OXFORD GENETICS

GENETICS, 2022, 221(4), iyac053
https://doi.org/10.1093/genetics/iyac053
Advance Access Publication Date: 7 April 2022
Investigation
Highlighted Article

Population genetics of polymorphism and divergence in rapidly evolving populations

Matthew J. Melissa [1,2,3,4] Benjamin H. Good [5,6] Daniel S. Fisher,[5,6] Michael M. Desai [1,2,3,4,*]

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA,
[2]Department of Physics, Harvard University, Cambridge, MA 02138, USA,
[3]Quantitative Biology Initiative, Harvard University, Cambridge, MA 02138, USA,
[4]NSF-Simons Center for Mathematical and Statistical Analysis of Biology, Harvard University, Cambridge, MA 02138, USA,
[5]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA,
[6]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

*Corresponding author: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. Email: mdesai@oeb.harvard.edu

Abstract

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.

Keywords: clonal interference; evolutionary dynamics; fixation probability; traveling waves

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, $p_{\mathrm{fix}}(s)$, 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 (the SFS, 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 independently of one another (i.e. by ignoring the physical *linkage* of mutations within the same genomic segment). The dynamics are then relatively

Received: June 28, 2021. **Accepted:** March 19, 2022

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 (Neher *et al.* 2013; Good *et al.* 2014; 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 (Miralles *et al.* 1999; Strelkowa and Lässig, 2012; Lieberman *et al.* 2014; Nourmohammad *et al.* 2019). 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 (Gerrish and Lenski, 1998; Kao and Sherlock, 2008; Lang *et al.*, 2013; Buskirk *et al.*, 2017).

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 variation 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) (Tsimring *et al.* 1996; Rouzine *et al.* 2003, 2008). 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 $p_{\mathrm{fix}}(s)$ 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 (Tsimring *et al.* 1996; Cohen *et al.* 2005; Hallatschek 2011; Neher and Hallatschek 2013) 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 $U\langle s^2 \rangle$ 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 *et al.* 2014; Good and Desai 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 *Analysis*). 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 *Analysis*. 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; Good *et al.* 2012; Desai *et al.* 2013; Fisher 2013; 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_cs and T_cU can be probed more directly. While dynamical interpretations of T_cs and T_cU 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_cs from observed levels of polymorphism and divergence among populations typically make the assumption that different loci evolve *independently* of one another (Sawyer and Hartl 1992; Bustamante *et al.* 2001). For example, a classic result shows that in a population with a constant size N, the ratio between selected and neutral site frequency spectra is given by

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

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

where $T_c = N$ is the coalescent timescale (Sawyer and Hartl 1992). Using this result, the magnitudes of the scaled selection coefficients (T_cs) 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 ($1 - \nu_c \sim 1/T_cs$) for beneficial mutations.

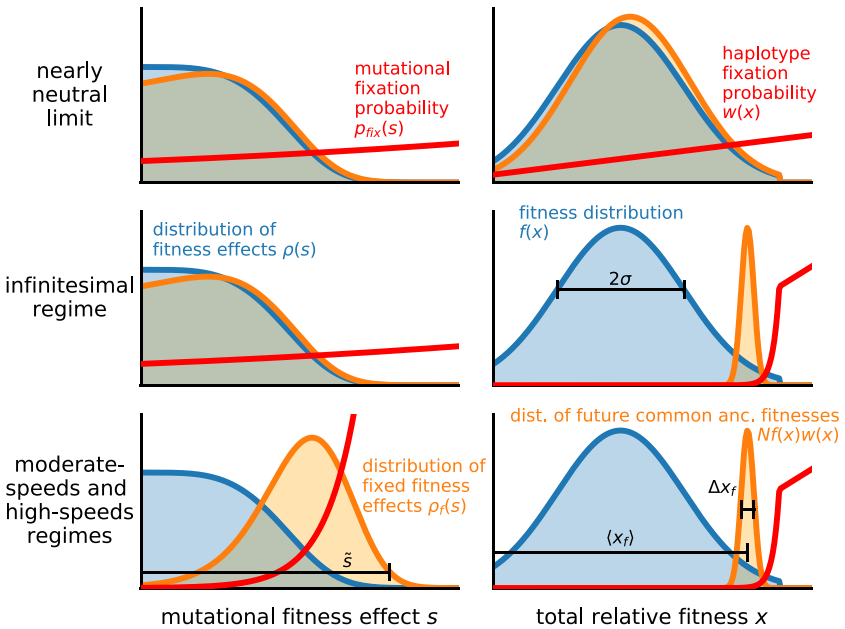

Fig. 1. Schematic depiction of different regimes. (Left) Schematic plots of the mutation fixation probability $p_{fix}(s)$, an example DFEs $\rho(s)$, and its corresponding distribution of fixed fitness effects $\rho_f(s)$, for regimes of the parameter space indicated by row labels. In the nearly neutral 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 *Analysis*.

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 mutations, 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. Sawyer and Hartl 1992; Hartl *et al.* 1994) 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 Fisher-Kolmogorov-Petrovsky-Piscunov (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 *Model*, we describe the model of the population dynamics and briefly summarize our simulation methods. In *The Traveling-Wave Approach*, we review previous approaches to model evolutionary dynamics using traveling wave theory. We begin *Analysis* 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 distribution of fitness effects (DFEs), and compare our analytical results to the results of simulations. In *Statistics of Genetic Diversity*, 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 *Key Fitness Scales and Timescales: Simple Heuristics*, 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.

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 $x \equiv X - \overline{X}$, whereby $\overline{X}$ we denote the population-wide mean fitness. We write down the particular stochastic birth/death process we carry out in simulations in *Simulation Methods*; in our analytical treatment, we assume a slightly different birth/death process, which we write down in *The Traveling-Wave Approach*. We assume the fitness effect of each new mutation is drawn at random from a time-independent DFEs, $\rho(s)$. 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 (1) all mutations are beneficial, (2) all mutations are deleterious, and (3) 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 affect various aspects of the evolutionary dynamics. Furthermore, 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).

Simulation methods

To test our analytical predictions, we conduct individual-based Wright–Fisher simulations. These simulations, which we describe in more detail in Supplementary 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 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 Equation (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 Equation (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 Equation (3) falls off more slowly than exponentially with large s (though for reasons we describe below, we do not consider this case extensively here).

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, $p_{\mathrm{fix}}(s)$, 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),

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

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

The last term in Equation (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 $d\overline{X}/dt$ 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 $v = \langle d\overline{X}/dt \rangle$; 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

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

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

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

The solution to Equation (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, Equation (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 (Neher *et al.* 2010; Good *et al.* 2012; Fisher 2013; Good and Desai 2014). Each of these lineages founds its own stochastic fitness wave, $g(x, t)$, whose dynamics are described by a related differential equation,

$$-\partial_t \phi = x\phi - \phi^2 + \int \mu(s)[\phi(x+s) - \phi(x)] \tag{9}$$

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

$$\begin{aligned} v\partial_x w(x) &\approx x w(x) \\ &+ \int \mu(s)[w(x+s) - w(x)] ds - w^2(x). \end{aligned} \tag{10}$$

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 $w(x)$, Equation (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,

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

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

A central complication is that Equation (10) is nonlinear. Furthermore, the nonlinear w^2 term is the only term in Equation (10), which reflects the stochasticity of births and deaths. As a result, dropping the w^2 term in Equation (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 (Neher *et al.* 2010; Good *et al.* 2012; Fisher 2013; Good and Desai 2014). 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 Equation (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 Equation (10) and solve the linear equation

$$\begin{aligned} v\partial_x w(x) &= x w(x) \\ &+ \int \mu(s)[w(x+s) - w(x)] ds. \end{aligned} \tag{12}$$

An approximation where the large-x and small-x solutions are simply *patched* together is then widely valid (Good *et al.* 2012; Fisher 2013; Good and Desai 2014). 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 Equation (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-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 Equation (5) leads to negative values for $f(x)$ at sufficiently large x (Tsimring *et al.* 1996; Rouzine *et al.* 2008; Fisher 2013). 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 $x > x_{cut}$ (Tsimring *et al.* 1996; Rouzine *et al.* 2008, 2003). 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 that 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 Equations (5) and (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

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

and

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

in Equations (5) and (12), respectively. The resulting $f(x)$ and $w(x)$ can be obtained using the Airy equation (Tsimring *et al.* 1996; Cohen *et al.* 2005; Hallatschek 2011). 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 Equations (5) and (12) by a set of local equations, used by Fisher (2013) to study the "high-speeds" regime. In *Analysis*, 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 an 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 BSC, and analytically describe departures from the BSC that are apparent in the low-frequency portion of the site frequency spectrum. Finally, we conclude *Analysis* 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 1.

Analysis

As noted above, our approach is based on a key approximation employed by Fisher (2013) to study the "high-speeds" regime. The key idea is that we can approximate $f(x-s)$ and $w(x+s)$ in Equations (5) and (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; Equation (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), Equation (17) implies that $T_c s_b < 1$. We will later show in *Statistics of Genetic Diversity* that, within the MSSM regime, the quantity T_c as defined in Equation (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.

Table 1. Notation used.

	Symbol	Quantity
Input parameters	N	Population size
	U	Mutation rate
	$\rho(s)$	DFEs
	$\mu(s) \equiv U\rho(s)$	Mutational fitness spectrum
Traveling wave quantities	$x \equiv X - \overline{X}$	Relative fitness
	$f(x)$	Distribution of relative fitnesses
	$w(x)$	Haplotype fixation probability
	$v \equiv d\overline{X}/dt$	Rate of change of meanfitness
	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)

However, we emphasize that T_c is a derived quantity, defined by Equation (17), which we will relate to the underlying parameters N and $\mu(s)$ in *The Relation Between T_c and N*. [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 Equations (15) and (16) into Equations (5) and (10), we find that for $x < x_c$, $\psi(x)$ and $\xi(x)$ satisfy

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

and

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

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

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

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

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

and

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

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

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

In Table 2, 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 2 (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 Equation (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 *The Traveling-Wave Approach*, 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 Supplementary 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 Equations (25) and (26), as well as x_c in Equation (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, $\sigma^2 \equiv v - \int \mu(s)sds$ corresponds to the population-wide fitness variance, and $D \equiv \frac{1}{2}\int \mu(s)s^2ds$ corresponds to a mutational

Table 2. Key definitions.

Quantity	Definition	Interpretation
T_c	$v \equiv \int \mu(s)e^{T_c s}sds$	Log-slope of $f(x)$ near $x = c$. Corresponds with sweep timescale and coalescence timescale $\langle T_2\rangle/2$.
b	$2b^3 \equiv \int \mu(s)s^2e^{T_c s}ds$	Width in distribution of fixed relative fitnesses $\sim\Delta x_f$.
c	$c \equiv \int \mu(s)[T_c se^{T_c s} + 1 - e^{T_c s}]ds$	Average background fitness of fixed individuals $\sim\langle x_f\rangle$.
D	$2D \equiv \int \mu(s)s^2ds$	Effective diffusion constant in infinitesimal regime.
$\langle s_f\rangle$	$\langle s_f\rangle \equiv \int \rho_f(s)sds$	Average fixed fitness effect.
Δs_f	$(\Delta s_f)^2 \equiv \int \rho_f(s)(s - \langle s_f\rangle)^2ds$	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 $w(x)$.

diffusion constant (an interpretation of which we provide in *Key Fitness Scales and Timescales: Simple Heuristics*). 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.

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 Equation (11), using the approximate $f(x)$ and $w(x)$ in Equations (25) and (26), respectively. We emphasize that the expressions in Equations (25) and (26) are only *local* approximations, valid within some range of x. In the next subsection, we will obtain and discuss conditions which ensure that Equations (25) and (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 Equations (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 Equation (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 Equation (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 Supplementary 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 between T_c and N, $\mu(s)$ and x_c,

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

(Supplementary Appendix B). By combining Equations (27) and (31), along with the definitions for T_c, b and c in Equations (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 Equations (27) and (31) for a given choice of N and $\mu(s)$ (Supplementary Appendix I). Although we discuss analytical approximations to this solution in *Specific Example DFEs*, we use this numerical solution for comparison with simulations throughout the rest of this paper. We note that Equation (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 Equation (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

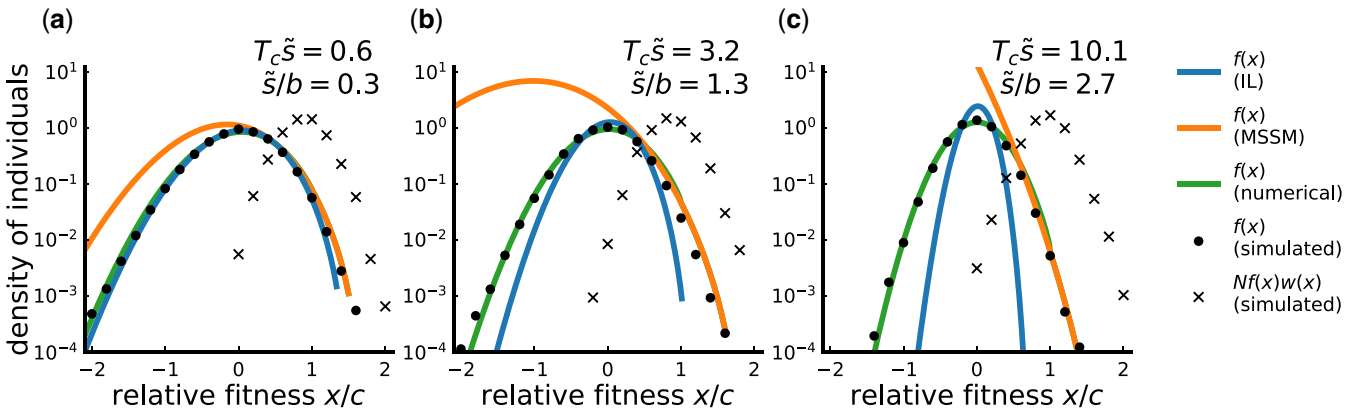

Fig. 2. Averaged distributions of relative fitnesses, $f(x)$, for simulated populations subject to beneficial mutations with an exponential DFE, with values of $T_c\tilde{s}$ and $\tilde{s}/b$ denoted above. A precise definition of $\tilde{s}$ is given in *Conditions of Validity: a "Moderate Selection, Strong Mutation Regime"*. 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 Supplementary 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 $\int f(x)w(x)dx$.

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 Equation (30) matches simulation results well in the region dominating $\int f(x)w(x)dx$. Outside, this region, however, the prediction in Equation (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 Supplementary Appendix I), as well as a numerical saddle point approximation we present in Supplementary 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 Equation (30) throughout the remainder of this article.

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

The approximate solutions $f(x)$ and $w(x)$ given in Equations (25) and (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 *Statistics of Genetic Diversity* 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 $\rho_f(s)$,

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

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

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

follows from Equation (32), and $\rho_f(s) \propto e^{T_c s}\rho(s)$ (Supplementary 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 Supplementary 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 Equations (25) and (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

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

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

$$s \ll b \tag{35}$$

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 Equation (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

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

These conditions have dynamical interpretations which we discuss in *Key Fitness Scales and Timescales: Simple Heuristics*. 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 an 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.]

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 $T_c s_b \ll 1$, Equation (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 (Neher and Hallatschek 2013; Good *et al.* 2014). We can see that Equation (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 $\log(N^3 U s_b^2)$.

In the opposite regime $T_c s_b \gg 1$, Equation (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) (Supplementary 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 Equations (42) and (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 $\mu(s) = U_e \delta(s - s_e)$ (Hegreness *et al.* 2006; Desai and Fisher 2007; Good *et al.* 2012; Fisher 2013). 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 (Supplementary Appendix G).

On the other hand, if $\rho(s)$ falls off slower than exponentially with large positive s, the integrals in Equation (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

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

(Supplementary 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 Equation (46) differ from those for a single beneficial fitness effect in Equation (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 (Tsimring *et al.* 1996; Cohen *et al.* 2005; Hallatschek 2011; Neher and Hallatschek 2013) 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 Equation (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 Equation (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 Equation (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 nonequilibrium 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 Equations (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 (Söderberg and Berg 2007; Good *et al.* 2014). We emphasize that the effective population size in Equation (47) does *not* denote a coalescence timescale, as is often implied in the literature. Indeed, the asymptotic expressions in Equations (40) and (37) show that T_c will typically be much less than N_e in the MSSM regime.

Together with Equation (47), the conditions of validity in Equation (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 Equation (37), with $N_e = Ne^{-U_d/s_d}$.

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 (Neher *et al.* 2013; Neher and Hallatschek 2013; Good *et al.* 2014), this scenario is well-described by the infinitesimal limit when $T_c s_d \ll 1$, with s_d replacing s_b in Equation (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; Good *et al.* 2014; Cvijović *et al.* 2018), which predicts that the large-N limit of the background selection model approaches a nearly neutral regime, with a coalescent timescale,

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

Interestingly, Equation (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 Equation (31). In 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.

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 Supplementary Appendix H; in Supplementary Fig. 1, 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 Supplementary Fig. 2, 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

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

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 Equation (33). For each of these quantities, we observe highly quantitative agreement

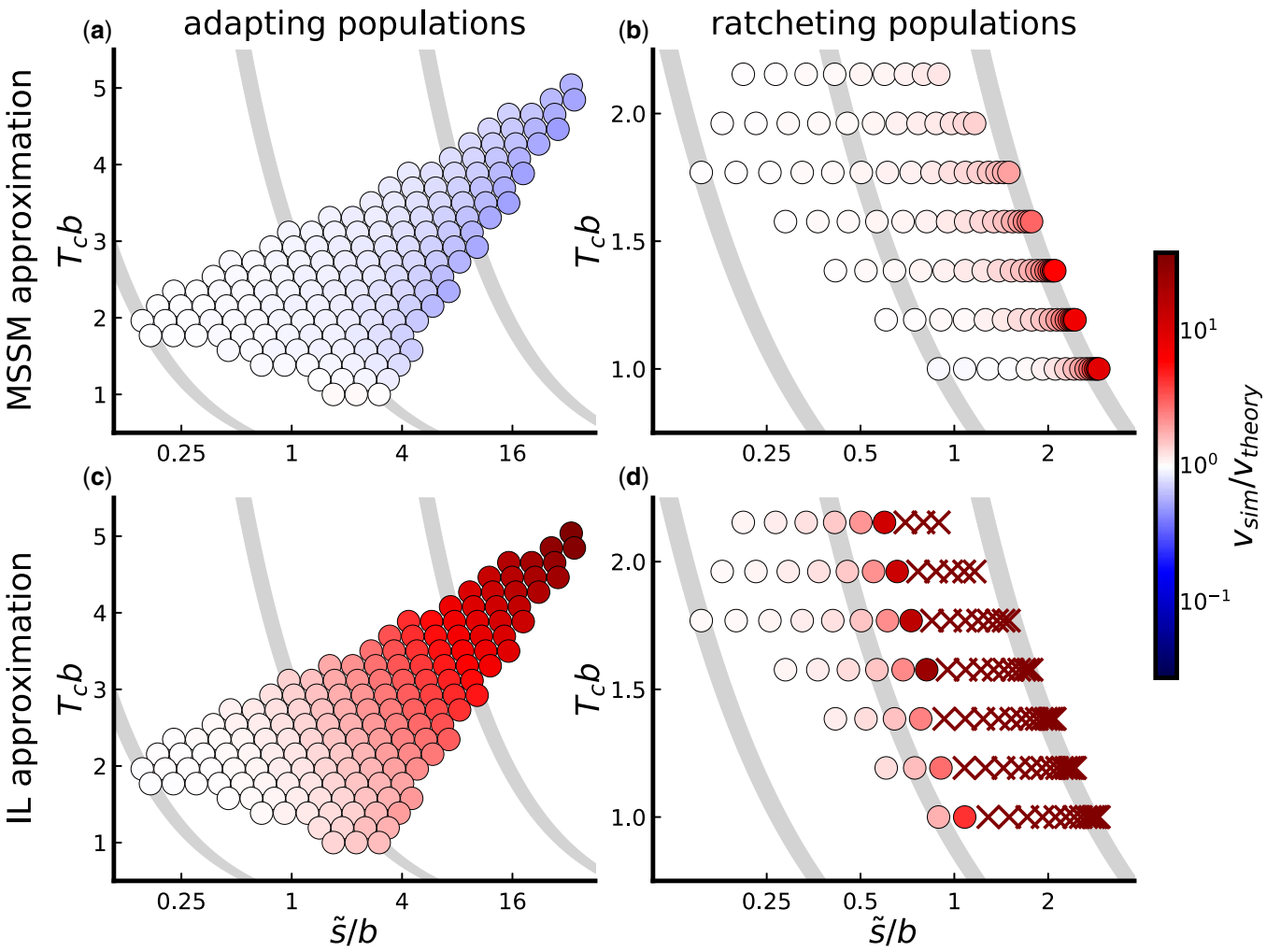

Fig. 3. Comparison of theoretical predictions for the rate of change in mean fitness, v, to the corresponding rates observed in simulations, for populations subject to an exponential DFE of beneficial mutations or deleterious mutations. The color of each point denotes the ratio of simulated to predicted 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 Supplementary Fig. 1. The "x" markers in D denote populations for which the infinitesimal approximation predicts a rate of fitness change of the incorrect sign.

between simulations and theory for small and moderate values of $\tilde{s}/b$.

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

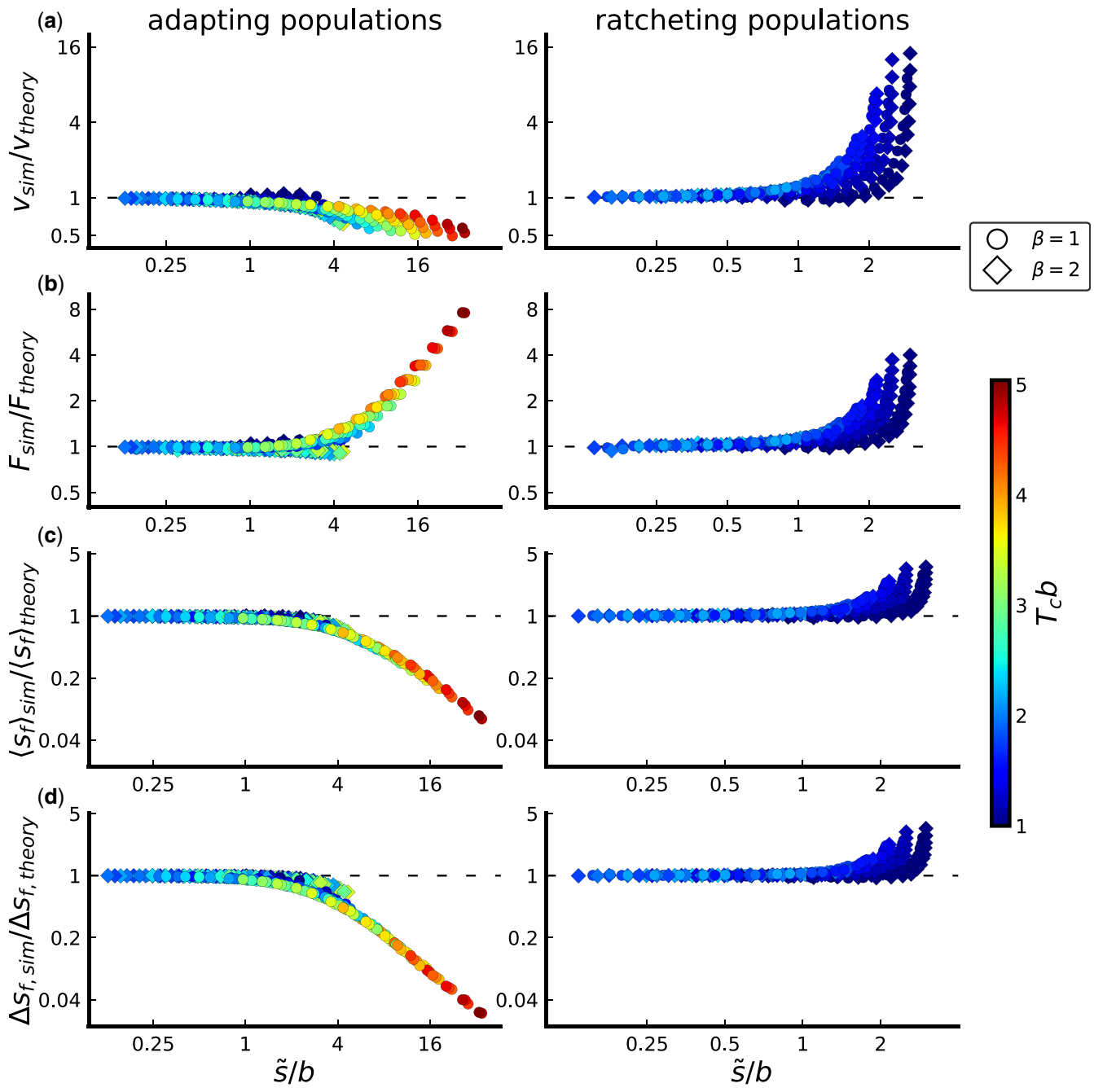

Fig. 4. Comparison between simulated and predicted rate v of fitness change (a), rate F of mutation accumulation (b), average fixed effect $\langle s_f \rangle$ (c), and standard deviation Δs_f in fixed effects (d). Populations in the left column are subject to only beneficial mutations; populations 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).

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 $h_s(\nu)$ (Sawyer and Hartl 1992; Hartl *et al.* 1994), which are lacking for rapidly evolving populations.

In the following subsections, we develop analytical predictions for these neutral and selected site frequency spectra in the MSSM regime, noting important departures from the classical intuition in Equation (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 BSC, 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 Equation (17). In the *Discussion*, we consider implications of our results in the context of the population genetic inference methods described above.

The site frequency spectrum: basic formalism

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

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

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:

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

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:

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

where μ is the per-site mutation rate and

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

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. Equation (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 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 Supplementary 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 Equation (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 Equation (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 *Analysis*. The identity $\Gamma(m) C^{-m} = \int_0^\infty dz z^{m-1} e^{-zC}$ then yields

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

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 Supplementary Appendix A,

$$\langle e^{-z n(t)} \rangle \approx e^{-\Phi_0(z,t)}, \tag{61}$$

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

$$\partial_t \phi_z(x,t) = -v \partial_x \phi_z(x,t) + x \phi_z(x,t) + \int \mu(s) [\phi_z(x+s,t) - \phi_z(x,t)] ds - \phi_z^2(x,t), \tag{62}$$

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

$$\langle e^{-z n_s(t)} \rangle \approx e^{-\Phi_s(z,t)}, \tag{63}$$

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

A detailed analysis of Equation (62) is conducted by Fisher (2013), which we review and extend in Supplementary 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 Equation (199), and that for $t < T_c - \mathcal{O}(1/b)$, $\Phi_0(z,t)$ is approximately linear in z. In Supplementary Appendix E, we show, using the properties of $\phi_z(x,t)$ discussed in Supplementary Appendix A along with our result for $f(x)$ in Equation (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 Equation (58) and integrating over time.

The ratio between the selected and neutral SFSs

The simple behavior of Equation (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 Equation (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 Equation (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, Equation (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 Equation (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 Equation (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 Supplementary Fig. 3. 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 Equation (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 Equation (1) (Sawyer and Hartl 1992; Hartl *et al.* 1994). 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 Equation (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

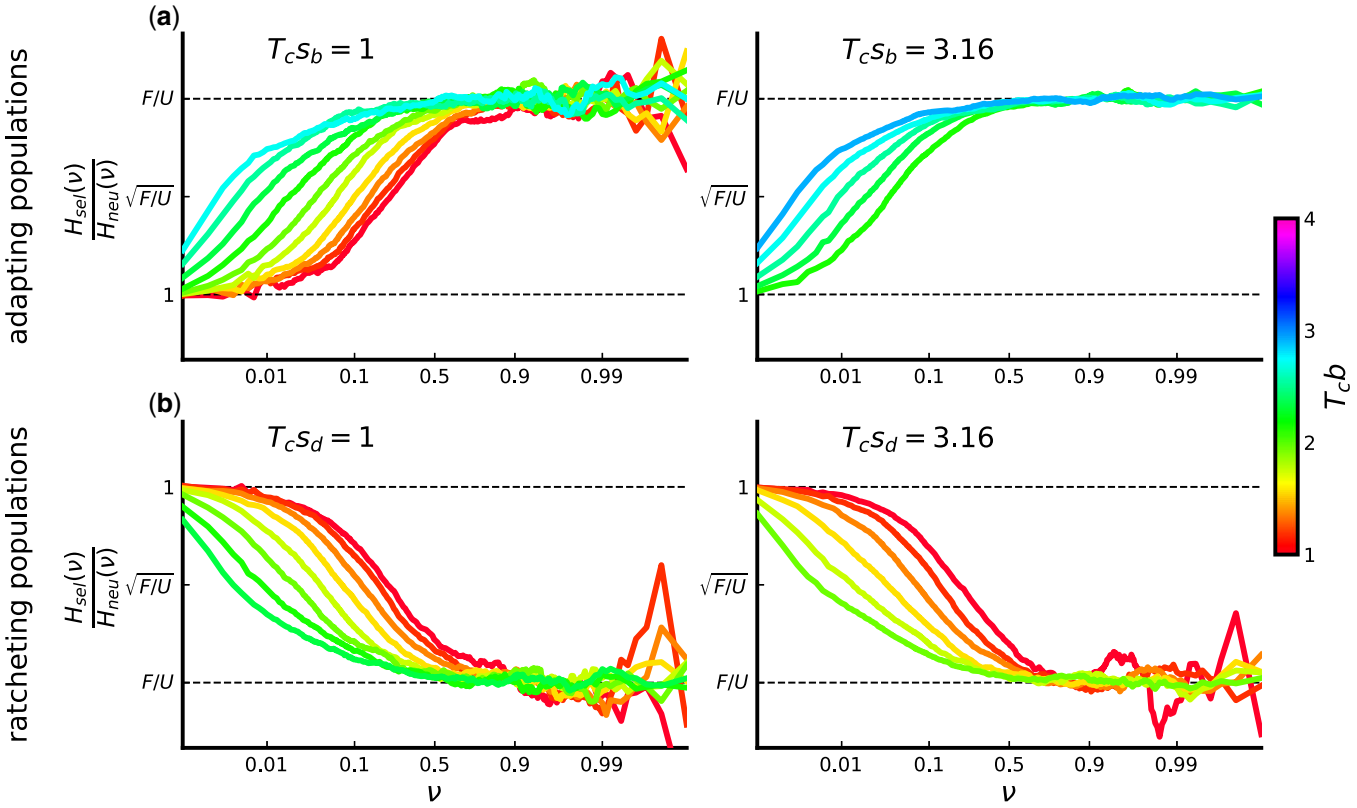

Fig. 5. Ratio between selected and neutral SFSs, scaled by the ratio between selected and neutral mutation rates, for adapting populations (a) and ratcheting populations (b) subject to a single-effect DFE with $T_c s$ values denoted in each panel. Ratio curves are plotted on a log scale and are scaled 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 Supplementary Fig. 3; for each panel, the represented $T_c b$ values are linearly spaced.

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 mediated 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 Supplementary Fig. 4, 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_{cs} 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_{cb} and/or $N\sigma$ observed above for fixed values of $T_c\langle s \rangle$.

The neutral and selected site frequency spectra

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

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

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 Equation (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), Equation (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 Equation (69) is valid. We will refer to the large-m limit of $mP_0(\nu m|m)$ with $P_0(k|m)$ given by Equation (69) as $h^{\text{BSC}}_{\text{neu}}(\nu)$ and provide an expression for $h^{\text{BSC}}_{\text{neu}}(\nu)$ in terms of special functions in Supplementary Appendix E.

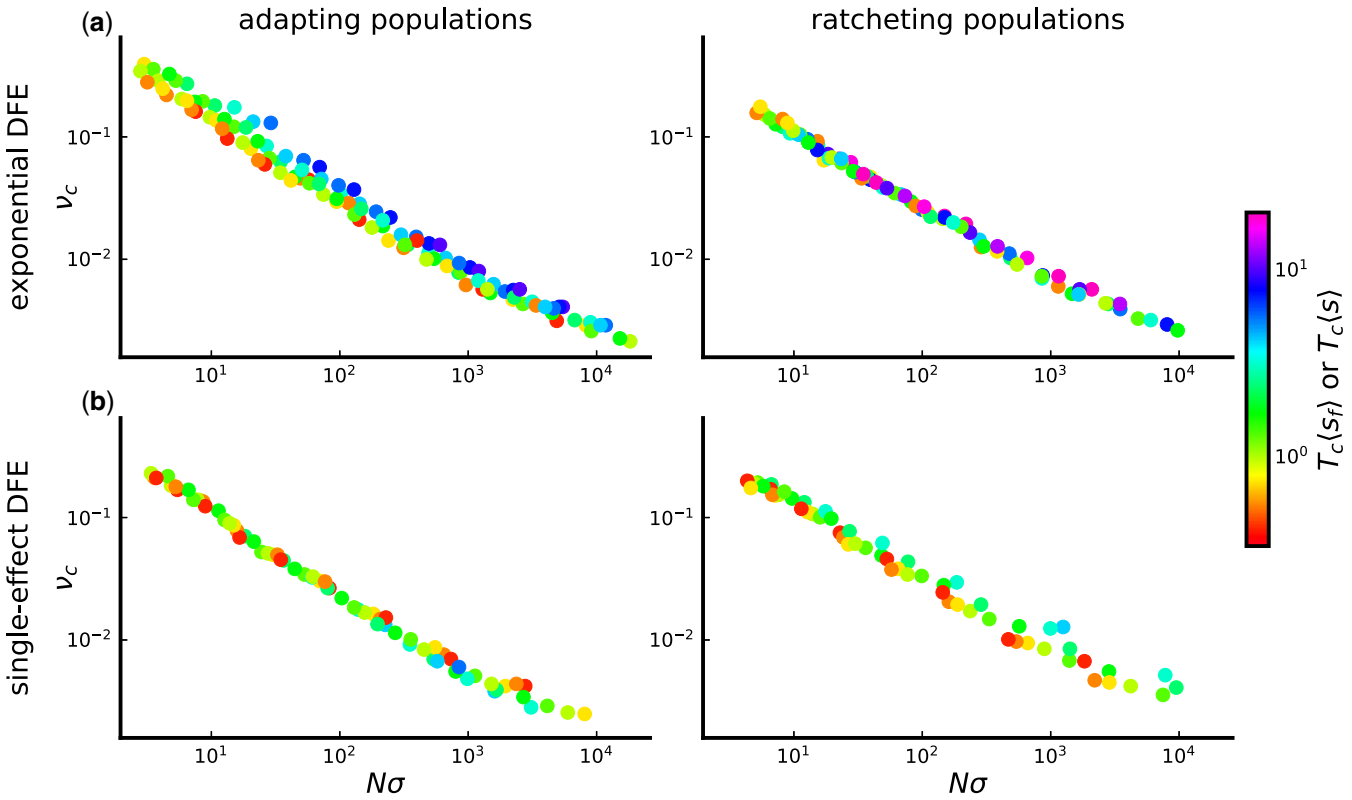

Fig. 6. Dependence of the crossover frequency ν_c on $N\sigma$. Simulated populations in (a) are subject to an exponential DFE, while simulated populations in (b) are subject to a single-effect DFE. For adapting populations (left) points are colored by their values of $T_c\langle s_f \rangle$, while 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 Supplementary Fig. 1. 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 Supplementary Fig. 3. 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).

The quantity $h^{\mathrm{BSC}}_{\mathrm{neu}}(\nu)$ well approximates the actual SFS $H_{\mathrm{neu}}(\nu)$ only for ν such that typical observed mutations have ages t satisfying $b(t-T_c) \gg 1$. The time integral yielding Equation (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 Equation (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 $\langle e^{-zn(t|x)}\rangle \approx e^{-\phi_z(x,t)}$. In Supplementary 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 Supplementary Appendix E, we show how this generating function can be inverted to obtain an approximation for $h_{\mathrm{neu}}(\nu)$,

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

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

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

which is precisely the asymptotic behavior of $h^{\mathrm{BSC}}_{\mathrm{neu}}(\nu)$ as $\nu \to 0$; thus the two functions $h^{\mathrm{Airy}}_{\mathrm{neu}}(\nu)$ and $h^{\mathrm{BSC}}_{\mathrm{neu}}(\nu)$ have a smooth crossover at $\frac{1}{T_c b}\log(1/\nu) \sim 1$. In principle, our predictions $h^{\mathrm{Airy}}_{\mathrm{neu}}(\nu)$ and $h^{\mathrm{BSC}}_{\mathrm{neu}}(\nu)$ could be connected by asymptotic matching. Notably, $h^{\mathrm{BSC}}_{\mathrm{neu}}(\nu)$ 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, $h^{\mathrm{Airy}}_{\mathrm{neu}}(\nu)$ 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 $h^{\mathrm{Airy}}_{\mathrm{neu}}(\nu)$ and $h^{\mathrm{BSC}}_{\mathrm{neu}}(\nu)$, 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 $n_0 e^{xt-\sigma^2 t^2/2}$ (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

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

In Supplementary 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 $n(t|x) \approx n_0 e^{xt-\sigma^2 t^2/2}$ 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

$$h^{\mathrm{Deterministic}}_{\mathrm{neu}}(\nu) \propto \frac{1}{\nu^2}\frac{d\log\nu_{peak}}{dt_{peak}}\Big|_{\nu_{peak}=\nu} \tag{73}$$

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 Equation (73) is evaluated at t_{peak} such that $\nu_{peak} = \nu$. The applicability of Equation (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 Supplementary Appendix E, Equation (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 $h^{\mathrm{Deterministic}}_{\mathrm{neu}}(\nu)$ tends to $h^{\mathrm{Gaussian}}_{\mathrm{neu}}(\nu)$ given in Equation (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 and Kimura 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 $\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-1/(N\nu)}^{1/(N\nu)} e^{-x^2/2\sigma^2}dx = \mathrm{erf}\left(\frac{1}{\sqrt{2}N\sigma\nu}\right)$, 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

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

which has a relatively smooth crossover to $h^{\mathrm{Gaussian}}_{\mathrm{neu}}(\nu)$ 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 $h_{\mathrm{neu}}(\nu) \to 2N\mu$ 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,

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

where $h^{\mathrm{BSC}}_{\mathrm{neu}}(\nu)$, $h^{\mathrm{Airy}}_{\mathrm{neu}}(\nu)$, $h^{\mathrm{Gaussian}}_{\mathrm{neu}}(\nu)$, $h^{\mathrm{Drift}}_{\mathrm{neu}}(\nu)$ are given above. Note that mutations with ages $t < T_c$ (accounted for by $h^{\mathrm{Gaussian}}_{\mathrm{neu}}(\nu)$) and with ages $t > T_c$ or $b(t - T_c) = \mathcal{O}(1)$ (accounted for by $h^{\mathrm{BSC}}_{\mathrm{neu}}(\nu)$ and $h^{\mathrm{Airy}}_{\mathrm{neu}}(\nu)$, respectively) contribute additively to the SFS, which motivates the inclusion of the sums in Equation (75). $h^{\mathrm{Gaussian}}_{\mathrm{neu}}(\nu)$ is important primarily for $\nu < e^{-T_c b}$—and over a broad range of $\log\nu$, both $h^{\mathrm{Gaussian}}_{\mathrm{neu}}(\nu)$ and $h^{\mathrm{Airy}}_{\mathrm{neu}}(\nu)$ 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 Equation (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 $H^{\mathrm{PW}}_{\mathrm{sel}}(\nu)$. This function is defined completely analogously to $H^{\mathrm{PW}}_{\mathrm{neu}}(\nu)$ in Equation (75), with analogous contributions $h^{\mathrm{Drift}}_{\mathrm{sel}}(\nu)$, $h^{\mathrm{Gaussian}}_{\mathrm{sel}}(\nu)$, $h^{\mathrm{Airy}}_{\mathrm{sel}}(\nu)$, and $h^{\mathrm{BSC}}_{\mathrm{sel}}(\nu)$ from selected mutations. The only differences are that μ is replaced by $\mu F/U$ in $h^{\mathrm{Airy}}_{\mathrm{sel}}(\nu)$, $h^{\mathrm{BSC}}_{\mathrm{sel}}(\nu)$, and in the upper limit to $h^{\mathrm{PW}}_{\mathrm{sel}}(\nu)$ imposed by $2N\mu$. These replacements are justified because the contributions to $h^{\mathrm{Airy}}_{\mathrm{sel}}(\nu)$ and $h^{\mathrm{BSC}}_{\mathrm{sel}}(\nu)$ 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 $H^{\mathrm{PW}}_{\mathrm{neu}}(\nu)$ and $H^{\mathrm{PW}}_{\mathrm{sel}}(\nu)$ with simulated site frequency spectra in Supplementary Figs. 5–8.

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

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

In Supplementary Fig. 9, we verify that Equation (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 $h_{\mathrm{sel}}(\nu)/h_{\mathrm{neu}}(\nu)$ at the lowest frequencies $\mathcal{O}\left(\frac{1}{N\sigma}\right)$, for future work.

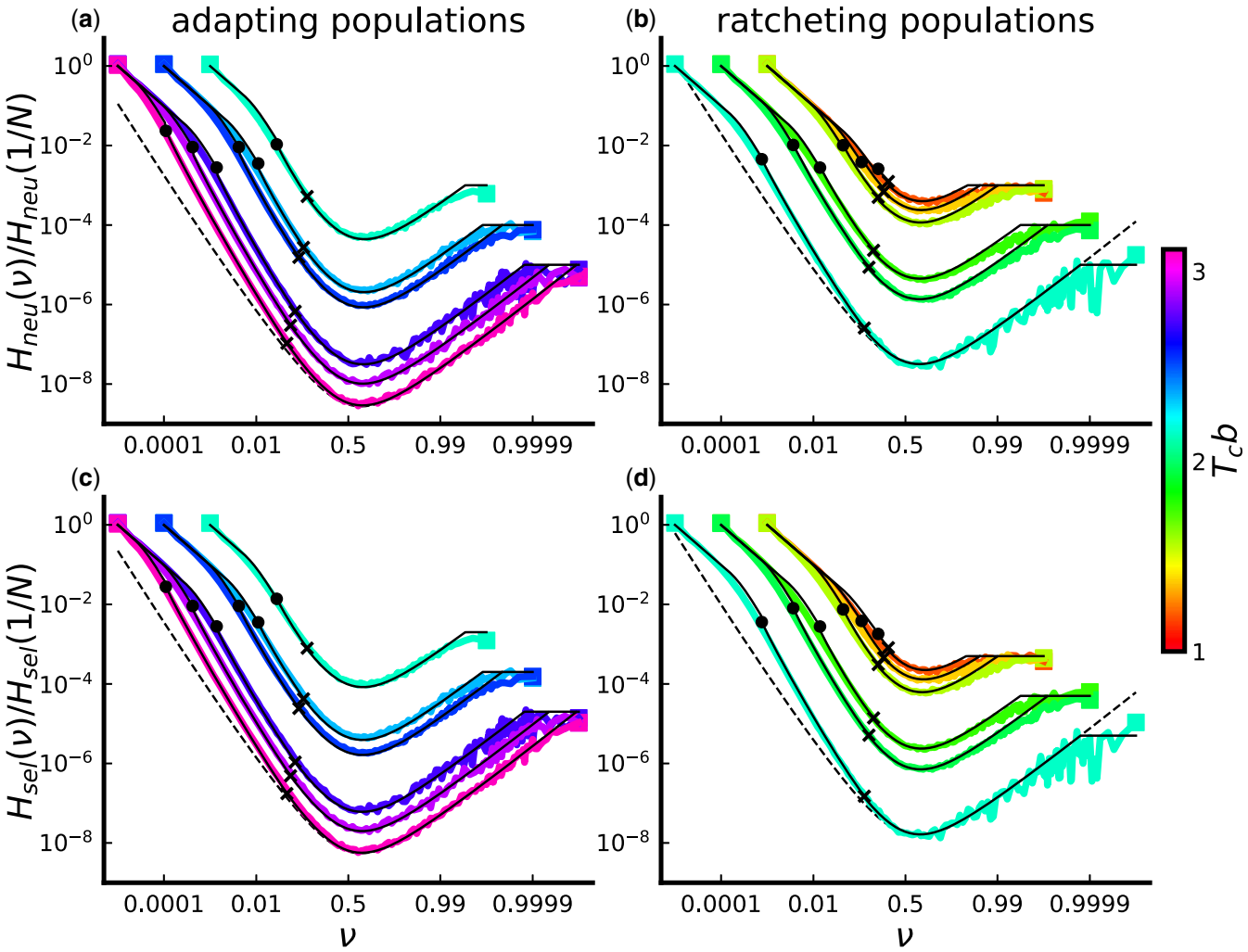

Fig. 7. Site frequency spectrum of neutral mutations (a and b) and of selected mutations (c and d). Colored lines denote SFSs observed in simulations, averaged over at least 800 epochs and smoothed using a moving average box kernel smoother. Solid black lines denote the corresponding theory predictions of the piecewise-defined function in Equation (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.

Pairwise heterozygosity and coalescence times

A special case of the neutral site frequency spectrum is the pairwise neutral heterozygosity $\pi_{neu} \equiv P_0(1|2)$. More generally, we can consider the pairwise heterozygosity $P_s(1|2)$ for a single selected site as well as the aggregate heterozygosity $\pi_{sel} \equiv \int P_s(1|2)\rho(s)ds$. With these definitions, π_{neu} and π_{sel} can both be computed using Equation (69) and the SFS ratio given in Equation (67). These expressions, however, are obtained under the assumption that typical contributing mutations have ages $t > T_c$. In reality, π_{neu} (as well as, rather generally, π_{sel}) receives a substantial contribution both from mutations with ages $t > T_c$ and from mutations with ages $t < T_c$. In Supplementary 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 Equation (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 Equation (77) reduces to

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

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

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

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 Supplementary 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 Supplementary 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 Equation (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 Equation (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 Supplementary Fig. 10; a similar level of agreement is obtained.

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 Supplementary Appendix C, we show how our approximate solutions $f(x)$ and $w(x)$ in Equations (25) and (26) can be considered local

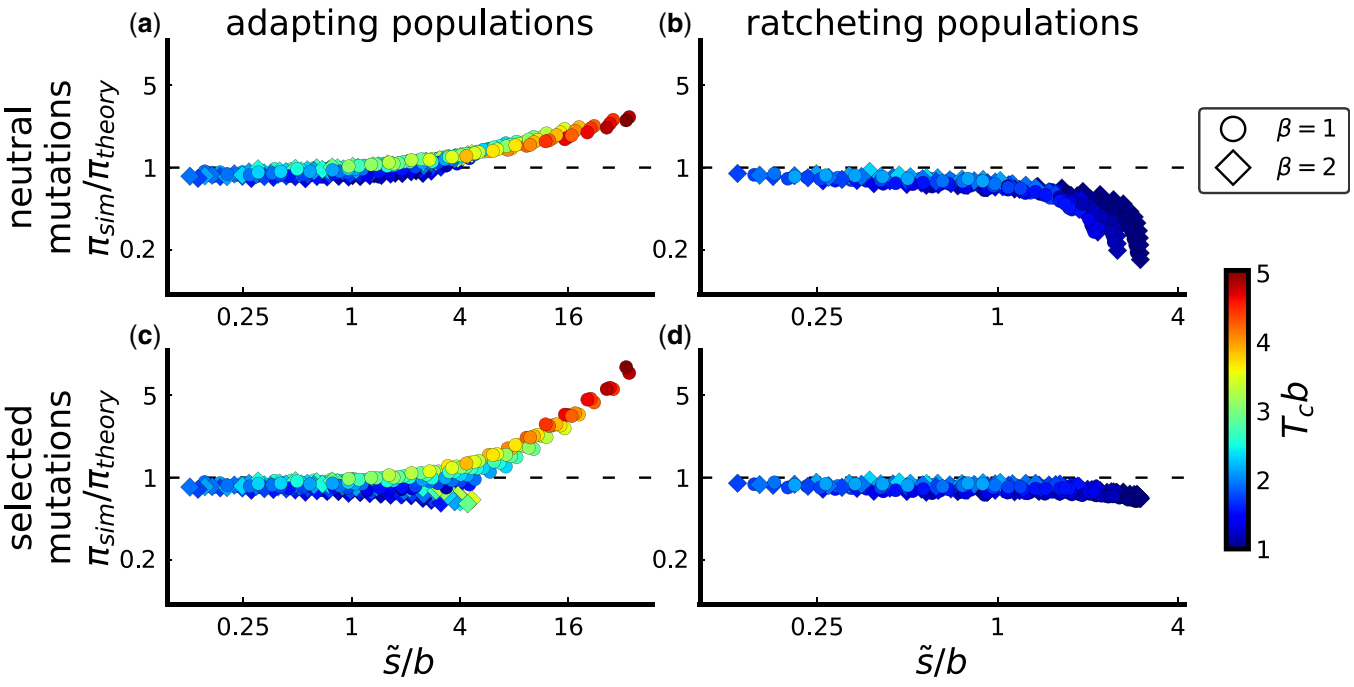

Fig. 8. Comparison between simulated and predicted heterozygosity π of neutral mutations (a and b), and between simulated and predicted heterozygosity π of selected mutations (c and d), for adapting populations (a and c) and ratcheting populations (b and d). Parameters are identical to those simulated in Fig. 4.

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.

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

Our condition of validity for the MSSM regime, expressed in Equation (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 Supplementary Appendix F), we can gain further intuition on the dynamics of lineages as they fall behind 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 $D \equiv \frac{1}{2} U \langle s^2 \rangle$, 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.* 2014). 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).

Interpretations of the timescale T_c

A key result of *Statistics of Genetic Diversity* is that $T_c \approx \langle T_2 \rangle / 2$. This motivates our interpretation of T_c [as defined in Equation (17)] as a *coalescence* timescale. The timescale T_c can also be interpreted in a few different ways. In Supplementary 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 Supplementary Appendix A,

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

with $\lim_{t\to\infty} A(x,t) = N f(x) w(x)$ 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 $A(x, T_c) \sim \mathrm{Ai}(\frac{c-x}{b})$ comes to resemble (and, subsequently, rapidly converges to on a timescale $\mathcal{O}(1/b)$) the eventual distribution of ancestor fitnesses $Nf(x)w(x) \propto \mathrm{Ai}^2(\frac{c-x}{b})$. 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.

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 eventually 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 $\langle e^{-\theta \nu_L(t)} \rangle$ is considered by Fisher (2013)—though in that work, $g(x, t)$ in Equation (81) is primarily taken as a fitness distribution of an entire *population* whose total size can fluctuate. In Supplementary Appendix E, we briefly review key features of $\langle e^{-\theta \nu_L(t)} \rangle$ 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 $\langle e^{-\theta \nu_L(t)} \rangle$ 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 $\langle e^{-\theta \nu_L(t)} \rangle$, 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

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

where $\alpha(t) \equiv e^{-t/T_c}$. We carry out this calculation in Supplementary Appendix E. Up to a change in timescale, $G_t(\nu_L(t),\nu_L(0))$ in Equation (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 an $1/n^2$ 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.

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; Ohta 1973; Akashi *et al.* 2012). 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 *Escherichia coli* (Hartl *et al.* 1994) and *Drosophila* (Akashi 1995; Zeng and Charlesworth 2009; Machado *et al.* 2020), 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 (Sawyer and Hartl 1992; Bustamante *et al.* 2001; 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 an 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 3.

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 4, we provide the (approximate) minimum N required for validity of the MSSM approximation, for four 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 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

Table 3. Qualitative properties of different regimes.

	MSSM regime	"Moderate speeds" regime	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.

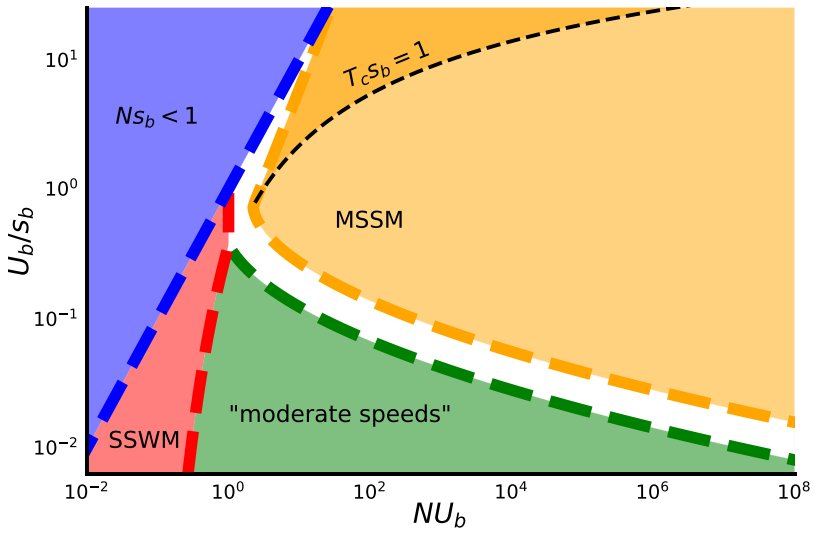

Fig. 9. Phase diagram for the case of a single beneficial effect s_b, in the space of dimensionless parameters U_b/s_b and NU_b. The blue line illustrates the $Ns_b = 1$ boundary which separates the neutral regime (blue region) from the SSWM regime (red region). The red line illustrates the $NU_b \log(Ns_b) = 1$ 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.

Table 4. 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)

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 relatively tractable, even for full DFEs, 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_{cs} 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 υ. 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.

Data availability

The authors state that all data necessary for confirming the conclusions presented in this article are represented fully within the article.

Supplemental material is available at *GENETICS* online.

Acknowledgments

The authors thank Ivana Cvijović and Jiseon Min for useful discussions and comments on the manuscript.

Funding

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.

Conflicts of interest

None declared.

Literature cited

Agarwala A, Fisher DS. Adaptive walks on high-dimensional fitness landscapes and seascapes with distance-dependent statistics. Theor Popul Biol. 2019;130:13–49.

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

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

Barton NH, Etheridge AM, Véber A. The infinitesimal model: definition, derivation, and implications. Theor Popul Biol. 2017;118:50–73.

Berestycki N. Recent progress in coalescent theory. Ensaios Matematicos. 2009;16:1–193.

Bolthausen E, Sznitman A-S. On Ruelle's probability cascades and an abstract cavity method. Commun Math Phys. 1998;197(2):247–276.

Brunet E, Derrida B, Mueller AH, Munier S. Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. Phys Rev E Stat Nonlin Soft Matter Phys. 2007;76(4 Pt 1):041104.

Buskirk SW, Peace RE, Lang GI. Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. Proc Natl Acad Sci U S A. 2017;114(31):8330–8335.

Bustamante CD, Wakeley J, Sawyer S, Hartl DL. Directional selection and the site-frequency spectrum. Genetics. 2001;159(4):1779–1788.

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

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

Cohen E, Kessler DA, Levine H. Front propagation up a reaction rate gradient. Phys Rev E Stat Nonlin Soft Matter Phys. 2005;72(6):066126.

Crow JF, Kimura M. An Introduction to Population Genetics Theory. New York: Harper and Row; 1970.

Cvijović I, Good BH, Desai MM. The effect of strong purifying selection on genetic diversity. Genetics. 2018;209(4):1235–1278.

Desai MM, Fisher DS. Beneficial mutation–selection balance and the effect of linkage on positive selection. Genetics. 2007;176(3):1759–1798.
Desai MM, Walczak AM, Fisher DS. Genetic diversity and the structure of genealogies in rapidly adapting populations. Genetics. 2013;193(2):565–585.
Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007;8(8):610–618.
Eyre-Walker A, Woolfit M, Phelps T. The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics. 2006;173(2):891–900.
Falconer D, Mackay T. Introduction to Quantitative Genetics.Essex (UK): Longman Group; 1996.
Fisher DS. Asexual evolution waves: fluctuations and universality. J Stat Mech. 2013;2013(1):P01011.
Fisher RA. XVII—the distribution of gene ratios for rare mutations. Proc R Soc Edinb. 1931;50:204–219.
Fogle CA, Nagle JL, Desai MM. Clonal interference, multiple mutations and adaptation in large asexual populations. Genetics. 2008;180(4):2163–2173.
Gerrish PJ, Lenski RE. The fate of competing beneficial mutations in an asexual population. Genetica. 1998;102:127.
Gillespie JH. A simple stochastic gene substitution model. Theor Popul Biol. 1983;23(2):202–215.
Gillespie JH. Molecular evolution over the mutational landscape. Evolution. 1984;38(5):1116–1129.
Good BH, Desai MM. Fluctuations in fitness distributions and the effects of weak linked selection on sequence evolution. Theor Popul Biol. 2013;85:86–102.
Good BH, Desai MM. Deleterious passengers in adapting populations. Genetics. 2014;198(3):1183–1208.
Good BH, Desai MM. The impact of macroscopic epistasis on long-term evolutionary dynamics. Genetics. 2015;199(1):177–190.
Good BH, Rouzine IM, Balick DJ, Hallatschek O, Desai MM. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. Proc Natl Acad Sci U S A. 2012;109(13):4950–4955.
Good BH, Walczak AM, Neher RA, Desai MM. Genetic diversity in the interference selection limit. PLoS Genet. 2014;10(3):e1004222.
Goyal S, Balick DJ, Jerison ER, Neher RA, Shraiman BI, Desai MM. Dynamic mutation–selection balance as an evolutionary attractor. Genetics. 2012;191(4):1309–1319.
Haigh J. The accumulation of deleterious genes in a population—Muller's ratchet. Theor Popul Biol. 1978;14(2):251–267.
Hallatschek O. The noisy edge of traveling waves. Proc Natl Acad Sci U S A. 2011;108(5):1783–1787.
Hallatschek O. Selection-like biases emerge in population models with recurrent jackpot events. Genetics. 2018;210(3):1053–1073.
Hartl DL, Moriyama EN, Sawyer SA. Selection intensity for codon bias. Genetics. 1994;138(1):227–234.
Hegreness M, Shoresh N, Hartl D, Kishony R. An equivalence principle for the incorporation of favorable mutations in asexual populations. Science. 2006;311(5767):1615–1617.
Held T, Klemmer D, Lässig M. Survival of the simplest in microbial evolution. Nat Commun. 2019;10(1):1–11.
Hudson RR, Kaplan NL. Deleterious background selection with recombination. Genetics. 1995;141(4):1605–1617.
Jerison ER, Desai MM. Genomic investigations of evolutionary dynamics and epistasis in microbial evolution experiments. Curr Opin Genet Dev. 2015;35:33–39.
Johnson MS, Martsul A, Kryazhimskiy S, Desai MM. Higher-fitness yeast genotypes are less robust to deleterious mutations. Science. 2019;366(6464):490–493.
Kao KC, Sherlock G. Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. Nat Genet. 2008;40(12):1499–1504.
Kaplan NL, Darden T, Hudson RR. The coalescent process in models with selection. Genetics. 1988;120(3):819–829.
Kersting G, Siri-Jégousse A, Wences AH. Site frequency spectrum of the Bolthausen-Sznitman coalescent. arXiv preprint arXiv:1910.01732. 2019. https://doi.org/10.48550/arXiv.1910.01732
Kimura M. Evolutionary rate at the molecular level. Nature. 1968;217(5129):624–626.
Kingman J. The coalescent. Stochastic Processes Appl. 1982;13(3):235–248.
Kosheleva K, Desai MM. The dynamics of genetic draft in rapidly adapting populations. Genetics. 2013;195(3):1007–1025.
Krone SM, Neuhauser C. Ancestral processes with selection. Theor Popul Biol. 1997;51(3):210–237.
Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science. 2014;344(6191):1519–1522.
Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, Desai MM. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature. 2013;500(7464):571–574.
Lieberman TD, Flett KB, Yelin I, Martin TR, McAdam AJ, Priebe GP, Kishony R. Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. Nat Genet. 2014;46(1):82–87.
Machado HE, Lawrie DS, Petrov DA. Pervasive strong selection at the level of codon usage bias in *Drosophila melanogaster*. Genetics. 2020;214(2):511–528.
McCandlish D, Stoltzfus A. Modeling evolution using the probability of fixation: history and implications. Q Rev Biol. 2014;89(3):225–252.
McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 1991;351(6328):652–654.
McVean GAT, Charlesworth B. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics. 2000;155(2):929–944.
Messer PW, Petrov DA. Frequent adaptation and the McDonald–Kreitman test. Proc Natl Acad Sci U S A. 2013;110(21):8615–8620.
Miralles R, Gerrish PJ, Moya A, Elena SF. Clonal interference and the evolution of RNA viruses. Science. 1999;285(5434):1745–1747.
Mohle M, Pitters H. A spectral decomposition for the block counting process of the Bolthausen-Sznitman coalescent. Electron Commun Probab. 2014;19:1–11.
Nachman MW. Deleterious mutations in animal mitochondrial DNA. Genetica. 1998;102:61.
Neher RA. Genetic draft, selective interference, and population genetics of rapid adaptation. Annu Rev Ecol Evol Syst. 2013;44(1):195–215.
Neher RA, Hallatschek O. Genealogies of rapidly adapting populations. Proc Natl Acad Sci U S A. 2013;110(2):437–442.
Neher RA, Kessinger TA, Shraiman BI. Coalescence and genetic diversity in sexual populations under selection. Proc Natl Acad Sci U S A. 2013;110(39):15836–15841.
Neher RA, Russell CA, Shraiman BI, McVean G. Predicting evolution from the shape of genealogical trees. eLife. 2014;3:e03568.
Neher RA, Shraiman BI. Genetic draft and quasi-neutrality in large facultatively sexual populations. Genetics. 2011;188(4):975–996.
Neher RA, Shraiman BI. Fluctuations of fitness distributions and the rate of Muller's ratchet. Genetics. 2012;191(4):1283–1293.
Neher RA, Shraiman BI, Fisher DS. Rate of adaptation in large sexual populations. Genetics. 2010;184(2):467–481.

Nielsen R. Molecular signatures of natural selection. Annu Rev Genet. 2005;39(1):197–218.
Nielsen R, Yang Z. Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. Mol Biol Evol. 2003;20(8):1231–1239.
Nolan JP. Stable Distributions - Models for Heavy Tailed Data. Boston: Birkhauser; 2018. Chapter 1. http://fs2.american.edu/jpnolan/www/stable/stable.html.
Nourmohammad A, Otwinowski J, Łuksza M, Mora T, Walczak AM. Fierce selection and interference in b-cell repertoire response to chronic hiv-1. Mol Biol Evol. 2019;36(10):2184–2194.
Ohta T. Slightly deleterious mutant substitutions in evolution. Nature. 1973;246(5428):96–98.
Orr HA. The distribution of fitness effects among beneficial mutations. Genetics. 2003;163(4):1519–1526.
Peck JR. A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. Genetics. 1994;137(2):597–606.
Pitman J. Exchangeable and partially exchangeable random partitions. Probab Th Rel Fields. 1995;102(2):145–158.
Pitman J. Coalescents with multiple collisions. Ann Probab. 1999;27(4):1870–1902.
Pitman J, Yor M. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. Ann Probab. 1997;25(2):855–900.
Rouzine IM, Brunet É, Wilke CO. The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. Theor Popul Biol. 2008;73(1):24–46.
Rouzine IM, Coffin JM. Evolution of human immunodeficiency virus under selection and weak recombination. Genetics. 2005;170(1):7–18.
Rouzine IM, Wakeley J, Coffin JM. The solitary wave of asexual evolution. Proc Natl Acad Sci U S A. 2003;100(2):587–592.
Sawyer SA, Hartl DL. Population genetics of polymorphism and divergence. Genetics. 1992;132(4):1161–1176.
Sawyer SA, Kulathinal RJ, Bustamante CD, Hartl DL. Bayesian analysis suggests that most amino acid replacements in drosophila are driven by positive selection. J Mol Evol. 2003;57:S154–S164.
Sawyer SA, Parsch J, Zhang Z, Hartl DL. Prevalence of positive selection among nearly neutral amino acid replacements in Drosophila. Proc Natl Acad Sci U S A. 2007;104(16):6504–6510.
Schiffels S, Szöllosi GJ, Mustonen V, Lässig M. Emergent neutrality in adaptive asexual evolution. Genetics. 2011;189(4):1361–1375.
Schweinsberg J. Coalescent processes obtained from supercritical Galton–Watson processes. Stochastic Processes Appl. 2003;106(1):107–139.
Slatkin M, Hudson RR. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics. 1991;129(2):555–562.
Söderberg RJ, Berg OG. Mutational interference and the progression of Muller's ratchet when mutations have a broad range of deleterious effects. Genetics. 2007;177(2):971–986.
Strelkowa N, Lässig M. Clonal interference in the evolution of influenza. Genetics. 2012;192(2):671–682.
Tsimring LS, Levine H, Kessler DA. RNA virus evolution via a fitness-space model. Phys Rev Lett. 1996;76(23):4440–4443.
Wakeley J. Coalescent Theory, an Introduction. Greenwood Village (CO): Roberts and Company; 2005.
Weissman DB, Barton NH. Limits to the rate of adaptive substitution in sexual populations. PLoS Genet. 2012;8(6):e1002740.
Weissman DB, Hallatschek O. The rate of adaptation in large sexual populations with linear chromosomes. Genetics. 2014;196(4):1167–1183.
Wright S. Evolution in Mendelian populations. Genetics. 1931;16(2):97–159.
Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 2000;15(12):496–503.
Yule GU. II—a mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. Philos Trans R Soc London Ser B. 1925;213(402–410):21–87.
Zeng K, Charlesworth B. Estimating selection intensity on synonymous codon usage in a nonequilibrium population. Genetics. 2009;183(2):651–662.

Communicating editor: N. Barton