

Evolutionary dynamics of one million barcoded lineages

Sasha Levy, Jamie Blundell,
Sandeep Venkataram, Dmitri Petrov,
Daniel Fisher, Gavin Sherlock

Laufer Center for Physical and Quantitative Biology
Stony Brook University

Departments of Genetics, Biology and
Applied Physics
Stanford University

Overview

I. Lineage tracking with random barcodes

- Why more is better
- Random barcode technology
- Serial batch evolution

II. Double barcodes and their uses

- Systematic analysis of epistasis
- Quantitative genetics
- Pairwise interactions (protein-protein)

Evolution of large asexual populations underlies diseases

Bacterial Pathogenesis



Salmonella



E. coli

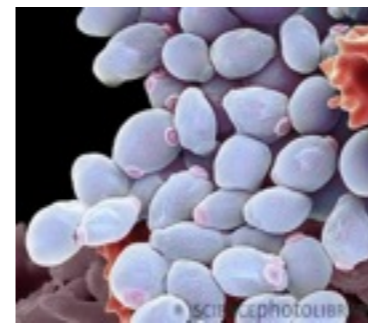


Staphylococcus



Tuberculosis

Yeast Pathogenesis



Candida albicans

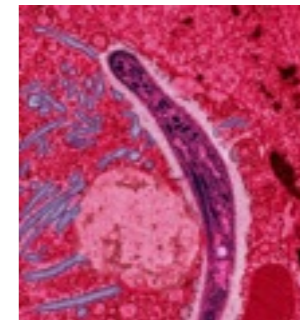


Candida glabrata

Protist Pathogenesis

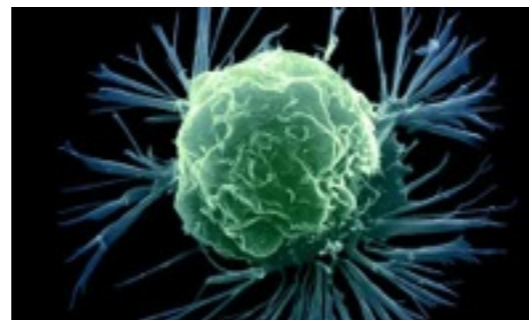


Giardia

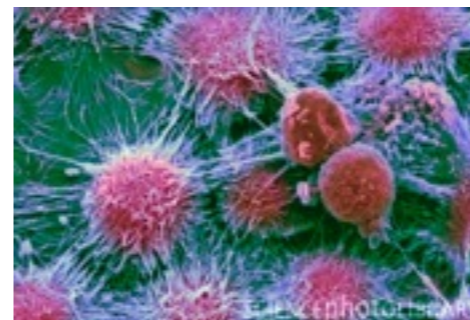


Malaria

Cancer Progression



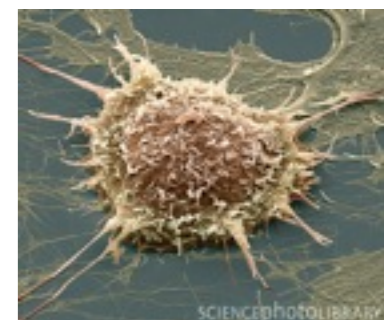
Breast



Kidney



Ovarian



Cervical

What factors impact the dynamics of evolution?

- 1) Distribution Mutation Rates to Each Fitness Effect
- 2) Population Size
- 3) Changing Environments
- 4) Population Structure
- 5) Epistasis
- 6) Sex
- 7) Ecology

Batch evolutions simplify the important factors

1) Distribution Mutation Rates to Each Fitness Effect

2) Population Size

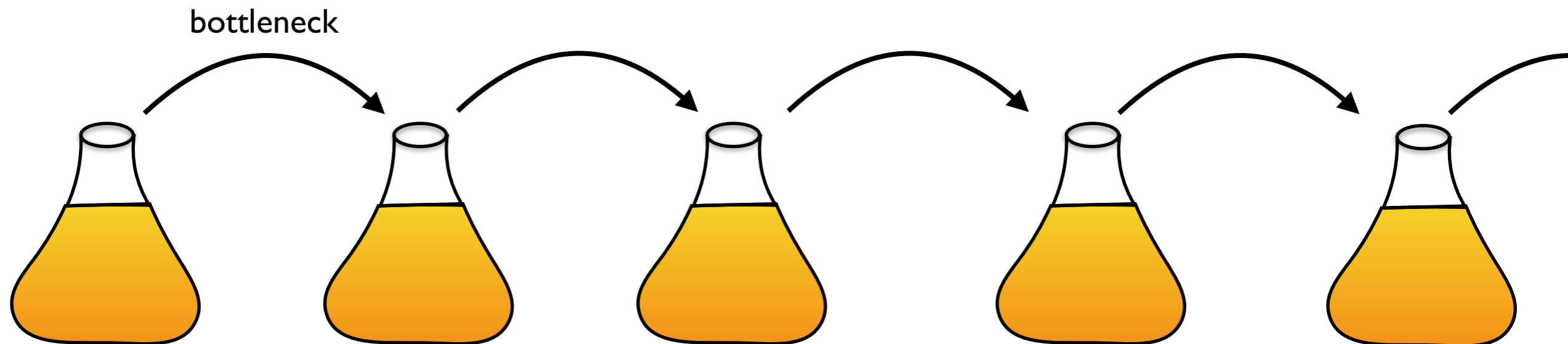
~~3) Changing Environments~~

~~4) Population Structure~~

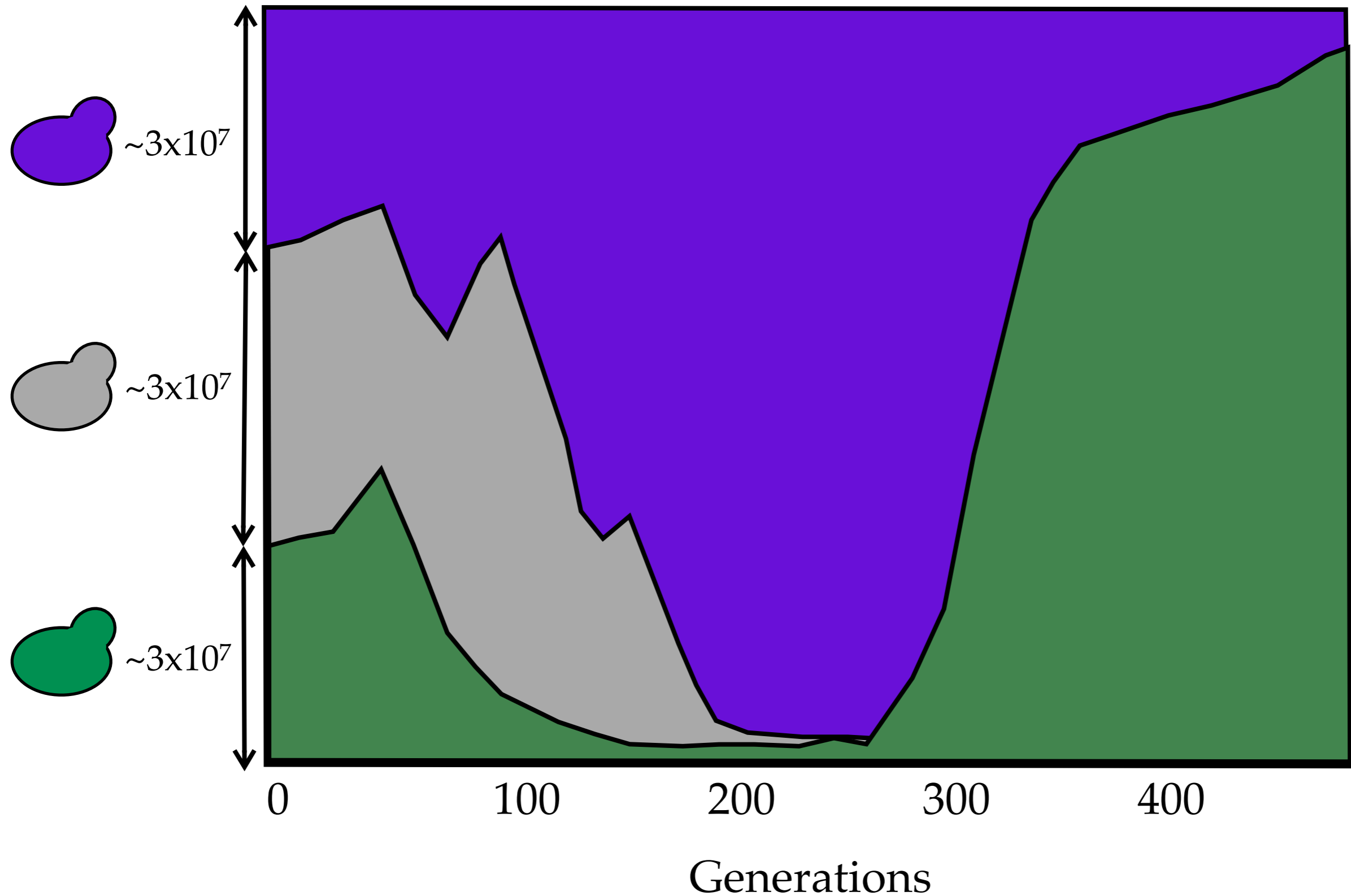
~~5) Epistasis~~

~~6) Sex~~

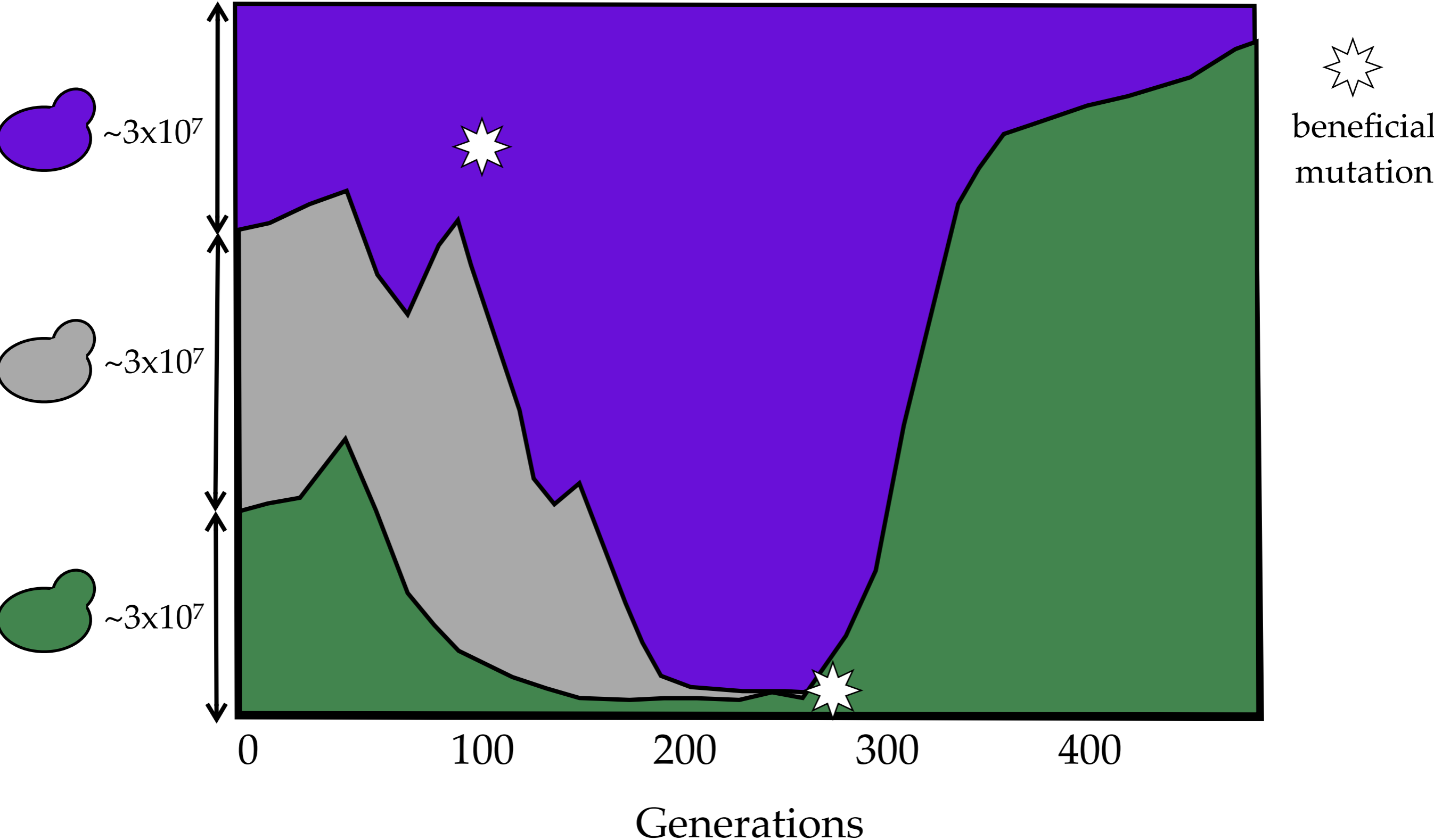
~~7) Ecology~~



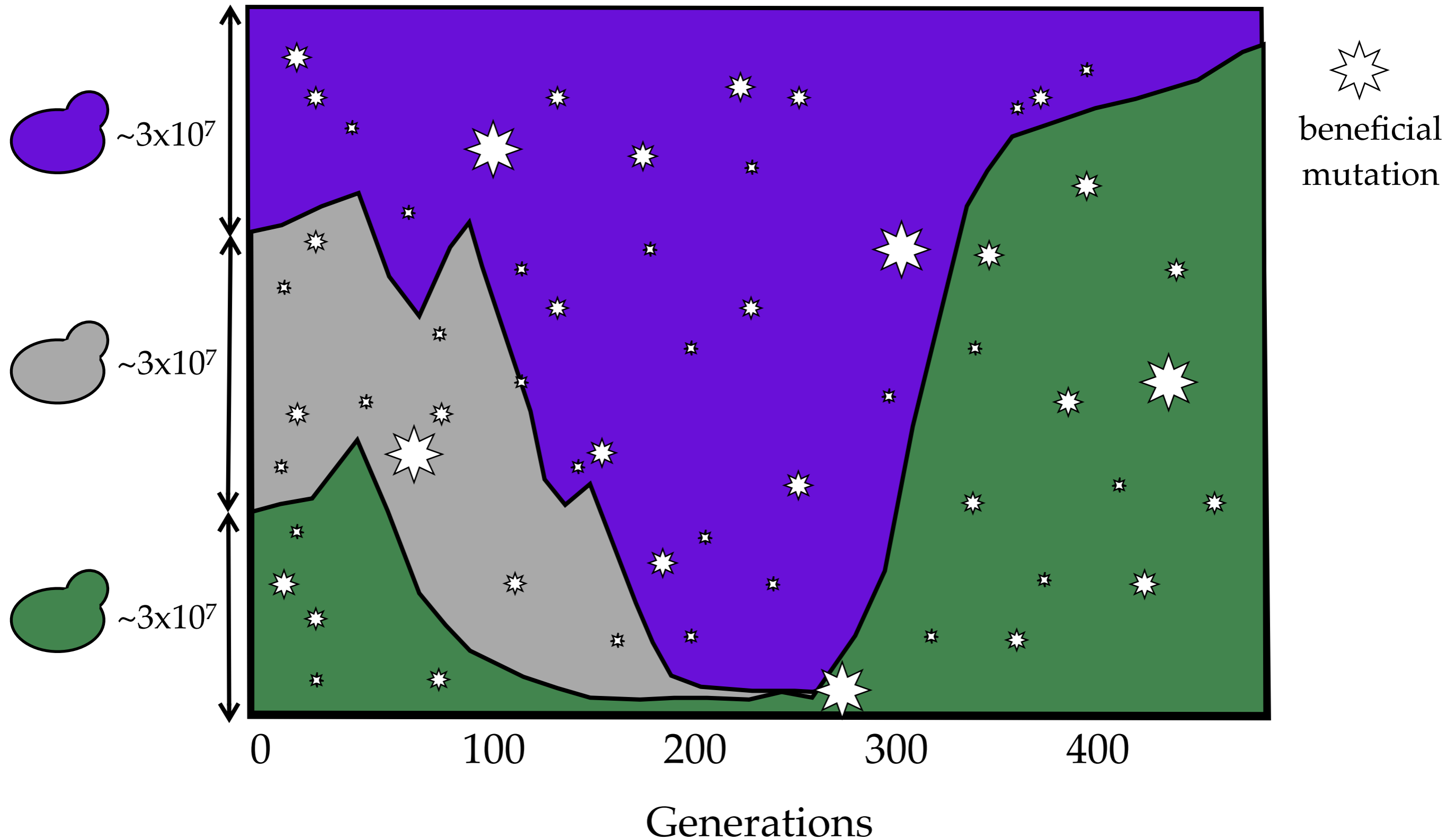
Dynamics of three large lineages of evolving yeast



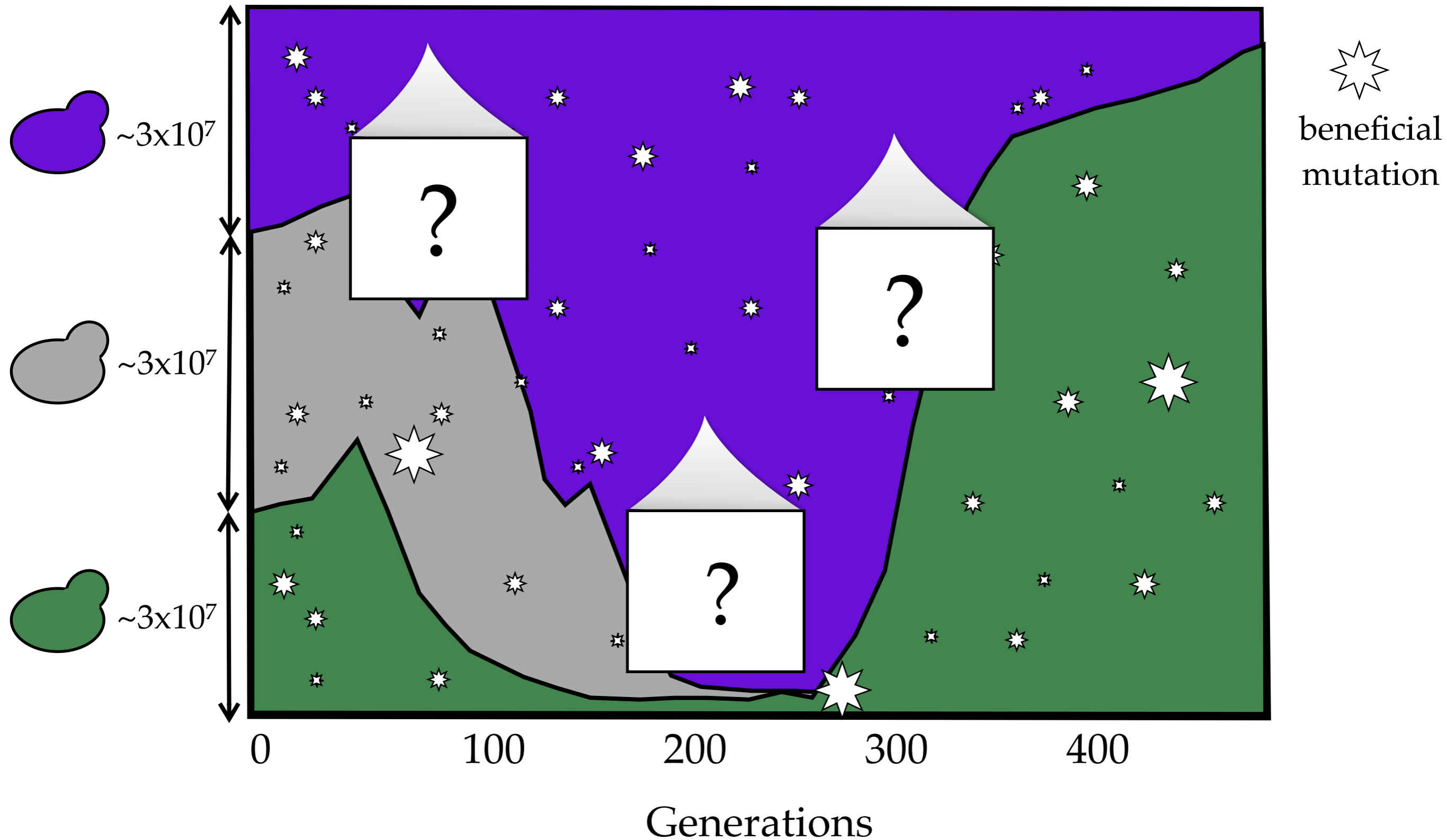
Beneficial mutations cause changes in lineage frequencies



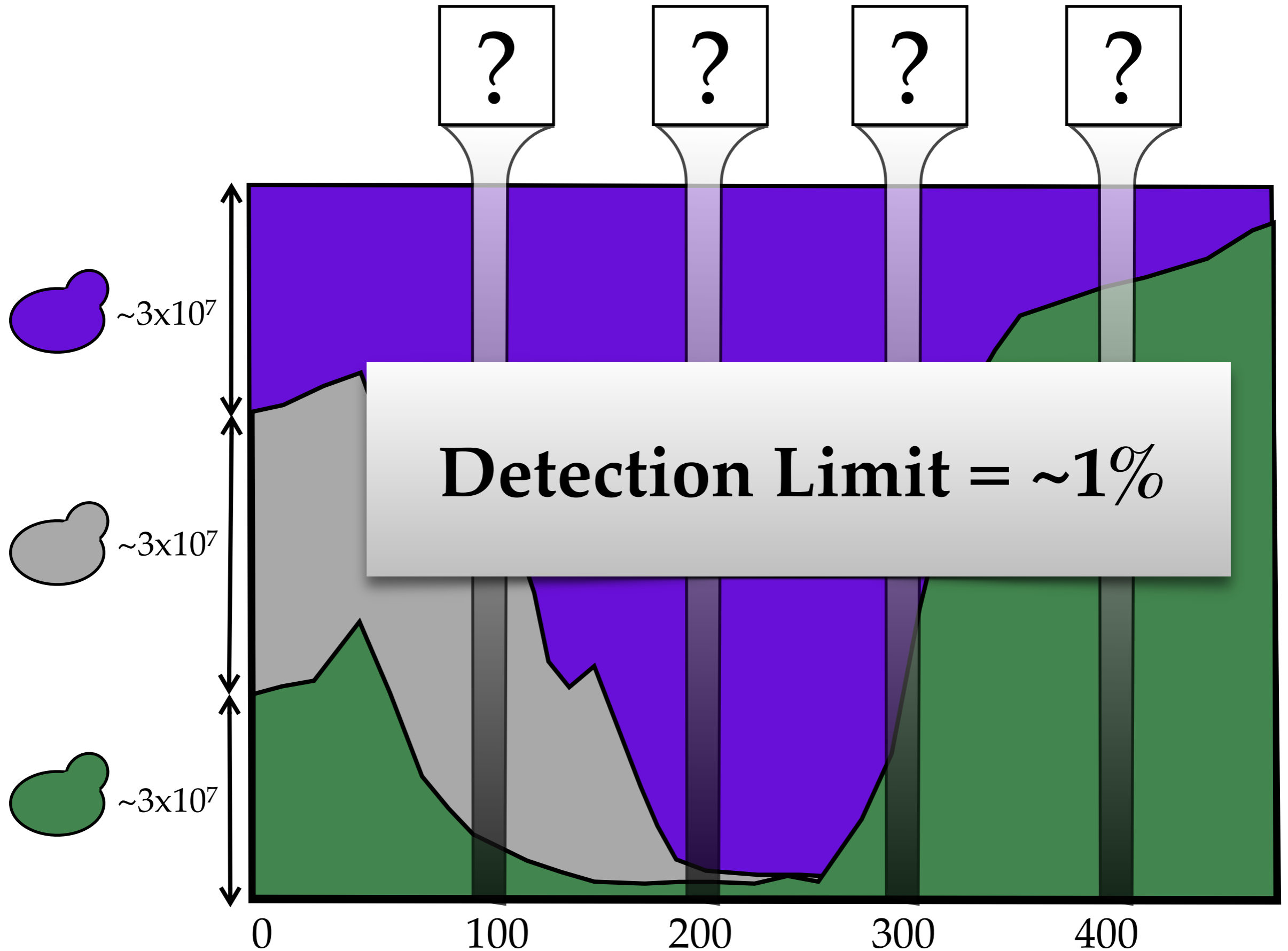
Clonal interference makes the dynamics difficult to understand



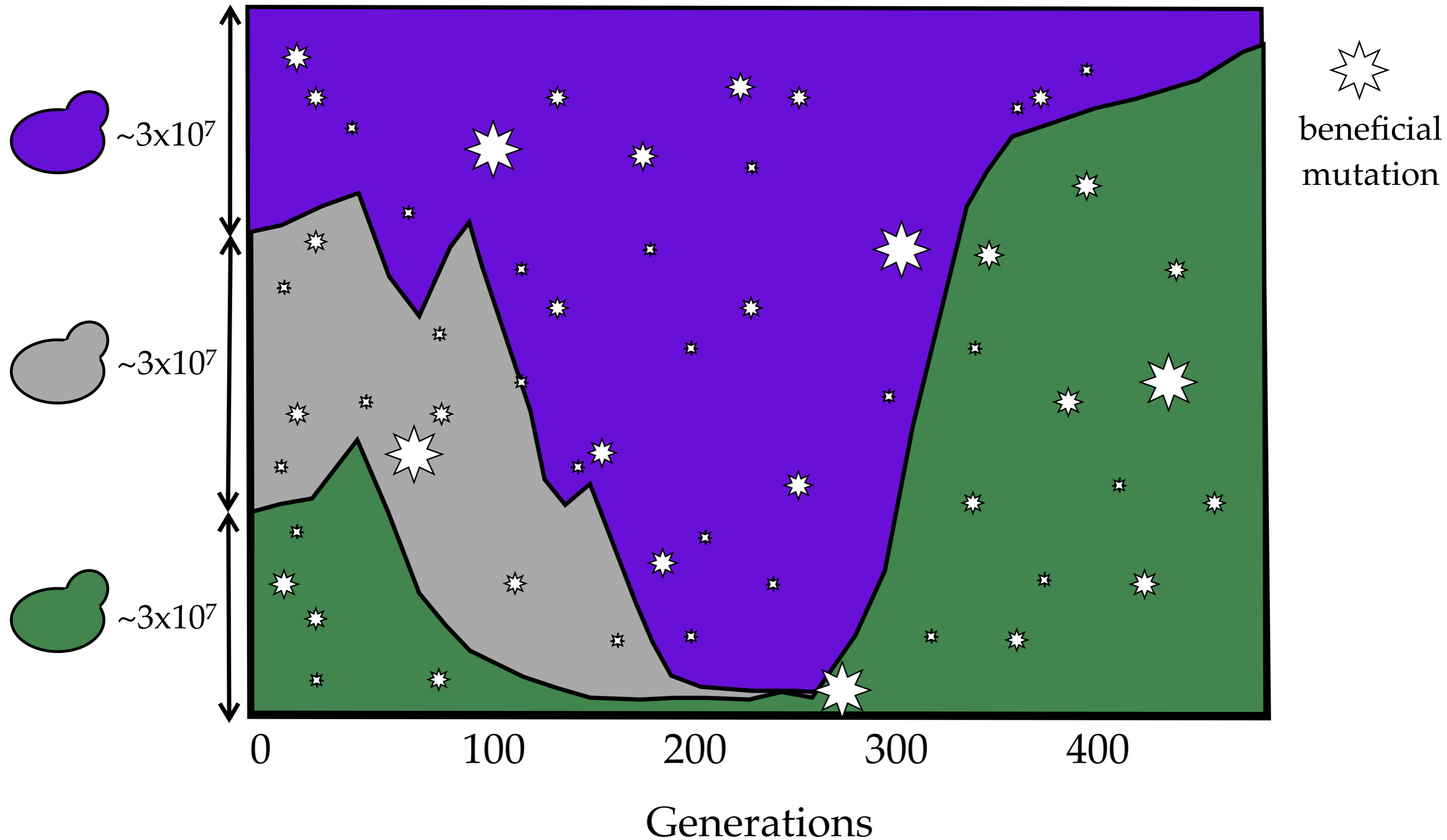
The random clone approach



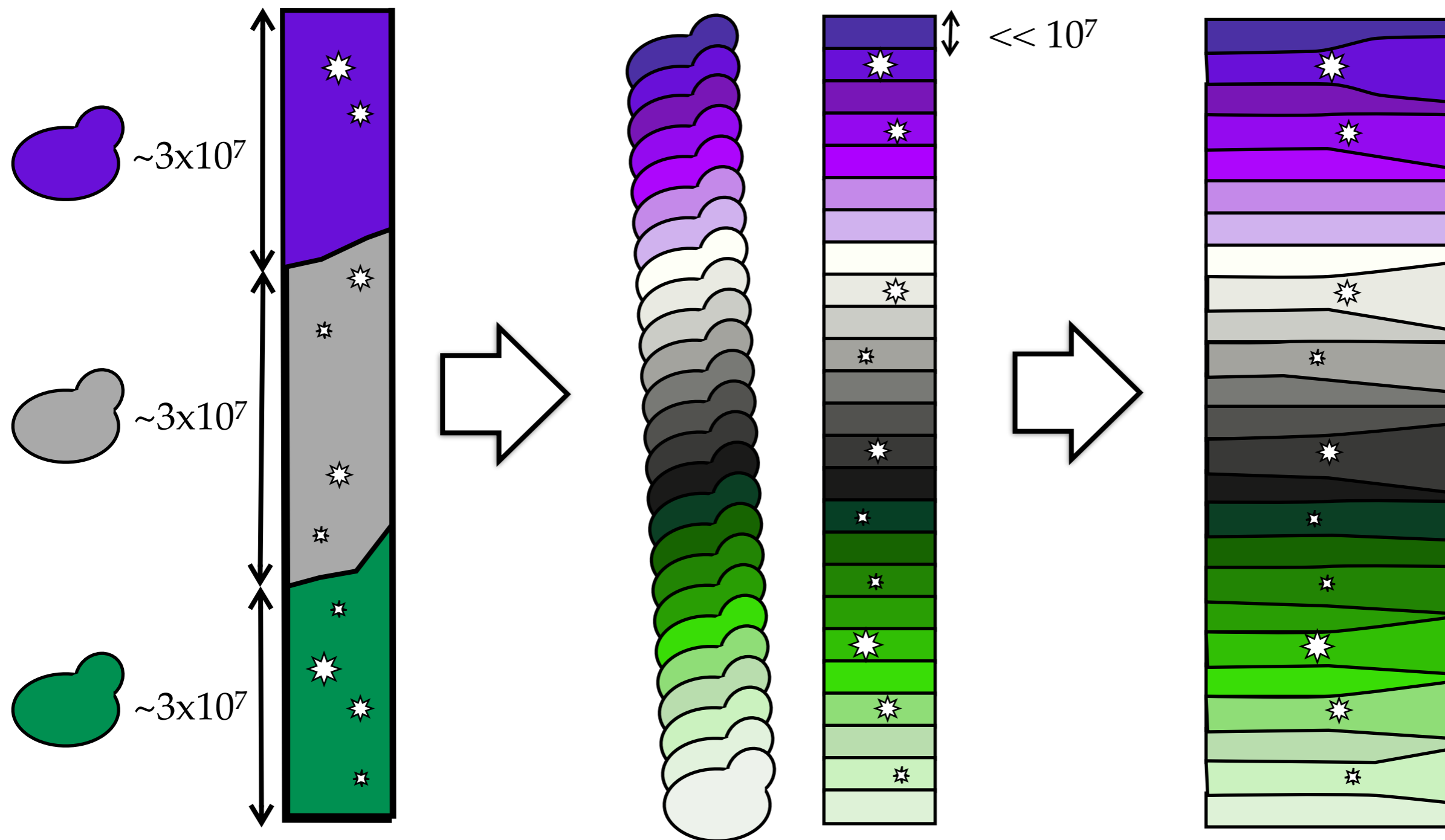
The population sequencing approach



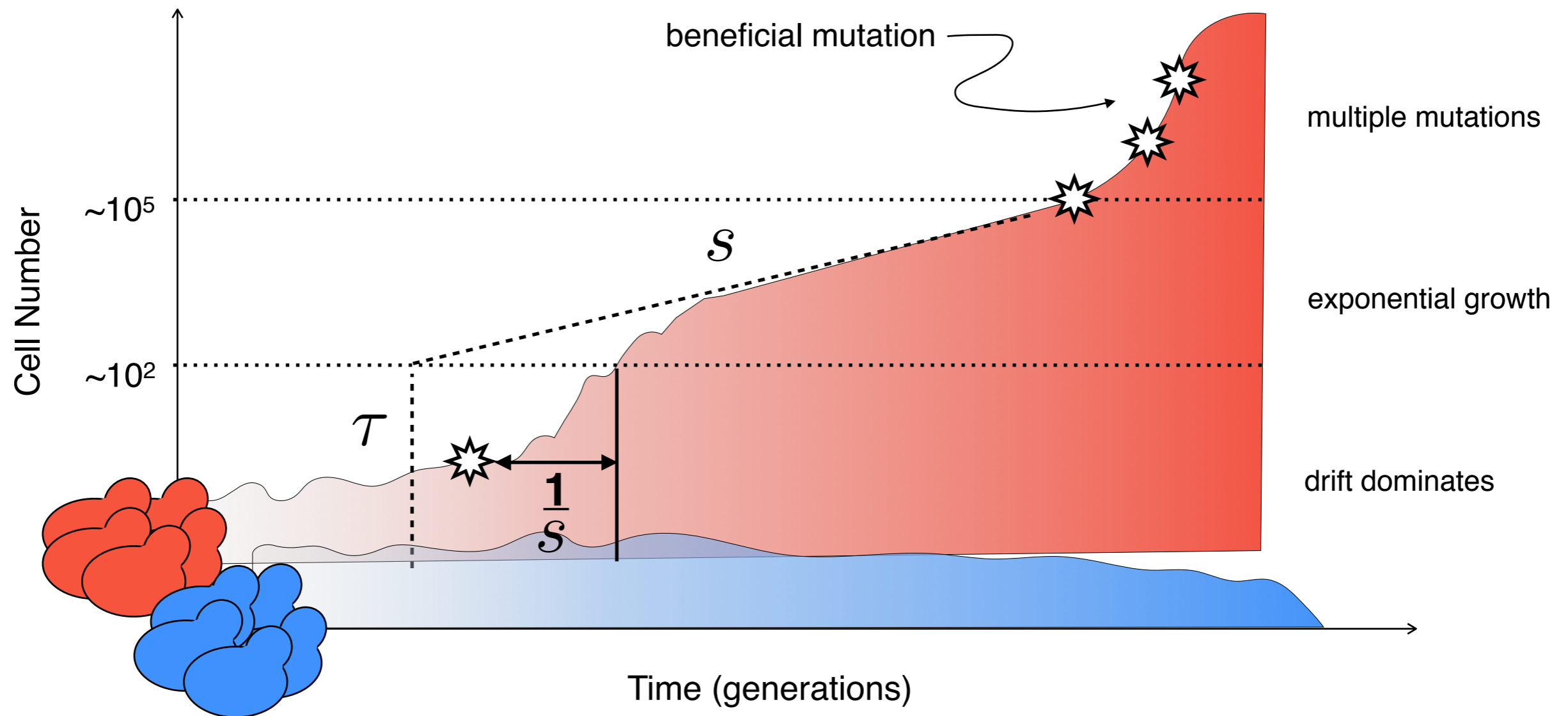
How do we study evolution more systematically?



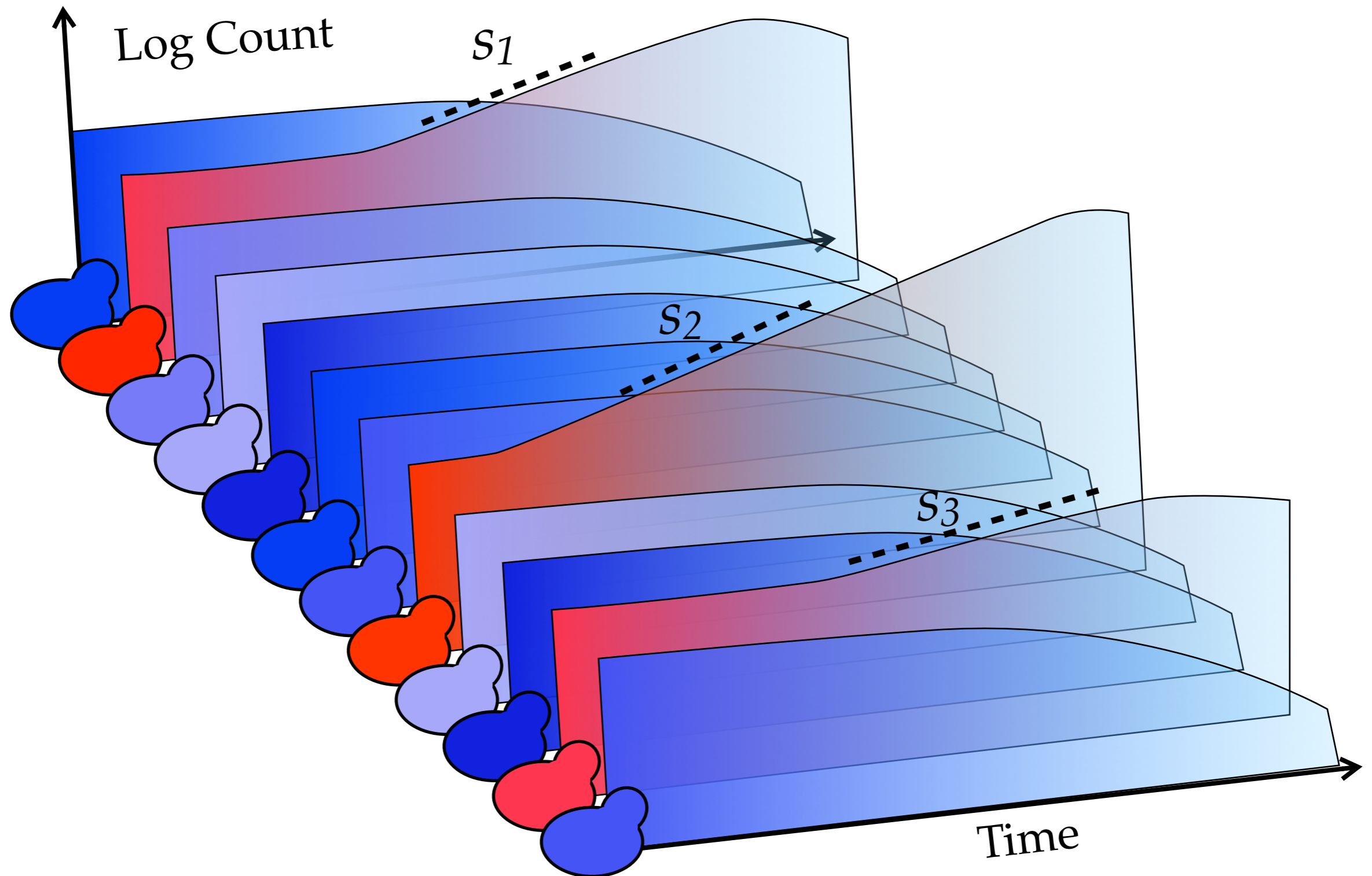
We could observe every beneficial mutation and its dynamics with smaller lineage sizes



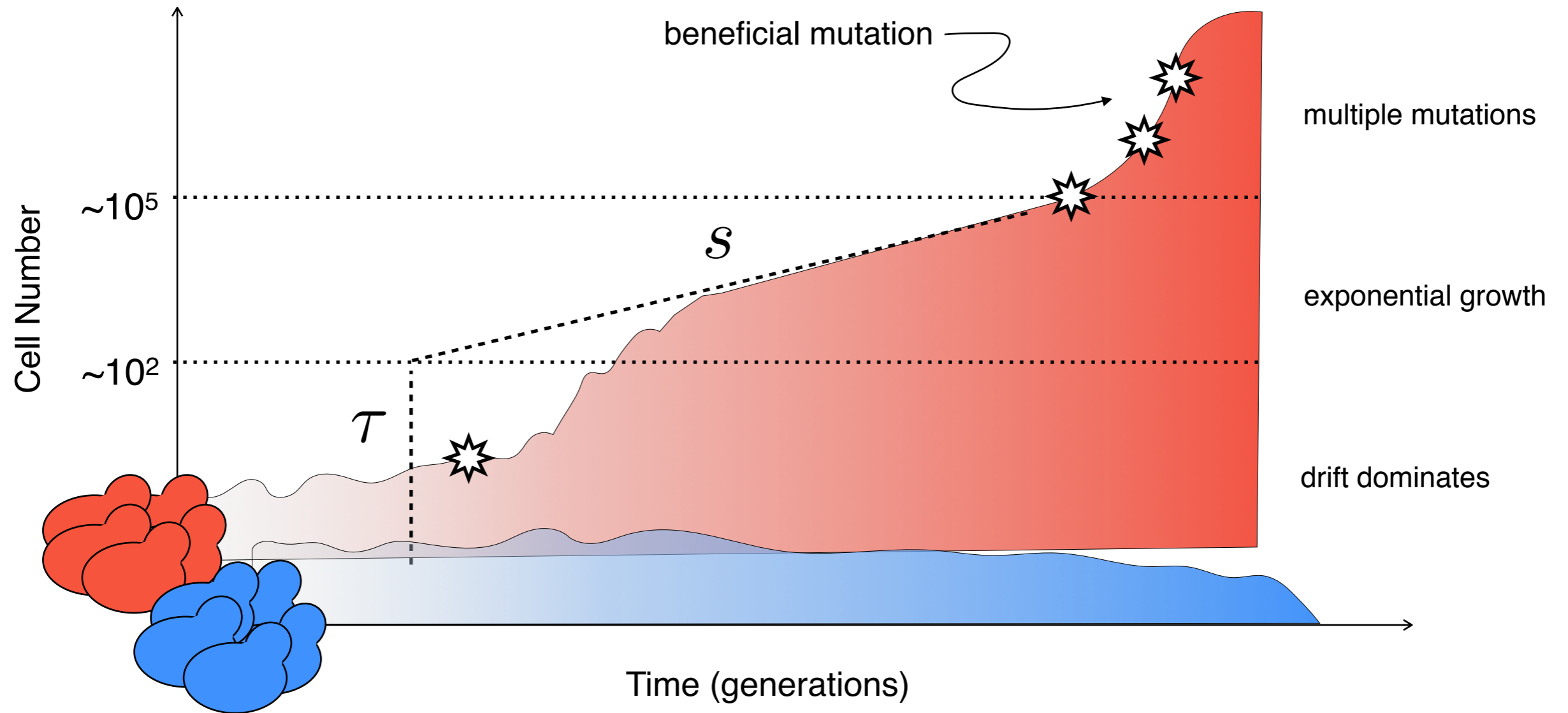
A window of exponential growth measures the fitness effect of a single beneficial mutation



Trajectories of many small lineages measures the distribution of fitness effects

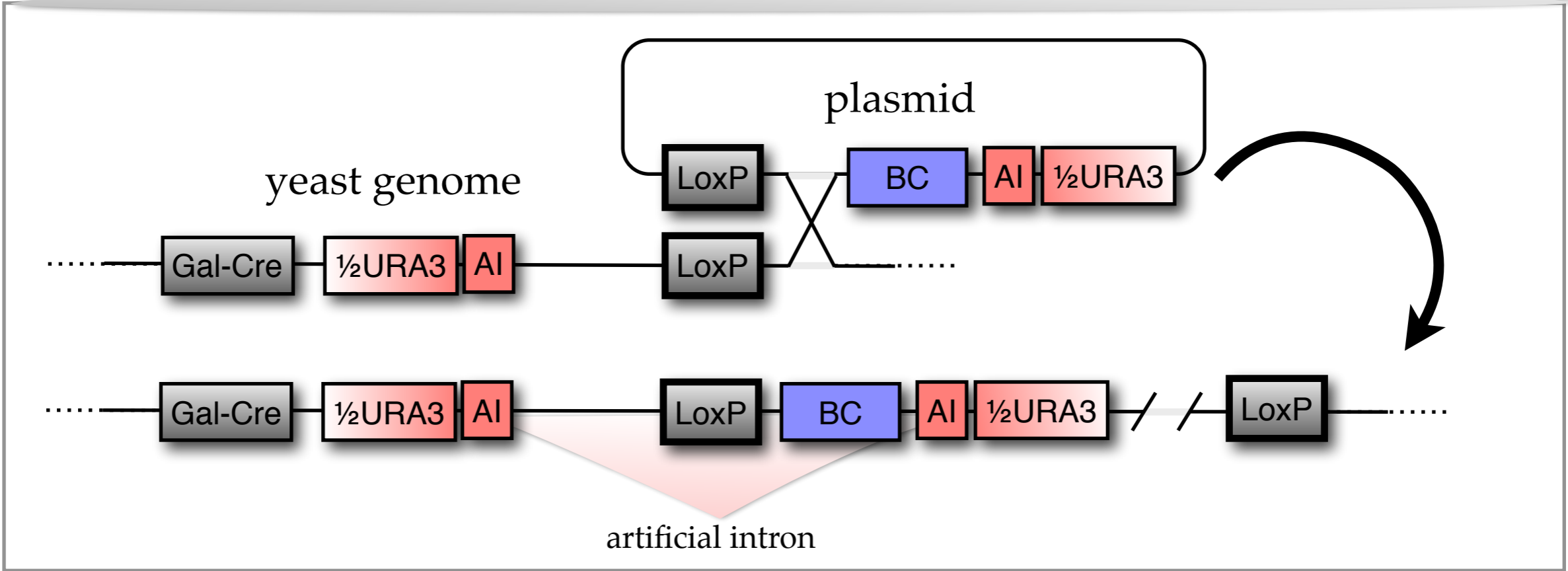
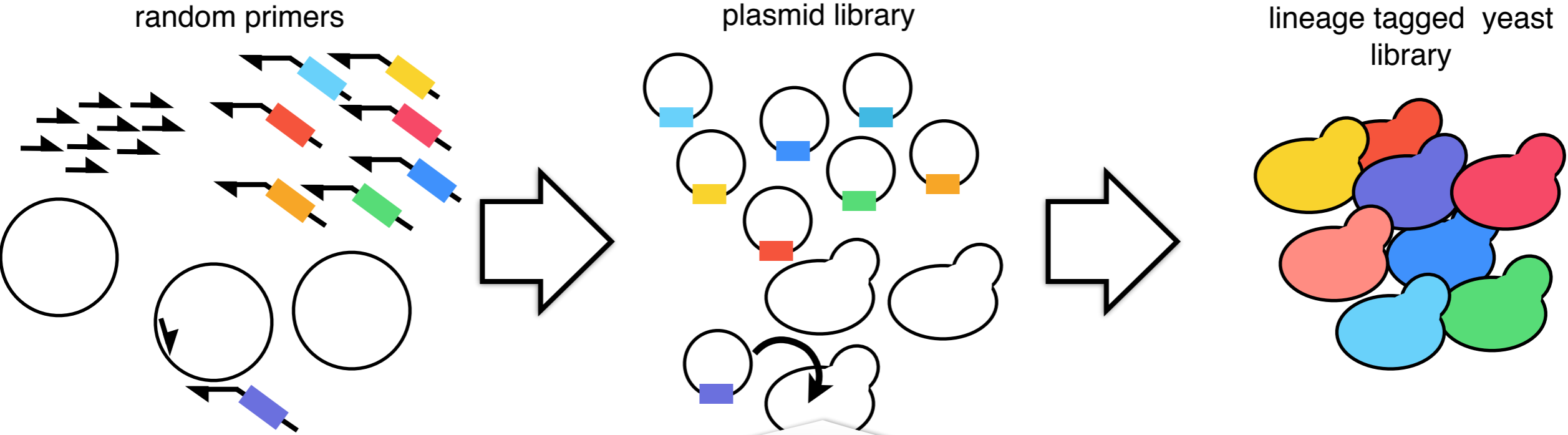


How many independent lineages are needed?

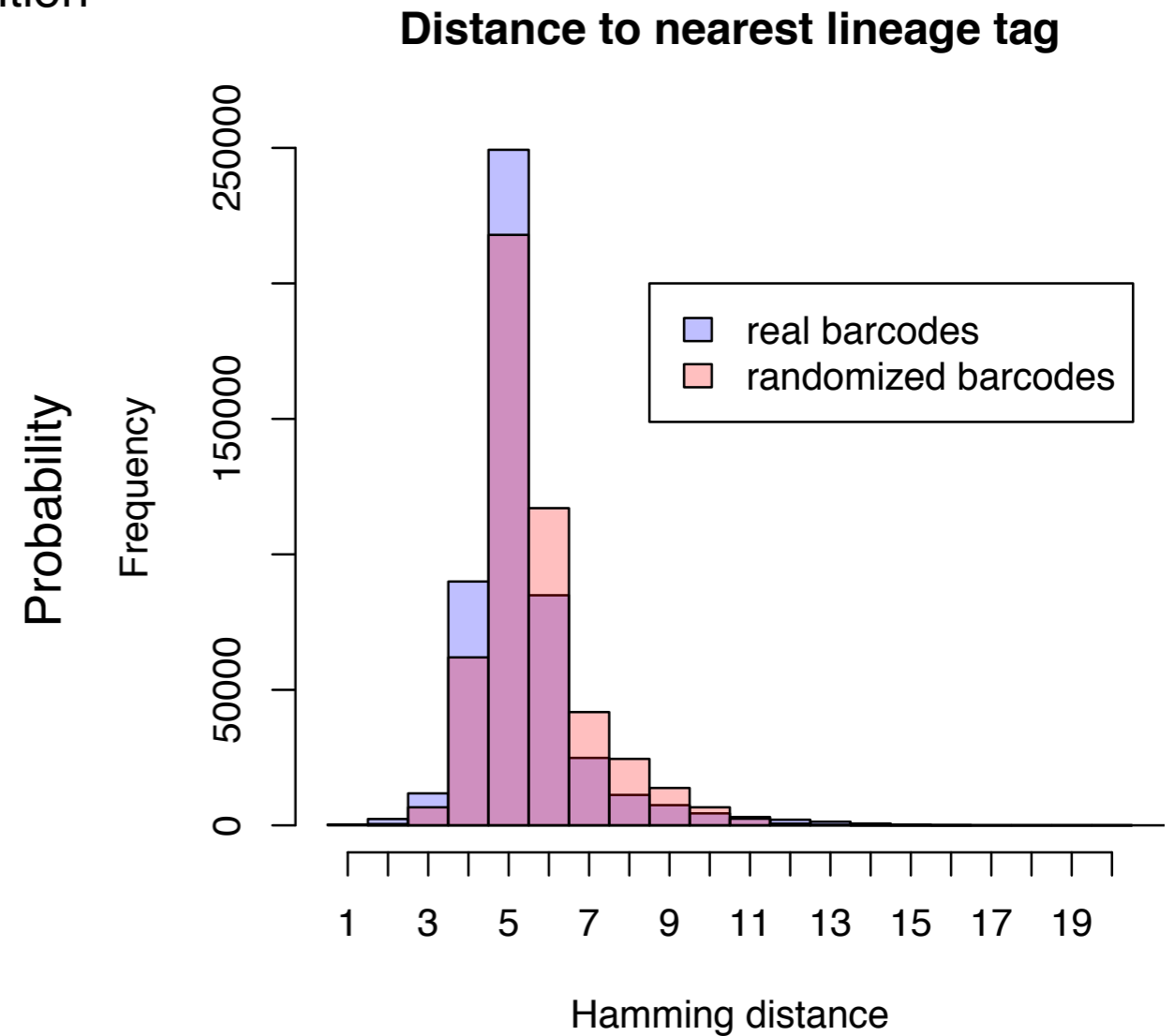
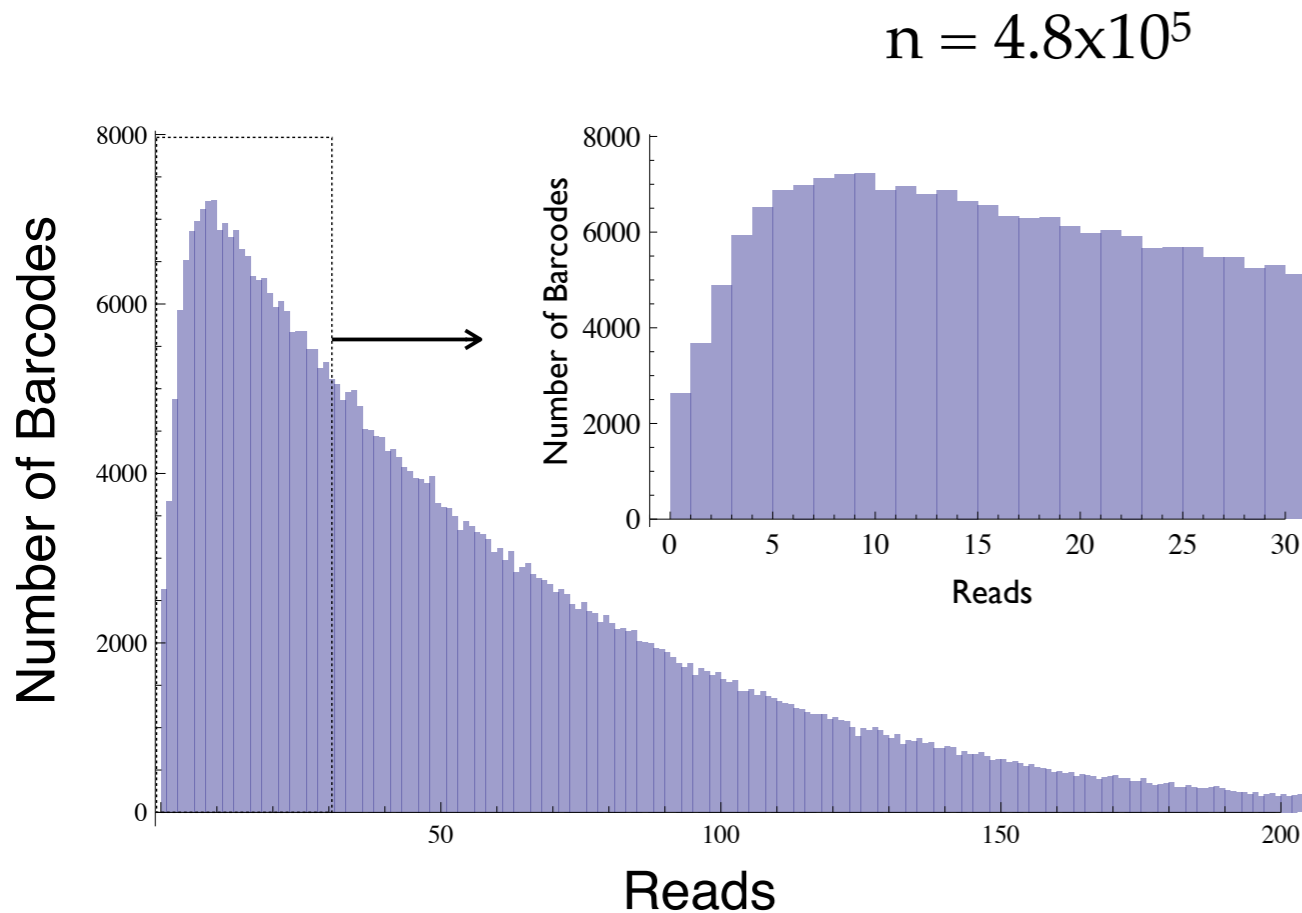
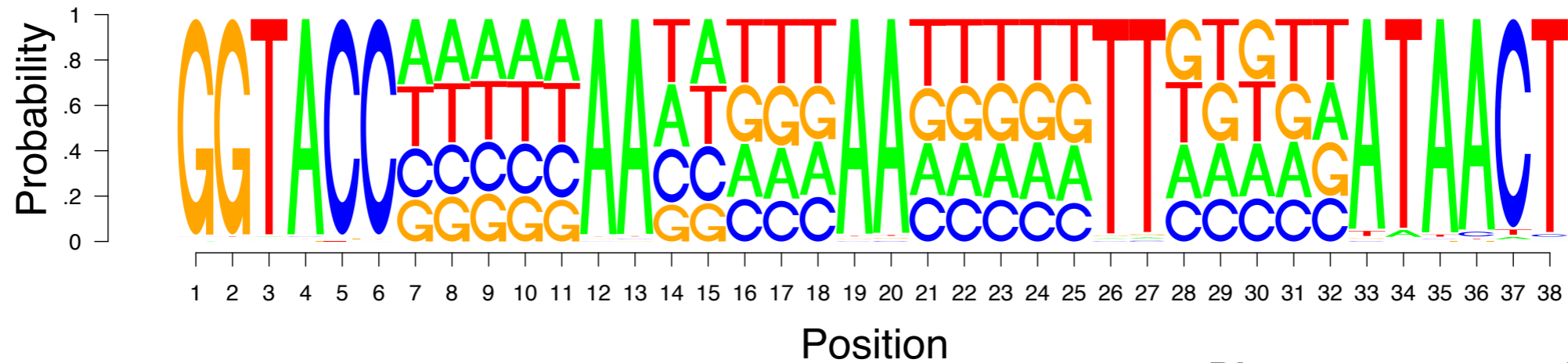


~Population size / 1000

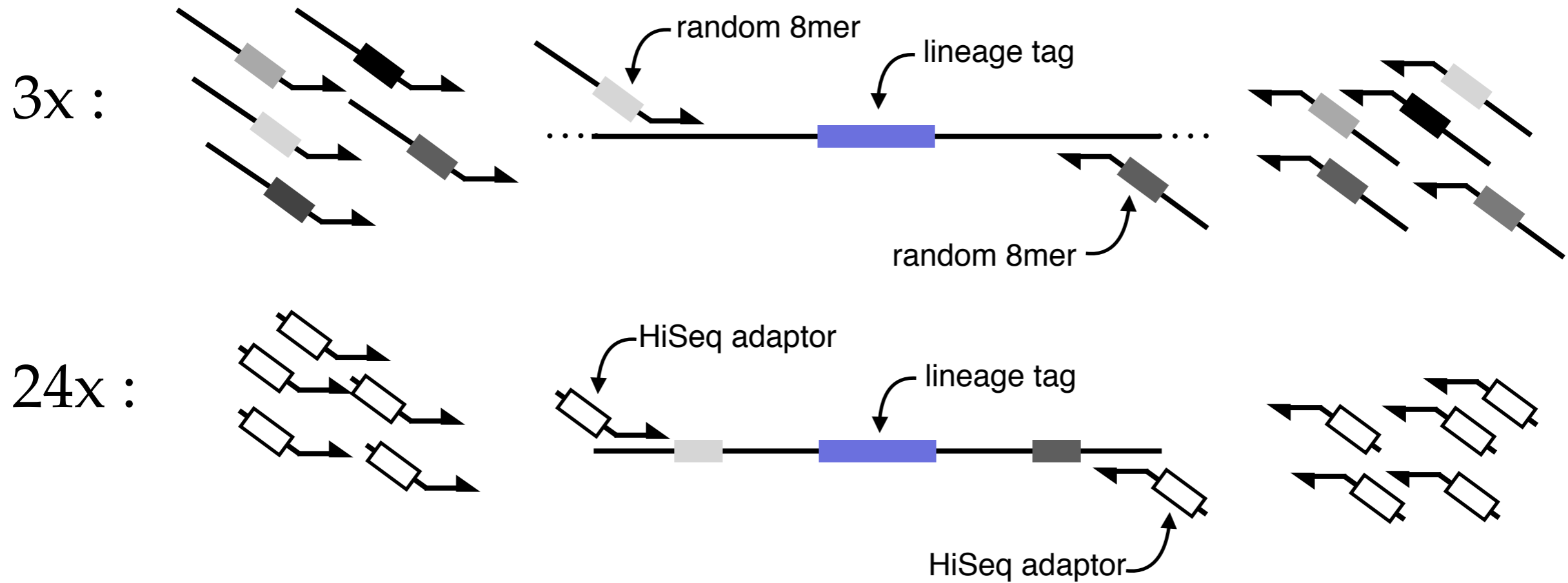
Construction of half a million lineage tags with random barcodes



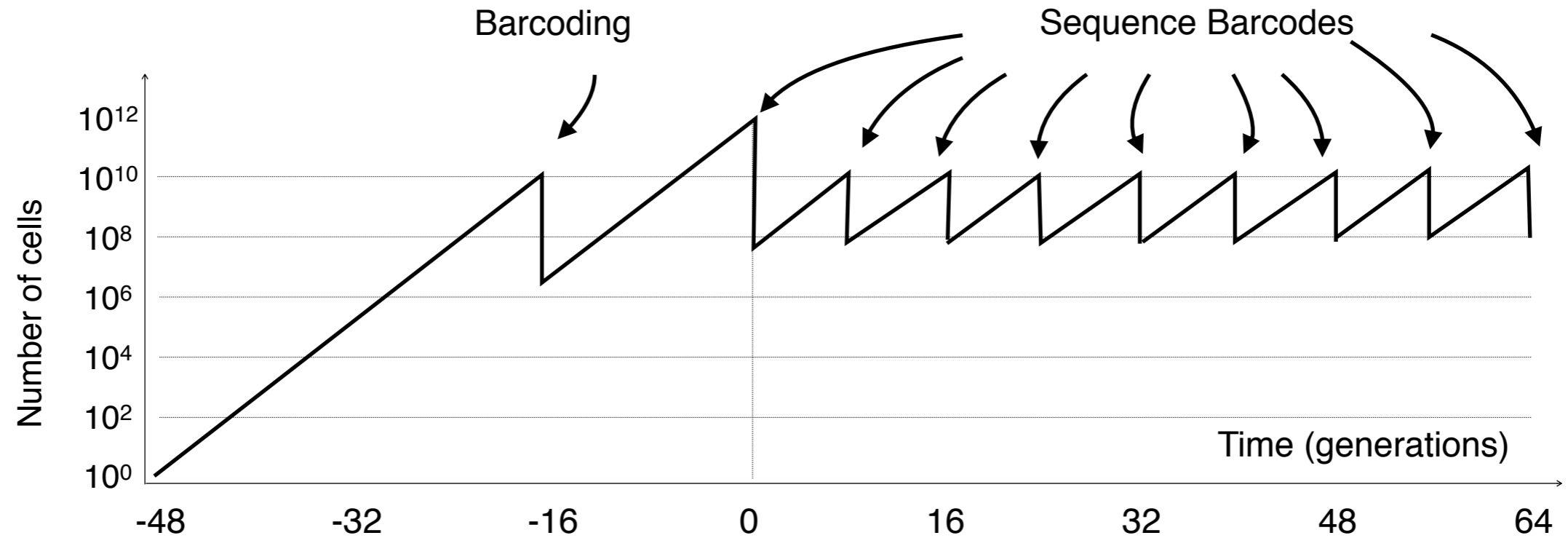
Lineage tag distribution and spacing



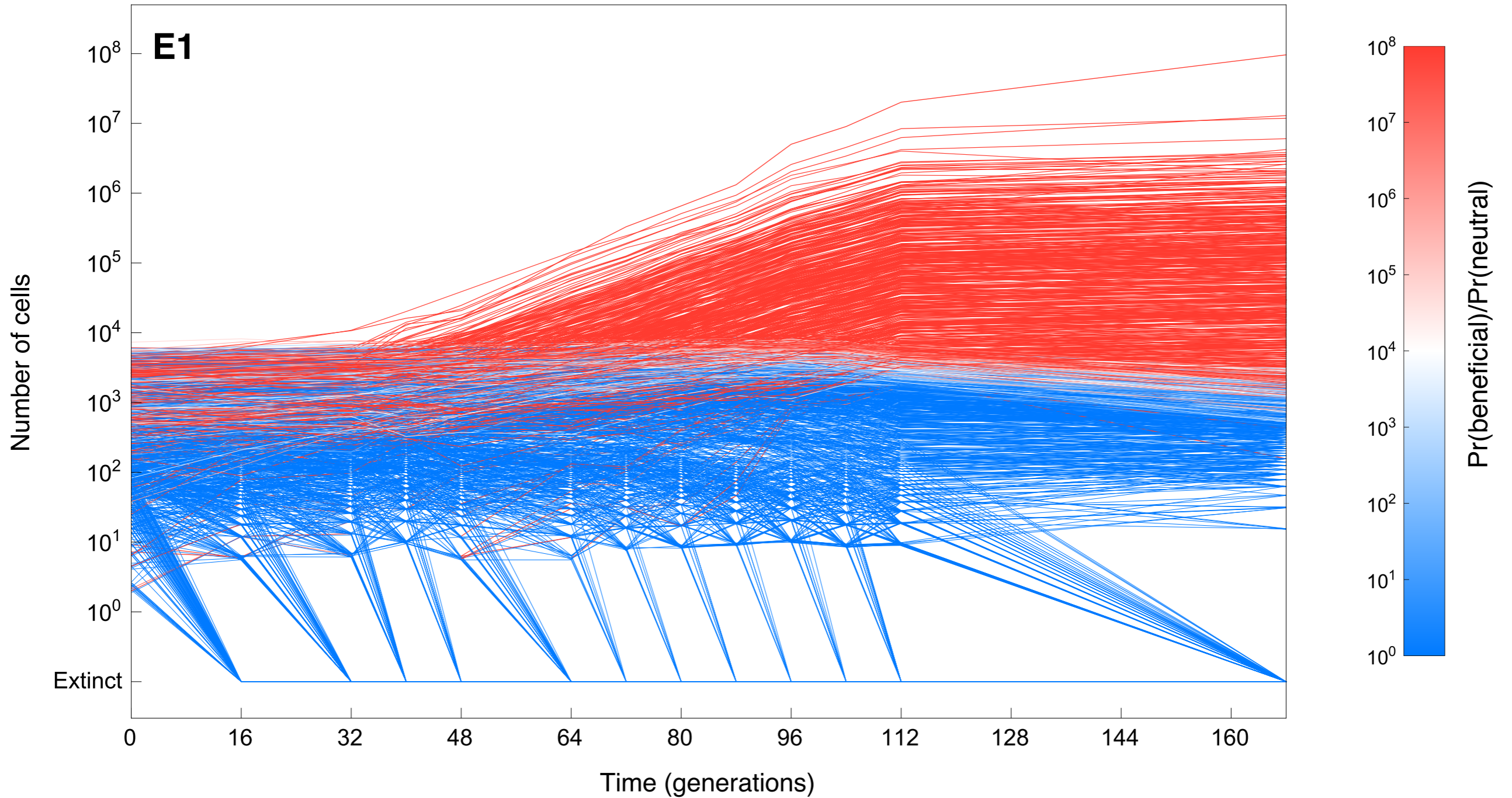
Accurate counting of lineage tags by next generation sequencing



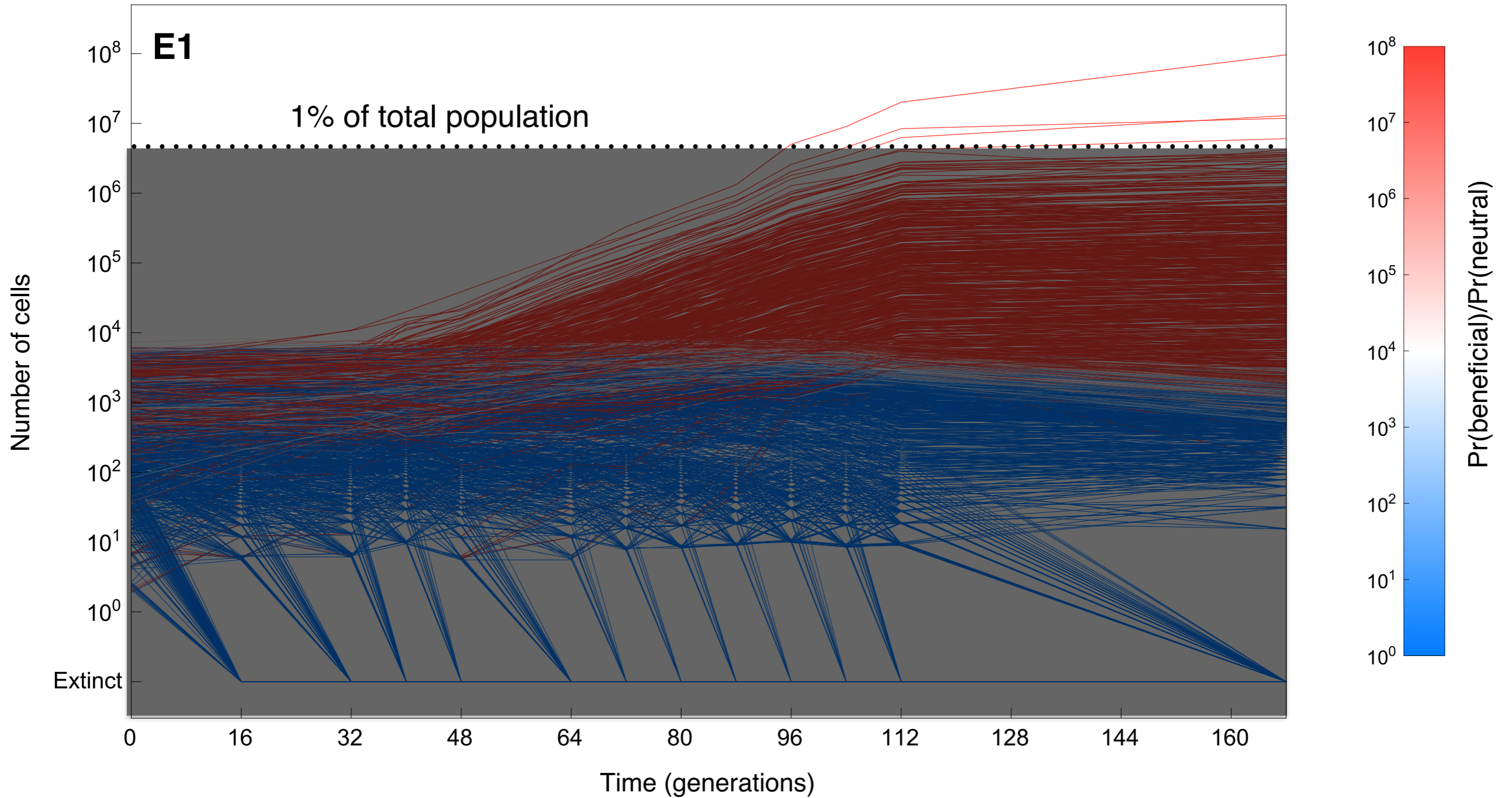
Experimental evolution



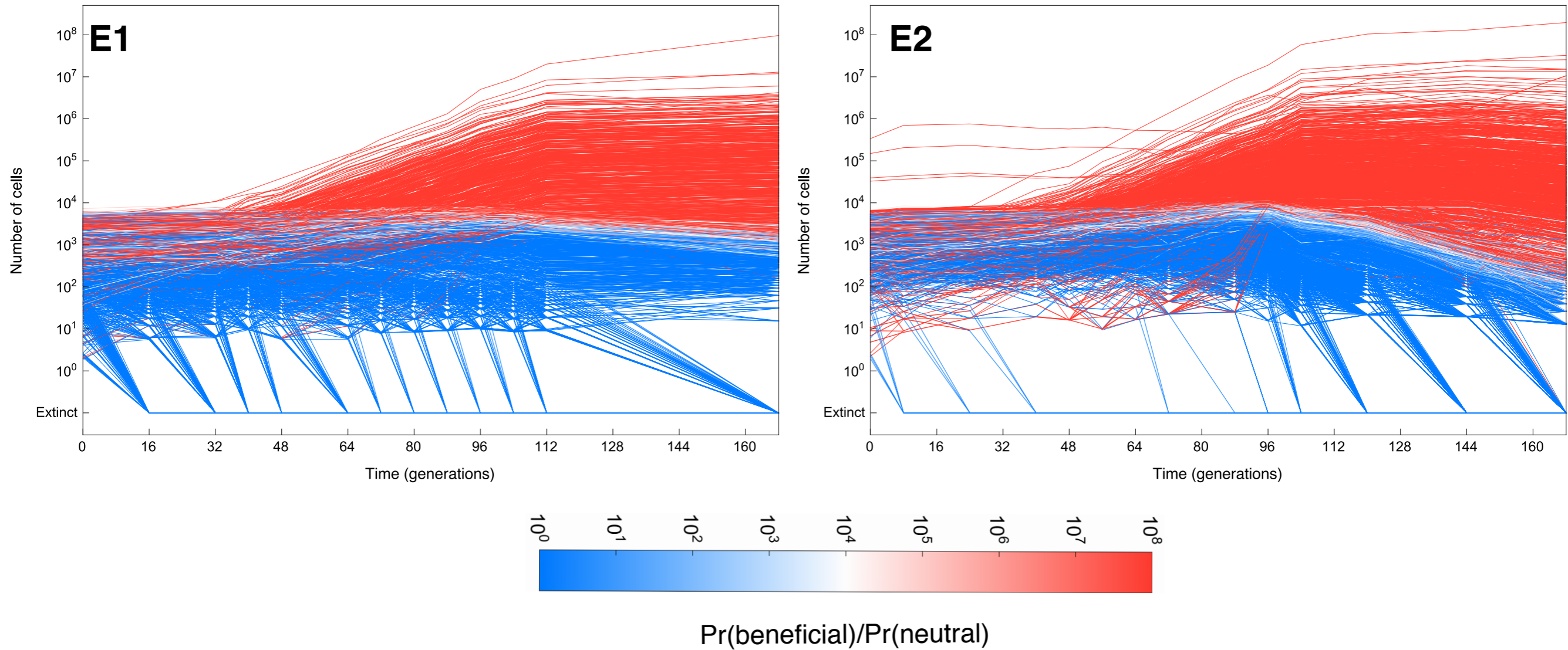
Trajectories of barcoded lineages



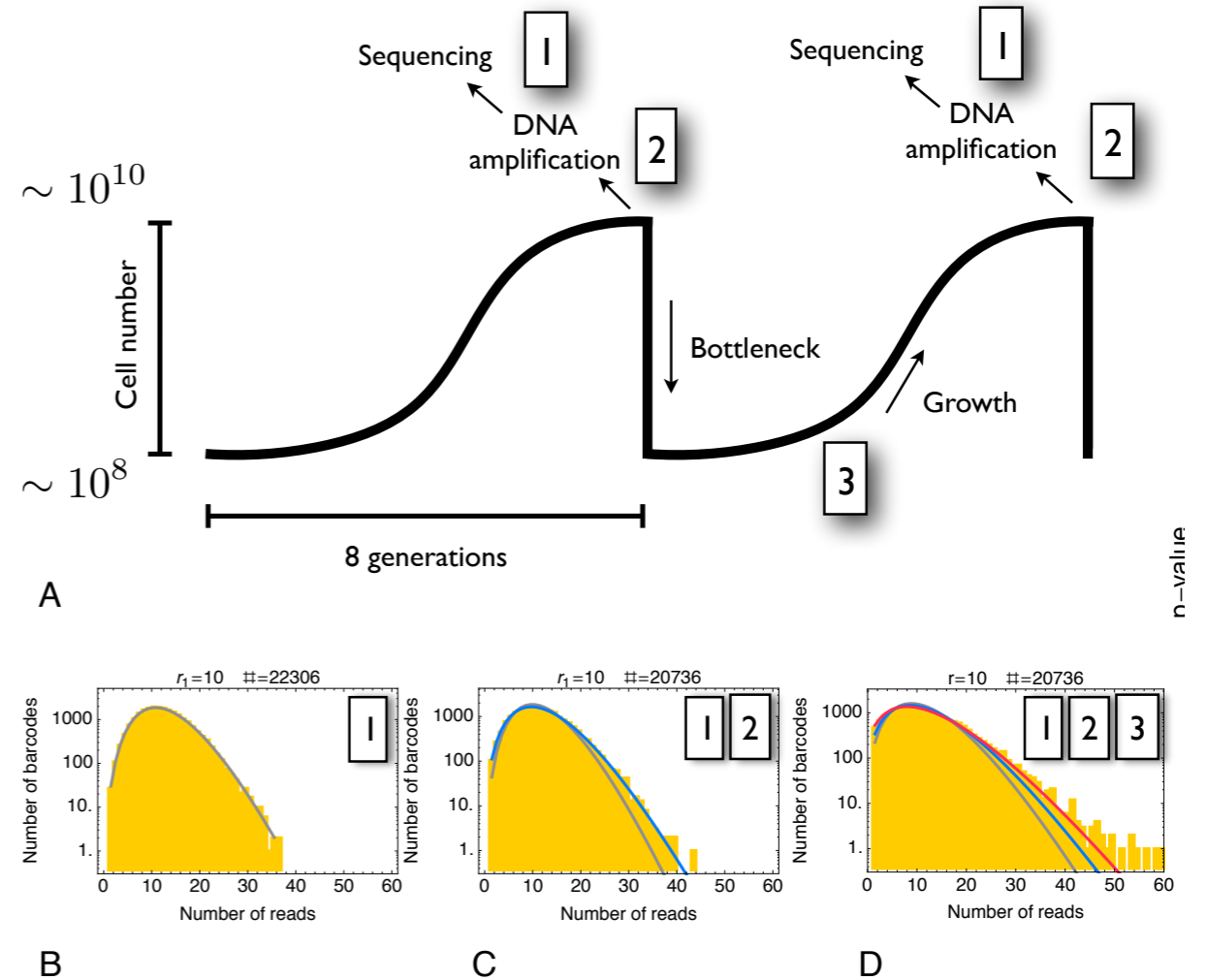
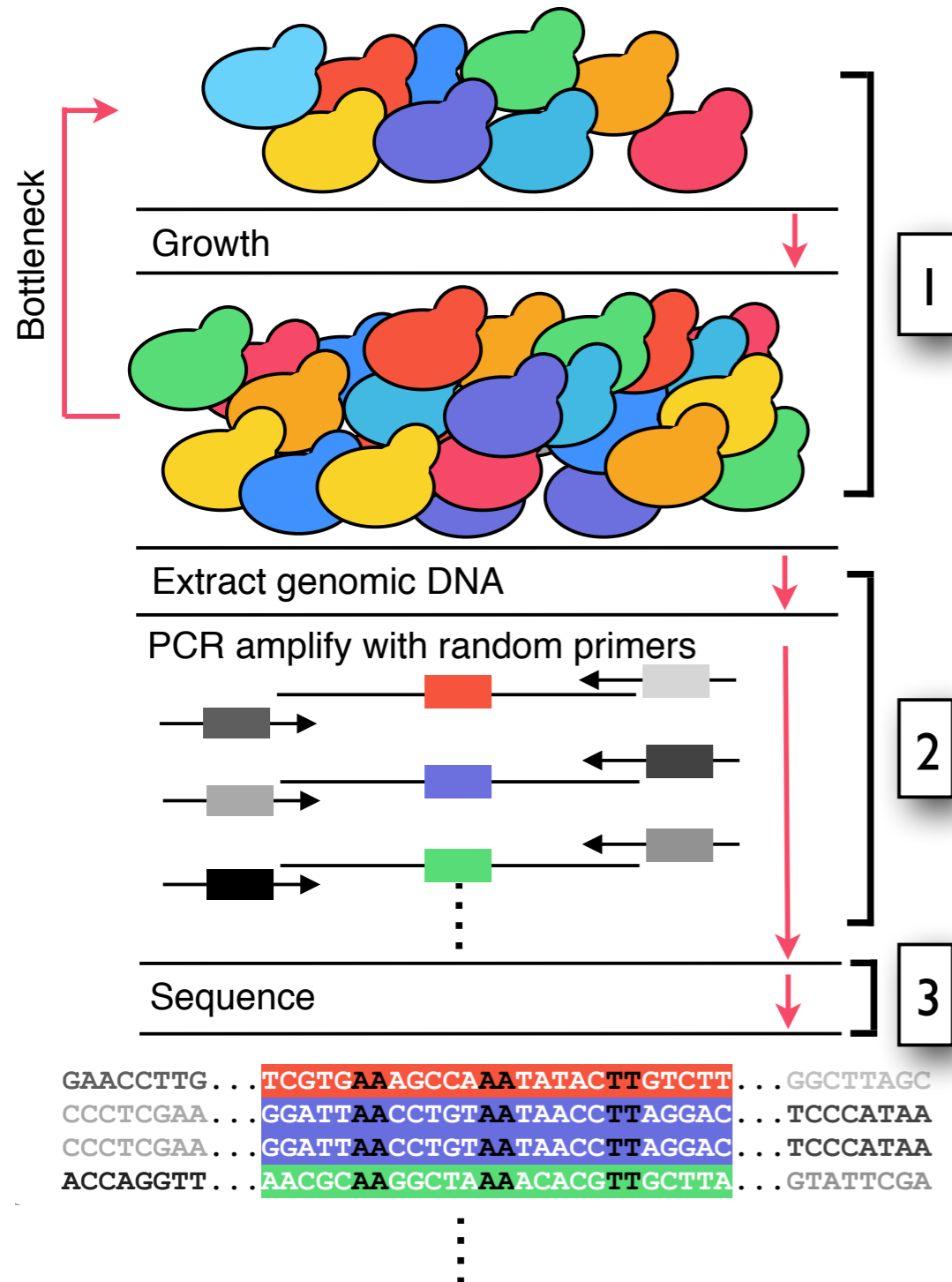
Trajectories of barcoded lineages



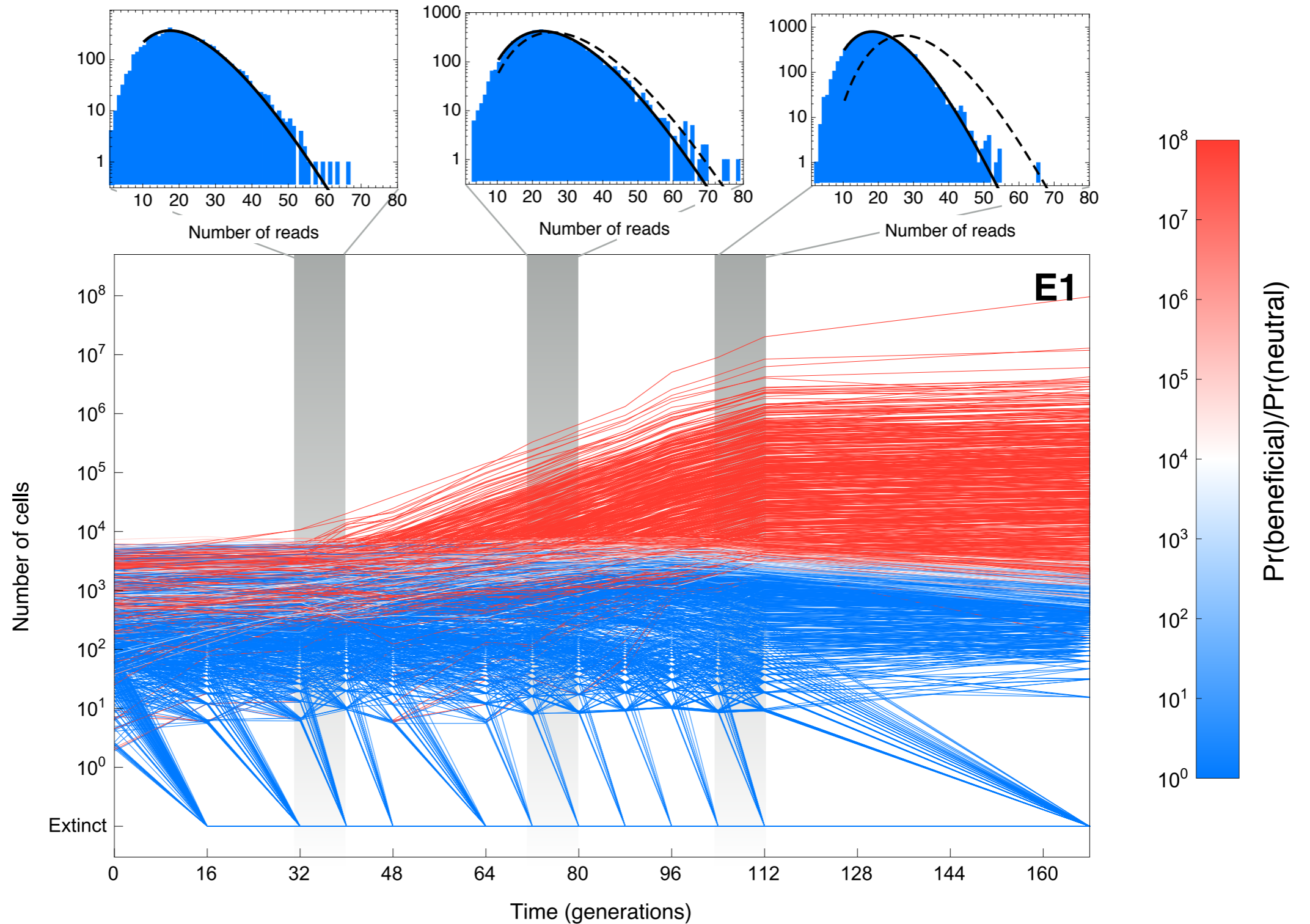
Evolutionary dynamics are grossly reproducible



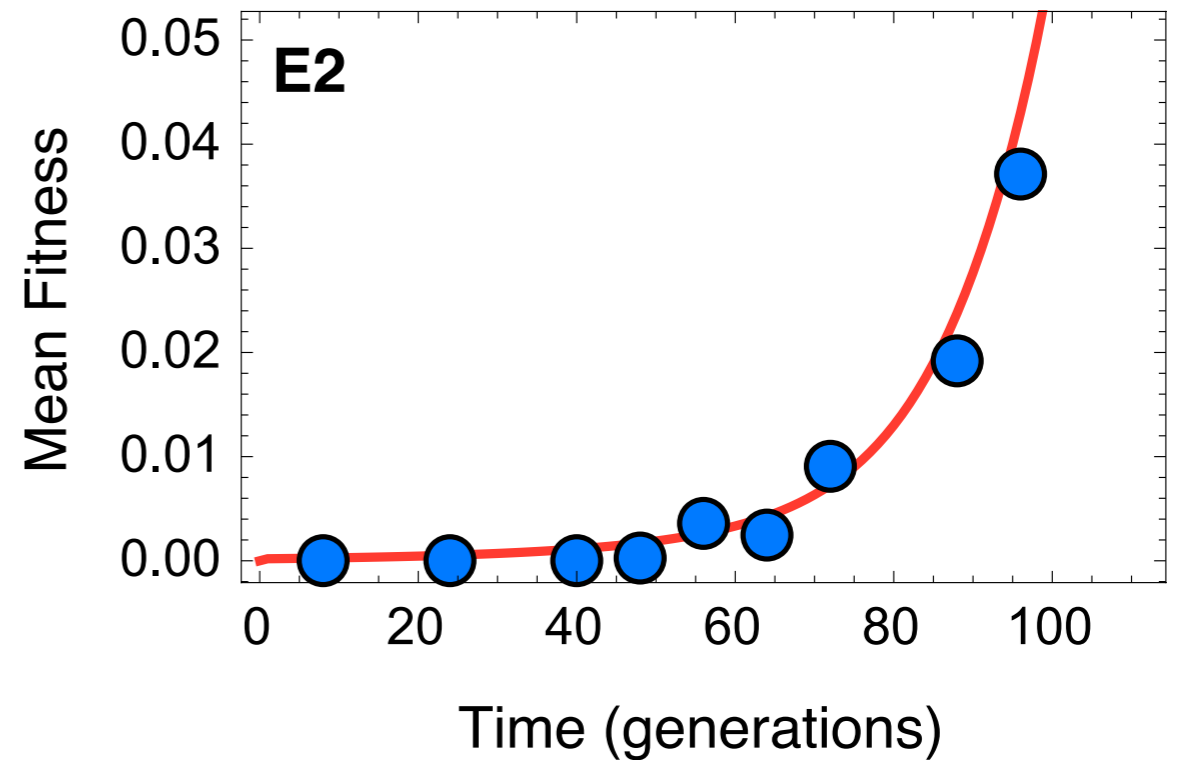
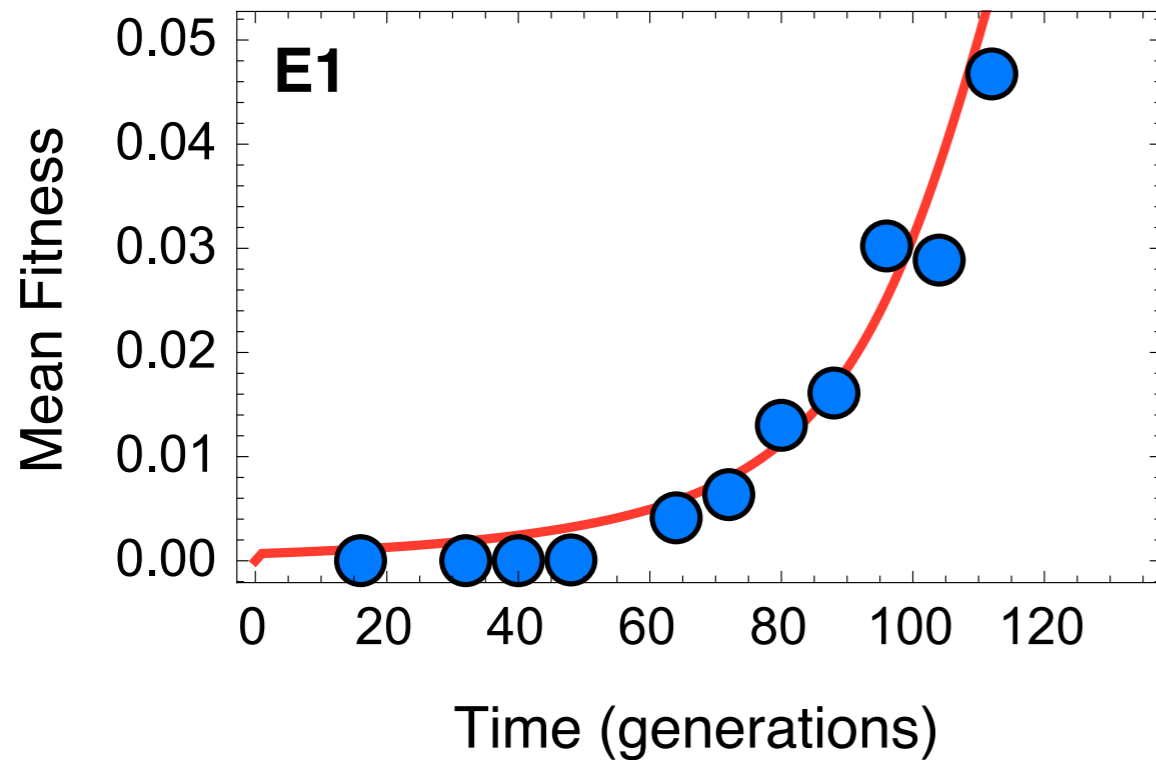
Estimating sources of variance in sequencing counts that are not due to fitness differences



Finding the mean fitness

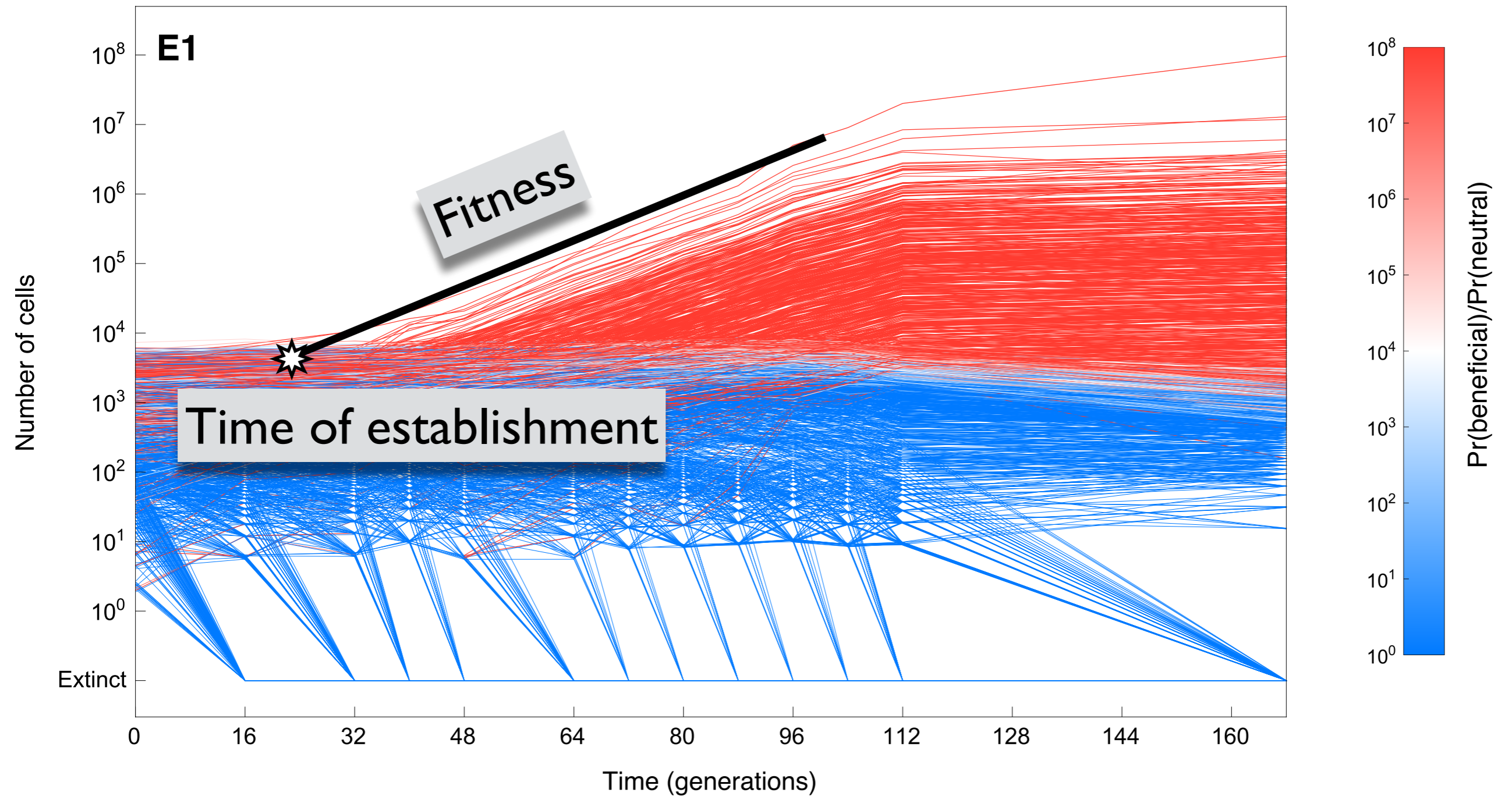


Finding the mean fitness

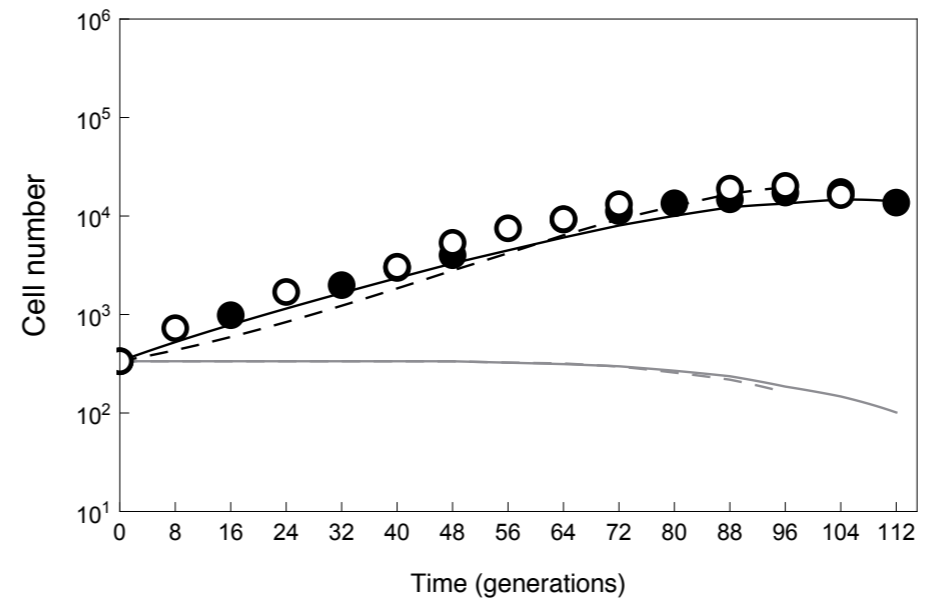
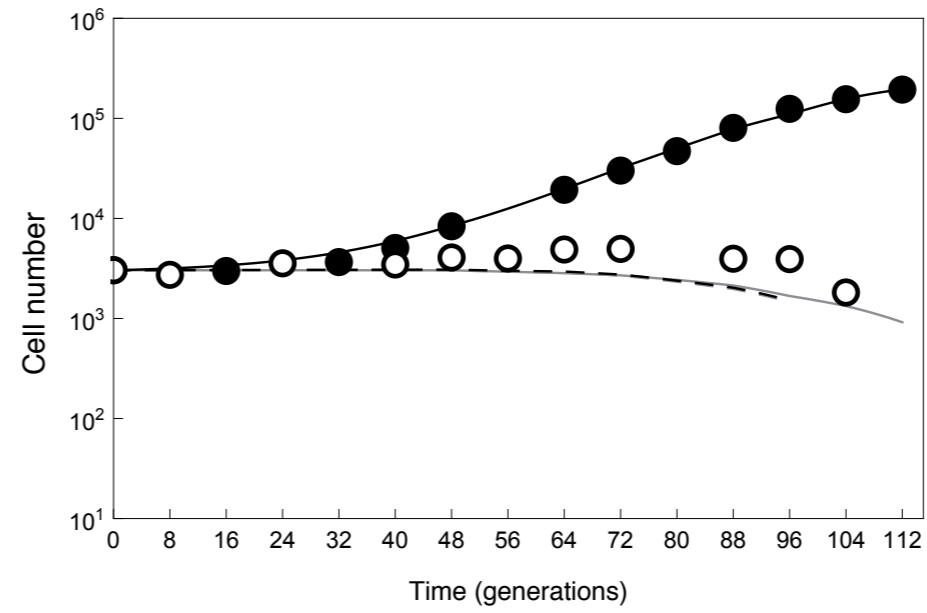
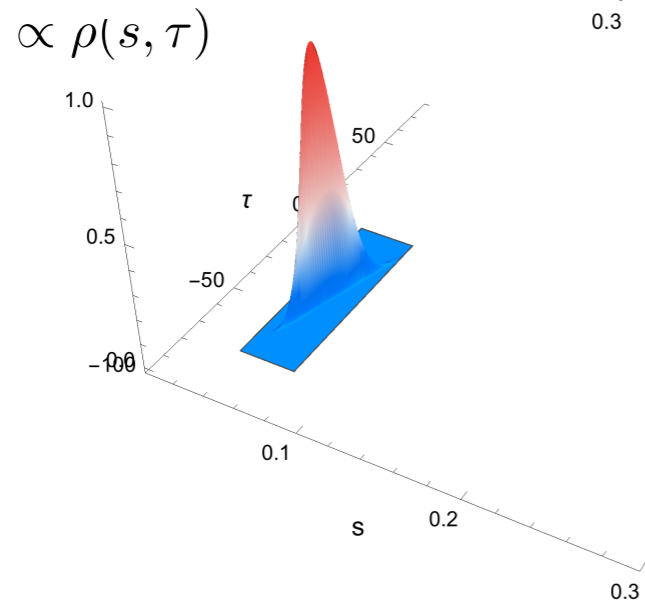
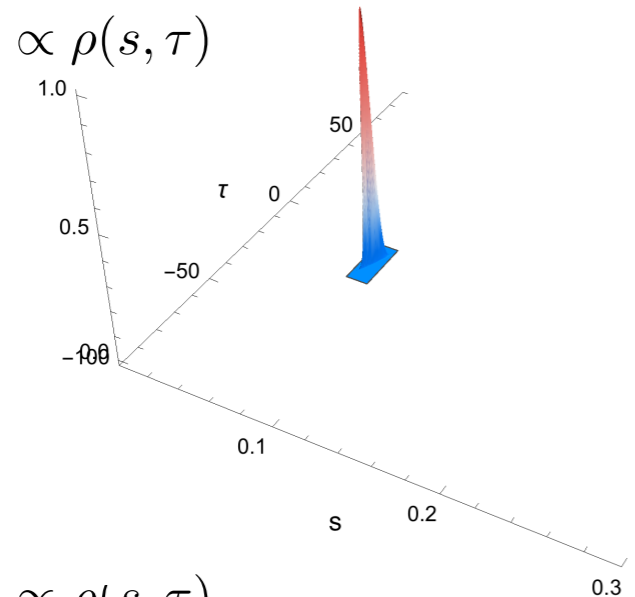


- Decline of neutral lineages
- Mean of all cells

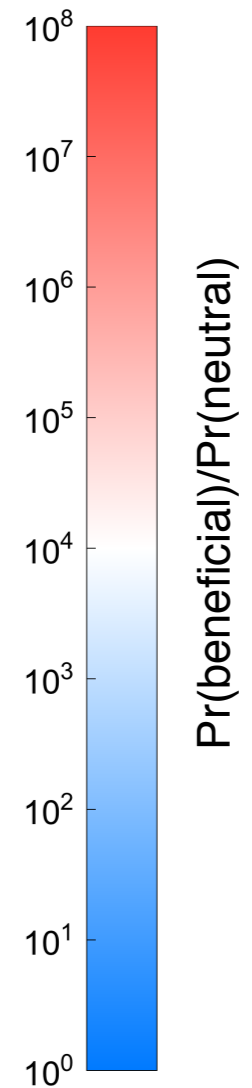
Identifying adaptive lineages



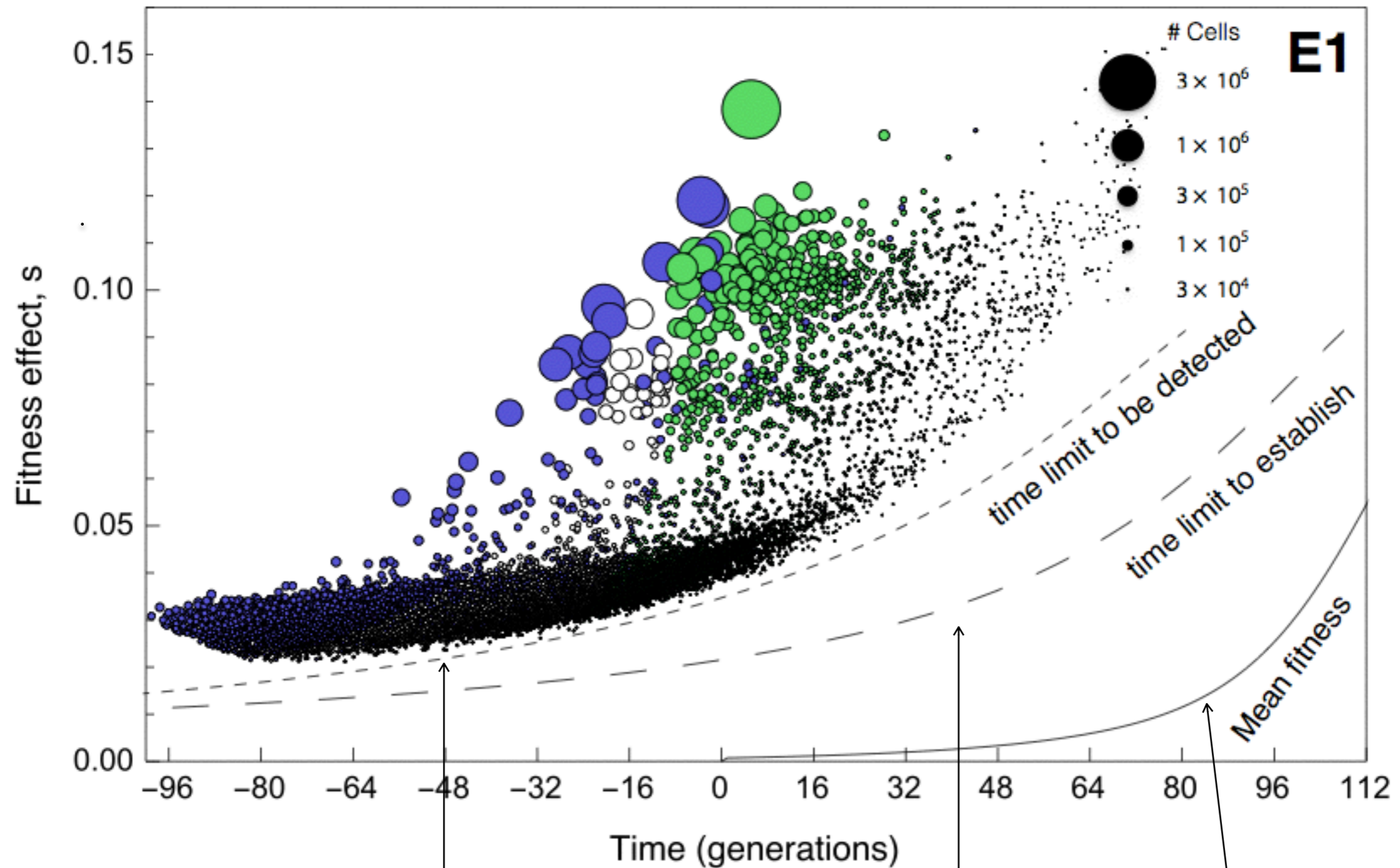
Finding beneficial mutations



- E1
- E1 fit
- E2
- - - E2 fit



Establishment times and fitness effects of beneficial mutations

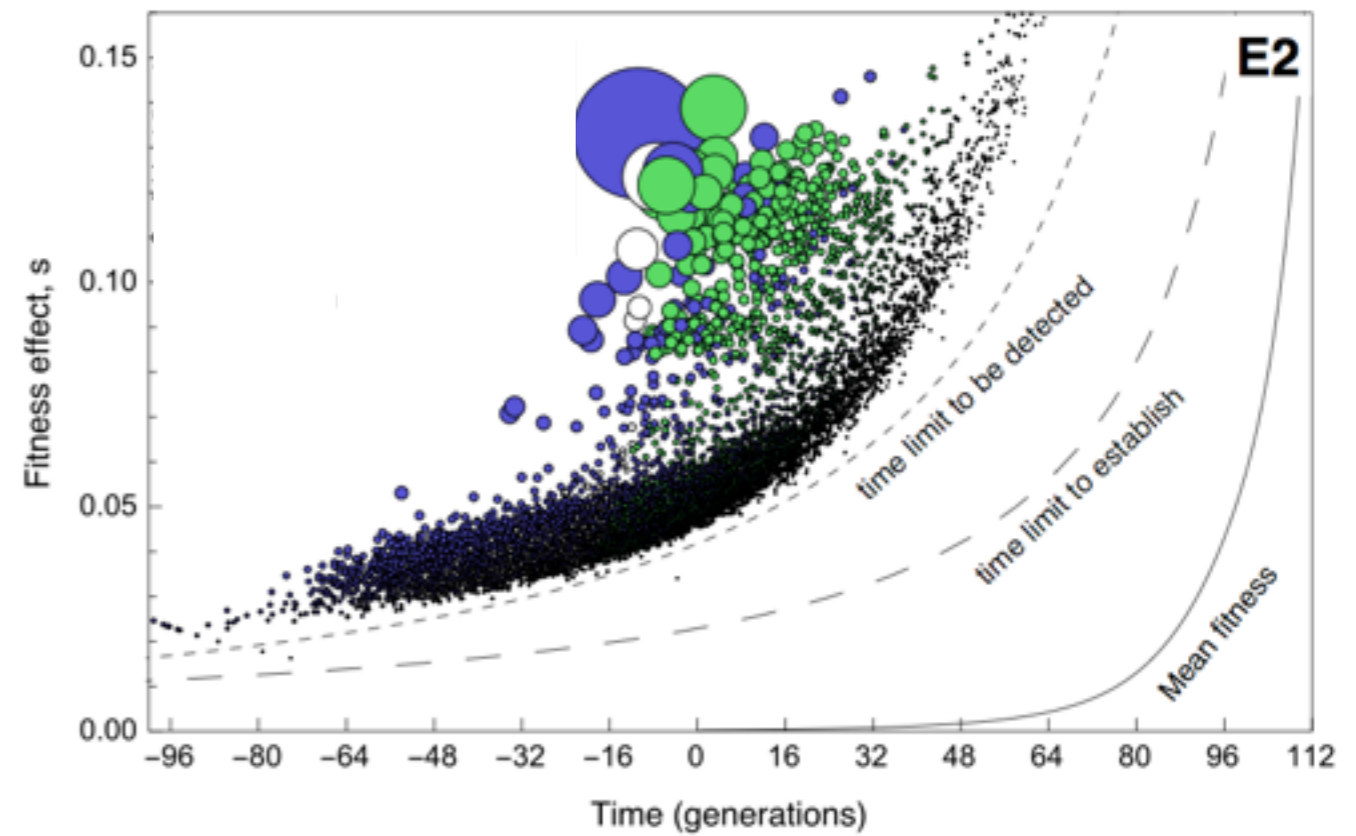
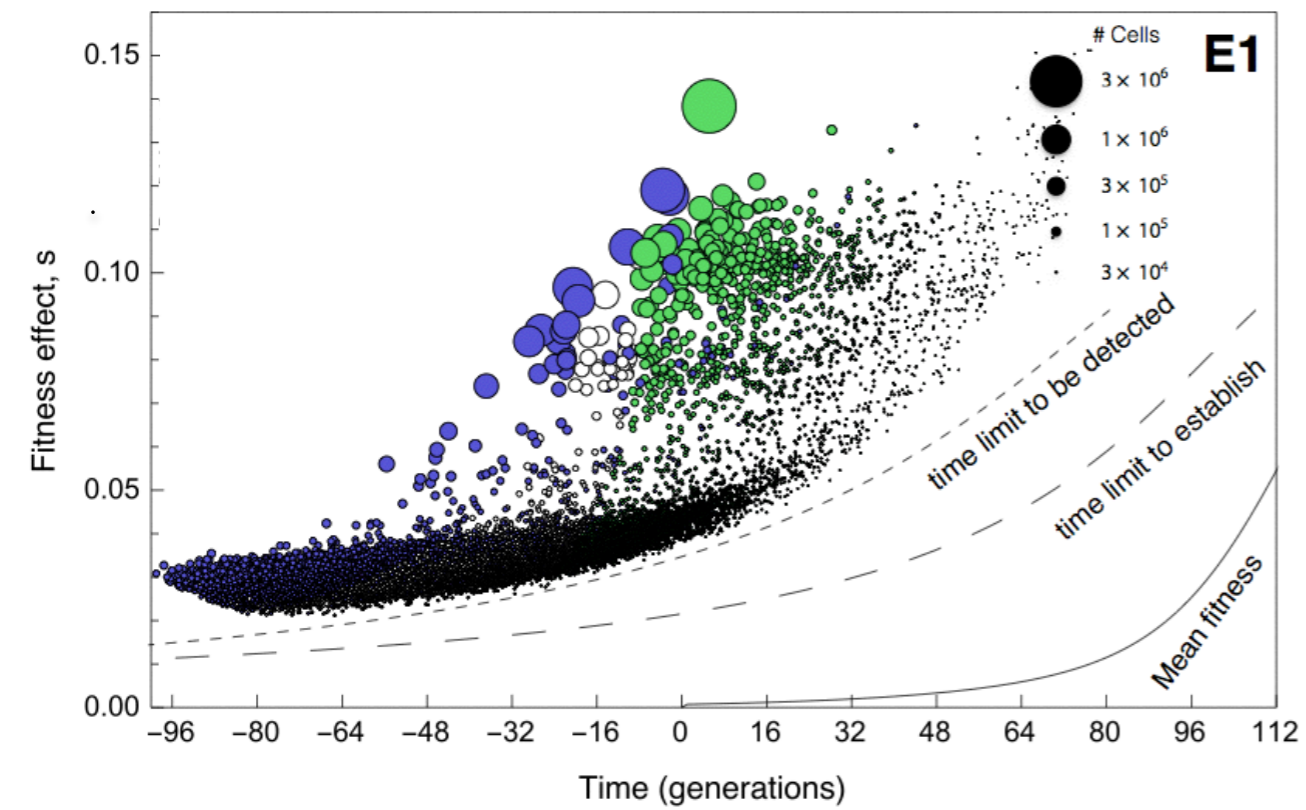


$$t(\bar{x} = s) - (1/s) \ln(n_0 s)$$

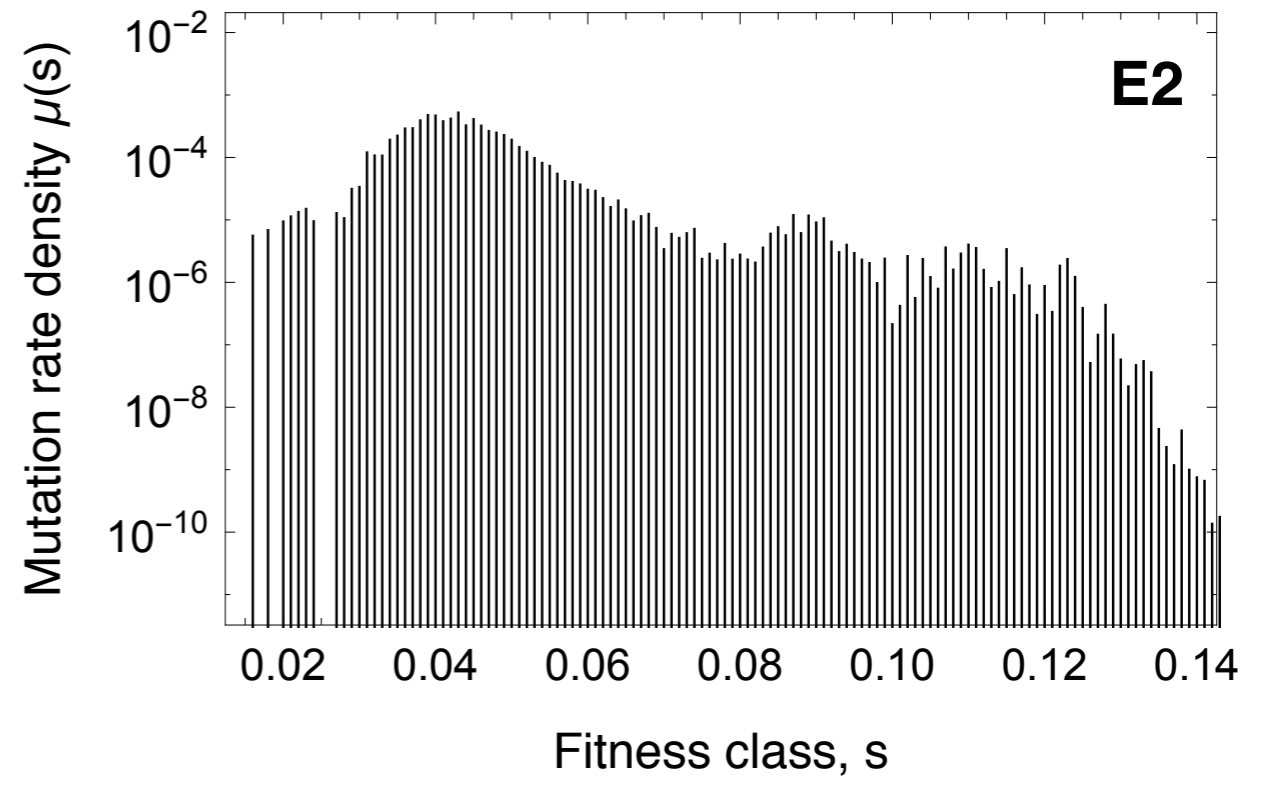
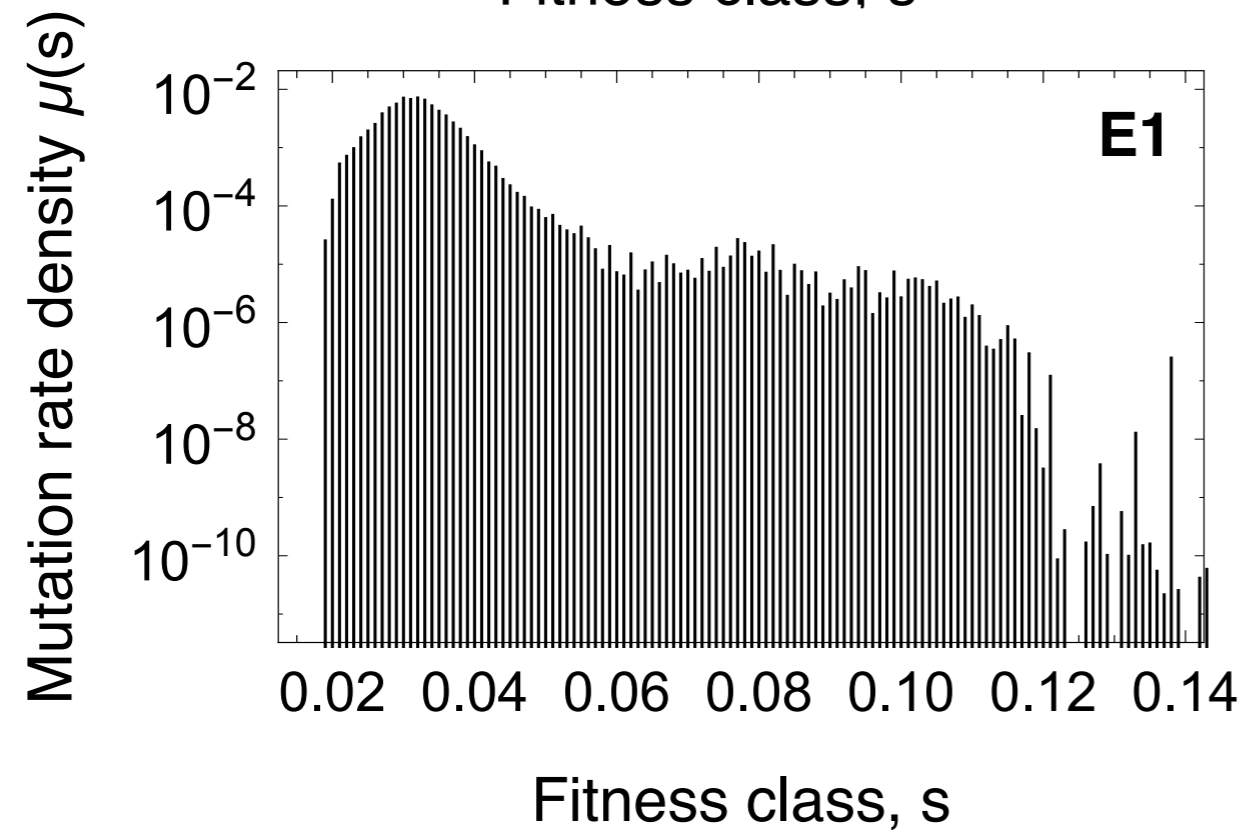
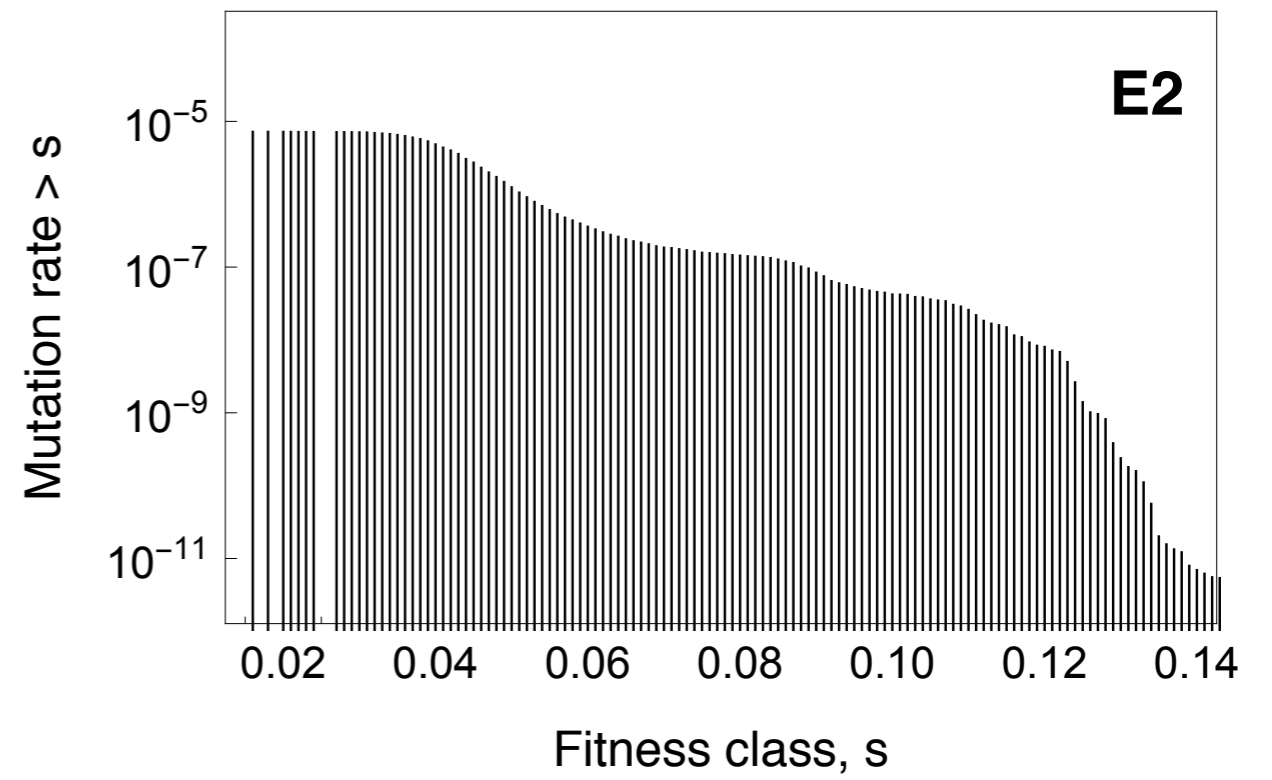
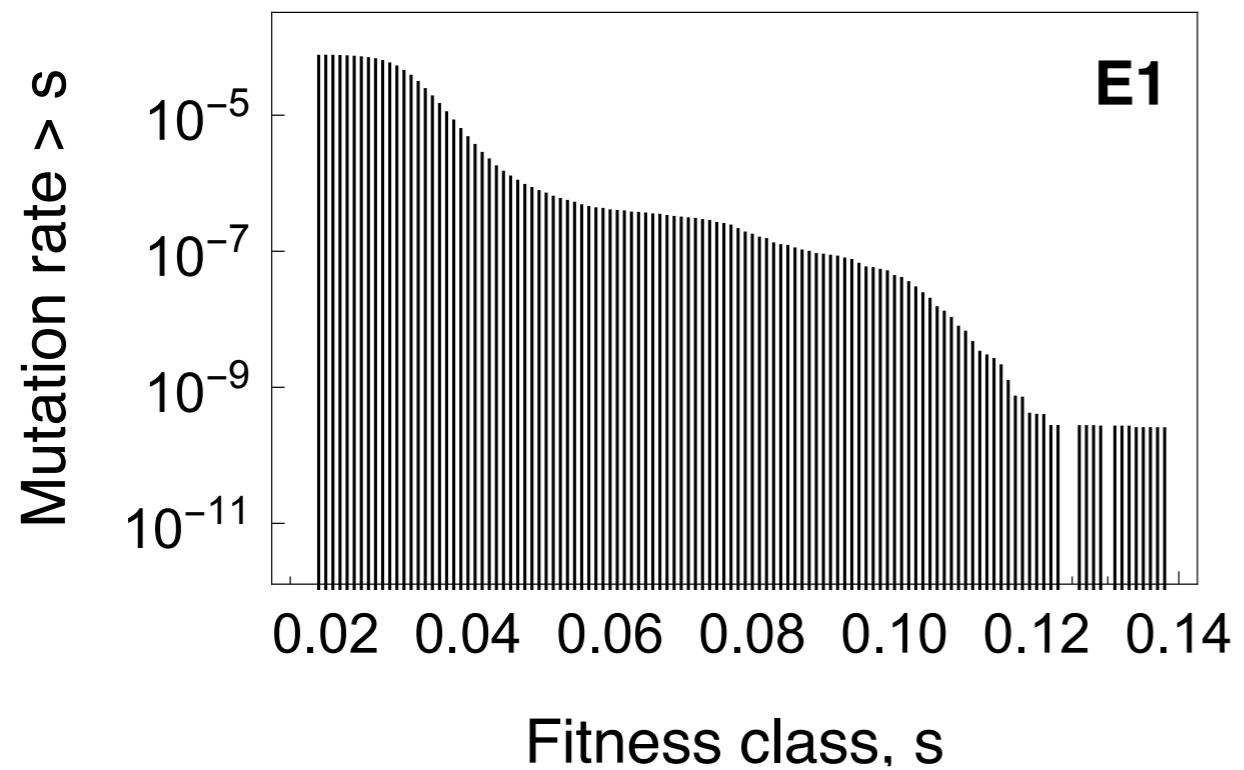
$$t(\bar{x} = s) - 1/s$$

$$t(\bar{x} = s)$$

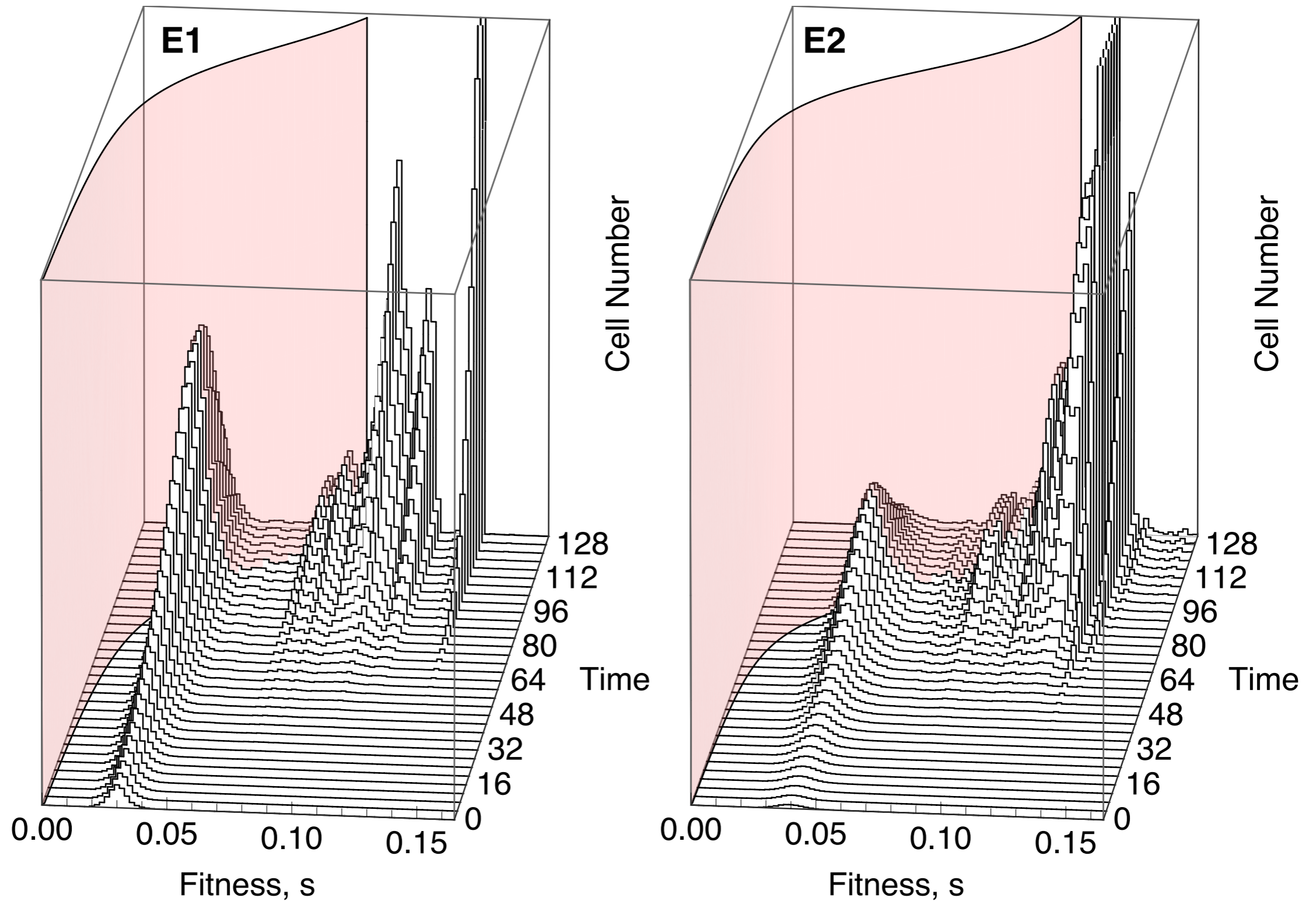
Establishment times and fitness effects of beneficial mutations



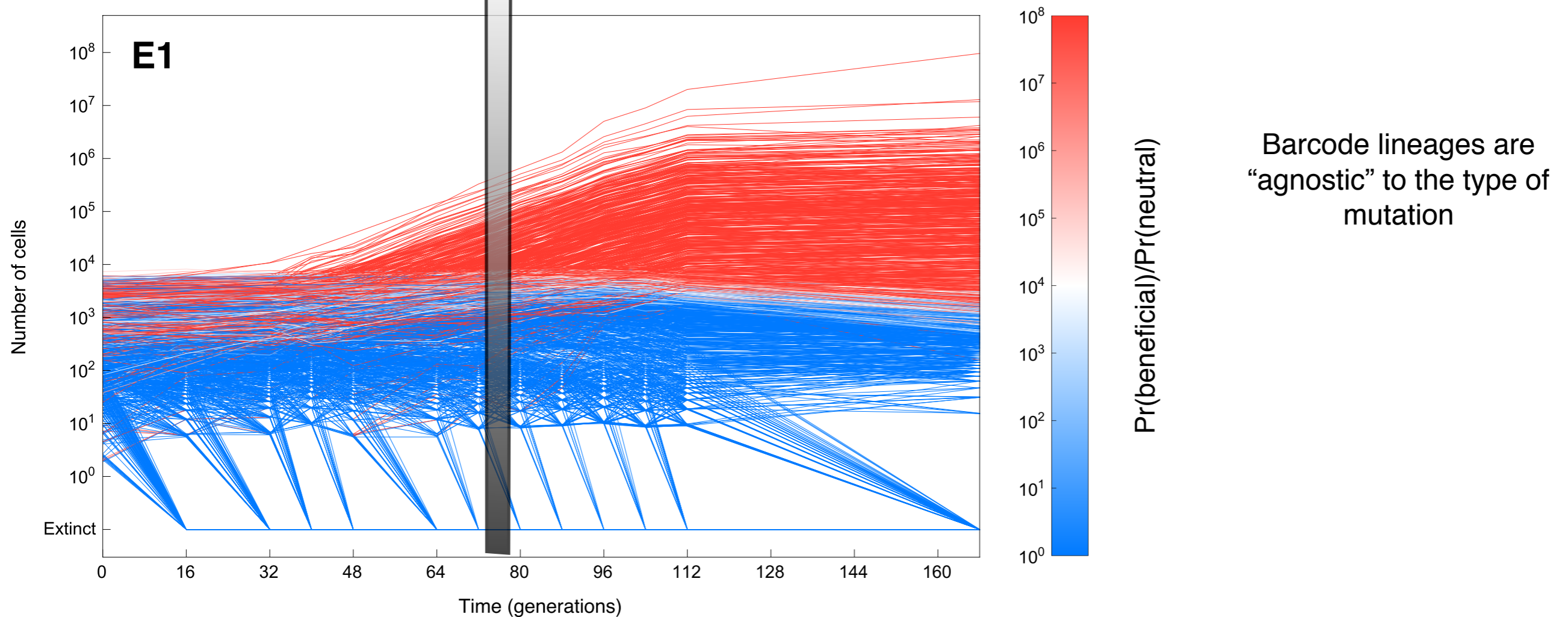
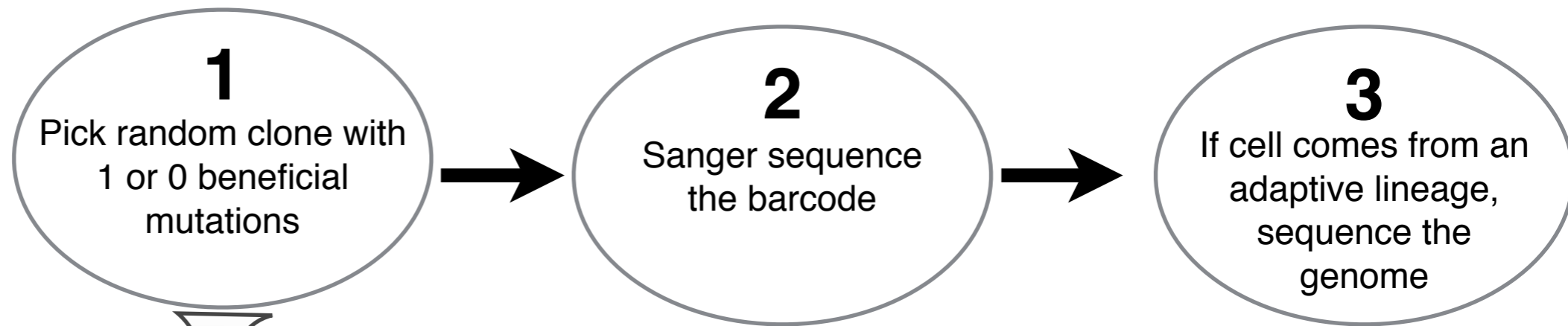
The distribution of mutation rates to each fitness effect is not monotonic (or exponential)



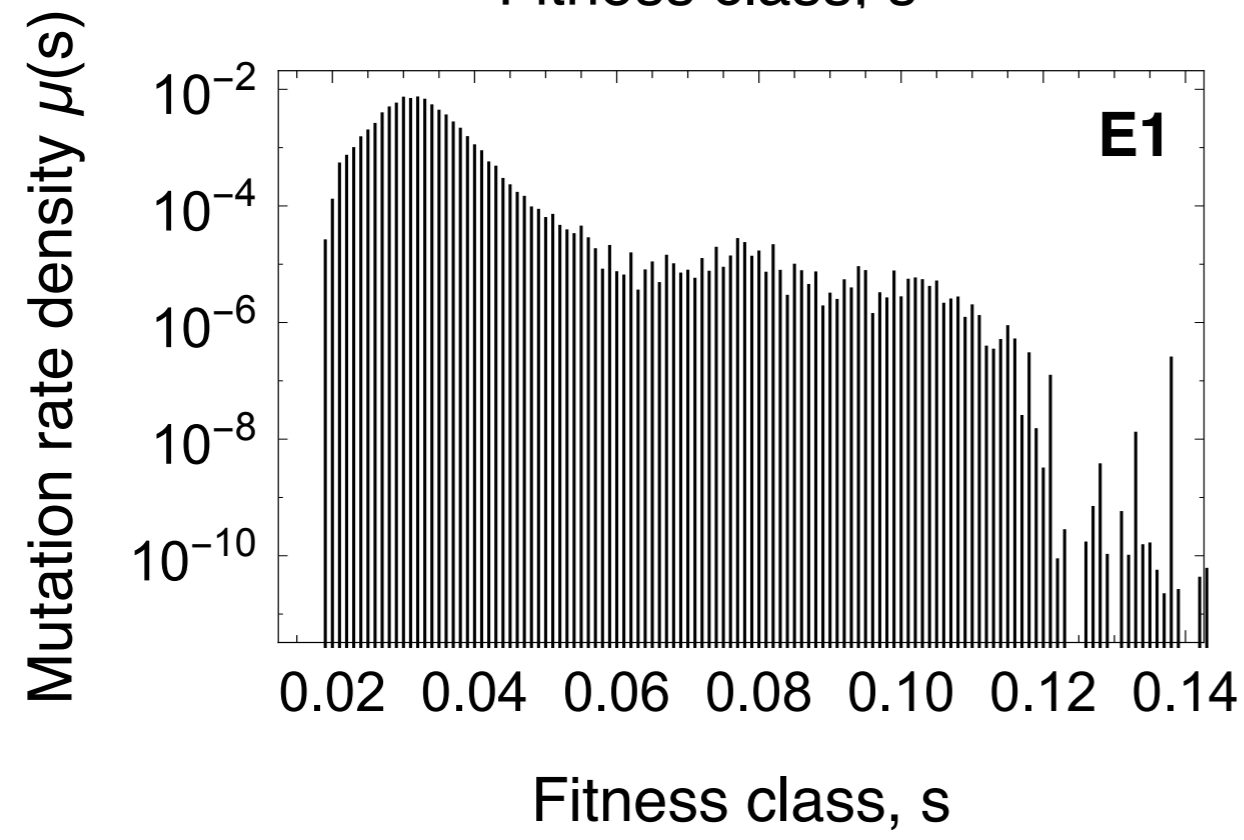
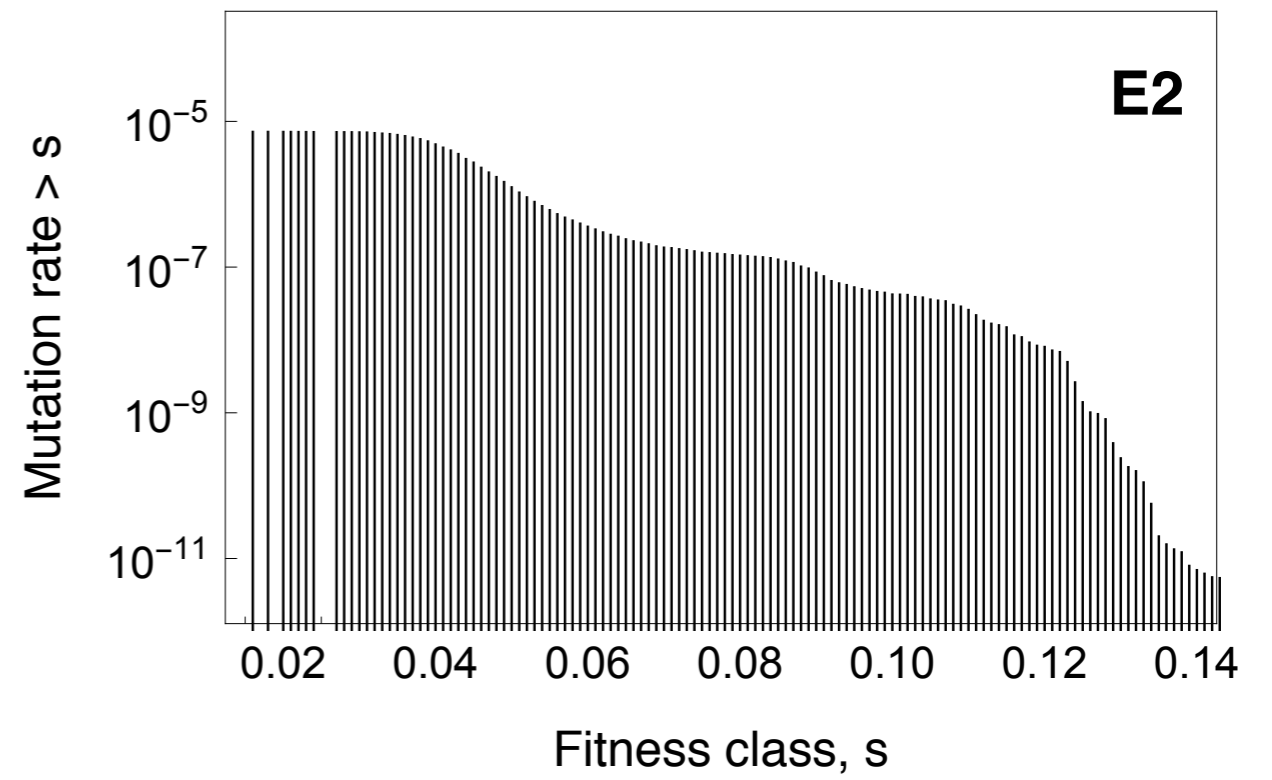
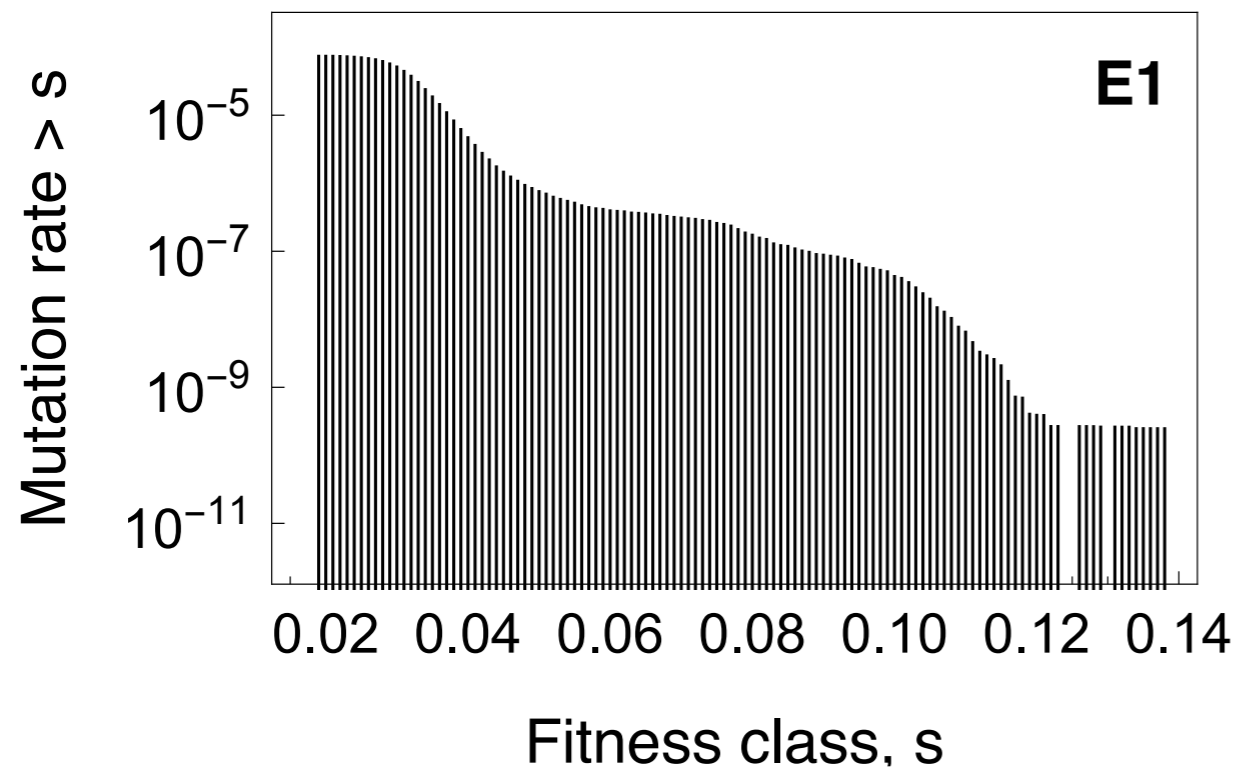
Frequency of cells at each fitness over time



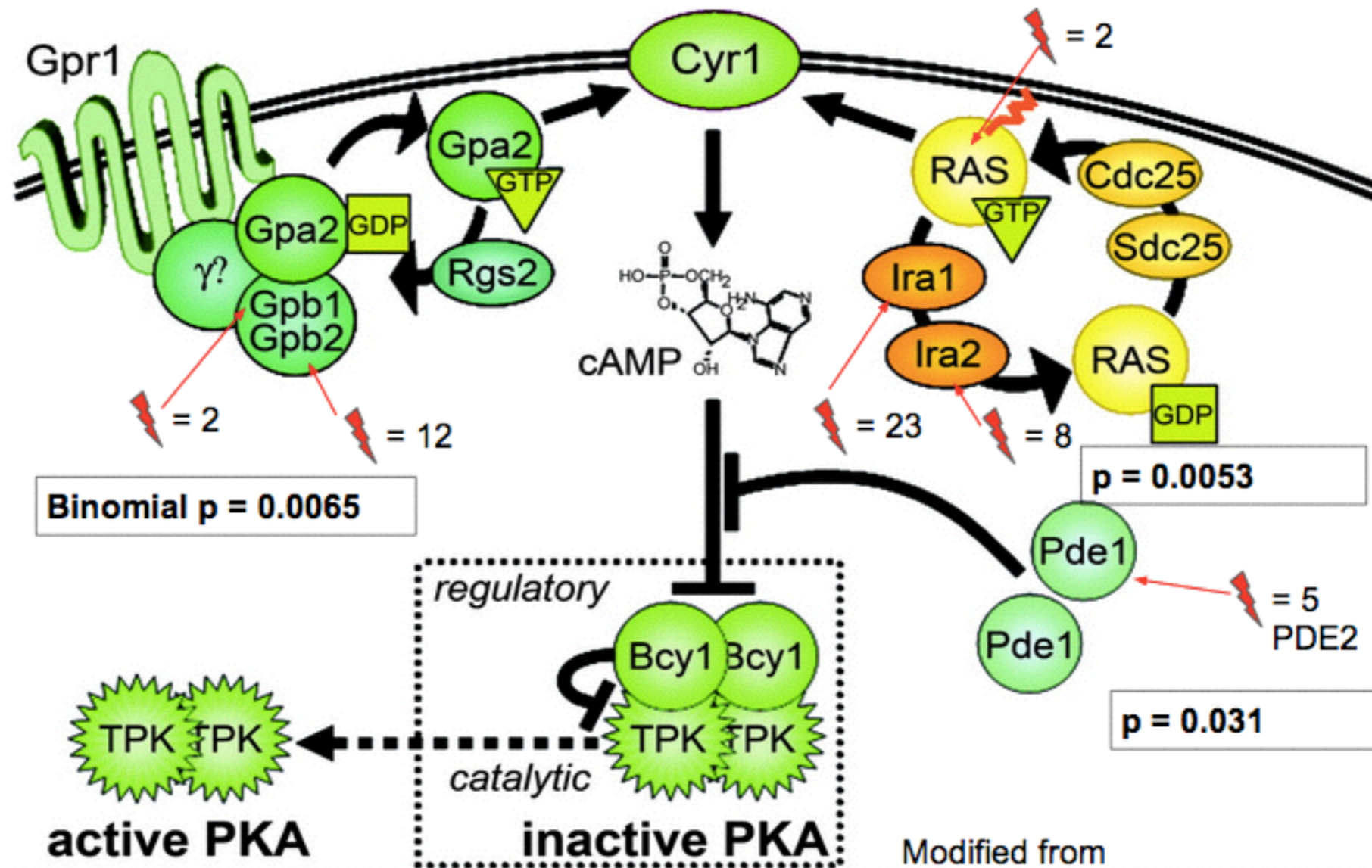
Barcodes help identify the mutational spectrum



The distribution of mutation rates to each fitness effect is not monotonic (or exponential)

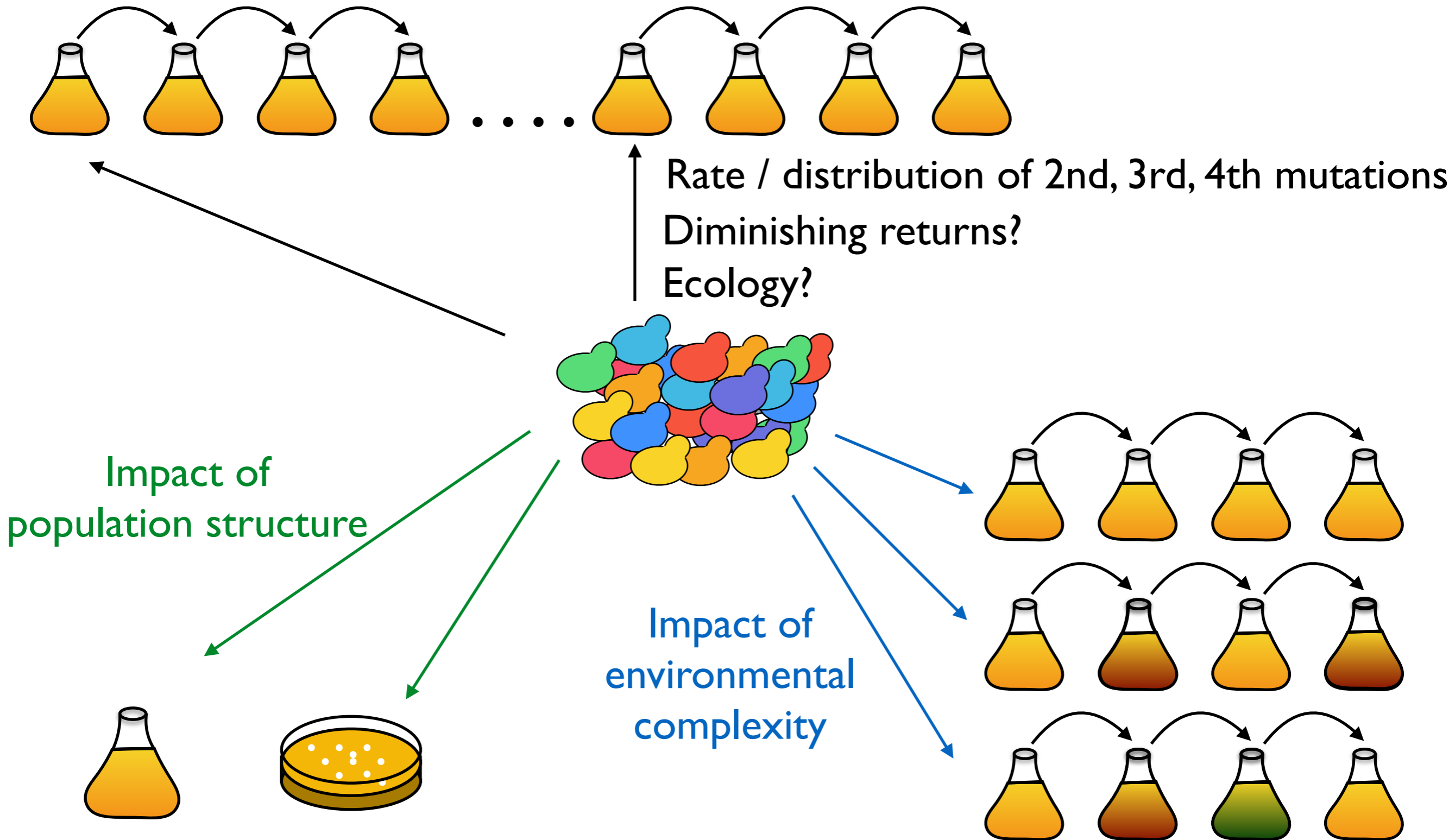


cAMP pathway is a common target



Modified from Santangelo GM. MMBR 2006

Future Directions - Single Barcodes



Conclusions

- I. Random barcodes are a powerful tool for studying evolutionary dynamics
- II. Massive clonal interference underlies the early dynamics of large populations
- III. Very early dynamics are a deterministic outcome of a large number of adaptive mutations
- IV. Later dynamics become more stochastic because they rely on the establishment times of rare large effect mutations
- V. Beneficial mutations have a time window in which to establish after which they will be outcompeted by the mean population fitness
- VI. The distribution of fitness effects has an idiosyncratic tail that is likely to stem from specific rare mutations
- VII. The population size and the distribution of fitness effects predicts the early dynamics well