The Structure of Genealogies in the Presence of Purifying Selection: a "Fitness-Class Coalescent"

The Structure of Genealogies in the Presence of Purifying Selection: A "Fitness-Class Coalescent"

Aleksandra M. Walczak[1,*], Lauren E. Nicolaisen[2,*], Joshua B. Plotkin[3], and Michael M. Desai[2]

[1] *CNRS-Laboratoire de Physique Théorique de l'École Normale Supérieure,*
[2] *Department of Organismic and Evolutionary Biology, Department of Physics, and FAS Center for Systems Biology, Harvard University*
[3] *Department of Biology, University of Pennsylvania*
[*] *These authors contributed equally to this work*

(Dated: November 29, 2011)

Abstract

Compared to a neutral model, purifying selection distorts the structure of genealogies and hence alters the patterns of sampled genetic variation. Although these distortions may be common in nature, our understanding of how we expect purifying selection to affect patterns of molecular variation 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 (a "fitness-class coalescent") 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 a description of the allele frequency spectrum in the presence of purifying selection with the structured coalescent approach of Kaplan *et al.* (1988), to trace the ancestry of individuals through the distribution of fitnesses within the population. We also derive our results using a more direct extension of the structured coalescent approach of Hudson and Kaplan (1994). We find that purifying selection leads to patterns of genetic variation that 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

Purifying selection acting simultaneously at many linked sites ("background selection") can substantially alter the patterns of molecular variation at these sites, and at linked neutral sites (Gordo *et al.*, 2002; Hill and Robertson, 1966; Hudson and Kaplan, 1994, 1995; Kaplan *et al.*, 1988; McVean and Charlesworth, 2000; O'Fallon *et al.*, 2010; Seger *et al.*, 2010). In recent years, evidence from sequence data points to the general importance of these selective forces among many linked variants in microbial and viral populations, and on short distance scales in the genomes of sexual organisms (Comeron *et al.*, 2008; Hahn, 2008; Seger *et al.*, 2010). In these situations, existing theory does not fully explain patterns of molecular evolution (Hahn, 2008).

It is difficult to incorporate negative selection at many linked sites into genealogical frameworks such as coalescent theory, because these frameworks typically 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).

Despite this difficulty, a number of productive approaches have been developed to predict how negative selection influences patterns of molecular variation and to infer selection pressures from data. Charlesworth *et al.* (1993) introduced the background selection model and showed that strong purifying selection reduces the effective population size relevant for linked neutral sites (Charlesworth, 1994; Charlesworth *et al.*, 1995). However, weaker selection also distorts patterns of variation, in a way that cannot be completely described by a neutral model with any effective population size (Comeron and Kreitman, 2002; McVean and Charlesworth, 2000), a phenomenon often referred to as Hill-Robertson interference (Hill and Robertson, 1966). Several theoretical frameworks have been developed to analyze this situation. The ancestral selection graph of Neuhauser and Krone (1997) and Krone and Neuhauser (1997) provides an elegant formal solution to the 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. An alternative

approach is based on the structured coalescent, and views the population as subdivided into different fitness classes, tracing the genealogies of individuals as they move between classes. This approach was first introduced by Kaplan *et al.* (1988) and further developed by Hudson and Kaplan (1994, 1995) in the case where fluctuations in the size of each fitness class can be neglected. This structured coalescent approach has been been the basis for computational methods developed by Gordo *et al.* (2002), Seger *et al.* (2010), and Zeng and Charlesworth (2011), and analytical approaches such as those of Barton and Etheridge (2004), Hermisson *et al.* (2002) and O'Fallon *et al.* (2010).

In this paper, we build on the structured coalescent framework by introducing the idea of a "fitness-class coalescent." Rather than considering the coalescence process in real time, 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. We show that the coalescent probabilities in this fitness-class coalescent can be computed using an approach based on the Poisson Random Field method of Sawyer and Hartl (1992), or equivalently can be derived as an extension of the structured coalescent approach of Hudson and Kaplan (1994).

Our fitness-class coalescent theory can be precisely mapped to a coalescence theory in which certain quantities (e.g. coalescence times) have different meanings than in the traditional theory. We can then invert this mapping to determine the structure of genealogies and calculate statistics describing expected patterns of genetic variation. This approach requires certain approximations, but it also has several advantages. Most importantly, we are able to derive relatively simple analytic expressions for coalescent probabilities and distributions of simple statistics such as heterozygosity. Consistent with earlier work, we find that the effects of purifying selection are broadly similar to an effective population size that changes as time recedes into the past. Our analysis makes this intuition precise and quantitative: we can compute the exact form of this time-varying effective population size, as defined by the rate of pairwise coalescence. We also show that this intuition has important limitations: for example, different pairs of individuals have different time-varying effective population size histories, meaning that in principle it is possible to distinguish selection from changing population size. Our approach also makes it possible to calculate the diversity of selected alleles themselves, which may be important when selection is common (Williamson and

Orive, 2002).

We begin in the next section by describing the fitness-class coalescent idea which underlies our approach. We then describe the details of our model and analyze two ways to implement the fitness-class coalescent. The first relies on the Poisson Random Field method of Sawyer and Hartl (1992) to describe the frequency distribution of distinct lineages within each fitness class. We show how this lineage structure can be used to compute coalescence probabilities in each fitness class. The second approach is based on tracing the ancestry of individuals in the order that events occur as described by Hudson and Kaplan (1994), and implemented numerically by Gordo *et al.* (2002). We show how we can sum over all possible ancestral paths to compute equivalent coalescence probabilities in each fitness class. The two approaches provide different and complementary intuitive pictures of the process, and depend on various approximations in somewhat different ways.

After computing coalescence probabilities with both approaches, we show how these probabilities can be used to analyze the structures of genealogies, and we calculate various statistics describing genetic variation in these populations, which we compare to numerical simulations. We then discuss the relationship between our results, neutral theory, and earlier work on selection, and we explore how various approximations limit our approach. The most important of these approximations is that we neglect fluctuations in the size of each fitness class, analogous to earlier work (Hudson and Kaplan, 1994), which restricts our analysis to the case of strong selection (relative to inverse population size). This approximation also means that we neglect Muller's ratchet. We describe this and related approximations and describe their regime of validity in the Discussion. Finally, in the Appendices we explore these approximations in more detail and describe how they inform the relationship between our work and earlier approaches.

THE FITNESS-CLASS COALESCENT

In this section, we outline the main ideas underlying our fitness-class coalescent approach. We begin our analysis by considering the balance between mutations at many linked sites and negative selection against the mutants, which leads to an equilibrium distribution of fitnesses within a population (Haigh, 1978). We illustrate this in Fig. 1, for the case in which all deleterious mutations have the same fitness cost. Each individual is characterized

by the number k of deleterious mutations it contains. Each fitness class k contains many distinct lineages, each of which arose from deleterious mutations in more-fit individuals, as illustrated in Fig. 2. Neutral mutations also occur, but we consider these later.

Hudson and Kaplan (1994) observed that individuals move between fitnesses by deleterious mutations, and that when two individuals are in the same fitness class they could be from the same lineage and hence coalesce. Our fitness-class coalescent exploits this observation to define an effective genealogical process that completely bypasses the ancestral process in real time. Instead, we treat each fitness class as a "generation," and we count time in deleterious mutations: each deleterious mutation moves us from one "generation" to the next. In this way, we can trace the ancestry of individuals through the fitness distribution. For example, there is some probability that two individuals chosen from fitness class k are genetically identical (i.e. come from the same lineage). If not, they each arose from mutations within fitness class $k-1$. If both those mutations occurred in individuals in the same lineage in fitness class $k-1$, we say the two individuals "coalesced" in class $k-1$. If not, they came from different mutations from class $k-2$, and could have coalesced there, and so on. In this way, we can construct a fitness-class coalescent tree describing the relatedness of two individuals, as illustrated in Fig. 2.

In this paper we show that the probability that two randomly chosen individuals who are currently in fitness classes k and k' coalesce in class $k-\ell$, $P_c^{k,k'\to k-\ell}$, is approximately

$$P_c^{k,k'\to k-\ell} = \frac{1}{2n_{k-\ell}s_{k-\ell}} A_\ell^{k,k'}, \tag{1}$$

where n_k is the population size of fitness class k, s_k is an effective selection pressure against these individuals, and

$$A_\ell^{k,k'} = \frac{\binom{k'}{k-\ell}\binom{k}{k-\ell}}{\binom{k+k'}{2\ell+k'-k}}. \tag{2}$$

This coalescent probability is inversely proportional to the population size of the fitness class, $n_{k-\ell}$, and the effective selection coefficient within that class, $s_{k-\ell}$, modified by the combinatoric coefficient $A_\ell^{k,k'}$. As we will see, this has a clear intuitive interpretation. Fitness class $k-\ell$ has size $n_{k-\ell}$, so the coalescence probability per real generation is $\frac{1}{n_{k-\ell}}$. We will see that each lineage spends of order $s_{k-\ell}$ generations in that class, so the total coalescence probability in this class has the form $\frac{1}{n_{k-\ell}}\frac{1}{s_{k-\ell}}$. This is multiplied by $A_\ell^{k,k'}/2$, which we will show describes the probability that the two individuals are in class $k-\ell$ at the same time.

In other words, the probability coalescence occurs in a class equals the inverse population size of the class times the number of generations lineages spend together in that class. In the following sections of this paper we derive Eq. 1 in the two alternative ways mentioned in the Introduction: by explicitly considering the lineage frequency distribution and by following the path summation method of Hudson and Kaplan (1994) and Gordo *et al.* (2002).

Calculating statistics describing sequence variation: Our approach of treating mutation events as timesteps, and computing coalescence probabilities at each timestep, allows us to make a precise mapping to coalescence theory in which certain quantities have a different meaning than in the traditional theory. In this framework, we can calculate a simple analytic expression for the probability two lineages sampled from particular fitness classes will coalesce in any other fitness class. These fitness-class coalescence probabilities allow us to explicitly calculate the structure of genealogies in this "mutation time." We can then compute the distribution of any statistic describing expected sequence variation by averaging over the fitness classes our original individuals come from. For a statistic x that depends on genealogies between two individuals, for example, we write expressions of the form

$$P(x) = \sum H(k,k')\text{Prob}[k,k' \text{ coalesce in } k-\ell]P(x|k,k',\ell), \tag{3}$$

where $H(k,k')$ describes the probability two individuals sampled at random from the population come from classes k and k' respectively.

From the form of these expressions and our simple result for the coalescence probabilities, we can immediately see the main effect of selection on the structure of genealogies. The discussion following Eq. (1) implies that the effect of negative selection is similar to that of an effective population size that changes as time recedes into the distant past — i.e. some $N_e(t)$. This intuition has been suggested by earlier work (see e.g. Seger *et al.* (2010)). As we will see, our analysis describes the precise form of $N_e(t)$: it follows the distribution $n_{k-\ell}$ as ℓ increases further to the past, modified by the coefficient $A_\ell^{k,k'}$. We will also see that this picture of time-varying population size has limits: different pairs of individuals have a different $N_e(t)$. As is clear from Eq. (3), these different histories are averaged according to the distribution $H(k,k')$. While it is the average $N_e(t)$ between pairs that determines the distribution of pairwise statistics, this suggests that statistical power may exist in larger samples to distinguish negative selection from neutral population expansion. We explore

these general conclusions of our analysis in detail in the Discussion.

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 fitness-class coalescent, by contrast, the coalescence times *are* the mutations. To avoid confusion, from here on we will refer to the effective "generations" in our model as "steps," and refer to the fitness-class coalescent "times" as the "steptimes." We will reserve the word "time" to refer to the actual coalescent time, measured in actual generations.

After determining a fitness-class coalescent tree, we can invert our mapping to determine the structure of genealogies in real time. We will do this by calculating how the steptime in our fitness-class 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.

Our fitness-time coalescent requires a number of approximations which limit its applicability. Most importantly, we neglect Muller's ratchet, and more generally ignore the effects of fluctuations in the size of each fitness class. We discuss these approximations in more detail below. We find that within a broad and biologically relevant parameter regime they lead to systematic but small corrections to our results. Despite these limitations, our approach also has several advantages relative to previous work. The fitness-time coalescent approach makes many otherwise difficult analytic calculations tractable, allows us to compute the diversity at the selected sites in addition to linked neutral sites, and may offer a useful basis for practical methods of coalescent simulation and inference.

MODEL

We imagine a finite haploid population of constant size N. Each haploid genome has a large number of sites, which begin in some ancestral state and mutate at a constant rate. Each mutation is either neutral or confers some fitness disadvantage s (where by convention $s > 0$). We assume an infinite-sites framework, so there is negligible probability that two mutations segregate simultaneously at the same site. We assume that there is no epistasis for fitness, and that each deleterious mutation carries fitness cost s, so that the fitness of an

individual with k deleterious mutations is $w_k = (1-s)^k$. Since we assume that $s \ll 1$, we will often approximate w_k by $1 - sk$.

The population dynamics are assumed to follow the diffusion limit of the standard Wright-Fisher model. That is, we assume that deleterious mutations occur at a genome-wide rate U_d per individual per generation (with deleterious mutations assumed to be decoupled from selection). We define $\theta_d/2 \equiv NU_d$, the per-genome scaled deleterious mutation rate. Similarly, neutral mutations occur at a rate U_n per individual per generation, and we analogously define $\theta_n/2 \equiv NU_n$. We assume that each newly arising mutation occurs at a site at 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). In this case, our model is equivalent to that considered by Hudson and Kaplan (1994).

We believe that this is the simplest possible 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. It is essentially equivalent to the model described by Charlesworth *et al.* (1993) and Hudson and Kaplan (1994), which has formed the basis for much of the analysis of background selection (Charlesworth *et al.*, 1993; Gordo *et al.*, 2002; Seger *et al.*, 2010).

Our analysis will develop a fitness-class 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. Since in our model all deleterious mutations have the same fitness cost s, 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).

HAIGH (1978) showed 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,

$$h_k = \frac{e^{-U_d/s}}{k!}\left(\frac{U_d}{s}\right)^k. \tag{4}$$

This means that the average fitness in the population is $1 - U_d$, and that $\bar{k} = \frac{U_d}{s}$.

Throughout our analysis, we will assume that the population exists in this steady state mutation-selection balance. In particular, we neglect the fact that in a finite population there will be fluctuations around this h_k. This approximation is central to our approach, and we make it in subtly different ways in both our lineage-structure and our sum of ancestral paths calculations of the fitness-class coalescence probabilities. It will typically be valid in the bulk of the fitness distribution when selection is strong ($Ns \gg 1$); our analysis is limited to this strong selection case and breaks down when $Ns \lesssim 1$. We discuss this approximation in more detail in the Discussion and in Appendix B. We note that this approximation also implies that we assume that Muller's ratchet can be neglected. We will return to the question of the importance of Muller's ratchet in more detail in the Discussion.

We will later need to understand the distributions of timings, $Q_k^{k-1}(t)$, at which an individual mutates from class $k-1$ to class k. We can calculate this 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

$$\frac{NU_d h_{k-1}}{Nh_k[1 - U_d - s(k - \bar{k})] + NU_d h_{k-1}}. \tag{5}$$

Substituting in the steady state values for the h_k, and noting that these mutation events are a Poisson process, we find

$$Q_k^{k-1}(t) = ske^{-skt}. \tag{6}$$

Note that this calculation is identical to the equivalent distribution of mutation timings computed by GORDO *et al.* (2002) following the approach of HUDSON and KAPLAN (1994).

LINEAGE STRUCTURE AND THE FITNESS-CLASS COALESCENCE PROBABILITIES

In general, the individuals in a particular fitness class k will not be genetically identical. Rather, there will be a number of different lineages within this class, each lineage created by a deleterious mutation from class $k-1$. We now consider the structure of lineage diversity amongst individuals within a given fitness class in the mutation-selection balance. Note that for our purposes here, we only consider deleterious mutations in defining lineages; we consider the diversity at neutral sites separately below.

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 deleterious mutations. In our infinite-alleles model, 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.

New lineages are founded in class k at a rate $\theta_k/2$, where

$$\theta_k = 2Nh_{k-1}U_d. \tag{7}$$

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

$$s_k \equiv -U_d - s(k-\bar{k}). \tag{8}$$

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 define

$$\gamma_k = Ns_k. \tag{9}$$

Using these definitions, we can compute the steady state probability distribution of lineages using the Poisson Random Field model of Sawyer and Hartl (1992). The essential result

is that the number of distinct lineages in class k with a frequency between a and b (in the total population) is Poisson distributed with mean $\int_a^b f_k(x)dx$, where

$$f_k(x) = \frac{\theta_k}{x(1-x)} \frac{1 - e^{-2\gamma_k(1-x)}}{1 - e^{-2\gamma_k}}. \tag{10}$$

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 $\int_0^1 dx x^2 f_k(x)/h_k^2$.

We note that the PRF result involves various implicit approximations, and is valid within a specific parameter regime. Most importantly, we neglect fluctuations in the sizes of each fitness class. This has two main effects. First, it means that we neglect the corresponding fluctuations in the distribution of lineage frequencies $f_k(x)$. Second, it means we are implicitly neglecting the fact that, given a lineage of size x exists in class k, the actual h_k is on average not at its steady state value (e.g. if a high-frequency lineage exists, h_k will tend to be larger). We explain these approximations in detail in Appendix B, and describe an alternative branching process formulation for the lineage structure that corrects for the second effect described above.

The Fitness-class Coalescent Probabilities: We can now calculate the degree of relatedness between two individuals sampled from the population. Our goal is to understand the probability distribution of the fitness-class coalescence steptimes for two individuals chosen at random from the population. We begin by calculating the coalescence probability in each step.

First, imagine that by chance we pick two individuals from the same fitness class k. If the two individuals are from the same lineage, they coalesce within this class. In this case, they are genetically identical and the coalescence steptime is 0. If not, we want to calculate the probability they coalesce in class $k-1$, $P_c^{k,k\to k-1}$. 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

$$P_c^{k,k\to k-1} = \int dt_1 dt_2 Q_{k,k}^{k-1}(t_1,t_2) \frac{x f_{k-1}(x)}{h_{k-1}} \frac{y}{h_{k-1}} G_{k-1}(y \to x, |t_2 - t_1|). \tag{11}$$

Here $Q_{k,k}^{k-1}(t_1, t_2)$ is the joint distribution of t_1 and t_2, x/h_k is the probability one of the individuals came from a lineage of size x given that the lineage exists, $f_k(x)$ is the probability that the lineage exists, and $G_{k-1}(y \to x, |t_2 - t_1|)$ is the probability a lineage in class $k-1$ changes in frequency from x to y in time $|t_2 - t_1|$ (where y could be 0, corresponding to a lineage that has already mutated back to class $k-2$ by the time the second individual mutates to class $k-1$). The forms of Q and G are described in Appendix A.

If the two individuals coalesced in this first step, the coalescent steptime is 1. If not (which occurs with probability $1 - P_c^{k,k\to k-1}$), 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 $k-1$), $P_c^{k,k\to k-2}$, and so on.

So far we have imagined 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

$$H(k,k') = \begin{cases} 2h_k h_{k'} & \text{if } k \neq k' \\ h_k^2 & \text{if } k = k' \end{cases} \tag{12}$$

For convenience we choose $k \leq k'$. We define $P_c^{k,k'\to k-\ell}$ to be the probability that two individuals from classes k and k' coalesce in class $k-\ell$. Note that $P_c^{k,k'\to k-\ell} = 0$ for $\ell < 0$. For $\ell \geq 0$ we have

$$P_c^{k,k'\to k-\ell} = \int dx dy dt_1 dt_2 Q_{k,k'}^{k-\ell}(t_1,t_2) \frac{x f_{k-\ell}(x)}{h_{k-\ell}} \frac{y G_{k-\ell}(y \to x, |t_2 - t_1|)}{h_{k-\ell}}. \tag{13}$$

From the set of coalescence probabilities Eq. (13), we can calculate the probability distribution of coalescence steptimes between two individuals. We describe these steptimes by the distribution of classes in which coalescence occurs; given that we pick two individuals from classes k and k' (with $k < k'$ by convention) the probability that they coalesce in class $k-\ell$ is simply

$$\phi_k^{k'}(\ell) = P_c^{k,k'\to k-\ell} \prod_{j=0}^{\ell-1} \left[1 - P_c^{k,k'\to k-j}\right]. \tag{14}$$

We note that this expression contains an implicit approximation, as described in Appendix A.

Computing 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 by evaluating the integrals in Eq. (13). We explain the details of this calculation in Appendix A. We find

$$P_c^{k,k'\to k-\ell} = \frac{1}{1+2Nh_{k-\ell}s(k-\ell)} A_\ell^{k,k'}, \tag{15}$$

where $A_\ell^{k,k'}$ is a numerical coefficient which depends on k, k', and ℓ but not on the population parameters,

$$A_\ell^{k,k'} = \frac{\binom{k'}{k-\ell}\binom{k}{k-\ell}}{\binom{k+k'}{2\ell+k'-k}}. \tag{16}$$

In Fig. 3 we show examples of these coalescence probabilities for different population parameters. We see that the probability of coalescence decreases with increasing selection coefficients and population size.

Eq. (15) is the complete solution for coalescent probabilities in the non-conditional approximation. 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 approximately proportional to one over the population size of this class times the number of generations individuals spend in this class. The additional 1 in the denominator captures the fact that the individuals might mutate away from the class before coalescing there (which reduces the average time they spend in the class together). The numerical factor multiplying this basic scaling, $A_\ell^{k,k'}$ comes from the integrals over the probability distribution of mutant timings (i.e. the dt_1 and dt_2 integrals). It reflects the probability that the ancestors of the two individuals we are considering were both in class $k-\ell$ at the same time, since they could not otherwise coalesce there.

From this result, we can also form an intuitive picture of the shape of genealogies in the presence of negative selection. We have just seen that the coalescence probability per actual generation depends on the parameters as $\frac{1}{Nh_{k-\ell}}$, where the relevant value of ℓ 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. We return to this point in the section on the structure of genealogies below.

A SUM OF ANCESTRAL PATHS APPROACH

We have just computed the fitness-class coalescence probabilities by considering the lineage structure within each fitness class. KAPLAN *et al.* (1988) proposed a somewhat different way to look at the same problem: they considered a sample of individuals and, without explicitly describing lineage structure, computed the relative probabilities that the next event to occur backwards in time would involve a mutation or coalescent event. For example, if two individuals are in the same fitness class, the next event could be either coalescence within that class or a mutation event. The rates at which these events occur determines their relative probabilities.

In its original form, this approach used diffusion equations to account for fluctuations in the frequencies of each fitness class h_k. BARTON and ETHERIDGE (2004) used this framework to provide a complete solution for the effect of selection at a single site on the structure of genealogies. However, it has not yet proven possible to solve these equations in the more general case of selection at many linked sites. Instead, HUDSON and KAPLAN (1994) made progress by neglecting fluctuations in the frequencies h_k, the same approximation that is central to our approach. Using this approximation, they derived a recursion relation for the mean time to a common ancestor, their Eq. (12). GORDO *et al.* (2002) used this equation as the basis for a coalescent simulation.

Recursion relations of the HUDSON and KAPLAN (1994) form can be solved numerically, and have been used to generate data describing coalescent statistics, but have not yet led to an analytic description of the structure of genealogies. We now demonstrate that these numerical methods are equivalent to our lineage-based formalism above, by showing that the HUDSON and KAPLAN (1994) approach can be used to derive identical analytical formulas for the fitness-class coalescent probabilities. We refer to this as a "sum of ancestral paths" approach, because it relies on summing over all possible paths of individual ancestry through the fitness distribution. The equivalence of this approach to our lineage-structure

calculations means that our analytical results in this paper match earlier numerical and simulation results based on the HUDSON and KAPLAN (1994) formulation.

In order to calculate the coalescence probabilities for a sample of two individuals, we consider the set of all possible ancestral paths these individuals may have followed. Each path is represented by an ordered set of events, backwards in time. These events may either be deleterious mutation events, which move one of the ancestral lineages to the previous fitness class, or coalescence events, which merge the two ancestral lineages. In order for two individuals to coalesce in class $k-\ell$, each ancestral lineage must undergo a series of deleterious mutation events, bringing them from their initial classes to class $k-\ell$. The lineages must then coalesce before any additional deleterious mutations occur. For example, in order for two individuals sampled from class k to coalesce in class $k-1$, the first event, backwards in time, must be a deleterious mutation. This mutation can occur in either individual. After this event, one of the ancestral lineages is still in class k, while the other is in class $k-1$. The second event, backwards in time, must be a deleterious mutation event in the ancestral lineage that remains in class k. Both ancestral lineages are now in class $k-1$. Finally, the third event must be a coalescent event. Note that there are a total of two paths, since either individual may have been the first to mutate.

The probability of any particular ancestral path is the product of the probability of each event in the path. We saw above that deleterious mutations occur in an individual in class k at rate sk. If the two individuals are in different classes, they are not able to coalesce as the next event. Thus the probability of each possible event is simply:

$$P(\text{1st Event is Del. Mut. in } k|k,k') = \frac{sk}{sk+sk'} \tag{17}$$

$$P(\text{1st Event is Del. Mut. in } k'|k,k') = \frac{sk'}{sk+sk'}. \tag{18}$$

If the two individuals are in the same class, the next event may either be a coalescent event or a deleterious mutation. Within each class, coalescence is a neutral process that occurs with rate $1/Nh_k$. Therefore, we have

$$P(\text{1st Event is Coal.}|k,k) = \frac{1/(Nh_k)}{sk+sk+1/(Nh_k)} = \frac{1}{1+2Nh_k sk} \tag{19}$$

$$P(\text{1st Event is Del. Mut.}|k,k) = \frac{2sk}{sk+sk+1/(Nh_k)} = \frac{2Nh_k sk}{1+2Nh_k sk}. \tag{20}$$

These probabilities are analogous to those used by Gordo *et al.* (2002), derived from the framework of Hudson and Kaplan (1994).

Using these probabilities, we can easily calculate the probability of any particular path. In general, in order for two individuals sampled from classes k' and k to coalesce in class $k-\ell$, the ancestral paths must consist of some order of $k'-k+2\ell$ events which include $k'-k+\ell$ deleterious mutation events in the ancestral lineage that began in k', and ℓ deleterious mutation events in the ancestral lineage that began in k. The path must then conclude with a final coalescent event. Note that there are a total of $\binom{k'-k+2\ell}{l}$ possible paths, reflecting the number of ways to order the mutation events in one lineage with those in the other. To calculate the coalescence probability, we sum the probabilities of each path that results in this particular coalescence event.

We can carry out this sum in the general case by dividing up the $\binom{k'-k+2\ell}{l}$ possible paths according to whether or not the ancestral lineages ever coexisted in each class before class $k-\ell$. Each case leads to a different path probability, and these probabilities can be exactly summed. We carry out this calculation in detail in Appendix A. We find that to leading order in $\frac{1}{1+2Nh_{k-\ell}s(k-\ell)}$, we have

$$P_c^{k,k'\to k-\ell} = \frac{1}{1+2Nh_{k-\ell}s(k-\ell)} A_\ell^{k,k'}, \tag{21}$$

which exactly matches our expression for the coalescence probabilities in our PRF approach, Eq. (15).

We note that in deriving this result, we have made the same approximations we used in our lineage structure based approach. Thus the results from the PRF method and the sum of ancestral paths are exactly equivalent in the regime where they are valid. However, there are subtle differences in the results to higher orders of the approximations, which provide useful intuition about the process. For example, in the sum of ancestral paths approach it is more natural to calculate $\phi_k^{k'}(\ell)$ directly, without first calculating $P_c^{k,k'\to k-\ell}$, and doing so allows us to compute certain higher-order corrections to the coalescence probabilities. We discuss these details of the correspondence between the approximations used in the two methods in Supplementary Appendix D.

THE STRUCTURE OF GENEALOGIES AND 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 fitness-class 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 π. This is the only statistic relevant to a sample of two individuals. In larger samples, the coalescent probabilities between any pair of sampled individuals are independent of those between any other pair that does not share the same most recent common ancestor, so the distribution of per-site heterozygosity we expect within such a sample is closely related to the ensemble distribution of π we calculate here.

In our fitness-class 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 π_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 π_n. In analyzing data, we will of course typically not know *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 selected sites, so that the total expected per-site heterozygosity is $\pi = \pi_d + \pi_n$.

Distribution of steptimes and π_d: We begin by imagining that we sample two individuals at random from the *same* fitness class k. If they coalesce in class $k - \ell$, they each acquired ℓ different deleterious mutations to reach class k. Thus the number of negatively selected sites at which they will be polymorphic is twice their coalescent steptime, $\pi_d = 2\ell$.

We therefore have

$$\rho(\pi_d = 2\ell) = \phi_k^k(\ell), \tag{22}$$

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

More generally, if two individuals sampled from classes k and k' coalesce in class $k - \ell$, we have $\pi_d = 2\ell + k' - k$. This means we have

$$\rho(\pi_d = 2\ell + k' - k | k, k') = \phi_k^{k'}(\ell). \tag{23}$$

We can average this over the distributions of k and k' to find the distribution of π_d amongst individuals sampled at random from the population. We find

$$\rho(\pi_d) = \sum_{\ell} \sum_{k=0}^{\infty} H(k, k' = k + \pi_d - 2\ell) \phi_k^{k'=k+\pi_d-2\ell}(\ell), \tag{24}$$

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

These results for the distributions of genealogy lengths and of π_d involve several sums. 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: 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 two individuals sampled from classes k and k' that coalesced in class $k - \ell$, $\psi(t|k, k', \ell)$. We compute this distribution in Supplementary Appendix E, and find

$$\psi(t|k', k, \ell) = \sum_{i=0}^{\pi_d - 1} s\pi_d (-1)^{\pi_d - i - 1} \binom{\pi_d - 1}{i} \binom{k' + k}{\pi_d} \frac{B}{A - B} \left(e^{-sBt} - e^{-sAt} \right), \tag{25}$$

where we have defined $A \equiv k' + k - i$ and $B \equiv 2\,(k-\ell) + \frac{1}{Nsh_{k-\ell}}$.

Note that when $Nh_{k-\ell}s(k-\ell) \gg 1$ (the same condition required to neglect fluctuations in h_k, see Appendix B), this expression can be simplified; we find

$$\psi(t|k',k,\ell) = s(\pi_d+1)e^{-s(k'+k)t}(e^{st}-1)^{\pi_d}\binom{k'+k}{\pi_d+1}. \tag{26}$$

However, it is important to note that while this approximation may be valid in the bulk of the distribution, it will always fail when coalescence occurs in the zero-class, where $s(k-\ell)=0$. In this case, we must use the more complex expression Eq. (25) (or in the case when the coalescence time within the 0-class can be neglected compared to the time taken to descend from the 0-class, the simpler expression described in Eq. (39) below).

Averaging over the possible values of k, k', and ℓ, we find the overall distribution of actual coalescent time between two randomly chosen individuals,

$$\psi(t) = \sum_{k'\geq k}\sum_{k=0}^{\infty}\sum_{\ell=0}^{k}\psi(t|k,k',\ell)\phi_k^{k'}(\ell)H(k,k'), \tag{27}$$

where the distributions $H(k,k')$, $\phi_k^{k'}(\ell)$, and $\psi(t|k,k',\ell)$ are as given above. However, as we will see below, in calculating neutral diversity we will typically find it easier to work directly with $\psi(t|k,k',\ell)$ rather than this unconditional distribution for $\psi(t)$.

The neutral heterozygosity π_n: From the distributions of real times to a common ancestor described above, we can calculate the distribution of π_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, π_n follows a Poisson distribution with mean $U_n t$, where t is drawn from the distribution of coalescence times, Eq. (27). We have

$$\rho(\pi_n) = \int_0^\infty \frac{[2U_n t]^{\pi_n}}{\pi_n!}e^{-2U_n t}\psi(t)dt. \tag{28}$$

In Fig. 5, we compare this distribution of neutral heterozygosity to simulations. We find good general agreement to the shape of the distribution, though there are slight systematic errors (consistent with the 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 model or effective population size approximation would predict; we describe these comparisons further in the Discussion.

We note that an alternative way to compute neutral heterozygosity is to further extend the sum of ancestral paths approach which we used above to provide an alternative derivation of the coalescence probabilities. In this formulation, we do not make any connection to real times. However, this approach provides an alternative way to compute the distribution of neutral heterozygosity, $\rho(\pi_n)$. We carry out this computation in Supplementary Appendix G, and show that it leads to results identical to our analysis above.

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

$$\rho(\pi_n|k,k',\ell) = \int_0^\infty \frac{[2U_n t]^{\pi_n}}{\pi_n!} e^{-2U_n t} \psi(t|k,k'\ell) dt. \tag{29}$$

We can compute this integral; we find

$$\rho(\pi_n|k',k,\ell) = \sum_{i=0}^{\pi_d-1} \pi_d (-1)^{\pi_d-i-1} \binom{\pi_d-1}{i} \binom{k'+k}{\pi_d} \frac{B}{A-B} \left(\frac{(\frac{2U_n}{s})^{\pi_n}}{(\frac{2U_n}{s}+B)^{\pi_n+1}} - \frac{(\frac{2U_n}{s})^{\pi_n}}{(\frac{2U_n}{s}+A)^{\pi_n+1}} \right). \tag{30}$$

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

$$\rho(\pi_n|\pi_d) = \sum_{\pi_d = k'-k+2\ell} \rho(\pi_n|k,k',\ell). \tag{31}$$

This describes the joint distribution of selected and neutral variation, which is of interest in situations where we know in advance which sites are likely to be neutral and which are selected (e.g. when analyzing the joint distribution of synonymous and non-synonymous variation). It implies a particular relationship between the observed diversity at selected sites and the reduction in linked neutral variation.

In many situations, however, we will not know which alleles are selected and which are neutral. In this case, we want to understand the distribution of total heterozygosity π, which is given by

$$\rho(\pi) = \sum_{\pi_n+\pi_d=\pi} \rho(\pi_d)\rho(\pi_n|\pi_d). \tag{32}$$

This is no more difficult to calculate than $\rho(\pi_n)$, since it involves analogous sums. In Fig. 6, we compare this predicted distribution of total heterozygosity to simulations. As with the

other aspects of heterozygosity, we find good general agreement to the simulations, with the slight systematic errors that are consistent with the effects of Muller's ratchet.

The mean pairwise heterozygosity: Above we have calculated the distribution of heterozygosity for both neutral and deleterious mutations, as well as total heterozygosity. It is straightforward to average these results to calculate the mean pairwise heterozygosity for both neutral and deleterious mutations; the mean total pairwise heterozygosity is simply the sum of these. In Fig. 7 and Fig. 8 we show how this mean heterozygosity depends on population size, mutation rate, and selection strength, for neutral and deleterious mutations respectively. We see that the dependence of $\langle \pi_d \rangle$ on the population size is fairly weak. While it increases roughly linearly with N in the weak selection regime, 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. The dependence of $\langle \pi_n \rangle$ on the parameters is also interesting: this depends weakly on the parameters for small N or U_d/s, but for larger N becomes roughly linear. These results make intuitive sense, particularly in light of the "mutation-time" approximation that we introduce in the Discussion, where we discuss these figures in more detail.

Statistics in larger samples: The distributions of π_n and π_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. 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. As we will see, it is unwieldy to calculate closed form expressions for these quantities in our framework, so here we merely lay out a prescription for calculating S_3.

We first consider the distribution of S_3^d, the number of segregating negatively selected sites among three randomly sampled individuals. In order to calculate the probability a sample has a particular S_3^d, we imagine picking three individuals at random from the population and calculate the probability of the coalescence events that lead to that S_3^d. We illustrate

such a situation where three individuals are sampled from classes k, k', and k'' in Fig. 9. Two of these three lineages coalesced in class k_1. We call the steptime at which two of the three lineages coalesced τ_3 (see Fig. 9). We next need to calculate the distribution of τ_2, the total steptime to common ancestry of the three individuals. This time of course cannot be smaller than τ_3. Given values of τ_3 and τ_2, it is clear from Fig. 9 that the total number of segregating negatively selected sites is $S_3^d = 2\tau_2 + \tau_3 - (k'' - k) - (k'' - k')$..

Calculating the joint distribution of τ_2 and τ_3 is tedious, because we must sum over all possible orderings of the coalescence events, but it can be computed using either our lineage structure method or the sum of ancestral paths approach. The basic result is analogous to our results for the coalescence steptime between a pair of individuals: coalescence probabilities within a given class are proportional to the inverse size of that class times the number of real generations the ancestors of given individuals typically spend in that class, times a factor that reflects the time that the ancestors of sampled individuals are present in each class at the same time.

Given a particular value of S_3^d, there is a relationship between the steptimes and actual times (analogous to Eq. (25)), which we could use to find the distribution of the total number of segregating neutral sites S_3^n. More complex statistics involving even larger samples can be computed using similar methods.

However, while this analysis provides a prescription for calculating the distribution of S_3^d and S_3^n, it is clear that the full distributions are opaque. In the Discussion we provide a simple approximation for S_n in a specific parameter regime we refer to as the "mutation-time" regime, but the complexities of the general calculation are tangential to the ideas behind our framework, so we do not pursue them further here. However, these issues will be important to explore in future work aiming to use this framework for data analysis, and our approach here can be used as the basis for genealogical simulations. Further, since our methods allow us to quickly compute the probability of a given genealogical history and to draw a particular genealogy from the appropriate distribution, they may provide a useful basis for importance sampling or MCMC methods to infer selection pressures from data.

NUMERICAL SIMULATIONS OF THE GENETIC DIVERSITY

We compare the predictions of our fitness-class 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 $\frac{1}{s}\ln(U_d/s)$ or N generations (whichever was larger), 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 order to produce the empirical distributions of π_d, and π_n shown in Fig. 4 and Fig. 5, we averaged across at least 300 independent populations for each parameter set.

Our simulations allow for random fluctuations in the frequencies of each fitness class, and for Muller's ratchet. In most of the parameter regimes we explored, the ratchet proceeded during the simulation, so that the least loaded class at the end of each simulation typically contained anywhere from no deleterious mutations (typical for $U_d/s = 2$) to of order ten (typical for $U_d/s = 4$). We see that despite these effects, our theory agrees well with the simulations, although there are small systematic errors that are consistent with effects of the ratchet. Generally speaking these errors increase as we increase U_d/s, but become less

severe for larger N or s. We consider these effects of Muller's ratchet in more detail in the Discussion.

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 (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. 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 work has addressed aspects of this problem, as described in the Introduction. The key insight of our approach is that instead of following the true ancestral process, we develop a *fitness-class* 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 of Sawyer and Hartl (1992), rather than by considering the actual reproductive process within that class. By extending formulas originally computed by Hudson and Kaplan (1994), we showed that these coalescent probabilities can also be computed using a summation of ancestral paths based on the structured coalescent described by Kaplan *et al.* (1988). Hence the conclusions from our analysis also describe the simulations of Gordo *et al.* (2002) and are consistent with all other results based on this structured coalescent approach. Our work is also closely related to recent work in a continuous-fitness model by O'Fallon *et al.* (2010), which uses a similar framework to analyze the weak-selection regime but not the $Ns \gg 1$ situation we study here. We explore the relationship between our analysis and earlier work in more detail in Appendix C.

Our approach leads to simple expressions for the coalescent probability at each step in

our fitness-class 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 a fitness-class 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 fitness-class genealogies and then converting these genealogies into actual coalescence times.

In this paper, we have used this fitness-class approach to calculate simple statistics describing genetic variation, in particular the distribution of pairwise heterozygosity. This leads to analytic expressions for the quantities of interest, although these expressions involve 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.

An Intuitive Picture of the Structure of Genealogies: The most important aspect of our analysis is not the specific results for heterozygosity, which match the conclusions of earlier simulations. Rather, the fitness-class coalescent approach allows us to draw several important general conclusions about how negative selection distorts the structure of genealogies. For two individuals drawn from particular fitness classes, the effect of negative selection is similar to that of an effective population size that changes as time recedes into the past. This is consistent with suggestions from earlier work (e.g. the simulation study of Williamson and Orive (2002) and the work of Seger *et al.* (2010)). However, this is not a population size that decreases in a simple way into the past. Our analysis shows the exact form of this time dependent population size. Further, it is clear from our analy-

sis that this is not the only effect of negative selection on genealogies. There are two key complications. First, the statistics of genetic variation (particularly at the deleterious sites themselves) depend on the structure of genealogies differently in our fitness-class coalescent than in the standard neutral coalescent. Second, the time-varying rate of coalescence between a pair of individuals depends on the fitness classes they were sampled from. In other words, different pairs of individuals have a different time-varying effective population size. This suggests that genetic diversity cannot be represented by a single time-varying effective $N_e(t)$ for the whole population, which means that it may be possible to develop statistical tests to distinguish negative selection from population size. All of these general intuitive conclusions about the structure of genealogies in our fitness-class coalescent are illustrated in Fig. 10.

We now pause to make this intuitive picture of the shape of typical genealogies more precise. In general the probability that two individuals will coalesce within class k has the form $P_c \approx \frac{A}{2}\frac{1}{n_k sk}$, where n_k is the population size of that class, sk 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. We have seen that each lineage spends on average $\frac{1}{sk}$ generations in class k. Thus we can think of each individual as seeing a historical effective population size as shown in Fig. 10c: it starts in some class k with size n_k and spends $\frac{1}{sk}$ generations in that class before moving to class $k-1$, and so on.

If we sample two individuals, however, they will not always be in the same class at the same time. This effect reduces the coalescence probabilities in each class, as captured by the factor $A/2$. This factor is the average fraction of the $\frac{1}{sk}$ generations each lineage spends in class k that the two lineages spend there together. Alternatively, we can think of this factor as consisting of two parts: A is the probability that the two lineages are ever in the same class at the same time, and $\frac{1}{2sk}$ is the average amount of time that they coexist in the class if they coexist at all (they each spend on average $\frac{1}{sk}$ generations there, but on average overlap for only half this time if they overlap at all). While the two lineages are in the class at the same time, the per-generation coalescent probability is $\frac{1}{n_k}$.

This logic implies that genealogies in the presence of purifying selection look like neutral genealogies with a specific type of historical population size dependence. Imagine for example

we picked two individuals from the same fitness class k. They each spend on average $\frac{1}{sk}$ generations in class k, and during that time they have a probability $\frac{A}{2}\frac{1}{n_k}$ per (real) generation of coalescing (this probability includes the fact that on average they are both in the class simultaneously for only a fraction of the mean time each spends there). So roughly speaking, they have an effective population size of $N_e \sim 2n_k/A_{\ell=0}^{k,k}$ for the first $\frac{1}{sk}$ generations. If they fail to coalesce, they then move to class $k-1$, where they spend $\frac{1}{s(k-1)}$ generations and have a probability $\frac{A}{2}\frac{1}{n_{k-1}}$ per generation of coalescing, and hence an effective population size $N_e \sim 2n_{k-1}/A_{\ell=1}^{k,k}$ for this time. If they again fail to coalesce, they move to class $k-2$, and so on.

So far, this picture of a time-dependent population size is rather crude, but we can make it more precise. Specifically, we can write the coalescence probability between two individuals sampled from class k and k' as a function of time in generations as

$$\psi(t|k,k') = \sum_{\ell=0}^{k} \phi_k^{k'}(\ell)\psi(t|k,k',\ell). \tag{33}$$

We can then define the time-dependent effective population size between these individuals, $N_e(t)$, as the inverse probability of coalescence at time t given that coalescence has not yet occurred,

$$\frac{1}{N_e(t)} = \frac{\psi(t|k,k')}{1-\int_0^t \psi(t'|k,k')dt'}. \tag{34}$$

In other words, the $N_e(t)$ is defined as usual as the inverse of the probability that the two individuals will coalesce at time t given that they have not yet done so.

We illustrate this precise time-dependent population size $N_e(t)$ in Fig. 10d. We see that for two individuals sampled from the same fitness class, $N_e(t)$ typically increases into the recent past and then decreases into the more distant past. This reflects the fact that the two individuals are becoming less likely to be in the same fitness class in the recent past, but that as time recedes into the distant past they are likely to be in the highly fit classes which have smaller n_k. For two individuals sampled from classes near but not identical to each other, $N_e(t)$ starts high and then drops before exhibiting a pattern similar to that among individuals sampled from the same class. This reflects the fact that it takes at least a short time before the two individuals have any chance of being in the same class. Finally, for two individuals sampled from more distant classes, $N_e(t)$ simply declines into the past, both

because longer ago they were more likely to be in the same class and more likely to be in the small classes near the high-fitness tail.

Averaging over the whole population, Fig. 10d shows the precise time-dependent population size $N_e(t)$ for two randomly sampled individuals. This average $N_e(t)$ initially stays roughly constant as time recedes into the past before decreasing thereafter. For these two randomly sampled individuals, selection is indistinguishable from this particular historically varying population size. 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. We note that the average $N_e(t)$ shown in Fig. 10d implies that recent branches of genealogies will typically be longer relative to ancient branches than we would expect under neutrality. Thus background selection will lead to an excess of low-frequency variants, and hence lead to negative values of Tajima's D, consistent with expectations from previous work (Charlesworth *et al.*, 1995; Fu, 1997; Gordo *et al.*, 2002).

However, a key difference from a neutral population of time-varying size is that, as is clear in Fig. 10d, 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. If we restrict ourselves to pairwise statistics such as π, the average $N_e(t)$ across pairs of individuals will accurately describe the genetic diversity. However, when we consider appropriately defined statistics in larger samples, the fact that there is no single $N_e(t)$ for the whole population could be important. It remains an interesting question for future work to explore how to exploit this fact to develop statistical tests to distinguish the effects of purifying selection from that of a historically varying effective population size.

Approximations underlying our approach: Our analysis relies on several key approximations. First, both our lineage-structure and our sum of ancestral paths methods assume that we can neglect fluctuations in the total frequency h_k of each class. 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. In Appendix B, we analyze these approximations in detail and show that they will hold in class k whenever $Nh_k sk \gg 1$. In practice, this condition will often break down in the high and low-fitness tails of the fitness distribution. Fortunately, provided it holds in the bulk of the distribution in which most individuals will be sampled (which will typically be true provided $Ns \gg 1$), our approach will still be a good approximation. We have also made several other more technical approximations in computing the fitness-class coalescent probabilities. We discuss these in detail in Supplementary Appendices A and D.

Our final and most important approximation is that we assume that Muller's ratchet can be neglected. The ratchet occurs when h_0 fluctuates to 0, so we can think of this approximation as an extreme aspect of neglecting fluctuations in the sizes of each fitness class. This approximation can sometimes be 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 regardless of whether s is strong or weak compared to U_d. However, when selection is sufficiently strong ($Ns \gg 1$ and $U_d/s < 1$), then an effective population size approximation accurately describes the patterns of genetic variation, as we describe below. Thus our methods are primarily useful for situations where selection is weak compared to mutation rates.

Relationship with an effective population size approximation: Charlesworth *et al.* (1993) considered how selection against many linked deleterious mutations affects linked neutral diversity in a model identical to ours. These authors found that when selection is sufficiently strong, the shape of genealogies and hence the statistics of variation at linked neutral sites is identical to the neutral case, with a reduced effective population size. We refer to this as the effective population size (EPS) approximation.

The idea behind the EPS approximation is that when selection is strong, 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 EPS 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 $N_e = Ne^{-U_d/s}$.

The EPS approximation is valid provided that the neutral coalescence time within the 0-class, t_{neut}, is large compared to the time it takes for a typical individual to have descended from the 0-class, t_{desc}. We know $t_{neut} \sim Ne^{-U_d/s}$, and since a typical individual comes from fitness class $k \sim U_d/s$, we have that $t_{desc} \sim \sum_{j=1}^{U_d/s} \frac{1}{js} \sim \frac{1}{s} \ln\left(\frac{U_d}{s}\right)$. This means that the EPS approximation will be valid provided

$$Nse^{-U_d/s} \gg \ln\left(\frac{U_d}{s}\right). \tag{35}$$

Because of the exponential term on the left hand side of this expression, it is clear that the EPS approximation is a strong-selection, weak-mutation limit. It will tend to be valid provided that $Ns > 1$ and $U_d < s$. However, whenever U_d becomes much larger than s, it will typically break down even in enormous populations, as has been suggested by Nordborg *et al.* (1996) and Kaiser and Charlesworth (2009).

Our analysis describes the effects of background selection beyond the EPS approximation. 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 the EPS approximation, it is not inconsistent with it. However, we have focused primarily on situations where the EPS 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 our approach is most useful.

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 the EPS approximation can act simultaneously, each for different classes of mutations. 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 $Ns_1 \gg 1$ so that the EPS approximation 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 the EPS approximation breaks down for these mutations. In this case, the genetic diversity we expect to see will be characteristic of our fitness-class coalescent theory (with $U_d = U_2$ and $s = s_2$), but with a reduced effective population size $N_e = Ne^{-U_1/s_1}$. In other words, the strongly selected mutations reduce the effective population size because all individuals are very recently descended from an individual that had no large-effect mutations, but the coalescence time through the distribution of weakly selected mutations cannot be neglected.

A "Mutation-time" Approximation: We have seen that our analysis accounts for two effects missing from the EPS approximation: 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 a generalization of the EPS approximation. Rather than considering primarily the diversity generated within the most-fit background, we focus instead on the diversity that accumulates while lineages move between different less-fit backgrounds. Hence we term this approach a "mutation-time approximation" (MTA) for short. In this approximation, we assume that all individuals coalesce within the 0-class, as with the EPS approximation. However, unlike the EPS approximation, we consider the time it took for individuals to descend from the 0-class in addition to the coalescence time within the 0-class. This approximation is valid for large N (when even Nh_1 is enormous compared to $\frac{1}{s}$) so that coalescence always occurs in the 0-class.

In this mutation-time approximation our results become much simpler and provide a useful intuitive picture of the structure of genealogies and genetic variation. Consider the deleterious heterozygosity π_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 $H(k, k')$, the distribution of π_d in the population as a whole is extremely simple: we have

$$\rho(\pi_d) = \sum_{k=\pi_d - k'} H(k, k') = e^{-2U_d/s} \frac{1}{\pi_d!} \left(\frac{2U_d}{s}\right)^{\pi_d}. \tag{36}$$

This simple approximation makes it clear why the distribution of π_d looks the way it does,

and explains how it varies with U_d/s and with N, both in this mutation-time approximation 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 π_d has roughly the same shape as the distribution of fitness within the population. The mean π_d is $2U_d/s$, since the average individual comes from class $k = U_d/s$. Smaller and larger π_d are less likely; the distribution of fitness in the population has variance equal to the mean, so the variance of the distribution of π_d is also roughly equal to its mean. As N gets smaller, there is sometimes coalescence outside of the 0-class. This reduces π_d given k and k'. Hence as we reduce N, the distribution of π_d shifts somewhat leftwards, with a peak somewhat below $2U_d/s$, and has slightly more variance relative to the mean since there is a less definite correspondence between k, k', and π_d. Since π_n is determined by π_d, this also explains why the distribution of π_n has the peaked form we observe, and how it depends on U_d/s and N (note that for π_n the coalescence time within the 0-class, which increases linearly with N, must also be included). All of these intuitive expectations are reflected in our results, as shown in Fig. 4, Fig. 5, Fig. 7, and Fig. 8. Note for example that in Fig. 4, the peak of π_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 mutation-time 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

$$\rho(S_n^d = x) = \sum_{k_1, k_2, \ldots k_n} h_{k_1} h_{k_2} \ldots h_{k_n}, \tag{37}$$

where the sum is over sets of the k_i that sum to x. We find

$$\rho(S_n^d = x) = e^{-nU_d/s} \frac{1}{x!} \left(\frac{nU_d}{s} \right)^x. \tag{38}$$

This is a distribution which is peaked around a mean value of $\frac{nU_d}{s}$, for the same reasons the distribution of π_d looks as it does. We note however that as we increase the sample size n the population size N must be even larger for this MTA approximation to hold.

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

$$\psi(t|k,k') = s(k+k')e^{-s(k+k')t}\left(e^{st}-1\right)^{k+k'-1}. \tag{39}$$

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

$$\psi(t) = 2U_d e^{-st-2(U_d/s)e^{-st}}. \tag{40}$$

The distribution of coalescence times once within the 0-class is $\psi_0(t) = \frac{1}{Nh_0}e^{-t/(Nh_0)}$. From this distribution of real coalescence times, we can find the distribution of neutral heterozygosity π_n in the usual way,

$$\rho(\pi_n) = \int_0^\infty \frac{[2U_n t]^{\pi_n}}{\pi_n!}e^{-2U_n t}\psi(t)dt. \tag{41}$$

We can immediately see that the average coalescence time in this MTA approximation is $t \approx \sum_0^{2U_d/s}\frac{1}{si} + Nh_0 \approx \frac{1}{s}\ln(2U_d/s) + Nh_0$. We therefore expect that the neutral heterozygosity will on average be

$$\langle\pi_n\rangle \sim \frac{2U_n}{s}\ln\left(\frac{2U_d}{s}\right) + 2Nh_0U_n. \tag{42}$$

The first term in this expression comes from the time to descend through the fitness distribution, while the second term comes from the time to coalesce within the 0-class. If this latter term is large compared to the former, the EPS approximation applies. In the opposite case where the time to descend through the distribution dominates, we can see from the MTA approximation that, as with π_d, the shape of this distribution of π_n is primarily determined by the shape of $H(k,k')$. In this case, the peak in h_k at $k = U_d/s$ leads to a peak in the distribution of real times and hence a peak in the distribution of π_n. The width of the distribution of π_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 π_n even given a particular real time.

This average heterozygosity would correspond to an effective population size of

$$N_e \sim \frac{1}{s} \ln\left(\frac{2U_d}{s}\right) + Nh_0, \tag{43}$$

but as we have seen this effective population size cannot correctly describe the full distribution of π_n nor its relationship to other statistics describing the genetic diversity. For smaller values of N where the mutation-time approximation breaks down, the average π_n would be somewhat lower than the MTA predicts, and its distribution somewhat broader.

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 (Charlesworth and Charlesworth, 1997; 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 several related problems within our theoretical framework. First, it causes the values of h_k to change with time, and means they may not always follow a Poisson distribution. 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 an additional 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 π_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 $\frac{1}{s}\ln(U_d/s)$ 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 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, provided the ratchet does not typically move a substantial fraction of the width of the fitness distribution in the coalescence time of two random individuals, it will be a small correction to π_d, and neglecting it is a reasonable first approximation. In practice we find in our simulations that for the parameter regimes we consider, π_d is at most of order 2 larger than our theoretical predictions, which would correspond roughly to the effect of 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 fitness-class 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 fitness-class 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 fitness-class 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 fitness-class 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 mutation-time approximation we discussed above. Of course, we can easily imagine a population which is large enough that the mutation-time approximation applies, and *then* take the standard coalescent limit.

Conclusion: Our fitness-class coalescent approach provides a framework in which we can compute distributions of genealogical structures in situations where many linked negatively selected sites distort patterns of genetic variation. 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 fitness-class 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 fitness-class coalescent approach. MMD acknowledges support from the James S. McDonnell Foundation and the Harvard Milton Fund. AMW thanks the Princeton Center for Theoretical Science at Princeton University, where she was a fellow during some of her work on this paper. LEN is supported 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). Many of the computations in this paper were run on the Odyssey cluster supported by the FAS Sciences Division Research Computing Group at Harvard University.

APPENDIX A: THE FITNESS-CLASS COALESCENT PROBABILITIES

PRF Lineage-Structure Approach: In the main text, we used our PRF lineage-structure approach to write an integral expression for the probability $P_c^{k,k'\to k-\ell}$ that two individuals sampled from fitness classes k and k' coalesce in class $k-\ell$, Eq. (13) above. In this Appendix, we evaluate this integral to calculate the coalescent probabilities.

Eq. (13) depends 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 appears in the context of an integral

$$I_G=\int yG_{k-\ell}(y\to x,|t_2-t_1|)dy, \tag{44}$$

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

$$I_G=xe^{-s(k-\ell)|t_2-t_1|}. \tag{45}$$

Note this deterministic solution simply reflects the exponential decline in frequency of a rare deleterious allele. Substituting Eq. (45) into Eq. (13), we find

$$P_c^{k,k'\to k-\ell}=\int dxdt_1dt_2Q_{k,k'}^{k-\ell}(t_1,t_2)\frac{x^2f_{k-\ell}(x)}{h_{k-\ell}^2}e^{-s(k-\ell)|t_2-t_1|}. \tag{46}$$

The x integral can be evaluated using standard asymptotic methods; we find

$$\int_0^1 dxx^2f_{k-\ell}(x)\equiv I_x^{k-\ell}=\frac{1}{1+2Nh_{k-\ell}s(k-\ell)}. \tag{47}$$

Note that this and all further expressions for $I_x^{k-\ell}$ incorporate the branching process correction for fluctuations in h_k described in Appendix B. Plugging in this result, we find

$$P_c^{k,k'\to k-\ell}=I_x^{k-\ell}\int dt_1dt_2Q_{k,k'}^{k-\ell}(t_1,t_2)e^{-s(k-\ell)|t_2-t_1|}. \tag{48}$$

To make further progress, we must understand $Q_{k,k'}^{k-\ell}(t_1,t_2)$, the joint distribution of the times at which individuals sampled from fitness classes k and k' originally mutated from class $k-\ell$ to class $k-\ell+1$. In general, t_1 and t_2 are not independent, since in order for the two lineages to have coalesced in class $k-\ell$ they must not have coalesced in any

earlier classes, which makes them less likely to have been in those classes at the same time. In Supplementary Appendix A, we analyze these distortions and their effects on the coalescence probabilities. 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 a particular class does not shift the distribution of mutation timings much. We therefore neglect the complications associated with the probability distributions of the mutant timings conditional on non-coalescence. We refer to this as the non-conditional approximation, and discuss its validity further in Supplementary Appendix A.

In the non-conditional approximation, the times t_1 and t_2 are independent, $Q_{k,k'}^{k-\ell}(t_1,t_2) = Q_k^{k-\ell}(t_1)Q_k^{k-\ell}(t_2)$. We calculate these distributions of mutant timings $Q_k^{k-\ell}(t)$ in Supplementary Appendix B. Plugging these in, and evaluating the integrals as described in Supplementary Appendix C, we find

$$\int dt_1 dt_2 Q_{k,k'}^{k-\ell}(t_1,t_2)e^{-s(k-\ell)|t_2-t_1|} = \frac{\binom{k'}{k-\ell}\binom{k}{k-\ell}}{\binom{k+k'}{2\ell+k'-k}} \equiv A_\ell^{k,k'}. \tag{49}$$

Plugging this result into Eq. (48), we find $P_c^{k,k'\to k-\ell} = I_x^{k-\ell}A_\ell^{k,k'}$, the result quoted in the main text. We note that $e^{-s(k-\ell)|t_2-t_1|}$ is the probability the ancestor of the first individual to mutate into class $k-\ell$ is still there when the ancestor of the second individual mutated into that class. Thus $A_\ell^{k,k'}$ is the probability that the ancestors of the two individuals were in class $k-\ell$ at the same time, while $I_x^{k-\ell}$ is the probability that they coalesce if so, as described in the main text.

Sum of ancestral paths approach: In the main text, we considered the probability of any particular ancestral path in the history of a sample of two individuals. In this section, we sum over the probabilities of all possible ancestral paths to compute the fitness-class coalescence probabilities. First, we consider sampling two individuals from the same fitness class k. In order for these two individuals to coalesce in class k, the first event must be a coalescent event. Using the event probabilities computed in the main text, we find $P_c^{k,k\to k} = I_x^k$, equivalent to our earlier lineage-based result. In order for these individuals to coalesce in class $k-1$, the first event must be a deleterious mutation event. Since both individuals' ancestral lineages are currently in class k, the probability the first event is a deleterious mutation event is $1-I_x^k$. After this event, there is now one ancestral lineage in class $k-1$, and one in class k. The next event must be a deleterious mutation in the latter,

which occurs with probability $\frac{k}{2k-1}$. Finally, the third event must be a coalescent event. This implies

$$\phi_k^k(1) = (1 - I_x^k) I_x^{k-1} \frac{k}{2k-1}. \tag{50}$$

Note that this logic has given us an expression for the probability that the coalescent steptime is 1, $\phi_k^k(1)$, and not the probability of coalescence in this class given that coalescence has not yet occurred, $P_c^{k,k\to k-\ell}$, because we have already included the probability that the coalescence event does not happen in class ℓ.

We can continue to extend this logic to subsequent fitness classes. For example, for coalescence to occur in class $k-2$, there are six possible paths. We can label them as AABBc, BBAAc, ABABc, ABBAc, BABAc, and BAABc, where A corresponds to a mutation in the first individuals' ancestral lineage, B corresponds to a mutation in the second individuals' ancestral lineage, and c corresponds to a coalescent event. We can calculate the probability of each path. For example,

$$P(AABBc) = \left(\frac{1 - I_x^k}{2}\right)\left(\frac{k-1}{2k-1}\right)\left(\frac{k}{2k-2}\right)\left(\frac{k-1}{2k-3}\right) I_x^{k-2}. \tag{51}$$

The probability of path BBAAc is identical, since it has the same probabilities at each step. However, the remaining four paths have a different probability, because the ancestral lineages exist together in the $k-1$ class at the same time. This distorts the probability of mutations at that step, since coalescence could also have occurred. For paths of this type, we have

$$P(ABABc) = \left(\frac{1 - I_x^k}{2}\right)\left(\frac{k}{2k-1}\right)\left(\frac{1 - I_x^{k-1}}{2}\right)\left(\frac{k-1}{2k-3}\right) I_x^{k-2}. \tag{52}$$

We add up each path to find

$$\phi_k^k(2) = I_x^{k-2} \frac{k(k-1)}{4(2k-1)(2k-3)} \left(2\left(1 - I_x^k\right) + 4\left(1 - I_x^k\right)\left(1 - I_x^{k-1}\right)\right) \tag{53}$$

$$= I_x^{k-2} \frac{3k(k-1)}{2(2k-1)(2k-3)} \left(1 - I_x^k - \frac{2}{3} I_x^{k-1} + \frac{2}{3} I_x^k I_x^{k-1}\right). \tag{54}$$

It is informative to consider the form of this result. The I_x^{k-2} factor is the probability that the two ancestral lineages coalesce in class $k-2$, given that they existed in class $k-2$ at the same time. The remaining factors represent the probability that the two ancestral lineages existed at the same time in class $k-2$. This consists of a leading order term $\frac{k(k-1)}{4(2k-1)(2k-3)}$ (identical to our earlier result for $A_{\ell=2}^k$), multiplied by a correction due to the distortion in paths from the possibility of coalescence in previous steps.

We can continue on to consider the probability of coalescence in class $k-3$. There are now a total of $\binom{6}{3}$ possible paths. These can be split into four types, depending upon whether the two ancestral lineages coexisted in both classes $k-1$ and $k-2$ (e.g. ABABABc), in class $k-1$ only (e.g. ABAABBc), in class $k-2$ only (e.g. AABBABc), or in neither (e.g. AAABBBc). The probability of each type of path is identical, except for a distortion factor $(1-I_x^{k-i})$ for each class $k-i$ in which the two ancestral lineages were together at the same time. The probabilities can be calculated as before, and summed to yield $\phi_k^k(3)$. Using similar logic, we can extend this approach to the situation where two individuals are sampled from different classes, k' and k.

In Supplementary Appendix D, we describe the details of carrying out this summation over all possible paths to determine the coalescent probabilities. We find

$$\phi_k^{k'}(\ell) = I_x^{k-\ell}\frac{\binom{k'}{k-\ell}\binom{k}{k-\ell}}{\binom{k'+k}{k'-k+2\ell}}\left[1-\sum_{i=0}^{\ell-1}\frac{\binom{k'-k+2i}{i}\binom{2\ell-2i}{\ell-i}}{\binom{k'-k+2\ell}{\ell}}I_x^{k-i}+\right. \tag{55}$$

$$\left.\sum_{i=0}^{\ell-2}\sum_{j>i}^{\ell-1}\frac{\binom{k'-k+2i}{i}\binom{2j-2i}{j-i}\binom{2\ell-2j}{\ell-j}}{\binom{k'-k+2\ell}{\ell}}I_x^{k-i}I_x^{k-j}-\ldots\right], \tag{56}$$

where as always we have assumed $k \leq k'$ by convention. The form of this solution is intuitive. The factor $I_x^{k-\ell}$ is the probability of coalescence in class $k-\ell$, given that the two ancestral lineages existed in this class at the same time. The remaining factors reflect the probability that the two lineages are together in class $k-\ell$ at some point. This consists of a leading order term, which is identical to the $A_\ell^{k,k'}$ calculated previously, times a correction. The correction represents the distortion in the paths due to the possibility that coalescence could have occurred at previous steps. There are a total of $l+1$ terms in the correction, each of which is known and calculable.

Provided that $2Nh_k sk \gg 1$, we can neglect the higher-order terms in Eq. (56). This is equivalent to calculating the probability of coalescence in a given class, without considering the possibility that coalescence events could have occurred in previous classes. Thus it converts our expression for $\phi_k^{k'}(\ell)$ into an expression for $P_c^{k,k'\to k-\ell}$. Neglecting these terms also implicitly makes the non-conditional approximation, as we did in the PRF method, because it assumes that the fact that coalescence did not occur in previous classes does not distort the likelihood of taking particular paths. Making this approximation, we find

$$P_c^{k,k'\to k-\ell} = \frac{1}{1+2Nh_{k-\ell}s(k-\ell)}A_\ell^{k,k'}, \tag{57}$$

which exactly matches our expression for the coalescence probabilities in the non-conditional approximation in our PRF approach, Eq. (15).

The condition $2Nh_k sk \gg 1$ is the condition we are already assuming in treating the frequencies of each class, h_k as constant (see Appendix B). Thus the results from the PRF method and the sum of ancestral paths are exactly equivalent in the regime where they are valid. We discuss the correspondence between approximations in the sum of ancestral paths method as compared to the PRF method in more detail in Supplementary Appendix D.

APPENDIX B: FLUCTUATIONS IN H_K

Throughout our analysis, we have neglected fluctuations in the frequencies of each frequency class h_k. This approximation was necessary to write our PRF expressions for lineage structure, $f_k(x)$, which depend on h_k. Similarly, it was necessary for us to compute the probabilities of each possible ancestral event in our sum of ancestral paths method. In this Appendix, we examine this approximation in detail and analyze its regime of validity.

Fluctuations in the fitness class frequencies affect the coalescence probability within class k in three different ways. First, fluctuations in h_{k-1} affect the rate at which mutations enter class k. When h_{k-1} is larger than average, more mutations occur. Within the PRF method, this means that there will be more small lineages than the steady state $f_k(x)$ accounts for, which reduces the coalescence probability. In the sum of ancestral paths method, this means that the probability of mutation events increases relative to the probability of coalescence events, which similarly reduces the coalescence probability. When h_{k-1} is smaller than average, less mutations occur, and the reverse is true.

Second, fluctuations in h_k affect the coalescence rates within this class. Consider the case where h_k is larger than average. Within the PRF method, this means that the probability that two individuals randomly sampled from class k come from a given lineage of size x is less than our assumption of $\frac{x^2}{h_k^2}$. This reduces the coalescence probability. In the sum of ancestral paths method, this means that the probability of coalescence events decreases relative to mutation events, which similarly reduces the coalescence probability. As before, when h_k is smaller than average, the reverse is true.

The third effect of fluctuations is specific to the PRF method, in which we assumed that the probability two individuals in class k come from a lineage of frequency x (given that

the lineage exists) is $\frac{x^2}{h_k^2}$. This implicitly assumes that the fact that there exists a lineage of frequency x in fitness class k does not affect the expected frequency of the class h_k. This is not strictly true: given that there exists a lineage at high frequency, it is likely that h_k is larger than average, and vice versa. In other words, there is a correlation between the size of a lineage and the frequency of the class, so the probability that two individuals picked from a class come from the a lineage of frequency x is not precisely $\frac{x^2}{h_k^2}$. When x is large, this expression overestimates the probability two individuals are from the same lineage, since given that those high-frequency lineages exist, h_k will be larger than average. Similarly (though less dramatically), when x is small our expression underestimates the probability two individuals are from the same lineage.

Note that this third effect of fluctuations is distinct from the second effect above. The second effect describes fluctuations in h_k that are uncorrelated to the frequency of a particular lineage. It thus applies to both the PRF and sum of ancestral paths methods; it reflects the general fact that when h_k is larger coalescence is less likely. The third effect, on the other hand, reflects the fact that if we assume we sample an individual from a lineage of size x, this biases the value of h_k. Since our sum of ancestral paths method never makes any references to lineages, this third effect of fluctuations only applies to the PRF method.

These three effects all depend on the size of the fluctuations relative to the average size of the each fitness class. Thus neglecting fluctuations will be a good approximation provided that the fluctuations in h_k are small compared to h_k. To determine when this will hold, we note that each lineage in class k can reach, at most, a maximum size of order $\frac{1}{s_k}$ individuals (selection prevents any individual lineage from becoming more common than this). The total number of individuals in the class is on average Nh_k. This means that, provided that $Nh_k \gg \frac{1}{s_k}$, each fitness class is made up of many individual lineages. Thus we would expect that the fluctuations in the sizes of each one would tend to cancel, and the overall fluctuations in h_k should be negligible provided that this condition holds.

To make this intuition more precise, we must calculate the variance in h_k and compare it to h_k. In principle this information is contained in our PRF expressions, but it is much simpler to compute using a continuous-time branching process method. That is, rather than use a diffusion approximation to describe the dynamics of each lineage, we use a continuous-time branching process. As before, we imagine that new lineages in class k are created at

a rate $\theta_k/2$. In steady state there will be some time-independent probability that there are n total individuals across all the lineages in the class, $P(n)$. Note that on average we must have $n/N = h_k$, and that $P(n)$ contains information on the fluctuations in the h_k. We first compute the generating function for $P(n)$,

$$H(z) \equiv \sum_{n=0}^{\infty} P(n) z^n. \tag{58}$$

To do so, we start by computing the generating function for the probability distribution of the number of individuals from each lineage, as described by Eqs. (7-9) of Desai and Fisher (2007). We substitute this expression into Eq. (24) of Desai and Fisher (2007) and integrate. We find

$$H(z) \equiv \sum_{n=0}^{\infty} P(n,t) z^n \equiv \langle z^n \rangle = \left[\frac{s}{1 - z(1-s)} \right]^{\frac{\theta}{2(1-s)}}, \tag{59}$$

where angle brackets denote expectation values, and we have suppressed the k subscripts. Note that this calculation is based on a continuous-time branching process, in which individuals have a different distribution of offspring number than in a Wright-Fisher process, leading to a transient distribution of the frequencies of individual lineages that is half as large as in the Wright-Fisher model for lineages of substantial frequency. Thus to make comparisons with the Wright-Fisher process, we have to take $\theta \to 2\theta$ (as we would in comparing Wright-Fisher to Moran models), as described by Desai and Fisher (2007).

Eq. (59) describes the fluctuations in the size of an individual fitness class: the mean, variance, and higher moments of n can be easily computed by taking derivatives of $H(z)$. Thus we can immediately compute $Var(h_k)/h_k$ using standard generating function methods. We find that in fact the fluctuations in h_k are indeed negligible provided that

$$N h_k s_k \gg 1. \tag{60}$$

In practice, this condition will often break down in the high and low-fitness tails of the fitness distribution. Fortunately, provided it holds in the bulk of the distribution in which most individuals will be sampled, which will typically be true provided $Ns \gg 1$, our approach will still be a good approximation.

Correcting for correlations between the size of a lineage and the frequency of the fitness class: All three effects of fluctuations in h_k described above are negligible

in the same parameter regime, $Nh_k sk \gg 1$. However, the fact that the third effect applies only to our PRF result obscures the precise relationship between our two approaches, and the relationship to earlier work. Further, relaxing this approximation provides a useful comparison of the subtle differences between the assumptions underlying the approaches. Thus we describe here an alternative approach to understanding the lineage structure in a fitness class which allows us to account for these correlations between the size of a lineage, x, and the frequency of the fitness class, h_k.

We first note that, in his original calculation of the neutral ESF, Ewens (1972) used a diffusion result, $f(x)$, roughly analogous to our PRF expression to describe the probability that there exists a lineage with frequency x in the population at a given time. However, Ewens' $f(x)$ was derived as the solution to the diffusion approximation to the K-allele Wright-Fisher process, in the limit of infinite alleles. This process explicitly imposes the constraint that the sum of all lineages in the population at a given time must add to 1. This means that there is no correlation between the size of a lineage and the total number of individuals in the population.

The PRF calculation of the lineage structure does not involve this explicit constraint. This is what makes it possible to compute a simple analytical expression for $f_k(x)$. This lack of constraint means that the PRF result admits fluctuations in h_k, which lead to corresponding correlations between x and h_k. We could partially avoid this by defining $\gamma_k = Nh_k s_k$, rather than Nh_k, as we have so far. This would effectively mean that each lineage is assumed to be diffusing between 0 and h_k rather than between 0 and 1, and forbid any lineage from reaching a frequency larger than h_k. Thus it reduces the discrepancies associated with the correlations between x and h_k. However, even with this redefinition, there is no constraint that the lineages in a given class all add to precisely h_k, and so correlations still exist.

To correct exactly for the effects of correlations between x and h_k, we extend the continuous-time branching process model introduced above. We now imagine that there are B sites in the genome, each of which can mutate to create a new lineage in class k. In the large-B limit, each distinct lineage in class k arose from a mutation at a different site in the genome (and we will later make the infinite-sites assumption $B \to \infty$, which makes this exactly true). The rate at which new mutations found lineages in class k due to mutations at a specific one of these B sites is $\frac{\theta_k}{2B}$. This means that, analogous to Eq. (59),

the generating function for the probability that there are n mutations at a particular site i in class k is

$$H_i(z) = \left[\frac{s}{1-z(1-s)}\right]^{\frac{\theta}{B(1-s)}}, \tag{61}$$

where again we have suppressed the k subscripts and we have taken $\theta \to 2\theta$ to match to the Wright-Fisher model as described above.

If we define $n_{i,k}$ to be the total number of mutants at site i in class k, we have that

$$\sigma_k \equiv \sum_{i=1}^{B} n_{i,k} \tag{62}$$

is the total number of individuals in the class (note that on average we expect $\sigma_k = Nh_k$). We now imagine that we sample some number m individuals from class k. The probability that they are all from the same lineage is

$$J_m^{(k)} = \left\langle \sum_{i=1}^{B} \frac{n_{i,k}^m}{\sigma_k^m} \right\rangle = \left\langle \frac{n_{1,k}^m}{(n_{1,k}+\ldots n_{1,B})^m} + \frac{n_{2,k}^m}{(n_{1,k}+\ldots n_{1,B})^m} + ... + \frac{n_{B,k}^m}{(n_{1,k}+..n_{1,B})^m} \right\rangle. \tag{63}$$

Note this has the same form as our PRF expression, except we are averaging over $\frac{n_i^m}{\sigma^m}$ rather than averaging over n_i^m and *then* dividing by the average σ^m. In other words, we are explicitly accounting for the correlations between x and h_k.

We can rewrite Eq. (63) using the identity

$$\frac{1}{\sigma_k^m} = \int_0^\infty \frac{x^{m-1}}{(m-1)!} e^{-x\sigma_k} dx. \tag{64}$$

This identity can easily be verified by integrating the RHS by parts. Using this, and noting that lineages at each of the B sites are independent, we find

$$\begin{aligned} J_m^{(k)} &= \left\langle \sum_{i=1}^{B} n_i^m \int_0^\infty \frac{x^{m-1}}{(m-1)!} e^{-x\sigma_k} dx \right\rangle \\ &= B \int_0^\infty \frac{x^{m-1}}{(m-1)!} \langle n_1^m e^{-x\sigma_k} \rangle dx \\ &= B \int_0^\infty \frac{x^{m-1}}{(m-1)!} \langle e^{-xn_i} \rangle^{B-1} \langle n_1^m e^{-xn_1} \rangle dx. \end{aligned} \tag{65}$$

The first expectation value inside the integral can be computed by noting that

$$\langle e^{-xn_i} \rangle = H(z = 1-x) = \left[1 + x\frac{1-s}{s}\right]^{\frac{\theta}{B(1-s)}}. \tag{66}$$

Differentiating this result m times with respect to x results in an expression for $\langle n_1^m e^{-xn_1} \rangle$. Plugging these results in and integrating, taking the limit $B \to \infty$, and neglecting higher order terms in s, we find

$$J_m^{(k)} = \theta \sum_{j=0}^{m-1} (-1)^j \binom{m-1}{j} \frac{1}{\theta + j} = \frac{(m-1)!}{\prod_{j=1}^{m-1}(\theta + j)} = \frac{1}{\binom{\theta+m-1}{\theta}}. \tag{67}$$

If we were to use the original PRF result to calculate the probability two individuals sampled simultaneously from class k are from the same lineage, we would find $\int_0^1 \left(\frac{x}{h_k}\right)^2 f_k(x)dx = \frac{1}{\theta}$. Using our branching process result for $J_2^{(k)}$, we see that correcting the PRF result for the third effect of fluctuations in h_k yields the modified probability $\frac{1}{1+\theta_k}$. As expected, the branching process result precisely matches the sum of ancestral paths approach, which is also unaffected by this third effect of fluctuations in the h_k. All of the formulae quoted in the main text and shown in the figures incorporate this correction, which appropriately handles the correlations between the frequency of an individual lineage and the size of the fitness class.

APPENDIX C: RELATION TO PREVIOUS WORK

In this Appendix we compare our analysis to related work, and summarize the key approximations that we and others have used. We have presented two main approaches to calculating coalescence probabilities in this paper. The first approach is based on the lineage structure within each fitness class, described using a PRF-based method. The second approach involves summing over all possible ancestral paths, based on the structured coalescent framework introduced by Kaplan *et al.* (1988) and Hudson and Kaplan (1994, 1995). We show in this paper that both approaches involve closely related approximations and yield equivalent expressions for the coalescence probabilities.

Historically, attempts to describe the coalescent process in the presence of selection go back to the structured coalescent introduced by Kaplan *et al.* (1988). These authors considered a sample of individuals from given fitness classes and computed the relative probabilities that the next event to occur backwards in time would involve a mutation or coalescent event, without explicitly describing lineage structure. In their original work, Kaplan *et al.* (1988) used a full stochastic description of the frequencies of each fitness class, in which one keeps track of the probability distribution of these frequencies to account

	this work	Hudson & Kaplan 88	Hudson & Kaplan 94,95	Gordo et al 02	Charlesworth et al 93	Barton & Etheridge 04	Seger et al.10	O'Fallon et al. 10
analytical expressions for genealogy structure	x				x	x		x
accounts for frequency class fluctuations (valid for $Ns \sim 1$)		x				x	x	x*
valid for $Nse^{-U/s} << \ln[U/s]$	x	x	x	x		x	x	x
valid for $Ns \gg 1$	x	x	x	x	x	x	x	
valid for many classes	x	x	x	x	x		x	x
accounts for Muller's ratchet		x				x†	x	
discrete fitness classes	x	x	x	x	x	x	x	

TABLE I A summary of related approaches to the coalescence process in the presence of purifying selection. *Addresses $Ns \sim 1$ situation, but assumes deterministic fitness distribution. †Within a two-class framework.

for selection. They derived diffusion equations for the transition probabilities between states. This approach is very general, but as a result is complex and requires numerical evaluation. Barton and Etheridge (2004) developed this diffusion approach to compute the effect of selection on genealogies in a system in which selection acts only on a single locus.

Hudson and Kaplan (1994) later simplified their original structured coalescent approach to describe the case where fluctuations in the frequencies of fitness classes can be neglected. In this deterministic approximation, they showed that one can compute very simple expressions for the relative probabilities of the next event to occur backwards in time in the history of a sample. In this manner, Hudson and Kaplan (1994) were able to generate a simple recursion relation for the mean time to a common ancestor, their Eq. (12). Gordo *et al.* (2002) used this equation as the basis for a coalescent simulation, and Zeng and Charlesworth (2011) recently extended this method to describe the joint effects of recombination and background selection.

Recursion relations of the Hudson and Kaplan (1994) form can be solved numerically, and have been used to generate data describing coalescent statistics, but have not yet led to an analytic description of the structure of genealogies in the presence of negative selection at many linked sites. In this paper we have shown that one can sum over ancestral paths within this framework, to derive analytical formulas for the coalescence probabilities which are equivalent to those computed from our lineage-based formalism. This equivalence means

that our analytical results in this paper match earlier numerical and simulation results based on the HUDSON and KAPLAN (1994) formulation. However, like the HUDSON and KAPLAN (1994) framework, neither of our approaches in this paper account for fluctuations in the frequencies of fitness classes.

In reality, the frequency of each fitness class will fluctuate due to genetic drift. As we have described in Appendix B, these fluctuations are substantial in classes whose deterministic size is small compared to the inverse of the effective selection pressure against individuals in that class, $Nh_k sk < 1$. This leads to important effects on the structure of genealogies if most fitness classes through the bulk of the fitness distribution fluctuate substantially. This will occur whenever $Ns \lesssim 1$, so fluctuations must therefore be taken into account for small Ns. While the diffusion approach of KAPLAN *et al.* (1988) in principle provides a complete solution to this problem for all values of Ns, this formalism and the related results of BARTON and ETHERIDGE (2004) are computationally strenuous. There remains a need for further work on accurate but more analytically tractable approaches which are able to account for the frequency fluctuations.

We note that the work of O'FALLON *et al.* (2010) and of HERMISSON *et al.* (2002) introduced analytical approaches valid for the case of $Ns \sim 1$, although these methods are not based on a model related to the ideas of KAPLAN *et al.* (1988). We also note that the problem of fluctuating fitness class sizes has been considered in the case of other problems (for example, forward selection (COOP and GRIFFITHS, 2004)), but a detailed discussion is outside the scope of this work.

Neglecting the fluctuations in fitness class frequencies is in principle reasonable when $Ns \gg 1$. However, we note that even when $Ns \gg 1$, the sizes of the smallest fitness classes near the tails of the distribution may still fluctuate substantially. Muller's ratchet is one aspect of this general effect. Recently SEGER *et al.* (2010) extended the simulation scheme of GORDO *et al.* (2002) to address this problem by first doing a forward-time simulation, recording the fluctuations in the classes (including Muller's ratchet) from this simulation, and then putting these fluctuations into a backwards simulation by hand. Our methods do not account for these effects. They are therefore less general than the work of SEGER *et al.* (2010), and break down due to fluctuation effects more quickly as Ns decreases. On the other hand, our analysis does not rely on forward simulations and is able to compute simple

analytic expressions for coalescence probabilities.

We also note that although we consider the large Ns approximation, our approach has a broader range of applicability than the effective population size approximation, which assumes that the coalescence time is dominated by the time to coalescence within the most-fit class. For the EPS approximation to be valid requires that this latter time ($\sim Ne^{-U_d/s}$) is small compared to the time average individuals took to descend from the most-fit class ($\sim \frac{1}{s}\ln Ns$). Thus for the EPS approximation to hold, we require $Ne^{-U_d/s} \gg \frac{1}{s}\ln[U_d/s]$, not just $Ns \gg 1$. Thus we can easily have $Ns \gg 1$, yet $Nse^{-U_d/s} \ll \ln[U_d/s]$, in which case the EPS approximation breaks down and yet our approach is still valid.

LITERATURE CITED

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

Charlesworth, B., 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. Genetical Research **63**: 213–227.

Charlesworth, B. and D. Charlesworth, 1997 Rapid fixation of deleterious alleles can be caused by muller's ratchet. Genetical Research **70**: 63–73.

Charlesworth, B., M. T. Morgan, and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134**: 1289–1303.

Charlesworth, D., B. Charlesworth, and M. T. Morgan, 1995 The pattern of neutral molecular variation under the background selection model. Genetics **141**: 1619–1632.

Comeron, J. M. and M. Kreitman, 2002 Population, evolutionary and genomic consequences of interference selection. Genetics **161**: 389–410.

Comeron, J. M., A. Williford, and R. M. Kliman, 2008 The hill-robertson effect: Evolutionary consequences of weak selection and linkage in finite populations. Heredity **100**: 19–31.

Coop, G. and R. C. Griffiths, 2004 Ancestral inference on gene trees under selection. Theroetical Population Biology **66**: 219–232.

Desai, M. M. and D. S. Fisher, 2007 Beneficial mutation-selection balance and the effect of linkage on positive selection. Genetics **176**: 1759–1798.

Ewens, W. J., 1972 The sampling theory of selectively neutral alleles. Theoretical Population Biology **3**: 87–112.

Ewens, W. J., 2004 *Mathematical Population Genetics: I. Theoretical Introduction*. Springer, New York, NY.

Fu, Y., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147**: 915–925.

Gessler, D. D. G., 1995 The constraints of finite size in asexual populations and the rate of the ratchet. Genetical Research **66**: 241–253.

Gordo, I. and B. Charlesworth, 2000a The degeneration of asexual haploid populations and the speed of muller's ratchet. Genetics **154**: 1379–1387.

Gordo, I. and B. Charlesworth, 2000b On the speed of muller's ratchet. Genetics **156**: 2137–2140.

Gordo, I., A. Navarro, and B. Charlesworth, 2002 Muller's ratchet and the pattern of variation at a neutral locus. Genetics **161**: 835–848.

Hahn, M. W., 2008 Toward a selection theory of molecular evolution. Evolution **62**: 255–265.

Haigh, J., 1978 The accumulation of deleterious genes in a population-muller's ratchet. Theoretical Population Biology **14**: 251–267.

Hermisson, J., O. Redner, H. Wagner, and E. Baake, 2002 Mutation-selection balance: Ancestry, load, and maximum principle. Theoretical Population Biology **62**: 9–46.

Hill, W. and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genetical Research **8**: 269–294.

Hudson, R. and N. Kaplan, 1994 Gene trees with background selection. In *Non-neutral evolution: Theories and molecular data*, edited by B. Golding, pp. 140–153, Chapman and Hall, New York.

Hudson, R. and N. Kaplan, 1995 Deleterious background selection with recombination. Genetics

141: 1605–1617.

Kaiser, V. B. and B. Charlesworth, 2009 The effects of deleterious mutations on evolution in non-recombining genomes. Trends in Genetics **25**: 9–12.

Kaplan, N., T. Darden, and R. Hudson, 1988 The coalescent process in models with selection. Genetics **120**: 819–829.

Kim, Y. and W. Stephan, 2002 Recent applications of diffusion theory to population genetics. In *Modern Developments in Theoretical Population Genetics: The Legacy of Gustave Malecot*, edited by M. Slatkin and M. Veuille, Oxford University Press, Oxford, UK.

Kimura, M., 1955 Stochastic processes and distribution of gene frequencies under natural selection. Cold Spring Harbor Symposia on Quantitative Biology **20**: 33–53.

Kingman, J. F. C., 1982 The coalescent. Stochastic Processes and their Applications **13**: 235–248.

Krone, S. M. and C. Neuhauser, 1997 Ancestral processes with selection. Theoretical Population Biology **51**: 210–237.

McVean, G. A. T. and B. Charlesworth, 2000 The effects of hill-robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics **155**: 929–944.

Neuhauser, C. and S. M. Krone, 1997 The genealogy of samples in models with selection. Genetics **145**: 519–534.

Nordborg, M., B. Charlesworth, and D. Charlesworth, 1996 The effect of recombination on background selection. Genetical Research **67**: 159–174.

O'Fallon, B. D., J. Seger, and F. R. Adler, 2010 A continuous-state coalescent and the impact of weak selection on the structure of gene genealogies. Mol Biol Evol **27**: 1162–1172.

Przeworski, M., B. Charlesworth, and J. Wall, 1999 Genealogies and weak purifying selection. Mol Biol Evol **16**: 246–252.

Sawyer, S. A. and D. L. Hartl, 1992 Population genetics of polymorphism and divergence. Genetics **132**: 1161–1176.

Seger, J., W. A. Smith, J. J. Perry, J. Hunn, Z. A. Kaliszewska, L. L. Sala, L. Pozzi, V. J. Rowntree, and F. R. Adler, 2010 Gene genealogies strongly distorted by weakly interfering mutations in constant environments. Genetics **184**: 529–545.

Tavare, S., 2004 Ancestral inference in population genetics. In *Lectures on Probability Theory and Statistics*, edited by J. Picard, volume 1837, pp. 1–188, Springer, Berlin.

Wakeley, J., 2009 *Coalescent Theory, an Introduction*. Roberts and Company, Greenwood Village, CO.

Williamson, S. and M. E. Orive, 2002 The genealogy of a sequence subject to purifying selection at multiple sites. Molecular Biology and Evolution **19**: 1376–1384.

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

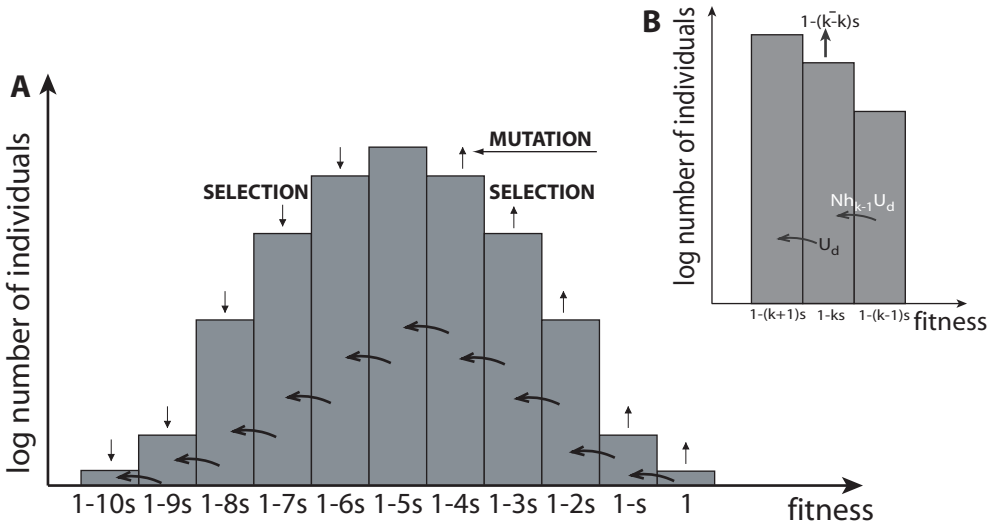

FIG. 1 The distribution of the fraction of the population in each fitness class. **(a)** The distribution of the number of individuals as a function of fitness, where the most beneficial class is arbitrarily defined to have fitness 1, and each deleterious mutation introduces a fitness disadvantage of 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.

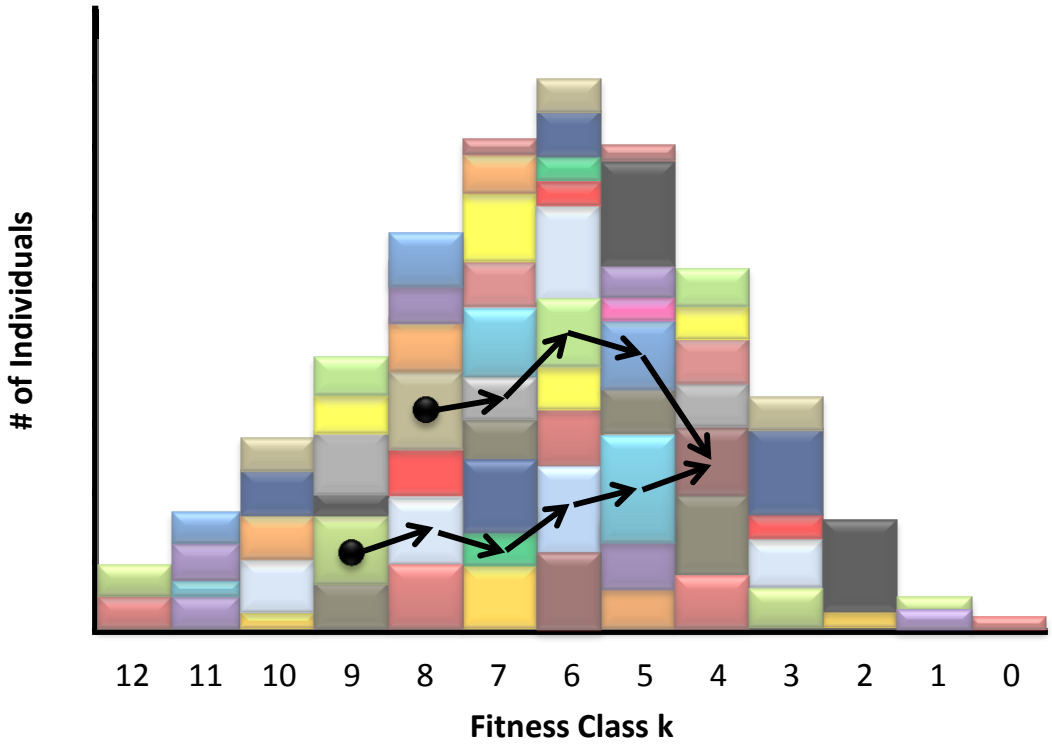

FIG. 2 Each fitness class in the population is composed of many lineages, each of which was created by a single mutation and is (in our infinite-sites model) genetically unique. Shown is a schematic cartoon in which each lineage is depicted in a different color. The arrows denote an example of the fitness-class coalescence process for two individuals sampled from classes 8 and 9. These individuals came from different lineages, 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.

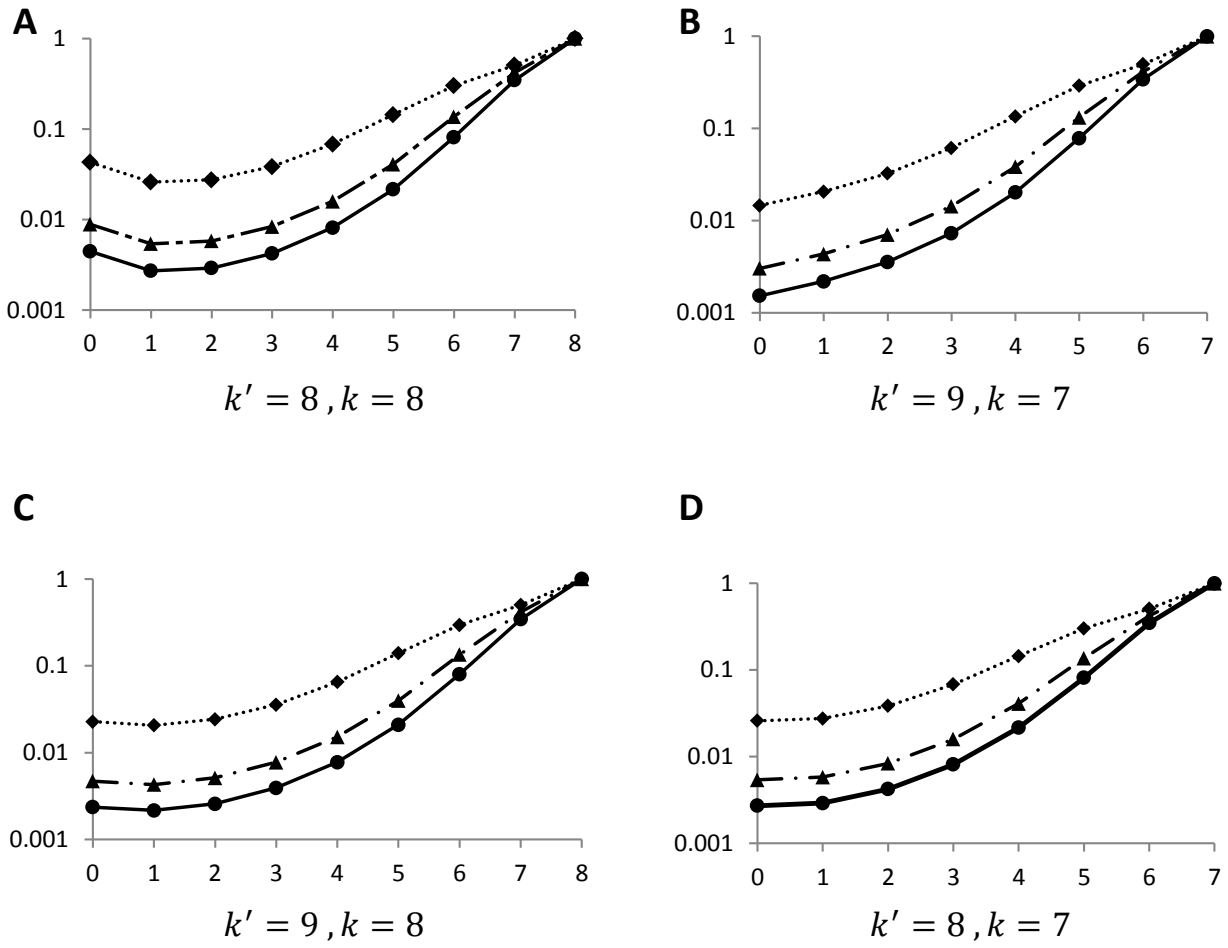

FIG. 3 Examples of the coalescence probabilities $P_c^{k,k'\to k-\ell}$ for two individuals sampled from fitness classes k and k' to coalesce in class $k-\ell$, shown as a function of ℓ. Here $U_d/s = 8$, $s = 10^{-3}$, and results are shown for $Ns = 10$ (dotted lines), $Ns = 50$ (dashed lines), and $Ns = 100$ (solid lines).

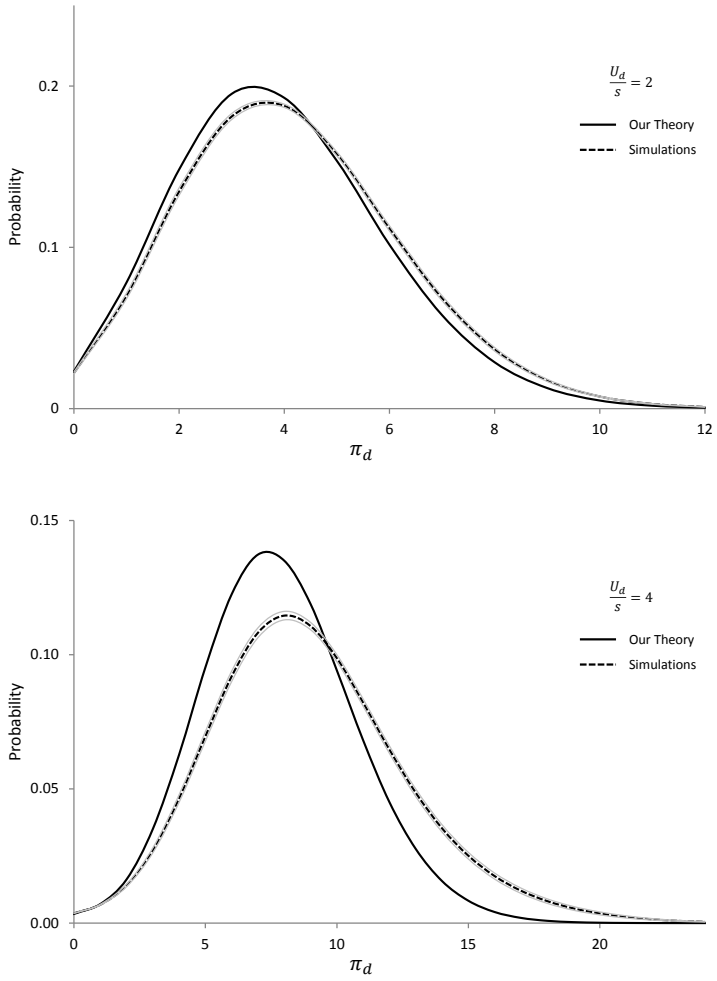

FIG. 4 Characteristic examples of the distribution of π_d. Here $N = 5 \times 10^4$, $s = 10^{-3}$ and in **(a)** $U_d/s = 2$, while in **(b)** $U_d/s = 4$. Theoretical predictions are shown as a solid line, simulation results as a dashed line. Simulation results are averaged across at least 300 independent simulations for each parameter set; shaded regions show one standard error in the simulation results. The fit to simulations is good, but we tend to slightly underestimate π_d, and this tendency is worse for larger U_d/s. This is consistent with the effects of Muller's ratchet, which becomes more problematic as we increase U_d/s. This systematic underestimate becomes less severe (for all values of U_d/s) as N increases, as expected, but comprehensive simulations for much larger N are computationally prohibitive.

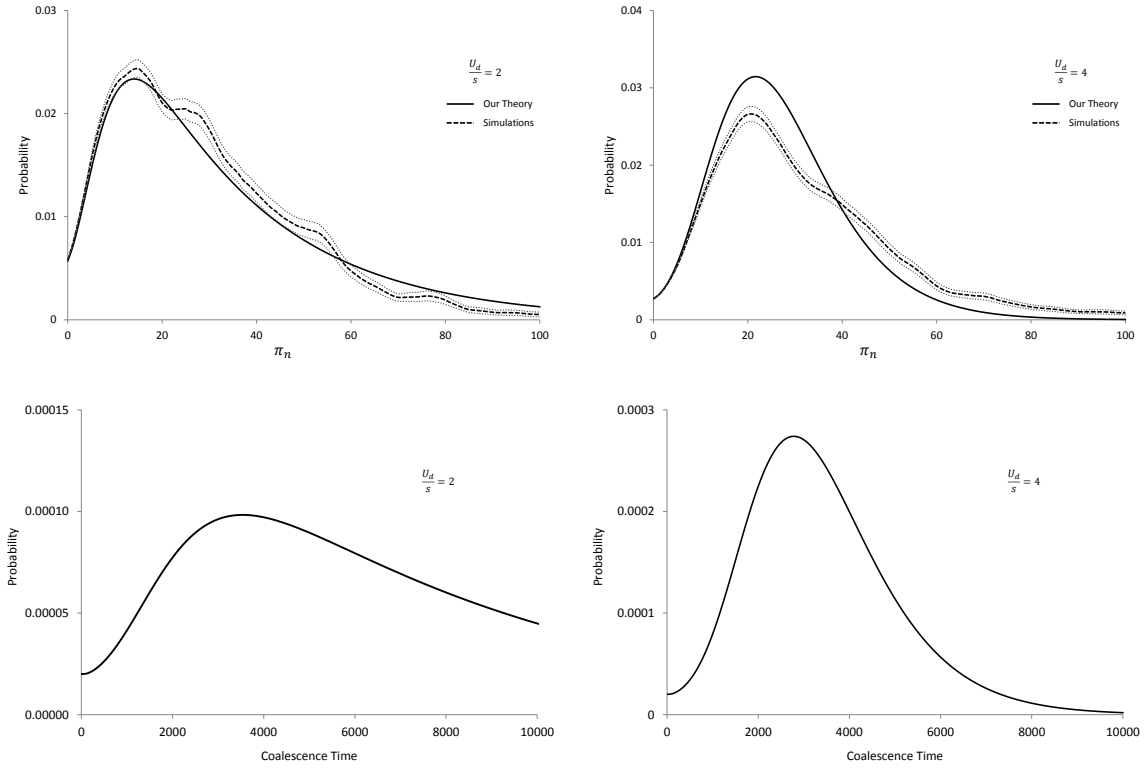

FIG. 5 Characteristic examples of the distributions of π_n and the real coalescent times. **(a)** Theoretical predictions for the distribution of π_n for $U_d/s = 2$, compared to simulation results. **(b)** Theoretical predictions for the distribution of π_n for $U_d/s = 4$, compared to simulation results. Simulation results are averaged across at least 300 independent simulations for each parameter set; shaded regions show one standard error in the simulation results. **(c)** Theoretical predictions for the distribution of real coalescence times for $U_d/s = 2$; note these simply mirror the distribution of π_n, as expected. **(d)** Theoretical predictions for the distribution of real coalescence times for $U_d/s = 4$. In all panels we have $N = 5 \times 10^4$ and $s = 10^{-3}$. Our theory agrees well with the simulations, but note that, as with π_d, we tend to systematically underestimate π_n, and this tendency is worse for larger U_d/s. This is consistent with Muller's ratchet, and as expected becomes more problematic for larger U_d/s. This systematic underestimate becomes less severe (for all values of U_d/s) as we increase N, as expected, but comprehensive simulations for much larger N are computationally prohibitive.

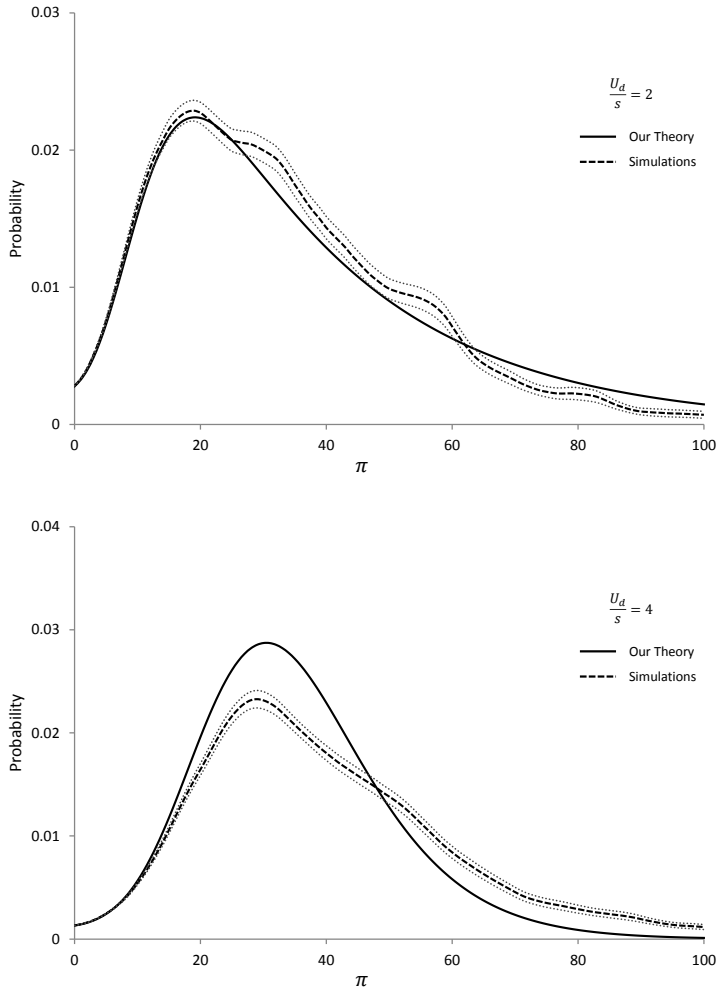

FIG. 6 Characteristic examples of the distribution of total heterozygosity π. Here $N = 5 \times 10^4$, $s = 10^{-3}$ and in **(a)** $U_d/s = 2$, while in **(b)** $U_d/s = 4$. Theoretical predictions are shown as a solid line, simulation results as a dashed line. Simulation results are averaged across at least 300 independent simulations for each parameter set; shaded regions show one standard error in the simulation results. The fit to simulations is good, but we tend to slightly underestimate π, and this tendency is worse for larger U_d/s. This is for the same reasons as in the distributions of π_n and π_d.

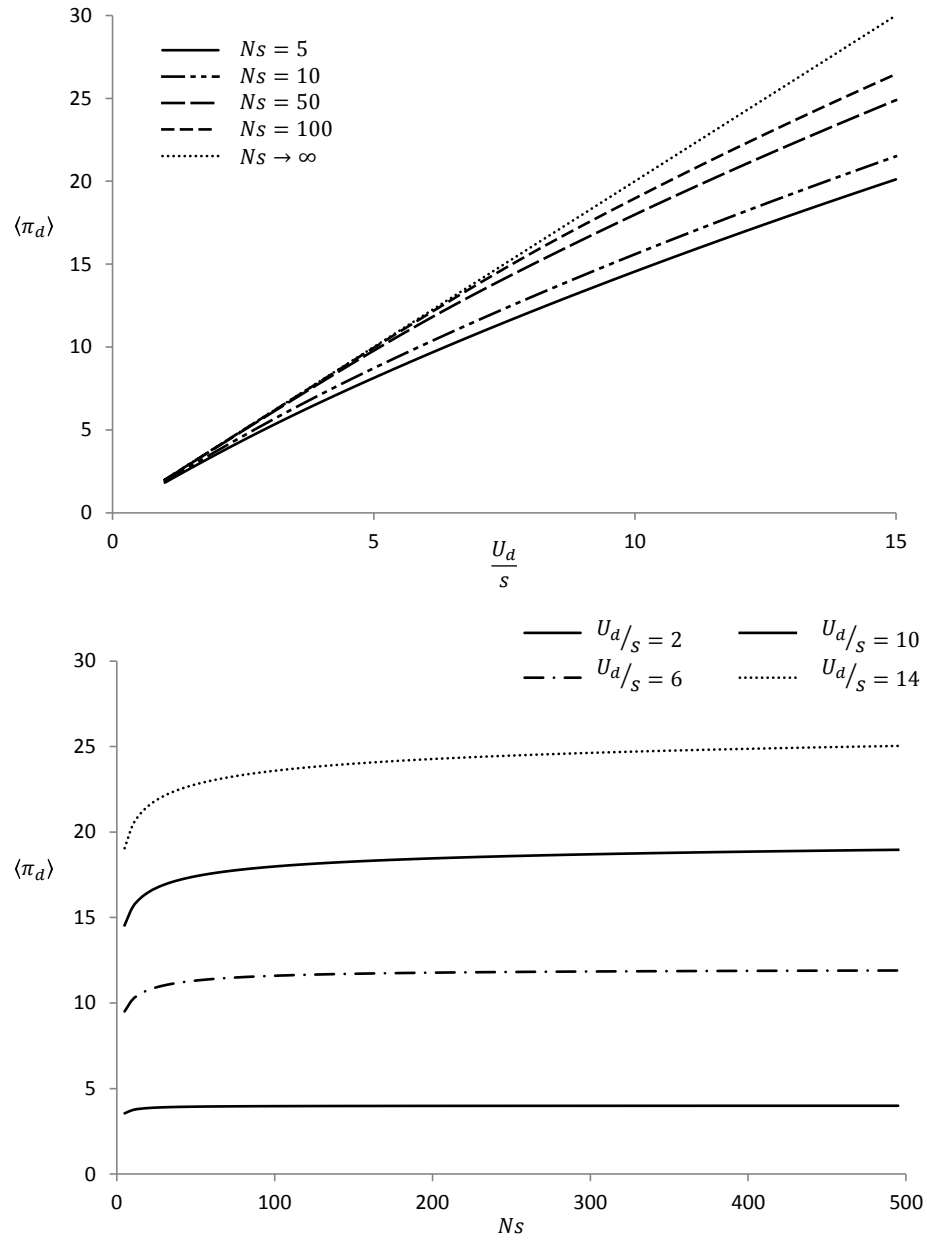

FIG. 7 Theoretical predictions for the mean pairwise heterozygosity at negatively selected sites, $\langle \pi_d \rangle$, as a function of the parameters. **(a)** $\langle \pi_d \rangle$ as a function of U_d/s for several values of Ns. In the "mutation-time" approximation we expect this to be linear 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 mutation-time approximation. As Ns decreases, coalescence within the bulk of the fitness distribution becomes more likely, and hence $\langle \pi_d \rangle$ decreases.

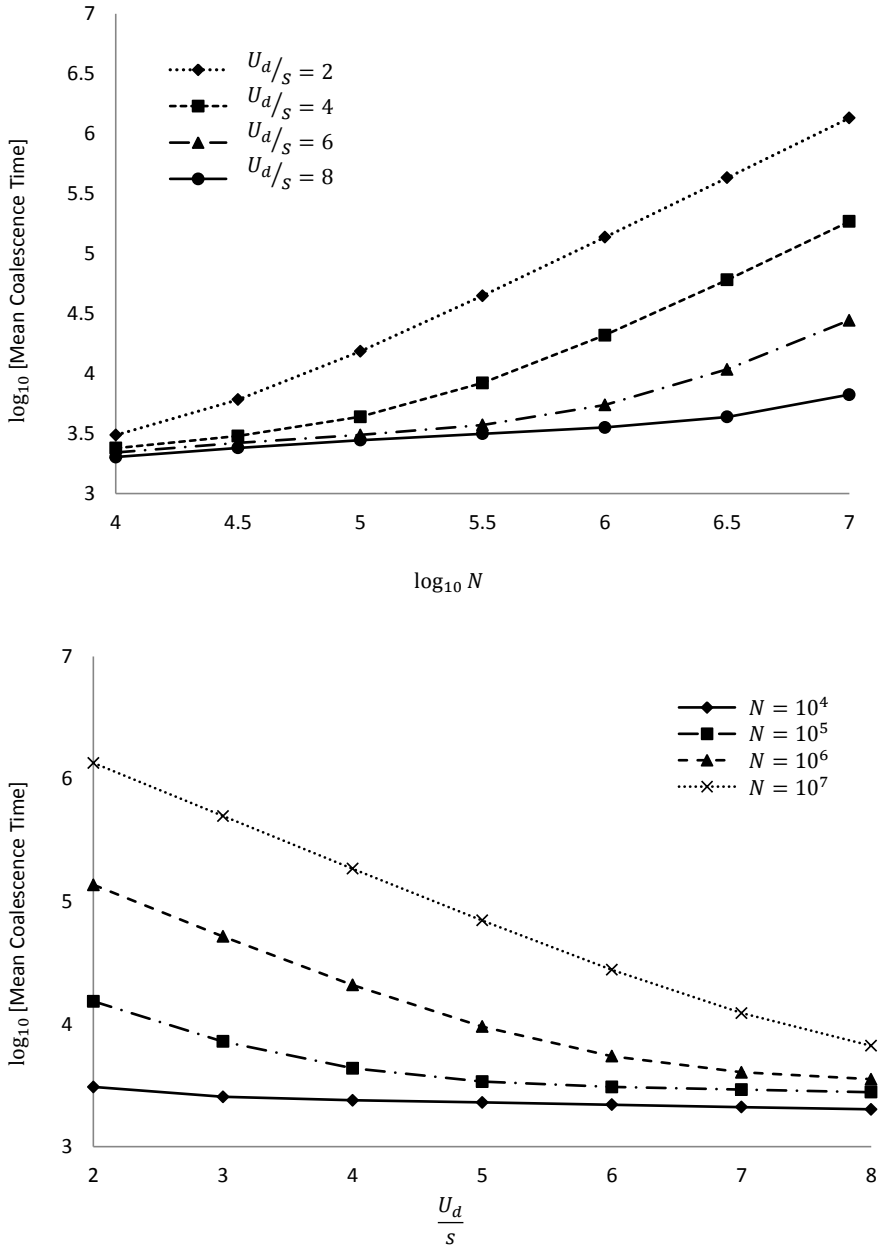

FIG. 8 Theoretical predictions for the mean real coalescence time $\langle t \rangle$. In this figure we fix $s = 10^{-3}$ and show the dependence of the mean pairwise heterozygosity on N and on U_d/s. The mean pairwise heterozygosity at neutral sites, $\langle \pi_n \rangle$ is simply $\langle \pi_n \rangle = 2U_n \langle t \rangle$. **(a)** Mean coalescence time as a function of N for various values of U_d/s. We see that $\langle t \rangle$ increases slowly with N until for large enough N the EPS approximation applies and $\langle t \rangle$ becomes linear in N. **(b)** Mean coalescence time as a function of U_d/s for several values of N. For large N, the dependence is roughly linear, consistent with the EPS approximation. For smaller N, coalescence can occur in the bulk of the fitness distribution, reducing the mean coalescence time.

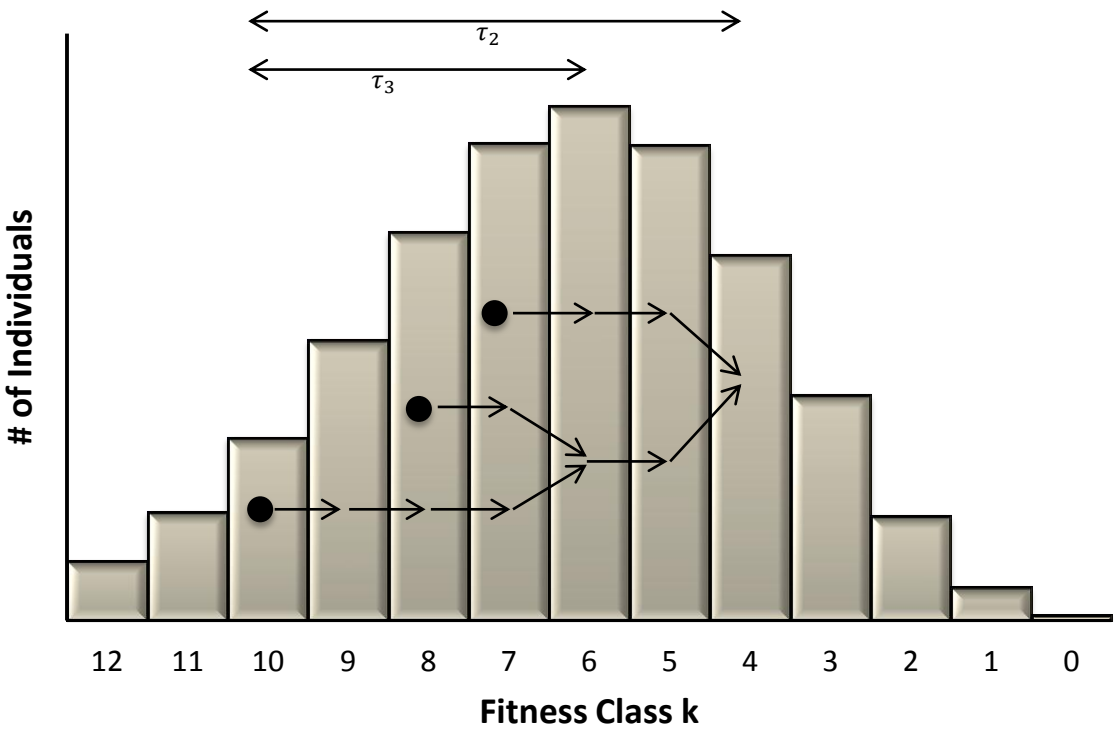

FIG. 9 The fitness-class coalescence process for three individuals, A, B and C, where A and B coalesced τ_3 steptimes ago and C coalesced with the other two τ_2 steptimes ago.

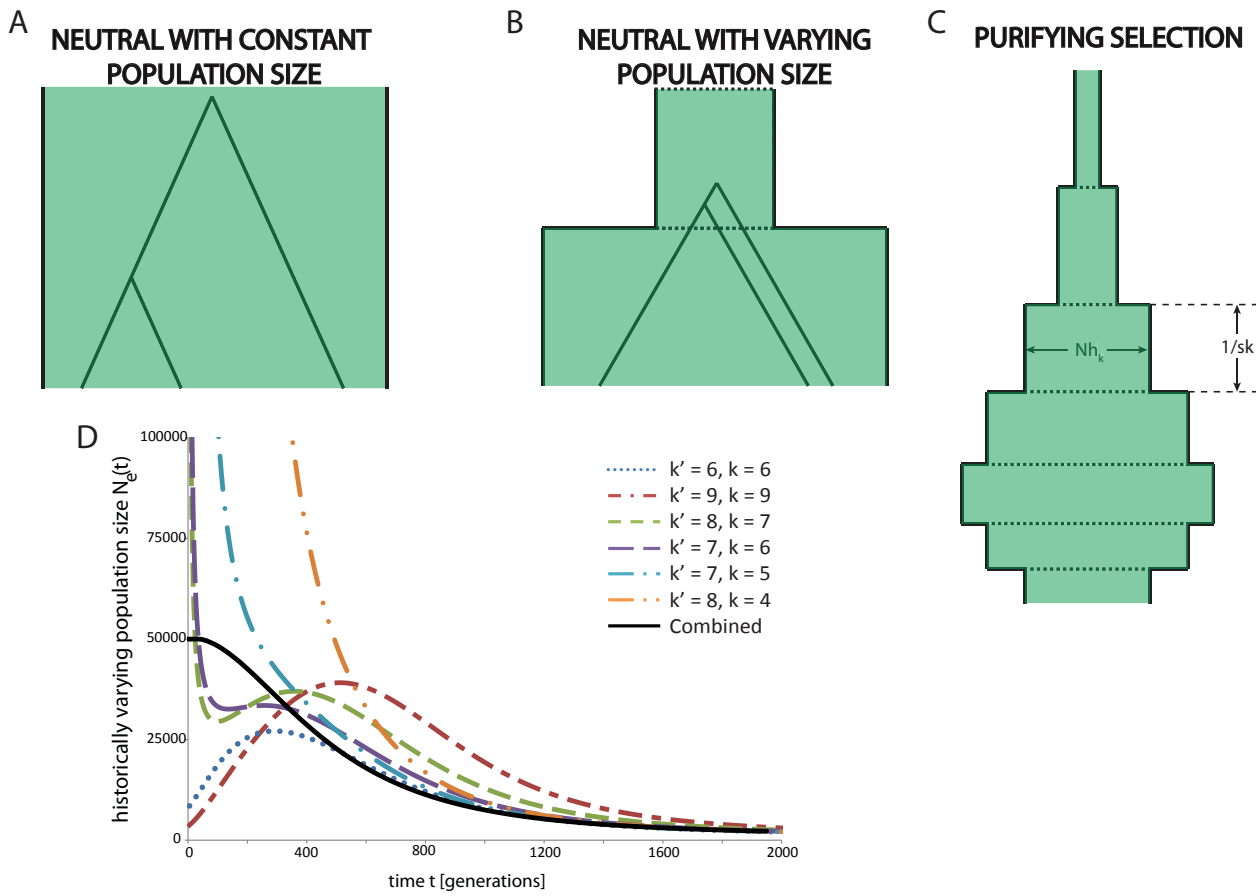

FIG. 10 Relationship between our results and an effective population size approximation. **(a)** A typical coalescent tree in a neutral population of constant size. The coalescent probability per generation between a random pair of individuals is the inverse population size. Time runs from the past at the top to the present at the bottom. **(b)** An example of a neutral coalescent tree in a population which was smaller in the past than the present. The population size is shown as the width in green. Coalescence events are more likely to occur when the population size is smaller. **(c)** The effective population size history for an individual experiencing purifying selection according to our model. The individual spends on average $\frac{1}{sk}$ generations in class k, which has a total size Nh_k. Note that pairs of individuals are sampled from different classes k (i.e. they are not all sampled from the bottom of this picture). Further, the coalescence probabilities also include a factor of $A/2$, which reflects the probability that two lineages are in the same class at the same time. **(d)** The historically varying effective population size $N_e(t)$ for a pair of individuals sampled from classes k and k', as defined in the text, for several values of k and k'. The $N_e(t)$ for two individuals sampled at random from the whole population is also shown. Here $N = 5 \times 10^4$, $U_d/s = 6$, and $s = 10^{-3}$.

Supplementary Information
The Structure of Genealogies in the Presence of Purifying Selection: A "Fitness-Class Coalescent"

Aleksandra M. Walczak[1,*], Lauren E. Nicolaisen[2,*], Joshua B. Plotkin[3], and Michael M. Desai[2]
[1] *CNRS-Laboratoire de Physique Théorique de l'École Normale Supérieure,*
[2] *Department of Organismic and Evolutionary Biology, Department of Physics, and FAS Center for Systems Biology, Harvard University*
[3] *Department of Biology, University of Pennsylvania*
[*] *These authors contributed equally to this work*

SUPPLEMENTARY APPENDIX A: THE FULL CONDITIONAL CALCULATION

In the main text, we focused primarily on the non-conditional approximation to the coalescence probabilities, which led to our simple expression for the coalescence probabilities, Eq. (15). In this Supplementary Appendix, we show how this approximation can be relaxed in our lineage-structure framework by carrying out the full conditional calculation for some of 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 note that the full conditional result can also be obtained from the sum of ancestral paths approach by keeping the higher order terms in Eq. (56) of Appendix A, as described in Supplementary Appendix D, and the validity of the non-conditional approximation can be directly assessed with that approach.

We begin by considering the full conditional result for the probability that two individuals both sampled from class k coalesce in class $k-2$. From Appendix A of the main text, we have

$$P_c^{k,k\to k-2} = I_x^{k-2} \int Q_{k,k}^{k-2}(t_1,t_2) \exp\left[-s(k-2)|t_1-t_2|\right] dt_1 dt_2. \tag{S.1}$$

In order to evaluate this integral, we need to determine the probability distribution of mutant timings $Q_{k,k}^{k-2}(t_1,t_2)$. The time t_1 is 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 $k-1$. To illustrate the main point, we neglect the distortion in the mutant timings due to the fact that individuals did not coalesce in class k and focus only on the distortions due to the fact that coalescence did not occur in class

$k-1$; if desired, the former distortion can also be included using analogous methods. 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

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

Thus the distribution of t_1 and t_2 is given by

$$Q_{k,k}^{k-2}(t_1,t_2) = Q_{k,k}^{k-1}(t_1,t_2|nc) \star Q_{1step}^{k-2}(t_1,t_2), \tag{S.3}$$

where $\star$ 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

$$Q_{k,k}^{k-1}(t_1,t_2|nc) = \frac{Q_{k,k}^{k-1}(t_1,t_2) - Q_{k,k}^{k-1}(t_1,t_2|c)P_c^{k,k\to k-1}}{1-P_c^{k,k\to k-1}}, \tag{S.4}$$

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. (13), we find that

$$Q_{k,k}^{k-1}(t_1,t_2|c) = \frac{I_x^{k-1}}{P_c^{k,k\to k-1}} Q_{k,k}^{k-1}(t_1,t_2) e^{-s(k-1)|t_1-t_2|}. \tag{S.5}$$

We therefore find

$$Q_{k,k}^{k-1}(t_1,t_2|nc) = \frac{1}{1-P_c^{k,k\to k-1}}\left[(sk)^2 e^{-sk(t_1+t_2)} - I_x^{k-1}(sk)^2 e^{-2k(t_1+t_2)} e^{-s(k-1)|t_1-t_2|}\right]. \tag{S.6}$$

Plugging this into our convolution formula for $Q_{k,k}^{k-2}(t_1,t_2)$ and evaluating the integrals by separating out the possible time orderings, we find

$$Q_{k,k}^{k-2}(t_1,t_2) = \frac{k^2\left[s(k-1)\right]^2}{1-P_c^{k,k\to k-1}} e^{-s(k-1)(t_1+t_2)}\left[\left(1-e^{-st_1}\right)\left(1-e^{-2t_2}\right) - \frac{I_x^{k-1}}{k-2}B\right], \tag{S.7}$$

where we have defined

$$B = \frac{1}{(k-2)}\left[1 - e^{-2s\min(t_1,t_2)} - \frac{2}{k}\left(1 - e^{-sk\min(t_1,t_2)}\right) + \frac{1}{k}\left(1 - e^{-2k|t_1-t_2|}\right)\left(e^{-2s\min(t_1,t_2)} - e^{-sk\min(t_1,t_2)}\right)\right]. \tag{S.8}$$

We can now use this expression in Eq. (S.1) to calculate the coalescence probability $P_c^{k,k\to k-2}$. 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 $k-3$, $P_c^{k,k\to k-3}$. This will be given by

$$P_c^{k,k\to k-3} = \int Q_{k,k}^{k-3}(t_1,t_2)\frac{x^2}{h_{k-3}^2} f_{k-3}(x) e^{-s(k-3)|t_1-t_2|} dt_1 dt_2 dx, \tag{S.9}$$

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,

$$Q_{k,k}^{k-3}(t_1,t_2) = Q_{k,k}^{k-2}(t_1,t_2|nc) \star Q_{1step}^{k-3}(t_1,t_2), \tag{S.10}$$

where analogously to the expression in the previous step

$$Q_{k,k}^{k-2}(t_1,t_2|nc) = \frac{1}{1-P_c^{k,k\to k-2}}\left[Q_{k,k}^{k-2}(t_1,t_2) - Q_{k,k}^{k-2}(t_1,t_2|c)P_c^{k,k\to k-2}\right]. \tag{S.11}$$

We note that $Q_{k,k}^{k-2}(t_1,t_2)$ is the expression in Eq. (S.7) we calculated above. As before, we have

$$Q_{k,k}^{k-2}(t_1,t_2|c)P_c^{k,k\to k-2} = I_x^{k-2} Q_{k,k}^{k-2}(t_1,t_2) e^{-s(k-2)|t_1-t_2|}, \tag{S.12}$$

hence we can write

$$Q_{k,k}^{k-2}(t_1,t_2|nc) = \frac{Q_{k,k}^{k-2}(t_1,t_2)}{1-P_c^{k,k\to k-2}}\left[1 - I_x^{k-2} e^{-s(k-2)|t_1-t_2|}\right]. \tag{S.13}$$

Plugging the above expression back into Eq. (S.10), we obtain

$$Q_{k,k}^{k-3}(t_1,t_2) = \frac{s^2(k-1)^2k^2s^2(k-2)^2}{(1-P_c^{k,k\to k-1})(1-P_c^{k,k\to k-2})} e^{-s(k-2)(t_1+t_2)} \int_0^{t_2}\int_0^{t_1} e^{s(k-2)(y+z)} e^{s(k-1)(y+z)} \times\left[1 - I_x^{k-2} e^{-s(k-z)|y-z|}\right]\left[(1-e^{-sy})(1-e^{-sz}) - \frac{I_x^{k-1}}{k-2}B\right]. \tag{S.14}$$

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 ℓ 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 ℓ, 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).

In practice, it is easier to make these full conditional calculations within the sum of ancestral paths approach. As we show in Supplementary Appendix D, that approach leads naturally to a power series in I_x of exactly the form described above, in which the leading order term is the non-conditional approximation and the additional terms represent the conditional corrections. This calculation shows that provided $I_x \ll 1$, which is true provided our usual condition that $Nh_k sk \gg 1$ holds, these higher order terms are all small, and our non-conditional approximation is valid.

These full conditional results are, however, very complex and unilluminating. Therefore we focus here on understanding the general structure of these results, and on showing why the non-conditional approximation is good description of the distribution of mutation timings. 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

$$Q_{k,k}^{k-\ell}(t_1,t_2) = \frac{1}{1-P_c^{k,k\to k-\ell}}\left[Q_{k,k}^{k-\ell}(t_1,t_2) - P_c^{k,k\to k-\ell} Q_{k,k}^{k-2}(t_1,t_2|c)\right]. \tag{S.15}$$

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, $P_c^{k,k\to k-\ell}$. The nature of the correction can be seen by plugging in the distribution of times conditional on coalescence, giving

$$Q_{k,k}^{k-\ell}(t_1,t_2) = \frac{Q_{k,k}^{k-\ell}(t_1,t_2)}{1-P_c^{k,k\to k-\ell}}\left[1 - I_x^{k-\ell}e^{-s(k-\ell)|t_1-t_2|}\right]. \tag{S.16}$$

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 ℓ 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 the 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. All of these heuristic conclusions are reflected in the fact that the full conditional result we calculate in the sum of ancestral paths approach is equal to the non-conditional result plus corrections that are small provided $I_x \gg 1$.

SUPPLEMENTARY APPENDIX B: 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. (48). Specifically, we need to calculate

$$Q_k^{k-\ell}(t) = Q_k^{k-1}(t) \star Q_{k-1}^{k-2}(t) \star Q_{k-2}^{k-3}(t) \star \ldots \star Q_{k-\ell+1}^{k-\ell}(t), \tag{S.17}$$

where $\star$ refers to a convolution and

$$Q_{k-\ell+1}^{k-\ell}(t) = s(k-\ell+1)e^{-s(k-\ell+1)t}, \tag{S.18}$$

as given by Eq. (6). In general, the convolution of n exponential distributions with parameters $\lambda_1 \ldots \lambda_n$ is given by

$$\sum_{i=0}^{n-1} \lambda_i e^{-\lambda_i t} \prod_{j=0,\neq i}^{n-1} \frac{\lambda_j}{\lambda_j - \lambda_i}. \tag{S.19}$$

Applying this identity with $\lambda_i = s(k-i)$, we find

$$Q_k^{k-\ell}(t) = \sum_{i=0}^{\ell-1} s e^{-s(k-i)t} \left(\frac{\prod_{j=0}^{\ell-1} k-j}{\prod_{j=0,\neq i}^{\ell-1} i-j} \right) \tag{S.20}$$

We can simplify this expression by noting that

$$\prod_{j=0}^{\ell-1} (k-j) = \frac{k!}{(k-\ell)!}, \tag{S.21}$$

and similarly that

$$\prod_{j=0,\neq i}^{\ell-1} (i-j) = i!(\ell-1-i)!(-1)^{\ell-1-i}. \tag{S.22}$$

This means we have

$$Q_k^{k-\ell}(t) = \sum_{i=0}^{\ell-1} s\ell e^{-s(k-i)t}(-1)^{\ell-i-1}\binom{\ell-1}{i}\binom{k}{k-\ell}. \tag{S.23}$$

We can evaluate this sum by recognizing the binomial expansion formula

$$(1+x)^n = \sum_{i=0}^{n} x^i \binom{n}{i}, \tag{S.24}$$

where we identify $x = -e^{st}$. We find

$$Q_k^{k-\ell}(t) = s\ell\binom{k}{\ell} e^{-skt}\left(e^{st}-1\right)^{\ell-1}. \tag{S.25}$$

More generally, we have

$$Q_a^b(t) = s(a-b)\binom{a}{b} e^{-sat}\left(e^{st}-1\right)^{a-b-1}. \tag{S.26}$$

SUPPLEMENTARY APPENDIX C: 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. (48) can be rewritten as

$$P_c^{k,k'\to k'-\ell} = \frac{1}{1+2Nh_{k-\ell}s(k-\ell)}\left[I_1+I_2\right], \tag{S.27}$$

where we have defined

$$I_1 = \int_0^\infty Q_{k'}^{k-\ell}(t_1)e^{-s(k-\ell)t_1}\int_0^{t_1} Q_k^{k-\ell}(t_2)e^{s(k-\ell)t_2}dt_2dt_1 \tag{S.28}$$

$$I_2 = \int_0^\infty Q_k^{k-\ell}(t_2)e^{-s(k-\ell)t_2}\int_0^{t_2} Q_{k'}^{k-\ell}(t_1)e^{s(k-\ell)t_1}dt_1dt_2. \tag{S.29}$$

Note that both I_1 and I_2 involve integrals of the form

$$I_a = \int_0^t Q_a^b(t')e^{sbt'}dt'. \tag{S.30}$$

Plugging in the results for the non-conditional distributions of mutant timings, Eq. (S.26), and making use of the binomial expansion formula for $(1+x)^n$ noted in Supplementary

Appendix B, we find this integral becomes

$$I_a = s(a-b)\binom{a}{b}\int_0^t e^{s(b-a)t'}\left(e^{st'}-1\right)^{a-b-1}dt' \tag{S.31}$$

$$= s(a-b)\binom{a}{b}\sum_{i=0}^{a-b-1}(-1)^{a-b-1+i}\binom{a-b-1}{i}\int_0^t e^{s(b-a+i)t'}dt' \tag{S.32}$$

$$= (a-b)\binom{a}{b}(-1)^{a-b}\sum_{i=0}^{a-b-1}\frac{(-1)^i}{a-b}\binom{a-b}{i}\left(e^{s(b-a+i)t}-1\right) \tag{S.33}$$

$$= \binom{a}{b}(-1)^{a-b}\sum_{i=0}^{a-b}(-1)^i\binom{a-b}{i}\left(e^{s(b-a+i)t}-1\right) \tag{S.34}$$

$$= \binom{a}{b}(-1)^{a-b}e^{s(b-a)t}\sum_{i=0}^{a-b}\left(-e^{st}\right)^i\binom{a-b}{i} \tag{S.35}$$

$$= \binom{a}{b}e^{s(b-a)t}\left(e^{st}-1\right)^{a-b}. \tag{S.36}$$

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

$$I_b = \int_0^\infty Q_a^b(t)\binom{c}{b}e^{-sct}\left(e^{st}-1\right)^{c-b}dt. \tag{S.37}$$

Using similar manipulations to those above, we find

$$I_b = (a-b)\binom{a}{b}\binom{c}{b}\int_0^\infty e^{-s(a+c)t}\left(e^{st}-1\right)^{a+c-2b-1}dt \tag{S.38}$$

$$= s(a-b)\binom{a}{b}\binom{c}{b}(-1)^{a+c-1}\sum_{i=0}^{a+c-2b-1}\binom{a+c-2b-1}{i}(-1)^i\int_0^\infty e^{-s(a+c-i)t}dt \tag{S.39}$$

$$= (a-b)\binom{a}{b}\binom{c}{b}(-1)^{a+c-1}\sum_{i=0}^{a+c-2b-1}(-1)^i\binom{a+c-2b-1}{i}\frac{1}{a+c-i}. \tag{S.40}$$

Using the partial fraction decomposition

$$\frac{1}{\binom{n+x}{n}} = \sum_{i=1}^{n}(-1)^{i-1}\binom{n}{i}\frac{i}{x+i}, \tag{S.41}$$

we find

$$I_b = \frac{\frac{a-b}{a+c-2b}\binom{a}{b}\binom{c}{b}(-1)^{a+c}}{\binom{-2b-1}{a+c-2b}} = \frac{\frac{a-b}{a+c-2b}\binom{a}{b}\binom{c}{b}(-1)^{2b}}{\binom{a+c}{a+c-2b}}. \tag{S.42}$$

We can now use this result for I_b to determine I_1 and I_2, and hence compute $P_c^{k,k'\to k'-\ell}$. We find

$$P_c^{k,k'\to k'-\ell} = \frac{1}{1+2Nh_{k-\ell}s(k-\ell)}\frac{\binom{k'}{k-\ell}\binom{k}{k-\ell}}{\binom{k+k'}{2\ell+k'-k}}. \tag{S.43}$$

As we noted in the main text, this is just

$$P_c^{k,k'\to k-\ell} = \frac{1}{1+2Nh_{k-\ell}s(k-\ell)} A_\ell^{k,k'}, \tag{S.44}$$

with $A_\ell^{k,k'}$ as defined in Eq. (16). Note that when $k = k'$, this result simplifies to $P_c^{k,k\to k-\ell}$ as defined in the main text, as expected.

SUPPLEMENTARY APPENDIX D: COMPUTING SUMS OF ANCESTRAL PATHS

In this appendix, we describe the calculation of $\phi_k^{k'}(\ell)$ using the sum of ancestral paths approach.

Calculation of $\phi_k^k(3)$: We begin by considering a simpler specific case, where $k = k'$ and $\ell = 3$. There are a total of $\binom{6}{3} = 20$ possible ancestral paths by which two individuals sampled from class k can coalesce in class $k-3$. These can be separated into four types, according to whether the two ancestral lineages were ever together in classes $k-1$ or $k-2$. We can list all paths of each type, using the notation that A is a mutation event in the first lineage, and B is a mutation event in the second lineage. We have

$$\underbrace{\begin{pmatrix} ABABAB \\ ABABBA \\ ABBAAB \\ ABBABA \\ BAABAB \\ BAABBA \\ BABAAB \\ BABABA \end{pmatrix}}_{\binom{2}{1}\binom{2}{1}\binom{2}{1}=8 \text{ ways}} \quad \underbrace{\begin{pmatrix} ABAABB \\ ABBBAA \\ BAAABB \\ BABBAA \end{pmatrix}}_{\binom{2}{1}\left(\binom{4}{2}-\binom{2}{1}\binom{2}{1}\right)=4 \text{ ways}} \quad \underbrace{\begin{pmatrix} AABBAB \\ AABBBA \\ BBAAAB \\ BBAABA \end{pmatrix}}_{\binom{2}{1}\left(\binom{4}{2}-\binom{2}{1}\binom{2}{1}\right)=4 \text{ ways}} \quad \underbrace{\begin{pmatrix} AAABBB \\ AABABB \\ BBBAAA \\ BBABAA \end{pmatrix}}_{\binom{6}{3}-\text{others}=4\text{ways}}.$$

The probabilities of all paths of a particular type are identical. We can calculate the probability of each of the four types of paths using the same logic as outlined in the main

text. We find

$$P(AAABBBc) = I_x^{k-3}\frac{k(k-1)(k-2)}{8(2k-1)(2k-3)(2k-5)}\left(1-I_x^k\right), \tag{S.45}$$

$$P(AABBABc) = I_x^{k-3}\frac{k(k-1)(k-2)}{8(2k-1)(2k-3)(2k-5)}\left(1-I_x^k\right)\left(1-I_x^{k-1}\right), \tag{S.46}$$

$$P(ABAABBc) = I_x^{k-3}\frac{k(k-1)(k-2)}{8(2k-1)(2k-3)(2k-5)}\left(1-I_x^k\right)\left(1-I_x^{k-2}\right), \tag{S.47}$$

$$P(ABABABc) = I_x^{k-3}\frac{k(k-1)(k-2)}{8(2k-1)(2k-3)(2k-5)}\left(1-I_x^k\right)\left(1-I_x^{k-1}\right)\left(1-I_x^{k-2}\right) \tag{S.48}$$

Summing over all the possible paths, we find

$$\phi_k^k(3) = I_{k-3}\frac{\binom{k}{k-3}\binom{k}{k-3}}{\binom{2k}{6}}\left[1-\frac{\binom{2}{1}\binom{4}{2}}{\binom{6}{3}}I_{k-1}-\frac{\binom{2}{1}\binom{4}{2}}{\binom{6}{3}}I_{k-2}+\frac{\binom{2}{1}\binom{2}{1}\binom{2}{1}}{\binom{6}{3}}I_{k-1}I_{k-2}\right]. \tag{S.49}$$

We now pause to consider the form of the probabilities of each type of ancestral path. These probabilities differ only by factors of $(1-I_x^{k-i})$. One such factor arises each time the two ancestral lineages are together in class $k-i$. In other words, we can rewrite the probability of each path as the probability of an undistorted path (defined to be a path in which the contributions due to the possibility of coalescence in previous classes are neglected), times a correction for each class in which the two lineages are together:

$$P(AAABBBc) = P(\text{Undistorted Path})\left(1-I_x^k\right) \tag{S.50}$$

$$P(AABBABc) = P(\text{Undistorted Path})\left(1-I_x^k\right)\left(1-I_x^{k-1}\right) \tag{S.51}$$

$$P(ABAABBc) = P(\text{Undistorted Path})\left(1-I_x^k\right)\left(1-I_x^{k-2}\right) \tag{S.52}$$

$$P(ABABABc) = P(\text{Undistorted Path})\left(1-I_x^k\right)\left(1-I_x^{k-1}\right)\left(1-I_x^{k-2}\right). \tag{S.53}$$

By definition, the "undistorted path" probability is the probability neglecting the contributions due to the possibility of coalescence in previous steps, and is therefore the same for all paths. We have

$$P(\text{Undistorted Path}) = \frac{k(k-1)(k-2)k(k-1)(k-2)}{2k(2k-1)(2k-2)(2k-3)(2k-4)(2k-5)}I_x^{k-\ell} \tag{S.54}$$

$$= \frac{\frac{k!}{(k-3)!}\frac{k!}{(k-3)!}}{\frac{2k!}{(2k-6)!}}I_x^{k-\ell}. \tag{S.55}$$

Using these results, we can write $\phi_k^k(3)$ as

$$\phi_k^k(3) = [\#\text{ of Paths}]\,P(\text{Undistorted Path})\left[F_k(1-I_x^k)+F_{k,k-1}(1-I_x^k)(1-I_x^{k-1})\right.$$
$$\left.+F_{k,k-2}(1-I_x^k)(1-I_x^{k-2})+F_{k,k-1,k-2}(1-I_x^k)(1-I_x^{k-1})(1-I_x^{k-2})\right], \tag{S.56}$$

where we have defined $F_{\{a\}}$ to be the fraction of paths that are together in the set of classes $\{a\}$ (and are not together in any other class).

Calculation of $\phi^k_{k'}(\ell)$: We now use this approach to calculate the coalescence probability in the general case. The probability of any particular ancestral path from k and k' to $k-\ell$ is the product of the individual probabilities of each mutational step that makes up this path. Each such individual probability consists of three parts: a numerator, which depends only on the current class of the lineage that mutates, divided by a denominator, which depends only on the sum of the current set of classes for both lineages, times a correction factor of $(1 - I_x^{k-i})$ if the two lineages are in the same class at that step.

Although in each ancestral path the mutations will occur in a different order, all paths will ultimately consist of the same set of mutations ($k' \to k'-1 \to \ldots \to k-\ell$ and $k \to k-1 \to \ldots \to k-\ell$). Therefore, regardless of the path taken, the product of the numerators from each step will be identical. Similarly, the sum of the current set of classes will begin at $k'+k$, and decrement by one each time a deleterious mutation occurs, until both lineages are in the final class ($k'+k \to k'+k-1 \to \ldots \to 2k-2\ell$). Therefore, regardless of the path taken, the product of the denominators from each step will also be identical. Therefore, the paths will differ only by the correction factor $(1-I_x^{k-i})$ for each class in which the two ancestral lineages are together. This means that, analogous to the case of $\phi^k_k(3)$ we described above, the probability of each path is the probability of an "undistorted path" times the appropriate correction factor. The probability of the undistorted path is

$$P(\text{Undistorted Path}) = \frac{k'(k'-1)\ldots(k-\ell+1)k(k-1)\ldots(k-\ell+1)}{(k'+k)(k'+k-1)\ldots(2k-2\ell+1)} I_x^{k-\ell}. \tag{S.57}$$

We can now sum up all possible paths to obtain

$$\begin{aligned}\phi^k_{k'}(\ell) = [\#\text{ of Paths}]\, P(\text{Undistorted Path}) \Bigg[& F_\emptyset + \sum_{i=0}^{\ell} F_{k-i}(1-I_x^{k-i}) \\ & + \sum_{i=0}^{\ell-1}\sum_{j>i}^{\ell} F_{k-i,k-j}(1-I_x^{k-i})(1-I_x^{k-j}) \\ & + \sum_{i=0}^{\ell-2}\sum_{j>i}^{\ell-1}\sum_{m>j}^{\ell} F_{k-i,k-j,k-m}(1-I_x^{k-i})(1-I_x^{k-j})(1-I_x^{k-m}) + \ldots \Bigg],\end{aligned} \tag{S.58}$$

where as before $F_{\{a\}}$ is the fraction of paths that are together in the set of classes $\{a\}$ (and are not together in any other class). Note that there are a total of $\ell+1$ terms in this

equation, representing the possibility that the two lineages can be together in anywhere from 0 to ℓ of the classes. We can rearrange these terms to write

$$\begin{aligned}\phi_{k'}^{k}(\ell) = [\#\text{ of Paths}]\, P(\text{Undistorted Path}) \Bigg[1 - \sum_{i=0}^{\ell} G_{k-i} I_x^{k-i} \\ + \sum_{i=0}^{\ell-1} \sum_{j>i}^{\ell} G_{k-i,k-j} I_x^{k-i} I_x^{k-j} \\ - \sum_{i=0}^{\ell-2} \sum_{j>i}^{\ell-1} \sum_{m>j}^{\ell} G_{k-i,k-j,k-m} I_x^{k-i} I_x^{k-j} I_x^{k-m} + \ldots \Bigg],\end{aligned} \tag{S.59}$$

where we have defined $G_{\{a\}}$ to be the fraction of paths that are together in *at least* the set of classes $\{a\}$.

We can evaluate each of these factors of G. For example, the fraction of paths that are together in class $k-i$ equals the number of ways for the two lineages to descend from classes k' and k to be together in class $k-i$, $\binom{k'-k+2i}{i}$, times the number of ways for the two lineages to descend from class $k-i$ to be together in class $k-\ell$, $\binom{2i-2\ell}{i-\ell}$, divided by the total number of ways for the two lineages to descend from classes k' and k to be together in $k-\ell$, $\binom{k'-k+2\ell}{\ell}$. Using this logic, we find

$$\begin{aligned}\phi_{k'}^{k}(\ell) = [\#\text{ of Paths}]\, P(\text{Undistorted Path}) \\ \times \left[1 - \sum_{i=0}^{\ell-1} \frac{\binom{k'-k+2i}{i}\binom{2\ell-2i}{\ell-i}}{\binom{k'-k+2\ell}{\ell}} I_x^{k-i} + \sum_{i=0}^{\ell-2} \sum_{j>i}^{\ell-1} \frac{\binom{k'-k+2i}{i}\binom{2j-2i}{j-i}\binom{2\ell-2j}{\ell-j}}{\binom{k'-k+2\ell}{\ell}} I_x^{k-i} I_x^{k-j} \ldots\right].\end{aligned} \tag{S.60}$$

The total number of paths is $\binom{k'-k+2\ell}{\ell}$, so we finally find that the full probability of coalescence in class $k-\ell$ is

$$\begin{aligned}\phi_{k}^{k'}(\ell) = I_x^{k-\ell} \frac{\binom{k'}{k-\ell}\binom{k}{k-\ell}}{\binom{k'+k}{k'-k+2\ell}} \left[1 - \sum_{i=0}^{\ell-1} \frac{\binom{k'-k+2i}{i}\binom{2\ell-2i}{\ell-i}}{\binom{k'-k+2\ell}{\ell}} I_x^{k-i} + \right. \\ \left. \sum_{i=0}^{\ell-2} \sum_{j>i}^{\ell-1} \frac{\binom{k'-k+2i}{i}\binom{2j-2i}{j-i}\binom{2\ell-2j}{\ell-j}}{\binom{k'-k+2\ell}{\ell}} I_x^{k-i} I_x^{k-j} - \ldots\right].\end{aligned} \tag{S.61}$$

This is Eq. (56) from the main text. Note that it equals our non-conditional result for $P_c^{k,k'\to\ell}$ times a correction factor. There are a total of $\ell+1$ terms in this correction factor. This full correction factor can be arbitrarily complex for large ℓ, so we do not write out a general form here. However, it is straightforward to calculate for any values of k, k', and ℓ; a Mathematica script to do so is available on request.

SUPPLEMENTARY APPENDIX E: THE CORRESPONDENCE BETWEEN STEPTIMES AND REAL TIMES

In this Supplementary Appendix, we calculate the correspondence between steptimes and the actual times measured in generations. Our goal is to calculate the probability distribution of real coalescence times, $\psi(t|k,k',\ell)$, given that individuals were initially in classes k and k' and coalesced in class $k-\ell$.

To begin, we neglect the coalescence time within class $k-\ell$, and consider the distribution of the time at which an ancestor of one of the two sampled individuals first mutated from class $k-\ell$ to class $k-\ell+1$. We refer to this as $\psi_1(t|k,k',\ell)$. We first calculate the joint distribution of the times at which both ancestors mutated out of the class, $R^{k-\ell}_{k,k'}(t_1,t_2)$. Conditional on coalescence in class $k-\ell$, $R^{k-\ell}_{k,k'}(t_1,t_2)$, is given by the probability of t_1 and t_2 and coalescence divided by the total probability of coalescence. That is,

$$R(t_1,t_2) = \frac{P(coal|t_1,t_2)P(t_1,t_2)}{P(coal)}. \tag{S.62}$$

Substituting in the relevant expressions from the main text, this gives

$$R^{k-\ell}_{k,k'}(t_1,t_2) = \frac{1}{A^{k,k'}_{\ell}} Q^{k-\ell}_{k,k'}(t_1,t_2) e^{-s(k-\ell)|t_1-t_2|}. \tag{S.63}$$

The time at which the first ancestor mutated out of class $k-\ell$ is the longer of the two times t_1 and t_2,

$$\psi(t|k,k',\ell) = \left[\int_0^t R^{k-\ell}_{k,k'}(t_1,t)dt_1 + \int_0^t R^{k-\ell}_{k,k'}(t,t_2)dt_2\right]. \tag{S.64}$$

Substituting in our expression for $R^{k-\ell}_{k,k'}(t_1,t_2)$ and carrying out the integrals as in Supplementary Appendix C, we find

$$\psi_1(t|k,k',\ell) = s\pi_d e^{-s(k'+k)t}(e^{st}-1)^{\pi_d-1}\binom{k'+k}{\pi_d}, \tag{S.65}$$

where we have used $\pi_d = k'-k+2\ell$.

We can alternatively calculate $\psi_1(t|k,k',\ell)$ using our sum of ancestral paths approach. As before, we imagine two individuals sampled from classes k and k' and condition on them coalescing in class $k-\ell$. Consider a case where $k \neq k'$. Then the first event in the history of these two individuals must be a deleterious mutation. Since these mutations happen at

rate sk and sk' in each lineage, the distribution of times since this mutation occurred in one of the two ancestral lineages is

$$P(t) = s(k+k')e^{-s(k+k')t}. \tag{S.66}$$

With probability $\frac{k'}{k+k'}$, this mutation is in the lineage sampled from class k', in which case the two lineages are now in classes k and $k'-1$. Alternatively, the mutaion occurred in the lineage sampled from k and the lineages are in classes $k-1$ and k'.

We can now consider the time to the next event backwards in time. If the two lineages are in the same class (but not yet in class $k-\ell$), the distribution of times to the next deleterious mutation event is somewhat shorter, because we are conditioning on coalescence not occuring. However, provided that $2sk_1 \gg \frac{1}{Nh_k}$ (the condition we are already making elsewhere), this shortening of the time will be a small correction and neglecting it is a good approximation.

Making this approximation, the rate at which the next deleterious mutation event occurs when the two lineages are in classes k_1 and k_2 is just $s(k_1+k_2)$. Regardless of the order in which these mutations happen between the two lineages, this sum is simply decreased by s at each step. This will continue until the both ancestral lineages are in class $k-\ell$. Therefore, the distribution of times until the original mutation out of class $k-\ell$ is given by:

$$\psi_1(t|k',k,\ell) = s(k'+k)e^{-s(k'+k)t} \star s(k'+k-1)e^{-s(k'+k-1)t} \star \ldots \star s(2k-2\ell+1)e^{-s(2k-2\ell+1)t}. \tag{S.67}$$

This can be written as

$$\psi_1(t|k',k,\ell) = \lambda_0 e^{-\lambda_0 t} \star \lambda_1 e^{-\lambda_1 t} \star \ldots \star \lambda_{k'-k+2\ell-1} e^{-\lambda_{k'-k+2\ell-1} t}, \tag{S.68}$$

where we have defined:

$$\lambda_i = s(k'+k-i). \tag{S.69}$$

We can compute this convolution as in Supplementary Appendix B (compare to Eq. (S.17) for $Q^{2k-2\ell}_{k+k'}(t)$). We find

$$\psi_1(t|k,k',\ell) = s\pi_d e^{-s(k'+k)t}(e^{st}-1)^{\pi_d-1}\binom{k'+k}{\pi_d}, \tag{S.70}$$

identical to the result of our lineage structure calculation above.

Distribution of Coalescence Times: To calculate the correspondence between step-times and real times, we now need to add the time it takes two individuals two coalesce in class $k-\ell$, which we refer to as $\psi_2(t|k,k',\ell)$, to the time it took them both to get to that class, $\psi_1(t|k,k',k-\ell)$. The rate of coalescence once in class $k-\ell$ is $\frac{1}{Nh_{k-\ell}}$, so we have

$$\psi_2(t|k',k,\ell) = (2s(k-\ell) + 1/Nh_{k-\ell})\, e^{-[2s(k-\ell)+1/Nh_{k-l}]t}. \tag{S.71}$$

Putting this together, the full distribution of times since coalescence is

$$\psi(t|k',k,\ell) = \psi_1(t|k',k,\ell) \star \psi_2(t|k',k,\ell). \tag{S.72}$$

Carrying out this convolution (and expanding the binomial factor $(e^{st}-1)^{\pi_d-1}$ in ψ_1), we find

$$\psi(t|k',k,\ell) = \sum_{i=0}^{\pi_d-1} s\pi_d(-1)^{\pi_d-i-1}\binom{\pi_d-1}{i}\binom{k'+k}{\pi_d}\frac{B}{A-B}\left(e^{-sBt}-e^{-sAt}\right), \tag{S.73}$$

where we have defined $A \equiv k'+k-i$ and $B \equiv 2\,(k-\ell) + \frac{1}{Nsh_{k-\ell}}$.

SUPPLEMENTARY APPENDIX F: AN ALTERNATIVE APPROACH TO NEUTRAL DIVERSITY

Instead of calculating the distribution of neutral heterozygosity by first computing the distribution of real times, we could alternatively incorporate neutral mutations directly into the sum of ancestral paths framework. This completely bypasses the correspondence with real coalescence times. To do this, we characterize ancestral paths not only by the ordering of deleterious mutation and coalescence events, but also by the ordering of neutral mutations. This means that if we sample two individuals A and B, there are five types of events that can happen in their ancestral paths: a deleterious mutation (DM) in A or in B, a neutral mutation (NM) in either A or in B, and or a coalescence (C) event (if A and B are currently in the same class).

We now imagine that we sample two individuals from classes k and k', and that they coalesce in class $k-\ell$. Our goal is to calculate the probability distribution of π_n given k, k', and ℓ, $\rho(\pi_n|k,k',\ell)$. We will find it helpful to divide the five types of events that can occur into two classes: neutral mutations on the one hand, and deleterious mutations or coalescence (which we call "steps") on the other. We begin by computing the probability

that a given number of NMs occur before the next DM or C events (i.e. the number of neutral mutations that occur at this "step"). We have

$$P(\text{a NMs, then DM in } k' \text{ or } k'|k',k) = \left(\frac{\frac{2U_n}{s}}{k'+k+\frac{2U_n}{s}}\right)^a \frac{k+k'}{k'+k+\frac{2U_n}{s}}, \tag{S.74}$$

where we have made our usual assumption that $Nh_k sk \gg 1$, allowing us to neglect the rates of coalescence events (when $k = k'$) in writing this expression.

This probability only depends on the sum of the current classes the individulas are in. At each subsequent step, regardless of the path taken, this sum of the classes will decrease by one. Therefore, the probability that a_i neutral mutations occur at step i is independent of the path taken. This observation allows us to calculate the probability that a given total number of neutral mutations have occurred since coalescence. We first calculate the probability that a given number of neutral mutations have occurred since the first deleterious mutation out of the $k-\ell$ class. We will add in the additional neutral mutations once in the $k-\ell$ class at the end.

In order for π_n neutral mutations to have occurred since the first deleterious mutation out of class $k-\ell$, we require that a_0 mutations occurred at the first step, a_1 mutations occurred at the second step, and so on, such that $a_0 + a_1 + \ldots + a_{k'-k+2\ell-1} = \pi_n$. This gives

$$\rho(\pi_n = X|k',k,\ell) = \frac{\frac{(k'+k)!}{(2k-2\ell)!}}{\frac{(\frac{2U_n}{s}+k'+k)!}{(\frac{2U_n}{s}+2k-2\ell)!}} \sum_{|\vec{a}|=X} \left(\frac{2U_n/s}{2U_n/s+k+k'}\right)^{a_0} \cdots \left(\frac{2U_n/s}{2U_n/s+2k-2l+1}\right)^{a_{k'-k+2l-1}}. \tag{S.75}$$

We can define $x \equiv 2U_n/s + k + k'$, recognize $\pi_d = k' - k + 2\ell$, and relabel the a_i as

$$a_0 \to X - b_0, \quad a_1 \to b_0 - b_1, \quad \ldots \quad a_{\pi_d-2} \to b_{\pi_d-3} - b_{\pi_d-2}, \quad a_{\pi_d-1} \to b_{\pi_d-2}. \tag{S.76}$$

This gives

$$\begin{aligned}\rho(\pi_n = X|k',k,\ell) = \frac{\binom{k'+k}{\pi_d}}{\binom{\frac{2U_n}{s}+k'+k}{\pi_d}} \left(\frac{2U_n}{s}\right)^X \left(\frac{1}{x}\right)^X \sum_{b_0=0}^{X} \left(\frac{x}{x-1}\right)^{b_0} \\ \sum_{b_1=0}^{b_0} \left(\frac{x-1}{x-2}\right)^{b_1} \cdots \sum_{b_{\pi_d-2}=0}^{b_{\pi_d-3}} \left(\frac{x-\pi_d+2}{x-\pi_d+1}\right)^{b_{\pi_d-2}}.\end{aligned} \tag{S.77}$$

To simnplify this expression, it is helpful to define a function f such that:

$$\mathbf{f}\left(A,B\right)\equiv\left(\frac{1}{x}\right)^{X}\sum_{b_0=0}^{X}\left(\frac{x}{x-1}\right)^{b_0}\sum_{b_1=0}^{b_0}\left(\frac{x-1}{x-2}\right)^{b_1}\cdots\sum_{b_{A-1}=0}^{X}\left(\frac{x-A+1}{x-A}\right)^{b_0}\sum_{b_A=0}^{b_{A-1}}\left(\frac{x-A}{x-B}\right)^{b_A} \tag{S.78}$$

In other words, $\mathbf{f}\left(A,B\right)$ is a set of A nested sums, each of the same form, except for the final sum, which can have a different denominator. Using this definition, we have

$$P(\pi_n=X|k',k,\ell)=\frac{\binom{k'+k}{\pi_d}}{\binom{\frac{2U_n}{s}+k'+k}{\pi_d}}\left(\frac{2U_n}{s}\right)^{X}\mathbf{f}\left(\pi_d-2,\pi_d-1\right). \tag{S.79}$$

The virtue of this definition is that this sum can be solved recursively. We have

$$\sum_{b_A=0}^{b_{A-1}}\left(\frac{x-A}{x-B}\right)^{b_A}=\frac{x-B}{A-B}-\frac{x-A}{A-B}\left(\frac{x-A}{x-B}\right)^{b_{A-1}}. \tag{S.80}$$

Therefore we have

$$\mathbf{f}\left(A,B\right)=\frac{x-A}{B-A}\mathbf{f}\left(A-1,B\right)-\frac{x-B}{B-A}\mathbf{f}\left(A-1,A\right). \tag{S.81}$$

Repeatedly inserting this result yields:

$$\begin{aligned}
\mathbf{f}\left(A,A+1\right)&\rightarrow\frac{(x-A)(x-A-1)}{1}\left(\frac{\mathbf{f}\left(A-1,A+1\right)}{x-A-1}-\frac{\mathbf{f}\left(A-1,A\right)}{x-A}\right)\\
\mathbf{f}\left(A,A+1\right)&\rightarrow\frac{(x-A+1)(x-A)(x-A-1)}{2}\left[\frac{\mathbf{f}\left(A-2,A+1\right)}{x-A-1}-\frac{2\mathbf{f}\left(A-2,A\right)}{x-A}+\frac{\mathbf{f}\left(A-2,A-1\right)}{x-A+1}\right]\\
&\vdots\\
\mathbf{f}\left(A,A+1\right)&\rightarrow(m+1)\binom{x-A-1+m}{m+1}\sum_{i=0}^{m}\frac{(-1)^{i+m}}{x-A-1+i}\binom{m}{i}\mathbf{f}\left(A-m,A+1-i\right).
\end{aligned} \tag{S.82}$$

Note that $\mathbf{f}(-1,B)=1/B^X$, since there are no more sums to compute. Thus, for $m=A+1$ we have

$$\mathbf{f}\left(A,A+1\right)=(A+2)\binom{x}{A+2}\sum_{i=0}^{A+1}\frac{(-1)^{i+A+1}}{(x-A-1+i)^{X+1}}\binom{A+1}{i}. \tag{S.83}$$

Relabeling the sum and taking $A=\pi_d-2$, we have

$$\mathbf{f}\left(\pi_d-2,\pi_d-1\right)=\pi_d\binom{x}{\pi_d}\sum_{i=0}^{\pi_d-1}\frac{(-1)^{i}}{(x-i)^{X+1}}\binom{\pi_d-1}{i}. \tag{S.84}$$

We can now substitute these results into our expression for π_n, to find

$$\rho_1(\pi_n = X|k', k, \ell) = \pi_d \binom{k'+k}{\pi_d} \left(\frac{2U_n}{s}\right)^X \sum_{i=0}^{\pi_d - 1} \frac{(-1)^i}{(2U_n/s + k + k' - i)^{X+1}} \binom{\pi_d - 1}{i} \quad \text{(S.85)}$$

Note, however, that this is only the distribution of neutral mutations since the first deleterious mutation out of class $k - l$. It is also possible for neutral mutations to occur prior to the coalescence event. Adding in this factor, we find

$$\begin{aligned} \rho(\pi_n = X|k', k, \ell) = \pi_d \binom{k'+k}{\pi_d} \sum_{i=0}^{\pi_d - 1} (-1)^i \binom{\pi_d - 1}{i} \quad \text{(S.86)} \\ \times \sum_{X=0}^{\pi_n} \frac{(2U_n/s)^X}{(2U_n/s + k + k' - i)^{X+1}} \left(\frac{2N_{k-l}U_n}{1 + 2N_{k-l}U_n + 2N_{k-l}s(k-l)}\right)^{\pi_n - X}. \end{aligned}$$

Rearranging this expression gives

$$\rho(\pi_n|k', k, \ell) = \sum_{i=0}^{\pi_d - 1} \pi_d (-1)^{\pi_d - i - 1} \binom{\pi_d - 1}{i} \binom{k'+k}{\pi_d} \frac{B}{A - B} \left(\frac{(\frac{2U_n}{s})^{\pi_n}}{(\frac{2U_n}{s} + B)^{\pi_n + 1}} - \frac{(\frac{2U_n}{s})^{\pi_n}}{(\frac{2U_n}{s} + A)^{\pi_n + 1}} \right), \quad \text{(S.87)}$$

where we have defind

$$A = k' + k - i, \qquad B = 2(k - \ell) + \frac{1}{Nsh_{k-l}}, \quad \text{(S.88)}$$

identical to our earlier result.