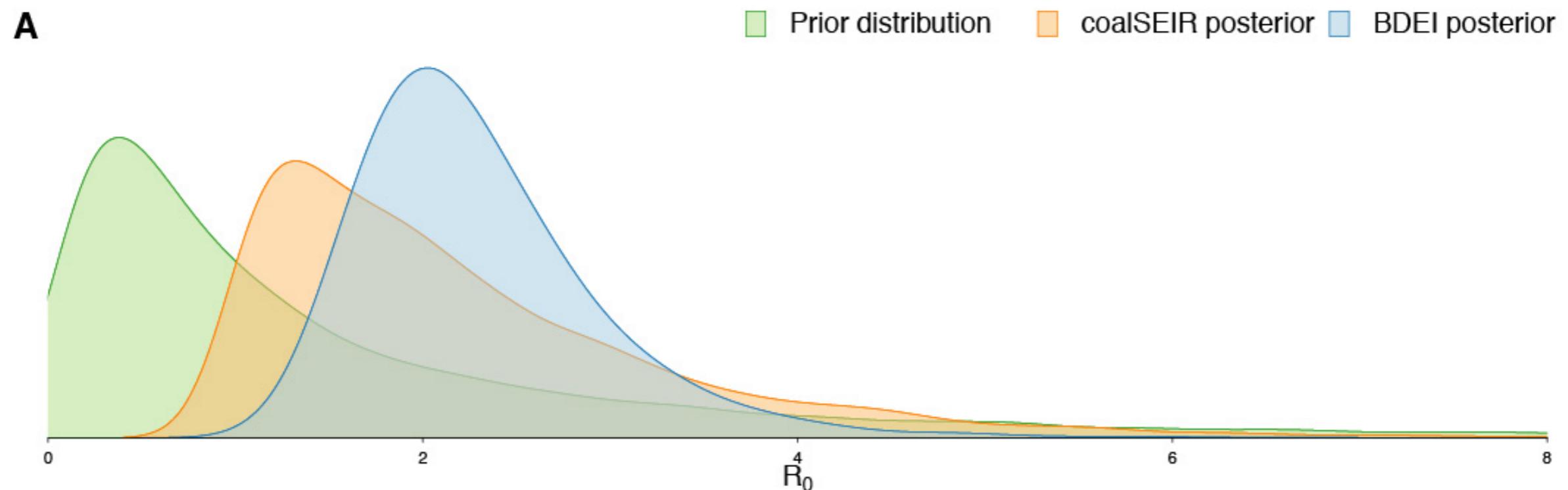
## Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak

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 Gnirke,<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>

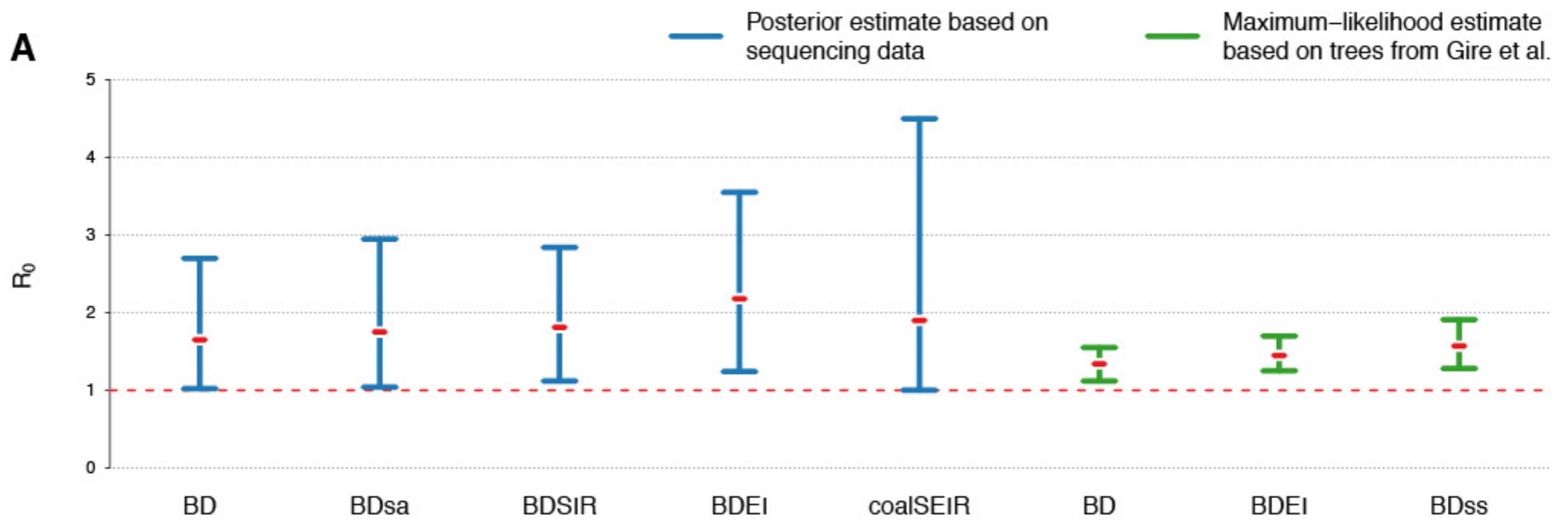


# Estimation of epidemic spread for Sierra Leone (72 genomes)

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



# Estimation of epidemic spread for Sierra Leone (72 genomes)



# Estimation of epidemic spread for Sierra Leone

---

Infected periods

Expected incubation period 4.92 days (1.37-11.54)

Expected infectious period 2.58 days (1.00-5.37)

# Estimation of epidemic spread for Sierra Leone

---

Infected periods

Expected incubation period 4.92 days (1.37-11.54)

Expected infectious period 2.58 days (1.00-5.37)

Sampling proportion

65% (20%-100%)

- ▶ Confirms independent estimates based on incidences

# Estimation of epidemic spread for Sierra Leone

---

Infected periods

Expected incubation period 4.92 days (1.37-11.54)

Expected infectious period 2.58 days (1.00-5.37)

Sampling proportion

65% (20%-100%)

- ▶ Confirms independent estimates based on incidences

Population structure

Half the population transmits very much ( $R_0 > 2$ )

Other half barely transmits ( $R_0 < 1$ )

- ▶ Key knowledge to public health officials

# How to improve understanding of epidemic spread?

---

Epidemic today

We only have sequences up to June!

- ▶ *Thus we cannot quantify the reproductive number and number of unobserved cases today.*

# How to improve understanding of epidemic spread?

---

Epidemic today

We only have sequences up to June!

- ▶ *Thus we cannot quantify the reproductive number and number of unobserved cases today.*

Infected periods

We estimated large confidence intervals.

Simulations show that more data allows us to get more precise estimates!

- ▶ *Based on incubation period, we can evaluate whether a 3-day shutdown is helpful!*

# How to improve understanding of epidemic spread?

---

Epidemic today

We only have sequences up to June!

- ▶ *Thus we cannot quantify the reproductive number and number of unobserved cases today.*

Infected periods

We estimated large confidence intervals.

Simulations show that more data allows us to get more precise estimates!

- ▶ *Based on incubation period, we can evaluate whether a 3-day shutdown is helpful!*

Data situation

No new sequencing data

- ▶ *Ebola needs to be isolated in BSL-4 labs - only 57 labs world-wide!*

# Main challenge in data collection

---

Scienceexpress

## 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 Gnirke,<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>

***In memoriam:*** Tragically, five co-authors, who contributed greatly to public health and research efforts in Sierra Leone, contracted EVD in the course of their work and lost their battle with the disease before this manuscript could be published. We wish to honor their memory.

# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

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

## 2. ...for identifying epidemiological changes over time

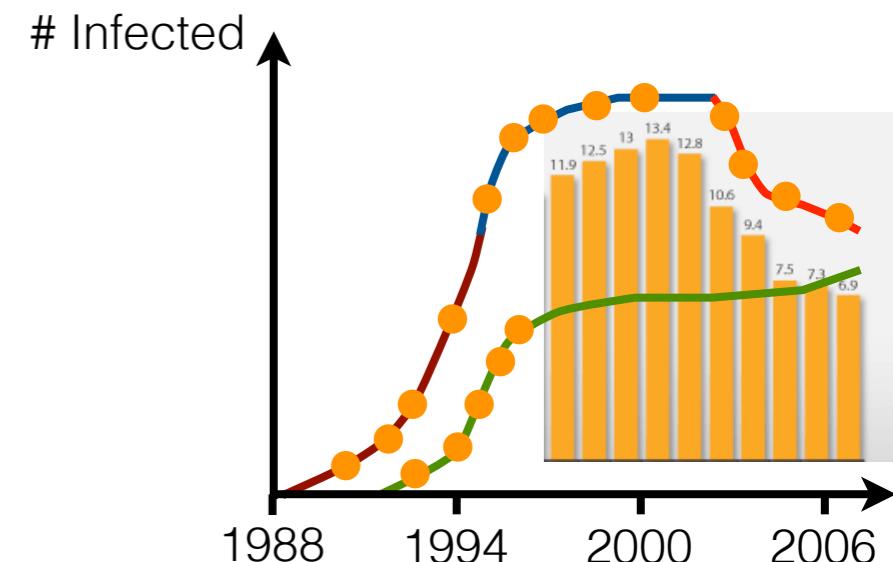
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

## 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 host population structure

- ▶ C: Volz (*Genetics*, 2012)
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013)



# Phylogenetic methods for different epidemic dynamics

## 1. ...for analyzing epidemic outbreaks

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

## 2. ...for identifying epidemiological changes over time

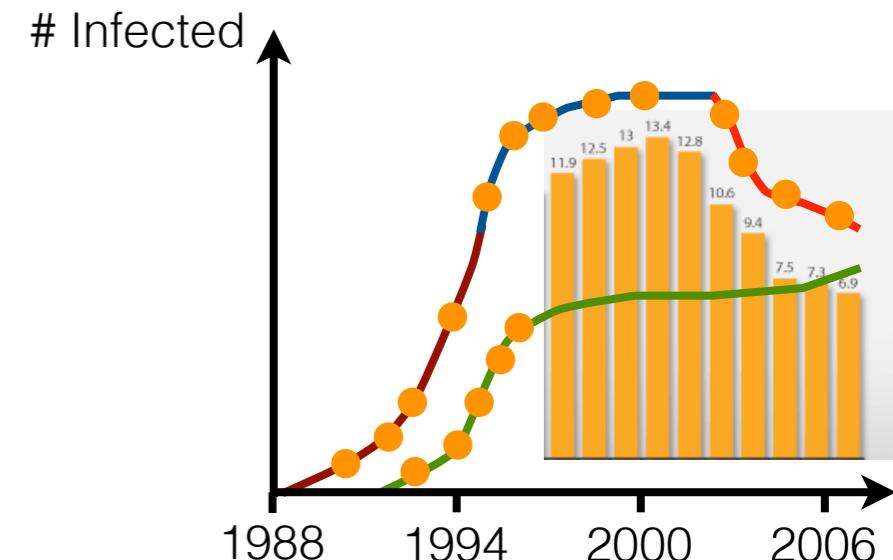
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

## 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 host population structure

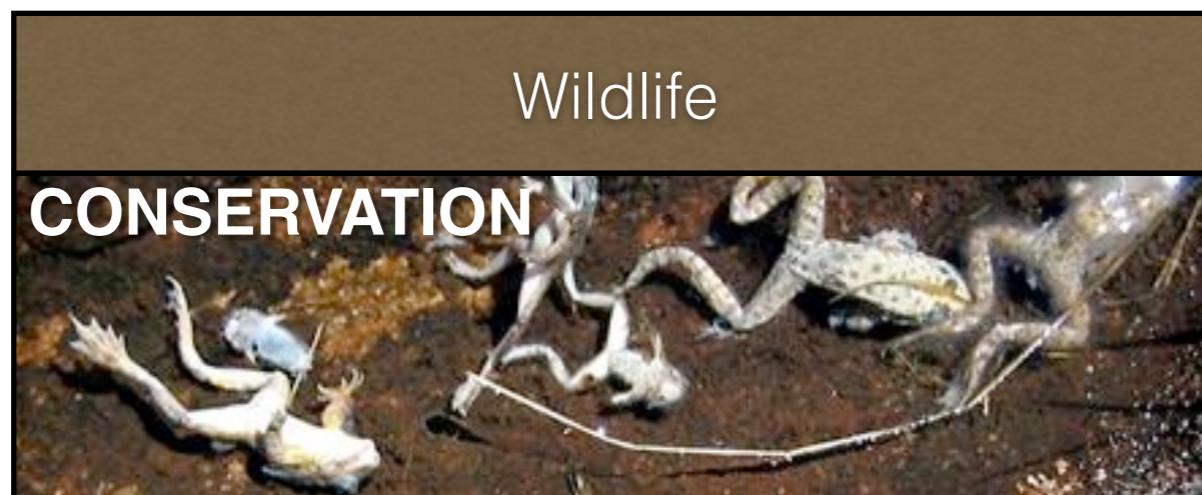
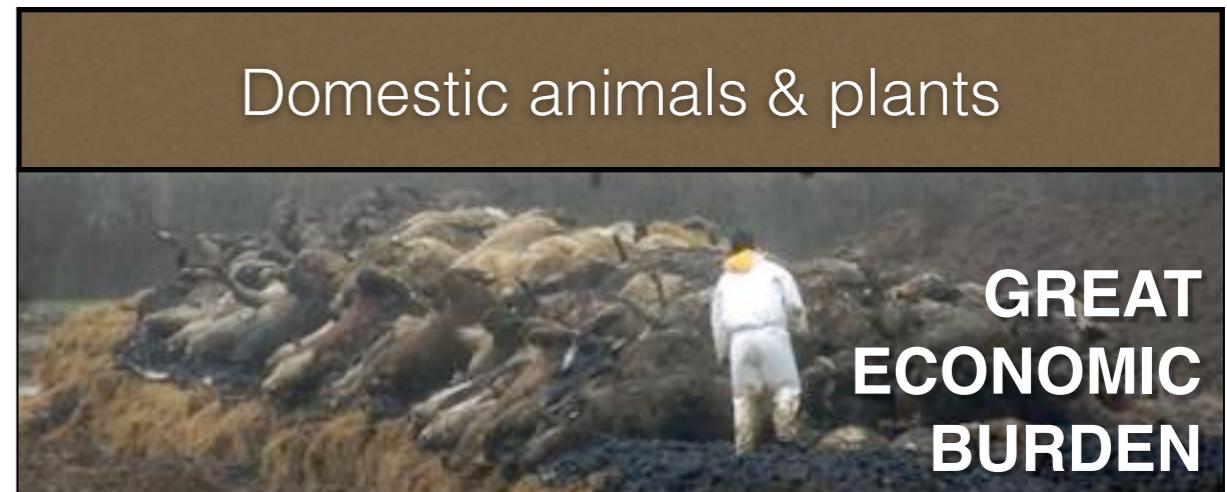
- ▶ C: Volz (*Genetics*, 2012)
- ▶ BD: Stadler & Bonhoeffer (*Phil. Trans. Roy. Soc. B*, 2013)



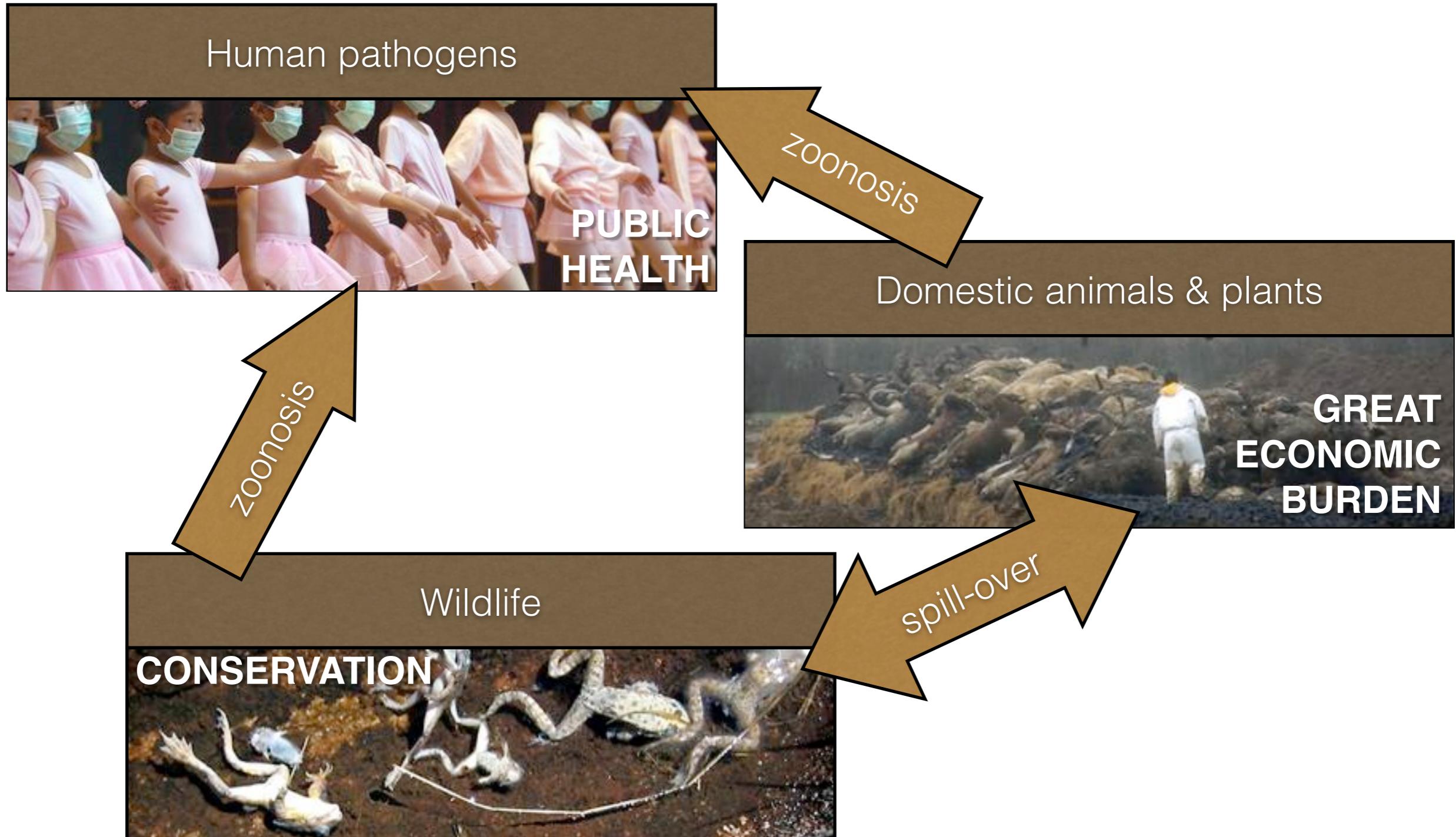
Implemented in Beast v2.0



# Key questions in infectious disease ecology approached through phylogenetics



# Key questions in infectious disease ecology approached through phylogenetics



# Key questions in infectious disease ecology approached through phylogenetics

