File S1

Purifying selection, drift and reversible mutation with arbitrarily high mutation rates

Supporting Information

Brian Charlesworth and Kavita Jain

Analytical approximations for the fraction of segregating sites

The proportion p_{seg} of segregating sites is given by

$$p_{seg} = 1 - p(n) - p(0) \tag{S1.1}$$

where

$$p(k) = \int_0^1 dq \binom{n}{k} q^k (1-q)^{n-k} \phi(q) \tag{S1.2}$$

and

$$\phi(q) = C e^{\gamma q} q^{\beta-1} (1-q)^{\alpha-1} \tag{S1.3}$$

Moments of the frequencies q and p: We first consider $p(n) = \overline{q^n}$ which is given by

$$\begin{aligned} p(n) &= \frac{(\beta)_n}{(\alpha+\beta)_n} \frac{{}_1F_1(n+\beta, n+\alpha+\beta, \gamma)}{{}_1F_1(\beta, \alpha+\beta, \gamma)} && \text{(S1.4)} \\ &= \frac{(\beta)_n}{(\alpha+\beta)_n} \frac{1 + \sum_{j=1}^{\infty} G_j^{(n)} \frac{\gamma^j}{j!}}{1 + \sum_{j=1}^{\infty} G_j^{(0)} \frac{\gamma^j}{j!}} && \text{(S1.5)} \end{aligned}$$

where

$$G_j^{(n)} = \frac{(n+\beta)_j}{(n+\alpha+\beta)_j} \tag{S1.6}$$

and $(a)_j$ is the Pochhammer's symbol. For $\alpha, \beta \to 0$ with $\kappa = \alpha/\beta$ finite, we have

$$G_j^{(n)} \approx \begin{cases} 1 - \alpha(H_{n+j-1} - H_{n-1}) \ , \ n > 0 \\ \frac{1}{1+\kappa} \ (1 - \alpha H_{j-1}) \ , \ n = 0 \end{cases} \tag{S1.7}$$

where $H_j = \sum_{k=1}^{j} (1/k)$ is the jth Harmonic number. Also, we can write

$$\frac{(\beta)_n}{(\alpha+\beta)_n} \approx \frac{1}{1+\kappa} \ (1 - \alpha H_{n-1}) \tag{S1.8}$$

Substituting the above approximations in the expression for $p(n)$ and keeping terms to order α, we finally obtain

$$p(n) \approx \frac{1}{1+\kappa e^{-\gamma}} \left(1 - \alpha H_{n-1} e^{-\gamma} - \frac{\alpha e^{-\gamma}}{1+\kappa e^{-\gamma}}(S_1(n) + \kappa S_2(n))\right) \quad \text{(S1.9)}$$

where

$$S_1(n) = \sum_{j=1}^{\infty} (H_{n+j-1} - H_{j-1}) \frac{\gamma^j}{j!} \overset{\gamma \gg 1}{\sim} \frac{e^{\gamma}}{\gamma} (n + c_1 \gamma^{-1}) \quad \text{(S1.10)}$$

$$S_2(k) = e^{-\gamma} \sum_{j=1}^{\infty} H_{n+j-1} \frac{\gamma^j}{j!} \overset{\gamma \gg 1}{\sim} \ln \gamma \quad \text{(S1.11)}$$

We note that the dependence on n appears at order α. Thus in the infinite sites model where these terms are neglected, all the moments of fraction q are equal. Setting $n = 1$ and 2 in the above equations reproduces the results for $\bar{q}$ in (7a) and (7b), and for $\overline{q - q^2}$ in (13) (after dividing by 2) given in the main text. In the neutral case, we have

$$p(n) \approx \frac{1 - \alpha H_{n-1}}{1+\kappa} \quad \text{(S1.12)}$$

while in the strong selection limit, using the asymptotic results for the sums $S_1(n)$ and $S_2(n)$, we get

$$1 - p(n) = \frac{1}{1+\kappa^{-1} e^{\gamma}} \left(1 + \frac{\alpha}{\kappa}(H_{n-1} + \frac{n e^{\gamma}}{\gamma})\right) \quad \text{(S1.13)}$$

For $\gamma \to \infty$, the above expression shows that $1 - p(n) \to \alpha n/\gamma$.

We next consider $p(0) = \overline{(1-q)^n}$ which is given by

$$p(0) = \frac{(\alpha)_n}{(\alpha+\beta)_n} \frac{{}_1F_1(\beta, n+\alpha+\beta, \gamma)}{{}_1F_1(\beta, \alpha+\beta, \gamma)} \quad \text{(S1.14)}$$

For $\alpha, \beta \to 0$ but arbitrary n and j, we can write

$$\frac{(\beta)_j}{(n+\alpha+\beta)_j} \approx \beta \frac{(j-1)!(n-1)!}{(n+j-1)!} \quad \text{(S1.15)}$$

Using the above approximation and as before, keeping terms to order α, we find that

$$p(0) \approx \frac{\kappa e^{-\gamma}}{1+\kappa e^{-\gamma}} \left(1 - \frac{\alpha}{\kappa} H_{n-1} + \frac{\alpha(n-1)!}{\kappa} S_3(n) + \frac{\alpha S_2(0)}{1+\kappa e^{-\gamma}}\right) \tag{S1.16}$$

where

$$S_3(n) = \sum_{j=1}^{\infty} \frac{\gamma^j}{j(n+j-1)!} \overset{\gamma \gg 1}{\sim} \frac{e^{\gamma}}{\gamma^n} \tag{S1.17}$$

In the case of neutrality, we have

$$p(0) = \frac{\kappa - \alpha H_{n-1}}{1+\kappa} \tag{S1.18}$$

and in the strong selection limit, we get

$$p(0) = \frac{1}{1+\kappa^{-1} e^{\gamma}} \left(1 - \frac{\alpha}{\kappa} (H_{n-1} - \frac{(n-1)! e^{\gamma}}{\gamma^n})\right) \tag{S1.19}$$

For $\gamma \to \infty$, the fraction $p(0) \to \alpha(n-1)!/\gamma^n$.

Segregating site fraction (p_{seg}): Using the above results, we can now look at the behavior of p_{seg}. For $\gamma = 0$, both $p(0)$ and $1 - p(n)$ contribute equally (in magnitude) to give

$$p_{seg} = \frac{2\alpha H_{n-1}}{1+\kappa} \tag{S1.20}$$

Since $H_n \sim \ln n + \gamma_{EM}$ for large n, the proportion of segregating sites increases logarithmically with the sample size in the neutral case. For $\beta = 0.02$, the above expression gives $p_{seg} = 0.094$ and 0.156 for $n = 20$ and 200 respectively which are close to the data in Table 1 of the main text. In the strong selection limit, for large γ, we have

$$p_{seg} \approx \frac{\alpha n}{\gamma} \tag{S1.21}$$

which increases linearly with the sample size.

One can also look at the $\beta \to \infty$ limit. For the neutral case, we have

$$p_{seg} = 1 - \frac{(\alpha)_n + (\beta)_n}{(\alpha+\beta)_n} \tag{S1.22}$$

For $n \ll \alpha, \beta$, we can write

$$\frac{(\alpha)_n}{(\alpha+\beta)_n} \approx \left(\frac{\kappa}{1+\kappa}\right)^n \tag{S1.23}$$

while for $n \gg \alpha, \beta$, using Stirling's approximation $s! \sim \sqrt{2\pi s}(s/e)^s$, we get

$$\frac{(\alpha)_n}{(\alpha+\beta)_n} \approx \frac{(\alpha+\beta-1)!}{(\alpha-1)!}\; n^{-\beta} \tag{S1.24}$$

Using these approximations, we find that

$$1 - p_{seg} = \begin{cases} \frac{1+\kappa^n}{(1+\kappa)^n}\;,\; n \ll \alpha, \beta \\ (\alpha+\beta-1)!\; \left(\frac{1}{(\alpha-1)! n^{\beta}} + \frac{1}{(\beta-1)! n^{\alpha}}\right)\;,\; n \gg \alpha, \beta \end{cases} \tag{S1.25}$$

Thus in small samples (relative to scaled mutation rates), p_{seg} approaches unity exponentially fast while for larger samples, the approach is algebraic.