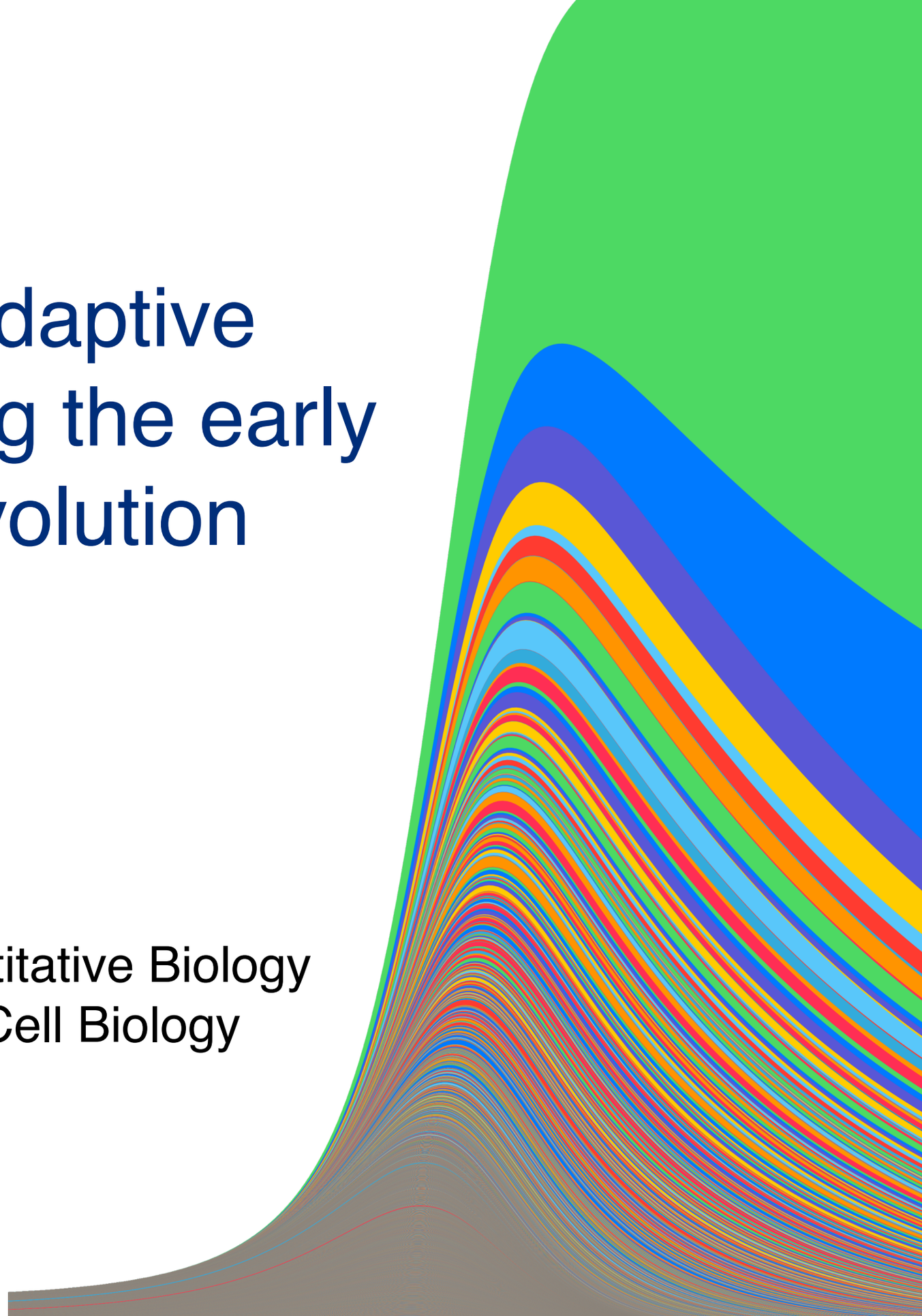
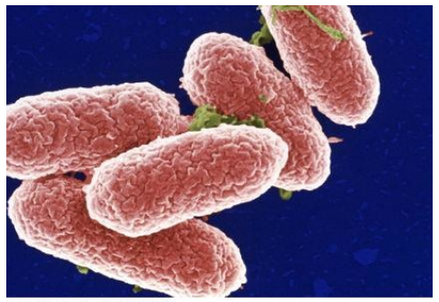
# The dynamics of adaptive genetic diversity during the early stages of clonal evolution

Sasha Levy
Laufer Center for Physical and Quantitative Biology
Department of Biochemistry and Cell Biology
Stony Brook University

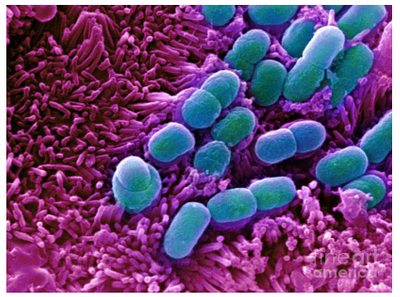


# Cellular evolution and disease

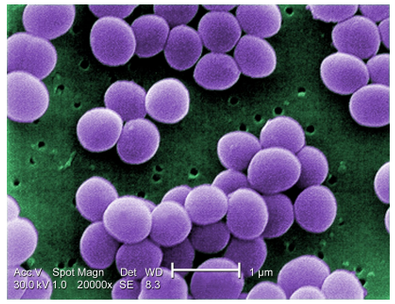
## Bacterial Pathogenesis/ Drug Resistance



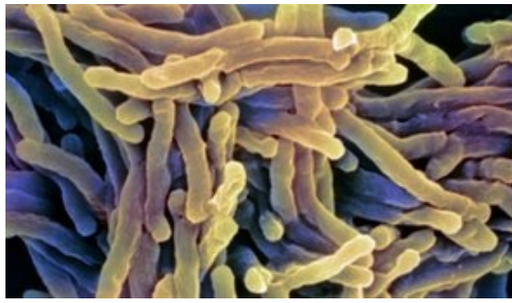
Salmonella



E. coli



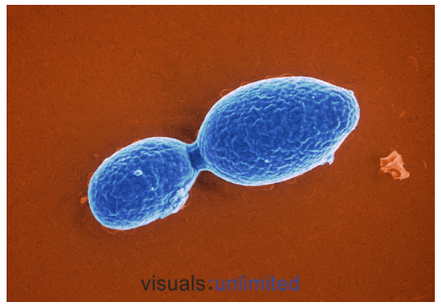

Staphylococcus



Tuberculosis

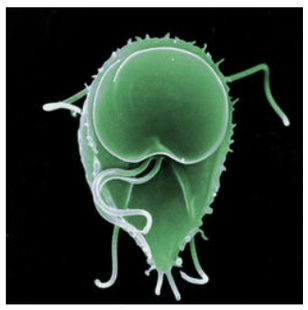
## Yeast Pathogenesis/ Drug Resistance

Candida albicans

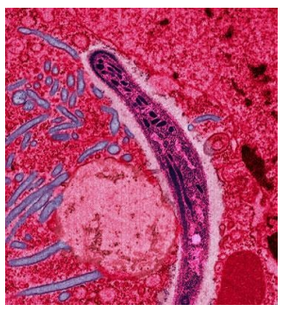



Candida glabrata

## Protist Pathogenesis/ Drug Resistance

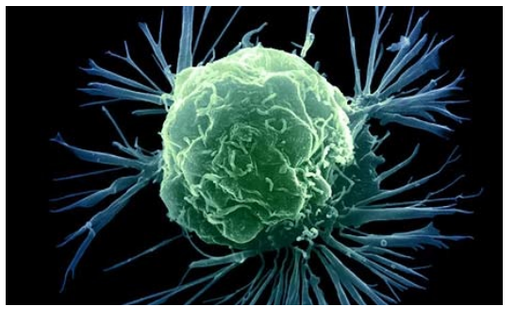


Giardia

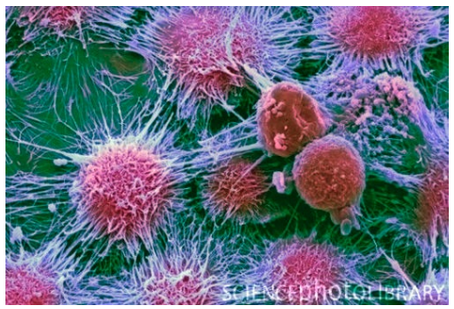


Malaria

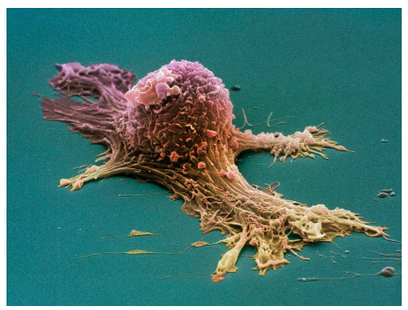
## Cancer Progression/ Drug Resistance



Breast



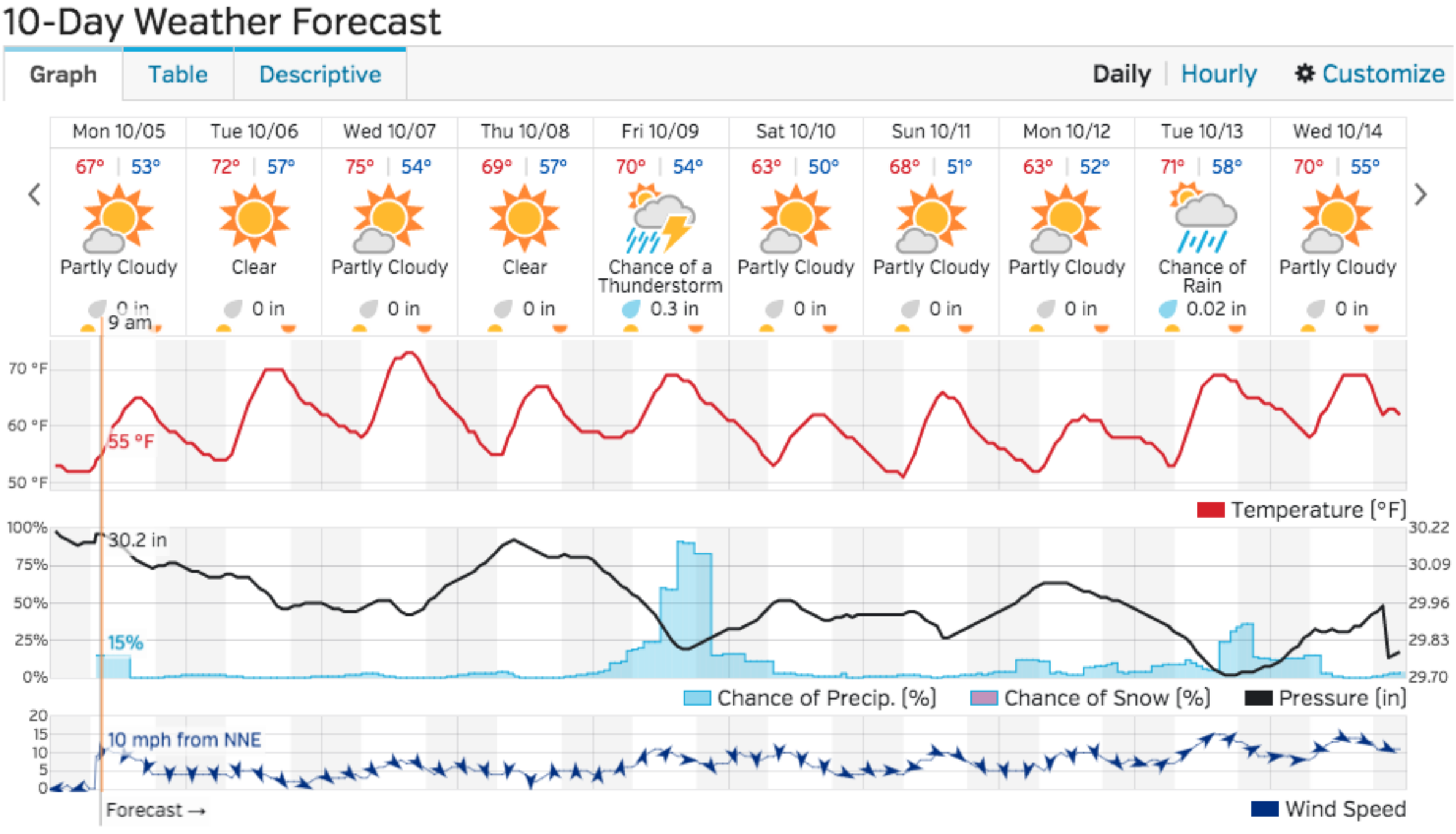

Kidney



Ovarian


Cervical

# Cloudy with a chance of adaptation



# A probabilistic approach to understanding evolution

1. High resolution quantitative measurement of the properties that control evolution
2. Mathematical modeling and numerical simulations using this data to assign probabilities to various outcomes

# What factors impact the dynamics of evolution?

1) Distribution of Fitness Effects (DFE)

2) Epistasis

3) Population Size

4) Population Structure

5) Changing Environments

6) Sex

7) Ecology

# Experimental evolution limits the unknowns

1) Distribution of Fitness Effects (DFE)

2) Epistasis

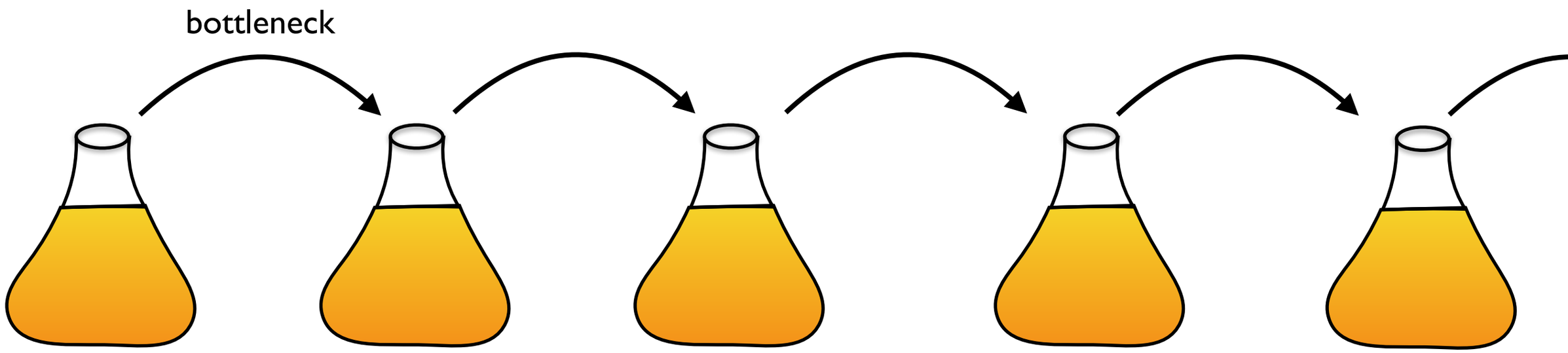
~~3) Population Size~~

~~4) Population Structure~~

~~5) Changing Environments~~

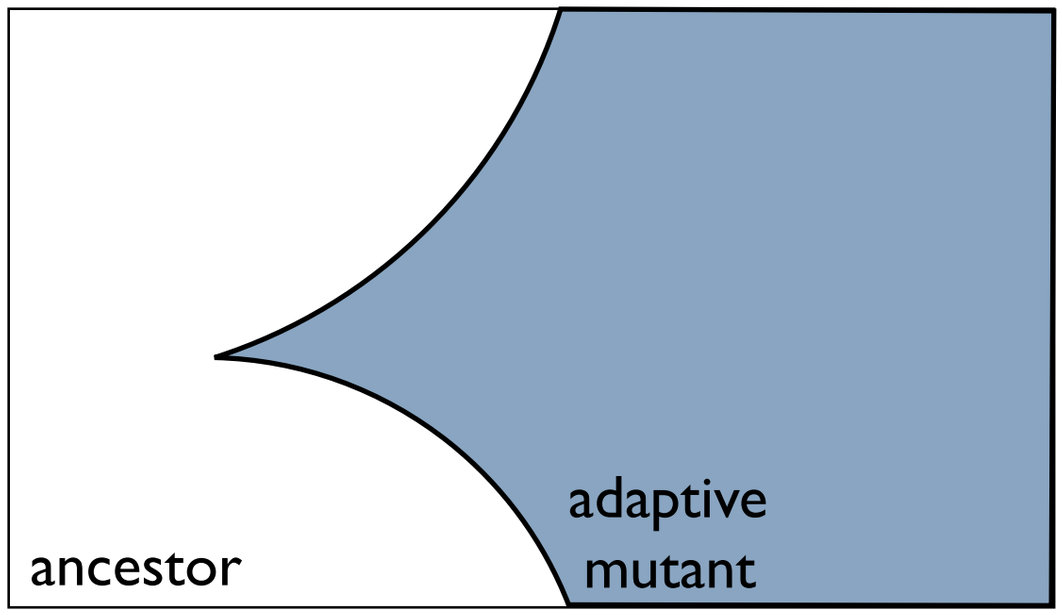
~~6) Sex~~

~~7) Ecology~~

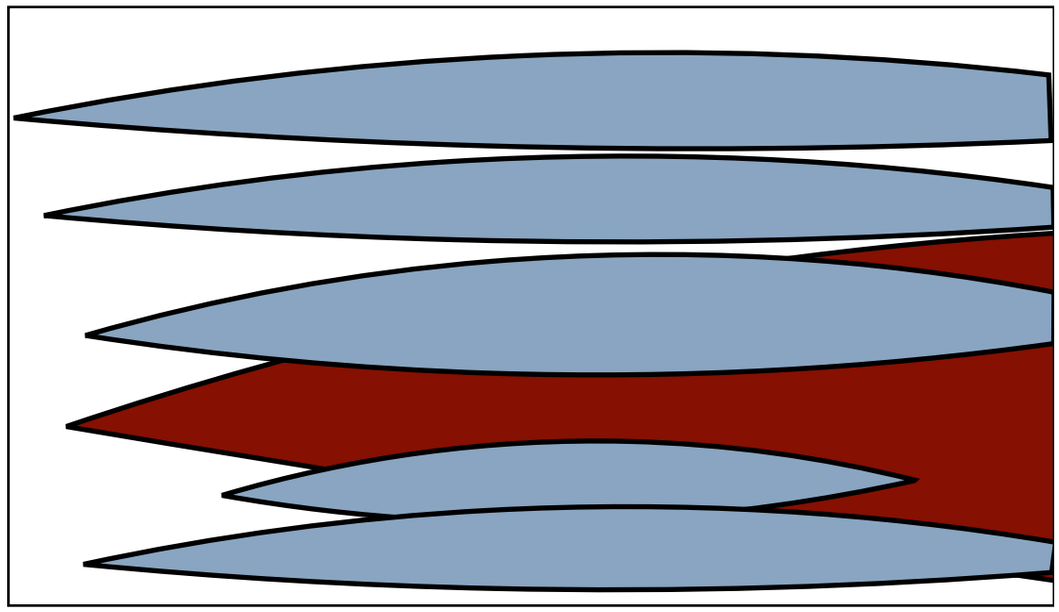


# Evolution of small and large populations

Small




Large



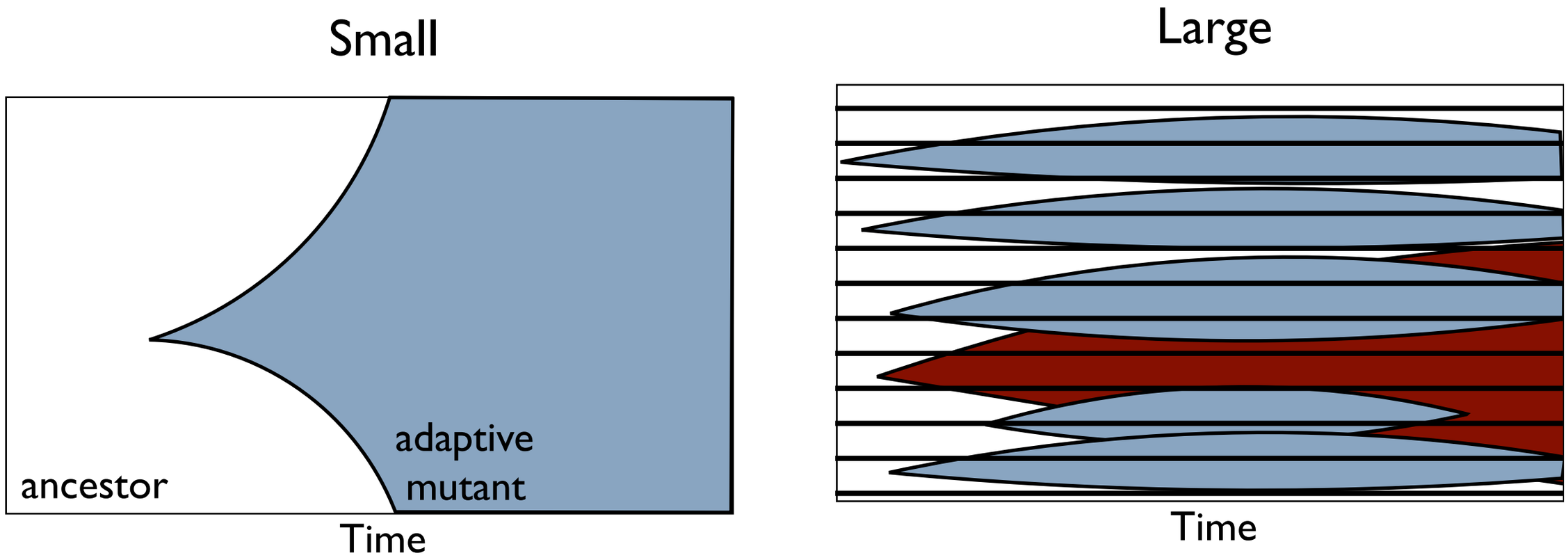

Selective Sweep

Clonal Interference

$NU_b < 1$

$NU_b >> 1$

# Small lineages within large populations

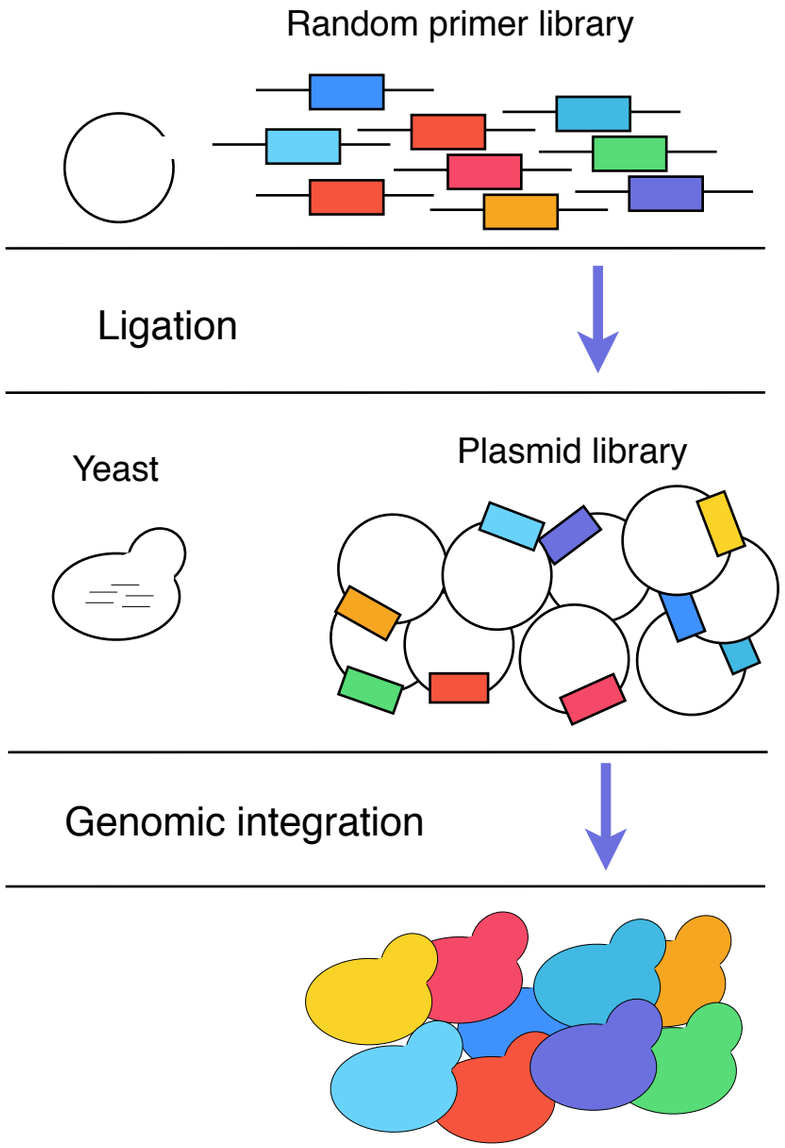



Selective Sweep

Clonal Interference

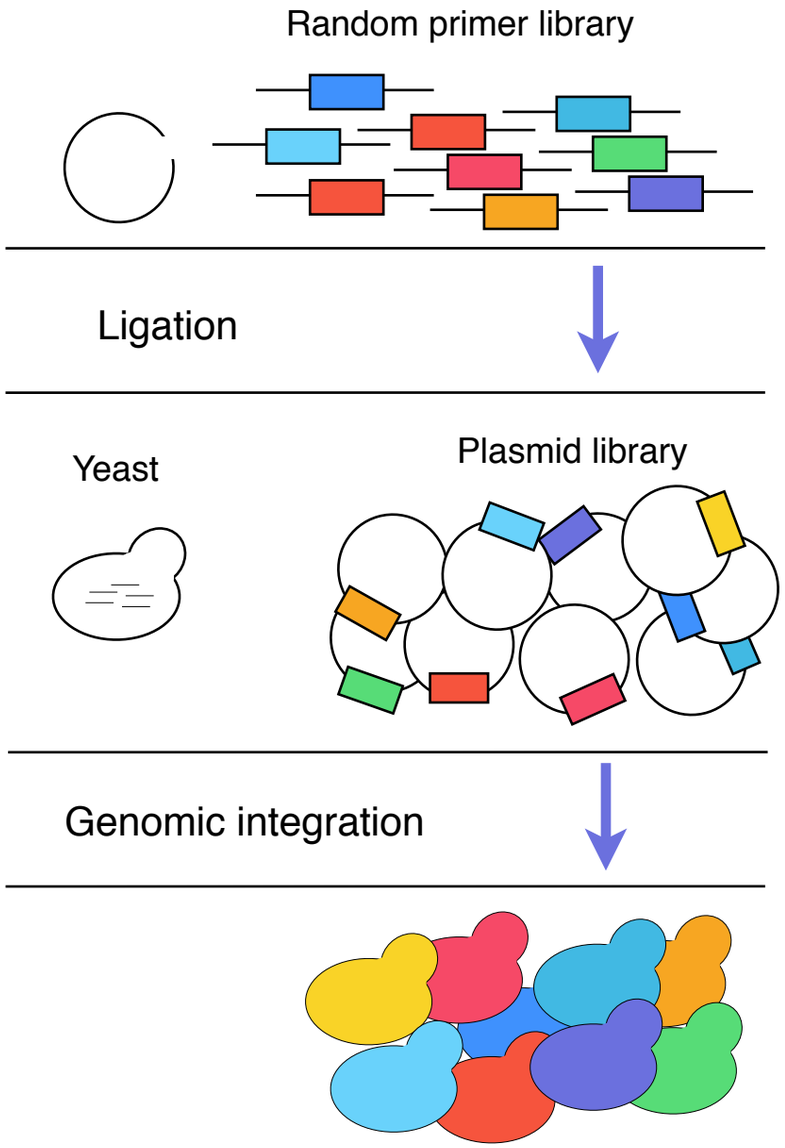
Break up a large population into small lineages

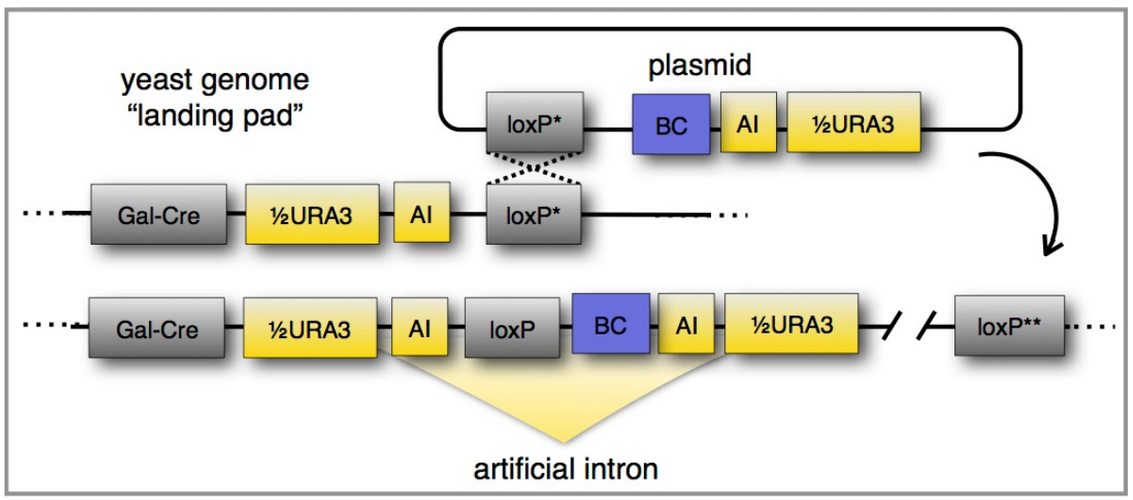
# Our approach: high-resolution barcode sequencing




Levy, Blundell et al., *Nature*, 2015

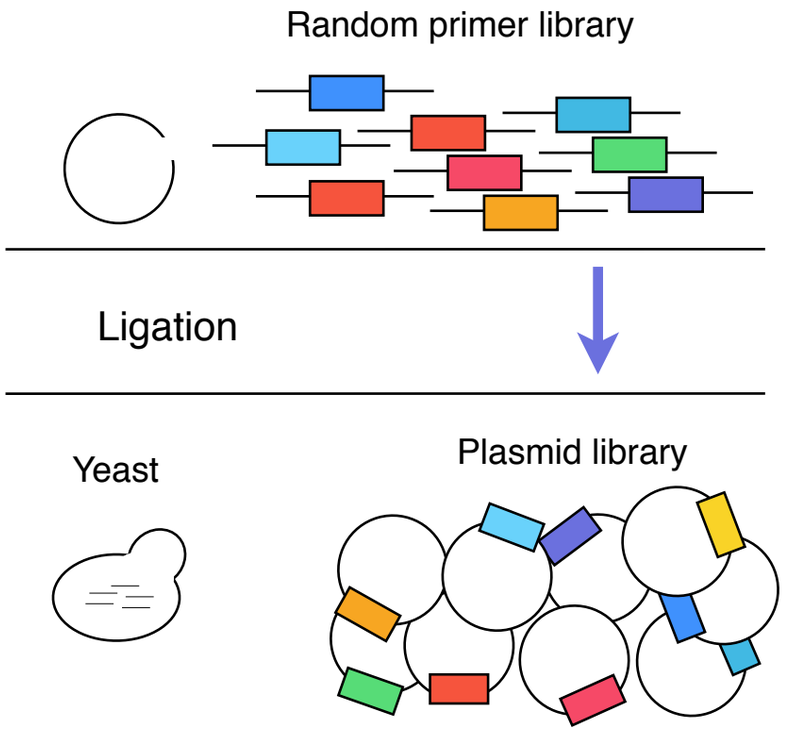
# Our approach: high-resolution barcode sequencing

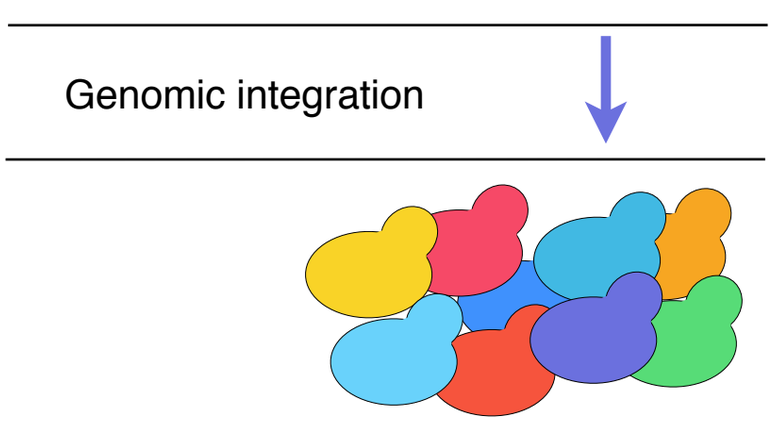


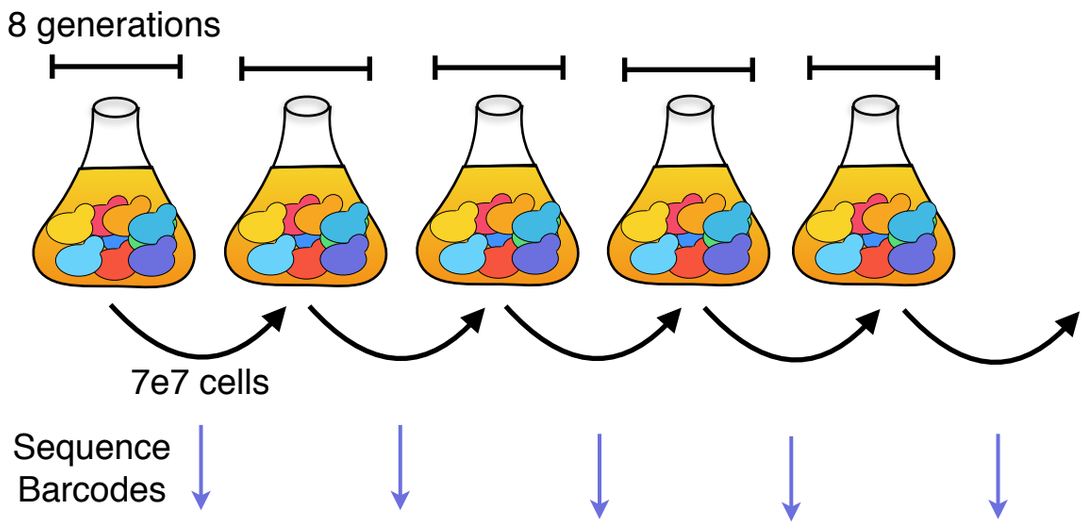



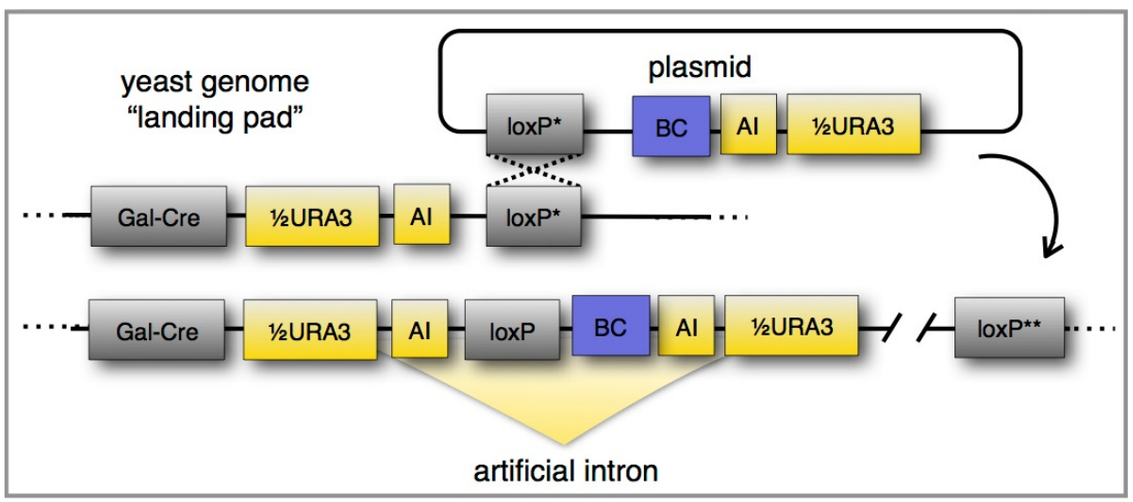

Levy, Blundell et al., *Nature*, 2015

# Our approach: high-resolution barcode sequencing



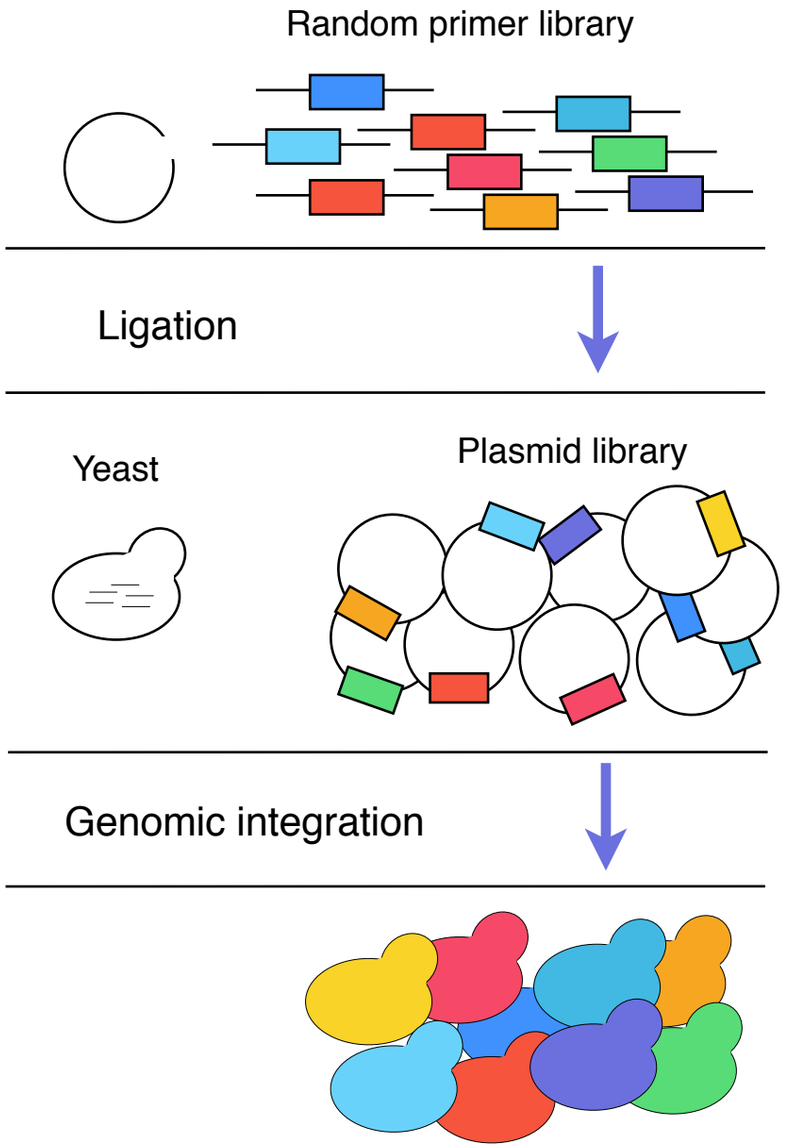



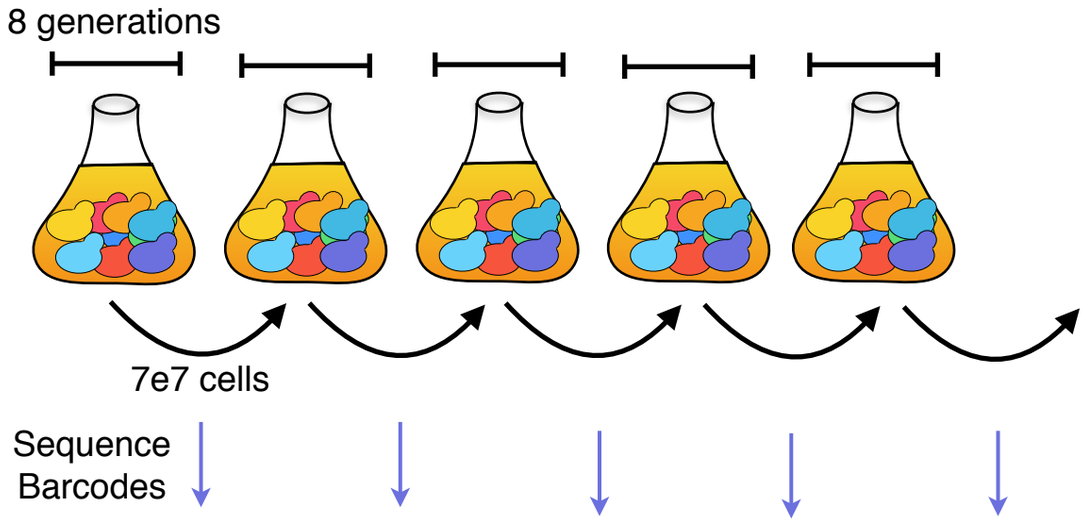



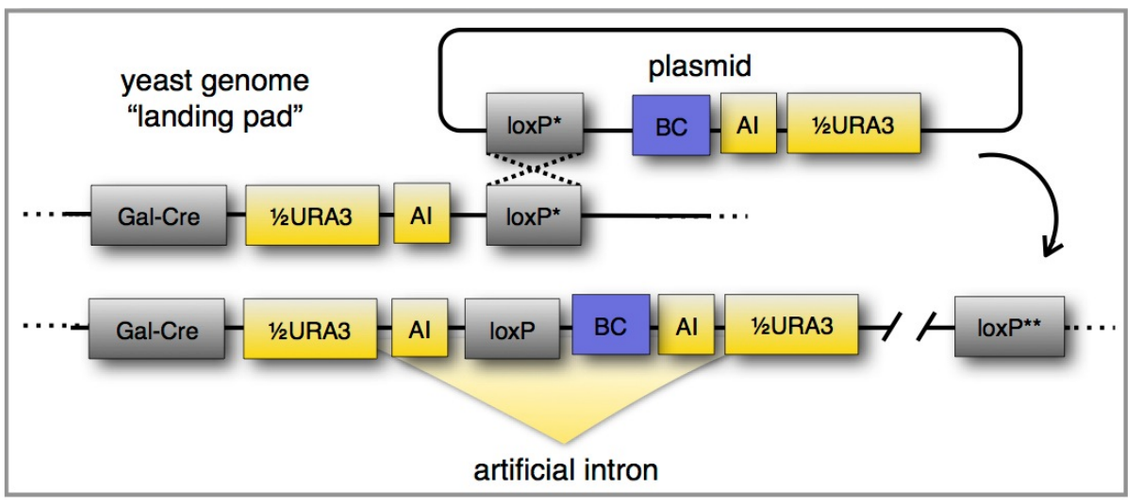



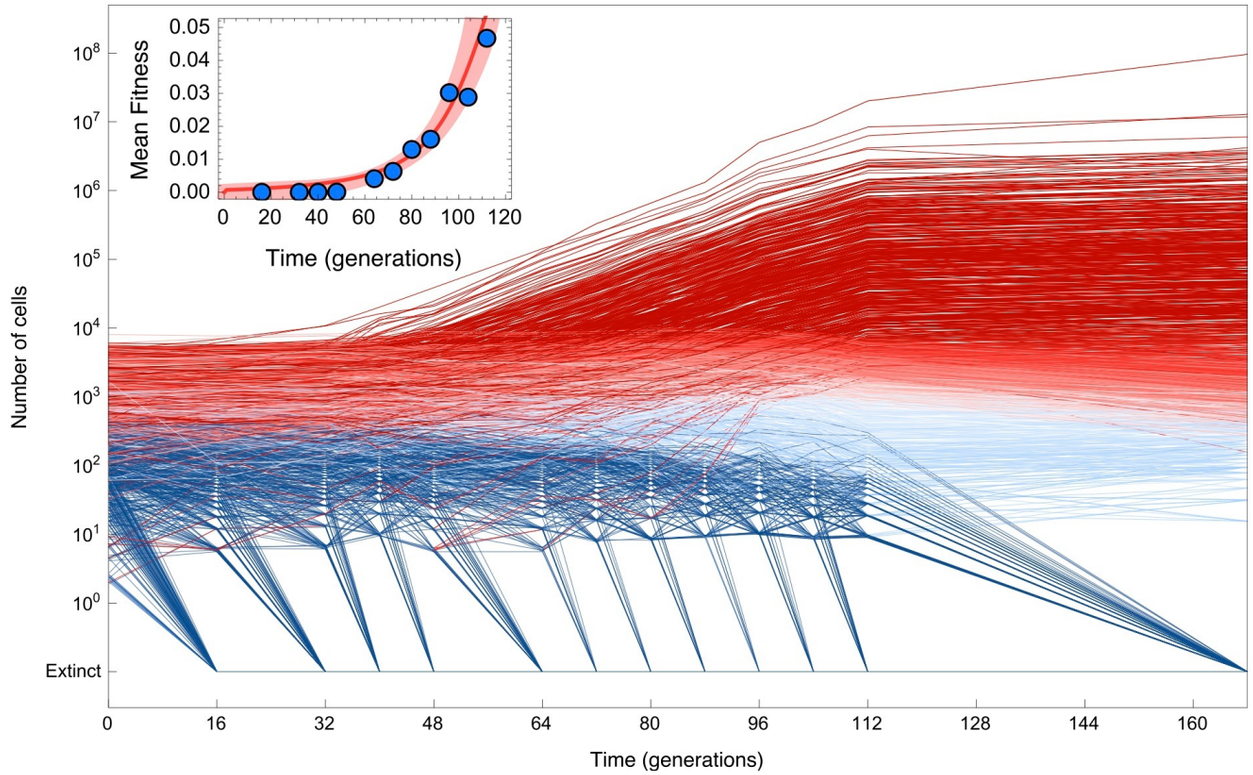

Levy, Blundell et al., *Nature*, 2015

# Our approach: high-resolution barcode sequencing



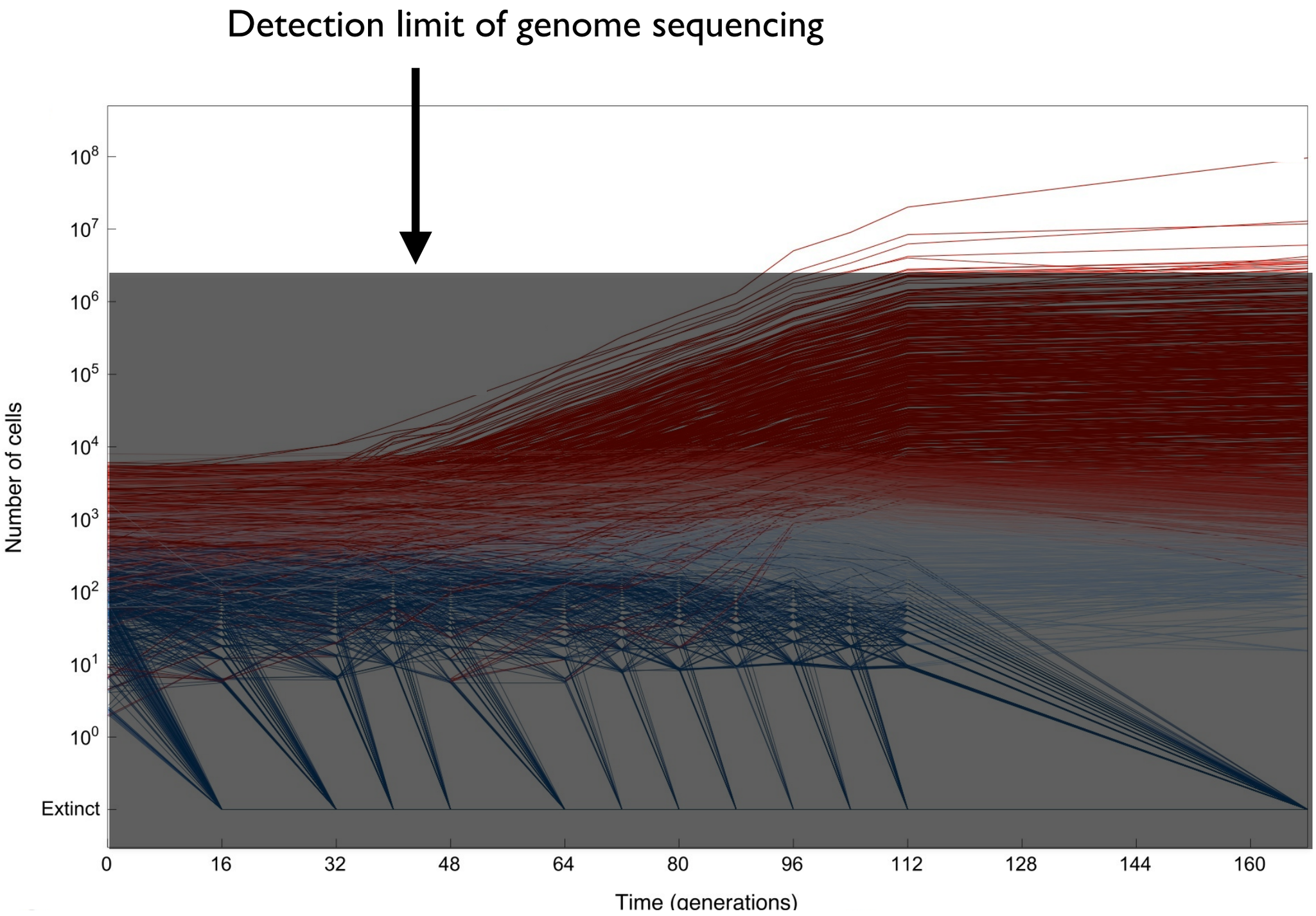



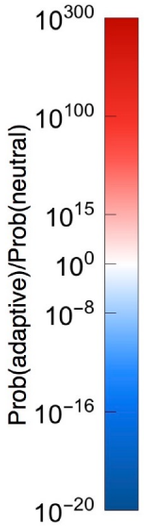



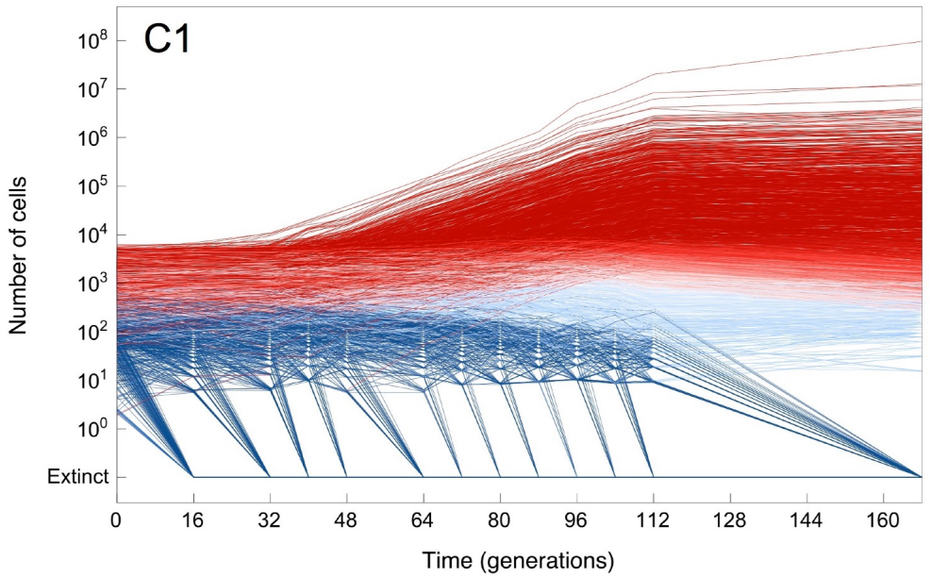

Levy, Blundell et al., *Nature*, 2015

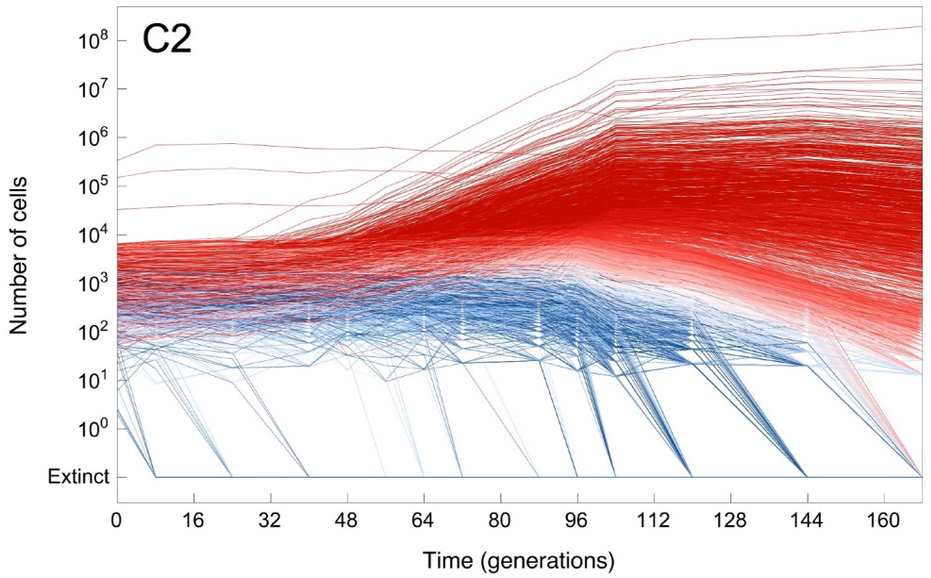
# The advantage of bar-seq: high signal / noise

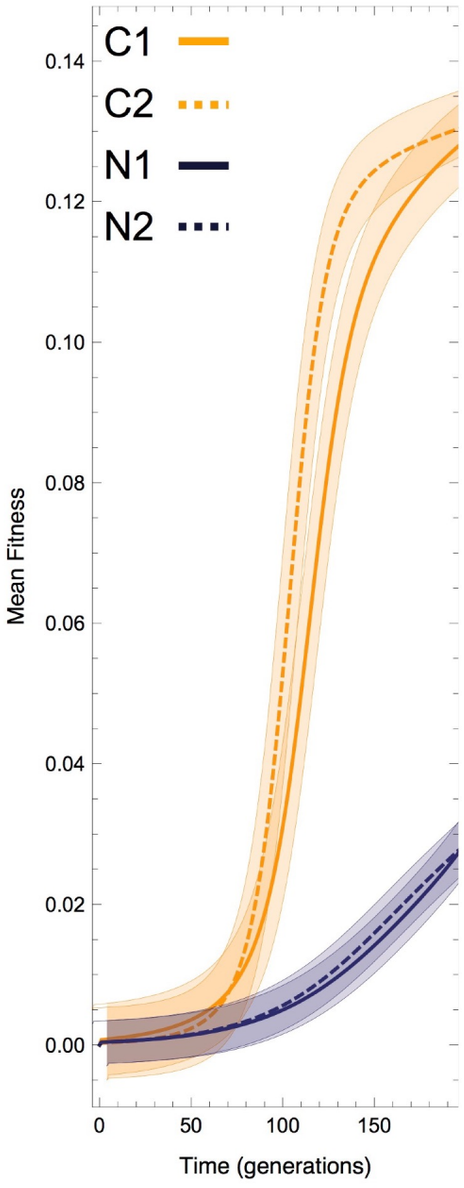


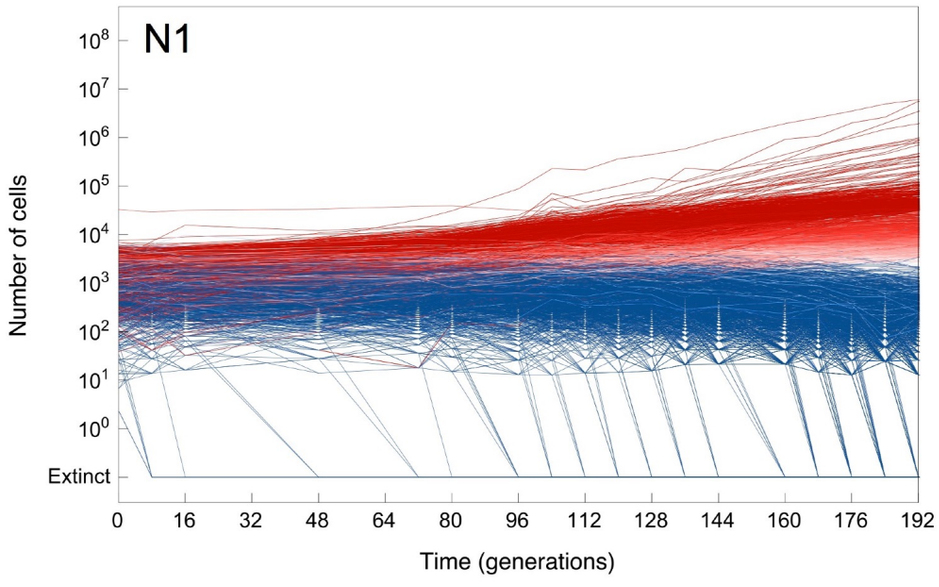
# Replicate evolutions in two environments

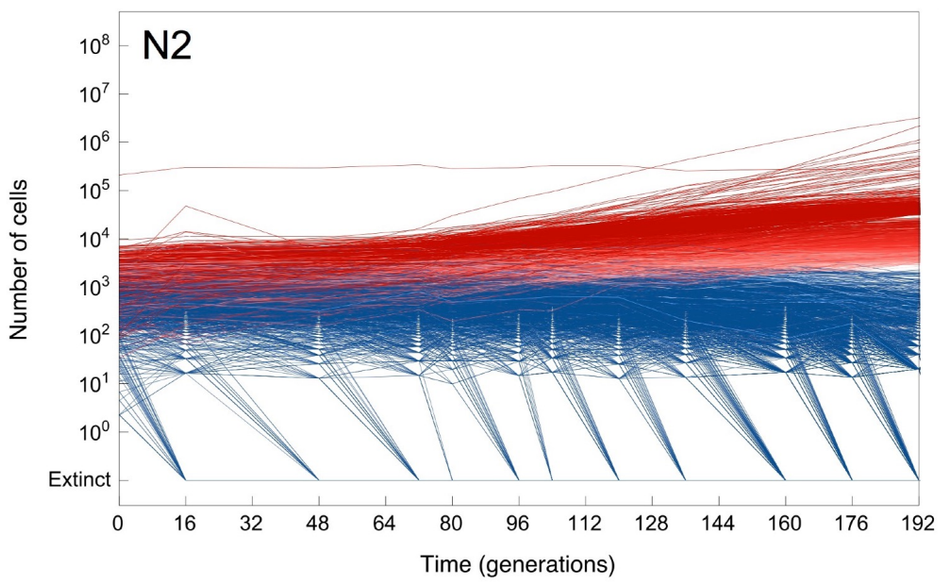


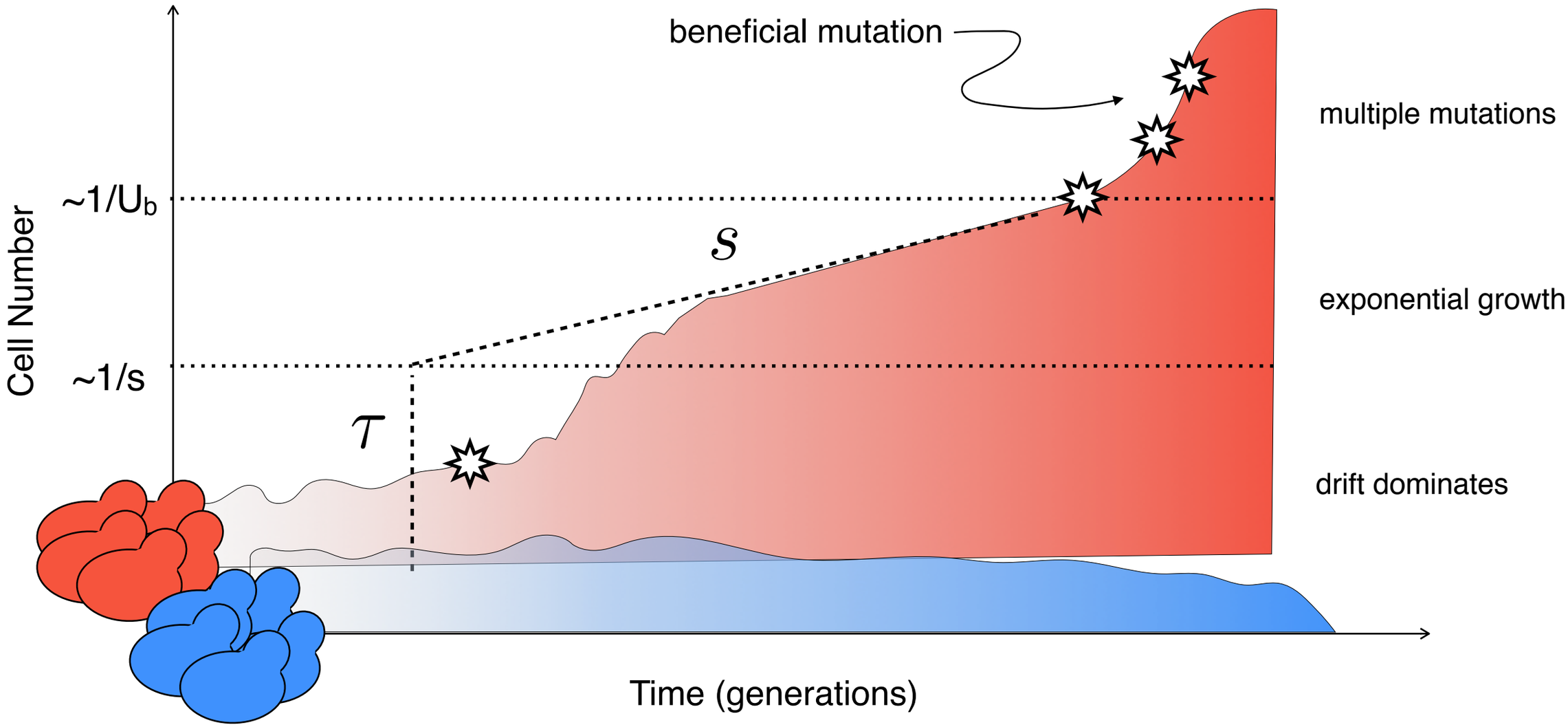



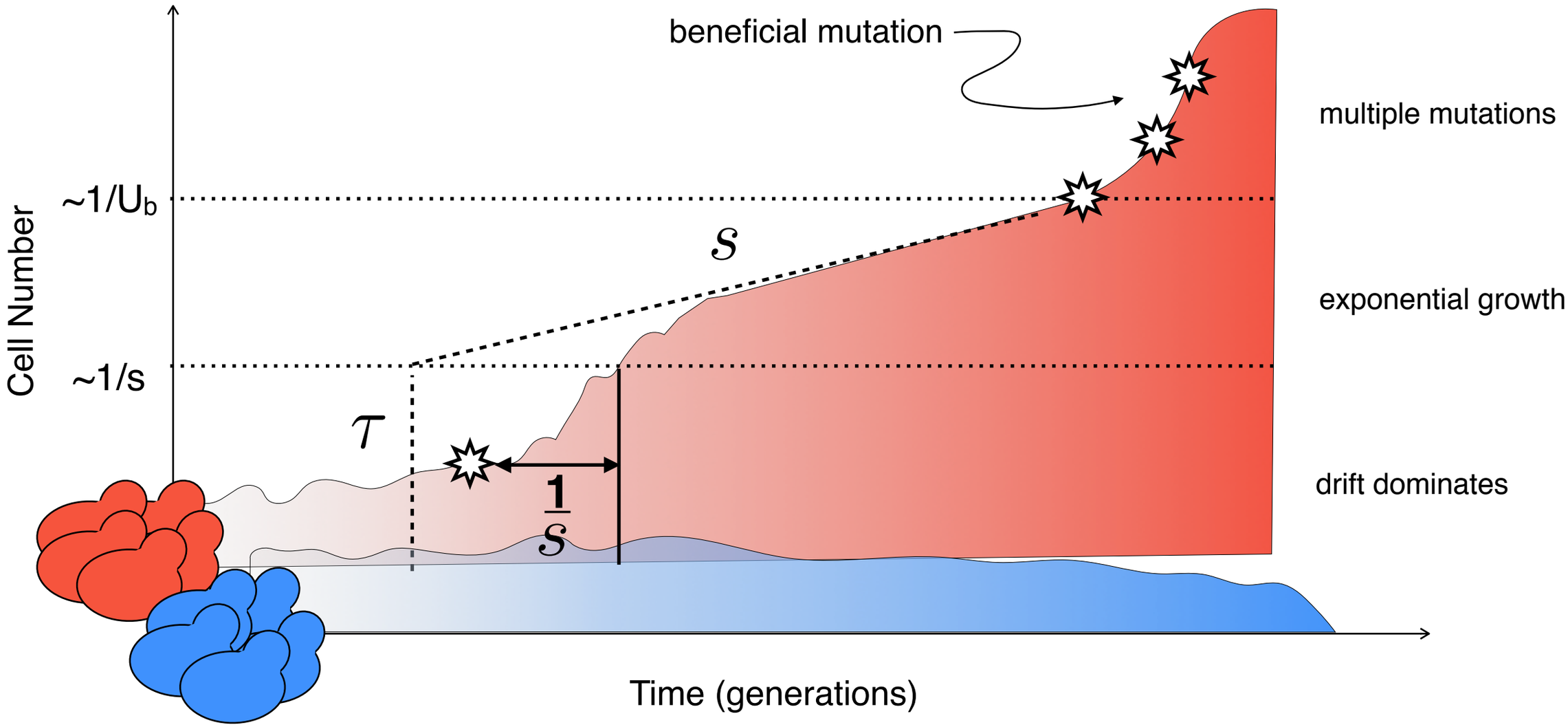



# A window of exponential growth measures the fitness effect of a single beneficial mutation



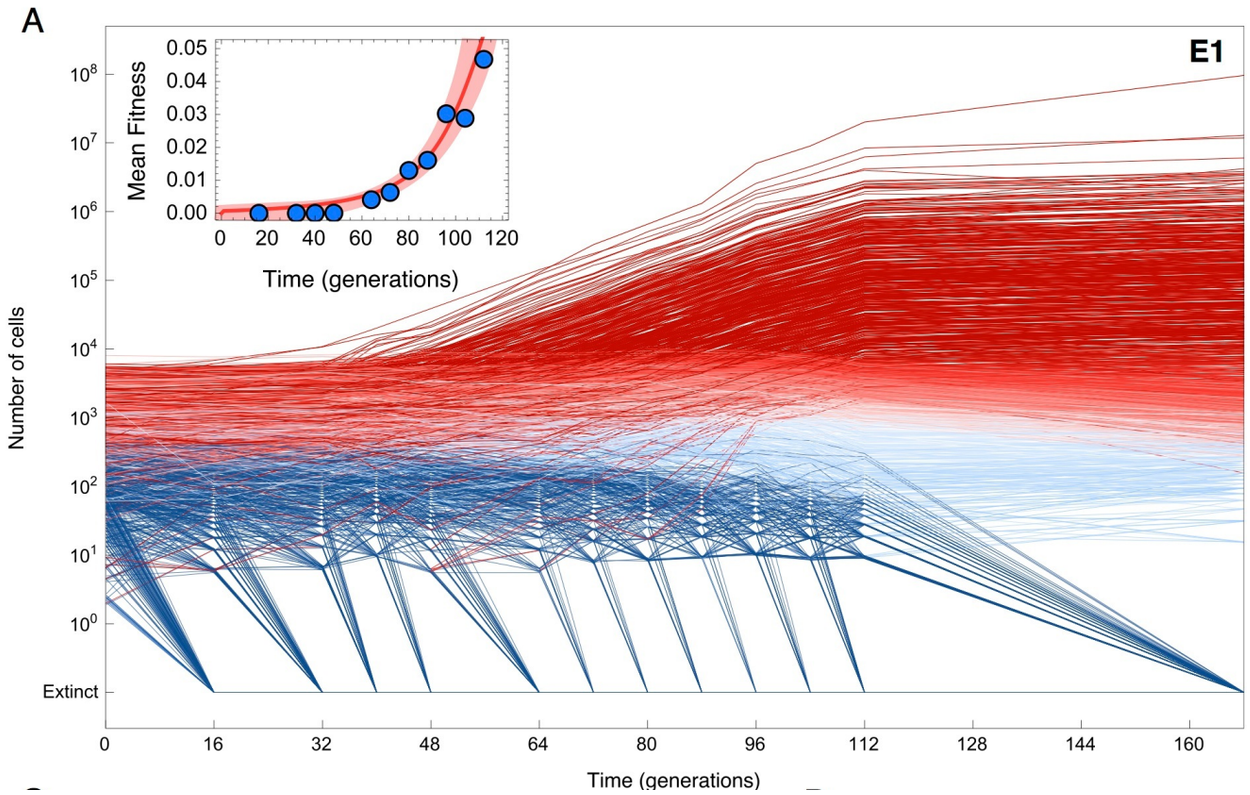

Michael Desai & Daniel Fisher

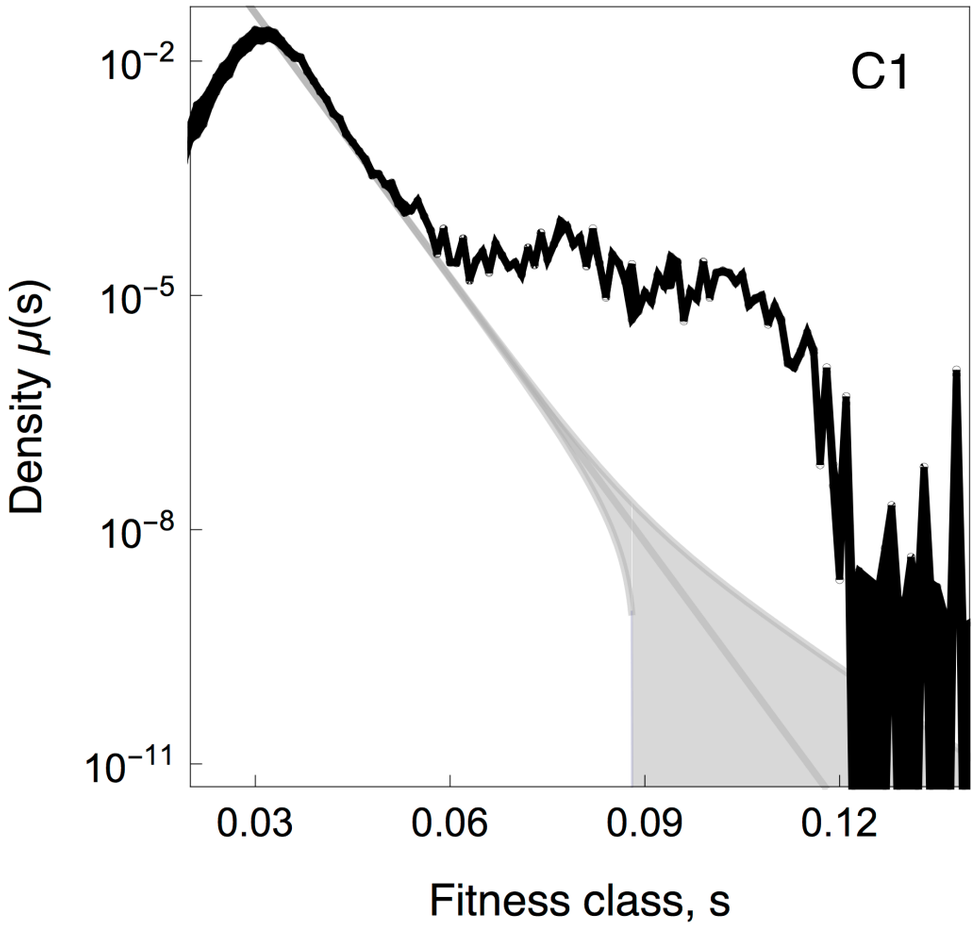
# Waiting time depends the the fitness effect

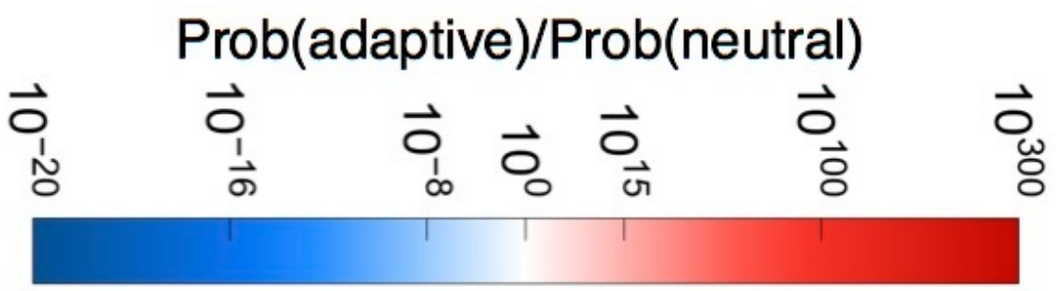



Michael Desai & Daniel Fisher

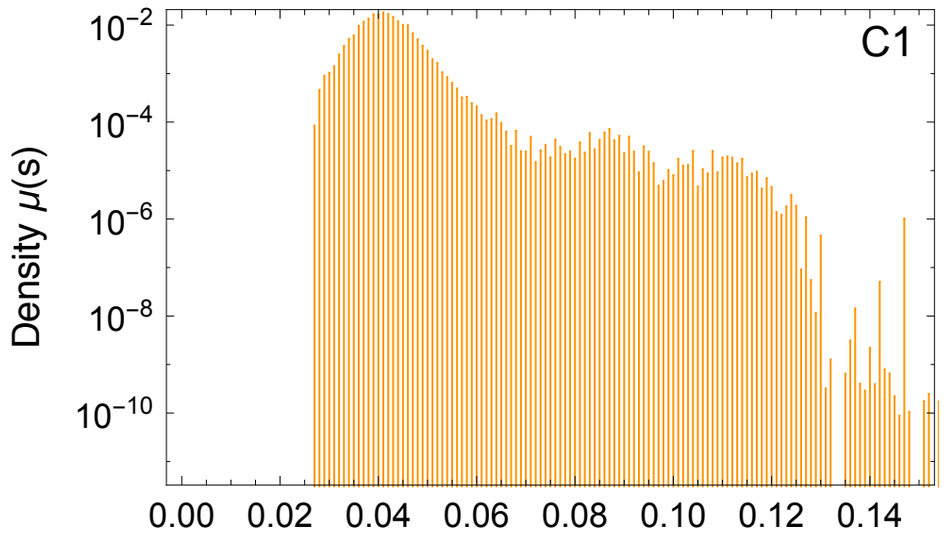
# Lineage trajectories of barcodes can be used to infer the distribution of beneficial fitness effects

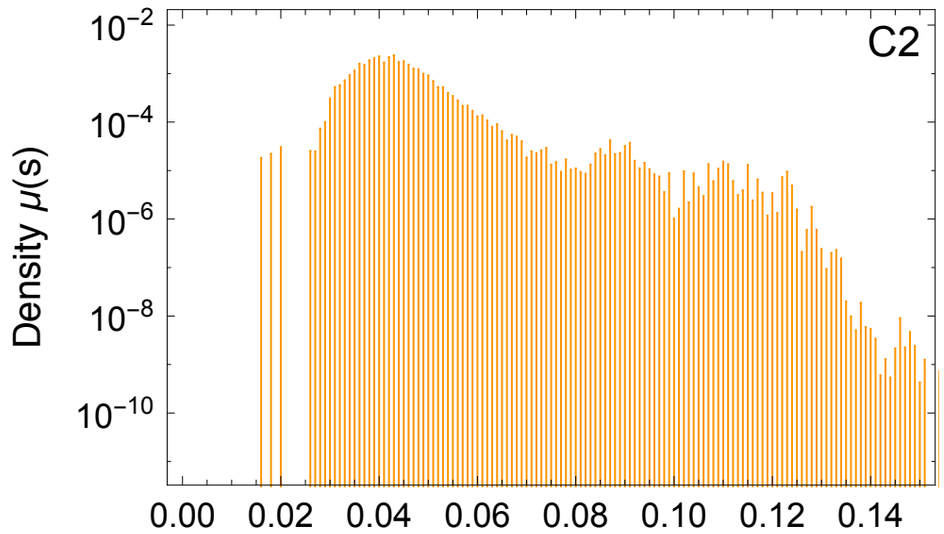


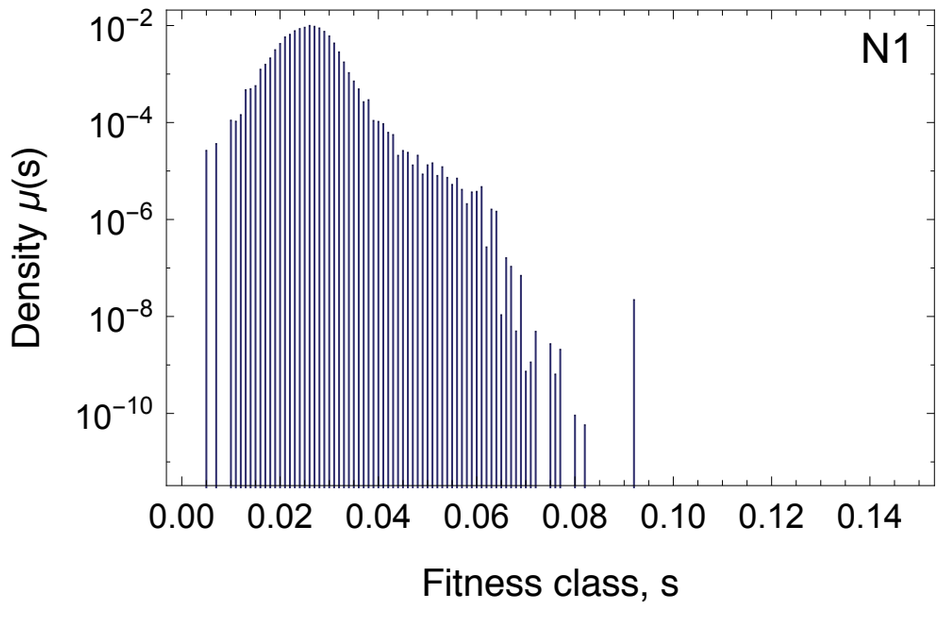



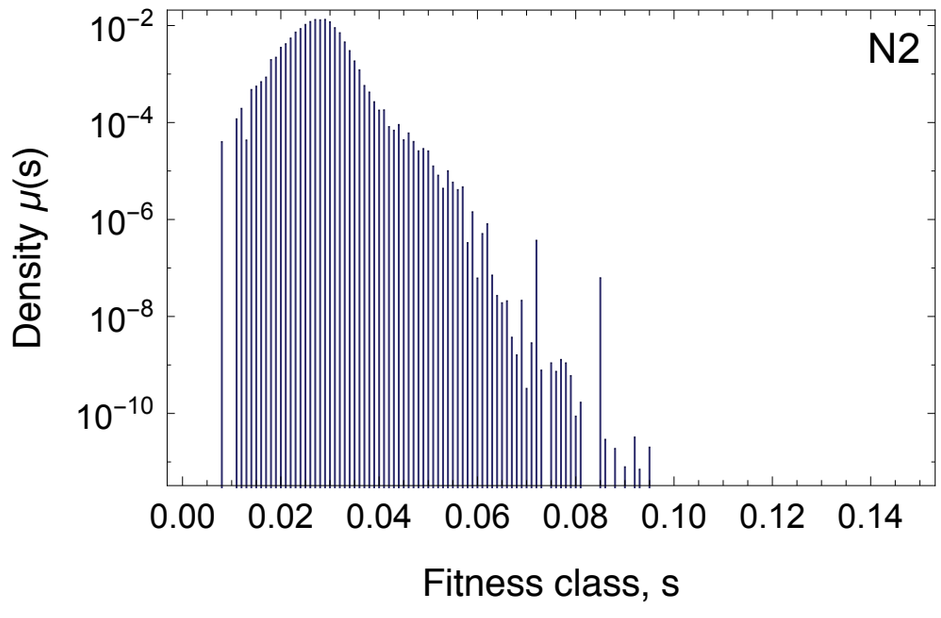



# DFEs in replicate evolutions

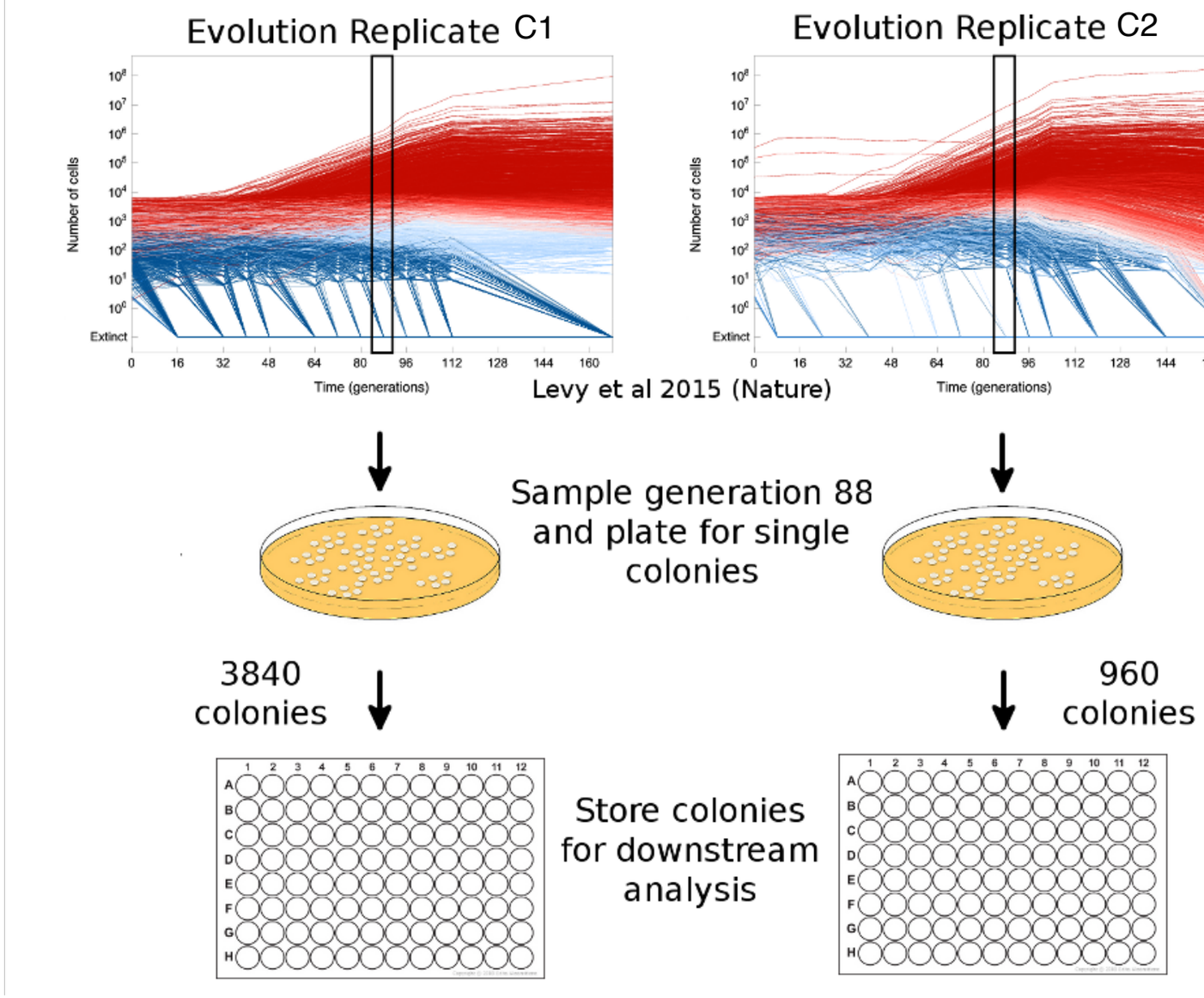


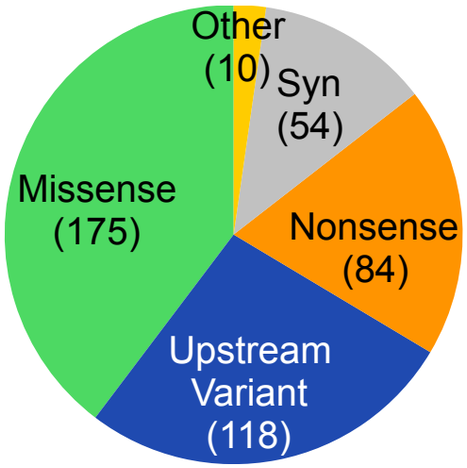
# Barcode-directed sequencing can be used characterize the adaptive mutational spectrum

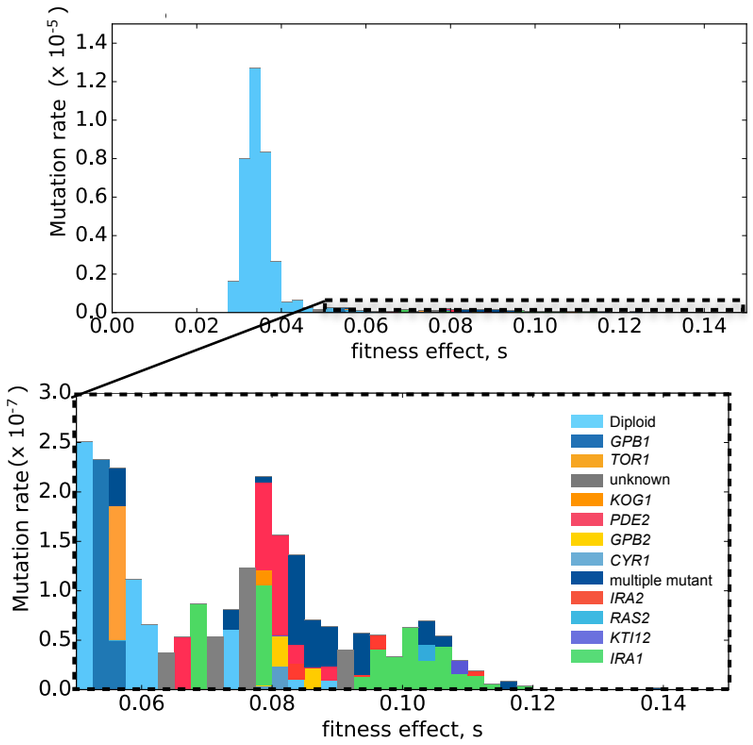



Venkataram et al., *Cell*, 2016

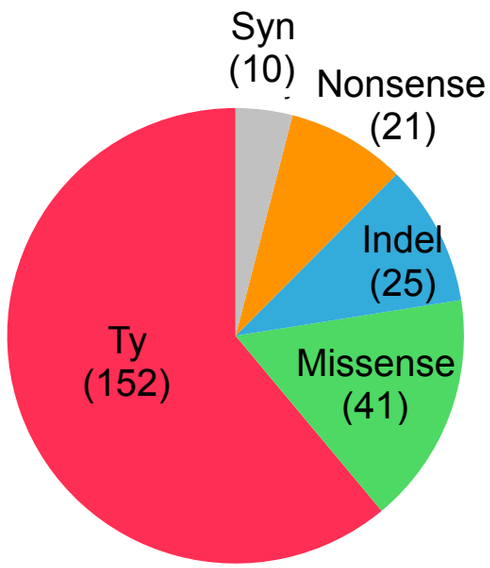
# The DFE and mutational spectrum depend on the environment

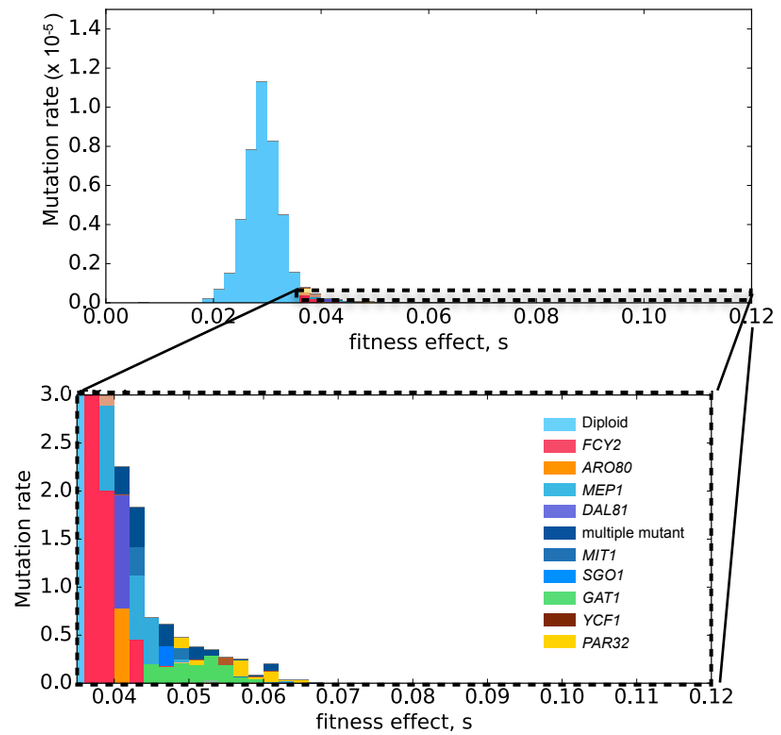
## Carbon Limitation




## Nitrogen Limitation

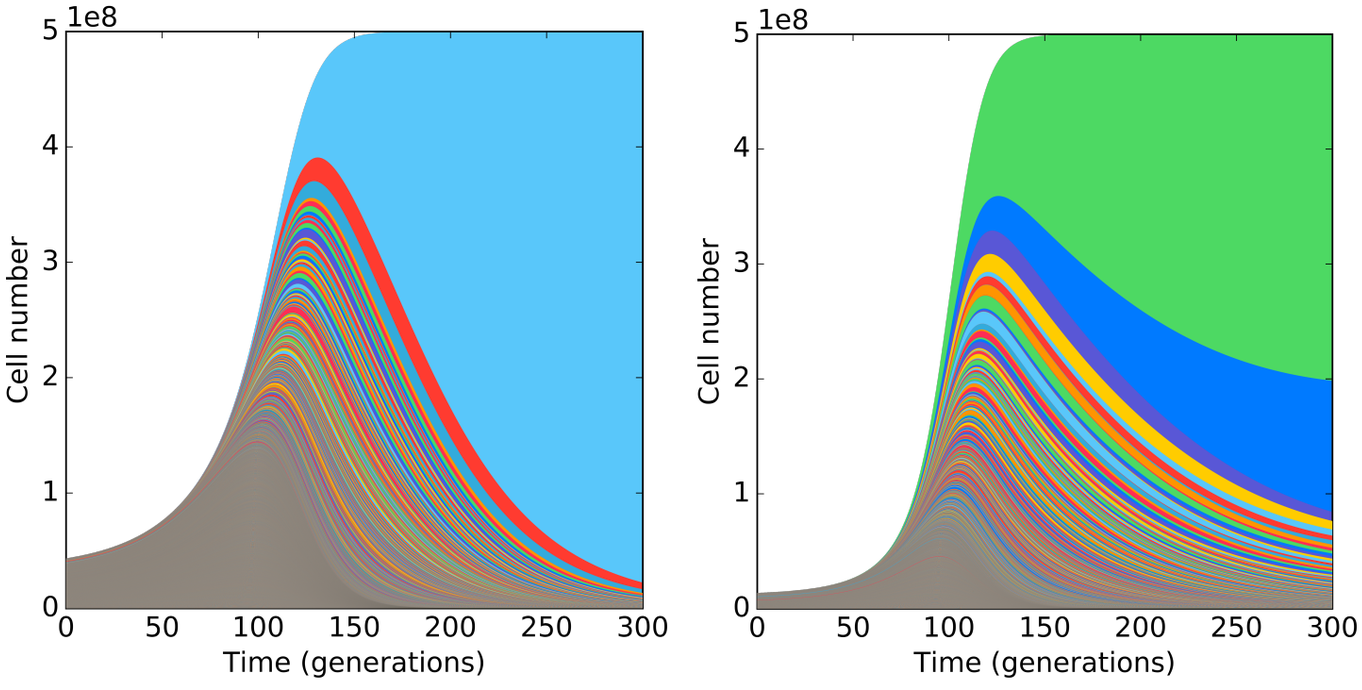



Blundell et al., *in review*

# What happens?

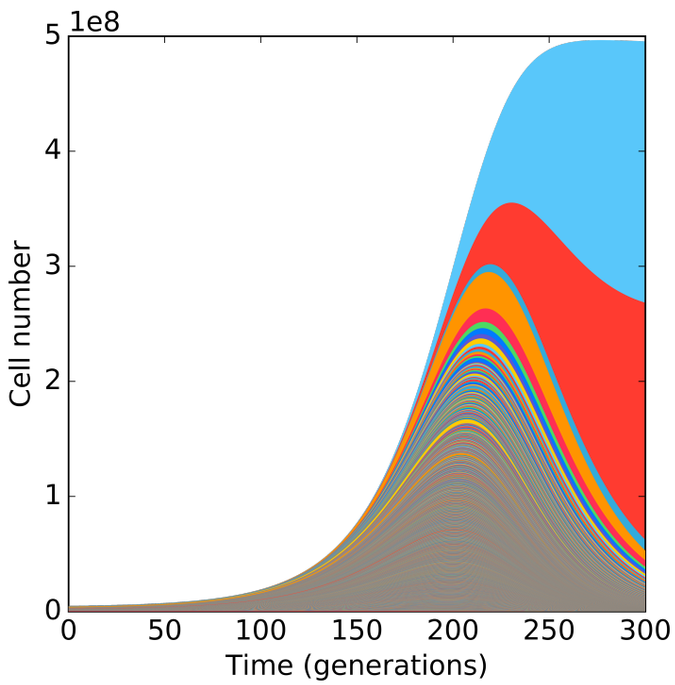
C-lim

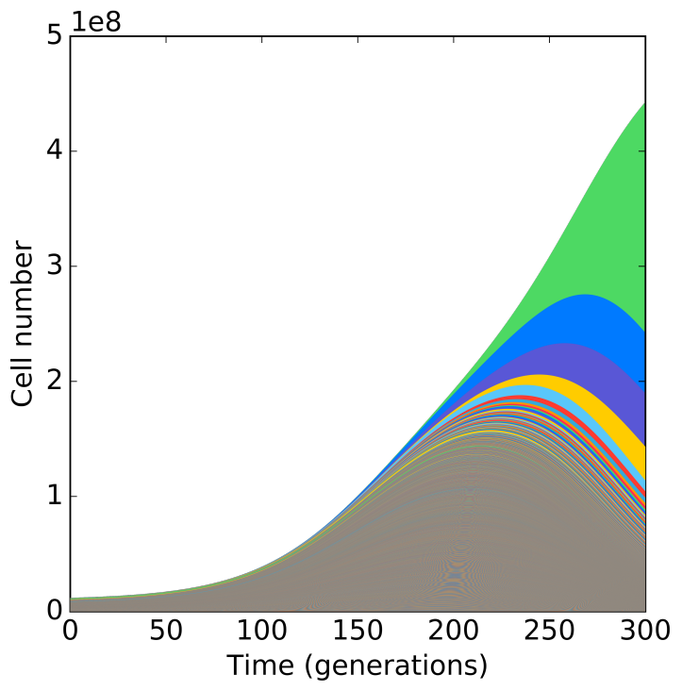






N-lim

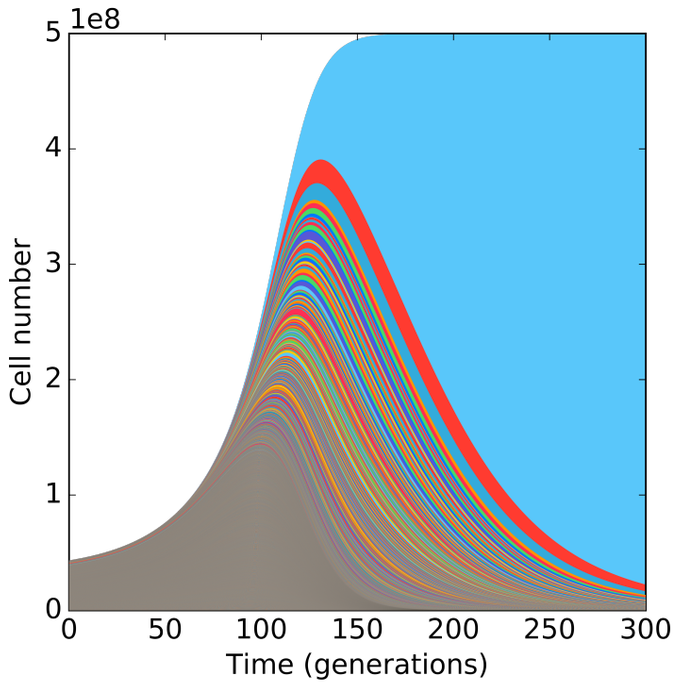


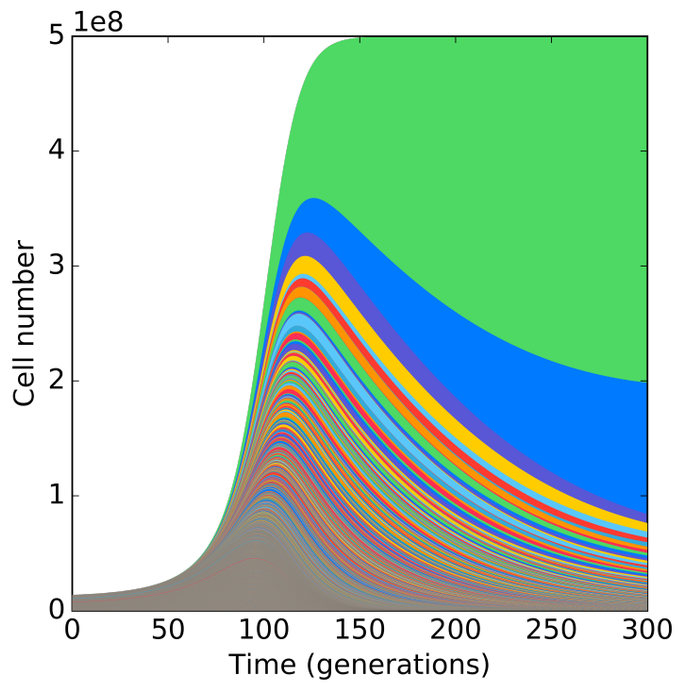




Blundell et al., *in review*

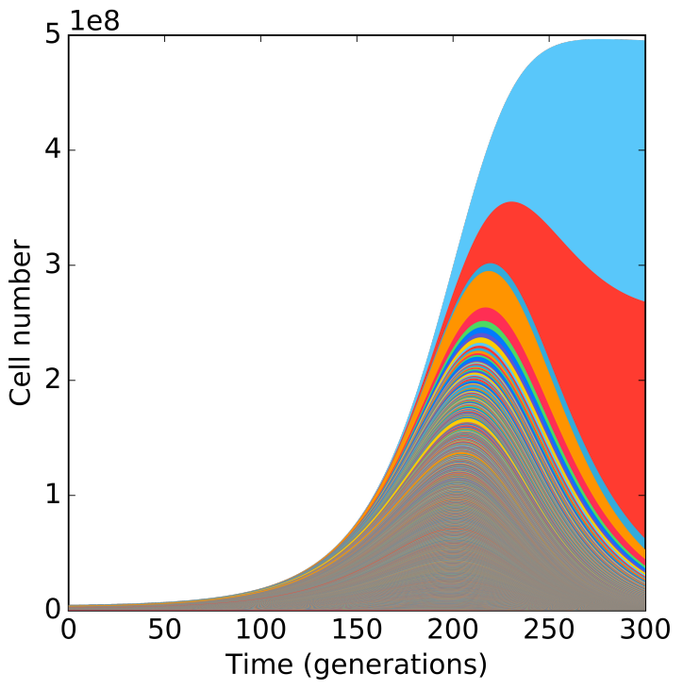
# What happens?

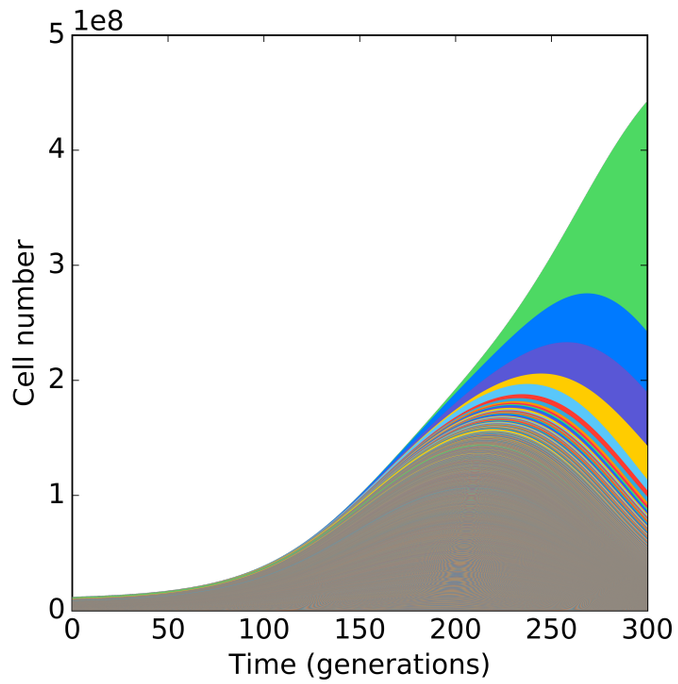
C-lim




N-lim




## Some common features

1) First, many lineages expand, increasing diversity quasi-deterministically

2) With time, the expansion becomes more stochastic

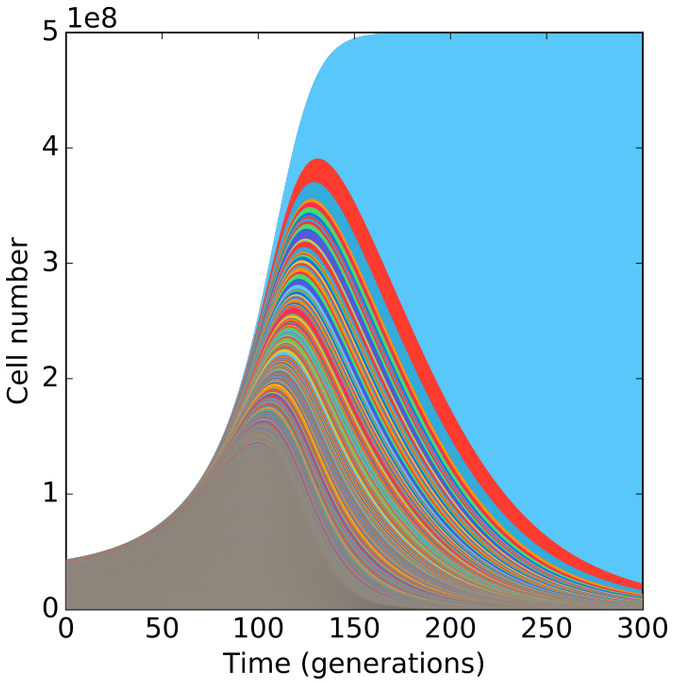
3) A highly stochastic diversity crash follows

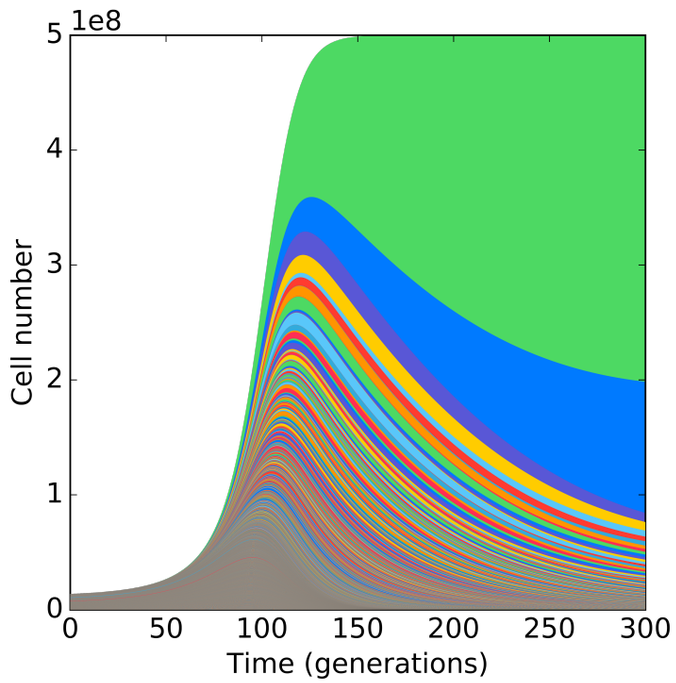
4) A handful of lineages dominate late

Blundell et al., *in review*

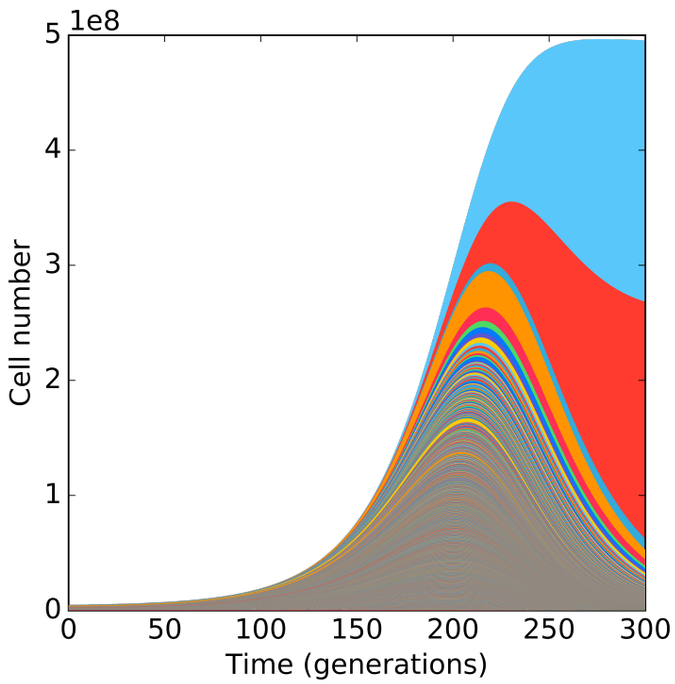
# What happens?

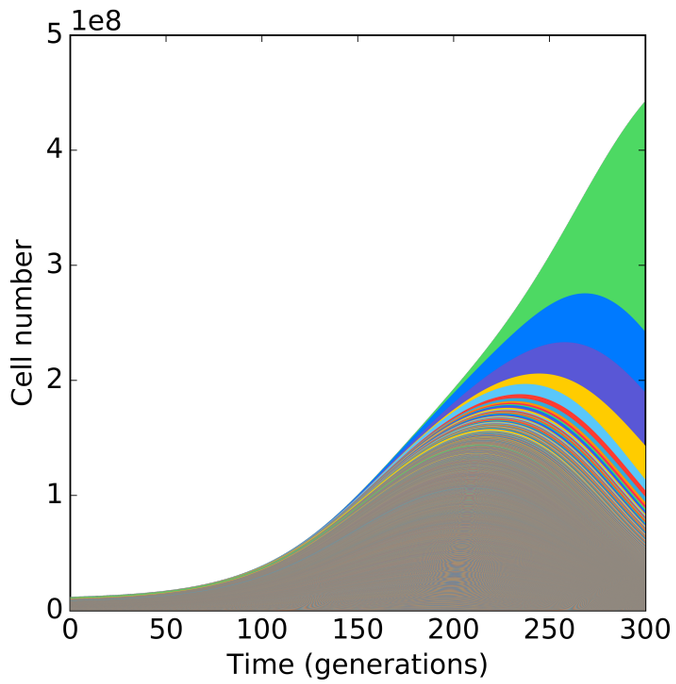
C-lim




N-lim




## Some common features

1) *First, many lineages expand, increasing diversity quasi-deterministically*

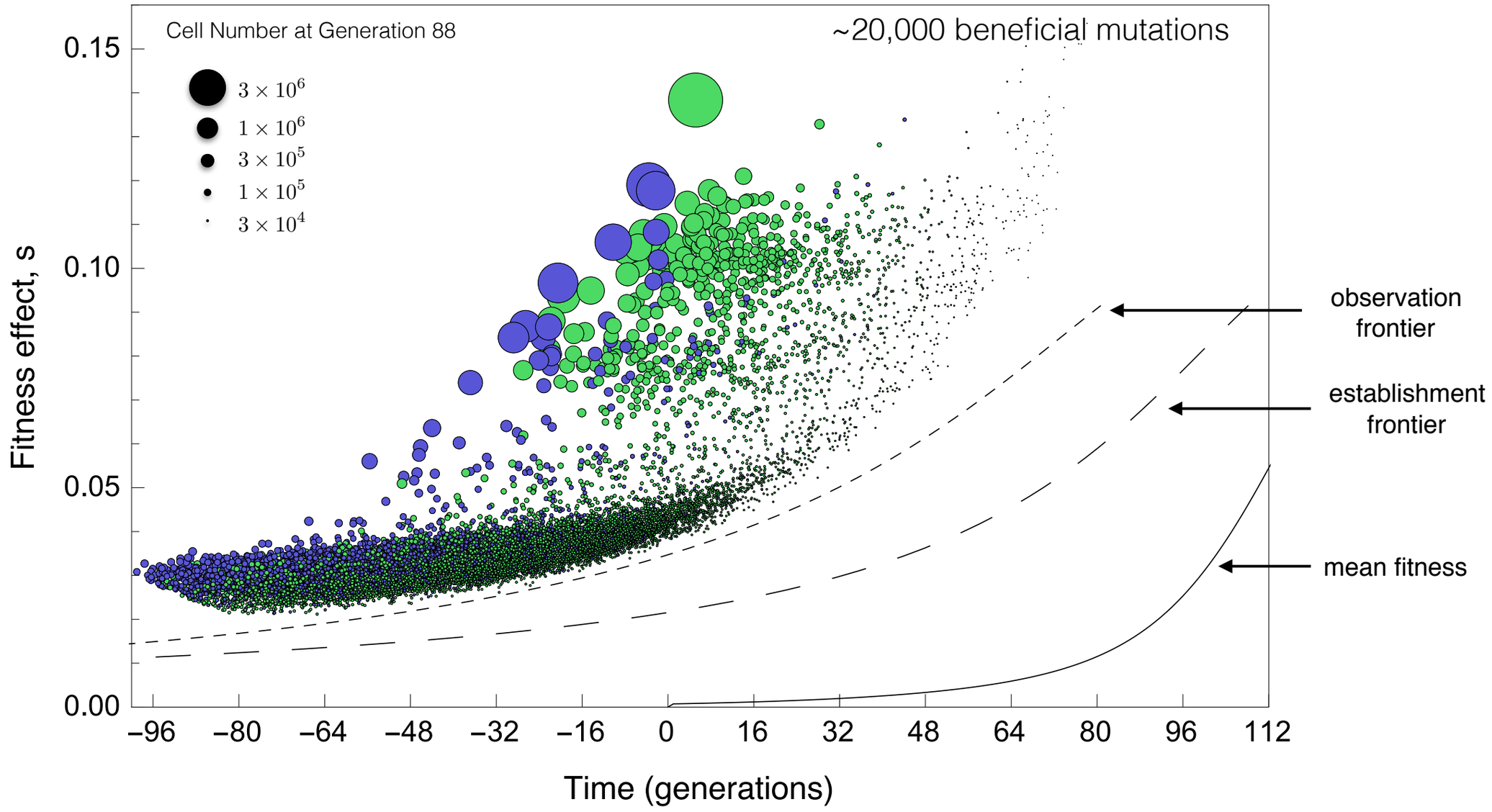
2) *With time, the expansion becomes more stochastic*

3) A highly stochastic diversity crash follows

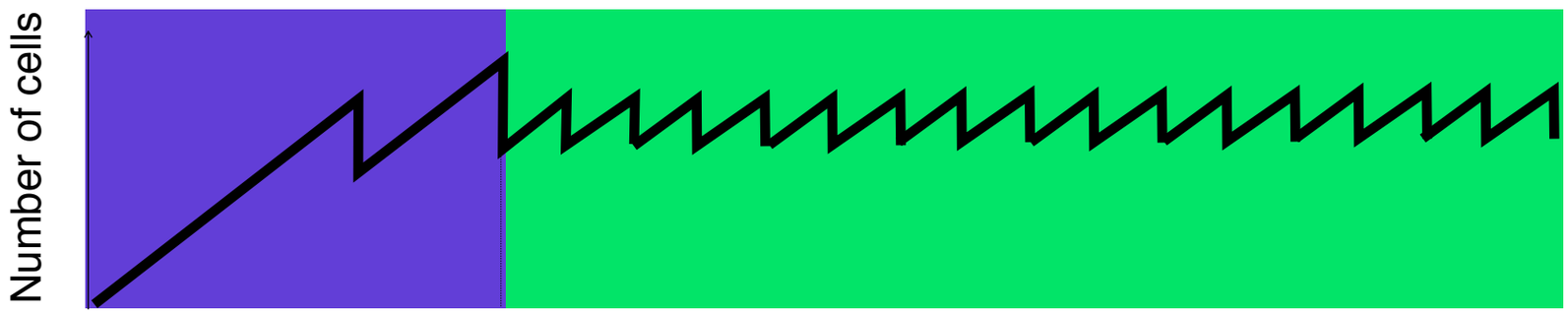
4) A handful of lineages dominate late

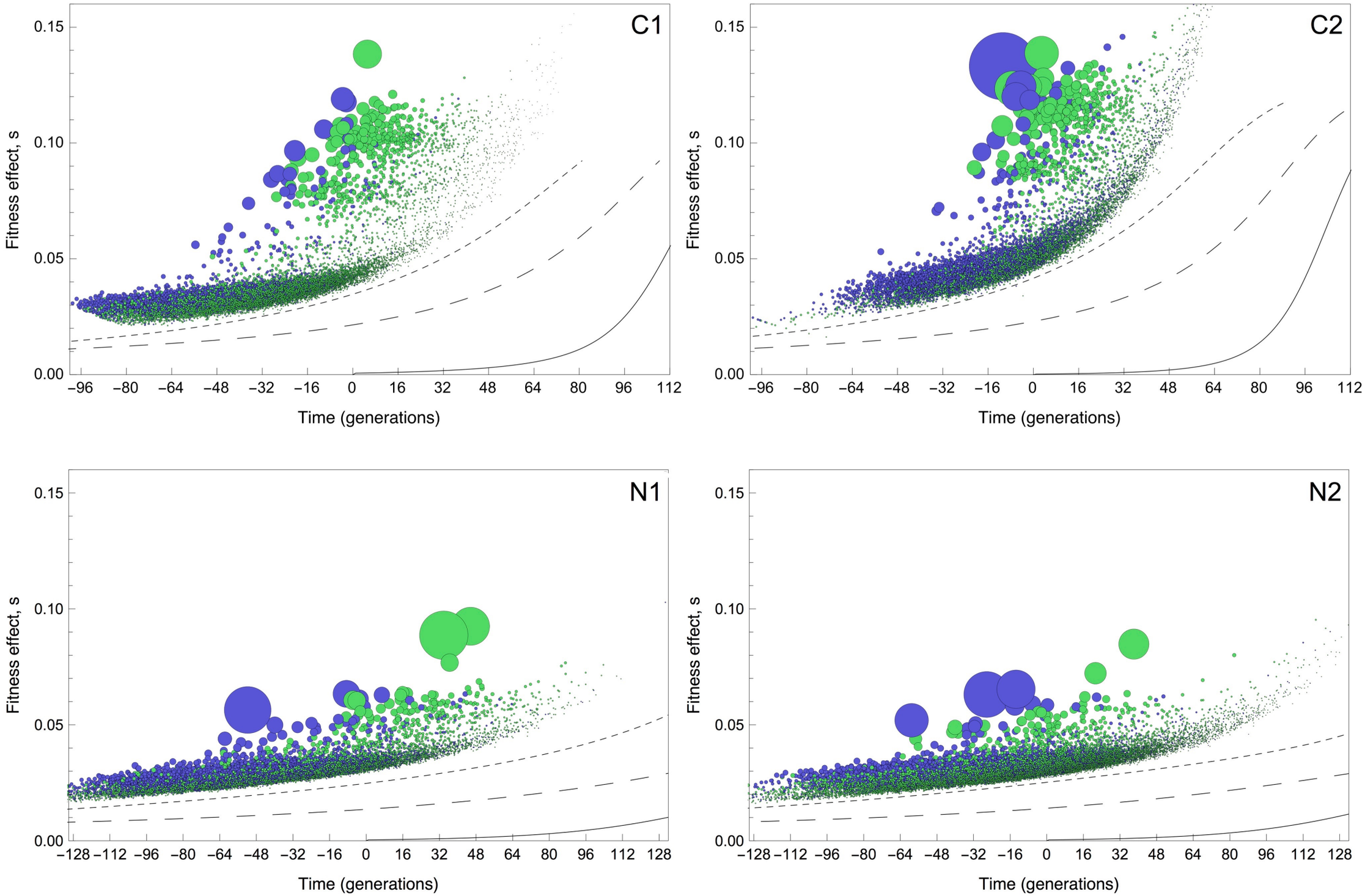
Blundell et al., *in review*

# The effective beneficial mutation rate determines whether the diversity expansion is deterministic or stochastic



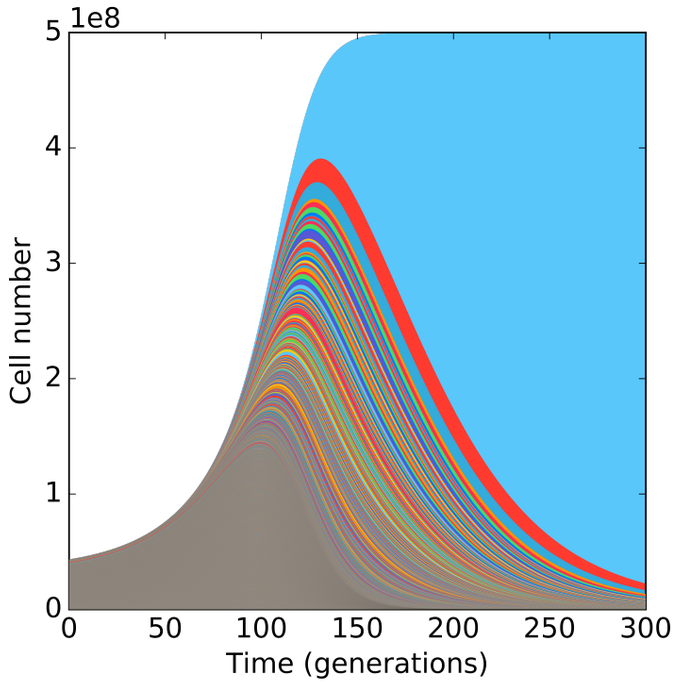

**As the mean fitness increases, fewer and fewer mutations have high enough fitness effects to establish

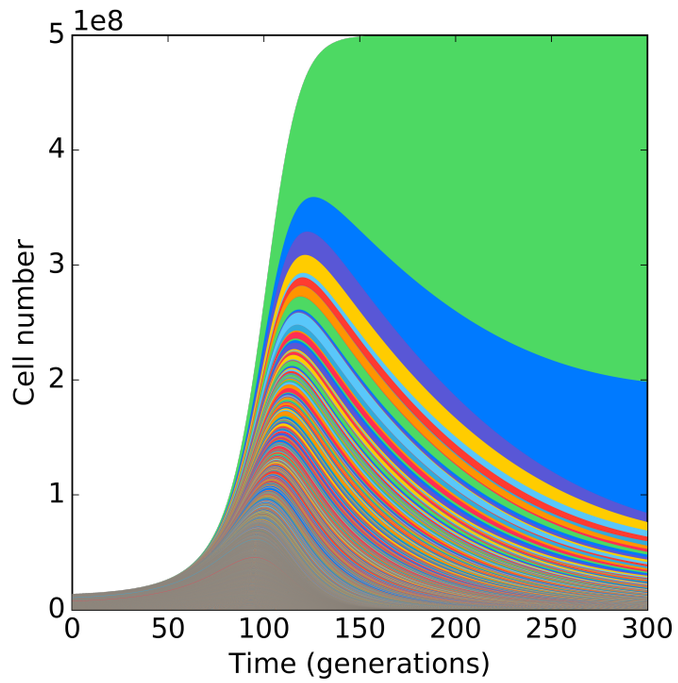



C1
C2
N1
N2
Fitness effect, s
0.00
0.05
0.10
0.15
−128 −112 −96 −80 −64 −48 −32 −16 0 16 32 48 64 80 96 112 128
Time (generations)

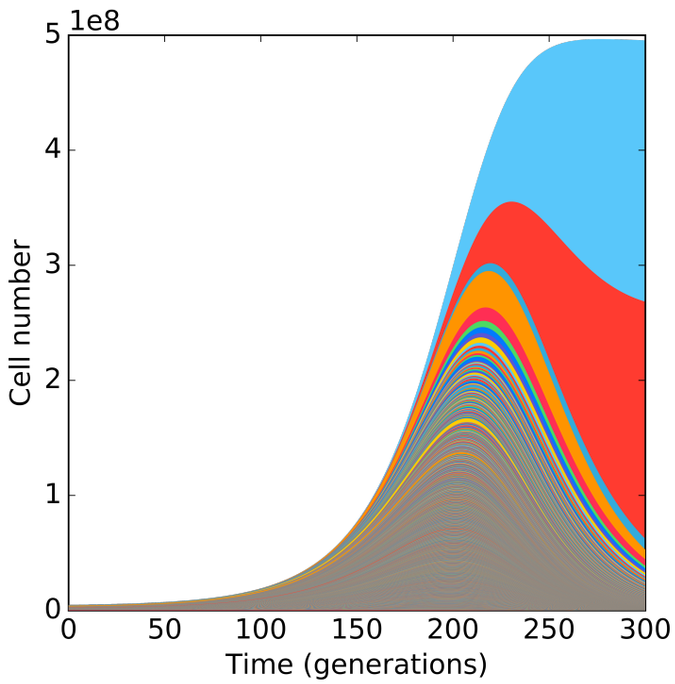
# What happens?

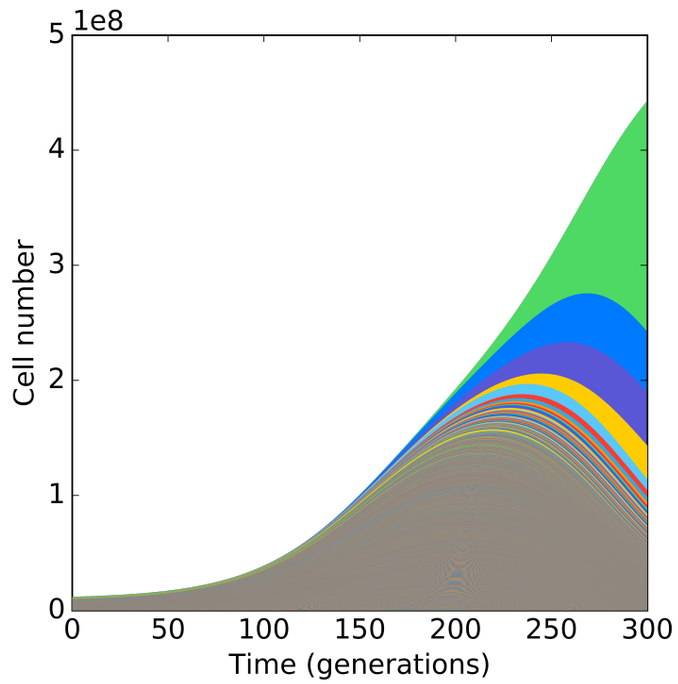
C-lim




N-lim




## Some common features

1) First, many lineages expand, increasing diversity quasi-deterministically

2) With time, the expansion becomes more stochastic

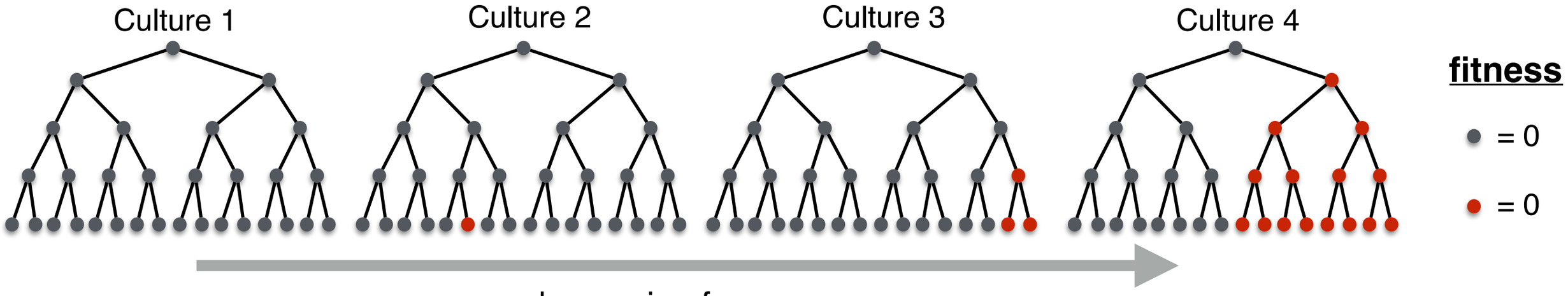
*3) A highly stochastic diversity crash follows*

*4) A handful of lineages dominate late*

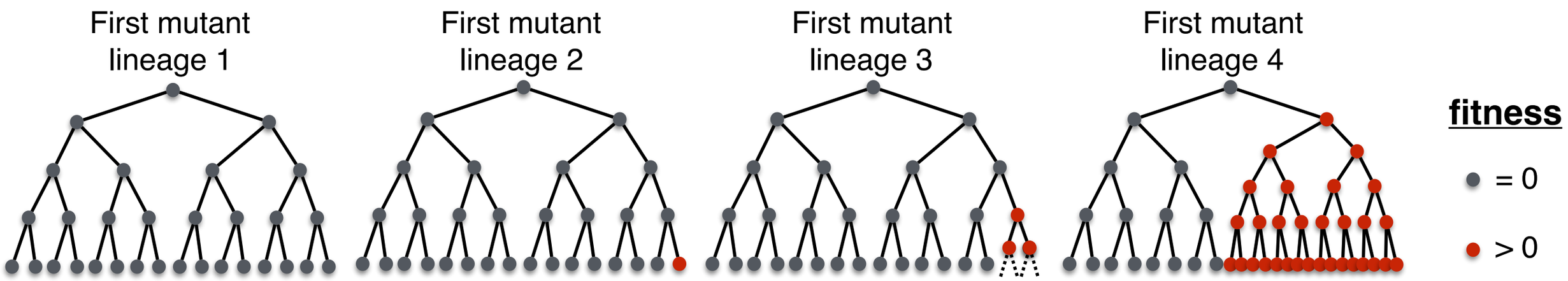
Blundell et al., *in review*

# Luria-Delbruck vs. Clonal Evolution

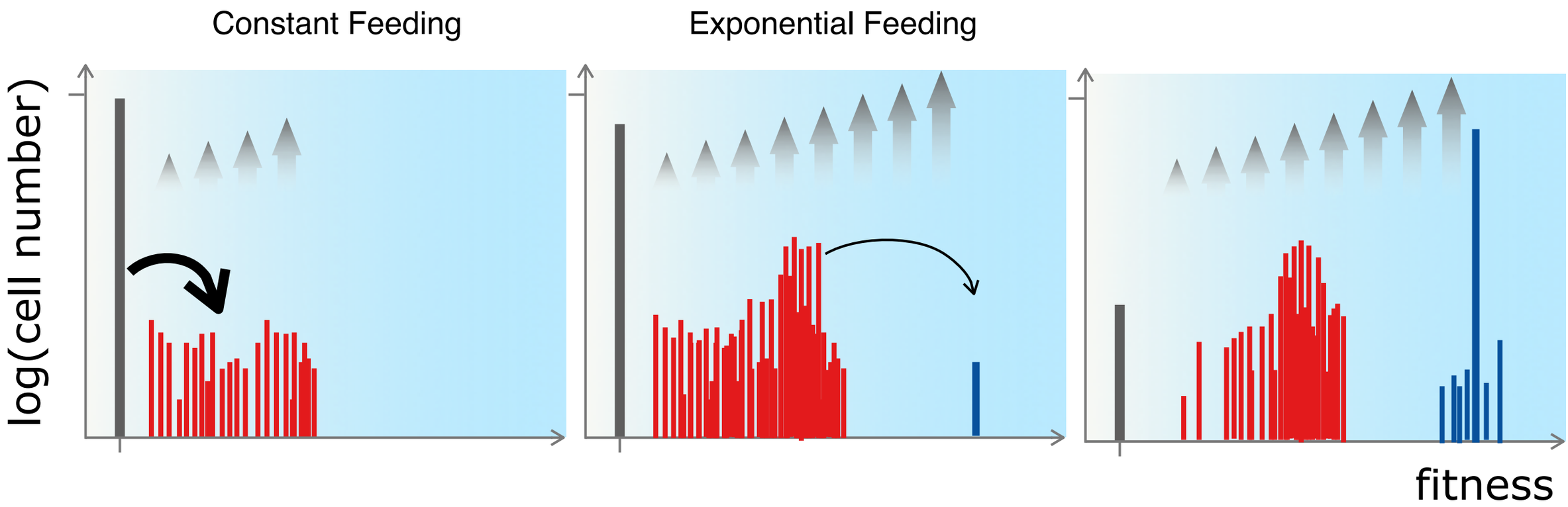
## Luria-Delbruk—mutations grow at the same rate




## Clonal evolution—mutations grow faster

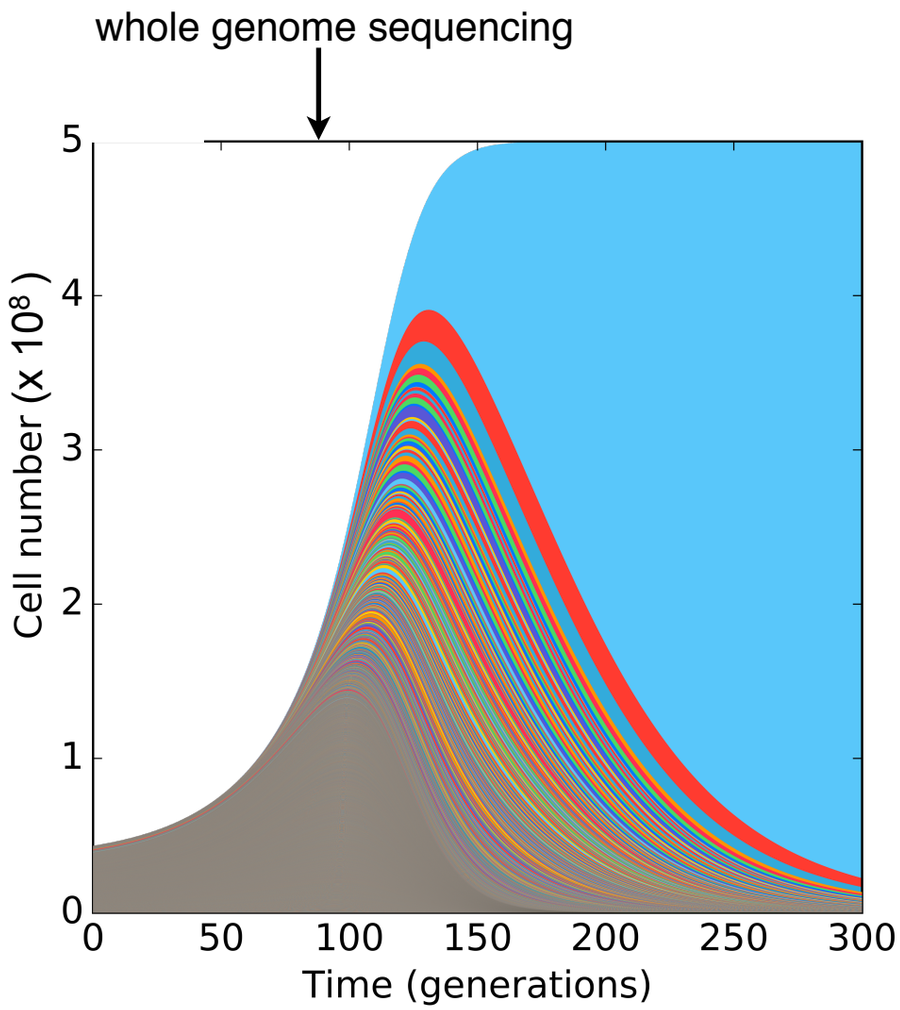


# Dynamics of first and second mutations



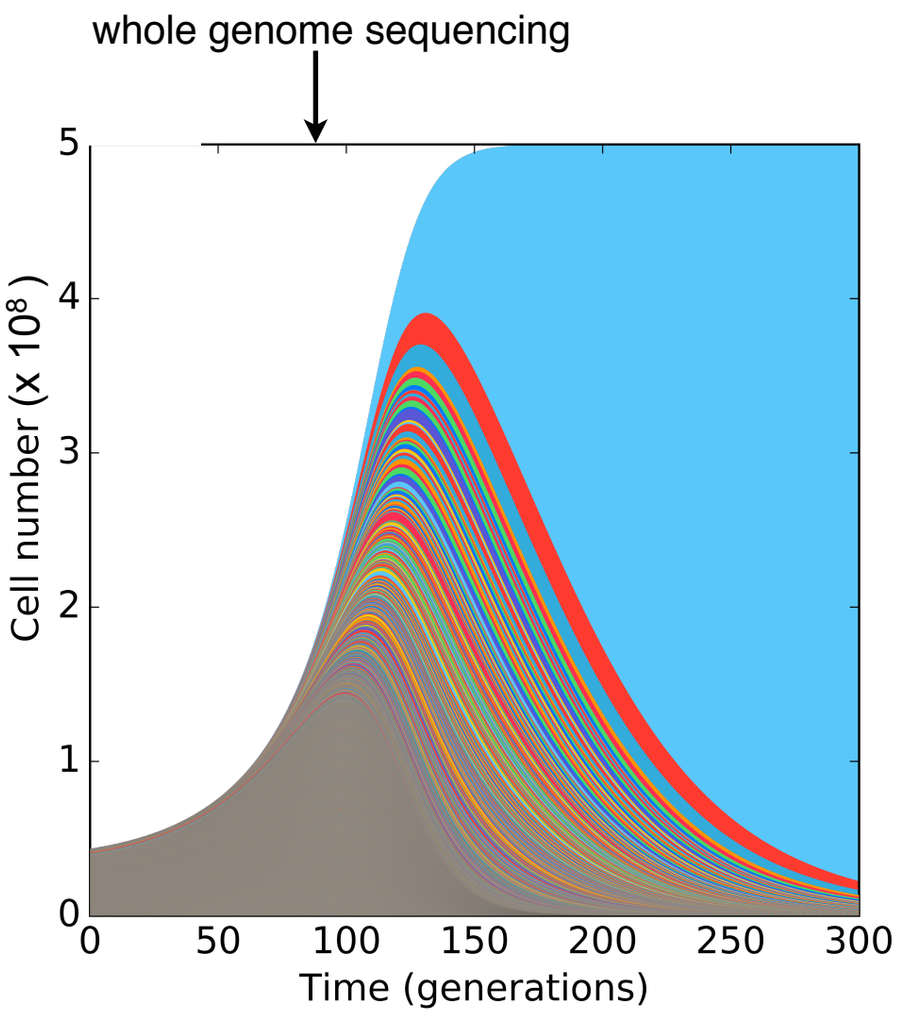

Blundell et al., *in review*

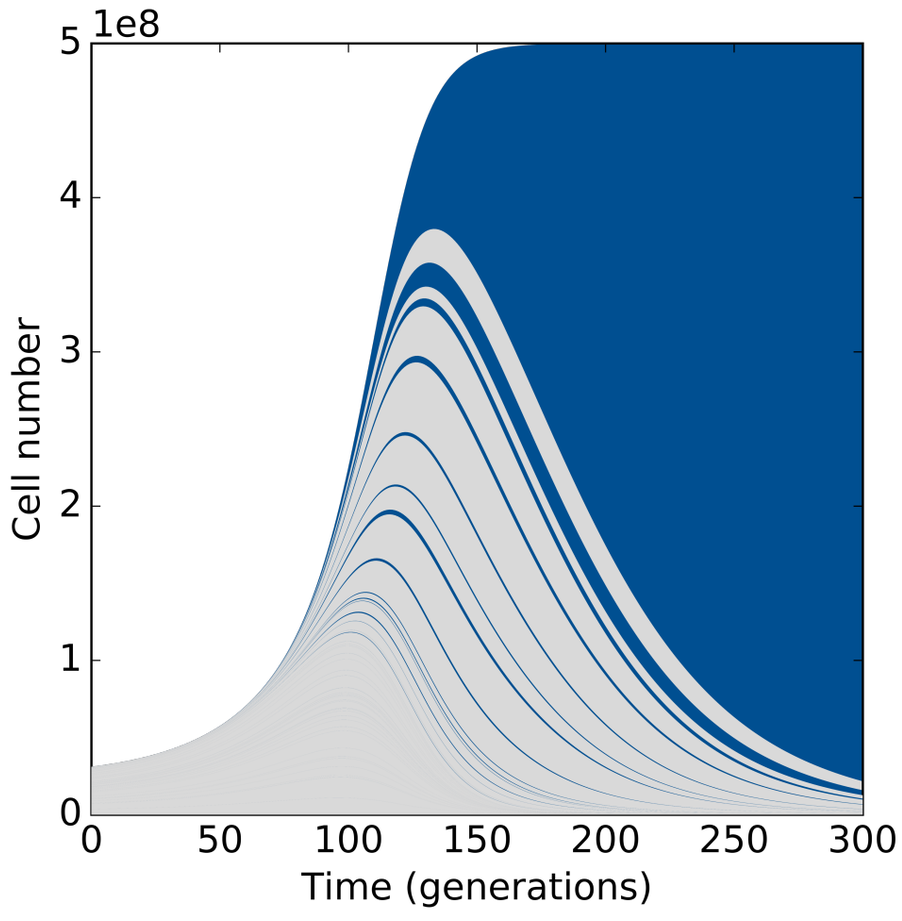
# Prediction: dominant lineages should contain early second mutations




Blundell et al., *in review*

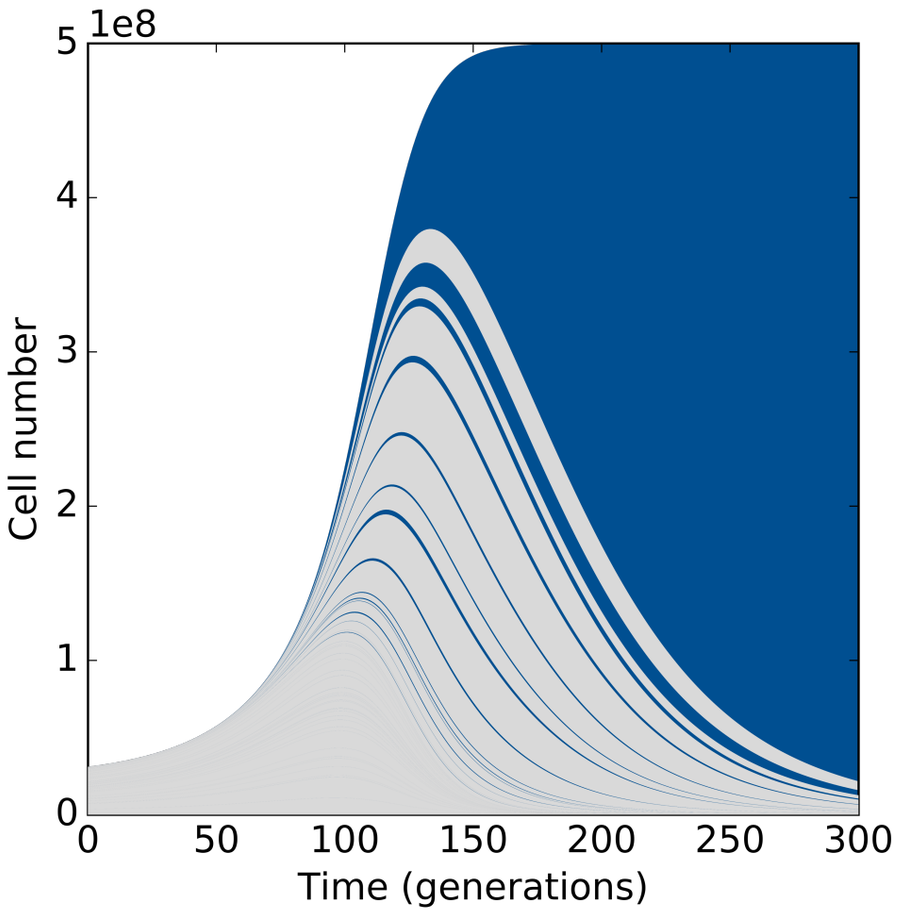
# Prediction: dominant lineages should contain early second mutations




Blundell et al., *in review*

# Some thoughts on mutational cohorts



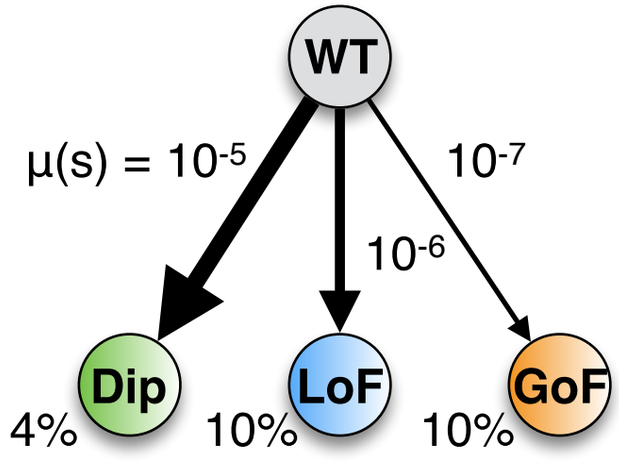

1) Adaptive cohorts are a natural feature of the adaptive mutation accumulation process

2) Population size controls the cohort size

Blundell et al., *in review*

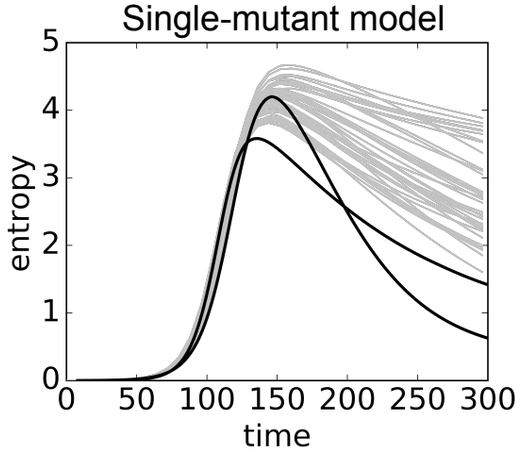
# Simulations using the DFE can assign outcome probabilities

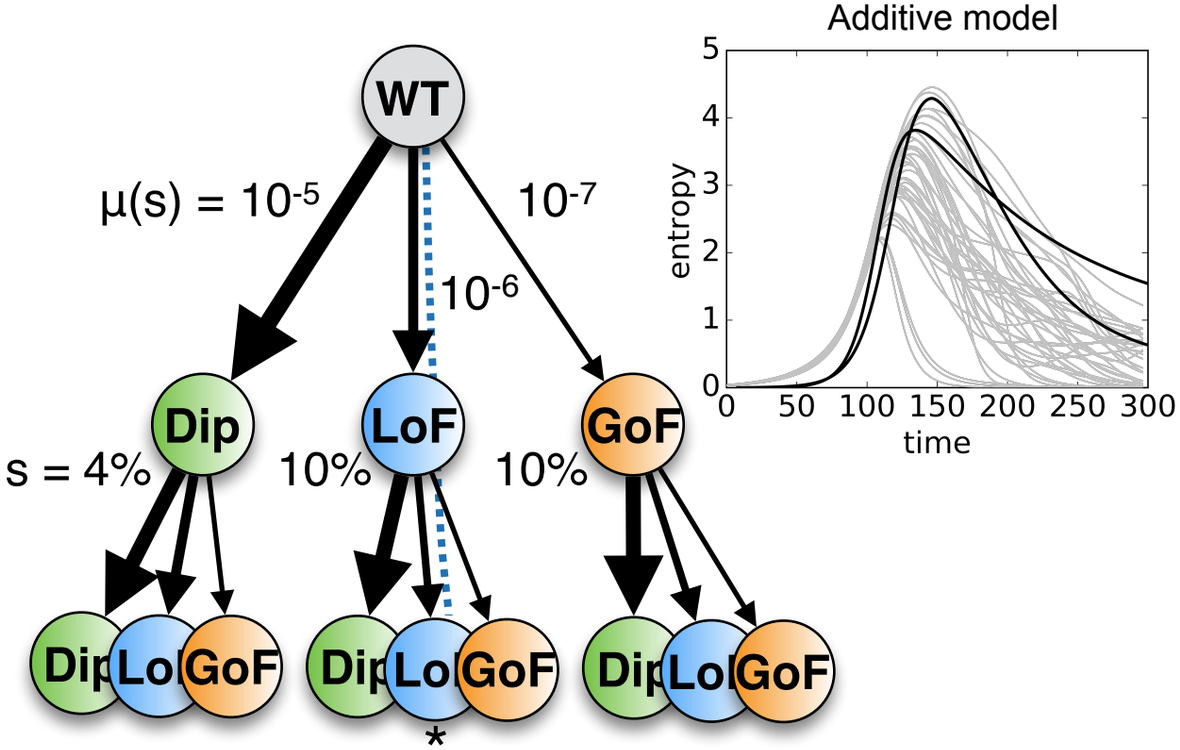
## fine print: epistasis matters

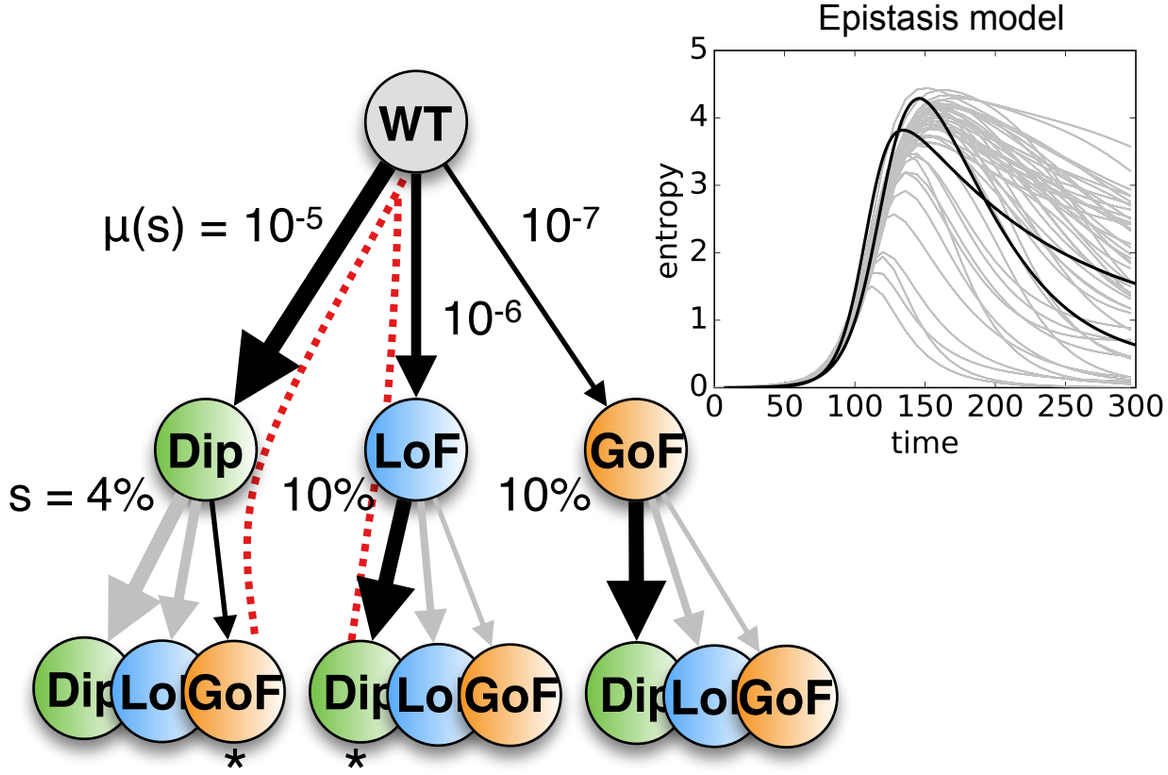



simulation

experiment



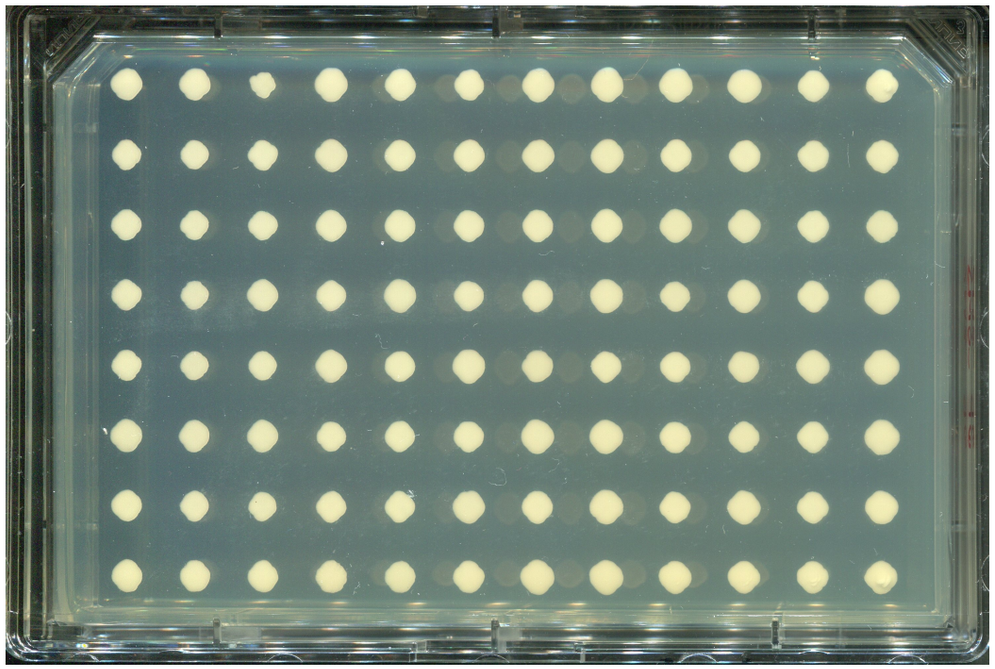




* Most likely dominant double mutants

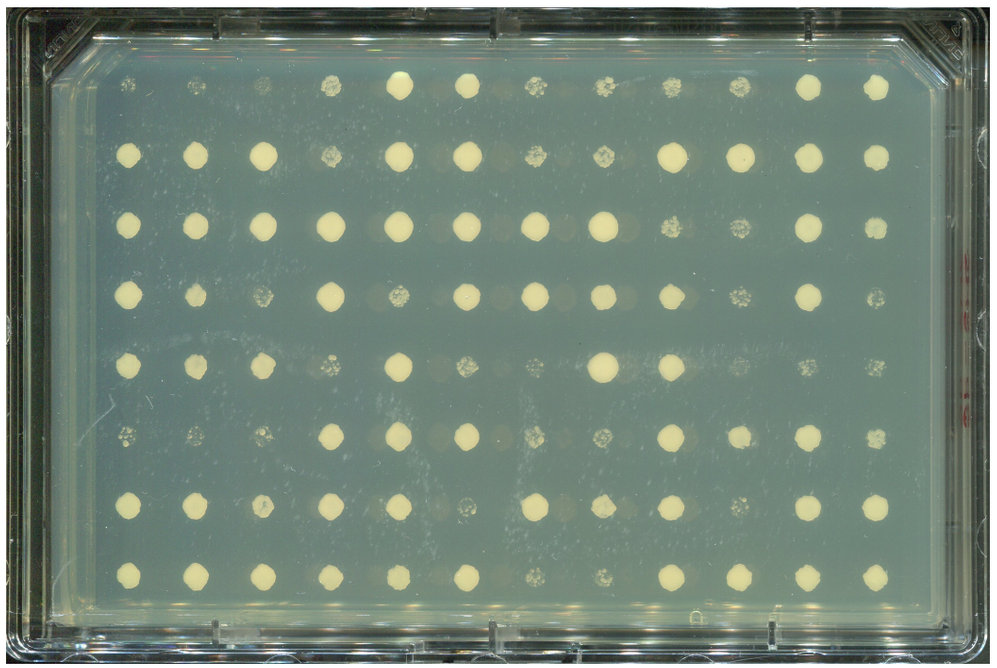
# Testing adaptation models with diploid trajectories

a



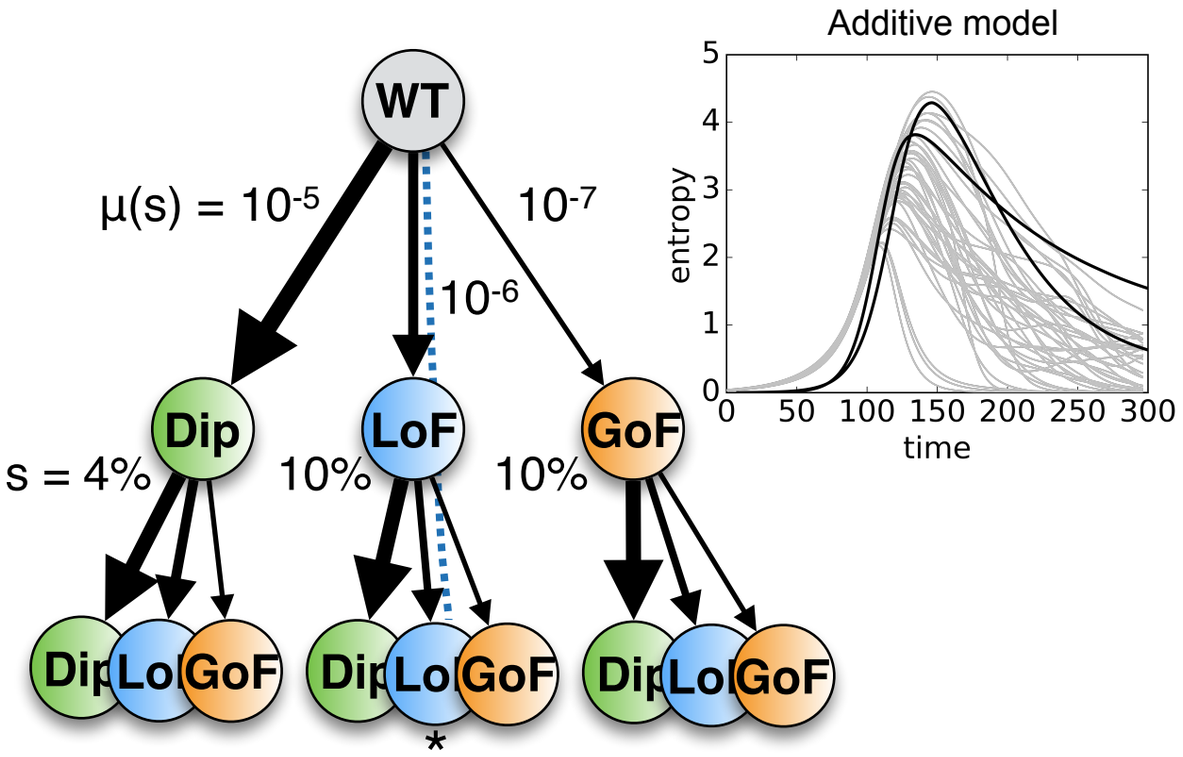
- Benomyl

b



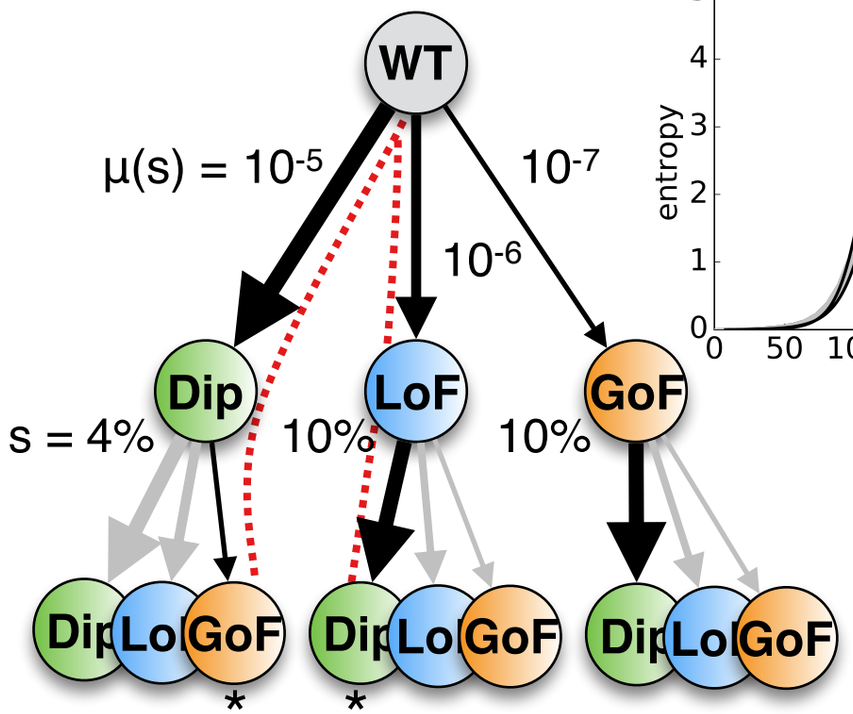
+ Benomyl

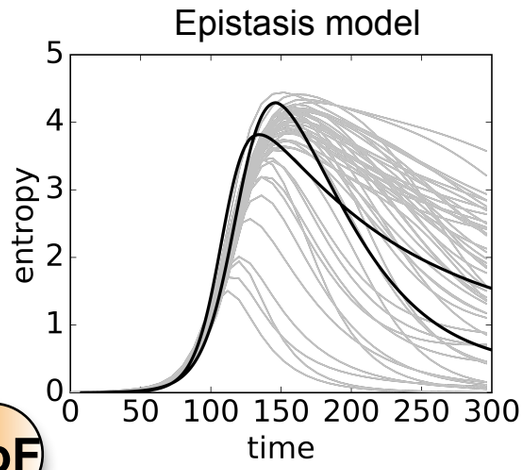
# Testing adaptation models with diploid trajectories

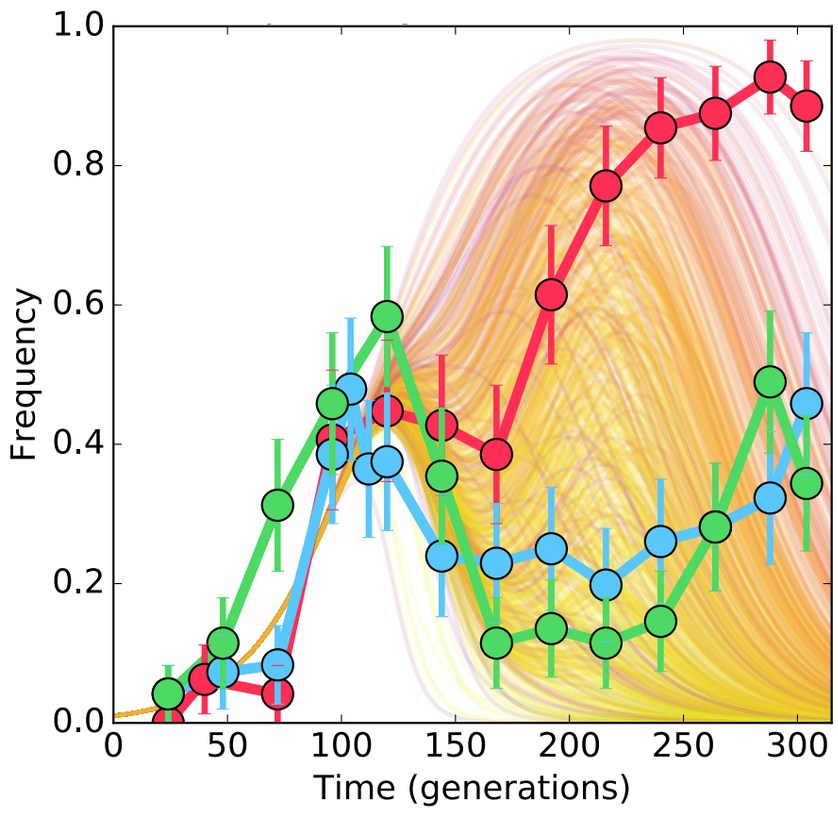


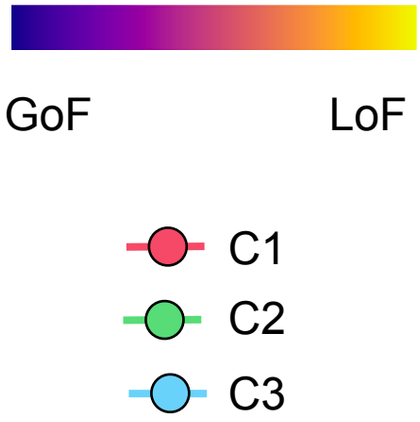


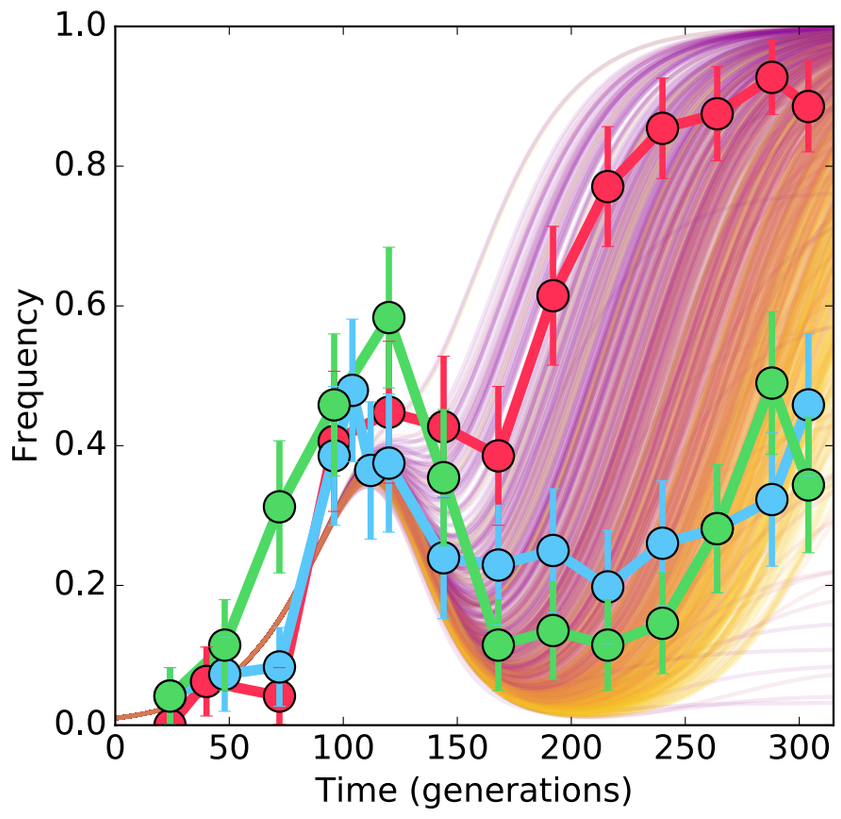




# Conclusions and Insights

1. Barcode-based lineage tracking measures the distribution of beneficial fitness effects

2. Barcode-directed sequencing discovers the mutational spectrum of adaptation-driving mutations

3. Early evolution of large populations is predictable, and is set by distribution of beneficial fitness effects in a particular environment

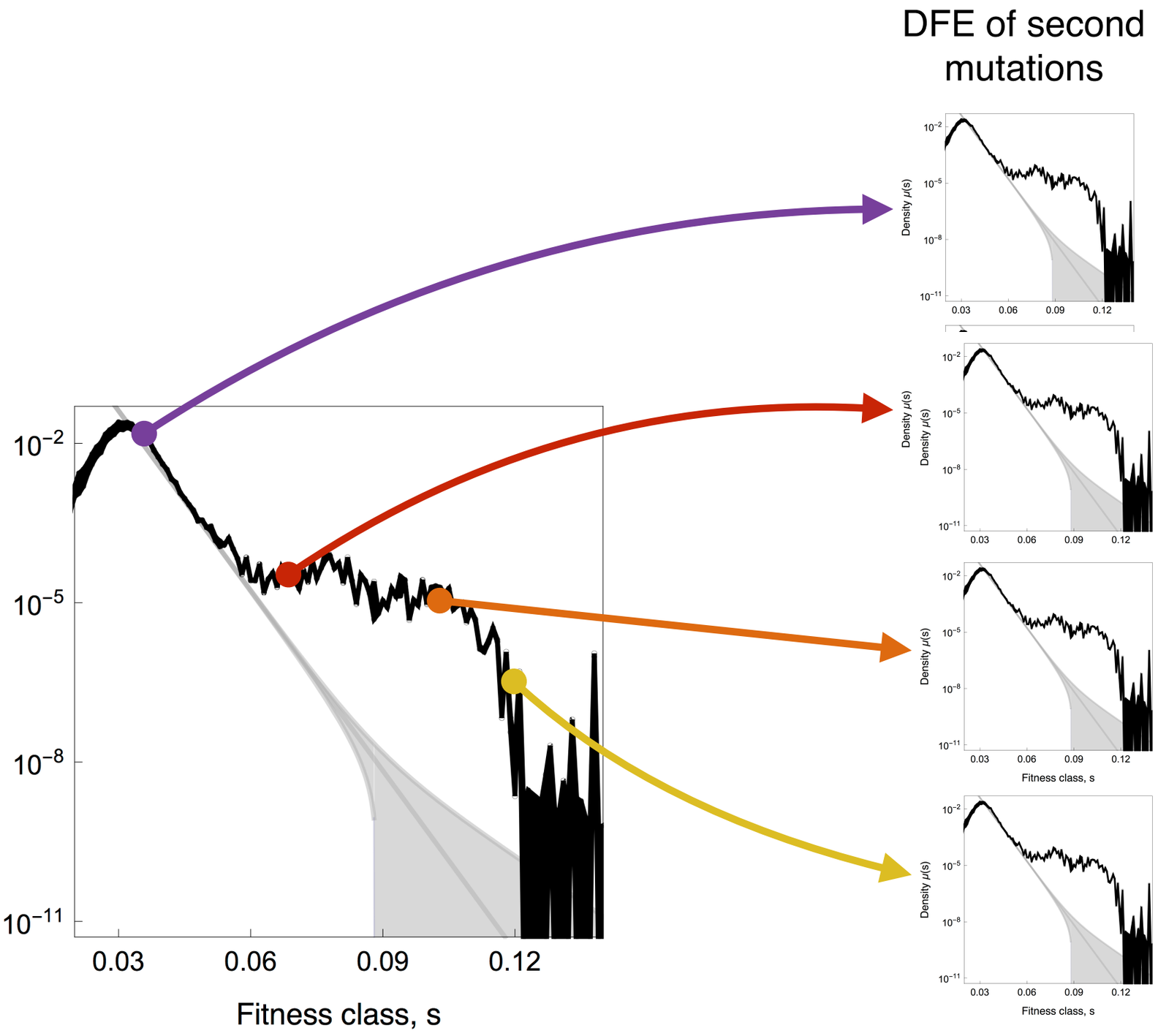
4. Later evolution is less predictable
    - The effective beneficial mutation rate decreases
    - Stochastic occurrence of multiple mutants

5. Diversity crashes will occur either because of a single-mutant selective sweep or because of anomalously early multiple-mutants

6. Adaptive cohorts are a natural feature of the mutation accumulation process

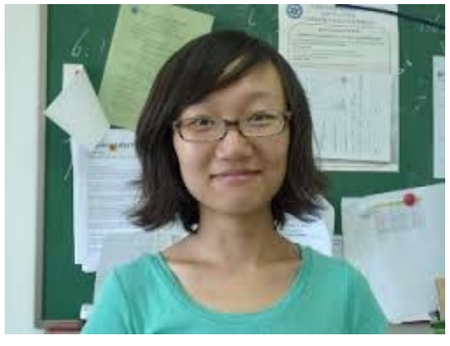
7. High resolution measurement of the DFE in combination with statistical modeling provides a means by which to forecast evolutionary outcomes

# How does the DFE of second mutations depend on the first mutation?



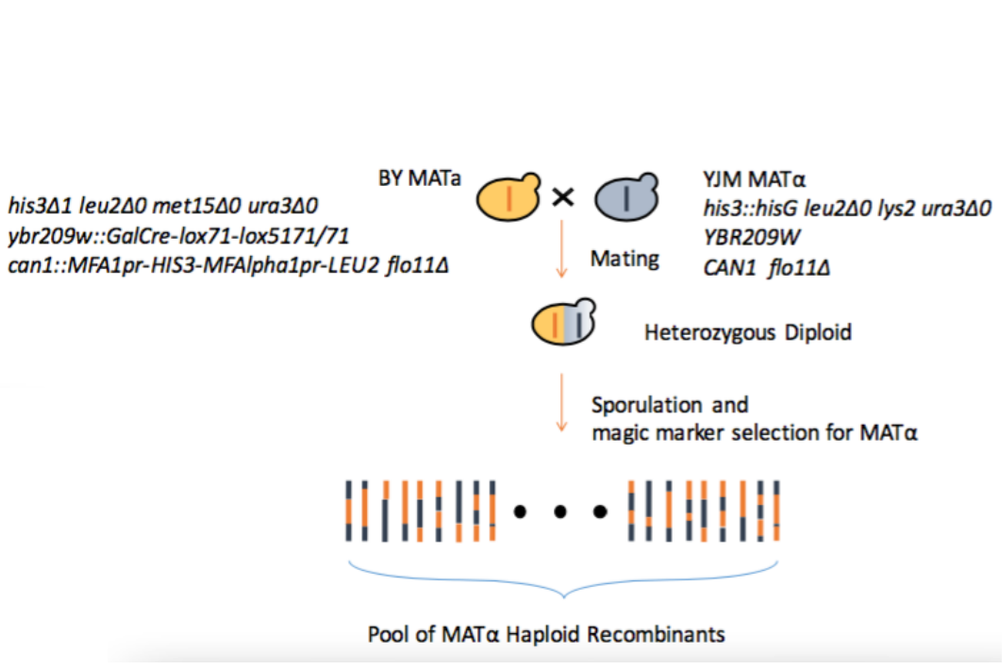

DFE of second mutations

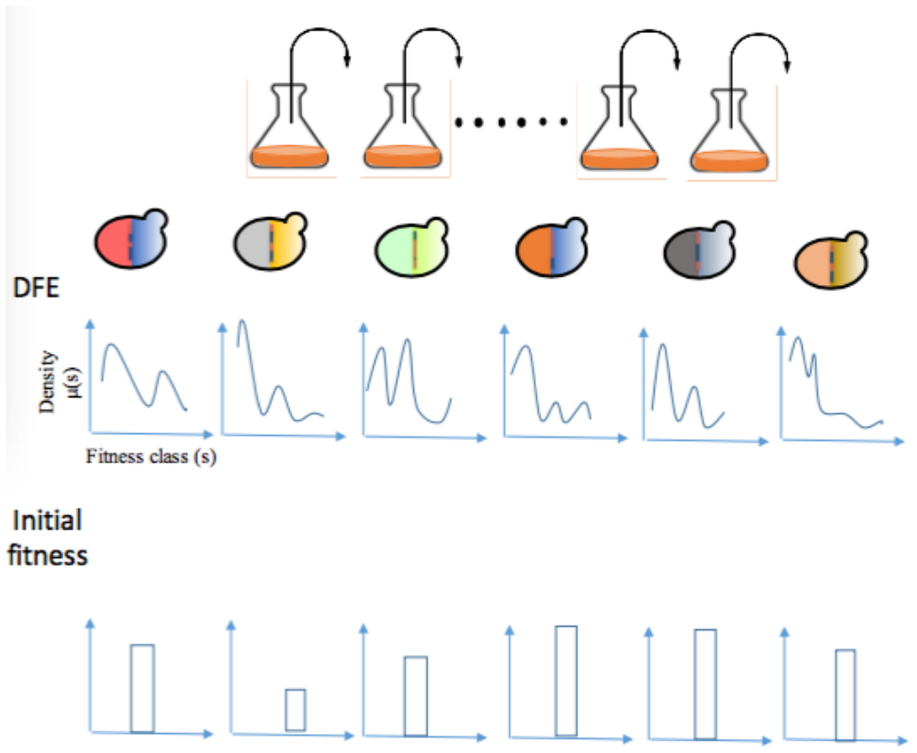


Fangfei Li

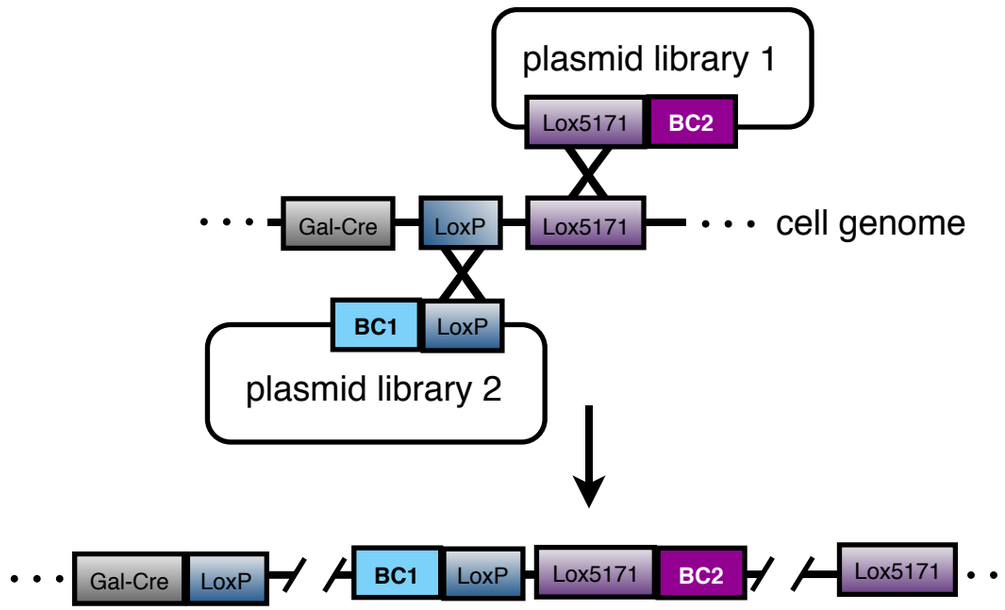
# How does the DFE of mutations change across genotype space?




Xianan Liu

# Multiple barcodes



# Thanks to...

Jamie Blundell

Gavin Sherlock

Daniel Fisher

Dmitri Petrov

Sandeep Venkataram

Katja Schwartz

## Levy Lab @ SBU

- Jamie Blundell (Cambridge)
- Fangfei Li
- Xianan Liu
- Zhimin Liu
- Danielle Francois
- Adam Dziulko

## Other Collaborators

- Ulrich Schlecht (Roche)
- Bob St. Onge (Stanford)
- Mia Jaffe (Stanford)
- Justin Smith (Stanford)
- Barbara Dunn (Stanford)
- Michael Desai (Harvard)
- Ian Ehrenreich (USC)
- Evangelos Coutsais (SBU)
- Frank Rosenzweig (G Tech)
- Dan Weinreich (Brown)
- David Gresham (NYU)

**Postdoc positions available!!**

LAUFER CENTER FOR PHYSICAL AND QUANTITATIVE BIOLOGY

Stony Brook University

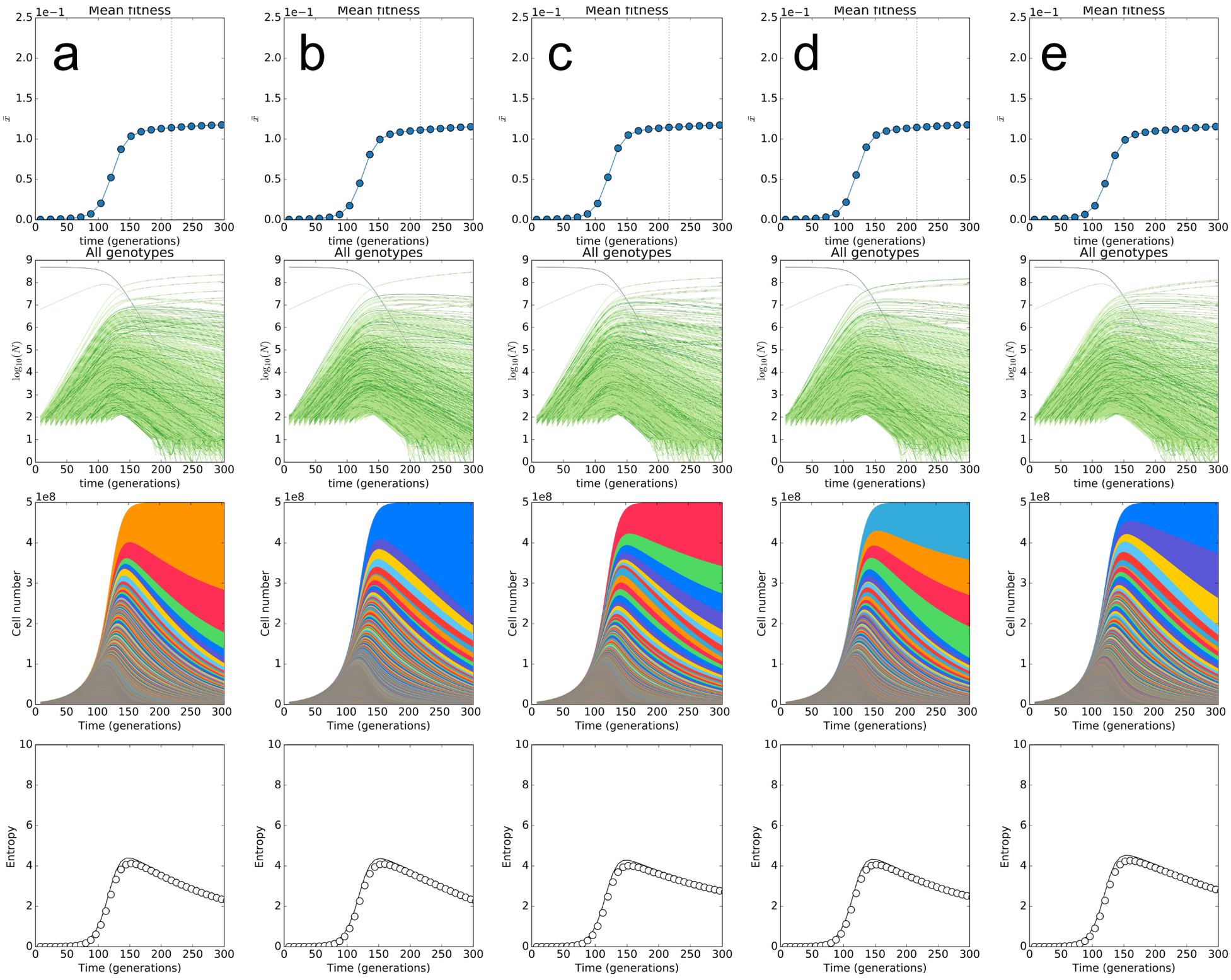
National Human Genome Research Institute

Carol M. Baldwin BREAST CANCER RESEARCH FUND, INC.

NEW YORK STATE Department of Health

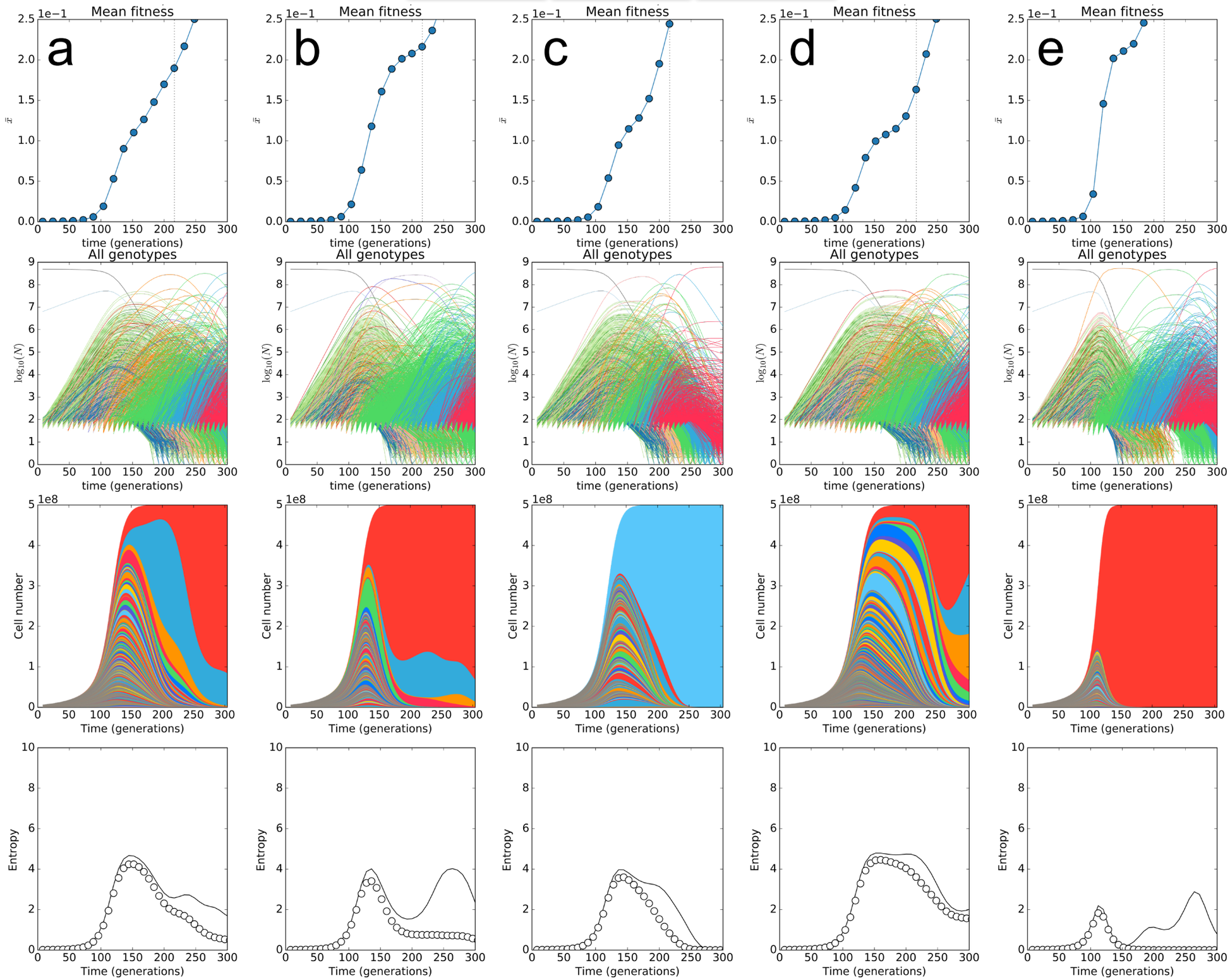
SBU BNL CSHL FIMR Long Island Bioscience Hub

Dip | LoF | GoF | Dip+Dip | Dip +LoF | Dip +GoF | Lof+Dip | LoF +LoF | LoF +GoF | GoF +Dip | GoF +LoF | GoF +GoF

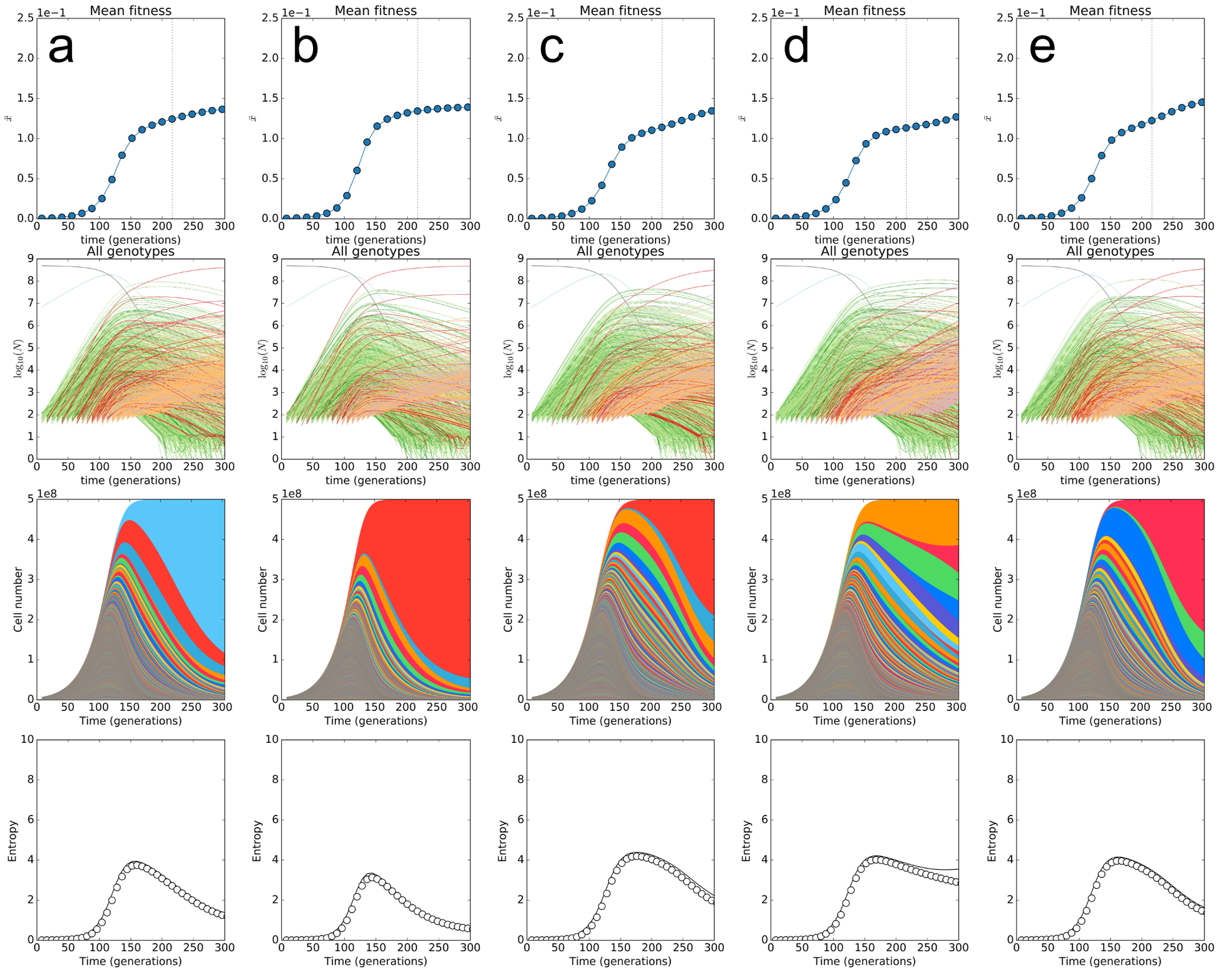


Dip | LoF | GoF | Dip+Dip | Dip +LoF | Dip +GoF | Lof+Dip | LoF +LoF | LoF +GoF | GoF +Dip | GoF +LoF | GoF +GoF

triple- | quadruple- | quintuple-



Dip | LoF | GoF | Dip+Dip | Dip +LoF | Dip +GoF | Lof+Dip | LoF +LoF | LoF +GoF | GoF +Dip | GoF +LoF | GoF +GoF



# Evolution in drug