

Negative Selection

Ryan D. Hernandez

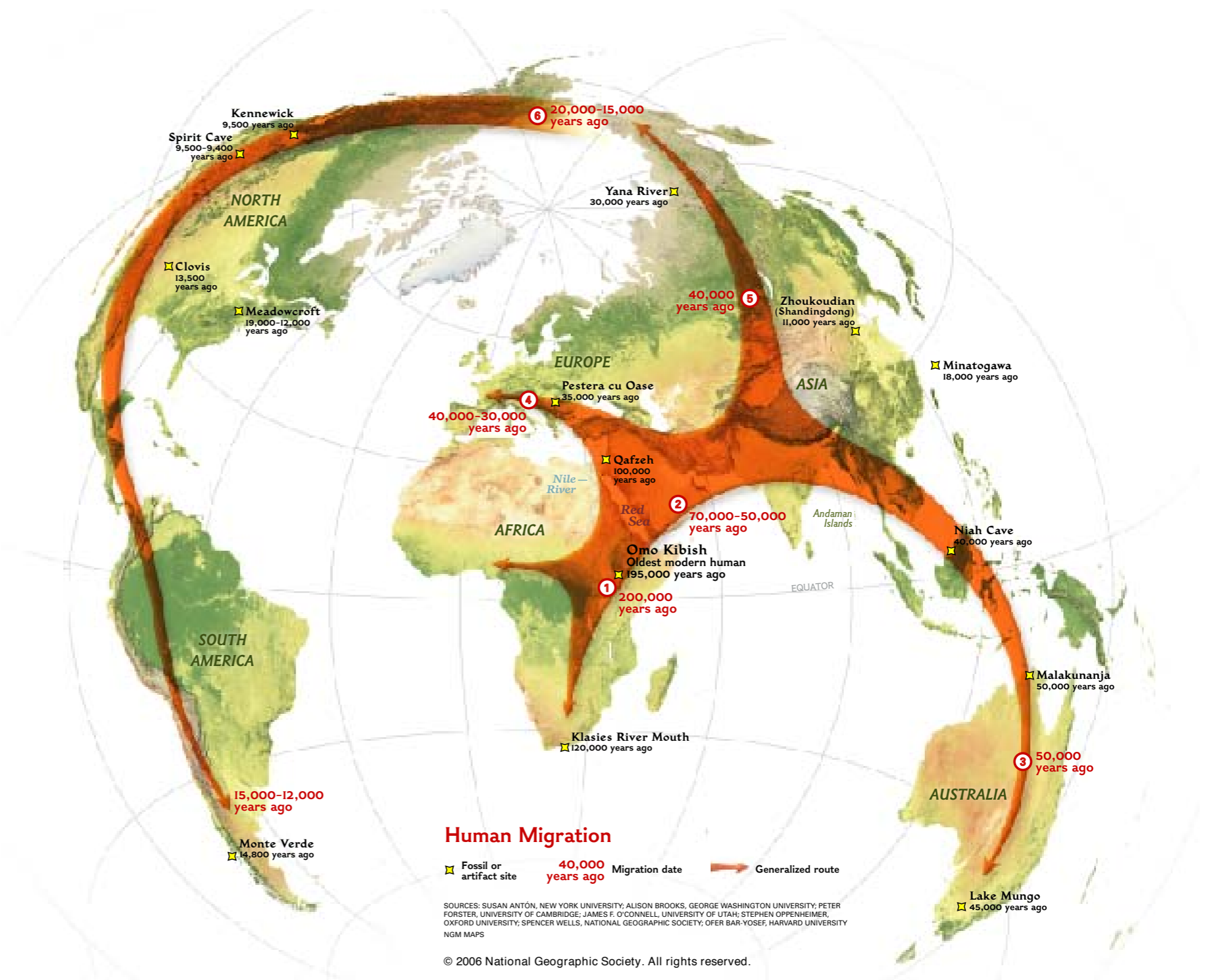
ryan.hernandez@ucsf.edu¹



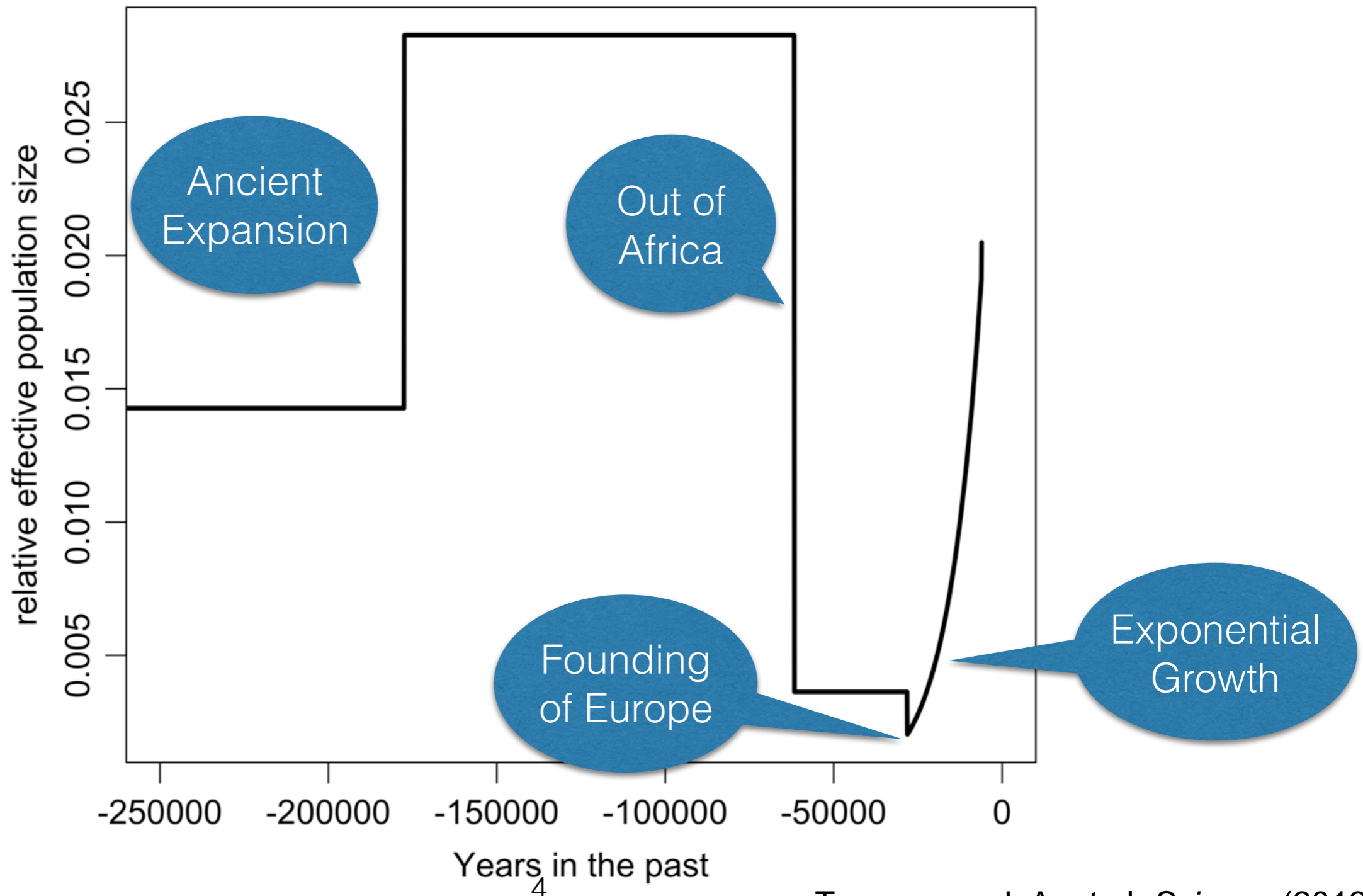
Department of Bioengineering and Therapeutic Sciences
a joint department of the UCSF Schools of Pharmacy and Medicine



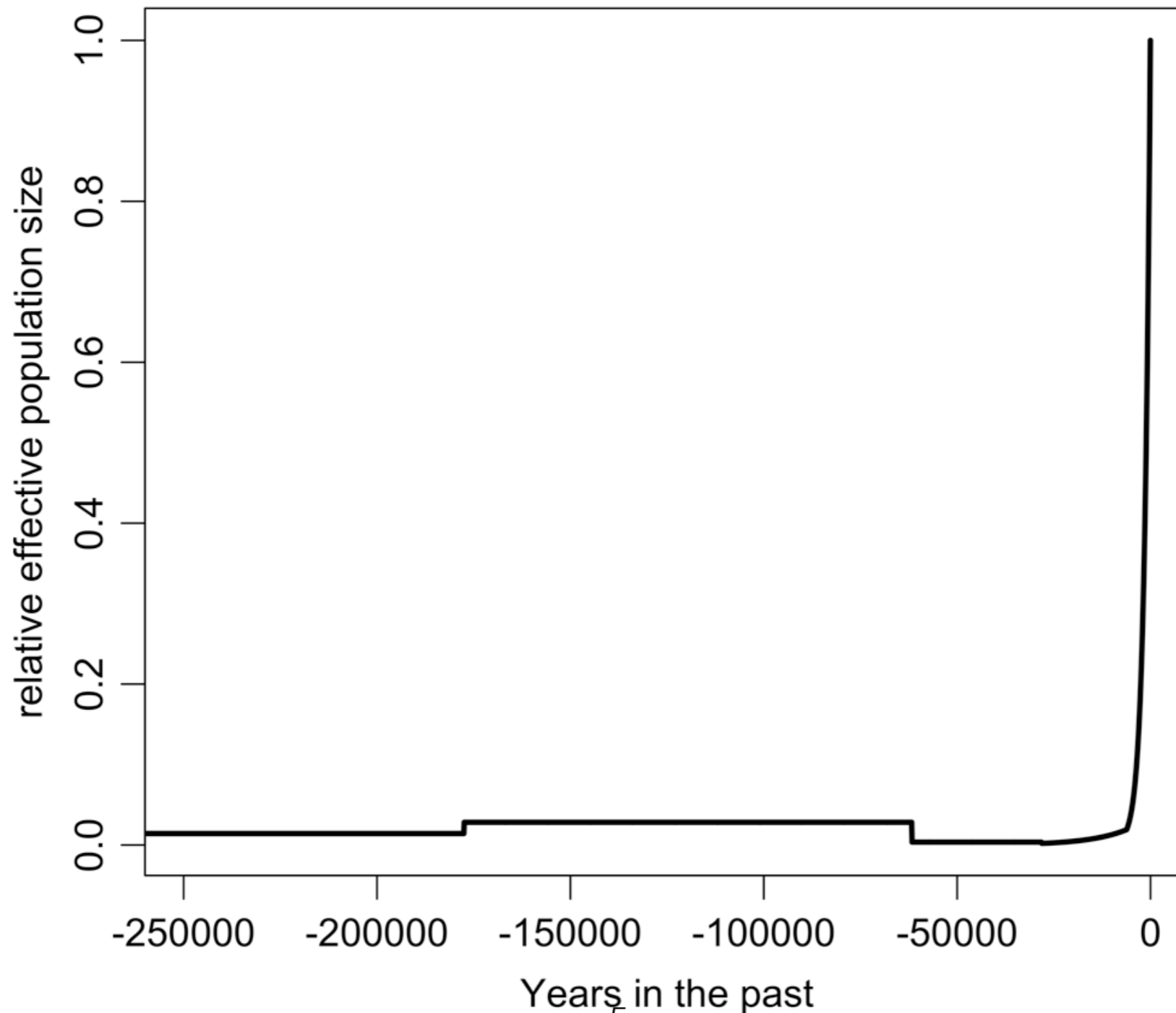
Human colonization of the world



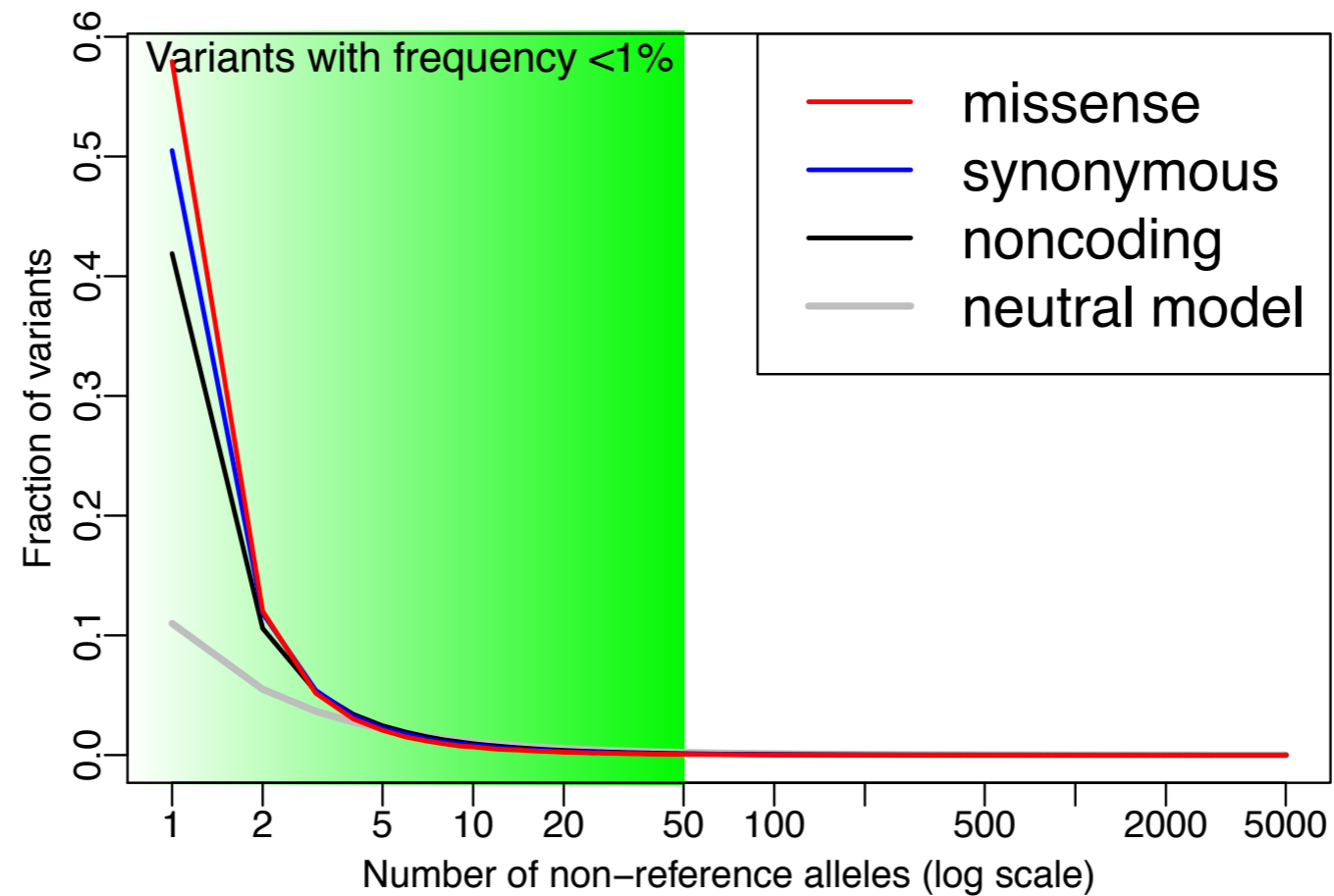
Growth of European “Effective Population Size”



Growth of European “Effective Population Size”



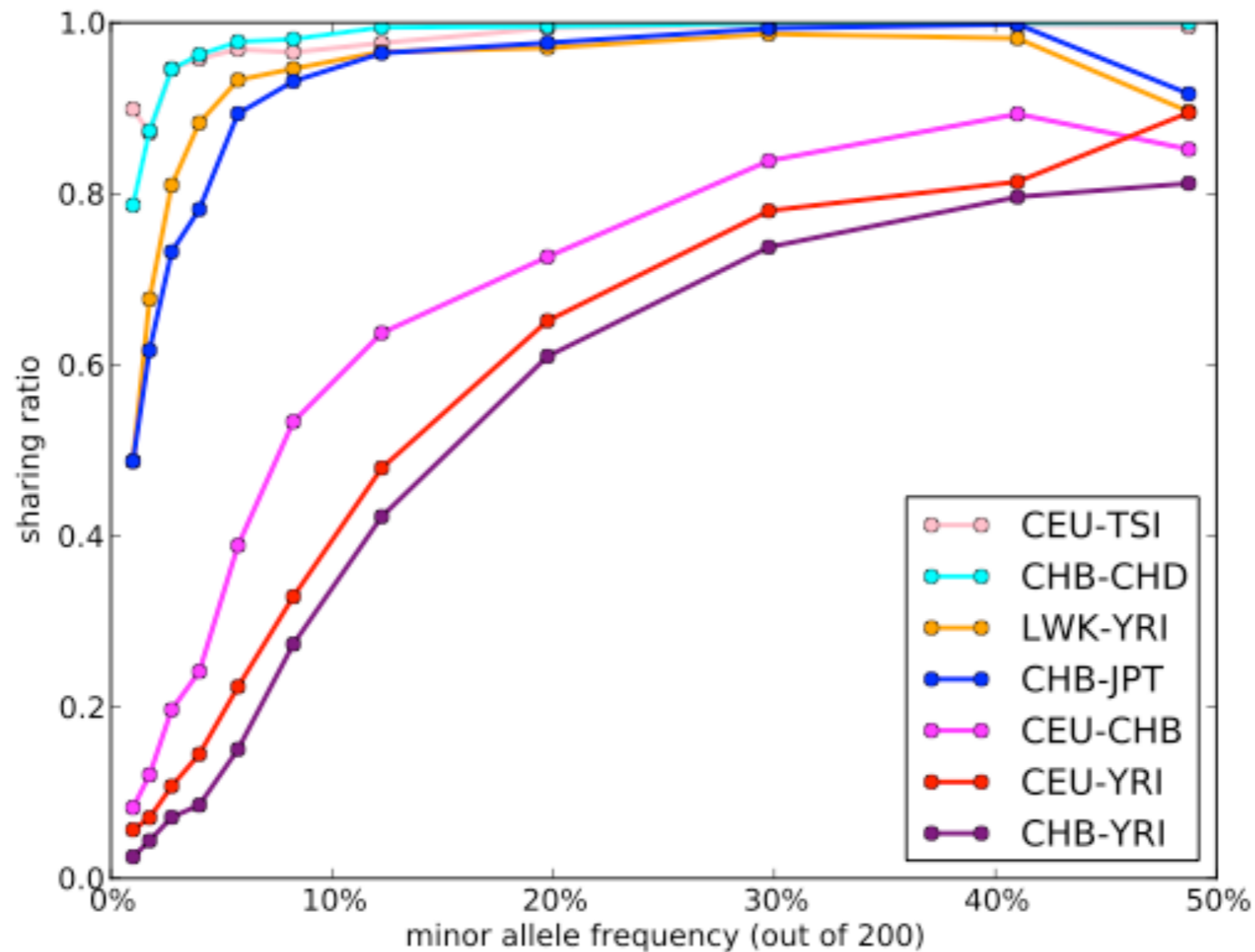
Vast Majority of human genetic variation is rare



Class	Fraction of variants
Missense	92.6%
Synonymous	88.5%
Non-coding	82.3%



Rare variation is highly population-specific



1000 Genomes exon pilot experiment

The Effect of Positive Selection

Adaptive

Neutral

Nearly Neutral

Mildly Deleterious

Fairly Deleterious

Strongly Deleterious



The Effect of Positive Selection

Adaptive

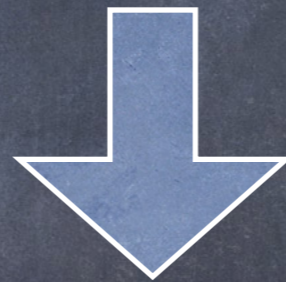
Neutral

Nearly Neutral

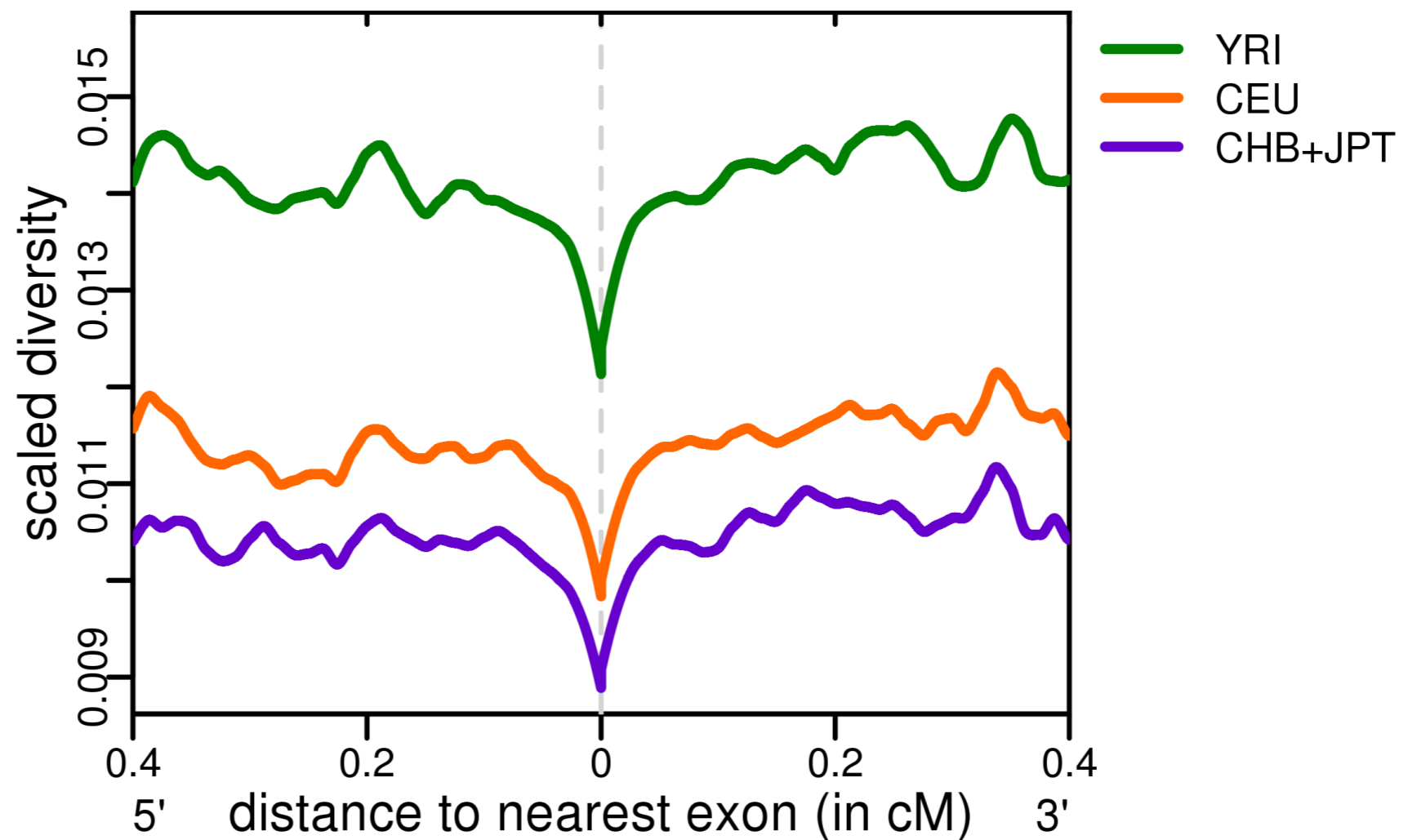
Mildly Deleterious

Fairly Deleterious

Strongly Deleterious



Coding regions tend to have the lowest levels of diversity in the genome



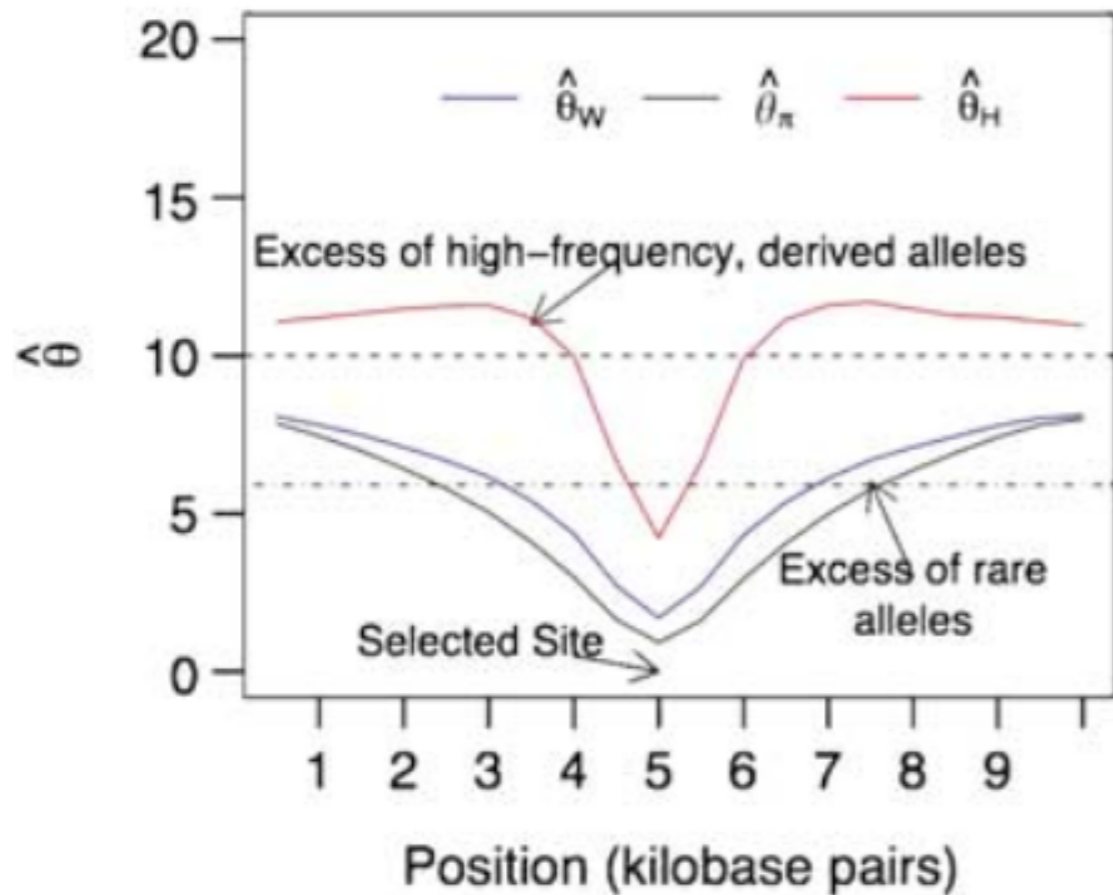
What are the predominant evolutionary forces driving human genomes?!

Eyre-Walker & Keightley (2009) ~**40%** of amino acid substitutions were **advantageous**

Boyko et al (2008) **10-20%** of amino acid substitutions were **advantageous**

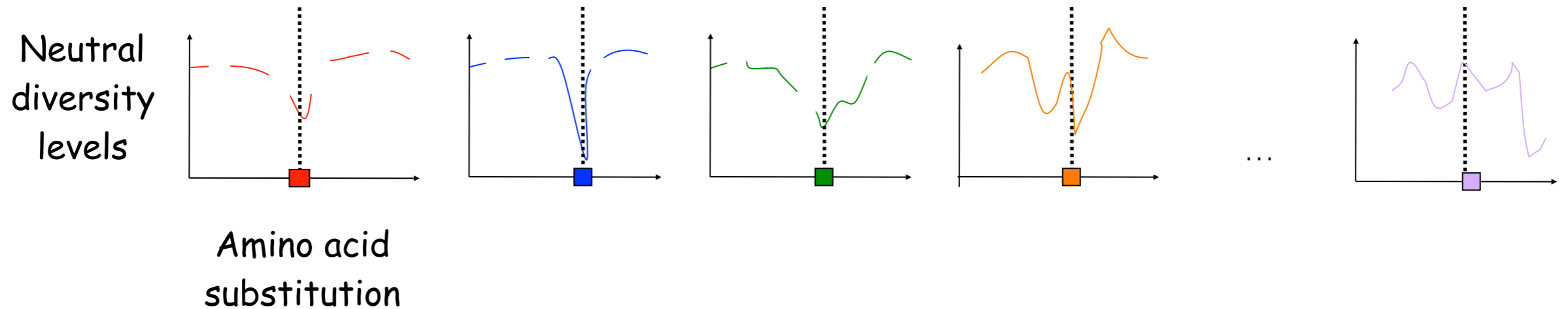
Williamson et al (2007) **10%** of the genome affected by **selective sweeps**

Diversity levels around a selective sweep

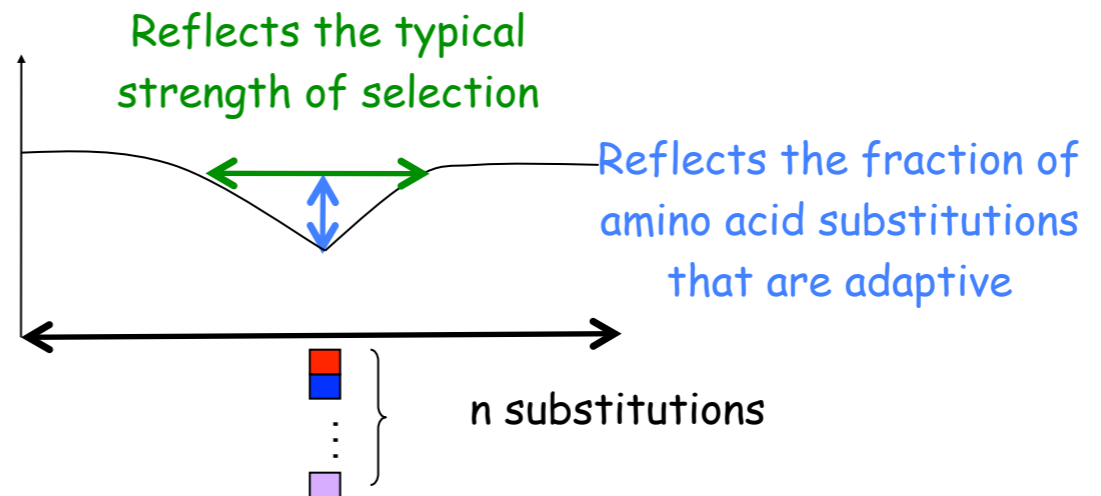


Thornton et al (2007): Simulation of patterns of **neutral** diversity around a **selective sweep**

The footprint of adaptive amino acid substitutions

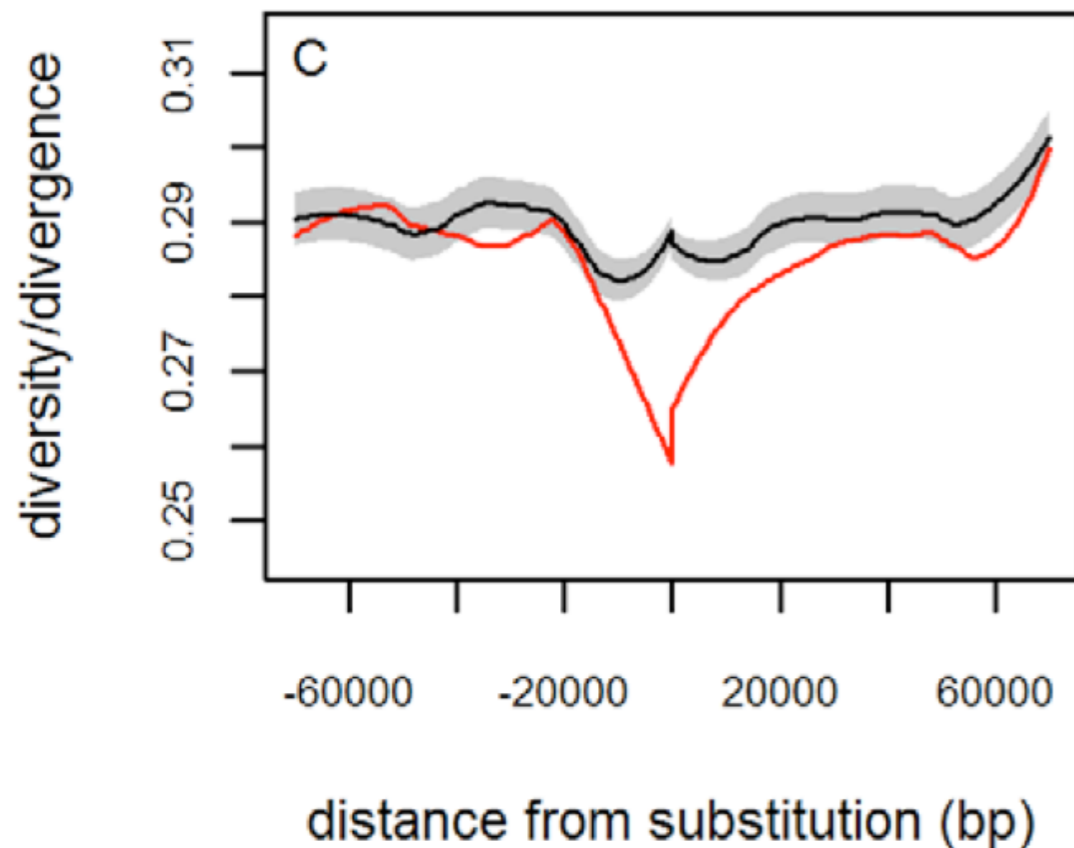


- Goal: compare the pattern around **amino acid substitutions** to the pattern around **synonymous substitutions**.



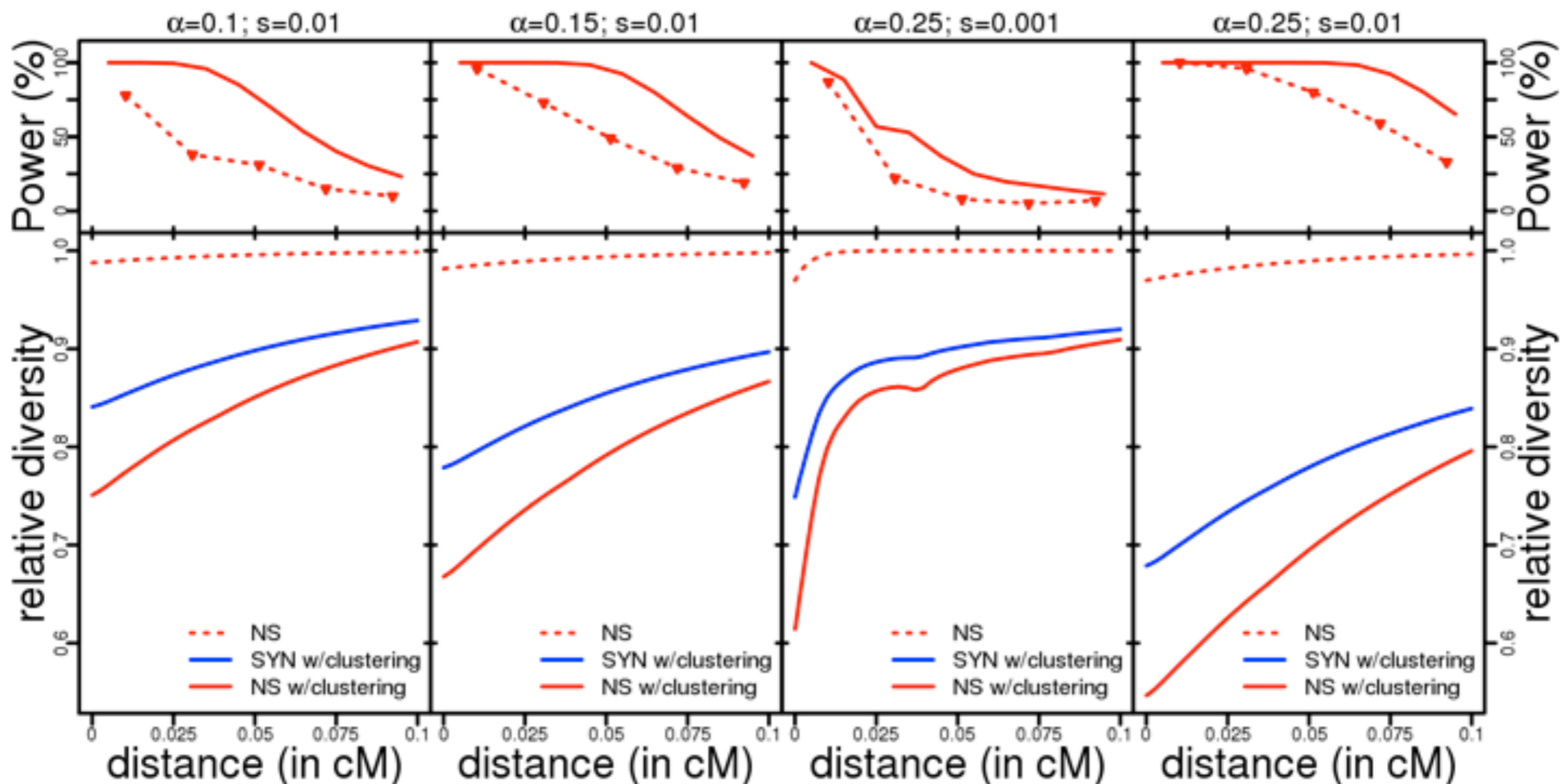
Other organisms...

Drosophila

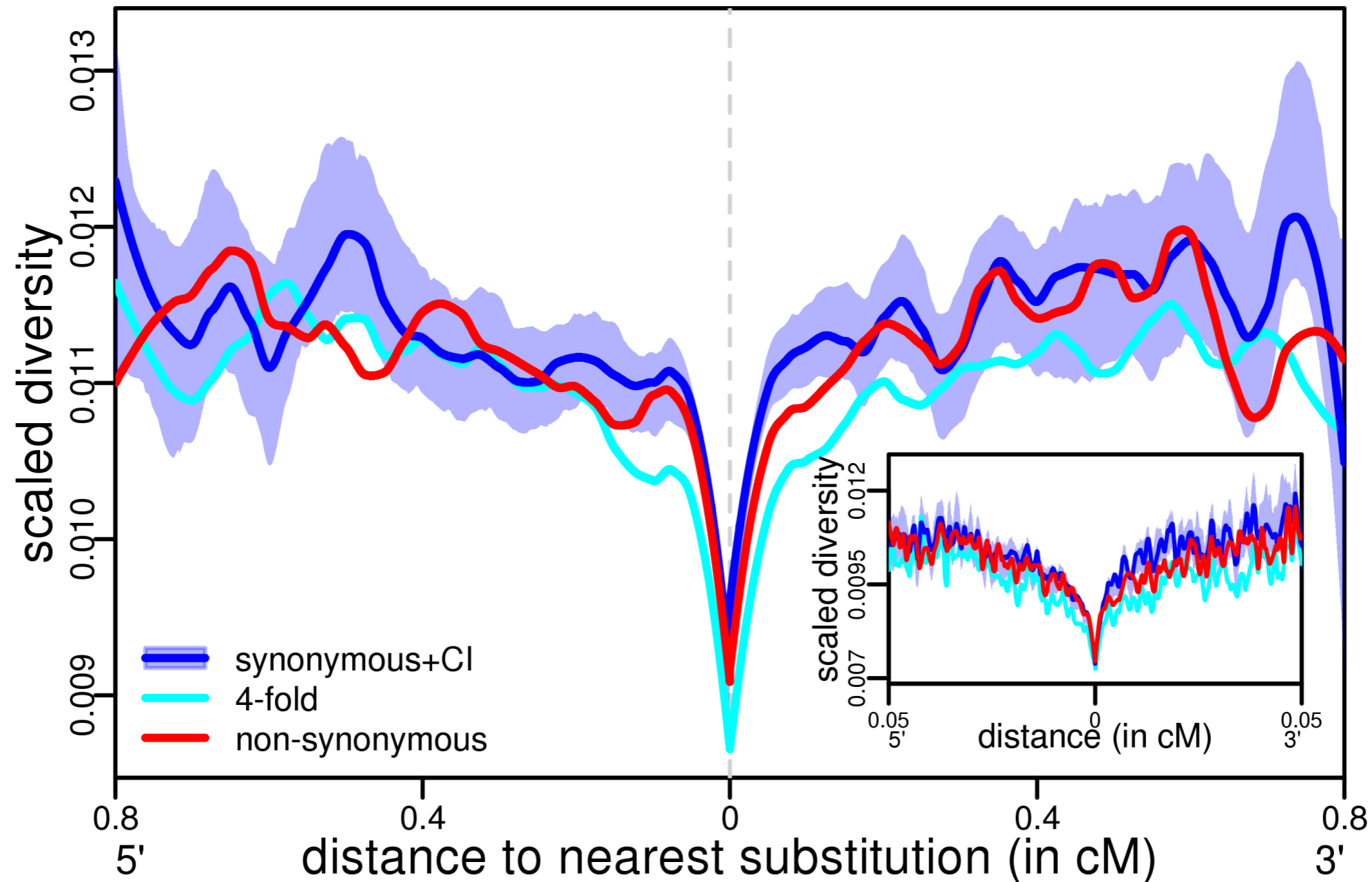


Sattath et al (2011) estimate
~13% of amino acid
substitutions were adaptive.

Our approach is highly powered to detect positive selection



Observed Patterns of Diversity Around Human Substitutions



The Effect of Negative Selection

Adaptive

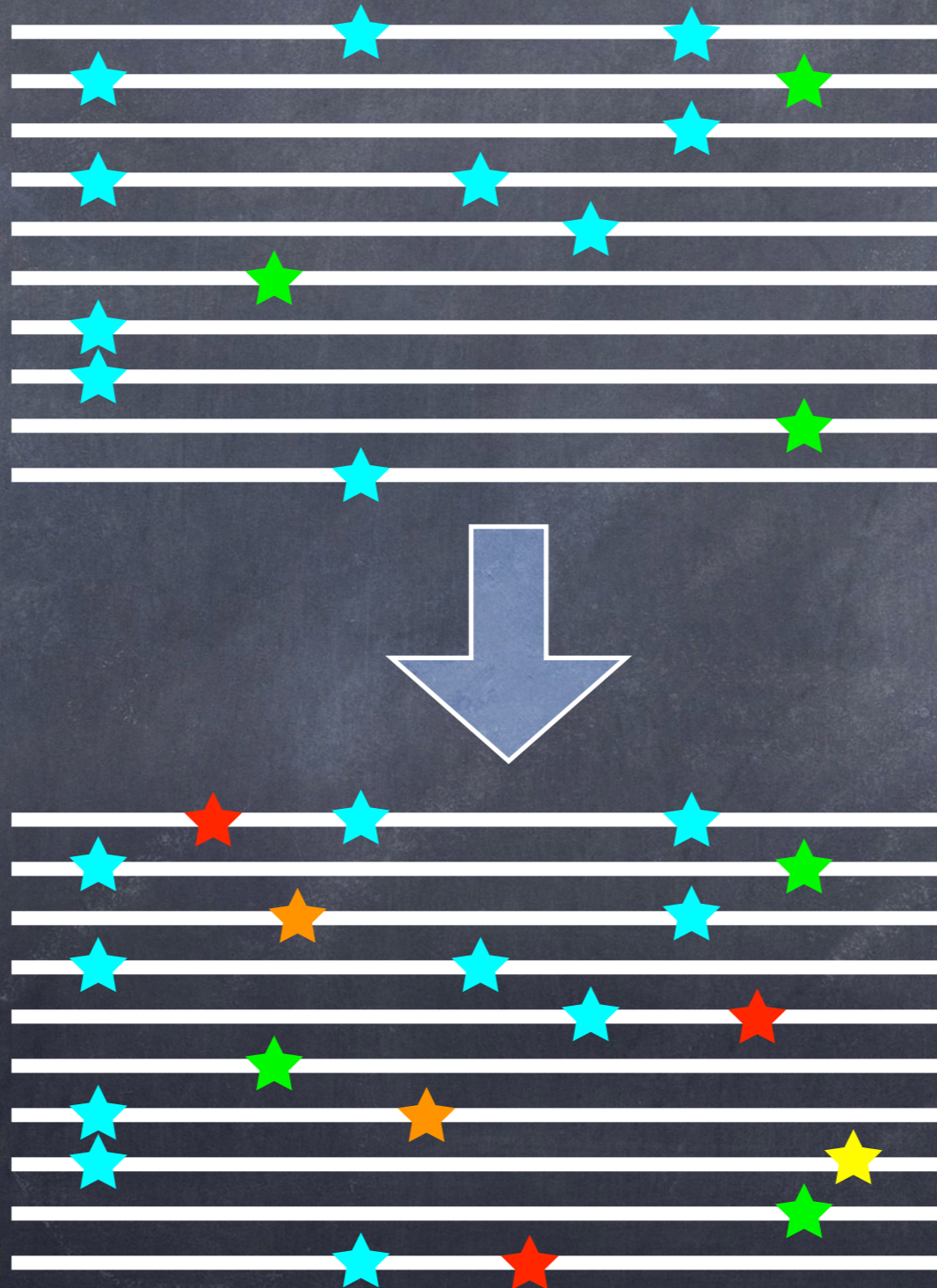
Neutral

Nearly Neutral

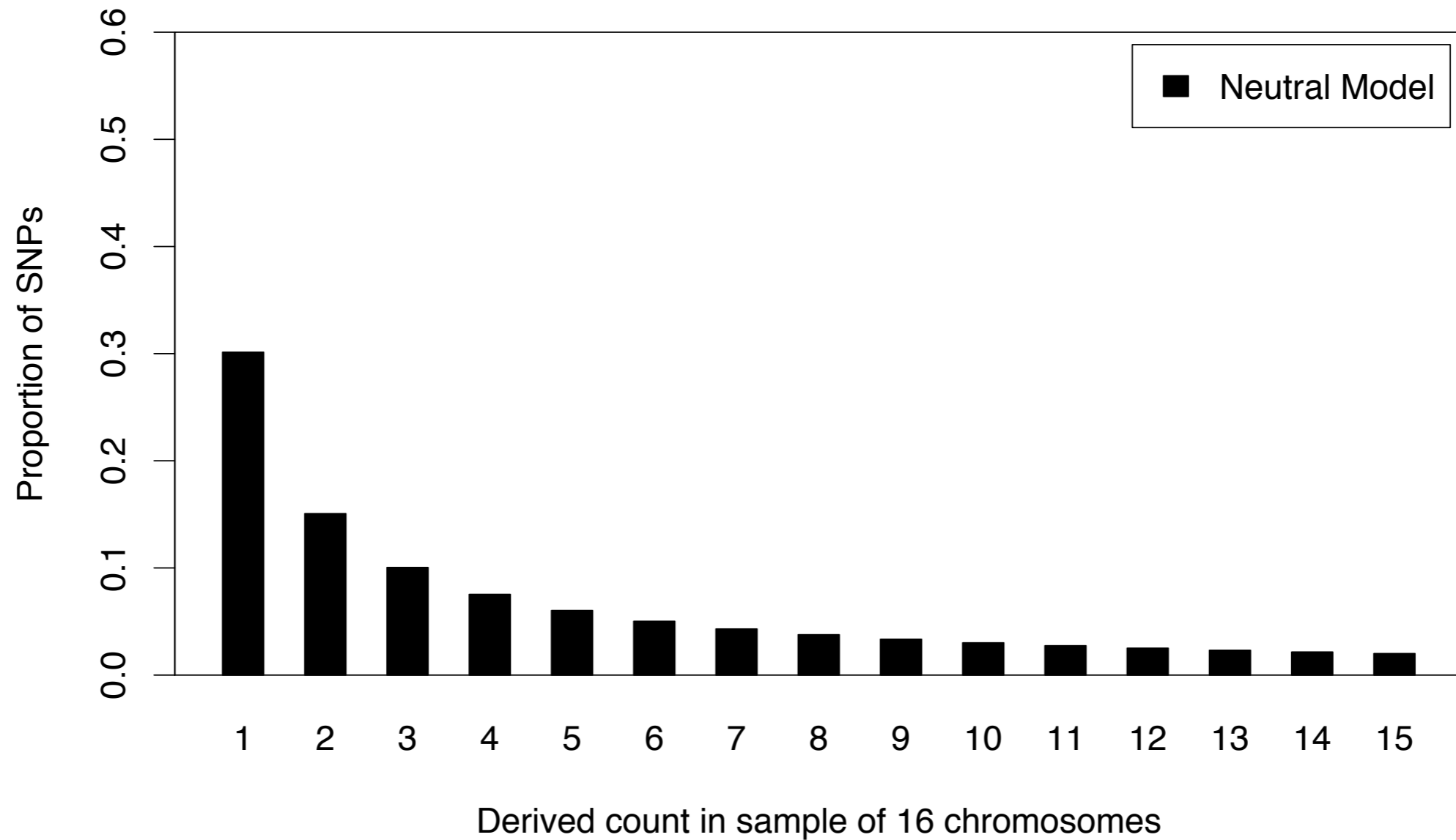
Mildly Deleterious

Fairly Deleterious

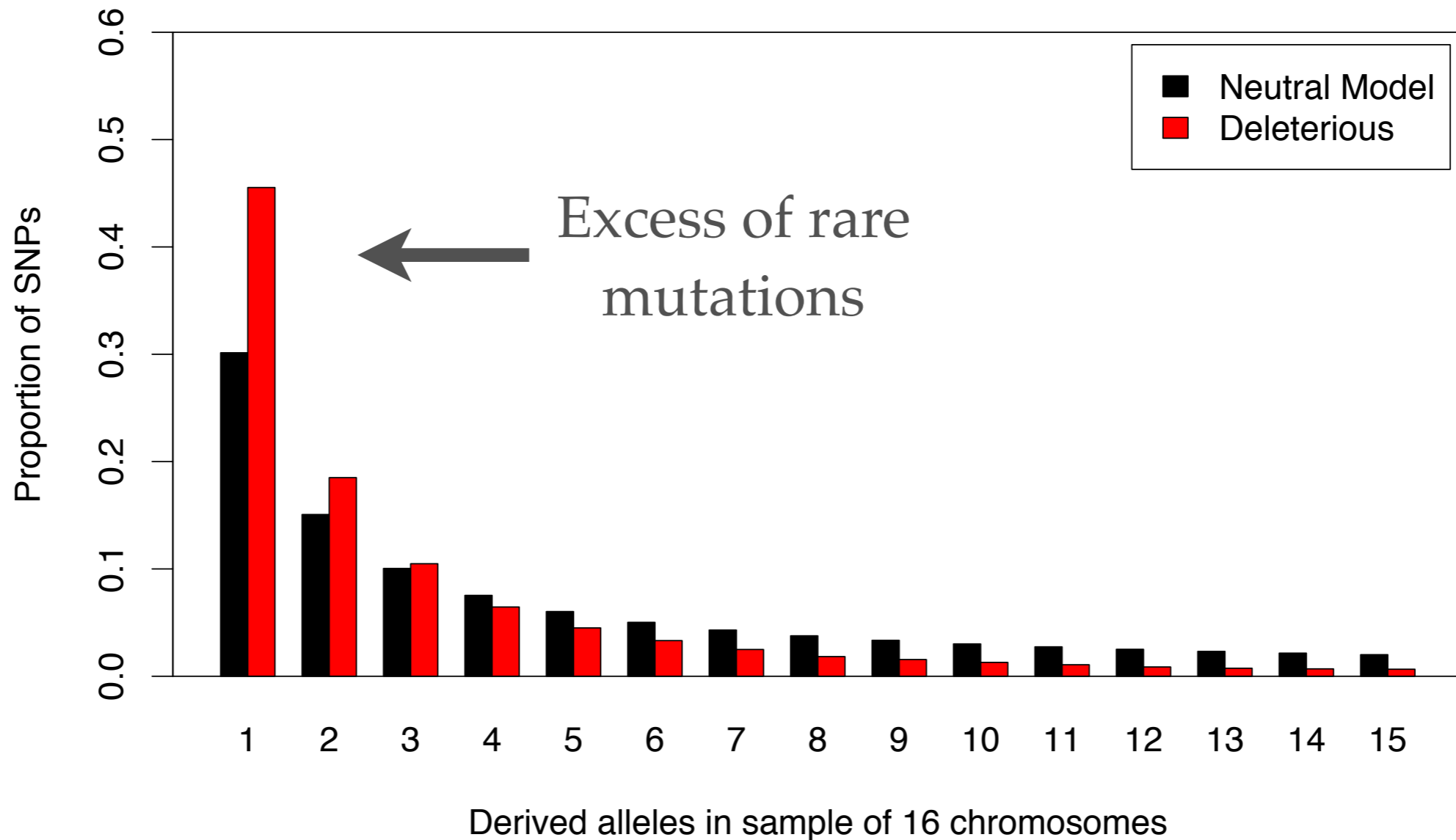
Strongly Deleterious



Site-Frequency Spectrum

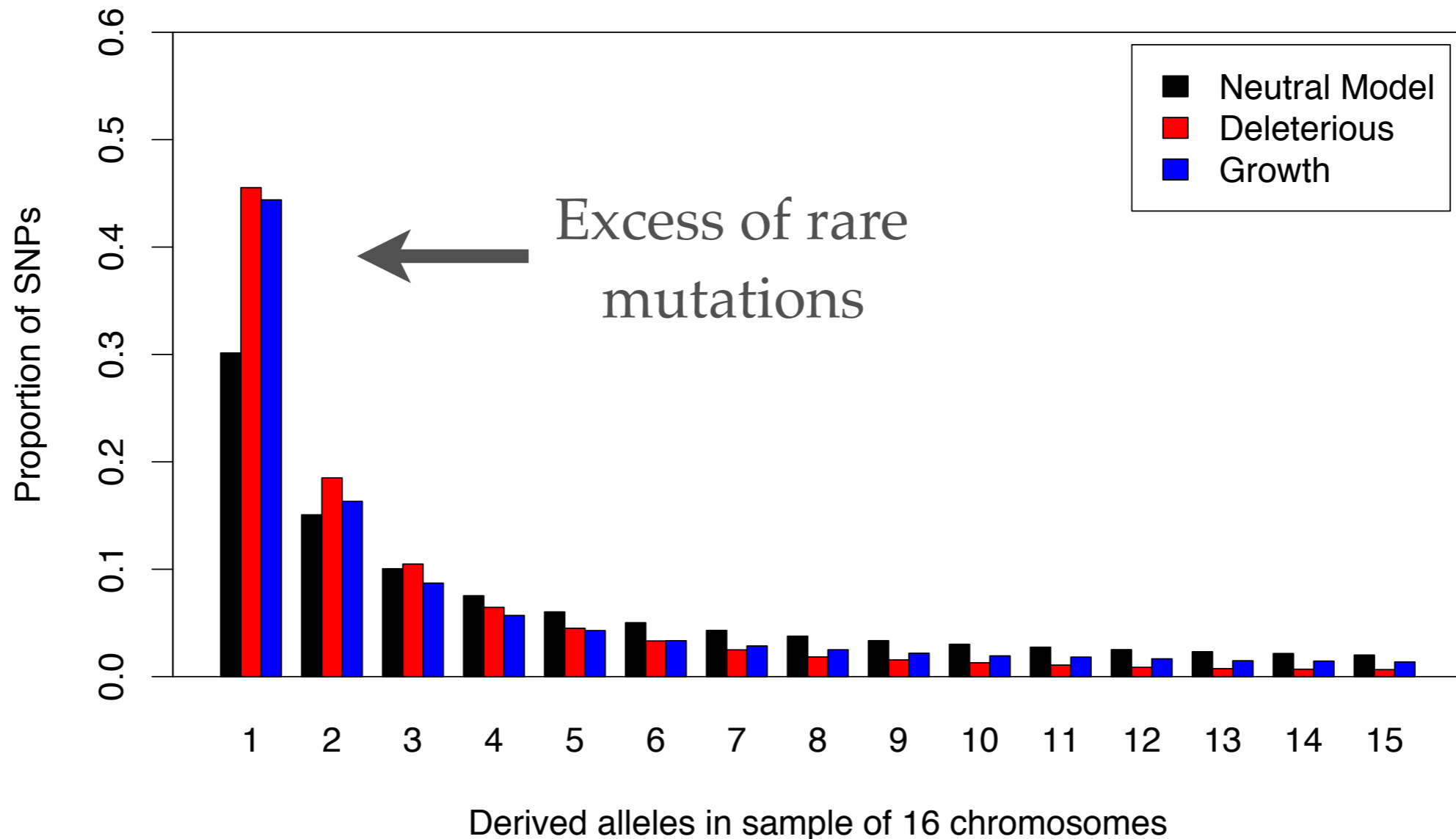


Site-Frequency Spectrum



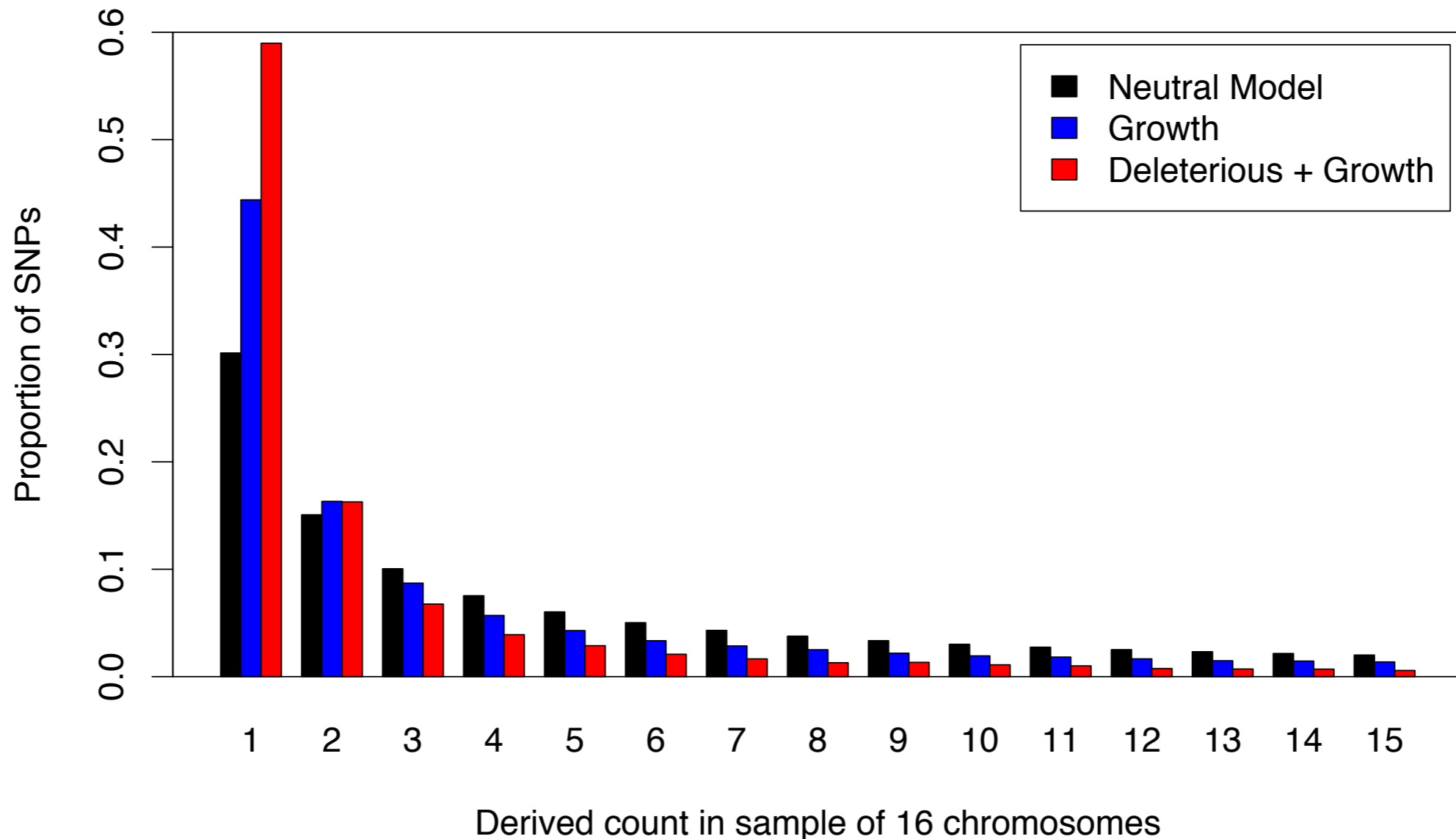
Site-Frequency Spectrum

Both population expansions as well as negative selection causes an excess of rare alleles.

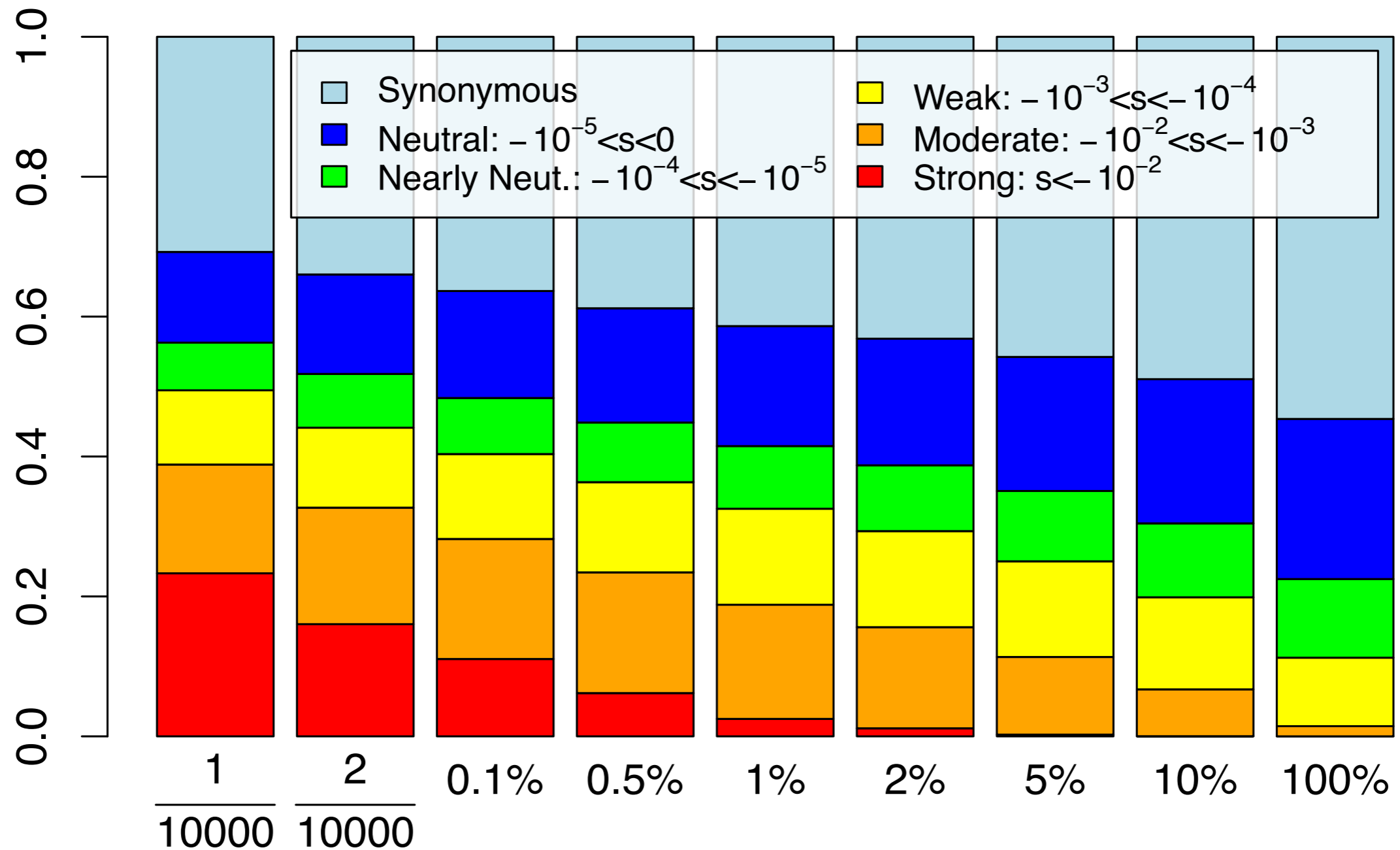


Site-Frequency Spectrum

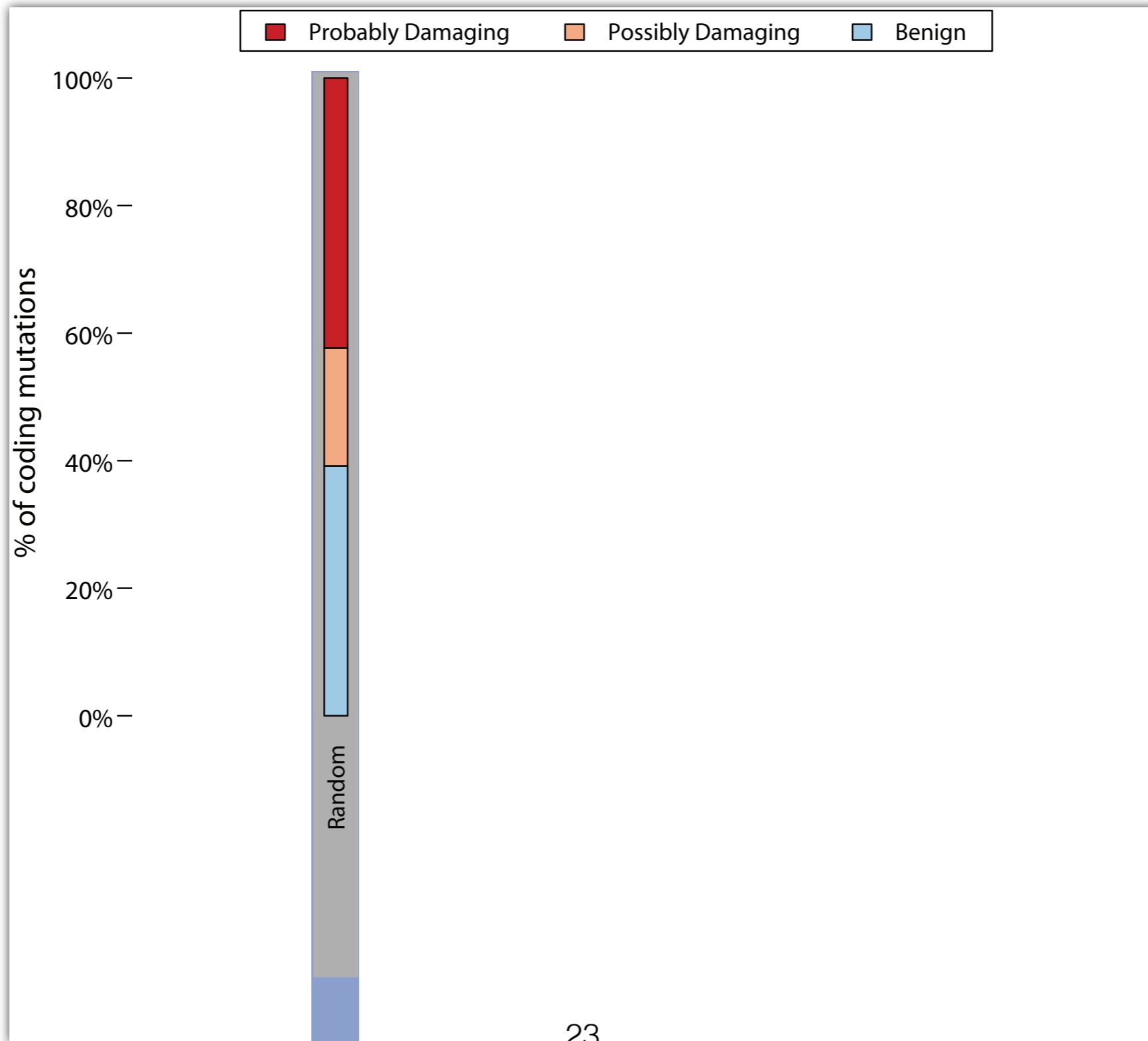
Selection is more efficient in large populations, and more deleterious mutations are introduced.



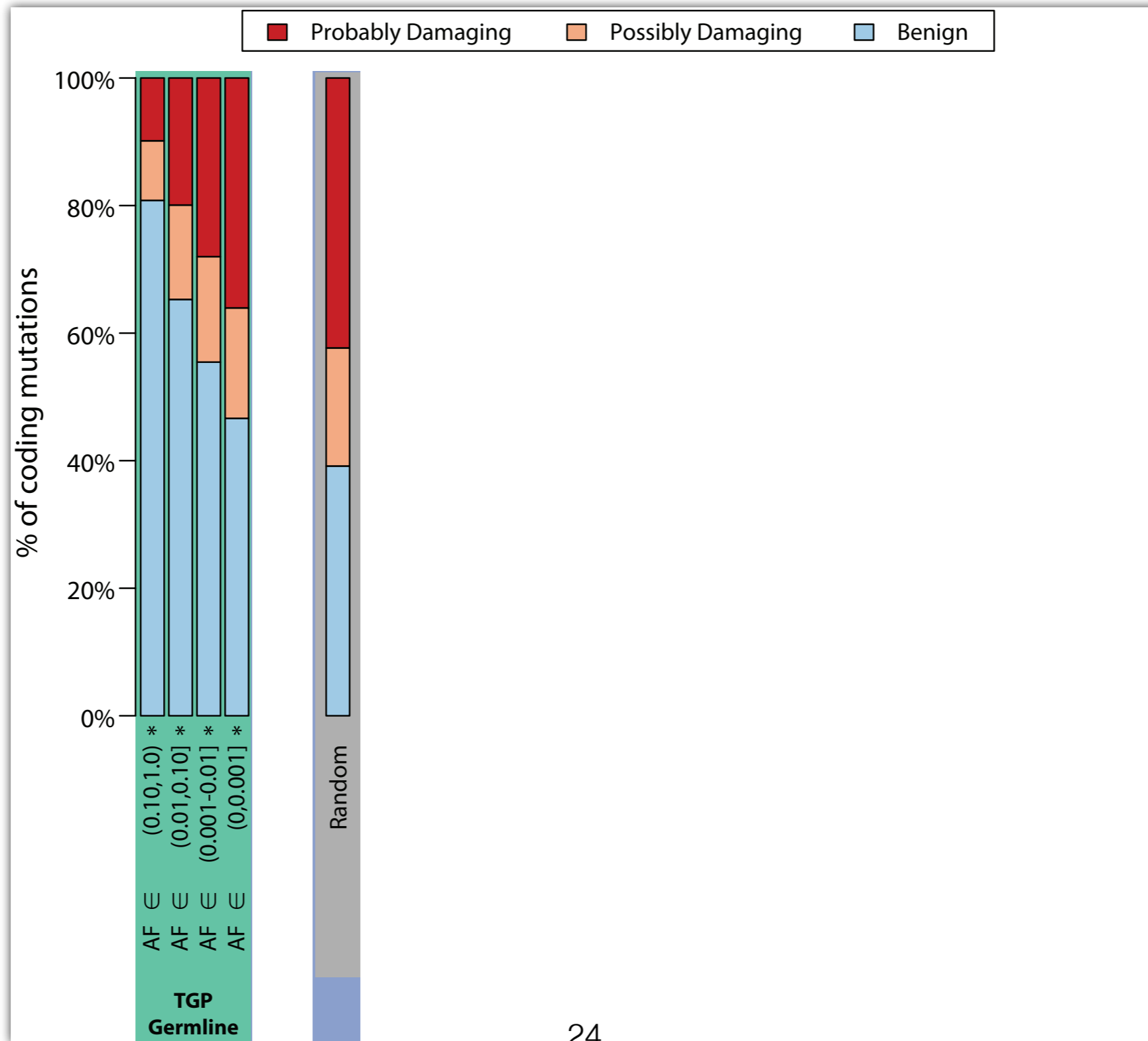
Inferred Distribution of Fitness Effects



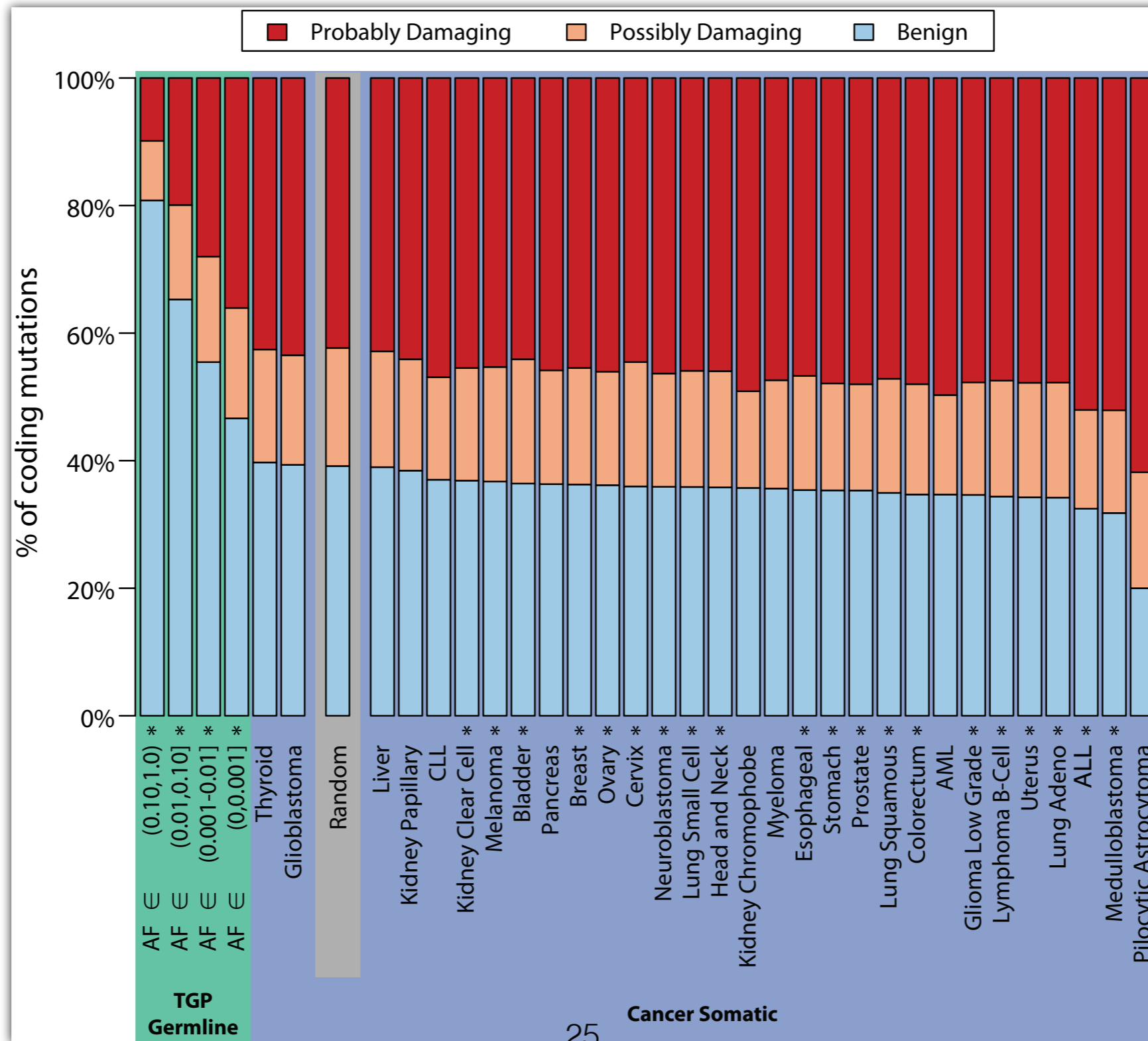
Observed Effect of Selection



Observed Effect of Selection



Observed Effect of Selection



The Effect of Negative Selection

Consequences:

- ④ Some proportion of chromosomes eliminated each generation
 - ➡ Decreased effective population size ($f_0 N_e$)
 - ➡ Decreased neutral variation ($f_0 \pi$)
- ④ While neutral variation can be lost, some neutral mutations may increase in frequency

Background
selection

Background selection (BGS)

- Definition: The reduction of diversity at a **neutral** locus due to the effects of linked deleterious selection
- Can estimate the effect of BGS by comparing **observed** diversity at neutral sites compared to the level of diversity you would **expect** under neutrality!
- π/π_0

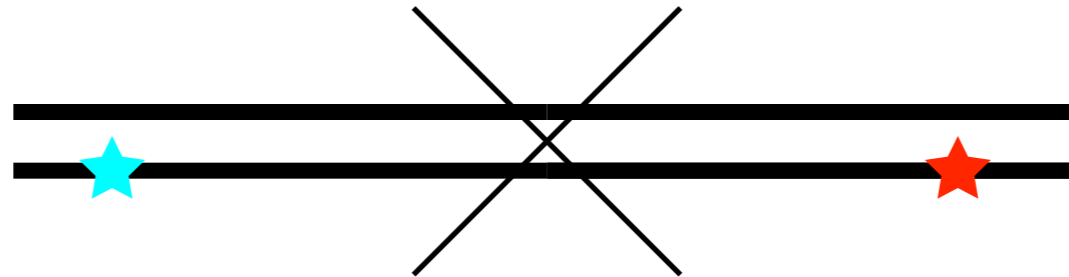
Earlier Theoretical Work

Hudson & Kaplan (1995)

$$f_0 = \exp\left(-\frac{U}{s + R}\right)$$

- U = deleterious mutation rate
- s = selection coefficient
- R = recombination rate

Effect of Recombination



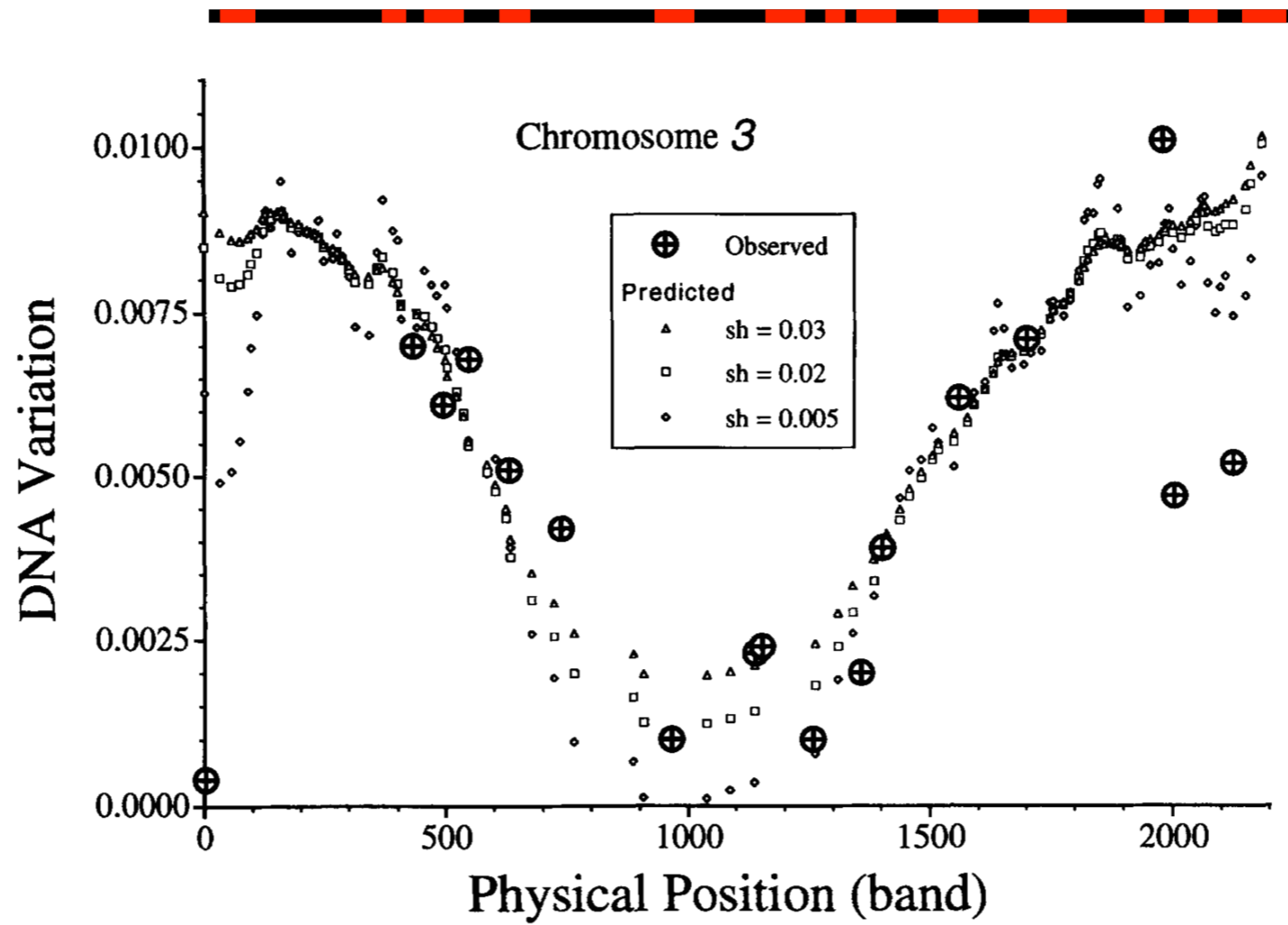
With recombination, neutral mutations can escape the grip of deleterious mutations.

Multiple Targets of Deleterious Mutations



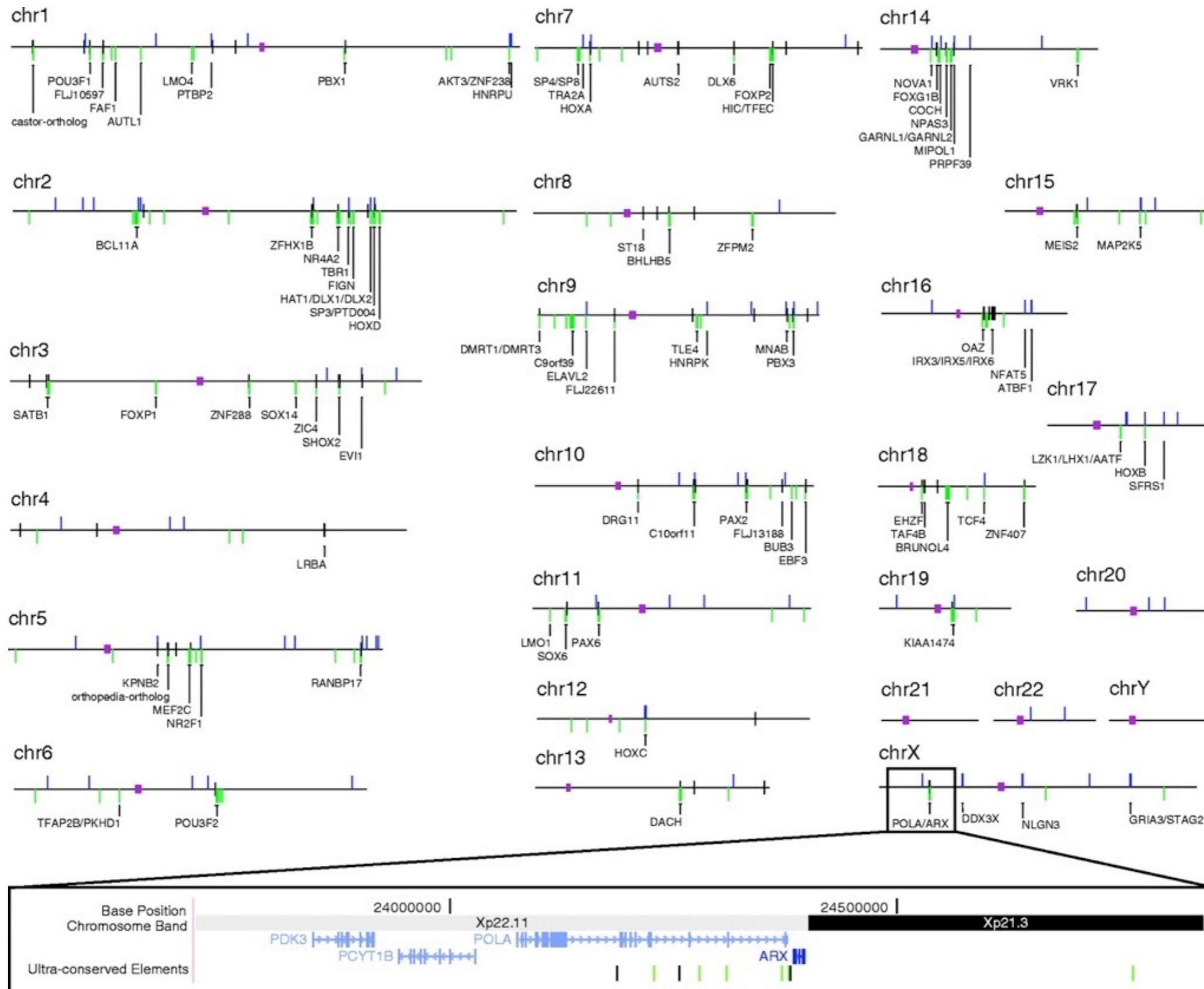
Consider a chromosome composed of neutral loci and deleterious loci

Drosophila



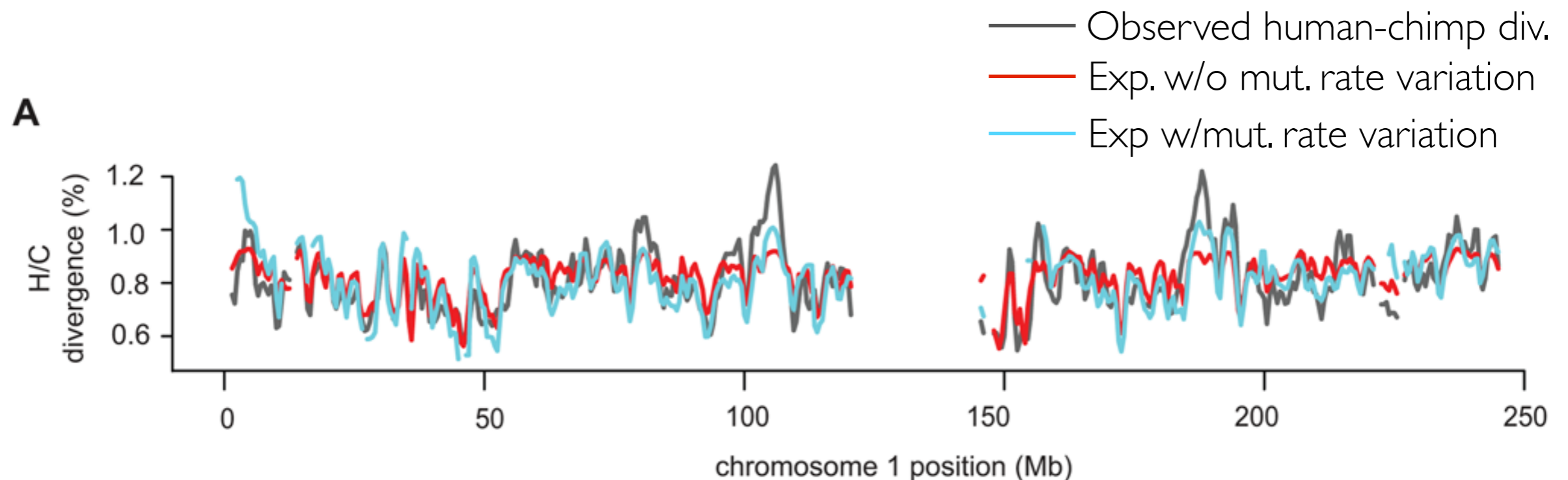
Hudson & Kaplan (1995)

Distribution of Ultraconserved Elements in the Human Genome

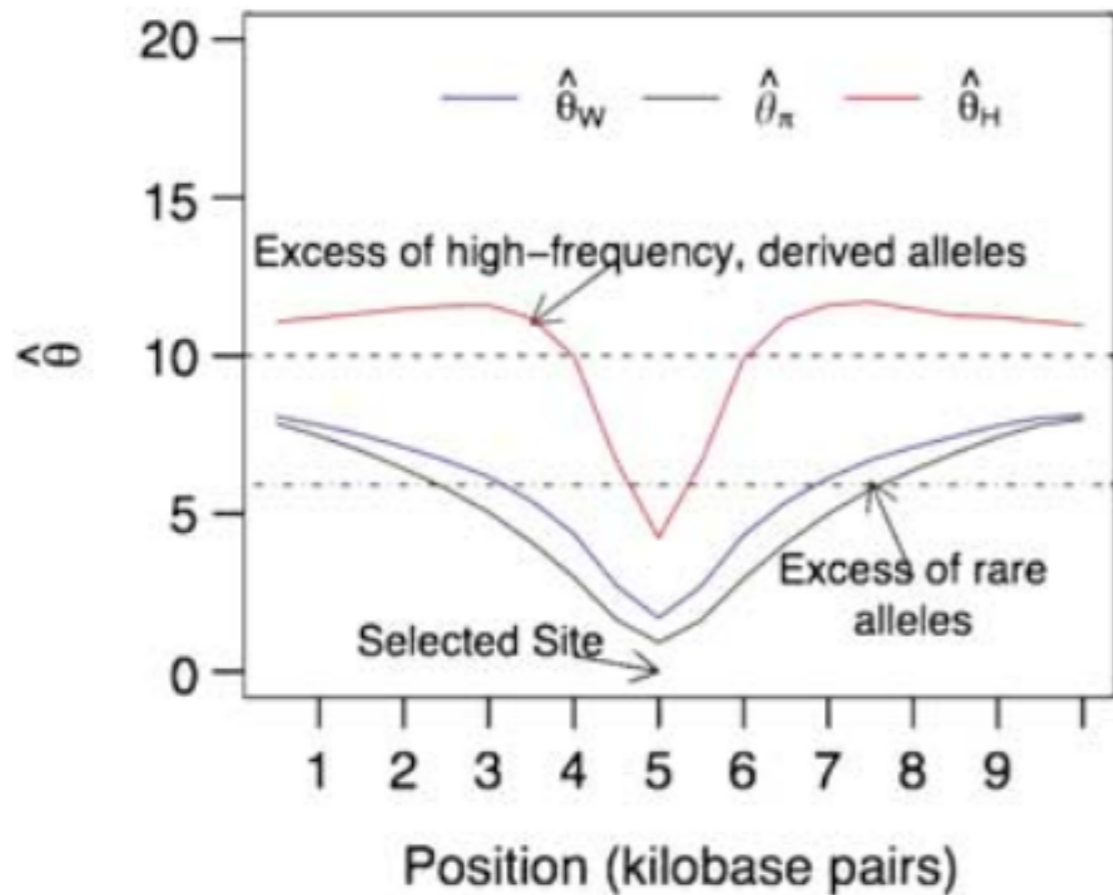


Background Selection

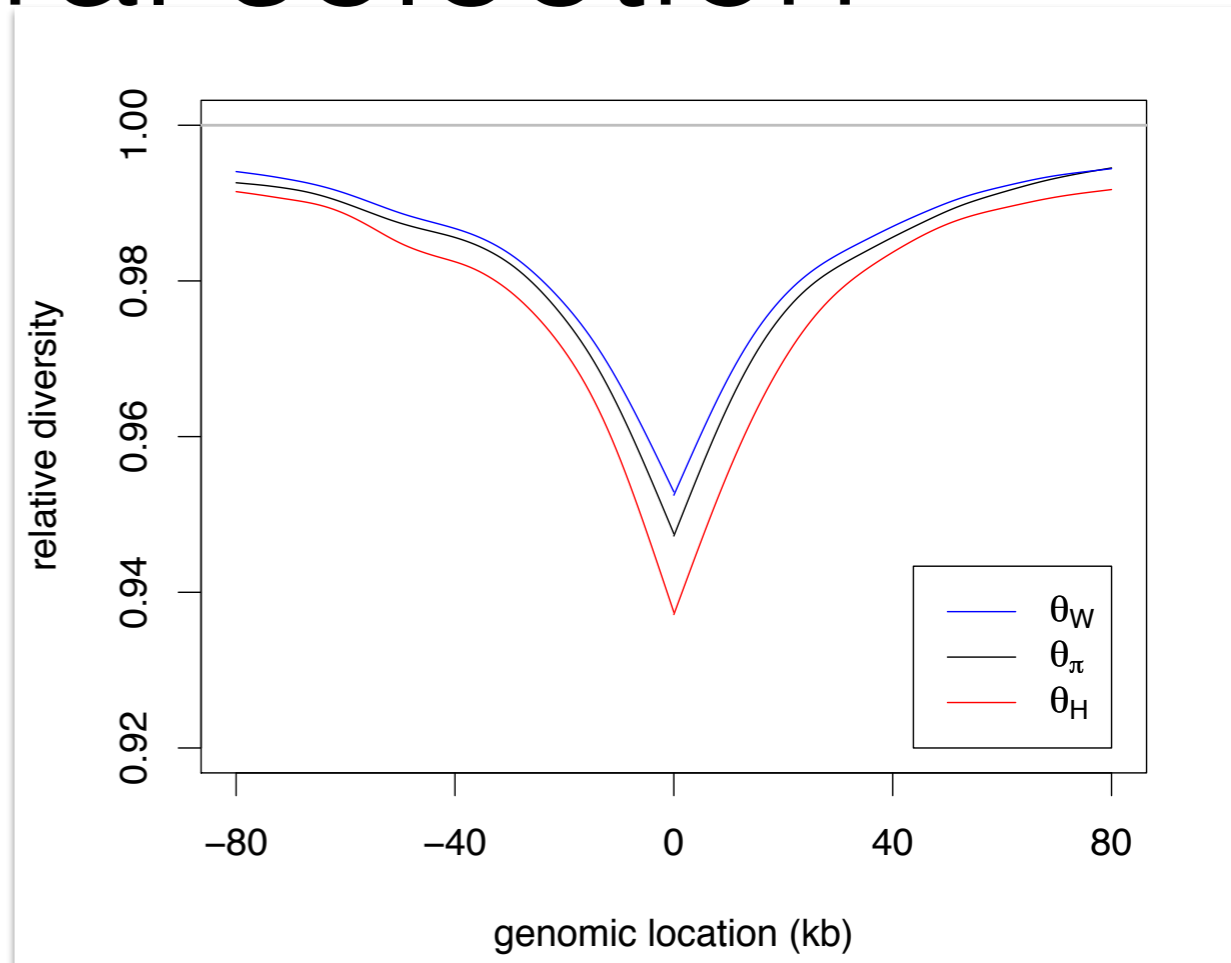
- The effects of the continual removal of deleterious mutations by natural selection on variability at linked sites.



Diversity levels around sites subject to natural selection

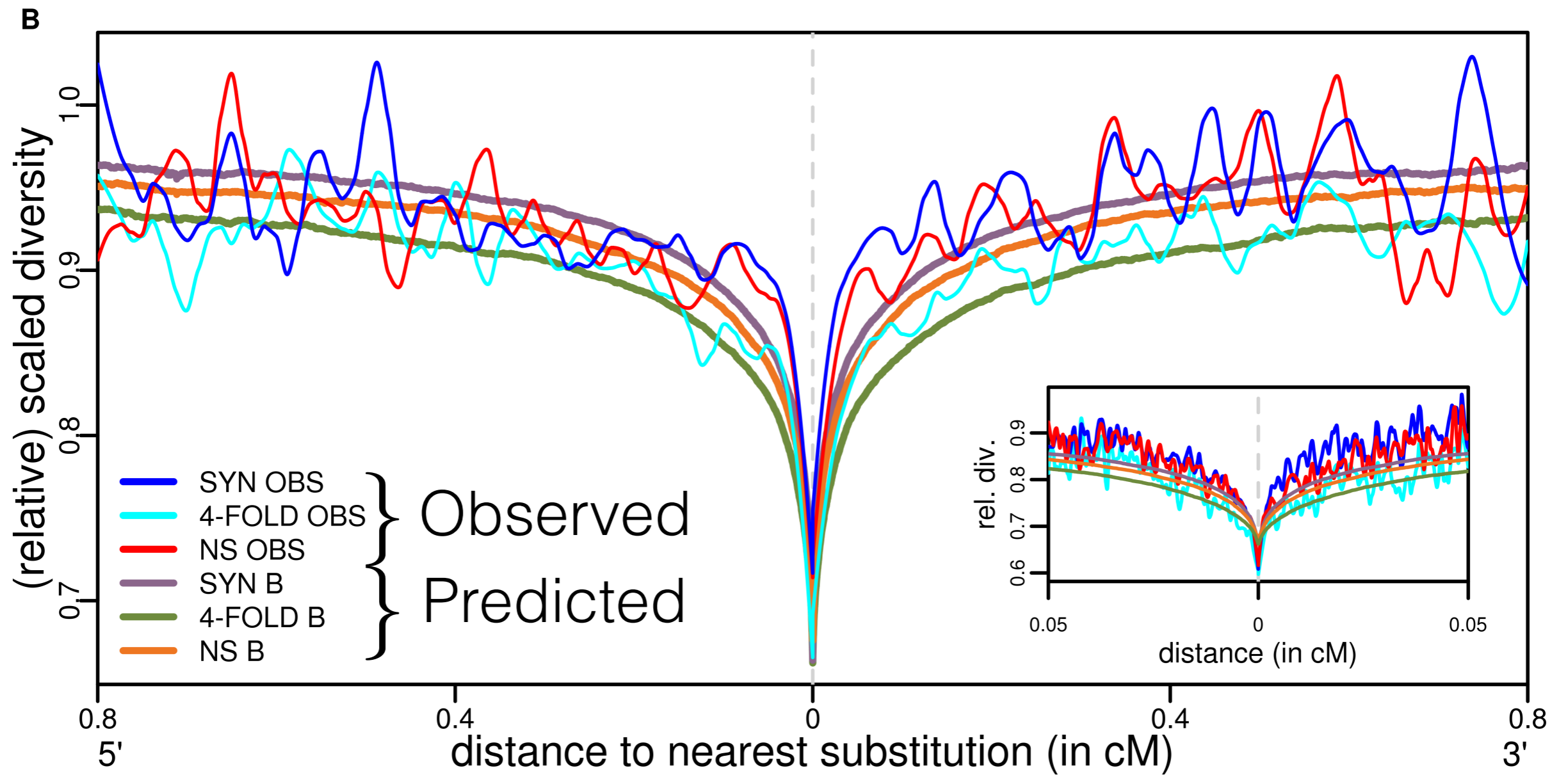


Thornton et al (2007): Simulation of patterns of **neutral** diversity around a **selective sweep**



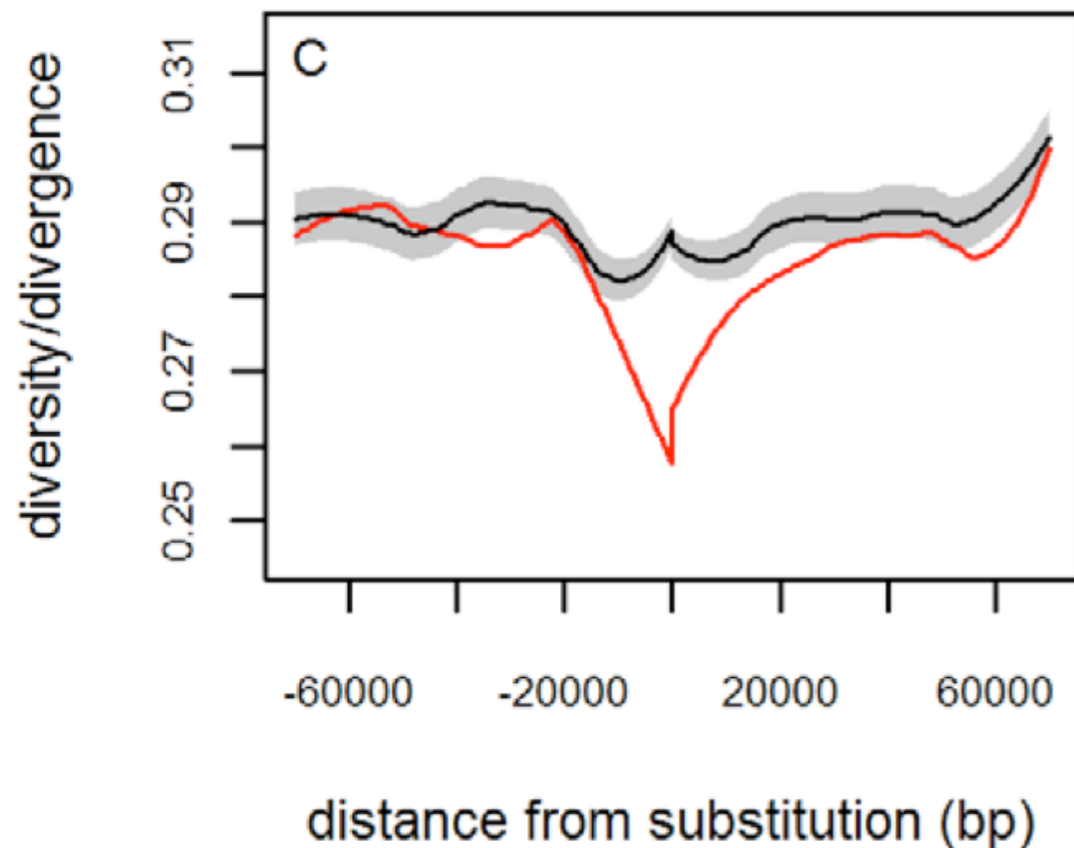
Simulation of patterns of **neutral** diversity around a 700bp **deleterious locus** with $\gamma=-5$.

Modeling the data



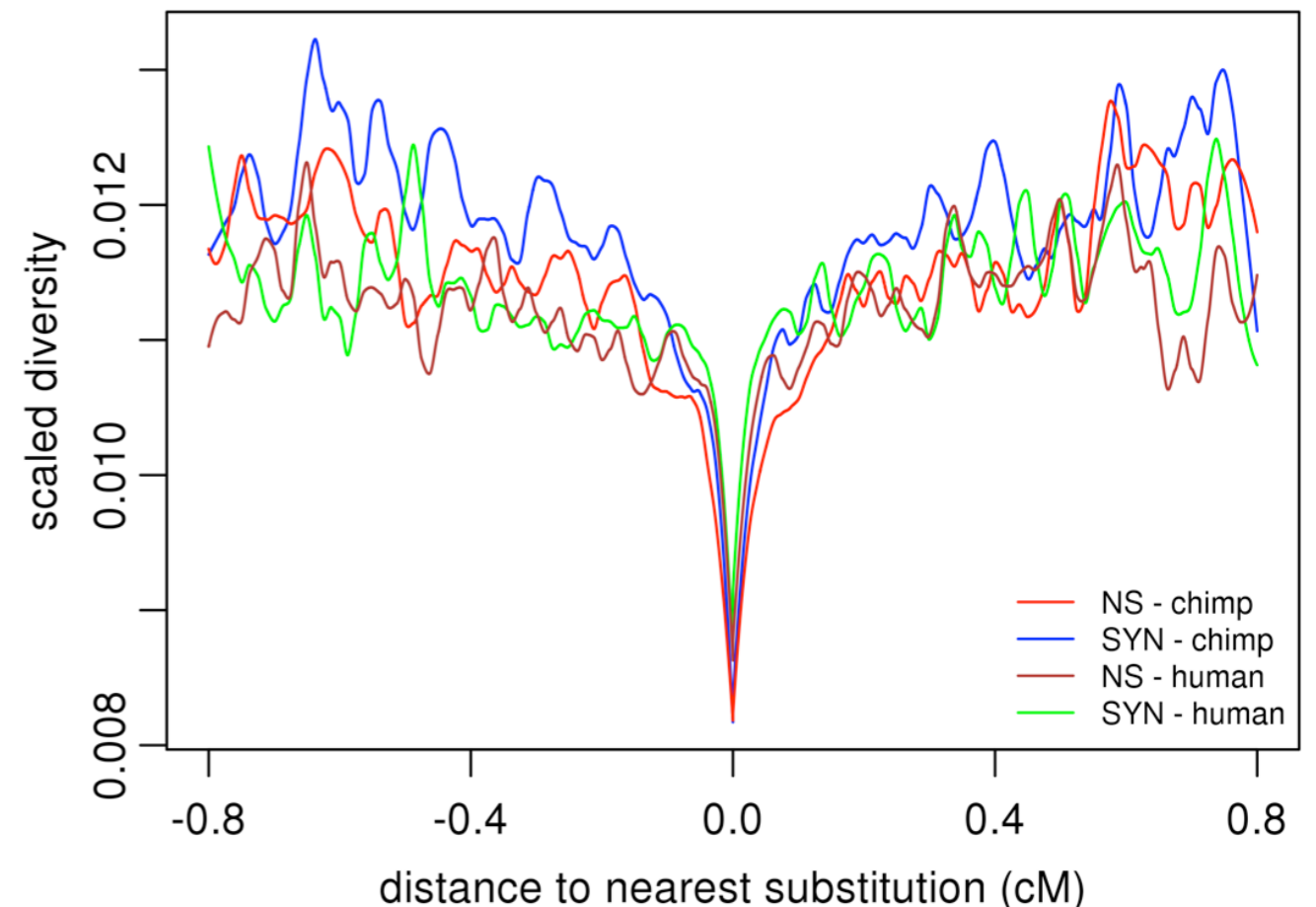
Other organisms...

Drosophila



Sattath et al (2011) estimate
~13% of amino acid
substitutions were adaptive.

Chimpanzee



PanMap Project: full genome
sequencing on 10 western
chimps yields identical results

Genetic Load

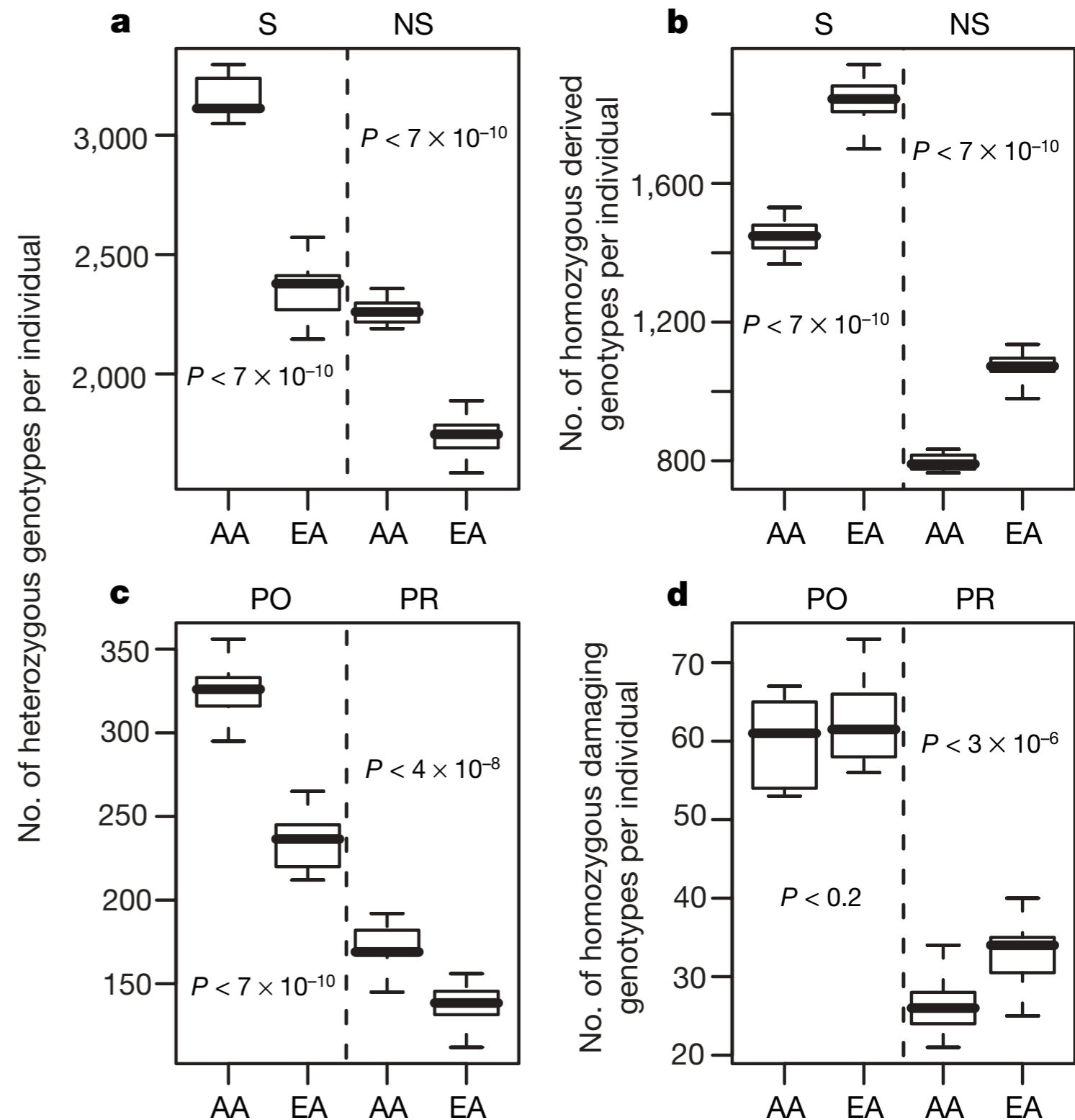
- Genetic load is the reduction in population mean fitness due to deleterious mutations compared to a (hypothetical) mutation-free population.
- Load is the outcome of the evolutionary process of a population.
- But, unlike other features of genetic variation, it cannot be directly observed.
- Must be indirectly inferred.

Inferring Genetic Load

- Empirical counting approaches:
 - Under an additive model, the number of derived deleterious alleles will be proportional to genetic load
 - Under a recessive model, the number of homozygous derived genotypes will be proportional to load

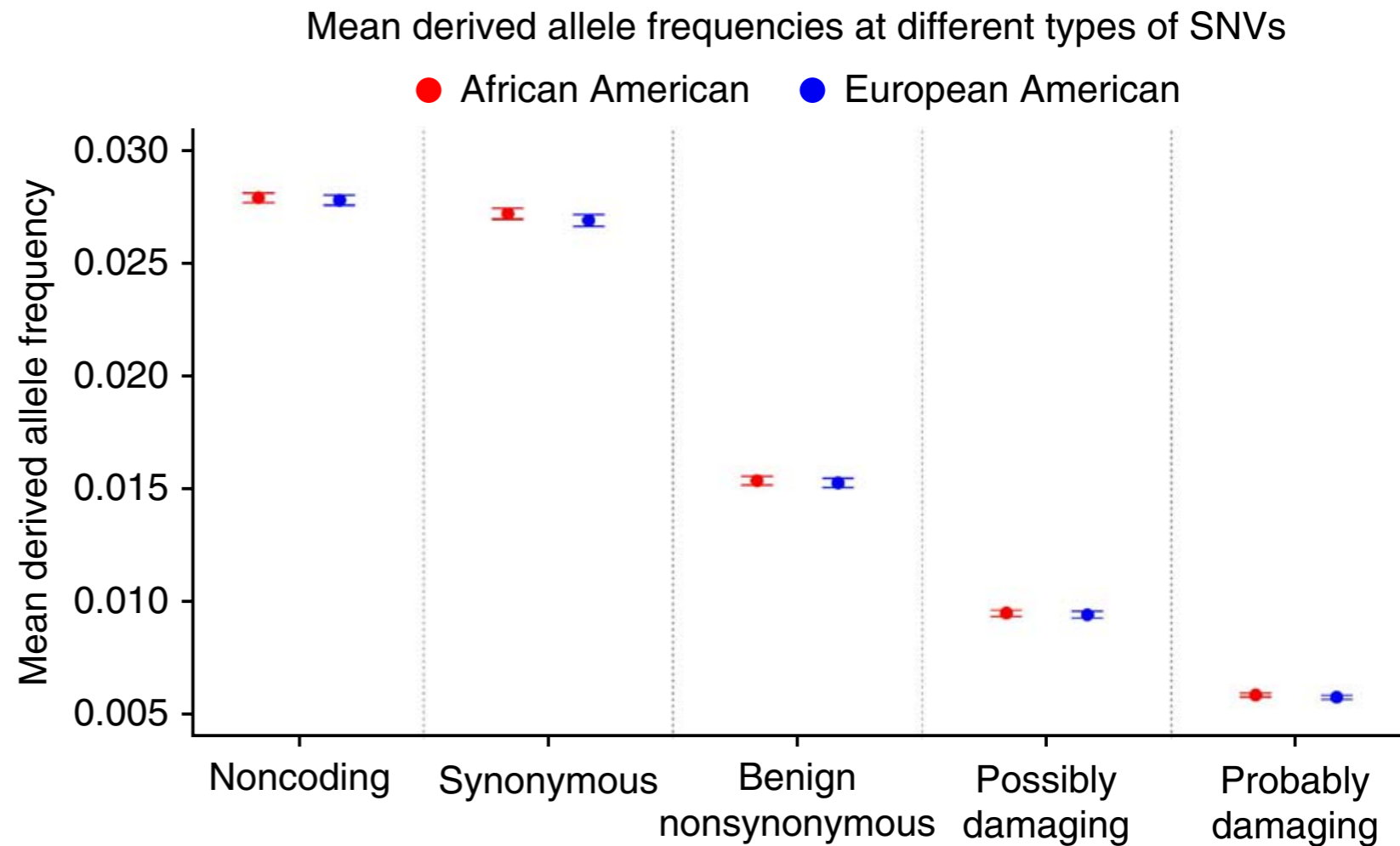
Inferring Genetic Load

- It is widely appreciated that African ancestry individuals have more variation overall than individuals with European ancestry.
- However, European individuals have more homozygous variation.
- Increased load?



Inferring Genetic Load

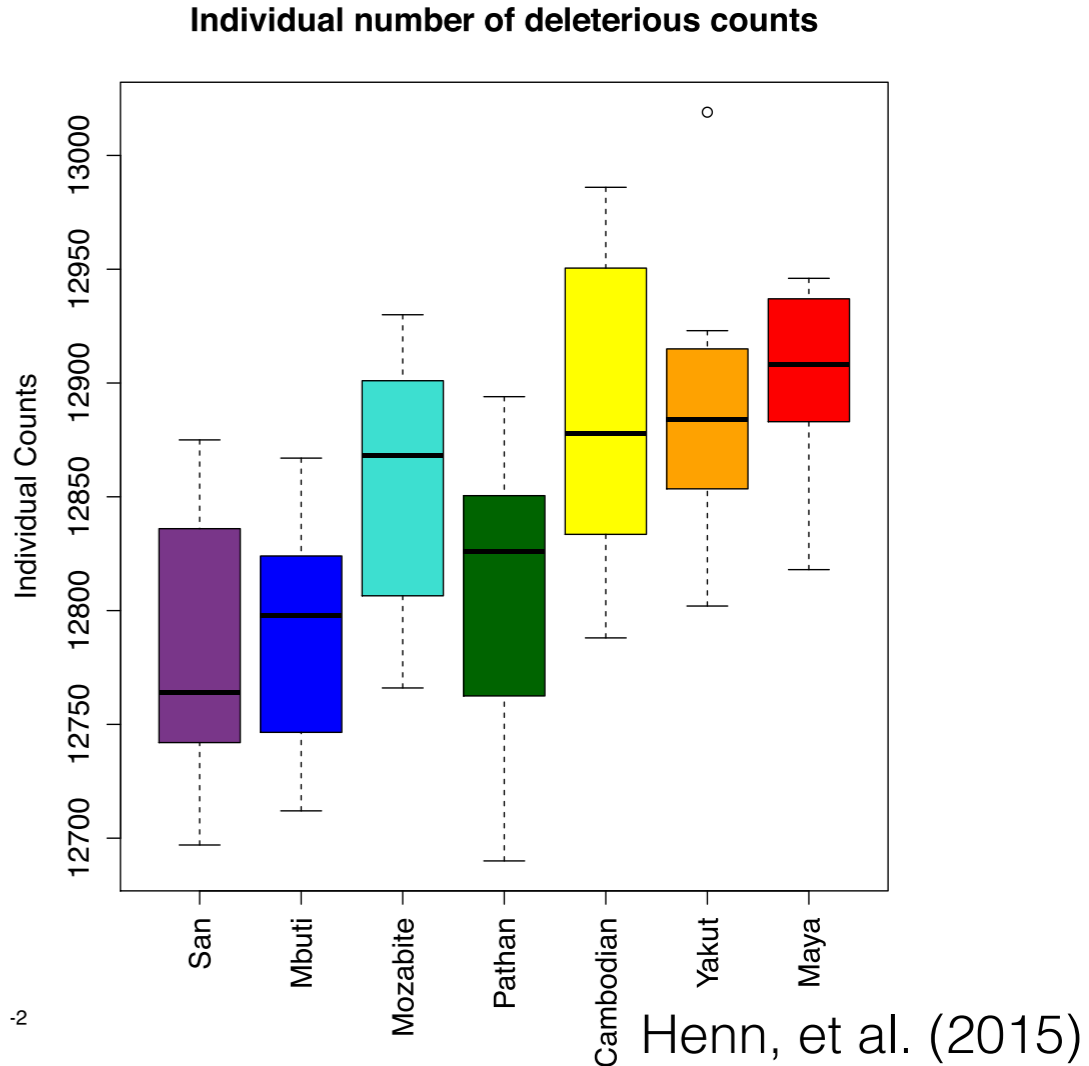
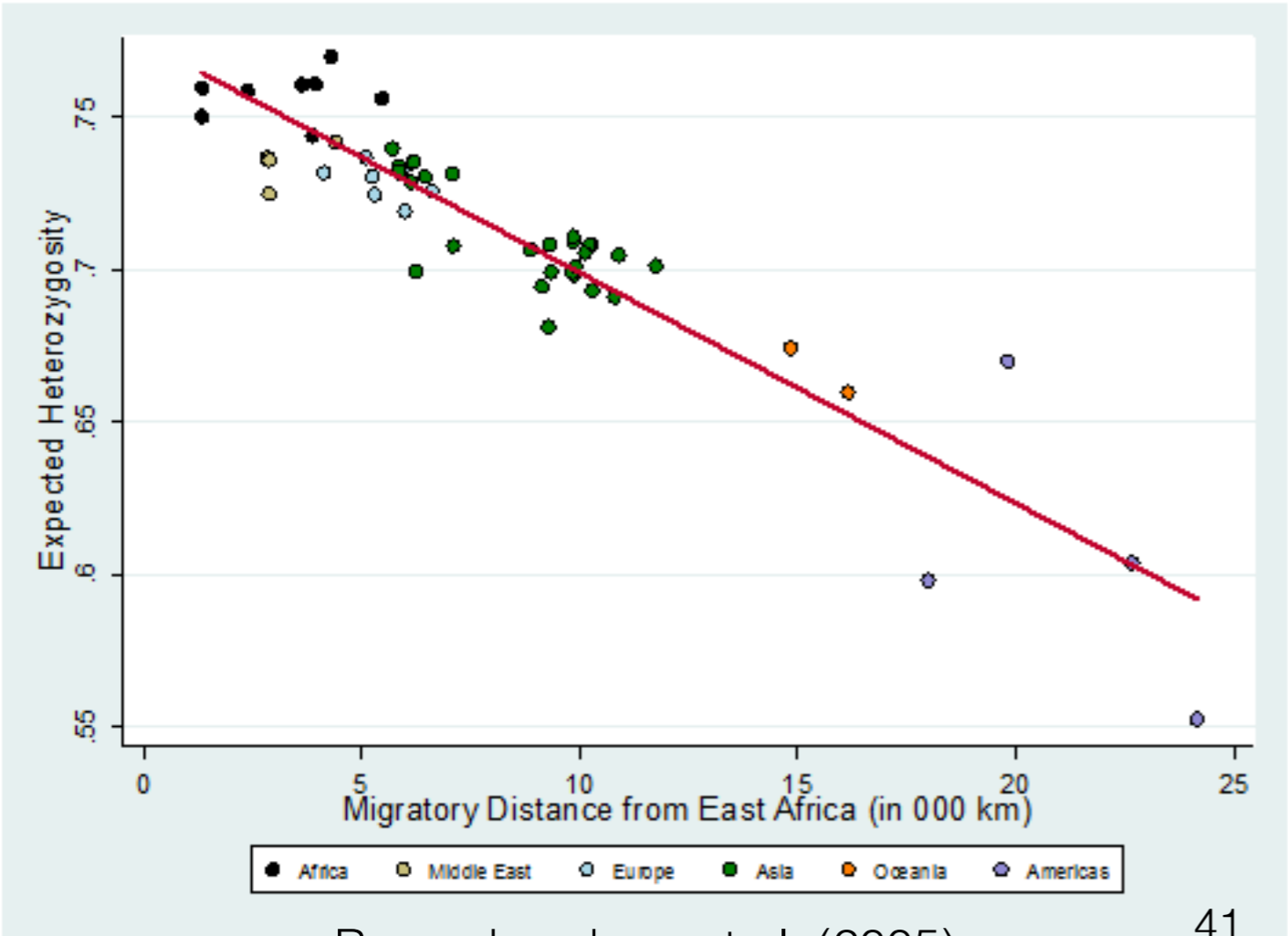
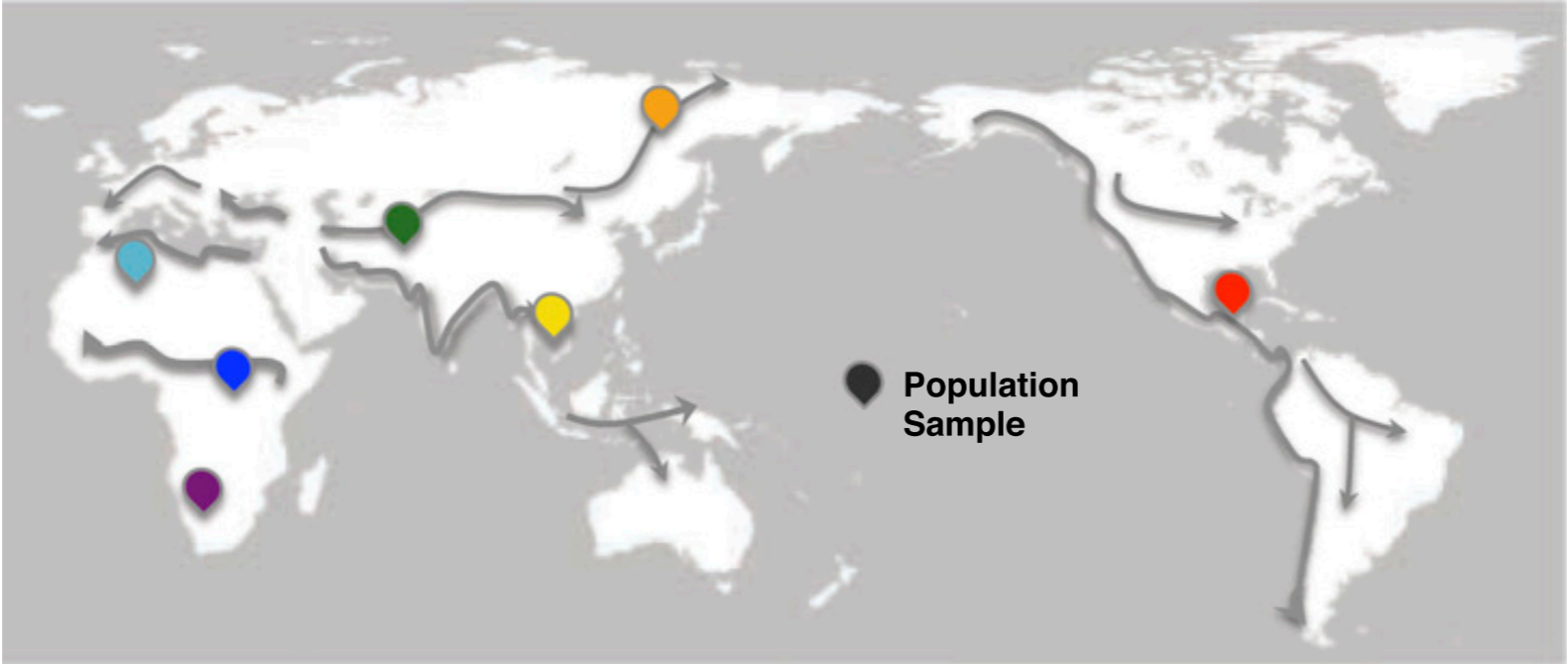
- However, het. and hom. derived alleles appear to balance between African and European Americans.



	Noncoding	Synonymous	Benign nonsynonymous	Possibly damaging	Probably damaging
Number per individual, AA:	21,421	15,401	5,373	1,695	2,002
Number per individual, EA:	21,345	15,231	5,338	1,682	1,969

- All individuals have same number of derived alleles!

Serial Founder Effects on Genetic Load



The Effect of Negative Selection

Adaptive

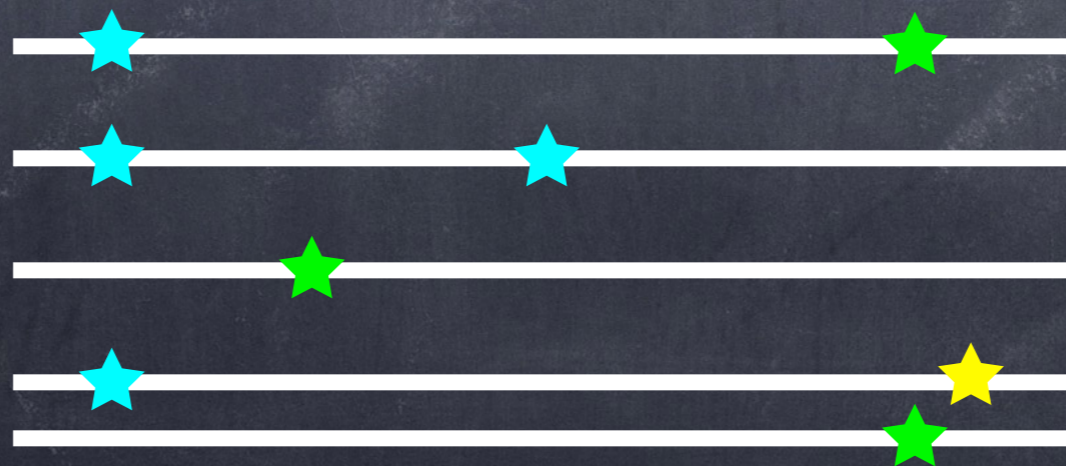
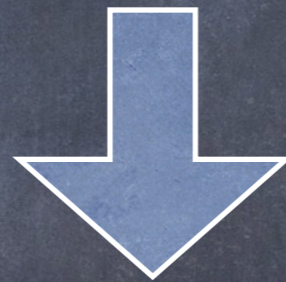
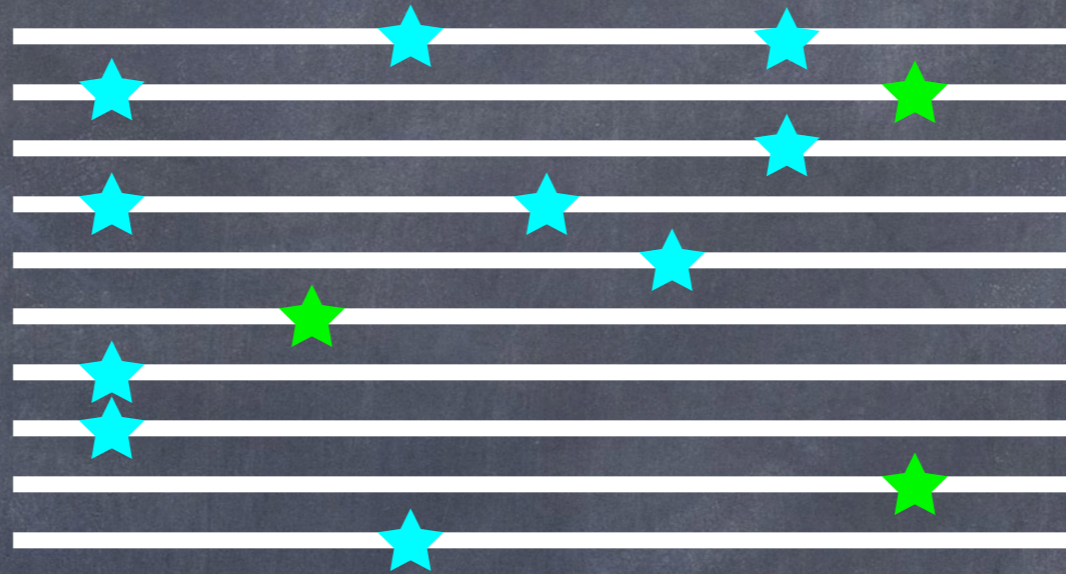
Neutral

Nearly Neutral

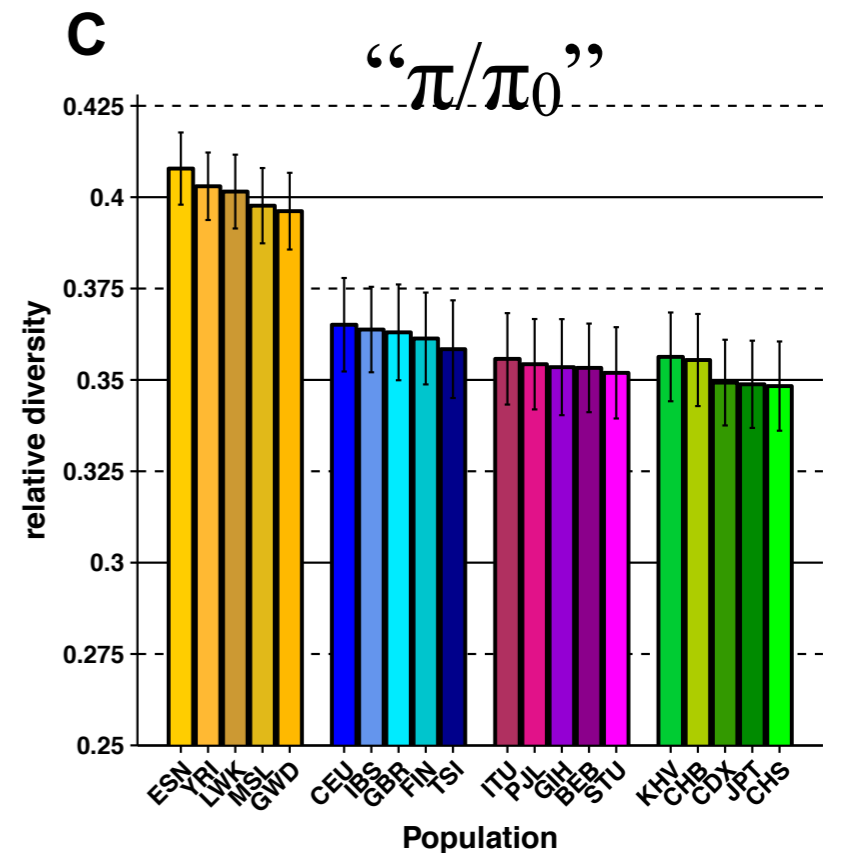
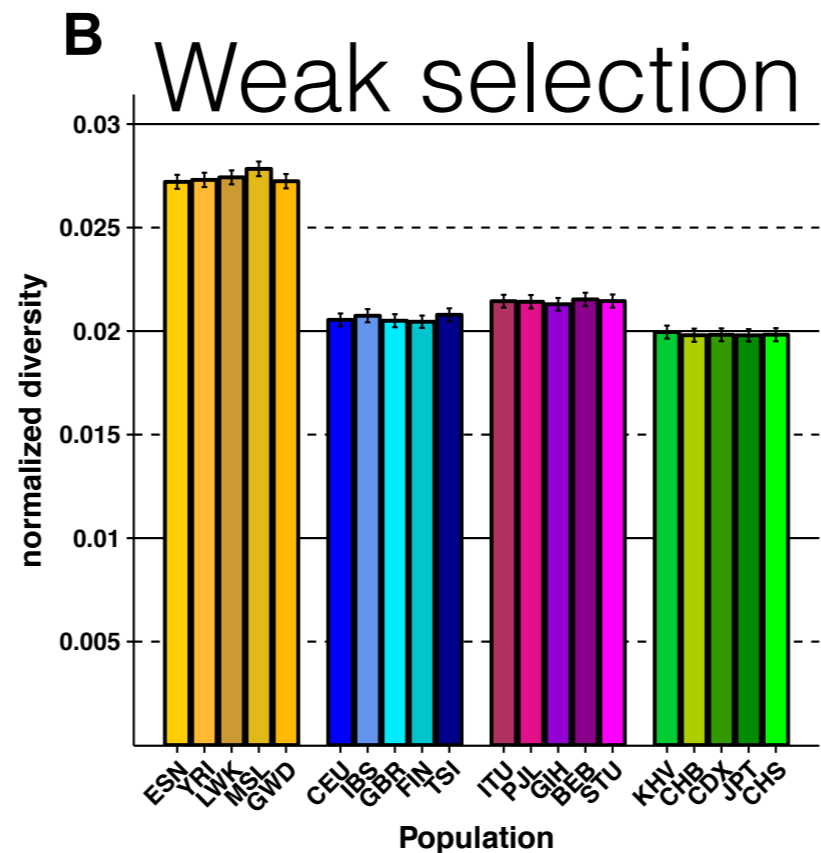
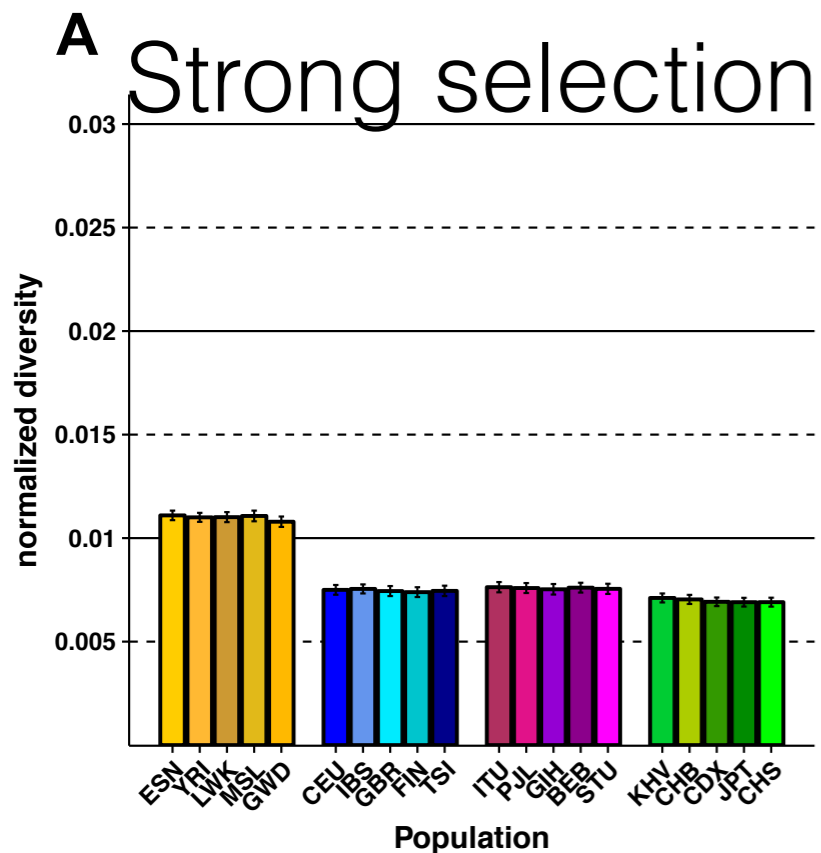
Mildly Deleterious

Fairly Deleterious

Strongly Deleterious



BGS Features



- Neutral sites in 1000 Genomes Project data: 20 non-admixed populations
- The strength of background selection varies across populations!

Background Selection & Disease?

Background selection drives patterns of genetic variation.

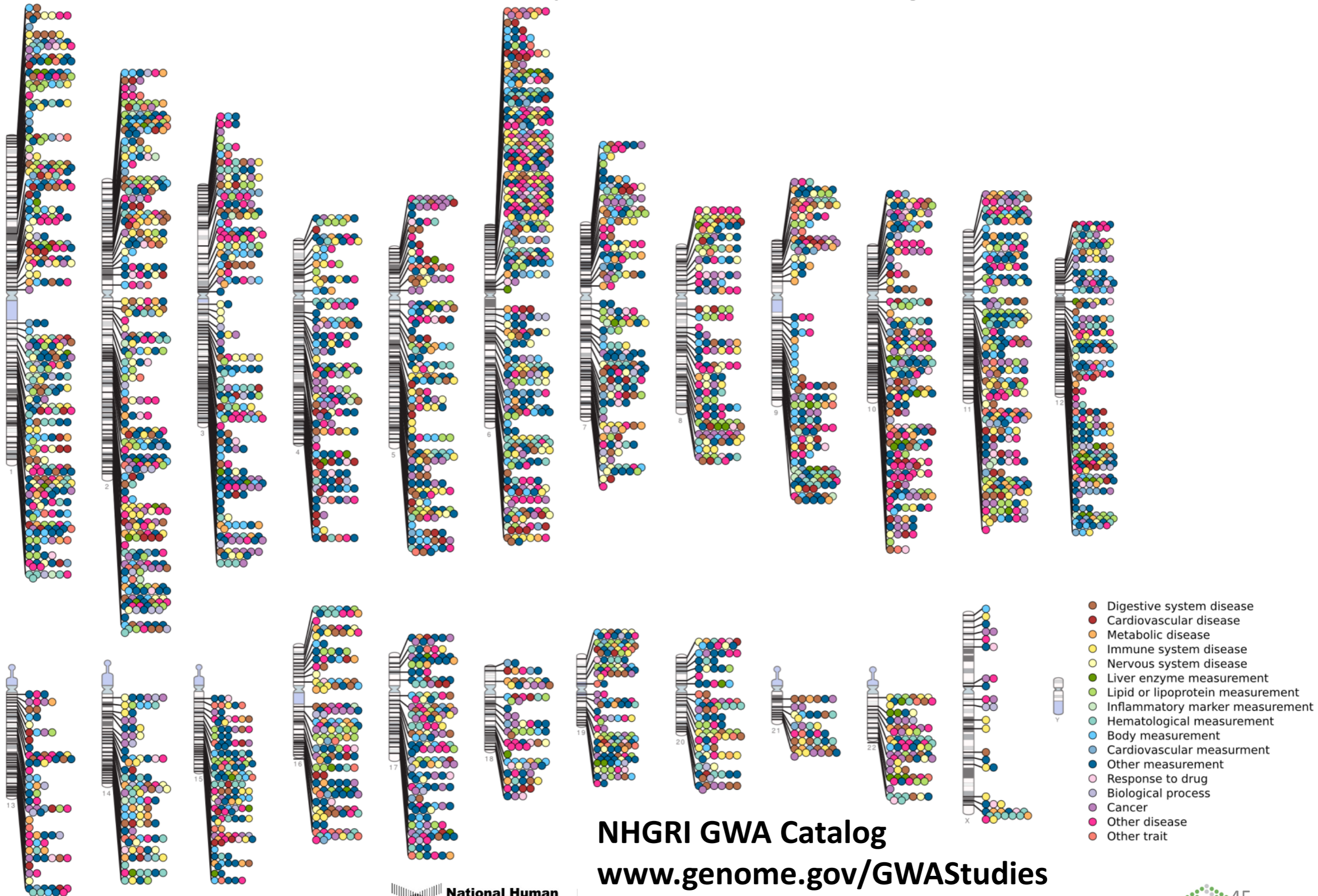
- But does it matter?
- Does it have implications for studying disease?

To find out, we looked at the NHGRI GWAS database:

www.genome.gov/gwastudies/

Published Genome-Wide Associations through 12/2013

Published GWA at $p \leq 5 \times 10^{-8}$ for 17 trait categories

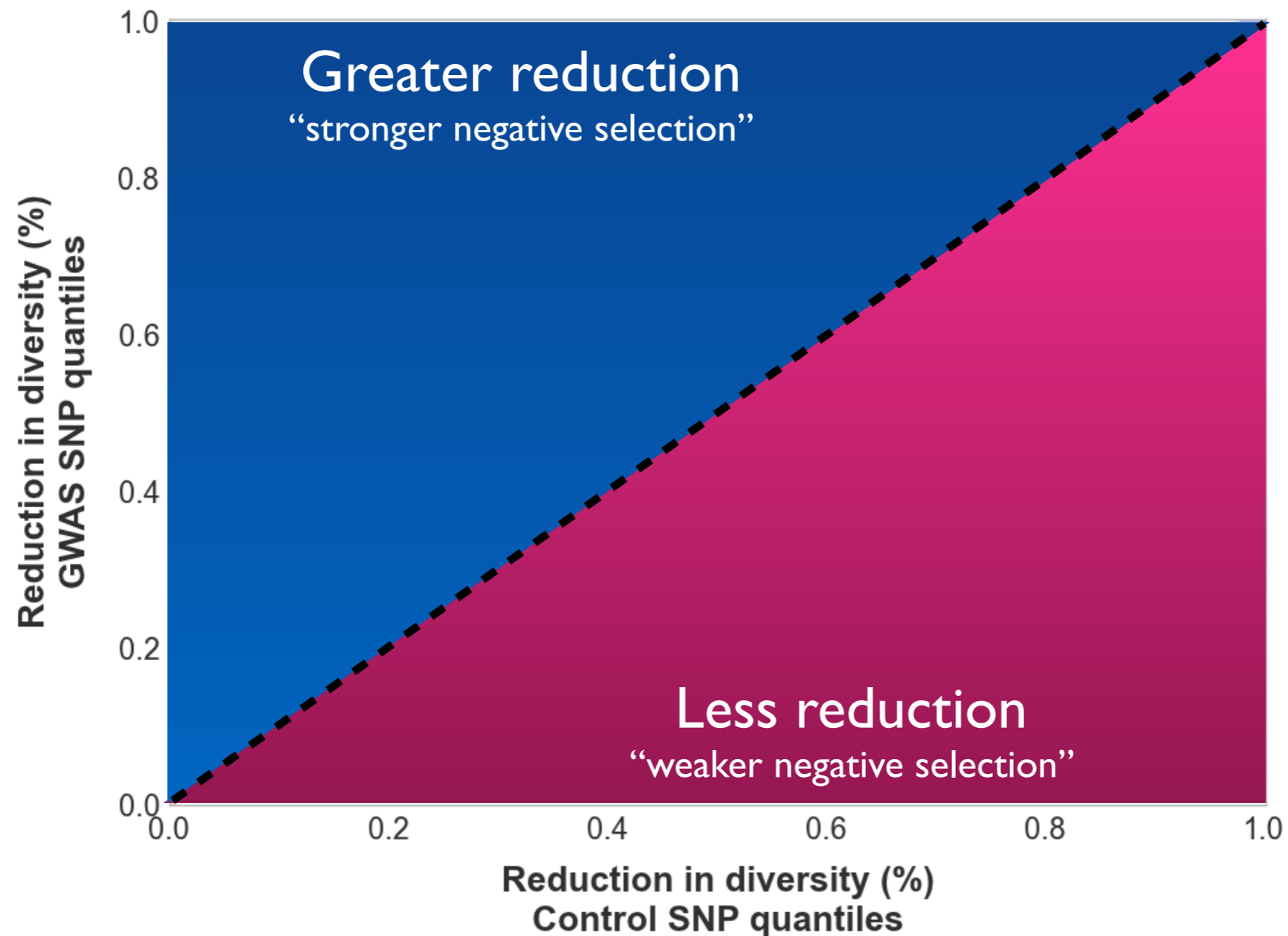


NHGRI GWA Catalog

www.genome.gov/GWAStudies

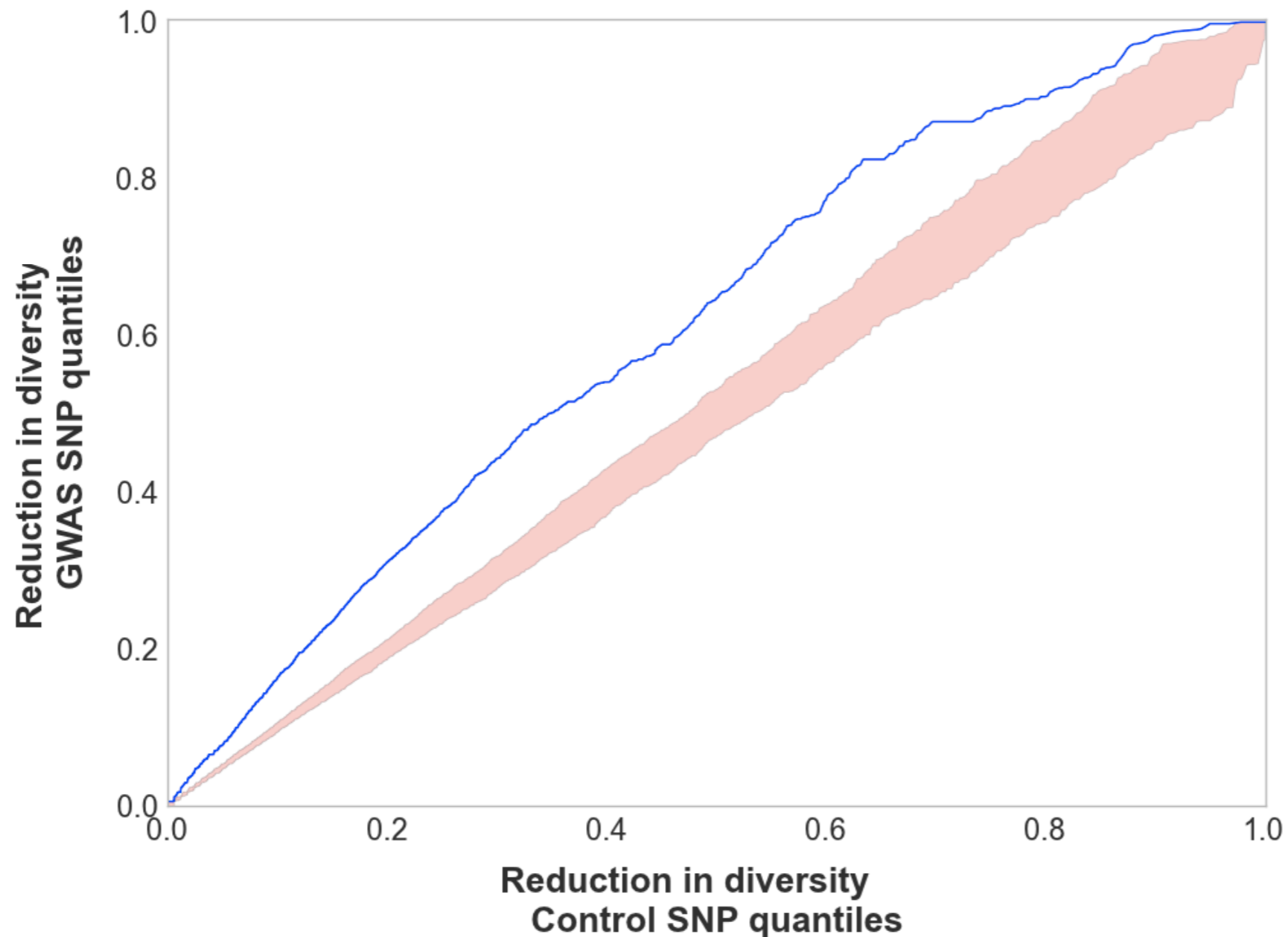
www.ebi.ac.uk/fgpt/gwas/

Effects of Linked Selection



- QQ-plot of the reduction in diversity around GWAS hits compared to background.

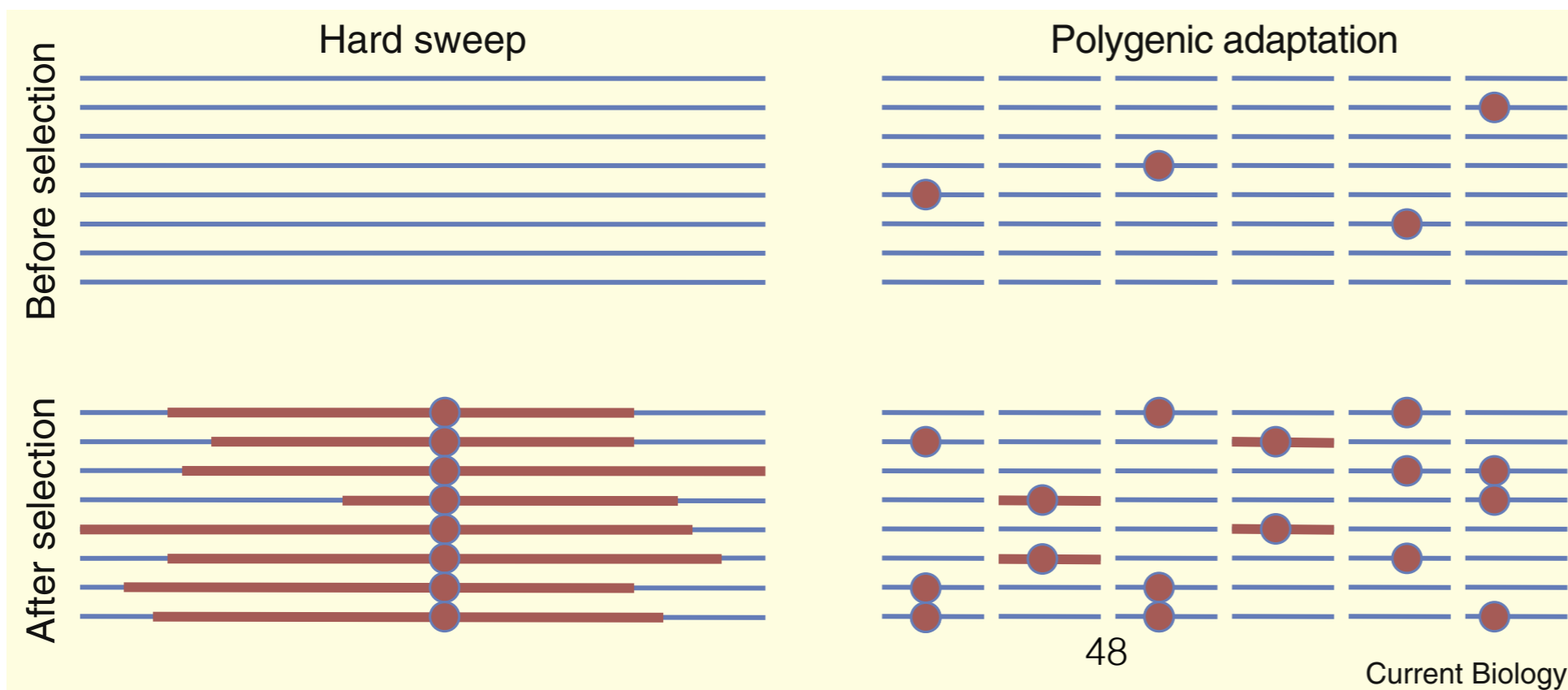
Effects of Linked Selection



- Greater reduction in diversity around GWAS hits indicates a strong, local burden of negative selection.

Implications

- Human evolution has **not** occurred by single, large-effect mutations.
- Natural selection operates on complex traits (e.g., disease), not simply mendelian genetics!
- Next: Apply detailed evolutionary models to disease!



Importance of Simulations

- When studying population genetics, simulations are an incredibly useful tool:
 - Gain insight into evolutionary processes
 - Improve null models of evolution
 - Develop “goodness of fit” tests for data
- For complex simulations, the forward simulator `sfs_code` can be a useful tool!
 - <http://sfscode.sourceforge.net/>