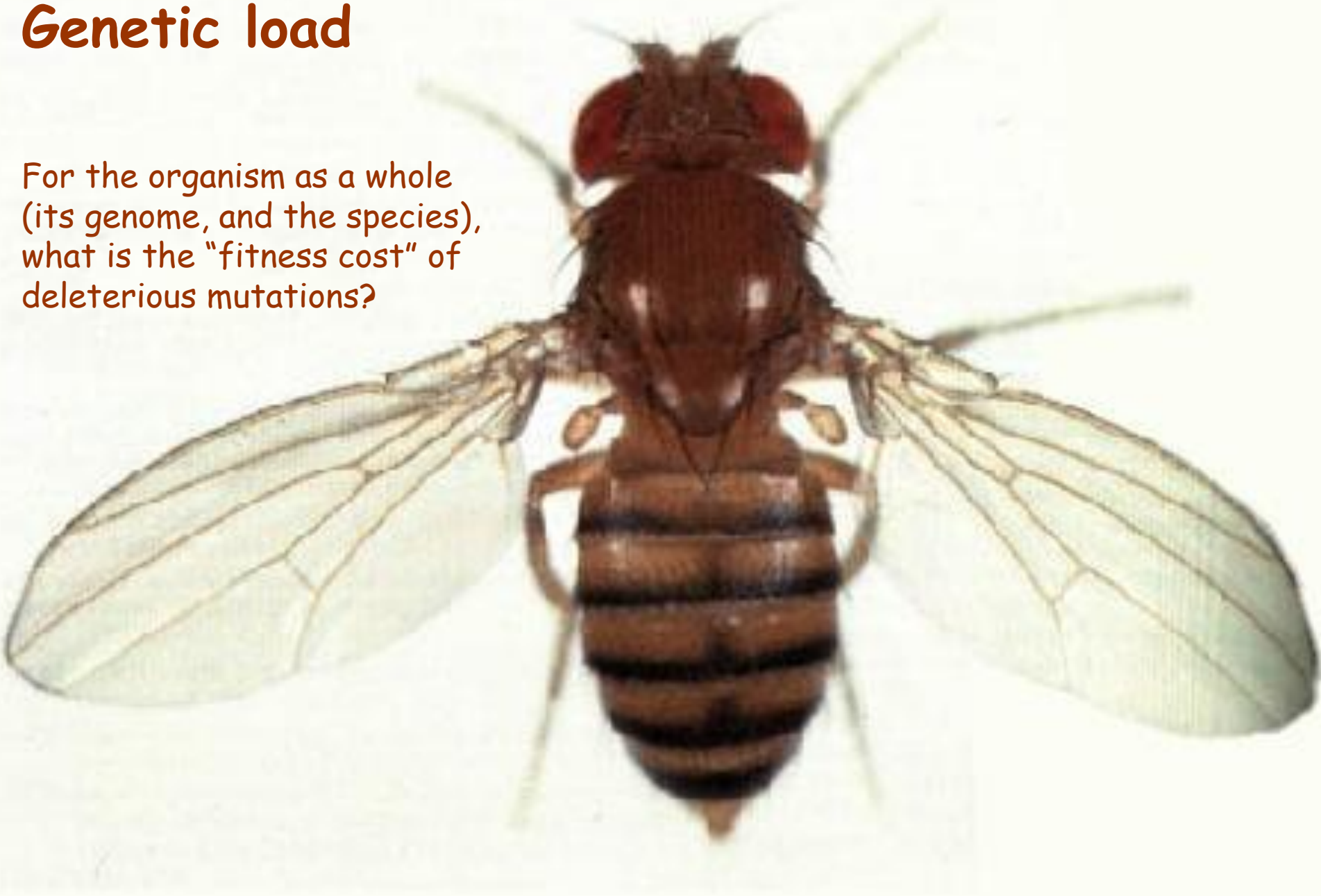
Genetic load

For the organism as a whole (its genome, and the species), what is the "fitness cost" of deleterious mutations?



We saw that the expected frequency of deleterious alleles is

$$\boxed{\hat{q} \approx \frac{u}{hs}}$$

How much is *average fitness* depressed, as a result?

$$\bar{w} = 1 - 2\hat{p}\hat{q}hs - \hat{q}^2 s$$

$$\approx 1 - 2\hat{q}hs \qquad \text{(because } q^2 \approx 0 \text{ and } p \approx 1\text{)}$$

$$\approx 1 - 2u \qquad \text{(because } qhs \approx u\text{)}$$

So the dominance (h) and the selection coefficient (s) don't matter!

Just the (diploid) rate of deleterious mutation ($2u$)!

Why?

The *load* is the reduction relative to an "unloaded" genotype.

$$L = \frac{w_{\max} - \bar{w}}{w_{\max}}$$

Thus at any given locus where the mutation probability is *u* per copy,

$$L = \frac{1 - (1 - 2u)}{1} = 2u$$

In other words, the load is equal to the diploid mutation rate!

Now, what about the genome (the organism) *as a whole*?

It depends on how the genotypes at different loci ***combine*** to determine fitness.

If each locus has an *independent* effect, then fitnesses will *multiply:*

$$\overline{W} = \prod_{i=1}^{n} \bar{w}_i \qquad = (1\text{-}2u)^n \text{ , if } u \text{ is a constant.}$$

In other words,

$$\overline{W} \approx e^{-U}$$

Where ***U*** is the ***total genomic mutation rate:***

$$U = 2\sum_{i=1}^{n} u_i$$

Thus, over the genome as a whole,

$$L = 1 - e^{-U}$$

And the load could be *crushing!* (Mean fitness approaches *zero* for $U \gg 1$, which *must* be true for mutations of very small effect!)

But the multiplicative (independent-effects) model is just one of many!

It's pretty, but not well supported by logic or evidence!

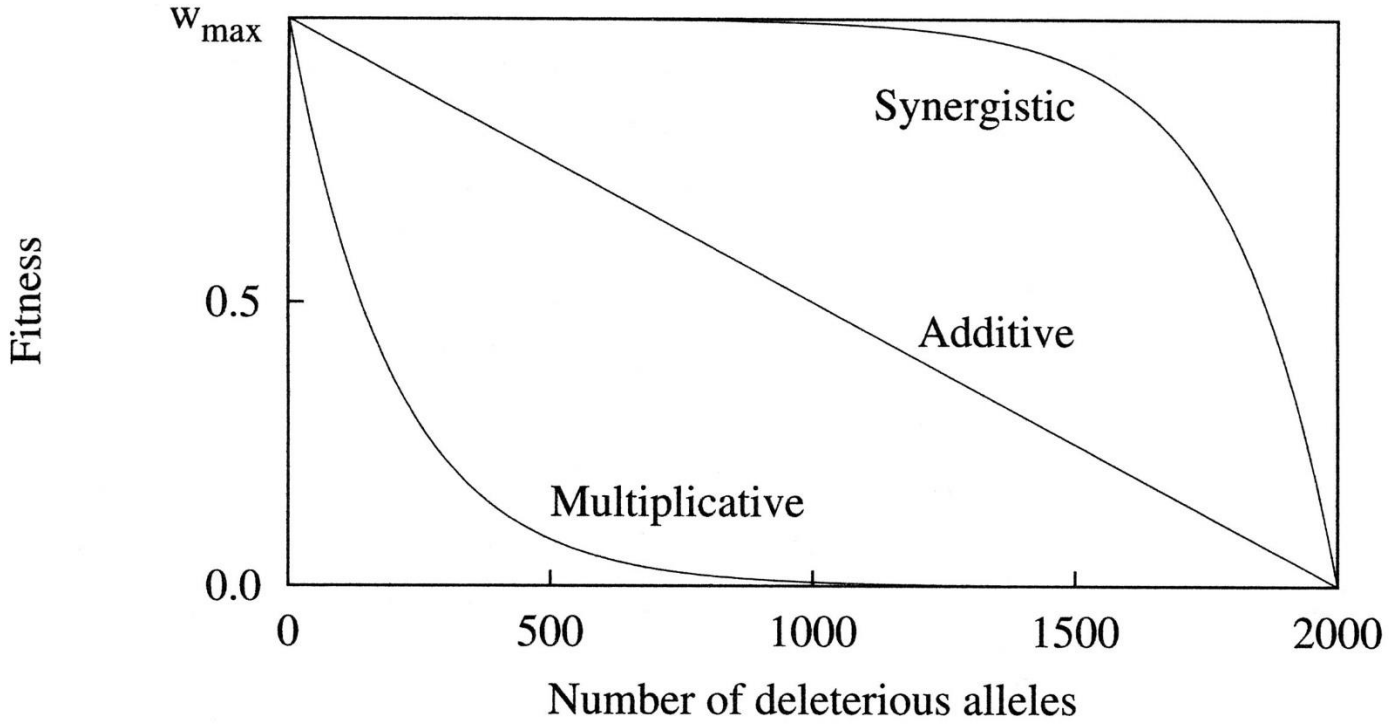


Figure 3.5: The fitness of a genotype as a function of the numbers of deleterious alleles for three models of epistasis.

Here "epistasis" means "how the effect of a bad allele at one locus depends on the number of bad alleles at other loci".

What's wrong with the multiplicative fitness model?

No biological justification for constant proportional effects of mutations.

Some data support the synergistic model (*increasing* proportional effects).

But in any case, if the mean mutation number per genome is large, then the ***fittest possible genotypes will never occur!***

The genome-wide mutation number will be ~*binomially* distributed, so most individuals will carry similar numbers and have similar fitnesses.

Gillespie's example: assume *additive* epistasis and $h = \frac{1}{2}$ (to keep the math simple).

The fitness variance contributed by each site is therefore $pqs^2/2$, and the variance contributed by n sites is $npqs^2/2$.

Suppose there are 10^9 nucleotide positions where $s = 10^{-5}$.

Then the total *fitness variance* can't exceed 0.0125 (standard deviation ≈ 10%).

Not catastrophic for the species, but what would this mean *biologically*?

Perhaps a great deal of variation in "health and happiness", *none* of it caused by genes with large effects!

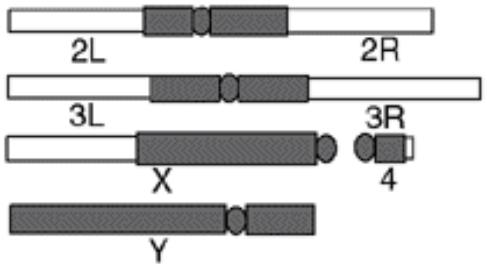
Estimating the load

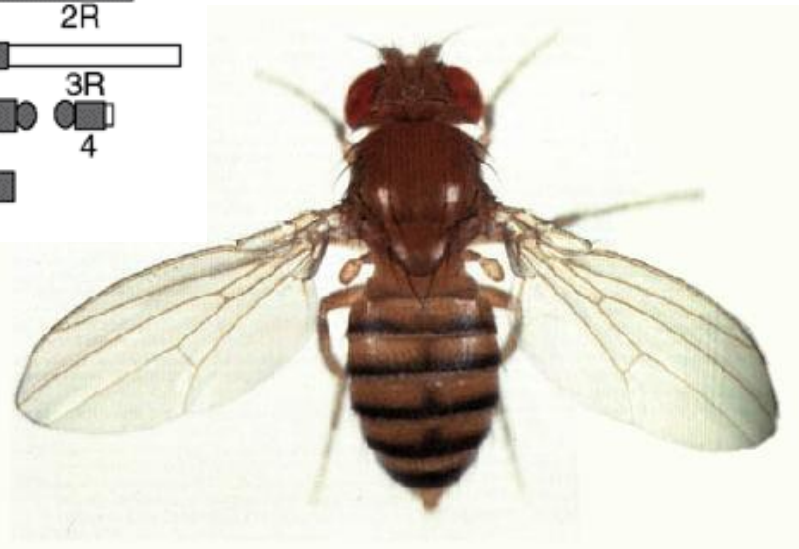
Greenberg and Crow make many *Drosophila* lines that are **homozygous** for different wild second chromosomes.

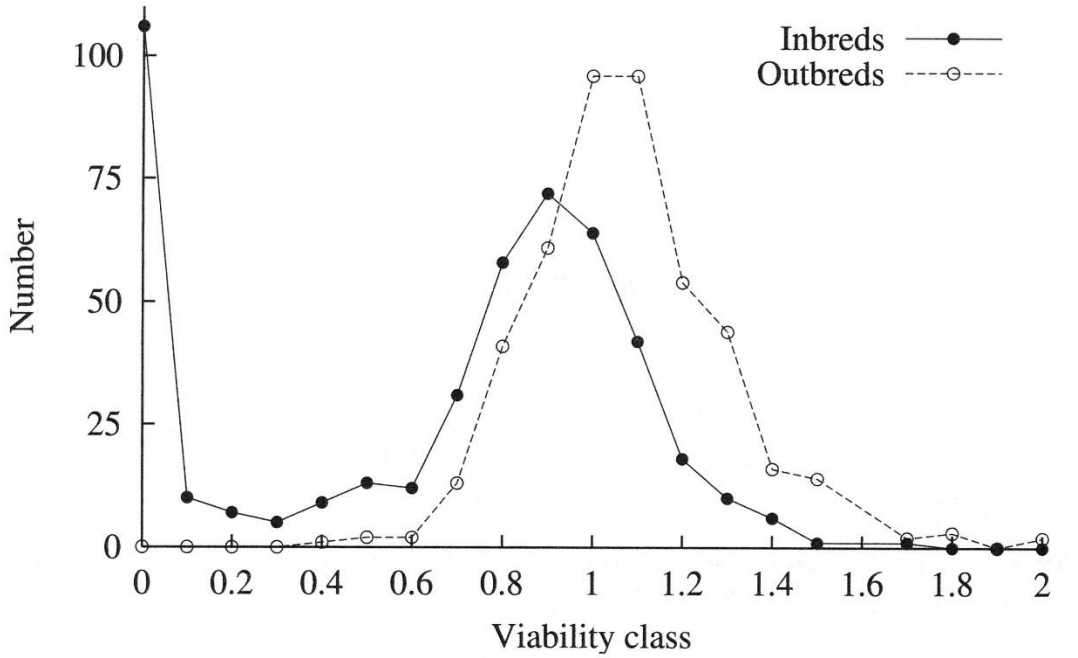
Compared to flies that are **heterozygous** for their second chromosomes, what is the distribution of the homozygotes' egg-to-adult *viabilities?*

Some chromosomes carry recessive lethal alleles (creating the zero-viability class).

And the others show a distribution shifted ~10% lower.







What does this tell us about the average degree of dominance (*h*)?

Genotype:	A_1A_1	A_1A_2	A_2A_2
Relative fitness:	1	$1-hs$	$1-s$
Inbred frequencies:	p	0	q
Outbred frequencies:	p^2	$2pq$	q^2

The mean viability of the "inbred" flies (homozygous 2nd chromosomes) is

$$(p \times 1) + [q \times (1-s)] = 1 - qs$$

The mean viability of the "outbred" flies is just W-bar. This will be greater than that of the inbred flies if

$$1 - 2pqhs - q^2 s > 1 - qs$$

which implies that

$$\boxed{h < 1/2}$$

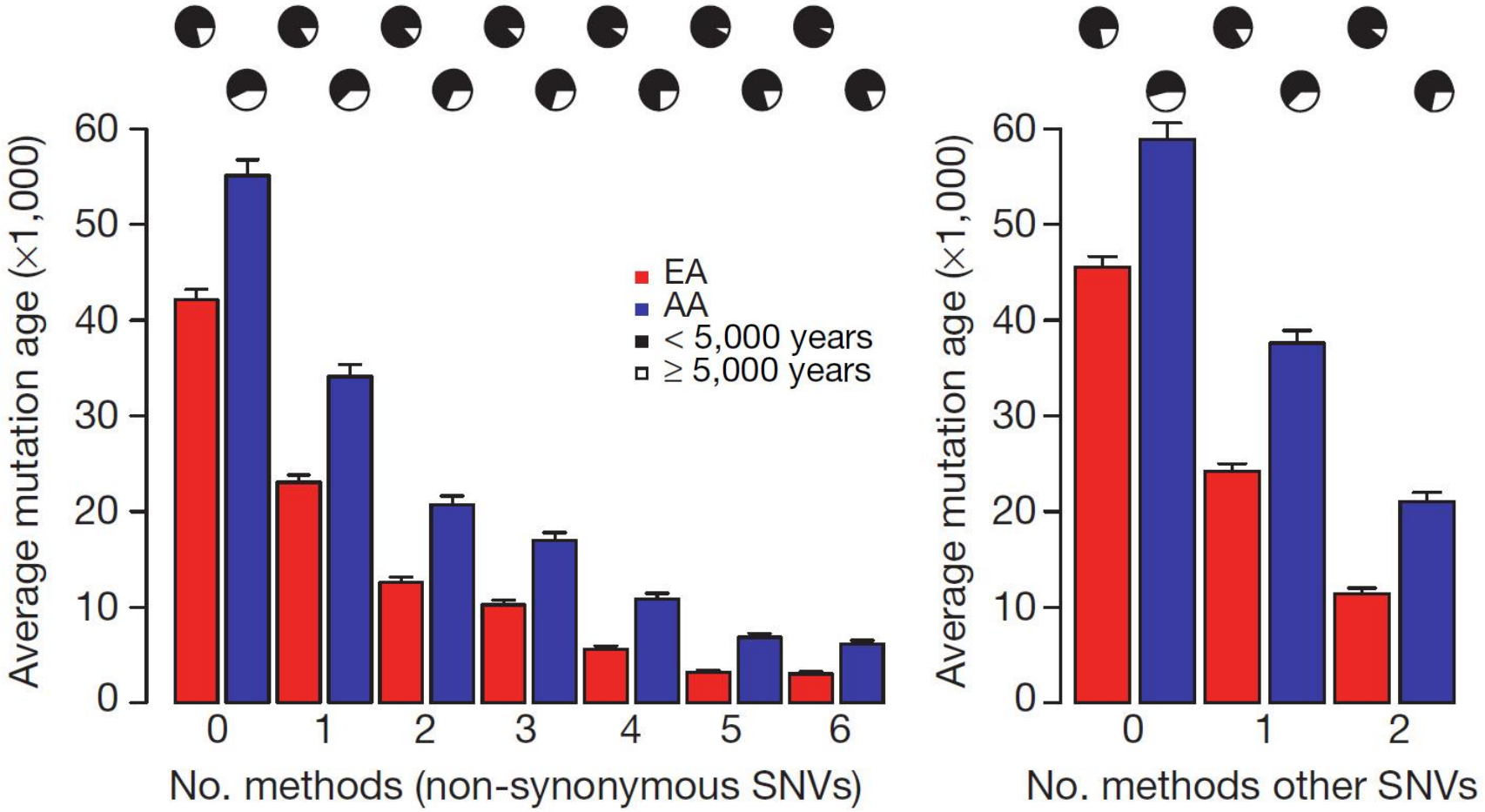
Greenberg and Crow infer that $h \approx 0$ for the lethal mutations, but that h tends to increase toward $\frac{1}{2}$ as s becomes smaller. (See Gillespie for details.)

Evidence from the ages of human mutations

Fu *et al.* surveyed single-nucleotide variants (SNVs) in 6,500 coding "exomes".

At over a million polymorphic sites, mutations become ***younger***, on average, as more "functional annotation" methods classify them as ***deleterious***.

(Pie charts show the proportions arising less than 5,000 years ago.)



Fu et al. (2013) *Nature* 493, 216-220

More evidence from the ages of mutations

Mathieson and McVean studied the ages of doubleton ("f_2") mutations in world-wide "1000 Genomes" data.

They're all young, but more so if they are annotated as harmful or even merely as "functional".

As a control, M&McV asked whether between-population f_2 variants show the same pattern.

They don't! Harmful, functional and unannotated mutants show the ***same*** distribution of ages, which appears to be the distribution of times when their populations separated.

These patterns strongly imply that ***many*** mutations in ***many*** genes are viable, but ***deleterious***.

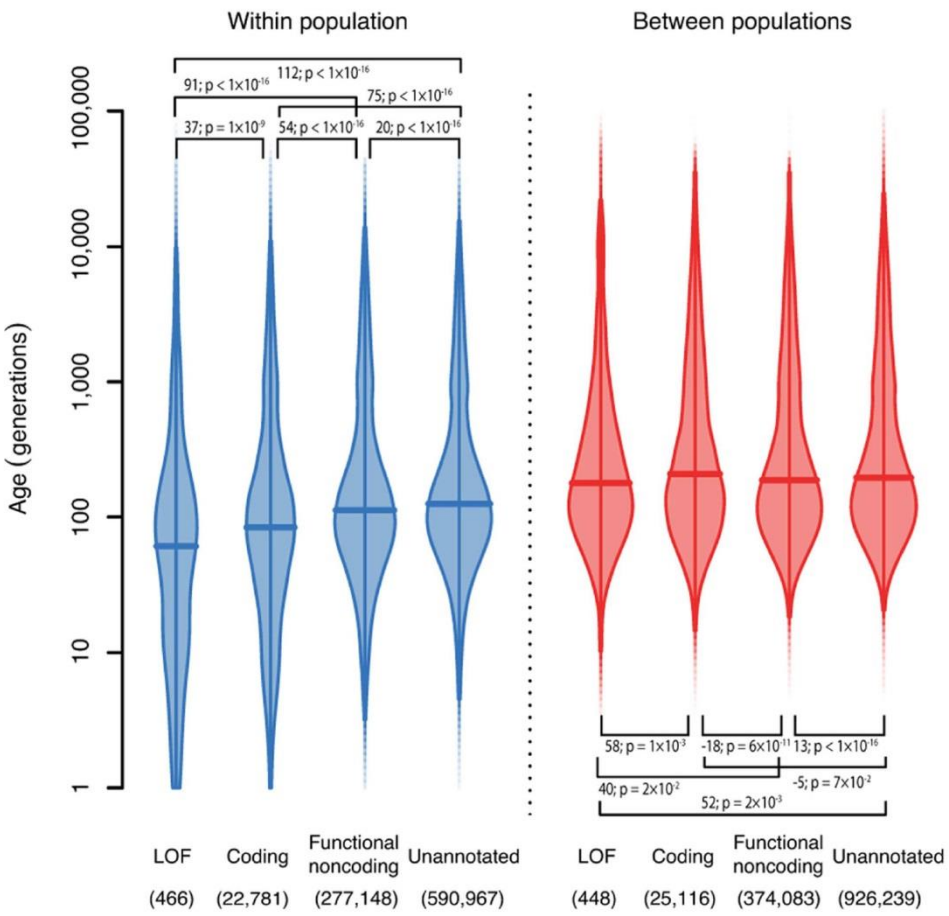


Figure 4. The ages of haplotypes around f_2 variants with different functional annotations. Density is indicated by the width of the shape, and horizontal bars show the median. We show separately the densities for f_2 variants shared within a population (left, blue), and f_2 variants shared between populations (right, red). Numbers in brackets show the number of variants in each class. Bars show the pairwise differences in means, and t test p-values for a difference in log means between groups.
doi:10.1371/journal.pgen.1004528.g004

Mathieson and McVean (2014) *PLoS Genetics*

Upshot:

We *all* may have millions of pre-existing conditions.

But good news: they're survivable!