

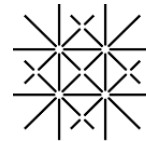
# Influenza: Limited predictability of evolution Ecology of host and pathogen

Pierre Barrat-Charlaix



**Politecnico  
di Torino**

Collaborators:  
**Richard Neher**  
**John Huddleston**  
**Trevor Bedford**



**Universität  
Basel**

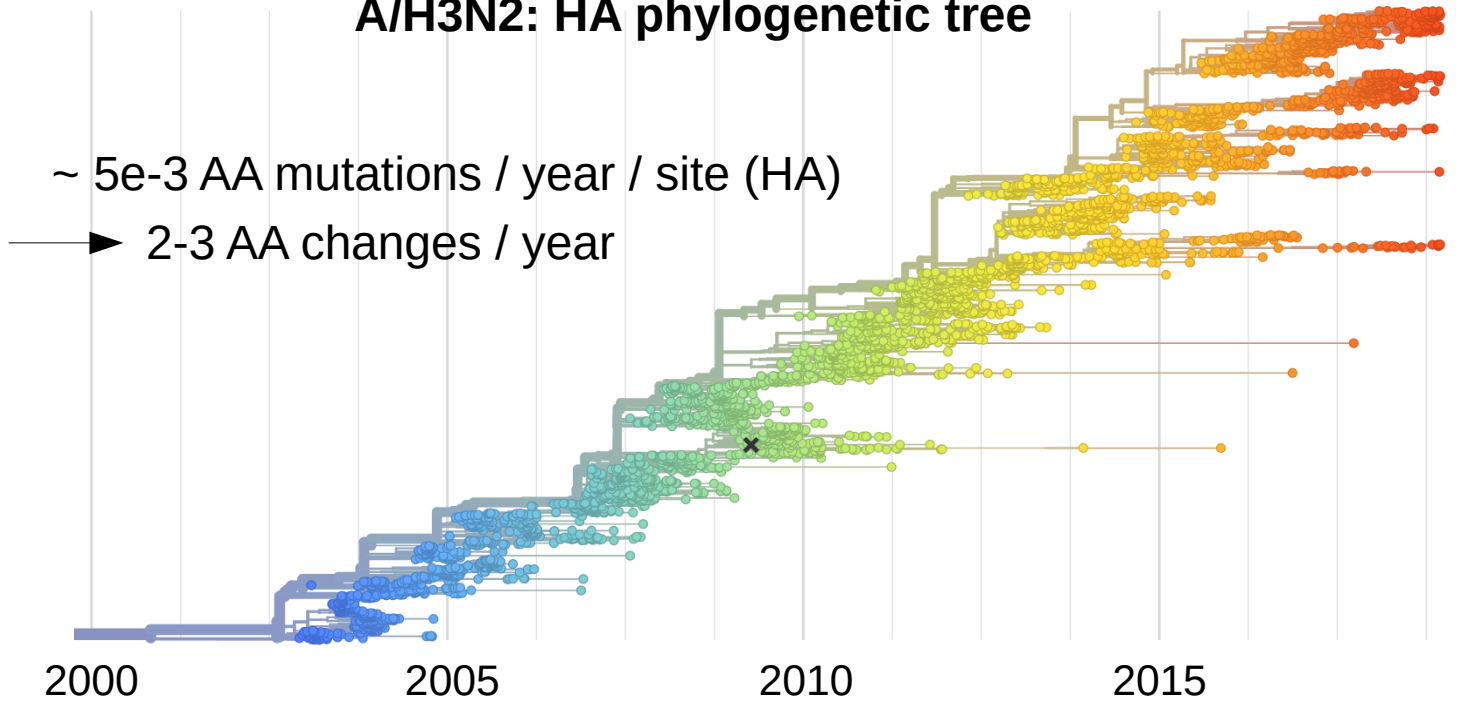
**BIOZENTRUM**  
Universität Basel  
The Center for  
Molecular Life Sciences

# Human seasonal influenza virus

~ hundreds of million cases / year → 5-10 % of humans  
In constant evolution (especially surface proteins HA & NA)

Generation time ~ 1 week  
→ Pop. size ~ 10e6 - 10e7

**A/H3N2: HA phylogenetic tree**



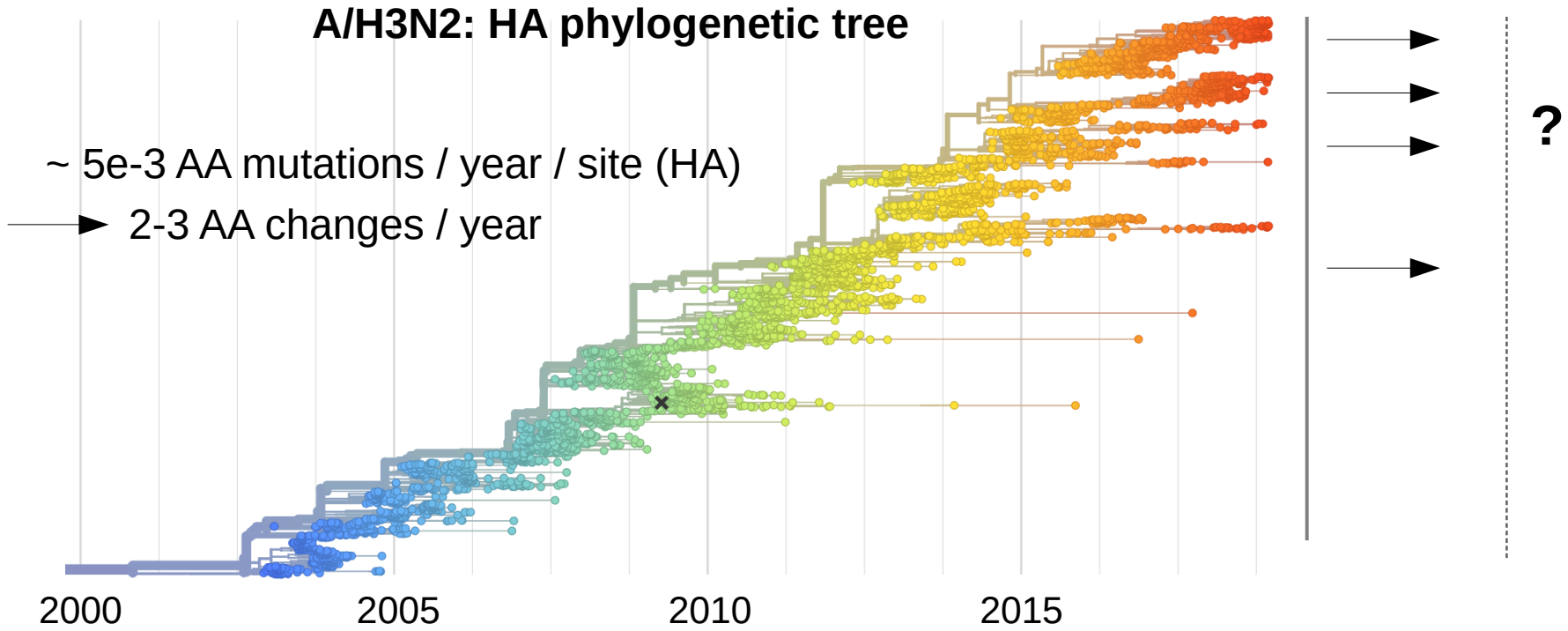
~ 5e-3 AA mutations / year / site (HA)  
→ 2-3 AA changes / year

Variability in the present population

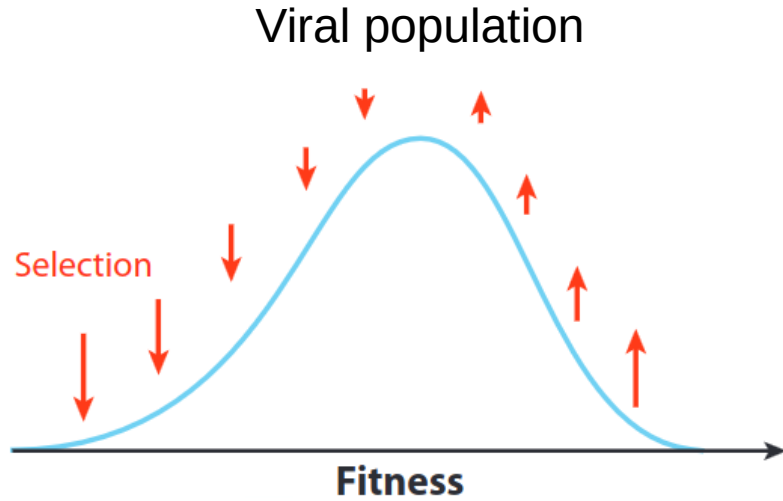
# Human seasonal influenza virus

~ hundreds of million cases / year → 5-10 % of humans  
In constant evolution (especially surface proteins HA & NA)

Generation time ~ 1 week  
→ Pop. size ~ 10e6 - 10e7



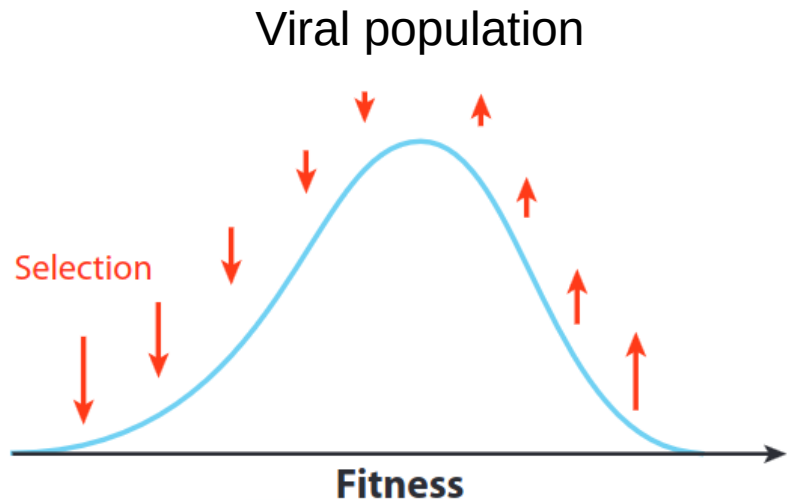
# Traditional approach: travelling fitness wave



- Mutations have a fixed fitness effect
- Fitness determines the fate of a mutant
- Extra-complexity: competition between mutants

→ **Some degree of predictability**

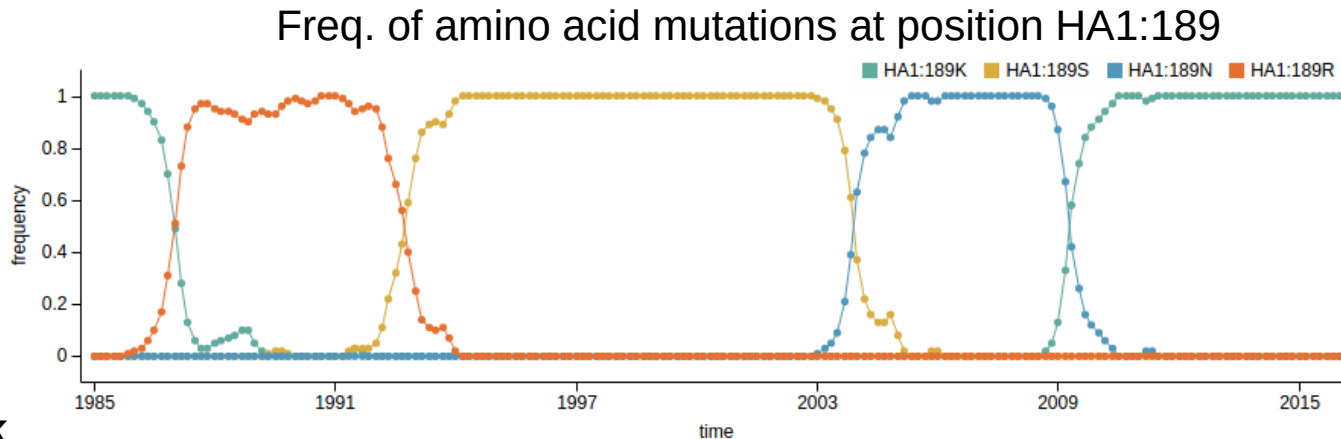
# Traditional approach: travelling fitness wave



- Mutations have a fixed fitness effect
- Fitness determines the fate of a mutant
- Extra-complexity: competition between mutants

→ **Some degree of predictability**

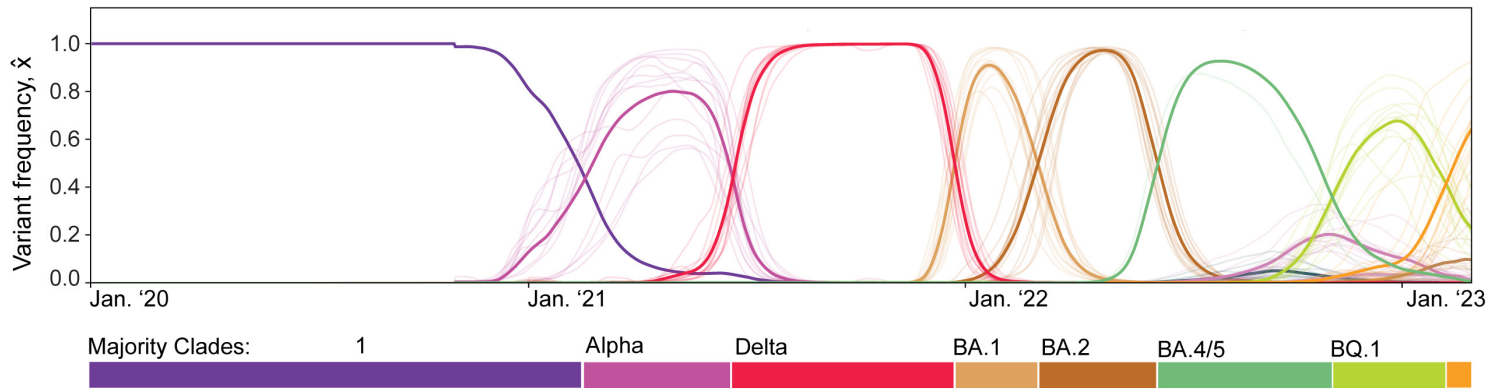
Selective sweeps



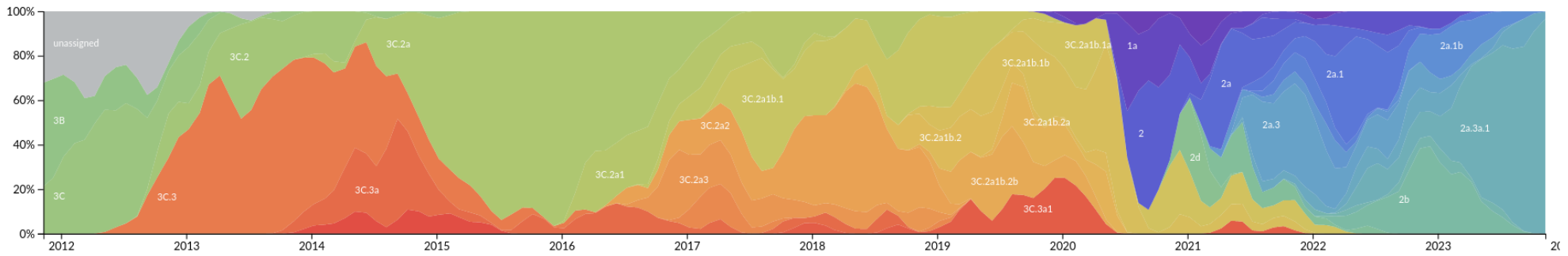
**Is this the right way to think about the process?**

# Clade frequencies: SARS-CoV-2 & Influenza

## Covid

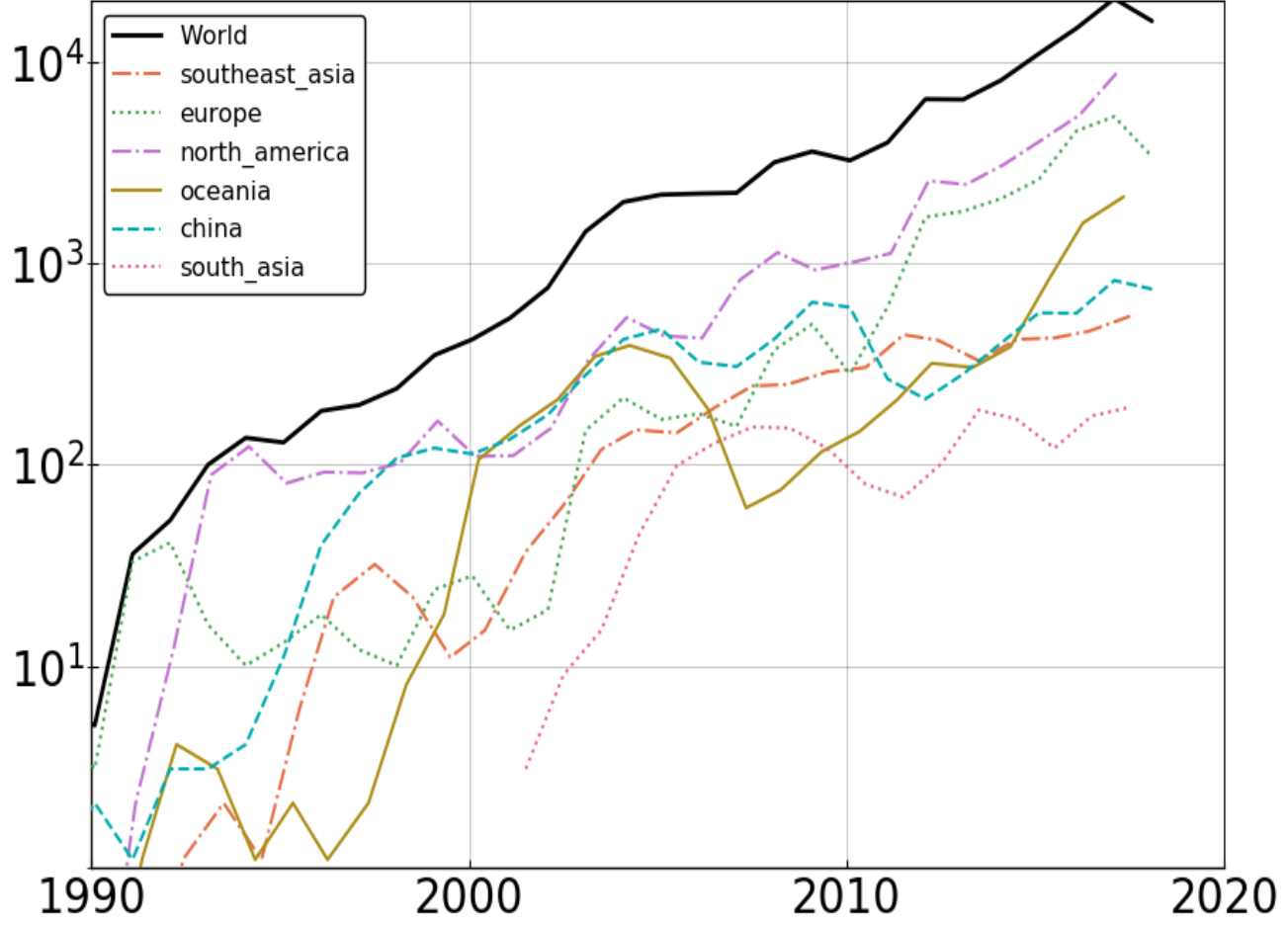


## H3N2 Influenza



# Data

# of seqs per year

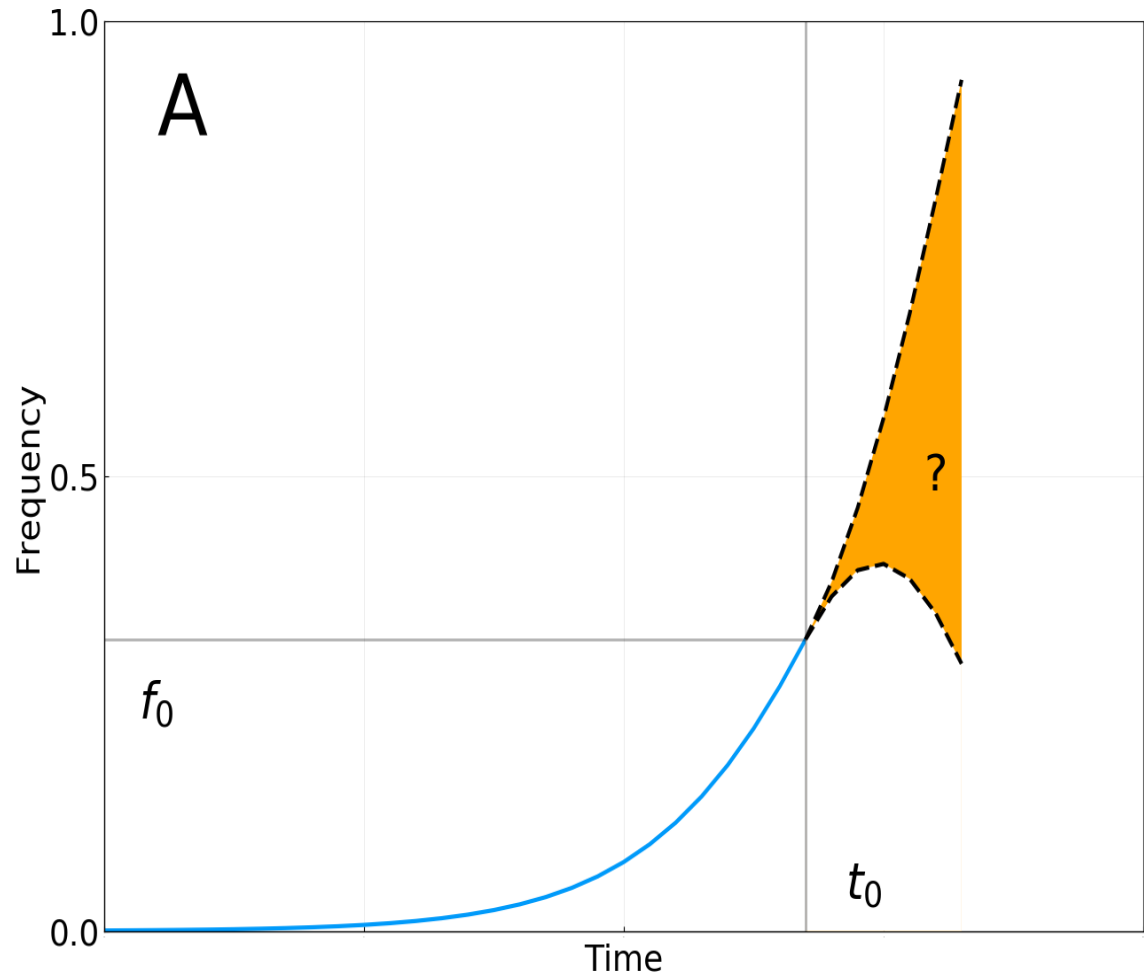


Time-binning of the past sequences by 1 month intervals

—▶ **Snapshots of the population**

—▶ **Frequency trajectories**

# Simple analysis: predictability of influenza



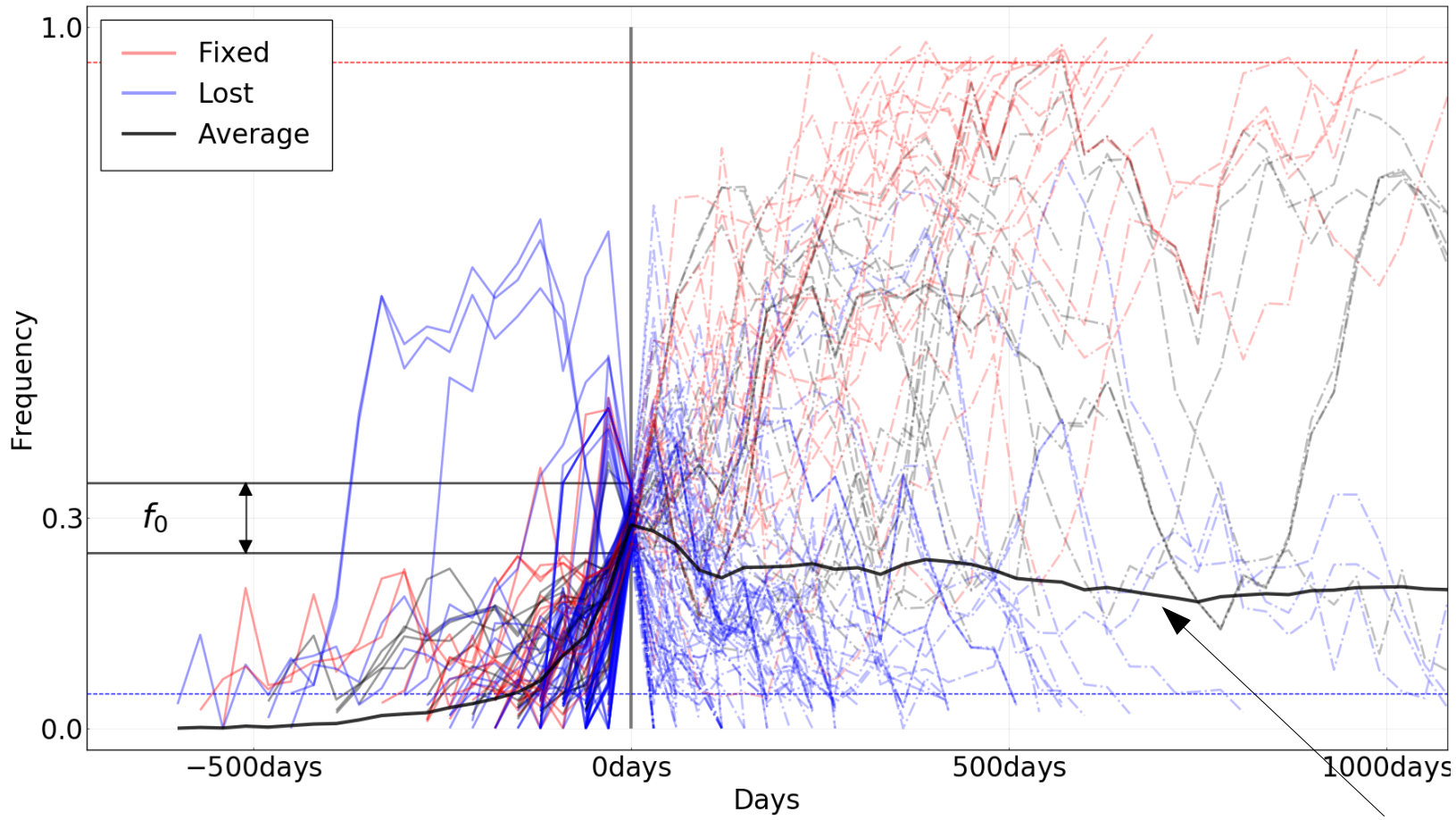
**Frequency trajectories of amino acid mutations**

Statistics from all sequences since 2000 (~20 years of data)



# Inertia of trajectories

Influenza H3N2, HA protein

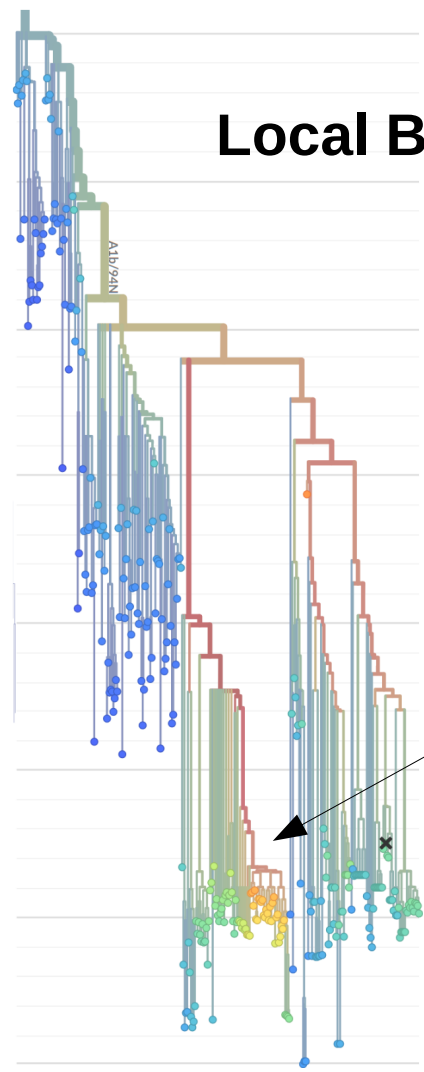
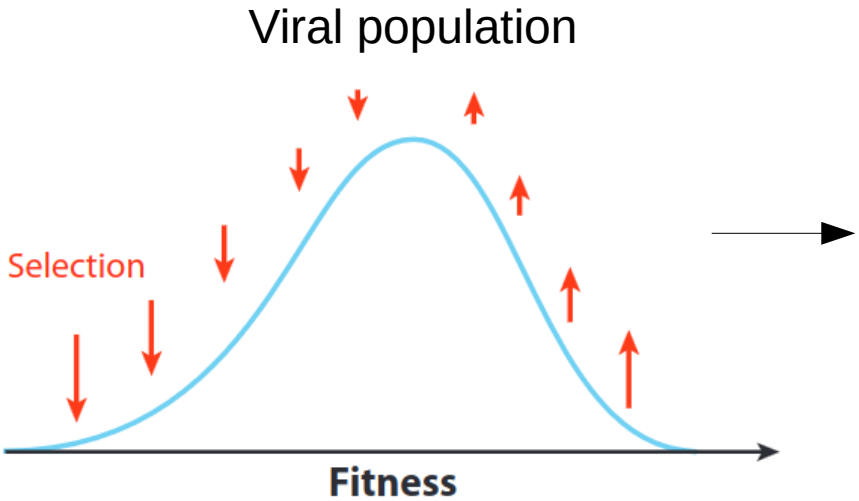


- Mutations:**
- Absent in the past
  - Seen around  $f_0=30\%$

► **No inertia**

**Average**

# Using a proxy for fitness



Neher, 2014

## Local Branching Index (LBI)

Strains with high fitness have more offsprings

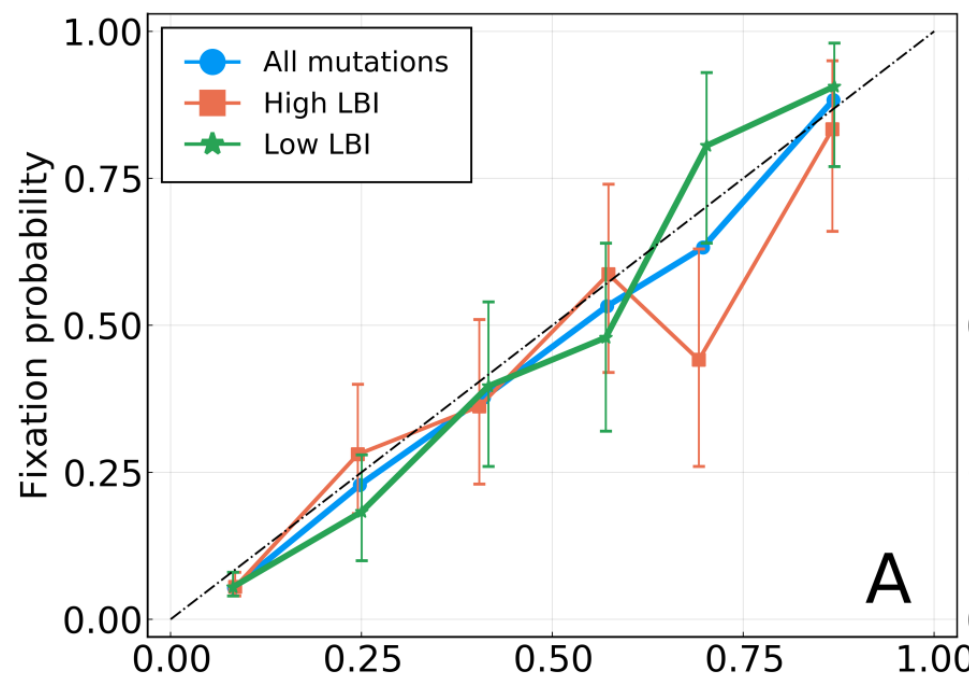


Dense branching



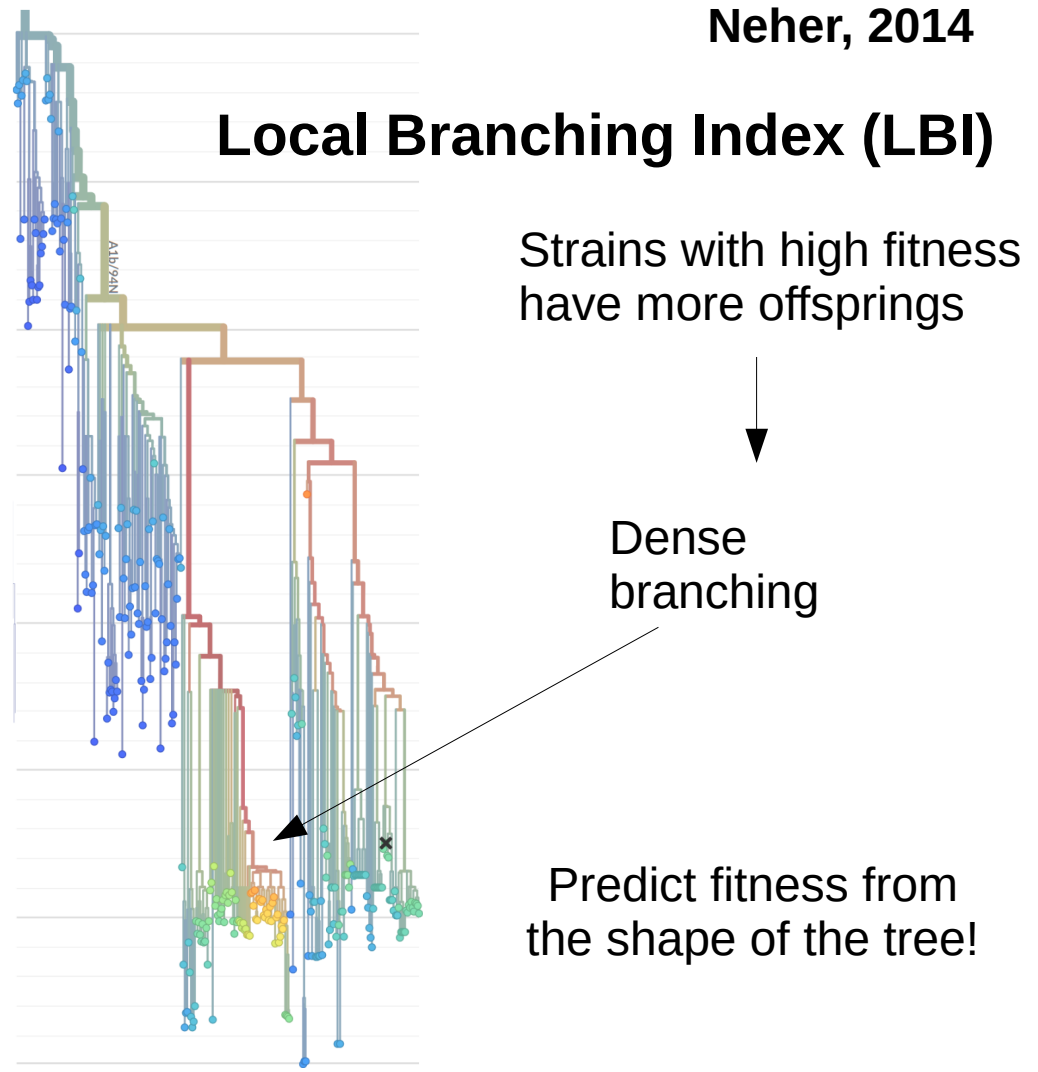
Predict fitness from the shape of the tree!

# Using a proxy for fitness



A

Neher, 2014



## Local Branching Index (LBI)

Strains with high fitness have more offsprings

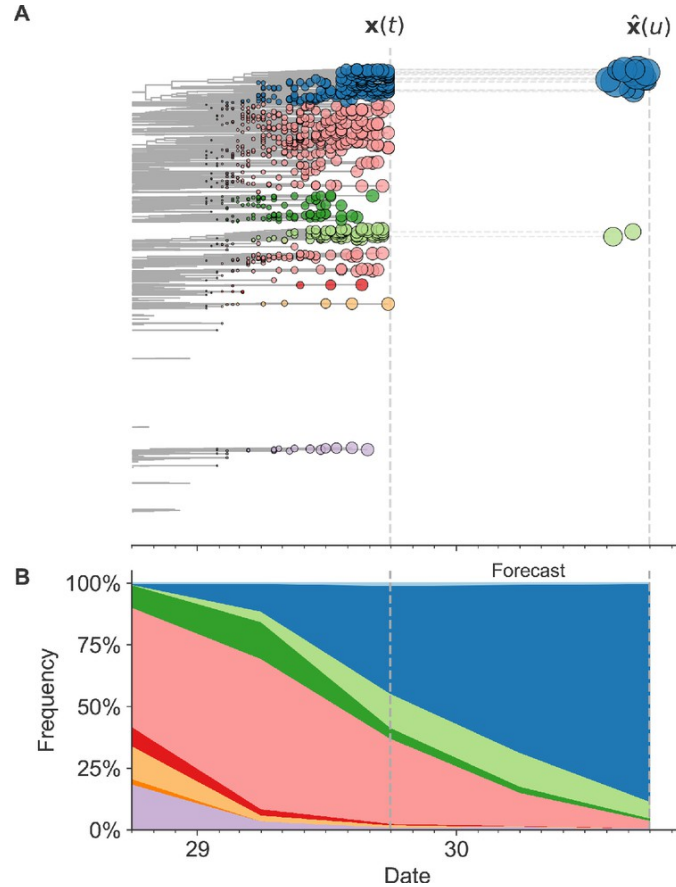
Dense branching

Predict fitness from the shape of the tree!

# Strain level forecast

Huddleston et. al. 2020

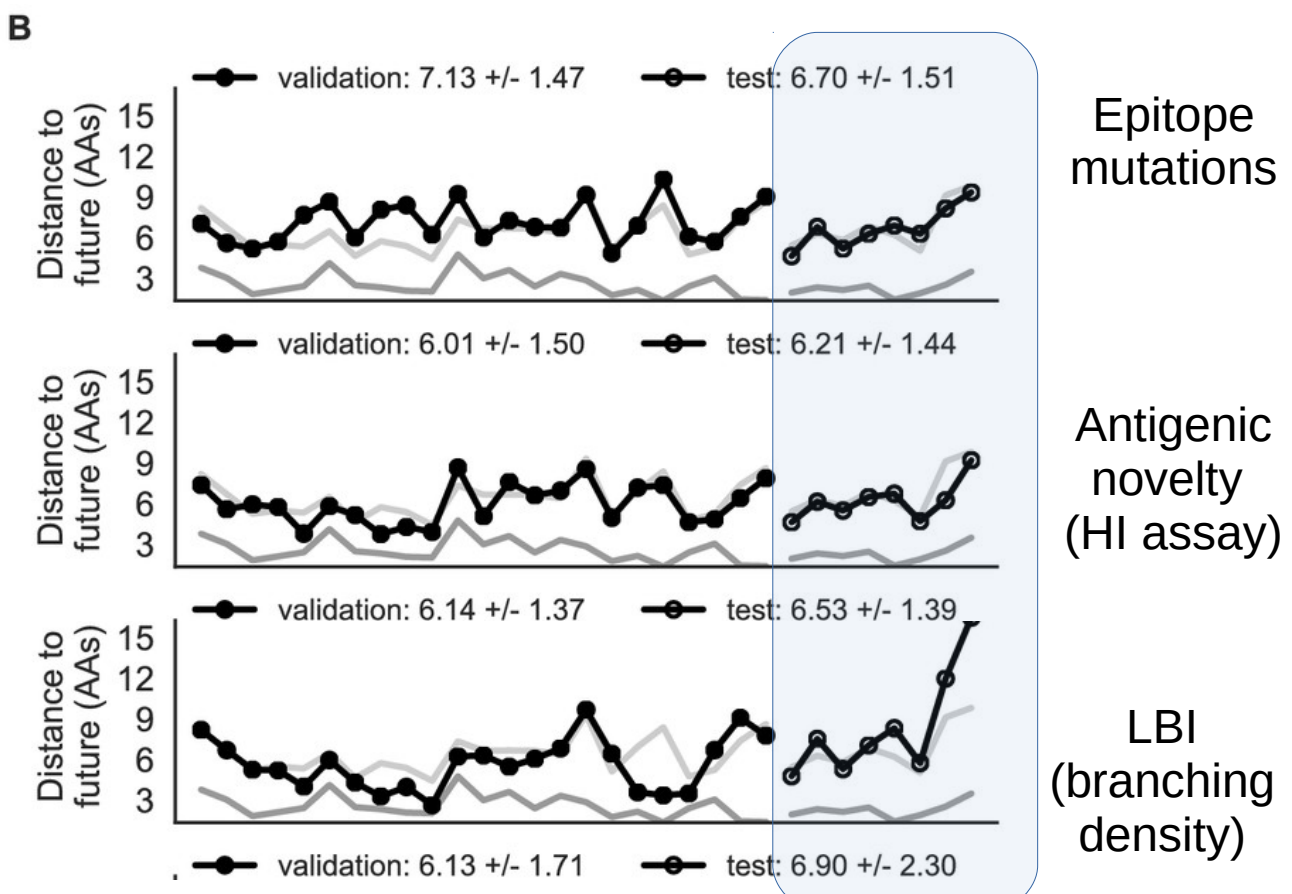
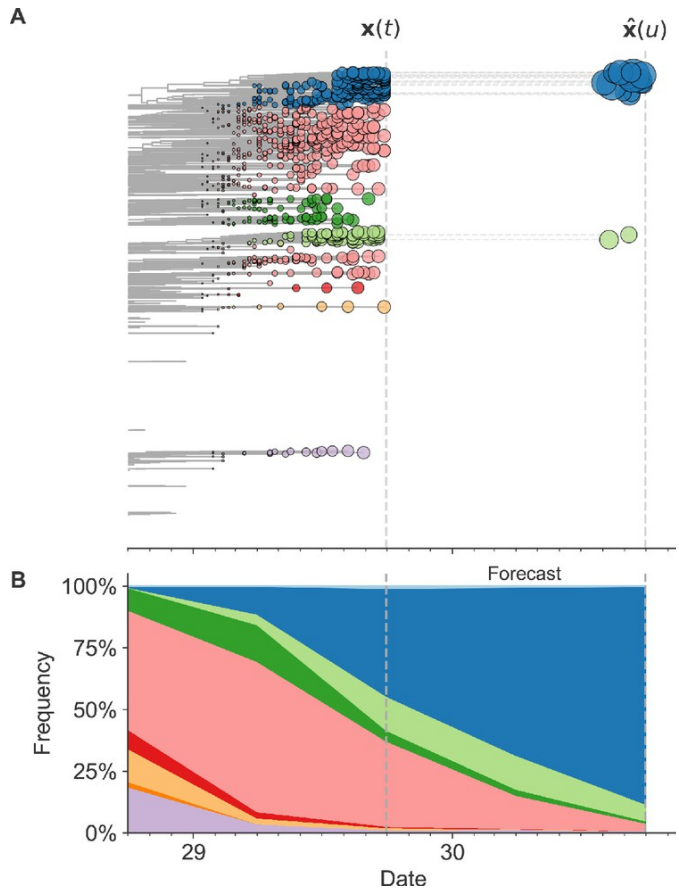
- Predict a fitness for each strain --> fit model to data (LBI, antigenic novelty, ...)
- Forecast future population 1 year ahead



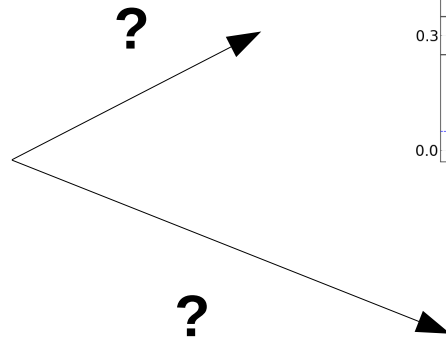
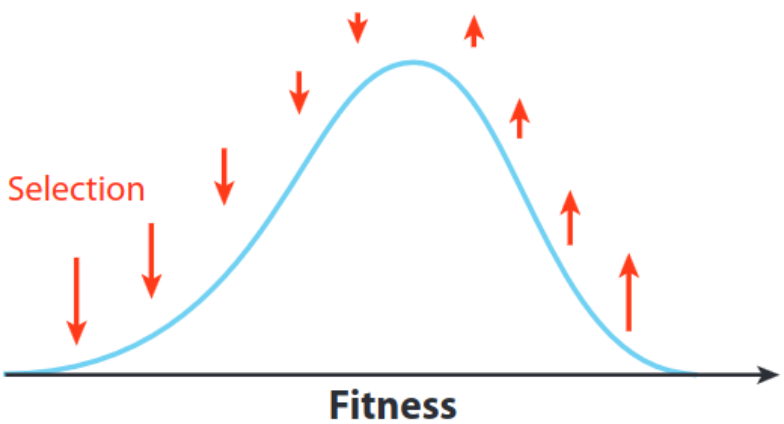
# Strain level forecast

Huddleston et. al. 2020

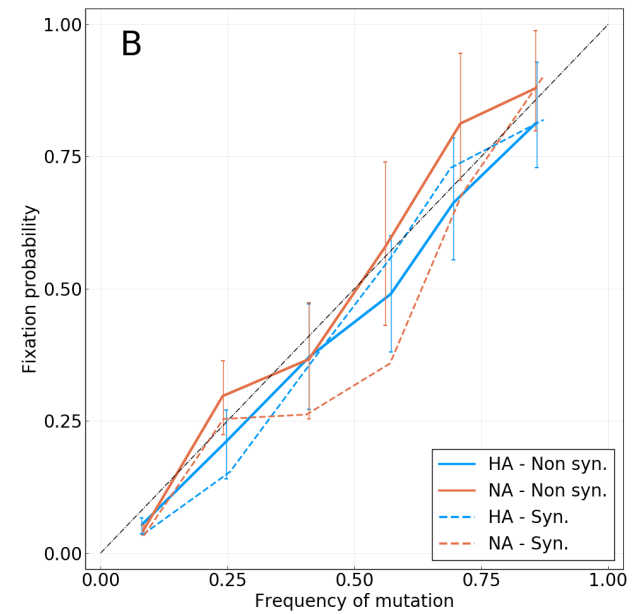
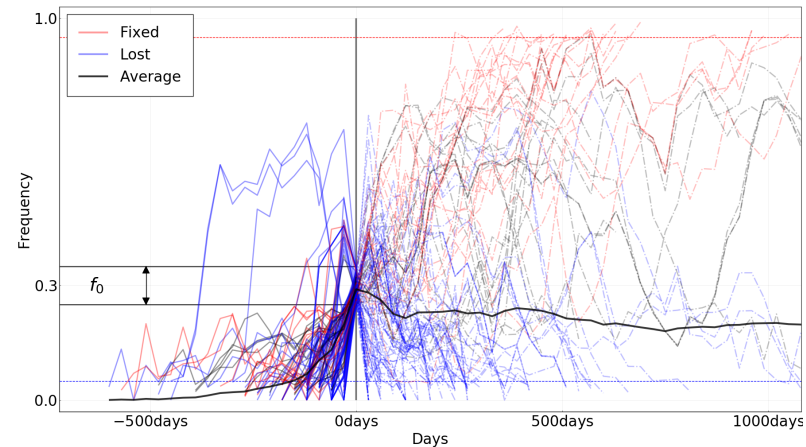
- Predict a fitness for each strain --> fit model to data (LBI, antigenic novelty, ...)
- Forecast future population 1 year ahead



# Limited predictability



**Qualitative difference between model and observations**



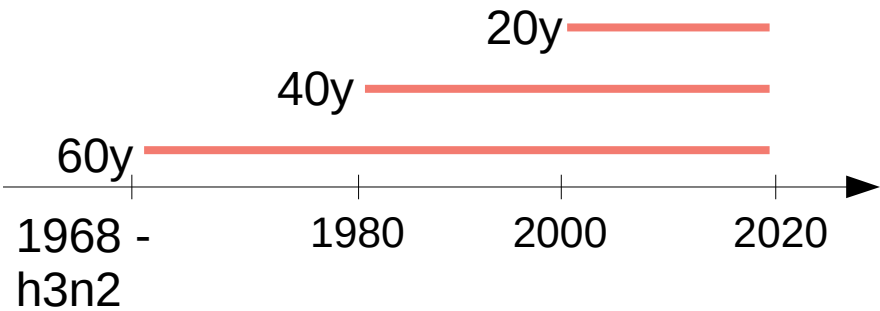
# Adaptive immunity and expiring fitness

## Adaptation is driven by immunity

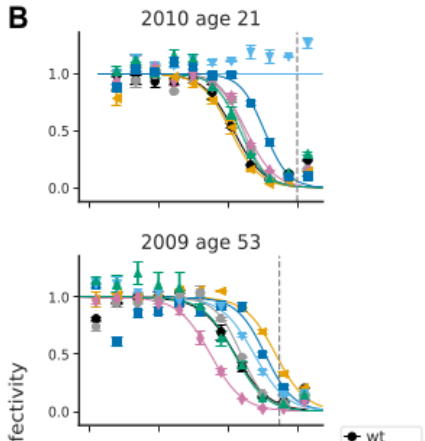
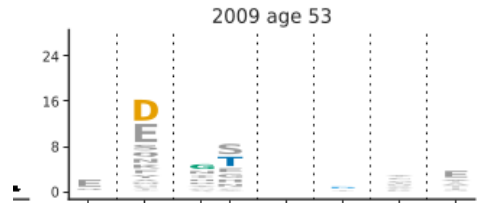
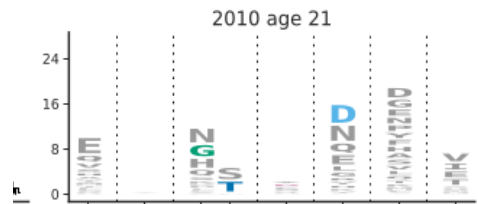
- Most adaptive mutations escape immunity
- They only escape a fraction of the host population

## Adaptive immunity of hosts

- fitness advantage expires before fixation
- “ecology”: organisms shape their environment



→ **Model hosts' immunity explicitly**



[Lee et. al., eLife 2019]

# Susceptible – Infected model

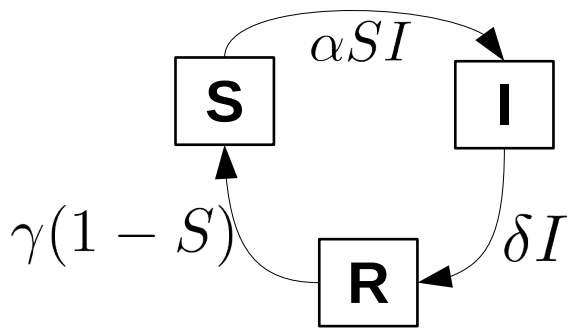
One viral strain

$$\dot{S} = -\alpha SI + \gamma(1 - S)$$

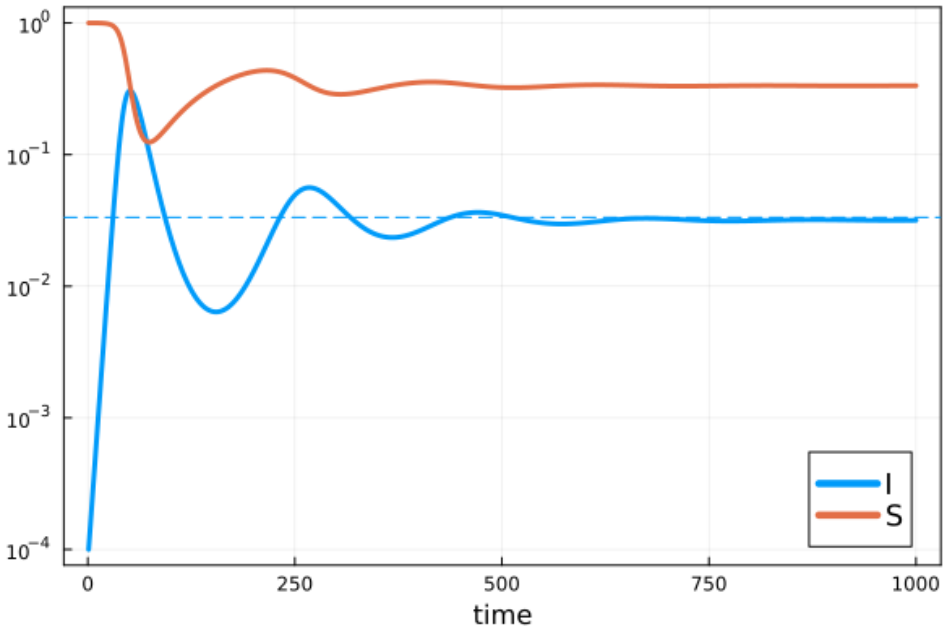
$$\dot{I} = \alpha SI - \delta I$$

Host population

$$S + I + R = 1$$



Initially naive pop.



$$I_{eq} = \frac{\gamma}{\delta} (1 - \delta \alpha^{-1})$$



# SI model

## Two strains: wild-type and mutant

$$\dot{S}^a = -\alpha S^a \sum_{b \in \{wt, m\}} K^{ab} I^b + \gamma(1 - S^a)$$

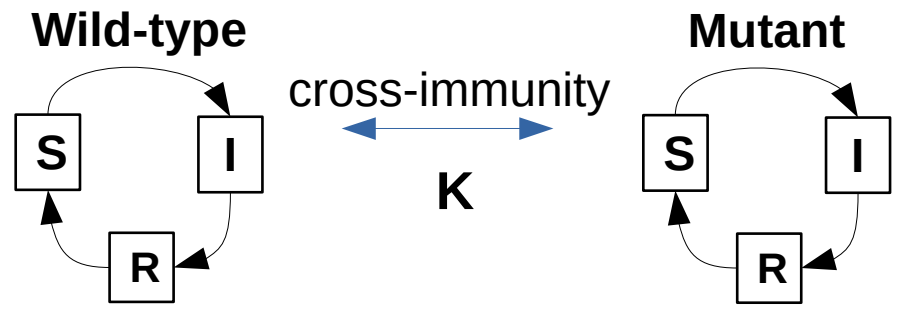
$$\dot{I}^a = \alpha S^a I^a - \delta I^a$$



### Equilibrium

$$S = \frac{\delta}{\alpha}$$

$$I = \frac{\gamma}{\delta} (1 - \delta/\alpha) \mathbf{K}^{-1} \vec{1}$$



### cross-immunity

$$K = \begin{bmatrix} 1 & b \\ f & 1 \end{bmatrix}$$

immune to wt after infection by mut.

immune to mut after infection by wt

- No notion of **“fitness”**
- Mutant is **different**, not better

# Partial sweep with SI model

( $b = 0.7, f = 0.8$ )

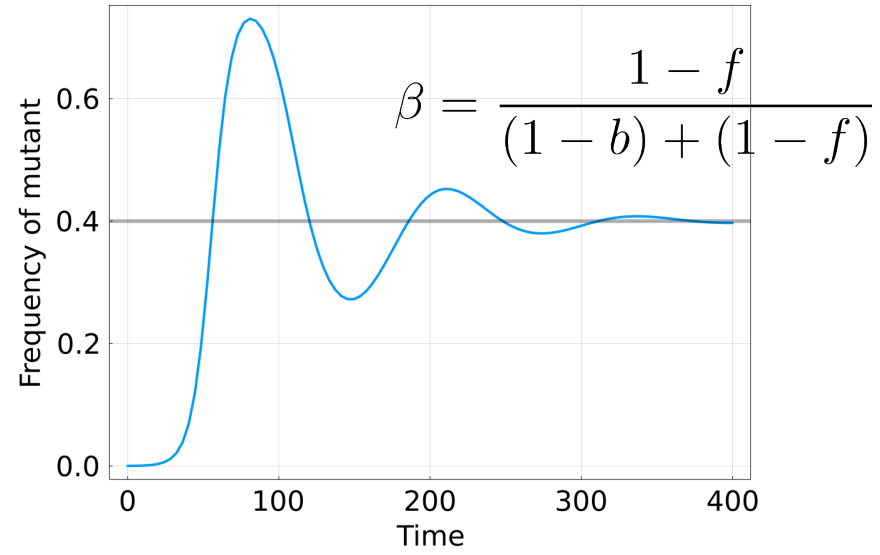
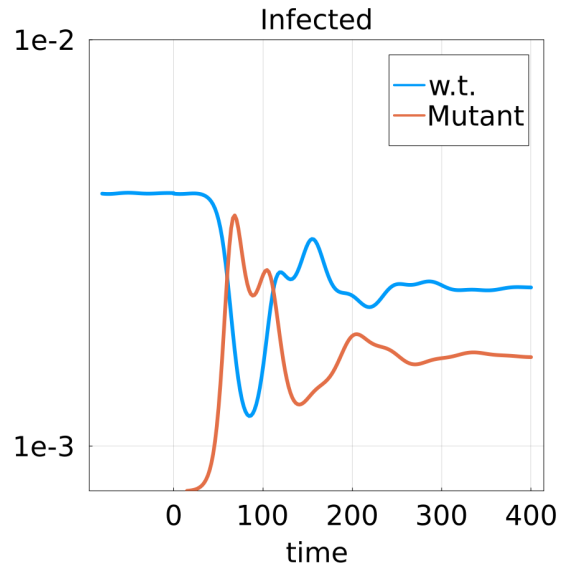
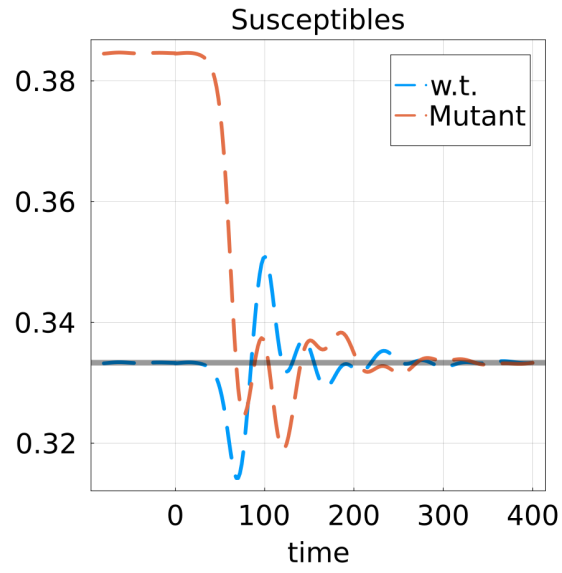
One **wild-type** virus, one **mutant**

$$\longrightarrow K = \begin{bmatrix} 1 & b \\ f & 1 \end{bmatrix}$$

**cross-immunity**

Initially, **no mutant**  $I^{mut} = 0$

- At  $t = 0$ , introduce mutant



# Immune groups: regularize the dynamics

Hosts split into  $\mathbf{M}$  immune groups, with different cross-immunity

Mutant is adaptive in group 1

$$K_1 = \begin{bmatrix} 1 & b \\ f & 1 \end{bmatrix}$$

Very close to w.t. in other groups

$$K_{i>1} = \begin{bmatrix} 1 & 1 - \varepsilon \\ 1 - \varepsilon & 1 \end{bmatrix}$$

# Immune groups: regularize the dynamics

Hosts split into  $\mathbf{M}$  immune groups, with different cross-immunity

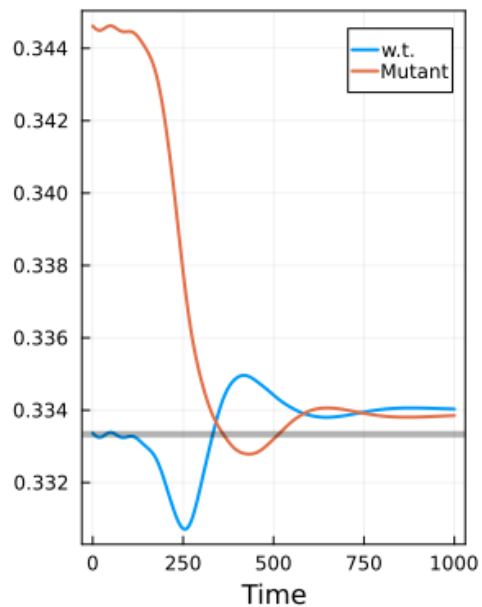
Mutant is adaptive in group 1

$$K_1 = \begin{bmatrix} 1 & b \\ f & 1 \end{bmatrix}$$

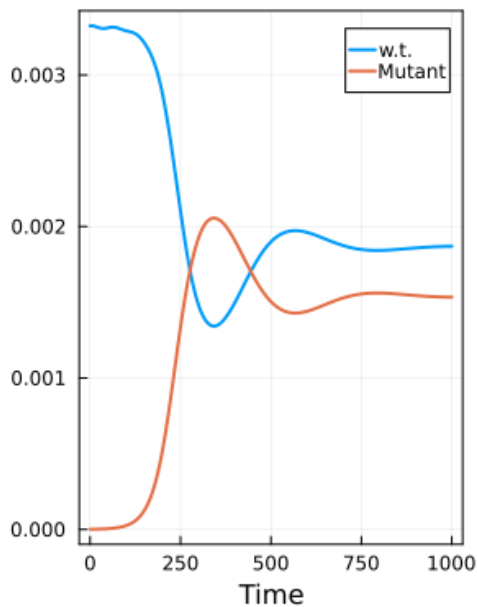
Very close to w.t. in other groups

$$K_{i>1} = \begin{bmatrix} 1 & 1 - \varepsilon \\ 1 - \varepsilon & 1 \end{bmatrix}$$

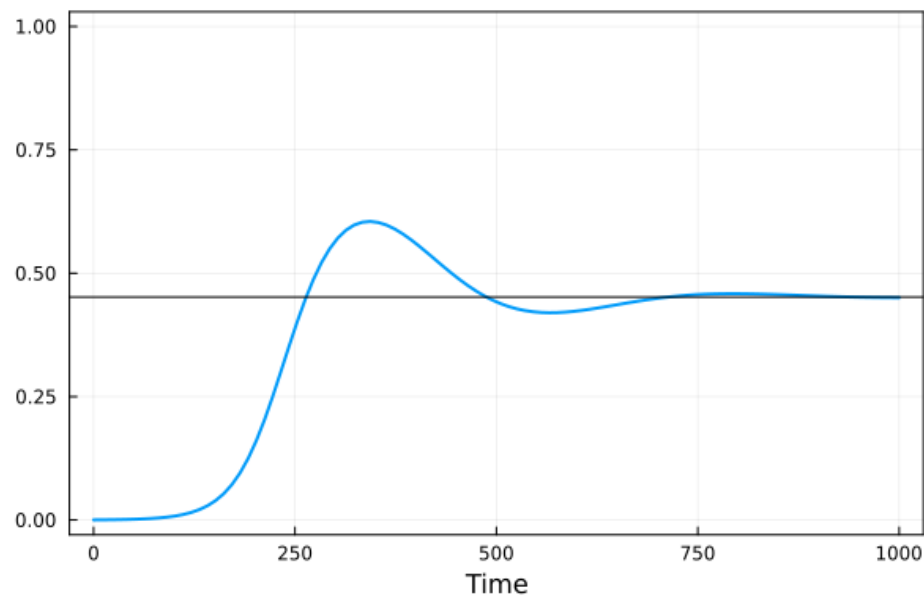
Susceptibles (all groups)



Infectious (all groups)



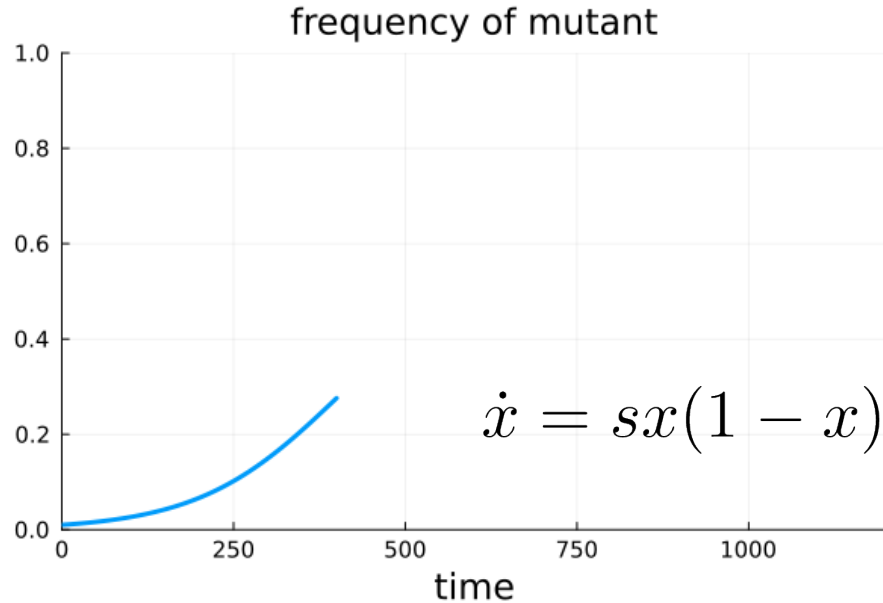
Mutant frequency



# Adaptive immunity and expiring fitness

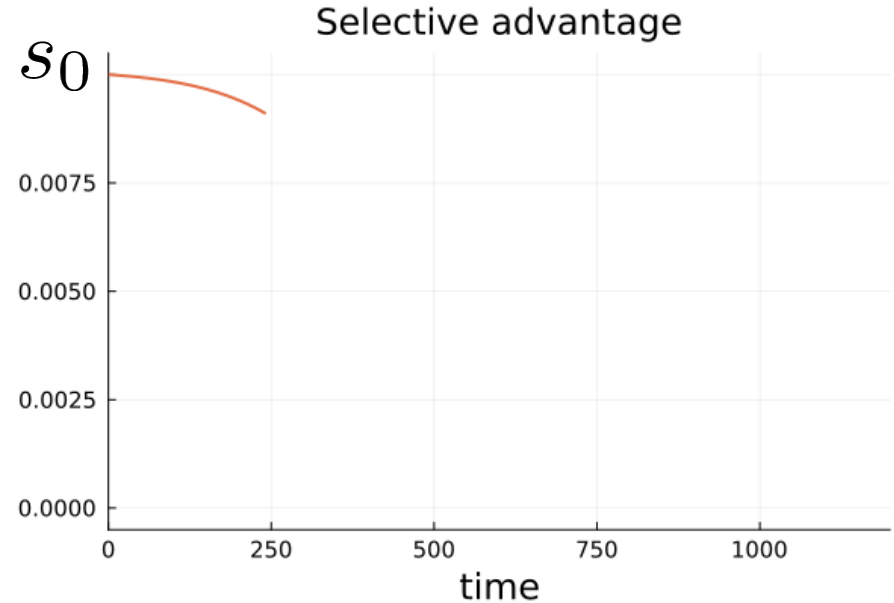
## Adaptation is driven by immunity

- Most adaptive mutations escape immunity
- They only escape a fraction of the host population



## Adaptive immunity of hosts

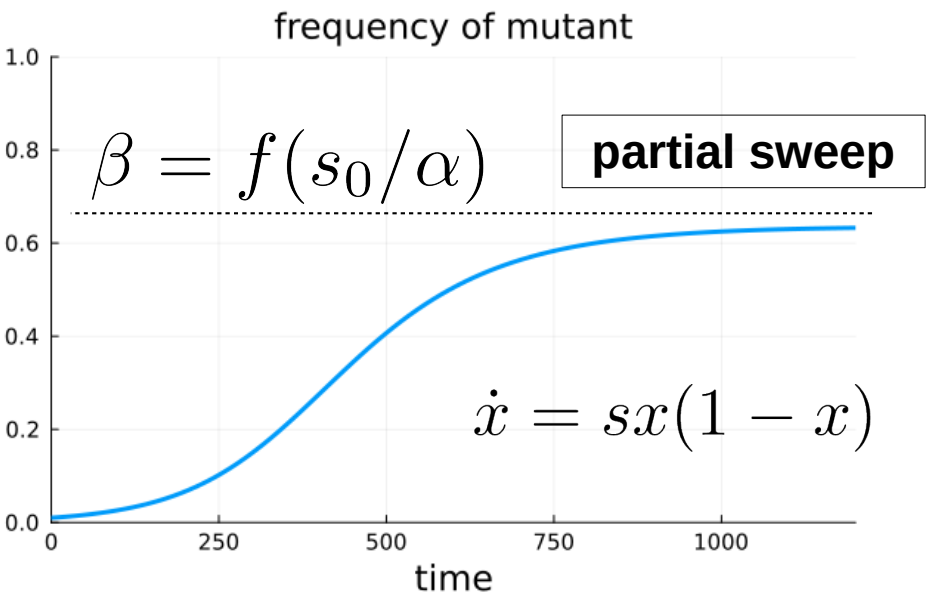
- fitness advantage expires before fixation
- “ecology”: organisms shape their environment



# Adaptive immunity and expiring fitness

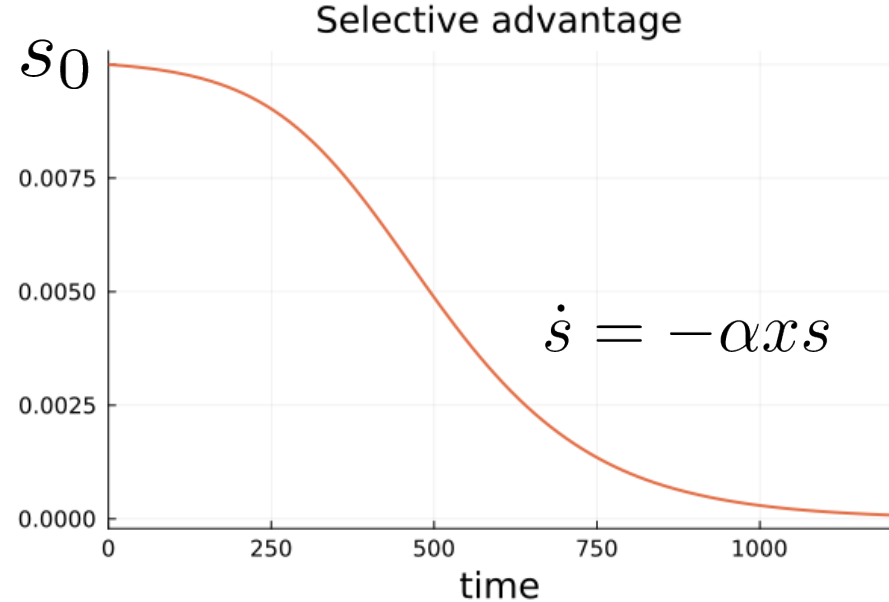
## Adaptation is driven by immunity

- Most adaptive mutations escape immunity
- They only escape a fraction of the host population



## Adaptive immunity of hosts

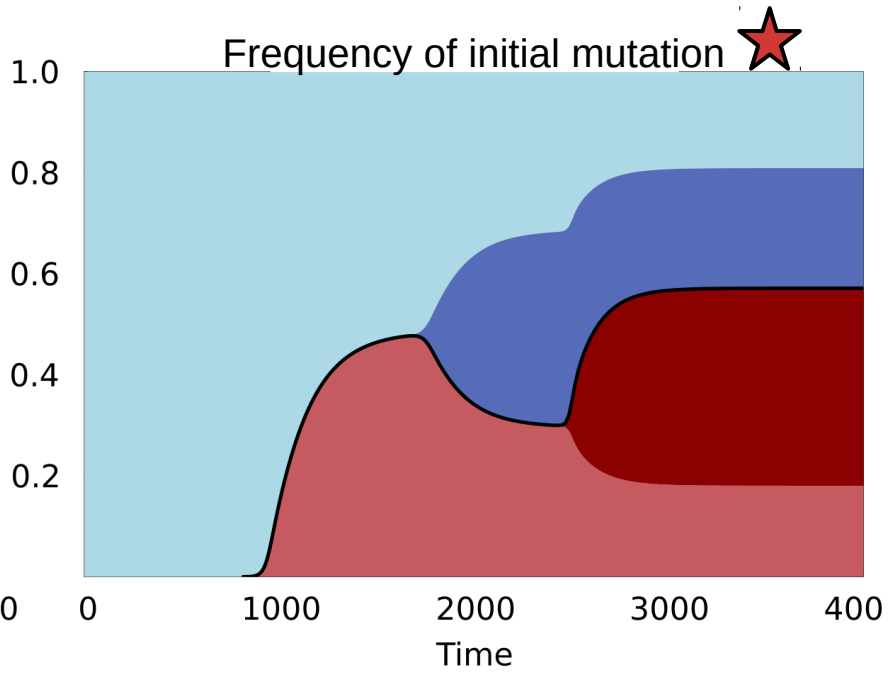
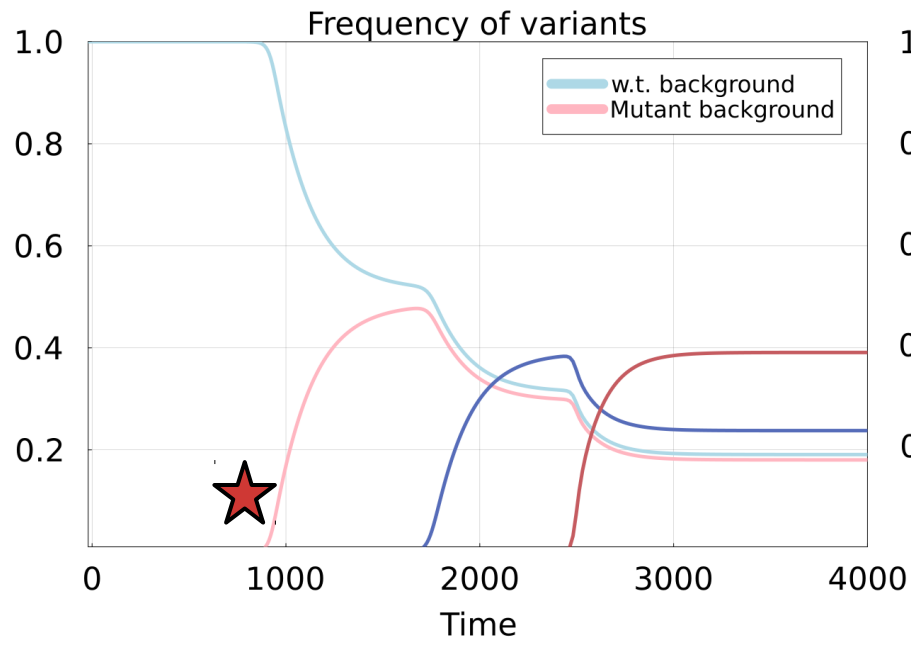
- fitness advantage expires before fixation
- “ecology”: organisms shape their environment



Used to model covid evolutionary dynamics [Meijers et. al., Cell 2023]

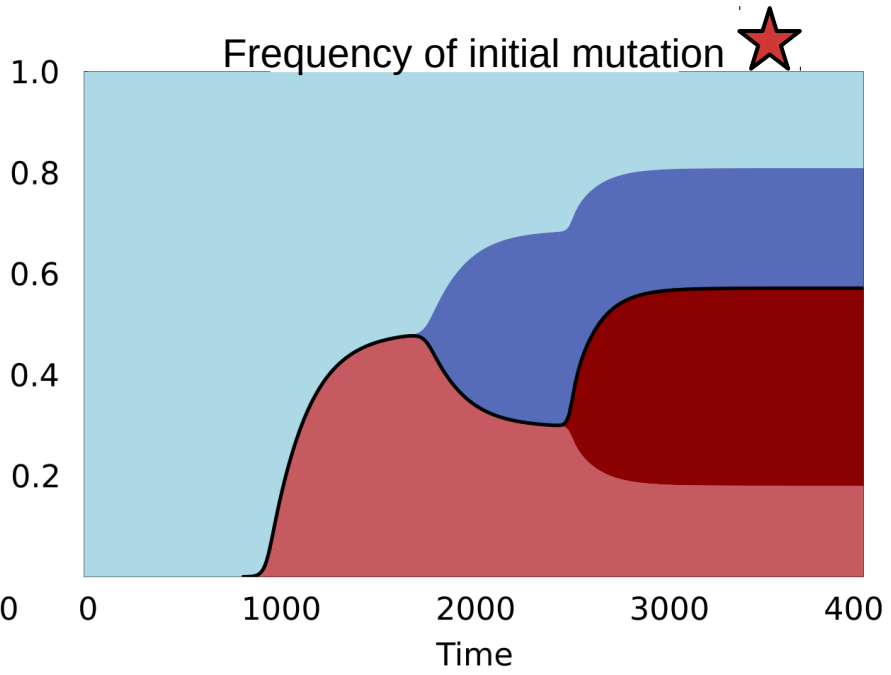
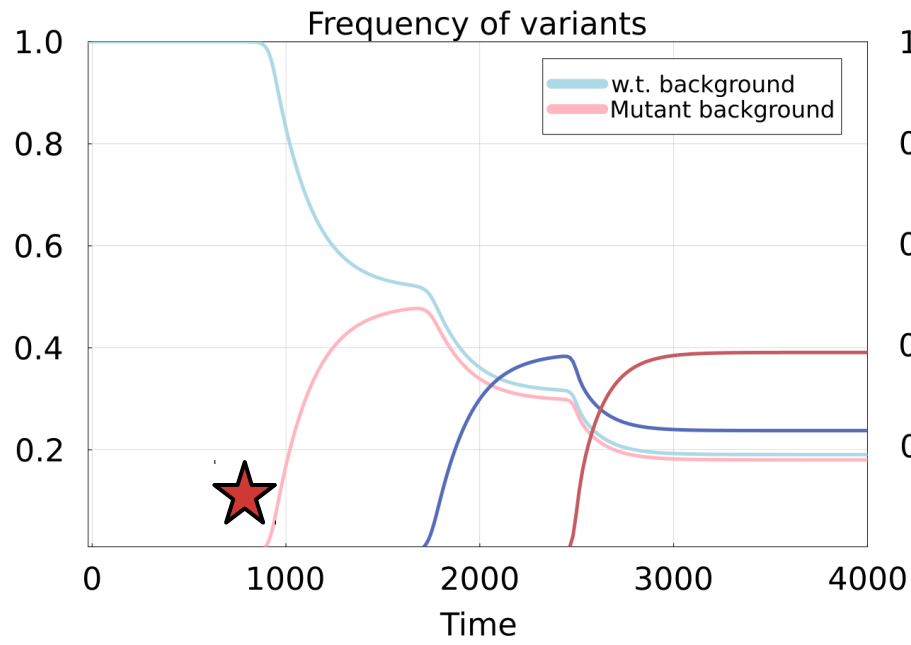
# Evolution with partial sweeps

New variants appear at rate  $\rho$



# Evolution with partial sweeps

New variants appear at rate  $\rho$



Limit case: non overlapping partial sweeps

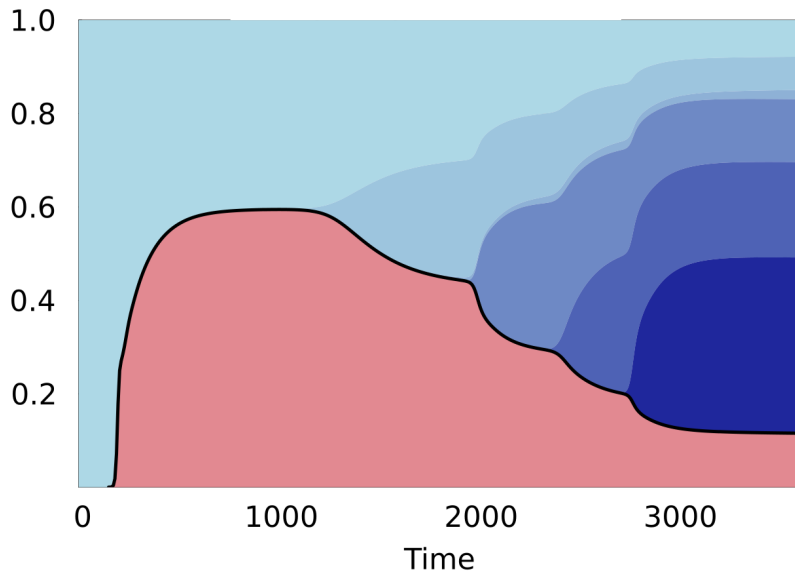
$$x_{t+1} = x_t + \begin{cases} \beta(1 - x_t) & \text{with prob. } x_t, \\ -\beta x_t & \text{with prob. } 1 - x_t, \end{cases}$$



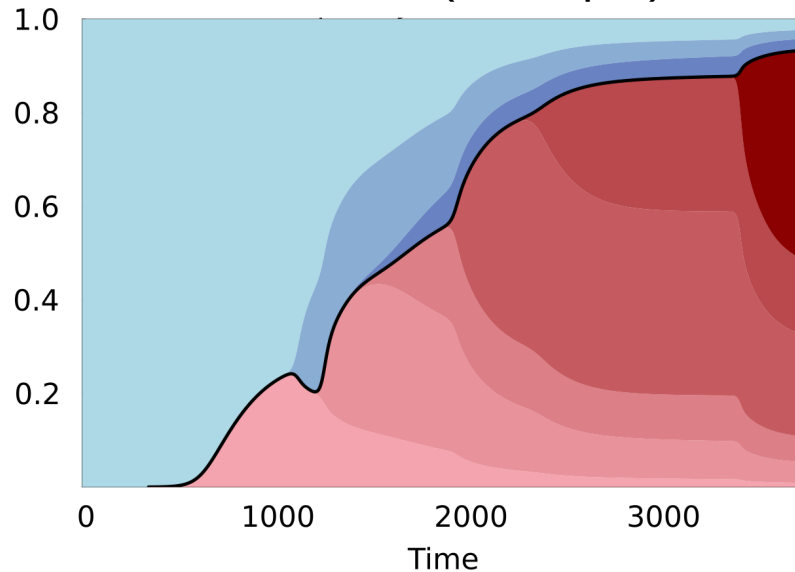
# Evolution with partial sweeps

New variants appear at rate  $\rho$

Mutant vanishes (“sweeps”)



Mutant fixes (“sweeps”)



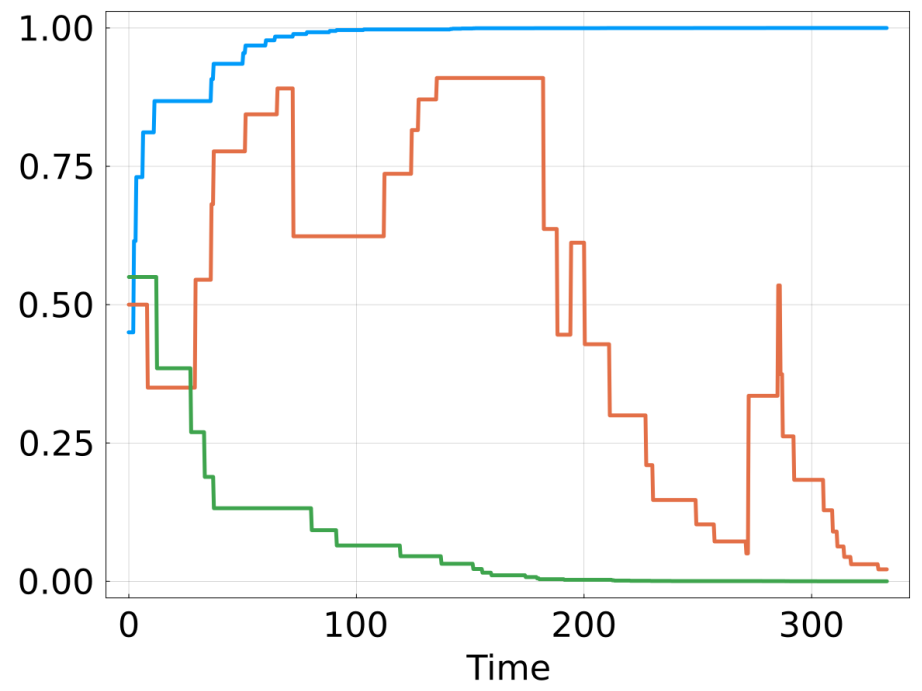
Limit case: non overlapping partial sweeps

$$x_{t+1} = x_t + \begin{cases} \beta(1 - x_t) & \text{with prob. } x_t, \\ -\beta x_t & \text{with prob. } 1 - x_t, \end{cases}$$

# Evolution with partial sweeps

$$x_{t+1} = x_t + \begin{cases} \beta(1 - x_t) & \text{with prob. } x_t, \\ -\beta x_t & \text{with prob. } 1 - x_t, \end{cases}$$

$\langle \Delta x \rangle = 0 \rightarrow$  **Low predictability**

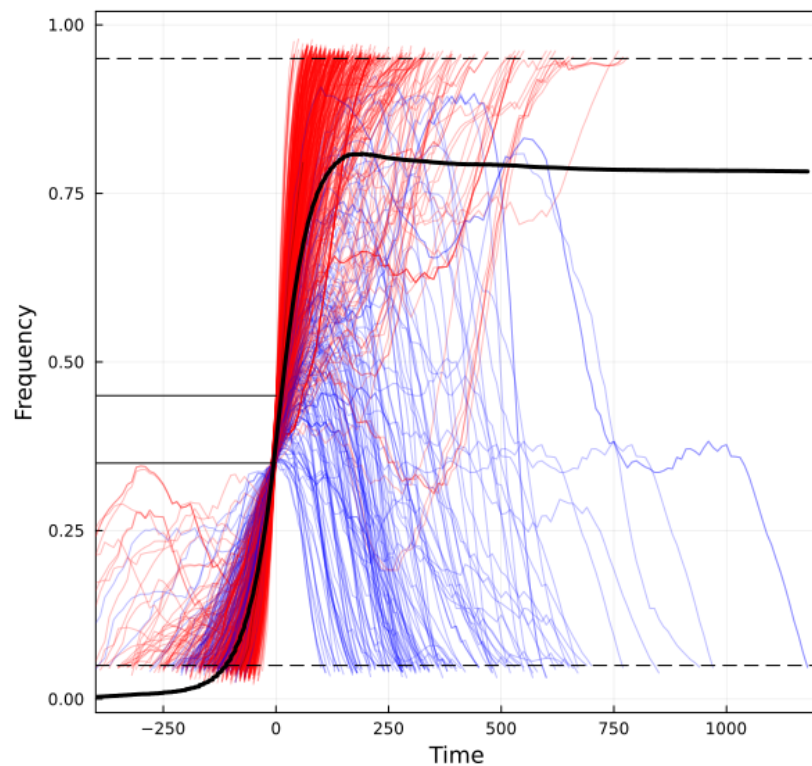


# Inertia of frequency trajectories

Simulation of a viral population (Wright-Fisher style)

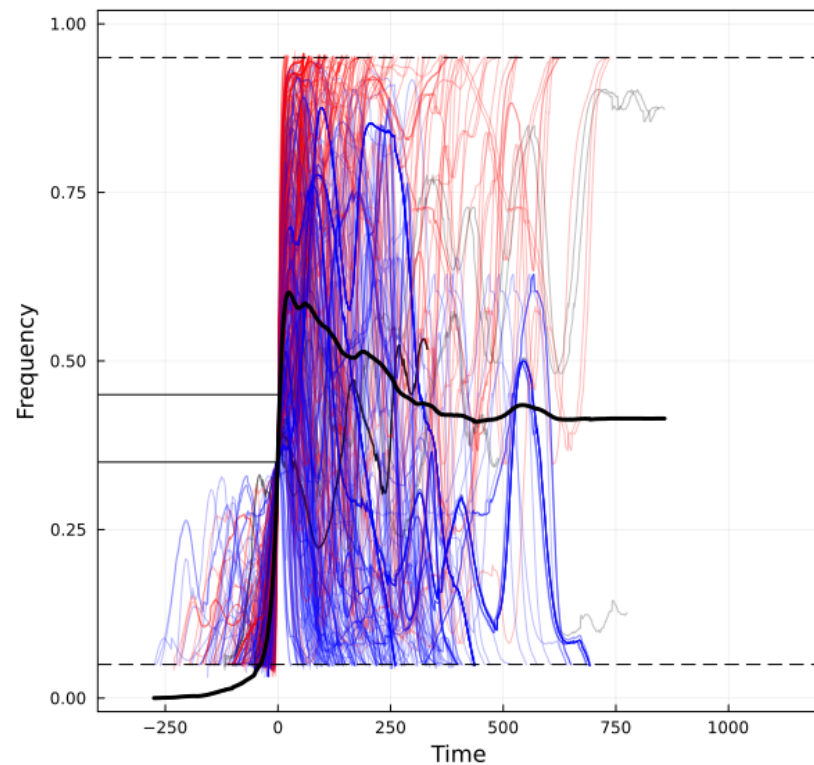
## Fixed fitness effects

$s = 0.01$ ,  $\rho = 0.083$ ,  $s/\rho = 0.12$ ,  $\Delta t = 10$



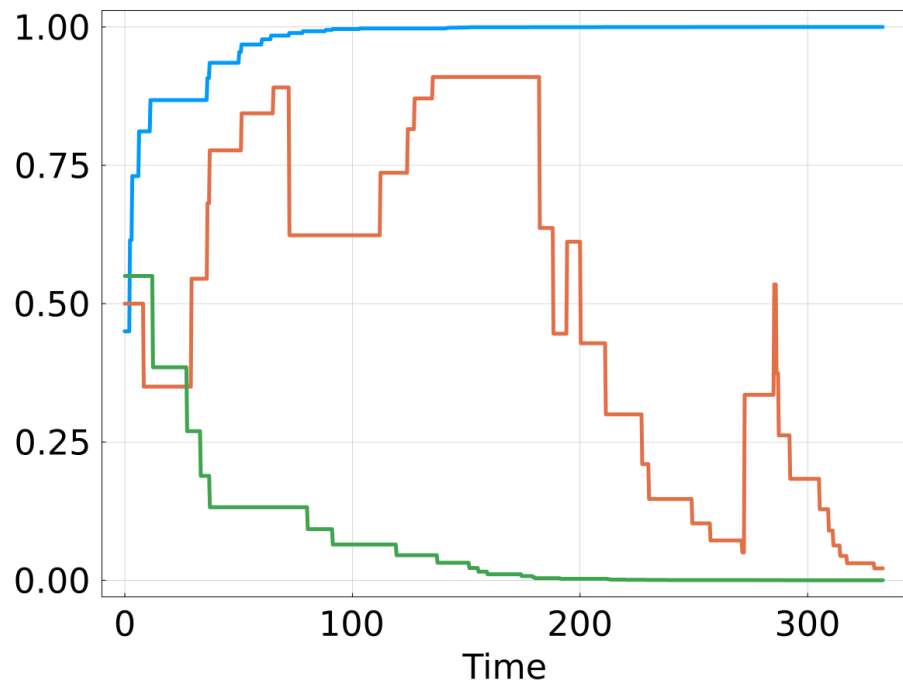
## Expiring fitness

$\langle \beta \rangle = 0.24$ ,  $\alpha = 0.1$ ,  $\Delta t = 1$



# Evolution with partial sweeps

$$x_{t+1} = x_t + \begin{cases} \beta(1 - x_t) & \text{with prob. } x_t, \\ -\beta x_t & \text{with prob. } 1 - x_t, \end{cases}$$



$$\langle \Delta x \rangle = 0 \quad \longrightarrow \quad \text{Low predictability}$$

## Timescales

$$\langle \Delta x^2 \rangle = \rho \langle \beta^2 \rangle x(1 - x)$$

## Analogy with neutral evolution

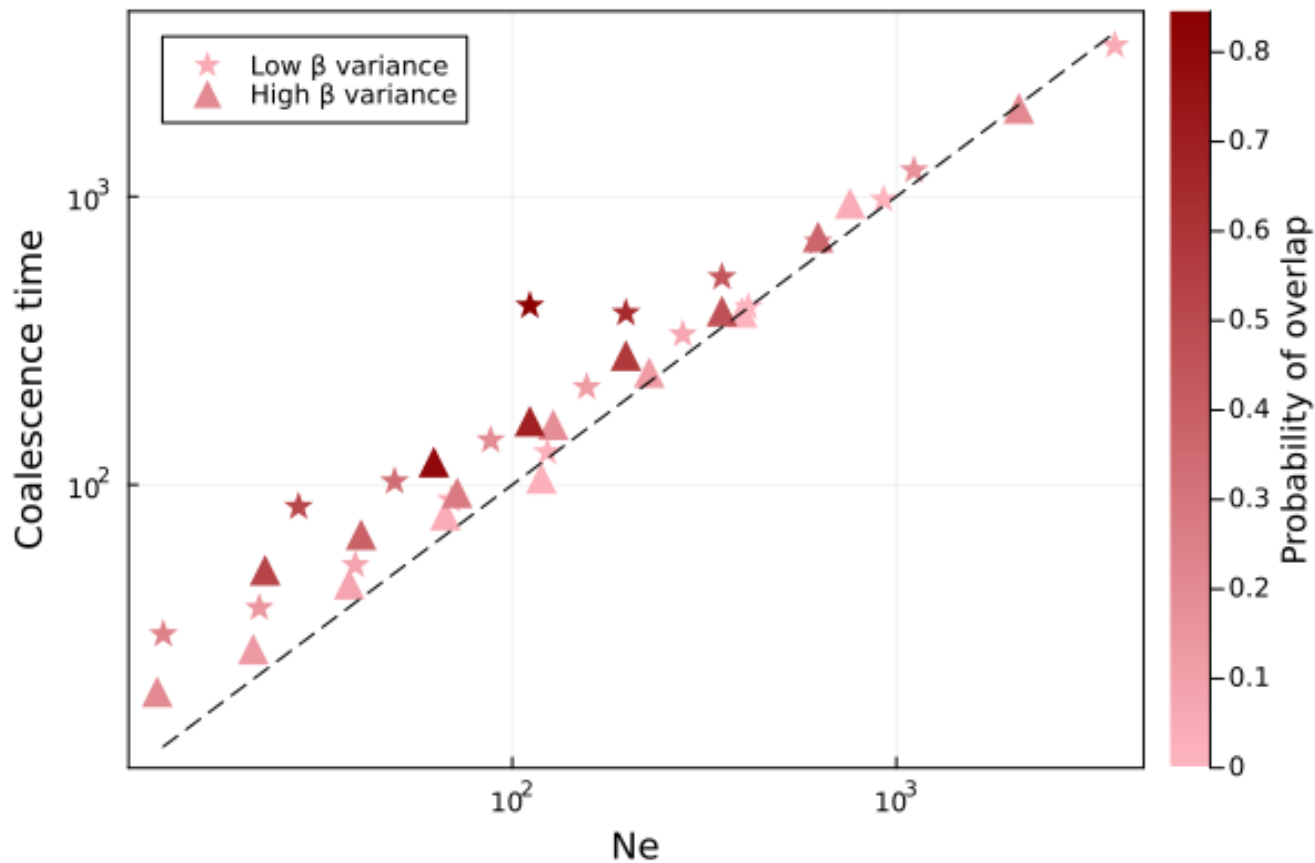
$$\text{timescale } N_e^{-1} = T^{-1} \sim \rho \langle \beta^2 \rangle$$

$$\rho^{-1} \sim 15 \quad (\sim 4 \text{ partial sweeps/year})$$

$$\beta \sim 0.3 \quad \longrightarrow \quad T \sim 150 \sim 3y$$

$$\text{H3N2 influenza} \quad \longrightarrow \quad T_{MRC A} \sim 6y$$

# Time to Most Recent Common Ancestor



$$N_e^{-1} = T^{-1} \sim \rho \langle \beta^2 \rangle$$

Simulate a population (Wright-Fisher) with

- different distributions for  $\beta$
- different values for  $\rho$

# Summary

H3N2Influenza:

- **Predictability** of evolution is surprisingly low
- Travelling wave models

## Partial sweeps

- **Adaptation** of host immunity
- Fitness advantage of mutant **expires** before full sweep
- Modeled by a multi-strain Susceptible-Infected model

Evolution with **partial sweeps**

- Driven by fitness
- Low predictability
- Qualitatively closer to data

**Thank you!**

# Summary

H3N2Influenza:

- **Predictability** of evolution is surprisingly low
- Qualitatively different from models

## Partial sweeps

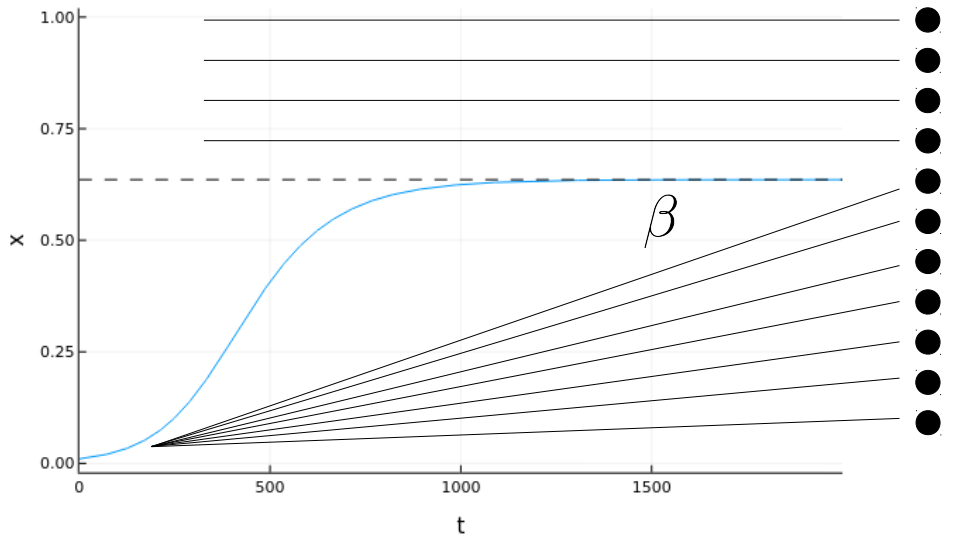
- **Adaptation** of host immunity
- Fitness advantage of mutant **expires** before full sweep

Evolution with **partial sweeps**

- Driven by fitness
- Low predictability
- Qualitatively closer to data

**Thank you!**

# Shape of the phylogeny: multiple mergers

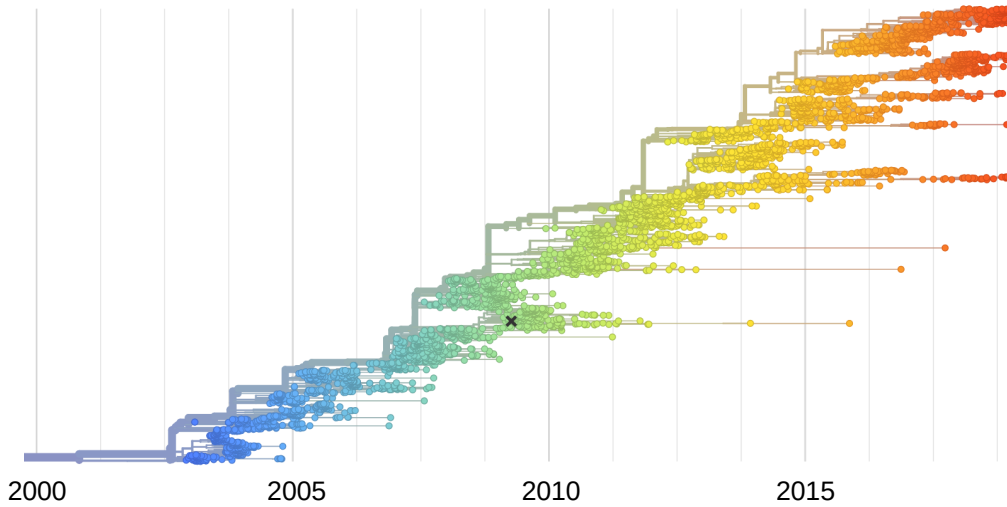


- Multiple mergers
- Ladder-like phylogenies

Rate of **k**-coalescence among **n** lineages

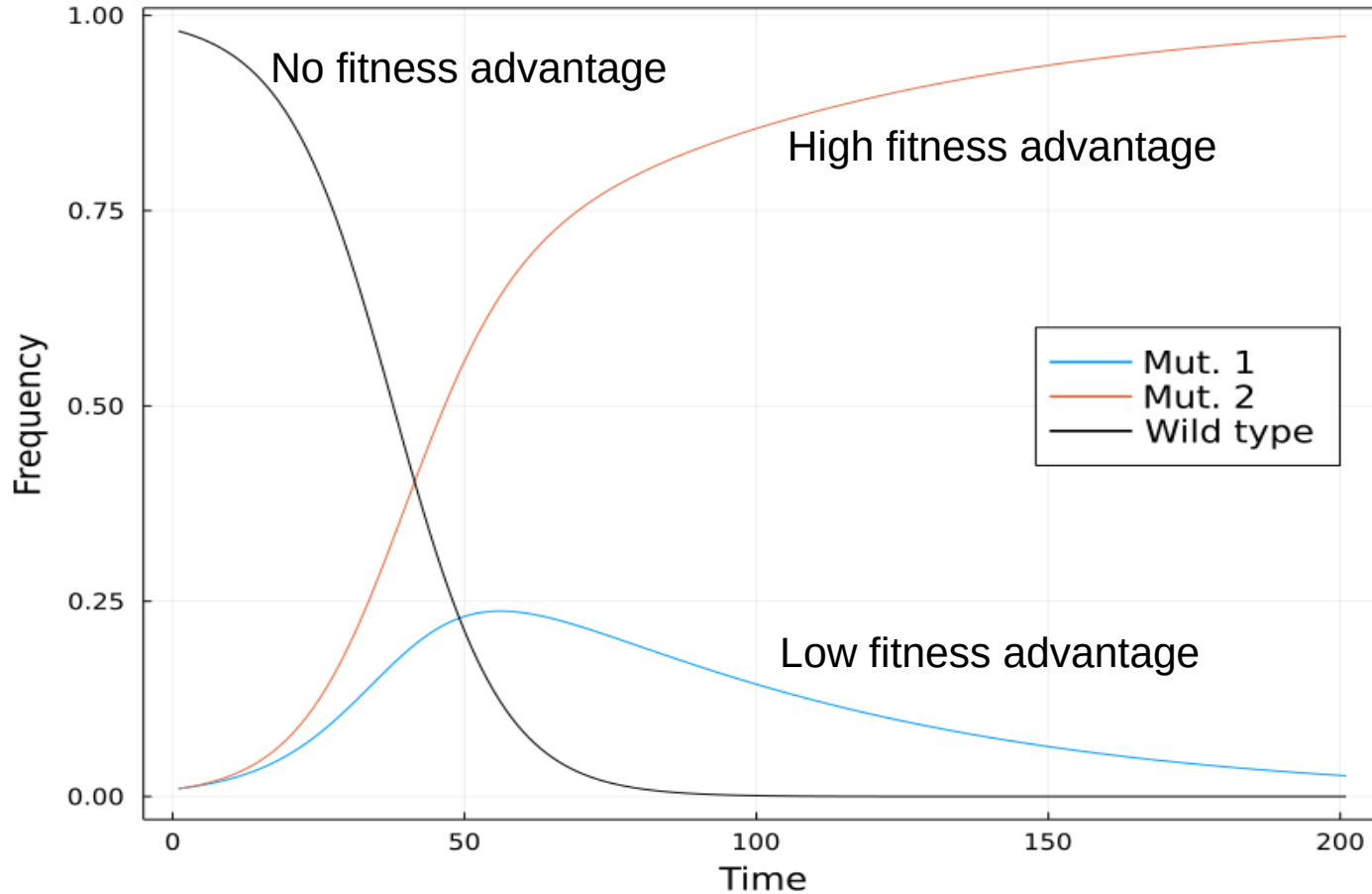
$$\Lambda_n(k) = \rho \binom{n}{k} \beta^k (1 - \beta)^{n-k}$$

Lambda coalescent





# Is this expected? Clonal interference



Adaptive mutations  
appearing on  
different individuals



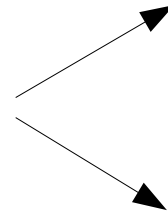
**Competition**

# Genetic linkage: toy model

Simulate a population

Simple fitness landscape  $f(\vec{s}) = \sum_{i=1}^L h_i s_i$

Change the fitness landscape periodically

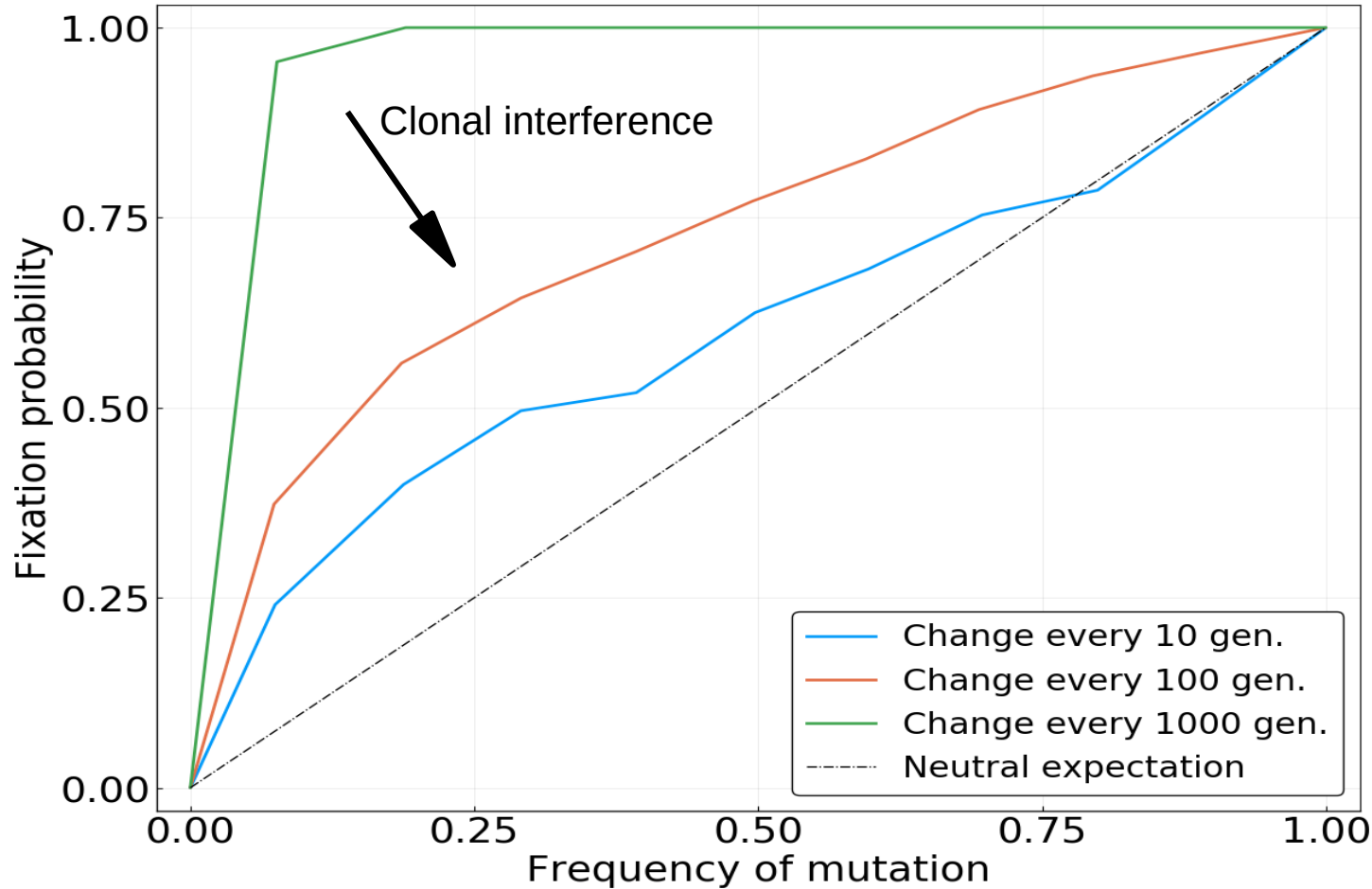


Slow rate of change  
**Clean sweeps**

High rate of change  
**Clonal interference**

# Genetic linkage: toy model

Sweep time ~400 generations  
(vs ~3 years for flu)



**It's hard to mimic neutrality!**