Multi-type birth-death models and adaptive molecular evolution

Molecular Epidemiology of Infectious Diseases
Lecture 9

March 9th, 2026

Adaptive pathogen evolution

So far we have mainly 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.

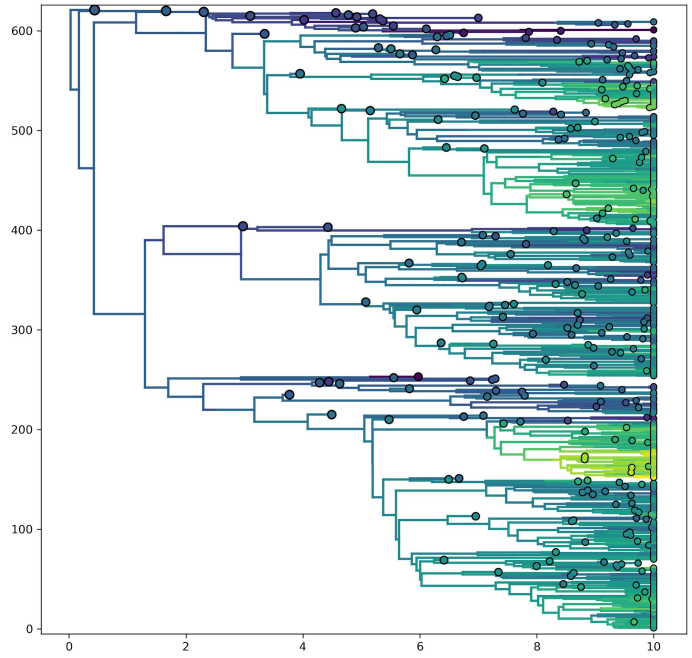
Between hosts, the birth rate can be thought of as a transmission rate and the death rate as the rate infected hosts are removed. The growth rate *r* therefore quantifies a pathogen's epidemic potential at the host-population level.

Growth rates can also be related back to selection coefficients in classic population genetics where $s = r_{mut} - r_{wt}$.

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



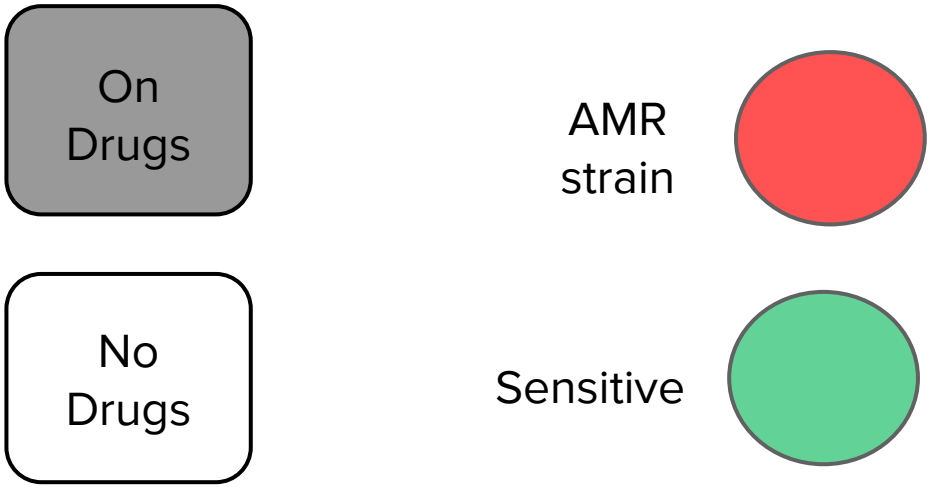
Selection shapes pathogen phylogenies

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

Grenfell *et al.* (Science, 2004)

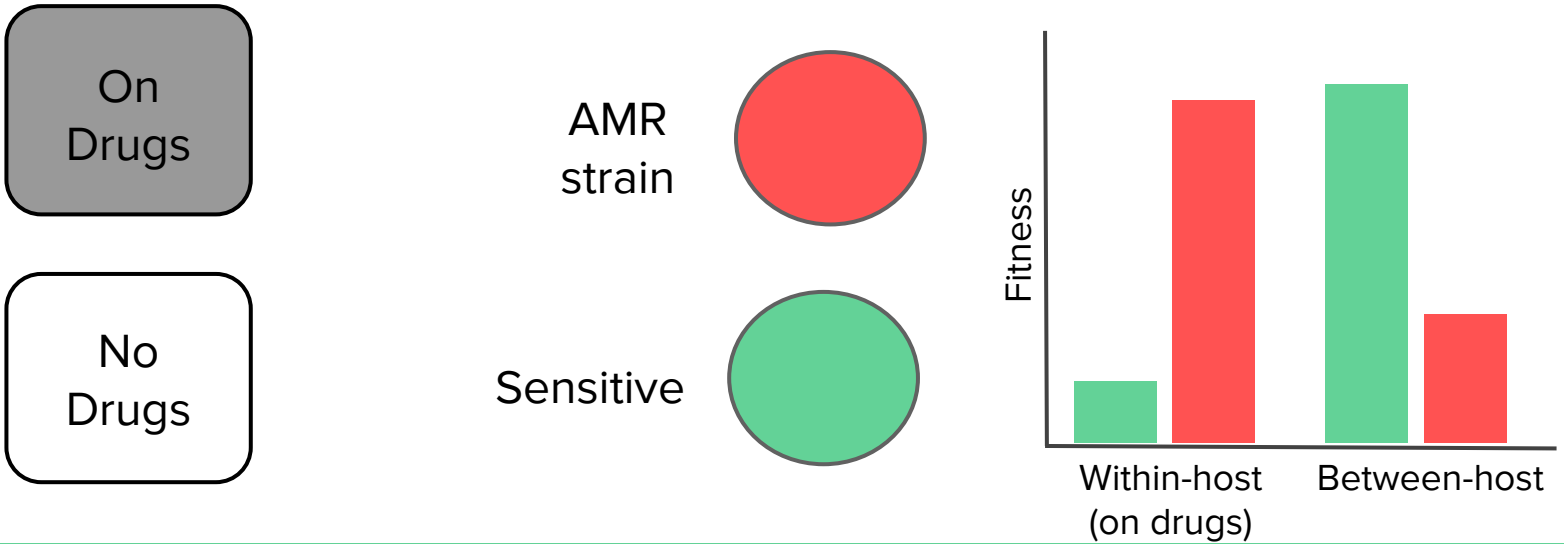
Motivation: antimicrobial resistance

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



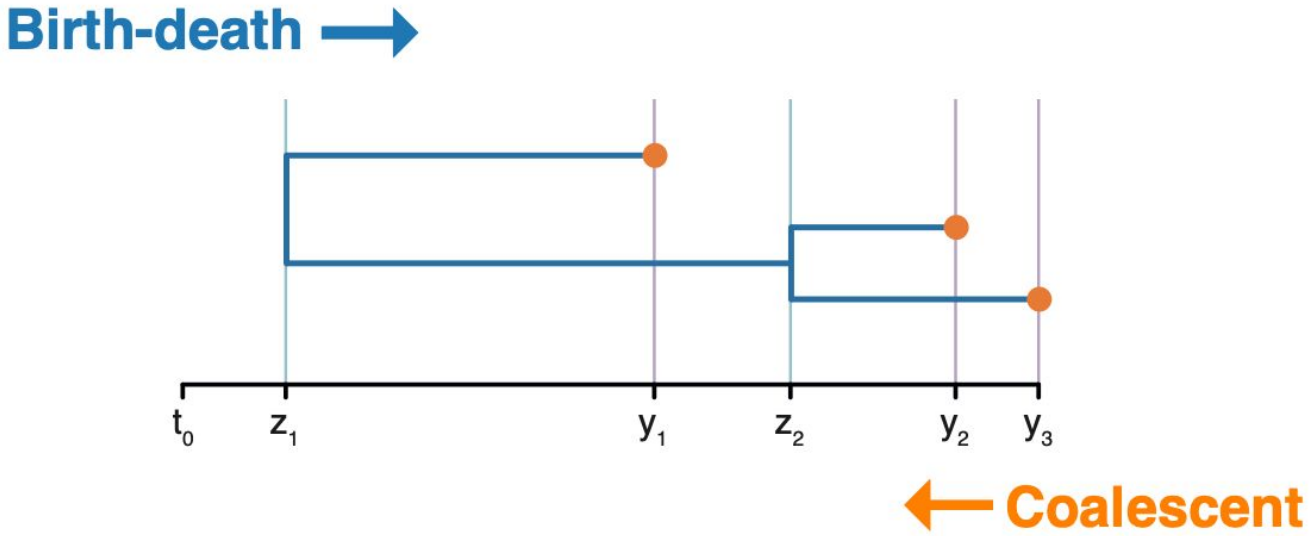
Motivation: antimicrobial resistance

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.

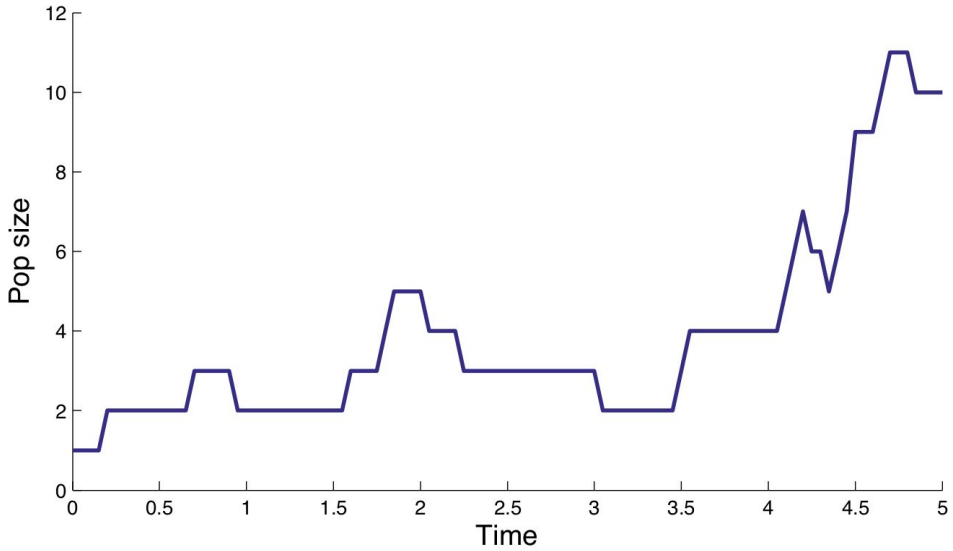
In the most basic model individuals reproduce and die. Nothing else happens.

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.

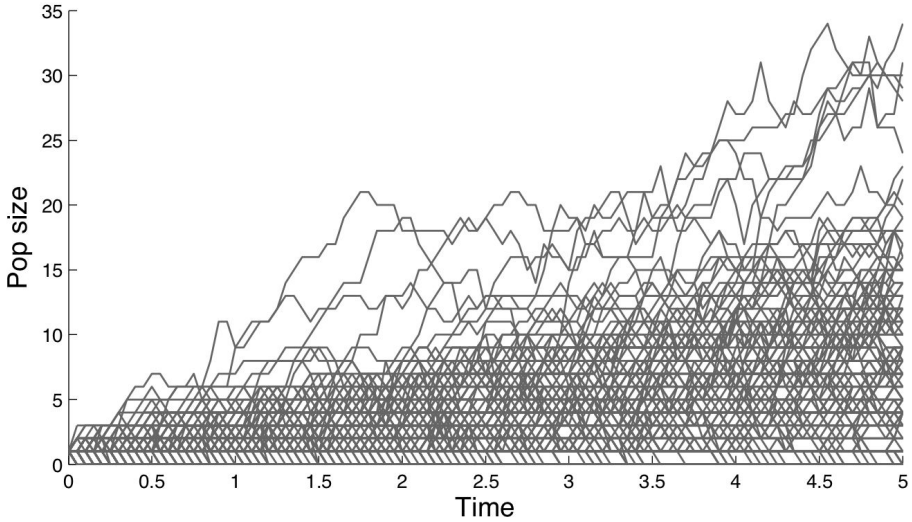
The birth-death process

A single stochastic realization of the birth-death process:



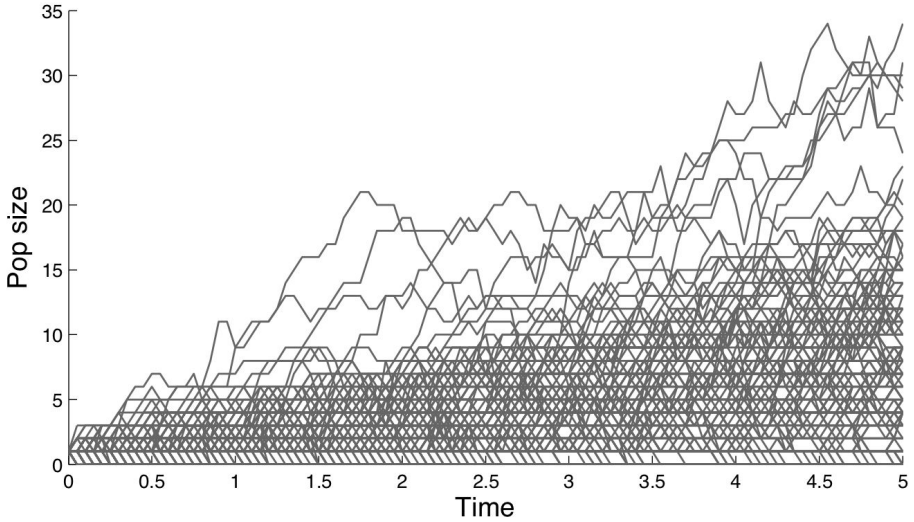
The birth-death process

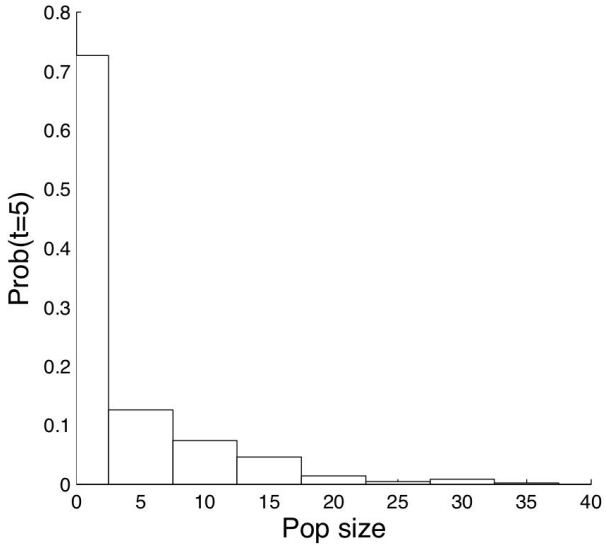
An ensemble of stochastic realizations



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

$\lambda = \text{birth rate}$

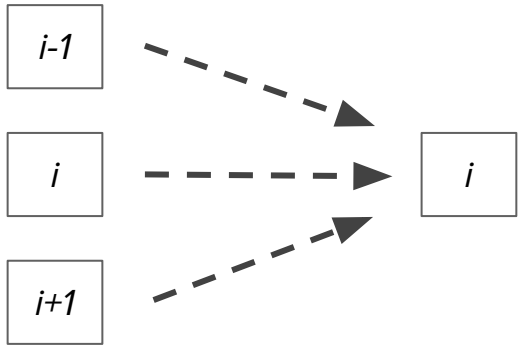
$\mu = \text{death rate}$

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

The birth-death process

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

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

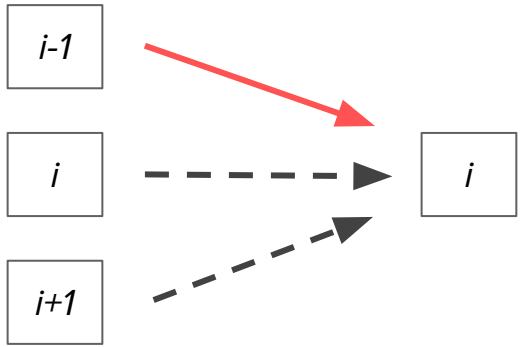


The birth-death process

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

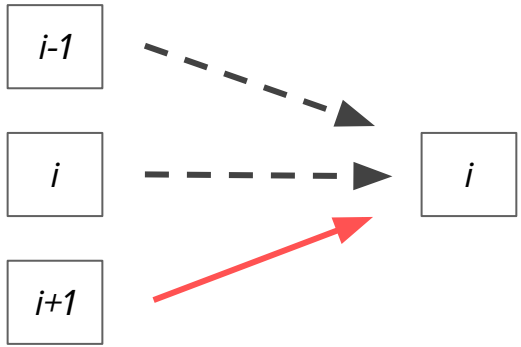


The birth-death process

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

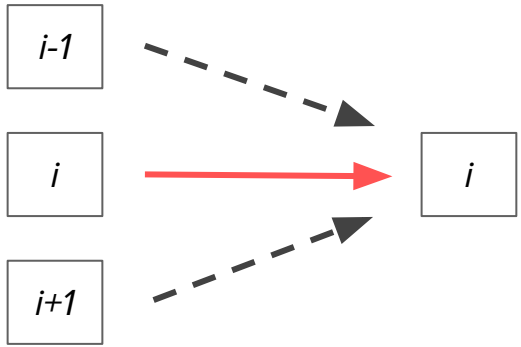


The birth-death process

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:

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

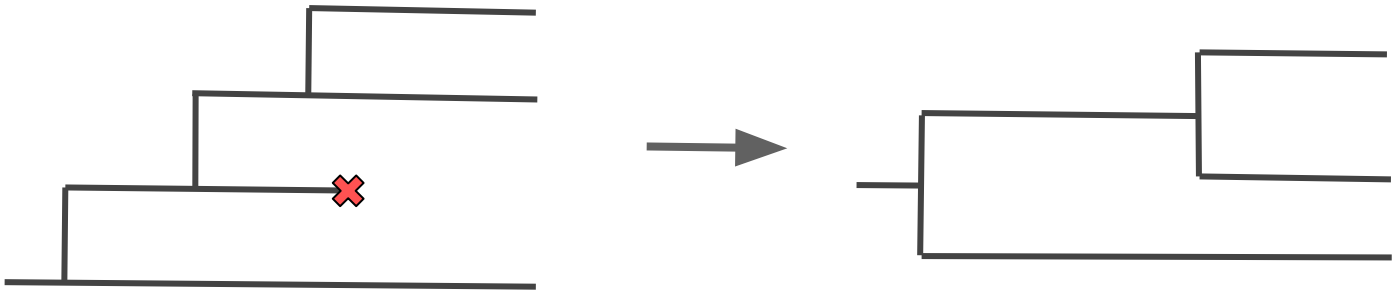
The probability that there is exactly one living individual at time t:

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

Why is this useful?

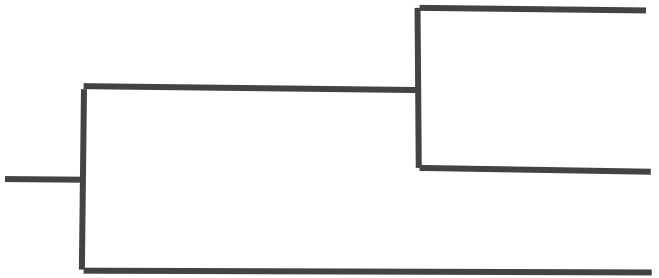
Birth-death processes as trees

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)

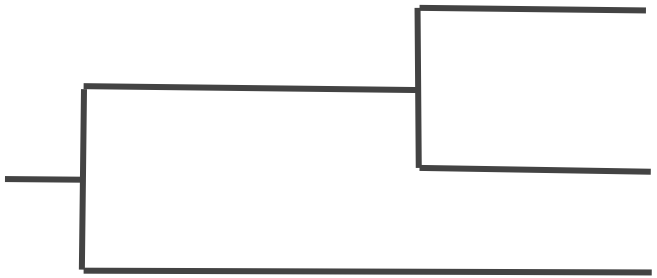
BD tree likelihood: complete sampling

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

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

λ = birth rate

μ = death rate



Thompson (1975); Rannala & Yang (1996)

BD tree likelihood: incomplete sampling

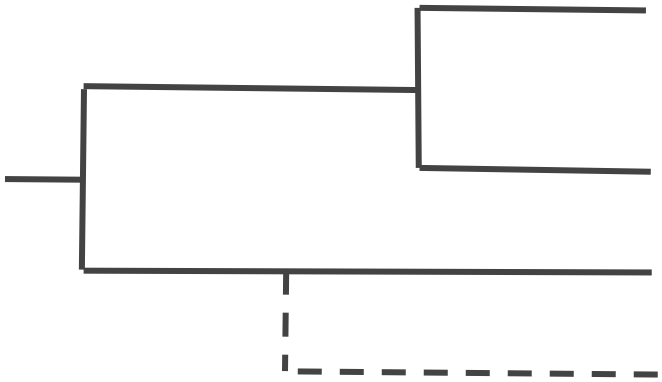
$$L(T|\lambda,\mu,\rho) \propto (\lambda\rho)^{n-1}\prod_{i=1}^{n-1} p_1(t_i)$$

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

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.

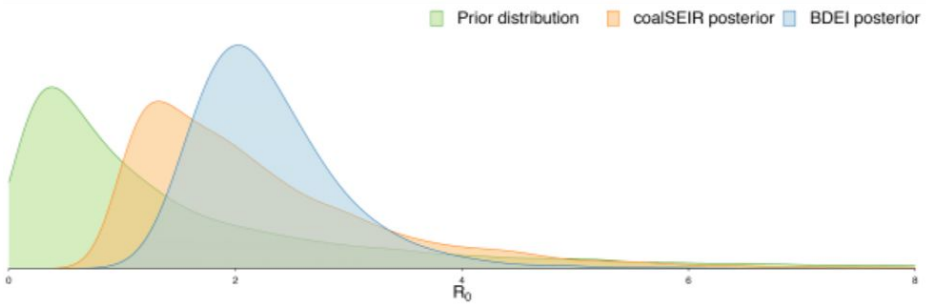
We can estimate both the birth and death rate if we know the sampling fraction.

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

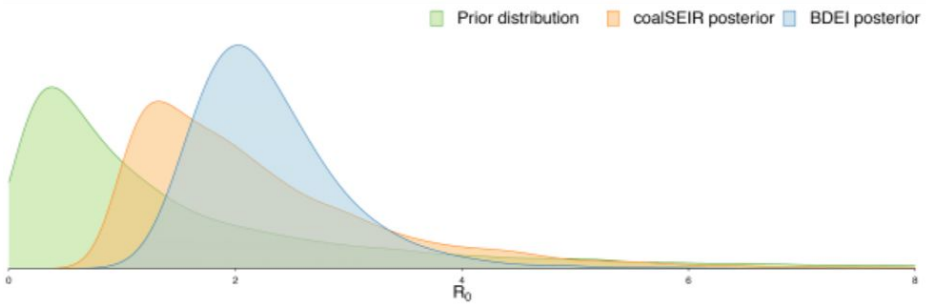
The key epidemiological parameter R_0, the number of secondary infections caused by an infected individual, can be estimated from pathogen phylogenies.



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.



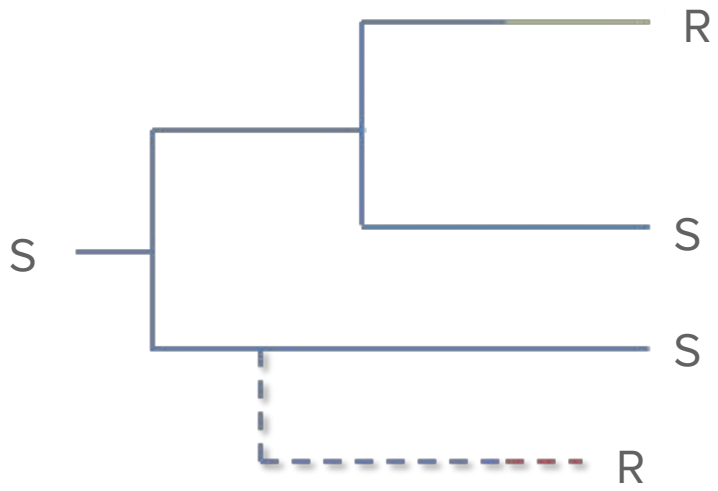
WHO Response Team Estimate:
R0 = 2.02 (95% CI: 1.79-2.26)

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

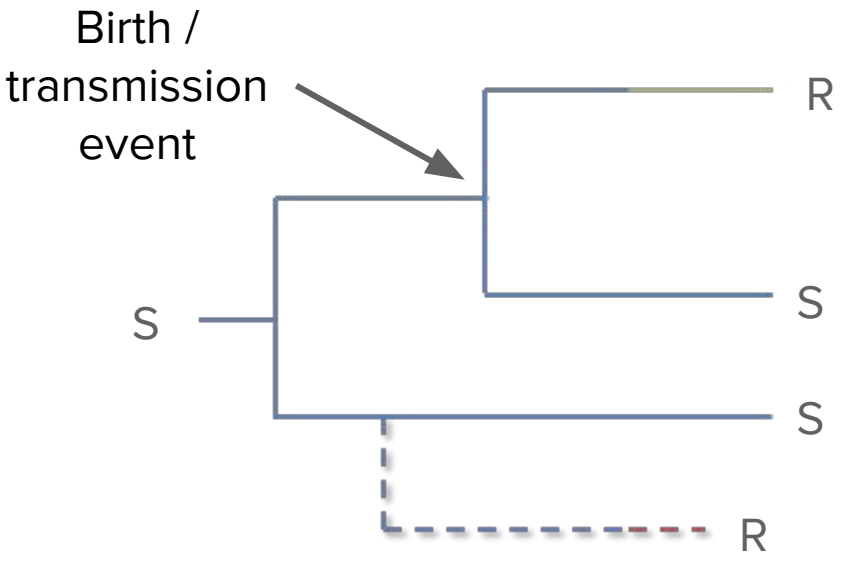
Stochastic branching processes that allow for different types of individuals (e.g. genotypes) to vary in their birth or death rates and therefore their fitness values.



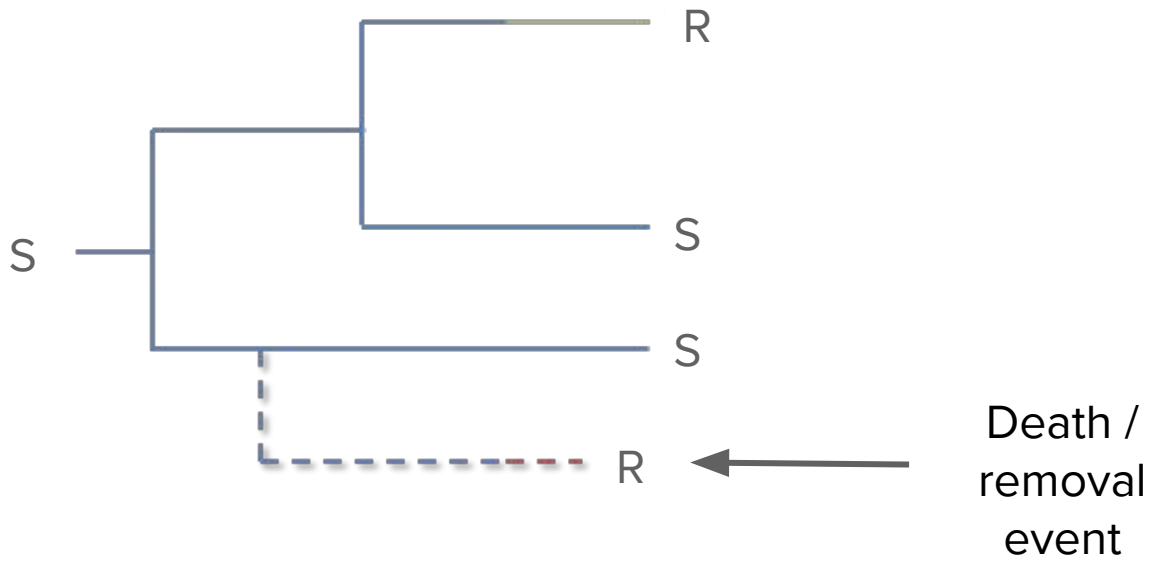
S = sensitive genotype

R = resistant genotype

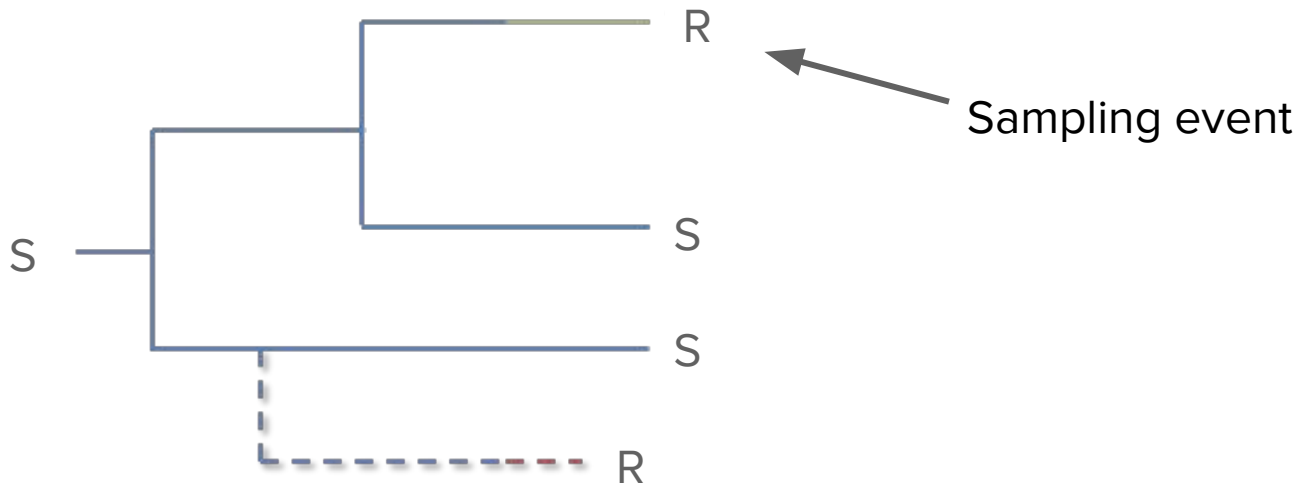
Multi-type birth-death models



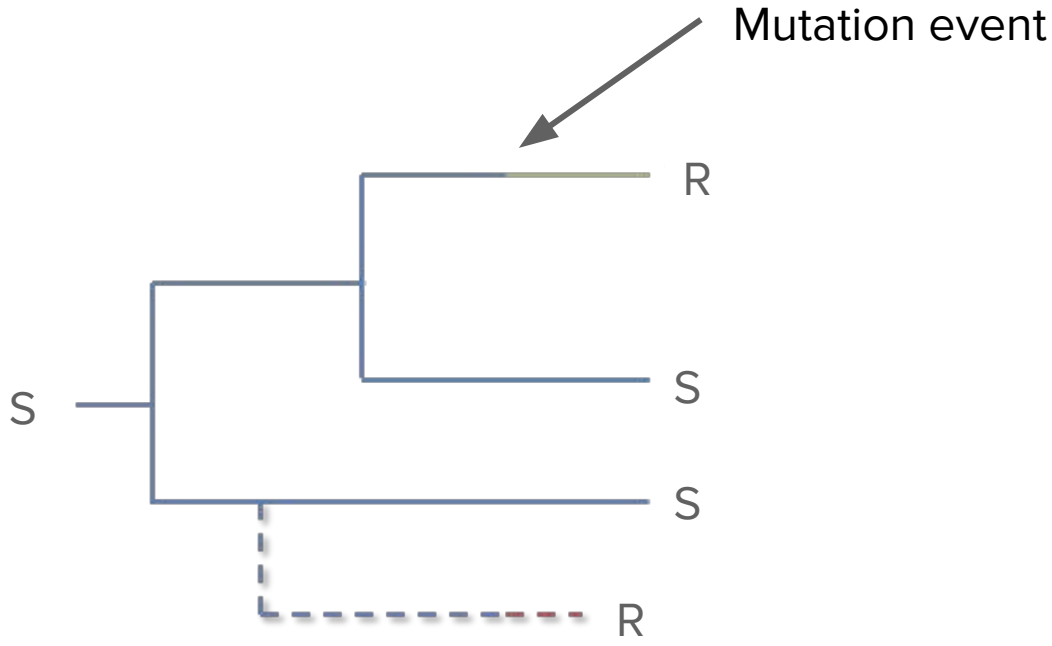
Multi-type birth-death models



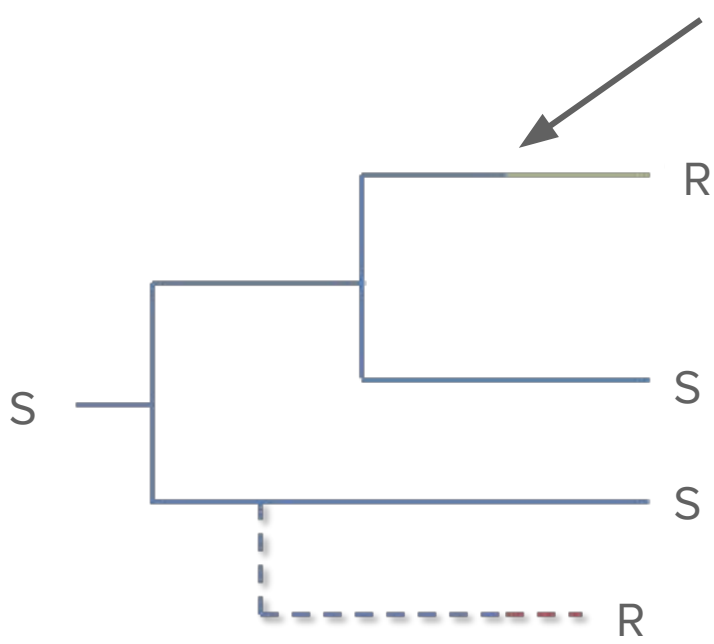
Multi-type birth-death models

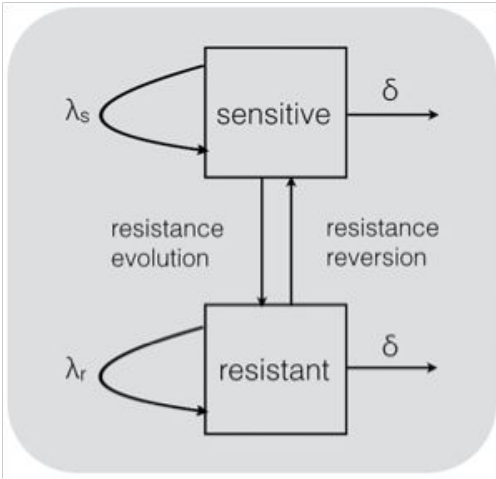


Multi-type birth-death models



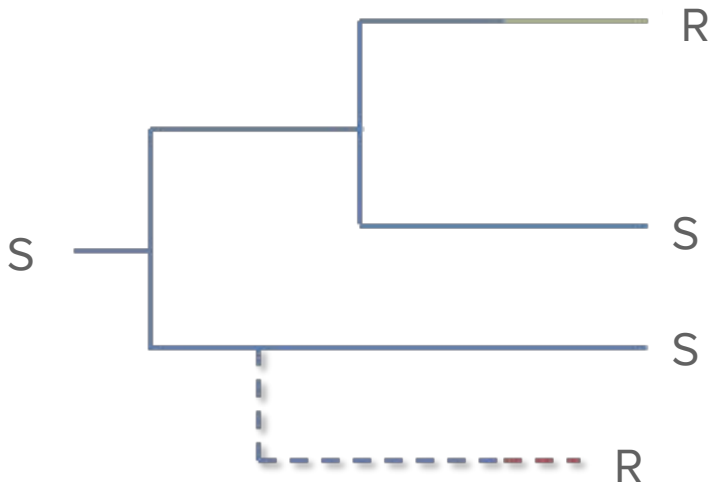
Multi-type birth-death models



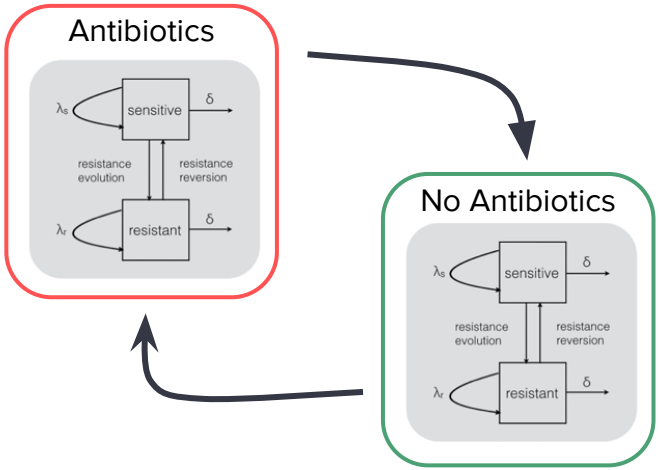


Kühnert *et al.* (PLoS Pathogens, 2018)

Multi-type birth-death models

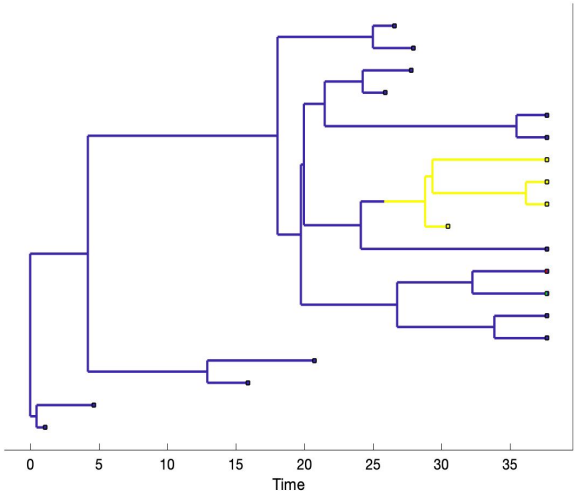


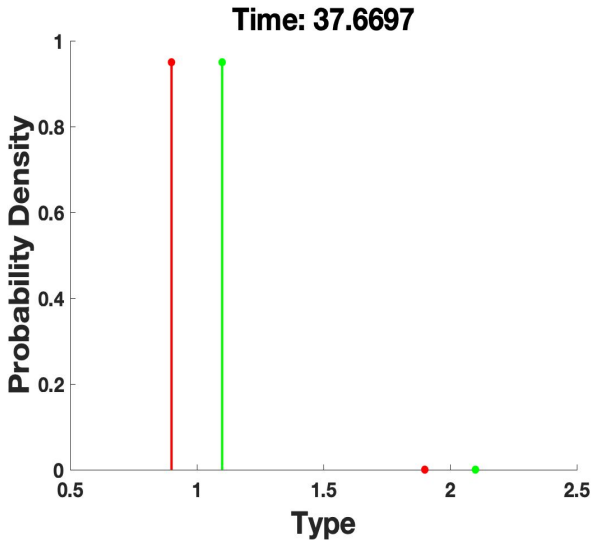
Types can represent intrinsic (genotype) or extrinsic (environment) features.



Multi-type birth-death models

MTBD models allow us to compute the likelihood that the tree evolved exactly as observed given the birth/death rate of each type.

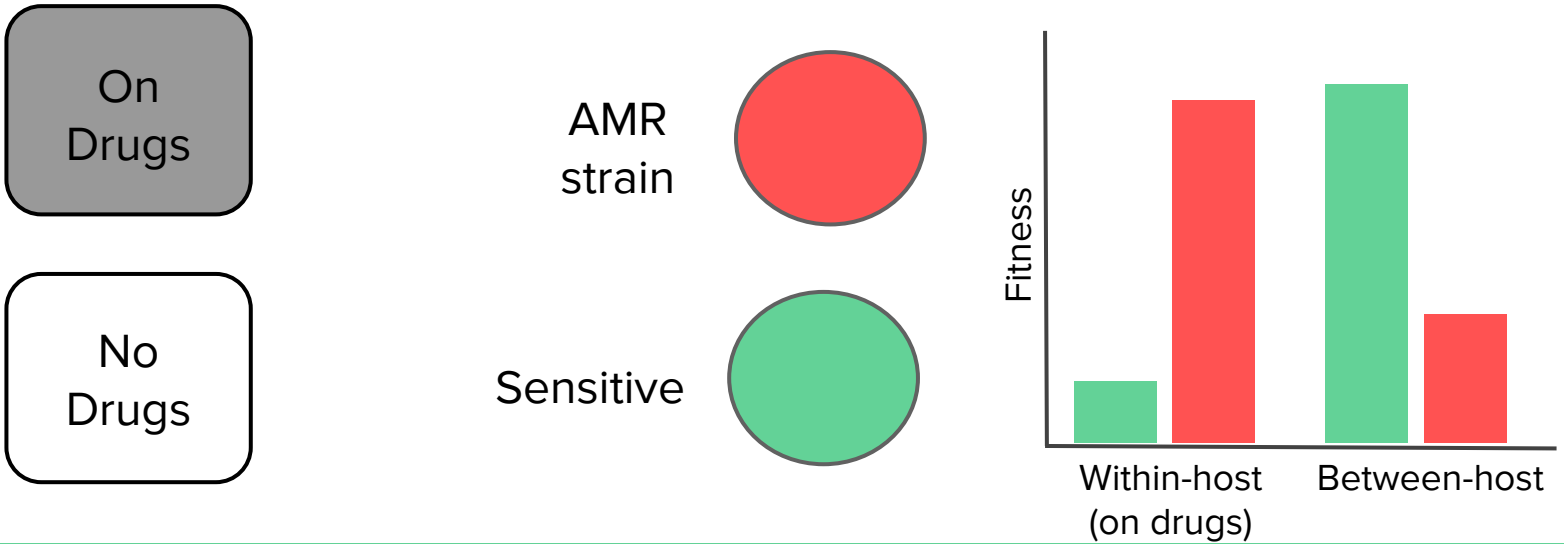




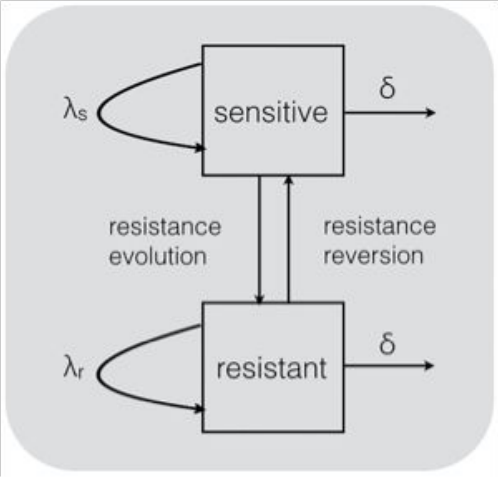
GIF courtesy of Marco Hamins-Puertolas

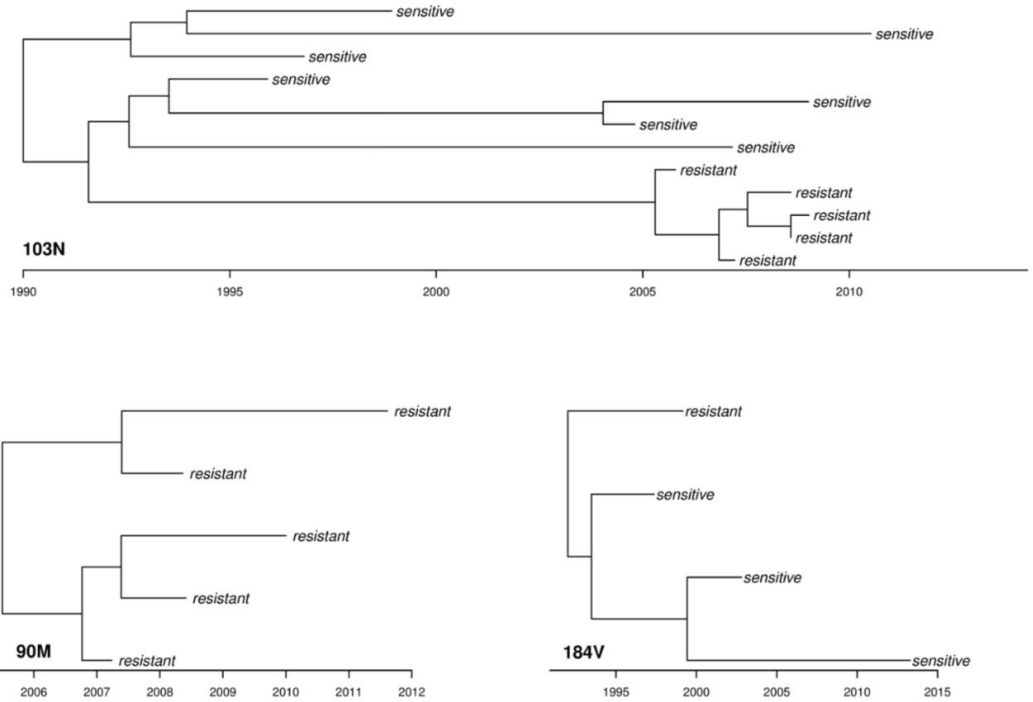
Motivation: antimicrobial resistance

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



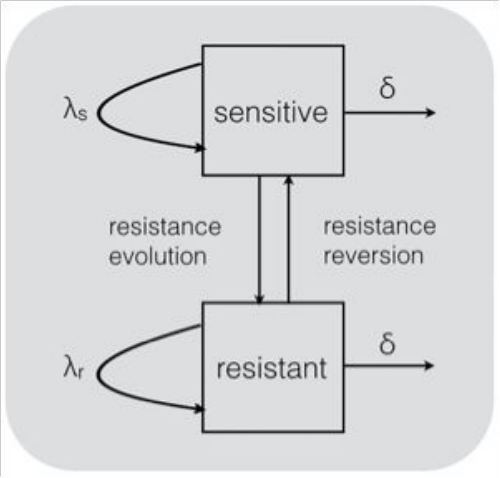
Fitness of HIV drug resistance mutations





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

Fitness of HIV drug resistance mutations



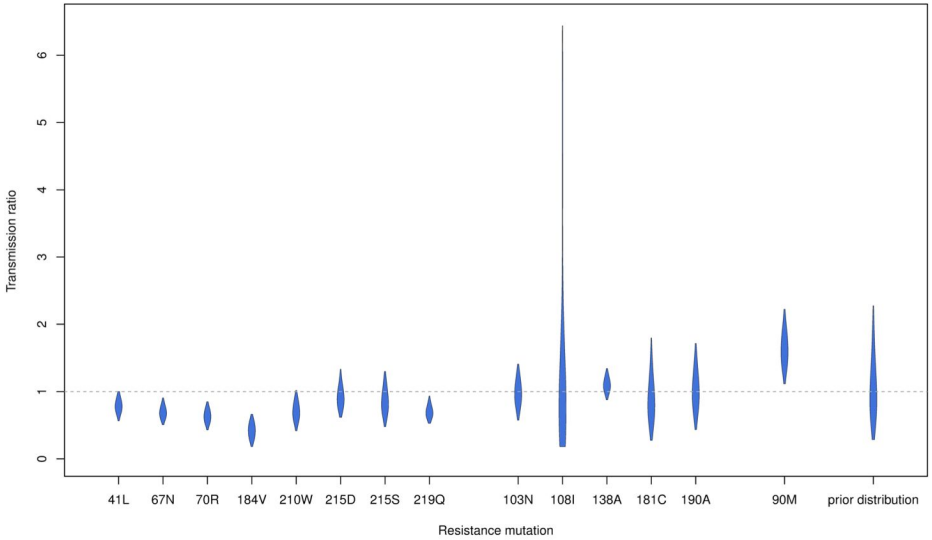


Table 1. Resistance mutations with numbers of corresponding clusters and samples, related drugs and drug usage dates within Switzerland.

	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

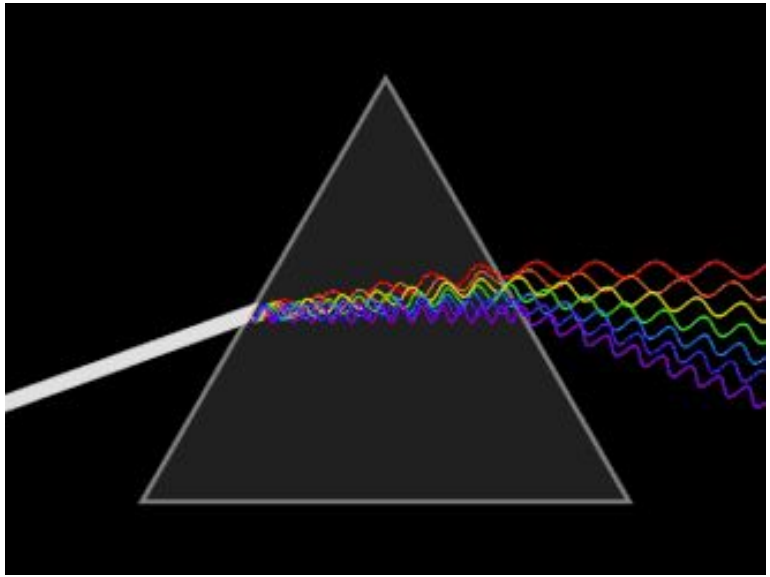
Kühnert *et al.* (PLoS Pathogens, 2018)

But pathogen fitness is a multifaceted trait determined by many different components

What determines pathogen fitness?

Can we decompose fitness into its component parts in order to learn what pathogen features are most important to overall fitness?

Fitness



https://en.wikipedia.org/wiki/Dispersive_prism

Mutation A

Mutation B

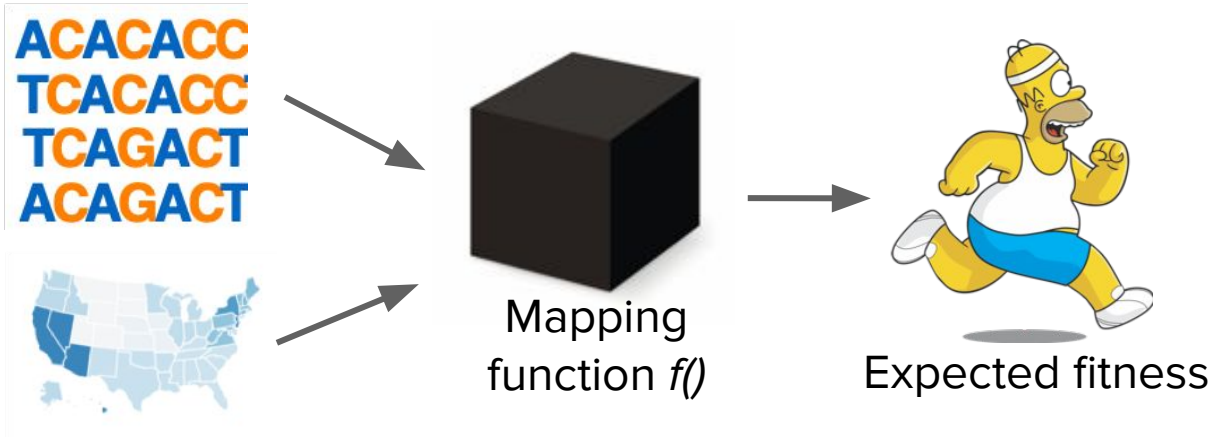
Mutation C

Environment A

Environment B

Mapping pathogen features to fitness

We want to learn how many different ***features*** including a pathogen's genotype and environment determine its fitness



Goal: learn the ***fitness mapping function*** that predicts a pathogen's expected fitness given its features (i.e. predictor variables).

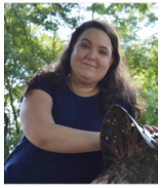
What drives *E. coli* ST 131's fitness?

ST131 causes the a substantial fraction of blood and urinary tract *E. coli* infections.

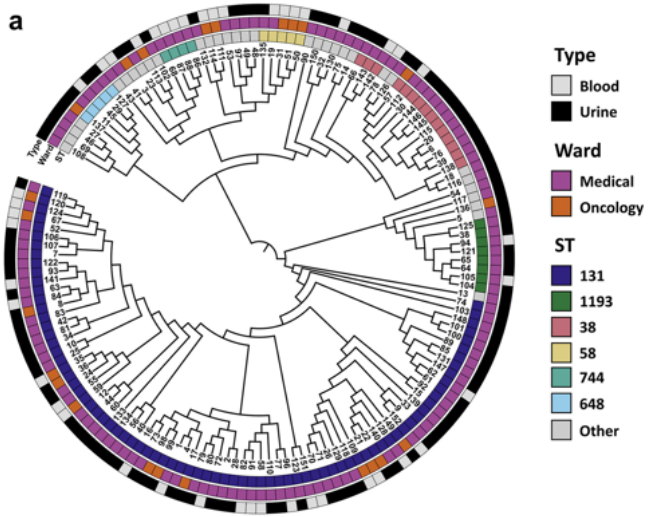
Rapid growth believed to be driven by ESBLs and resistance to other antibiotics like fluoroquinolones.

Success could also be attributable to virulence factors that allow for colonization/persistence

The fitness costs/benefits of AMR and their contribution to ST 131's overall fitness remains unclear

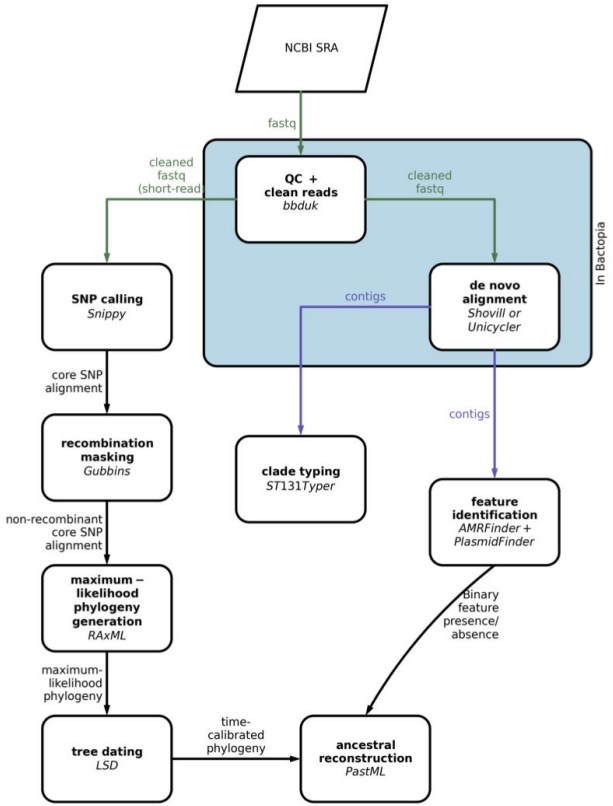


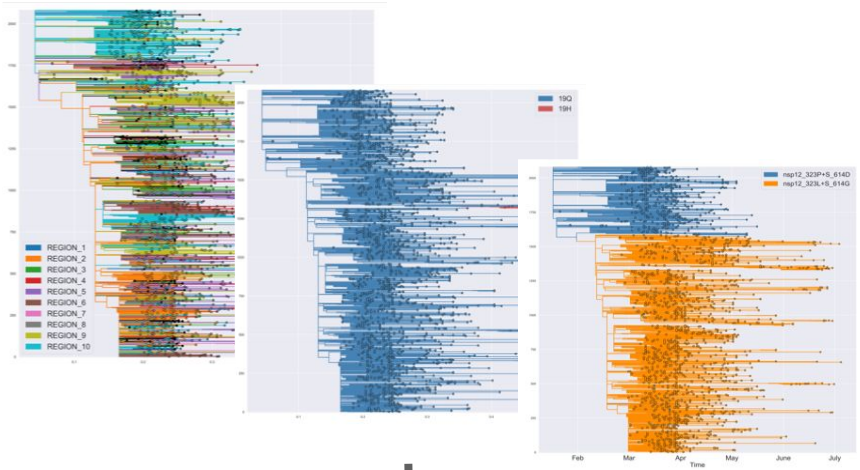
Lenora Kepler (NCSU Bioinformatics PhD student)

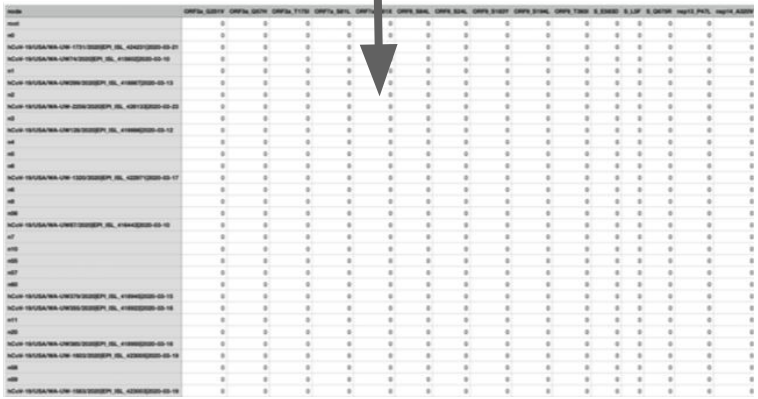


Mahmud *et al.* (mSystems, 2022)

ST 131 workflow







Predictors of ST 131 fitness

69 genetic features including AMR mutations/genes, virulence factors and plasmids

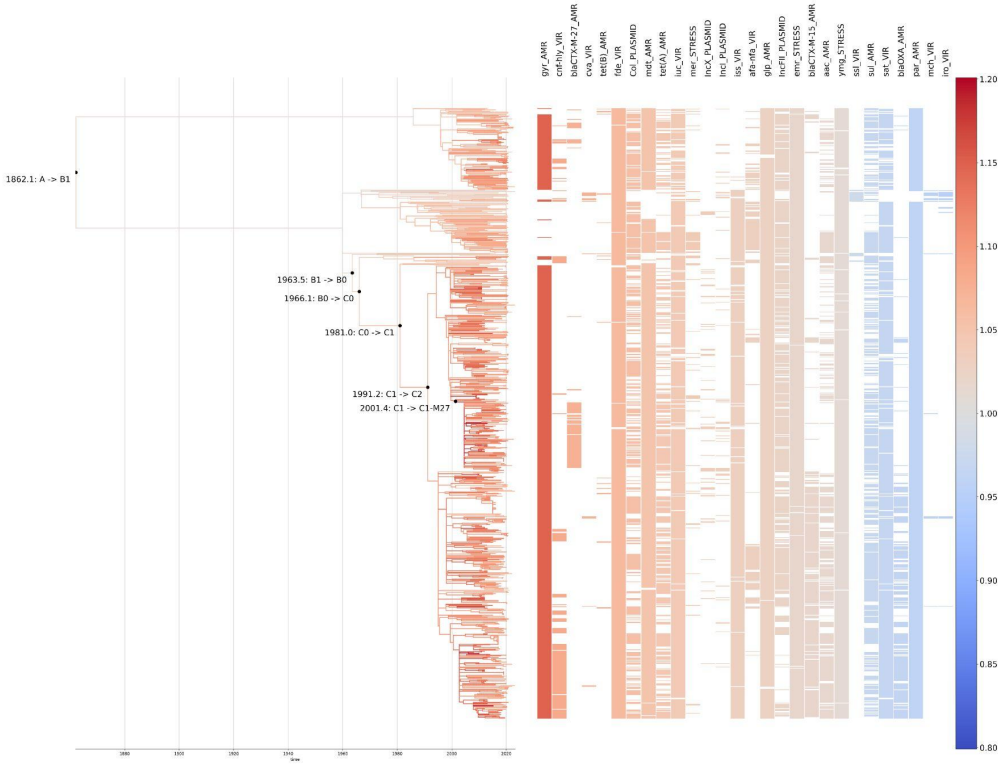
Time-varying background effects to account for changing epidemic dynamics

Study/bioproject effects to account for differential sampling between studies

Branch-specific random effects to account for additional unmodeled sources of fitness variation

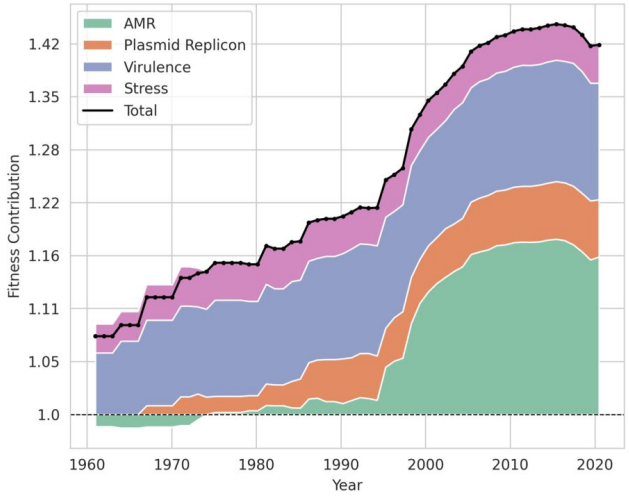
Fitness mapping function: $$\log(F(x)) = \sum_{i \in \mathcal{X}} \log(\beta_i) x_{n,i} + \log(u_n).$$

The determinants of *E. coli* ST131 fitness



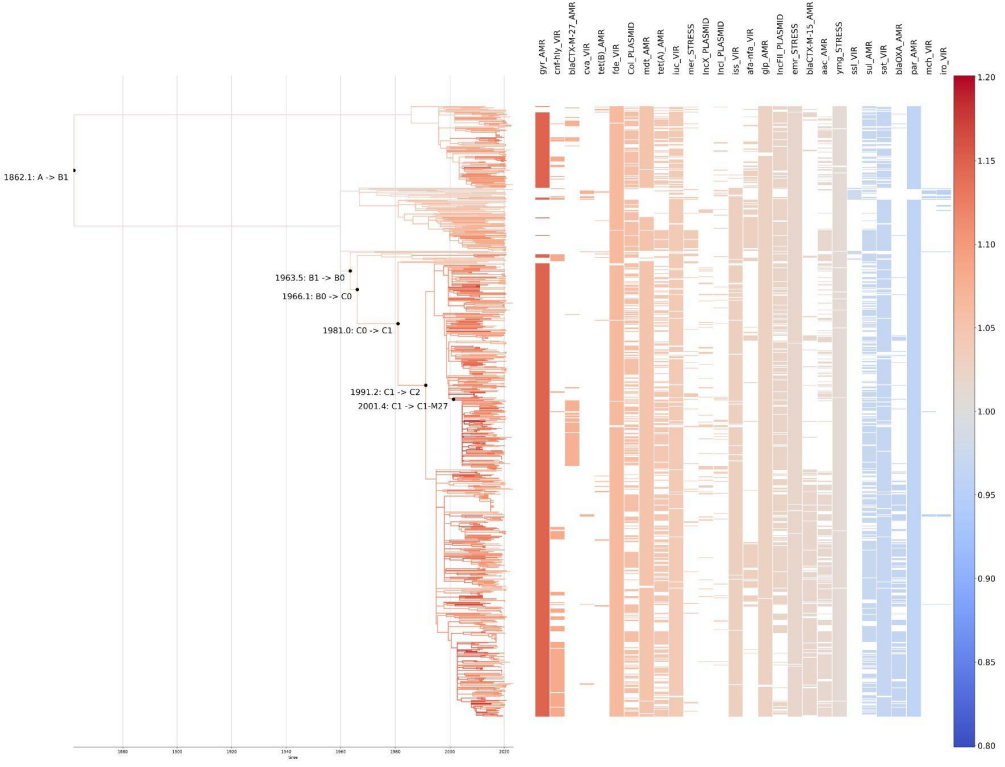
Kepler *et al.* (bioRxiv, 2024)

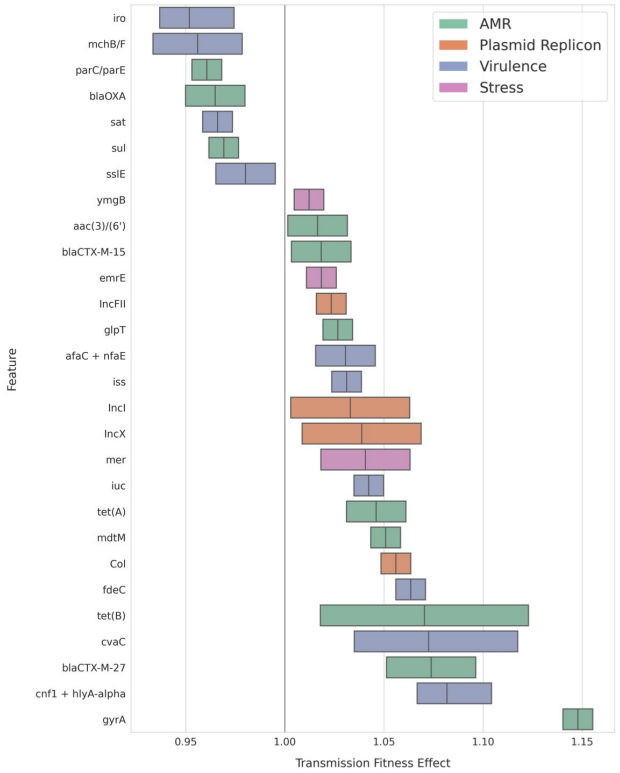
The determinants of *E. coli* ST131 fitness



Kepler *et al.* (bioRxiv, 2024)

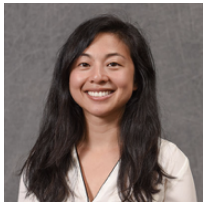
The determinants of *E. coli* ST131 fitness





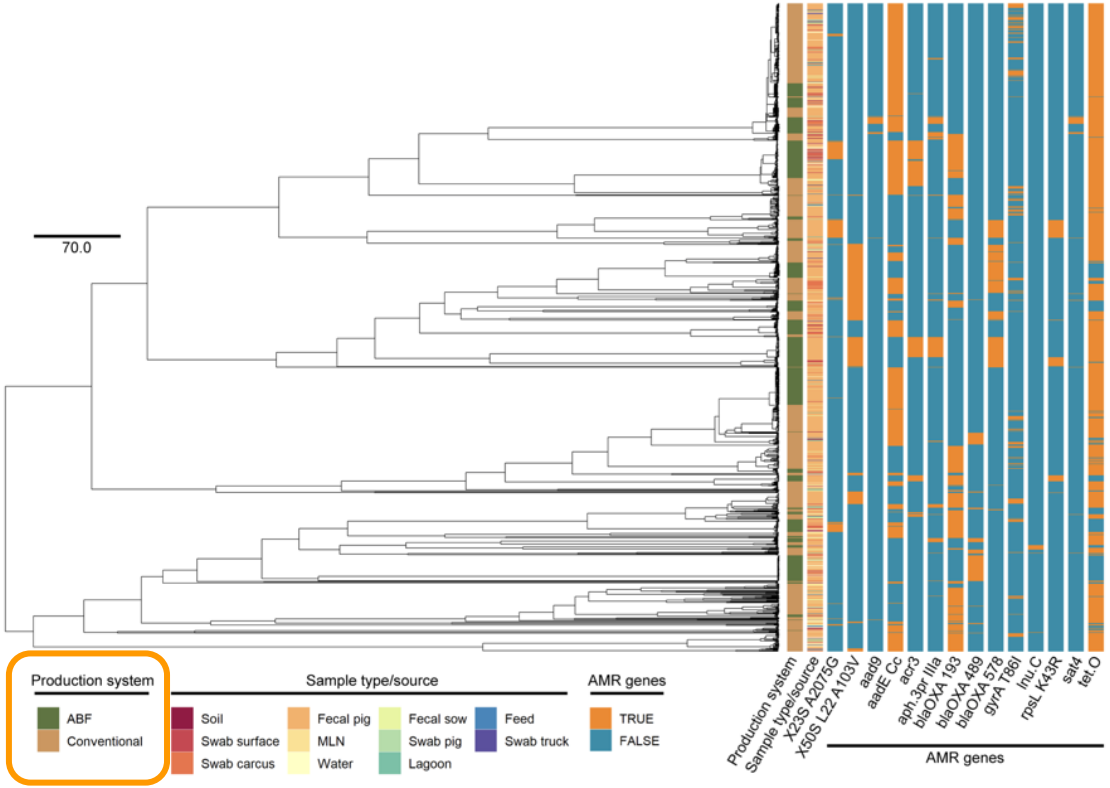
Kepler *et al.* (bioRxiv, 2024)

AMR fitness effects in *Campylobacter*

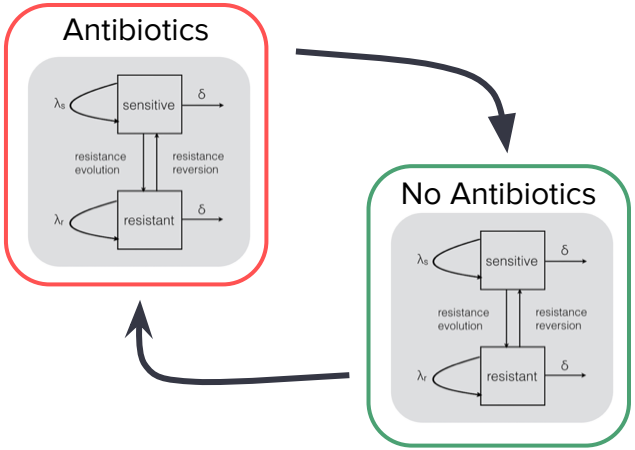


Annie Wang (NCSU DVM-PhD student)

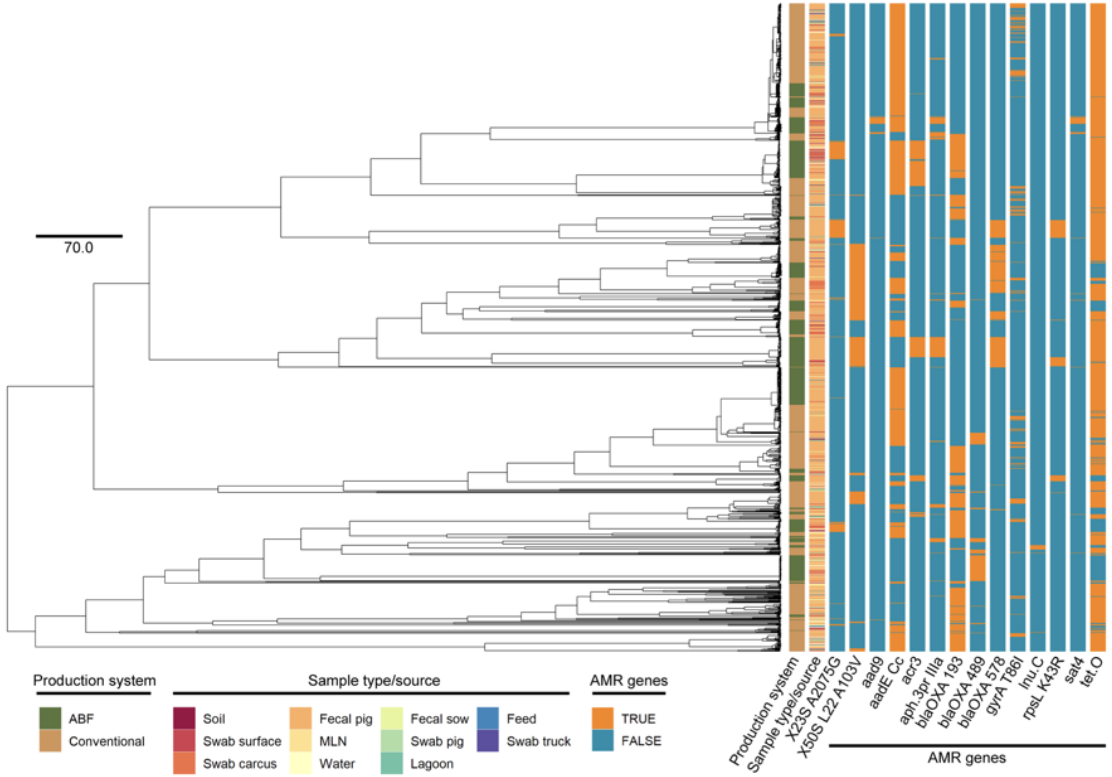
AMR fitness effects in *Campylobacter*

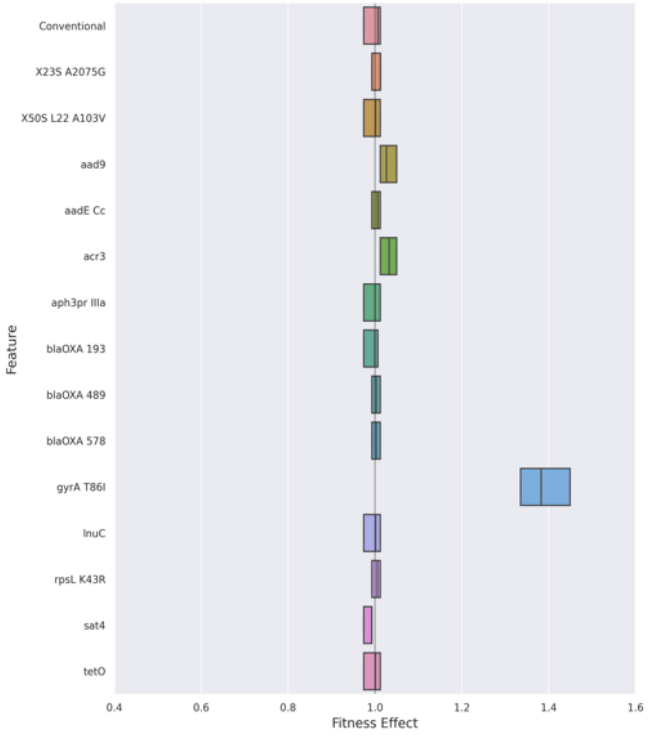


Can estimate fitness benefits of AMR as well as costs on antibiotic free farms!



AMR fitness effects in *Campylobacter*





What’s going on with fluoroquinolones?

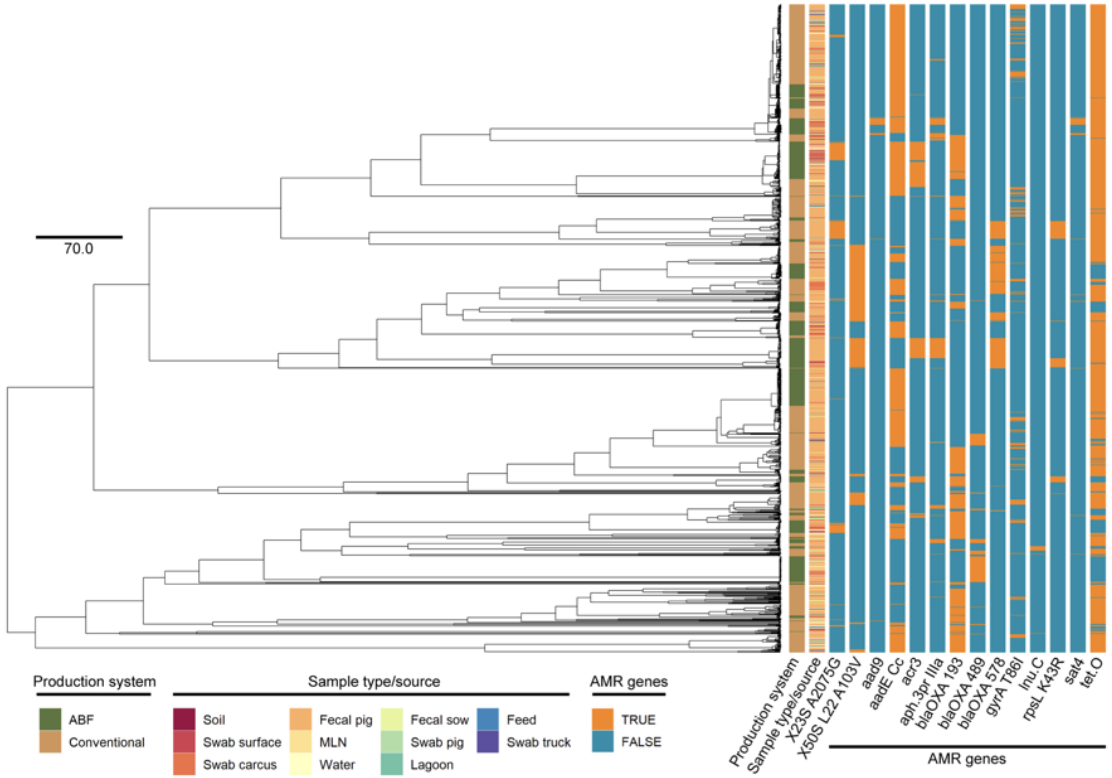
gyrA mutations conferring resistance to fluoroquinolones are consistently found to have very large beneficial fitness effects.

Other studies have found gyrA mutations can be beneficial even in the absence of antibiotics (Luo *et al*. 2005; Baker *et al*., 2013).

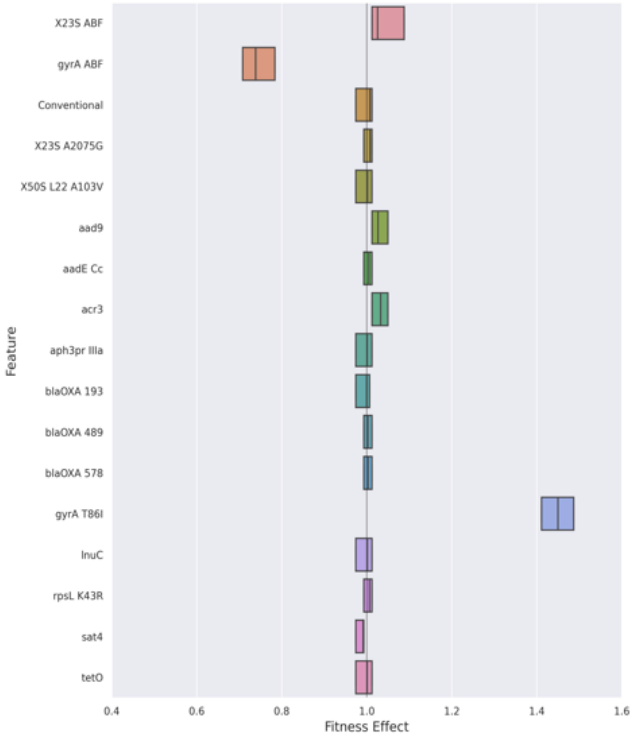
gyrA mutations may alter DNA supercoiling allowing for adaptive changes in gene expression in many different stressful environments.

Do gyrA mutations carry a fitness cost in the absence of antibiotics?

AMR fitness effects in *Campylobacter*



With gyrA x farm-type interactions



Multi-type birth-death models

MTBD models allow us to compute the joint likelihood that both the tree and the sampled tip genotypes evolved exactly as observed..

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

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

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