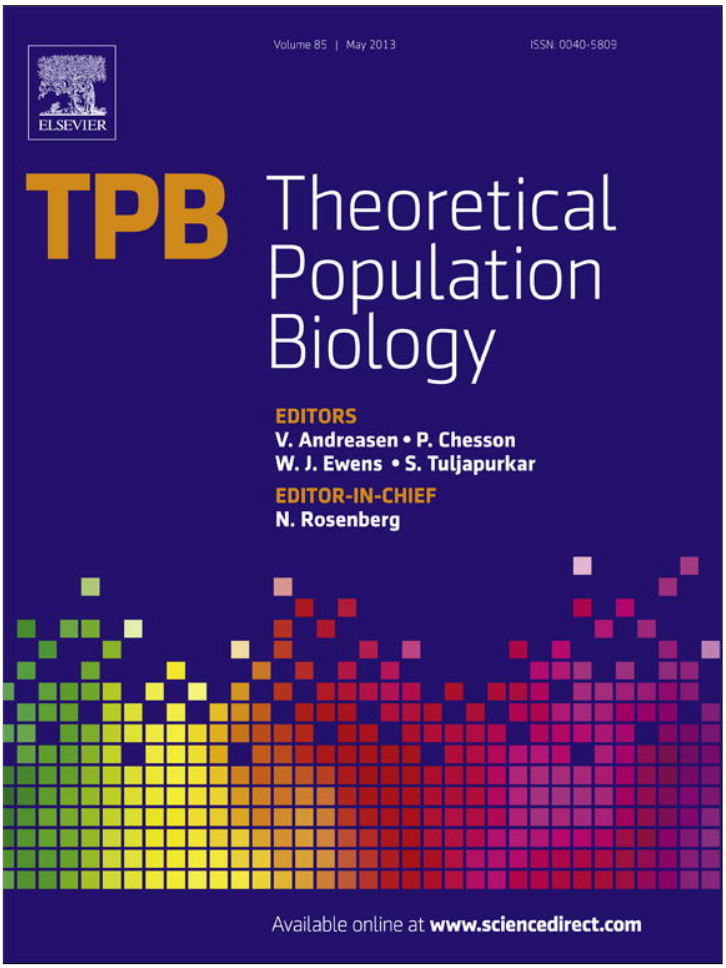

Theoretical Population Biology 85 (2013) 86–102

Contents lists available at SciVerse ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Fluctuations in fitness distributions and the effects of weak linked selection on sequence evolution

Benjamin H. Good, Michael M. Desai *

Department of Organismic and Evolutionary Biology, Department of Physics, and FAS Center for Systems Biology, Harvard University, United States

ARTICLE INFO

Article history:
Received 30 November 2012
Available online 18 January 2013

Keywords:
Linked selection
Fitness distribution
Interference
Adaptation
Muller's ratchet

ABSTRACT

Evolutionary dynamics and patterns of molecular evolution are strongly influenced by selection on linked regions of the genome, but our quantitative understanding of these effects remains incomplete. Recent work has focused on predicting the distribution of fitness within an evolving population, and this forms the basis for several methods that leverage the fitness distribution to predict the patterns of genetic diversity when selection is strong. However, in weakly selected populations random fluctuations due to genetic drift are more severe, and neither the distribution of fitness nor the sequence diversity within the population are well understood. Here, we briefly review the motivations behind the fitness-distribution picture, and summarize the general approaches that have been used to analyze this distribution in the strong-selection regime. We then extend these approaches to the case of weak selection, by outlining a perturbative treatment of selection at a large number of linked sites. This allows us to quantify the stochastic behavior of the fitness distribution and yields exact analytical predictions for the sequence diversity and substitution rate in the limit that selection is weak.

1. Introduction

A central goal of modern population genetics is to predict the diversity and fate of DNA sequences within a population, accounting for the joint effects of mutation, recombination, and natural selection at the sequence level. Genetic diversity is a fundamental feature on these genomic scales, since mutation rates in most organisms are so large that many sequence variants coexist within the population at any given time (Begun et al., 2007; Kreitman, 1983; Lewontin and Hubby, 1966; Nelson et al., 2012; Nik-Zinal et al., 2012; Rambaut et al., 2008). Some fraction of these mutations will have a negligible effect on reproductive fitness, but mounting empirical evidence suggests that a significant number are influenced by natural selection as well (see Hahn (2008) for a recent review). It is therefore imperative that our models of sequence evolution should be able to describe a large number of variants at disparate sites within the genome, possibly with different effects on the reproductive fitness of each individual.

This picture of extensive diversity at the sequence level stands in contrast to the large body of population genetics theory developed during the first half of the 20th century, which typically focused on the fate of a single mutant allele (relative to the wildtype) at a single genetic locus. Numerous mathematical models have been proposed, even for this highly simplified scenario, which correspond to different underlying assumptions about the mechanisms of natural selection, the reproductive lifecycle of the organism, and so on (Ewens, 2004). Fortunately, many of the differences between these models become negligible in large populations, and in this case an elegant theoretical description of the two-allele, single-locus system can be obtained from the standard diffusion limit (Feller, 1951). The frequency f of a mutant allele with fitness effect s in a population of size N is assumed to satisfy the stochastic differential equation

$$\frac{\partial f}{\partial t} = \underbrace{s[f(1-f)]}_{\text{selection}} + \underbrace{\sqrt{\frac{f(1-f)}{N}}\,\eta(t)}_{\text{genetic drift}}, \tag{1}$$

where $\eta(t)$ is a stochastic noise term that will be defined in Section 2. This model relates the rate of change in f to the deterministic action of selection and the random effects of genetic drift; it is formally equivalent to the more traditional diffusion equation for the probability distribution of f (Feller, 1951). Although the full solution to Eq. (1) is quite complicated (Kimura, 1955; Song and Steinrücken, 2012), this diffusion model is simple enough to admit a number of useful and exact results, including the well-known formula for the probability of fixation of a new mutant,

$$p_{\text{fix}} = \frac{2s}{1 - e^{-2Ns}}, \tag{2}$$

* Corresponding author.
E-mail address: mdesai@oeb.harvard.edu (M.M. Desai).

0040-5809/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2013.01.005

and (in the limit of low mutation rate μ) the average pairwise heterozygosity,

$$\pi = 2\left(\frac{\mu}{s}\right)\left[\frac{e^{-2Ns} + 2Ns - 1}{1 - e^{-2Ns}}\right]. \quad (3)$$

The historical impact of this diffusion model cannot be overstated, and these simple results played a large role in illuminating both the qualitative and quantitative effects of genetic drift arising from the finite size of the population. However, extending these single-locus results to an explicitly sequence-based setting proves to be quite challenging when selection is present.

In principle, one can treat the entire genome as a single locus with each possible genotype represented by a unique allele. A genome of length L would therefore require 2^L separate alleles and a corresponding system of diffusion equations relating the $2^L - 1$ independent allele frequencies. This clearly becomes unwieldy for large genomes since the number of alleles grows exponentially with L, and the sparse mutational connectivity between the different sequences and their varying fitnesses removes much of the desired symmetry from the problem (Ethier and Kurtz, 1987). Even for a genome with just $L = 2$ sites, exact solutions can only be found for a few special cases, and one must often resort to numerical calculations (Barton and Etheridge, 2004) or Monte-Carlo simulations (Hill and Robertson, 1966).

A popular alternative approach is to treat each site in the genome as a separate locus and assume some sort of quasi-independent evolution among the various loci, so that the single-locus model in Eq. (1) applies to the *marginal* nucleotide frequencies at each site (Sawyer and Hartl, 1992). This independent-sites approximation, which is exact in the limit of infinite recombination, reflects a historical perception of linkage as an infrequent and generally small correction to an otherwise freely-recombining set of loci, as is often the case for a quantitative trait with genetic contributions from several distant sites (Barton and Turelli, 1991; Falconer, 1960; Neher and Shraiman, 2011b). But given the typical recombination rates in most organisms, this assumption is likely to break down on local genomic scales, and effectively asexual selection on particular haplotype blocks may be a more accurate description (Franklin and Lewontin, 1970; Slatkin, 1972). Moreover, it has been shown that selection within these linked regions leads to large deviations from the predictions assuming independent evolution between the various sites, even after adjusting for possible reductions in the effective population size (Bustamante et al., 2001; Charlesworth et al., 1993; Comeron and Kreitman, 2002; Good and Desai, 2012; Messer and Petrov, 2012). Correctly accounting for the effects of selection on local genomic scales remains one of the major outstanding problems in population genetics, and is a necessary prerequisite if we wish to take full advantage of the increasing availability of DNA sequence data in order to make inferences about the evolutionary forces acting within a population (Pool et al., 2010).

Recent advances in this area have employed a third approach – situated somewhere between the genotypes-as-alleles and sites-as-loci schemes – in which the distribution of fitnesses in the population plays a central role (Desai and Fisher, 2007; Goyal et al., 2012; Haigh, 1978; Hallatschek, 2011; Neher et al., 2010; Ohta and Kimura, 1973; Park and Krug, 2007; Rouzine et al., 2003; Tsimring et al., 1996). Although the fitness distribution may seem to be rather tangential to the sequence-oriented questions introduced above, this quantity turns out to play an important role in mediating the effects of linked selection within the population, and several promising methods predict the behavior of individual sequences based on their interactions with this population-wide distribution (Good et al., 2012; Hudson and Kaplan, 1994; Neher and Shraiman, 2011a; O'Fallon et al., 2010; Walczak et al., 2012; Zeng and Charlesworth, 2011). Instead of tracking the frequencies of all possible genotypes or just the marginal frequencies at each site, this approach requires an explicit model for the frequency of individuals at each possible fitness, otherwise known as a *fitness-class*. Here too, the interactions between mutation, recombination, drift, and selection can be quite complex, and significant progress has been made only in the case where genetic drift is negligible compared to these other evolutionary forces. This can often be a reasonable approximation in many populations, since the effects of genetic drift are typically less severe for the fitness classes than for the frequencies of the underlying genotypes.

Nevertheless, even in this fitness-class picture the effects of genetic drift cannot be excised completely, since they play a crucial role in the high-fitness "nose" of the fitness distribution that often controls the behavior in the rest of the population (Brunet et al., 2008; Desai and Fisher, 2007; Goyal et al., 2012; Hallatschek, 2011; Neher and Shraiman, 2012). Various ad-hoc methods have been devised to account for this drift-dominated nose and its relation to the deterministic behavior in the bulk population, which yield accurate predictions for simple quantities such as the average rate of adaptation and the fixation probability of new mutations. Yet because of their ad-hoc nature, it is not entirely clear when these approximations are likely to be valid, or whether they remain appropriate for more complicated quantities of interest. Furthermore, in populations with weaker selection this separation between the drift-dominated nose and the deterministic bulk starts to break down, and the random nature of genetic drift becomes important throughout the entire fitness distribution.

In the present work, we follow an approach that is orthogonal to both the weak-drift limit of this fitness-class description as well as the weak-mutation limit implicit in the standard single-locus treatment. Rather, we seek a fitness-class description for a regime with weak selection at a large number of completely linked sites. Suitably defined, the neutral limit of the population "fitness distribution" is exactly solvable, and the corrections in the presence of selection can be calculated order by order as a perturbation series in powers of the selection strength. The resulting expressions may have relevance to sequence data obtained from natural populations [particularly in the context of the nearly-neutral theory of evolution (Ohta, 1992)], but their primary value is qualitative. The zeroth-order neutral description offers a valuable window into the stochastic aspects of the population fitness distribution in the absence of the complicating effects of selection, while the higher-order terms give the exact corrections from interference at a large number of linked sites and help illuminate the previously obscure transition to neutrality. The exact nature of these selective corrections provides a valuable check on a number of common heuristic assumptions in the literature, which should agree with our asymptotic results when selection becomes weak.

2. Fitness classes and the population fitness distribution

The distribution of fitnesses within the population is itself a random object which changes in time and reflects the inherent stochasticity of the evolutionary process. Two populations with the same genetic composition and the same set of available mutations will typically possess different fitness distributions after evolving independently for the same amount of time, although these distributions will be related in some statistical sense. Like the stochastic frequency of a single mutant allele discussed above, the statistical properties of the fitness distribution can be described by a generalization of the diffusion model in Eq. (1) that makes both the large population and long genome limits explicit. We consider a population of N haploid individuals that acquire new mutations at a total rate U per generation. We assume that these mutations occur over a large number of loci, each

with relatively small contributions to the total fitness, so that a mutation of effect s arising in an individual with (log) fitness X increases its fitness to $X + s$. Furthermore, we assume that the number of loci is sufficiently large, and epistasis sufficiently weak, that the set of available mutations can be approximated by a continuous distribution of fitness effects $\rho(s)$ that remains constant throughout the relevant time interval.

The random arrival of new mutations and the effects of genetic drift are treated by a continuous-time stochastic model similar to the one introduced in Hallatschek (2011). Let $f(X,t)$ denote the relative frequency of individuals with absolute fitness X at time t, normalized so that $\int dX\, f(X,t) = 1$. In some infinitesimal time δt, these frequencies are incremented according to the stochastic update rule

$$\begin{aligned} f(X, t+\delta t) \propto\ & f(X,t) + Xf(X,t)\delta t \\ & + U\int ds\, [f(X-s,t) - f(X,t)]\,\delta t \\ & + \sqrt{\frac{f(X,t)\delta t}{N}}\,\eta(X,t), \end{aligned} \tag{4}$$

where $\eta(X,t)$ denotes a set of independent Gaussian noise terms with zero mean and unit variance (Gardiner, 1985), and the constant of proportionality is chosen to satisfy the population size constraint

$$\int dX\, f(X, t+\delta t) = 1. \tag{5}$$

This yields a familiar Langevin equation for the fitness distribution

$$\begin{aligned} \frac{\partial f(X)}{\partial t} =\ & \underbrace{\left[X - \overline{X}(t)\right] f(X)}_{\text{selection}} \\ & + \underbrace{U\int ds\, \rho(s)\,[f(X-s) - f(X)]}_{\text{mutation}} \\ & + \underbrace{\int dX'\, \left[\delta(X'-X) - f(X)\right]\sqrt{\frac{f(X')}{N}}\,\eta(X')}_{\text{genetic drift}}, \end{aligned} \tag{6}$$

where $\overline{X}(t) = \int dX\, Xf(X,t)$ is the mean fitness of the population (see Fig. 1). Like the diffusion approximation at a single locus, this stochastic model is thought to describe the universal behavior that emerges in the limit that $N \to \infty$ and $L \to \infty$, while the per-site mutation rate μ and the relevant fitnesses X tend to zero in such a way that the scaled quantities $NU \equiv NL\mu$ and NX completely determine the dynamics. This scaling behavior provides an important check on our intuition (as well as our algebra), since it implies that any effects that depend on $1/N$, X, or $N\mu$ alone are completely negligible in this model unless we explicitly relax one of these assumptions (e.g., the finite site effects in Appendix A).

Two of the defining features of this stochastic model arise from the population size constraint in Eq. (5) that connects Eqs. (4) and (6): the resulting selection term becomes a nonlinear function of $f(X,t)$, and the previously simple noise terms acquire a complicated correlation structure. Such features are inherent in any model that imposes a population size constraint in this manner. The only fitnesses that matter are the relative fitnesses $x = X - \overline{X}(t)$, which depend not just on the properties of a particular DNA sequence, but also on the global behavior of all of the other sequences in the population. Moreover, the action of genetic drift is correlated among the various fitness classes in order to respect the constant population size.

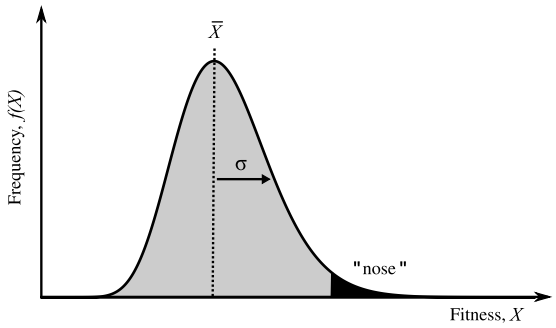

Fig. 1. A schematic depiction of the population fitness distribution, $f(X,t)$, which is obtained by grouping together genotypes with the same absolute fitness X. Important features of this distribution include the mean fitness $\overline{X}$ and the standard deviation σ, which is proportional to the typical fitness difference between two individuals in the population. We have also highlighted the high-fitness "nose" of the distribution, where genetic drift continues to dominate even in extremely large populations.

Yet despite the complex correlation structure of this drift term, we have constructed our stochastic model so that its average effect vanishes at any particular instant in time. Thus, we are led to examine the average profile $\langle f(X,t)\rangle$, which represents the expected value of the fitness distribution averaged over many independent populations. Taking the expectation of both sides of Eq. (6), we find that this average profile $\langle f(X,t)\rangle$ is governed by the *deterministic* differential equation

$$\begin{aligned} \frac{\partial \langle f(X)\rangle}{\partial t} =\ & X\langle f(X)\rangle - \int dX'\, X'\langle f(X) f(X')\rangle \\ & + \int ds\, \rho(s)\,[\langle f(X-s)\rangle - \langle f(X)\rangle]\,. \end{aligned}$$

However, this moment equation does not close: the nonlinear selection term in Eq. (6) implies that the future behavior of $\langle f(X)\rangle$ does not just depend upon its own value in the present, but also on the two-point correlation function $\langle f(X)f(X')\rangle$. The time evolution of this two-point correlation function will in turn depend on the three-point correlation function, and so on. One must therefore solve an infinite hierarchy of these moment equations in order to obtain predictions for the mean behavior. A similar hierarchy arises for the single-locus diffusion in Eq. (1), but in that case the simplicity of the drift term permits an exact solution. In contrast, the lack of closure among the moments of the fitness distribution is arguably the primary obstacle for a quantitative description of large numbers of interfering mutations, and many statistical properties of the fitness distribution remain unknown as a result.

Much of the existing work in this field has essentially focused on various ways to approximate this correlated selection term. This is often achieved through some sort of approximate factorization of the form

$$\langle (X - \overline{X}) f(X)\rangle \approx (X - \langle \overline{X}\rangle)\langle f(X)\,\rangle \tag{7}$$

so that the statistical aspects of the nonlinearity are marginalized.[1] Since the magnitude of genetic drift is proportional to $1/N$, one regime where this approximation appears quite naturally is in the strong-selection limit, $N|X - \overline{X}| \gg 1$, when the genetic drift term can be neglected in Eq. (6) and all higher correlations vanish. The canonical example of such a regime is the deleterious mutation–selection balance attained under strong purifying selection (Haigh, 1978), which has been intensely studied

[1] Strictly speaking, this sort of approximation is often more appropriate when this simple ensemble average is replaced with some other averaging scheme (see below) or an alternative measure of the "typical" behavior (Fisher, 2013).

in the context of Muller's ratchet (Etheridge et al., 2007; Gessler, 1995; Gordo and Charlesworth, 2000; Higgs and Woodcock, 1995; Jain, 2008; Muller, 1964; Neher and Shraiman, 2012; Stephan et al., 1993; Waxman and Loewe, 2010) and background selection (Charlesworth et al., 1993; Gordo et al., 2002; Hudson and Kaplan, 1994; Nicolaisen and Desai, 2012; Walczak et al., 2012). Neher and Shraiman (2012) have demonstrated that the deterministic limit becomes exact in this particular case when selection is infinitely strong. However, if beneficial mutations are present, or if some of the deleterious mutations are weakly selected, then even in this extreme limit the factorization in Eq. (7) does not hold for all X, since it starts to break down near the high-fitness "nose" of the distribution (see Fig. 1), where a relatively small number of individuals have an outsized chance of taking over the population. Thus, in this strong-selection limit one often speaks of a division of the population into a drift-dominated nose (where stochasticity is extremely important) and a deterministic bulk where Eq. (7) holds. When the distinction between these regions is sufficiently sharp, a number of highly successful (although somewhat ad-hoc) approximations have been developed to treat the stochasticity in the nose and to self-consistently match this behavior with the deterministic bulk of the population (Desai and Fisher, 2007; Goyal et al., 2012; Neher and Shraiman, 2012; Rouzine et al., 2003; Tsimring et al., 1996). Several alternative approaches are based on a modification of the stochastic dynamics in Eq. (6), which is chosen in a particular way so that the nonlinearity in the selection term vanishes by design (Fisher, 2013; Hallatschek, 2011). These models may be more appropriate when the boundary between the stochastic nose and the deterministic bulk is less pronounced, but their relation (and relevance) to the standard evolutionary model in Eq. (6) must be justified on an ad-hoc basis.

These methods are by far the most promising candidates for describing the evolutionary dynamics in a strong-selection regime relevant to many laboratory evolution experiments, microbial populations, or other rapidly adapting organisms. Yet from a purely theoretical standpoint, they suffer from a major shortcoming in that they attempt to describe a parameter regime for which no exact asymptotic description has been found. Although these methods were devised to approximate this asymptotic behavior, their correctness (apart from self-consistency) can only be validated by numerical comparisons to Monte-Carlo simulations of Eq. (6) for particular parameter values. This can make it difficult to test the individual assumptions that enter into these approximations or to compare different approximation methods, and it offers little direct information about which quantities or parameter regimes fall outside their domain of validity. On a more practical level, there may be many populations that are dominated by a large number of *weakly* selected mutations where these strong-selection methods do not apply. In this case, few quantitative descriptions exist apart from assuming strict neutrality, and our knowledge of the relevant processes in this regime is extremely limited.

Thus, while previous approaches have attempted to reconcile the joint effects of selection and drift by mostly neglecting the latter, our approach here will be exactly the opposite. Rather than focus on those regimes where the selection term can be factored like Eq. (7), we consider the weak-selection limit where this selection term can be neglected entirely, or at least treated as a small perturbation. Unlike the strong-selection limit, the zeroth order solution in this nearly-neutral regime can be treated exactly and the full statistical behavior can be elucidated, which leads to a natural (and similarly exact) perturbation expansion in the presence of selection.

3. The neutral limit

There are a variety of ways we could define the neutral limit of Eq. (6), but we are interested in one which does not lead to a trivial description of the resulting "fitness distribution". For example, there is a naive limit in which the fitness effects of all new mutations have $s = 0$, which implies that the entire population is confined to a single "fitness class" with fitness $X = 0$ for all time, i.e.

$$f(X) = \delta(X).$$

However, we can maintain much more of the interesting multi-locus behavior by ignoring the absolute fitness of each individual for the moment and concentrating instead on the number of mutations k that each individual possesses. In this limit, a population-wide "mutation number distribution" $f(k, t)$ emerges in the same way that a fitness distribution $f(X, t)$ arises from Eq. (6). The stochastic dynamics in this case are governed by the Langevin equation

$$\frac{\partial f(k)}{\partial t} = Uf(k-1) - Uf(k) + \sum_{k'} [\delta_{kk'} - f(k)] \sqrt{\frac{f(k')}{N}} \eta(k', t). \tag{8}$$

These dynamics are similar to the charge-ladder model introduced by Ohta and Kimura (1973), which was initially created to model the early electrophoresis measurements of allelic diversity. This model later played an important role in the development of the neutral coalescent, which has since largely superseded it (Kingman, 1976, 1982, 2000; Moran, 1975). Our approach below will have much in common with this standard neutral result, although some quantities are more convenient to calculate in one framework than the other. However, our description in terms of fitness classes will lead to a natural generalization in the presence of selection, which is difficult to incorporate into the standard coalescent model (Neuhauser and Krone, 1997).

The stochastic dynamics in Eq. (8) are free of the nonlinearities that plagued our earlier analysis of Eq. (6), and the resulting equation for the average profile $\langle f(k, t)\rangle$ closes:

$$\frac{\partial \langle f(k)\rangle}{\partial t} = U\langle f(k-1)\rangle - U\langle f(k)\rangle.$$

This differential equation is straightforward to solve, and under the assumption that all individuals start with zero mutations at time $t = 0$, we find that

$$\langle f(k, t)\rangle = \frac{(Ut)^k}{k!} e^{-Ut}. \tag{9}$$

Thus, the mean of this distribution accumulates mutations at a constant rate U, which agrees with the standard calculation that assumes that each neutral mutation fixes independently. As $t \to \infty$, the width of this distribution grows larger and larger, and in order to conserve probability, $\langle f(k, t)\rangle$ approaches the trivial solution

$$\lim_{t\to\infty} \langle f(k, t)\rangle = 0.$$

A similar observation was made previously in the context of the charge-ladder model, which reflects the fact that this earlier model and the one defined by Eq. (8) have no true stationary distribution. Intuitively, this degenerate behavior is an artifact of the averaging process we used in order to calculate $\langle f(k, t)\rangle$. While the average rate of mutation accumulation is simply the mutation rate U, the actual rate for any particular population will tend to fluctuate around this value, and the location $\bar{k}(t)$ will become increasingly uncertain with time. By calculating the

average $\langle f(k,t)\rangle$ as $t \to \infty$, we are effectively averaging many independent distributions whose centers are distributed across a large region of k, and hence the average number of individuals at any particular k tends to zero. This line of reasoning is not specific to the neutral limit considered in this section, but is in fact a general property of any fitness distribution whose absolute location is subject to stochastic fluctuations.

In all of these cases, the average distribution at long times is a poor summary of the typical distribution found in a random population. For example, while the width of the *average distribution* in Eq. (9) increases without bound, a simple argument from neutral coalescent theory shows the *average width* of the population fitness distribution has a finite extent as $t \to \infty$. A random pair of individuals in the population will typically share a common ancestor $T_2 \sim N$ generations ago, so the difference between the number of mutations accumulated since the common ancestor is on the order of NU. We can see this in our current framework by simply measuring the number of mutations in each individual relative to the mean number of mutations in the population at any given time. In particular, we can examine the variance in the number of mutations within the population, which is defined by

$$\sigma_k^2 = \sum_k (k-\bar{k})^2 f(k).$$

Due to the presence of the $\bar{k}$ terms within this definition, σ_k^2 is not just a simple linear function of the class sizes $f(k)$, and the rate of change of the average variance $\langle\sigma_k^2\rangle$ cannot be written as a function of the average class sizes $\langle f(k)\rangle$ alone. Nevertheless, we can use the stochastic dynamics in Eq. (8) to show that the differential equation for $\langle\sigma_k^2\rangle$ does close on *itself*, and we find that

$$\frac{\partial\langle\sigma_k^2\rangle}{\partial t} = U - \frac{\langle\sigma_k^2\rangle}{N}.$$

Again, assuming that all individuals start out with zero mutations at time $t = 0$, this equation yields the simple solution

$$\langle\sigma_k^2(t)\rangle = NU\left(1-e^{-t/N}\right) \approx \begin{cases} Ut & \text{if } t \ll N, \\ NU & \text{if } t \gg N. \end{cases}$$

Thus, we see that the variance attains an equilibrium value $\langle\sigma_k^2\rangle = NU$, as expected from our coalescent arguments, and it does so on the coalescent timescale $T_2 \sim N$. For $t \gg N$, the population continues to accumulate mutations at the same steady-state rate U, but it does so with the relatively constant shape dictated by this mutation–drift balance (see Fig. 2). On the other hand, for $t \ll N$ the average variance is essentially given by the deterministic estimate Ut obtained from Eq. (9). We argued earlier that this average distribution becomes unreliable when the uncertainty in the location of the mean becomes comparable to the width of a typical distribution. Our Langevin framework allows us to make this argument more explicit, since we can directly show that the variance in $\bar{k}$ obeys the differential equation

$$\frac{\partial \mathrm{Var}(\bar{k})}{\partial t} = \frac{1}{N}\langle\sigma_k^2\rangle,$$

and hence

$$\mathrm{Var}(\bar{k}) = Ut - NU\left(1-e^{-t/N}\right) \approx \begin{cases} \frac{Ut}{2}\left(\frac{t}{N}\right) & \text{if } t \ll N, \\ NU\left(\frac{t}{N}\right) & \text{if } t \gg N. \end{cases}$$

Thus, when $t \sim N$, the uncertainty in $\bar{k}$ is on the order of the variance σ_k^2 within a typical population. Beyond this threshold, the uncertainty in $\bar{k}$ starts to dominate the width of the average distribution $\langle f(k,t)\rangle$. On much longer timescales, the distribution of mutations within a typical population will have a relatively tight width, $\langle\sigma_k^2\rangle = NU$, and a mean $\bar{k}$ which increases at rate U with a diffusion constant $D = U/2$ (see Fig. 2).

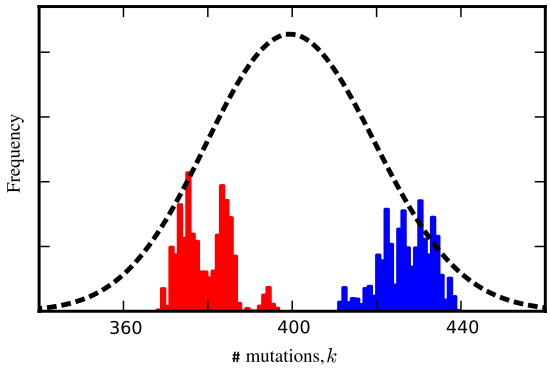

Fig. 2. The distribution of the number of neutral mutations within the population after $t = 8N$ generations when $NU = 50$. The colored bars denote the results of two independent realizations of the stochastic dynamics in Eq. (8), and the dashed line is proportional to the average profile $\langle f_k(t)\rangle$ from Eq. (9). [The vertical scale has been adjusted to improve visibility; by definition all three distributions have unit area.].

3.1. Higher moments and correlations

Of course, the substitution rate and the steady-state variance $\langle\sigma_k^2\rangle$ can be calculated by other means, without the need for the complicated machinery of the Langevin equation in Eq. (8). The real utility of this approach is that it allows us to calculate higher moments of the distribution of mutations that are inaccessible by these other methods. Motivated by our discussion of the variance in mutation number, we consider the whole family of central moments

$$M_m = \left\langle \sum_k (k-\bar{k})^m f(k) \right\rangle,$$

where by definition $M_2 = \langle\sigma_k^2\rangle$ from before. For $m > 2$, these moments measure the higher-order features of the distribution $f(k)$ after controlling for the uncertainty in the mean. Using the stochastic dynamics in Eq. (8) and rescaling time by $\tau = t/N$, it is straightforward to show that the central moments obey the compact equation

$$\frac{\partial M_m}{\partial\tau} = \binom{m}{2}\left\langle \sum_k (k-\bar{k})^2 f(k) \cdot \sum_k (k-\bar{k})^{m-2} f(k)\right\rangle - mM_m + NU\sum_{\ell=0}^{m-2}\binom{m}{\ell} M_\ell, \tag{10}$$

which we derive in Appendix C. Given the nonlinear term on the right, this system of equations for M_m does not close for $m > 3$ unless we also include more complicated "replica" moments of the form

$$M_{m_1,\ldots,m_J} = \left\langle \prod_{j=1}^{J}\left[\sum_k (k-\bar{k})^{m_j} f(k,t)\right]\right\rangle, \tag{11}$$

where $M_{m,n} \neq M_m \cdot M_n$. A general procedure for obtaining the equations governing these replica moments is outlined in Appendix C. At lower orders these dynamics are relatively simple, and were first analyzed by Higgs and Woodcock (1995). In our present notation, they showed that

$$\frac{\partial M_2}{\partial\tau} = NU - M_2, \tag{12a}$$

$$\frac{\partial M_3}{\partial\tau} = NU - 3M_3, \tag{12b}$$

$$\frac{\partial M_4}{\partial\tau} = NU + 6NUM_2 + 6M_{2,2} - 4M_4, \tag{12c}$$

$$\frac{\partial M_{2,2}}{\partial\tau} = 2NUM_2 - 3M_{2,2} + M_4, \tag{12d}$$

although one can technically include the fifth order moments before triple products of the form M_{m_1,m_2,m_3} start to appear. This system of first-order linear differential equations can be solved with standard Laplace transform methods in order to obtain a full time-dependent solution. However, if we are merely interested in the steady-state behavior as $t \to \infty$, then the time derivatives on the left-hand side vanish and the resulting *algebraic* system can be easily solved:

$$M_2 = NU, \qquad M_4 = \frac{10(NU)^2 + NU}{2},$$
$$M_3 = \frac{NU}{3}, \qquad M_{2,2} = \frac{14(NU)^2 + NU}{6}. \tag{13}$$

In addition to these central moments M_m, we could also try to describe the shape of $f(k)$ through its cumulants C_m, which are defined by

$$C_m = \left\langle \frac{\partial_z^m \log\left(\sum_k e^{zk} f(k)\right)}{\partial z^m} \right\rangle \Bigg|_{z=0}.$$

Note that by definition, $C_2 = M_2$ and $C_3 = M_3$, while $C_4 = M_4 - 3M_{2,2}$. Although the definition of a cumulant is less intuitive than its corresponding central moment, many statistical distributions can be described more compactly in terms of their cumulants.[2] Moreover, given that C_4 already contains one of the replica products, $M_{2,2}$, it could be the case that the replica moments in Eq. (10) only arose because we did not write everything in terms of the cumulants from the beginning. We can check this hypothesis by deriving an analogous set of equations for the cumulants in Appendix D, which shows that

$$\frac{\partial C_m}{\partial \tau} = NU - \sum_{\vec{m}} \binom{m}{\vec{m}} \frac{\prod_j (2^{m_j} - 1)}{\dim \vec{m}!} C_{\vec{m}}, \tag{14}$$

where the sum over $\vec{m}$ denotes a sum over all integer partitions of m. Thus, we can see that while the mutation term is much simpler for the cumulants, replica products are required in this case as well, and they appear in a much more complicated form than they do for the central moments M_m. In this respect, both the cumulants and the central moments provide a rather unnatural basis for describing the shape of $f(k)$ when genetic drift is taken into account. At higher order in m, genetic drift starts to mix in contributions from the replica products, which only become more numerous as m increases.

Yet while these replica products make calculations of M_m and C_m more onerous, they provide important information about the variability of these quantities between independent populations. For example, the second order product $M_{2,2}$ can be used to calculate the variance in the characteristic width σ_k^2 from the relation

$$\mathrm{Var}(\sigma_k^2) = M_{2,2} - M_2^2 = \frac{8(NU)^2 + NU}{6}.$$

This shows that the long-term uncertainty in σ_k^2 remains larger than its expected value $\langle \sigma_k^2 \rangle$ even in the limit that $NU \to \infty$. Thus, as argued by Higgs and Woodcock (1995), the variance in the number of mutations within the population is not a self-averaging quantity, since σ_k^2 does not "settle-down" to some fixed value in large populations as one might naively expect if the individuals were completely independent. Instead, the genealogical relationships between individuals create large fluctuations in the mutational diversity within the population. The typical lifetime of these fluctuations can be measured from the autocorrelation function

$$\lim_{\tau \to \infty} \left[\langle \sigma_k^2(\tau) \sigma_k^2(\tau + \Delta\tau) \rangle - \langle \sigma_k^2(\tau) \rangle \langle \sigma_k^2(\tau + \Delta\tau) \rangle \right]$$
$$= \mathrm{Var}(\sigma_k^2) e^{-\Delta\tau},$$

which implies that these correlations decay in a simple manner on the coalescent timescale $T_2 \sim N$.

Continuing the system in Eqs. (12a)–(12d) to central moments with $m > 5$ starts to become complicated, since the moment equations for M_m start to involve more and more of the replica moments in Eq. (11). Nevertheless, the algebraic structure of these moment equations is sufficiently simple that the exact solution can be found with the help of a computer. The most important property of these equations is already apparent in Eqs. (12a)–(12d): replica products $M_{\vec{m}}$ with total order $\sum_j m_j = m$ only depend on other products with order $\leq m$. Thus, at any given order we have a finite system of linear equations that determine the values of the moments, and this immediately suggests a simple iterative algorithm for extracting these solutions. Given values for the replica moments at order $\leq m$, we can calculate the moments at order $m + 1$ by solving the matrix equation

$$\mathbf{A}_{m+1} \cdot \vec{M}_{m+1} = \vec{b}(NU, \vec{M}_1, \ldots, \vec{M}_m). \tag{15}$$

Here, $\vec{M}_m$ is the collection of all replica moments with total order $\sum_j m_j = m$, $\mathbf{A}_m$ is a matrix of constants (independent of NU or any of the moments), and $\vec{b}$ is a vector-valued function of NU and the lower-order moments. The entries of the matrix $\mathbf{A}_m$ can be determined directly by inspection from the system of equations in Appendix C. The resulting matrix must only be inverted once for each m, and then the analytical solutions for the various moments can be obtained by simple matrix multiplication. An implementation of this iterative algorithm is available from the authors upon request.

One can in principle use this algorithm to calculate the moments for arbitrary m, which will be in the form of some polynomial in NU similar to what we found for the first few moments in Eq. (13). Typically, we will be interested in the limiting behavior for large NU, which we can access most easily by defining the rescaled moments $\tilde{M}_{\vec{m}} = M_{\vec{m}}/(\sqrt{NU})^m$. In the limit that $NU \to \infty$, the equations for the rescaled moments become independent of NU, so $\tilde{M}_m$ can only depend on the order m. These rescaled moments can be calculated from the same iterative scheme outlined above, and the results for the first thirty moments are shown in Fig. 3. For large m, these moments obey the approximate scaling relation

$$M_m \sim \left(\frac{m!}{m^p}\right) (NU)^{m/2}$$

where the exponent $p \approx 2$ can be extracted from the plot in Fig. 3. Although these central moments grow more quickly than the corresponding central moments of a Gaussian distribution, they grow sufficiently slowly that the centered distribution of mutations remains a light-tailed distribution. This is in contrast to the analogous neutral limit for Fisher–KPP waves, where Hallatschek and Korolev (2009) showed that the propagating front displays a power-law shape due to the periodic formation of smaller waves at the tip. However, many of these technical details are beyond the scope of the present paper. The main point of this discussion for $m \geq 4$ is simply that *all* of the central moments of this neutral distribution are *exactly* solvable, as long as one is willing to devote the time and computing power necessary to implement the iterative scheme described above.

[2] For example, a Poisson distribution is uniquely characterized by the fact that $C_m = \lambda$ for all m, while a normal distribution has $C_m = 0$ for all $m > 2$.

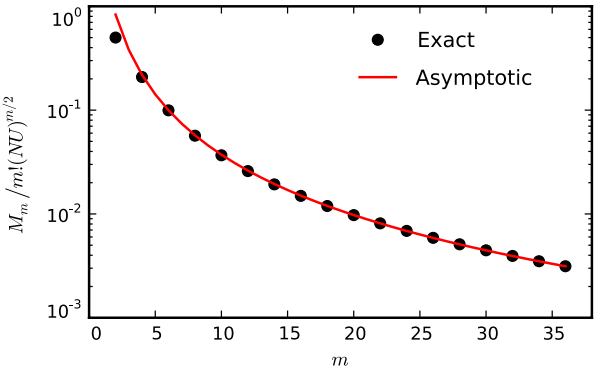

Fig. 3. The central moments M_m as a function of m in the limit that $NU \to \infty$. Symbols denote the exact numerical results calculated using the iterative scheme outlined in the text for $m = 2, \ldots, 30$. The red line denotes the approximate scaling form, $M_m \sim m!(NU)^{m/2}/m^p$ where $p = 1.93$.

4. Perturbation theory for selected mutations

While certain properties of $f(k)$ may be interesting in their own right, previously developed methods like the neutral coalescent usually offer a simpler and more direct way to quantify the genetic diversity at the sequence level. The true utility of this neutral "fitness distribution" is that unlike many of these earlier coalescent approaches, it can be easily generalized to calculate the corrections that arise when selection is present. As an example, suppose the neutral mutations in the previous section now have a constant fitness effect s. Selection on these mutations leads to an additional term $+s(k - \bar{k})$ on the right-hand side of the stochastic dynamics in Eq. (8). Under these modified dynamics, the substitution rate $R = d\langle k\rangle/dt$ for these mutations is now given by

$$R = U + s \cdot C_2,$$

which now depends on the variance in k in addition to the neutral accumulation rate U. The variance C_2 will in turn depend upon the skew C_3, and so on in the infinite hierarchy of moment equations mentioned earlier. However, if the strength of selection is weak and $Ns \ll 1$, the contribution to the variance from C_3 will be small, and the variance will be approximately equal to the neutral result $C_2 = NU$. Thus, by neglecting the selection term in the calculation of C_2, we obtain an approximate expression for the substitution rate,

$$R \approx U\,(1 + Ns)\,,$$

first introduced by Higgs and Woodcock (1995) using a similar argument. This expression is valid in the $Ns \to 0$ limit, where the second term is a small *perturbative* correction to the neutral result $R = U$. In this way, the exact solution for the neutral wave can therefore serve as a basis for a perturbative analysis of the effects of selection in increasing powers of Ns.

However, in order to apply this perturbative scheme to populations with more general distributions of fitness effects [as described in Eq. (6)], we must define the neutral limit a bit more precisely. In our analysis above, it was most natural to divide the population into fitness classes according to the discrete number of mutations k in each individual. For more general distributions of fitness effects, it will be most natural to introduce a typical selection strength s and a continuous variable k such that the fitness is given by the product $X = sk$. We can therefore rewrite $\rho(s)$ as a distribution of "k-effects", $\rho(\Delta k)$, so that we can hold the shape of the distribution constant as we modulate the overall strength of selection s. With these definitions, it is straightforward to show that the average rate of fitness change $v = d\langle \bar{X}\rangle/dt$ is given by

$$v = Us\left[\langle \Delta k\rangle + \frac{Ns}{NU}\cdot C_2\right], \tag{16}$$

while the equation for the cumulants in Eq. (14) becomes

$$\frac{\partial C_m}{\partial \tau} = Ns \cdot C_{m+1} + NU\langle \Delta k^m\rangle - \sum_{\vec{m}} \binom{m}{\vec{m}} \frac{\prod_j (2^{m_j} - 1)}{\dim \vec{m}!} C_{\vec{m}}, \tag{17}$$

where $\langle \Delta k^m\rangle = \int d(\Delta k)\,(\Delta k)^m \rho(\Delta k)$ denotes mth moment of the distribution of fitness effects. Analogous (but slightly more complicated) equations for the central moments and the corresponding replica products are derived in Appendix C. These equations all share a common feature with the general expression in Eq. (17): the moments at order m now include terms that depend on the moments at order $m + 1$. The resulting system of equations cannot be solved at any fixed order because it always depends on the moments at a still-higher order.

While this lack of closure among the moment equations makes it difficult to obtain a closed-form solution for any particular moment, it naturally suggests a perturbative approach similar to the $R \approx U(1 + Ns)$ approximation above. In particular, we assume that in the limit $Ns \to 0$, each of the central moments M_m admits an asymptotic expansion of the form

$$M_m \sim \sum_{j=0}^{\infty} M_m^{(j)} (Ns)^j \quad (Ns \to 0), \tag{18}$$

where $M_m^{(j)}$ is a numerical coefficient that depends only on NU. A similar expansion is assumed for the replica products in Eq. (11), as well as the cumulants C_m and their corresponding replica products. After substituting these expressions into the moment equations and grouping terms in powers of Ns, we can obtain a generalized system of equations for the coefficients $M_m^{(j)}$. The important feature of these equations is that unlike the case for the full moments M_m, the equations for the coefficients $M_m^{(j)}$ close for a given order j and m. The algebraic properties of these equations are again sufficiently simple that they only lead to a slightly more complicated version of Eq. (15):

$$\mathbf{A}_{m+1} \cdot \vec{M}_{m+1}^{(j)} = \vec{b}\left(NU, \vec{M}_1^{(j)}, \ldots, \vec{M}_m^{(j)}\right) + \vec{c}\left(\vec{M}_{m+1}^{(j-1)}\right). \tag{19}$$

Here, $\mathbf{A}$ and $\vec{b}$ are the same as in the neutral case, and $\vec{c}$ is a vector-valued function of the moments at the next-lowest order in j. Each of these functions $\mathbf{A}$, $\vec{b}$, and $\vec{c}$ is the same for all j. Thus, the iterative procedure outlined for the neutral case can be easily generalized to calculate the coefficients $\vec{M}_m^{(j)}$ order-by-order for arbitrary m and j. An implementation of this algorithm is available from the authors upon request.

As an alternative to this explicit order-by-order calculation, we note that the coefficients in the asymptotic expansion in Eq. (18) are unique (Hinch, 1991), so we can also obtain these coefficients simply by dropping the selection term in the moment hierarchy at the desired order, solving the resulting finite system of equations, and then reexpanding the solution in powers of Ns. Applying this procedure to the first few orders of Eq. (17) (i.e., the selected versions of Eqs. (12a)–(12d) listed in Appendix C), we obtain the first few corrections to the population variance in fitness

$$M_2 = NU\langle \Delta k^2\rangle + \left(\frac{NU\langle \Delta k^3\rangle}{3}\right) Ns - \left(\frac{2[NU\langle \Delta k^2\rangle]^2}{3}\right)(Ns)^2 + O(Ns)^3, \tag{20}$$

and the corresponding rate of fitness change

$$v = Us\left[\langle \Delta k\rangle + Ns\langle \Delta k^2\rangle + \frac{(Ns)^2\langle \Delta k^3\rangle}{3} - \frac{2[NU\langle \Delta k^2\rangle (Ns)^2][Ns\langle \Delta k^2\rangle]}{3}\right] + O(Ns)^4. \tag{21}$$

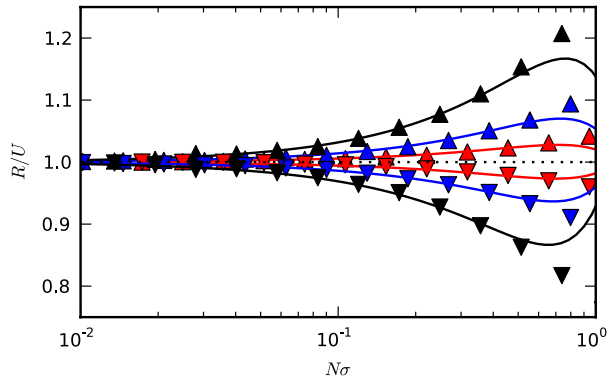

Fig. 4. The scaled substitution rate R as a function of the zeroth-order fitness variance $(N\sigma)^2 = NU(Ns)^2$. Symbols denote the results of forward-time simulations for $NU = 10$ (black), $NU = 50$ (blue), and $NU = 300$ (red), and the solid lines give the predictions from Eq. (21). Upper triangles denote populations with a purely beneficial distribution of fitness effects $\rho(\Delta k) = \delta(\Delta k - 1)$, while the lower triangles give the corresponding deleterious distribution $\rho(\Delta k) = \delta(\Delta k + 1)$. Our predictions start to diverge near $N\sigma = 1$, when we expect our perturbation expansion to break down. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The first three terms in this expansion are exactly what one would obtain by assuming that the individual sites evolve independently, in which case the rate of fitness increase would simply be the sum of the single-locus adaptation rates that can be calculated from Eq. (2). The fourth term in this expansion represents a fundamentally new correction that arises solely from the accumulation of selected mutations over many different sites. This term is proportional to the variance in fitness within the population, so we see that the rate of adaptation is reduced in populations with a larger number of selected mutations (see Fig. 4) because many of these mutations will be lost to clonal interference before they can fix. Similarly, in populations which are accumulating deleterious mutations due to Muller's ratchet, this interference term leads to an increase in the rate of Muller's ratchet (again, see Fig. 4) due to the increased importance of fluctuations in the high-fitness nose of the population.

In addition to the mean rate of adaptation, we can also use this perturbative scheme to calculate the *fluctuations* in the rate of adaptation, which have so far been accessible only through heuristic arguments (Brunet et al., 2006; Desai and Fisher, 2007; Fisher, 2013; Hallatschek, 2011). Using the Langevin dynamics in Eq. (6) and the rules of our stochastic calculus, we can construct similar moment hierarchy for $\mathrm{Var}(\bar{k})$ and its relatives, the first few orders of which are given in Appendix E. At long times, we find that the mean fitness continues to fluctuate diffusively (i.e., $\mathrm{Var}(\bar{k}) \sim 2Dt$) with a diffusion constant

$$D = \frac{U\langle \Delta k^2\rangle}{2}\left[1 + \frac{\langle \Delta k^3\rangle}{\langle \Delta k^2\rangle}Ns - \frac{2NU\langle \Delta k^2\rangle (Ns)^2}{3} + \frac{NU\langle \Delta k^4\rangle (Ns)^2}{3}\right] + O(Ns)^3. \tag{22}$$

Again, we see that interference between the various lineages leads to a reduction in the diffusivity of the mean fitness that is proportional to the variance in fitness within the population. This leads to a novel prediction in the case of the dynamic mutation–selection balance discussed in Goyal et al. (2012), where a balance between beneficial and deleterious substitutions halts any global fitness change in the population. The results in Eq. (22) show that even though the average fitness change in these populations is zero, we still expect the distribution to wander diffusively around this fixed point, with diffusion constant

$$D(\epsilon_c) = \frac{U}{2}\left[1 + Ns(1 - 2\epsilon_c) - \frac{NU(Ns)^2}{3}\right] + O(Ns)^3,$$

where ϵ_c is the critical ratio of beneficial to deleterious mutations. This scenario is discussed in more detail in Appendix A.

5. Application to sequence evolution

Our analysis so far has focused on population-wide properties of the fitness distribution and important aspects of the evolutionary dynamics such as the rate of adaptation. In the introduction however, we were primarily interested in predicting evolutionary fates and diversity at the sequence level, which we have so far neglected. In the present section, we will demonstrate how this fitness-class description can also provide a window into the evolutionary dynamics of sequences *within* a particular population.

5.1. Fate of a focal lineage

As an example, we consider the fate of some clonal lineage within the population that has an initial frequency p and fitness $X_0 = sk_0$ at time $t = 0$. This lineage could consist of a group of individuals that share a common point mutation at a particular site, or it could alternatively represent some fluorescently labeled "marker population" whose dynamics we wish to follow. Of course, if subsequent mutations could be neglected then the fate of this lineage could be predicted using the ordinary single-locus model in Eq. (1). What makes this problem so difficult in general is that we must account for additional mutations in the focal lineage as well as those in the background population. Each of these two "subpopulations" can develop substantial variation in fitness not unlike the fluctuating distributions analyzed in the previous section. Now in addition to tracking the fitness of each individual in the population, we must also keep track of whether these individuals are descended from this particular lineage or from the background population. We can accomplish this by dividing the fitness classes $f(X,t)$ into two classes $f_1(X,t)$ and $f_0(X,t)$ which contain individuals descended from the focal lineage and the background population respectively. This division requires a small modification to our original dynamics in Eq. (6):

$$\frac{\partial f_i(X)}{\partial t} = (X - \overline{X})f_i(X) + U\int ds\,\rho(s)\,[f_i(X - s) - f_i(X)] + \sum_{j,X'}\left[\delta_{ij}\delta_{X,X'} - f_i(X)\right]\sqrt{\frac{f_j(X')}{N}}\eta_j(X'). \tag{23}$$

One can verify that these dynamics reduce to the classic model in Eq. (1) if further fitness-altering mutations can be neglected in both lineages. With these new dynamics, we can introduce a new set of "focal" moments

$$F_m = \left\langle \int dk\,(k - \bar{k})^m f_1(k,t)\right\rangle,$$

that are specific to the focal lineage, while the population-wide central moments M_m can be defined the same way as before. The lowest focal moment $F_0(t)$ represents the average total fraction of the population that is descended from the focal lineage (e.g., the frequency of a particular SNP within the population). At long times, one of two things can happen: either the focal lineage is outcompeted by the background and $\int dk f_1(k,t) = 0$, or its descendants take over the population and $\int dk f_1(k,t) = 1$. Thus, the probability of fixation is given by

$$p_{\mathrm{fix}} = \lim_{t\to\infty} F_0(t). \tag{24}$$

Using the dynamics in Eq. (23) and the rules of our stochastic calculus, we obtain an additional moment hierarchy for the focal

moments F_m. As we show in Appendix F, this again requires us to consider generalized replica products of the form

$$F_{\vec{m};\vec{n}} = \left\langle \prod_{j=1}^{J_m} \left[\int dk\, (k-\bar{k})^{m_j} f_1(k,t) \right] \times \prod_{j=1}^{J_n} \left[\sum_i \int dk\, (k-\bar{k})^{n_j} f_i(k,t) \right] \right\rangle \quad (25)$$

where the m indices denote moments specific to the focal lineage and the n indices denote the population-wide moments discussed earlier. The first few orders of the moment hierarchy for F_0 are listed in Appendix F. We are interested in the limiting behavior at long times, and this is most easily obtained via the Laplace transformed moments $\tilde{F}_{\vec{m};\vec{n}}(z) = \int e^{-z\tau} F_{\vec{m};\vec{n}}(\tau)\, d\tau$, which satisfy the asymptotic relation

$$F_0(\tau) \sim F_0(0) + Ns \cdot \tilde{F}_1(1/\tau), \quad (\tau \to \infty).$$

After transforming the first few equations in the moment hierarchy and truncating the selection terms at order $O(Ns)^4$, we find that

$$p_{\text{fix}} = F_0(0) + NsF_1(0) + \frac{(Ns)^2}{3}\left[F_2(0) - F_{0;2}(0)\right] - \frac{(Ns)^3}{3}\left[F_{1;2}(0) + NU\langle \Delta k^2\rangle\right] + O(Ns)^4.$$

For a lineage created by a spontaneous mutation, the initial frequency is just $p = 1/N$, so the initial conditions for the various moments are given by

$$F_m(0) = \frac{1}{N}\left[k_0 - (\bar{k})_{\text{bg}}\right]^m + O(N^{-2}),$$
$$M_m(0) = (M_m)_{\text{bg}} + O(N^{-1}).$$

Here $(\bar{k})_{\text{bg}}$ and $(M_m)_{\text{bg}}$ denote the mean fitness and the central moments of the background population, whose averages will coincide with the steady-state moments derived in the previous section. The initial fitness $X_0 = sk_0$ for the mutant lineage is obtained as a random draw from the background fitness distribution $f_0(X)$, plus the fitness effect $s\Delta k$ of the mutation itself. Averaging over the possible fitness backgrounds that this mutation could have arisen on, we find that

$$F_m(0) = \frac{1}{N}\sum_{\ell=0}^{m} \binom{m}{\ell} (\Delta k)^\ell M_{m-\ell} + O(N^{-2}),$$
$$M_m(0) = M_m + O(N^{-1}).$$

The fixation probability for a spontaneous mutation with effect Δk is therefore given by

$$p_{\text{fix}}(\Delta k) = \frac{1}{N}\left[1 + Ns\Delta k + \frac{(Ns\Delta k)^2}{3} - \frac{2[NU\langle \Delta k^2\rangle (Ns)^2][Ns\Delta k]}{3}\right] + O(Ns)^4. \quad (26)$$

Again, we see that the first three terms are identical to the single-locus result in Eq. (2). We obtain the lowest-order "interference correction" in the fourth term, which reduces the probability of fixation of a beneficial mutation in a way that is directly proportional to the average variance in fitness within the population at the time of the mutation. For a deleterious mutation, this correction term actually increases the fixation probability because the mutant could find itself on an anomalously fit background (thus mitigating some of the effect of the deleterious mutation). We recover the standard neutral fixation probability $p_{\text{fix}} = 1/N$ when $\Delta k = 0$.

Noting the similarity between the fixation probability in Eq. (26) and the rate of adaptation in Eq. (21), we see that the relation

$$v = \int NU \cdot s\Delta k \cdot p_{\text{fix}}(\Delta k) \cdot \rho(\Delta k)\, d(\Delta k)$$

holds at least through the first few orders in Ns. This relation has formed the basis for several studies of the evolutionary dynamics under strong selection (Good et al., 2012; Hallatschek, 2011; Neher et al., 2010), with the additional "mean-field" ansatz

$$p_{\text{fix}}(\Delta k) \approx \int dx\, \langle f(x - s\Delta k)\rangle \cdot p_{\text{fix}}(x|\langle f(x)\rangle). \quad (27)$$

Here, $p_{\text{fix}}(x|\langle f(x)\rangle)$ denotes the fixation probability for a new mutation with relative fitness x, given that the centered fitness distribution of the rest of the population is $\langle f(x)\rangle$. However, we see that in the present regime, this mean-field ansatz is not quite correct. Instead, we require the slightly more complicated average

$$p_{\text{fix}}(\Delta k) \approx \int dx\, \langle f(x - s\Delta k) \cdot p_{\text{fix}}(x|f(x))\rangle, \quad (28)$$

which jointly considers the fluctuations in the fitness background of the mutant as well as the fluctuations in the fitnesses of its competing lineages. Like the other correlated quantities we have considered in the present work, this average more or less decouples in the strong selection limit $Ns \to \infty$, and we recover the "mean-field" ansatz in Eq. (27). But in the weak selection regime considered here, these correlated fluctuations start to become more important, and Eq. (28) is required in order to correctly account for the population-level dynamics.

5.2. Diversity at a focal site

In addition to predicting the ultimate fate of a sequence, these focal lineage dynamics can also be used to predict the average heterozygosity at a particular site along the genome and therefore the overall levels of sequence diversity in the population. We consider a particular site within the genome with a per-site mutation rate μ and scaled fitness effect Δk. This site will be polymorphic in a randomly sampled pair of individuals if and only if (1) this site mutated at some time t in the past and (2) exactly one member of the pair was drawn from the mutant lineage, which has size $\int dk f_1(k,t)$ in the present. After averaging overall possible mutation times (and taking note of the fact that the backward-time mutation process is Poisson), we find that

$$\pi = \left\langle \int_0^\infty dt\, N\mu e^{-N\mu t} \times 2\left(\int dk f_1(k)\right)\left(1 - \int dk f_1(k)\right)\right\rangle.$$

In the infinite-sites limit where $N\mu \to 0$, this yields the well-known relation

$$\pi = 2N\mu \int_0^\infty NH(\tau) d\tau,$$

where we have defined the heterozygosity function

$$H(\tau) = \left\langle \left(\int dk f_1(k,t)\right)\left(1 - \int dk f_1(k,t)\right)\right\rangle = F_0(\tau) - F_{0,0}(\tau).$$

Again, we can use the dynamics in Eq. (23) to construct a similar hierarchy of moment equations for the heterozygosity, the first few orders of which are listed in Appendix F. Truncating the hierarchy and solving the Laplace transformed equations, we find that

$$\pi = 2N\mu\left[1 + \frac{Ns\Delta k}{3} - \frac{2NU\langle \Delta k^2\rangle (Ns)^2}{9}\right] + O(Ns)^3. \quad (29)$$

Again, the first two terms in this expansion are equivalent to the single-locus results in Eq. (3). At second order in Ns, we obtain the lowest-order correction due to interference at neighboring sites, which reduces the diversity at a particular site according to the variance in fitness within the population. This reduction in diversity is seen even at putatively neutral or synonymous sites that are not otherwise selected on their own, and this helps explain why the interference corrections enter at a lower order here than they did for p_{fix} and v above. It is clear that the lowest order interference corrections must couple to the variance in fitness within the population, which is already $O(Ns)^2$. Interference corrections for p_{fix} require an additional factor of Ns arising from the fitness effect of the mutation itself, since it is required that these corrections vanish when the mutation itself is neutral. In contrast, no such requirement is imposed on the diversity at neutral sites, which explains the stronger interference effect observed in Eq. (29).

6. Discussion

Although natural selection acts on the genome as a whole, the effects of selection at a large number of linked sites are only beginning to be characterized. Recent studies have identified the distribution of fitnesses within the population as a key mediator for these effects, but our understanding of this distribution remains limited to a few special cases where the strength of selection is strong and genetic drift is correspondingly weak. Here, we have introduced a general method for analyzing the effects of selection at many linked loci, which incorporates linkage and drift exactly while treating the global strength of selection as a perturbative correction. This framework allows us to investigate the stochastic behavior of the fitness distribution in a regime where the fluctuations due to drift are especially strong, and it fills an important gap in our theoretical understanding of linked selection in the approach to the neutral limit.

This perturbative approach, which builds on earlier work by Higgs and Woodcock (1995), bears a superficial resemblance to a number of other studies of weak selection that also utilize series expansions in powers of Ns. Chief among these is Kimura's classic solution for the time-dependent diffusion equation in Eq. (1), although other examples arise in the quantitative genetics literature (Bürger, 1991; Nagylaki, 1993; Turelli and Barton, 1990). However, these previous studies differ from ours in a crucial way, since they assume that genetic linkage between sites plays a minor or nonexistent role. Kimura's solution concerns the frequency of a selected allele at a single genetic locus, and it is limited to weak selection only because of the difficulty in obtaining the full probability distribution of this frequency. However, if one is only interested in derived quantities like the fixation probability or the average diversity, the simpler formulas in Eqs. (2) and (3) are valid for the entire range of selection strengths. Similarly, weak selection analyses in the quantitative genetics literature typically assume that linkage is weak compared to the strength of selection, so that the various sites in the genome evolve more or less independently. In contrast, our analysis here focuses on a regime where interference between selected mutations is explicit, and this allows us to investigate many of the effects of linked selection that are absent from these earlier models.

The evolution of the fitness distribution described by Eq. (17) is also reminiscent of the *deterministic* cumulant equations,

$$\frac{\partial C_m}{\partial t} = s \cdot C_{m+1} + U\langle s^m \rangle,$$

which are often used to describe the dynamics of "infinitely large" or otherwise strongly-selected populations (Bürger, 1991; Desai and Fisher, 2011; Goyal et al., 2012; Rouzine et al., 2008). Our present analysis shows that in the presence of genetic drift, these deterministic dynamics are modified according to

$$\frac{\partial C_m}{\partial t} = s \cdot C_{m+1} + U\langle s^m \rangle + \frac{\sum \text{replica products}}{N},$$

so that C_m picks up contributions from all other replica products with an overall constant of proportionality set by N^{-1}. The only moment that remains unmodified is the mean fitness $\langle \overline{X} \rangle$, which continues to evolve according to Fisher's fundamental theorem of natural selection, $\partial_t \langle \overline{X} \rangle = U\langle s \rangle + s^2 \cdot C_2$, independent of the relative strengths of drift and selection. Our present analysis makes it clear that the robustness of this classic result arises from the linear nature of $\langle \overline{X}(t) \rangle$. Higher moments of the fitness distribution involve multiple factors of $f(X, t)$ which are more sensitive to the fluctuations induced by drift or rare mutational events.

As a quantitative theory, the present framework suffers from several shortcomings that may limit its direct applicability to data from natural populations. Our perturbative approach gives predictions for various quantities in terms of an asymptotic series in the limit that $Ns \to 0$, which means that for a fixed number of terms in this series, the resulting formulas will only become valid once Ns is sufficiently small. Moreover, these asymptotic expressions are *nonuniform* as a function of the mutation rate NU, and in general for larger mutation rates we require ever smaller values of Ns for our expressions to remain accurate. In reality, these asymptotic series could be more accurately described as an expansion in powers of the typical fitness variance $N\sigma \approx Ns\sqrt{NU}$, valid in the limit that $N\sigma \lesssim 1$.

It remains an open question exactly what values of $N\sigma$ are relevant for natural populations. Indeed, one of the major motivating factors behind this quantitative approach to linked selection is to eventually use these theoretical tools to infer $N\sigma$ directly from DNA polymorphism data in sampled populations. Because the vast majority of new mutations are thought to be either neutral or weakly deleterious, there has been speculation that evolution at the sequence level is dominated by these "nearly-neutral" mutations with $Ns \lesssim 1$ (Ohta, 1992), although it is unclear whether these selection coefficients lead to an $N\sigma$ that is sufficiently small for our results to apply. In principle, the range of applicability of our expressions can be improved by including more terms in the expansion, but there is typically an upper limit to the radius of convergence that can be achieved this way (Hinch, 1991). However, because we have outlined a method for calculating successively higher-order terms programmatically, series improvement methods could potentially be used to extend the radius of convergence, even for $N\sigma > 1$ (Song and Steinrücken, 2012; Van Dyke, 1974). This constitutes an interesting direction for future work.

While the experimental applicability of these perturbation methods may be limited, they nevertheless provide a valuable *qualitative* window into the effects of selection at many linked sites, and the exact nature of the selective corrections allows us to address a number of longstanding assumptions in the population genetics literature. Chief among these is the independent-sites assumption that is frequently used to model selection at individual sites along the genome. Somewhat surprisingly, we have demonstrated here that this assumption is valid not only in the purely neutral case, but also frequently through the first-order selective correction. At higher-orders, however, we start to obtain terms that depend on NU, and more generally, the variance in fitness maintained within the population. These terms represent corrections that arise solely from the interactions between the selected sites, and cannot be predicted from any single-locus theory.

Of course, the standard assumption is not that linked sites evolve in this strictly independent fashion, but that they evolve

independently at a reduced *effective population size* N_e, which is supposed to encapsulate the effects of selection at neighboring sites (Hill and Robertson, 1966). However, several recent studies have begun to challenge this assumption (Comeron and Kreitman, 2002; Good and Desai, 2012; Santiago and Caballero, 1998), often on the grounds that a different N_e must be defined for every quantity we wish to predict. This shortcoming is apparent from our present analysis as well, and our analytical corrections provide an explicit demonstration.

The effective population size is most commonly measured from the diversity at putatively neutral or synonymous sites. To lowest order in Ns, our analysis of the pairwise heterozygosity yields a corresponding effective population size

$$N_e^{\pi} = N\left[1 - \frac{2NU(Ns)^2}{9}\right], \tag{30}$$

which is reduced in the presence of selection as expected. To lowest order, this same N_e correctly predicts the reduction of heterozygosity at selected sites as well. Alternatively, we could define N_e by measuring the divergence (i.e., frequency of nucleotide substitutions) at various sites under selection, which depends on the fixation probability of a new mutant. To lowest order in Ns, this yields an effective population size

$$N_e^{p_{\rm fix}} = N\left[1 - \frac{2NU(Ns)^2}{3}\right], \tag{31}$$

which is also reduced by selection at linked sites, but at slightly faster rate than in Eq. (30). Thus, we require a different N_e to account for linkage depending upon whether we wish to predict π, $p_{\rm fix}$, or some other sequence-based statistic. While the effective population size can still be used in the technical sense on a per-quantity basis, these results imply that its predictive or explanatory power is greatly reduced, and that the effects of linked selection are more complicated than a simple increase in genetic drift would suggest.

In this way, the selective corrections obtained here can be extremely useful from a model-building standpoint, even when we wish to ultimately apply these models in regions where the perturbative approach breaks down. These corrections are straightforward to calculate for any quantity with a well-defined neutral limit, and because they are exact, any other model describing weakly selected mutations should recover these expressions as $Ns \to 0$. Many aspects of natural selection at the sequence level remain poorly understood, and exact results are few and far between. It is our hope that the methods outlined in the present work can be used as a stepping-stone to identify and evaluate those approximations which will lead to further progress on this important problem.

Acknowledgments

We thank Richard Neher for useful discussions. This work was supported in part by the James S. McDonnell Foundation, the Alfred P. Sloan Foundation, and the Harvard Milton Fund. B.H.G. acknowledges support from a National Science Foundation Graduate Research Fellowship. Simulations in this paper were performed on the Odyssey cluster supported by the Research Computing Group at Harvard University.

Appendix A. Finite site effects

In the main text, we worked exclusively in the large-genome limit, where the number of sites was so large (and the per-site mutation rate so low) that we could focus on an intermediate asymptotic regime where back mutations could be neglected, and the mutation rate and distribution of fitness effects was independent of the previous mutations within a particular genome. In the present section we now consider what happens when we start to relax these assumptions. In particular, we assume that the genome has some finite size L and that the per-site mutation rates are now sufficiently large that the scaled product $N\mu$ at each site is finite. For simplicity, we also assume a constant fitness effect for mutations at each site. This is similar to the model analyzed in Woodcock and Higgs (1996) and Rouzine et al. (2003).

In this case, the population can still be partitioned according to the number of mutations $k = 0, \ldots, L$ in each individual, but now we must account for the fact that the distribution of k-effects depends on k in addition to the fitness. An individual with k mutations can mutate at another site at rate $\mu(L-k)$, in which case $k \to k+1$. This individual can also experience a back-mutation at one of its k mutated sites at rate μk, in which case $k \to k-1$. The total rate for one of these two events to happen is of course just $U = \mu L$. Thus, the fitness classes $f(k)$ evolve according to the stochastic dynamics

$$\begin{aligned}\frac{\partial f(k)}{\partial \tau} &= Ns(k-\bar{k})f(k) + N\mu(L-k+1)f(k-1) \\ &\quad + N\mu(k+1)f(k+1) - N\mu L f(k) \\ &\quad + \sum_{k'}[\delta_{kk'} - f(k)]\sqrt{f(k')}\eta(k').\end{aligned} \tag{A.1}$$

In contrast to the $L \to \infty$ case analyzed in the main text, the behavior of the average profile $\langle f(k)\rangle$ in the neutral limit no longer degenerate, and we find that

$$\lim_{\tau\to\infty}\langle f(k,\tau)\rangle = \binom{L}{k}2^{-L}. \tag{A.2}$$

This is just a binomial distribution with mean $L/2$ and variance $L/4$, which is consistent with the intuition that the population at long times consists of individuals with independent and identically distributed mutations along the L sites in the genome. Although there is no longer any infinite-width "red-flag" to suggest that fluctuations may play an important role here, the absence of any $N\mu$ dependence in Eq. (A.2) is suspicious, since we would expect that the typical width of the distribution should vanish as $N\mu \to 0$.

Thus, we are lead to consider the behavior of the mean $\langle\bar{k}\rangle$ and the central moments M_m that we analyzed in the infinite-sites model. Using the dynamics in Eq. (A.1) it is straightforward to show that

$$\begin{aligned}\frac{\partial\langle\bar{k}\rangle}{\partial\tau} &= NsM_2 + N\mu\left\langle\sum_k[k(L-k+1)f(k-1) + k \right. \\ &\quad \left. + 1f(k+1) - Lf(k)]\right\rangle \\ &= NsM_2 + N\mu L\left[1 - \frac{2\bar{k}}{L}\right],\end{aligned} \tag{A.3}$$

so that in the neutral limit the population reaches mutation–reversion balance when the average number of mutations in each genome is $\langle\bar{k}\rangle = L/2$, just like the average profile in Eq. (A.2). We also see that this equilibrium point is reached on a timescale $t_{eq} \sim 1/2\mu$. For the central moments M_m, it is straightforward to show that the equations become

$$\begin{aligned}\frac{\partial M_m}{\partial\tau} &= N\mu L\sum_{\ell=0}^{m-2}\binom{m}{\ell}M_\ell + \binom{m}{2}M_{2,m-2} \\ &\quad - mM_m + Ns\left(M_{m+1} - mM_{2,m-1}\right) \\ &\quad - 2mN\mu M_m - N\mu\sum_{\ell=0}^{m-2}\binom{m}{\ell}\left[M_{\ell+1} + \langle\bar{k}M_\ell\rangle\right] \\ &\quad \times\left[1 + (-1)^{m-\ell+1}\right],\end{aligned} \tag{A.4}$$

where the last two terms arise from the k-dependent mutation rates. The first few orders of the moment hierarchy are given by

$$\frac{\partial M_2}{\partial \tau} = N\mu L - [1+4N\mu]M_2 + NsM_3, \tag{A.5a}$$

$$\frac{\partial M_3}{\partial \tau} = N\mu\left[L - 2\langle\bar{k}\rangle\right] - 3[1+2N\mu]M_3 + Ns\left[M_4 - 3M_{2,2}\right], \tag{A.5b}$$

$$\frac{\partial M_4}{\partial \tau} = N\mu L + [6N\mu L - 8N\mu]M_2 + 6M_{2,2} - 4[1+2N\mu]M_4 + O(Ns), \tag{A.5c}$$

$$\frac{\partial M_{2,2}}{\partial \tau} = 2N\mu L M_2 - [3+8N\mu]M_{2,2} + M_4 + O(Ns). \tag{A.5d}$$

Thus, in the neutral limit, the actual variance in k within the population is given by

$$M_2 = \frac{N\mu L}{1+4N\mu}, \tag{A.6}$$

which only approaches the deterministic value of $L/4$ when $N\mu \gg 1$. For smaller per-site mutation rates, the width of the fitness distribution can be much smaller than this, and for $N\mu \ll 1$ it approaches the infinite-sites limit $M_2 = N\mu L$ that we analyzed in the main text. We can apply our perturbative approach to this moment hierarchy as well, which shows that when the mutants are weakly beneficial the equilibrium value of $\langle\bar{k}\rangle$ is given by

$$\langle\bar{k}\rangle = \frac{L}{2}\left[1 + \frac{Ns}{1+4N\mu} - \frac{2Ns}{3}\left(\frac{Ns}{1+4N\mu}\right)^2 \times \left(\frac{N\mu L + \frac{1}{2} + 4N\mu + 16(N\mu)^2}{3 + \frac{34}{3}N\mu + \frac{88}{3}(N\mu)^2 + \frac{64}{3}(N\mu)^3}\right)\right] + O(Ns)^4. \tag{A.7}$$

Following Woodcock and Higgs (1996) and Goyal et al. (2012), we define ϵ_c to be the fraction of all the possible mutations that are beneficial at this equilibrium point, or

$$\epsilon_c = \frac{(L - \langle\bar{k}\rangle)\mu}{L\mu} = 1 - \frac{\langle\bar{k}\rangle}{L}. \tag{A.8}$$

This allows us to rewrite Eq. (A.7) in terms of this critical ratio as

$$\epsilon_c = \frac{1}{2}\left[1 - \frac{Ns}{1+4N\mu} + \frac{2Ns}{3}\left(\frac{Ns}{1+4N\mu}\right)^2 \times \left(\frac{N\mu L + \frac{1}{2} + 4N\mu + 16(N\mu)^2}{3 + \frac{34}{3}N\mu + \frac{88}{3}(N\mu)^2 + \frac{64}{3}(N\mu)^3}\right)\right] + O(Ns)^4. \tag{A.9}$$

In the limit that $N\mu \to 0$ and $L \to \infty$ with $NU = N\mu L$ fixed, we recover the infinite sites prediction

$$\epsilon_c = \frac{1}{2}\left[1 - Ns + \frac{(Ns)^3}{3} + \frac{2NU(Ns)^3}{3}\right] + O(Ns)^4, \tag{A.10}$$

which provides a more accurate expression for the critical fraction as $Ns \to 0$ compared to the corresponding expression in Goyal et al. (2012).

Appendix B. Stochastic calculus

In this section, we outline the stochastic (Itô) calculus that is used to derive moment equations from the stochastic dynamics in Eqs. (1), (6) and (8). For concreteness, we will restrict our attention to the neutral dynamics in Eq. (8), but these results will apply more generally. Let $\phi(\{f_k\})$ be some arbitrary function of the fitness classes, $f_k(t)$. We wish to find an expression for the time-evolution of the mean $\langle\phi(\{f_k\})\rangle$ using the definition

$$\frac{\partial\langle\phi(\{f_k\})\rangle}{\partial t} = \lim_{\delta t\to 0}\frac{\langle\phi(\{f_k(t+\delta t)\})\rangle - \langle\phi(\{f_k(t)\})\rangle}{\delta t}. \tag{B.1}$$

The dynamics in Eq. (8) are essentially just a shorthand notation for calculating $f_k(t+\delta t)$ conditioned on $f_k(t)$. Taking care to note that the drift term in Eq. (8) is of the Itô form, the dynamics in Eq. (8) imply that

$$f_k(t+\delta t) = f_k(t) + \delta t\,[Uf(k-1,t) + Uf(k,t)] + \sqrt{\delta t}\left[\frac{1}{\sqrt{N}}\sum_{k'}[\delta_{kk'} - f(k,t)]\sqrt{f(k',t)}\eta(k',t)\right]. \tag{B.2}$$

The value of $\phi(\{f_k(t+\delta t)\})$ can then be found by Taylor expansion in powers of δt. If $\phi(\{f_k\})$ is just a singleton function $\phi(\{f_k\}) = f_k(t)$, then

$$\langle\phi(\{f_k(t+\delta t)\})\rangle = \langle f_k(t)\rangle + \delta t\,\langle Uf(k-1,t) + Uf(k,t)\rangle \tag{B.3}$$

$$= \langle f_k(t)\rangle + \delta t\left\langle\left(\frac{\partial f_k}{\partial t}\right)_{\text{det}}\right\rangle, \tag{B.4}$$

where $(\partial f_k/\partial t)_{\text{det}}$ is simply the deterministic part of the dynamics in Eq. (8). On the other hand, if $\phi(\{f_k\})$ is a pairwise product of the form

$$\phi(\{f_k\}) = f_{k_1}(t)f_{k_2}(t), \tag{B.5}$$

then things start to become more complicated. Organizing all the terms in powers of δt, we see that

$$f_{k_1}(t+\delta t)f_{k_2}(t+\delta t) = f_{k_1}f_{k_2} + \delta t\left[\left(Uf_{k_1-1} - Uf_{k_1}\right)f_{k_2} + f_{k_1}\left(Uf_{k_2-1} - Uf_{k_2}\right) + \frac{1}{N}\sum_{j_1,j_2}\left[\delta_{k_1,j_1} - f_{k_1}\right]\left[\delta_{k_2,j_2} - f_{k_2}\right]\sqrt{f_{j_1}f_{j_2}}\eta_{j_1}\eta_{j_2}\right] + \sqrt{\delta t}\,O(\eta). \tag{B.6}$$

The term proportional to $\sqrt{\delta t}$ is linear in η, so upon averaging this term vanishes. The term proportional to δt also contains a term involving η, but this time as a quadratic function rather than a linear function, which yields

$$\left\langle\sum_{j_1,j_2}\left[\delta_{k_1,j_1} - f_{k_1}\right]\left[\delta_{k_2,j_2} - f_{k_2}\right]\sqrt{f_{j_1}f_{j_2}}\eta_{j_1}\eta_{j_2}\right\rangle = \left\langle\sum_{j}\left[\delta_{k_1,j} - f_{k_1}\right]\left[\delta_{k_2,j} - f_{k_2}\right]f_j\right\rangle \tag{B.7}$$

$$= \left[\delta_{k_1,k_2} - f_{k_2}\right]f_{k_1}, \tag{B.8}$$

where we have used the fact that the η_k are uncorrelated for different k. Thus, taking the average of Eq. (B.6), we obtain

$$\frac{\partial\langle f(k_1,t)f(k_2,t)\rangle}{\partial t} = \left\langle f(k_1,t)\left(\frac{\partial f(k_2,t)}{\partial t}\right)_{\text{det}}\right\rangle + \left\langle\left(\frac{\partial f(k_1,t)}{\partial t}\right)_{\text{det}}f(k_2,t)\right\rangle + \frac{1}{N}\langle f(k_1,t)\star f(k_2,t)\rangle, \tag{B.9}$$

where we have defined a new operation $\star$ such that

$$f(k_1,t)\star f(k_2,t) \equiv \left[\delta_{k_1,k_2} - f_{k_2}\right]f_{k_1}. \tag{B.10}$$

This is the generalized product rule of the Itô calculus, which arises from the combination of two $\sqrt{\delta t}$ terms in the expansion of the Langevin equation. More complicated functions of the f_k can be analyzed recursively with the help of the sum and product rules

$$\frac{\partial\langle\phi(f)+\psi(f)\rangle}{\partial t} = \left\langle\frac{\partial\phi(f)}{\partial t}\right\rangle + \left\langle\frac{\partial\psi(f)}{\partial t}\right\rangle \tag{B.11}$$

and

$$\frac{\partial\langle\phi(f)\psi(f)\rangle}{\partial t} = \left\langle\frac{\partial\phi(f)}{\partial t}\psi(f)\right\rangle + \left\langle\phi(f)\frac{\partial\psi(f)}{\partial t}\right\rangle + \frac{1}{N}\langle\phi(f)\star\psi(f)\rangle, \tag{B.12}$$

where $\star$ is defined in terms of the underlying fitness classes f_k and satisfies additivity and the distributive property. As an example, we have

$$\bar{k}\star f(k,t) = \sum_{k'} k'\left[f(k,t)\star f(k',t)\right] \tag{B.13}$$

$$= \sum_{k'} k'\left[\delta(k-k') - f(k)\right]f(k') \tag{B.14}$$

$$= (k-\bar{k})f(k,t), \tag{B.15}$$

and

$$\bar{k}\star\bar{k} = \sum_k k[\bar{k}\star f(k,t)] \tag{B.16}$$

$$= \sum_k k(k-\bar{k})f(k,t) \tag{B.17}$$

$$= \sum_k (k-\bar{k})^2 f(k,t). \tag{B.18}$$

These rules can be used to rapidly generate equations of motion for the generalized moments $M_{\vec{m}}$, $C_{\vec{m}}$, and $F_{\vec{m};\vec{n}}$ analyzed in the text (see Appendices C, D and F).

Appendix C. Central moments

In this section, we use the rules of the stochastic calculus in Appendix B to derive equations of motion for the mean "fitness"

$$\bar{k} = \int dk\, k f(k,t) \tag{C.1}$$

and the central moments

$$M_m = \left\langle\int dk\,(k-\bar{k})^m f(k,t)\right\rangle. \tag{C.2}$$

Without loss of generality, we will restrict our attention to the full dynamics in Eq. (6), which can be rewritten in terms of the scaled parameters as

$$\frac{\partial f(k)}{\partial\tau} = Ns(k-\bar{k})f(k) + NU\int d(\Delta k)\,\rho(\Delta k)\left[f(k-\Delta k) - f(k)\right] + \int dk'\left[\delta(k'-k) - f(k)\right]\sqrt{f(k')}\eta(k'). \tag{C.3}$$

Directly from the Langevin equation, we can see that

$$\frac{\partial\langle\bar{k}\rangle}{\partial\tau} = \int dk\, k\left\langle\frac{f(k,\tau)}{\partial\tau}\right\rangle = \int dk\, k\left\langle NU\int d(\Delta k)\,\rho(\Delta k)\left[f(k-\Delta k) - f(k)\right] + Ns(k-\bar{k})f(k)\right\rangle = NU\langle\Delta k\rangle + Ns\cdot M_2. \tag{C.4}$$

For the central moments M_m, we can use the product rule in Eq. (B.12) to show that

$$\frac{\partial M_m}{\partial\tau} = \left\langle\int dk\,(k-\bar{k})^m\frac{\partial f(k)}{\partial t} - m\left[\int dk\,(k-\bar{k})^{m-1}f(k)\right]\frac{\partial\bar{k}}{\partial\tau}\right\rangle + \left\langle -m\int dk\,(k-\bar{k})^{m-1}\left[\bar{k}\star f(k)\right]\right\rangle + \binom{m}{2}\left\langle\left[\int dk\,(k-\bar{k})^{m-2}f(k)\right]\bar{k}\star\bar{k}\right\rangle, \tag{C.5}$$

or

$$\frac{\partial M_m}{\partial\tau} = NU\sum_{\ell=0}^{m-2}\binom{m}{\ell}\langle\Delta k^{m-\ell}\rangle M_\ell + \binom{m}{2}M_{2,m-2} - mM_m + Ns\left(M_{m+1} - mM_{2,m-1}\right). \tag{C.6}$$

This equation can also be rewritten in terms of the scaled moments $\tilde{M}_m = M_m/(NU\langle\Delta k^2\rangle)^{m/2}$ in order to remove the zeroth order dependence on the variance of $f(k)$. This yields

$$\frac{\partial\tilde{M}_m}{\partial\tau} = \sum_{j=0}^{m-2}\frac{\binom{m}{2+j}\frac{\langle\Delta k^{2+j}\rangle}{\langle\Delta k^2\rangle}\tilde{M}_{m-\ell-2}}{(NU\langle\Delta k^2\rangle)^{j/2}} + \binom{m}{2}\tilde{M}_{2,m-2} - m\tilde{M}_m + N\sigma\left[\tilde{M}_{m+1} - m\tilde{M}_{2,m-1}\right], \tag{C.7}$$

where $N\sigma \equiv NU\langle\Delta k^2\rangle(Ns)^2$ is the zeroth order estimate of the variance in fitness within the population. Thus, in the limit that $NU\to\infty$, this reduces to

$$\frac{\partial\tilde{M}_m}{\partial\tau} = \binom{m}{2}\tilde{M}_{m-2} + \binom{m}{2}\tilde{M}_{2,m-2} - m\tilde{M}_m + N\sigma\left[\tilde{M}_{m+1} - m\tilde{M}_{2,m-1}\right],$$

which only depends on the underlying distribution of fitness effects through the scaled parameter $N\sigma$.

In order to obtain equations of motion for the generalized products $M_{m_1,\ldots,m_J}$, we can again appeal to the product rule in Eq. (B.12) which shows that (with some abuse of notation)

$$\frac{\partial M_{m,n}}{\partial\tau} = \left\langle\frac{\partial M_m}{\partial\tau}M_n\right\rangle + \left\langle M_m\frac{\partial M_n}{\partial\tau}\right\rangle + \langle M_m\star M_n\rangle. \tag{C.8}$$

This requires us to compute the higher-order "star products"

$$\bar{k}\star M_m = \int dk\,(k-\bar{k})^m[\bar{k}\star f(k,t)] - mM_{m-1}(\bar{k}\star\bar{k}) = M_{m+1} - mM_{2,m-1} \tag{C.9}$$

and

$$M_m\star M_n = \int dk_1\,dk_2\,(k_1-\bar{k})^m(k_2-\bar{k})[f(k_1)\star f(k_2)] - mM_{m-1}\int dk\,(k-\bar{k})^n[\bar{k}\star f(k)] - nM_{n-1}\int dk\,(k-\bar{k})^m[\bar{k}\star f(k)] + mn(\bar{k}\star\bar{k})M_{m-1}M_{n-1} = M_{m+n} - M_{m,n} - mM_{m-1,n+1} - nM_{m+1,n-1} + mnM_{2,m-1,n-1}. \tag{C.10}$$

Combining these formulas with the dynamics in Eq. (C.6), we can easily write down the first few equations in the moment hierarchy

for M_m, several of which were previously obtained by Higgs and Woodcock (1995) and Etheridge et al. (2007):

$$\frac{\partial M_2}{\partial \tau} = NU\langle \Delta k^2\rangle - M_2 + NsM_3, \tag{C.11a}$$

$$\frac{\partial M_3}{\partial \tau} = NU\langle \Delta k^3\rangle - 3M_3 + Ns\left(M_4 - 3M_{2,2}\right), \tag{C.11b}$$

$$\frac{\partial M_4}{\partial \tau} = NU\langle \Delta k^4\rangle + 6NU\langle \Delta k^2\rangle M_2 + 6M_{2,2} - 4M_4 + Ns\left(M_5 - 4M_{2,3}\right), \tag{C.11c}$$

$$\frac{\partial M_{2,2}}{\partial \tau} = 2NU\langle \Delta k^2\rangle M_2 + M_4 - 3M_{2,2} + 2NsM_{2,3}, \tag{C.11d}$$

$$\frac{\partial M_5}{\partial \tau} = NU\langle \Delta k^5\rangle + 10NU\left[\langle \Delta k^3\rangle M_2 + \langle \Delta k^2\rangle M_3\right] + 10M_{2,3} - 5M_5 + O(Ns), \tag{C.11e}$$

$$\frac{\partial M_{2,3}}{\partial \tau} = NU\left[\langle \Delta k^3\rangle M_2 + \langle \Delta k^2\rangle M_3\right] - 8M_{2,3} + M_5 + O(Ns). \tag{C.11f}$$

Higher orders start to include contributions from replica moments with more than two terms, and the corresponding equations of motion are slightly more complicated. Nevertheless, these dynamics can be obtained by repeated application of the product rule in Eq. (B.12) using the expressions we have already derived. This yields the general formula

$$\begin{aligned}\frac{\partial M_{\vec{m}}}{\partial \tau} = &\sum_{j=1}^{\dim \vec{m}}\Bigg[NU\sum_{\ell=0}^{m_j-2}\binom{m_j}{\ell}\langle \Delta k^{m_j-\ell}\rangle M_{\vec{m}+(\ell-m_j)\hat{e}_j} \\ &+ \binom{m_j}{2} M_{2,\vec{m}-2\hat{e}_j} - m_j M_{\vec{m}} + Ns\left(M_{\vec{m}+\hat{e}_j} - m_j M_{2,\vec{m}-\hat{e}_j}\right)\Bigg] \\ &+ \sum_{j=1}^{\dim \vec{m}}\sum_{i>j}^{\dim \vec{m}}\Big[M_{\vec{m}+m_i(\hat{e}_j-\hat{e}_i)} - M_{\vec{m}} - m_j M_{\vec{m}-\hat{e}_j+\hat{e}_i} \\ &- m_i M_{\vec{m}+\hat{e}_j-\hat{e}_i} + m_j m_i M_{2,\vec{m}-\hat{e}_j-\hat{e}_i}\Big],\end{aligned} \tag{C.12}$$

where $\hat{e}_j$ denotes the jth "unit vector", which has a 1 in entry j and a zero everywhere else.

Appendix D. Cumulants

In this section, we derive a hierarchy of equations for the cumulants of the fitness distribution, which provide an alternative representation of the central moment hierarchy derived in Appendix C. The cumulants C_m are most easily defined in terms of the cumulant generating function

$$g(z) = \log\left(\int dk\, e^{zk} f(k)\right), \tag{D.1}$$

using the relation

$$C_m = \left.\frac{\partial^m \langle g(z)\rangle}{\partial z^m}\right|_{z=0}. \tag{D.2}$$

For more general replica products $C_{\vec{m}}$, the corresponding definition is simply

$$C_{m_1,\ldots,m_J} = \left\langle \prod_{i=1}^{J}\left.\frac{\partial^{m_i} g(z)}{\partial z^{m_i}}\right|_{z=0}\right\rangle. \tag{D.3}$$

Using the rules of the stochastic calculus in Appendix B, we can try to derive an equation of motion for the average cumulant generating function $\langle g(z)\rangle$. This yields

$$\begin{aligned}\frac{\partial \langle g(z)\rangle}{\partial \tau} &= \left\langle \frac{\int dk\, e^{zk}\frac{\partial f(k)}{\partial \tau}}{\int dk\, e^{zk} f(k)}\right\rangle - \frac{1}{2}\left\langle \frac{\int dk\, dk' e^{zk} e^{zk'} f(k)\star f(k')}{\left(\int dk\, e^{zk} f(k)\right)^2}\right\rangle, \\ &= Ns\left\langle \frac{\int dk\, e^{zk}(k-\bar{k})f(k)}{\int dk\, e^{zk} f(k)}\right\rangle \\ &\quad + NU\int d(\Delta k)\,\rho(\Delta k)\left\langle \frac{\int dk\, e^{zk}[f(k-\Delta k)-f(k)]}{\int dk\, e^{zk} f(k)}\right\rangle \\ &\quad - \frac{1}{2}\left\langle \frac{\int dk\, e^{2zk} f(k)}{\left(\int dk\, e^{zk} f(k)\right)^2} - \frac{\left(\int dk\, e^{zk} f(k)\right)\left(\int dk'\, e^{zk'} f(k')\right)}{\left(\int dk\, e^{zk} f(k)\right)^2}\right\rangle, \\ &= Ns\frac{\partial \langle g(z)\rangle}{\partial z} - Ns\langle \bar{k}\rangle + NU\int d(\Delta k)\,\rho(\Delta k)\left[e^{z\Delta k}-1\right] \\ &\quad - \frac{1}{2}\left\langle e^{g(2z)-2g(z)}-1\right\rangle.\end{aligned} \tag{D.4}$$

Thus, we see that because of the last term on the right, the equation for $\langle g(z)\rangle$ does not close. Nevertheless, we will still be able to use Eq. (D.4) to obtain equations of motion for the cumulants C_m by expanding in powers of z. For instance, expanding the left side of Eq. (D.4) yields the power series

$$\frac{\partial \langle g(z)\rangle}{\partial \tau} = \sum_{m=1}^{\infty}\frac{z^m}{m!}\frac{\partial C_m}{\partial \tau}, \tag{D.5}$$

so the equation of motion for C_m can be obtained by examining the mth power of z on the right side. The first three terms on the right are simple. These yield

$$Ns\frac{\partial \langle g(z)\rangle}{\partial z} = Ns\sum_{m=0}^{\infty}\frac{z^m}{m!}C_{m+1}, \tag{D.6}$$

$$Ns\langle \bar{k}\rangle = O(z^0), \tag{D.7}$$

$$NU\int d(\Delta k)\,\rho(\Delta k)\left[e^{z\Delta k}-1\right] = NU\sum_{m=1}^{\infty}\frac{z^m}{m!}\langle \Delta k^m\rangle. \tag{D.8}$$

The last term on the right is slightly more difficult, but can be attacked by recursive applications of exponential, binomial, and Taylor series:

$$\begin{aligned}\left\langle e^{g(2z)-2g(z)}-1\right\rangle &= \left\langle \sum_{J=1}^{\infty}\frac{1}{J!}\left[\sum_{k=1}^{\infty}\frac{(2z)^m - 2z^m}{m!}\frac{\partial^m g(z)}{\partial z^m}\right]^J\right\rangle, \\ &= \sum_{J=1}^{\infty}\sum_{m_1=1}^{\infty}\cdots\sum_{m_J=1}^{\infty}\left(\frac{2C_{m_1,\ldots,m_J}}{J!}\right)\left(z^{\sum_i m_i}\right)\prod_{i=1}^{J}\left(\frac{2^{m_i-1}-1}{m_i!}\right), \\ &= \sum_{m=0}^{\infty}\frac{z^m}{m!}\sum_{\vec{m}}\frac{2C_{\vec{m}}}{(\dim \vec{m})!}\binom{m}{\vec{m}}\prod_{i=1}^{\dim \vec{m}}\left(2^{m_i-1}-1\right),\end{aligned} \tag{D.9}$$

where $\sum_{\vec{m}}$ denotes a sum over all partitions of m and $\dim \vec{m}$ is the number of entries in $\vec{m}$. Combining these series expansions and equating like powers of z, we see that the cumulants satisfy the compact equation

$$\frac{\partial C_m}{\partial \tau} = Ns\cdot C_{m+1} + NU\langle \Delta k^m\rangle - \sum_{\vec{m}}\binom{m}{\vec{m}}\frac{\prod_i(2^{m_i-1}-1)}{(\dim \vec{m})!}C_{\vec{m}}. \tag{D.10}$$

Thus, we see that just like the central moments above, the equations of motion for the cumulants also involve the corresponding replica products $C_{\vec{m}}$ for sufficiently large m. In order to obtain equations of motion for these replica products, it remains for us to evaluate $C_m \star C_n$ so that we can utilize the product rule in Appendix B. This can be done using a similar power series approach, making use of the identity

$$g(x) \star g(y) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{x^m y^n}{m!n!} C_m \star C_n. \tag{D.11}$$

Using the rules of the stochastic calculus in Appendix B, we can show that

$$\begin{aligned} g(x) \star g(y) &= \frac{\int dk\, dk' e^{xk} e^{yk'} f(k) \star f(k')}{(\int dk\, e^{xk} f(k))(\int dk'\, e^{yk'} f(k'))}, \\ &= \exp\left[g(x+y) - g(x) - g(y)\right] - 1, \\ &= \sum_{J=1}^{\infty} \frac{1}{J!} \left[\sum_{k=1}^{\infty} \frac{C_k}{k!} \sum_{\ell=1}^{k-1} \binom{k}{\ell} x^{\ell} y^{k-\ell}\right]^J, \\ &= \sum_{J=1}^{\infty} \frac{1}{J!} \sum_{k_1=1}^{\infty} \cdots \sum_{k_J=1}^{\infty} \sum_{\ell_1=1}^{k_1-1} \cdots \sum_{\ell_J=1}^{k_J-1} C_{\vec{k}} \left[x^{\sum \ell_i} y^{\sum k_i - \sum \ell_i}\right] \\ &\quad \times \prod_{i=1}^{J} \frac{\binom{k_i}{\ell_i}}{k_i!}, \\ &= \sum_{m,n} \frac{x^m y^n}{m!n!} \left[\sum_{\vec{k}} \frac{\binom{n+m}{\vec{k}}}{\binom{n+m}{m}} \frac{A(\vec{k}, m)}{(\dim \vec{k})!} C_{\vec{k}}\right], \end{aligned} \tag{D.12}$$

where $\sum_{\vec{k}}$ denotes a sum over all partitions of $m+n$ and $A(\vec{k}, m)$ is a counting factor defined by

$$A(\vec{k}, m) = \sum_{\substack{\ell_1, \ldots, \ell_J \\ 0 < \ell_i < k_i \\ \sum \ell_i = m}} \prod_i \binom{k_i}{\ell_i}. \tag{D.13}$$

Thus, we conclude that

$$C_m \star C_n = \sum_{\vec{k}} \frac{\binom{n+m}{\vec{k}}}{\binom{n+m}{m}} \frac{A(\vec{k}, m)}{(\dim \vec{k})!} C_{\vec{k}}, \tag{D.14}$$

and a similar calculation shows that

$$\bar{k} \star C_m = C_{m+1}. \tag{D.15}$$

The equations of motion for general replica products $C_{\vec{m}}$ can then be obtained by combining Eqs. (D.10) and (D.14) with the product rule derived in Eq. (B.12).

Appendix E. Fluctuations in the mean fitness

In this section, we use the rules of the stochastic calculus in Appendix B to derive equations of motion for the fluctuations in the mean fitness of the population,

$$\mathrm{Var}(\bar{k}) = \langle \bar{k}^2 \rangle - \langle \bar{k} \rangle^2. \tag{E.1}$$

Starting from the product rule in Eq. (B.12), we have

$$\begin{aligned} \frac{\partial \mathrm{Var}(\bar{k})}{\partial \tau} &= \left\langle 2\bar{k} \frac{\partial \bar{k}}{\partial \tau} \right\rangle + \langle \bar{k} \star \bar{k} \rangle - 2\langle \bar{k} \rangle \frac{\partial \langle \bar{k} \rangle}{\partial t}, \\ &= C_2 + 2Ns \cdot \mathrm{Cov}(\bar{k}, C_2), \end{aligned} \tag{E.2}$$

where we have substituted our expressions for $\partial_\tau \bar{k}$ and $\bar{k} \star \bar{k}$ derived above. Thus, we see that selection couples the variance in the mean fitness to its *covariance* with the central cumulants. Using our previously derived results, we can derive the general formula

$$\begin{aligned} \frac{\partial \mathrm{Cov}(\bar{k}, C_m)}{\partial \tau} &= \left\langle \frac{\partial \bar{k}}{\partial \tau} C_m \right\rangle + \left\langle \bar{k} \frac{\partial C_m}{\partial \tau} \right\rangle + \langle \bar{k} \star C_m \rangle \\ &\quad - \frac{\partial \langle \bar{k} \rangle}{\partial \tau} C_m - \bar{k} \frac{\partial C_m}{\partial \tau}, \\ &= C_{m+1} + Ns\left[\mathrm{Cov}(\bar{k}, C_{m+1}) + \mathrm{Cov}(C_2, C_m)\right] \\ &\quad - \sum_{\vec{m}} \frac{\prod_i (2^{m_i - 1} - 1)}{(\dim \vec{m})!} \binom{m}{\vec{m}} \mathrm{Cov}(\bar{k}, C_{\vec{m}}). \end{aligned} \tag{E.3}$$

The first few orders of the moment hierarchy are therefore given by

$$\frac{\partial \mathrm{Var}(\bar{k})}{\partial \tau} = C_2 + 2Ns\mathrm{Cov}(\bar{k}, C_2), \tag{E.4a}$$

$$\begin{aligned} \frac{\partial \mathrm{Cov}(\bar{k}, C_2)}{\partial \tau} &= -\mathrm{Cov}(\bar{k}, C_2) + C_3 + Ns\left[\mathrm{Var}(C_2) + \mathrm{Cov}(\bar{k}, C_3)\right], \end{aligned} \tag{E.4b}$$

$$\frac{\partial \mathrm{Cov}(\bar{k}, C_3)}{\partial \tau} = -3\mathrm{Cov}(\bar{k}, C_3) + C_4 + O(Ns). \tag{E.4c}$$

Appendix F. Focal lineage moments

In this section, we use the rules of the stochastic calculus in Appendix B to derive equations of motion for the focal lineage moments

$$F_m = \left\langle \int dk\, (k - \bar{k})^m f_1(k, t) \right\rangle \tag{F.1}$$

discussed in the main text. Starting from the product rule in Eq. (B.12), we have

$$\begin{aligned} \frac{\partial F_m}{\partial \tau} &= \left\langle \int dk\, (k - \bar{k})^m \frac{\partial f_1(k)}{\partial t} - m\left[\int dk\, (k - \bar{k})^{m-1} f_1(k)\right] \frac{\partial \bar{k}}{\partial \tau} \right\rangle \\ &\quad + \left\langle -m \int dk\, (k - \bar{k})^{m-1} \left[\bar{k} \star f_1(k)\right] \right. \\ &\quad \left. + \binom{m}{2} \left[\int dk\, (k - \bar{k})^{m-2} f_1(k)\right] \bar{k} \star \bar{k} \right\rangle. \end{aligned} \tag{F.2}$$

In order to progress further, we must extend the $\star$ operation to the case where the fitness classes $f_i(k, t)$ are labeled according to which lineage they descend from. It is a straightforward matter to show that

$$f_{i_1}(k_1, t) \star f_{i_2}(k_2, t) = \left[\delta_{i_1 i_2} \delta(k_1 - k_2) - f_{i_2}(k_2)\right] f_{i_1}(k_1). \tag{F.3}$$

Thus, we have

$$\begin{aligned} \bar{k} \star f_1(k, t) &= \sum_i \int dk'\, k' [f_i(k', t) \star f_1(k, t)], \\ &= \sum_i \int dk'\, k' \left[\delta(k - k') \delta_{i1} - f_1(k, t)\right] f_i(k', t), \\ &= (k - \bar{k}) f_1(k, t), \end{aligned} \tag{F.4}$$

and we can immediately conclude that

$$\begin{aligned} \frac{\partial F_m}{\partial \tau} &= NU \sum_{\ell=0}^{m-2} \binom{m}{\ell} \langle \Delta k^{m-\ell} \rangle F_\ell \\ &\quad + \binom{m}{2} F_{m-2;2} - m F_m + Ns\left(F_{m+1} - m F_{m-1;2}\right). \end{aligned} \tag{F.5}$$

In order to derive the equations of motion for the general products $F_{\vec{m};\vec{n}}$ considered in the text, we simply need to calculate $\star$ products of the form $F_m \star M_n$ and $F_m \star F_n$. Starting from

$$\begin{aligned}\bar{k} \star F_m &= -mF_{m-1}(\bar{k} \star \bar{k}) + \int dk\,(k-\bar{k})^m[\bar{k} \star f_1(k,t)], \\ &= F_{m+1} - mF_{m-1;2}, \end{aligned} \tag{F.6}$$

we can easily see show that

$$\begin{aligned}F_m \star M_n &= mnF_{m-1}M_{m-1}[\bar{k} \star \bar{k}] \\ &\quad + \sum_i \int dk_1\,dk_2\,(k_1-\bar{k})^m(k_2-\bar{k})^n[f_1(k_1) \star f_i(k_2)] \\ &\quad - mF_{m-1}\sum_i \int dk\,(k-\bar{k})^m[\bar{k} \star f_i(k)] \\ &\quad - nM_{n-1}\int dk\,(k-\bar{k})^m[\bar{k} \star f_1(k)], \\ &= F_{m+n} - F_{m;n} - mF_{m-1;n+1} - nF_{m+1;n-1} \\ &\quad + mnF_{m-1;n-1,2}, \end{aligned} \tag{F.7}$$

and

$$\begin{aligned}F_m \star F_n &= \int dk_1\,dk_2\,(k_1-\bar{k})^m(k_2-\bar{k})^n[f_1(k_1) \star f_1(k_2)] \\ &\quad - mF_{m-1}\int dk\,(k-\bar{k})^n[\bar{k} \star f_1(k)] \\ &\quad - nF_{n-1}\int dk\,(k-\bar{k})^m[\bar{k} \star f_1(k)] + mnF_{m-1}F_{n-1}[\bar{k} \star \bar{k}], \\ &= F_{m+n} - F_{m,n} - mF_{m-1,n+1} - nF_{m+1,n-1} + mnF_{m-1,n-1;2}. \end{aligned} \tag{F.8}$$

The equations of motion for generalized products with more than two terms follow from the product rule in Eq. (B.12). Using these formulas, we can derive the first few equations in the moment hierarchy for F_0:

$$\frac{\partial F_0}{\partial \tau} = NsF_1, \tag{F.9a}$$

$$\frac{\partial F_1}{\partial \tau} = -F_1 + Ns\left(F_2 - F_{0;2}\right), \tag{F.9b}$$

$$\frac{\partial F_2}{\partial \tau} = NU\langle \Delta k^2\rangle F_0 + F_{0;2} - 2F_2 + Ns\left(F_3 - 2F_{1;2}\right), \tag{F.9c}$$

$$\frac{\partial F_{0;2}}{\partial \tau} = NU\langle \Delta k^2\rangle F_0 + F_2 - 2F_{0;2} + Ns\left(F_{1;2} + F_{0;3}\right), \tag{F.9d}$$

$$\frac{\partial F_3}{\partial \tau} = NU\langle \Delta k^3\rangle F_0 + 3NU\langle \Delta k^2\rangle F_1 - 3F_3 + 3F_{1;2} + O(Ns), \tag{F.9e}$$

$$\frac{\partial F_{0;3}}{\partial \tau} = NU\langle \Delta k^3\rangle F_0 + F_3 - 4F_{0;3} - 3F_{1;2} + O(Ns), \tag{F.9f}$$

$$\frac{\partial F_{1;2}}{\partial \tau} = NU\langle \Delta k^2\rangle F_1 + F_3 - F_{0;3} - 3F_{1;2} + O(Ns), \tag{F.9g}$$

as well as the corresponding moment hierarchy for the heterozygosity H:

$$\frac{\partial H}{\partial \tau} = -H + Ns\left(F_1 - 2F_{0,1}\right), \tag{F.10a}$$

$$\begin{aligned}\frac{\partial (F_1 - 2F_{0,1})}{\partial \tau} &= -3\left(F_1 - 2F_{0,1}\right) + Ns\left(F_2 - 2F_{0,2} - F_{0;2}\right) \\ &\quad + Ns\left(2F_{0,0;2} - 2F_{1,1}\right), \end{aligned} \tag{F.10b}$$

$$\begin{aligned}\frac{\partial F_{0,2}}{\partial \tau} &= NU\langle \Delta k^2\rangle\,(F_0 - H) + F_2 + F_{0,0;2} - 3F_{0,2} - 2F_{1,1} \\ &\quad + O(Ns), \end{aligned} \tag{F.10c}$$

$$\begin{aligned}\frac{\partial F_{0,0;2}}{\partial \tau} &= NU\langle \Delta k^2\rangle\,(F_0 - H) + F_{0;2} - 4F_{0,0;2} \\ &\quad + 2F_{0,2} + O(Ns), \end{aligned} \tag{F.10d}$$

$$\frac{\partial F_{1,1}}{\partial \tau} = -3F_{1,1} + F_2 - 2F_{0,2} + F_{0,0;2} + O(Ns). \tag{F.10e}$$

For arbitrary replica moments $F_{\vec{m};\vec{n}}$, we can use the product rule in Eq. (B.12) to derive a general formula similar to Eq. (C.12) for the central moments. This yields

$$\begin{aligned}\frac{\partial F_{\vec{m};\vec{n}}}{\partial \tau} &= \sum_{j=1}^{\dim \vec{m}}\Bigg[NU\sum_{\ell=0}^{m_j-2}\binom{m_j}{\ell}\langle \Delta k^{m_j-\ell}\rangle F_{\vec{m}+(\ell-m_j)\hat{e}_j;\vec{n}} \\ &\quad + \binom{m_j}{2}F_{\vec{m}-2\hat{e}_j;2,\vec{n}} - m_jF_{\vec{m};\vec{n}} + Ns\left(F_{\vec{m}+\hat{e}_j;\vec{n}} - m_jF_{\vec{m}-\hat{e}_j;2,\vec{n}}\right)\Bigg] \\ &\quad \times \sum_{j=1}^{\dim \vec{n}}\Bigg[NU\sum_{\ell=0}^{n_j-2}\binom{n_j}{\ell}\langle \Delta k^{n_j-\ell}\rangle F_{\vec{m};\vec{n}+(\ell-n_j)\hat{e}_j} \\ &\quad + \binom{n_j}{2}F_{\vec{m};2,\vec{n}-2\hat{e}_j} - n_jF_{\vec{m};\vec{n}} + Ns\left(F_{\vec{m};\vec{n}+\hat{e}_j} - n_jF_{\vec{m};2,\vec{n}-\hat{e}_j}\right)\Bigg] \\ &\quad + \sum_{j=1}^{\dim \vec{m}}\sum_{i>j}^{\dim \vec{m}}\Big[F_{\vec{m}+m_i(\hat{e}_j-\hat{e}_i);\vec{n}} - F_{\vec{m};\vec{n}} - m_jF_{\vec{m}-\hat{e}_j+\hat{e}_i;\vec{n}} \\ &\quad - m_iF_{\vec{m}+\hat{e}_j-\hat{e}_i;\vec{n}} + m_jm_iF_{\vec{m}-\hat{e}_j-\hat{e}_i;2,\vec{n}}\Big] \\ &\quad + \sum_{j=1}^{\dim \vec{n}}\sum_{i>j}^{\dim \vec{n}}\Big[F_{\vec{m};\vec{n}+n_i(\hat{e}_j-\hat{e}_i)} - F_{\vec{m};\vec{n}} - n_jF_{\vec{m};\vec{n}-\hat{e}_j+\hat{e}_i} \\ &\quad - n_iF_{\vec{m};\vec{n}+\hat{e}_j-\hat{e}_i} + n_jn_iF_{\vec{m};2,\vec{n}-\hat{e}_j-\hat{e}_i}\Big] \\ &\quad + \sum_{j=1}^{\dim \vec{m}}\sum_{i=1}^{\dim \vec{n}}\Big[F_{\vec{m}+n_i\hat{e}_j;\vec{n}-n_i\hat{e}_i} - F_{\vec{m};\vec{n}} - m_jF_{\vec{m}-\hat{e}_j;\vec{n}+\hat{e}_i;} \\ &\quad - n_iF_{\vec{m}+\hat{e}_j;\vec{n}-\hat{e}_i} + m_jn_iF_{\vec{m}-\hat{e}_j;2,\vec{n}-\hat{e}_i}\Big]. \end{aligned} \tag{F.11}$$

We can collect all of the $F_{\vec{m};\vec{n}}$ terms on the right side to obtain the single term

$$-\left[\sum_j m_j + \sum_j n_j + \binom{\dim \vec{m} + \dim \vec{n}}{2}\right]F_{\vec{m};\vec{n}}. \tag{F.12}$$

Otherwise all of the terms on the right hand side are unique (unless there are duplicate entries in $\vec{m}$ or $\vec{n}$). A Python script that programmatically outputs these equations in a form suitable for import into Mathematica is available from the authors upon request.

References

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

Barton, N.H., Turelli, M., 1991. Natural and sexual selection at many loci. Genetics 127, 229–255.

Begun, D.J., Holloway, A.K., Stevens, K., Hillier, L.W., Poh, Y.-P., et al., 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol. 5, e310.

Brunet, E., Derrida, B., Mueller, A.H., Munier, S., 2006. A phenomenological theory giving the full statistics of the position of fluctuating pulled fronts. Phys. Rev. E 73, 056126.

Brunet, E., Rouzine, I.M., Wilke, C.O., 2008. The stochastic edge in adaptive evolution. Genetics 179, 603–620.

Bürger, R., 1991. Moments, cumulants, and polygenic dynamics. J. Math. Biol. 30, 199–213.

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

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

Comeron, J.M., Kreitman, M., 2002. Population, evolutionary, and genomic consequences of interference selection. Genetics 161, 389–410.
Desai, M.M., Fisher, D.S., 2007. Beneficial mutation selection balance and the effect of genetic linkage on positive selection. Genetics 176, 1759–1798.
Desai, M.M., Fisher, D.S., 2011. The balance between mutators and nonmutators in asexual populations. Genetics 188, 997–1014.
Etheridge, A., Pfaffelhuber, P., Wakolbinger, A., 2007. How often does the ratchet click? facts, heuristics, and asymptotics. In: Blath, P.M.J., Scheutzow, M. (Eds.), Trends in Stochastic Analysis. Cambridge University Press, pp. 365–390.
Ethier, S.N., Kurtz, T.G., 1987. The infinitely-many-alleles model with selection as a measure valued diffusion. In: Lecture Notes in Biomathematics, vol. 70. pp. 72–87.
Ewens, W.J., 2004. Mathematical Population Genetics, second edition. Springer-Verlag, New York.
Falconer, D.S., 1960. Introduction to Quantitative Genetics. Oliver & Boyd, Edinburgh/London.
Feller, W., 1951. Diffusion processes in genetics. In: Proc. Second Berkeley Symp. on Math. Statist. and Prob. Univ. of Calif. Press, pp. 227–256.
Fisher, D.S., 2013. Asexual evolution waves: fluctuations and universality. J. Stat. Mech. P01011. http://dx.doi.org/10.1088/1742-5468/2013/01/P01011.
Franklin, I., Lewontin, R.C., 1970. Is the gene the unit of selection? Genetics 65, 707–734.
Gardiner, C., 1985. Handbook of Stochastic Methods. Springer, New York.
Gessler, D., 1995. The constraints of finite size in asexual populations and the rate of the ratchet. Genet. Res. 73, 119–131.
Good, B.H., Desai, M.M., 2012. The equivalence between weak and strong purifying selection. arXiv:1210.4500.
Good, B.H., Rouzine, I.M., Balick, D.J., Hallatschek, O., Desai, M.M., 2012. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. Proc. Natl. Acad. Sci. 109, 4950–4955.
Gordo, I., Charlesworth, B., 2000. On the speed of Muller's ratchet. Genetics 156, 2137–2140.
Gordo, I., Navarro, A., Charlesworth, B., 2002. Muller's ratchet and the pattern of variation at a neutral locus. Genetics 161, 835–848.
Goyal, S., Balick, D.J., Jerison, E.R., Neher, R.A., Shraiman, B.I., Desai, M.M., 2012. Dynamic mutation-selection balance as an evolutionary attractor. Genetics 191, 1309–1319.
Hahn, M.W., 2008. Toward a selection theory of molecular evolution. Evolution 62, 255–265.
Haigh, J., 1978. The accumulation of deleterious genes in a population. Theor. Popul. Biol. 14, 251–267.
Hallatschek, O., 2011. The noisy edge of traveling waves. Proc. Natl. Acad. Sci. USA 108, 1783–1787.
Hallatschek, O., Korolev, K.S., 2009. Fisher waves in the strong noise limit. Phys. Rev. Lett. 103, 108103.
Higgs, P., Woodcock, G., 1995. The accumulation of mutations in asexual populations and the structure of genealogical trees in the presence of selection. J. Math. Biol. 33, 677–702.
Hill, W.G., Robertson, A., 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8, 269–294.
Hinch, E.J., 1991. Perturbation Methods. Cambridge University Press, New York.
Hudson, R.R., Kaplan, N.L., 1994. Gene trees with background selection. In: Golding, B. (Ed.), Non-Neutral Evolution: Theories and Molecular Data. Chapman & Hall, London, pp. 140–153.
Jain, K., 2008. Loss of least-loaded class in asexual populations due to drift and epistasis. Genetics 179, 2125–2134.
Kimura, M., 1955. Stochastic processes and distribution of gene frequencies under natural selection. Cold Spring Harb. Symp. Quant. Biol. 20, 33–53.
Kingman, J.F.C., 1976. Coherent random walks arising in some genetical models. Proc. R. Soc. Lond. Ser. A 351, 19–31.
Kingman, J.F.C., 1982. The coalescent. Stochastic Processes Appl. 13, 235–248.
Kingman, J.F.C., 2000. Origins of the coalescent: 1974–1982. Genetics 156, 1461–1463.
Kreitman, M., 1983. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature 304, 412–417.
Lewontin, R.C., Hubby, J.L., 1966. A molecular approach to the study of genic heterozygosity in natural populations. ii. amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics 54, 595–609.
Messer, P.W., Petrov, D.A., 2012. The Mcdonald–Kreitman test and its extensions under frequent adaptation: problems and solutions. arXiv:1211.0060.
Moran, P.A.P., 1975. Wandering distributions and the electrophoretic profile. Theor. Popul. Biol. 8, 318–330.
Muller, H.J., 1964. The relation of recombination to mutational advance. Mutat. Res. 1, 2–9.
Nagylaki, T., 1993. The evolution of multilocus systems under weak selection. Genetics 134, 627–647.
Neher, R., Shraiman, B.I., 2011a. Genetic draft and quasi-neutrality in large facultatively sexual populations. Genetics 188, 975–976.
Neher, R.A., Shraiman, B.I., 2011b. Statistical genetics and evolution of quantitative traits. Rev. Modern Phys. 83, 1283–1300.
Neher, R., Shraiman, B.I., 2012. Fluctuations of fitness distributions and the rate of Muller's ratchet. Genetics 191, 1283–1293.
Neher, R., Shraiman, B.I., Fisher, D.S., 2010. Rate of adaptation in large sexual populations. Genetics 184, 467–481.
Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., et al., 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science 337, 100–104.
Neuhauser, C., Krone, S.M., 1997. The genealogy of samples in models with selection. Genetics 145, 519–534.
Nicolaisen, L.E., Desai, M.M., 2012. Distortions in genealogies due to purifying selection. Mol. Biol. Evol. 29, 3589–3600.
Nik-Zinal, S., Loo, P.V., Wedge, D.C., Alexandrov, L.B., et al., 2012. The life history of 21 breats cancers. Cell 149, 994–1007.
O'Fallon, B.D., Seger, J., Adler, F.R., 2010. A continuous-state coalescent and the impact of weak selection on the structure of genealogies. Mol. Biol. Evol. 27, 1162–1172.
Ohta, T., 1992. The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. 23, 263–286.
Ohta, T., Kimura, M., 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. 22, 201–204.
Park, S., Krug, J., 2007. Clonal interference in large populations. Proc. Natl. Acad. Sci. USA 104, 18135–18140.
Pool, J.E., Hellmann, I., Jensen, J.D., Nielson, R., 2010. Population genetic inference from geomic sequence variation. Genome Res. 20, 291–300.
Rambaut, A., Pybus, O.G., Nelson, M.I., Viboud, C., et al., 2008. The genomic and epidemiological dynamics of human influenza a virus. Nature 453, 617–620.
Rouzine, I., Brunet, E., Wilke, C., 2008. The traveling-wave approach to asexual evolution: Muller's ratchet and the speed of adaptation. Theor. Popul. Biol. 73, 24–46.
Rouzine, I.M., Wakeley, J., Coffin, J.M., 2003. The solitary wave of asexual evolution. Proc. Natl. Acad. Sci. USA 100, 587–592.
Santiago, E., Caballero, A., 1998. Effective size and polymorphism of linked neutral loci in populations under directional selection. Genetics 149, 2105–2117.
Sawyer, S.A., Hartl, D.L., 1992. Population genetics of polymorphism and divergence. Genetics 132, 1161–1176.
Slatkin, M., 1972. On treating the chromosome as the unit of selection. Genetics 72, 157–168.
Song, Y.S., Steinrücken, M.A., 2012. A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. Genetics 190, 1117–1129.
Stephan, W., Chao, L., Smale, J.G., 1993. The advance of Muller's ratchet in a haploid asexual population: approximate solutions based on diffusion theory. Genet. Res. 61, 225–231.
Tsimring, L., Levine, H., Kessler, D., 1996. RNA virus evolution via a fitness-space model. Phys. Rev. Lett. 90, 088103.
Turelli, M., Barton, N.H., 1990. Dynamics of polygeneic characters under selection. Theor. Popul. Biol. 38, 1–57.
Van Dyke, M., 1974. Analysis and improvement of perturbation series. Quart. J. Mech. Appl. Math. 27, 423–450.
Walczak, A.M., Nicolaisen, L.E., Plotkin, J.B., Desai, M.M., 2012. The structure of genealogies in the presence of purifying selection: a fitness-class coalescent. Genetics 190, 753–779.
Waxman, D., Loewe, L., 2010. A stochastic model for a single click of Muller's ratchet. J. Theoret. Biol. 264, 1120–1132.
Woodcock, G., Higgs, P.G., 1996. Population evolution on a multiplicative single-peak fitness landscape. J. Theoret. Biol. 179, 61–73.
Zeng, K., Charlesworth, B., 2011. The joint effects of background selection and genetic recombination on local gene genealogies. Genetics 189, 251–266.