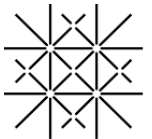


Influenza: Limited predictability of evolution Ecology of host and pathogen

Pierre Barrat-Charlaix



Politecnico
di Torino



Universität
Basel

BIOZENTRUM

Universität Basel
The Center for
Molecular Life Sciences

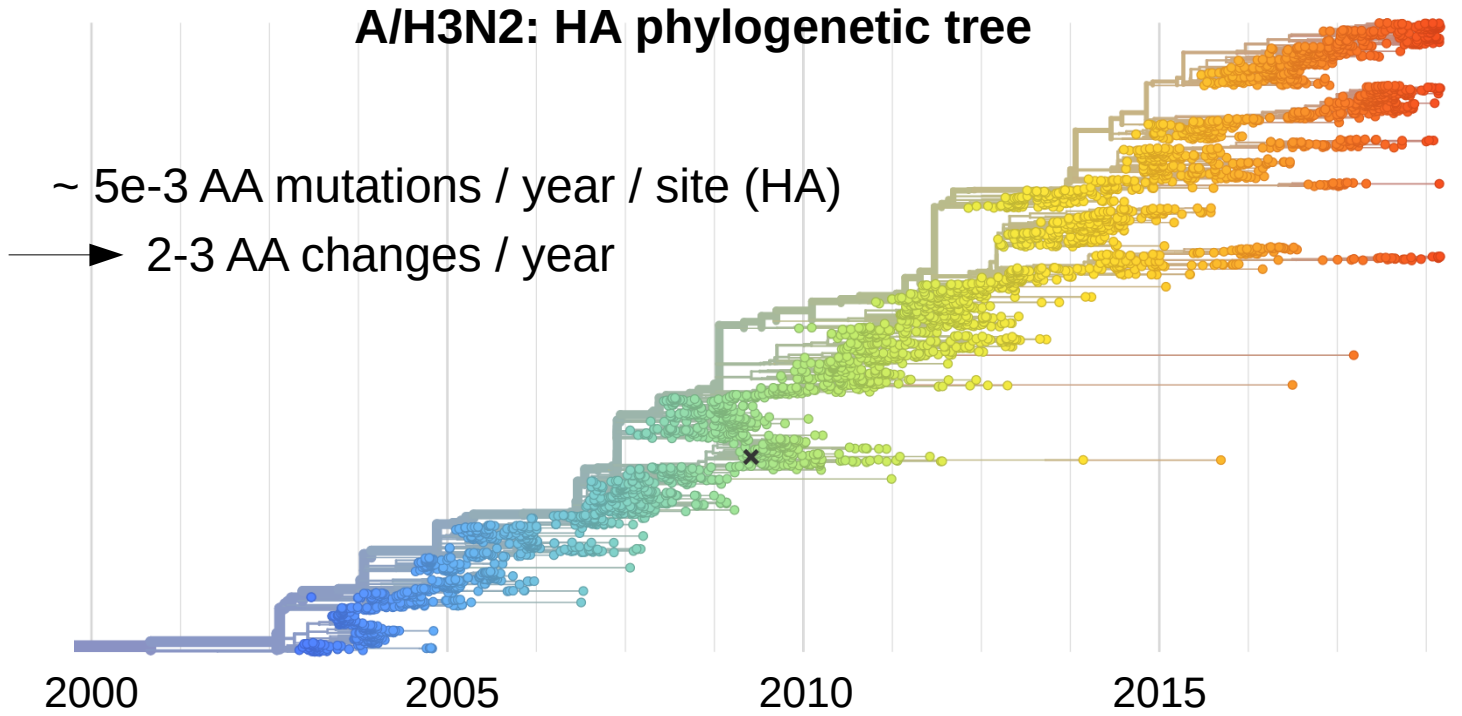
Group of **Richard Neher**

Collaborators:
John Huddleston
Trevor Bedford

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

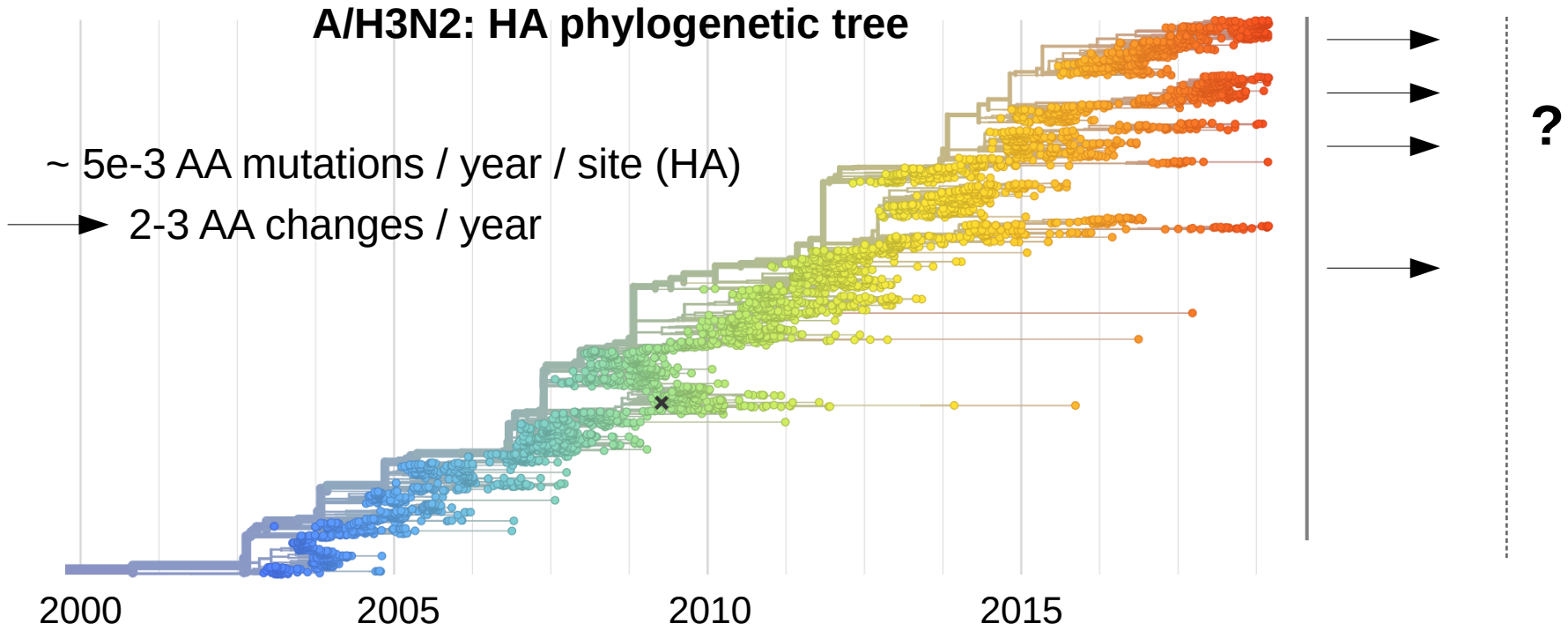


Variability in the present population

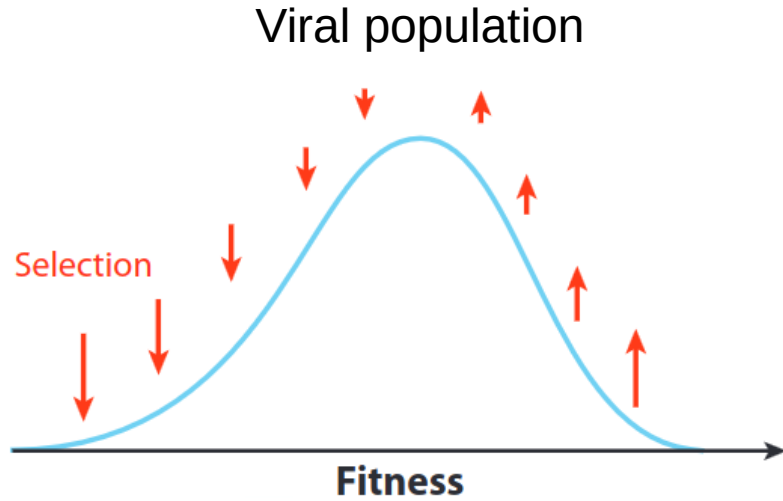
Human seasonal influenza virus

~ 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



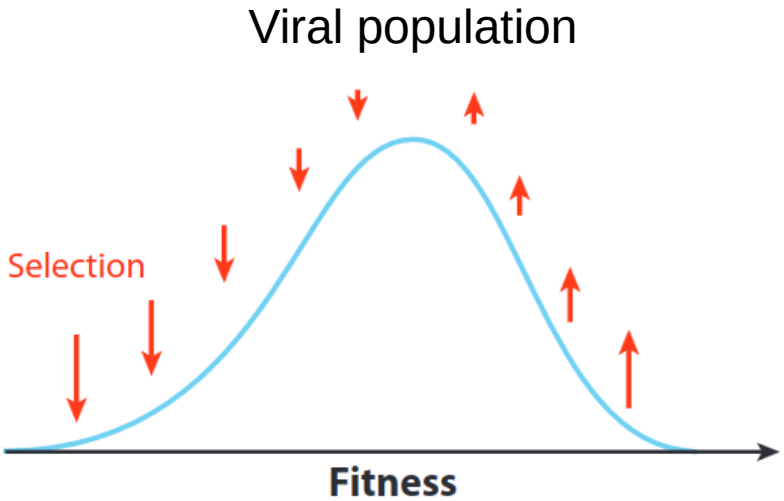
Traditional approach: travelling fitness wave



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

→ **Some degree of predictability**

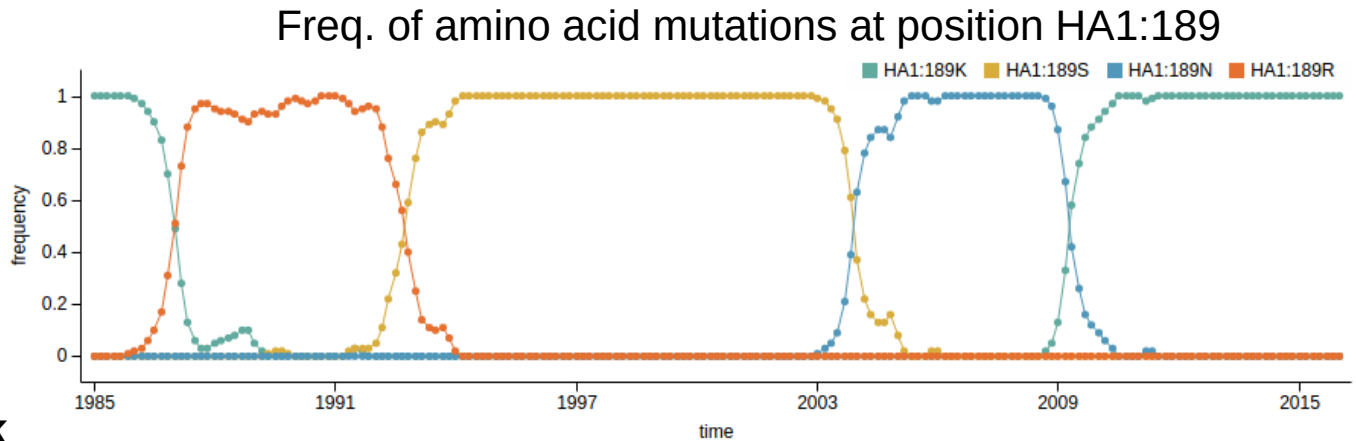
Traditional approach: travelling fitness wave



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

→ **Some degree of predictability**

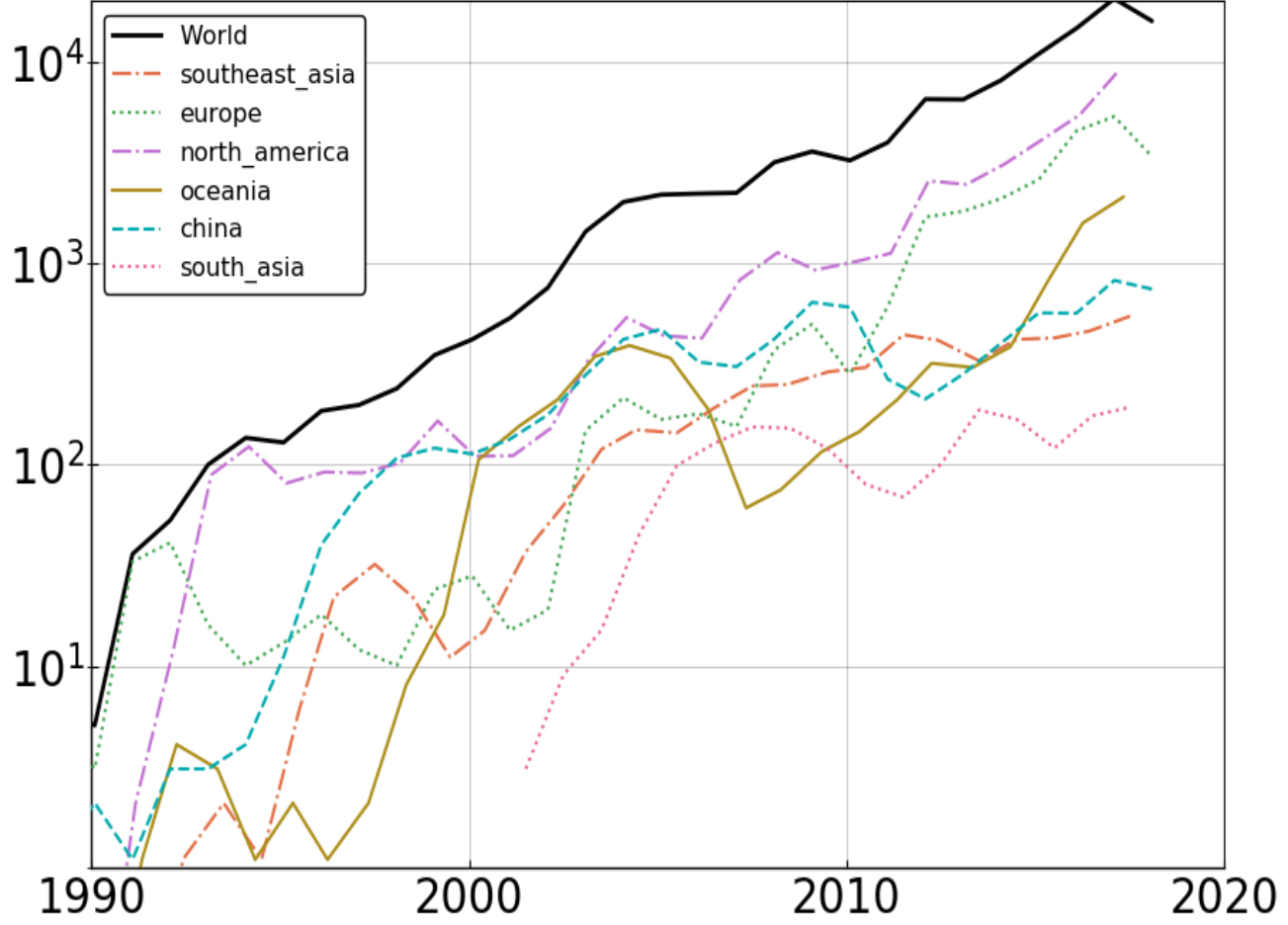
Selective sweeps



Is this the right way to think about the process?

Data

of seqs per year

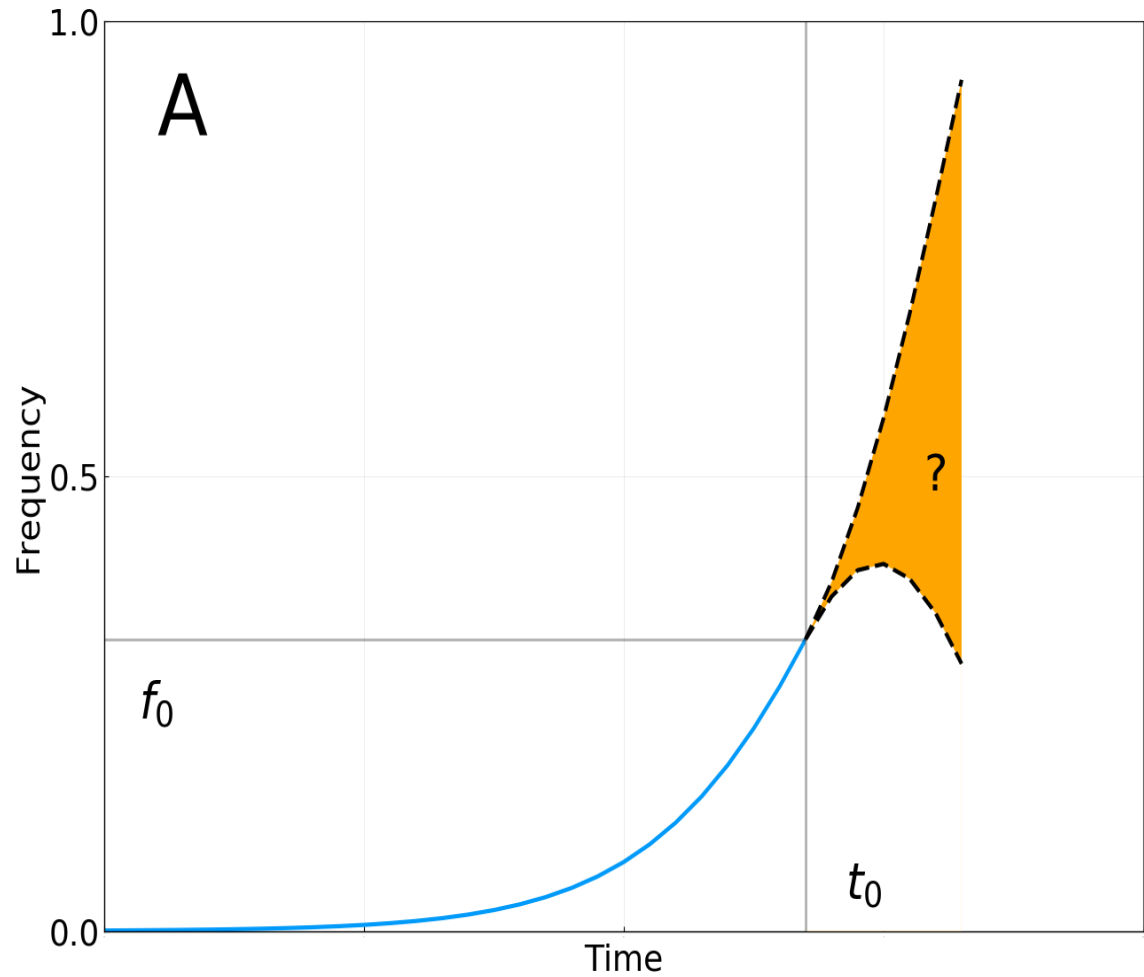


Time-binning of the past sequences by 1 month intervals

—▶ **Snapshots of the population**

—▶ **Frequency trajectories**

Simple analysis: predictability of influenza

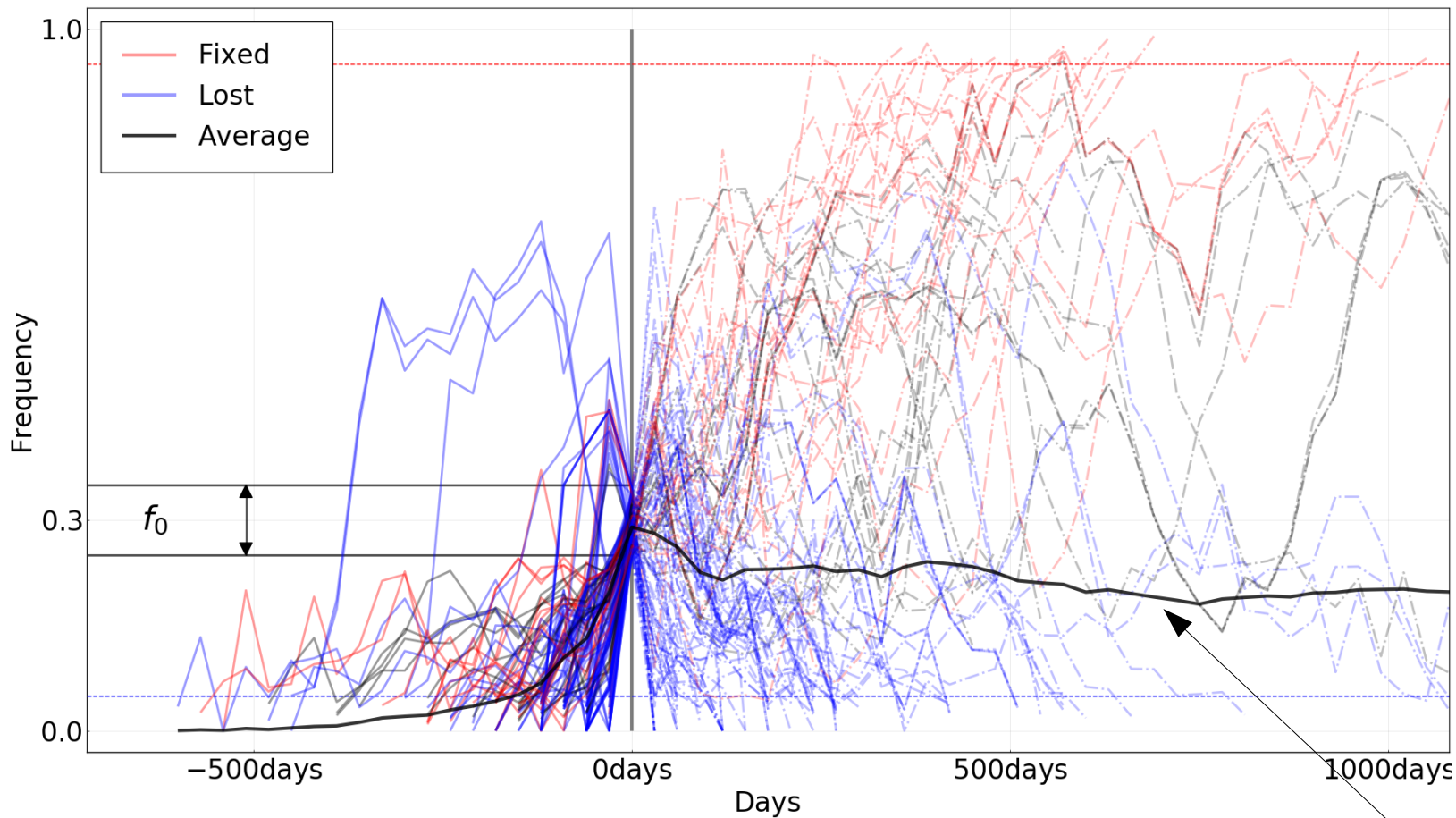


Frequency trajectories of amino acid mutations

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

Inertia of trajectories

Influenza H3N2, HA protein



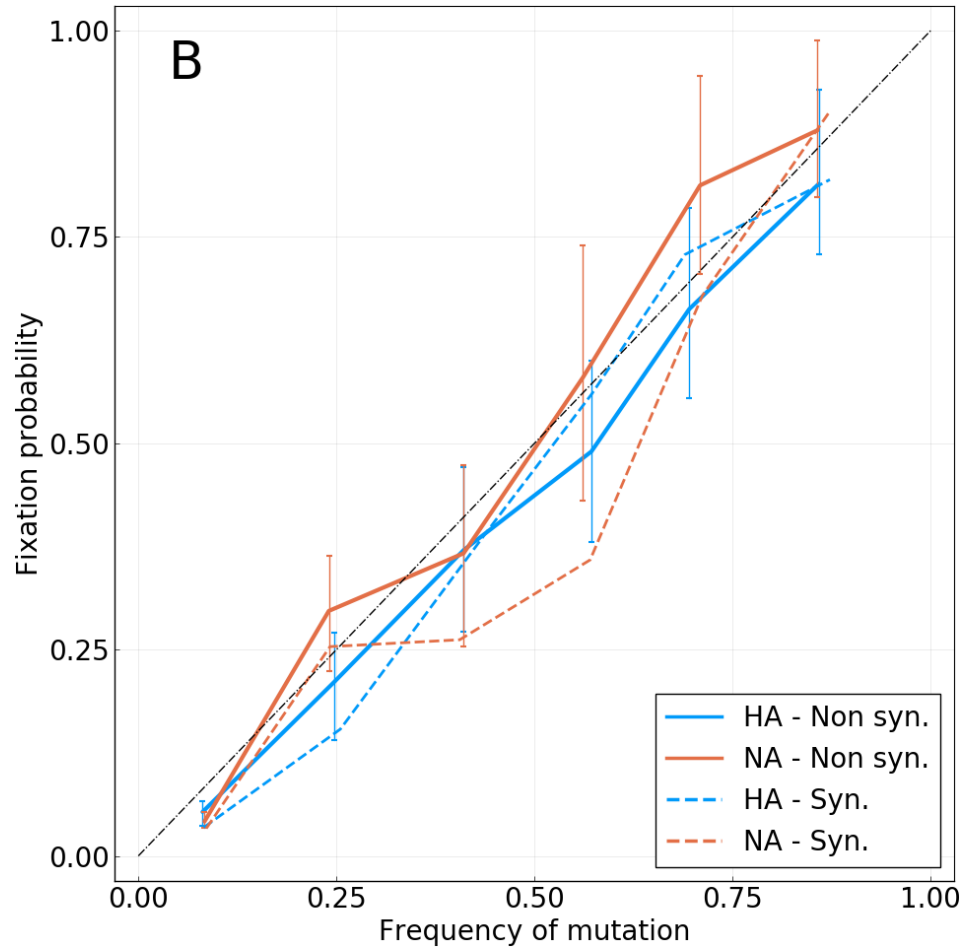
Mutations:

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

—▶ **No inertia**

Average

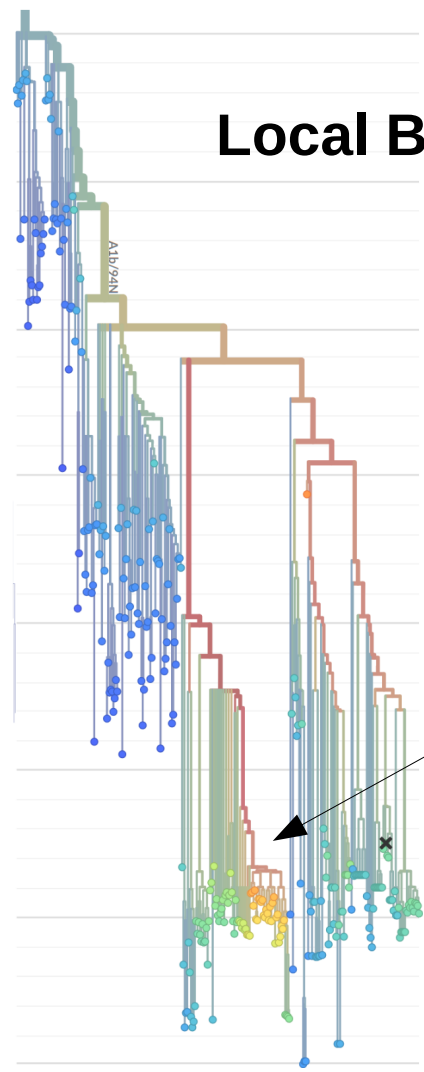
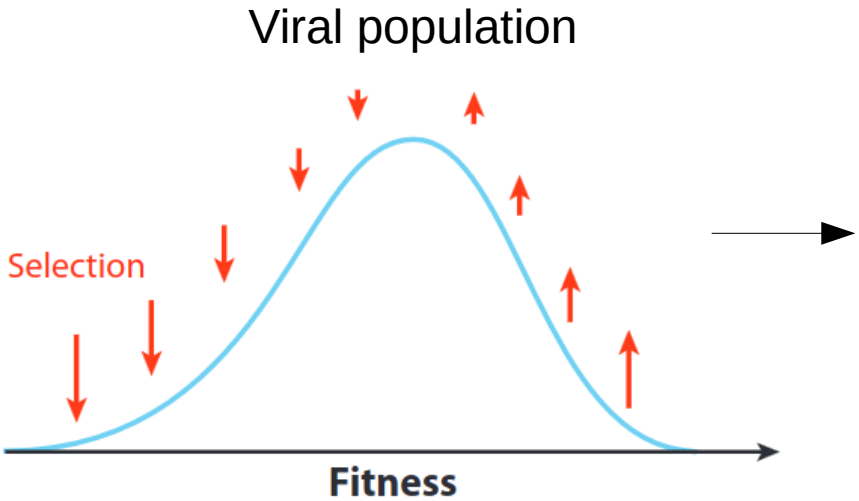
Probability of fixation



For rising trajectories

No signs of selection !

Using a proxy for fitness



Neher, 2014

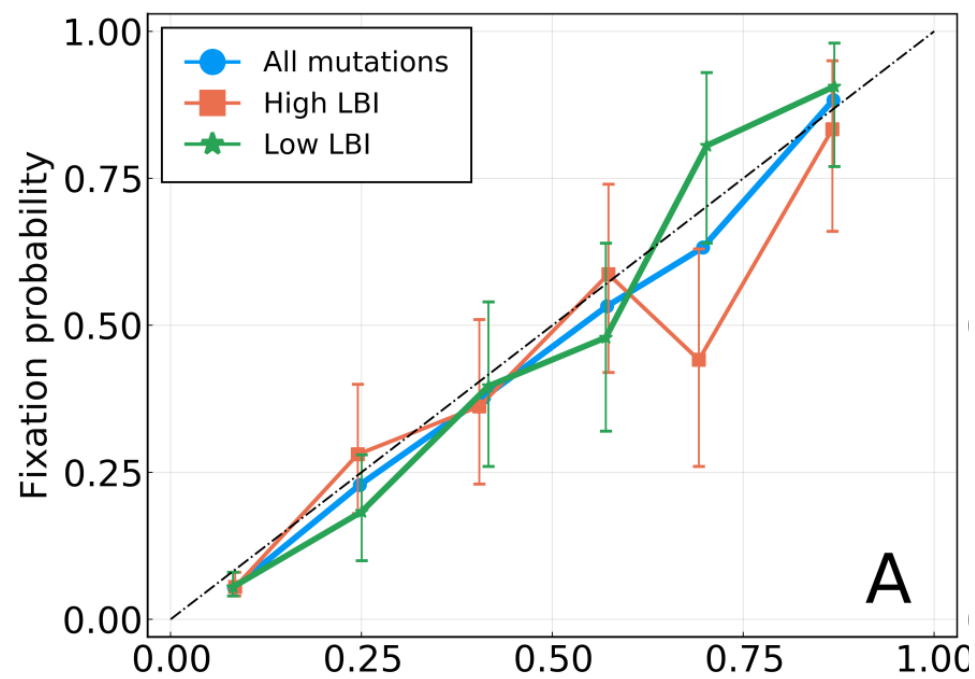
Local Branching Index (LBI)

Strains with high fitness have more offsprings

Dense branching

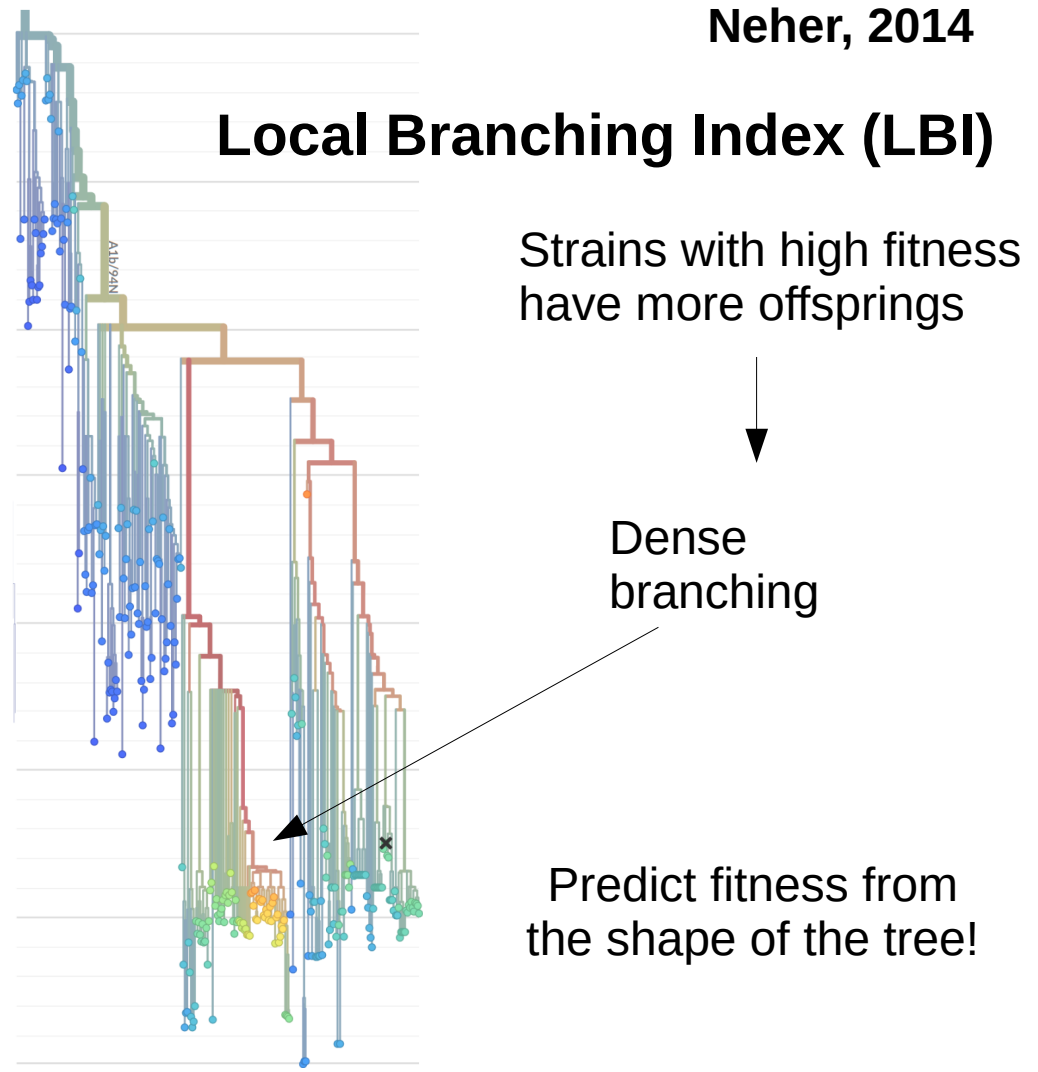
Predict fitness from the shape of the tree!

Using a proxy for fitness



A

Neher, 2014



Local Branching Index (LBI)

Strains with high fitness have more offsprings

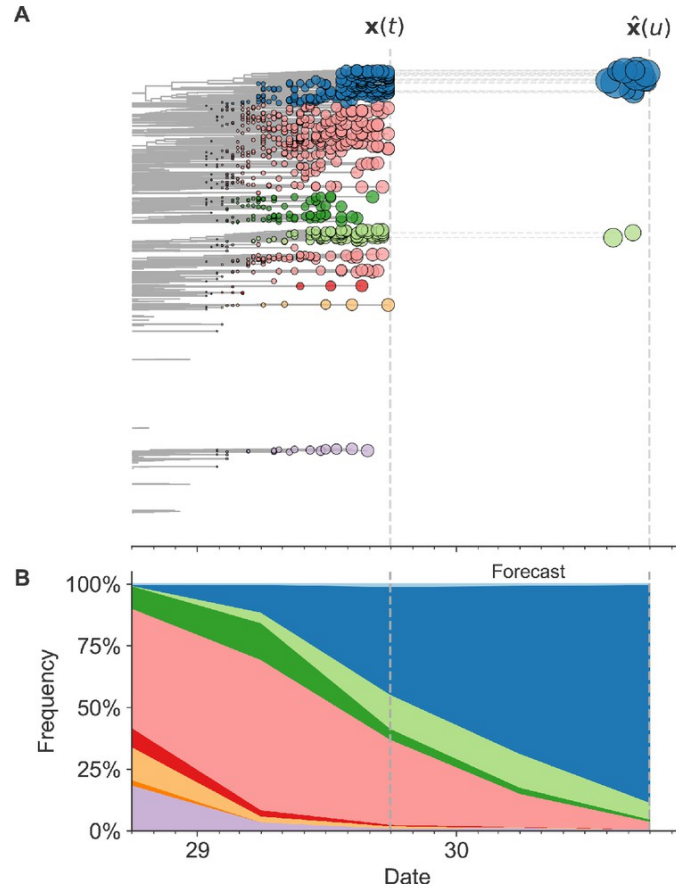
Dense branching

Predict fitness from the shape of the tree!

Strain level forecast

Huddleston et. al. 2020

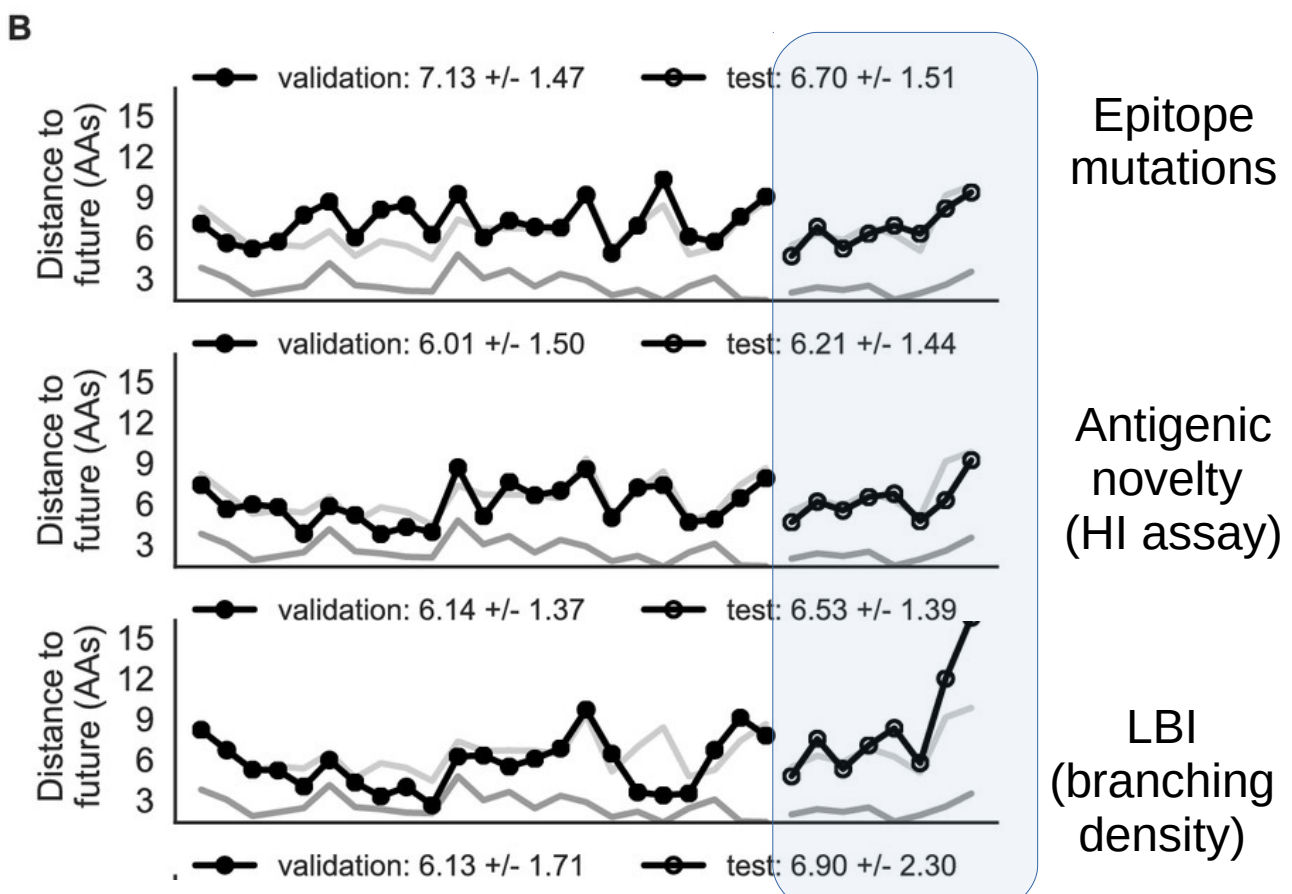
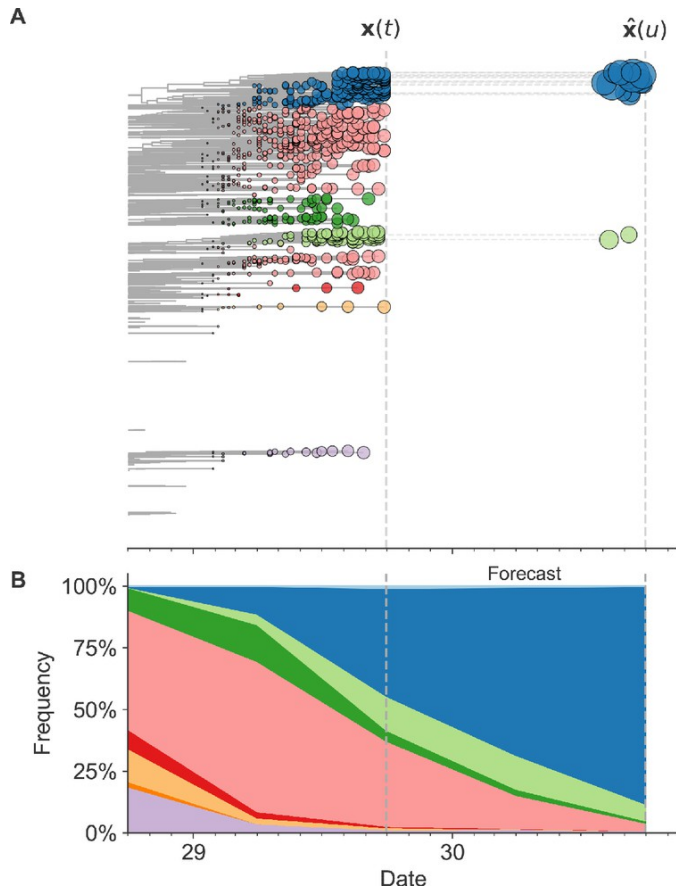
- Predict a fitness for each strain --> fit model to data (LBI, antigenic novelty, ...)
- Forecast future population 1 year ahead



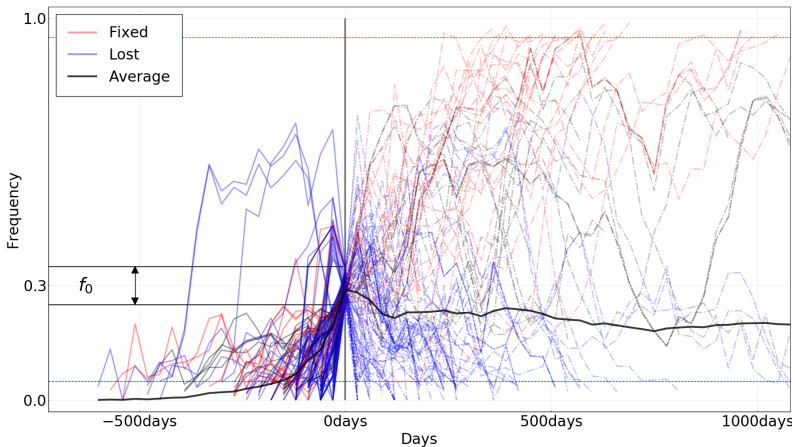
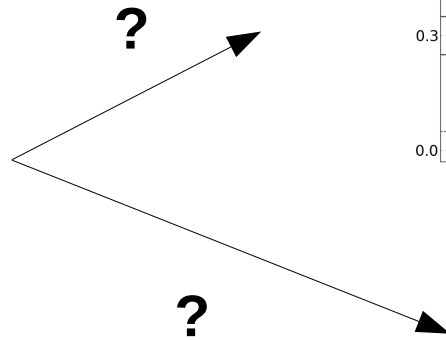
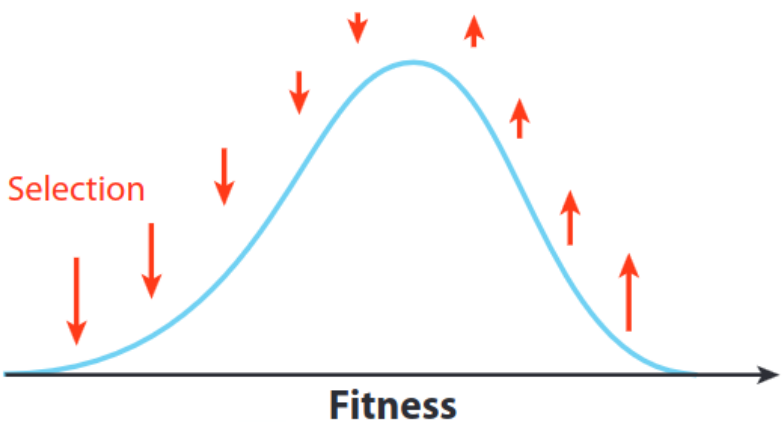
Strain level forecast

Huddleston et. al. 2020

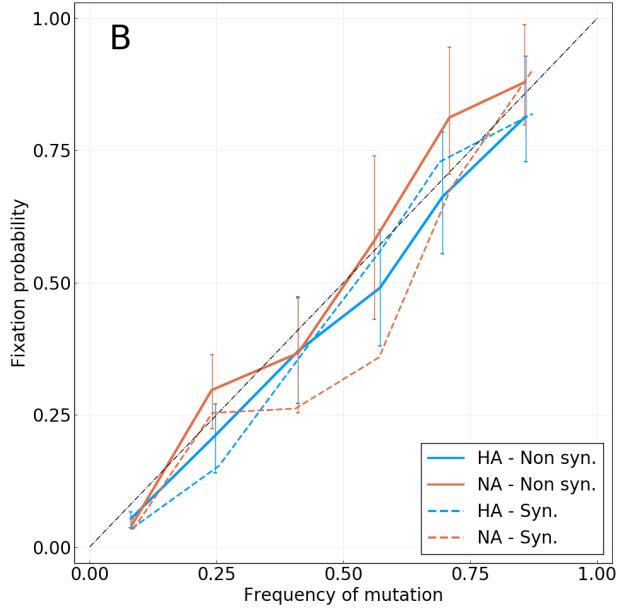
- Predict a fitness for each strain --> fit model to data (LBI, antigenic novelty, ...)
- Forecast future population 1 year ahead



Limited predictability



Qualitative difference between model and observations



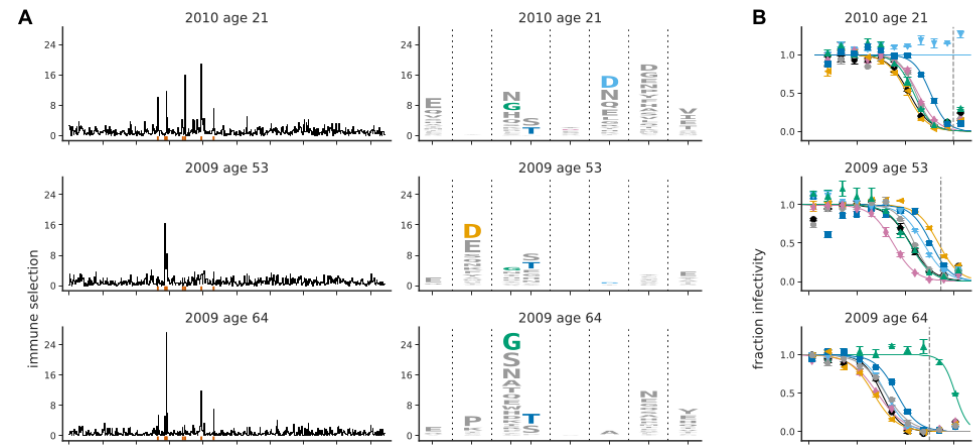
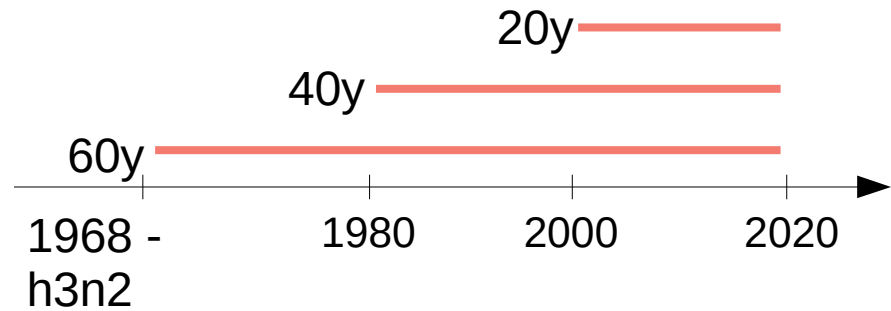
Adaptive immunity and expiring fitness

Adaptation is driven by immunity

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

Adaptive immunity of hosts

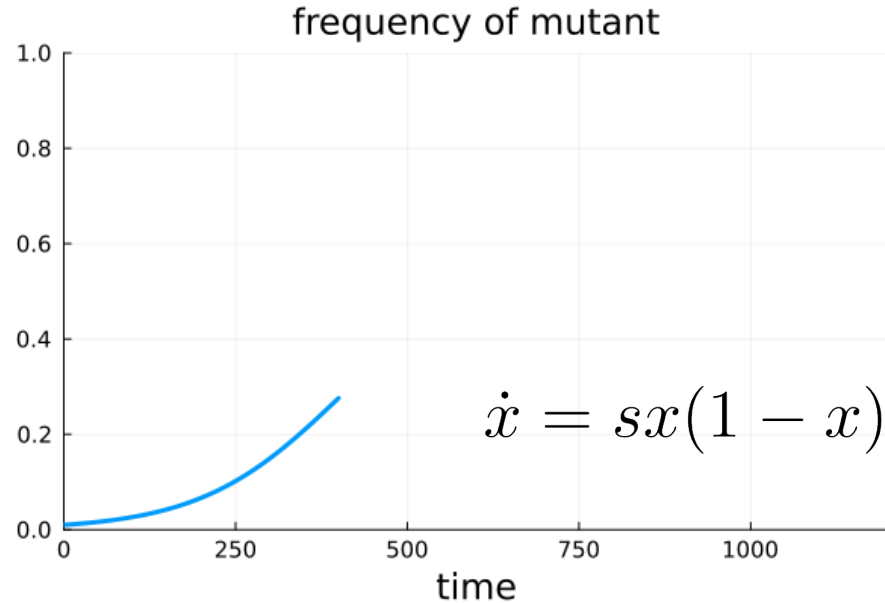
- fitness advantage expires before fixation
- “ecology”: organisms shape their environment



Adaptive immunity and expiring fitness

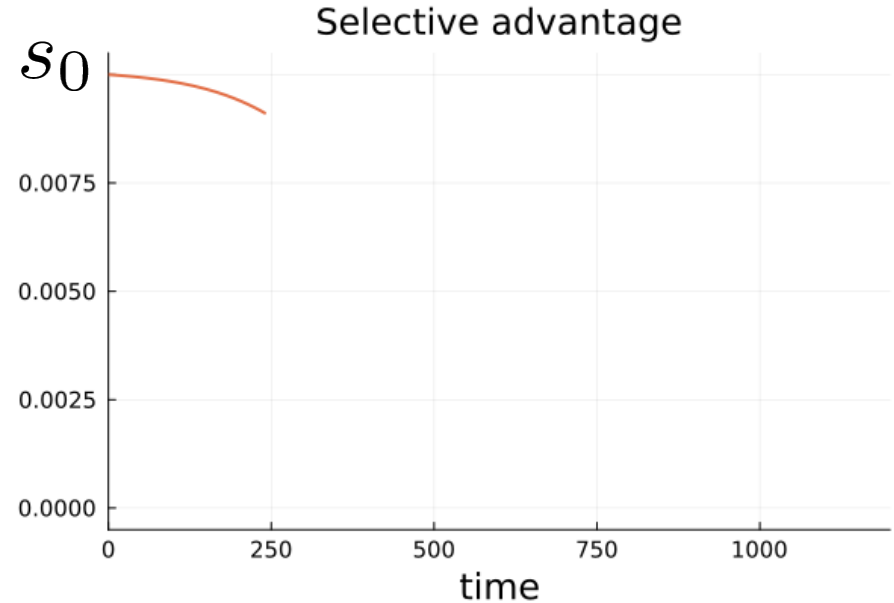
Adaptation is driven by immunity

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



Adaptive immunity of hosts

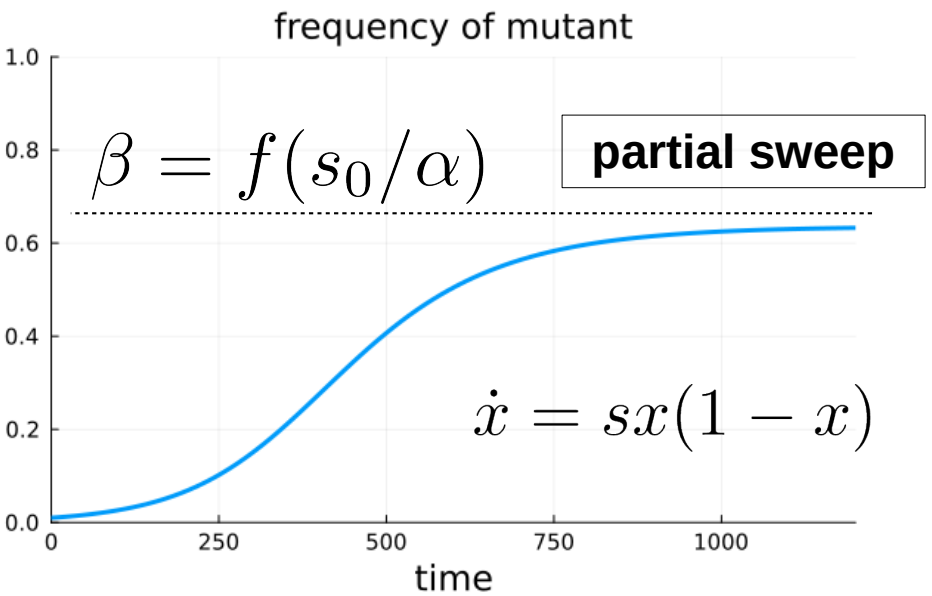
- fitness advantage expires before fixation
- “ecology”: organisms shape their environment



Adaptive immunity and expiring fitness

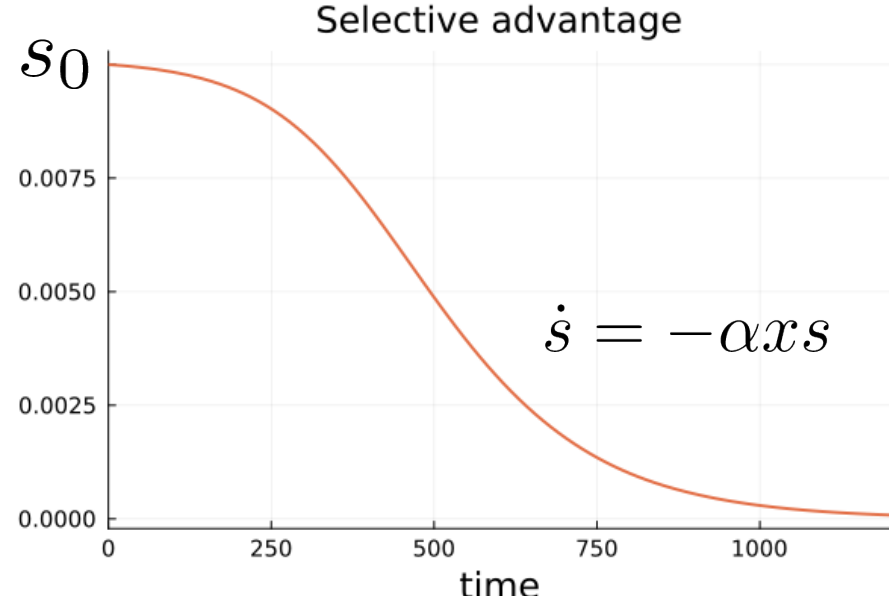
Adaptation is driven by immunity

- 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



Explicit model: SIR model

- N strains: a, b in $(1 \dots N)$
- For each strain, hosts are susceptible, infected or immune

$$\dot{S}^a = -\alpha S^a \sum_{b=1}^N K_{ab} I^b + \gamma(1 - S^a),$$

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

α \rightarrow rate of infection

δ \rightarrow duration of infection (rate)

γ \rightarrow population turnover

K_{ab} \rightarrow cross-immunity

Explicit model: SIR model

- N strains: a, b in $(1 \dots N)$
- For each strain, hosts are susceptible, infected or immune

$$\dot{S}^a = -\alpha S^a \sum_{b=1}^N K_{ab} I^b + \gamma(1 - S^a),$$

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

α \rightarrow rate of infection

δ \rightarrow duration of infection (rate)

γ \rightarrow population turnover

K_{ab} \rightarrow cross-immunity

Equilibrium

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

Partial sweep with SIR model

($b = 0.7, f = 0.8$)

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

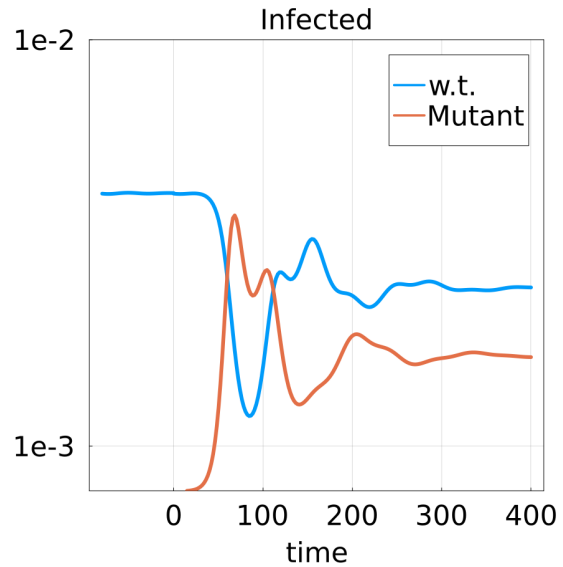
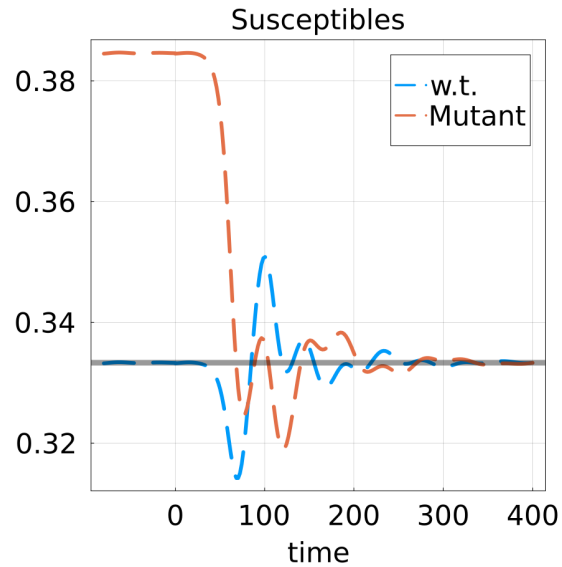


$$K = \begin{bmatrix} 1 & b \\ f & 1 \end{bmatrix}$$

cross-immunity

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

- At $t = 0$, introduce mutant



Partial sweep with SIR model

($b = 0.7, f = 0.8$)

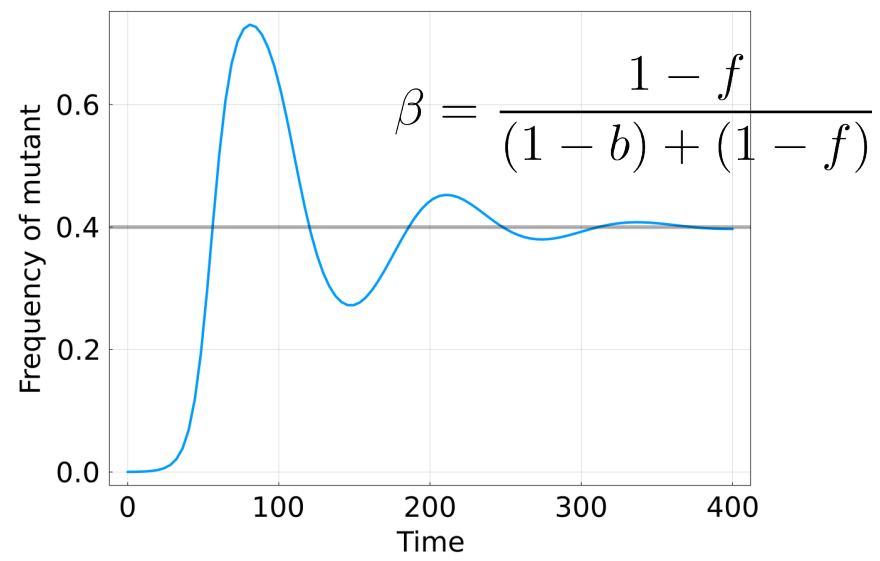
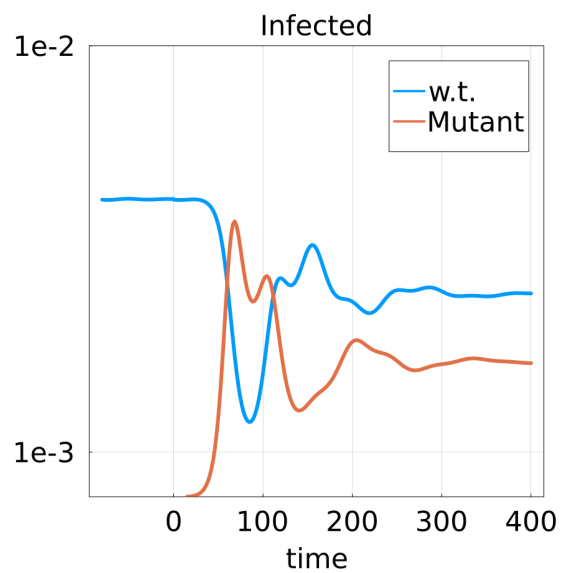
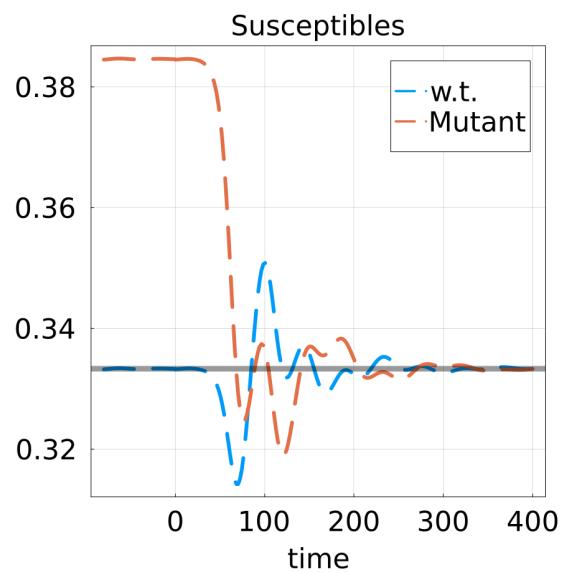
One **wild-type** virus, one **mutant**

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

cross-immunity

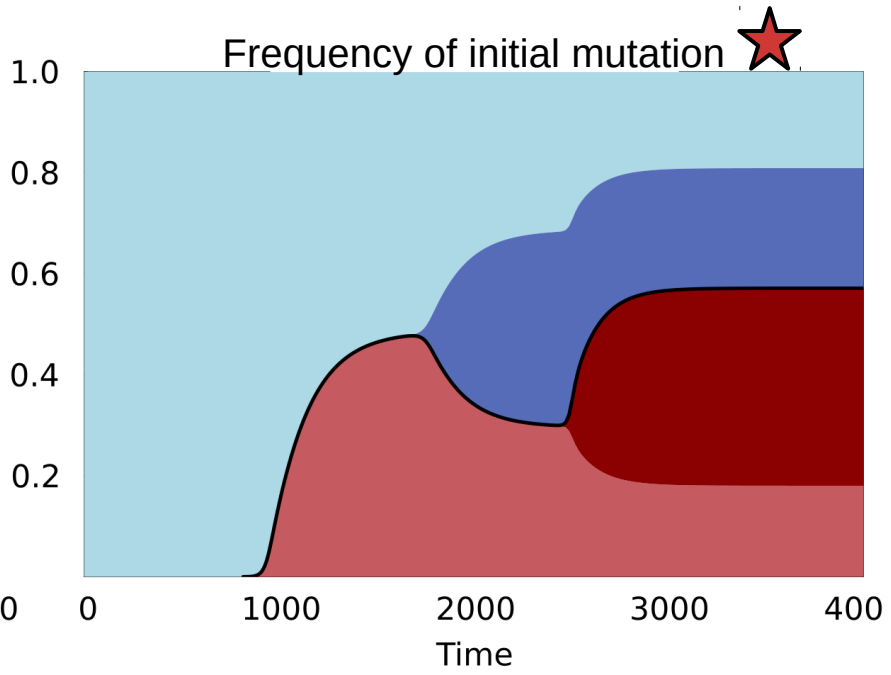
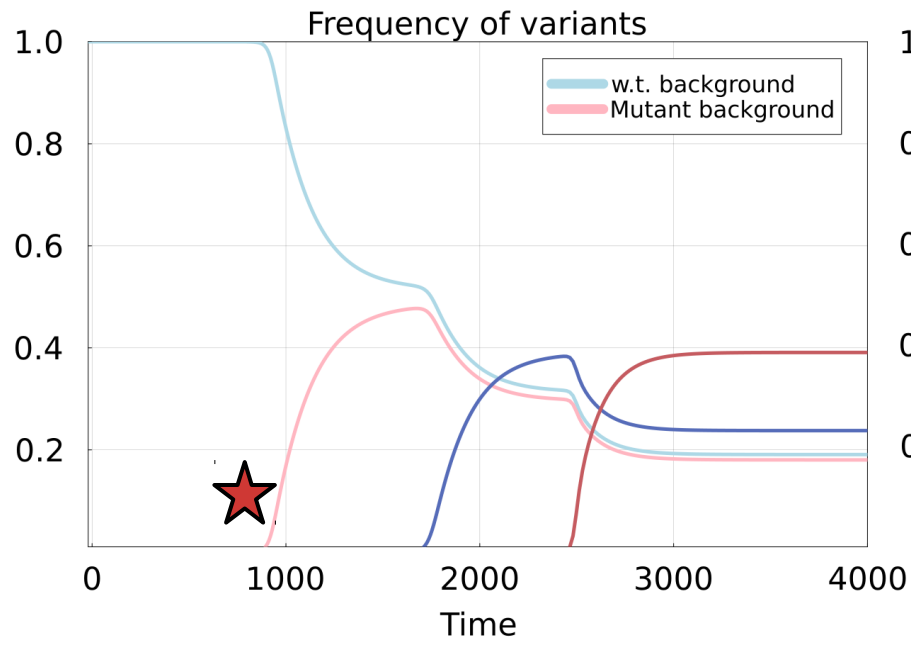
Initially, **no mutant** $I^{mut} = 0$

- At $t = 0$, introduce mutant



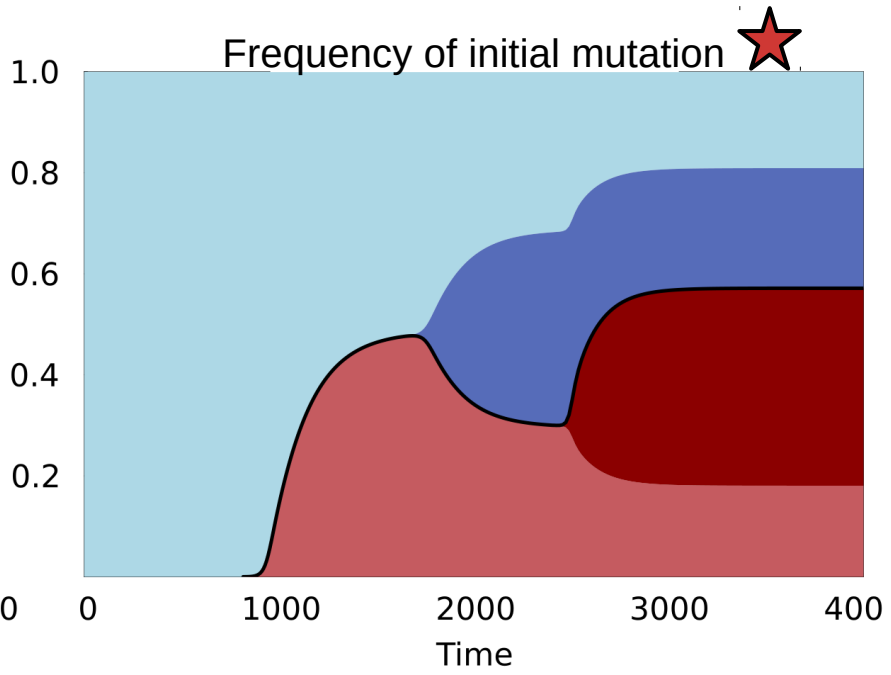
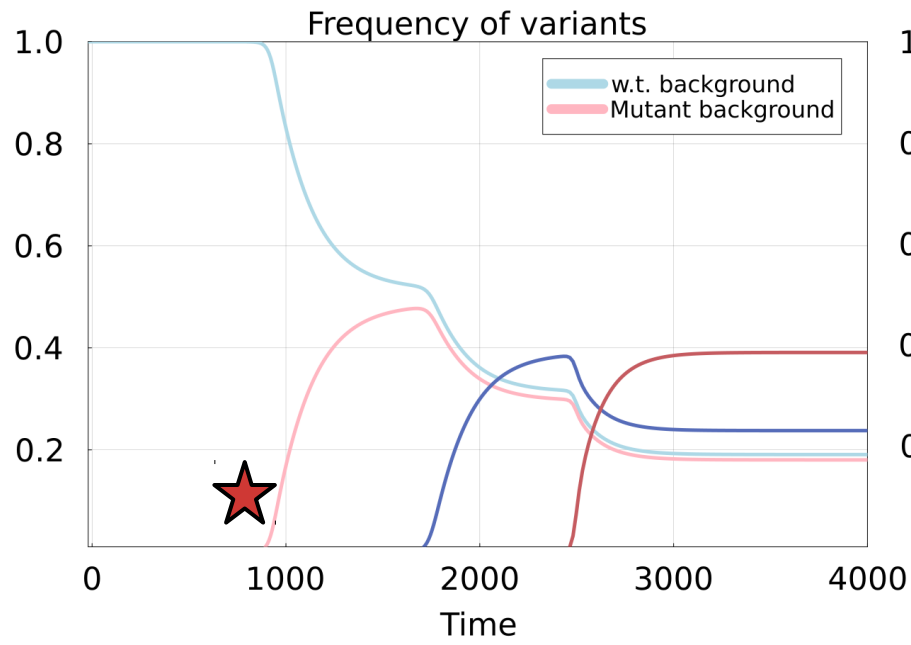
Evolution with partial sweeps

New variants appear at rate ρ



Evolution with partial sweeps

New variants appear at rate ρ



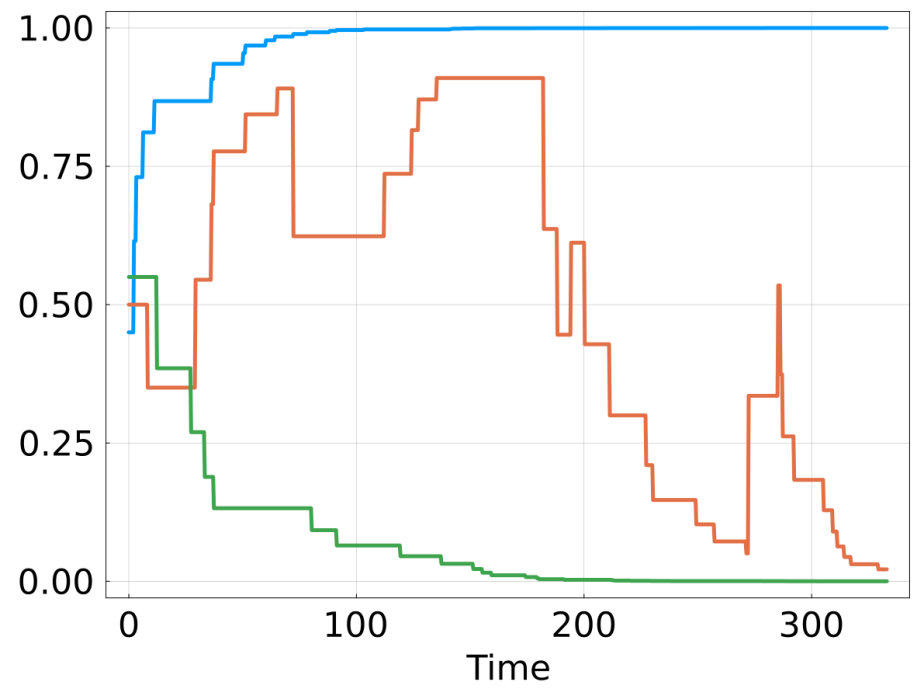
Limit case: non overlapping partial sweeps

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

Evolution with partial sweeps

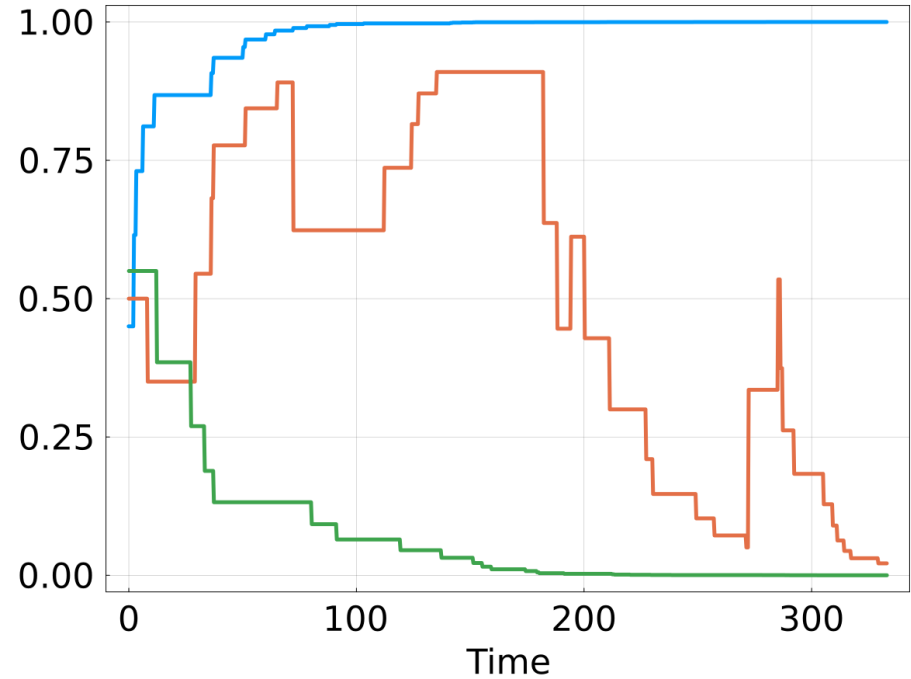
$$x_{t+1} = x_t + \begin{cases} \beta(1 - x_t) & \text{with prob. } x_t, \\ -\beta x_t & \text{with prob. } 1 - x_t, \end{cases}$$

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



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**

Timescales

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

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

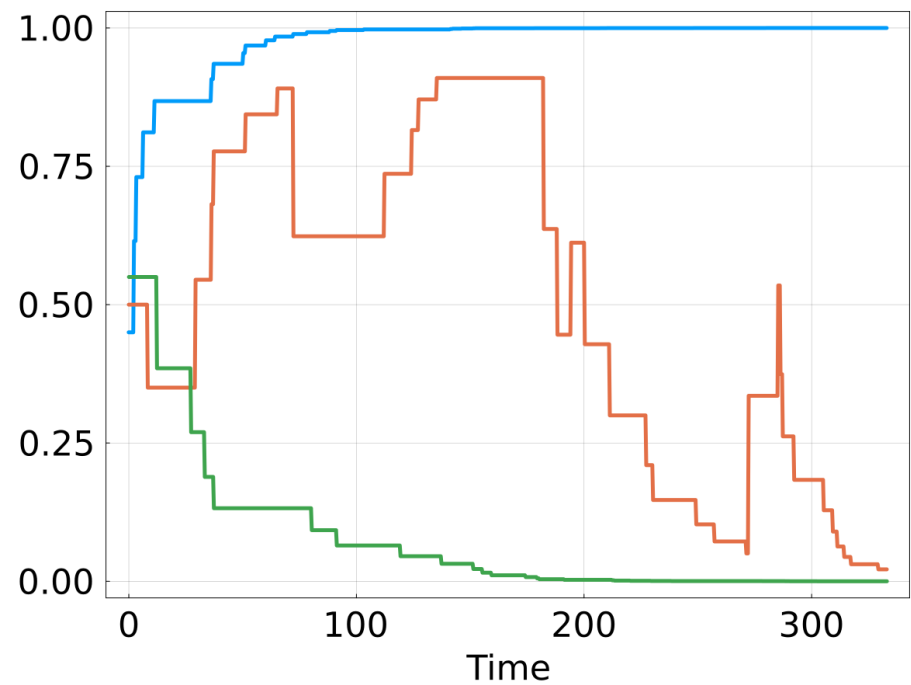
$\rho^{-1} \sim 15$ (~4 partial sweeps/year)

$\beta \sim 0.3 \rightarrow T \sim 150 \sim 3y$

H3N2 influenza $\rightarrow T_{MRC A} \sim 6y$

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**

Timescales

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

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

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

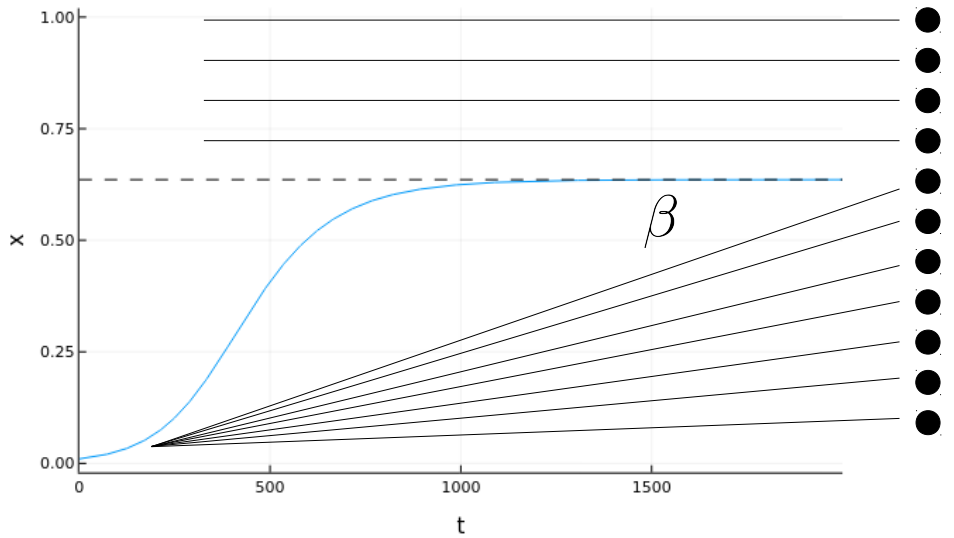
$\beta \sim 0.3 \rightarrow T \sim 150 \sim 3y$

H3N2 influenza $\rightarrow T_{MRC A} \sim 6y$

Higher moments: similar to neutral drift

$$\langle \Delta x^3 \rangle = \rho \langle \beta^3 \rangle x(1 - x)(1 - 2x)$$

Shape of the phylogeny: multiple mergers

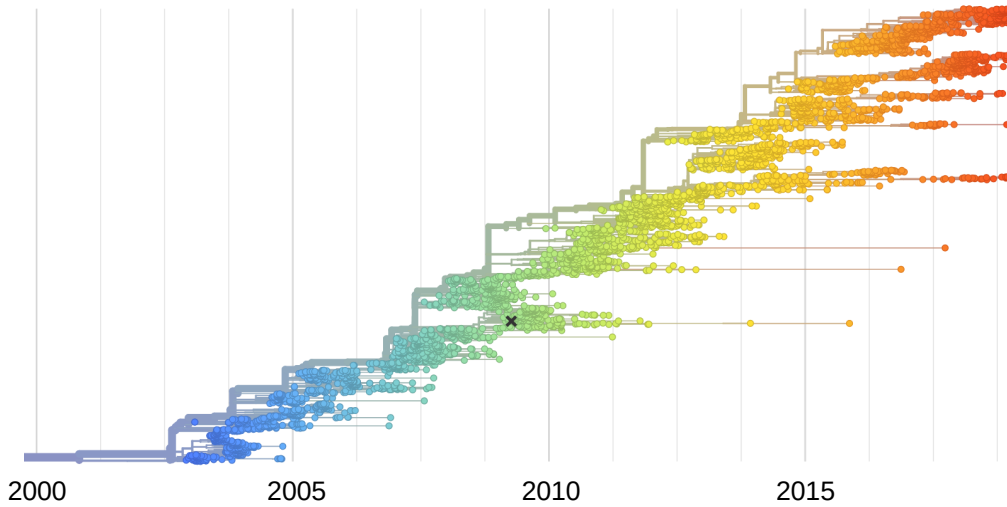


- Multiple mergers
- Ladder-like phylogenies

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

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

Lambda coalescent



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!

Summary

H3N2Influenza:

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

Partial sweeps

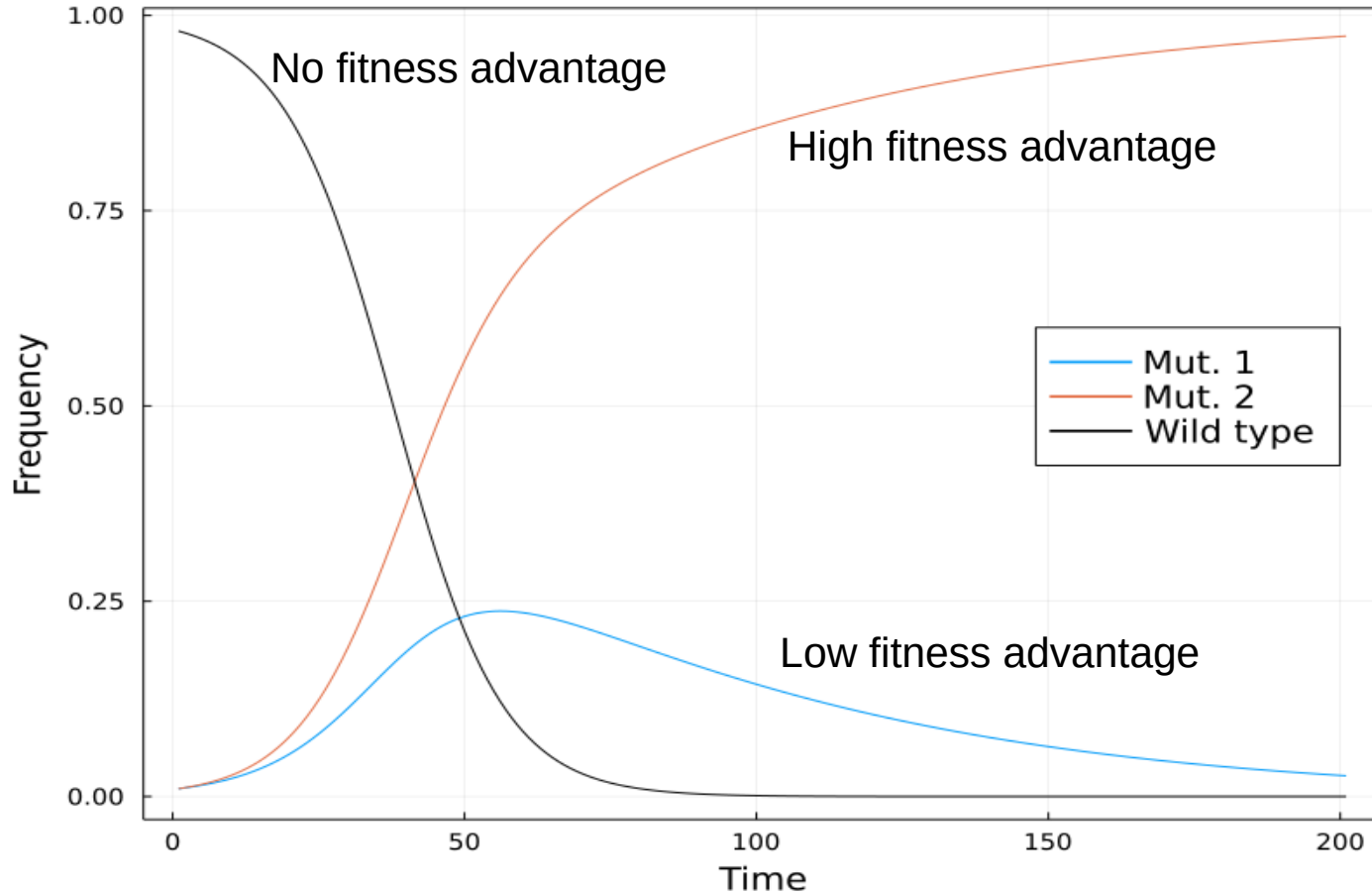
- **Adaptation** of host immunity
- Fitness advantage of mutant **expires** before full sweep

Evolution with **partial sweeps**

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

Thank you!

Is this expected? Clonal interference



Adaptive mutations
appearing on
different individuals



Competition

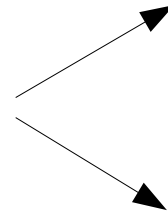
Genetic linkage: toy model

Simulate a population

Simple fitness landscape

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

Change the fitness landscape periodically

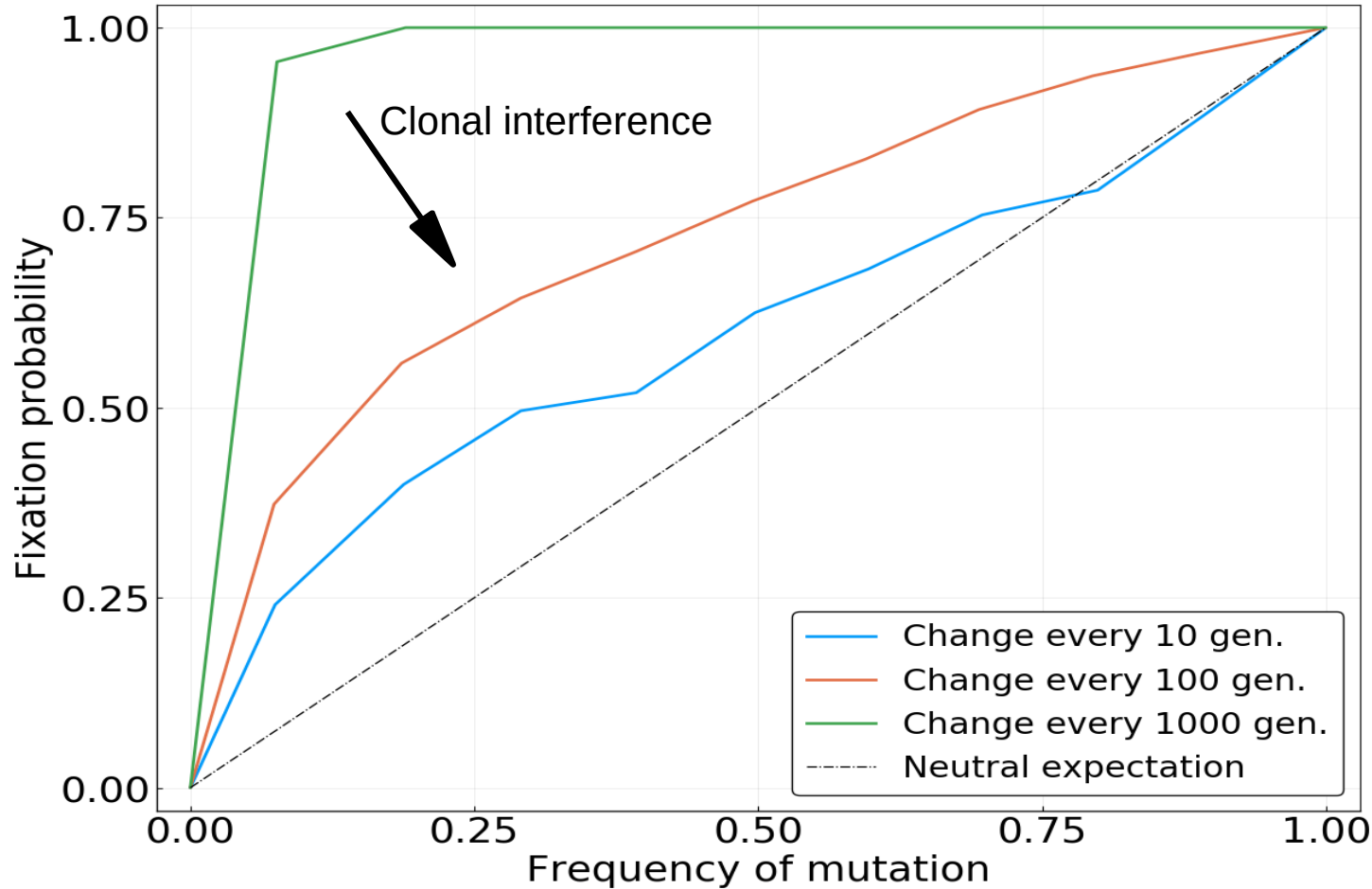


Slow rate of change
Clean sweeps

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!