

- $s_{A \rightarrow B}$ is the selection coefficient of allele B in a population of A .
- w_A is the Wrightian fitness (survival \times fecundity) of allele A .
- $f_A = \ln(w_A)$ is the Malthusian fitness of allele A .

$$\begin{aligned}
 s_{A \rightarrow B} &= \frac{w_B - w_A}{w_A} \\
 &= \frac{w_B}{w_A} - 1 \\
 &\simeq \ln\left(\frac{w_B}{w_A}\right) \\
 &= \ln(w_B) - \ln(w_A) \\
 &= f_B - f_A.
 \end{aligned}$$

$s_{A \rightarrow B} = 0.01$ means that B is favored with a increased chance of survival or fecundity by 1%.

For a diploid population of size N_e the initial frequency of a mutation with selection coefficient s is $p = 1/2N_e$. Its probability of fixation denoted $\mathbb{P}_{\text{fix}}(s)$ is given by:

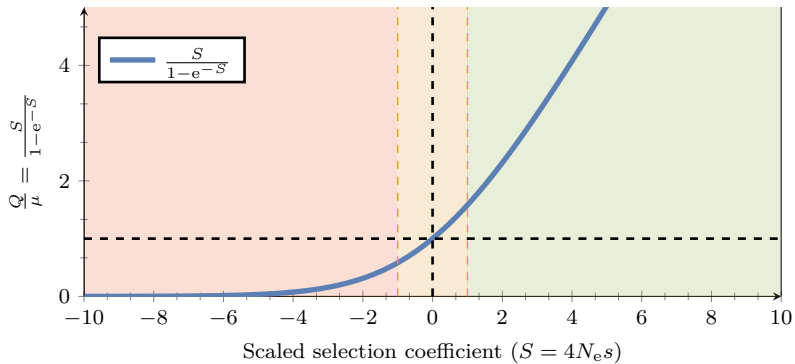
$$\begin{aligned}\mathbb{P}_{\text{fix}}(s) &= \frac{1 - e^{-2s}}{1 - e^{-4N_e s}} \\ &\simeq \frac{2s}{1 - e^{-4N_e s}}.\end{aligned}$$

The special case of a neutral allele can be obtained by taking the limit $s \rightarrow 0$:

$$\begin{aligned}\mathbb{P}_{\text{fix}}(0) &= \lim_{s \rightarrow 0} \mathbb{P}_{\text{fix}}(s), \\ &= \lim_{s \rightarrow 0} \frac{2s}{1 - e^{-4N_e s}} \\ &= \lim_{s \rightarrow 0} \frac{2s}{1 - (1 - 4N_e s)} \\ &= \lim_{s \rightarrow 0} \frac{2s}{4N_e s} \\ &= \frac{1}{2N_e}.\end{aligned}$$

- N_e is the number of diploid individuals.
- q is the substitution rate of new alleles.
- μ is the mutation rate of new alleles.
- $\mathbb{P}_{\text{fix}}(s)$ is the probability of fixation of new allele, with selection coefficient s .

$$\begin{aligned}
 q &= 2N_e \times \mu \times \mathbb{P}_{\text{fix}}(s), \\
 &= 2N_e \times \mu \times \frac{2s}{1 - e^{-4N_e s}}, \\
 &= \mu \times \frac{S}{1 - e^{-S}} \text{ with } S = 4N_e s.
 \end{aligned}$$



- $q_{\text{ATT} \rightarrow \text{ATG}}$ is the substitution rate from codon **ATT** to **ATG**.
- $\mu_{\text{T} \rightarrow \text{G}}$ is the mutation rate from nucleotide **T** to **G**.
- $S_{\text{Ile} \rightarrow \text{Met}} = 4N_e s_{\text{Ile} \rightarrow \text{Met}}$ is the scaled selection coefficient from **Ile** to **Met**.
- $F_{\text{Ile}} = 4N_e f_{\text{Ile}}$ is the scaled fitness of Isoleucine.
- $F_{\text{Met}} = 4N_e f_{\text{Met}}$ is the scaled fitness of Methionine.

$$\begin{aligned}
 q_{\text{ATT} \rightarrow \text{ATG}} &= \mu_{\text{T} \rightarrow \text{G}} \frac{S_{\text{Ile} \rightarrow \text{Met}}}{1 - e^{-S_{\text{Ile} \rightarrow \text{Met}}}} \\
 &= \mu_{\text{T} \rightarrow \text{G}} \frac{F_{\text{Met}} - F_{\text{Ile}}}{1 - e^{F_{\text{Ile}} - F_{\text{Met}}}}
 \end{aligned}$$

- $q_{i \rightarrow j}$ is the substitution rate from codon i to j .
- $\mu_{i \rightarrow j}$ is the mutation rate from codon i to j .
- F_i is the scaled fitness of codon i (F_j for codon j).

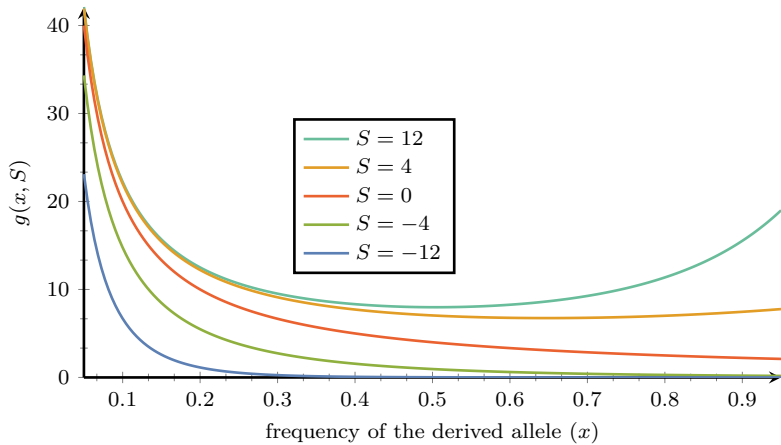
$$\begin{cases} q_{i \rightarrow j} = \mu_{i \rightarrow j} & \text{if synonymous,} \\ q_{i \rightarrow j} = \mu_{i \rightarrow j} \frac{F_j - F_i}{1 - e^{F_i - F_j}} & \text{if non-synonymous.} \end{cases}$$

Before fixation or extinction, the probability of an allele to be at a certain frequency can be related to its selection coefficient (s) and to the effective population size (N_e).

- $g(x)dx$ is the expected time for which the population frequency of derived allele is in the range $(x, x + dx)$ before eventual absorption

$$g(x, S) = \frac{(1 - e^{-2s}) (1 - e^{-4N_e s(1-x)})}{s(1 - e^{-4N_e s})x(1 - x)}$$

$$\approx \frac{2(1 - e^{-S(1-x)})}{(1 - e^{-S})x(1 - x)}$$



- n is the number of sampled chromosomes (twice the number of individuals).
- $1 \leq i \leq n - 1$ is number of chromosomes with the derived allele.
- β_i is the number of SNPs with derived allele carried by i chromosomes.
- $\beta'_i = \beta_i \times i$ is the scaled β_i .

$$H = \frac{2}{n(n-1)} \left(\sum_{i=1}^{n-1} \beta_i \times (n-i) \times i - \sum_{i=1}^{n-1} \beta_i \times i^2 \right)$$

- $H = 0$: no evidence of deviation from neutrality.
- $H > 0$: deficit of SNPs at high frequency (negative selection).
- $H < 0$: excess of SNPs at high frequency (positive selection).

$$\mathbb{P}\left[S_0\mid S\right]$$

$$\mathbb{P}\left[S\mid S_0\right]$$

$$\mathbb{P}\left[S_0\right]$$

$$\mathbb{P}\left[S\right]$$

$$\mathbb{P}\left[S_0\mid S\right]=\frac{\mathbb{P}\left[S\mid S_0\right]\mathbb{P}\left[S_0\right]}{\mathbb{P}\left[S\right]}$$

$$\mathbb{P}\left[\textcolor{blue}{S_0}>\textcolor{blue}{1}\mid S>1\right]$$

$$\mathbb{P}\left[S>1\mid \textcolor{blue}{S_0}>\textcolor{blue}{1}\right]$$

$$\mathbb{P}\left[\textcolor{blue}{S_0}>\textcolor{blue}{1}\right]$$

$$\mathbb{P}\left[S>1\right]$$

$$\mathbb{P}\left[\textcolor{blue}{S_0}>\textcolor{blue}{1}\mid S>1\right]=\frac{\mathbb{P}\left[S>1\mid \textcolor{blue}{S_0}>\textcolor{blue}{1}\right]\mathbb{P}\left[\textcolor{blue}{S_0}>\textcolor{blue}{1}\right]}{\mathbb{P}\left[S>1\right]}$$

$$\mathcal{D} := S < -1$$

$$\mathcal{N} := -1 < S < 1$$

$$\mathcal{B} := S > 1$$

$$\mathbb{P}[\mathcal{D}]$$

$$\mathbb{P}[\mathcal{N}]$$

$$\mathbb{P}[\mathcal{B}]$$

$$\textcolor{red}{\mathcal{D}}_0 := S_0 < -1$$

$$\textcolor{green}{\mathcal{N}}_0 := -1 < S_0 < 1$$

$$\textcolor{blue}{\mathcal{B}}_0 := S_0 > 1$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{Predicted Positive}}$$

$$\frac{\mathbb{P}[\mathcal{B} \cap \textcolor{blue}{\mathcal{B}}_0]}{\mathbb{P}[\textcolor{blue}{\mathcal{B}}_0]} = \mathbb{P}[\mathcal{B} \mid \textcolor{blue}{\mathcal{B}}_0]$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{Positive}}$$

$$\frac{\mathbb{P}[\mathcal{B} \cap \textcolor{blue}{\mathcal{B}}_0]}{\mathbb{P}[\mathcal{B}]} = \mathbb{P}[\textcolor{blue}{\mathcal{B}}_0 \mid \mathcal{B}]$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\frac{\mathbb{P}[S > 1 \cap \textcolor{blue}{S}_0 > \textcolor{blue}{1}]}{\mathbb{P}[\textcolor{blue}{S}_0 > \textcolor{blue}{1}]} = \mathbb{P}[S > 1 \mid \textcolor{blue}{S}_0 > \textcolor{blue}{1}]$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\frac{\mathbb{P}[S > 1 \cap \textcolor{blue}{S}_0 > \textcolor{blue}{1}]}{\mathbb{P}[S > 1]} = \mathbb{P}[\textcolor{blue}{S}_0 > \textcolor{blue}{1} \mid S > 1]$$