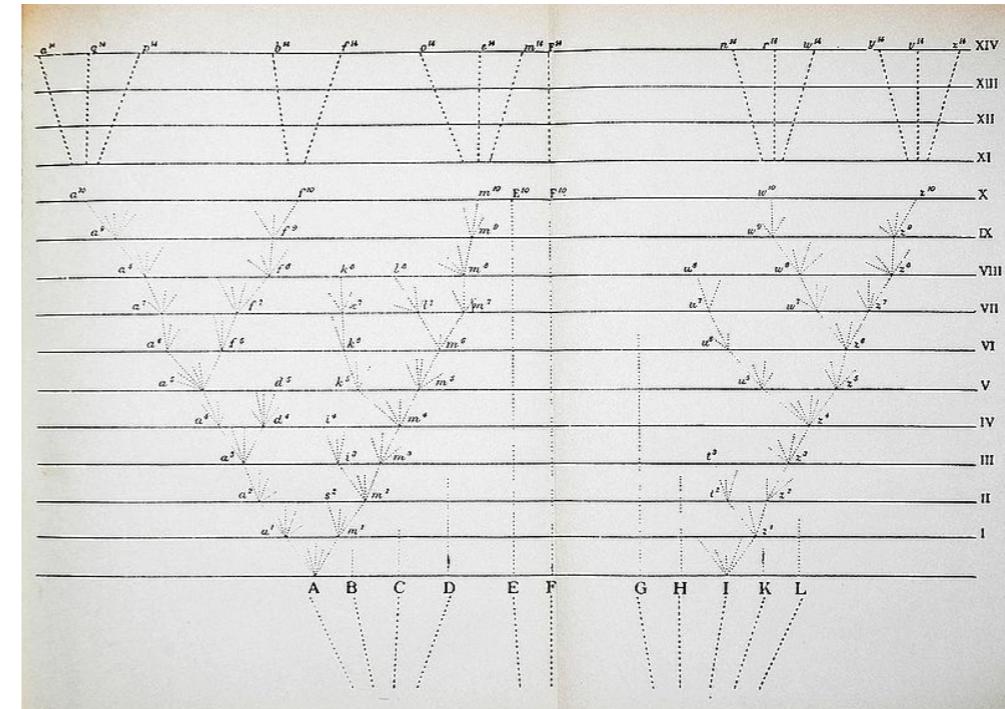
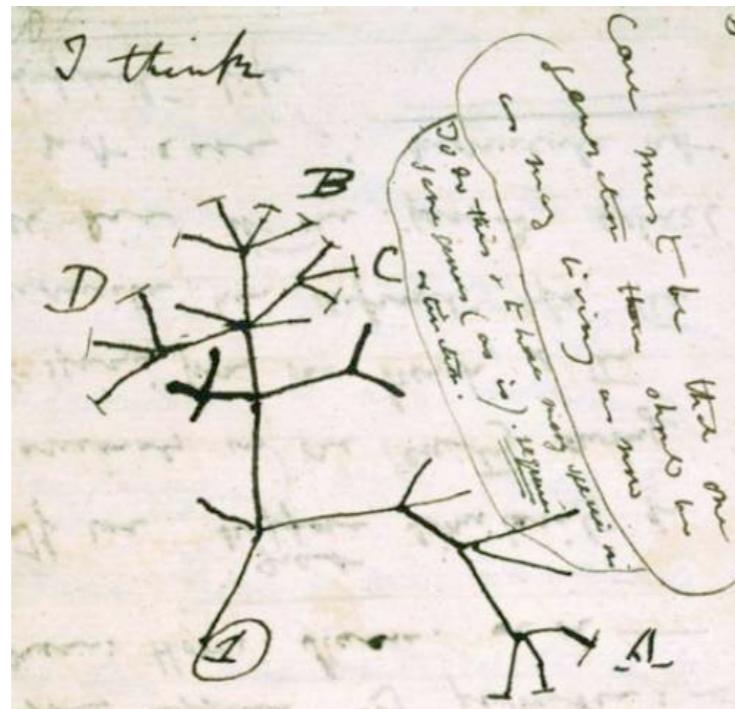


Phylogenetics

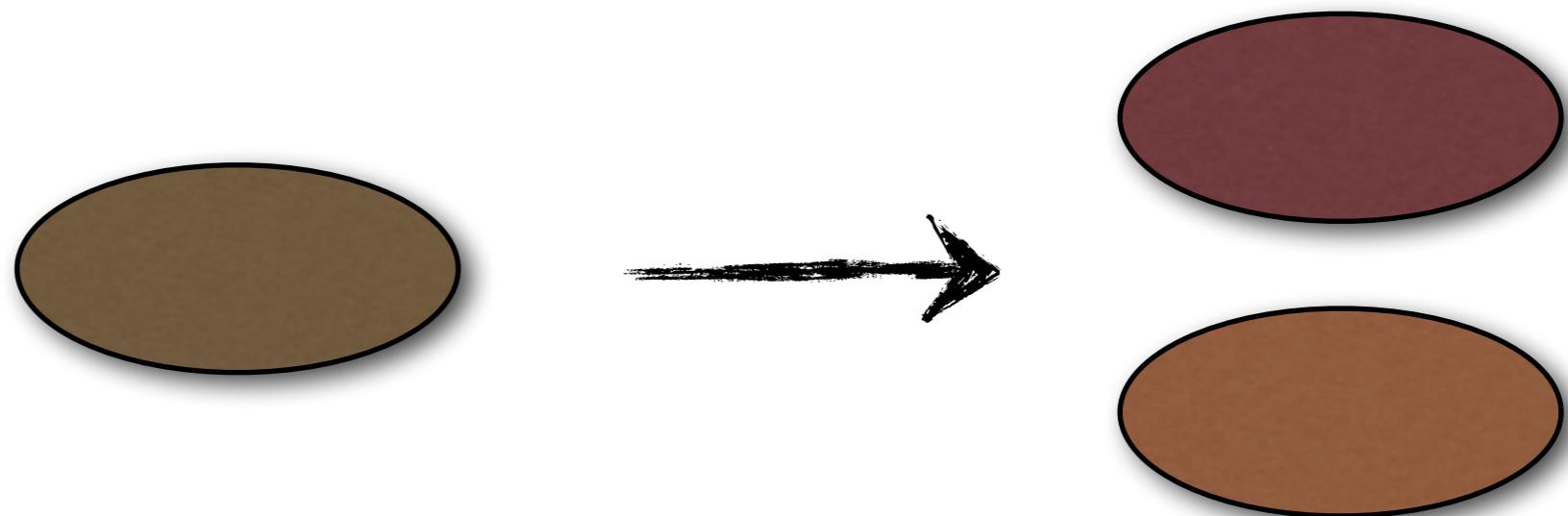
Tree priors based on
the birth-death model and the coalescent



Basis of biology: Reproduction & genetic change

Scale:

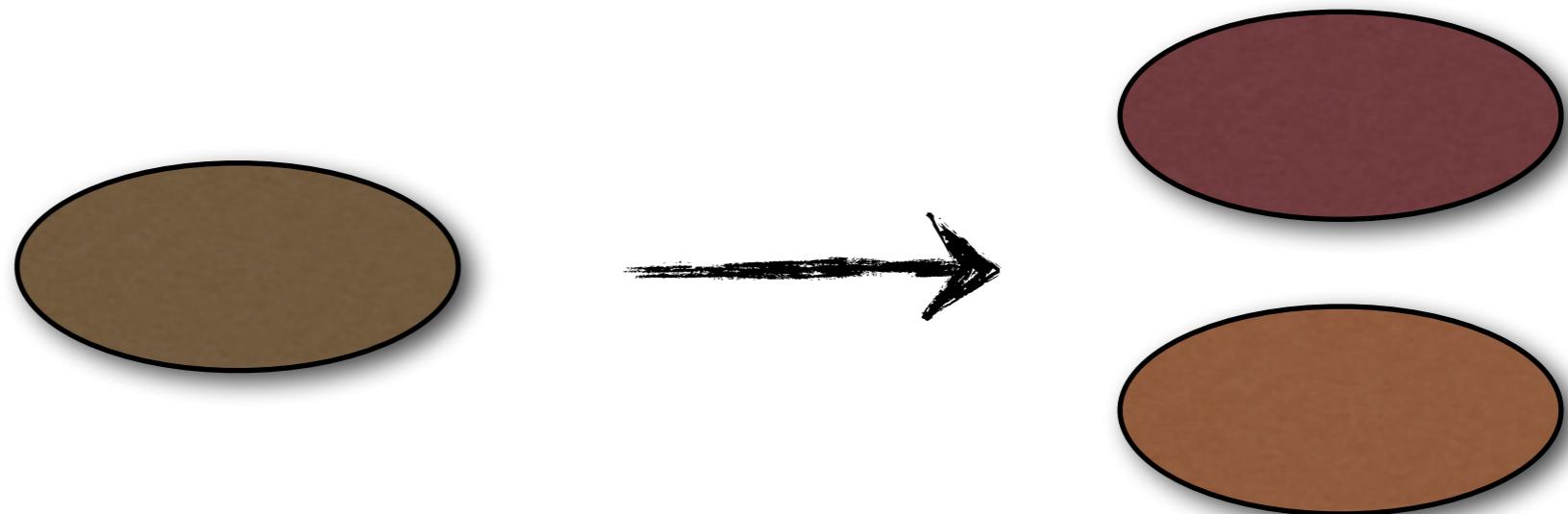
DNA
Virion
Prokaryote
Eukaryotic cell
Eukaryote
Species
Infected host
Immune cells



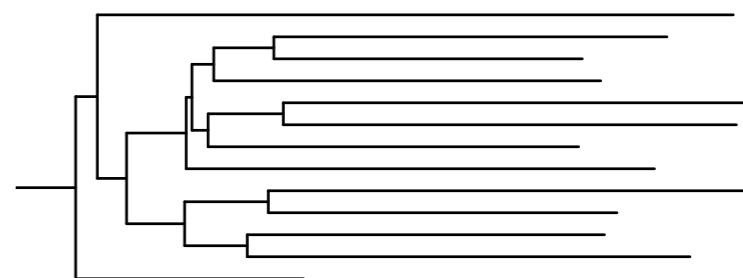
Basis of biology: Reproduction & genetic change

Scale:

DNA
Virion
Prokaryote
Eukaryotic cell
Eukaryote
Species
Infected host
Immune cells



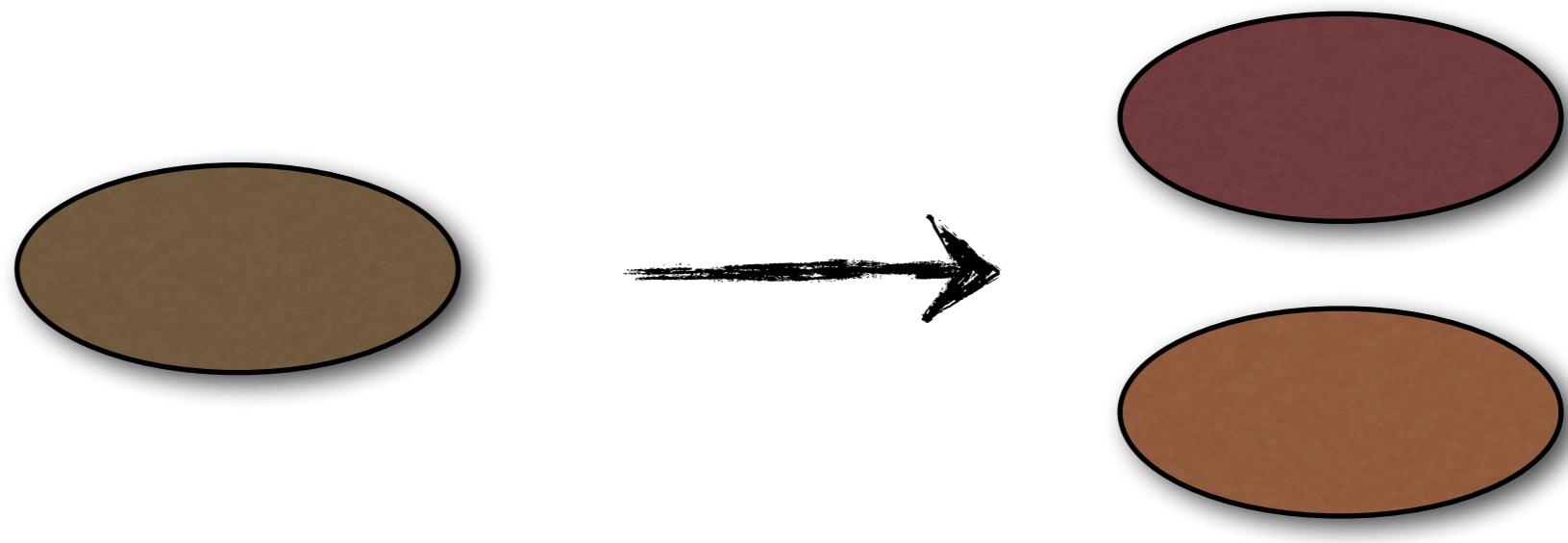
- ▶ Lineage tree (=phylogeny) represents **reproduction** process on any of these scales
- ▶ **Genetic** information changes along the tree



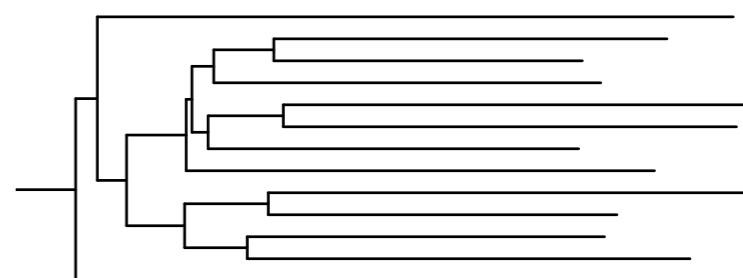
Basis of biology: Reproduction & genetic change

Scale:

DNA
Virion
Prokaryote
Eukaryotic cell
Eukaryote
Species
Infected host
Immune cells



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



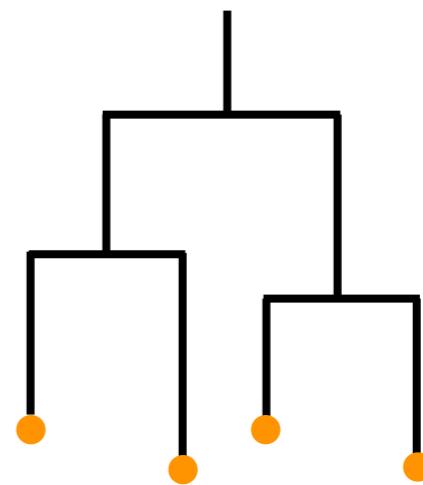
Phylogenetic model

Population
model:
phylodyn.
likelihood

Growth of tree

(Population dynamics - replication)

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



Phylogenetic model

Population model:
phyldyn.
likelihood

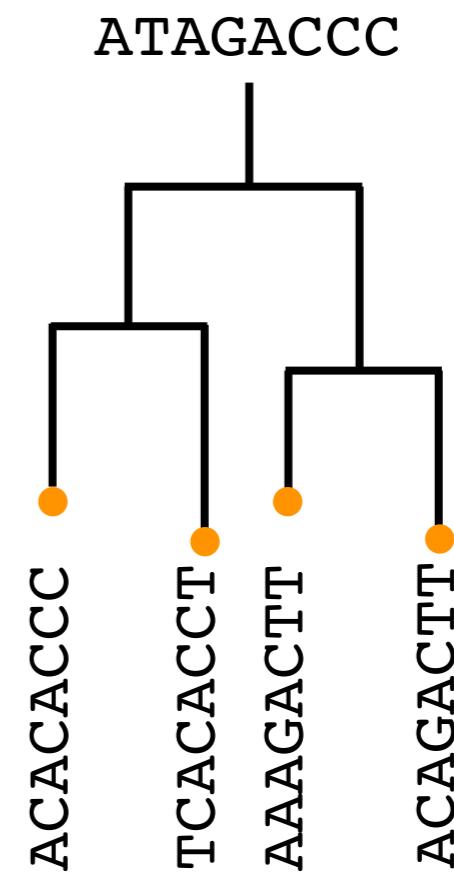
Growth of tree
(Population dynamics - replication)

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

Evolution model:
phylogen.
likelihood

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

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



Phylogenetic model

Population model:
phyldyn.
likelihood

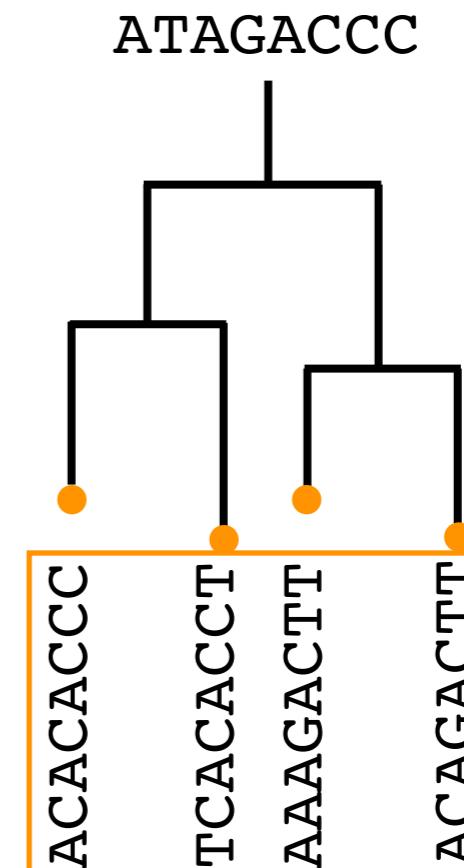
Growth of tree
(Population dynamics - replication)
► parameters η ($\rightarrow p[\text{tree} | \eta]$)

Evolution model:
phylogen.
likelihood

Evolution of sequences along tree
(Evolutionary dynamics - genetic change)
► parameters θ ($\rightarrow p[\text{seq} | \theta, \text{tree}]$)

Data

Sequence alignment



Phylogenetic model

Population model:
phylogen. likelihood

Growth of tree
(Population dynamics - replication)

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

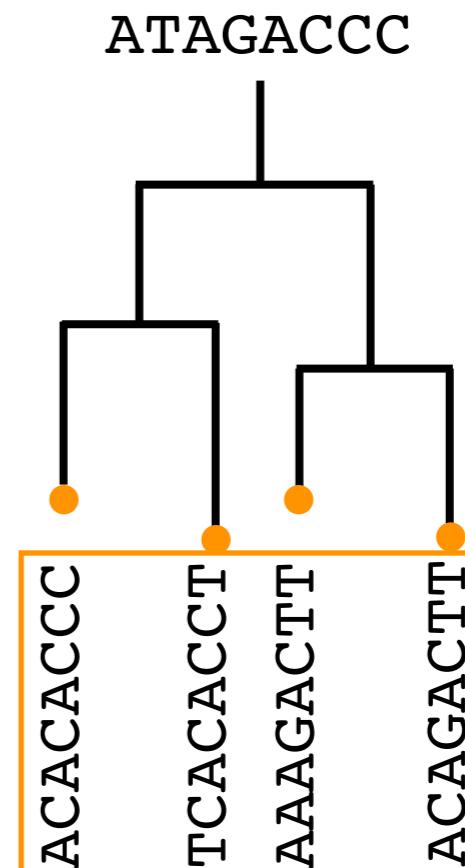
Evolution model:
phylogen. likelihood

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

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

Data

Sequence alignment



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

=

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

Phylogenetic model

Population model:
phylogen. likelihood

Growth of tree
(Population dynamics - replication)

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

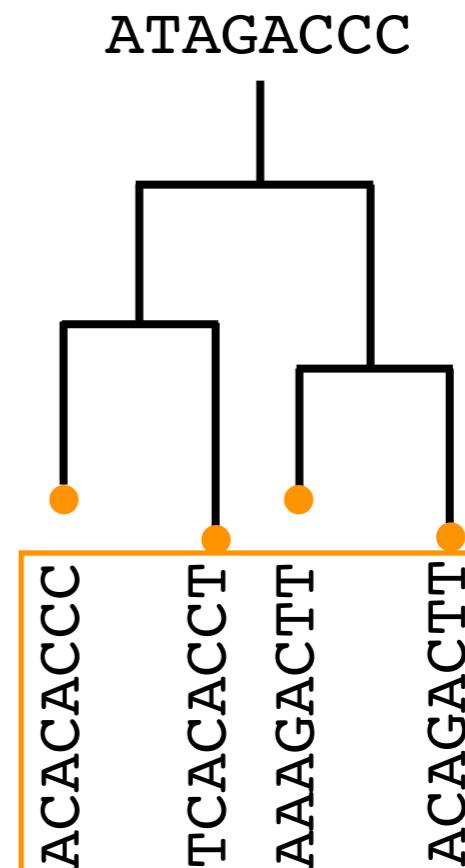
Evolution model:
phylogen. likelihood

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

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

Data

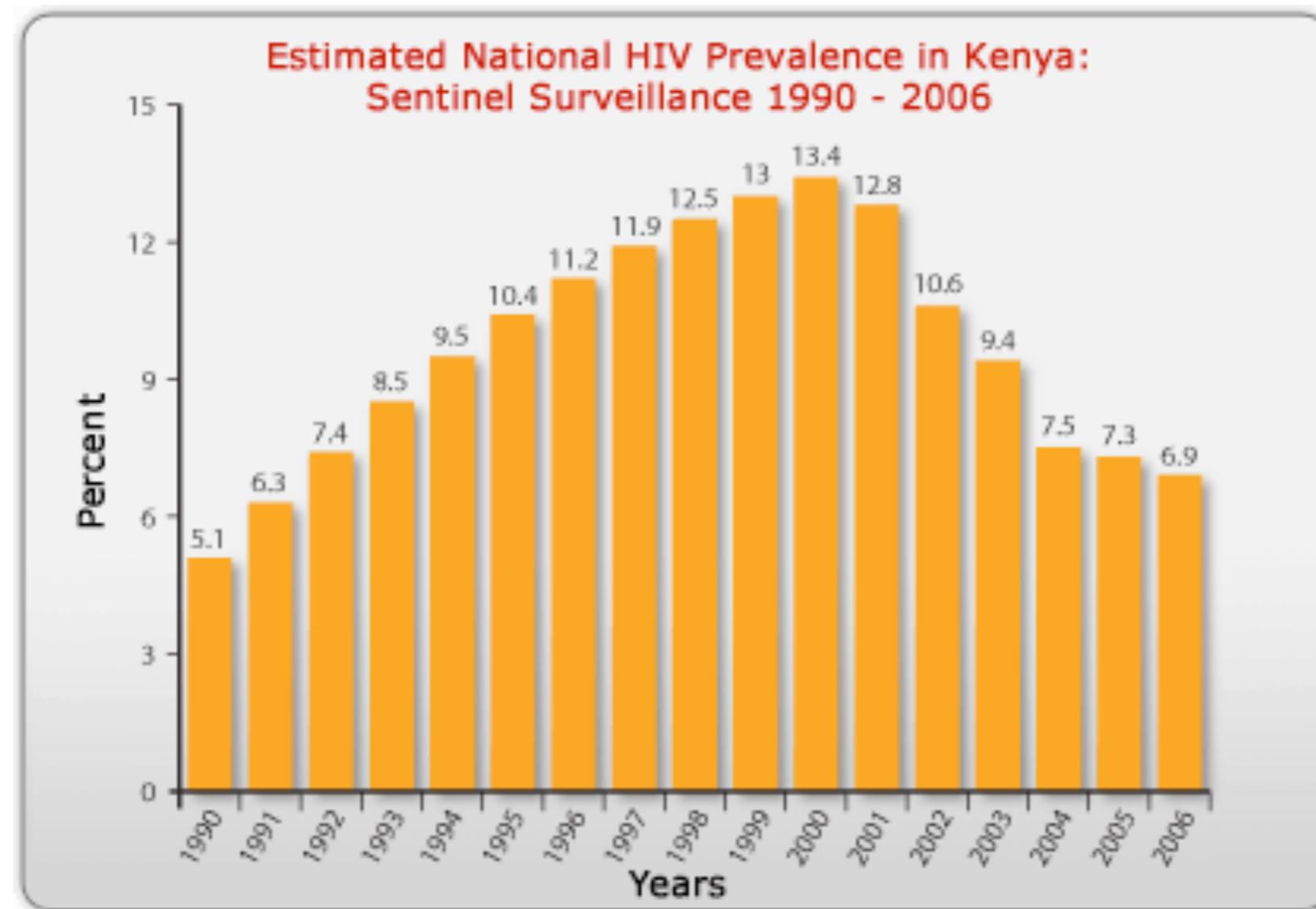
Sequence alignment



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

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

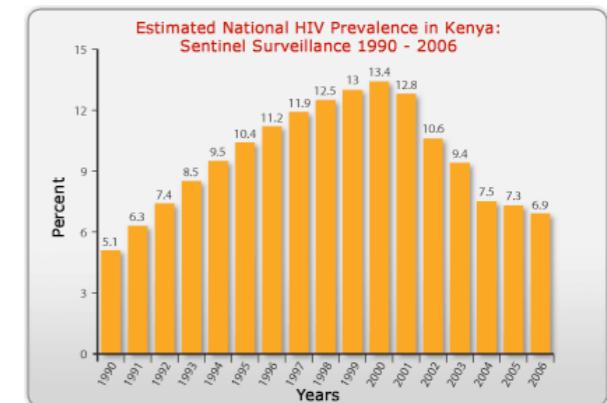
Phylogenetics in epidemiology



Phylogenetics in epidemiology

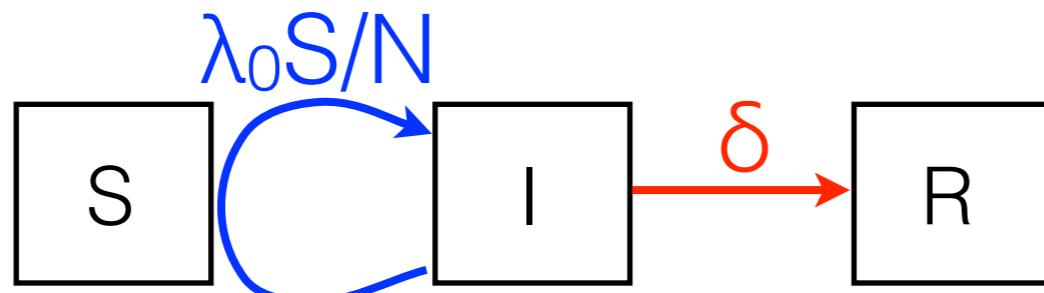
Data

of infected individuals through time



Models

Population dynamics described by SIR models:

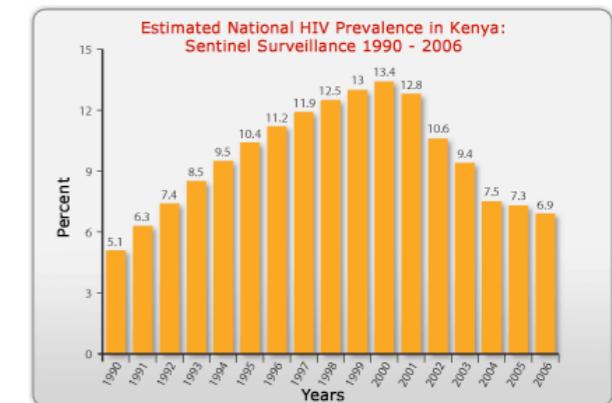


$$\begin{aligned}\frac{dS}{dt} &= -\lambda_0 IS/N \\ \frac{dI}{dt} &= \lambda_0 IS/N - \delta I \\ \frac{dR}{dt} &= \delta I\end{aligned}$$

Phylogenetics in epidemiology

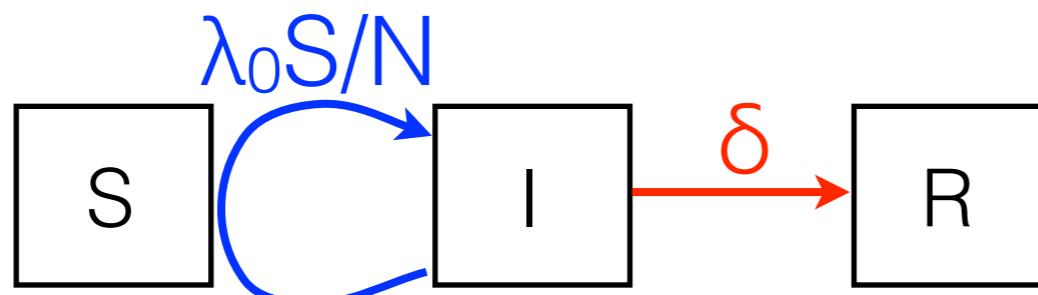
Data

of infected individuals through time



Models

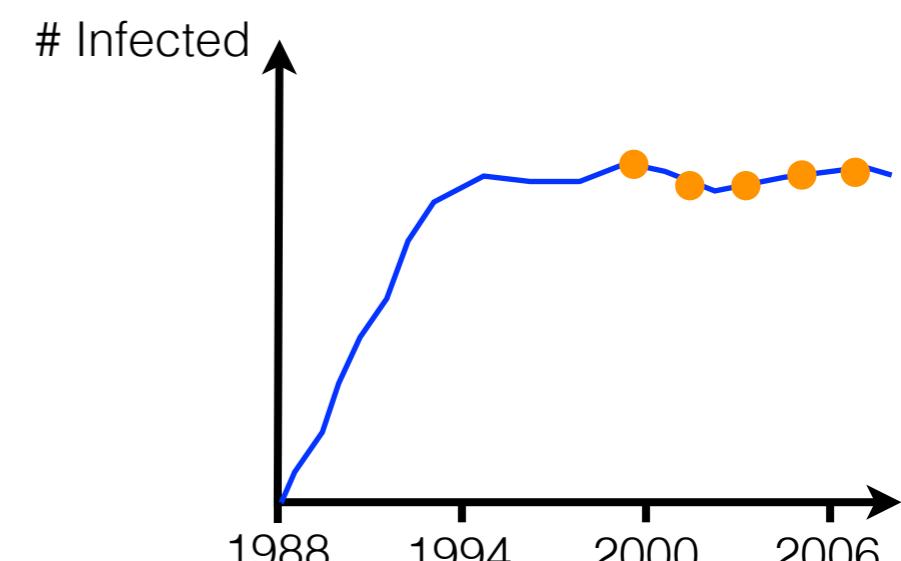
Population dynamics described by SIR models:



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

Limitations

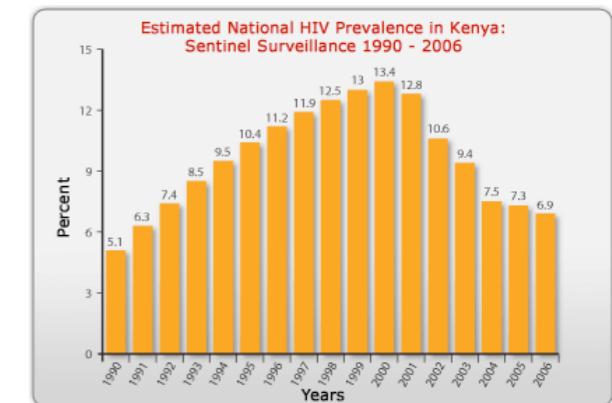
If sampling in early epidemic was missed:



Phylogenetics in epidemiology

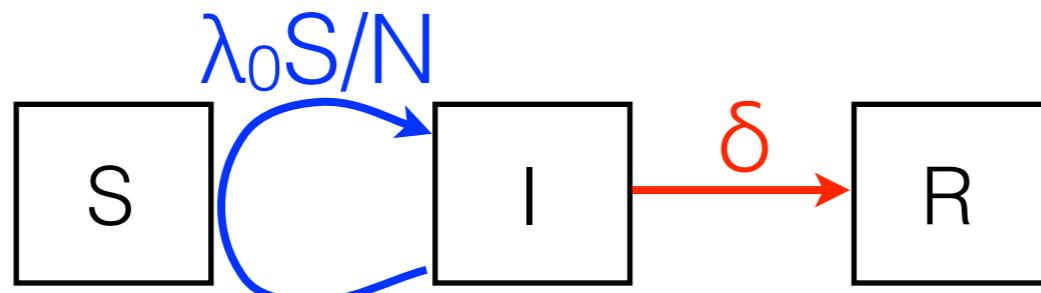
Data

of infected individuals through time



Models

Population dynamics described by SIR models:

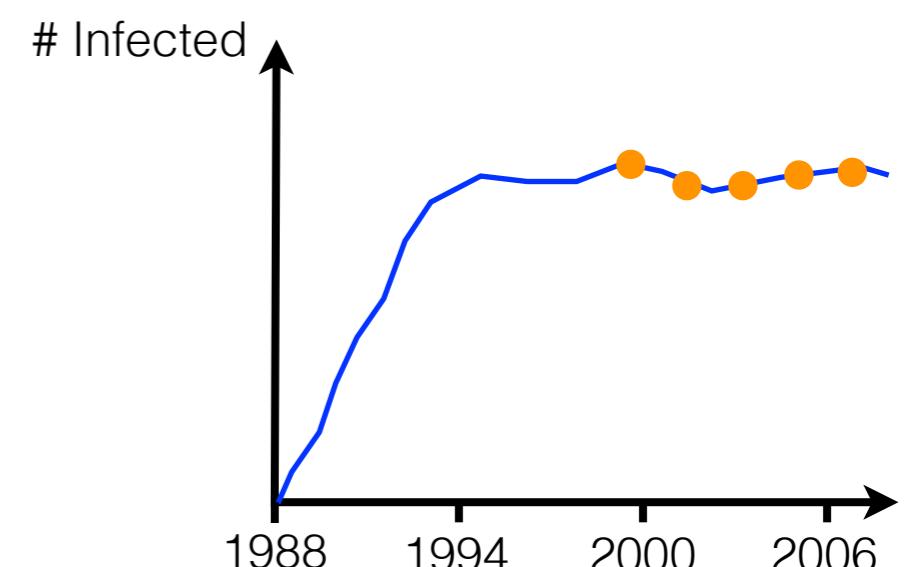


$$\begin{aligned}\frac{dS}{dt} &= -\lambda_0 IS/N \\ \frac{dI}{dt} &= \lambda_0 IS/N - \delta I \\ \frac{dR}{dt} &= \delta I\end{aligned}$$

Limitations

If sampling in early epidemic was missed:

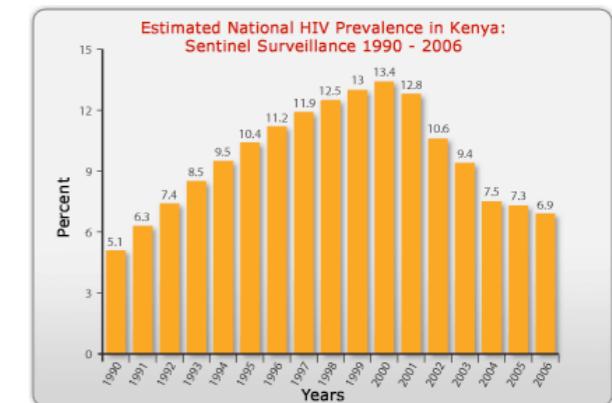
- ▶ **Time of epidemic outbreak?**
- ▶ **Basic reproductive number R_0 ?**



Phylogenetics in epidemiology

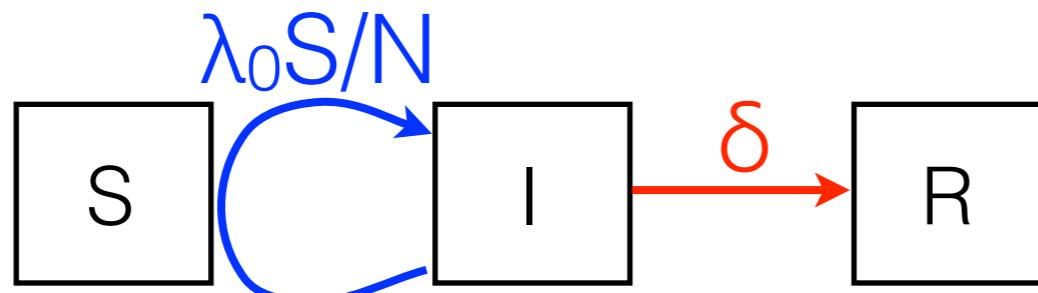
Data

of infected individuals through time



Models

Population dynamics described by SIR models:



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

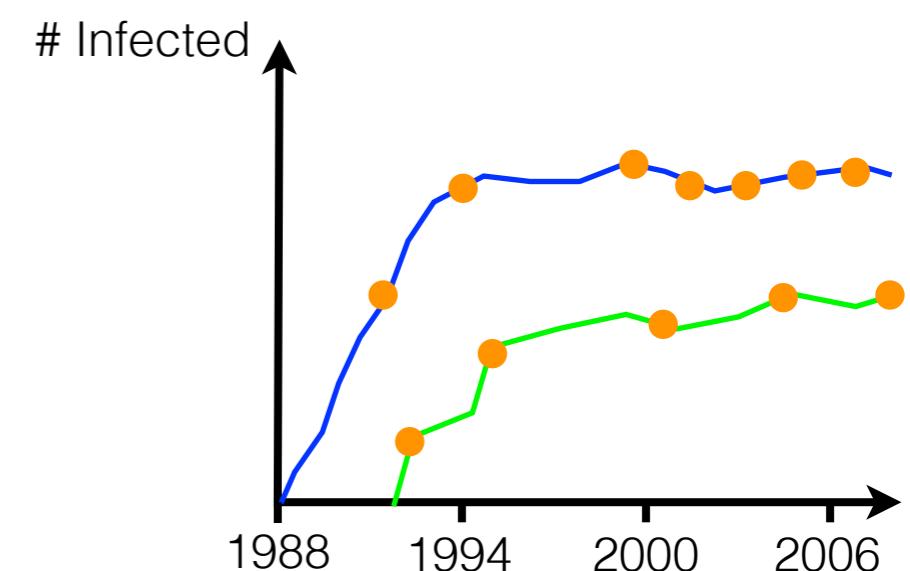
Limitations

If sampling in early epidemic was missed:

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

Data does not tell who infected whom:

- ▶ **Population structure?**



Genetic data entering epidemiology

Phylo-
genetics

Inf1 ACACACCC
Inf2 TCACACCT
Inf3 AAAGACTT
Inf4 ACAGACTT

Genetic data entering epidemiology

Phylo-
genetics

Input:

Pathogen genetic data from different hosts

Inf1	ACACACCC
Inf2	TCACACCT
Inf3	AAAGACTT
Inf4	ACAGACTT

Genetic data entering epidemiology

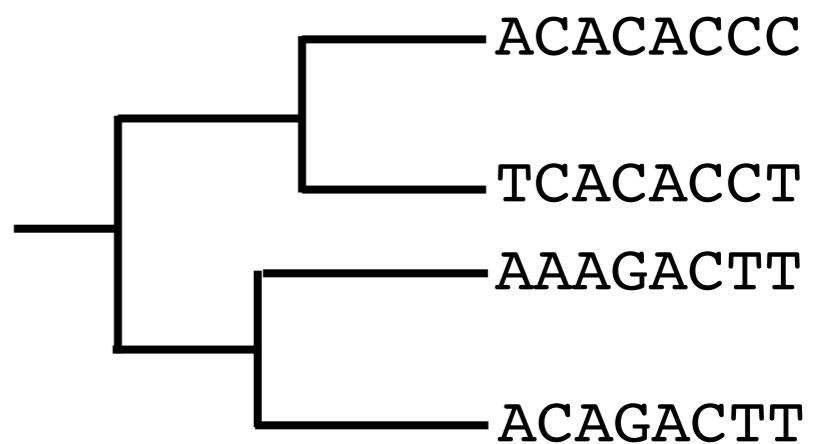
Phylogenetics

Input:

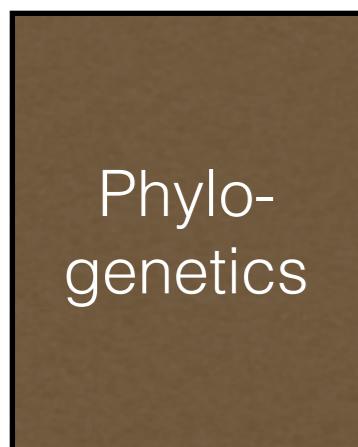
Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)



Genetic data entering epidemiology



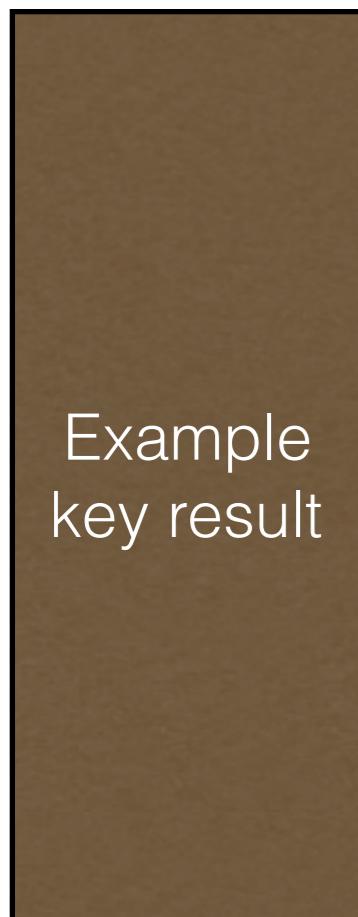
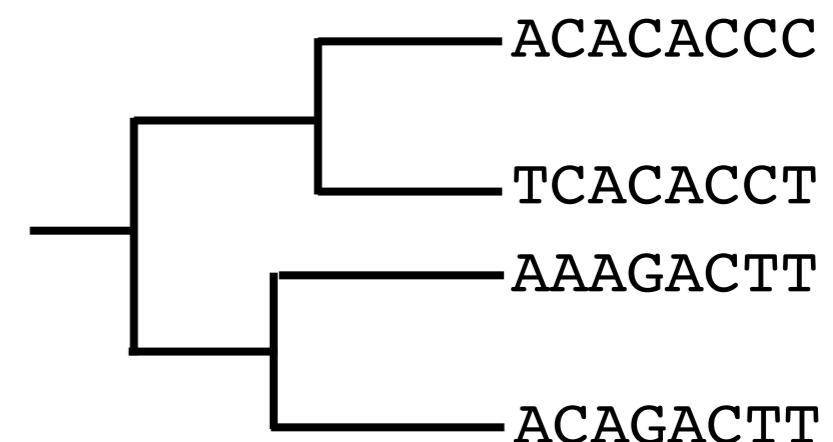
Input:

Pathogen genetic data from different hosts

Phy-
genetics

Output:

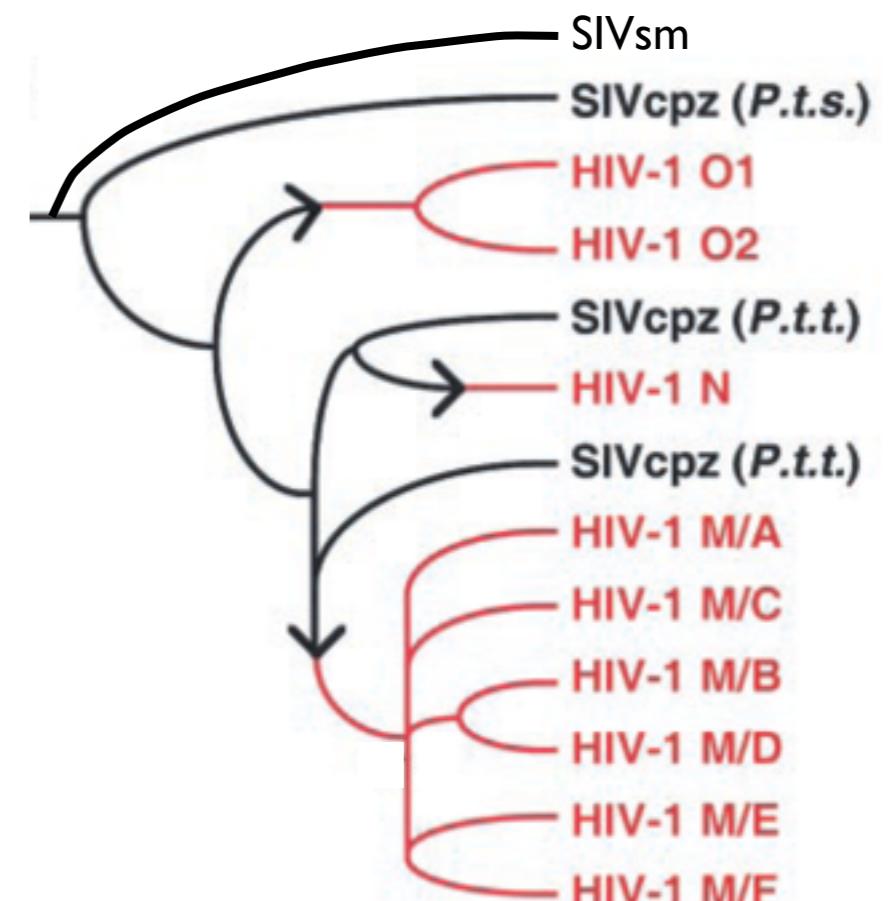
Transmission chain (who infected whom)



Emergence of HIV

Epidemic identified in 1980s

Example
key result



adopted from Hahn et al. (Science, 2000)

Genetic data entering epidemiology

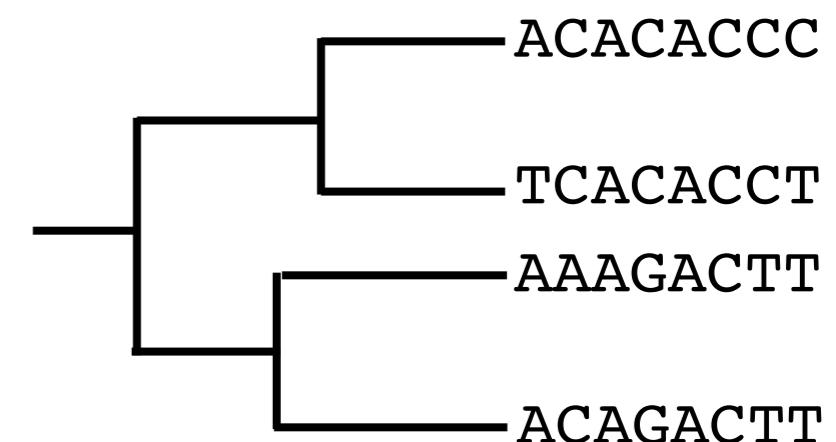
Phylogenetics

Input:

Pathogen genetic data from different hosts

Output:

Transmission chain (who infected whom)



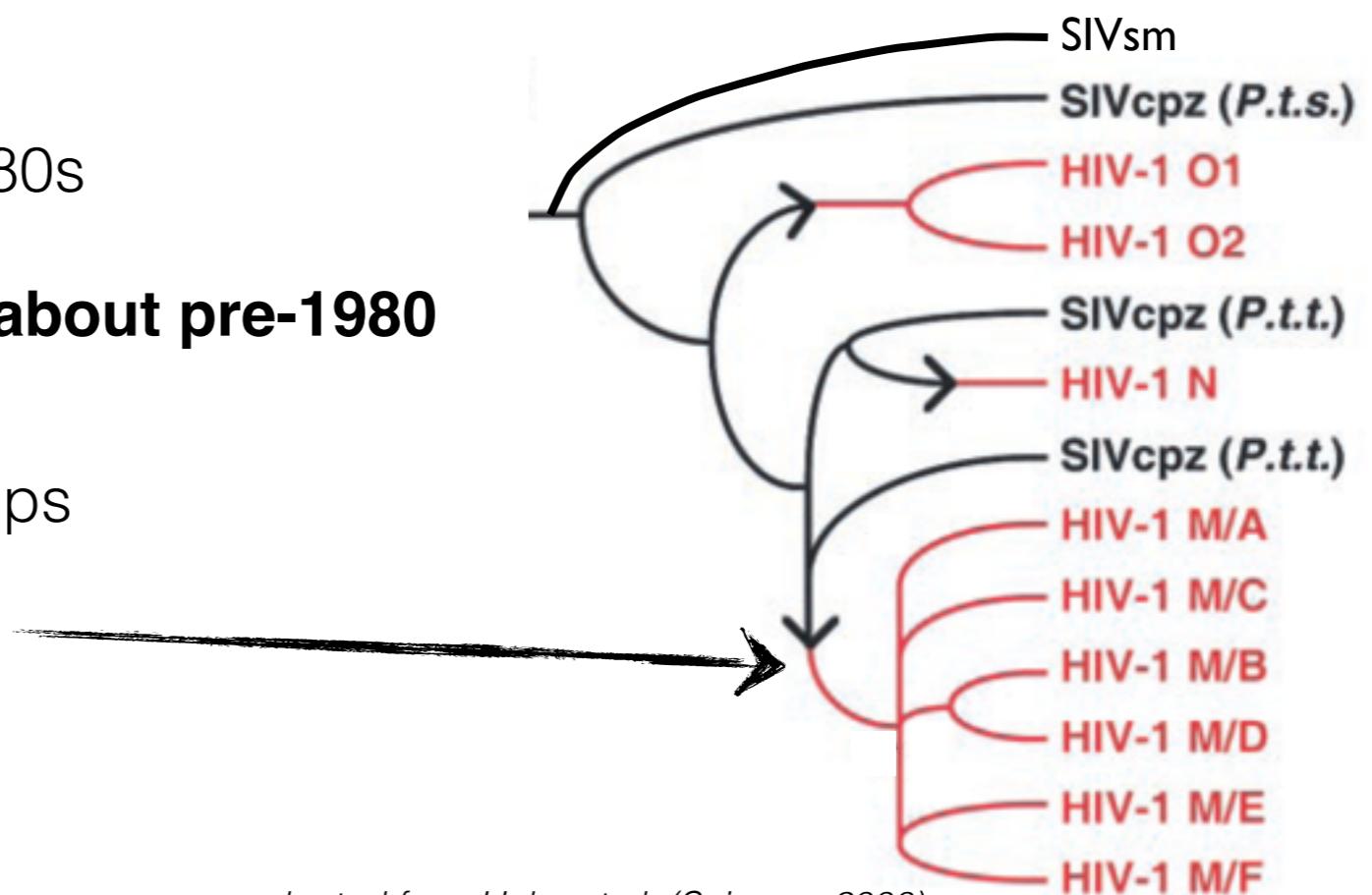
Example key result

Emergence of HIV

Epidemic identified in 1980s

Sequence data tells us about pre-1980

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



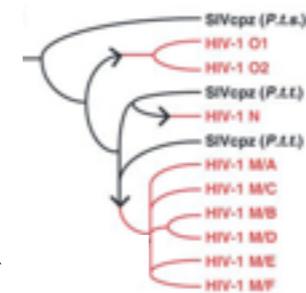
adopted from Hahn et al. (Science, 2000)

Genetic data contains information about the epidemiological dynamics

Phylogenetics

State of process

- ▶ Time and geographic location of outbreak



Emergence of a pathogen (here HIV)

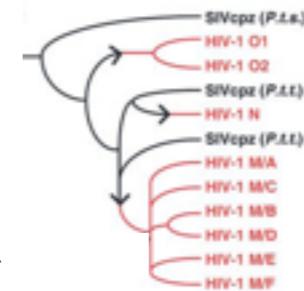
Hahn et al. (Science, 2000)

Genetic data contains information about the epidemiological dynamics

Phylogenetics

State of process

- ▶ Time and geographic location of outbreak



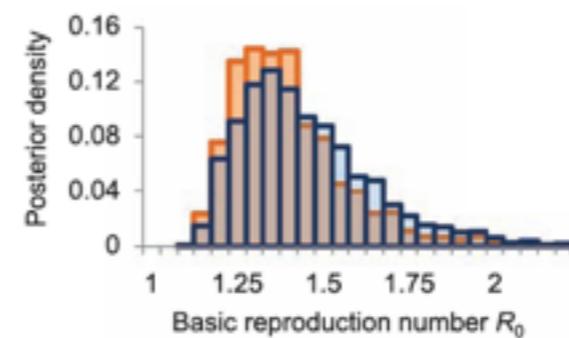
Emergence of a pathogen (here HIV)

Hahn et al. (Science, 2000)

Phylogenetics

Dynamics of process

- ▶ Transmission process



Basic reproductive number of an emerging pathogen (here H1N1)

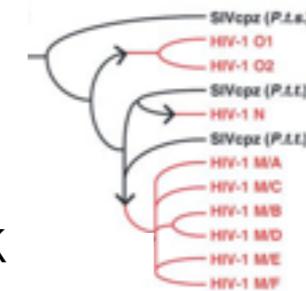
Fraser et al. (Science, 2009)

Genetic data contains information about the epidemiological dynamics

Phylogenetics

State of process

- ▶ Time and geographic location of outbreak



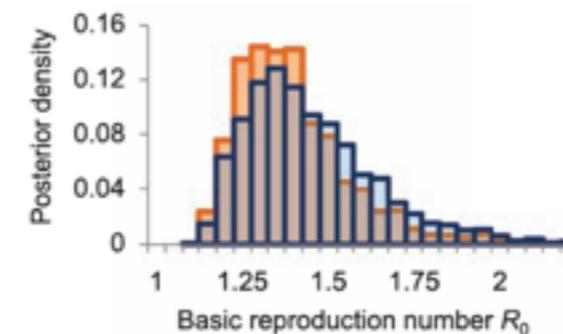
Emergence of a pathogen (here HIV)

Hahn et al. (Science, 2000)

Phylogenetics

Dynamics of process

- ▶ Transmission process



Basic reproductive number of an emerging pathogen (here H1N1)

Fraser et al. (Science, 2009)

Epidemiological models

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

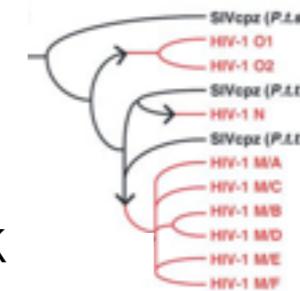
- ▶ Quantification of all model parameters!

Genetic data contains information about the epidemiological dynamics

Phylogenetics

State of process

- ▶ Time and geographic location of outbreak



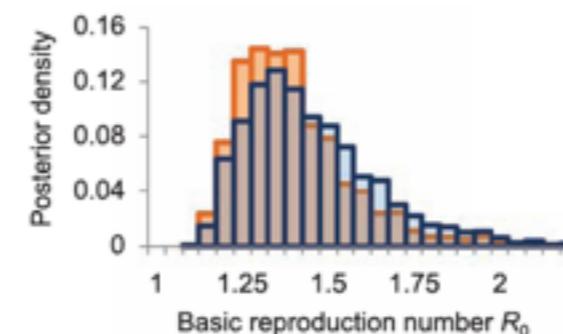
Emergence of a pathogen (here HIV)

Hahn et al. (Science, 2000)

Phylogenetics

Dynamics of process

- ▶ Transmission process



Basic reproductive number of an emerging pathogen (here H1N1)

Fraser et al. (Science, 2009)

Epidemiological models

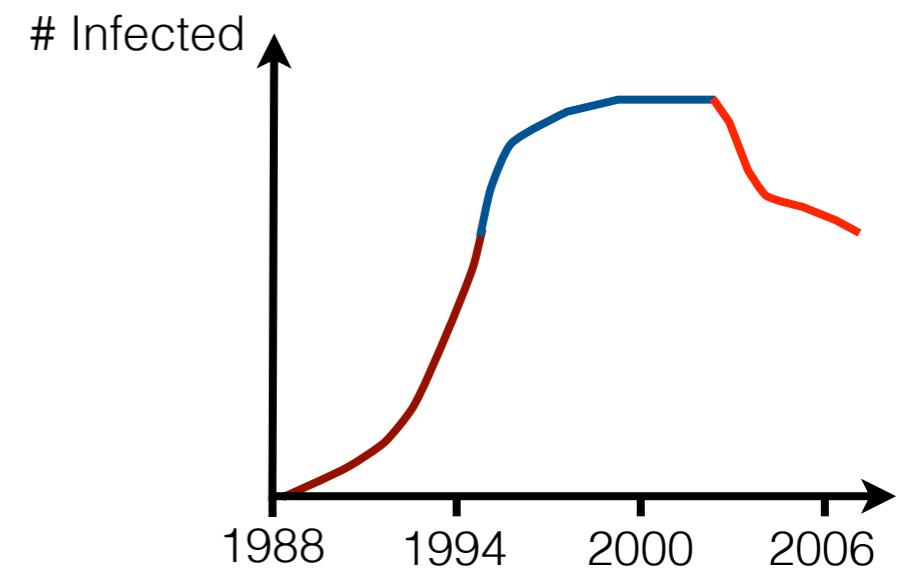
Ideally: assume epidemiological model in phylogenetics (birth-death model)

- ▶ Quantification of all model parameters!

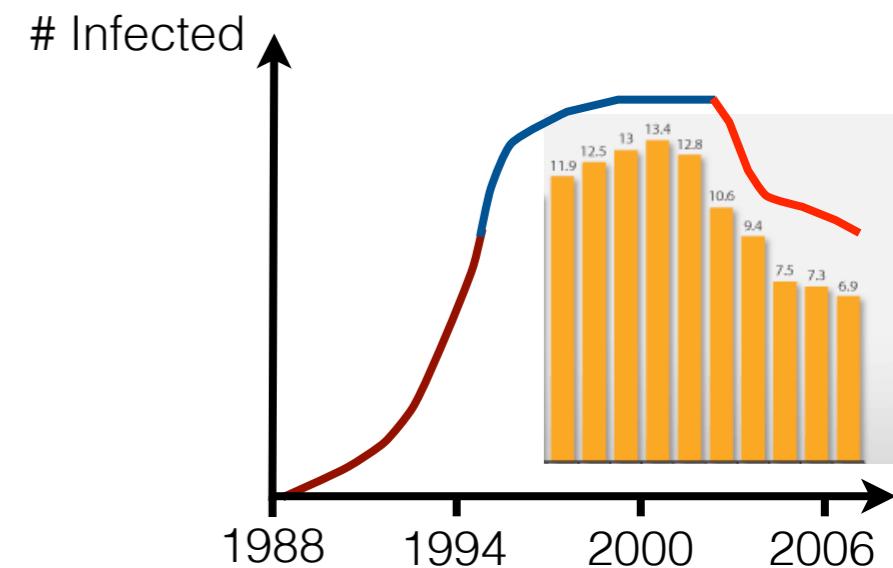
Approximate: by population genetic model (coalescent)

- ▶ Deterministic population size is parameterized!

Phylogenetic methods for different epidemic dynamics



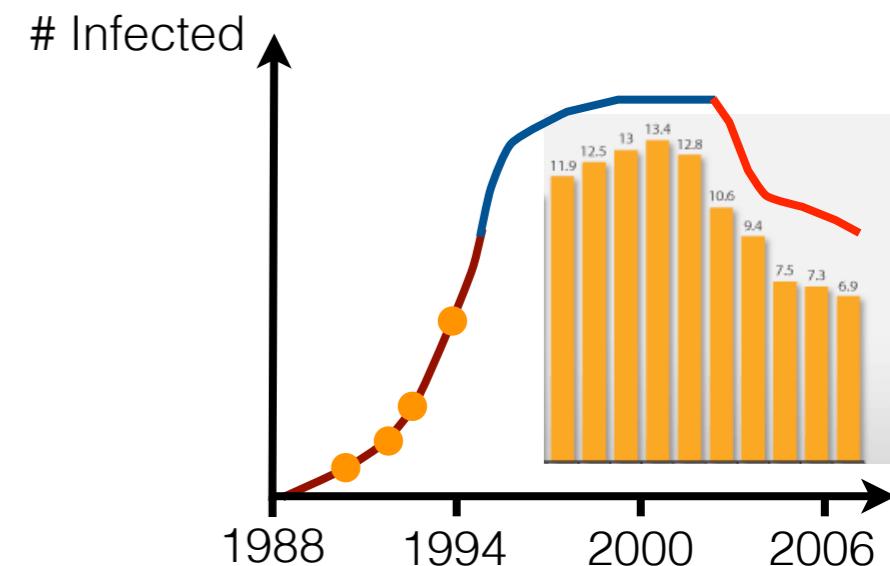
Phylogenetic methods for different epidemic dynamics



Phylogenetic methods for different epidemic dynamics

1. ...for analyzing epidemic outbreaks

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



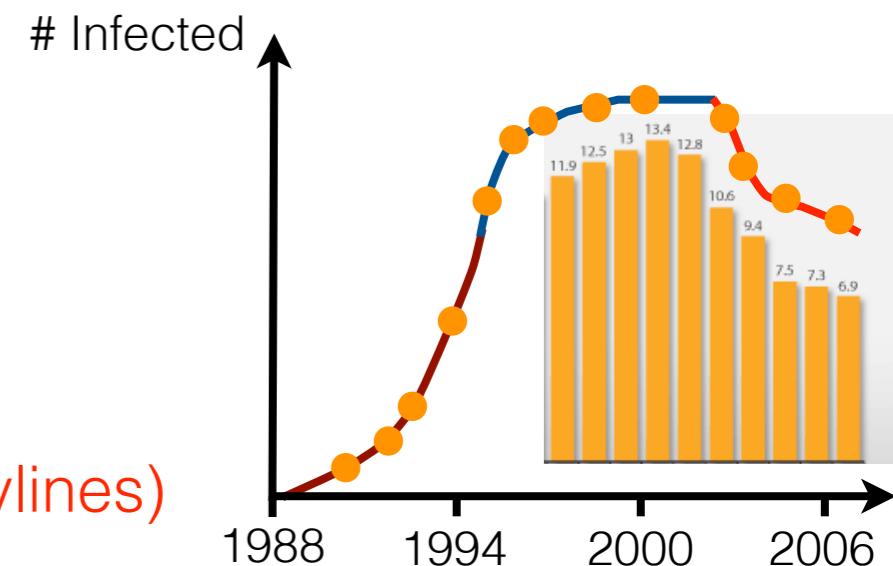
Phylogenetic methods for different epidemic dynamics

1. ...for analyzing epidemic outbreaks

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

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

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



Phylogenetic methods for different epidemic dynamics

1. ...for analyzing epidemic outbreaks

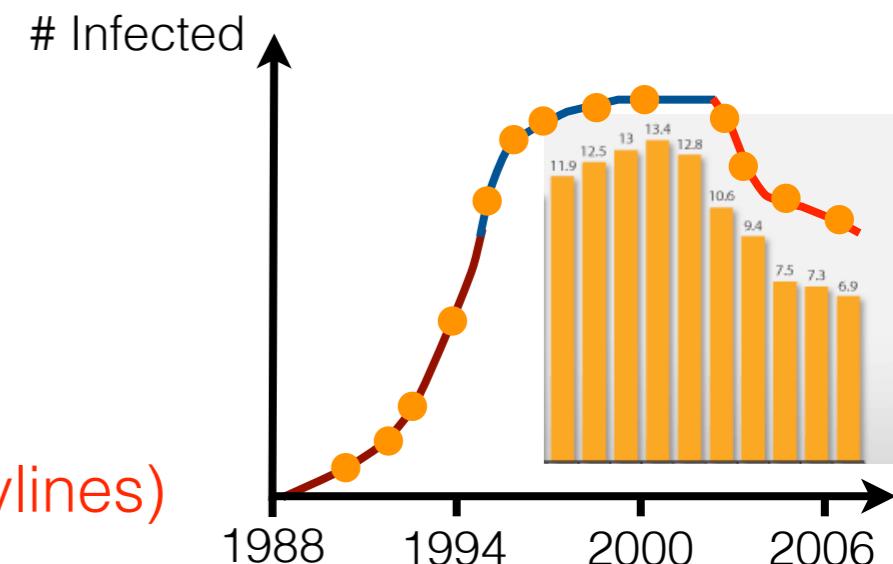
- ▶ C: Drummond et al. (*Genetics*, 2002)
- ▶ BD: Stadler et al. (*MBE*, 2012)

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

- ▶ 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), , Vaughan et al. (*submitted*)



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 (skylines)

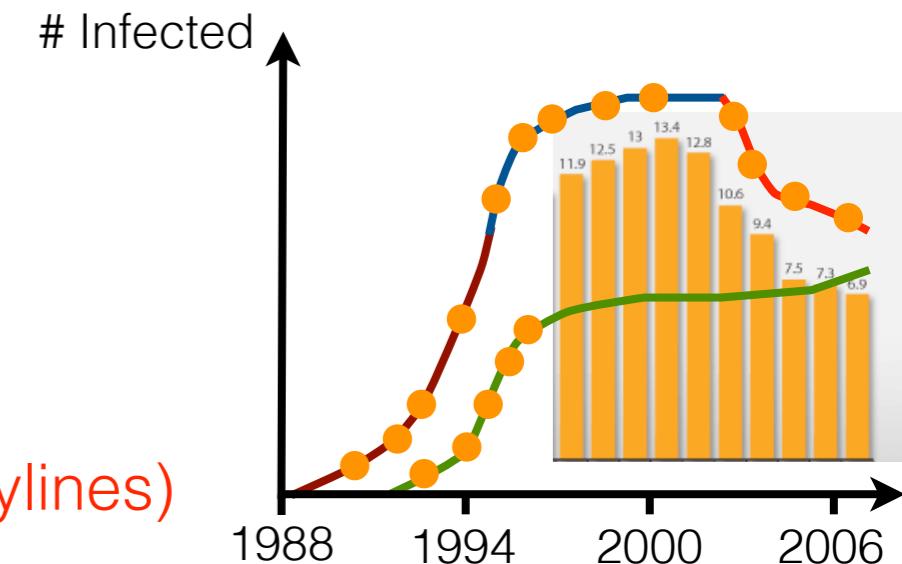
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

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), , Vaughan et al. (submitted)

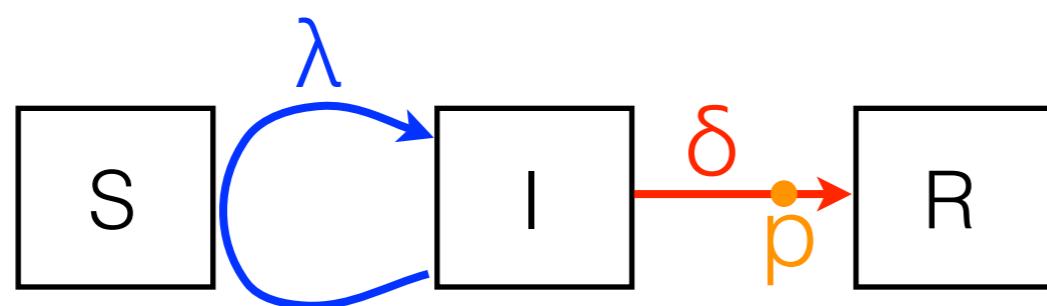
4. ...accounting for population structure

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



Birth-death model as a model for epidemic spread

Birth-death
model

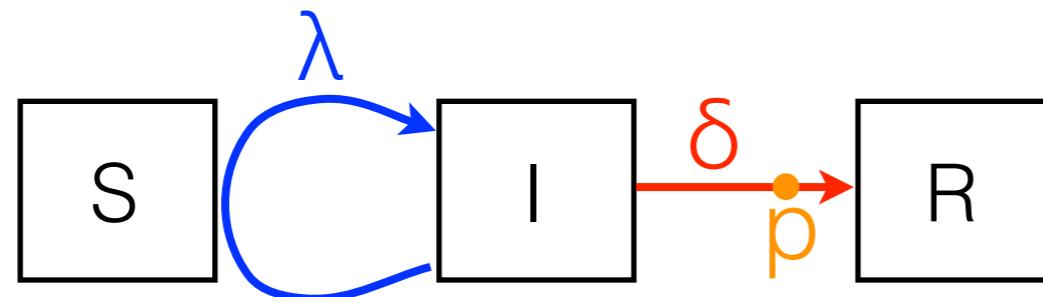


- transmission rate λ
- “becoming-non-infectious” rate δ
- sampling probability p

Parameters may depend on:

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

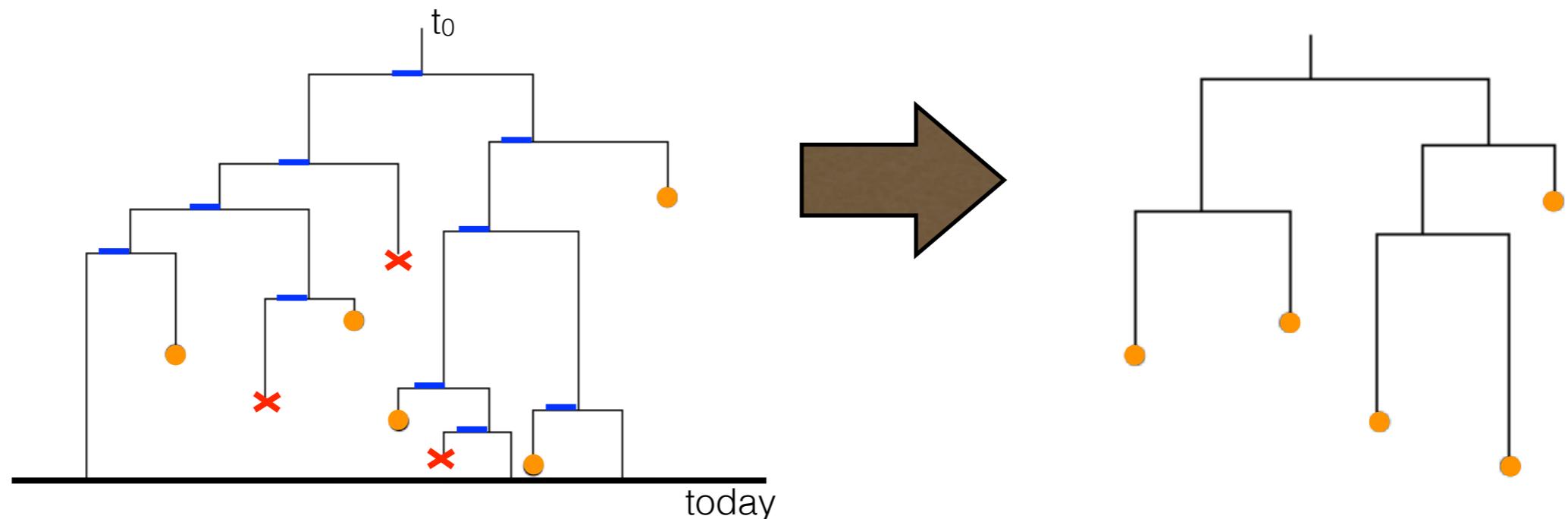
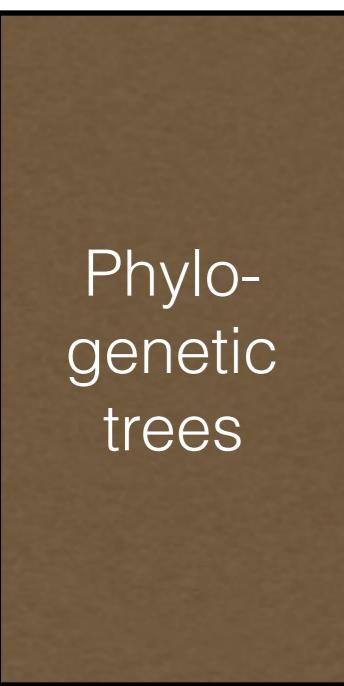
Birth-death model as a model for epidemic spread



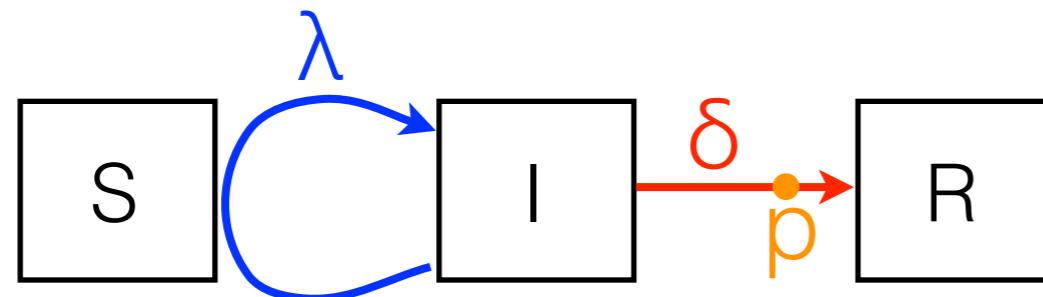
- transmission rate λ
- “becoming-non-infectious” rate δ
- sampling probability p

Parameters may depend on:

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



Birth-death model as a model for epidemic spread

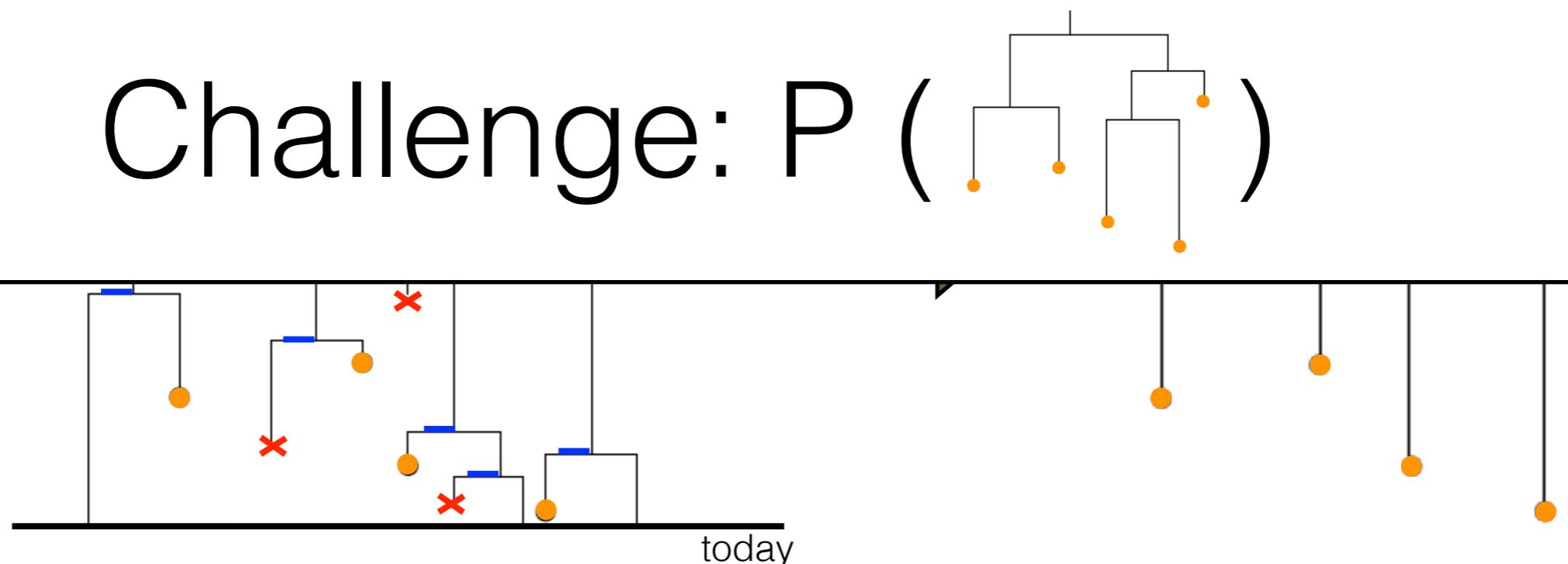
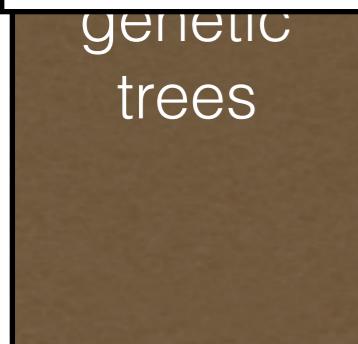


- transmission rate λ
- “becoming-non-infectious” rate δ
- sampling probability p

Parameters may depend on:

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

Challenge: $P(\text{ })$



Phylogenetic methods for different epidemic dynamics

1. ...for analyzing epidemic outbreaks

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

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

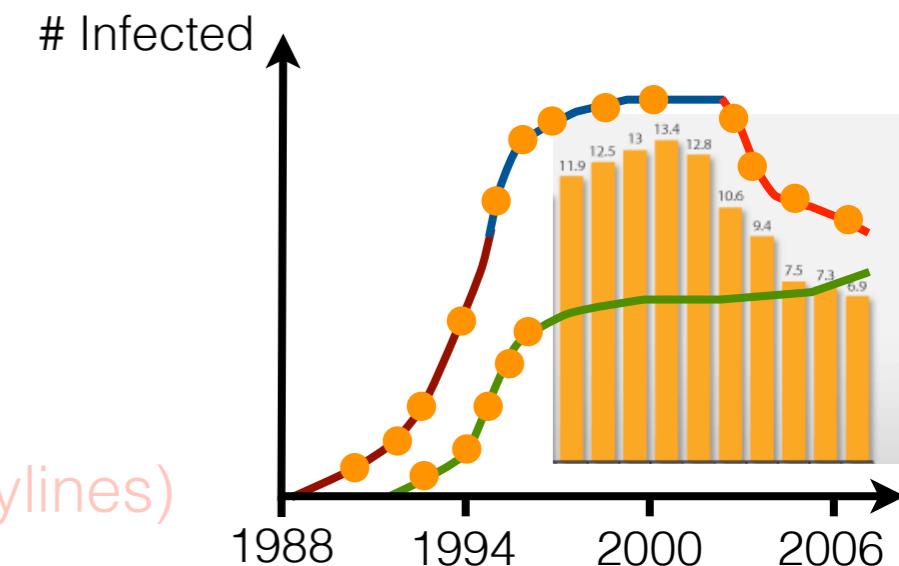
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

3. ...accounting for SIR model dynamics

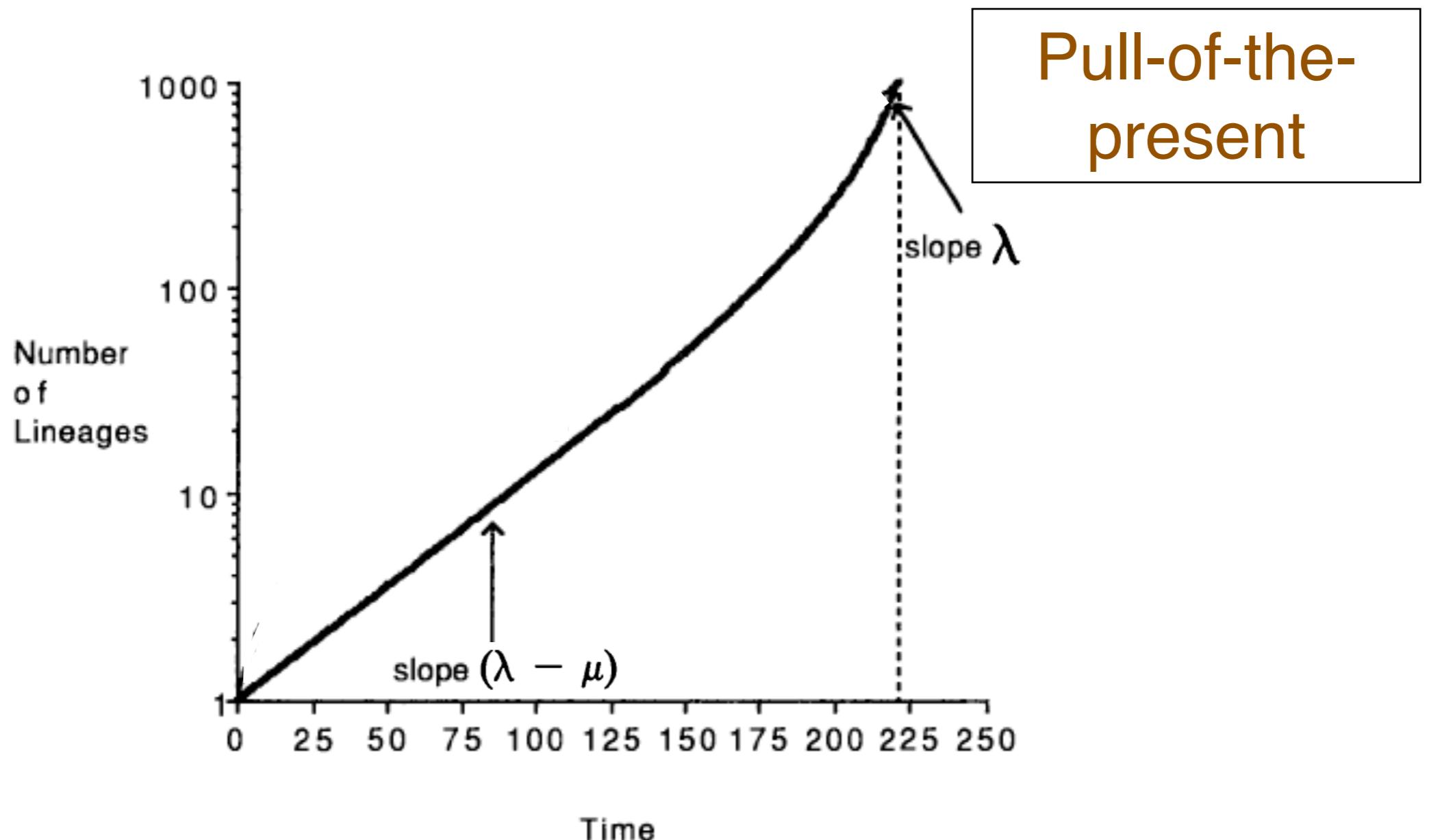
- ▶ C: Volz et al. (*Genetics*, 2009)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014), Leventhal et al. (*MBE*, 2014), , Vaughan et al. (submitted)

4. ...accounting for population structure

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

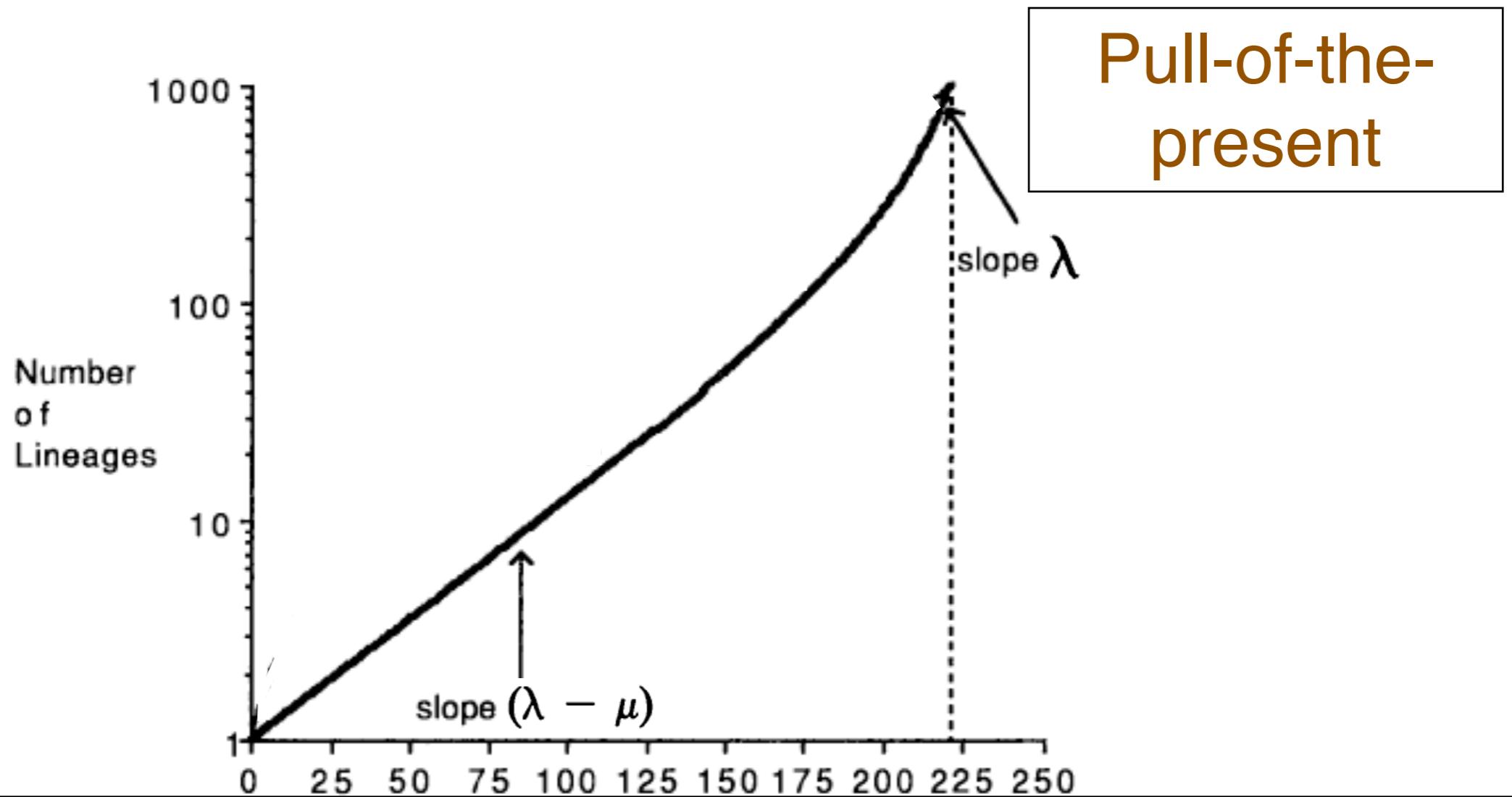


Fingerprint of the birth-and death rates in phylogenies



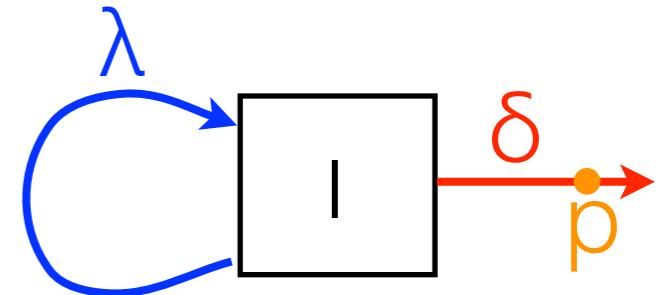
Harvey et al (1994, Evolution)

Fingerprint of the birth-and death rates in phylogenies



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

Tree probability density



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

where

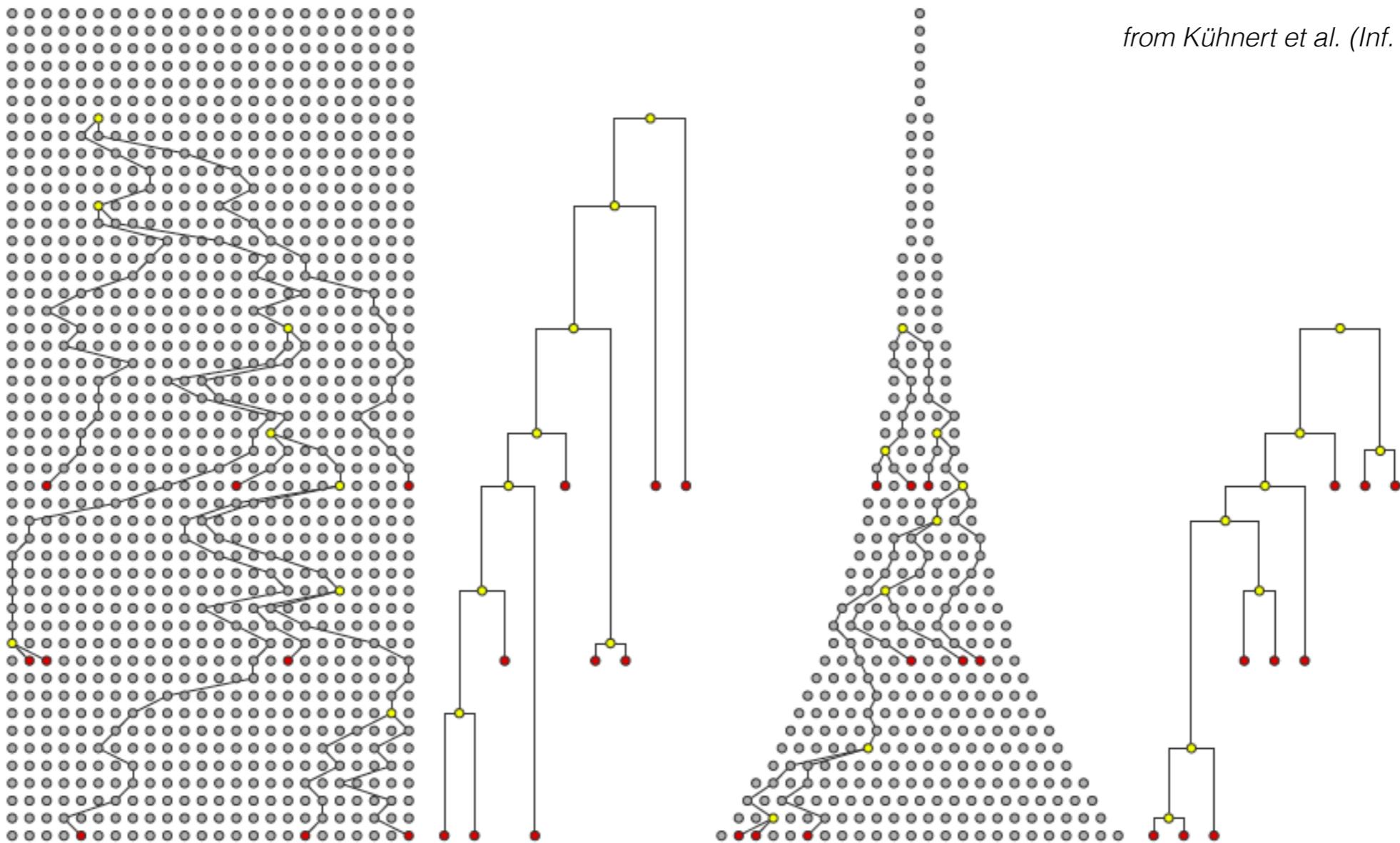
$$q(t) = 2(1 - c_2^2) + e^{-c_1 t}(1 - c_2)^2 + e^{c_1 t}(1 + c_2)^2$$

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

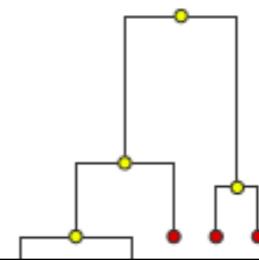
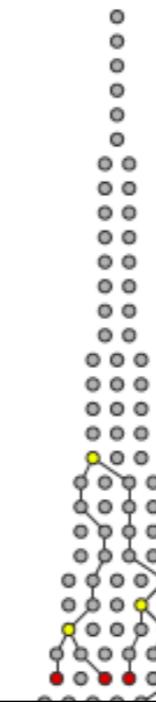
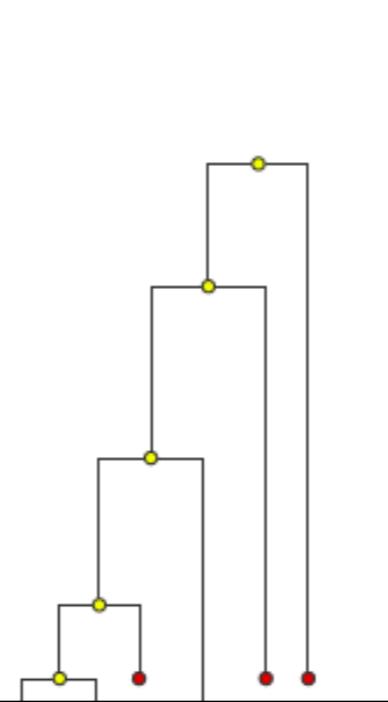
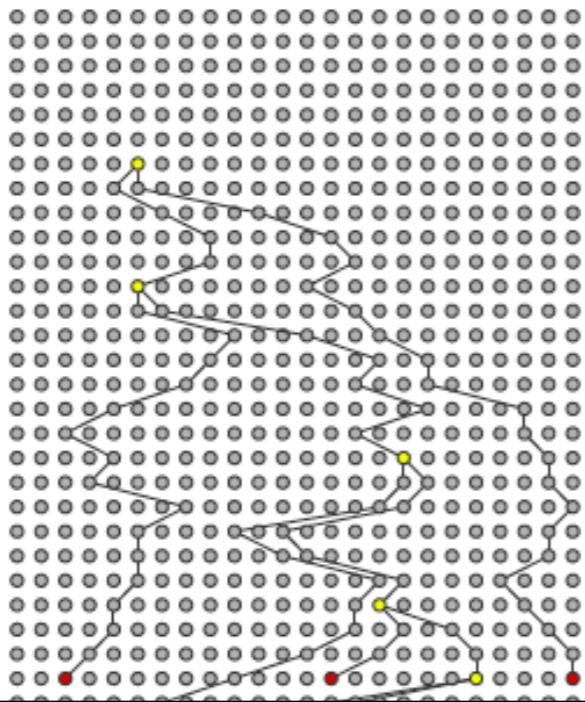
$$c_2 = -\frac{\lambda - \delta}{c_1}$$

and x_i branching times, y_i sampling times.

Coalescent as an alternative model for epidemic spread

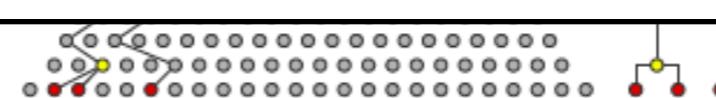
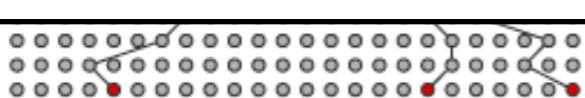


Coalescent as an alternative model for epidemic spread

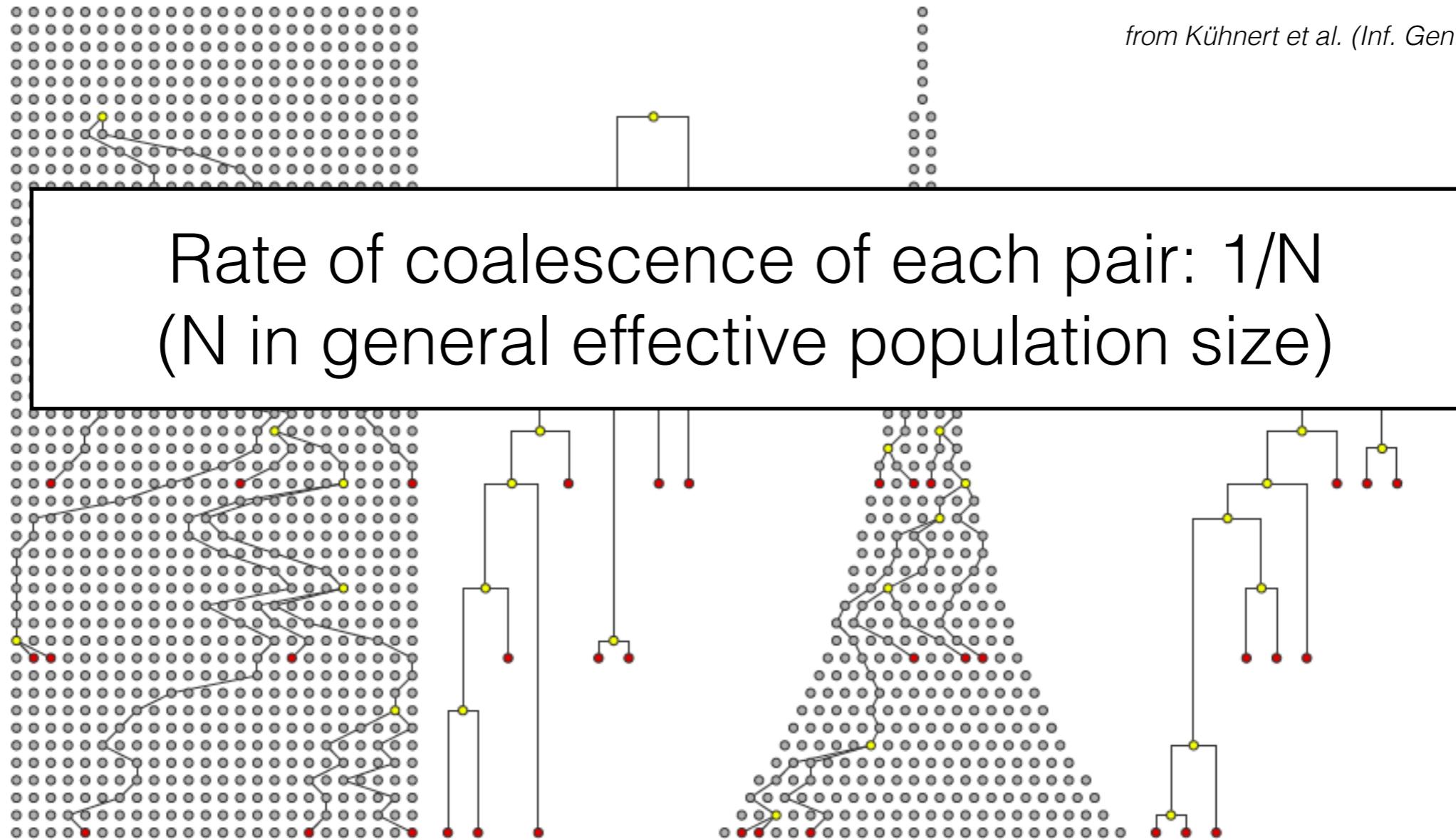


from Kühnert et al. (*Inf. Gen. Evol.*, 2011)

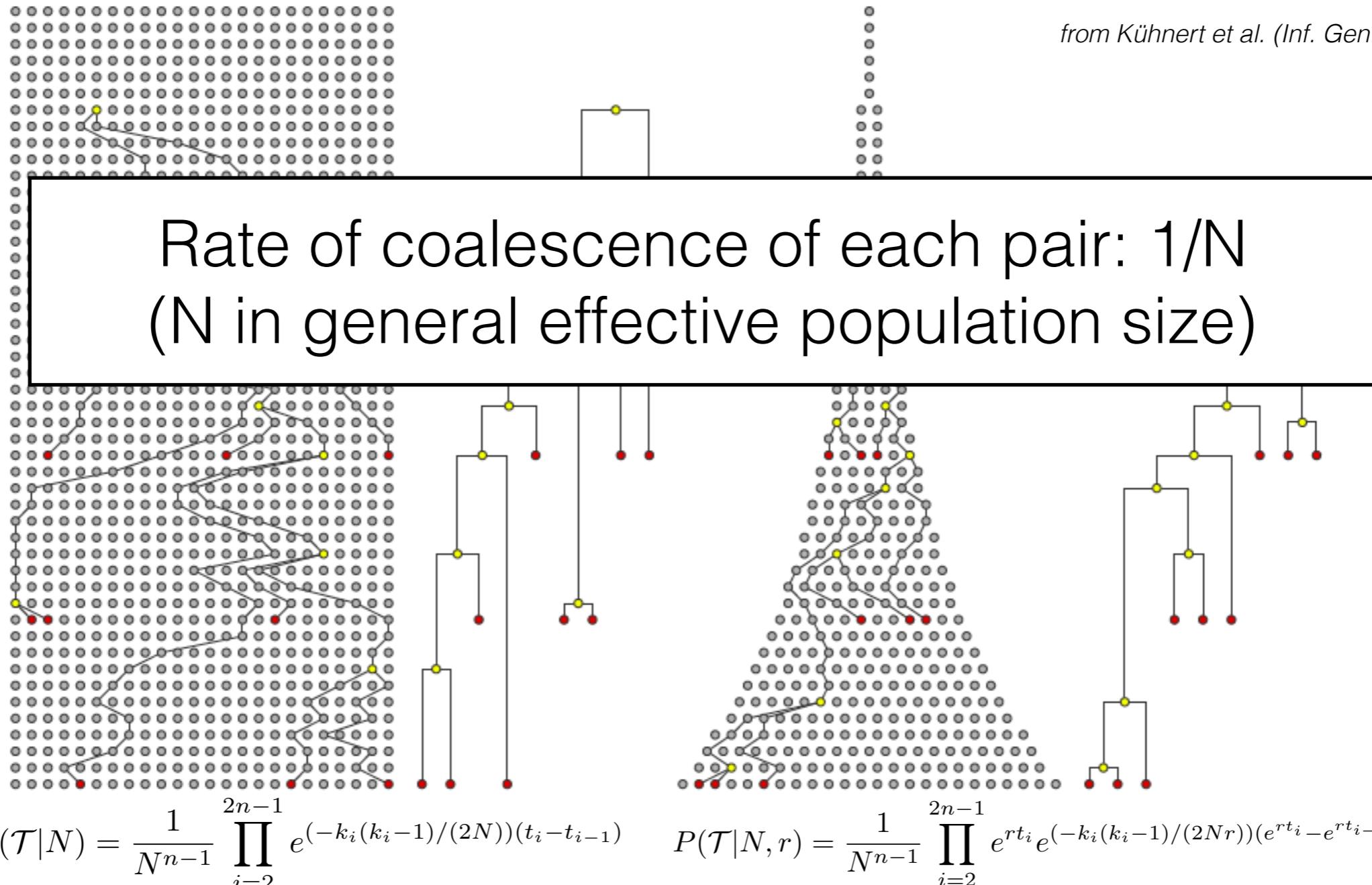
$$P \left(\text{ } \right) ?$$



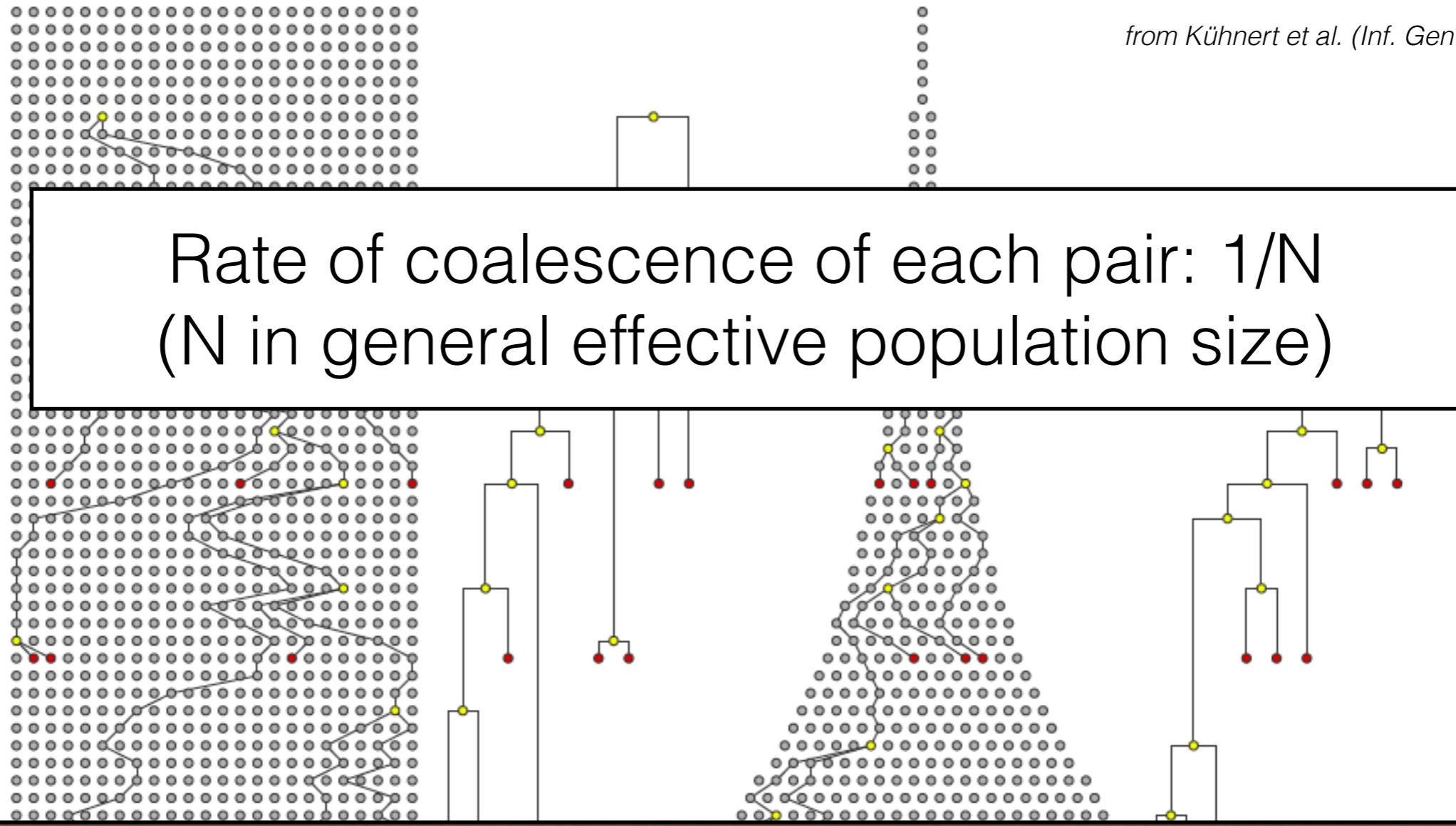
Coalescent as an alternative model for epidemic spread



Coalescent as an alternative model for epidemic spread



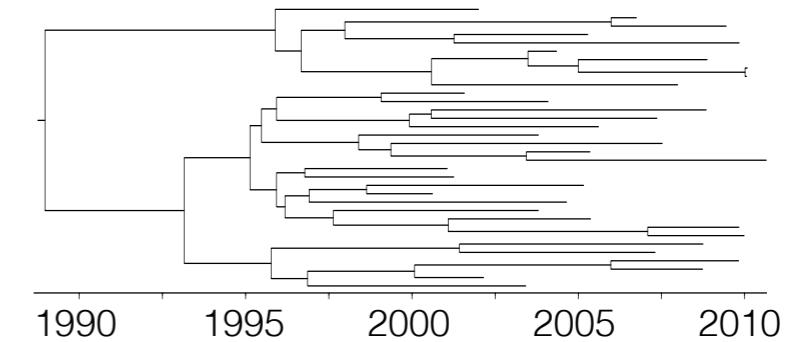
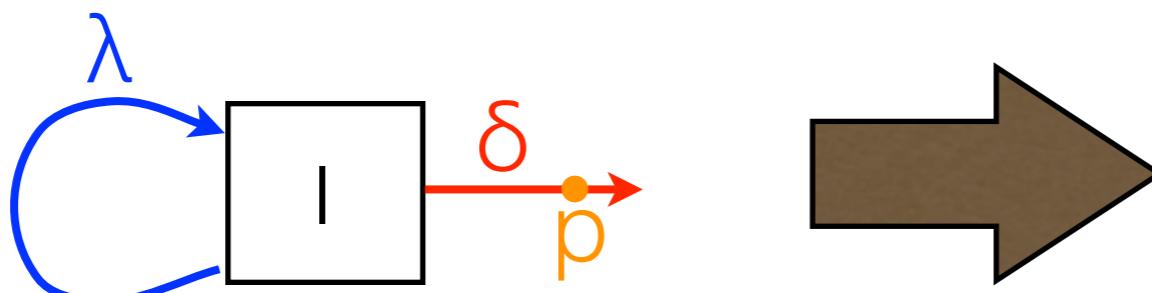
Coalescent as an alternative model for epidemic spread



Many short branches indicate small population!

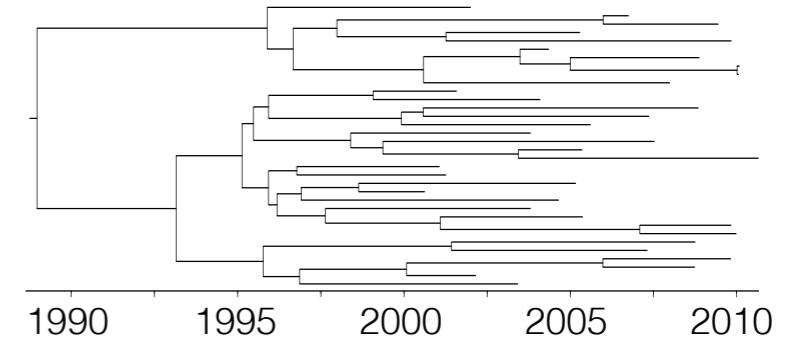
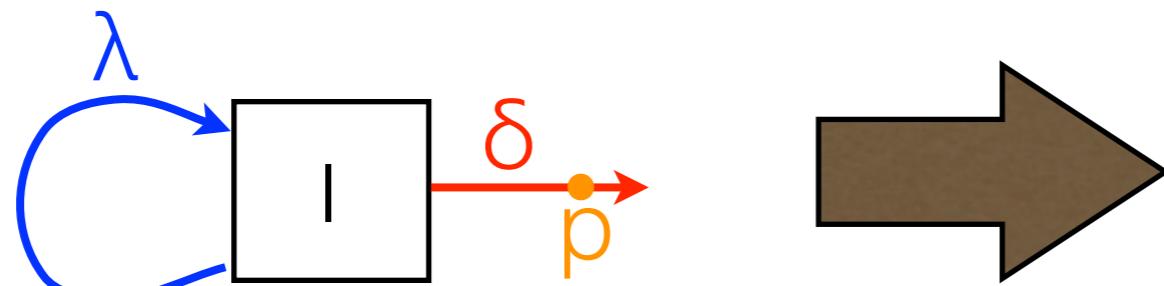
Simulation study: Comparing birth-death model to coalescent

Simulating
epidemic
outbreak



Simulation study: Comparing birth-death model to coalescent

Simulating
epidemic
outbreak



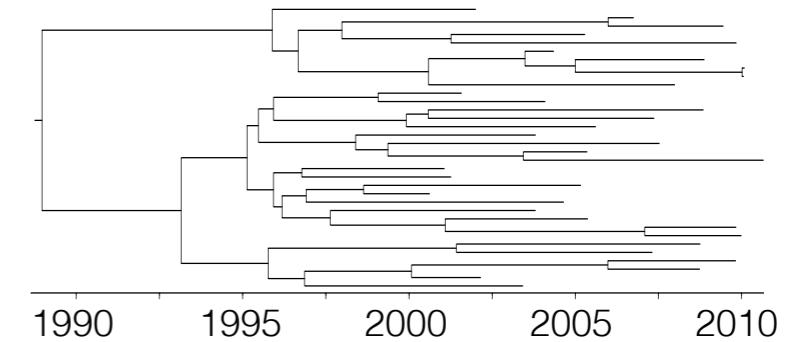
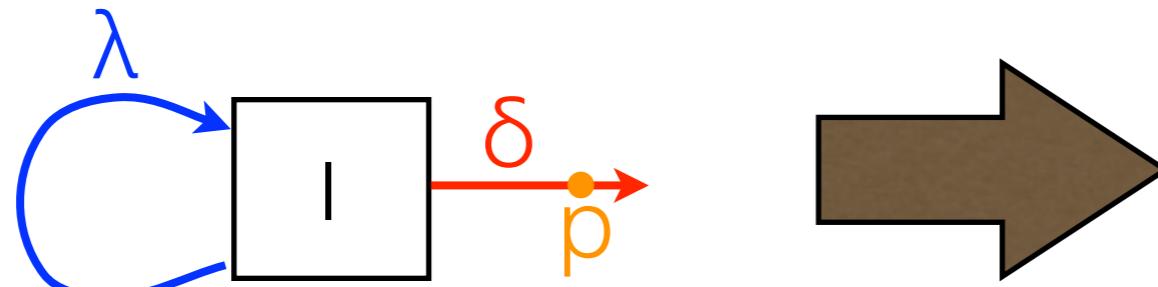
Exact
method

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

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

Simulation study: Comparing birth-death model to coalescent

Simulating
epidemic
outbreak



Exact
method

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

Coales-
cent
approxi-
mation

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

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 (skylines)

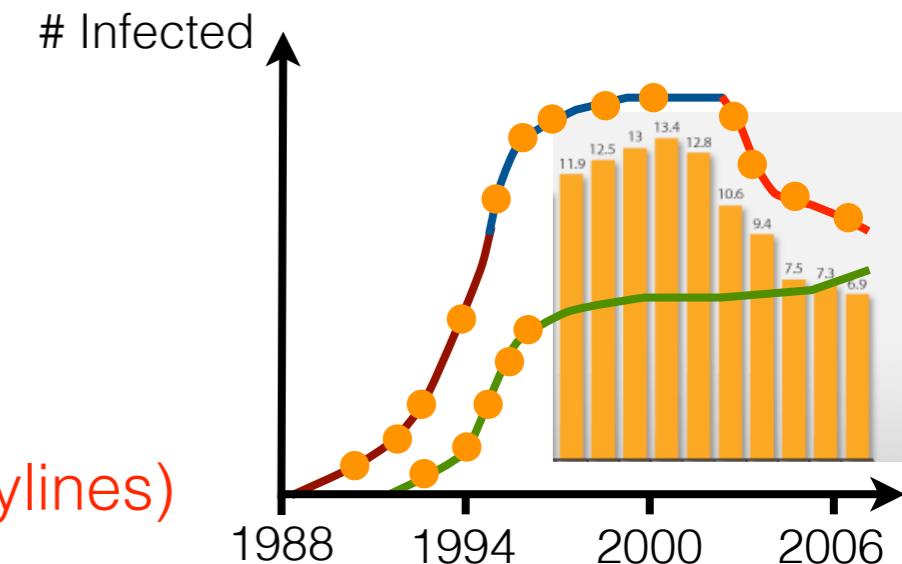
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

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), , Vaughan et al. (submitted)

4. ...accounting for population structure

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



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 (skylines)

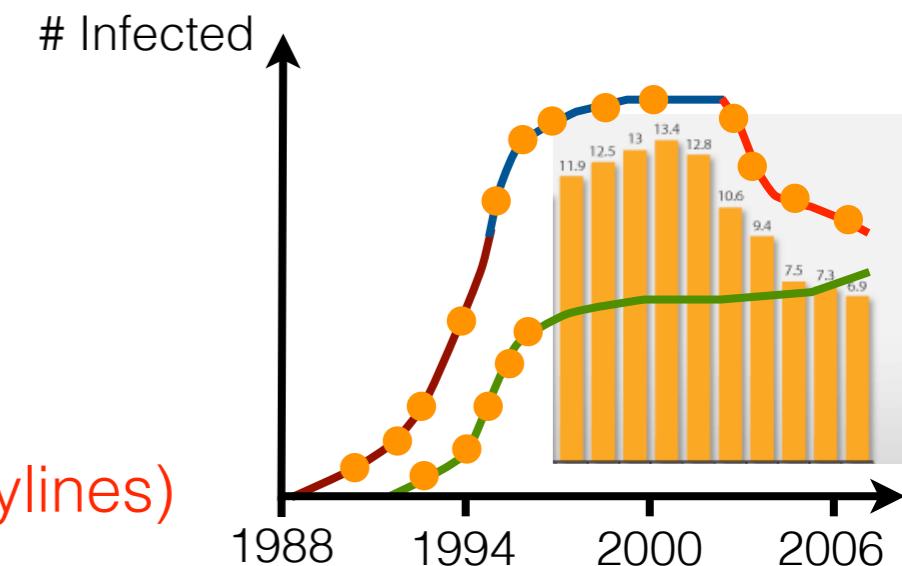
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

3. ...accounting for SIR model dynamics

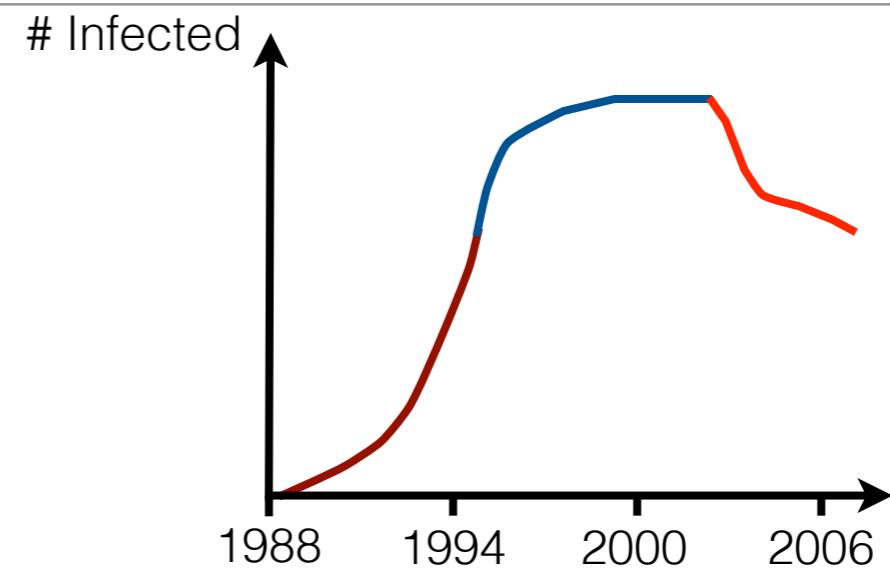
- ▶ C: Volz et al. (*Genetics*, 2009)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014), Leventhal et al. (*MBE*, 2014), , Vaughan et al. (submitted)

4. ...accounting for population structure

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



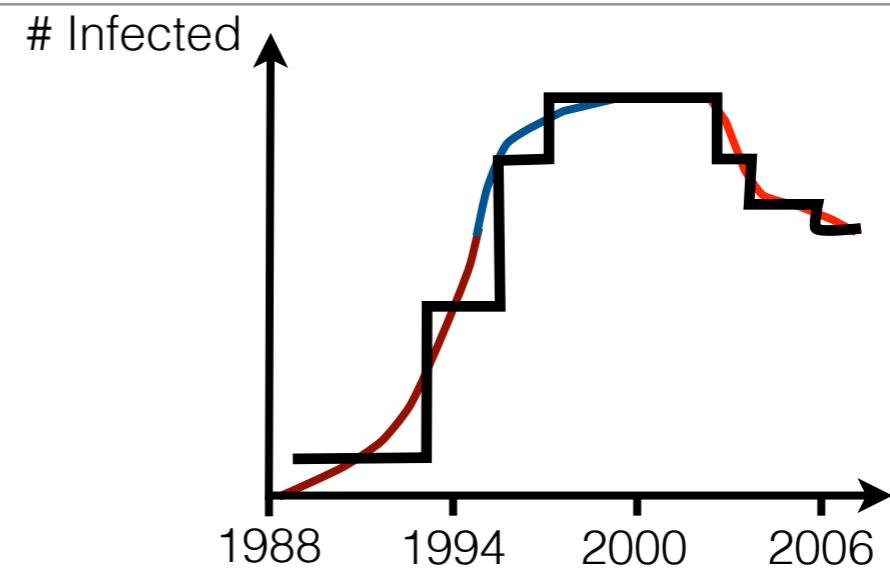
Bayesian skyline plot relaxes constant growth rate



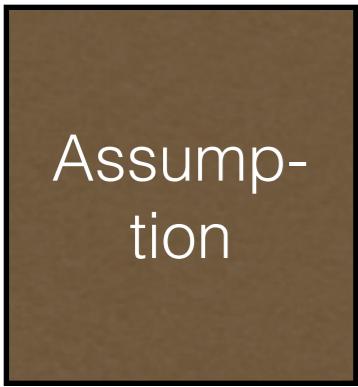
Bayesian skyline plot relaxes constant growth rate



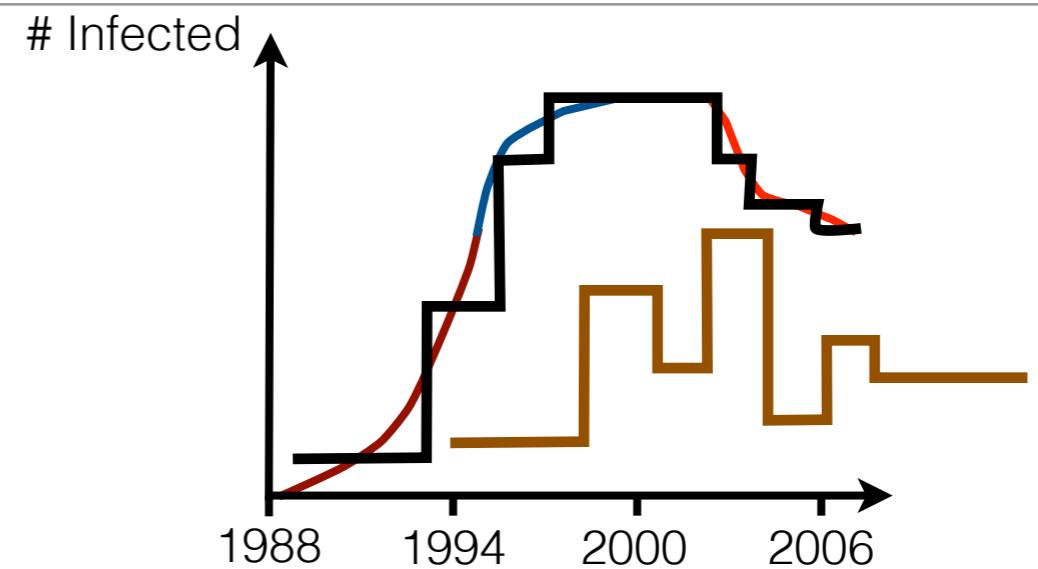
Piecewise constant population size



Bayesian skyline plot relaxes constant growth rate



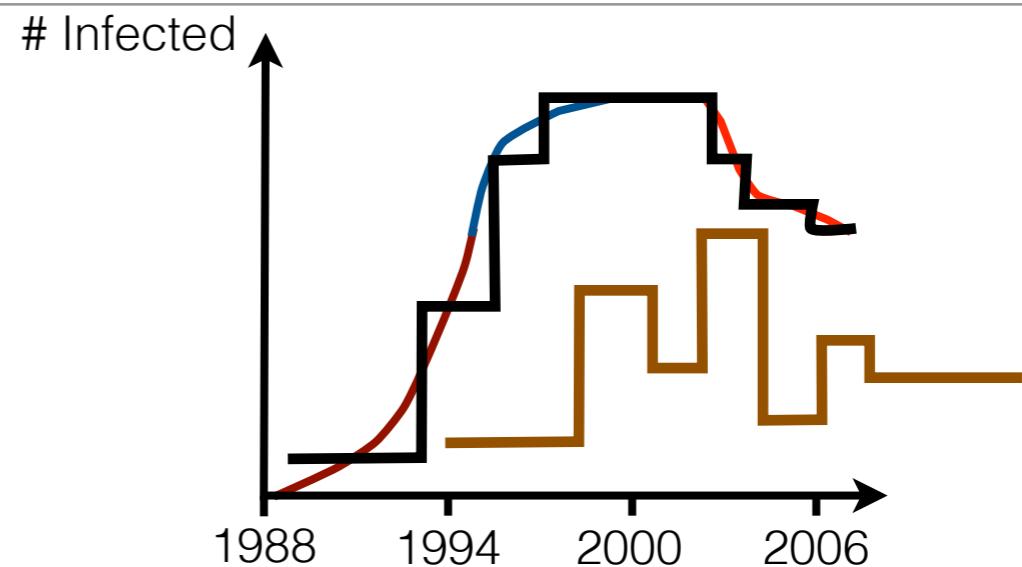
Piecewise constant population size



Bayesian skyline plot relaxes constant growth rate

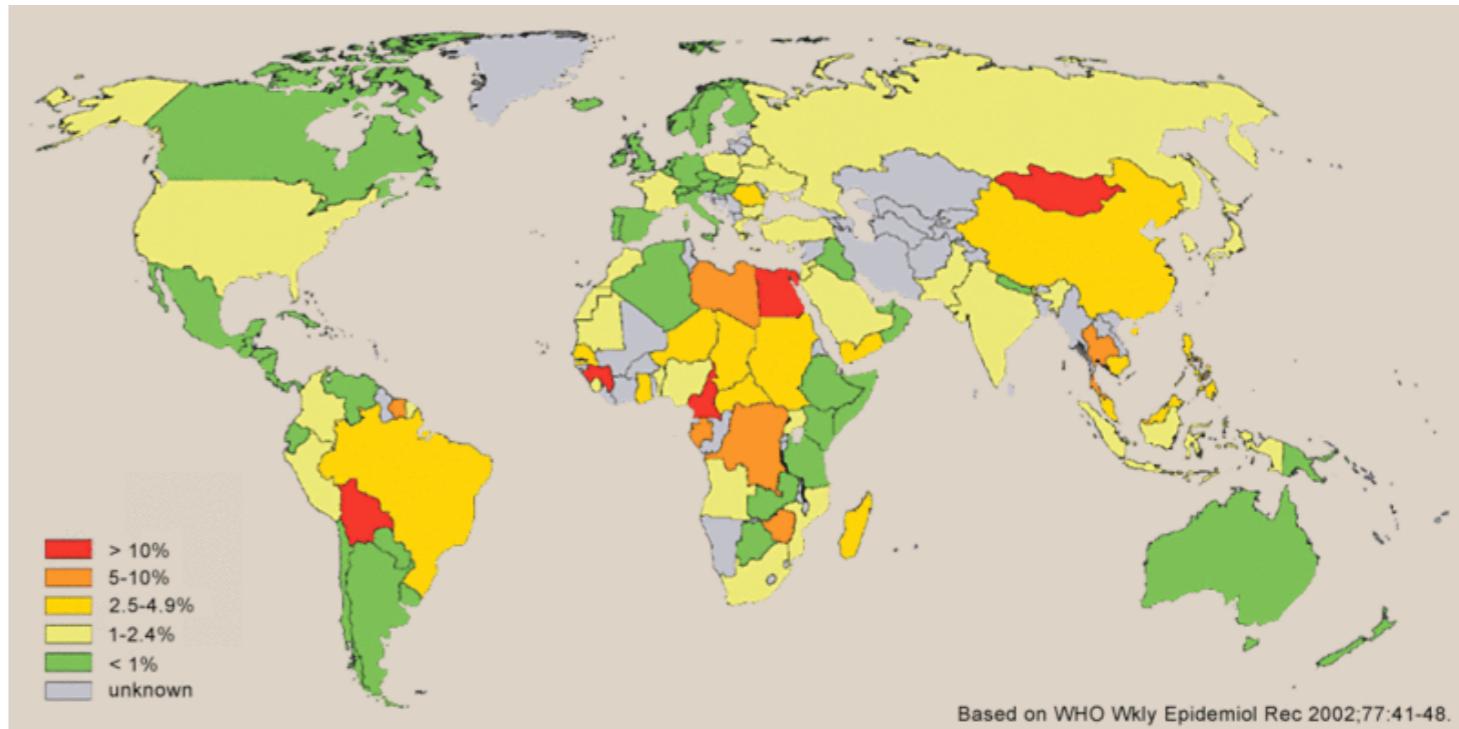
Assump-
tion

Piecewise constant
population size



HCV in
Egypt

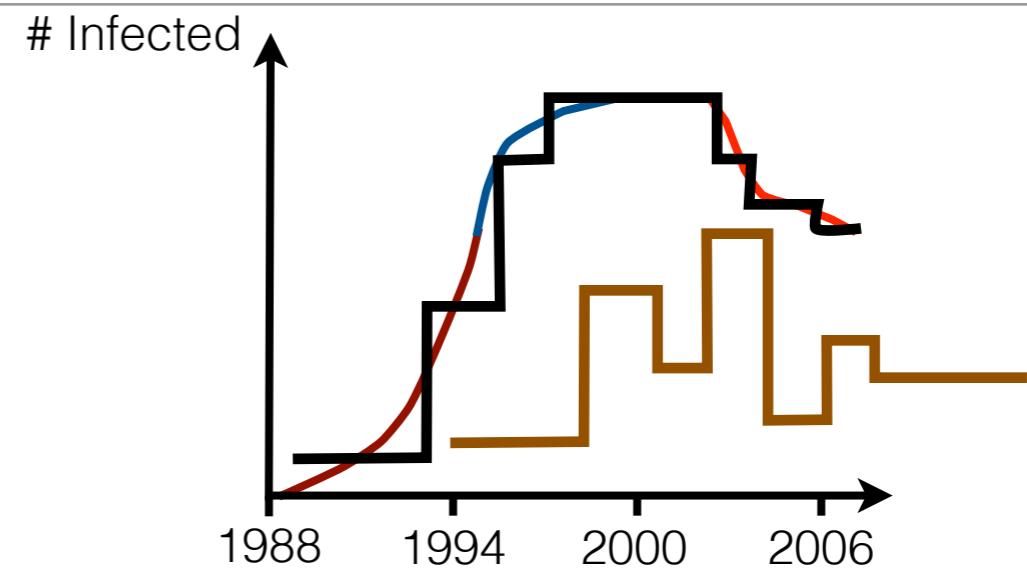
- ▶ Egypt highest HCV prevalence worldwide (about 13 %)
- ▶ Neighboring countries low prevalence
- ▶ When did the high prevalence emerge?
- ▶ Why?



Bayesian skyline plot relaxes constant growth rate

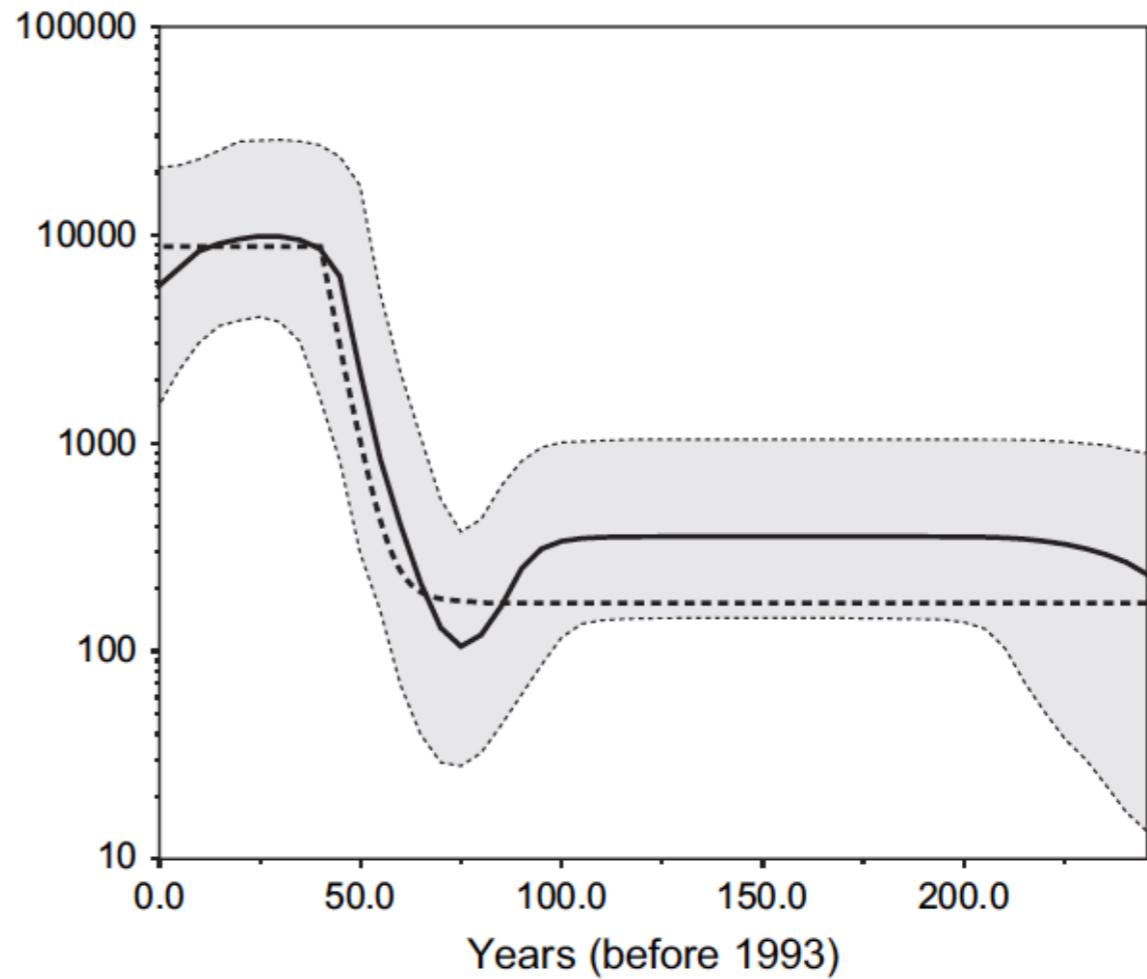
Assump-
tion

Piecewise constant
population size



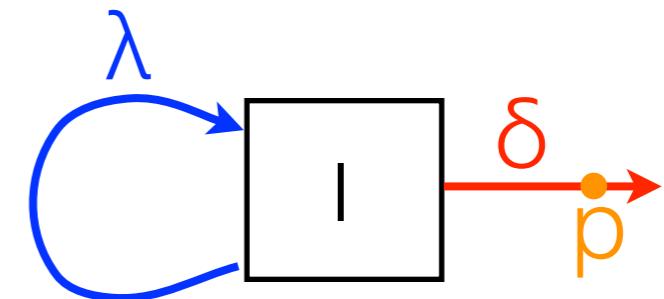
HCV in
Egypt

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



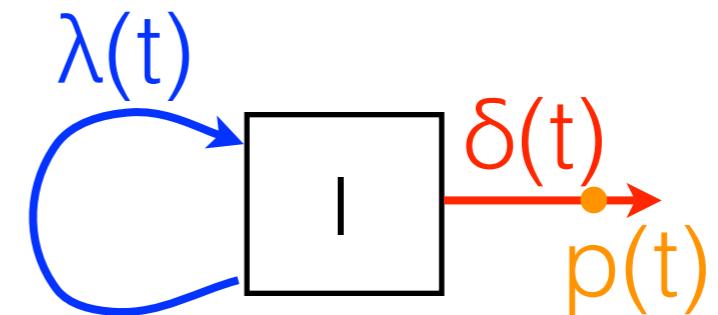
Birth-death-skyline plot

Epidemiological rates may change through time:



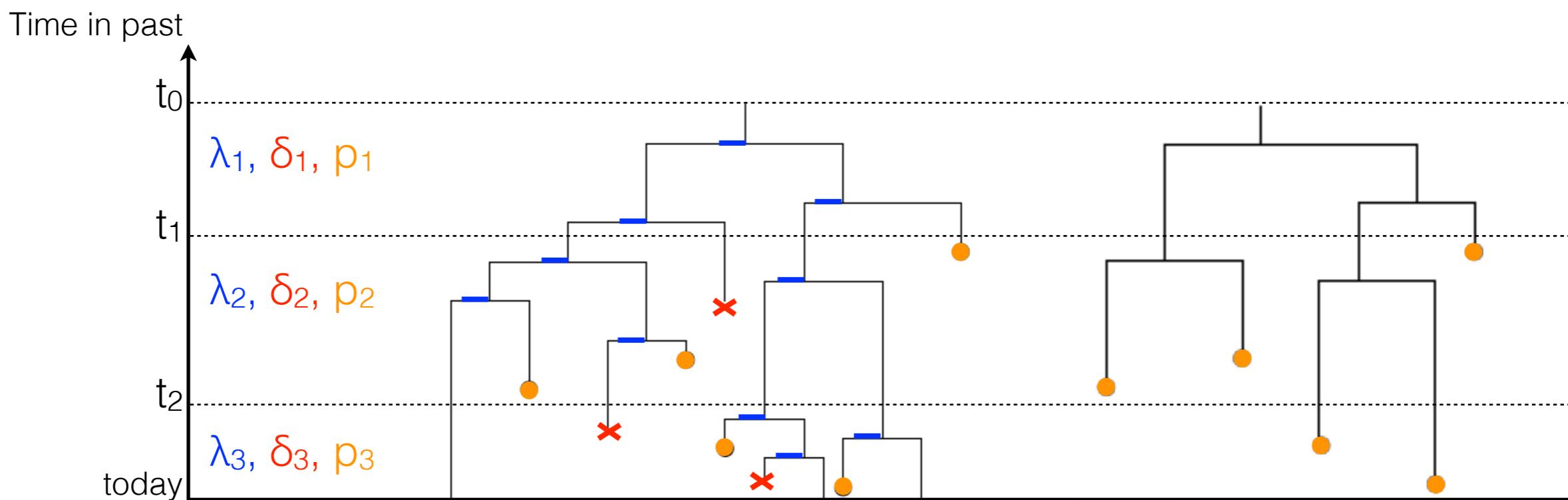
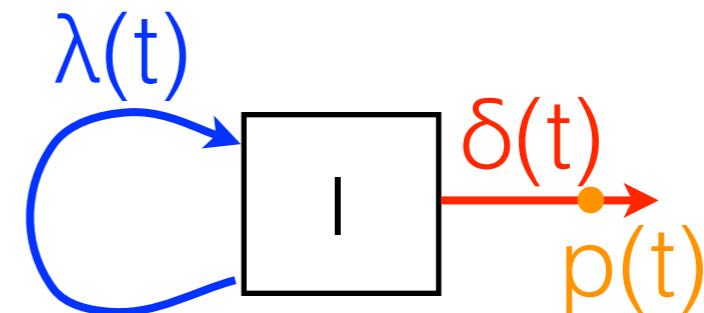
Birth-death-skyline plot

Epidemiological rates may change through time:



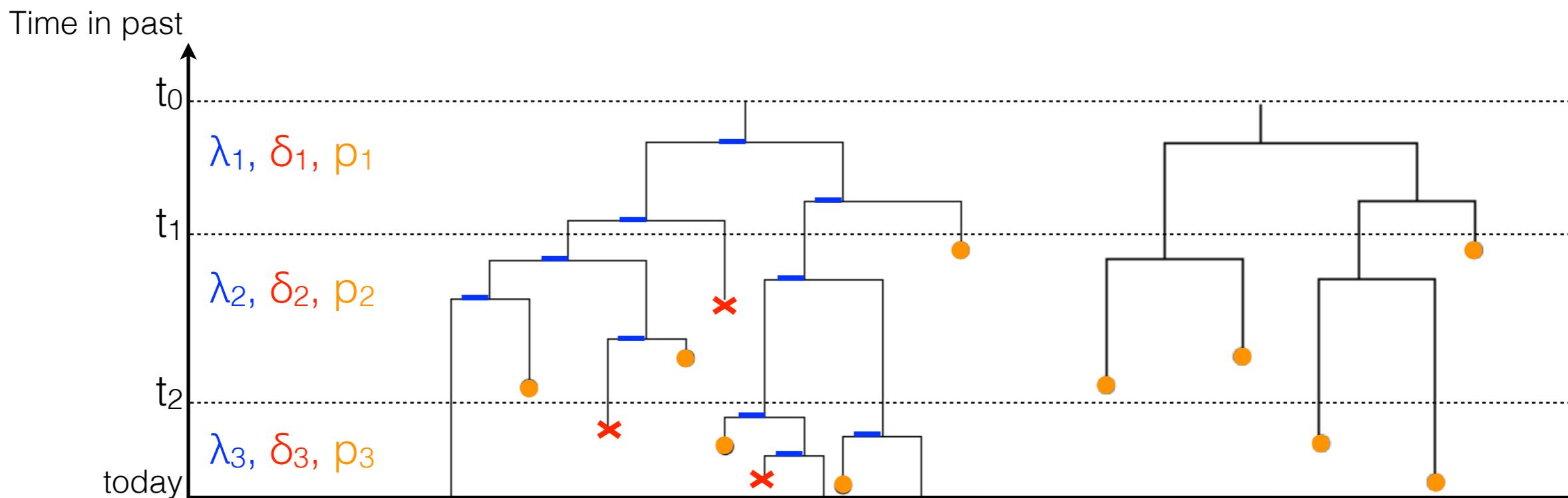
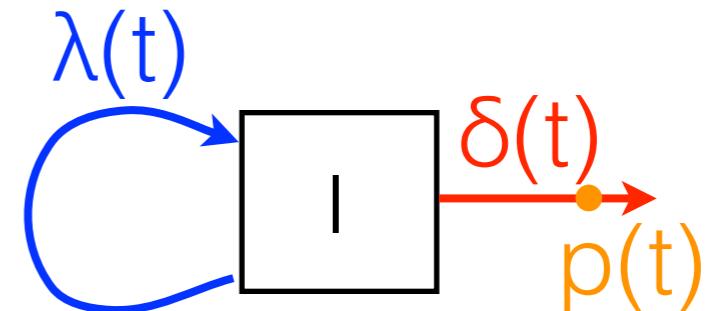
Birth-death-skyline plot

Epidemiological rates may change through time:



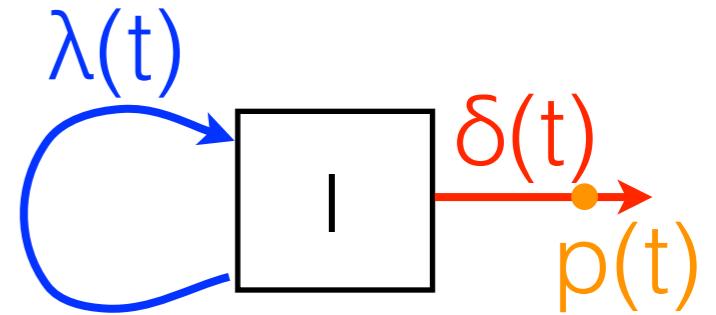
Birth-death-skyline plot

Epidemiological rates may change through time:



BEAST add-on BDSKY

Hepatitis C virus in Egypt

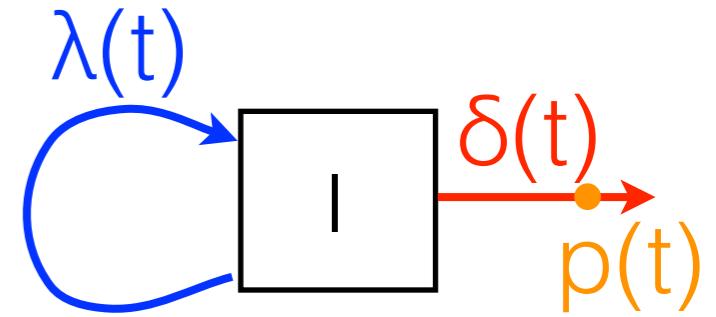


Effective reproductive number $R_e(t)$

generalizes

Basic reproductive number $R_0 = R_e(0)$

Hepatitis C virus in Egypt



Effective reproductive number $R_e(t)$
generalizes

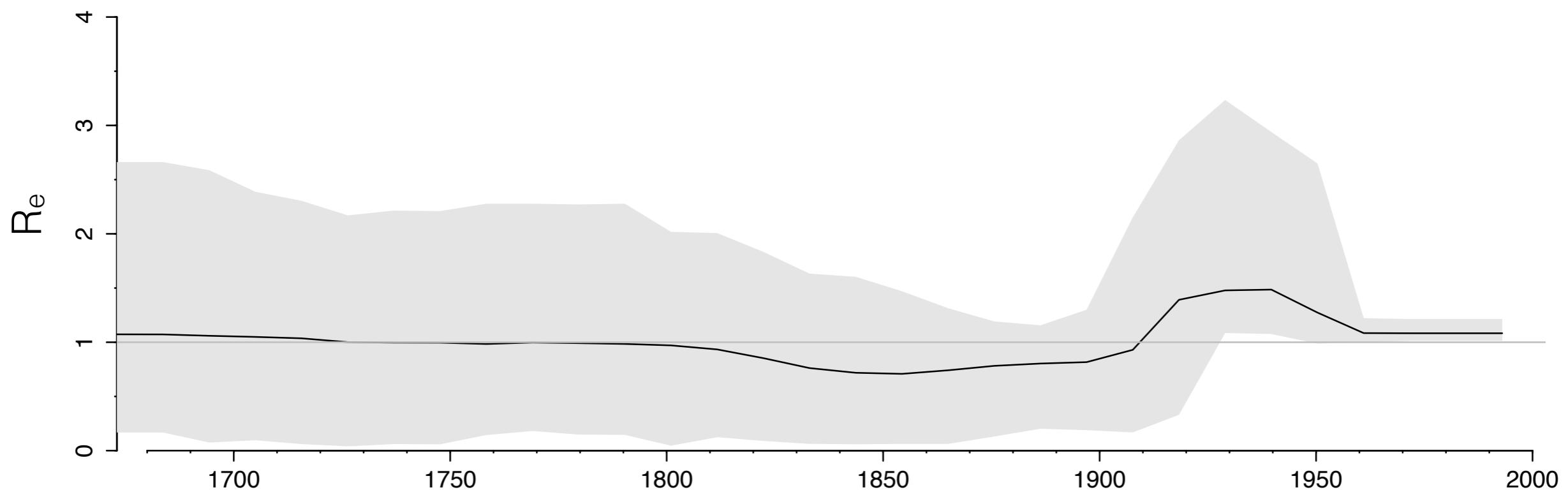
Basic reproductive number $R_0 = R_e(0)$

Analysis of 63 seq:

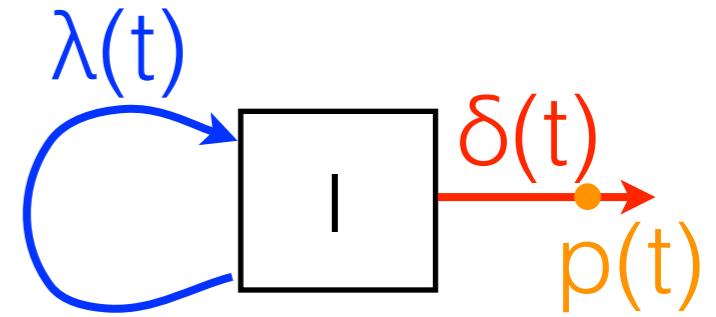
$$R_e(t) = \lambda(t)/\delta(t)$$

median

95% HPD interval



Hepatitis C virus in Egypt



Effective reproductive number $R_e(t)$
generalizes

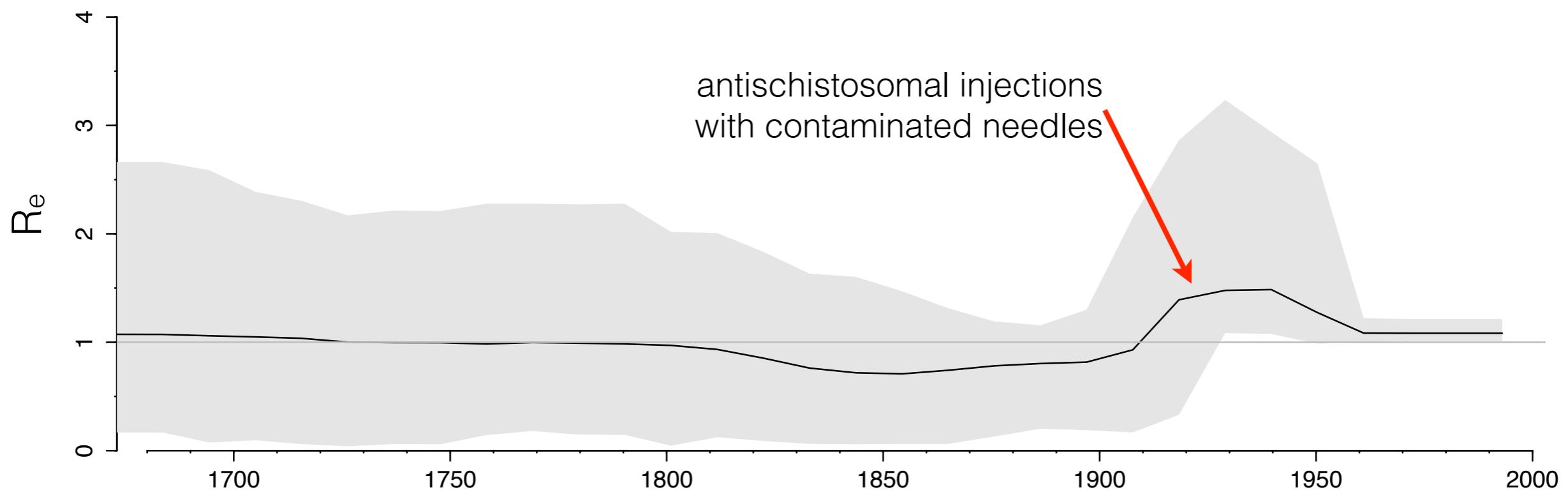
Basic reproductive number $R_0 = R_e(0)$

Analysis of 63 seq:

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

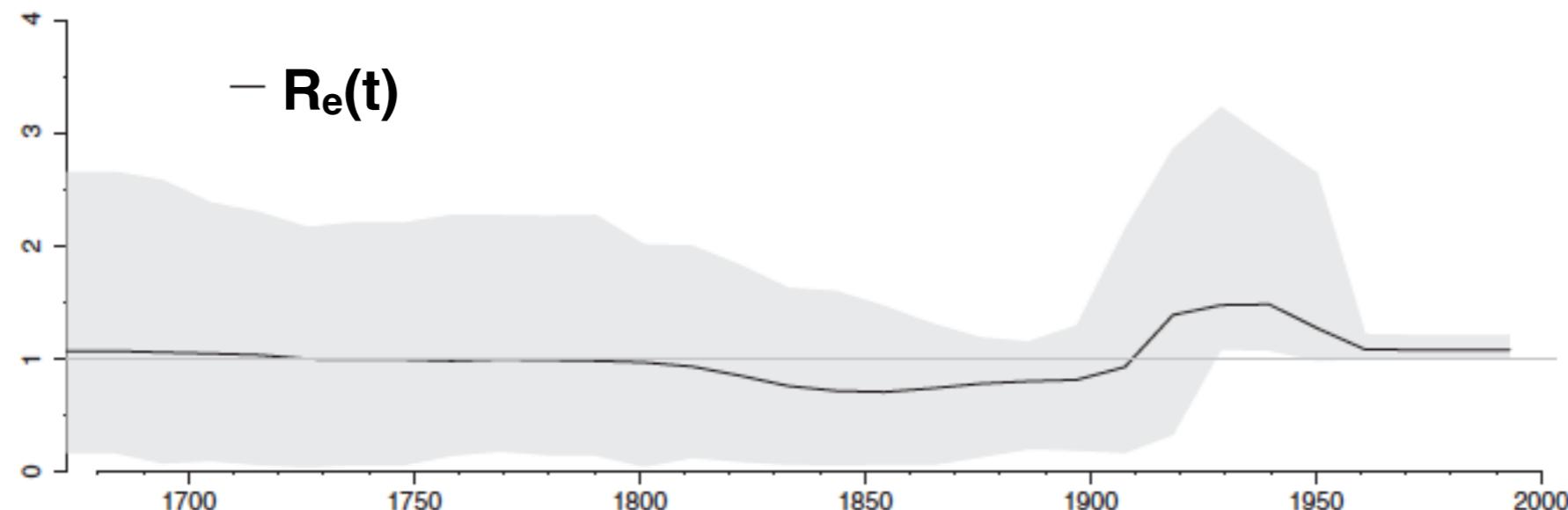
median

95% HPD interval



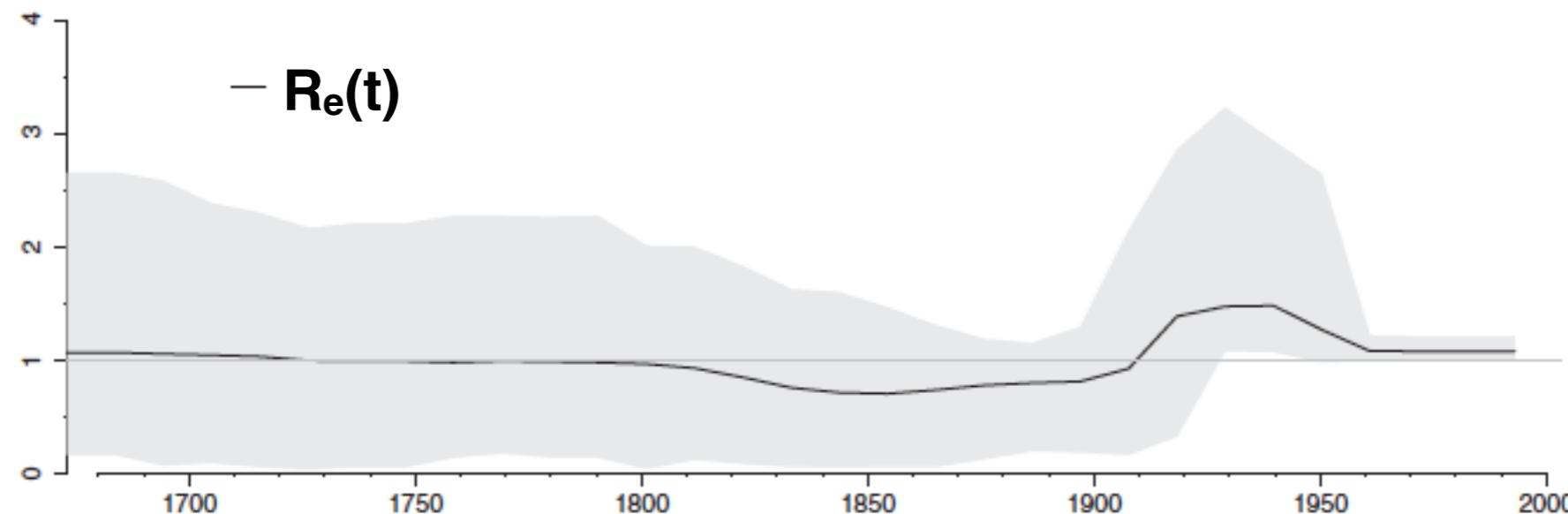
The two skyline models

Birth-death skyline plot: effective reproductive number

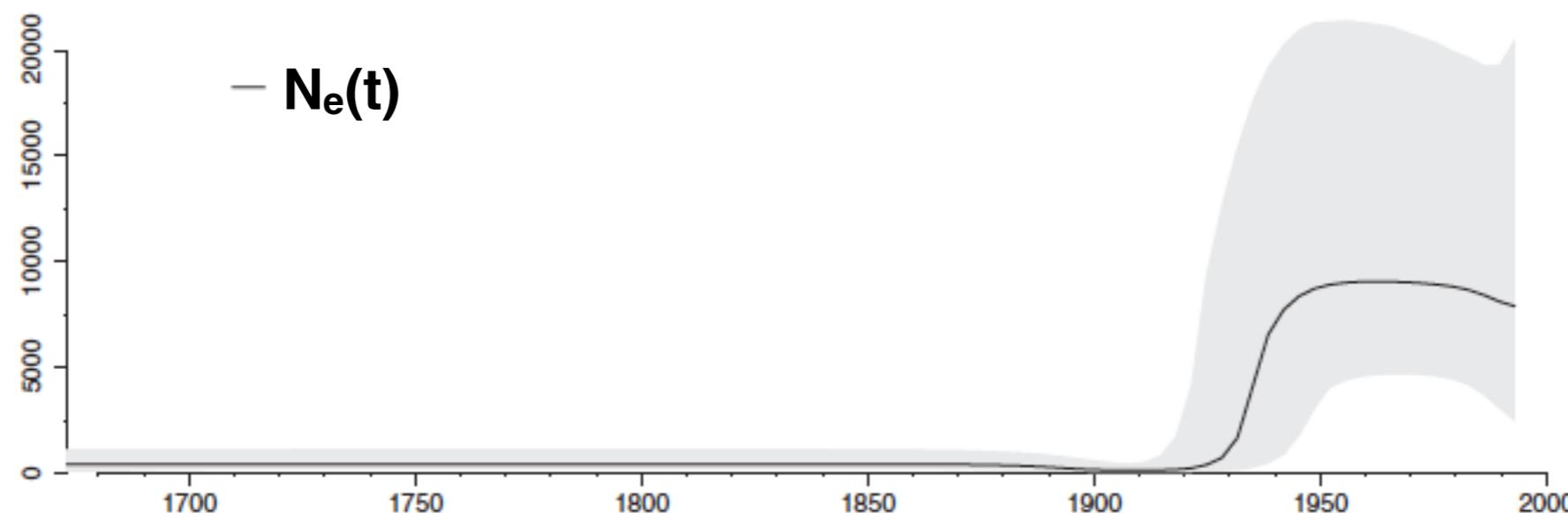


The two skyline models

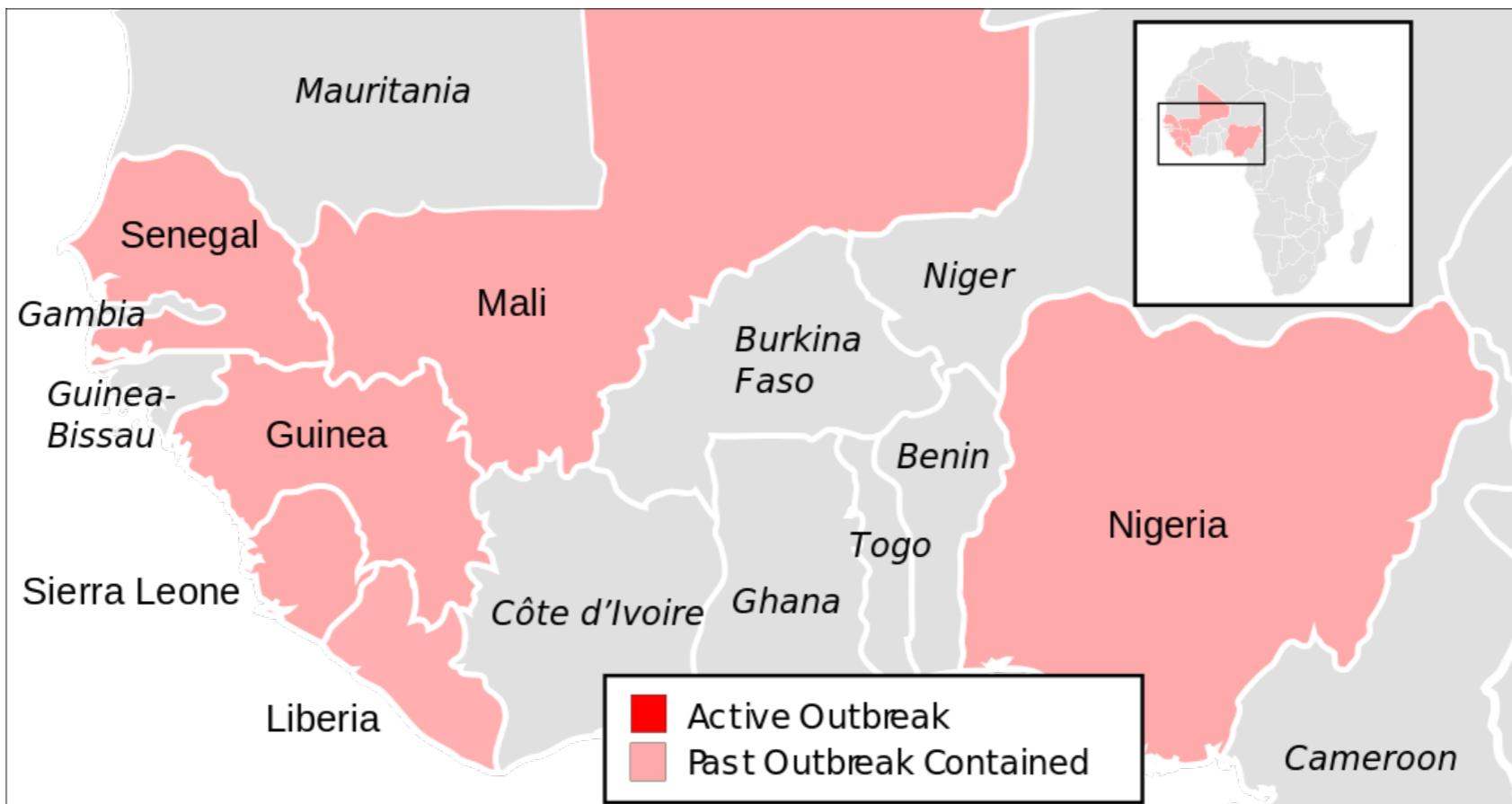
Birth-death skyline plot: effective reproductive number



Coalescent skyline plot: effective population size



Quantifying the epidemic spread of Ebola in West Africa



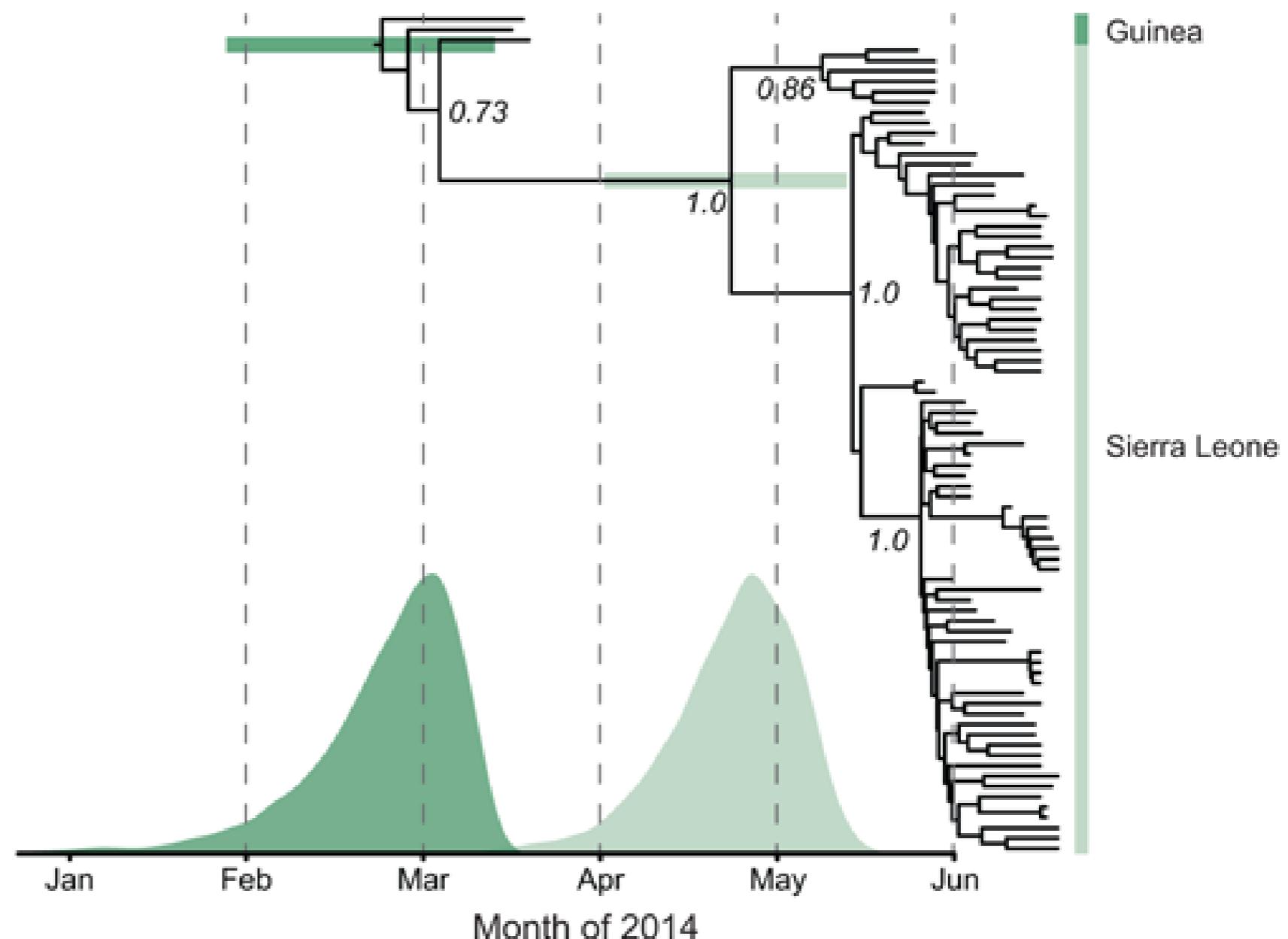
28657 suspected cases, 11325 deaths (probably underestimate)

Single zoonosis in December caused the whole epidemic

Scienceexpress

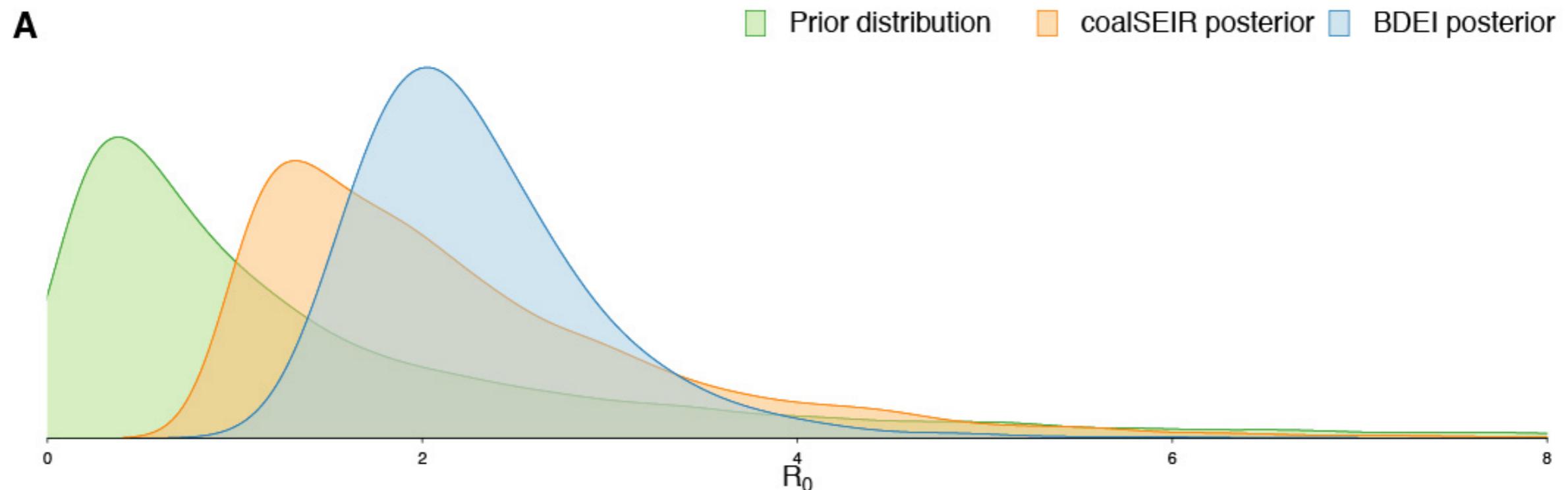
Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak

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



Quantification the epidemic spread in Sierra Leone (72 genomes)

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



Major challenge in data collection

Scienceexpress

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

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

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.

Major challenge in data collection

Scienceexpress

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

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

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.

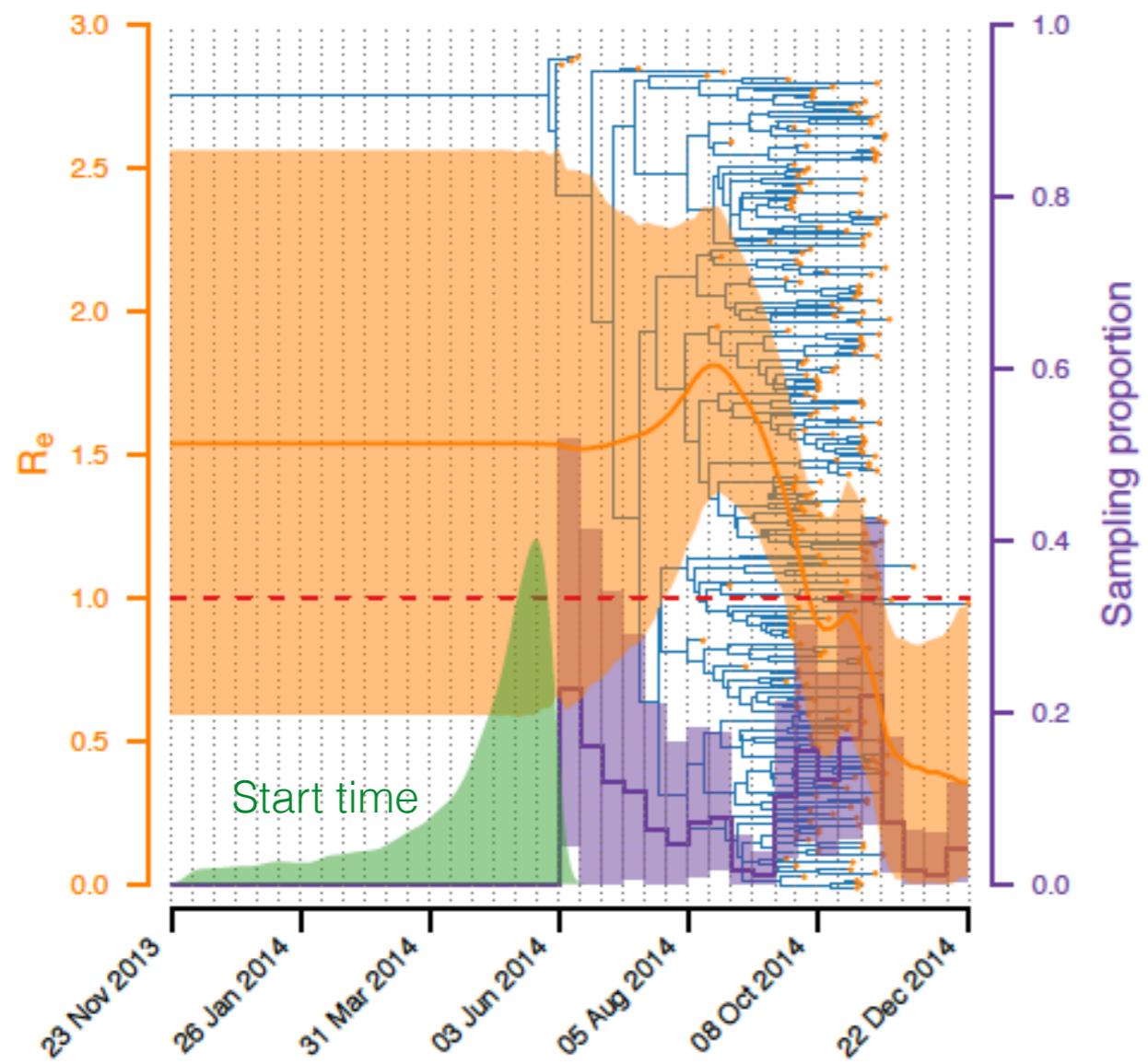
Eventually 1600 sequences became available
What do they tell about the epidemic spread?

Sierra Leone

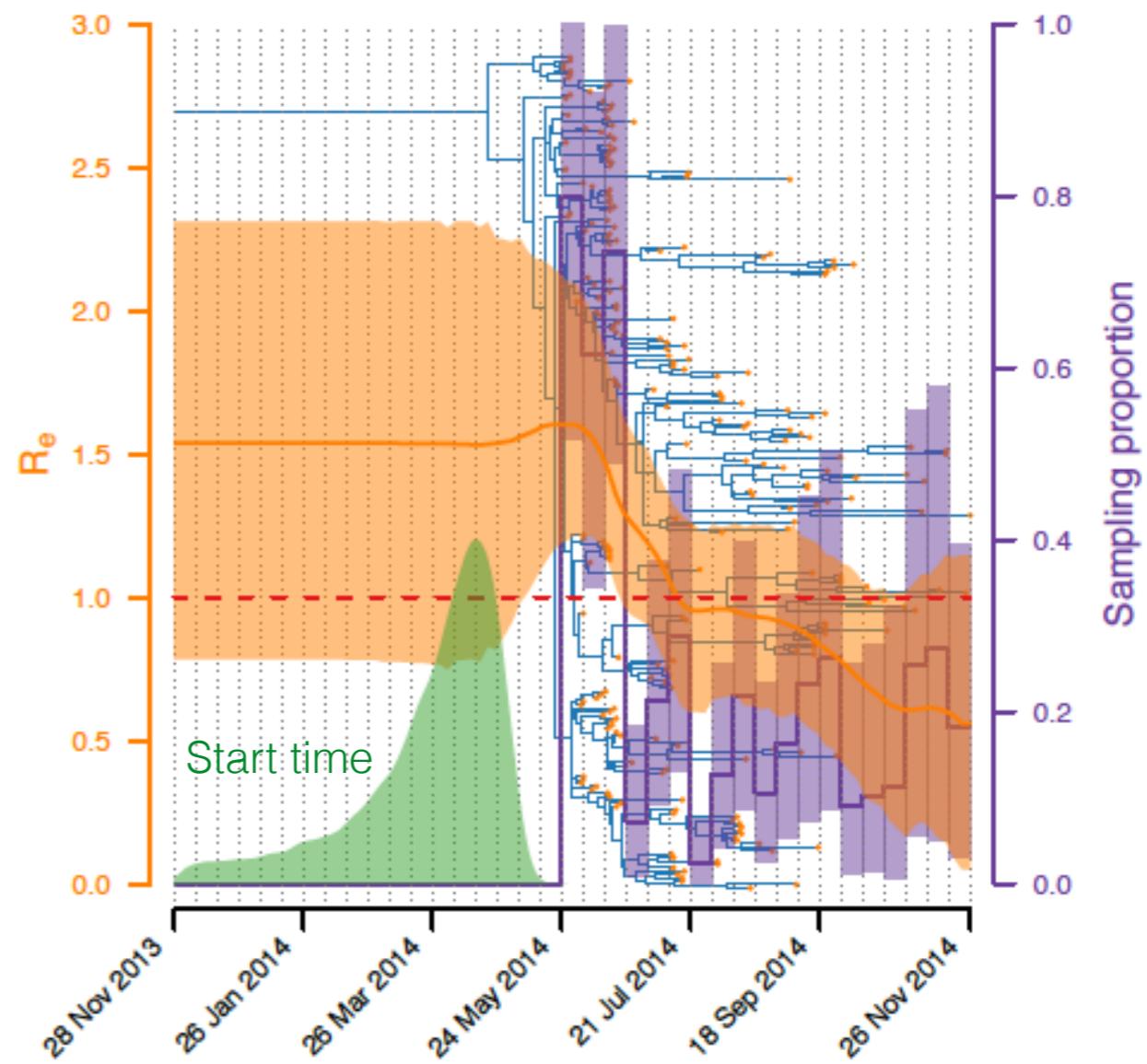
(Based on **507** genomes sampled between **25 May** and **22 Dec 2014**)



(C) Sierra Leone (Northwest)



(D) Sierra Leone (Southeast)

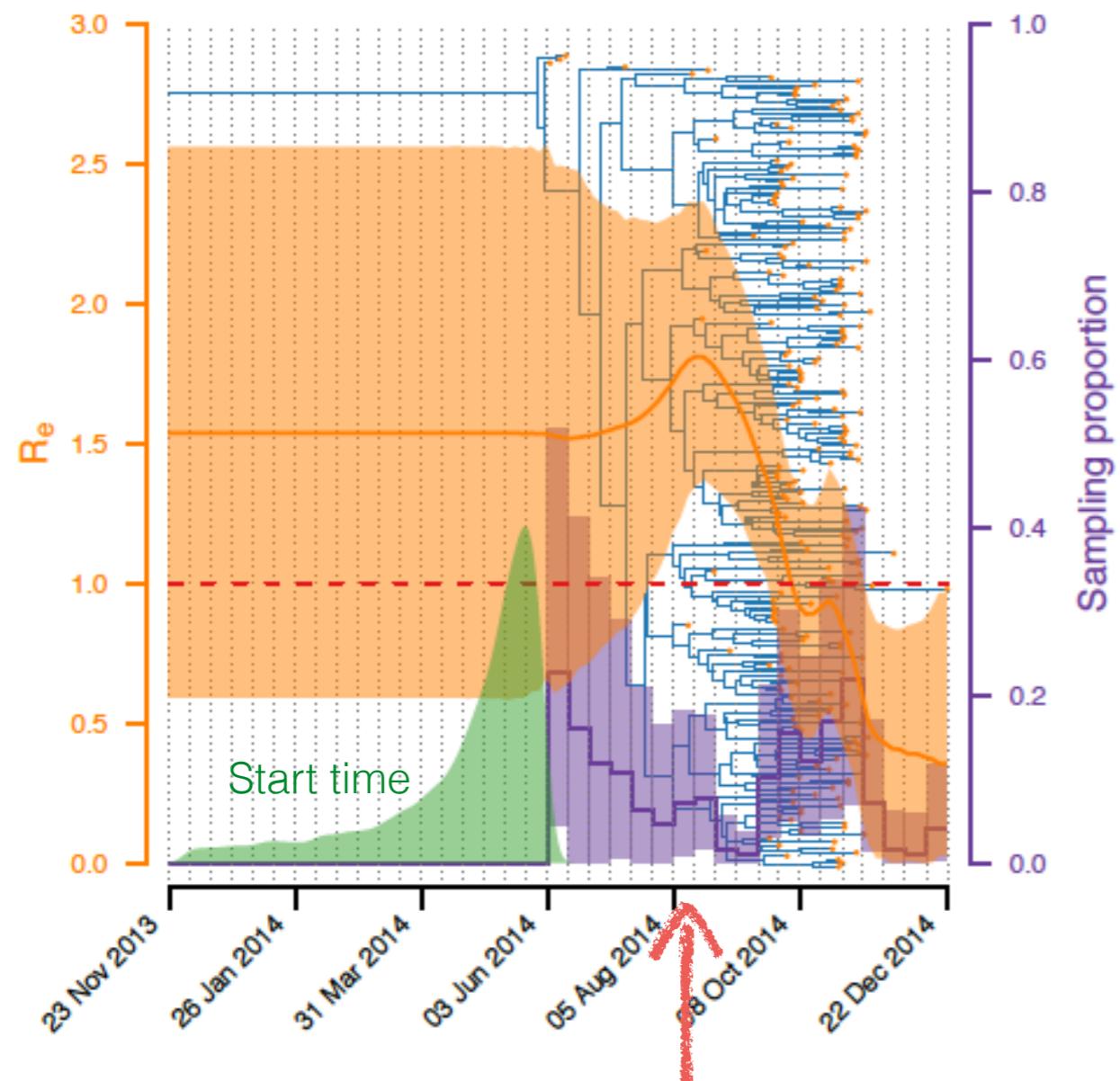


Sierra Leone

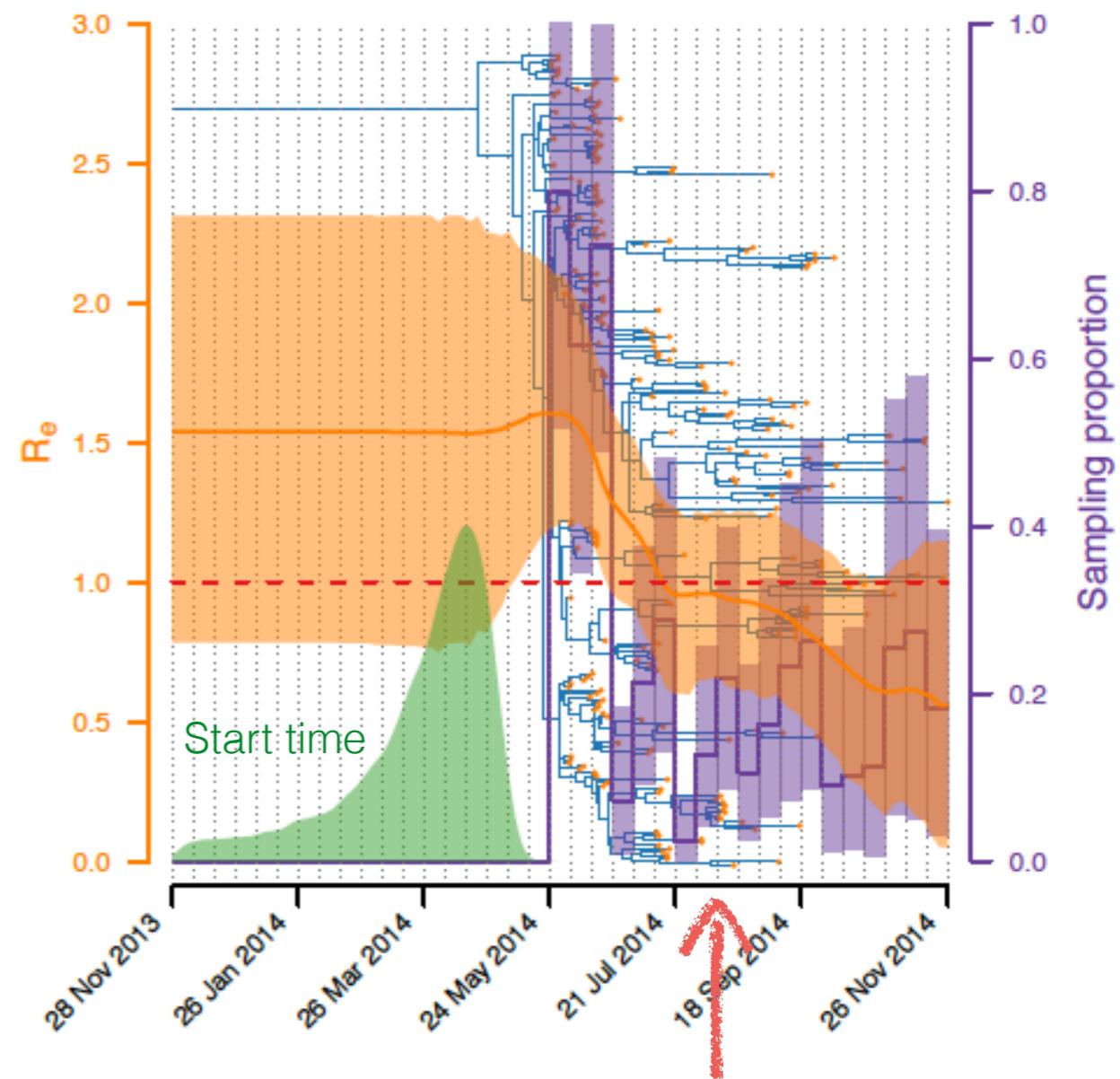
(Based on **507** genomes sampled between **25 May** and **22 Dec 2014**)



(C) Sierra Leone (Northwest)

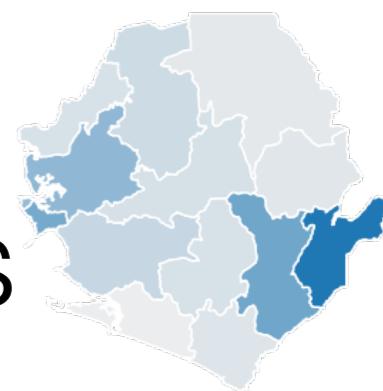


(D) Sierra Leone (Southeast)

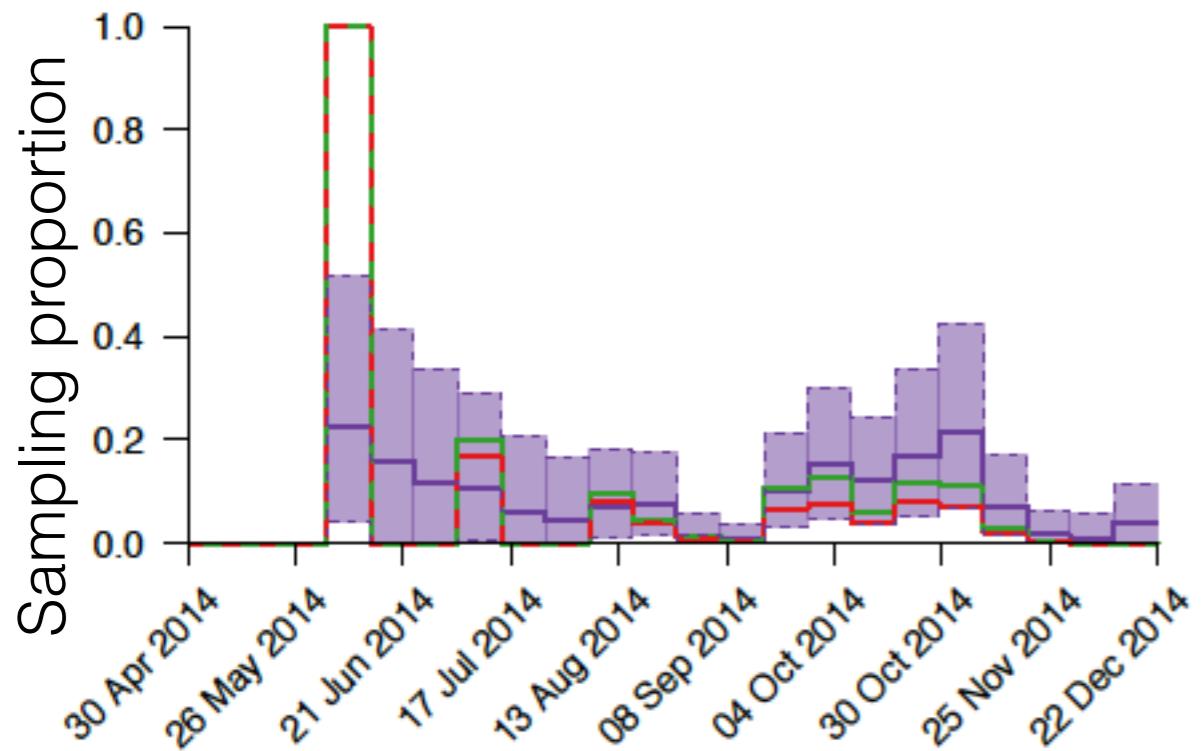


Global public health emergency declared

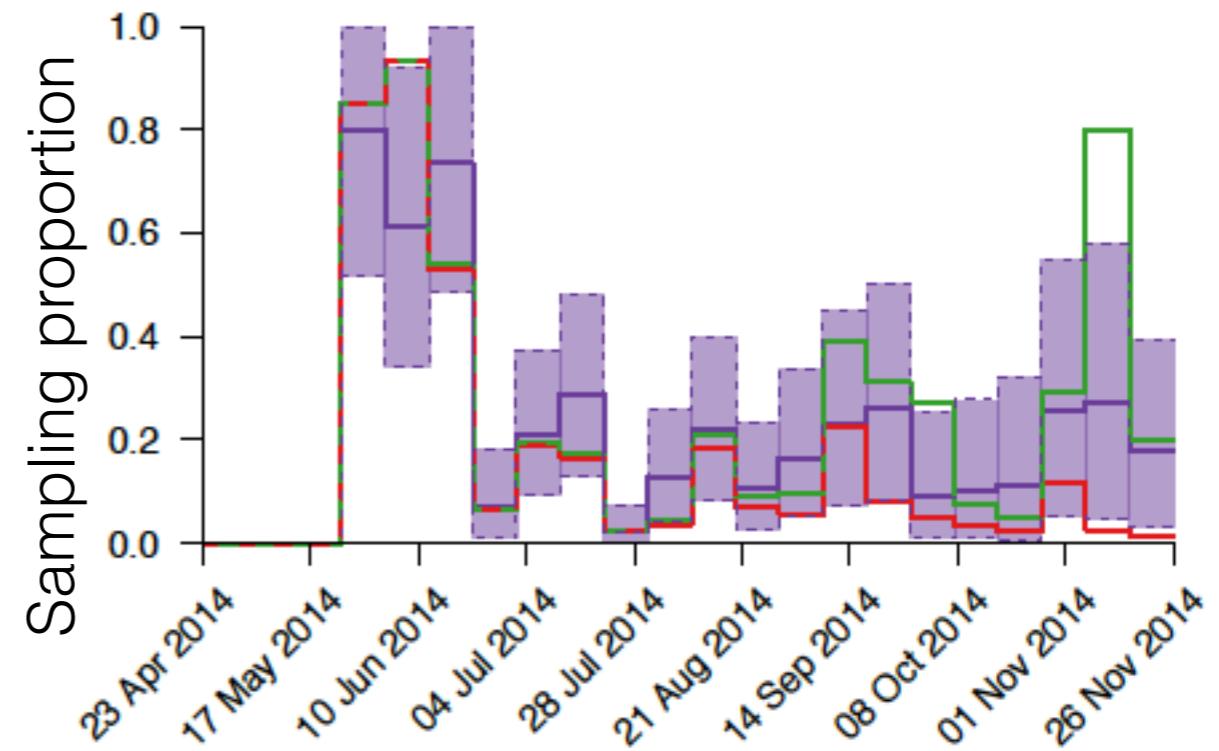
Estimating infected population sizes



(C) Sierra Leone (Northwest)



(D) Sierra Leone (Southeast)



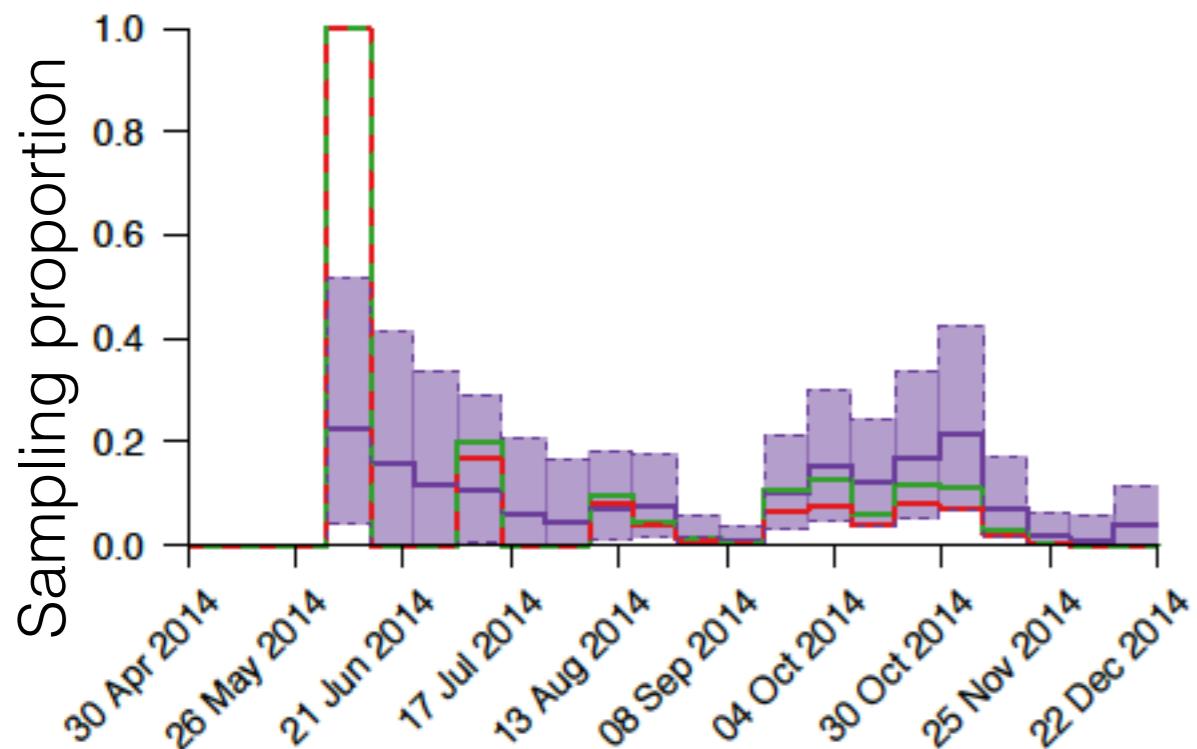
#Infected \approx #sequenced / (sampling proportion)

- Estimated sampling proportion
- Ratio (newly sequenced / new cases)
- Ratio (newly sequenced / new cases)
(highly sampled areas)

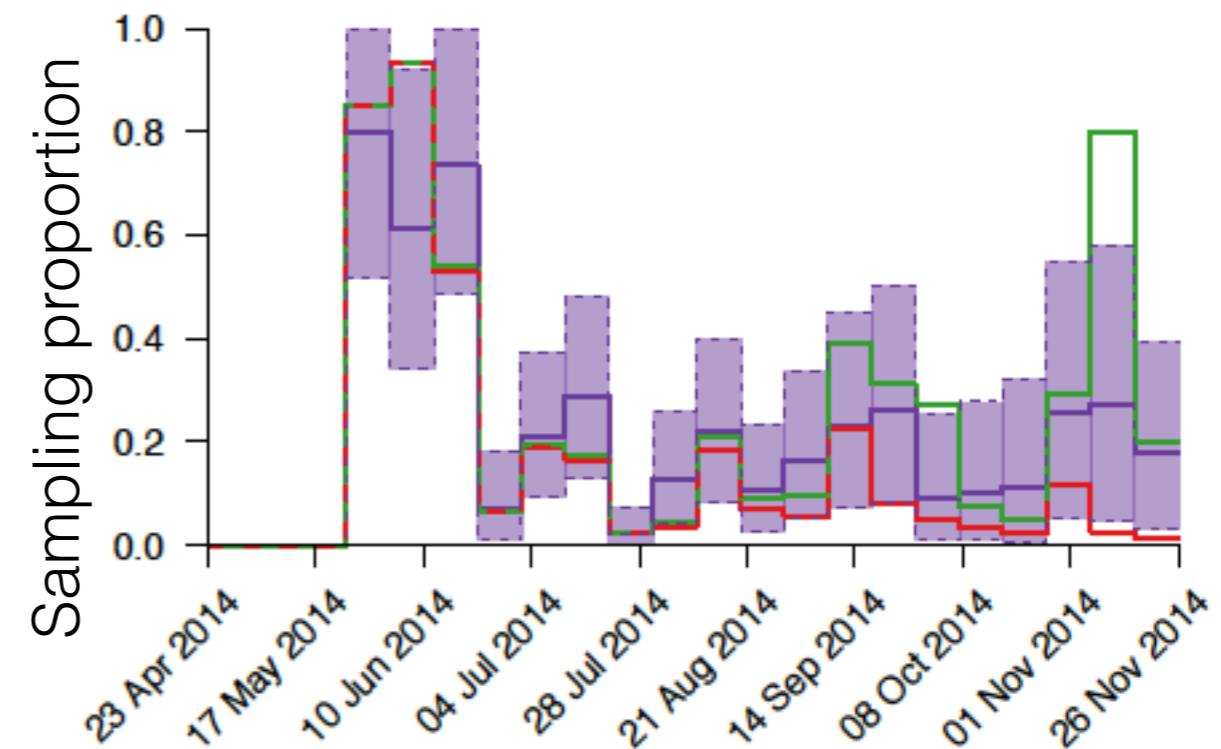
Estimating infected population sizes



(C) Sierra Leone (Northwest)



(D) Sierra Leone (Southeast)



— Estimated sampling proportion

Phylogenetic methods (purple) estimate well empirical sampling proportion (green) and thus infected pop. size!

Estimating infected population sizes directly

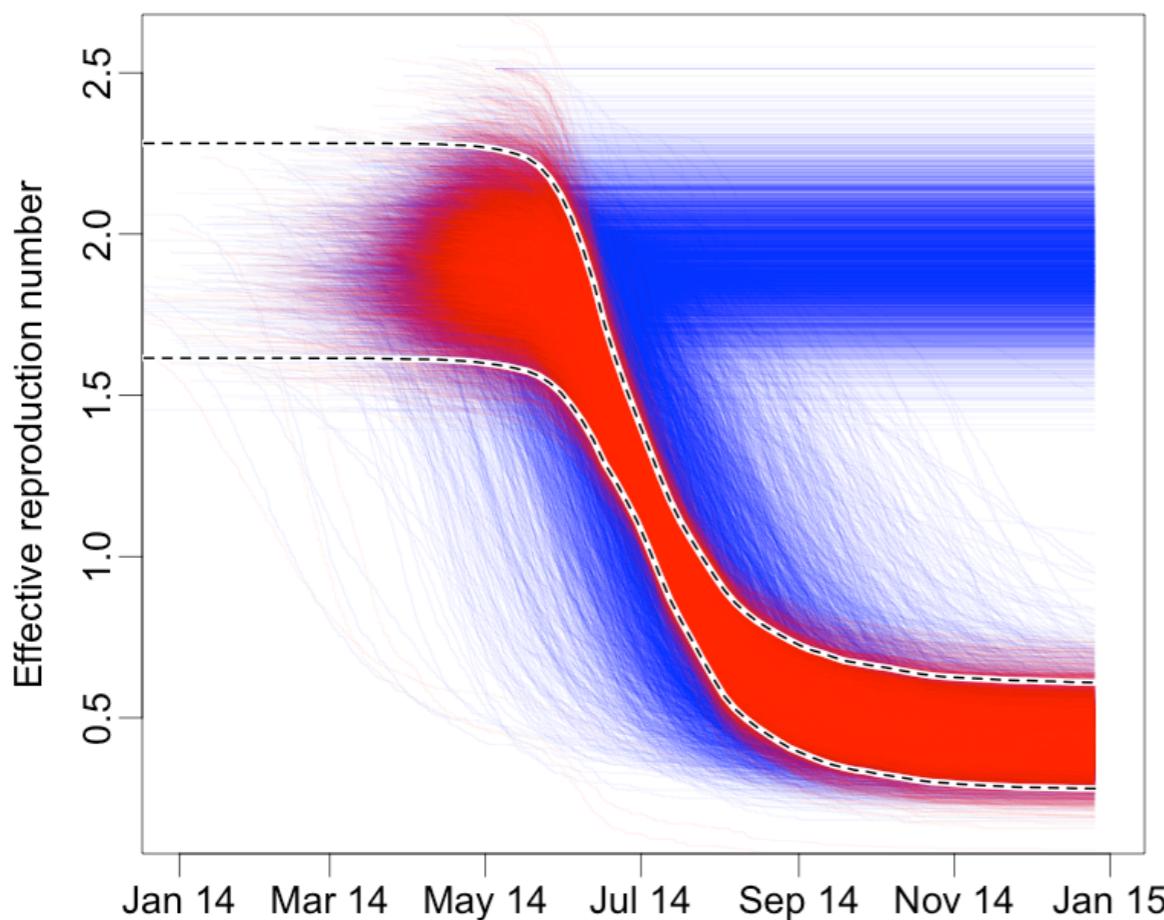
Estimating infected population sizes directly

BEAST add-on Epilnf

Estimating infected population sizes directly

Ebola in Kailahun, Sierra Leone

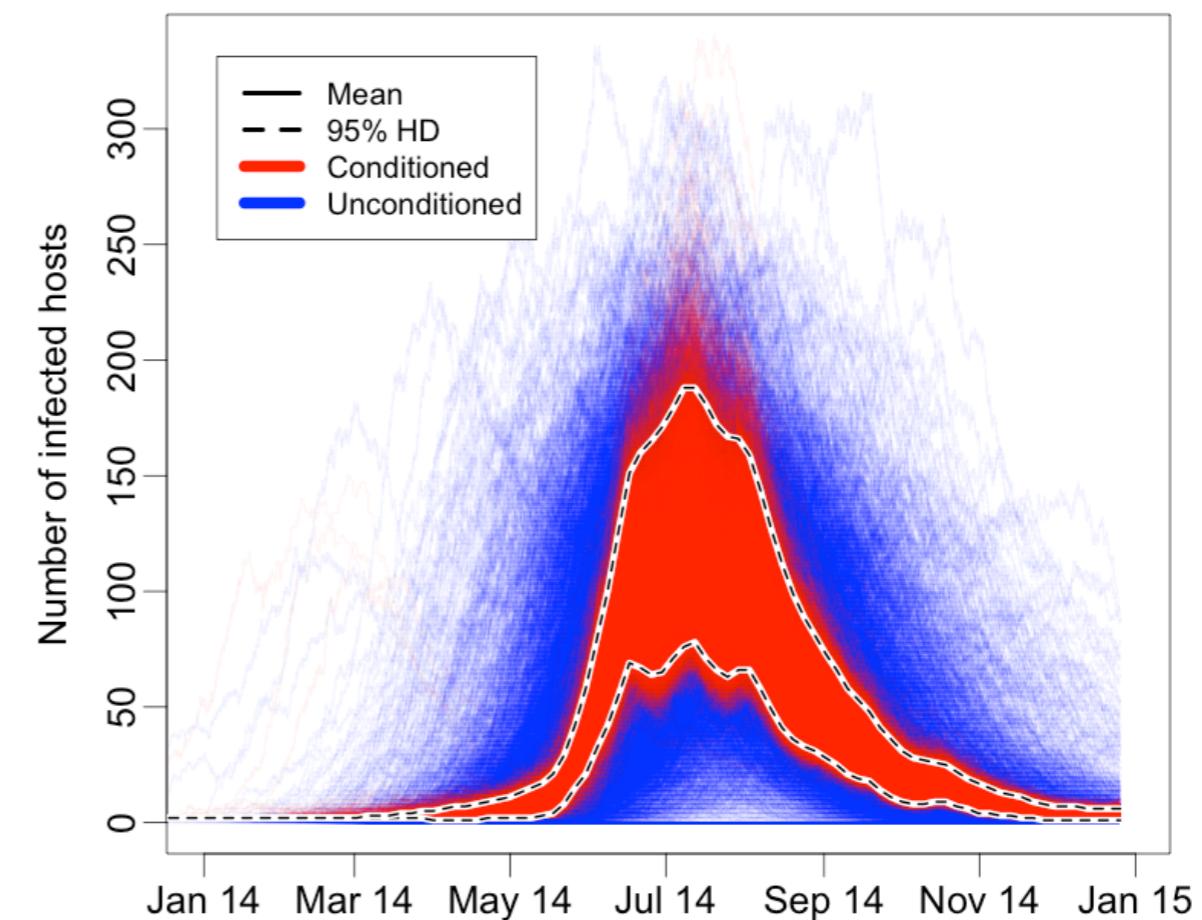
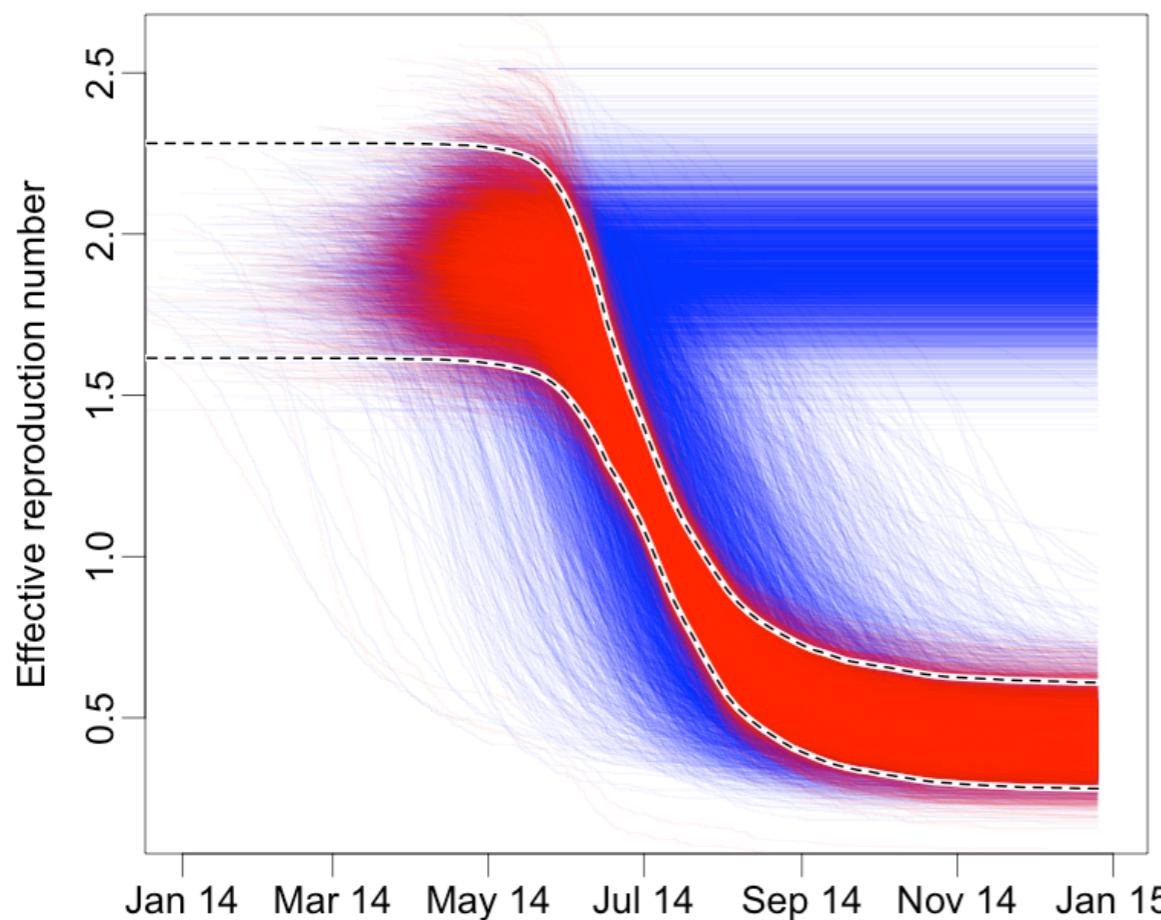
BEAST add-on Epilnf



Estimating infected population sizes directly

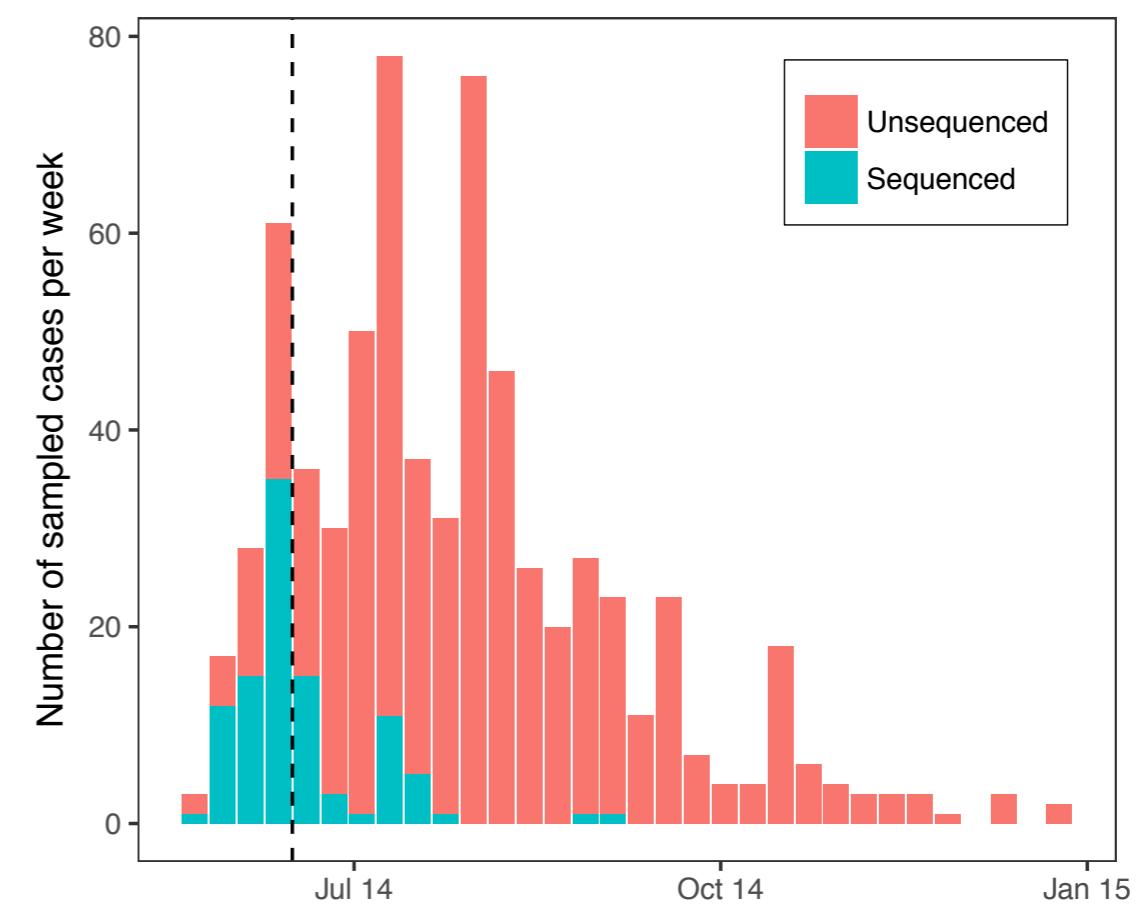
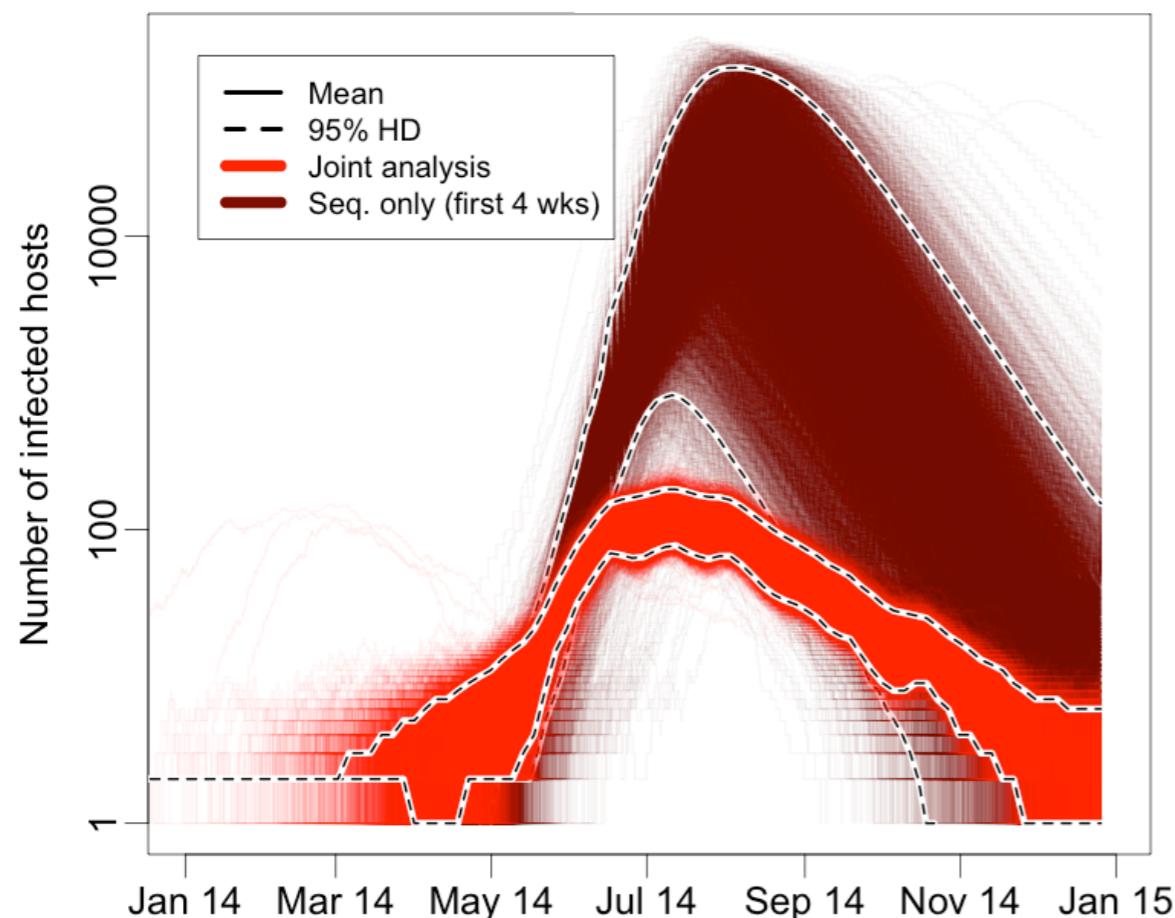
Ebola in Kailahun, Sierra Leone

BEAST add-on Epilnf



Analysis of case counts in addition to sequencing data

Ebola in Kailahun, Sierra Leone



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 (skylines)

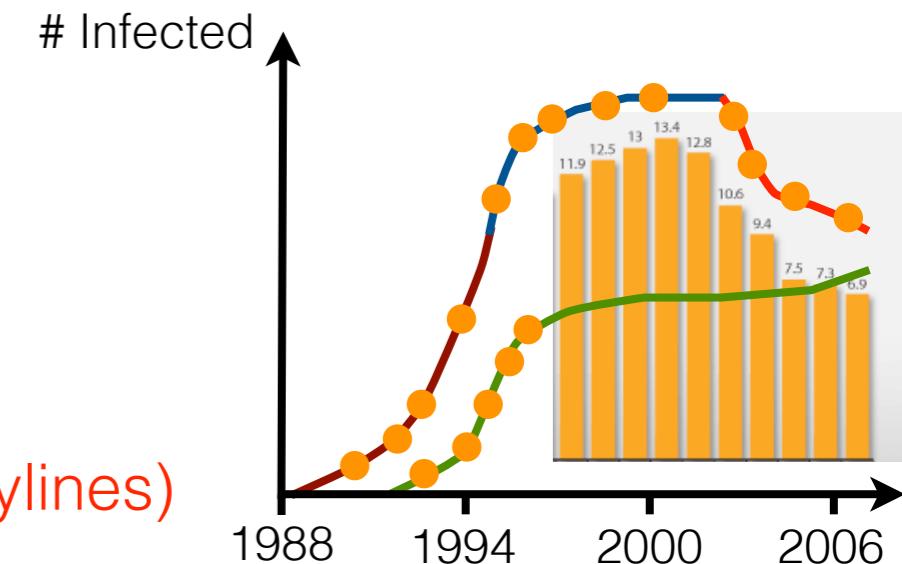
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

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), , Vaughan et al. (submitted)

4. ...accounting for population structure

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



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 (skylines)

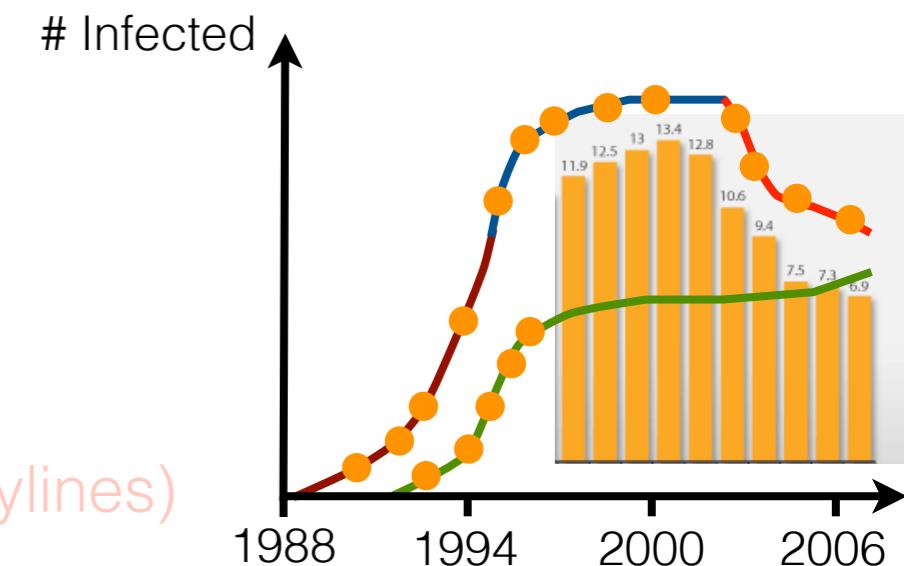
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

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), , Vaughan et al. (submitted)

4. ...accounting for population structure

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



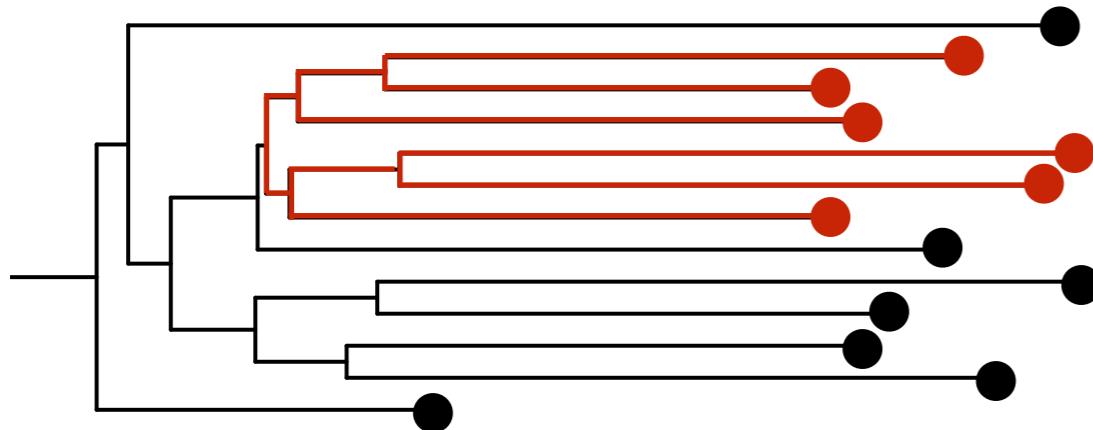
Pathogen population structure leaves pattern in the phylogenetic tree

Is drug resistance transmitted or evolves *de novo*?

Pathogen population structure leaves pattern in the phylogenetic tree

Is drug resistance transmitted or evolves *de novo*?

Transmitted drug resistance:

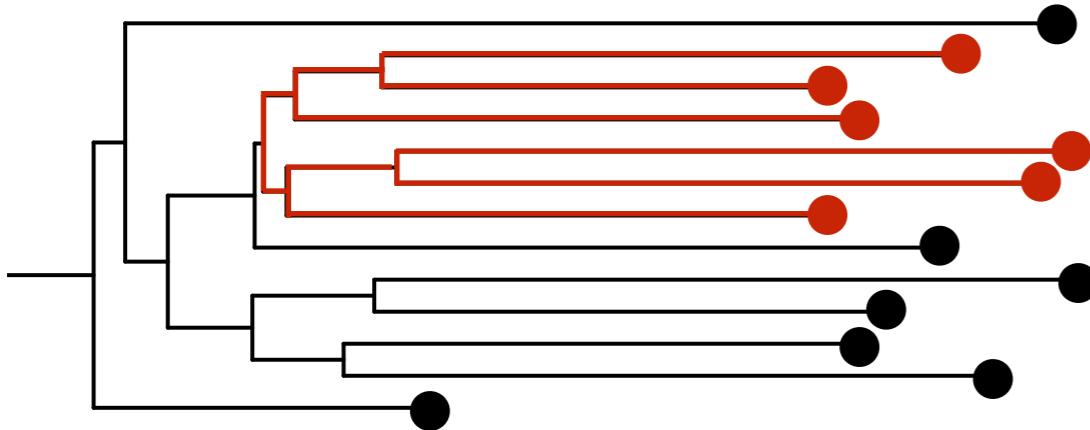


- drug resistant
- drug sensitive

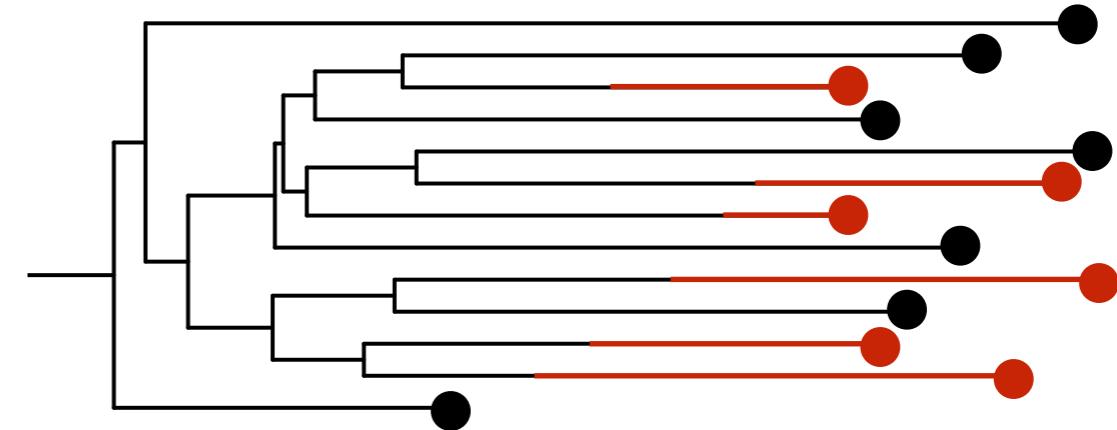
Pathogen population structure leaves pattern in the phylogenetic tree

Is drug resistance transmitted or evolves *de novo*?

Transmitted drug resistance:



De novo drug resistance:

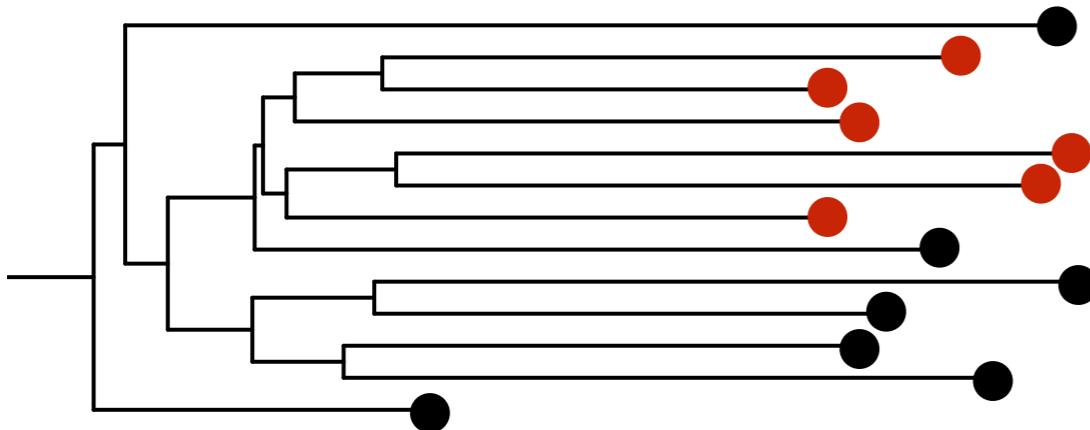


- drug resistant
- drug sensitive

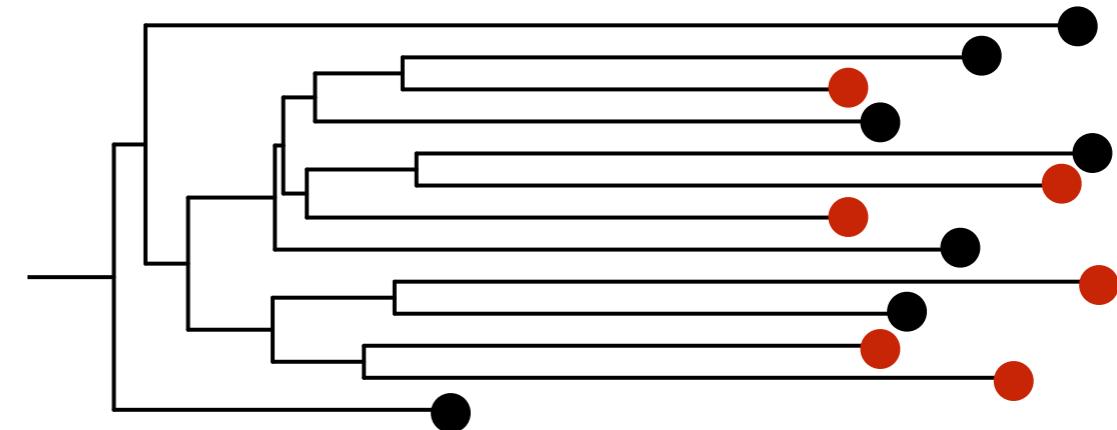
Pathogen population structure leaves pattern in the phylogenetic tree

Is drug resistance transmitted or evolves *de novo*?

Transmitted drug resistance:



De novo drug resistance:

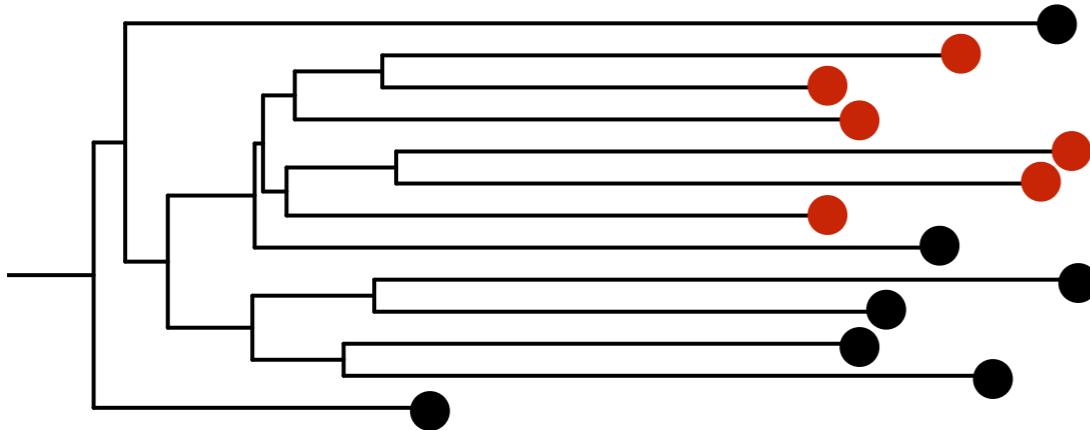


- drug resistant
- drug sensitive

Pathogen population structure leaves pattern in the phylogenetic tree

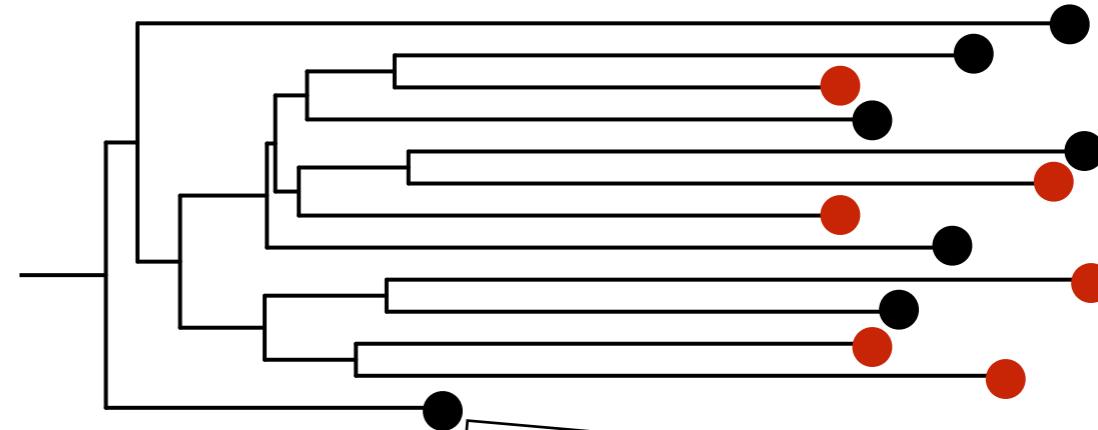
Is drug resistance transmitted or evolves *de novo*?

Transmitted drug resistance:



- drug resistant
- drug sensitive

De novo drug resistance:



Quantifying
selection

Host population structure leaves pattern in the phylogenetic tree

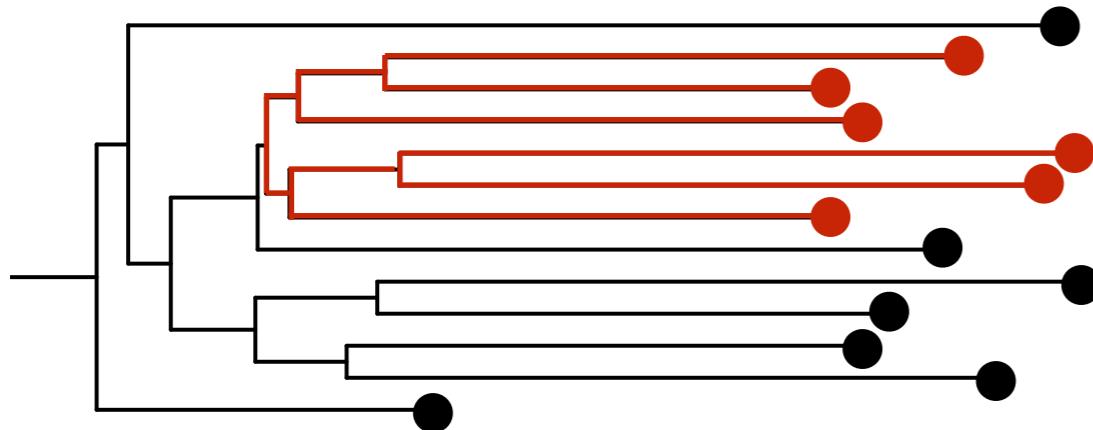
Are new infections explained through ongoing transmission or migration from another area?

-> identifying source demes!

Host population structure leaves pattern in the phylogenetic tree

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

Transmission in red deme:

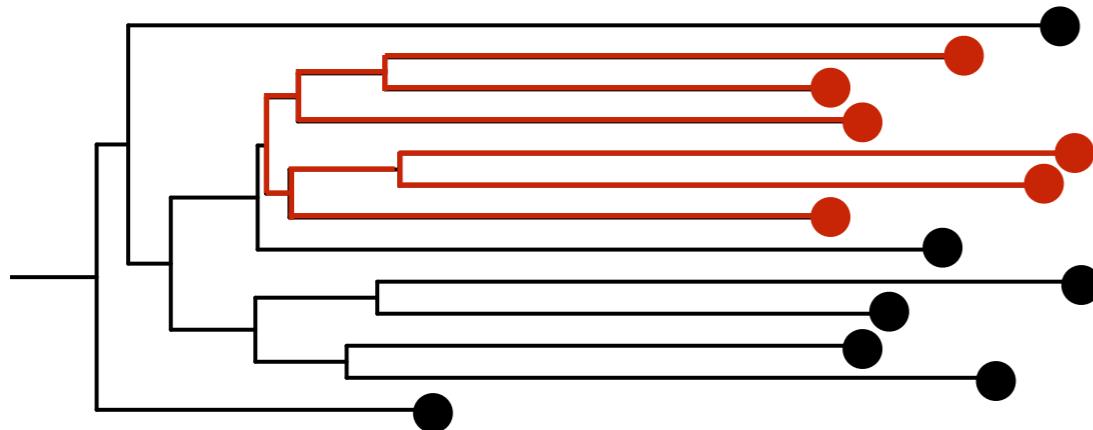


- Red deme
- Black deme

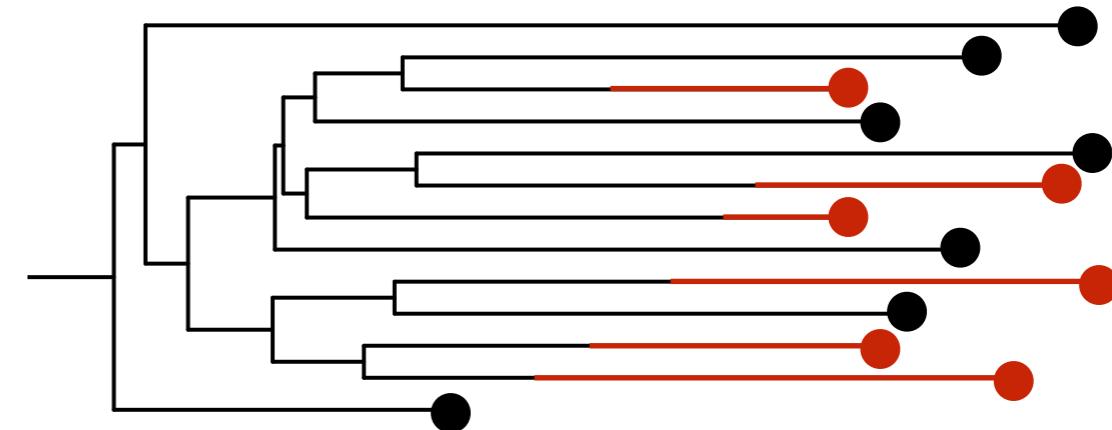
Host population structure leaves pattern in the phylogenetic tree

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

Transmission in red deme:



Migration into red deme:

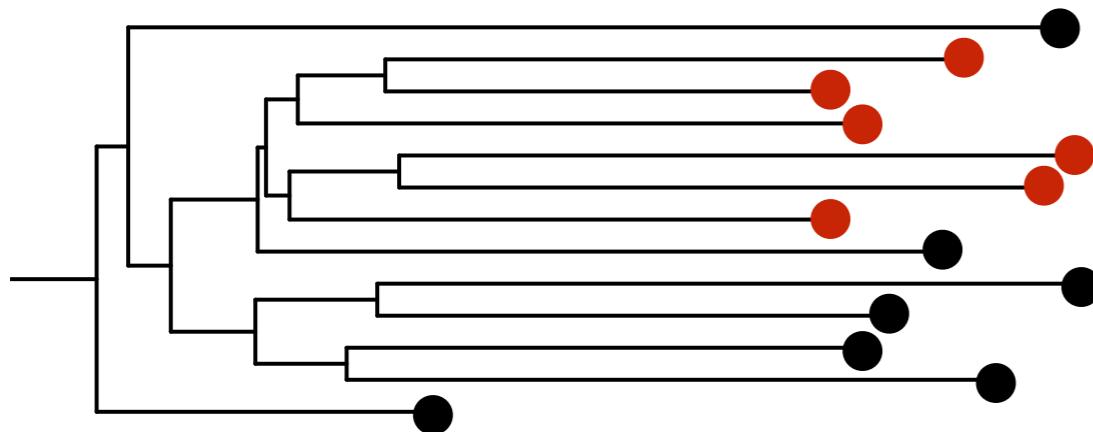


- Red deme
- Black deme

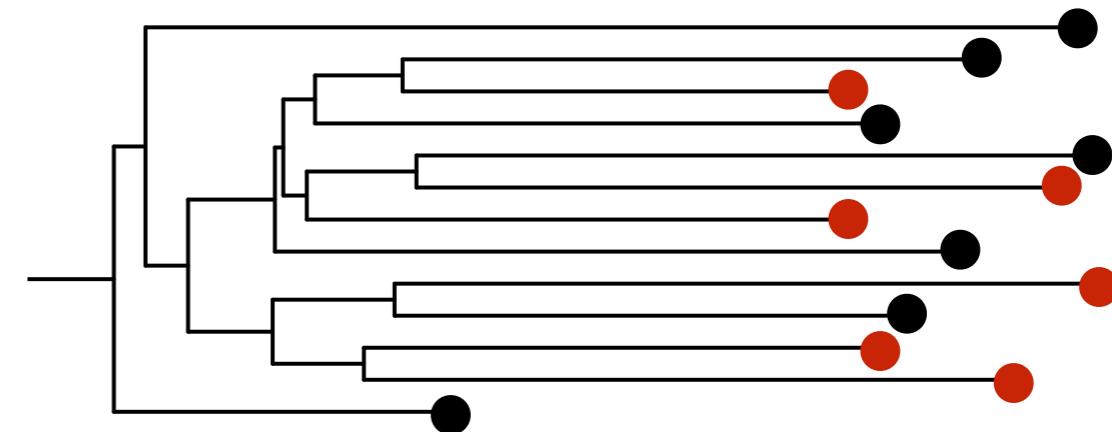
Host population structure leaves pattern in the phylogenetic tree

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

Transmission in red deme:

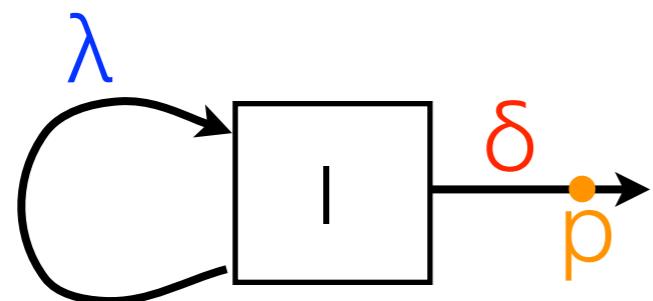


Migration into red deme:

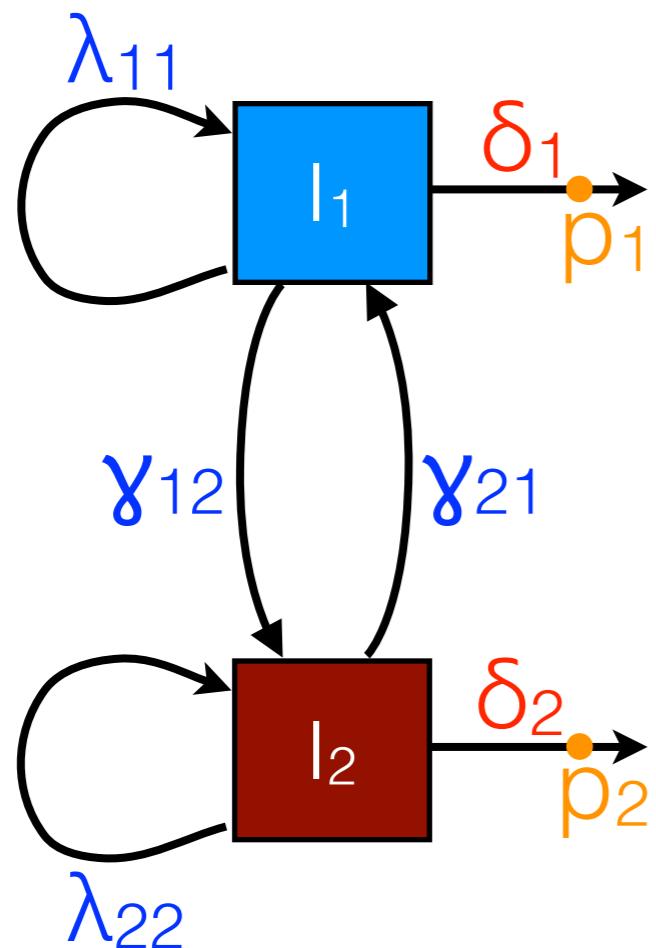


- Red deme
- Black deme

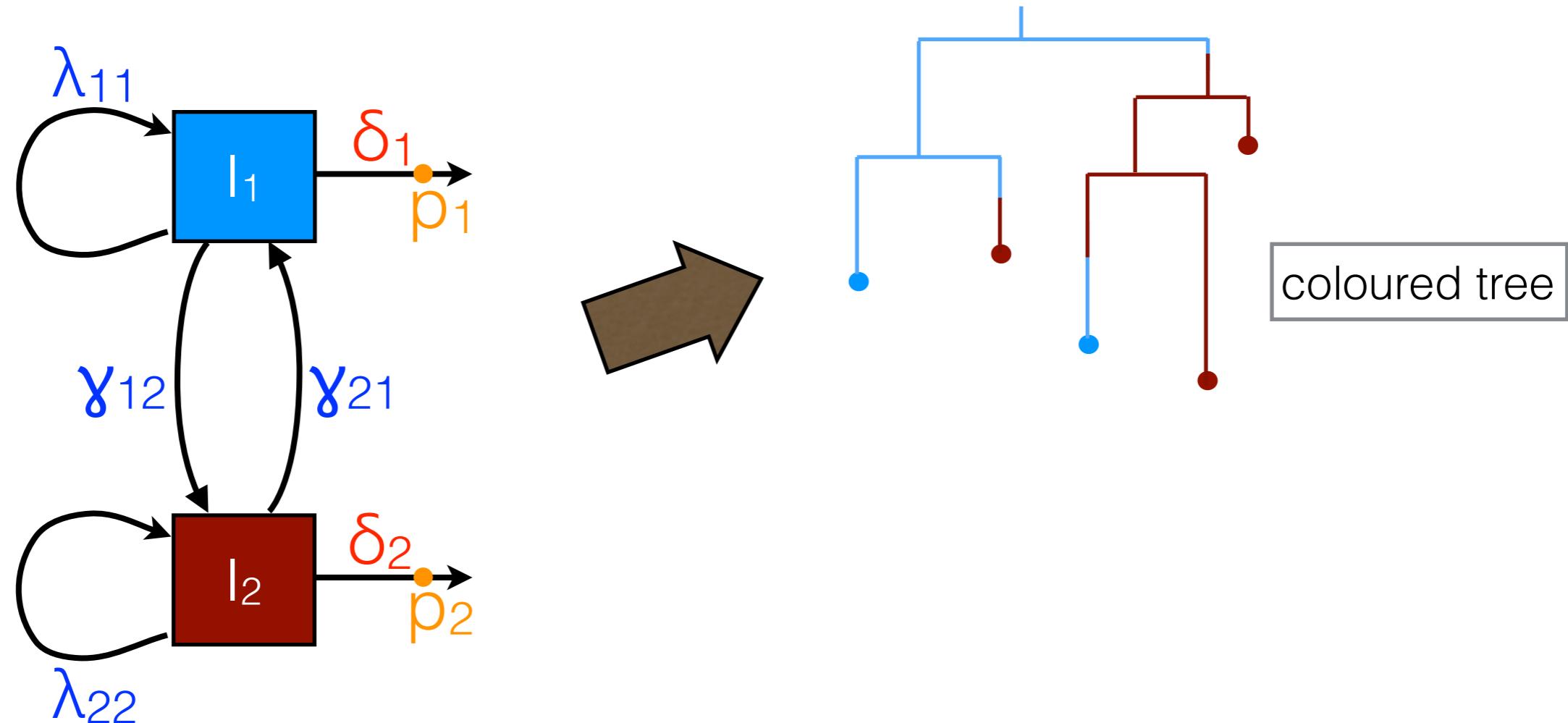
Modelling population structure



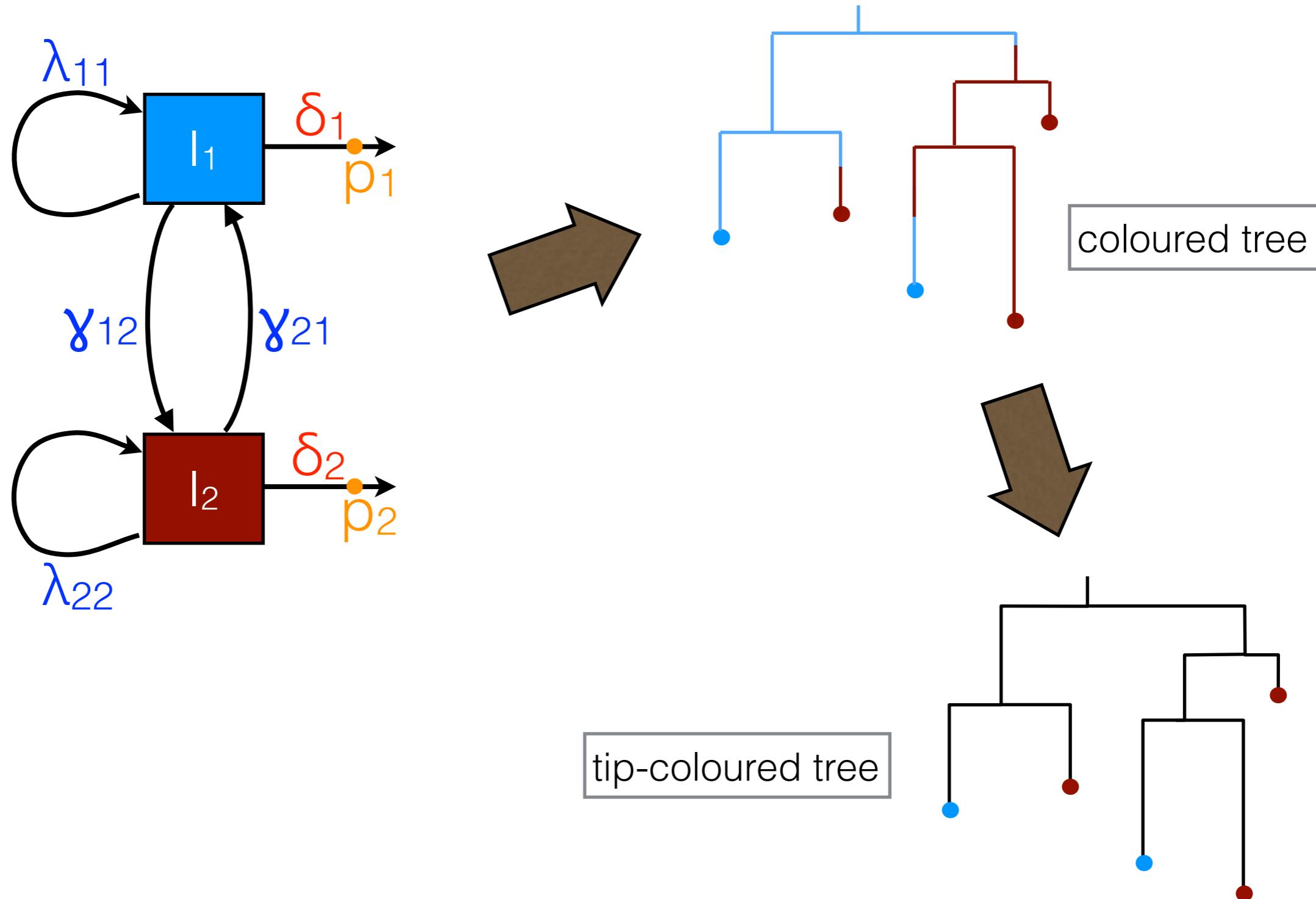
Modelling population structure



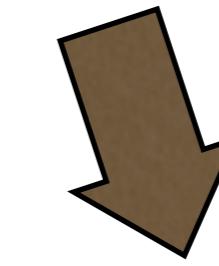
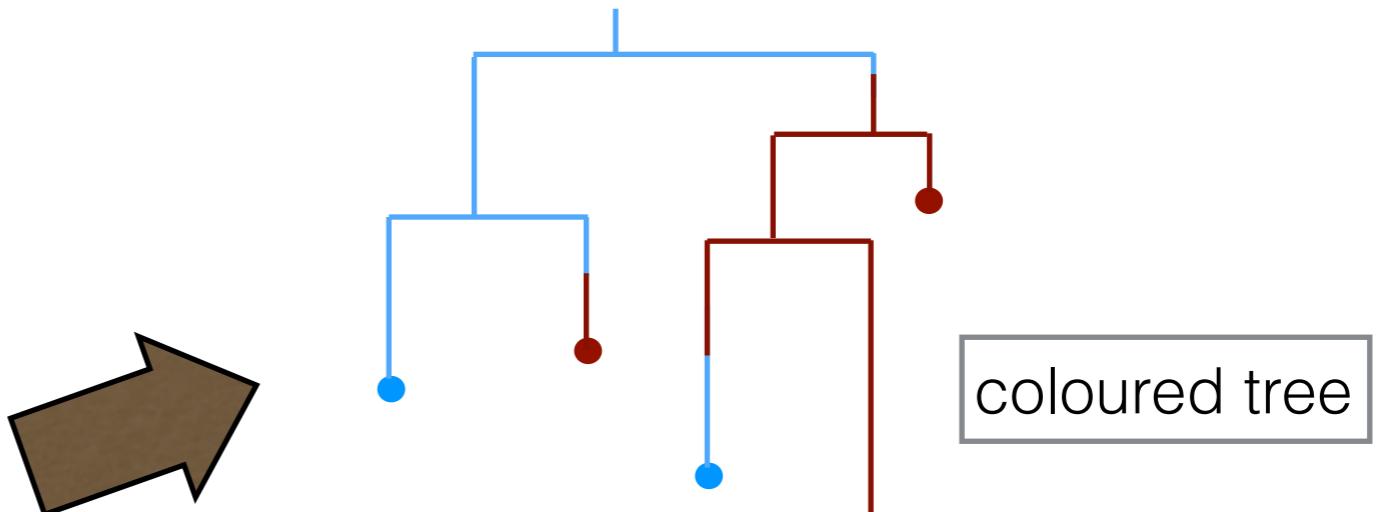
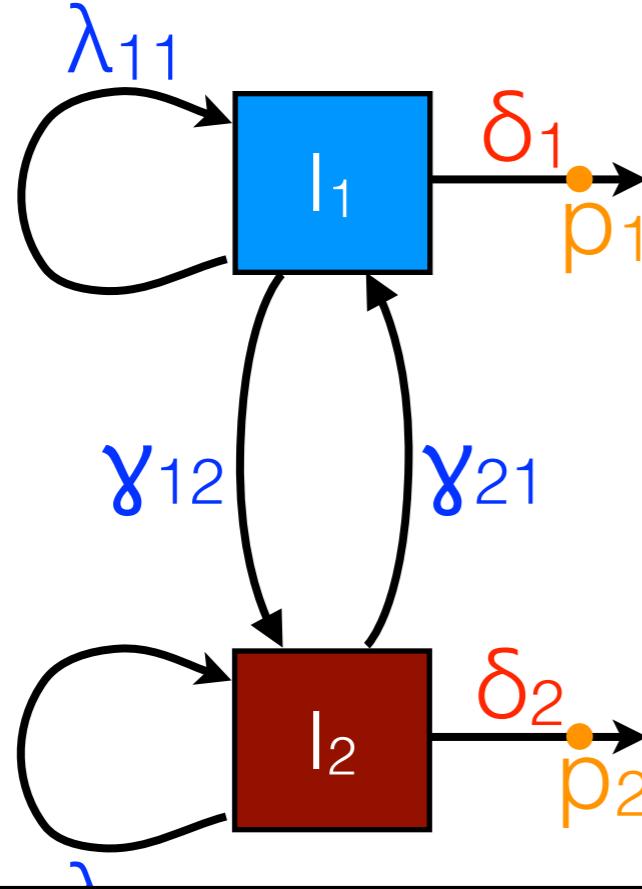
Modelling population structure



Modelling population structure

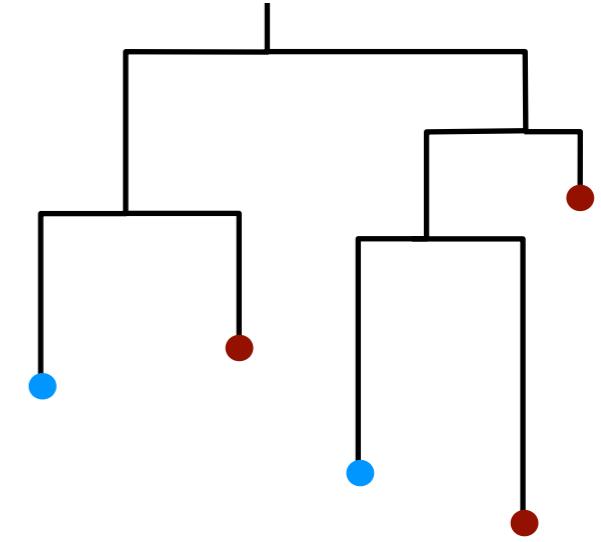


Modelling population structure



Estimate transmission / become noninfectious / migration / sampling rates across demes (eg. space; strains)

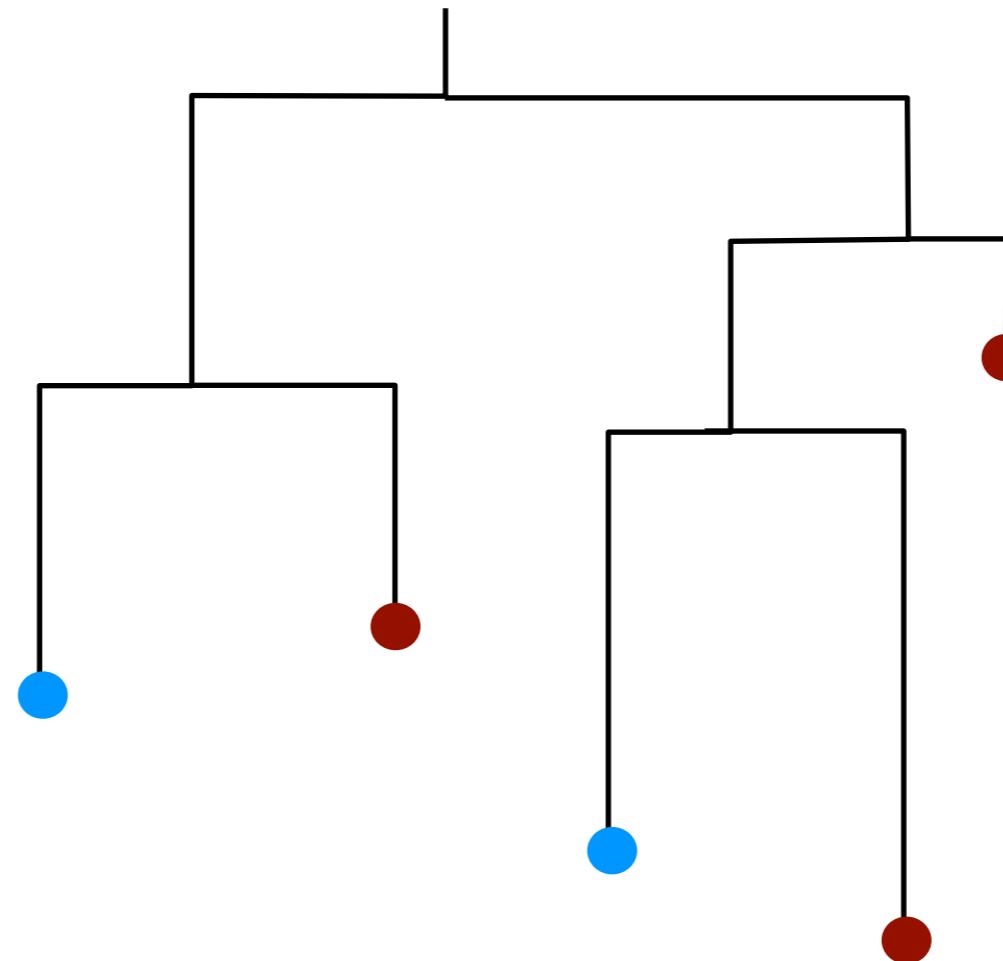
red tree



How can we quantify the rates?

Multi-type birth-death model

Structured coalescent

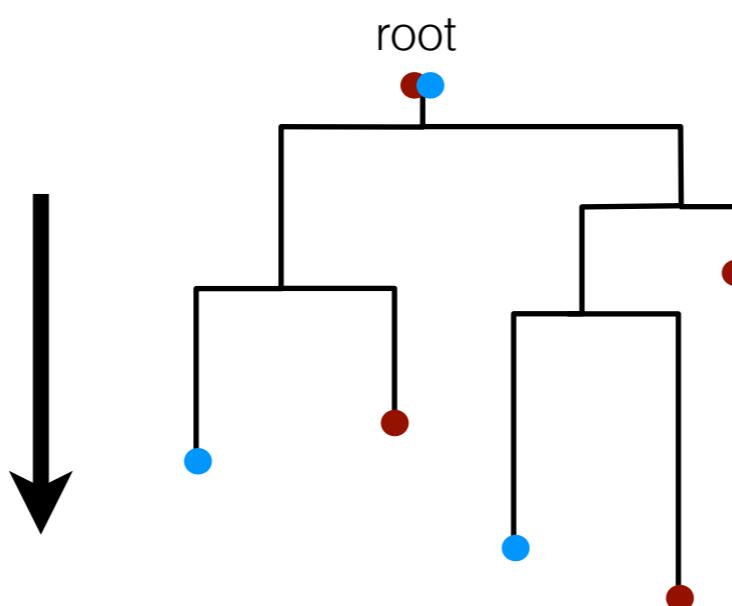


How can we quantify the rates?

Multi-type birth-death model

Structured coalescent

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



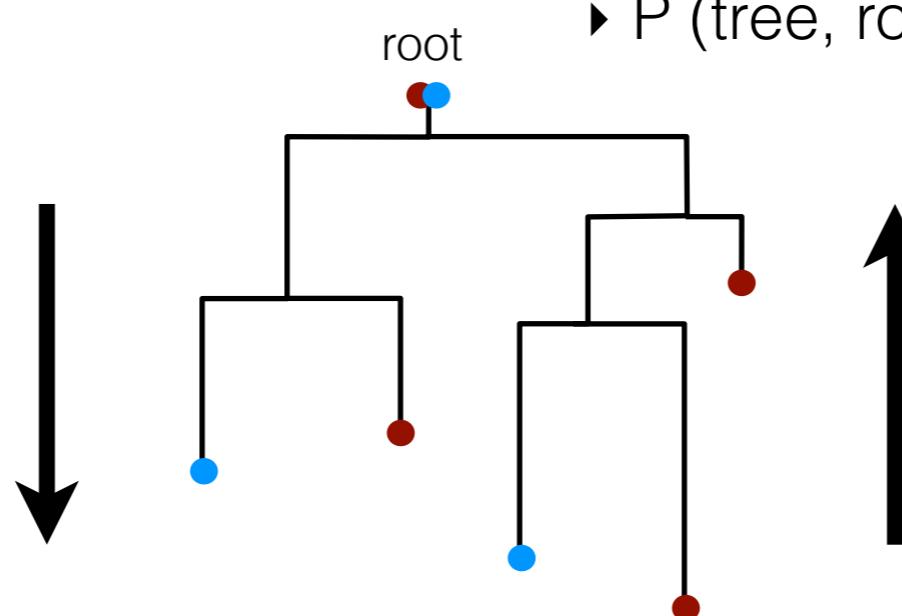
How can we quantify the rates?

Multi-type birth-death model

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

Structured coalescent

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

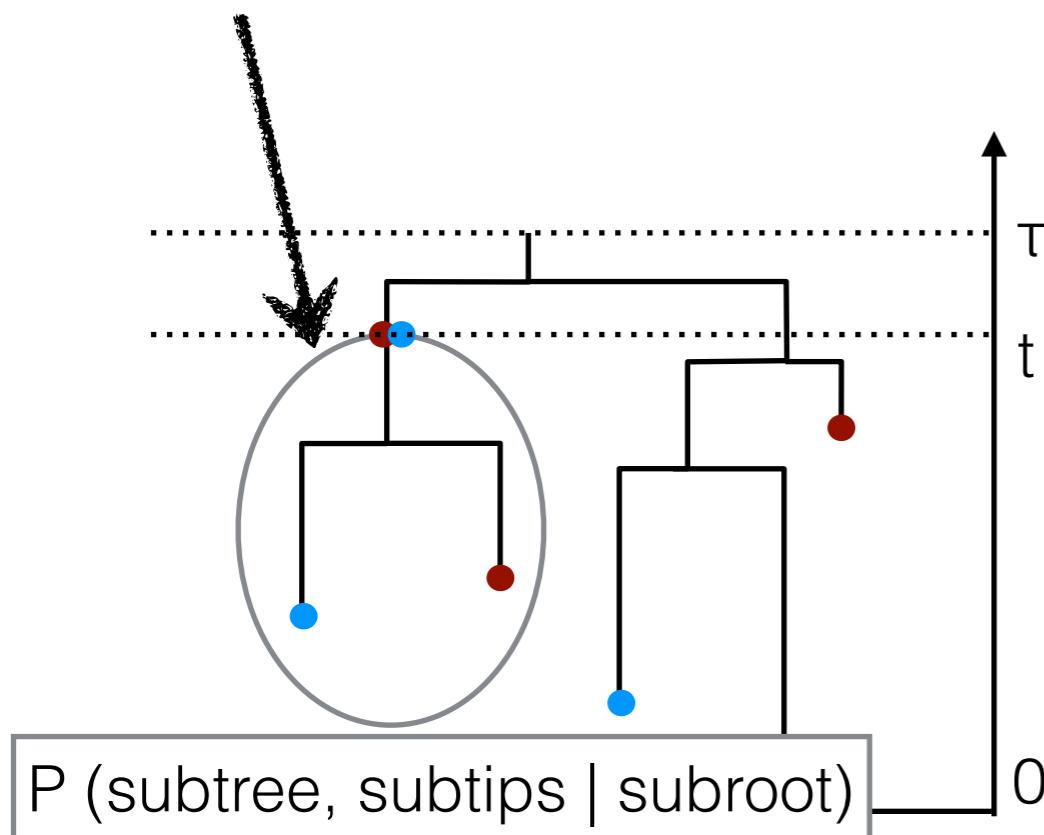


Inferring tip-coloured trees

Multi-type birth-death model
(Stadler & Bonhoeffer, 2013)

Marginal approx. of the structured coalescent (Müller et al., 2017)

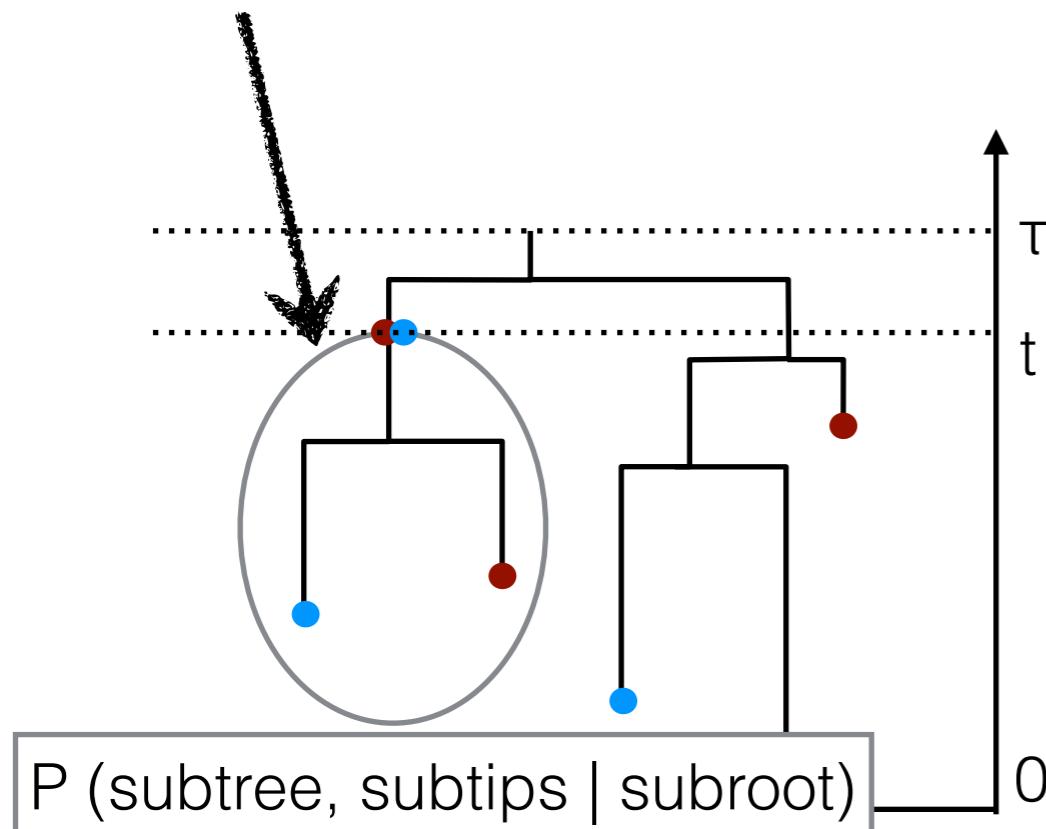
- ▶ $P(\text{tree, tips} \mid \text{root}; n)$
- ▶ Integrate along each branch (backward in time)



Inferring tip-coloured trees

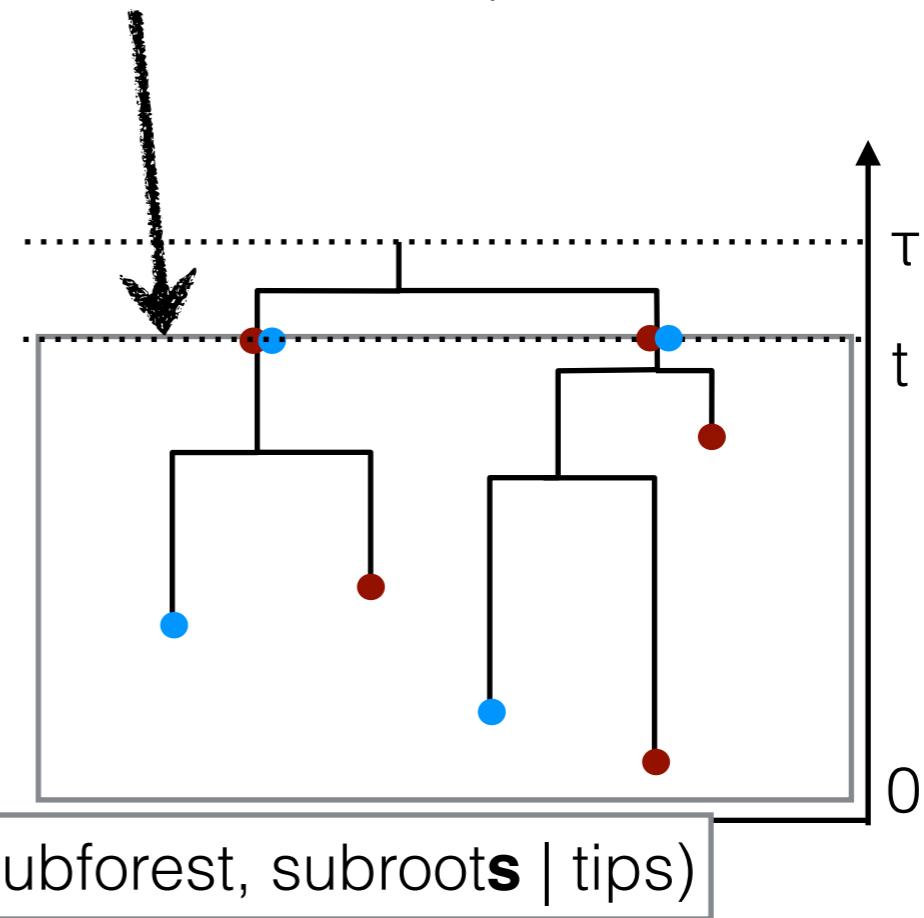
Multi-type birth-death model
(Stadler & Bonhoeffer, 2013)

- ▶ $P(\text{tree, tips} \mid \text{root}; n)$
- ▶ Integrate along each branch (backward in time)



Marginal approx. of the structured coalescent (Müller et al., 2017)

- ▶ $P(\text{tree, root} \mid \text{tips}; n)$
- ▶ Integrate along all branches (backward in time)

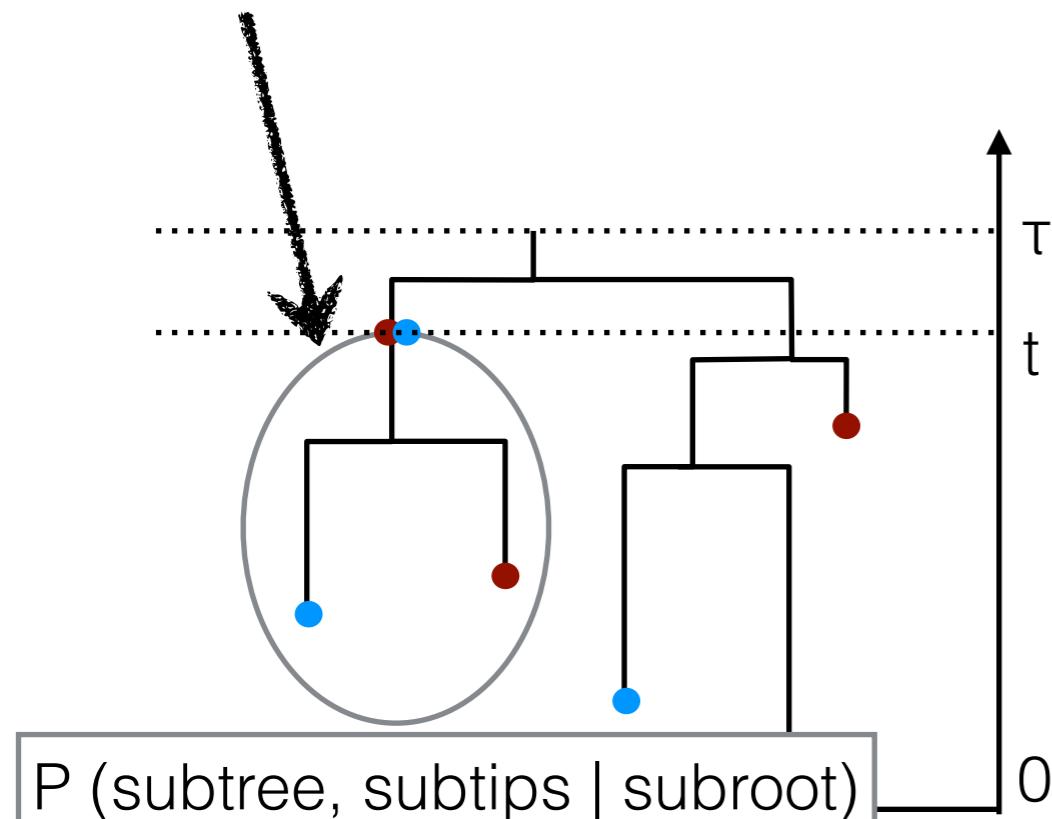


Inferring tip-coloured trees

Multi-type birth-death model
(Stadler & Bonhoeffer, 2013)

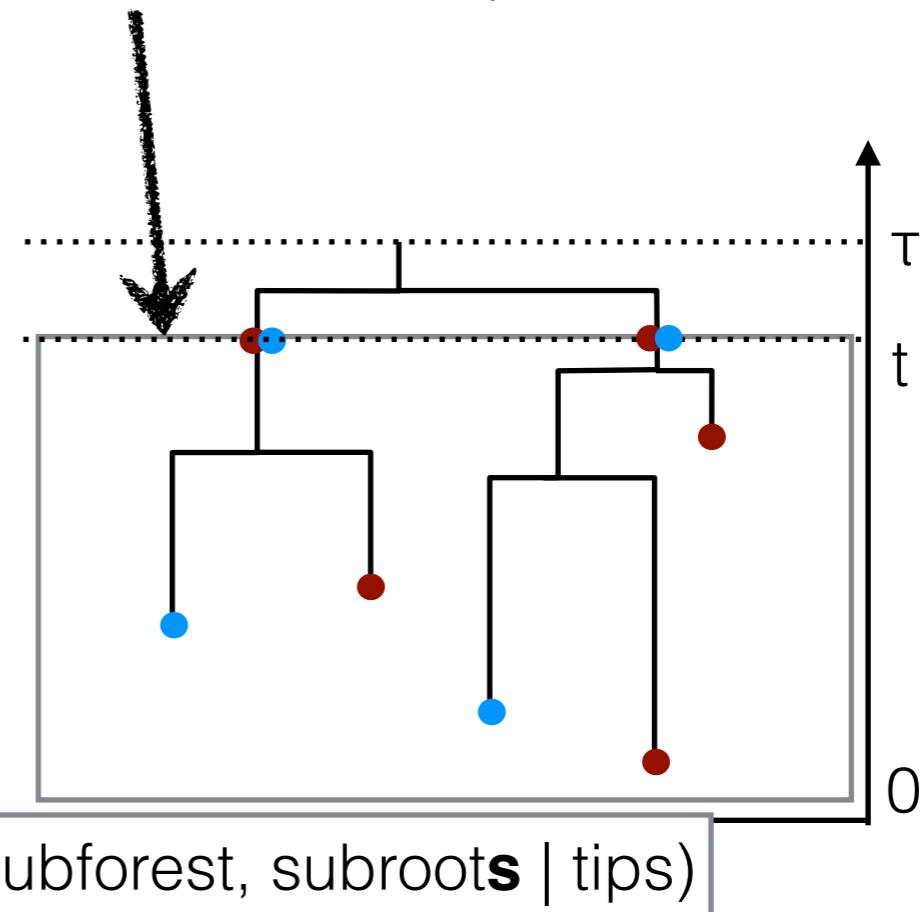
BEAST add-on bdmm

- ▶ $P(\text{tree}, \text{tips} | \text{root})$
- ▶ Integrate along each branch (backward in time)



Marginal approx. of the structured coalescent (Müller et al., 2017)

- ▶ $P(\text{tree, root} | \text{tips}; n)$
- ▶ Integrate along all branches (backward in time)

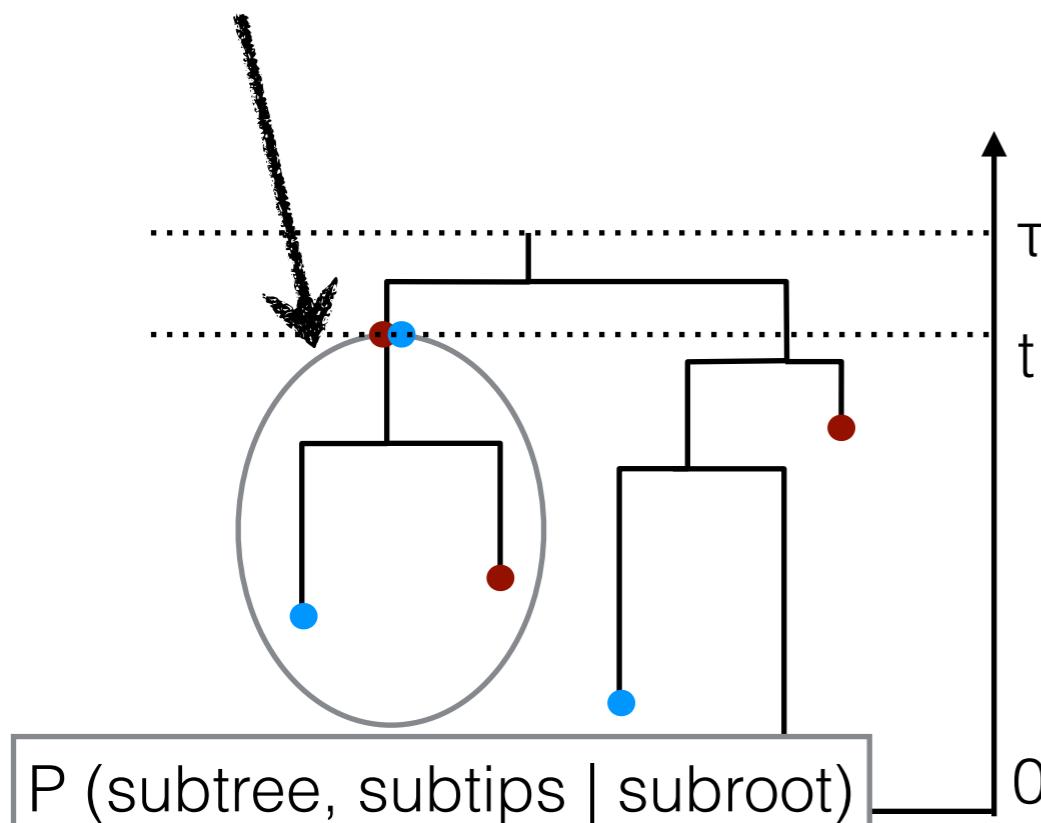


Inferring tip-coloured trees

Multi-type birth-death model
(Stadler & Bonhoeffer, 2013)

BEAST add-on bdmm

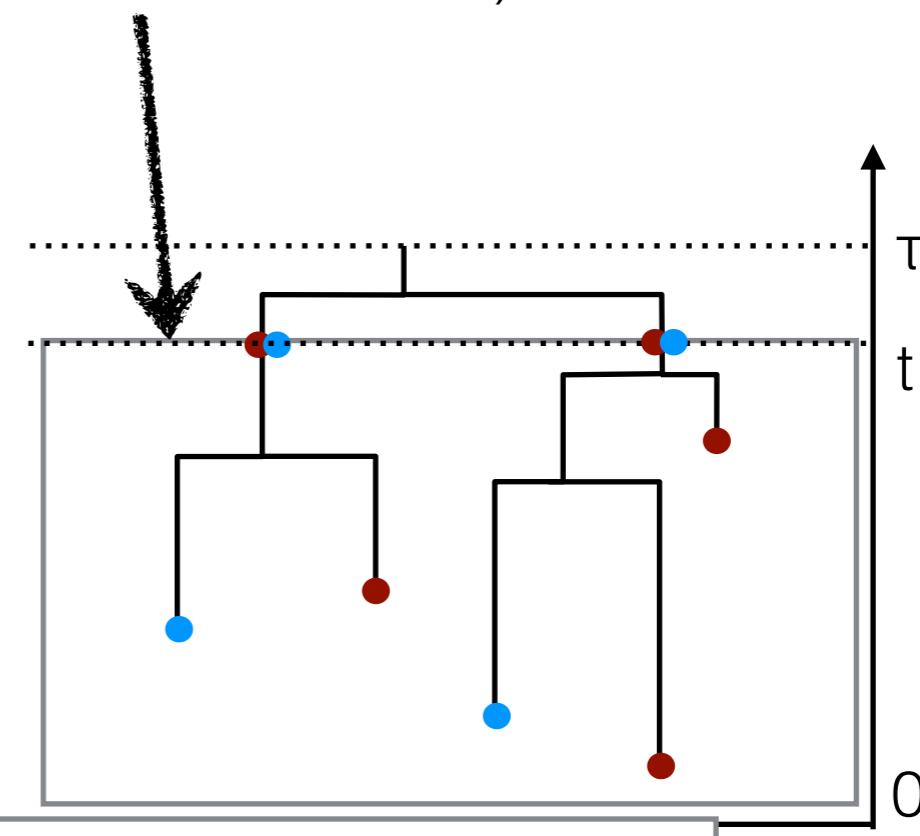
- ▶ $P(\text{tree}, \text{tips} | \text{subroot})$
- ▶ Integrate along each branch
(backward in time)



Marginal approx. of the structured coalescent (Müller et al., 2017)

BEAST add-on Mascot

- ▶ $P(\text{tree}, \text{tips} | \text{subroot})$
- ▶ Integrate along all branches
(backward in time)

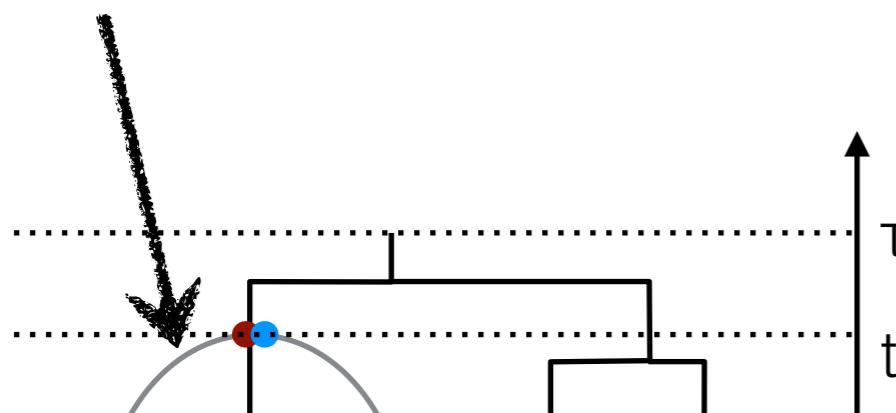


Inferring tip-coloured trees

Multi-type birth-death model
(Stadler & Bonhoeffer, 2013)

BEAST add-on bdmm

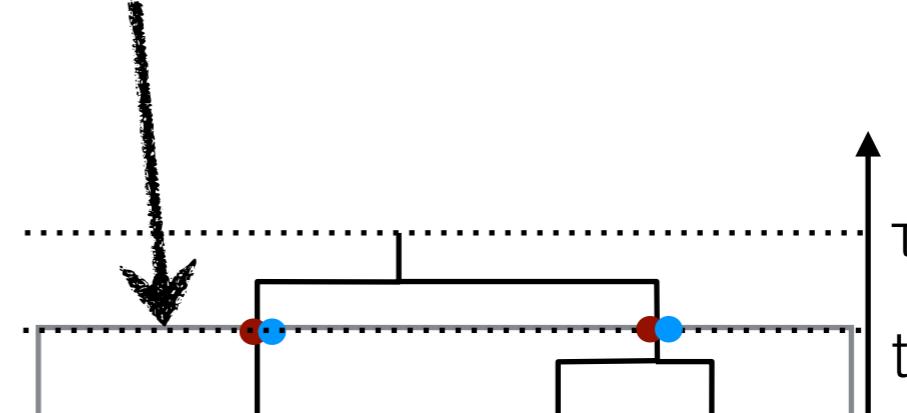
- ▶ $P(\text{tree}, \text{tips} | \text{subroot})$
- ▶ Integrate along each branch
(backward in time)



Marginal approx. of the structured coalescent (Müller et al., 2017)

BEAST add-on Mascot

- ▶ $P(\text{tree}, \text{tips} | \text{subroot})$
- ▶ Integrate along all branches
(backward in time)



Linear number (in demes) of ODEs to solve

$P(\text{subtree, subtips} | \text{subroot})$

$P(\text{subforest, subroots} | \text{tips})$

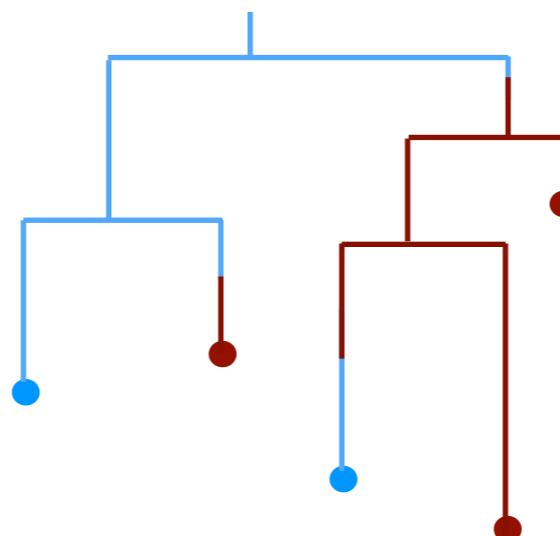
Inferring branch-coloured trees

Multi-type birth-death model
(Kühnert et al, 2016)

- ▶ $P(\text{coloured_tree}, \text{tips} | \text{root})$
- ▶ Simpler system of ODEs
- ▶ Direct estimate of probability of colour for each lineage

Structured coalescent
(Vaughan et al, 2014:)

- ▶ $P(\text{coloured_tree}, \text{root} | \text{tips})$
- ▶ No ODEs
- ▶ Direct estimate of probability of colour for each lineage



Inferring branch-coloured trees

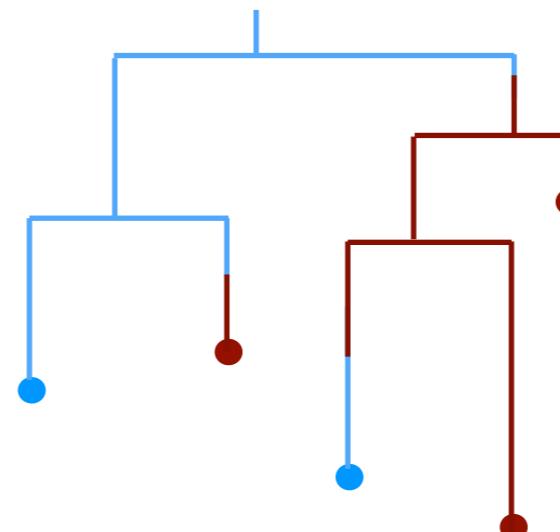
Multi-type birth-death model
(Kühnert et al, 2016)

BEAST add-on bdmm

- ▶ P (coloured_tree, root | tips)
- ▶ Simpler system of ODEs
- ▶ Direct estimate of probability of colour for each lineage

Structured coalescent
(Vaughan et al, 2014:)

- ▶ P (coloured_tree, root | tips)
- ▶ No ODEs
- ▶ Direct estimate of probability of colour for each lineage



Inferring branch-coloured trees

Multi-type birth-death model
(Kühnert et al, 2016)

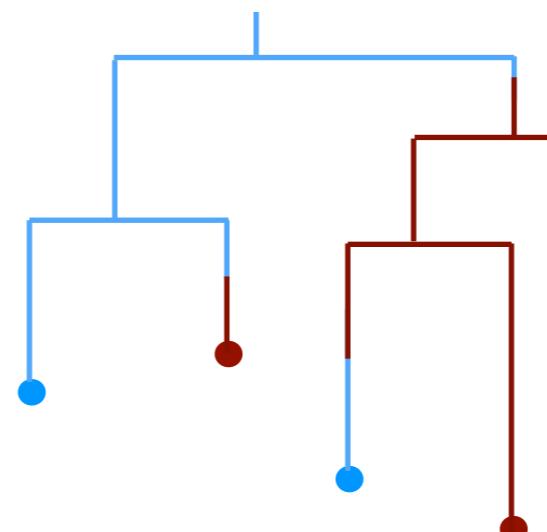
BEAST add-on **bdmm**

- ▶ ~~Parameterised~~
- ▶ Simpler system of ODEs
- ▶ Direct estimate of probability of colour for each lineage

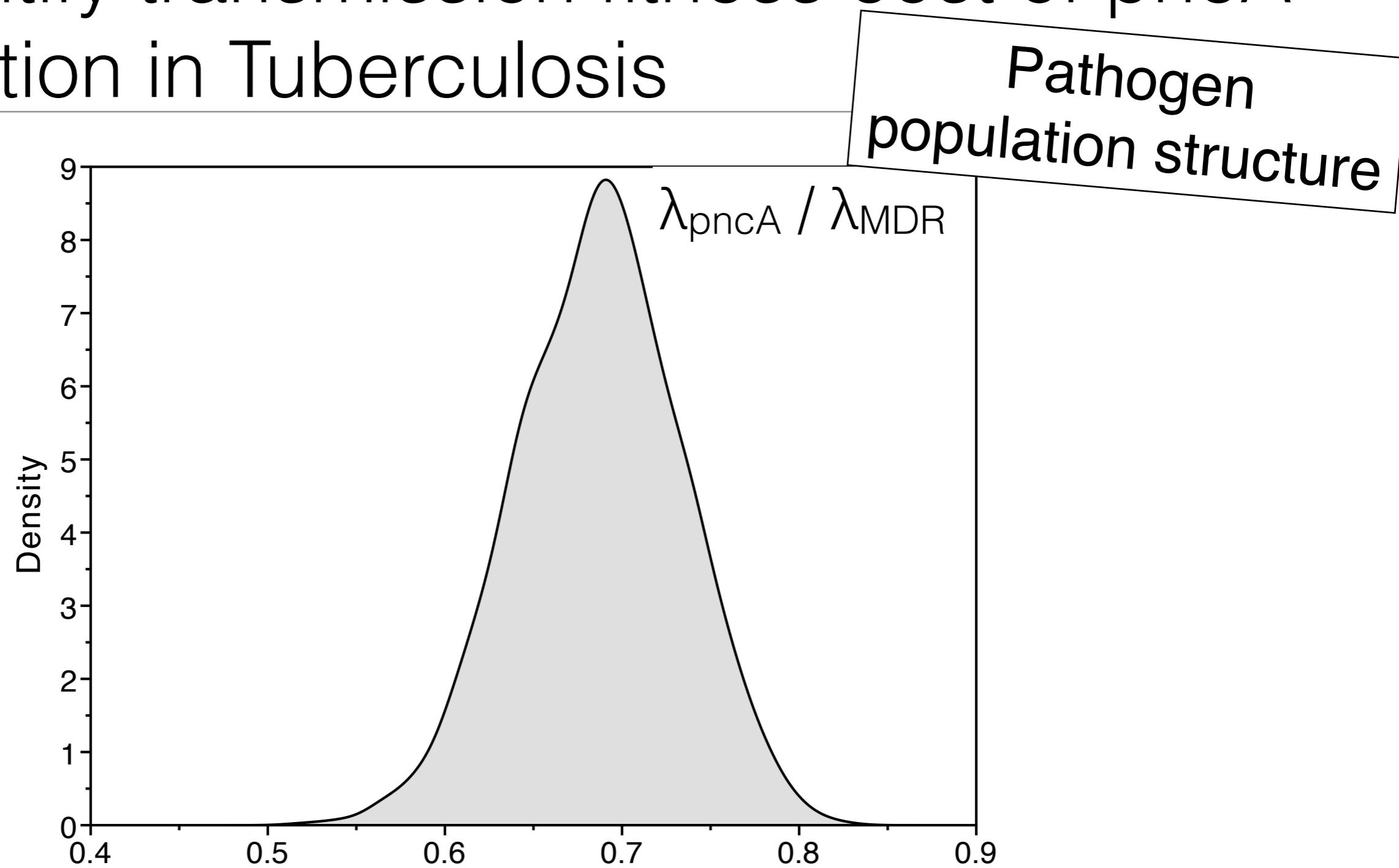
Structured coalescent
(Vaughan et al, 2014:)

BEAST add-on **MultiTypeTree**

- ▶ Direct estimate of probability of colour for each lineage

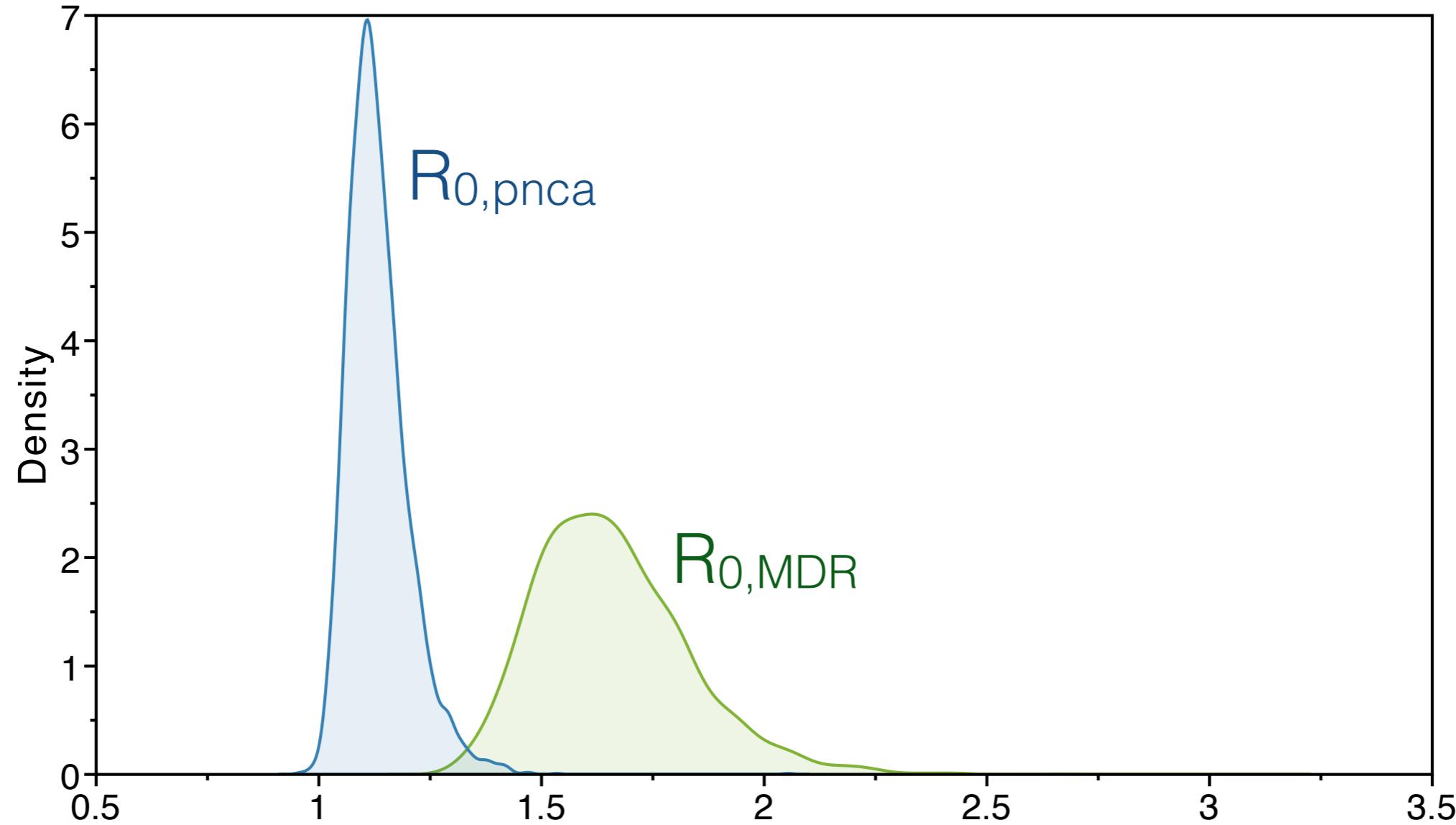


Quantify transmission fitness cost of pncA mutation in Tuberculosis



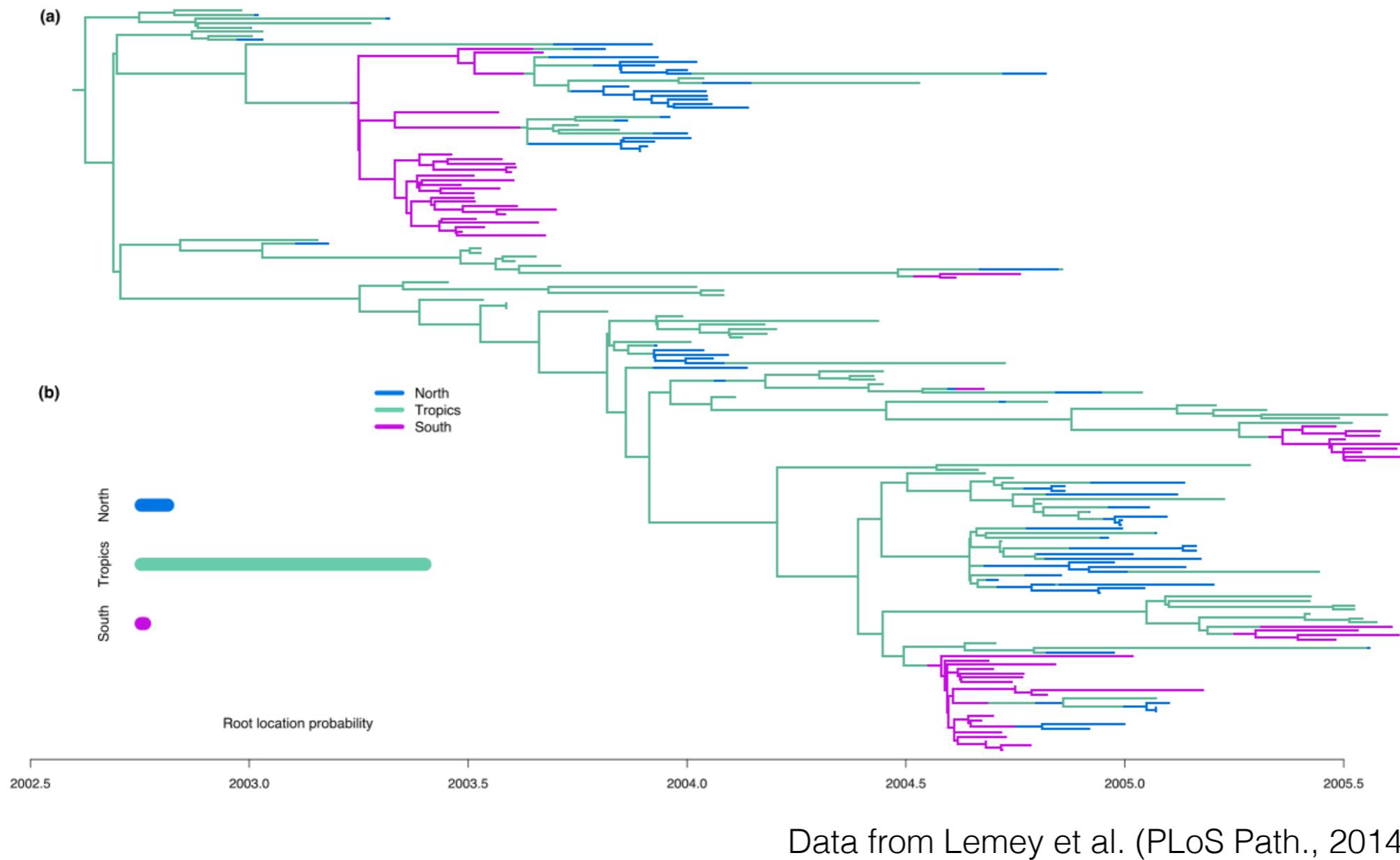
Two analyses of 150 MDR genomes from Kinshasa each. A third of them have the pncA mutation.

Quantify transmission fitness cost of pncA mutation in Tuberculosis

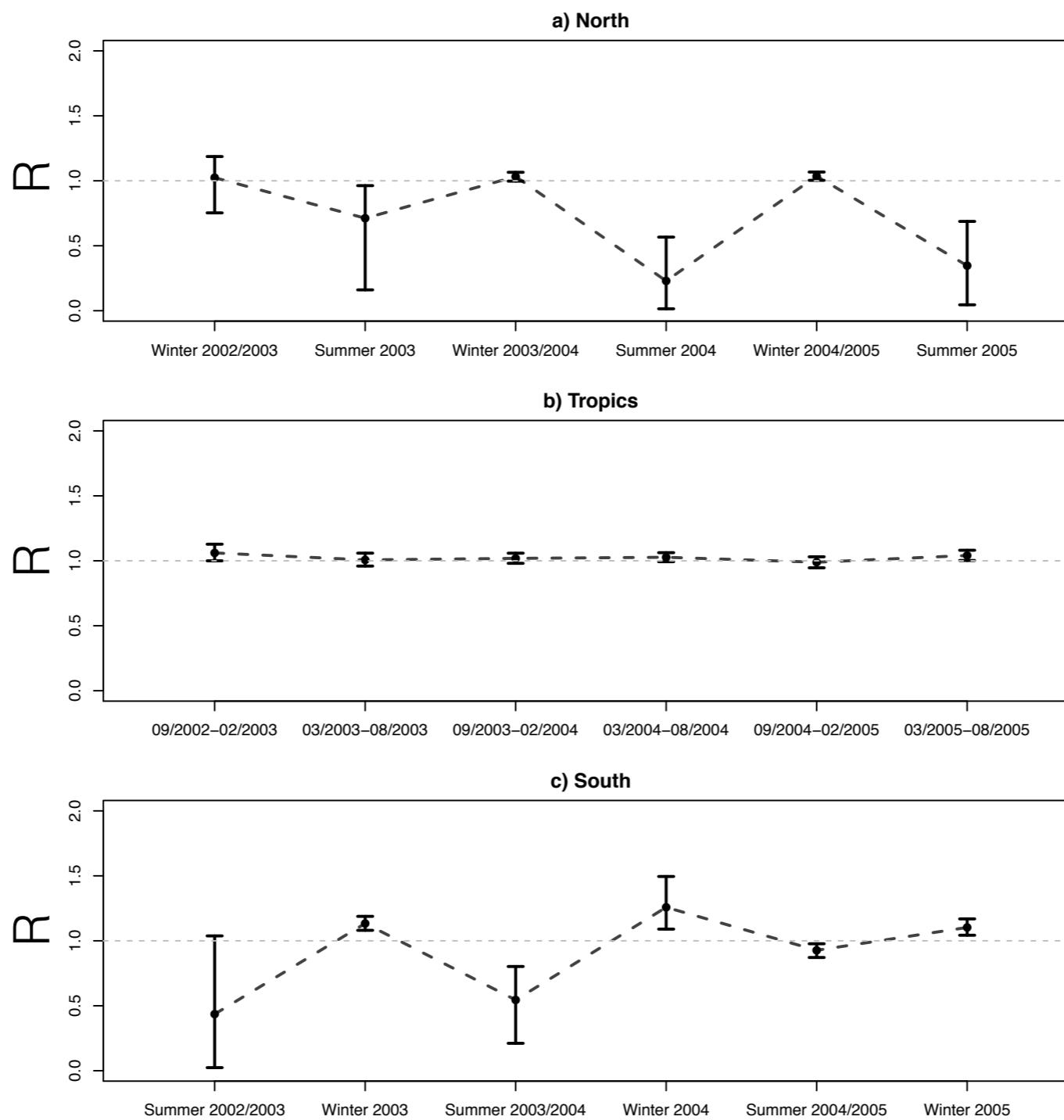


Global spread of seasonal influenza: coloured tree

Host
population structure



Global spread of seasonal influenza: parameter estimates



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 (skylines)

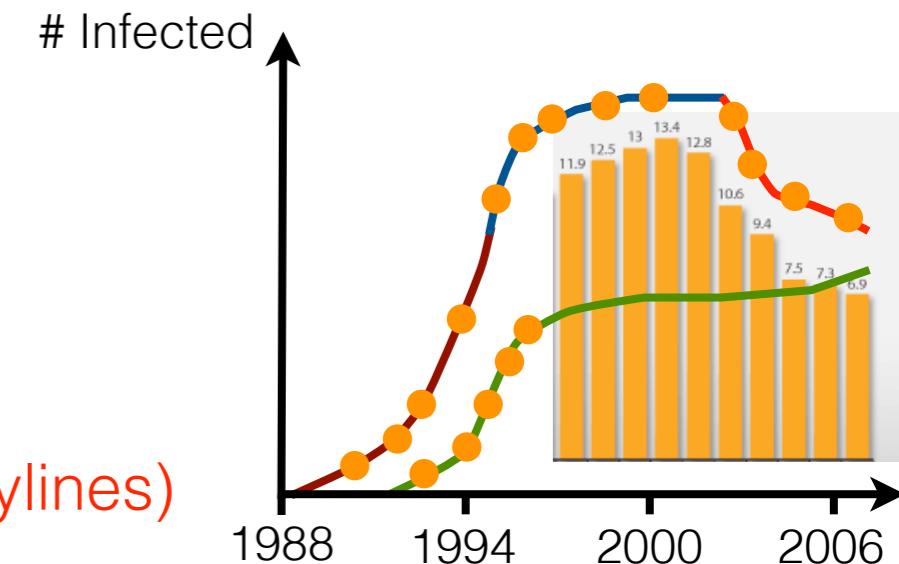
- ▶ C: Drummond et al. (*MBE*, 2005)
- ▶ BD: Stadler, Kühnert et al. (*PNAS*, 2013)

3. ...accounting for SIR model dynamics

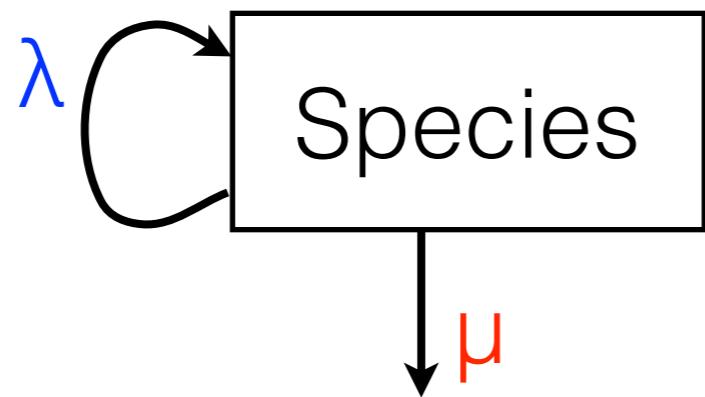
- ▶ C: Volz et al. (*Genetics*, 2009)
- ▶ BD: Kühnert et al. (*Roy. Soc. Int.*, 2014), Leventhal et al. (*MBE*, 2014), , Vaughan et al. (submitted)

4. ...accounting for population structure

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



Outlook: The birth-death model as a model for speciation and extinction

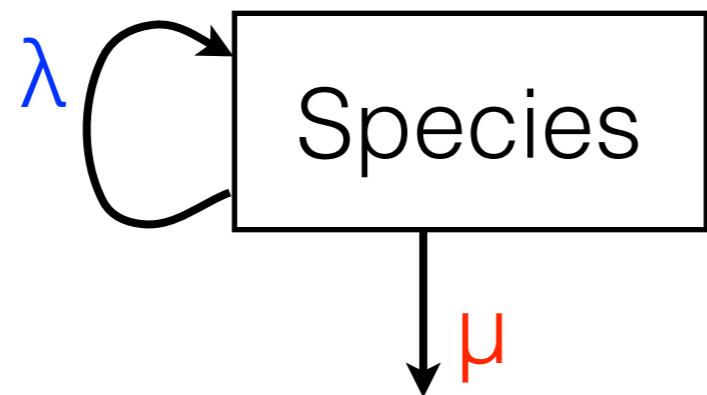
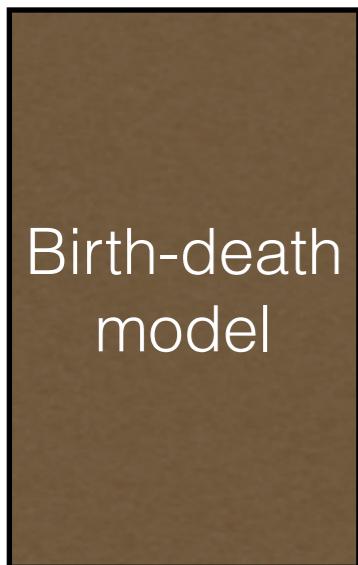


- stem age of a clade t
- speciation rate λ
- extinction rate μ
- sampling probability p

Rates may depend on:

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

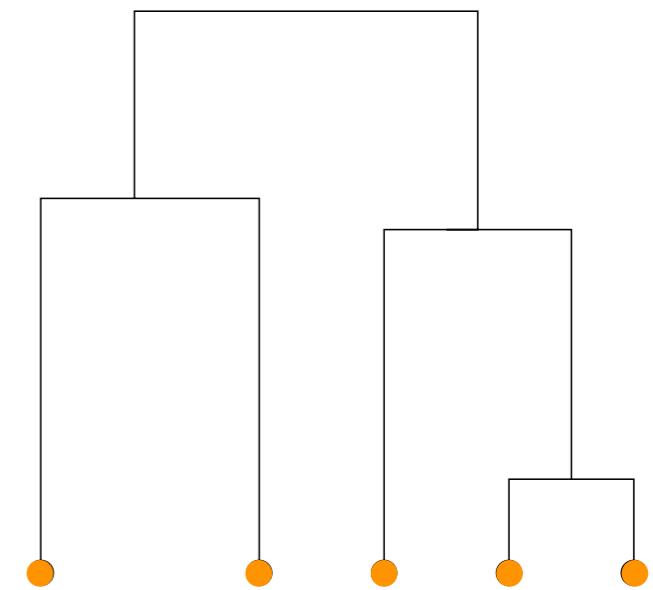
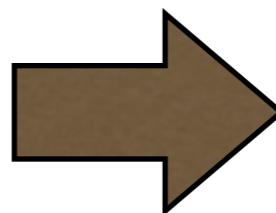
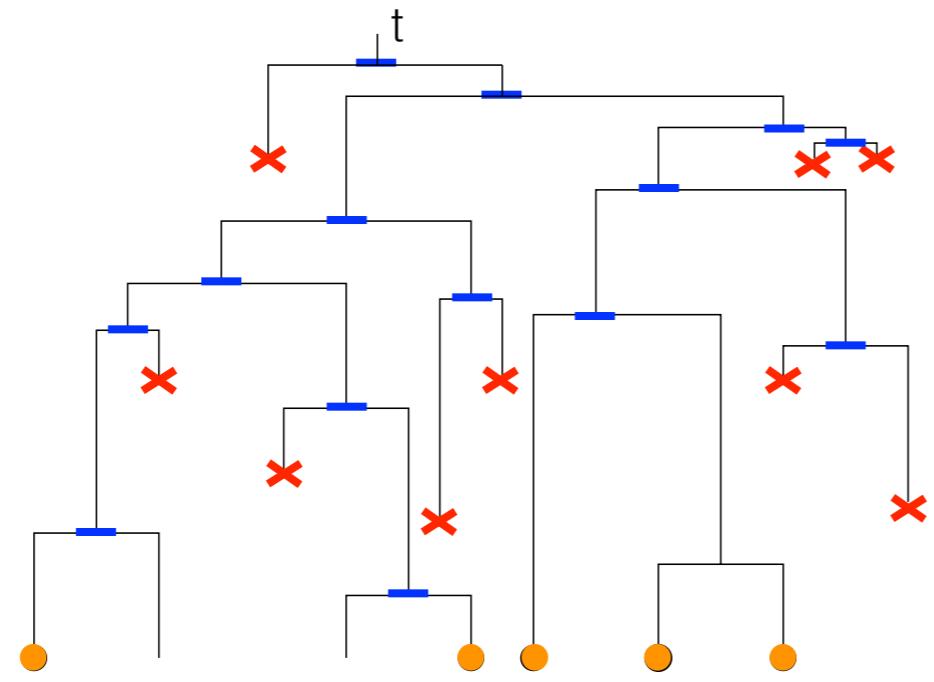
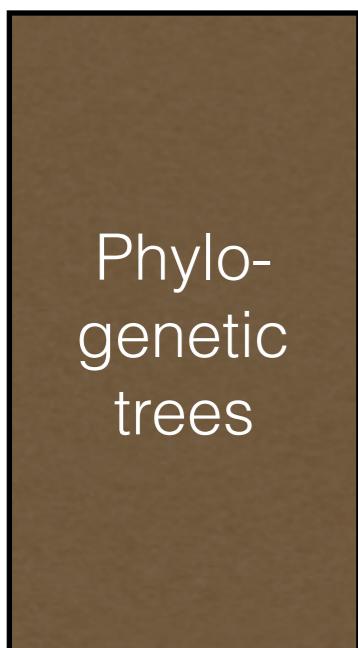
Outlook: The birth-death model as a model for speciation and extinction



- stem age of a clade t
- speciation rate λ
- extinction rate μ
- sampling probability p

Rates may depend on:

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



Phylogenetic methods for different population dynamics

1. ...for analyzing epidemic outbreaks

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

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

- ▶ 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), , Vaughan et al. (submitted)

4. ...accounting for population structure

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