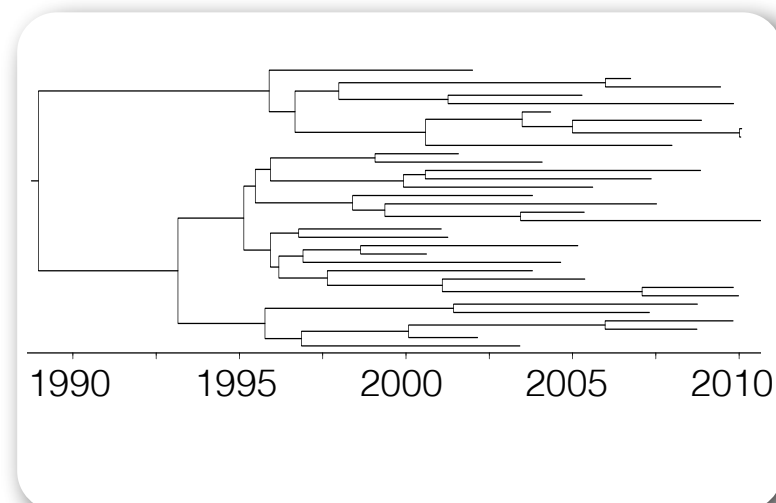# Infectious Disease Dynamics in RevBayes
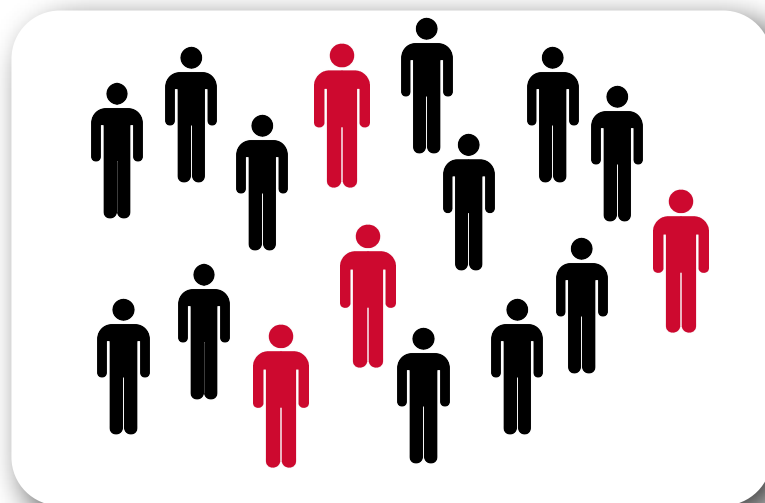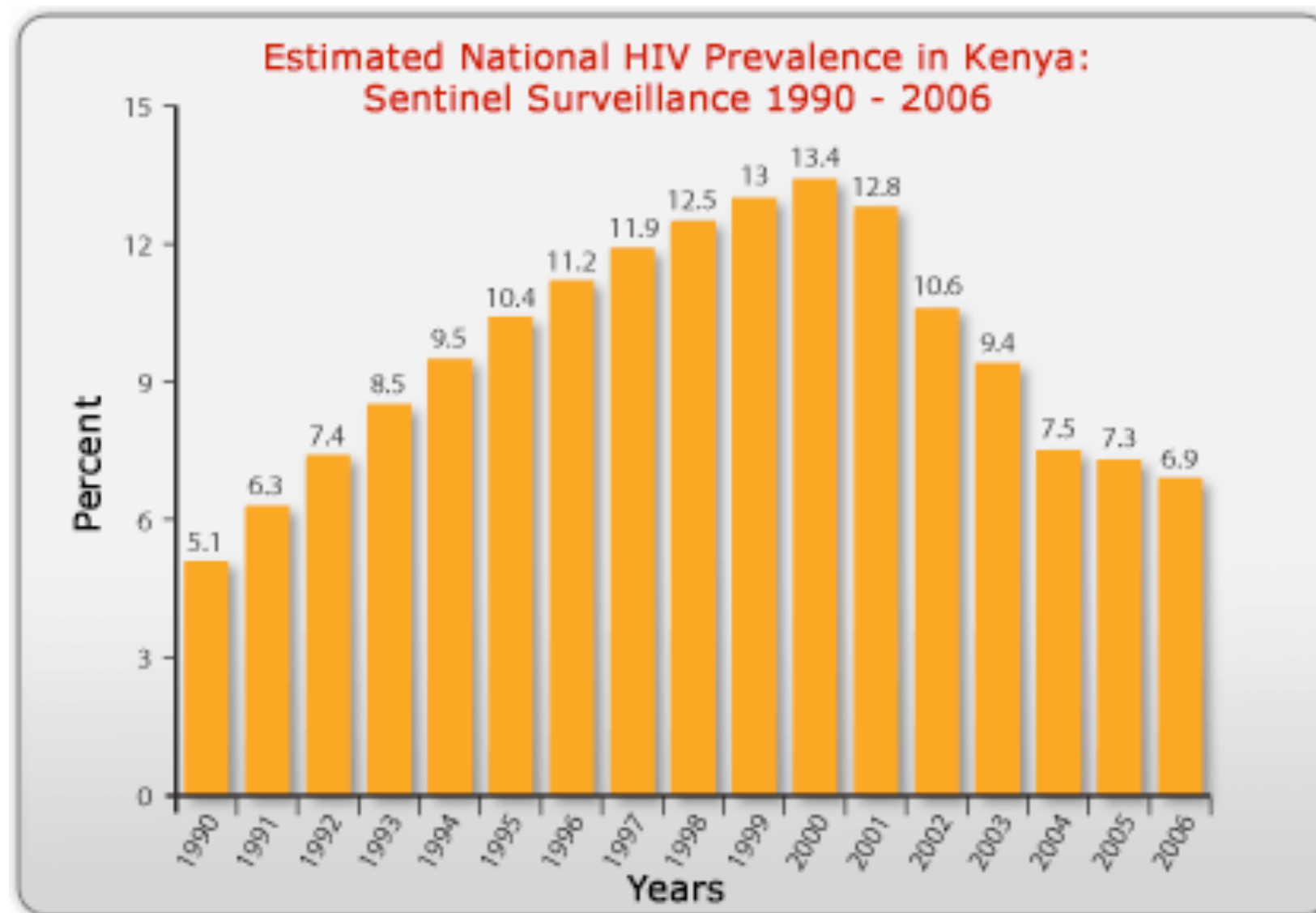
Inferring epidemiological dynamics based on genetic sequence data

Tracy Heath & Tanja Stadler

Tanja Stadler
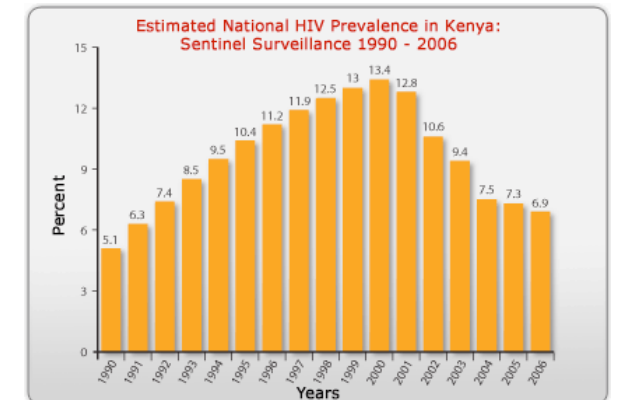Computational Evolution group — Department of Biosystems Science and Engineering
tanja.stadler@bsse.ethz.ch

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Epidemiology

Estimated National HIV Prevalence in Kenya: Sentinel Surveillance 1990 - 2006
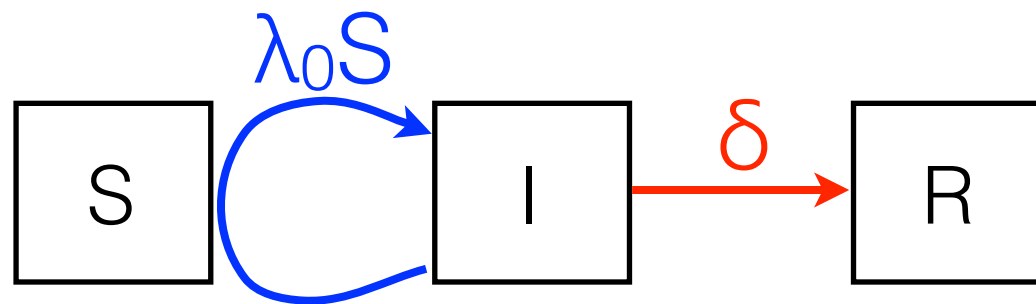
Epidemiology

Data

\# of infected individuals through time

Models

Population dynamics described by SIR models:

$$\frac{dS}{dt} = -\lambda_0 IS/N$$

$$\frac{dI}{dt} = \lambda_0 IS/N - \delta I$$

$$\frac{dR}{dt} = \delta I$$

Epidemiology

Data

\# of infected individuals through time

Estimated National HIV Prevalence in Kenya:
Sentinel Surveillance 1990 - 2006
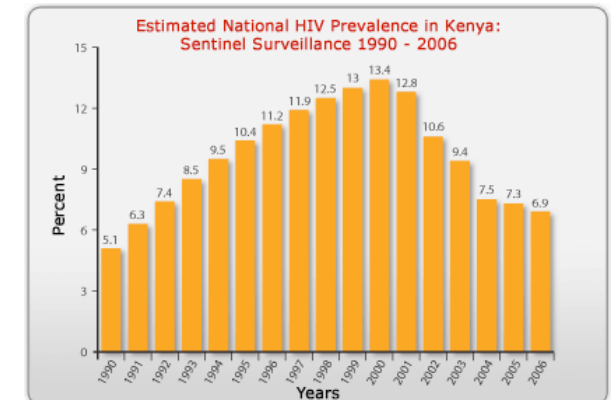
Models

Population dynamics described by SIR models:

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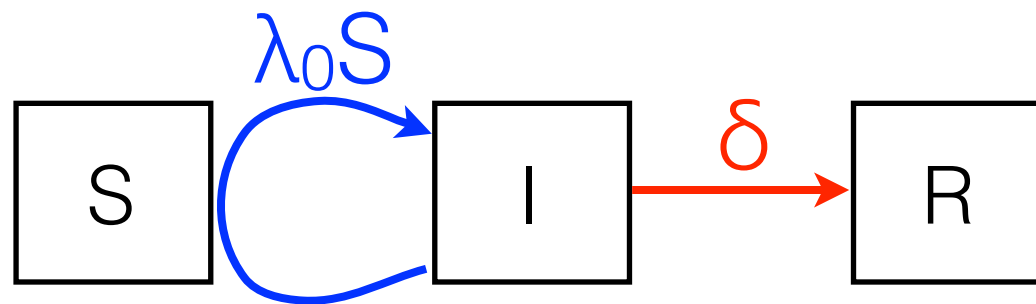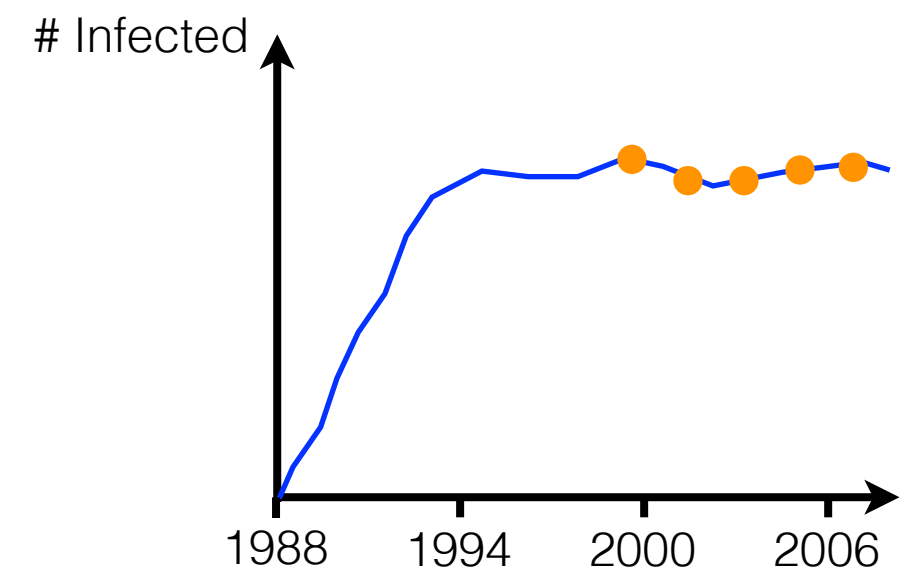\frac{dR}{dt} &= \delta I
\end{aligned}
$$

$\lambda_0 S$
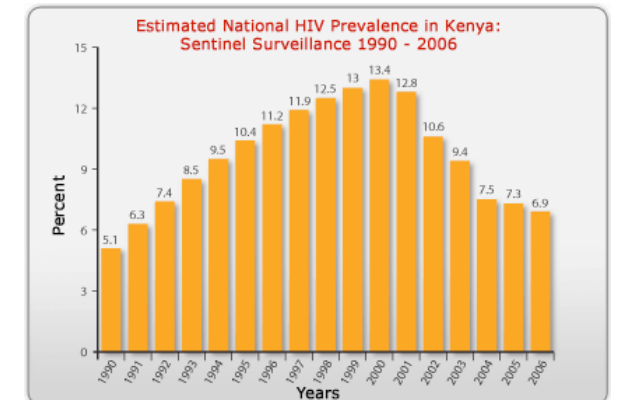
δ

S   I   R

Limitations

If sampling in early epidemic was missed:

\# Infected

1988   1994   2000   2006

Epidemiology

Data

\# of infected individuals through time

Estimated National HIV Prevalence in Kenya:
Sentinel Surveillance 1990 - 2006

Models

Population dynamics described by SIR models:

$$\frac{dS}{dt} = -\lambda_0 IS/N$$

$$\frac{dI}{dt} = \lambda_0 IS/N - \delta I$$

$$\frac{dR}{dt} = \delta I$$

$\lambda_0 S$

S   I δ R

Limitations

If sampling in early epidemic was missed:

‣ **Time of epidemic outbreak?**

‣ **Basic reproductive number R_0?**

\# Infected

1988   1994   2000   2006

Epidemiology

Data

\# of infected individuals through time

Estimated National HIV Prevalence in Kenya: Sentinel Surveillance 1990 - 2006

Models

Population dynamics described by SIR models:

$$\lambda_0 S$$

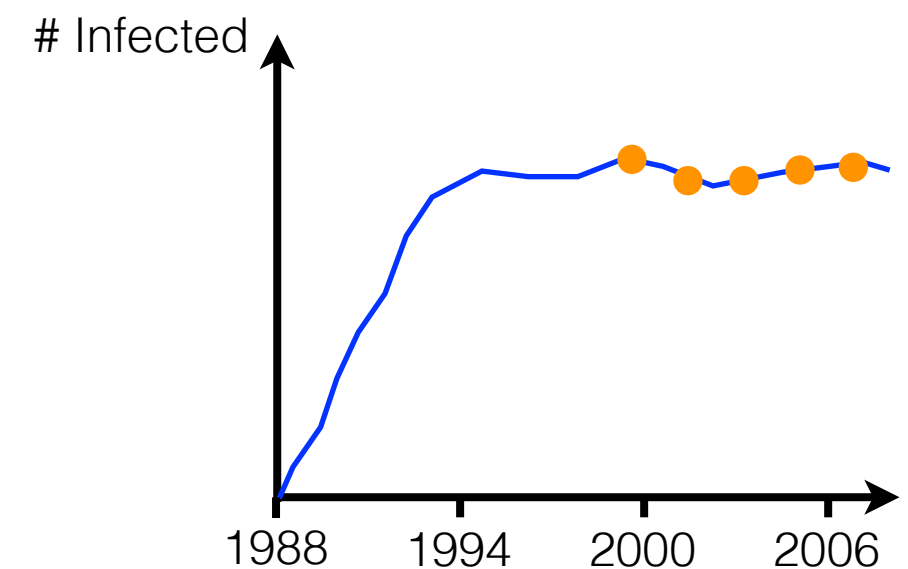$$S \quad \xrightarrow{\quad} \quad I \quad \xrightarrow{\delta} \quad R$$

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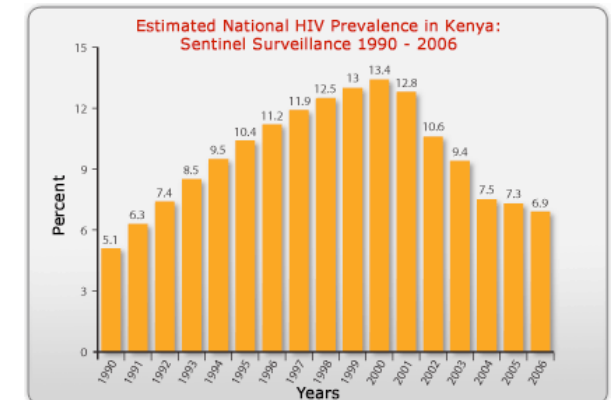
Limitations

If sampling in early epidemic was missed:

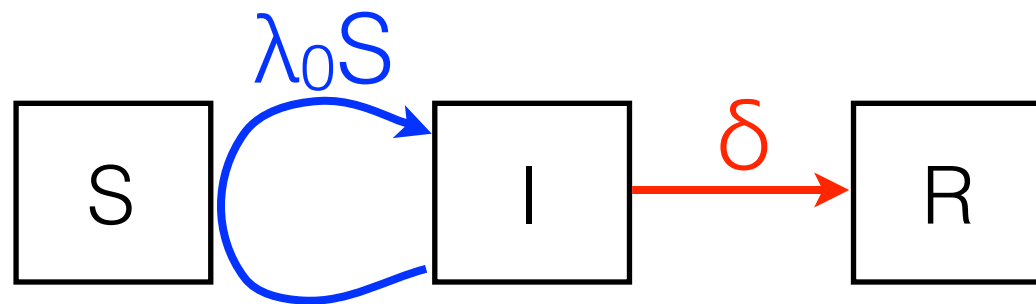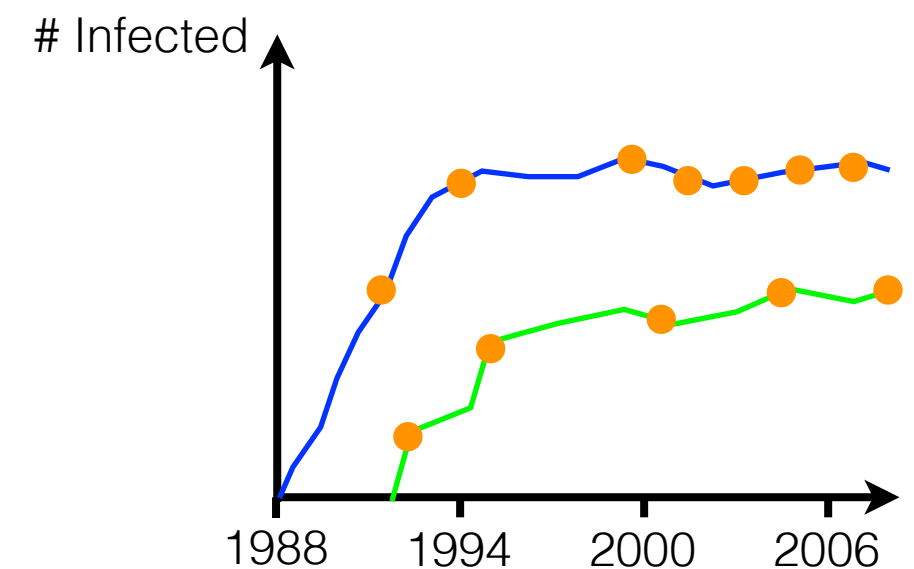▸ **Time of epidemic outbreak?**

▸ **Basic reproductive number R_0?**

Data does not tell who infected whom:

▸ **Population structure?**

\# Infected

1988   1994   2000   2006

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`
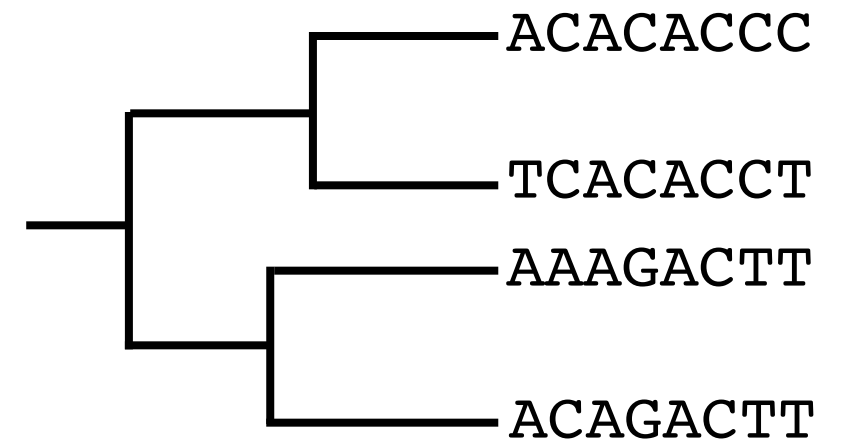
Genetic data entering epidemiology
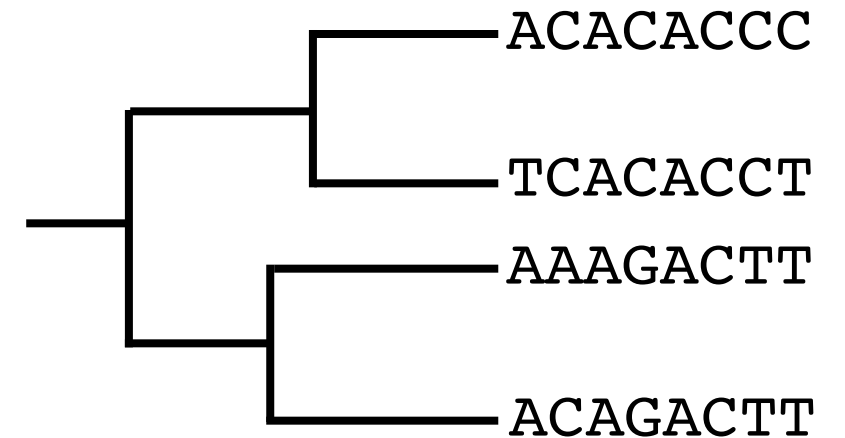
Phylo-
genetics

Input:

Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)

ACACACCC

TCACACCT

AAAGACTT

ACAGACTT

Genetic data entering epidemiology

Phylo-genetics

Input:

Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)

```
                    ┌──── ACACACCC
              ┌─────┤
    ┌─────────┤     └──── TCACACCT
    │         │
────┤         └───── AAAGACTT
    │         ┌────
    └─────────┤
              └──── ACAGACTT
```

Example key result

Emergence of HIV

Epidemic identified in 1980s

Genetic data entering epidemiology

Phylo-genetics

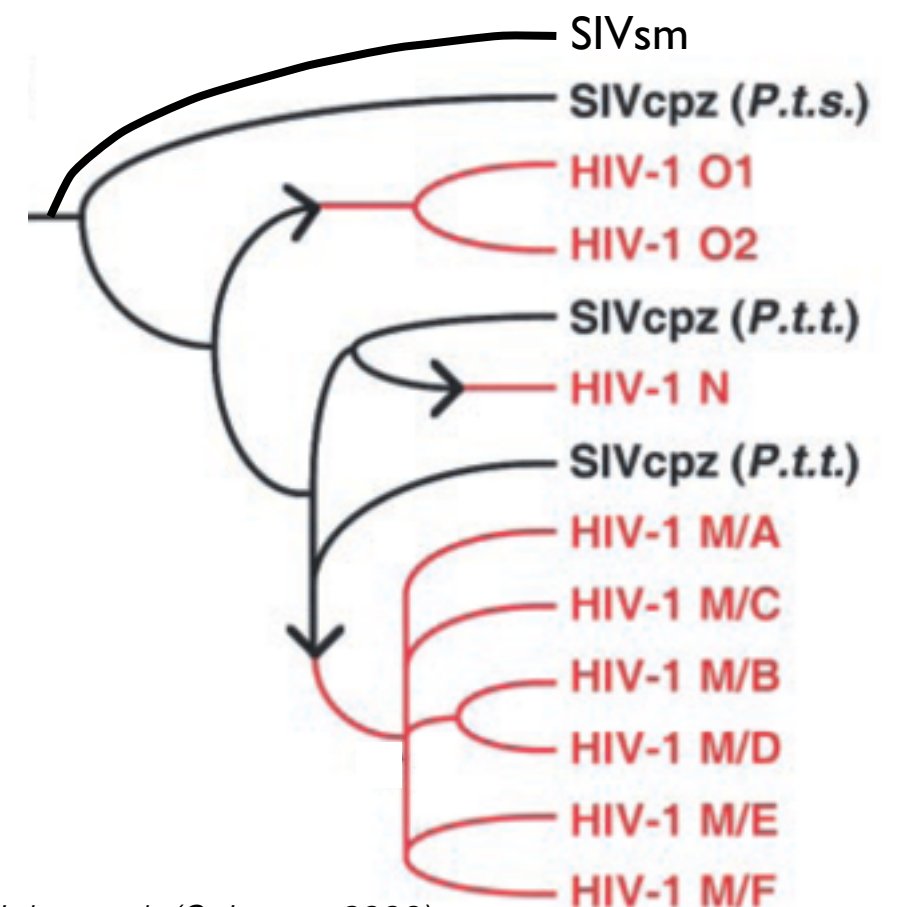Input:

Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)


```
                    ┌──── ACACACCC
              ┌─────┤
              │     └──── TCACACCT
        ──────┤
              │     ┌──── AAAGACTT
              └─────┤
                    └──── ACAGACTT
```

Example key result

Emergence of HIV

Epidemic identified in 1980s

- SIVsm
- SIVcpz (*P.t.s.*)
- HIV-1 O1
- HIV-1 O2
- SIVcpz (*P.t.t.*)
- HIV-1 N
- SIVcpz (*P.t.t.*)
- HIV-1 M/A
- HIV-1 M/C
- HIV-1 M/B
- HIV-1 M/D
- HIV-1 M/E
- HIV-1 M/F

adopted from Hahn et al. (Science, 2000)

Genetic data entering epidemiology

Phylo-genetics

Input:

Pathogen genetic data from different hosts

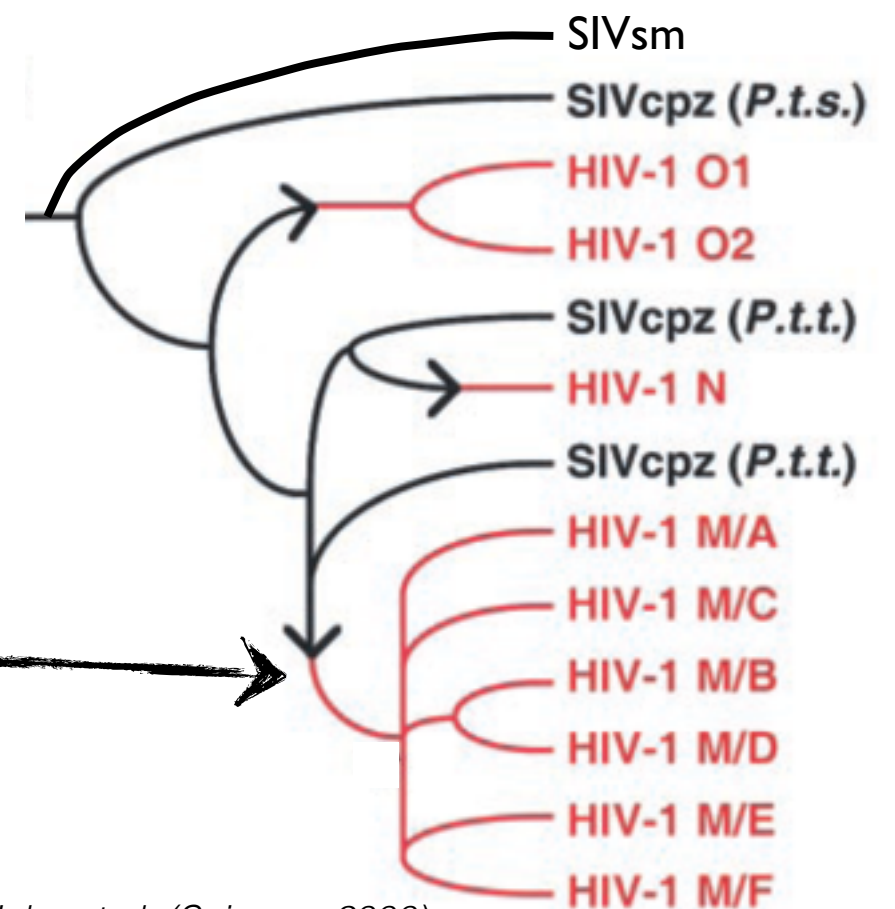Output:

Transmission chain (who infected whom)

```
                        ┌──── ACACACCC
                  ┌─────┤
            ┌─────┤     └──── TCACACCT
            │     │
  ──────────┤     ┌──────────── AAAGACTT
            └─────┤
                  └──────────── ACAGACTT
```
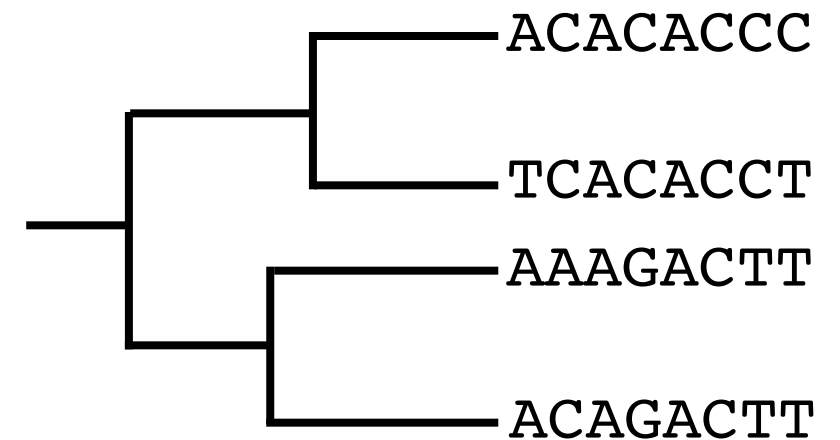
Example key result

Emergence of HIV

Epidemic identified in 1980s

Sequence data tells us about pre-1980

‣ Zoonosis from chimps

SIVsm
SIVcpz (*P.t.s.*)
HIV-1 O1
HIV-1 O2
SIVcpz (*P.t.t.*)
HIV-1 N
SIVcpz (*P.t.t.*)
HIV-1 M/A
HIV-1 M/C
HIV-1 M/B
HIV-1 M/D
HIV-1 M/E
HIV-1 M/F

adopted from Hahn et al. (Science, 2000)

3

Genetic data contains information about the epidemiological dynamics

Phylo-genetics

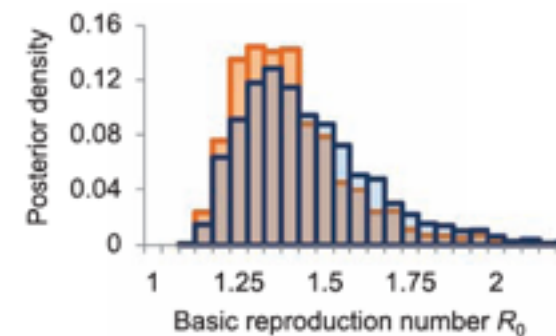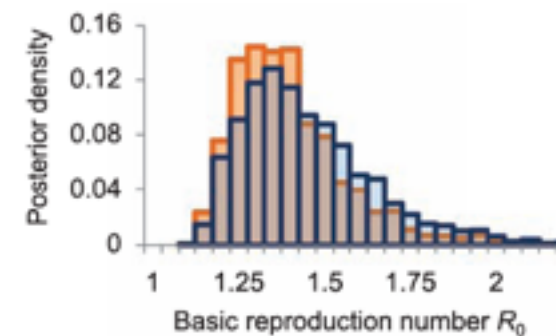State of process

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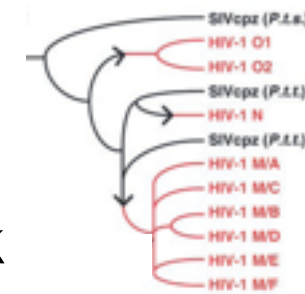
Emergence of a pathogen
(here HIV)
Hahn et al. (Science, 2000)

Phylo-dynamics

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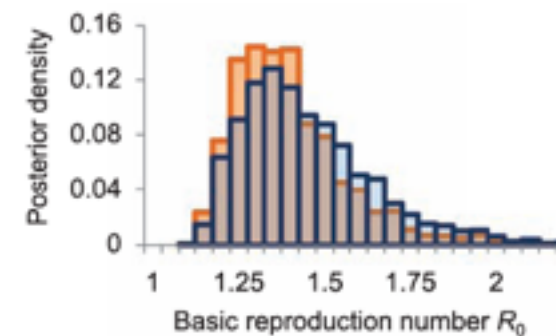
Basic reproductive number of an emerging pathogen
(here H1N1)
Fraser et al. (Science, 2009)

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
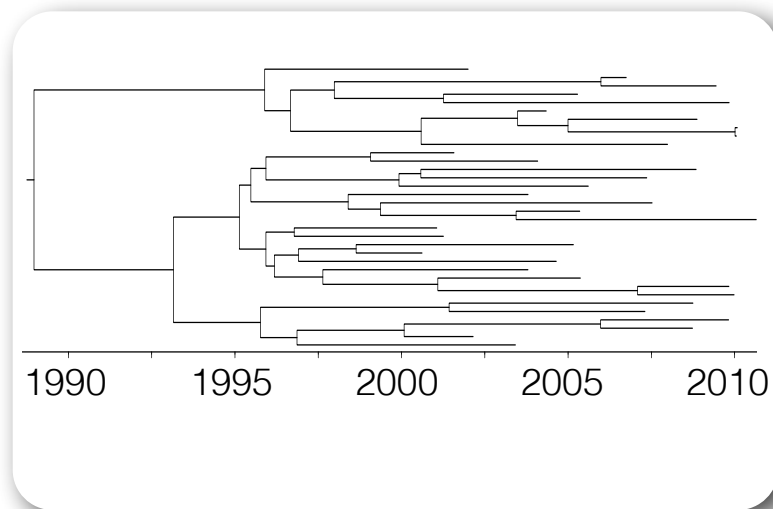
Methodology to:

Simultaneously infer phylogenetic trees and epidemiological parameters

Merge epidemiology and phylogenetics
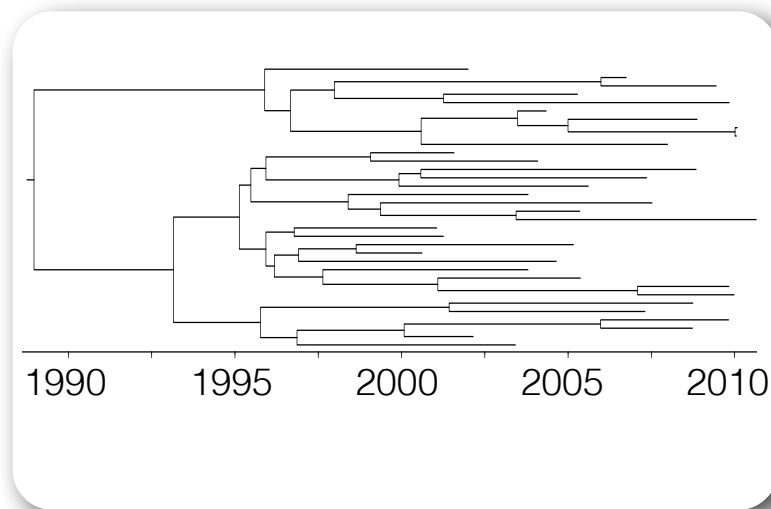
Methodology to:

Simultaneously infer phylogenetic trees and epidemiological parameters
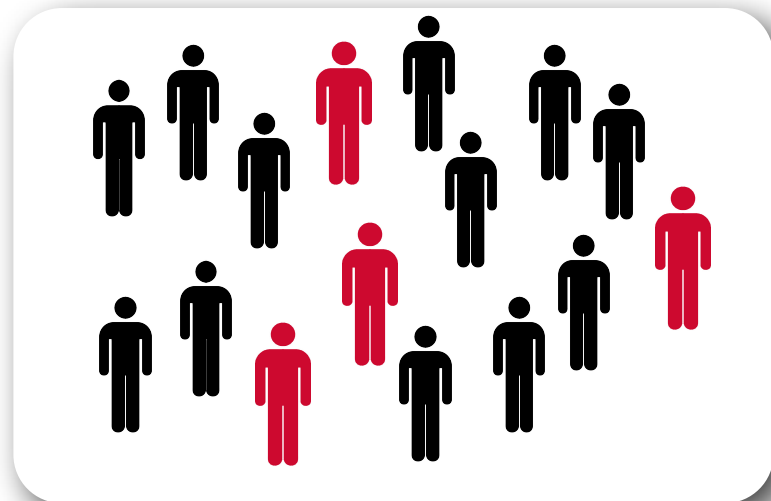
Quantification of spread of infectious diseases?

Factors governing transmission dynamics?
(e.g. population structure, environment, vaccines)

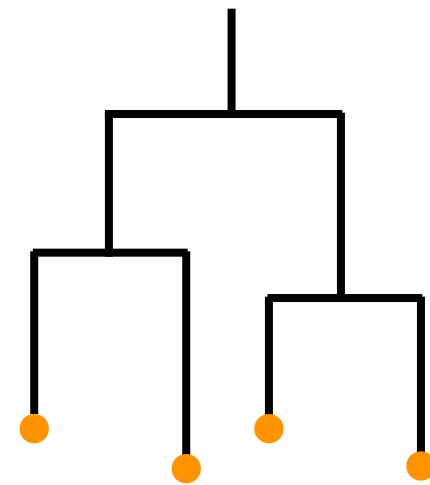# General modeling framework

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

```
ATAGACCC
```

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

ATAGACCC

ACACACCC TCACACCT AAAGACTT ACAGACTT

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
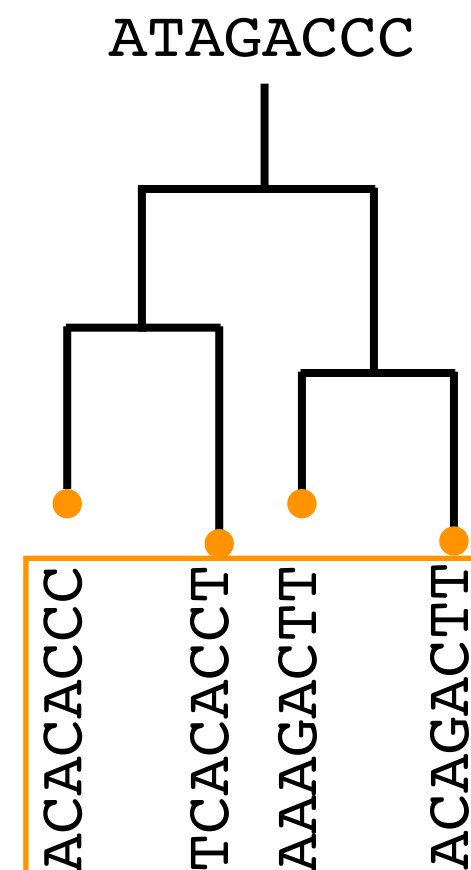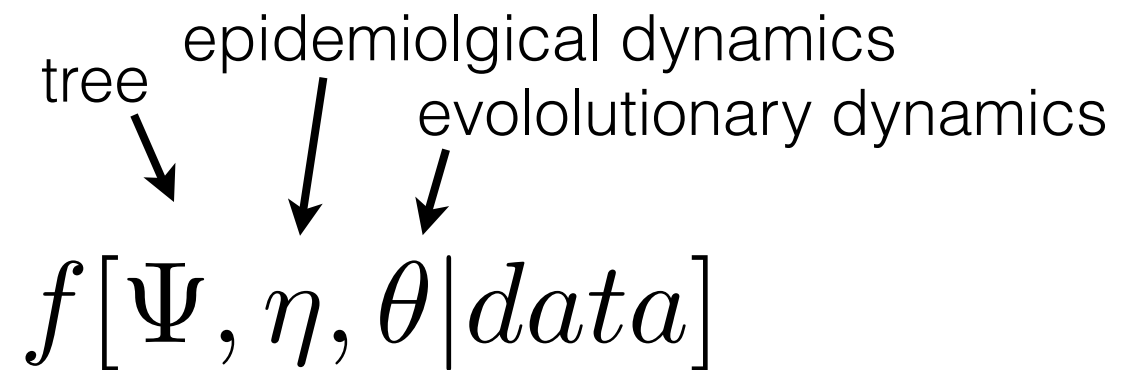
ATAGACCC

ACACACCC TCACACCT AAAGACTT ACAGACTT

Bayesian approach for estimating epidemiological parameters
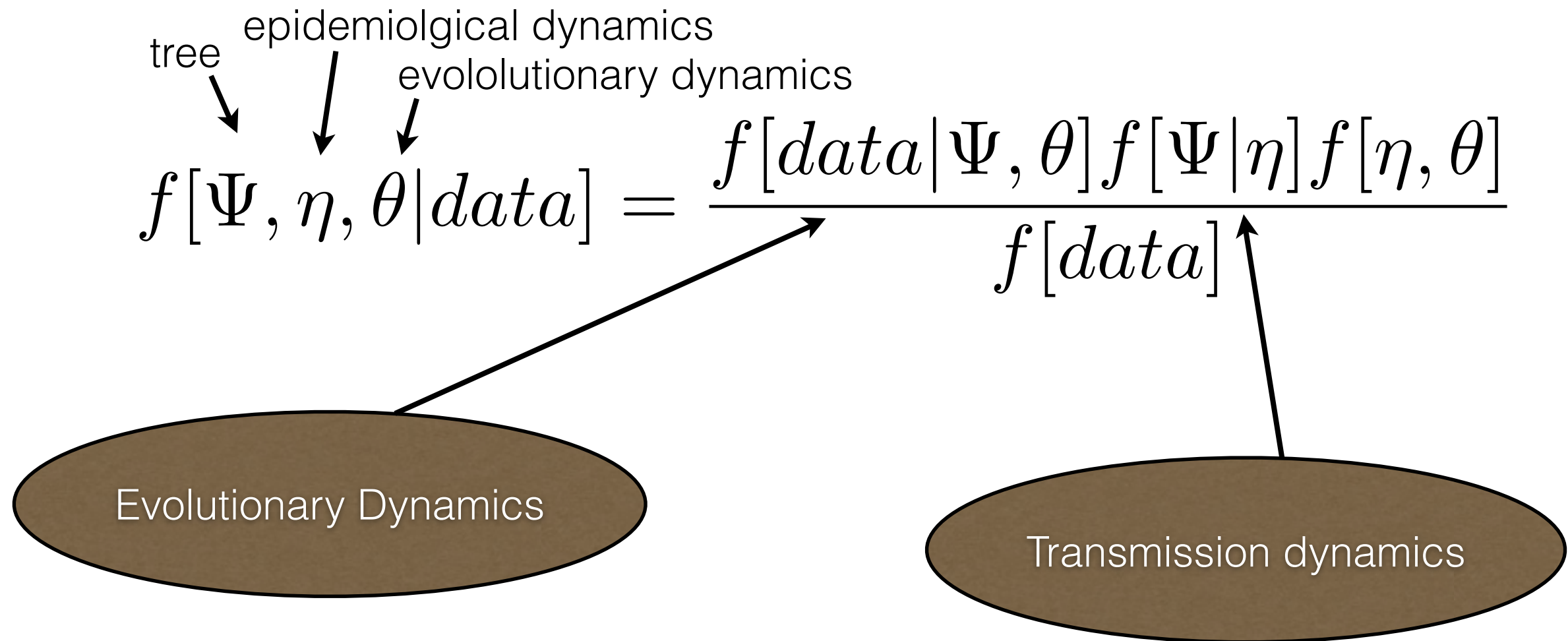
tree epidemiolgical dynamics

evololutionary dynamics

$$f\left[\Psi, \eta, \theta | data\right]$$

Bayesian approach for estimating epidemiological parameters

tree

epidemiolgical dynamics

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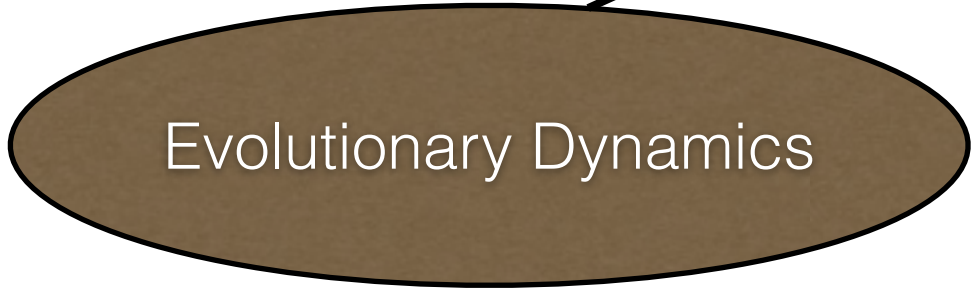
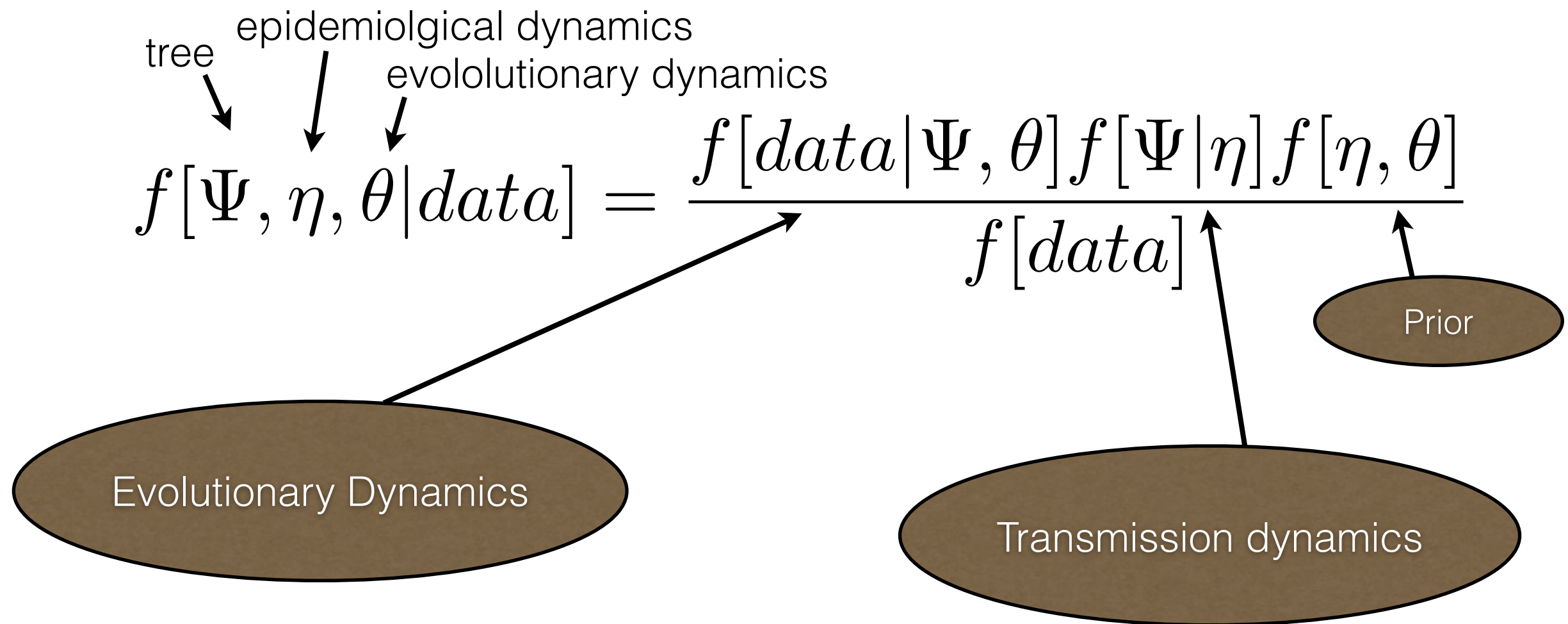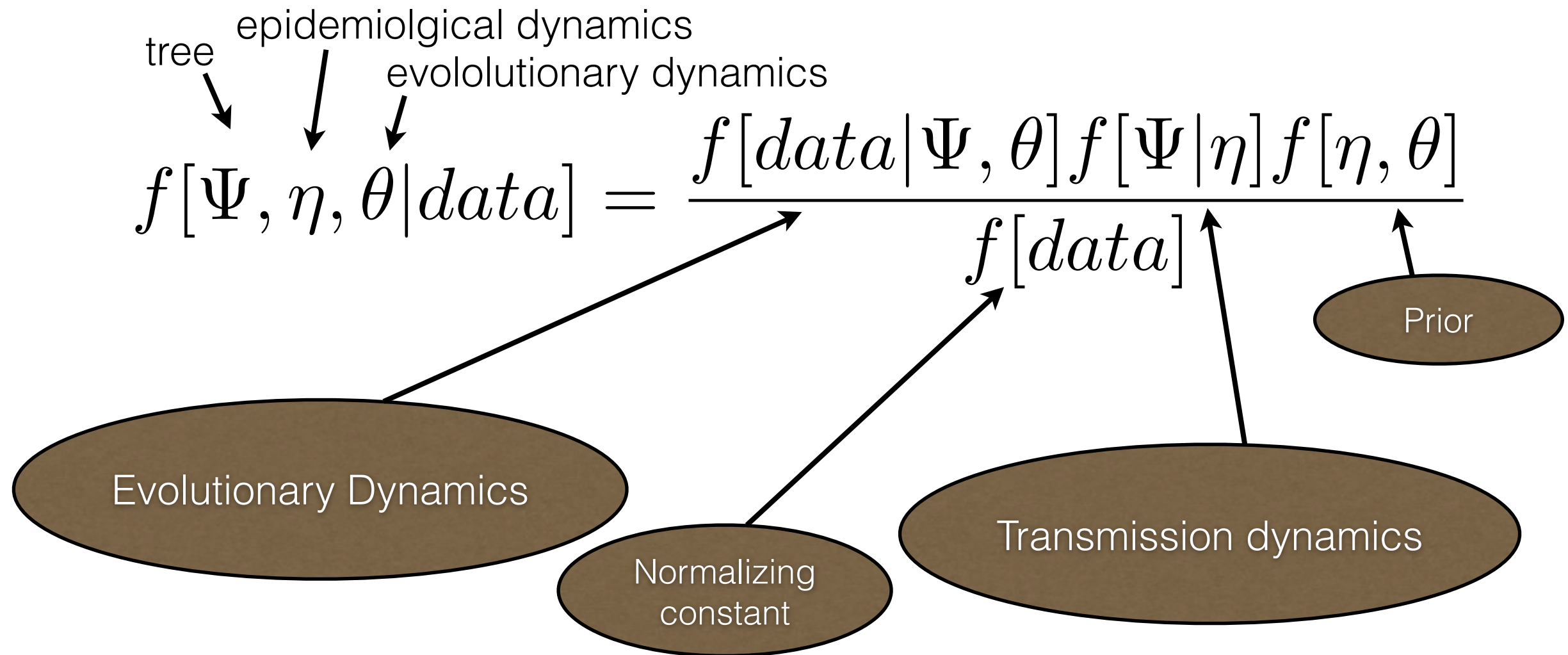$$f[\Psi, \eta, \theta | data] = \frac{f[data | \Psi, \theta] f[\Psi | \eta] f[\eta, \theta]}{f[data]}$$

Bayesian approach for estimating epidemiological parameters

tree

epidemiolgical dynamics
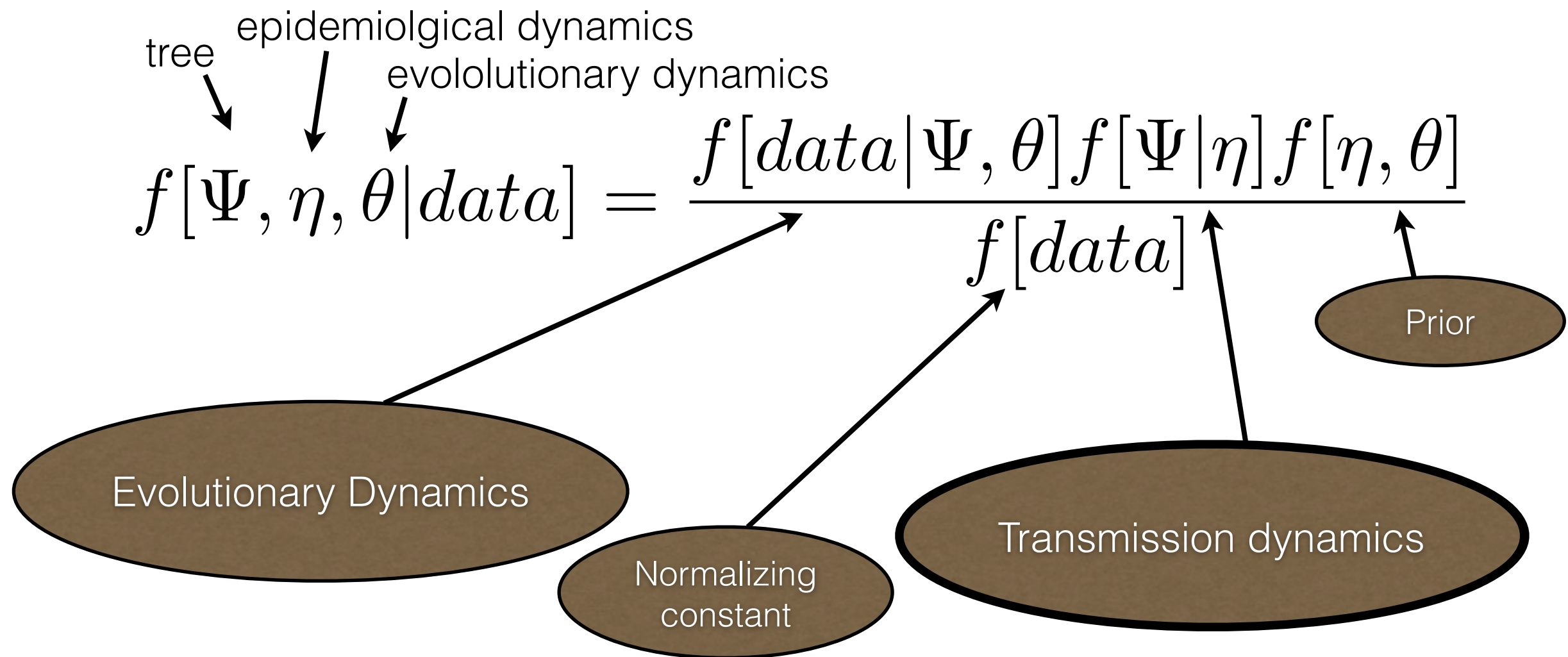
evololutionary dynamics

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

Evolutionary Dynamics

Bayesian approach for estimating epidemiological parameters

tree

epidemiolgical dynamics

evololutionary dynamics

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

Evolutionary Dynamics

Transmission dynamics

Bayesian approach for estimating epidemiological parameters

tree

epidemiolgical dynamics

evololutionary dynamics

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

Evolutionary Dynamics

Transmission dynamics

Prior

Bayesian approach for estimating epidemiological parameters

tree

epidemiolgical dynamics

evololutionary dynamics

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

Prior

Evolutionary Dynamics

Normalizing constant

Transmission dynamics

Bayesian approach for estimating epidemiological parameters

tree

epidemiolgical dynamics

evololutionary dynamics

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

Prior

Evolutionary Dynamics

Normalizing constant

Transmission dynamics

Bayesian approach for estimating epidemiological parameters
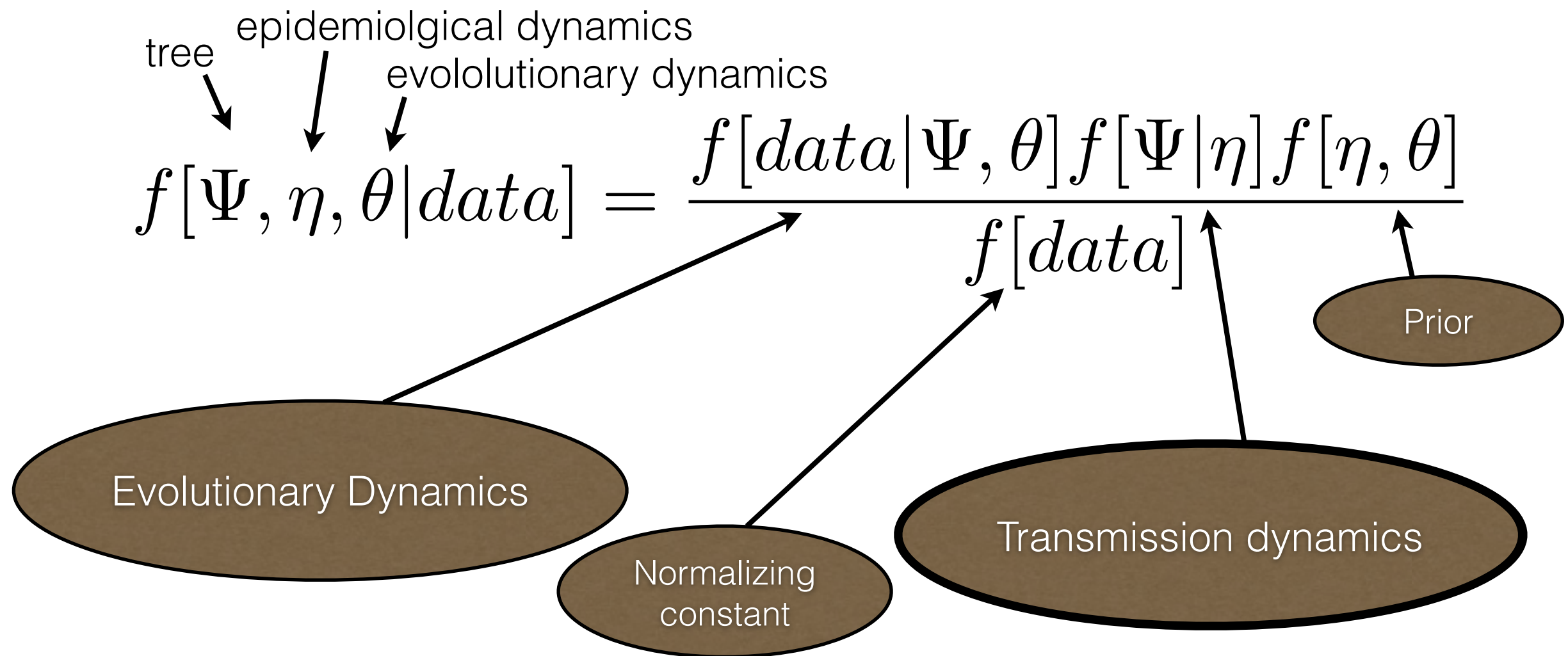
RevBayes samples the posterior distribution of trees and parameters

Part A
Coalescent as a model for transmission

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

Coalescent as a model for transmission

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

$$f(\quad)\,?$$

Part A
Coalescent as a model for transmission
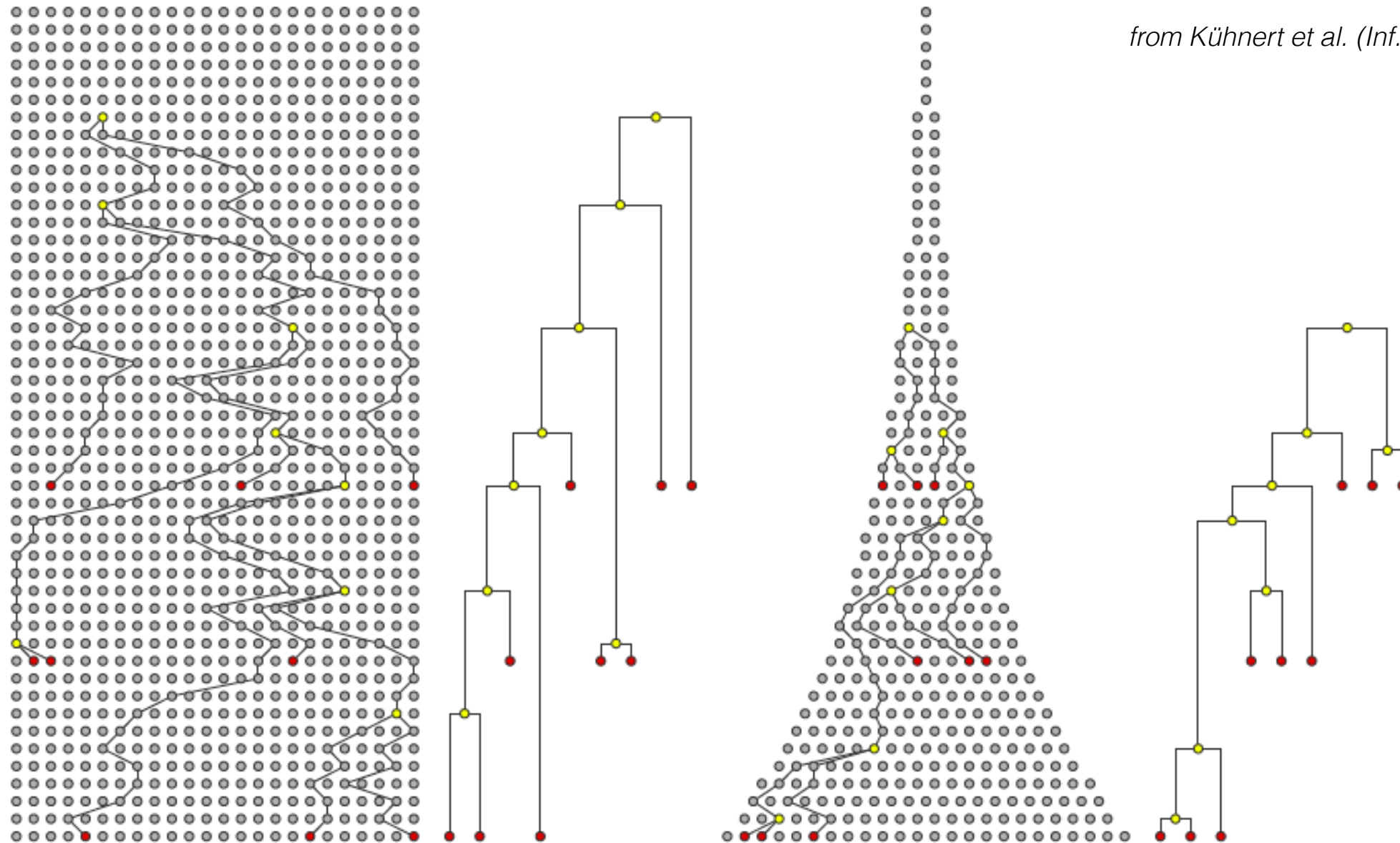
from Kühnert et al. (Inf. Gen. Evol., 2011)
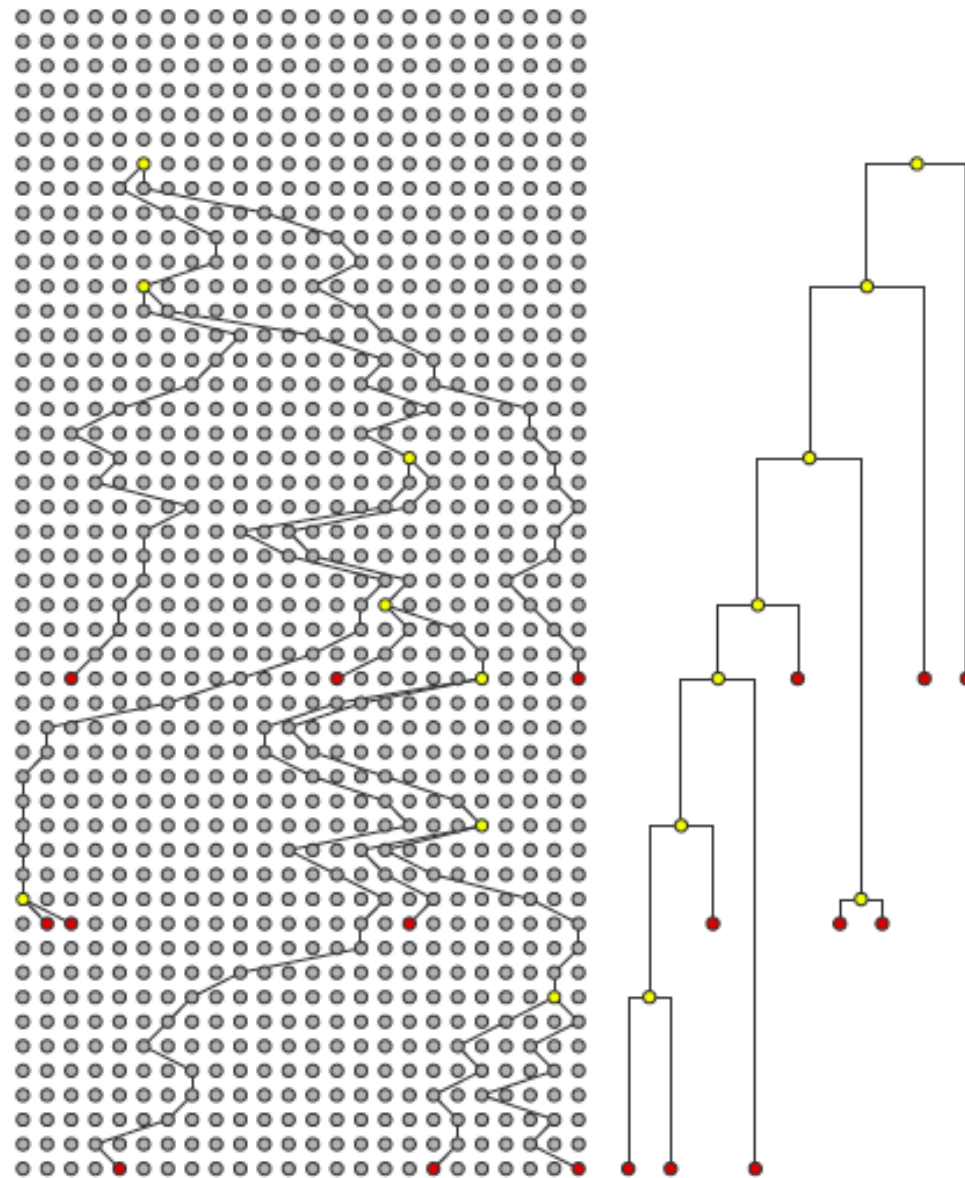
$$f_G(g|\theta) = \frac{1}{\theta^{n-1}} \cdot \prod_{i=2}^{2n-1} e^{(-k_i(k_i-1)/2\theta)(t_i-t_{i-1})}$$
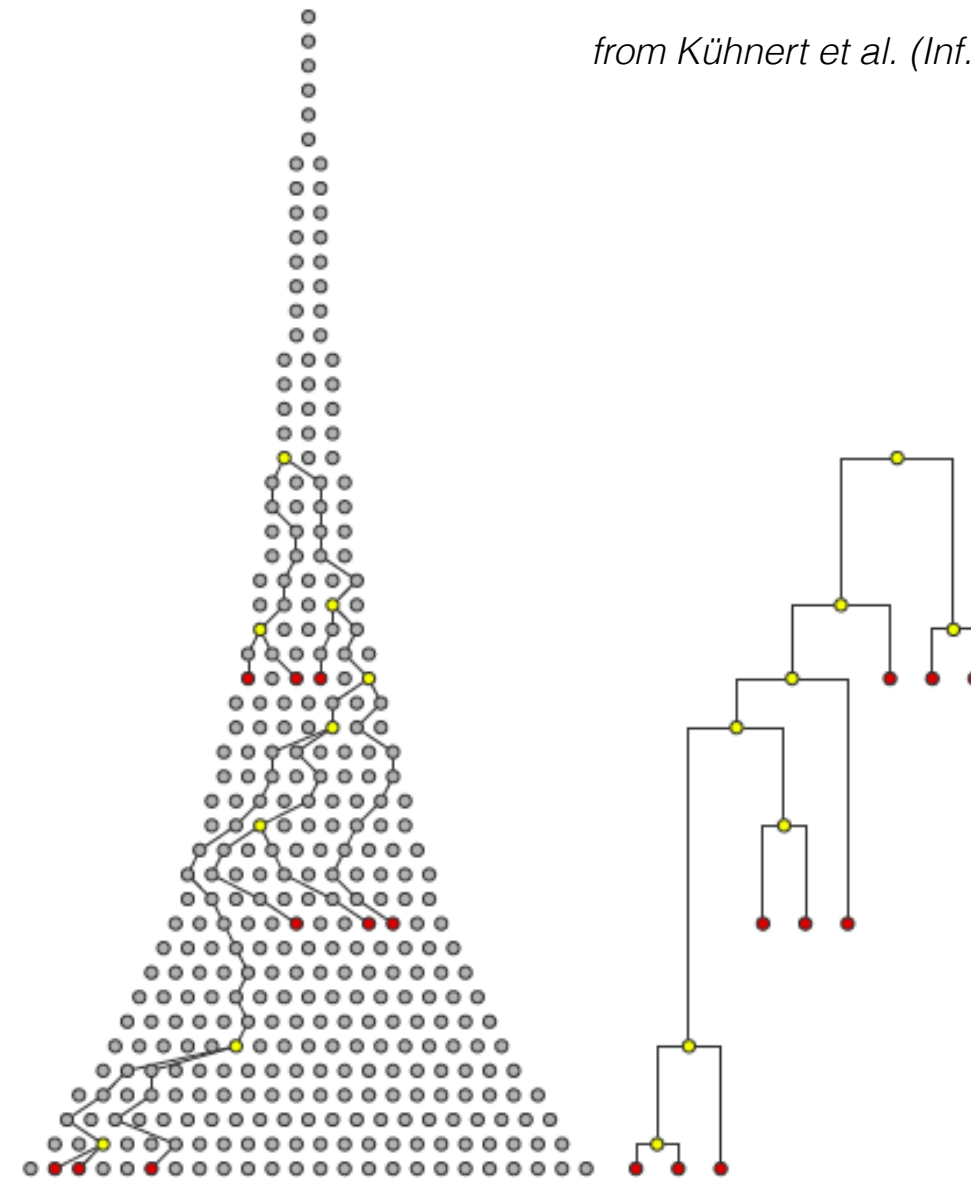
$$f_G(g|\theta, r) = \frac{1}{\theta^{n-1}} \cdot \prod_{i=2}^{2n-1} e^{rt_i} e^{(-k_i(k_i-1)/2\theta r)(e^{rt_i}-e^{rt_{i-1}})}$$

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. η=r)

$R_0 = \lambda/\delta = rD+1$ (where D is expected time of infectiousness)
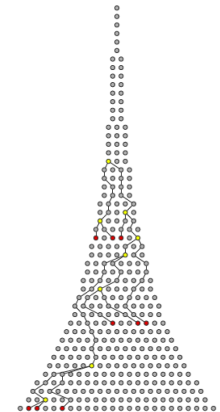
Estimating R_0 from sequencing data

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. η=r)

$R_0 = \lambda/\delta = rD+1$ (where D is expected time of infectiousness)
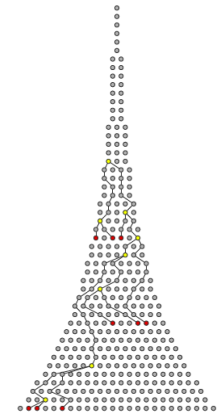
Results
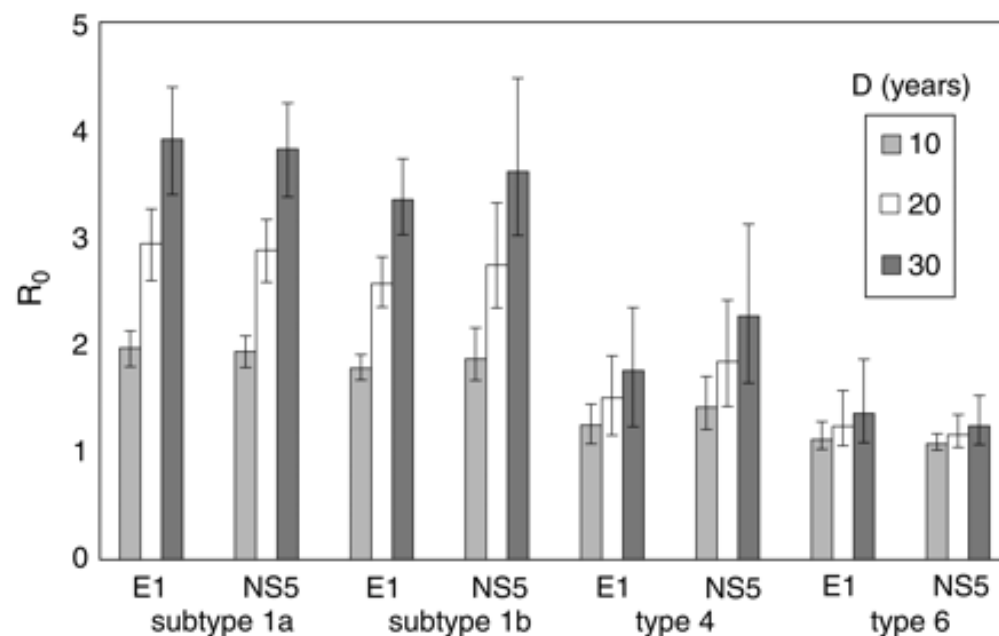
Bayesian skyline plot relaxes constant growth rate
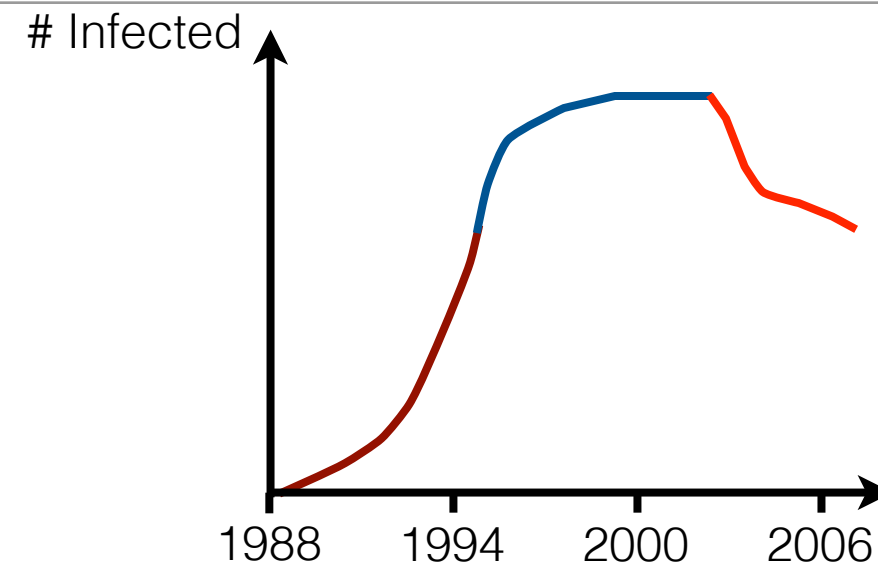
Infected

1988 1994 2000 2006

10

Bayesian skyline plot relaxes constant growth rate

| Assump-tion | Piecewise constant population size |

Infected

1988   1994   2000   2006
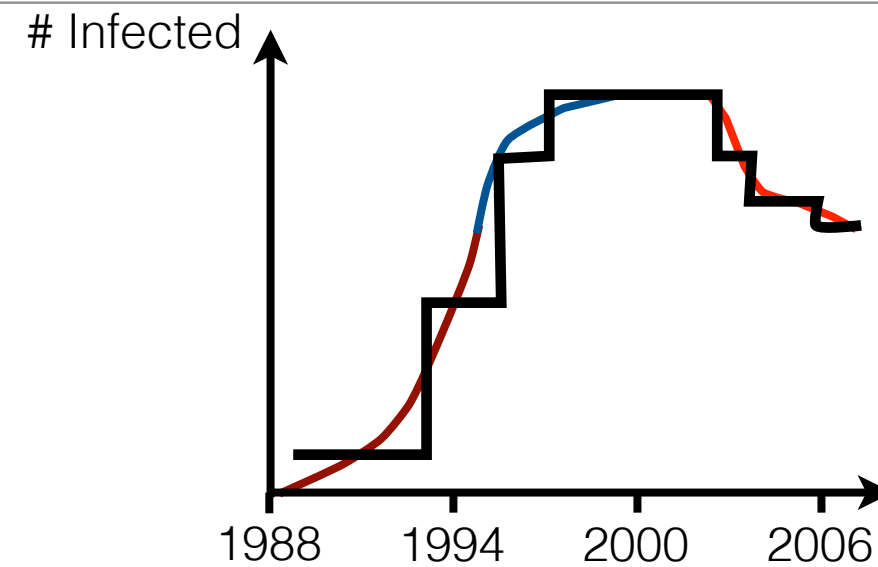
Bayesian skyline plot relaxes constant growth rate

Assumption

Piecewise constant population size

Bayesian skyline plot relaxes constant growth rate
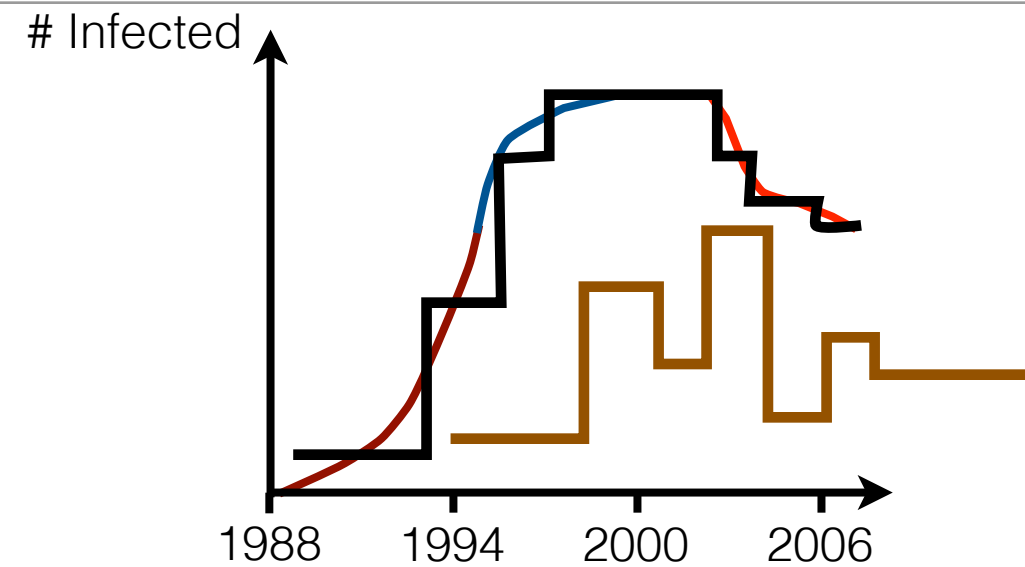
Assumption

Piecewise constant population size
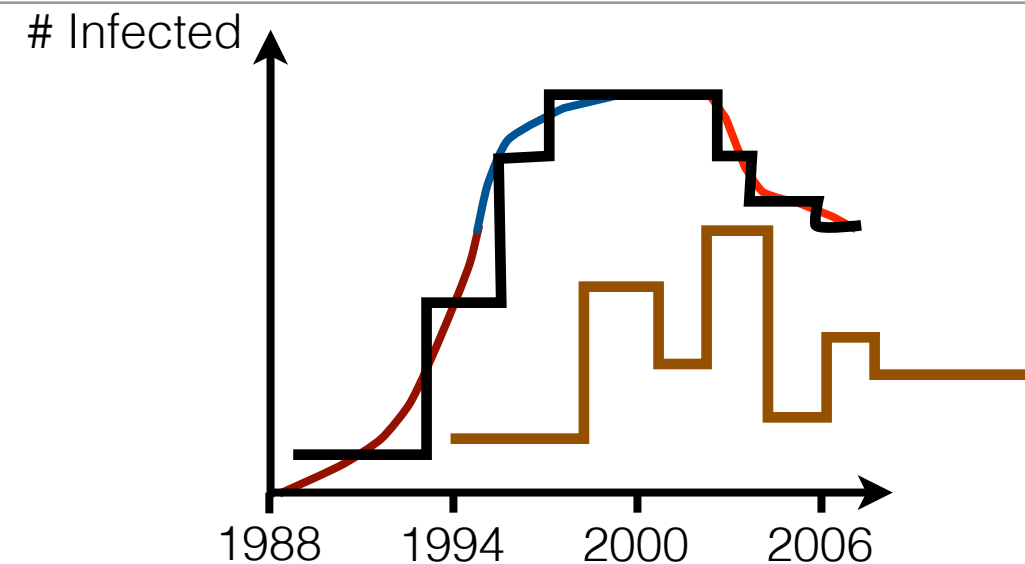
HCV in Egypt

‣ 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

Assumption

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?

from Drummond et al. (MBE, 2005)
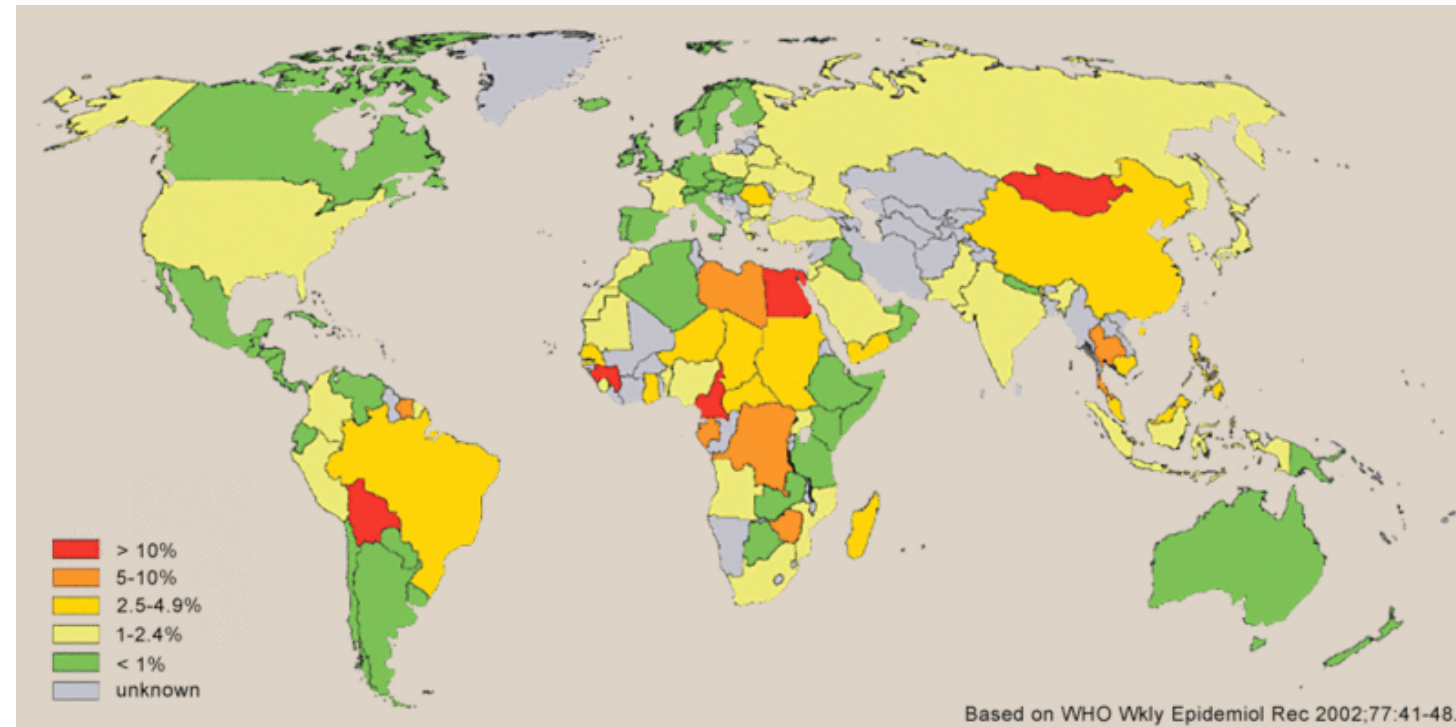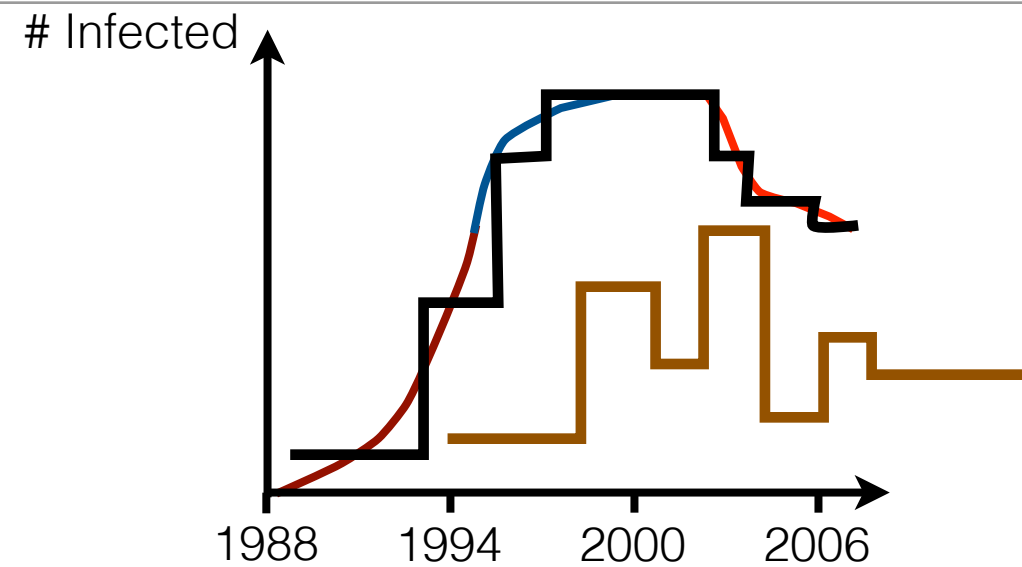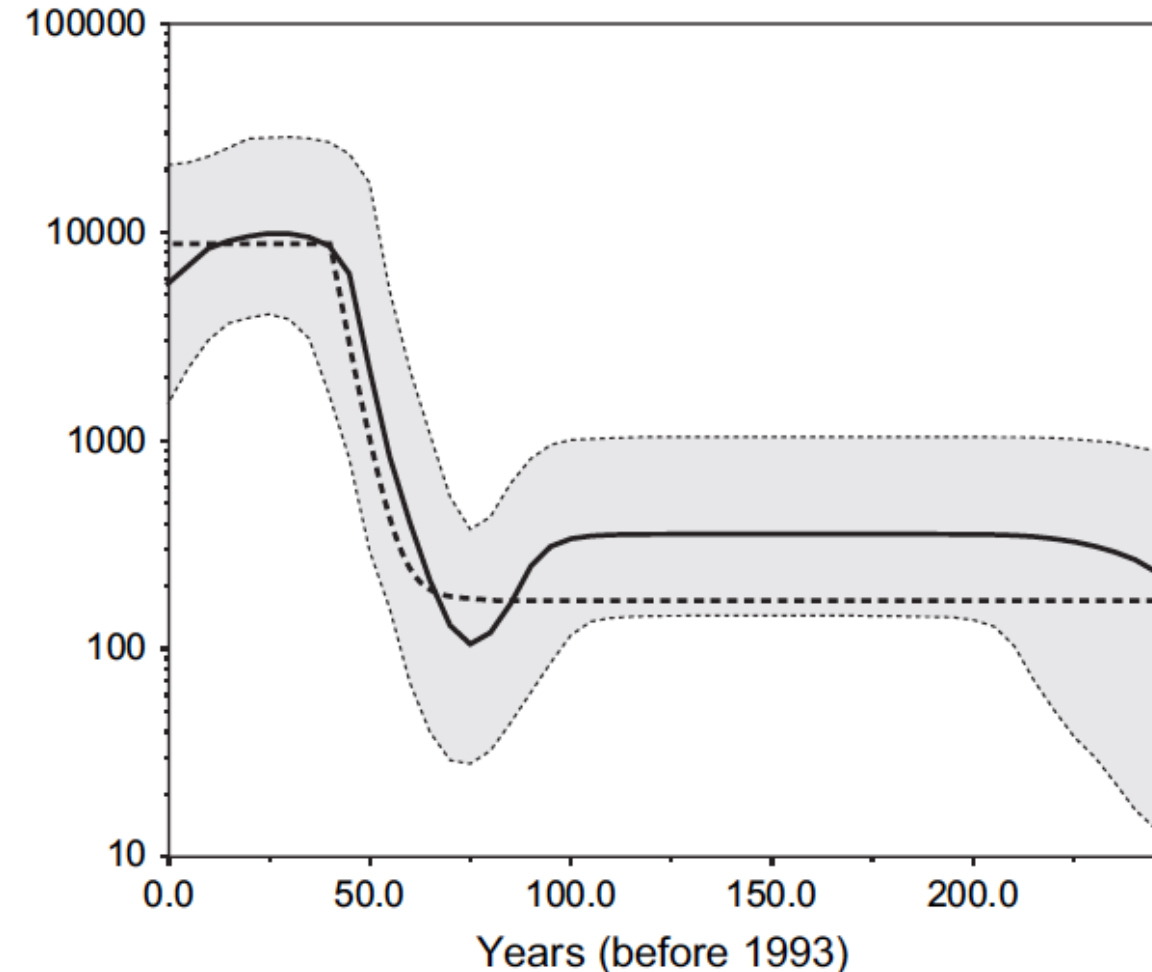
Comparing coalescent times under the birth-death model and the coalescent
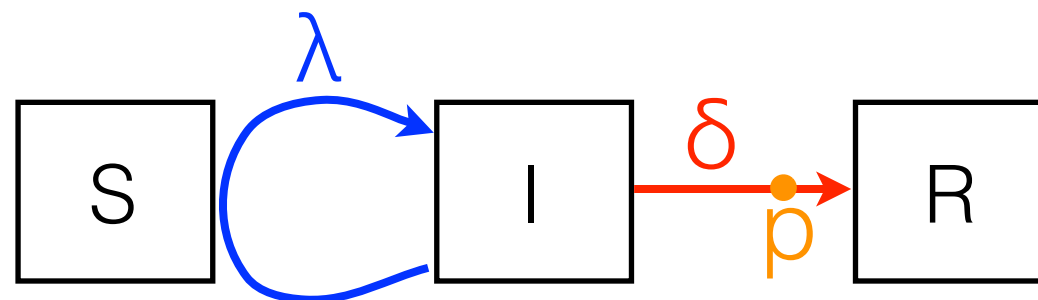
N = 10000

— birth-death

— coalescent

$R_0 = 1.05$

cumulative probability distribution

time

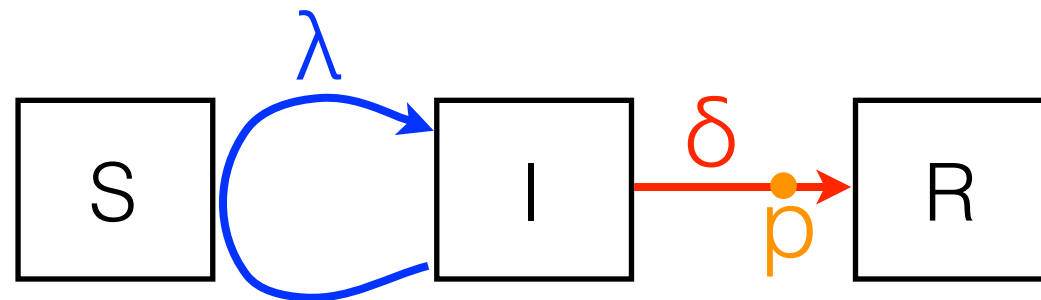Coalescent overestimates growth rate for small (realistic) R_0

Part 2
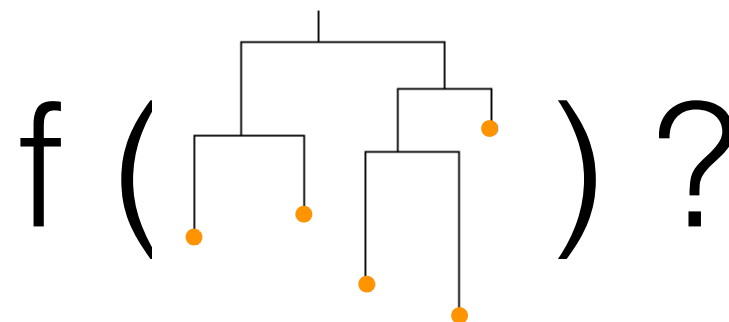# Birth-death model as model for transmission
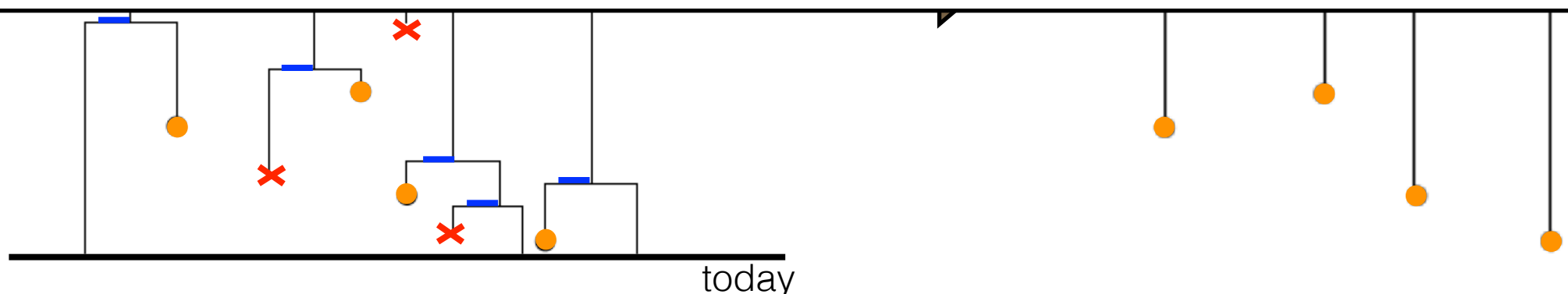

Birth-death
model

- transmission rate λ
- "becoming-non-infectious" rate δ
- sampling probability p

Parameters may depend on:
   1) constant (epidemic outbreak)
   2) time (environmental effect)
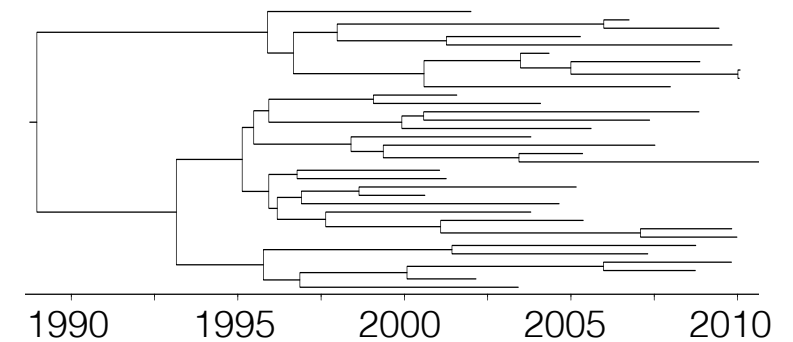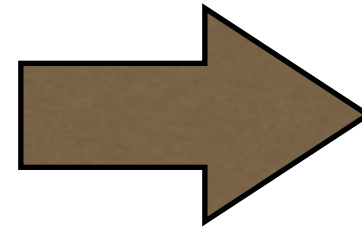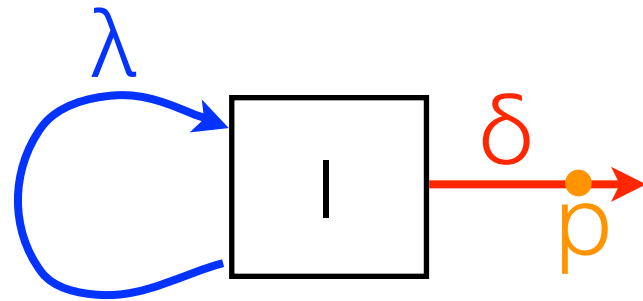   3) # of susceptibles (SIR model)
   4) type of infected individual

Part 2
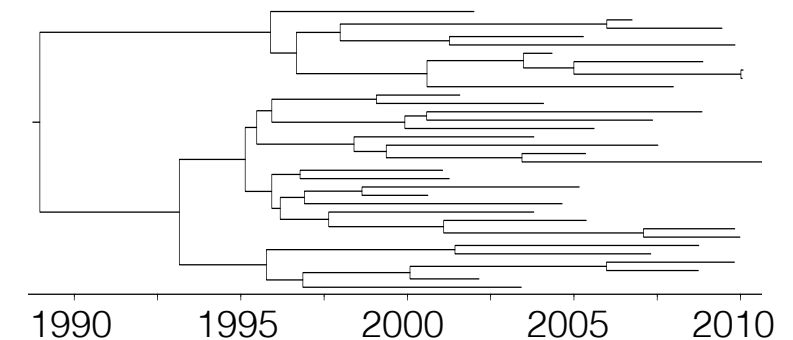# Birth-death model as model for transmission

Birth-death model

Phylo-genetic trees

λ

S I δ R

p

• 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

t_0

today

Stadler (JTB, 2010)

Part 2
Birth-death model as model for transmission

Birth-death
model

genetic
trees

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 ( ) ?

today

Stadler (JTB, 2010)

Simulation study:
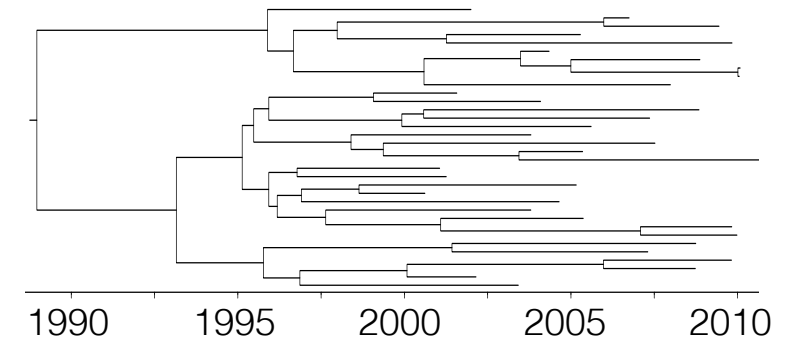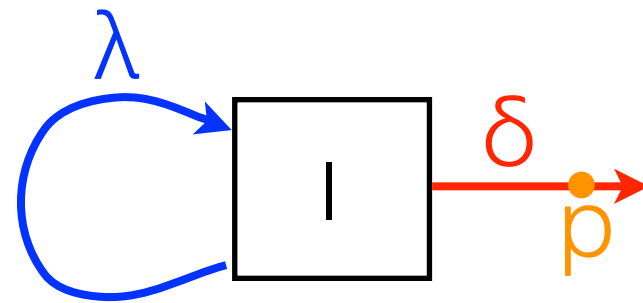# Comparing exact method to coalescent

Simulation study:
Comparing exact method to coalescent

λ-δ is population size growth rate

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

Stadler et al. (Mol. Biol. Evol., 2012)

Simulation study:
Comparing exact method to coalescent

Simulating epidemic outbreak

λ-δ is population size growth rate

Exact method

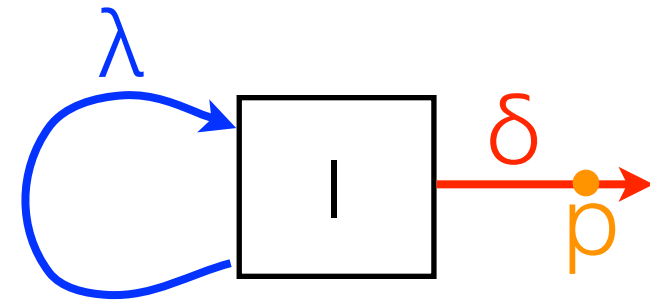▸   95% *HPD* interval contains true value in **97%** of the simulated trees

Coales-cent approxi-mation

Parameterizes deterministic population size growth

▸   *95% HPD* interval contains true value in **55%** of the simulated trees
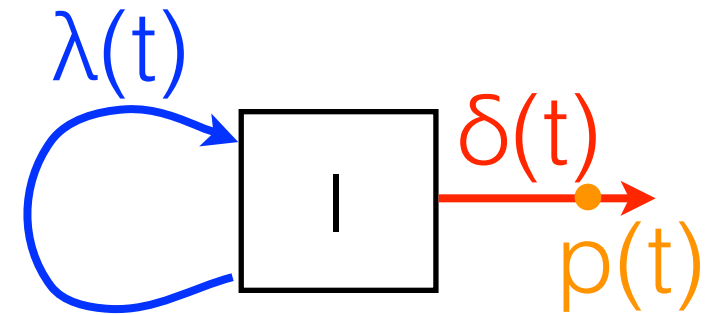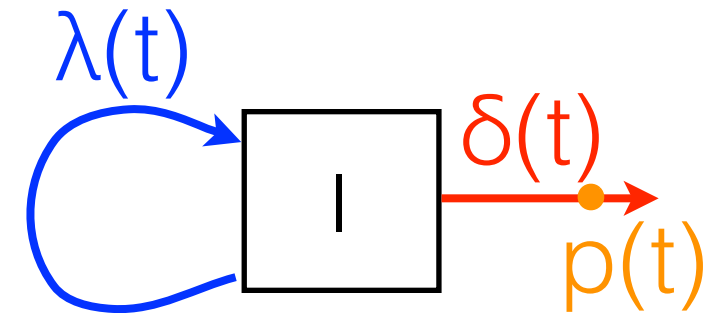
Stadler et al. (Mol. Biol. Evol., 2012)

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

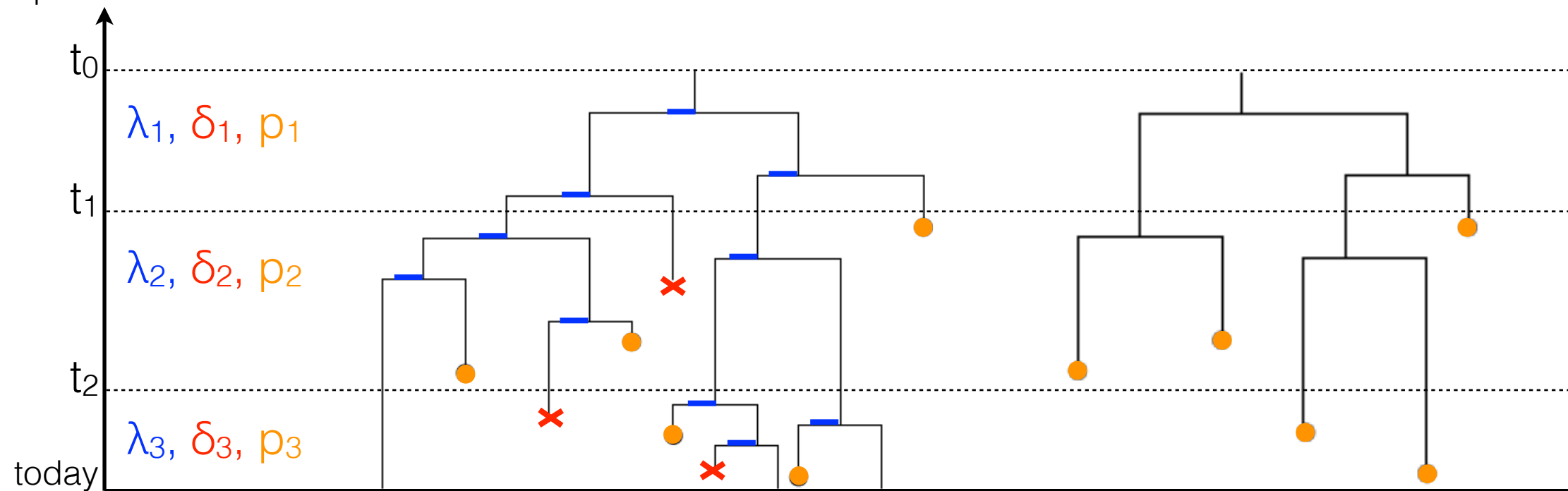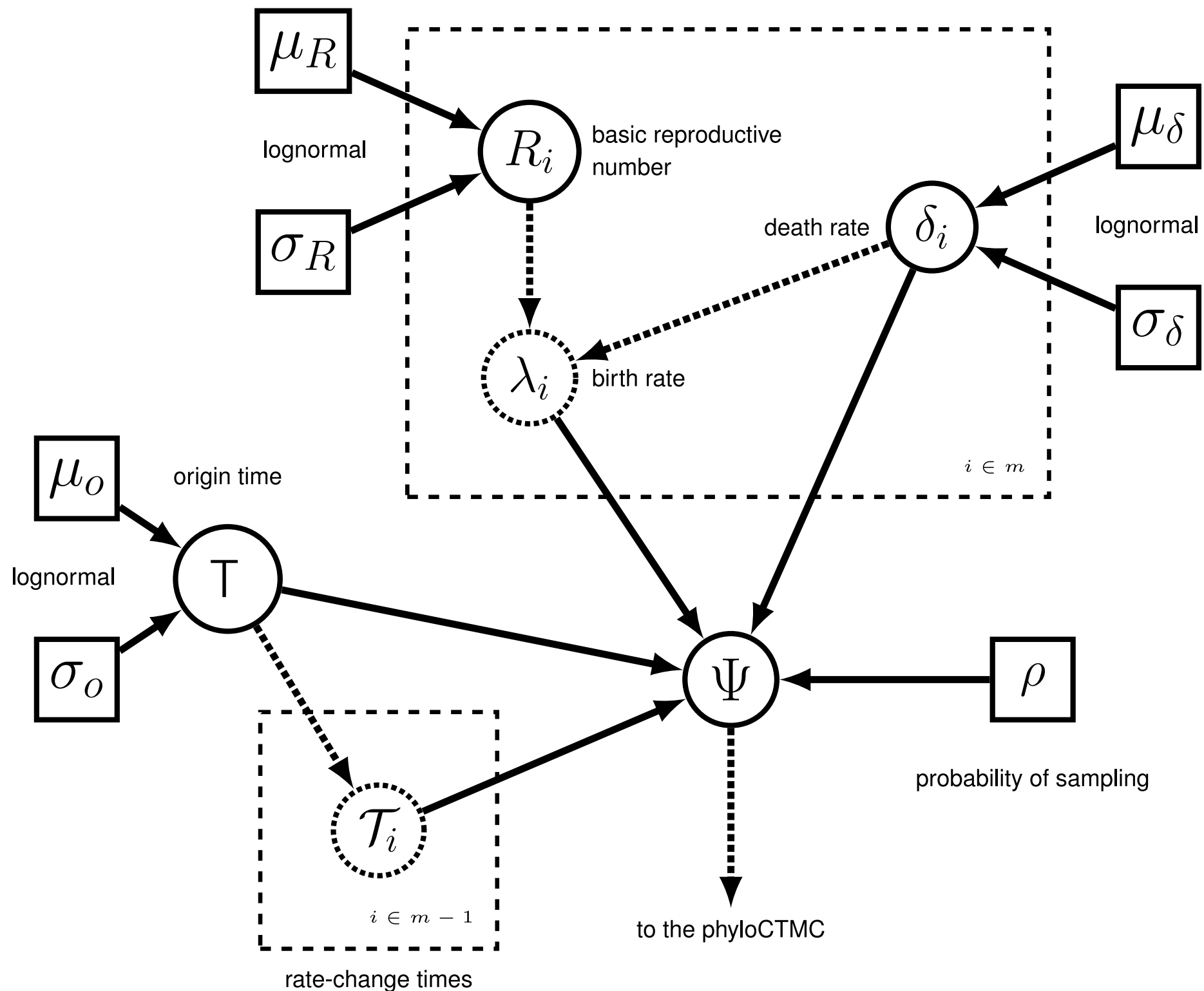Epidemiological rates may change through time:

$$\lambda(t)$$
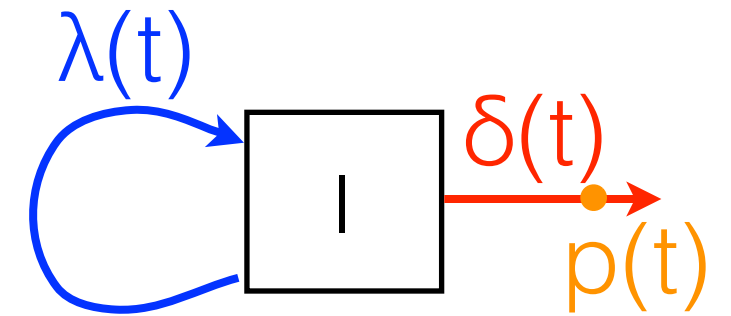
$$\delta(t)$$

$$\rho(t)$$

I

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

Epidemiological rates may change through time:

$\lambda(t)$

$\delta(t)$

$p(t)$

Time in past

t_0

λ_1, δ_1, p_1

t_1

λ_2, δ_2, p_2

t_2

λ_3, δ_3, p_3

today

Graphical model

Hepatitis C virus in Egypt

$$\lambda(t)$$

$$\delta(t)$$

I

$$p(t)$$

Effective reproductive number $R_e(t)$

 generalizes

Basic reproductive number $R_0 = R_e(0)$

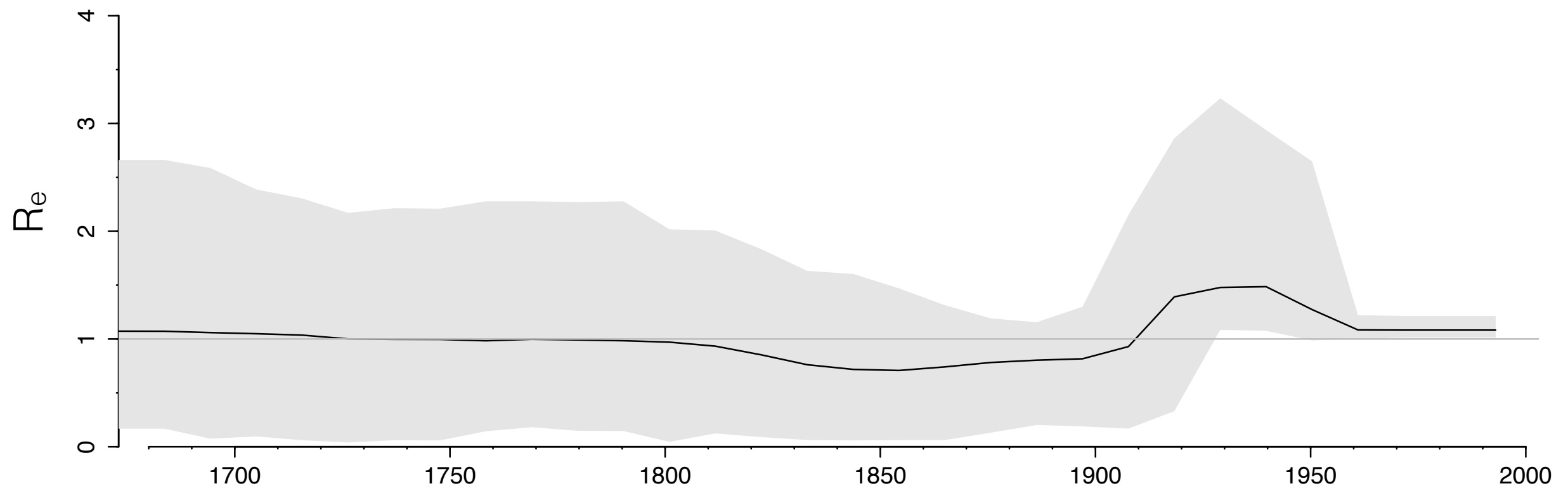*Stadler, Kühnert et al. (PNAS, 2013)*

Hepatitis C virus in Egypt

Effective reproductive number $R_e(t)$
generalizes

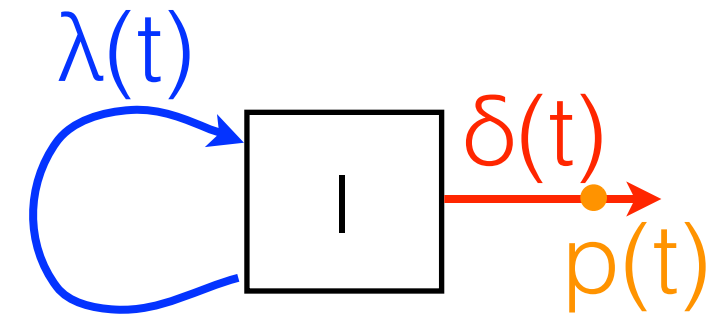Basic reproductive number $R_0 = R_e(0)$

Analysis of 63 seq:

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

——— median

▨ 95% HPD interval

Stadler, Kühnert et al. (PNAS, 2013)

Hepatitis C virus in Egypt

$\lambda(t)$

$\delta(t)$

I

$p(t)$
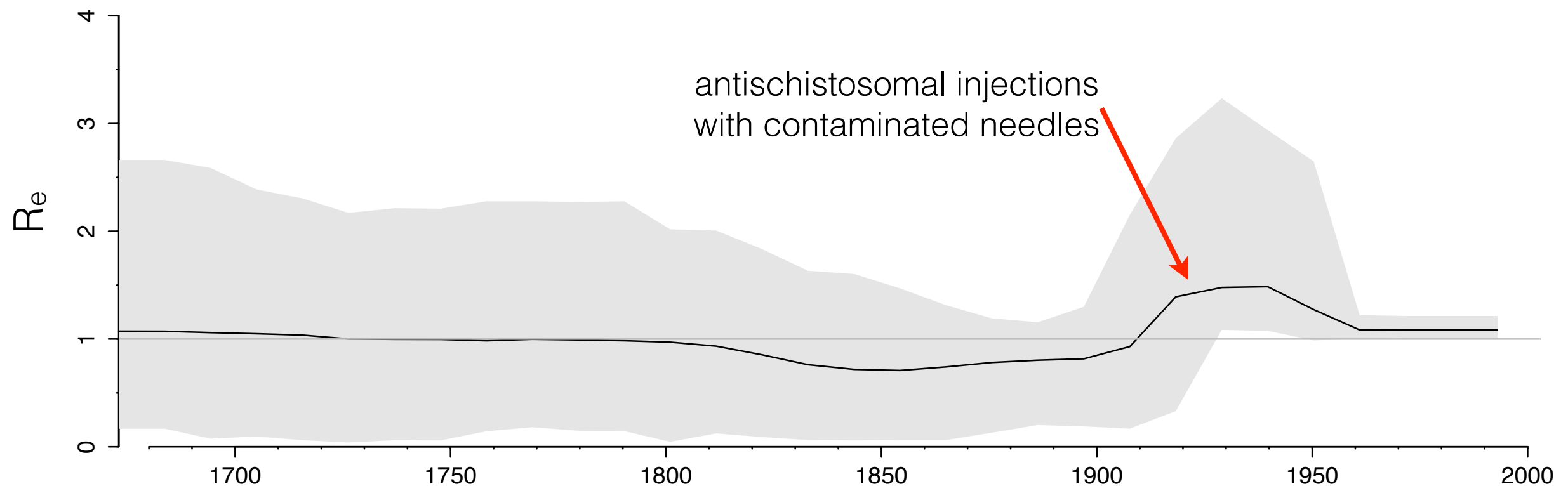
Effective reproductive number $R_e(t)$
generalizes
Basic reproductive number $R_0 = R_e(0)$
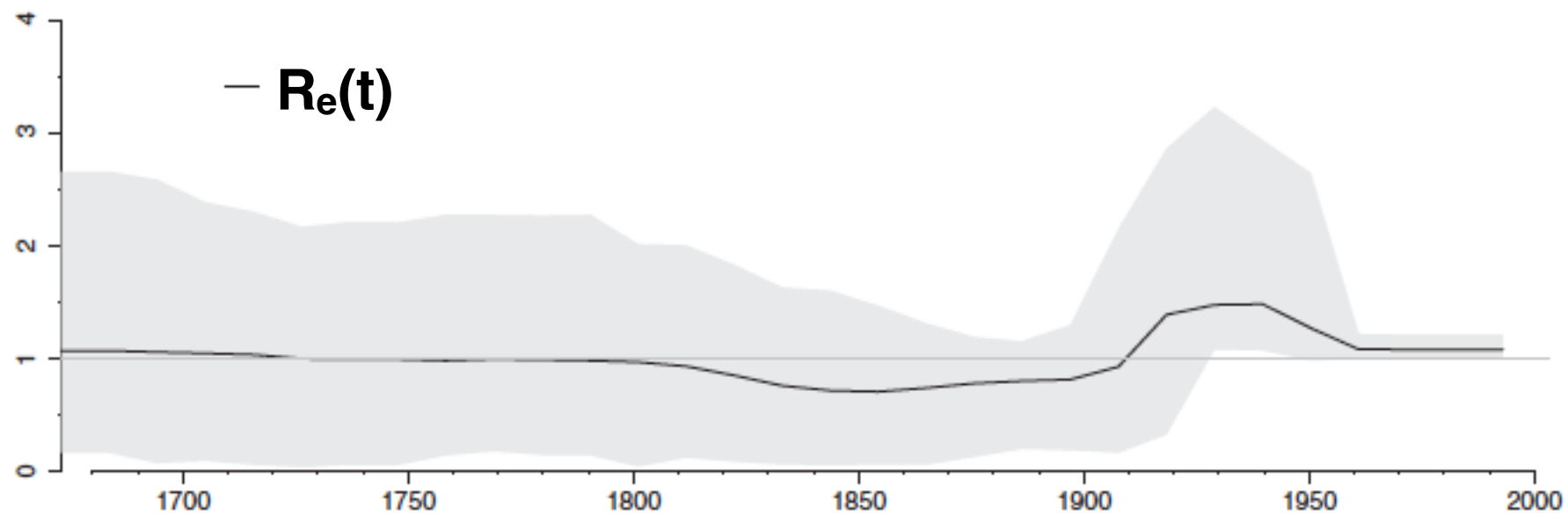
Analysis of 63 seq:

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

—— median

▨ 95% HPD interval

antischistosomal injections
with contaminated needles

R_e

1700   1750   1800   1850   1900   1950   2000
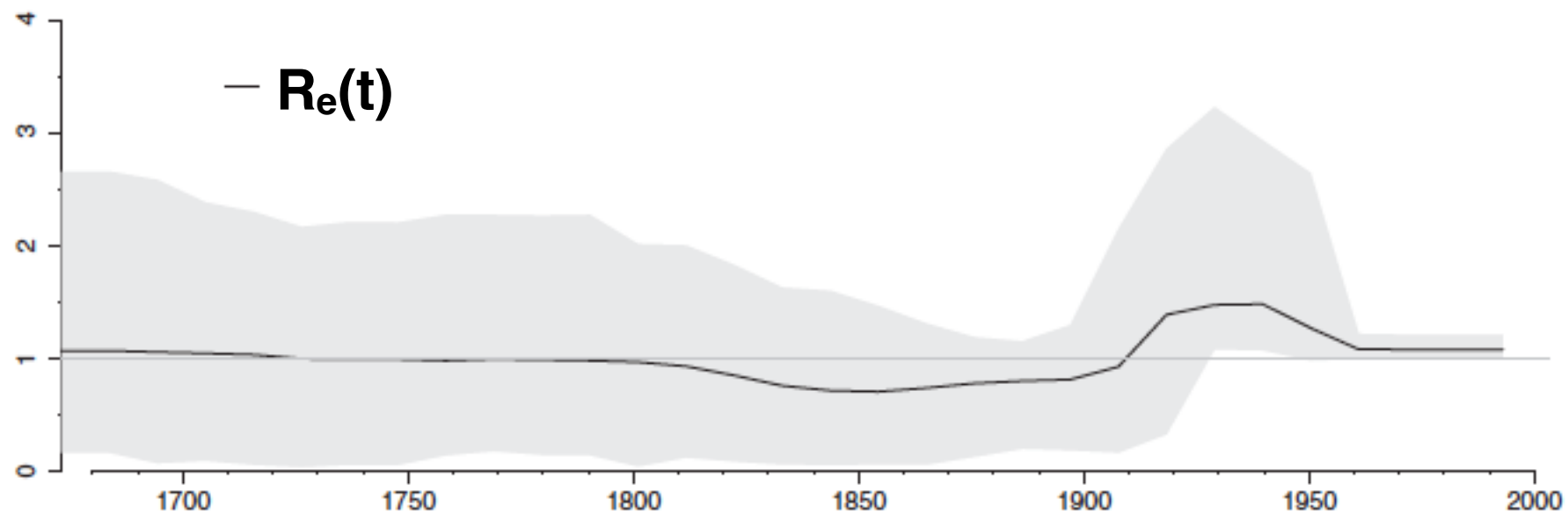
Stadler, Kühnert et al. (PNAS, 2013)

The two skyline models

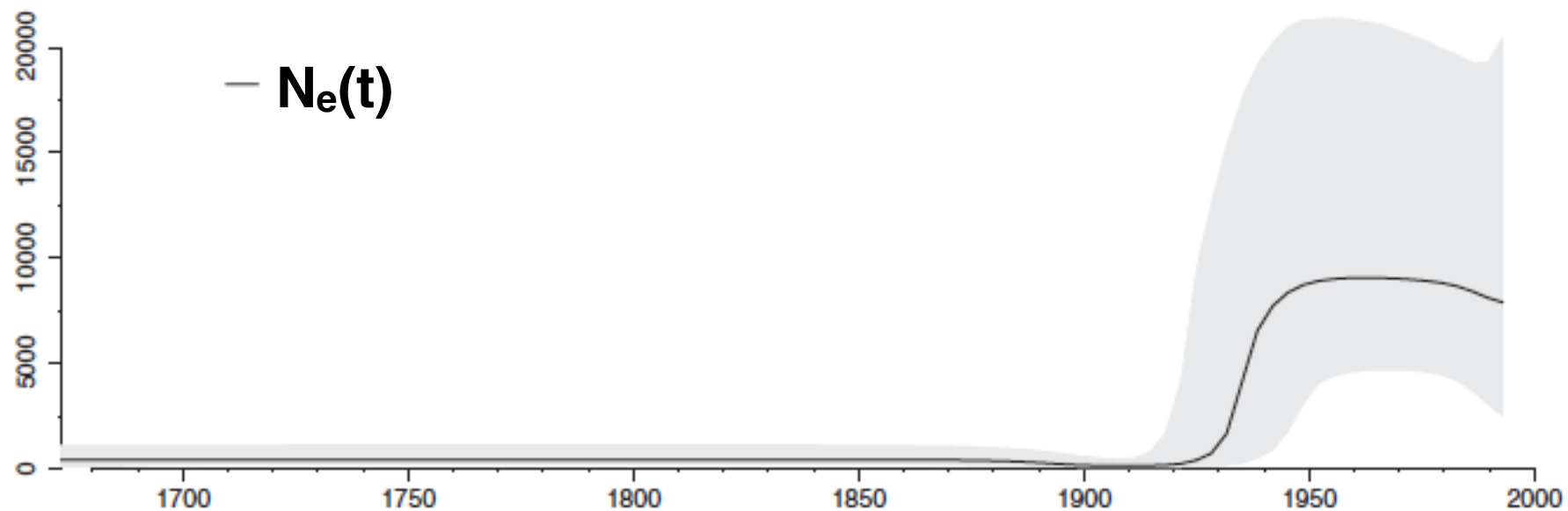Birth-death skyline plot: effective reproductive number

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