Genetic Diversity in the Interference Selection Limit

How does pervasive natural selection alter patterns of genetic diversity?

Michael Desai, Harvard University

[Influenza, Neher 2013]

[Experimental Yeast Populations]

Ben Good
Standard methods describe neutral evolution

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 $f(i) \propto \frac{1}{i}$
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 reduces effective population size. Exact in the limit $Ns \to \infty$ while holding $NU/Ns$ constant.

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

When $Nse \downarrow -U/s \sim 1$ 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

- $(NR=0)$

- $\frac{\pi}{\tau_0} = 0.5$
- $\frac{\pi}{\tau_0} = 0.2$
- $\frac{\pi}{\tau_0} = 0.1$
- $\frac{\pi}{\tau_0} = 0.05$
- $\frac{\pi}{\tau_0} = 0.02$
- $0.01$

Mutations, $NU$

Selection, $Ns$

- Interference selection regime
- Background selection regime
Collapse with $U/s$ (BGS) or with $N\sigma$ (IS)
Interference Selection collapse holds generally

A

Rel. fraction of sites

$10^0$

$10^{-1}$

$10^{-2}$

$10^{-3}$

0.0

0.5

1.0

Derived allele frequency

Neutral expectation

$Ns = 30, NU = 300$

B

Legend:

- Asexual
- $NR = 10$
- $NR = 100$

C

Average MAF (%)

0

2

4

6

8

10

12

14

D

Tajima's $D / D_{min}$

-0.8

-0.6

-0.4

-0.2

0.0

Non-monotonicity, $\mathcal{I}$

0.0

0.5

1.0

1.5

2.0

Scaled diversity, $\pi / \pi_0$

$10^{-2}$

$10^{-1}$

$10^0$

E

Large $N_\sigma$ limit

Coarse-grained predictions

Neutral expectation
Two limits: background selection and interference selection

Background selection: \( Ns \rightarrow \infty \) while holding \( NU/Ns \) constant

Interference selection: \( Ns \rightarrow 0 \) while holding \( N\sigma \) constant
Intuition: “coarse-graining” the fitness distribution

![Graph showing fitness distribution with fractions and relative fitness values.]

- Fraction of individuals
- Relative fitness: $-\sigma$, $0$, $\sigma$, $\chi_c$

![Bar chart showing deleterious load distribution.]

- Number of individuals
- Deleterious load: $-42s$, $-41s$, $-40s$, $-39s$, $-38s$, $-37s$, $-36s$, $-35s$
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

![Graph showing relationship between scaled diversity and deleterious load](image)

- Scaled diversity, $\pi/\pi_0$
- Deleterious load, $U/s$
- Stddev fitness, $N\sigma$
- $N_s$ scale

Legend:
- Eq. (2)
- Neutral expectation
- Background selection
- Interference selection
- Coarse-grained predictions
Coarse-Grained Predictions

A

![Graph showing the relationship between derived allele frequency and relative fraction of sites, with lines for neutral expectation and Coarse-grained predictions labeled with parameters $N_s = 30$, $NU = 300$.]

B

![Tree diagrams for Neutral and Interference scenarios, with nodes labeled as Asexual, $NR = 10$, and $NR = 100$, and dashed lines for large $N_s$ limit.]

Legend:

- Asexual
- Neutral expectation
- $NR = 10$
- Coarse-grained predictions
- $NR = 100$
- Large $N_s$ limit

C

![Graph showing average MAF (%) with markers for different $N_s$ values, including $N_s = 30$, $NU = 300$, and $NR = 10$, $NR = 100$, with dashed lines for large $N_s$ limit.]

D

![Graph showing Tajima's $D/D_{\text{min}}$ with markers for different $N_s$ values, including $N_s = 30$, $NU = 300$, and $NR = 10$, $NR = 100$, with dashed lines for large $N_s$ limit.]

E

![Graph showing non-monotonicity, $\Upsilon$, with markers for different $N_s$ values, including $N_s = 30$, $NU = 300$, and $NR = 10$, $NR = 100$, with dashed lines for large $N_s$ limit.]

Scaled diversity, $\pi/\pi_0$
A Linkage-Block Approximation for Recombining Genomes

\[ \Delta R \cdot T \downarrow MRCA \ll 1 \quad \text{(Effectively asexual)} \]

\[ \Delta R \cdot T \downarrow MRCA \gg 1 \quad \text{Independent} \]

\[ L \downarrow b \cdot r \cdot T \downarrow MRCA \approx 1 \]

Scaled Diversity, \( \pi / \pi_0 \)

- Neutral expectation
- Eq. (6)
- Structured coalescent

Average MAF

- Coarse-grained predictions
- Asymptotic limit
Distributions of Fitness Effects

Weak

\[ s_e = \sqrt{\langle s^2 \rangle} \]

Strong

\[ N_e = N e^{-\langle U/s \rangle} \]

Deleterious fitness effect, \( s \)
Interference Selection Still Applies

- Sexual ($NR=10$), $Ns=30$, $NU=354$
- TruncatedExp($s_{max}/\bar{s}=3$), $N\bar{s}=10$, $NU=2230$
- Uniform($0,s_{max}$), $N_{s_{max}}=28.5$, $NU=1000$
- 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
Acknowledgements

Graduate Students and Postdocs:
- Elizabeth Jerison
- Ben Good
- Lauren Nicolaisen
- Katya Kosheleva
- Sebastian Akle
- Mike McDonald
- Gabriel Perron
- David van Dyken

Undergraduate Students:
- Henry Shull
- Ian Ochs
- Sadik Shahidain
- Nisreen Shiban

Collaborators:
- Joshua Plotkin
- Alex Walczak
- Greg Lang
- Richard Neher
- Oskar Hallatschek

Funding:
- McDonnell Foundation
- Milton Fund
- Sloan Foundation
- NIH/NIGMS
- NSF PoLS