How does pervasive natural selection alter patterns of genetic diversity?

Michael Desai, Harvard University
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 $Ns \rightarrow \infty$ while holding $NU/Ns$ constant.

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

When $Nse \frac{U}{s} \sim 1$ 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

A

\[ \text{Interference Selection collapse holds generally} \]

B

C

D

E

Legend:
- Asexual
- \( NR = 10 \)
- \( NR = 100 \)
- Neutral expectation
- Coarse-grained predictions
- Large \( N\sigma \) limit

\[ Ns = 30, \ Nu = 300 \]
Two limits: background selection and interference selection

Background selection: $Ns \to \infty$ while holding $NU/Ns$ constant
Interference selection: $Ns \to 0$ while holding $N\sigma$ constant
Intuition: “coarse-graining” the fitness distribution

A

Fraction of individuals

Relative fitness

$-\sigma$ $0$ $\sigma$ $\chi_c$

$MRC$

Number of individuals

Deleterious load

$-42s$ $-41s$ $-40s$ $-39s$ $-38s$ $-37s$ $-36s$ $-35s$

$U_d'$ $U_n'$

$S'$
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

Scaled diversity, $\pi/\pi_0$

Deleterious load, $U/s$

Scaled diversity, $\pi/\pi_0$

Stddev fitness, $N\sigma$

Eq. (2)

Neutral expectation

Background selection

Interference selection

Coarse-grained predictions

$Ns$

$10^3$

$10^2$

$10^1$

$10^0$

$10^{-1}$

$10^{-2}$

$10^{-3}$
Coarse-Grained Predictions

Relative SFS, $Q_n(i)$

Derived allele frequency, $i/n$

- Single-s, $N_s = 30$, $NU = 300$
- Coarse-grained predictions
- Large $N\sigma$ limit
- Neutral expectation
Coarse-Grained Predictions

A

- **Neutral expectation**
- **$Ns = 30$, $NU = 300$**

B

Legend:
- Asexual
- $NR = 10$
- $NR = 100$
- Neutral expectation
- Large $N_\sigma$ limit

C

- Average MAF (%)

D

- Tajima's $D/D_{min}$

E

- Non-monotonicity, $\Upsilon$

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

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

\[ \Delta R \cdot T_{MRCA} \gg 1 \quad \text{(Effectively free recombination)} \]

\[ L \downarrow b \cdot r \cdot T_{MRCA} \approx 1 \]

“Linkage block”

Independent

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

Mutation density, \( U/R \)

Average MAF

Mutation density, \( U/R \)
Distributions of Fitness Effects

Weak

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

Strong

\[ N_e = N e^{-\langle u \rangle_s / s} \]

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

![Graph showing the Relative SFS, $Q_n(i)$, for different scenarios.

- **Sexual** ($NR=10$), $N_s=30$, $N_U=354$
- **TruncatedExp** ($s_{\text{max}}/s=3$), $N_s=10$, $N_U=2230$
- **Uniform** ($0,s_{\text{max}}$), $N_{s_{\text{max}}}=28.5$, $N_U=1000$
- **Finite sites** ($L=10^5$), $N_s=21.4$, $N_U=600$
- **Single-s**, $N_s=30$, $N_U=300$

- 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:

Sergey Kryazhimskiy
Dan Rice
Evgeni Frenkel
Julia Piper

Collaborators:

Joshua Plotkin
Alex Walczak
Greg Lang
Richard Neher
Oskar Hallatschek

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