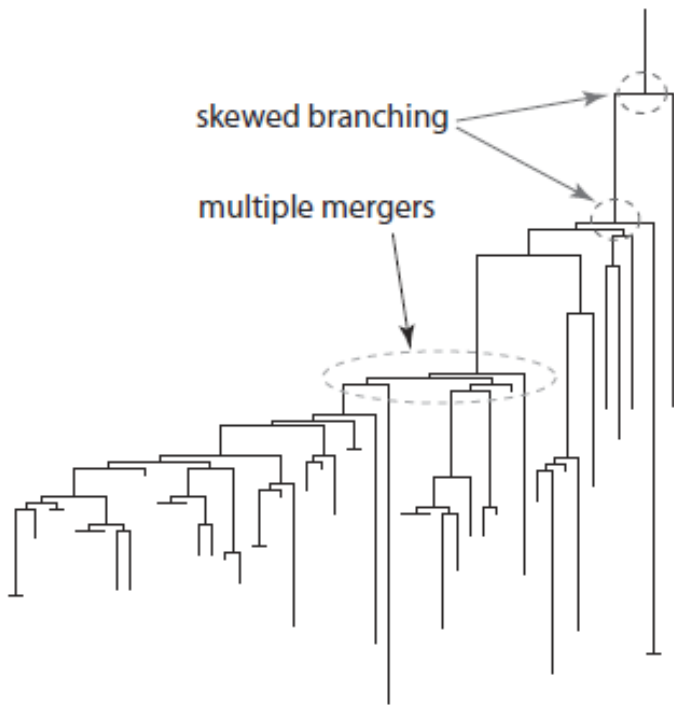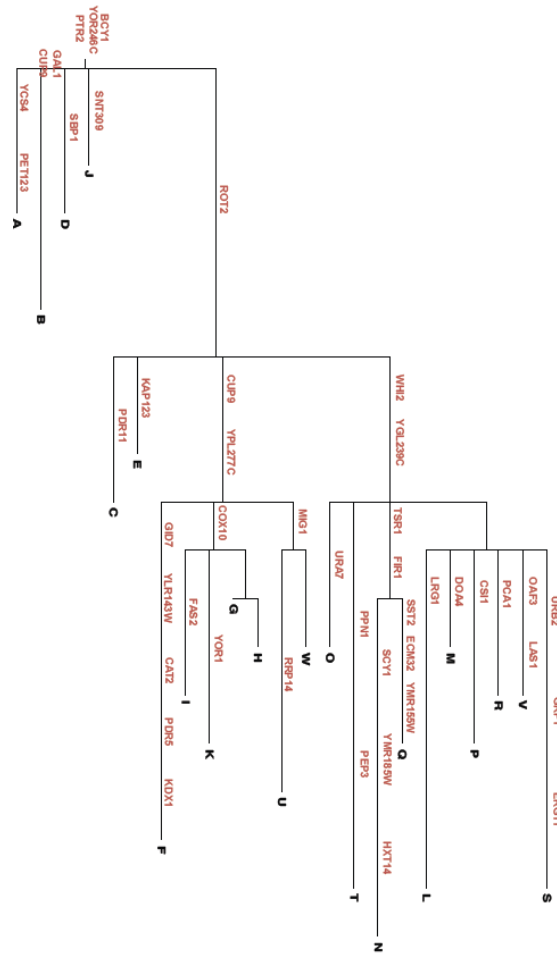# Genetic Diversity in the Interference Selection Limit

[Influenza, Neher 2013]

[Experimental Yeast Populations]
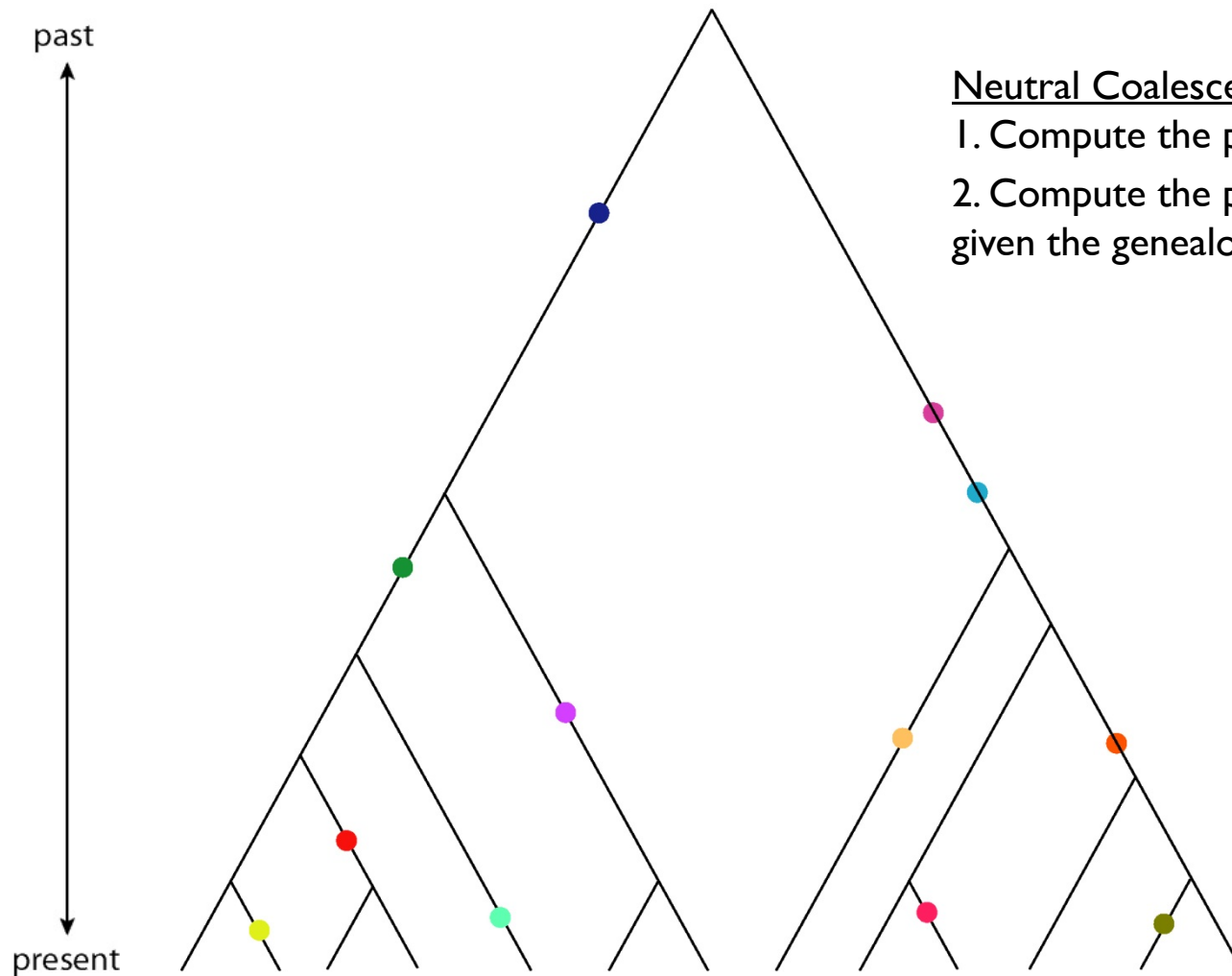
Ben Good

How does pervasive natural selection alter patterns of genetic diversity?

Michael Desai, Harvard University

Standard methods describe neutral evolution
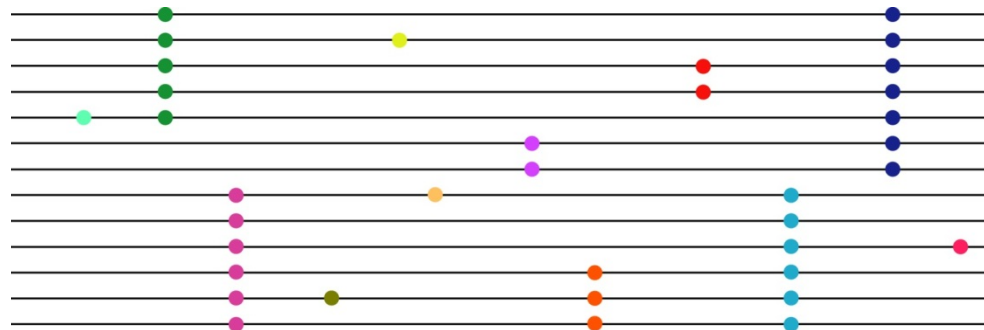

past

present
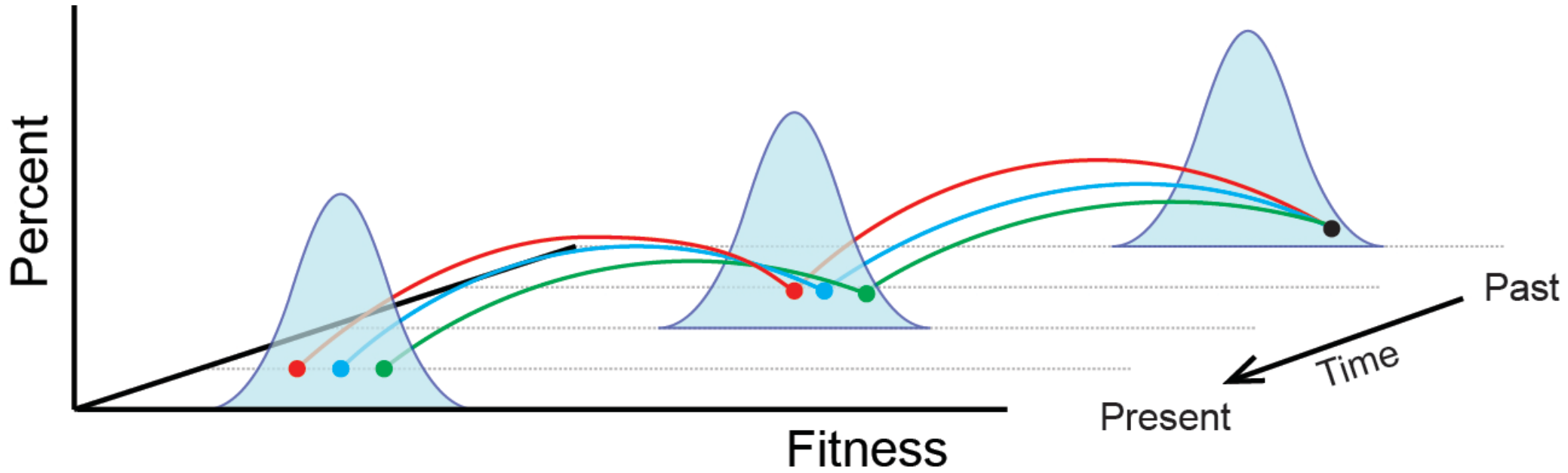
Neutral Coalescent Theory:
1. Compute the probability of a genealogy
2. Compute the probability of observed diversity given the genealogy

Key Predictions:

1. Diversity $\pi \propto N$

2. Frequency spectrum $f(i) \propto 1$

Selection and the Shape of Genealogies

Trace individual lineages through the fitness distribution:
   -Present individuals are descended from the fittest ancestors.

How does purifying selection shape diversity?

A simple model:

Population size:	N
Mutation rate:	U
Fitness effects:	$\rho(s)$
Recombination rate:	R

An even simpler model:

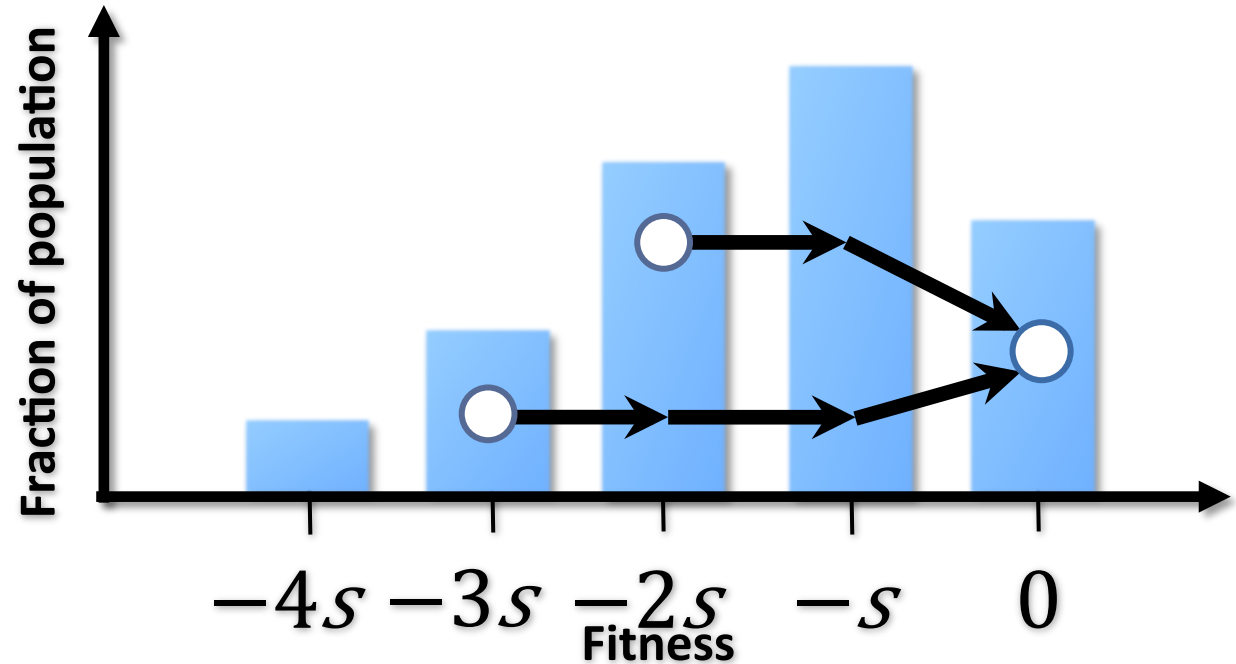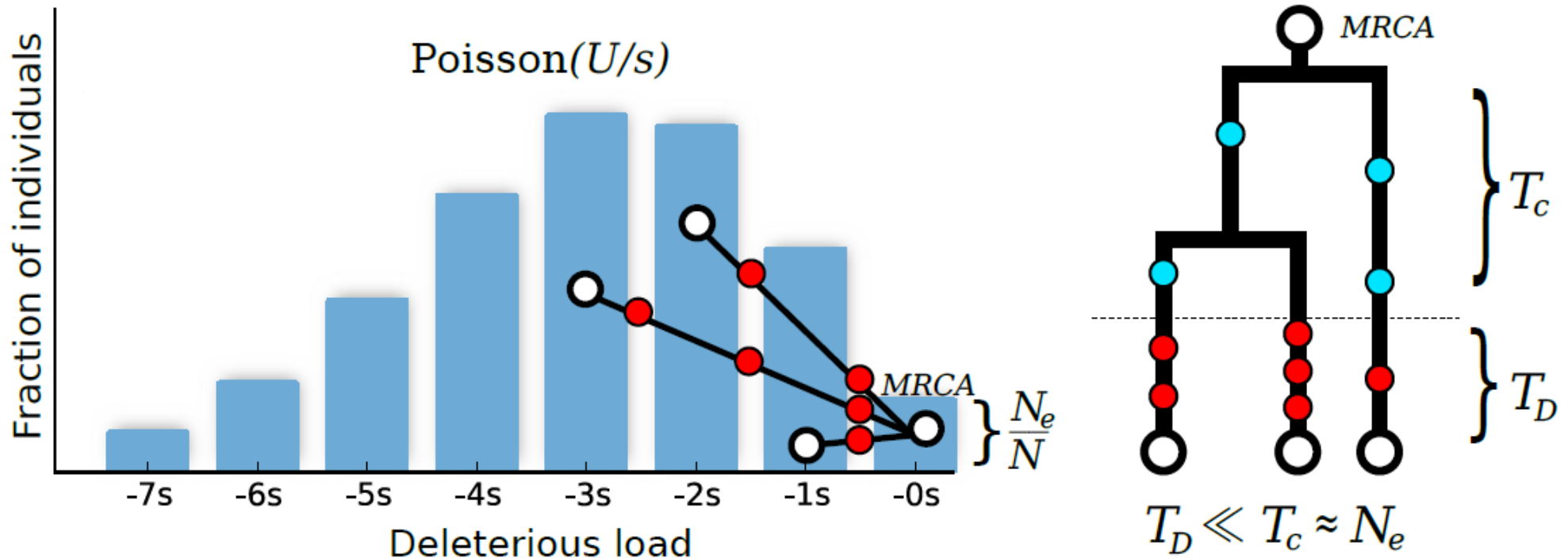|	
Population size:	N
Neutral mutation rate:	U_n
Deleterious mutation rate:	U_d
Fitness effect:	s

Structured Coalescent:
Steady state distribution of fitness within the population.
"Migrate" between fitness classes by mutations.
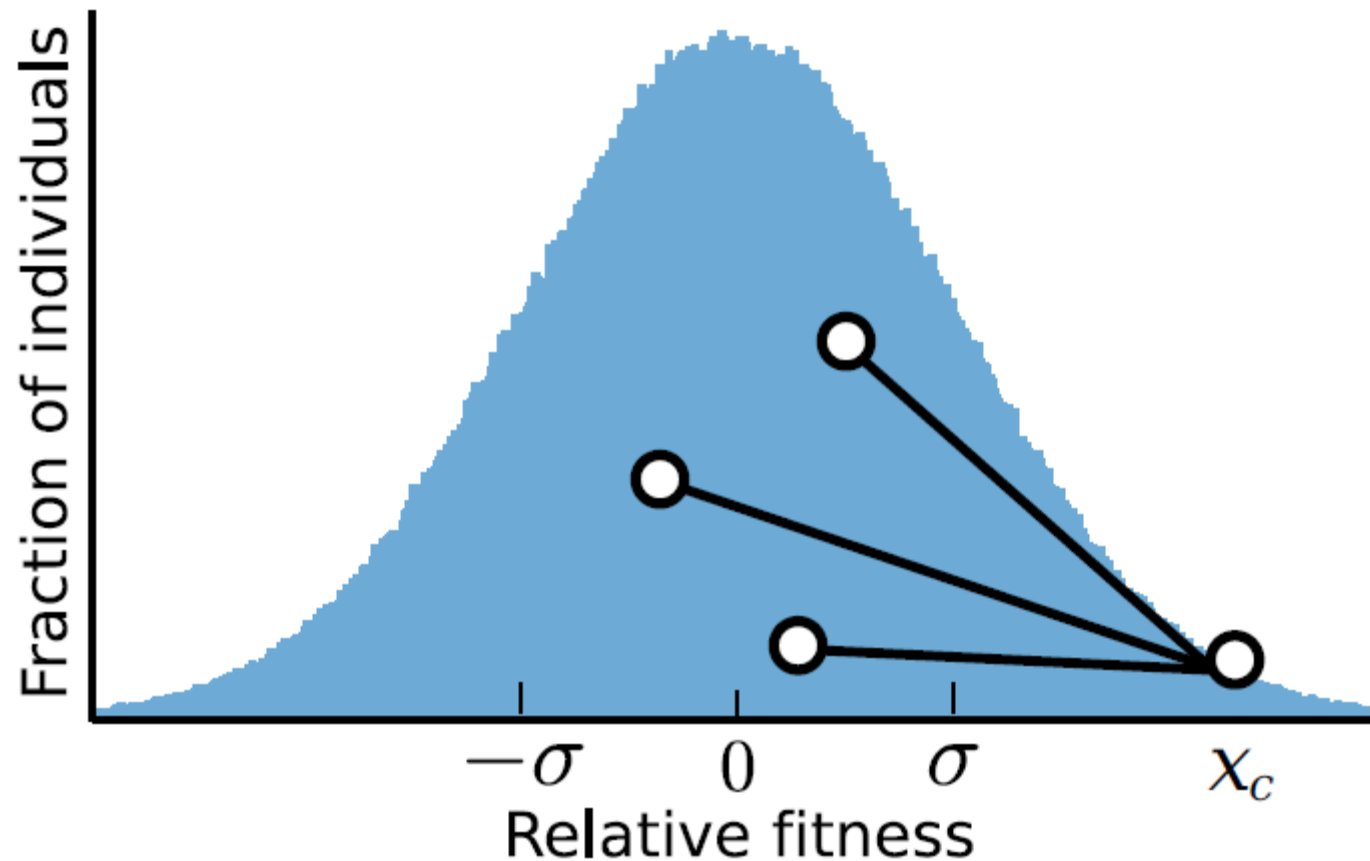Exchangeability *within* each fitness class.

Strong purifying selection: "Background Selection"

Strong purifying selection reduces effective population size.
Exact in the limit $Ns \rightarrow \infty$ while holding NU/Ns constant.

Corrections for large but finite Ns from the *Structured Coalescent*
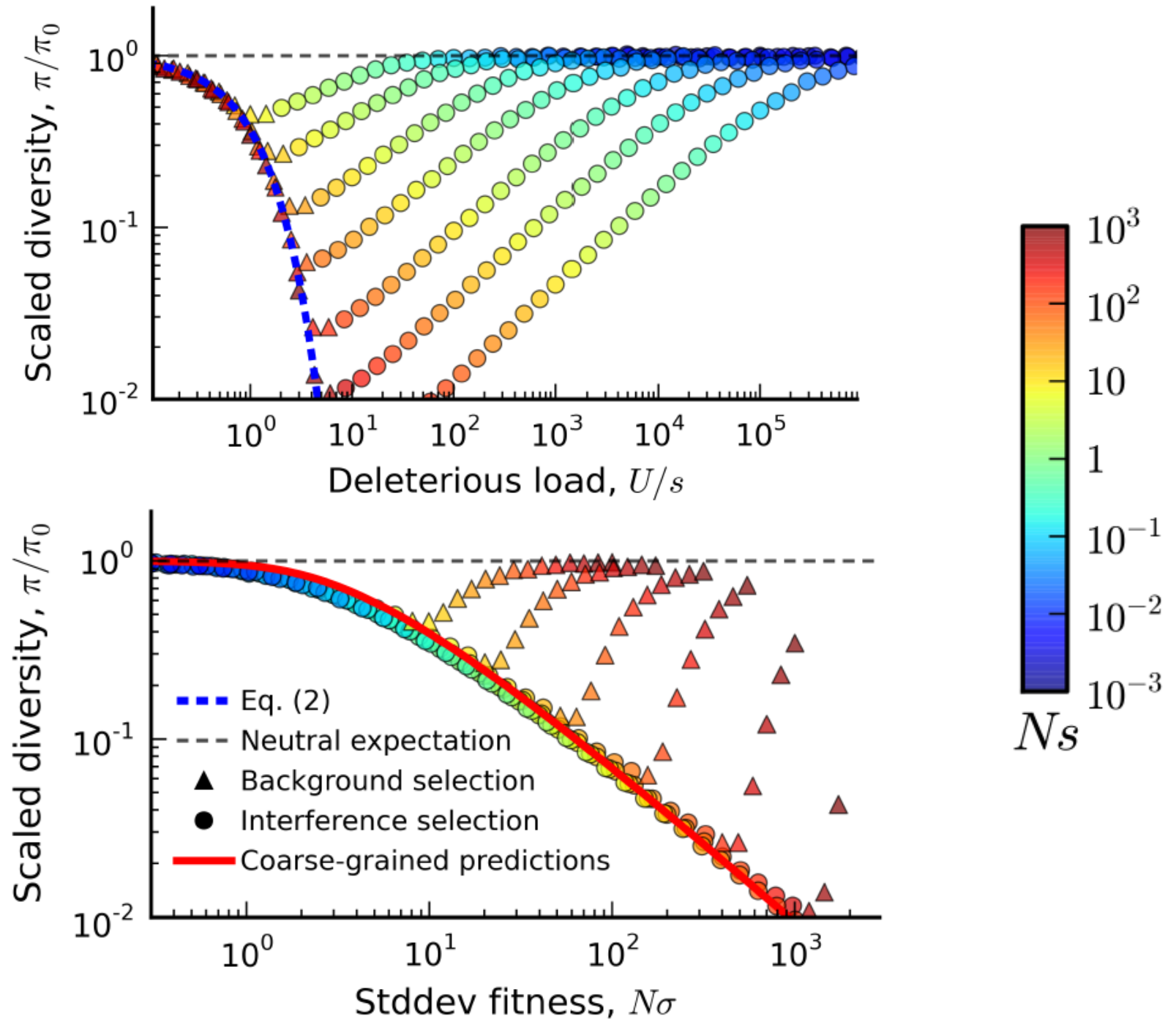
What about weak or pervasive selection?

When $Nse^{-U/s} \sim 1$ or less, the distribution fluctuates too much underneath, so the structured coalescent does not make sense.
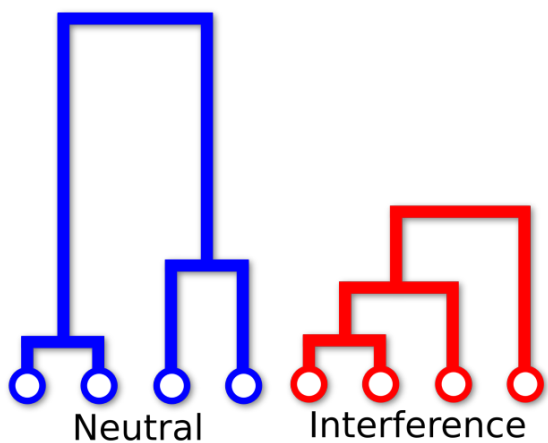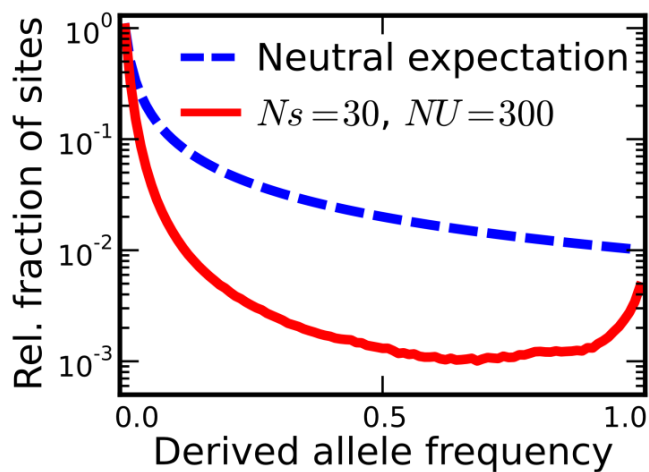
[Good, Walczak, Neher, Desai *PLoS Genetics* 2014]
[Good and Desai *TPB* 2013]

BGS/Structured Coalescent Break Down for Weak Selection
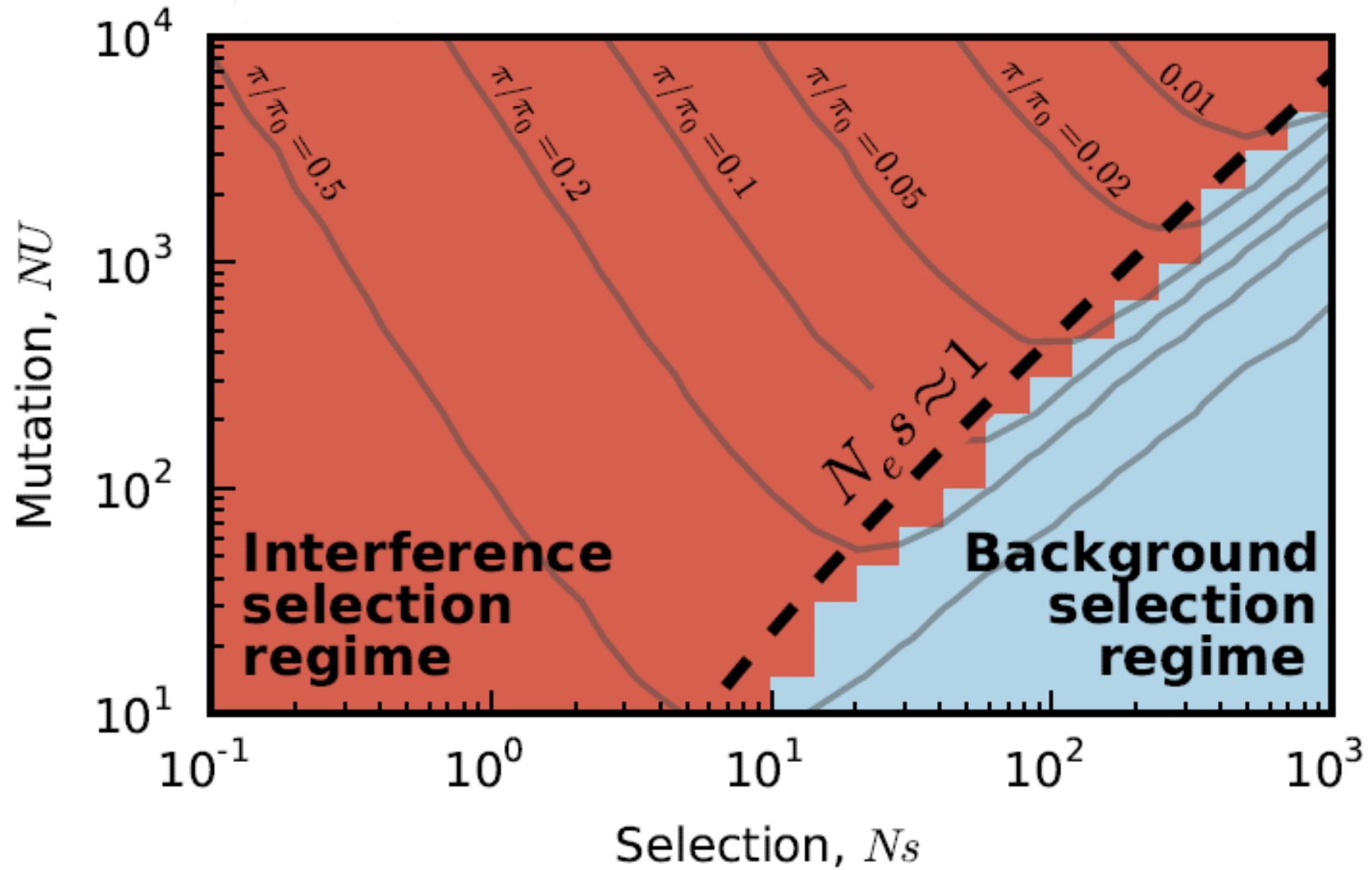
Collapse with U/s (BGS) or with Nσ (IS)

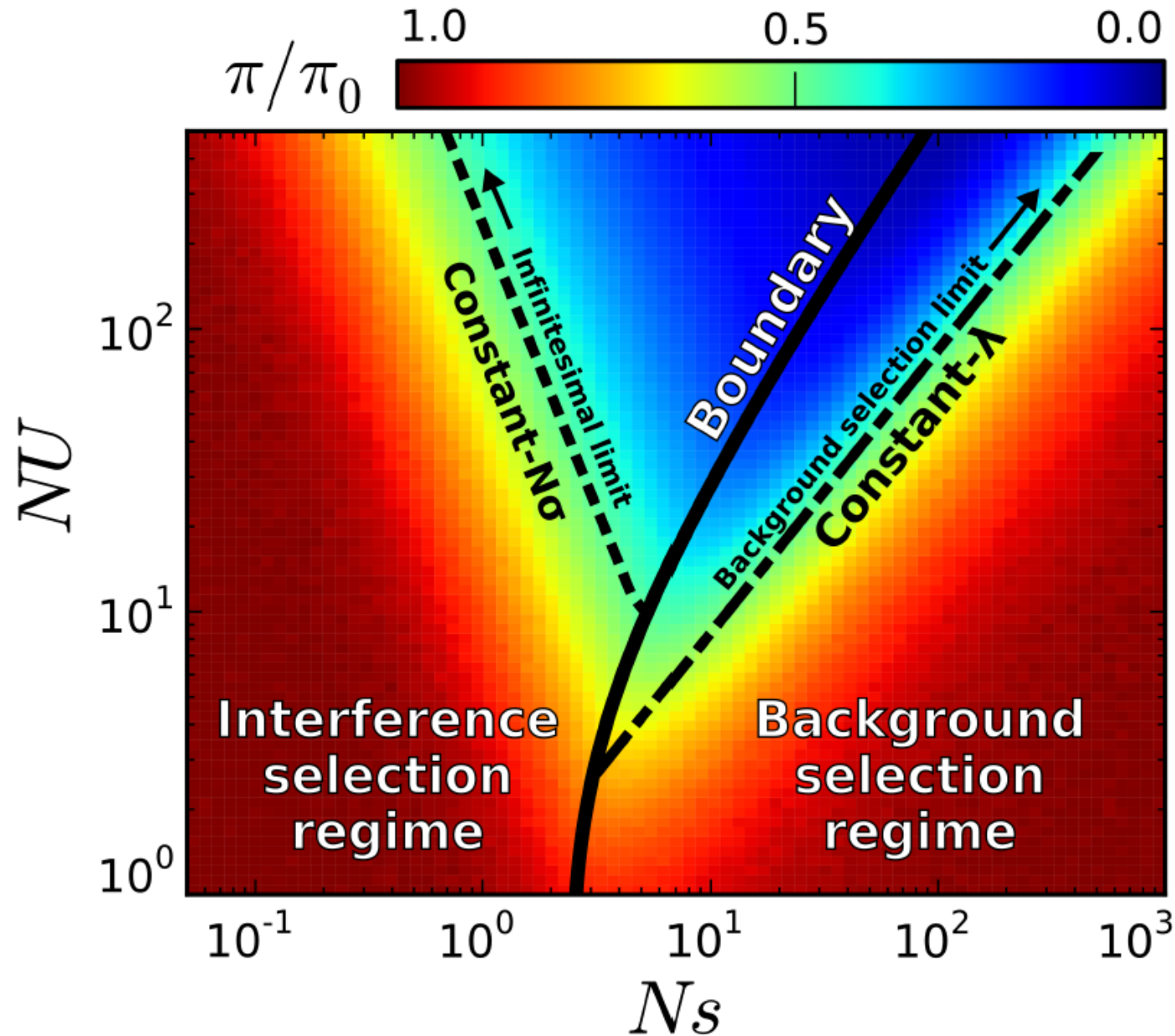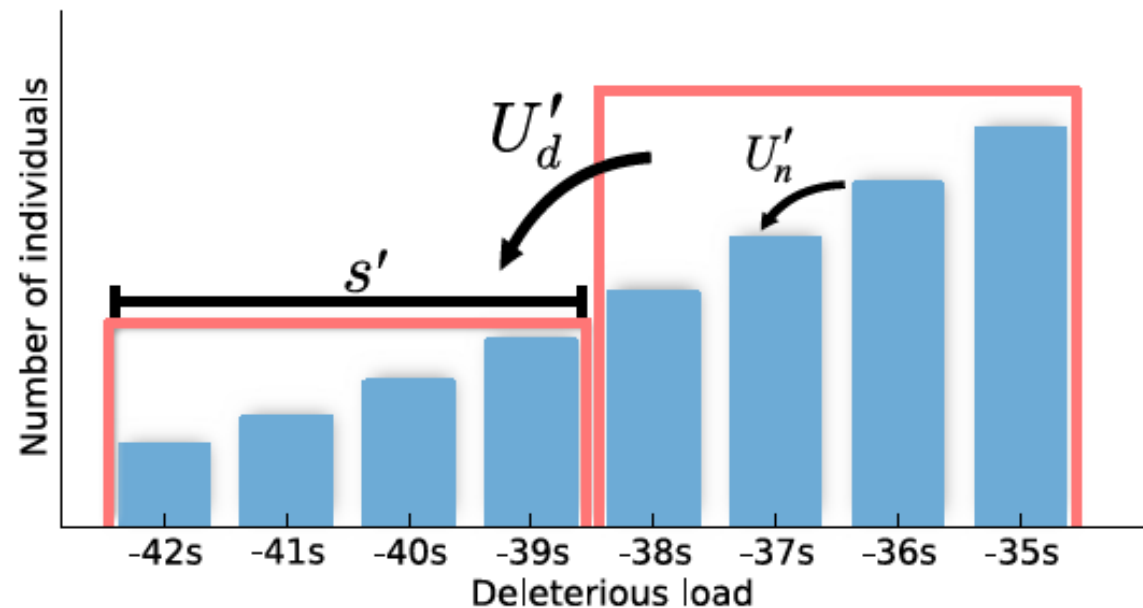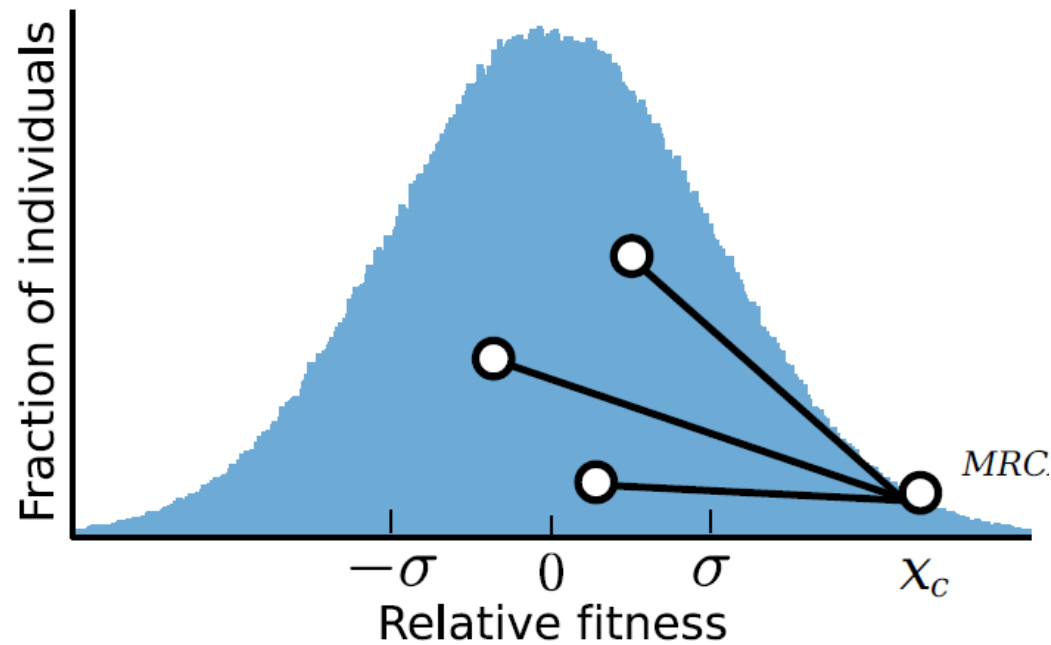# Interference Selection collapse holds generally

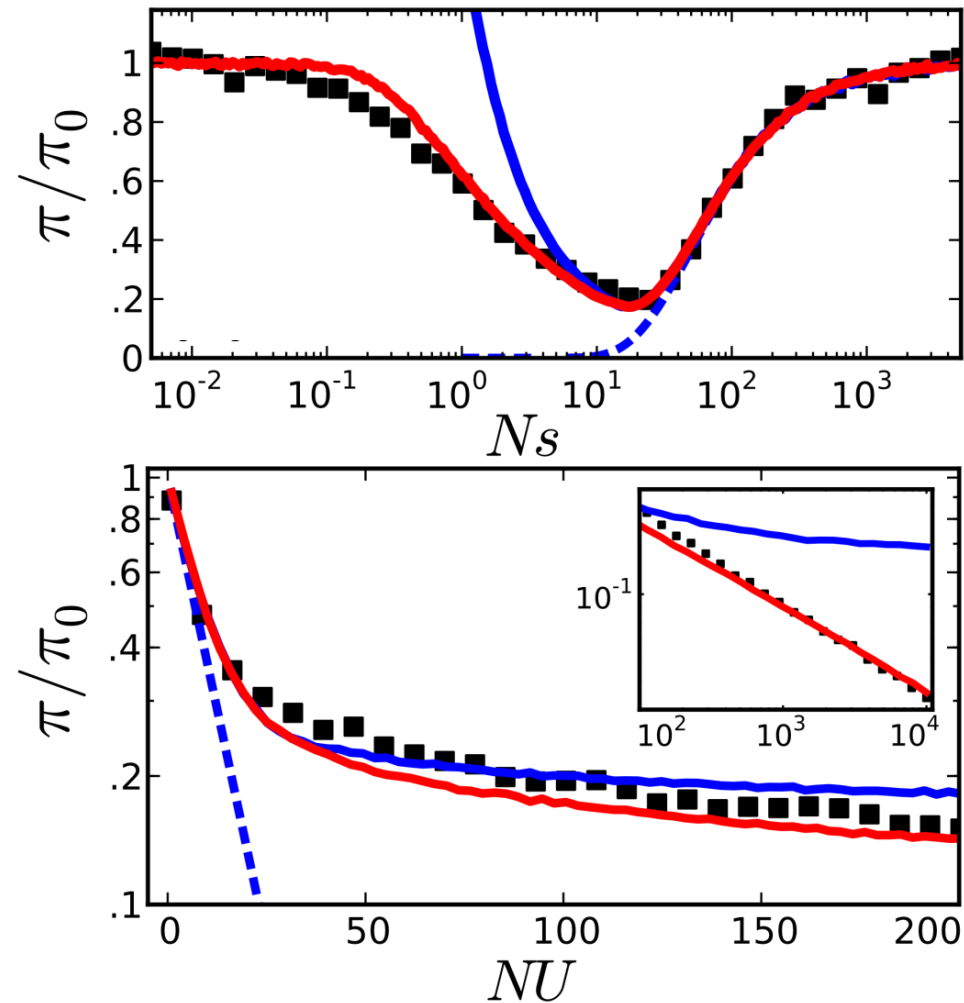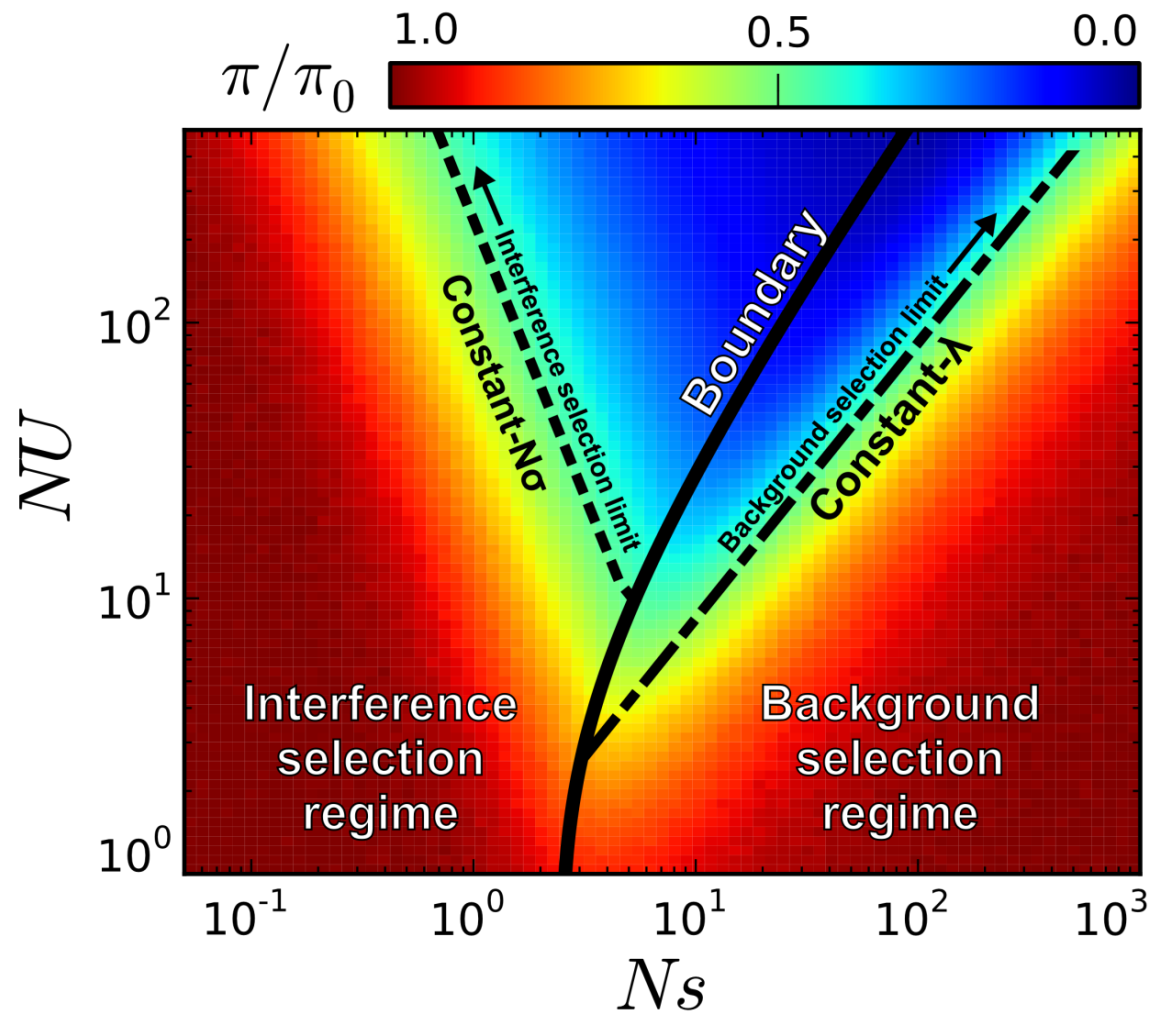# Two limits: background selection and interference selection

Background selection: $Ns \rightarrow \infty$ while holding NU/Ns constant
Inteference selection: $Ns \rightarrow 0$ while holding $N\sigma$ constant

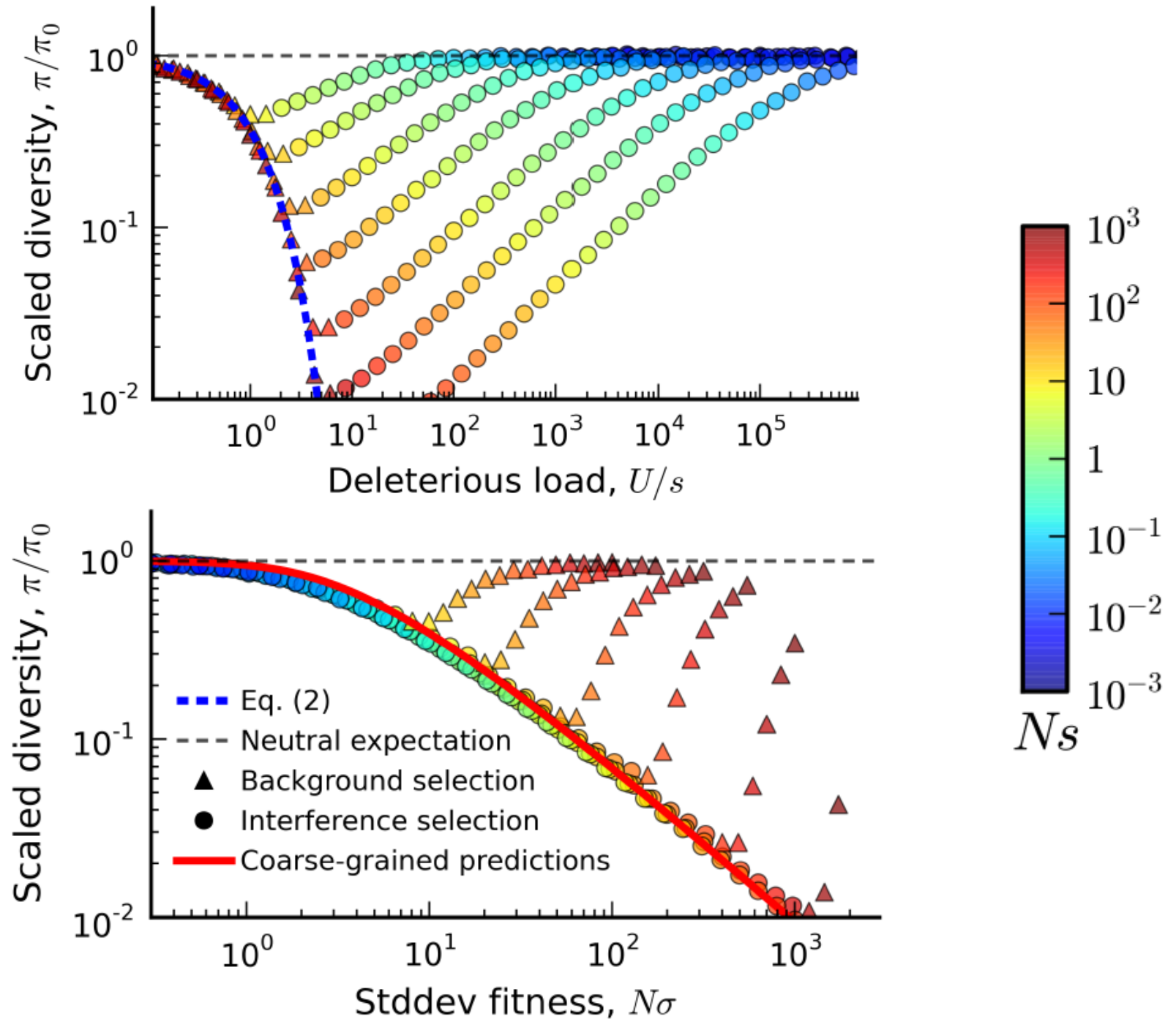# Intuition: "coarse-graining" the fitness distribution
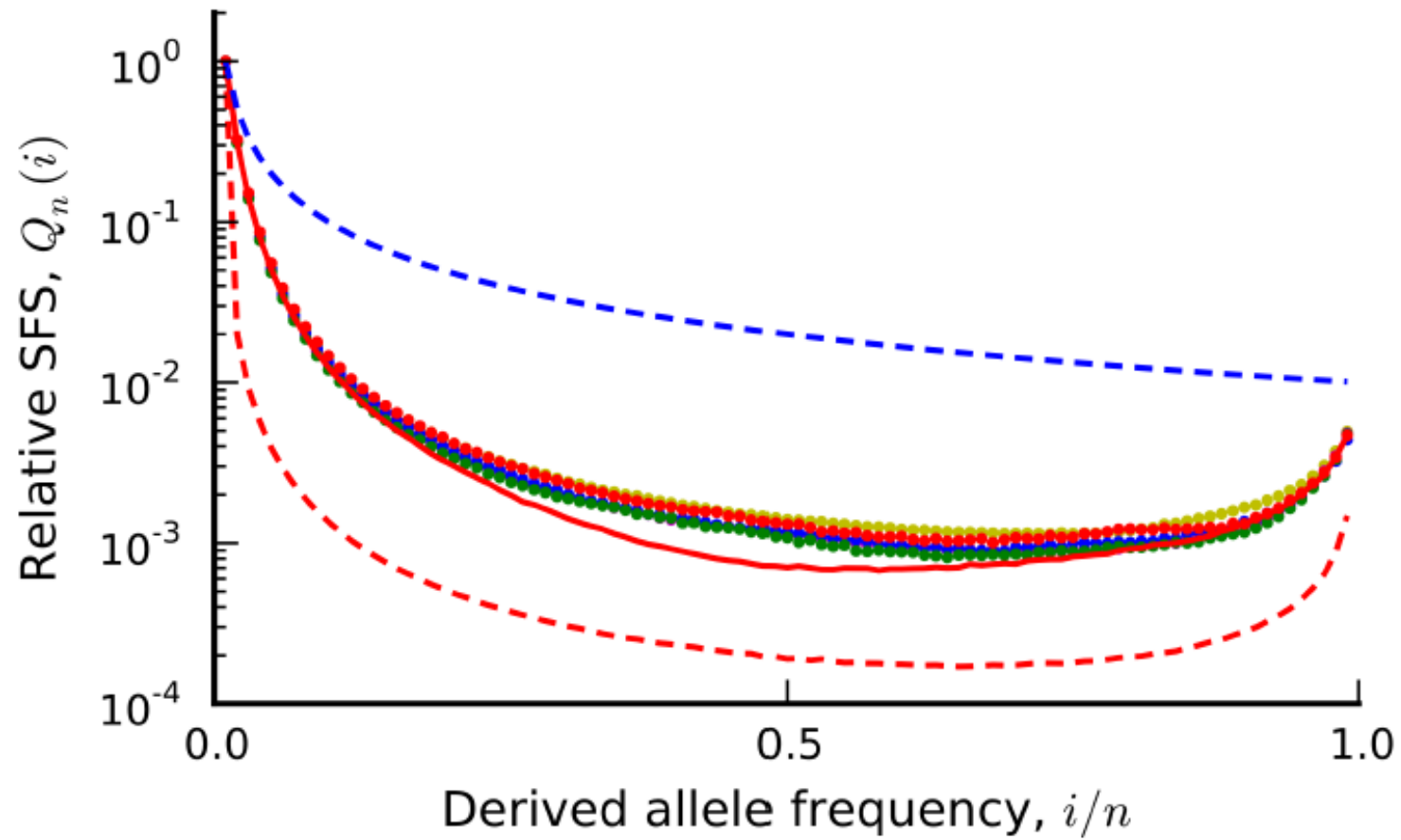
This allows us to predict diversity

There is a fundamental problem of identifiability:
Many different parameter values lead to *identical* patterns of diversity.
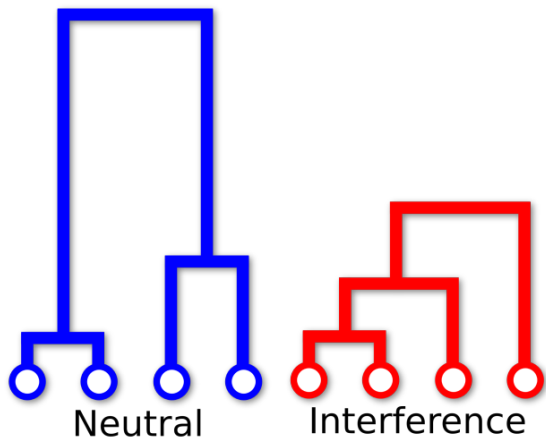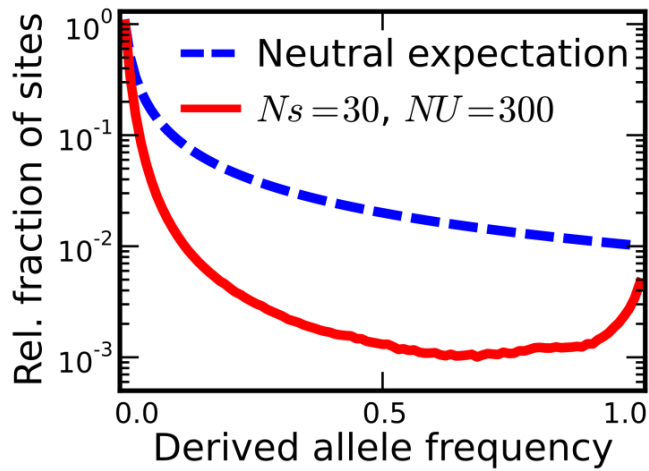
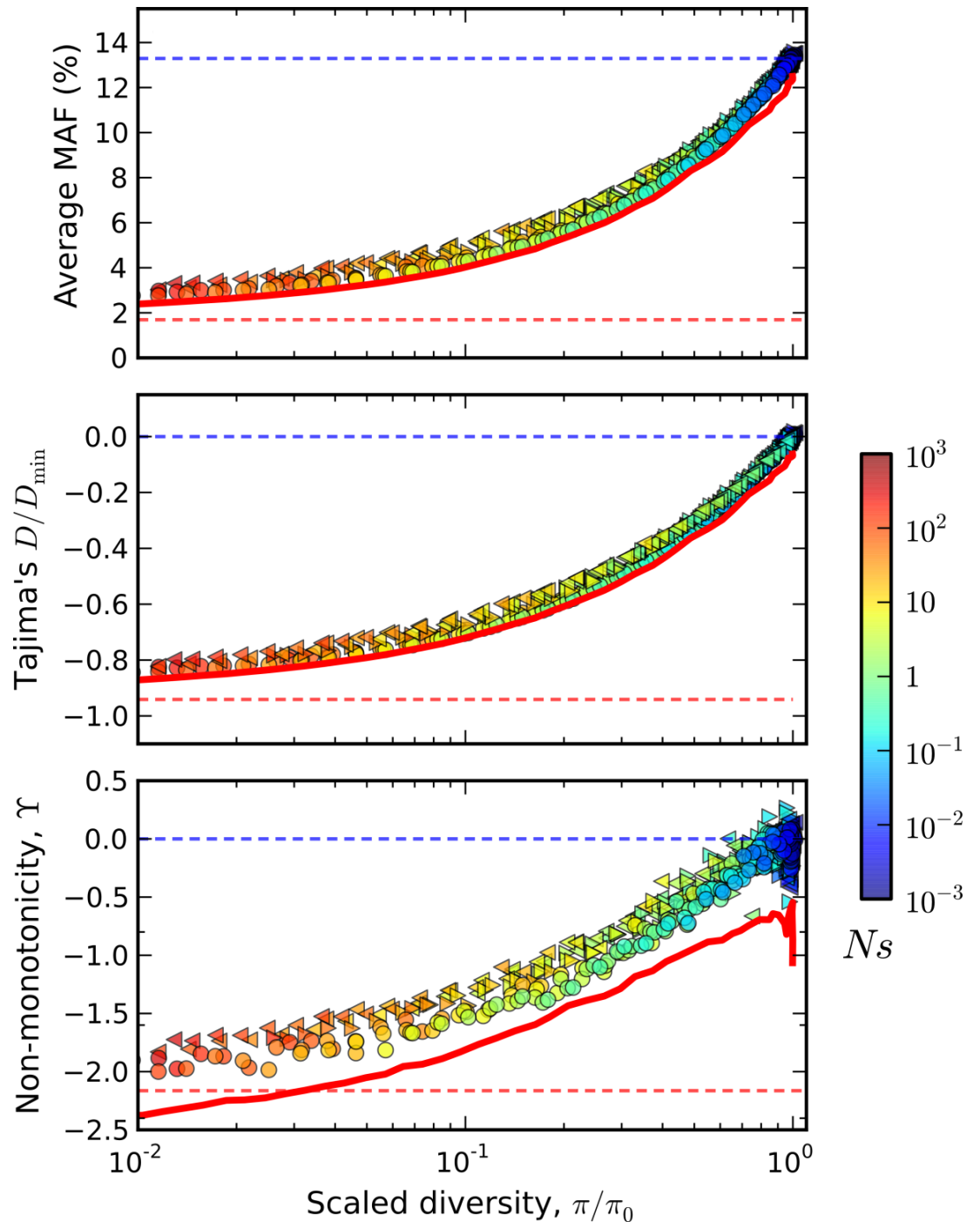# Coarse-Grained Predictions

Coarse-Grained Predictions

Legend:
- —— Coarse-grained predictions
- - - - Large $N\sigma$ limit
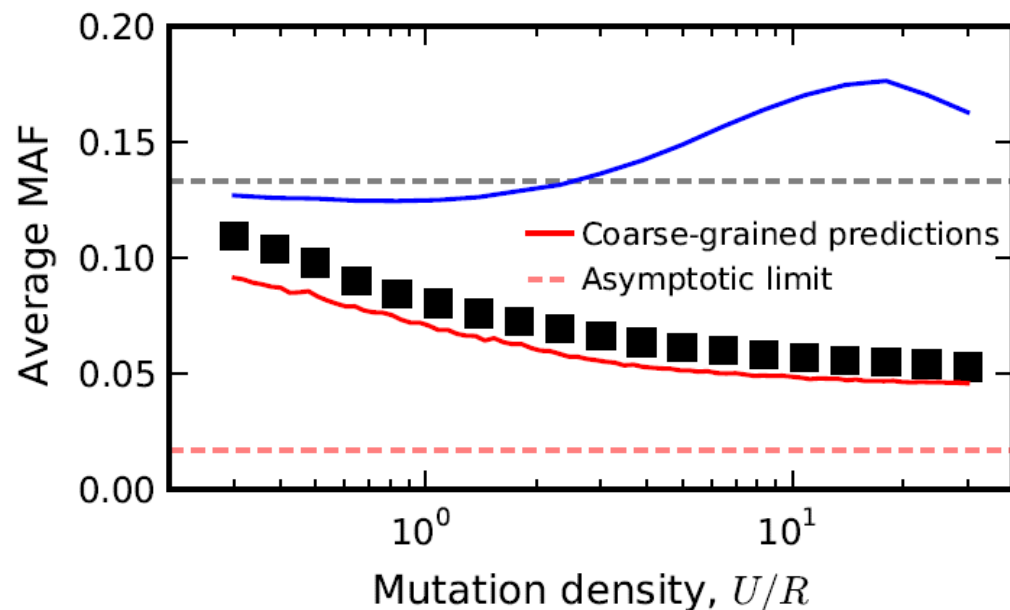- - - - Neutral expectation
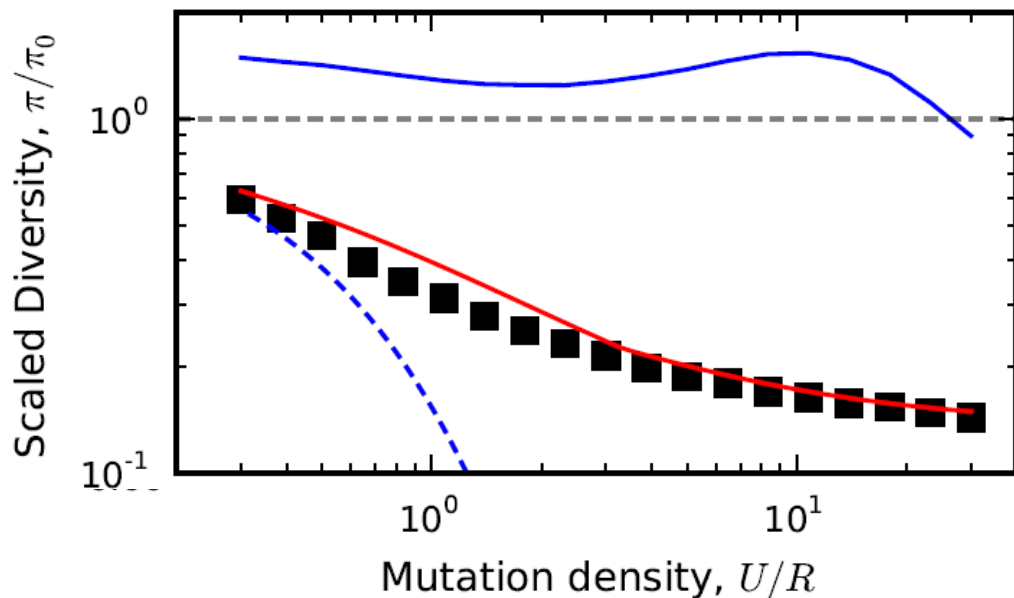
Coarse-Grained Predictions

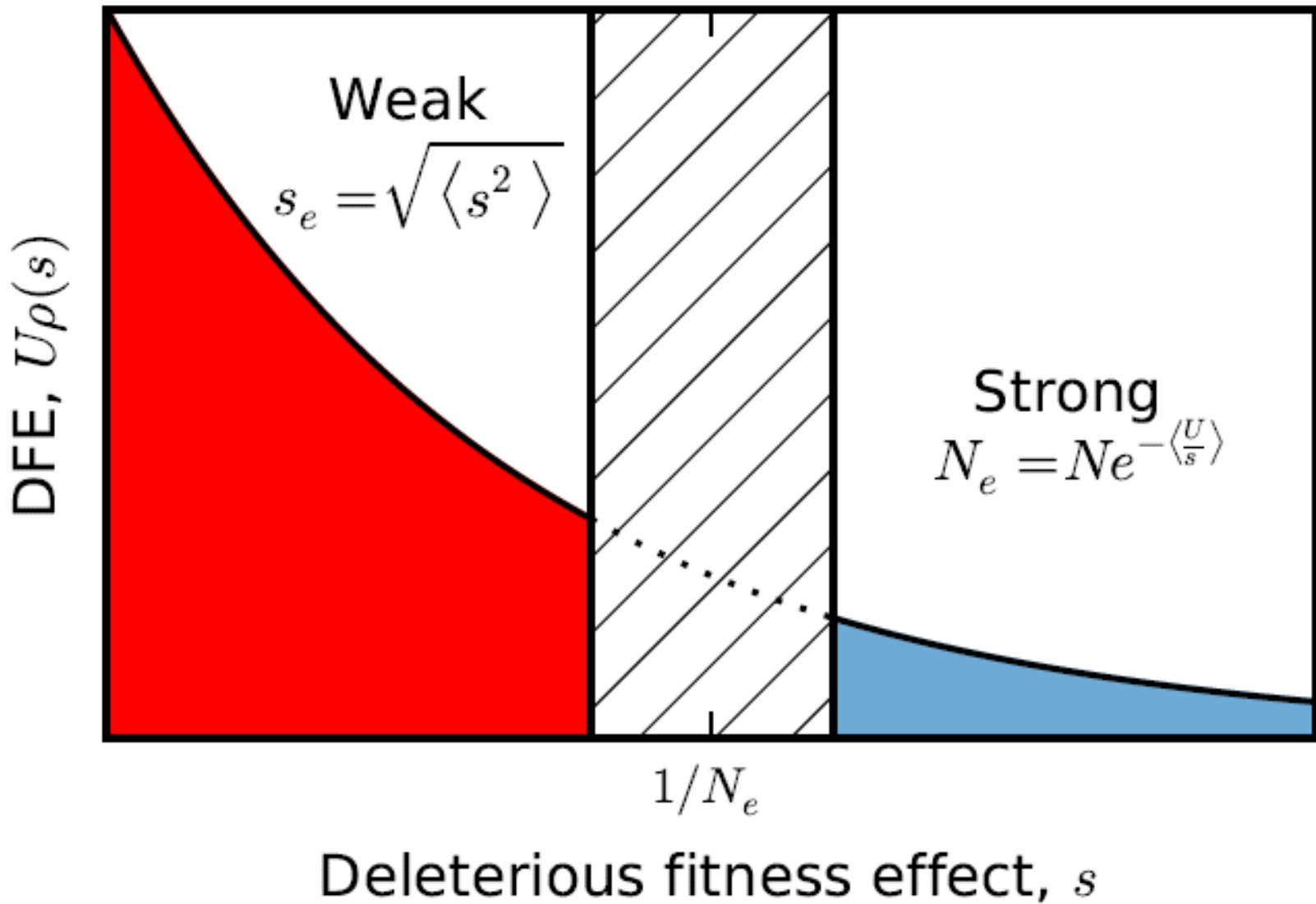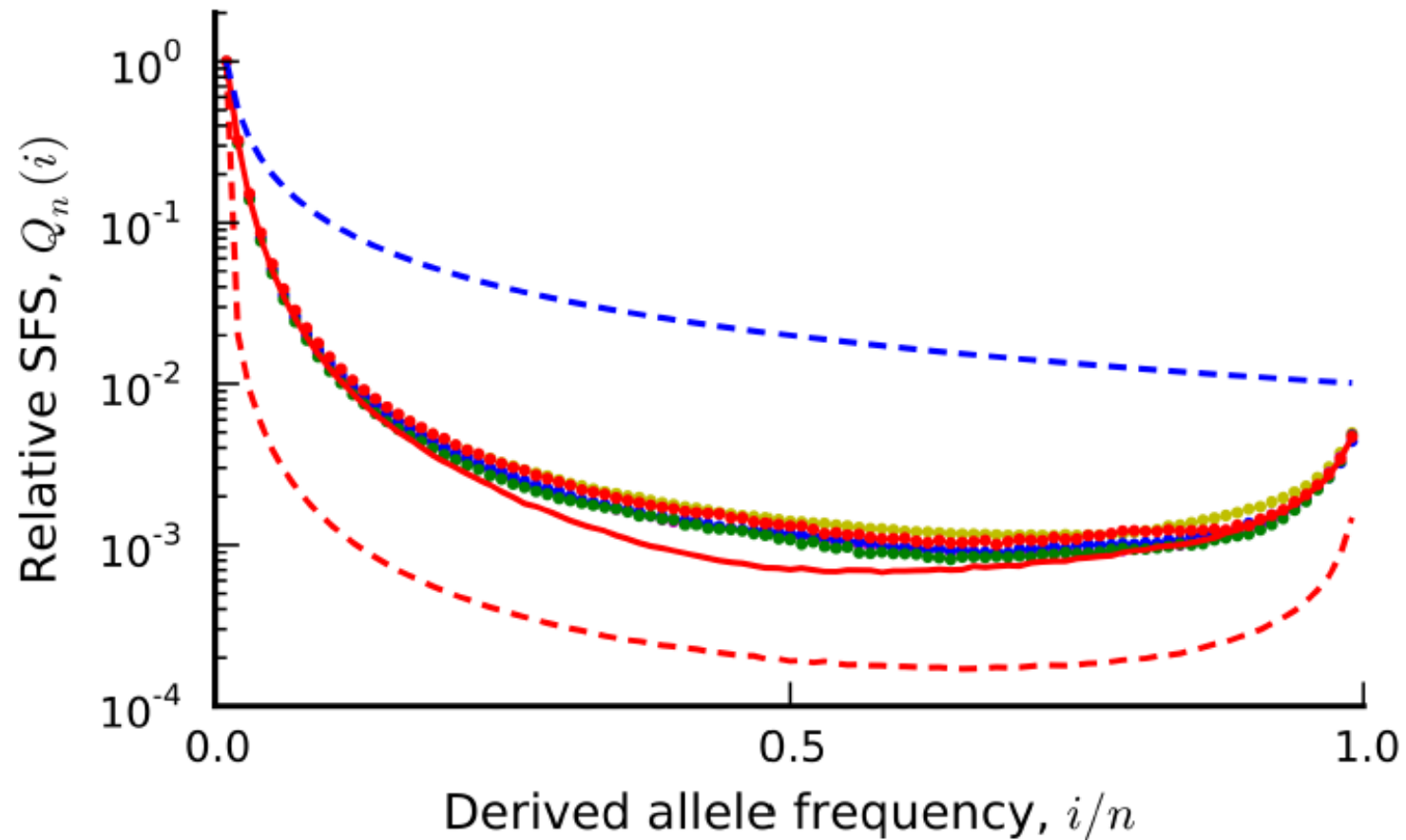# A Linkage-Block Approximation for Recombining Genomes
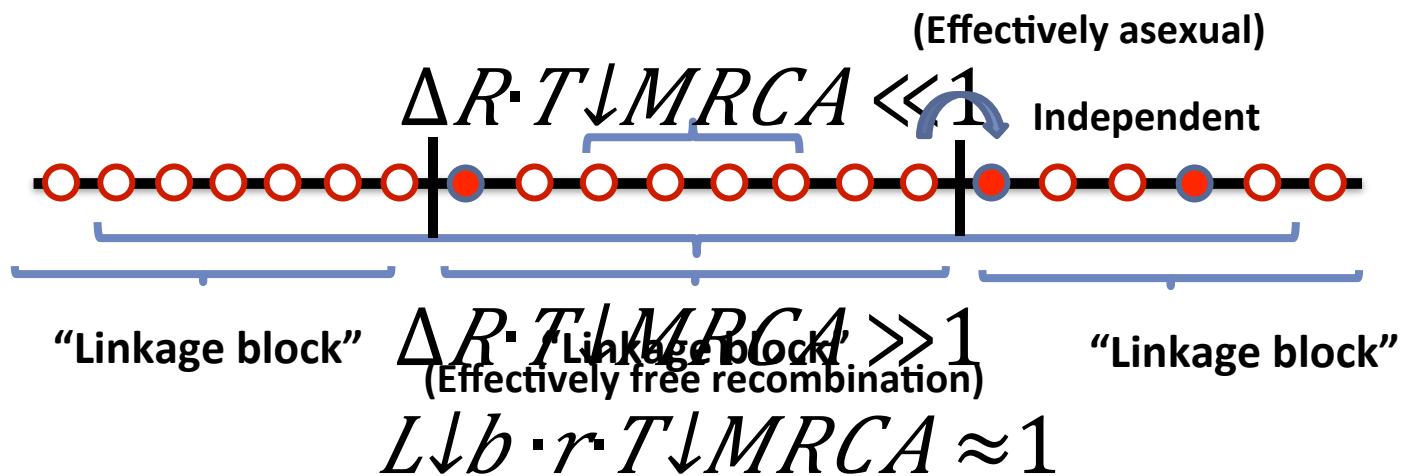
Distributions of Fitness Effects

Interference Selection Still Applies

Legend:
- Sexual ($NR=10$), $Ns=30$, $NU=354$
- TruncatedExp($s_{max}/\bar{s}=3$), $N\bar{s}=10$, $NU=2230$
- Uniform($0, s_{max}$), $Ns_{max}=28.5$, $NU=1000$
- Finite sites ($L=10^5$), $Ns=21.4$, $NU=600$
- Single-s, $Ns=30$, $NU=300$
- Coarse-grained predictions
- Large $N\sigma$ limit
- Neutral expectation

Axes:
- Relative SFS, $Q_n(i)$
- Derived allele frequency, i/n

Acknowledgements

Graduate Students and Postdocs:
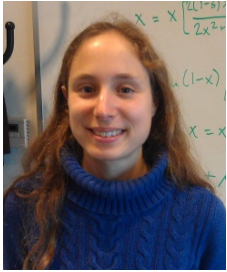
Elizabeth Jerison

Ben Good

Lauren Nicolaisen

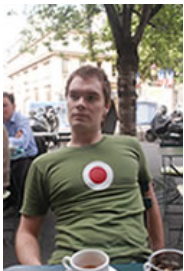Katya Kosheleva

Sebastian Akle

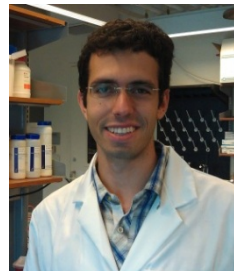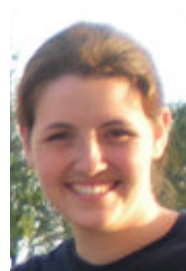Mike McDonald

Gabriel Perron

David van Dyken

Sergey Kryazhimskiy

Dan Rice

Evgeni Frenkel

Julia Piper

Collaborators:

Joshua Plotkin

Alex Walczak

Greg Lang

Richard Neher

Oskar Hallatschek

Undergraduate Students:

Henry Shull
Ian Ochs
Sadik Shahidain
Nisreen Shiban