

Genetic load and efficacy of selection on admixed populations

A Diffusion Processes Approach

Jonatas Cesar

University of São Paulo & University of Chicago

Table of contents

1. Genetic Load and Efficacy of Selection
2. Diffusion simulation
3. Further Steps

Genetic Load and Efficacy of Selection

Definition

Genetic load is the relative reduction in the population fitness compared to the theoretical "perfectly adapted" genotype.

$$L = \frac{W_{max} - W}{W_{max}}$$

Causes:

- Mutation Load: Influx of new deleterious mutation
- Change in environment
- Inbreeding
- ...

Equilibrium Rationale

Genotype	aa	aA	AA
Frequency	$(1 - x)^2$	$2x(1 - x)$	x^2
Average Fitness	1	$1 - hs$	$1 - s$
Mutation Rate	$a \xrightleftharpoons[u]{v} A$		

Fitness

$$W(x) = 1 - 2hsx(1 - x) - sx^2$$

Eq. dist. of x

$$\phi(x) = CW^{2N}x^{4Nu-1}(1 - x)^{4Nv-1}$$

Genetic load

$$\bar{L} = \frac{\int_0^1 dx(1 - W)\phi(x)}{\int_0^1 dx\phi(x)}$$

Small Populations

Load is high since drift increases the frequency of deleterious mutations.

Large Populations

Selection decreases the genetic load.

$$\bar{L} \propto u$$

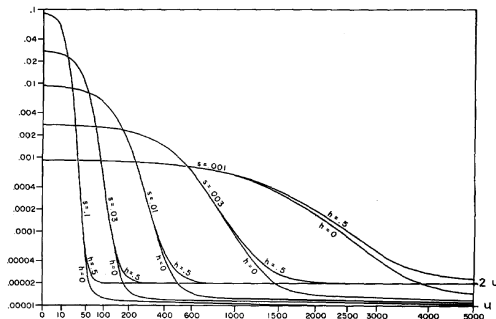
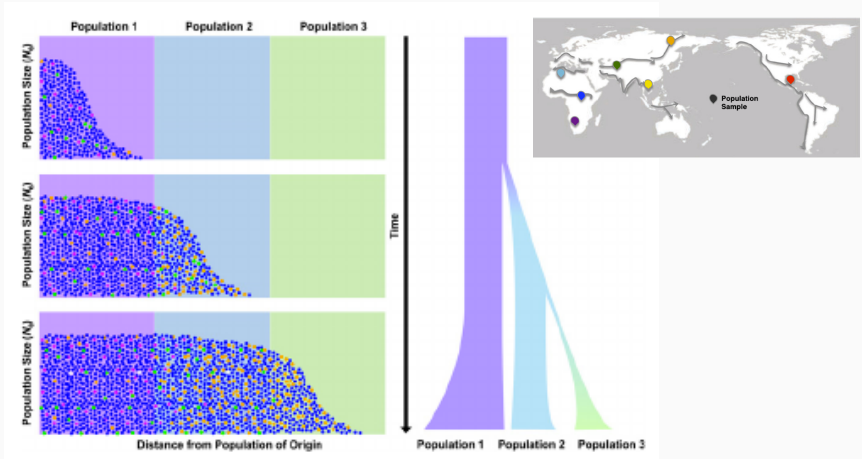


FIGURE 1.—The mutation load (ordinate) as a function of the effective population number (abscissa) for various values of the selection coefficient, s , and dominance, h . The ordinate is on a logarithmic scale; the abscissa is on a scale proportional to the square root. The forward mutation rate, u , is 10^{-5} ; the reverse rate v , is 10^{-6} .

M. Kimura, T. Maruyama & J. Crow; The Mutation Load in Small Populations. Genetics (1963)

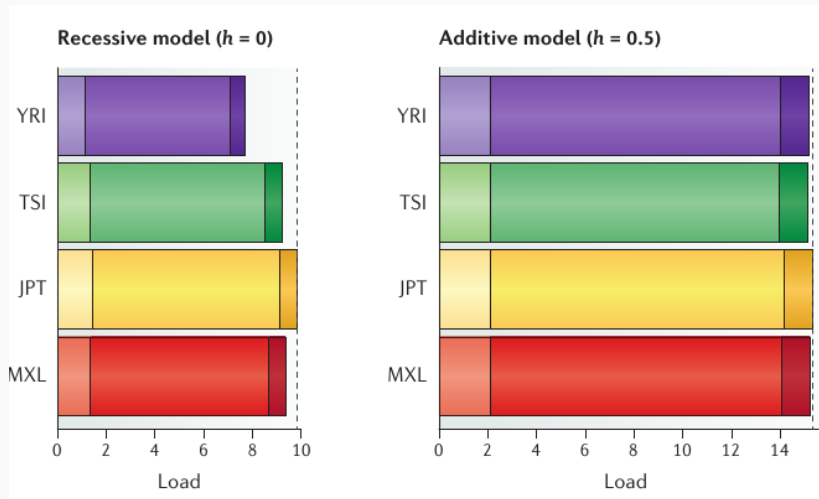
There are differences in mutation load between human populations?



R. McCoy & J. Akey, Pnas 2016

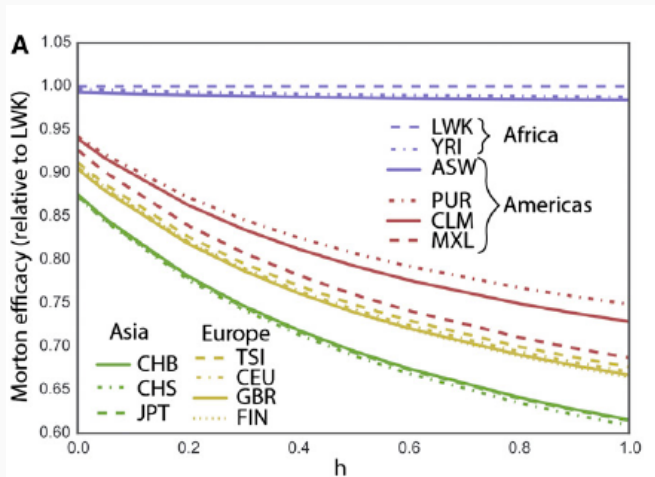
B. Henn, *et.al* Pnas 2016

It depends



B. Henn *et al*/ Nature Review 2016: for snps with GERP 2:4, 4:6 and >6 where assigned $s = -4.5 \times 10^{-4}$, $s = -4.5 \times 10^{-3}$ and $s = -1.0 \times 10^{-1}$ using Boyko et al DFE

But there are differences in the present day efficacy of selection



S. Gravel, Genetics 2016

Definition

The efficacy of selection is defined as the contribution of selection on the population load change $\partial L / \partial t$.

Derivation

Given the average fitness $W = s[2h\mu_1 + (1 - 2h)\mu_2]$ where $\mu_k \equiv \int_0^1 dx x^k \phi(x, t)$ we want to calculate

$$\frac{d}{dt} W = s[2h\dot{\mu}_1 + (1 - 2h)\dot{\mu}_2]$$

using: $\dot{\mu}_k = \frac{\partial}{\partial t} \int_0^1 dx x^k \phi(x, t) = \int_0^1 dx x^k \frac{\partial}{\partial t} \phi(x, t)$

Derivation outline

Diffusion Approximation

Let $\phi(x, t)dx$ the expected number of alleles with frequency between x and $x + dx$ at time t .

$$\begin{aligned}\frac{\partial}{\partial t}\phi(x, t) &\approx \frac{1}{4N} \frac{\partial^2}{\partial x^2} x(1-x)\phi(x, t) && \left. \vphantom{\frac{\partial}{\partial t}} \right\} \text{drift} \\ &- s \frac{\partial}{\partial x} [h + (1-2h)x]x(1-x)\phi(x, t) && \left. \vphantom{\frac{\partial}{\partial t}} \right\} \text{selection} \\ &+ 2Nu\delta\left(x - \frac{1}{2N}\right) && \left. \vphantom{\frac{\partial}{\partial t}} \right\} \text{mutation} \quad (1)\end{aligned}$$

Change in moments ($k > 0$)

$$\dot{\mu}_k = \frac{k(k-1)}{8N} \pi_{k-1} + \frac{sk}{4} \Gamma_{k,h} + \frac{u}{(2N)^{k-1}}$$

where $\Gamma_{k,h} = 2[h\pi_k + (1-2h)\pi_{k+1}]$ and $\pi_k = (\mu_k - \mu_{k+1})$

Single Population Efficacy of Selection

Change in Fitness

$$\begin{aligned}\dot{W} &= \dot{W}_u + \dot{W}_N + \dot{W}_s \\ &= s \left\{ \underbrace{2hu}_{\text{mutation}} + \underbrace{\frac{(1-2h)\pi_1}{4N}}_{\text{drift}} + \underbrace{\frac{s}{2}[h\Gamma_{1,h} + (1-2h)\Gamma_{2,h}]}_{\text{selection}} \right\} \quad (2)\end{aligned}$$

Efficacy of Selection

- Fit Efficacy (from Fitness increase theorem)

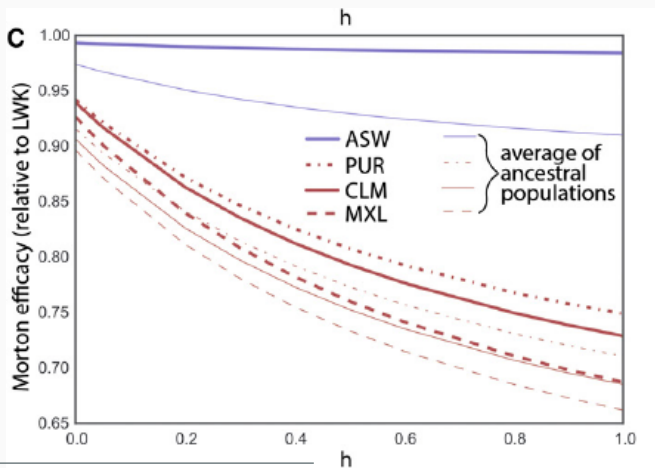
$$\dot{W}_s = \frac{s^2}{2} h\Gamma_{1,h} + (1-2h)\Gamma_{2,h}$$

- Morton Efficacy (from Morton, Crow and Mueller mutational damage)

$$\dot{\mu}_1 = \frac{s^2}{2} h\Gamma_{1,h}$$

Efficacy of selection of Admixed populations

$$\dot{W}_{s,admix} > \sum_i p_i \dot{W}_{s,i}$$



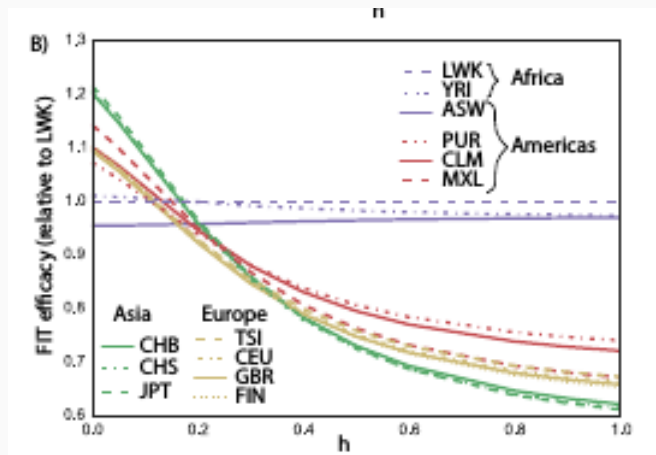
Allele Frequency Distribution

$$\phi(x_3) = \int_0^1 \int_0^1 dx_1 dx_2 \phi(x_1, x_2) \delta(x_3 - \alpha_1 x_1 - \alpha_2 x_2)$$

Moments of The Admixed Population

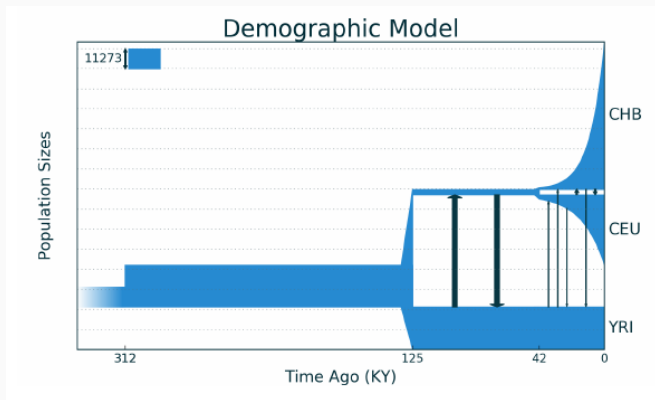
$$\begin{aligned} \mu_{3,k} &= \int_0^1 \int_0^1 \int_0^1 dx_1 dx_2 dx_3 x_3^k \phi(x_1, x_2) \delta(x_3 - \alpha_1 x_1 - \alpha_2 x_2) \\ &= \int_0^1 \int_0^1 dx_1 dx_2 (\alpha_1 x_1 + \alpha_2 x_2)^k \phi(x_1, x_2) \\ &= \sum_{j=1}^k \binom{j}{k} \alpha_1^j \alpha_2^{k-j} \int_0^1 \int_0^1 dx_1 dx_2 x_1^j x_2^{k-j} \phi(x_1, x_2) \\ &= \sum_{j=1}^k \binom{j}{k} \alpha_1^j \alpha_2^{k-j} \mu_{j,k-j}^{1,2} \end{aligned} \tag{3}$$

Present day efficacy of selection - $CADD > 2$

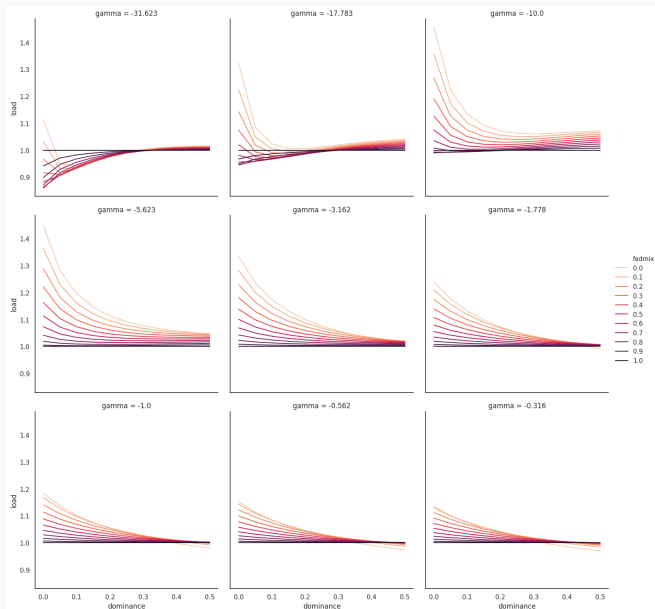


Diffusion simulation

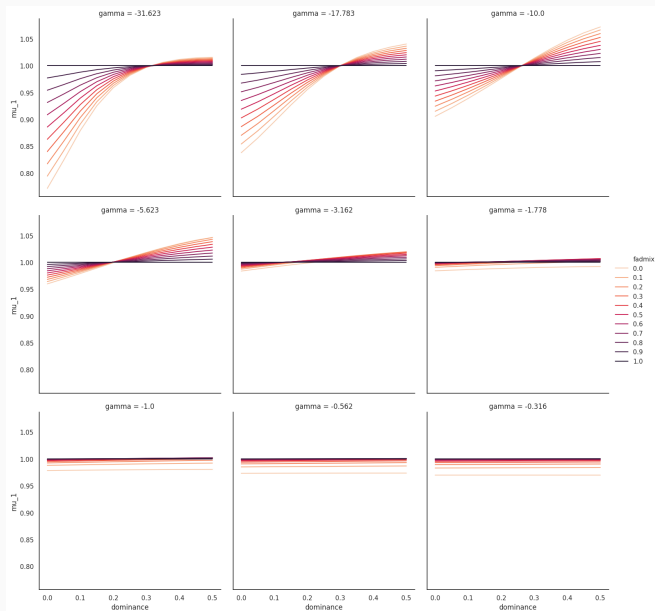
Out of Africa Simulation



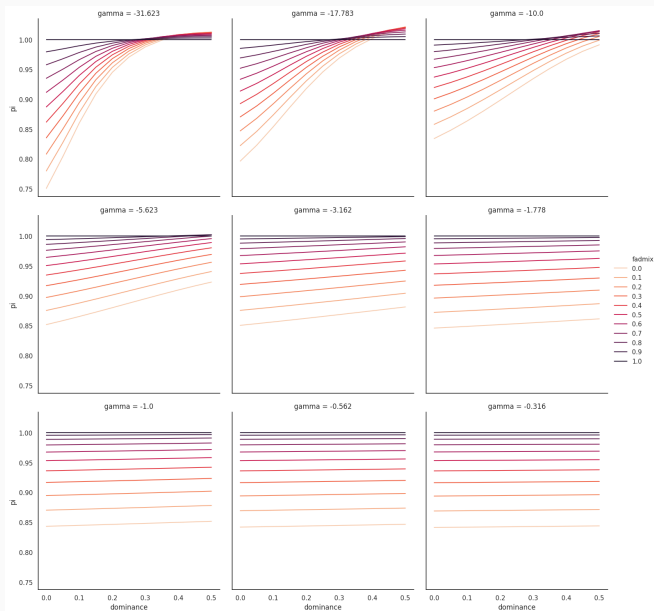
Genetic Load



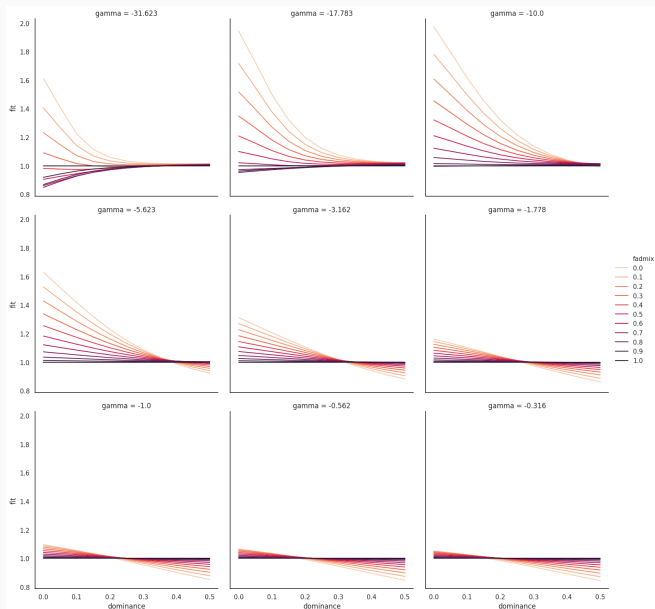
Number of deleterious mutations



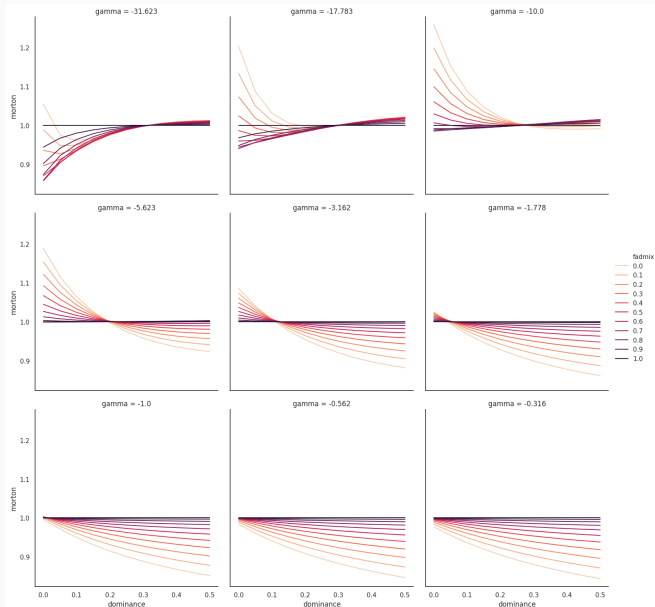
Nucleotide diversity



Fit efficacy



Morton Efficacy



Further Steps

From near to not so near future

- More realistic simulations with Slim3
 - Recombination
 - Distribution of Fitness Effects
 - Dominance models
- Real data analysis
 - 1000G
 - GnomAD
 - AbraOM (1.2K Brazilians)
- Impact of dominance and admixture on GWAS
 - Eyre-Walker (2010)
 - Simons (2017)