FILE S1

APPROXIMATIONS

In our derivation of the time-dependent effective population size, we have made two key approximations. First, we have assumed that lineages and allele frequencies may be treated as effectively deterministic. Second, we have assumed that the ancestral fitness distributions at different sites may be treated as independent. These two approximations are prevalent in the history of background selection, and form the basis for many of the strong-selection results currently in use (Charlesworth, 2012; Charlesworth *et al.*, 1993; Hudson and Kaplan, 1995). In this Supplemental Information, we discuss these two approximations in detail.

The Deterministic Approximation

One of the central assumptions of background selection is that the population may be treated as approximately deterministic. This implies that frequencies may be assumed to be at mutation-selection balance, and that lineages may be described using deterministic equations such as that used to derive Eq. (1). In general, this assumption will hold when the strength of selection is sufficiently strong that it dominates the effects of drift (or analogously, when lineages are selected against sufficiently strongly that they never grow to a substantial fraction of the population). As a result, we expect the deterministic approximation to hold roughly when $Nse^{-U_d/(s+R/2)} \gg 1$. This approximation forms the foundation for previous results in background selection, including the structured coalescent results of Zeng and Charlesworth (2011) and the original background selection formulae from Charlesworth *et al.* (1993) and Hudson and Kaplan (1995).

The main difference between the classic background selection analysis and our analysis is that we include the transient period during which deleterious alleles may segregate in the population prior to being removed by selection. The traditional analysis assumes that this time-period is sufficiently small relative to the total coalescence time that it can be neglected. In general, the time-scale of this transition is roughly of order $1/s$, and therefore, by definition, should be small relative to the typical coalescence times, $\approx N_e$, whenever the deterministic approximation holds.

However, in practice, as seen in Figures 3-5, the deterministic approximation is still reasonable even when the time-scale of the transition begins to represent a significant fraction of the total coalescence times. Thus, by incorporating this transition time, we are able to more accurately describe the distribution of coalescence times and other statistics. This allows us to capture the distortions that begin to arise as a consequence of this transition period, and thus to qualitatively understand how selection distorts the shapes of genealogies, and how this depends upon the parameters involved. Even when this effect is small, by taking advantage of the fact that, in the presence of recombination, sites far away from one another become effectively independent, it may be possible to detect even small differences with enough sequence data. We note, however, that our method is only able to account for the distortions that arise due to this transition period, and not the additional effects that arise from fluctuations. As $N_e s$ becomes smaller, our analysis begins to break down as fluctuations in the population become very strong. When this happens, additional distortions (including topological distortions) arise which we are not able to capture with our analysis.

The breakdown of the deterministic approximation when $N_e s \approx 1$ has been discussed in several notable studies considering the weak selection regime (Barton and Etheridge, 2004; O'Fallon *et al.*, 2010). Earlier studies have suggested that the deterministic approximation is reasonable for the calculation of pairwise coalescence times when $N_e s > 3$ (Barton and Etheridge, 2004; Charlesworth, 2012), which is consistent with our findings in Figure 4. However, it is unclear whether such a precise threshold would remain accurate for more extreme parameter combinations, where additional logarithmic corrections could arise.

The Independent-Sites Approximation

The second key approximation made in the main text is that we may treat the ancestral fitness distribution at each site as independent. In other words, we assume that the joint ancestral fitness distribution across all sites is equal to the product of the ancestral fitness distribution at each site, $P_{k_1,k_2,k_3\ldots k_L}(t) = P_{k_1}(t)P_{k_2}(t)\ldots P_{k_L}(t)$, where k_i is either 0 or 1, indicating whether a mutation exists at site i.

In an asexual population, this holds whenever the deterministic approximation is valid. However, in the presence of recombination, correlations will exist between neighboring sites. This is a consequence of the fact that, when an ancestral recombination event occurs between the focal site and multiple index sites, all of those sites will now be randomly chosen from the population at the same time, and thus will all be 'reset' to the steady state mutation-selection balance simultaneously. Thus, sites that share the same history will be correlated.

However, this effect will be small provided that the deterministic approximation is valid ($N_e s \gg 1$) and that the probability of a mutation at any given site is small ($\mu/s \ll 1$). This approximation is prominent in previous literature on background selection, and is discussed in detail in the appendix of Hudson and Kaplan (1995). In order to justify this approximation, we will show that, provided the conditions stated above hold, the joint fitness distribution at two loci are approximately independent, i.e. $P_{k_1,k_2}(t) = P_{k_1}(t)P_{k_2}(t) + \mathcal{O}(\frac{\mu^2}{s^2})$. The same argument can then be extended to additional loci.

We denote the ancestral fitness distribution of an individual as $P_{ij}(t)$, where i and j represent whether a mutation exists at two sites of distances x_1 and x_2 from the focal site, respectively. We know from the main text that, to first order in μ/s:

$$\begin{aligned}
P_{00}(t) + P_{01}(t) &= 1 - \frac{\mu}{s}\left(\frac{rx_1}{rx_1+s} + \frac{s}{rx_1+s}e^{-st-rx_1t}\right)\\
P_{00}(t) + P_{10}(t) &= 1 - \frac{\mu}{s}\left(\frac{rx_2}{rx_2+s} + \frac{s}{rx_2+s}e^{-st-rx_2t}\right).
\end{aligned}$$

We can now write out the backwards-in-time master equation for $P_{00}(t)$, again keeping only first-order terms in μ, s, rx_1, and rx_2:

$$\begin{aligned}
P_{00}(t+1) = P_{00}(t)&(1 - rx_1(1-f_{00}) - r(x_2-x_1)(1-f_{00}-f_{10}))\\
&+P_{01}(t)\frac{f_{00}}{f_{01}}(\mu + rx_1f_{01} + r(x_2-x_1)(f_{01}+f_{11}))\\
&+P_{10}(t)\frac{f_{00}}{f_{10}}(\mu + rx_1f_{10})\\
&+P_{11}(t)\frac{f_{00}}{f_{11}}(rx_1f_{11}).
\end{aligned}$$

Making the continuous approximation this becomes:

$$\begin{aligned}
\frac{dP_{00}(t)}{dt} &= -rx_2P_{00}(t) + rx_1f_{00}(P_{00}(t)+P_{01}(t)+P_{10}(t)+P_{11}(t))\\
&\quad+r(x_2-x_1)\Big(P_{00}(t)(f_{00}+f_{10}) + \frac{f_{00}}{f_{01}}P_{01}(t)(f_{01}+f_{11})\Big) + \mu f_{00}\left(\frac{P_{01}(t)}{f_{01}} + \frac{P_{10}(t)}{f_{10}}\right)\\
&= -(rx_2 + 2s - 2\mu)P_{00}(t) + rx_1\left(1-\frac{\mu}{s}\right)^2\\
&\quad+r(x_2-x_1)\left(1-\frac{\mu}{s}\right)\left(1-\frac{\mu}{s}\left(\frac{rx_1}{rx_1+s}+\frac{s}{rx_1+s}e^{-st-rx_1t}\right)\right)\\
&\quad+s\left(1-\frac{\mu}{s}\right)\left(2-\frac{\mu}{s}\left(\frac{rx_1}{rx_1+s}+\frac{s}{rx_1+s}e^{-st-rx_1t}\right)-\frac{\mu}{s}\left(\frac{rx_2}{rx_2+s}+\frac{s}{rx_2+s}e^{-st-rx_2t}\right)\right).
\end{aligned}$$

Solving this to first order in μ/s:

$$
\begin{aligned}
P_{00}(t) &= 1 - \frac{\mu}{s}\left(\frac{rx_1}{rx_1+s} + \frac{s}{rx_1+s}e^{-st-rx_1t}\right) - \frac{\mu}{s}\left(\frac{rx_2}{rx_2+s} + \frac{s}{rx_2+s}e^{-st-rx_2t}\right) + \mathcal{O}\left(\frac{\mu^2}{s^2}\right)\\
P_{01}(t) &= \frac{\mu}{s}\left(\frac{rx_2}{rx_2+s} + \frac{s}{rx_2+s}e^{-st-rx_2t}\right) + \mathcal{O}\left(\frac{\mu^2}{s^2}\right)\\
P_{10}(t) &= \frac{\mu}{s}\left(\frac{rx_1}{rx_1+s} + \frac{s}{rx_1+s}e^{-st-rx_1t}\right) + \mathcal{O}\left(\frac{\mu^2}{s^2}\right)\\
P_{11}(t) &= \mathcal{O}\left(\frac{\mu^2}{s^2}\right).
\end{aligned}
$$

Thus, we see that $P_{ij}(t) = P_i(t)P_j(t) + \mathcal{O}\left(\mu^2/s^2\right)$, such that the sites are approximately independent. We note, however, that this independence does not hold to higher-order in $\frac{\mu}{s}$, and corrections would be required to accurately capture the joint ancestral probability at those orders. Thus, the independence approximation will only strictly hold when $\mu/s \ll 1$, and when the deterministic approximation holds.

We note that this approximation is discussed in detail in the appendix of Hudson and Kaplan (1995). They provide an analogous derivation of the joint mutation probability at two loci (see Equation A10), and similarly find that sites may be treated as independent provided the deterministic approximation holds and $\mu/s \ll 1$.

FILE S2

INCORPORATING BACK MUTATIONS

In our derivation of the time-dependent effective population size, we have neglected the effect of back mutations. In practice, back mutations only introduce terms of higher-order in μ/s, and thus are of negligible contribution in the regime we consider. However, it is straightforward to incorporate these terms into our analysis, which we do here.

First, we consider the steady-state distribution of mutations at a single site. This is determined by the solution to the equations:

$$\begin{aligned} f_1 &= \frac{f_1(1-s)}{\overline{\omega}}(1-\mu_b) + \frac{f_0}{\overline{\omega}}\mu_f \\ f_0 &= \frac{f_1(1-s)}{\overline{\omega}}\mu_b + \frac{f_0}{\overline{\omega}}(1-\mu_f), \end{aligned}$$

where μ_f and μ_b are the forward and back mutation rates, respectively. This yields:

$$\begin{aligned} f_1 &= \frac{s+\mu_b(1-s)+\mu_f-\sqrt{(s+\mu_b(1-s)+\mu_f)^2-4s\mu_f}}{2s} \\ f_0 &= \frac{s-\mu_b(1-s)-\mu_f+\sqrt{(s+\mu_b(1-s)+\mu_f)^2-4s\mu_f}}{2s}. \end{aligned}$$

When $\mu_b = 0$, these reduce to the usual mutation-selection balance results, $f_1 = \mu_f/s$ and $f_0 = 1 - \mu_f/s$. Furthermore, if we define $\mu_f \equiv \mu$ and $\mu_b \equiv c\mu$, and expand this result in orders of μ/s, we see that:

$$\begin{aligned} f_1 &= \frac{\mu}{s} - \frac{\mu^2}{s^2}c(1-s) + \frac{\mu^3}{s^3}(c^2(1-s)^2 - c(1-s))\ldots \\ f_0 &= 1 - \frac{\mu}{s} + \frac{\mu^2}{s^2}c(1-s) - \frac{\mu^3}{s^3}(c^2(1-s)^2 - c(1-s))\ldots \end{aligned}$$

Thus, we see that incorporating back mutations leads to a correction of order μ^2/s^2. As a consequence, the effect of back mutations is negligible in the regime we consider. However, we may derive Equation 2 from the main text including them. We have that:

$$\frac{dP_{mut}(t)}{dt} = -\left(rx + \frac{\mu_f N f_0}{N f_1} + \frac{\mu_b N f_1}{N f_0}\right) P_{mut}(t) + rx f_1 + \frac{\mu_b N f_1}{N f_0}$$

Solving this yields:

$$P_{mut}(x,t) = \frac{rx f_1 + \frac{\mu_b f_1}{f_0}}{rx + \frac{\mu_f f_0}{f_1} + \frac{\mu_b f_1}{f_0}} + \frac{\mu_f f_0 - \mu_b f_1}{rx + \frac{\mu_f f_0}{f_1} + \frac{\mu_b f_1}{f_0}} e^{-\left(rx + \frac{\mu_f f_0}{f_1} + \frac{\mu_b f_1}{f_0}\right)t}.$$

which replaces Equation 2 in the main text. Similarly, Equation 1 may be recovered by substituting $rx \to r(x_i, x_f)$. We note that these equations are identical to those given in the main text to leading-order in μ/s, and thus back mutations represent only a small correction to our results in the regime we consider.

LITERATURE CITED

Barton, N. H. and A. M. Etheridge, 2004. The effect of selection on genealogies. Genetics 166: 1115–1131.

Charlesworth, B., 2012. The effects of deleterious mutations on evolution at linked sites. Genetics 190: 5–22.

Charlesworth, B., M. Morgan, and D. Charlesworth, 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134(4): 1289.

Hudson, R. and N. Kaplan, 1995. Deleterious background selection with recombination. Genetics 141: 1605–1617.

O'Fallon, B., J. Seger, and F. Adler, 2010. A continuous-state coalescent and the impact of weak selection on the structure of gene genealogies. Mol. Biol. Evol. 27(5): 1162.

Zeng, K. and B. Charlesworth, 2011. The joint effects of background selection and genetic recombination on local gene genealogies. Genetics 189(1): 251–266.