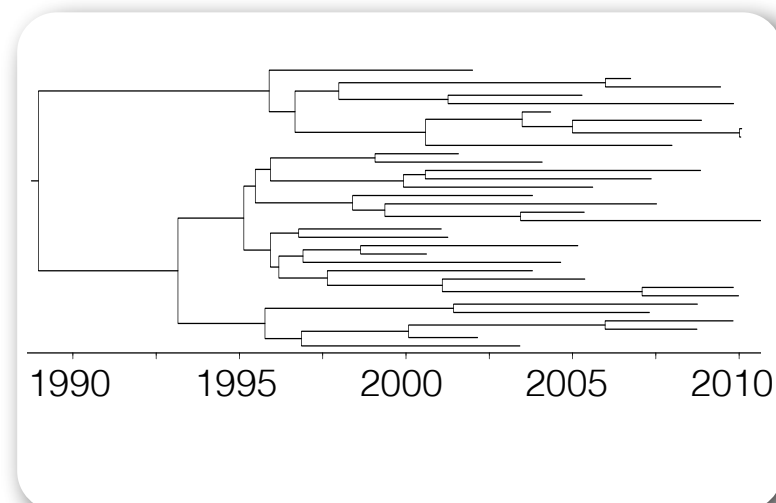
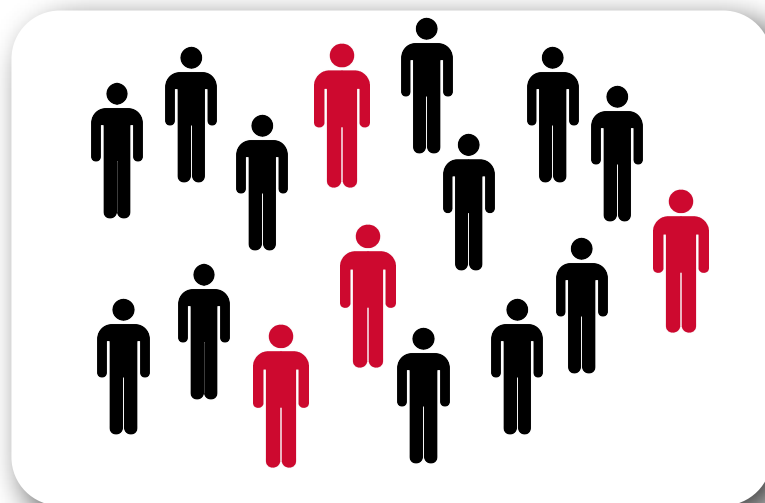


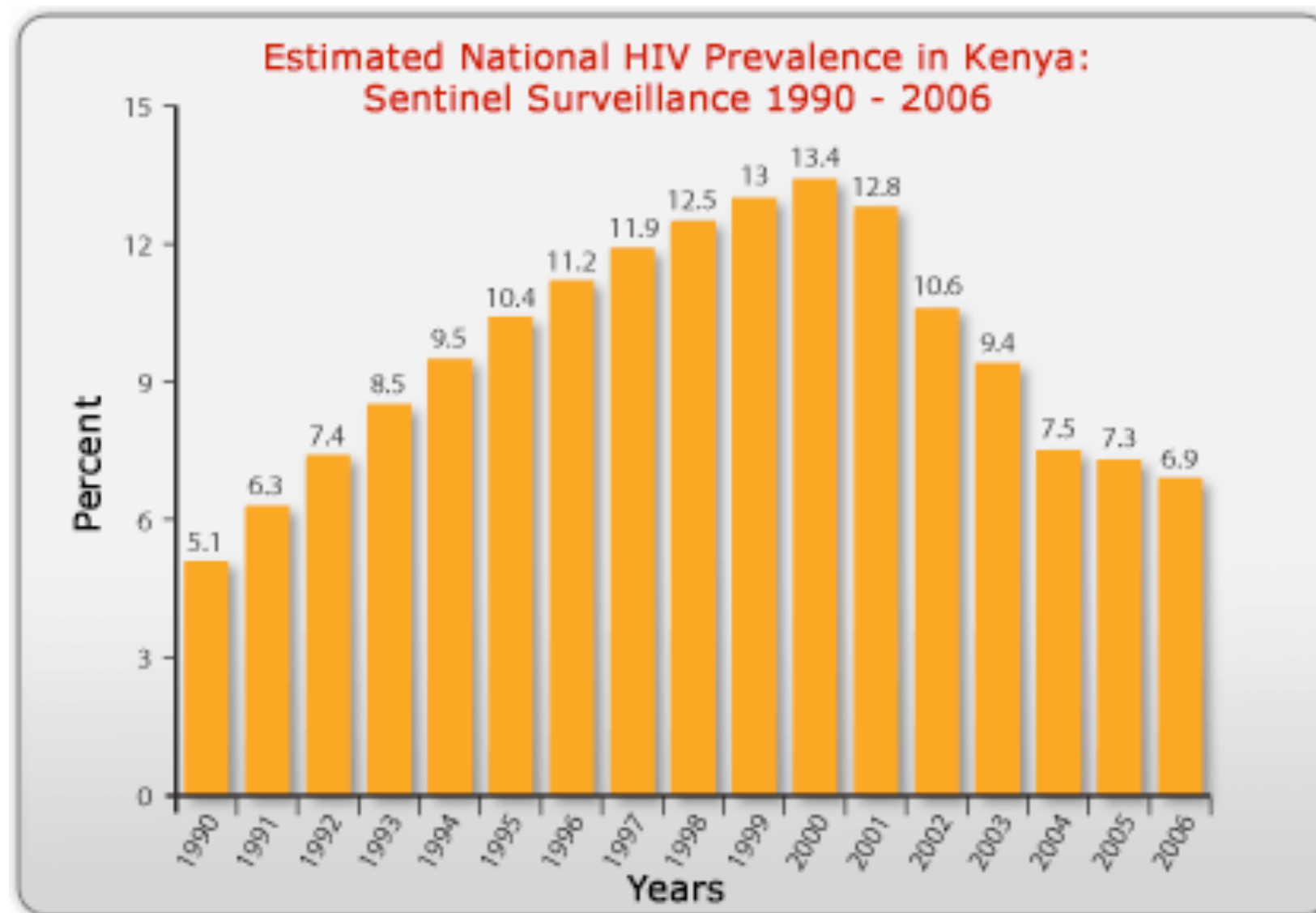
Infectious Disease Dynamics in RevBayes

Inferring epidemiological dynamics based on genetic sequence data

Tracy Heath & Tanja Stadler



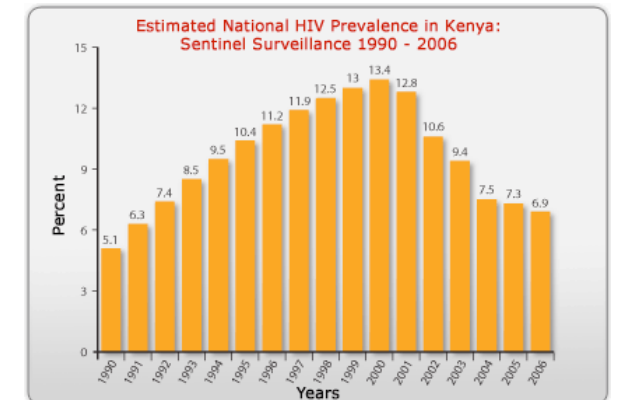
Epidemiology



Epidemiology

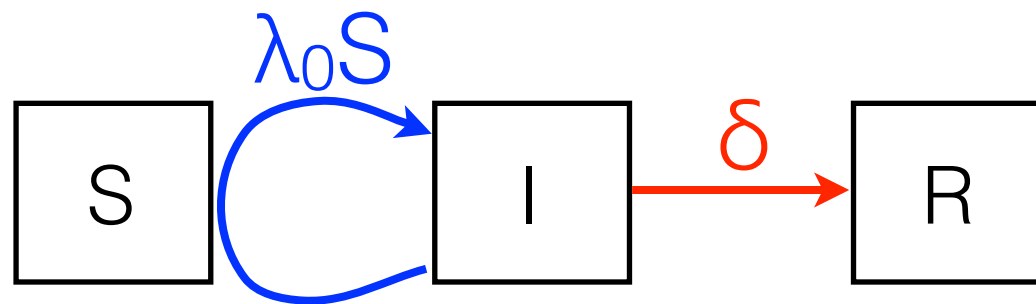
Data

of infected individuals through time



Models

Population dynamics described by SIR models:

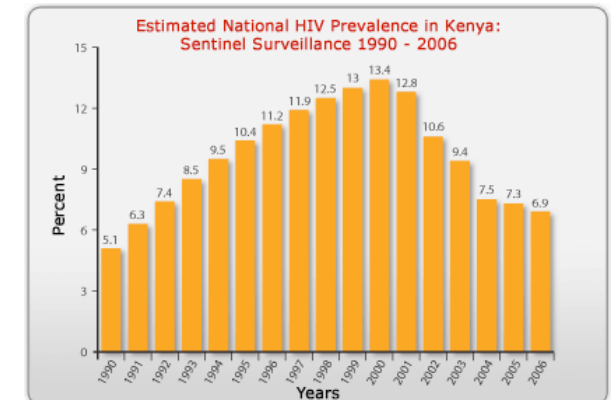


$$\begin{aligned}\frac{dS}{dt} &= -\lambda_0 IS/N \\ \frac{dI}{dt} &= \lambda_0 IS/N - \delta I \\ \frac{dR}{dt} &= \delta I\end{aligned}$$

Epidemiology

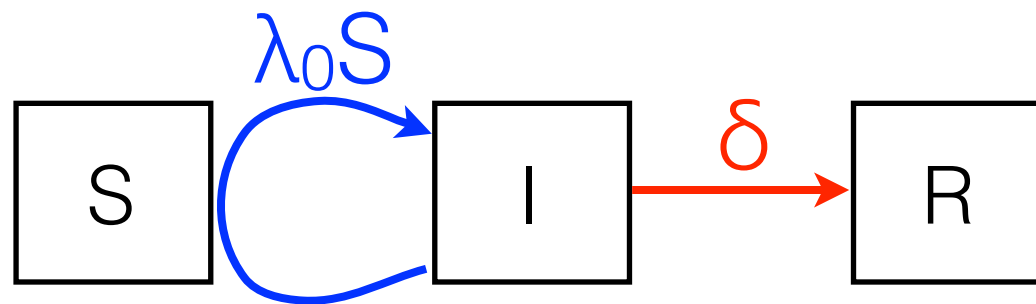
Data

of infected individuals through time



Models

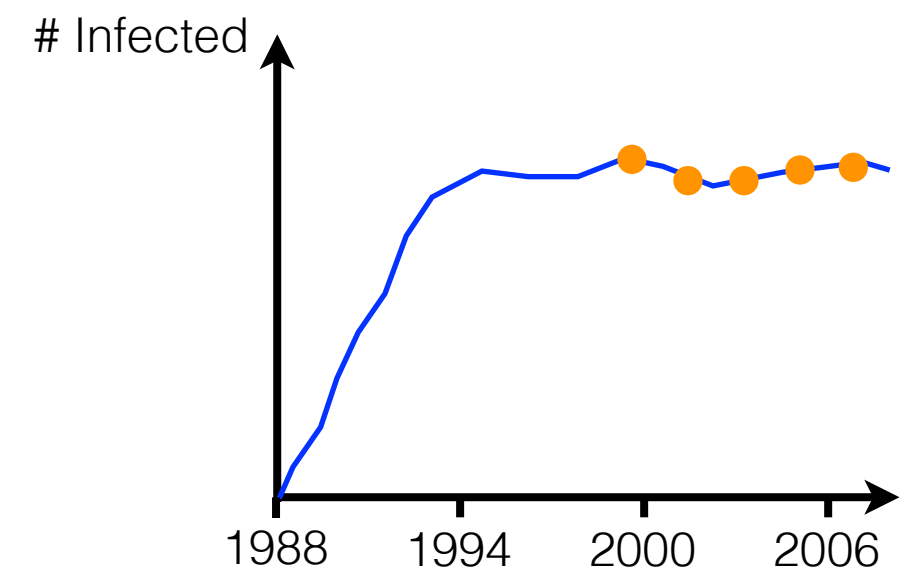
Population dynamics described by SIR models:



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

Limitations

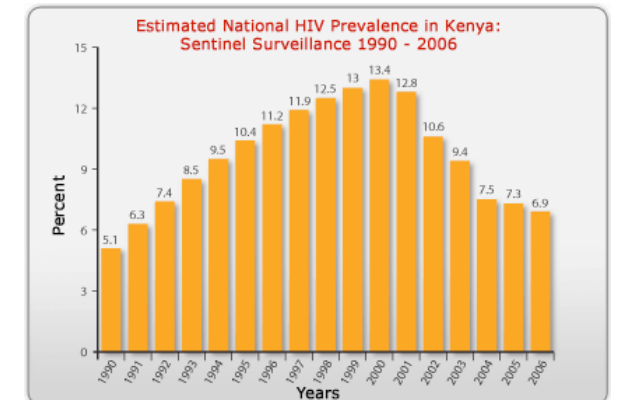
If sampling in early epidemic was missed:



Epidemiology

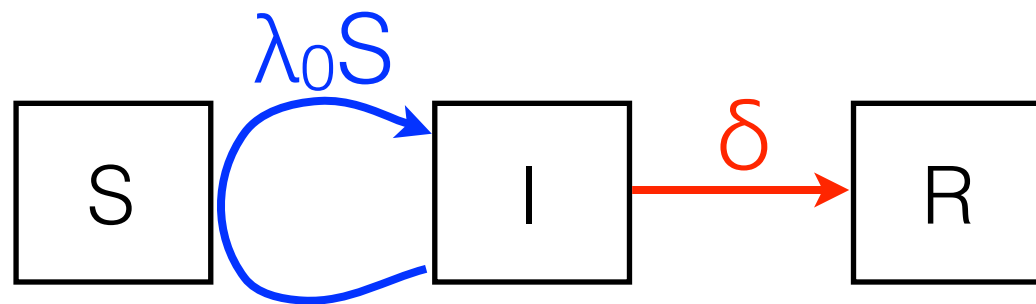
Data

of infected individuals through time



Models

Population dynamics described by SIR models:

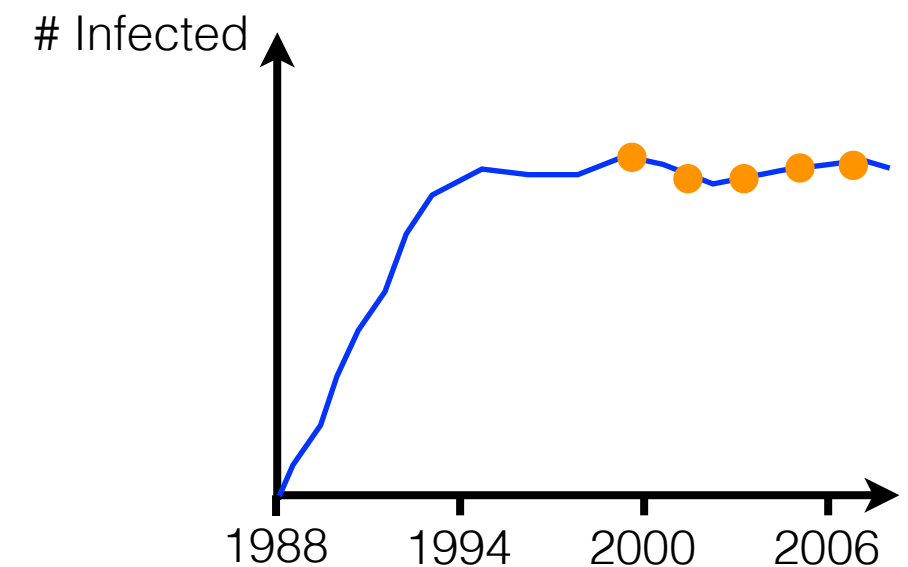


$$\begin{aligned}\frac{dS}{dt} &= -\lambda_0 IS/N \\ \frac{dI}{dt} &= \lambda_0 IS/N - \delta I \\ \frac{dR}{dt} &= \delta I\end{aligned}$$

Limitations

If sampling in early epidemic was missed:

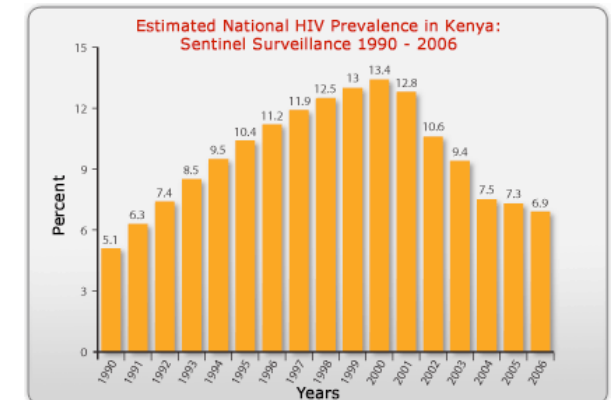
- ▶ **Time of epidemic outbreak?**
- ▶ **Basic reproductive number R_0 ?**



Epidemiology

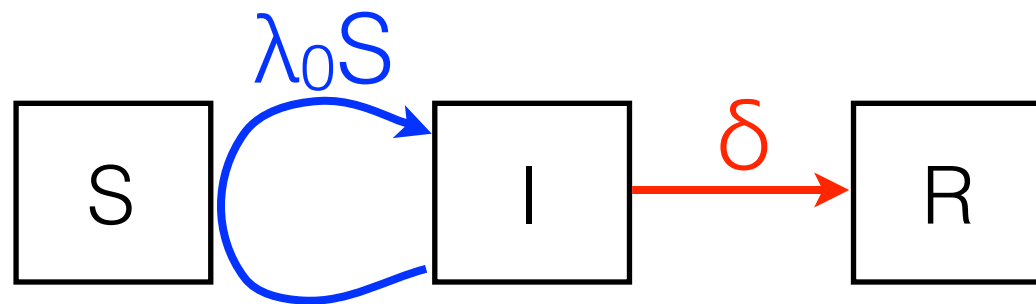
Data

of infected individuals through time



Models

Population dynamics described by SIR models:



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

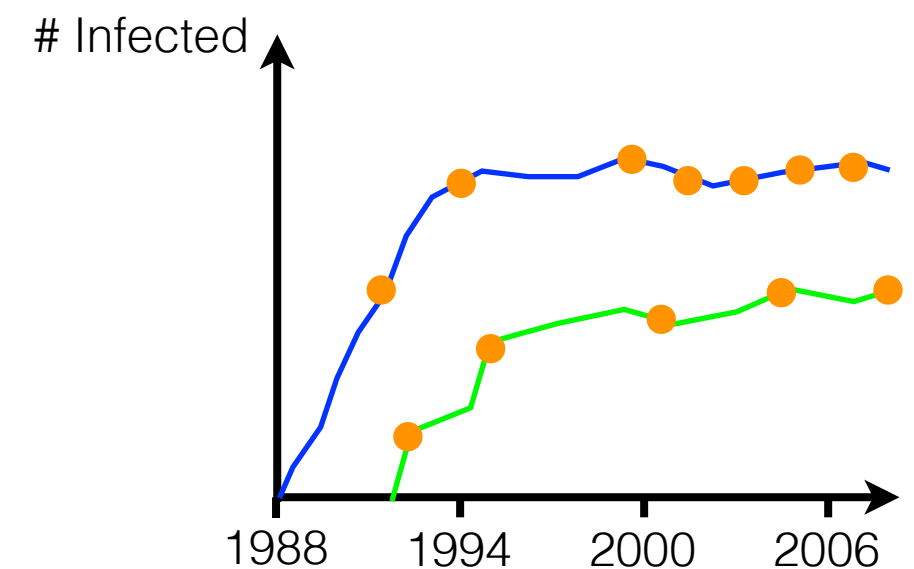
Limitations

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 ACACACCC

Inf2 TCACACCT

Inf3 AAAGACTT

Inf4 ACAGACTT

Genetic data entering epidemiology

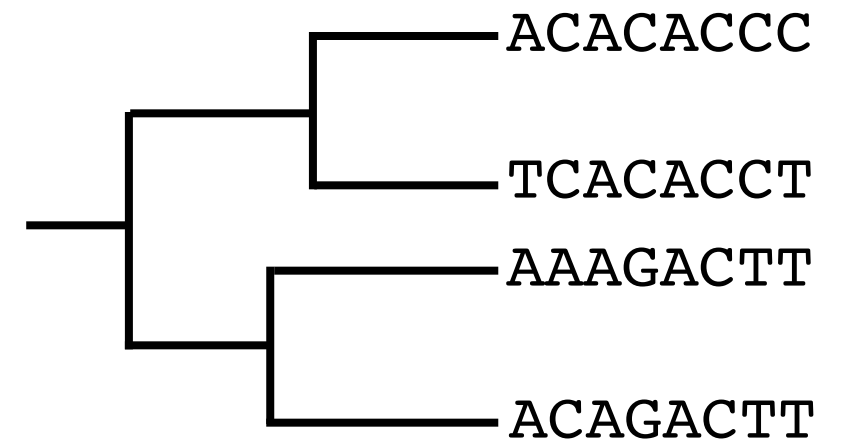
Phylo-
genetics

Input:

Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)



Genetic data entering epidemiology

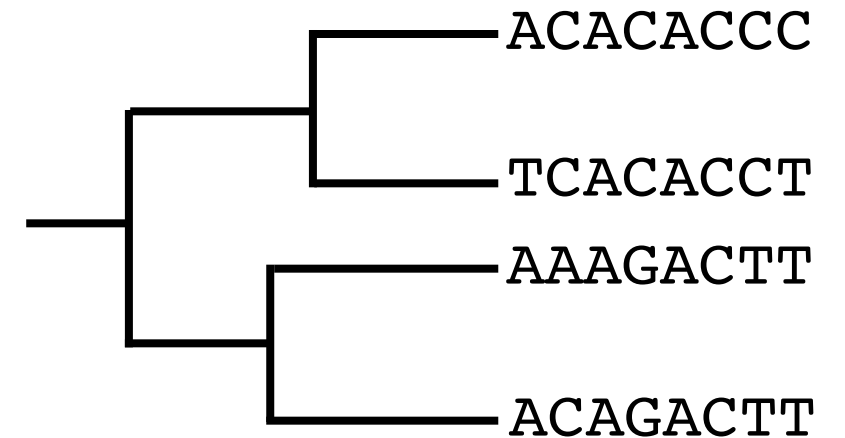
Phylo- genetics

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

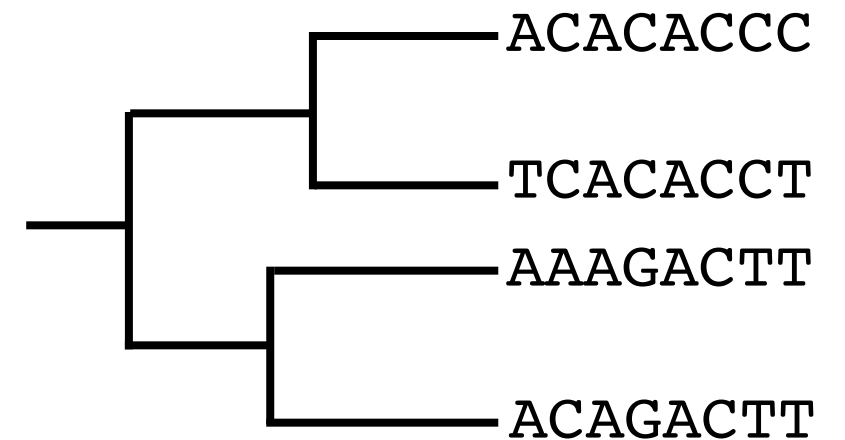
Phylo- genetics

Input:

Pathogen genetic data from different hosts

Output:

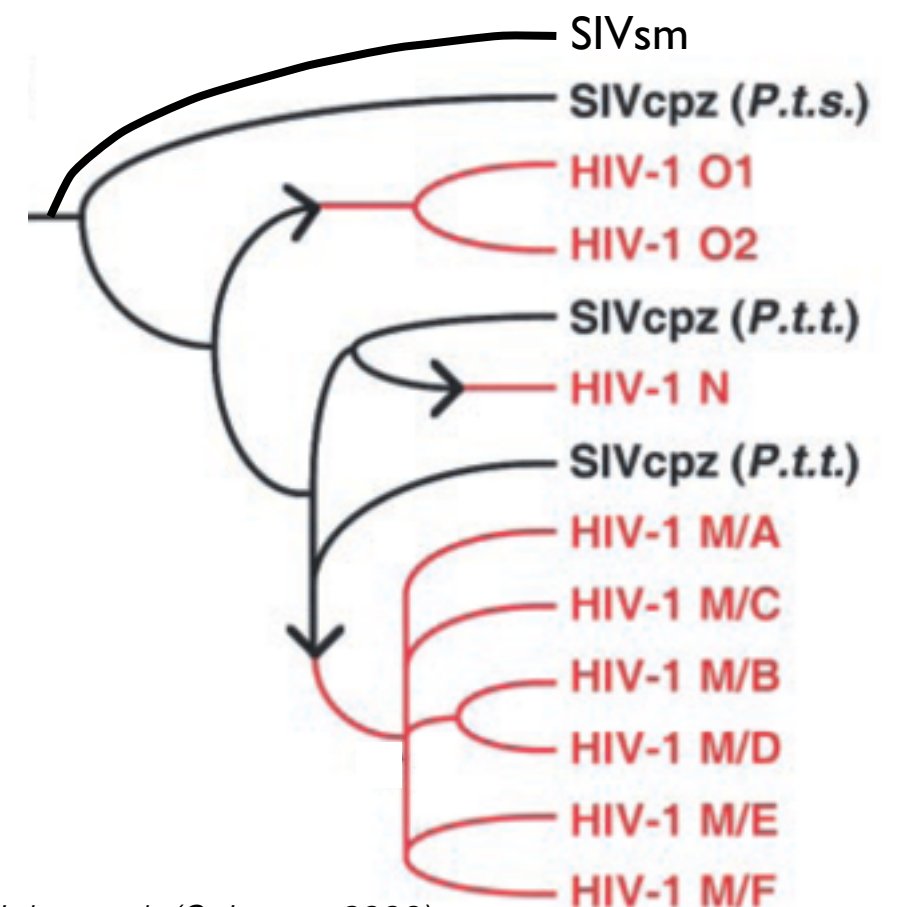
Transmission chain (who infected whom)



Example key result

Emergence of HIV

Epidemic identified in 1980s



Genetic data entering epidemiology

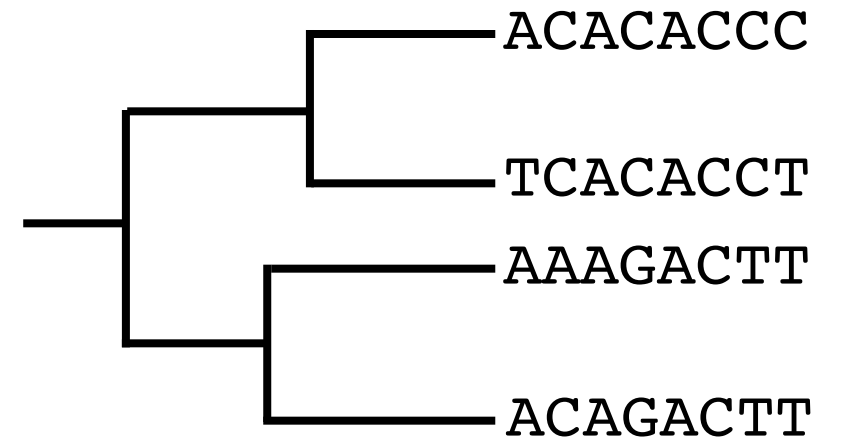
Phylo- genetics

Input:

Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)



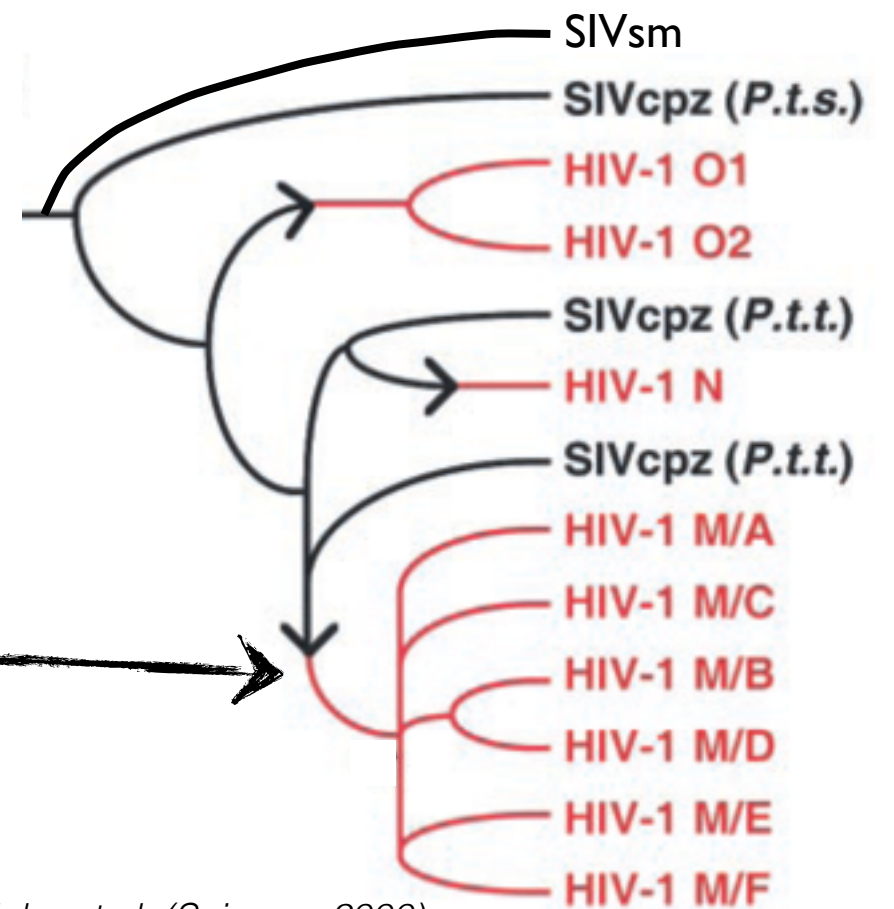
Example
key result

Emergence of HIV

Epidemic identified in 1980s

Sequence data tells us about pre-1980

- ▶ Zoonosis from chimps



adopted from Hahn et al. (Science, 2000)

Genetic data contains information about the epidemiological dynamics

Phylo-
genetics

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

Phylo- genetics

State of process

- ▶ Time and geographic location of outbreak



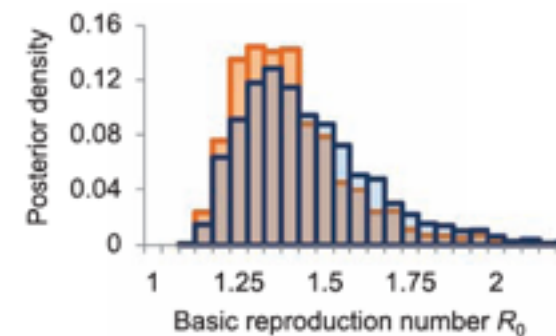
Emergence of a
pathogen
(here HIV)

Hahn et al. (Science, 2000)

Phylo- dynamics

Dynamics of process

- ▶ Transmission process



Basic reproductive
number of an
emerging pathogen
(here H1N1)

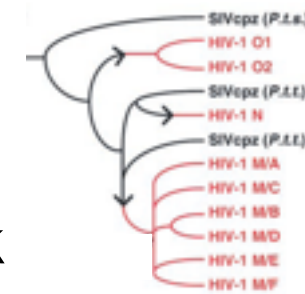
Fraser et al. (Science, 2009)

Genetic data contains information about the epidemiological dynamics

Phylo- genetics

State of process

- ▶ Time and geographic location of outbreak

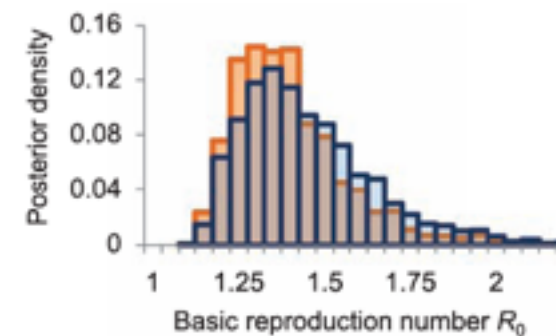


Emergence of a
pathogen
(here HIV)
Hahn et al. (Science, 2000)

Phylo- dynamics

Dynamics of process

- ▶ Transmission process



Basic reproductive
number of an
emerging pathogen
(here H1N1)
Fraser et al. (Science, 2009)

Epidemio- logical models

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

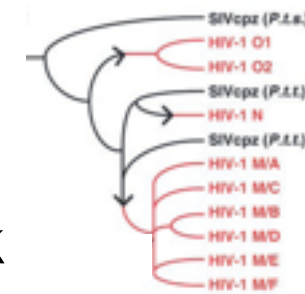
- ▶ Quantification of all model parameters!

Genetic data contains information about the epidemiological dynamics

Phylo- genetics

State of process

- ▶ Time and geographic location of outbreak



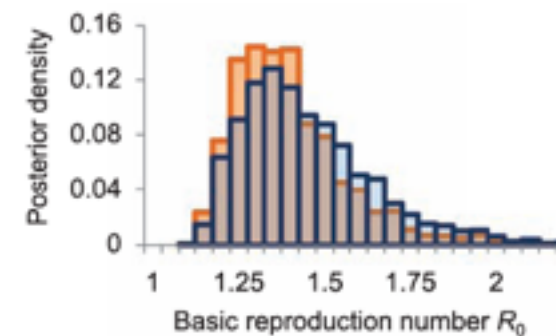
Emergence of a
pathogen
(here HIV)

Hahn et al. (Science, 2000)

Phylo- dynamics

Dynamics of process

- ▶ Transmission process



Basic reproductive
number of an
emerging pathogen
(here H1N1)

Fraser et al. (Science, 2009)

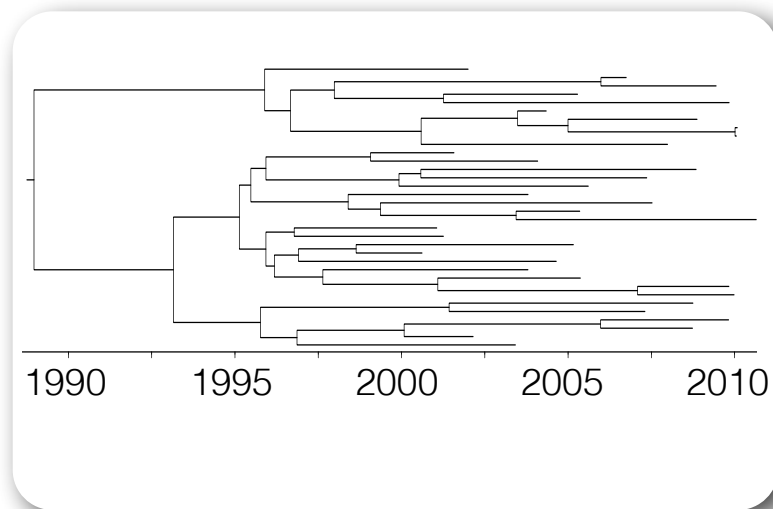
Epidemio- logical models

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

- ▶ Quantification of all model parameters!

Approximate: by population genetic model (coalescent)

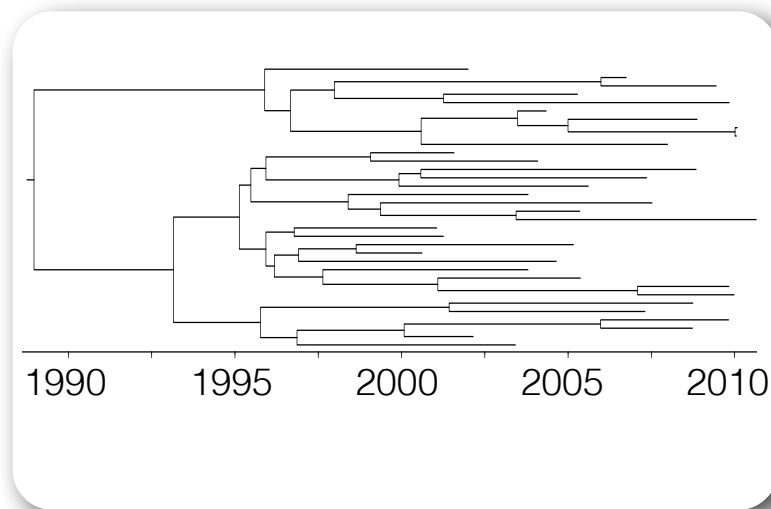
Merge epidemiology and phylogenetics



Methodology to:

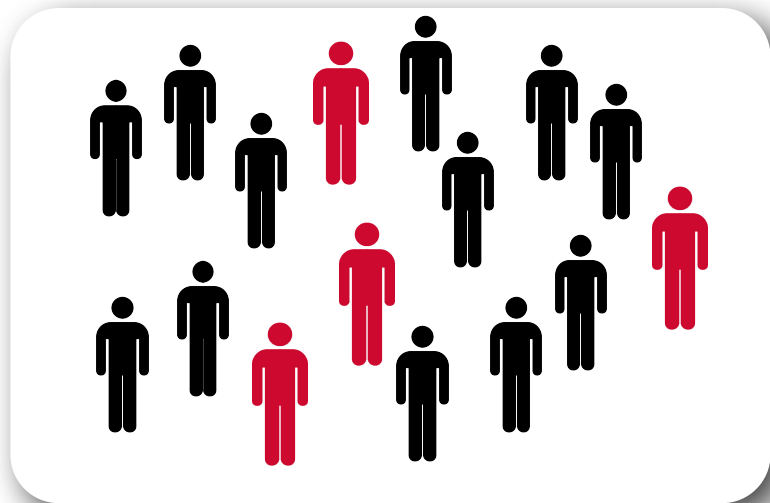
Simultaneously infer phylogenetic trees and epidemiological parameters

Merge epidemiology and phylogenetics



Methodology to:

Simultaneously infer phylogenetic trees and epidemiological parameters



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 η

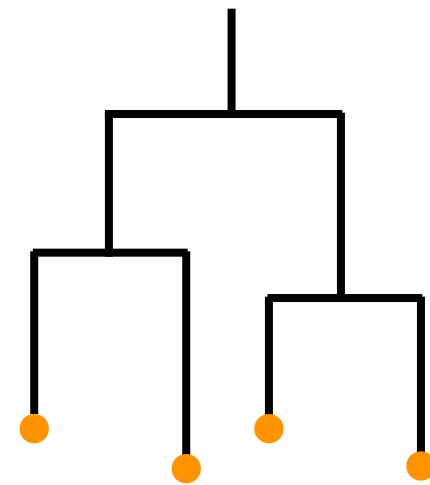


General modeling framework

Epidemiological
model

Growth of tree

- parameters η



General modeling framework

Epidemiological model

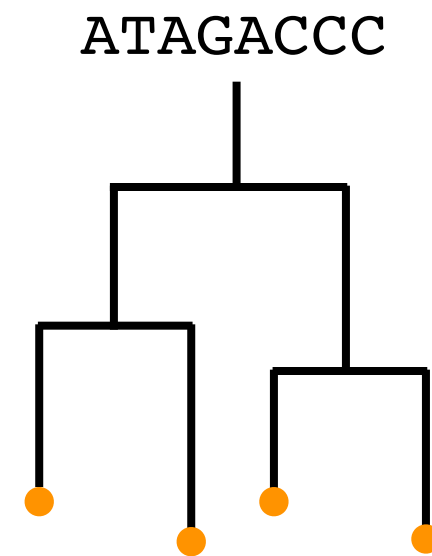
Growth of tree

- ▶ parameters η

Evolutionary model

Evolution of sequences along tree

- ▶ parameters θ



General modeling framework

Epidemiological model

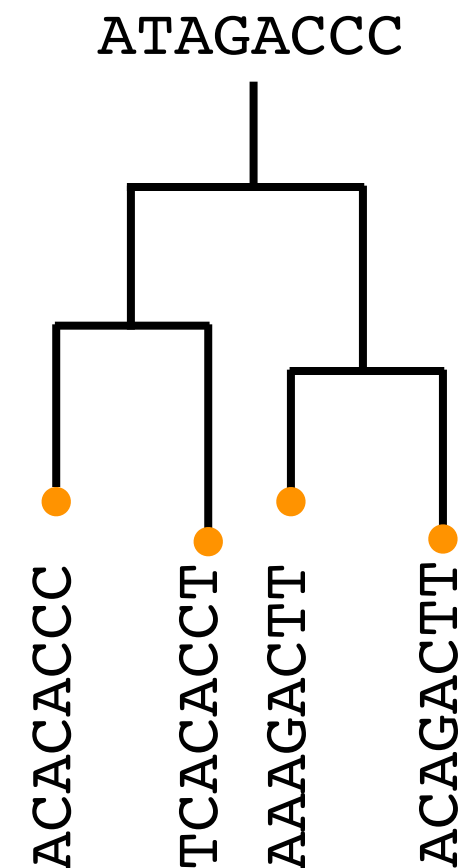
Growth of tree
▶ parameters η

Evolutionary model

Evolution of sequences along tree
▶ parameters θ

Tree distribution

Defined through η and θ



General modeling framework

Epidemiological model

Growth of tree
▸ parameters η

Evolutionary model

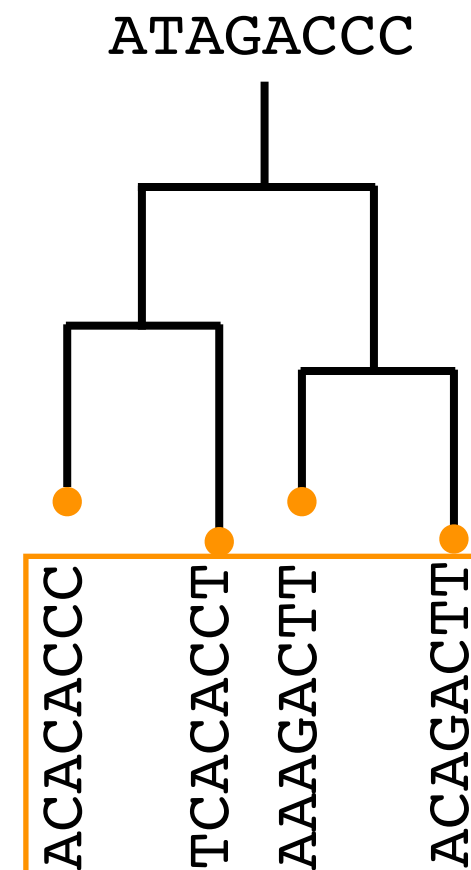
Evolution of sequences along tree
▸ parameters θ

Tree distribution

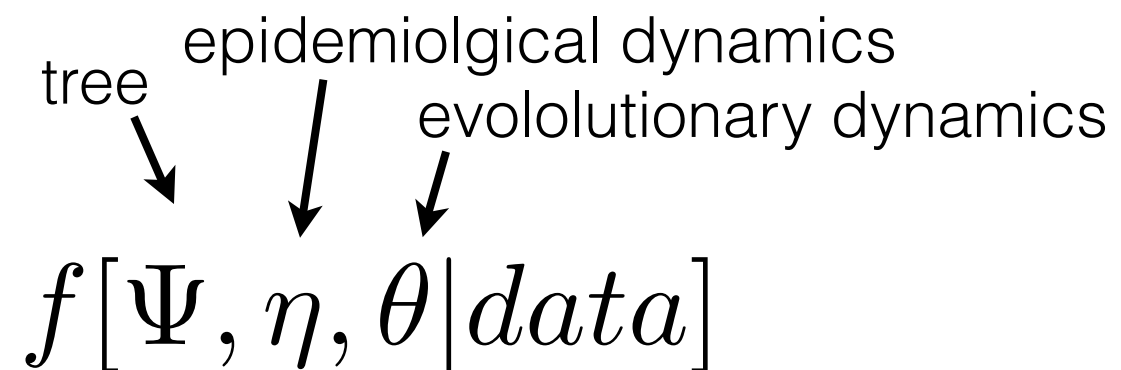
Defined through η and θ

Data

Sequence alignment



Bayesian approach for estimating epidemiological parameters



Bayesian approach for estimating epidemiological parameters

tree epidemiological dynamics evolutionary dynamics

↓ ↓ ↓

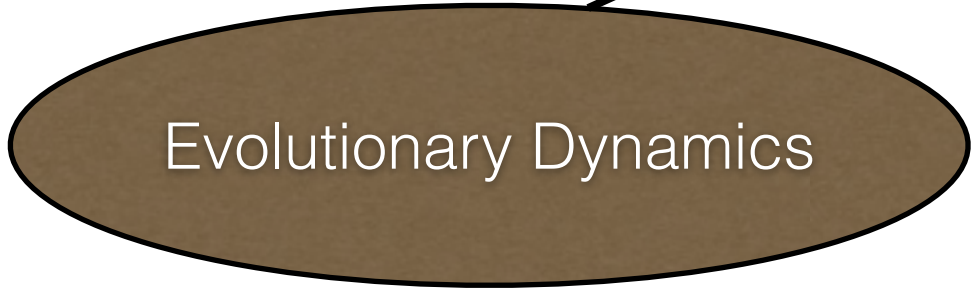
$$f[\Psi, \eta, \theta | data] = \frac{f[data | \Psi, \theta] f[\Psi | \eta] f[\eta, \theta]}{f[data]}$$

Bayesian approach for estimating epidemiological parameters

tree epidemiolglcal dynamics evololutionary dynamics

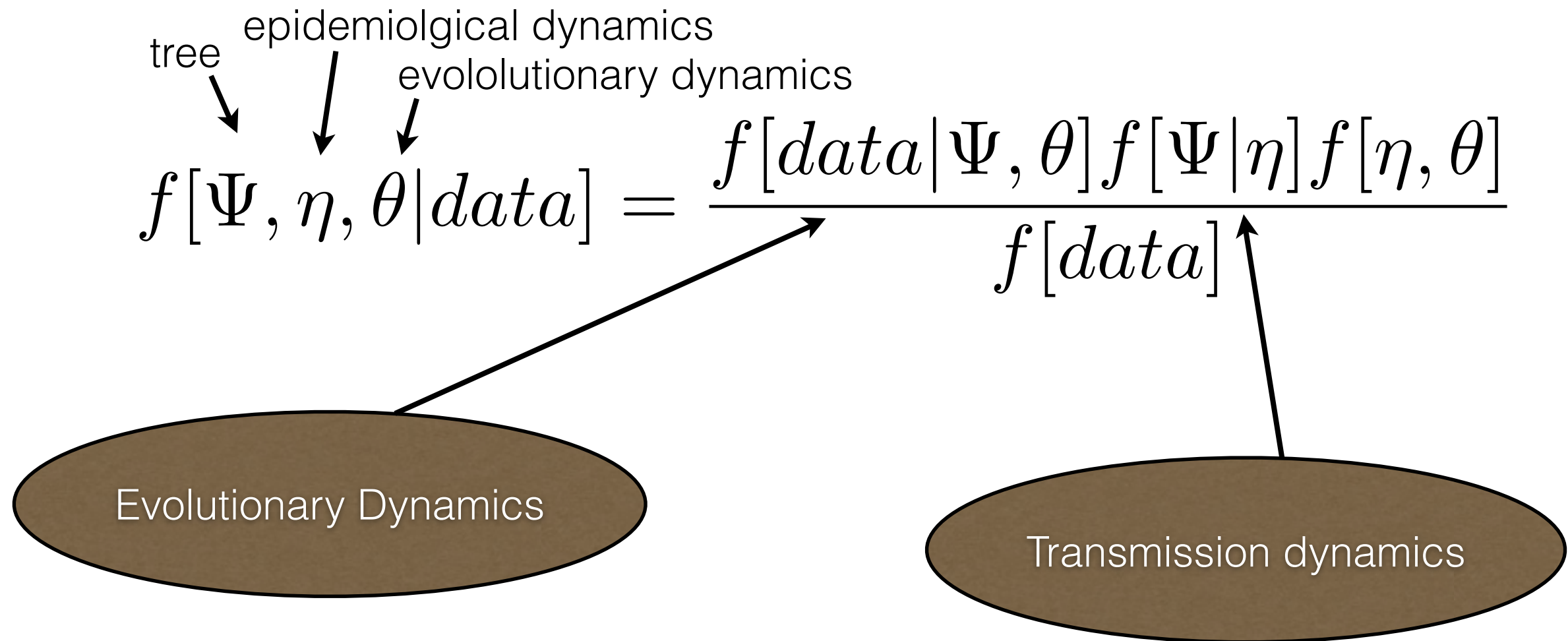
↓ ↓ ↓

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

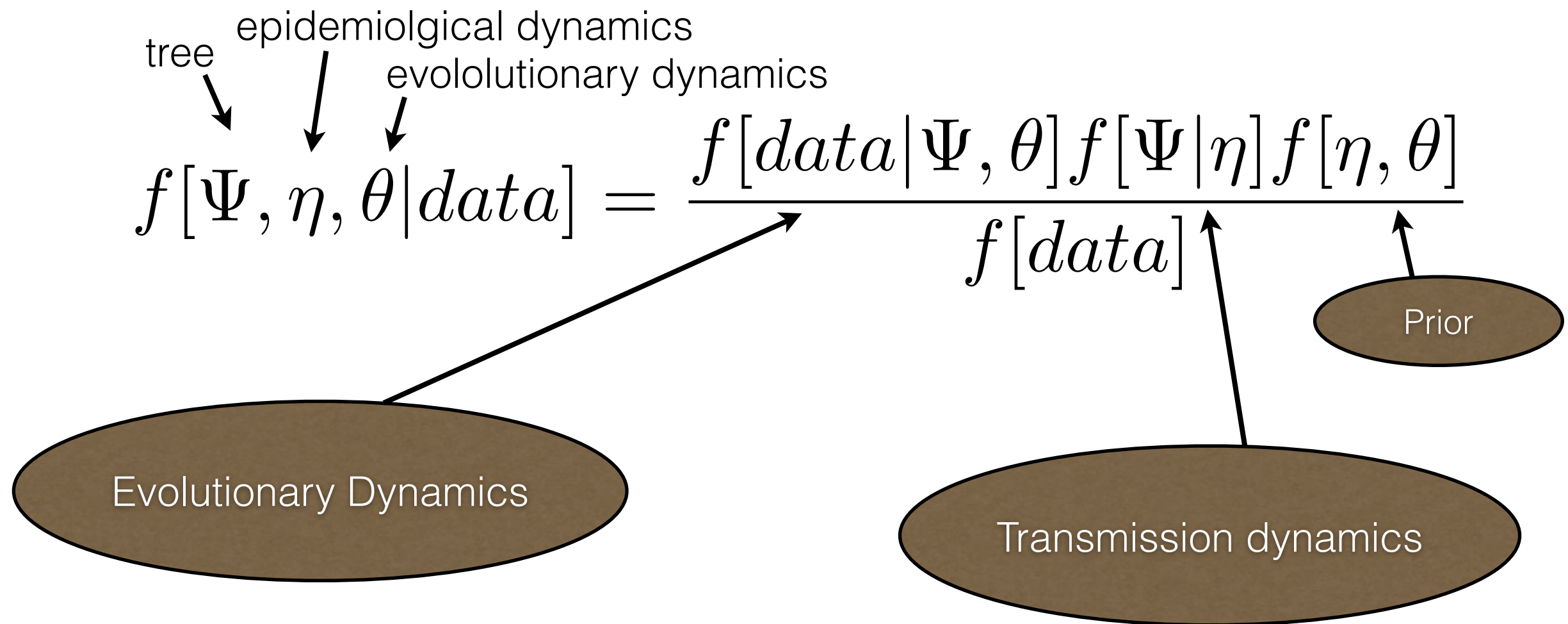


Evolutionary Dynamics

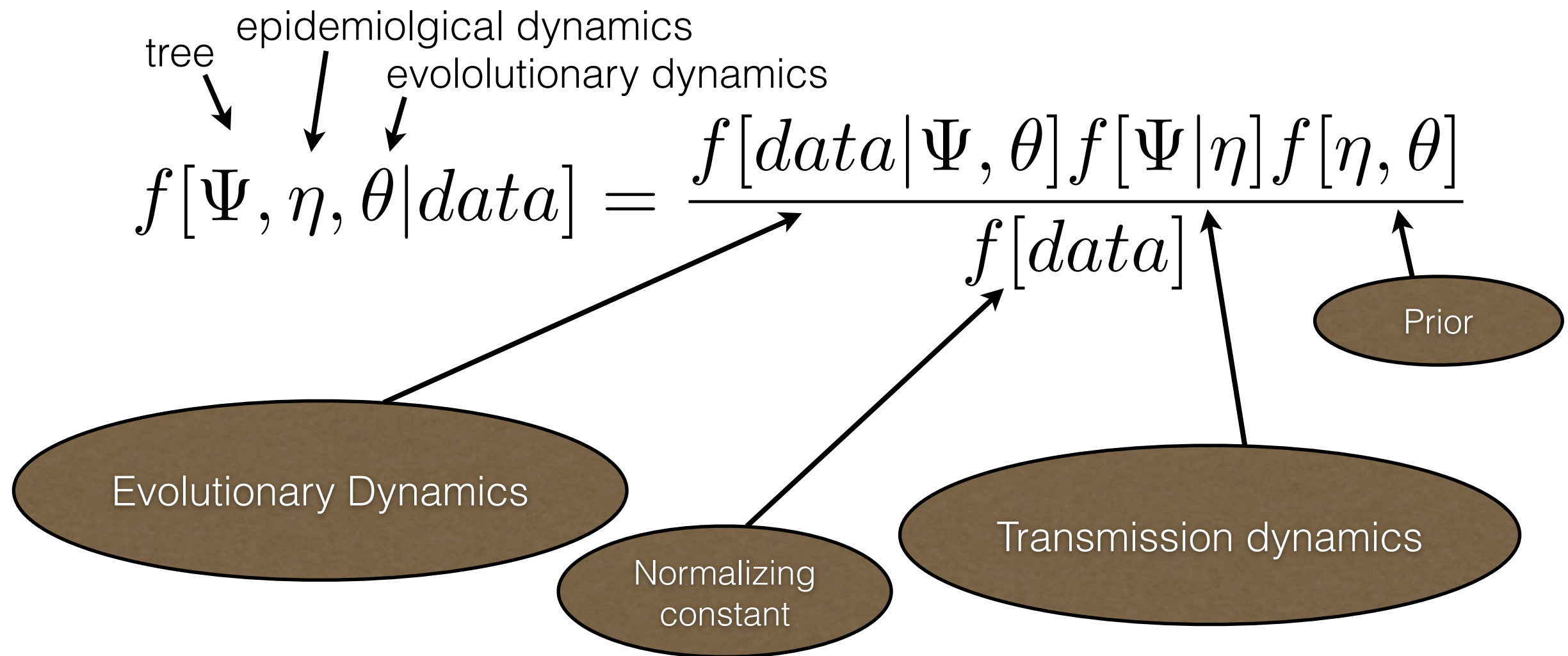
Bayesian approach for estimating epidemiological parameters



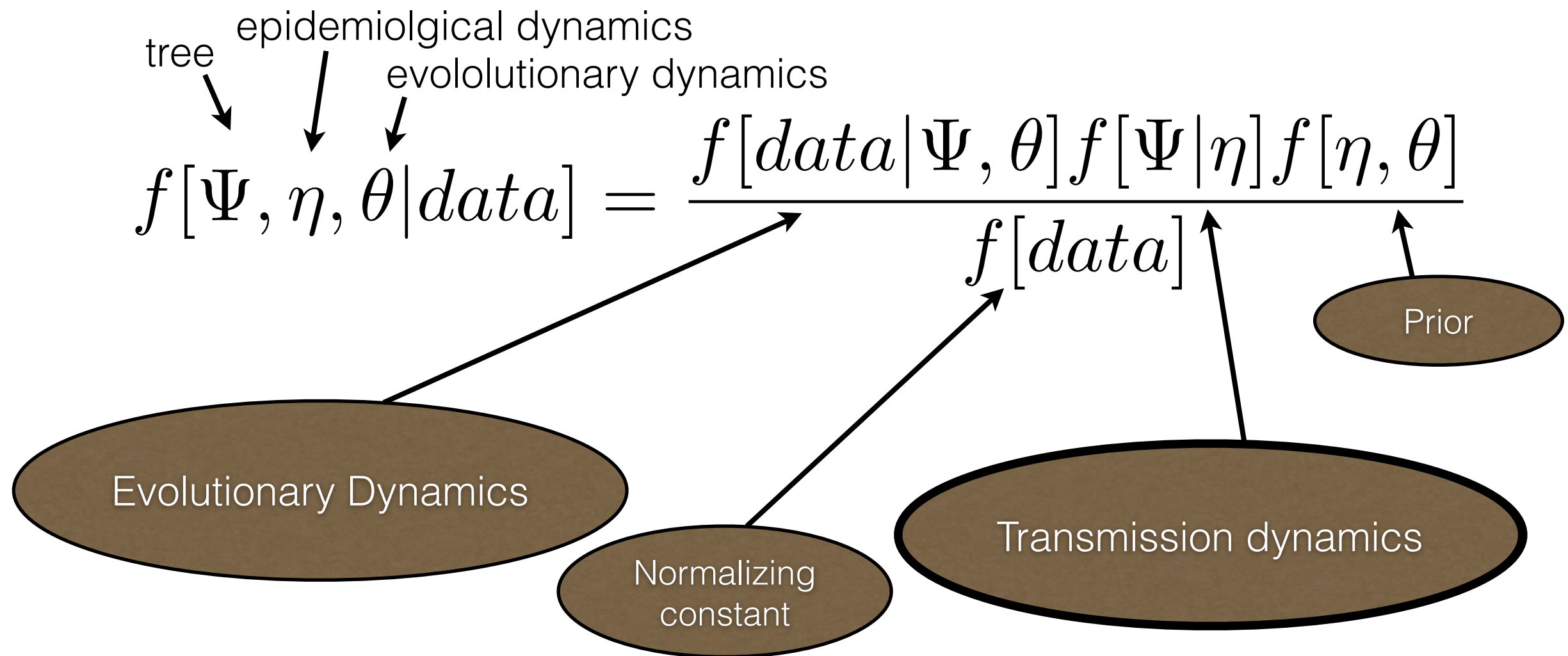
Bayesian approach for estimating epidemiological parameters



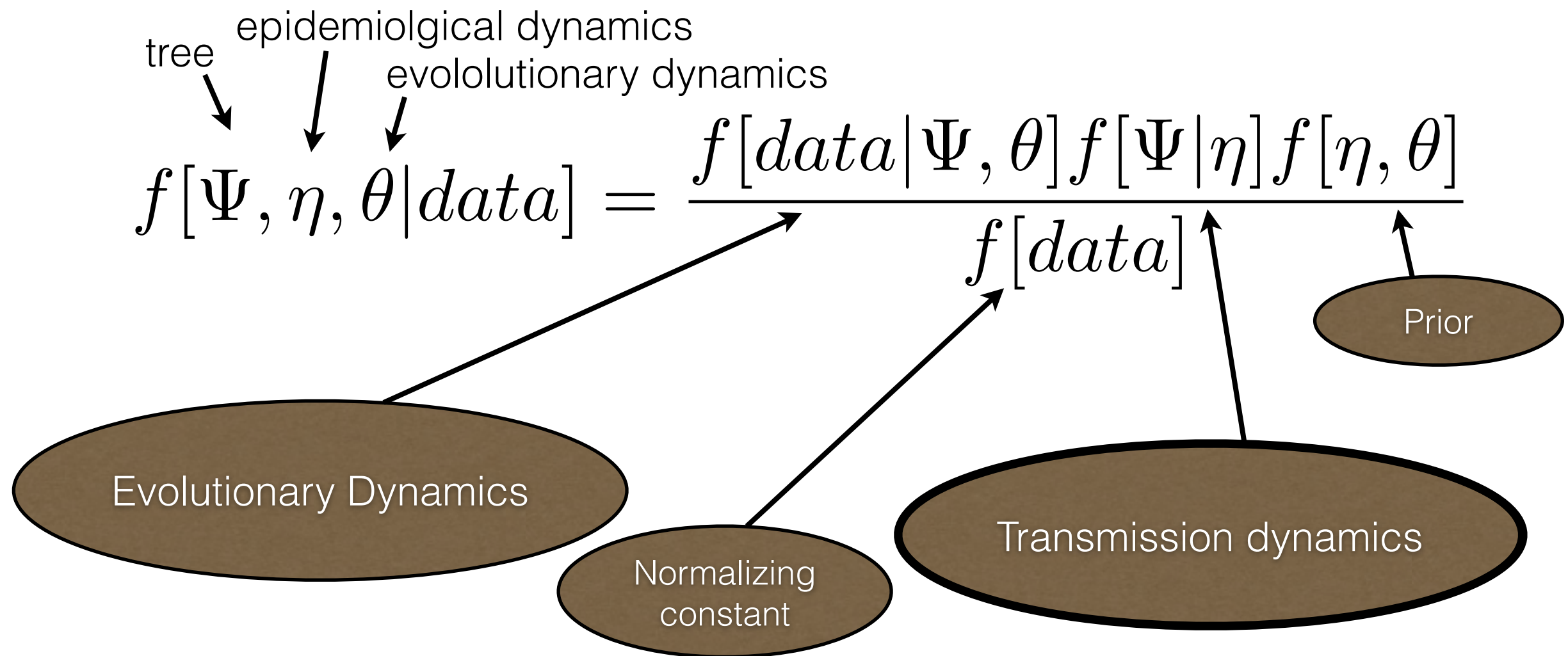
Bayesian approach for estimating epidemiological parameters



Bayesian approach for estimating epidemiological parameters



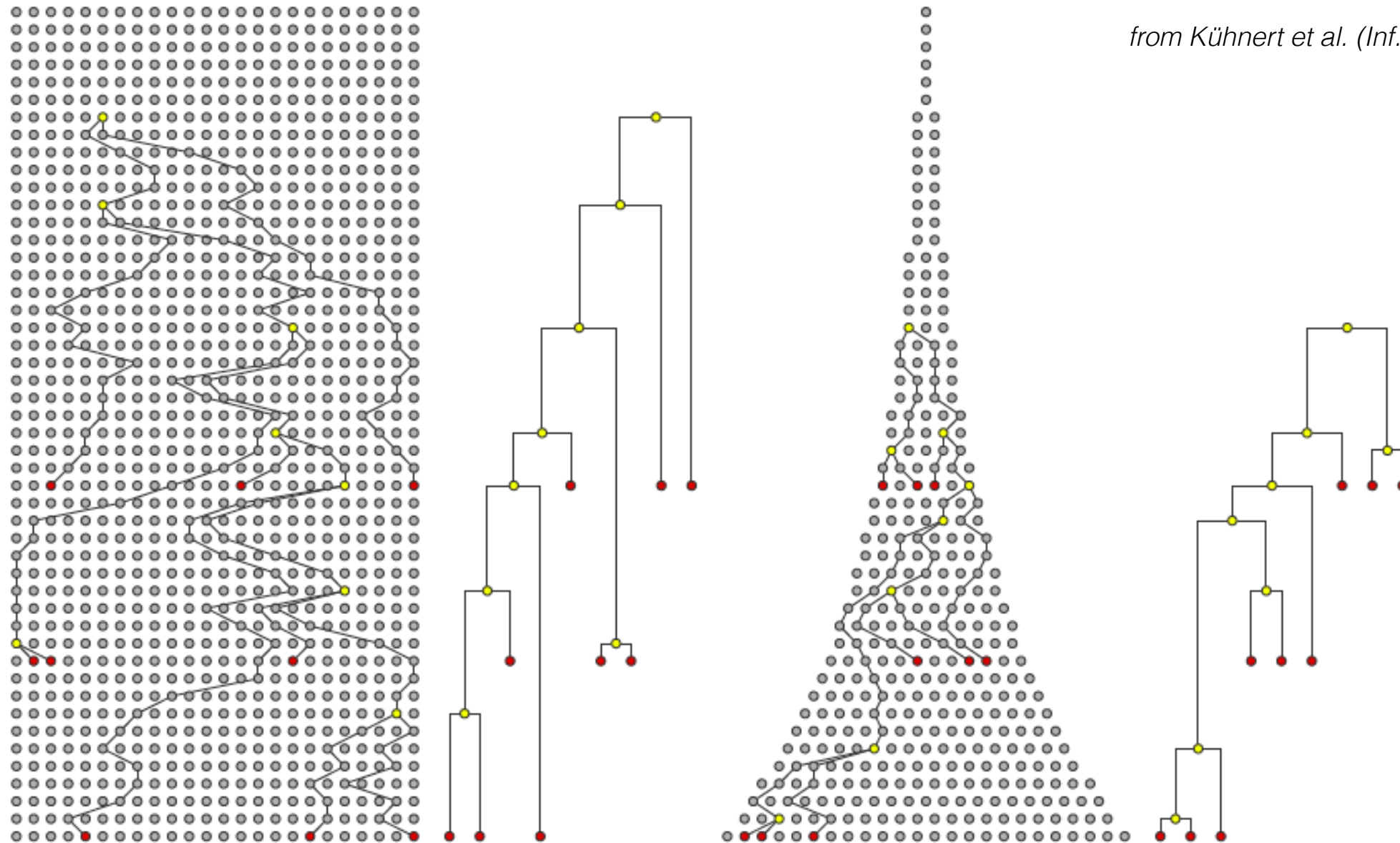
Bayesian approach for estimating epidemiological parameters



RevBayes samples the posterior distribution of trees and parameters

Part A

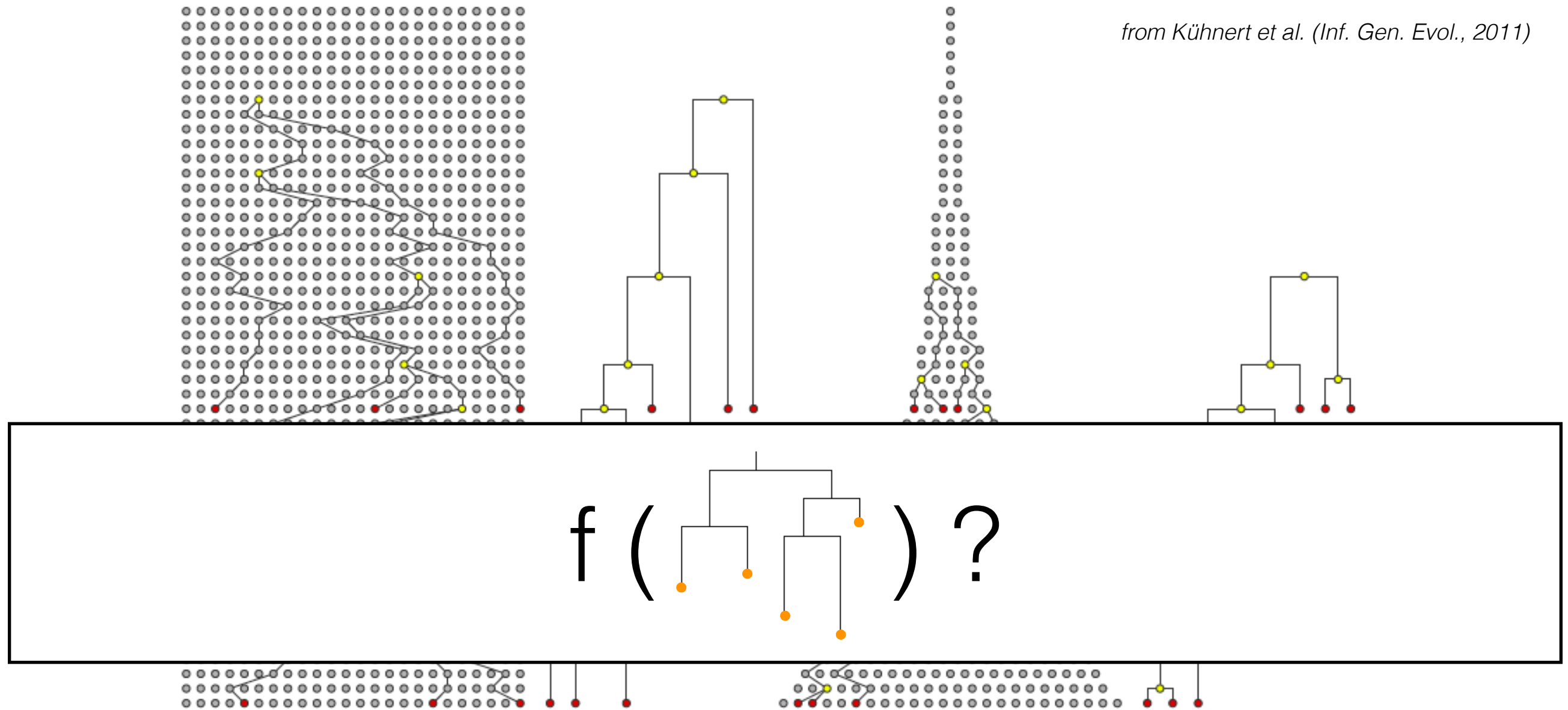
Coalescent as a model for transmission



Part A

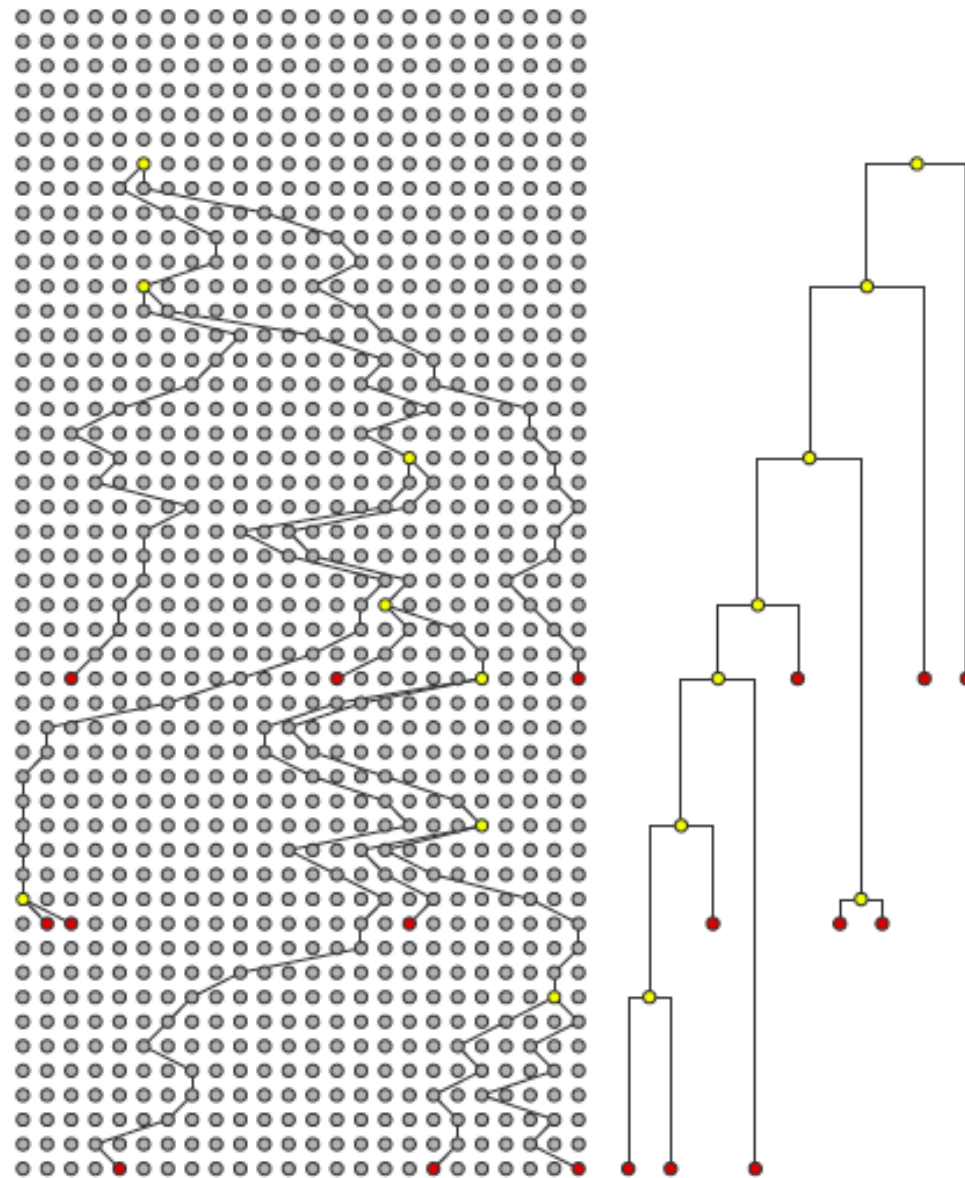
Coalescent as a model for transmission

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

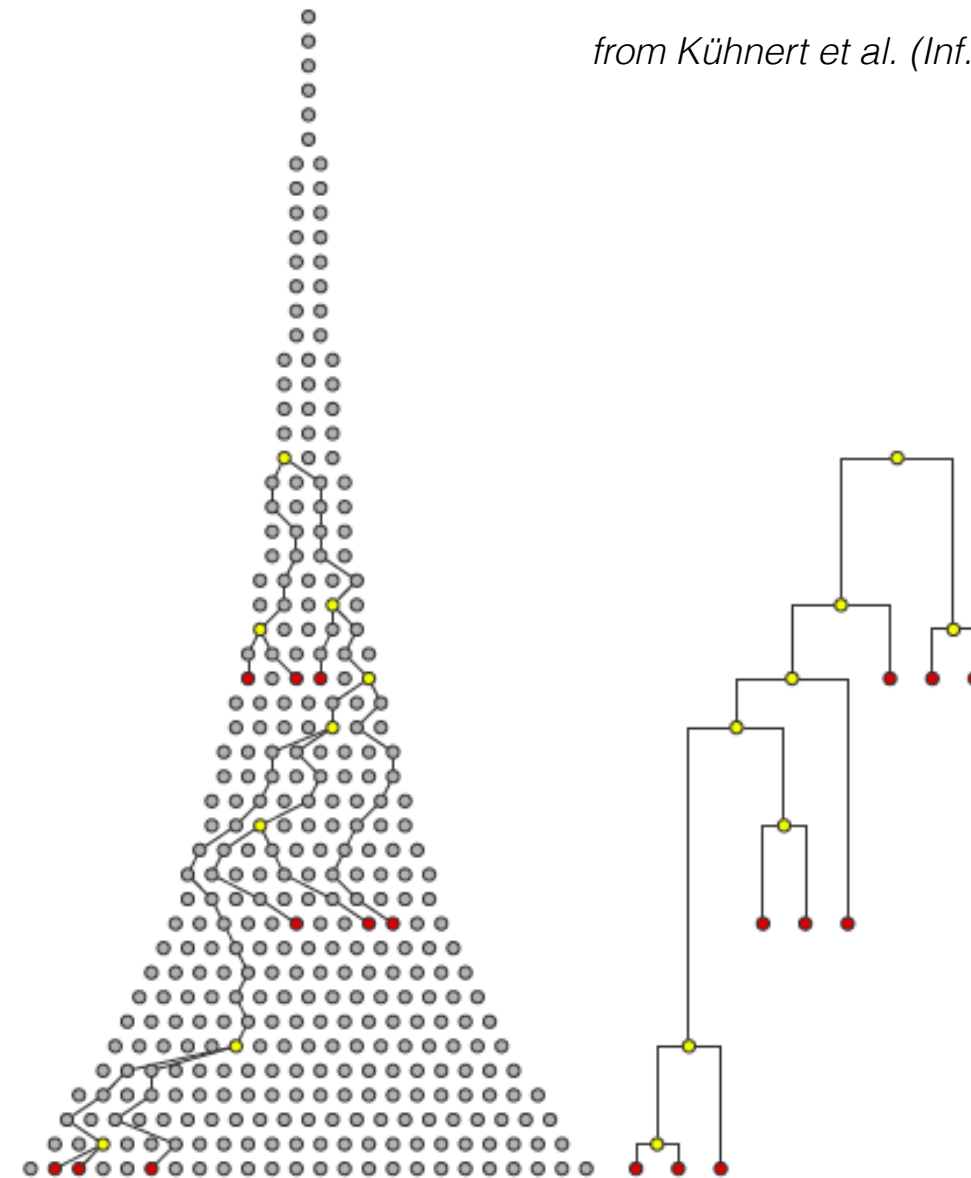


Part A

Coalescent as a model for transmission



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



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

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

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 ?

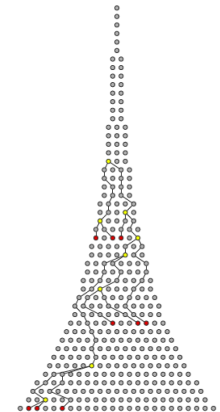
Estimating R_0 from sequencing data

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. $\eta=r$)
 $R_0 = \lambda/\delta = rD + 1$ (where D is expected time of infectiousness)



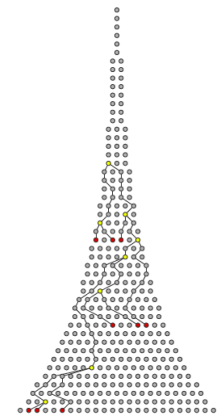
Estimating R_0 from sequencing data

Hepatitis C

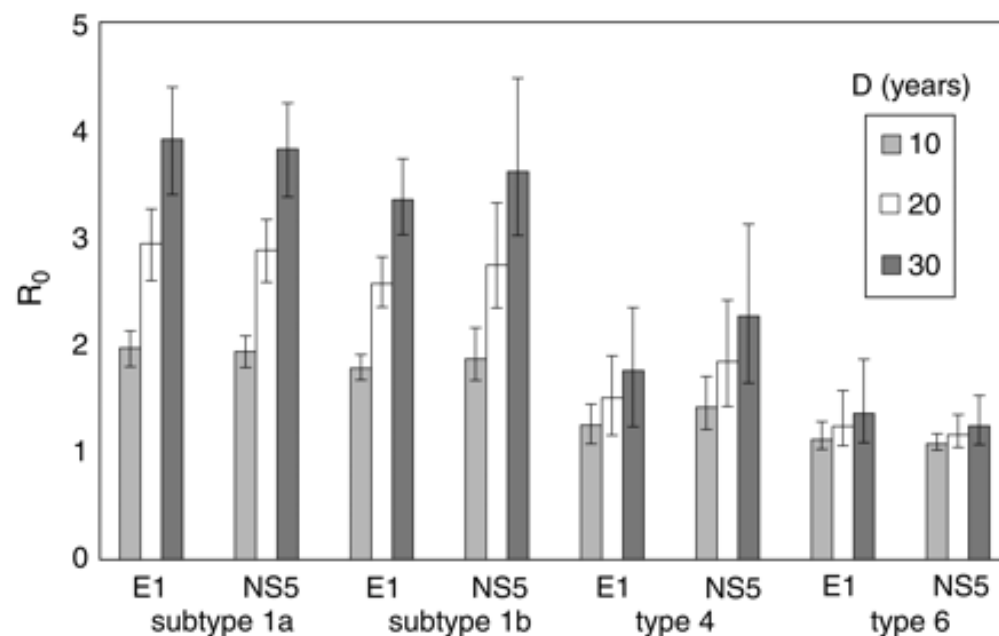
Little known before first isolation, ie. pre-1989.
How can we determine basic reproductive number R_0 ?

Inference

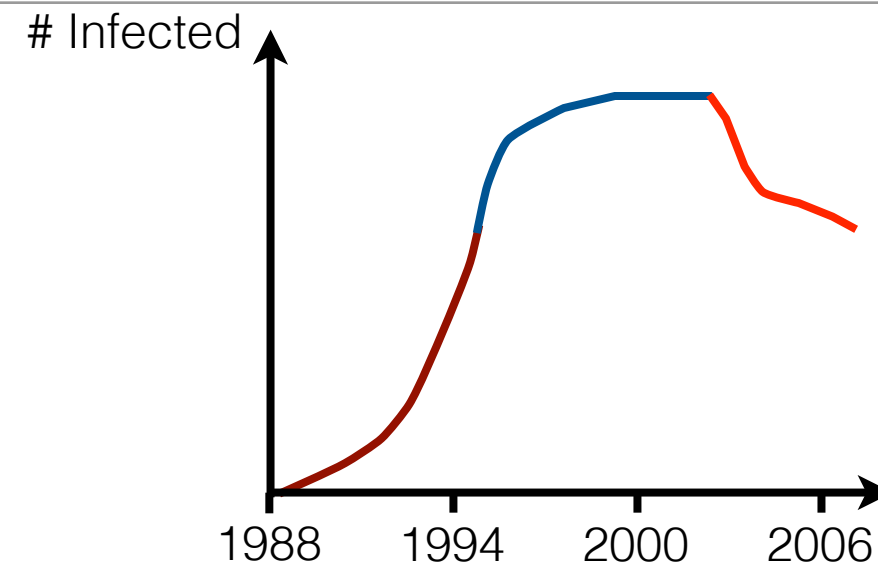
40-100 sequences per subtype
Fit exponential-growth coalescent to sequences and obtain growth rate r (i.e. $\eta=r$)
 $R_0 = \lambda/\delta = rD + 1$ (where D is expected time of infectiousness)



Results



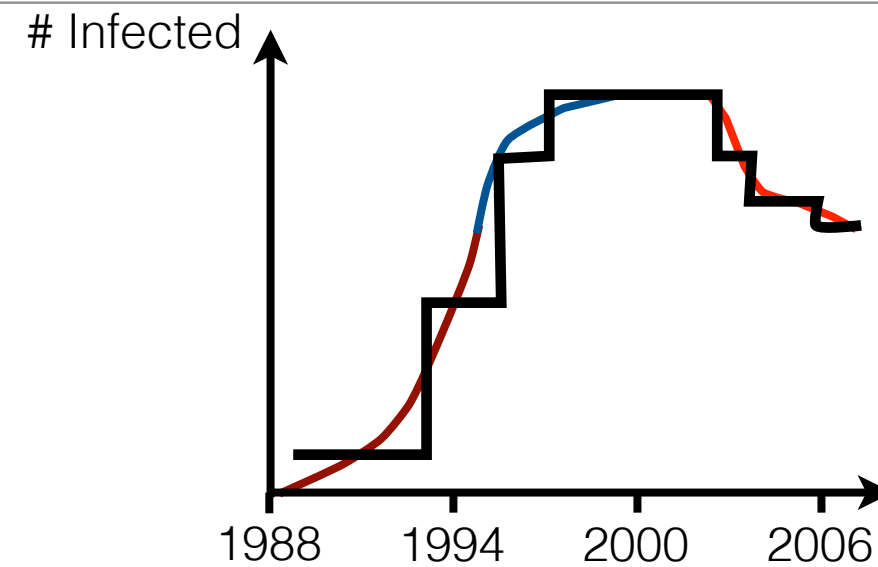
Bayesian skyline plot relaxes constant growth rate



Bayesian skyline plot relaxes constant growth rate

Assump-
tion

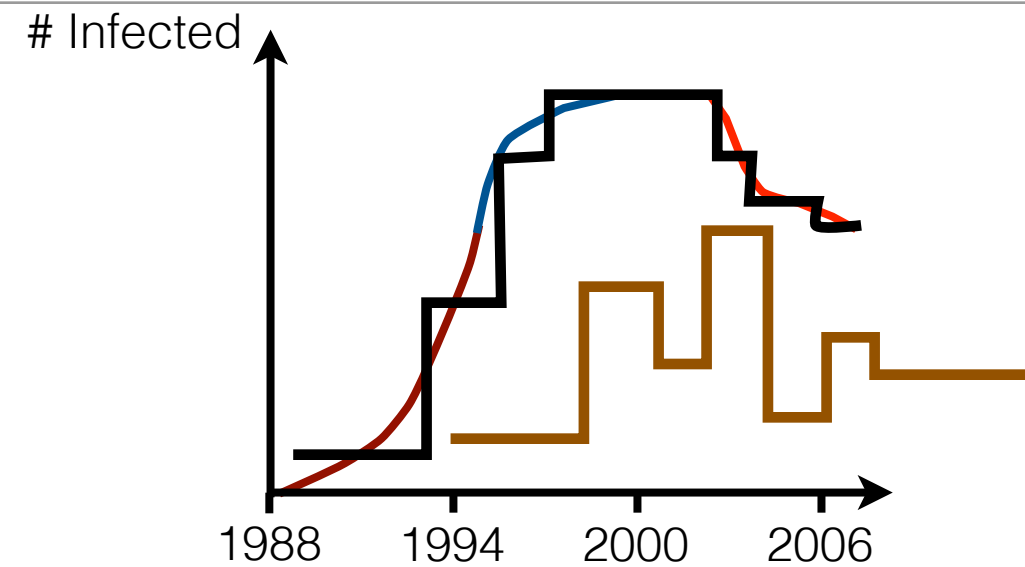
Piecewise constant
population size



Bayesian skyline plot relaxes constant growth rate

Assump-
tion

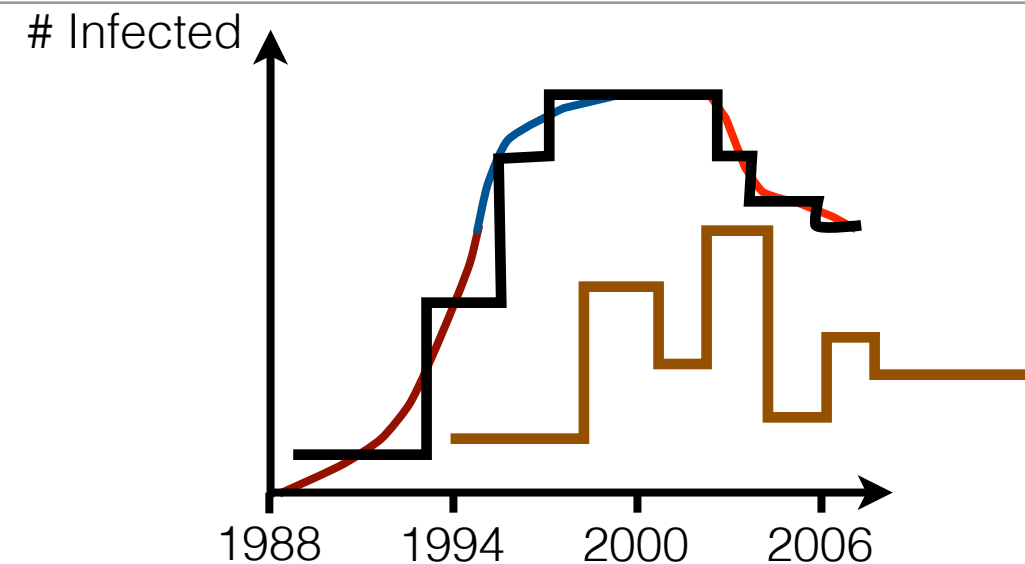
Piecewise constant
population size



Bayesian skyline plot relaxes constant growth rate

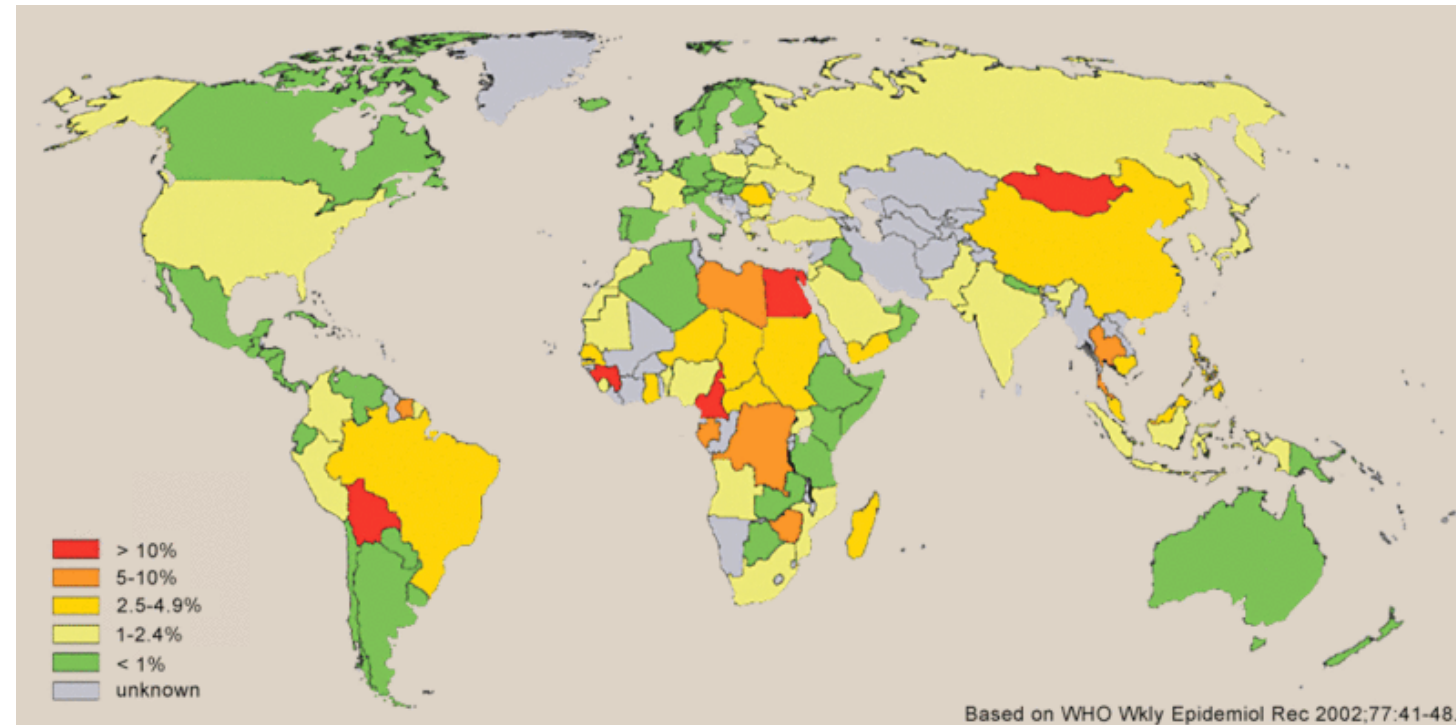
Assump-
tion

Piecewise constant
population size



HCV in
Egypt

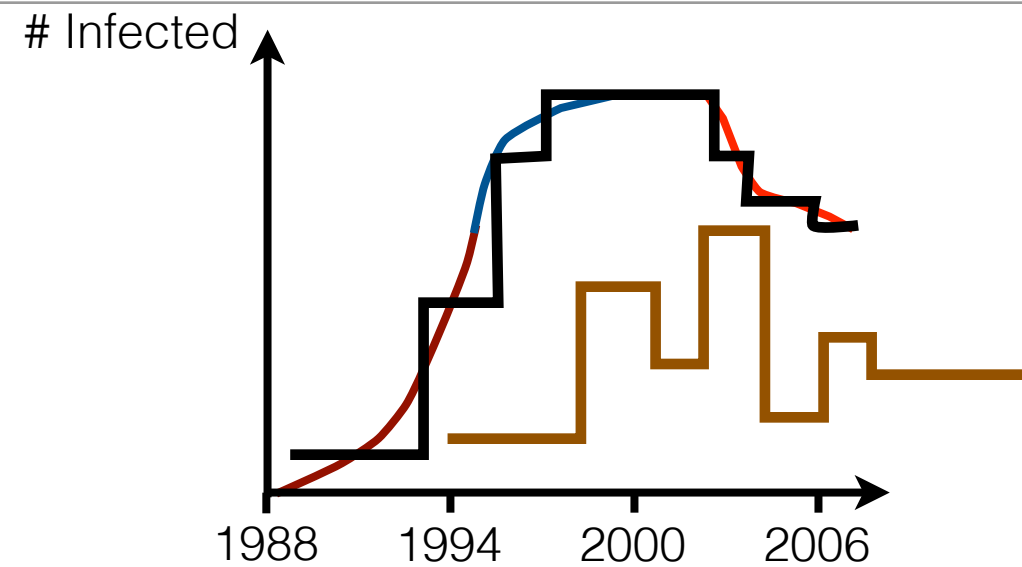
- ▶ Egypt highest HCV prevalence worldwide (about 13 %)
- ▶ Neighboring countries low prevalence
- ▶ When did the high prevalence emerge?
- ▶ Why?



Bayesian skyline plot relaxes constant growth rate

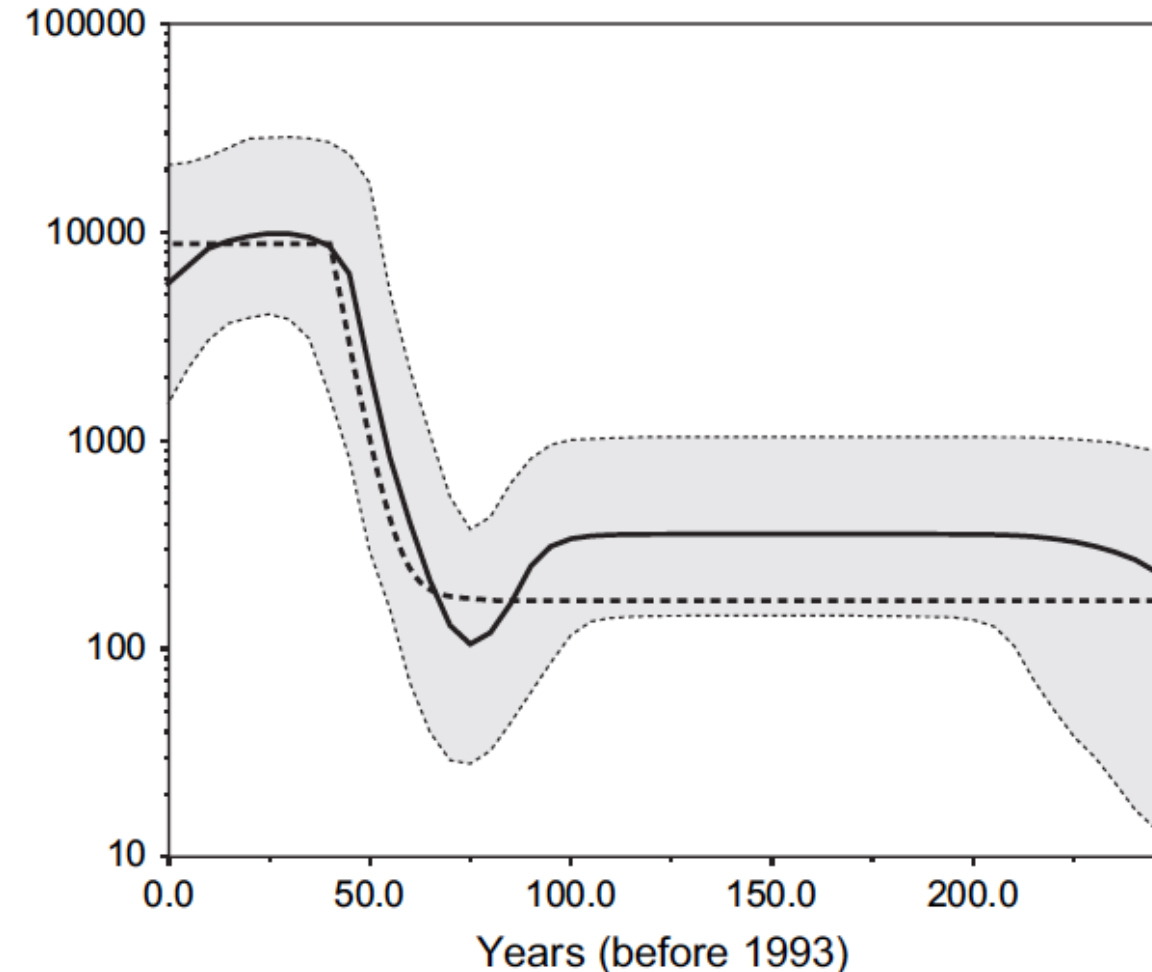
Assump-
tion

Piecewise constant
population size

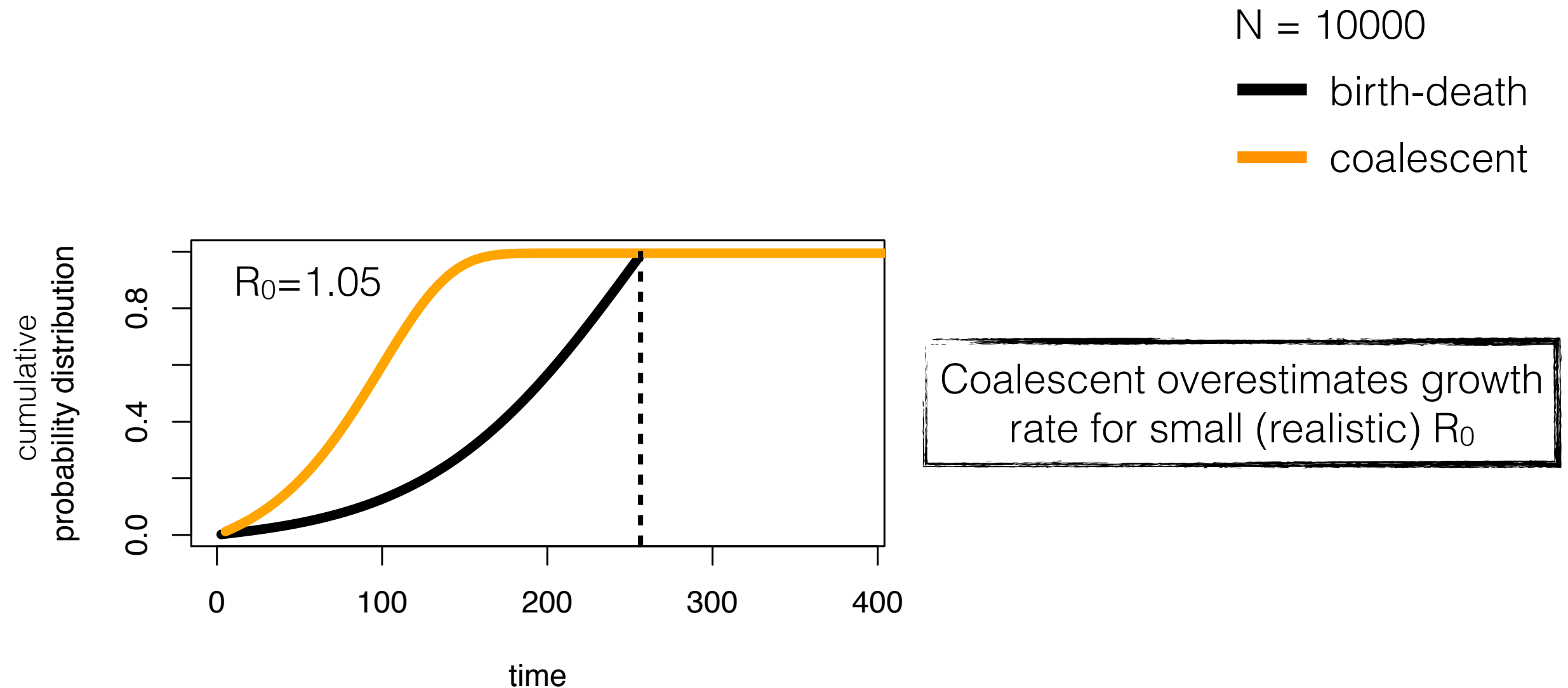


HCV in
Egypt

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



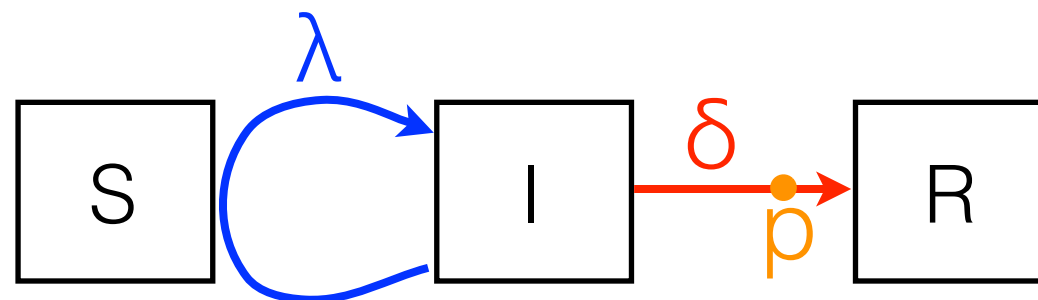
Comparing coalescent times under the birth-death model and the coalescent



Part 2

Birth-death model as model for transmission

Birth-death
model



- transmission rate λ
- “becoming-non-infectious” rate δ
- sampling probability p

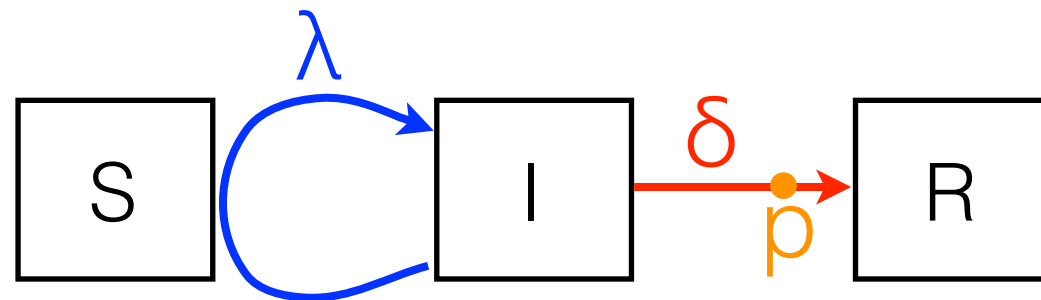
Parameters may depend on:

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

Part 2

Birth-death model as model for transmission

Birth-death model

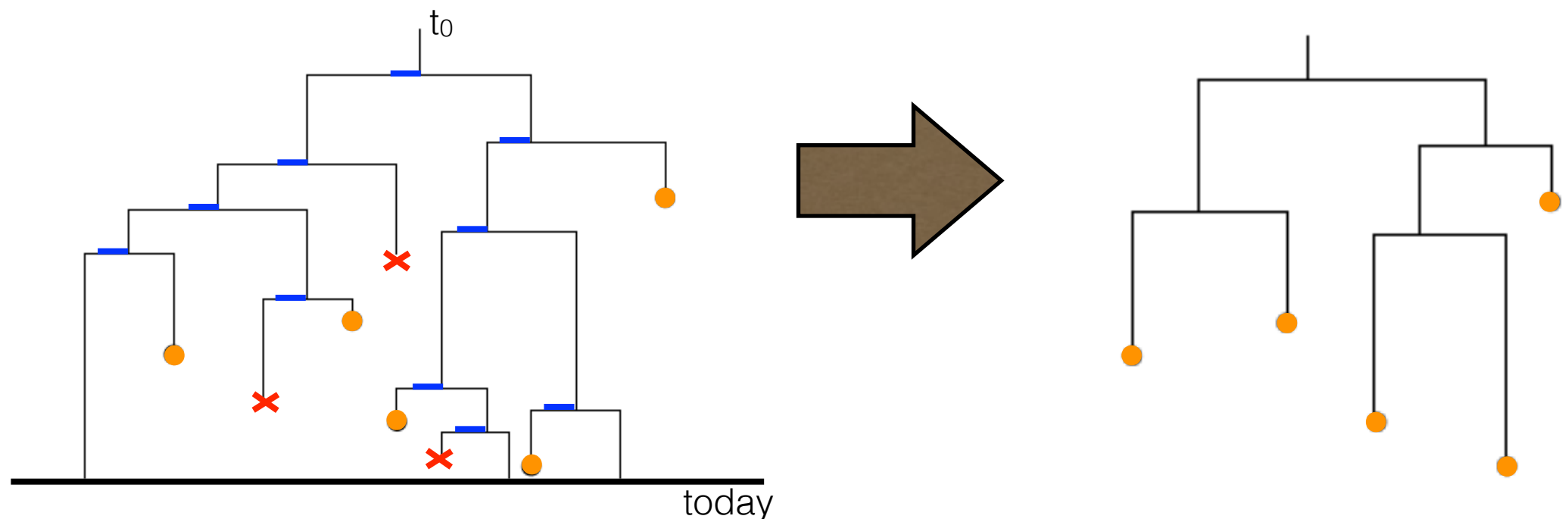


- transmission rate λ
- “becoming-non-infectious” rate δ
- sampling probability p

Parameters may depend on:

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

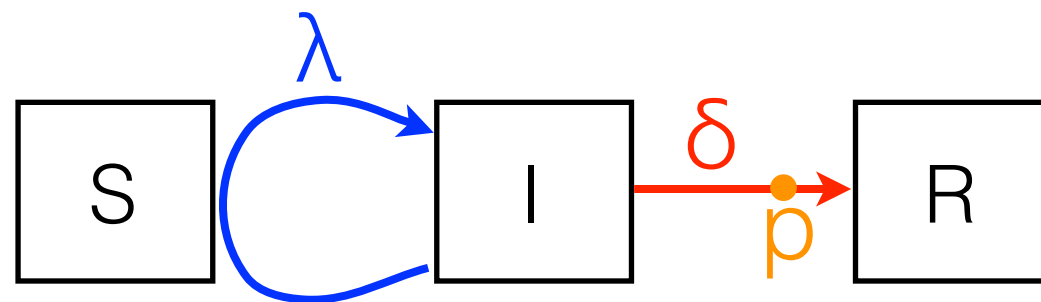
Phylogenetic trees



Part 2

Birth-death model as model for transmission

Birth-death model



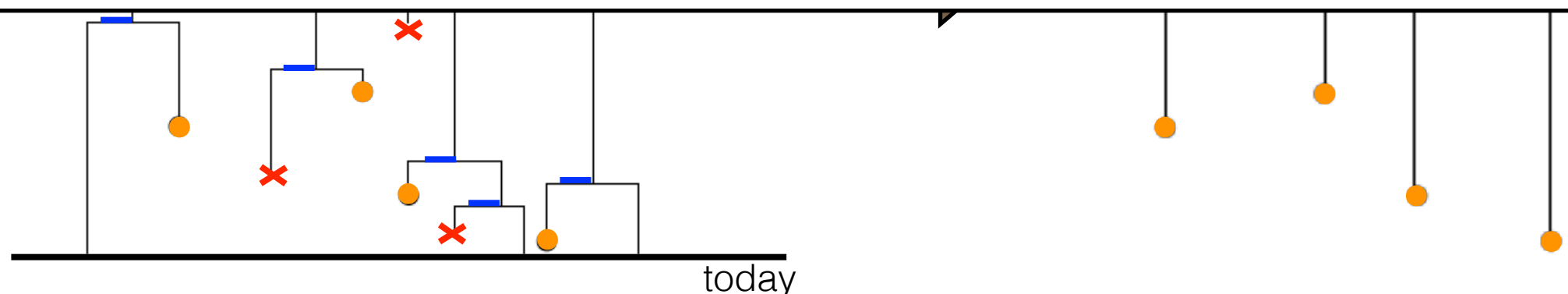
- transmission rate λ
- “becoming-non-infectious” rate δ
- sampling probability p

Parameters may depend on:

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

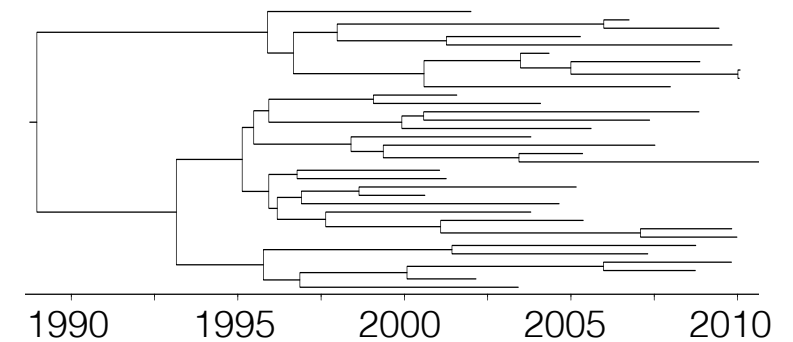
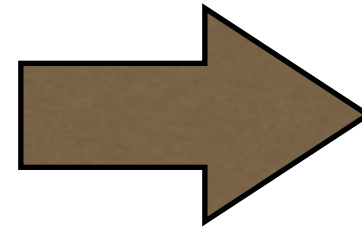
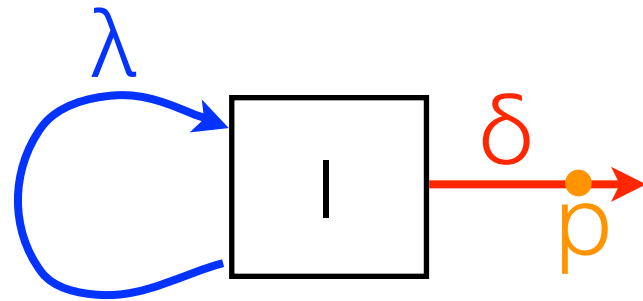
$$f(\text{tree}) ?$$

genetic trees



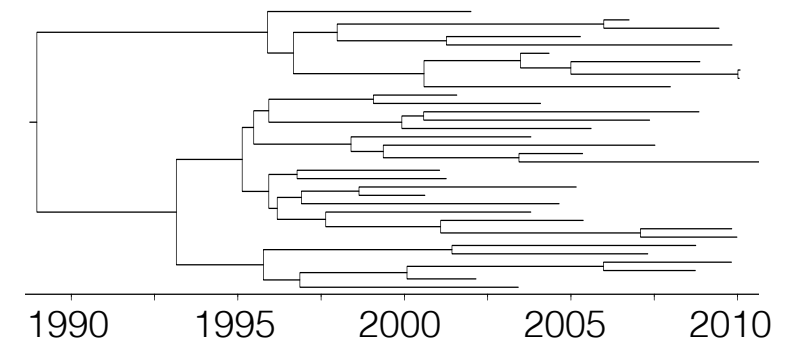
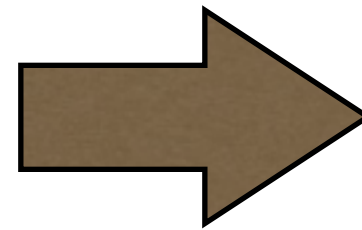
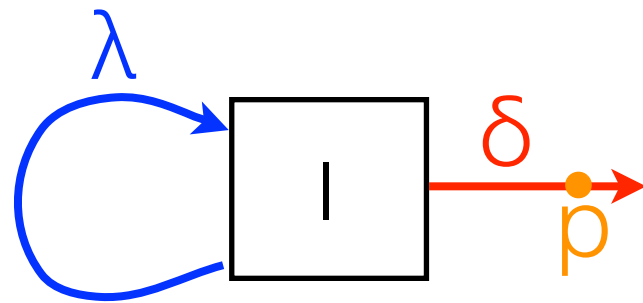
Simulation study: Comparing exact method to coalescent

Simulating
epidemic
outbreak



Simulation study: Comparing exact method to coalescent

Simulating
epidemic
outbreak



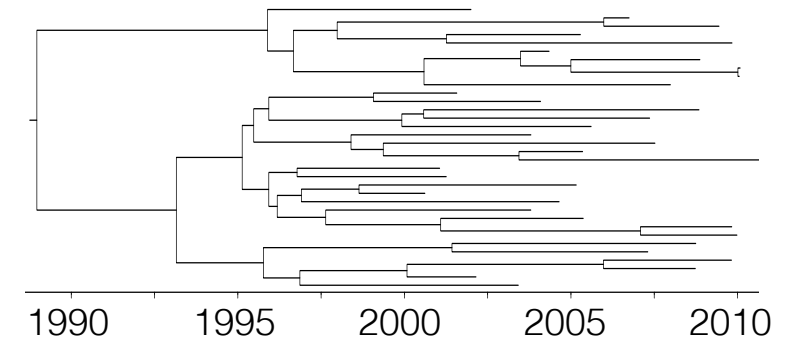
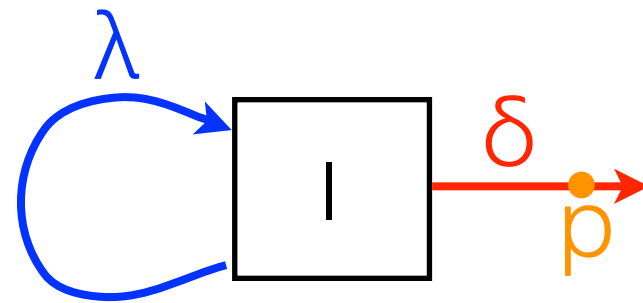
Exact
method

$\lambda - \delta$ 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

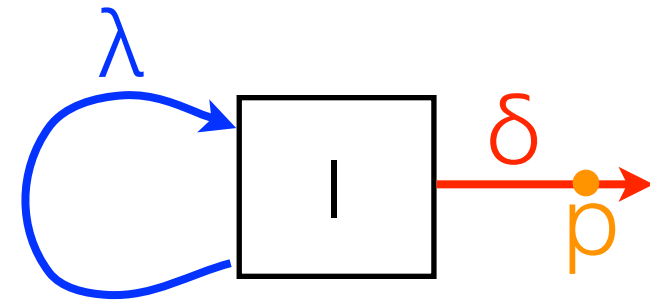
Coales-
cent
approx-
imation

Parameterizes deterministic population size growth

- ▶ 95% *HPD* interval contains true value in **55%** of the simulated trees

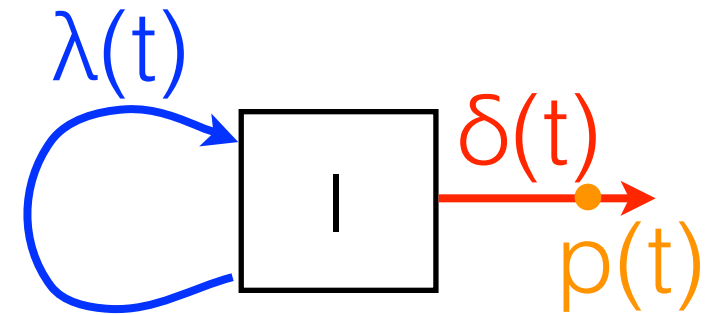
Piecewise constant birth-death model, or: Birth-death skyline plot

Epidemiological rates may change through time:



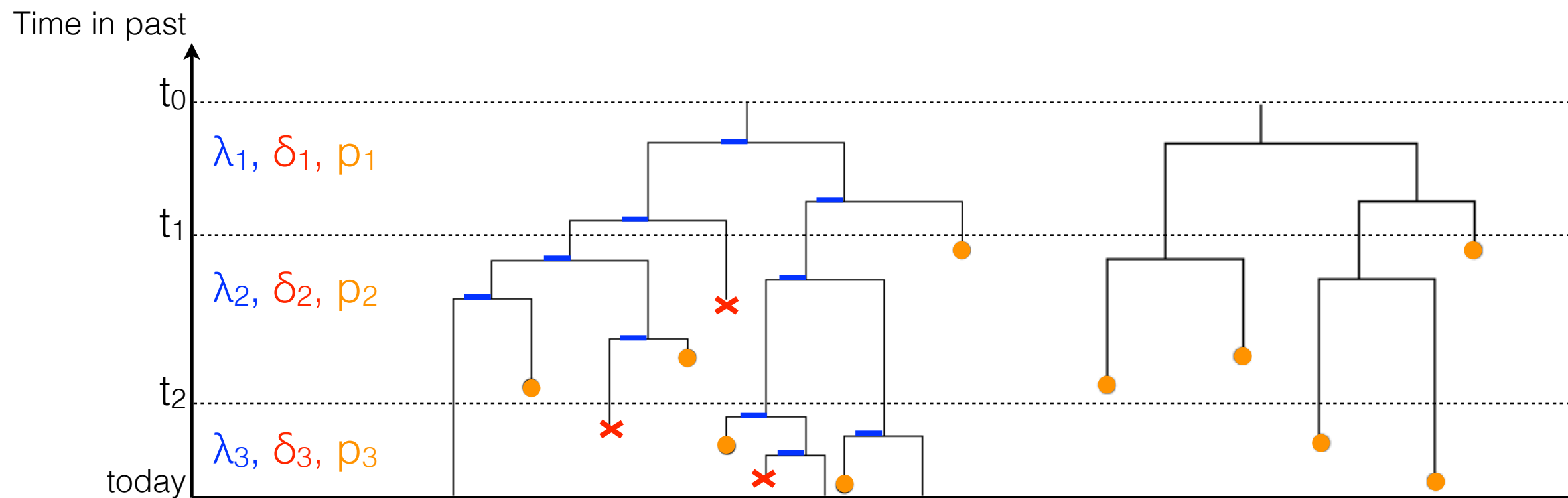
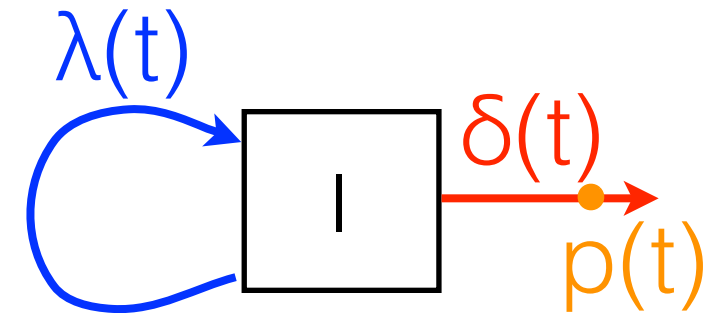
Piecewise constant birth-death model, or: Birth-death skyline plot

Epidemiological rates may change through time:

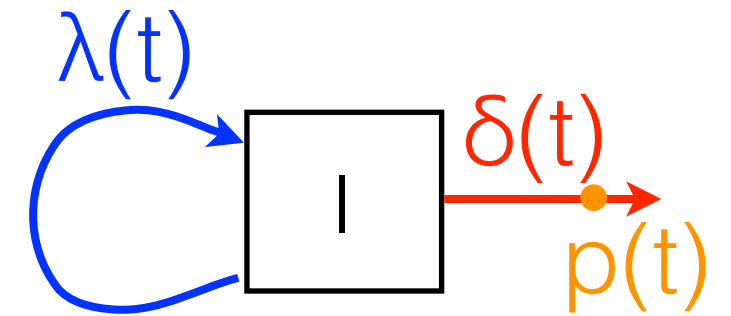


Piecewise constant birth-death model, or: Birth-death skyline plot

Epidemiological rates may change through time:



Hepatitis C virus in Egypt

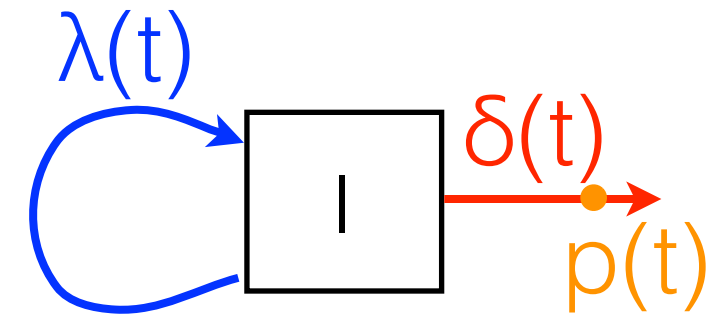


Effective reproductive number **$R_e(t)$**

generalizes

Basic reproductive number **$R_0 = R_e(0)$**

Hepatitis C virus in Egypt



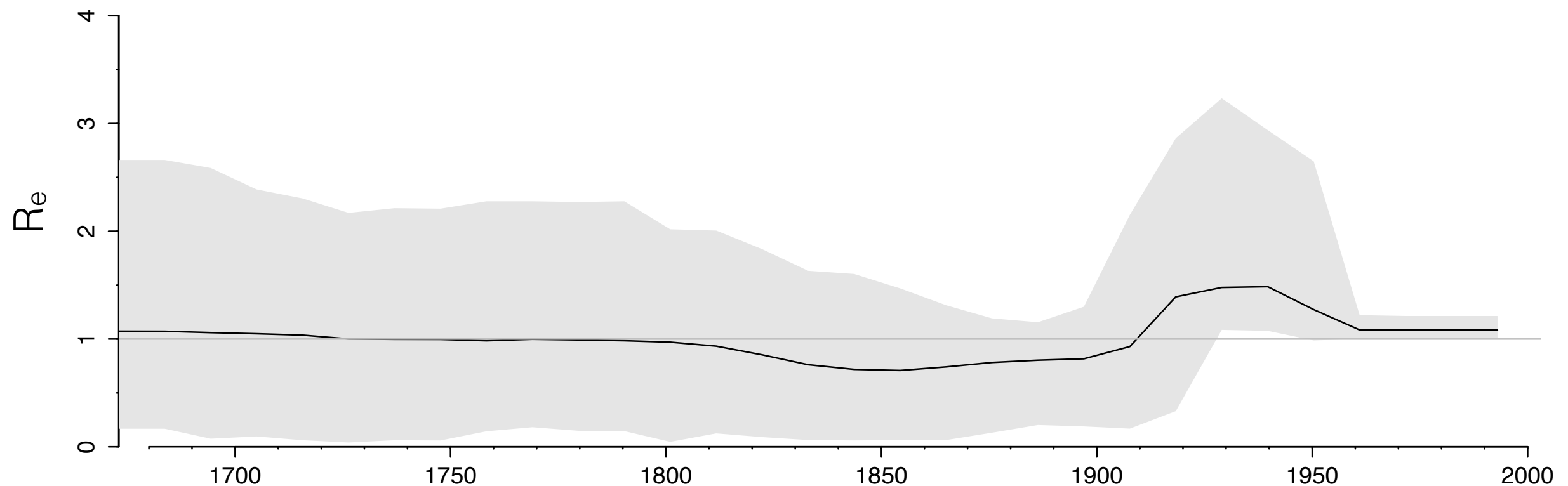
Effective reproductive number $R_e(t)$
generalizes
Basic reproductive number $R_0 = R_e(0)$

Analysis of 63 seq:

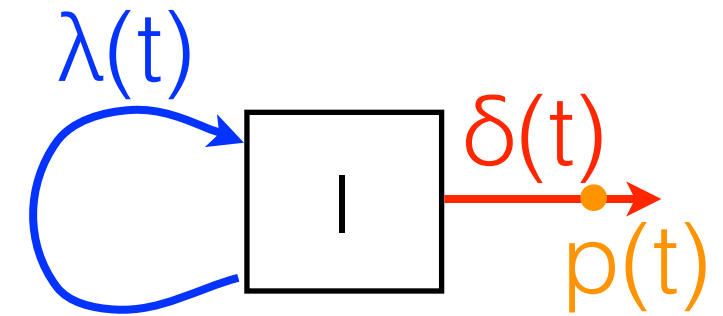
$$R_e(t) = \lambda(t)/\delta(t)$$

— median

■ 95% HPD interval



Hepatitis C virus in Egypt



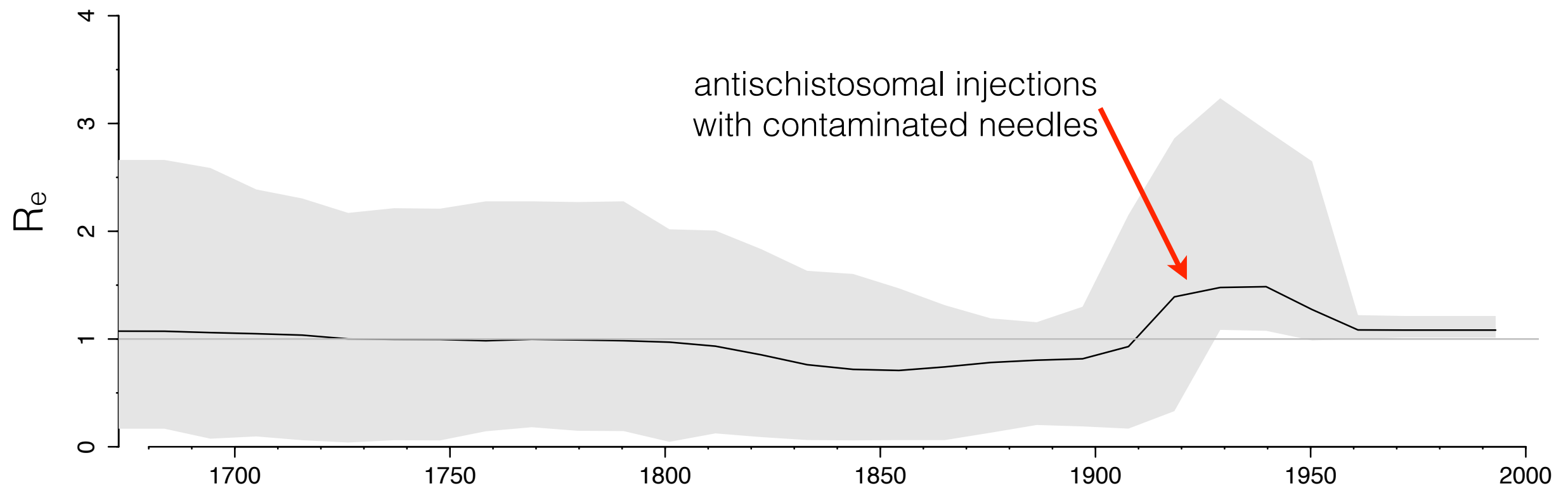
Effective reproductive number **$R_e(t)$**
generalizes
Basic reproductive number **$R_0 = R_e(0)$**

Analysis of 63 seq:

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

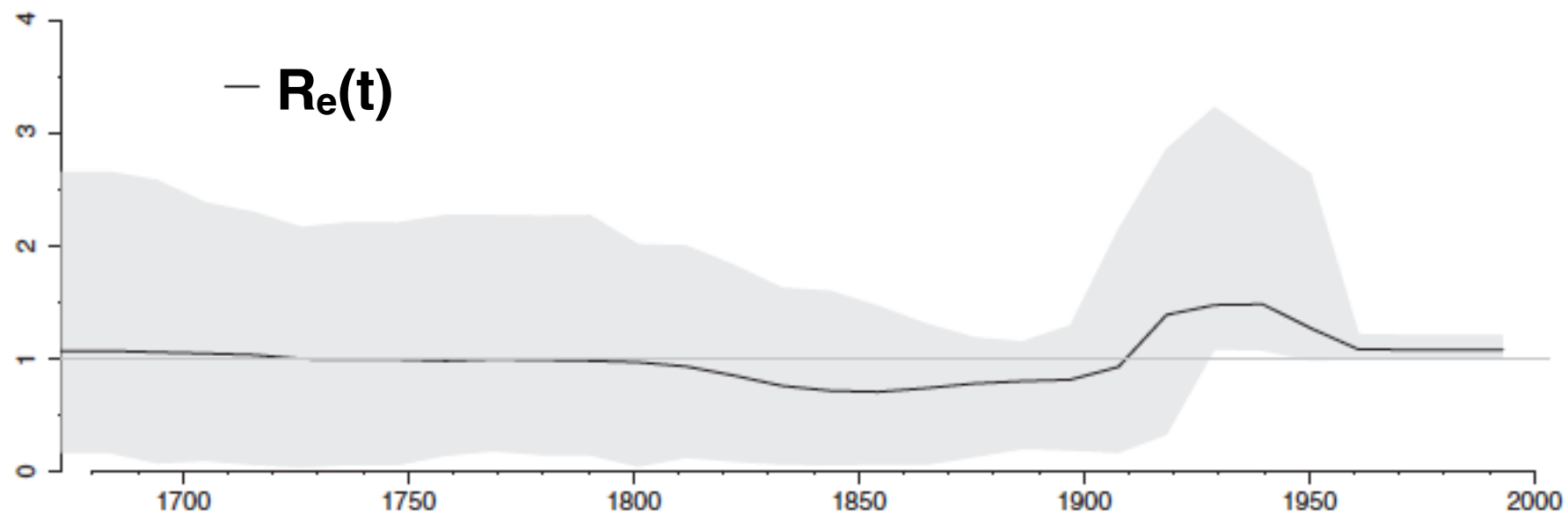
— median

■ 95% HPD interval



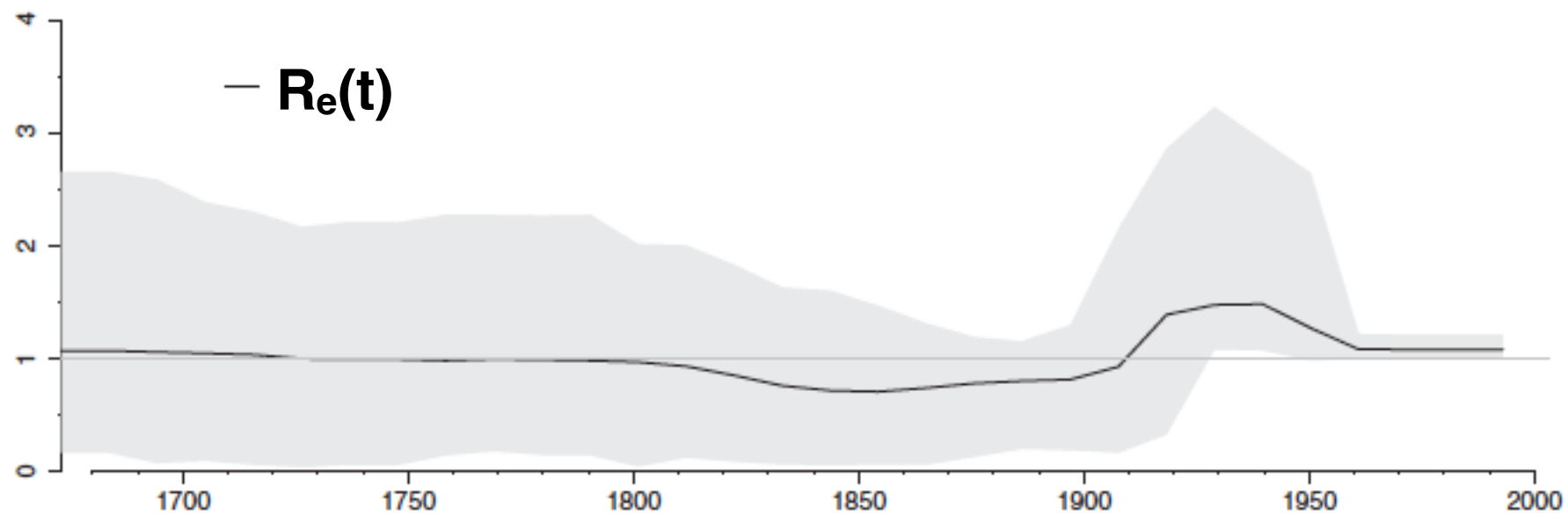
The two skyline models

Birth-death skyline plot: effective reproductive number

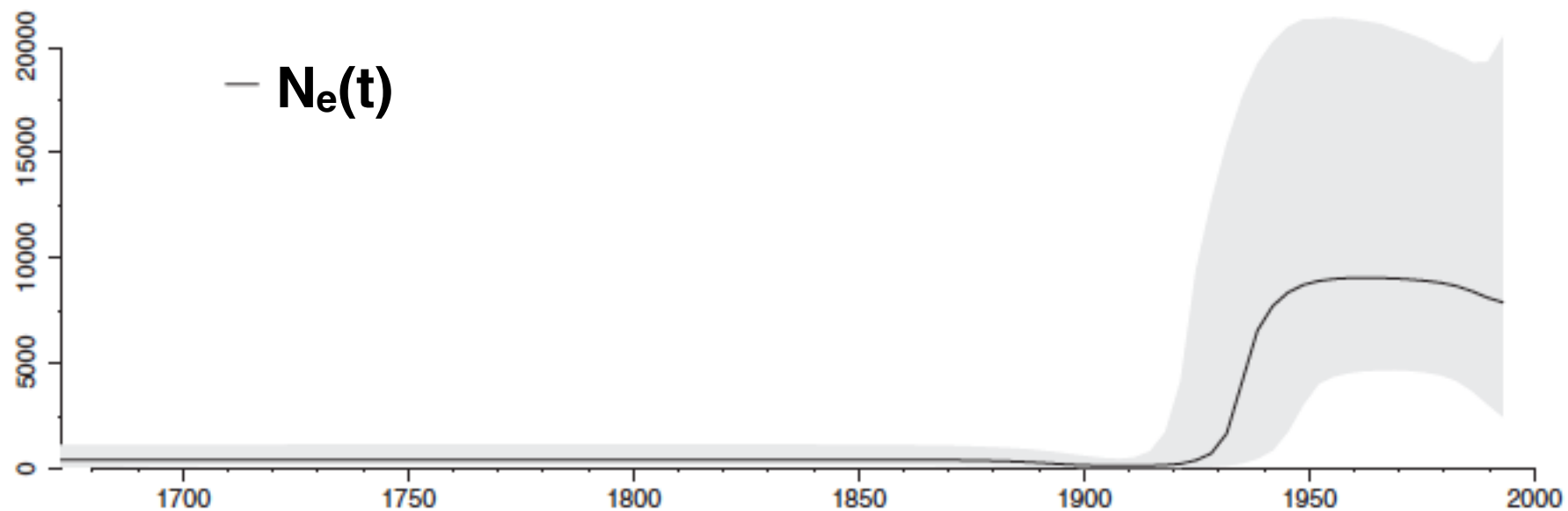


The two skyline models

Birth-death skyline plot: effective reproductive number



Coalescent skyline plot: effective population size



Exercises: Analysis of HCV dataset using RevBayes

Data

HCV sequencing data from 1993

MCMC

Infer trees and parameters using the birth-death skyline model