The Structure of Genealogies in the Presence of Purifying Selection: An Effective Coalescent Theory

Michael M. Desai\textsuperscript{1,2,3}, Aleksandra M. Walczak\textsuperscript{4}, Lauren E. Nicolaisen\textsuperscript{2,3}, and Joshua B. Plotkin\textsuperscript{5}

\textsuperscript{1}Department of Organismic and Evolutionary Biology, \textsuperscript{2}Department of Physics, and \textsuperscript{3}FAS Center for Systems Biology, Harvard University
\textsuperscript{4}CNRS-Laboratoire de Physique Théorique de l’École Normale Supérieure, 24 rue Lhomond, 75005 Paris, France,
\textsuperscript{5}Department of Biology, University of Pennsylvania

(Dated: October 12, 2010)
Abstract

Purifying selection distorts the structure of genealogies and hence alters the patterns of genetic variation we observe in sequence data. Although these distortions may be common, our understanding of how we expect purifying selection to affect patterns of molecular evolution remains incomplete. Genealogical approaches such as coalescent theory have proven difficult to generalize to situations involving selection at many linked sites, unless selection pressures are extremely strong. Here, we introduce an effective coalescent theory to describe the structure of genealogies in the presence of purifying selection at many linked sites. We use this effective theory to calculate several simple statistics describing the expected patterns of variation in sequence data, both at the sites under selection and at linked neutral sites. Our analysis combines our earlier description of the allele frequency spectrum in the presence of purifying selection (Desai et al., 2010) with the structured coalescent approach of Nordborg (1997), to trace the ancestry of individuals through the distribution of fitnesses within the population. We find that purifying selection leads to patterns of genetic variation which are related but not identical to a neutrally evolving population in which population size has varied in a specific way in the past.
Running Head: Coalescent Theory with Purifying Selection

Keywords: Coalescent, Purifying Selection, Genealogies, Linkage

Corresponding Author:
Michael M. Desai

Departments of Organismic and Evolutionary Biology and of Physics
FAS Center for Systems Biology
Harvard University
435.20 Northwest Labs
52 Oxford Street
Cambridge, MA 02138
617-496-3613
mdesai@oeb.harvard.edu
INTRODUCTION

Selection acting simultaneously at many linked sites can substantially alter the patterns of molecular evolution at these sites, and of linked neutral variation (Felsenstein, 1974; Hahn, 2008; Hill and Robertson, 1966; McVean and Charlesworth, 2000). Attention has traditionally focused on the statistics of purely neutral variation (Kimura, 1983), or on how a relatively few strongly selected mutations affect linked neutral sites (Barton, 1998; Barton and Etheridge, 2004; Gillespie, 2001, 2000; Ohta and Kimura, 1975; Smith and Haigh, 1974). But in recent years, evidence from sequence data points to the general importance of weak selective forces among many linked variants in microbial and viral populations, and on short distance scales in the genomes of sexual organisms (Betancourt et al., 2009; Comeron et al., 2008; Hahn, 2008; Seger et al., 2010). In these situations, existing theory does not explain patterns of molecular evolution (Hahn, 2008).

A vast body of work provides an excellent understanding of purely neutral variation, amongst both recombining and tightly linked sites. This work is based primarily on genealogical approaches such as coalescent theory, which provides a complete and elegant framework for understanding genetic variation in the absence of recombination and selection (Wakeley, 2009). While many extensions to coalescent theory have been developed to account for various complicating factors, including arbitrary degrees of linkage between sites (Griffiths and Marjoram, 1997), there remains no broadly useful way of handling selection within coalescent theory. The problem is fundamental to genealogical frameworks, which rely on characterizing the space of possible genealogical trees before considering the possibility of mutations at various locations on these trees. When selection operates, the probabilities of particular trees cannot be defined independently of the mutations, and the approach breaks down (Tavare, 2004; Wakeley, 2009). The ancestral selection graph of Neuhauser and Krone (1997) and Krone and Neuhauser (1997) provides an elegant formal solution to this problem, but unfortunately it requires extensive numerical calculations (Przeworski et al., 1999). These limit the intuition we can draw from this method, and make it impractical as the basis for inference from most modern sequence data. Alternative approaches such as Poisson Random Field models allow us to understand how selection affects the evolution and population genetics of many unlinked loci (Bus-
We also know how a few selected mutations affect linked neutral loci (Barton and Etheridge, 2004; Gillespie, 2001). This is often sufficient to understand strongly selected traits in small populations. Yet when weak selection is common in larger populations, existing theory falls short (Hahn, 2008).

The existing methods summarized above provide an understanding of genetic variation among neutral sites which are arbitrarily linked, or among selected sites which are completely unlinked. But we have little understanding of the genetic diversity we expect to see among many linked selected sites, with or without linked neutral sites. This makes it difficult to form a coherent basis for inferring if or how selection has influenced the patterns of variation we observe in sequence data. Instead, we understand what sequence data would look like if all mutations were neutral, and there are various ways to look for deviations from this expectation that may suggest different selective forces (Ewens, 2004). But with a few limited exceptions, there are no models of what sequence variation should look like in the presence of selection on many linked selected sites. Thus when we look for selection, we do not know precisely what we are looking for. This makes it hard to identify the most powerful ways to distinguish selection from other evolutionary forces. Even simple null models of this process would be useful in forming precise predictions, which may help us to develop more powerful methods to distinguish competing possibilities for which the intuitively expected departures from the neutral expectations are similar.

In this paper, we study this situation where selection acts on a large number of linked sites. We focus on the case of purifying selection among many perfectly linked sites, and study the simplest null model that describes this situation. Specifically, we imagine a number of sites at which either deleterious or neutral mutations can occur. We assume these sites are within an asexual genome, or close enough together on a sexual genome that recombination can be neglected. Charlesworth et al. (1993) proposed an approximation for studying genetic diversity in exactly this situation, which has become known as “background selection” (BGS) (Charlesworth, 1994; Charlesworth et al., 1995). Our method is an extension of the BGS analysis to weaker selection or larger deleterious mutation rates. By “weaker selection” we mean the regime $Ns \gg 1$, but with the total deleterious mutation rate larger than the selection pressure against individual mutants; we consider the precise relationship between
our work and BGS in the Discussion. The even weaker selection regime where \( N_s \sim 1 \) has recently been studied by O’Fallon et al. (2010) using a somewhat different continuous-fitness model.

In regimes in which BGS does not apply, simulation studies have shown that selection distorts patterns of genetic variation in a way that cannot be reduced to a simple neutral model with a modified effective population size (Comeron and Kreitman, 2002; McVean and Charlesworth, 2000; Seger et al., 2010). This effect is sometimes referred to as Hill-Robertson interference (Hill and Robertson, 1966). In this paper we propose an analytical framework for understanding the expected genetic diversity in the presence of Hill-Robertson interference from many linked negatively selected mutations. We find that indeed selection distorts variation, in a way that is related but not identical to a neutrally evolving population in which population size has varied in a specific way in the past.

Our analysis is inspired in some ways by the structured coalescent of Nordborg (1997), in that we think of the population as subdivided into different fitness classes and we trace the genealogies of individuals as they move between classes. In this sense our approach is similar to the recent work by O’Fallon et al. (2010). However, we stress that our method is an “effective” coalescent theory, not an actual one. We do not study coalescence in real time. Rather, we treat each fitness class as a “generation” and trace how individuals have descended by mutations through fitness classes, moving from one “generation” to the next by subsequent mutations. This analogy allows us to make a precise mapping to coalescence theory, in which certain quantities (e.g. coalescence times) have a different meaning than in the traditional theory. We can then invert aspects of this mapping to determine the structure of genealogies and calculate statistics describing expected patterns of genetic variation. This approach has several advantages. Most importantly, it makes the entire analysis possible in a situation where more traditional structured coalescent approaches have proven intractable. Our approach also makes it possible to calculate the diversity created by the selected sites themselves, which may be important when selection is common, and is impossible to determine in a traditional structured coalescence approach.

We begin in the next section by describing the details of our model. We next explain the formal structure of our effective coalescent approach, and we calculate the coalescence probabilities. This work relies heavily on the framework developed in Desai et al. (2010)
to calculate the frequency distribution of distinct lineages within each fitness class. We then show how this effective coalescent determines the structures of genealogies, and we calculate various statistics describing genetic variation in these populations, which we compare to numerical simulations. We finally discuss the relationship between our results and neutral theory and background selection, and we explore how various approximations (most importantly the fact that we neglect Muller’s ratchet) limit our approach.

MODEL

We consider in this paper a model identical to that in Desai et al. (2010). That is, we imagine a finite haploid population of constant size $N$. Each individual has a genome composed of a large number of sites. Each site is assumed to begin in some ancestral state, and can mutate with some constant rate. Each mutation is assumed to be either neutral or to confer some fitness disadvantage $s$ (where by convention $s > 0$). We work within an infinite-sites approximation, where the probability that two mutations at the same site segregate simultaneously within the population is negligible.

We assume that there is no epistasis for fitness, so each deleterious mutation contributes multiplicatively to the fitness of each individual. We assume that all deleterious mutations carry the same fitness cost $s$, and that $s \ll 1$, so that the fitness of an individual with $k$ deleterious mutations is approximately $w_k = 1 - sk$.

The dynamics of competing individuals are assumed to follow the diffusion limit of the standard Wright-Fisher model. In each generation an individual acquires a new deleterious mutation, somewhere in its genome, with probability $U_d$. Thus, $\theta_d/2 \equiv N U_d$ is the per-genome scaled deleterious mutation rate. Similarly, neutral mutations occur at a rate $U_n$ per individual per generation, and we define $\theta_n/2 \equiv N U_n$. Whenever a mutation arises, it is assumed to arise at site for which there are no other segregating polymorphisms in the population (the infinite-sites assumption). We focus exclusively on the case of perfect linkage, where we imagine that all the sites we are considering are in an asexual genome or within a short enough distance in a sexual genome that recombination can be entirely neglected. Although our model is defined for haploids, this assumption means that our analysis also applies to diploid populations provided that there is no dominance (i.e. being
homozygous for the deleterious mutation carries twice the fitness cost as being heterozygous).

For the bulk of this paper, we will assume that Muller’s ratchet can be neglected. While this assumption presented minimal problems in the context of the allele-based analysis in Desai et al. (2010), it is more problematic here. Thus we will return to the question of the importance of Muller’s ratchet in more detail in the Discussion.

We believe that our model is the simplest possible null model based on a concrete picture of mutations at individual sites that can describe the effects of a large number of linked negatively selected sites on patterns of genetic variation. In Desai et al. (2010) we discuss its relationship with other models which have been introduced in earlier related work.

**ALLELIC DIVERSITY IN THE DELETERIOUS MUTATION-SELECTION BALANCE**

In this paper, we will develop an effective coalescent theory that involves tracing the ancestry of individuals as they change in fitness by acquiring deleterious mutations. In order to do this, we need to first understand the distribution of fitnesses within the population and the structure of lineage diversity amongst individuals within a given fitness class. We have analyzed these topics in detail in Desai et al. (2010). Here we merely summarize the results relevant for our subsequent coalescent analysis.

In our model all deleterious mutations have the same fitness cost $s$, and so we can classify individuals based on their Hamming class, $k$, relative to the wildtype (which by definition has $k = 0$). That is, individuals in class $k$ have $k$ deleterious mutations more than the most-fit individuals in the population. Note that not all individuals in class $k$ have the same set of $k$ deleterious mutations. Furthermore, $k$ refers only to the number of deleterious mutations an individual has; individuals with the same $k$ can have different numbers of neutral mutations. We normalize fitness such that by definition all individuals in class $k = 0$ have fitness 1. Individuals in class $k$ then have fitness $1 - ks$ (Fig. 1).

We showed in Desai et al. (2010) that the balance between mutation and selection leads to a steady state in which the fraction of the population in fitness class $k$, which we call $h_k$, is given by a Poisson distribution with mean $U_d/s$,

\[
h_k = e^{-U_d/s} \frac{U_d^k}{k! s^k}.
\]
This is consistent with the earlier work by Haigh (1978), and means that the average fitness in the population is $1 - U_d$, and that $\bar{k} = \frac{U_d}{s}$.

Consider a fitness class $k$, which has an overall frequency $h_k$ (Fig. 1b). The frequency $h_k$ is maintained by a stochastic process in which the class is constantly receiving new individuals from class $k - 1$ due to mutations. In our infinite-alleles approximation, each such mutation creates a lineage which is an allele that is unique within the population. Each lineage fluctuates in frequency for a while before eventually dying out, perhaps after acquiring additional mutations that found new lineages in fitness class $k + 1$. At any given moment, there is some frequency distribution of lineages in each class $k$ (see Fig. 2). While the identity of these lineages changes over time, there is a probability distribution that at any moment there is a given frequency distribution of lineages. In steady state, this probability distribution does not change with time.

In Desai et al. (2010), we calculated this steady state probability distribution of the frequency distribution of lineages. For our purposes here, it is most useful to consider these results in the absence of neutral mutations; we will consider the diversity at neutral sites separately below. In the absence of neutral mutations, we noted that new lineages are founded in class $k$ at a rate $\theta_k/2$, where

$$\theta_k = 2Nh_{k-1}U_d. \quad (2)$$

These individuals are then removed from class $k$ at a per capita rate

$$s_k \equiv -U_d - s(k - \bar{k}). \quad (3)$$

We refer to $s_k$ as the effective selection coefficient against an allele in class $k$, because it is the rate at which any particular lineage in class $k$ loses individuals, and we defined

$$\gamma_k = Ns_k. \quad (4)$$

Our model then reduced to the situation studied by the Poisson Random Field model of Sawyer and Hartl (1992) and Hartl and Sawyer (1994). Thus the frequency distribution of lineages (alleles) in fitness class $k$ follows a Poisson Random Field (PRF) with effective parameters $\theta_k$ and $\gamma_k$. That is, the number of distinct lineages in class $k$ with a frequency between $a$ and $b$ (relative to the total size of the population $N$) is Poisson distributed
with mean
\[ \int_a^b f_k(x)dx, \]  
where
\[ f_k(x) = \frac{\theta_k}{x(1-x)} \frac{1 - e^{-2\gamma_k(1-x)}}{1 - e^{-2\gamma_k}}. \]

This is equivalent to saying that the probability that there exists a lineage in class \( k \) with frequency between \( x \) and \( x + dx \) is \( f_k(x)dx \), for infinitesimal \( dx \). Note that this analysis involves various implicit approximations, and the results are valid within a specific parameter regime. We describe these approximations and limitations in detail in Desai et al. (2010), and return to them as relevant for the present work in the Discussion.

Most importantly for our subsequent analysis, note that our Poisson Random Field result implies that on average the sum of all the frequencies of all the alleles in fitness class \( k \) is simply
\[ h_k = \int_0^1 x f_k(x)dx, \]
and that the probability that two individuals chosen at the same time at random from fitness class \( k \) both come from the same lineage is
\[ \frac{1}{h_k^2} \int_0^1 x^2 f_k(x)dx. \]

**AN EFFECTIVE COALESCENT PROCESS**

We have just seen that within each fitness class \( k \) there are various lineages (each genetically unique in our infinite-sites framework) with a frequency distribution as described by the PRF result \( f_k(x) \) (with appropriate effective \( \gamma_k \) and \( \theta_k \)). We now imagine picking two individuals at random from the population, and attempt to calculate their degree of relatedness. If they happen to be from the same lineage, they are genetically identical. If they are from two different lineages, we want to know how many mutations separate them (i.e. how related the two lineages are). That is, we want to know the distribution of the per-site heterozygosity \( \pi \), the number of sites at which the two individuals have a different nucleotide. For now, we focus on the \( \pi \) among deleterious sites only, which we will call \( \pi_d \), neglecting neutral variation. We will later relate this to the distribution of coalescence times between these
two individuals, from which we can calculate the distribution of neutral variation between these sequences.

To calculate \( \pi_d \), we will trace the ancestry of lineages through the fitness distribution. The idea is that two lineages in fitness class \( k \) came from mutations within individuals in fitness class \( k - 1 \). If both those mutations came from individuals in the same lineage in fitness class \( k - 1 \), the two lineages in fitness class \( k \) differ only at two sites: those at which the mutations taking them from class \( k - 1 \) to class \( k \) occurred. We thus have \( \pi_d = 2 \). In this case, we say that the two individuals in fitness class \( k \) (which are in different lineages there) coalesce in fitness class \( k - 1 \). If they did not coalesce in fitness class \( k - 1 \), they must have come from mutations in lineages within class \( k - 1 \). We can then ask whether or not these lineages coalesced in fitness class \( k - 2 \) (in which case \( \pi_d = 4 \)), and so on.

In this way, we can construct an effective coalescent tree describing the relatedness of two individuals from fitness class \( k \), as illustrated in Fig. 2. In this effective coalescent, each “generation” represents the deleterious mutations taking individuals from one fitness class to the next. The coalescence probability between two individuals in two different lineages in a given fitness class is the probability that the two mutations that created these lineages came from individuals within the same lineage (and hence were genetically identical) in the previous fitness class \( k - 1 \). We will call this coalescence probability \( P^{k,k-1}_{c} \). If they did not coalesce in class \( k - 1 \), they came from two different lineages within that class, which may have coalesced in class \( k - 2 \), and so on. In general, we will define \( P^{k',k-1}_{c} \) as the probability that two individuals chosen at random from fitness class \( k' \) coalesce in class \( k \) (i.e. they are descendants of two different mutations in individuals from within the same lineage within class \( k \)).

Note that in the standard neutral coalescent, one first calculates the distribution of coalescence times and then imagines mutations occurring as a Poisson process throughout the coalescent tree, with rates proportional to branch lengths. In our effective coalescent, by contrast, the coalescence times are the mutations. Specifically, \( \pi_d \) equals twice the coalescent “time” in our effective model, and hence the distribution of \( \pi_d \) is given directly by the distribution of coalescent “time.” To avoid confusion, from here on we will refer to the effective “generations” in our model as “steps”, and refer to the effective coalescent “times” as the “steptimes.” We will reserve the word time to refer to the actual coalescent time,
measured in actual generations.

As is the case in the standard neutral coalescent, we will find that the probability of three or more lineages coalescing in a single step is small compared to the probability of two lineages coalescing. That is, we can neglect triplet (or higher) coalescent events. This means that once we calculate the coalescence probabilities in each step, we can in principle calculate everything about the probability distribution of coalescent trees, and hence describe any aspect of the genetic diversity between any number of individuals.

In addition to the diversity at negatively selected sites, we will also want to understand the diversity at linked neutral sites. We will do this by calculating how the steptime in our effective coalescent model translates into an actual time in generations. This will allow us to relate the distribution of branch lengths in steptimes to an actual coalescent tree in generations. We can then treat neutral mutations as is usually done in the standard coalescent: as a Poisson process with probabilities proportional to branch lengths. However, for now we neglect neutral mutations and focus on formulating the effective coalescent framework; we defer the calculations of neutral diversity to a later section.

The Coalescence Probabilities:

Our goal is to understand the probability distribution of the effective coalescence steptimes for two individuals chosen at random from the population. We begin in this section by calculating the coalescence probability in each step.

First, imagine that by chance we pick two individuals from the same fitness class \( k \). This class has a total frequency \( h_k \) as given in Eq. (1), and within the class there is a probability \( f_k(x) \) as given in Eq. (6) that there exists a lineage with frequency \( x \). Thus there is probability

\[
P_c^{k,k \rightarrow k} = \int_0^\infty x^2 f_k(x) / h_k^2
\]  

(9)

that these two chosen individuals come from the same lineage (note this contains an implicit approximation, see Appendix A for details). If so, they are genetically identical and the coalescence steptime is 0. If not, we want to calculate the probability they coalescence in class \( k - 1 \), \( P_c^{k,k \rightarrow k-1} \). If the lineage of individual \( A \) in class \( k \) was founded by a mutation from class \( k - 1 \) a time \( t_1 \) ago, and the lineage of individual \( B \) in class \( k \) was founded by a mutation a time \( t_2 \) ago, the probability the two individuals came from a common lineage in
class $k - 1$ is
\[ P^{k,k \rightarrow k-1}_c = \int dx dy dt_1 dt_2 Q^{k-1}_{k,k}(t_1, t_2) \frac{x f_{k-1}(x) y G_{k-1}(y \rightarrow x, |t_2 - t_1|)}{h_{k-1}}. \]  

(10)

Here $Q^{k-1}_{k,k}(t_1, t_2)$ is the joint distribution of $t_1$ and $t_2$, and $G_{k-1}(y \rightarrow x, |t_2 - t_1|)$ is the probability a lineage in class $k - 1$ changes in frequency from $x$ to $y$ in time $|t_2 - t_1|$. We return to the forms of these functions below. We assume the distribution $h_k$ is constant in time. This is the same assumption we used in calculating $f_k(x)$, and requires either $NU_d \gg 1$ or $Ns \gg 1$ (Desai et al., 2010). Note however that additional complications can arise from Muller’s ratchet, which we neglect here and will address in the Discussion. These formulas also assume that the probability a single lineage represents a substantial fraction of the size of a fitness class can be neglected; we discuss this approximation in more detail in Appendix A.

If the two individuals coalesced in this first step, the coalescent steptime is 1. If not (which occurs with probability $1 - P^{k,k \rightarrow k-1}_c$), we have to consider the probability they coalesce at the next step (i.e. in the mutations that took them from class $k - 2$ to $k - 1$). This probability is
\[ P^{k,k \rightarrow k-2}_c = \int dx dy dt_1 dt_2 Q^{k-2}_{k,k}(t_1, t_2) \frac{x f_{k-2}(x) y G_{k-2}(y \rightarrow x, |t_2 - t_1|)}{h_{k-2}}. \]  

(11)

Here $t_1$ is the time the ancestor of individual $A$ in class $k$ mutated from class $k - 2$ to $k - 1$, and analogously for $t_2$; $Q^{k-2}_{k,k}(t_1, t_2)$ is the joint distribution of these times, and $f_{k-2}(x)$ and $G_{k-2}$ are defined as above. If the two individuals did not coalesce in this step, we can continue in the same vein and calculate $P^{k,k \rightarrow k-3}_c$, and so on.

So far we have imaged that both individuals that we originally selected from the population came from the same class $k$. This will not generally be true. Rather, when we pick two individuals at random, they will come from classes $k$ and $k'$ with probability
\[ H(k, k') = 2h_ikh_{k'}, \quad \text{if } k \neq k' \]
\[ H(k, k) = h_k^2 \]  

(12)

For convenience we choose $k \leq k'$. We define $P^{k,k' \rightarrow k-\ell}_c$ to be the probability that two individuals from classes $k$ and $k'$ coalesce in class $k - \ell$. Note that $P^{k,k' \rightarrow k-\ell}_c = 0$ for $\ell < 0$. For $\ell \geq 0$ we have
\[ P^{k,k' \rightarrow k-\ell}_c = \int dx dy dt_1 dt_2 Q^{k-\ell}_{k,k'}(t_1, t_2) \frac{x f_{k-\ell}(x) y G_{k-\ell}(y \rightarrow x, |t_2 - t_1|)}{h_{k-\ell}}. \]  

(13)
Of course the fact that \( k' > k \) means that typically \( t_1 \) will be larger than \( t_2 \), and have a broader distribution. Note that in this case when coalescence occurs in class \( k - \ell \), we have \( \pi_d = 2\ell + (k' - k) \)

From the set of coalescence probabilities Eq. (13), we can calculate the probability distribution of coalescence steptimes between two individuals, and hence the distribution of per-site heterozygosity at negatively selected sites, \( \pi_d \). Further, assuming that the probability that three lineages coalesce in a given step is negligible, we can in principle calculate the distribution of coalescent tree shapes and branch lengths in steptimes for a sample of any number of individuals.

**CALCULATING THE COALESCENCE PROBABILITIES**

We now have a formal structure describing the structure of coalescent genealogies in the presence of negative selection. It remains, however, to evaluate the coalescent probabilities in each step, and to use these probabilities to calculate the probability distribution of genealogies.

We begin by noting that the coalescent probabilities all depend on the transition probability for the change in the frequency of a lineage from \( x \) to \( y \) in a time \( |t_1 - t_2| \) in class \( k - \ell \), \( G_{k-\ell}(y \to x, |t_2 - t_1|) \). This transition probability was calculated by Kimura (1955) and can be expressed as an infinite sum of Gegenbauer polynomials. Fortunately, it always appears in the context of an integral

\[
I_G = \int yG_{k-\ell}(y \to x, |t_2 - t_1|)dy,
\]

which is simply the average of \( y \) over \( G_{k-\ell} \). Hence it is given by the deterministic result for the change in the frequency of the lineage,

\[
I_G = xe^{-s(k-\ell)|t_2 - t_1|}.
\]

This simple expression for \( I_G \) makes our approach analytically tractable.

**The coalescence probability in the first step:**

We begin by evaluating the probability that two individuals chosen from fitness class \( k \) coalesce in class \( k - 1 \). Applying Eq. (15) to Eq. (10), we have

\[
P_{c}^{k,k\to k-1} = \int dxdt_1dt_2Q_{k,k}^{k-1}(t_1, t_2) \frac{x}{(h_{k-1})^2}f_{k-1}(x)xe^{-s(k-1)|t_1 - t_2|}.
\]
Since the two individuals mutated independently from class $k - 1$, the joint distribution $Q_{k,k-1}^{k-1}(t_1, t_2) = Q_{k}^{k-1}(t_1)Q_{k-1}^{k}(t_2)$. The distribution $Q_{k}^{k-1}(t)$ can be calculated by noting that the probability that an individual in class $k$ arose from a mutation in an individual in class $k - 1$ rather than a reproduction event from an individual in class $k$ is

$$\frac{NUdh_{k-1}}{Nh_{k}(1-\gamma_d) + NUdh_{k-1}}.$$  

Substituting in the steady state values for the $h_{k}$, this becomes

$$\frac{1}{1 + \frac{1}{k} (\frac{1}{s} - \frac{\gamma_d}{s})} \approx \frac{1}{1 + \frac{1}{sk}} \approx sk$$  

This means that we have

$$Q_{k}^{k-1}(t) = ske^{-skt}.$$  

Using this, and substituting for $f_{k-1}(x)$ from Eq. (6), we find

$$P_{k,k\rightarrow k-1}^{k,k} = \frac{(sk)^{2}a_{k-1}}{(e_{k-1}^a - 1)h_{k-1}} \int dx \frac{x}{1-x} \left[ e^{a_{k-1}(1-x)} - 1 \right] \int dt_1 dt_2 \exp[-sk(t_1 + t_2) - s(k-1)|t_1 - t_2|],$$

where we have defined $a_{k-1} \equiv -2\gamma_{k-1} = 2Ns(k-1)$. We can do the time integral by noting that

$$\int_{0}^{\infty} \int_{0}^{\infty} dt_1 dt_2 \exp[-sk(t_1 + t_2) - s(k-1)|t_1 - t_2|] =$$

$$= 2 \int_{0}^{\infty} dt_1 \int_{0}^{t_1} dt_2 \exp[-sk(t_1 + t_2) - s(k-1)(t_1 - t_2)] = \frac{1}{s^2 k (2k-1)}.$$  

The $dx$ integral is more complex; we discussed how to compute integrals of this form in Appendix A of Desai et al. (2010). Plugging in the result we found there, we have

$$P_{c}^{k,k\rightarrow k-1} = \frac{k}{2Nh_{k-1}s(k-1)(2k-1)}.$$  

**Coalescence probabilities in subsequent steps:**

We now wish to calculate the probability two individuals both chosen from fitness class $k$ coalesce in an arbitrary class $k - \ell$. First consider the probability of coalescence in class $k - 2$. This is given by

$$P_{c}^{k,k\rightarrow k-2} = \int Q_{k,k}^{k-2}(t_1, t_2) \frac{x^2}{(h_{k-2})^2} f_{k-2}(x) \exp[-s(k-2)|t_1 - t_2|] dt_1 dt_2 dx$$

$$= R_{x}^{k-2} \int Q_{k,k}^{k-2}(t_1, t_2) \exp[-s(k-2)|t_1 - t_2|] dt_1 dt_2,$$
where we have defined \( I_x^{k-2} = \frac{1}{2Nh_x^{k-2}} \).

The time \( t_1 \) is now the sum of the time for one individual to have mutated from class \( k - 2 \) to class \( k - 1 \) plus the time for it to have mutated from class \( k - 1 \) to class \( k \), and analogously for \( t_2 \). However, in order for the two lineages to coalesce in class \( k - 2 \), they must not have coalesced in class \( k - 1 \). We refer to the probability distribution of the times when these individuals mutated from class \( k - 1 \) to class \( k \) conditional on them not having coalesced in class \( k - 1 \) as \( Q_{k,k}^{k-1}(t_1, t_2|nc) \). The distribution of the times for these individuals to then have mutated from class \( k - 2 \) to class \( k - 1 \) is then given by

\[
Q_{1\text{step}}^{k-2} = [s(k - 1)]^2 e^{-s(k-1)(t_1 + t_2)},
\]

as in the first step. Thus the distribution of \( t_1 \) and \( t_2 \) is given by

\[
Q_{k,k}^{k-2}(t_1, t_2) = Q_{k,k}^{k-1}(t_1, t_2|nc) \ast Q_{1\text{step}}^{k-2}(t_1, t_2),
\]

where \( \ast \) indicates a convolution. Note that much of the time when the individuals did coalesce in class \( k - 1 \), they did so because \( t_1 \) happened to be close to \( t_2 \) (since this increases the chance the two individuals mutated from the same lineage). Thus in \( Q_{k,k}^{k-1}(t_1, t_2|nc) \), \( t_1 \) and \( t_2 \) are on average further apart than in \( Q_{k,k}^{k-1}(t_1, t_2) \), and \( t_1 \) and \( t_2 \) are no longer independent random variables.

We now need to calculate \( Q_{k,k}^{k-1}(t_1, t_2|nc) \). We have

\[
Q_{k,k}^{k-1}(t_1, t_2|nc) = \frac{Q_{k,k}^{k-1}(t_1, t_2) - Q_{k,k}^{k-1}(t_1, t_2|c) P_{c}^{k,k\rightarrow k-1}}{1 - P_{c}^{k,k\rightarrow k-1}},
\]

where \( Q_{k,k}^{k-1}(t_1, t_2|c) \) is the distribution of timings of mutations from class \( k - 1 \) to \( k \) given that the lineages do coalesce in class \( k - 1 \). Applying the general probability identity

\[
P(t_1, t_2|c) = \frac{1}{P(c)} P(c|t_1, t_2) P(t_1, t_2),
\]

and reading off the coalescence probability given \( t_1 \) and \( t_2 \) from Eq. (16), we find that

\[
Q_{k,k}^{k-1}(t_1, t_2|c) = \frac{I_x^{k-1}}{P_{c}^{k,k\rightarrow k-1} Q_{k,k}^{k-1}(t_1, t_2)} e^{-s(k-1)|t_1 - t_2|}.
\]

Plugging Eq. (29) and the results from the previous section for \( Q_{k,k}^{k-1}(t_1, t_2) \) and \( P_{c}^{k,k\rightarrow k-1} \) into Eq. (28), then plugging Eq. (28) into Eq. (27), and finally plugging Eq. (27) into Eq. (25), we can now calculate the coalescence probability in the second step, \( P_{c}^{k,k\rightarrow k-2} \). We can then repeat this analysis to calculate the coalescence probabilities in subsequent steps.
We discuss this full calculation in Appendix B. Here we make use of a simpler approximation: since the coalescence probability in each step will turn out to be small, conditioning on not coalescing in class \( k - 1 \) does not shift the distribution of mutation timings much. To be precise, we see in Eq. (28) that \( Q_{k,k}^{k-1}(t_1, t_2 | nc) \) differs from \( Q_{k,k}^{k-1}(t_1, t_2) \) only by a factor proportional to \( P_{c}^{k,k \rightarrow k-1} \). In what follows, we will therefore neglect the complications associated with the probability distributions of the mutant timings conditional on non-coalescence, and use the simpler distributions of unconditional timings. We refer to this as the non-conditional approximation, and discuss its validity further in Appendix B.

The non-conditional approximation:

In the non-conditional approximation, we can write the probability that two individuals both chosen from fitness class \( k \) coalesce in an arbitrary class \( k - \ell \) as

\[
P_{c}^{k,k \rightarrow k-\ell} = \int Q_{k,k}^{k-\ell}(t_1, t_2) \frac{x^2}{h_k^{k-\ell}} f_{k-\ell}(x)e^{-s(k-\ell)(t_1-t_2)}\, dt_1 dt_2 dx,
\]

(30)

where in our approximation \( Q_{k,k}^{k-\ell}(t_1, t_2) \) is the unconditional distribution of the times at which the two individuals sampled in class \( k \) originally moved from class \( k - \ell \) to class \( k - \ell + 1 \) by acquiring a deleterious mutation.

Since \( t_1 \) and \( t_2 \) are independent in the non-conditional approximation, we have \( Q_{k,k}^{k-\ell}(t_1, t_2) = Q_{k}^{k-\ell}(t_1)Q_{k}^{k-\ell}(t_2) \). Using this, we find

\[
P_{c}^{k,k \rightarrow k-\ell} = \frac{1}{2Nh_{k-\ell} s(k-\ell)} \int_{0}^{\infty} Q_{k}^{k-\ell}(t_1) e^{-s(k-\ell)t_1} \int_{0}^{t_1} Q_{k}^{k-\ell}(t_2) e^{s(k-\ell)t_2} \, dt_2 \, dt_1.
\]

(31)

We calculate the distributions of mutant timings \( Q_{k}^{k-\ell}(t) \) in Appendix C. Plugging these in, and evaluating the integrals as described in Appendix D, we find

\[
P_{c}^{k,k \rightarrow k-\ell} = \frac{1}{2Nh_{k-\ell} s(k-\ell)} \frac{\binom{k}{\ell}^2}{\binom{2k}{2\ell}},
\]

(32)

where \( \binom{a}{b} \equiv \frac{a!}{b!(a-b)!} \). This is our final result for the coalescence probability in class \( k - \ell \) of two individuals chosen from the same class \( k \). Note that the dependence on the parameters of the evolutionary process is entirely contained in the factor \( \frac{1}{2Nh_{k-\ell} s(k-\ell)} \). Thus the result Eq. (32) is simply

\[
P_{c}^{k,k \rightarrow k-\ell} = \frac{1}{2Nh_{k-\ell} s(k-\ell)} A_{k}^{\ell},
\]

(33)

where \( A_{k}^{\ell} \) is a numerical coefficient which depends on \( k \) and \( \ell \) but not on the population parameters.
It is interesting to explicitly calculate a few specific cases. For $\ell = 1$, we have simply

$$P_{c}^{k,k\rightarrow k-\ell} = \frac{1}{2Nh_{k-1} s(k-1)} \frac{k}{2k - 1}, \quad (34)$$

as we found earlier above. For $\ell = 2$ we find

$$P_{c}^{k,k\rightarrow k-2} = \frac{1}{2Nh_{k-2} s(k-2)} \frac{3k^2(k-1)^2}{k(2k - 1)(2k - 2)(2k - 3)}. \quad (35)$$

In general, we see that the coalescent probability in class $k-\ell$ is $\frac{1}{2Nh_{k-\ell} s(k-\ell)}$ times a numerical factor which depends only on $k$ and $\ell$.

This general form for the coalescence probabilities makes intuitive sense. $Nh_{k-\ell}$ is the population size of class $k - \ell$, and $\frac{1}{s(k-\ell)}$ is the average number of generations that an individual spends in class $k - \ell$ before mutating away. Since the per-generation coalescent probability in a population of size $n$ is proportional to $\frac{1}{n}$, it makes sense that the coalescent probability in class $k - \ell$ is proportional to one over the population size of this class times the number of generations individuals spend in this class. The numerical factor multiplying this basic scaling comes from the integrals over the probability distribution of mutant timings (i.e. the $dt_1$ and $dt_2$ integrals). It reflects the fact that the larger the $\ell$ (i.e. the further back in time we look for a coalescence event) the more likely it is that $t_1$ and $t_2$ are far apart, which makes it likely that the ancestors of the two individuals we are considering were not both in class $k - \ell$ at the same time, and hence could not coalesce there.

From this result, we can also form an intuitive picture of the shape of genealogies in the presence of negative selection. Since the coalescent probability per steptime is $\frac{1}{2Nh_{k-\ell} s(k-\ell)} A_{\ell}^k$ and there are typically of order $\frac{1}{s(k-\ell)}$ actual generations per step, the coalescent probability per actual generation depends on the parameters as $\frac{1}{Nh_{k-\ell}}$, where the relevant value of $\ell$ increases as we go back in time. Thus the structure of genealogies in the presence of negative selection is similar to having a variable population size as we go back in time. The precise nature of this variable population size is encoded in the fitness distribution $h_{k-\ell}$.

For example, if we imagine sampling two individuals from the same below-average fitness class, the probability distribution of their genealogies is like having a population size that initially increases and then decreases as we look backwards in time. Of course, this analogy only goes so far. Most importantly, the coalescent steptimes are related to the statistics describing genetic diversity in a different way from how normal coalescent times are usually
related to these statistics. Further, in general we will not happen to sample two individuals in the same fitness class, a complication we now turn to.

**General coalescence probabilities in the non-conditional approximation:**

Thus far we have focused on the coalescence probabilities starting from a sample of two individuals from the same fitness class \( k \). However, when we sample two individuals from the population at random, it is likely that they come from different fitness classes. In general, the probability that two individuals sampled at random from the population come from classes \( k \) and \( k' \) respectively is \( H(k, k') \), as defined in Eq. (12).

Given that we sample two individuals from classes \( k \) and \( k' \), where by convention we choose \( k' > k \), the coalescence probability in the non-conditional approximation is

\[
P_c^{k,k'\rightarrow k-\ell} = \int Q_{k-\ell}^{k}(t_1)Q_{k'}^{k-\ell}(t_2) \frac{x^2}{h_{k-\ell}} f_{k-\ell}(x)e^{-s(k-\ell)|t_1-t_2|} dx dt_1 dt_2. \tag{36}
\]

We introduce the notation \( k' \equiv k + m \), substitute in our expressions for \( Q_{k-\ell}^{k}(t) \), and evaluate the integrals in Appendix D; we find

\[
P_c^{k,k+m\rightarrow k-\ell} = \frac{1}{2Nh_{k-\ell}s(k-\ell)} A_{k,m}^{k,m}, \tag{37}
\]

where

\[
A_{k,m}^{k,m} = \frac{(k')!}{(k-\ell)!} \frac{(k-\ell)!}{(k+k' - k)!} \frac{(2\ell+k' - k)!}{(2\ell+k' - k)!}. \tag{38}
\]

Eq. (37) is the complete solution for coalescent probabilities in the non-conditional approximation. As in the previous subsection, the parameter dependence is simple and the probability of coalescence in a given fitness class is proportional to the inverse population size of the fitness class and the time an average individual spends in that fitness class. This is multiplied by a \( k, \ell, \) and \( m \)-dependent numerical factor which decreases \( m \) increases, reflecting the fact that the larger \( m \) is, the less likely the ancestors of the two sampled individuals are to have been in a given fitness class at the same time. The dependence of \( A_{k,m}^{k,m} \) on \( \ell \) is more complex, but reflects the probability that the ancestors of the two individuals we are considering were in class \( k - \ell \) at the same time.

In Fig. 3 we show examples of coalescence probabilities calculated from our theoretical framework within the non-conditional approximation for different population parameters. We see that the probability of coalescence steadily increases for longer steptimes (classes with larger fitness), and decreases with increasing selection coefficients and population size.
NUMERICAL SIMULATIONS OF THE GENETIC DIVERSITY

We compare the predictions of our effective coalescence analysis to Monte Carlo simulations of the Wright-Fisher model. In our simulations, we consider a population of constant size $N$ and we keep track of the frequencies of all genotypes over successive, discrete generations. In each generation, $N$ individuals are sampled with replacement from the preceding generation, according to the standard Wright-Fisher multinomial sampling procedure (Ewens, 2004) in which the chance of sampling an individual is determined by its fitness relative to the population mean fitness.

In our simulations, each genotype is characterized by the set of sites at which it harbors deleterious mutations and the set of sites at which it harbors neutral mutations. In each generation, a Poisson number of deleterious mutations are introduced, with mean $NU_d$, and a Poisson number of neutral mutations are introduced, with mean $NU_n$; each new mutation is ascribed to a novel site, indexed by a random number. The mutations are distributed randomly and independently among the individuals in the population (so that a single individual might receive multiple mutations in a given generation). The simulations record the time (in generations) at which each distinct genotype was first introduced.

Starting from a monomorphic population, all simulations were run for at least $\frac{1}{s} \ln(NU_d/s)$ generations, to ensure relaxation both to the steady-state mutation-selection equilibrium and to the PRF equilibrium of allelic frequencies within each fitness class. The final state of the population — i.e. the frequencies of all surviving genotypes — was recorded at the last generation. In most of the parameter regimes we explored, Muller’s ratchet proceeded during the simulation, so that the least loaded class at the end of each simulation typically contained at least 10 deleterious mutations, and often more.

In order to produce the empirical distributions of $\pi_d$ and $\pi_n$ shown in Fig. 4 and Fig. 5, respectively, we sampled 20 pairs of individuals from the final simulated population. The number of deleterious and neutral sites that differed in each sampled pair was recorded. For each parameter set, we simulated at least 100 independent populations and sampled 20 pairs of individuals from each of these replicates, in order to produce the empirical distribution of $\pi$ values shown in Fig. 4 and Fig. 5. Fig. 5 also shows the empirical distribution of real coalescence times, which was produced in the same way — by sampling 20 individuals from
each replicate population and recording the time at which their coalescent genotype first arose.

In the Monte Carlo simulations, unlike in our analytic theory, some deleterious sites may segregate in the least-loaded class, due to the action of Muller’s ratchet. Therefore, to judge the accuracy of our approximation of neglecting the ratchet, we produced two versions of the empirical $\pi_d$ distribution — choosing either to ignore or not to ignore such sites when comparing sampled individuals (dashed and dotted lines in Fig. 4).

THE STRUCTURE OF GENEALOGIES AND THE STATISTICS OF GENETIC DIVERSITY

We can now use the coalescence probabilities described above to calculate the structure of genealogies in the presence of negative selection. We can then use these genealogies to calculate various statistics describing the genetic diversity within the population. We know the coalescent probabilities in each step of our effective coalescent process, so in principle we can calculate the probability of any genealogy relating an arbitrary number of individuals using methods analogous to those used in standard neutral coalescent theory. This would then allow us to calculate the distribution of any statistic describing the genetic diversity among these individuals, again using methods analogous to neutral coalescent theory.

Here we will focus on the simplest genealogical relationship: the distribution of the time to the most recent common ancestor of two individuals, which demonstrates the main ideas in the simplest context. This allows us to calculate the distribution of the per-site heterozygosity $\pi$. This is the only statistic relevant to a sample of two individuals. In larger samples, provided the total number of individuals sampled is not too large, the coalescent probabilities between any pair of sampled individuals are independent to those between any other pair. Thus the distribution of per-site heterozygosity $\pi$ we expect in such a sample is equivalent to the distribution of $\pi$ we calculate here.

In our effective coalescent framework, it is natural to consider diversity at the negatively selected sites separately from diversity at linked neutral sites. We focus first on the distribution of coalescent steptimes and $\pi_d$, the per-site heterozygosity at negatively selected sites alone, ignoring neutral mutations. We will then turn to the connection between steptimes.
and actual times in generations, which will enable us to calculate the distribution of neutral diversity, including the per-site heterozygosity at neutral sites $\pi_n$. In analyzing data, we will of course typically not know \textit{a priori} which sites are neutral and which are negatively selected. In such a situation, we merely add up the expected diversity at neutral sites and negatively sites, so that the total expected per-site heterozygosity is $\pi = \pi_d + \pi_n$.

**Distribution of steptimes and $\pi_d$ for individuals in the same fitness class:**

We begin by imagining that we sample two individuals at random from the \textit{same} fitness class $k$. We wish to know the distribution of steptimes $\ell$ before they coalesce. By construction, the number of negatively selected sites at which they will be polymorphic is twice this coalescent steptime, $\pi_d = 2\ell$.

We have seen above that the probability that the two individuals are genetically identical at negatively selected sites (i.e. $\pi_d = 0$, and the coalescent steptime is 0) is $P_c^{k,k\rightarrow k} = \int_0^\infty x^2 f_k(x)/h_k^2$. The probability that the steptime is 1, $\phi_k^k(\tau = 1)$ (and hence $\pi_d = 2$) is the probability that two individuals are not identical times the probability that they do coalesce at the first step, which is

$$\phi_k^k(\tau = 1) = (1 - P_c^{k,k\rightarrow k})P_c^{k,k\rightarrow k-1}. \quad (39)$$

In general, the probability that two individuals both sampled from class $k$ coalesce a steptime $\ell$ ago is the probability that they coalesced in class $k - \ell$ times the probability that they did not coalesce in any class before this. Hence we have

$$\phi_k^k(\tau = \ell) = \rho(\pi_d = 2\ell) = P_c^{k,k\rightarrow k-\ell} \prod_{j=0}^{\ell-1}(1 - P_c^{k,k\rightarrow k-j}), \quad (40)$$

where $\rho(\pi_d = 2\ell)$ is the probability $\pi_d = 2\ell$.

**General distribution of steptimes and $\pi_d$:**

In general, if we sample two individuals at random from the population, they will not come from the same fitness class. Instead, one will come from class $k$ and one from class $k'$. We arbitrarily choose $k' > k$ and define $k' = k + m$ as before. The distribution of $k$ and $k'$ is $H(k,k')$, as given in Eq. (12).

Given $k$ and $m$, the two individuals coalesce in class $k - \ell$ with probability

$$\phi_k^{k+m}(\tau = \ell) = P_c^{k,k+m\rightarrow k-\ell} \prod_{j=0}^{\ell-1}(1 - P_c^{k,k+m\rightarrow k-j}). \quad (41)$$
We refer to this as a coalescent steptime of $\ell$, even though it also involves $m$ additional steps for one of the two individuals. We have $\pi_d = 2\ell + m$.

We can combine the distributions of $k$ and $k'$ with the distribution of coalescent steptimes given $k$ and $m$ to find the distribution of steptimes for the coalescence of two randomly chosen individuals from the population. We have

$$\phi(\tau = \ell) = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} H(k, k + m) \phi^k_{k+m}(\tau = \ell).$$

(42)

The distribution of $\pi_d$ has a slightly different form,

$$\rho(\pi_d) = \sum_{\ell=0}^{\pi_d/2} \sum_{k=0}^{\infty} H(k, k + m = k + \pi_d - 2\ell) \phi^k_{k+m}(\tau = \ell),$$

(43)

where the first sum runs from $\ell = 0$ to the largest integer less than or equal to $\pi_d/2$. Note that in practice we only have to evaluate the sum over $k$ from 0 to a multiple of $U_d/s$, since $H(k, k + m)$ will be negligible for larger $k$.

These results for the distributions of genealogy lengths and of $\pi_d$ involve several sums that must be computed numerically. However, all the terms in these sums are straightforward and the numerical evaluations of their values are simple and fast. In Fig. 4 we show a representative example of the predicted distribution of the per-site heterozygosity at negatively selected sites, $\rho(\pi_d)$, compared to simulation results. We explore the significance of the shape of the distribution $\rho(\pi_d)$, how this distribution depends on the parameter values, and the source of the small but systematic deviations between the theoretical predictions and the simulation results in the Discussion.

The relationship between steptimes and time in generations, and the neutral heterozygosity $\pi_n$:

So far we have focused on the genealogies measured in steptimes, which allowed us to calculate the distribution of heterozygosity among negatively selected sites. We would now like to relate the steptimes to actual times in generations. To do this, we consider the probability that a coalescence event occurred at time $t$ given that the individuals were initially in classes $k$ and $k + m$ and coalesced in class $k - \ell$. Conditional on coalescence in class $k - \ell$, the distribution of times $t_1$ and $t_2$ since the ancestors of the two individuals originally mutated from class $k - \ell$ to class $k - \ell + 1$ is given by

$$R^{k-\ell}_{k,k+m}(t_1, t_2) = KQ^{k-\ell}_{k,k+m}(t_1, t_2)e^{-s(k-\ell)|t_1-t_2|},$$

(44)
where $K$ is a normalization factor,
\[
\frac{1}{K} = A_m^{k,\ell}.
\] (45)

The actual time at which the two individuals coalesced is approximately the longer of the two times ago at which this original mutation happened, plus the time for them to coalesce within class $k - \ell$, which is approximately equal to the time at which this lineage mutated from class $k - \ell - 1$. Thus the distribution of actual coalescence time of two individuals from classes $k$ and $k + m$ conditional on their coalescing in class $k - \ell$ is approximately
\[
\psi(t|k,k',\ell) = \int_0^t R_{k,k'}^{k-\ell}(t_1,t) dt_1 + \int_0^t R_{k,k'}^{k-\ell}(t,t_2) dt_2 \ast Q_{k-\ell-1}^{k-\ell}(t),
\] (46)

where $\ast$ refers to a convolution. We carry out these integrals in Appendix E, and find
\[
\psi(t|k,k',\ell) = \frac{s e^{-s(k+k')t} \left( e^{st} - 1 \right)^{2\ell+k'-k-1} [k+k']!}{[2\ell + k' - k - 1]! [2k - 2\ell]!} \ast \left[ s(k - \ell) e^{-s(k-\ell)t} \right].
\] (47)

Carrying out this convolution gives
\[
\psi(t|k,k',\ell) = \frac{s(k - \ell) e^{-s(k-\ell)t} \left( -1 \right)^{2\ell+k'-k-1} [k+k']!}{[2\ell + k' - k - 1]! [2k - 2\ell]!} \times \left[ \sum_{i=0}^{2\ell+k'-k-1} (-1)^i \binom{2\ell + k' - k - 1}{i} \frac{1 - e^{-st(k'+\ell-i)}}{k'+\ell-i} \right].
\] (48)

Evaluating this expression in practice requires numerical computation of the sum, but this does not present any numerical difficulties; it is fast and straightforward. If a simpler analytical approximation is desired, for moderate to large $U_d/s$ the time to coalesce within class $k - \ell$ can be neglected. In this approximation $\psi(t|k,k',\ell)$ is given simply by the term preceding the convolution in Eq. (47).

Note that Eq. (47) and Eq. (48) do not apply when $\ell = k$; in this case coalescence occurs in the class of individuals with 0 deleterious mutations and the convolution in Eq. (47) does not apply. We will primarily be interested in the situation when the coalescence time within the 0-class is small compared to the coalescence time through the fitness distribution. Thus we simply neglect this coalescence time, and in the case where coalescence occurs in the 0-class we have
\[
\psi(t|k,k',\ell = k) = s(k + k') e^{-s(k+k')t} \left( e^{st} - 1 \right)^{k+k'-1}.
\] (49)

In the alternative regime where the coalescence time in the 0 class is not short, this coalescence is a neutral process within a class of size $N h_0$, so we could generalize Eq. (49) by
convolving it with \( Q_0(t) = N h_0 e^{-t/(N h_0)} \). However, as noted in the Discussion, this is the regime when the background selection approximation is accurate anyway and our detailed analysis is not necessary, so we do not pursue this further here.

Averaging over the possible values of \( k, m, \) and \( \ell \), we find the overall distribution of actual coalescent time between two randomly chosen individuals,

\[
\psi(t) = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \sum_{\ell=0}^{k} \psi(t|k,k',\ell) \phi_k^{k+m} \psi(t) \phi_k^{k+m}(\tau = \ell) H(k,k + m),
\]

where the distributions \( H(k,k + m), \phi_k^{k+m}(\tau = \ell), \) and \( \psi(t|k,k',\ell) \) are as given above.

From this distribution of times to common ancestor for two randomly chosen individuals, we can calculate the distribution of \( \pi_n \), the neutral heterozygosity. Since the neutral mutations occur as a Poisson process with rate \( U_n \), and there are a total of \( 2t \) generations in which these mutations can occur, \( \pi_n \) follows a Poisson distribution with mean \( U_n t \), where \( t \) is drawn from the distribution of coalescence times, Eq. (50). We have

\[
\rho(\pi_n) = \int_0^{\infty} \frac{[2U_n t]^{\pi_n}}{\pi_n!} e^{-2U_n t} \psi(t) dt.
\]

In Fig. 5, we compare this distribution of neutral heterozygosity (as modified by the corrections described in Appendix A) to direct simulations. We find good general agreement to the shape of the distribution, though the theory slightly underestimates the mean \( \pi_n \) (presumably due to effects of Muller’s ratchet, which we explore further in the Discussion). Note that, like our results for the diversity at negatively selected sites, these results differ dramatically from the exponential distribution a neutral or background selection model would predict; we describe these comparisons further in the Discussion.

We note that to calculate the distribution of total heterozygosity \( \pi = \pi_n + \pi_d \), we must account for the fact that \( \pi_d \) and \( \pi_n \) are not independent: large \( \pi_d \) means a large coalescent steptime and hence makes a large \( \pi_n \) more likely. The distribution of \( \pi_d \) is independent of \( \pi_n \), and is given by \( \rho(\pi_d) \) above. Above we found \( \psi(t|k,k',\ell) \), which implies that

\[
\rho(\pi_n|k,k',\ell) = \int_0^{\infty} \frac{[2U_n t]^{\pi_n}}{\pi_n!} e^{-2U_n t} \psi(t|k,k',\ell) dt.
\]

Since \( \pi_d = 2\ell + k - k' \), this implies

\[
\rho(\pi_n|\pi_d) = \sum_{\pi_d=2\ell+k-k'} \rho(\pi_n|k,k',\ell).
\]
The distribution of $\pi$ is then given by
\[
\rho(\pi) = \sum_{\pi_n + \pi_d = \pi} \rho(\pi_d) \rho(\pi_n | \pi_d).
\]
(54)

This is no more difficult to calculate than $\rho(\pi_n)$, since it involves analogous sums. However, while the distribution of $\pi$ is clearly important in analyzing sequence data, in this paper we focus on the distributions of $\pi_n$ and $\pi_d$ separately, which provides a more complete picture of the source of all aspects of the genetic variation.

**The mean pairwise heterozygosity:**

Above we have calculated the distribution of heterozygosity for both neutral and deleterious mutations. It is straightforward to average these results to calculate the mean pairwise heterozygosity for both neutral and deleterious mutations. In Fig. 6 and Fig. 7 we show how this mean heterozygosity depends on population size, mutation rate, and selection strength, for neutral and deleterious mutations respectively. We see that in contrast to the purely neutral case, the dependence on the population size is fairly weak: while both $\langle \pi_d \rangle$ and the mean real coalescence time (and hence $\langle \pi_n \rangle$) increase roughly linearly with $N$ in the weak selection regime $Ns \sim 1$, this quickly saturates and for $Ns$ substantially greater than 1 the mean heterozygosity becomes almost independent of population size. The dependence on $U_d/s$, by contrast, is much stronger. These results make intuitive sense, particularly in light of the “foreground selection” approximation that we introduce in the Discussion, where we discuss these figures in more detail.

**Statistics in larger samples:**

The distributions of $\pi_n$ and $\pi_d$ described above are very different from the distributions of heterozygosity expected in the absence of selection. We could certainly measure the distribution of pairwise heterozygosity from a sample of many individuals from a population, and use this to infer the action of selection. However, it may also be useful to understand the expected distribution of other statistics describing the variation in larger samples. The relationship between these different statistics will typically be different than expected in the neutral case, making them useful in constructing other statistical tests for selection.

One statistic often used to describe variation in larger samples is the total number of segregating sites among a sample of $n$ individuals, $S_n$. Here we describe how our framework allows us to calculate the distribution of $S_3$; similar methods can be used to calculate the
distribution of $S_n$ for larger $n$. One common test for neutrality, Tajima’s $D$, is based on a comparison between the observed values of $\pi$ and $S_n$; our results for $S_3$ could in principle be used to show how this statistic should be expected to behave in the presence of purifying selection. As we will see, this is unwieldy to calculate in our framework, so here we merely lay out a prescription for calculating $S_3$.

We first consider the distribution of $S_d^3$, the number of segregating negatively selected sites among three randomly sampled individuals. In order to calculate the probability a sample has a particular $S_d^3$, we imagine picking three individuals at random from the population and calculate the probability of the coalescence events that lead to that $S_d^3$.

To illustrate our approach, we start with the special case where all three individuals are selected from the same fitness class $k$. This is illustrated in Fig. 8a. Two of these three lineages coalesced after steptime $\ell$, in class $k - \ell$. We call this steptime at which two of the three lineages coalesced $\tau_3$. Since all three lineages are equivalent, the probability that two of the three coalesce in step $\ell$ is $\binom{3}{2} P_{c}^{k,k \rightarrow k-\ell}$. Thus the distribution of $\tau_3$ is given by

$$\chi(\tau_3 = \ell) = \binom{3}{2} P_{c}^{k,k \rightarrow k-\ell} \prod_{j=0}^{\ell-1} \left(1 - \binom{3}{2} P_{c}^{k,k \rightarrow k-j}\right).$$

(55)

We next need to calculate the distribution of $\tau_2$, the total steptime to common ancestry of the three individuals (see Fig. 8a). This time of course cannot be smaller than $\tau_3$. Given a particular value of $\tau_3$, the probability the remaining two lineages coalesced in class $\tau_3 - 1$ is just $P_{c}^{k,k \rightarrow k-\tau_3-1}$, and so on. Thus we have

$$\chi(\tau_2 = r | \tau_3 = \ell) = P_{c}^{k,k \rightarrow k-r} \prod_{j=\ell}^{r-1} \left(1 - P_{c}^{k,k \rightarrow k-j}\right),$$

(56)

valid for $r \geq \ell$. For $r < \ell$ this probability is simply 0. Given values of $\tau_3$ and $\tau_2$, it is clear from Fig. 8a that the total number of segregating negatively selected sites is $S_d^3 = 2\tau_2 + \tau_3$.

We now turn to the general situation where three individuals are selected at random from the population, as illustrated in Fig. 8b. We adopt the notation that the individuals came from fitness classes $k, k'$, and $k''$, where by convention we choose $k'' \geq k' \geq k$. The
probability the three classes are \( k'' \), \( k' \), and \( k \) is \( H(k'', k', k) \), where

\[
H(k'', k', k) = \begin{cases} 
  h_k^3 & \text{if } k'' = k' = k \\
  3h_k'h_k^2 & \text{if } k' = k 
eq k'' \\
  3h_k'h_k & \text{if } k'' = k' 
eq k \\
  6h_k''h_k'h_k & \text{otherwise}
\end{cases}
\]  

(57)

Analogous to before, we define \( \tau_3 \) to be the steptime for coalescence of the first two lineages (measured from 0 in class \( k'' \), see Fig. 8b). It is clear that no two lineages can coalesce above class \( k' \), and hence \( \tau_3 \) must be at least \( k'' - k' \). The probability that \( \tau_3 = k'' - k' + \ell \) is given by

\[
\chi_<(\tau_3 = k'' - k' + \ell) = P_{\ell+1}^c P_{k'' \rightarrow k'}^{k'' - \ell} \prod_{j=0}^{k-k'} P_{c}^{k'' \rightarrow k'-j}(1 - P_{c}^{k'' \rightarrow k'-j})(1 - P_{c}^{k,k'' \rightarrow k'-j})
\]

valid for \( \ell < k' - k \). For \( \ell \geq k' - k \), there are three possible coalescence events. The total coalescence probability at each step is the sum of the probabilities of each of these events. Thus we have

\[
\chi_>(\tau_3 = k'' - k' + \ell) = \left[ 1 - \left( (1 - P_{c}^{k,k'' \rightarrow k'-j})(1 - P_{c}^{k,k'' \rightarrow k'-\ell})\right) \right] \\
\times \prod_{i=0}^{k'-k-1} (1 - P_{c}^{k',k'' \rightarrow k'-i}) \\
\times \prod_{j=k' - k}^{\ell-1} (1 - (P_{c}^{k',k'' \rightarrow k'-j})(1 - P_{c}^{k'',k \rightarrow k'-j}))(1 - P_{c}^{k,k'' \rightarrow k'-j})
\]

(59)

Note that Eq. (59) reduces to Eq. (55) when \( k'' = k' = k \), as we would expect.

Putting these results together, we see that the distribution of \( \tau_3 \) conditional on the values of \( k'' \), \( k' \), and \( k \) is given by

\[
\chi(\tau_3 | k'', k', k) = \begin{cases} 
  \chi_<(\tau_3 = k'' - k' + \ell) & \text{for } 0 \leq \ell < k' - k \\
  \chi_>(\tau_3 = k'' - k' + \ell) & \text{for } k' - k \leq \ell \leq k'
\end{cases}
\]  

(60)

Averaging over the values of \( k'' \), \( k' \), and \( k \), we see that the overall distribution of \( \tau_3 \) is given by

\[
\chi(\tau_3) = \sum_{k=0}^{\infty} \sum_{k''=k}^{\infty} \sum_{k'=k}^{\infty} \chi(\tau_3 | k'', k', k) H(k'', k', k).
\]

(61)

We now wish to calculate the distribution of time \( \tau_2 \). In general \( \tau_2 \) will depend on \( \tau_3 \), because by definition \( \tau_2 \geq \tau_3 \). In addition, \( \tau_3 \) will depend on \emph{which} of the three lineages coalesced first. There are four possible scenarios, illustrated in Fig. 9. The situation in Fig.
9a always applies whenever $\tau_3 < k'' - k$. When $\tau_3 \geq k'' - k$, the situation in Fig. 9b applies with probability

$$p_b = \frac{P_{c}^{k',k''\rightarrow k''-\tau_3}}{P_{c}^{k',k''\rightarrow k''-\tau_3} + P_{c}^{k,k''\rightarrow k''-\tau_3} + P_{c}^{k,k'\rightarrow k''-\tau_3}}.$$  \hspace{1cm} (62)

The situations in Fig. 9c and Fig. 9d occur with analogously defined probabilities $p_c$ and $p_d$, respectively.

Given $\tau_3$ and the coalescence ordering scenario which applies, we can calculate the probability distribution of $\tau_2$. For example, in scenario b, $\tau_2$ is at least $k$ and takes on a value $k + \ell$ with a probability equal to the probability that the individual starting in class $k''$ coalesced with the individual starting in class $k$ in class $k - \ell$, conditional on them not coalescing before class $k'' - \tau_3$. Similar results hold for the other scenarios. This is complicated because the probability of coalescence of the $k''$ individual with the $k$ individual depends on $\tau_3$, since the fact that the $k''$ individual coalesced with the $k'$ individual in class $k'' - \tau_3$ affects the probability distribution of the time at which ancestors of the $k''$ individual were in class $k'' - \tau_3$. This in turn affects the probability the $k''$ individual coalesces with the $k$ individual in any particular fitness class. Thus in order to calculate the distribution of $\tau_3$, we must know the probability distribution of the time at which the ancestor of the $k''$ individual mutated from class $k'' - \tau_3$, conditional on it having coalesced with the $k'$ individual in that class. We have already calculated this time; it is given by Eq. (47), with appropriate values of $k, k', \text{ and } \ell$. Using this, we can calculate the distribution of $\tau_2$ in each scenario, and by averaging over the probabilities of each scenario we obtain the overall distribution of $\tau_2$.

The number of segregating sites $S_{3d}$ is given by

$$S_{3d} = \tau_3 + 2\tau_2 - (k'' - k) - (k'' - k').$$  \hspace{1cm} (63)

Thus using the distributions of $\tau_3$ and $\tau_2$ conditional on $k''$, $k'$, and $k$ as described above, we can calculate the full distribution of $S_{3d}$. Given a particular value of $S_{3d}$, there is a relationship between the steptimes and actual times (analogous to Eq. (47)), which we could use to find the distribution of the total number of segregating neutral sites $S_{3n}$. However, while this analysis provides a prescription for calculating the distribution of $S_{3d}$ and $S_{3n}$, it is clear that the full distributions are opaque and involve extensive numerical calculations. These computational complexities are tangential to the ideas behind our framework, so we do not pursue them further here, though they will be important to explore in future work aiming
to use this framework for data analysis. However, in the Discussion we do provide a simple approximation for $S_n$ in a specific parameter regime we refer to as the “foreground selection” regime.

**DISCUSSION**

In recent years, both experimental studies and sequence data have pointed to the general importance of selective forces among many linked variants in microbial and viral populations, and on short distance scales in the genomes of sexual organisms (Betancourt, 2009; Bollback and Huelsenbeck, 2007; de Visser et al., 1999; Desai et al., 2007; Hahn, 2008). Our analysis provides a framework for understanding how one particular type of selection — pervasive purifying (i.e. negative) selection against deleterious mutations — affects the structure of genetic variation at the negatively selected sites themselves and at linked neutral loci. In other words, our work provides a way to understand the statistics of genealogies in the presence of Hill-Robertson interference. This type of selection is presumably widespread in many populations, in which there is a selective pressure to maintain existing genotypes and mutations away from these genotypes at a variety of loci are deleterious.

A variety of earlier approaches have addressed aspects of this problem. The Poisson Random Field method of Sawyer and Hartl (1992) provides a complete description of the statistics of genetic variation at negatively selected sites, but assumes that these sites are completely unlinked from any other variation. The ancestral selection graph (ASG) introduced by Neuhauser and Krone (1997), by contrast, provides a logically complete framework for computing the structure of genealogies in the presence of selection on many linked sites. However, this approach has proven to be numerically intractable (Przeworski et al., 1999). Alternative approaches have studied how particular types of strong selection affect the structure of genealogies (Kaplan et al., 1988, 1989), how one or a few selected sites affect linked neutral variation (Barton, 1998; Barton and Etheridge, 2004; Gillespie, 2001, 2000; Ohta and Kimura, 1975; Smith and Haigh, 1974), or how selection on many very weakly linked sites affects genetic variation (Barton, 1995; Otto and Barton, 1997). However, other than the background selection approximation (Charlesworth et al., 1993), which we discuss in detail below, and the recent work in a continuous-fitness
model by O’Fallon et al. (2010), which works well in the weak-selection regime but not the $Ns \gg 1$ situation we study here, none of these approaches provides a way to efficiently calculate the statistics of genetic variation or the structure of genealogies when selection acts simultaneously among many strongly linked sites.

These earlier analyses have either attempted to treat selected mutations independently (or mostly independently) from one another, or they have modified the statistics of real ancestral coalescent processes (e.g. the ASG) to account for the distortions caused by selection. Instead of following the true ancestral process of each individuals, our key insight is to develop an *effective* genealogical approach which focuses on how individuals “move” through the fitness distribution. Here each mutation plays the role of a reproductive event that moves individuals through the fitness distribution, and each fitness class is a “generation” in which coalescence can occur with some probability. We calculate this probability using a simple approximation based on the PRF model, which we developed in Desai et al. (2010), rather than by considering the actual reproductive process within that class. This takes advantage of the insight that while the frequencies of mutations at different sites are not independent, the frequencies of genetically distinct alleles *are* (approximately). In this paper we have determined the relationships between alleles, which depends on their frequencies (and hence makes frequencies of mutations at individual sites correlated). We have calculated the distribution of genetic diversity at a per-site level despite these correlations by starting with the independent frequencies of different alleles, and then combining this with the genealogical relationships between them.

Our approach leads to simple expressions for the coalescent probability at each step in our effective genealogical process. This makes it a complete effective coalescent theory: using these probabilities, we can calculate the probability that a sample of individuals has any particular ancestral relationship. Our coalescent probabilities are different from those in the standard Kingman coalescent (Kingman, 1982), so the structure of genealogies has a different form.

Of course, since our process is an effective rather than an actual coalescent, the relationship between an effective genealogy and the expected statistics of genetic variation given that genealogy is different than in the standard neutral coalescent. Given a particular genealogy measured in steptimes, the numbers of deleterious mutations *are* the coalescent times, and
to calculate the statistics of neutral variation we have to make use of the relationship between steptimes and actual coalescence times. This contrasts with the Kingman coalescent, where numbers of neutral mutations are typically Poisson-distributed variables with means proportional to coalescence times (Wakeley, 2009). However, we can account for these differences by starting with the distribution of effective genealogies and then converting these genealogies into actual coalescence times.

In this paper, we have calculated simple statistics describing genetic variation, in particular the distribution of pairwise heterozygosity. This leads to an analytic expression for the quantities of interest, although this expression involves sums which are most easily calculated numerically. These are easy to compute, and do not become harder to evaluate in larger populations, and hence are more efficient to evaluate than either simulations or calculations within the ancestral selection graph.

**Approximations underlying our approach:**

Our analysis relies on three key approximations. First, we assume that the PRF formulas describe the frequencies of lineages within each fitness class. This requires that each lineage is approximately independent of the others. In Desai et al. (2010), we showed that this will generally hold in class $k$ whenever $\gamma_k \gg 1$. This always holds provided that $Ns \gg 1$, but will also be reasonable within the bulk of the fitness distribution provided $NU_d \gg 1$ even when $Ns \sim 1$. In other words, our approach is valid provided either $Ns \gg 1$ or the total genome-wide $\theta_d$ is large (or region-wide $\theta_d$ is large if we are considering a smaller fully linked part of a sexual genome). This is consistent with our focus on situations involving many linked selected sites. Related to this approximation, we have also implicitly assumed that the probability a lineage in class $k$ reaches a frequency close to $h_k$ can be neglected. This will typically be true in the bulk of the fitness distribution, but can break down in the high-fitness tail; we discuss this problem and the methods we use to handle it in Appendix A.

Our second key approximation is the non-conditional approximation, which we discuss in more detail in Appendix B. Finally, we assume that Muller’s ratchet can be neglected. This final assumption is more problematic; we discuss it in detail below.

Although we have focused primarily on situations when selection is weak compared to total deleterious mutation rates, our approach is also valid for both strong and weak selective
pressures. However, when selection is sufficiently strong \((Ns \gg 1 \text{ and } U_d/s < 1)\), then background selection accurately describes the patterns of genetic variation (see below). Thus our methods are primarily useful for situations where selection is weak compared to mutation rates (both when \(Ns \gg 1\) and \(Ns \lesssim 1\)).

**An Intuitive Picture of the Structure of Genealogies:**

Our numerical formulas for the statistics of genetic variation can be somewhat opaque, so we pause now to develop an intuitive picture of the shape of typical genealogies. In general the probability that two individuals will coalesce within class \(k\) has the general form \(P_c = A n_k/|s_k|\), where \(n_k\) is the population size of that class, \(s_k\) is the effective selection pressure against individuals within that class, and \(A\) is a constant that depends on which classes the lineages began in, but not on any of the population parameters. Since the lineages spend roughly \(1/|s_k|\) generations in each class, this means that the per-generation coalescent probability within class \(k\) is proportional to \(1/n_k\). This leads to a simple intuitive picture of the coalescent process. Imagine we picked two individuals from the same fitness class \(k\). They spend \(1/|s_k|\) generations in class \(k\), and during that time they have a probability proportional to \(1/n_k\) per generation of coalescing. If they fail to coalesce, they then move to class \(k - 1\), where they spend \(1/s_{k-1}\) generations together (times the appropriate \(A\) factor) and have a probability proportional to \(1/n_{k-1}\) per generation of coalescing. If they again fail to coalesce, they move to class \(k - 2\), and so on.

This picture suggests that genealogies in the presence of purifying selection look like neutral genealogies with a specific type of historical population size dependence. For two individuals sampled from some fitness class \(k\) less fit than the mean, \(k > U_d/s\), the distribution of coalescence times is like that in a neutral coalescent with a population size that was initially small, then increased, and then decreased again in the more recent past. To be specific, it is as if the population size was \(n_k\) for the first \(1/|n_k|\) generations, \(n_{k-1}\) for the next \(1/s_{k-1}\) generations, and so on, where \(n_k\) is proportional to a Poisson distribution (as a function of \(k\)) with mean \(U_d/s\). For two individuals more fit than the mean, it is as if the population size began large and decreased moving backward in time, again having size \(n_k\) for \(1/|n_k|\) generations. For these two individuals sampled from this class, selection is indistinguishable from this particular historically varying population size (although this particular type of variation in population size is presumably rather unusual). The distribution of coalescence
times between this pair of individuals looks the same as neutral coalescent histories with this specific population size history. The deleterious mutation rates and selection pressures only matter in that they determine the form of this population size history.

However, the key difference from a neutral population of time-varying size is that pairs of individuals do not typically come from the same fitness class. Rather, they come at random from different parts of the fitness distribution, and those that come from different places have ancestries characterized by different historically varying population sizes. The total distribution of ancestry is the sum of all of these. In other words, the genetic variation within the population is like that in a population where some individuals had one type of historical population size history, while others had another. This leads to qualitative differences from neutral expectations, which we now explore in more detail.

**Comparison with neutrality:**

In neutral coalescent theory, the distribution of coalescence times between any number of individuals is exponentially distributed. The mean of this exponential distribution is proportional to the population size \( N \). This means that the distribution of neutral heterozygosity \( \pi_n \) is also exponentially distributed with mean \( 2NU_n \) (in a haploid population; an additional factor of 2 applies in diploids). Note that the most likely value of \( \pi_n \) is 0, with larger values of \( \pi_n \) always being less likely than smaller values.

Since the statistics of purely neutral diversity depend only on a single parameter, \( \theta = 2NU_n \), the expected mean \( \pi_n \) corresponds to a specific expected distribution of all other aspects of genetic diversity. These expected relationships between different statistics describing genetic variation have led to a number of statistical tests comparing the observed values of these quantities to check for deviations from neutrality. Some simple types of selection lead only to a shift in the effective population size, which means that distributions of all statistics describing genetic variation are identical to the neutral case, but are parameterized by some \( N_e \) which does not correspond to the actual population size. When this is true, it is impossible to distinguish this form of selection from neutral evolution in a population of a different size. We have seen here that this is not the case for pervasive purifying selection. This type of selection fundamentally changes the form of the distributions of statistics describing genetic variation, and alters the relationship between different statistics, all in ways that cannot be reduced to a shift in effective population size. This
means that in principle it would be possible to develop statistical tests to infer this type of selection pressure from sequence data. Our theoretical framework allows us to calculate the form we expect deviations from neutral expectations to take. It remains for future work to use this as a basis for finding more powerful methods to detect negative selection in sequence data.

One of the most striking differences between our results and neutral expectations is that the distribution of heterozygosity (both $\pi_n$ and $\pi_d$) has a nonzero peak (see Fig. 4 and Fig. 5). That is, it is very unlikely that two individuals are extremely closely related. This reflects the fact that the distribution of coalescence times has a peak at of order $\frac{1}{s} \ln \left( \frac{U_d}{s} \right)$, and coalescence times much shorter or longer than this are very unlikely. This stands in stark contrast with the case of the neutral coalescent, where the distribution of coalescence times is exponential. It remains true that selection against linked deleterious mutations on average tends to make individuals more closely related (and hence reduces heterozygosity). But while selection reduces average coalescence times it also changes the shape of the distribution of relatedness, so that it is less likely that that two individuals are very closely related than that they have common ancestors in the medium-term past.

**Relationship with background selection:**

CHARLESWORTH *et al.* (1993) considered how selection against many linked deleterious mutations affects linked neutral diversity in a model identical to ours, developing an approach that has become known as background selection (BGS). Background selection makes a simple assertion: in the presence of selection against many linked deleterious variants, the shape of genealogies is identical to the neutral case, with a reduced effective population size $N_e = N_e - U_d/s$. This means that the variation at linked neutral sites is also characteristic of purely neutral evolution, but with reduced effective population size.

The idea behind the background selection hypothesis is that deleterious mutations are quickly eliminated from the population by selection. Thus if we sample individuals from the population, they must have very recently descended from individuals within the class of individuals which had no deleterious mutations (the 0-class). The BGS approximation assumes that the time for this to happen can be neglected, and that individuals never coalesce before it does. These individuals then coalesce within the 0-class as a neutral process with effective population size equal to the size of that 0-class, which is $Ne^{-U_d/s}$. Thus the genetic
diversity within the population is identical to that in a neutral population of reduced size $N e^{-U_d/s}$.

We can estimate the conditions required for the background selection approximation to be valid. An individual in fitness class $k$ has typically taken a time $\sum_{j=1}^{k} \frac{1}{s_k}$ to descend from the 0-class to class $k$. The average value of $k$ is $\frac{U_d}{s}$, so an average individual will have taken a time

$$t \sim \frac{1}{s} \ln \left( \frac{U_d}{s} \right)$$

(64)

to descend through the distribution (assuming $U_d > s$). Within the 0-class, the ancestor of this individual will coalesce with other individuals at a rate given by the inverse size of this 0-class; the typical coalescence time will be $N e^{-U_d/s}$. Background selection will be accurate when this coalescence time within the 0-class is long compared to the time it has taken for an individual to have descended from the 0-class. This means BGS requires

$$N s e^{-U_d/s} \gg \ln \left( \frac{U_d}{s} \right).$$

(65)

Because of the exponential term on the left hand side of this expression, it is clear that background selection is a strong-selection, weak-mutation limit. It will tend to be valid provided that $Ns > 1$ and $U_d < s$, but whenever $U_d$ becomes much larger than $s$, it will typically break down even in enormous populations.

Our analysis is an extension of the background selection approach. We study exactly the same model, but we do not assume that the coalescence time through the fitness distribution is small compared to the coalescence times within the 0-class, or that coalescence cannot occur among individuals carrying deleterious mutations. It is precisely these two effects that lead to distortions away from the neutral expectations, making it impossible to describe genealogies using neutral theory with a revised effective population size.

Although our analysis is a generalization of background selection, it is not inconsistent with it. We have focused primarily on situations where the background selection approximation breaks down, and coalescence times through the fitness distribution are large compared to those in the 0-class, because this is the situation where a generalization of BGS is most useful. Because of this focus, we have often neglected the coalescence times within the 0-class, if coalescence occurs there. We described above how we could generalize Eq. (49) to include this time; this addition causes our results to reduce to the BGS predictions in the
limits where we expect BGS to be valid.

Note also that in many situations it may be the case that there are many linked weakly selected mutations and many linked strongly selected mutations. In such circumstances, the process we consider and background selection can act simultaneously. Imagine we had one class of mutations with fitness cost $s_1$ which occur with mutation rate $U_1$, where $U_1 < s_1$ and $N s_1 \gg 1$ so that background selection applies. At the same time, imagine another class of mutations with fitness cost $s_2$ which occur with mutation rate $U_2$, where $U_2 \gg s_2$ so that background selection breaks down for these mutations. In this case, the genetic diversity we expect to see will be characteristic of our effective coalescent theory (with $U_d = U_2$ and $s = s_2$), but with a reduced effective population size $N_e = N e^{-U_1/s_1}$. In other words, background selection against the strongly selected mutations means that all individuals are very recently descended from an individual that had no large-effect mutations, but that the coalescence time through the distribution of weakly selected mutations cannot be neglected.

**A “Foreground Selection” Approximation:**

We have seen that our analysis accounts for two effects missing from background selection: coalescence events outside the 0-class, and the time it takes for individuals to have descended from the 0-class. Whenever $U_d/s$ and $N$ are both sufficiently large, the former effect can be neglected while the latter is still important, because the number of lineages in each fitness class becomes large and hence coalescence events are very unlikely to occur outside of the 0-class. This leads to an approximation which we can think of as the opposite of BGS, or “foreground selection” (FGS) for short. In this approximation, we assume that all individuals coalesce within the 0-class, as with background selection. However, unlike BGS, in this regime the coalescence time within the 0-class can be neglected and the time is instead determined entirely by the time it took for those individuals to descend from the 0-class. This approximation is valid for large $U_d/s$ in the limit of large $N$ (provided always $N s e^{-U_d/s} \ll \ln(U_d/s)$).

In this foreground selection limit our results become much simpler and provide a useful intuitive picture of the structure of genealogies and genetic variation. Consider the deleterious heterozygosity $\pi_d$ of two individuals sampled from fitness classes $k$ and $k'$. In this approximation, these two individuals always coalesce in the 0-class so we always have $\pi_d = k + k'$. Since two individuals are sampled from classes $k$ and $k'$ with probability
\[ \rho(\pi_d = r) = \sum_{k=r-k'} H(k, k') = e^{-2U_d/s} \frac{1}{r!} \left( \frac{2U_d}{s} \right)^r. \] (66)

This simple approximation makes it clear why the distribution of \( \pi_d \) looks the way it does, and explains how it varies with \( U_d/s \) and with \( N \), both in this foreground selection limit and more generally. For large \( N \), when coalescence outside the 0-class can be neglected, two individuals from class \( k \) and \( k' \) have \( \pi_d = k + k' \). Thus the distribution of \( \pi_d \) has roughly the same shape as the distribution of fitness within the population. The mean \( \pi_d \) is \( 2U_d/s \), since the average individual comes from class \( k = U_d/s \). Smaller and larger \( \pi_d \) are less likely; the distribution of fitness in the population has variance equal to the mean, so the variance of the distribution of \( \pi_d \) is also roughly equal to its mean. As \( N \) gets smaller, there is sometimes coalescence outside of the 0-class. This reduces \( \pi_d \) given \( k \) and \( k' \). Hence as we reduce \( N \), the distribution of \( \pi_d \) shifts somewhat leftwards, with a peak somewhat below \( 2U_d/s \), and has slightly more variance since there is a less definite correspondence between \( k, k' \), and \( \pi_d \). Since \( \pi_n \) is determined by \( \pi_d \), this also explains why the distribution of \( \pi_d \) has the peaked form we observe, and how it depends on \( U_d/s \) and \( N \). All of these intuitive expectations are reflected in our results, as shown in Fig. 4, Fig. 5, Fig. 6, and Fig. 7. Note for example that in Fig. 4, the peak of \( \pi_d \) is slightly below \( 2U_d/s \) (reflecting the finite population size) and has variance about equal to its mean; we have verified that as \( N \) increases the shape of the distribution remains roughly the same, but the mean increases towards \( 2U_d/s \) and the variance decreases slightly.

More complex statistics of sequence variation are similarly straightforward to calculate in the foreground selection approximation. When considering larger samples, the genetic diversity is determined by the fitness classes these individuals come from, which is always simple since the probability a given individual is sampled from fitness class \( k \) is just the Poisson-distributed \( h_k \). This approximation may therefore prove useful in developing simple and intuitive expressions for various statistics. For example, we can use this approximation to calculate a simple expression for the distribution of the total number of segregating negatively selected sites in a sample of size \( n \), \( S_n^d \), which as we have seen above is otherwise rather involved. We have

\[ \rho(S_n^d = x) = \sum_{k_1, k_2, \ldots, k_n} h_{k_1} h_{k_2} \ldots h_{k_n}, \] (67)
where the sum is over sets of the $k_i$ that sum to $x$. We find

$$\rho(S_n^d = x) = e^{-nU_d/s} \frac{1}{x!} \left( \frac{nU_d}{s} \right)^x. \quad (68)$$

This is a distribution which is peaked around a mean value of $\frac{nU_d}{s}$, for the same reasons the distribution of $\pi_d$ looks as it does.

We can also calculate the distributions of actual coalescence times and hence the distributions of statistics describing neutral diversity in the foreground selection approximation. Consider the distribution of the real coalescence time between two individuals chosen from classes $k$ and $k'$. In the foreground selection approximation where the coalescence time within the 0-class can be neglected, the actual coalescence time is as given in Eq. (49),

$$\psi(t|k,k') = s(k + k') e^{-s(k+k')t} \left( e^{st} - 1 \right)^{k+k'-1}. \quad (69)$$

Averaging over the values of $k$ and $k'$, we have

$$\psi(t) = \sum_{k=0}^{k'} \sum_{k'=0}^{\infty} H(k,k') \psi(t|k,k'). \quad (70)$$

From this distribution of real coalescence times, we can find the distribution of neutral heterozygosity $\pi_n$ in the usual way,

$$\rho(\pi_n) = \int_0^\infty \frac{[2U_nt]^{\pi_n}}{\pi_n!} e^{-2U_nt} \psi(t) dt. \quad (71)$$

As with $\pi_d$, the shape of this distribution of $\pi_n$ is primarily determined by the shape of $H(k,k')$; the peak in $h_k$ at $k = U_d/s$ leads to the peak in the distribution of real times and hence the peak in the distribution of $\pi_n$. The width of the distribution of $\pi_n$ is somewhat wider, however, since even given individuals coming from fitness classes near the mean, there is a broad distribution of possible real times, and a broad distribution of $\pi_n$ even given a particular real time. However, we see that since individuals at the average fitness class $k = U_d/s$ have on average descended from the 0-class in a time $t \approx \sum_0^{U_d/s} \frac{1}{s} \approx \frac{1}{s} \ln (U_d/s)$, we expect the neutral heterozygosity to have a distribution peaked around an average value

$$\langle \pi_n \rangle \sim \frac{2U_n}{s} \ln \left( \frac{2U_d}{s} \right), \quad (72)$$

valid in the large-$N$ “foreground selection” approximation. Note that the factor of two inside the logarithm arises because one of the two individuals sampled will have taken longer to descend from the 0-class; this individual will have taken a slightly longer than average time.
This average heterozygosity would correspond to an effective population size of

\[ N_e \sim \frac{1}{s} \ln \left( \frac{2U_d}{s} \right), \]  

but as we have seen this effective population size cannot correctly describe the full distribution of \( \pi_n \) nor its relationship to other statistics describing the genetic diversity. For smaller values of \( N \) where the foreground selection approximation breaks down, the average \( \pi_n \) would be somewhat lower than FGS predicts, and its distribution somewhat broader.

Our results for the mean coalescence time as a function of \( U_d/s \) and of \( N \) reflect this intuitive discussion. As is apparent in Fig. 7, as we increase \( N \), the expected coalescence time (and hence \( \langle \pi_n \rangle \)) increases, because as we increase \( N \) it becomes less likely that coalescence occurs in the bulk and more likely it occurs at the forward tail of the fitness distribution. This increase with \( N \) continues until we reach the FGS regime. Above this FGS limit, the mean coalescence times and \( \langle \pi_n \rangle \) become independent of \( N \), as expected. Similarly, we see in Fig. 7 that for large \( N \) the mean coalescence time increases roughly logarithmically with \( U_d/s \), since the time for individuals sampled from middle of the fitness distribution to have descended from the most-fit tail of the distribution increases logarithmically with \( U_d/s \).

**Muller’s Ratchet:**

We have neglected Muller’s ratchet throughout our analysis, and assumed that the fitness distribution \( h_k \) is fixed. Yet Muller’s ratchet will certainly occur, and in some circumstances could have a significant impact on genetic diversity (Gordo et al., 2002; Seger et al., 2010). Thus this is a potentially important omission from our theory. In this section we discuss some of the complications associated with Muller’s ratchet that are important to keep in mind when considering our approach. We discuss the parameter regimes where neglecting Muller’s ratchet should be reasonable, and those where it is likely to cause more serious problems. We provide rough estimates of how large we expect these problems to be, and suggest a few possible ways in which future work might incorporate Muller’s ratchet into our general framework.

Muller’s ratchet causes two related problems within our theoretical framework. First, it causes the values of \( h_k \) to change with time. This changes the distribution of lineage frequencies within each class, and hence changes the coalescence probabilities. After a “click” of the ratchet, the whole distribution \( h_k \) shifts in a complicated way, eventually
reaching a new state where it is shifted left (so the class that was originally at frequency $h_k$ is now at frequency $h_{k-1}$, and so on). In a similarly complex way, the PRF distribution of lineage frequencies in class $k$ shifts from $f_k$ to $f_{k-1}$, and so on. This naturally changes the coalescence probabilities in each class. Fortunately, since the coalescence probabilities in class $k$ are generally very similar to those in classes $k + 1$ or $k - 1$, this effect is unlikely to lead to major inaccuracies provided the ratchet does not click many times within a coalescent time. This is true except when we start considering coalescence in classes close to the 0-class, where the $k$-dependence becomes significant. This can be thought of as the second problem associated with Muller’s ratchet, and is associated with the fact that the ratchet shifts the whole fitness distribution. This effect is easiest to see with an example: imagine we sample two individuals within the $k$-class, and that these individuals did not coalesce before their ancestors were both in the 0-class. At the time (in the past) when these individuals’ ancestors were in the 0-class, this current 0-class might have been the 1-class or 2-class (or higher). Thus these two individuals within the 0-class might not coalesce until, for example, their ancestors were in what is currently the “−2”-class. This clearly means that we might in fact have $\pi_d > 2k$, which our analysis assumes is impossible. In fact, we observe precisely this effect in simulations, and it is the reason why we commonly observe systematic deviations where the simulated values of $\pi_d$ are larger than our theory predicts.

From this discussion it is clear that the key factor in determining whether Muller’s ratchet can reasonably be neglected is how many times the ratchet “clicks” in a coalescence time. We have seen above that an average individual coalesces through the fitness distribution in a time at most of order $\frac{1}{4} \ln (U_d/s)$ generations. Once within the 0-class, coalescence times are of order $Ne^{-U_d/s}$. We must compare these times to the time it takes for the ratchet to “click.” The rate of the ratchet is a complex issue that has been analyzed in by GORDO and CHARLESWORTH (2000a), GORDO and CHARLESWORTH (2000b), and Kim and Stephan (2002) in the regime where $Ne^{-U_d/s} > 1$ and by GESSLER (1995) in the regime where $Ne^{-U_d/s} < 1$. No general analytic expressions exist which are valid across all parameter regimes. However, by definition the ratchet can never move a substantial fraction of the width of the fitness distribution in the coalescence time of just two random individuals. Thus the ratchet is always a small correction to $\pi_d$, and neglecting it is a reasonable first approximation. In practice we find using simulations that the ratchet causes $\pi_d$ to be at
most of order 2 larger than our theoretical predictions, corresponding roughly to a single “click” of the ratchet during a typical coalescence time.

The discussion above suggests a way to incorporate Muller’s ratchet within our theoretical framework, albeit in an ad-hoc way. The ratchet shifts the distribution \( h_k \) underneath the effective coalescent process. The details of this shift are complicated, but on average every click of the ratchet shifts the distribution one step to the left. We can define \( k_{\min} \) to be the number of deleterious mutations (relative to the optimal genotype) in the most-fit individual at any given time. For the case where \( Ne^{-U_d/s} > 1 \), the rest of the distribution will be approximately a Poisson distribution, but with \( h_k \) replaced by \( h_{k-k_{\min}} \). Muller’s ratchet can then be thought of as a process by which \( k_{\min} \) increases over time. This increase is a random process, but has some average rate, leading to an average \( k_{\min}(t) \). As we look backwards in time during the effective coalescent process, the value of \( k_{\min} \) is decreasing due to Muller’s ratchet. This suggests a simple approximation: we replace the actual value of \( k \) with an “effective” value of \( k \) that accounts for the fact that \( k_{\min} \) decreases as we look backwards in time. For each step through the fitness distribution, we imagine that \( k_{\min} \) has decreased by the appropriate amount, and hence the effective value of \( k \) in the new fitness class is decreased by less than 1 compared to the old fitness class. When \( Ne^{-U_d/s} < 1 \) the ratchet is an almost deterministic process, so a similar approximation may prove useful, but in this case the distribution \( h_k \) is on average shifted from the Poisson form (Gessler, 1995).

To incorporate the ratchet into our analysis in this situation, we first must recalculate the relevant coalescence probabilities given the expected average form of \( h_k \), and then carry out the above program. These and other methods to account for Muller’s ratchet remain an interesting topic for future work.

Despite the potential relevance of Muller’s ratchet in practical situations, we note that it does not affect our results in the standard coalescent limit. As is apparent from our general expressions for the coalescence probabilities, the structure of our effective coalescent theory does not depend on all three parameters \( N, U_d, \) and \( s \) independently. Rather, it depends only on the combinations \( NU_d \) and \( Ns \). Thus our theory makes sense in the standard limit where \( NU_d \) and \( Ns \) are held constant while we take \( N \to \infty \). In this limit, Muller’s ratchet does not occur. Whether this means we can neglect the ratchet for large but finite \( N \) depends on the convergence properties of the coalescent limit. This is a difficult limit to explore with
simulations, because it requires large population sizes. However, we have used simulations to verify in a few cases that, as expected, increasing $N$ while keeping $NU_d$ and $Ns$ constant does not change the predicted structure of genealogies but decreases some of the systematic differences between theoretical predictions and the simulations which are suggestive of the effect of the ratchet. Note that while this ratchet-free limit does not change the structure of genealogies in our effective coalescent, the distribution of real coalescent times does change, since all real timescales are proportional to $s$. Thus, as might be expected, we must also take $NU_n$ constant as $N \to \infty$ if we wish neutral diversity to also remain unaffected in this limit.

Note that this ratchet-free limit, while fairly standard in coalescent theory, is somewhat different from the foreground selection approximation discussed. Of course, we can easily imagine a population which is large enough that the foreground selection approximation applies, and then take the standard coalescent limit.

**Conclusion:**

Previous work in population genetics has struggled to understand the patterns of genetic variation in situations where many linked negatively selected sites distort patterns of genetic variation. Our effective coalescent approach addresses precisely this situation, providing a framework in which we can calculate distributions of genealogical structures. We have used this framework to calculate the distributions of a few simple statistics describing sequence variation. It remains for future work to use this effective coalescent approach to compute a wide array of statistics to better understand the details of how purifying selection on many linked sites distorts patterns of genetic variation. The eventual goal will be to use our results to help interpret the increasing amounts of sequence data which seem to point to the importance of negative selection on many linked sites.

**ACKNOWLEDGMENTS**

We thank Daniel Fisher and John Wakeley for many useful discussions, which inspired our effective coalescent approach. We thank Lucy Colwell for her insight into the effects of Muller’s ratchet. AMW thanks the Princeton Center for Theoretical Science at Princeton University, where she was a fellow during the bulk of her work on this paper. LEN is sup-
ported by the Department of Defense through the National Defense Science and Engineering Graduate Fellowship Program, and also acknowledges support from an NSF graduate research fellowship. JBP acknowledges support from the James S. McDonnell Foundation, the Alfred P. Sloan Foundation, the David and Lucille Packard Foundation, the Burroughs Wellcome Fund, Defense Advanced Research Projects Agency (HR0011-05-1-0057), and the US National Institute of Allergy and Infectious Diseases (2U54AI057168).
APPENDIX A: APPROXIMATIONS IN THE COALESCENCE PROBABILITIES

In Eq. (9) we wrote that the probability that two individuals picked from class \( k \) came from the same lineage is

\[ P_{c}^{k,k \rightarrow k} = \int_{0}^{1} \left( \frac{x}{h_{k}} \right)^{2} f_{k}(x) \, dx. \]  

(74)

The idea behind this equation is that \( f_{k}(x) \, dx \) is the probability that there exists a lineage in class \( k \) at frequency \( x \), while \( h_{k} \) is the total frequency of class \( k \), so the probability that an individual in class \( k \) comes from this lineage (given the lineage exists) is \( x/h_{k} \). Similarly the probability two randomly chosen individuals come from this lineage is \( x^{2}/h_{k}^{2} \), and summing up over all possible lineages (times the probability each exists) gives

\[ \int_{0}^{1} x^{2} f_{k}(x)/h_{k}^{2} \, dx. \]

This expression works well in classes where \( 2N h_{k} s k \gg 1 \). In these classes no lineage ever reaches a substantial proportion of \( h_{k} \). However, even when the conditions for our PRF approximation to be valid are met (e.g. \( N s \gg 1 \)), when \( 2N h_{k} s k \lesssim 1 \) a single lineage can sometimes dominate \( h_{k} \). This can be true despite the fact that no lineage can ever reach a frequency of order 1, and hence lineages are independent and \( \int_{0}^{\infty} x f(x) \approx h_{k} \). In this case our PRF method is consistent and the average \( h_{k} \) is indeed correct, but this average \( h_{k} \) consists of some time periods when \( h_{k} \) is smaller than average and other times when there is a large lineage and \( h_{k} \) is larger than average. This means that given that a lineage of frequency \( x \) exists, the expected frequency of class \( k \) is not precisely \( h_{k} \). The most striking consequence of this is that sometimes a lineage will exist at a frequency \( x > h_{k} \), in which case the above expressions will incorrectly predict that the probability two individuals come from the same lineage is larger than 1.

To correct this, we could imagine replacing Eq. (9) with

\[ P_{c}^{k,k \rightarrow k} = \int_{0}^{\infty} \left( \frac{x}{h_{k} + x} \right)^{2} f_{k}(x). \]  

(75)

This expression accounts for the fact that if a large-frequency lineage exists, \( h_{k} \) will tend to be larger than average (by the frequency of that lineage). Since we have \( \int_{0}^{1} \frac{x}{h_{k}} f_{k}(x) = 1 \), and \( \frac{x}{h_{k} + x} < 1 \), Eq. (75) will always be less than 1. Of course, this revised expression is also not exact, because it implies that the average frequency of the class is not precisely \( h_{k} \). The problem is that this expression does not account for the fact that when a high-frequency
lineage is *not* present in the fitness class, the total frequency of the class will be less than $h_k$. Thus while our original expression was an overestimate of the probability that two individuals came from the same lineage, this modified expression is an underestimate.

The same approximations arise in calculating coalescent probabilities more generally. In Eq. (13), we wrote

$$P_{c}^{k,k'} \rightarrow k-\ell = \int dx dy dt_{1} dt_{2} Q_{k,k'}^{k-\ell}(t_{1}, t_{2}) \frac{x f_{k-\ell}(x) y G_{k-\ell}(y \to x, |t_{2} - t_{1}|)}{h_{k-\ell}}.$$  \hspace{1cm} (76)$$

In classes where $2Nh_{k}sk \gg 1$, the fact that a lineage of frequency $x$ exists does not significantly affect the expected size of the fitness class, so this expression is approximately correct. However, for the same reasons discussed above, it can be inaccurate when $2Nh_{k}sk \lesssim 1$, and lead to coalescence probabilities that are larger than 1. We could try to correct the coalescence probabilities as described above by replacing Eq. (13) with

$$P_{c}^{k,k'} \rightarrow k-\ell = \int dx dy dt_{1} dt_{2} Q_{k,k'}^{k-\ell}(t_{1}, t_{2}) \frac{x f_{k-\ell}(x) y G_{k-\ell}(y \to x, |t_{2} - t_{1}|)}{h_{k-\ell} + x}.$$  \hspace{1cm} (77)$$

However, as above, while our original Eq. (13) was an overestimate of the coalescence probability, this revised expression is an underestimate.

These effects associated with a single lineage in class $k-\ell$ reaching a substantial fraction of $h_{k-\ell}$, which become important in fitness classes near the most-fit tail of the distribution where $2Nh_{k-\ell}s(k-\ell) \lesssim 1$, imply that these fitness classes fluctuate in size substantially. They are thus closely related to Muller’s ratchet, and a full analysis of their effects would require a stochastic description of the fitness distribution and a better understanding of how this stochasticity affects the frequency distributions of individual lineages. While this is an interesting topic for future work, it is beyond the scope of the present analysis. Fortunately, by definition this problem only arises in the most-fit tail of the fitness distribution, and does not affect coalescent probabilities in most classes. We thus focus here on simple ad-hoc methods to roughly account for these effects.

The coalescence probabilities in class $k-\ell$ have the general form $I_{x}^{k-\ell} A_{t}^{k,m}$, where $I_{x}^{k-\ell}$ is the probability that two individuals are sampled at the same time from class $k-\ell$ are from the same lineage, and $A_{t}^{k,m}$ is a factor which reflects the decrease in this probability because the two individuals are not sampled at the same time. Since $I_{x}^{k-\ell} = \frac{1}{2Nh_{k-\ell}s(k-\ell)}$, whenever $2Nh_{k-\ell}s(k-\ell) < 1$, we have $I_{x}^{k-\ell} > 1$. This is clearly artificial, and occurs because there is
a substantial probability a lineage has frequency greater than the total average frequency of
the class. In fact, in classes where \(2Nh_{k-\ell}s(k-\ell) < 1\), there is typically only one dominant
lineage in the class at a given time. This means that the coalescence probability between
two individuals sampled from these classes at the same time is approximately 1. We thus
make the simple ad-hoc approximation that \(I_x^{k-\ell} = 1\) in the fitness classes near the most-fit
tail of the distribution where \(2Nh_{k-\ell}s(k-\ell) < 1\).

It remains to consider the second factor, \(A_{k,m}^{k,\ell}\), which is based on the assumption that
the coalescence probability decreases as \(e^{-s(k-\ell)|t_2-t_1|}\) because the lineage the first mutation
came from decreases in frequency relative to the total frequency of the class at this rate.
If the time between the two mutations is short enough that the lineage that was dominant
at \(t_1\) is still dominant at \(t_2\), then this factor should be neglected. This will occur whenever
\(xe^{-s(k-\ell)|t_2-t_1|}\) is also large compared to \(h_{k-\ell}\). Alternatively, if \(xe^{-s(k-\ell)|t_2-t_1|} < h_{k-\ell}\), then
our existing expression for \(A_{k,m}^{k,\ell}\) accurately reflects the decrease in probability that the second
mutation came from the same lineage as the first due to the difference between \(t_2\) and \(t_1\).
Since the distribution of the difference between \(t_2\) and \(t_1\) is dominated by times that are
multiples of \(\frac{1}{s}\), the latter situation is much more common (and when it is not, \(t_2\) and \(t_1\) are by
definition close enough that the exponential factor is of order 1 and hence the contribution
to \(A_{k,m}^{k,\ell}\) is of order 1 anyway).

We thus have a simple prescription which captures the relevant aspects of this effect,
albeit in an ad-hoc way: we simply set \(\frac{1}{2Nh_{k-\ell}s(k-\ell)} = 1\) in all of our coalescent formulas
whenever it would otherwise be greater than 1, leaving all other factors unchanged. All of
our figures and comparisons with simulation reflect this correction to the formulas in the
main text, and we see that our ad-hoc approximation works reasonably well.

We note that for deleterious diversity, the details of this approximation are of limited
importance. Most individuals sampled at random from the population come from near the
center of the fitness distribution, and the deleterious diversity between them is dominated
by their coalescence properties through the bulk of the distribution. Thus these details of
exactly where they coalesce in the high-fitness tail (in those cases where they do not coalesce
in the bulk) can be at most a minor correction to the negatively selected diversity.

However, this effect is more important when considering the distribution of real coales-
cence times and neutral diversity. This is because the real time taken to move between
fitness classes is of order $\frac{1}{\Delta}$, so the classes near the forward tail of the fitness distribution have a disproportionate effect on neutral diversity. While the ad-hoc approach described above gives a reasonable approximation for the neutral diversity, there is an alternative approximation we could make. In the high-fitness tail of the population, those classes where $2Nh_{k-\ell}s(k-\ell) < 1$, the average total size of the population is small compared to the inverse selection strength. This means these classes evolve approximately neutrally. Thus for the purpose of calculating the real coalescence times and the neutral diversity, we can simply approximate the evolution within these classes as a neutral process within a population of size $\sum Nh_j$, where the sum is over all classes in this high-fitness tail. That is, the distribution of real coalescence times (and hence neutral diversity) is as given by our theory through the bulk of the fitness distribution, plus (when coalescence does not happen in the bulk) an exponentially distributed time with mean $\sum Nh_j$ (with the sum over the high-fitness tail where $2Nh_{k-\ell}s(k-\ell) < 1$ only). In practice, we find that both this neutral approximation and the ad-hoc prescription laid out above give similar results.

APPENDIX B: THE FULL CONDITIONAL CALCULATION

In the main text, we focused primarily on the non-conditional approximation to the coalescence probabilities. In this Appendix, we carry out the full conditional calculation for the simplest possible cases. We use this to understand the structure of the conditional results and discuss the validity of the non-conditional approximation.

We begin by considering the full conditional result for the probability that two individuals both sampled from class $k$ coalesce in class $k-2$. In the main text we found that this coalescence probability, $P_{c}^{k,k\rightarrow k-2}$, was given by Eq. (25). In order to evaluate this integral, we need to determine the probability distribution of mutant timings $Q_{c}^{k-2}(t_1,t_2)$. In the main text, we showed that

$$Q_{c}^{k-2}(t_1,t_2) = Q_{c}^{k-1}(t_1,t_2|nc) \ast Q_{1step}^{k-2}(t_1,t_2), \quad (78)$$

where $\ast$ denotes a convolution.

Here $Q_{1step}^{k-2}(t_1,t_2)$ refers to the distribution of timings of the first mutational step from the class we are calculating the probability of coalescence in (in this case, class $k-2$) to the
next most fit class (in this case to class $k - 1$). Since this is the first mutational step, this is always the simple unconditional result

$$Q_{1,\text{step}}^{-1}(t_1, t_2) = sje^{-sj_{t_1}}, sj e^{-sj_{t_2}}. \quad (79)$$

It remains to calculate $Q_{k,k}^{-1}(t_1, t_2|nc)$. We showed in the main text that this conditional probability was given by

$$Q_{k,k}^{-1}(t_1, t_2|nc) = \frac{1}{1 - P_{c}^{k,k\rightarrow k-1}} \left[ Q_{k,k}^{-1}(t_1, t_2) - Q_{k,k}^{-1}(t_1, t_2|c) P_{c}^{k,k\rightarrow k-1} \right]. \quad (80)$$

Here

$$Q_{k,k}^{-1}(t_1, t_2) = (sk)^2 e^{-sk(t_1 + t_2)} \quad (81)$$

is the probability distribution of timings of mutations from class $k - 1$ to class $k$, and $P_{c}^{k,k\rightarrow k-1}$ is given by Eq. (23). Note that while both of these expressions have a simple unconditional form in this case, they will be more complex when we consider larger values of $\ell$. Finally, we saw that

$$Q_{k,k}^{-1}(t_1, t_2|c) = \frac{I_{x}^{-1}}{1 - P_{c}^{k,k\rightarrow k-1}} Q_{k,k}^{-1}(t_1, t_2)e^{-s(k-1)|t_1-t_2|}, \quad (82)$$

where we defined

$$I_{x}^{j} \equiv \frac{1}{2Nh_{j}s_{j}}. \quad (83)$$

Putting these results together, we first find

$$Q_{k,k}^{-1}(t_1, t_2|nc) = \frac{1}{1 - P_{c}^{k,k\rightarrow k-1}} \left[ (sk)^2 e^{-sk(t_1 + t_2)} - I_{x}^{-1} (sk)^2 e^{-2k(t_1 + t_2)} e^{-s(k-1)|t_1-t_2|} \right]. \quad (84)$$

Plugging this into our convolution formula for $Q_{k,k}^{-2}(t_1, t_2)$, we find

$$Q_{k,k}^{-2}(t_1, t_2) = \int_{0}^{t_2} \int_{0}^{t_1} \frac{1}{1 - P_{c}^{k,k\rightarrow k-1}} \left[ (sk)^2 e^{-sk(y+z)} - I_{x}^{-1} (sk)^2 e^{-2k(y+z)} e^{-s(k-1)|y-z|} \right] \times \right.$$

$$\times [s(k - 1)]^2 e^{-s(k-1)(t_1-z+t_2-y)} dx dy \quad (85)$$

$$= \frac{(sk)^2 [s(k - 1)]^2}{1 - P_{c}^{k,k\rightarrow k-1}} e^{-s(k-1)(t_1+t_2)} \int_{0}^{t_2} f_{0}^{t_1} \left[ e^{-sz} e^{-sy} - I_{x}^{-1} e^{-sz} e^{-sy} e^{-s(k-1)|y-z|} \right] d(yz) \quad (86)$$

$$= \frac{(sk)^2 [s(k - 1)]^2}{1 - P_{c}^{k,k\rightarrow k-1}} e^{-s(k-1)(t_1+t_2)} \times \right.$$

$$\times \left[ \frac{1}{s^2} (1 - e^{-st_1}) (1 - e^{-st_2}) - I_{x}^{-1} e^{-sz} e^{-sy} e^{-s(k-1)|y-z|} \right] dy dz \quad (88)$$

In order to evaluate this expression, we need to do the integral

$$B \equiv s^2 \int_{0}^{t_2} \int_{0}^{t_1} e^{-sz} e^{-sy} e^{-s(k-1)|y-z|} dy dz. \quad (90)$$
We begin by considering the case where \( t_1 > t_2 \); in this case we note that

\[
\int_0^{t_2} \int_0^{t_1} dydz = \int_0^{t_2} \int_0^z dydz + \int_0^{t_2} \int_y^{t_1} dydz + \int_0^{t_2} \int_0^{t_2} dydz.
\] (91)

Applying this separation of the integrals, we find

\[
B = s^2 \left[ \frac{1}{s^2(k - 2)} \left( 1 - e^{-2st_2} - \frac{2}{k}(1 - e^{-skt_2}) \right) + \int_{t_2}^{t_1} \int_0^{t_2} e^{-sky} e^{s(k - 2)y} dydz \right]
\] (92)

\[
= \frac{1}{(k - 2)} \left[ 1 - e^{-2st_2} - \frac{2}{k}(1 - e^{-skt_2}) + \frac{1}{k} \left( e^{-skt_2} - e^{-skt_1} \right) \left( e^{s(k - 2)t_2} - 1 \right) \right].
\] (93)

In the alternative case where \( t_2 > t_1 \), we have an identical calculation, but with \( t_1 \) and \( t_2 \) interchanged. This means that the general result is

\[
B = \frac{1}{(k - 2)} \left[ 1 - e^{-2s \min(t_1, t_2)} - \frac{2}{k} \left( 1 - e^{-sk \min(t_1, t_2)} \right) + \frac{1}{k} \left( 1 - e^{-2k|t_1 - t_2|} \right) \left( e^{-2s \min(t_1, t_2)} - e^{-sk \min(t_1, t_2)} \right) \right].
\] (94)

Substituting in our result for \( B \), we find

\[
Q_{k,k}^{k-2}(t_1, t_2) = \frac{k^2 \left[ s(k - 1) \right]^2}{1 - r_c^{k,k \rightarrow k-1}} e^{-s(k - 1)(t_1 + t_2)} \left[ (1 - e^{-st_1}) \left( 1 - e^{-2t_2} \right) - \frac{s(k - 3)}{k - 2} B \right].
\] (95)

We can now use this expression in Eq. (25) to calculate the coalescence probability \( P_{c}^{k,k \rightarrow k-2} \).

Since the result is tedious and does not further illuminate the structure of the full conditional calculation, we do not do so explicitly here, but the integrals are straightforward to evaluate with the methods we have used above.

To motivate the validity of the non-conditional approximation, we need to consider the full calculation going back one additional step. Thus we consider the probability that two individuals both sampled from class \( k \) coalesce in class \( k - 3 \), \( P_{c}^{k,k \rightarrow k-3} \). This will be given by

\[
P_{c}^{k,k \rightarrow k-3} = \int Q_{k,k}^{k-3}(t_1, t_2) \frac{x^2}{r_{k-3}} f_{k-3}(x) e^{-s(k - 3)(t_1 - t_2)} dt_1 dt_2 dx.
\] (96)

where here \( Q_{k,k}^{k-3}(t_1, t_2) \) is the distribution of the time at which the ancestors of the two sampled individuals originally mutated from class \( k - 3 \) to class \( k - 2 \), conditional on them not coalescing in classes \( k - 2 \) or \( k - 1 \).

We can calculate \( Q_{k,k}^{k-3}(t_1, t_2) \) in the same way we calculated \( Q_{k,k}^{k-2}(t_1, t_2) \). Explicitly,

\[
Q_{k,k}^{k-3}(t_1, t_2) = Q_{k,k}^{k-2}(t_1, t_2|nc) \ast Q_{1step}^{k-3}(t_1, t_2),
\] (97)
where analogously to the expression in the previous step

\[ Q_{k,k}^{k-2}(t_1, t_2|nc) = \frac{1}{1 - P_c^{k,k\rightarrow k-2}} \left[ Q_{k,k}^{k-2}(t_1, t_2) - Q_{k,k}^{k-2}(t_1, t_2|c) P_c^{k,k\rightarrow k-2} \right]. \quad (98) \]

We note that \( Q_{k,k}^{k-2}(t_3, t_2) \) is the expression in Eq. (95) we calculated above. As before, we have

\[ Q_{k,k}^{k-2}(t_1, t_2|c) P_c^{k,k\rightarrow k-2} = I_x e^{-s(k-2)|t_1-t_2|}, \quad (99) \]

hence we can write

\[ Q_{k,k}^{k-2}(t_1, t_2|nc) = \frac{Q_{k,k}^{k-2}(t_1, t_2)}{1 - P_c^{k,k\rightarrow k-2}} \left[ 1 - I_x e^{-s(k-2)|t_1-t_2|} \right]. \quad (100) \]

Plugging the above expression back into Eq. 97, we obtain

\[ Q_{k,k}^{k-3}(t_1, t_2) = \frac{s^2(k-1)^2 k^2 s^2(k-2)^2}{(1 - P_c^{k,k\rightarrow k-1})(1 - P_c^{k,k\rightarrow k-2})} e^{-s(k-2)(t_1+t_2)} \int_0^{t_2} \int_0^{t_1} e^{s(k-2)(y+z)} e^{s(k-1)(y+z)} \times \left[ 1 - I_x e^{-s(k-z)|y-z|} \right] \left[ (1 - e^{-sy})(1 - e^{-sz}) - \frac{I_x^{k-1}}{k-2} B \right]. \quad (101) \]

We could evaluate the integrals in the above expression for \( Q_{k,k}^{k-3}(t_1, t_2) \) in the same way that we did in our calculation for \( Q_{k,k}^{k-2}(t_1, t_2) \). We would then substitute this result for \( Q_{k,k}^{k-3}(t_1, t_2) \) into an analogous calculation of \( Q_{k,k}^{k-4}(t_1, t_2) \), and so on. In this way we can build up the full conditional results. The most useful way to go about this is to separate the results into powers of \( I_x \), which is a small parameter related to the coalescent probability in each step. We see from the expression for \( Q_{k,k}^{k-3}(t_1, t_2) \) that there is a term in \( (I_x)^0 \), which is exactly the non-conditional approximation. There are two terms involving \( (I_x)^1 \), and a single term involving \( (I_x)^2 \). In general, in the expression for \( Q_{k,k}^{k-\ell}(t_1, t_2) \), we will have one \( (I_x)^0 \) term (which equals the result in the non-conditional approximation) plus \( \ell \) terms proportional to \( I_x \), \( \binom{2}{\ell} \) terms proportional to \( (I_x)^2 \), and so on. Fortunately, the dependence on the population parameters is entirely contained within these powers of \( I_x \). That is, the coefficients of these various powers of \( I_x \) depend only on \( k \) and \( \ell \), and not at all on the population parameters \( N \), \( s \), and \( U_d \). Thus we could simply calculate a table of coefficients once, and then would be able to understand all the distributions of mutant timings (and from this all the coalescent probabilities).

However, these results rapidly become very complex and unilluminating. Thus rather than carry out the above program, we focus here on understand the general structure of
these results, and on the validity of the non-conditional approximation. We can see that at each step back through the fitness distribution, the probability distribution of times shifts from the non-conditional results by a factor which is roughly proportional to the coalescence probability at that step. That is, in general we have

\[ Q_{k,k}^{k-\ell}(t_1, t_2) = \frac{1}{1 - P_c^{k,k \rightarrow k-\ell}} \left[ Q_{k,k}^{k-\ell}(t_1, t_2) - P_c^{k,k \rightarrow k-\ell} Q_{k,k}^{k-2}(t_1, t_2|c) \right]. \]  

(102)

The first term in square brackets reflects the fact that the probability distribution at a given step conditional on non-coalescence at that step is almost equal to the unconditional probability distribution at that step. The second term represents the correction: note that it is proportional to the coalescence probability in that step, \( P_c^{k,k \rightarrow k-\ell} \). The nature of the correction can be seen by plugging in the distribution of times conditional on coalescence, giving

\[ Q_{k,k}^{k-\ell}(t_1, t_2) = \frac{Q_{k,k}^{k-\ell}(t_1, t_2)}{1 - P_c^{k,k \rightarrow k-\ell}} \left[ 1 - I_x^{k-\ell} e^{-s(k-\ell)|t_1-t_2|} \right]. \]  

(103)

We see that the correction acts to reduce the probability that \(|t_1 - t_2|\) is small — that is, it makes it more likely that \(t_1\) and \(t_2\) are further apart, because this is more likely to be the case given that coalescence did not occur.

Since at each step the shift in the distribution of mutant timings is proportional to the coalescence probability, and the coalescence probability at each step is small, it seems clear that the non-conditional approximation where we simply ignore this shift in mutant timings is reasonable. However there is one potential caveat we must consider: although the shift in the distribution of mutation timings due to conditioning on non-coalescence is small in each step, we typically take many steps before the lineages coalesce. In fact, since the shift in mutation timings is proportional to the coalescence probability, and we typically go back a number of steps of order one over the coalescence probability, in principle the shifts in mutation timings could add up to a substantial shift.

Fortunately, there are three factors which prevent this from happening. First, the shift in mutation timings at each step is always to reduce the probability of times \(t_1\) and \(t_2\) where \(|t_1 - t_2| \lesssim \frac{1}{(k-\ell)s}\). Since at each step \(\ell\) is increasing, and the range of separations between mutation timings at which coalescence can happen is also increasing, the shifts in mutation timings from many steps ago are not a huge factor in determining coalescence probabilities in a particular step. That is, though the shifts in mutation timings add up over many steps,
the shifts most relevant to the coalescent probability in a given step do not. Second, the coalescence probabilities at each step are different. This reduces the chance that we take enough steps to shift the overall mutation timings substantially by the time we coalesce. Finally, and most importantly, we will see that there is a substantial probability that the ancestors of the two individuals sampled do not coalesce until they are in the most-fit class. This means that the total sum of coalescence probabilities (and hence the total possible weight in the shift of mutation timings) remains small even in the worst case where the two lineages do not coalesce for the maximum possible number of steps. The non-conditional approximation will always be good in the regime where this is true.

While this discussion makes clear why we expect the non-conditional approximation to be reasonable for the parameter regimes we consider, it does not constitute a formal proof. Ultimately we rely on simulations to show that the approximation holds, as described in the main text.

APPENDIX C: THE NON-CONDITIONAL DISTRIBUTIONS OF MUTANT TIMINGS

Within the non-conditional approximation we need to calculate the distribution of mutant timings, as used in Eq. (31) and Eq. (36). Specifically, we need to calculate

\[ Q_{k-\ell}^{k}(t) = Q_{k}^{k-1}(t) \ast Q_{k-1}^{k-2}(t) \ast Q_{k-2}^{k-3}(t) \ast \ldots \ast Q_{k-\ell+1}^{k-\ell}(t), \]  

(104)

where \( \ast \) refers to a convolution and

\[ Q_{k-\ell+1}^{k-\ell}(t) = s(k-\ell+1)e^{-s(k-\ell+1)t}, \]  

(105)

as motivated in Eq. (19). To evaluate the convolution we rewrite the one step time distributions in Eq. (105) in Laplace space as

\[ \tilde{Q}_{a}^{a-1}(z) = \frac{sa}{sa + z}. \]  

(106)

Defining \( y = z/s \), the convolution in Eq. (104) now becomes a product

\[ \tilde{Q}_{k-\ell}^{k}(z) = \frac{k!}{(k-\ell)!} \frac{1}{k+y} \frac{1}{k-1+y} \ldots \frac{1}{k-\ell+1+y}. \]  

(107)
\[
\frac{k!}{(k-\ell)!} \left[ C_0^{\ell} \frac{1}{k+y} + C_1^{\ell} \frac{1}{k-1+y} + \ldots + C_{\ell-1}^{\ell} \frac{1}{k-\ell+1+y} \right]
\]

\[
= \frac{k!}{(k-\ell)!} \sum_{j=0}^{\ell-1} C_j^{\ell} \frac{1}{k-j+y},
\]

where we define

\[
C_a^\ell = \prod_{i=0, i \neq a}^{\ell-1} \frac{1}{a-i} = \frac{(-1)^{\ell-1-a}}{a!(l-1-a)!} = \frac{(-1)^{\ell-1-a}}{(\ell-1)!} \binom{\ell-1}{a}.
\]

Converting back to real space, we obtain the distribution of mutant timings

\[
Q_k^{k-\ell}(t) = \frac{sk!}{(k-\ell)!} \sum_{j=0}^{\ell-1} C_j^{\ell} e^{-s(k-j)t}.
\]

We can evaluate this sum to simplify the expression by recognizing the binomial expansion formula

\[
(1 + x)^n = \sum_{i=0}^{n} x^i \binom{n}{i},
\]

where we identify \( x = -e^{st} \). We find

\[
Q_k^{k-\ell}(t) = sk \binom{k}{\ell} e^{-skt} \left( e^{st} - 1 \right)^{\ell-1}.
\]

More generally, Eq. (111) can be written as

\[
Q_a^b(t) = \frac{sa!}{bl!} \sum_{i=0}^{a-b-1} C_i^{a-b} e^{-s(a-i)t},
\]

which can be simplified as

\[
Q_a^b(t) = s(a-b) \binom{a}{b} e^{-sat} \left( e^{st} - 1 \right)^{a-b-1}.
\]

**APPENDIX D: GENERAL COALESCENCE PROBABILITIES IN THE NON-CONDITIONAL APPROXIMATION**

The probability of coalescence for two individuals originally in two different classes \( k \) and \( k' \), as defined in Eq. (36) can be rewritten as

\[
P_{c, k \rightarrow k'}^{k, k'} = \frac{1}{2Nh_{k-\ell}s(k-\ell)} [I_1 + I_2],
\]

\[
I_1 = \int_0^{\infty} Q_{k-\ell}(t_1) e^{-s(k-\ell)t_1} \int_0^{t_1} Q_{k-\ell}(t_2) e^{s(k-\ell)t_2} dt_2 dt_1,
\]

\[
I_2 = \int_0^{\infty} Q_k^{k-\ell}(t_2) e^{-s(k-\ell)t_2} \int_0^{t_2} Q_k^{k-\ell}(t_1) e^{s(k-\ell)t_1} dt_1 dt_2.
\]
Throughout this section we adopt the notation $k' = k + m$. Note that both $I_1$ and $I_2$ involve integrals of the form

$$I_a = \int_0^t Q_a^k(t')e^{st'}dt'.$$  \hspace{1cm} (119)

Plugging in the results for the non-conditional distributions of mutant timings, Eq. (115), and making use of the binomial expansion formula for $(1 + x)^n$ noted in Appendix C, we find this integral becomes

$$I_a = s(a - b)\left(\begin{array}{c} a \\ b \end{array}\right)\int_0^t e^{s(b-a)t'} \left( e^{st'} - 1 \right)^{a-b-1}dt' \hspace{1cm} (120)$$

$$= s(a - b)\left(\begin{array}{c} a \\ b \end{array}\right)\sum_{i=0}^{a-b-1} (-1)^i \binom{a-b-1+i}{i} \int_0^t e^{s(b-a+i) t'} dt' \hspace{1cm} (121)$$

$$= (a - b)\left(\begin{array}{c} a \\ b \end{array}\right) (-1)^{a-b} \sum_{i=0}^{a-b-1} \binom{a-b-1}{i} (a-b-I) (e^{s(b-a+i) t} - 1) \hspace{1cm} (122)$$

$$= (a - b)\left(\begin{array}{c} a \\ b \end{array}\right) (-1)^{a-b} e^{s(b-a) t} \sum_{i=0}^{a-b-1} (-e^{st})^i \binom{a-b}{i} \hspace{1cm} (123)$$

$$= (a - b)\left(\begin{array}{c} a \\ b \end{array}\right) e^{s(b-a) t} (e^{st} - 1)^{a-b}. \hspace{1cm} (125)$$

We now substitute this result for $I_a$ into our expressions for $I_1$ and $I_2$. We note that both have terms of the form

$$I_b = \int_0^\infty Q_b^b(t)\left(\begin{array}{c} c \\ b \end{array}\right)e^{-st} \left( e^{st} - 1 \right)^{c-b} dt. \hspace{1cm} (126)$$

Using similar manipulations to those above, we find

$$I_b = (a - b)\left(\begin{array}{c} a \\ b \end{array}\right)\left(\begin{array}{c} c \\ b \end{array}\right)\int_0^\infty e^{-s(a+c) t} \left( e^{st} - 1 \right)^{a+c-2b-1} dt \hspace{1cm} (127)$$

$$= s(a - b)\left(\begin{array}{c} a \\ b \end{array}\right)\left(\begin{array}{c} c \\ b \end{array}\right) (-1)^{a+c-1} \sum_{i=0}^{a+c-2b-1} \binom{a+c-2b-1}{i} (a+c-2b-1-I) (-1)^i \int_0^\infty e^{-s(a+c-i) t} dt \hspace{1cm} (128)$$

$$= (a - b)\left(\begin{array}{c} a \\ b \end{array}\right)\left(\begin{array}{c} c \\ b \end{array}\right) (-1)^{a+c-1} \sum_{i=0}^{a+c-2b-1} (-1)^i \binom{a+c-2b-1}{i} \frac{1}{a+c-i}. \hspace{1cm} (129)$$

Using the partial fraction decomposition

$$\frac{1}{\binom{n+x}{n}} = \sum_{i=1}^{n} (-1)^{i-1} \binom{n}{i} \frac{i}{x+i}, \hspace{1cm} (130)$$
we find
\[ I_b = \frac{a-b}{a+c-2b} \frac{(a)}{(b)} (\hat{\psi} (-1)^{a+c} = \frac{a-b}{a+c-2b} \frac{(a)}{(b)} (\hat{\psi} (-1)^{2b}) \tag{131} \]

We can now use this result for \( I_b \) to determine \( I_1 \) and \( I_2 \), and hence compute \( P^{k,k'\rightarrow k'-\ell}_c \). We find
\[ P^{k,k'\rightarrow k'-\ell}_c = \frac{1}{2Nh_{k-\ell}s(k-\ell)} \frac{(k')^{k-k'}}{(k-\ell)^{k+k'-\ell}} A_{k,m} \tag{132} \]

As we noted in the main text, this is just
\[ P^{k,k+m\rightarrow k-\ell}_c = \frac{1}{2Nh_{k-\ell}s(k-\ell)} A_{k,m} \tag{133} \]

with \( A_{k,m} \) as defined in Eq. (38). Note that when \( m = 0 \) (i.e. \( k = k' \)) this result simplifies to \( P^{k,k\rightarrow k-\ell}_c \) as defined in the main text, as expected.

**APPENDIX E: THE DISTRIBUTION OF REAL COALESCENCE TIMES.**

The distribution of real coalescence times involves the integral
\[ \psi(t|k,k',\ell) = \left[ \int_0^t R^{k-\ell}_k(t_1,t)dt_1 + \int_0^t R^{k-\ell}_k(t_2,t)dt_2 \right] \right] \ast Q^{k-\ell-1}_k(t), \tag{134} \]
where \( \ast \) refers to a convolution. We begin by considering
\[ \psi_1(t|k,k',\ell) = \left[ \int_0^t R^{k-\ell}_k(t_1,t)dt_1 + \int_0^t R^{k-\ell}_k(t_2,t)dt_2 \right] \tag{135} \]

Substituting in our expressions for \( R^{k-\ell}_k \), we find
\[ \psi_1(t|k,k',\ell) = K \left[ Q^{k-\ell}_k(t)e^{-s(k-\ell)t} \right] \int_0^t Q^{k-\ell}_k(t')e^{s(k-\ell)t'}dt' + Q^{k-\ell}_k(t)e^{-s(k-\ell)t} \right] \int_0^t Q^{k-\ell}_k(t')e^{s(k-\ell)t'}dt' \tag{136} \]

This contains integrals of the form
\[ \int_0^t Q^{b}_a(t')e^{at'dt'}, \tag{137} \]
which we evaluated as \( I_a \) in Appendix D. Using these results, we find
\[ \psi_1(t|k,k',\ell) = K \left[ Q^{k-\ell}_k(t)e^{-s(k-\ell)t} \right] \left( \frac{k}{k-\ell} \right) e^{-st} e^{s(k-\ell)t} \left( e^{st} - 1 \right) ^{\ell} \tag{138} \]
\[ + Q^{k-\ell}_k(t)e^{-2(k-\ell)t} \left( \frac{k'}{k-\ell} \right) e^{s(k-\ell-k')t} \left( e^{st} - 1 \right) ^{\ell+k'-k} \]
Substituting in our result for $Q^b_a(t)$ from Appendix C, we find

$$\psi_1(t|k, k', \ell) = \left[ se^{-s(k+k')t} (e^{st} - 1)^{2\ell + k' - k - 1} \frac{[k + k']!}{[2\ell + k' - k - 1]! [2k - 2\ell]!} \right].$$

(139)

We can now carry out the convolution with $Q^{k-\ell-1}_{k-\ell}(t)$ to find $\psi(t|k, k', \ell)$. We find

$$\psi(t|k, k', \ell) = \frac{s(k - \ell)e^{-s(k - \ell)t}(-1)^{2\ell + k' - k - 1} [k + k']!}{[2\ell + k' - k - 1]! [2k - 2\ell]!} \times \sum_{i=0}^{2\ell + k' - k - 1} (-1)^i \binom{2\ell + k' - k - 1}{i} \frac{1 - e^{-st(k' + \ell - i)}}{k' + \ell - i},$$

which is Eq. (48) in the main text.


WAKELEY, J., 2009 Coalescent Theory, an Introduction. Roberts and Company, Greenwood Village, CO.
FIG. 1 The distribution of the fraction of the population in each fitness class. (a) The distribution of the number of individuals as a function of fitness, where the most beneficial class is arbitrarily defined to have fitness 1, and each deleterious mutation introduces a fitness disadvantage of $s$. Mutations move individuals to less-fit classes, and selection balances this by favoring the classes more fit than average. The shape of the depicted steady state distribution is a result of this mutation–selection balance. The inset (b) shows the processes which lead to this balance within a given fitness class; this is explored in more detail in Desai et al. (2010).
FIG. 2 Each fitness class in the population is composed of many lineages, each of which was created by a single mutation and is (in our infinite-sites model) genetically unique. In Desai et al. (2010) we described the distribution of lineage frequencies within each fitness class. Shown is a schematic cartoon in which each lineage is depicted in a different shade of gray. The arrows denote an example of the effective coalescence process for two individuals sampled from the class second from left. These individuals came from different lineages within that fitness class, and these lineages were created by mutations from different lineages within the next most-fit class (as shown by the arrows). The arrows trace the ancestry of the two individuals back through the different lineages that successively founded each other, until they finally coalesce in the class third from right.
FIG. 3 Examples of the coalescence probabilities $P_{c}^{k,k \rightarrow \ell}$ for two individuals sampled from fitness class $k$ to coalesce in class $\ell$, shown as a function of $\ell$. Here $U_d/s = 10$, and results are shown for $N_s$ ranging from 10 to 100. (a) Results for $k = 9$. (b) Results for $k = 14$. Note the kinks in the graph are due to the issues that arise in classes where $2Nh_{k-\ell}S(k-\ell) < 1$, as described in Appendix A.
FIG. 4 Characteristic examples of the distribution of $\pi_d$. Here $N s = 10$ and in (a) $U_d/s = 6$, while in (b) $U_d/s = 8$. Theoretical predictions are shown as a solid line, simulation results as a dashed line. The fit to simulations is good, but we tend to slightly underestimate the coalescence times, and this tendency is worse for larger $U_d/s$. This is due to Muller’s ratchet, which becomes more problematic as we increase $U_d/s$. This systematic underestimate becomes less severe as $N$ increases, as expected, but comprehensive simulations for much larger $N$ are computationally prohibitive. We can control somewhat for the effects of Muller’s ratchet by neglecting sites that segregate in the most-fit class of individuals in the simulation in computing $\pi_d$, as described in the text. We show the results of simulations neglecting this diversity in the most-fit class as dotted lines. As expected, this somewhat reduces our tendency to underestimate $\pi_d$. 
FIG. 5 Characteristic examples of the distributions of $\pi_n$ and the real coalescent times. (a) Theoretical predictions (solid curve) for the distribution of real coalescence times for $U_d/s = 6$, compared to simulation results (dotted line). (b) Theoretical predictions for the distribution of real coalescence times for $U_d/s = 8$, compared to simulation results. (c) Theoretical predictions for the distribution of $\pi_n$ for $U_d/s = 6$, compared to simulation results. (d) Theoretical predictions for the distribution of real coalescence times for $U_d/s = 8$, compared to simulation results. In all panels we have $Ns = 10$. Our theory agrees well with the simulations, but note that, as with $\pi_d$, we tend to systematically underestimate the coalescence times, and this tendency is worse for larger $U_d/s$. This is due to Muller’s ratchet and the related complications discussed in Appendix A, which (for fixed $Ns$) both become more problematic for larger $U_d/s$. This systematic underestimate becomes less severe as we increase $N$, as expected, but comprehensive simulations for much larger $N$ are computationally prohibitive.
FIG. 6 Theoretical predictions for the mean pairwise heterozygosity at negatively selected sites, $\langle \pi_d \rangle$, as a function of the parameters. (a) $\langle \pi_d \rangle$ as a function of $U_d/s$ for several values of $Ns$. In the “foreground selection” approximation we expect this to be linear with a slope of 2, since on average individuals are sampled from the mean class at $k = U_d/s$ and coalesce in the 0-class, and hence have $\pi_d = 2U_d/s$. We see that as expected this approximation becomes more and more accurate as $Ns$ increases. For smaller $N$, there is substantial probability of coalescence in the bulk of the fitness distribution, which is greater for larger $U_d/s$. Thus the slope of $\langle \pi_d \rangle$ as a function of $U_d/s$ decreases as $Ns$ decreases, and has a downwards curvature. (b) $\langle \pi_d \rangle$ as a function of $Ns$ for several values of $U_d/s$. We see that as $Ns$ becomes large, $\langle \pi_d \rangle$ approaches $2U_d/s$, again consistent with the foreground selection approximation. As $Ns$ decreases, coalescence within the bulk of the fitness distribution becomes more likely, and hence $\langle \pi_d \rangle$ decreases.
FIG. 7 Theoretical predictions for the mean real coalescence time $\langle t \rangle$. All real coalescence times in our analysis scale linearly with $\frac{1}{s}$ (for fixed $N$ and $U_d/s$), so in this figure we fix $s = 10^{-3}$ and show the dependence of the mean pairwise heterozygosity on $N$ and on $U_d/s$. The mean pairwise heterozygosity at neutral sites, $\langle \pi_n \rangle$ is simply $\langle \pi_n \rangle = 2U_n(t)$. (a) Mean coalescence time as a function of $N$ for various values of $U_d/s$. We see that $\langle t \rangle$ increases with $N$ until it approaches a constant value consistent with the foreground selection approximation. (b) Mean coalescence time as a function of $U_d/s$ for several values of $N$. For large $N$, the dependence is roughly logarithmic, consistent with the foreground selection approximation. For smaller $N$, coalescence can occur in the bulk of the fitness distribution, reducing the mean coalescence time.
FIG. 8 The effective coalescence process for three individuals, A, B and C, where A and B coalesced $\tau_3$ steptimes ago and C coalesced with the other two $\tau_2$ steptimes ago. (a) This special case where $k = k' = k''$. (b) An example of the more general case where the three sampled individuals came from three different fitness classes $k'' < k' < k$. 
FIG. 9 The four possible coalescence ordering scenarios relevant to the calculation of the distribution of $\tau_2$ (as part of the calculation of $S_3$). (a) The situation applicable whenever $\tau_3 < k'' - k$. (b-d) The alternative coalescence orderings when $\tau_3 \geq k'' - k$. 