Evolution of influenza: A molecular red queen race

Michael Lässig
Institute for Theoretical Physics, University of Cologne

joint work with Natalja Strelkowa, Imperial College, London

An evolutionary biologist's dream

- **Influenza is a well-observed evolutionary system:**

  - **viral genome**: consists of 8 recombining segments
    haemagglutinin coding sequence: data available over 40 years

  - **functional information** on host interactions (via epitope)

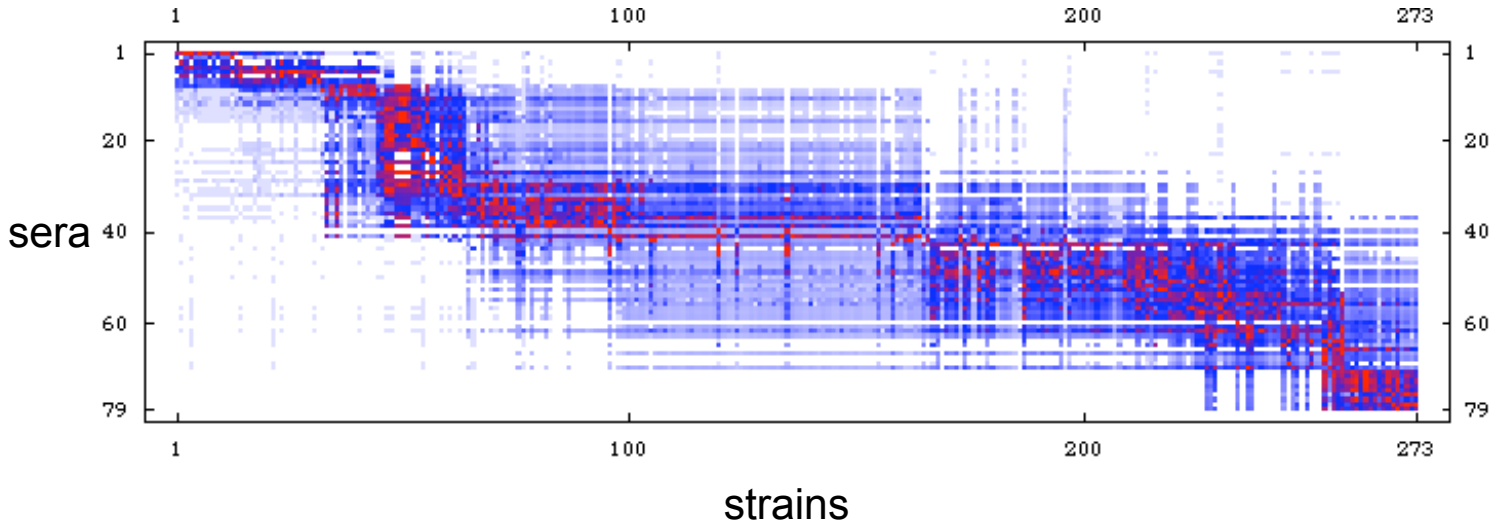
viral epitope
(containing receptor-
binding site)

viral haemagglutinin
(surface protein)

human antibody
(containing antigen-binding
fragment)

An evolutionary biologist's dream

- phenotypic data: **strain-serum interactions**
  measured by haemagglutinin inhibition tests



[Smith et al., Science 2004]

-> **in vivo time series on adaptation to a changing environment.**

An theorist's nightmare

- **Influenza is a strong-coupling system:**

  - linkage within the HA1 domain sequence

  - strong and complex selection with explicit time- and frequency dependence

  - high mutation rate

- **Can we quantify natural selection acting on this system?**
- **Can we predict its antigenic evolution?**

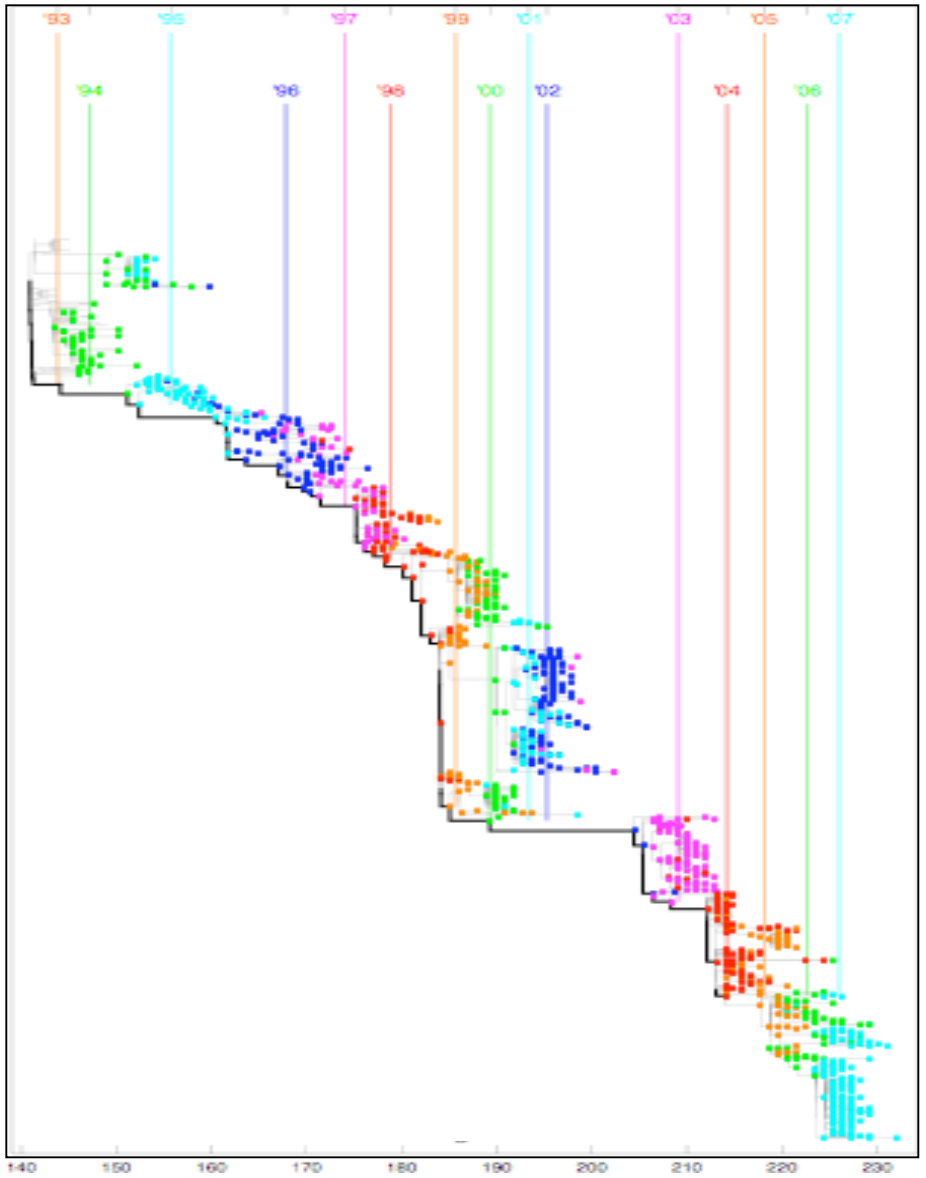
1. Reconstructing the evolution of the influenza genome

Genetic trees

- **Coalescent tree** based on HA1 sequences of 1971 influenza H3N2 strains**:**

- Genome evolution follows an approximate **molecular clock** of 4 substitutions/year in the HA1 domain

- **Sequence diversity** is limited, **coalescent times** are ~ 3.5 yrs



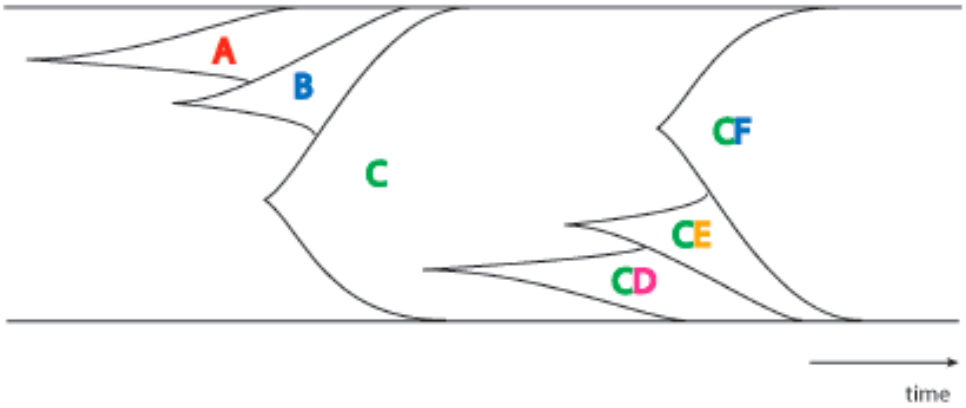
mutational distance

2. Effects of genomic linkage and high mutation rate

Linkage effects

- In a **small population**, fixations of benefical mutations are rare and independent:

- **Clonal interference: competion of linked beneficial mutations** in a **large asexual population**

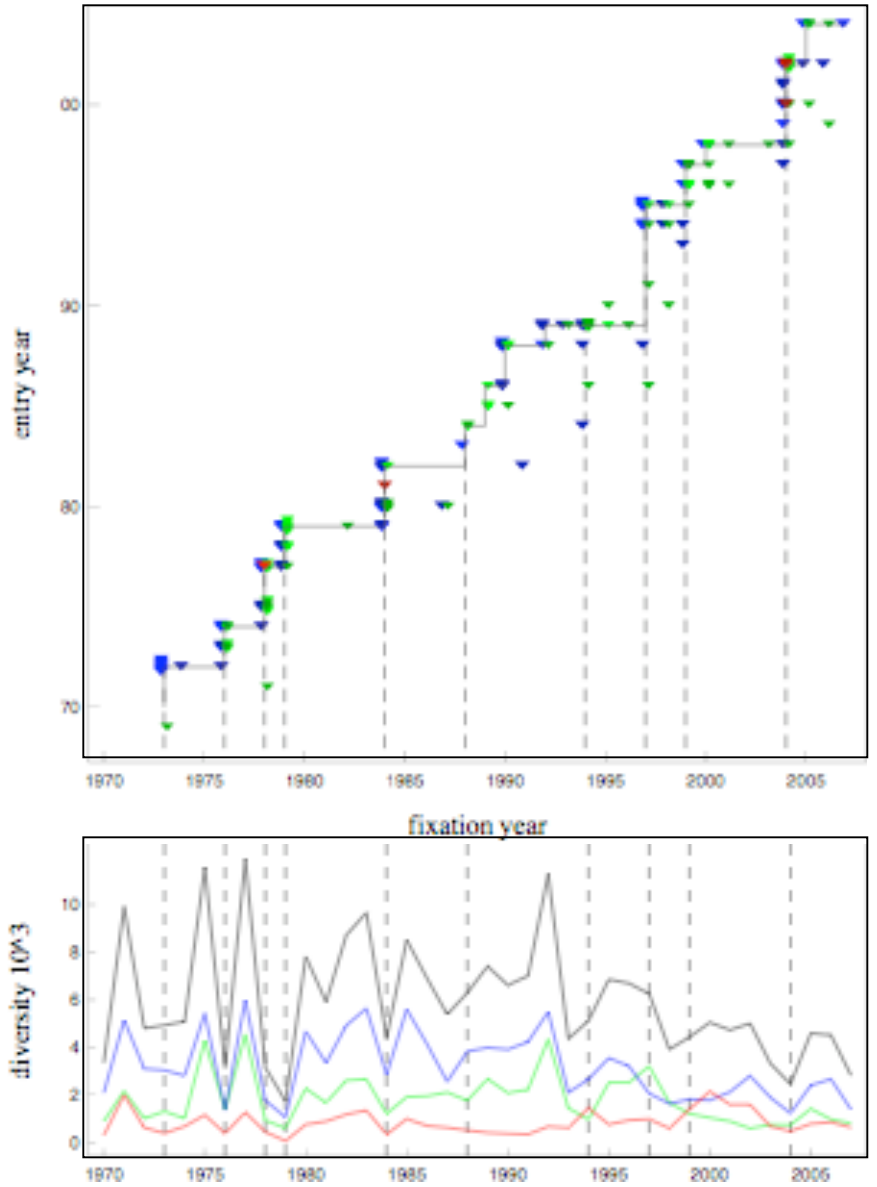


[Gerrish and Lenski,
Park and Krug,
Desai and Fisher]

Clonal interference leads to
-- the same **temporal ordering** of origination and fixation events
-- **clustering** of fixation times.

Linkage effects

- **Clonal interference in the influenza HA1 domain:**
  10 co-fixation clusters account for 75% of the substitutions.



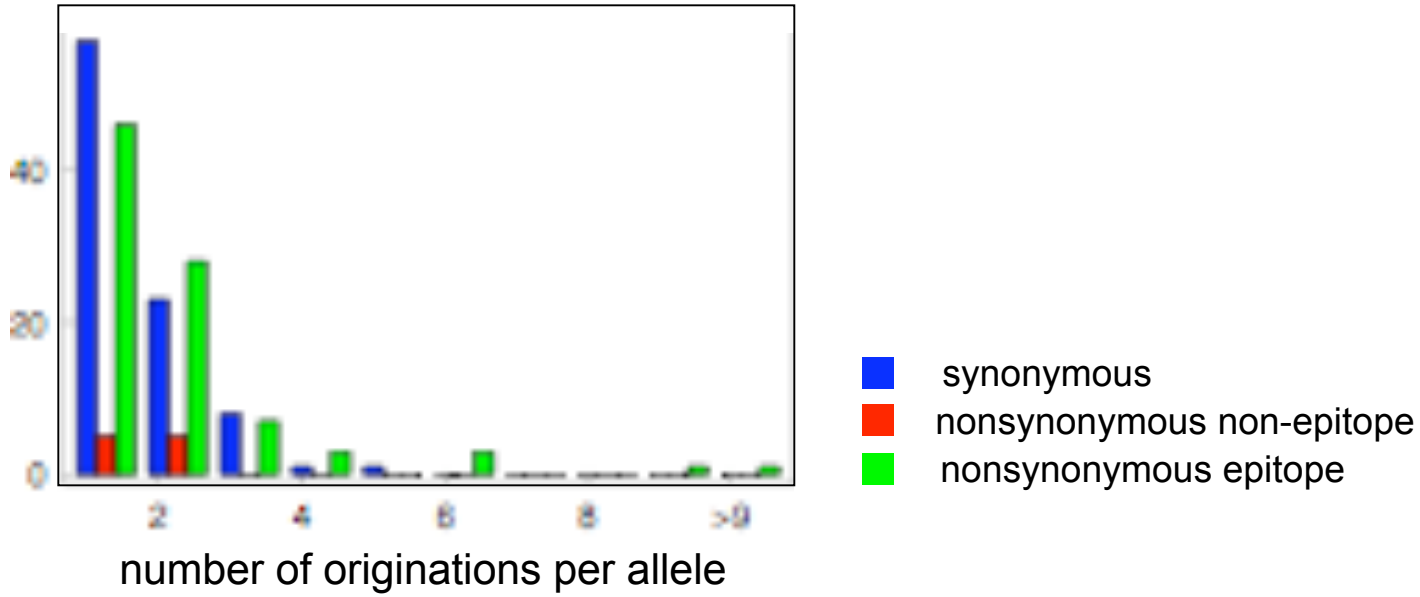
- Dips in **sequence diversity** correlate with fixation clusters.

Linkage effects

- Influenza has an extremely **high mutation rate:**

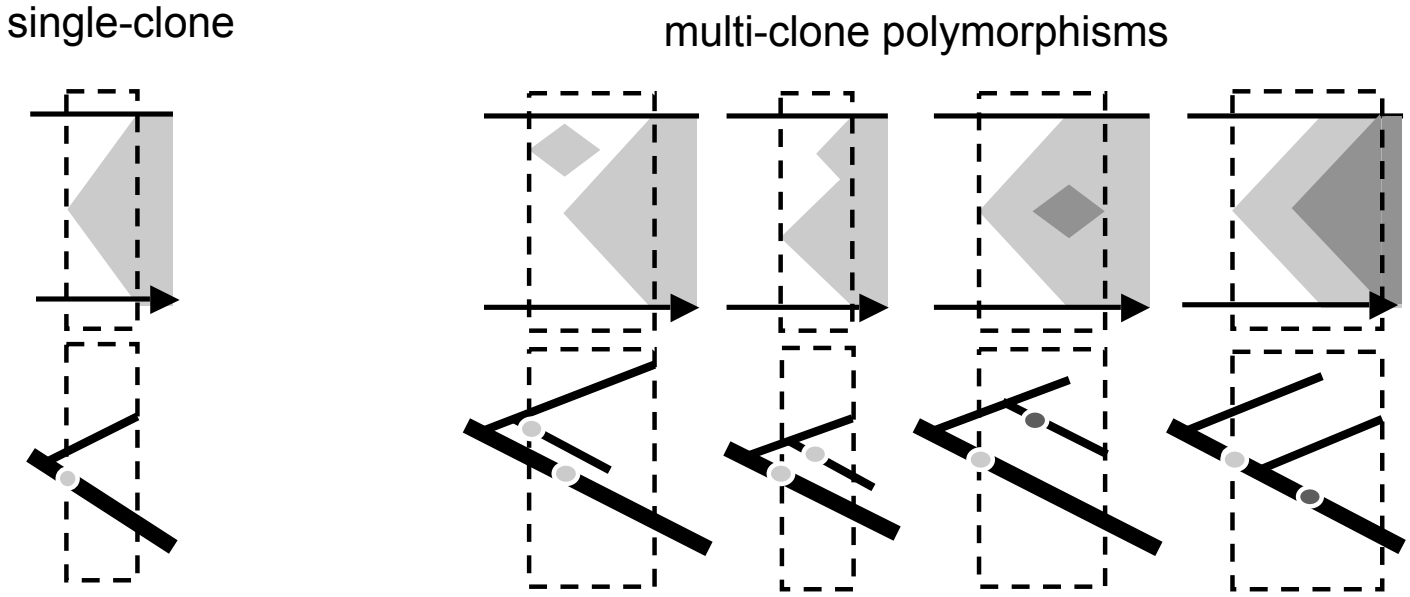
  5.8×10^{-3} / bp yr (synonymous substitutions)

- This leads to **multiple originations** of the same mutation in **competing clones**:



Linkage effects

- **Stochastic clonal interference:** multiple originations partially randomize the temporal order of originations and fixations.



- These processes are in influenza:

present **absent** present present

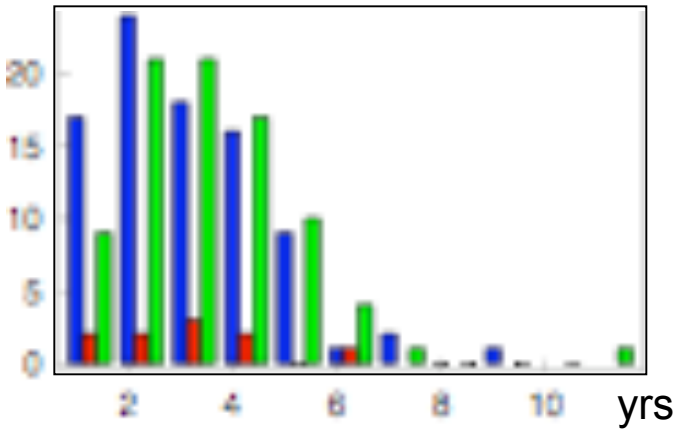
The **fixation** of a polymorphism always occurs in a **single** most successful clone**.**

3. Inference of selection

Selection inference

- Clonal interference reduces differences between sequence classes:

  - distribution of **fixation times**:



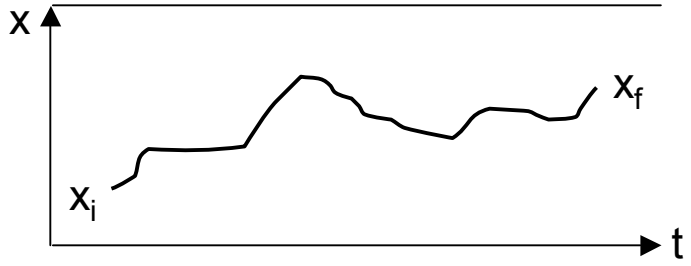
  - **polymorphism spectra**, …

Selection inference

- Selection inference from **allele frequency time series:**

  - frequency propagator

$$G (x_f | x_i)$$



  - propagator ratio between sequence classes

$$\phi (x_f) = G (x_f | x_i) / G_0 (x_f | x_i) \qquad (x_i << 1)$$

Selection inference

- **Non-epitope** aminoacid changes are under **negative selection:**

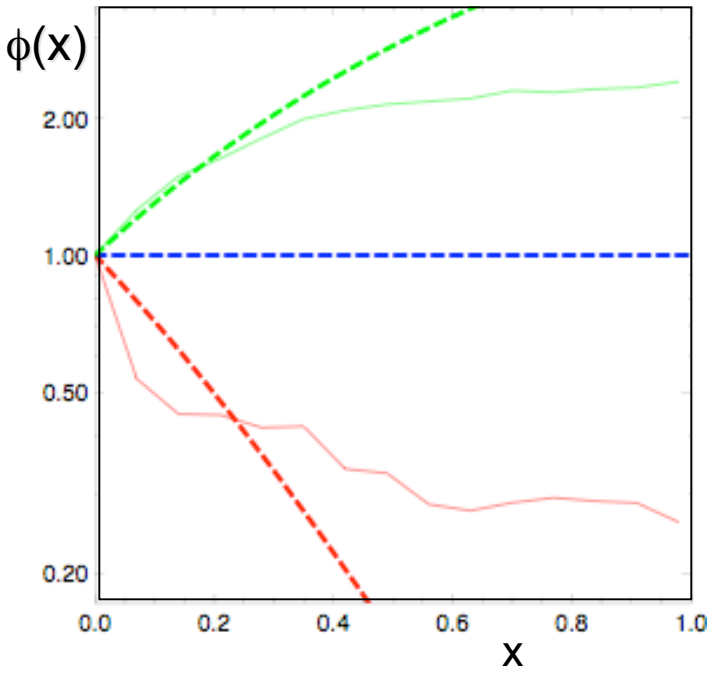
- fraction of selected sites:

$> \phi(1) = 80\%$

- strength of selection:

$s < - s^* = - 1/3$ yrs

$s^* \sim 3$ yrs: characteristic sweep time



→ $\mu / s^* = 0.02$ **is (not far) below the error threshold.**

Selection inference

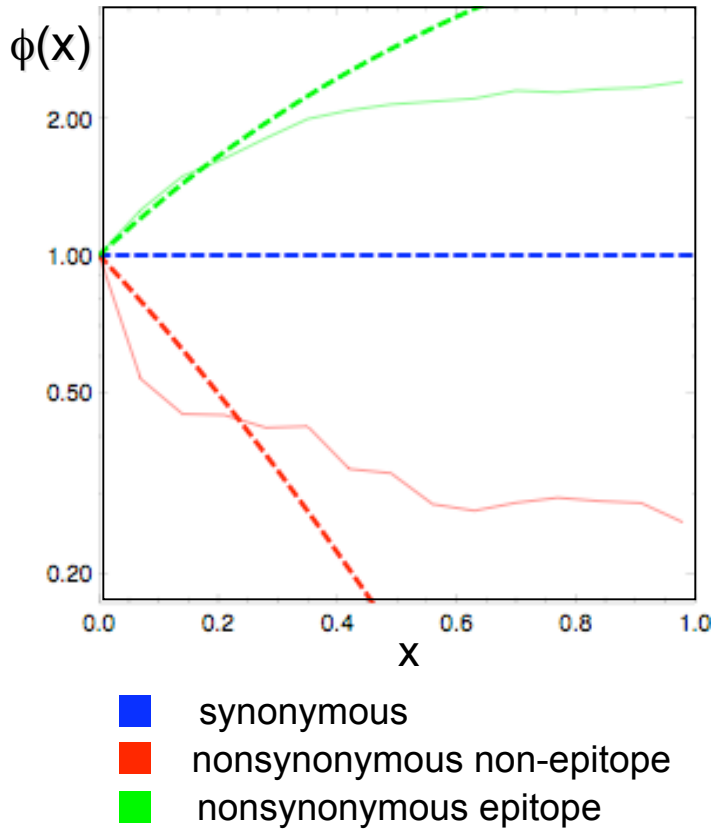
- **Epitope** aminoacid changes are under **positive selection:**

- fraction of selected substitutions:

$> \phi(1) - 1 = 50\%$

- strength of selection:

$s > s^* = 1/3$ yrs

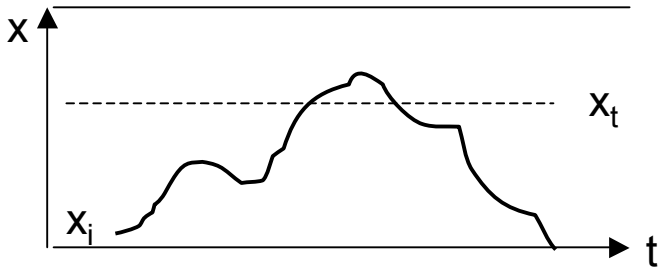


→ **sweeps are driven by combinations of beneficial mutations.**

Selection and viral ecology

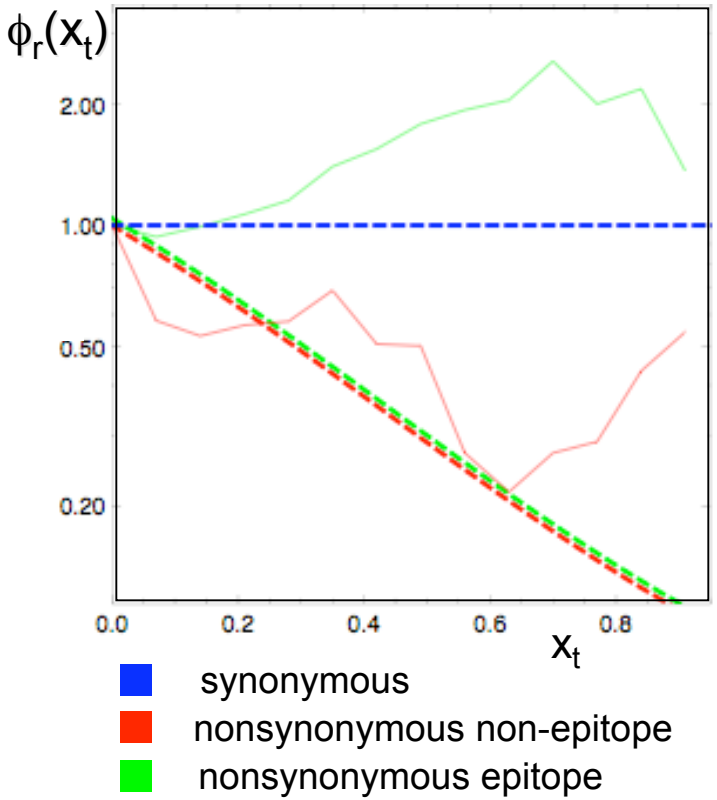
- **Microevolutionary fitness changes**
  are measured by **return propagator** of lost alleles

$$G_r (x_t \mid x_i)$$



propagator ratio between sequence classes

$$\phi_r (x_t) = G_r (x_t \mid x_i) / G_{0r} (x_t \mid x_i) \quad (x_i << 1)$$



- **Overrepresentation of high frequencies**
  at intermediate times ($\phi_r > 1$)
  - is incompatible with directional selection
  - shows **red queen effect**:
    decay of fitness due to host immunity.

Conclusion

- Influenza evolution is shaped by
  - **strong selection,**
  - **linkage,**
  - **high mutation rates.**

- **Stochastic clonal interference:**
  strong sweeps governed by **multiple driving mutations.**

- **Selection is dynamic:** host immunity generates **red queen race.**

- **This system will lead to new concepts in population genetics and should become a model organism.**

- **Red queen race:** micro-evolutionary predictability or epistasis?