Supplementary materials

Contents

1 ω susceptibility after a change in N_e

1.1 Genotype to phenotype map

Define n as the number of sites in the genotype sequence. Each site can be in one of $K \geq 2$ states, where only 1 state is defined the stable state, and $K-1$ states are unstable. For a given genotype sequence, define phenotype $0 \leq x \leq 1$ as the current proportion of sites in the unstable state. After a mutation, given that only one site can change at a time, the absolute change of x is either 0 or $\delta x = 1/n$. Define $\rho_x(\delta x)$ as the probability to get a change of phenotype equal to δx, if the current phenotype is x:

$$\begin{cases} \delta x & \text{with probability } \rho_x(\delta x) = 1 - x, \\ 0 & \text{with probability } \rho_x(0) = x\left[1 - \frac{1}{K-1}\right], \\ -\delta x & \text{with probability } \rho_x(-\delta x) = \frac{x}{K-1}. \end{cases} \tag{1}$$

1.2 Selection coefficient

$s(x, \delta x)$ is the selection coefficient of an effect δx if the current phenotype is x:

$$s(x, \delta x) = \frac{f(x + \delta x) - f(x)}{f(x)}, \tag{2}$$

$$\simeq \frac{1}{f(x)} \frac{\partial f(x)}{\partial x} \delta x, \tag{3}$$

$$\simeq \frac{\partial \ln f(x)}{\partial x} \delta x, \tag{4}$$

where $f(x)$ is the Wrightian fitness of phenotype x. And thus we also have:

$$s(x, -\delta x) \simeq -s(x, \delta x) \text{ from eq. 4}, \tag{5}$$

$$\iff S(x, -\delta x) \simeq -S(x, \delta x), \tag{6}$$

where $S(x^*, \delta x) = 4N_\mathrm{e} s(x^*, \delta x)$ is the scaled selection coefficient.

1.3 Probability of fixation

The probability of fixation of a mutation with effect δx, for a resident phenotype x is :

$$p_\mathrm{fix}(x, \delta x) = \frac{1 - \mathrm{e}^{-2s(x,\delta x)}}{1 - \mathrm{e}^{-4N_\mathrm{e} s(x,\delta x)}}, \tag{7}$$

$$\simeq \frac{2s(x, \delta x)}{1 - \mathrm{e}^{-4N_\mathrm{e} s(x,\delta x)}}, \tag{8}$$

$$= \frac{2s(x, \delta x)}{1 - \mathrm{e}^{-S(x,\delta x)}}. \tag{9}$$

And in the case of neutral mutations, the probability of fixation is:

$$p_\mathrm{fix}(x, 0) = \frac{1}{2N_\mathrm{e}}. \tag{10}$$

And the ratio of probability of fixation between selected and neutral mutations is:

$$\frac{p_\mathrm{fix}(x, \delta x)}{p_\mathrm{fix}(x, 0)} = \frac{2N_\mathrm{e} 2s(x, \delta x)}{1 - \mathrm{e}^{-S(x,\delta x)}} \text{ from eq. 9 and 10}, \tag{11}$$

$$= \frac{S(x, \delta x)}{1 - \mathrm{e}^{-S(x,\delta x)}}. \tag{12}$$

1.4 Equilibrium phenotype

At equilibrium phenotype x^*, the expected selection coefficient of mutation that reached fixation must be 0:

$$0 = \mathbb{E}_{\delta x} \left[s(x^*, \delta x) p_\mathrm{fix}(x^*, \delta x) \right], \tag{13}$$

$$\iff 0 = \frac{2s(x^*, \delta x)^2}{1 - \mathrm{e}^{-S(x^*,\delta x)}} \rho_{x^*}(\delta x) + s(x^*, 0) \frac{\rho_{x^*}(0)}{2N_\mathrm{e}} + \frac{2s(x^*, -\delta x)^2}{1 - \mathrm{e}^{-S(x^*,-\delta x)}} \rho_{x^*}(-\delta x) \text{ from eq. 9 and 10}, \tag{14}$$

$$\implies \frac{2s(x^*, \delta x)^2}{1 - \mathrm{e}^{-S(x^*,\delta x)}} \rho_{x^*}(\delta x) \simeq \frac{-2s(x^*, \delta x)^2}{1 - \mathrm{e}^{S(x^*,\delta x)}} \rho_{x^*}(-\delta x) \text{ from eq. 6}, \tag{15}$$

$$\iff \frac{\rho_{x^*}(\delta x)}{\rho_{x^*}(-\delta x)} \simeq \mathrm{e}^{-S(x^*,\delta x)} \frac{\mathrm{e}^{-S(x^*,\delta x)} - 1}{\mathrm{e}^{-S(x^*,\delta x)} \left(1 - \mathrm{e}^{S(x^*,\delta x)}\right)}, \tag{16}$$

$$\iff \ln\left(\frac{1 - x^*}{x^*}\right) + \ln(\mathrm{K} - 1) \simeq -S(x^*, \delta x) \text{ from eq. 1}, \tag{17}$$

$$\iff \lambda_K(x^*) \simeq -S(x^*, \delta x), \tag{18}$$

where $\lambda_K(x^*) = \ln\left(\frac{1-x^*}{x^*}\right) + \ln(\mathrm{K} - 1)$.

1.5 Substitution rate bias (ω) at equilibrium

The substitution rate of selected mutations relative to neutral mutations (ω) is by definition:

$$\omega = \mathbb{E}_{\delta x}\left[\frac{p_{\text{fix}}(x,\delta x)}{p_{\text{fix}}(x,0)}\right], \tag{19}$$

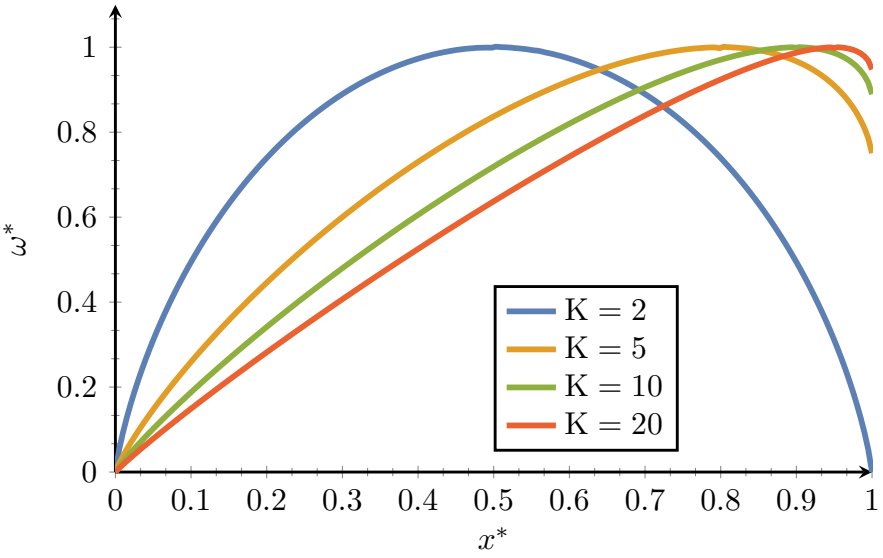
$$= (1-x)\frac{S(x,\delta x)}{1-\mathrm{e}^{-S(x,\delta x)}} + x\left(\frac{\mathrm{K}-2}{\mathrm{K}-1}\right) + \frac{x}{\mathrm{K}-1}\frac{S(x,-\delta x)}{1-\mathrm{e}^{-S(x,-\delta x)}} \text{ from eq. 1, 9 and 10,} \tag{20}$$

$$= (1-x)\frac{S(x,\delta x)}{1-\mathrm{e}^{-S(x,\delta x)}} - \frac{x}{\mathrm{K}-1}\frac{S(x,\delta x)}{1-\mathrm{e}^{S(x,\delta x)}} + x\left(\frac{\mathrm{K}-2}{\mathrm{K}-1}\right) \text{ from eq. 6.} \tag{21}$$

ω^* at equilibrium is then determined by the phenotype at equilibrium x^*:

$$\omega^* = (1-x^*)\frac{S(x^*,\delta x)}{1-\mathrm{e}^{-S(x^*,\delta x)}} - \frac{x^*}{\mathrm{K}-1}\frac{S(x^*,\delta x)}{1-\mathrm{e}^{S(x^*,\delta x)}} + x^*\left(\frac{\mathrm{K}-2}{\mathrm{K}-1}\right), \tag{22}$$

$$= x^*\left[\frac{2(x^*-1)\lambda_K(x^*)}{\mathrm{K}(x^*-1)+1} + \frac{\mathrm{K}-2}{\mathrm{K}-1}\right] \text{ from eq. 17.} \tag{23}$$



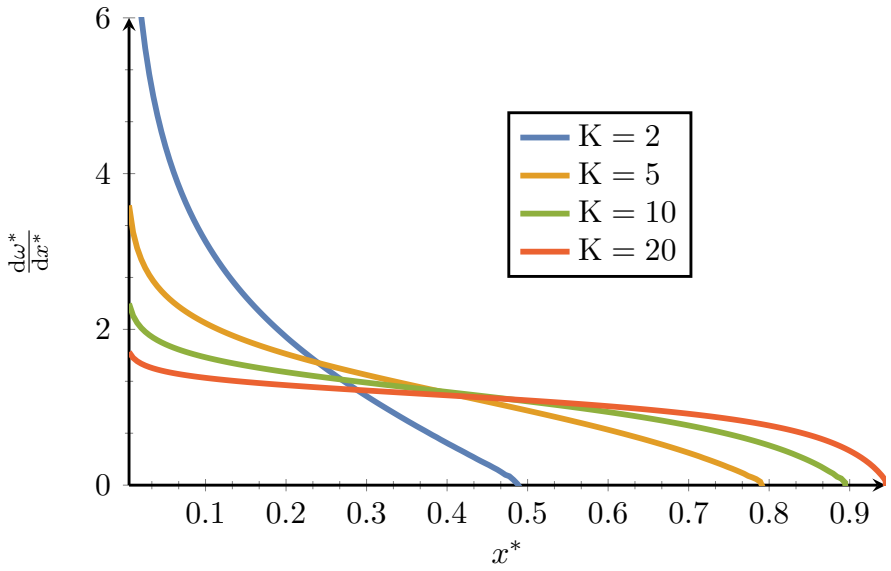
Moreover, given that the number of state if large enough $\mathrm{K} \gg 1$, the equilibrium ω can be approximated as:

$$\omega^* = x^*\left[\frac{2(x^*-1)\lambda_K(x^*)}{\mathrm{K}(x^*-1)+1} + \frac{\mathrm{K}-2}{\mathrm{K}-1}\right], \tag{24}$$

$$\simeq x^* \tag{25}$$

And the derivative of ω^* w.r.t to x^* is:

$$\frac{\mathrm{d}\omega^*}{\mathrm{d}x^*} = 2\left[\frac{\mathrm{K}(x^*-1)+1+\left[\mathrm{K}(x^*-1)^2+2x^*-1\right]\lambda_K(x^*)}{(\mathrm{K}(x^*-1)+1)^2}\right] + \frac{\mathrm{K}-2}{\mathrm{K}-1}. \tag{26}$$



Moreover, given that the number of state if large enough $\mathrm{K} \gg 1$, the response in equilibrium ω due to change in phenotype can be approximated as:

$$\frac{\mathrm{d}\omega^*}{\mathrm{d}x^*} = 2\left[\frac{\mathrm{K}(x^*-1)+1+\left[\mathrm{K}(x^*-1)^2+2x^*-1\right]\lambda_K(x^*)}{(\mathrm{K}(x^*-1)+1)^2}\right] + \frac{\mathrm{K}-2}{\mathrm{K}-1}, \tag{27}$$

$$\simeq \frac{2\lambda_K(x^*)}{\mathrm{K}} + 1, \tag{28}$$

$$\simeq 1. \tag{29}$$

1.6 ω susceptibility after a change in N_e

Define the function $G(x, N_\mathrm{e})$ as:

$$G(x, N_\mathrm{e}) \equiv \lambda_K(x^*) + 4N_\mathrm{e}s(x, \delta x), \tag{30}$$

The equilibrium equation (eq. 17) states that $G(x^*, N_\mathrm{e}) = 0$, meaning that x^* is implicitly a function of N_e:

$$G(x^*(N_\mathrm{e}), N_\mathrm{e}) = 0, \tag{31}$$

$$\implies \frac{\partial G(x^*, N_\mathrm{e})}{\partial x^*}\frac{\mathrm{d}x^*}{\mathrm{d}N_\mathrm{e}} + \frac{\partial G(x^*, N_\mathrm{e})}{\partial N_\mathrm{e}} = 0, \tag{32}$$

$$\iff \left[\frac{\partial \lambda_K(x^*)}{\partial x^*} + 4N_\mathrm{e}\frac{\partial s(x^*, \delta x)}{\partial x^*}\right]\frac{\mathrm{d}x^*}{\mathrm{d}N_\mathrm{e}} + 4s(x^*, \delta x) = 0, \tag{33}$$

$$\iff \left[\frac{\partial \lambda_K(x^*)}{\partial x^*} + 4N_\mathrm{e}\frac{\partial^2 \ln f(x^*)}{\partial {x^*}^2}\delta x\right]\frac{\mathrm{d}x^*}{\mathrm{d}N_\mathrm{e}} = -4\frac{\partial \ln f(x^*)}{\partial x^*}\delta x \text{ from eq. 4}, \tag{34}$$

$$\iff 4\delta x\left[\frac{1}{4\delta x N_\mathrm{e}}\frac{\partial \lambda_K(x^*)}{\partial x^*} + \frac{\partial^2 \ln f(x^*)}{\partial {x^*}^2}\right]N_\mathrm{e}\frac{\mathrm{d}x^*}{\mathrm{d}N_\mathrm{e}} = -4\delta x\frac{\partial \ln f(x^*)}{\partial x^*}, \tag{35}$$

$$\iff \frac{\mathrm{d}x^*}{\mathrm{d}\ln(N_\mathrm{e})} = -\frac{\frac{\partial \ln f(x^*)}{\partial x^*}}{\frac{1}{4\delta x N_\mathrm{e}}\frac{\partial \lambda_K(x^*)}{\partial x^*} + \frac{\partial^2 \ln f(x^*)}{\partial {x^*}^2}}. \tag{36}$$

Giving the equation for the response of phenotype at equilibrium after a change of effective population size. Together, the response of substitution rate at equilibrium, after a change of effective population size can be obtain as:

$$\frac{\mathrm{d}\omega^*}{\mathrm{d}\ln(N_\mathrm{e})} = \frac{\mathrm{d}\omega^*}{\mathrm{d}x^*}\frac{\mathrm{d}x^*}{\mathrm{d}\ln(N_\mathrm{e})}, \tag{37}$$

$$= -\frac{\mathrm{d}\omega^*}{\mathrm{d}x^*}\frac{\frac{\partial \ln f(x^*)}{\partial x^*}}{\frac{1}{4\delta x N_\mathrm{e}}\frac{\partial \lambda_K(x^*)}{\partial x^*} + \frac{\partial^2 \ln f(x^*)}{\partial x^{*2}}} \text{ from eq. 36.} \tag{38}$$

Moreover, with the approximation that $\left|4N_\mathrm{e}\frac{\partial s(x^*,\delta x)}{\partial x^*}\right| \gg \left|\frac{\partial \lambda_K(x^*)}{\partial x^*}\right|$, meaning that a change in phenotype causes a higher change in scaled selection coefficient than mutational bias, we have:

$$\frac{\mathrm{d}x^*}{\mathrm{d}\ln(N_\mathrm{e})} = -\frac{\frac{\partial \ln f(x^*)}{\partial x^*}}{\frac{1}{4\delta x N_\mathrm{e}}\frac{\partial \lambda_K(x^*)}{\partial x^*} + \frac{\partial^2 \ln f(x^*)}{\partial x^{*2}}}, \tag{39}$$

$$\implies \frac{\mathrm{d}x^*}{\mathrm{d}\ln(N_\mathrm{e})} \simeq -\frac{\frac{\partial \ln f(x^*)}{\partial x^*}}{\frac{\partial^2 \ln f(x^*)}{\partial x^{*2}}}. \tag{40}$$

Together, these approximations leads to the following susceptibility in equilibrium ω after change in N_e as:

$$\frac{\mathrm{d}\omega^*}{\mathrm{d}\ln(N_\mathrm{e})} \simeq -\frac{\frac{\partial \ln f(x^*)}{\partial x^*}}{\frac{\partial^2 \ln f(x^*)}{\partial x^{*2}}} \tag{41}$$

2 Models for the log-fitness function

2.1 Folded fraction

All phenotype-fitness functions considered below are log-concave, and as a result, $\frac{\partial \ln f(x^*)}{\partial x^*}$ is an decreasing function of x; the less stable the protein already is, the stronger the purifying selection against additional destabilizing mutations. More precisely, fitness functions depends on the folded fraction of the protein of interest, which is given by the Fermi-Dirac distribution:

$$P_F(x) \quad = \quad \frac{1}{1+e^{\beta(\alpha+\gamma \mathrm{n} x)}} \tag{42}$$

where x is the fraction of destabilizing mutations, each contributing a $\Delta\Delta G = \gamma$, and $\beta = 1/kT$. Thus, $\alpha = \Delta G_0 < 0$ is the difference in free energy between folded and unfolded state when all sites are stable. $\mathrm{n}\gamma$ is the expected change in ΔG when all sites are unstable. The misfolded fraction is $P_U = 1 - P_F$. In addition, P_F is typically close to 1 (or $P_U \ll 1$), so that we can use a first-order approximation:

$$P_F(x) \quad = \quad 1 - P_U(x) \tag{43}$$

$$\simeq \quad 1 - e^{\beta(\alpha+\gamma \mathrm{n} x)} \tag{44}$$

or equivalently

$$P_U(x) \quad \simeq \quad e^{\beta(\alpha+\gamma \mathrm{n} x)} \tag{45}$$

2.2 Fitness equal to folded fraction

A first model is to assume that the fitness is equal to the folded fraction (Goldstein, 2013):

$$f(x) = \frac{1}{1+\mathrm{e}^{\beta(\alpha+\mathrm{n}\gamma x)}}. \tag{46}$$

The derivative of fitness w.r.t to phenotype is:

$$\frac{\partial \ln(f(x))}{\partial x} = -\frac{\partial \ln\left(1+\mathrm{e}^{\beta(\alpha+\mathrm{n}\gamma x)}\right)}{\partial x} \text{ from eq. 46,} \tag{47}$$

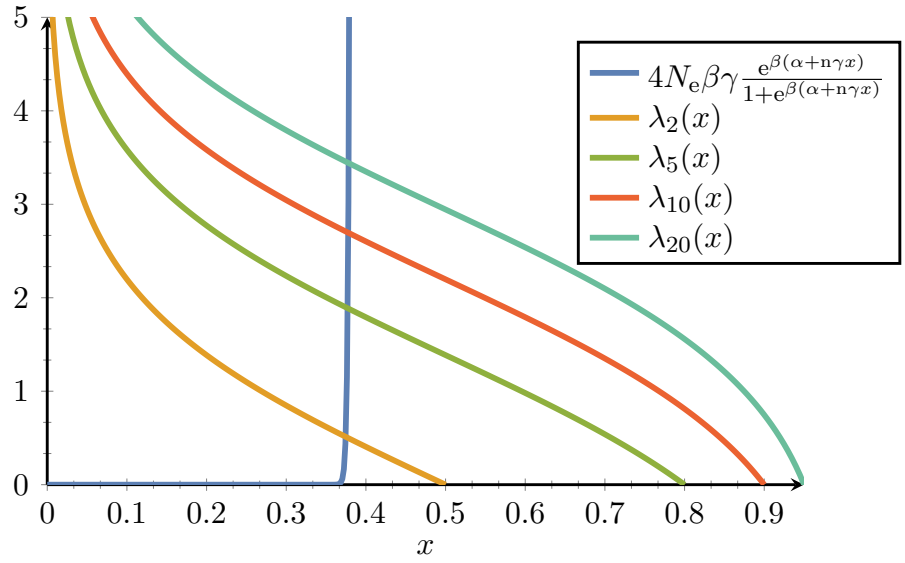
$$= -\beta \mathrm{n}\gamma \frac{\mathrm{e}^{\beta(\alpha+\mathrm{n}\gamma x)}}{1+\mathrm{e}^{\beta(\alpha+\mathrm{n}\gamma x)}}, \tag{48}$$

$$\simeq -\beta \mathrm{n}\gamma \mathrm{e}^{\beta(\alpha+\mathrm{n}\gamma x)}. \tag{49}$$

The equilibrium phenotype (x^*) is :

$$\lambda_K(x^*) = 4N_e\beta\gamma\frac{e^{\beta(\alpha+n\gamma x^*)}}{1+e^{\beta(\alpha+n\gamma x^*)}} \text{ from eq. 18 and 48.} \tag{50}$$

Using $N_e = 10^4$, $\beta = 1.686$, $\alpha = -118$, n $= 300$, $\gamma = 1$, we have the following :



Where in this example we can visually appreciate that the a change in phenotype causes a higher change in scaled selection coefficient than mutational bias (eq. 40). And the second derivative of fitness w.r.t to phenotype is:

$$\frac{\partial^2 \ln(f(x))}{\partial x^2} = -\beta n\gamma\frac{\partial}{\partial x}\left(\frac{e^{\beta(\alpha+n\gamma x)}}{1+e^{\beta(\alpha+n\gamma x)}}\right) \text{ from eq. 48,} \tag{51}$$

$$= -\beta n\gamma\beta n\gamma\frac{e^{\beta(\alpha+n\gamma x)}}{\left(1+e^{\beta(\alpha+n\gamma x)}\right)^2}, \tag{52}$$

$$= \frac{\beta n\gamma}{1+e^{\beta(\alpha+n\gamma x)}}\frac{\partial \ln f(x)}{\partial x} \text{ from eq. 48,} \tag{53}$$

$$\simeq \beta n\gamma\frac{\partial \ln f(x)}{\partial x} \tag{54}$$

Finally, ω susceptibility after a change in N_e is simply:

$$\frac{d\omega^*}{d\ln(N_e)} \simeq -\frac{1}{\beta n\gamma} \text{from eq. 53 and 29,} \tag{55}$$

which is independent of x^*, meaning ω is linearly decreasing with N_e in log space. This model, however, does not express the fact that selection is typically stronger for proteins characterized by higher levels of expression.

2.3 Selective cost proportional to amount of misfolded protein

A slight variation is to assume that the selective cost itself is proportional to the total amount of misfolded protein (Drummond *et al.*, 2005; Wilke and Drummond, 2006; Drummond and Wilke, 2008; Serohijos *et al.*, 2012). For a given protein with expression level y:

$$\ln f(x) = -AyP_U(x) \tag{56}$$

where A is the cost per misfolded macromolecule. Then,

$$\frac{\partial \ln(f(x))}{\partial x} \simeq -Ay\beta\gamma \mathrm{n} e^{\beta(\alpha+\gamma \mathrm{n} x)}. \tag{57}$$

Under this model, the phenotype at equilibrium is given by:

$$\lambda_K(x^*) = 4N_\mathrm{e} y A \beta\gamma \mathrm{n} e^{\beta(\alpha+\gamma \mathrm{n} x^*)} \text{ from eq. 18 and 48.} \tag{58}$$

And the susceptibility of ω after a change in N_e is the same as before:

$$\frac{\mathrm{d}\omega^*}{\mathrm{d}\ln(N_\mathrm{e})} \simeq -\frac{1}{\beta \mathrm{n}\gamma}. \tag{59}$$

Since N_e and y are confounded factors, meaning they only appear in the equation as a product between the two, implicit derivation leads to the same result whenever the derivation is w.r.t N_e or y, leading to same compact susceptibility:

$$\frac{\mathrm{d}\omega^*}{\mathrm{d}\ln(y)} = \frac{\mathrm{d}\omega^*}{\mathrm{d}\ln(N_\mathrm{e})} \simeq -\frac{1}{\beta \mathrm{n}\gamma}. \tag{60}$$

2.4 Translational errors

An other variant account for translational errors. Translational errors occur at a rate ρ per residue. These errors contribute additional destabilizing mutations, each with effect size $\delta x = 1/n$. The total number of translational errors per macromolecule is approximately Poisson distributed:

$$\pi_k = e^{-\rho \mathrm{n}} \frac{(\rho \mathrm{n})^k}{k!} \tag{61}$$

and the total selective cost is now an average over all possible values of k:

$$\begin{aligned}
\ln f(x) &= -Ay \sum_k \pi_k e^{\beta(\alpha+\gamma \mathrm{n} x+\gamma k)} && (62)\\
&= -Aye^{\beta(\alpha+\gamma \mathrm{n} x)} \sum_k e^{-\rho n} \frac{(\rho n)^k}{k!} e^{\beta\gamma k} && (63)\\
&= -Aye^{\beta(\alpha+\gamma \mathrm{n} x)+\rho \mathrm{n}(e^{\beta\gamma}-1)} && (64)\\
&\simeq -Aye^{\beta(\alpha+\gamma \mathrm{n} x)+\rho\beta\gamma \mathrm{n}} && (65)\\
&= -Aye^{\beta(\alpha+\gamma \mathrm{n}(x+\rho))} && (66)
\end{aligned}$$

In words, the fitness function is the same the previous model, except that the trait x (fraction of destabilizing mutations) is shifted by ρ, the mean fraction of additional mutations contributed by translation errors. This additional factor is independent of x, and as a result, the scaled selection strength is essentially the same, up to a proportionality constant (contributed by the shift):

$$\begin{aligned}
4N_\mathrm{e} \frac{\partial \ln(f(x))}{\partial x} &\propto -4N_\mathrm{e} y e^{\beta(\alpha+\gamma \mathrm{n}(x+\rho))} && (67)\\
&\propto -4N_\mathrm{e} y e^{\beta(\alpha+\gamma \mathrm{n} x)} && (68)
\end{aligned}$$

Moreover, ω susceptibility after a change in N_e is again the same as before:

$$\frac{\mathrm{d}\omega^*}{\mathrm{d}\ln(N_\mathrm{e})} \simeq -\frac{1}{\beta \mathrm{n}\gamma}. \tag{69}$$

2.5 Cost-benefit argument

The cost-benefit argument (Beaulieu *et al.*, 2018) is based on two assumptions

1. the expression level is regulated so that the total number of *functional* macromolecules is maintained at a target level y;

2. the log-fitness is proportional to the ratio of the *total* cost of expression over the benefit contributed by the protein .

Specifically, the protein is assumed to be regulated so as to reach a level of expression of functional proteins of y, and contributes a total benefit B (which depends on its specific function). Given that only a fraction $P_F(x) = 1 - P_U(x)$ of the total amount of protein expressed by the cell is functional, the total cost of expression C is then equal to:

$$\begin{aligned} C(x) &= \frac{y}{P_F(x)} && (70)\\ &\simeq y(1 + P_U(x)) && (71)\end{aligned}$$

Then, the log-fitness is given by:

$$\begin{aligned} \ln f(x) &= -A\frac{y}{B}\left(1 + e^{\beta(\alpha+\gamma \mathrm{n} x)}\right) && (72)\\ &= -by(1 + e^{\beta(\alpha+\gamma \mathrm{n} x)}) && (73)\end{aligned}$$

where $b = A/B$. Compared to model 2 and 3, the log-fitness now has an additional term that depends on the target expression level y, but not on trait x. The scaled strength of selection on mutations affecting x has thus the same functional form as for the two previous models:

$$4N_{\mathrm{e}}\frac{\partial \ln(f(x))}{\partial x} \propto -4N_{\mathrm{e}} y e^{\beta(\alpha+\gamma \mathrm{n} x)} \quad (74)$$

Alternative cost-expression models could also be used, allowing for a non-linear cost function for expression or for some susceptibility of the realized equilibrium expression level, as a function of the number of mutations. Under these models, the strength of selection is still expected to be an increasing function of y, although not linear, e.g.

$$4N_{\mathrm{e}}\frac{\partial \ln(f(x))}{\partial x} \propto -4N_{\mathrm{e}} g(y) e^{\beta(\alpha+\gamma \mathrm{n} x)} \quad (75)$$

where g is some function of y. Moreover, ω susceptibility after a change in N_{e} is again the same as before:

$$\frac{\mathrm{d}\omega^*}{\mathrm{d}\ln(N_{\mathrm{e}})} \simeq -\frac{1}{\beta \mathrm{n}\gamma}. \quad (76)$$

3 Model of protein-protein interactions

The proteome is assumed to be composed of m protein species, all with same abundance C. Each macromolecule may either be in free form or engaged in a non-specific interaction. Only pairwise interactions are considered, and higher-order interactions are ignored. The equilibrium is characterized by:

$$[ij] = \frac{[i][j]}{C_0} e^{\beta E_{ij}} \quad (77)$$

where $[i]$ and $[j]$ are the concentrations of protein species i and j, and $[ij]$ is the concentration of their (non-specific) dimer. Here, E_{ij} is the interaction free energy, which can itself be decomposed as a sum of three terms:

$$\begin{aligned} E_{ij} &= \alpha + E_i + E_j && (78)\\ &= \alpha + \gamma \mathrm{n}(x_i + x_j) && (79)\end{aligned}$$

where we assume that each protein has $n = 100$ residues at its surface, x_i stands for the fraction of hydrophobic residues at the surface of protein i, and each hydrophobic residue makes an additive contribution of γ to the total.

By conservation of the total number of molecules:

$$\begin{aligned} C &= [i] + \sum_{j\neq i}[ij] && (80)\\ &= [i] + \sum_{j\neq i}\frac{[i][j]}{C_0} e^{\beta E_{ij}} && (81)\end{aligned}$$

and we note:

$$\epsilon_i = \sum_j [ij] \tag{82}$$

the fraction of protein i sequestered in non-specific interactions. We assume that the log fitness is proportional to the total amount of protein sequestered in non-specific interactions:

$$\ln f(x) \quad = \quad -b \sum_i \epsilon_i \tag{83}$$

where $b > 0$ is a parameter determining the overall stringency of selection against non-specific interactions.

3.1 Mean field, weak-interaction limit

To make the model tractable and compact, we assume that non-specific interactions are weak, i.e. $\epsilon_i << 1$ for all i. We then make a first-order approximation in the ϵ_i's. In addition, we use a mean-field approximation, such that, when considering a specific protein species i, we assume that all other proteins have the same fraction $\bar{x}$ of hydrophobic residues at their surface. The value of $\bar{x}$ could in principle be found using a self-consistent argument, essentially by (1) explicitly calculating the net substitution flux for protein i with fraction x_i, under mean field $\bar{x}$, and (2) expressing the constraint that this substitution process for protein i is stationary at $x_i = \bar{x}$. This derivation is not conducted here, as it is not needed. Using these approximations, we can re-express the conservation of total mass as:

$$C \quad = \quad [i] + (m-1)[i] \frac{C}{C_0} e^{\beta(\alpha + \gamma \mathrm{n}(\bar{x} + x_i))} \tag{84}$$

Here, we have used the fact that $[j] = C(1 - \epsilon_j)$ can be approximated as $[j] \simeq C$ since it is involved in a term already of the order of ϵ_i. As a result, all $m - 1$ terms of the sum over $j \neq i$ are identical. Next, solving for $[i]$ gives:

$$\begin{aligned}
[i] \quad &= \quad \frac{C}{1 + (m-1)\frac{C}{C_0} e^{\beta(\alpha + \gamma \mathrm{n}(\bar{x} + x_i))}} && (85) \\
&\simeq \quad C \left(1 - m \frac{C}{C_0} e^{\beta(\alpha + \gamma \mathrm{n}(\bar{x} + x_i))} \right) && (86) \\
&= \quad C(1 - \epsilon_i) && (87)
\end{aligned}$$

and thus ϵ_i can be identified with:

$$\epsilon_i \quad = \quad m \frac{C}{C_0} e^{\beta(\alpha + \gamma \mathrm{n}(\bar{x} + x_i))} \tag{88}$$

Now, assume that the system is at equilibrium (thus $x_i = \bar{x}$). The strength of selection acting on mutations occurring at the surface of protein i, of effect size $\delta x = \pm 1/n$, is given by $s = \kappa \delta x$ where:

$$\begin{aligned}
\kappa_i \quad &= \quad b \frac{\mathrm{d}\epsilon_i}{\mathrm{d}x_i} && (89) \\
&= \quad b\beta\gamma \mathrm{n} m \frac{C}{C_0} e^{\beta(\alpha + \gamma \mathrm{n}(\bar{x} + x_i))} && (90)
\end{aligned}$$

and thus:

$$\ln \kappa_i \quad = \quad \ln \left(b\beta\gamma \mathrm{n} m \frac{C}{C_0} \right) \; + \; \beta(\alpha + \gamma \mathrm{n} \bar{x}) \; + \; \beta\gamma \mathrm{n} x_i \tag{91}$$

where only the last term depends on x_i. Finally, applying the main result of this work to the present case allows us to express the susceptibility of ω as a function of N_e as:

$$\begin{aligned}
\chi \quad &= \quad \frac{\mathrm{d}\omega}{\mathrm{d} \ln N_\mathrm{e}} && (92) \\
&= \quad 2(\lambda - 1) \frac{\mathrm{d} \ln \kappa_i}{\mathrm{d}x_i} && (93) \\
&= \quad 2(\lambda - 1) \frac{1}{\beta\gamma n} && (94)
\end{aligned}$$

Note that, here, we have used $K = 2$ (hydrophobic and polar residues are roughly equally likely to occur by mutation), and assumed $x^* << 1$. A more accurate formula could be used without this latter assumption. In any case, χ is now dependent on x^*, through λ.

3.2 Empirical calibration

Based on empirical estimates (Zhang *et al.*, 2008). The mean fraction of hydrophobic residues at the surface of proteins is 0.22 ± 0.06. With $n = 100$ residues, this makes 22 ± 6. The mean value for E_{ij} is 7 kT, with a standard deviation of $\sigma = 1.8$ kT. Assuming that this standard deviation of ± 1.8 kT is contributed by ± 6 mutations gives $\gamma = 1.8/6 = 0.3$ kT or 0.18 kcal per mole. Also, with $x = 0.22$, $\lambda \simeq 4$, and thus $\chi = 6/30 = 0.2$, thus a much stronger response than under the model based on conformational stability.

4 Empirical estimation of χ

Substitution rate as a function of expression level compiled by (Zhang and Yang, 2015). In (Brevet and Lartillot, 2019), the covariance matrix with $\ln(N_e)$ and $\ln(\omega)$ as entries allow us to approximate χ:

Type	Specie	$\hat{\chi}$	r^2
Plant	Oryza sativa	-0.008	0.047
Plant	Arabidopsis thaliana	-0.012	0.128
Archaea	Sulfolobus solfataricus	-0.037	0.097
Archaea	Thermococcus kodakarensis	-0.026	0.058
Fungi	Saccharomyces cerevisiae	-0.029	0.211
Fungi	Aspergillus nidulans	-0.034	0.124
Bacteria	Escherichia coli	-0.021	0.151
Bacteria	Bacillus subtilis	-0.046	0.151
Animal	Caenorhabditis elegans	-0.026	0.039
Animal	Drosophila melanogaster	-0.005	0.021
Animal	Mus musculus	-0.008	0.085
Animal	Homo sapiens	-0.004	0.031

$$\widehat{\frac{\mathrm{d}\ln(\omega)}{\mathrm{d}\ln(N_e)}} = \frac{\mathrm{Cov}[\ln(\omega), \ln(N_e)]}{\mathrm{Var}[\ln(N_e)]} \text{ from linear regression.} \tag{95}$$

$$\Rightarrow \widehat{\frac{\mathrm{d}\omega}{\omega \mathrm{d}\ln(N_e)}} = \frac{\mathrm{Cov}[\ln(\omega), \ln(N_e)]}{\mathrm{Var}[\ln(N_e)]} \tag{96}$$

$$\Rightarrow \widehat{\chi} \simeq \widehat{\omega} \frac{\mathrm{Cov}[\ln(\omega), \ln(N_e)]}{\mathrm{Var}[\ln(N_e)]} \tag{97}$$

$$\Rightarrow \chi \simeq 0.2 \frac{-0.45}{4.45} \tag{98}$$

$$\Rightarrow \chi \simeq -0.02 \tag{99}$$

5 Simulation using the 3d structure of protein

We simulated substitutions in the protein phosphatase (n = 300 codon sites). From a DNA sequence $\mathbb{S}$ after t substitutions, we compute the free energy of the folded state $G_{\mathrm{F}}(\mathbb{S})$, using the 3-dimensional structure of the folded state and pair-wise contact energies between neighboring amino-acid residues:

$$G_{\mathrm{F}}(\mathbb{S}) = \sum_{1 \leq \mathrm{i} \leq \mathrm{n}} \sum_{r \in \mathcal{N}(\mathrm{i})} I(\mathbb{S}(\mathrm{i}), \mathbb{S}(r)), \tag{100}$$

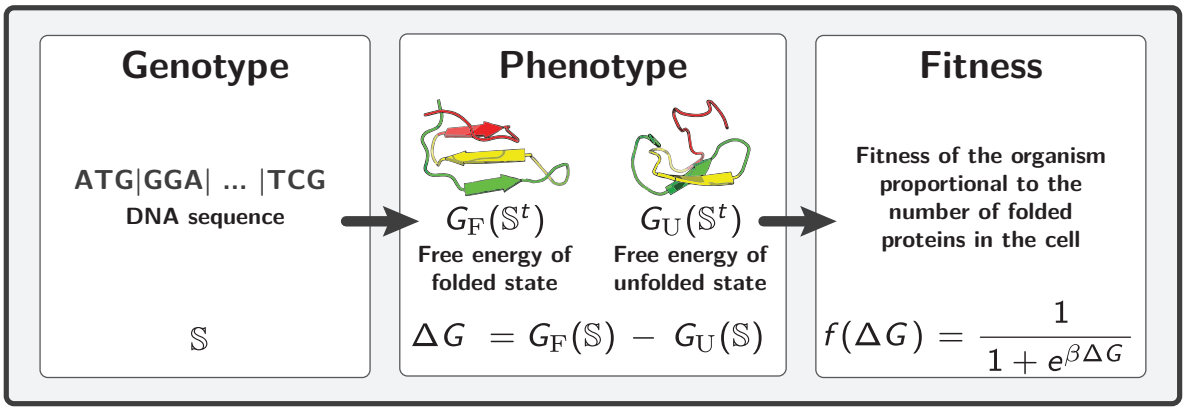
where $I(a, b)$ is the pair-wise contact energies between amino-acid a and b, using contact potentials estimated by Miya-zawa and Jernigan, and $\mathcal{N}(\mathrm{i})$ are the neighbor residues of site i (closer than 7Å) in the 3D structure.

The free energy of unfolded states $G_{\mathrm{U}}(\mathbb{S})$ is approximated using 55 decoy 3D structures that supposedly represent a sample of possible unfolded states:

$$G_{\mathrm{U}}(\mathbb{S}) = \langle G(\mathbb{S}) \rangle - kT \ln(1.0\mathrm{E}^{160}) - \frac{2\left[\langle G(\mathbb{S})^2 \rangle - \langle G(\mathbb{S}) \rangle^2\right]}{kT} \tag{101}$$

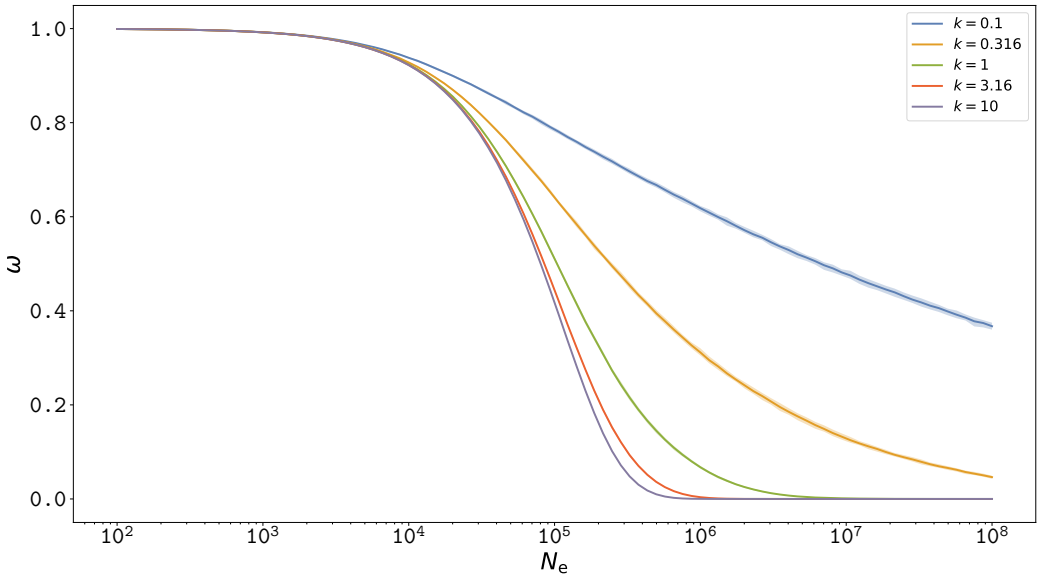
where the average $\langle . \rangle$ runs other the 55 decoy 3D structures, and k is the Boltzmann constant and T the temperature in Kelvin.
From the energy of folded and unfolded states, we can compute the difference in free energy between the states:

$$\Delta G(\mathbb{S}) = G_{\mathrm{F}}(\mathbb{S}) - G_{\mathrm{U}}(\mathbb{S}) \tag{102}$$



6 Simulated ω susceptibility to changes in N_{e}

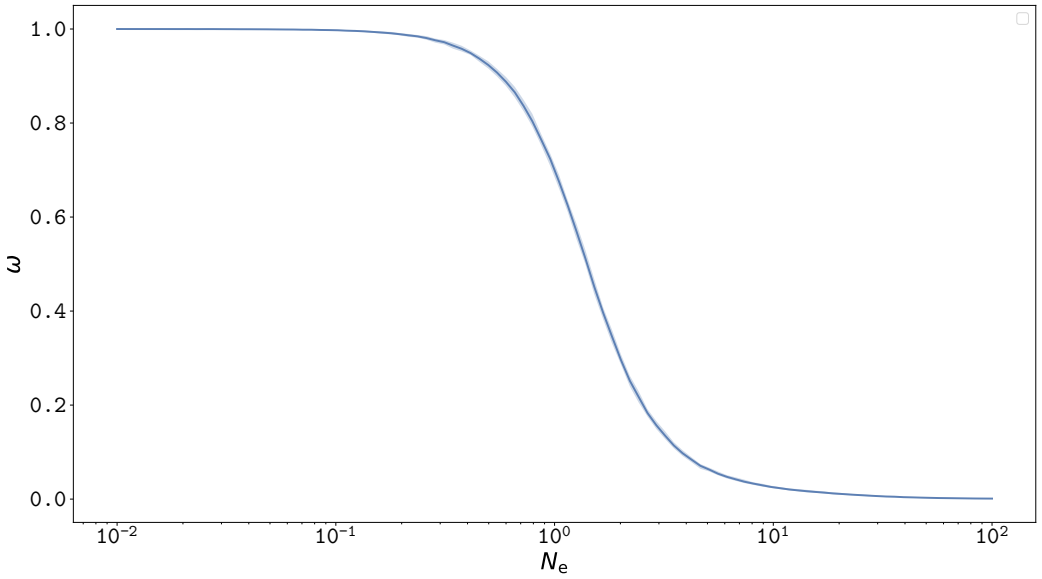
6.1 ω susceptibility with gamma distributed selection coefficient



ω at equilibrium as a function of $\boldsymbol{N_{\mathrm{e}}}$ (log scale). For each population size, 200 simulations were performed and the average (solid line) and 90% confidence interval (shaded area) are shown. In the model of gamma distributed fitness effect, ω at equilibrium is strongly dependent on log-N_{e} where the slope correlation is proportional to the inverse of the shape parameter of the gamma distribution (Welch *et al.*, 2008).

6.2 ω susceptibility with amino-acid fitness profiles

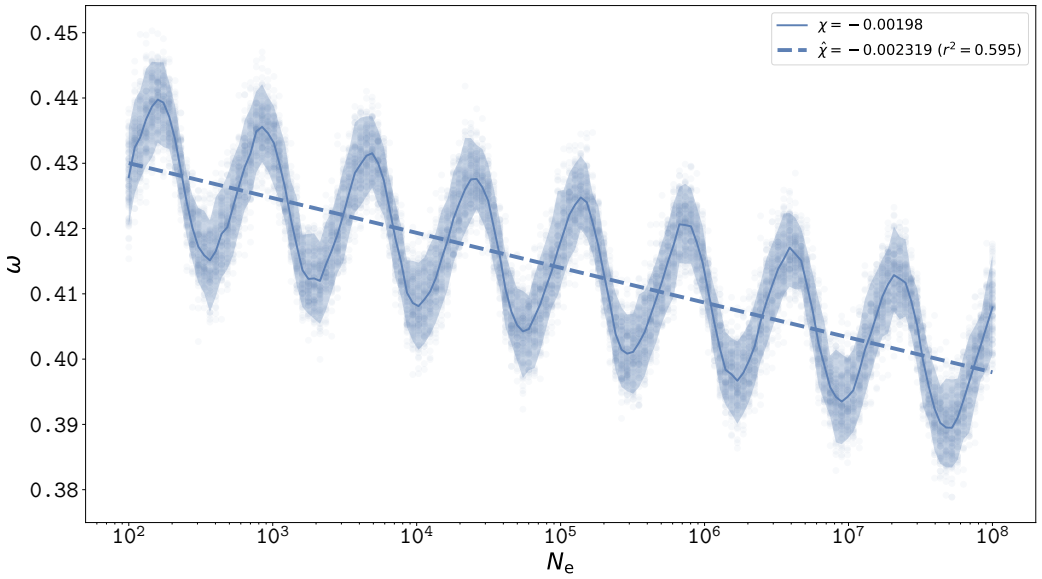
.



ω at equilibrium as a function of $\boldsymbol{N_e}$ (relative). For each population size, 200 simulations were performed and the average (solid line) and 90% confidence interval (shaded area) are shown. In the model of site-wise amino-acid fitness profiles taken from (Bloom, 2017), ω at equilibrium is strongly dependent on log-N_e.

6.3 ω susceptibility with additive free energy of folding

.

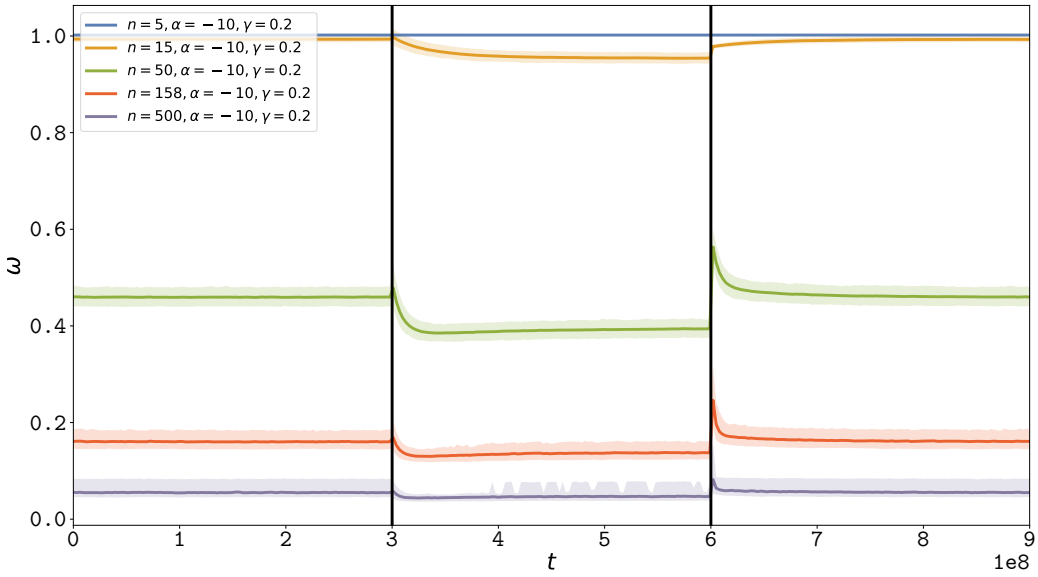


ω at equilibrium as a function of $\boldsymbol{N_e}$ (log scale). For each population size, 200 simulations were performed and the average (solid line) and 90% confidence interval (shaded area) are shown. The fixed parameters are $\alpha = -118$, $\gamma = 1$, n $= 300$, $\beta = 1.686$. The simulations of our additive free energy model match the theoretical prediction that the slope of the linear relation (dashed line) is equal to $\beta \mathrm{n} \gamma)^{-1} = 0.00198 \simeq = 0.00199$. The non-monotony is suspected to be due to the discrete number of sites and states, such that the changes in ΔG after a mutation is either -1, 0 or 1. Such non-monotony is not observed with the Grantham model, in which the ω is lower and the slope of susceptibility lower, closer to the empirical 3D model.

7 Simulated relaxation time of ω

7.1 Relaxation time of ω dependence on n

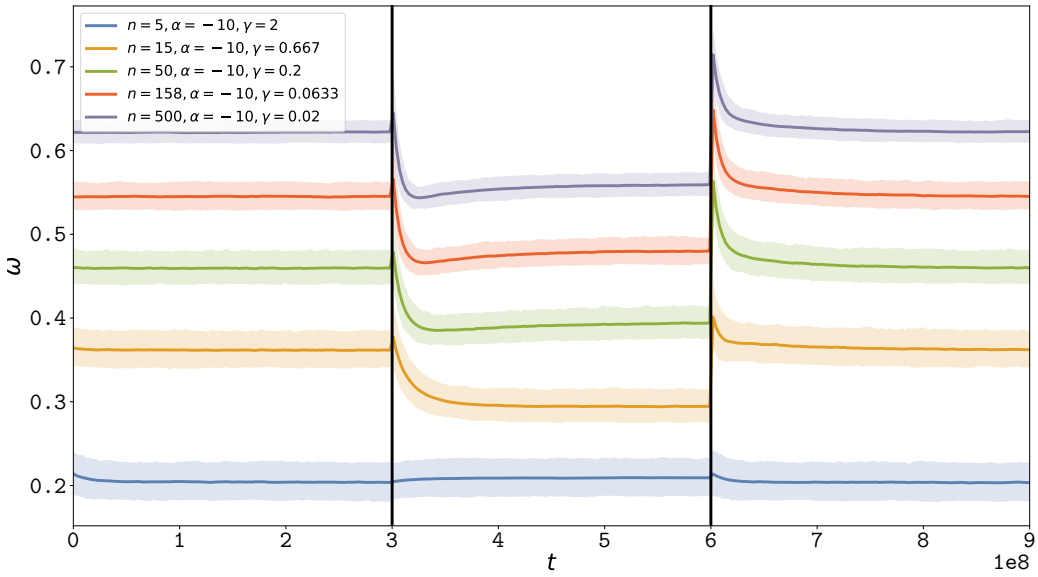
.



ω Relaxation after a brutal change in N_e, the left and right panel correspond to low N_e ($1e^5$) and the middle panel corresponds to high N_e ($2e^6$). Solid line corresponds to the average over replicates (r) and the shaded area correspond to the 90% interval among replicates. The mutation rate (μ) is $1e{-}8$ per year per site, and the total time of the computation is 900 million years. $\beta = 1.686$, $\gamma = 0.2$ and $\alpha = -10$ for all simulations. The number of sites is changed from n = 15 to n = 158, and the number of replicates is changed accordingly such that the total number of sites ($n * r$) is kept constant. Increasing n implies a higher ω at equilibrium, a lower susceptibility of the ω to changes in N_e and a higher rate of relaxation.

7.2 Relaxation time of ω dependence on n, while correction for γ

.

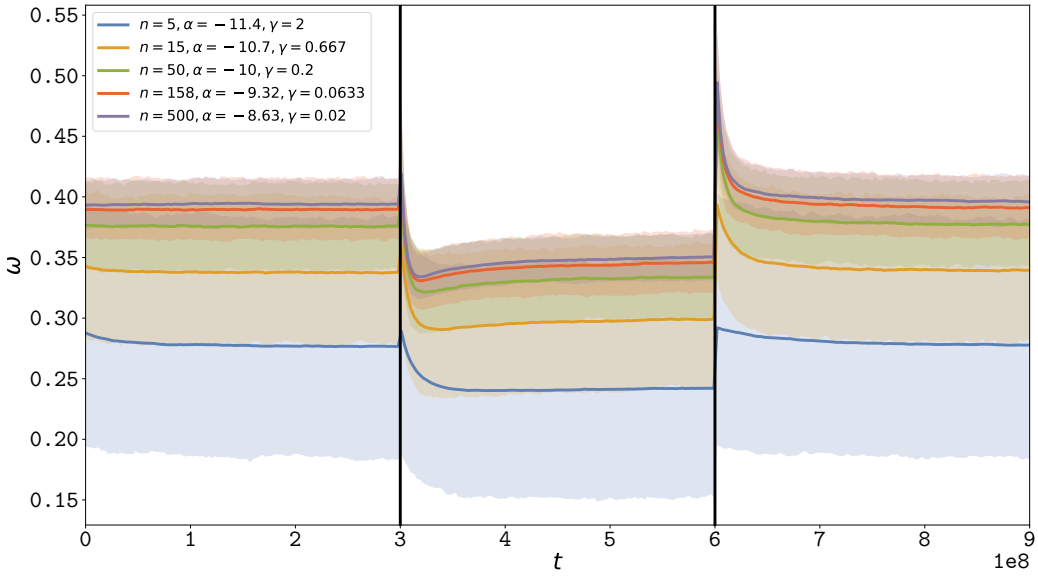


ω Relaxation after a brutal change in N_e, the left and right panel correspond to low N_e ($1e^5$) and the middle panel corresponds to high N_e ($2e^6$). Solid line corresponds to the average over replicates (r) and the shaded area

correspond to the 90% interval among replicates. The mutation rate (μ) is $1e{-}8$ per year per site, and the total time of the computation is 900 million years. $\beta = 1.686$, $\alpha = -10$ for all simulations. The number of sites is changed from n = 15 to n = 158, and the number of replicates is changed accordingly such that the total number of sites ($n * r$) is kept constant. Moreover, γ is changed according to n such that the product γn is kept constant, thus the susceptibility of the ω to changes in N_e is kept constant. Increasing n implies a higher ω at equilibrium, and a higher rate of relaxation.

7.3 Relaxation time of ω for the Grantham model

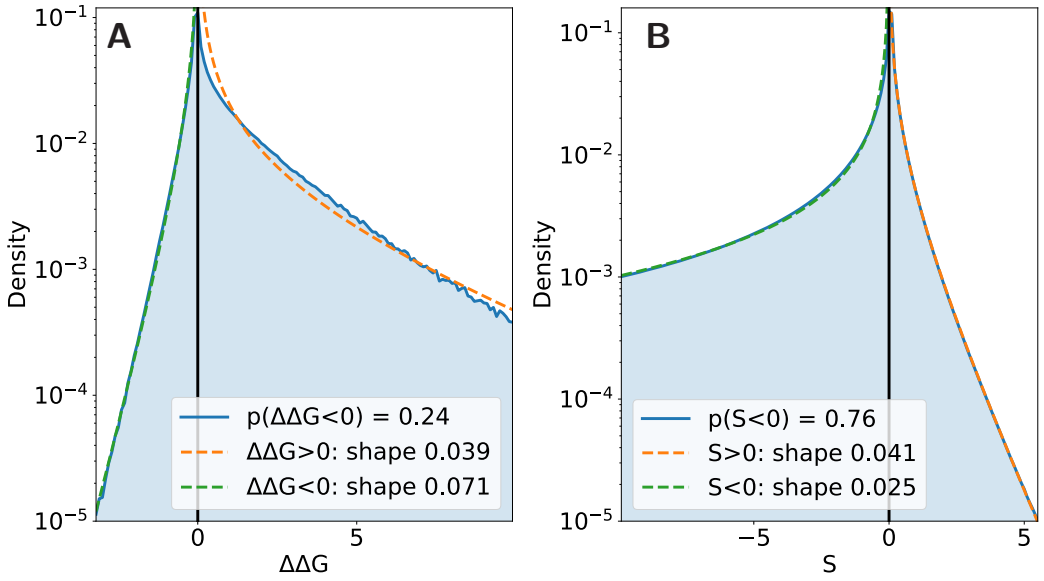
.



ω Relaxation after a brutal change in N_e, the left and right panel correspond to low N_e ($1e^5$) and the middle panel corresponds to high N_e ($2e^6$). Solid line corresponds to the average over replicates (r) and the shaded area correspond to the 90% interval among replicates. The mutation rate (μ) is $1e{-}8$ per year per site, and the total time of the computation is 900 million years. $\beta = 1.686$, $\gamma = -10$ for all simulations. The number of sites is changed from n = 15 to n = 158, and the number of replicates is changed accordingly such that the total number of sites ($n * r$) is kept constant. Moreover, γ is changed according to n such that the product γn is kept constant, thus the susceptibility of the ω to changes in N_e is kept constant. Finally, α is changed according to n and γ such that the equilibrium value x^* is kept constant, by solving numerically equation 17. Increasing n implies a higher rate of relaxation.

8 Distribution of fitness effects

DNA mutations changing a genotype can result in a change of phenotype, and ultimately a change in fitness. From a specific genotype, all the possible mutations thus result in a distribution a phenotypic effect (DPE) and fitness effects (DFE). The DPE and DFE are not known a priori, but are the resulting consequence of the mutation-selection-drift balance. Empirically, these distributions are of particular importance since they can be obtained experimentally or inferred with other data. As an example, DFE can be inferred from polymorphism dataset (Eyre-walker and Keightley, 2007; Galtier, 2016). Moreover, the distribution of $\Delta\Delta G$ for novel mutations can be obtained experimentally.



Distribution of fitness effects and phenotypic effect for novels non-synonymous mutations observed along a simulation at the mutation-selection balance. $\alpha = -118$, $\gamma = 1$, n = 300, $\beta = 1.686$, and for each non-optimal amino-acid, γ is scaled by the Grantham distance to the optimal amino-acid. Each side of the distribution is fitted to a gamma distribution, shown in dotted line. Panel A. Distribution of observed $\Delta\Delta G$, which fit adequately the gamma distribution for negative $\Delta\Delta G$ (stabilizing mutations). Panel A. Distribution of observed selection coefficient, which fit adequately the gamma distribution for both positive and negative selection coefficient. However the shape parameter estimated is not the same for positive and negative selection coefficients.

References

Beaulieu, J. M., O'Meara, B. C., Zaretzki, R., Landerer, C., Chai, J., and Gilchrist, M. A. 2018. Population Genetics Based Phylogenetics Under Stabilizing Selection for an Optimal Amino Acid Sequence: A Nested Modeling Approach. *Molecular Biology and Evolution*, 36(4): 834–851.

Bloom, J. D. 2017. Identification of positive selection in genes is greatly improved by using experimentally informed site-specific models. *Biology Direct*, 12(1): 1.

Brevet, M. and Lartillot, N. 2019. Reconstructing the history of variation in effective population size along phylogenies. *bioRxiv*, page 793059.

Drummond, D. A. and Wilke, C. O. 2008. Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. *Cell*, 134(2): 341–352.

Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O., and Arnold, F. H. 2005. Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences of the United States of America*, 102(40): 14338–14343.

Eyre-walker, A. and Keightley, P. D. 2007. The distribution of fitness effects of new mutations. *Nature*, 8(August).

Galtier, N. 2016. Adaptive protein evolution in animals and the effective population size hypothesis. *PLoS Genetics*, pages 1–23.

Goldstein, R. A. 2013. Population Size Dependence of Fitness Effect Distribution and Substitution Rate Probed by Biophysical Model of Protein Thermostability. *Genome Biology and Evolution*, 5(9): 1584–1593.

Serohijos, A. W., Rimas, Z., and Shakhnovich, E. I. 2012. Protein Biophysics Explains Why Highly Abundant Proteins Evolve Slowly. *Cell Reports*, 2(2): 249–256.

Welch, J. J., Eyre-Walker, A., and Waxman, D. 2008. Divergence and polymorphism under the nearly neutral theory of molecular evolution. *Journal of Molecular Evolution*, 67(4): 418–426.

Wilke, C. O. and Drummond, D. A. 2006. Population Genetics of Translational Robustness. *Genetics*, 173(1): 473 LP – 481.

Zhang, J. and Yang, J. R. 2015. Determinants of the rate of protein sequence evolution.

Zhang, J., Maslov, S., and Shakhnovich, E. I. 2008. Constraints imposed by non-functional protein–protein interactions on gene expression and proteome size. *Molecular Systems Biology*, 4(1): 210.