File S1

Supporting Information

S1 Equating the distribution of variant frequencies across sites and the probability distribution of variant frequencies

This section concerns the adequacy of the approximation of equating the mean and variance of the trait for a given population, determined by the distribution of variant frequencies among m sites, to their expectations obtained from the overall probability distribution of variant frequencies generated by the stochastic process of mutation, selection and drift. For a set of m independent exchangeable sites, the mean of q for a given population is $\bar{q} = (\Sigma_i\, q_i)/m$, where q_i is the frequency of A_2 at a given site i. Let q have expectation q^* and variance σ^2_q; $\bar{q}$ has expectation q^* and variance σ^2_q/m. From Equation 4, the mean phenotypic value is given by $\bar{z} = a\Sigma_i\,(2q_i - 1)$, with expectation $ma(2q^* - 1)$ and variance $\sigma^2_z = 4ma^2\sigma^2_q$.

The ratio of the standard deviation of $\bar{q}$ to q^* is thus equal to $\sigma_q/(q^*\sqrt{m})$, which becomes indefinitely small as m increases, provided that σ_q/q^* is of order one, which must be true when q^* is non-zero. This means that random fluctuations in $\bar{q}$ relative to its expected value become extremely small as m increases. It is therefore seems reasonable to treat $\bar{q}$ as equivalent to q^* when m is large, as is implicitly done in the standard models of codon usage bias that assume directional selection, mutation and drift (Li 1987; Bulmer 1991; McVean and Charlesworth 1999). This argument can also be applied to the model of mutation and directional selection with epistasis examined in the main text, where the mean value of the trait, $\bar{n}$, is equal to $2m(1 - \bar{q})$. In this case, the ratio of the standard deviation of $\bar{n}$ to its expectation, n^*, also approaches zero as m increases, so that $\bar{n}$ can be equated to n^* with large m.

The genetic variance, V_g, has expectation $V_g^* = 2a^2\,\mathrm{E}\{\Sigma_i\, q_i(1- q_i)\} = ma^2\pi^*$, where π^* is the expectation of the diversity at a given site i, given by $\pi_i = 2q_i(1 - q_i)$. In the case of neutrality, the result that $\pi^* = 8Nu\kappa/(1 + \kappa)$ for the state-dependent mutation model at the infinite sites limit (Charlesworth and Charlesworth 2010, p.274) yields $V_g^* = 8Nu\kappa ma^2/(1 + \kappa)$ at stationarity. Furthermore, the variance of V_g is equal to $ma^4\sigma^2_\pi$, where σ^2_π is the variance of π over the probability distribution of q; $\sigma^2_\pi \approx \pi^*/3$ in the case of stationarity and neutrality (Tajima 1983).

The ratio of the standard deviation of V_g to V_g* is therefore equal to $\sigma_\pi/(\pi^*\sqrt{m})$. Provided that σ_π/π^* is of order 1, these results suggest that it is reasonable to equate V_g to V_g* for large m, since the neutral expression gives a good approximation to the expected variance when selection is weak, as can be in Tables 1-3. This result applies both to the stabilizing selection and purifying selection models considered here.

These arguments show that the mean and variance of the allele frequencies across sites in a given population, the main quantities of interest for this paper, are close to those generated by the overall probability distribution of allele frequencies, provided that the assumption of independence among sites within a population is met. As mentioned in the main text, simulations of multi-locus models support the assumption of only a minor effect of linkage disequilibrium among variants within a population, provided that recombination rates among nearby sites are sufficiently high in relation to the strength of selection (Bürger 2000, p.276). However, the question of what happens when recombination is rare or absent is relevant to the important general problem of the effect of restricted recombination on evolutionary processes (Charlesworth *et al.* 2010), and will probably require simulation studies.

Even with linkage equilibrium, however, the population mean at a given time enters into the expression for the change in variant frequencies at each site, for both the stabilizing selection model and for the purifying selection model with epistasis. This means that variant frequencies are not strictly independent of each other in terms of the overall evolutionary process, even with linkage equilibrium. But with a large number of sites, the state of a given site i has only a small effect on the trait mean, of the order of $1/m$. This suggests that, with sufficiently large values of m, the population mean can be treated as independent of the value of the allele frequency at a given site, so it should be valid to ignore this source of non-independence (see Section S2 for further discussion of this point).

In addition, in the case of stabilizing selection, there is the problem that the term in brackets in Equation 2 of the main text involves $\delta = (z_0 - \bar{z})/a$, whose expected value is close to zero when NSa^2 is sufficiently large (see section S4). This implies that fluctuations in δ around its expectation, δ^*, could be so large that we cannot legitimately replace δ by δ^*, as was done in the derivation following Equations 6 and 12.

Examination of this question requires knowledge of the variance in $z_0 - \bar{z}$ generated by drift. An expression for this is given in section S3 below, using the approach of Lande (1976) and Bürger and Lande (1994), which assumes that there is an indefinitely large number of sites influencing the trait (the "infinitesimal model"). There is then a stationary, normal distribution of the values of $z_0 - \bar{z}$ among independent realizations of the stochastic process, with approximate standard deviation $1/\sqrt{(4NS)}$. This implies that

$$\sigma_\delta \approx \frac{1}{\sqrt{(4NSa^2)}} \qquad \text{(S1.1)}$$

The stationary distribution of δ, $\phi^*(\delta)$, is normal, with expectation δ^* and a standard deviation given by Equation S1.1. Importantly, σ_δ is independent of m, in contrast to the result for the mean allele frequency, showing that fluctuations in δ may indeed be important regardless of the number of sites involved. The argument used in the main text showed that, for the state-dependent mutational model of stabilizing selection, we have $\delta^* \approx [\ln(\kappa) + 2b]/(8NSma^2)$ (see Equations A3b and A7). Fluctuations around δ^* will be unimportant if this quantity is several times σ_δ, since then δ and δ^* will always be close. This condition is satisfied when

$$\frac{[\ln(\kappa) + 2b]}{4\sqrt{NSa^2}} \gg 1 \qquad \text{(S1.2)}$$

For the state-independent mutational model of stabilizing selection, the term in $2b$ is omitted (Equation 13b).

The cases shown in Tables 1 and 2 with $N = 50$ satisfy this requirement, whereas it is violated for the others. Nevertheless, there is still good agreement between the formulae based on equating δ and δ^* and both the simulation and matrix results. This raises the question of why this should be, which is examined in the next section.

S2 Use of δ^* instead of δ in the analytical and numerical models of stabilizing selection

Consider an ensemble of independently evolving populations, each with a potentially different value of δ at any given time t. Assume that the probability density of value $\delta = \delta_k$ at time t for the kth population is $\phi(\delta_k, t)$; for this population, let the probability that a random site i has allele frequency q_i be $f_k(q_i, t)$. The probability of transition from q_i to $q_i + \varepsilon_i$ is a binomial deviate, with parameter q_i plus the deterministic change in allele frequency (given by Equations 2 and 5 of the main text); this change is dependent on δ_k, which in turn is determined by the set of allele frequencies over all the sites for the population in question, as given by Equation 3.

The overall probability density of finding frequency $q_i + \varepsilon_i$ at time $t + 1$ is thus obtained by summing the transition probabilities for all δ_k values, and multiplying each of these by the probability density of finding allele frequency q_i at time t in population k, denoted by $g_k(q_i, t)$. But, by the argument made in Section S1, the state of a single site has a negligible effect on the population mean when m is very large, so that we can write

$$g_k(q_i, t) \approx \phi_k(\delta_k) g(q_i, t) \qquad \text{(S2.1)}$$

where $g(q_i, t)$ is the overall probability density for q_i at time t.

Hence, the transition probability for q_i at time t changing to $q_i + \varepsilon_i$ at time $t + 1$ is given by the integral over the distribution of δ of the transition probabilities q_i to $q_i + \varepsilon_i$ for each δ_k, each weighted by $\phi(\delta_k)$. This is what should properly be used in the calculations, instead of the fixed value involving the expectation δ^*.

The following argument shows, however, that the use of δ^* is legitimate, provided that the usual assumptions of diffusion theory are met (i.e., all evolutionary forces are sufficiently weak that their second-order terms are negligible– Ewens 2004, Chapter 4). The subscript i can be dropped, since the sites are exchangeable. The forward diffusion equation for population k is then

$$g_k(q, t) - g_k(q, t-1) \approx -\frac{\partial(g_k \Delta q_k)}{\partial q} + \frac{1}{(4N)} \frac{\partial^2 [q(1-q) g_k]}{\partial q^2} \qquad \text{(S2.2)}$$

where Δq_k is the expected change in allele frequency in population k, given current frequency q.

Only the first term on the right-hand side of Equation S2.2 depends on Δq_k and hence δ_k. The relevant partial derivative can be written as

$$g_k \frac{\partial \Delta q_k}{\partial q} + \Delta q_k \frac{\partial g_k}{\partial q}$$

Writing $g_k(q, t) = \phi(\delta_k)g(q, t)$, Equation S.2.2 becomes

$$\phi(\delta_k)\{g(q)\frac{\partial \Delta q_k}{\partial q} + \Delta q_k \frac{\partial g(q)}{\partial q}\} \qquad \text{(S2.3)}$$

From Equation 2 of the main text, the term in braces can be seen to be a linear function of δ_k. This establishes that the diffusion operator is linear in δ_k; hence, if we take its expectation over the distribution of δ_k, we obtain an expression that depends only on δ^*. It follows that the use of δ^* in the analytical approximations and the matrix equation will be accurate, under the usual conditions for the validity of diffusion equations. The same argument applies to the backward diffusion equation, which is used to obtain the expressions for fixation probabilities and diversities used in the main text.

An alternative argument can be applied to the matrix equation used in the numerical calculations, as described in the Appendix. Since this is equivalent to the diffusion equation when the conditions for the latter to be valid are satisfied, the results must also apply to the latter. For a given value of δ, we can write the transition matrix as $\mathbf{A}(\delta)$. The dependence on δ is mediated by the set of deterministic changes in allele frequencies across sites, given by Equation 2. Let $\Delta q_s(\delta, q)$ be the expected change in allele frequency due to selection, for a given frequency q. We can expand $\mathbf{A}$ in a Taylor series around the neutral value, $\mathbf{A}_0$, for which $\Delta q_s = 0$ for all q. This expansion involves the sum over n and all permissible values of q (i.e., 0, $1/(2N)$, … , 1) of the product of $\Delta q_s(\delta, q)^n/n!$ and the nth order partial differential coefficient of $\mathbf{A}$ with respect to $\Delta q_s(\delta, q)$. If selection is sufficiently weak, we should be able to ignore all such terms for $n > 1$.

From Equation 2, the linear dependence of Δq on δ means that we can write

$$\Delta q_s(\delta,\ q) = \Delta q_s(\delta^*, q) + (\delta\ - \delta^*)2q(1-q)Sa^2$$

so that

$$\mathbf{A}(\delta) \approx \mathbf{A}_0 + \sum_q [\Delta q_s(\delta^*, q) + (\delta - \delta^*)2q(1-q)Sa^2] \left(\frac{\partial \mathbf{A}}{\partial \Delta q_s}\right)_0 \qquad \text{(S2.4)}$$

where the summation is taken over all permissible values of q, and the derivative is evaluated at $Sa^2 = 0$.

Using the discrete probability equivalent of Equation S2.1, the matrix that represents the net change between generations in the probability vector **f** is the expectation of **A**(δ) over the distribution of δ values. The linearity of Equation S2.4 in δ – δ* immediately implies that that only the terms in $\mathbf{A}_0$ and $\Delta q_s(\delta,\ q)$ remain after taking this expectation. It follows that the only substantial contribution to the Taylor expansion of **A** around $\mathbf{A}_0$ is the first-order term given by the sum over sites of the terms in $\Delta q_s(\delta^*)$; this is equivalent to **A**(δ*) to the order of the approximations used here. This implies that we can replace δ by δ* in **A** in order to generate the probability distribution of allele frequencies without significant error, as has been done when generate the numerical results from the matrix method displayed in Tables 1, 2 and 4.

Very similar reasoning can be used to arrive at similar conclusions for the case of purifying selection with epistasis represented by the quadratic model of Equation 13, since the selection coefficient for a individual variant is a linear function of $\bar{n}$ (see Equation A11a).

S3 Use of the infinitesimal model to obtain results on the outcome of drift, mutational bias and stabilizing selection

An alternative approach to the models with stabilizing selection is to use the infinitesimal model of Lande (1976) (see Section S1). The state-independent mutational model will be considered first, since it is somewhat simpler to analyze. A forward diffusion equation for the trait mean in a given generation, $\bar{z}$ can be derived, using the expected change in mean for a given value of $\bar{z}$,

$M_{\delta z}$, and the variance in $\bar{z}$ generated by one generation of drift, $V_{\delta z}$. In the present case, and using Equation 2 (but neglecting the terms in $2q_i - 1$ compared with 2δ), we have

$$M_{\delta z} = 2V_g\, S(z_0 - \bar{z}) + 2mua(1 - \kappa) \qquad \text{(S3.1)}$$

where the second term on the right-hand side represents the effect of mutational bias on the trait mean over one generation (under the state-independent mutation model, the expected rate of occurrence of mutations that each increase the current mean by a when heterozygous is $2mu$ per individual per generation, and the expected rate for mutations that each decrease it by a is $2m\kappa u$).

Following Lande (1976), we have $V_{\delta z} = V_g/N$, where V_g is the current value of the genetic variance. Further progress requires making the assumption that fluctuations in V_g can be ignored, so that we can replace with its expectation V_g*. The diffusion representation is then approximated by an Ornstein-Uhlenbeck process, with variance V_g*$/N$, and change in mean given by replacing V_g with V_g* in Equation S3.1 (Lande 1976; Lande and Bürger 1994). Standard results for the Ornstein-Uhlenbeck process imply that the stationary distribution of $\bar{z}$ is normal, with variance $1/(4NSa^2)$ and expectation

$$\bar{z}^* = z_0 + \frac{mua(1 - \kappa)}{(V_g\, S)} \qquad \text{(S3.2)}$$

An upper bound to V_g* is provided by the neutral case; with the state-independent mutational model at the infinite sites limit, $V_g^* = 4Nmu(1 + \kappa)a^2$ (see main text). Use of this expression in Equation S3.2 gives

$$\bar{z}^* = z_0 + \frac{(1 - \kappa)}{(1 + \kappa)(4NSa)} \qquad \text{(S3.3a)}$$

which implies that

$$\delta^* = \frac{(\kappa - 1)}{(1 + \kappa)(4NSa^2)} \qquad \text{(S3.3b)}$$

Similarly, the variance of δ is

$$\sigma_\delta^2 = \frac{1}{4NSa^2} \qquad \text{(S3.4)}$$

At first sight, the result for δ* is very different from that in Equation 13b, δ* = ln(κ)/(8NSa^2). We can, however, write (κ – 1) = η, so that 1 + κ = 2(1 + η/2) yielding (κ – 1)/(1 + κ) = [η – 0.5η^2 + 0.25 η^3 – ....]/2 when η < 2. This is close to ln(κ)/2 when η < 1, and is slightly smaller than ln(κ)/2 in general. For example, with κ = 2 and 4, (κ – 1)/[(1 + κ)] = 0.33 and 0.60, respectively, instead of 0.34 and 0.70 for ln(κ)/2.

There is therefore reasonably good agreement between the expressions for δ*, derived using these two different methods of approximation, and with the results of the stochastic simulations, although the results derived in the main text fit the simulation results considerably better than those from the infinitesimal model. The reason for the discrepancies is unclear, but presumably reflects the neglect of the contributions from the sum of terms involving $2q_i - 1$in Equation 2 to the change in $\bar{z}$ in the infinitesimal model, and the use of the neutral expectation for V_g. The qualitative behaviors of the two expressions for δ* as functions of κ and NSa^2 are, however, very similar.

A similar argument can be used for the state-dependent mutational model. The mutational term in $M_{\delta z}$ is, however, more complex. Using the infinite sites limit, let the overall frequency of sites fixed for A_2 be q*; at these sites, mutations to A_1 occur at rate κu per site. From Equation 4, these cause an expected change in $\bar{z}$ of $-2m\kappa uaq$*. Similarly, mutations at sites fixed for A_1 occur at rate u, resulting in an expected change of $2mua(1 - q^*)$. The net expected mutational change in $\bar{z}$ is thus $2mua[1 - (1 + \kappa)q^*]$. As before, we can approximate V_g by its neutral expected value, which in this case is equal to $8Nu\kappa ma^2/(1 + \kappa)$ (see section S1). This gives the equivalent of Equation S3.4 as

$$\delta^* = \frac{[(1 + \kappa)q^* - 1](1 + \kappa)}{8N\kappa Sa^2} \qquad \text{(S3.5a)}$$

For large m, $q^* \approx (1 + b)/2$, where $b = z_0/(ma)$ (see Equation 8). Substituting this into Equation S3.6a, we find that

$$\delta^* \approx \frac{(1+\kappa)}{\kappa} \frac{[(1+\kappa)b + \kappa - 1]}{16NSa^2} \qquad \text{(S3.5b)}$$

The behavior of δ^* is similar to that predicted by Equations A3, although this expression somewhat overestimates δ^* compared with these approximations and with the simulations.

The behavior of the infinitesimal model also yields some insights into why equating δ and δ^* seems to work so well. When mutational bias is absent, Bürger and Lande (1994) used established properties of the Ornstein-Uhlenbeck process to show that the timescale over which the temporal autocorrelation in the mean decays is of the order of $1/(2SV_g^*)$, using the present notation. This provides a timescale over which the fluctuations in δ will tend to average out, denoted by T_δ. Using the neutral approximation for V_g^*, for the state-independent mutational model, we obtain

$$T_\delta \approx \frac{\delta^*}{mu(1+\kappa)\ln(\kappa)} \qquad \text{(S3.6)}$$

With quasi-neutrality, the timescale over which variant frequencies change is $T_d \approx 4N$. The ratio of the two timescales is thus

$$\frac{T_d}{T_\delta} \approx \frac{4Nmu(1+\kappa)\ln(\kappa)}{\delta^*} \qquad \text{(S3.7)}$$

When T_d/T_δ is approximately 1 or more, variants are likely to experience the whole range of fluctuations of δ around δ^* during their sojourn in the population, so that we can expect the effects of these to average out when affecting its fixation probability. This condition is met for the cases with N of 100 or more in Table 2, but not for $N = 50$. However, in the latter case, the argument presented in section S1 shows that the standard deviation of δ is considerably smaller

than δ* (as can be seen in Table 2), so that the fluctuations would be expected to have relatively small effects compared with those for the other11N values.

A similar relation can be derived for the state-dependent model, except that the approximation derived above for the mutation term implies that $2Nmu[(1 + \kappa)b + \kappa - 1][2b + \ln(k)]$ is used in the numerator of the equivalent of Equation S3.7.

S4 Values of δ and V_g for large values of NSa^2

This section examines the values of δ and V_g for values of NSa^2 that are sufficiently large that most sites are skewed to a high frequencies of either A_1 or A_2 type variants, when there is a predominance of stabilizing selection ($2\delta < 1$). The case of state-independent mutations will be considered first. Here, A_2-type mutations occur at rate u each generation, and A_1-type mutations occur at rate κu. Under the infinite sites assumption, a site will segregate for at most one of these two types of mutation. Rare A_2-type mutations are selected against with net selection coefficient $Sa^2(1 - 2\delta)$, since the directional selection component opposes the effect of stabilizing selection in Equation 2; rare A_1-type mutations are selected against with net selection coefficient $Sa^2(1 + 2\delta)$, since directional and stabilizing selection reinforce each other. The changes in frequencies of rare A_2-type mutations due to the stabilizing selection component of the right hand side of Equation 2 cause a net change in δ of approximately $2\bar{q}_2\,(Sa^2)$, where $\bar{q}_2$ is their mean frequency. Similarly, rare A_1-type mutations cause a net change in δ due to stabilizing selection of $-2\bar{q}_1\,(Sa^2)$, where $\bar{q}_1$ is their corresponding mean frequency. We also need to include the change in δ caused by the directional selection component of Equation 2 as well as the mutational component, as was done in deriving Equation S3.1. We obtain

$$\Delta\delta \approx -2V_g S\,\delta + 2m(Sa^2)\,(\bar{q}_2 - \bar{q}_1) + 2mu(\kappa - 1) \qquad \text{(S4.1)}$$

If the infinite sites assumption holds, the mean variant frequencies will be close to their infinite population equilibrium values under selection and mutation, $q_2^* = u/[(1 - 2\delta)(Sa^2)]$ and $q_1^* = \kappa u/[(1 + 2\delta)(Sa^2)]$ In addition, the expected diversity at each class of site can approximated by

the appropriate deterministic formula for mutation selection balance (see Charlesworth and Charlesworth 2010, p.278, Equation B6.7.3); multiplication of these by a^2 yields the expected variance, V_g*, as before. Taking each class of mutation into account, we obtain the following expression

$$V_g^* \approx \frac{2mu[\kappa(1-2\delta)+1+2\delta]}{S(1-4\delta^2)} \qquad \text{(S4.2)}$$

Substituting the equilibrium expressions for $\bar{q}_1$ and $\bar{q}_2$ into Equation S4.1, setting $\Delta\delta$ to zero, multiplying top and bottom by $1 - 4\delta^2$, and cancelling common factors, we obtain the equilibrium equation

$$0 = -[\kappa(1-2\delta)+1+2\delta]\delta + [1+2\delta-\kappa(1-2\delta)] + (\kappa-1)(1-4\delta^2) \qquad \text{(S4.3)}$$

The constant term in this quadratic expression in δ is equal to zero. It therefore has one root of zero, and the other given by the remaining terms, which yields the alternative equilibrium solution $\delta^* = (\kappa + 1)/[2(\kappa - 1)]$. However, this implies $\delta^* > ½$ with $\kappa > 1$, and so δ^* lies outside the permissible range for the present analysis.

The equilibrium $\delta^* = 0$ is locally stable, as can be seen informally as follows. Consider what happens when δ is perturbed upwards from the equilibrium with $\delta = 0$, with an accompanying arbitrary small perturbation to V_g. This has the effect of introducing a negative first term into the expression for $\Delta\delta$ given by Equation S4.1. Similarly, an increase in δ implies a decrease in the contribution from the term in $\bar{q}_2 - \bar{q}_1$, so that this quantity is reduced below its equilibrium value for $\delta = 0$. Since the mutational term is unchanged, the net result is to cause $\Delta\delta$ to become negative, and so δ will move back towards zero.

A similar approach can be used for the state-dependent model. Applying the approach in the Appendix for obtaining Equations A1 and A7, with large m and NSa^2 the fractions of sites with mutations with high frequencies of A_1 and A_2 can be approximated by $(1 - b)/2$ and $(1 + b)/2$, respectively; these sites generate rare A_2-type and rare A_1-type mutations at rates u and κu. The corresponding net changes in δ at equilibrium due to stabilizing selection at each type of site

are then $mu(1 - b)/(1 - 2\delta)$ and $-(1 + b)\kappa u/(1 + 2\delta)$, respectively, yielding a total contribution of $mu[(1 - b)(1 + 2\delta) - \kappa(1 + b)(1 - 2\delta)]/(1 - 4\delta^2)$. The equilibrium variance is now given by

$$V_g^* \approx \frac{mu[\kappa(1 + b)(1 - 2\delta) + (1 - b)(1 + 2\delta)]}{S(1 - 4\delta^2)} \qquad \text{(S4.4)}$$

and the mutational term (as in the derivation of Equations S3.5) is equal to $mu[\kappa - 1 + b(1 + \kappa)]$.

The equation for equilibrium analogous to Equation S4.3 is now

$$0 = -[\kappa(1 + b)(1 - 2\delta) + (1 - b)(1 + 2\delta)]\delta + [(1 - b)(1 + 2\delta) - \kappa(1 + b)(1 - 2\delta)] + [\kappa - 1 + b(1 + \kappa)](1 - 4\delta^2)] \qquad \text{(S4.5)}$$

A similar analysis to the above shows that the constant terms again sum to zero, so that there is a root $\delta^* = 0$. The other root is $\delta^* = [\kappa(1 + b) + 1 - b]/\{2[\kappa(1 + b) + b - 1]\}$. Similar remarks apply to the existence and stability of these equilibria as in the state-independent case.

S5 Conditions for validity of the approximations for the stabilizing selection model

An important issue concerns the conditions under when the approximations described in the main text for obtaining the the results presented in Tables 1 and 2 break down. This is expected to happen when the stabilizing selection term in Equation 2, $2q_i - 1$, becomes dominant over the term in $2\delta^*$. Since the magnitude of $2q_i - 1$ is always < 1, Equation 13b for the state-independent model implies that a sufficient condition for this is $\ln(\kappa)/(4NSa^2) < 1$. The parameter $\zeta = 4NSa^2$ should thus play a critical role in controlling the outcome of the process; if $\zeta > \ln(\kappa)$, there is the potential for net selection against A_2 over the part of the distribution of q_i values for which $q_i << 1$, and selection in favor of A_2 over the rest of the distribution. None of the parameters shown in Tables 1 or 2 satisfy this condition. Even when $\zeta > \ln(\kappa)$,the argument leading up to Equation A7 shows that the formulae for δ^* in terms of κ and NSa^2 should still apply, provided that the approximation of treating δ as fixed at δ^* is valid; with $NSa^2 >> 1$, the deterministic formulae for V_g^* should provide good approximations.

Values of N, S and a that cause ζ to fall well above the critical value were therefore chosen for further numerical study by stochastic simulations (Table S2). Values of γ are not shown, because a single γ value is not meaningful in these cases. Because the fluctuations in δ relative to δ^* are very large here, 4000 replicate runs were carried out for each parameter set. It will be seen that the deterministic approximations for V_g^* are quite accurate, although there is a tendency for them to slightly underestimate the true values, as would expected from the effects of drift; the analytical approximations for the state-dependent model also tend to underestimate δ^* somewhat.

The small values of δ^* in these cases compared with the results in Tables 1 and 2 (between 0.02 and 0.06) are consistent with the argument given in the Supplementary Information, Section S4, as well as with the results of Waxman and Peck (2003) and Zhang and Hill (2008). The intuitive basis for this approach of δ^* to zero with large NSa^2 is that an examination of the contributions to the net change in δ per generation from the two components of the bracketed term on the right hand-side of Equation 2 (i.e., from 2δ and from $2q_i - 1$) shows that the contribution from the second (stabilizing selection) term approximately counteracts the contribution from mutation. This leaves a net contribution from the directional selection term, which is equal to $-2V_gS\delta$ (see Equation S4.1). Hence, for an equilibrium to be achieved in a infinite population, δ must be close to zero.

Selection in these cases is largely driven by the stabilizing selection component of Equation 2. This implies that γ_1 (for rare A_1 mutations) and γ_2 (for rare A_2 mutations) are both negative. But with mutational bias there is a predominance of rare A_1 mutations segregating in the population, as opposed to rare A_2 mutations. This causes the shape of the pooled frequency spectrum to differ considerably from the U-shape with pure stabilizing selection (Kimura 1983, p.147), although there is a slight upturn in the frequency spectrum at low frequencies of A_2 mutations with a low level of mutational bias. An example is shown in Figure S1. In contrast, the unfolded frequency spectra for derived A_1 and A_2 variants show selection against each of them. These differences from the case when there is net directional selection or pure stabilizing selection should be informative in applications to data from natural populations.

Literature Cited

Bulmer, M. G., 1991 The selection-mutation-drift theory of synonomous codon usage. Genetics **129:** 897-907.

Bürger, R., 2000 *The Mathematical Theory of Selection, Recombination and Mutation*. John Wiley, Chichester.

Bürger, R., and R. Lande, 1984 On the distribution of the mean and variance of a quantitative trait under mutation-selection-drift balance. Genetics **138:** 901-912.

Charlesworth, B., and D. Charlesworth, 2010 *Elements of Evolutionary Genetics*. Roberts and Company, Greenwood Village, CO.

Charlesworth, B., A. J. Betancourt, V. B. Kaiser and I. Gordo, 2010 Genetic recombination and molecular evolution. Cold Spring Harb. Symp. Quant. Biol. **74:** 177-186.

Ewens, W. J., 2004 *Mathematical Population Genetics. 1. Theoretical Introduction*. Springer, New York.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Lande, R., 1976 Natural selection and random genetic drift in phenotypic evolution. Evolution **30:** 314-334.

Li, W.-H., 1987 Models of nearly neutral mutations with particular implications for non-random usage of synonymous codons. J. Mol. Evol. **24:** 337-345.

Tajima, F., 1983. Evolutionary relationship of DNA sequences in a finite population. Genetics **23** 28-64.

McVean, G. A. T., and B. Charlesworth, 1999 A population genetic model for the evolution of synonymous codon usage: patterns and predictions. Genet. Res. **74:** 145-158.

Table S1 Properties of the site frequency spectra for state-independent mutations

	Sites fixed for A_1	Sites fixed for A_2	Mean frequency of A_2
$\kappa = 2$			
$N = 50$	0.366 (0.015)	0.653 (0.016)	0.607 (0.063)
100	0.320 (0.014)	0.640 (0.015)	0.606 (0.046)
200	0.309 (0.014)	0.614 (0.015)	0.604 (0.032)
400	0.288 (0.014)	0.575 (0.015)	0.604 (0.025)
$\kappa = 4$			
$N = 50$	0.193 (0.013)	0.777 (0.013)	0.685 (0.050)
100	0.187 (0.012)	0.755 (0.013)	0.680 (0.034)
200	0.176 (0.011)	0.716 (0.013)	0.679 (0.026)
400	0.161 (0.011)	0.653 (0.014)	0.674 (0.018)

The entries display the mean proportions (over 500 replicate simulations) of sites in a sample of 20 alleles that are fixed for A_1- and A_2-type variants, respectively, together with the mean frequency of A_2-type variants in the sample. Standard deviations are shown in brackets. The selection and mutation parameters of Table 2 were used, with an optimum of zero.

Table S2 Simulation values of parameters under stabilizing selection with large Nsa^2

	State-dependent mutations				State-independent mutations			
	δ (x10)		V_g		δ (x10)		V_g	
	Sim. Mean	**s.d.**	**App.**	**Sim. Mean**	**s.d.**	**Stab. Sel.**	**Sim. Mean**	**s.d.**	**App.**	**Sim. Mean**	**s.d.**	**Stab. Sel.**
κ = 2												
$z_0 = 0$	0.298	2.61 (0.041)	0.217	0.381	0.076 (0.001)	0.300	0.244	2.60 (0.041)	0.212	0.635	0.100 (0.002)	0.600
$z_0 = 20$	0.388	2.61 (0.041)	0.257	0.382	0.079 (0.001)	0.320	0.212	2.58 (0.041)	0.212	0.634	0.099 (0.002)	0.600
κ = 4												
$z_0 = 0$	0.489	2.59 (0.041)	0.434	0.588	0.095 (0.002)	0.500	0.460	2.59 (0.041)	0.433	0.922	0.113 (0.002)	1.00
$z_0 = 20$	0.610	2.57 (0.041)	0.473	0.605	0.097 (0.002)	0.560	0.422	2.62 (0.041)	0.433	0.921	0.114 (0.002)	1.00

$N = 400$; $u = 1 \times 10^{-5}$; $m = 1000$, $S = 0.1$, $a = 0.316$ ($Sa^2 = 0.01$).

The entries headed 'Sim.' were obtained from stochastic simulations with 4000 replicates; the entries for δ headed 'App.' were obtained from Equations A3a (state-dependent model) and 13b (state-independent model), and the entries headed 'Stab. Sel.' from the formulae for V_g* with large population size (Equations S4.2 and S4.4, setting δ = 0). Results from matrix iterations are not shown, since problems with convergence were experienced.

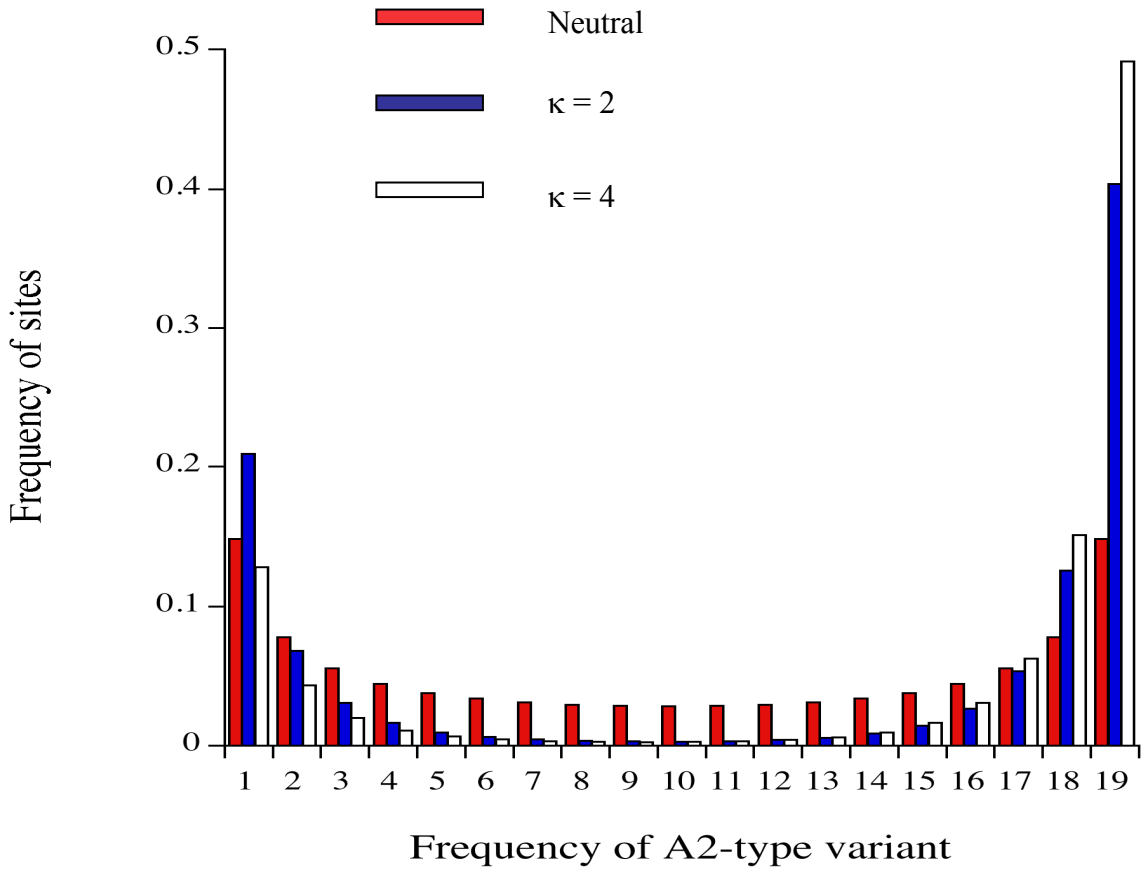

Figure S1 Site frequency spectra in a sample of 20 alleles with a predominance of stabilizing selection, from the results of pooling 4000 stochastic simulations. The histograms show the spectra for the cases of neutrality (red bars) and stabilizing selection with state-independent mutations and two different levels of mutational bias (blue and white bars) with $N = 400$, $S = 0.1$, $m = 1000$, $a = 0.316$, $z_0 = 0$ and $u = 1 \times 10^{-5}$.