Multi-type birth-death models and adaptive molecular evolution

Molecular Epidemiology of Infectious Diseases
Lecture 7

March 2nd, 2020

Adaptive pathogen evolution

So far we have considered neutral evolutionary models where all pathogen genotypes have the same fitness — mutations do not impact fitness.

Now we will consider fitness variation in pathogen populations where different genotypes may have different fitness values.

In this case, selection can act on fitness differences between strains, allowing pathogen populations to adapt to their environment.

What do we mean by fitness?

For our purposes, we can define a pathogen's fitness in terms of its growth rate *r*:

r = birth rate - death rate

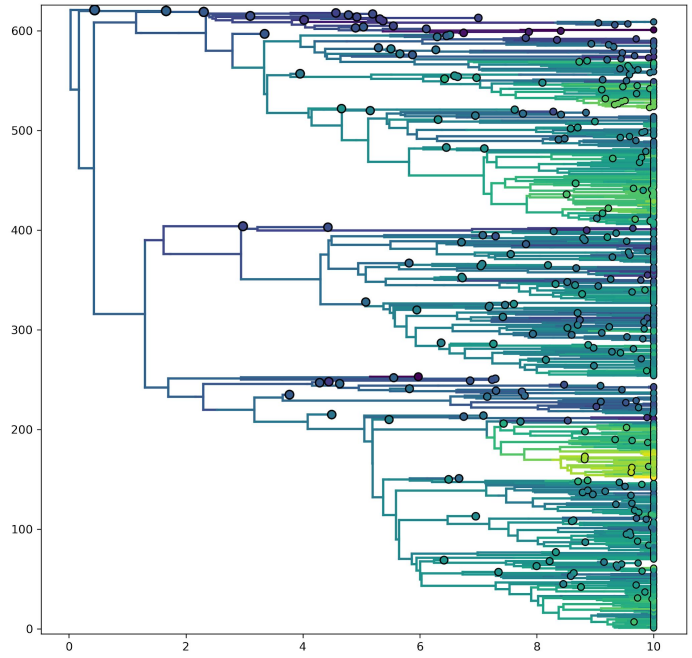
Within hosts, the birth rate can be thought of as the pathogen's replication rate.

Between hosts, the birth rate can be thought of as a transmission rate and the death rate as the rate infected hosts are removed. The growth rate *r* therefore quantifies a pathogen's epidemic potential at the host-population level.

Selection shapes pathogen phylogenies

More fit lineages will have higher growth rates and therefore branch more often... leaving behind more sampled descendents in a phylogeny.

branching = birth/transmission events



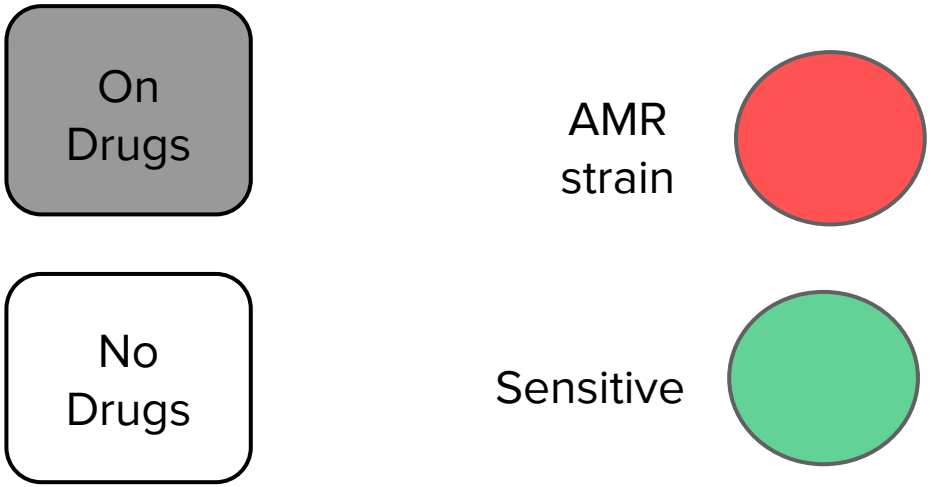
Selection shapes pathogen phylogenies

Selection for better adapted strains strongly shapes the phylogenetic history of many different pathogens.

Grenfell *et al.* (Science, 2004)

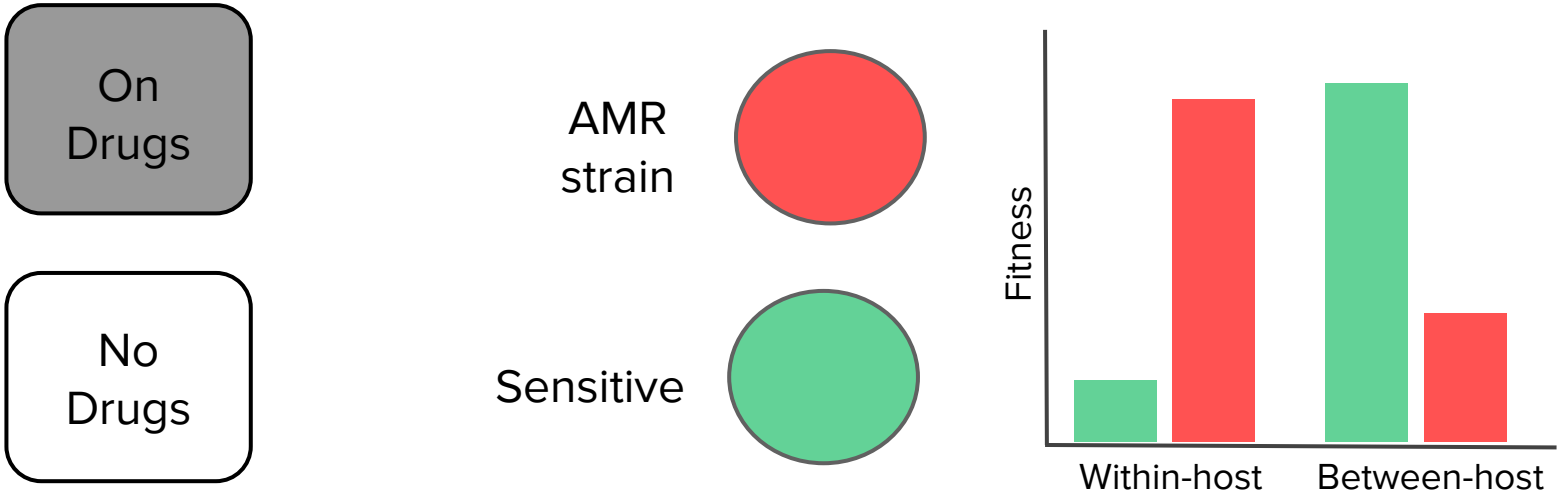
Motivation: antimicrobial resistance

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



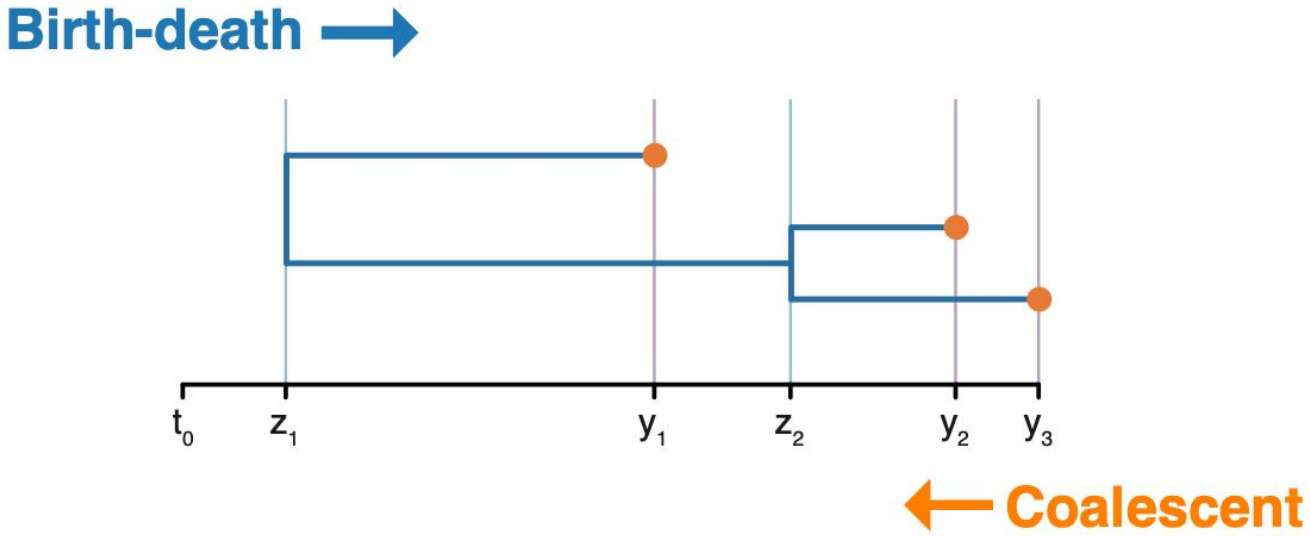
Motivation: antimicrobial resistance

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



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

Two types of phylodynamic models



Birth-death models

Population dynamics are viewed forward in time starting at some point in the past.

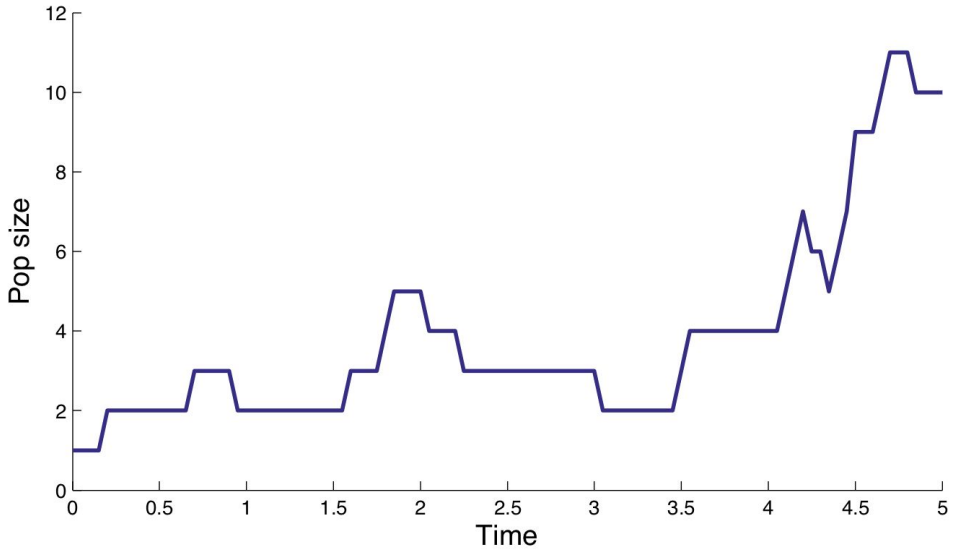
In the most basic model individuals reproduce and die. Nothing else happens.

This is a stochastic process: the number and timing of birth and death events are viewed as random variables.

Mathematically: a continuous-time Markov process on the space of positive integers.

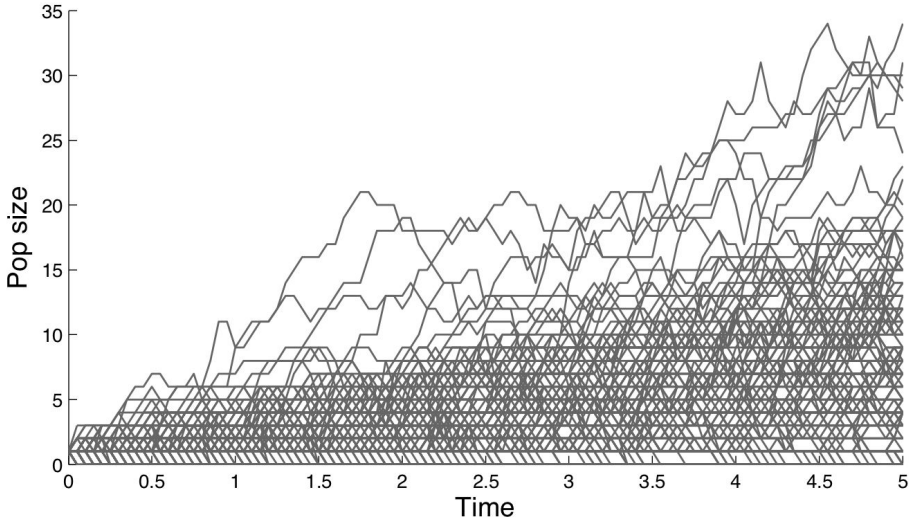
The birth-death process

A single stochastic realization of the birth-death process:



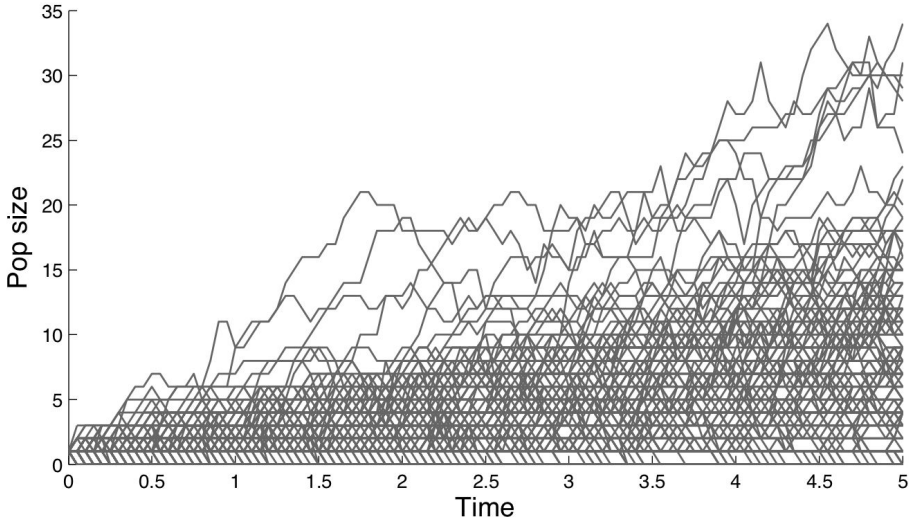
The birth-death process

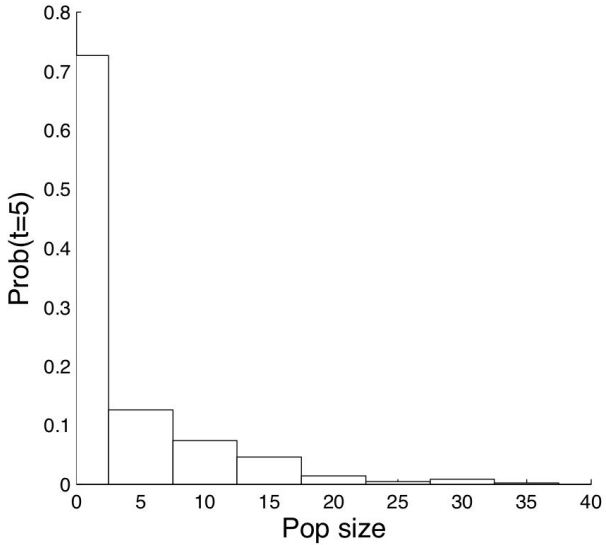
An ensemble of stochastic realizations



The birth-death process

Let's consider the probability $p_i(t)$ that i individuals are alive at time t





The birth-death process

We can analytically compute $p_i(t)$

The ***transition probabilities*** over a small interval of time Δt

$\lambda = \text{birth rate}$

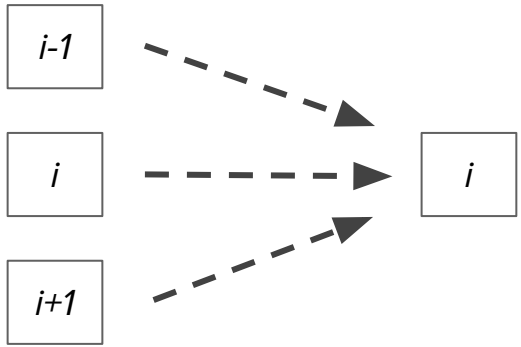
$\mu = \text{death rate}$

$$
\begin{aligned}
p_{i \to i-1}(\Delta t) &= \mu i \Delta t + o(\Delta t) \\
p_{i \to i+1}(\Delta t) &= \lambda i \Delta t + o(\Delta t) \\
p_{i \to i}(\Delta t) &= 1 - (\lambda + \mu) i \Delta t + o(\Delta t)
\end{aligned}
$$

The birth-death process

Given $p_i(t)$, we can compute $p_i(t+\Delta t)$:

$$p_i(t+\Delta t) = \lambda(i-1)p_{i-1}(t)\Delta t + \mu(i+1)p_{i+1}(t)\Delta t - (\lambda+\mu)ip_i(t)\Delta t$$

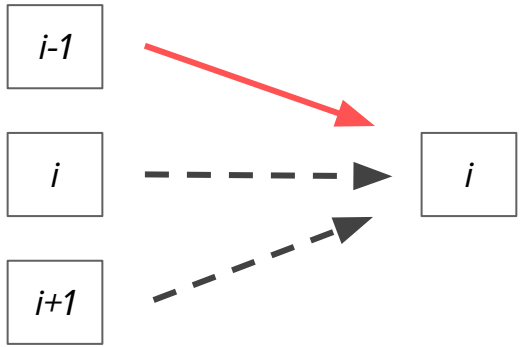


The birth-death process

Given $p_i(t)$, we can compute $p_i(t+\Delta t)$:

$$p_i(t+\Delta t) = \lambda(i-1)p_{i-1}(t)\Delta t + \mu(i+1)p_{i+1}(t)\Delta t - (\lambda+\mu)ip_i(t)\Delta t$$

Birth

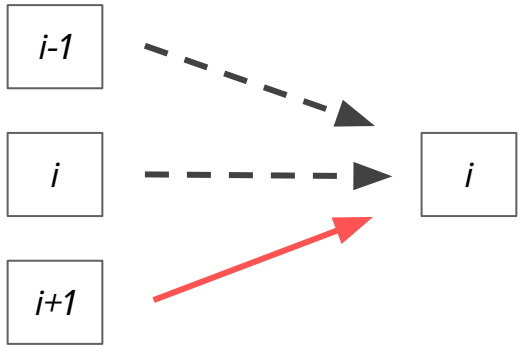


The birth-death process

Given $p_i(t)$, we can compute $p_i(t+\Delta t)$:

$$p_i(t+\Delta t) = \lambda(i-1)p_{i-1}(t)\Delta t + \mu(i+1)p_{i+1}(t)\Delta t - (\lambda+\mu)ip_i(t)\Delta t$$

Death

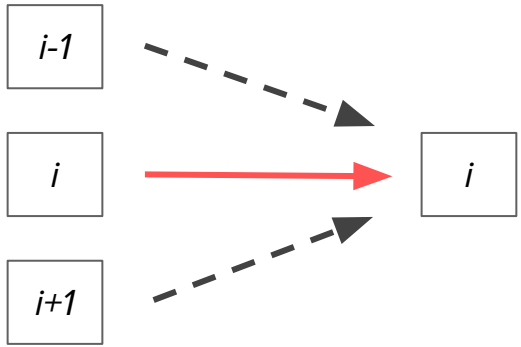


The birth-death process

Given $p_i(t)$, we can compute $p_i(t+\Delta t)$:

$$p_i(t+\Delta t) = \lambda(i-1)p_{i-1}(t)\Delta t + \mu(i+1)p_{i+1}(t)\Delta t - (\lambda+\mu)ip_i(t)\Delta t$$

No event



The birth-death process

Given $p_i(t)$, we can compute $p_i(t+\Delta t)$:

$$p_i(t + \Delta t) = \lambda(i-1)p_{i-1}(t)\Delta t + \mu(i+1)p_{i+1}(t)\Delta t - (\lambda + \mu)ip_i(t)\Delta t$$

Letting Δt go to zero:

$$\frac{dp_i(t)}{dt} = \lambda(i-1)p_{i-1}(t) + \mu(i+1)p_{i+1}(t) - (\lambda + \mu)ip_i(t)$$

Special cases

The probability that there are no living individuals after time t:

$$p_0(t) = \frac{\mu - \mu e^{(\mu-\lambda)t}}{\lambda - \mu e^{(\mu-\lambda)t}}$$

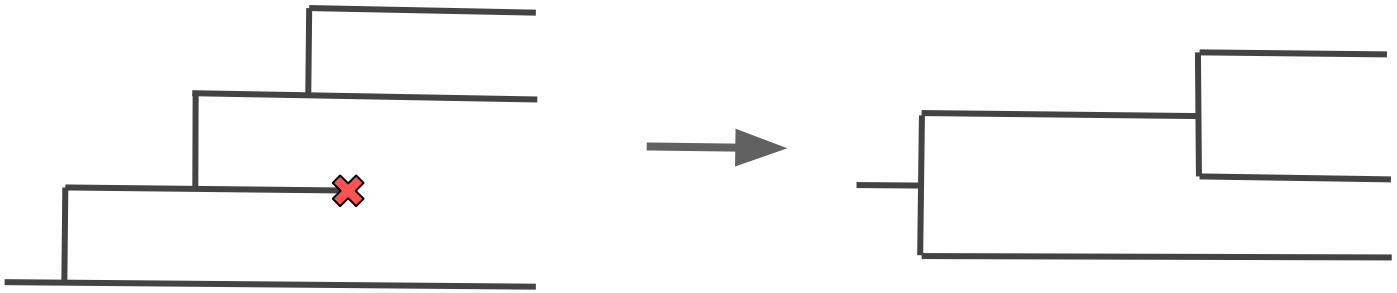
The probability that there is exactly one living individual at time t:

$$p_1(t) = \frac{(\lambda - \mu)^2 e^{-(\lambda-\mu)t}}{(\lambda - \mu e^{(\mu-\lambda)t})^2}$$

Why is this useful?

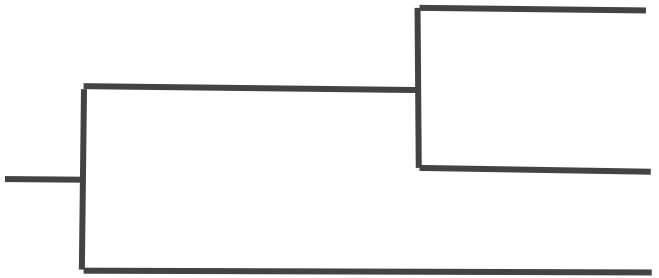
Birth-death processes as trees

The birth-death process can also be thought of as branching process that generates a tree-like structure



BD tree likelihood: complete sampling

$$L(T|\lambda,\mu) \propto \lambda^{n-1} \prod_{i=1}^{n-1} p_1(t_i)$$



λ = birth rate

μ = death rate

Thompson (1975); Rannala & Yang (1996)

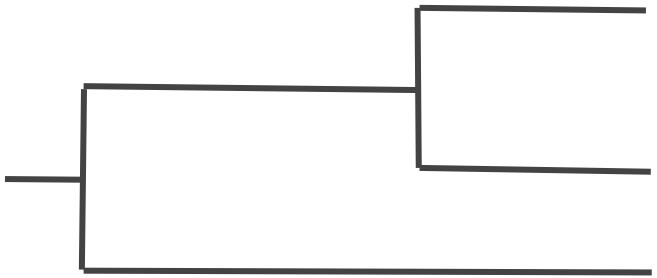
BD tree likelihood: complete sampling

$$L(T|\lambda,\mu) \propto \lambda^{n-1} \prod_{i=1}^{n-1} p_1(t_i)$$

$$p_1(t) = \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{(\mu-\lambda)t})^2}$$

λ = birth rate

μ = death rate



Thompson (1975); Rannala & Yang (1996)

BD tree likelihood: incomplete sampling

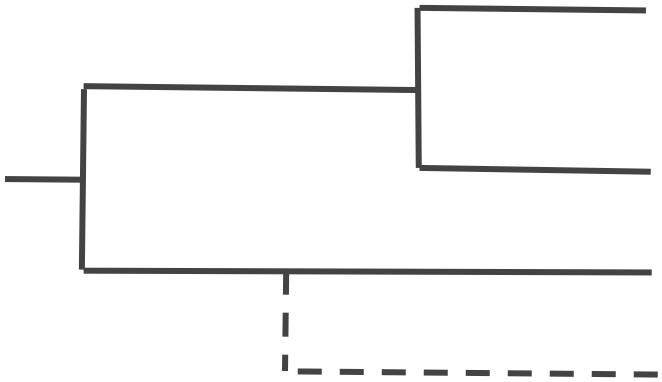
$$L(T|\lambda,\mu,\rho) \propto (\lambda\rho)^{n-1}\prod_{i=1}^{n-1} p_1(t_i)$$

$$p_1(t) = \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)t}}{((\rho\lambda+\lambda(1-\rho)-\mu)e^{(\mu-\lambda)t})^2}$$

λ = birth rate

μ = death rate

ρ = sampling fraction



Stadler (2009)

BD tree likelihoods

We can compute the likelihood of a phylogeny having evolved as observed given the parameters of our birth-death model, even in the presence of incomplete sampling!

This means we can directly estimate birth (transmission) and death (removal) rates as well as sampling proportions from phylogenies!

Parameters we can estimate from trees

The catch: we can only estimate two of the three parameters (λ, μ and ρ) in the model, i.e. an increased birth rate can always be compensated by a decreased sampling fraction and vice versa.

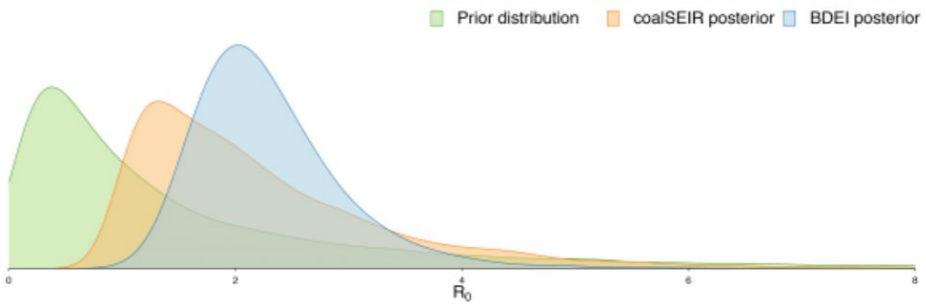
We can estimate both the birth and death rate if we know the sampling fraction.

For pathogens, we generally don't know the sampling fraction but have prior information about the removal rate μ, so we can estimate the transmission rate λ.

We can therefore estimate R_0 = (λ / μ) and the sampling fraction.

Estimating R_0 from pathogen phylogenies

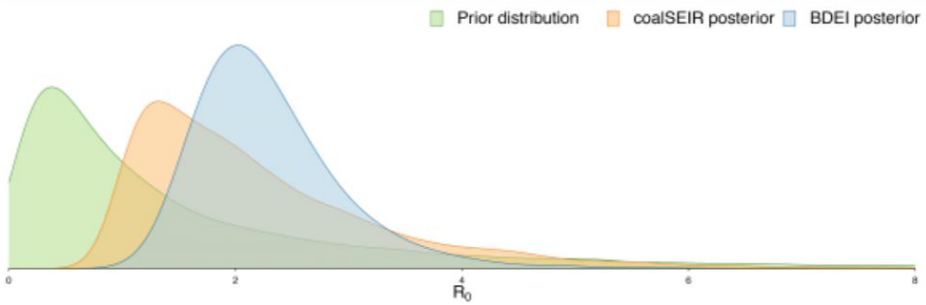
The key epidemiological parameter R_0, the number of secondary infections caused by an infected individual, can be estimated from pathogen phylogenies.



Stadler *et al.*, (PLoS Currents, 2014); WHO Ebola Response Team (NEJM, 2014)

Estimating R_0 from pathogen phylogenies

The key epidemiological parameter R_0, the number of secondary infections caused by an infected individual, can be estimated from pathogen phylogenies.



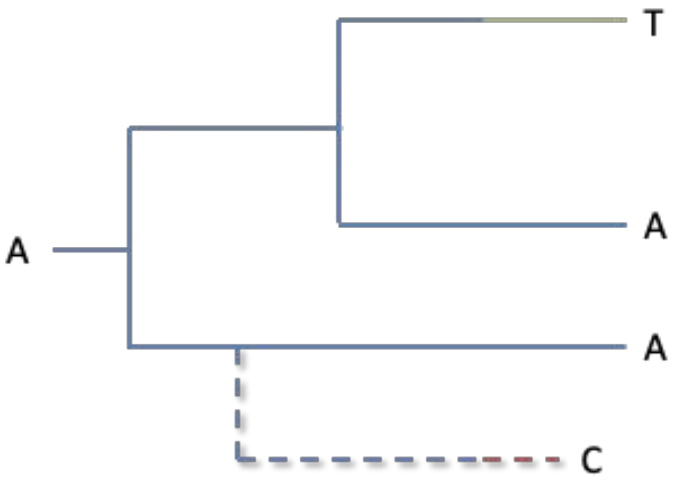
WHO Response Team Estimate:
R0 = 2.02 (95% CI: 1.79-2.26)

Stadler *et al.*, (PLoS Currents, 2014); WHO Ebola Response Team (NEJM, 2014)

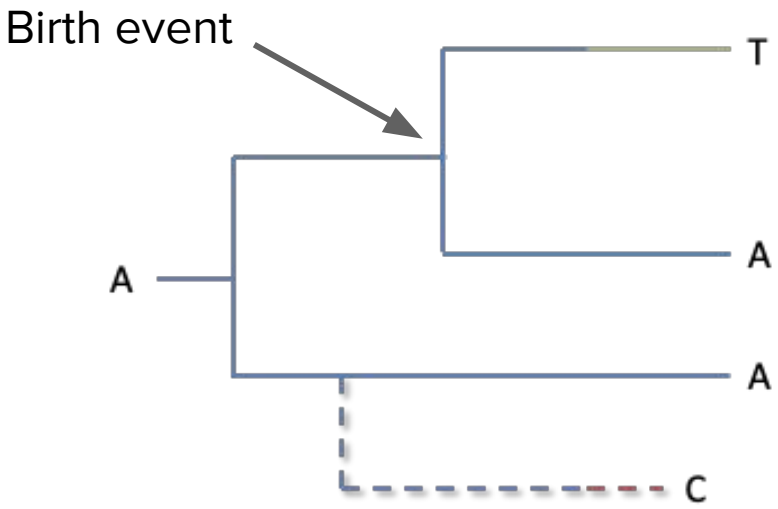
What if there is more than one type of pathogen?

Multi-type birth-death models

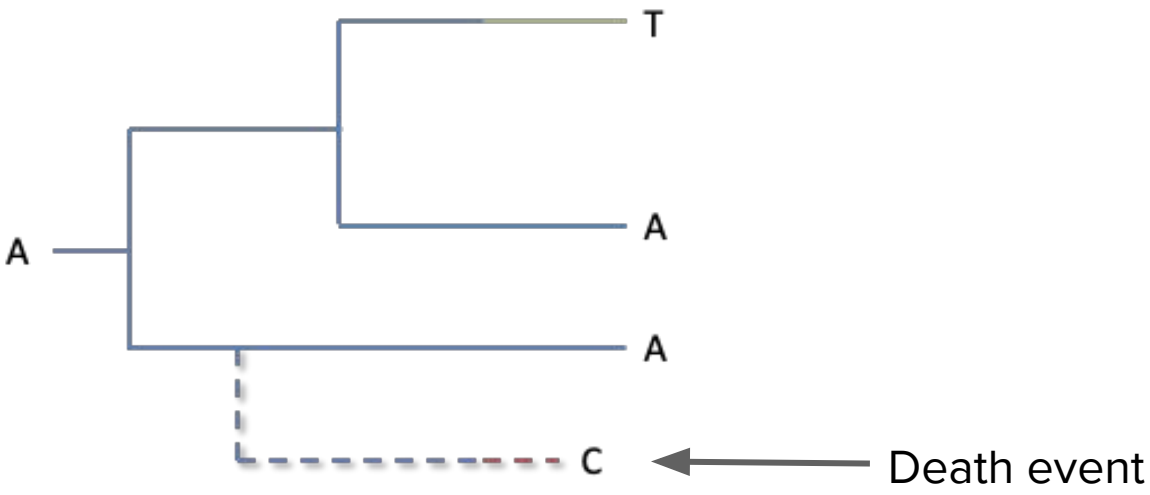
Allows for different types of individuals (e.g. genotypes) that can vary in their birth or death rates and therefore their fitness values.



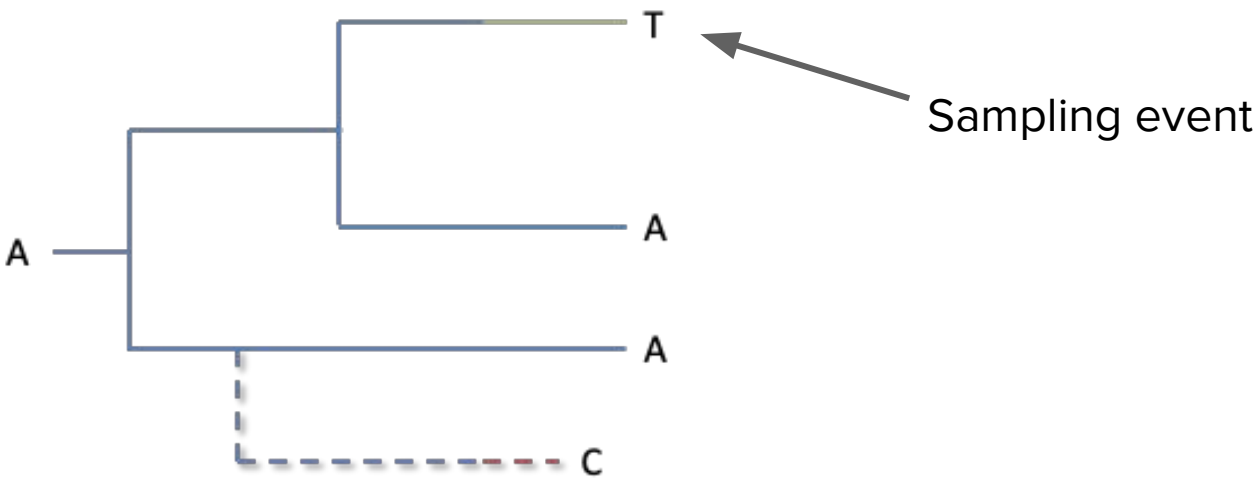
Multi-type birth-death models



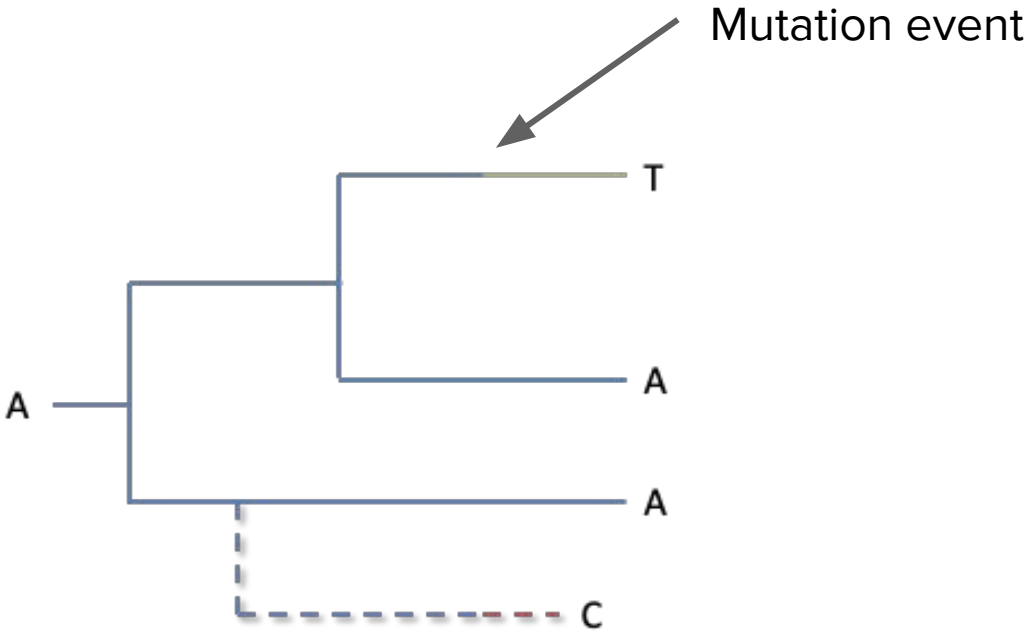
Multi-type birth-death models



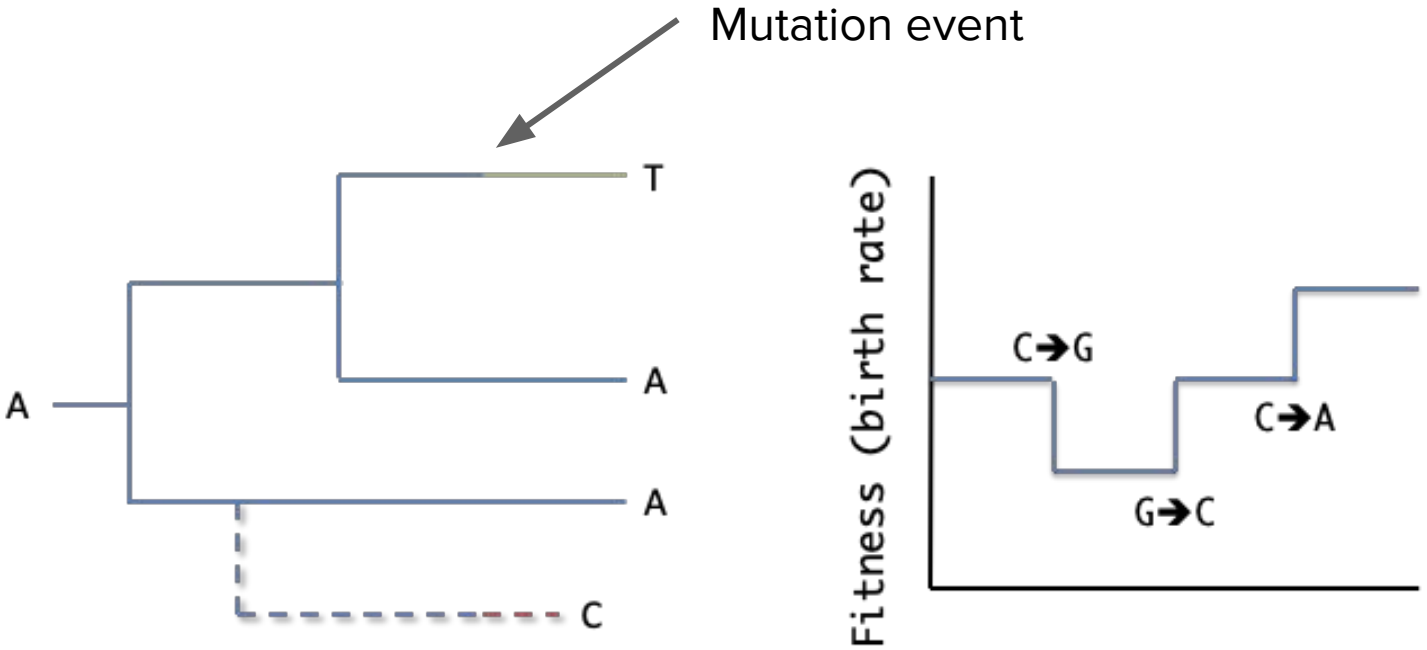
Multi-type birth-death models



Multi-type birth-death models

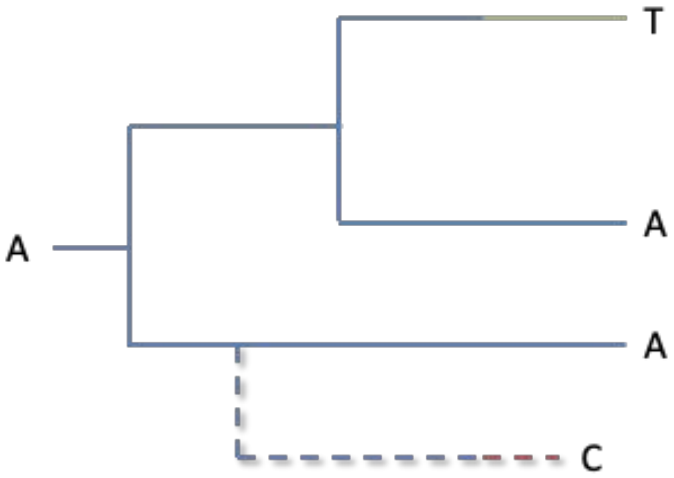


Multi-type birth-death models



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).



Multi-type birth-death models

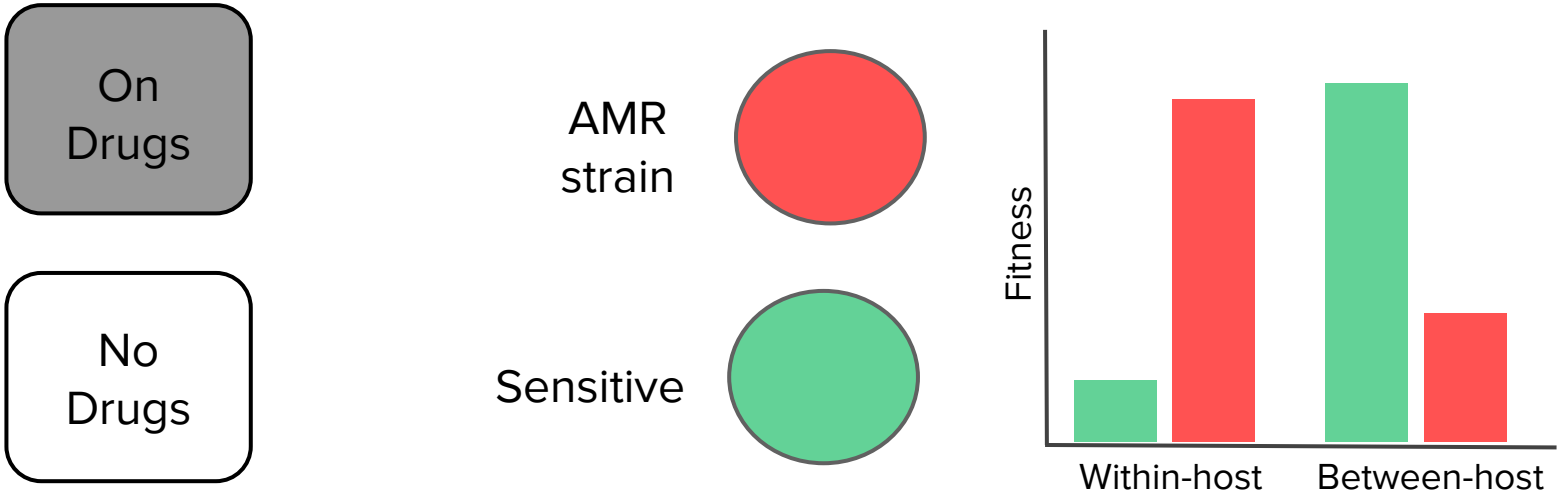
MTBD models allow us to compute the **joint likelihood** that both the tree and the observed tip genotypes evolved exactly as observed (Stadler and Bonhoeffer, 2013).

This allows us to estimate the birth/death rates of each type and the transition rates between types from a phylogeny.

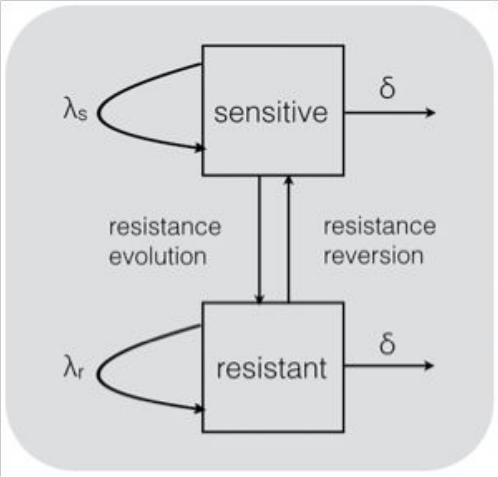
Model is implemented in the BDMM package in BEAST 2 (Kühnert *et al.*, MBE, 2016)

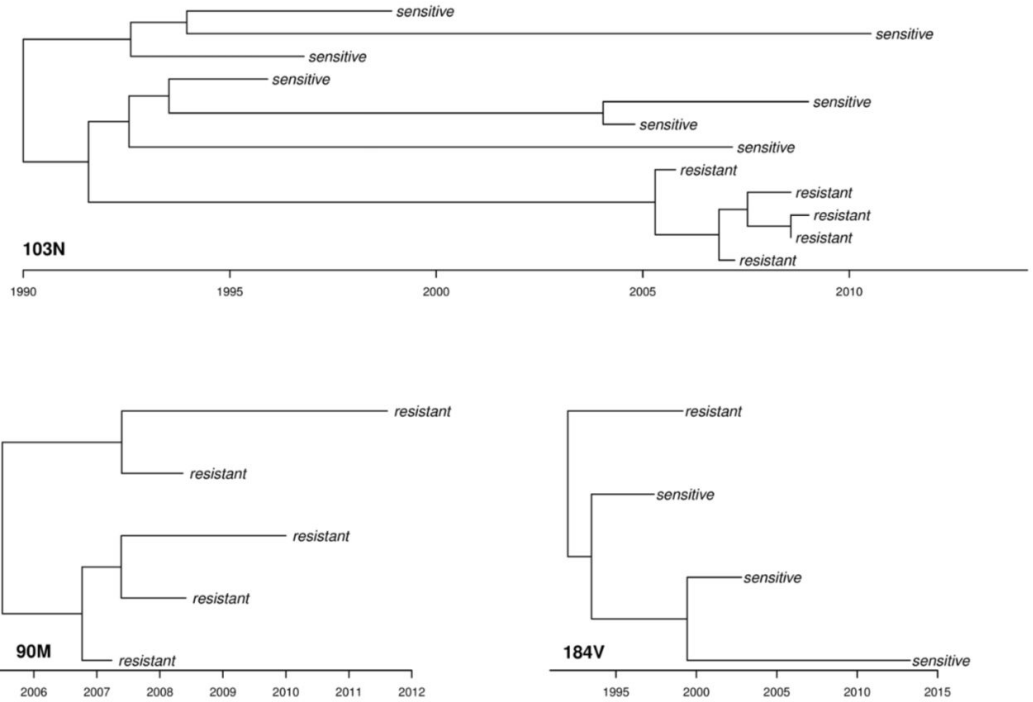
Motivation: antimicrobial resistance

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



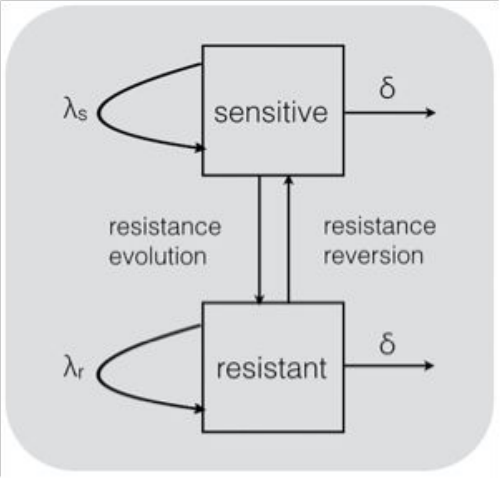
Fitness of HIV drug resistance mutations





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

Fitness of HIV drug resistance mutations



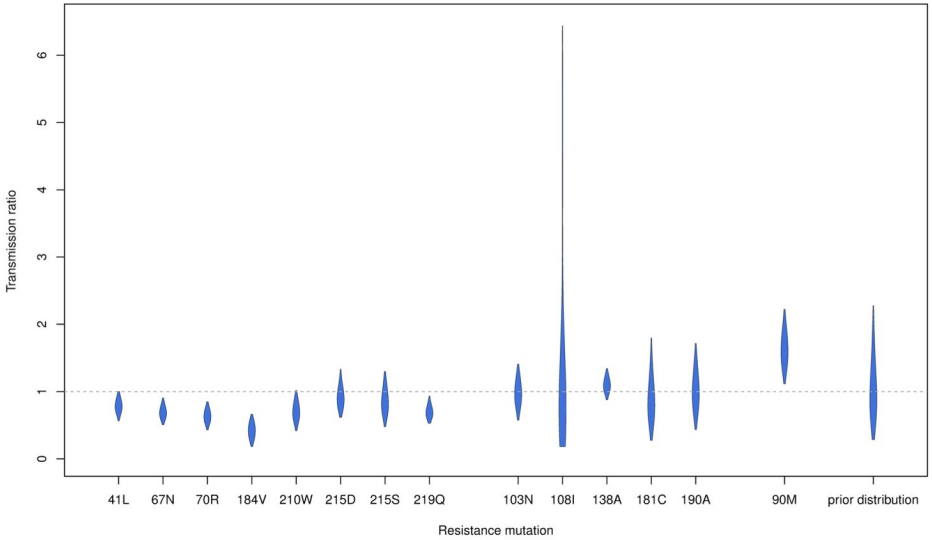


Table 1. Resistance mutations with numbers of corresponding clusters and samples, related drugs and drug usage dates within Switzerland.

	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)

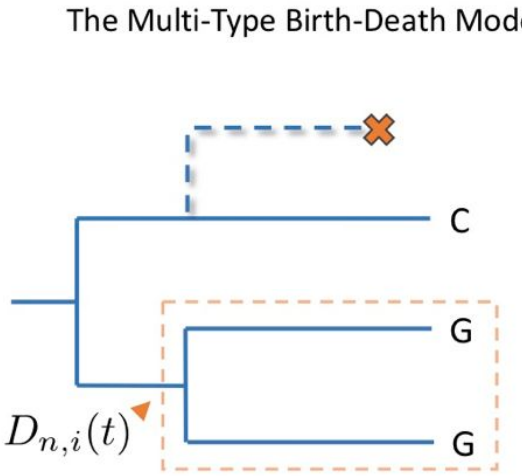
Modeling adaptive molecular evolution

If we want to consider mutations at multiple sites, we need to track all possible genotypes under the MTBD model.

The number of genotypes increases exponentially with the number of sites L (e.g. 4^L for a nucleotide model).

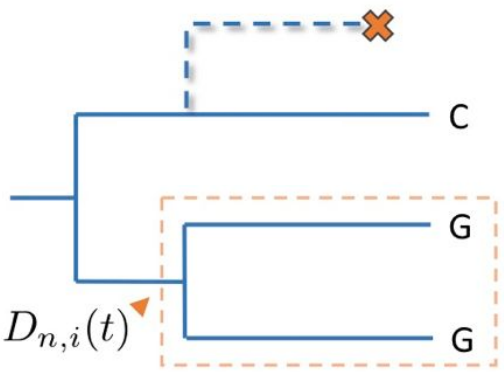
It therefore becomes prohibitively computationally expensive to track molecular evolution for more than a few evolving sites.

Marginal fitness birth-death models

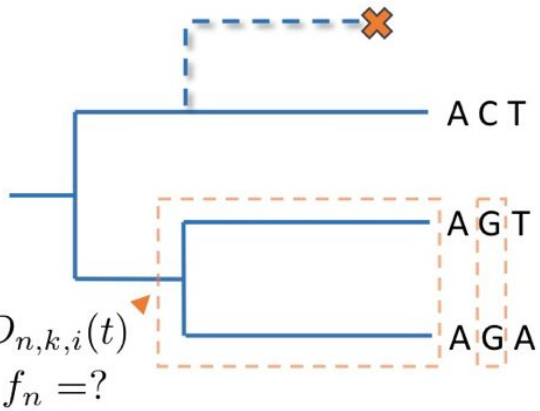


Marginal fitness birth-death models

The Multi-Type Birth-Death Model



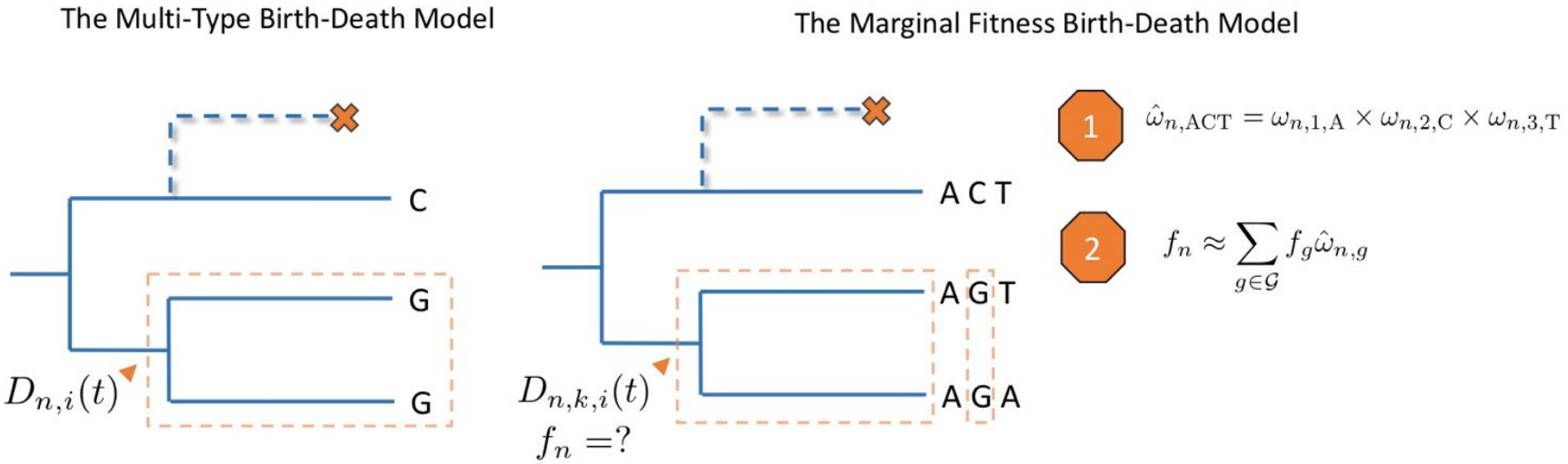
The Marginal Fitness Birth-Death Model



1 $\hat{\omega}_{n,\mathrm{ACT}} = \omega_{n,1,\mathrm{A}} \times \omega_{n,2,\mathrm{C}} \times \omega_{n,3,\mathrm{T}}$

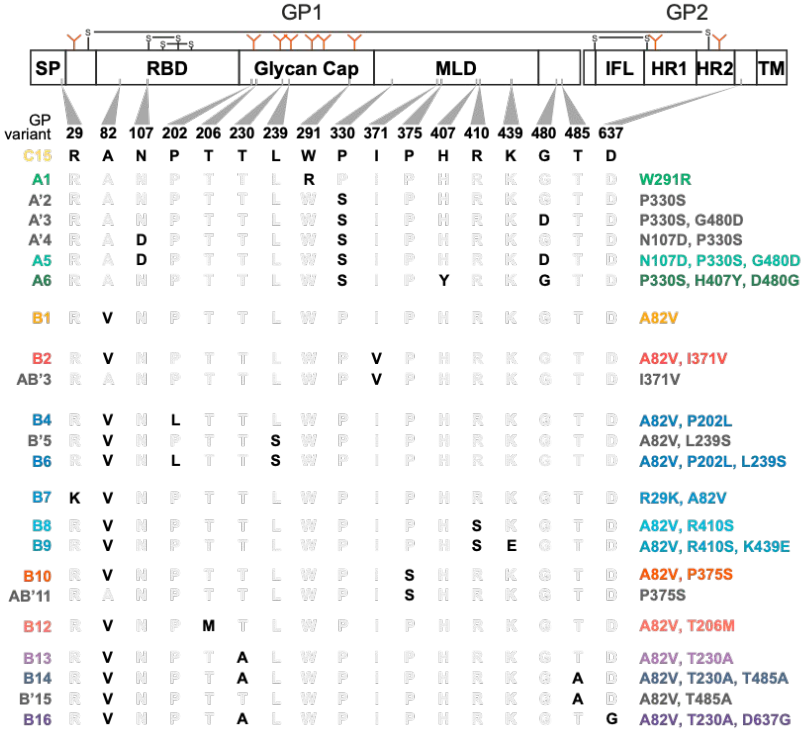
2 $f_n \approx \sum_{g \in \mathcal{G}} f_g \hat{\omega}_{n,g}$

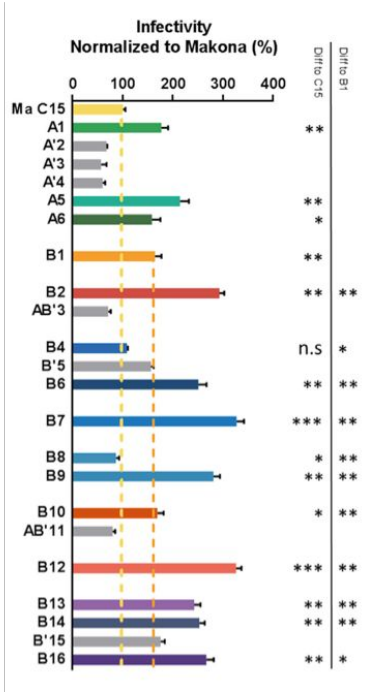
Marginal fitness birth-death models



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

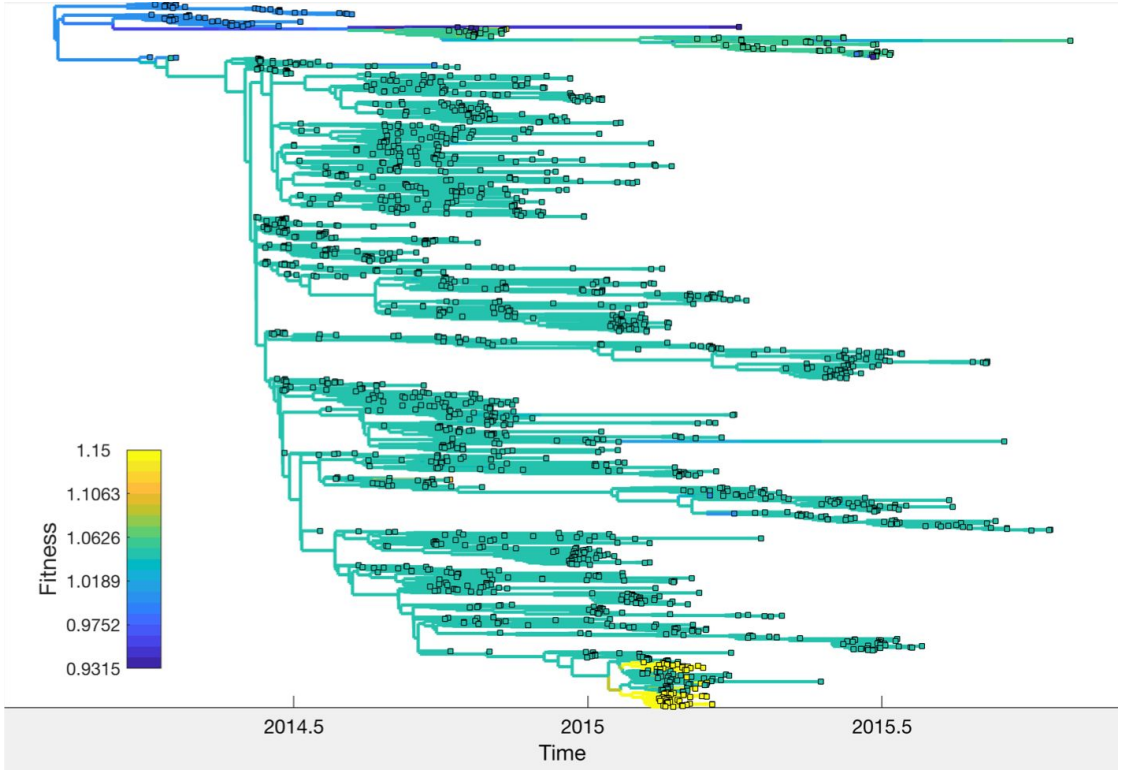
Adaptation of Ebola virus to humans





Urbanowicz *et al.* (Cell, 2016)

Adaptation of Ebola virus to humans



Rasmussen and Stadler (eLife, 2019)

Multi-type birth-death models in BEAST

The marginal fitness birth death model model is implemented in BEAST2 as the *LUMIERE* package.

On Wednesday, we will use the BDMM package to estimate the transmission fitness of AMR mutations.