

Introduction to infectious disease phylodynamics

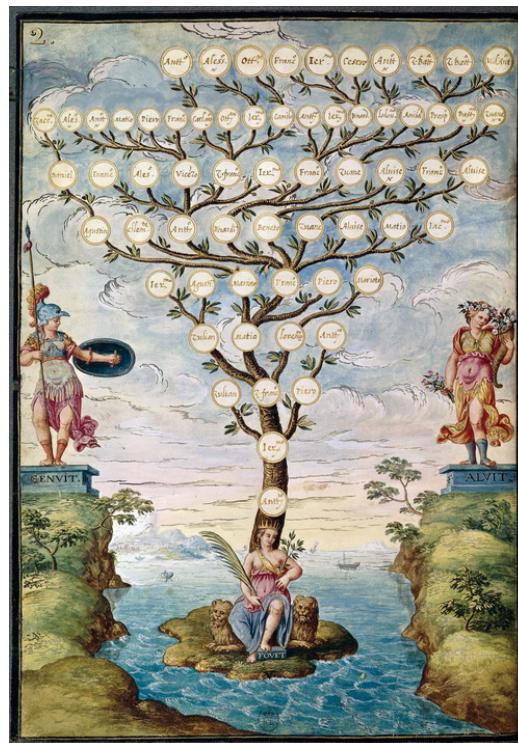
David Rasmussen

Dept. of Entomology and Plant Pathology
Bioinformatics Research Center
NC State University

Taming the BEAST Eh!
August 12th, 2019

Introduction

- **Phylogenetics** reconstructs the ancestral relationships among a set of individuals or taxa.



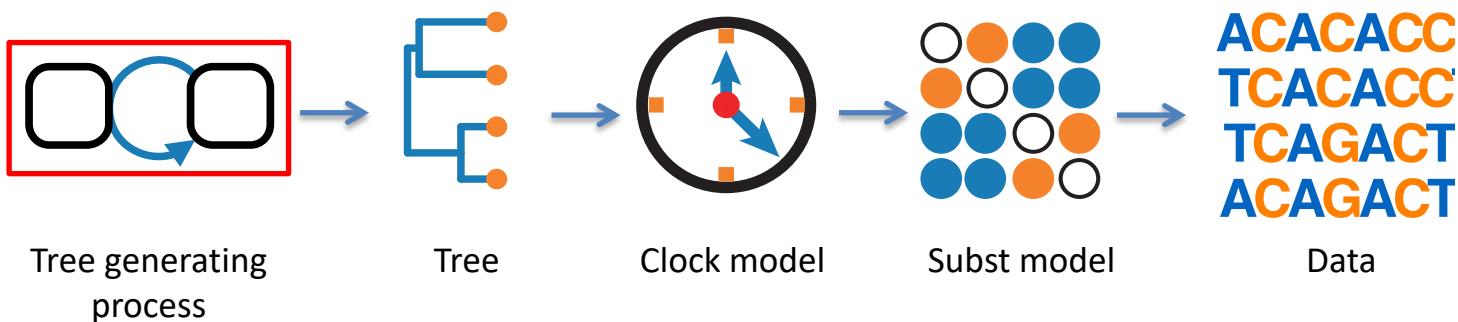
From 'The Book of Trees' Manuel Lima (2014)

Introduction

- **Phylogenetics** reconstructs the ancestral relationships among a set of individuals or taxa.
- **Phyldynamics** studies how ecological and evolutionary processes act or interact to shape the phylogenetic history of a population (Grenfell *et al.*, 2004; Volz *et al.*, 2014).

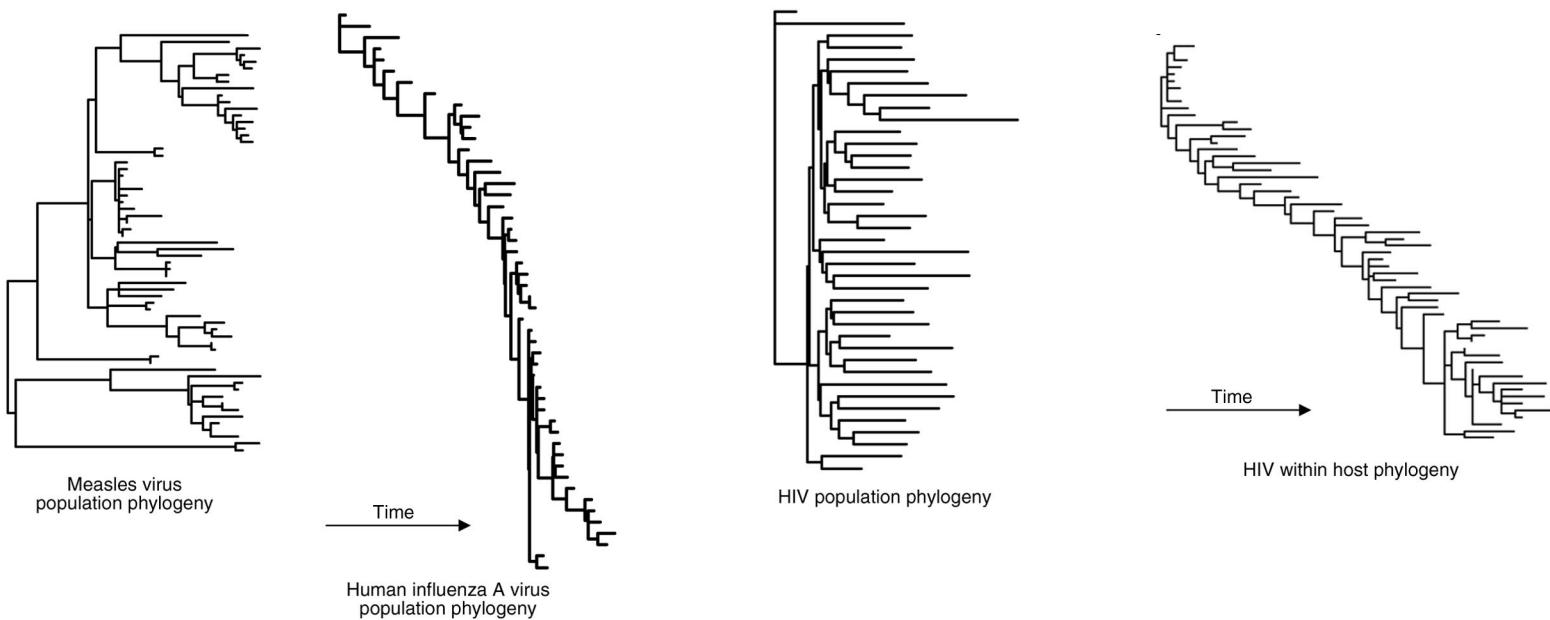
Introduction

- **Phylogenetics** reconstructs the ancestral relationships among a set of individuals or taxa.
- **Phyldynamics** studies how ecological and evolutionary processes act or interact to shape the phylogenetic history of a population (Grenfell *et al.*, 2004; Volz *et al.*, 2014).



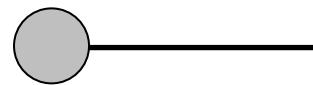
Introduction

- What processes shape the phylogenetic history of different pathogens?

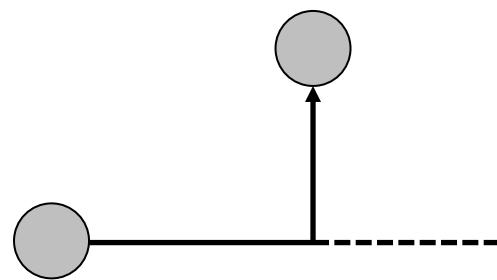


Trees from Grenfell *et al.* (Science, 2004)

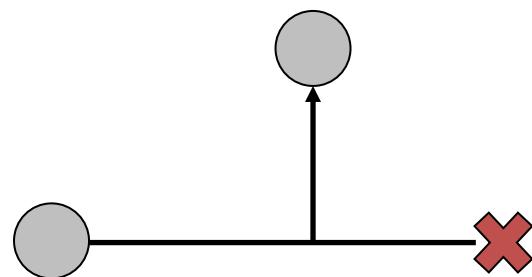
A simple epidemic example



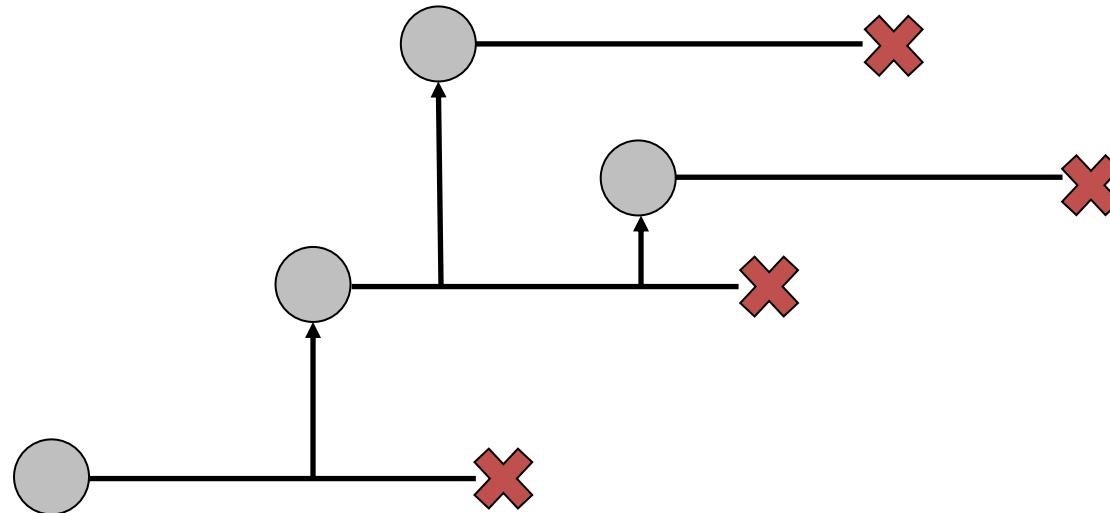
A simple epidemic example



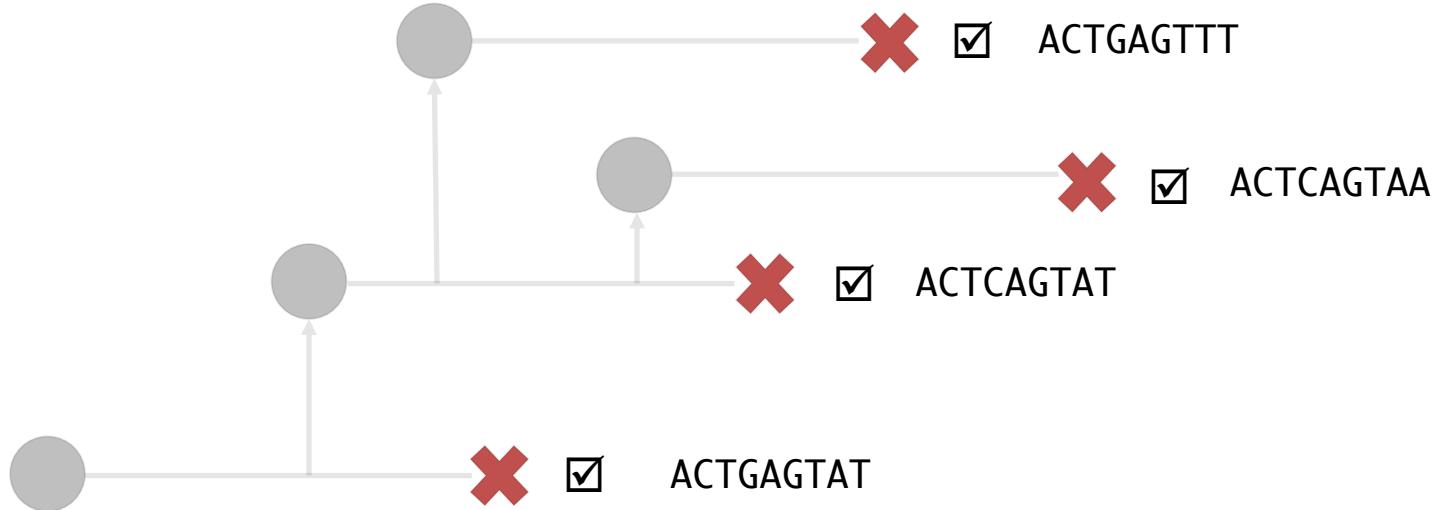
A simple epidemic example



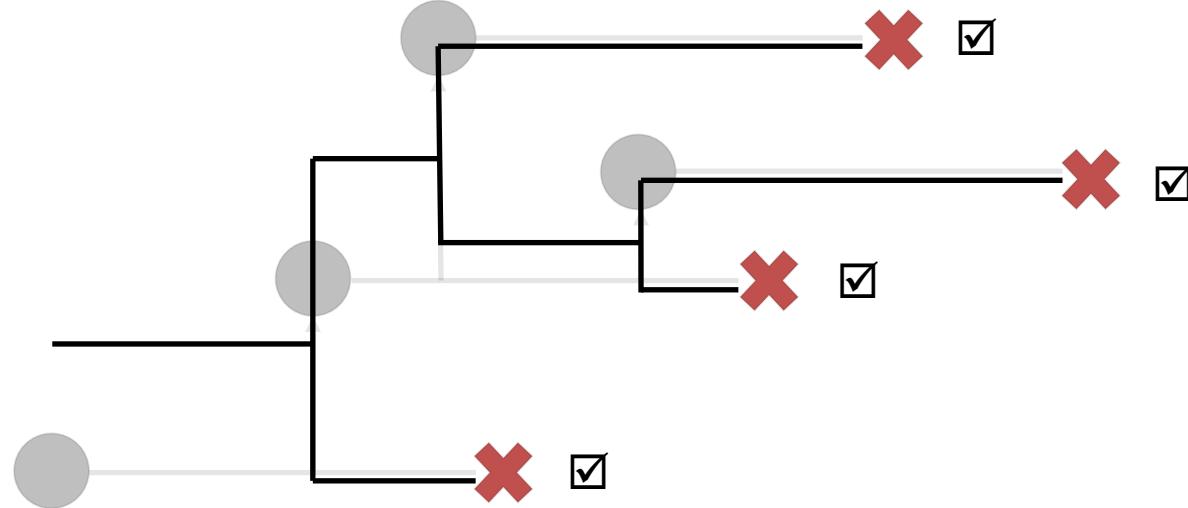
A simple epidemic example



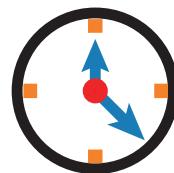
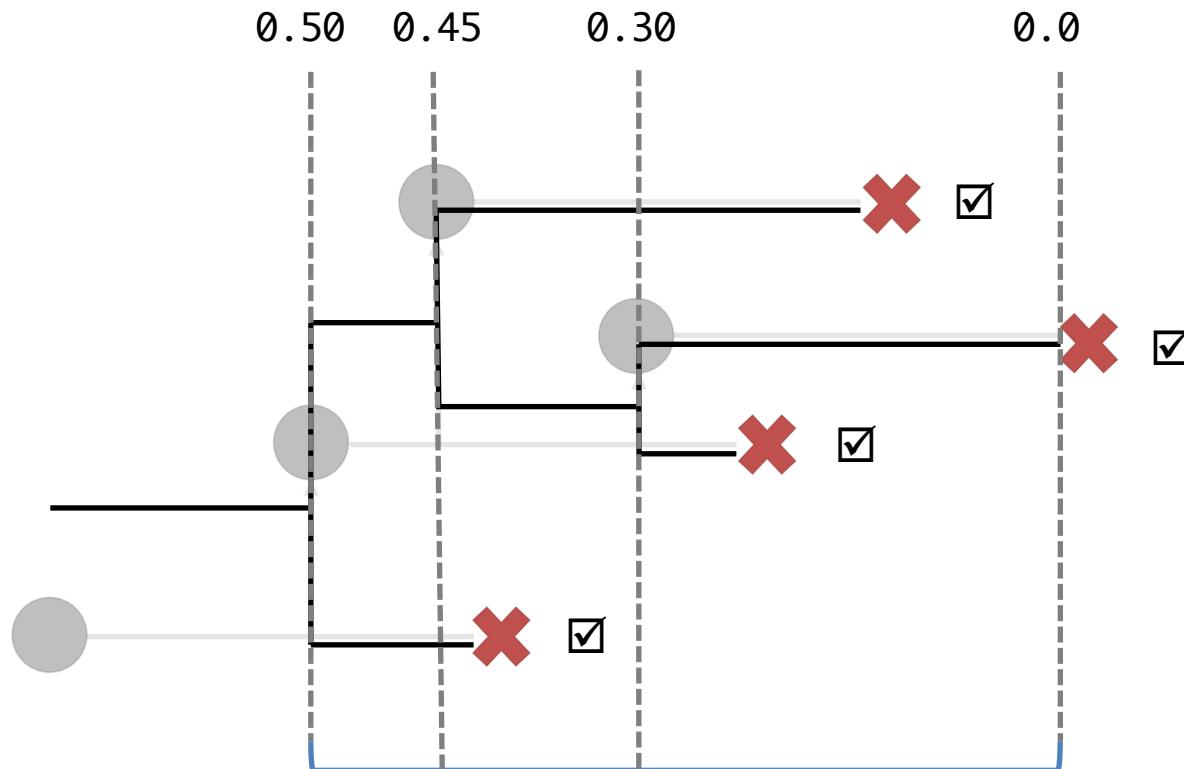
A simple epidemic example



A simple epidemic example



With a molecular clock



Real time = genetic distance
clock rate

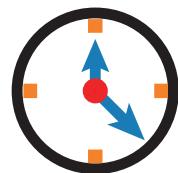
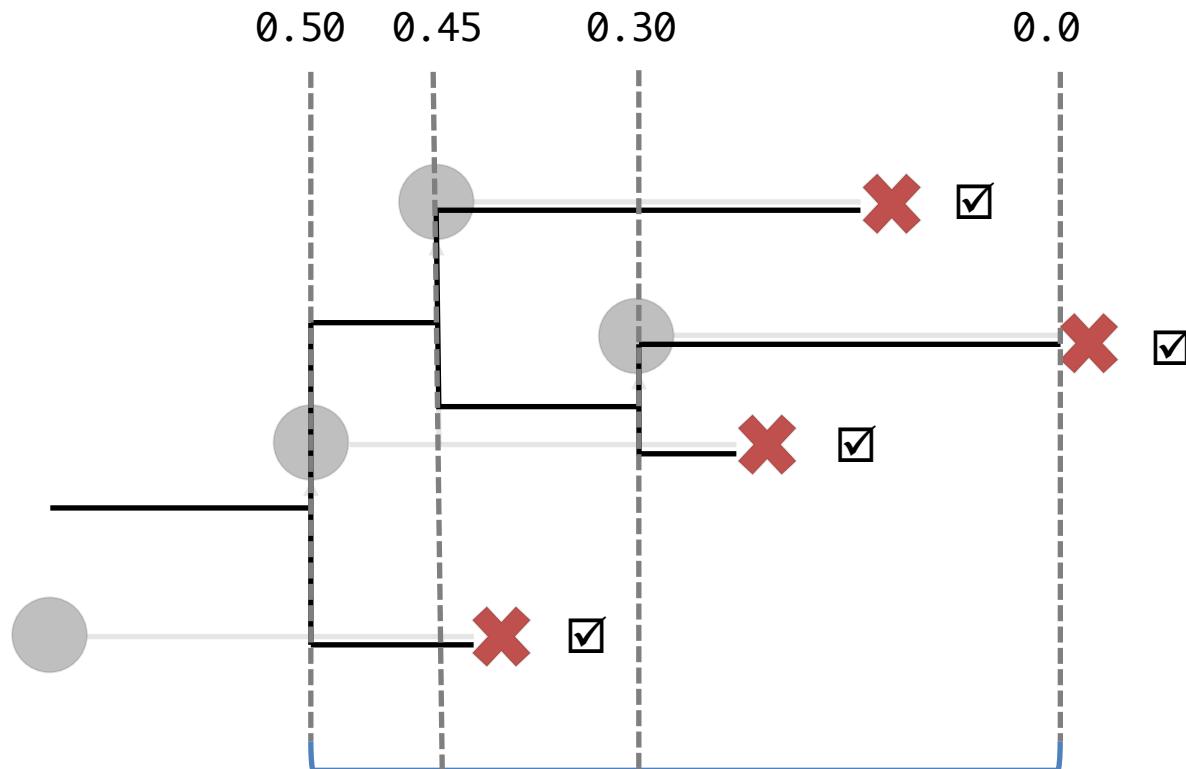
Phylogenies can tell us about:

- The origins of epidemics and new strains
- Linkage – the sources of transmission
- Past epidemic/population dynamics
- Pathogen fitness and adaptation

Phylogenies can tell us about:

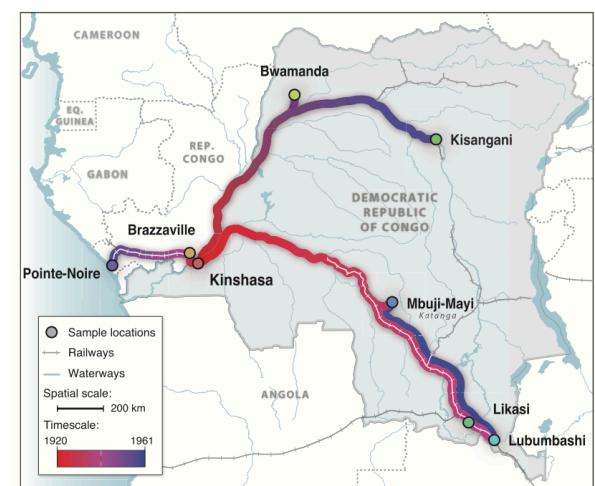
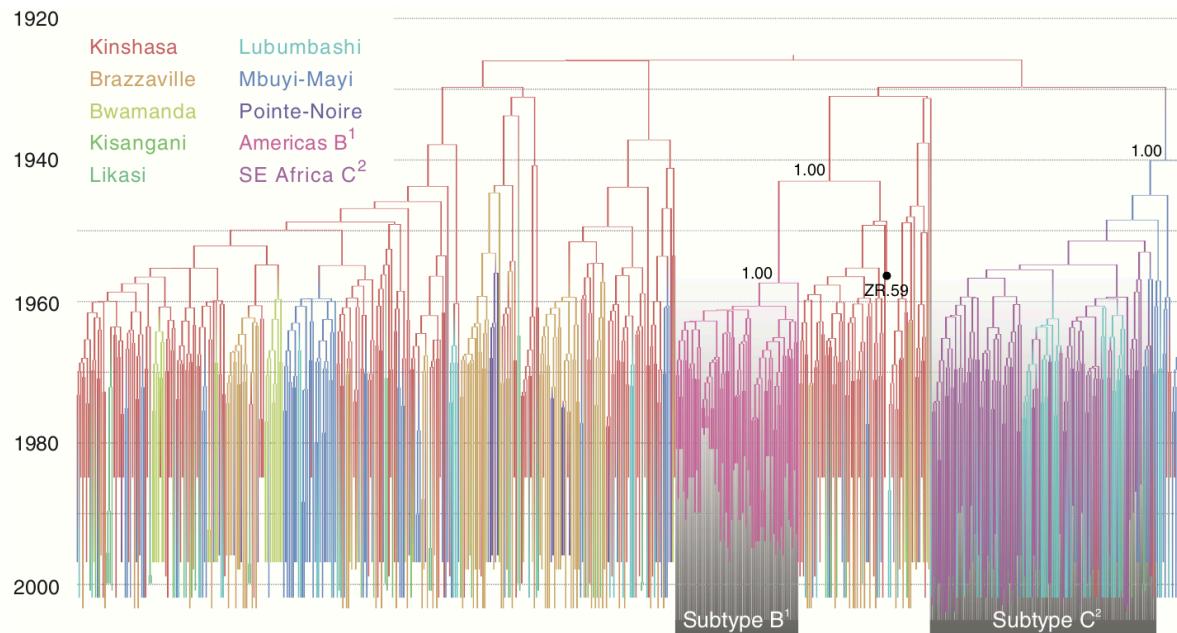
- The origins of epidemics and new strains
- Linkage – the sources of transmission
- Past epidemic/population dynamics
- Pathogen fitness and adaptation

With a molecular clock



Real time = genetic distance
clock rate

Origins of the HIV-1 epidemic

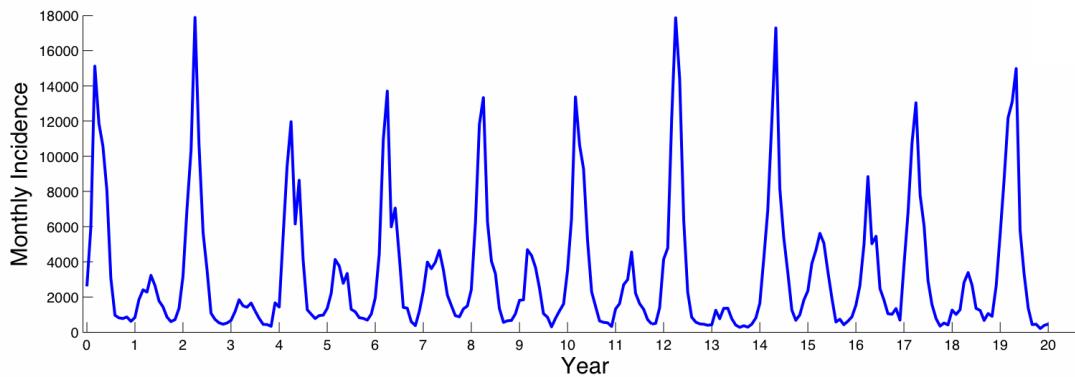


Phylogenies can tell us about:

- The origins of epidemics and new strains
- Linkage – the sources of transmission
- Past epidemic/population dynamics
- Pathogen fitness and adaptation

Revealing the source of infections

- Classic epidemiological data are typically not informative about the source of new infections.

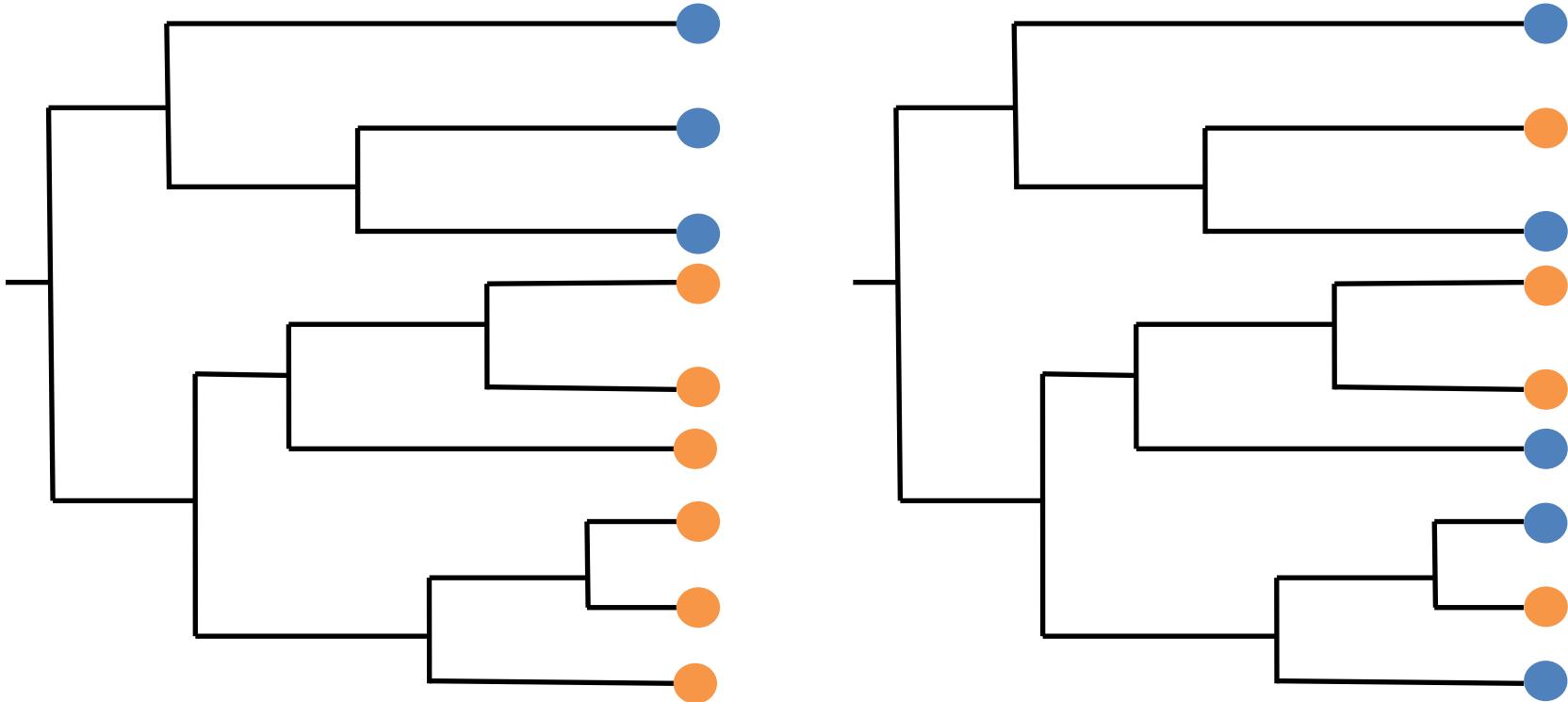


- Phylogenetics can help reveal the movement of pathogens and the source of new infections.

Revealing the source of infections

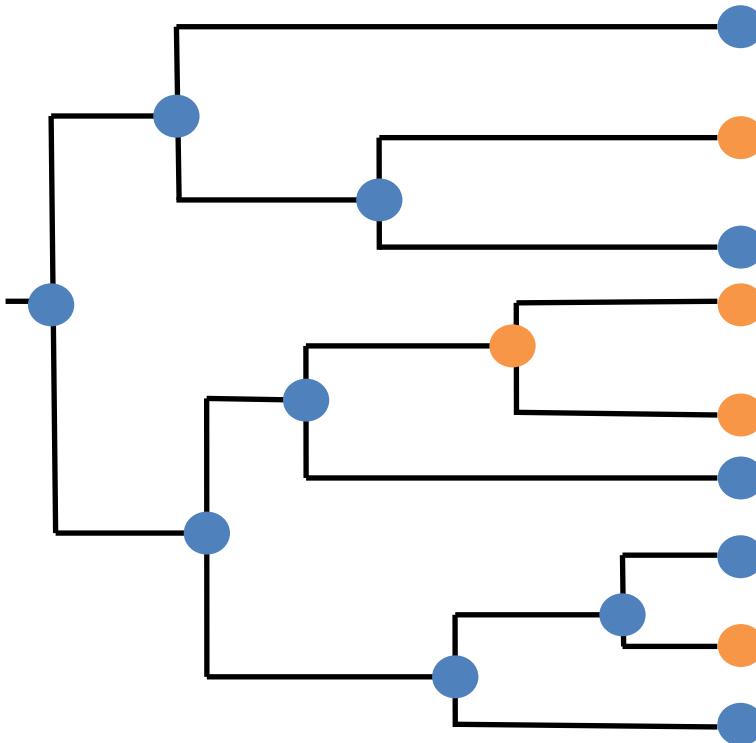
External

Local



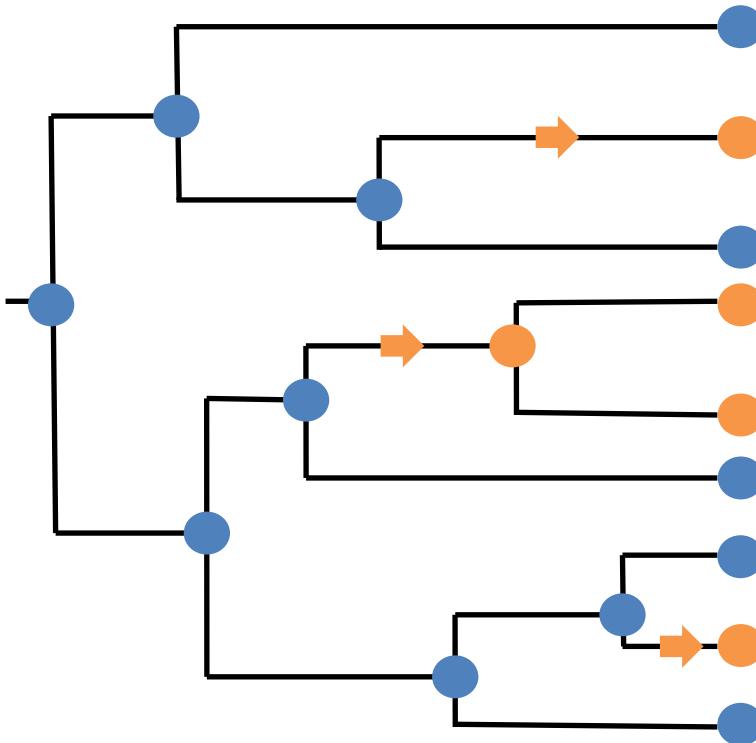
Ancestral state reconstruction

● External ● Local



Ancestral state reconstruction

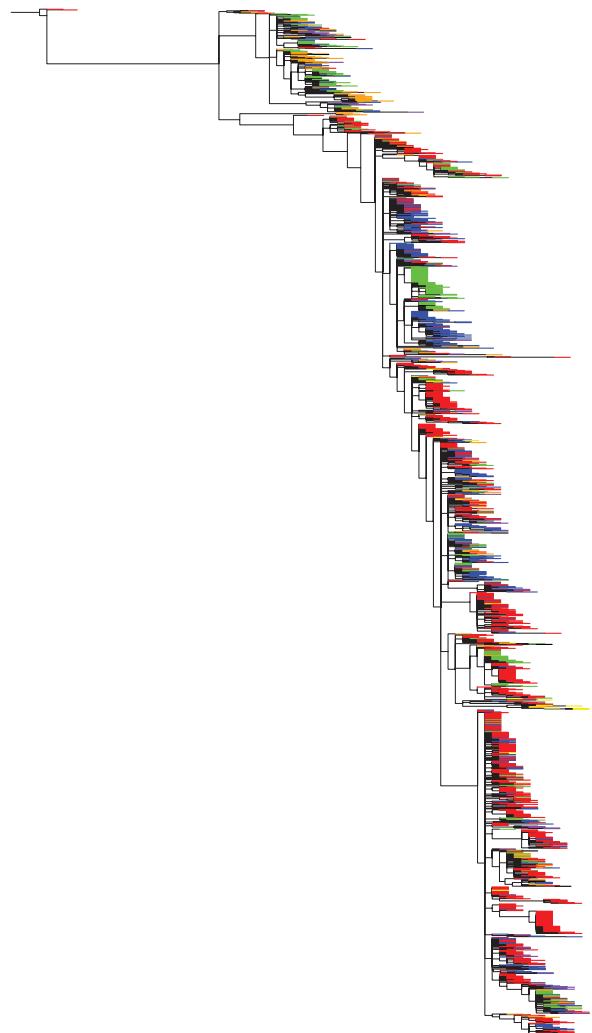
● External ● Local



Phylogeography

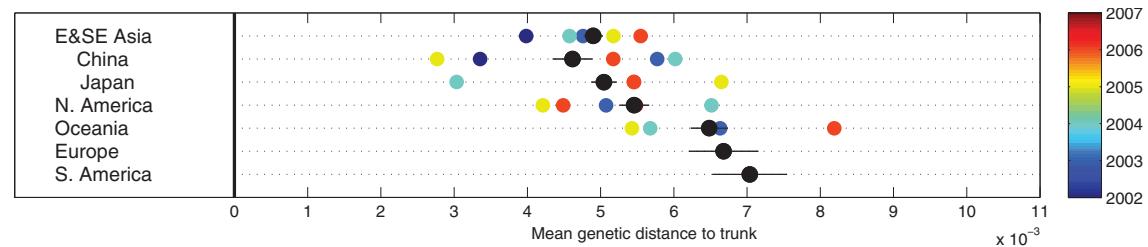
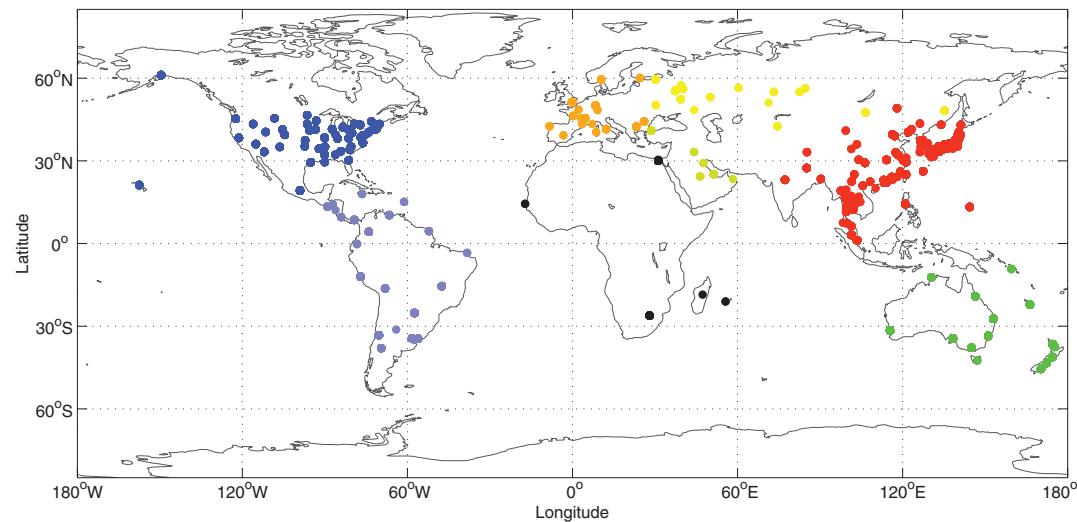
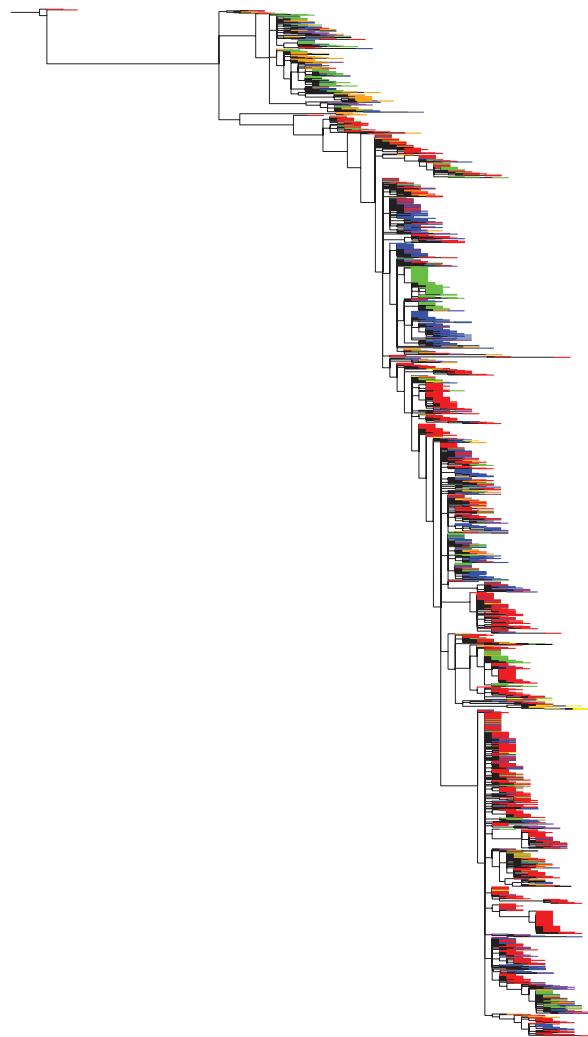
- Phylogeography combines phylogenetic methods with geographic information (e.g. sampling locations) to reconstruct the spatial spread of pathogens.
- Often used to estimate the geographic origins of emerging epidemics:
 - Pandemic influenza H1N1 (Smith et al., 2009)
 - Ebola in West Africa (Gire et al., 2014)
 - HIV-1 in Africa (Faria et al., 2014)
- But can also be used to study much more complex spatiotemporal dynamics of endemic pathogens.

Case study: Influenza A H3N2

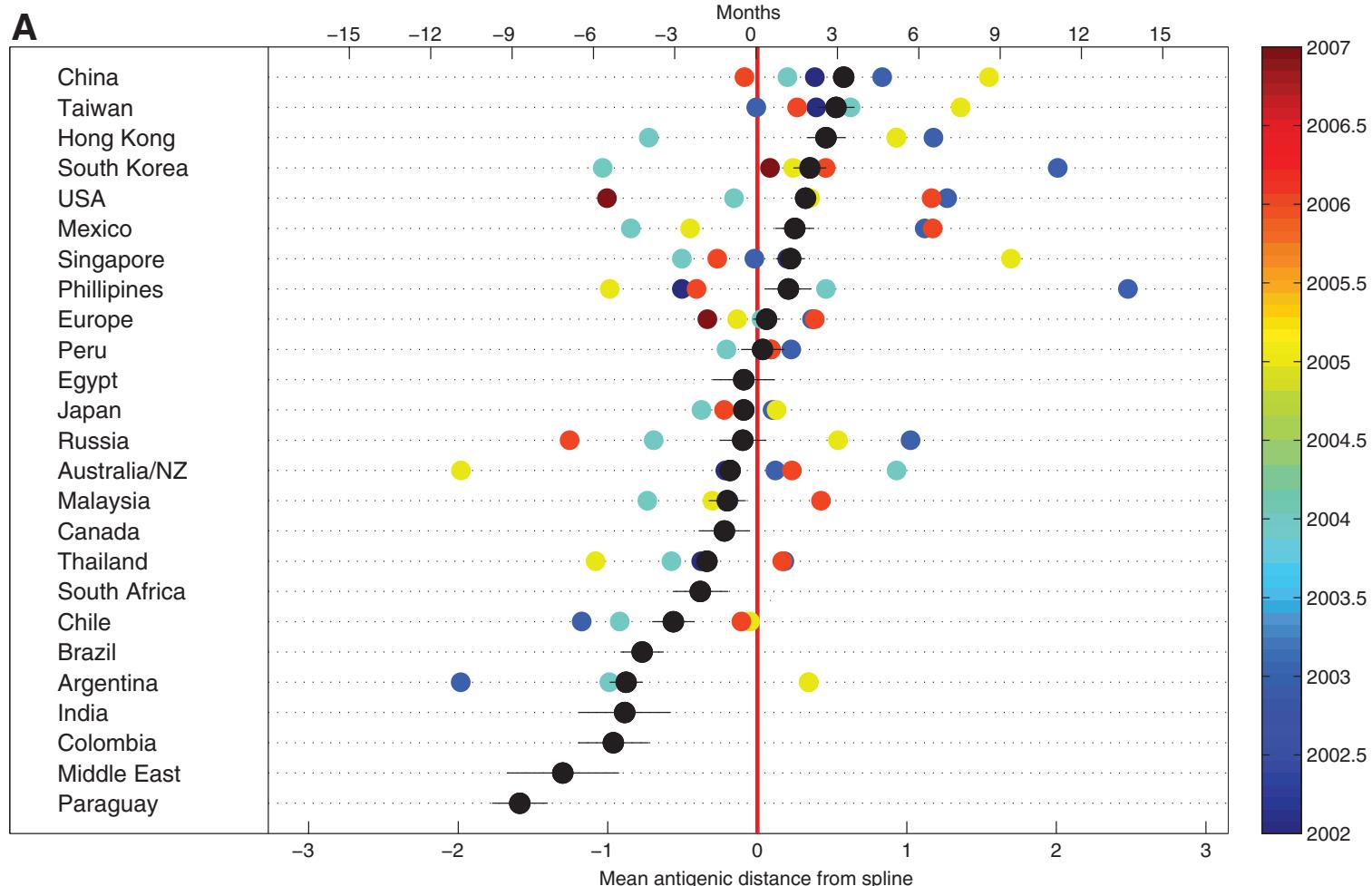


- New antigenic variants periodically replace older strains
- This creates antigenic-mismatch with vaccines.
- Antigenic drift leads to a ladder-like structure with a prominent “trunk” lineage.
- Where do new antigenic variants arise?

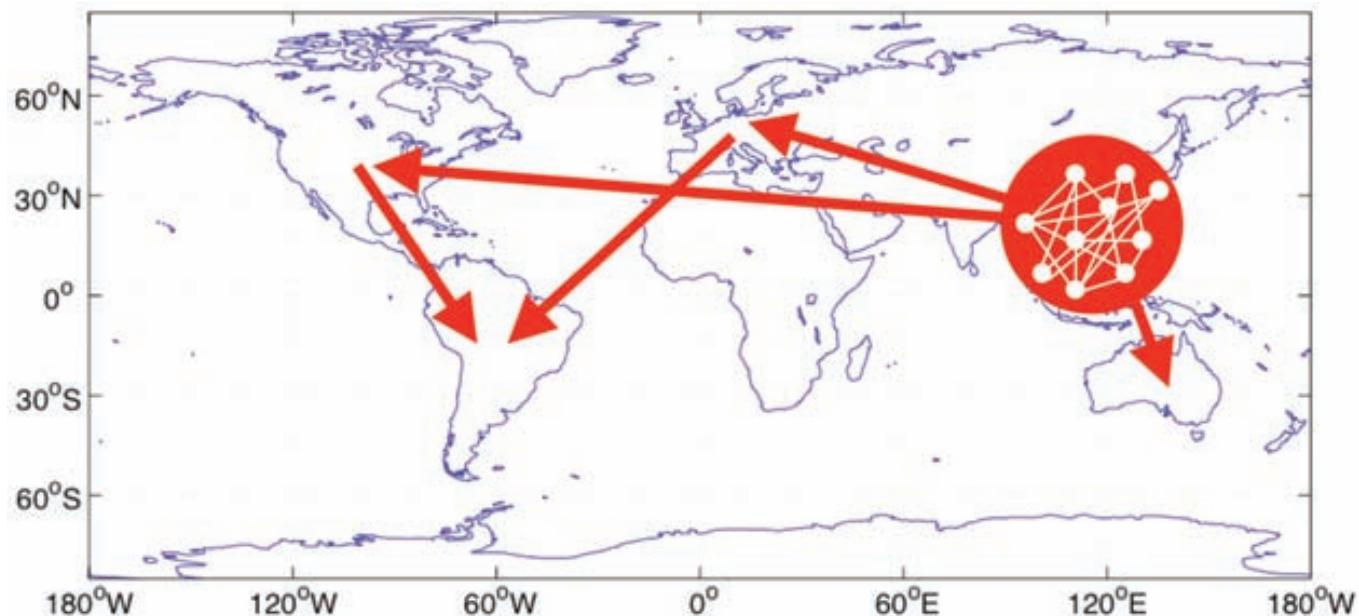
A potential SE Asian source



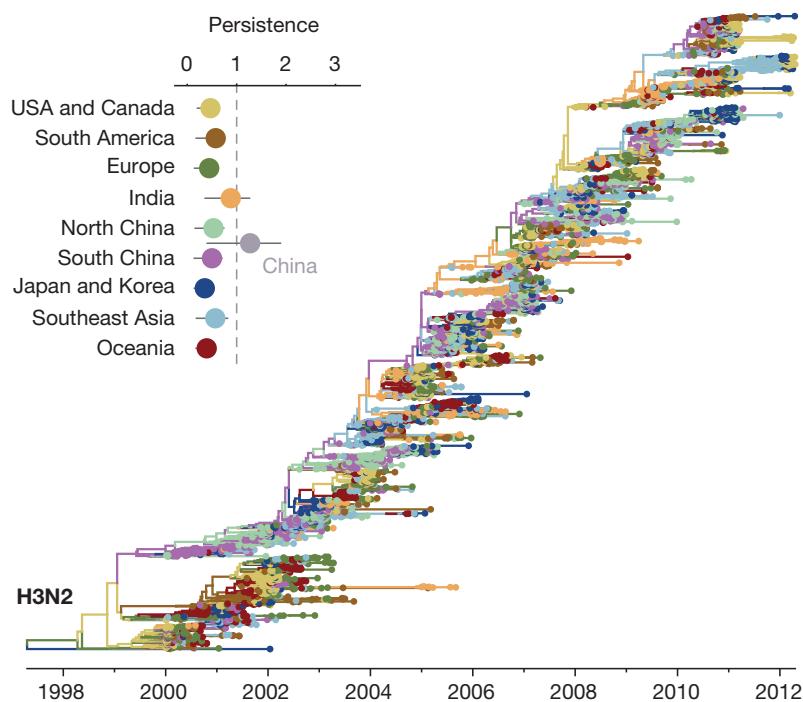
Antigenic leading in SE/East Asia



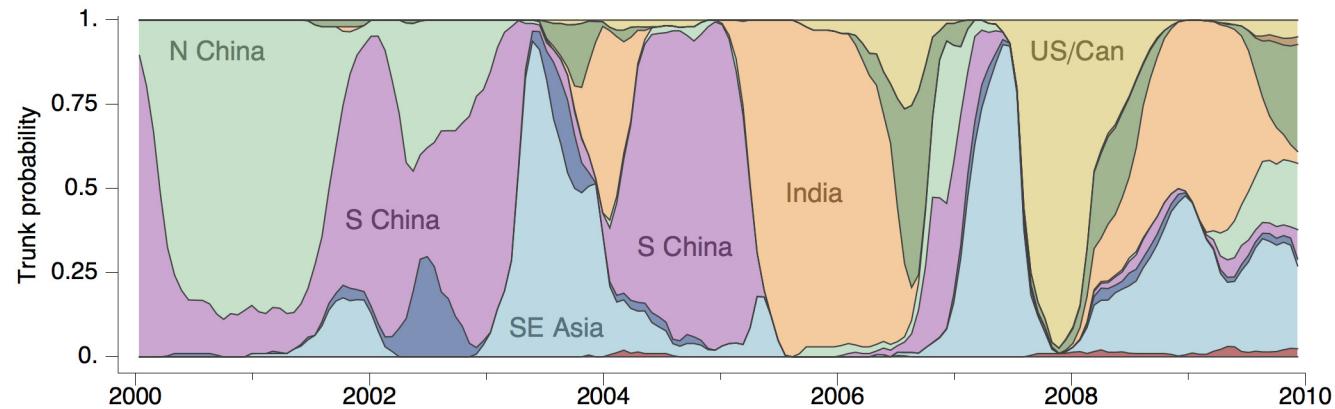
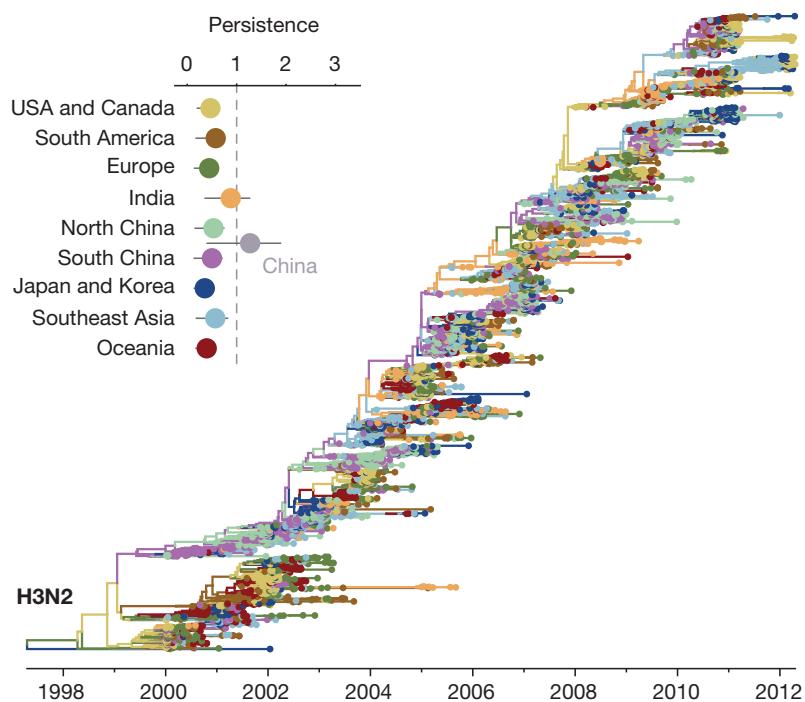
A potential SE Asian source



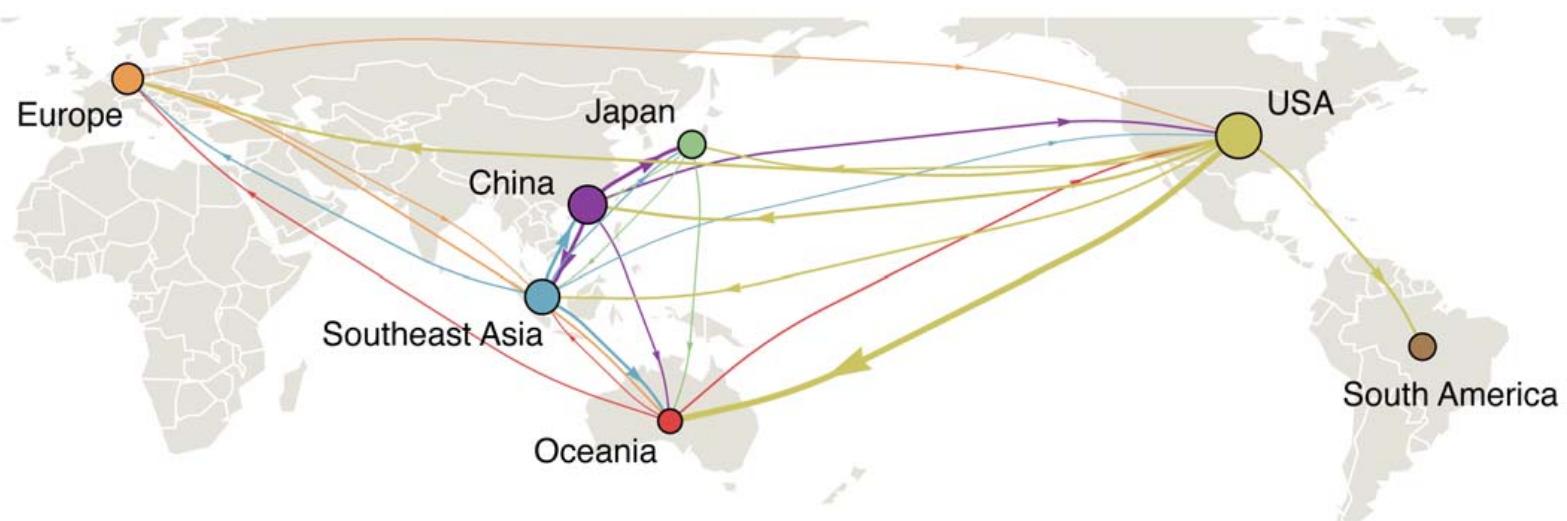
Dependence on sampling



Dependence on sampling



Global influenza circulation network

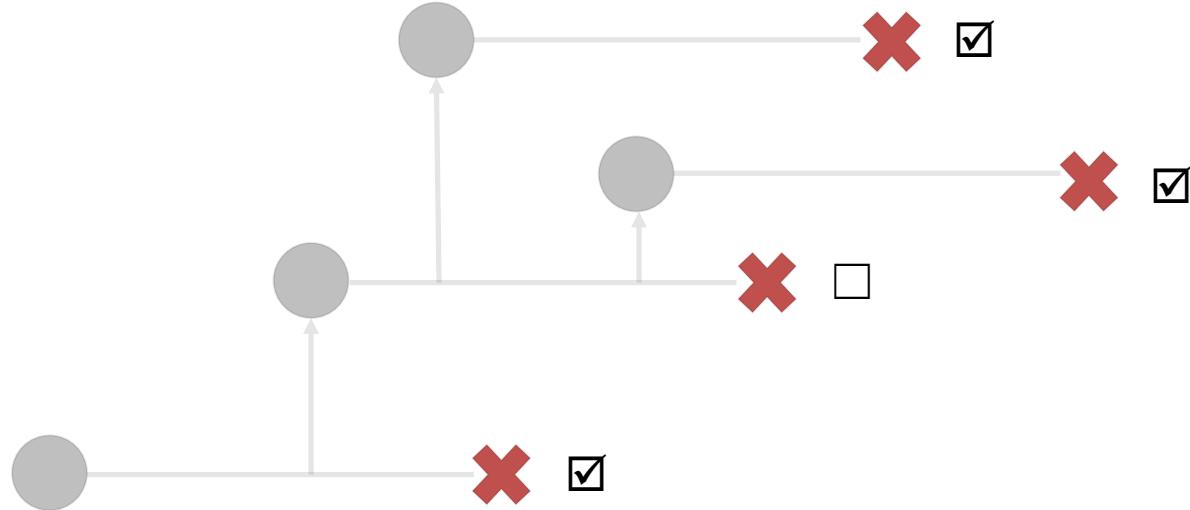


Influenza H3N2 summary

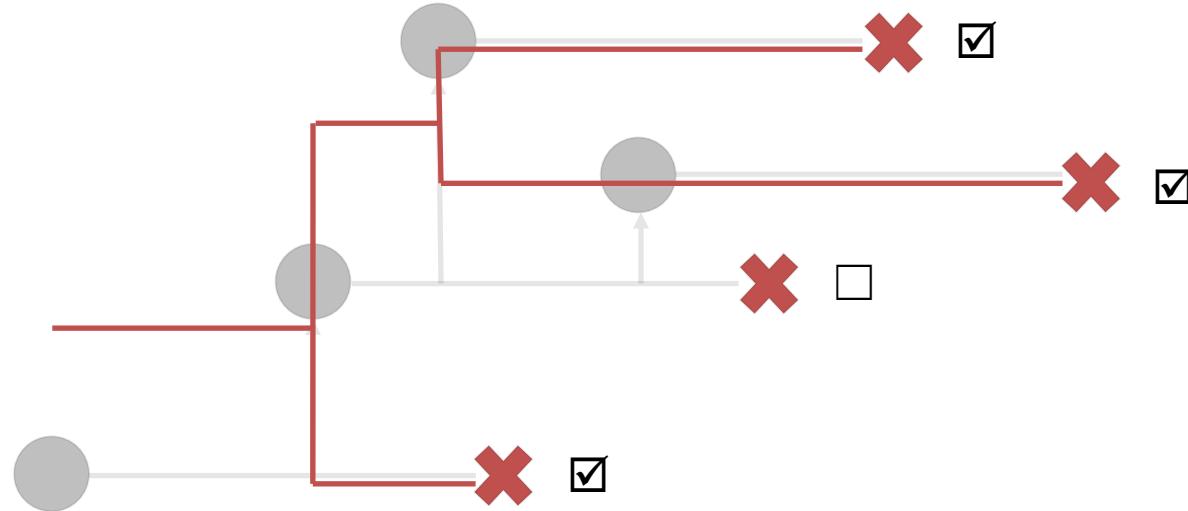
- Phylogeography has helped reveal the global circulation patterns of H3N2
- East and SE Asia likely play a central role in the circulation network but no single location acts as a constant global source but rather there is a “globally circulating metapopulation” (Bahl et al, PNAS 2011).
- Phylogenetic analyses should never be viewed as definitive – new evidence can always overturn previous hypotheses.

Phylodynamic modeling

A simple epidemic example – with incomplete sampling



A simple epidemic example – with incomplete sampling

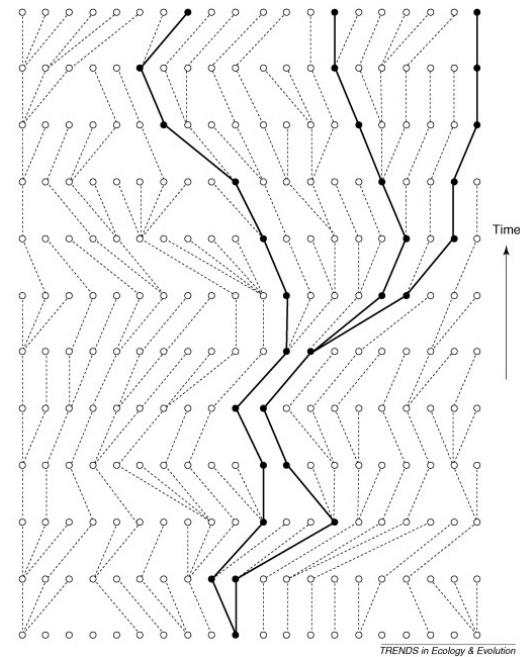


Phylodynamic modeling in a nutshell

- Phylogenies will only contain sampled lineages.

Phylodynamic modeling in a nutshell

- Phylogenies will only contain sampled lineages.
- The sampled phylogeny is embedded within the full ancestral history of the population.



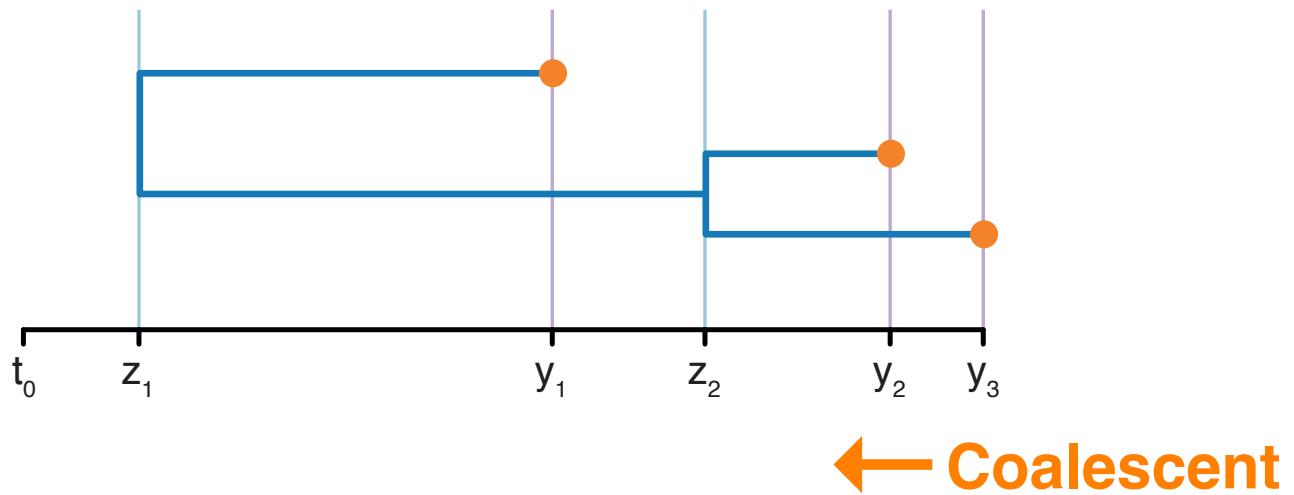
Kuhner (2008)

Phylodynamic modeling in a nutshell

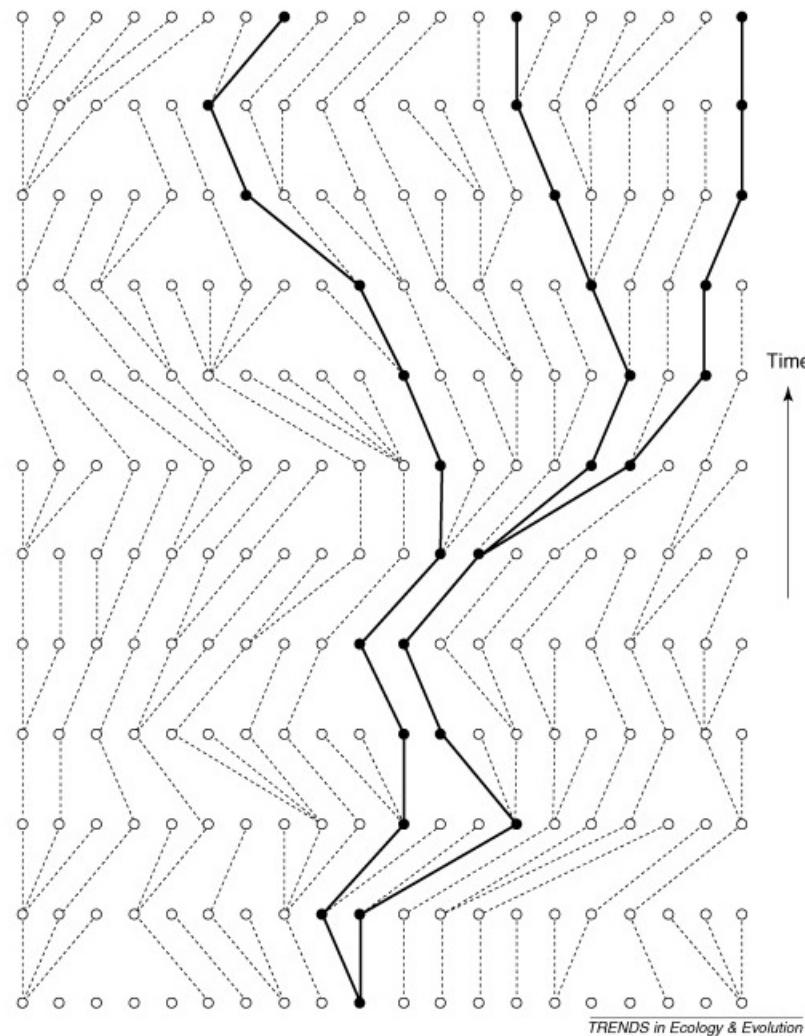
- Phylogenies will only contain sampled lineages.
- The sampled phylogeny is embedded within the full ancestral history of the population.
- We need a statistical model that allows us to infer the most likely population history from the sampled phylogeny.

Two flavors of phylodynamic models

Birth-death →

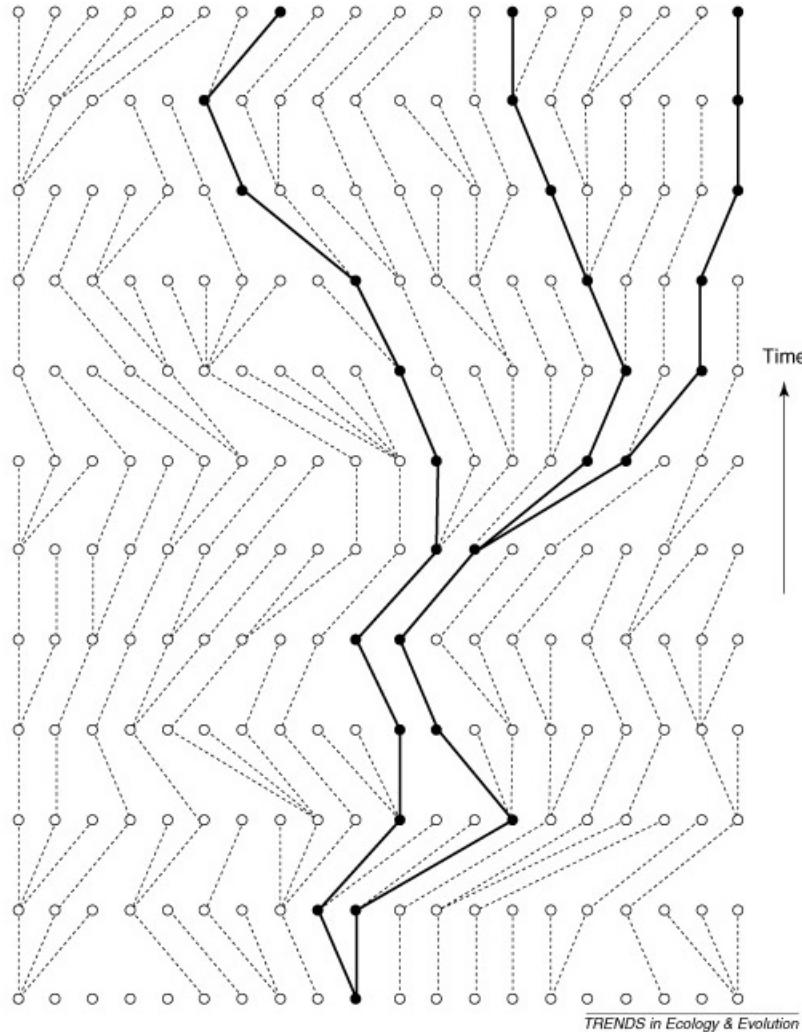


Coalescent theory



Kuhner (2008)

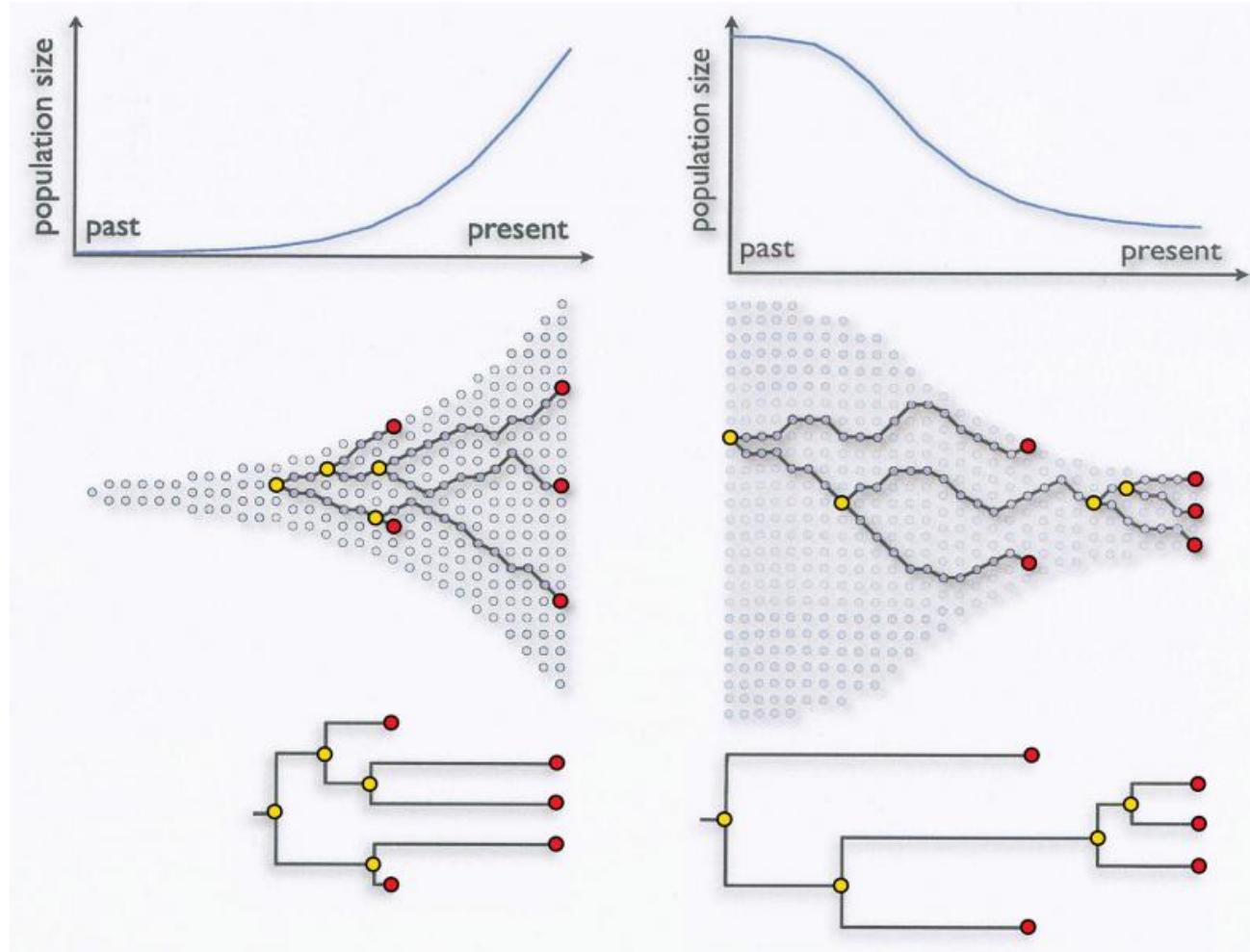
Coalescent theory



Probability of coalescing per generation:

$$p_{coal} = \frac{1}{N}$$

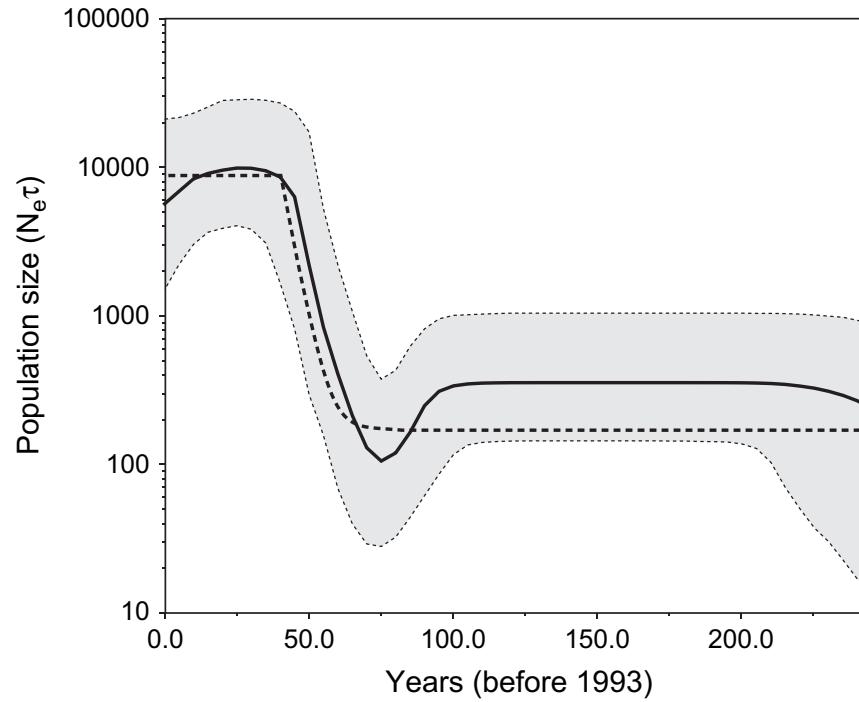
Reconstructing population dynamics



Phylogenies can tell us about:

- The origins of epidemics and new strains
- Linkage – the sources of transmission
- Past epidemic/population dynamics
- Pathogen fitness and adaptation

Reconstructing past dynamics: Hepatitis C



Pybus *et al.* (2003)

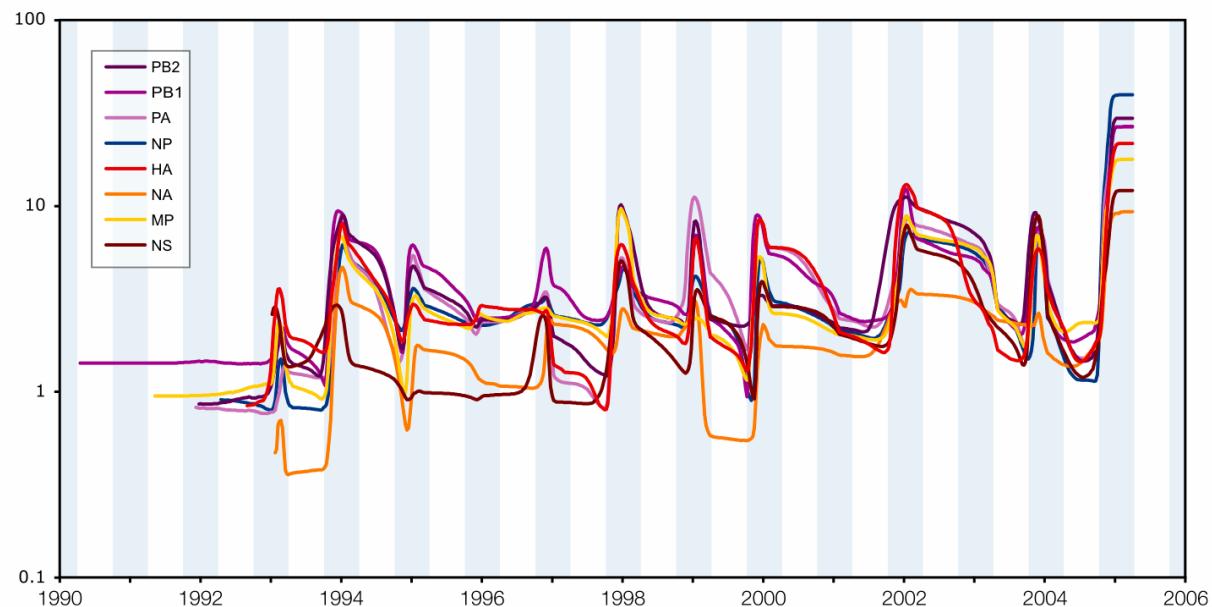
Drummond *et al.* (2005)

Reconstructing past dynamics

The genomic and epidemiological dynamics of human influenza A virus

Andrew Rambaut¹, Oliver G. Pybus², Martha I. Nelson³, Cecile Viboud⁴, Jeffery K. Taubenberger⁵
& Edward C. Holmes^{3,4}

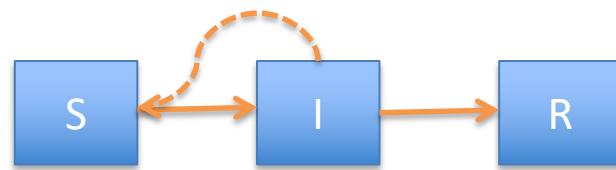
The evolutionary interaction between influenza A virus and the human immune system, manifest as 'antigenic drift' of the viral haemagglutinin, is one of the best described patterns in molecular evolution. However, little is known about the genome-scale evolutionary dynamics of this pathogen. Similarly, how genomic processes relate to global influenza epidemiology, in which the A/H3N2 and A/H1N1 subtypes co-circulate, is poorly understood. Here through an analysis of 1,302 complete viral genomes sampled from temperate populations in both hemispheres, we show that the genomic evolution of influenza A virus is characterized by a complex interplay between frequent reassortment and periodic selective sweeps. The A/H3N2 and A/H1N1 subtypes exhibit different evolutionary dynamics, with diverse lineages circulating in A/H1N1, indicative of weaker antigenic drift. These results suggest a sink-source model of viral ecology in which new lineages are seeded from a persistent influenza reservoir, which we hypothesize to be located in the tropics, to sink populations in temperate regions.



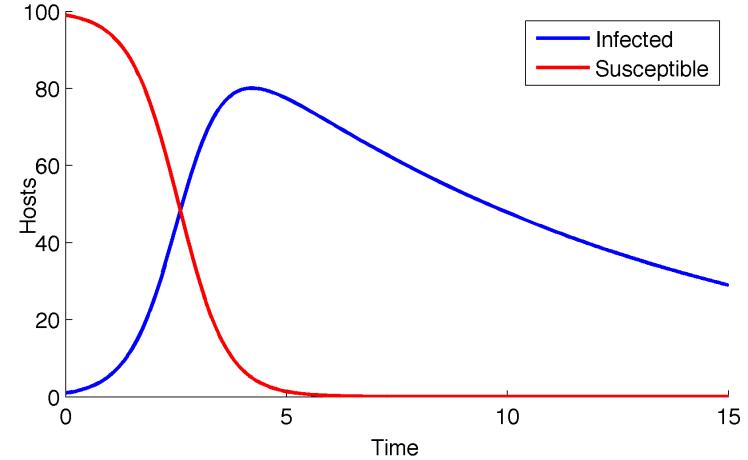
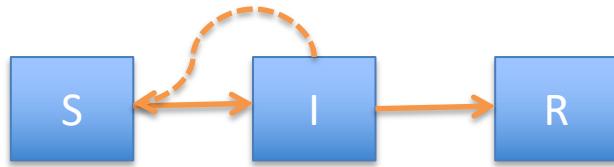
(Non-)Parametric phylodynamic models

- Non-parametric methods reconstruct effective population size (N_e) without assuming an explicit population dynamic model
 - Bayesian Skyline (Drummond et al., 2005)
 - Bayesian Skygrid (Minin et al., 2008)
- Alternatively, we can often derive a new phylodynamic model from a mechanistic population dynamic model with relevant epidemiological parameters.

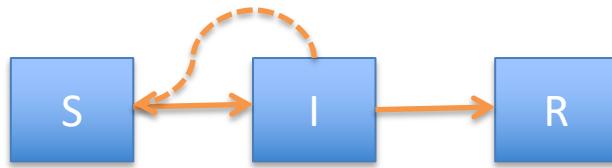
Epidemiological models



Epidemiological models



Epidemiological models



$$\frac{dS}{dt} = -\beta \frac{S}{N} I$$

$$\frac{dI}{dt} = \boxed{\beta \frac{S}{N} I} - \nu I$$

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

SIR-type coalescent model

- The total rate at which transmission/coalescent events occur in the population depends on the incidence:

$$f(t) = \beta S(t)I(t)$$

- Given a transmission event, the probability that two particular lineages coalesce is:

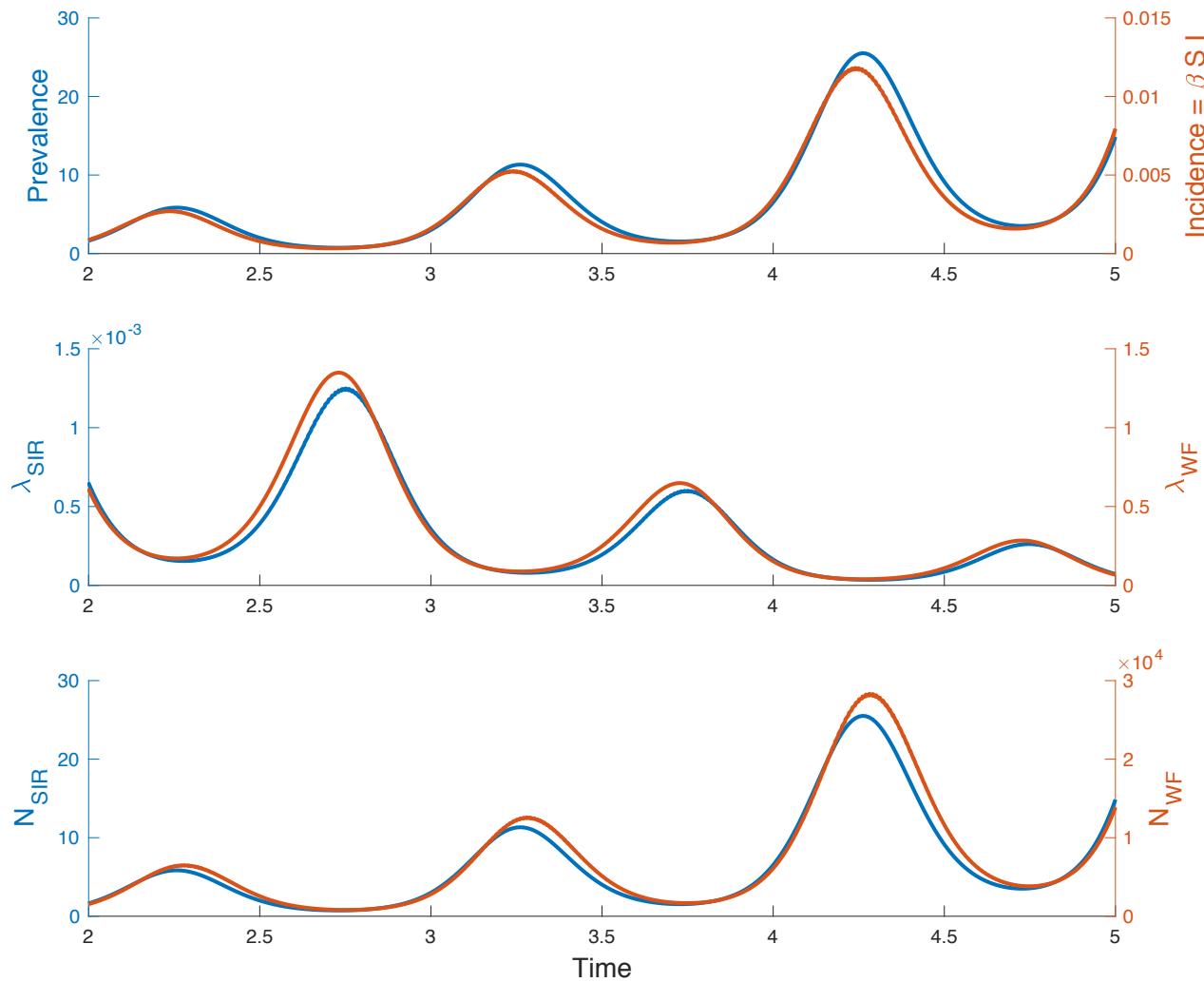
$$p_{coal} = \frac{2}{Y(t)^2} = \frac{2}{I(t)^2}$$

- The pairwise coalescent rate is then:

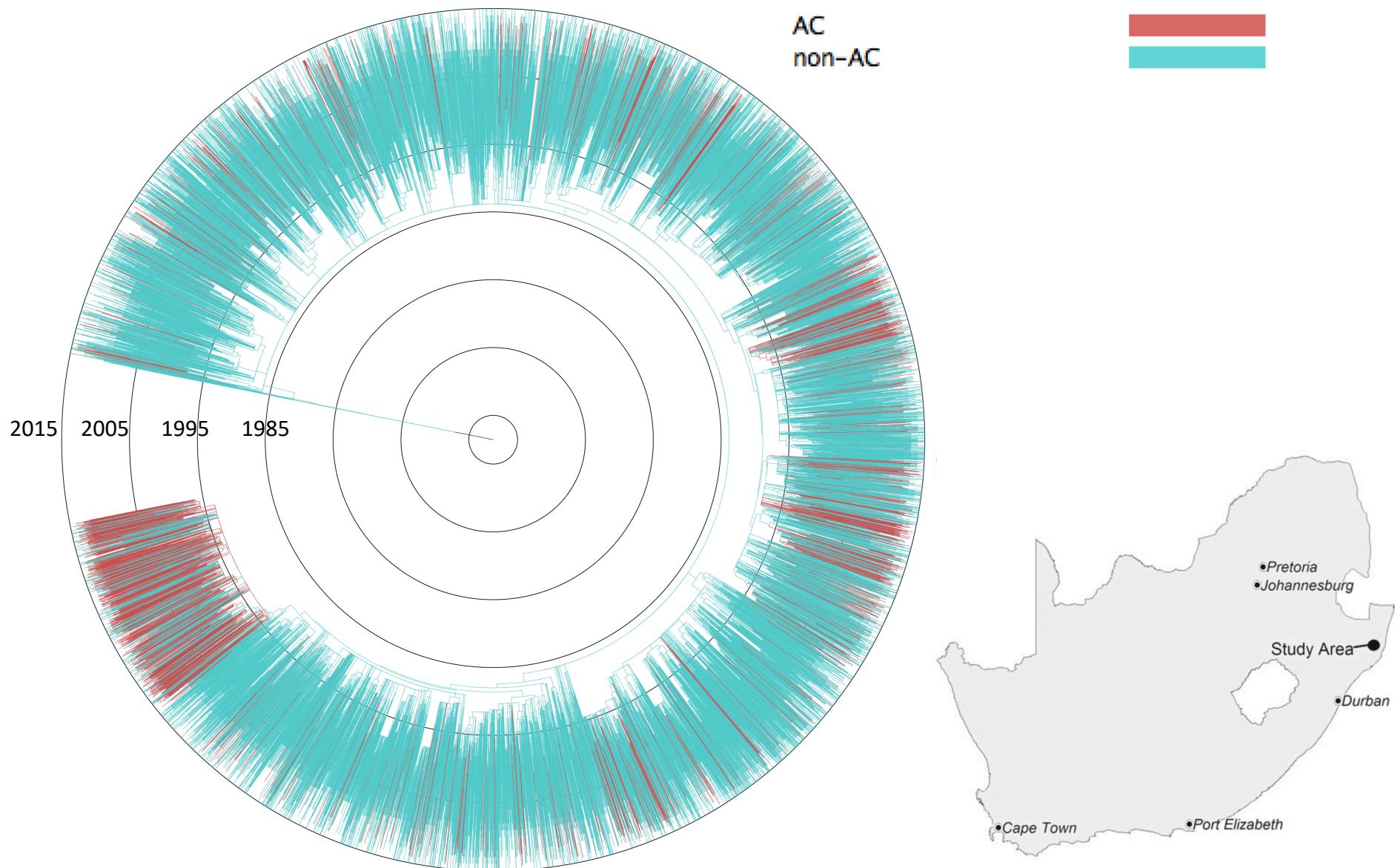
$$\lambda(t) = \frac{2f(t)}{Y(t)^2} = \frac{2\beta S(t)I(t)}{I(t)^2} = \frac{2\beta S(t)}{I(t)}$$

Volz *et al.* (2009);
Volz (2012)

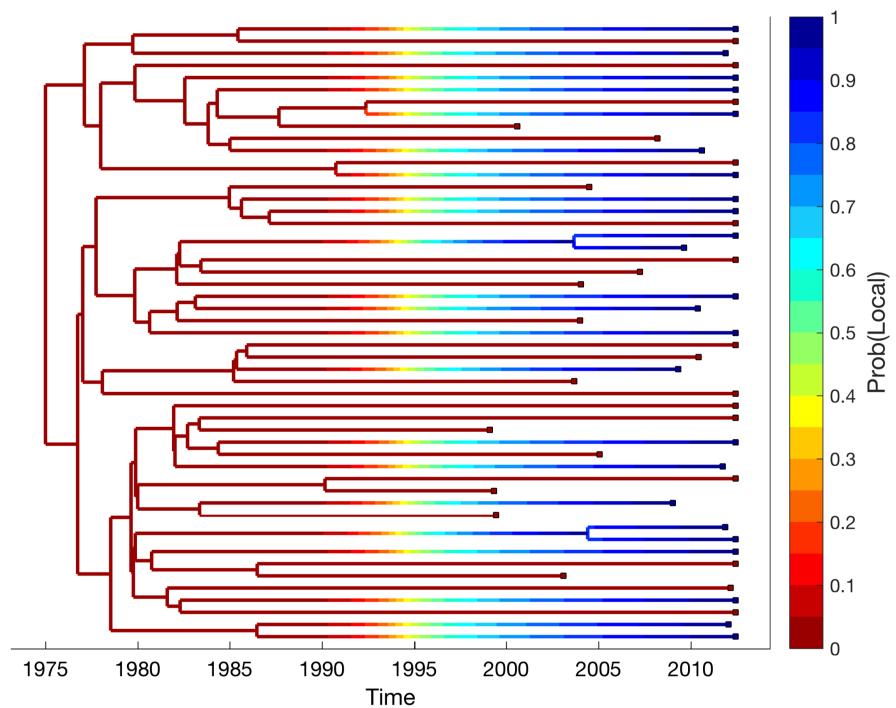
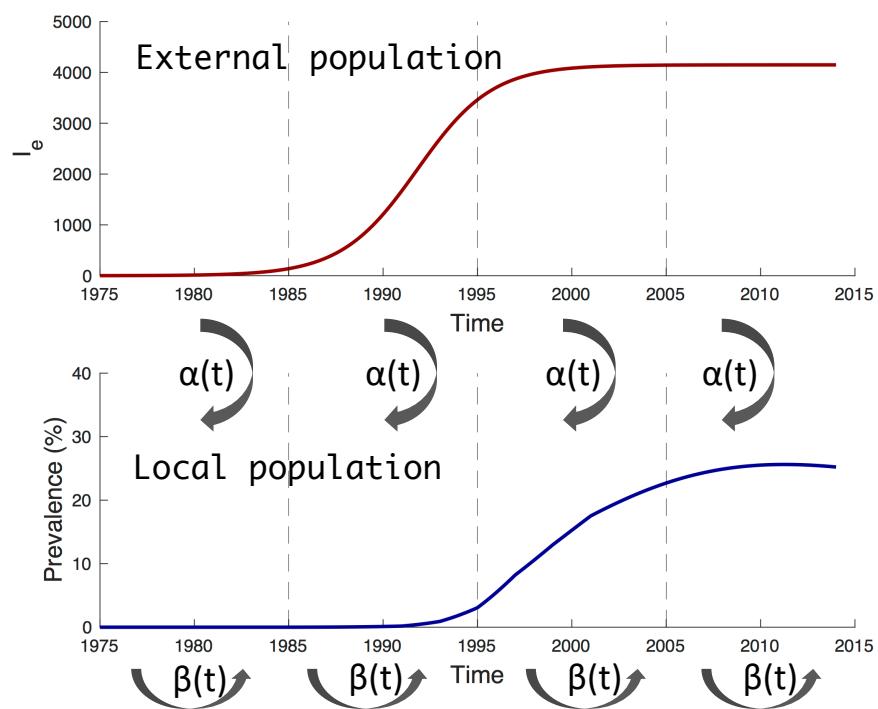
Coalescent model comparisons



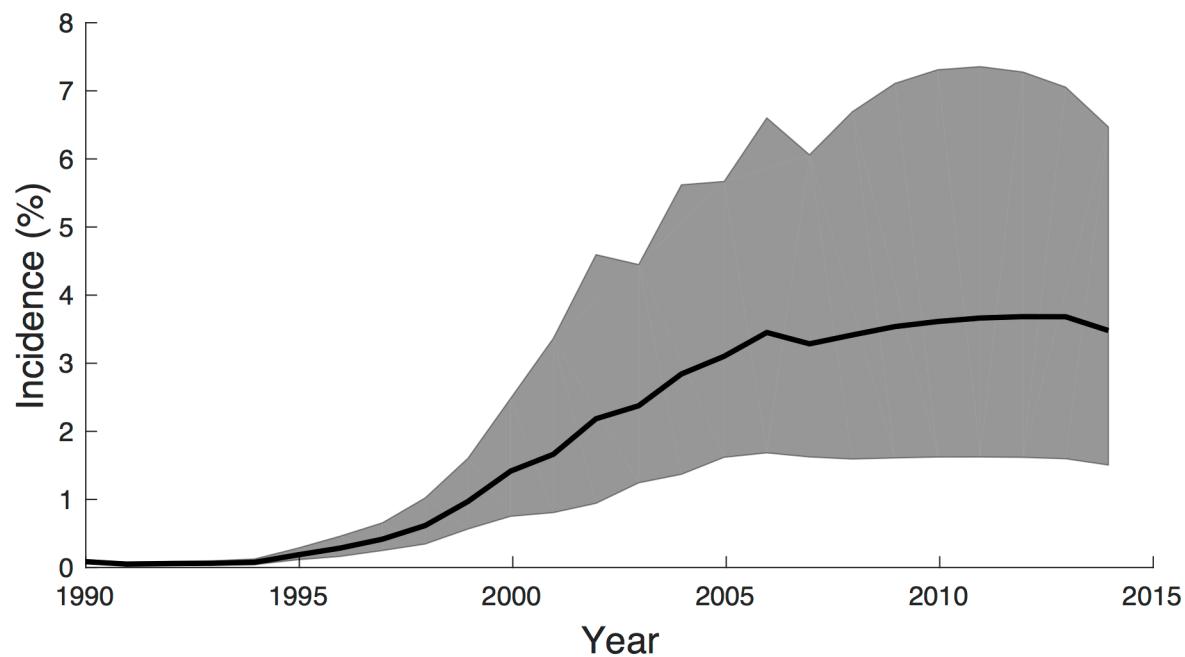
HIV-1 in rural Kwa-Zulu Natal



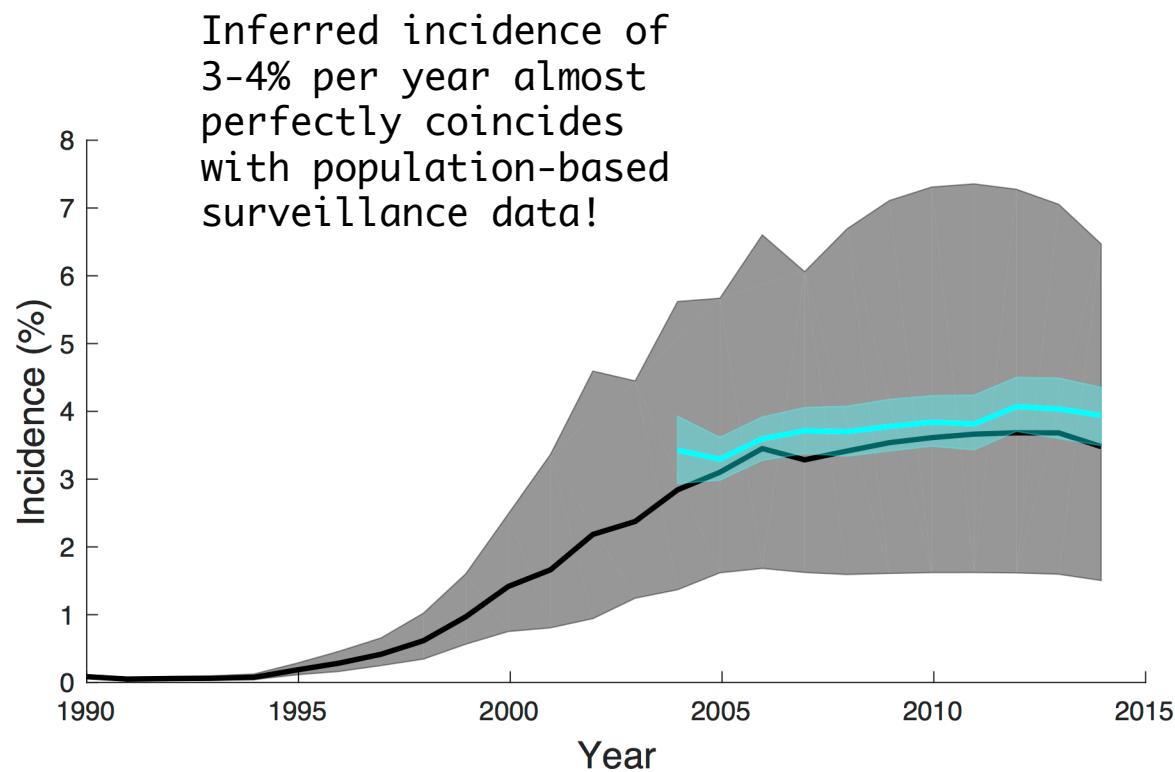
Tracking epidemic dynamics



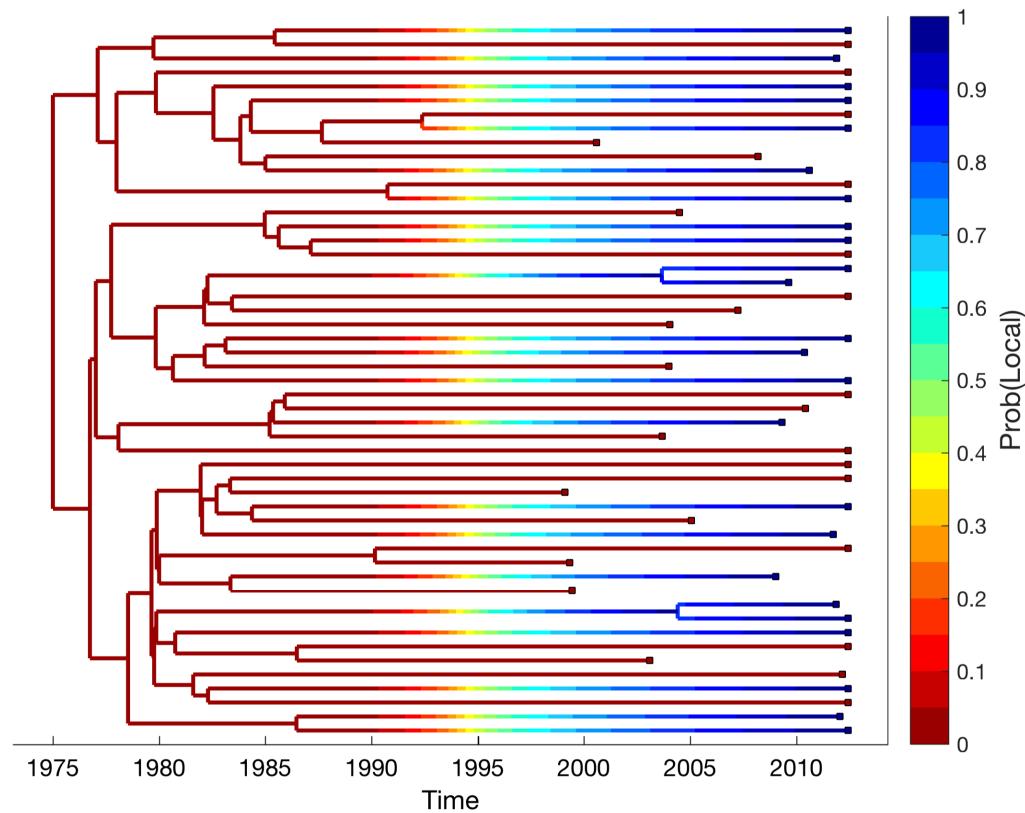
Estimated total incidence



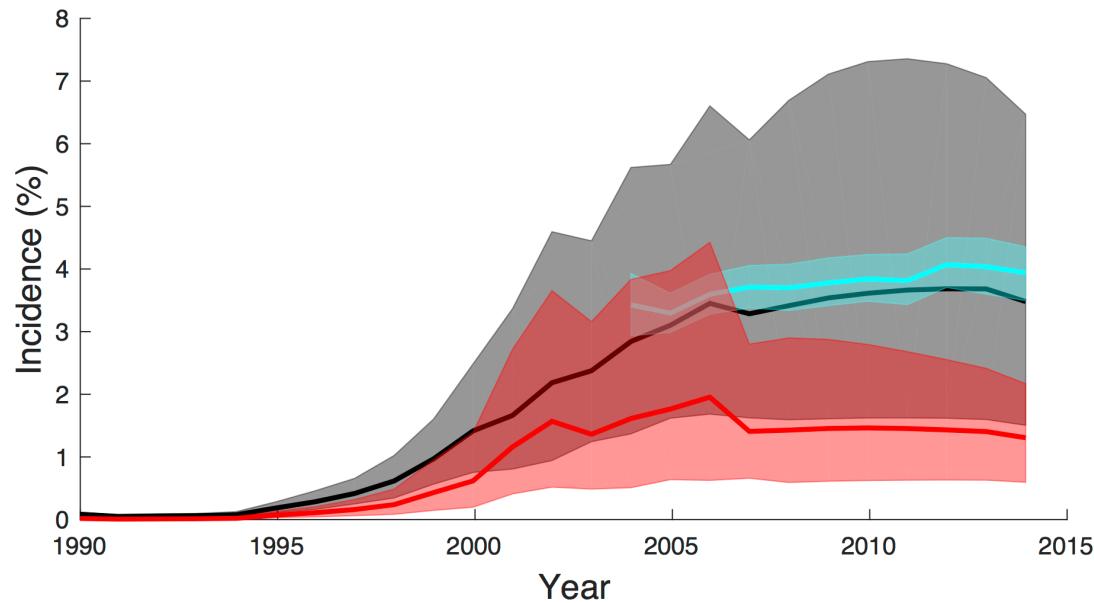
Estimated total incidence



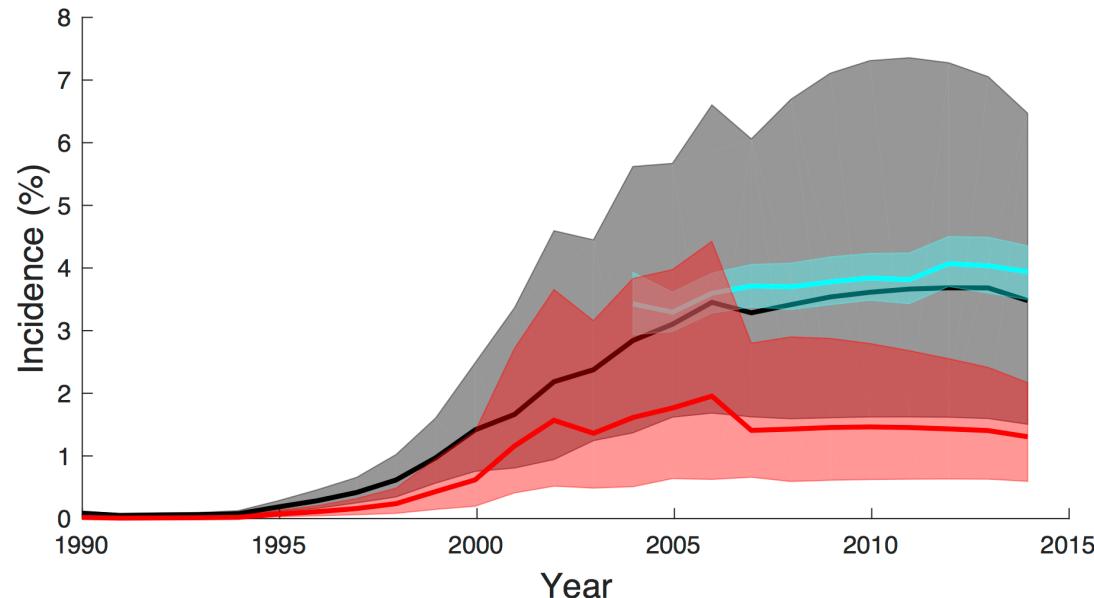
Tracking lineage movement



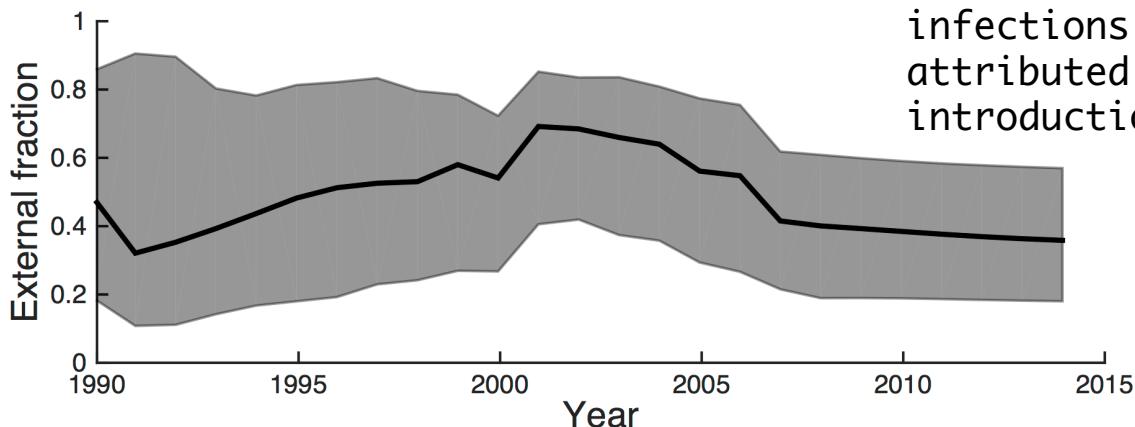
Estimated external incidence



Estimated external incidence



As of 2014, 35% of new infections were attributed to external introductions.

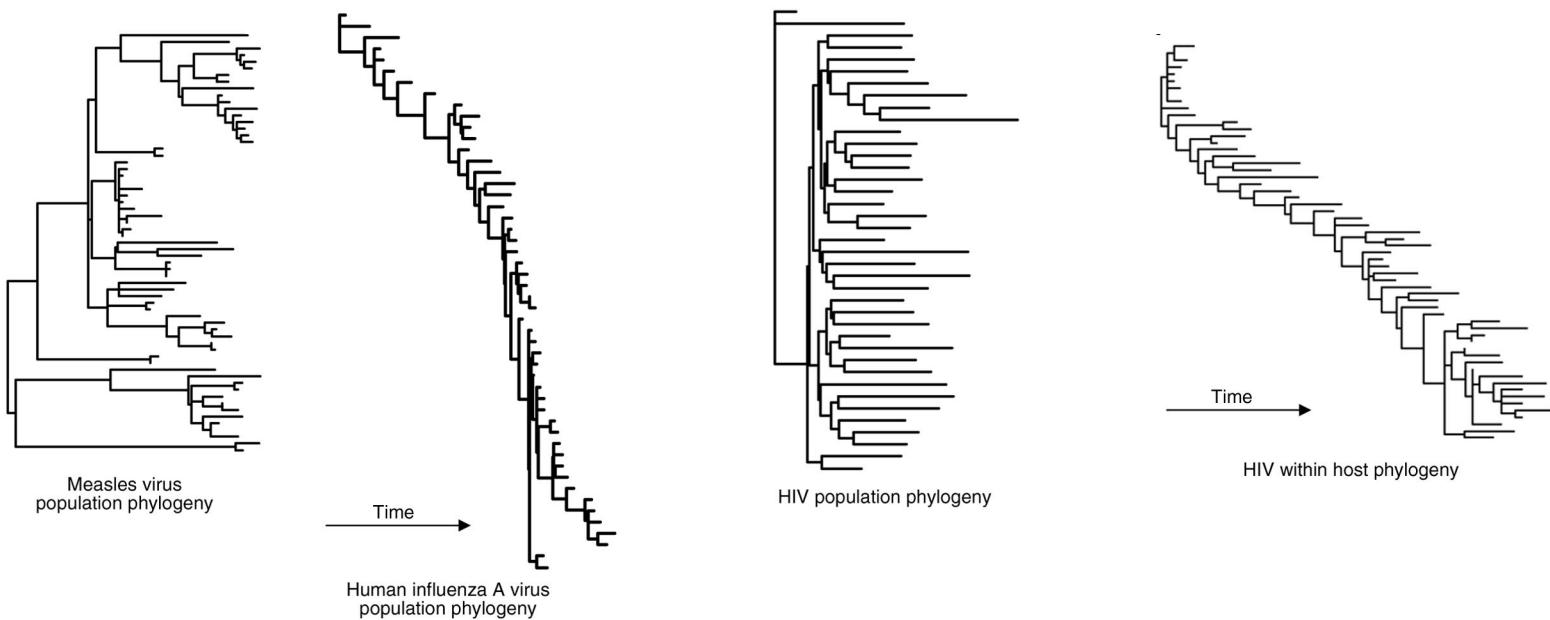


Phylogenies can tell us about:

- The origins of epidemics and new strains
- Linkage – the sources of transmission
- Past epidemic/population dynamics
- Pathogen fitness and adaptation

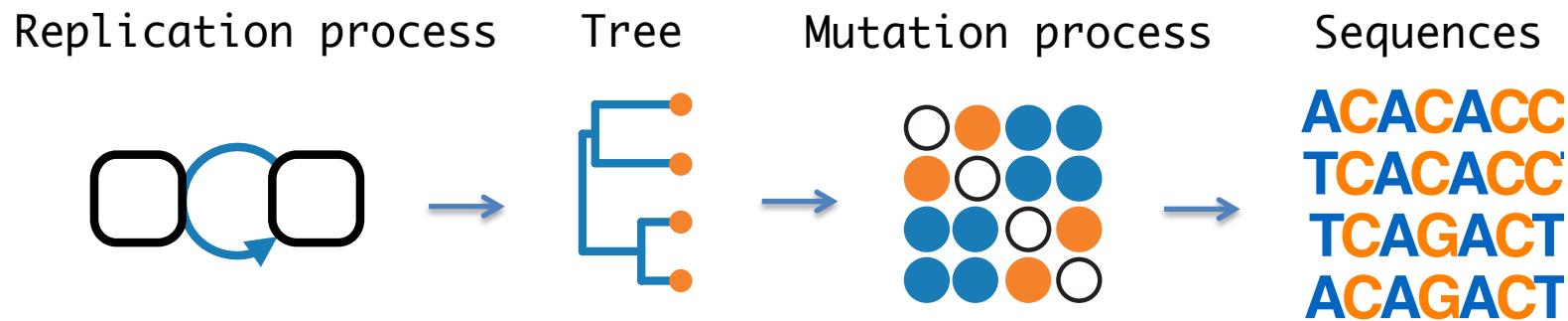
Introduction

- What processes shape the phylogenetic history of different pathogens?

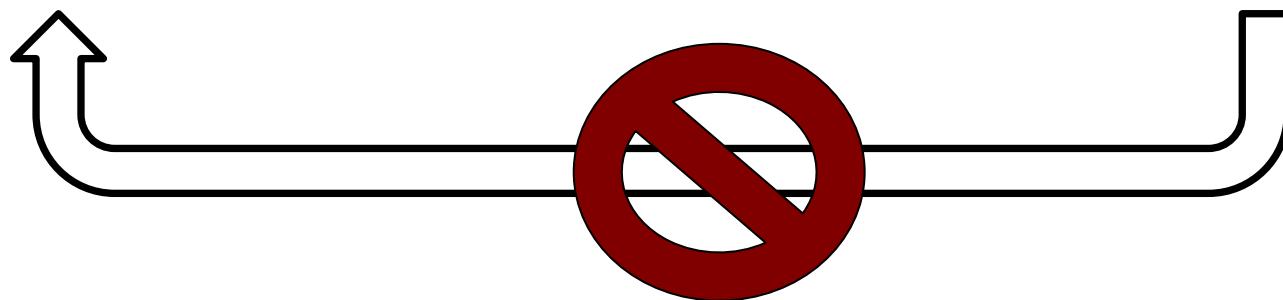
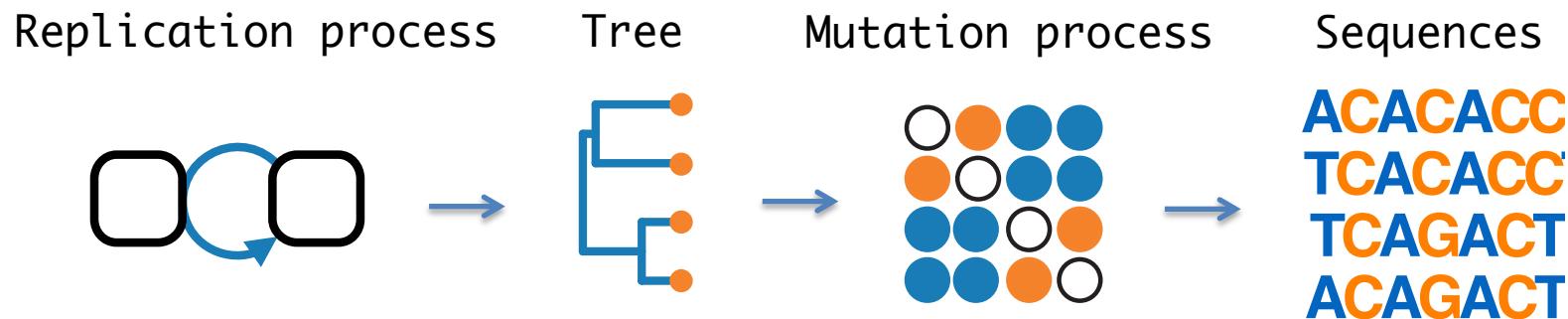


Trees from Grenfell *et al.* (Science, 2004)

The big assumption of phylogenetic models



The big assumption of phylogenetic models

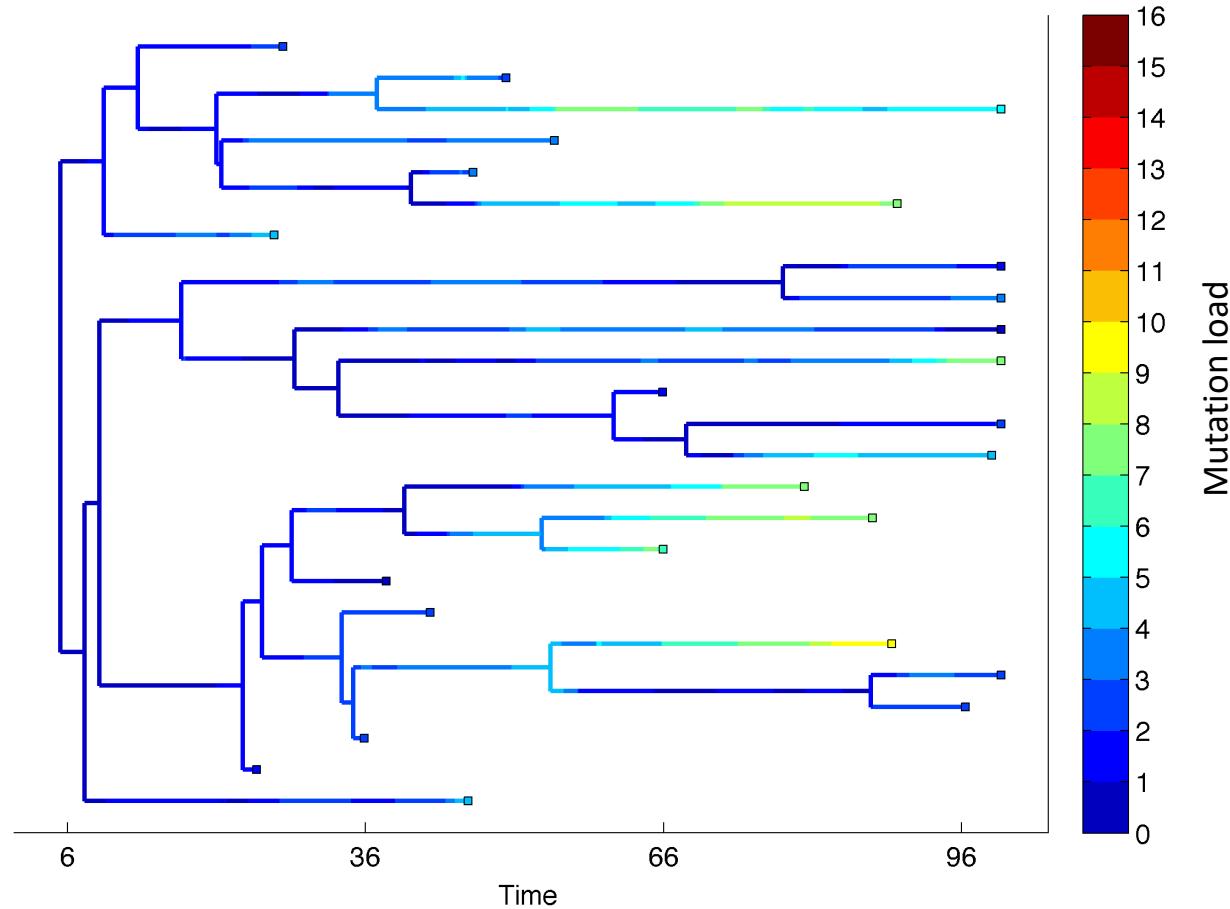


The independence assumption

- This independence assumption allows us to factor the joint likelihood of a tree T and sequence data S into terms we can easily compute:

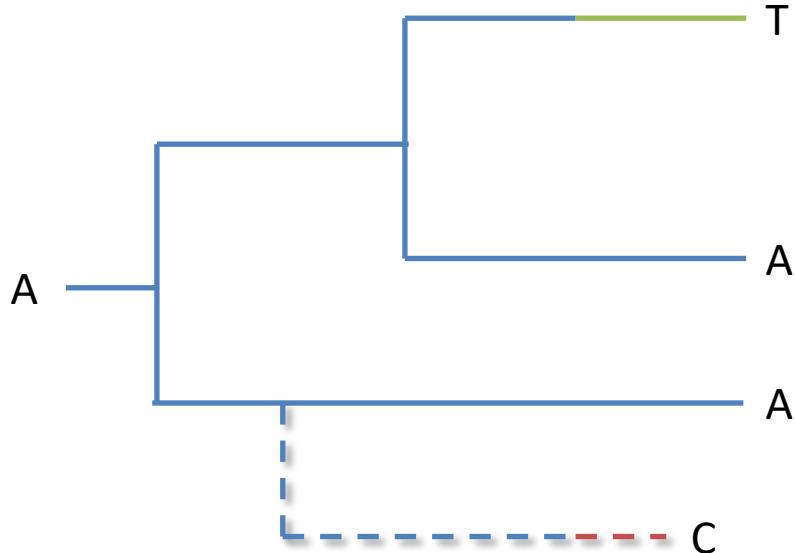
$$L(S, T | \mu, \theta) = L(S|T, \mu)p(T|\theta).$$

Example: deleterious mutation load



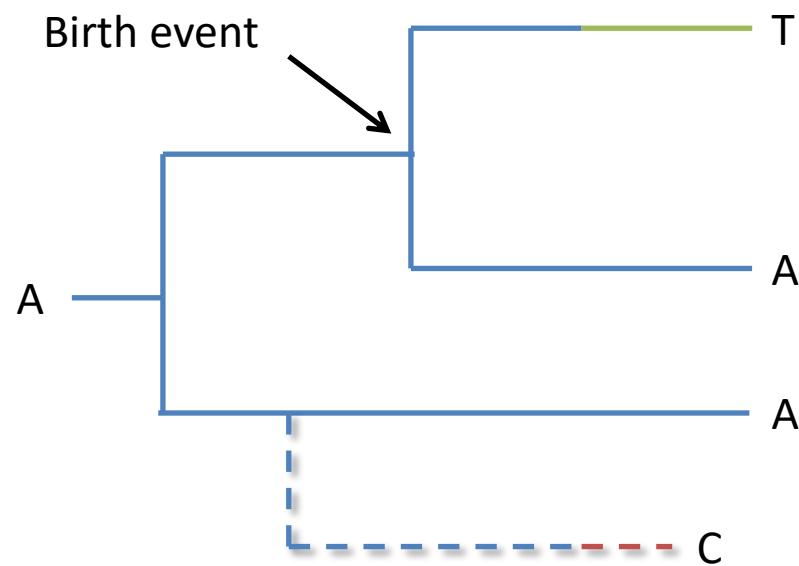
The multi-type birth-death model

- Allows us to compute the joint likelihood that a tree and genotype data at a single loci evolved under a non-neutral model (Stadler & Bonhoeffer, 2013).

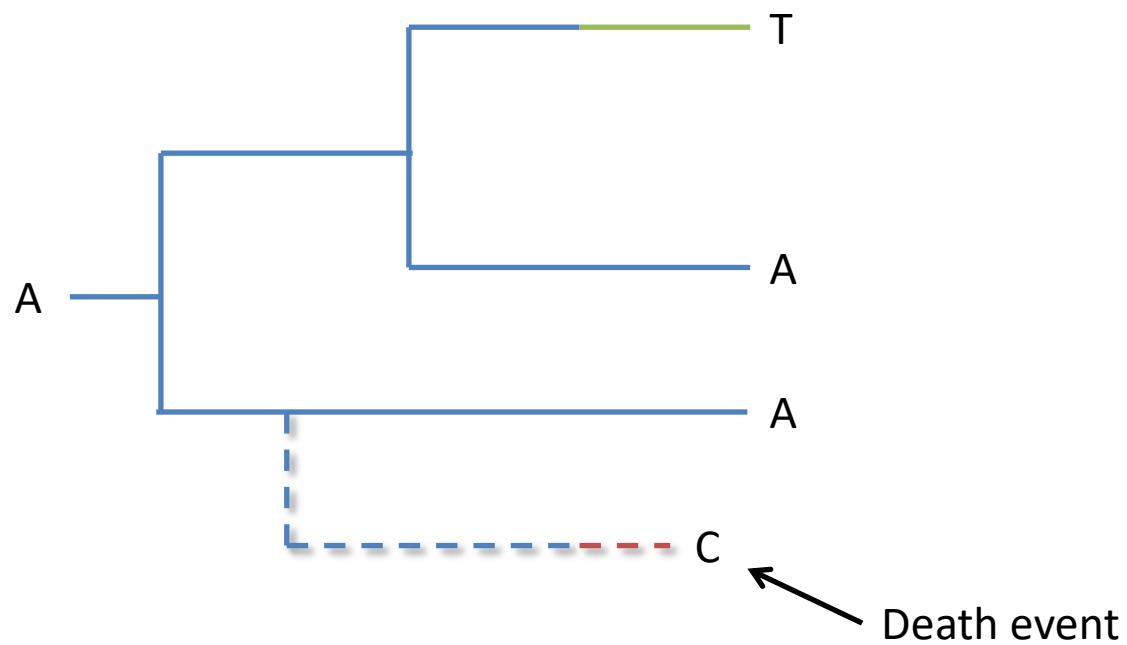


Tanja Stadler

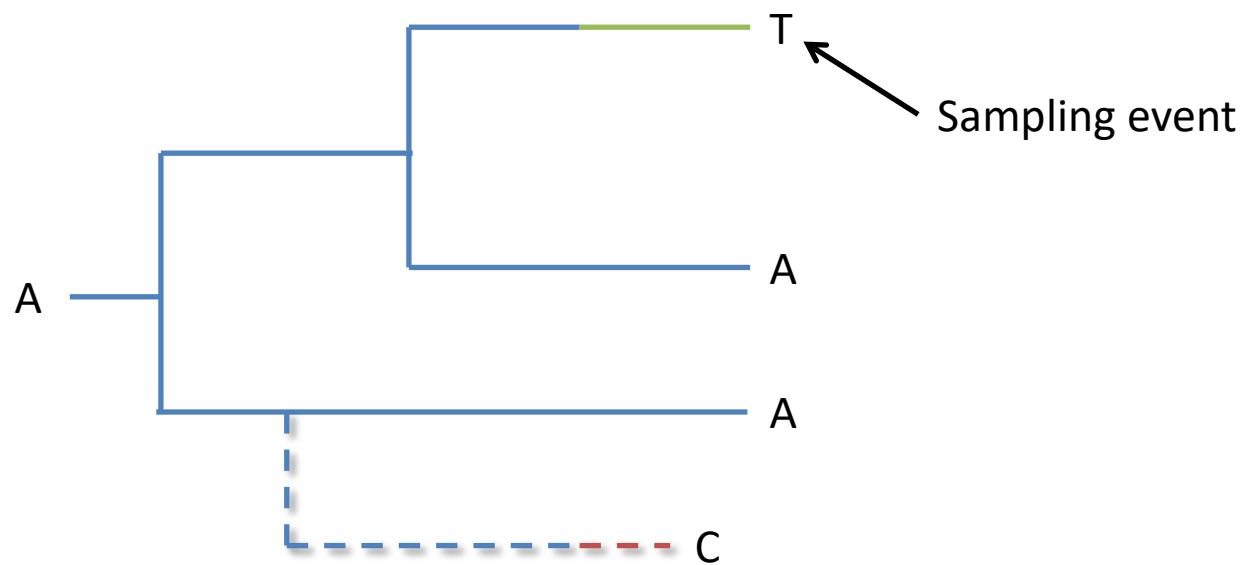
The multi-type birth-death model



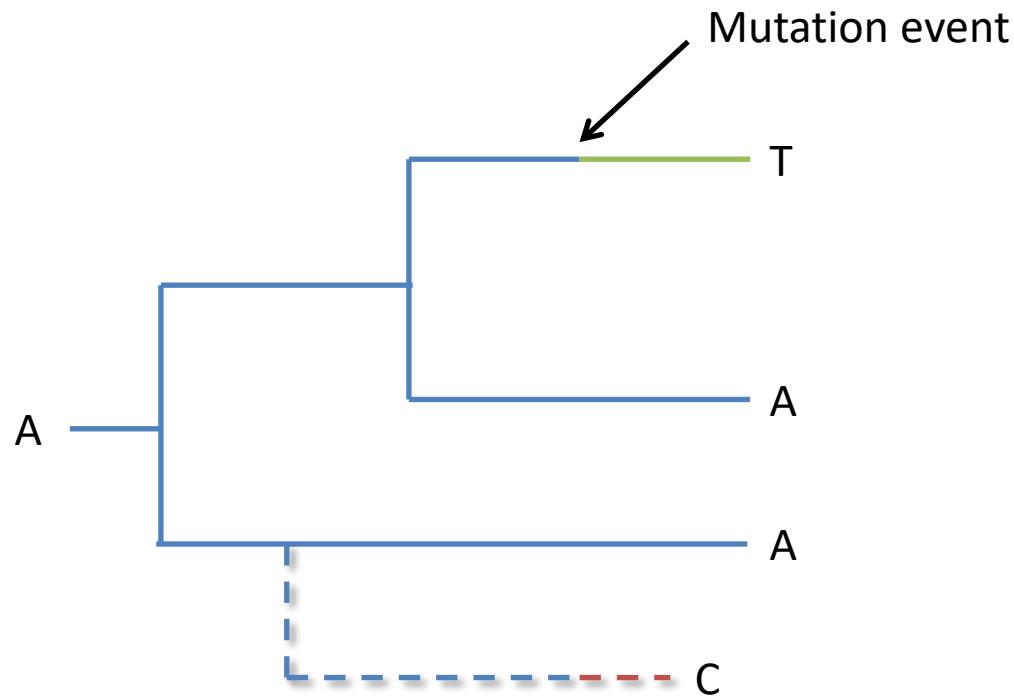
The multi-type birth-death model



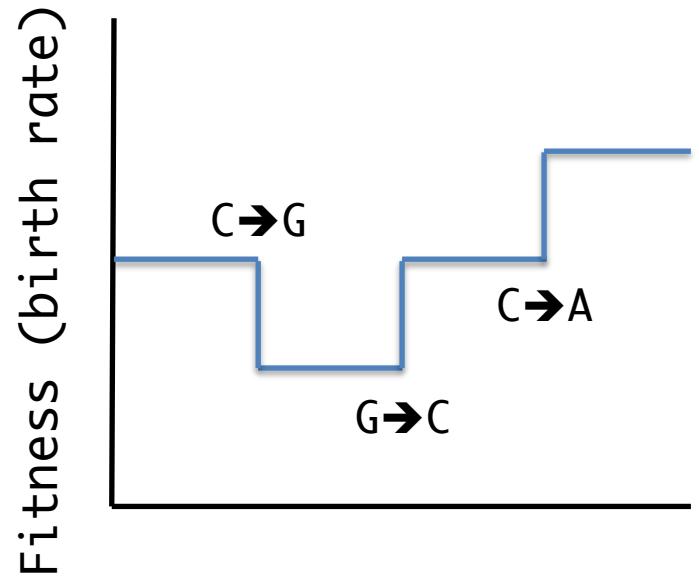
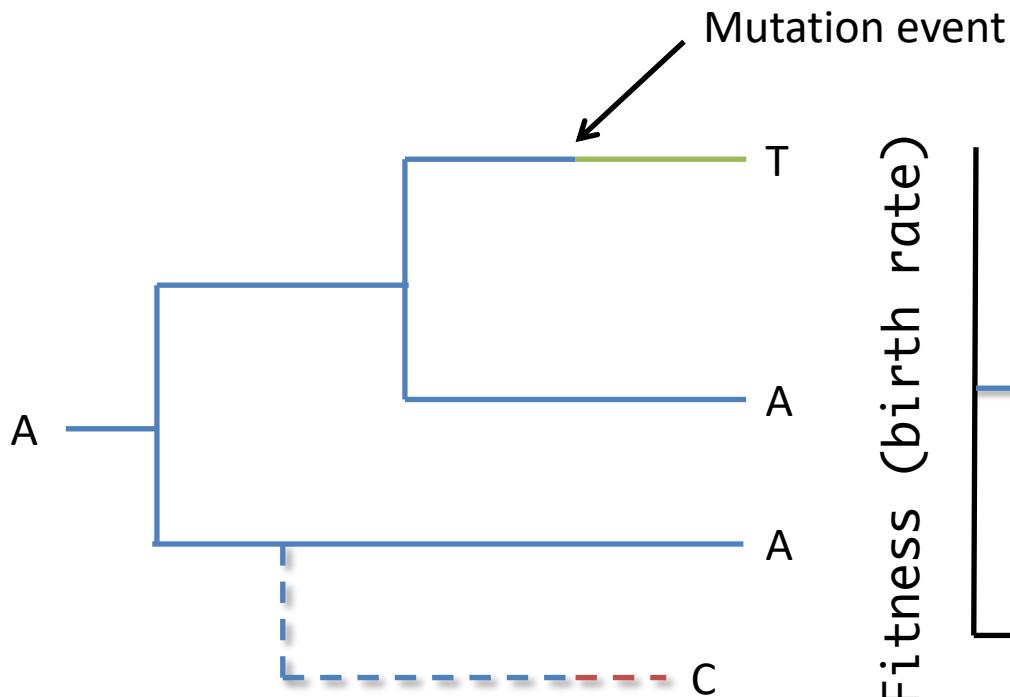
The multi-type birth-death model



The multi-type birth-death model

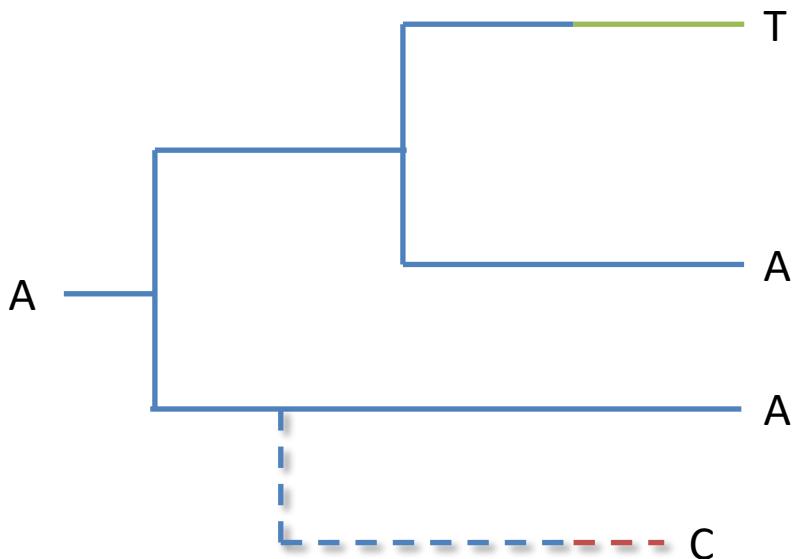


The multi-type birth-death model



The multi-type birth-death model

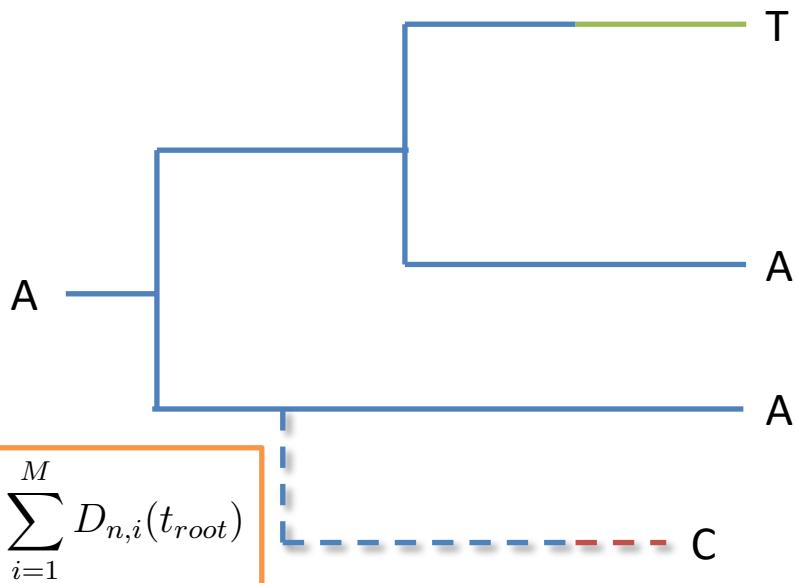
- At a single evolving site, we can compute the joint likelihood of a tree and the ‘sequence’ at each tip using a multi-type birth-death model (Stadler & Bonhoeffer, 2013).



$$\begin{aligned}\frac{d}{dt}D_{n,i}(t) = & -(\lambda_i + \sum_{j=1}^M \gamma_{i,j} + d_i)D_{n,i}(t) \\ & + 2\lambda_i E_i(t)D_{n,i}(t) \\ & + \sum_{j=1}^M \gamma_{i,j} D_{n,j}(t).\end{aligned}$$

The multi-type birth-death model

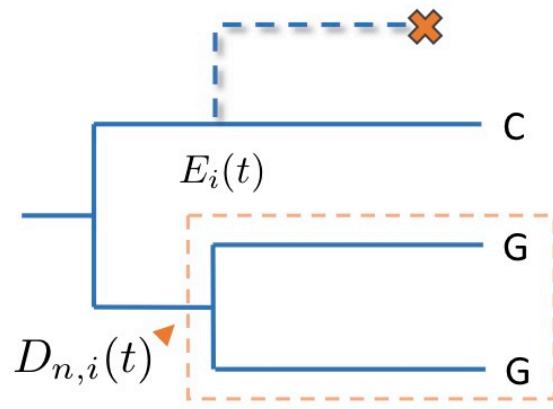
- At a single evolving site, we can compute the joint likelihood of a tree and the ‘sequence’ at each tip using a multi-type birth-death model (Stadler & Bonhoeffer, 2013).



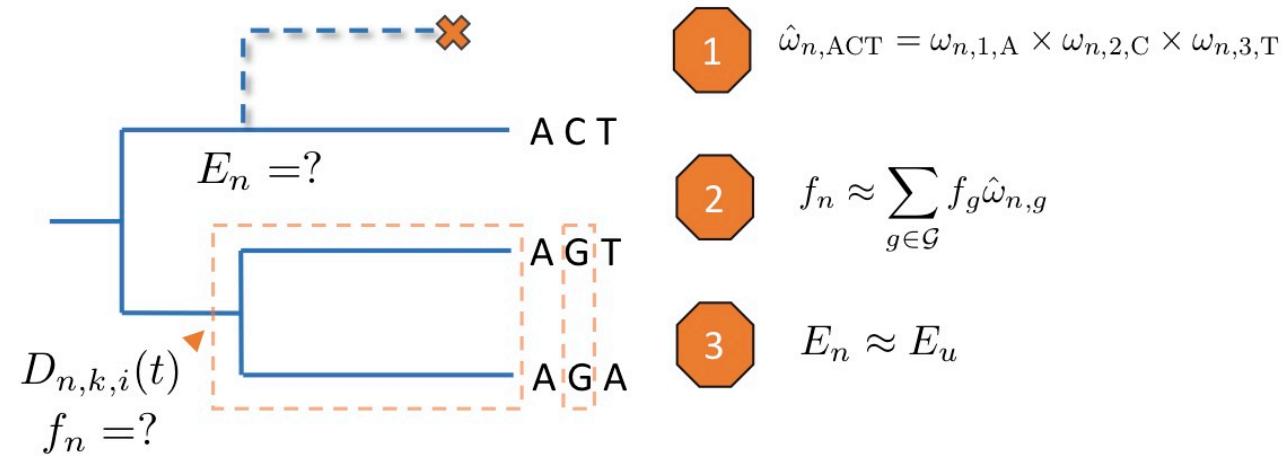
$$\begin{aligned}\frac{d}{dt} D_{n,i}(t) = & - (\lambda_i + \sum_{j=1}^M \gamma_{i,j} + d_i) D_{n,i}(t) \\ & + 2\lambda_i E_i(t) D_{n,i}(t) \\ & + \sum_{j=1}^M \gamma_{i,j} D_{n,j}(t).\end{aligned}$$

The marginal fitness birth-death model

The Multi-Type Birth-Death Model

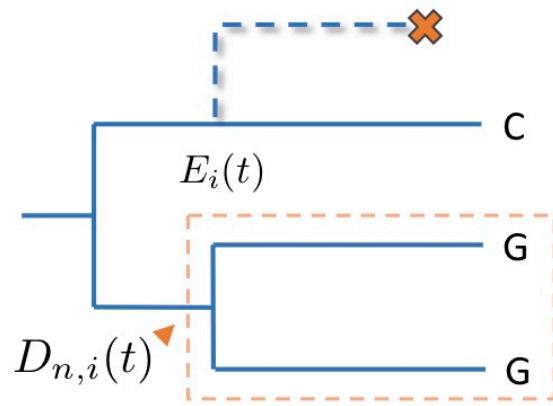


The Marginal Fitness Birth-Death Model

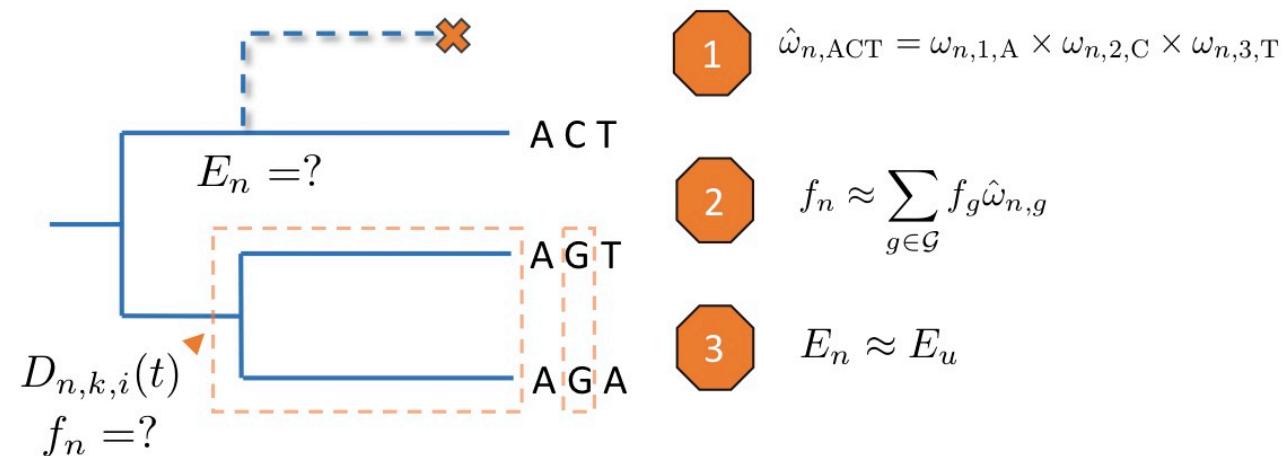


The marginal fitness birth-death model

The Multi-Type Birth-Death Model

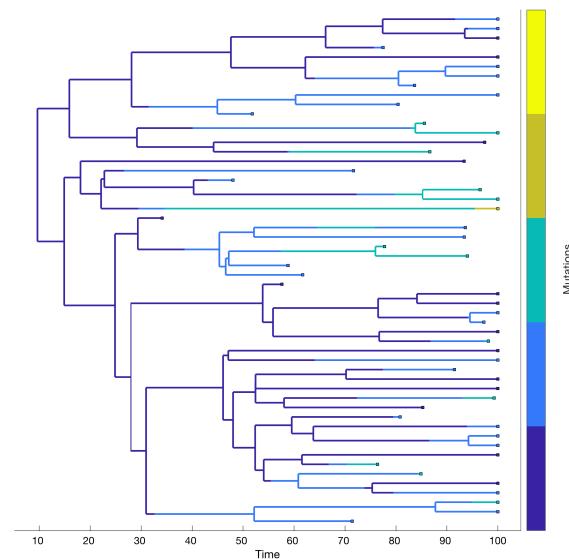


The Marginal Fitness Birth-Death Model

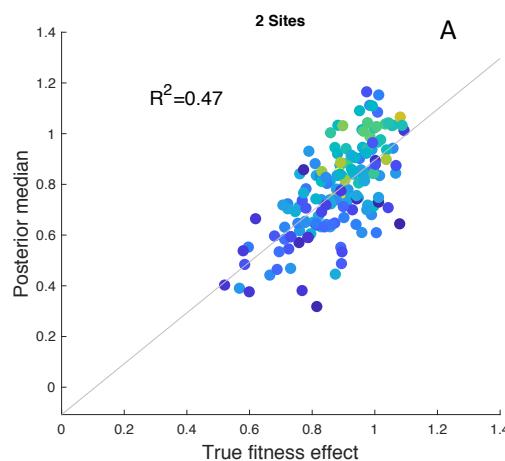


The important part: The MFBD model allows us to consider how selection shapes sequence evolution at multiple sites while considering how mutations act together to shape the fitness of a lineage.

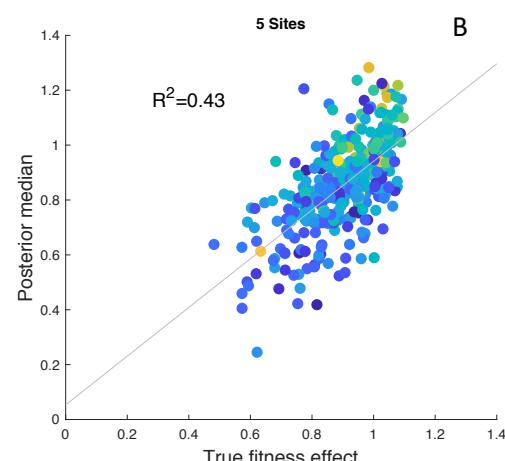
Results: quantifying site-specific effects



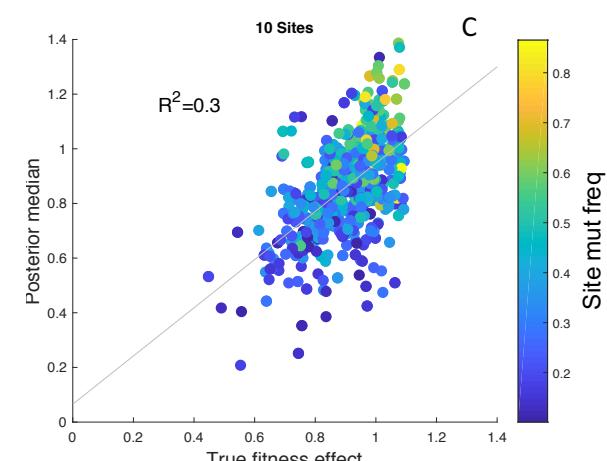
2 Sites



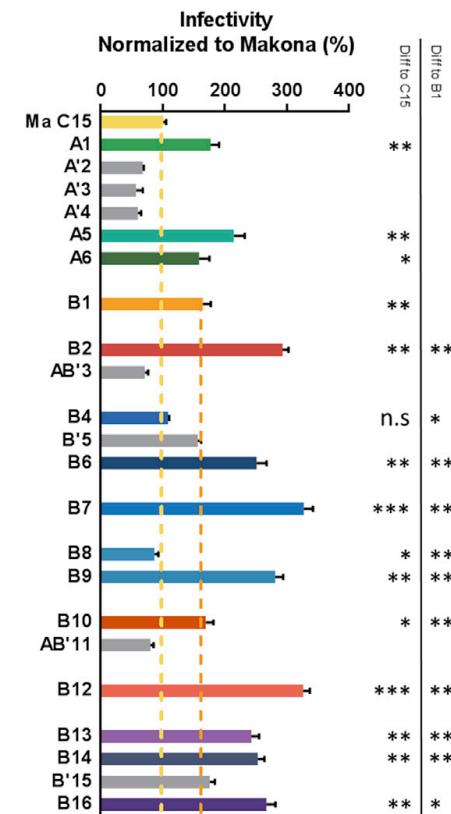
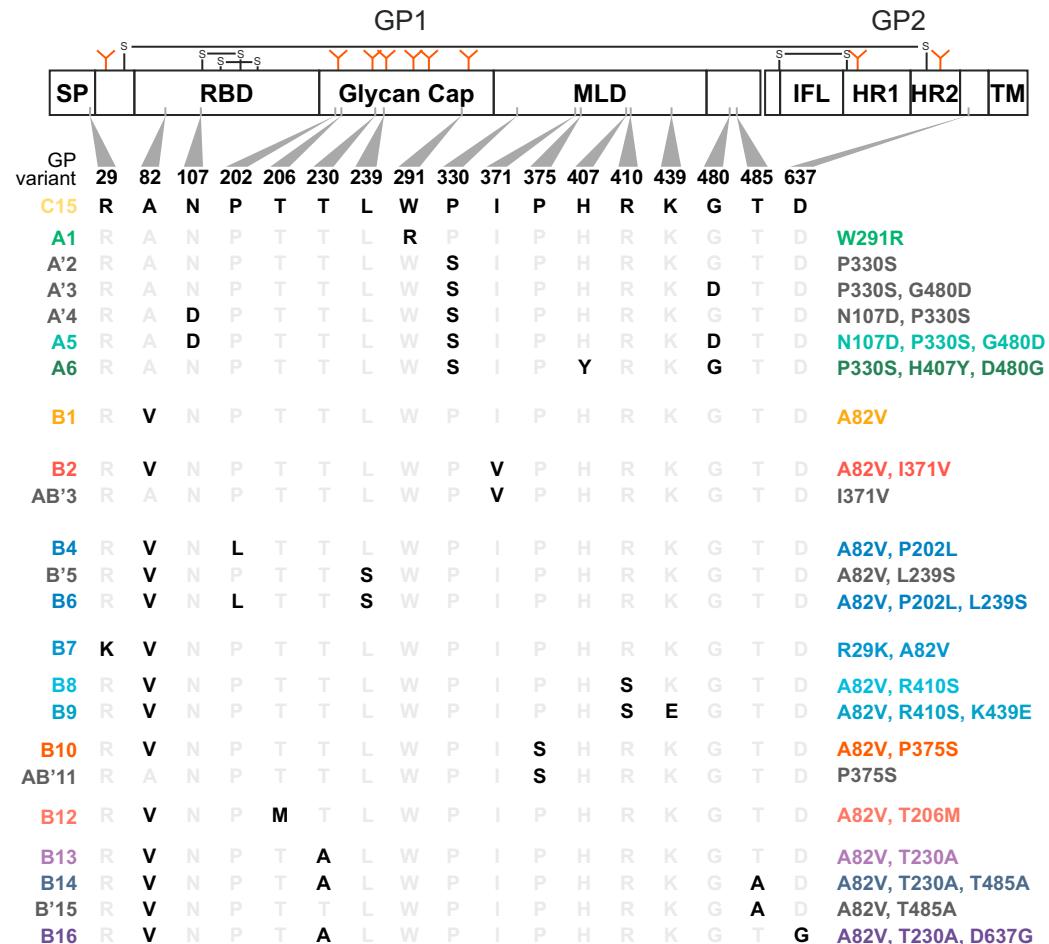
5 Sites



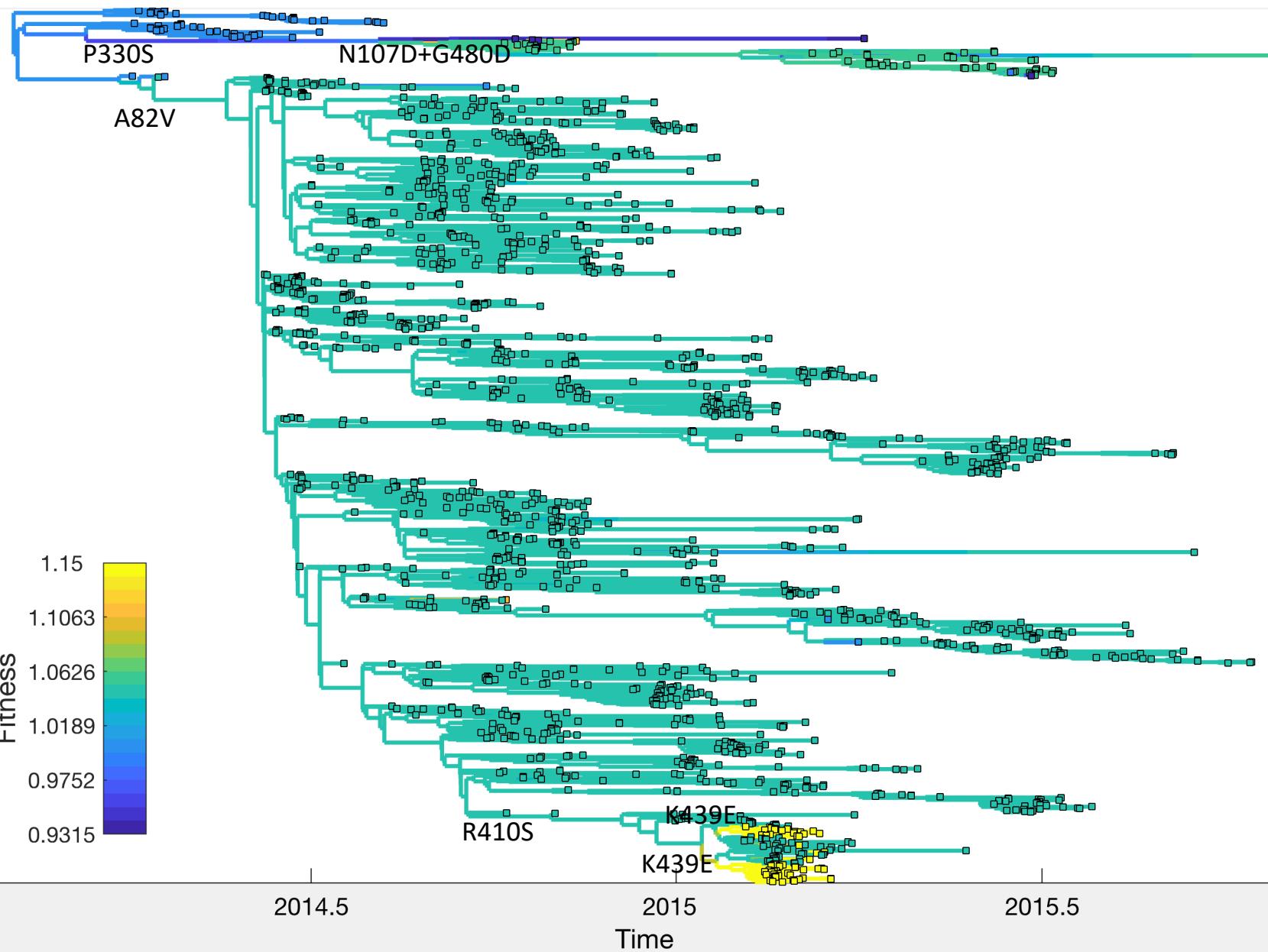
10 Sites



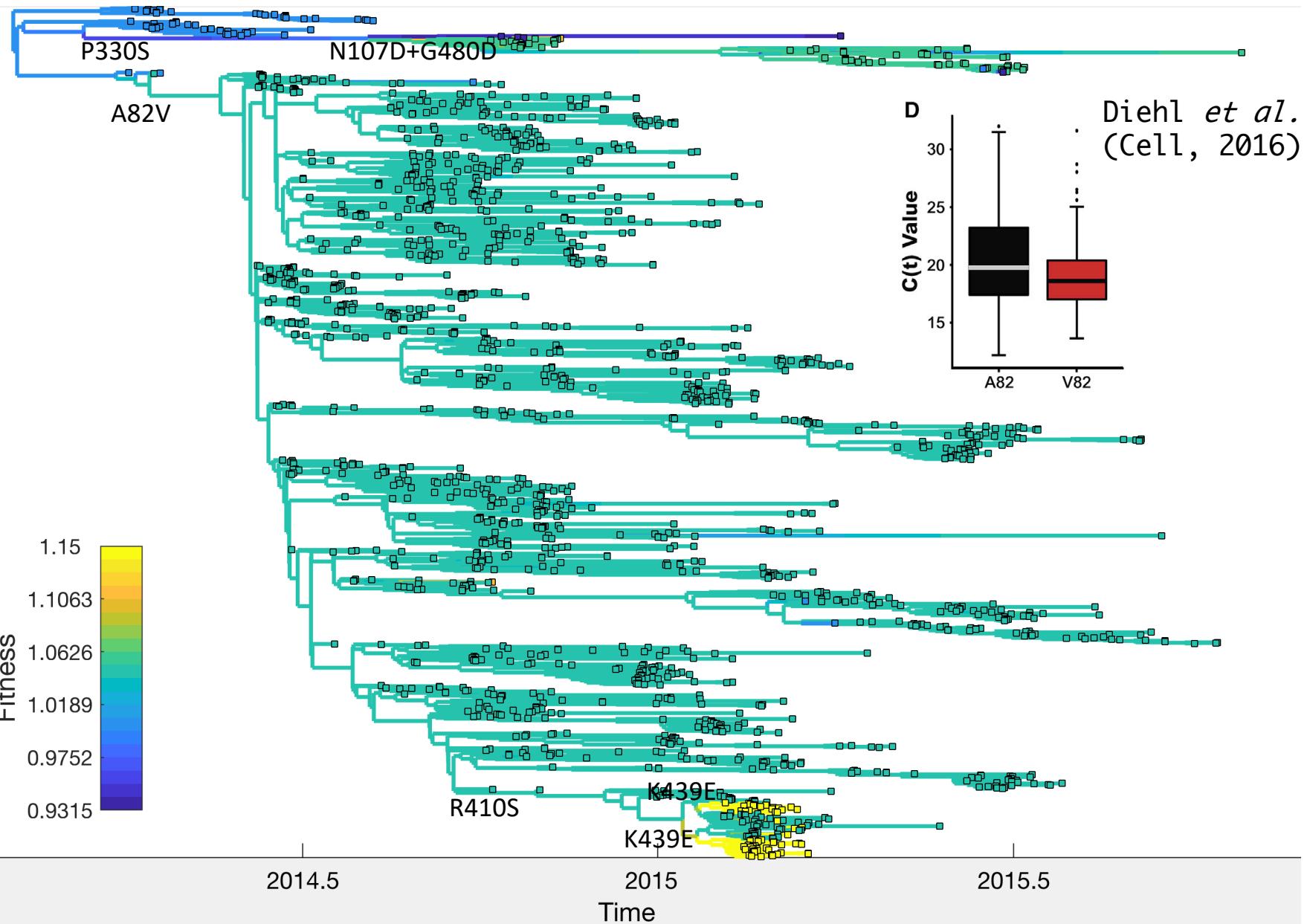
Adaptation of Ebola virus to humans



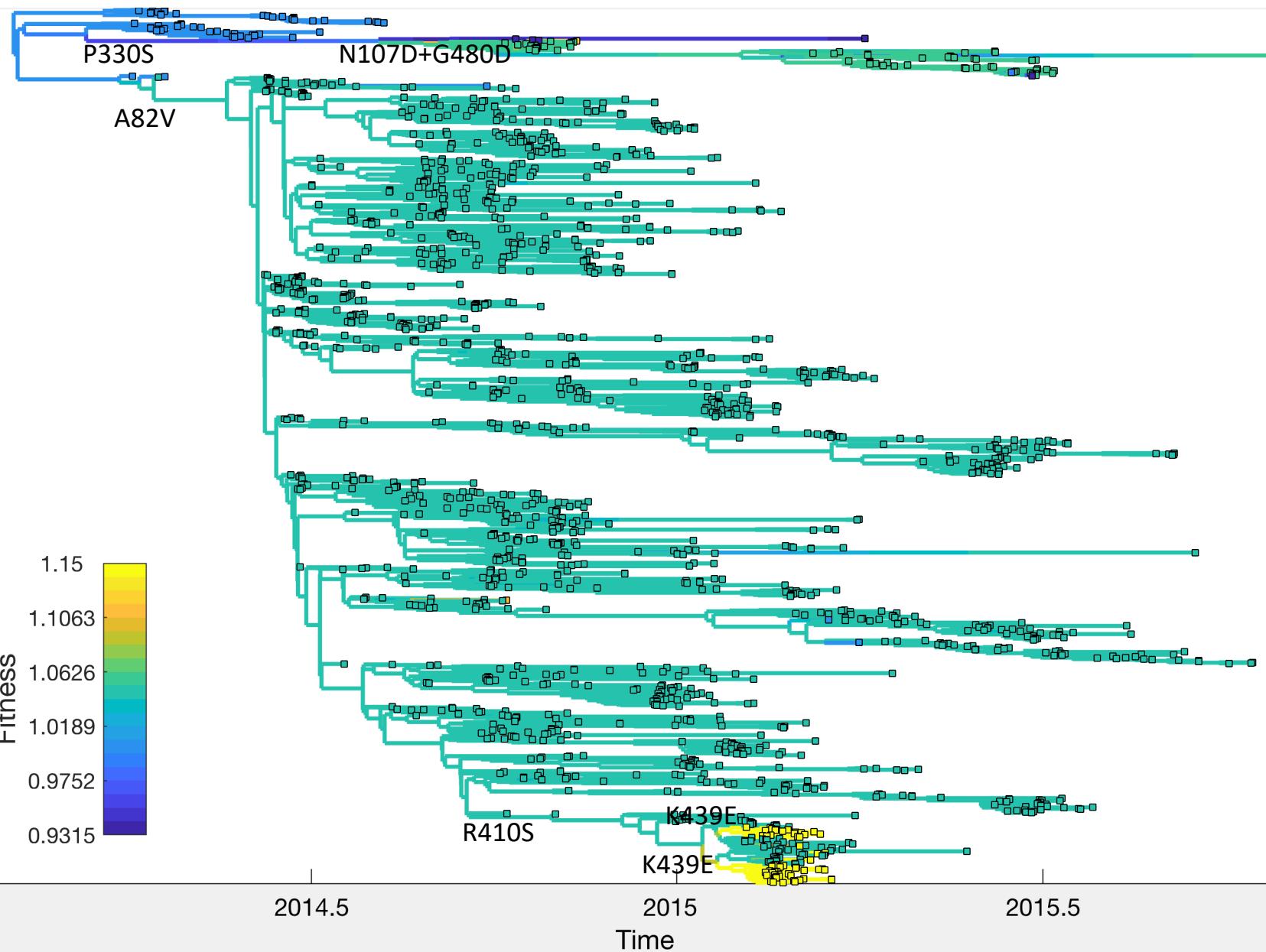
Adaptation of Ebola virus to humans



Adaptation of Ebola virus to humans



Adaptation of Ebola virus to humans



Marginal fitness birth-death model

- Marginal Fitness Birth Death model is implemented in *LUMIERE* in BEAST2.



- Code is available on GitHub but user-friendly add-on package is still in progress.
- Also provides a way to fit multi-trait, multi-type birth-death models.