NC STATE UNIVERSITY

Phylogenetic insights into infectious disease epidemiology

Molecular Epidemiology of Infectious Diseases

Lecture 1

January 8th, 2024

Genomic data has given us new power to track the spread of infectious pathogens

Course overview

“This course will focus on how phylogenetic and population genomic methods are used to track the spread of infectious diseases using pathogen genomic data. ***We will explore how models and methods can be adapted to the epidemiology and natural history of different pathosystems, including viral, bacterial and fungal pathogens in plants, animals and humans.*** Topics include reconstructing epidemic dynamics, spatial movement (phylogeography), transmission networks, recombination and adaptive evolution.”

Hourglass format of course

Starting from very different backgrounds

Core phylogenetic methods applicable across systems

More targeted applications and team projects

Weekly course structure

The course will meet twice per week.

The Monday session will generally be a lecture or discussion.

The Wednesday session will be tutorial-based and provide the opportunity to apply methods to real data with a few optional coding exercises.

Coursework and grades

“Everyone should get a A”

There is no graded work other than a team project focusing on a pathogen and dataset of your choice during the second half of the semester.

But please do:

- Look at the suggested readings.
- Participate in class discussions and tutorials
- Come to class ready to ask questions and discuss problems

Genomic data has given us new power to track the spread of infectious pathogens

The importance of phylogenies

While there are many methods for analyzing pathogen genomic data, this lecture and most of the first half of the semester will examine phylogenetic methods.

Phylogenies describe the ancestral (parent-child) relationships among individuals or taxa in terms of shared descent.

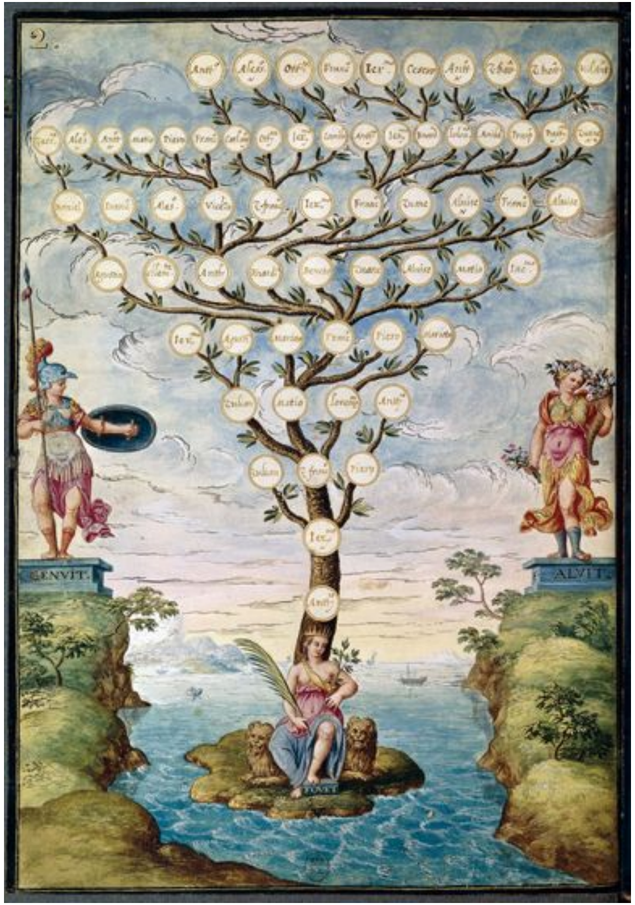


Image from *The Book of Trees* (Manuel Lima, 2014)

Why phylogenies?

1. The branching structure of pathogen phylogenies can be directly related back to the transmission process.

2. Thinking phylogenetically can help us understand how epidemic dynamics shape genetic variation in a pathogen population.

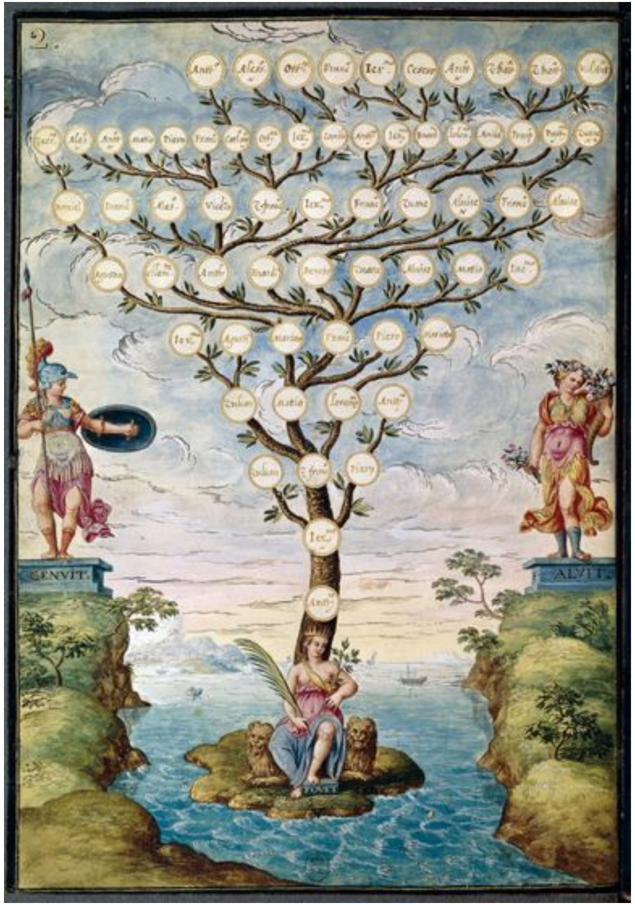
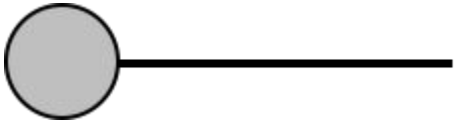


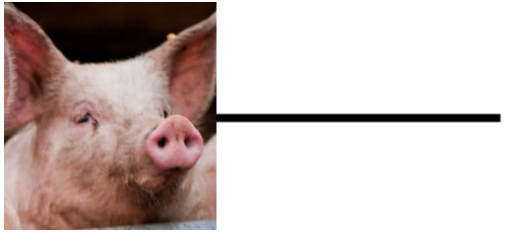
Image from *The Book of Trees* (Manuel Lima, 2014)

Let’s start by considering a small epidemic spreading through a host population

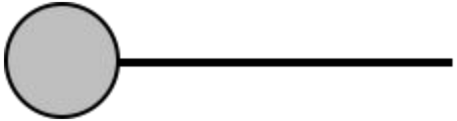
A simple epidemic example



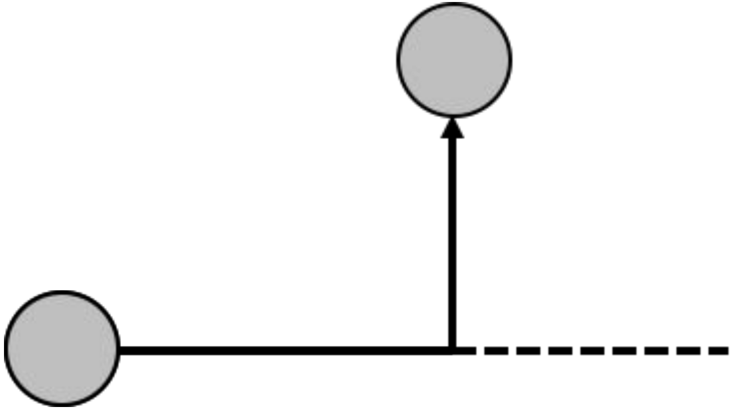
A simple epidemic example



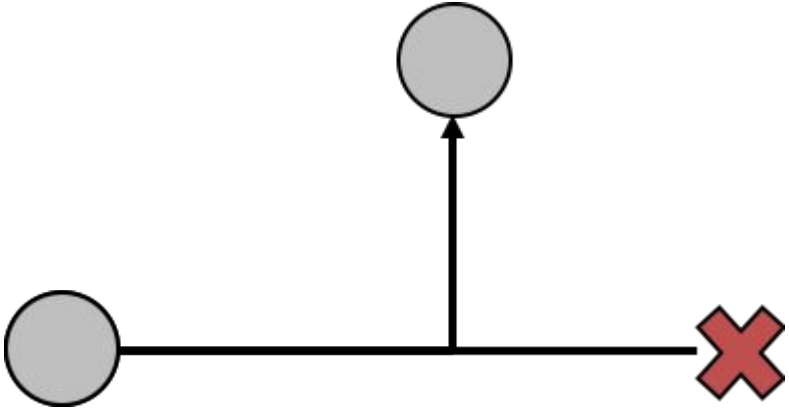
A simple epidemic example



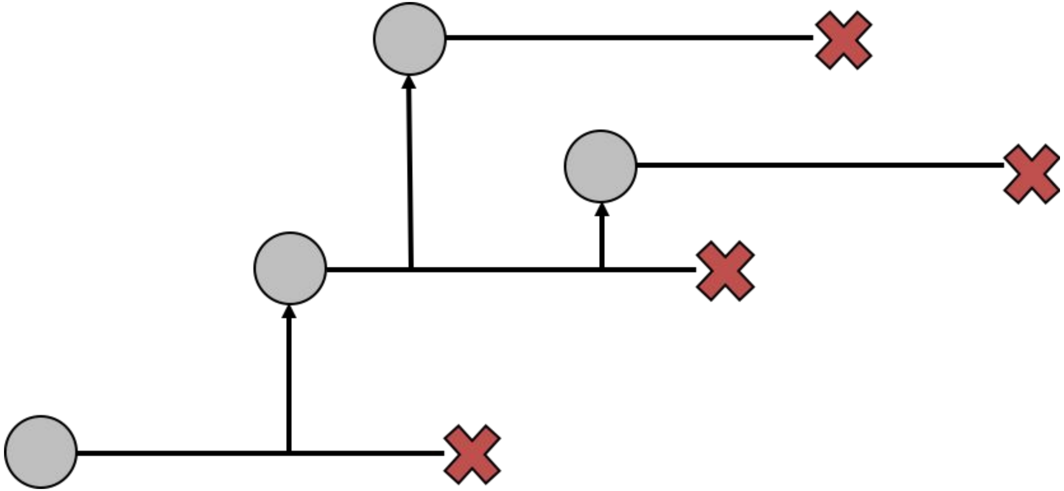
A simple epidemic example



A simple epidemic example

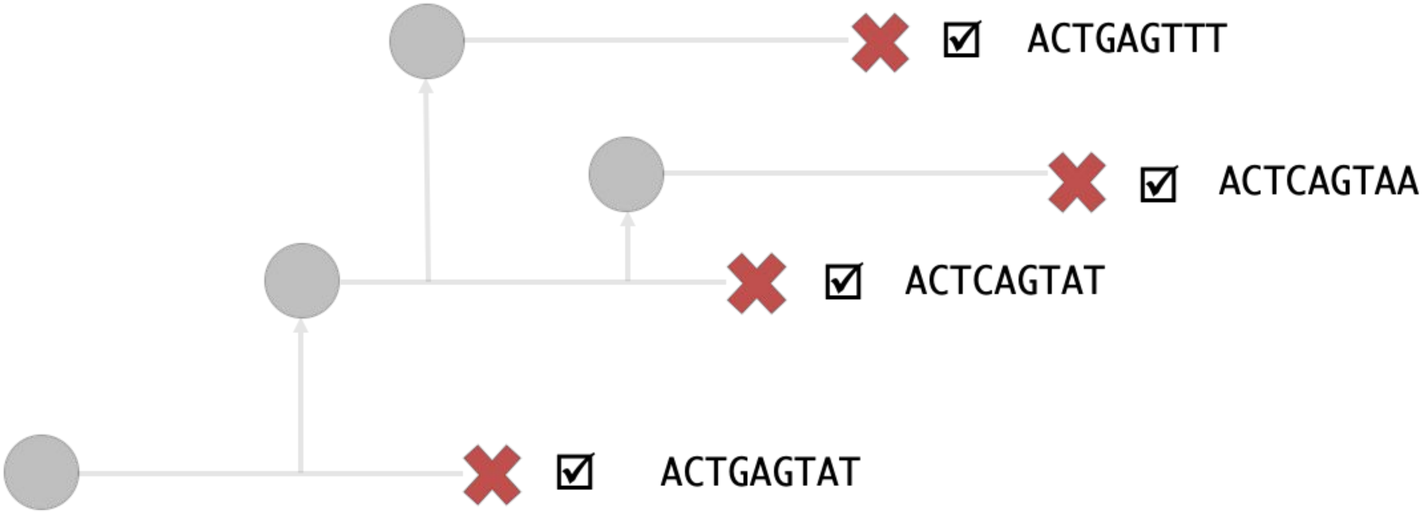


A simple epidemic example

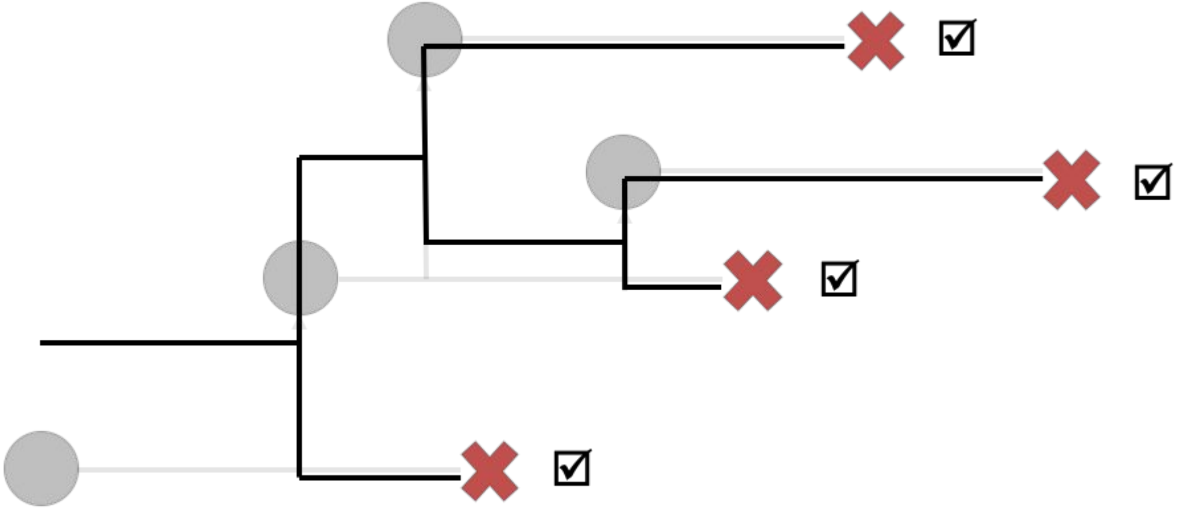


Transmission tree

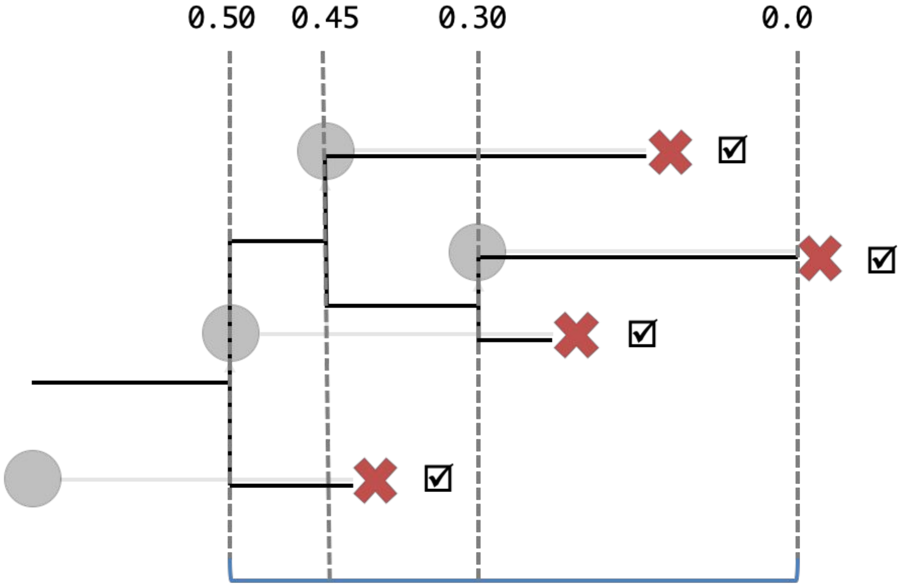
A simple epidemic example



A simple epidemic example



A simple epidemic example



$$\text{Real time} = \frac{\text{genetic distance}}{\text{clock rate}}$$

Phylogenies can tell us about:

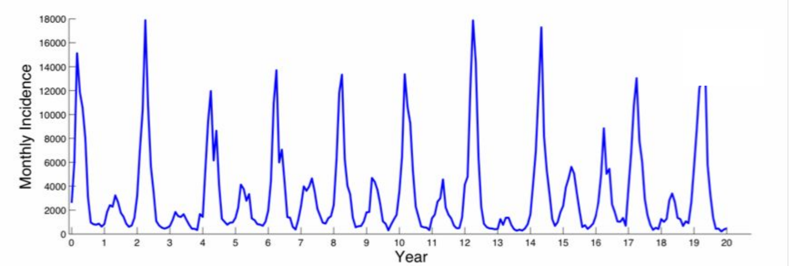
- Linkage and the sources of transmission
- The origins of epidemics and new strains
- Past epidemic dynamics
- Pathogen fitness and adaptation

Phylogenies can tell us about:

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

Revealing the source of infections

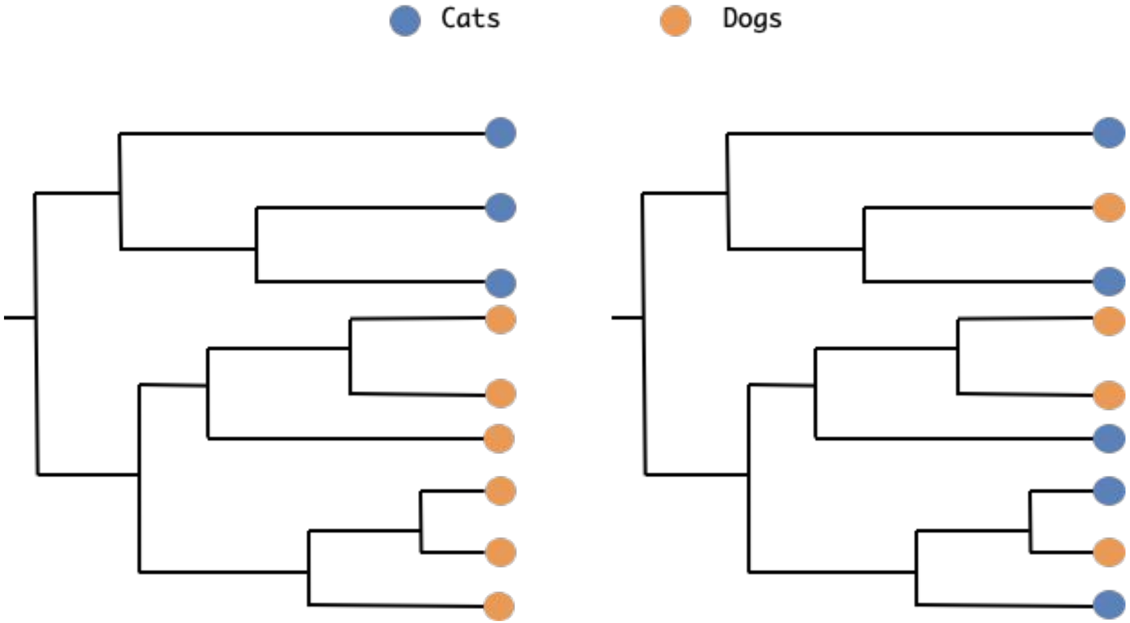
Classic sources of epidemiological data like time series of case reports are typically not informative about the sources of new infections



The genetic relatedness of pathogens sampled from different hosts or environments provides us with information about possible transmission routes including **the source of new infections.**

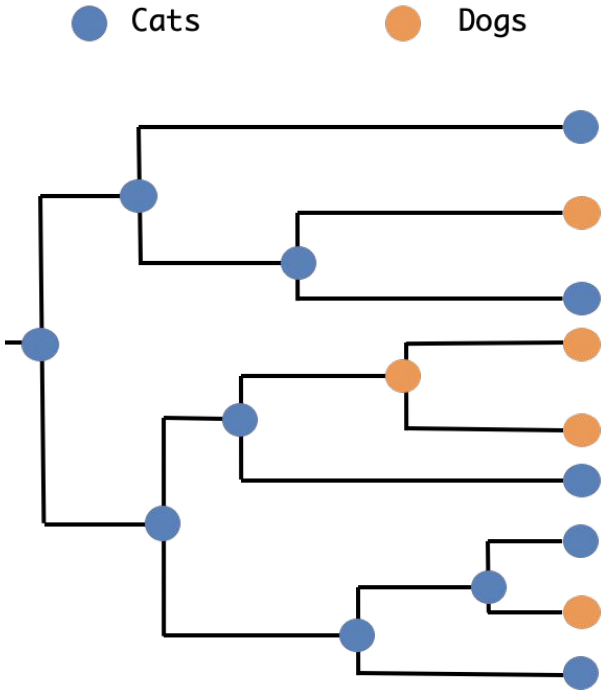
Phylogenetic linkage

We can “link” or connect infections to determine who might be infecting whom based on phylogenetic relationships.



Ancestral state reconstruction

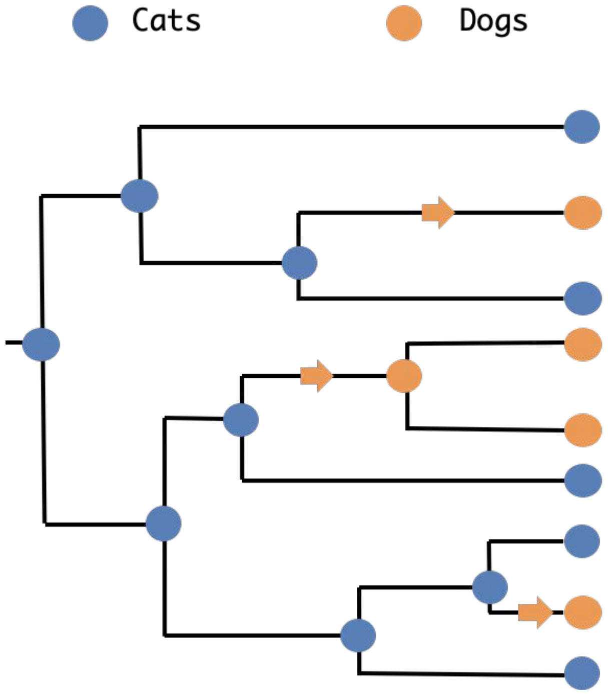
Ancestral state reconstruction allows us to infer the location/host of past transmission events.



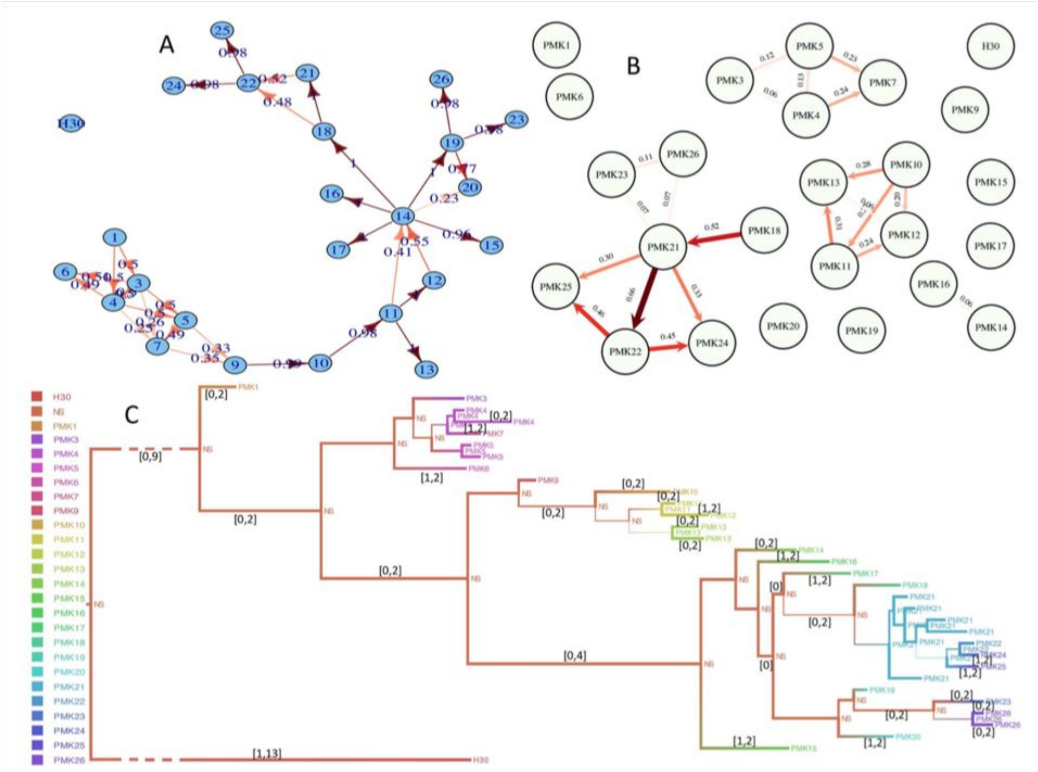
Ancestral state reconstruction

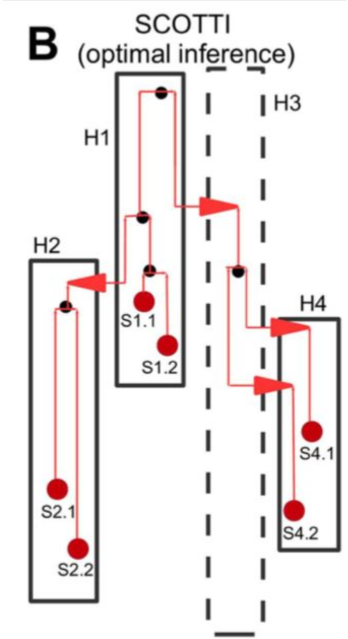
Ancestral state reconstruction allows us to infer the location/host of past transmission events.

Ancestral states can therefore allow us to infer the direction of infection.



Klebsiella transmission trees





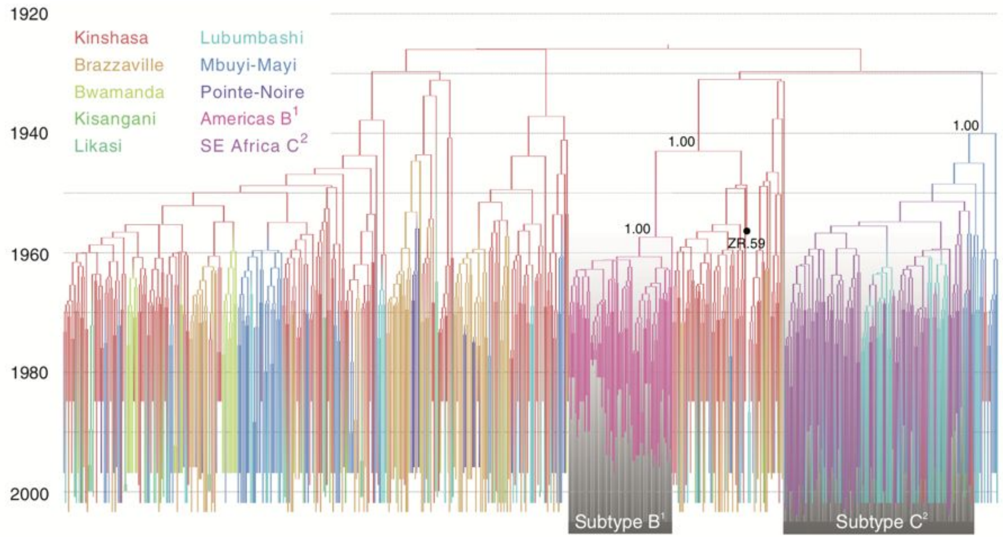
De Maio *et al.* (PCB, 2016)

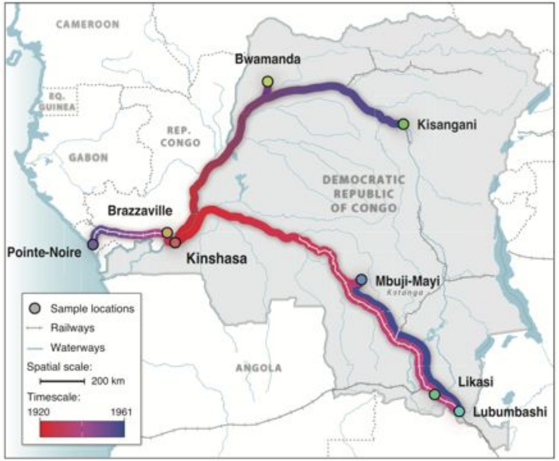
Phylogenies can tell us about:

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

Origins of the HIV-1 epidemic

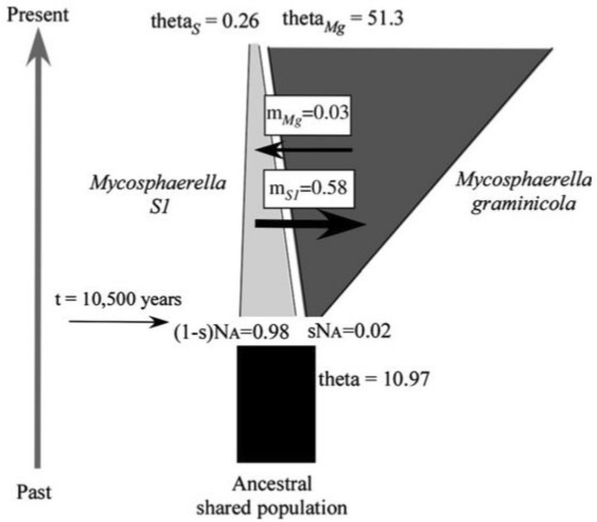
Faria *et al.* (Science, 2014) traced the origins of the HIV-1 epidemic back to the 1920’s and 30’s in Kinshasa, DRC.





Origins of *Mycosphaerella graminicola*

Stukenbrock *et al.* (MBE, 2006) traced the fungal pathogen causing septoria leaf blotch on wheat back to 8,000 to 9,000 BC in the Fertile Crescent.



M. graminicola on wheat (Wikipedia)
Now named ***Zymoseptoria tritici***

Neolithic origins of other agro-pathogens

Supports idea that many agriculturally important pathogens today arose during the Neolithic transition to farming.

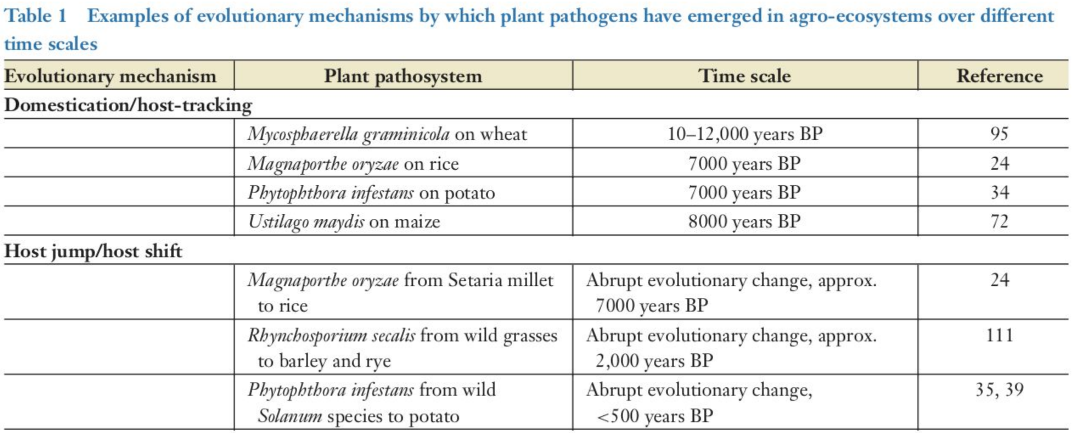
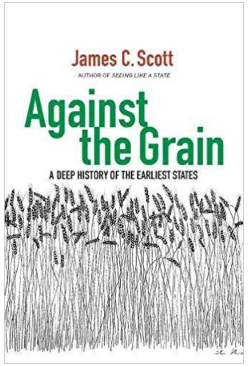


Table 1 Examples of evolutionary mechanisms by which plant pathogens have emerged in agro-ecosystems over different time scales

Evolutionary mechanism	Plant pathosystem	Time scale	Reference
Domestication/host-tracking			
	Mycosphaerella graminicola on wheat	10–12,000 years BP	95
	Magnaporthe oryzae on rice	7000 years BP	24
	Phytophthora infestans on potato	7000 years BP	34
	Ustilago maydis on maize	8000 years BP	72
Host jump/host shift			
	Magnaporthe oryzae from Setaria millet to rice	Abrupt evolutionary change, approx. 7000 years BP	24
	Rhynchosporium secalis from wild grasses to barley and rye	Abrupt evolutionary change, approx. 2,000 years BP	111
	Phytophthora infestans from wild *Solanum* species to potato	Abrupt evolutionary change, <500 years BP	35, 39

Stukenbrock and McDonald (Annu. Rev. Phyto., 2008)

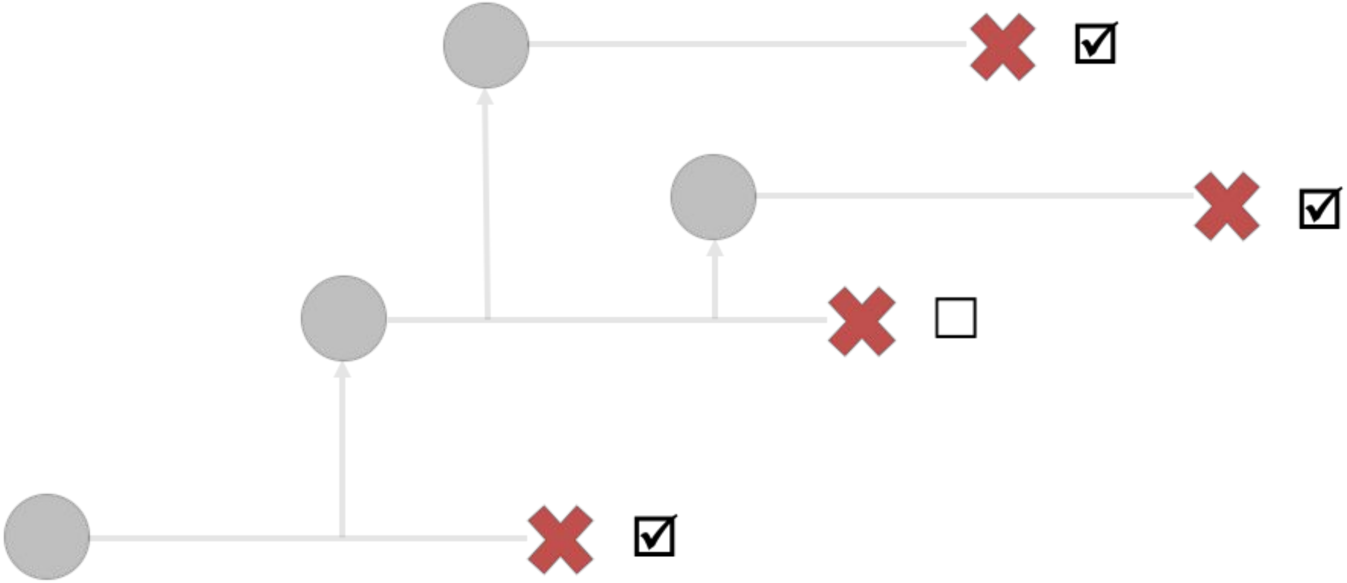
“Neolithic pathogen relocation camps”


James C. Scott
AUTHOR OF SEEING LIKE A STATE
Against the Grain
A DEEP HISTORY OF THE EARLIEST STATES

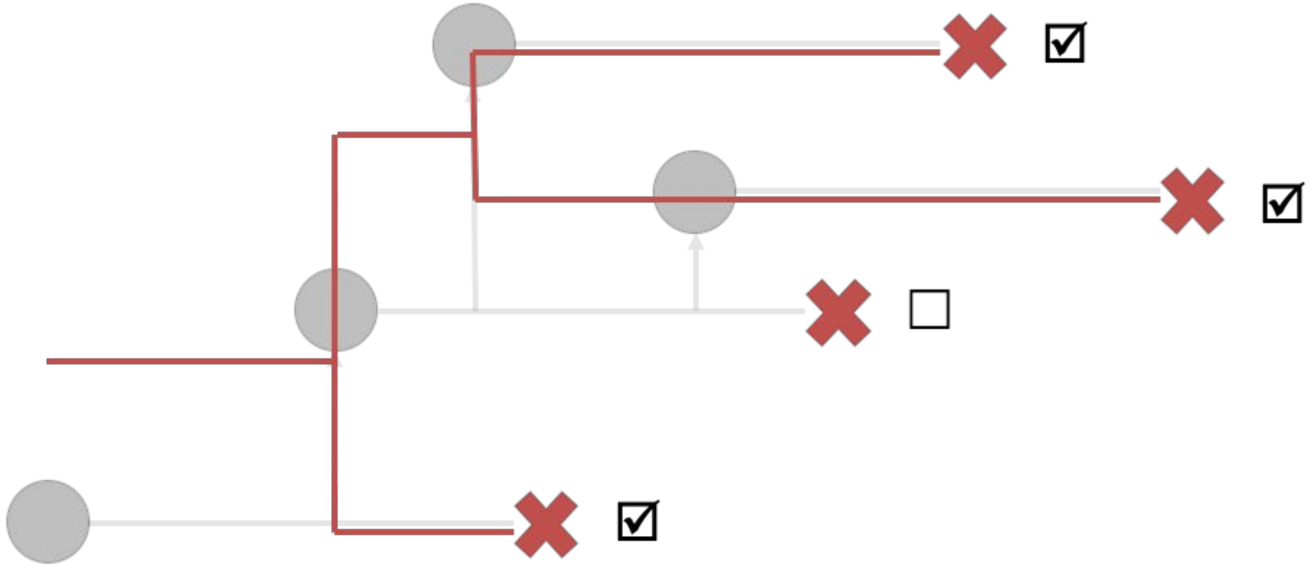
Phylogenies can tell us about:

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

A simple epidemic example with incomplete sampling



A simple epidemic example with incomplete sampling



We only observe transmission events as branching events if we sample both the parent and child lineage descending from the transmission event

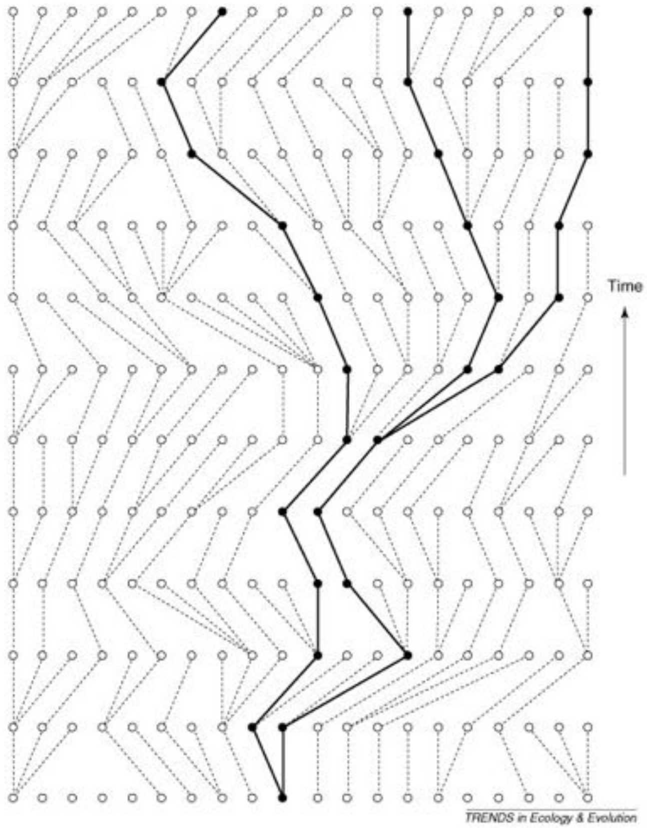
This brings us to phylodynamic modeling

Phylodynamic modeling in a nutshell

Phylogenies will only contain sampled lineages.

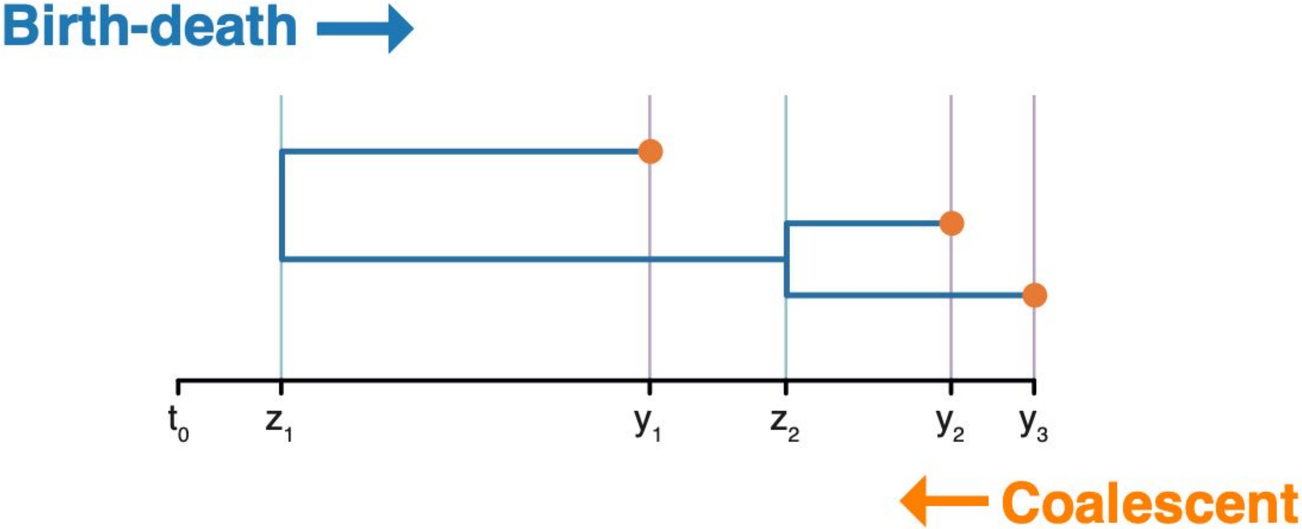
The sampled lineages are embedded within the full ancestral history of the population.

We need a statistical model that allows us to infer the most likely population history from the sampled phylogeny.



Kuhner *et al.* (2008)

Two types of phylodynamic models



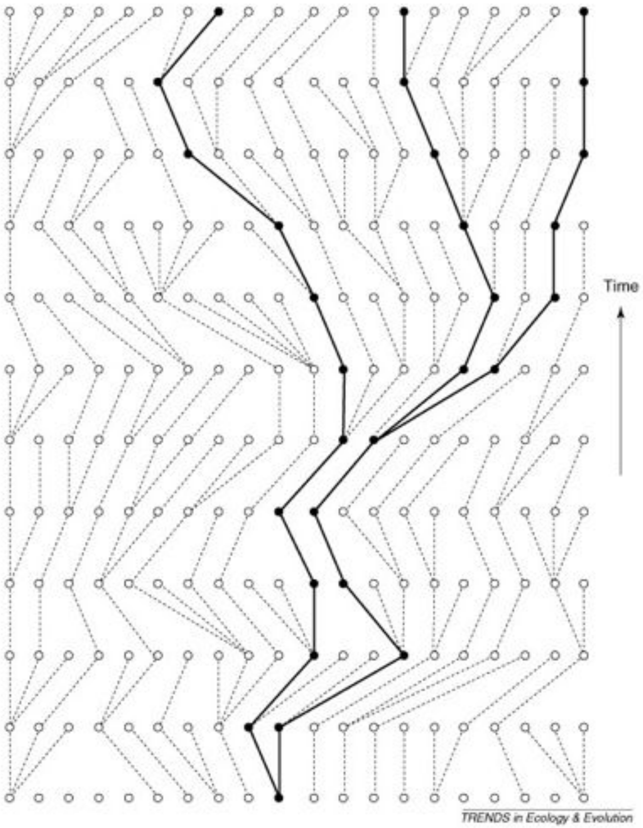
Coalescent theory

The coalescent traces the ancestry of sampled individuals back in time.

Allows us to relate events observed in the tree to the larger history of a population

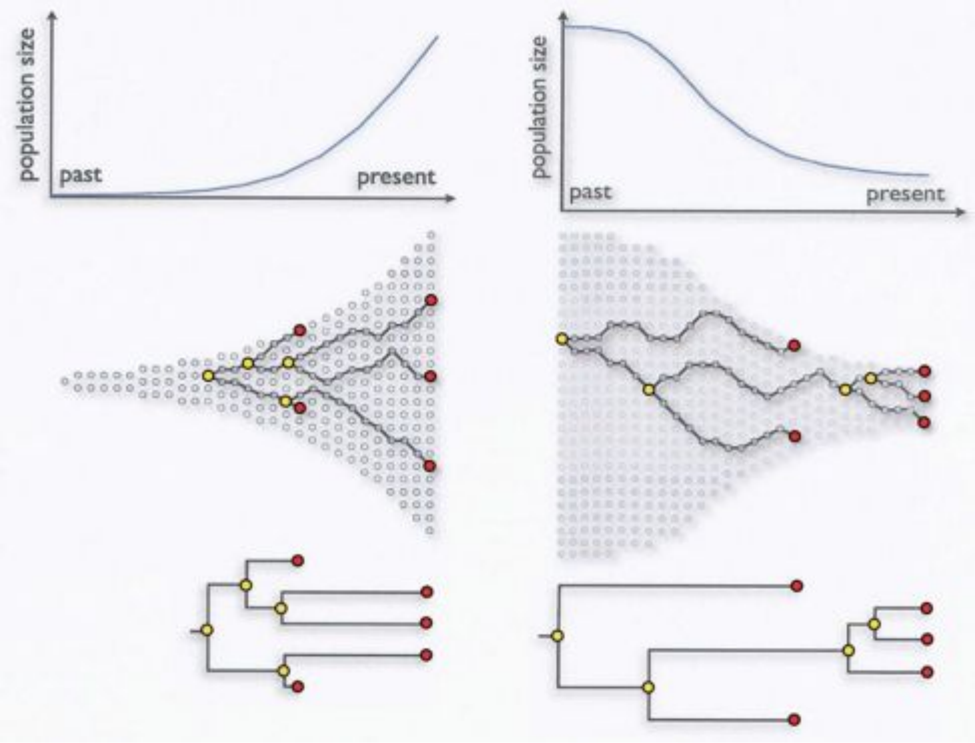
Probability of two lineages coalescing per generation is:

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

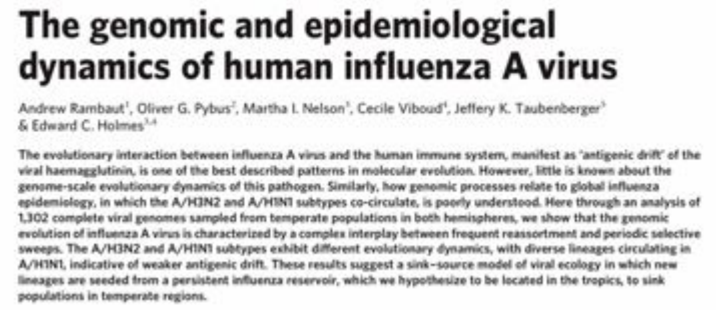


Kuhner *et al.* (2008)

Reconstructing population dynamics



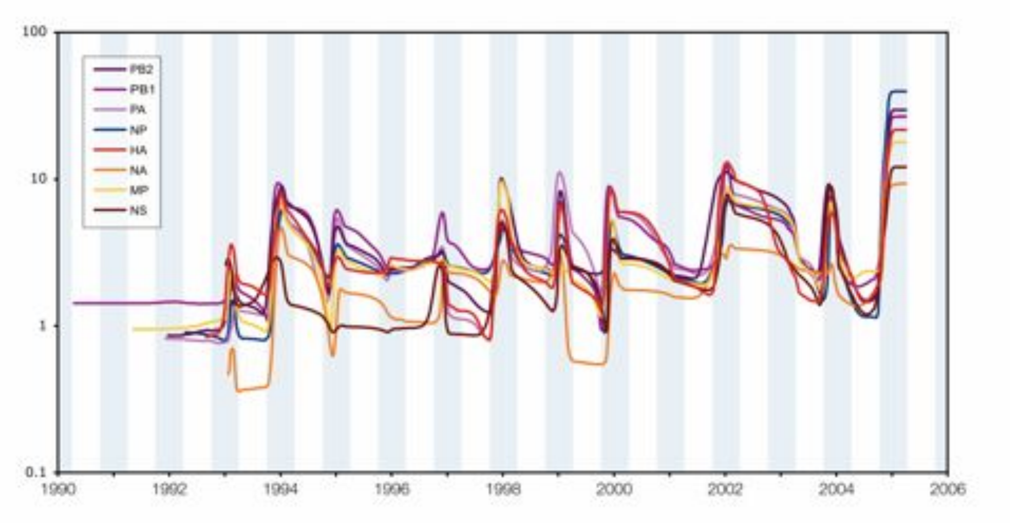
Reconstructing dynamics: influenza A



The genomic and epidemiological dynamics of human influenza A virus

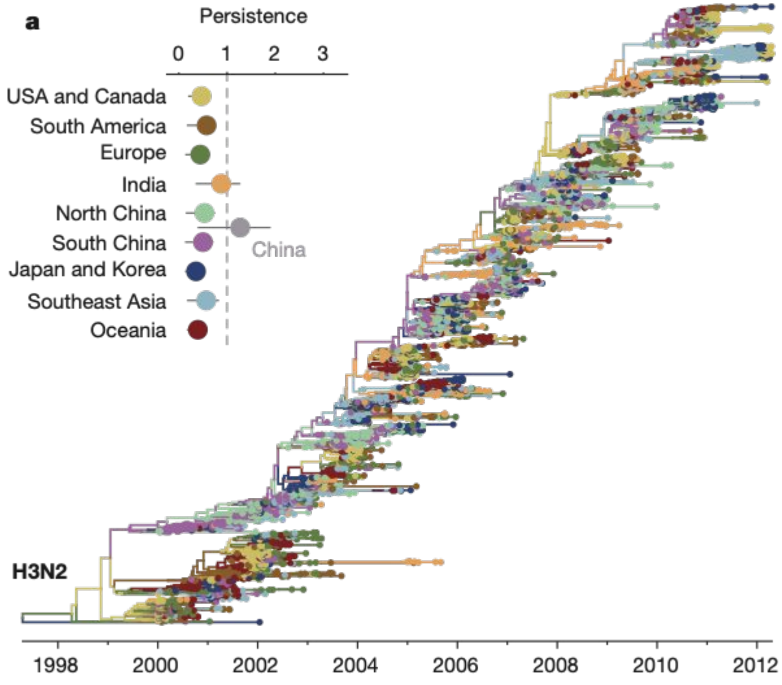
Andrew Rambaut[1], Oliver G. Pybus[2], Martha I. Nelson[3], Cecile Viboud[4], Jeffery K. Taubenberger[5]
& Edward C. Holmes[3,4]

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.

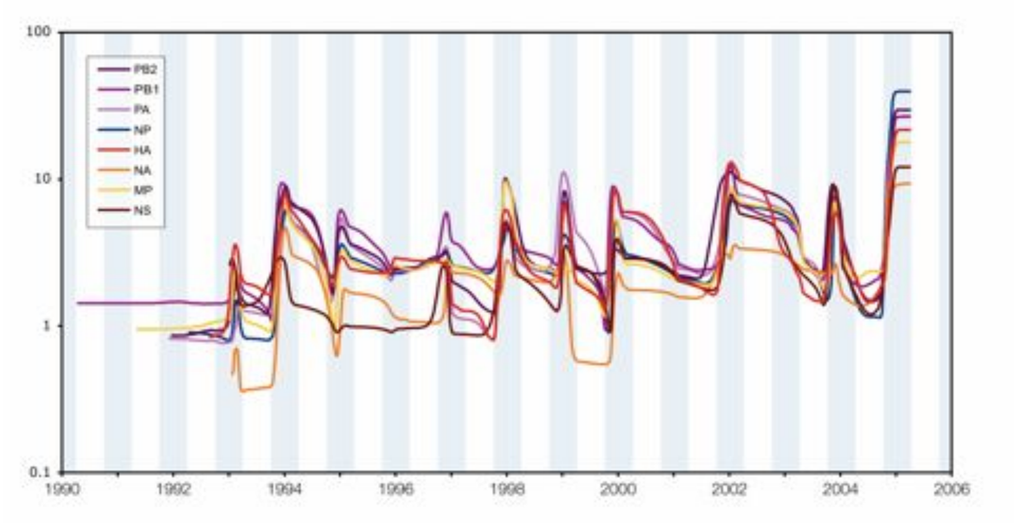


Rambaut *et al.* (2008)

Reconstructing dynamics: influenza A



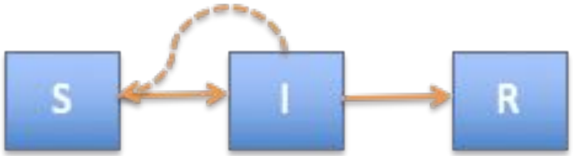
Bedford *et al.* (Nature, 2015)

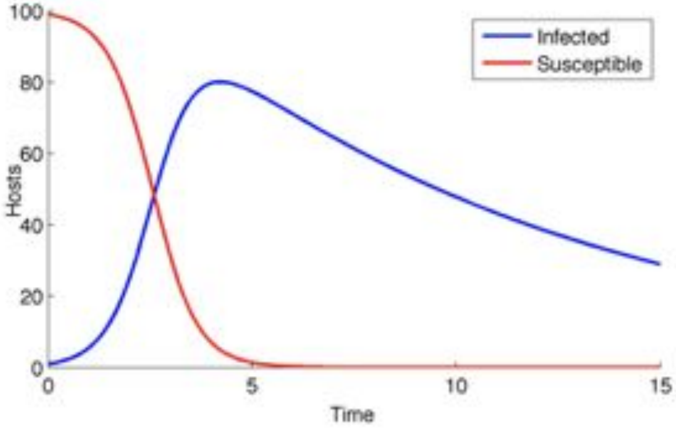


Rambaut *et al.* (2008)

Coupling epidemiological models to trees

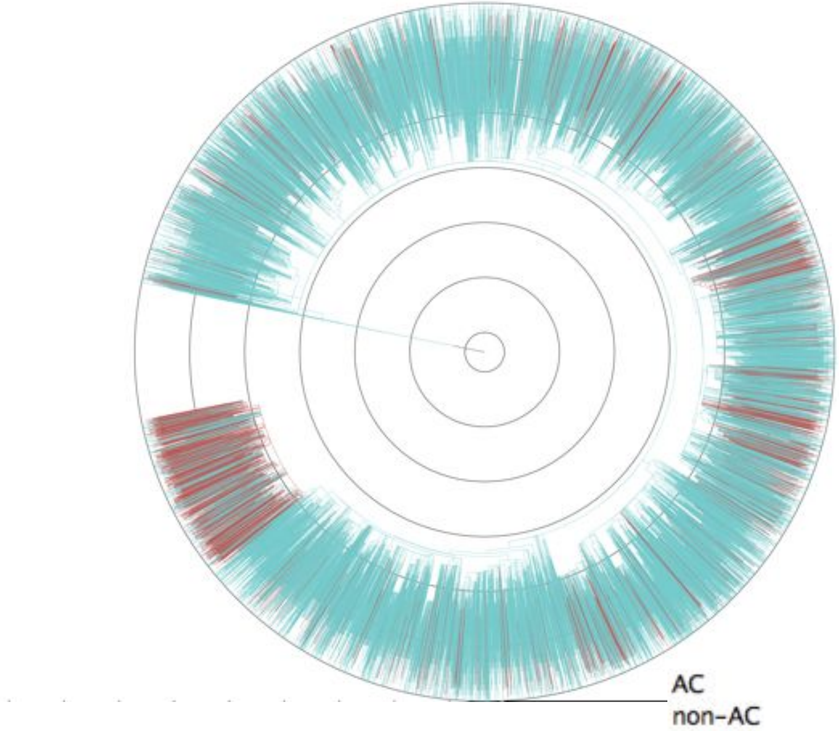
We can use phylodynamic modeling to couple phylogenetic methods with more traditional epidemiological models

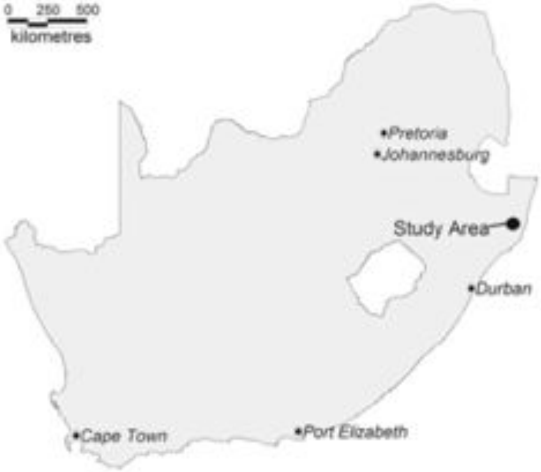




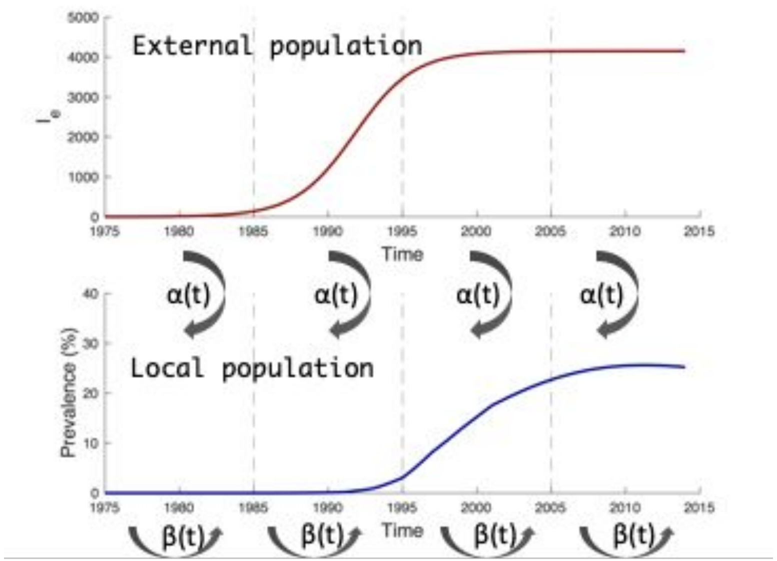
We can formulate epidemic models that we can then fit to phylogenies to estimate parameters of interest.

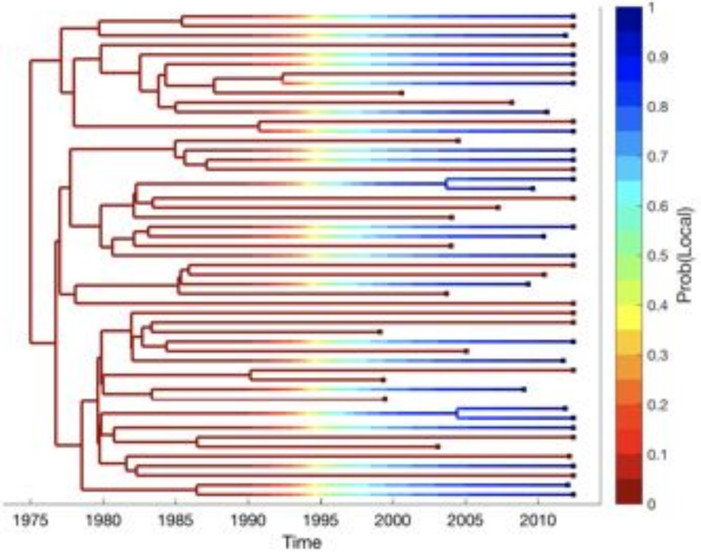
HIV in rural Kwa-Zulu Natal



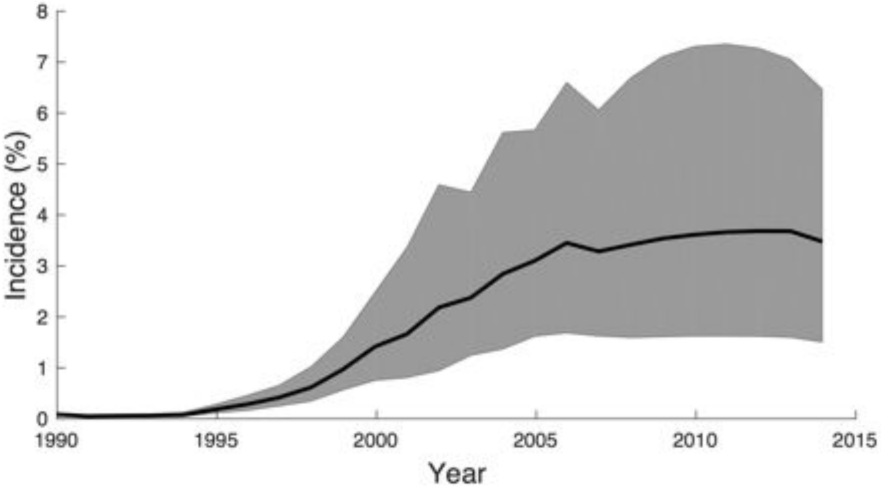


A simple two-patch SIR model for HIV





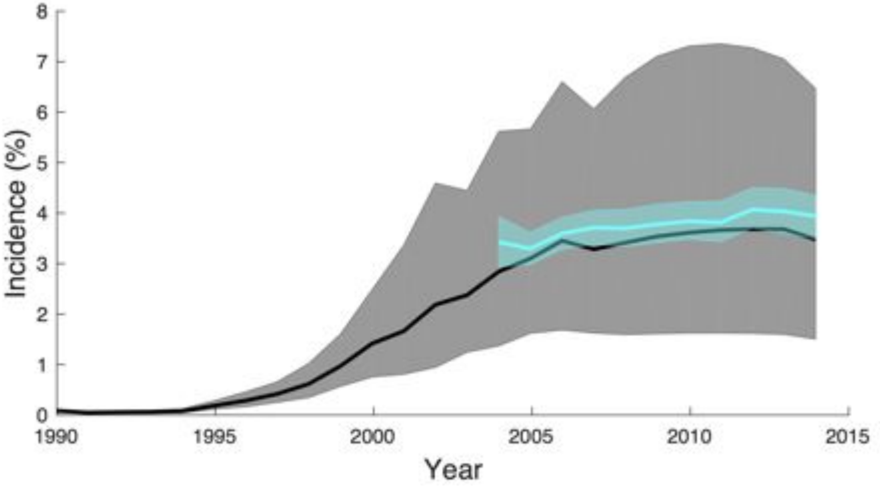
Phylodynamic estimates of HIV incidence



Rasmussen et al. (Virus Evolution, 2018)

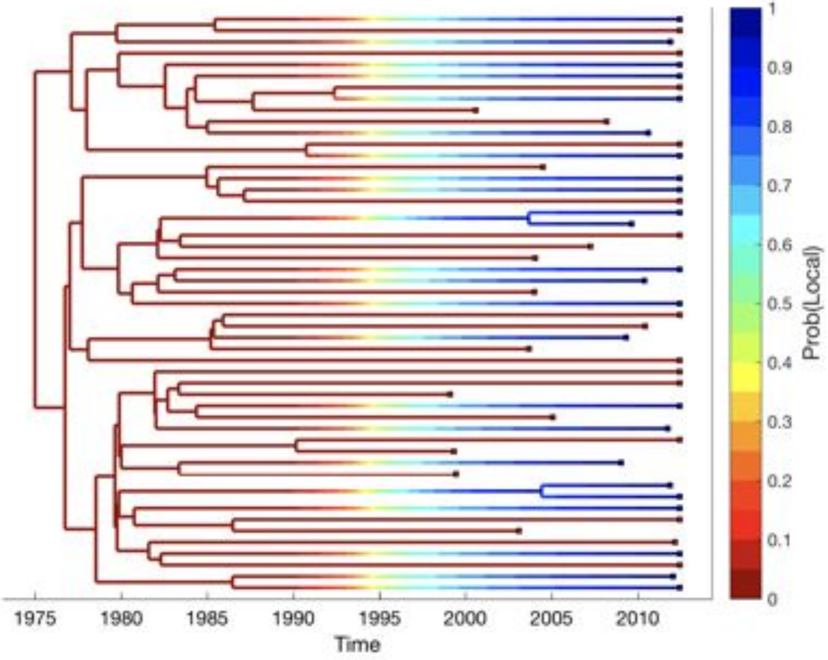
Phylodynamic estimates of HIV incidence

Inferred incidence of 3-4% per year almost perfectly coincides with population-based surveillance data.

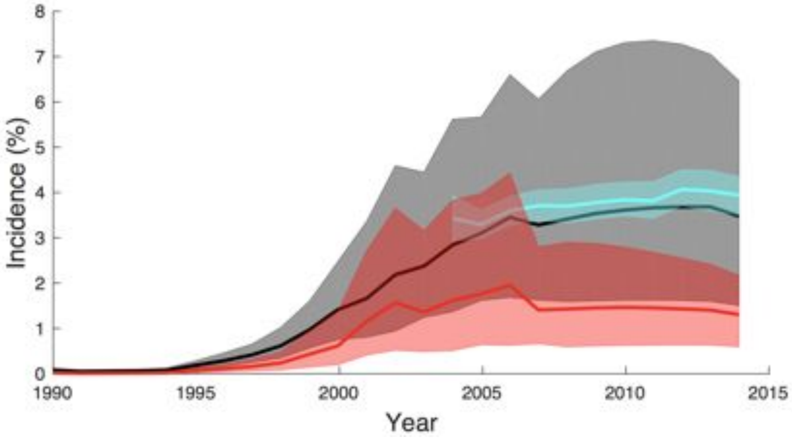


Rasmussen et al. (Virus Evolution, 2018)

Tracking lineage movement

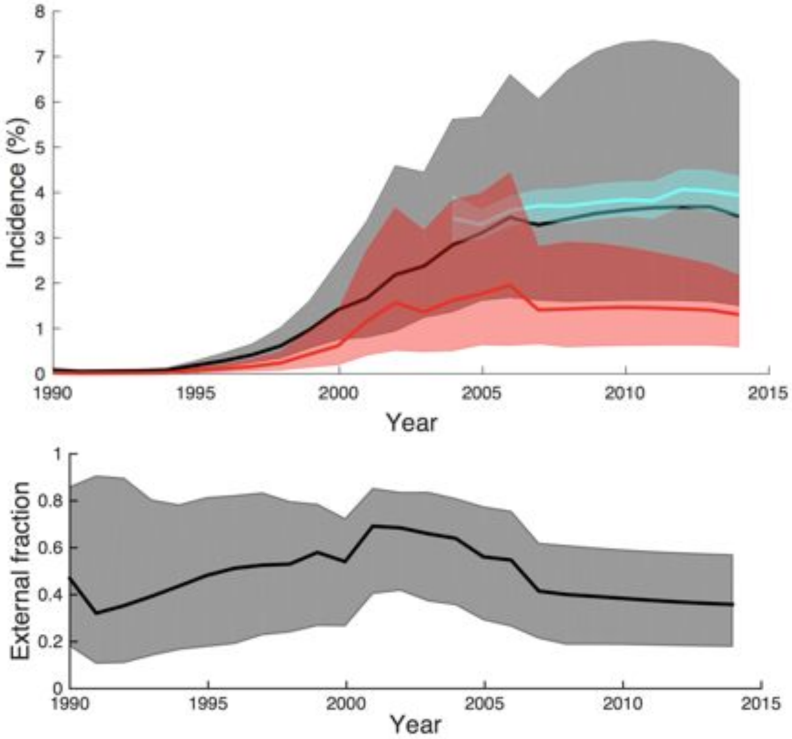


Incidence due to external introductions



Rasmussen et al. (Virus Evolution, 2018)

Incidence due to external introductions



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

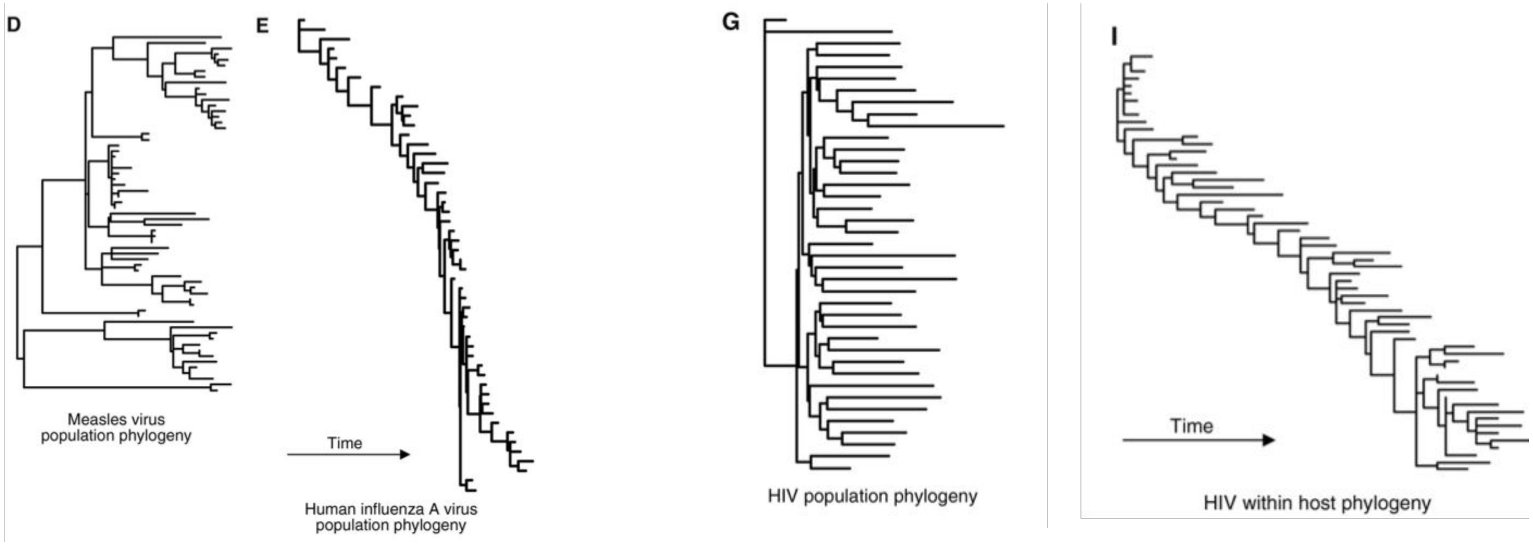
Rasmussen et al. (Virus Evolution, 2018)

Phylogenies can tell us about:

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

Phylodynamics with selection

Selection for higher fitness strains strongly shapes the phylogenetic history of many different pathogens.

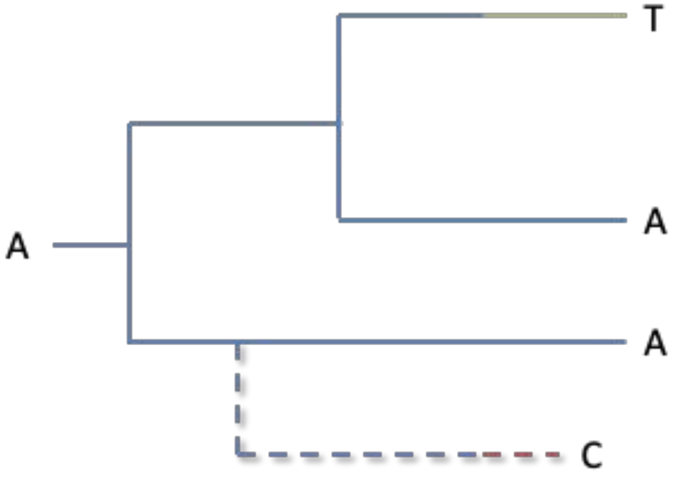

D
Measles virus population phylogeny
E
Time
Human influenza A virus population phylogeny
G
HIV population phylogeny
I
Time
HIV within host phylogeny

Grenfell *et al.* (Science, 2004)

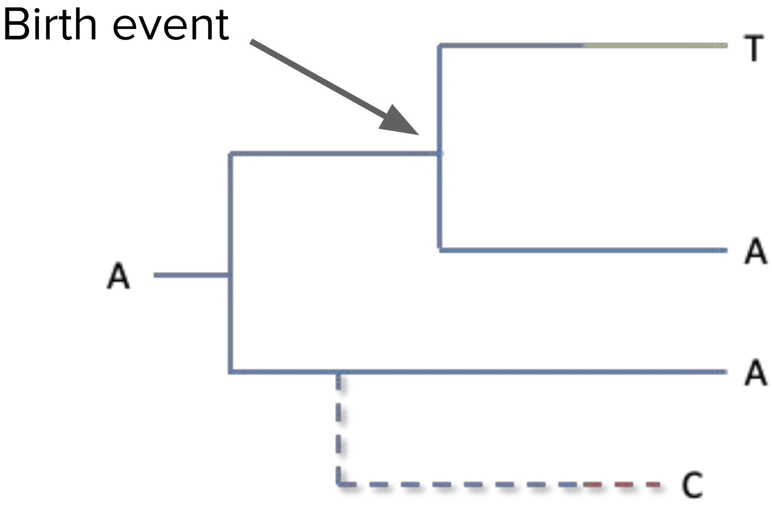
We therefore need *phylodynamic* models that allow selection to shape trees

Multi-type birth-death models

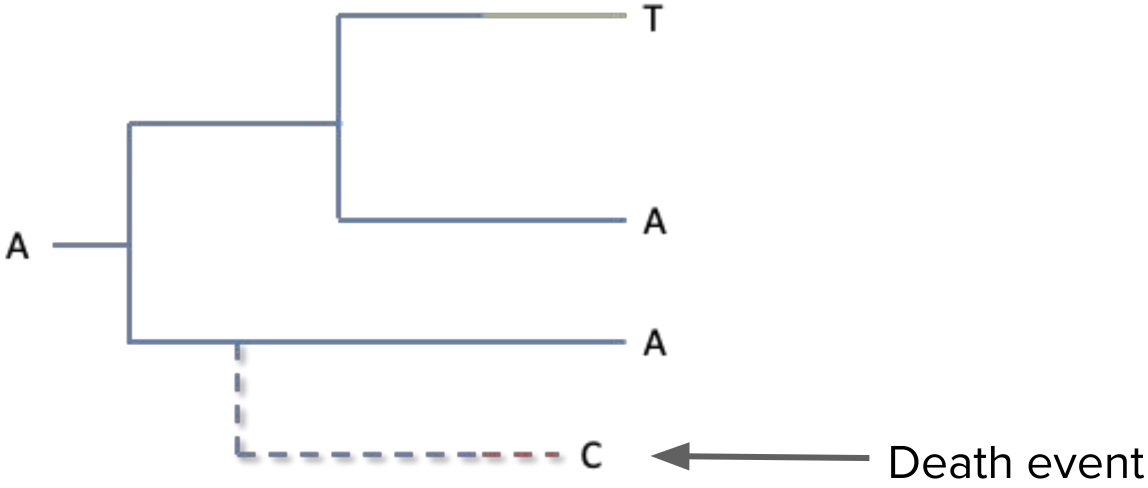
Provide one way of incorporating adaptive (non-neutral) evolution into phylogenetic models.



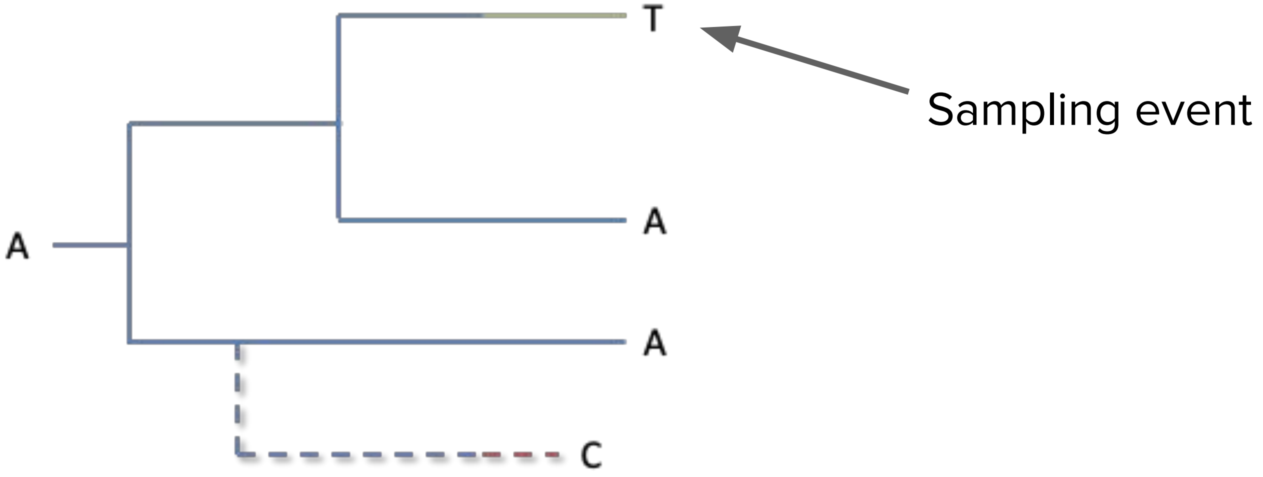
Multi-type birth-death models



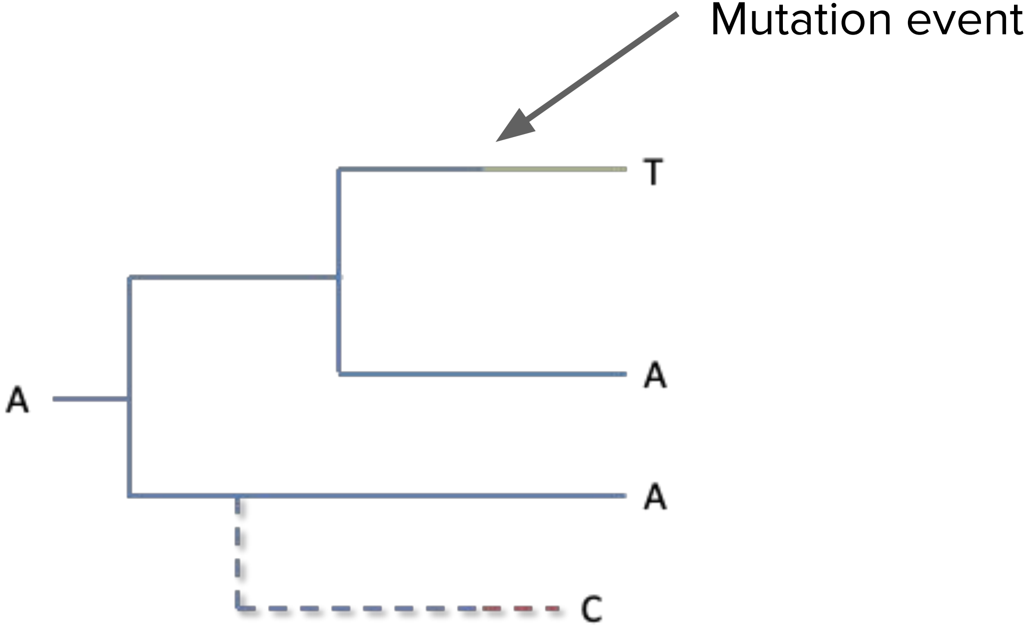
Multi-type birth-death models



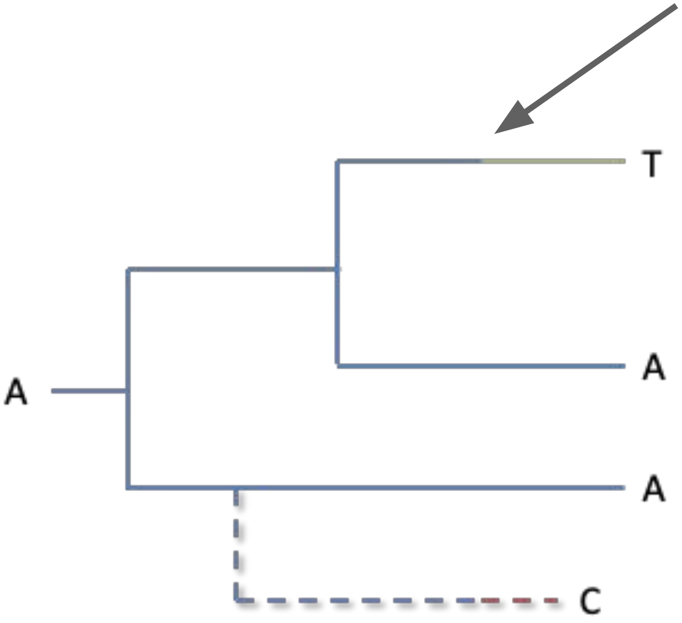
Multi-type birth-death models



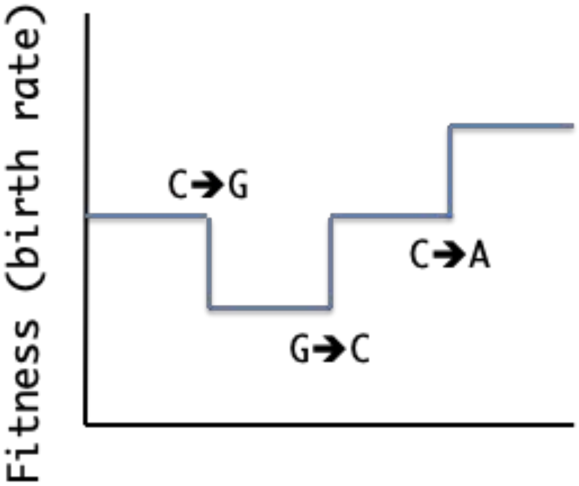
Multi-type birth-death models



Multi-type birth-death models

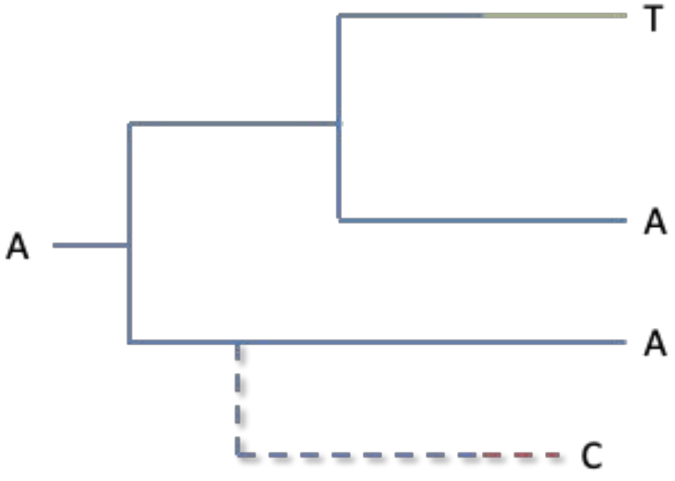


Mutation event



Multi-type birth-death models

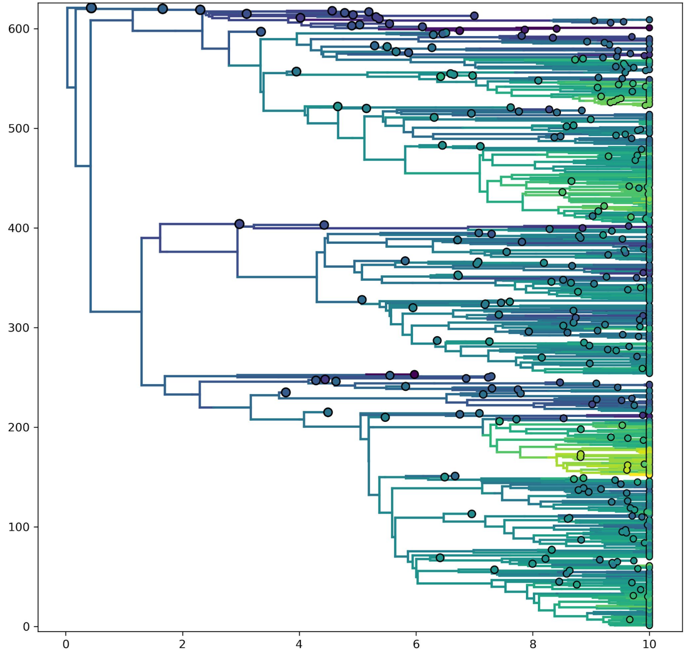
MTBD models allow us to compute the **joint likelihood** that both the tree and the observed tip genotypes evolved exactly as observed (Stadler and Bonhoeffer, 2013).



Fitness shapes trees

More fit lineages will be transmitted (branch) more often and leave behind more sampled descendants than less fit lineages.

Estimating transmission rates from the branching structure of phylogenies using MTBD provides us with one way to directly estimate pathogen fitness from genomic data.



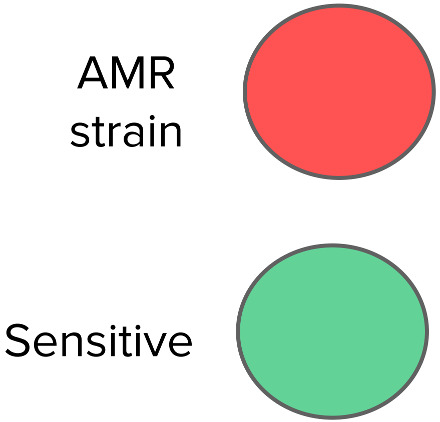
Warmer colors = More fit

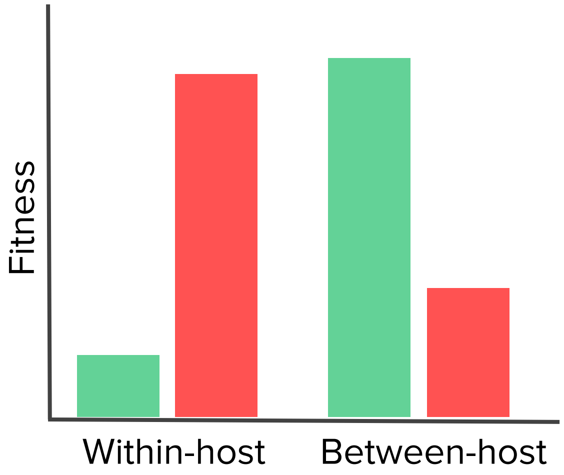
Fitness effects of antimicrobial resistance

We will consider fitness differences between drug-sensitive and antimicrobial resistant (AMR) strains of a pathogen.

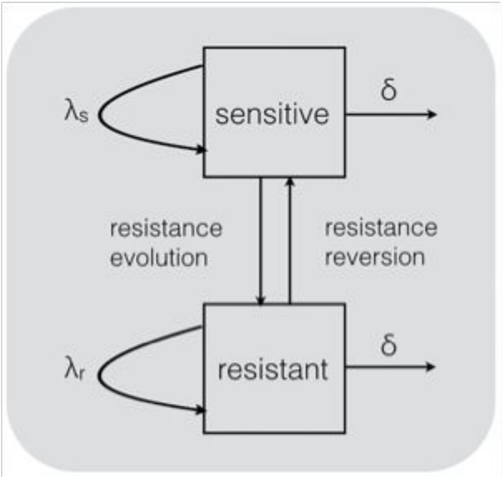
On Drugs

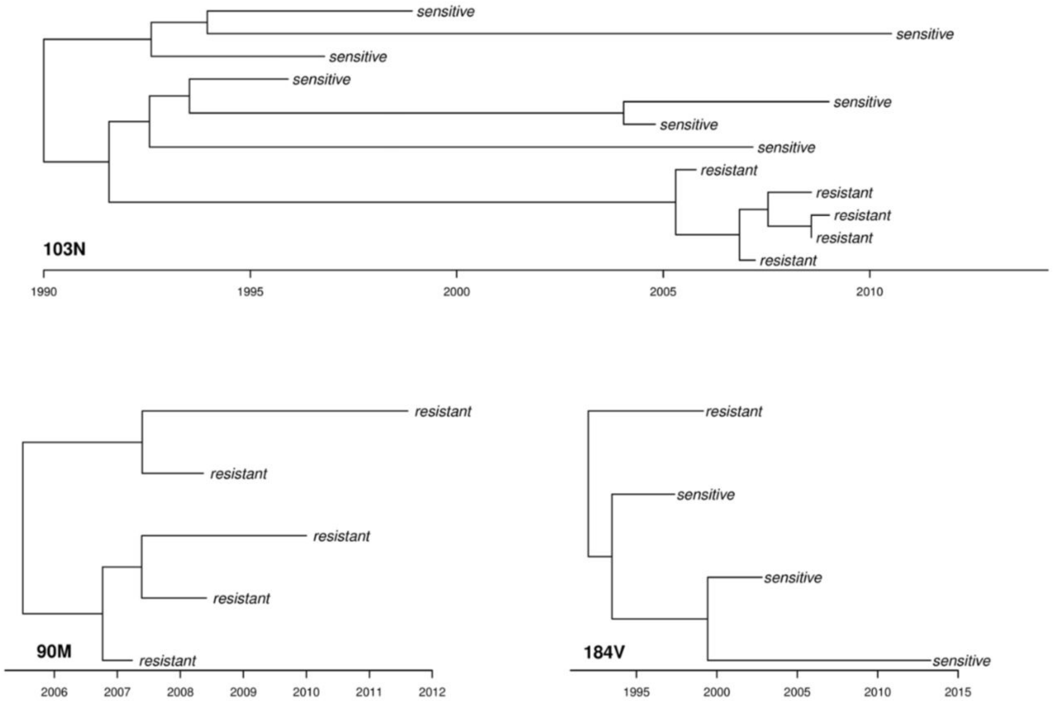
No Drugs





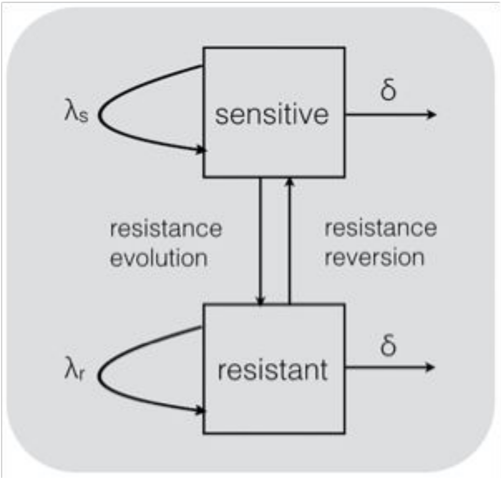
Fitness of HIV drug resistance mutations

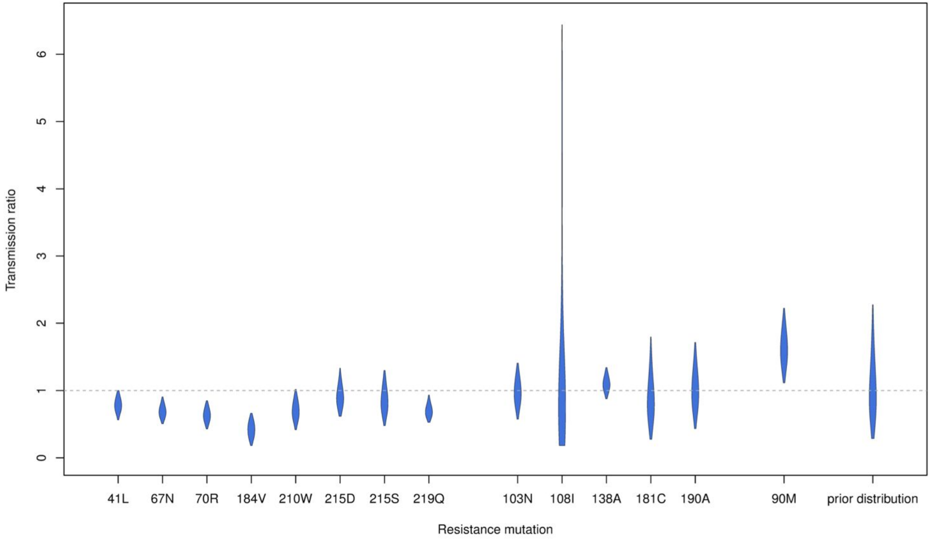




Kühnert *et al.* (PLoS Pathogens, 2018)

Fitness of HIV drug resistance mutations





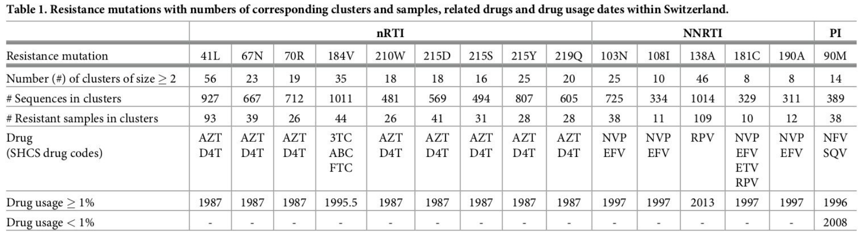


Table 1. Resistance mutations with numbers of corresponding clusters and samples, related drugs and drug usage dates within Switzerland.

	nRTI									NNRTI					PI
Resistance mutation	41L	67N	70R	184V	210W	215D	215S	215Y	219Q	103N	108I	138A	181C	190A	90M
Number (#) of clusters of size ≥ 2	56	23	19	35	18	18	16	25	20	25	10	46	8	8	14
# Sequences in clusters	927	667	712	1011	481	569	494	807	605	725	334	1014	329	311	389
# Resistant samples in clusters	93	39	26	44	26	41	31	28	28	38	11	109	10	12	38
Drug (SHCS drug codes)	AZT D4T	AZT D4T	AZT D4T	3TC ABC FTC	AZT D4T	AZT D4T	AZT D4T	AZT D4T	AZT D4T	NVP EFV	NVP EFV	RPV	NVP EFV ETV RPV	NVP EFV	NFV SQV
Drug usage ≥ 1%	1987	1987	1987	1995.5	1987	1987	1987	1987	1987	1997	1997	2013	1997	1997	1996
Drug usage < 1%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2008

Kühnert *et al.* (PLoS Pathogens, 2018)

Phylogenies can tell us about:

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

What do you want to learn from this class?

For Wednesday

On Wednesday we’ll start with a tutorial that should help us ease into working with sequence data and trees.

Please have your laptops ready!

Try to install RAxML ahead of time

If you’re interested in doing the Python exercises, install Python (with Anaconda) and Biopython.