Population genetics of polymorphism and divergence in rapidly evolving populations

Matthew J. Melissa[1], Benjamin H. Good[2], Daniel S. Fisher[2], and Michael M. Desai[1*]

[1] *Department of Organismic and Evolutionary Biology, Department of Physics, Quantitative Biology Initiative, and NSF-Simons Center for Mathematical and Statistical Analysis of Biology, Harvard University*
[2] *Department of Applied Physics and Department of Bioengineering, Stanford University*
[*] *mdesai@oeb.harvard.edu*

(Dated: July 21, 2023)

In rapidly evolving populations, numerous beneficial and deleterious mutations can arise and segregate within a population at the same time. In this regime, evolutionary dynamics cannot be analyzed using traditional population genetic approaches that assume that sites evolve independently. Instead, the dynamics of many loci must be analyzed simultaneously. Recent work has made progress by first analyzing the fitness variation within a population, and then studying how individual lineages interact with this traveling fitness wave. However, these "traveling wave" models have previously been restricted to extreme cases where selection on individual mutations is either much faster or much slower than the typical coalescent timescale T_c. In this work, we show how the traveling wave framework can be extended to intermediate regimes in which the scaled fitness effects of mutations ($T_c s$) are neither large nor small compared to one. This enables us to describe the dynamics of populations subject to a wide range of fitness effects, and in particular, in cases where it is not immediately clear which mutations are most important in shaping the dynamics and statistics of genetic diversity. We use this approach to derive new expressions for the fixation probabilities and site frequency spectra of mutations as a function of their scaled fitness effects, along with related results for the coalescent timescale T_c and the rate of adaptation or Muller's ratchet. We find that competition between linked mutations can have a dramatic impact on the proportions of neutral and selected polymorphisms, which is not simply summarized by the scaled selection coefficient $T_c s$. We conclude by discussing the implications of these results for population genetic inferences.

I. INTRODUCTION

Models of evolutionary dynamics and population genetics aim to predict how observable properties of the distribution of genotypes within a population are shaped by evolutionary forces such as mutations, natural selection, and genetic drift. We focus here on the simplest possible models of these three core evolutionary forces, which consider a population consisting of N individuals, with each individual subject to new mutations at rate U per generation, and with the fitness effect of each mutation drawn from some distribution $\rho(s)$. More complicated models of evolutionary dynamics can incorporate the effects of additional evolutionary forces (e.g., spatial structure and migration, environmental fluctuations in time or space, and ecological interactions). However, the dynamics arising from even the simplest models—incorporating only the forces of mutation, selection, genetic drift—are already rather complex.

Within the context of these simple models, we wish to understand how several types of observables depend on the key parameters. Specifically, we aim to predict the probability, $p_{\text{fix}}(s)$, that a mutation with fitness effect s will fix. From this quantity, we can compute the rate at which mutations will accumulate (i.e. the rate of genotypic divergence from an ancestor) and the rate, v, at which the mean fitness of the population changes over time (McCandlish and Stoltzfus, 2014). In addition to these properties of long-term divergence, we wish to predict expected patterns of genetic diversity within the population at any given time. For example, we aim to compute the coalescence timescale T_c (the typical time since individuals are related by common ancestry, which determines the expected overall level of genetic diversity in the population), as well as other readily observable quantities such as the site frequency spectrum (which describes the relative abundances of polymorphisms at different allele frequencies). Obtaining predictions for patterns of genetic diversity is of particular interest because of the potential to use these predictions, in combination with measured levels of contemporary sequence diversity, to infer the strength of the various evolutionary forces which have acted on a population.

Much classical work proceeds by assuming the different loci in the genome evolve completely indepen-

dently of one another (i.e., by ignoring the physical *linkage* of mutations within the same genomic segment). The dynamics are then relatively straightforward to analyze (Fisher, 1931; Wright, 1931). This assumption is reasonable in sexually evolving populations where recombination is sufficiently rapid compared to the timescales on which mutation, selection and genetic drift generate correlations between loci (Falconer and Mackay, 1996). In this case, a given mutation is shuffled onto a large number of genetic backgrounds over the course of its evolutionary trajectory, and any correlations in the fates of mutations at different loci become negligible.

In asexual populations, however, mutations are fully linked and do not evolve independently of one another. Even in the genomes of obligately sexual organisms, loci that are physically close within a chromosome will be broken up by recombination only over very long timescales, and cannot necessarily be treated as independently evolving (Weissman and Barton, 2012). A useful opposite regime to consider is then the limit of a genomic region that is small enough that recombination can be entirely neglected. One can imagine that on sufficiently short genomic distance scales (within so-called "linkage blocks"), evolution proceeds entirely asexually, while on larger genomic distance scales evolution acts on some collection of recombining blocks. It is not entirely clear how this division between asexual linkage blocks and larger-scale regions consisting of multiple recombining blocks is best modeled, though some recent work has begun to address this question (Good et al., 2014; Neher et al., 2013; Weissman and Hallatschek, 2014). However, we focus in this paper on an even simpler question: how to model the evolution within a single nonrecombining linkage block. As we will see, even within the context of the simple models described above, evolutionary dynamics even within perfectly asexual linkage blocks are surprisingly complex, and remain incompletely understood.

There is a long history of efforts to analyze the dynamics of purely asexual evolution, and numerous approaches have been proposed to analyze the dynamics in different limiting regimes of the parameter space. Extensive work has focused on the *neutral limit*, in which natural selection can be neglected entirely (Kimura, 1968). In this case, genealogical approaches such as coalescent theory can provide a complete description of the expected rates of divergence and patterns of diversity (Kingman, 1982; Wakeley, 2005). However, the backwards-time nature of the coalescent approach makes it difficult to incorporate the effects of selection: genealogies cannot be considered independently of the selected mutations which occur on their branches (Kaplan et al., 1988). Efforts to do so have largely been limited to simulation-based or essentially numerical approaches (see e.g., Krone and Neuhauser (1997)), though analytical progress has been made in certain cases, particularly in the presence of purifying selection on deleterious mutations (Charlesworth, 1994; Hudson and Kaplan, 1995).

To model the effects of selection on beneficial mutations, much work has instead been done using forward-time approaches. Broadly speaking, these approaches seek to characterize the trajectories of mutant lineages in a probabilistic sense. Provided selection is sufficiently strong and selected mutations are sufficiently rare (more precisely, if $Ns \gg 1$ and $NU \log Ns < 1$), a beneficial mutation typically either sweeps to fixation or is purged before another such mutation arises within the same linkage block (Desai and Fisher, 2007). Thus, within this *strong-selection weak-mutation* (SSWM) regime, multiple selected mutations are unlikely to segregate simultaneously within the block. In this case, the dynamics of each selected mutation can be treated independently of one another (Gillespie, 1983). However, recent work has shown that in a wide range of microbial and viral populations, and potentially in many linked regions of the genomes of obligately sexual organisms such as humans, multiple beneficial mutations often segregate simultaneously (Lieberman et al., 2014; Miralles et al., 1999; Nourmohammad et al., 2019; Strelkowa and Lässig, 2012). In these *rapidly adapting* populations, clonal interference (i.e. competition between multiple distinct adaptive lineages) and genetic hitchhiking can be critical to the dynamics, and analyzing the evolution of multiple linked loci simultaneously is critical (Buskirk et al., 2017; Gerrish and Lenski, 1998; Kao and Sherlock, 2008; Lang et al., 2013).

Over the past two decades, numerous authors have analyzed the evolutionary dynamics of many linked loci in rapidly adapting populations (reviewed by Neher (2013)). Similar ideas have also been used to analyze the dynamics of populations rapidly *declining* in fitness due to Muller's ratchet (Neher and Shraiman, 2012), as well as those maintained in a dynamic steady-state balance between beneficial and deleterious mutations (Goyal et al., 2012). Collectively, we can think of this body of work as studying the evolutionary dynamics of rapidly *evolving* populations—those populations in which there are typically multiple linked mutations, either beneficial or deleterious, in the population at once. The central idea of this work is to first analyze how the collective effects of many linked mutations generate vari-

ation in fitness within the population. This leads to a time-dependent within-population fitness distribution that can be described as a *traveling wave*, and which maintains a constant steady-state shape $f(x)$ while its mean fitness changes at a constant rate v (potentially with fluctuations) (Rouzine et al., 2008, 2003; Tsimring et al., 1996). Extensive work has characterized how the velocity and steady-state shape of the wave depends on the key parameters, which can be used to compute observables related to long-term divergence such as $p_{\text{fix}}(s)$ or the distribution $\rho_f(s)$ of *fixed* fitness effects (Good et al., 2012). More recent studies have then used this traveling wave picture as the basis for tracing the genealogical history of individual mutations (Desai et al., 2013; Neher and Hallatschek, 2013) or tracking their trajectories forward in time (Kosheleva and Desai, 2013) to calculate diversity statistics such as the coalescence timescale or the site frequency spectrum.

While this work has led to substantial progress, it has been focused primarily on two limiting cases. One body of work (Cohen et al., 2005; Hallatschek, 2011; Neher and Hallatschek, 2013; Tsimring et al., 1996) has analyzed the *infinitesimal limit*, in which the selective effects of individual mutations are infinitesimal, but selected mutations occur extremely frequently (more precisely, in which $s \to 0$ and $U \to \infty$ with the product $U\left\langle s^2\right\rangle$ held fixed). The corresponding *infinitesimal approximation* is often thought to be valid when $s \ll U$, although a more accurate condition of its validity is that $T_c s \ll 1$ (Good and Desai, 2014; Good et al., 2014); in both conditions, the relevant s can be taken as the root-mean-squared effect $\sqrt{\langle s^2\rangle}$. As a result, within the infinitesimal regime, selection on individual mutations can be neglected: the timescale $1/s$ on which selection can substantially alter the fate of a mutation with effect size s is longer than the coalescent timescale T_c on which common ancestry is determined (and which we define more precisely in Section IV). At the same time, the population as a whole maintains substantial variance in fitness, σ^2, resembling in some ways infinitesimal models in quantitative genetics (Barton et al., 2017); selection on *haplotypes* in the infinitesimal regime can be strong. This work has led to analytical results for both divergence-related quantities and, more recently, diversity-related quantities (Neher and Hallatschek, 2013), valid for populations subject to beneficial mutations, to deleterious mutations, or to some combination of the two.

A second body of work has analyzed the opposite limit of *strong* selection on individual mutations, such that $T_c\tilde{s} \gg 1$. Here $\tilde{s}$ denotes the typical effect size of mutations which fix; $\tilde{s}$ is defined more precisely in Section IV. A schematic depiction of $\tilde{s}$ is provided in Fig. 1, which illustrates key qualitative differences between regimes. When selection is strong on deleterious mutations, Muller's ratchet clicks slowly and beneficial mutations can much more easily dominate the dynamics. While some work has been done on the rate of Muller's ratchet when $T_c\tilde{s} \gg 1$ (Neher and Shraiman, 2012), the majority of this work has focused exclusively on beneficial mutations (Desai and Fisher, 2007; Desai et al., 2013; Fisher, 2013; Good et al., 2012; Kosheleva and Desai, 2013) or on the case in which deleterious mutations can be considered a perturbative correction (Good and Desai, 2014). This work can be further divided into two regimes: the "moderate-speeds" regime, in which $v \ll \tilde{s}^2$, as well as the "high-speeds" regime, in which $v \gg \tilde{s}^2$ (Fisher, 2013). Qualitatively, the "moderate-speeds" and "high-speeds" regimes can be distinguished based on the range of background fitnesses Δx_f which typically produces an eventual common ancestor of the population (see Fig. 1 for a depiction of Δx_f). In the "moderate-speeds" regime, $\Delta x_f \sim \tilde{s}$, while in the "high-speeds" regimes, $\Delta x_f \gg \tilde{s}$. That is, within the "moderate-speeds" regime, the individual which will eventually fix is likely within one mutational effect $\tilde{s}$ of the "nose" (the high-fitness edge) of the fitness distribution; in the "high-speeds" regime, individuals can catch up and fix from further behind by rapidly acquiring multiple mutations. Analytical results for divergence-related quantities have been obtained within both the "moderate-speeds" and "high-speeds" regimes (Fisher, 2013; Good and Desai, 2014); diversity-related quantities have also been studied within the "moderate-speeds" regime (Desai et al., 2013; Kosheleva and Desai, 2013).

Although the parameters N, U and s are natural quantities to use in specifying a model of the evolutionary dynamics, in many applied settings the combined quantities $T_c s$ and $T_c U$ can be probed more directly. While dynamical interpretations of $T_c s$ and $T_c U$ are less immediately clear, these can be considered independent properties of the evolution that together can describe many aspects of a population. Existing methods to infer $T_c s$ from observed levels of polymorphism and divergence among populations typically make the assumption that different loci evolve *independently* of one another (Bustamante et al., 2001; Sawyer and Hartl, 1992). For example, a classic result shows that in a population with a constant size N, the ratio between selected

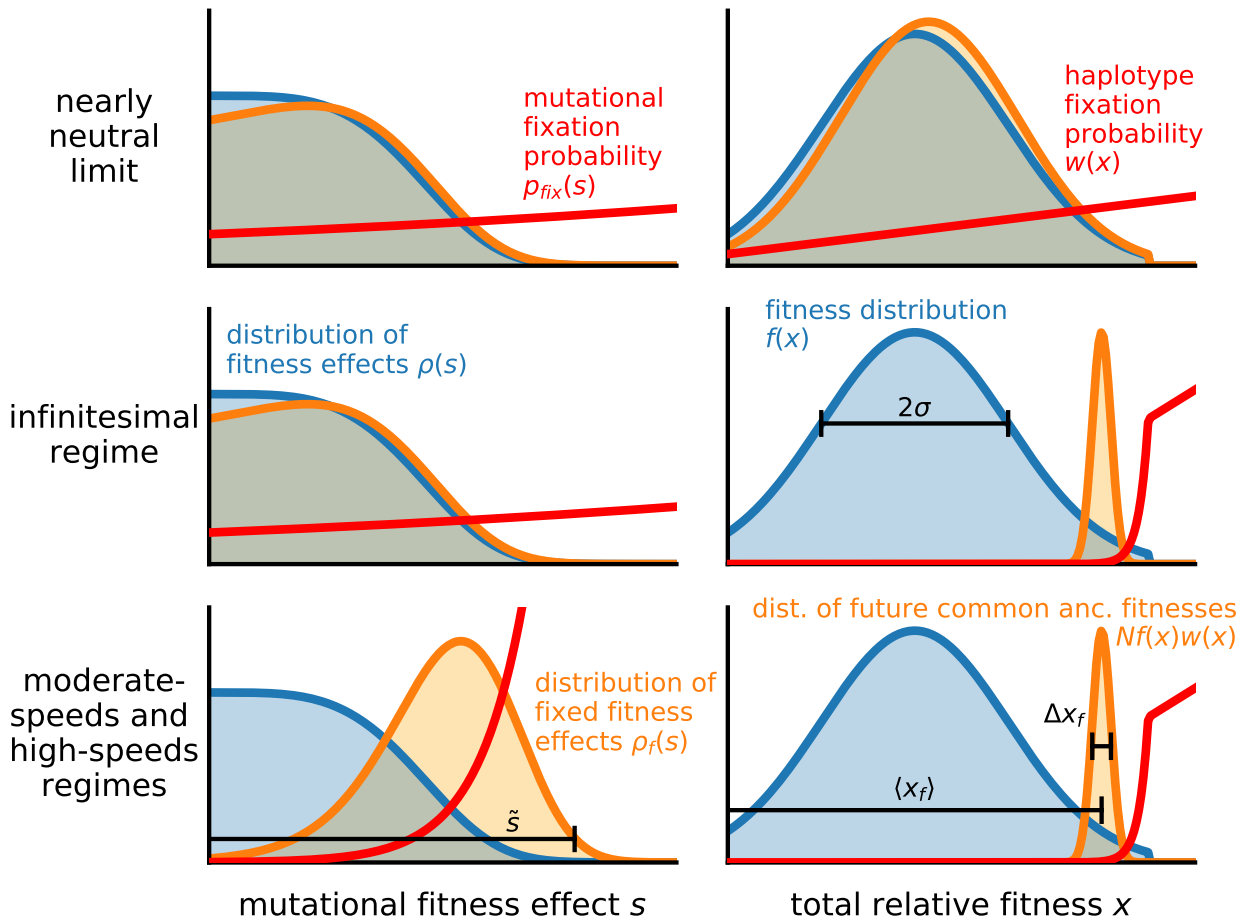

FIG. 1 Schematic depiction of different regimes. (Left) Schematic plots of the mutation fixation probability $p_{\text{fix}}(s)$, an example distribution of fitness effects $\rho(s)$, and its corresponding distribution of fixed fitness effects $\rho_f(s)$, for regimes of the parameter space indicated by row labels. In the nearly neutral limit and in the infinitesimal regime, selection acts weakly on single mutations, and $\rho_f(s)$ closely matches $\rho(s)$. In the "moderate-speeds" and "high-speeds" regimes, $\rho_f(s)$ differs substantially from $\rho(s)$. Note that $\tilde{s}$ in the bottom panel denotes the scale of a typical fixed mutational effect. (Right) Schematic plots of the haplotype fixation probability $w(x)$, the fitness distribution $f(x)$ and the distribution of future common ancestor fitnesses $Nf(x)w(x)$. In the infinitesimal regime as well as the "moderate-speeds" and "high-speeds" regimes, selection acts strongly on haplotypes, and the distribution $Nf(x)w(x)$ is concentrated within a narrow range of high background fitnesses. The width and average of the distribution $Nf(x)w(x)$ are indicated by Δx_f and $\langle x_f \rangle$, respectively. Both the infinitesimal regime and the "high-speeds" regime, but not the "moderate-speeds" regime, are part of the MSSM regime; $\Delta x_f \gg \tilde{s}$ in the infinitesimal regime and "high-speeds" regime, but $\Delta x_f \sim \tilde{s}$ in the "moderate-speeds" regime. In the MSSM regime, Δx_f and $\langle x_f \rangle$ will be seen to correspond roughly to the quantities b and c, respectively, defined in Section IV.

and neutral site frequency spectra is given by

$$\frac{h_s(\nu)}{h_0(\nu)} = \frac{1 - e^{-2T_c s(1-\nu)}}{(1-\nu)(1 - e^{-2T_c s})} \tag{1}$$

$$\approx \begin{cases} e^{-2T_c|s|\nu} & \text{if } s < 0;\ T_c|s| \gg 1 \\ 1 & \text{if } T_c|s| \ll 1 \\ \frac{1-e^{-2T_c s(1-\nu)}}{1-\nu} & \text{if } s > 0;\ T_c s \gg 1 \end{cases} \tag{2}$$

where $T_c = N$ is the coalescent timescale (Sawyer and Hartl, 1992). Using this result, the magnitudes of the scaled selection coefficients ($T_c s$) can be identified from the transitions in the site frequency spectra at low frequencies ($\nu_c \sim 1/T_c|s|$) for deleterious mutations, or at high frequencies ($1 - \nu_c \sim 1/T_c s$) for beneficial mutations.

However, previous work has shown that these classical results can lead to substantial errors in the inferred selection strengths when natural selection is widespread (Messer and Petrov, 2013). This makes it difficult to identify the relevant values of $T_c s$ in rapidly evolving populations. For deleterious muta-

tions, the intermediate regime where $T_c s \sim 1$ is particularly critical: it is precisely mutations with effect sizes on the order of $1/T_c$ which are expected to have the largest impact on patterns of genetic diversity within a population (Good et al., 2014). Thus when there is a relatively broad distribution of effect sizes of new mutations (including strong, intermediate, and weak-effect mutations), it may often be the case that the intermediate-effect mutations with $T_c s \sim 1$ have the largest impact on observed patterns of diversity. More generally, an understanding of the intermediate regime in which $T_c s \sim 1$ is critical for understanding the transition between strong selection and neutral-like behavior in site frequency spectra. Our understanding of these effects is limited, however: except for special cases—such as when a single strong beneficial fitness effect is available (Kosheleva and Desai, 2013)—we lack analytical predictions for the diversity statistics of selected mutations in rapidly evolving populations. In particular, we lack predictions for how mutations with a distribution of beneficial and/or deleterious selective effects contribute to the site frequency spectrum, which has in turn limited our ability to apply commonly used inference methods (e.g. of Hartl et al. (1994); Sawyer and Hartl (1992)) to rapidly evolving populations.

To address this gap, in this work we reexamine the approach used by Fisher (2013) to study the "high-speeds" regime. We demonstrate that the key ideas of that approach can be applied more generally to the case in which $T_c \tilde{s} \ll 1$ (as well as the intermediate regime in which $T_c \tilde{s} \sim 1$). As a result, we argue that the infinitesimal regime and the "high-speeds" regime—previously studied separately—can be unified into a single *moderate selection, strong mutation* (MSSM) regime, which includes populations subject to a distribution of beneficial mutations, deleterious mutations, or some combination of the two. The key requirements for validity of our MSSM approximation are that $\tilde{s} \ll \Delta x_f$ (which ensures that selection is moderate, or weak, on single mutations) and that $T_c \Delta x_f \gg 1$ (which ensures that mutation is strong, and that selection on haplotypes is strong). Dynamically, these conditions imply that mutational "leapfrogging" is important to the dynamics: *some* individuals routinely catch up to the high-fitness "nose" despite an initial fitness disadvantage, relative to the nose, of several mutations. These requirements are satisfied in the limit $N \to \infty$ with $U\rho(s)$ held fixed—as long as $\rho(s)$ falls off faster than exponentially with large positive s—as well as in other cases we describe. In particular, we show that this approach can model the dynamics of deleterious mutations with $T_c \tilde{s} \sim 1$, to which patterns of diversity and divergence are particularly sensitive.

Using this approach, we compute divergence-related statistics such as the fixation probabilities $p_{\text{fix}}(s)$ of new mutations and the average rate v of adaptation, or—if $v < 0$—of Muller's ratchet. We also compute diversity-related statistics such as the coalescence timescale T_c and the site frequency spectrum. We show that at high frequencies, the neutral site frequency spectrum corresponds to that of the Bolthausen-Sznitman coalescent (BSC) (Bolthausen and Sznitman, 1998), which has previously been found to describe genealogies of populations in the infinitesimal regime (Neher and Hallatschek, 2013) and "moderate-speeds" regime (Desai et al., 2013), as well as in evolving populations modeled as FKPP waves (Brunet et al., 2007). We identify the frequency scale above which this correspondence holds and analytically describe departures from the BSC at lower frequencies. Importantly, we find that the low-frequency portion of the neutral site frequency spectrum is much more useful for distinguishing different parameter combinations than the high-frequency portion, which depends on the parameters N, U and $\rho(s)$ via a single overall scale factor. Over a broad range of intermediate and high frequencies—extending beyond the frequency range at which a correspondence with the BSC exists—we demonstrate that the neutral and selected site frequency spectra are simply proportional to one another, with a constant of proportionality equal to the ratio of fixation rates of the two types of mutations. This proportionality reflects the fact that, in the presence of widespread linkage, the fates of even strongly selected mutations can be considered *conditionally neutral* at long times $t > T_c$: at this point, their initial fitness effects have been "forgotten", and their lineage trajectories are indistinguishable from those of neutral mutations with the same age (and frequency at age T_c). We discuss implications of our results for inferring the strength and/or frequency of selection in natural populations using observed levels of polymorphism and divergence.

Outline of this Paper: The remainder of this work is organized as follows. In Section II, we describe the model of the population dynamics and briefly summarize our simulation methods. In Section III, we review previous approaches to model evolutionary dynamics using traveling wave theory. We begin Section IV by presenting our MSSM approximation, which we use to obtain results for the steady-state distribution of fitnesses, fixation probabilities, and the rate of adaptation. We then provide and discuss conditions of validity of the MSSM approximation, apply the MSSM approximation to

several specific examples of DFEs, and compare our analytical results to the results of simulations. In Section V, we use our MSSM approximation to analyze statistics of genetic diversity, with particular focus on the site frequency spectra of neutral mutations and of selected (non-neutral) mutations. In Section VI, we provide heuristic interpretations of several fitness scales and timescales that emerge in our analysis, and compare dynamical aspects of populations described by the MSSM approximation to those of populations in other regimes. In the Discussion, we consider implications of our results for making inferences from population genetic data.

II. MODEL

We aim to understand the evolutionary dynamics of an asexual population, given the simplest possible model incorporating the effects of mutations, natural selection, and genetic drift. To this end, we consider a population of a constant number N asexual haploid individuals, within which new mutations arise at a rate U per genome per generation. We assume that all mutations have additive effects on (log) fitness, X (i.e., that the log fitness of each individual is the sum of the fitness effects of all the mutations it or its ancestors have acquired). An individual's offspring number distribution is determined by its relative fitness $x \equiv X - \bar{X}$, where by $\bar{X}$ we denote the population-wide mean fitness. We write down the particular stochastic birth/death process we carry out in simulations in Subsection II.A; in our analytical treatment, we assume a slightly different birth/death process, which we write down in Section III. We assume the fitness effect of each new mutation is drawn at random from a time-independent distribution of fitness effects (DFE), $\rho(s)$. Our model thus assumes that while *epistasis* may exist among individual mutations, there are no overall differences in the DFE among different genotypes, so we can treat $\rho(s)$ as a constant that remains the same as the population evolves. Because most dependence on the parameters U and $\rho(s)$ will be mediated by the product $\mu(s)$, we define $\mu(s) \equiv U\rho(s)$ as the *mutational fitness spectrum*. We assume neutral mutations occur at rate U_n (not included in U).

The actual DFEs relevant to natural populations may be broad and complex, with mutations conferring a wide range of fitness effects that occur at a variety of different rates, and these empirical distributions are difficult to measure precisely (Eyre-Walker and Keightley, 2007). Much of our analysis is conducted for an arbitrary DFE (provided it meets our conditions of validity). For concreteness, we also focus on a few simplified forms of $\mu(s)$ in order to gain intuition. For example, we consider the cases in which (i) all mutations are beneficial, (ii) all mutations are deleterious, and (iii) all mutations confer the same effect size. The analysis of these simplified forms of $\mu(s)$ can provide useful intuition about how mutations of different types and effect sizes affects various aspects of the evolutionary dynamics. Further, we find that most of our results are sensitive to the assumed $\mu(s)$ only within a limited region of s. That is, given values of the other parameters, the evolutionary dynamics are often dominated by a subset of mutations that have a narrow range of effect sizes, and hence can be predicted based on simplified forms of $\mu(s)$ (Hegreness et al., 2006).

A. Simulation methods

To test our analytical predictions, we conduct individual-based Wright-Fisher simulations. These simulations, which we describe in more detail in Appendix H, separately track all of the individuals in the population, as well as the mutations they have acquired. These simulations make it possible to measure divergence-related and diversity-related quantities in the population, including the rate of change in the mean fitness, the rate of mutation accumulation, the heterozygosity of neutral mutations and of selected (i.e. non-neutral) mutations, and the site frequency spectrum.

Simulations consist of a series of two steps repeated each generation. First, each individual acquires a Poisson-distributed number of mutations (with mean number U, and with the fitness effects of those mutations each drawn from $\rho(s)$). Mutations increment an individual's log-fitness X by an amount s. Second, in the selection/reproduction step, individuals in the population are resampled (with replacement) to form the population in the following generation. Each individual is sampled with with probability proportional to its (exponential) fitness e^X.

Each simulated population is initialized clonally. We record the number of generations until a mutation has fixed within the population, which we define as the length of an *epoch*. At the conclusion of each subsequent epoch, we record the state of the population, including its mean fitness, site frequency spectrum, and fixed mutations. Simulations are run for a prespecified number of epochs, with results from the first 10 epochs discarded.

Simulated populations are subject to stretched exponential DFEs of the form

$$\rho(s) = \frac{1}{\chi\Gamma(1+\beta^{-1})} e^{-(s/\chi)^{\beta}}, \tag{3}$$

or to mutations consisting of a single fitness effect (while our analysis applies for more general $\rho(s)$). DFEs of the form in Eq. (3) have been considered in previous theoretical work (Desai and Fisher, 2007; Fogle et al., 2008; Good et al., 2012). Note that if $\beta > 1$, $\rho(s)$ in Eq. (3) falls off faster than exponentially with large s, while $\rho(s)$ is an exponential distribution if $\beta = 1$. For $\beta < 1$, $\rho(s)$ in Eq. (3) falls off more slowly than exponentially with large s (though for reasons we describe below, we do not consider this case extensively here).

III. THE TRAVELING-WAVE APPROACH

In this Section, we review the basic traveling-wave approach which underlies our work, and which has been used to study the dynamics of rapidly evolving populations in multiple regimes of the parameter space. Readers familiar with the traveling-wave literature may wish to skip directly to the Analysis section. We emphasize that the details of the traveling-wave approach have differed across studies—particularly in the treatment of fluctuations—and that our review is not entirely comprehensive. Our presentation, which largely follows the approach of Good and Desai (2014), neglects some of the fluctuations which are discussed at length in Fisher (2013), but enables us to streamline computations of average quantities such as v, $p_{\text{fix}}(s)$, T_c and the site frequency spectrum.

A key quantity of interest is the distribution of fitnesses within the population, $f(X,t)$, which gives the fraction of the population at absolute (log) fitness X. In our model, this fluctuating distribution evolves in time according to the nonlinear stochastic differential equation (Good and Desai, 2013),

$$\begin{aligned}\frac{\partial f(X,t)}{\partial t} &= \left(X - \bar{X}(t)\right) f(X,t) \\ &\quad + \int \mu(s)\left[f(X-s,t) - f(X,t)\right] ds \\ &+ \int dX' \left[\delta(X'-X) - f(X,t)\right] \sqrt{\frac{f(X',t)}{N}}\eta(X',t),\end{aligned} \tag{4}$$

where $\bar{X}(t) \equiv \int X f(X,t) dX$, and $\eta(X,t)$ is a Brownian noise term (Appendix A).

The last term in Eq. (4), which captures birth/death number fluctuations, is important. Without this term, even in the simple case of a single beneficial fitness effect, the solution $f(X,t)$ would have a rate of fitness increase $d\bar{X}/dt$ that grows without bound (Tsimring et al., 1996). In contrast, in stochastic simulations, Tsimring et al. (1996), Rouzine et al. (2003) and others have found that after an initial transient period, the distribution of fitnesses in a population attains a steady state "traveling wave" profile, $f(X-vt)$ which moves through fitness space at a roughly constant rate $v = \langle d\bar{X}/dt\rangle$; the steady state can be understood as a type of "mutation-selection balance" in which fitness variation, purged by selection, is repleted by new mutations (Desai and Fisher, 2007). In large populations, this average (or more precisely, *typical*) profile $f(x)$ can then be approximated using a "quasi-deterministic" approach, in which

$$\begin{aligned}-v\partial_x f(x) &= x f(x) \\ &\quad + \int \mu(s)\left[f(x-s) - f(x)\right] ds,\end{aligned} \tag{5}$$

for x less than some cutoff x_{cut}, and $f(x) = 0$ for $x > x_{cut}$ (Fisher, 2013; Good et al., 2012; Neher et al., 2010; Rouzine et al., 2008; Rouzine and Coffin, 2005; Rouzine et al., 2003; Tsimring et al., 1996). This defines a distribution which is normalized such that

$$\int f(x)dx = 1. \tag{6}$$

The solution to Eq. (5) will depend on the unknown average rate of mean fitness change v. To determine this quantity, we can express v in terms of the stochastic accumulation of new mutations,

$$v = N \int \mu(s) s p_{\text{fix}}(s) ds, \tag{7}$$

where $p_{\text{fix}}(s)$ describes the fixation probability of a mutation with fitness effect s. In our traveling wave framework, it is useful to express this fixation probability as

$$p_{\text{fix}}(s) = \int f(x) w(x+s) dx, \tag{8}$$

where $w(x)$ denotes the fixation probability of a lineage founded by a single individual with relative fitness x. In this way, Eq. (8) averages over all of the possible fitness backgrounds on which a mutation can occur.

A key simplification is that $w(x)$ can often be approximated by modeling the population as a collection of independent branching processes, which compete with each other only through the average rate of fitness change v (Fisher, 2013; Good and Desai, 2014; Good et al., 2012; Neher et al., 2010). Each of these lineages founds its own stochastic fitness wave, $g(x,t)$, whose dynamics are described by a related differential equation,

$$-\partial_t \phi = x\phi - \phi^2 + \int \mu(s)\left[\phi(x+s) - \phi(x)\right] \quad (9)$$

for the generating functional $\langle e^{-N\int \phi(x',0)g(x',t)dx'}\rangle = e^{-\phi(x,t)}$ (Appendix A). The fixation probability $w(x)$ then follows from

$$v\partial_x w(x) \approx xw(x) + \int \mu(s)\left[w(x+s) - w(x)\right] ds - w^2(x). \quad (10)$$

Together, Equations (5) and (10) can be solved given a particular $\mu(s)$ and an assumed rate of adaptation v. Using these solutions $f(x)$ and $w(x)$, Eq. (7) can be enforced as a self-consistency relation to determine v in terms of N and $\mu(s)$. In practice, it will be useful to enforce a related self-consistency condition,

$$p_{\text{fix}}(0) = \int f(x)w(x)dx \approx 1/N, \quad (11)$$

which ensures that neutral mutations fix with unbiased probability $1/N$. Approximating the solutions to Eq. (5) and Eq. (10), or similar equations, given the parameters N and $\mu(s)$, has thus been the subject of extensive theoretical work (Fisher, 2013; Good and Desai, 2014; Good et al., 2012; Hallatschek, 2011; Neher et al., 2010).

A central complication is that Eq. (10) is nonlinear. Furthermore, the nonlinear w^2 term is the only term in Eq. (10) which reflects the stochasticity of births and deaths. As a result, dropping the w^2 term in Eq. (10) essentially amounts to a deterministic approximation, and leads to fundamentally wrong predictions. However, a *dominant balance* approach can be taken to treat this nonlinearity (Fisher, 2013; Good and Desai, 2014; Good et al., 2012; Neher et al., 2010). The basic idea is that because the fixation probability w must be an increasing function of x, at very large relative fitness (i.e. large x), xw and w^2 are the dominant terms in Eq. (10). Thus at large x we have $w(x) \approx x$. In contrast, for sufficiently small fitness, the nonlinear term must be negligible compared to the other terms. Thus at these small x we can neglect the w^2 term in Eq. (10) and solve the linear equation

$$v\partial_x w(x) = xw(x) + \int \mu(s)\left[w(x+s) - w(x)\right] ds. \quad (12)$$

An approximation where the large-x and small-x solutions are simply *patched* together is then widely valid (Fisher, 2013; Good and Desai, 2014; Good et al., 2012). Specifically, a boundary x_c is identified such that the large-x result ($w(x) = x$) is valid for $x > x_c$ and the small-x result (the solution of Eq. (12)) is valid for $x < x_c$, without any intervening shoulder region (or more precisely, with a shoulder region that is narrow on relevant scales). A schematic of this patched solution $w(x)$ is depicted in Fig. 1. The details of how to determine the boundary x_c can still be quite complicated and case-

TABLE I Notation Used

	Symbol	Quantity
Input Parameters	N	Population size
	U	Mutation rate
	$\rho(s)$	Distribution of fitness effects
	$\mu(s) \equiv U\rho(s)$	Mutational fitness spectrum
Traveling Wave Quantities	$x \equiv X - \bar{X}$	Relative fitness
	$f(x)$	Distribution of relative fitnesses
	$w(x)$	Haplotype fixation probability
	$v \equiv d\bar{X}/dt$	Rate of change of mean fitness
	σ^2	Population-wide fitness variance
	$p_{\text{fix}}(s)$	Mutational fixation probability
	$\rho_f(s)$	Distribution of fixed fitness effects
	$\langle s_f \rangle$	Average fixed fitness effect
	Δs_f	Standard deviation in fixed fitness effects
	$\langle x_f \rangle$	Average fixed relative fitness
	Δx_f	Standard deviation in fixed relative fitnesses
Diversity and Divergence Statistics	F	Fixation rate (of non-neutral mutations)
	T_2	(Random) time since pairwise coalescence
	π_{neu}	Pairwise heterozygosity (of neutral mutations)
	π_{sel}	Pairwise heterozygosity (of selected mutations)
	$H_{\text{neu}}(\nu)$	Distribution of allele frequencies (of neutral mutations)
	$H_{\text{sel}}(\nu)$	Distribution of allele frequencies (of selected mutations)

dependent (Fisher, 2013); however, in the regime of interest to us, a simple heuristic—ensuring that both $w(x)$ and its derivative are continuous at the boundary x_c—is adequate (Good and Desai, 2014). The boundary x_c is often referred to as the *interference threshold*, since individuals with relative fitness $x > x_c$ fix with probabilities largely unaffected by interference (i.e. with their establishment probabilities $\approx x$).

We note that solving Eq. (5) leads to negative values for $f(x)$ at sufficiently large x (Fisher, 2013; Rouzine et al., 2008; Tsimring et al., 1996). This is an artifact of neglecting the fact that a finite set of fitnesses are represented in a population at any given time. To avoid this pathology, a common approach is to implement a "cutoff" by assuming that $f(x) = 0$ for $x > x_{cut}$ (Rouzine et al., 2008, 2003; Tsimring et al., 1996). The value x_{cut} can then be interpreted as a typical maximum relative fitness of the individuals in a population, or as the fitness advantage of the "nose" of the population-wide fitness distribution. Consistent with Good and Desai (2014), Fisher (2013), and others, we take $x_{cut} = x_c$ throughout. At a heuristic level, this choice can motivated by the fact that established lineages cannot typically exist above the interference threshold (if so, they would interfere); a more rigorous justification is given by Fisher (2013). Imposing a cutoff in $f(x)$ can also be motivated using the tunable constraint framework introduced in Hallatschek (2011). In that work, the underlying model is changed in such a way to yield moment closure of the equations for the stochastic time-dependent fitness distribution; the resulting analog of $f(x)$ does not display the above-mentioned pathological behavior, and instead exhibits an exponential decay beyond a fitness scale related to x_c (suggesting that imposing a cutoff at x_c is reasonable). We note further that the "tunable constraint" imposed by Hallatschek (2011) turns out to be equivalent to the requirement $p_{\text{fix}}(0) = 1/N$ that we enforce to ensure a self-consistent rate of adaptation v (Good et al., 2012). Ultimately, our results are relatively insensitive to whether a cutoff is taken in $f(x)$ or a tunable constraint model is assumed.

Analysis of the linearized equations (5) and (12) is still not straightforward, however, since both equations contain mutation terms which are *nonlocal*. While exact solutions for suitably defined Laplace transforms $\tilde{f}(z)$ and $\tilde{w}(z)$ can be obtained (Fisher, 2013), the subsequent inversion of these transforms requires approximation. A related strategy has been to approximate Eq. (5) and Eq. (12) by a set of *local* differential equations, which can then be solved straightforwardly. For example, in the infinitesimal regime, the relevant fitness effects s are assumed sufficiently small to approximate

$$f(x-s) \approx f(x) - s\partial_x f(x) + \frac{1}{2}s^2\partial_x^2 f(x), \quad (13)$$

and

$$w(x+s) \approx w(x) + s\partial_x w(x) + \frac{1}{2}s^2\partial_x^2 w(x), \quad (14)$$

in Eq. (5) and Eq. (12), respectively. The resulting $f(x)$ and $w(x)$ can be obtained using the Airy equation (Cohen et al., 2005; Hallatschek, 2011; Tsimring et al., 1996). Rouzine et al. (2003) and Rouzine et al. (2008) employ a similar approach, instead Taylor approximating the logarithm of $f(x-s)$.

Our focus in this article is on a related approach to approximate Eq. (5) and Eq. (12) by a set of local equations, used by Fisher (2013) to study the "high-speeds" regime. In Section IV, we begin by reviewing this approximate calculation of $f(x)$ and $w(x)$. We proceed to provide novel solutions to a number of questions within a *moderate selection, strong mutation* (MSSM) regime, within which this approximate calculation is valid, and which includes both the "high-speeds" regime and the infinitesimal regime as special cases. We first solve for divergence-related quantities including fixation probabilities $p_{\text{fix}}(s)$ and the average rate of fitness change v, and discuss conditions of validity of our approach that are obtained by ensuring $f(x)$ and $w(x)$ are well-approximated in the region of x making a dominant contribution to these quantities. We then consider statistics of genetic diversity within the MSSM regime, with a focus on the site frequency spectra of selected mutations and of neutral mutations. By calculating the neutral site frequency spectrum, we demonstrate a partial correspondence between genealogies within the MSSM regime and those of the Bolthausen-Sznitman coalescent (BSC), and analytically describe departures from the BSC that are apparent in the low-frequency portion of the site frequency spectrum. Finally, we conclude Section IV by discussing the various fitness and time scales that emerge in our analysis, and provide a heuristic description of the dynamics of lineages within the MSSM regime.

We summarize some of our key notation used throughout in Table I .

IV. ANALYSIS

As noted above, our approach is based on a key approximation employed by Fisher (2013) to study

TABLE II Key Definitions

Quantity	Definition	Interpretation
T_c	$v \equiv \int \mu(s) e^{T_c s} s ds$	Log-slope of $f(x)$ near $x = c$. Corresponds with sweep timescale and coalescence timescale $\langle T_2 \rangle /2$.
b	$2b^3 \equiv \int \mu(s) s^2 e^{T_c s} ds$	Width in distribution of fixed relative fitnesses $\sim \Delta x_f$.
c	$c \equiv \int \mu(s) \left[T_c s e^{T_c s} + 1 - e^{T_c s}\right] ds$	Average background fitness of fixed individuals $\sim \langle x_f \rangle$.
D	$2D \equiv \int \mu(s) s^2 ds$	Effective diffusion constant in infinitesimal regime.
$\langle s_f \rangle$	$\langle s_f \rangle \equiv \int \rho_f(s) s ds$	Average fixed fitness effect.
Δs_f	$(\Delta s_f)^2 \equiv \int \rho_f(s) \left(s - \langle s_f \rangle\right)^2 ds$	Standard deviation in fixed fitness effects.
$\tilde{s}$	$\tilde{s} \equiv \langle s_f \rangle + 2\Delta s_f$	Maximum typical fixed fitness effect.
s^*	$s^* \equiv \mathrm{argmax}_s\, \rho_f(s)$	Most likely fixed fitness effect; "predominant-s".
$\psi(x)$	$f(x) \equiv e^{-T_c x} \psi(x)$	Scaled profile of $f(x)$.
$\xi(x)$	$w(x) \equiv e^{T_c x} \xi(x)$	Scaled profile of $w(x)$.

the "high-speeds" regime. The key idea is that we can approximate $f(x-s)$ and $w(x+s)$ in Eq. (5) and Eq. (10), respectively, by first pulling out rapid exponential prefactors varying at the appropriate rate, and then performing Taylor approximations of the remaining factors. To do so, we define

$$f(x) \equiv e^{-T_c x} \psi(x), \tag{15}$$

and

$$w(x) \equiv e^{T_c x} \xi(x), \tag{16}$$

with T_c defined by

$$v \equiv \int \mu(s) s e^{T_c s} ds. \tag{17}$$

Note that T_c is not a parameter of our model; Eq. (17) uniquely defines a specific T_c given the model parameters N and $\mu(s)$. If $\mu(s)$ falls off *slower* than exponentially with s at large positive s, then T_c is not well-defined and our approach breaks down; in the marginal case of an exponential DFE of beneficial mutations (with mean effect s_b), Eq. (17) implies that $T_c s_b < 1$. We will later show in Section V that, within the MSSM regime, the quantity T_c as defined in Eq. (17) is approximately equal to $\langle T_2 \rangle /2$—one half the average time since pairwise coalescence among individuals in a population—which motivates our interpretation of T_c as a *coalescence* timescale. However, we emphasize that T_c is a derived quantity, defined by Eq. (17), which we will relate to the underlying parameters N and $\mu(s)$ in Section IV.A. [For example, for the case of single beneficial fitness effect, $\mu(s) = U\delta(s - s_b)$, we will have $T_c \approx \frac{1}{s_b} \log\left(\frac{s_b}{U} \log NU\right)$ provided that $T_c s_b \gg 1$.]

Plugging Eq. (15) and Eq. (16) into Eq. (5) and Eq. (10), we find that for $x < x_c$, $\psi(x)$ and $\xi(x)$ satisfy

$$-v\partial_x \psi(x) = (x - T_c v)\psi(x) + \int \mu(s) \left[e^{T_c s}\psi(x-s) - \psi(x)\right] ds, \tag{18}$$

and

$$v\partial_x \xi(x) = (x - T_c v)\xi(x) + \int \mu(s) \left[e^{T_c s}\xi(x+s) - \xi(x)\right] ds, \tag{19}$$

respectively. Taylor expanding $\psi(x-s)$ in Eq. (18) and keeping the lowest two nonzero orders in s, we find

$$-v\partial_x \psi = \left[x - T_c v + \int \mu(s)(e^{T_c s} - 1) ds\right] \psi - \left[\int \mu(s) s e^{T_c s} ds\right] \partial_x \psi + \left[\frac{1}{2}\int \mu(s) s^2 e^{T_c s} ds\right] \partial_x^2 \psi. \tag{20}$$

It will be useful to define the fitness scales b and c according to

$$b^3 \equiv \int \mu(s) s^2 e^{T_c s}, \tag{21}$$

and

$$c \equiv T_c v - \int \mu(s) ds \left[e^{T_c s} - 1\right], \tag{22}$$

respectively, which enable us to rewrite Eq. (20) in the compact form

$$b^3 \partial_x^2 \psi = (c - x)\psi. \tag{23}$$

In Table II, we list key defined quantities used throughout, along with definitions and interpretations of those quantities. [Note that we make

a distinction between quantities defined purely by the definitions in Table II (e.g., b and c) and the phenomenological quantities to which they approximately correspond (Δx_f and $\langle x_f \rangle$, as we later show).]

Similar manipulations can be carried out to approximate $\xi(x)$ in Eq. (19), with both equations solved by

$$\psi(x) \propto \xi(x) \propto \mathrm{Ai}\left(\frac{c-x}{b}\right). \tag{24}$$

This implies that for $x < x_c$,

$$f(x) \propto e^{-T_c x}\mathrm{Ai}\left(\frac{c-x}{b}\right), \tag{25}$$

$$w(x) \propto e^{T_c x}\mathrm{Ai}\left(\frac{c-x}{b}\right). \tag{26}$$

As discussed in Section III, we will take $w(x) \approx x$ and $f(x) \approx 0$ for $x > x_c$. Provided that $T_c b \gg 1$, the transition between these solutions will occur over a narrow boundary layer, such that x_c can be determined by enforcing continuity of $w(x)$ and $w'(x)$ at $x = x_c$ (as done by Good et al. (2014), for instance). This calculation is presented in Appendix B, with the result

$$x_c \approx c + |z_0|b - 1/T_c, \tag{27}$$

where $z_0 \approx -2.34$ is the least negative zero of $\mathrm{Ai}(z)$. We note that x_c is obtained by Fisher (2013) in the "high-speeds" regime using an alternative "solvability condition" for $w(x)$ which yields the same result.

Relation to infinitesimal approximation. We note that in the limit $T_c s \to 0$ for relevant fitness effects s (with $T_c b \gg 1$ fixed), the solutions $f(x)$ and $w(x)$ in Eq. (25) and Eq. (26), as well as x_c in Eq. (27), reduce to the corresponding quantities obtained using the infinitesimal approximation. In this limit,

$$\begin{aligned} T_c &\to \sigma^2/2D \\ b &\to D^{1/3} \\ c &\to \sigma^4/4D, \end{aligned} \tag{28}$$

where here, $\sigma^2 \equiv v - \int \mu(s) s ds$ corresponds to the population-wide fitness variance, and $D \equiv \frac{1}{2}\int \mu(s) s^2 ds$ corresponds to a mutational *diffusion* constant (an interpretation of which we provide in Subsection VI). The above approximation can be considered a *generalized infinitesimal* approximation; we further discuss the relation between our approximation detailed above and the infinitesimal approximation in the following Subsections.

A. The relation between T_c and N

Our derivation above made use of the phenomenological quantities T_c, b, c and v, which are functions of the underlying parameters N and $\mu(s)$. We now derive an additional equation relating these quantities, which allows us to solve for T_c (and thus v, as well as b and c) in terms of N and $\mu(s)$. To do so, we enforce the condition $\int f(x)w(x)dx = 1/N$ in Eq. (11), using the approximate $f(x)$ and $w(x)$ in Eq. (25) and Eq. (26), respectively. We emphasize that the expressions in Eq. (25) and Eq. (26) are only *local* approximations, valid within some range of x. In the next Subsection, we will obtain and discuss conditions which ensure that Eq. (25) and Eq. (26) are approximately valid within the important region dominating $1/N = \int f(x)w(x)dx$. The quantity $Nf(x)w(x)$ can be interpreted as a distribution of fitnesses of future common ancestors, and given $f(x)$ and $w(x)$ in Eqs. (25) and (26), evaluates to

$$Nf(x)w(x) \propto \mathrm{Ai}^2\left(\frac{c-x}{b}\right). \tag{29}$$

The integral $\int f(x)w(x)dx$ thus receives a dominant contribution from the region $|c-x| \sim \mathcal{O}(b)$. This motivates us to refer to the region $|c-x| \sim \mathcal{O}(b)$ probed by our condition of validity as the *fixation class*, since collectively, this region of fitness space produces a future common ancestor of the population with probability $\mathcal{O}(1)$, despite comprising a small fraction of the total population.

To evaluate the integral in Eq. (11), we need the overall constants of proportionality of $f(x)$ and $w(x)$. Recall that the constant of proportionality for $w(x)$ was fixed by matching to the solution $w(x) \approx x$ at $x = x_c$. On the other hand, the overall constant of proportionality of $f(x)$ must be determined by the normalization condition $\int f(x)dx = 1$. However, while our local approximation for $f(x)$ in Eq. (25) is valid near the nose, it is not necessarily valid throughout the entire normalization integral $\int f(x)dx$, which is dominated by fitnesses near the mean. Instead, we show in Appendix C that an analogous local approximation can be applied to the Laplace transform of $f(x)$, which allows us to obtain the relevant normalization,

$$f(x) \approx \exp\left[\frac{vT_c^2}{2} - \int \frac{\mu(s)ds}{s}\left(e^{T_c s} - T_c s - 1\right)\right] \times \frac{e^{-T_c x}}{b}\mathrm{Ai}\left(\frac{c-x}{b}\right). \tag{30}$$

Using this expression, we can obtain a relation be-

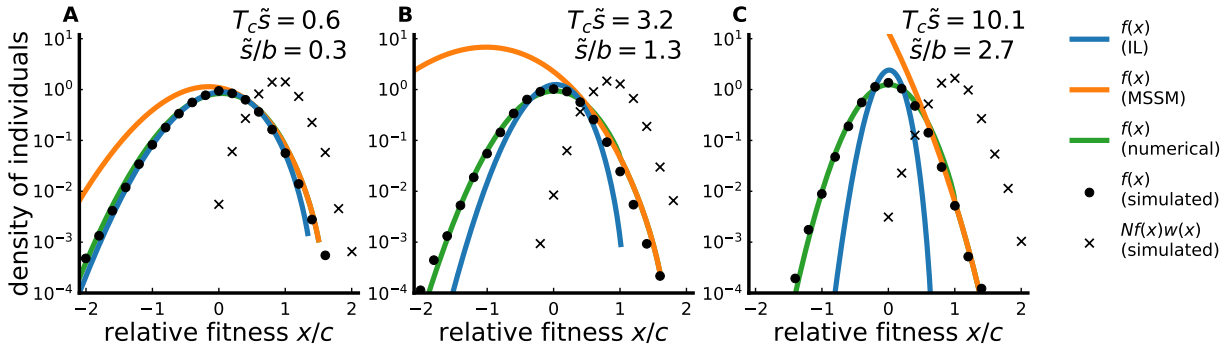

FIG. 2 Averaged distributions of relative fitnesses, $f(x)$, for simulated populations subject to beneficial mutations with an exponential DFE, with values of $T_c\tilde{s}$ and $\tilde{s}/b$ denoted above. A precise definition of $\tilde{s}$ is given in Subsection IV.B. Note the logarithmic scale of the vertical axis. Filled circles represent the simulated distribution of relative fitnesses, $f(x)$, as obtained by averaging over measurements from 490 different epochs. Black 'x' markers represent the simulated distribution of future common ancestor relative fitnesses, $Nf(x)w(x)$. Blue and orange lines denote theoretical predictions for $f(x)$ obtained with the infinitesimal (IL) and MSSM approximations, respectively; green lines denote predictions for $f(x)$ obtained using the numerical saddle point approximation detailed in Appendix C. In all cases $f(x)$ is best-predicted by the numerical saddle point approximation, but the MSSM approximation adequately predicts the behavior of $f(x)$ within the region of x dominating $\int f(x)w(x)dx$. In panel A the infinitesimal approximation adequately predicts $f(x)$ for all x, which is expected since $T_c\tilde{s} < 1$. Panels B and C correspond to larger values of $T_c\tilde{s}$ for which the infinitesimal approximation begins to break down, particularly in the region of x dominating $\int f(x)w(x)dx$.

tween T_c and N, $\mu(s)$ and x_c,

$$T_c(x_c - U) - \frac{vT_c^2}{2} + \int \frac{\mu(s)ds}{s}\left(e^{T_c s} - 1\right) \approx \log\left[Nx_cT_cb\right]. \quad (31)$$

(Appendix B). By combining Eq. (27) and Eq. (31), along with the definitions for T_c, b and c in Eqs. (17), (21) and (22) it is possible to solve for T_c in terms of N and $\mu(s)$. We therefore defined a simple numerical routine to solve Eq. (27) and Eq. (31) for a given choice of N and $\mu(s)$ (Appendix I). Although we discuss analytical approximations to this solution in Subsection IV.C, we use this numerical solution for comparison with simulations throughout the rest of this paper. We note that Eq. (31) can be rearranged to provide the relationship between T_c/N and the combinations T_cU and the distribution of scaled effects T_cs. In this way, our approximation is well suited to describing the dynamics corresponding to given values of the quantities T_cU and T_cs, parameter combinations more often inferred in natural settings (albeit often problematically, as our analyses show) than NU and Ns.

We illustrate the fact that Eq. (30) is only a locally valid approximation—and the importance of normalizing $f(x)$ appropriately—in Fig. 2. There, we compare our predicted $f(x)$ with the same quantity observed in simulations, for three example parameter choices. In the same Figure, we also plot histograms of future common ancestors fitnesses–that is, the empirically measured distribution $Nf(x)w(x)$—obtained from simulations. We can see that the prediction for $f(x)$ in Eq. (30) matches simulation results well in the region dominating $\int f(x)w(x)dx$. Outside, this region, however, the prediction in Eq. (30) breaks down, and would yield an incorrect constant of proportionality if normalized "directly"—particularly in the region of the MSSM regime that lies outside the infinitesimal regime. For comparison, we also plot predictions for $f(x)$ obtained using the infinitesimal approximation (for details, see Appendix I), as well as a numerical saddle point approximation we present in Appendix C. While our numerical saddle point approximation does yield improved accuracy in predicting the simulated $f(x)$ throughout its "bulk", it has a negligible impact on global quantities of interest such as v and $p_{\text{fix}}(s)$, which are dominated by the behavior of $f(x)$ and $w(x)$ in the fixation class. We therefore use only our analytical prediction for $f(x)$ in Eq. (30) throughout the remainder of this article.

B. Conditions of validity: a "moderate selection, strong mutation regime"

The approximate solutions $f(x)$ and $w(x)$ given in Eq. (25) and Eq. (26) are valid only within a limited region of fitness space. Our primary motivation in solving for $f(x)$ and $w(x)$ is that these quantities can be used to compute other dynamical quantities such as v and $\rho_f(s)$. These computations involve integrals of the form $\int f(x)w(x+s)dx$ (and our computations of genetic diversity statistics in Section V involve similar integrals as well). We can thus obtain conditions of validity of our approach by demanding that both $f(x)$ and $w(x)$ are well-approximated in the region of fitness space dominating these integrals. We will find that these conditions are most readily expressed in terms of $\rho(s)$ and the derived quantities T_c and b, although these conditions can straightforwardly be expressed in terms of the model parameters N and $\mu(s)$ using the relation between T_c and N given above.

In particular, as noted in the previous Subsection, $\int f(x)w(x)dx$ is dominated by the region $|c-x| \sim \mathcal{O}(b)$. The same region dominates the integral used to compute $\rho_f(s)$,

$$\int f(x)w(x+s)dx \propto e^{T_c s}\int \mathrm{Ai}\left(\frac{c-x}{b}\right)\mathrm{Ai}\left(\frac{c-x-s}{b}\right)dx, \quad (32)$$

as long as $s \ll b$. For $s \ll b$,

$$p_{\mathrm{fix}}(s) \approx \frac{e^{T_c s}}{N} \quad (33)$$

follows from Eq. (32), and $\rho_f(s) \propto e^{T_c s}\rho(s)$ (Appendix D). We therefore obtain a condition of validity of our approximation by ensuring that $f(x)$ and $w(x)$ are well-approximated within the region $|c-x| \sim \mathcal{O}(b)$. We will see below that this condition requires that $s \ll b$ for typical fixed effects s (i.e. for the region of s which dominates $\int s^2\rho_f(s)ds$), so that $f(x)$ and $w(x)$ are also well-approximated within the region dominating $\int f(x)w(x+s)dx$ for relevant s.

We obtain this condition of validity in Appendix C. The basic idea is to ensure that the inverse Laplace transforms of the linearized $f(x)$ and $w(x)$ reduce to the approximate expressions in Eq. (25) and Eq. (26). For relevant x, this will be true provided that a particular term can be neglected. By ensuring, for $|c-x| \sim \mathcal{O}(b)$, that this additional term yields small corrections near saddle points of the inverse Laplace integral, we obtain the condition

$$S_v' \equiv \frac{\int \mu(s)e^{T_c s}\left(e^{s/b}-1-(s/b)-\frac{1}{2}(s/b)^2\right)ds}{\int \mu(s)e^{T_c s}\frac{(s/b)^2}{2}ds} \ll 1. \quad (34)$$

Note that the condition $T_c b \gg 1$ is required to justify the dominant balance approximation for $w(x)$. These conditions can be applied to determine the suitability of the above approximations for populations subject to beneficial mutations, deleterious mutations, or some combination of the two.

As anticipated above, under relatively mild assumptions on $\rho(s)$, the condition $S_v' \ll 1$ is satisfied if

$$s \ll b \quad (35)$$

for s dominating $\int \rho_f(s)s^2ds$ (i.e. if typical fixed fitness effects s are much smaller than b in magnitude). For bookkeeping purposes, it is useful to define a maximum typical fixed fitness effect, $\tilde{s} \equiv \langle s_f\rangle + 2\Delta s_f$, where $\langle s_f\rangle$ and Δs_f are the average and standard deviation, respectively, of fixed mutation effects. From Eq. (29), we also note that $\Delta x_f \sim b$ (where Δx_f^2 denotes the fitness variance of future common ancestors). Given these definitions, the conditions $S_v' \ll 1$ and $T_c b \gg 1$ can then be written more compactly as

$$\tilde{s} \ll \Delta x_f, \quad T_c\Delta x_f \gg 1 \quad (36)$$

These conditions have dynamical interpretations which we discuss in Subsection VI. Notably, because $T_c\tilde{s} < 1$ is not required by our conditions, selection on single mutations need not be weak. Because $1/T_c < \tilde{s} \ll \Delta x_f$ (or $\tilde{s} < 1/T_c \ll \Delta x_f$) is permitted, we refer to the regime of validity of our approximation as a *moderate selection, strong mutation* (MSSM) regime. ["Moderate selection" refers to the requirement that $\tilde{s} < \Delta x_f$, while "strong mutation" refers to the requirement that clonal interference is strong.]

C. Specific example DFEs

We can gain more intuition about the MSSM regime by solving for T_c (and therefore v) for a few concrete example DFEs. In doing so, we can obtain expressions for the validity of the MSSM approximation in terms of the more experimentally accessible parameters N, U and $\rho(s)$, although in doing so we make further assumptions on the specific DFE considered.

Single beneficial fitness effect. The simplest

such scenario is that of a single beneficial fitness effect s_b. We begin by noting that when $T_c s_b \ll 1$ Eq. (31) reduces to the well-known results

$$T_c \approx \frac{1}{s_b}\left[2\left(\frac{s_b}{U}\right)\log\left(N^3 U s_b^2\right)\right]^{1/3}, \tag{37}$$

and

$$b \approx \left(\frac{U s_b^2}{2}\right)^{1/3} \tag{38}$$

from the infinitesimal limit (Good et al., 2014; Neher and Hallatschek, 2013). We can see that Eq. (37) is self-consistently valid (i.e. $T_c s_b \ll 1$, $T_c b \gg 1$ and $s_b \ll b$) when

$$\frac{U}{s_b} \gg \log\left(N^3 U s_b^2\right) \gg 1. \tag{39}$$

Note that this condition requires more than just $U \gg s_b$; U/s_b must also be larger than the large parameter combination $\log\left(N^3 U s_b^2\right)$.

In the opposite regime $T_c s_b \gg 1$, Eq. (31) reduces to

$$T_c \approx \frac{1}{s_b}\log\left[\frac{s_b}{U}\log NU\right] \tag{40}$$

and

$$b \approx s_b\left(\frac{\log NU}{\log^2\left(\frac{s_b}{U}\log NU\right)}\right)^{1/3} \tag{41}$$

which coincides with the "high-speeds" regime considered by Fisher (2013) (Appendix B). This is self-consistently valid when

$$\log NU \gg \log^2\left[\frac{s_b}{U}\log NU\right] \gg 1. \tag{42}$$

By combining Eq. (42) and Eq. (40), we can see that there is a relatively smooth crossover between the infinitesimal and "high-speeds" regimes when $U/s_b \sim \log NU$, where $T_c s_b \sim \mathcal{O}(1)$. For fixed U and s_b, the infinitesimal limit always breaks down for sufficiently large N, while the "high-speeds" regime is eventually valid for any s_b and U.

Distributions of beneficial fitness effects.

More generally, as long as the DFE $\rho(s)$ falls off *faster* than exponentially with large positive s, convergence to the MSSM regime is obtained in the limit $N \to \infty$. In this limit, previous work has shown that the relevant integrals over $\mu(s)e^{T_c s}$ become increasingly sharply peaked around a characteristic value, such that

$$\int s^k e^{sT_c}\mu(s)ds \approx U_e s_e^k e^{s_e T_c}, \tag{43}$$

where s_e coincides with the predominant fitness effect $s^* \equiv \mathrm{argmax}_s \rho_f(s)$ (Fisher, 2013). This shows that short-tailed DFEs can be approximated by an effective DFE with $\mu(s) = U_e\delta(s - s_e)$ (Desai and Fisher, 2007; Fisher, 2013; Good et al., 2012; Hegreness et al., 2006). Moreover, under rather general conditions, one can show that s^* becomes increasingly small compared to b as N (and therefore T_c) increases, so that the conditions of validity of the MSSM approximation will be satisfied (Appendix G).

On the other hand, if $\rho(s)$ falls off slower than exponentially with large positive s, the integrals in Eq. (43) no longer converge, and the MSSM approximation cannot be applied. The case of an exponential DFE (with mean effect size $\langle s\rangle$) is a marginal case, since convergence will depend on the relative values of T_c and $\langle s\rangle$, or equivalently, of T_c and $\langle s_f\rangle$. When $T_c\langle s_f\rangle \ll 1$, the MSSM approximation reduces to the infinitesimal approximation, as in the case of a single beneficial effect. In the opposite case that $T_c\langle s_f\rangle \gg 1$, we find that

$$T_c\langle s_f\rangle \approx \sqrt{\frac{2\langle s\rangle \log N\langle s\rangle}{U}} \tag{44}$$

and

$$\frac{b}{\langle s_f\rangle} \approx \left(\frac{U}{\langle s\rangle}\right)^{1/3}. \tag{45}$$

(Appendix G). The conditions $T_c\langle s_f\rangle \gg 1$, $\langle s_f\rangle \ll b$ and $T_c b \gg 1$ are then jointly satisfied, and the MSSM approximation is valid, when

$$\log N\langle s\rangle \gg \frac{U}{\langle s\rangle} \gg 1. \tag{46}$$

Note that since $T_c\langle s_f\rangle \approx T_c\langle s\rangle/(1 - T_c\langle s\rangle)$ for an exponential DFE, $(1 - T_c\langle s\rangle)$ is a small and positive quantity in this case, such that the relevant integrals converge. Note that the conditions in Eq. (46) differ from those for a single beneficial fitness effect in Eq. (42) in that $U \gg \langle s\rangle$ is now explicitly required, no matter how large the value of N. The opposite case, in which $\langle s\rangle > U$, cannot be described using the MSSM approximation, and remains only partially understood; for a discussion of this case, see Fisher (2013).

Deleterious mutations in adapting populations.

The previous two examples have mostly recapitulated earlier results for beneficial mutations in the infinitesimal (Cohen et al., 2005; Hallatschek, 2011; Neher and Hallatschek, 2013; Tsimring et al., 1996)

and high-speed regimes (Fisher, 2013). A key advantage of our MSSM approximation is that it can also be applied to scenarios with large numbers of deleterious mutations, where few analytical results are currently available.

As an example, we first consider a scenario where beneficial and deleterious mutations have the same effect size s, and occur at rates U_b and U_d respectively. The infinitesimal approximation will once again apply in the limit that $T_c s \ll 1$, but qualitatively new behavior starts to emerge when $T_c s \gg 1$. In this case, a useful simplification occurs in Eq. (31): the contributions to the left-hand side are exponentially suppressed for deleterious mutations, with the exception of the U_d/s term. As a result, the overall solution for T_c reduces to the purely beneficial case in Eq. (40), but with an *effective population size*

$$N_e \equiv N e^{-U_d/s}. \tag{47}$$

in place of N.

The right-hand side of Eq. (47) is the classical prediction for the number of mutation-free individuals that would exist at mutation-selection balance (Haigh, 1978), prominent in discussions of background selection (Charlesworth et al., 1993). Our results suggest that this simple approximation continues to apply in certain non-equilibrium settings as well, even when the deleterious fitness costs are small compared to the typical fitness variation in the population. Although these strongly deleterious mutations are unable to fix, they can still affect the evolutionary dynamics through the size of the unloaded class, similar to the "background selection" (Charlesworth, 1994) or "ruby in the rubbish" (Peck, 1994) behavior that has been observed for single beneficial mutations. We emphasize that this effective population size approximation was a direct consequence of our mathematical expressions in Eqs. (21) and (31). This provides a more rigorous justification for previous ad-hoc approaches that assumed that strongly deleterious mutations can be treated in this way (Good et al., 2014; Söderberg and Berg, 2007). We emphasize that the effective population size in Eq. (47) does *not* denote a coalescence timescale, as is often implied in the literature. Indeed, the asymptotic expressions in Eqs. (40) and (37) show that T_c will typically be much less than N_e in the MSSM regime.

Together with Eq. (47), the conditions of validity in Eq. (42) imply that the MSSM approximation will be valid for sufficiently large N for any choice of U_b, U_d and s. Note that while we have assumed that $s_d = s_b = s$ for simplicity above, this same argument can be generalized to unequal selection strengths as long as $T_c s_d \gg 1$. For example, if $T_c s_b \ll 1$, then T_c will instead be described by the infinitesimal approximation in Eq. (37), with $N_e = N e^{-U_d/s_d}$.

Background selection.

Finally, we can also apply our framework to the case of purely deleterious mutations (often referred to as "background selection"). For simplicity, we will focus on the well-studied case where deleterious mutations have a single fitness cost s_d. As noted in previous studies (Good et al., 2014; Neher and Hallatschek, 2013; Neher et al., 2013), this scenario is well-described by the infinitesimal limit when $T_c s_d \ll 1$, with s_d replacing s_b in Eq. (37). In the opposite case where $T_c s_d \gg 1$, our MSSM approximations yield an alternative solution, in which

$$T_c \approx \frac{1}{s_d} \log \left[\frac{U/s_d}{\log\left(\frac{1}{NUe^{-U/s_d}}\right)} \right] \tag{48}$$

and

$$b \approx s_d \left[\frac{1}{2} \log \left(\frac{1}{NUe^{-U/s_d}} \right) \right]^{1/3}. \tag{49}$$

This solution is self-consistently valid ($T_c s_d \gg 1$, $s_d \ll b$ and $T_c b \gg 1$) when

$$1 \ll \log \left(\frac{1}{NUe^{-U/s_d}} \right) \ll \frac{U}{s_d}. \tag{50}$$

For fixed values of U and s_d, this condition will always be violated for sufficiently large population sizes ($NUe^{-U/s_d} \gtrsim 1$). This breakdown is consistent with previous theory (Charlesworth, 1994; Cvijović et al., 2018; Good et al., 2014) which predicts that the large-N limit of the background selection model approaches a nearly neutral regime, with a coalescent timescale,

$$T_c = N e^{-U/s_d}. \tag{51}$$

Interestingly, Eq. (50) shows that the MSSM approximation can remain valid even for arbitrarily large values of $T_c s_d$, as long as the corresponding values of U/s_d are also sufficiently large. Thus, even for purely deleterious mutations, the MSSM approximation does not necessarily require that all mutations have weak effects (i.e. that $T_c s_d \ll 1$). In principle, strong clonal interference can occur even for large scaled fitness effects ($T_c s_d \gtrsim 1$), but only in an increasingly narrow region of the underlying parameter space.

This last point suggests an alternative way of looking at the relation between T_c and N in Eq. (31). In

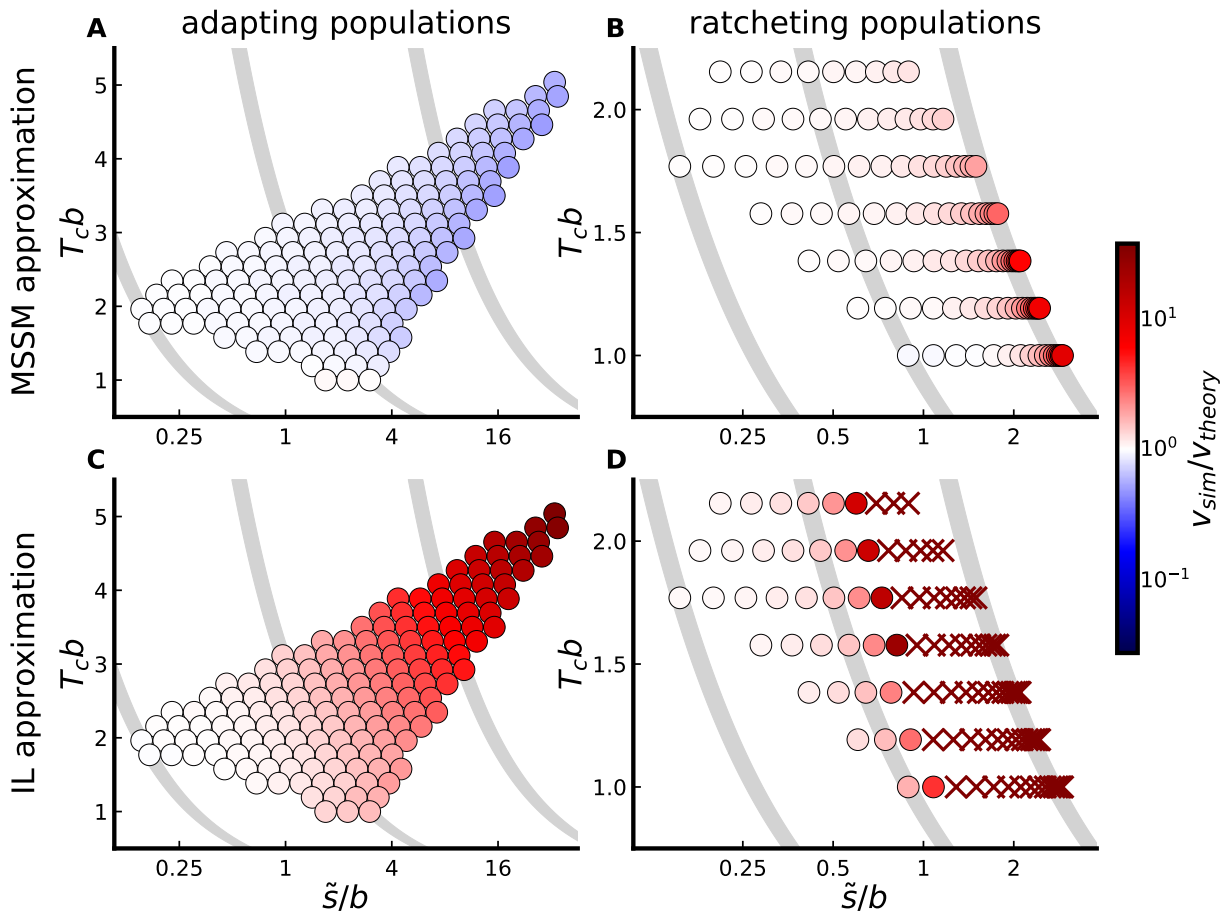

FIG. 3 Comparison of theoretical predictions for the rate of change in mean fitness, v, to the corresponding rates observed in simulations, for populations subject to an exponential DFE of beneficial mutations or deleterious mutations. The color of each point denotes the ratio of simulated to predicted v for a set of parameters (with red and blue indicating that theory underestimates or overestimates v, respectively, according to the scale at right). In panels A and B, predictions are obtained using the MSSM approximation, while in panels C and D, predictions are obtained using the infinitesimal (IL) approximation. The location of a point along the vertical axis denotes the value of $T_c b$ for its corresponding set of parameters (with T_c and b computed using the MSSM approximation in all panels), and the horizontal axis denotes the value of $\tilde{s}/b$ for that set of parameters (with $\tilde{s} \equiv \langle s_f \rangle + 2\Delta s_f$); curves of constant $T_c \tilde{s}$ are denoted in gray. Panels A and C involve simulations of populations subject only to beneficial mutations (adapting populations), and panels B and D involve simulations of populations subject only to deleterious mutations (ratcheting populations). Simulated parameters are those lying on the constrained grid described below, and depicted in the space of NU vs. $N\langle s \rangle$ in Fig. S1. The 'x' markers in D denote populations for which the infinitesimal approximation predicts a rate of fitness change of the incorrect sign.

addition to solving for T_c as a function of the bare parameters N, U and s_d, it is also possible to directly solve for NU as a function of the phenomenological variables $T_c s_d$ and $T_c U$:

$$\log NU \approx \frac{U}{s_d}\left(1 - \frac{(T_c s_d)^2 e^{-T_c s_d}}{2}\right). \tag{52}$$

This expression gives the underlying value of NU that would be required for the MSSM regime to apply for a given value of $T_c s_d$ and $T_c U$. This alternative view of the parameter space, in which $T_c s_d$ and $T_c U$ are the "independent" parameters, is often more natural in applied settings, where the phenomenological parameters are usually estimated from contemporary patterns of genetic diversity. This perspective will be useful for our discussion of nonsynonymous site frequency spectra below.

D. Simulation results

To complement our analytical predictions, we simulated populations subject either to only beneficial mutations (adapting populations) or only deleterious mutations (ratcheting populations). Simulated parameters N and $\mu(s)$ are chosen to correspond with a grid with linearly spaced $T_c b$ values, and with logarithmically spaced $T_c \langle s_f \rangle$ values (for adapting populations) or $T_c \langle s \rangle$ values (for ratcheting populations), subject to the constraints $T_c b > 1$, $1 < NU < 10^5$, $1 < N \langle s \rangle < 10^{3.5}$ and $U/\langle s \rangle \leq 10^4$. For ratcheting populations we add a further constraint that $U/\langle s \rangle > 1$. Simulated populations have size $10^3 < N < 1.2 \times 10^5$. These constraints are chosen both to limit attention to the MSSM regime (and the region of parameter space where it begins to break down) and to ensure feasibility of our individual-based simulations. For each point on these constrained grids, we (separately) simulated populations subject to an exponential DFE, and to a stretched-exponential DFE with steepness parameter $\beta = 2$. Details of the numerical implementation of these simulations and choice of parameters are given in Appendix H; in Fig. S1 we plot our constrained grid of simulation parameters in the space of NU vs. $N \langle s \rangle$, colored by values of $T_c b$.

In Fig. 3, we plot our constrained grid in the space of $T_c b$ vs. $\tilde{s}/b$, with the color of each point denoting the accuracy of either the MSSM approximation or infinitesimal approximation in predicting the rate of change v in the mean fitness. For clarity, we include only populations subject to an exponential DFE in Fig. 3. Plotting our simulation grid in the space of $T_c b$ vs. $\tilde{s}/b$ enables us to verify that predictions of the MSSM approximation are accurate when our conditions of validity are met (roughly speaking, when $T_c b \gg 1$ and $\tilde{s} \ll b$). From Fig. 3, we can see that our predictions for v are reasonably accurate even for $\tilde{s}/b \sim \mathcal{O}(1)$, as long as $T_c b > 1$. Moreover, we can see that the infinitesimal approximation breaks down even for small $\tilde{s}/b$ when $T_c \tilde{s}$ is large. In Fig. S2, we plot the same grid of simulated populations in the space of NU vs $N \langle s \rangle$, colored by $v_{\text{sim}}/v_{\text{theory}}$.

To visualize the quantitative agreement between simulations and our predictions more directly, we plot the ratio $v_{\text{sim}}/v_{\text{theory}}$ as a function of the single quantity $\tilde{s}/b$ in Fig. 4, with points colored according to their values of $T_c b$. We include, in Fig. 4, simulated populations subject to a $\beta = 2$ stretched-exponential DFE, as well as simulated populations subject to an exponential DFE. In the same Figure, we compare theoretical predictions of the *fixation rate* of new mutations, given by

$$F = N \int \mu(s) p_{\text{fix}}(s) \approx \int \mu(s) e^{T_c s} ds, \tag{53}$$

as well as the average fixed effect $\langle s_f \rangle$ and standard deviation Δs_f in fixed effects, to measurements of these quantities in simulations. Predictions for the quantities $\langle s_f \rangle$ and Δs_f are obtained using the simulated $\rho(s)$ and $p_{\text{fix}}(s)$ in Eq. (33). For each of these quantities, we observe highly quantitative agreement between simulations and theory for small and moderate values of $\tilde{s}/b$.

V. STATISTICS OF GENETIC DIVERSITY

We now consider statistics of genetic diversity within the MSSM regime. Our central quantity of interest will be the frequency spectrum $h_s(\nu)$ for a single site with fitness effect s, which is defined such that

$$H_{\text{sel}}(\nu) d\nu \equiv \int h_s(\nu) \rho(s) ds \cdot d\nu \tag{54}$$

gives the expected number of selected (non-neutral, including both beneficial and deleterious) mutations per site with frequencies between ν and $\nu + d\nu$. We will also consider an analogous aggregate quantity

$$H_{\text{neu}}(\nu) \equiv h_0(\nu) \tag{55}$$

defined for a subset of putatively neutral sites (e.g. synonymous sites, short introns, etc). These site frequency spectra are important statistics that are often used to make inferences regarding the evolutionary forces acting within a population (Nielsen, 2005). Several empirical studies have drawn inferences on the presence and strength of selection acting in a population based on differences in the site frequency spectrum among synonymous (i.e. mostly neutral) and nonsynonymous (i.e. more selected) mutations (Eyre-Walker et al., 2006). The key idea underlying these approaches is simple: because synonymous and nonsynonymous sites (which are interdigitated throughout the genome) share the same demographic history, differences in their patterns of diversity can be attributed to (positive or negative) selection on nonsynonymous mutations (McDonald and Kreitman, 1991). The ability to make inferences using this line of thinking requires predictions for $h_s(\nu)$ (Hartl et al., 1994; Sawyer and Hartl, 1992), which are lacking for rapidly evolving populations.

In the following subsections, we develop analytical predictions for these neutral and selected site

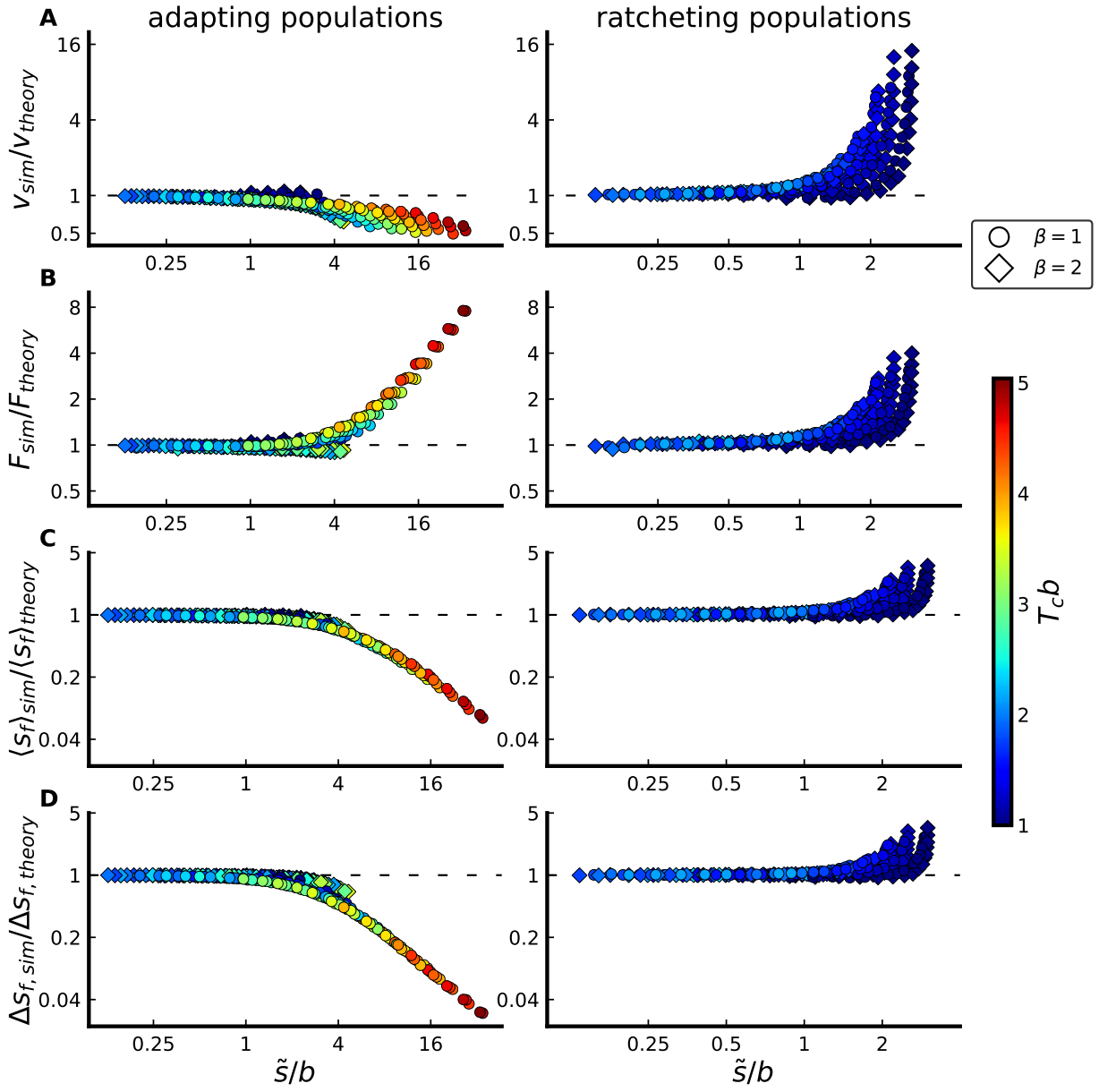

FIG. 4 Comparison between simulated and predicted rate v of fitness change (A), rate F of mutation accumulation (B), average fixed effect $\langle s_f \rangle$ (C), and standard deviation Δs_f in fixed effects (D). Populations in the left column are subject to only beneficial mutations; populations in the right column are subject to only deleterious mutations. Predictions are obtained using the MSSM approximation; quantitative agreement between simulations and predictions is obtained as long as $\tilde{s}/b \lesssim \mathcal{O}(1)$. Simulated sets of parameters lie on the same (constrained) grid considered in Fig. 3, although here we include, for each point on that grid, both an exponential DFE (circles) and $\beta = 2$ stretched-exponential DFE (diamonds).

frequency spectra in the MSSM regime, noting important departures from the classical intuition in Eq. (1). We demonstrate that $h_s(\nu)$ and $h_0(\nu)$ are simply related by a constant factor above a characteristic frequency $\nu_c \ll 1$, which is not simply summarized by the scaled selection coefficient $T_c s$. Building on previous work (Desai et al., 2013; Fisher, 2013; Neher and Hallatschek, 2013), we also demonstrate a partial correspondence between the genealogies in our model and the Bolthausen-Sznitman coalescent, and we analytically describe the departures from the BSC SFS that can be observed at low frequencies.

Compared to the BSC SFS, we find that the functional form of these departures provides substantially greater power to distinguish different parameter combinations. Finally, we compute the pairwise heterozygosity for selected and neutral mutations, and demonstrate a correspondence between the pairwise coalescence time $\langle T_2 \rangle$ and the quantity T_c defined in Eq. (17). In the Discussion, we consider implications of our results in the context of the population genetic inference methods described above.

1. The site frequency spectrum: basic formalism

To calculate $h_s(\nu)$, it will be useful to first consider the discrete version,

$$P_s(k|m) \equiv \int \binom{m}{k} \nu^k (1-\nu)^{m-k} h_s(\nu) d\nu, \quad (56)$$

which gives the normalized probability of observing a mutation in exactly k individuals in a random sample of size m. The continuous version can be recovered by taking the limit of large sample sizes:

$$h_s(\nu) \equiv \lim_{m\to\infty} m P_s(\nu m|m). \quad (57)$$

One advantage of switching to $P_s(k|m)$ is that it can be rewritten as an average over lineages defined at different times in the past:

$$P_s(k|m) = \mu \int_0^\infty \binom{m}{k} \Lambda^s_{m,k}(t) dt, \quad (58)$$

where μ is the per-site mutation rate and

$$\Lambda^s_{m,k}(t) = N \left\langle \left(\frac{n_s(t)}{n_s(t) + \sum_{i=1}^{N-1} n_i(t)} \right)^k \times \left(\frac{\sum_{i=1}^{N-1} n_i(t)}{n_s(t) + \sum_{i=1}^{N-1} n_i(t)} \right)^{m-k} \right\rangle. \quad (59)$$

In this equation, $n_s(t)$ denotes the present-day size of a lineage founded by a mutation with effect size s that occurred t generations ago, while the $\{n_i(t)\}_{i=1}^{N-1}$ represent the lineages founded by the remaining $N-1$ individuals that were alive at that time. Eq. (58) can be interpreted as integrating over the possible times a mutation last occurred at a given site. This mutation could have occurred on any of the N possible genetic backgrounds in the population. Given that it did occur, it will be observed in k individuals in the present with probability $\binom{m}{k}\Lambda^s_{m,k}(t)/N$.

We note that in the special case of a neutral site ($s=0$), the quantity $\Lambda^0_{m,k}(t)$ can be interpreted as a *merger probability* within a corresponding coalescent model. That is, $\Lambda^0_{m,k}(t)$ gives the probability that in a sample of size m, a particular set of k individuals share a common ancestor at t generations into the past, and the remaining $m-k$ individuals do *not* trace back to that same ancestor. Related merger probabilities (defined slightly differently) are considered by Neher and Hallatschek (2013) to show a correspondence between genealogies in the infinitesimal regime and those of the Bolthausen-Sznitman coalescent (BSC). Below, we follow a similar approach to simplify the quantities $\Lambda^s_{m,k}(t)$ used in our calculation of the selected site frequency spectrum. In Appendix E, we use our calculated $\Lambda^0_{m,k}(t)$ to explicitly demonstrate a (partial) correspondence with the BSC in the MSSM regime, and highlight key ways in which our results differ from those of Neher and Hallatschek (2013).

To simplify Eq. (59), we make the key approximation that each of $n_i(t)$ and $n_s(t)$ evolve as a collection of independent branching processes described by Eq. (9) (Desai et al., 2013; Fisher, 2013; Neher and Hallatschek, 2013). That is, we assume the fates of lineages are coupled only through the average rate of adaptation, which is shaped by interference as calculated in Section IV. The identity $\Gamma(m)C^{-m} = \int_0^\infty dz z^{m-1} e^{-zC}$ then yields

$$\Lambda^0_{m,k}(t) \approx (-1)^m \frac{N}{\Gamma(m)} \times \int_0^\infty dz z^{m-1} \partial_z^k \left\langle e^{-zn(t)} \right\rangle \partial_z^{m-k} \left\langle e^{-zn(t)} \right\rangle^{N-1}, \quad (60)$$

where, since the $\{n_i(t)\}_{i=1}^{N-1}$ are identically distributed, we have dropped the subscripts i in writing $n(t)$. As we discuss in Appendix A,

$$\left\langle e^{-zn(t)} \right\rangle \approx e^{-\Phi_0(z,t)}, \quad (61)$$

where $\Phi_0(z,t) \equiv \int f(x)\phi_z(x,t)dx$, and where $\phi_z(x,t)$ satisfies

$$\partial_t \phi_z(x,t) = -v\partial_x \phi_z(x,t) + x\phi_z(x,t) + \int \mu(s) \left[\phi_z(x+s,t) - \phi_z(x,t)\right] ds - \phi_z^2(x,t), \quad (62)$$

with initial condition $\phi_z(x,0) = z > 0$; more generally,

$$\left\langle e^{-zn_s(t)} \right\rangle \approx e^{-\Phi_s(z,t)}, \quad (63)$$

with $\Phi_s(z,t) \equiv \int f(x-s)\phi_z(x,t)dx$.

A detailed analysis of Eq. (62) is conducted by Fisher (2013), which we review and extend in Appendix A. The key results we use here are that

$$\Phi_0(z,t) \propto z^{e^{1-t/T_c}}, \tag{64}$$

for $b(t-T_c) \gg 1$, over the region of z that will turn out to dominate the integral in Eq. (199), and that for $t < T_c - \mathcal{O}(1/b)$, $\Phi_0(z,t)$ is approximately linear in z. In Appendix E, we show, using the properties of $\phi_z(x,t)$ discussed in Appendix A along with our result for $f(x)$ in Eq. (25), that

$$\frac{\Lambda^s_{m,k}(t)}{\Lambda^0_{m,k}(t)} \approx \frac{\Phi_s(z,t)}{\Phi_0(z,t)} \approx \begin{cases} e^{st} & t < T_c \\ e^{sT_c} & t > T_c \end{cases}, \tag{65}$$

provided that $|s| \ll b$. The relevant site frequency spectra then can then be obtained by substituting these expressions into Eq. (58) and integrating over time.

2. The ratio between the selected and neutral SFSs

The simple behavior of Eq. (65) suggests that the *ratios* of neutral and selected site frequency spectra may take on a particularly simple form in certain regimes. For example, if we consider frequencies $\nu = k/m$ that are sufficiently large that the integrals in Eq. (58) are dominated by times greater than T_c (we discuss the minimum frequencies required for this assumption below), then the time-independent behavior of Eq. (65) immediately implies that

$$\frac{h_s(\nu)}{h_0(\nu)} = \lim_{m\to\infty} \frac{P_s(\nu m|m)}{P_0(\nu m|m)} \approx e^{T_c s}, \tag{66}$$

when $|s| \ll b$. Since $p_{\text{fix}}(s) \approx \frac{1}{N} e^{T_c s}$, we can also write this as

$$\frac{h_s(\nu)}{h_0(\nu)} \approx \frac{p_{\text{fix}}(s)}{p_{\text{fix}}(0)}. \tag{67}$$

In this range of frequencies, Eq. (67) predicts that neutral and selected site frequency spectra are simply proportional to one another, and that the constant of proportionality is equal to the ratio of their fixation probabilities.

By summing over sites, we can obtain an analogous result for the aggregate site frequency spectra,

$$\frac{H_{\text{sel}}(\nu)}{H_{\text{neu}}(\nu)} \approx \frac{F}{U}, \tag{68}$$

where $F \equiv N \int \mu(s) p_{\text{fix}}(s) ds$ is the total fixation rate of the selected mutations defined in Eq. (53). Recall that we have defined these aggregate quantities so that a (rough) analogy can be drawn between $H_{\text{neu}}(\nu)$ and the distribution of *synoynmous* site frequencies, and between $H_{\text{sel}}(\nu)$ and the distribution of *nonsynonymous* site frequencies. Under this analogy, the right-hand side of Eq. (68) corresponds to dN/dS—the ratio of nonsynonymous to synonymous divergence rates (adjusted, as usual, for differences in mutation rates among the two types of mutations) (Yang and Bielawski, 2000).

In Fig. 5, we plot the ratio between $H_{\text{sel}}(\nu)$ and $H_{\text{neu}}(\nu)$ as measured in simulations for a subset of the populations considered in Fig. S3. In all cases, the ratio $H_{\text{sel}}(\nu)/H_{\text{neu}}(\nu)$ approaches F/U as $\nu \to 1$. This can be understood as a consequence of the fact that mutations already present at the very highest frequencies will drift neutrally to (or away from) fixation. In contrast, as $\nu \to 0$, $H_{\text{sel}}(\nu)/H_{\text{neu}}(\nu) \to 1$ for all cases depicted in Fig. 3 (and thus Eq. (67) breaks down). This is also to be expected: mutations observed at sufficiently low frequencies will have occurred at short enough times into the past that their fates have not yet been substantially impacted by selection.

The behavior we see in these extreme limits is also observed in the independent sites model in Eq. (1) (Hartl et al., 1994; Sawyer and Hartl, 1992). However, these classical results predict that the high frequency limit only applies for extremely large frequencies ($1-\nu \ll 1/|T_c s| \ll 1$). In contrast, Fig. 5 shows that $H_{\text{sel}}(\nu)/H_{\text{neu}}(\nu) \approx F/U$ over a much broader range of intermediate frequencies in the MSSM regime. This has important consequences for the interpretation of population-genetic data, and in particular for application of the "asymptotic alpha" approach introduced by Messer and Petrov (2013); we comment on this further in the Discussion. Similarly, the independent sites model in Eq. (1) predicts that the deleterious site frequency spectrum will start to differ from its neutral counterpart when $\nu \gtrsim 1/T_c|s|$. In contrast, Fig. 5 shows that the frequency scale at which $H_{\text{sel}}(\nu)$ and $H_{\text{neu}}(\nu)$ start to differ is not given by $1/T_c|s|$ in the MSSM regime; we discuss related implications for estimates of $T_c s$ in the Discussion.

Importantly, from Fig. 5 it is clear that the shapes of SFS ratio curves—and in particular, the frequency scale on which they transition from 1 to F/U—do not depend simply on $T_c \langle s \rangle$ (or, equivalently, on $T_c \langle s_f \rangle$); a given value of $T_c \langle s \rangle$ is compatible with SFS ratio curves which differ substantially, and the parameter $T_c b$ is equally if not more important in determining the crossover frequency scale. Empirically, we can see that variation of SFS ratio curves (and the crossover frequency scale) is largely medi-

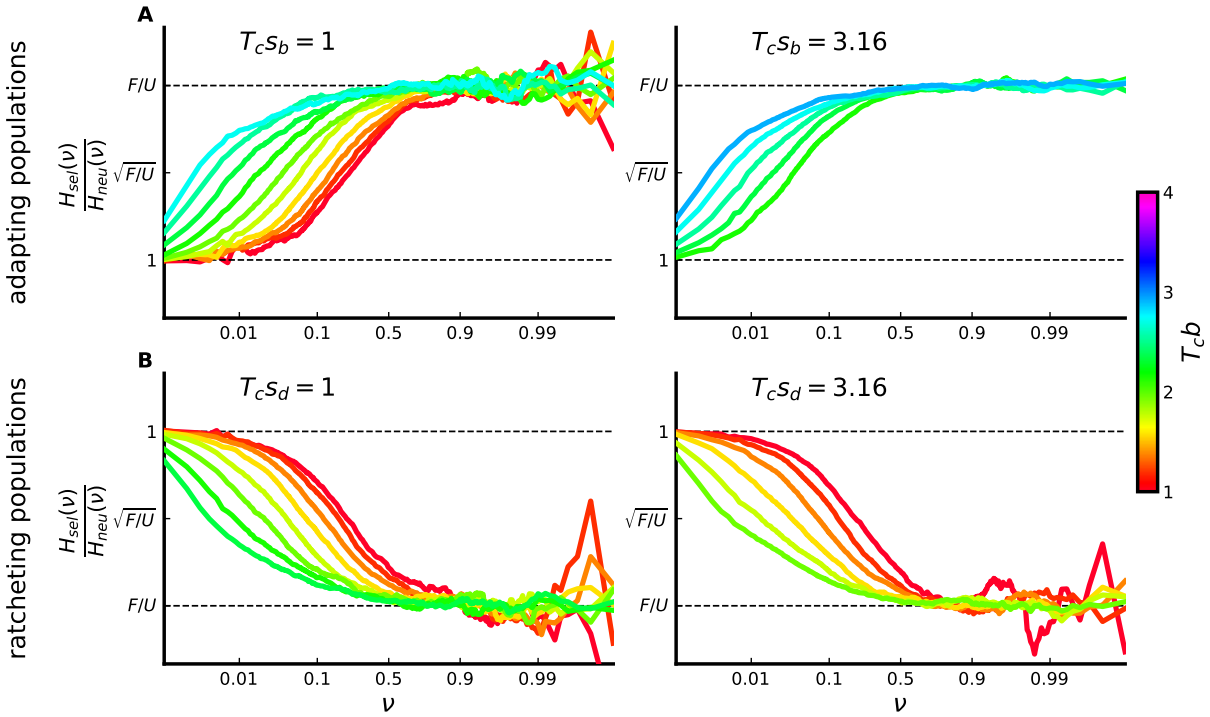

FIG. 5 Ratio between selected and neutral SFSs, scaled by the ratio between selected and neutral mutation rates, for adapting populations (A) and ratcheting populations (B) subject to a single-effect DFE with $T_c s$ values denoted in each panel. Ratio curves are plotted on a log scale, and are scaled such that tick marks correspond to the values of $\sqrt{F/U}$ and F/U averaged over simulation runs, for each curve (with F the fixation rate of selected mutations, either beneficial or deleterious). Each curve corresponds to a simulated population with parameters lying on the constrained grid depicted in Fig. S3; for each panel, the represented $T_c b$ values are linearly spaced.

ated by the quantity $N\sigma$, another population-level quantity; note that $\sigma^2 = v - U\langle s\rangle$ denotes the population-wide fitness variance. To see this across a broad range of parameters, for each observed SFS ratio curve we identified, using spline interpolation, the frequency ν_c at which $H_{\text{sel}}(\nu)/H_{\text{neu}}(\nu)$ reaches $\sqrt{F/U}$ (the half-maximum of $H_{\text{sel}}(\nu)/H_{\text{neu}}(\nu)$ in log-space). In Fig. 6, we plot these crossover frequencies ν_c a function of our MSSM approximation predictions for $N\sigma$, with points colored by $T_c\langle s_f\rangle$ values (for adapting populations) or $T_c\langle s\rangle$ values (for ratcheting populations). We can see a simple power law dependence of ν_c on $N\sigma$; parameter combinations with similar $N\sigma$ values have similar ν_c values, even if their values of $T_c\langle s\rangle$ differ substantially. This echoes findings of Good et al. (2014) in the infinitesimal regime; in that work, a similar collapse is found for the dependence of the neutral heterozygosity. In Fig. S4, we plot the full SFS ratio curves for the same set of parameter combinations, colored by $N\sigma$ values; there we can see that variation in full SFS ratio curves (in addition to the crossover frequency ν_c) is largely mediated by variation in $N\sigma$.

Together, these results suggest that efforts to infer the distribution of scaled effects $T_c s$ using existing approaches (e.g Bustamante et al. (2001); Eyre-Walker et al. (2006)) which "fit" the neutral and selected SFSs may fail when applied to rapidly evolving populations such as those considered here. In particular, these classical results cannot explain the marked dependence of SFS ratio curves on $T_c b$ and/or $N\sigma$ observed above for fixed values of $T_c\langle s\rangle$.

3. The neutral and selected site frequency spectra

We now proceed to compute the neutral and selected SFSs directly. In Appendix E, we use the merger probabilities $\Lambda^0_{m,k}(t)$ calculated using Eq. (60) and Eq. (64) to simplify Eq. (58), with the result

$$P_0(k|m) \approx T_c\mu\binom{m}{k} \times \int_0^1 \frac{\sin \pi y}{\pi y} B\left(k-y, m-k+y\right) dy, \quad (69)$$

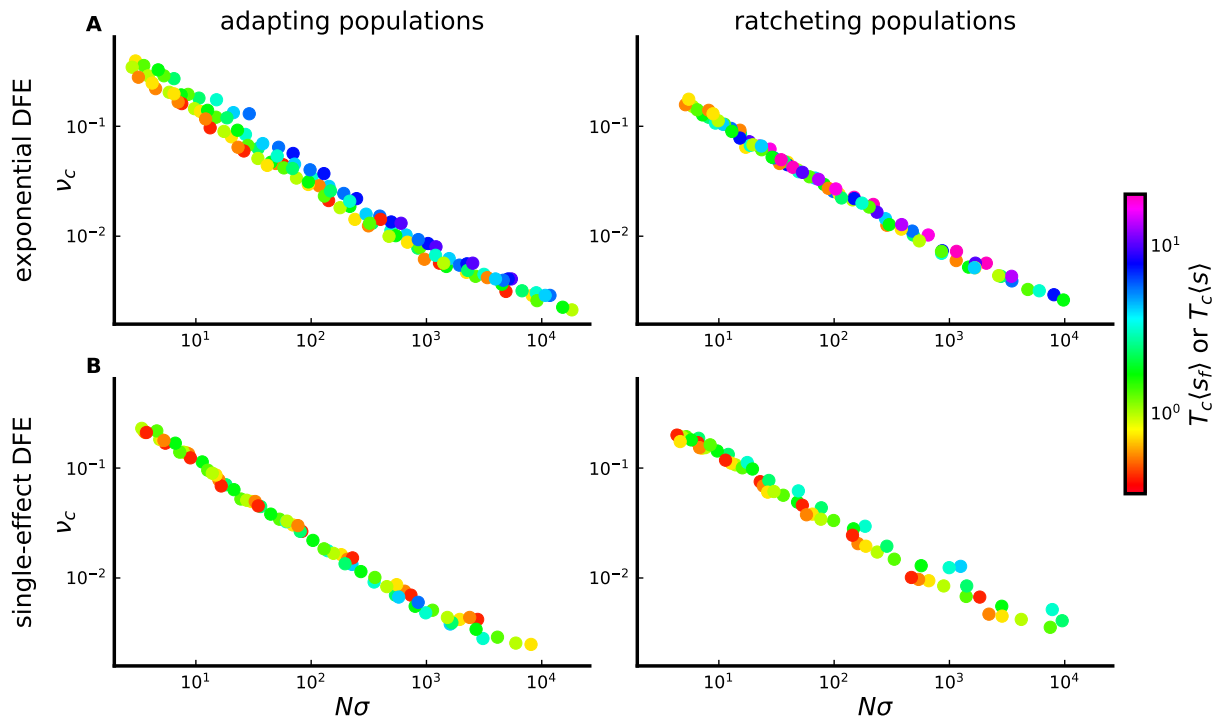

FIG. 6 Dependence of the crossover frequency ν_c on $N\sigma$. Simulated populations in (A) are subject to an exponential DFE, while simulated populations in (B) are subject to a single-effect DFE. For adapting populations (left) points are colored by their values of $T_c\langle s_f\rangle$, while for ratcheting populations (right) points are colored according to their values of $T_c\langle s\rangle$. Values of $N\sigma$ are computed using the MSSM approximation with $\sigma^2 = v - U\langle s\rangle$. For the exponential DFE case, parameters correspond to those on the constrained grid described above and depicted in Fig. S1. For the single-effect DFE case, parameters correspond to the points on a similar constrained grid in the space of $T_c b$ vs. $T_c s$, depicted in the space of NU vs. Ns in Fig. S3. For clarity, we have displayed only points corresponding to simulated populations with $\langle s_f\rangle < 3b$ (such that the MSSM approximation does not break down) and $T_c\langle s_f\rangle > 1/4$ (such that the neutral and selected SFSs differ substantially at high frequencies).

where $B(x,y)$ is the Beta function satisfying $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$. Up to an overall scale factor, $P_0(k|m)$ in Eq. (69) matches the SFS corresponding to the BSC, recently calculated by Kersting et al. (2019) directly from the BSC partition structure. Thus, similar to previous work (Desai et al., 2013; Kosheleva and Desai, 2013; Neher and Hallatschek, 2013), Eq. (69) implies a correspondence between genealogies in the MSSM regime and those of the BSC—at least for aspects of genealogies which determine the average SFS at the moderate to high frequencies for which Eq. (69) is valid. We will refer to the large-m limit of $mP_0(\nu m|m)$ with $P_0(k|m)$ given by Eq. (69) as $h_{\rm neu}^{\rm BSC}(\nu)$, and provide an expression for $h_{\rm neu}^{\rm BSC}(\nu)$ in terms of special functions in Appendix E.

The quantity $h_{\rm neu}^{\rm BSC}(\nu)$ well approximates the actual SFS $H_{\rm neu}(\nu)$ only for ν such that typical observed mutations have ages t satisfying $b(t-T_c) \gg 1$. The time integral yielding Eq. (69) is dominated by times $b(t-T_c) \gg 1$ only when $\log(1/\nu) \ll T_c b$. At lower frequencies such that $\log(1/\nu) \gg T_c b$, mutations with ages $b(t-T_c) = \mathcal{O}(1)$ make a dominant contribution to the SFS. Fortunately at these lower frequencies, fluctuations in the size of a focal lineage make a small contribution to the denominator of Eq. (59). As a result, the SFS can be calculated by considering the marginal distribution of a single lineage, which is encoded by the generating function $\left\langle e^{-zn(t|x)}\right\rangle \approx e^{-\phi_z(x,t)}$. In Appendix A we carry out an analysis of $\phi_z(x,t)$ on fitness scales $\mathcal{O}(b)$ and time scales $\mathcal{O}(1/b)$, extending the previous analysis of $\phi_z(x,t)$ conducted by Fisher (2013). In Appendix E we show how this generating function can be inverted to obtain an approximation for $h_{\rm neu}(\nu)$,

$$h_{\rm neu}^{\rm Airy}(\nu) \equiv -\frac{T_c\mu}{\nu^2(T_c b)^2} \times [\log {\rm Ai}]''\left(-|z_0| + \frac{1}{T_c b}\log\frac{1}{\nu}\right). \quad (70)$$

In the limit $\frac{1}{T_c b}\log\frac{1}{\nu} \ll 1$,

$$h_{\text{neu}}^{\text{Airy}}(\nu) \to \frac{T_c\mu}{\nu^2\log^2(1/\nu)}, \tag{71}$$

which is precisely the asymptotic behavior of $h_{\text{neu}}^{\text{BSC}}(\nu)$ as $\nu \to 0$; thus the two functions $h_{\text{neu}}^{\text{Airy}}(\nu)$ and $h_{\text{neu}}^{\text{BSC}}(\nu)$ have a smooth crossover at $\frac{1}{T_c b}\log(1/\nu) \sim 1$. In principle, our predictions $h_{\text{neu}}^{\text{Airy}}(\nu)$ and $h_{\text{neu}}^{\text{BSC}}(\nu)$ could be connected by asymptotic matching. Notably, $h_{\text{neu}}^{\text{BSC}}(\nu)$ depends on the parameters N and $\mu(s)$ only via the overall scale factor $T_c\mu$, which reflects the overall fixation rate of neutral mutations over the timescale T_c of coalescence. In contrast, $h_{\text{neu}}^{\text{Airy}}(\nu)$ depends separately on the quantities $T_c\mu$ (which again sets an overall scale factor) and $T_c b$. This suggests that for the purposes of distinguishing between (and potentially inferring) evolutionary parameters, an understanding of the low-frequency portion of the SFS is particularly important.

Both $h_{\text{neu}}^{\text{Airy}}(\nu)$ and $h_{\text{neu}}^{\text{BSC}}(\nu)$, however, neglect the contribution of mutations with ages t such that $b(t-T_c)$ is large and negative. To capture the contribution of mutations with ages $t < T_c$, a deterministic approximation is useful. Under such an approximation, the frequency of an observed mutation—and its age—precisely determine the relative fitness of the ancestral background on which the mutation must have arisen. The contribution to the SFS from ages $t < T_c$ can then be obtained by integrating over the times at which a mutation may have occurred, weighted by the corresponding probabilities with which the mutation arose on a background with the respective requisite fitness. A particularly simple approximation, introduced by Neher and Shraiman (2011), can be made by assuming that mutations arise within the "bulk" of the fitness distribution (which is well-described as a Gaussian $f(x) \propto e^{-x^2/2\sigma^2}$, with variance $\sigma^2 = v - U\langle s\rangle$) and by approximating $n(t|x)$—the deterministic lineage size at time t, given a founding background fitness x—as $n_0 e^{xt-\sigma^2t^2/2}$ (where n_0 denotes the size of a lineage upon *establishing*, at which point its lineage begins to grow deterministically, under this approximation). This yields a resulting SFS approximated by

$$h_{\text{neu}}^{\text{Gaussian}}(\nu) \equiv \frac{\mu}{\sigma\nu^2\sqrt{2\log N\sigma\nu}}. \tag{72}$$

In Appendix E, we generalize the above argument to the regime in which mutations arise outside the "bulk" of the fitness distribution, or at times such that the approximation $n(t|x) \approx n_0 e^{xt-\sigma^2t^2/2}$ breaks down by the time the mutation is observed. A key simplification arises because of a relation between the deterministic lineage sizes $n(t|x)$ and the Laplace transform $\tilde{f}(z)$ of $f(x)$. We find that more generally, the deterministic contribution to the SFS is given by

$$h_{\text{neu}}^{\text{Deterministic}}(\nu) \propto \frac{1}{\nu^2}\left.\frac{d\log\nu_{peak}}{dt_{peak}}\right|_{\nu_{peak}=\nu} \tag{73}$$

where $\nu_{peak}(t_{peak})$ is the frequency at which a lineage peaks in size, given its peak size occurs at time t_{peak} (which in turn can be expressed as a function of the lineage's initial relative fitness x); the derivative in Eq. (73) is evaluated at t_{peak} such that $\nu_{peak} = \nu$. The applicability of Eq. (73) reflects the fact that the contribution of any particular lineage to the time-averaged SFS is typically dominated by the time the lineage spends near its peak in size; this intuition has been used to calculate the SFS in the presence of purifying selection, for example (Cvijović et al., 2018). As we show in Appendix E, Eq. (73) can easily be approximated when $st_{peak} \ll 1$ for relevant s in $\mu(s)$, or when $b(t_{peak}-T_c) \ll 1$, and can be evaluated more generally by numerically solving a simple equation for t_{peak} in terms of ν_{peak}. In the limit $st_{peak} \ll 1$ (which occurs for sufficiently low ν), we find that $h_{\text{neu}}^{\text{Deterministic}}(\nu)$ tends to $h_{\text{neu}}^{\text{Gaussian}}(\nu)$ given in Eq. (72), as expected.

Finally, at the very lowest (and highest) frequencies, the SFS is dominated by completely neutral genetic drift. A well-known result is that $h(\nu) = \frac{2N_e\mu}{\nu}$ for a neutrally evolving population with effective population size N_e (Crow et al., 1970); at sufficiently low frequencies, mutations contribute in the same way to the SFS, since selection has not yet had sufficient time to alter their fates substantially (Cvijović et al., 2018). As a rough heuristic, we might expect this result to hold with N_e/N equated to $\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-1/(N\nu)}^{1/(N\nu)} e^{-x^2/2\sigma^2}dx = \text{erf}\left(\frac{1}{\sqrt{2}N\sigma\nu}\right)$, the fraction of individuals in the "bulk" of the fitness distribution which will typically reach a frequency ν in the population *before* establishing (i.e., reach a frequency ν primarily by genetic drift as opposed to by deterministic forces, contingent on reaching a frequency ν). We thus define

$$h_{\text{neu}}^{\text{Drift}}(\nu) \equiv \frac{2N\mu}{\nu}\text{erf}\left(\frac{1}{\sqrt{2}N\sigma\nu}\right), \tag{74}$$

which has a relatively smooth crossover to $h_{\text{neu}}^{\text{Gaussian}}(\nu)$ at $\nu \approx 2/(N\sigma)$, roughly the threshold frequency which most mutations founded in the "bulk" of the fitness distribution will not reach before establishing. A similar argument suggests that $h_{\text{neu}}(\nu) \to 2N\mu$ as $\nu \to 1$, although we have not

worked out a heuristic for the dependence on ν in this limit.

The transitions between these frequency regimes are smoothly varying crossovers. To obtain concrete predictions, it is useful to consider a piecewise approximation,

$$H_{\rm neu}^{\rm PW}(\nu) \equiv \begin{cases} h_{\rm neu}^{\rm Drift}(\nu) & \nu < 2/(N\sigma) \\ h_{\rm neu}^{\rm Airy}(\nu) + h_{\rm neu}^{\rm Gaussian}(\nu) & 2/(N\sigma) < \nu < e^{-T_c b} \\ h_{\rm neu}^{\rm BSC}(\nu) + h_{\rm neu}^{\rm Gaussian}(\nu) & e^{-T_c b} < \nu < 1/2 \\ \min\left[h_{\rm neu}^{\rm BSC}(\nu) + h_{\rm neu}^{\rm Gaussian}(\nu), 2N\mu\right] & \nu > 1/2 \end{cases}, \tag{75}$$

where $h_{\rm neu}^{\rm BSC}(\nu)$, $h_{\rm neu}^{\rm Airy}(\nu)$, $h_{\rm neu}^{\rm Gaussian}(\nu)$, $h_{\rm neu}^{\rm Drift}(\nu)$ are given above. Note that mutations with ages $t < T_c$ (accounted for by $h_{\rm neu}^{\rm Gaussian}(\nu)$) and with ages $t > T_c$ or $b(t - T_c) = \mathcal{O}(1)$ (accounted for by $h_{\rm neu}^{\rm BSC}(\nu)$ and $h_{\rm neu}^{\rm Airy}(\nu)$, respectively) contribute additively to the SFS, which motivates the inclusion of the sums in Eq. (75). $h_{\rm neu}^{\rm Gaussian}(\nu)$ is important primarily for $\nu < e^{-T_c b}$—and over a broad range of $\log \nu$, both $h_{\rm neu}^{\rm Gaussian}(\nu)$ and $h_{\rm neu}^{\rm Airy}(\nu)$ are important—but is retained for $\nu > e^{-T_c b}$ to ensure a smooth piecewise curve. In Fig. 7, we compare the predicted SFSs given by Eq. (75) to neutral SFSs observed in simulations. In the same Figure, we compare simulated *selected* SFSs to predicted selected SFSs obtained using a piecewise-defined function $H_{\rm sel}^{\rm PW}(\nu)$. This function is defined completely analogously to $H_{\rm neu}^{\rm PW}(\nu)$ in Eq. (75), with analogous contributions $h_{\rm sel}^{\rm Drift}(\nu)$, $h_{\rm sel}^{\rm Gaussian}(\nu)$, $h_{\rm sel}^{\rm Airy}(\nu)$, and $h_{\rm sel}^{\rm BSC}(\nu)$ from selected mutations. The only differences are that μ is replaced by $\mu F/U$ in $h_{\rm sel}^{\rm Airy}(\nu)$, $h_{\rm sel}^{\rm BSC}(\nu)$, and in the upper limit to $h_{\rm sel}^{\rm PW}(\nu)$ imposed by $2N\mu$. These replacements are justified because the contributions to $h_{\rm sel}^{\rm Airy}(\nu)$ and $h_{\rm sel}^{\rm BSC}(\nu)$ from a mutation with fitness effect s both involve overall factors of $e^{T_c s}$ (which, integrated over $\rho(s)$, yield a factor F/U). We provide further comparison of $H_{\rm neu}^{\rm PW}(\nu)$ and $H_{\rm sel}^{\rm PW}(\nu)$ with simulated site frequency spectra in Fig. S5, Fig. S6, Fig. S7 and Fig. S8.

Based on these considerations, ν_c should lie at a frequency such that $h_{\rm sel}^{\rm Gaussian}(\nu_c)$ is of comparable magnitude to $h_{\rm sel}^{\rm Airy}(\nu_c)$, and should obey the approximate bound

$$T_c b \lesssim \log\left(\frac{1}{\nu_c}\right) \lesssim \log N\sigma. \tag{76}$$

In Fig. S9 we verify that Eq. (76) is satisfied for the populations shown in Fig. 6 for large $N\sigma$. We leave a more complete analytical description of the crossover frequency ν_c, as well as the precise frequency-dependence of $h_{\rm sel}(\nu)/h_{\rm neu}(\nu)$ at the lowest frequencies $\mathcal{O}\left(\frac{1}{N\sigma}\right)$, for future work.

4. Pairwise heterozygosity and coalescence times

A special case of the neutral site frequency spectrum is the pairwise neutral heterozygosity $\pi_{\rm neu} \equiv P_0(1|2)$. More generally, we can consider the pairwise heterozygosity $P_s(1|2)$ for a single selected site as well as the aggregate heterozygosity $\pi_{\rm sel} \equiv \int P_s(1|2)\rho(s)ds$. With these definitions, $\pi_{\rm neu}$ and $\pi_{\rm sel}$ can both be computed using Eq. (69) and the SFS ratio given in Eq. (67). These expressions, however, are obtained under the assumption that typical contributing mutations have ages $t > T_c$. In reality, $\pi_{\rm neu}$ (as well as, rather generally, $\pi_{\rm sel}$) receives a substantial contribution both from mutations with ages $t > T_c$ and from mutations with ages $t < T_c$. In Appendix E we show, using the approximate behavior of $\Phi_s(z,t)$ for $t > T_c$ and for $t < T_c$, that

$$P_s(1|2)ds \approx 2T_c\mu\left(e^{T_c s} + \frac{e^{T_c s} - 1}{T_c s}\right), \tag{77}$$

where the second term in Eq. (77) can be considered the contribution to $P_s(1|2)$ from mutations with ages $t < T_c$. This term, which can alternatively be written as $2\mu\int_0^{T_c} e^{st}dt$, can be interpreted as follows: mutations with ages $t \ll T_c$ are present at expected frequency e^{st}/N, with a negligible fraction of mutations shared among sampled individuals. We note that the $s \to 0$ case of Eq. (77) reduces to

$$\pi_{\rm neu} = 4T_c\mu \tag{78}$$

Since $\pi_{\rm neu}$ is related to the mean time $\langle T_2\rangle$ to pairwise *coalescence* according to $\pi_{\rm neu} = 2U_n\langle T_2\rangle$, Eq. (78) then implies that

$$\langle T_2\rangle \approx 2T_c. \tag{79}$$

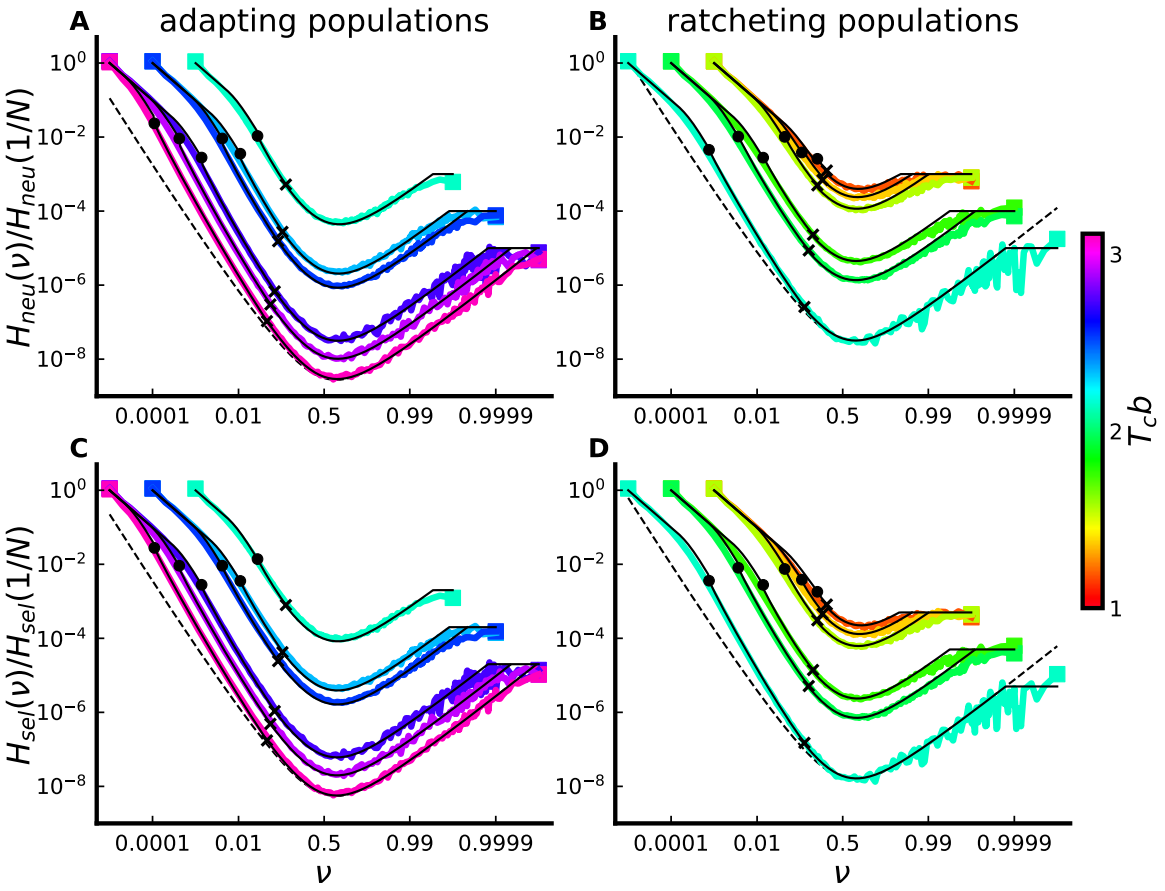

FIG. 7 Site frequency spectrum of neutral mutations (A and B) and of selected mutations (C and D). Colored lines denote SFSs observed in simulations, averaged over at least 800 epochs and smoothed using a moving average box kernel smoother. Solid black lines denote the corresponding theory predictions of the piecewise-defined function in Eq. (75) or its generalization to the selected SFS; for each theory curve, the 'x' marker denotes the point at which $\nu = e^{-T_c b}$, and the circle marker denotes the point at which $\nu = 2/(N\sigma)$. Dashed black lines denote the BSC prediction $h^{\mathrm{BSC}}(\nu)$ for the parameter combination with the largest value of $T_c b$ in each panel (with BSC predictions for other parameter combinations simply shifted by a constant factor). All SFSs and theory curves are normalized by our theoretical prediction for $h(1/N)$, which is $2N^2U$ for selected mutations and $2N^2U_n$ for neutral mutations. In all cases populations are driven by an exponential DFE. Simulated parameters are chosen with $T_c \langle s_f \rangle = 1$ for adapting populations, and with $T_c \langle s \rangle = 1$ for ratcheting populations; in both cases $T_c b$ values are linearly spaced and denoted by the color of each curve. Note that each simulated SFS terminates at the frequencies $\nu = 1/N$ and $\nu = 1 - 1/N$ at which we denote simulated SFS values by square markers.

That is, the defined quantity T_c corresponds with (one-half) the average time $\langle T_2 \rangle$ to pairwise coalescence, motivating our interpretation of T_c as a coalescence timescale. In Appendix E, we obtain the same result by considering the time-dependent pairwise merger probability $Q_2(t) = \Lambda_{2,2}(t)$. Our calculation in Appendix E also yields the *distribution* of times to pairwise coalescence (and relatedly, the distribution of the pairwise neutral heterozygosity); we find the same exponential distribution of pairwise coalescence times, following an initial delay period of time T_c during which coalescence events are negligible, observed by Neher and Hallatschek (2013) in the infinitesimal regime (with a different overall timescale).

In Fig. 8, we compare our predictions in Eq. (77) for π_{neu} and π_{sel} to averages of these quantities measured in simulations. We find good agreement between simulations and our prediction, provided the MSSM approximation conditions of validity are met, though agreement appears to require larger values of $T_c b$ than for the quantities considered in Fig. 4. We

note also that a separate prediction for π can be obtained from the piecewise approximation to the site frequency spectrum in Eq. (75), using the relation $\pi = 2\int h(\nu)\nu(1-\nu)d\nu$. We compare this prediction to the observed values of π_{neu} and π_{sel} in Fig. S10; a similar level of agreement is obtained.

VI. KEY FITNESS SCALES AND TIMESCALES: SIMPLE HEURISTICS

We now turn to summarize the different interpretations that can be given to the quantities T_c, b and c defined above, and provide a heuristic description of the dynamics within the MSSM regime. We begin by noting that the distribution $Nf(x)w(x)$ of future common ancestors fitnesses is peaked with width $\Delta x_f \sim \mathcal{O}(b)$ around $x_f^* \approx c+b \sim c$. This is our motivation for defining the region $x = c+\mathcal{O}(b)$ as the *fixation class*, since collectively, this region (of width approximately $|z_0|b$) fixes with probability $\mathcal{O}(1)$, despite its small size relative to the total population. In Appendix C, we show how our approximate solutions $f(x)$ and $w(x)$ in Eq. (25) and Eq. (26) can be considered local approximations, valid precisely around $x = c$; we obtain a condition of validity for the MSSM regime by ensuring the region of validity of this *local* approximation encompasses the entire fixation class.

1. Competition among the fixation class, and the fates of "doomed" lineages

Our condition of validity for the MSSM regime, expressed in Eq. (34), essentially requires that $s \ll \Delta x_f$ throughout the region dominating $\int s^2\rho_f(s)ds$. In this sense, mutational effects can be considered "infinitesimal" in the MSSM regime: the typical fixed fitness effect $\tilde{s}$ must be much smaller in magnitude than the range of fitness space which typically produces a future common ancestor (i.e., the width of the fixation class). Dynamically, this means that individuals routinely fix despite having fitness disadvantages, compared to the most-fit individuals in the population, of several times (up to $\Delta x_f/\tilde{s}$ times) the typical fixed fitness effect $\tilde{s}$. These individuals catch up and fix not by acquiring a single large-effect beneficial mutation, but rather by rapidly acquiring several mutations (and/or by *avoiding* deleterious mutations) so as to "leapfrog" above their more-fit competitors. Less-fit individuals are exponentially less likely to fix—with $1/T_c$ the fitness scale on which fixation probabilities vary—but there are also exponentially more of such individuals; over the fitness scale $\Delta x_f \approx b$, the two exponential factors cancel, leaving a relatively broad range of background fitnesses which routinely supply a future common ancestor. Intuitively, it makes sense that this behavior is obtained for high mutation rates (relative to the size of selective effects), but as we have shown, this behavior also arises for sufficiently large population sizes (as long as $\rho(s)$ falls off faster than exponentially with large, positive s). The infinitesimal regime constitutes a special case of the MSSM regime in which mutations are "infinitesimal" in a more narrow sense: in the infinitesimal regime, selection acts in a negligible way on any individual mutation.

In contrast, in the "moderate-speeds" regime, $\Delta x_f \sim \tilde{s}$: nearly all future common ancestors come from within one "predominant" effect $\tilde{s}$ of the nose (Fisher, 2013; Good and Desai, 2014). The lineages of individuals lying more than one multiple of $\tilde{s}$ below the nose are essentially "doomed" to eventually go extinct, and the chance with which these individuals leapfrog above their more-fit competitors can be ignored. As a result, in analyzing the "moderate-speeds" regime several simplifications can be made. For instance, in treating a population as a set of discrete fitness classes (separated by predominant-effect mutations), mutations are only important only via their potential to establish a new "lead" fitness class (Desai and Fisher, 2007), and can thus be ignored in fitness classes below the current lead class. Once a new lead class is established, the frequencies of lineages, within the previous lead class, can be treated as "frozen". Each prior lead class will grow in size—while losing relative fitness at rate v—and will eventually comprise a large fraction of the total population, and at that point the frequencies of individual lineages within the class will have changed negligibly. This approximation has been used to simplify calculations of genetic diversity statistics and sojourn/fixation times of mutations (Desai et al., 2013; Kosheleva and Desai, 2013).

In the MSSM regime, it is less useful to model a population as a set of discrete fitness classes, since lineage frequencies are at no point "frozen" within fitness classes. Instead, lineages continue to acquire mutations, and *spread out* in relative fitness as they fall behind the nose. To connect the dynamics of the fixation class with the dynamics of the bulk—which we have done by ensuring $f(x)$ is appropriately normalized, in enforcing the condition $1/N = \int f(x)w(x)dx$—accounting for this spread is important. Using a deterministic approximation (carried out in Appendix F) we can gain further intuition on the dynamics of lineages as they fall be-

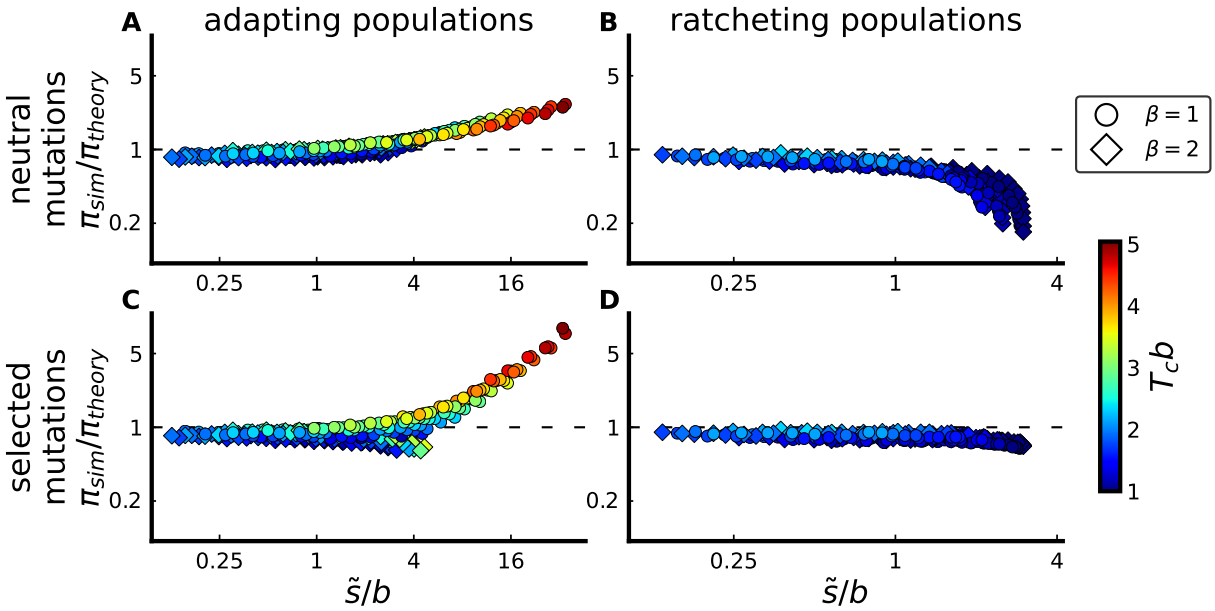

FIG. 8 Comparison between simulated and predicted heterozygosity π of neutral mutations (A and B), and between simulated and predicted heterozygosity π of selected mutations (C and D), for adapting populations (A and C) and ratcheting populations (B and D). Parameters are identical to those simulated in Fig. 4.

hind the nose. In the special case of the infinitesimal regime, we review how lineage-wide relative-fitness distributions evolve according to a *reaction-diffusion* equation with diffusion constant $D \equiv \frac{1}{2}U\left\langle s^2\right\rangle$, advection rate $-\sigma^2$, and local growth rate x. The infinitesimal regime has thus long been recognized as a limit of "mutational diffusion" (Tsimring et al., 1996); along the line of descent, fitness follows a biased random walk with diffusion constant D (Neher et al., 2014a). A schematic depiction of how fit lineages spread out in relative fitness over time is provided by Neher and Hallatschek (2013). More generally in the MSSM regime, we can see that a diffusion approximation may *not* be adequate in describing the trajectories of lineages. In particular if $T_c\tilde{s} > 1$ then at later times, further mutations acquired by the lineage will have begun to shape its fitness distribution in a nondiffusive way (i.e. its fitness variance will grow faster than linearly with time).

2. Interpretations of the timescale T_c

A key result of the Section V is that $T_c \approx \langle T_2\rangle/2$. This motivates our interpretation of T_c (as defined in Eq. (17)) as a *coalescence* timescale. The timescale T_c can also be interpreted in a few different ways. In Appendix F, we show that under a deterministic approximation, the descendants of the fixation class—initially consisting of $N\int_c^{x_c} f(x)dx$ individuals—will collectively sweep through and comprise an $\mathcal{O}(1)$ fraction of the population at T_c generations. This motivates us to interpret T_c as a *sweep* timescale. In the "high-speeds regime," a related interpretation has been given to T_c as the time required for fluctuations near the high-fitness nose of the fitness distribution to substantially affect its bulk dynamics (Fisher, 2013).

Since the (linearized) operators governing the forward-time and backward-time dynamics are adjoints of one another, it is not particularly surprising that the timescale T_c can also be interpreted as a *delay* timescale of coalescence: looking backwards in time, a given pair of individuals is unlikely to coalesce until T_c generations have elapsed. We can understand this correspondence heuristically by viewing the fixation class as an exponentially expanding subpopulation, among the total population, within which all coalescence events must occur (since the fixation class is destined to eventually fix). It is well known that in rapidly *expanding* populations, coalescence events occur primarily at the very beginning of exponential growth (i.e., when the population was small in size) and that genealogical trees are starlike, with long terminal branches (Slatkin and Hudson, 1991); thus, looking backward in time, we expect few coalescence events to occur until near the end of

the sweep timescale.

We can also gain intuition on the delay timescale T_c by considering, for a randomly chosen individual, the distribution $A(x,t)$ of its ancestor's relative fitness x as the time t recedes into the past. As we describe in Appendix A,

$$A(x,t) \propto f(x)\phi_{1/N}(x,t), \tag{80}$$

with $\lim_{t\to\infty} A(x,t) = Nf(x)w(x)$ corresponding to the distribution of fitnesses of *eventual* common ancestors described above. Importantly, we can see that at T_c generations into the past, the distribution $A(x,T_c) \sim \mathrm{Ai}\left(\frac{c-x}{b}\right)$ comes to resemble (and, subsequently, rapidly converges to on a timescale $\mathcal{O}(1/b)$) the eventual distribution of ancestor fitnesses $Nf(x)w(x) \propto \mathrm{Ai}^2\left(\frac{c-x}{b}\right)$. Thus, the ancestors of typical individuals (which start out near "bulk" of the fitness distribution) migrate upward in relative fitness and reach the fixation class on the timescale T_c. After this point, the ancestors of a typical sample of individuals begin to coalesce *within* the fixation class. This matches the interpretation given by Neher and Hallatschek (2013) for the delay timescale in the infinitesimal regime, using a heuristic argument: in that case, ancestors migrate upwards in fitness at initial rate σ^2, slowing down at later times to reach a fitness $\sigma^4/4D$ at the delay timescale $\sigma^2/2D$.

It is not entirely clear why the delay timescale should match the coalescence timescale of the ensuing BSC process—that is, why coalescence *within* the fixation class requires approximately the same amount of time as that required to *reach* the fixation class. The correspondence between these two timescales appears to be a relatively universal feature of rapidly evolving populations, observed in the moderate-speeds regime (Desai et al., 2013; Kosheleva and Desai, 2013) and the infinitesimal regime (Neher and Hallatschek, 2013). However, this is not generically the case: for instance, in the adapting populations modeled by Brunet et al. (2007) as FKPP waves, the delay timescale is much shorter than the coalescence timescale, even though the genealogies of those populations are described by the BSC. The key difference is that in the model considered by Brunet et al. (2007), the growth rate of a lineage does not depend linearly on its fitness advantage x. Instead, all individuals lying above minimum fitness cutoff survive until the next generation, and individuals therefore have a reduced benefit of being much fitter than average.

3. Conditional neutrality at long times

The simple exponential dependence $p_{\mathrm{fix}}(s) \propto e^{T_c s}$ suggests that in the MSSM regime, selection acts in a substantial way to amplify the frequency of a mutation only over the timescale T_c. This in turn suggests a picture of *conditional neutrality* in the fates of mutations at times $t > T_c$. This conditional neutrality can also be seen through our result for the SFS of selected mutations. As we have seen, the SFS of selected mutations with effects between s and $s + ds$ (and ages $t > T_c$) is simply scaled by an overall factor $e^{T_c s}$, relative to the neutral SFS (as well as a scaling factor reflecting the difference in the mutation rates). Thus, we can think of selection acting to amplify the probability with which a mutation is observed only over timescale T_c. At least up to the information contained in the SFS, frequency trajectories of selected mutations at times $t > T_c$ are indistinguishable from those of neutral mutations with the same frequency at time T_c. We note that a *stronger* picture of conditional neutrality is seen in the infinitesimal regime. In the infinitesimal regime, $T_c \ll 1/s$ for relevant s, so $p_{\mathrm{fix}}(s) \approx e^{T_c s}/N$ reduces to $1/N$ (plus small corrections). Even during the initial period $t < T_c$, the fate of a typical mutation (or even of a typical *fixed* mutation) is influenced by its selective effect in a negligible way. In the MSSM regime, the selective effect of a mutation can influence its fate in a substantial way, but only during the first T_c generations of the mutation's lifetime.

At a heuristic level, it makes sense that T_c is the relevant timescale on which selection acts. To see this, we note that looking backward in time, a given pair of individuals which happen to be sampled from the fixation class will coalesce at average time T_c (because they are sampled from the fixation class, they skip the initial delay period). Looking forward in time, then, some individual will have grown to comprise a macroscopic fraction of the fixation class on the same timescale T_c. Note that individuals do not *fix* within the fixation class over the timescale T_c—fixation requires a time $\mathcal{O}(\log\log N)$ multiples of T_c, given the distributions of times to coalescence of a large sample under the BSC (Berestycki, 2009; Desai et al., 2013). At the timescale T_c, however, lineages originally founded in the fixation class (or more precisely, the portions of these lineages still *in* the fixation class) have spread out enough in fitness so that initial fitness differences of size $s \ll b$ have been "forgotten"; selection thus no longer acts to amplify the frequency of a mutation which occurred in that region. Because the fixation class will even-

tually take over the population, this implies conditional neutrality in the full population on the same timescale.

A similar period of conditional neutrality—after an initial delay—is noted by Kosheleva and Desai (2013) in the "moderate-speeds" regime. In that case, Kosheleva and Desai (2013) note that a mutation observed at a macroscopic frequency x_0—large enough to essentially guarantee the mutation occurred in the "nose" class—will go on to fix with the same probability x_0. This is explained by the fact that lineage frequencies are "frozen" within fitness classes in that case, so that a mutation currently at macroscopic frequency x_0 in the population was once present in the "nose" class at the same frequency x_0, a delay time T_d ago. The fraction of the nose class occupied by any given lineage evolves *neutrally* as a Markov process (in that the *future* fraction of the nose class occupied by any given lineage depends only on the fraction of the nose class it *currently* occupies) and so a period of conditional neutrality emerges after the delay timescale T_d.

In the MSSM regime, the same argument does not quite apply, even upon replacing the "nose" class with the fixation class. The key difference is that the fraction of the fixation class occupied by a given lineage does not evolve as a Markov process: fitness differences among individuals in the fixation class are important in predicting their eventual fates (since $T_c b \gg 1$). However, the same argument *does* hold if we instead consider the *effective lead frequency* ν_L of a lineage, defined as

$$\nu_L(t) \equiv N \int g(x,t) w(x) dx, \tag{81}$$

where $g(x,t)$ is the relative fitness distribution of the lineage (normalized such that its total size at time t equals $N \int g(x,t)dx$). The generating function $\left\langle e^{-\theta \nu_L(t)} \right\rangle$ is considered by Fisher (2013)—though in that work, $g(x,t)$ in Eq. (81) is primarily taken as a fitness distribution of an entire *population* whose total size can fluctuate. In Appendix E, we briefly review key features of $\left\langle e^{-\theta \nu_L(t)} \right\rangle$ discussed by Fisher (2013). An implication of these results is that the frequency $n(t)/N$ of a lineage in a population closely mirrors its effective lead frequency $\nu_L(t - T_c)$ a time T_c into the past (provided the lineage was founded in the fixation class). Furthermore, the effective lead frequency $\nu_L(t)$ evolves as a neutral Markov process, in that $\left\langle e^{-\theta \nu_L(t)} \right\rangle$ can be obtained from ν_L at any earlier time (i.e., the distribution of ν_L at later times is mediated by its current value, and does not explicitly depend on the full lineage-wide distribution $g(x,t)$). Together these results give rise to a type of conditional neutrality at long times similar to that observed by Kosheleva and Desai (2013).

Using the generating function $\left\langle e^{-\theta \nu_L(t)} \right\rangle$, we can obtain the *transition density* $G_t(\nu_L(t)|\nu_L(0))$—the probability a lineage has effective lead frequency $\nu_L(t)$ at time t, given its effective lead frequency $\nu_L(0)$ at time 0—with the result

$$G_t(\nu_L(t), \nu_L(0)) \approx \frac{\sin\left[\pi\alpha(t)\right] \nu_L(0)\left(1-\nu_L(0)\right)}{\pi \nu_L(t)\left(1-\nu_L(t)\right)} \times \frac{1}{\left[\left(1-\nu_L(0)\right)^2 \left(\frac{\nu_L(t)}{1-\nu_L(t)}\right)^{\alpha(t)} + \nu_L(0)^2 \left(\frac{1-\nu_L(t)}{\nu_L(t)}\right)^{\alpha(t)} + 2\nu_L(0)(1-\nu_L(0)) \cos\left(\pi\alpha(t)\right)\right]}, \tag{82}$$

where $\alpha(t) \equiv e^{-t/T_c}$. We carry out this calculation in Appendix E. Up to a change in timescale, $G_t(\nu_L(t), \nu_L(0))$ in Eq. (82) is the same as the transition density for the *actual* lead frequency (i.e the frequency in the "nose" class) found in the "moderate-speeds" regime (Desai et al., 2013; Kosheleva and Desai, 2013). The distributions of sojourn times found by Kosheleva and Desai (2013) using the transition density thus carry over to the MSSM regime as well, after making an overall change in the timescale, and replacing the actual lead frequency with the effective lead frequency. Interestingly, the same transition density is obtained by Hallatschek (2018) in a purely neutral model of a population with a $1/n^2$ offspring number distribution from one generation to the next (and which can be considered a forward-time dual to the BSC). As discussed by Hallatschek (2018), this transition density manifests as an apparent *frequency-dependent selection*—a typical bias favoring majority alleles is balanced by rare compensating jumps of low-frequency alleles to maintain overall neutrality. We

thus expect to see a similar apparent frequency-dependence among mutational trajectories in the MSSM regime at long times.

Upon incorporating selection into the model described above, Hallatschek (2018) finds the same exponential dependence $p_{\text{fix}}(s) \propto e^{T_c s}$ that we have shown is obtained in the MSSM regime. In the model considered by Hallatschek (2018), a mutation with effect s is shown to effectively increment the logit frequency $\log[\nu/(1-\nu)]$ of a lineage by an amount s (i.e., a lineage which just acquired a mutation with effect s behaves identically to a neutral lineage whose logit frequency is larger by an amount s). In our case, a mutation of effect s increments $\log \nu_L$ by an amount $T_c s$, provided it occurs within the fixation class (and has effect $s \ll b$). Thus, the model considered by Hallatschek (2018) replicates quite well the long-time dynamics of the MSSM regime, coarse-grained on a timescale of T_c generations. A key difference, however, is that in our model, the timescale T_c—the length of an effective generation—is shaped by selection in a complicated way; in the model considered by Hallatschek (2018), T_c is a fixed input parameter, independent of the action of selection.

The above results emphasize that while the effective frequency of a mutation in the fixation class evolves as a neutral Markov process, it does *not* evolve as a neutral Wright-Fisher diffusion process. Thus, even during the conditionally neutral period, we do not expect mutational trajectories to resemble those of a purely neutral Wright-Fisher population—we only expect that lineages have "forgotten" their initial fitness effect. This can be contrasted with a type of quasi-neutrality discussed by Cvijović et al. (2018) in a model of strong purifying selection ($|T_c s| \gg 1$). In that model, the fitness advantage of individuals in the "lead" class is balanced, on average, by the rate of mutations *out* of the "lead" class. As a result, fluctuations in lineage frequencies within the "lead" class closely resemble those in a purely neutral Wright-Fisher population. The same fluctuations, smoothed on a certain timescale, are then mirrored by the fluctuations in overall mutational frequencies, after a delay period. Relatedly, the SFS for this model has a "quasi-neutral" region scaling as $1/\nu$ for moderate frequencies ν, and differs from the SFS of the BSC in important ways (Cvijović et al., 2018); in contrast, in the MSSM regime deviations from the SFS of the BSC are only observed at very low frequencies with $\log(1/\nu) > T_c b$ (and to some extent, at the very highest frequencies). This difference highlights the fact that the presence of multiple selected mutations in a population at once—which is the case for the model considered by Cvijović et al. (2018)—is not sufficient to give rise to the seemingly universal correspondence to the BSC we have described above. Rather, the struggle among fit lineages to increase fitness through new mutations, and the jackpot dynamics this gives rise to, appear to be important features in giving rise to this correspondence.

VII. DISCUSSION

As we have seen, evolutionary dynamics within asexual genomes can be complex, even within the simplest models that include only the effects of mutations, natural selection, and genetic drift. The central difficulty is that when the mutation rate is sufficiently high and the population size is sufficiently large, multiple selected mutations often segregate simultaneously and their dynamics are not independent. We refer to this scenario as *rapid evolution*, because evolution is not primarily limited by the waiting time for new mutations to arise. Instead, numerous mutations arise in a variety of linked combinations, and selection can only act on these combinations as a whole. The resulting complex dynamics of clonal interference and hitchhiking can limit the efficiency of natural selection, and dramatically alter evolutionary dynamics and population genetics.

We note that rapid evolution does not necessarily have to involve adaptation. The key components of rapid evolution are simply that numerous selected mutations segregate simultaneously within a linkage block—such that the population maintains substantial variation in fitness—and that the population moves through fitness space over time as selected mutations arise and fix. It can therefore involve both beneficial and deleterious mutations, and in particular can result when the accumulation of beneficial and deleterious mutations balances so that $v = 0$, and the population on average neither increases or decreases in fitness (Goyal et al., 2012). It can even occur in scenarios where only deleterious mutations are possible (and hence the rate of change in mean fitness, v, will be negative), as long as deleterious mutations routinely fix (Neher and Hallatschek, 2013).

In the past two decades, many authors have analyzed evolutionary dynamics in rapidly evolving populations using *traveling wave* models. However, previous work on these models has largely been focused on two limiting cases: the case in which selection is *strong* on single mutations (the "moderate-speeds" regime and the "high-speeds" regime), and the case

in which selection is *weak* on single mutations (the infinitesimal limit). These two limits correspond to the cases where $T_c\tilde{s} \gg 1$ and $T_c\tilde{s} \ll 1$ respectively, where T_c is the coalescence timescale and $\tilde{s}$ is the typical fitness effect of a fixed mutation. In other words, this previous work has assumed a strong separation of scales between the timescale T_c on which common ancestry is determined, and the timescale $1/\tilde{s}$ on which selection can act on a typical fixed mutational effect $\tilde{s}$.

In reality, however, any population is likely to experience mutations with a wide range of selective effects, including many in the intermediate regime between these two extremes. Our lack of understanding of the dynamics of those mutations for which $T_c s \sim 1$ thus represents an important gap. This is particularly problematic because it is natural to expect that "nearly neutral" mutations with effects on the order of the inverse coalescence timescale (i.e. for which $T_c s \sim 1$) may have the largest impact on patterns of genetic diversity (they are strong enough that their effects are felt, but not so strong that they immediately sweep or are purged) (Akashi et al., 2012; Ohta, 1973). Furthermore, recent theoretical work has found, in a model coupling both interference and local epistasis, that the prevalence of mutations with $T_c s \sim 1$ may be an emergent property of the evolutionary process (Held et al., 2019).

The expectation that mutations confer selective effects on a wide range of scales is broadly consistent with numerous empirical studies that have attempted to infer distributions of selection coefficients in natural populations based on population genetic data (Eyre-Walker and Keightley, 2007). These studies typically do not infer selection strengths directly, but rather the product $T_c s$ (Sawyer and Hartl, 1992; Sawyer et al., 2003) (this is often referred to as $N_e s$ under the assumption that T_c corresponds to an effective population size, but we have avoided this terminology since the dynamics differ in many important ways from those of a neutral population with an appropriately sized effective population size (Neher, 2013)). Many of these studies find mutations with $T_c s \sim 1$ are quite prevalent and comprise a large proportion of fixed mutations. This includes mutations involved in viral and mitochondrial DNA (Nielsen and Yang, 2003), amino-acid substitutions in *Drosophila* (Sawyer et al., 2007), synonymous mutations affecting codon usage in *E. coli* (Hartl et al., 1994) and *Drosophila* (Akashi, 1995; Machado et al., 2020; Zeng and Charlesworth, 2009), as well as among mutations occurring within animal mitochondria (Nachman, 1998).

It has remained unclear whether these results actually imply that typical selective coefficients are of order $1/T_c$, or whether emergent aspects of the evolutionary dynamics tend to generate patterns of variation that are most sensitive to the subset of mutations in this regime. In addition, because these inference approaches typically assume free recombination (Bustamante et al., 2001; Sawyer and Hartl, 1992; Sawyer et al., 2003), it is unclear whether interference may confound these results (see e.g. McVean and Charlesworth (2000)).

These considerations highlight the importance of understanding the evolutionary dynamics and population genetics of rapidly evolving populations in cases where a relatively broad range of selective effects is relevant, including effects s with $T_c s \sim 1$. In this work, we make progress in this direction by extending existing methods to apply within a broader regime of the population-genetic parameter space, which we refer to as a *moderate selection, strong mutation* (MSSM) regime. In the MSSM regime, no assumption is made about the magnitude of $T_c\tilde{s}$; instead, we require that $T_c \Delta x_f \gg 1$ and $\tilde{s} \ll \Delta x_f$, where Δx_f is the standard variation of fitness advantages among individuals which eventually take over the population. The first of these conditions is also required within the infinitesimal regime, and implies that selection is strong among *haplotypes* competing for fixation. The second of these conditions differs from the condition $T_c\tilde{s} \ll 1$ required within the infinitesimal regime. Instead of requiring that typical fixed mutational effects are weak (and fix essentially neutrally), within the MSSM regime we require only that typical fixed mutational effects are weak compared to the scale of fitness variation among potential future common ancestors. Qualitative features of the MSSM and other regimes are summarized in Table III.

In Fig. 9, we provide a phase diagram depicting the MSSM and other regimes, where for concreteness we assume the special case of a single beneficial effect, parameterized by the dimensionless quantities NU_b and U_b/s_b. In Table IV, we provide the (approximate) minimum N required for validity of the MSSM approximation, for 4 representative sets of parameters (mutation rate U_b and single beneficial effect s_b). This helps illustrate examples of relevant populations for which the MSSM regime may be of particular importance. For example, while validity of the MSSM approximation for wild-type yeast with strong selection may require an unreasonably large population size, this regime is more plausible for RNA viruses, mutator bacteria or yeast, or wild-type yeast with weaker selection. In general, however, our knowledge of the effect size distributions

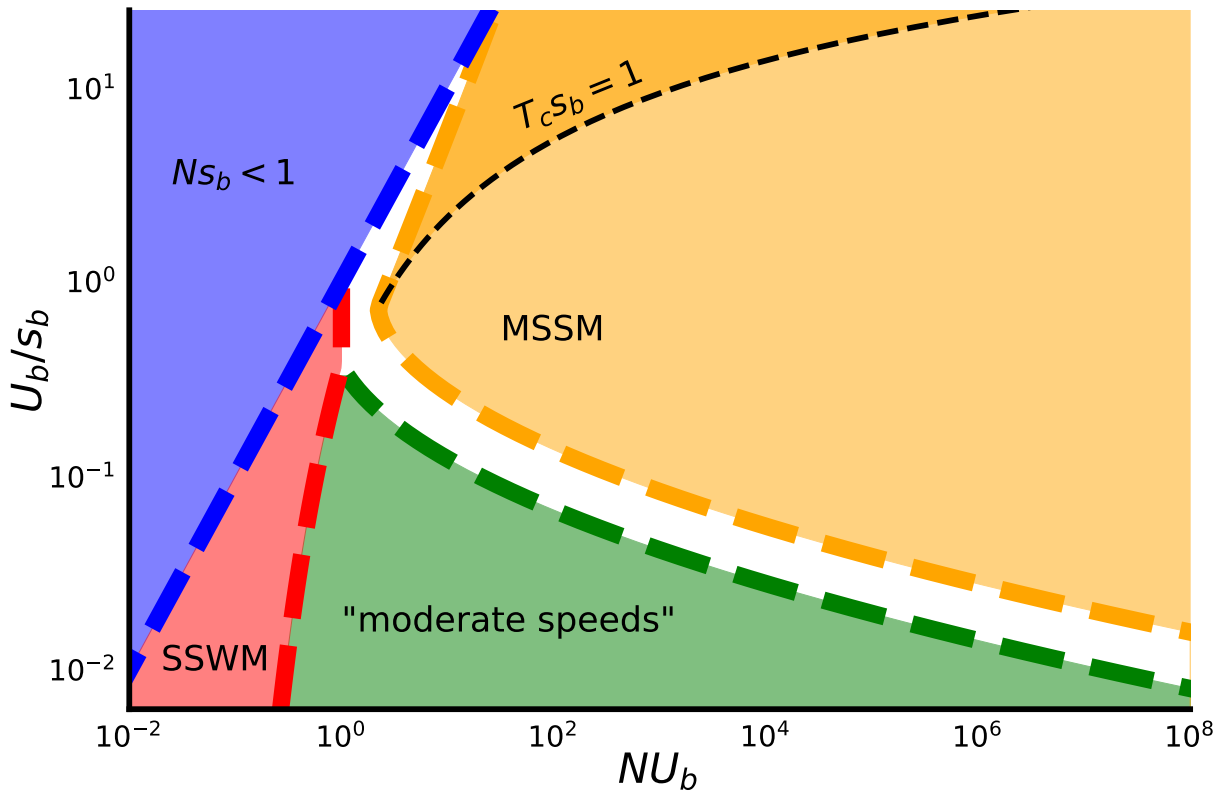

FIG. 9 Phase diagram for the case of a single beneficial effect s_b, in the space of dimensionless parameters U_b/s_b and NU_b. The blue line illustrates the $Ns_b = 1$ boundary which separates the neutral regime (blue region) from the strong-selection weak-mutation (SSWM) regime (red region). The red line illustrates the $NU_b \log(Ns_b) = 1$ boundary which separates the SSWM regime from the "moderate speeds" regime (green region). The green line illustrates the other boundary of the "moderate speeds" regime, at which $\log^2(s_b/U_b) = \log(Ns_b)$ (Fisher, 2013; Good and Desai, 2014). The orange line denotes the boundary of the MSSM regime at which $s_b = b$ (the portion of orange line below the dotted line, where $s_b < b$ is the limiting condition of validity of the MSSM regime) or $T_c b = 1$ (the portion of orange line above the dotted line, where $T_c b > 1$ is the limiting condition). The black dotted line shows where $T_c s_b = 1$, with the infinitesimal regime the region shaded darker orange and the rest of the MSSM regime (including the "high speeds" regime) the region shaded lighter orange.

in specific empirical systems remains rather limited. We also note that the quantities T_c, Δx_f and $\tilde{s}$—which determine validity of the MSSM regime as well as qualitative aspects of the dynamics—can at least in principle by probed experimentally. For example, T_c could be estimated using the typical fixation time of a new mutation as measured from mutational frequency trajectories, $\tilde{s}$ could be measured by assaying the fitness effects of fixed mutations, and Δx_f could obtained by conducting relative fitness assays on the individuals which eventually fix.

Our results are not a complete solution to the problem, since our analysis does make other assumptions, and in particular it does not apply to populations in the "moderate-speeds" regime. However, in combination with earlier work, our analysis helps to provide a more complete picture of how mutations with effects on a wide range of scales shape the evolutionary dynamics of rapidly evolving populations. By explicitly demonstrating an approximate correspondence between genealogies in the MSSM regime and those of the BSC, our analysis further supports previous claims that certain aspects of the evolutionary dynamics of rapidly evolving populations are relatively universal (Neher and Hallatschek, 2013). Given this apparent universality, we expect that many of our qualitative conclusions (e.g. for selected and neutral site frequency spectra) may apply in rapidly evolving populations more generally. With this in mind, the MSSM regime is a particularly useful regime of the parameter space to study: as in the infinitesimal regime, the evolutionary dynamics are

	Moderate-selection strong-mutation (MSSM) regime	"Moderate speeds" regime	Strong-selection weak-mutation (SSWM) regime	Nearly neutral limit
Broad steady state distribution of fitnesses ($\sigma \gg s$)	Yes	Yes	No	No; Broad distribution of mutational classes possible.
Strong selection on single mutations ($T_c s \gg 1$)	Possible, but not required. Not possible in the infinitesimal regime; required in the "high speeds" regime.	Yes	Yes	No
Strong selection on haplotypes ($T_c \sigma \gg 1$)	Yes	Yes	N/A; Transient strong selection on established lineages.	No
Substantial mutational leapfrogging ($\Delta x_f \gg s$)	Yes	No	No	No; Leapfrogging across mutational classes possible.

TABLE III Qualitative Properties of Different Regimes

TABLE IV Example Parameters

Mutation Rate U_b (per genome, per generation)	Fitness Effect s_b	Minimum N Required	Possible Representative System
10^{-3}	10^{-2}	10^6	RNA viruses
10^{-4}	10^{-3}	10^7	Mutator bacteria or yeast
10^{-5}	10^{-4}	10^8	Wild-type yeast (weak selection)
10^{-5}	10^{-3}	10^{15}	Wild-type yeast (stronger selection)

relatively tractable, even for a full distribution of fitness effects, but unlike in the infinitesimal regime, the dynamics of neutral mutations and of selected mutations differ in a substantial way.

In particular, in this work we have computed the selected and neutral site frequency spectra in the MSSM regime, from which predictions for dN/dS, π_N/π_S and related statistics can readily be obtained. While these quantities are used extensively to infer the strength and presence of selection in natural populations, our analytical understanding of these quantities has previously remained limited when linked selection is widespread. These quantities are considered by Kosheleva and Desai (2013) in the "moderate-speeds" regime, although that analysis is limited to the case in which mutations each confer a single strongly beneficial effect s_b (with $T_c s_b \gg 1$). Our present results allow us to calculate these polymorphism and divergence statistics for full distributions of beneficial or deleterious fitness effects, including those with mutations near $T_c s \sim \mathcal{O}(1)$. This revealed that the MSSM regime can produce dramatic departures from existing intuition based on independently evolving sites.

At sufficiently low frequencies, the ratio between nonsynonymous and synonymous site frequency spectra, $h_N(\nu)/h_S(\nu)$, approaches the ratio of the underlying mutation rates, as expected under neutrality. For deleterious mutations, the frequency scale at which the shape of $h_N(\nu)$ starts to deviate from $h_S(\nu)$ is often used as a rough estimate of $1/|T_c s|$. Here, we have found that in the MSSM regime, this transition does not occur for frequencies $\nu \sim 1/|T_c s|$, but instead depends strongly on population-level quantities such as $T_c \Delta x_f$ or $N\sigma$. This suggests that naive estimates of $T_c s$ based on deviations of synonymous from nonsynonymous site frequency spectra may severely overestimate the underlying selection strengths.

We have also shown that over a broad range of higher frequencies, nonsynonymous and synonymous frequency spectra are again related by a constant

factor, equal to the ratio of the fixation rates of the two types of mutations (F_N and F_S). As a consequence, the frequency-resolved McDonald-Kreitman statistic,

$$\alpha(\nu) \equiv 1 - \frac{F_S}{F_N}\frac{h_N(\nu)}{h_S(\nu)}, \tag{83}$$

approaches 0 for moderate to large frequencies ν, even for large $|T_c s|$, and regardless of the rate of adaptation v. This quantity has previously been used to estimate the fraction of fixed mutations which were strongly beneficial, based on the value to which $\alpha(\nu)$ asymptotes at large frequencies (Messer and Petrov, 2013). Our present results suggest that in the MSSM regime, this "asymptotic alpha" approach may severely underestimate the actual fraction of adaptive substitutions, and might even give the impression that a population is evolving nearly neutrally.

Throughout our analysis, we have assumed that the DFE $\mu(s)$ does not change as the population evolves. This will be true provided that there is no *microscopic epistasis* between individual mutations, such that the fitness effect of one mutation depends on the presence or absence of the other. However, our analysis can also apply even in the presence of extensive microscopic epistasis between individual mutations, provided that the overall DFE $\mu(s)$ does not vary across genotypes — that is, provided there is no *macroscopic epistasis* (Good and Desai, 2015).

Recent experimental work suggests that both microscopic and macroscopic epistasis are widespread, at least in the evolution of laboratory microbial populations (Jerison and Desai, 2015). For example, several recent studies have found general patterns of diminishing returns epistasis, where fitness effects of beneficial mutations systematically decline in more-fit genetic backgrounds (Kryazhimskiy et al., 2014). Recent work has also shown an analogous pattern where the fitness costs of deleterious mutations become more severe in more-fit genetic backgrounds (Johnson et al., 2019). These patterns of macroscopic epistasis suggest that $\mu(s)$—and particularly $\rho(s)$—may change substantially and systematically as a population evolves.

Our analysis, like most previous work on traveling wave models, does not directly address these effects of changing $\mu(s)$. However, provided that $\mu(s)$ changes slowly compared to timescale T_c on which mutations sweep through the population, we expect our analysis to provide an accurate description of the evolutionary dynamics at any given moment (given the appropriate $\mu(s)$ at that moment). Thus we can potentially model the effects of diminishing returns and increasing cost epistasis (or any other systematic variation in $\mu(s)$), provided only that these changes in $\mu(s)$ are sufficiently slow. Analogously, our approach does not explicitly consider situations where changes in environmental condition lead to temporal fluctuations in the DFE (e.g., a time-varying fitness *seascape* (Agarwala and Fisher, 2019; Schiffels et al., 2011)). However, provided that these temporal fluctuations in $\mu(s)$ are sufficiently slow, our approach will appropriately describe the evolutionary dynamics at any given moment. Of course, if $\mu(s)$ changes rapidly, either due to dramatic epistatic effects or environmental shifts, a transient period may occur where the traveling wave has not reached its steady-state shape. These transient dynamics are understood only coarsely at a theoretical level (Fisher, 2013), and we have not analyzed their effects here. If these shifts are sufficiently common that the transients play a significant role in the overall evolutionary dynamics, the traveling-wave approach will break down.

ACKNOWLEDGMENTS

We thank Ivana Cvijović and Jiseon Min for useful discussions and comments on the manuscript. MJM acknowledges support from the NSF-Simons Center for Mathematical and Statistical Analysis of Biology at Harvard University (NSF grant DMS-1764269) and the Harvard FAS Quantitative Biology Initiative. This work was also supported in part by the NIH (R01-GM104239), the Simons Foundation (Grant 376196), and the NSF (PHY-1914916). Computational work was performed on the Cannon cluster supported by the Research Computing Group at Harvard University.

REFERENCES

Agarwala, A. and Fisher, D. S. (2019). Adaptive walks on high-dimensional fitness landscapes and seascapes with distance-dependent statistics. *Theoretical Population Biology*, 130:13–49.

Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in drosophila dna. *Genetics*, 139(2):1067–1076.

Akashi, H., Osada, N., and Ohta, T. (2012). Weak selection and protein evolution. *Genetics*, 192(1):15–31.

Barton, N. H., Etheridge, A. M., and Véber, A. (2017). The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, 118:50–73.

Berestycki, N. (2009). Recent progress in coalescent theory. *Ensaios Matematicos*, 16.

Bolthausen, E. and Sznitman, A.-S. (1998). On ruelle's probability cascades and an abstract cavity method. *Communications in Mathematical Physics*, 197(2):247–276.

Brunet, E., Derrida, B., Mueller, A. H., and Munier, S. (2007). Effect of selection on ancestry: An exactly soluble case and its phenomenological generalization. *Phys. Rev. E*, 76:041104.

Buskirk, S. W., Peace, R. E., and Lang, G. I. (2017). Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. *Proceedings of the National Academy of Sciences*, 114(31):8330–8335.

Bustamante, C. D., Wakeley, J., Sawyer, S., and Hartl, D. L. (2001). Directional selection and the site-frequency spectrum. *Genetics*, 159(4):1779–1788.

Charlesworth, B. (1994). The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetics Research*, 63(3):213–227.

Charlesworth, B., Morgan, M., and Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, 134(4):1289–1303.

Cohen, E., Kessler, D. A., and Levine, H. (2005). Front propagation up a reaction rate gradient. *Phys. Rev. E*, 72:066126.

Crow, J. F., Kimura, M., et al. (1970). An introduction to population genetics theory. *An introduction to population genetics theory.*

Cvijović, I., Good, B. H., and Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. *Genetics*, 209(4):1235–1278.

Desai, M. M. and Fisher, D. S. (2007). Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics*, 176(3):1759–1798.

Desai, M. M., Walczak, A. M., and Fisher, D. S. (2013). Genetic diversity and the structure of genealogies in rapidly adapting populations. *Genetics*, 193(2):565–585.

Eyre-Walker, A. and Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8:610 EP –.

Eyre-Walker, A., Woolfit, M., and Phelps, T. (2006). The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics*, 173(2):891.

Falconer, D. and Mackay, T. (1996). Introduction to quantitative genetics. essex. *UK: Longman Group.*

Fisher, D. S. (2013). Asexual evolution waves: fluctuations and universality. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01):P01011.

Fisher, R. A. (1931). Xvii.?the distribution of gene ratios for rare mutations. *Proceedings of the Royal Society of Edinburgh*, 50:204?219.

Fogle, C. A., Nagle, J. L., and Desai, M. M. (2008). Clonal interference, multiple mutations and adaptation in large asexual populations. *Genetics*, 180(4):2163–2173.

Gardiner, C. W. et al. (1985). *Handbook of stochastic methods*, volume 3. springer Berlin.

Gerrish, P. J. and Lenski, R. E. (1998). The fate of competing beneficial mutations in an asexual population. *Genetica*, 102(0):127.

Gillespie, J. H. (1983). A simple stochastic gene substitution model. *Theoretical Population Biology*, 23(2):202–215.

Gillespie, J. H. (1984). Molecular evolution over the mutational landscape. *Evolution*, pages 1116–1129.

Good, B. H. and Desai, M. M. (2013). Fluctuations in fitness distributions and the effects of weak linked selection on sequence evolution. *Theoretical Population Biology*, 85:86–102.

Good, B. H. and Desai, M. M. (2014). Deleterious passengers in adapting populations. *Genetics*, 198(3):1183–1208.

Good, B. H. and Desai, M. M. (2015). The impact of macroscopic epistasis on long-term evolutionary dynamics. *Genetics*, 199(1):177–190.

Good, B. H., Rouzine, I. M., Balick, D. J., Hallatschek, O., and Desai, M. M. (2012). Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proceedings of the National Academy of Sciences*, 109(13):4950–4955.

Good, B. H., Walczak, A. M., Neher, R. A., and Desai, M. M. (2014). Genetic diversity in the interference selection limit. *PLOS Genetics*, 10(3):1–1.

Goyal, S., Balick, D. J., Jerison, E. R., Neher, R. A., Shraiman, B. I., and Desai, M. M. (2012). Dynamic mutation–selection balance as an evolutionary attractor. *Genetics*, 191(4):1309–1319.

Haigh, J. (1978). The accumulation of deleterious genes in a population—muller's ratchet. *Theoretical Population Biology*, 14(2):251 – 267.

Hallatschek, O. (2011). The noisy edge of traveling waves. *Proceedings of the National Academy of Sciences*, 108(5):1783–1787.

Hallatschek, O. (2018). Selection-like biases emerge in population models with recurrent jackpot events. *Genetics*, 210(3):1053.

Hartl, D. L., Moriyama, E. N., and Sawyer, S. A. (1994). Selection intensity for codon bias. *Genetics*, 138(1):227–234.

Hegreness, M., Shoresh, N., Hartl, D., and Kishony, R. (2006). An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science*, 311(5767):1615–1617.

Held, T., Klemmer, D., and Lässig, M. (2019). Survival of the simplest in microbial evolution. *Nature communications*, 10(1):1–11.

Hudson, R. R. and Kaplan, N. L. (1995). Deleterious background selection with recombination. *Genetics*, 141(4):1605–1617.

Jerison, E. R. and Desai, M. M. (2015). Genomic investigations of evolutionary dynamics and epistasis in microbial evolution experiments. *Current Opinion in Genetics & Development*, 35:33–39.

Johnson, M. S., Martsul, A., Kryazhimskiy, S., and Desai, M. M. (2019). Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science*, 366(6464):490–493.

Kao, K. C. and Sherlock, G. (2008). Molecular characterization of clonal interference during adaptive evolution in asexual populations of saccharomyces cerevisiae. *Nat Genet*, 40(12):1499–1504.

Kaplan, N. L., Darden, T., and Hudson, R. R. (1988). The coalescent process in models with selection. *Genetics*, 120(3):819–829.

Kersting, G., Siri-Jégousse, A., and Wences, A. H. (2019). Site frequency spectrum of the bolthausen-sznitman coalescent. *arXiv preprint arXiv:1910.01732*.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626.

Kingman, J. (1982). The coalescent. *Stochastic Processes and their Applications*, 13(3):235 – 248.

Kosheleva, K. and Desai, M. M. (2013). The dynamics of genetic draft in rapidly adapting populations. *Genetics*, 195(3):1007–1025.

Krone, S. M. and Neuhauser, C. (1997). Ancestral processes with selection. *Theoretical Population Biology*, 51(3):210–237.

Kryazhimskiy, S., Rice, D. P., Jerison, E. R., and Desai, M. M. (2014). Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*, 344(6191):1519–1522.

Lang, G. I., Rice, D. P., Hickman, M. J., Sodergren, E., Weinstock, G. M., Botstein, D., and Desai, M. M. (2013). Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, 500(7464):571–574.

Lieberman, T. D., Flett, K. B., Yelin, I., Martin, T. R., McAdam, A. J., Priebe, G. P., and Kishony, R. (2014). Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nature genetics*, 46(1):82–87.

Machado, H. E., Lawrie, D. S., and Petrov, D. A. (2020). Pervasive strong selection at the level of codon usage bias in drosophila melanogaster. *Genetics*, 214(2):511–528.

McCandlish, D. and Stoltzfus, A. (2014). Modeling evolution using the probability of fixation: History and implications. *The Quarterly review of biology*, 89:225–52.

McDonald, J. H. and Kreitman, M. (1991). Adaptive protein evolution at the adh locus in drosophila. *Nature*, 351(6328):652–654.

McVean, G. A. T. and Charlesworth, B. (2000). The effects of hill-robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics*, 155(2):929.

Messer, P. W. and Petrov, D. A. (2013). Frequent adaptation and the mcdonald–kreitman test. *Proceedings of the National Academy of Sciences*, 110(21):8615.

Miralles, R., Gerrish, P. J., Moya, A., and Elena, S. F. (1999). Clonal interference and the evolution of rna viruses. *Science*, 285(5434):1745–1747.

Mohle, M. and Pitters, H. (2014). A spectral decomposition for the block counting process of the bolthausen-sznitman coalescent. *Electron. Commun. Probab.*, 19.

Nachman, M. W. (1998). Deleterious mutations in animal mitochondrial dna. *Genetica*, 102(0):61.

Neher, R. A. (2013). Genetic draft, selective interference, and population genetics of rapid adaptation. *Annual Review of Ecology, Evolution, and Systematics*, 44(1):195–215.

Neher, R. A. and Hallatschek, O. (2013). Genealogies of rapidly adapting populations. *Proceedings of the National Academy of Sciences*, 110(2):437–442.

Neher, R. A., Kessinger, T. A., and Shraiman, B. I. (2013). Coalescence and genetic diversity in sexual populations under selection. *Proceedings of the National Academy of Sciences*, 110(39):15836–15841.

Neher, R. A., Russell, C. A., Shraiman, B. I., and McVean, G. (2014a). Predicting evolution from the shape of genealogical trees. *eLife*, 3:e03568.

Neher, R. A., Russell, C. A., Shraiman, B. I., and McVean, G. (2014b). Predicting evolution from the shape of genealogical trees. *eLife*, 3:e03568.

Neher, R. A. and Shraiman, B. I. (2011). Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics*, 188(4):975–996.

Neher, R. A. and Shraiman, B. I. (2012). Fluctuations of fitness distributions and the rate of muller's ratchet. *Genetics*, 191(4):1283–1293.

Neher, R. A., Shraiman, B. I., and Fisher, D. S. (2010). Rate of adaptation in large sexual populations. *Genetics*, 184(2):467–481.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, 39(1):197–218.

Nielsen, R. and Yang, Z. (2003). Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral dna. *Molecular Biology and Evolution*, 20(8):1231–1239.

Nolan, J. P. (2018). *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, Boston. In progress, Chapter 1 online at http://fs2.american.edu/jpnolan/www/stable/stable.html.

Nourmohammad, A., Otwinowski, J., Łuksza, M., Mora, T., and Walczak, A. M. (2019). Fierce selection and interference in b-cell repertoire response to chronic hiv-1. *Molecular Biology and Evolution*, 36(10):2184–2194.

Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature*, 246(5428):96–98.

Orr, H. A. (2003). The distribution of fitness effects among beneficial mutations. *Genetics*, 163(4):1519.

Peck, J. R. (1994). A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics*, 137(2):597.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields*, 102(2):145–158.

Pitman, J. (1999). Coalescents with multiple collisions. *Ann. Probab.*, 27(4):1870–1902.

Pitman, J. and Yor, M. (1997). The two-parameter poisson-dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 25(2):855–900.

Rouzine, I. M., Brunet, É., and Wilke, C. O. (2008). The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. *Theoretical Population Biology*, 73(1):24 – 46.

Rouzine, I. M. and Coffin, J. M. (2005). Evolution of human immunodeficiency virus under selection and weak recombination. *Genetics*, 170(1):7–18.

Rouzine, I. M., Wakeley, J., and Coffin, J. M. (2003). The solitary wave of asexual evolution. *Proceedings of*

the National Academy of Sciences, 100(2):587–592.

Sawyer, S. A. and Hartl, D. L. (1992). Population genetics of polymorphism and divergence. *Genetics*, 132(4):1161–1176.

Sawyer, S. A., Kulathinal, R. J., Bustamante, C. D., and Hartl, D. L. (2003). Bayesian analysis suggests that most amino acid replacements in drosophila are driven by positive selection. *Journal of Molecular Evolution*, 57(1):S154–S164.

Sawyer, S. A., Parsch, J., Zhang, Z., and Hartl, D. L. (2007). Prevalence of positive selection among nearly neutral amino acid replacements in drosophila. *Proceedings of the National Academy of Sciences*, 104(16):6504.

Schiffels, S., Szöllősi, G. J., Mustonen, V., and Lässig, M. (2011). Emergent neutrality in adaptive asexual evolution. *Genetics*, 189(4):1361.

Schweinsberg, J. (2003). Coalescent processes obtained from supercritical galton–watson processes. *Stochastic Processes and their Applications*, 106(1):107–139.

Slatkin, M. and Hudson, R. R. (1991). Pairwise comparisons of mitochondrial dna sequences in stable and exponentially growing populations. *Genetics*, 129(2):555–562.

Söderberg, R. J. and Berg, O. G. (2007). Mutational interference and the progression of muller's ratchet when mutations have a broad range of deleterious effects. *Genetics*, 177(2):971.

Strelkowa, N. and Lässig, M. (2012). Clonal interference in the evolution of influenza. *Genetics*, 192(2):671.

Tsimring, L. S., Levine, H., and Kessler, D. A. (1996). Rna virus evolution via a fitness-space model. *Phys. Rev. Lett.*, 76:4440–4443.

Wakeley, J. (2005). *Coalescent Theory, an Introduction*. Roberts and Company, Greenwood Village, CO.

Weissman, D. B. and Barton, N. H. (2012). Limits to the rate of adaptive substitution in sexual populations. *PLOS Genetics*, 8(6):e1002740–.

Weissman, D. B. and Hallatschek, O. (2014). The rate of adaptation in large sexual populations with linear chromosomes. *Genetics*, 196(4):1167–1183.

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16(2):97–159.

Yang, Z. and Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, 15(12):496–503.

Yule, G. U. (1925). Ii.—a mathematical theory of evolution, based on the conclusions of dr. j. c. willis, f. r. s. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 213(402-410):21–87.

Zeng, K. and Charlesworth, B. (2009). Estimating selection intensity on synonymous codon usage in a nonequilibrium population. *Genetics*, 183(2):651–662.

Appendix A CHARACTERISTIC EQUATION FOR $\phi(x,\tau)$

In this Appendix, we review much of the key formalism used by Good and Desai (2014), Neher and Hallatschek (2013) and Fisher (2013) to analyze traveling wave models of asexual evolution. In particular, we derive equations for the density $f(x)$ of individuals with relative fitness x and the fixation probability $w(x)$ of an individual with relative fitness x. We then compute the generating function $\left\langle e^{-zn(t|x)}\right\rangle$ for the stochastic size $n(t|x)$ of a lineage t generations after its foundation, given the relative fitness x at which it was founded. In Appendix E, we use properties of $\left\langle e^{-zn(t|x)}\right\rangle$—as well as $\left\langle e^{-zn(t)}\right\rangle$, where $n(t)$ does not condition on the relative fitness x at which a lineage is founded—discussed here to compute statistics of genetic diversity such as the distribution of times to pairwise coalescence and the site frequency spectrum.

To begin, following the approach taken by Good and Desai (2014), we consider a lineage with initial absolute fitness distribution $g(X,0)$ at time $t=0$. Both the size and fitness distribution of the lineage may vary with time; we denote by $g(X,t)$ the time-dependent absolute fitness distribution of the lineage, normalized such that the lineage consists of $n(t)\equiv N\int g(X,t)dX$ individuals at time t. The time-dependent trajectory of the lineage is coupled to the trajectory of the rest of the population, whose time-dependent absolute fitness distribution we denote by $f(X,t)$. These two quantities obey the following equations:

$$\frac{\partial f(X,t)}{\partial t} = \left[X-\bar{X}(t)\right]f(X,t) + \int \mu(s)ds\left[f(X-s,t)-f(X,t)\right] + \sqrt{\frac{f(X,t)}{N}}\eta_f(X) - f(X,t)\int\left(\sqrt{\frac{f(X',t)}{N}}\eta_f(X') + \sqrt{\frac{g(X',t)}{N}}\eta_g(X')\right)dX', \quad (84)$$

$$\frac{\partial g(X,t)}{\partial t} = \left[X-\bar{X}(t)\right]g(X,t) + \int \mu(s)ds\left[g(X-s,t)-g(X,t)\right] + \sqrt{\frac{g(X,t)}{N}}\eta_g(X) - g(X,t)\int\left(\sqrt{\frac{f(X',t)}{N}}\eta_f(X') + \sqrt{\frac{g(X',t)}{N}}\eta_g(X')\right)dX', \quad (85)$$

with the stochastic nature of births and deaths captured by the last two terms in Eq. (84) and Eq. (85). Here η_g denotes a Gaussian random variable satisfying $\langle\eta_g(X)\rangle = 0$ and $\langle\eta_g(X)\eta_g(X')\rangle = 2\delta(X-X')$ (and likewise for η_f) (Gardiner et al., 1985). By construction, these noise terms ensure that $\int\left[f(X,t)+g(X,t)\right]dX$ is time-independent, such that a constant population size is maintained. The quantity $\bar{X}(t) = \int\left[f(X,t)+g(X,t)\right]XdX$ denotes the mean absolute fitness of the population at time t, such that $\left[X-\bar{X}(t)\right]f(X,t)$ denotes the rate at which $f(X,t)$ changes due to selection alone. Similarly, $\int\mu(s)ds\left[f(X-s,t)-f(X,t)\right]$ denotes the rate at which $f(X,t)$ changes due to mutation alone. We note that in Eq. (84) and Eq. (85), $f(X,t)$ and $g(X,t)$ are treated on a completely equal footing (in the sense that no assumption is made that $\int f(X,t)dX \gg \int g(X,t)dX$ or vice versa). In what follows, however, we focus on the case $\int g(X,t)dX \ll 1$ (and thus $\int f(X,t)dX \approx 1$), motivated by our key assumption that the fates of relevant lineages are determined while rare.

Throughout, we follow previous authors in making a "mean-field" approximation to Eq. (84) (Fisher, 2013; Good and Desai, 2014; Good et al., 2012; Neher et al., 2010). That is, we assume that at steady state, fluctuations shape the distribution $f(X,t)$ only via their effects on the mean fitness $\bar{X}(t)$ of the population, which can be approximated as increasing at a constant rate v. The population-wide fitness distribution $f(X,t)$ can then be described as a fixed-profile traveling wave, with $f(X,t) = f(X-vt,0)\equiv f(x)$. The quantity $f(x)$ can be interpreted as the distribution of *relative* fitnesses x in the population, with $x\equiv X-vt$. Without loss of generality, we take $\bar{X}(0)=0$; we further assume that $f(X,t)$ has already attained a steady-state traveling wave profile at $t=0$, potentially following an initial transient period. Given the rate v, $f(x)$ obeys

$$-v\partial_x f(x) = xf(x) + \int\mu(s)ds\left[f(x-s)-f(x)\right]. \quad (86)$$

As discussed in the main text, the solution to Eq. (86) attains negative values at large x (Fisher, 2013). To alleviate this pathological behavior, we follow previous authors in implementing a cutoff in $f(x)$, taking $f(x)\approx 0$ for $x > x_c$.

The rate of change of mean fitness, v, can be obtained by considering the probabilities with which new mutations eventually fix. This in turn can be done using a stochastic treatment of newly seeded lineages, which are described by Eq. (85). We make the key assumption that the fates of relevant lineages are determined while rare, so that in treating Eq. (85), we can focus on the case $\int g(X,t)dX \ll 1$. Following the approach taken in Good and Desai (2014), this implies that the last term in Eq. (85) can be neglected, so that $g(X,t)$ approximately satisfies

$$\frac{\partial g(X,t)}{\partial t} = (X - vt)\, g(X,t) + \int \mu(s)ds\, [g(X-s,t) - g(X,t)] + \sqrt{\frac{g(X,t)}{N}}\eta_g(X). \tag{87}$$

The *relative* fitness distribution $g(x,t)$ of a lineage approximately then satisfies

$$\frac{\partial g(x,t)}{\partial t} = v\partial_x g(x,t) + xg(x,t) + \int \mu(s)ds\, [g(x-s,t) - g(x,t)] + \sqrt{\frac{g(x,t)}{N}}\eta_g(x). \tag{88}$$

Here, by $Ng(x,t)$ we denote the number density at time t of individuals in the lineage with relative fitness x (or, equivalently, with absolute fitness $vt + x$). Given Eq. (87) or Eq. (88), fluctuations in lineage-wide fitness distributions can be examined, under the assumption that the background population advances in fitness at a constant rate v. In reality, fluctuations in lineage sizes are coupled to fluctuations in the rate of adaptation; this is particularly important at later times, when a lineage may comprise a macroscopic fraction of the total population with substantial probability. In assuming a constant rate v at which the population adapts, we neglect this coupling between the rate of adaptation and lineage sizes. Importantly, however, we do ensure that on average, fluctuations in lineage sizes (either to fixation or to extinction) give rise to a self-consistent rate of adaptation v.

To examine fluctuations in $g(x,t)$, we consider the generating functional $H[\phi(x),t]$, defined as

$$H[\phi(x),t] \equiv \left\langle \exp\left[-\int N\phi(x)g(x,t)dx\right]\right\rangle. \tag{89}$$

As discussed by Good and Desai (2013), we interpret (88) in the Itô sense, such that

$$\begin{aligned} H[\phi(x),t+dt] = \Big\langle \exp\Big(-\int N\phi(x) \\ \times\left[g(x,t) + \left(v\partial_x g(x,t) + xg(x,t) + \int \mu(s)ds\,[g(x-s,t) - g(x,t)]\right)dt + \sqrt{\frac{g(x,t)dt}{N}}\eta_g(x)\right]dx\Big)\Big\rangle. \end{aligned} \tag{90}$$

Retaining only terms $\mathcal{O}(dt)$ and larger, Eq. (90) simplifies to

$$\begin{aligned} H[\phi(x),t+dt] = \Bigg\langle \exp\left[-\int N\phi(x)g(x,t)dx\right] \\ \times\left[1 - dt\int N\phi(x)\left(v\partial_x g(x,t) + xg(x,t) + \int \mu(s)ds\,[g(x-s,t) - g(x,t)]\right)dx\right] \\ \times\left[1 - \sqrt{dt}\int N\phi(x)\sqrt{\frac{g(x,t)}{N}}\eta_g(x)dx + \frac{dt}{2}\left(\int N\phi(x)\sqrt{\frac{g(x,t)}{N}}\eta_g(x)dx\right)^2\right]\Bigg\rangle. \end{aligned} \tag{91}$$

Eq. (91) can be expressed as a differential equation for H, which, exploiting the independence of g and η_g, takes on the form

$$\begin{aligned} \frac{\partial H}{\partial t} &= \Bigg\langle \exp\left[-\int N\phi(x)g(x,t)dx\right]\Bigg\langle \\ &\times\left[-\int N\phi(x)\left(v\partial_x g(x,t) + xg(X,t) + \int \mu(s)ds\,[g(x-s,t) - g(x,t)]\right)dx + \int N^2\phi^2(x)\frac{g(x,t)}{N}dx\right]\Bigg\rangle \\ &= \int\left[-v\partial_x\phi(x) + x\phi(x) - \phi^2(x)\right]\frac{\partial H}{\partial\phi(x)}dx + \int \mu(s)ds\int dx\phi(x)\left[\frac{\partial H}{\partial\phi(x-s)} - \frac{\partial H}{\partial\phi(x)}\right] \\ &= \int\left[-v\partial_x\phi(x) + x\phi(x) + \int \mu(s)ds\,[\phi(x+s) - \phi(x)] - \phi^2(x)\right]\frac{\partial H}{\partial\phi(x)}dx, \end{aligned} \tag{92}$$

where the second equality is obtained using integration by parts, along with the fact that $\frac{\partial H}{\partial \phi(X)} = \left\langle \exp\left[-\int N\phi(X)g(X,t)dX\right](-Ng(X,t))\right\rangle$.

This enables us to identify characteristic curves $\phi(x,\tau)$ satisfying

$$\frac{\partial \phi(x,\tau)}{\partial \tau} = -v\partial_x \phi(x,\tau) + x\phi(x,\tau) + \int \mu(s)ds\left[\phi(x+s,\tau) - \phi(x,\tau)\right] - \phi^2(x,\tau), \tag{93}$$

with the property that $H\left[\phi(x),t\right] = H\left[\phi(x,\tau),t-\tau\right]$ for any $\tau > 0$. In particular,

$$H\left[\phi(x),t\right] = H\left[\phi(x,t),0\right] = \left\langle e^{-\int N\phi(x,t)g(x,0)dx}\right\rangle, \tag{94}$$

where the averaging in Eq. (94) is only over possible initial conditions—that is, possible instantiations of $g(x,0)$—and not over the stochasticity of the evolutionary process. The problem of calculating $H\left[\phi(x),t\right]$ is therefore reduced to solving Eq. (93) for $\phi(x,\tau)$ given the initial condition $\phi(x,0) = \phi(x)$.

A.1 Interpretation of $\phi_z(x,t)$

In the remaining analysis, we will denote by $\phi_z(x,t)$ the solution $\phi(x,t)$ to Eq. (93) that follows from the initial condition $\phi(x,0) = z$. Then

$$e^{-\phi_z(x,t)} = \left\langle e^{-zn(t|x)}\right\rangle, \tag{95}$$

where $n(t|x)$ is the (stochastic) size of a lineage founded at initial relative fitness x, after t generations (i.e., we can obtain a generating function for $n(t|x)$ by solving for $\phi_z(x,t)$). We note that in general, the initial condition $\phi(x,0)$ need not be independent of x. In Appendix E, we will briefly discuss and apply results of Fisher (2013) for the case $\phi(x,0) = \theta w(x)$, which yields a generating function for a lineage's effective lead frequency $\nu_L(t) = \int g(x,t)w(x)dx$.

A solution for the long-time limit of $\phi_z(x,t)$ is particularly useful, since for $z > 0$, $\lim_{t\to\infty}\phi_z(x,t)$ converges to $w(x)$ (Fisher, 2013; Good and Desai, 2014). To see this, we note that the above analysis of H (and thus of $\phi_z(x,t)$) is done under the branching process approximation, which assumes that lineages are small enough such that population-size constraints are unimportant. Consequently, fixation corresponds to the case in which $n(t|x)$ tends to infinity (rather than the total population size N). At sufficiently long times, $e^{-zn(t|x)}$ can only approach 1 (provided that the lineage goes extinct) or 0 (provided that the lineage fixes). That is,

$$\lim_{t\to\infty}\left\langle e^{-zn(t|x)}\right\rangle = 1 - w(x), \tag{96}$$

where $w(x)$ is defined as the fixation probability of an individual with relative fitness x. Given a solution $\phi_z(x,t)$ to Equation (93), it follows from (94) that

$$w(x) = 1 - \lim_{t\to\infty} e^{-\phi_z(x,t)} \approx \lim_{t\to\infty}\phi_z(x,t), \tag{97}$$

where the last approximation is employed under the assumption that $w(x) \ll 1$ (which holds for $x \ll 1$, as has been previously assumed in our analysis). Note, then, that $w(x)$ is the unique long-time limit of $\phi_z(x,t)$; at sufficiently long times, $\phi_z(x,t)$ does not depend on the initial condition $\phi(x,0) = z$, as long as $z > 0$. The fixation probability $w(x)$ is therefore the solution $\phi(x,t)$ to Eq. (93) with the time-derivative term neglected, and satisfies

$$v\partial_x w(x) = xw(x) + \int \mu(s)\left[w(x+s) - w(x)\right] - w^2(x). \tag{98}$$

This equation can also be obtained using standard branching process techniques (see e.g. Good et al. (2012); Neher et al. (2010)). In the main text, we discuss approximate solutions to Eq. (98) and Eq. (86) within the MSSM regime.

As discussed by Neher et al. (2014a) in the infinitesimal regime, $\phi_z(x,t)$ can additionally be interpreted as a lineage *sampling probability*. In particular, upon sampling Nz individuals at time t, $\phi_z(x,t)$ gives the

probability at least one of the sampled individuals is a descendant of a certain individual with relative fitness x at time 0. To see this, note that

$$\phi_z(x,t) \approx 1 - e^{-\phi_z(x,t)} \approx 1 - \left\langle \left(1 - \frac{Nz}{N}\right)^{n(t|x)} \right\rangle \tag{99}$$

where the right-hand side is precisely the above-described lineage sampling probability (with the expectation value on the right-hand side taken over $n(t|x)$). The quantity $\phi_z(x,t)$ with $z = 1/N$ is particularly useful to consider, since this gives the probability that a single sampled individual is the descendant of a given individual with relative fitness x a time t into the past. From a simple application of Bayes' theorem, we can obtain the probability distribution $A(x,t)$ for the relative fitness x of the *ancestor* of a randomly chosen individual (at t generations into the past):

$$A(x,t) \approx \frac{f(x)\phi_{1/N}(x,t)}{\int f(x)\phi_{1/N}(x,t)dx}, \tag{100}$$

where $f(x)$ can be interpreted as a prior probability distribution for the ancestor's relative fitness. Note that

$$\lim_{t\to\infty} A(x,t) = Nf(x)w(x), \tag{101}$$

so that, as recognized by Hallatschek (2011), $Nf(x)w(x)$ can be considered a distribution of relative fitnesses of future common ancestors of the population. In the main text, we discuss how Eq. (100), together with our solutions for $f(x)$ and $\phi_z(x,t)$ in the MSSM regime, can be used to address the timescale on which individuals in the "bulk" of the fitness distribution descend from its high-fitness "nose".

We will also consider the quantity $\Phi_0(z,t) \equiv \int f(x)\phi_z(x,t)dx$. By calculating $\Phi_0(z,t)$, we can obtain the generating function for the size $n(t)$ of a lineage founded by a randomly chosen individual in the population, after t generations. That is,

$$\left\langle e^{-zn(t)} \right\rangle = \int f(x)e^{-\phi_z(x,t)}dx \approx e^{-\Phi_0(z,t)}, \tag{102}$$

where in the second equality, we used the fact that $\phi_z(x,t) \ll 1$.

A.2 Review of properties of $\phi_z(x,t)$

We now review some key properties of the family of time-dependent solutions $\phi_z(x,t)$. These properties are discussed by Fisher (2013) in the context of populations subject only to beneficial mutations, in either the moderate-speeds ($v \ll \tilde{s}^2$) or high-speeds ($v \gg \tilde{s}^2$) regimes, although the key ideas hold more generally, including throughout the MSSM regime. In the following Subsection, we further analyze $\phi_z(x,t)$, explicitly demonstrating that these properties hold within the MSSM regime and extending the analysis conducted by Fisher (2013) to intermediate times that are particularly relevant to our calculation of the site frequency spectrum.

For short times ($t < T_c$), a dominant balance approach is useful (Fisher, 2013). For sufficiently large x, the $x\phi_z$ and ϕ_z^2 terms balance, and $\phi_z(x,t) \approx \phi_z^s(x,t) \equiv x$. The superscript denotes that ϕ_z^s is the approximate solution taken within the *saturation* regime, in which $\partial_x\phi_z(x,t)$ saturates approximately to 1. For smaller x, the nonlinear term ϕ^2 is subdominant. We denote by $\phi_z^l(x,t)$ the solution to the linearized equation

$$\frac{\partial\phi_z^l(x,t)}{\partial t} = -v\partial_x\phi_z^l(x,t) + x\phi_z^l(x,t) + \int \mu(s)ds\left[\phi_z^l(x+s,t) - \phi_z^l(x,t)\right], \tag{103}$$

which is solved exactly by

$$\phi_z^l(x,t) = z\exp\left[xt - \frac{1}{2}vt^2 + \int \frac{\mu(s)ds}{s}(e^{st} - 1 - st)\right], \tag{104}$$

given the initial condition $\phi_z(x,0) = z$. This suggests the approximate solution

$$\phi_z(x,t) \approx \begin{cases} z \exp\left[xt - \frac{1}{2}vt^2 + \int \frac{\mu(s)ds}{s}(e^{st} - 1 - st)\right] & x < P_z^l(t) \\ x & x > P_z^l(t), \end{cases} \tag{105}$$

where $P_z^l(t)$ is defined such that $\phi_z^l(P_z^l(t),t) \equiv P_z^l(t)$. The quantity $P_z^l(t)$ can be interpreted as the location of the time-dependent shoulder of the solution given in Eq. (105).

For $t > T_c - \mathcal{O}\left(\frac{1}{b}\right)$, a dominant balance approach is not suitable, and Eq. (105) is not a good approximation. Even if the nonlinear term is subdominant—that is, even if $x \ll P_z^l(t)$, and $\phi_z^l(x,t) \ll x$—the linearized solution $\phi_z^l(x,t)$ may be a very poor approximation to the actual solution $\phi_z(x,t)$. The reason is that in obtaining Eq. (104) from Eq. (103), $\phi_z^l(x+s,t)$ is also assumed to satisfy the same (linearized) equation. For $x > P_z(t) - s$, the actual $\phi_z(x+s,t)$ may be better approximated by $\phi_z^s(x+s,t) = x+s$ as opposed to $\phi_z^l(x+s,t)$. This discrepancy results in a time-derivative $\partial_t \phi^l(x,t)$ which differs from that given by Eq. (103). Over time, this discrepancy propagates down to lower x, and for $t > T_c$, the actual solution $\phi_z(x,t)$ will differ qualitatively from the linearized solution $\phi_z^l(x,t)$ even for x well below $P_z^l(t)$.

This issue is considered at length by Fisher (2013), in which the following properties of the solution $\phi_z(x,t)$ are identified for $t > T_c$:

1. Shortly after $t \approx T_c$, on a timescale $\mathcal{O}(1/b)$, $\phi_z(x,t)$ converges to $\phi_{P_z(t)}(x)$, with $\phi_{P_z(t)}(x)$ closely resembling the long-time fixed-point $w(x) \equiv \lim_{t\to\infty} \phi_{P_z(t)}(x)$, but with a shoulder at $P_z(t)$ rather than at $x_c = \lim_{t\to\infty} P_z(t)$.

2. The location $P_z(t)$ of the time-dependent shoulder converges to its fixed-point value x_c with rate $1/T_c$:
$$\frac{dP_z}{dt} = \frac{1}{T_c}\left(x_c - P_z(t)\right). \tag{106}$$

3. At long times, the solution $\phi_z(x,t)$ depends on z only via the initial condition $P_z(T_c)$ to Eq. (106). This can be approximated by taking $P_z(T_c) \approx P_z^l(T_c)$, with
$$\frac{P_z^l(T_c)}{x_c} = 1 + \frac{1}{T_c x_c} \log\left(\frac{\tilde{z} P_z^l(T_c)}{z x_c},\right), \tag{107}$$
where $\tilde{z}$ is defined such that $P_{\tilde{z}}^l(T_c) = x_c$. $P_z(T_c)$ can be further approximated by
$$P_z(T_c) \approx x_c + \frac{1}{T_c} \log\left(\frac{\tilde{z}}{z}\right). \tag{108}$$

4. The integral $\Phi_0(z,t) \equiv \int f(x)\phi_z(x,t)dx$ can be approximated as
$$\Phi_0(z,t) \approx e^{T_c(x_c - P_z(t))} \int f(x)w(x)dx = \frac{e^{T_c(x_c - P_z(t))}}{N}. \tag{109}$$
We note that Fisher (2013) considers a model in which the population size is not fixed, and the analog of $\int f(x)w(x)dx$ in Eq. (109) can fluctuate; here, we assume $\int f(x)w(x)dx = 1/N$ throughout, since the population size is assumed constant in our model.

These properties imply that for $b(t - T_c) \gg 1$,

$$\Phi_0(z,t) \approx \frac{1}{N}\left(\frac{z}{\tilde{z}}\right)^{e^{1-t/T_c}} \tag{110}$$

which is a key result we will use in computing statistics of genetic diversity in the MSSM regime (in particular, in computing the site frequency spectrum at moderate and high frequencies). Most of our results follow from the fact $\Phi_0(z,t)$ is a simple monomial of z with power e^{1-t/T_c} (i.e., that $\Phi_0(z,t) \approx C(t) z^{e^{1-t/T_c}}$, with $C(t)$ independent of z). The full expression in Eq. (110) is useful for ensuring that this simple monomial dependence is valid over the appropriate region. We will see that our calculations of genetic diversity statistics in Appendix E involve integrals over $u \equiv N\Phi_0(z,t)$ which are dominated by the region $u \sim \mathcal{O}(1)$. For use in this application, it will therefore be sufficient that Eq. (109) holds for the range $|P_z(t) - x_c| \lesssim \mathcal{O}(b)$ (over which u varies from $e^{-T_c b}$ to $e^{T_c b}$).

A.3 Analysis of $\phi_z(x,t)$

Here, we provide a detailed analysis of $\phi_z(x,t)$, explicitly demonstrating that the above properties hold for $b(t-T_c) \gg 1$, and extending this analysis to the case $b(t-T_c) = \mathcal{O}(1)$. In both cases, we can analyze $\phi_z(x,t)$ in a way similar to our analysis of $f(x)$ and $w(x)$ by defining

$$\phi_z(x,t) \equiv \frac{x_c e^{T_c(x-x_c)}}{\mathrm{Ai}\left(\frac{c-x_c}{b}\right)} r(x,t), \tag{111}$$

and approximating the resulting differential equation for $r(x,t)$. Note that the constant factors in our definition of $r(x,t)$ are chosen such that $\lim_{t\to\infty} r(x,t) = \mathrm{Ai}\left(\frac{c-x}{b}\right)$; for simplicity in notation, we have suppressed the dependence of $r(x,t)$ on z. Making the approximation $r(x+s,t) \approx r(x) + s\partial_x r(x,t) + \frac{1}{2}s^2\partial_x^2 r(x,t)$ (which can be justified by noting that the resulting solution varies on the scale $\mathcal{O}(b)$ for relevant fitnesses x) and neglecting the nonlinear term $\phi_z^2(x,t)$, we have

$$\partial_t r(x,t) = b^3\partial_x^2 r(x,t) - (c-x)r(x,t). \tag{112}$$

Our solution incorporates the nonlinearity in the equation for $\phi_z(x,t)$ through the boundary conditions

$$r(P_z(t) + 1/T_c, t) = 0, \tag{113}$$

and

$$\partial_x r\left(P_z(t) + 1/T_c, t\right) = -\frac{T_c P_z e^{-T_c(P_z - x_c)}}{x_c} \mathrm{Ai}\left(\frac{c-x_c}{b}\right). \tag{114}$$

These boundary conditions ensure that $\phi_z(P_z(t),t) \approx P_z(t)$, and that rapid variation in $e^{T_c x}$ is approximately canceled by variation in $r(x,t)$ at $x = P_z(t)$, such that $\phi_z(P_z(t),t)$ is approximately linear at $x = P_z(t)$.

An approximate solution to Eq. (112) can be written as

$$r(x,t) \approx \sum_{j=0}^{\infty} y_j \exp\left[h(t) + b(t-T_c)(z_j - z_0) - a(P_z(t) - x_c)\right] \mathrm{Ai}\left(\frac{P_z(t)-x}{b} + z_j + \frac{1}{T_c b}\right), \tag{115}$$

where $h(t)$ satisfies

$$T_c \frac{dP_z}{dt} = x_c - P_z(t) + \frac{dh}{dt}, \tag{116}$$

and the y_j are arbitrary coefficients. Here, by z_j we denote the (all negative) zeros of the Airy function satisfying $\mathrm{Ai}(z_j) = 0$, in order of increasing magnitude with j. To be more precise, Eq. (115) with $h(t)$ satisfying Eq. (116) approximately solves Eq. (112), provided that $\mathrm{Ai}'\left(\frac{P_z(t)-x}{b} + z_j + \frac{1}{T_c b}\right) \ll ab\mathrm{Ai}\left(\frac{P_z(t)-x}{b} + z_j + \frac{1}{T_c b}\right)$, which is satisfied in the region of interest. The boundary condition $r(P_z + 1/T_c, t) = 0$ is automatically satisfied, while the boundary condition in Eq. (114) implies that

$$e^{-h(t)} \approx \frac{x_c}{P_z(t)} \sum_{j=0}^{\infty} y_j \frac{\mathrm{Ai}'(z_j)}{\mathrm{Ai}'(z_0)} e^{b(t-T_c)(z_j - z_0)}, \tag{117}$$

which, under the assumption $P_z(t) - x_c \ll x_c$, can be differentiated to yield

$$\frac{dh}{dt} \approx -\frac{\sum_{j=0}^{\infty} b(z_j - z_0) y_j \mathrm{Ai}'(z_j) e^{z_j b(t-T_c)}}{\sum_{j=0}^{\infty} y_j \mathrm{Ai}'(z_j) e^{z_j b(t-T_c)}}. \tag{118}$$

Long times: $b(t-T_c) \gg 1$

The analysis simplifies considerably at long times such that $b(t-T_c) \gg 1$. The sums in Eq. (115) and Eq. (117) are dominated by $j=0$ terms, with remaining terms exponentially suppressed. We therefore have

$$T_c \frac{dP_z}{dt} \approx x_c - P_z, \tag{119}$$

which is solved by

$$P_z(t) - x_c \approx [P_z(T_c) - x_c]\, e^{1-t/T_c}, \tag{120}$$

and the quantity $r(x,t)$ simplifies to

$$r(x,t) \approx \frac{P_z(t)}{x_c} e^{-T_c(P_z(t)-x_c)} \mathrm{Ai}\left(\frac{c-x}{b} + \frac{P_z(t)-x_c}{b}\right). \tag{121}$$

From Eq. (121) it follows that $\phi_z(x,t) \approx \frac{P_z(t)}{x_c} w\left(x + x_c - P_z(t)\right)$; the quantity $\Phi_0(z,t) \equiv \int f(x)\phi_z(x,t)dx$ can therefore easily be evaluated using the same calculation carried out in Appendix D to obtain $p_{\mathrm{fix}}(s)$, with the result

$$\Phi_0(z,t) \approx \frac{e^{T_c(x_c - P_z(t))}}{N}, \tag{122}$$

provided that $|P_z(t) - x_c| \ll b$. Using the $P_z(T_c)$ in Eq. (108), Eq. (122) further simplifies to

$$\Phi_0(z,t) \approx \frac{1}{N}\left(\frac{z}{\tilde{z}}\right)^{e^{1-t/T_c}}. \tag{123}$$

Intermediate times $b(t-T_c) \lesssim \mathcal{O}(1)$

At intermediate times such that $b(t-T_c) \lesssim \mathcal{O}(1)$, multiple terms contribute substantially to the sums in Eq. (115) and Eq. (117). As a result, we need to determine the appropriate coefficients y_j in these sums, which can be done by matching our solution in Eq. (115) onto the solution for $t < T_c$ given in Eq. (105). To do so, we will integrate $\mathrm{Ai}\left(\frac{P_z - x}{b} + z_i + \frac{1}{T_c b}\right) r(x,t)$ over the x-interval $(-\infty, P_z(t) + 1/T_c)$, using $r(x,t)$ from both of these solutions, and exploit the orthogonality and completeness of appropriately shifted Airy functions. We will equate the two resulting integrals evaluated at a time $t = t_0$ satisfying $1 \ll (T_c - t_0)b \ll T_c b$, and see that our final result will not depend on the precise choice of t_0.

In particular, from the Airy equation it follows (using integration by parts) that

$$\int_{x_3}^{\infty} \mathrm{Ai}(x+x_1)\mathrm{Ai}(x+x_2)dx = \frac{1}{x_2 - x_1}\left[\mathrm{Ai}\left(x_3 + x_2\right)\mathrm{Ai}'\left(x_3 + x_1\right) - \mathrm{Ai}\left(x_3 + x_1\right)\mathrm{Ai}'\left(x_3 + x_2\right)\right] \tag{124}$$

for arbitrary x_1, x_2 and x_3. A further consequence is that $\int_{z_i}^{\infty} \mathrm{Ai}^2(x)dx = \left[\mathrm{Ai}'(z_i)\right]^2$, and also that

$$\int_0^{\infty} \mathrm{Ai}\left(x + z_i\right)\mathrm{Ai}\left(x + z_j\right) = \delta_{ij}\left[\mathrm{Ai}'(z_i)\right]^2, \tag{125}$$

which establishes the orthogonality of the set of functions $\mathrm{Ai}(x + z_j)$ on the interval $(0, \infty)$. Using these properties, we can carry out the above-described matching, with the result

$$y_i\left[\mathrm{Ai}'(z_i)\right]^2 e^{h(t_0)} \approx \frac{P_z(t_0)e^{-(x_c - c)(t_0 - T_c)}}{x_c}\mathrm{Ai}\left(\frac{c - x_c}{b}\right)\int_{z_i}^{\infty} \mathrm{Ai}(u)e^{b(T_c - t_0)u}du, \tag{126}$$

where, in obtaining the right-hand side, we made use of the relations $P_z(t) = ze^{P_z(t)t - \frac{1}{2}vt^2 + \int \frac{\mu(s)ds}{s}\left(e^{st} - st - 1\right)}$ and $x_c \approx c - bz_0 - 1/T_c$. For sufficiently large $b(t_0 - T_c)$, the integral on the right-hand side is dominated by $u > z_0$, so that we can extend its lower limit of integration to $-\infty$; the integral thus evaluates approximately to $e^{[b(T_c - t_0)]^3/3}$. We can therefore take

$$y_i = \frac{1}{\left[\mathrm{Ai}'(z_i)\right]^2}, \tag{127}$$

with

$$e^{h(t_0)} \approx \frac{P_z(t_0)e^{-(x_c - c)(t_0 - T_c)}}{x_c}\mathrm{Ai}\left(\frac{c - x_c}{b}\right)e^{[b(T_c - t_0)]^3/3}. \tag{128}$$

To simplify Eq. (128), we use the relation between T_c, N and x_c given in Eq. (31). This yields

$$e^{h(t_0)} \approx \left[\mathrm{Ai}'(z_0)\right]^2 N z e^{(P_z(t_0)-x_c)t_0 + c(t_0-T_c) + [b(T_c-t_0)]^3/3 + J(T_c) - J(t_0)}, \tag{129}$$

where $J(t) \equiv vt^2/2 - \int \frac{\mu(s)}{s}(e^{st} - st - 1)ds$. Expanding $J(t_0) - J(T_c) \approx J'(T_c)(t_0 - T_c) + \frac{1}{2}J''(T_c)(t_0 - T_c)^2 + \frac{1}{6}J^{(3)}(T_c)(t_0 - T_c)^3$, we then have

$$e^{h(t_0)} \approx \left[\mathrm{Ai}'(z_0)\right]^2 N z e^{(P_z(t_0)-x_c)t_0}, \tag{130}$$

for $-1 \lesssim b(t_0 - T_c) < 0$.

Eq. (127) can be substituted into Eq. (118) to yield

$$\frac{dh}{dt} \approx -b \frac{\sum_{j=0}^{\infty}(z_j - z_0)\left[1/\mathrm{Ai}'(z_j)\right] e^{z_j b(t-T_c)}}{\sum_{j=0}^{\infty}\left[1/\mathrm{Ai}'(z_j)\right] e^{z_j b(t-T_c)}} \tag{131}$$

which is valid for intermediate times $0 < b(t-T_c) \lesssim \mathcal{O}(1)$ and long times $b(t-T_c) \gg 1$, and which determines the time-dependence of $P_z(t)$ through Eq. (116). Importantly, from Eq. (131) it follows that $dh/dt \sim \mathcal{O}(b)$ for $b(t - T_c) \lesssim \mathcal{O}(1)$; from Eq. (116), it then follows that

$$\frac{P_z(t) - x_c}{b} \approx \frac{P_z(T_c) - x_c}{b} \approx \frac{1}{T_c b} \log\left(\frac{\tilde{z}}{z}\right), \tag{132}$$

for times $b(t - T_c) \lesssim \mathcal{O}(1)$. Eq. (116) can thus be integrated to yield, for times $b(t - T_c) \lesssim \mathcal{O}(1)$,

$$e^{h(t)} \approx \left[\mathrm{Ai}'(z_0)\right]^2 N z e^{(P_z(T_c)-x_c)t}, \tag{133}$$

which does not depend on the choice t_0. Substituting Eq. (133) and Eq. (127) into Eq. (115), we thus have

$$r(x,t) = zN \sum_{j=0}^{\infty} \left(\frac{\mathrm{Ai}'(z_0)}{\mathrm{Ai}'(z_j)}\right)^2 e^{b(t-T_c)[z_j - z_0 + (P_z(T_c)-x_c)/b]} \mathrm{Ai}\left(\frac{c-x}{b} + \frac{P_z(T_c) - x_c}{b} + (z_j - z_0)\right), \tag{134}$$

for $b(t - T_c) \lesssim \mathcal{O}(1)$. We can further simplify Eq. (134) by replacing $(P_z(T_c) - x_c)/b$ with $\frac{1}{T_c b}\log\frac{\tilde{z}}{z}$. In particular, evaluating $\phi_z(x,t)$ at $z = \zeta/N$, we have

$$\frac{\phi_{\zeta/N}(x,t)}{\zeta} \approx \left(\frac{x_c e^{T_c(x-x_c)}}{\mathrm{Ai}\left(\frac{c-x_c}{b}\right)}\right) \times \sum_{j=0}^{\infty} \left(\frac{\mathrm{Ai}'(z_0)}{\mathrm{Ai}'(z_j)}\right)^2 e^{b(t-T_c)\left[z_j - z_0 - \frac{1}{T_c b}\log\zeta\right]} \mathrm{Ai}\left(\frac{c-x}{b} + (z_j - z_0) - \frac{1}{T_c b}\log\zeta\right), \tag{135}$$

since (as discussed below) $N\tilde{z} \approx 1$ to logarithmic accuracy.

A.4 Average rate of adaptation

Finally, we conclude this Appendix by briefly reviewing the procedure for computing v in terms of N introduced by Fisher (2013). While some of that work considers a model in rate of adaptation v is fixed and the population size can vary, Fisher (2013) notes that for a fixed population size N,

$$N = 1/\tilde{z}, \tag{136}$$

where $\tilde{z}$ depends implicitly on v, and can be defined such that given the initial condition $\phi(x,0) = \tilde{z}$, $\phi(x,t)$ converges to the eventual fixed-point $w(x)$ with zero amplitude of the slowest eigenvector. Given the properties of the solutions $\phi_z(x,t)$ discussed in the previous Subsection, it follows from this definition that

$P_{\tilde{z}}^l(T_c) = x_c$, where $P_z^l(t)$ is the location of the time-dependent shoulder obtained by matching the linearized and saturated solutions in Eq. (105). That is,

$$\tilde{z}e^{T_c x_c - \frac{1}{2}vT_c^2 + \int \frac{\mu(s)ds}{s}\left(e^{T_c s} - 1 - T_c s\right)} = x_c, \tag{137}$$

which yields

$$Nx_c = e^{T_c x_c - \frac{1}{2}vT_c^2 + \int \frac{\mu(s)ds}{s}\left(e^{T_c s} - 1 - T_c s\right)}. \tag{138}$$

Apart from a factor of $T_c b \mathrm{Ai}'(z_0)$ multiplying N, Eq. (138) precisely matches our result obtained for the MSSM regime. Because of the logarithmic dependence of most quantities of interest on N, this is a relatively minor difference (even though $T_c b \gg 1$). As in our case, to solve for v using Eq. (138) requires another relation between x_c and T_c (as well as $\mu(s)$). Within the "high-speeds" regime, Fisher (2013) obtains the same relation $x_c \approx c + b|z_0| - 1/T_c$ that we obtain within the MSSM regime, using a solvability condition for $w(x)$ which is exact asymptotically up to ambiguities of order $1/T_c$ due to fluctuations.

Appendix B DETERMINATION OF v AND x_c

In this Appendix, we enforce the relevant self-consistency conditions to solve for v and x_c, given the approximate solutions $f(x)$ and $w(x)$ obtained in Appendix C. Our presentation largely parallels the analysis presented by Tsimring et al. (1996) for the determination of v within the infinitesimal regime. We reproduce this derivation in detail because with a slight modification, this computation can be directly applied to yield v and x_c within the MSSM regime.

In particular, we make use the following properties of $f(x)$ identified within both limits: (i) $w(x) \approx x$ for $x > x_c$, and (ii) within the region dominating $\int_{-\infty}^{x_c} f(x)w(x)dx$, $f(x) \approx B_1 e^{-T_c x}\mathrm{Ai}\left(\frac{c-x}{b}\right)$ and $w(x) \approx C_1 e^{T_c x}\mathrm{Ai}\left(\frac{c-x}{b}\right)$, with constant factors B_1 and C_1 independent of x. In describing the infinitesimal approximation, (ii) is to be interpreted along with the substitutions $T_c \to \frac{\sigma^2}{2D}$, $b \to D^{1/3}$, and $c \to \frac{\sigma^4}{4D}$; in describing the MSSM regime, (ii) is to be interpreted with the definitions for T_c, b, and c used throughout the main text. These substitutions allow for an analysis of the two limits in parallel. In both cases, we assume $T_c b \gg 1$. In the MSSM approximation, $T_c b \gg 1$ is explicitly assumed, while in the infinitesimal approximation, $T_c b \gg 1$ follows from the additional assumption $ND^{1/3} \gg 1$. The majority of work on the infinitesimal regime has focused on the case $ND^{1/3} \gg 1$; we thus focus on this case in our consideration of the infinitesimal regime. An analysis of the $ND^{1/3} \to 0$ limit, as well as a numerical consideration of the arbitrary $ND^{1/3}$ case, is conducted by Good and Desai (2013) and Good et al. (2014). Finally, we make use of the normalization constants B_1 obtained in Appendix C. With the substitutions assumed here,

$$B_1 = \frac{e^{\frac{2}{3}(T_c b)^3}}{b} \tag{139}$$

within the infinitesimal approximation, and

$$B_1 = \frac{e^{\frac{vT_c^2}{2} - \int \frac{\mu(s)}{s}\left(e^{T_c s} - T_c s - 1\right)ds}}{b} \tag{140}$$

within the MSSM approximation.

The constant C_1 and the interference threshold x_c can be determined (in terms of T_c, b and c) by enforcing the continuity of $w(x)$ and $w'(x)$ at $x = x_c$ (Good and Desai, 2014). Continuity of $w(x)$ at x_c requires that

$$C_1 = \frac{x_c e^{-T_c x_c}}{\mathrm{Ai}\left(\frac{c-x_c}{b}\right)}, \tag{141}$$

and continuity of $w'(x)$ at $x = x_c$ requires that

$$1 = T_c x_c \left[1 - \frac{1}{T_c b}\frac{\mathrm{Ai}'\left(\frac{c-x_c}{b}\right)}{\mathrm{Ai}\left(\frac{c-x_c}{b}\right)}\right]. \tag{142}$$

Given T_c and $\mu(s)$, x_c can be obtained as a numerical solution to Eq. (142). An explicit expression for x_c in terms of T_c and $\mu(s)$ can be obtained by making an ansatz that $T_c x_c \gg 1$, such that

$$\frac{\mathrm{Ai}'\left(\frac{c-x_c}{b}\right)}{\mathrm{Ai}\left(\frac{c-x_c}{b}\right)} \approx T_c b. \tag{143}$$

We will assume that $x_c < c + b|z_0|$, where $z_0 \approx -2.34$ is the least negative zero of $\mathrm{Ai}(z)$; this ensures that the patched $w(x)$, as well as the cutoff $f(x)$, are both nonnegative for all x. For $y > z_0$, $\mathrm{Ai}'(y)/\mathrm{Ai}(y)$ is only large and positive for y close to z_0, motivating an expansion of $\mathrm{Ai}\left(\frac{c-x_c}{b}\right)$ around $\frac{c-x_c}{b} = z_0$. Retaining only the lowest-order term of this expansion yields

$$x_c \approx c + b|z_0| - \frac{1}{T_c}. \tag{144}$$

Eq. (144) is consistent with our ansatz $T_c x_c \gg 1$, given our additional assumption that $T_c b \gg 1$ (along with the positivity of c implied by its definition), and is therefore justified.

The consistency condition $\frac{1}{N} = \int f(x)w(x)dx$, which enforces that a neutral mutation fixes with an unbiased probability $1/N$, can be used to relate the quantities T_c, b and c—and thus v—to N and $\mu(s)$. This consistency condition can be expressed as

$$\frac{1}{N} = \frac{B_1 x_c e^{-T_c x_c}}{\mathrm{Ai}\left(\frac{c-x_c}{b}\right)} \int_{-\infty}^{x_c} \mathrm{Ai}^2\left(\frac{c-x}{b}\right) dx, \tag{145}$$

which evaluates to

$$\frac{1}{N} = \frac{B_1 b x_c e^{-T_c x_c}}{\mathrm{Ai}\left(\frac{c-x_c}{b}\right)} \left[\mathrm{Ai}'\left(\frac{c-x_c}{b}\right)^2 - \left(\frac{c-x_c}{b}\right)\mathrm{Ai}^2\left(\frac{c-x_c}{b}\right)\right], \tag{146}$$

or, approximately, to

$$\frac{1}{N} \approx B_1 b x_c e^{-T_c x_c}(T_c b)\mathrm{Ai}'(z_0) \tag{147}$$

given the location of x_c identified in Eq. (27). Substituting in the appropriate normalization constant B_1 for the infinitesimal approximation, Eq. (147) simplifies to

$$N x_c = \frac{e^{(T_c b)^3/3 + |z_0| T_c b - 1}}{T_c b \mathrm{Ai}'(z_0)}, \tag{148}$$

while for the MSSM approximation, Eq. (147) reduces to

$$N x_c = \frac{e^{T_c x_c - \frac{v T_c^2}{2} + M_{-1}(T_c)}}{T_c b \mathrm{Ai}'(z_0)}. \tag{149}$$

These are the key equations, in combination with Eq. (142), which relate T_c (and thus v) to N and $\mu(s)$ in the two approaches. We discuss the numerical solution of these equations in Appendix I.

B.1 An asymptotic approximation in the $N \to \infty$ limit

Finally, we conclude this Appendix by analyzing Eq. (149) to obtain an asymptotic expression for T_c in the limit $N \to \infty$. We focus on the case in which the DFE consists of a single beneficial effect—that is, we assume $\mu(s) = U\delta(s - s_b)$. In the limit $N \to \infty$, $T_c x_c \to T_c v$ and we have

$$\log NU \sim \frac{T_c^2 U s_b e^{T_c s_b}}{2}, \tag{150}$$

which has the solution

$$T_c s_b \sim 2W\left(\sqrt{\frac{s_b \log NU}{2U}}\right) \tag{151}$$

where W is the Lambert-W function satisfying $W(z)e^{W(z)} = z$. Using the asymptotic approximation $W(z) \sim \log z - \log\log z$ for large z, we have

$$T_c \sim \frac{1}{s_b}\log\left(\frac{s_b}{U}\log NU\right) \tag{152}$$

$$v \sim \frac{2s_b^2 \log NU}{\log^2\left(\frac{s_b}{U}\log NU\right)}, \tag{153}$$

$$b \sim \frac{s_b \log^{1/3} NU}{\log^{2/3}\left(\frac{s_b}{U}\log NU\right)}, \tag{154}$$

and

$$c \sim \frac{2s_b \log NU}{\log\left(\frac{s_b}{U}\log NU\right)}. \tag{155}$$

Note that in the limit $N \to \infty$, the conditions of validity $s_b \ll b$ and $T_c b \gg 1$ of the MSSM approximation are in fact satisfied.

Appendix C APPROXIMATION OF $f(x)$ AND $w(x)$ USING TRANSFORM METHODS

In this Appendix, we use transform methods to approximate $f(x)$ and $w(x)$ within both the infinitesimal and MSSM regimes. To do so, we begin by reviewing the solutions, provided by Fisher (2013), for the Laplace transforms $\tilde{f}(z) = \int_{-\infty}^{\infty} e^{-zx} f(x)dx$ and $\tilde{w}(z) = \int_{-\infty}^{\infty} e^{-zx} w(x)dx$. We then demonstrate that the infinitesimal approximation solutions $f(x)$ and $w(x)$ can be obtained from inverse Laplace transforms of $\tilde{f}(z)$ and $\tilde{w}(z)$, respectively, by approximation of the relevant contour integrals. Using a different approximation of these contour integrals, we reproduce the MSSM approximate solutions $f(x)$ and $w(x)$ found in the main text. In doing so, our main purpose is to obtain precise conditions of validity for both the infinitesimal and MSSM approximations, expressed as integrals involving $\mu(s)$.

Note that in the definitions we take of $\tilde{f}(z)$ and $\tilde{w}(z)$, $f(x)$ and $w(x)$ are assumed to satisfy

$$-v\partial_x f(x) = xf(x) + \int \mu(s)ds\,[f(x-s) - f(x)] \tag{156}$$

and

$$v\partial_x w(x) = xw(x) + \int \mu(s)ds\,[w(x+s) - w(x)] \tag{157}$$

respectively. Here $f(x)$ and $w(x)$ denote the formal solutions to Eq. (156) and Eq. (157), respectively. That is, within the definition of $\tilde{f}(z)$, $f(x)$ does not possess a "cutoff" at x_c; similarly, within the definition of $\tilde{w}(z)$, $w(x)$ does not tend to x for $x \gg x_c$ (which is the case for the solution to Eq. (10)). The transform methods used here are employed only to obtain approximate solutions Eq. (156) and Eq. (157). These approximate solutions can then be patched with the appropriate behavior for $x > x_c$ to obtain the full approximate solutions $f(x)$ and $w(x)$ which are used to compute quantities of interest.

The transforms $\tilde{f}(z)$ and $\tilde{w}(z)$ satisfy

$$-vz\tilde{f}(z) = -\partial_z \tilde{f}(z) + \left[\int \mu(s)(e^{-zs} - 1)ds\right]\tilde{f}(z) \tag{158}$$

and

$$vz\tilde{w}(z) = -\partial_z \tilde{w}(z) + \left[\int \mu(s)(e^{zs} - 1)ds\right] \tilde{w}(z). \tag{159}$$

In obtaining Eq. (158) from Eq. (5) we assume that $e^{-zx}f(x) \to 0$ as $x \to \infty$ and as $x \to -\infty$. Likewise, in obtaining Eq. (159) from Eq. (10) we assume that $e^{-zx}w(x) \to 0$ as $x \to \infty$ and as $x \to -\infty$. Together, these conditions restrict the range of z for which Eq. (158) and Eq. (159), and thus our solutions $\tilde{f}(z)$ and $\tilde{w}(z)$, are valid. However, if $f(x)$ and $w(x)$ are reasonably well-behaved, these conditions are satisfied when $\tilde{f}(z)$ and $\tilde{w}(z)$ (which are defined by integrals over $e^{-zx}f(x)$ and $e^{-zx}w(x)$, respectively) take on finite values; this is in fact satisfied by our solutions $\tilde{f}(z)$ and $\tilde{w}(z)$, within the region of z of interest to us.

Eq. (158) and Eq. (159) can each be integrated to yield

$$\tilde{f}(z) = \exp\left[\frac{vz^2}{2} - \int \frac{\mu(s)ds}{s}\left(e^{-zs} + zs - 1\right)\right] \tag{160}$$

and

$$\tilde{w}(z) = C_1 \exp\left[-\frac{vz^2}{2} + \int \frac{\mu(s)ds}{s}(e^{zs} - zs - 1)\right] \tag{161}$$

(Fisher, 2013), where the constant of integration for $\tilde{f}(z)$ is set by the requirement that $\int f(x)dx = 1$. The constant of integration for $\tilde{w}(z)$ will be determined along with the location x_c of the interference threshold by enforcing that $w(x)$ and $w'(x)$ are continuous at x_c.

To enforce this condition—as well as to determine v and other dynamical quantities of interest—the functional forms of $f(x)$ and $w(x)$ are needed (as opposed to their transforms $\tilde{f}(z)$ and $\tilde{w}(z)$). The distribution $f(x)$ follows from $\tilde{f}(z)$, using the standard formula for an inverse Laplace transform, as

$$f(x) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \exp\left[xz + \frac{vz^2}{2} - M_{-1}(-z)\right] dz \tag{162}$$

and

$$w(x) \propto \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \exp\left[xz - \frac{vz^2}{2} + M_{-1}(z)\right] dz, \tag{163}$$

where, as defined by Fisher (2013), $M_{-1}(z) \equiv \int \frac{\mu(s)}{s}\left(e^{zs} - zs - 1\right) ds$. We will also define

$$M_p(z) \equiv \frac{d^{p+1}M_{-1}(z)}{d(z)^{p+1}} \tag{164}$$

for $p \geq 0$, so that $M_0(z) \equiv \int \mu(s)\left(e^{zs} - 1\right) ds$ and $M_p(z) = \int \mu(s)s^p e^{zs} ds$ for $p \geq 1$.

C.1 Approximation of $f(x)$ and $w(x)$ within the Infinitesimal Approximation

We now turn to approximate Eq. (162) and Eq. (163) within the infinitesimal regime. Eq. (162) can be rewritten as

$$f(x) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \exp\left[xz + \frac{\sigma^2 z^2}{2} + \frac{Dz^3}{3} - \sum_{k=4}^{\infty}\frac{(-1)^k z^k}{k!}M_{k-1}(0)\right] dz \tag{165}$$

by expanding the exponential e^{-zs} within the integral over s, and defining $D \equiv \frac{1}{2}\int \mu(s)s^2 ds$, and $\sigma^2 \equiv v - U\langle s\rangle$. As in the rest of this article, we limit our attention to the case of the infinitesimal regime in which $ND^{1/3} \gg 1$, or equivalently, in which $\sigma \gg D^{1/3}$ (for a brief discussion of the case in which $ND^{1/3}$ is not large, see Appendix B). By making the substitution $u = -\frac{2Dz}{\sigma^2}$, (165) can be alternatively expressed as

$$f(x) = \frac{\sigma^2/2D}{2\pi i}\int_{-i\infty}^{i\infty} \exp\left[-\frac{\sigma^6}{8D^2}\left(\frac{xu}{\sigma^4/4D} - u^2 + \frac{u^3}{3}\right) - I_v(u)\right] du, \tag{166}$$

where we have defined

$$I_v(u) \equiv \left[\left(\frac{\sigma^2}{2D}\right) D^{1/3}\right]^3 \frac{\int \mu(s) \frac{2D}{\sigma^2 s}\left[e^{\frac{\sigma^2 su}{2D}} - 1 - \frac{\sigma^2 su}{2D} - \frac{1}{2}\left(\frac{\sigma^2 su}{2D}\right)^2 - \frac{1}{6}\left(\frac{\sigma^2 su}{2D}\right)^3\right] ds}{\frac{1}{2}\int \mu(s)(\frac{\sigma^2 s}{2D})^2 ds}. \tag{167}$$

To approximate $f(x)$ within the infinitesimal approximation, we neglect the contribution of $I_v(u)$ to the exponent of Eq. (166); the remaining integral can be evaluated, yielding

$$f(x) \approx \frac{e^{\frac{\sigma^6}{12D^2}}}{D^{1/3}} \exp\left[-\frac{\sigma^2 x}{2D}\right] \mathrm{Ai}\left(\frac{\sigma^4/4D - x}{D^{1/3}}\right). \tag{168}$$

This can be justified by ensuring that $I_v(u) \ll 1$ throughout the region around saddle points dominating the integral in Eq. (166). To see this, note that the integration contour can be deformed to pass through one or more saddle points along a path of steepest descent, with the integral then dominated by a relatively narrow region around those saddle points. In general the suitability of this approximation may depend sensitively on x; below we focus on the region $-\frac{\sigma^4}{4D} < x < \frac{\sigma^4}{4D} + \mathcal{O}(D^{1/3})$, which will comprise the "bulk" of $f(x)$ as well as the "fixation class" (the region dominating $\int f(x)w(x)dx$).

The relevant saddle points are located at u^* which satisfy

$$\frac{\sigma^6}{8D^2}\left(\frac{x}{\sigma^4/4D} - 2u^* + (u^*)^2\right) + I_v'(u^*) = 0, \tag{169}$$

and which can be approximated as

$$u_\pm^* \approx 1 \pm \frac{1}{\left(\frac{\sigma^2}{2D}\right) D^{1/3}} \sqrt{\frac{\sigma^4/4D - x}{D^{1/3}}} \tag{170}$$

after further assuming that $|I_v'(1)| \ll \frac{\sigma^6}{8D^2}$ (or more precisely, that $|I_v'(u)| \ll \frac{\sigma^6}{8D^2}$ for $|u| \lesssim \mathcal{O}(1)$, along with the assumption $\left(\frac{\sigma^2}{2D}\right) D^{1/3} \gg 1$, which turns out to be equivalent to the assumption $ND^{1/3} \gg 1$). Upon making the appropriate deformation of the integration contour, the integral in Eq. (166) is dominated by a region of width Δ around u_-^* (for $x < \frac{\sigma^4}{4D}$) or a region of width Δ around *both* u_-^* and u_+^* (for $x > \frac{\sigma^4}{4D}$), where

$$\Delta \sim \frac{1}{\left(\frac{\sigma^2}{2D}\right) D^{1/3}} \min\left[\left(\frac{\sigma^4/4D - x}{D^{1/3}}\right)^{-1/4}, 1\right]. \tag{171}$$

It follows that the region $|u| \lesssim \mathcal{O}(1)$ dominates the integral in Eq. (166), throughout the region of x considered here. The conditions $|I_v(u)| \ll 1$ and $|I_v'(u)| \ll \frac{\sigma^6}{8D^2}$ for $u \lesssim \mathcal{O}(1)$ then justify the approximations yielding Eq. (168); these conditions are closely approximated by the conditions $|I_v(1)| \ll 1$ and $|I_v'(1)| \ll \frac{\sigma^6}{8D^2}$.

A similar approximation can be made to obtain $w(x)$; Eq. (163) can be rewritten as

$$w(x) \propto \frac{\sigma^2/2D}{2\pi i} \int_{-i\infty}^{i\infty} \exp\left[\frac{\sigma^6}{8D^2}\left(\frac{xu}{\sigma^4/4D} - u^2 + \frac{u^3}{3}\right) + I_v(u)\right] du \tag{172}$$

with $I_v(u)$ again defined by Eq. (167), which simplifies to

$$w(x) \propto \exp\left[\frac{\sigma^2 x}{2D}\right] \mathrm{Ai}\left(\frac{\sigma^4/4D - x}{D^{1/3}}\right) \tag{173}$$

provided, again, that $I_v(u)$ is negligible within the region that dominates the integral in Eq. (172). For a particular x, this region may differ from the region of u dominating the integral in Eq. (166): for $x > \sigma^4/4D$, these regions are similar, while for $x < \sigma^4/4D$, the integral is dominated by the region around u_+^* (as opposed to the region around u_-^*, for $f(x)$). However, throughout the region $-\sigma^4/4D < x < \sigma^4/4D + \mathcal{O}(D^{1/3})$ of x

considered, the same conditions $|I_v(1)| \ll 1$ and $|I_v'(1)| \ll \frac{\sigma^6}{8D^2}$ justify the neglect of $I_v(u)$ within the region dominating Eq. (172).

Note that the condition $|I_v'(1)| \ll \frac{\sigma^6}{8D^2}$ can be written

$$I_v'(1)/\left(\frac{\sigma^6}{8D^2}\right) = \frac{\int \mu(s)ds\left[e^{\frac{\sigma^2 s}{2D}} - 1 - \frac{\sigma^2 s}{2D} - \frac{1}{2}\left(\frac{\sigma^2 s}{2D}\right)^2\right]}{\frac{1}{2}\int \mu(s)ds\left(\frac{\sigma^2 s}{2D}\right)^2} \ll 1 \tag{174}$$

which is, roughly speaking, satisfied if $\sigma^2 s/2D \ll 1$ throughout the region dominating $\int \mu(s)s^2 ds$. Because, in the infinitesimal regime, $T_c \approx \sigma^2/2D$ (Neher and Hallatschek, 2013), Eq. (174) essentially requires that $T_c s \ll 1$ for a substantial majority of available fitness effects s, as stated in the main text. The other condition $|I_v(1)| \ll 1$ has a similar interpretation.

C.2 Approximation of $f(x)$ and $w(x)$ within the MSSM regime

As in the previous Subsection, here we approximate the integrals in Eq. (162) and Eq. (163) by first approximating the integrand in the vicinity of its saddle points, and then employing an integral representation of the Airy function. The integrand in Eq. (162) possesses saddle points whose locations $z_s(x)$ satisfy

$$x + vz_s(x) + M_0\left[-z_s(x)\right] = 0. \tag{175}$$

Validity of the infinitesimal approximation requires that the relevant saddle points z_s are small enough in magnitude to justify a Taylor approximation of Eq. (175) to second-order in z_s. Within the MSSM regime, this Taylor approximation is not necessarily justified. Instead, in the next Subsection we make a local approximation to Eq. (175), valid for x within $\mathcal{O}(b)$ of c, and obtain approximate expressions for $f(x)$ and $w(x)$ within this fixation class. Using a similar approach, we then turn to obtain approximate solutions for $f(x)$ for x within the "bulk" of the fitness distribution.

Solution within the Fixation Class: $|c - x| \lesssim \mathcal{O}(b)$

To approximate Eq. (162) for x within the fixation class, we shift our integration contour to the line from $-T_c - i\infty$ to $-T_c + i\infty$, and expand the exponent of our integrand around its value at $z = -T_c$, the relevant (and only) saddle point of the integrand for $x = c$. That is,

$$f(x) = \frac{1}{2\pi i}\int_{-T_c - i\infty}^{-T_c + i\infty} \exp\left[xz + \frac{vz^2}{2} - M_{-1}(-z)\right] dz, \tag{176}$$

which can be rewritten as

$$f(x) = \frac{e^{-T_c x + \frac{vT_c^2}{2} - M_{-1}(T_c)}}{2\pi i b}\int_{-i\infty}^{i\infty} \exp\left[\left(\frac{c-x}{b}\right)u - u^3/3 - S_v(u)\right] du, \tag{177}$$

with

$$S_v(u) \equiv \frac{\int \mu(s)e^{T_c s}\frac{b}{s}\left[e^{us/b} - 1 - \frac{us}{b} - \frac{1}{2}\left(\frac{us}{b}\right)^2 - \frac{1}{6}\left(\frac{us}{b}\right)^3\right] ds}{\int \mu(s)e^{T_c s}\left[\frac{1}{2}\left(\frac{s}{b}\right)^2\right] ds}. \tag{178}$$

Note that Eq. (177) is obtained using the expansion $M_{-1}(-z) = \sum_{k=0}^{\infty}\frac{(-1)^k}{k!}M_{k-1}(T_c)(z+T_c)^k$ and the substitution $u \equiv -b(z + T_c)$. An integral representation of the Airy function, $\mathrm{Ai}(x) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} e^{xt - t^3/3}dt$, can then be applied to yield

$$f(x) = \frac{e^{\frac{vT_c^2}{2} - M_{-1}(T_c)}}{b}e^{-T_c x}\mathrm{Ai}\left(\frac{c-x}{b}\right), \tag{179}$$

provided that $S_v(u)$ can be neglected throughout the region (within the vicinity of saddle points) that dominates the integral in Eq. (177). Saddle points are located at u^* which satisfy

$$(u^*)^2 + S_v'(u^*) = \frac{c-x}{b} \tag{180}$$

with

$$S_v'(u) = \frac{\int \mu(s) e^{T_c s}\left[e^{us/b} - 1 - \frac{us}{b} - \frac{1}{2}(\frac{us}{b})^2\right] ds}{\int \mu(s) e^{T_c s} \frac{1}{2}(\frac{s}{b})^2 ds}. \tag{181}$$

We approximate Eq. (180) by

$$u_\pm^* \approx \pm\sqrt{\frac{c-x}{b}} \tag{182}$$

which, provided that $|S_v'(u)| \ll 1$ for $|u| \lesssim \mathcal{O}(1)$, holds self-consistently throughout the region $|c-x| \lesssim \mathcal{O}(b)$ of x considered. The region of u within $\Delta \sim \min\left[1, \left(\frac{c-x}{b}\right)^{-1}\right]$ of u^* dominates the integral in Eq. (177); the approximation yielding Eq. (179) is then justified throughout the region of x considered if $|S_v(1)| \ll 1$ and $|S_v'(1)| \ll 1$. A similar calculation yields

$$w(x) \propto \frac{e^{T_c x - \frac{vT_c^2}{2} + M_{-1}(T_c)}}{2\pi i b} \int_{-i\infty}^{i\infty} \exp\left[\left(\frac{c-x}{b}\right) u - u^3/3 + S_v(-u)\right] du, \tag{183}$$

which we approximate by $w(x) \propto e^{T_c x}\mathrm{Ai}\left(\frac{c-x}{b}\right)$ when $S_v(-u)$ can be neglected. The same conditions $|S_v(1)| \ll 1$ and $|S_v'(1)| \ll 1$ identified above can be shown to ensure that $S_v(-u)$ is negligible in the relevant regions.

Solution within the Bulk Class: $c - x \gtrsim \mathcal{O}(b)$

Here, we obtain an approximate solution for $f(x)$ for x *below* the fixation class. Similar manipulations can be carried out to yield an approximate solution for $w(x)$ within this region. Within this region, approximating the locations of saddle points—solutions $z_s(x)$ to Eq. (175)—as in Eq. (180) is not justified. Instead we obtain approximations to the integrals in Eq. (162) and Eq. (163) in terms of $z_s(x)$, which must be obtained numerically. Note that for each $x < c$ there are two real solutions $z_s(x)$ to Eq. (175): one in which $z_s(x) < -T_c$ and another in which $z_s(x) > -T_c$. Throughout the rest of this section, we denote by z_s the solution $z_s(x)$ to Eq. (175) such that $z_s(x) > -T_c$, which will be the relevant saddle point through which it is useful to deform the integration contour.

After shifting the integration contour to the line from $z_s - i\infty$ to $z_s + i\infty$ and making the substitution $iu \equiv \left(\frac{M_2(-z_s)}{2}\right)^{1/3}(z - z_s)$, Eq. (162) can be rewritten

$$f(x) = \frac{e^{z_s x + \frac{vz_s^2}{2} - M_{-1}(-z_s)}}{2\pi\left(\frac{M_2(-z_s)}{2}\right)^{1/3}} \int_{-\infty}^{\infty} \exp\left[-\left(\frac{v - M_1(-z_s)}{2^{1/3} M_2^{2/3}(-z_s)}\right) u^2 - \frac{iu^3}{3} - T(u)\right] du, \tag{184}$$

where $T(u) \equiv \sum_{k=4}^{\infty} \frac{M_{k-1}(-z_s)}{k!}\left(-\frac{2^{1/3} iu}{M_2^{1/3}(-z_s)}\right)^k$. We wish to approximate $f(x)$ by neglecting $T(u)$. The integral representation $\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ru^2 - iu^3/3} du = e^{2r^3/3}\mathrm{Ai}(r^2)$—which assumes $\mathrm{Re}(r) \geq 0$—can then be applied to yield

$$f(x) \approx \left(\frac{2}{M_2(-z_s)}\right)^{1/3} \exp\left(z_s x + \frac{vz_s^2}{2} - M_{-1}(-z_s) + \frac{[v - M_1(-z_s))]^3}{3M_2^2(-z_s)}\right) \mathrm{Ai}\left(\left[\frac{v - M_1(-z_s)}{2^{1/3} M_2^{2/3}(-z_s)}\right]^2\right). \tag{185}$$

When $v - M_1(-z_s) \gg M_2^{2/3}(-z_s)$, an asymptotic approximation of the Airy function can be applied to yield

$$f(x) \approx \frac{1}{\sqrt{2\pi\left[v - M_1(-z_s)\right]}} \exp\left(z_s x + \frac{v z_s^2}{2} - M_{-1}(-z_s)\right). \tag{186}$$

For x sufficiently close to 0, Eq. (186) reduces to the expression for $f(x)$ obtained within the infinitesimal approximation—that is, Eq. (168). In particular, for sufficiently small x,

$$f(x) \approx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}. \tag{187}$$

Note that Eq. (187) requires an approximation for σ^2. Taking $\sigma^2 = v - U\langle s\rangle$, with v computed using the MSSM approximation as described in the main text, should yield a good approximation to $f(x)$ within its bulk (provided the conditions of validity of the MSSM approximation are satisfied).

Appendix D MUTATIONAL FIXATION PROBABILITIES

In this Appendix, we evaluate

$$Np_{\text{fix}}(s) = \frac{\int f(x)w(x+s)dx}{\int f(x)w(x)dx} \tag{188}$$

using the approximate expressions for $f(x)$ and $w(x)$ identified within the MSSM regime. In particular, we make use of our result that $f(x) \propto e^{-T_c x}\text{Ai}\left(\frac{c-x}{b}\right)$, and $w(x) \propto e^{T_c x}\text{Ai}\left(\frac{c-x}{b}\right)$, within the region of x dominating $\int f(x)w(x)dx$ (in which $|c-x| \lesssim \mathcal{O}(b)$). We focus throughout on the case $|s| \ll b$; as discussed in the main text, validity of the MSSM approximation requires that $|s| \ll b$ for a substantial majority of fixed mutations. Given the assumption $|s| \ll b$, the region dominating $\int f(x)w(x+s)dx$ is approximately the same as the region dominating $\int f(x)w(x)dx$, justifying the use of our approximate results for $f(x)$ and $w(x)$.

We first consider the case in which $s < 0$, so that

$$Np_{\text{fix}}(s) \approx \frac{\int^{x_c} e^{-T_c x}\text{Ai}\left(\frac{c-x}{b}\right) e^{T_c(x+s)}\text{Ai}\left(\frac{c-x-s}{b}\right) dx}{\int^{x_c} \text{Ai}^2\left(\frac{c-x}{b}\right) dx} \approx \frac{e^{T_c s}\int_{z_0+\frac{1}{T_c b}}^{\infty} \text{Ai}(u)\text{Ai}(u-\frac{s}{b})du}{\int_{z_0+\frac{1}{T_c b}}^{\infty} \text{Ai}^2(u)du}, \tag{189}$$

where in the last equality we used our approximate result for x_c in Eq. (27). Expanding $\text{Ai}(u-\frac{s}{b})$ to second order in $\frac{s}{b}$,

$$\frac{Np_{\text{fix}}(s)}{e^{T_c s}} \approx 1 - \frac{s}{2b}\frac{\int_{z_0+\frac{1}{T_c b}}^{\infty} \frac{d}{du}\text{Ai}^2(u)du}{\int_{z_0+\frac{1}{T_c b}}^{\infty} \text{Ai}^2(u)du} + \frac{1}{2}\left(\frac{s}{b}\right)^2 \frac{\int_{z_0+\frac{1}{T_c b}}^{\infty} u\text{Ai}^2(u)du}{\int_{z_0+\frac{1}{T_c b}}^{\infty} \text{Ai}^2(u)du} \tag{190}$$

which simplifies to

$$\frac{Np_{\text{fix}}(s)}{e^{T_c s}} \approx 1 + \frac{1}{2(T_c b)^2}\left(\frac{s}{b}\right) - \frac{|z_0|}{6}\left(\frac{s}{b}\right)^2. \tag{191}$$

The $s > 0$ case is a bit more complicated, since $\int f(x)w(x+s)dx$ receives a contribution from the region of x in which $x + s > x_c$. In this case,

$$Np_{\text{fix}}(s) = \frac{\int^{x_c - s} e^{-T_c x}\text{Ai}\left(\frac{c-x}{b}\right) e^{T_c(x+s)}\text{Ai}\left(\frac{c-x-s}{b}\right) dx}{\int^{x_c} \text{Ai}^2\left(\frac{c-x}{b}\right) dx} + \frac{\int_{x_c - s}^{x_c} e^{-T_c x}\text{Ai}\left(\frac{c-x}{b}\right)(x+s)dx}{C_1 \int^{x_c} \text{Ai}^2\left(\frac{c-x}{b}\right) dx} \tag{192}$$

$$\approx \frac{e^{T_c s}\int_{z_0+\frac{1}{T_c b}}^{\infty} \text{Ai}(u)\text{Ai}(u+\frac{s}{b})du}{\int_{z_0+\frac{1}{T_c b}}^{\infty} \text{Ai}^2(u)du} + \frac{x_c e^{-T_c c}}{C_1}\frac{\int_{z_0+\frac{1}{T_c b}}^{z_0+\frac{1}{T_c b}+\frac{s}{b}} \text{Ai}(u)\, e^{T_c b u}du}{\int_{z_0+\frac{1}{T_c b}}^{\infty} \text{Ai}^2(u)\, du} \tag{193}$$

with $x_c = C_1 e^{T_c x_c} \mathrm{Ai}\left(\frac{c-x_c}{b}\right)$. The first term above can be easily evaluated using our $s < 0$ result. The second term can be approximated by noting that the integrand $\mathrm{Ai}(u)e^{T_c bu}$ peaks at $u = u^* \approx (T_c b)^2 \gg z_0 + \frac{1}{T_c b} + \frac{s}{b}$. An application of Watson's Lemma then yields

$$Np_{\mathrm{fix}}(s) \approx e^{T_c s}\left[1 - \frac{1}{2(T_c b)^2}\left(\frac{s}{b}\right) - \frac{|z_0|}{6}\left(\frac{s}{b}\right)^2 + \frac{1}{(T_c b)^3}\left(1 + T_c s\right)\left(1 - e^{-T_c s}\right)\right]. \tag{194}$$

In both the $s > 0$ and $s < 0$ cases, then,

$$p_{\mathrm{fix}}(s) \approx \frac{e^{T_c s}}{N} \tag{195}$$

given the assumption $T_c b \gg 1$ made throughout (as well as the assumption $s \ll b$ made here).

We note that we have defined T_c according to $N\int \mu(s) s e^{T_c s} ds \equiv v = N\int \mu(s) s p_{\mathrm{fix}}(s) ds$. Unless $\mu(s)$ consists of only a single effect size, this definition does not necessarily imply that $Np_{\mathrm{fix}}(s) = e^{T_c s}$. We have shown above that $Np_{\mathrm{fix}}(s) \approx e^{T_c s}$ in fact holds for $s \ll b$; together with the fact that $\int \mu(s) p_{\mathrm{fix}}(s) ds$ is dominated by the region of $s \ll b$, this implies that approximate self-consistency of $N\int \mu(s) s e^{T_c s} ds \equiv v = N\int \mu(s) s p_{\mathrm{fix}}(s) ds$ is achieved. Our analysis only explicitly enforced that $p_{\mathrm{fix}}(0) = 1/N$, and yet a consequence of this requirement is that a self-consistent rate v is obtained. The equivalence of the self-consistency conditions $p_{\mathrm{fix}}(0) = 1/N$ and $v = N\int \mu(s) s p_{\mathrm{fix}}(s) ds$ has been identified by Good et al. (2012) to hold *exactly* in the "tunable constraint" models introduced in Hallatschek (2011), provided that an analog of $\int f(x) w(x) dx = 1/N$ is taken as the "tuned constraint"; here these conditions are approximately equivalent.

Appendix E STATISTICS OF GENETIC DIVERSITY

In this Appendix, we provide derivations of the merger probabilities $\Lambda^0_{m,k}(t)$ that follow from our evolutionary model. We use these results to conclude that, after an initial delay period during which negligible coalescence events occur, genealogies are well-described by the Bolthausen-Sznitman coalescent (Bolthausen and Sznitman, 1998). In a similar calculation, we demonstrate a more direct correspondence with the BSC through the *partition structure* (which we define below) following from our evolutionary model. We then provide an explicit calculation of the neutral and selected site frequency spectra (SFS), separately obtaining the contribution to the SFS from mutations with ages t such that $b(t - T_c) \gg 1$, and from mutations with ages t such that $b(t - T_c) \lesssim \mathcal{O}(1)$. Finally, we conclude by briefly discussing how results from Fisher (2013) can be applied to obtain a *transition density* of the effective lead frequency $\nu_L(t)$ defined in the main text.

E.1 Merger Probabilities

The BSC is a type of Λ-coalescent, which in turn is a coalescent process that can be defined based on the merger rates $\lambda_{m,k}$ at which any given k-tuple of m blocks coalesce (Berestycki, 2009). In the BSC,

$$\lambda^{\mathrm{BSC}}_{m,k} \equiv \frac{(k-2)!(m-k)!}{(m-1)!}, \tag{196}$$

which can be contrasted to the Kingman coalescent, in which $\lambda_{m,k} = \delta_{k,2}$ (i.e. only pairwise merger events are allowed) (Berestycki, 2009). Following an approach similar to that taken by Neher and Hallatschek (2013), we can obtain the merger rates $\lambda_{m,k}$ for our evolutionary model by first calculating the merger *probabilities* $\Lambda^0_{m,k}(t)$. By $\Lambda^0_{m,k}(t)$, we denote the probability that of a sample of size m, a particular set of k individuals shares a common ancestor at t generations into the past, and the remaining $m - k$ individuals do *not* trace back to that same ancestor. These can be computed as

$$\Lambda^0_{m,k}(t) = N\left\langle \left(\frac{n_0(t)}{n_0(t) + \sum_{i=1}^{N-1} n_i(t)}\right)^k \left(\frac{\sum_{i=1}^{N-1} n_i(t)}{n_0(t) + \sum_{i=1}^{N-1} n_i(t)}\right)^{m-k} \right\rangle, \tag{197}$$

where the factor of N accounts for the N possible common ancestors of the k individuals which were alive at time t before the present generation. Here, the $\{n_i(t)\}_{i=0}^{N-1}$ are the (stochastic) lineage sizes of the N

individuals after a time t, which are identically distributed. Note also that our definition of $\Lambda^0_{m,k}(t)$ differs from the definition of $\Lambda^0_{m,k}(t)$ taken by Neher and Hallatschek (2013). In that work, $\Lambda^0_{m,k}(t)$ is defined as the probability that of a sample of m individuals, a given subset k of those individuals share a common ancestor, and that the remaining $m-k$ individuals trace back to $m-k$ *distinct* ancestors, at t generations. We take our definition because it enables us to obtain the same correspondence with the BSC observed by Neher and Hallatschek (2013), and because defined as such, the probabilities $\Lambda^0_{m,k}(t)$ can be used to compute the site frequency spectrum.

To simplify Eq. (197), we employ the identity $\Gamma(m)C^{-m} = \int_0^\infty dz z^{m-1} e^{-zC}$ with $C = n_0(t) + \sum_{i=1}^{N-1} n_i(t)$, which yields

$$\Lambda^0_{m,k}(t) = \frac{N}{\Gamma(m)} \int_0^\infty dz z^{m-1} \left\langle n_0^k e^{-zn_0} \left(\sum_{i=1}^{N-1} n_i \right)^{m-k} e^{-z\sum_{i=1}^{N-1} n_i} \right\rangle. \tag{198}$$

Assuming $n_0(t)$ and $\sum_{i=1}^N n_i(t)$ are independent of one another,

$$\Lambda^0_{m,k}(t) \approx (-1)^m \frac{N}{\Gamma(m)} \int_0^\infty dz z^{m-1} \partial_z^k \left(e^{-\Phi_0(z,t)} \right) \partial_z^{m-k} \left(e^{-(N-1)\Phi_0(z,t)} \right). \tag{199}$$

since, as we discuss in Appendix A, $e^{-\Phi_0(z,t)} \approx \left\langle e^{-zn(t)} \right\rangle$ (note that $\Phi_0(z,t) \equiv \int f(x)\phi_z(x,t)dx$). We discuss the behavior of $\Phi_0(z,t)$ in Appendix A. The key results we use here are that, for $t < T_c$, $\Phi_0(z,t)$ is approximately linear in z, and for $t > T_c$,

$$\Phi_0(z,t) \approx \frac{1}{N} \left(\frac{z}{\tilde{z}} \right)^{e^{1-t/T_c}} \tag{200}$$

within the relevant region of z (note that for $t > T_c$, our results primarily require only that $\Phi_0(z,t) \approx A(t) z^{e^{1-t/T_c}}$ for some z-independent function $A(t)$).

Linearity of $\Phi_0(z,t)$ in z implies that $\Lambda^0_{m,k}(t) \approx \delta_{k,1}$ and thus that merger events are negligible for $t < T_c$. For $t > T_c$, Eq. (199) can be simplified by noting that $\Phi_0(z,t) \ll 1$ in the relevant region of z, so that

$$\Lambda^0_{m,k}(t) \approx (-1)^{m+1} \frac{N}{\Gamma(m)} \int_0^\infty dz z^{m-1} \partial_z^k \Phi_0(z,t) \partial_z^{m-k} \left(e^{-(N-1)\Phi_0(z,t)} \right). \tag{201}$$

Substituting Eq. (200), and integrating by parts $b-k$ times, yields

$$\Lambda^0_{m,k}(t) \approx \frac{(-1)^{1-k}}{\Gamma(m)} \prod_{i=0}^{k-1} \left(e^{1-t/T_c} - i \right) \prod_{j=1}^{m-k} \left(e^{1-t/T_c} + j - 1 \right) \int_0^\infty \frac{dz}{z} N \Phi_0(z,t) e^{-(N-1)\Phi_0(z,t)}. \tag{202}$$

Using the substitution $u \equiv N\Phi_0(z,t)$, the z-integral in Eq. (202) can be evaluated to give

$$\Lambda^0_{m,k}(t) \approx \frac{\Gamma(k - e^{1-t/T_c})\Gamma(m-k+e^{1-t/T_c})}{\Gamma(m)\Gamma(1-e^{1-t/T_c})\Gamma(e^{1-t/T_c})}, \tag{203}$$

which, using the Euler reflection formula for $\Gamma(z)$, can be further simplified to

$$\Lambda^0_{m,k}(t) \approx \frac{\Gamma(k - e^{1-t/T_c})\Gamma(m-k+e^{1-t/T_c})}{\pi\Gamma(m)} \sin\left(\pi e^{1-t/T_c} \right). \tag{204}$$

From Eq. (204) the merger *rates* $\lambda_{m,k} \equiv d\Lambda^0_{m,k}/dt$ at $t = T_c$ (just after merger events occur at a non-negligible rate) follow as

$$\lambda_{m,k}(T_c) \approx \frac{(k-2)!(b-k)!}{a(b-1)!}. \tag{205}$$

These are the BSC merger rates $\lambda^{\rm BSC}_{m,k}$, scaled by the overall timescale T_c. The same correspondence is noted by Neher and Hallatschek (2013) to argue that genealogies within the infinitesimal regime resemble those

of the BSC (although in that case, the overall timescale identified is $\sigma^2/2D$, to which T_c reduces in the infinitesimal regime), after an initial delay period during which few coalescence events occur. The initial delay period can be interpreted, looking backward in time, as the time required for the ancestors of typical individuals—those sampled from the "bulk" of the fitness distribution—to reach the high-fitness edge of the fitness distribution; once this occurs, coalescence proceeds as described by the BSC.

The probability $Q_k(t) \equiv \Lambda_{k,k}(t)$ gives the probability with which k randomly chosen individuals share a common ancestor within t generations, and simplifies to

$$Q_k(t) \approx \frac{\Gamma(k - e^{1-t/T_c})}{\Gamma(k)\Gamma(1 - e^{1-t/T_c})}, \tag{206}$$

for $t > T_c$. Apart from the replacement of t by $(t - T_c)/T_c$, these are precisely the full time-dependent $Q_k(t)$ which are obtained in the BSC (Mohle and Pitters, 2014; Pitman, 1999). Thus the correspondence with the BSC in our model extends beyond the correspondence of the instantaneous merger rates at T_c generations. Note that $q_k(T_k) \equiv dQ_k/dt$ can be interpreted a probability distribution for the coalescence times T_k of a sample of k individuals. In particular,

$$q_2(T_2) = \begin{cases} 0 & T_2 < T_c \\ T_c^{-1} e^{1-T_2/T_c} & T_2 \geq T_c \end{cases}, \tag{207}$$

and therefore

$$T_c \equiv \langle T_2 \rangle / 2 \approx T_c. \tag{208}$$

That is, the timescale T_c of coalescence is the same as the *delay* timescale—the initial time during which negligible coalescence events occur. In the infinitesimal regime, the same correspondence (as well as the exponential distribution of pairwise coalescence times found in Eq. (207)) is inferred based on the instantaneous merger rates calculated at T_c generations, and observed in simulations, by Neher and Hallatschek (2013).

We note that although our overall conclusions are quite similar to those of Neher and Hallatschek (2013), our calculation differs in an important way. In obtaining $\Phi_0(z,t)$, Neher and Hallatschek (2013) approximate the solution $\phi_z(x,t)$ to Eq. (62) using a dominant balance argument—in particular, by taking

$$\phi_z(x,t) \approx \begin{cases} \phi_z^l(x,t) & x < P_z(t) \\ x & x > P_z(t) \end{cases}, \tag{209}$$

where $\phi_z^l(x,t)$ is the solution to Eq. (62) with the $\phi_z^2(x,t)$ term neglected, and $\phi_z^l(P_z(t)) \equiv P_z(t)$ (so that the two solutions match up at $x = P_z(t)$). Fisher (2013) argues that Eq. (209) is valid only for $t < T_c$, and describes the $t > T_c$ behavior in detail (which we review in Appendix A). Neher and Hallatschek (2013) analytically demonstrate a correspondence with the BSC by calculating the instantaneous merger rates at $t = T_c$ (more precisely, at $t = \sigma^2/2D \approx T_c$ since they work within the infinitesimal regime), so that Eq. (209) may still be valid; however, extending their calculation to later times yields various pathological results (e.g diverging expectation values for coalescence times including $\langle T_2 \rangle$). Our calculation, which makes use of behavior of $\phi_z(x,t)$ described by Fisher (2013), yields well-behaved predictions for the full time-dependence of the quantities $Q_k(t)$ and $\Lambda^0_{m,k}(t)$, and therefore the site frequency spectrum. As a result, we are able to analytically demonstrate aspects of the correspondence with the BSC which are observed in simulations, and justified using heuristic arguments, by Neher and Hallatschek (2013).

We emphasize that in our demonstration of a correspondence with the BSC, a key step is noting that Eq. (200) holds over the region of z which dominates the relevant integrals. That is, we take

$$\left\langle e^{-zn(t)} \right\rangle \approx \exp\left[-\frac{1}{N} \left(\frac{z}{\tilde{z}} \right)^{\alpha(t)} \right], \tag{210}$$

with $\alpha(t) = e^{1-t/T_c}$. The same generating function is used by Fisher (2013) to describe fluctuations in the total size $N(t)$ of a population, assuming its mean fitness increases at a fixed rate v (which can then be

used to describe fluctuations in the rate of adaptation at fixed population size). As noted by Fisher (2013) for the case of a fluctuating population size, from Eq. (210) it follows that $n(t)$ is drawn from a one-sided Lévy stable distribution with parameter $\alpha(t)$; the offspring number distribution—the pdf of $n(t)$—then falls off as $p(n) \sim 1/n^{1+\alpha(t)}$ (Nolan, 2018). In particular, at T_c generations the offspring distribution falls off as $1/n^2$. This offspring number distribution—which is the same as that obtained for the Luria-Delbrück process (Yule, 1925)—is observed (on different timescales) for rapidly adapting populations lying in several regimes of the parameter space (Desai and Fisher, 2007; Kosheleva and Desai, 2013), and has been explained as arising from the exponential amplification of fit lineages (Neher and Hallatschek, 2013).

It is well-understood that the offspring number distribution $p(n) \sim 1/n^2$ is related to the BSC (Schweinsberg, 2003); Hallatschek (2018) recently described a precise (neutral) stochastic process, involving an offspring number distribution of $p(n) \sim 1/n^2$, which can be considered a forward-time dual of the BSC (much like Wright-Fisher diffusion can be considered a forward-time dual of the Kingman coalescent). This duality has been used to argue that, after coarse graining to the timescale on which $p(n) \sim 1/n^2$, the genealogies of rapidly adapting populations can be described by the BSC. Interestingly, we have found above that Eq. (210) implies a correspondence with the BSC that goes beyond a correspondence on a coarse-grained timescale: after making reasonable approximations, the full time-dependence of the BSC partition structure follows from our evolutionary model. We provide further discussion of the relation between the dynamics under our model in the MSSM regime, and the dynamics under the model considered by Hallatschek (2018), in Subsection VI..3.

We note that the total population size N is fixed within our model, and therefore a $1/n^2$ offspring number distribution cannot strictly speaking apply. At minimum, the true offspring number distribution must have a cutoff at the total population size N. However, because only the relative *fractions* of the population comprised by any given lineage enter into our calculation of diversity statistics, we expect that Eq. (210) is still appropriate. We note also that Eq. (210) is only valid within a limited range of z; for instance, Eq. (210) yields pathological results both in the limit $z \to \infty$ (which gives extinction probabilities) and in the limit $z \to 0$ (which gives moments of $n(t)$). As we discuss in Appendix A, however, Eq. (210) is valid in the region of z which dominates integrals used to compute statistics of genetic diversity—roughly speaking, for z corresponding to $n(T_c)$ in the range $e^{-T_c b} \ll n(T_c)/N \ll e^{T_c b}$. Lineages of this size—while atypically large, relative to the expected value $\langle n(T_c)\rangle = 1$—make the dominant contribution to coalescence probabilities.

Selected Merger Probabilities

Using this framework, we can also compute the probability $\Lambda^s_{m,k}(t)/N$ that, of a sample of m individuals, a given set of k individuals trace back to a *specific* common ancestor at t generations (and that the remaining $m-k$ individuals do not trace back to that ancestor), conditioned on that common ancestor of the k individuals having acquired a mutation of effect size s at time t in the past. As noted in the main text, the probability $\Lambda^s_{m,k}(t)/N$ can be directly applied to compute the selected site frequency spectrum. For $t > T_c$, we have that

$$\frac{\Lambda^s_{m,k}(t)}{N} \approx (-1)^m \frac{1}{\Gamma(m)} \int_0^\infty dz z^{m-1} \partial_z^k \left(e^{-\Phi_s(z,t)}\right) \partial_z^{m-k} \left(e^{-(N-1)\Phi_0(z,t)}\right), \tag{211}$$

where Eq. (211) differs from Eq. (199) only by the replacement of $\partial_z^k \left(e^{-\Phi_0(z,t)}\right)$ by $\partial_z^k \left(e^{-\Phi_s(z,t)}\right)$. Here

$$\Phi_s(z,t) \equiv \int f(x-s)\phi_z(x,t)dx, \tag{212}$$

such that $e^{-\Phi_s(z,t)} \approx \left\langle e^{-zn_s(t)}\right\rangle$; that is $e^{-\Phi_s(z,t)}$ is the generating function for the (stochastic) size of a lineage seeded by a mutation effect s, after t generations. The properties of $\phi_z(x,t)$ and $\Phi(z,t)$ discussed in Appendix A imply that $\Phi_s(z,t) \approx e^{T_c s}\Phi_0(z,t)$ for $s \ll b$. Note that even with an additional factor of $e^{T_c s}$ (compared to $\Phi_0(z,t)$) we still have that $\Phi_s(z,t) \ll 1$ in the relevant region of z, and thus that $\Lambda^s_{m,k}(t) \approx e^{T_c s}\Lambda^0_{m,k}(t)$.

The $t < T_c$ and $k = 1$ case is also relevant for computing the heterozygosity. To simplify $\Lambda^s_{b,1}(t)$ for $t < T_c$, we note that $\phi_z(x,t)$ has x-dependence $\phi_z(x,t) \propto e^{xt}$ for the region of x and z of interest (at least until just

before $t = T_c$). As a result, we can make the approximation $\Phi_s(z,t) \propto e^{st}\Phi_0(z,t)$, which yields $\Lambda^s_{b,1}(t) \approx e^{st}$. That is, an individual descends from an ancestor at time t (which acquired a mutation of effect s at that same time) with probability $\Lambda^s_{b,1}(t)/N \approx e^{st}/N$.

E.2 Partition Structure

We can see a more direct correspondence with the BSC by considering the *partition structure* implied by our model. In doing so, we will define the *partition* $\Pi(t)$ of a sample of m individuals, at a time t into the past, as follows. The partition $\Pi(0)$ consists of m blocks of size 1 (with each block corresponding to a different individual). As the time t recedes into the past, blocks of $\Pi(t)$ merge together once their corresponding individuals share a common ancestor, with their sizes summing up (so that at time t, each block of $\Pi(t)$ consists of a group of individuals related by common ancestry at t generations). Thus once the entire sample shares a common ancestor, $\Pi(t)$ will consist of a single block of size m. For a sample of size m, the *partition structure* is then given by the probability $p(l_1, l_2, ..., l_m; t)$ that the partition $\Pi(t)$ of the sample consists of l_1 blocks of size 1, l_2 blocks of size 2, and so on.

To compute the partition structure $p(l_1, l_2, ..., l_m; t)$, we first compute the coalescence configuration probabilities

$$E(\{h_j\}_{j=1}^k) = \frac{N!}{(N-k)!}\left\langle \prod_{j=1}^k \left(\frac{n_j}{\sum_{i=1}^N n_i}\right)^{h_j}\right\rangle. \tag{213}$$

The quantity $E(\{h_j\}_{j=1}^k)$ gives the probability that at time t, the sample traces back to a total of k distinct ancestors, and further, that a particular set of h_1 individuals trace back to one of those ancestors, a particular set of h_2 individuals trace back to another, and so on (with $\sum_{j=1}^k h_j = m$). These can be calculated using manipulations similar to those used to calculate $\Lambda^0_{m,k}(t)$; note that a similar calculation is also performed by Desai et al. (2013), though that work is primarily focused on tracing the ancestors of a sample of individuals as they move from one discrete fitness class to another by acquiring a single beneficial mutation.

The probability E is given by

$$E(\{h_j\}_{j=1}^k, t) = \frac{N!}{(N-k)!}\left\langle \prod_{j=1}^k \left(\frac{n_j}{\sum_{i=1}^N n_i}\right)^{h_j}\right\rangle \tag{214}$$

Note that the combinatorial factor $N!/(N-k)!$ is included because we do not specify the common ancestors of the k "families" which share common ancestry. Using the identity $\Gamma(m)C^{-m} = \int_0^\infty z^{m-1}e^{-zC}dz$ with $C = \sum_{i=1}^N n_i$, and assuming the n_i are independent of one another, yields

$$E(\{h_j\}_{j=1}^k, t) \approx \frac{N!}{(N-k)!(m-1)!}\int_0^\infty dz z^{m-1}\left\langle e^{-zn_1}\right\rangle^{N-k}\prod_{j=1}^k\left\langle n_j^{h_j}e^{-zn_j}\right\rangle, \tag{215}$$

which can be expressed as

$$E(\{h_j\}_{j=1}^k, t) \approx \frac{(-1)^{m+k}N!}{(N-k)!(m-1)!}\int_0^\infty dz z^{m-1}e^{-N\Phi_0}\prod_{j=1}^k \partial_z^{h_j}\Phi_0(z,t). \tag{216}$$

As in our computation of $\Lambda^0_{m,k}(t)$, we use the properties of $\Phi_0(z,t)$ discussed by Fisher (2013). For $t < T_c$, linearity of $\Phi_0(z,t)$ in z implies that E is negligible unless $h_j = 1$ for all j (which describes the case in which each individual in the sample traces back to a distinct common ancestor). For $t > T_c$, Eq. (200) can be simplified to

$$E(\{h_j\}_{j=1}^k, t) \approx \frac{N!(e^{1-t/T_c})^{k-1}}{N^k(N-k)!(m-1)!}\prod_{j=1}^k\left(\frac{\Gamma(h_j - e^{1-t/T_c})}{\Gamma(1-e^{1-t/T_c})}\right)\int_0^\infty du e^{-u}u^{k-1}. \tag{217}$$

by using the $\Phi_0(z,t)$ in Eq. (200) and making the substitution $u \equiv N\Phi_0$. The integral in Eq. (217) evaluates to $\Gamma(k)$; using Stirling's approximation to approximate $N!/(N-k)! \approx N^k$ then yields

$$E(\{h_j\}_{j=1}^k, t) \approx \frac{(k-1)!(e^{1-t/T_c})^{k-1}}{(m-1)!} \prod_{j=1}^{k} \frac{\Gamma(h_j - e^{1-t/T_c})}{\Gamma(1-e^{1-t/T_c})}. \tag{218}$$

Note that E depends only on the *block counts* l_i, where l_i counts the number of blocks of size i (so that $k = \sum_{i=1}^m l_i$ and $m = \sum_{i=1}^m i l_i$). We can thus rewrite E as

$$E(\{h_j\}_{j=1}^k, t) \approx \frac{(k-1)!(e^{1-t/T_c})^{k-1}}{(m-1)!} \prod_{j=1}^{m} \left[\frac{\Gamma(j - e^{1-t/T_c})}{\Gamma(1-e^{1-t/T_c})}\right]^{l_j}. \tag{219}$$

The quantity E gives the probability of a particular partition with block counts $\{l_i\}_{i=1}^m$. The total number of partitions with l_1 blocks of size 1, l_2 blocks of size 2, and so on, is $m!/\prod_j \left[(j!)^{l_j} l_j!\right]$ (note that by definition a partition is an *unordered* collection of blocks). The probability $p(l_1, l_2, ..., l_m; t)$ then follows as

$$p(l_1, l_2, ..., l_m; t) \approx m(k-1)!(e^{1-t/T_c})^{k-1} \prod_{j=1}^{m} \frac{1}{l_j!} \left[\frac{\Gamma(j - e^{1-t/T_c})}{\Gamma(j+1)\Gamma(1-e^{1-t/T_c})}\right]^{l_j}. \tag{220}$$

The partition structure $p(l_1, l_2, ..., l_n; t)$ in Eq. (220) is precisely that of a partition drawn from a two parameter Poisson-Dirichlet distribution $PD(\alpha, \theta)$ with $\alpha = e^{1-t/T_c}$ and $\theta = 0$ (Pitman, 1995; Pitman and Yor, 1997). Importantly, the partition structure of the BSC is *also* described by a Poisson-Dirichlet distribution $PD(\alpha, 0)$ with $\alpha = e^{-t}$ (Berestycki, 2009). Thus, given the approximations made above, for $t > T_c$ the partition structure under our evolutionary model corresponds precisely to that under the BSC (up to the replacement of t by $(t - T_c)/T_c$, which reflects the delay period and coalescence timescale arising within our model).

The correspondence of the partition structure for our evolutionary model with that of the BSC does not imply that genealogies within our model are perfectly described by the BSC. For example, a further assumption of the BSC is that successive merger events are independent of one another; in our evolutionary model, heritable fitness variation among individuals may result in non-independence among merger events (though Neher and Hallatschek (2013) suggest, by a heuristic argument, that lineages within the appropriate region of fitness space equilibrate on a timescale much faster than the timescale of coalescence, such that successive merger events *are* largely independent of one another). However, the partition structure contains much more information than simply the instantaneous merger rates $\Lambda^0_{m,k}(T_c)$ or the time-dependent probabilities $Q_k(t)$, and its correspondence implies a relatively strong resemblance between genealogies of our model and those of the BSC.

E.3 The Site Frequency Spectrum

As discussed in the main text, the selected site frequency spectrum $P_s(k|m)$ can be computed as

$$P_s(k|m) = \mu \int_0^\infty \binom{m}{k} \Lambda^s_{m,k}(t) dt. \tag{221}$$

Substituting into Eq. (221) the $\Lambda^s_{m,k}(t)$ found above, we have

$$P_s(k|m) \approx T_c \mu \binom{m}{k} \left[\delta_{k,1}\left(\frac{e^{T_c s} - 1}{T_c s}\right) + e^{T_c s} \int_0^1 dy \frac{\sin \pi y}{\pi y} B(k-y, m-k+y)\right] \tag{222}$$

where we made the substitution $y \equiv e^{1-t/T_c}$ in the integral; note that $B(x_1, x_2)$ denotes the Beta function which satisfies $B(x_1, x_2) = \Gamma(x_1)\Gamma(x_2)/\Gamma(x_1 + x_2)$. The first term in Eq. (222) denotes a contribution from observed mutations with ages $t < T_c$; consistent with our approximate result that coalescence events do not occur until T_c generations into the past have elapsed, this term is only present (up to the level of

our approximation) for the case $k = 1$, corresponding to singletons. Apart from an overall scale factor, the contribution from $t > T_c$ (the second term in Eq. (222)) exactly matches the SFS of the true BSC, as calculated by Kersting et al. (2019). Kersting et al. (2019) computes the SFS of the true BSC directly from its known partition structure; a similar calculation is carried out by Neher and Hallatschek (2013) for the common-allele portion of the BSC SFS. For completeness, here we simplify Eq. (222) for the case $s = 0$ (which gives the neutral site frequency spectrum); results for the selected site frequency spectrum follow immediately.

First, we note that $P_0(1|m)$ simplifies to

$$P_0(1|m) \approx T_c \mu m \left[1 + \int_0^1 dy \frac{\sin \pi y}{\pi y} B(1-y, m-1+y)\right] \tag{223}$$

while $P_0(1|2)$, the pairwise heterozygosity, further simplifies to

$$P_0(1|2) \approx 4T_c\mu \approx 2 \langle T_2 \rangle \mu. \tag{224}$$

That is, we recover the well-known "molecular clock" result for $P_0(1|2)$: the average number of (neutral) pairwise differences among individuals is given by the average time since they share a common ancestor, times the per-locus mutation rate μ. Note that the contribution of $\Lambda_{2,1}(t)$ with $t < T_c$ is important for computing $P_0(1|2)$ (otherwise, our calculation would disagree with the "molecular clock" result by a factor of 2). For large m and $k = 1$, the contribution from $t < T_c$ dominates and $P_0(1|m) \sim T_c\mu m$; this is precisely the contribution of mutations which occur on the m terminal branches of the sample, none of which merge until T_c generations into the past have elapsed.

The BSC SFS in the large sample size limit

To analyze the more general case with $k > 1$, it is useful to consider the large m limit, such that the allele frequency $\nu \equiv k/m$ can be treated as a continuous variable. The contribution from $t < T_c$ vanishes for large N, and the distribution of neutral site frequencies $h_0(\nu)$ follows as

$$h_0(\nu) = \lim_{m\to\infty} m P_0(m\nu|m) \approx \lim_{m\to\infty} \frac{T_c\mu}{\nu(1-\nu)} \int_0^1 dy \frac{\sin \pi y}{\pi y} \frac{B(m\nu - y, m(1-\nu)+y)}{B(m\nu, m(1-\nu))}. \tag{225}$$

The integral on the right-hand side of Eq. (225) has been analyzed by Kersting et al. (2019); we briefly reproduce this analysis here. The right-hand side of Eq. (225) can be further simplified using Stirling's approximation, which yields

$$h_{\text{neu}}(\nu) \approx \frac{T_c U_n}{\nu(1-\nu)} \int_0^1 dy \frac{\sin \pi y}{\pi y} \left(\frac{1-\nu}{\nu}\right)^y. \tag{226}$$

The integral in Eq. (226) can be evaluated exactly in terms of functions arctan and the exponential integral Ei, yielding

$$h_0(\nu) \approx \frac{T_c\mu}{\nu(1-\nu)} \left(\frac{i}{2\pi}\left[\text{Ei}\left(-i\pi + \log\frac{1-\nu}{\nu}\right) - \text{Ei}\left(i\pi + \log\frac{1-\nu}{\nu}\right)\right] - \frac{1}{\pi}\arctan\left[\frac{\pi}{\log\frac{1-\nu}{\nu}}\right]\right) \tag{227}$$

where in Eq. (227), the range of arctan must be taken as $(0, \pi)$ instead of the usual $(-\pi/2, \pi/2)$. For rare alleles ($\log\frac{1-\nu}{\nu} \to \infty$) Eq. (227) simplifies to

$$h_0(\nu) \approx \frac{T_c\mu}{\nu(1-\nu)} \left(\frac{\left(\frac{1-\nu}{\nu}\right)}{\log^2\frac{1-\nu}{\nu}} - \frac{1}{\log\frac{1-\nu}{\nu}}\right) \approx \frac{T_c\mu}{\nu^2 \log^2(1/\nu)}, \tag{228}$$

while for common alleles ($\log\frac{1-\nu}{\nu} \to -\infty$),

$$h_{\text{neu}}(\nu) \approx \frac{T_c\mu}{(1-\nu)\log\left(\frac{1}{1-\nu}\right)}. \tag{229}$$

Note that the integral in Eq. (226)—and in particular, the region of y which dominates that integral—can be examined to identify the typical allelic ages of mutations observed at a given frequency. For example, if $\nu > 1/2$, the integrand peaks at $y = 0$ and the integral is dominated by the region of width $1/|\log \frac{1-\nu}{\nu}|$ above 0. As a result, alleles at a frequency $1 - \nu \ll 1$ have typical ages $t \gtrsim T_c \left(1 + \log\log \frac{1}{1-\nu}\right)$. Alternatively, if $\nu < 1/2$, the integrand peaks at an intermediate y^* between 0 and 1 which satisfies $1/y^* - \pi \cot \pi y^* = \log \frac{1-\nu}{\nu}$. In the limit $\nu \to 0$, $1 - y^* \to 1/\log(1/\nu)$ so that $t^* \to T_c(1 + 1/\log(1/\nu))$. In particular, $b(t^* - T_c) > 1$ when $\log(1/\nu) < T_c b$, or equivalently, when $\nu > e^{-T_c b}$.

Deviations from the BSC SFS at lower frequencies

For lower frequencies such that $\log(1/\nu) > T_c b$, the time integral yielding the SFS receives a substantial contribution from times t such that $b(t - T_c) \lesssim \mathcal{O}(1)$. At these times the approximation $\Phi_0(z, t) \propto z^{e^{1-t/T_c}}$ used to obtain $\Lambda^0_{m,k}(t)$, and thus the SFS, begins to break down. In Appendix A, we analyze the behavior of $\phi_z(x, t)$ at intermediate times such that $b(t - T_c) \lesssim \mathcal{O}(1)$, extending the analysis of $\phi_z(x, t)$ carried out by Fisher (2013) at long times such that $b(t - T_c) \gg 1$. Because the corresponding $\Phi_0(z, t)$ does not have a simple power law dependence on z, the manipulations carried out above to obtain the merger rates $\Lambda^0_{m,k}(t)$, and in turn the contribution to the SFS, are difficult to extend to these times. To obtain the contribution to the SFS of these mutations, we will instead directly extract $p(\nu; t, x)$, the probability that a lineage has frequency ν at time t, given it was seeded at time 0 with relative fitness x, from $\left\langle e^{-\zeta\nu(x,t)} \right\rangle$, the generating function for the frequency $\nu(x, t)$ of the lineage. The generating function $\left\langle e^{-\zeta\nu(x,t)} \right\rangle$ can in turn be obtained from the quantity $\phi_z(x, t)$ computed in Appendix A, according to

$$1 - \left\langle e^{-\zeta\nu(x,t)} \right\rangle = 1 - \left\langle e^{-(\zeta/N)n(t|x)} \right\rangle \approx \phi_{\zeta/N}(x, t). \tag{230}$$

We will then compute the contribution of these mutations to the SFS by integrating $p(\nu; t, x)$ over both times t and initial fitnesses x.

We will define the function $E[R; t, x]$ according to

$$E\left[\log(1/\nu)/(T_c b); t, x\right] \equiv \nu^2 p(\nu; t, x), \tag{231}$$

with the scaling variable $R \equiv \log(1/\nu)/(T_c b)$, so that

$$1 - \left\langle e^{-\zeta\nu(x,t)} \right\rangle = \int_0^1 \left(1 - e^{-\zeta\nu'}\right) \frac{E\left[\log(1/\nu')/(T_c b); t, x\right]}{(\nu')^2} d\nu', \tag{232}$$

which can be written as

$$\frac{1 - \left\langle e^{-\zeta\nu(x,t)} \right\rangle}{\zeta} = \int_0^\infty \left(\frac{1 - e^{-\zeta\nu'}}{\zeta\nu'}\right) E\left[\log(1/\nu')/(T_c b); t; x\right] d\log(1/\nu'), \tag{233}$$

by making a change of variables. For $\log(1/\nu) \ll \log\zeta$, the factor $(1 - e^{-\zeta\nu})/(\zeta\nu)$ is exponentially suppressed, while for $\log(1/\nu) \gg \log\zeta$, $(1 - e^{-\zeta\nu})/(\zeta\nu) \approx 1$. Upon making the ansatz that $E[R; t, x]$ varies on R scales of $\mathcal{O}(1)$ (which can later be checked to be self-consistent), we then have

$$\frac{1 - \left\langle e^{-\zeta\nu(x,t)} \right\rangle}{\zeta} \approx \int_{\log\zeta}^\infty E\left[\log(1/\nu')/(T_c b); t; x\right] d\log(1/\nu'), \tag{234}$$

and thus

$$E\left[\log(1/\nu)/(T_c b); t; x\right] \approx -\frac{\partial}{\partial\log\zeta} \left(\frac{1 - \left\langle e^{-\zeta\nu(x,t)} \right\rangle}{\zeta}\right)\Bigg|_{\log\zeta = \log 1/\nu}. \tag{235}$$

Using our result from Appendix A for $\phi_{\zeta/N}(x, t)$ when $b(t - T_c) \lesssim \mathcal{O}(1)$, given in Eq. (135), we can evaluate the derivative on the right-hand side of Eq. (235) as a contour integral. We then have

$$E\left[R; t, x\right] \approx x_c e^{T_c(x - x_c)} \mathrm{Ai}'(z_0) \left(\frac{1}{2\pi i} \int_{\mathcal{C}} e^{b(t - T_c)z} \frac{\mathrm{Ai}\left(\frac{c-x}{b} + z\right)}{\mathrm{Ai}^2\left(z + z_0 + R\right)} dz\right), \tag{236}$$

where the contour $\mathcal{C}$ runs from $e^{-2\pi i/3}\infty$ through 0 to $e^{2\pi i/3}\infty$. To see this, note that the integral in Eq. (236) has second order poles at $z = z_j - (z_0 + R)$ with corresponding residues

$$\frac{1}{\mathrm{Ai}'(z_j)^2}\,\frac{\partial}{\partial z}\left[e^{b(t-T_c)z}\mathrm{Ai}\left(\frac{c-x}{b}+z\right)\right]\Bigg|_{z=z_j-(z_0+R)}, \tag{237}$$

so that Eq. (236) follows from Eq. (135) and an application of the residue theorem. Integrating over times from $b(t - T_c) = -\infty$ to $b(t - T_c) = \infty$, we have

$$\int E(R;t,x)dt \approx \frac{x_c e^{T_c(x-x_c)}\mathrm{Ai}'(z_0)}{b}\,\frac{\mathrm{Ai}\left(\frac{c-x}{b}\right)}{\mathrm{Ai}^2\left(z_0+R\right)}, \tag{238}$$

which can then be integrated against the fitness distribution to yield

$$\frac{\nu^2 h_s(\nu)}{N\mu} = \int^{\min[x_c-bR,x_c+s]} f(x-s)\left(\int E(R;t,x)dt\right)dx, \tag{239}$$

the SFS of mutations with fitness effect s. We can consider the cases $s > -bR$ and $s < -bR$ separately, in each case using the expression for $f(x)$ given in Eq. (30). If $s > -bR$,

$$\frac{\nu^2 h_s(\nu)}{N\mu} \approx \frac{e^{T_c s}}{NT_c b^2\mathrm{Ai}^2(z_0+R)}\int_{z_0+\frac{1}{T_c b}+R}^{\infty}\mathrm{Ai}(u)\mathrm{Ai}(u+s/b)du, \tag{240}$$

which evaluates to

$$\begin{aligned}\frac{\nu^2 h_s(\nu)}{T_c\mu} &\approx \frac{e^{T_c s}}{(T_c b)^2}\\ &\times\left[\frac{\mathrm{Ai}(z_0+\frac{1}{T_c b}+R+s/b)\mathrm{Ai}'(z_0+\frac{1}{T_c b}+R)-\mathrm{Ai}(z_0+\frac{1}{T_c b}+R)\mathrm{Ai}'(z_0+\frac{1}{T_c b}+R+s/b)}{(s/b)\mathrm{Ai}^2(z_0+R)}\right],\end{aligned} \tag{241}$$

and, in the limit $s/b \to 0$, to

$$\frac{\nu^2 h_s(\nu)}{T_c\mu} \approx \frac{e^{T_c s}}{(T_c b)^2}\left(\left[\frac{\mathrm{Ai}'\left(z_0+\frac{1}{T_c b}+R(\nu)\right)}{\mathrm{Ai}\left(z_0+\frac{1}{T_c b}+R(\nu)\right)}\right]^2-\left(z_0+\frac{1}{T_c b}+R(\nu)\right)\right). \tag{242}$$

For small R (and large $T_c b$), Eq. (242) simplifies to

$$\frac{\nu^2 h_s(\nu)}{T_c\mu} \approx \frac{e^{T_c s}}{(T_c b)^2 R^2}, \tag{243}$$

and thus

$$h_s(\nu) \sim \frac{T_c\mu e^{T_c s}}{\nu^2\log^2(1/\nu)}, \tag{244}$$

which is precisely the asymptotic behavior of the BSC SFS at low frequencies. In the limit of large $R + z_0$, Eq. (242) simplifies to

$$\frac{\nu^2 h_s(\nu)}{T_c\mu} \approx \frac{e^{T_c s}}{2(T_c b)^2}\sqrt{\frac{1}{z_0+R}}, \tag{245}$$

and thus

$$h_s(\nu) \sim \frac{T_c\mu e^{T_c s}}{2(T_c b)^{3/2}\nu^2\sqrt{T_c b z_0+\log(1/\nu)}}. \tag{246}$$

Note, however, that when $z_0 + R \gg 1$, the contribution to the SFS is dominated by lineages with initial fitnesses $(c-x)/b \sim z_0+R \gg 1$; this can be seen from the integral in Eq. (240), which is boundary-dominated in this case. These fitnesses lie well below the fixation class, where the approximation $f(x) \propto e^{-T_c x}\mathrm{Ai}\left(\frac{c-x}{b}\right)$, used to obtain Eq. (240), begins to break down.

If $s < -bR$ (which requires that the mutation is deleterious) we have

$$h_s(\nu) \sim \frac{T_c \mu e^{T_c s}}{\nu^2 \log^2(1/\nu)}, \tag{247}$$

for small R, which again matches the low-frequency behavior of the BSC SFS.

The contribution to the SFS from deterministic lineage trajectories

To obtain the contribution to the SFS from mutations with ages $t < T_c$ (which will be important for low frequencies) we will treat the growth of lineages deterministically. For simplicity, we focus our attention on the neutral SFS $h_0(\nu)$, which can be computed as

$$h_0(\nu) = N\mu \int_0^\infty dt \int f(x) p_{\mathrm{est}}(x) \delta\left(\nu - \frac{n(t|x)}{N}\right) dx, \tag{248}$$

where $p_{\mathrm{est}}(x)$ denotes the *establishment* probability of a lineage founded at relative fitness x, and $n(t|x)$ denotes the size of a deterministically growing lineage at time t, given it was established at time 0 with relative fitness x. A solution for $n(t|x)$ is reviewed, and the dynamics of deterministically growing lineage trajectories are discussed, in Appendix F (see Eq. (275)). Here we will use the result that $n(t|x) = n_0(x)e^{xt-J(t)}$, where $J(t) \equiv vt^2/2 - \int \frac{\mu(s)}{s}\left(e^{st} - st - 1\right) ds$; $n_0(x) = 1/p_{\mathrm{est}}(x)$ is an additional factor we include here which denotes the size of an *established* lineage—the size of a lineage at the point such that its future growth is largely described by the deterministic forces of mutation and selection, as opposed to genetic drift.

The time integral in Eq. (248) can easily be carried out, yielding

$$h_0(\nu) = N\mu \int dx f(x) p_{\mathrm{est}}(x) / \left| \frac{\partial}{\partial t} \frac{n(t|x)}{N} \right|_{n(t|x)/N=\nu}, \tag{249}$$

where the time derivative in Eq. (249) is evaluated at t such that $n(t|x)/N = \nu$, for a given x and ν. To simplify Eq. (249), we note that the quantity $n(t|x)$ peaks at time t_x solving

$$\left.\frac{dJ}{dt}\right|_{t=t_x} = x, \tag{250}$$

with peak size $n_{\mathrm{peak}}(x)$ given by

$$\frac{n_{\mathrm{peak}}(x)}{n_0(x)} = e^{xt_x - J(t_x)} = e^{t_x J'(t_x) - J(t_x)}. \tag{251}$$

For t near the peak—which will turn out to often dominate the contribution of a lineage to the SFS—we have

$$\frac{n(t|x)}{n_{\mathrm{peak}}(x)} \approx e^{-\frac{1}{2}J''(t_x)(t-t_x)^2}, \tag{252}$$

from which it follows that

$$\left| \frac{\partial}{\partial t} \frac{n(t|x)}{N} \right|_{n(t|x)/N=\nu} \approx \nu \sqrt{2J''(t_x) \log\left(\frac{n_{\mathrm{peak}}(x)}{N\nu}\right)}. \tag{253}$$

Assuming the contribution of lineages near their peak dominates the SFS, we then have

$$h_0(\nu) \approx \frac{N\mu}{\nu} \int \frac{f(x) p_{\mathrm{est}}(x) dx}{\sqrt{2J''(t_x) \log(n_{\mathrm{peak}}(x)/N\nu)}}, \tag{254}$$

where the lower limit of integration in Eq. (254) is such that $n_{\rm peak}(x) > N\nu$. A simplification occurs in that

$$f(x) \approx \frac{e^{-xt_x + J(t_x)}}{\sqrt{2\pi J''(t_x)}} = \frac{n_0(x)}{n_{\rm peak}(x)\sqrt{2\pi J''(t_x)}}, \tag{255}$$

(see Eq. (186) in Appendix C) so that

$$h_0(\nu) \approx \frac{N\mu}{2\nu\sqrt{\pi}} \int \frac{n_0(x) p_{\rm est}(x)}{n_{\rm peak}(x) J''(t_x)\sqrt{\log(n_{\rm peak}(x)/(N\nu))}} dx. \tag{256}$$

The factors $n_0(x)$ and $p_{\rm est}(x)$ cancel one another in the numerator of Eq. (256), and thus contribute only logarithmic corrections to our final result. For simplicity, we therefore neglect the x-dependence of $n_0(x)$ throughout the rest of this calculation; at the end of our calculation we replace n_0 with the appropriate value based on the dominant range of x contributing to the SFS at a given frequency. Noting that $d\log n_{\rm peak}(x) = t_x dx$, Eq. (256) can be rewritten as

$$h_0(\nu) \approx \frac{N\mu}{2\nu\sqrt{\pi}} \int_{N\nu}^{\infty} \frac{dn_{\rm peak}}{n_{\rm peak}^2\sqrt{\log(n_{\rm peak}/(N\nu))}} \frac{1}{L(n_{\rm peak})}, \tag{257}$$

with

$$L(n_{\rm peak}) \equiv t_x J''(t_x) = \frac{d\log n_{\rm peak}(x)}{dt_x}. \tag{258}$$

Note that in Eq. (257), and in our definition of $L(n_{\rm peak})$, we have replaced the dependence of $n_{\rm peak}$ on x with its dependence on t_x. The quantity $L(n_{\rm peak})$ can be simplified to $t_x\left[v - \int \mu(s) ds s e^{st_x}\right]$, where t_x is the time a lineage size peaks, given it peaks at a size $n_{\rm peak}$. Changing the variable of integration to $u \equiv n_{\rm peak}/(N\nu)$, we have

$$h_0(\nu) \approx \frac{N\mu}{2N\nu^2\sqrt{\pi}} \int_1^{\infty} \frac{du}{u^2\sqrt{\log u}} \frac{1}{L(N\nu u)} \tag{259}$$

The integral over u is dominated by u close to 1 on the logarithmic scale over which L varies; we thus have

$$h_0(\nu) \propto \frac{N\mu}{2N\nu^2 L(N\nu)\sqrt{\pi}} \int_1^{\infty} \frac{du}{u^2\sqrt{\log u}} = \frac{U_n}{2\nu^2 L(N\nu)}. \tag{260}$$

For concreteness, we now consider two special cases. First, provided that $st_x \ll 1$ for relevant s, $L(n_{\rm peak}) \approx \sigma^2 t_x$ and $n_{\rm peak} \approx n_0 e^{\sigma^2 t_x^2/2}$, so $L(n_{\rm peak}) \approx \sigma\sqrt{2\log(n_{\rm peak}/n_0)}$. Mutations observed at these frequencies occurred in the "bulk" of the fitness distribution, with relative fitness $x_f \sim \sigma\sqrt{2\log(N\nu/n_0)}$; we can thus take $n_0 \sim 1/\sigma$, to logarithmic accuracy. We thus have

$$h(\nu) \propto \frac{1}{\nu^2\sqrt{\log N\sigma\nu}} \tag{261}$$

if $\nu \ll \frac{1}{N} e^{\sigma^2/(2s^2)}$ for relevant s. .

Alternatively, if $b(T_c - t_x) \ll 1$, then $L(n_{\rm peak}(x)) \approx 2T_c b^3 (T_c - t_x)$ and $b^2(t_x - T_c)^2 \approx (c - x)/b$. We then have that

$$\log(n_{\rm peak}(x)/n_0) \approx \log N x_c - T_c b^3 (t_x - T_c)^2 + T_c b z_0 + 1, \tag{262}$$

with $n_0 \sim 1/x_c$, and thus

$$L(N\nu) \approx 2T_c b^2 \sqrt{z_0 + \frac{1}{T_c b}\log\frac{1}{\nu}}. \tag{263}$$

It follows that

$$h(\nu) \sim \frac{1}{\nu^2\sqrt{T_c b z_0 + \log\frac{1}{\nu}}}, \tag{264}$$

and that the dominant relative fitness x_f at which observed mutations occurred is given by $(c - x_f)/b \sim z_0 + \frac{1}{T_c b}\log(1/\nu)$. Note also that this approach breaks down if $\log(1/\nu) < T_c b|z_0|$. Mutations observed at frequencies such that $\log(1/\nu) < T_c b|z_0|$ are likely to have originated at relative fitnesses $x_f > c$, for which trajectories monotonically increase with time, under a deterministic approximation.

We note that the above approach resembles an approach taken by Neher and Shraiman (2011) to compute the rare-allele portion of the SFS. In that work, lineages contributing to the SFS are assumed to be founded by individuals within the Gaussian portion of the fitness distribution $f(x) \sim e^{-x^2/2\sigma^2}$, with sizes growing deterministically according to $n(t|x) \propto e^{xt-\sigma^2 t^2/2}$. We can carry out a similar calculation to obtain the SFS at very low frequencies (such that the SFS is not necessarily dominated by mutations at the peaks of their deterministic trajectories). Given these assumptions, in contrast to the previous case we can more easily carry out the x integral in Eq. (248) before the t integral, which yields

$$h_0(\nu) \approx \frac{N\mu}{\nu\sqrt{2\pi\sigma^2 N\nu}} \int_0^\infty \frac{dt}{t} p_{\text{est}} \sqrt{n_0} e^{-\frac{\sigma^2 t^2}{8} - \frac{\log^2\left(\frac{N\nu}{n_0}\right)}{2\sigma^2 t^2}}. \tag{265}$$

As in the above case, both p_{est} and n_0 depend on t (via the x that yields a lineage of size $N\nu$ at time t); we will neglect this dependence and see that, under this assumption, the factor p_{est} in the numerator cancels with a factor $1/n_0$, leaving only a logarithmic dependence on n_0. We can then evaluate the integral in Eq. (265) exactly, yielding

$$h_0(\nu) \approx \frac{N\mu p_{\text{est}}}{\nu^{3/2}\sqrt{2\pi\sigma^2 N/n_0}} K_0\left(\frac{\log\frac{N\nu}{n_0}}{2}\right), \tag{266}$$

where K_0 is the modified Bessel function of the second kind. Using the asymptotic expansion $K_0(x) \sim \sqrt{\frac{\pi}{2x}}e^{-x}$ for large x then gives

$$h_0(\nu) \sim \frac{\mu}{\nu^2\sqrt{2\sigma^2 \log\frac{N\nu}{n_0}}} \approx \frac{\mu}{\nu^2\sqrt{2\sigma^2 \log N\sigma\nu}}, \tag{267}$$

if $N\sigma\nu \gg 1$. Note that in the final equality above we replaced n_0 by $1/\sigma$, since the SFS is dominated by mutations which arose at relative fitness $\mathcal{O}(\sigma)$ (i.e., in the "bulk" of the fitness distribution).

E.4 Transition Density of the Effective Lead Frequency

To obtain the transition density, we first consider the generating function

$$H[\theta w(x), t] = \left\langle e^{-N\theta\int w(x)g(x,t)dx}\right\rangle = \left\langle e^{-\theta\nu_L(t)}\right\rangle \tag{268}$$

for the effective lead frequency $\nu_L(t) \equiv N\int g(x,t)w(x)dx$. Note that the effective lead frequency ν_L can also be interpreted as the fixation probability of a lineage. Using the method of Appendix A, this can be done by solving Eq. (93) with the initial condition $\phi(x,0) = \theta w(x)$. Fisher (2013) considers essentially the same quantity as in Eq. (268), with the only difference being that in that work, $g(x,t)$ is replaced by the population-wide fitness distribution, which can fluctuate. In our case, $g(x,t)$ is the fitness distribution of a lineage (which can fluctuate), evolving in competition with a population whose mean fitness changes at rate v. We can thus directly apply the following result from Eq. (68) of Fisher (2013):

$$\left\langle e^{-\theta\nu_L(t)}\right\rangle \approx \left\langle e^{-\theta^{\alpha(\tau)}\nu_L(t-\tau)}\right\rangle, \tag{269}$$

with $\alpha(t) = e^{-t/T_c}$ (in our notation). As discussed by Fisher (2013), Eq. (269) implies that $\nu_L(t)$ is drawn from a (one-sided) Lévy stable distribution with a tail falling off as $1/\nu_L^{1+\alpha(\tau)}$. Eq. (269) is valid for $\tau > \log(1/\theta)/x_c$ and $\theta \lesssim \mathcal{O}(1)$. As a result, Eq. (269) does not adequately describe the probability distribution of ν_L for $\nu_L \ll 1$; our results below thus only apply to lineages with "macroscopic" effective lead frequency ν_L (i.e., with a significant probability of taking over the population).

The result for $\left\langle e^{-\theta\nu_L(t)}\right\rangle$ in Eq. (269) is essentially the same as the result for a similar quantity obtained by Kosheleva and Desai (2013) (though that case involves transitions in frequencies among discrete fitness classes at discrete points in times; the analog of $\alpha(\tau)$ also differs slightly). The derivation in Kosheleva and Desai (2013) can be carried over to our case, yielding, in our notation,

$$G_t(\nu_L(t),\nu_L(0)) = \frac{\sin\left[\pi\alpha(t)\right]\nu_L(0)\left(1-\nu_L(0)\right)}{\nu_L(t)\left(1-\nu_L(t)\right)\pi\left[\left(1-\nu_L(0)\right)^2\left(\frac{\nu_L(t)}{1-\nu_L(t)}\right)^{\alpha(t)}+\nu_L(0)^2\left(\frac{1-\nu_L(t)}{\nu_L(t)}\right)^{\alpha(t)}+2\nu_L(0)(1-\nu_L(0))\cos\left(\pi\alpha(t)\right)\right]}. \tag{270}$$

We note that a key aspect of the derivation of Eq. (270) is essential in allowing its application to our case. In particular, the calculation of Kosheleva and Desai (2013) computes Eq. (270) essentially by integrating over the transitions of $\nu_{L,1}(t)$ (given a starting value $\nu_L(0)$) as well as $\nu_{L,2}(t)$ (given a starting value, $1-\nu_L(0)$, equal to the complement of $\nu_L(0)$), with $\nu_L(t)$ then equal to $\nu_{L,1}/\left(\nu_{L,1}(t)+\nu_{L,2}(t)\right)$. This can be thought of as a way to explicitly enforce that the $\nu_L(t)$ of the different lineages in a population sum up to 1, as they should in our model, and resembles the way in which our calculation of genetic diversity statistics only involves *relative* sizes of lineages. In both cases, the generating functions used correspond to Lévy distributions with diverging mean and variance, and so care must be taken to deal only with "relative" quantities in considering a population of fixed size.

Appendix F DETERMINISTIC LINEAGE TRAJECTORIES

In this Appendix, we review the dynamics of a lineage whose fitness distribution evolves deterministically and in competition with a population steadily increasing in fitness at rate v. More precisely, we consider the size trajectory $n_l(x_0,t)$ of a lineage founded at relative fitness x_0 as well as its time-dependent distribution $g(X,t)$ of fitnesses. For clarity, we proceed by writing down and directly analyzing an equation for $g(X,t)$ with number fluctuations neglected; alternatively, the same results can be obtained using a solution for the generating function $\left\langle e^{-zn(t|x_0)}\right\rangle$ from Appendix A, which treats $n(t|x_0)$ as a random variable. These results facilitate a comparison of the behavior obtained within the infinitesimal and MSSM approximations (which are stochastic treatments) and the behavior obtained within a purely deterministic treatment, which is provided in the Discussion. These results are also useful in motivating an interpretation of the timescales T_c and $\sigma^2/2D$ as *sweep* timescales with the infinitesimal and MSSM regimes, respectively; interpretations of $\sigma^4/4D$ and c are also possible.

In treating the deterministic behavior of a lineage, apart from neglecting any stochasticity in births or deaths, we assume that the deterministic growth of the lineage does not perturb the steady advance of the mean fitness of the population from its expected value vt. Without loss of generality, we assume a population-wide mean fitness of 0 when $t=0$. Our focus is on the quantity $g(X,t)$, with $g(X,t)dX$ denoting the number of individuals within the lineage whose absolute fitness lies between X and $X+dX$, at a time t since the foundation of the lineage. The equation for $g(X,t)$ is then given by Eq. (87) with the noise term $\eta_g(X)$ further neglected:

$$\frac{\partial g(X,t)}{\partial t} = (X-vt)g(X,t) + \int \mu(s)\left[g(X-s,t)-g(X,t)\right]ds. \tag{271}$$

Our approach to analyzing Eq. (271) is similar to the approach carried out in Fisher (2013) to analyze the population-wide fitness distribution, assuming it evolves deterministically at fixed rate v (with the total population size allowed to vary in that model). We solve for the the Laplace transform $\tilde{g}(z,t) \equiv \int g(X,t)e^{-zX}dX$, which evolves according to

$$\frac{\partial\tilde{g}(z,t)}{\partial t} = -\frac{\partial\tilde{g}(z,t)}{\partial z} + \left[\int \mu(s)(e^{-zs}-1)ds - vt\right]\tilde{g}(z,t), \tag{272}$$

subject to the initial condition

$$\tilde{g}(z,0) = \frac{e^{-zx_0}}{N}, \tag{273}$$

assuming the lineage is founded by a single individual with relative fitness x_0. Eq. (272) can be solved using the method of characteristics, yielding

$$\tilde{g}(z,t) = \frac{1}{N}\exp\left[(t-z)x_0 + \int \frac{\mu(s)ds}{s}\left(e^{(t-z)s} - e^{-zs} - st\right) - \frac{vt^2}{2}\right]. \tag{274}$$

At time t, the total lineage size $n_l(x_0,t) = N\tilde{g}(0,t)$ is

$$n_l(x_0,t) = \exp\left[x_0 t + \int \frac{\mu(s)ds}{s}(e^{st} - 1 - st) - \frac{vt^2}{2}\right]. \tag{275}$$

The same result for $n_l(x_0,t)$ could be obtained using the formalism of Appendix A by substituting into $\left\langle e^{-zn(t|x)}\right\rangle = e^{-\phi_z(x,t)}$ the approximate solution for $\phi_z(x,t)$ obtained in Appendix E, and then evaluating $\partial_z \left\langle e^{-zn(t|x)}\right\rangle\big|_{z=0}$.

The mean relative fitness $\bar{x}_l(x_0,t)$ of the lineage, along with the lineage-wide fitness variance $\sigma_l^2(x_0,t)$, can be similarly extracted by taking z-derivatives of $\tilde{g}(z,t)$:

$$\bar{x}_l(x_0,t) = x_0 - vt + \int \mu(s)(e^{st} - 1)ds, \tag{276}$$

$$\sigma_l^2(t) = \int \mu(s)s\left(e^{st} - 1\right)ds, \tag{277}$$

and for $p \geq 2$,

$$\frac{\partial^p}{\partial(-z)^p}\log\tilde{g}(z,t) = \int \mu(s)s^{p-1}(e^{st} - 1)e^{-zs}ds. \tag{278}$$

Of course, a consequence of Eq. (271) is that the lineage size $n(t|x_0)$ grows as $\frac{dn(t|x_0)}{dt} = n(t|x_0)\bar{x}(x_0,t)$. That is, the instantaneous growth rate of the lineage is equal to its mean relative fitness. Similarly, the rate of change of the mean relative fitness of the lineages matches what would be expected from Fisher's fundamental theorem: $\frac{d}{dt}\bar{x}(x_0,t) = \sigma_l^2(t) + \int \mu(s)sds$. Of note is that the scale c, as defined in the main text, demarcates the region $x_0 > c$ for which lineage sizes increase in size indefinitely from the region $x_0 < c$ within which they do not. Lineages founded at $0 < x_0 < c$ will increase initially in size but attain a maximum size at some point t^*, with $t^* < T_c$ (and with T_c as defined in the main text by Eq. (17)). For x_0 sufficiently close to c, this local maximum will be attained at $t^* \approx T_c\left[1 - (1/T_c b)\sqrt{\frac{c-x_0}{b}}\right]$. Lineages founded at $x_0 < 0$ decrease in size indefinitely. We note also that $\sigma_l^2(t)$ is independent of x_0, and increases from 0 to match σ^2, the population-wide fitness variance, over the course of precisely T_c generations. For $t > T_c$, $\sigma_l^2(t)$ exceeds σ^2, suggesting that the approximations made above break down.

The above expressions for $n_l(x_0,t)$, $\bar{x}_l(x_0,t)$ and $\sigma_l^2(t)$ involve integrals over $\rho(s)$. In the next subsection, we restrict our attention to the infinitesimal regime, in which the above expressions can be simplified in terms of σ^2 and D. We then turn to the MSSM regime, in which simplifications can be made in terms of the scales T_c, b and c.

F.1 Lineage Trajectories in the Infinitesimal Regime

Assuming validity of the infinitesimal approximation, for $t \lesssim \frac{\sigma^2}{2D}$, Taylor approximations can be made within Eq. (275), Eq. (276) and Eq. (277), yielding

$$\log n_l(x_0,t) \approx x_0 t - \frac{\sigma^2 t^2}{2} + \frac{Dt^3}{3} \tag{279}$$

with the quantities σ^2 and D as defined in the main text (Neher et al., 2014b). The mean and variance in fitness of the lineage-wide fitness distribution can be computed in a similar way, yielding

$$\bar{x}_l(x_0,t) \approx x_0 - \sigma^2 t + Dt^2, \tag{280}$$

and

$$\sigma_l^2(t) \approx 2Dt. \tag{281}$$

Under the additional assumption $\frac{1}{U} \ll t \ll \frac{\sigma^2}{2D}$, the (relative) fitness density $g(x,t)$ of the lineage can be approximated as

$$g(x,t) \approx \frac{n_{l,inf}(x_0,t)}{\sqrt{2\pi\sigma_{l,inf}^2(t)}} \exp\left[-\frac{(x-\bar{x}_{l,inf}(x_0,t))^2}{2\sigma_{l,inf}^2(t)}\right]. \tag{282}$$

for $|x-\bar{x}_{l,inf}(x_0,t)| \lesssim \sigma_{l,inf}(x_0,t)$. This can be done by writing down $g(X,t)$ as the inverse Laplace transform of Eq. (274), identifying the regions dominating the integral (saddle points, which are located at $z \approx \frac{x-\bar{x}_{l,inf}}{\sigma_{l,inf}^2(t)}$), and arguing that the relevant terms can be neglected within those regions, provided that $Ut \gg 1$. Alternatively, we can note that Eq. (282) solves

$$\frac{\partial g(X,t)}{\partial t} = (X-vt)g(X,t) + D\frac{\partial^2 g(X,t)}{\partial X^2} - U\langle s\rangle \frac{\partial g(X,t)}{\partial X}, \tag{283}$$

which can be obtained from (271) by a Taylor approximation of $g(X-s,t)$ to second order in s (similar to the expansion of $f(x-s)$ in the usual presentation of the infinitesimal approximation). Within the region $|x-\bar{x}_{l,inf}(x_0,t)| \lesssim \sigma_{l,inf}(x_0,t)$ of interest, $g(X,t)$ varies on a scale of $\sigma_l(t)$, suggesting validity of this Taylor approximation when $s \ll \sigma_l(t)$ throughout the support of $\rho(s)$, or equivalently, when $Ut \gg 1$. Eq. (283) motivates the following description of the deterministic dynamics of lineages within the infinitesimal regime: lineages grow in size at at a rate given by their overall mean (relative) fitness advantage, advect upwards in absolute fitness at rate $U\langle s\rangle$, and diffuse outward in fitness with diffusion constant D.

The above analysis considers the behavior of a single lineage evolving deterministically, given its initial fitness x_0. We can easily extend our analysis to consider how, under the deterministic approximation, the collective set of descendants of a *class* of individuals evolves. In particular, given the results from the main text for the steady-state relative fitness distribution $f(x)$, we can compute the number $n(t;x_1,x_2)$ of descendants of the class of individuals with fitnesses between x_1 and x_2, after t generations. At the particular time $t=\sigma^2/2D$, $n(t;x_1,x_2)$ simplifies to

$$n(\sigma^2/2D;x_1,x_2) \equiv N\int_{x_1}^{x_2} f(x)n(x,\sigma^2/2D)dx \approx N\int_{\frac{\sigma^4/4D-x_2}{D^{1/3}}}^{\frac{\sigma^4/4D-x_1}{D^{1/3}}} \mathrm{Ai}(u)du. \tag{284}$$

As we discuss in the main text, within the infinitesimal regime, the region $\sigma^4/4D < x < x_c$ contributes a future common ancestor of the population with probability $\mathcal{O}(1)$ (which motivates us to refer to this region as the fixation class). Under the deterministic approximation, after $\sigma^2/2D$ generations, the lineage seeded collectively by the fixation class has reached a size of

$$n\left(\frac{\sigma^2}{2D};\frac{\sigma^4}{4D},x_c\right) = 0.94\times N. \tag{285}$$

individuals. This in turn motivates our interpretation of $\sigma^2/2D$ as a *sweep* timescale: over the course of $\sigma^2/2D$ generations, the fixation class collectively sweeps through to comprise a macroscopic fraction of the population, under the deterministic approximation. A timescale $\sigma^2/2D$ has also been interpreted as a sweep timescale by

As discussed in the main text, $D^{1/3}$ corresponds to the width of the fixation class. The above results results further suggest an interpretation of $D^{1/3}$ as a *diffusion* scale. To see this, we consider the ratio of lineage sizes seeded simultaneously with initial fitnesses x_0 and $x_1 > x_0$, which evaluates to $n(t|x_1)/n(t|x_0) = e^{(x_1-x_0)t}$. The lineage seeded at initial fitness x_1 is favored by selection on the timescale $\tau_{sel} = 1/(x_1-x_0)$. Over this same timescale, both lineages acquire fitness variance of the amount $\sigma_l^2(\tau_{sel}) = 2D/(x_1-x_0)$. The lineage-wide fitness distributions overlap substantially at times *before* $t=\tau_{sel}$ only if $\sqrt{2D/(x_1-x_0)} \gg (x_1-x_0)$, or equivalently, only if $(x_1-x_0) \ll D^{1/3}$. The fitness scale $D^{1/3}$ then corresponds roughly to the maximum fitness difference of two lineages such that their fitness distributions overlap substantially *before* selection acts to amplify the frequency of the more-fit lineage. The overlap of the fitness distributions can be thought of as resulting from *diffusion* of the lineages through fitness space; this diffusion is particularly important on fitness scales smaller than $D^{1/3}$.

F.2 Lineage Trajectories in the MSSM regime

Within the MSSM regime, analogous approximations can be made for $\log n(t|x_0)$ and its derivatives. Here, however, a Taylor approximation of e^{st} for $t < T_c$ is not necessarily justified, since $T_c s$ is not necessarily small throughout the bulk of $\mu(s)$. In fact, even the smaller quantity s/b need not be small throughout the bulk of $\mu(s)$; s/b is only required to be small within the region dominating $\int \mu(s)e^{T_c s}ds$. For deleterious mutations, the condition that $s/b \ll 1$ for s dominating $\int \mu(s)e^{T_c s}ds$ is strictly *weaker* than the condition that $s/b \ll 1$ for s dominating $\int \mu(s)ds$.) Instead, given that $s \ll b$ for s dominating $\int \mu(s)e^{T_c s}ds$, an approximation of Eq. (275) can be made for $|t - T_c| \lesssim \mathcal{O}(1/b)$, yielding

$$\log n_l(x_0, t) \approx \log n(x_0, T_c) + (x_0 - c)\,(t - T_c) + \frac{b^3(t - T_c)^3}{3}, \tag{286}$$

$$\bar{x}_l(x_0, t) \approx x_0 - c + b^3(t - T_c)^2, \tag{287}$$

and

$$\sigma_l^2(t) \approx \sigma^2 + 2b^3(t - T_c), \tag{288}$$

with

$$\log n_l(x_0, T_c) = T_c(x_0 - x_c) + \log T_c b \mathrm{Ai}'(z_0) N x_c, \tag{289}$$

where the last equality makes use of Eq. (31) from the main text.

As in the previous subsection, the quantity $n(T_c; x_1, x_2)$ can be evaluated, in this case by making use of Eq. (275) along with Eq. (179) for $f(x)$ obtained within the MSSM regime. Since Eq. (179) is accurate only for x within $\mathcal{O}(b)$ of c, x_1 and x_2 must also be within $\mathcal{O}(b)$ of c to yield reasonable accuracy in $n(T_c; x_1, x_2)$. Provided this condition is met,

$$n(T_c; x_1, x_2) = N \int_{\frac{c-x_2}{b}}^{\frac{c-x_1}{b}} \mathrm{Ai}(u)du, \tag{290}$$

and in particular,

$$n(T_c; c, x_c) \approx 0.94 \times N, \tag{291}$$

such that T_c can be interpreted as a *sweep* timescale, in the sense that the fixation class collectively sweeps to comprise a macroscopic fraction of the population over T_c generations. The timescale $1/b$ can be interpreted as the *rate* at which $n(T_c; c, x_c)$ approaches N; to see this, we note that for $c < x_0 < x_c$ and $|t - T_c| \lesssim \mathcal{O}(b)$, the quantity $n_l(x_0, t)$ varies on the timescale $1/b$. The fitness scale b also corresponds to the width of the fixation clas; however, the interpretation given in the previous subsection to $D^{1/3}$ as a diffusion fitness scale does not extend to this case.

Appendix G PARTICULAR DISTRIBUTIONS OF FITNESS EFFECTS

In this Appendix, we consider two classes of DFEs: stretched-exponential distributions, of the form

$$\rho(s) = \frac{1}{\chi} \frac{e^{-(s/\chi)^\beta}}{\Gamma(1 + \beta^{-1})} \tag{292}$$

considered by Desai and Fisher (2007), Good et al. (2012) and others, as well as (more briefly) gamma distributions, of the form

$$\rho(s) = \frac{\alpha}{\langle s \rangle \Gamma(\alpha)} \left(\frac{\alpha s}{\langle s \rangle}\right)^{\alpha - 1} e^{-\frac{\alpha s}{\langle s \rangle}}. \tag{293}$$

For these classes of DFEs, we simplify many of the expressions provided in the main text, focusing particular attention on the scale $\tilde{s}$ and its magnitude relative to the scales T_c^{-1} and b. In doing so, our primary goal is to evaluate the suitability of the infinitesimal and MSSM approximationes, particularly in the $N \to \infty$ limit. We limit our attention here to DFEs consisting only of beneficial mutations, although our analysis can easily be extended to consider DFEs consisting of deleterious mutations or a combination of the two.

We begin by considering the case of an exponential DFE. Attention on the case of a beneficial exponential DFE can be motivated by arguments from extreme value theory (Gillespie, 1984; Orr, 2003). We then briefly consider more general gamma distributions, and conclude by considering the case of a stretched-exponential DFE with $\beta > 1$. The requirement that $\beta > 1$ is a key requirement for validity of our approach; otherwise, the integrals $M_p(T_c)$, which we consider throughout, do not converge.

G.1 Beneficial Exponential DFE

The special case in which mutations are beneficial and exponentially-distributed yields particularly simple analytical results. We assume in this Subsection that $\mu(s) = \frac{U}{\langle s \rangle} e^{-s/\langle s \rangle}$. We will compute quantities of interest by first computing the moments $M_p(T_c)$, defined as $M_p(T_c) \equiv \int \mu(s) e^{T_c s} s^p ds$ for $p \geq 1$, $M_0(T_c) \equiv \int \mu(s) \left(e^{T_c s} - 1\right) ds$, and $M_{-1}(T_c) \equiv \int \frac{\mu(s)}{s} \left(e^{T_c s} - T_c s - 1\right)$. These evaluate to

$$M_p(T_c) = \begin{cases} \frac{U}{\langle s \rangle} \left[\log\left(\frac{1}{1 - T_c \langle s \rangle}\right) - T_c \langle s \rangle\right] & p = -1 \\ U \frac{T_c \langle s \rangle}{1 - T_c \langle s \rangle} & p = 0 \\ p! \frac{U \langle s \rangle^p}{(1 - T_c \langle s \rangle)^{p+1}} & p \geq 1. \end{cases} \tag{294}$$

In particular,

$$v \equiv M_1(T_c) = \frac{U \langle s \rangle}{(1 - T_c \langle s \rangle)^2}. \tag{295}$$

Consequently, $T_c \langle s \rangle$ approaches 1 from below as $N \to \infty$ (note this follows from the fact that v increases monotonically, without bound, with N). The fact that $T_c \langle s \rangle < 1$ ensures convergence of the $M_p(T_c)$ for all p, and implies that $\rho_f(s) \propto e^{-(1 - T_c \langle s \rangle) s / \langle s \rangle}$ is a strictly decreasing function of s on the interval $(0, \infty)$. The average effect size $\langle s_f \rangle$ of a fixed mutation simplifies to

$$\langle s_f \rangle = \frac{\langle s \rangle}{1 - T_c \langle s \rangle}, \tag{296}$$

such that

$$\frac{\langle s_f \rangle}{b} = \left(\frac{\langle s \rangle}{U}\right)^{1/3} \tag{297}$$

is independent of N. Since $\rho_f(s)$ is an exponential distribution, the standard deviation Δs_f of fixed fitness effects equates to $\langle s_f \rangle$, and thus $\tilde{s} = 3 \langle s_f \rangle$ given the definition of $\tilde{s}$ in Section II. Additionally,

$$T_c b = \left(\frac{U}{\langle s \rangle}\right)^{1/3} \frac{T_c \langle s \rangle}{1 - T_c \langle s \rangle}, \tag{298}$$

so that the conditions $\tilde{s} \ll b$ and $T_c b \gg 1$ are both met (implying suitability of our MSSM approximation) in the $N \to \infty$ limit, as long as $\langle s \rangle \ll U$.

Conditions of validity can be considered more precisely by considering the quantities $S_v(1)$ and $S_v'(1)$ discussed in Appendix C. These quantities evaluate to

$$S_v = \frac{U}{\langle s \rangle} \left(-\log\left[1 - \left(\frac{\langle s \rangle}{U}\right)^{1/3}\right] - \left[\left(\frac{\langle s \rangle}{U}\right)^{1/3} + \frac{1}{2}\left(\frac{\langle s \rangle}{U}\right)^{2/3} + \frac{1}{3}\left(\frac{\langle s \rangle}{U}\right)\right]\right) \tag{299}$$

and

$$S_v' = \frac{(\langle s\rangle/U)^{1/3}}{1-(\langle s\rangle/U)^{1/3}}, \tag{300}$$

with both S_v and S_v' tending to 0 as $\langle s\rangle/U \to 0$, and diverging for $\langle s\rangle \geq U$. We note that in the limit $\tilde{s}/b \to 0$, $S_v' \to \frac{\tilde{s}}{3b}$ (which is the same result for S_v' obtained when $\rho(s)$ is *not* an exponential distribution, but $\rho_f(s)$ is sufficiently sharply peaked—as in the case $\rho(s) = \delta(s - s_b)$, for instance). This can be seen as a motivation for taking the particular definition $\tilde{s} \equiv \langle s_f\rangle + 2\Delta s_f$; with this definition, the (arguably) simpler and more interpretable quantity $\tilde{s}/b$ can be used instead of S_v' in determining the suitability of the MSSM approximation. The quantity $\tilde{s}/b$ behaves the same way as S_v' in the opposing limiting cases that $\rho_f(s)$ is very sharply peaked, and that $\rho_f(s)$ is an exponential distribution. The same holds for the quantity S_v; $S_v \to \tilde{s}/12b$ in the limit $\tilde{s}/b \to 0$, both when $\rho_f(s)$ is sharply peaked and when $\rho_f(s)$ is an exponential distribution.

An asymptotic approximation for the $N \to \infty$ limit

We can obtain asymptotic results in the limit $N \to \infty$ at fixed $\langle s\rangle/U \ll 1$. As noted above, in this limit $T_c\langle s\rangle \to 1$, so this limit can alternatively be thought of as the limit $T_c b \to \infty$ at fixed $\langle s\rangle/U \ll 1$. We wish to obtain the leading behavior of $T_c b$ in terms of N by analyzing Eq. (31) in this limit. In doing so, it is useful to simplify Eq. (31) in terms of the large parameter $y \equiv (U/\langle s\rangle)^{1/3} T_c b$. Using the integrals evaluated in Eq. (294), we have, in the limit $T_c\langle s\rangle \to 1$,

$$\left(\frac{U}{\langle s\rangle}\right)^{1/3} \log\left[N\langle s\rangle\left(\frac{\langle s\rangle}{U}\right)^{1/3} y^2 T_c b\right] \approx \frac{y^2}{2} + |z_0| y \tag{301}$$

which can be further approximated as

$$y^2 \approx 2\left(\frac{U}{\langle s\rangle}\right)^{1/3} \log N\langle s\rangle\,. \tag{302}$$

if $y \gg 1$. This yields

$$1 - T_c\langle s\rangle \sim \sqrt{\frac{U}{2\langle s\rangle \log N\langle s\rangle}} \tag{303}$$

and therefore

$$v \sim 2\langle s\rangle^2 \log N\langle s\rangle \tag{304}$$

which is approximately valid if $\log N\langle s\rangle \gg U/\langle s\rangle$. Note that for $U/\langle s\rangle \gg 1$, an unreasonably large population size N may be required for this asymptotic approximation to be useful (and at the same time, $U/\langle s\rangle \gg 1$ is required for validity of the MSSM approximation). In particular, caution must be taken in interpreting Eq. (303) or Eq. (304) as a function of U or $\langle s\rangle$, since changes in either quantity can cause the asymptotic approximation to break down. In practice, the asymptotic approximation $v \sim 2(U\langle s\rangle^2)^{1/3}\langle s\rangle \log N\langle s\rangle$ generally provides better accuracy throughout the parameter regime we have considered in simulations. Additionally, in this limit

$$\frac{\langle s_f\rangle}{s_b} \sim \sqrt{\frac{2\langle s\rangle \log N\langle s\rangle}{U}}, \tag{305}$$

$$\frac{b}{\langle s\rangle} \sim \left(\frac{\langle s\rangle}{U}\right)^{1/6} \sqrt{2\log N\langle s\rangle}, \tag{306}$$

and

$$\frac{c}{\langle s\rangle} \sim 2\log N\langle s\rangle\,. \tag{307}$$

Single-s equivalence

Finally, we note that in the limit $N \to \infty$ the dynamics can be captured an effective mutational fitness spectrum $\mu_{eff}(s) = U_{eff}\delta(s - s_{eff})$, with

$$s_{eff} \equiv \frac{2\langle s\rangle}{1 - T_c\langle s\rangle} = 2\langle s_f\rangle \tag{308}$$

and

$$U_{eff} \equiv \frac{U}{2(1 - T_c\langle s\rangle)} e^{-2T_c\langle s\rangle/(1-T_c\langle s\rangle)} \tag{309}$$

With these definitions of U_{eff} and s_{eff}, $v = M_1(T_c) = U_{eff}s_{eff}e^{T_c s_{eff}}$ and $2b^3 = M_2(T_c) = U_{eff}s_{eff}^2 e^{T_c s_{eff}}$, matching the corresponding quantities that are obtained for $\mu(s) = U_{eff}\delta(s - s_{eff})$. In the limit $N \to \infty$, the quantities T_c and v depend on $\mu(s)$ only through $M_1(T_c)$ and $M_2(T_c)$; the dynamics can thus be described by the effective mutation rate U_{eff} and effective DFE $\delta(s - s_{eff})$.

G.2 Beneficial Gamma DFE

Much of the above analysis can be straightforwardly generalized to the case of gamma-distributed beneficial fitness effects, with

$$\mu(s) = \frac{\alpha\mu}{\langle s\rangle\Gamma(\alpha)}\left(\frac{\alpha s}{\langle s\rangle}\right)^{\alpha-1}\exp\left(-\frac{\alpha s}{\langle s\rangle}\right). \tag{310}$$

Here $\langle s\rangle$ denotes the mean effect size and $\alpha > 0$ denotes the shape parameter of the gamma distribution (such that the variance in effect sizes is $\langle s\rangle^2/\alpha$).

The moments $M_p(T_c)$ evaluate to

$$M_p(T_c) = \begin{cases} \frac{U}{\langle s\rangle}\left[\frac{\alpha}{\alpha-1}\left(\left(\frac{\alpha}{\alpha - T_c\langle s\rangle}\right)^{\alpha-1} - 1\right) - T_c\langle s\rangle\right] & p = -1 \\ U\left[\left(\frac{\alpha_b}{\alpha_b - T_c\langle s\rangle}\right)^{\alpha_b} - 1\right] & p = 0 \\ \frac{\Gamma(\alpha+p)}{\alpha^p\Gamma(\alpha)}\frac{U\langle s\rangle^p}{(1-T_c\langle s\rangle/\alpha)^{\alpha+p}} & p \geq 1 \end{cases} \tag{311}$$

with $v = M_1(T_c) \to \infty$ as $T_c\langle s\rangle \to \alpha$; consequently, $T_c\langle s\rangle$ approaches α from below as $N \to \infty$. (The expression for $M_{-1}(T_c)$ assumes $\alpha \neq 1$.)

The distribution of fixed effects in the MSSM regime is just $\rho_f(s) \propto \rho(s)e^{T_c s}$, which is a gamma distribution with shape parameter α and mean effect $\langle s_f\rangle = \langle s\rangle/(1 - T_c\langle s\rangle/\alpha)$; the variance in fixed effects is given by $\Delta s_f^2 = \langle s_f\rangle^2/\alpha$. The ratio $\langle s_f\rangle/b$ simplifies to

$$\frac{\langle s_f\rangle}{b} = \left(\frac{2}{1 + 1/\alpha}\right)^{1/3}\left(\frac{\langle s\rangle}{U}\right)^{1/3}\left(1 - \frac{T_c\langle s\rangle}{\alpha}\right)^{(\alpha-1)/3}, \tag{312}$$

and so as long as $\alpha > 1$, $\frac{\langle s_f\rangle}{b} \to 0$ (and thus $\tilde{s}/b \to 0$) in the $N \to \infty$ limit. Further, $T_c b \to \infty$ as $N \to \infty$, suggesting that the MSSM regime is approached in the $N \to \infty$ limit as long as $\alpha > 1$. In the opposite case $\alpha < 1$, the MSSM approximation is expected to break down as $N \to \infty$ since $\langle s_f\rangle/b \to \infty$. Note that this can be confirmed by simplifying the more precise conditions of validity $S_v'(1) \ll 1$ and $S_v(1) \ll 1$. Additionally, as might be expected, the infinitesimal approximation breaks down in the limit $N \to \infty$.

G.3 Beneficial Stretched-Exponential DFE

Here we consider beneficial stretched-exponential DFEs of the form

$$\rho(s) = \frac{1}{\chi\Gamma(1 + \beta^{-1})}e^{-(s/\chi)^\beta} \tag{313}$$

for $s > 0$, with $\beta > 1$. To do so involves evaluating integrals of the form $\int_0^\infty s^p e^{g(s)} ds$, with $g(s) \equiv T_c s - \left(\frac{s}{\chi}\right)^\beta$. In contrast to the $\beta = 1$ case, simple closed-form expressions for these integrals are not available when $\beta > 1$. In the limit that $T_c\chi \to 0$, these integrals can be approximated as

$$\int_0^\infty \rho(s) s^p e^{T_c s} ds \approx \int_0^\infty \rho(s) s^p ds = \frac{\Gamma\left(\frac{1+p}{\beta}\right)}{\Gamma(1/\beta)} \chi^p. \tag{314}$$

We focus here instead on the case in which $T_c\chi \gg 1$—which is obtained in the limit $N \to \infty$—in which these integrals and related expressions can be approximated using Laplace's method.

The quantity $g(s)$ peaks at s^* satisfying

$$s^* = \left(\frac{T_c\chi}{\beta}\right)^{\frac{1}{\beta-1}} \chi, \tag{315}$$

with

$$g(s^*) = T_c s^*(1 - 1/\beta), \tag{316}$$

and

$$g^{(k)}(s^*) = -\frac{\Gamma(1+\beta)}{\chi^k \Gamma(1+\beta-k)} \left(\frac{T_c s^*}{\beta}\right)^{1-k/\beta}, \tag{317}$$

for $k = 2$. The width $\Delta \equiv \left|g^{(2)}(s^*)\right|^{-1/2}$ of $e^{g(s)}$ scales, relative to s^*, as

$$\frac{\Delta}{s^*} \sim \frac{1}{\sqrt{\beta(\beta-1)}} \left(\frac{T_c\chi}{\beta}\right)^{-\frac{\beta}{2(\beta-1)}} = \frac{1}{\sqrt{(\beta-1)as^*}}, \tag{318}$$

and

$$\left|\frac{\left[g^{(2)}(s^*)\right]^{-1/2}}{\left[g^{(k)}(s^*)\right]^{-1/k}}\right|^{2k} = \left(\frac{T_c s^*}{\beta}\right)^{2-k} \frac{1}{\beta^{k-2}(\beta-1)^{k-2}} \left(\frac{\Gamma(\beta-1)}{\Gamma(\beta+1-k)}\right)^2, \tag{319}$$

motivating the use of Laplace's method when $T_c s^* \gg \beta$ and $\beta - 1 \gg 1$. Laplace's method yields

$$\int_0^\infty \rho(s) e^{T_c s} ds \approx \sqrt{2\pi\Delta^2}\rho(s^*) e^{T_c s^*} = \frac{e^{(\beta-1)\left(\frac{T_c\chi}{\beta}\right)^{\frac{\beta}{\beta-1}}}}{\Gamma(1+\beta^{-1})} \sqrt{\frac{2\pi}{\beta(\beta-1)} \left(\frac{T_c\chi}{\beta}\right)^{\frac{2-\beta}{\beta-1}}}, \tag{320}$$

and more generally,

$$\int_0^\infty \rho(s) s^p e^{T_c s} ds \approx \sqrt{2\pi\Delta^2}\rho(s^*)(s^*)^p e^{T_c s^*}. \tag{321}$$

Under this application of Laplace's method, $\langle s_f \rangle$ is approximated to precisely s^*. Further, this application of Laplace's method essentially implies a *single-s equivalence* of the type discussed in Good et al. (2012), whereby the dynamics arising from the full $\mu(s)$ can be approximated by the dynamics arising within a population subject to mutations which confer only a single effect $s_{eff} = s^*$, occurring at rate $U_{eff} = \sqrt{2\pi} U \rho(s^*)\Delta$.

The magnitude of s^* can be compared to b, with the result

$$\left(\frac{s^*}{b}\right)^3 \approx \sqrt{\frac{\beta(\beta-1)}{2\pi}} \frac{1}{\Gamma(1+\beta^{-1})} \frac{\chi}{U} \left(\frac{a\chi}{\beta}\right)^{\frac{\beta}{2(\beta-1)}} e^{-(\beta-1)\left(\frac{T_c\chi}{\beta}\right)^{\frac{\beta}{\beta-1}}} \tag{322}$$

From Eq. (322) it follows that s^*/b decreases with T_c (and therefore with N) at large T_c (and therefore at large N). Together with the fact that $T_c b$ increases monotonically with N without bound, this suggests that the MSSM regime is approached in the $N \to \infty$ limit. In contrast with the case of exponentially-distributed beneficial fitness effects ($\beta = 1$), no requirement needs to be made that $\langle s \rangle \ll U$.

Appendix H SIMULATION METHODS

In this Subsection, we provide details on the implementation of our Wright-Fisher simulations. In conducting individual-based simulations, we track the absolute fitness of, as well as the mutations carried by, each individual over the duration of the simulation. The population is initialized as a clonal population with initial absolute fitness 0. Each generation consists of a mutation step and and a stochastic birth/death step. In the mutation step, each individual acquires a Poisson-distributed number of mutations (with mean U), with the effect sizes of those mutations drawn from $\rho(s)$ and incrementing the log-fitness of an individual by s. Individuals are also subject to completely neutral mutations at rate $U_n = 0.1$ (per-individual, per-generation). In the stochastic birth/death step, N individuals are sampled with replacement from the population, with an individual's sampling probability proportional to its absolute exponential (as opposed to log) fitness.

Our simulations implement an "infinite-sites" model of mutation, in which each new mutation is assigned a unique identity upon its occurrence. Each time a mutation occurs, we record its fitness effect. When a mutation is shared by all individuals in the population, we mark the mutation as fixed. This enables us to empirically measure the distribution of fixed fitness effects, which can be compared to our theoretical predictions $\rho_f(s)$. Additionally, we measure the time elapsed before the *first* fixation of a neutral mutation, which we take as the length T_{eq} of an epoch. From that point on, we record the state of the population each epoch until the conclusion of the simulation. In particular, at each epoch, we record the mean fitness of the population (and in some cases a histogram of relative fitnesses present in the population), the mutations which have fixed (and their fitness effects), as well as a measurement of the population's site frequency spectra, both of neutral mutations and of selected mutations. From measured site frequency spectra we obtain measurements of the pairwise heterozygosity, both of neutral mutations and of selected mutations. We discard the first 10 recorded states of the population, which may still reflect the transient behavior of the population after its initialization as a clonal population. The code implementing our simulations is available at https://github.com/mjmel/mssm-sim.

Appendix I NUMERICAL METHODS

To obtain theoretical predictions for comparison with simulation results, we numerically solve the system

$$1 - T_c x_c \left[1 - \frac{1}{T_c b} \frac{\mathrm{Ai}'\left(\frac{c-x_c}{b}\right)}{\mathrm{Ai}\left(\frac{c-x_c}{b}\right)} \right] = 0 \tag{323}$$

$$\left[\frac{\left(\mathrm{Ai}'\left(\frac{c-x_c}{b}\right)\right)^2 - \left(\frac{c-x_c}{b}\right) \mathrm{Ai}^2\left(\frac{c-x_c}{b}\right)}{\mathrm{Ai}\left(\frac{c-x_c}{b}\right)} \right] N x_c - e^{(T_c b)^3/3 + |z_0| T_c b - 1} = 0 \tag{324}$$

within the infinitesimal approximation; in the MSSM approximation, we solve Eq. (323) along with

$$\left[\frac{\left(\mathrm{Ai}'\left(\frac{c-x_c}{b}\right)\right)^2 - \left(\frac{c-x_c}{b}\right) \mathrm{Ai}^2\left(\frac{c-x_c}{b}\right)}{\mathrm{Ai}\left(\frac{c-x_c}{b}\right)} \right] N x_c - e^{T_c x_c - \frac{v T_c^2}{2} + \int \frac{\mu(s) ds}{s} \left(e^{T_c s} - T_c s - 1\right)} = 0. \tag{325}$$

In doing so, the definitions for T_c, b and c differ between the two approaches; in the infinitesimal approximation,

$$T_c \equiv \frac{\sigma^2}{2D} \tag{326}$$

$$b \equiv D^{1/3} \tag{327}$$

$$c \equiv \frac{\sigma^4}{4D}, \tag{328}$$

with σ^2 and $D^{1/3}$ defined by $v \equiv \sigma^2 + U\langle s\rangle$ and $2D \equiv U\left\langle s^2\right\rangle$, respectively. In contrast, within the MSSM approximation,

$$v \equiv \int \mu(s) s e^{T_c s} ds \tag{329}$$

$$2b^3 \equiv \int \mu(s) s^2 e^{T_c s} ds \tag{330}$$

$$c \equiv \int \mu(s) \left(T_c s e^{T_c s} + 1 - e^{T_c s}\right) ds. \tag{331}$$

The parameters N and $\mu(s)$ can be thought to be the "known" inputs to a system of two equations with the two unknowns T_c and x_c.

Numerical solution is made more straightforward by enforcing upon the solution that $T_c > 0$ (to ensure that $v > \int \mu(s) s ds$) and $T_c < N$ (given the interpretation of T_c as a coalescence timescale, necessarily shorter than the timescale N of neutral pairwise coalescence). Additionally, we enforce that $x_c > 0$ (since otherwise, our approach would neglect the fitness density of all individuals with greater than average fitness) and $x_c < c - bz_0$ (since otherwise, our approach would predict negative values of fitness density for individuals with relative fitness exceeding $c - bz_0$).

To solve the relevant systems, an array of test T_c values, logarithmically spaced between 1 and N, are considered. For each T_c value, the corresponding x_c solving Eq. (323) is identified. This is done by evaluating the left-hand side of Eq. (323) for linearly spaced x_c between 0 and $c - bz_0$, identifying zero crossings and employing the scipy function *brentq*. In cases where multiple solutions x_c might be obtained, the solution closest to $c - bz_0$ is taken. This permits evaluation of the left-hand side of Eq. (324) or Eq. (325) for a particular T_c. Once this is done for the entire array of test T_c values, zero crossings are identified and the scipy function *brentq* is employed. In cases where no zero crossings are identified, the scipy function *fsolve* is employed. The initial guess is taken as the test T_c value which minimizes the absolute value of the left-hand side of Eq. (324) or Eq. (325). In cases where multiple zero crossings are identified, the scipy function *brentq* is employed for each zero crossing, and the solution with the largest value $T_c b$ is taken.

In practice, our simulations are carried out for parameter combinations lying on a grid with specified $T_c b$ and $T_c\langle s\rangle$ (or, equivalently, $T_c\langle s_f\rangle$) values. We thus invert the above equations to solve for a corresponding N and $\mu(s)$, given $T_c b$ and $T_c\langle s\rangle$. A given combination of $T_c b$ and $T_c\langle s\rangle$ values is compatible with a range of N values (with N, $T_c b$ and $T_c\langle s\rangle$ fully determining $\mu(s)$; note that properly scaled dynamical quantities depend on N, U and $\langle s\rangle$ only via the scaled quantities NU and $N\langle s\rangle$.) We thus choose a value of N that ensures feasibility of our individual-based simulations.

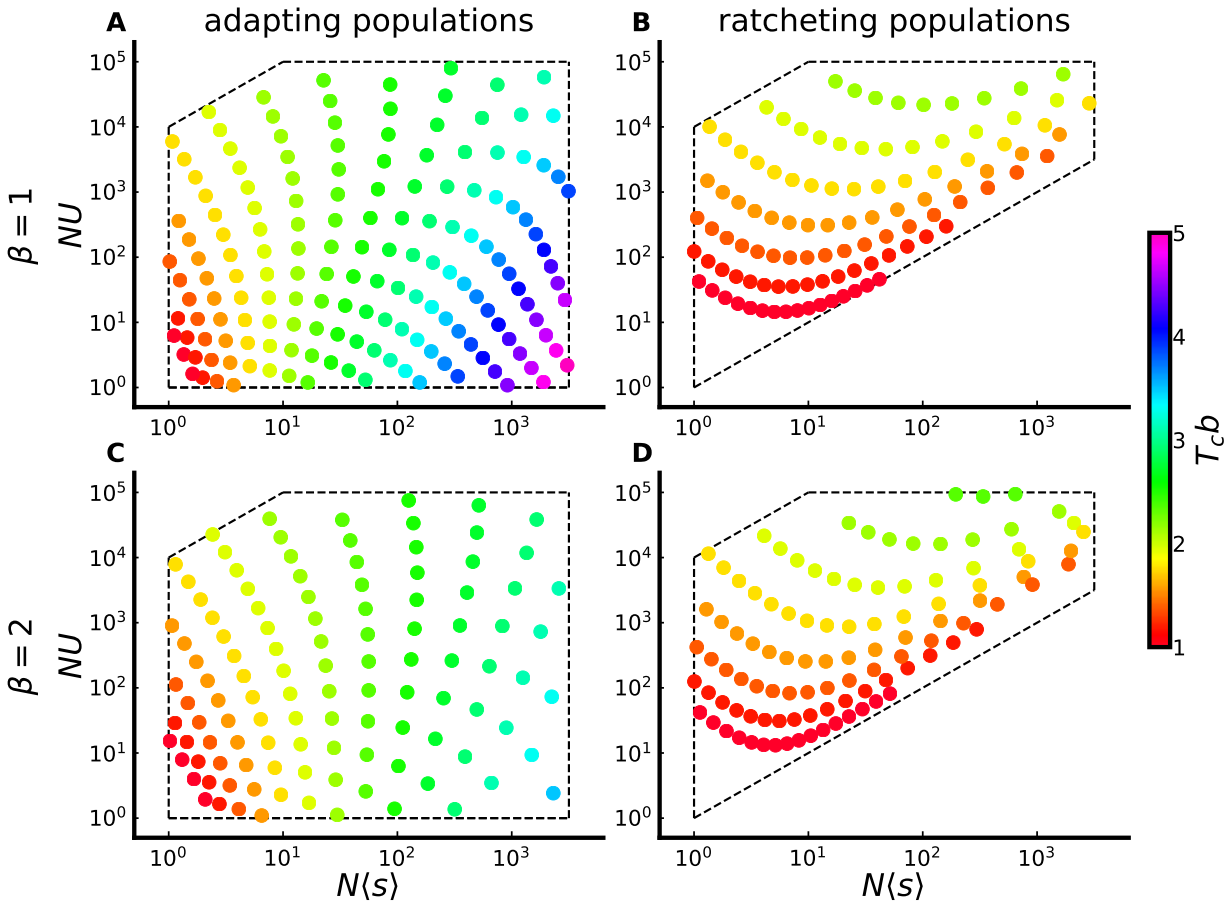

FIG. S1 Simulated parameter combinations with an exponential or $\beta = 2$ stretched exponential DFE, plotted as a function of NU and $N\langle s\rangle$, and colored by values of $T_c b$. Dashed lines represent constraints imposed on a grid of parameter combinations with linearly spaced $T_c b$ values, and logarithmically spaced $T_c\langle s_f\rangle$ values (for adapting populations) or $T_c\langle s\rangle$ values (for ratcheting populations). These constraints are chosen to limit attention to the MSSM regime (and the region of parameter space just beyond its regime of validity) and to ensure feasibility of individual-based simulations.

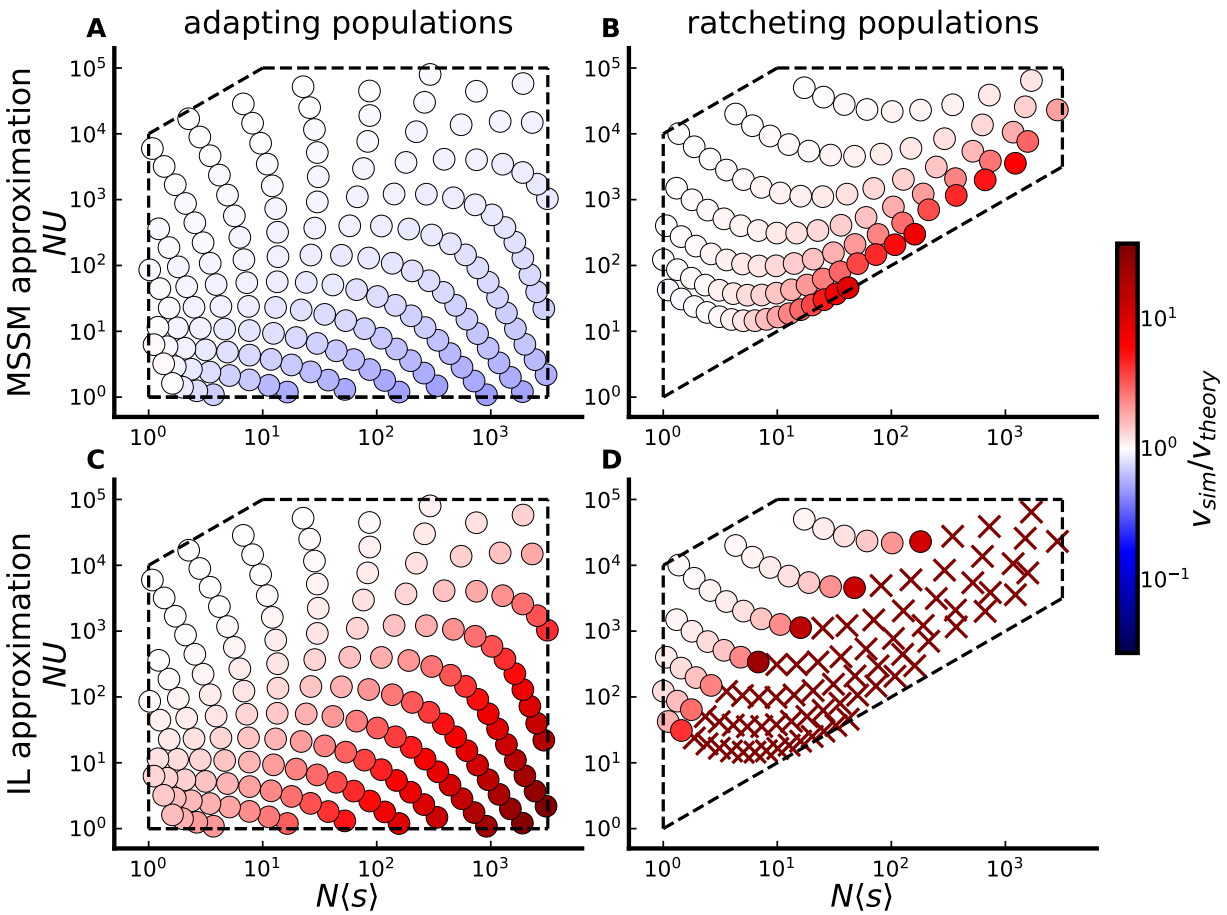

FIG. S2 Simulated parameter combinations with an exponential DFE, plotted as a function of NU and $N\langle s\rangle$, and colored by values of $v_{\text{sim}}/v_{\text{theory}}$. Dashed lines represent constraints imposed on a grid of parameter combinations with linearly spaced $T_c b$ values, and logarithmically spaced $T_c\langle s_f\rangle$ values (for adapting populations) or $T_c\langle s\rangle$ values (for ratcheting populations).

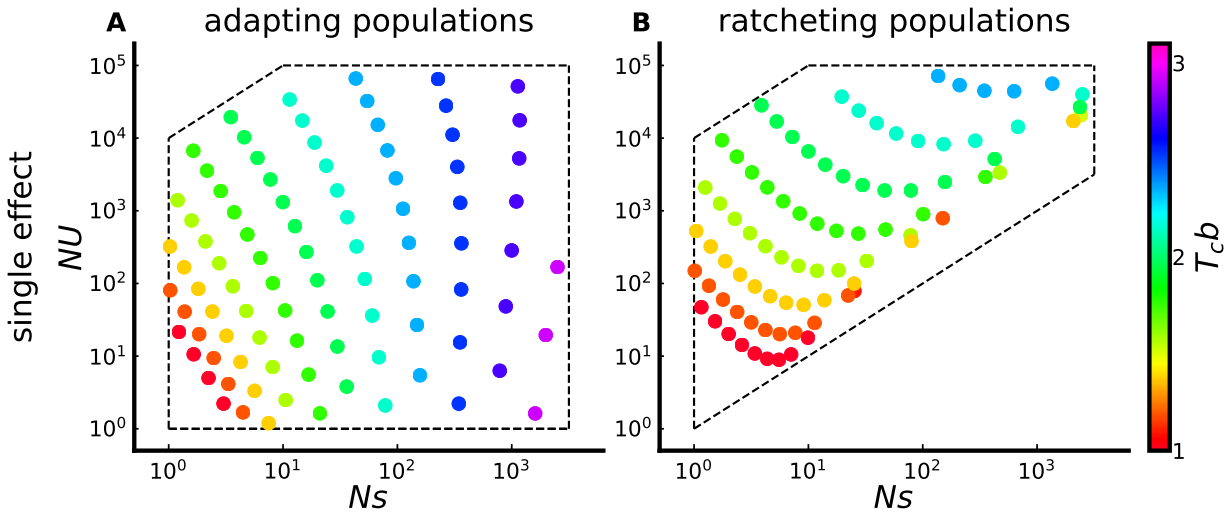

FIG. S3 Simulated parameter combinations with a single-effect DFE, plotted as a function of NU and Ns, and colored by values of $T_c b$. As in Fig. S1, dashed lines represent additional constraints imposed on the grid of parameter combinations.

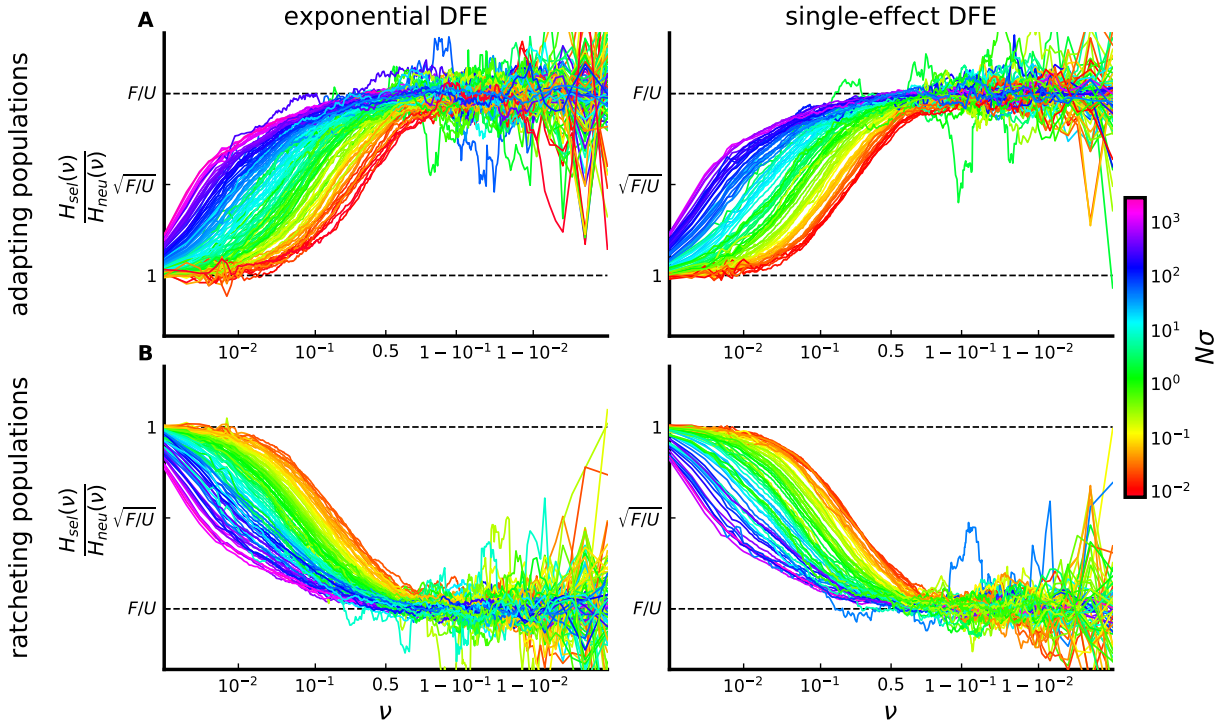

FIG. S4 Color plot of SFS ratio curves for adapting populations (A) and ratcheting populations (B). SFS ratio curves for the entire range of simulated $T_c \langle s \rangle$ values are plotted with the same scaling of the vertical axis as in Fig. 5, colored by $N\sigma$ values. Displayed curves lie on the constrained grids for exponential DFEs (left) or single-effect DFEs (right), with the additional restriction that $\langle s_f \rangle < 3b$.

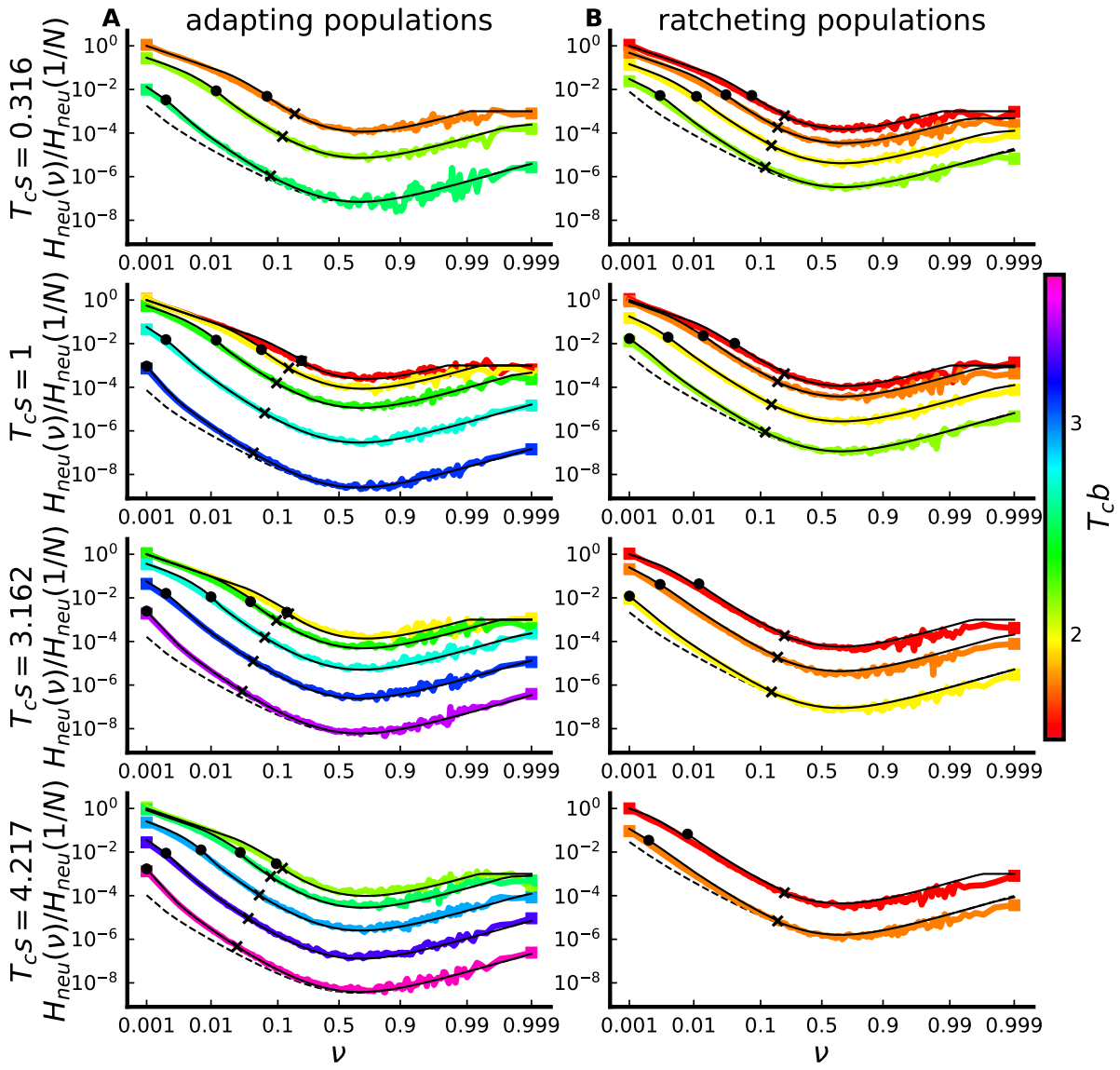

FIG. S5 Comparison between predicted and simulated neutral SFSs, for populations subject to an exponential DFE. Solid black lines denote theory predictions; dashed black lines denote predictions of the BSC SFS for one parameter combination in each panel. The left side corresponds to adapting populations, while the right column corresponds to ratcheting populations. The $T_c s$ values denoted on the left-hand side are the values of $T_c \langle s_f \rangle$ for adapting populations, and the values of $T_c \langle s \rangle$ for ratcheting populations. For computational reasons, SFSs are observed with a sample size of 1000 individuals, in some cases smaller than the population size N. The expected SFS depends on sample size (in addition to the frequency ν); theory curves are appropriately downsampled versions of the piecewise-defined function given in Eq. (75). Parameter combinations lie on the same grid used for comparison between simulation and theory in the main text.

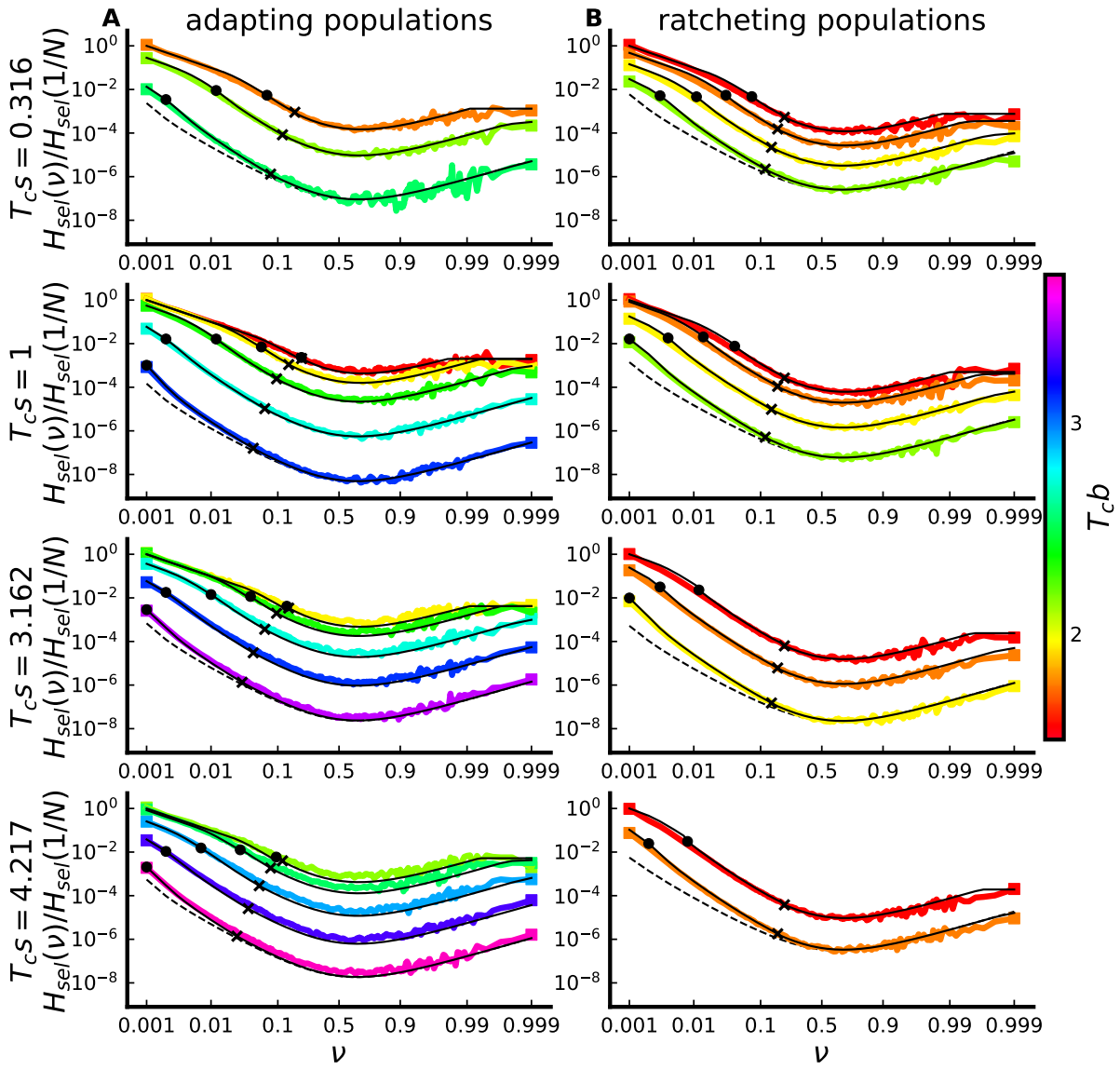

FIG. S6 Comparison between predicted and simulated selected SFSs, for populations subject to an exponential DFE. Solid black lines denote theory predictions; dashed black lines denote predictions of the BSC SFS for one parameter combination in each panel. As in Fig. S5, the $T_c s$ values denoted on the left-hand side are the values of $T_c \langle s_f \rangle$ for adapting populations, and the values of $T_c \langle s \rangle$ for ratcheting populations.

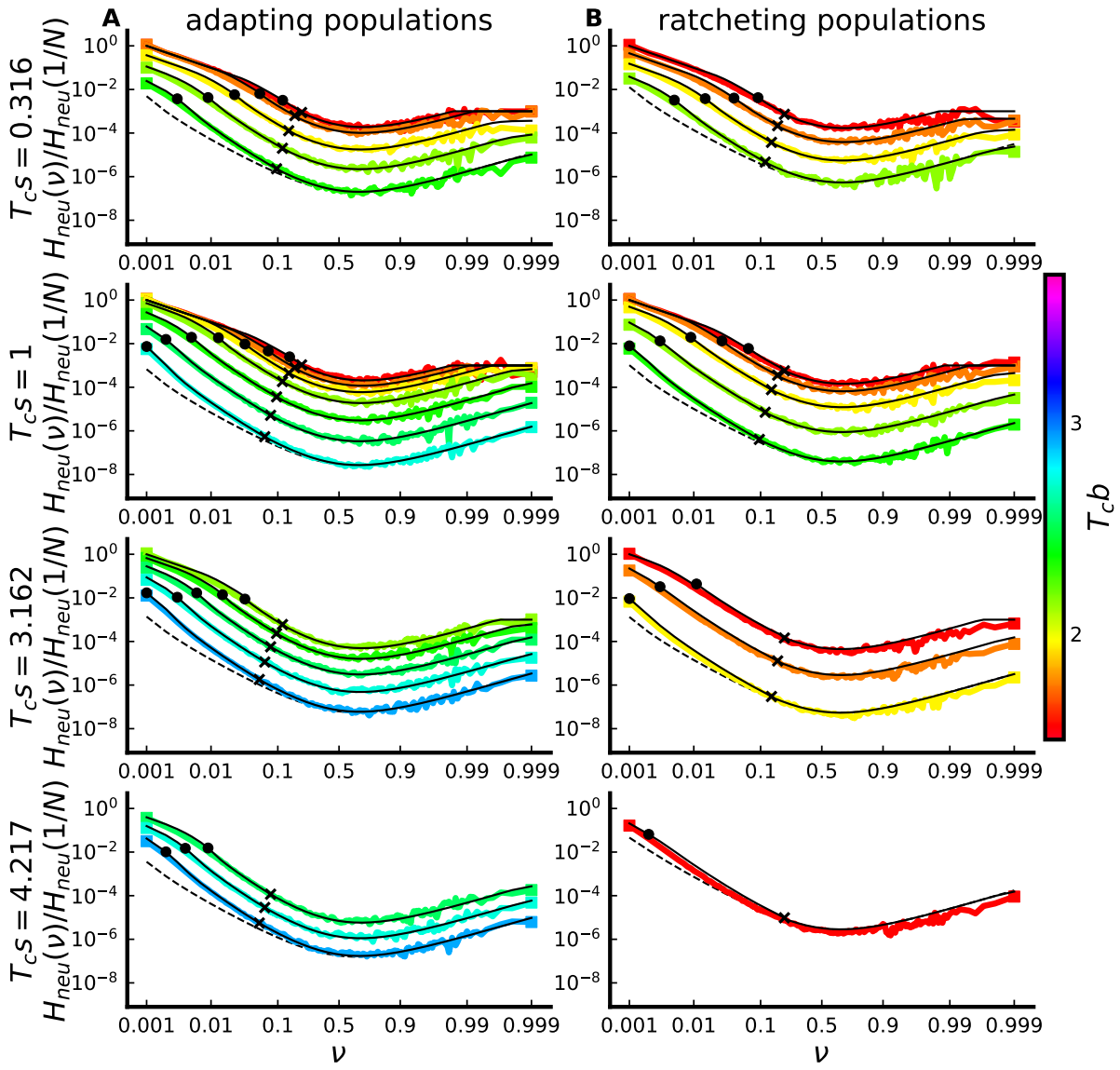

FIG. S7 Comparison between predicted and simulated neutral SFSs, for populations subject to an single-effect DFE. Solid black lines denote theory predictions; dashed black lines denote predictions of the BSC SFS for one parameter combination in each panel. As in Fig. S5, the $T_c s$ values denoted on the left-hand side are the values of $T_c \langle s_f \rangle$ for adapting populations, and the values of $T_c \langle s \rangle$ for ratcheting populations.

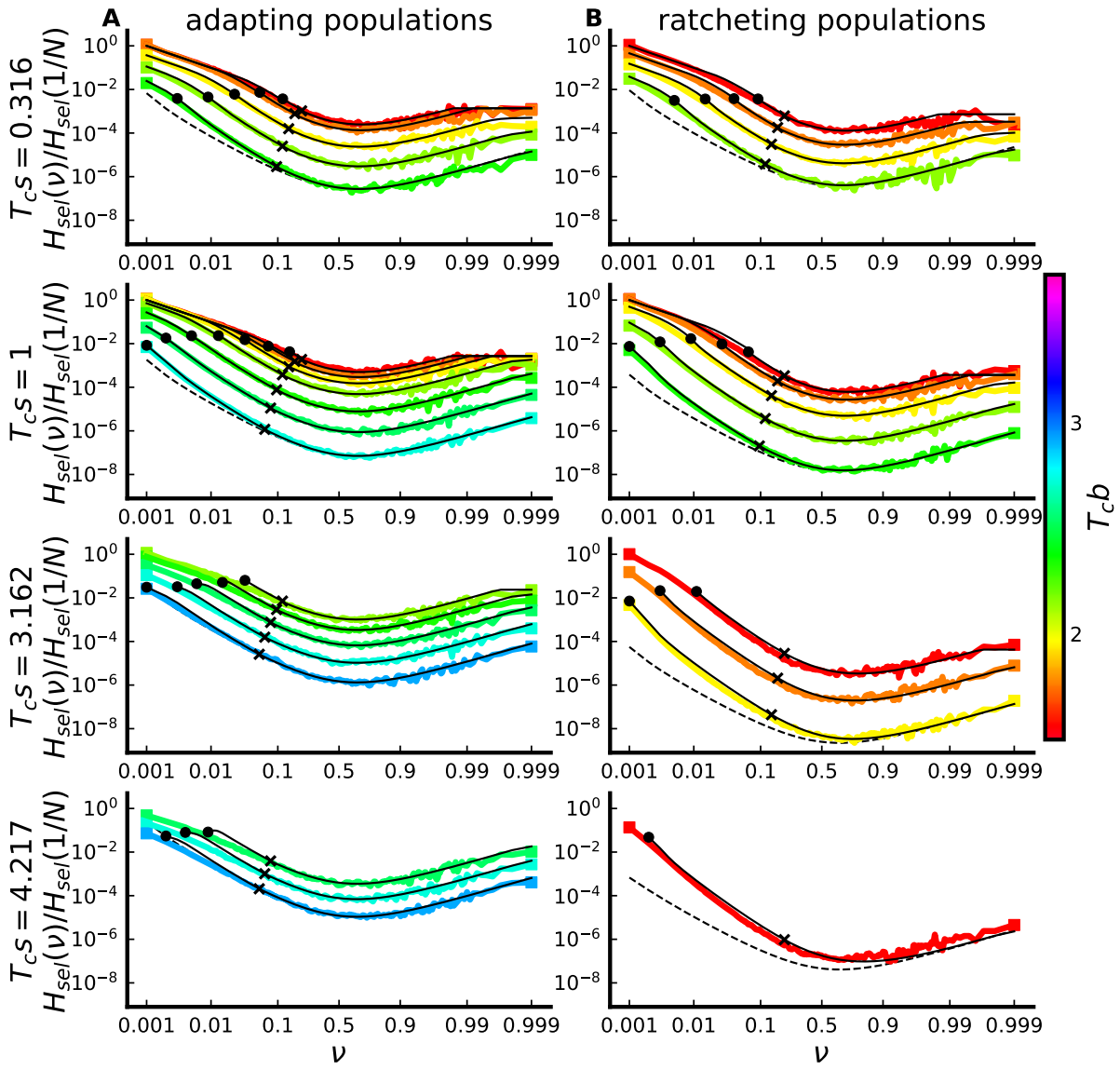

FIG. S8 Comparison between predicted and simulated selected SFSs, for populations subject to an single-effect DFE. Solid black lines denote theory predictions; dashed black lines denote predictions of the BSC SFS for one parameter combination in each panel. Theory predictions break down for $\nu < 2/(N\sigma)$ when $s > b$; for clarity, we have omitted these predictions. As in Fig. S5, the $T_c s$ values denoted on the left-hand side are the values of $T_c \langle s_f \rangle$ for adapting populations, and the values of $T_c \langle s \rangle$ for ratcheting populations.

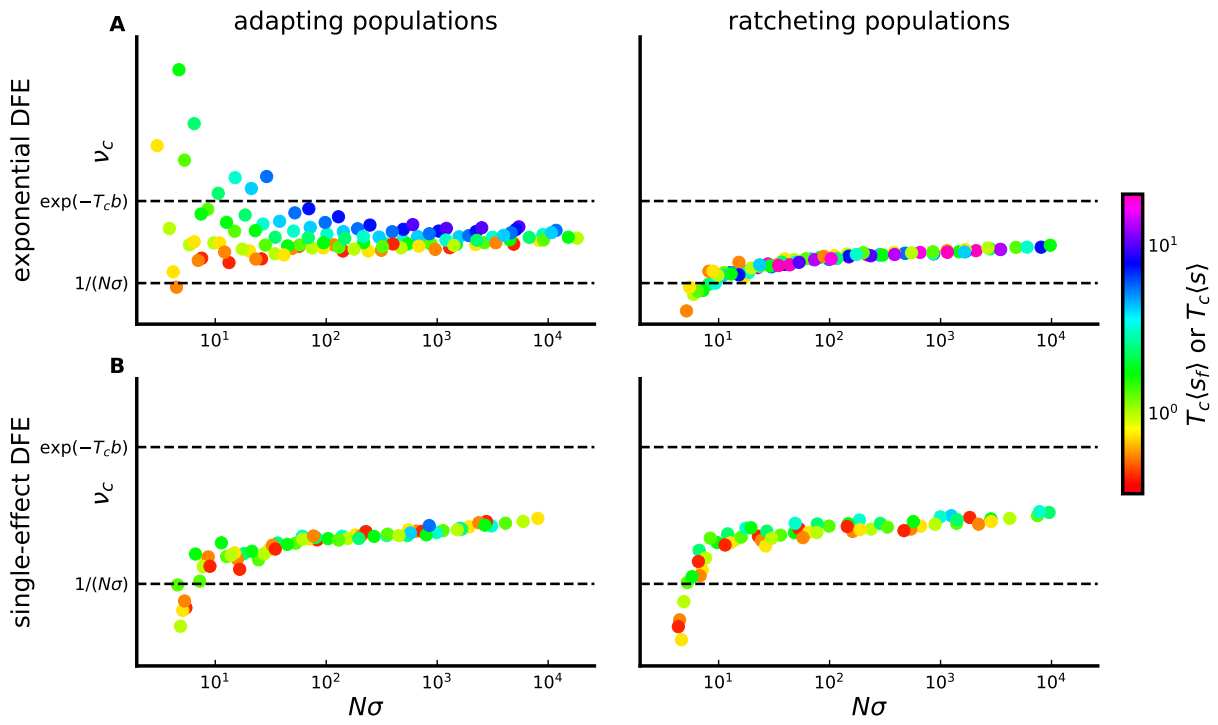

FIG. S9 Bounds on the crossover frequency ν_c. The location of a point along the vertical axis denotes the crossover frequency ν_c of that simulated population, extracted from its SFS ratio curve using spline interpolation. The vertical axis is on a logarithmic scale, scaled such that dashed lines correspond to $\nu_c = 1/(N\sigma)$ and $\nu_c = e^{-T_c b}$. Note that $T_c b$ differs even among simulated populations with the same value of $N\sigma$. For large $N\sigma$, we can see that $1/(N\sigma) < \nu_c < e^{T_c b}$. Points are colored according to their values of $T_c \langle s_f \rangle$ (for adapting populations, in A) or $T_c \langle s \rangle$ (for ratcheting populations, in B). Simulated populations are identical to those considered in Fig. 6.

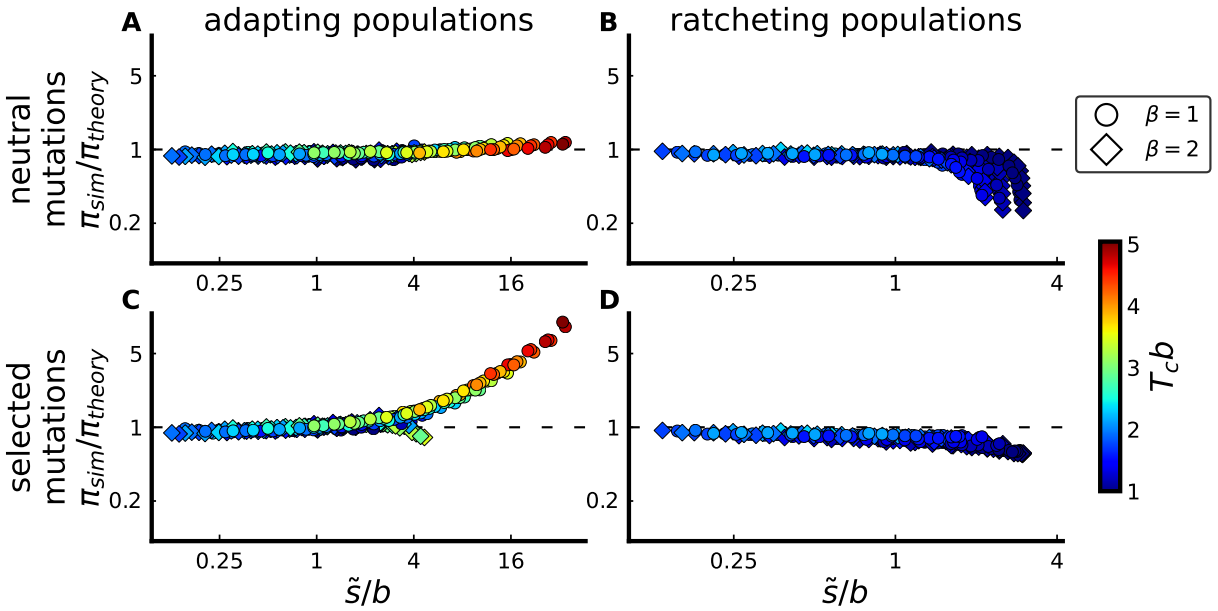

FIG. S10 Comparison between simulated and predicted heterozygosity π of neutral mutations (A and B), and between simulated and predicted heterozygosity π of selected mutations (C and D), for adapting populations (A and C) and ratcheting populations (B and D). Parameters are identical to those simulated in Fig. 4. Predictions are obtained by integrating over the piecewise-defined approximation to the SFS given in the main text.