When bad things happen to good genes: mutation *vs.* selection

Selection tends to increase the frequencies of alleles with higher marginal fitnesses.

Does this mean that genes are all maintained nearly "perfect"?

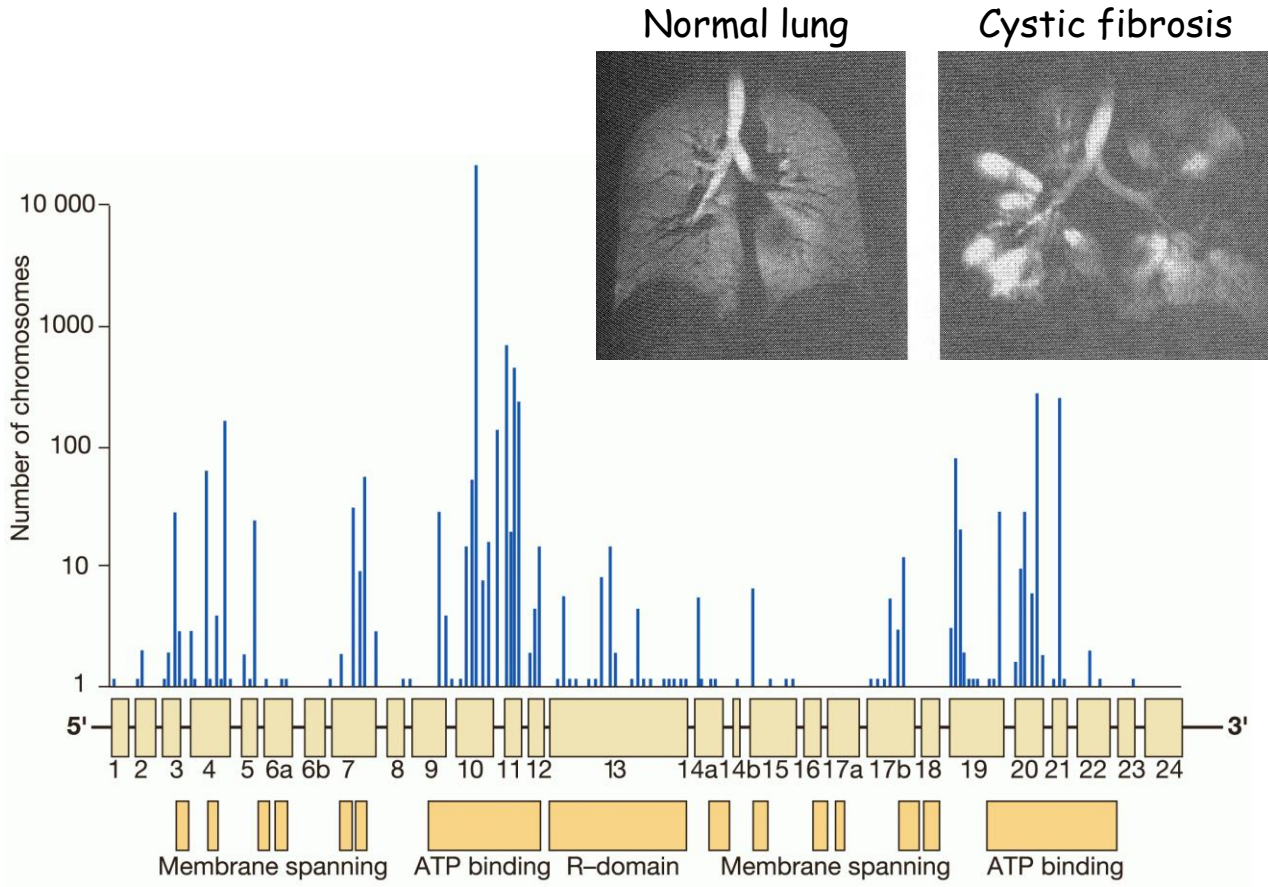
No, ***mutation*** can maintain damaged (even lethal) alleles at remarkably high frequencies, even in very large populations, especially if the mutations are recessive.

And in small populations, deleterious alleles may even fix!

More than 500 loss-of-function mutations have been identified in the cystic fibrosis transmembrane conductance regulator gene (CFTR).

People who carry two of these mutations (*i.e.*, no "wild-type" allele) develop cystic fibrosis.

Collectively, these loss-of-function mutations comprise $q = 0.02$ of all alleles in populations of European ancestry.



Can high frequencies of deleterious alleles be maintained by mutation?

Yes, if the bad alleles are ***fully recessive***.

Let **A** be the normal ("wild-type") allele, and let **a** be the class of deleterious recessive mutations that arise at a rate μ per generation.

Let the (collective) frequency of **a**-type mutations be q. Then the expected **change** in q that occurs each generation due to **mutation** is the frequency of **A** times μ:

$$\Delta q_{mut} \approx (1-q)\mu$$

The change in q due to **selection** is $\Delta q_{sel} = qp(\bar{W}_a - \bar{W}_A)/\bar{W}$

This is close to $-sq^2$, assuming $p \approx 1$ (because the **a** allele is rare), $W_A = 1$ and $W_{aa} = 1-s$.

The total change due both to mutation and selection is therefore

$$\Delta q = \Delta q_{mut} + \Delta q_{sel} = (1-q)\mu - sq^2$$

Mutation tends to *increase* q, and selection tends to *decrease* q. When these two processes or "forces" are in balance, there will be no net change in the frequency of **a**.

Setting $\Delta q = 0$ and solving for q, we get the equilibrium frequency $\hat{q} = (\mu/s)^{1/2}$

For example, suppose $\mu = 10^{-6}$ (on the low end of per-gene, per generation mutation rates). Then even if **a** is *lethal* in the homozygous state ($s = 1$), its equilibrium frequency will be $\hat{q} = 0.001$ (one person in 500 is a heterozygous carrier). And if $\mu = 10^{-4}$, then $\hat{q} = 0.01$.

What if the deleterious mutation is less than *fully* recessive?

Gillespie's way of writing the general selection model:

$$\Delta_s p = \frac{pq[p(w_{11} - w_{12}) + q(w_{12} - w_{22})]}{p^2 w_{11} + 2pqw_{12} + q^2 w_{22}}$$

Rewrite using relative-fitness notation:

Genotype:	A_1A_1	A_1A_2	A_2A_2
Relative fitness:	1	$1 - hs$	$1 - s$

$$\Delta_s p = \frac{pqs[ph + q(1 - h)]}{\bar{w}}$$

Expand and then *approximate* by assuming $q \approx 0$ ($p \approx 1$) and $h \gg q$.

Terms with q^2 disappear.

1 dominates anything with q.

$$\Delta_s p = \frac{pqs[ph + q(1 - h)]}{1 - 2pqhs - q^2 s} \approx qhs$$

That's the model of *selection* for $h > 0$ (where the mutant is partly dominant).

Now add the model of mutation:

$$A_1 \underset{v}{\overset{u}{\rightleftharpoons}} A_2$$

Thus

$$p' = p(1-u) + (1-p)(v) \quad ,$$

due to mutation. But again we can assume $p \approx 1$ and *approximate*:

$$p' \approx p - pu$$

So that

$$\Delta_u p \approx p' - p \approx -pu \approx \boxed{-u} \; .$$

At equilibrium, the effects of mutation and selection *offset each other* exactly,

$$0 = \Delta_u p + \Delta_s p$$
$$\approx \boxed{-u} + \boxed{qhs}$$

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

(generally much lower than for fully recessive mutations).

Some examples

Assume $u = 10^{-6}$

	$h = 0$	$h = 0.1$	$h = 0.5$
$s = 0.001$	$\hat{q} = 0.0316$	$\hat{q} = 0.01$	$\hat{q} = 0.002$
$s = 0.01$	$\hat{q} = 0.01$	$\hat{q} = 0.001$	$\hat{q} = 0.0002$

Deleterious mutation rates can be very high

Spinal muscular atrophy is caused by homozygosity for recessive mutations of the telomeric survival motor neuron gene (*telSMN*).

Seven of 340 patients were found to carry new mutations not present in either of their parents.

This implies a mutation rate of around 1×10^{-4}.

The selection coefficient (s) for homozygotes is estimated to be around 0.9.

The predicted allele frequency at mutation-selection equilibrium is therefore $\hat{q} = (0.0001/0.9)^{\frac{1}{2}} \approx 0.01$.

In Caucasian populations, roughly 1 in 10,000 infants is affected.

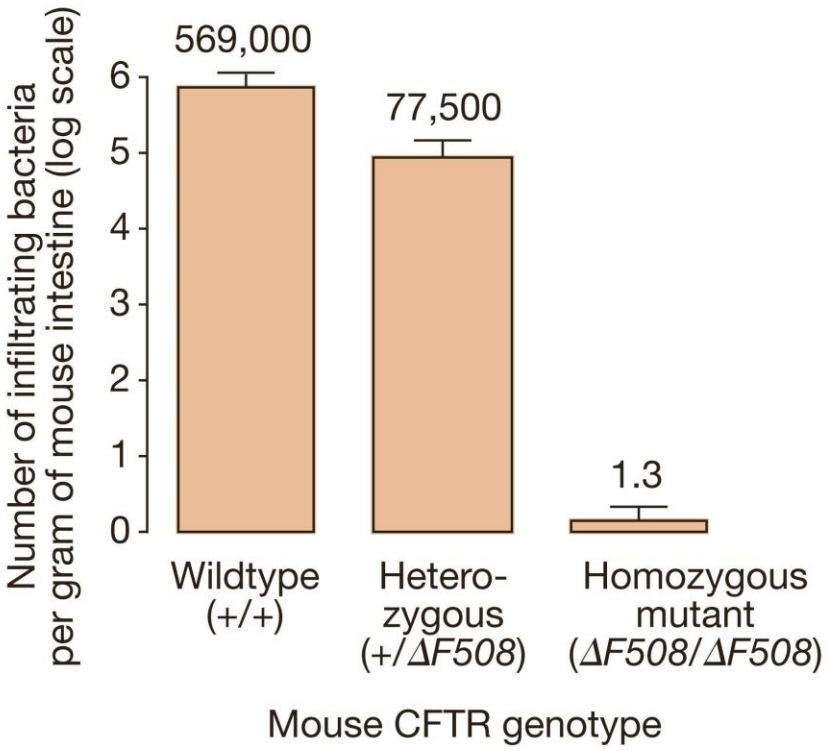
This implies $q^2 = 1/10{,}000 = 0.0001$, or $q = 0.01$, in excellent agreement with the mutation rate estimated directly from pedigrees and our simple model of mutation-selection equilibrium!

But the model *fails* for CFTR, where the deleterious allele frequency is *higher* ($q \approx 0.02$) and the estimated mutation rate is *lower* ($\mu \approx 7 \times 10^{-7}$).

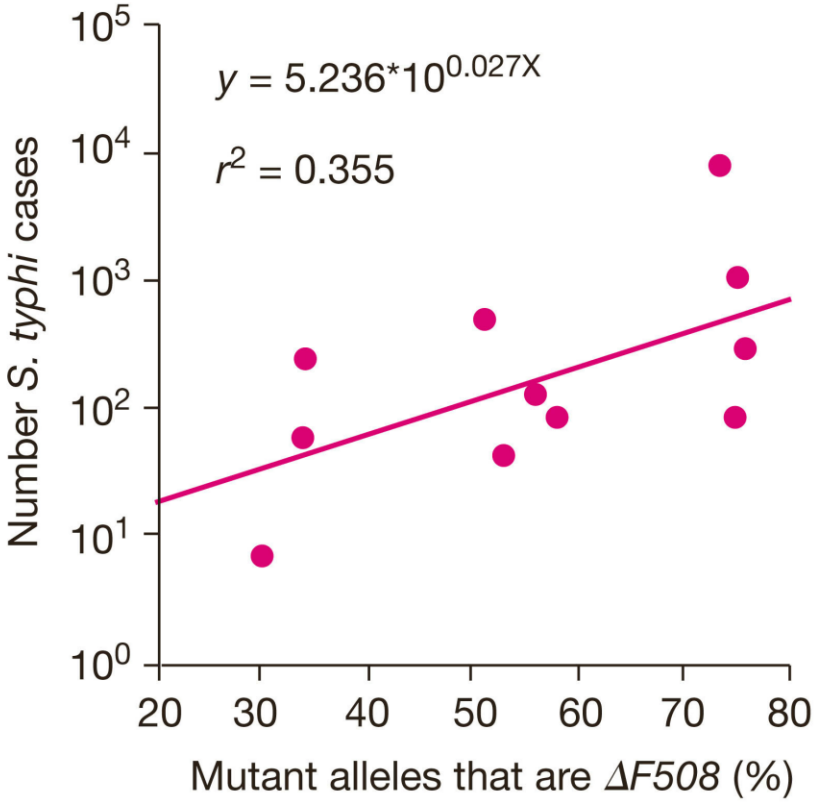
Why are the CF alleles so frequent? One hypothesis is that heterozygotes resist some disease, perhaps typhoid fever (*Salmonella enterica typhi*).

Some evidence that (weakly) supports the typhoid-fever hypothesis

Cultured mouse cells resist infiltration by *Salmonella typhi* if they are heterozygous (or homozygous) for the most common human CFTR mutation.



European countries with larger numbers of *S. typhi* outbreaks have higher relative proportions of the most common CFTR mutation (*ΔF508*).



Estimates of deleterious amino-acid substitutions in humans, from genome sequencing

Several methods have been developed to identify "deleterious mutations" by:

(1) their polymorphism at sites that are highly conserved in other mammals;

and (2) their predicted disruptive effects on protein structure.

Chun & Fay (2009) applied three of these methods to three individual human genome sequences with full reporting of heterozygous sites.

The methods estimate 500-1000 apparently deleterious mutations/genome.

Most occur in more than one of the genomes (*i.e.*, are fairly common).

However, they are enriched among singletons, and depleted on the X, consistent with the hypothesis that they really are deleterious, on average.

	Rare alleles		Common alleles[a]	
Genome	Tested	Deleterious[b]	Tested	Deleterious[b]
J. Craig Venter	766	213 (28%)	3066	583 (19%)
James D. Watson	940	303 (32%)	2632	513 (19%)
Han Chinese	703	186 (26%)	3438	651 (19%)

[a]Common alleles include those that are shared between any of the three genomes.
[b]Percent deleterious is shown in parentheses.

Genome Research 19, 1553-1561

1.1 million protein-coding single-nucleotide variants (SNVs) from 6,515 human exomes – mostly very young!

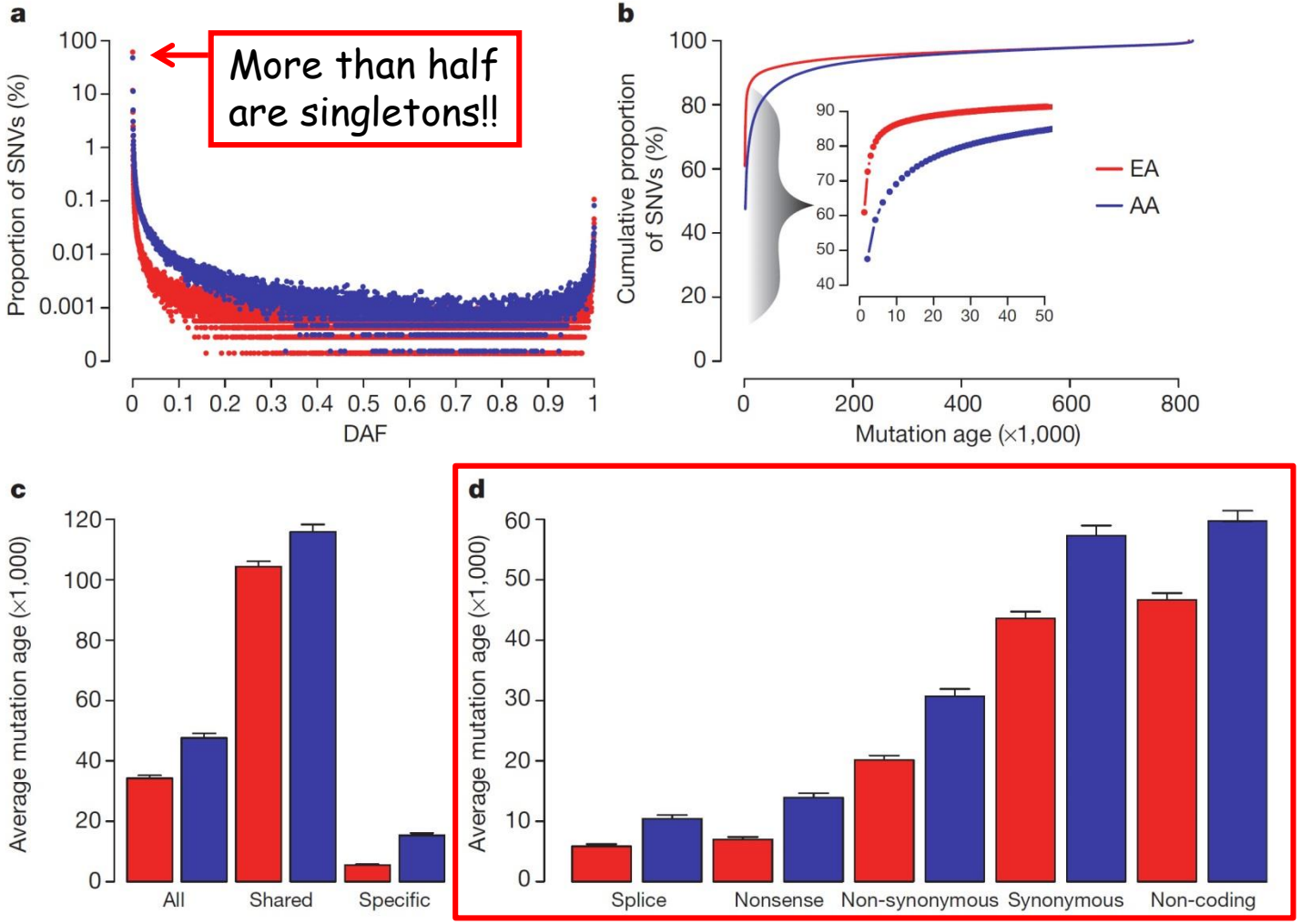
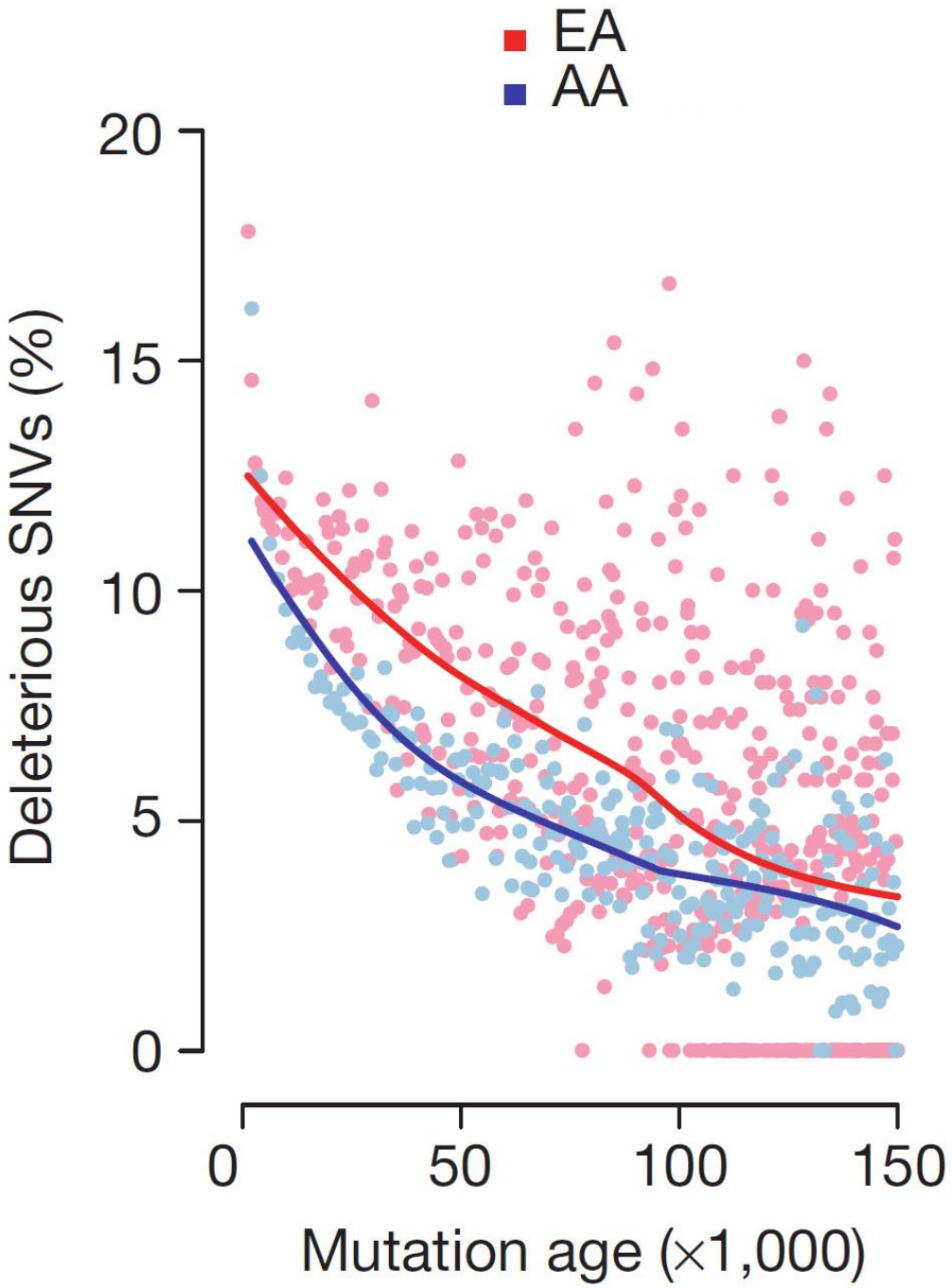


Figure 1 | The vast majority of protein-coding single-nucleotide variants arose recently. **a**, The site-frequency spectrum for European Americans (EAs, red) and African Americans (AAs, blue). DAF, derived allele frequency. **b**, Cumulative proportion of SNVs for a given allele age. The inset highlights the cumulative proportion of SNVs that are estimated to have arisen in the last 50,000 years. The *x* axis denotes allele age (×1,000) and the *y* axis indicates the the cumulative proportion of SNVs (%). **c**, Average age for all SNVs, SNVs found in both the European Americans and African Americans (shared), and SNVs found in only one population (specific). **d**, Average age for different types of variants. Error bars represent s.d.

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

Deleterious SNVs were estimated by agreement among several conservation- and structure- based methods, as in Chun and Fay (2009).

The graph shows the proportion of all SNVs of a given age that are predicted to be deleterious.



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

Conclusion

Theory predicts that in an outbred species, *every genome* should carry many bad genes, and genome sequencing suggests this is true!

Fully recessive ***lethals*** could have frequencies as high as 0.001 – 0.01.

Partly dominant deleterious mutations (where heterozygotes suffer slightly) could have frequencies just as high, depending on their degrees of dominance (h) and harmfulness to homozygotes.

The probability of being a heterozygous carrier at any such locus is *low*.

But if there are *many such loci*, then the expected number of heterozygous deleterious mutations per *genome* could be *high*.

Suppose there are 5000 loci with mutations that are lethal or severely deleterious when homozygous, at average frequencies like $q = 0.001$.

Then $2pq \approx 0.002$, and 5000*0.002 = 10 heterozygous loci, on average.

And the "load" of more mildly deleterious mutations could be vastly higher, as we will see in the next lecture.