



JOHNS HOPKINS
MEDICINE

An introduction to infectious disease phylogenetics

Michael A. Martin, PhD

mmart108@jhmi.edu

2023-08-15

An introduction to infectious disease phylodynamics

An introduction to infectious disease phylodynamics

Phylodynamics: the study of how epidemiological and evolutionary processes act and/or interact to shape the phylogenetic pattern of pathogen populations

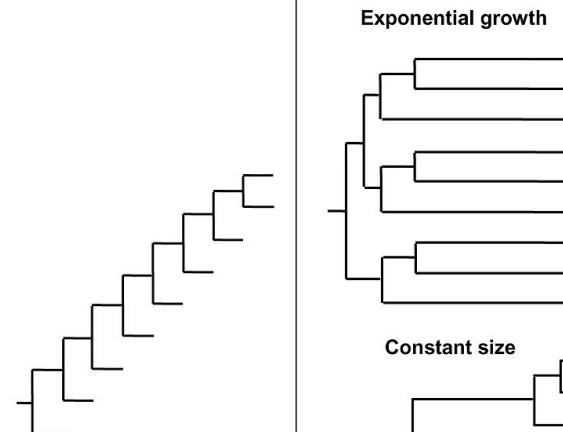
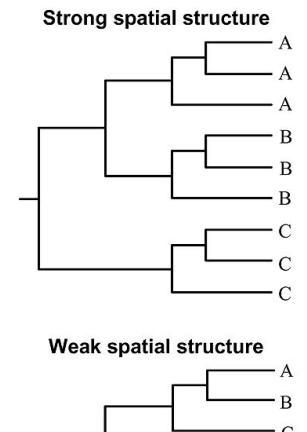
An introduction to infectious disease phylodynamics

REVIEW

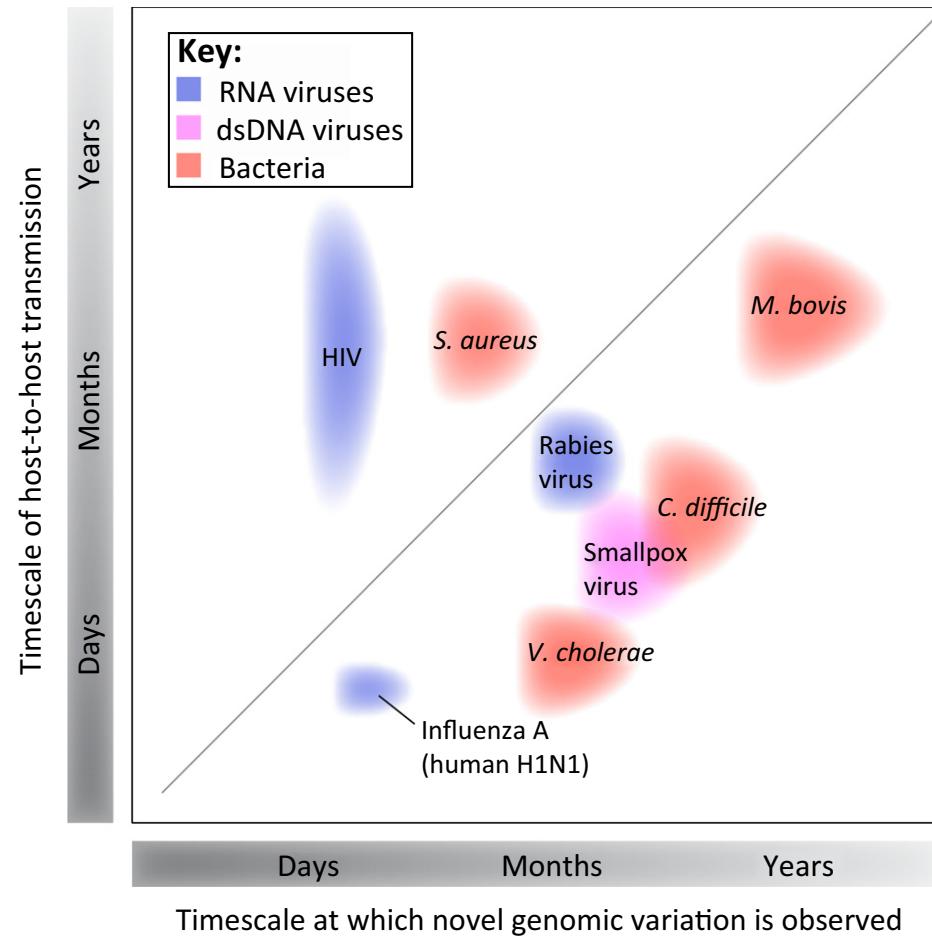
Unifying the Epidemiological and Evolutionary Dynamics of Pathogens

Bryan T. Grenfell,^{1*} Oliver G. Pybus,² Julia R. Gog,¹ James L. N. Wood,³ Janet M. Daly,³ Jenny A. Mumford,³ Edward C. Holmes²

Phylodynamics: the study of how epidemiological and evolutionary processes act and/or interact to shape the phylogenetic pattern of pathogen populations

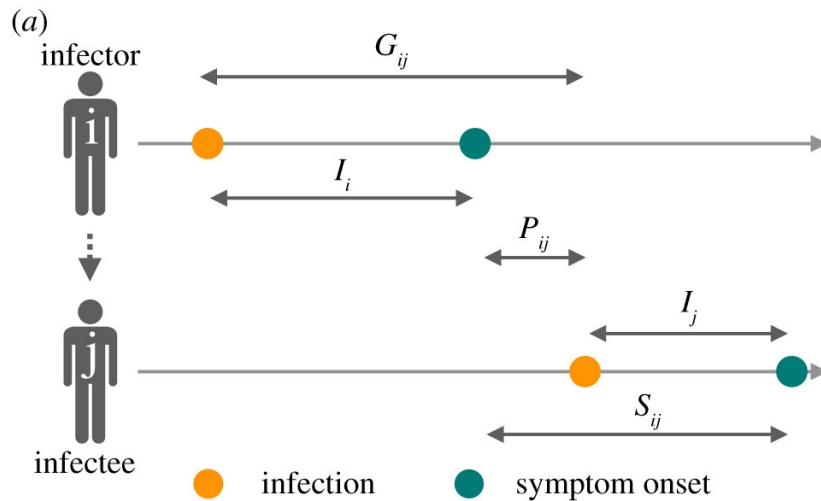
	Continual Immune Selection	Weak or Absent Immune Selection	
		Tree shape controlled by non-selective population dynamic processes	
Idealized Phylogeny Shapes		Population size dynamics  Time →	Spatial dynamics 
Examples	Human influenza A virus intra-host HIV	inter-host HIV inter-host HCV	Measles, rabies inter-host HIV
Tree Inferences	Detection of antigenic escape mutations	Estimation of population growth rates	Estimation of population migration rates

Infectious disease phylodynamics is possible because of (relatively) high substitution rates



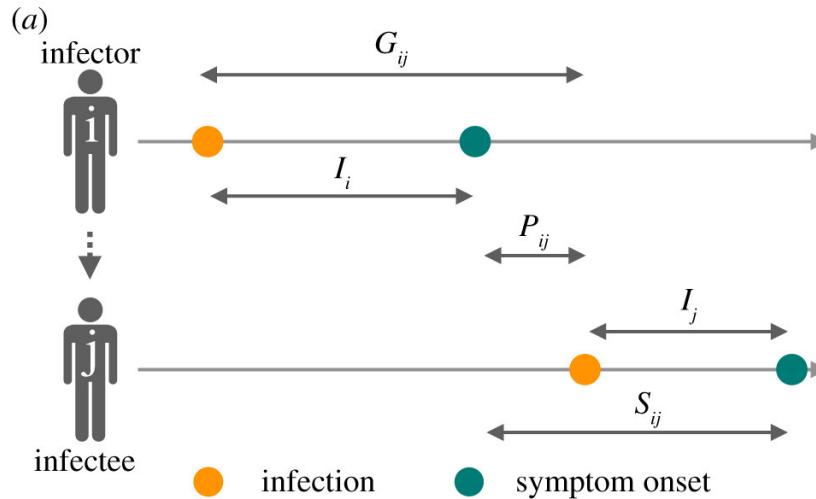
Infectious disease phylodynamics is possible because of (relatively) high substitution rates

Timescale of ecological dynamics is governed by the generation interval

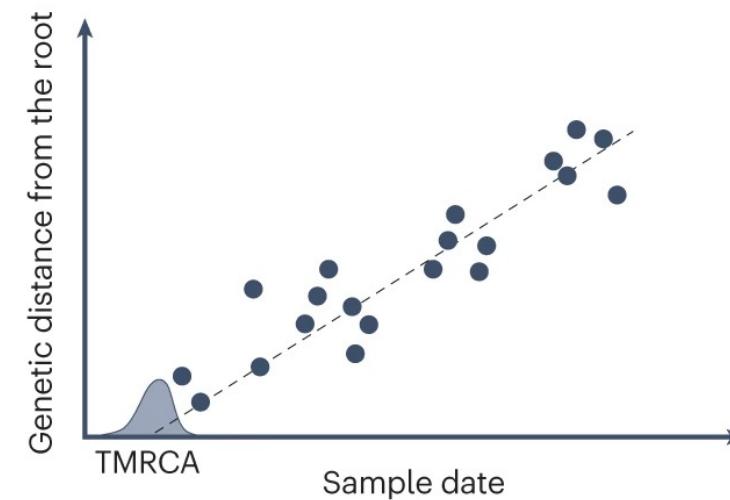


Infectious disease phylodynamics is possible because of (relatively) high substitution rates

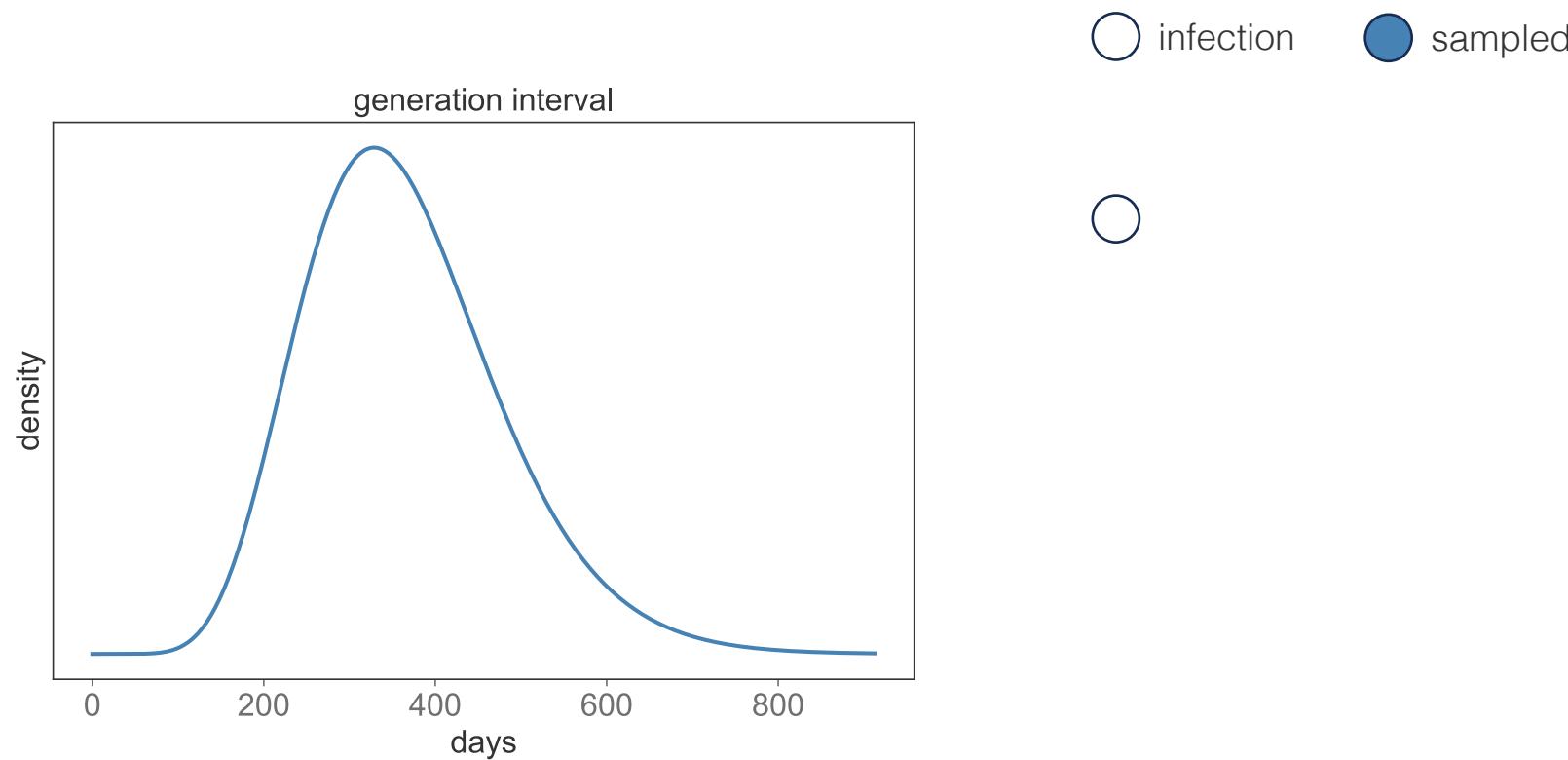
Timescale of ecological dynamics is governed by the generation interval



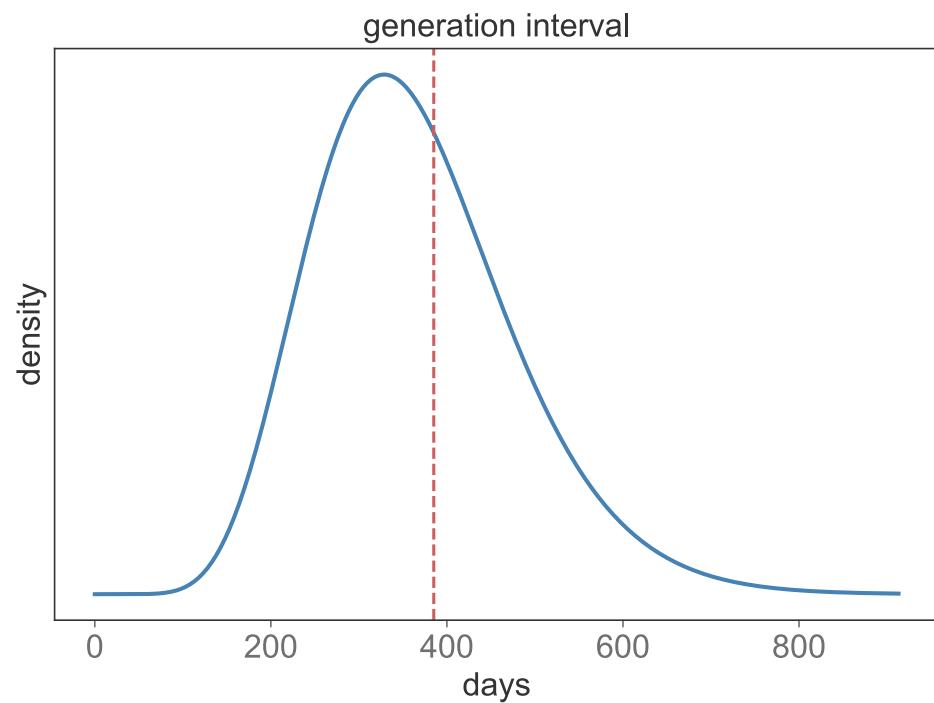
Timescale of evolutionary dynamics is governed by the substitution rate



Infectious disease phylodynamics is possible because of (relatively) high substitution rates



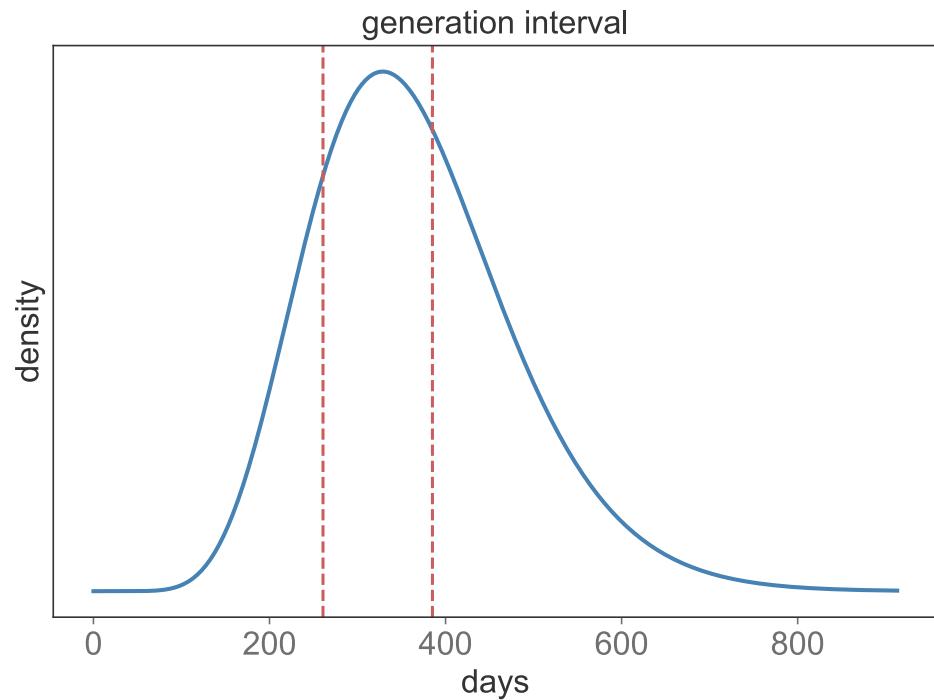
Infectious disease phylodynamics is possible because of (relatively) high substitution rates



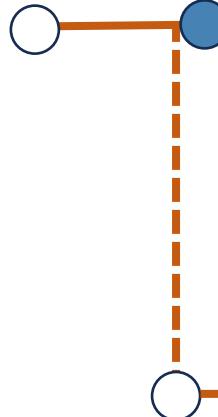
○ infection ● sampled



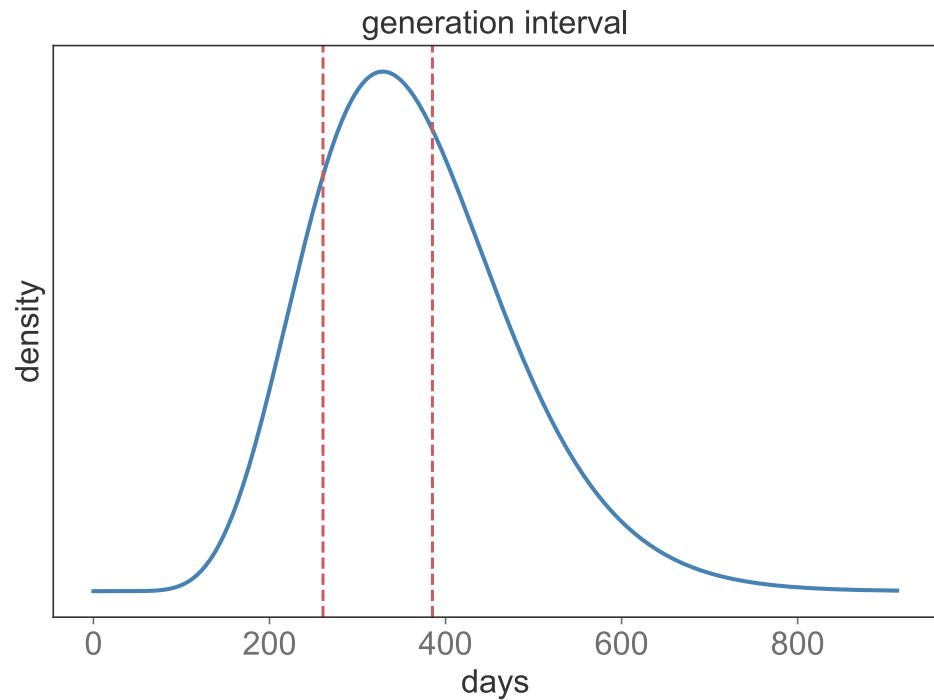
Infectious disease phylodynamics is possible because of (relatively) high substitution rates



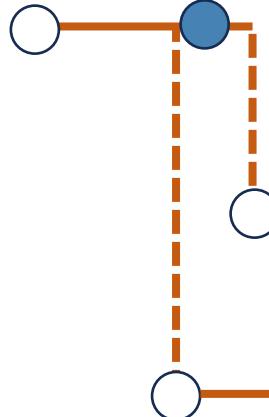
○ infection ● sampled



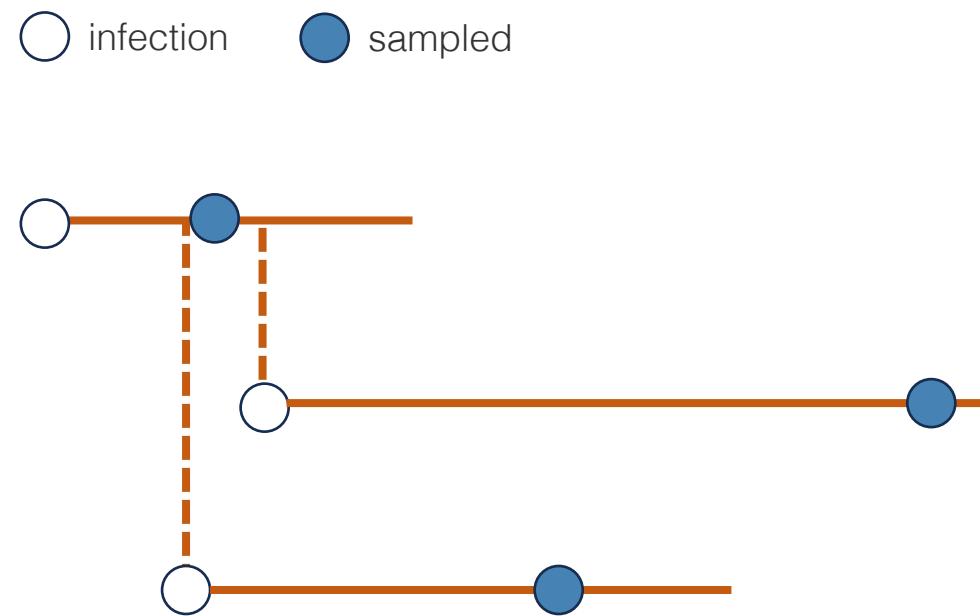
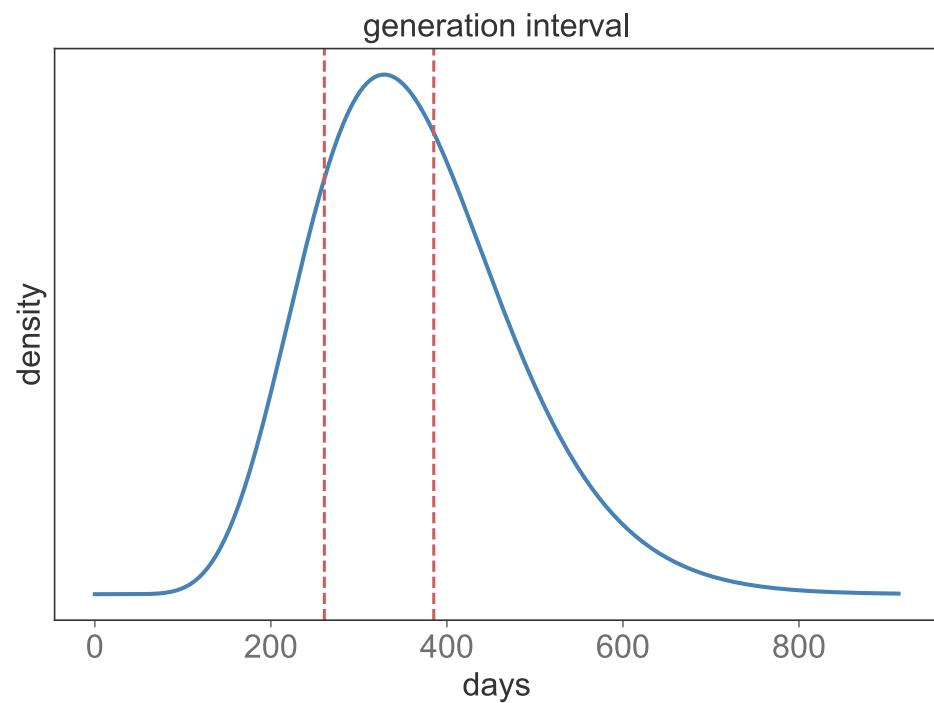
Infectious disease phylodynamics is possible because of (relatively) high substitution rates



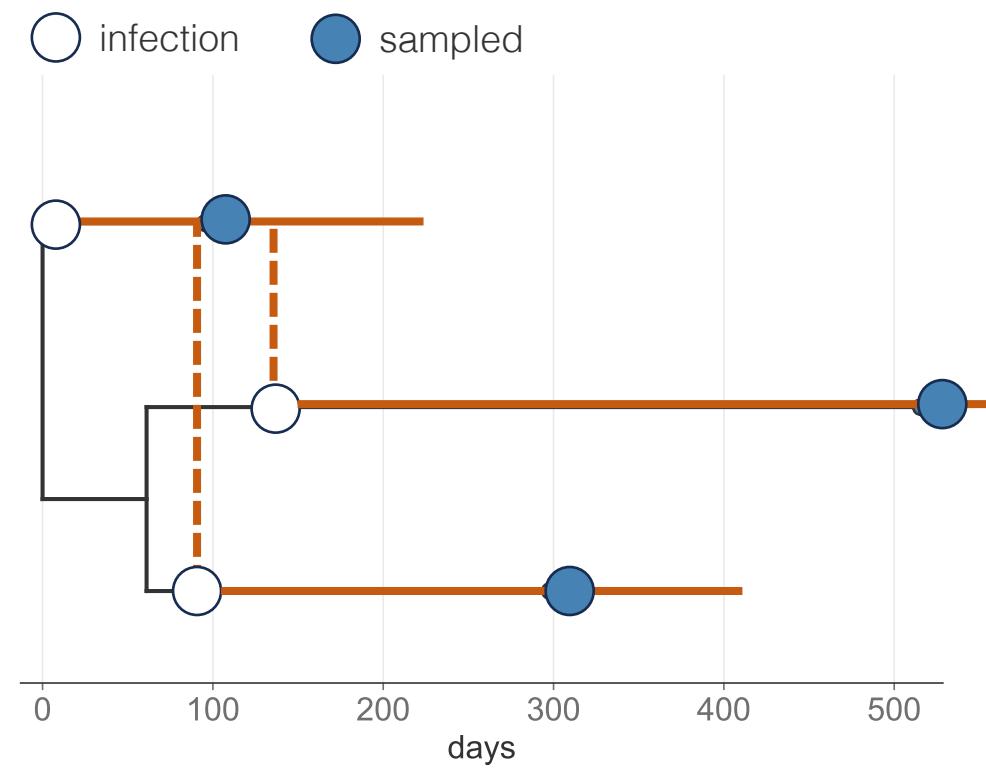
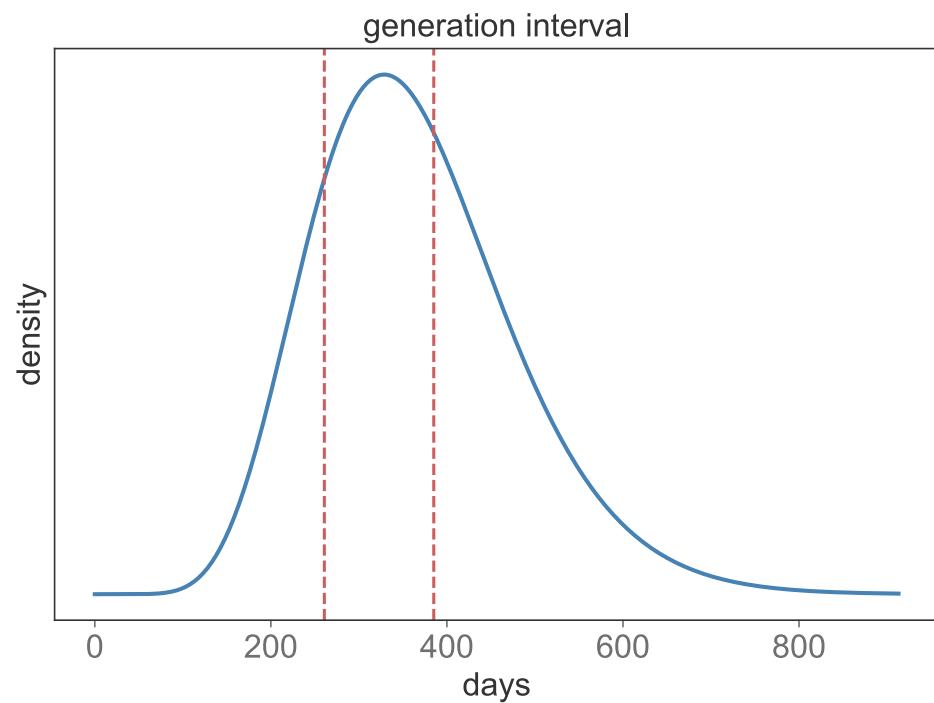
○ infection ● sampled



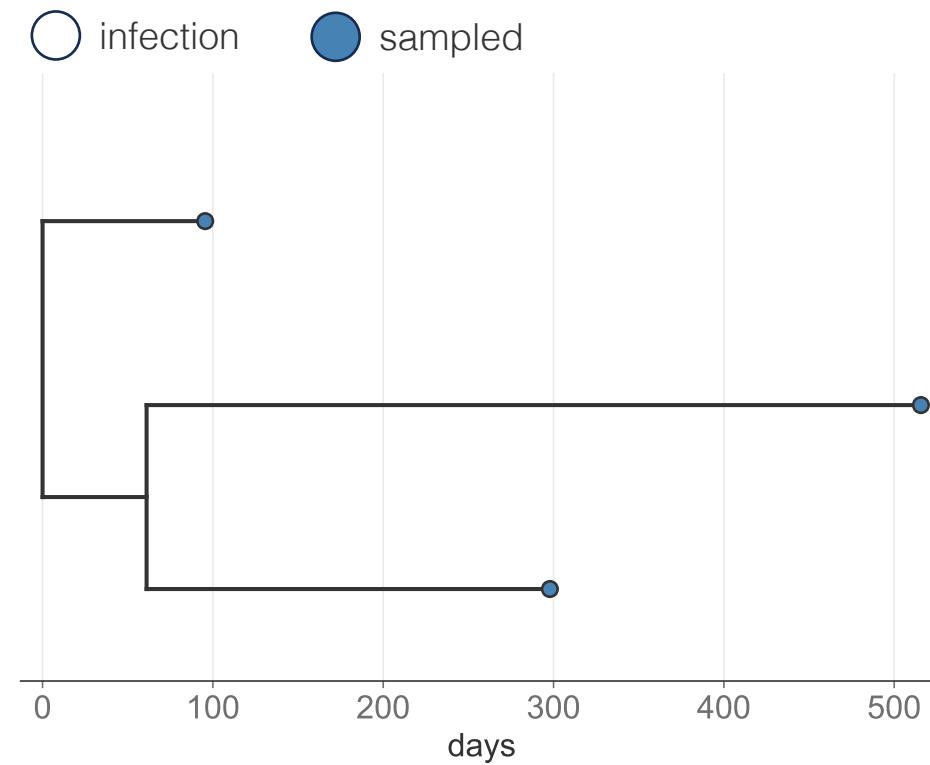
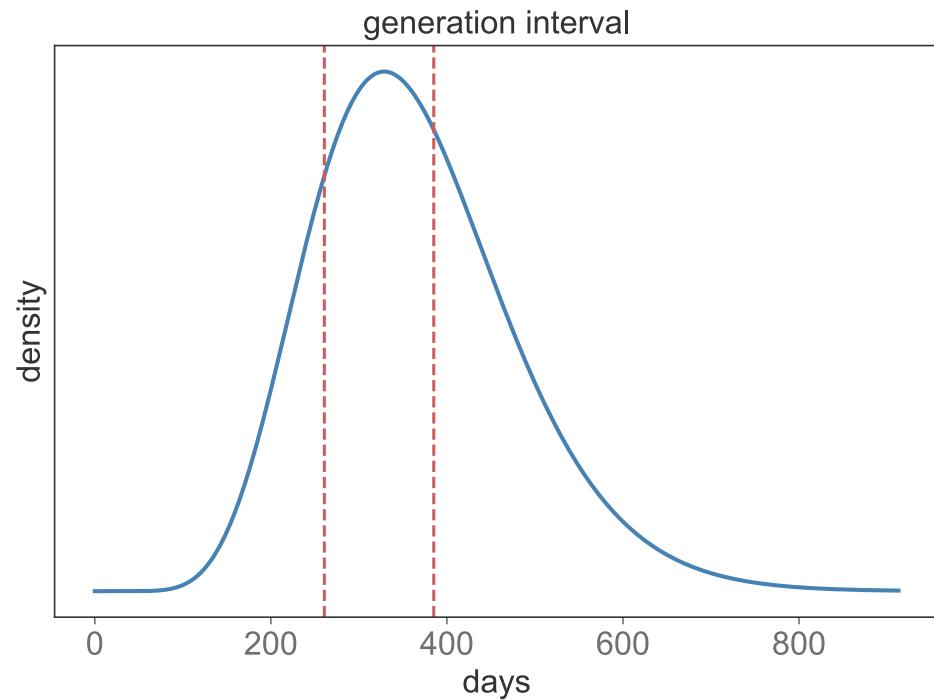
Infectious disease phylodynamics is possible because of (relatively) high substitution rates



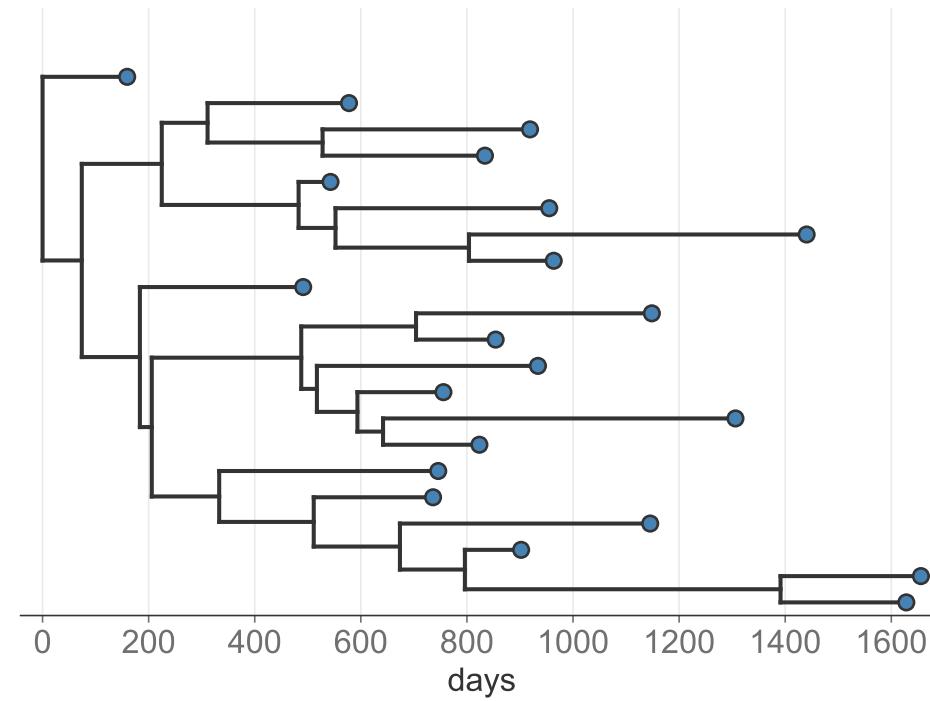
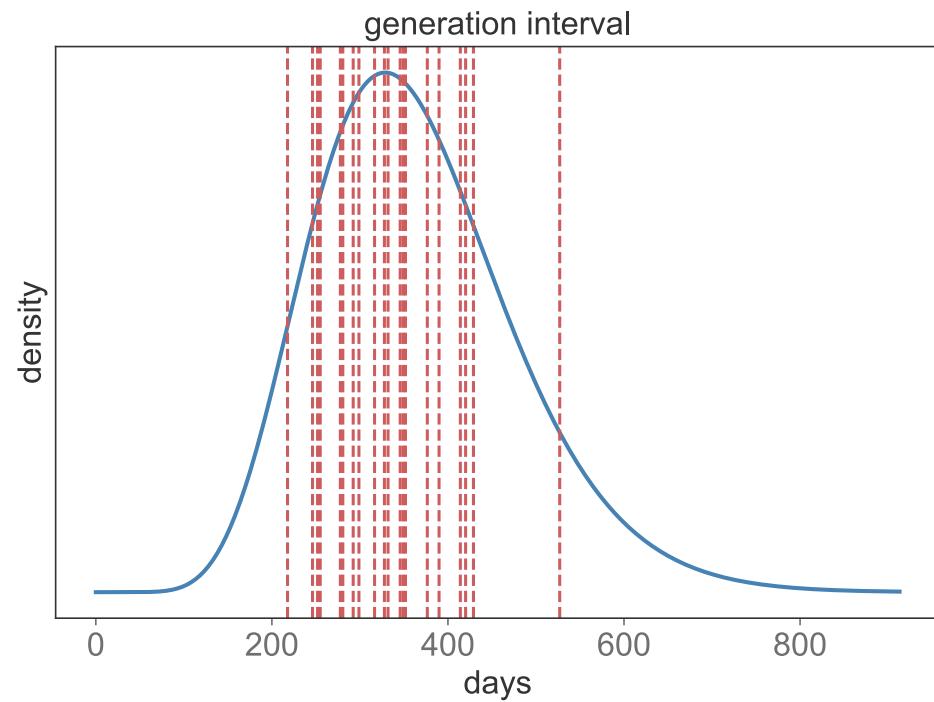
Infectious disease phylodynamics is possible because of (relatively) high substitution rates



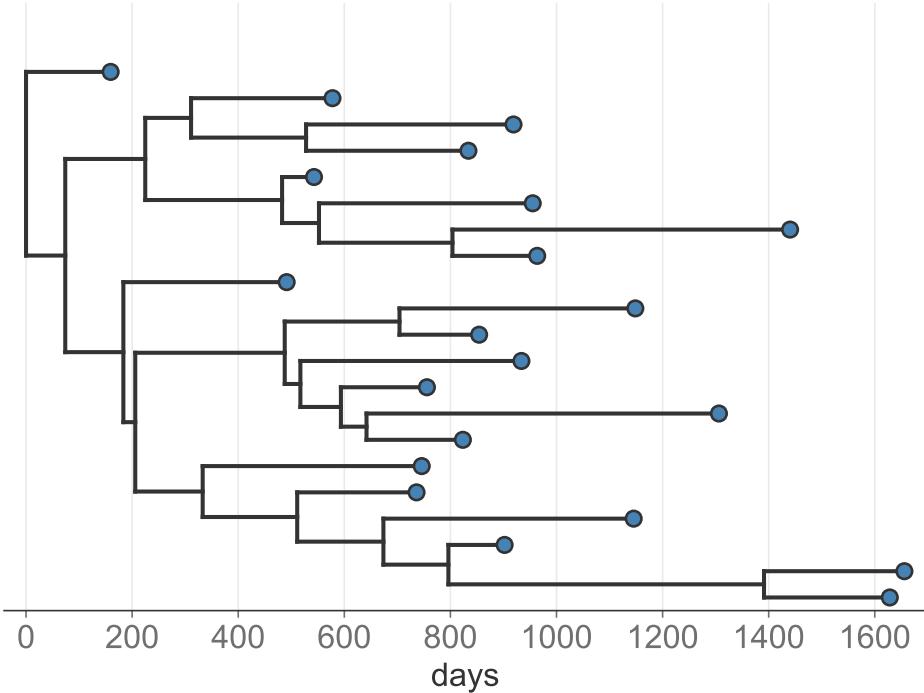
Infectious disease phylodynamics is possible because of (relatively) high substitution rates



Infectious disease phylodynamics is possible because of (relatively) high substitution rates

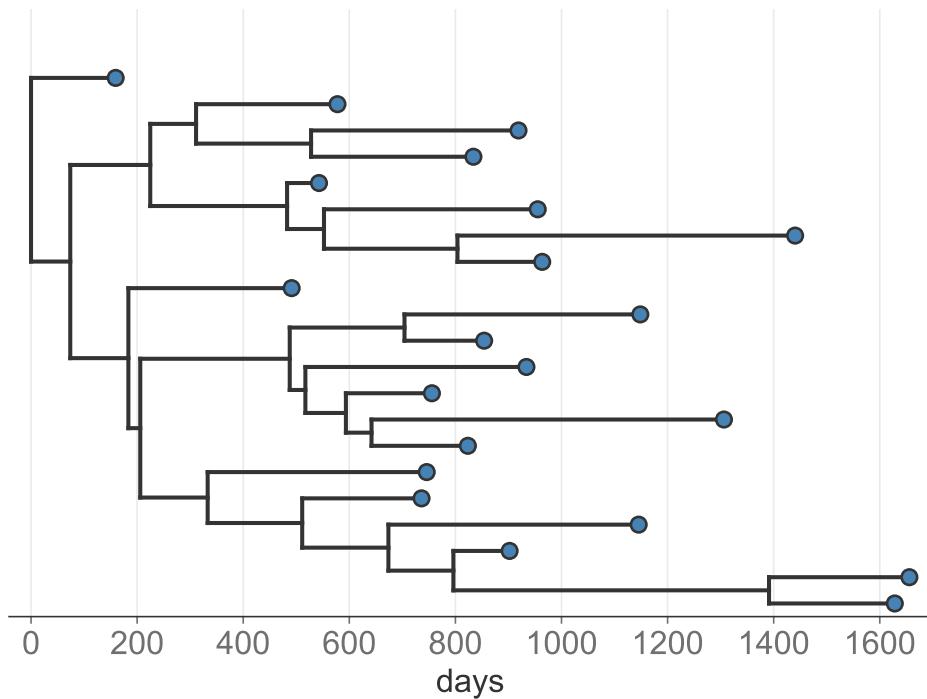


Infectious disease phylodynamics is possible because of (relatively) high substitution rates

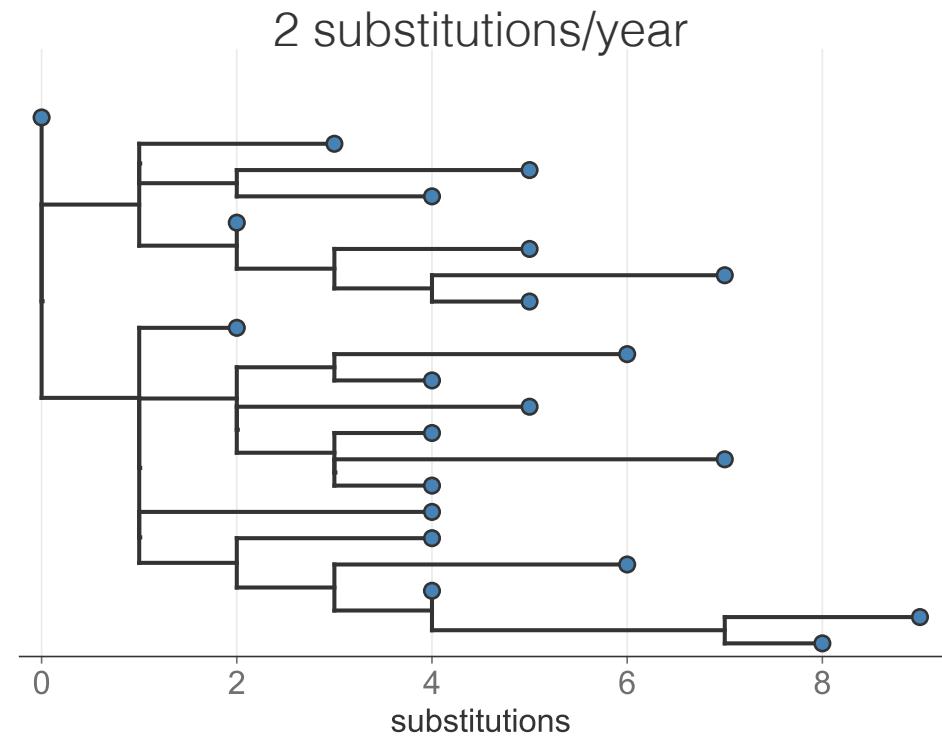


substitutions = calendar time \times substitution rate

Infectious disease phylodynamics is possible because of (relatively) high substitution rates

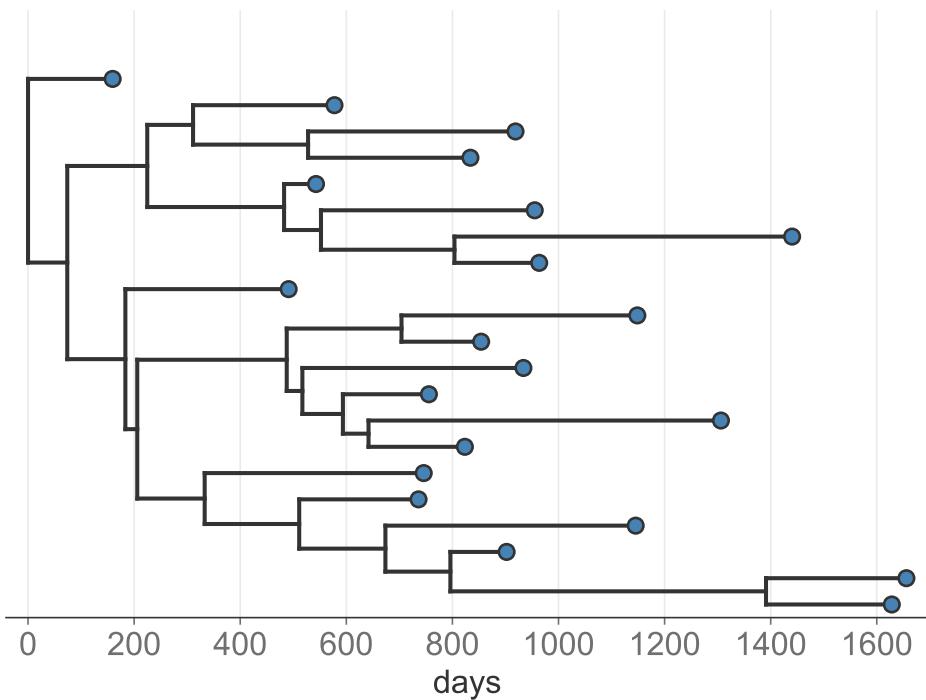


substitutions = calendar time \times substitution rate

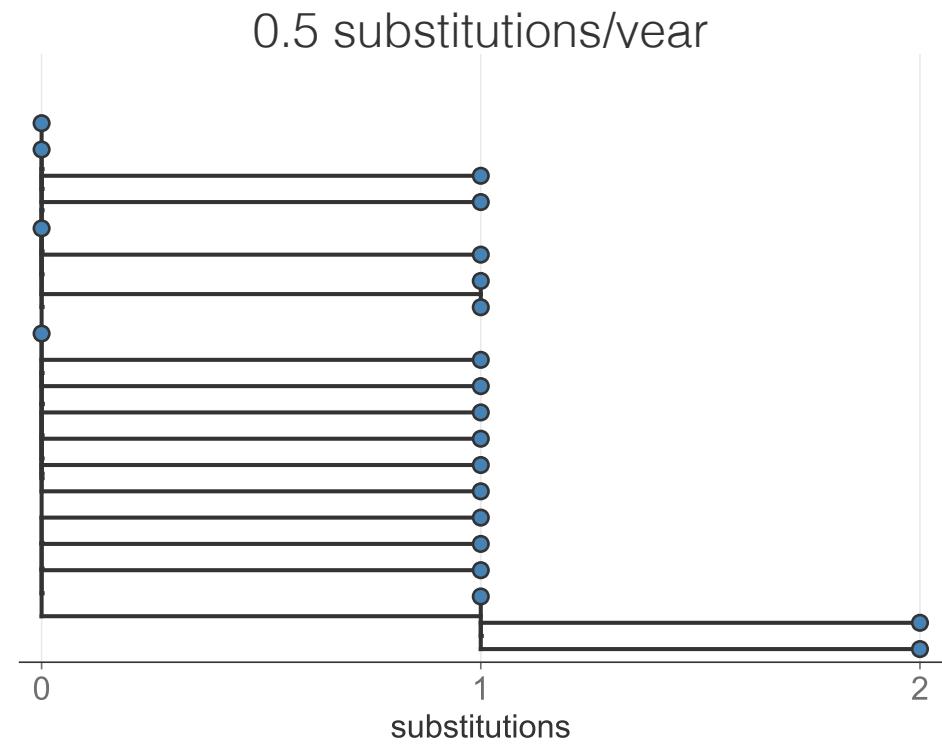


When the substitution rate is sufficiently high, genetic diversity is generated during an outbreak

Infectious disease phylodynamics is possible because of (relatively) high substitution rates



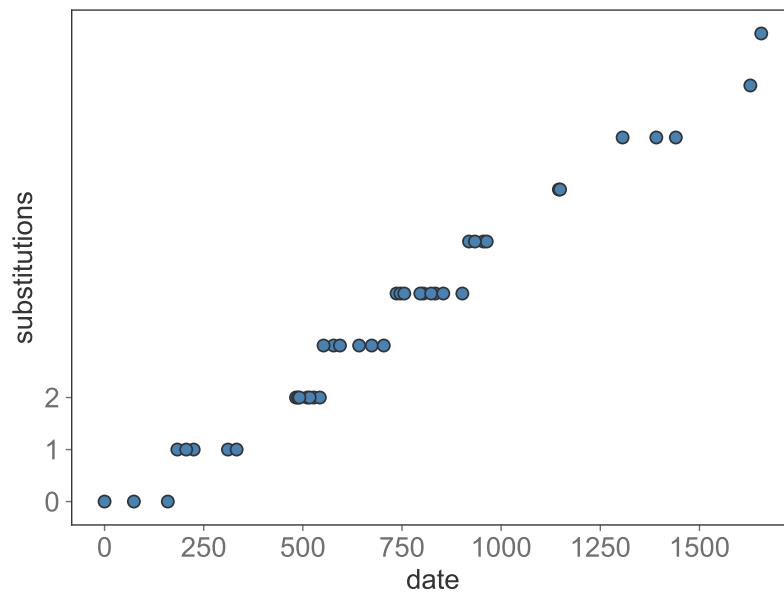
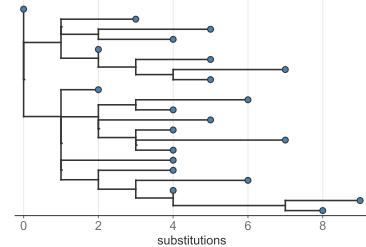
substitutions = calendar time \times substitution rate



When the substitution rate is low,
limited genetic diversity is generated
over the time scale of transmission

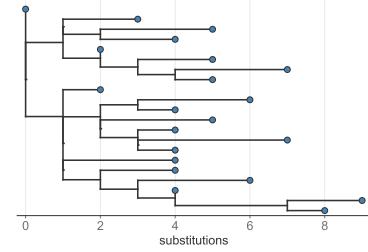
Infectious disease phylodynamics is possible because of (relatively) high substitution rates

2 substitutions/year

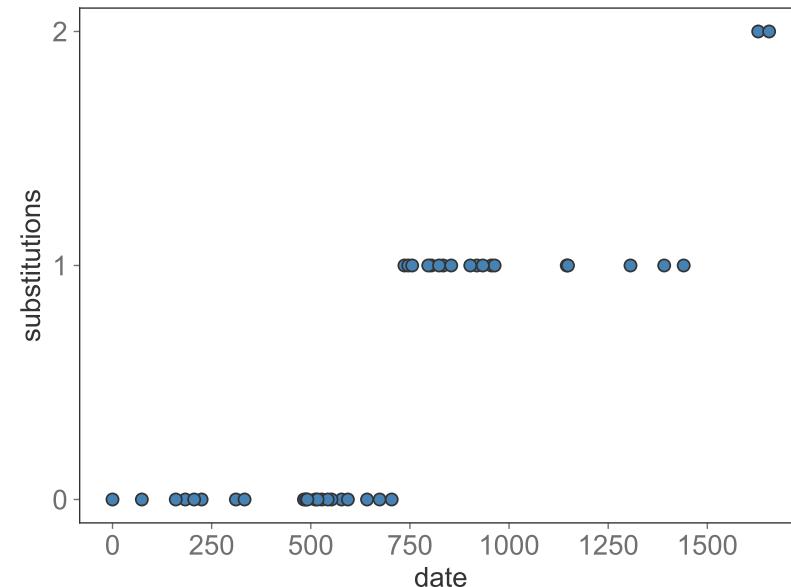
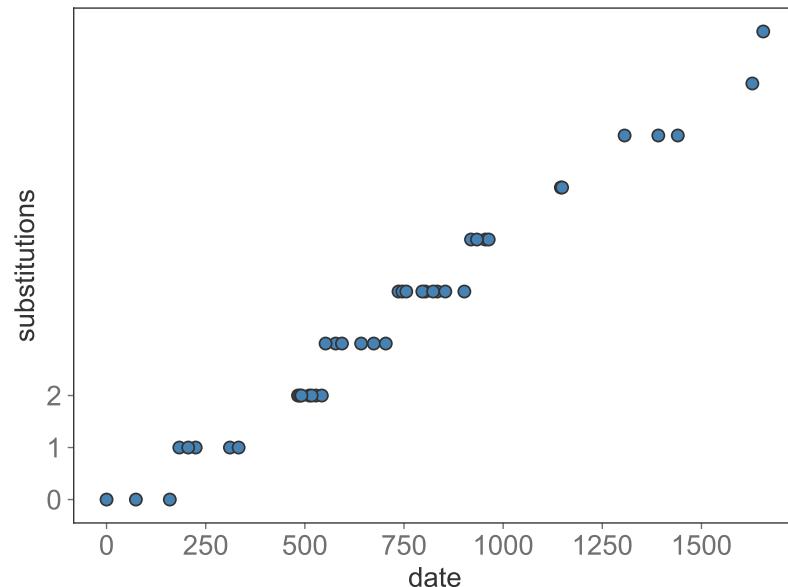
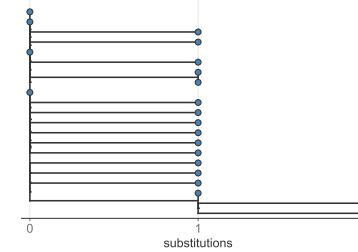


Infectious disease phylodynamics is possible because of (relatively) high substitution rates

2 substitutions/year



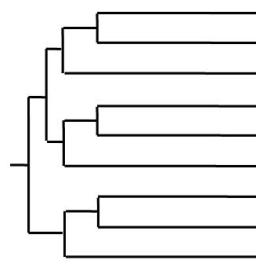
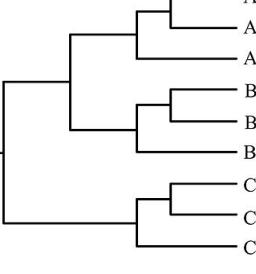
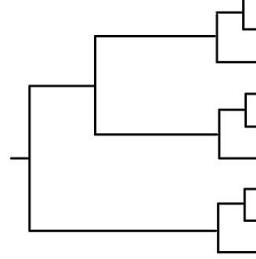
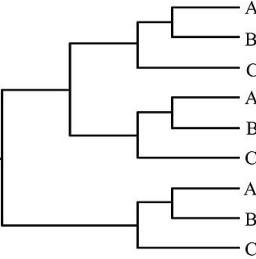
0.5 substitutions/year

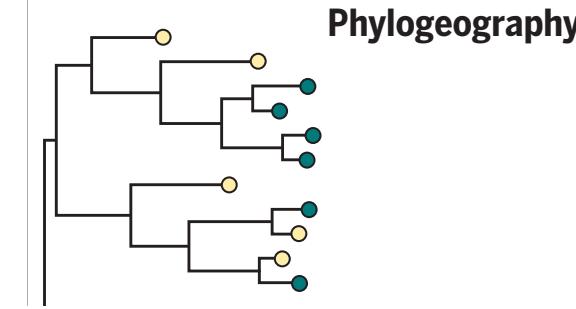


Identifying epidemiological dynamics using patterns of neutral evolution

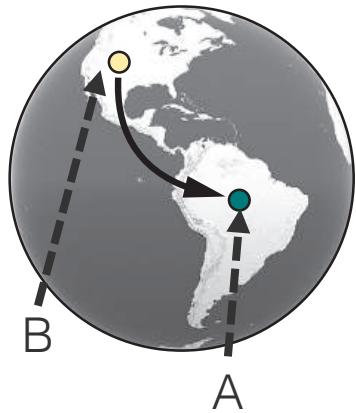
	Continual Immune Selection	Weak or Absent Immune Selection	
		Tree shape controlled by non-selective population dynamic processes	
Idealized Phylogeny Shapes	<p>Time →</p>	Population size dynamics Strong spatial structure Weak spatial structure 	
Examples	Human influenza A virus intra-host HIV	inter-host HIV inter-host HCV	Measles, rabies inter-host HIV
Tree Inferences	Detection of antigenic escape mutations	Estimation of population growth rates	Estimation of population migration rates

Identifying epidemiological dynamics using patterns of neutral evolution

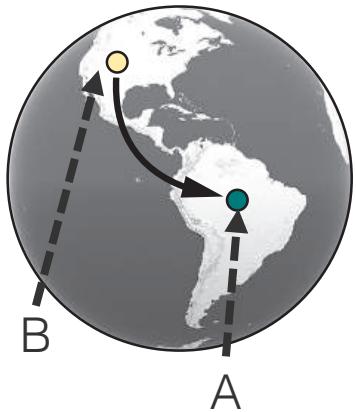
	Continual Immune Selection	Weak or Absent Immune Selection	
		Tree shape controlled by non-selective population dynamic processes	
Idealized Phylogeny Shapes		Population size dynamics	Spatial dynamics
		Exponential growth	Strong spatial structure
			
		Constant size	Weak spatial structure
			
Examples	Human influenza A virus intra-host HIV	inter-host HIV inter-host HCV	Measles, rabies inter-host HIV
Tree Inferences	Detection of antigenic escape mutations	Estimation of population growth rates	Estimation of population migration rates



Ancestral state reconstruction informs patterns of geographic spread

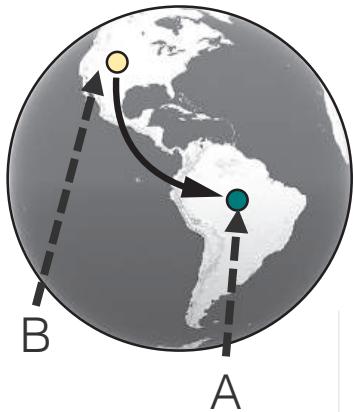


Ancestral state reconstruction informs patterns of geographic spread

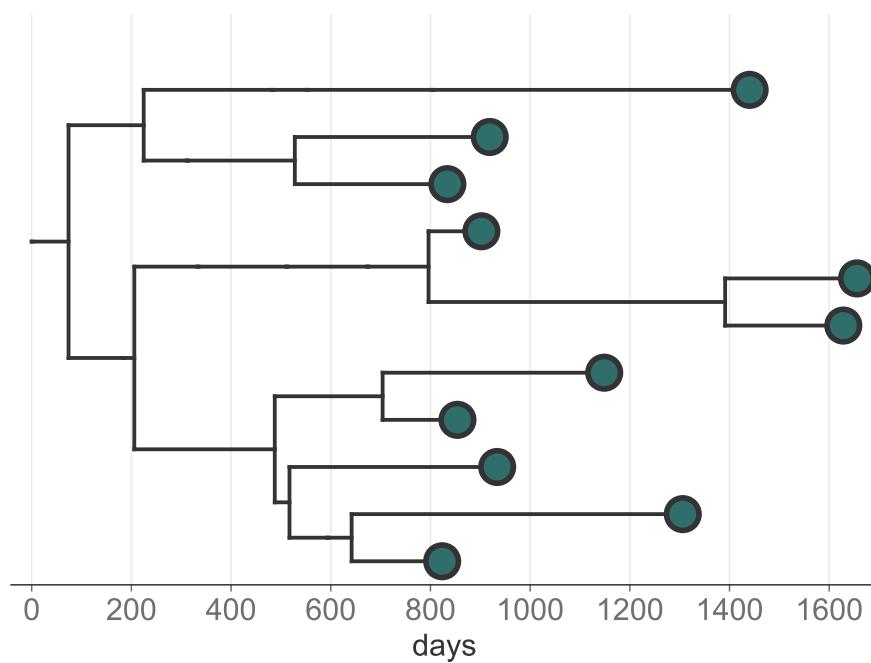


What is the relative contribution
of transmission within country A
v. importation from country B?

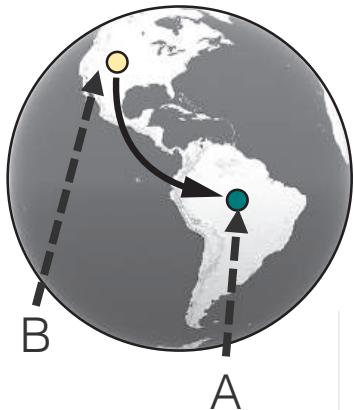
Ancestral state reconstruction informs patterns of geographic spread



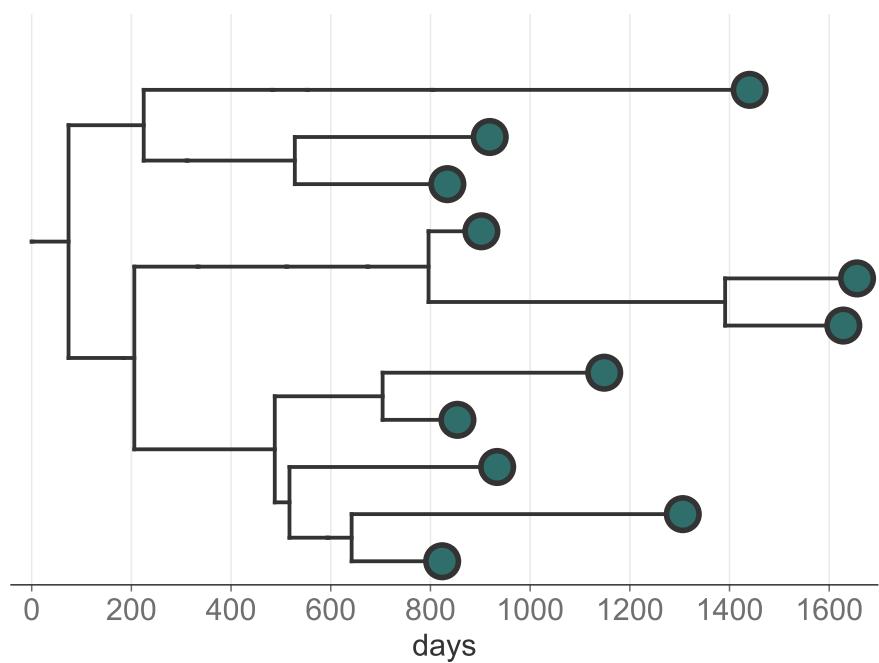
What is the relative contribution
of transmission within country A
v. importation from country B?



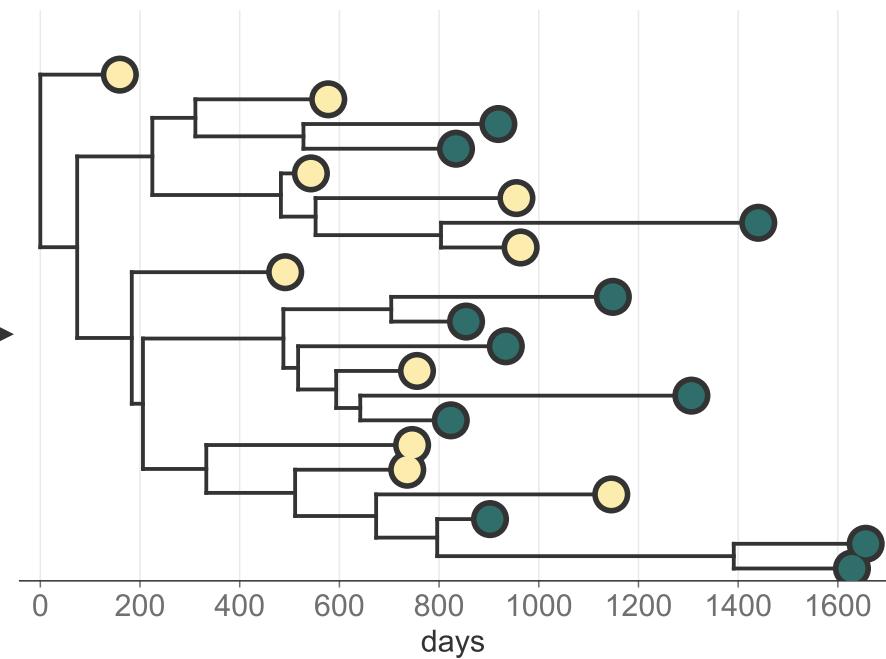
Ancestral state reconstruction informs patterns of geographic spread



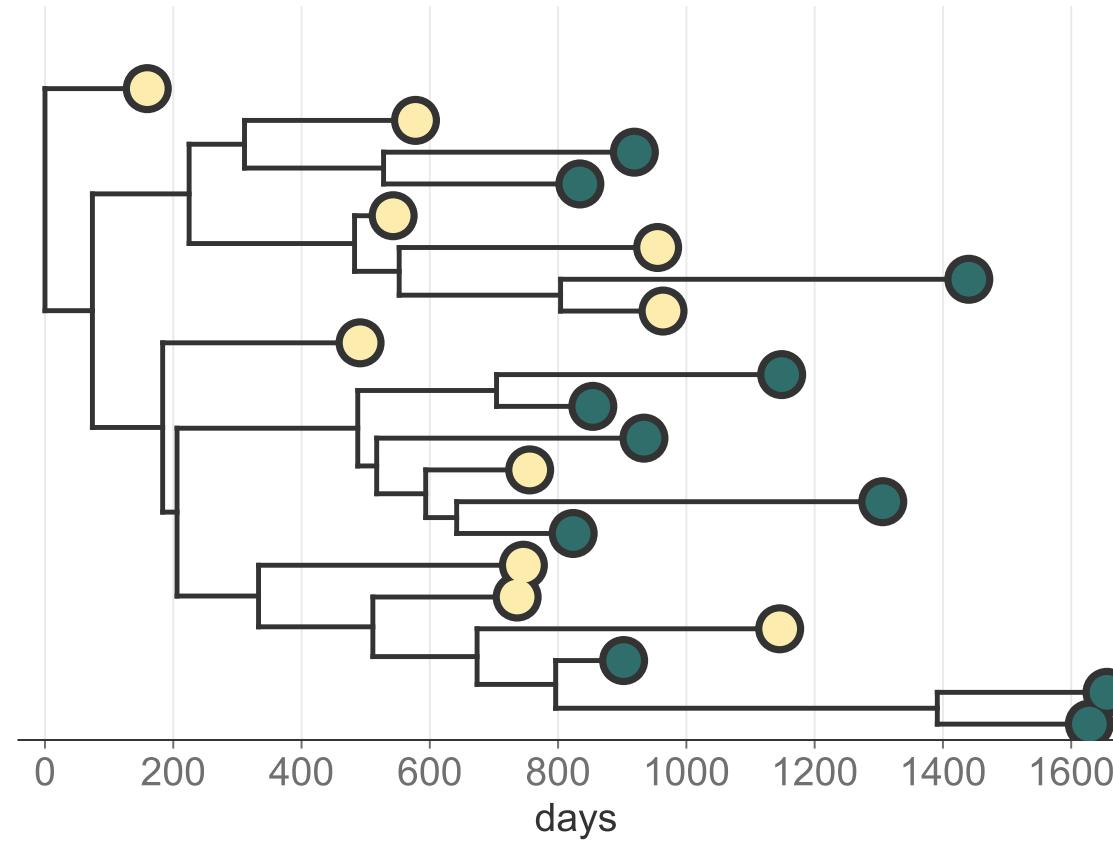
What is the relative contribution
of transmission within country A
v. importation from country B?



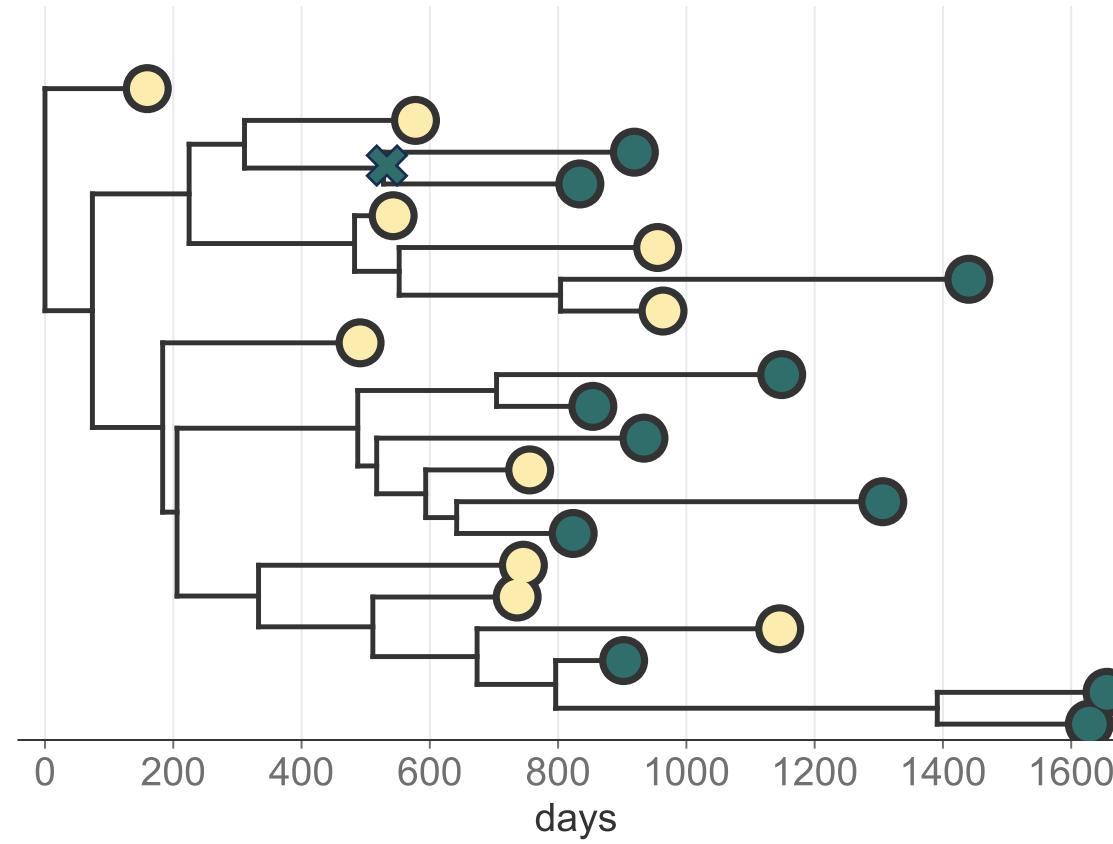
external
sampling



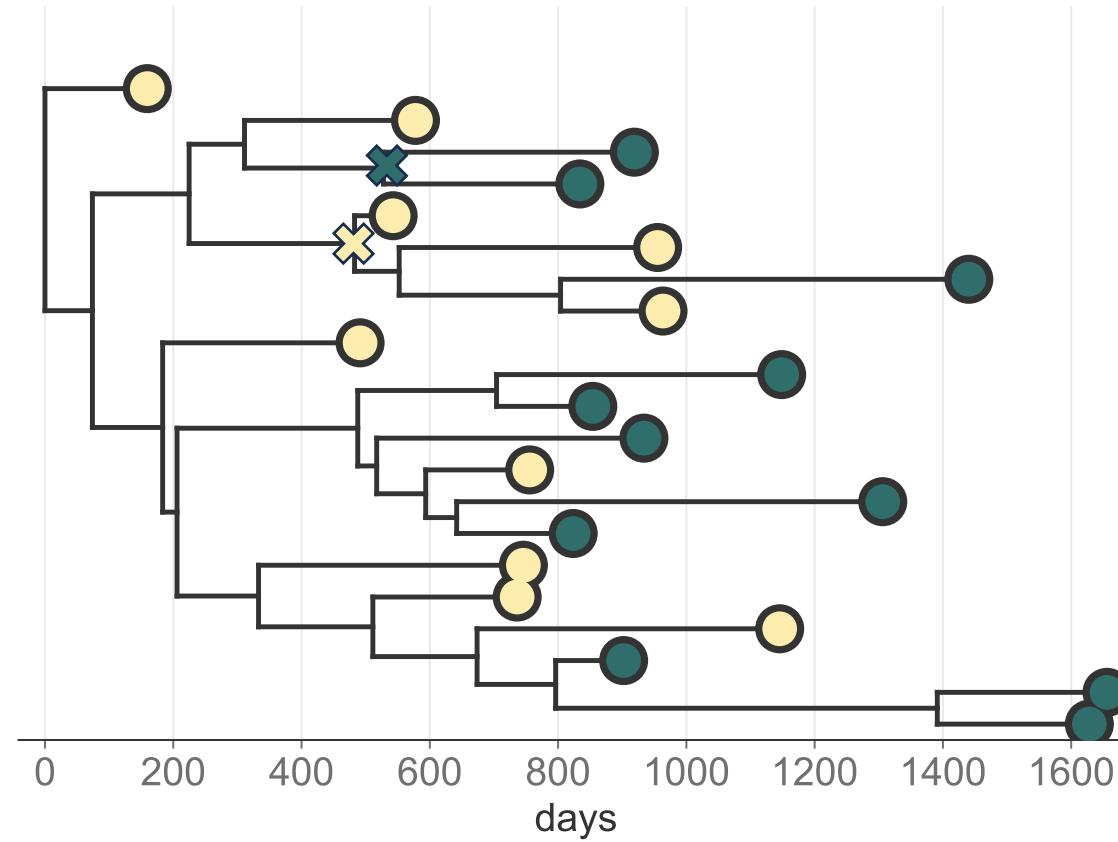
Ancestral state reconstruction informs patterns of geographic spread



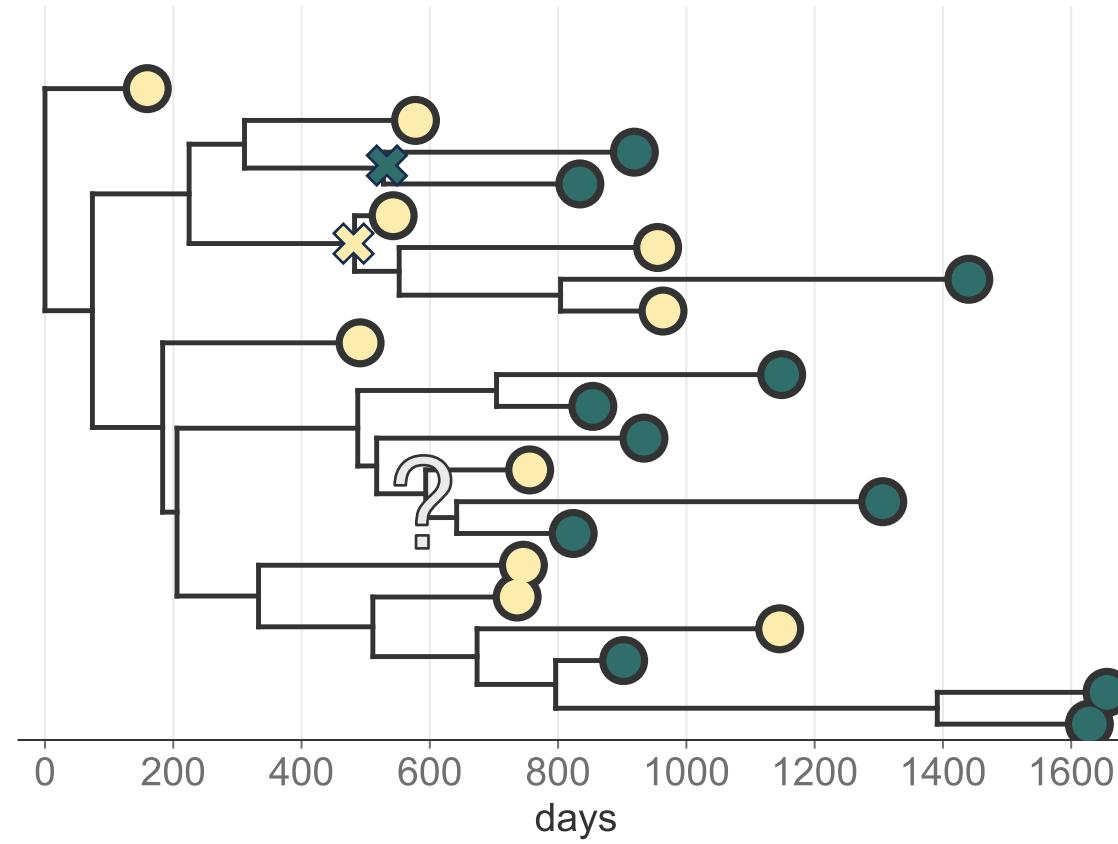
Ancestral state reconstruction informs patterns of geographic spread



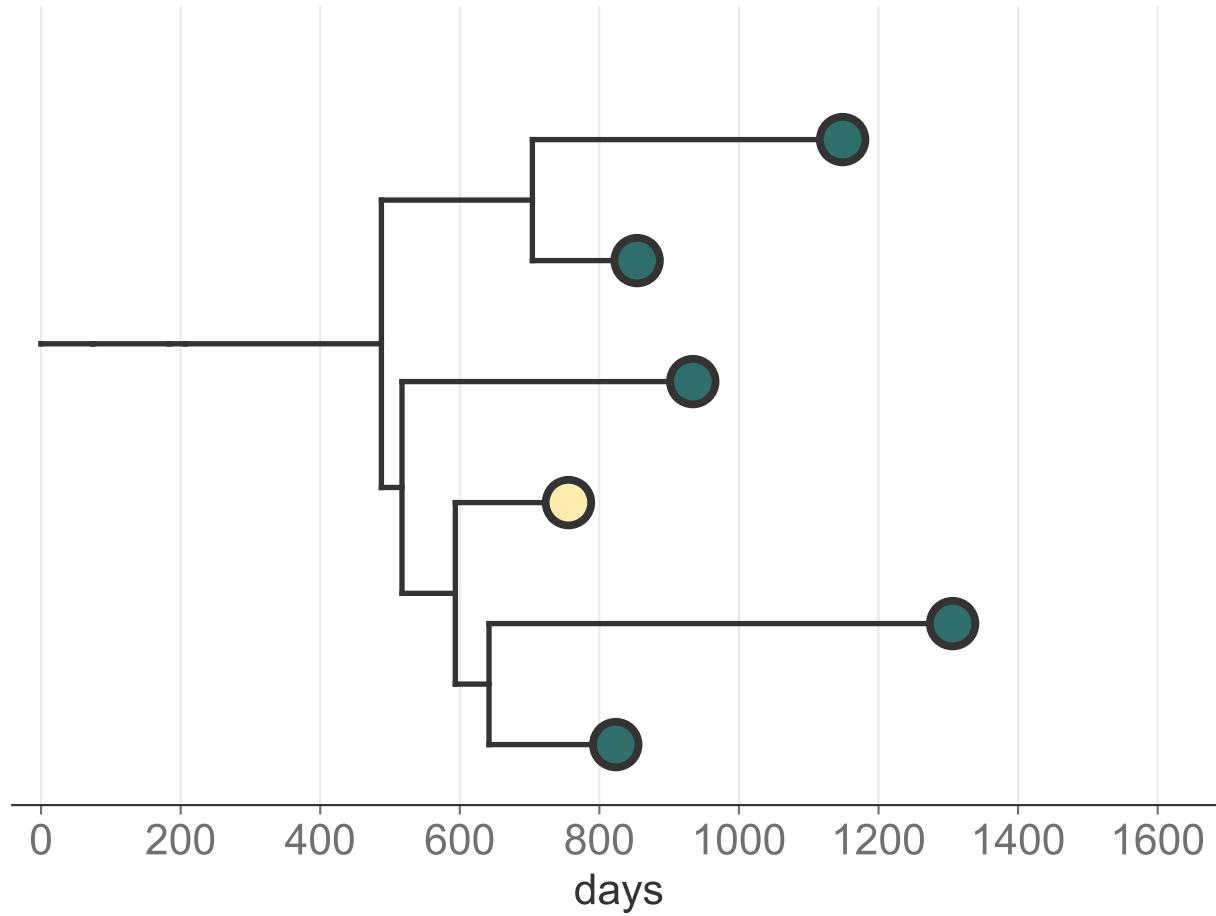
Ancestral state reconstruction informs patterns of geographic spread



Ancestral state reconstruction informs patterns of geographic spread



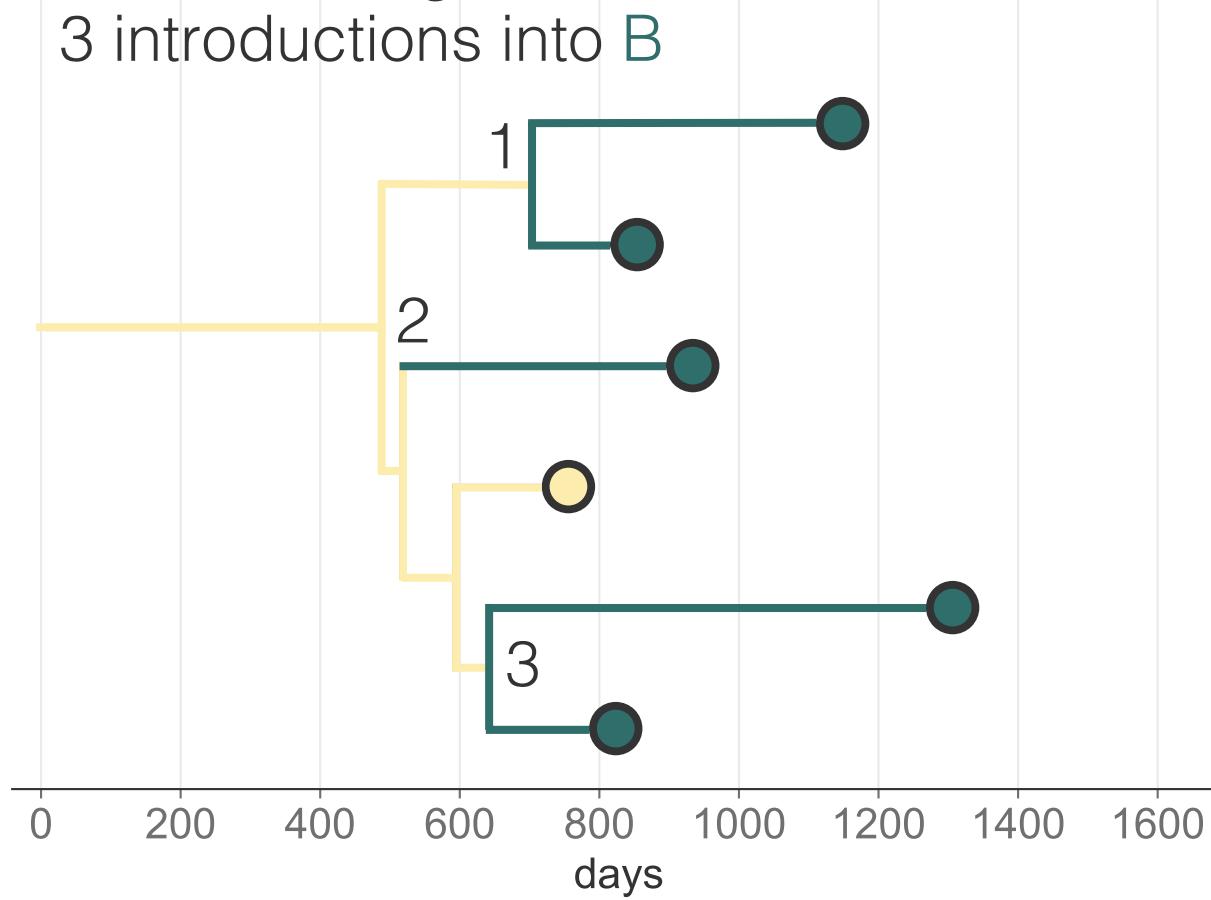
Maximum parsimony ancestral state reconstruction minimizes state changes



Maximum parsimony ancestral state reconstruction minimizes state changes

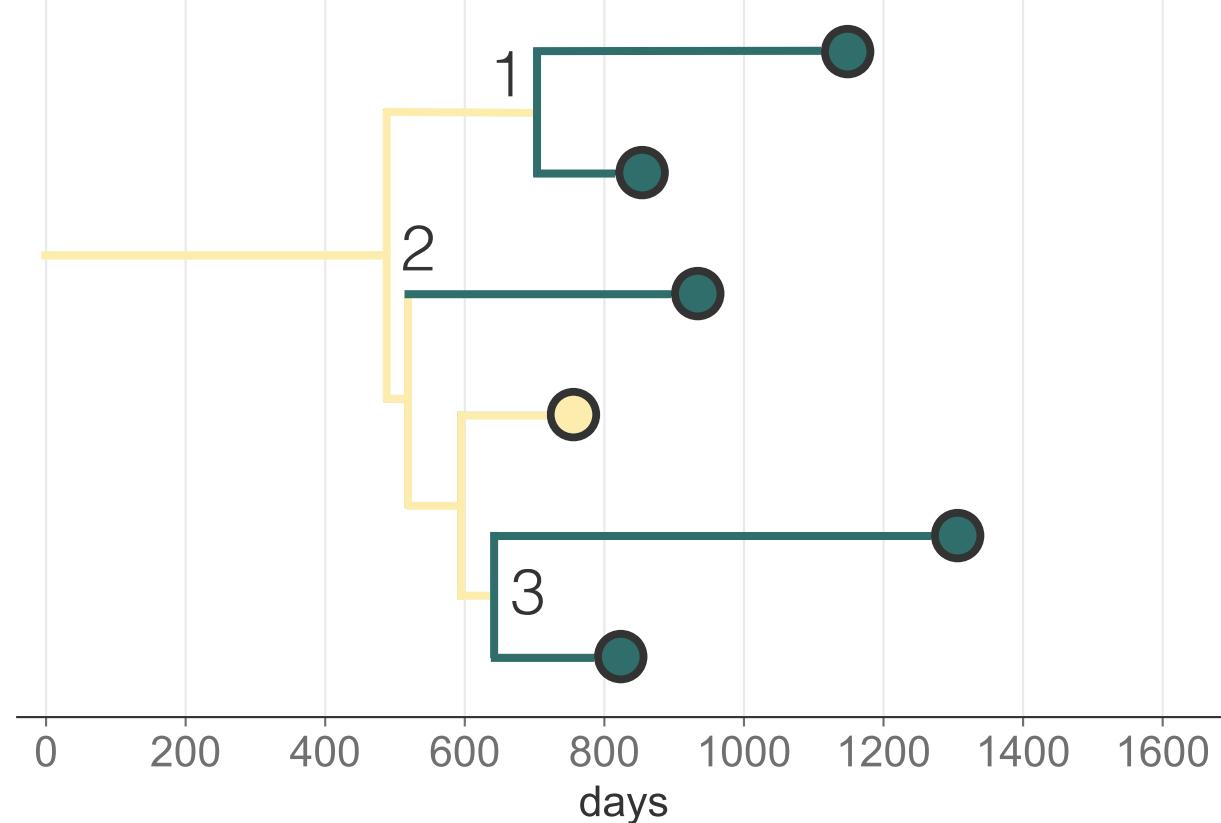
3 state changes

3 introductions into B

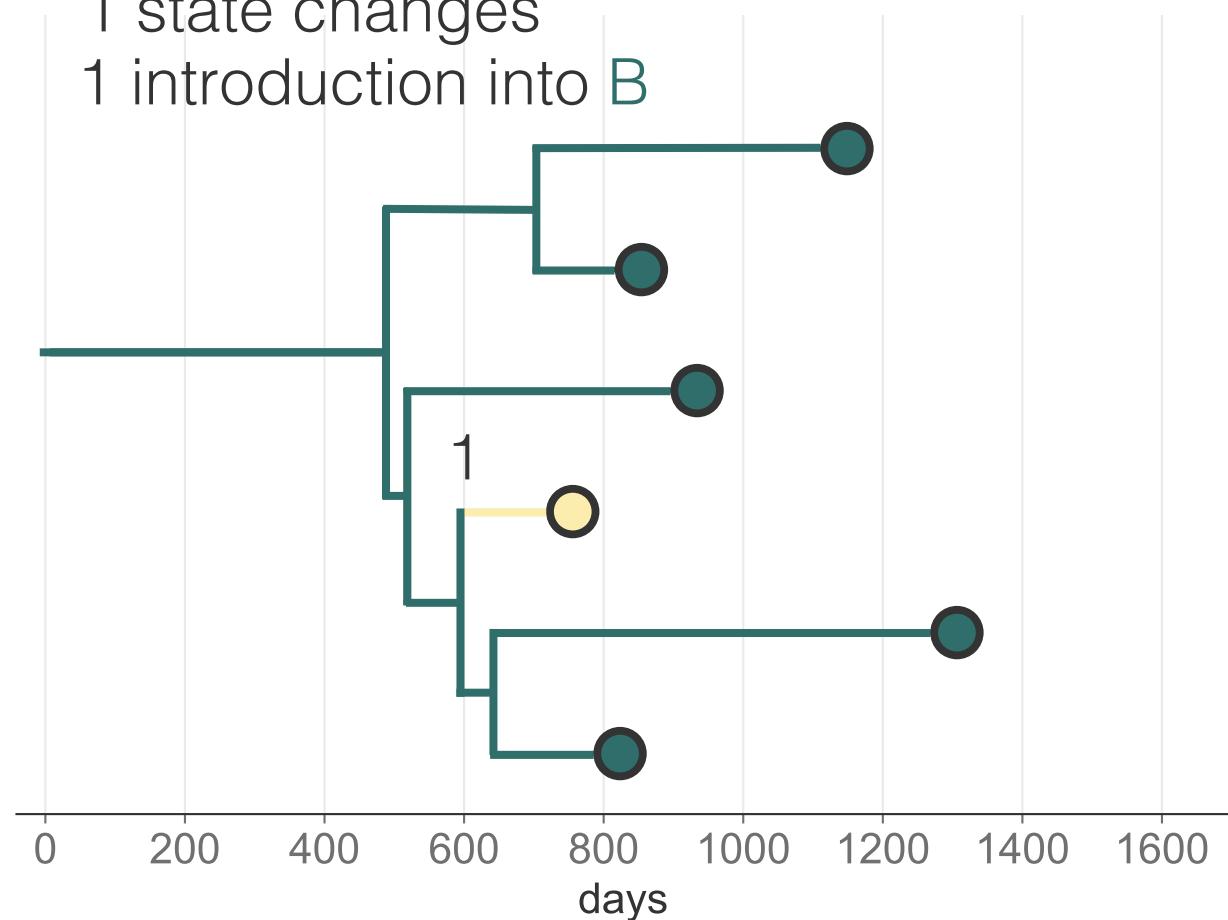


Maximum parsimony ancestral state reconstruction minimizes state changes

3 state changes
3 introductions into B

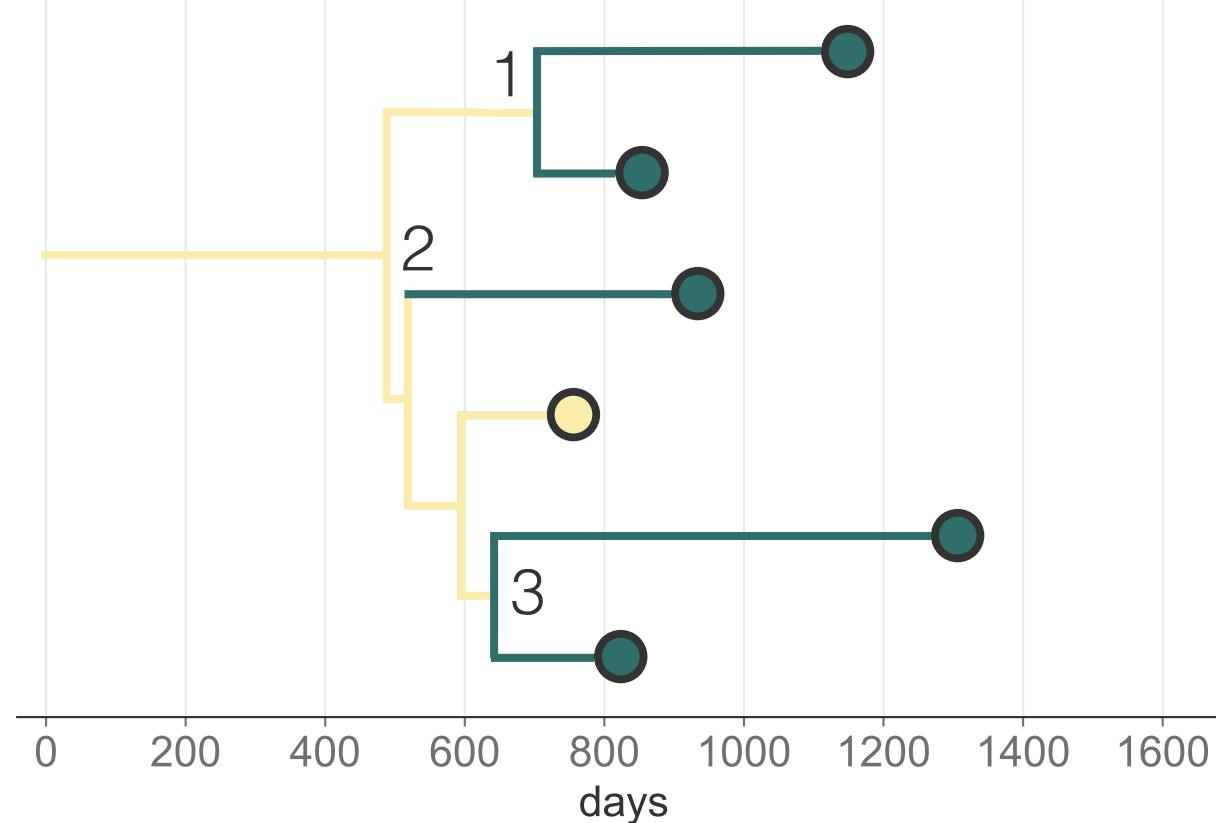


1 state changes
1 introduction into B

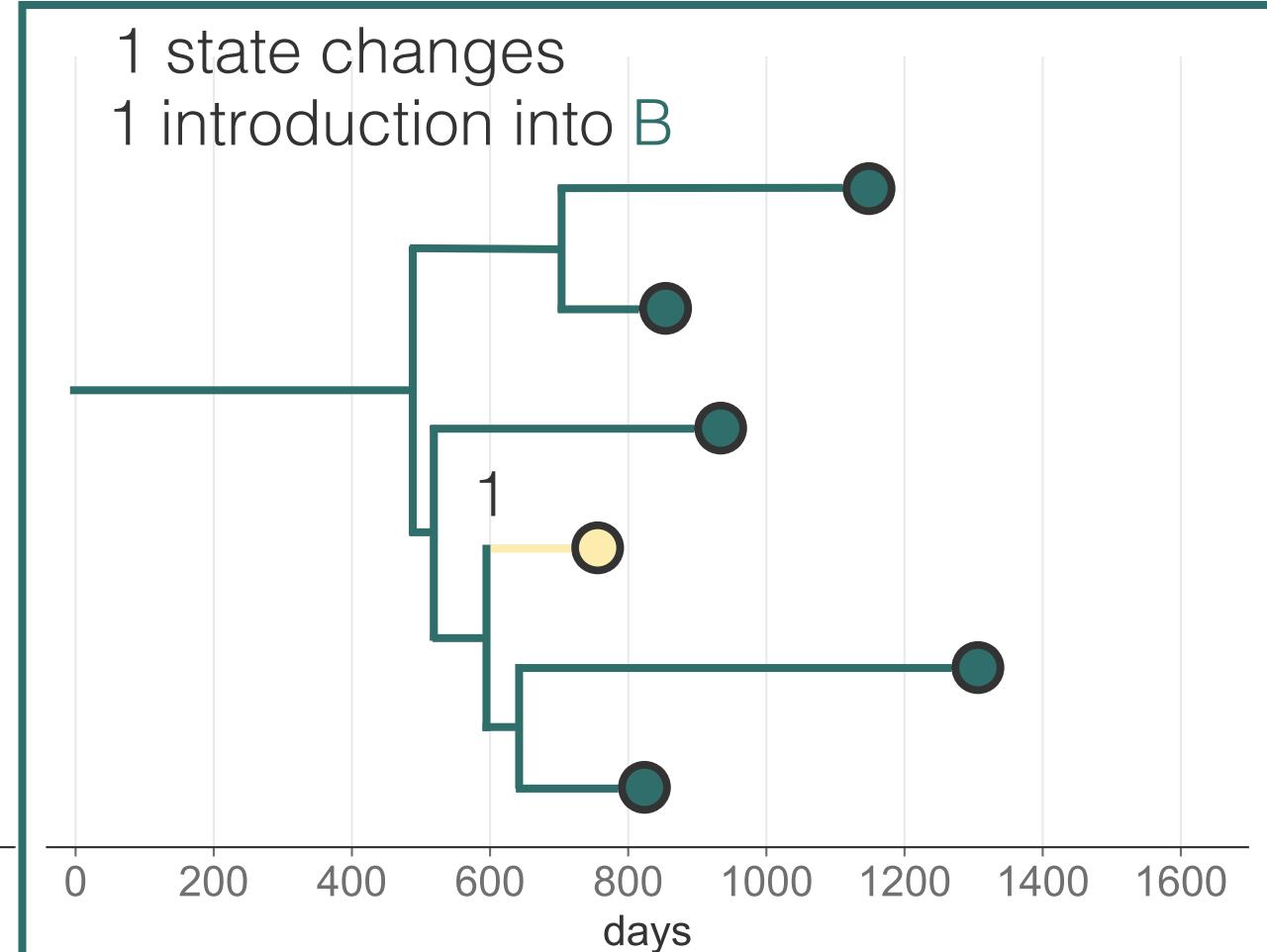


Maximum parsimony ancestral state reconstruction minimizes state changes

3 state changes
3 introductions into B

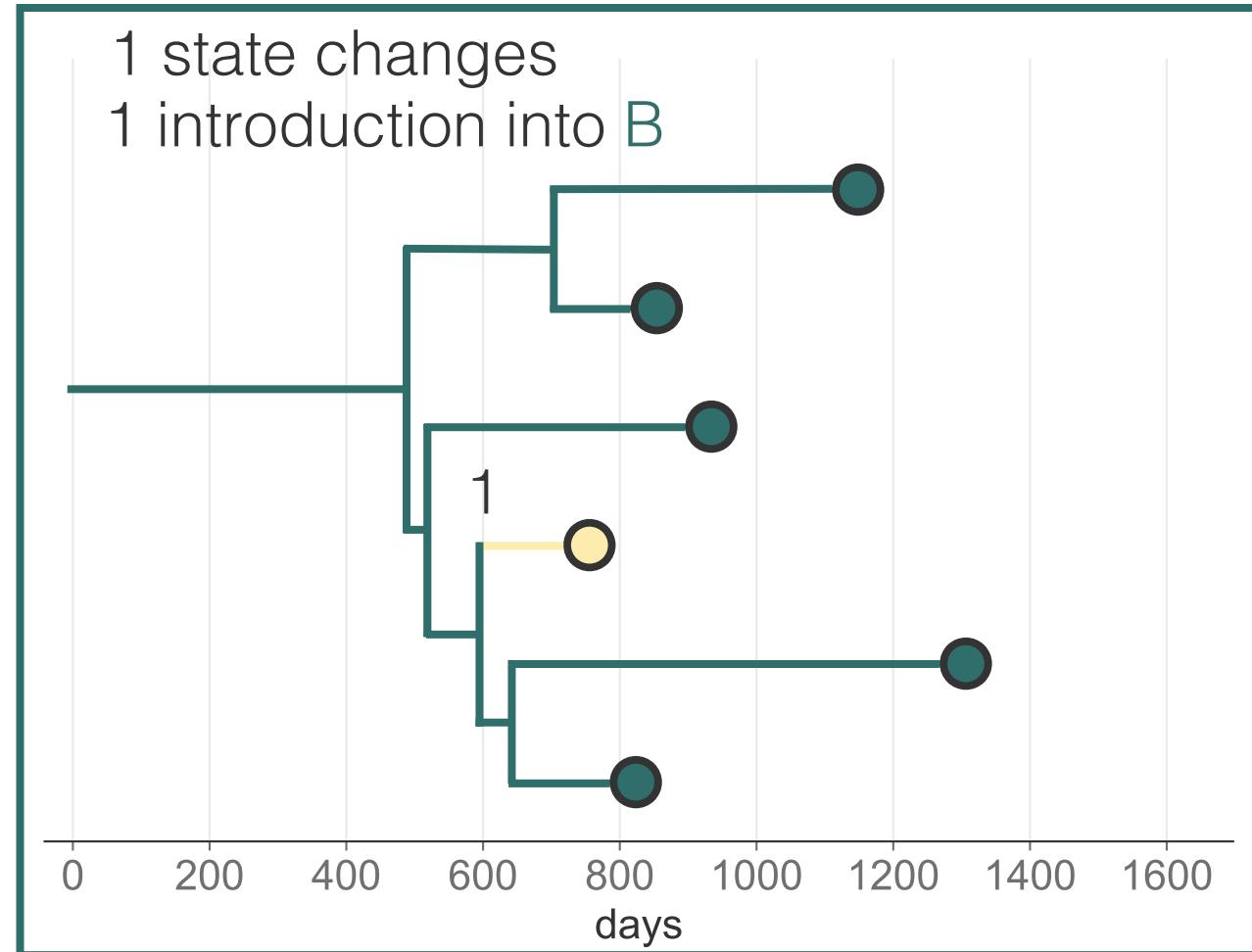


1 state changes
1 introduction into B



Maximum parsimony ancestral state reconstruction minimizes state changes

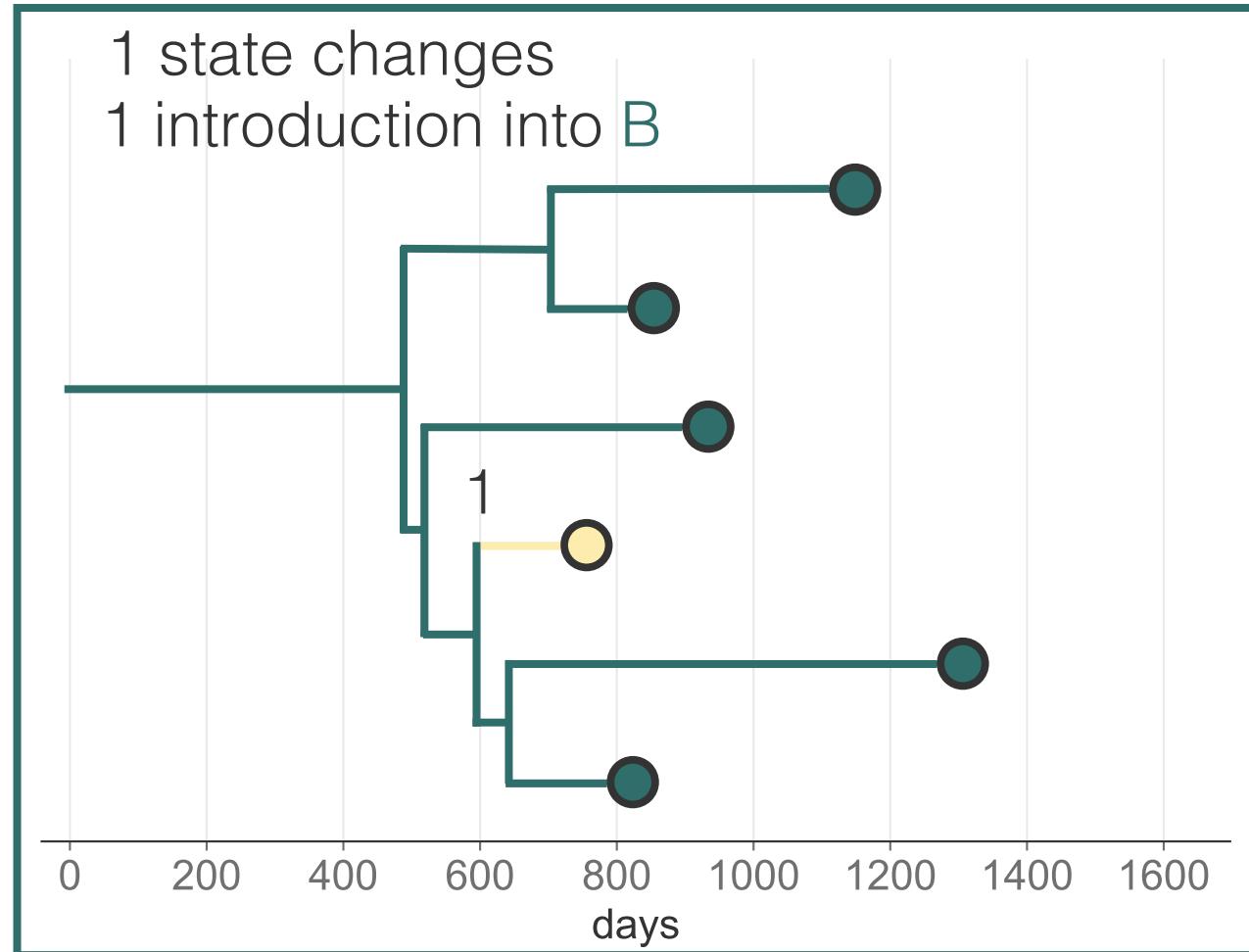
Because state changes always occur at nodes, parsimony-based ancestral state reconstruction implicitly assumes all branch lengths are equal.



Maximum parsimony ancestral state reconstruction minimizes state changes

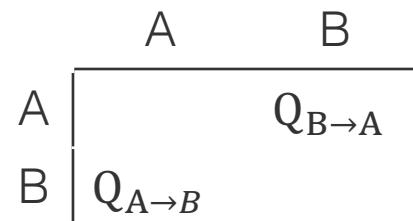
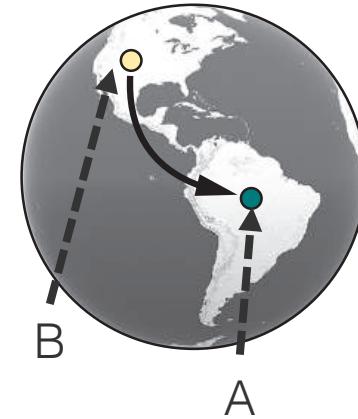
Because state changes always occur at nodes, parsimony-based ancestral state reconstruction implicitly assumes all branch lengths are equal.

This is rarely true.



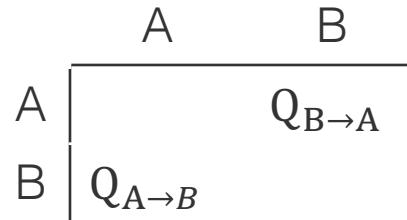
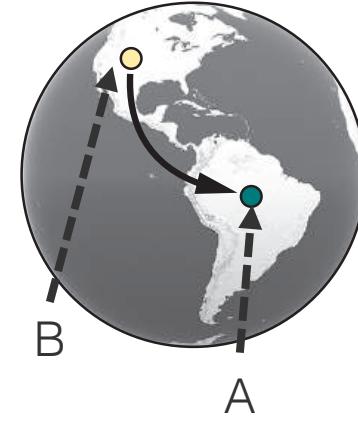
Ancestral state reconstruction through “mugration” models

- Models transition between discrete states analogously to transition between nucleotides (continuous time Markov chain)



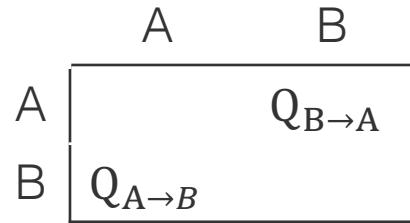
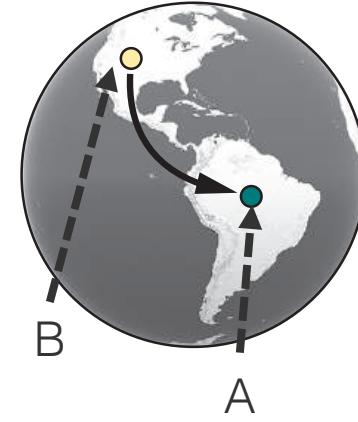
Ancestral state reconstruction through “mugration” models

- Models transition between discrete states analogously to transition between nucleotides (continuous time Markov chain)
- Allows for the incorporation of prior information about transition rates



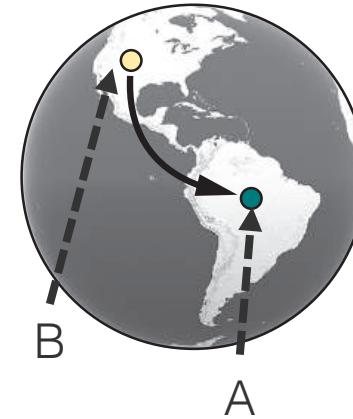
Ancestral state reconstruction through “mugration” models

- Models transition between discrete states analogously to transition between nucleotides (continuous time Markov chain)
- Allows for the incorporation of prior information about transition rates
- Allows for the inference of transition rates

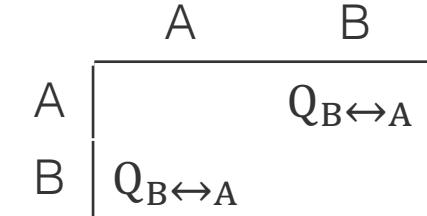
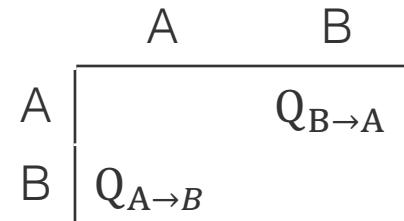


Ancestral state reconstruction through “mugration” models

- Models transition between discrete states analogously to transition between nucleotides (continuous time Markov chain)
- Allows for the incorporation of prior information about transition rates
- Allows for the inference of transition rates

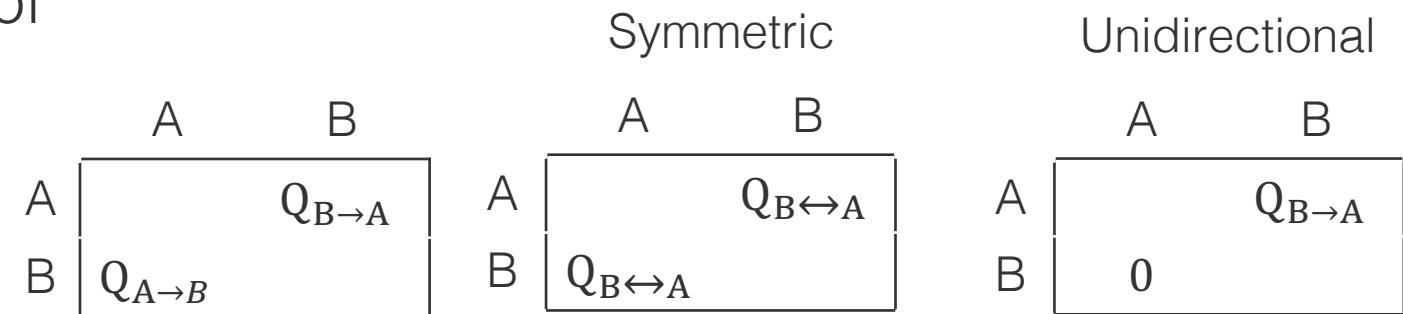
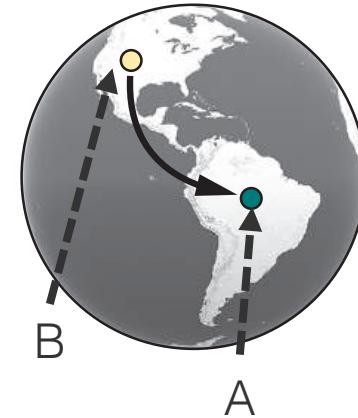


Symmetric

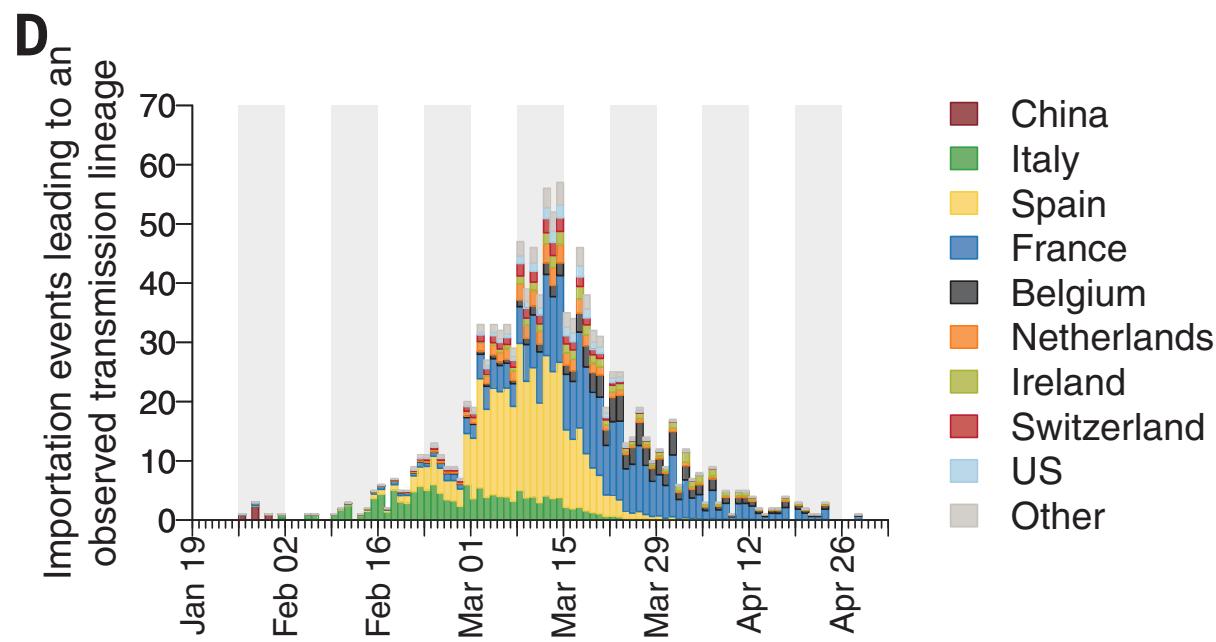


Ancestral state reconstruction through “mugration” models

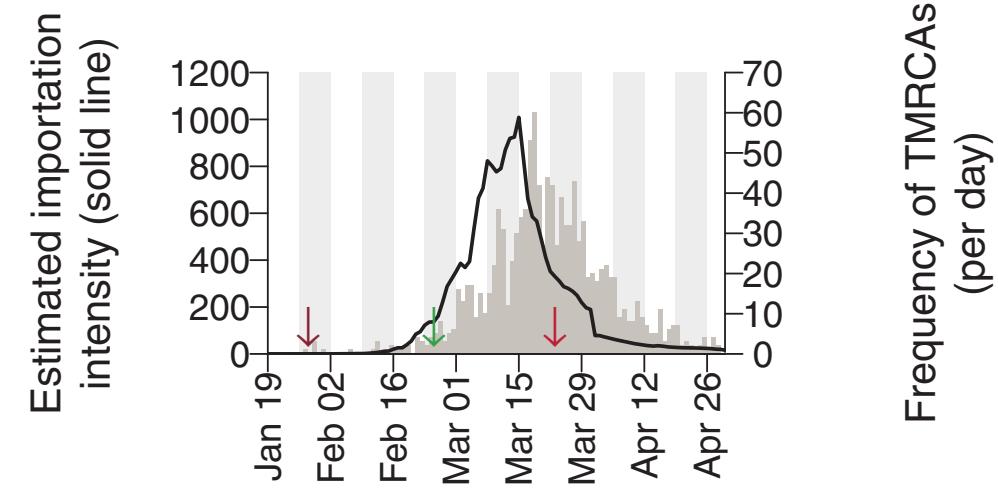
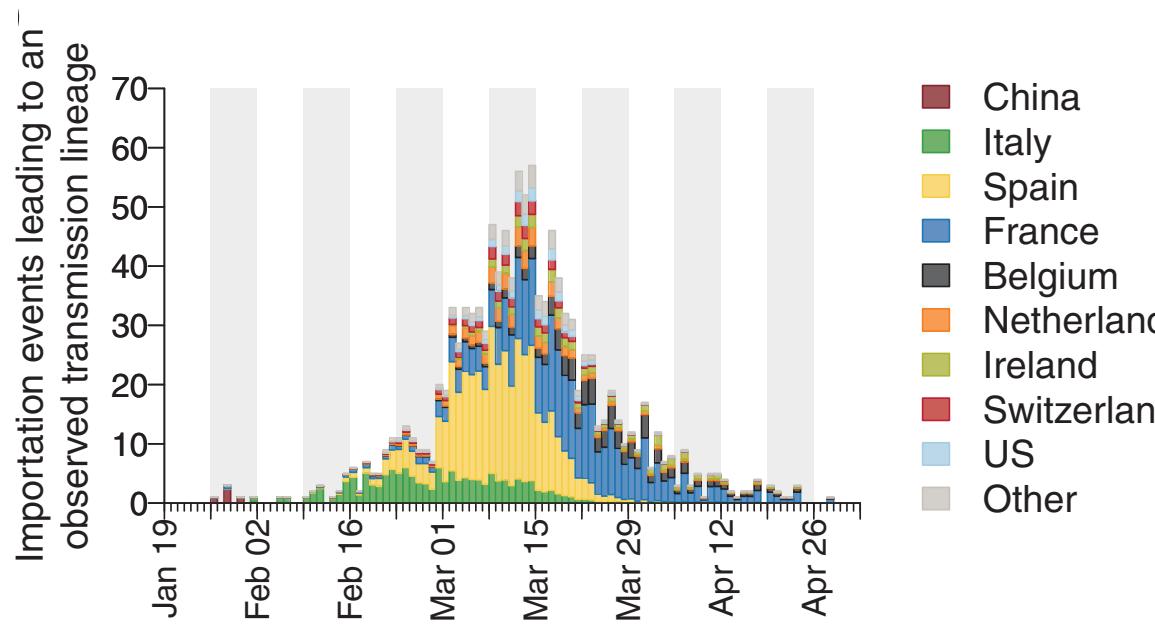
- Models transition between discrete states analogously to transition between nucleotides (continuous time Markov chain)
- Allows for the incorporation of prior information about transition rates
- Allows for the inference of transition rates



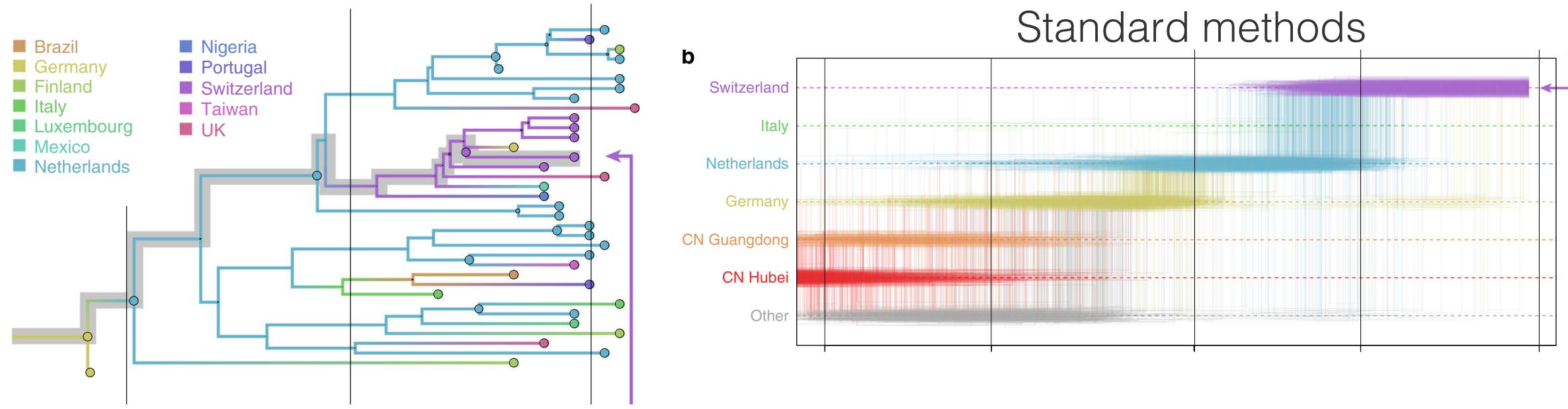
Establishment of SARS-CoV-2 in the U.K.



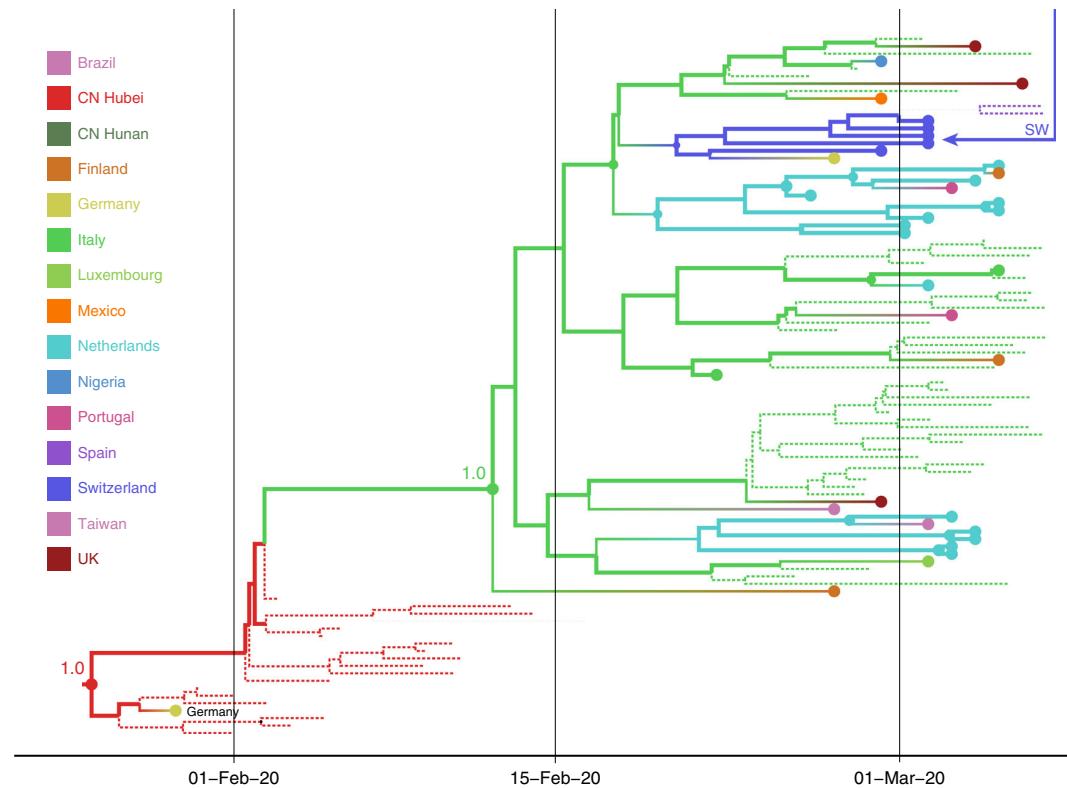
Establishment of SARS-CoV-2 in the U.K.



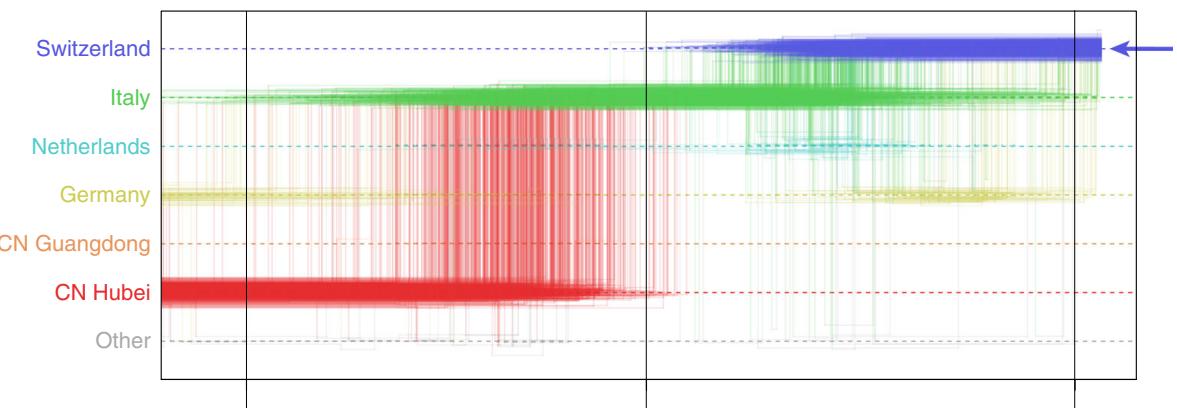
Incorporating unsampled taxa to account for sampling biases



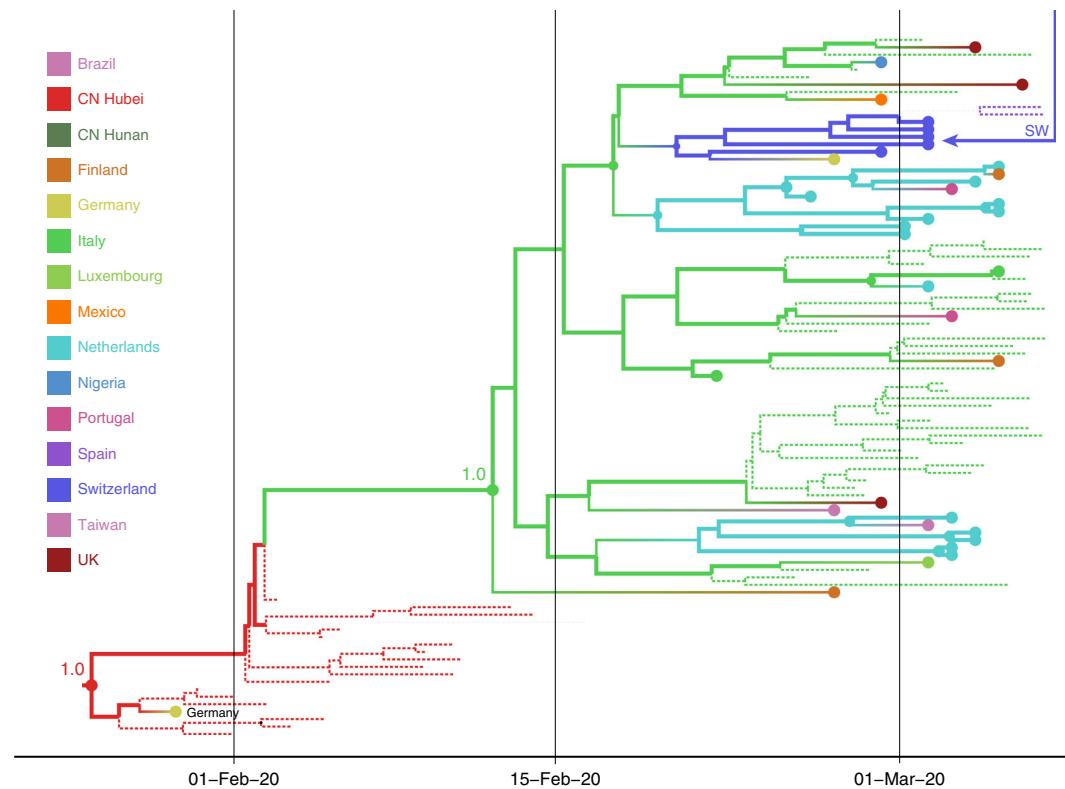
Incorporating unsampled taxa to account for sampling biases



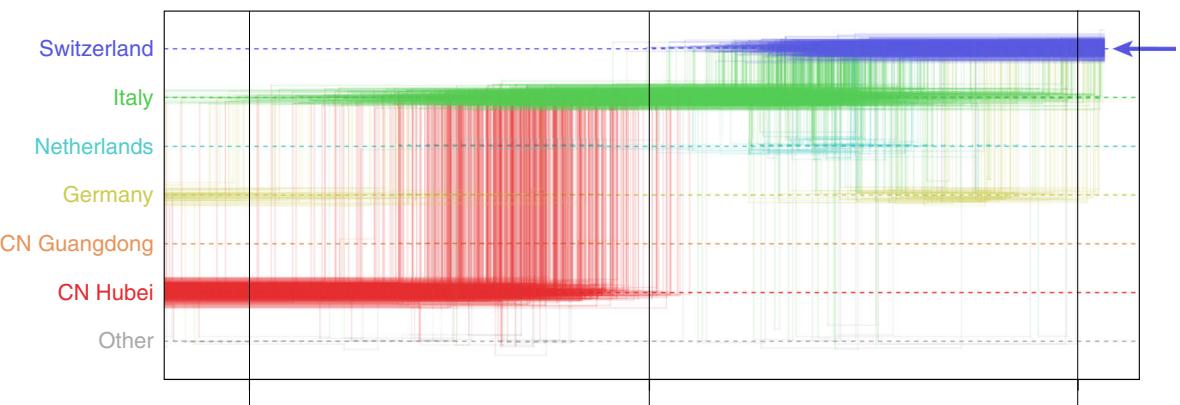
Incorporating unsampled taxa



Incorporating unsampled taxa to account for sampling biases

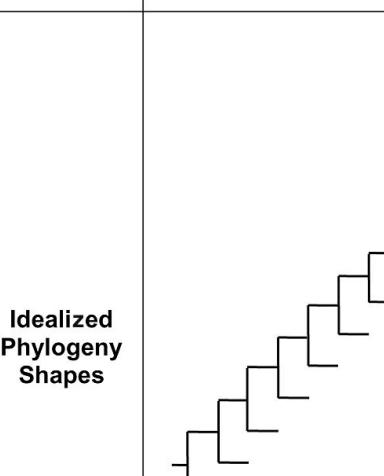
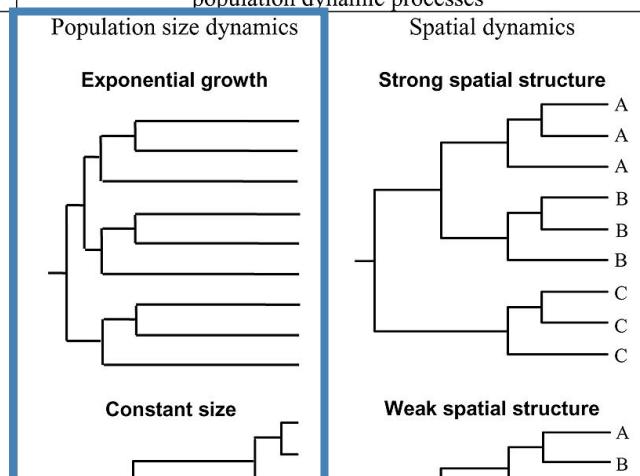


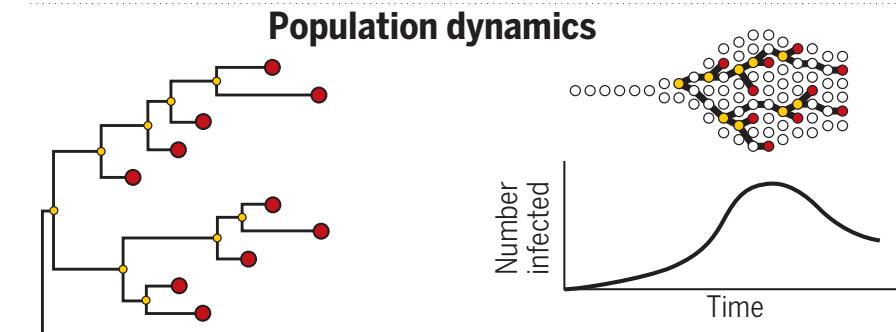
Incorporating unsampled taxa



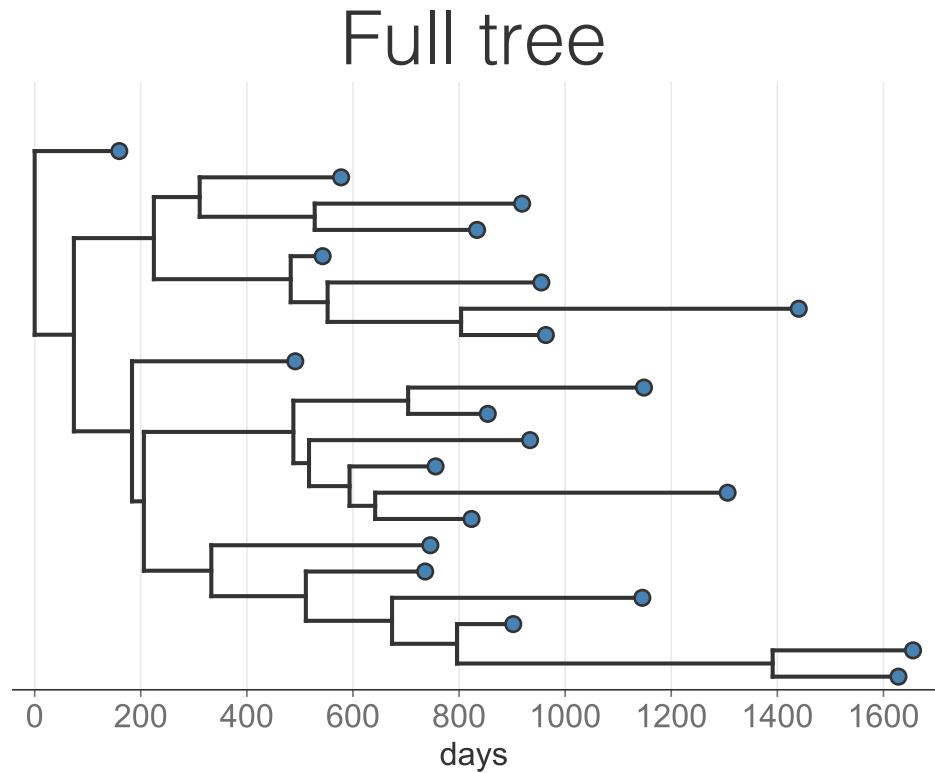
This method relies on reliable case counts!

Identifying epidemiological dynamics using patterns of neutral evolution

	Continual Immune Selection	Weak or Absent Immune Selection	
		Tree shape controlled by non-selective population dynamic processes	
Idealized Phylogeny Shapes			
Examples	Human influenza A virus intra-host HIV	inter-host HIV inter-host HCV	Measles, rabies inter-host HIV
Tree Inferences	Detection of antigenic escape mutations	Estimation of population growth rates	Estimation of population migration rates

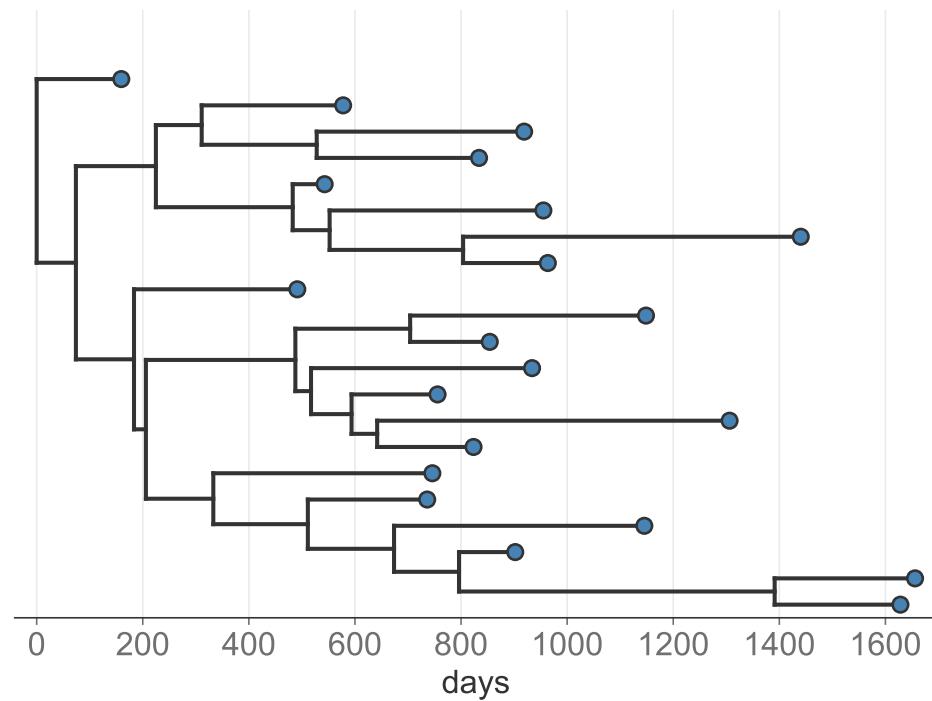


We (almost) always only sample a subset of the viral population

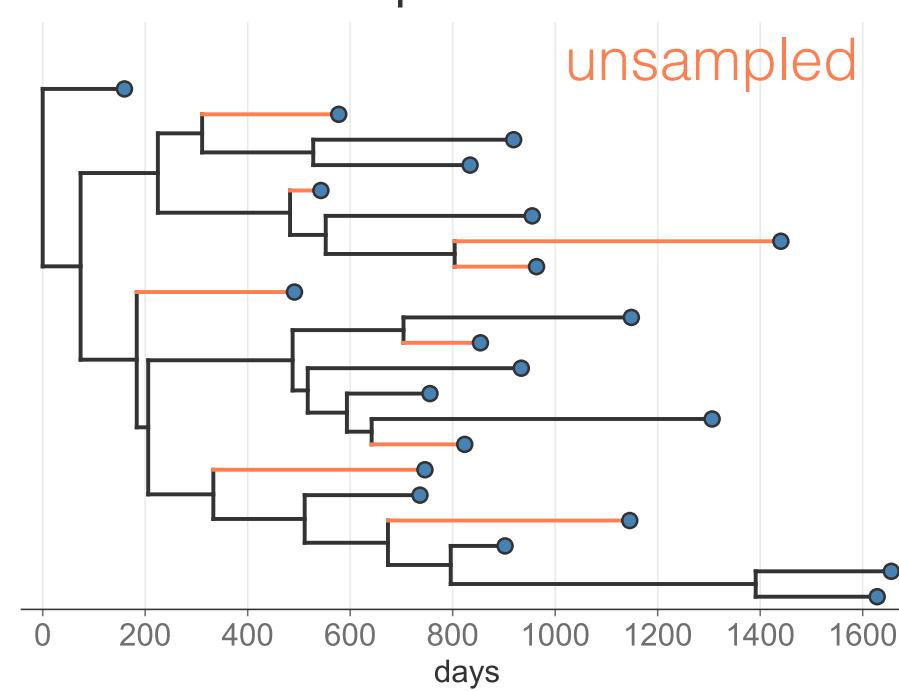


We (almost) always only sample a subset of the viral population

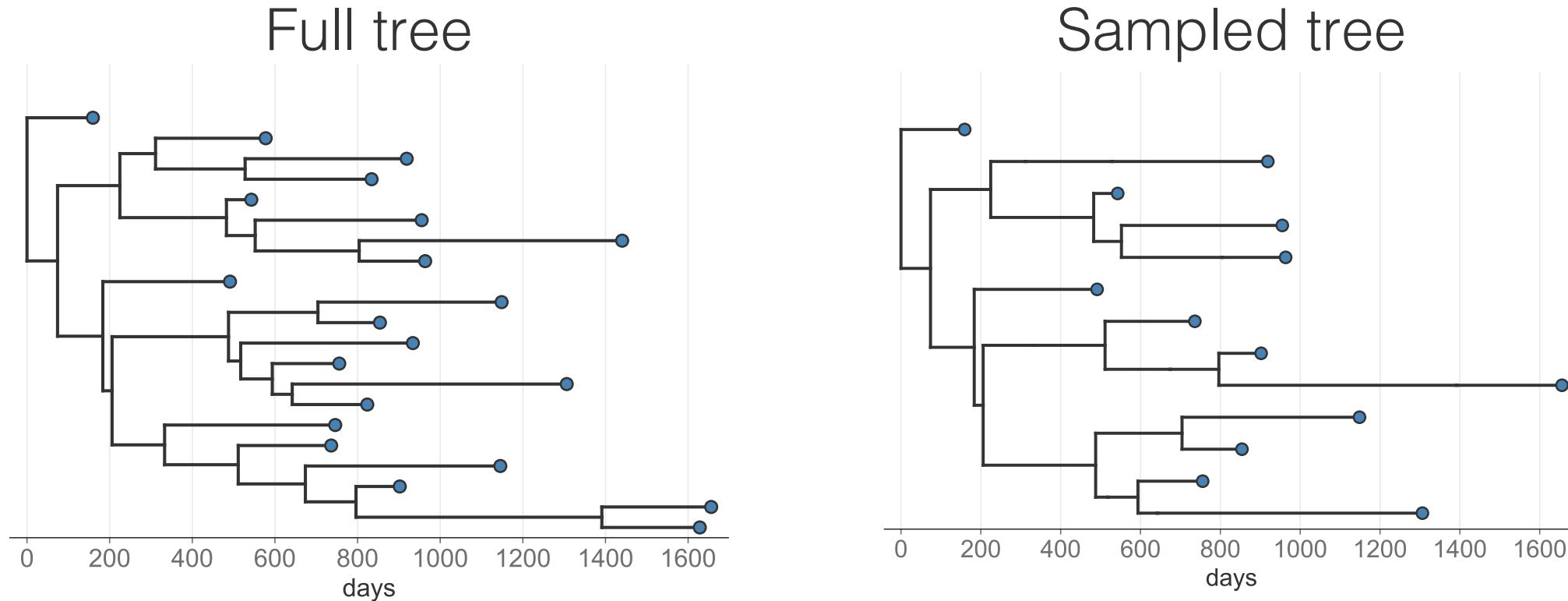
Full tree



Sampled tree

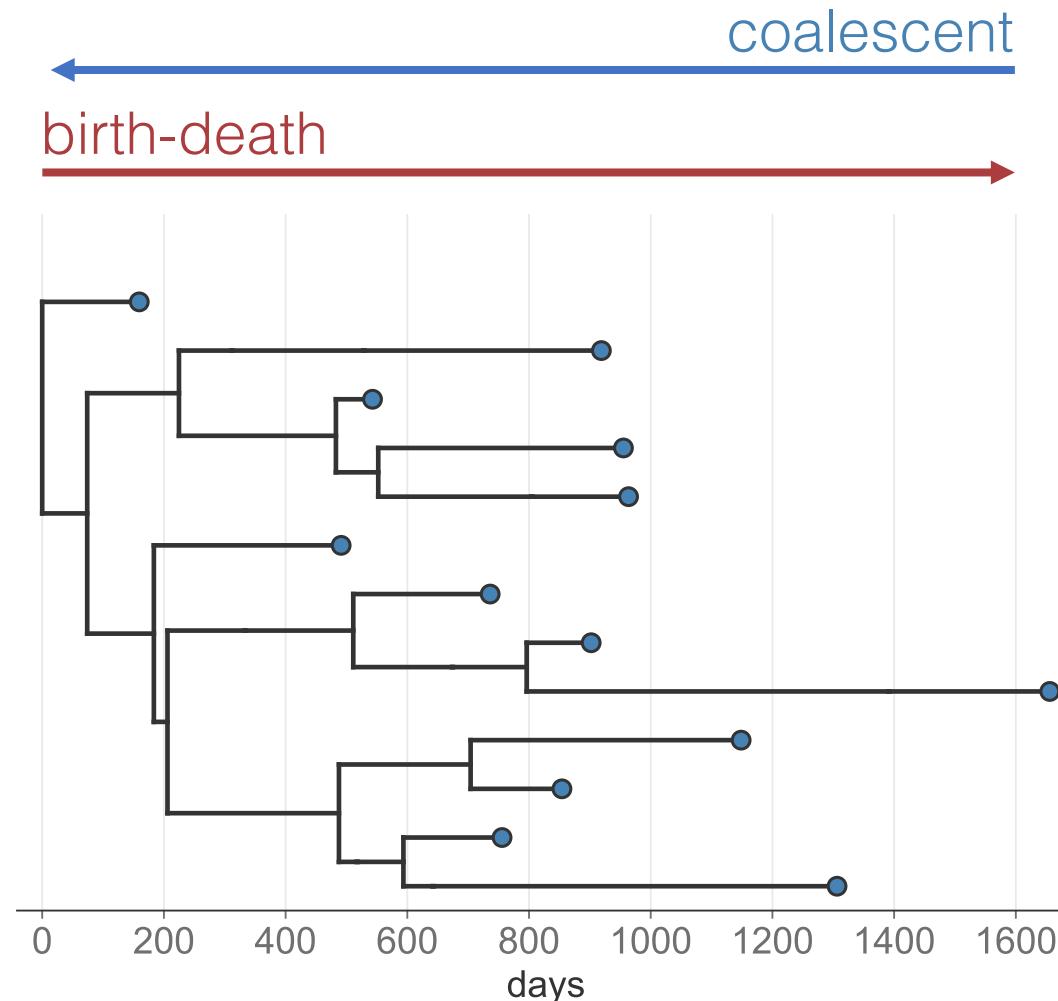


We (almost) always only sample a subset of the viral population



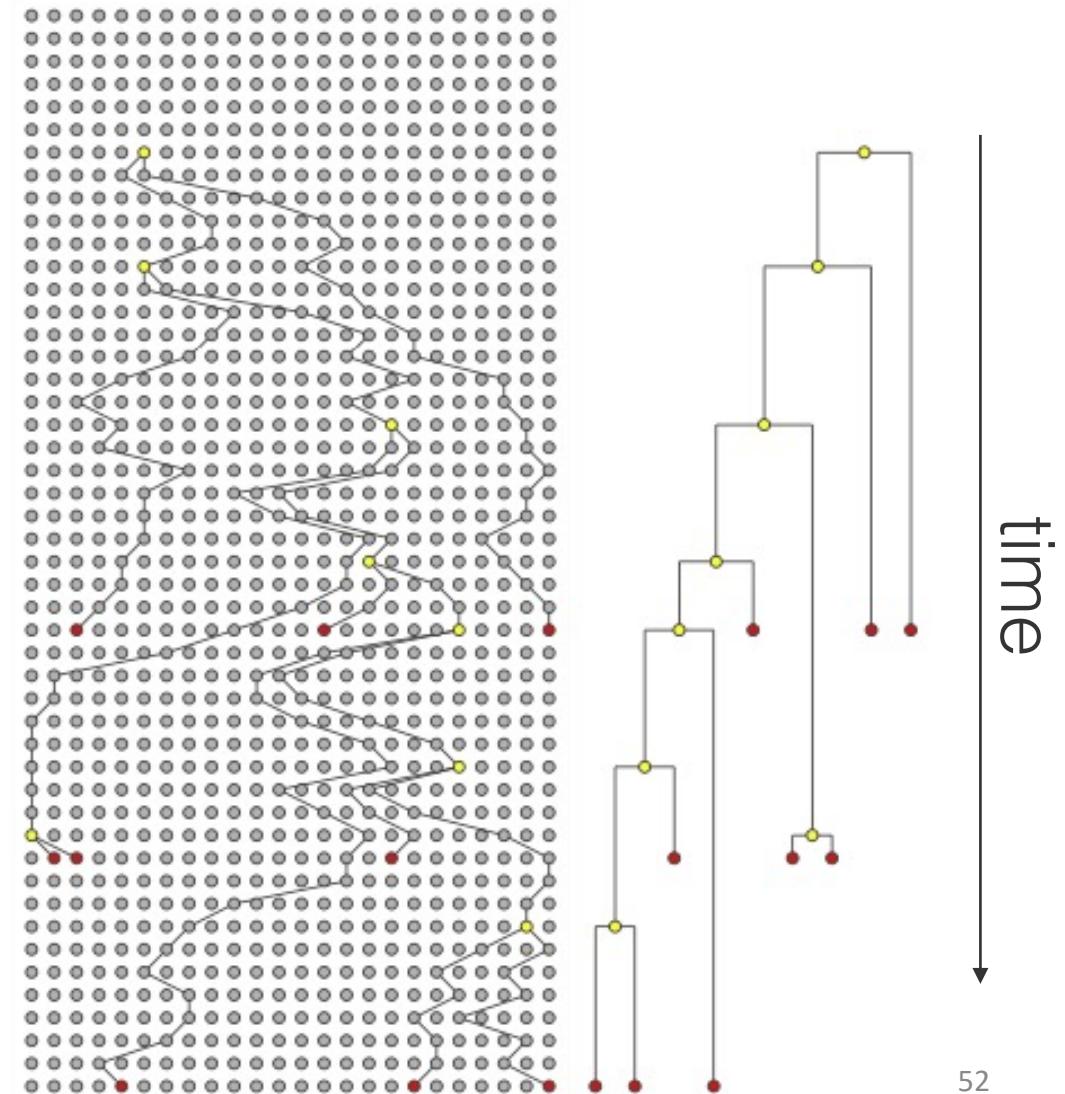
Demographic modeling using phylodynamics
infers underlying population dynamics using a
set of sampled sequences

Phylogenetic demographic models either look forward or backwards in time



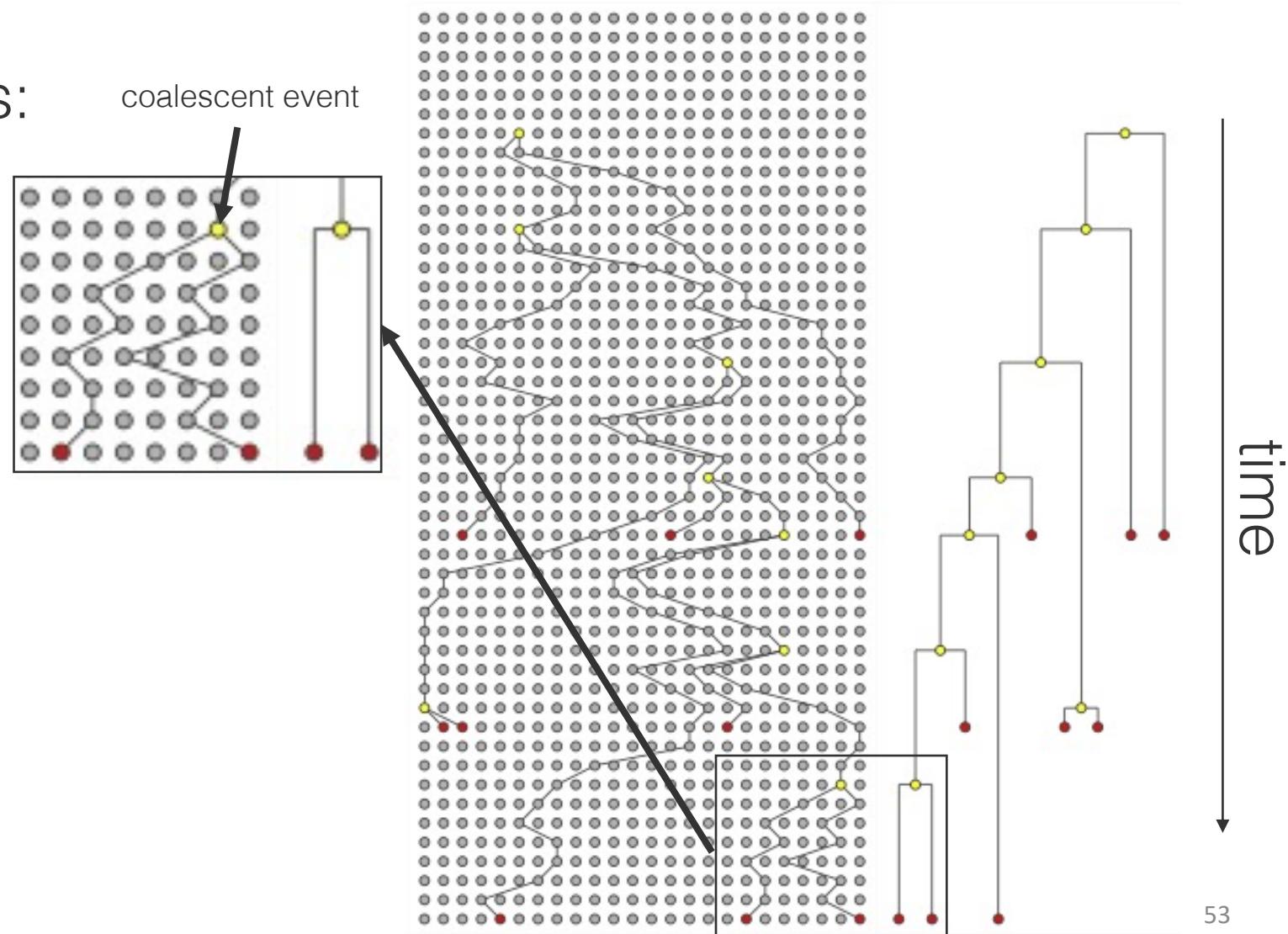
Coalescent models look backwards in time

Two lineages:



Coalescent models look backwards in time

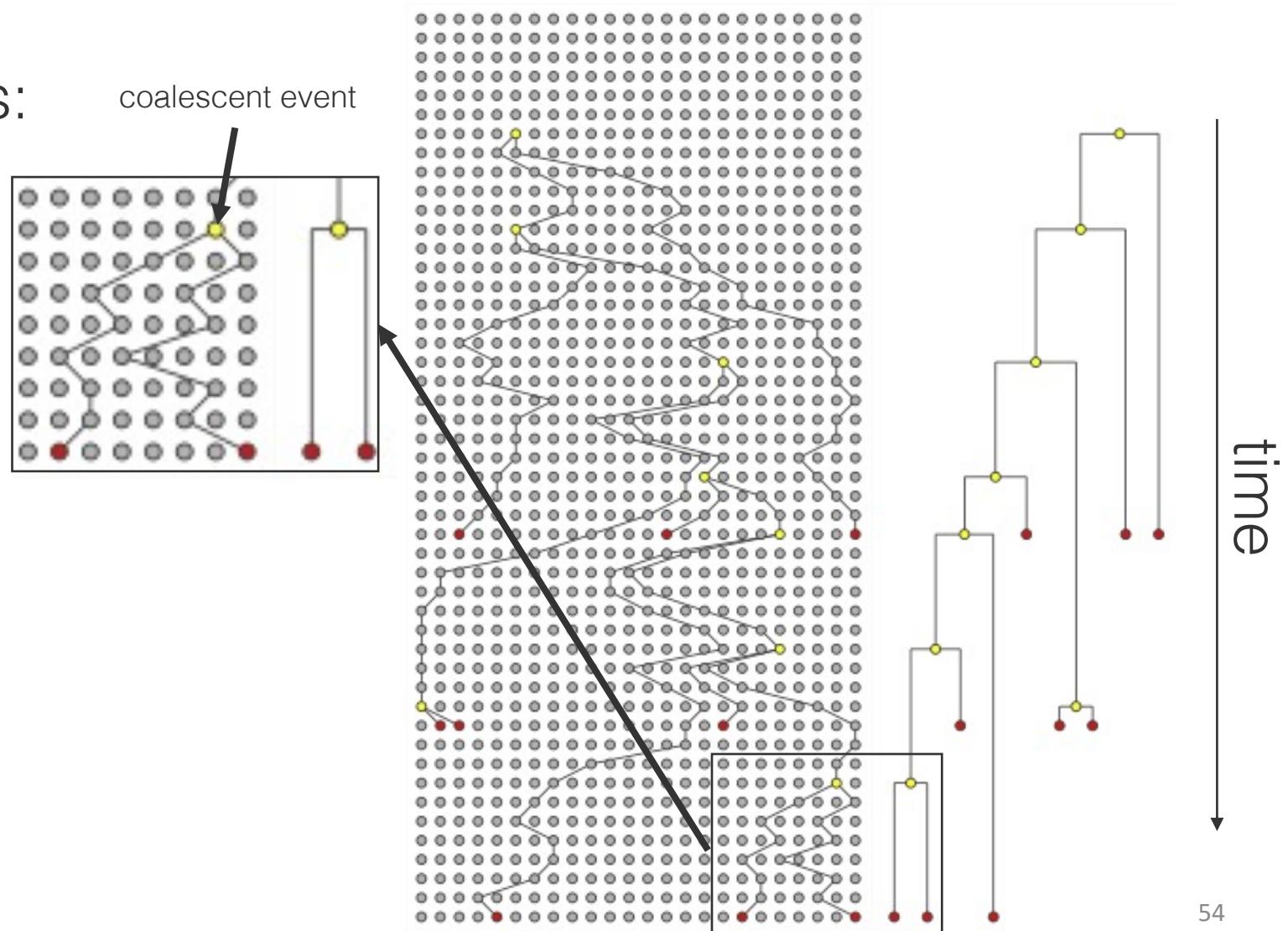
Two lineages:



Coalescent models look backwards in time

Two lineages:

$$P(\text{coalesce at time } t - 1) = \frac{1}{N}$$

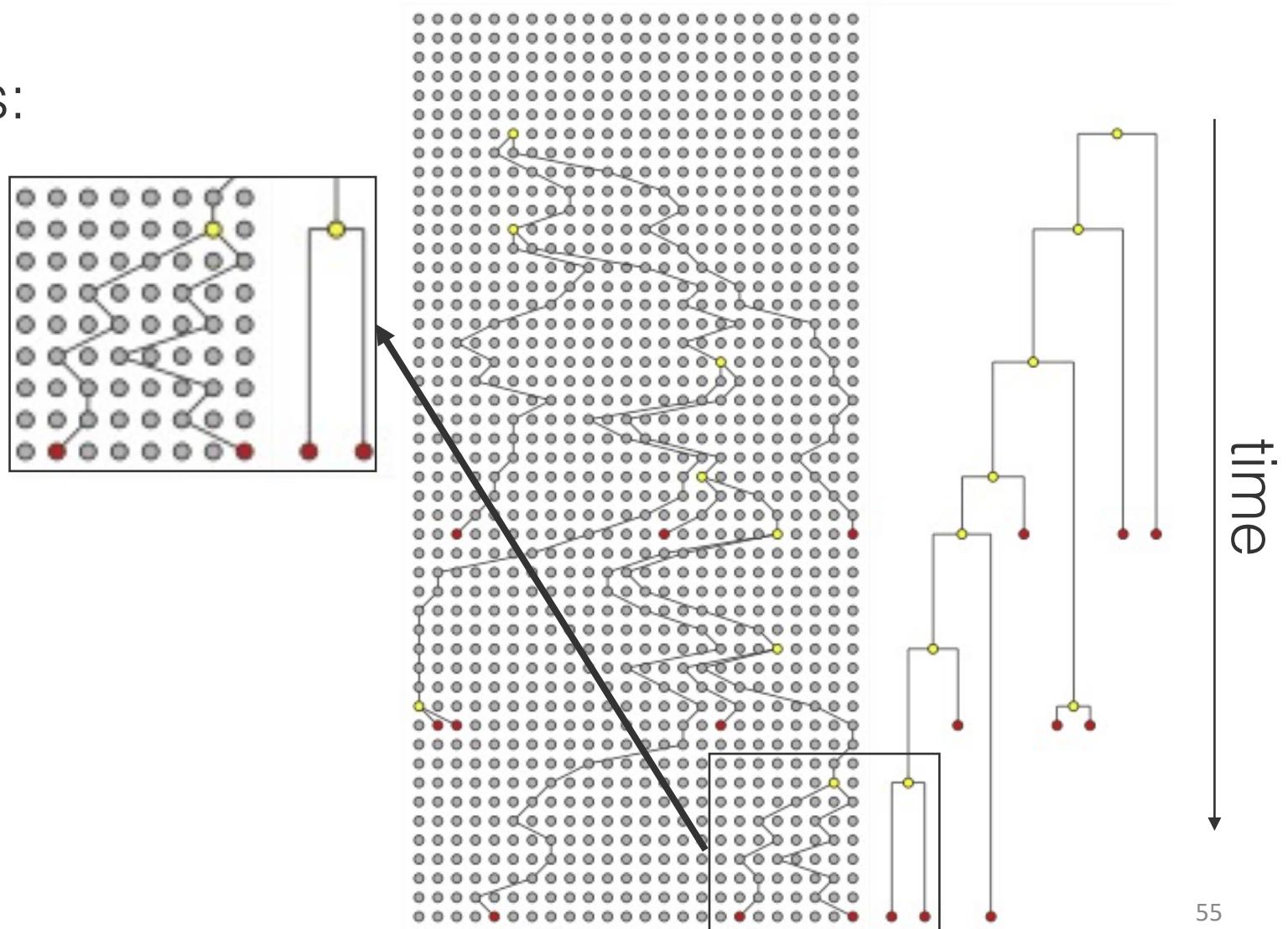


Coalescent models look backwards in time

Two lineages:

$$P(\text{coalesce at time } t - 1) = \frac{1}{N}$$

What is the average time until coalescence (time to most recent common ancestor, TMRCA)?



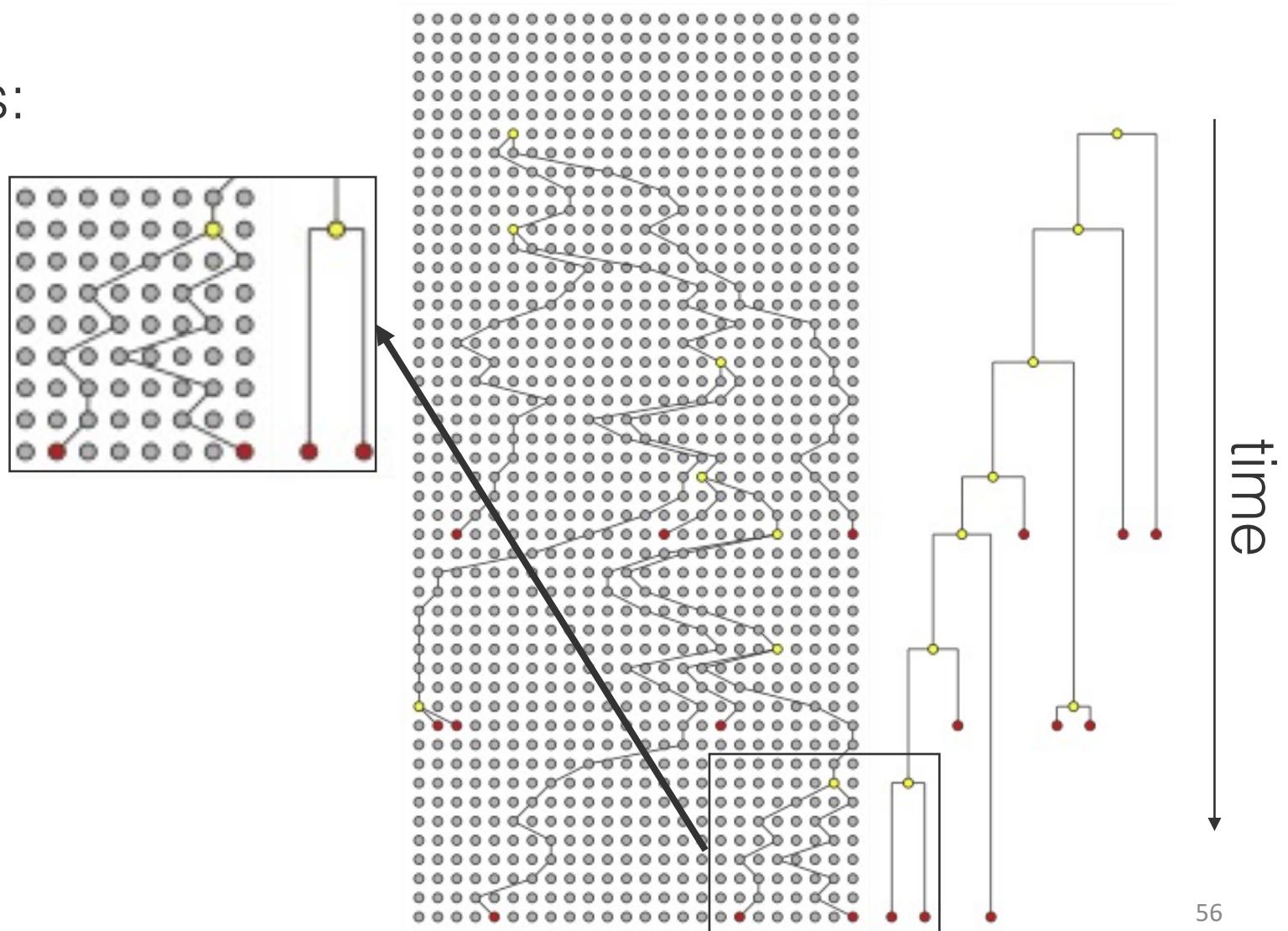
Coalescent models look backwards in time

Two lineages:

$$P(\text{coalesce at time } t - 1) = \frac{1}{N}$$

What is the average time until coalescence (time to most recent common ancestor, TMRCA)?

Geometric distribution with success p :
 $P(x) = (1 - p)^{x-1}p$



Coalescent models look backwards in time

Two lineages:

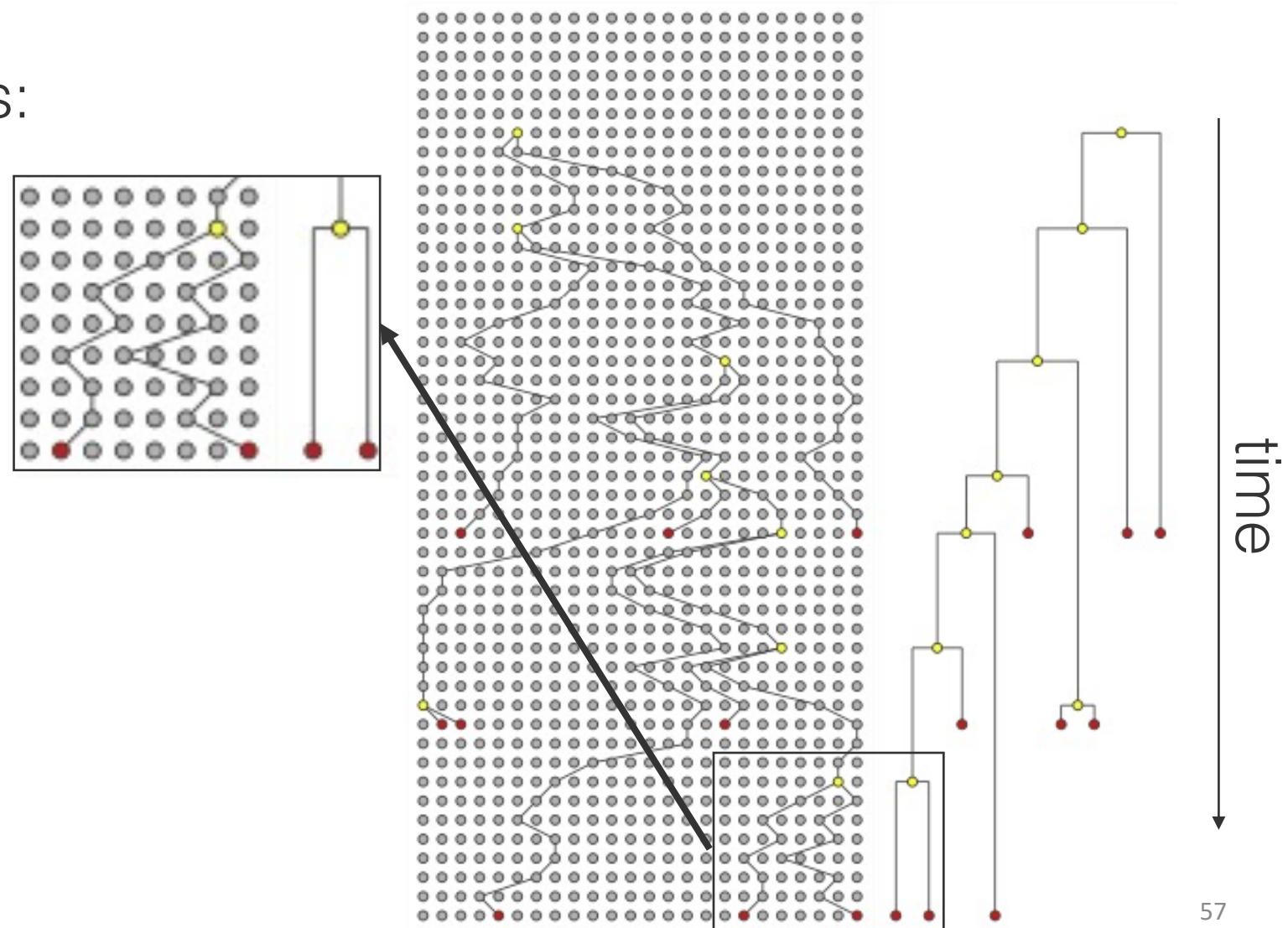
$$P(\text{coalesce at time } t - 1) = \frac{1}{N}$$

What is the average time until coalescence (time to most recent common ancestor, TMRCA)?

Geometric distribution with success p :

$$P(x) = (1 - p)^{x-1} p$$

$$E(x) = \sum_{x=1}^{\infty} xP(x) = 1/p$$



Coalescent models look backwards in time

Two lineages:

$$P(\text{coalesce at time } t - 1) = \frac{1}{N}$$

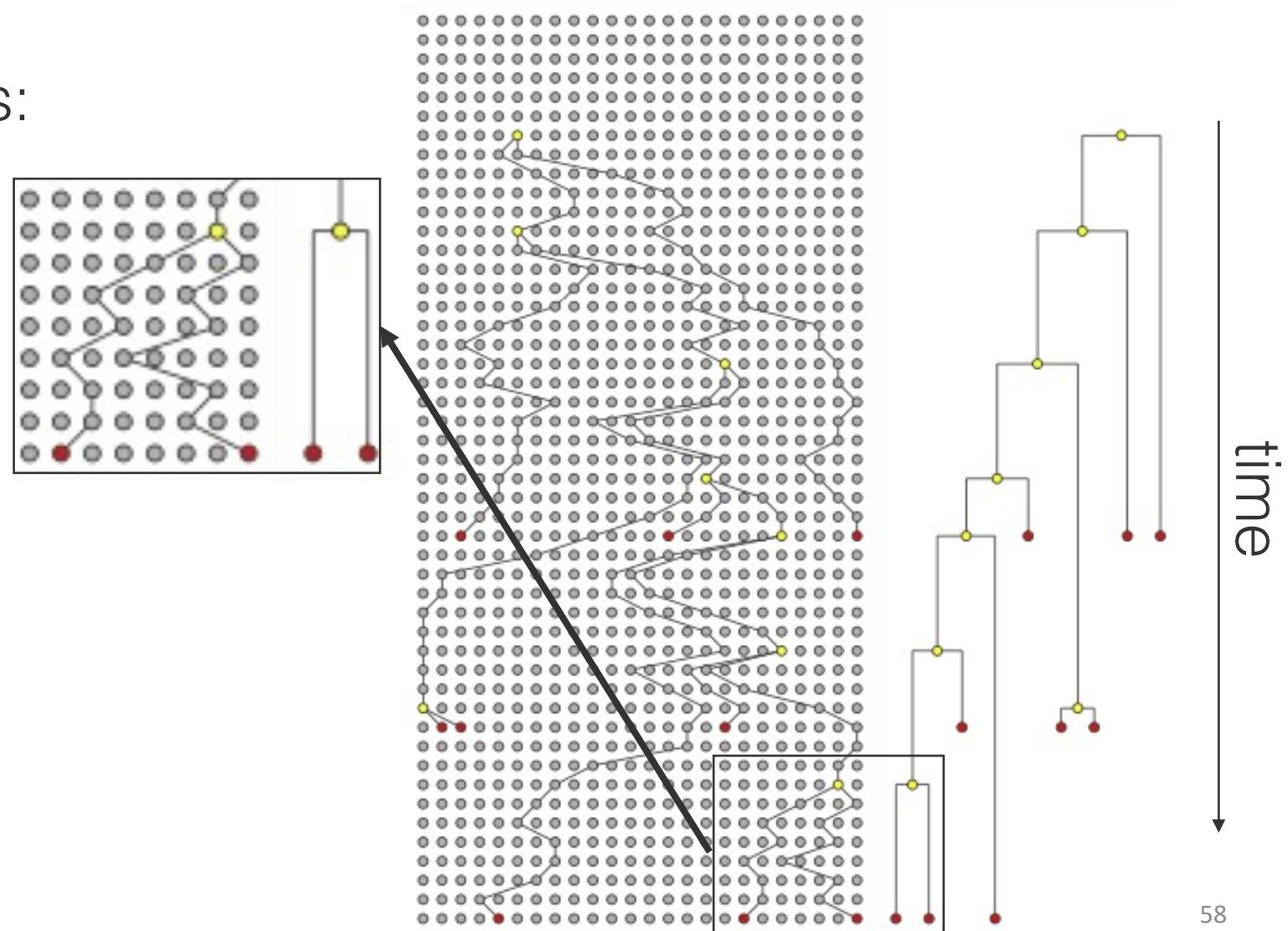
What is the average time until coalescence (time to most recent common ancestor, TMRCA)?

Geometric distribution with success p :

$$P(x) = (1 - p)^{x-1} p$$

$$E(x) = \sum_{x=1}^{\infty} xP(x) = 1/p$$

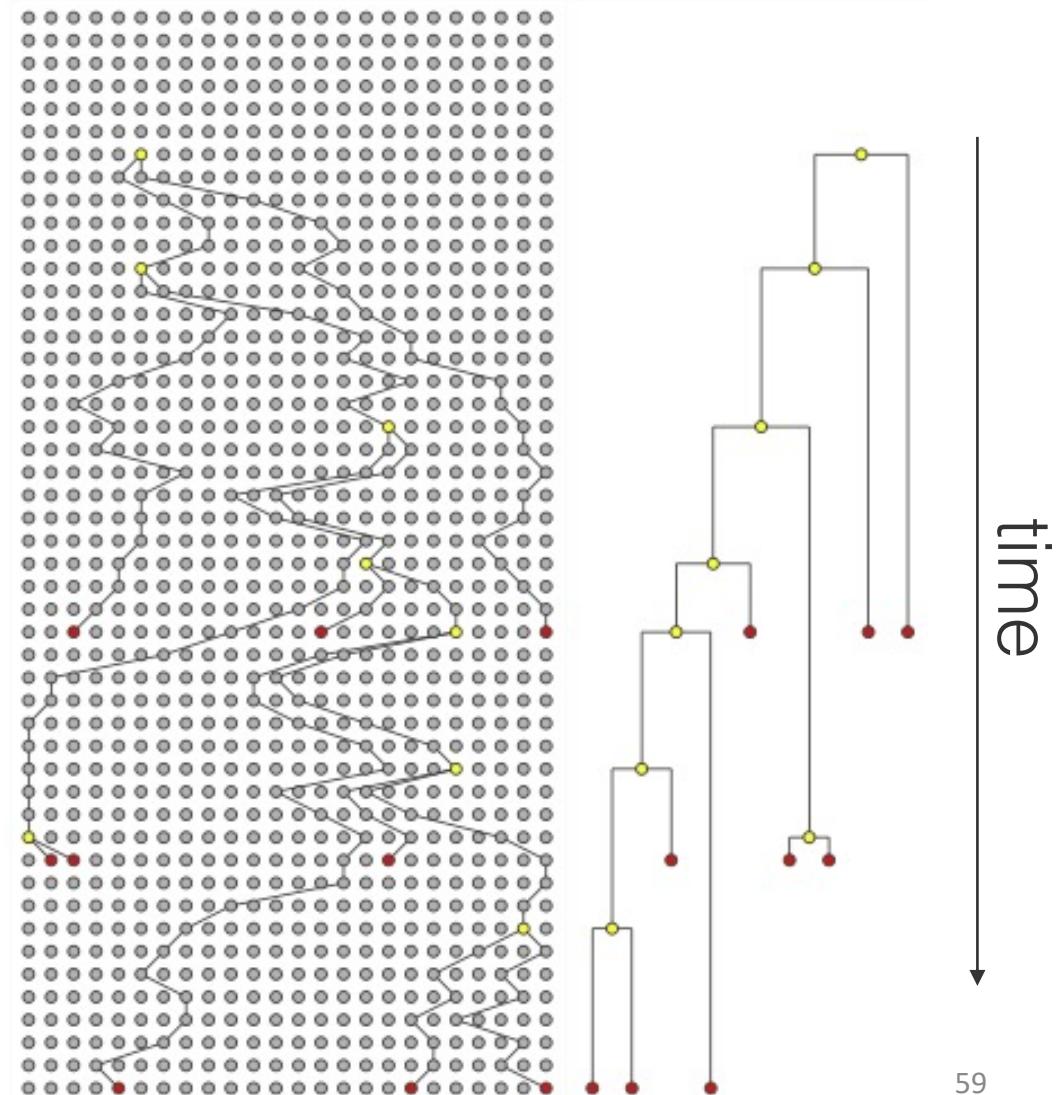
$$E(\text{TMRCA}) = \frac{1}{1/N} = N$$



Coalescent models look backwards in time

k lineages:

There are $\binom{k}{2} = \frac{k(k-1)}{2}$ possible combinations of lineages to coalesce in the $t - 1$ generation

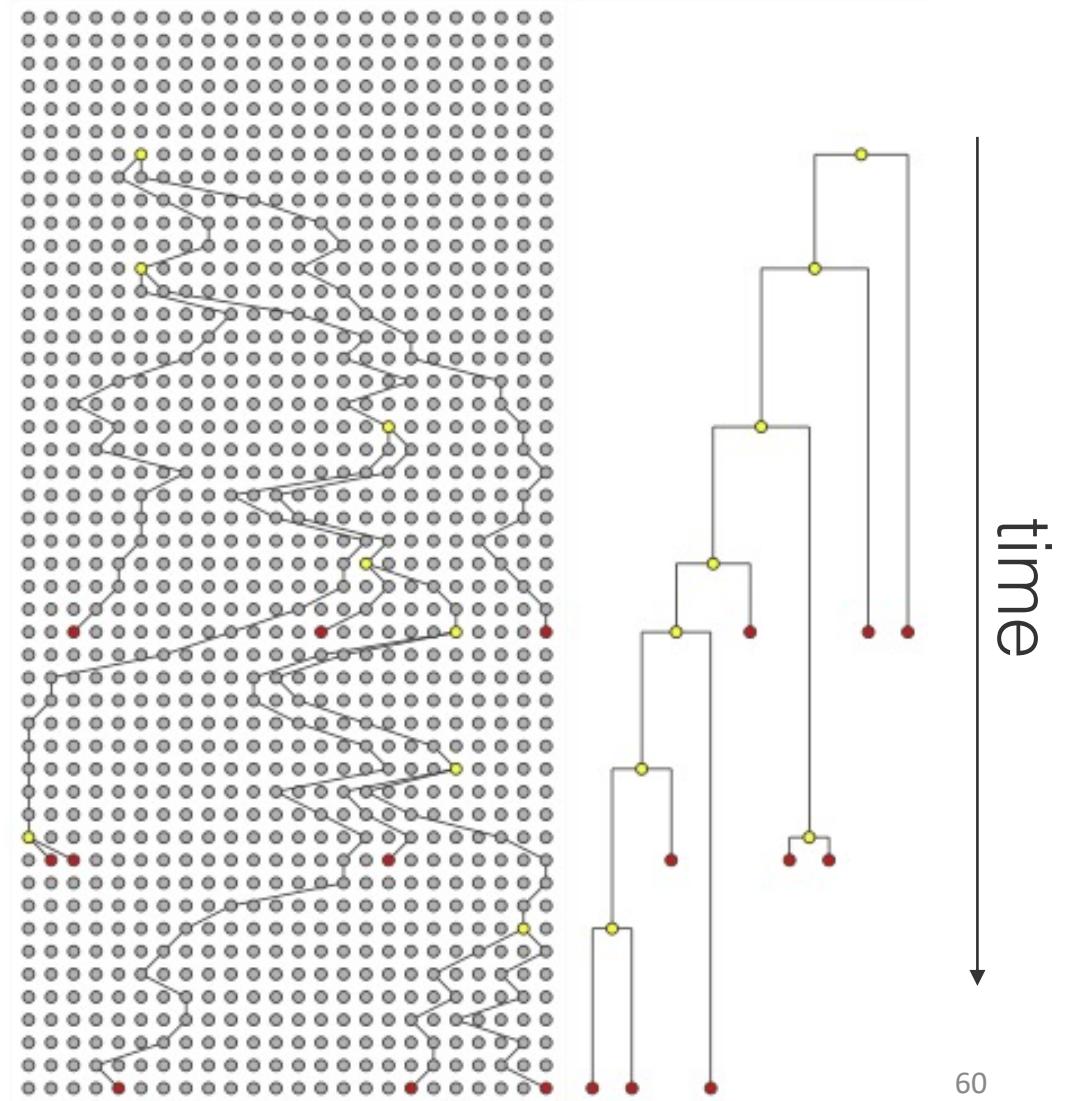


Coalescent models look backwards in time

k lineages:

There are $\binom{k}{2} = \frac{k(k-1)}{2}$ possible combinations of lineages to coalesce in the $t - 1$ generation

$$P(\text{coalesce at time } t - 1) = \frac{\binom{k}{2}}{N}$$



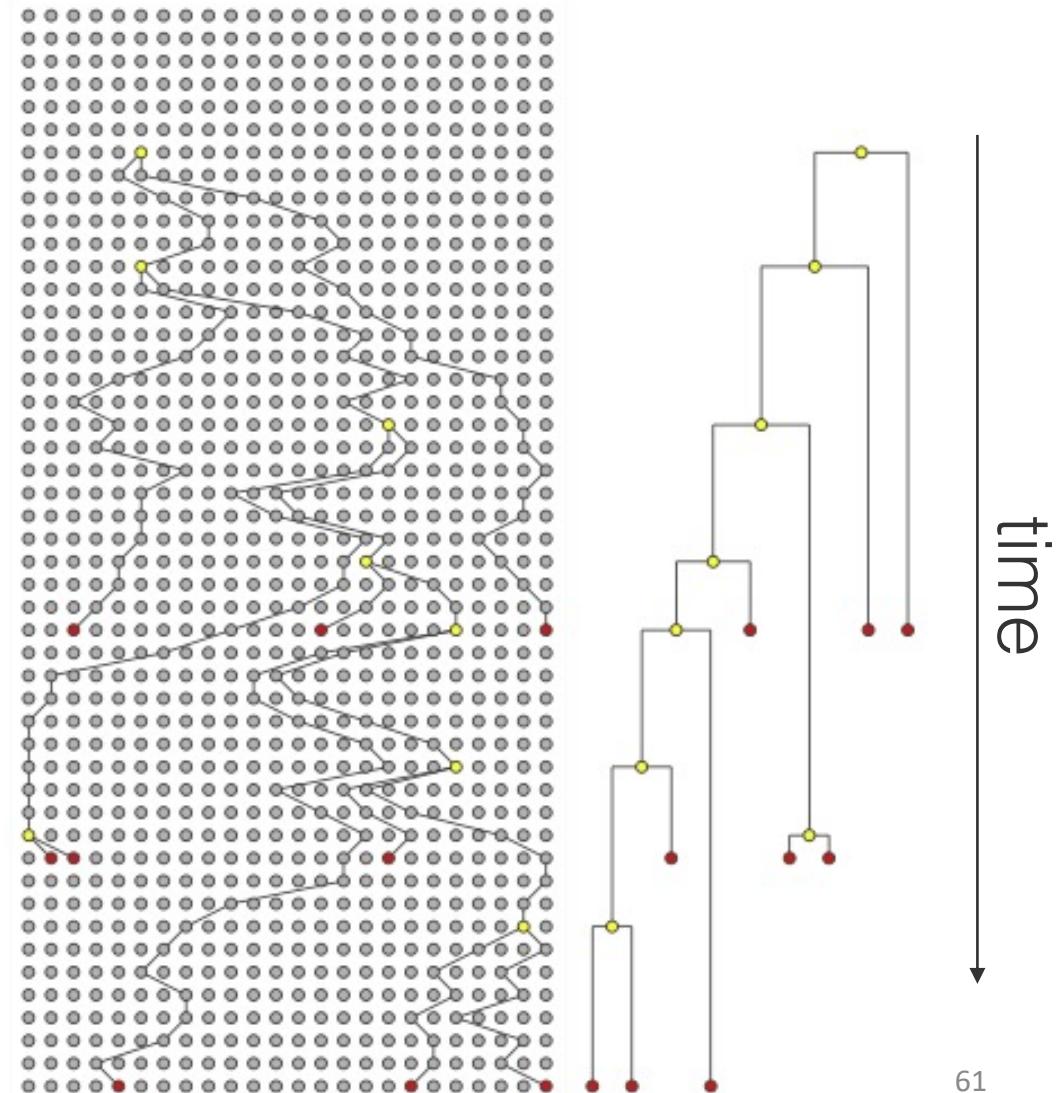
Coalescent models look backwards in time

k lineages:

There are $\binom{k}{2} = \frac{k(k-1)}{2}$ possible combinations of lineages to coalesce in the $t - 1$ generation

$$P(\text{coalesce at time } t - 1) = \frac{k(k-1)}{2} \frac{1}{N}$$

$$\text{Expected time to 1 coalescence } (T_k) = \frac{2N}{k(k-1)}$$



Coalescent models look backwards in time

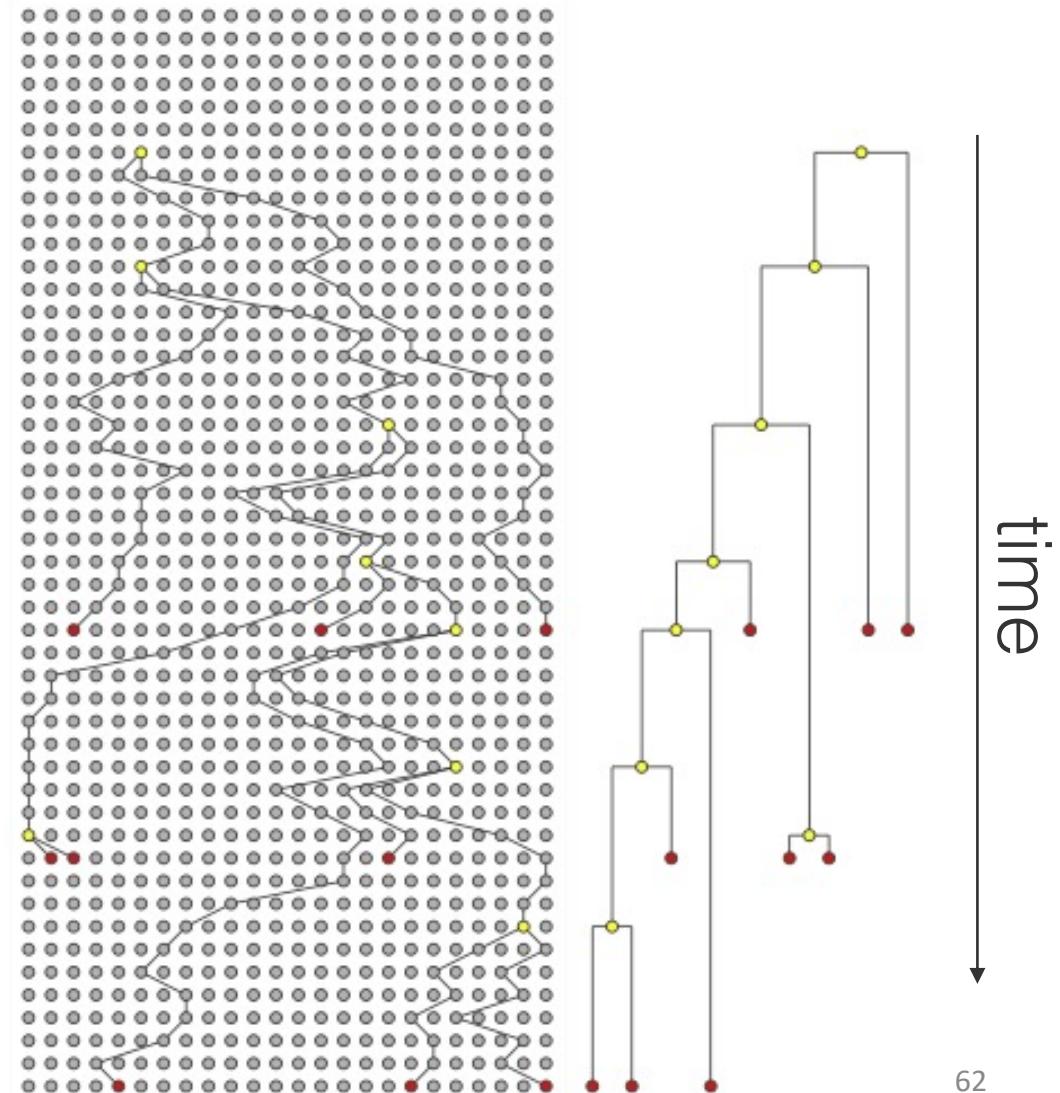
k lineages:

There are $\binom{k}{2} = \frac{k(k-1)}{2}$ possible combinations of lineages to coalesce in the $t - 1$ generation

$$P(\text{coalesce at time } t - 1) = \frac{k(k-1)}{2} \frac{1}{N}$$

$$\text{Expected time to 1 coalescence } (T_k) = \frac{2N}{k(k-1)}$$

$$\text{Expected time to 2 coalescences } (T_{k-1}) = \frac{2N}{(k-1)(k-2)}$$



Coalescent models look backwards in time

k lineages:

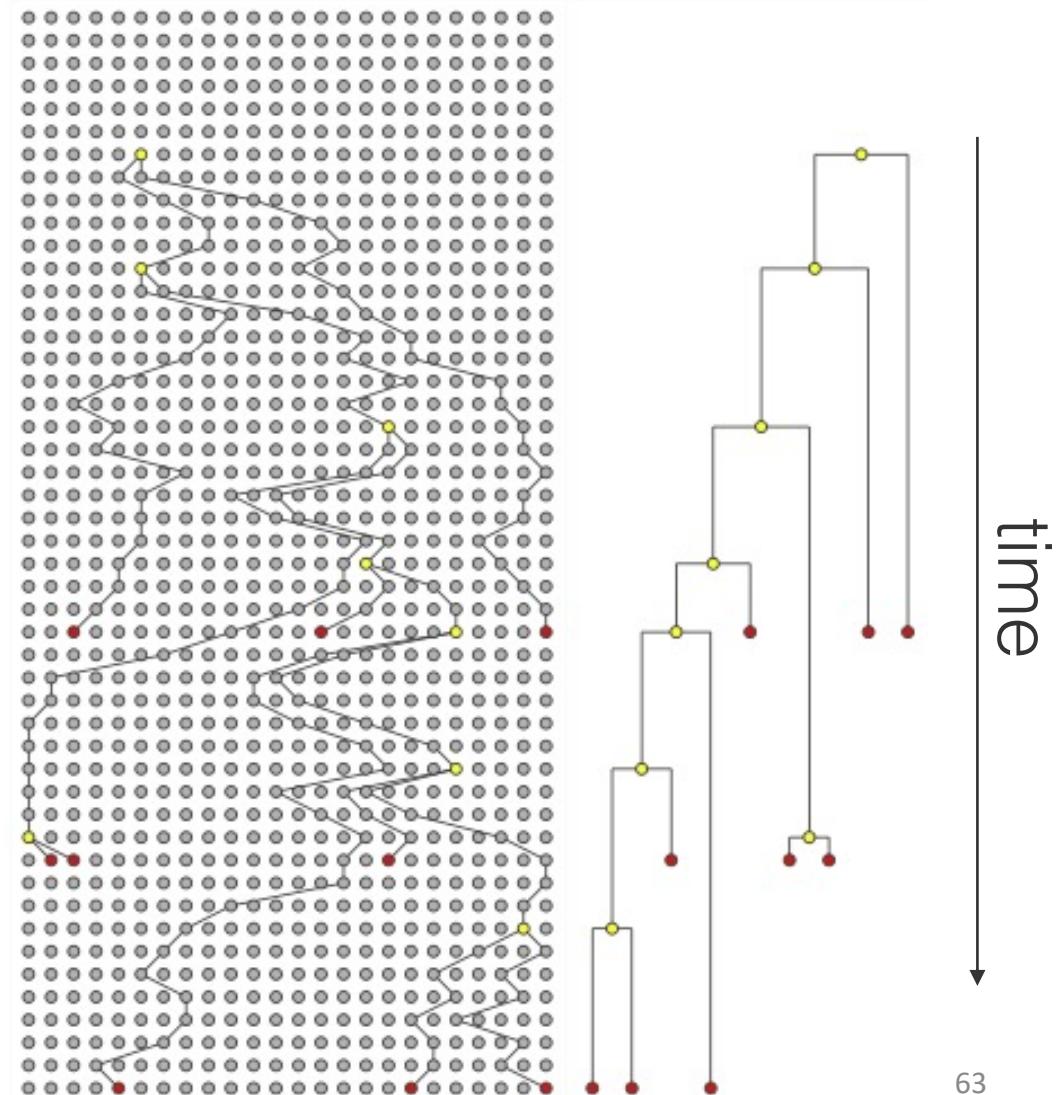
There are $\binom{k}{2} = \frac{k(k-1)}{2}$ possible combinations of lineages to coalesce in the $t - 1$ generation

$$P(\text{coalesce at time } t - 1) = \frac{k(k-1)}{2} \frac{1}{N}$$

$$\text{Expected time to 1 coalescence } (T_k) = \frac{2N}{k(k-1)}$$

$$\text{Expected time to 2 coalescences } (T_{k-1}) = \frac{2N}{(k-1)(k-2)}$$

$$\text{Expected time to } k \text{ coalescences} = T_k + T_{k-1} + \dots + T_2 \approx 2N$$



Coalescent models look backwards in time

k lineages:

There are $\binom{k}{2} = \frac{k(k-1)}{2}$ possible combinations of lineages to coalesce in the $t - 1$ generation

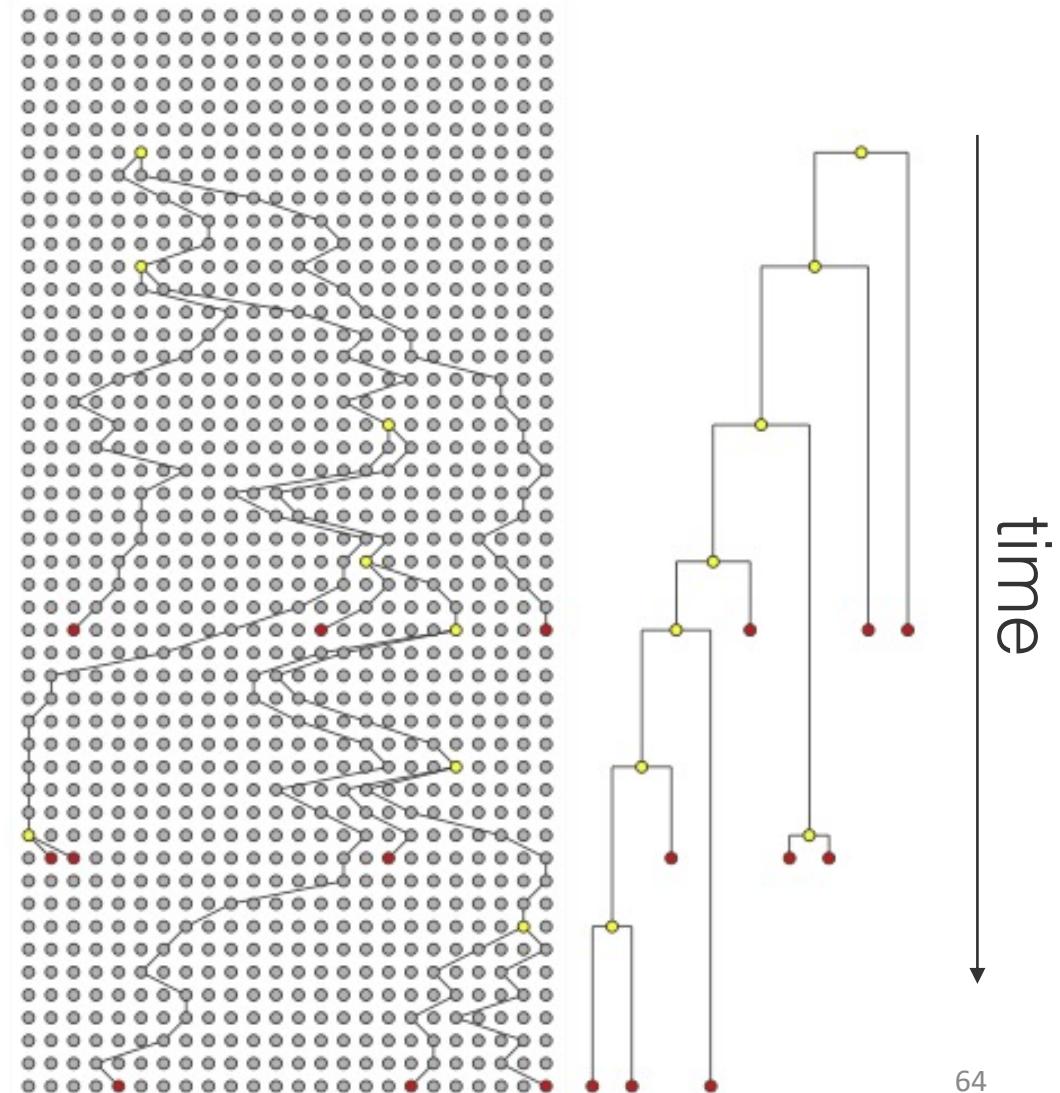
$$P(\text{coalesce at time } t - 1) = \frac{k(k-1)}{2} \frac{1}{N}$$

$$\text{Expected time to 1 coalescence } (T_k) = \frac{2N}{k(k-1)}$$

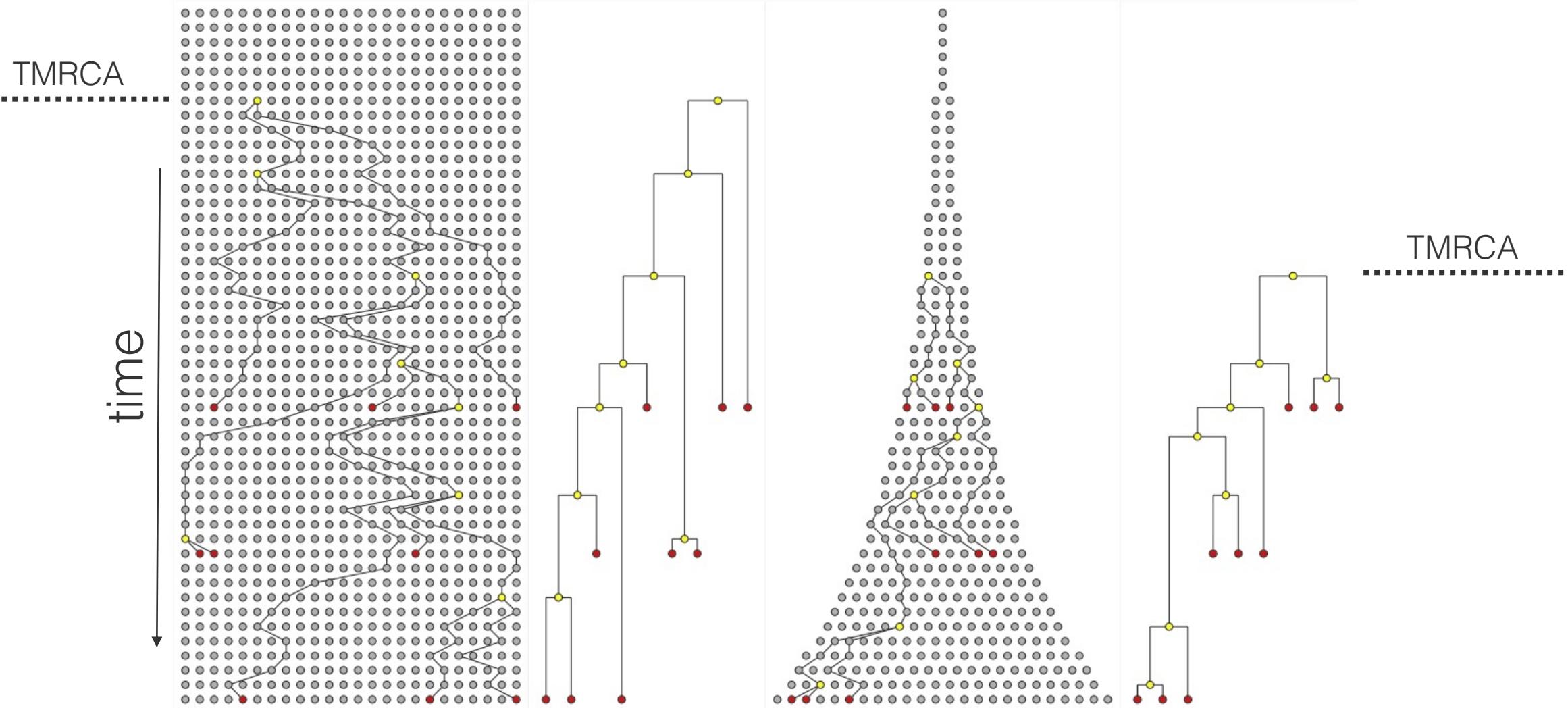
$$\text{Expected time to 2 coalescences } (T_{k-1}) = \frac{2N}{(k-1)(k-2)}$$

$$\text{Expected time to } k \text{ coalescences} = T_k + T_{k-1} + \dots + T_2 \approx 2N$$

The time to coalesce the last two lineages is 50% of the total TMRCA!



Population sizes are often dynamic

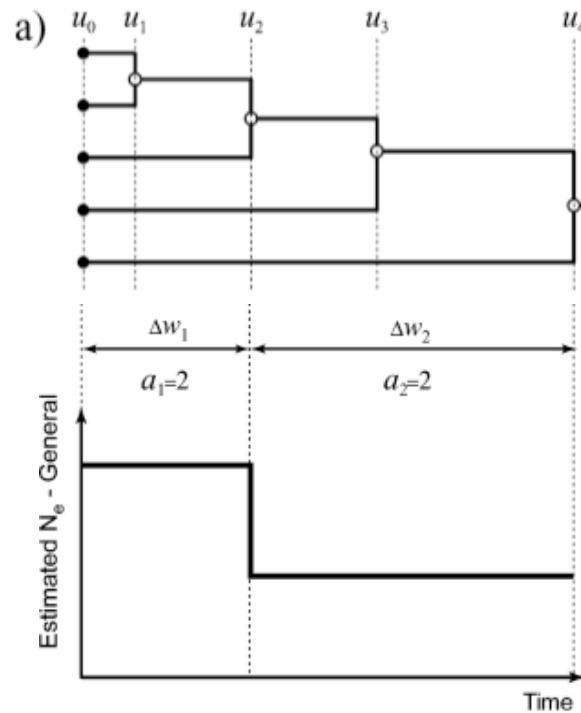


Non-parametric coalescent models

Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences

A. J. Drummond, A. Rambaut, B. Shapiro, and O. G. Pybus

Department of Zoology, University of Oxford, Oxford, United Kingdom

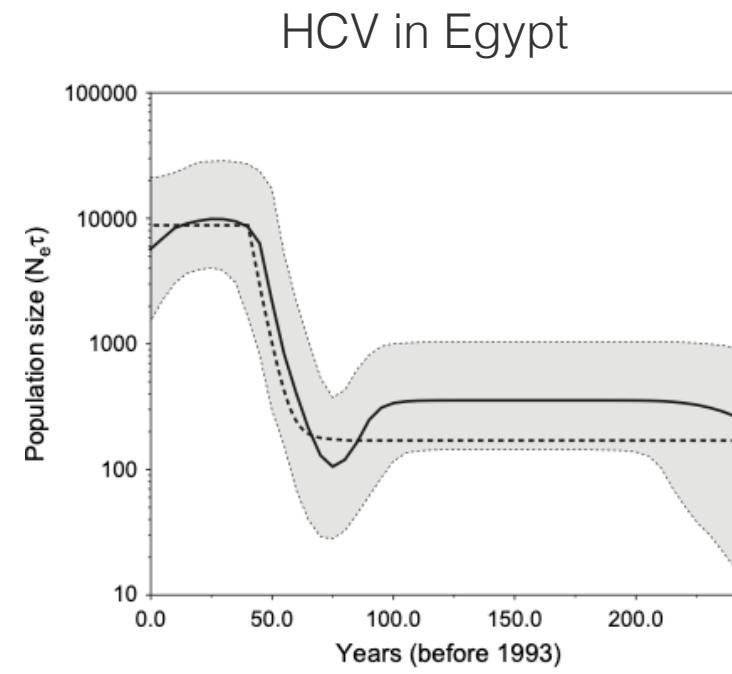
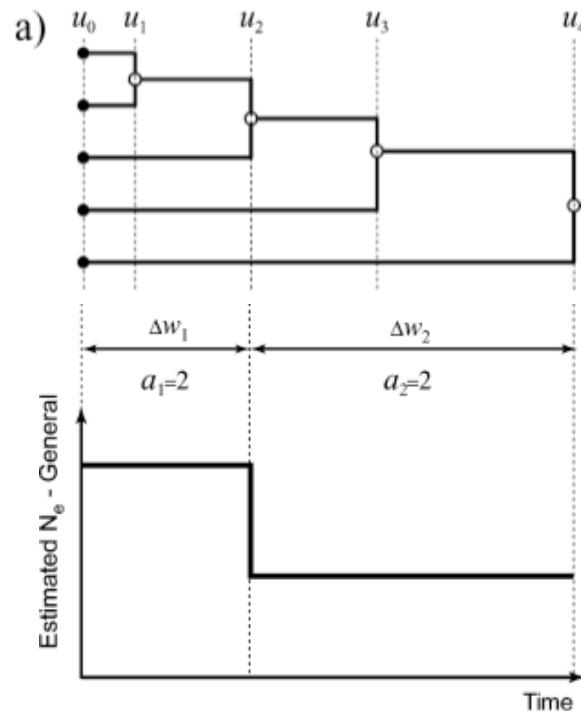


Non-parametric coalescent models

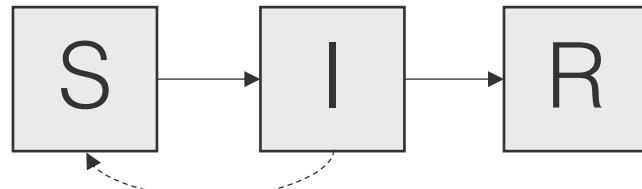
Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences

A. J. Drummond, A. Rambaut, B. Shapiro, and O. G. Pybus

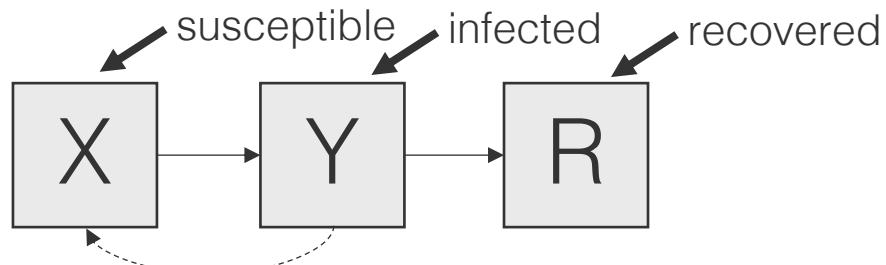
Department of Zoology, University of Oxford, Oxford, United Kingdom



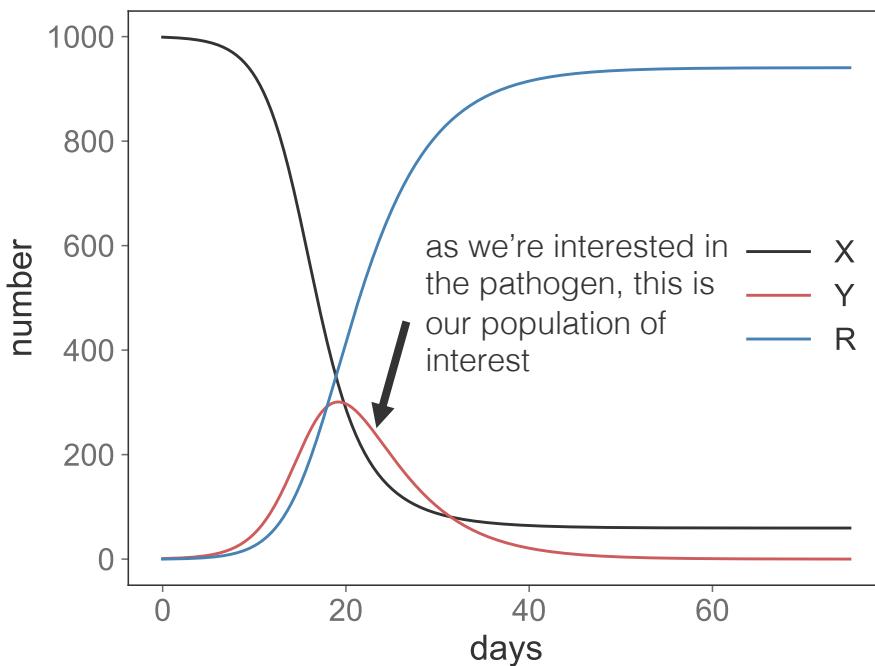
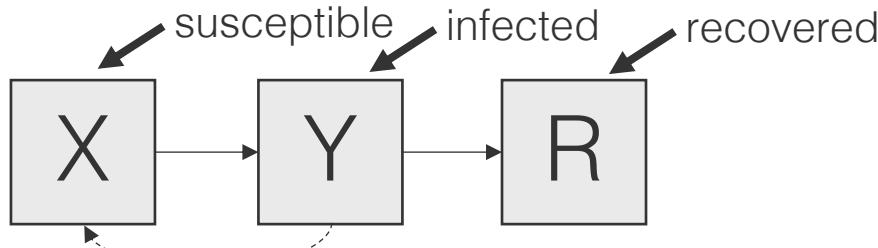
Structured coalescent models



Structured coalescent models



Structured coalescent models

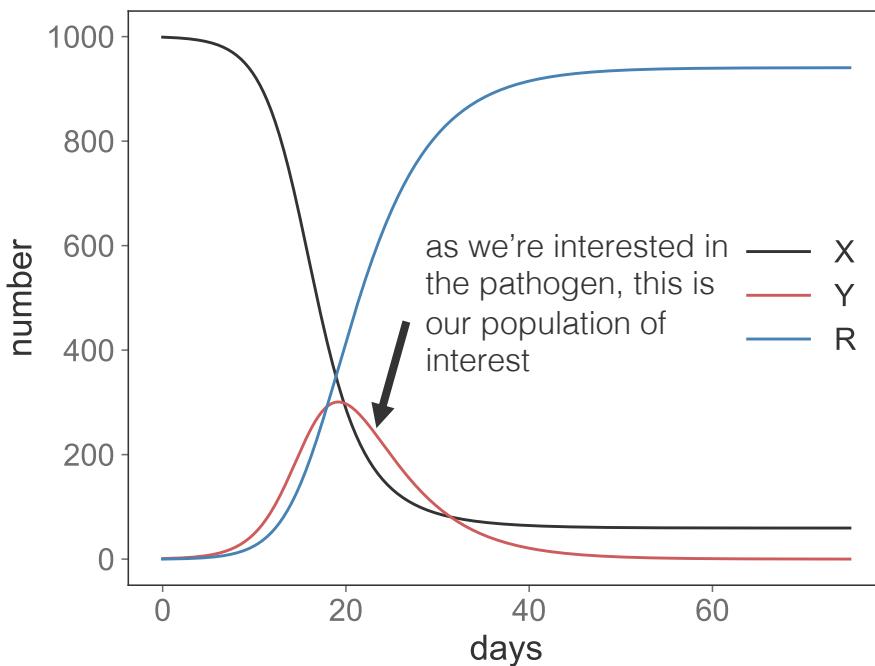
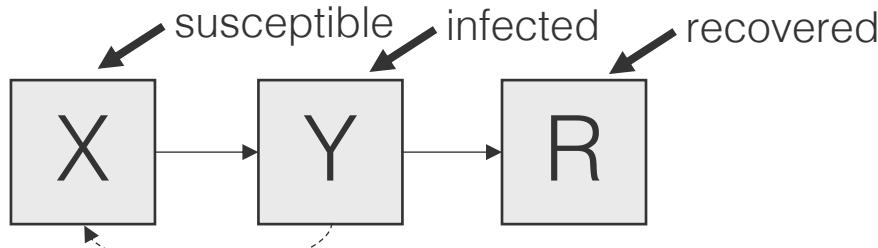


$$\frac{dX}{dt} = -\beta X(t)Y(t) = -f(t)$$

$$\frac{dY}{dt} = f(t) - \gamma Y(t)$$

$$\frac{dR}{dt} = \gamma Y(t)$$

Structured coalescent models



$$\frac{dX}{dt} = -\beta X(t)Y(t) = -f(t)$$

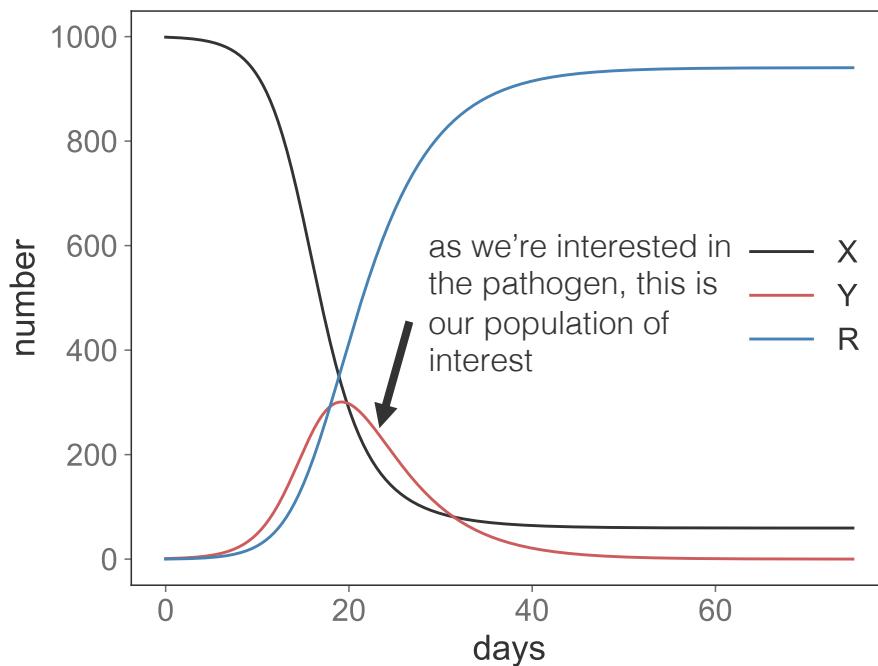
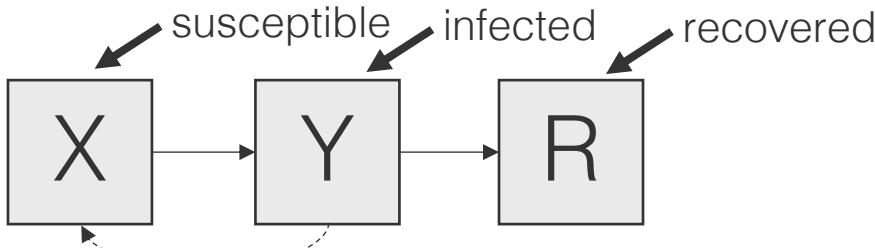
$$\frac{dY}{dt} = f(t) - \gamma Y(t)$$

$$\frac{dR}{dt} = \gamma Y(t)$$

The coalescent rate is not $\frac{1}{Y(t)}$

Need to account for changing number of offspring over time.

Structured coalescent models



$$\frac{dX}{dt} = -\beta X(t)Y(t) = -f(t)$$

$$\frac{dY}{dt} = f(t) - \gamma Y(t)$$

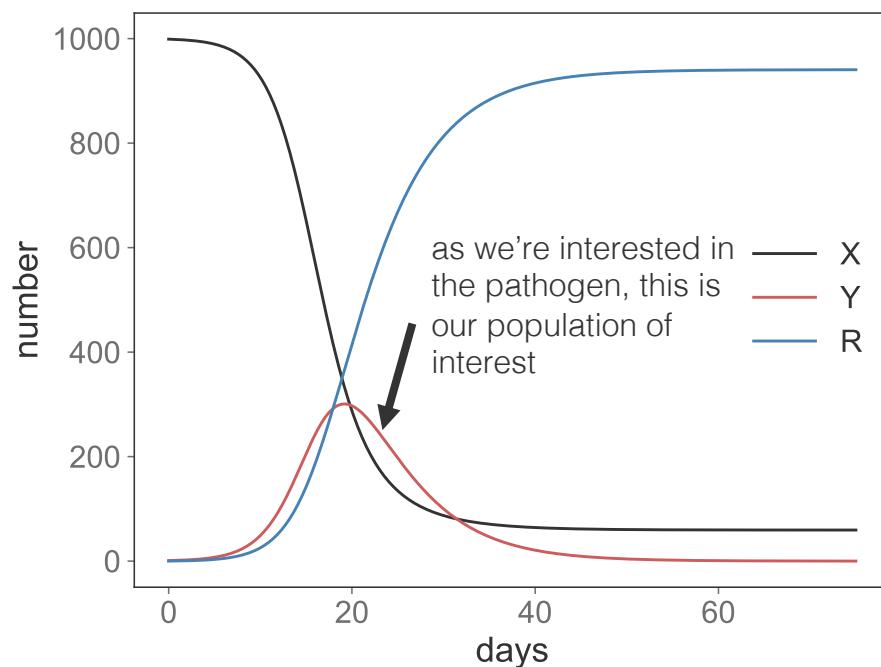
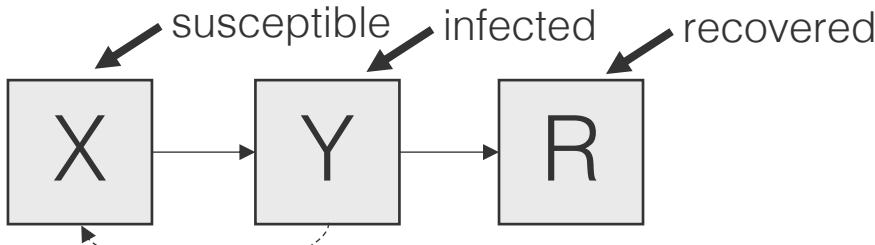
$$\frac{dR}{dt} = \gamma Y(t)$$

The coalescent rate is not $\frac{1}{Y(t)}$

Need to account for changing number of offspring over time.

$$\text{Coalescent rate } (\lambda) = \frac{2f(t)}{Y(t)^2} = \frac{2\beta X(t)}{Y(t)}$$

Structured coalescent models



$$\frac{dX}{dt} = -\beta X(t)Y(t) = -f(t)$$

$$\frac{dY}{dt} = f(t) - \gamma Y(t)$$

$$\frac{dR}{dt} = \gamma Y(t)$$

The coalescent rate is not $\frac{1}{Y(t)}$

Need to account for changing number of offspring over time.

$$\text{Coalescent rate } (\lambda) = \frac{2f(t)}{Y(t)^2} = \frac{2\beta X(t)}{Y(t)}$$

Generalizable to many deterministic scenarios

PhyDyn: Bayesian phylodynamics with complex models

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Bayesian phylodynamic inference with complex models

Erik M. Volz*, Igor Siveroni

Department of Infectious Disease Epidemiology and the MRC Centre for Global Infectious Disease Analysis,
Imperial College London, London, United Kingdom

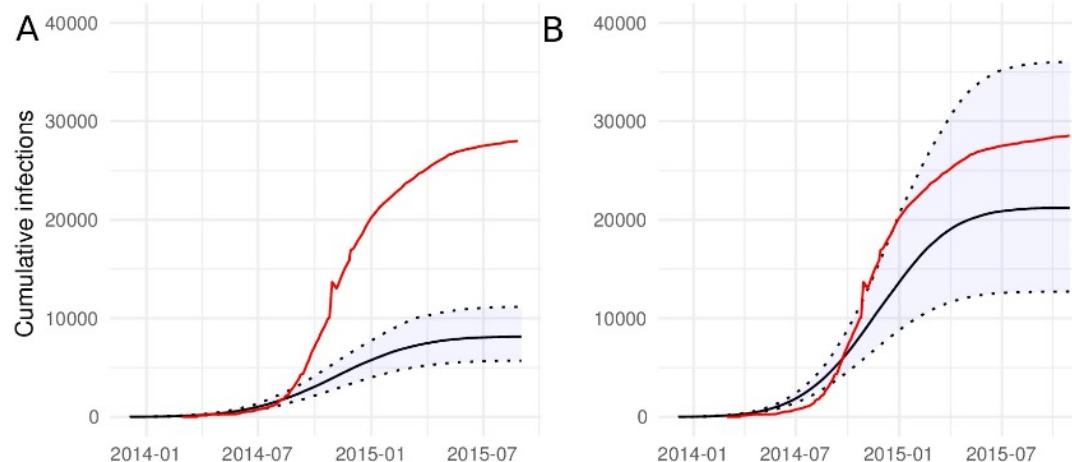


Fig 3. Model-based estimates of cumulative infections through time for the 2014-15 Ebola epidemic in Western Africa.
Estimates are shown for the SEIR model (A) and the model which includes super-spreading (B). The red line show the cumulative number of cases reported by WHO [35].

PhyDyn: Bayesian phylodynamics with complex models

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Bayesian phylodynamic inference with complex models

Erik M. Volz*, Igor Siveroni

Department of Infectious Disease Epidemiology and the MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, United Kingdom

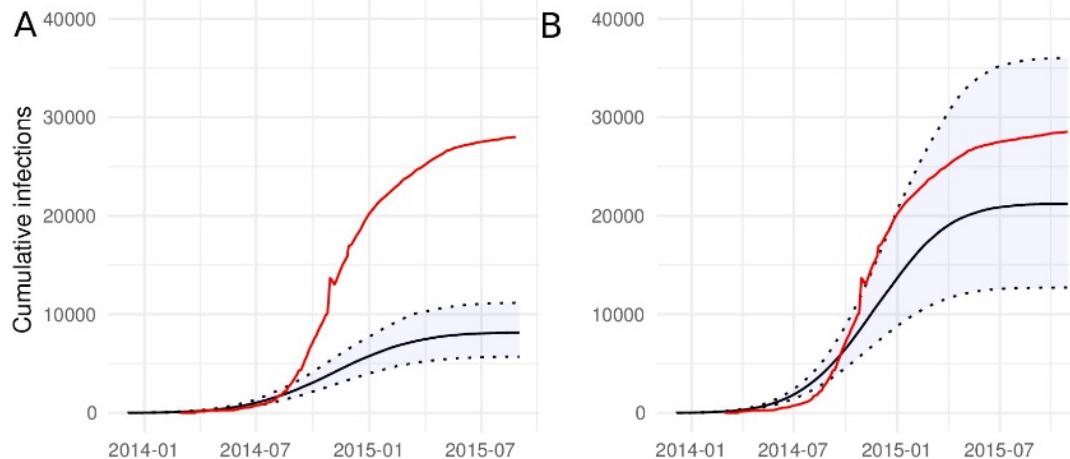


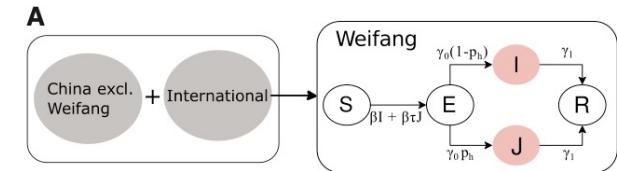
Fig 3. Model-based estimates of cumulative infections through time for the 2014–15 Ebola epidemic in Western Africa. Estimates are shown for the SEIR model (A) and the model which includes super-spreading (B). The red line shows the cumulative number of cases reported by WHO [35].

VIRUS EVOLUTION

Virus Evolution, 2021, 7(1): evesa102
doi: 10.5293/vesa.evesa102
Research Article

Genomic epidemiology of a densely sampled COVID-19 outbreak in China

Lily Geidelberg,^{1,*§} Olivia Boyd,¹ David Jorgensen,¹ Igor Siveroni,^{1,**} Fabrícia F. Nascimento,¹ Robert Johnson,¹ Manon Ragonnet-Cronin,¹ Han Fu,¹ Haowei Wang,¹ Xiaoyue Xi,² Wei Chen,³ Dehui Liu,³ Yingying Chen,³ Mengmeng Tian,³ Wei Tan,⁴ Junjie Zai,⁵ Wanying Sun,⁶ Jiandong Li,⁶ Junhua Li,⁶ Erik M. Volz^{1,*††} Xingguang Li^{7,*‡‡} and Qing Nie^{3,*‡‡}



PhyDyn: Bayesian phylodynamics with complex models

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Bayesian phylodynamic inference with complex models

Erik M. Volz*, Igor Siveroni

Department of Infectious Disease Epidemiology and the MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, United Kingdom

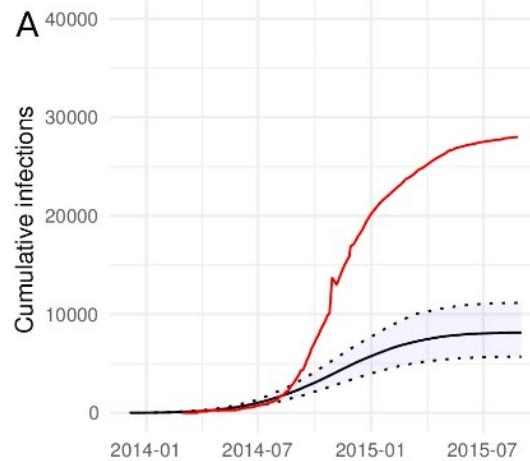


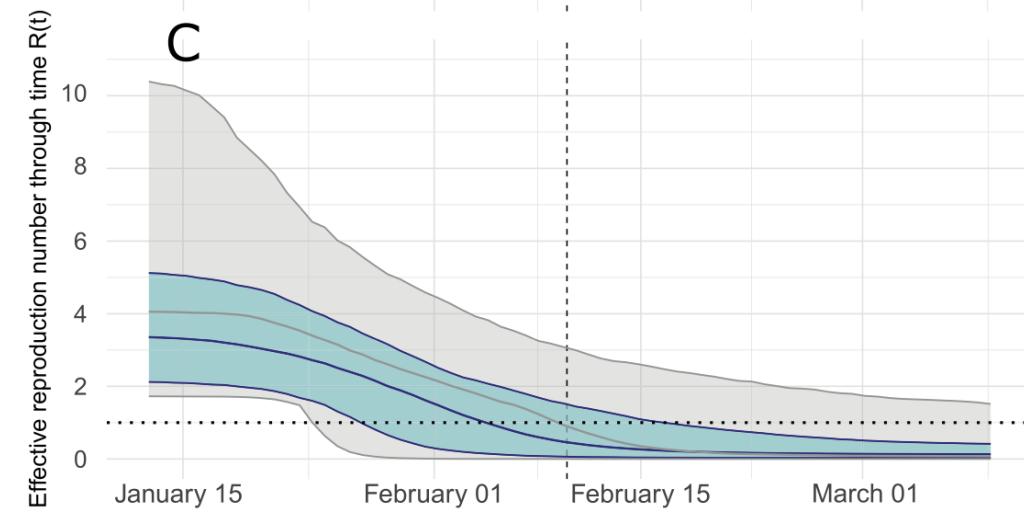
Fig 3. Model-based estimates of cumulative infections through time for the 2014–15 Ebola epidemic in Western Africa. Estimates are shown for the SEIR model (A) and the model which includes super-spreading (B). The red line show the cumulative number of cases reported by WHO [35].

VIRUS EVOLUTION

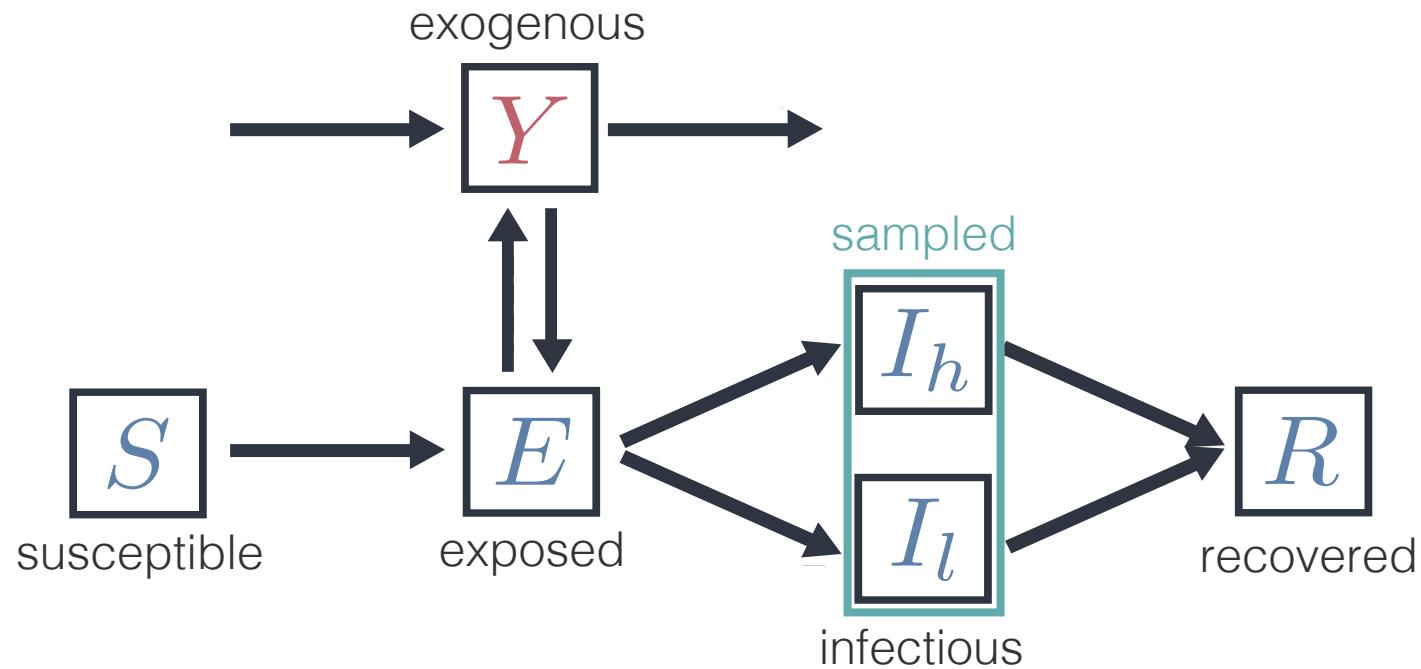
Virus Evolution, 2021, 7(1): eves102
doi: 10.5293/ves.eves102
Research Article

Genomic epidemiology of a densely sampled COVID-19 outbreak in China

Lily Geidelberg,^{1,*§} Olivia Boyd,¹ David Jorgensen,¹ Igor Siveroni,^{1,**} Fabrícia F. Nascimento,¹ Robert Johnson,¹ Manon Ragonnet-Cronin,¹ Han Fu,¹ Haowei Wang,¹ Xiaoyue Xi,² Wei Chen,³ Dehui Liu,³ Yingying Chen,³ Mengmeng Tian,³ Wei Tan,⁴ Junjie Zai,⁵ Wanying Sun,⁶ Jiandong Li,⁶ Junhua Li,⁶ Erik M. Volz^{1,*†‡} Xingguang Li^{7,*†‡} and Qing Nie^{3,*‡†‡}



Phylogenetics of SARS-CoV-2 in Israel, early 2020



Dr. Katia Koelle



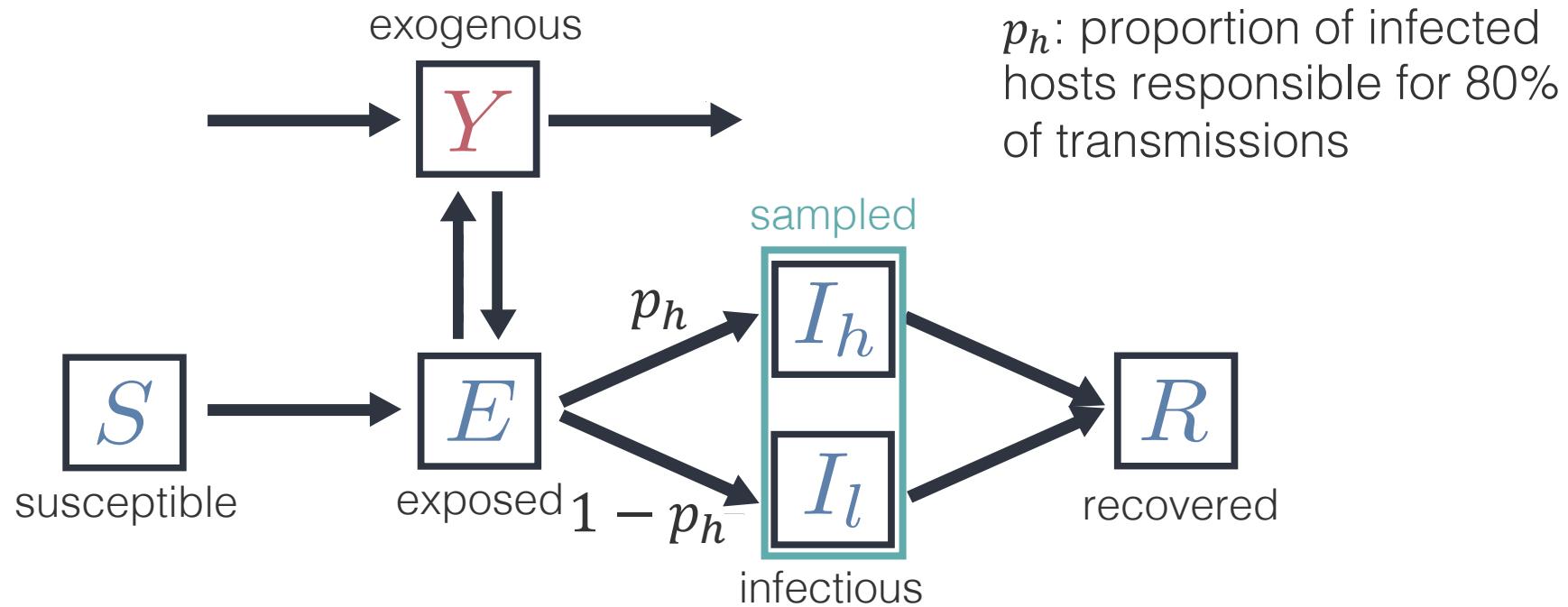
Dr. Adi Stern



Dr. Danielle Miller



Phylogenetics of SARS-CoV-2 in Israel, early 2020



Dr. Katia Koelle



Dr. Adi Stern

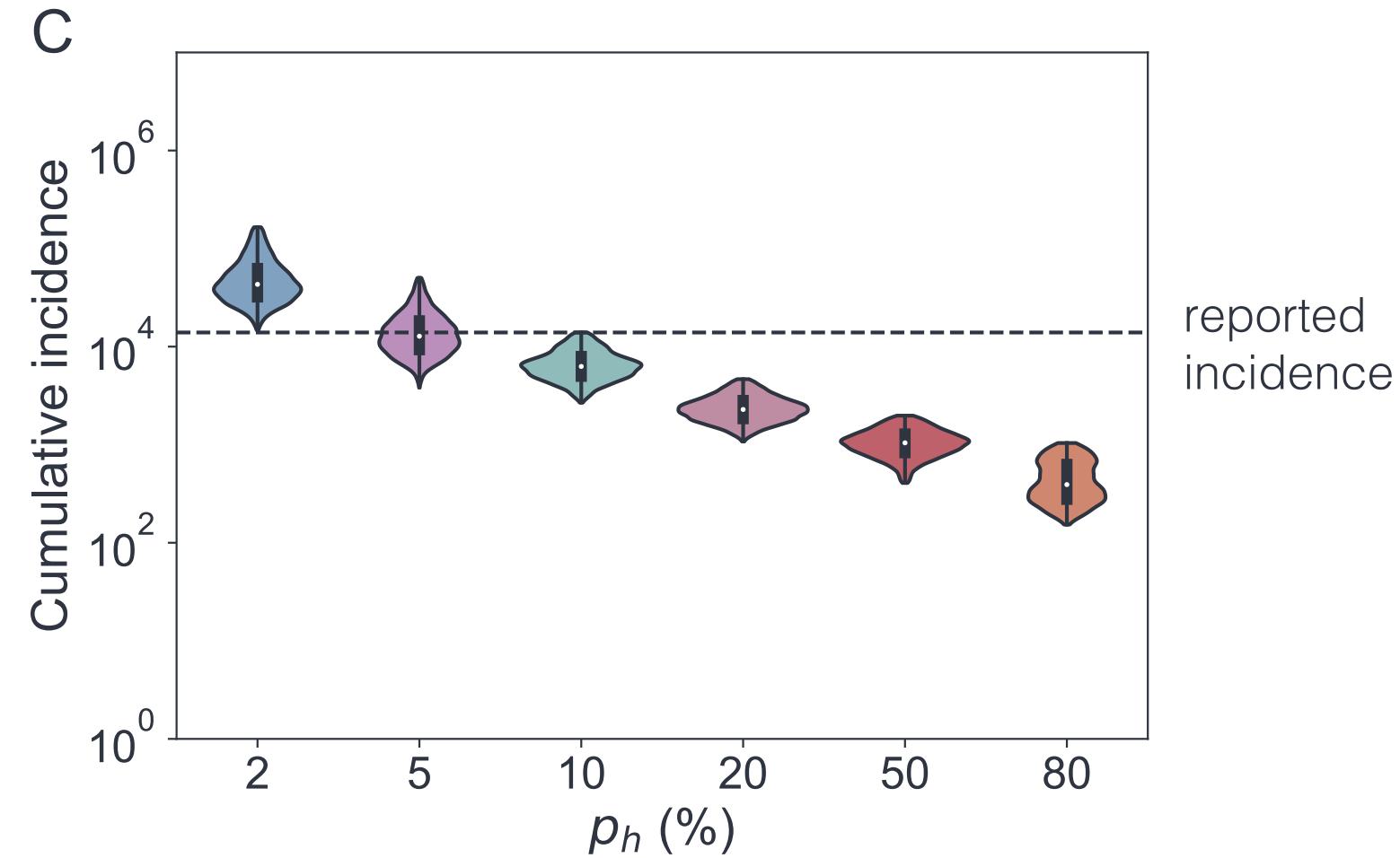


Dr. Danielle Miller



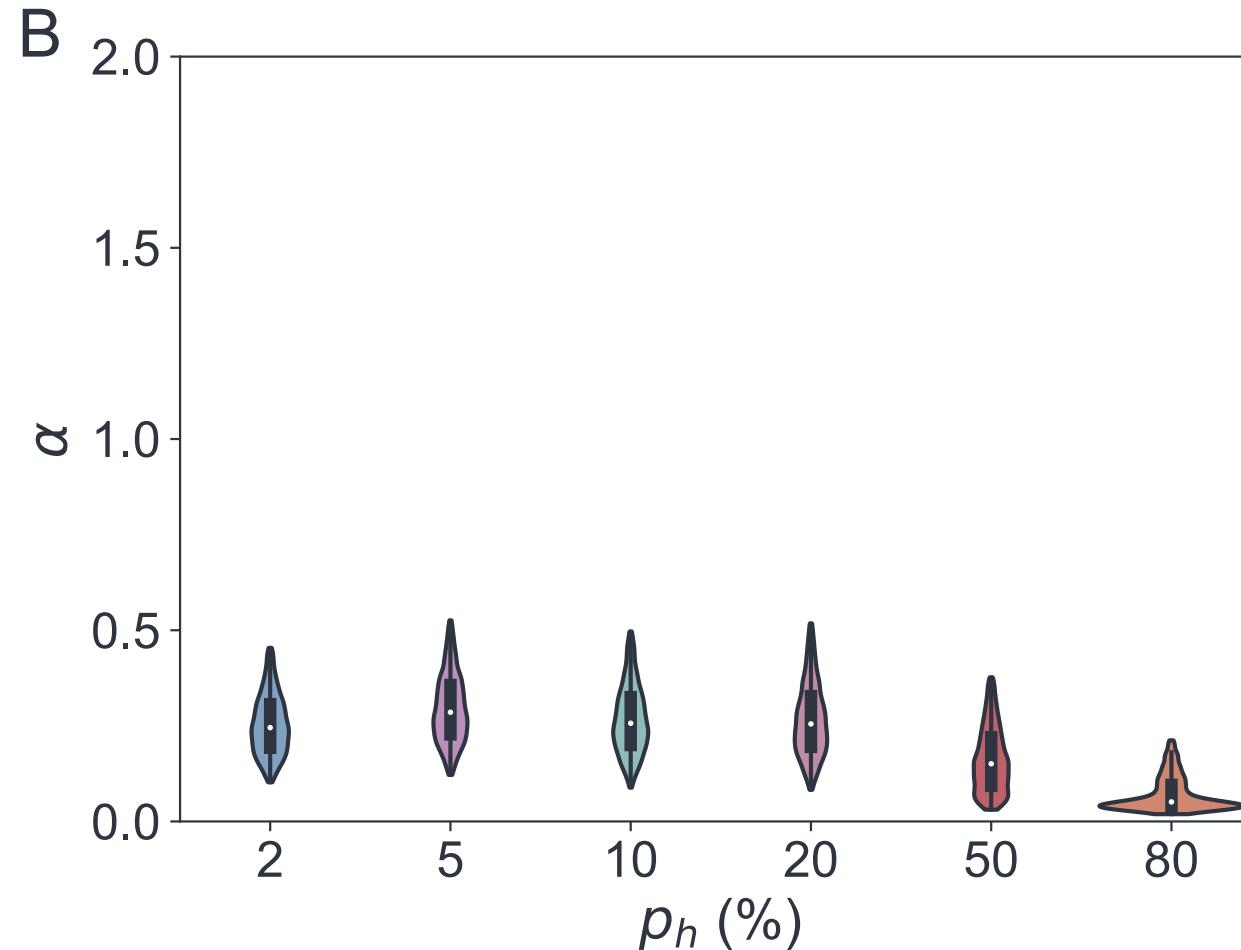
Phylogenetics of SARS-CoV-2 in Israel, early 2020

p_h : proportion of infected hosts responsible for 80% of transmissions

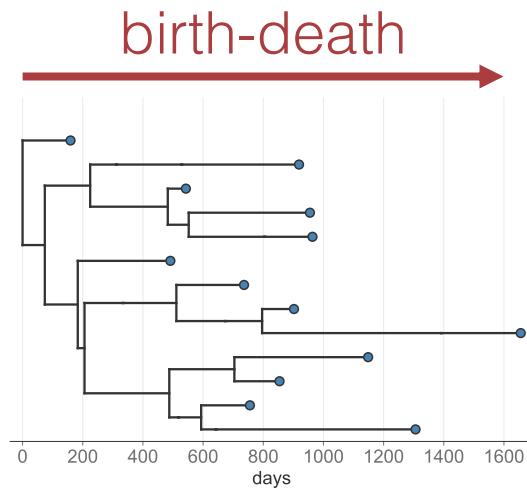


Phylogenetics of SARS-CoV-2 in Israel, early 2020

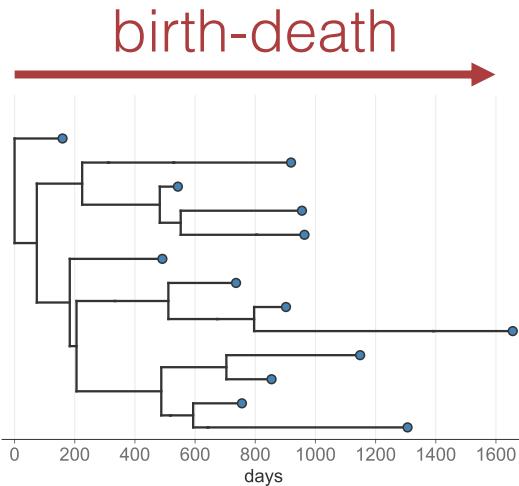
α : proportional
change in R_0 after
March 19th



Birth-death phylodynamic models look forward in time



Birth-death phylodynamic models look forward in time



x_i = infection events

y_i = sampling events

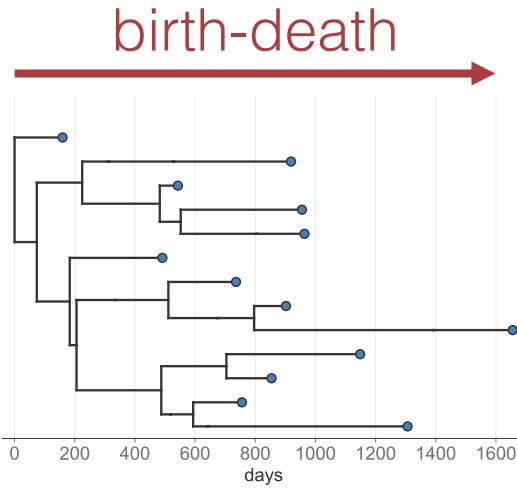
λ = transmission rate

ψ = sampling rate

μ = becoming infectious rate

$$R_0 = \frac{\lambda}{\psi + \mu}$$

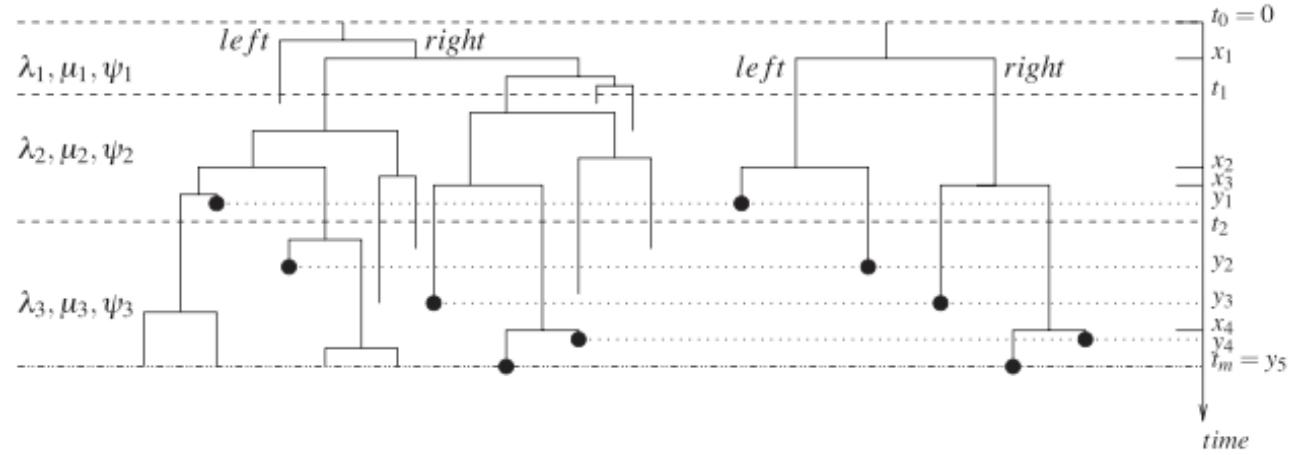
Birth-death phylodynamic models look forward in time



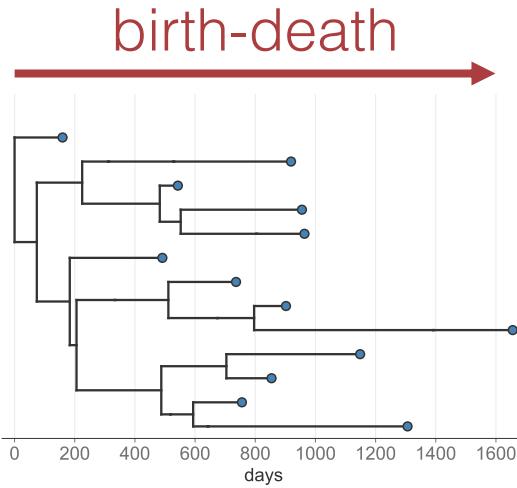
x_i = infection events
 y_i = sampling events

λ = transmission rate
 ψ = sampling rate
 μ = becoming infectious rate

$$R_0 = \frac{\lambda}{\psi + \mu}$$



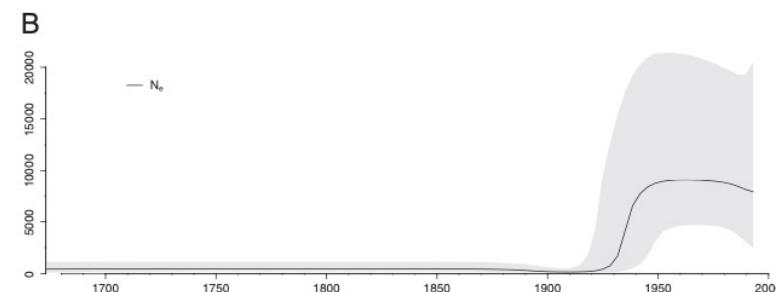
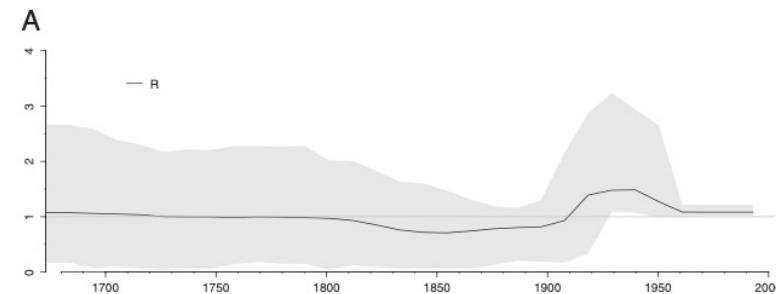
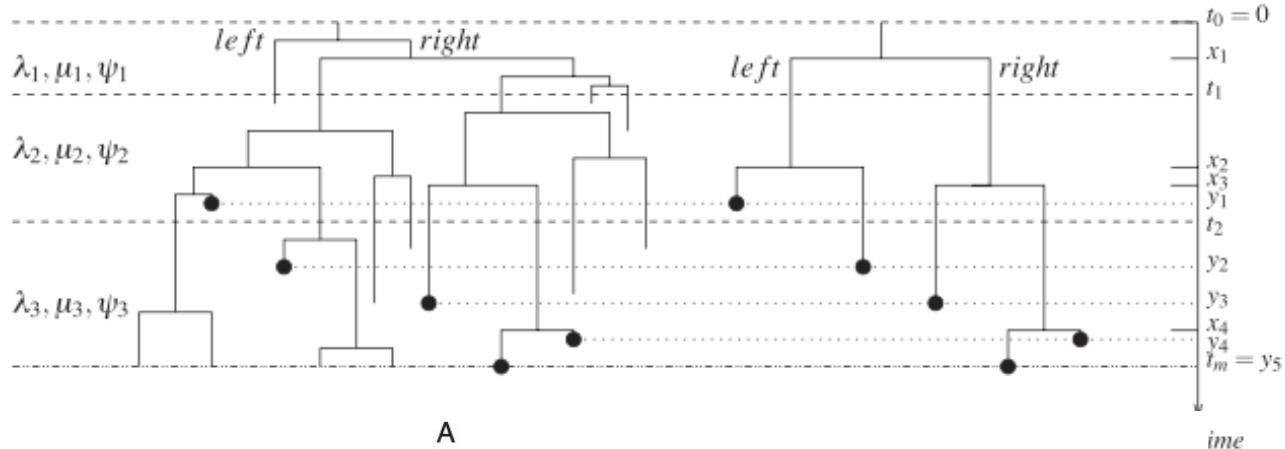
Birth-death phylodynamic models look forward in time



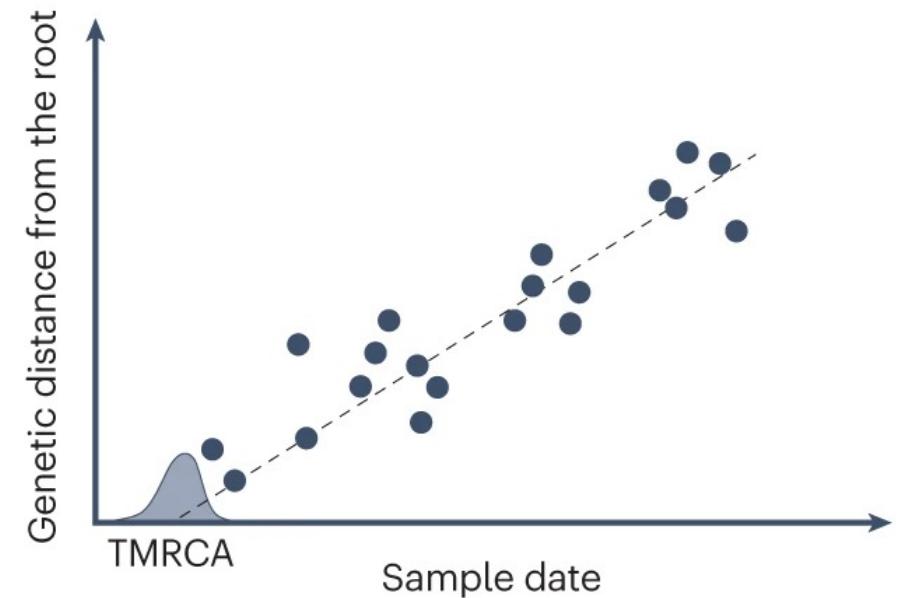
x_i = infection events
 y_i = sampling events

λ = transmission rate
 ψ = sampling rate
 μ = becoming infectious rate

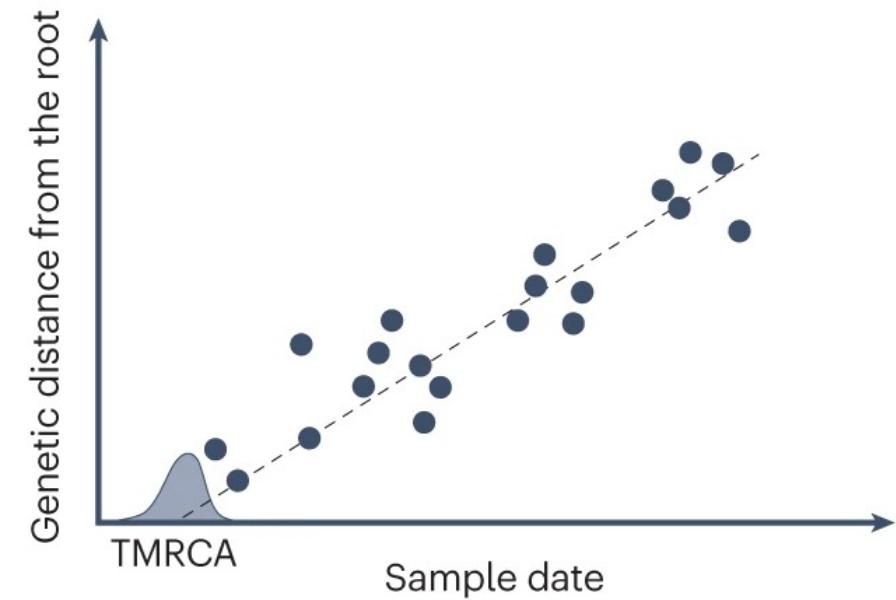
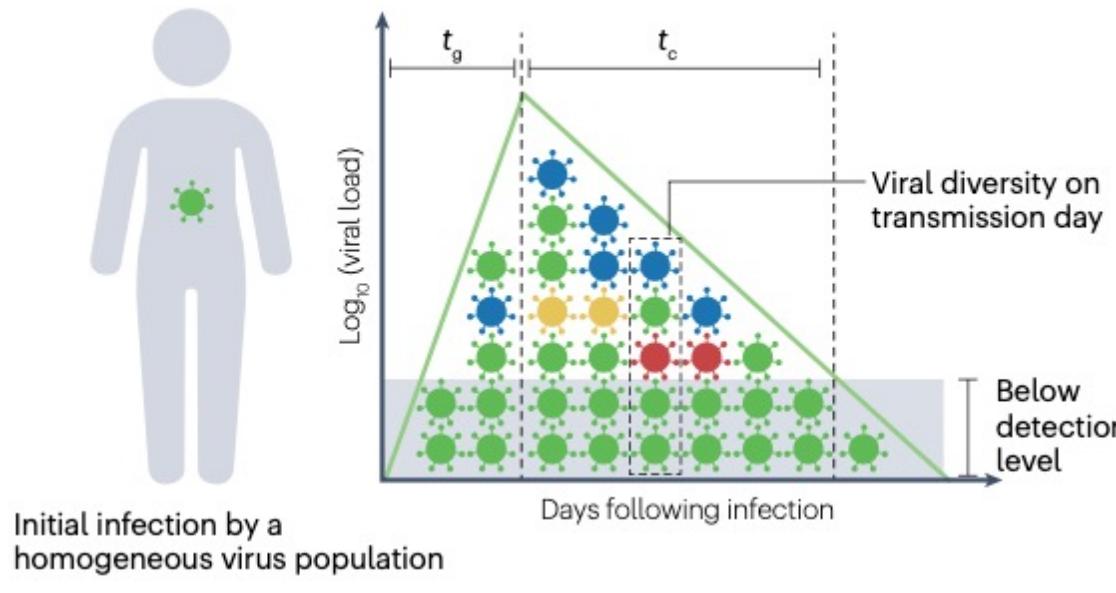
$$R_0 = \frac{\lambda}{\psi + \mu}$$



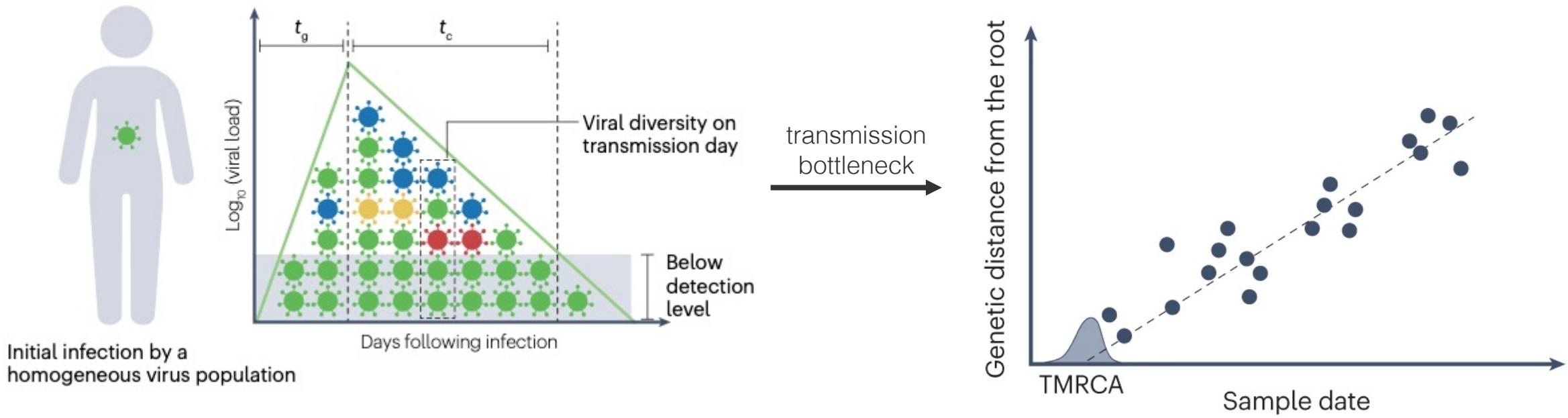
Mutations accumulate within and between hosts



Mutations accumulate within and between hosts



Mutations accumulate within and between hosts

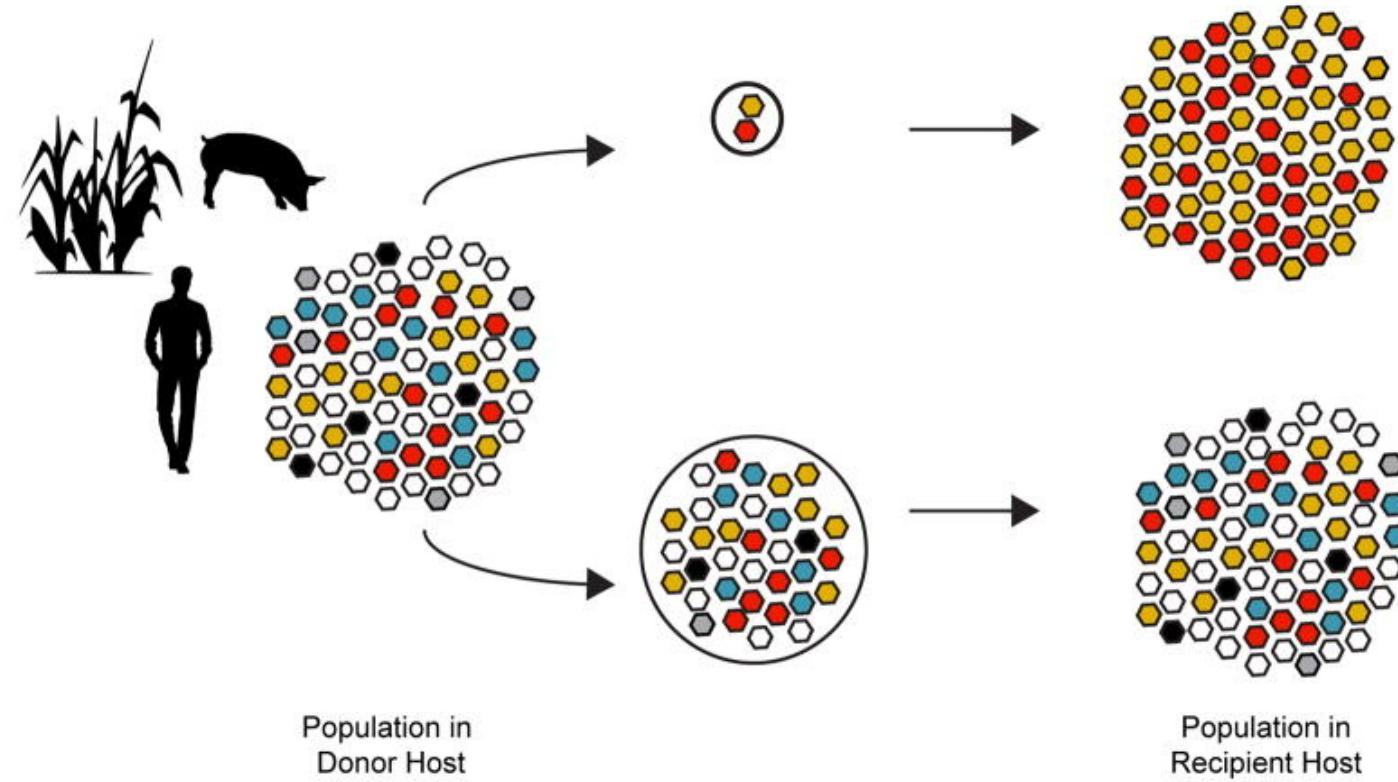


The transmission bottleneck modulates how genetic diversity is transmitted between hosts

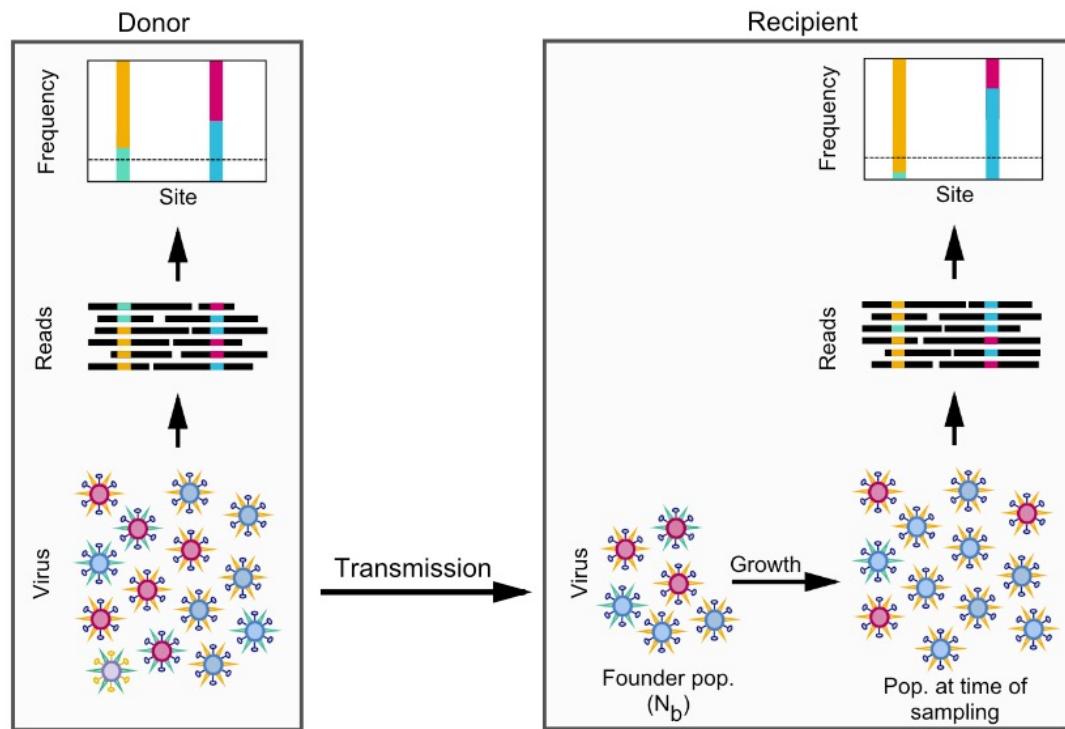
Transmission bottleneck: the number of virions (census) from a donor host that establish infection within a recipient host

The transmission bottleneck modulates how genetic diversity is transmitted between hosts

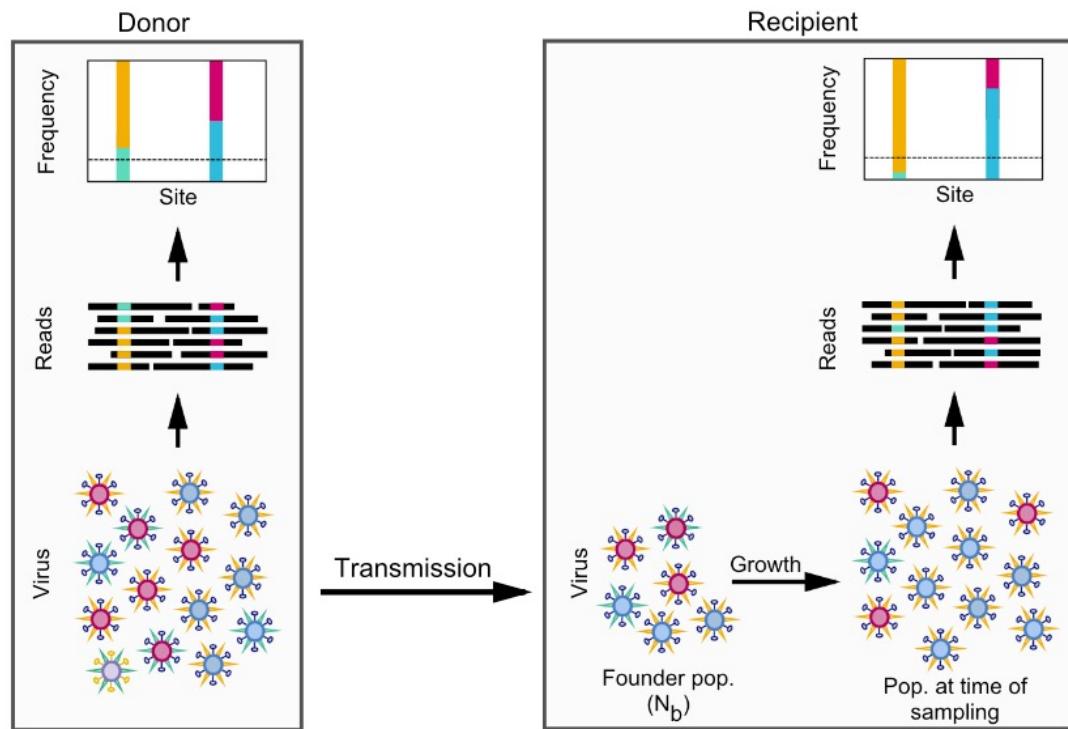
Transmission bottleneck: the number of virions (census) from a donor host that establish infection within a recipient host



Estimating transmission bottlenecks using the beta-binomial method



Estimating transmission bottlenecks using the beta-binomial method



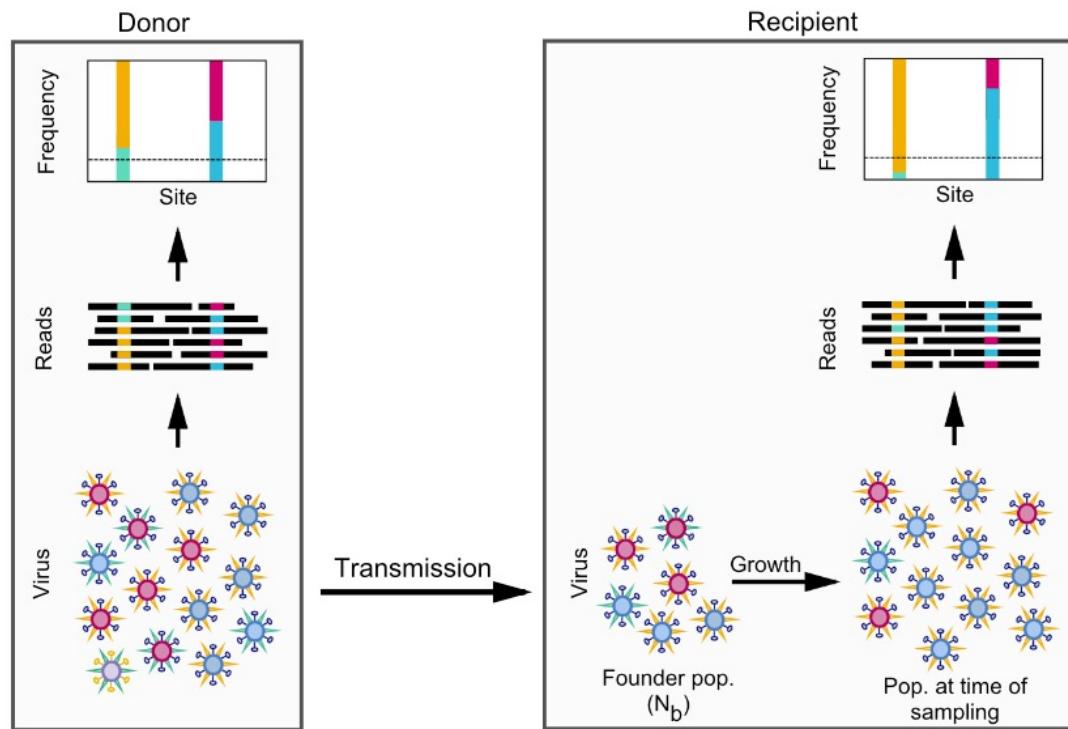
N_b = bottleneck size

k = founding virions
with variant i

v_d = donor variant i freq.

v_r = recipient variant i freq.

Estimating transmission bottlenecks using the beta-binomial method



N_b = bottleneck size

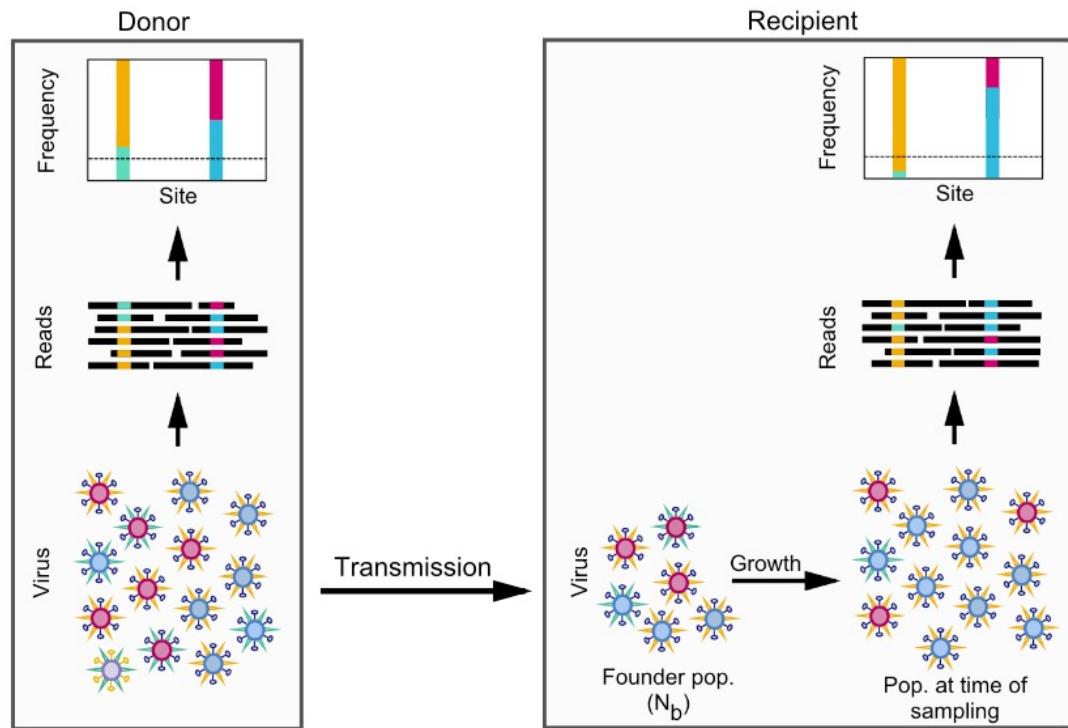
k = founding virions
with variant i

v_d = donor variant i freq.

v_r = recipient variant i freq.

$$k \sim Bin(N_b, v_d)$$

Estimating transmission bottlenecks using the beta-binomial method

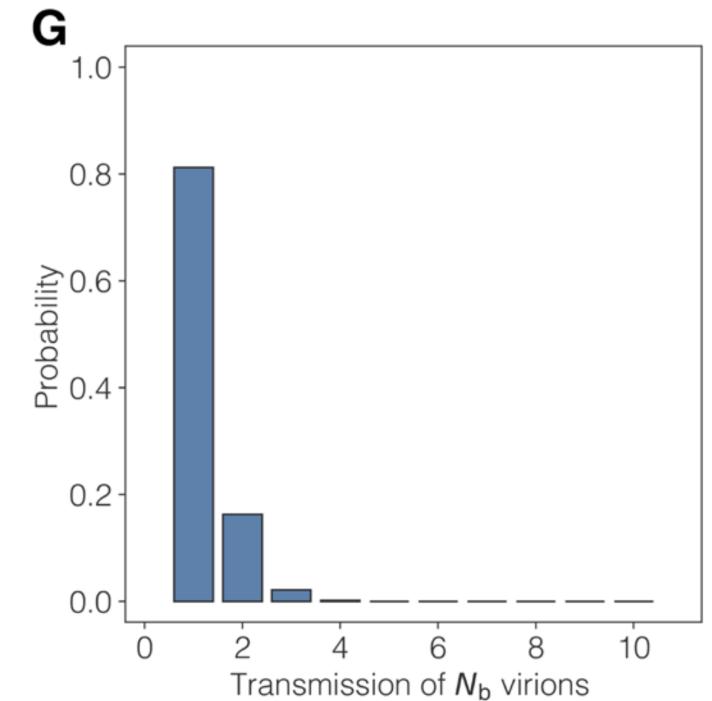
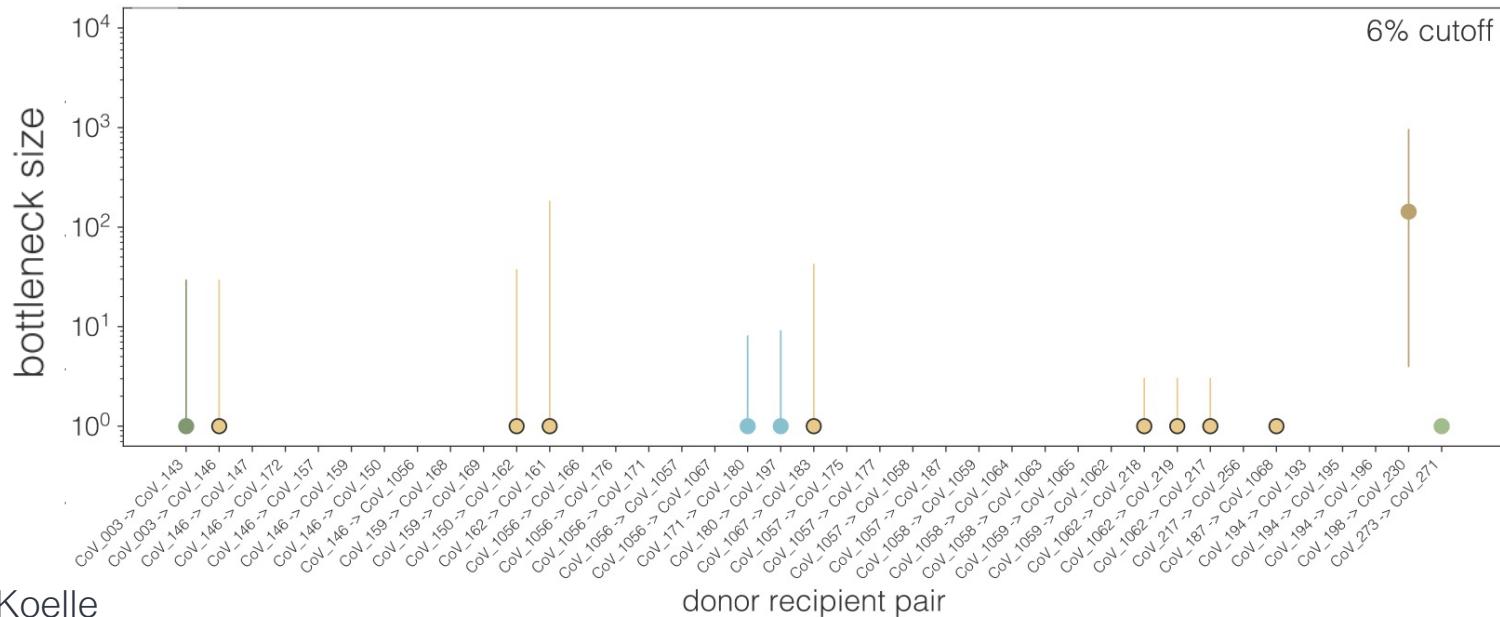


N_b = bottleneck size v_d = donor variant i freq.
 k = founding virions v_r = recipient variant i freq.
with variant i

$$k \sim Bin(N_b, v_d)$$

$$v_r \sim Beta(k, N_b - k)$$

The transmission bottleneck of SARS-CoV-2 is on the order of 1-3 virions



Dr. Katia Koelle



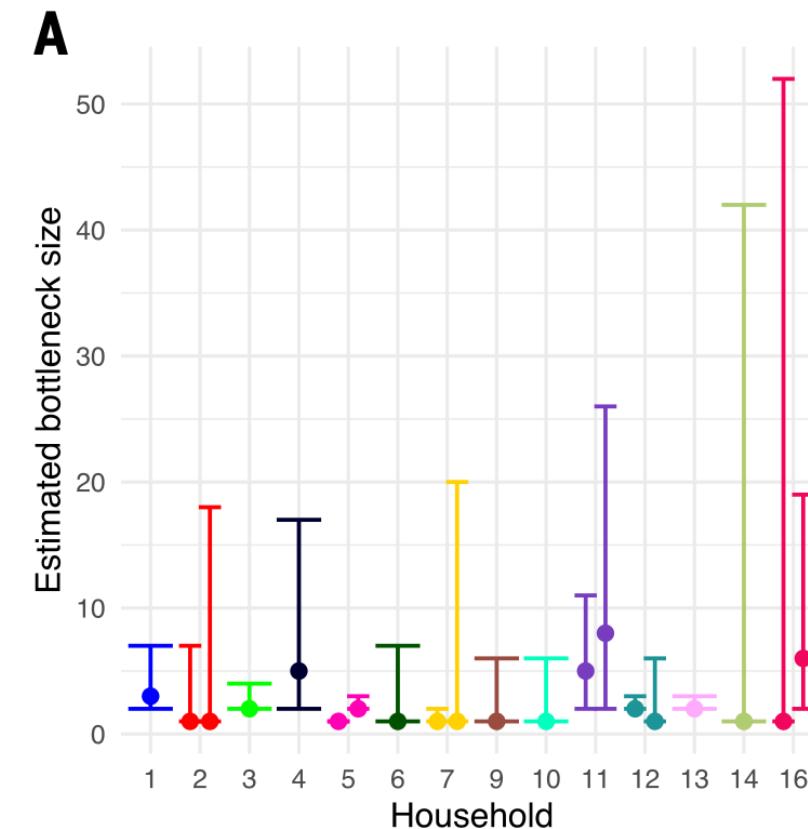
The transmission bottleneck of SARS-CoV-2 is on the order of 1-3 virions

RESEARCH ARTICLE

CORONAVIRUS

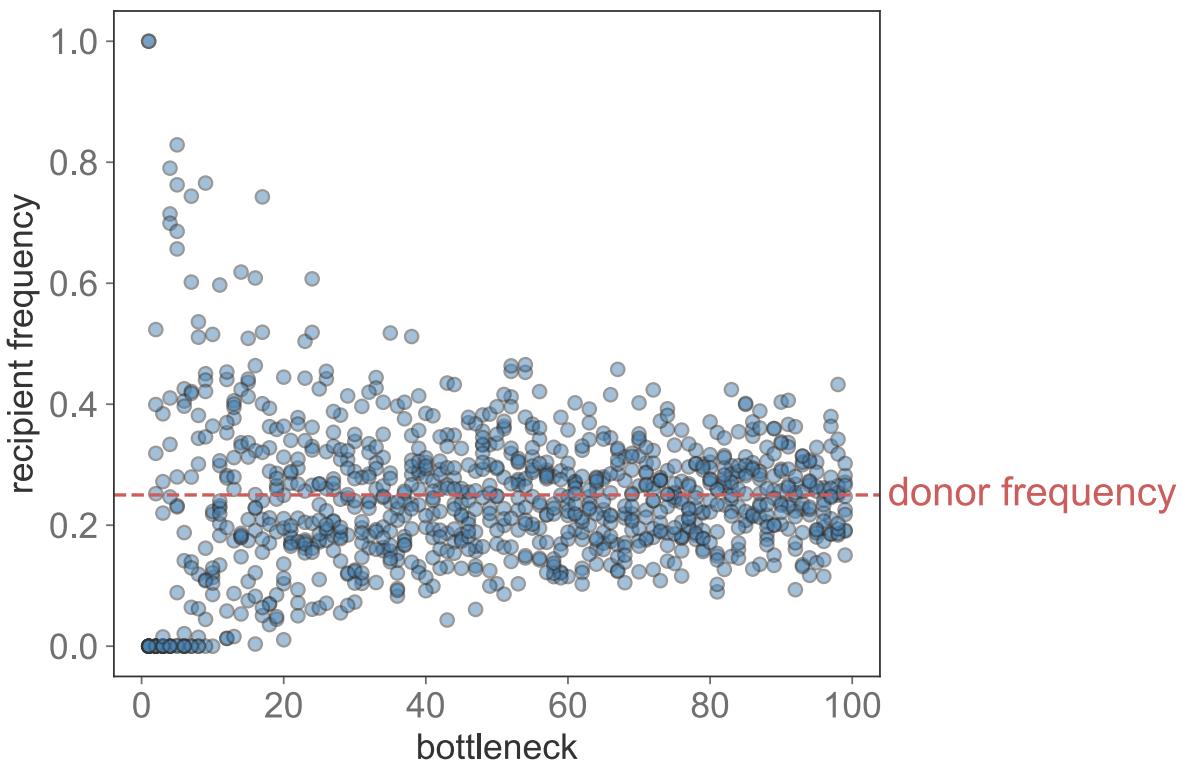
SARS-CoV-2 within-host diversity and transmission

Katrina A. Lythgoe^{1,2†*}, Matthew Hall^{1†*}, Luca Ferretti¹, Mariateresa de Cesare^{1,3}, George MacIntyre-Cockett^{1,3}, Amy Trebes³, Monique Andersson^{4,5}, Newton Otecko¹, Emma L. Wise^{6,7}, Nathan Moore⁶, Jessica Lynch⁶, Stephen Kidd⁶, Nicholas Cortes^{6,8}, Matilde Mori⁹, Rebecca Williams⁶, Gabrielle Vernet⁶, Anita Justice⁴, Angie Green³, Samuel M. Nicholls¹⁰, M. Azim Ansari¹¹, Lucie Abeler-Dörner¹, Catrin E. Moore¹, Timothy E. A. Peto^{4,12}, David W. Eyre^{4,13}, Robert Shaw⁴, Peter Simmonds¹¹, David Buck³, John A. Todd³ on behalf of the Oxford Virus Sequencing Analysis Group (OVSG)‡, Thomas R. Connor^{14,15}, Shirin Ashraf¹⁶, Ana da Silva Filipe¹⁶, James Shepherd¹⁶, Emma C. Thomson¹⁶, The COVID-19 Genomics UK (COG-UK) Consortium§, David Bonsall^{1,3,4}, Christophe Fraser^{1,3,17}, Tanya Golubchik^{1,2,*}



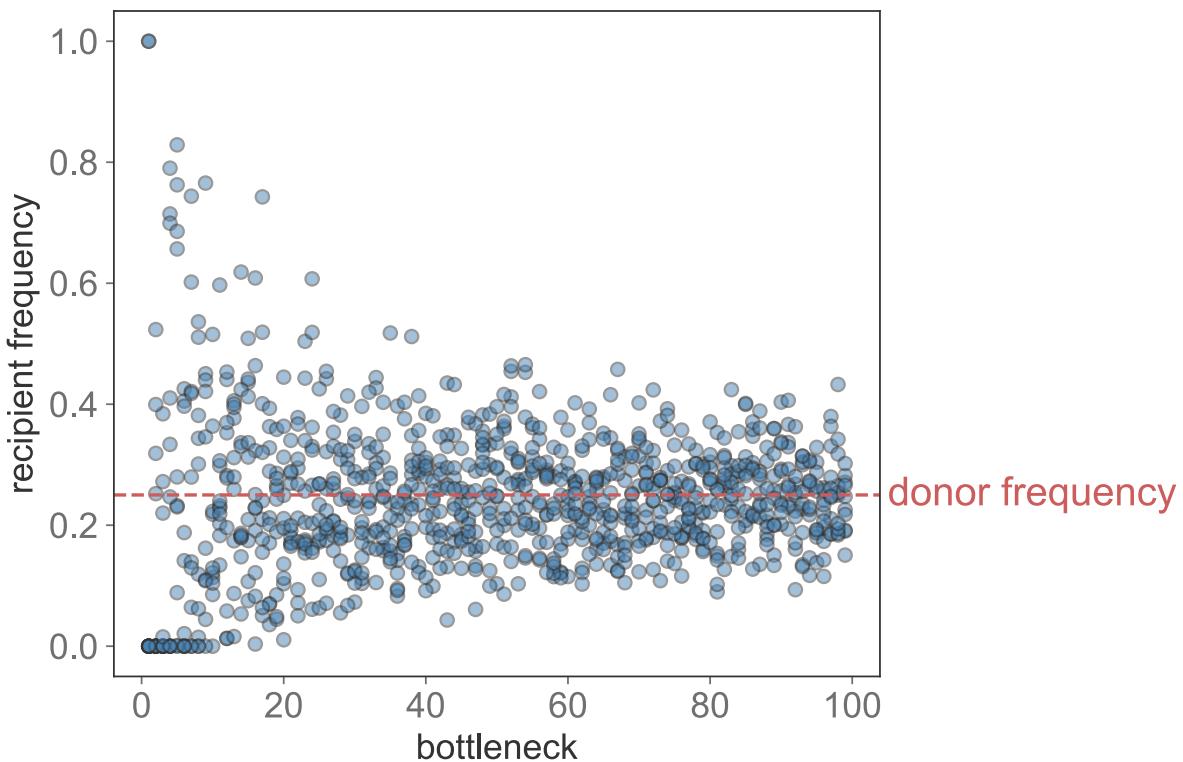
Tight transmission bottlenecks increase stochasticity at the point of transmission

Introduces noise into the frequency of transmitted donor variants



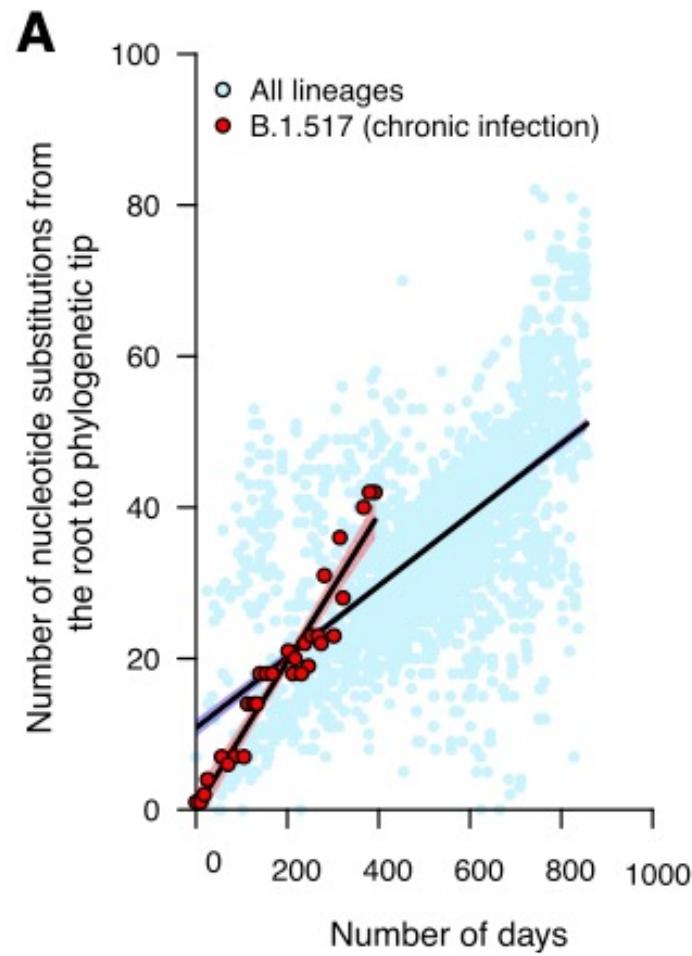
Tight transmission bottlenecks increase stochasticity at the point of transmission

Introduces noise into the frequency of transmitted donor variants

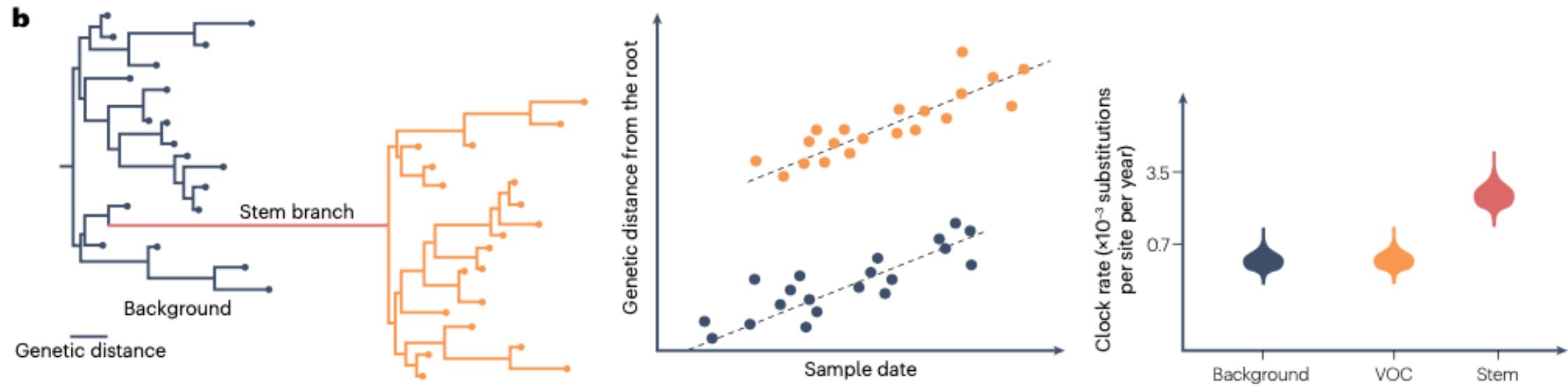


Stochasticity at the transmission bottleneck slows adaptation

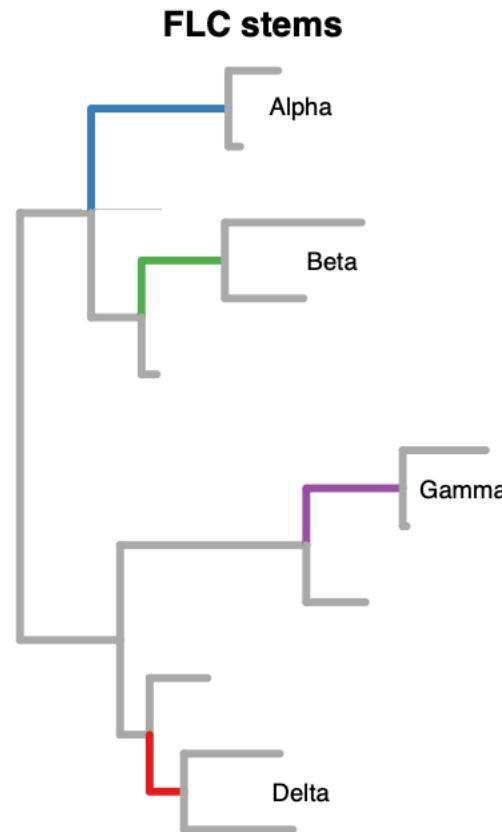
The lack of a bottleneck during chronic infections accelerates the substitution rate



Accelerated within-host evolution (maybe?) leaves characteristic phylogenetic signatures



Identification of accelerated evolution using local clock models



The Emergence of SARS-CoV-2 Variants of Concern Is Driven by Acceleration of the Substitution Rate

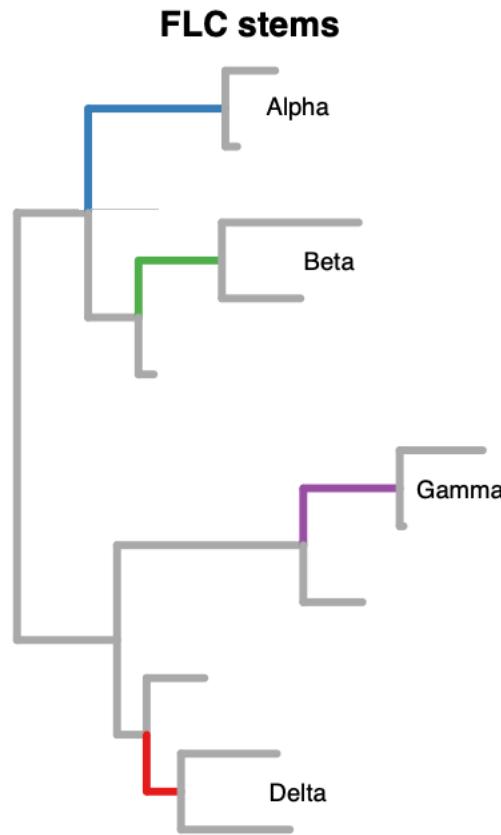
John H. Tay, Ashleigh F. Porter, Wytamma Wirth, and Sebastian Duchene *

Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC, Australia

*Corresponding author: E-mail: sduchene@unimelb.edu.au.

Associate editor: Thomas Leitner

Identification of accelerated evolution using local clock models



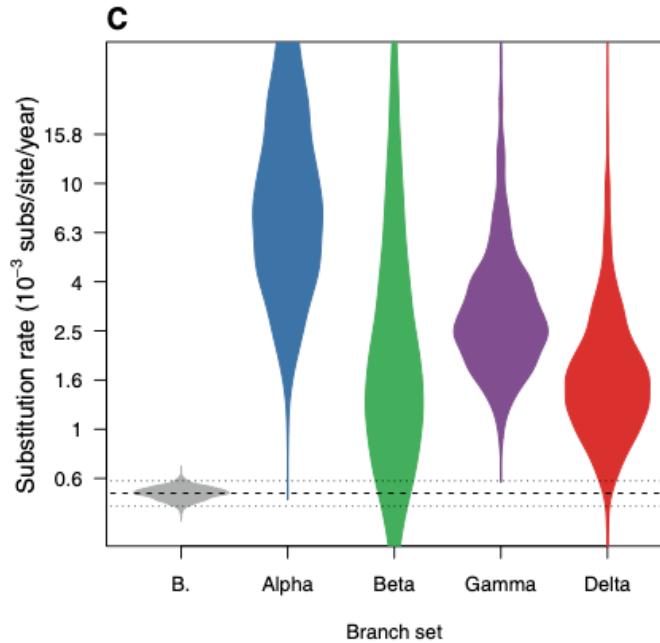
The Emergence of SARS-CoV-2 Variants of Concern Is Driven by Acceleration of the Substitution Rate

John H. Tay, Ashleigh F. Porter, Wytamma Wirth, and Sebastian Duchene *

Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC, Australia

*Corresponding author: E-mail: sduchene@unimelb.edu.au.

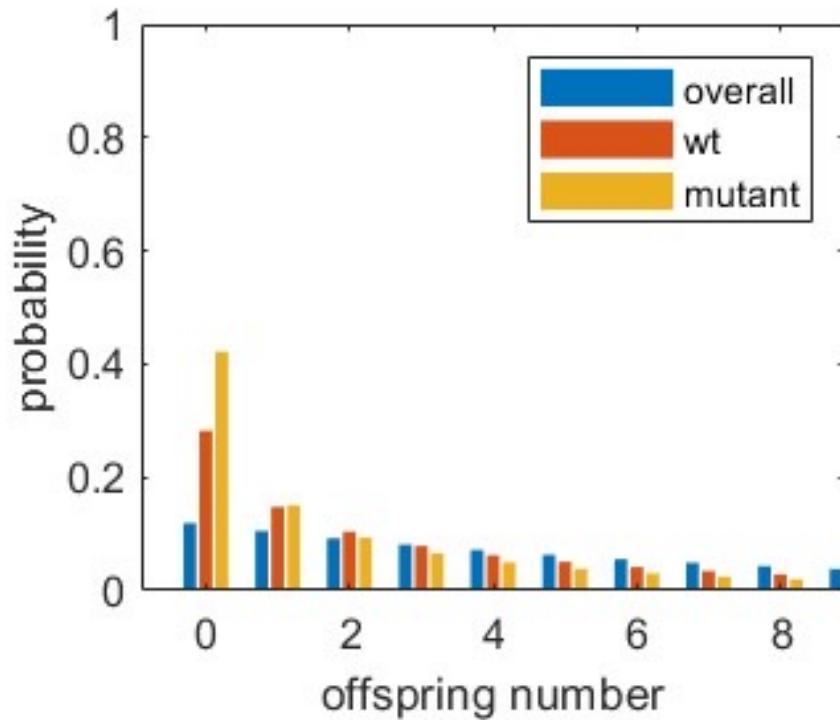
Associate editor: Thomas Leitner



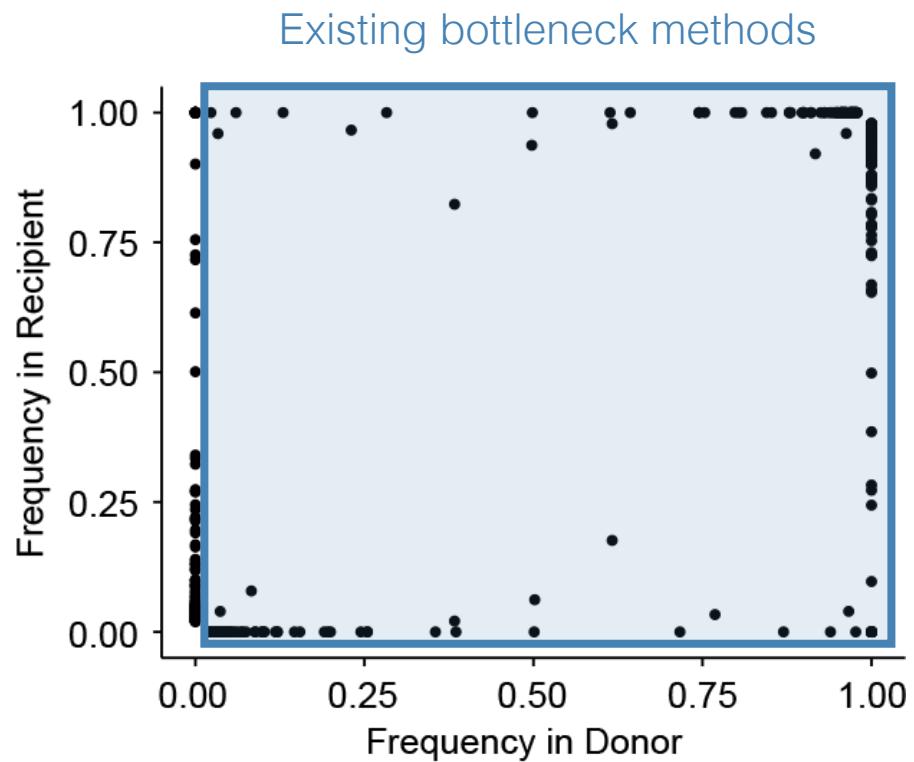
Tight transmission bottlenecks increase stochasticity at the point of transmission

Allows for the generation of new variants

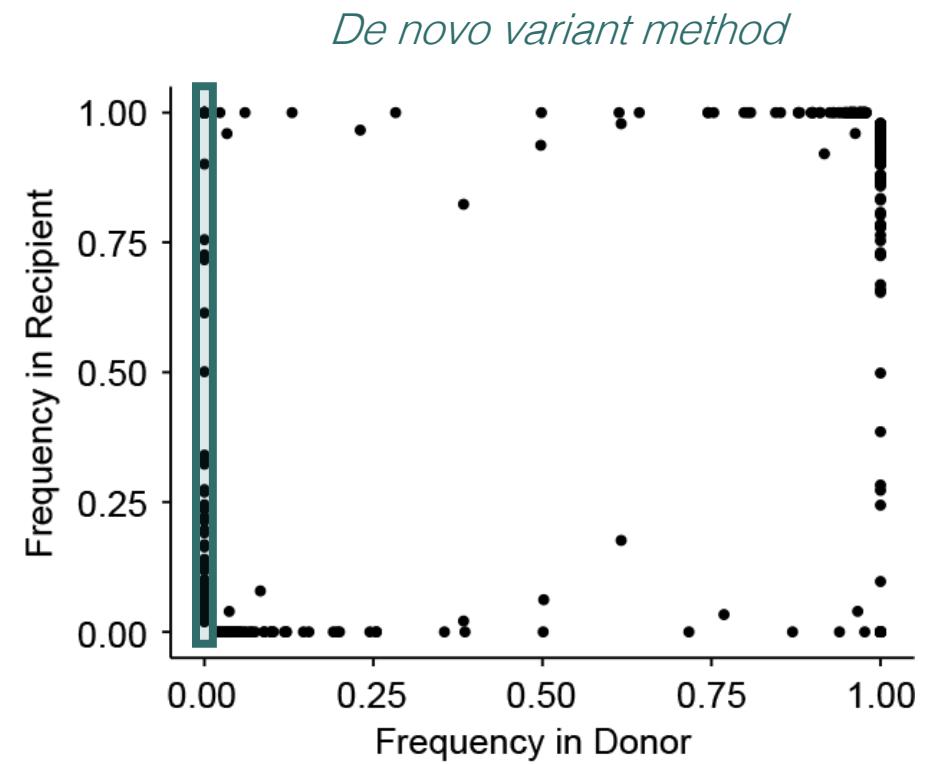
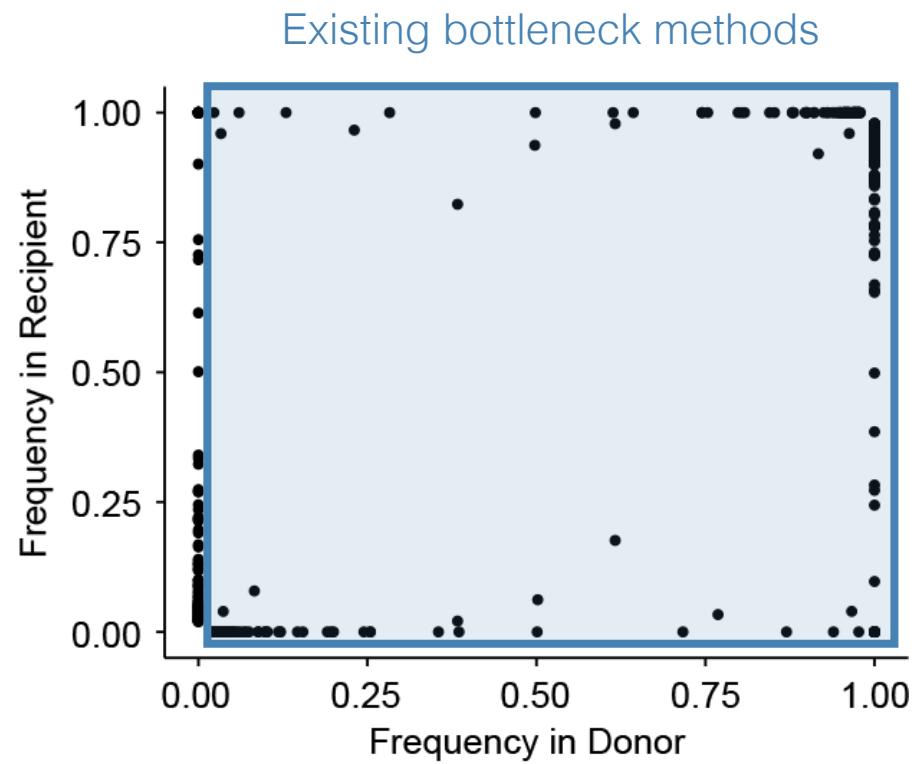
The bottleneck size affects the distribution of *de novo* variants in the recipient host



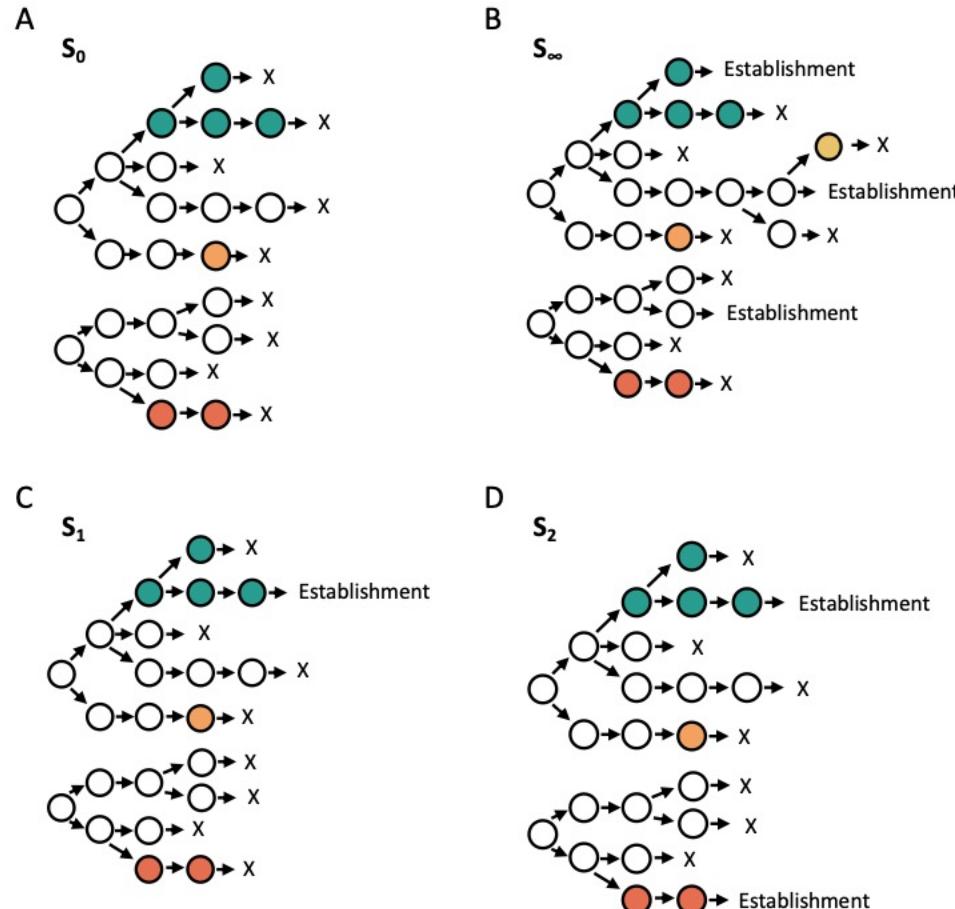
The bottleneck size affects the distribution of *de novo* variants in the recipient host



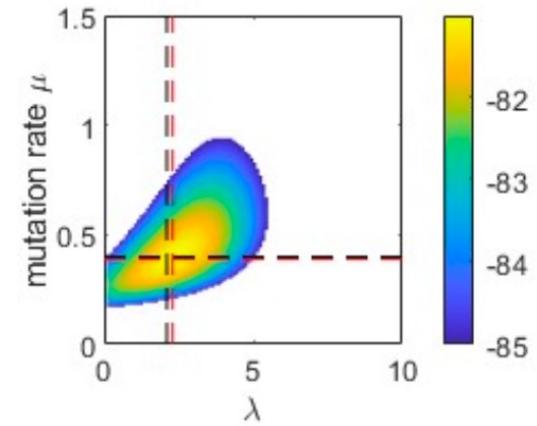
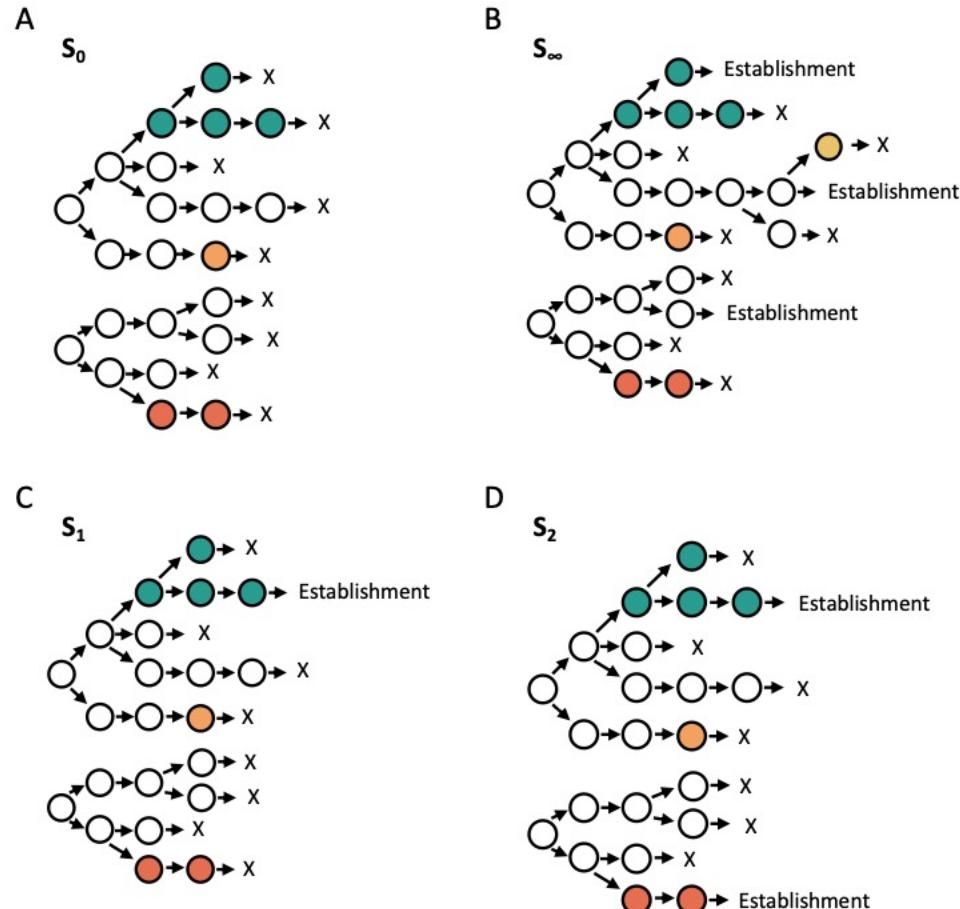
The bottleneck size affects the distribution of *de novo* variants in the recipient host



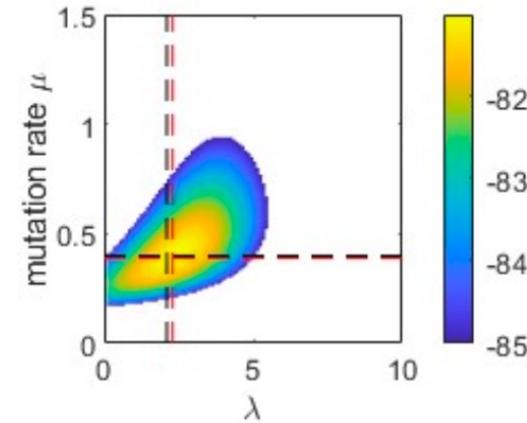
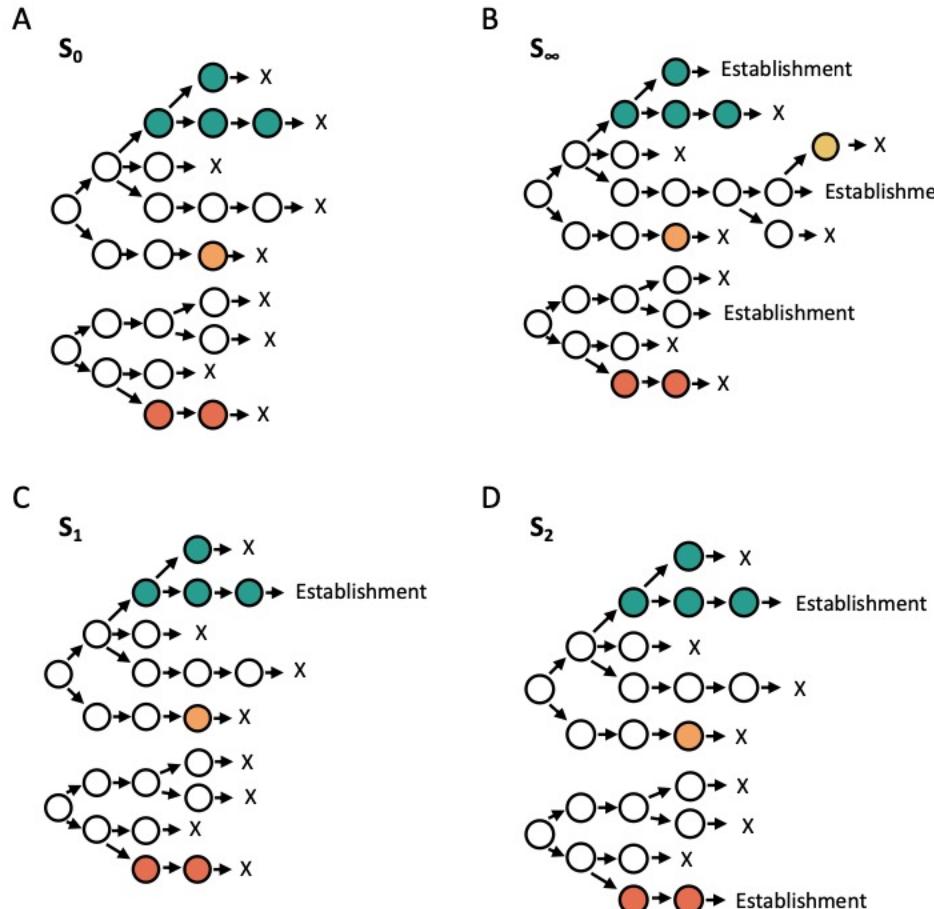
The bottleneck size affects the distribution of *de novo* variants in the recipient host



The bottleneck size affects the distribution of *de novo* variants in the recipient host



The bottleneck size affects the distribution of *de novo* variants in the recipient host



Transmission bottleneck size estimation from *de novo* viral genetic variation

Teresa Shi¹, Jeremy D. Harris^{1,2}, Michael A. Martin^{1,3,4}, Katia Koelle^{1,5,*}

1 Department of Biology, Emory University, Atlanta, GA, USA

2 Department of Mathematics, Rose-Hulman Institute of Technology, Terre Haute, IN, USA

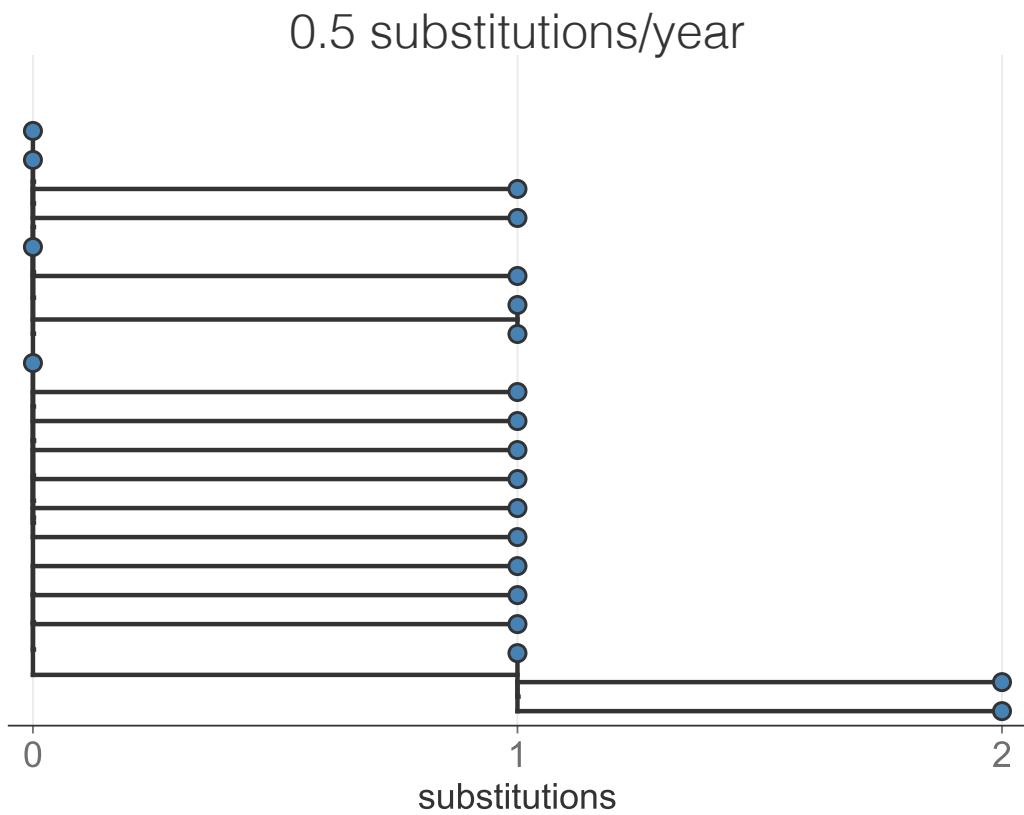
3 Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD, USA

4 Graduate Program in Population Biology, Ecology, and Evolution, Emory University, Atlanta, GA, USA

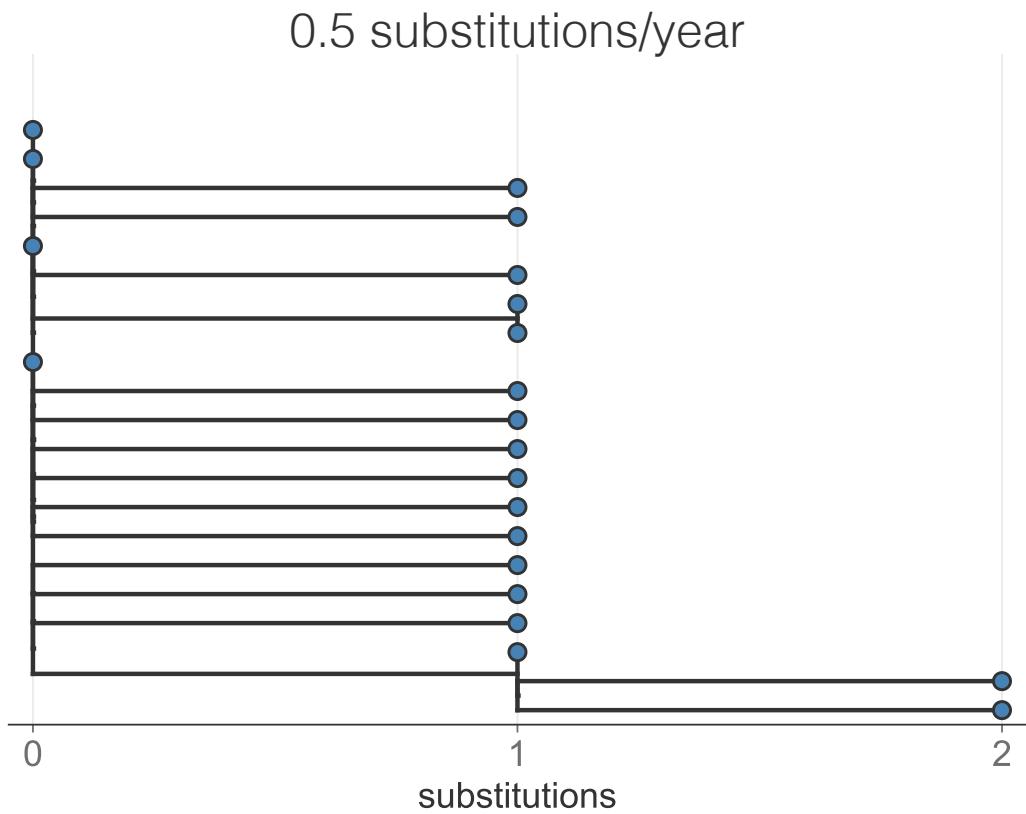
5 Emory Center of Excellence for Influenza Research and Response (CEIRR), Atlanta GA, USA

*katia.koelle@emory.edu

When genetic diversity is low, phylogenetic uncertainty is high



When genetic diversity is low, phylogenetic uncertainty is high



Uncertainty in phylogenetic structure & evolutionary model

=

Uncertainty in inferred demographic parameters

When genetic diversity is low, phylogenetic uncertainty is high

BIOINFORMATICS **APPLICATIONS NOTE**

Vol. 26 no. 10 2010, pages 1372–1373
doi:10.1093/bioinformatics/btq110

Phylogenetics

Advance Access publication March 12, 2010

DensiTree: making sense of sets of phylogenetic trees

Remco R. Bouckaert

Department of Computer Science, Auckland University

Associate Editor: Martin Bishop

When genetic diversity is low, phylogenetic uncertainty is high

BIOINFORMATICS APPLICATIONS NOTE

Vol. 26 no. 10 2010, pages 1372–1373
doi:10.1093/bioinformatics/btq110

Phylogenetics

Advance Access publication March 12, 2010

DensiTree: making sense of sets of phylogenetic trees

Remco R. Bouckaert

Department of Computer Science, Auckland University

Associate Editor: Martin Bishop



Alternative, phylogenetic-free approach

AGATGATCAT

AGATAATCAT

CGATGATCAA

AGATAATCAA

CGATAATCAA

S = 3

Dr. Katia Koelle



Yeongseon Park



Alternative, phylogenetic-free approach

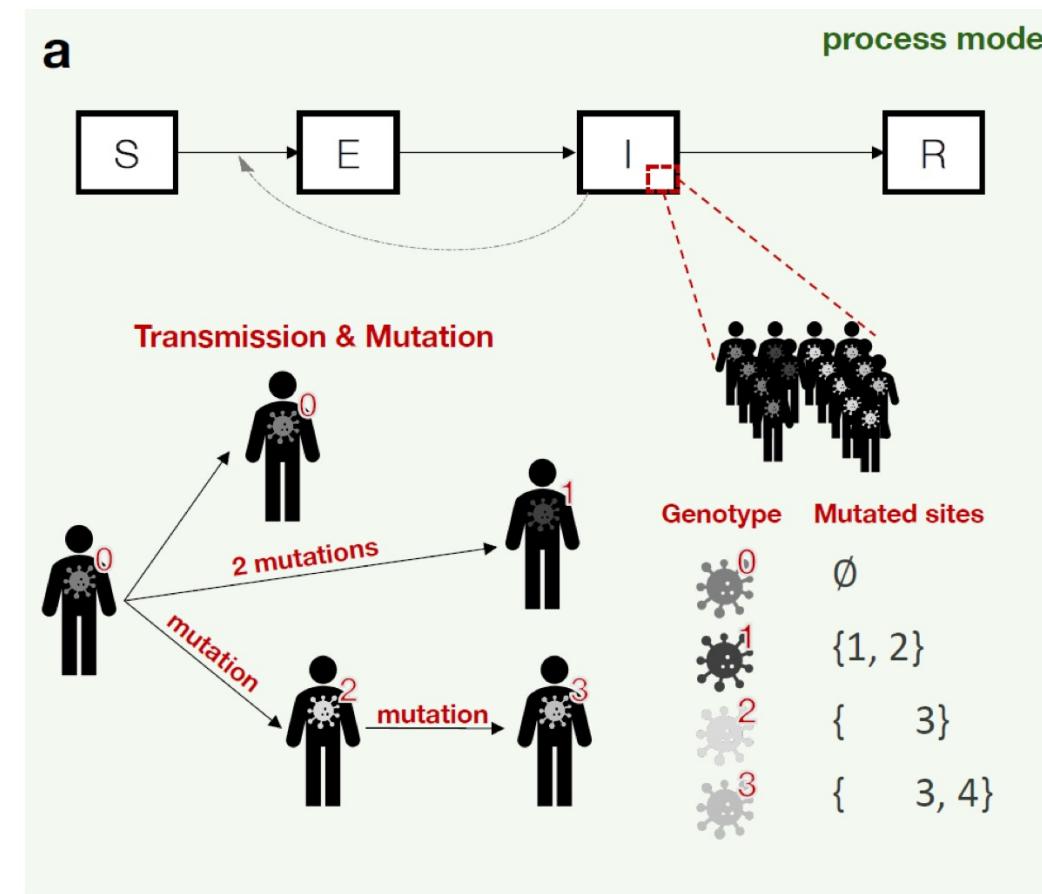
AGATGATCAT
AGATAATCAT
CGATGATCAA
AGATAATCAA
CGATAATCAA

$$S = 3$$

Dr. Katia Koelle



Yeongseon Park



Alternative, phylogenetic-free approach

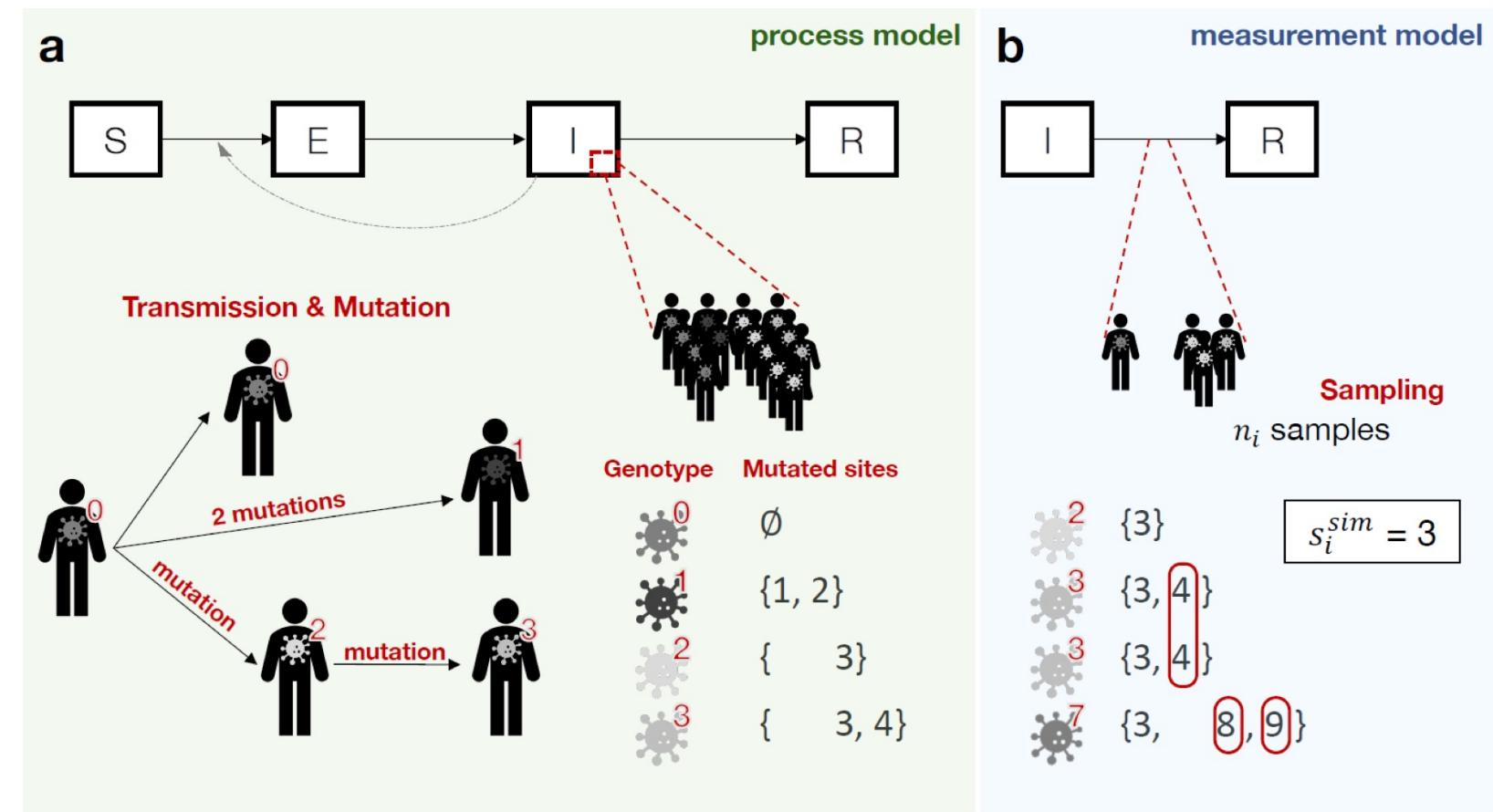
AGATGATCAT
AGATAATCAT
CGATGATCAA
AGATAATCAA
CGATAATCAA

$$S = 3$$

Dr. Katia Koelle

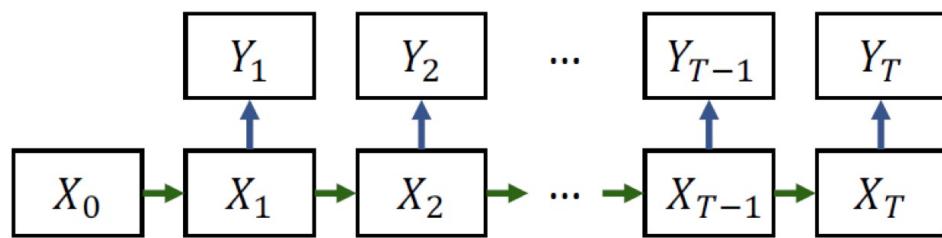
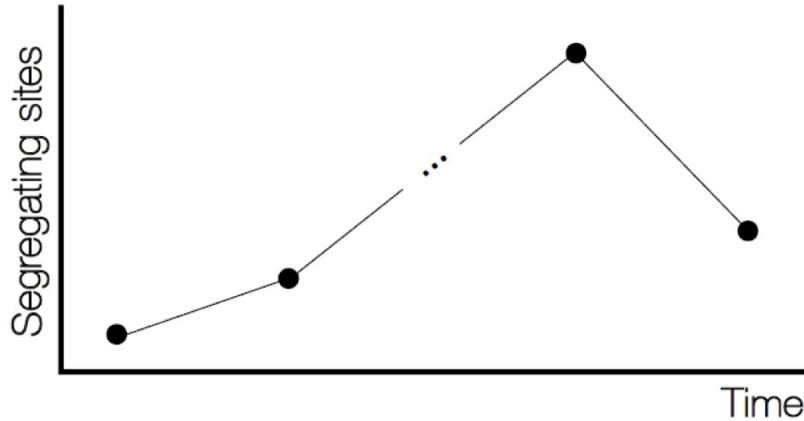


Yeongseon Park

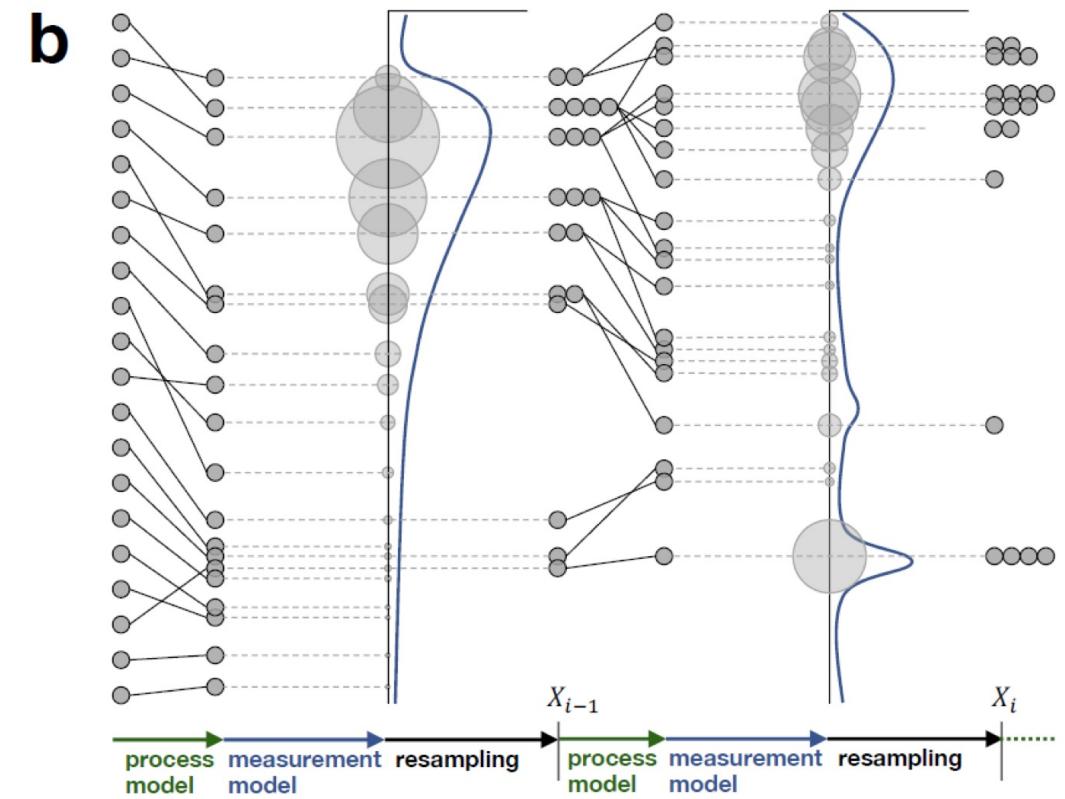
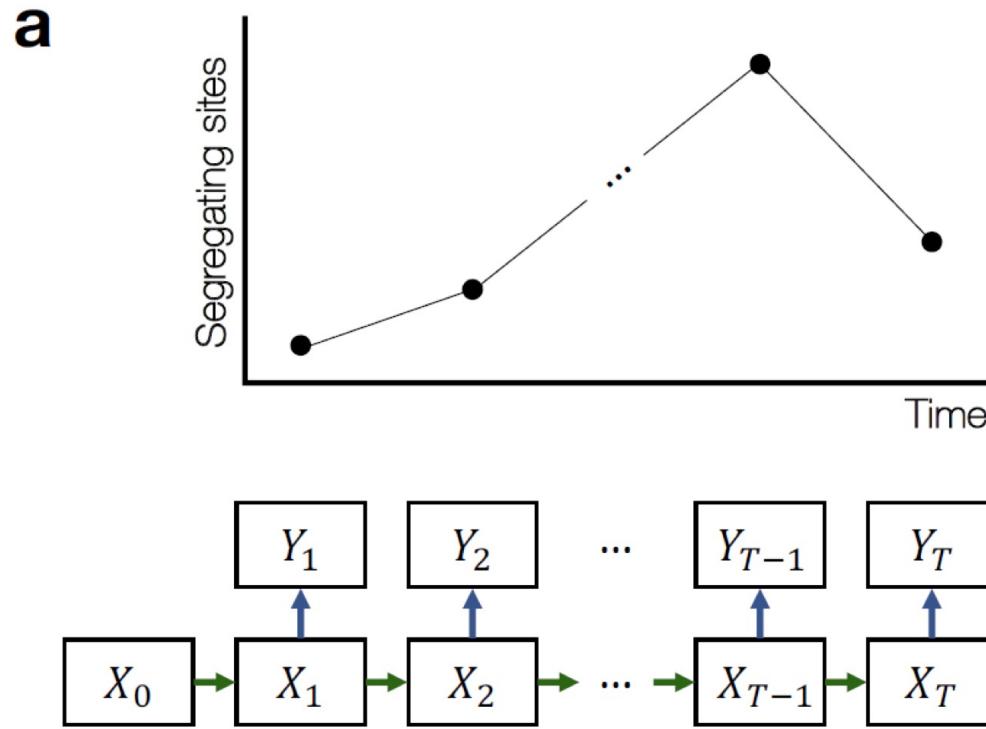


Alternative, phylogenetic-free approach

a

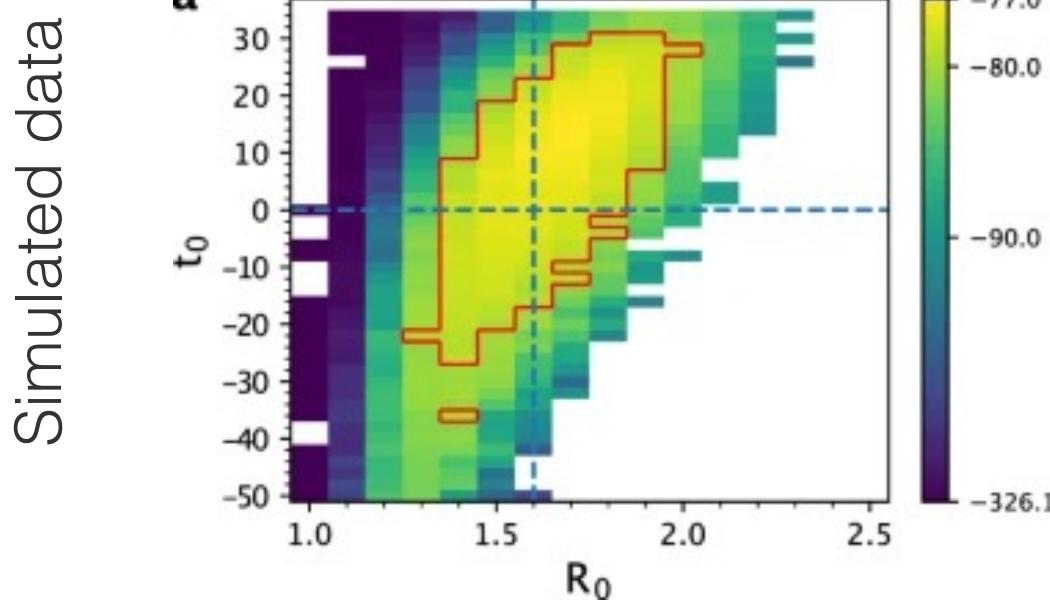


Alternative, phylogenetic-free approach

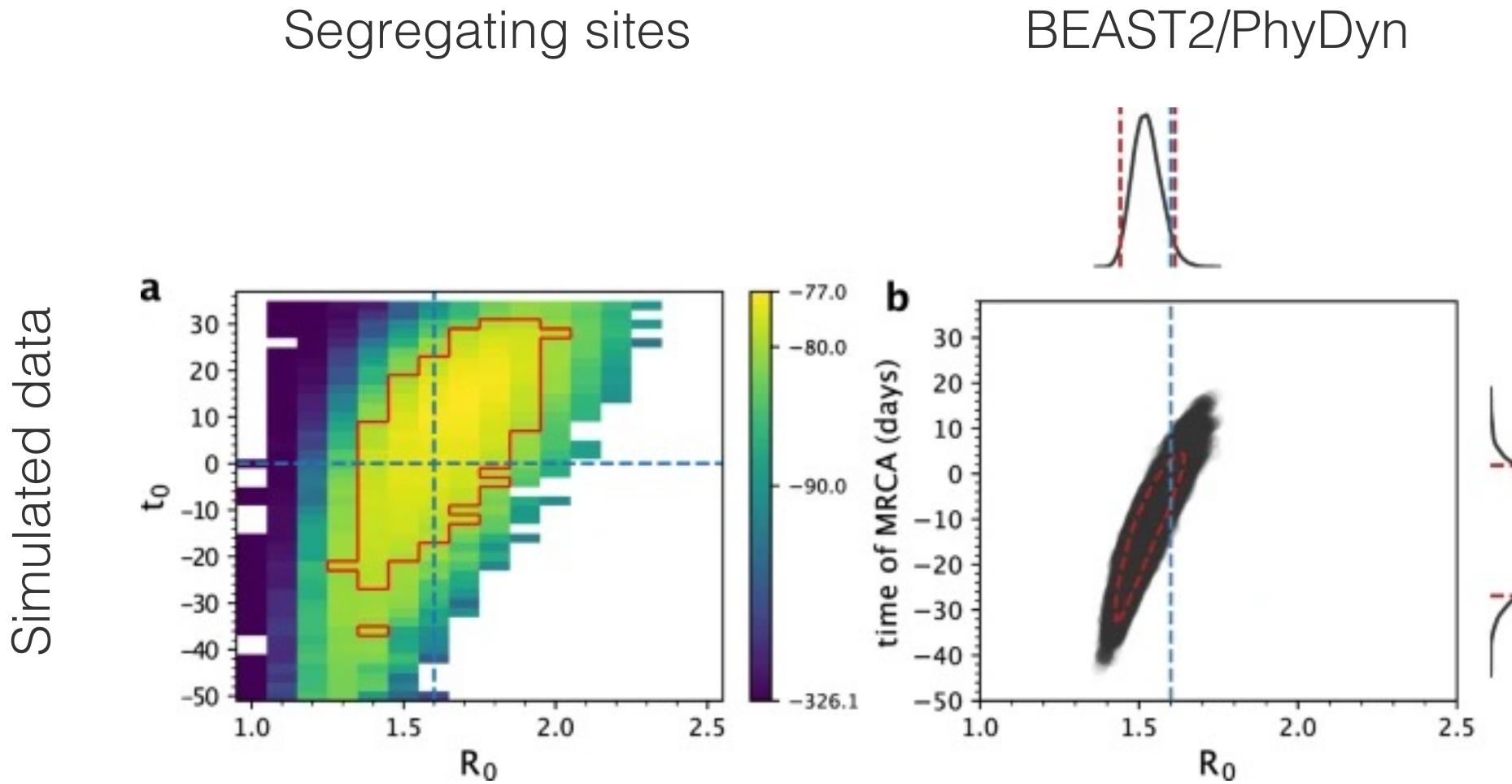


Modeling epidemiological dynamics using segregating sites

Segregating sites

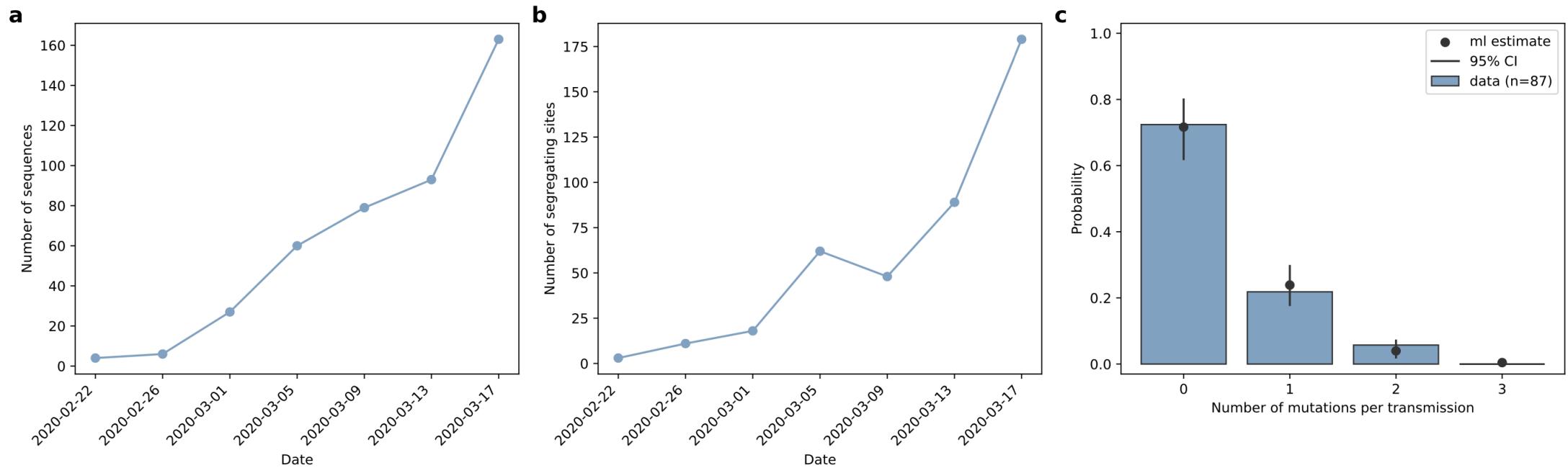


Modeling epidemiological dynamics using segregating sites

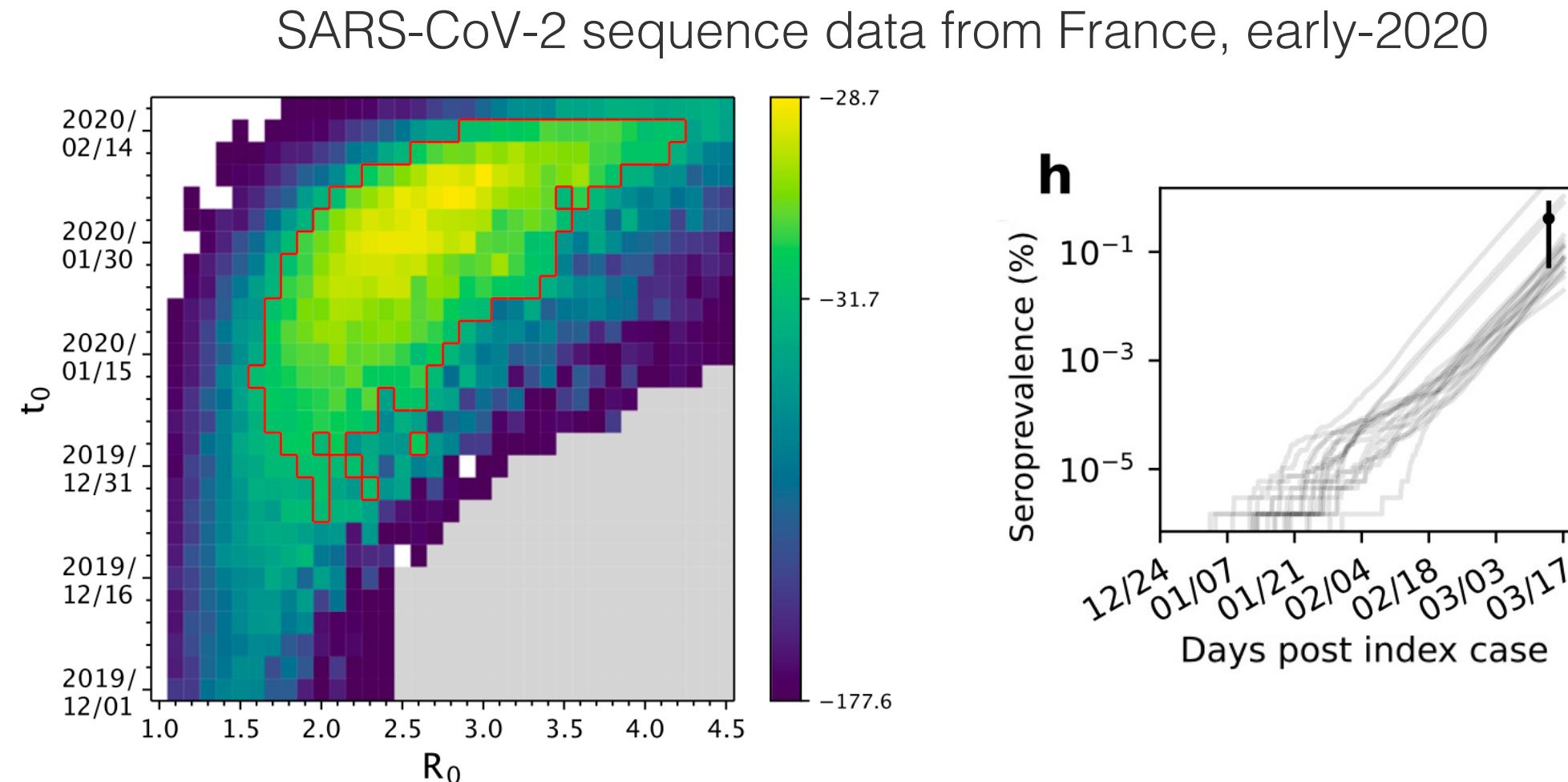


Modeling epidemiological dynamics using segregating sites

SARS-CoV-2 sequence data from France, early-2020



Modeling epidemiological dynamics using segregating sites



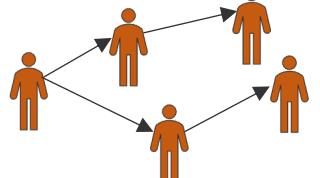
Thank you!

Questions?

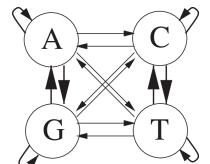
Ecological dynamics are inferred from observed data

Biology

Epidemiological processes

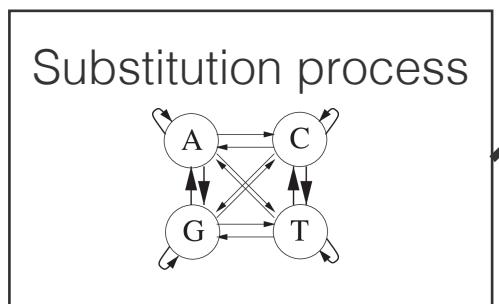
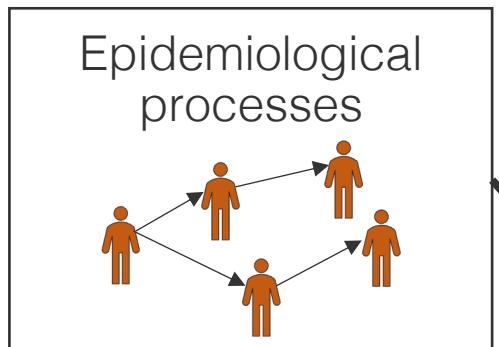


Substitution process



Ecological dynamics are inferred from observed data

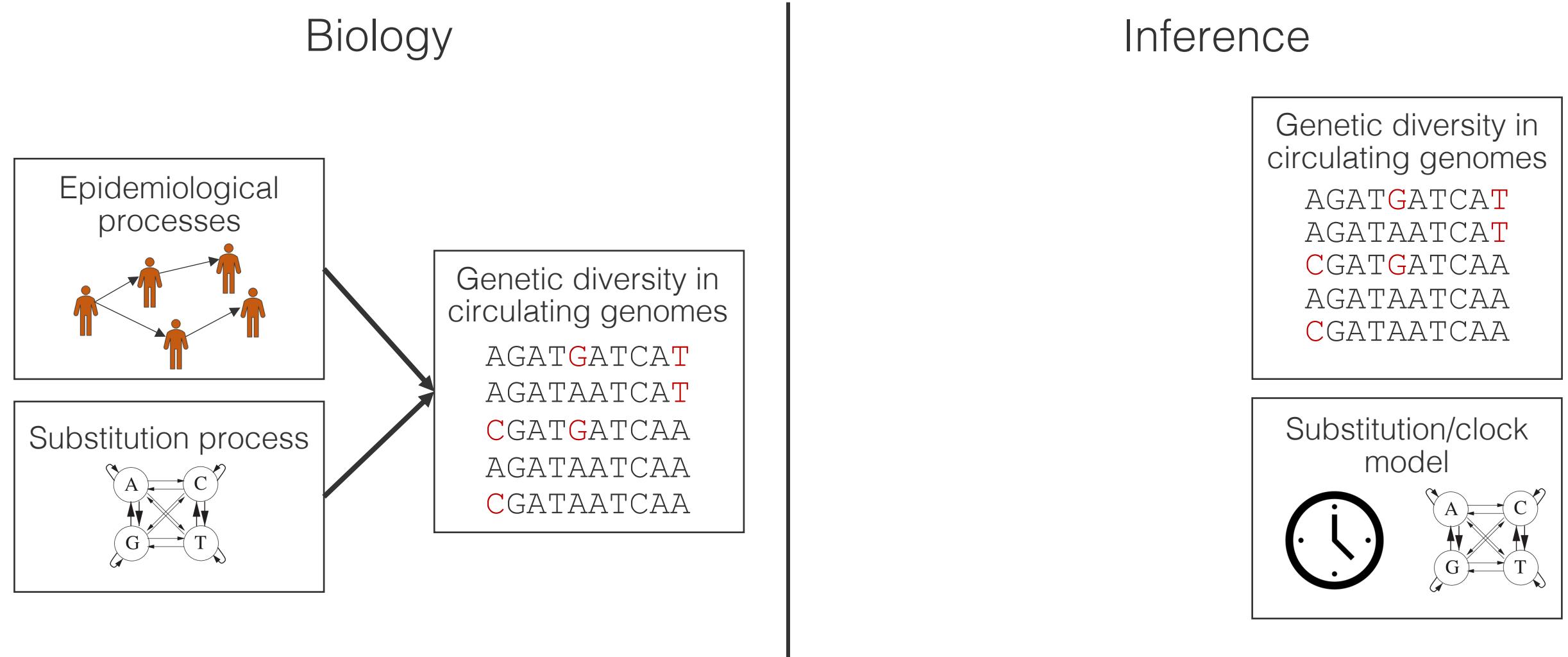
Biology



Genetic diversity in circulating genomes

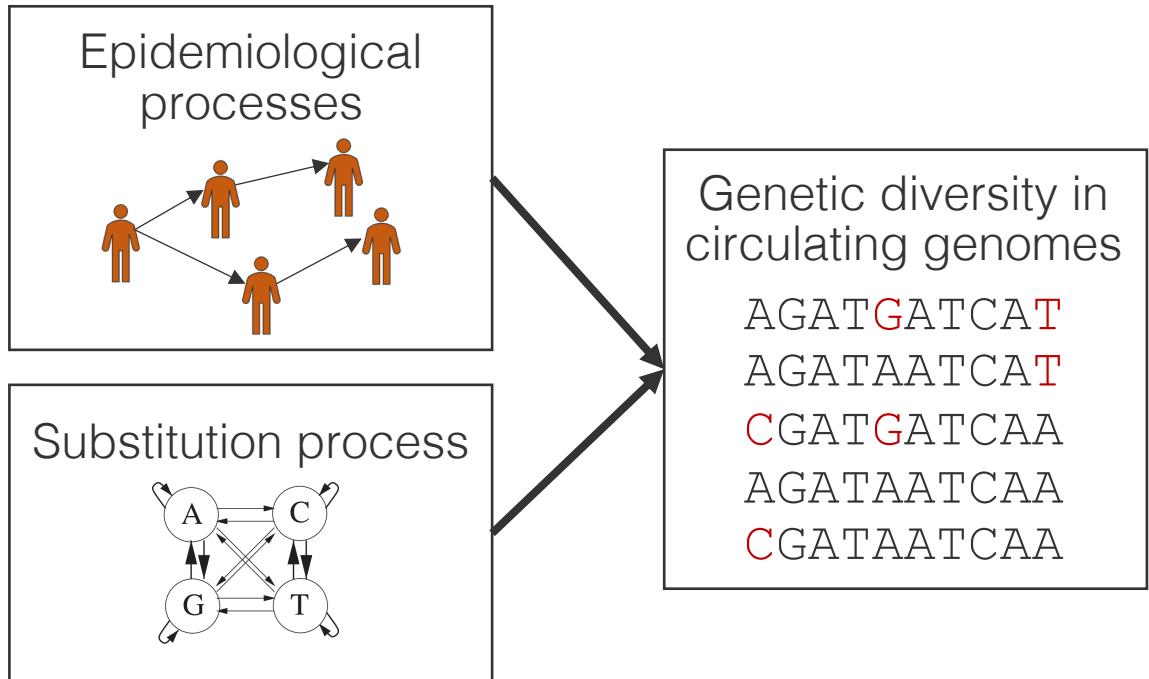
AGATGATCAT
AGATAATCAT
CGATGATCAA
AGATAATCAA
CGATAATCAA

Ecological dynamics are inferred from observed data



Ecological dynamics are inferred from observed data

Biology



Inference

