What limits the efficiency of natural selection?

Nick Barton
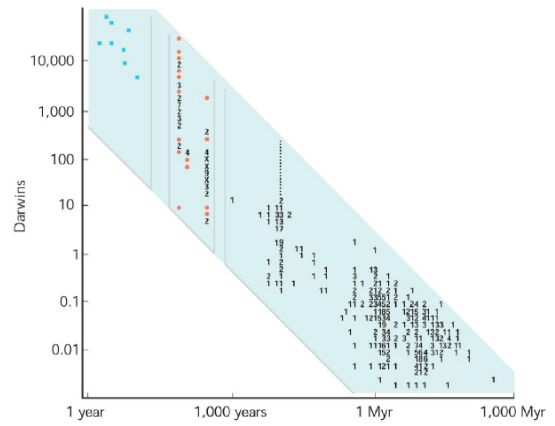
IST AUSTRIA

Complex genomes that code for complex organisms have evolved

Human genome: 3×10^9 bases, $> 10^8$ maintained by selection

Particular changes have happened quickly:

      - insecticide resistance in *Drosophila* (Karasov et al, 2010)

      - rapid morphological change (Gingerich, 1983)

Is the rate of evolution limited by mutation, selection, pop'ln size…?

Population genetics:

- genotype $\underline{X}$, genotype frequency $g[\underline{X}]$, allele frequency $\underline{p} = \mathbb{E}[\underline{X}]$

Quantitative genetics:

- traits $\underline{Z}$, mean & covariance $\overline{z}$, v

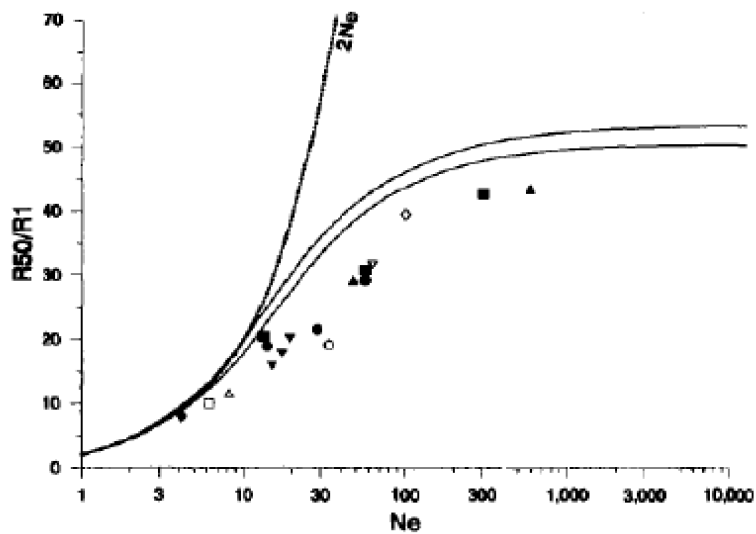In sexual populations, the *infinitesimal model* is accurate:

- offspring are normally distributed around the mid-parent
- covariance within families *independent* of selection
- increased by mutation, decreased by inbreeding

Consistent with additive model: $\underline{Z} = \sum_{i=1}^{n} \underline{\alpha}_i X_i$, $n \gg 1$

Selection experiments fit the predicted response after 50 generations

<div align="right">Weber & Diggins, 1990</div>

The infinitesimal model is *locally* accurate, even though $\underline{Z} = f[\underline{X}]$ is complex

What limits the efficiency of selection?

"Genetic load": loss of fitness relative to some ideal: $1 - \frac{\overline{W}}{W_{\max}}$

This leads to simple constraints:

mutation load $\sim U$ (Haldane, 1937)

substitution load $\sim \log\left[\frac{1}{p_0}\right]$ (Haldane, 1957; Kimura, 1961)

drift load $\sim \frac{1}{4 N}$ per allele or trait

(Kimura & Ohta, 1970; Lande, 1976)

However, these constraints become weaker when genes interact

Mutation load

With asexual reproduction, mean fitness is reduced by the chance of producing offspring with *no* deleterious mutations: $\frac{\overline{W}}{W_{max}} \sim e^{-U}$

In humans, $\mu \sim 10^8$ per base per generation $\Rightarrow U \sim 60$ per diploid genome. $U_{del} \sim 2$ per diploid genome per generation (??)

In a sexual population, the # of bad mutations $\sim \frac{U}{s}$, where $s = \frac{\partial \log(\overline{W})}{\partial k}$
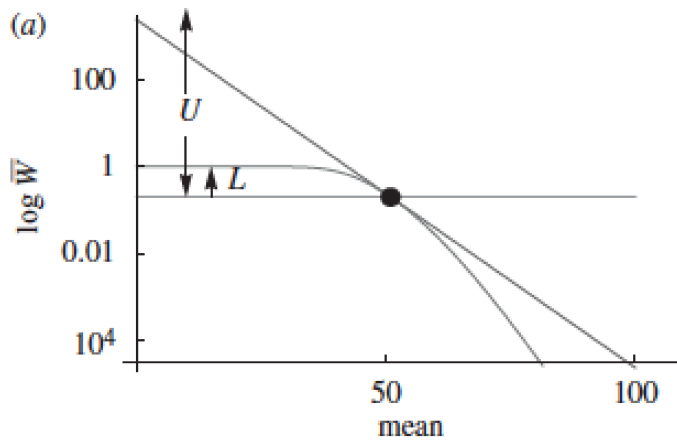
The mutation load is greatly reduced by *negative epistasis*

The *variance in fitness* is $U s$, which equals the rate of decline in fitness due to mutation.

Substitution load

$$p_t = p_0 \; \frac{W_1^*}{\overline{W}_1} \; \frac{W_2^*}{\overline{W}_2} \; \cdots \; \frac{W_t^*}{\overline{W}_t} \qquad \therefore \; \log\left[\frac{p_t}{p_0}\right] = \sum_{i=1}^{t} \log\left[\frac{W_i^*}{\overline{W}_i}\right] = \text{“load”}$$

This is true for asexuals, and with sex & multiplicative fitnesses

With sex, and selecting the best θ of the population, *all* rare variants will increase by $\frac{1}{\theta}$ per generation.

Substitutions at rate Λ require variance in fitness $\sim \Lambda \jmath$

For given fitness variance, weak selection, $N \jmath \sim 1$, maximises Λ

Drift load

Wright's (1937) distribution of allele frequencies:

$$P[\underline{p}] ~\sim~ \overline{W}^{2\,N} \left(\prod_{i=1}^{n} p_i^{4\,N\mu_i-1} ~ q_i^{4\,N\nu_i-1} \right)$$

assuming free recombination, allowing arbitrary interactions
$\Rightarrow$ distribution of trait means and covariance:

$$P[\overline{z},~v] ~\sim~ \overline{W}[\overline{z},~v]^{2\,N} \psi_0[\overline{z},~v]$$

Focus on the mean of a single trait, around an optimum at $z = 0$:

$$\overline{W} \sim \exp\left[-\tfrac{S}{2}\,z^2\right] \quad \Rightarrow \quad \overline{W}^{2\,N} \sim \exp[-N\,S\,z^2]$$

$$\Rightarrow \text{var}[z] \sim \tfrac{1}{2\,N\,S} \quad \Rightarrow \quad E\!\left[\tfrac{S}{2}\,z^2\right] \sim \tfrac{1}{4\,N}$$

variance in fitness is $\tfrac{1}{8\,N^2}$ $\quad \Rightarrow$ # of traits $< 8\,N^2$

Fitness flux limits accumulation of information

Mustonen and Lassig (2010), Jarzynski (1997)

$$\langle \exp[-2\,N\,\Phi + \Delta H]\rangle = 1$$

$$H \equiv \log\left[\frac{P}{P_0}\right] \qquad \Phi \equiv \int_0^T \sum_i s_i \frac{dp_i}{dt}\, dt$$

$$2\,N\,\langle\Phi\rangle \geq \langle\Delta H\rangle$$

$$\int_0^T \sum_i s_i\,(s_i\,p_i\,q_i + mut'\,n + drift)\, dt \leq \int_0^T \sum_i \left(s_i^2\,p_i\,q_i\right)\, dt$$

Total variance in fitness bounds the increase in information?

$$2\,N \int_0^T var\,(W)\, dt \geq \langle\Delta H\rangle$$

Summary

Quantitative genetics describes trait evolution:

      - the infinitesimal model is remarkably accurate

The genetic load appears to constrain genome size, rate of substitution, # of functional traits

…

If genes interact in the right way, constraints are relaxed:

      - variance in fitness limits rate of substitution, # of traits

Why should genes interact in this way?