Genetic Draft and Quasi-Neutrality in Large Facultatively Sexual Populations

R. A. Neher*,† and B. I. Shraiman*,‡,1

*Kavli Institute for Theoretical Physics, ‡Department of Physics, University of California, Santa Barbara, California 91306, and †Max-Planck Institute for Developmental Biology, 72070 Tübingen, Germany

ABSTRACT Large populations may contain numerous simultaneously segregating polymorphisms subject to natural selection. Since selection acts on individuals whose fitness depends on many loci, different loci affect each other's dynamics. This leads to stochastic fluctuations of allele frequencies above and beyond genetic drift—an effect known as *genetic draft*. Since recombination disrupts associations between alleles, draft is strong when recombination is rare. Here, we study a facultatively outcrossing population in a regime where the frequency of outcrossing and recombination, r, is small compared to the characteristic scale of fitness differences σ. In this regime, fit genotypes expand clonally, leading to large fluctuations in the number of recombinant offspring genotypes. The power law tail in the distribution of the latter makes it impossible to capture the dynamics of draft by an effective neutral model. Instead, we find that the fixation time of a neutral allele increases only slowly with the population size but depends sensitively on the ratio r/σ. The efficacy of selection is reduced dramatically and alleles behave "quasi-neutrally" even for $Ns \gg 1$, provided that $|s| < s_c$, where s_c depends strongly on r/σ, but only weakly on population size N. In addition, the anomalous fluctuations due to draft change the spectrum of (quasi)-neutral alleles from $f(\nu) \sim \nu^{-1}$, corresponding to drift, to $\sim \nu^{-2}$. Finally, draft accelerates the rate of two-step adaptations through deleterious intermediates.

THE genetic diversity of a population is determined by mutation, selection, recombination, and genetic drift, *i.e.*, the stochasticity inherent in reproduction. Understanding how genetic diversity depends on these elements of evolutionary dynamics is central to population genetics, since it allows us to make inferences about the past history and to predict how rapidly populations can adapt.

Population genetic inference focuses on the diversity at putatively neutral sites and assumes that the history of these sites is described by the neutral "coalescent" (Kingman 1982). Coalescent theory models the genealogy of asexual organisms or nonrecombining segments of a genome by positing that lineages merge at random, backward in time, due to common ancestry. Under this assumption, the mean time to the most recent common ancestor, T_C, of the extant N individuals, is $2N$ generations. The coalescence timescale is

very important, since the genetic diversity of the population is given by the number of mutations that occur in all lineages descending from the most recent common ancestor of the population. Genetic diversity is therefore controlled by T_C and hence, under the assumption of neutral evolution, proportional to N. [Coalescent theory has been extended to weak selection (Krone and Neuhauser 1997) and recombination (Hudson 1983; Griffiths and Marjoram 1996).]

However, the prediction that neutral genetic diversity is proportional to N is at odds with observations: Population sizes of different organisms differ by many orders of magnitude, while genetic variation among organisms is fairly constant (Lewontin 1974). To resolve this "paradox of variation", Maynard Smith and Haigh (1974) suggested that selection acting on linked loci might reduce diversity at a neutral locus. Rapid fixation of a novel mutation at a linked locus will perturb the allele frequencies. These perturbations can bring alleles to fixation and, more generally, reduce the coalescence time and hence the average genetic diversity (Kaplan *et al.* 1989; Barton 1998; Gillespie 2001). Such "hitchhiking" of neutral alleles on linked selected loci will dominate over genetic drift in large populations. Since hitchhiking leads to larger perturbations for more closely linked

loci, one expects genetic variation to correlate with the recombination rate, as is indeed observed in *Drosophila* (Begun and Aquadro 1992; Stephan and Mitchell 1992; Andolfatto and Przeworski 2001; Sella *et al.* 2009).

A related effect was described earlier by Hill and Robertson (1966), who studied the reduction in the fixation probability of a novel beneficial mutation because of selection acting at a linked locus. This effect is commonly known as Hill–Robertson interference (Felsenstein 1974). Hitchhiking and Hill–Robertson interference are different aspects of the same phenomenon, one focusing on the effects of linked selection on genetic diversity and the other on the efficacy of selection. While hitchhiking and Hill–Robertson effects are primarily associated with positive selection for novel alleles, purifying selection against deleterious mutations also affects genetic diversity. The elimination of (neutral) alleles linked deleterious mutations is known as background selection. The lower the recombination rates are, the larger is the target for linked deleterious mutations, resulting in stronger background selection (Charlesworth *et al.* 1993; Hudson and Kaplan 1995; Nordborg *et al.* 1996).

Most models used to study Hill–Robertson and hitchhiking effects between positively selected mutations consider only two loci. Deleterious mutations, however, are expected to be much more frequent, and background selection models typically consider many mutations with small deleterious effects. A systematic study of the effect of interference between many weakly selected sites in a mutation/selection/drift equilibrium was presented by McVean and Charlesworth (2000), who used computer simulations of a model of codon bias evolution. They showed that linkage-dependent interference between a large number of weakly selected sites has substantial effects on genetic diversity, fixation probability of mutations, and the degree of adaptation measured as the frequency of preferred codons. This and subsequent computational studies reinforced the understanding that the Hill–Roberson effect reduces the effectiveness of selection and made clear that a quantitative understanding of Hill–Robertson effects in multilocus systems requires an analysis that goes beyond two-locus models (Comeron and Kreitman 2002; Seger *et al.* 2010); see Comeron *et al.* (2008) for a recent review.

It is common to interpret the effect of linked selection in terms of increased variance in offspring number. In this interpretation, linked selection can be thought of as a stochastic force analogous to genetic drift and is often referred to as *genetic drift*—a term coined by Gillespie (2000). Following Hill and Robertson (1966) and Felsenstein (1974), the increased variance in offspring number is often captured by a reduction in the "effective population" size, N_e, in a neutral model (which means enhanced drift and accelerated coalescence). It has, however, been noted that a rescaled neutral model does not consistently explain all observables (Charlesworth *et al.* 1993; Braverman *et al.* 1995; Fay and Wu 2000; McVean and Charlesworth 2000; Barton and Etheridge 2004; Seger *et al.* 2010) and that

different effective population sizes need to be defined depending on the question and timescale of interest (Ewens 2004; Karasov *et al.* 2010). Furthermore, the dependence of N_e on the actual population size and other relevant parameters is not understood (Wiehe and Stephan 1993; Gillespie 2000; Lynch 2007).

Here, we provide analytic results on the effect of draft in a stochastic multilocus evolution model. Instead of a mutation/selection equilibrium considered in McVean and Charlesworth (2000), our focus here is a continuously adapting and facultatively sexual population, like human immunodeficiency virus (HIV) in coevolution with the host's immune system. Our model and its relation to the biology of HIV are described in more detail below. The scope of the model, however, extends beyond HIV and is equally applicable to scenarios where background selection is dominant. Many important and well-studied organisms such as influenza, yeast, and plants are facultatively sexual. Rice, for example, is a partly selfing species and strong selection has acted during its domestication (Caicedo *et al.* 2007). While dominance effects can render the selfing of diploid organisms more complicated than facultatively sexual propagation of haploid organisms (Charlesworth *et al.* 1991; Kelly and Williamson 2000), our analysis still provides a null model on top of which dominance effects can be investigated.

Using computer simulations of an adapting population, we first demonstrate how quantities such as the coalescence time, the fixation probability of beneficial or deleterious mutations, and the allele frequency spectra depend on the rate of outcrossing relative to selection. We also show that our *in silico* observations cannot be described by a neutral model with a reduced effective population size. This is because single genotypes can, through clonal expansion, give rise to a wildly fluctuating number of recombinant genotypes. The distribution is so broad that its variance diverges, which in turn makes an effectively neutral diffusion limit impossible. To provide an analytic understanding of the simulation results, we use a branching process model that allows us to study the stochastic dynamics of novel mutations (neutral, beneficial, and deleterious) as they spread through the population. We calculate fixation probabilities and the typical time to fixation, T_{fix} (and more generally, the probability of attaining n copies after time T), for a new mutant allele, making explicit the dependence on the rate of recombination, fitness variance, and the population size. An important consequence of genetic draft is a qualitatively different frequency spectrum of rare alleles, which we also calculate analytically. Finally, we show that empirical HIV allele frequency spectra are in agreement with our theoretical prediction, confirming the relevance of our model to the dynamics of HIV adaptation.

Model

Our model is inspired by the intrapatient evolution of HIV. We first outline briefly the biology of HIV and then describe

our computational model and the branching process approximation used to study the phenomena analytically.

Intrapatient evolution of HIV

After successful infection, the virus proliferates quickly to levels of $>10^6$ viral genomes per milliliter plasma. This acute phase lasts for several weeks, after which the immune system of the host reduces the number of viral genomes per milliliter plasma to $\sim 100-10^4$. The following chronic infection can last for many years and is largely asymptomatic up to the onset of AIDS. However, even in the chronic phase of infection the total number of viruses produced per day is estimated to be $\sim 10^{10}$, while the generation time is ~ 2 days (Perelson *et al.* 1996; Markowitz *et al.* 2003). The viral population is subject to clearance by cytotoxic T-lymphocyte (CTL)–mediated cell death and antibody-mediated destruction of virus (Paul 2008), which puts the virus under constant selection to change its proteins. CTL epitopes are found throughout the HIV genome (Korber *et al.* 2007) and escape mutations are often selected for with coefficients of a few percent per generation (Asquith *et al.* 2006). Antibodies bind primarily to the products of the envelope gene. In response to antibody recognition, numerous escape mutants arise in the envelope gene and spread through the population driven by selection coefficients of 1% to a few percent (Williamson 2003; Neher and Leitner 2010). Additional, even stronger, selection pressure is put on the viral population by drug treatment, resulting in rapid emergence of resistance during suboptimal therapy (Larder *et al.* 1989; Hedskog *et al.* 2010). Several hundred mutations are implicated in drug resistance (Rhee *et al.* 2003).

HIV carries two copies of its RNA genome (10^4 bp), from which one cDNA is produced and integrated into the host cell genome. Recombination in HIV occurs through frequent template switching of the reverse transcriptase in the process of cDNA (Levy *et al.* 2004) production. In other words, the two genomes of the "diploid" virion are combined to produce a "haploid" cDNA, from which all the viral proteins and the next generations genomes are produced. The rate of recombination is limited by co-infection of host cells with several viruses, which is necessary to produce a heterozygous virus (see Figure 1A). Estimates suggest an effective recombination rate of $\approx 1.5 \times 10^{-5}$ per base per generation (Neher and Leitner 2010), corresponding to a co-infection rate $\leq 10\%$. Viruses therefore undergo clonal amplification most of the time, while different parts of the genome are only weakly linked in the event of "outcrossing" (heterozygote formation, followed by template switching).

Computational model

On the basis of the discussion above our model must include the following elements: a large population, selection at many polymorphic loci, a constant supply of beneficial mutations, and facultative mating with substantial reassortment in the case of outcrossing. Models combining all of these elements have already been established (Rouzine and

Coffin 2005, 2010; Neher *et al.* 2010) and a similar model is used here. In our simulation a (nearly) constant variance of fitness in the population, σ^2, is maintained by a constant supply of beneficial mutations. This results in a continuously adapting population, with a rate of adaptation given approximately by the variance in fitness. Of course, one does not expect constant selection to persist indefinitely (see Mustonen and Lässig 2010 for a discussion of evolution in changing environments): steady conditions on the timescale of fixation of a single allele would suffice for the validity of our analysis.

To simulate population dynamics with the ingredients described above, we implemented discrete time (Fisher–Wright) dynamics of individuals with haploid genomes and a large number of loci that additively define the (log-) fitness x of an individual. Each individual contributes a Poisson-distributed number of gametes to the next generation with mean $\exp(x - \bar{x} - \alpha)$, where $\bar{x}$ is the current mean fitness and α is adjusted to keep the population size approximately constant at N. To implement facultative mating, a fraction r of the gametes are paired at random, the alleles at all of their loci reassorted, and the two resulting recombinant offspring are placed into the next generation. The remaining $1 - r$ fraction of gametes is placed unchanged into the next generation. New beneficial mutations with effect size s_0 ($Ns_0 \gg 1$) are introduced into a randomly chosen individual at rate $NU_b \gg 1$. After equilibration, the fitness distribution in the population is approximately Gaussian with a nearly constant variance and steadily increasing mean fitness. Into this adapting population wave, we introduce additional mutations, neutral, deleterious, or beneficial, and track their dynamics. In particular, we record after what time these mutations reach certain frequency thresholds and the fitness of genotypes on which successful mutations originate. We also measure the allele frequency spectra and the cumulative number of individuals that carried a particular mutation before it goes extinct. A further, and more detailed, description of the implementation is given in *Appendix B*.

Our focus here is on rapidly adapting populations with many concurrent sweeps responsible for the fitness variance σ^2. HIV, as any other organism, also suffers from deleterious mutations that vastly outnumber the beneficial mutations. Deleterious mutations also contribute to fitness variation and increase σ^2. For the present investigation, it does not matter whether fitness variation is due to sweeping beneficial mutations, deleterious mutations subject to purifying selection, or a combination of the two. The two extreme scenarios are illustrated in Figure 1B. We shall see that the fate of novel mutations depends on how the fitness of clones changes with respect to the mean fitness of the population. It is irrelevant whether this fitness difference decreases because the mean fitness increases or because the fitness of the clone decreases due to deleterious mutations.

Mathematical model

To understand the results of the computer simulations, we analyze analytically an idealized model, which describes
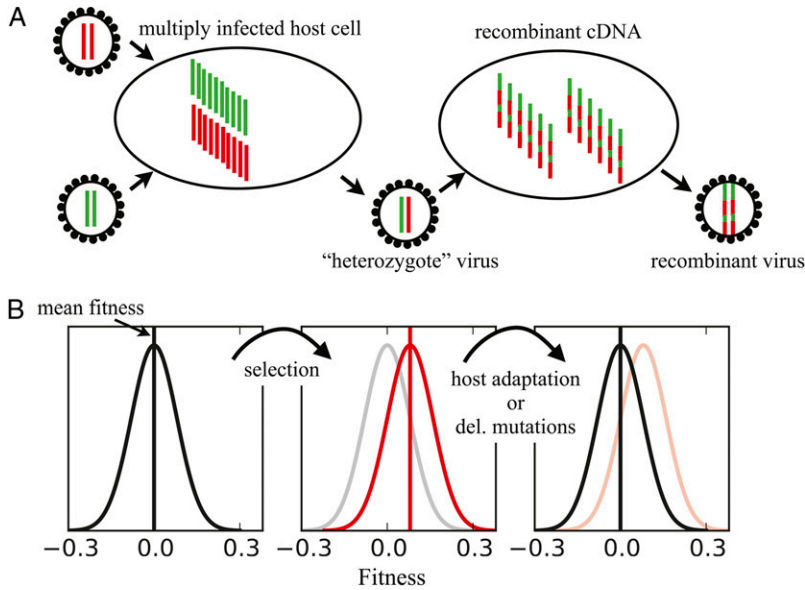
Figure 1 (A) The process of recombination in HIV. Diversifying recombination occurs only in "heterozygote" virions, which require multiple infections of the host cell in the previous generation. This is rare, leading to an effectively facultative sexual population. (B) The dynamics of the fitness distribution. Selection and mutation are shown as separate steps, for simplicity. Selection, acting between the left and center panels, results in a shift of the fitness distribution toward higher fitness. In a deleterious mutation/selection balance, this increase in fitness is compensated by deleterious mutations occurring during the step leading from the center to the right panel. If deleterious mutations are not the dominant force, but continuous adaptation to a coevolving host is modeled, the adaptation of the host counterbalances the fitness advance of the population.

the propagation of a new allele (under selection, drift, and occasional recombination onto a different genetic background) within a population with a fitness distribution approximated by a Gaussian traveling wave. This is justified if many beneficial mutations are sweeping through the population at the same time (which requires $NU_{\rm b} \gg 1$ and is true for HIV) and is supported by our numerical simulations. Gaussian fitness distributions are expected in large asexual populations (Tsimring *et al.* 1996; Rouzine *et al.* 2003; Desai and Fisher 2007) and in the context of background selection [the deterministic equilibrium distribution in the case of multiplicative selection is Poisson (Haigh 1978), which is well approximated by a Gaussian in the limit of interest]. Recombination stabilizes the Gaussian form if outcrossing rates are much larger than the inverse of the coalescence times (Cohen *et al.* 2005; Rouzine and Coffin 2005); see supporting information, Figure S1. Selection moves this Gaussian distribution toward higher fitness at a rate v equal to the variance in fitness σ^2 (Fisher's "fundamental theorem" of natural selection).

Branching process approximation: The fate of a novel allele is decided mainly at early times when the allele is rare and the finite population size constraint can be neglected. In this case the dynamics of the novel allele are governed by a stochastic branching process that in addition to the birth/death events also accounts for the recombination process that transfers the novel allele to different genetic backgrounds with different fitness x' (Barton 1995a; Neher *et al.* 2010). Within the branching process approximation the probability of finding n copies of the allele (with effect s on fitness) anywhere in the population at time T, given that there were k copies on genotypes with fitness x at time t, obeys the equation

$$
\begin{aligned}
-\partial_t p(n,T \,|\, k,t,x) \\
= {}&- k(B+D+r)p(n,T \,|\, k,t,x) \\
&+ kBp(n,T \,|\, k+1,t,x) + kDp(n,T \,|\, k-1,t,x) \\
&+ rk \sum_{n'} \int dx' K_{xx'} p(n-n',T \,|\, k-1,t,x)p(n',T \,|\, 1,t,x')
\end{aligned}
\tag{1}
$$

(see File S1), where $B = 1 + x - \bar{x}(t) + s$ is the birth rate, $D = 1$, the rate at which individuals die (used to set the unit of time), and r is the outcrossing rate. The first term describes the probability flux out of state (k, t). The second term corresponds to a birth, which happens with rate kB and in which case the final state (n, T) is reached with probability $p(n, T \,|\, k + 1, t, x)$. Analogously after a death, the probability of reaching (n, T) is $p(n, T \,|\, k - 1, t, x)$. The last term describes the process of outcrossing with the offspring fitness distribution parameterized by a recombination function $K_{xx'}$, which depends on the details of the model.

In a model of outcrossing where offspring are produced by mating two individuals at random and reassorting all of their genes, the recombination kernel $K_{xx'}$, after averaging over the mating partner, is given by

$$
K_{xx'} = \sqrt{\frac{2}{3\pi\sigma^2}} \exp\left[-\frac{\left(x' - \bar{x}(t) - (x - \bar{x}(t))/2\right)^2}{3\sigma^2/2} \right]
\tag{2}
$$

In this model, known as the *infinitesimal model*, the distribution of the offspring fitness is centered around half the fitness of the parent carrying the focal allele, measured relative to the instantaneous mean fitness $\bar{x}(t)$ (Bulmer 1980; Barton 2009) (the model was referred to as the *full recombination model* in Neher *et al.* 2010). (Note that the variance of offspring fitness relative to parental mean is $\sigma^2/2$. The variance of $3\sigma^2/4$ in Equation 2 is the result of averaging over the mate.) We assume that individuals are polygamous;

i.e., each offspring is the result of an independent mating event.

For most of the analytic calculations, we employ an even simpler recombination model, where offspring fitness is simply a random sample from the population, independent of the parents:

$$K_{xx'} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(x' - \bar{x}(t)\right)^2}{2\sigma^2} \right] \tag{3}$$

In this *communal recombination* model new offspring are assembled from the genetic diversity in the entire population; *i.e.*, the novel genotype is drawn from a product distribution with the current allele frequencies in the population (Barton 2009; Neher *et al.* 2010). Note, however, that this model does not lead to instantaneous elimination of correlations between loci (linkage disequilibrium), since only a fraction of the population outcrosses in a single generation. The infinitesimal and communal models have been shown to yield very similar results in Neher *et al.* (2010). We confirm this again by simulation of both models. Even though this model is a drastic idealization, it might resemble rather closely the recombination process in HIV. Since recombination in HIV depends on co-infection of T cells with several viruses, most recombination occurs in centers with high concentrations of virus (Jung *et al.* 2002). In these centers, lineages might undergo several successive outcrossing events, effectively producing a cloud of offspring that contain genetic material from a large number of parents.

Results

Simulation results

Figure 2 shows simulation results for the fixation time of neutral mutations and the fixation probability of beneficial and deleterious mutations for populations of different size and different outcrossing rates. At low r/σ, fixation times (Figure 2A) are reduced by orders of magnitude below the neutral expectation of $2N$ and reach the latter only for $r/\sigma \geqslant 1$, an effect already observed in Charlesworth *et al.* (1992). The scaling of the data in Figure 2A reveals that fixation times are not proportional to the population size at low r/σ, in which case the curves would lie on top of each other. Genetic draft has an equally dramatic effect on the efficacy of selection, which is shown in Figure 2B. The fixation probabilities of beneficial and deleterious mutations are only slightly perturbed from N^{-1} if $r/\sigma < 1$, even though $|Ns|$ ranges up to 80. The fact that genetic draft reduces fixation times and the efficacy of selection is of course well known and it is customary to describe this effect by a neutral model with a reduced population size, *i.e.*, treating draft as if it were just an enhancement of genetic drift. This, however, is often not appropriate, as draft causes fluctuations of a very different nature from those of genetic drift. To illustrate the problem, we define $2N_e$ as the fixation time of

neutral mutations T_{fix}. With this definition, a mutation with effect size s should fix with probability $sT_{\text{fix}}/N(1-e^{-sT_{\text{fix}}})$. However, the observed fixation probability does not follow this expectation if $r \leq \sigma$, as shown in Figure 3A (note the logarithmic scale).

Another striking feature where a neutral model with a reduced effective population size fails to capture the effects of draft is the allele frequency spectrum, shown in Figure 3B for various ratios of r/σ. For small r/σ, the frequency spectrum $f(\nu)$ of neutral alleles falls off much more rapidly than the prediction of the diffusion theory: it decays as ν^{-2} with frequency instead of ν^{-1}. This effect is often referred to as "excess of rare variants" Braverman *et al.* (1995), but should, in fact, more aptly be called a "lack of common variants". Very similar dependencies of the fixation time, fixation probabilities, and allele frequency spectra on
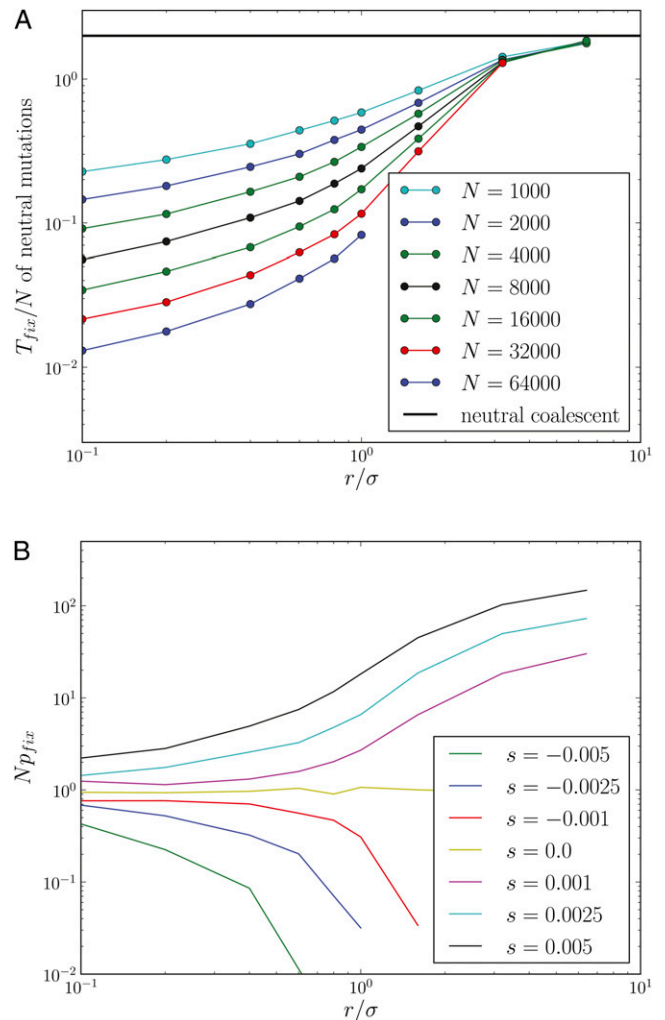

Figure 2 Genetic draft dramatically reduces fixation times and the efficacy of selection. (A) Mean fixation times T_{fix} of neutral mutations, normalized by the population size N, as a function of r/σ for different N. For $r/\sigma < 1$, T_{fix} increases only slowly (sublinear) with N. (B) The fixation probability of beneficial and deleterious mutations relative to that of neutral mutations. Even though $|Ns| > 50$, fixation probabilities are close to N^{-1} at low outcrossing rates ($N = 16,000$).
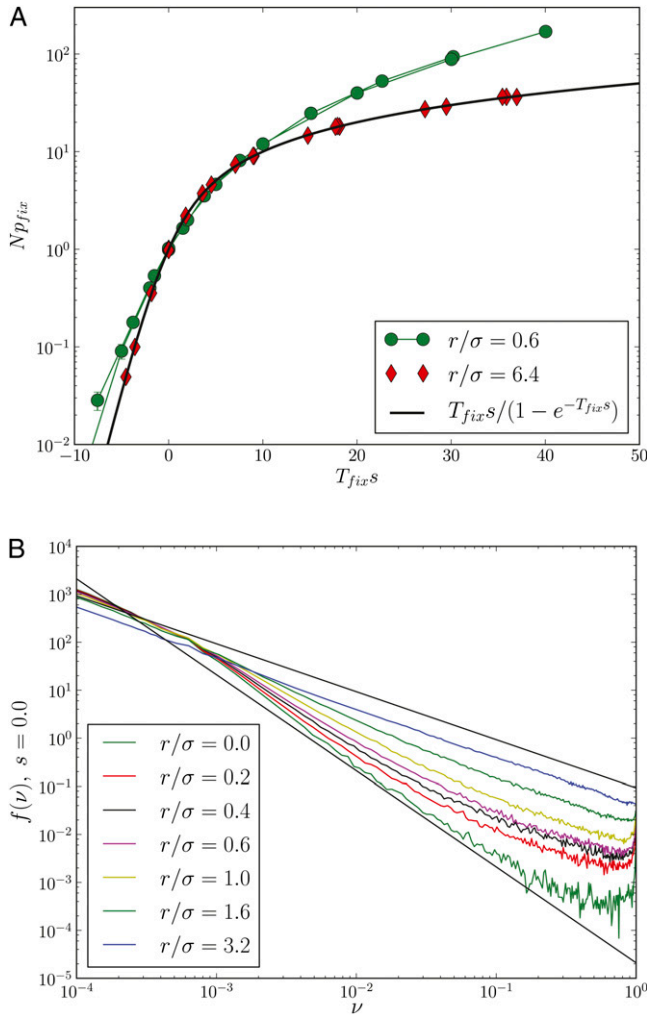
Figure 3 Fixation probabilities and allele frequency spectra are inconsistent with a neutral model with reduced effective population size. (A) The rescaled fixation probability Np_{fix} of mutations with effect size s as a function of $T_{fix}s$, where T_{fix} is the fixation time of neutral mutations. In a neutral model, $T_{fix} = 2N$, and hence it seems sensible to define an effective populations size $N_e = T_{fix}/2$. The fixation probability of a mutation in a neutral model is indicated as a black solid line. While the effective neutral model describes p_{fix} well for $r/\sigma = 6.4$, it fails for $r/\sigma = 0.6$ (note the log scale). (B) The distribution of derived allele frequencies $f(\nu)$ in a population of $N = 32,000$ individuals for different r/σ. For $r/\sigma < 1$, allele frequency spectra fall off much more steeply than the prediction $f(\nu) \sim \nu^{-1}$ of the neutral model. In fact, the distribution falls off more like $f(\nu) \sim \nu^{-2}$, which is the leading behavior predicted by Equation 8. All curves are normalized to unit area and the two power laws are indicated by straight lines.

r/σ are observed if fitness variation is not due to multiple selective sweeps but due to purifying selection. Simulation results for such a background selection scenario are shown in Figure S2 and Figure S3.

To rationalize the observed behavior we solved analytically a simplified branching process model describing the dynamics of alleles in an adapting population; see *Mathematical model*. This branching process solution shows explicitly how observables depend on population parameters and elucidates the difference between draft and drift.

Analytic results for the branching process model

The key to understanding the qualitatively different behavior at low r/σ compared to $r/\sigma \gg 1$ is the fact that genotypes with fitness x can expand clonally if $x - \bar{x} - r > 0$, *i.e.*, if their growth rate exceeds the rate at which they outcross. In contrast to the case of $r/\sigma \gg 1$, for $r/\sigma \leq 1$ the condition for clonal expansion is fulfilled for a substantial fraction of the genotypes in a finite population. A genotype establishes with probability $x - r$ [setting $\bar{x}(t_0) = 0$] and is clonally amplified so that its copy number subsequently grows as

$$n(T,x) \approx \frac{1}{x - r} e^{(x-r)T - \sigma^2 T^2/2}, \tag{4}$$

where T is the time since establishment (Desai and Fisher 2007). The term $\sigma^2 T^2/2 = \int_0^T dt\, \bar{x}(t)$ accounts for the increasing mean fitness, $\bar{x}(t) = \sigma^2 T$, of the rest of the population, competing with the clone. After establishment, a fit genotype therefore gives rise to a clone whose size has a Gaussian profile in time, as illustrated in Figure 4A. During its lifetime, an individual clone spawns an average of $\xi = r \int_0^\infty dt\, n(t)$ recombinant genotypes. For $x - r \gg \sigma$ one finds $\xi \approx \sigma(x-r)^{-1} e^{(x-r)^2/2\sigma^2}$, which increases rapidly with increasing $x - r$. In the *Discussion*, we develop this intuition into a simplified model of the genealogy of clones. Figure 4B shows the copy number of the mutant allele on different genetic backgrounds as a function of time in the continuous-time branching process model. The observed stochastic trajectory shows the features embodied in the simplified effective model: an anomalously fit genotype gives rise to many recombinant offspring genotypes, resulting in large fluctuations in the copy number of the mutant allele.

The dynamics of the distribution $p(n, T \mid x)$ of a mutant allele, given that it arose on a genotype with fitness x ($k = 1$, $t = 0$), can be described by the branching process defined in Equation 1. We proceed by analyzing the fixation probability, the typical time to fixation, and the allele frequency spectra for neutral, beneficial, and deleterious mutations for the case $r/\sigma \leq 1$. In the opposite limit of rapid outcrossing, the effect of selection during the lifetime of a genotype is small and Equation 1 can be analyzed perturbatively in σ/r. This perturbation analysis was performed in Neher *et al.* (2010). It was shown that the fixation probabilities of beneficial mutations are reduced by a factor $1 - \alpha\sigma^2/r^2$, where α is a numerical factor of order one and depends on the details of the recombination model. This is consistent with the classical argument of Robertson (1961) showing that the cumulative effect of draft is proportional to the square of the degree of linkage (Santiago and Caballero 1998; Barton 2009). The analysis and derivation of the results are mainly relegated to *Appendix A*; here we present and discuss the results.

Extinction and survival of neutral mutations

The fundamental quantity describing the stochastic dynamics of an allele is the distribution, $p(n, T)$, of the number of
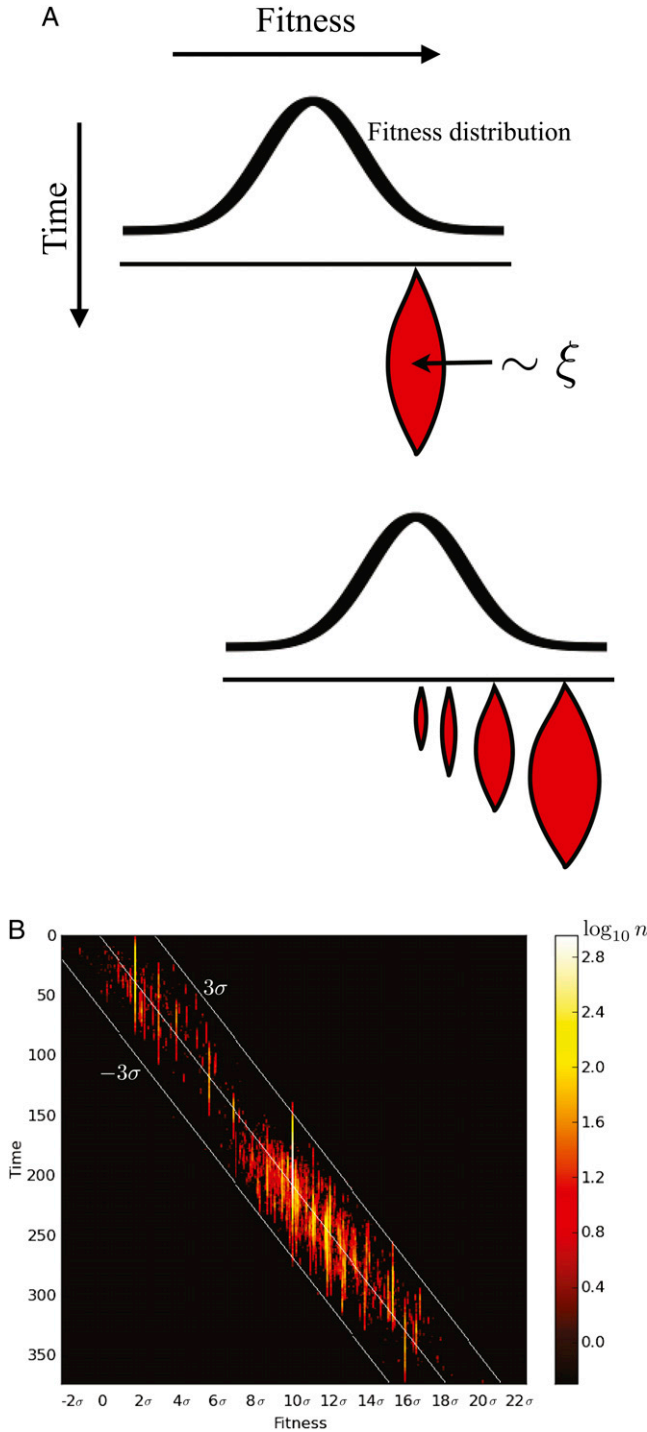
Figure 4 Clonal expansion results in large fluctuations. (A) A clone with high initial fitness can grow to large numbers before going extinct. Large clones give rise to many recombinant daughter clones. (B) A corresponding trajectory obtained by simulation of the branching process equation; see *Model*. The mutant allele originated on a background with fitness $x - \bar{x}(t) = 2\sigma$ at time $t = 0$. Each subclone of the population appears as a vertical line whose horizontal position indicates its fitness x. The size of a clone (color coded) rises as long as $x - \bar{x}(t) - r > 0$ and shrinks after the population wave has passed. The larger a subclone is, the more daughter clones it produces. The mean fitness $\bar{x}(t)$ as well as $\bar{x}(t) \pm 3\sigma$ indicated as white lines.

copies that exist T generations after the founding mutation. Note that the mutation could have arisen on genotypes with different background fitness x, and $p(n, T)$ is the average over all possible x at time $T = 0$. Of particular importance is the survival probability $P_S(T) = 1 - p(0, T)$. Figure 5 shows $P_S(T)$ for a neutral mutation for different outcrossing rates in the communal recombination model, obtained by a numerical solution of Equation A1. Initially, $P_S(T)$ decays as $(1 + T)^{-1}$, regardless of the outcrossing rate. This is the same behavior seen in homogenous or completely neutral populations. Selection on the heterogenous backgrounds starts influencing the fate of the allele at times $T > \sigma^{-1}$, *i.e.*, the reciprocal of the typical selection differentials in the population. After this initial transient, the survival probability depends strongly on the relative magnitude of the outcrossing rate and the fitness variance in the population, decaying faster for smaller r/σ. We show in *Appendix A* (Equation A8 and *Short-time solution* section), that the time dependence of the survival probability is given by

$$P_S(T) \sim \begin{cases} \sigma e^{-(\sigma^2/2r^2)\log^2(r^3\sigma^{-2}T)} & \sigma/\sqrt{\log N} < r < \sigma \\ (1+T)^{-1} & r \gg \sigma. \end{cases} \tag{5}$$

Consistent with our previous argument, we find that for $r \gg \sigma$ neutral mutations are unaffected by selection on the background, reproducing the familiar result $P_S(T) \sim T^{-1}$. However, a strong deviation from this neutral behavior occurs as soon as r/σ approaches 1. In the regime with $r < \sigma$, the ratio r/σ enters the dynamics of $P_S(T)$ as a prefactor of $\log^2 T$ in the exponent, resulting in faster extinction at lower r. (Yet another regime appears at even smaller r for $r/\sigma < 1/\sqrt{\log N}$ and is discussed in File S1). Equation 5 was derived using the communal recombination model. In Figure A1 we show, via simulation of the branching process, that the infinitesimal model exhibits the same asymptotic dependence on parameters.

Dynamics of surviving alleles and fixation times

Given that a new allele is not lost after T generations, what is the mean number of its copies? The answer to this question is surprisingly simple and will inform us about the scaling of fixation times of neutral alleles. The unconditional expected number of copies after time T is obtained by multiplying Equation 1 by n, summing over n, and averaging over the distribution of x at $T = 0$. One finds the simple expression

$$\overline{\langle n(T) \rangle} = e^{sT}, \tag{6}$$

which reduces to $\overline{\langle n(T) \rangle} = 1$ in the neutral case. Here, $\langle ... \rangle$ denotes the average over $p(n, T)$, while the overbar denotes the average over the fitness of the founding genotype. Hence, the expected number of copies depends neither on recombination nor on the speed of adaptation. It depends only on the intrinsic fitness effect of the novel allele. However, $\overline{\langle n(T) \rangle}$ is a product of the survival probability, times the average n conditioned on survival. For neutral mutations,
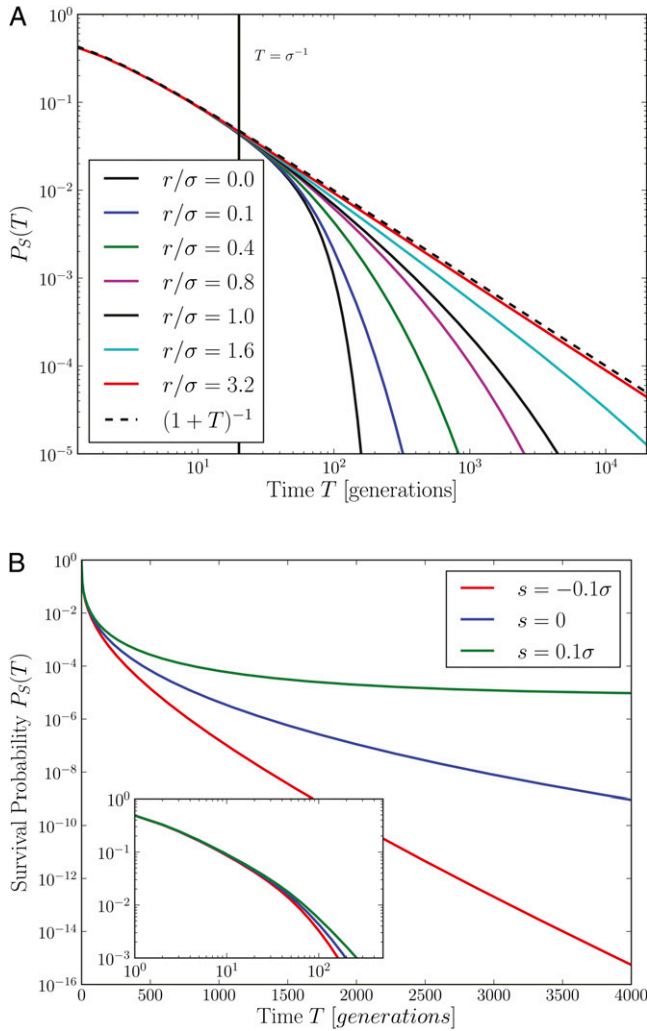
Figure 5 Survival of novel mutations in the branching process model. (A) The survival probabilities of neutral mutations in the communal recombination model for different outcrossing rates as a function of time (obtained through numerical solution for the generating function of the branching process). For large r/σ, the survival probability closely follows the result for a single neutral locus $P_S = (1 + T)^{-1}$ (note the log–log scale). For $r/\sigma \leq 1$, $P_S(T)$ decays much more rapidly as soon as $T \gg \sigma^{-1}$. (B) Illustration of how an intrinsic fitness effect s affects the survival probability for $r/\sigma = 0.4$ (note the log-linear scale). At short times, beneficial and deleterious mutations behave like neutral mutations. After a crossover time, which depends on s, the survival probability of deleterious mutations continues to decay exponentially with rate s, while the survival probability of beneficial mutations becomes time independent.

the mean n conditioned on survival is therefore the reciprocal of the survival probability, and when $P_S(T)$ is of order N^{-1}, a nonextinct allele has reached copy numbers of $\sim N$ and has essentially fixed. The first passage time, T_{fp}, to allele frequency v, i.e., to vN copies, is given asymptotically by

$$T_{\text{fp}} \sim \begin{cases} \sigma^2 r^{-3} e^{r\sigma^{-1}\sqrt{2\log\sigma vN}} & \sigma/\sqrt{\log N} < r < \sigma \\ vN & r \gg \sigma. \end{cases} \quad (7)$$

We see that the time to reach allele frequency v scales very differently with N in the two regimes. Instead of diffusive

dynamics with diffusion constant $\sim N^{-1}$, the rapid turnover of individuals by selection at $r < \sigma$ results in large fluctuations, which can rapidly propel a mutation to large numbers. Simulation results confirming the validity of the argument that inverse survival probability can be used as a proxy for typical copy numbers, as well as the adequacy of modeling allele frequency dynamics by a branching process, are shown in Figure A1.

A different approach to calculate fixation times in a model of continuous adaptation with facultative outcrossing has been developed in Rouzine and Coffin (2007, 2010). Rouzine and Coffin use a coalescence approach based on the clonal structure of the population to calculate fixation times and obtain the same exponential dependence on $r\sigma^{-1}\sqrt{2\log Nr}$.

Allele frequency spectra

In addition to changing the dependence of the fixation time on the population parameters, draft also results in a qualitatively different allele frequency spectrum. This has been noted previously (Braverman *et al.* 1995; Fay and Wu 2000; Gillespie 2000; McVean and Charlesworth 2000) but to our knowledge never calculated explicitly. Figure 3B shows the distribution of derived neutral allele frequencies $v = n/N$ at different outcrossing rates measured in individual-based computer simulations. At high outcrossing rates, we observe $f(v) \sim v^{-1}$ as expected for mutations subject to random drift alone. However, when draft is dominant, $f(v)$ decays more steeply as

$$f(v) \sim \frac{e^{r\sigma^{-1}\sqrt{2\log N\,v\sigma}}}{v^2\log N v\sigma}. \quad (8)$$

This steep decay persists up to a crossover frequency that depends on r/σ and the population size. Beyond this crossover frequency, the finite population size constraint becomes important. The calculation of the allele frequency spectra is detailed in *Appendix A, Distribution of the allele copy number* and the result for the copy number distribution is given in Equation A22.

The $f(v) \sim v^{-2}$ behavior can be understood heuristically by considering the sizes of clones containing the allele. Let us consider the asexual case with $r = 0$. To calculate the copy number distribution of a novel allele, T generations after it originated, we have to average over the initial fitness x of the genotype that it could have arisen on. Given that the allele arose on a genetic background with fitness x, its average number after T generations is given by Equation 4. Since we know that the distribution of x is $P(x, 0) = (2\pi\sigma^2)^{-1/2}e^{-x^2/2\sigma^2}$, we can calculate the expected distribution of n after time T within the deterministic approximation of Equation 4. One finds that $p(n, T) \sim n^{-3/2}\exp(-\sigma^2 T^2/8 - \log^2(n)/2\sigma^2 T^2)$. The allele frequency spectrum is an average over alleles of different age, T, and the above expression therefore has to be integrated over T, which yields $f(n) \sim n^{-2}$. In a facultatively sexual population with $r < \sigma$, allele frequency spectra are also driven by

the clonal expansion of individual genotypes. However, in contrast to the asexual case, additional clones are constantly seeded through outcrossing so that the novel allele resides in many different clones of different ages, similar to the average over asexual clones of different ages. For this reason, both the copy number spectrum of individual alleles and the frequency spectrum averaged over alleles of different ages have a leading behavior $p(n, T) \sim n^{-2}$ and $f(\nu) \sim \nu^{-2}$. This reasoning is confirmed by the branching process calculation given in *Appendix A*.

In addition to a much steeper spectrum at low frequencies, the spectra of neutral and deleterious mutations are nonmonotonic and increase again at frequencies near one (Fay and Wu 2000). On a linear chromosome with isolated sweeps that affect tightly linked neutral variants, this effect is easy to understand: A tightly linked variant is brought close to fixation, resulting in a pile-up of derived variants at frequencies close to one. In our case with many unlinked sweeps in a facultatively mating population the effect is less intuitive, but ultimately of similar origin.

Selection efficiency and quasi-neutrality

The rapid and erratic dynamics of alleles incurred by selection on other loci affect not only the time to fixation, but also the ability of selection to prune deleterious mutations and fix beneficial mutations. Mutations behave essentially neutrally, as long as selection on the intrinsic effect of the allele is outweighed by fluctuations. A beneficial mutation has established when it has risen to sufficient numbers that future extinction is improbable and the allele goes to fixation deterministically. Without draft, a beneficial mutation is established when it has reached copy numbers of $\sim s^{-1}$. With additional fluctuations through draft, however, the allele is more likely to go extinct and has to rise to much higher numbers to establish. Similarly, draft can propel deleterious mutations to copy numbers they would never reach under drift alone. This reduction of the efficacy of selection is a hallmark of Hill–Robertson interference (Hill and Robertson 1966) and was mainly studied in obligate sexuals (Barton 1994; McVean and Charlesworth 2000). Here, we explore these effects in facultative outcrossers.

Since an allele that has survived longer is likely to be present in more copies, the influence of stochastic fluctuations and hence the rate at which the mutant allele goes extinct decrease over time. This is explicit in the time derivative of the log survival probability, which in the regime $r < \sigma$ is given by

$$\partial_T \log P_S(T) \approx s - \frac{\sigma^2 \log(r^3 \sigma^{-2} T)}{r^2 T} \tag{9}$$

(see *Appendix A, Beneficial and deleterious mutations*). The first term s accounts for the deterministic bias due to the intrinsic effect of the allele, which causes accelerated extinction if $s < 0$ and preferential survival if $s > 0$ (compare with Figure 5B). The second term accounts for the stochastic

forces on the allele, whose importance decreases with T. The time after which the allele's fate is dominated by selection on its intrinsic fitness can be roughly determined by equating the two terms in Equation 9. In the regime of intermediate outcrossing rates $r < \sigma$, we find this crossover time to be $T^* \approx \sigma^2 \log(r/s)/r^2 s$. If fixation of a neutral mutation occurs before that time, selection has little effect and the fate of the mutation is dominated by fluctuations until fixation. Hence, we find a window of quasi-neutrality for mutations with selection coefficient smaller than s_c

$$|s| < s_c = \frac{\sigma^2 \log(r/s)}{r^2 T_{\text{fix}}} \sim r e^{-r\sigma^{-1}\sqrt{2\log \sigma N}}, \tag{10}$$

within which the fixation probability of a mutation is $p_{\text{fix}} \sim N^{-1}$. Beneficial mutations with effects larger than s_c have a chance of fixation $p_{fix} = r^{-1}\log(r/s)\exp[-\sigma^2\log^2(r/s)/2r^2]$ as already found in Neher *et al.* (2010). Propagation of deleterious mutations with $|s| < s_c$ is also dominated by fluctuations, resulting in allele frequency spectra similar to neutral mutations. Only for $|s| > s_c$ does the allele frequency spectrum $f(\nu)$ decay exponentially as expected for deleterious mutations. Fixation of deleterious mutations is exponentially suppressed with $p_{\text{fix}} \sim N^{-1}e^{-|s|T_{\text{fix}}}$. Simulation results, qualitatively showing this behavior, are presented in Figures 2B and 3.

A related phenomenon is observed in the context of a linear chromosome in the limit where only one sweep affects a polymorphism at any given time. Barton (1994) showed that a beneficial allele that is subject to many *sequential*, weakly linked interfering sweeps has little chance of fixation if its selection coefficient is below a critical value. The critical value reported by Barton is $s_c \sim \sigma^2/R$, where σ^2/R is the fitness variance of sweeps per map length. In this approximation, the frequency of the focal allele is repeatedly reduced by a factor that depends on the degree of linkage and the strength of the interfering sweeps: this results in an effective reduction in the growth rate of the frequency of the focal allele to $s - s_c$. In the case of numerous *concurrent* sweeps that we consider, the situation is different because the focal allele is spread over many different backgrounds subject to selection. The older the allele is, the more likely it is to be spread over many backgrounds, attenuating the effect of interference. Hence we find an additional dependence of s_c on the typical fixation timescale and the extent of recombination that occurs during this time.

Stochastic tunneling and complex adaptations

Consider an adaptation process where two mutations at the same locus are needed to confer a fitness advantage, but either mutation in isolation is neutral or deleterious. The rate of this process is given by the probability per unit time that the second mutation occurs in an individual already carrying the first mutation. More generally, the probability that the double-mutant genotype never existed up to a time T is

$$P(T) = e^{-\mu \int_0^T dt(n_1(t)+n_2(t))}, \tag{11}$$

where μ is the mutation rate and $n_i(t)$ is the number of individuals carrying mutation i. The deleterious single mutants form a transient subpopulation in the large background population, which we refer to as a mutant "bubble". The important quantity determining the rate of such secondary events is the distribution of the cumulative number of single mutants $w(T) = \int_0^T dt n(t)$, i.e., the integrated bubble size. For homogeneous background populations, where genetic drift dominates the dynamics of neutral alleles, this quantity has been calculated by Iwasa *et al.* (2004), Weissman *et al.* (2009, 2010), and Lynch (2010). In most cases of interest the rate of secondary events is small, so that the probability of tunneling in a single bubble is low. Hence tunneling takes longer than the lifetime of a typical bubble and only the long time limit of $w(T)$ is important. In analogy to the results presented above, we calculate the moment-generating function $\Phi(z) = \langle e^{-zw} \rangle$ of the bubble size w, where $\langle ... \rangle$ denotes the average over the bubble size distribution $P(w)$. The calculation given in *Appendix A, The double mutation probability: stochastic tunneling* yields for the $\Phi(z)$

$$\Phi(z) = \begin{cases} z e^{r\sigma^{-1}\sqrt{-2\log z}} & |s| \ll r e^{-r\sigma^{-1}\sqrt{-2\log z}} \\ \frac{z}{|s|} & |s| \gg r e^{-r\sigma^{-1}\sqrt{-2\log z}}. \end{cases} \tag{12}$$

The result for strongly deleterious mutations is the same as in homogenous populations (Weissman *et al.* 2009). The crossover between the two regimes occurs essentially at the quasi-neutrality threshold, with N replaced by z^{-1}. The generating function $\Phi(z)$, which is also the Laplace transform of $P(w)$, is of immediate practical relevance, since it is exactly the probability that a secondary event with rate z does not occur within a single bubble. From the Laplace transform, we calculate the distribution of $P(w)$, which in the draft-dominated regime we find to be

$$P(w) \sim \frac{1}{w^2 \sqrt{2\log w}}. \tag{13}$$

The latter has to be compared to the result for drift alone: $P(w) \sim w^{-3/2}$. Both of these behaviors are clearly seen in the simulation results shown in Figure 6. Draft causes large neutral bubbles to become rarer, but also dramatically enhances the probability of large bubbles for deleterious (quasi-neutral) alleles.

Discussion

The importance of hitchhiking and draft as a source of fluctuations was first discussed by Maynard Smith and Haigh (1974) as a possible explanation for the apparent insensitivity of genetic variation on the census population size, dubbed the paradox of variation (Lewontin 1974). Draft is expected to dominate over genetic drift in very large populations and hence is the factor determining the level of
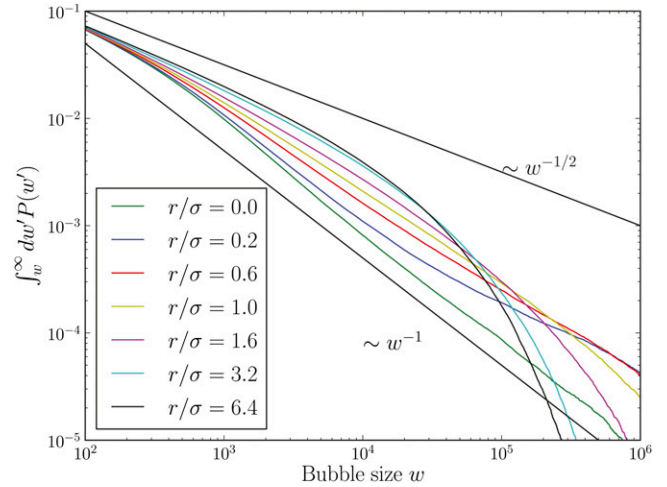

Figure 6 The cumulative distribution of "bubble sizes", i.e., the integrated copy number $w = \int_0^\infty dt n(t)$, of deleterious ($s = -0.005$) alleles. The simulated bubble size distributions for $r/\sigma \ll 1$ agree with the predicted behavior $P(w) \sim w^{-2}$, which corresponds to a cumulative distribution $\sim \omega^{-1}$. For $r \gg \sigma$ the size of bubbles of deleterious alleles is cut off exponentially. While small bubbles are rarer at $r \ll \sigma$ compared to $r \gg \sigma$, this relation is reversed for large bubbles when the exponential cuts below the power law ($N = 16,000$).

genetic diversity and the efficacy of selection (Barton 1995b; Gillespie 2001). It was shown to have appreciable effects on genetic diversity in *Drosophila* (Sella *et al.* 2009). Draft is expected to be even more pronounced in facultatively outcrossing organisms, where alleles stay associated with each other for longer periods of time. Such organisms include many plants, fungi, nematodes, and viruses, including the human pathogens influenza and HIV.

We have studied draft in continuously adapting facultatively outcrossing populations, using direct computer simulation of a model inspired by intrapatient evolution of HIV. We explain the simulation results by analyzing a simplified branching process model, which is amenable to analytic calculations that elucidate how fixation time, the efficiency of selection, and allele frequency spectra depend on fundamental parameters of the population dynamics. The simplification that makes these analytic results possible is a "large number" limit: due to the large number of concurrent sweeps, the distribution of fitness in the population assumes a Gaussian form and translates toward higher fitness with constant velocity. If the outcrossing rate, r, is comparable to or smaller than typical fitness differences, σ, between individuals in the population, clonal expansion of genotypes results in a heavy-tailed distribution of recombinant offspring with dramatic effects. In particular, we find that alleles with selection coefficients smaller than $|s_c| = \sigma^2 \log(r/s)/(r^2 T_{\text{fix}})$ behave *quasi-neutrally*: their fate is dominated by fluctuations and is nearly independent of the intrinsic allele fitness s. Yet because selection acting on the allele is masked not by random drift, but by *draft*— transient associations with chromosomal "backgrounds" with different fitness—the behavior is quantitatively different

from true neutrality. The conventional dynamics dominated by genetic drift are realized only in the limit of no linkage or complete neutrality (including the background). To emphasize this distinction we refer to the draft-dominated regime as *quasi-neutrality*.

In addition to a reduced efficacy of selection, draft reduces fixation and coalescence times, which instead of growing linearly with the population size, increase much more slowly. The probability that a given neutral locus is polymorphic with an allele at intermediate frequency is roughly given by the product of the neutral mutation rate, μ, and the coalescence time, T_C. Hence we expect the neutral diversity $\Theta \sim \mu\sigma^2 r^{-3} e^{r\sigma^{-1}\sqrt{2\log\sigma N}}$. Details of the expected heterozygosity, however, depend on the allele frequency spectrum. The latter is also the most informative quantity that can be evaluated from a static population snapshot. We have shown that in our model, allele frequency spectra decay as $\nu^{-2}\exp[r\sigma^{-1}\sqrt{2\log N\nu\sigma}]$ which is much steeper than ν^{-1} predicted by drift. As a consequence, in contrast to the classic neutral theory (Ewens 2004), one should expect the number of polymorphic loci to increase almost linearly (rather than logarithmically) with the sample size as one samples the population deeper and deeper.

Why draft is different from drift

Draft and hitchhiking are often accounted for by an effective neutral model with reduced population size N_e (Hill and Robertson 1966; Felsenstein 1974). This is possible if draft does not change the offspring distribution qualitatively but only increases its variance. The central limit theorem then guarantees the existence of a meaningful diffusion limit for the allele frequency dynamics with the diffusion constant that could, if one were so inclined, be interpreted as the inverse effective population size. This diffusion limit, however, is not possible in our case since the number of recombinant genotypes, ξ, that descend from a single clone has a very broad distribution with power-law tails and diverging variance. To see this let us consider a genotype seeded by recombination at fitness $x \gg \sigma$ above the mean, which establishes with probability $x - r$. After establishment, its copy number $n(t)$ initially grows almost deterministically; see Equation 4. Over time, the mean fitness increases such that the growth rate eventually turns negative and the genotype goes extinct. During its lifetime, the clone produces $\xi(x) \approx r \int_0^\infty dt \; n(t) \sim e^{(x-r)^2/2\sigma^2}$ recombinant offspring. Since the initial fitness x of the genotype is drawn from a Gaussian distribution $P(x) = e^{-x^2/2\sigma^2}/\sqrt{2\pi\sigma^2}$, the mean and second moment of the distribution of the number of recombinant offspring are given by $\langle\xi\rangle = \int dx P(x)\xi(x) = 1$ and $\langle\xi^2\rangle = \int dx P(x)\xi^2(x)$, respectively. While the mean exists, $\xi^2(x) \sim e^{(x-r)^2/\sigma^2}$ increases so rapidly with x that the second moment and hence the variance are divergent. Of course, the finite population cuts off the offspring distribution and strictly speaking prevents the divergence of the variance. However, this cutoff is on the order of the population size

and is irrelevant for the dynamics of rare alleles and the existence of a diffusion limit. Below this cutoff, one finds that the distribution of ξ behaves asymptotically as

$$P(\xi) \sim \frac{e^{-r\sigma^{-1}\sqrt{2\log\xi\sigma}}}{\xi^2} \qquad (14)$$

The implications of this heavy-tailed distribution of genotypes that descend from a single clone are best understood by slight abstraction of the model. Assume for a moment population dynamics where clones are seeded simultaneously and their recombinant offspring start growing only after the last clone of the previous round has died, as illustrated in Figure 4A. This corresponds to an effective discrete generation scheme, only that one generation corresponds to the rise and fall of a genotype, rather than a single individual. Each clone spawns a Poisson-distributed number of recombinant genotypes with mean ξ, where the ξ are drawn from the heavy-tailed distribution $P(\xi)$. The generating function of the distribution of the number of recombinant offspring is found to be $\hat{P}(\lambda) \approx e^{-(1-\lambda)\left(1-e^{-r\sigma^{-1}\sqrt{-2\log(1-\lambda)}}\right)}$. Starting from a single genotype, we calculate the distribution $P_n(m)$ of the number of genotypes carrying the mutant allele after n effective generations, which has the generating function

$$\hat{P}_n(\lambda) = 1 - \phi\circ\phi\cdots\phi(\lambda) = 1 - \Phi_n(\lambda), \qquad (15)$$

where $\phi(\lambda) = 1 - \hat{P}(\lambda)$ and $\circ$ denotes functional composition; *i.e.*, $f \circ g(x) = f(g(x))$. From this result, we derive a difference equation for $\Phi_n(\lambda)$,

$$\Phi_{n+1}(\lambda) - \Phi_n(\lambda) = -e^{-r\sigma^{-1}\sqrt{-2\log(\Phi_n(\lambda)r\sigma^{-1})}}\Phi_n(\lambda), \quad (16)$$

where it is assumed that $\Phi_n(\lambda)$ is small. This difference equation is the discrete analog of Equation A6, which is derived in *Appendix A* for the continuous-time model and from which most of our results follow. Thus the dynamics of mutations in a rapidly adapting population can be viewed as population genetics of genotypes, rather than of individuals, with a power-law tailed offspring distribution with diverging variance. This feature makes the description by an effective neutral Fisher–Wright model impossible, since no diffusion limit exists when increments have such a long-tailed distribution. A similar effect occurs in Gillespie's pseudo-hitchhiking model (Gillespie 2000), where a hitchhiking event can bring an allele to instantaneous fixation.

Coalescent processes with broad and skewed offspring distributions are an active field of research; see, for example, Mohle and Sagitov (2001) and Schweinsberg (2003). It is known that broad offspring distributions can result in simultaneous mergers of multiple lineages and have dramatic effects on the time to the most recent common ancestor, which can increase sublinearly with the population size. Whether and how our interpretation of the adapting population in terms of a coarse-grained coalescent corresponds to

a known universality class of coalescent models remains to be seen. Coalescent models with broad offspring distributions have been applied to diversity data of pacific oyster (Eldon and Wakeley 2006) and our results suggest that they might apply to a larger class of phenomena.

Adaptation vs. purifying selection

Discussions of Hill–Robertson interference typically focus on the effect of either linked beneficial or deleterious alleles: the former is referred to as hitchhiking (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989) and the latter as "background selection" (Charlesworth *et al.* 1993; Nordborg *et al.* 1996). Our approach addresses the effect of variation in genetic background fitness independent of its origin. Thus, while our analysis was set up in the context of continuous adaptation driven by many simultaneous selective sweeps, it is equally applicable to fitness variation dominated by deleterious mutations. The balance between deleterious mutations and selection in large populations gives rise to a steady Poisson distribution of fitness (Haigh 1978; Charlesworth *et al.* 1993). If the deleterious mutation rate U_d is much larger than the selection coefficient s_0 of mutations, the Poisson distribution is very close to Gaussian, as in the case of the adapting populations considered here. While the origin of fitness variation is very different in these two cases, the effect on the fate of mutations is similar. In the adaptation scenario, a genotype carrying the mutations in question stays at a certain point along the fitness axis while the mean fitness is increasing. When deleterious mutations dominate, the mean fitness is constant and set by a mutation–selection balance, while the fitness of asexual offspring of a particular genotype decreases due to accumulating weakly deleterious mutations. Since the dynamics are determined by fitness relative to mean fitness, there is little difference between these scenarios from the point of view of the focal mutation. We have simulated scenarios where fitness variation is due to purifying selection and found that the stochastic dynamics of novel mutations in a purifying selection scenario are very similar to the case where fitness variation is due to sweeping mutations. The analogs of Figures 2 and 3 for background selection are shown in Figure S2 and Figure S3.

In the asexual case, the most relevant quantity for background selection is the size of the least-loaded class given by $Ne^{-U_d/|s_0|}$, from which all future individuals descend and within which the dynamics are essentially neutral (Haigh 1978; Charlesworth *et al.* 1993). With recombination, beneficial and deleterious mutations can decouple from each other and individuals in the bulk of the distribution have a chance of contributing to the future population, albeit with a smaller and smaller chance as the distance from the fittest class increases. In the limit of rapid recombination, this effect can be described by a reduction in the effective population, which has been studied by Hudson and Kaplan (1995) and Nordborg *et al.* (1996). Our results imply that background selection changes the population genetics

more dramatically when the outcrossing rate is small compared to the standard deviation in fitness.

In any real-world scenario, there will be contributions to fitness variation from beneficial as well as deleterious mutations. In fact, in HIV the fitness variation due to deleterious mutations is expected to be substantial: with a mutation rate of ~ 0.2 per genome replication (Mansky and Temin 1995) and an average effect size of a mutation of $s_0 = -0.01$, we have $\sigma = \sqrt{U_d |s_0|} \approx 0.04$. The outcrossing rate is of the same order of magnitude (Neher and Leitner 2010), so our results are expected to apply for deleterious mutations alone.

Sequential vs. multiple sweeps

To illustrate the applicability of isolated *vs.* multiple-sweep regimes of selection we compare sweep frequencies in *Drosophila* and in HIV. Genetic divergence between different *Drosophila* species suggests a rate of amino acid substitution of about one every 200 generations (see Sella *et al.* 2009 for a review). While these estimates come with a rather large statistical and methodological uncertainty, they nevertheless indicate that sweeps are frequent, but do not interfere. Each sweep "occupies" a stretch of $\sim s/\rho$ nucleotides (ρ is the recombination rate per nucleotide) on a chromosome for $\sim s^{-1}$ generations. Since they come in at a rate of 0.005 per generation spread over the total map length, they should on average be far apart: expect on average $<1\%$ of genome length to be subject to draft at any given time. However, sites under weak selection might still cause substantial Hill–Robertson interference (McVean and Charlesworth 2000).

On the other hand, the multiple-sweep regime is likely to be relevant to organisms with a facultative sexual life cycle. In particular, pathogens like HIV are under constant selection pressure, resulting in a high rate of selective sweeps. Selection in HIV evolution is best characterized in the *pol* gene, the target of most antiretroviral drugs, and in the *env* gene, the target of neutralizing antibodies. On the order of 100 codons are implicated in drug resistance (Rhee *et al.* 2003; Chen *et al.* 2004) and several such mutations compete and sweep during the evolution of drug resistance in a single patient. The *env* gene frequently builds up a nucleotide diversity of 3–4% (Shankarappa *et al.* 1999), with frequent adaptive amino acid substitutions (Williamson 2003; Neher and Leitner 2010). The effective recombination rate is $\sim 10^{-5}$ per base and generation (Levy *et al.* 2004; Neher and Leitner 2010). With typical selection strength of a few percent per generation, the characteristic length of the segment affected by the sweep is ~ 1 kb, which is comparable to the size of the evolving genes: hence the dynamics are in the multiple-sweep regime. While *pol* and *env* genes constitute only a fraction of the HIV genome, CTL epitopes are found throughout the HIV genome (Korber *et al.* 2007) so that even more sweeps associated with CTL escapes are expected (Asquith *et al.* 2006). In a recent study, Hedskog *et al.* (2010) characterized the evolution of the *pol*
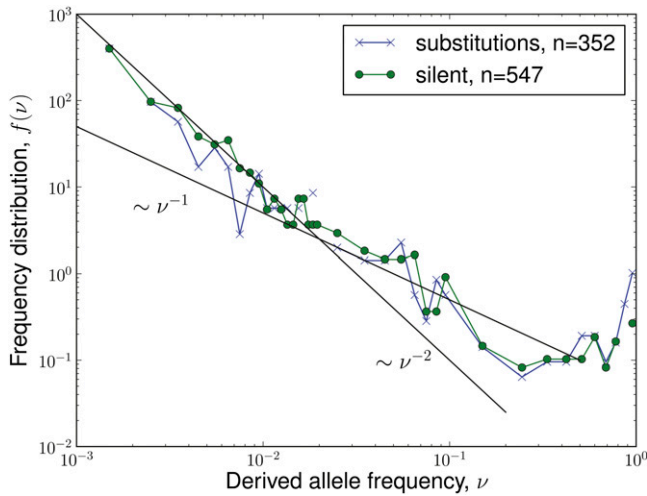
Figure 7 Derived allele frequency spectrum of silent and nonsynonymous mutations in the HIV *pol* gene, amino acids 180–220 (Hedskog *et al.* 2010). The black lines indicate the expectation $f(\nu) \sim \nu^{-2}$ with draft and $f(\nu) \sim \nu^{-1}$ in absence of draft. The neutral spectrum $\sim\nu^{-1}$ fails to capture the steep decay at low frequencies and fits the data only in an intermediate-frequency range. Note, however, that population expansions can give rise to allele frequency spectra similar to those expected under draft.

gene of HIV in six longitudinally sampled patients during antiretroviral treatment. Using their data, we measured the site frequency spectrum of derived alleles, shown in Figure 7. The allele frequency spectrum is indeed much steeper than expected for neutral evolution and compatible with ν^{-2}, rather than the neutral expectation of ν^{-1}. Furthermore, there is little difference between the spectra of silent and nonsynonymous mutations, consistent with our notion of quasi-neutrality. Note, however, that steep allele frequency spectra proportional to ν^{-2} are also expected in expanding populations. While the sequences used in Figure 7 are from chronically infected patients, changes in drug therapy resulted in shrinking and expanding population sizes and the effects of this expansion and draft cannot be disentangled at present.

In a recent study, Rouzine and Coffin (2010) present an analysis of adaptation of HIV, using a model similar to the one used here and in Neher *et al.* (2010). Rouzine and Coffin study the problem of selection on standing genetic variation and in addition to the speed of adaptation, they compute the rate of coalescence, which in their model is controlled by the probability that two individuals in the population are genetically identical, *i.e.*, are part of the same clone. In agreement with our result in Equation 7, they find that the coalescence rate is $\sim e^{-(rx_{max}/\sigma^2)}$, where x_{max} is the fitness of the fittest individuals in the population (relative to $\bar{x}$). The quantity x_{max} corresponds to $\sigma\sqrt{2\log\sigma N}$ in our Equation 7. Furthermore, Rouzine and Coffin (2010) study how benefical alleles at low frequency can be lost due to the lack of recombination. Our work complements Rouzine and Coffin (2010), as we study the stochastic dynamics of new alleles arising from mutations, rather than the evolution of

the fitness distribution, given standing variation. On the other hand, Rouzine and Coffin (2010) account for correlations between loci by allowing the distribution of recombinants to be broader than the population distribution (Rouzine and Coffin 2005)—an effect absent in our model. In the range of recombination rates where our results apply, the latter deviations are small.

Our analytic results are derived using a branching process approximation that assumes that the novel allele is a small fraction of the total population and the finiteness of the population does not yet affect its dynamics. The influence of the finite population size is apparent in Figure 3B, causing deviations from the branching process predictions at large frequencies. The branching process approach can be complemented by an effective theory for macroscopic frequencies, similar in spirit to diffusion theory for random drift. This theory, however, has to account for the very broad distribution of clone sizes. Throughout the article, we have assumed that the speed of adaptation is identical to the variance in fitness; *i.e.*, we have assumed that effects on fitness of different mutations are purely additive. If interactions between mutations contribute to fitness, the variance in fitness tends to be larger than the speed of adaptation, since coadapted combinations are destroyed by recombination. Including genetic interactions within our framework involves decoupling ν and σ, as well as changing the recombination functions. Genetic interaction might increase the importance of draft significantly, as interactions have a tendency to result in clonal population structures and couple the loci along the genome beyond physical linkage (Neher and Shraiman 2009). Another simplification we have made is to assume that all loci segregate independently in an outcrossing event. Our analysis can be extended to account for linear chromosomes by considering a hierarchy of recombination rates for chromosomal segments at different distances. These and other extensions are interesting projects, which we leave for future work.

Several large-scale resequencing projects are underway with the goal to characterize genetic diversity in populations of humans, *Drosophila*, and *Arabidopsis* at great depth. These upcoming data will allow us to measure allele frequency spectra to much greater depth, similar to the example from HIV shown in Figure 7, and reveal the mechanisms shaping genetic diversity in natural populations.

Acknowledgments

Literature Cited

Andolfatto, P., and M. Przeworski, 2001 Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. Genetics 158: 657–665.

Asquith, B., C. T. T. Edwards, M. Lipsitch, and A. R. McLean, 2006 Inefficient cytotoxic t lymphocyte-mediated killing of HIV-1-infected cells in vivo. PLoS Biol. 4: e90.

Barton, N., 1994 The reduction in fixation probability caused by substitutions at linked loci. Genet. Res. 64: 199–208.

Barton, N., 1998 The effect of hitch-hiking on neutral genealogies. Genet. Res. 72: 123–133.

Barton, N., 2009 Why sex and recombination? Cold Spring Harbor Symp. Quant. Biol. 74: 187–195.

Barton, N. H., 1995a A general model for the evolution of recombination. Genet. Res. 65: 123–145.

Barton, N. H., 1995b Linkage and the limits to natural selection. Genetics 140: 821–841.

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

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring dna polymorphism correlate with recombination rates in D. melanogaster. Nature 356: 519–520.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley, and W. Stephan, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140: 783–796.

Bulmer, M. G., 1980 *The Mathematical Theory of Quantitative Genetics*. Oxford University Press, Oxford.

Caicedo, A. L., S. H. Williamson, R. D. Hernandez, A. Boyko, A. Fledel-Alon *et al.*, 2007 Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet. 3: 1745–1756.

Charlesworth, B., M. Morgan, and D. Charlesworth, 1991 Multilocus models of inbreeding depression with synergistic selection and partial self-fertilization. Genet. Res. 57: 177–194.

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

Charlesworth, D., M. T. Morgan, and B. Charlesworth, 1992 The effect of linkage and population size on inbreeding depression due to mutational load. Genet. Res. 59: 49–61.

Chen, L., A. Perlina, and C. J. Lee, 2004 Positive selection detection in 40,000 human immunodeficiency virus (HIV) type 1 sequences automatically identifies drug resistance and positive fitness mutations in HIV protease and reverse transcriptase. J. Virol. 78: 3722–3732.

Cohen, E., D. A. Kessler, and H. Levine, 2005 Recombination dramatically speeds up evolution of finite populations. Phys. Rev. Lett. 94: 098102.

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

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

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

Eldon, B., and J. Wakeley, 2006 Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics 172: 2621–2633.

Ewens, W. J., 2004 *Mathematical Population Genetics*. Springer-Verlag, Berlin/Heidelberg, Germany/New York.

Fay, J. C., and C. I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics 155: 1405–1413.

Felsenstein, J., 1974 The evolutionary advantage of recombination. Genetics 78: 737–756.

Gillespie, D. T., 1977 Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81: 2340–2361.

Gillespie, J. H., 2000 Genetic drift in an infinite population. The pseudohitchhiking model. Genetics 155: 909–919.

Gillespie, J. H., 2001 Is the population size of a species relevant to its evolution? Evolution 55: 2161–2169.

Griffiths, R. C., and P. Marjoram, 1996 Ancestral inference from samples of DNA sequences with recombination. J. Comput. Biol. 3: 479–502.

Haigh, J., 1978 The accumulation of deleterious genes in a population—Muller's ratchet. Theor. Popul. Biol. 14: 251–267.

Hedskog, C., M. Mild, J. Jernberg, E. Sherwood, G. Bratt *et al.*, 2010 Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing. PLoS ONE 5: e11345.

Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genet. Res. 8: 269–294.

Hudson, R. R., 1983 Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. 23: 183–201.

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

Iwasa, Y., F. Michor, and M. A. Nowak, 2004 Stochastic tunnels in evolutionary dynamics. Genetics 166: 1571–1579.

Jung, A., R. Maier, J.-P. Vartanian, G. Bocharov, V. Jung *et al.*, 2002 Multiply infected spleen cells in HIV patients. Nature 418: 144.

Kaplan, N. L., R. R. Hudson, and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics 123: 887–899.

Karasov, T., P. W. Messer, and D. A. Petrov, 2010 Evidence that adaptation in Drosophila is not limited by mutation at single sites. PLoS Genet. 6: 1000924.

Kelly, J. K., and S. Williamson, 2000 Predicting response to selection on a quantitative trait: a comparison between models for mixed-mating populations. J. Theor. Biol. 207: 37–56.

Kingman, J., 1982 On the genealogy of large populations. J. Appl. Probab. 19: 27–43.

Korber, B. T. M., C. Brander, B. F. Haynes, R. Koup, and J. P. Moore, et al. (Editors), 2007 *HIV Molecular Immunology 2006/2007*. Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, NM.

Krone, S., and C. Neuhauser, 1997 Ancestral processes with selection. Theor. Popul. Biol. 51: 210–237.

Larder, B., G. Darby, and D. Richman, 1989 HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. Science 243: 1731.

Levy, D. N., G. M. Aldrovandi, O. Kutsch, and G. M. Shaw, 2004 Dynamics of HIV-1 recombination in its natural target cells. Proc. Natl. Acad. Sci. USA 101: 4204–4209.

Lewontin, R. C., 1974 *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

Lynch, M., 2007 *The Origins of Genome Architecture*. Sinauer Associates, Sunderland, MA.

Lynch, M., 2010 Scaling expectations for the time to establishment of complex adaptations. Proc. Natl. Acad. Sci. USA 107: 16577–16582.

Mansky, L. M., and H. M. Temin, 1995 Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J. Virol. 69: 5087–5094.

Markowitz, M., M. Louie, A. Hurley, E. Sun, M. D. Mascio *et al.*, 2003 A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and t-cell decay in vivo. J. Virol. 77: 5037–5038.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23–35.

McVean, G. A., and B. Charlesworth, 2000 The effects of Hill-Robertson interference between weakly selected mutations on

patterns of molecular evolution and variation. Genetics 155: 929–944.

Mohle, M., and S. Sagitov, 2001 A classification of coalescent processes for haploid exchangeable population models. Ann. Probab. 29: 1547–1562.

Mustonen, V., and M. Lässig, 2010 Fitness flux and ubiquity of adaptive evolution. Proc. Natl. Acad. Sci. USA 107: 4248–4253.

Neher, R. A., and T. Leitner, 2010 Recombination rate and selection strength in HIV intra-patient evolution. PLoS Comput. Biol. 6: e1000660.

Neher, R. A., and B. I. Shraiman, 2009 Competition between recombination and epistasis can cause a transition from allele to genotype selection. Proc. Natl. Acad. Sci. USA 106: 6866–6871.

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

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

Oliphant, T., 2007 Python for scientific computing. Comput. Sci. Eng. 9: 10–20.

Paul, W. E., 2008 Fundamental Immunology. Lippincott Williams & Wilkins, Philadelphia.

Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho, 1996 HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science 271: 1582–1586.

Rhee, S.-Y., M. J. Gonzales, R. Kantor, B. J. Betts, J. Ravela et al., 2003 Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res. 31: 298–303.

Robertson, A., 1961 Inbreeding in artificial selection programmes. Genet. Res. 2: 189–194.

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

Rouzine, I. M., and J. M. Coffin, 2007 Highly fit ancestors of a partly sexual haploid population. Theor. Popul. Biol. 71: 239–250.

Rouzine, I. M., and J. M. Coffin, 2010 Multi-site adaptation in the presence of infrequent recombination. Theor. Popul. Biol. 77: 189–204.

Rouzine, I. M., J. Wakeley, and J. M. Coffin, 2003 The solitary wave of asexual evolution. Proc. Natl. Acad. Sci. USA 100: 587–592.

Santiago, E., and A. Caballero, 1998 Effective size and polymorphism of linked neutral loci in populations under directional selection. Genetics 149: 2105.

Schweinsberg, J., 2003 Coalescent processes obtained from supercritical Galton-Watson processes. Stoch. Proc. Appl. 106: 107–139.

Seger, J., W. Smith, J. Perry, J. Hunn, Z. Kaliszewska et al., 2010 Gene genealogies strongly distorted by weakly interfering mutations in constant environments. Genetics 184: 529.

Sella, G., D. A. Petrov, M. Przeworski, and P. Andolfatto, 2009 Pervasive natural selection in the drosophila genome? PLoS Genet. 5: e1000495.

Shankarappa, R., J. B. Margolick, S. J. Gange, A. G. Rodrigo, D. Upchurch et al., 1999 Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J. Virol. 73: 10489–10502.

Stephan, W., and S. J. Mitchell, 1992 Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of Drosophila ananassae. Genetics 132: 1039–1045.

Tsimring, L., H. Levine, and D. Kessler, 1996 RNA virus evolution via a fitness-space model. Phys. Rev. Lett. 76: 4440–4443.

Weissman, D., M. M. Desai, D. Fisher, and M. Feldman, 2009 The rate at which asexual populations cross fitness valleys. Theor. Popul. Biol. 75: 286–300.

Weissman, D., M. Feldman, and D. Fisher, 2010 The rate of fitness-valley crossing in sexual populations. Genetics 186: 1389–1410.

Wiehe, T. H., and W. Stephan, 1993 Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from Drosophila melanogaster. Mol. Biol. Evol. 10: 842–854.

Williamson, S., 2003 Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. Mol. Biol. Evol. 20: 1318–1325.

Communicating editor: L. M. Wahl

Appendix A

Derivation of the results

To analyze the dynamics of the probability distribution, it is useful to consider the generating function $\hat{p}(\lambda, T, t, x) = \sum_n \lambda^n p(n, T | 1, t, x)$. From Equation 1, one obtains (see File S1) the following equation for $\phi(\lambda, T, t, x) = 1 - \hat{p}(\lambda, T, t, x)$,

$$- \partial_t \phi(\lambda, T, t, x) = r \int dx' K_{xx'} \phi(\lambda, T, t, x') + (x - vt + s - r)\phi(\lambda, T, t, x) - \phi^2(\lambda, T, t, x) \tag{A1}$$

with boundary condition $\phi(\lambda, T, T, x) = 1 - \lambda$. The survival probability is given by evaluating $\phi(\lambda, T, t, x)$ at $\lambda = 0$, while the moments and asymptotic behavior of $p(n, T, t)$ are determined by the behavior of $\phi(\lambda, T, t, x)$ in the vicinity of $\lambda = 1$. We have to solve for $\phi(\lambda, T, t, x)$ in these two limits, which show up repeatedly below. It is convenient to remove the explicit dependence on the initial condition and the velocity $v = \sigma^2$ by rescaling $\psi(T, t, x) = \phi(\lambda, T, t, x)/(\sigma\varepsilon)$ with $\varepsilon = 1 - \lambda$. The two limits of interest are now $\varepsilon \ll 1$ ($\lambda \approx 1$) and $\varepsilon \approx \sigma^{-1} \gg 1$ ($\lambda \approx 0$). After rescaling $\tilde{r} = \sigma^{-1} r$, $\tilde{s} = \sigma^{-1} s$, and $\chi = \sigma^{-1} x - \sigma T$ and $\tau = \sigma(T - t)$, Equation A1 takes the form

$$\partial_\tau \psi(\tau, \chi) = \tilde{r} \int d\chi' K_{\chi\chi'} \psi(\tau, \chi') + (\chi + \tau + \tilde{s} - \tilde{r})\psi(\tau, \chi) - \varepsilon \psi^2(\tau, \chi), \tag{A2}$$

where $\psi(\tau, \chi)$ is the rescaled generating function of the copy number distribution of a mutation that initially occurred on the genome with rescaled background fitness χ. The distribution of rescaled fitness χ in the population $P(\chi, \tau) = (1/\sqrt{2\pi})e^{-(\chi - \bar{\chi}(\tau))^2/2}$ is Gaussian around the moving mean fitness $\bar{\chi}(\tau)$. Averaging Equation A2 over $P(\chi, \tau)$ yields an equation for the scaled generating function $\Phi(\tau) = \int d\chi P(\chi, \tau)\psi(\tau, \chi)$:

$$\partial_\tau \Phi(\tau) = \partial_\tau \int d\chi P(\chi, \tau)\psi(\tau, \chi) = \tilde{s}\Phi(\tau) - \varepsilon \int d\chi P(\tau, \chi)\psi^2(\tau, \chi). \tag{A3}$$

For beneficial mutations, $\Phi(\tau)$ approaches a steady state where the two terms on the right balance each other. This long time limit was used as a solvability condition in Neher *et al.* (2010) to determine the fixation probability of beneficial mutations. The steady state is independent of the initial condition ε and therefore equals the rescaled survival probability. However, if $\tilde{s} \leq 0$, no such steady state exists, since neutral and deleterious mutations have zero chance of fixation in an infinite population. To calculate the probability that a mutation reaches n copies after time τ, we need to find the full time-dependent solution of Equation A3. For simplicity, we use the communal recombination model throughout and show that the infinitesimal model yields similar results by simulations of the latter. For the communal recombination model, the recombination contribution in Equation A2 reduces to $\tilde{r}\Phi(\tau)$.

In analogy to Neher *et al.* (2010), we solve Equation A2 in the two asymptotic regimes of large positive and large negative growth rate $\theta = \chi + \tau + \tilde{s} - \tilde{r}$. For $\theta < 0$ the nonlinearity in Equation A2 can be neglected since $\psi(\tau, \chi) \ll \varepsilon^{-1}$. Conversely, at large θ the dominant balance in Equation A2 is $\theta\psi(\tau, \chi) = \varepsilon\psi^2(\tau, \chi)$. The two asymptotic solutions are

$$\psi(\tau, \chi) = \begin{cases} \tilde{r}e^{\theta^2/2} \int_0^\tau d\tau' \Phi(\tau')e^{-\theta'^2/2} & \theta < \Theta_c(\tau) \\ \\ \theta/\varepsilon & \theta > \Theta_c(\tau). \end{cases} \tag{A4}$$

The crossover between the two solutions occurs in a narrow region at $\theta = \Theta_c(\tau)$, where $\Theta_c(\tau)$ is approximately the point where the two solutions cross. The evolution of $\Phi(\tau)$ is governed by Equation A3, which has to be self-consisted with Equation A4. We first solve for $\Phi(\tau)$ assuming that τ is large. In this case the dominant contribution to the integral in Equation A4 will come from a well-defined $\tau' \gg 0$ that is determined by the maximum of $\exp(-\theta'^2/2 + \log \Phi(\tau'))$ at $\tau' = \tau - \theta$. $\Phi(\tau)$ will turn out to change slowly and its time derivative can be neglected when determining the maximum of the exponent. Hence,

$$\psi(\tau) \approx \sqrt{2\pi}\tilde{r}\Phi(\tau)e^{\theta^2/2}. \tag{A5}$$

This solution is valid below $\Theta_c(\tau)$, where $\psi(\tau)$ crosses over the linear saturated form θ/ε. Substituting this solution into Equation A3 yields

$$\partial_\tau \Phi(\tau) = \tilde{s}\Phi(\tau) - \tilde{r}\,\Phi(\tau)e^{-\tilde{r}\,\Theta_c(\tau)-\tilde{r}^2/2}. \tag{A6}$$

We first consider neutral mutations, $\tilde{s} = 0$, in which case $\Phi(\tau)^{-1}\partial_\tau\Phi(\tau) = -\tilde{r}e^{-\tilde{r}\Theta_c(\tau)-\tilde{r}^2/2}$, confirming that $\log\Phi(\tau)$ changes slowly as long as $\tilde{r}\Theta_c \gg 1$ (the limit of $\tilde{r}\Theta_c \ll 1$ yields qualitatively different answers). Solving Equation A6 for $\Theta_c(\tau)$ and differentiating with respect to τ yields an equation for $\partial_\tau\Theta_c(\tau)$, which is readily solved for $\Theta_c(\tau)$ in the case $\tilde{r}\Theta_c \gg 1$,

$$\Theta_c(\tau) = \frac{1 + W\left(\tilde{r}^3(\tau-\tau_0)e^{-1}\right)}{\tilde{r}} \approx \frac{\log\left(\tilde{r}^3(\tau-\tau_0)+e^{\tilde{r}\Theta_0}\right)}{\tilde{r}}, \tag{A7}$$

where Θ_0 is the initial condition at $\tau = \tau_0$ and $W(x)$ is Lambert's W, i.e., the solution of $W(x)e^{W(x)} = x$. The approximation of $W(x)$ by a logarithm is accurate only at a very large time since there are significant $\log\log x$ corrections. The solution for the rescaled generating function then reads

$$\Phi(\tau) = \frac{\Theta_c(\tau)}{\varepsilon\tilde{r}}e^{-\Theta_c(\tau)^2/2} \sim e^{-\log^2\left(\tilde{r}^3(\tau-\tau_0)+e^{\tilde{r}\Theta_0}\right)/2\tilde{r}^2}, \tag{A8}$$

where the last terms include only the exponential factors. Apart from the rescaling, this solution is independent of the initial condition ε. The dependence on ε will reemerge through the matching to the short-time solution. Note that $\Phi(\tau)$ satisfies asymptotically for large τ,

$$\partial_\tau\Phi(\tau) \approx -\frac{\log\left(\tilde{r}^3\tau\right)}{\tilde{r}^2\tau}\Phi(\tau), \tag{A9}$$

which means that the relaxation rate of Φ decreases with time, but more slowly than it does in the case of pure drift (in the absence of draft), in which case $\partial_\tau\log\Phi(\tau) = -1/\tau$.

The long-time solution in Equation A8 has to be matched to the solution at small τ where the initial condition cannot be neglected.

Short-time solution

To analyze the short-time behavior, it is useful to split $\psi(\tau, \chi) = \psi_r(\tau, \chi) + \psi_0(\tau, \chi)$ into a contribution fed by recombination $\psi_r(\tau, \chi)$ and the contribution originating from the initial condition $\psi_0(\tau, \chi)$, where $\psi_r(0, \chi)$ is initially zero and $\psi_0(0, \chi) = 1$. In the communal recombination model, the term describing the production of novel recombinants reduces to $\tilde{r}\Phi(\tau)$ and the two contributions evolve according to

$$\begin{aligned}\partial_\tau\psi_0(\tau,\chi) &= \theta\psi_0 - \varepsilon\psi_0^2 \\ \partial_\tau\psi_r(\tau,\chi) &= \tilde{r}\Phi(\tau) + \theta\psi_r - \varepsilon\left(2\psi_0\psi_r + \psi_r^2\right).\end{aligned} \tag{A10}$$

Recombination enters the equation for ψ_0 only through a reduction of the growth rate and the solution for $\psi_0(\tau, \chi)$ is simply

$$\psi_0(\tau,\chi) = \frac{e^{\theta^2/2}}{\varepsilon\int_0^\tau d\tau' e^{\theta'^2/2} + e^{(\theta-\tau)^2/2}}. \tag{A11}$$

The integral $\Phi_0(\tau) = \int d\chi P(\chi, \tau)\psi_0(\tau, \chi)$ acts as a source for $\psi_r(\tau, \chi)$. The initial behavior of $\Phi_0(\tau)$ is quite different for $\varepsilon \gg 1$ and $\varepsilon \ll 1$ and we analyze these cases separately below.

Survival probability of neutral mutations

Here, we present the steps necessary to arrive at the result for the survival probability (Equation 5). The survival probability is given by the generating function of $p(n, \tau)$, evaluated at $\lambda = 0$. In our rescaled variables, the survival probability therefore corresponds to $\Phi(\tau)$ with $\varepsilon = \sigma^{-1} \gg 1$. The long-time solution for $\Phi(\tau)$ given in Equation A8 contains the free parameters τ_0 and Θ_0, defined by matching with the short-time solution that depends explicitly on the initial condition (see above). To establish the matching, we trace the short-time solution for $\varepsilon \gg 1$ through several regimes until the initial condition is forgotten. At small times $\tau \gg 1$, we have

$$\Phi(\tau) \approx \Phi_0(\tau) = \frac{1}{\varepsilon\tau+1}. \tag{A12}$$

This rapid initial decay corresponds to the early loss of the allele due to neutral processes. Selection starts to matter only after $\tau \sim 1$, yielding an accelerated decay of $\Phi_0(\tau)$:

$$\Phi_0(\tau)\frac{1}{\varepsilon\sqrt{2\pi}}\exp\left[-\frac{\tau^2}{8}-\frac{\tilde{r}\tau}{2}-\frac{\tilde{r}^2}{2}\right] \tag{A13}$$

The Gaussian decay of the survival probability of unrecombined alleles is due to the increasing mean fitness, *i.e.*, the adapting population wave moving past the genotype the mutation initially arose on. Hence $\Phi_0(\tau)$ does not contribute to the long-time behavior, which is dominated by the part of the solution driven by recombination. We therefore have to calculate how much weight is transferred from ψ_0 to ψ_r via the term $\tilde{r}\Phi(\tau)$ in Equation A10. In the case $\varepsilon \gg 1$, the dominant balance for $\psi_r(\tau, \chi)$ is $\tilde{r}\Phi_0(\tau) \approx 2\varepsilon\psi_r(\tau, \chi)\psi_0(\tau, \chi)$, resulting in an initially steady $\Phi_r(\tau) = \tilde{r}/2\varepsilon$. After the initial condition has decayed and $\Phi_0(\tau) \ll \Phi_r(\tau)$, the solution $\psi_r(\tau, \chi)$ acquires the form of Equation A4 with the only memory of the initial condition in $\Phi_r(\tau)$. The crossover time is obtained by equating $\Phi_0(\tau)$ with $\Phi_r(\tau)$, which yields $\tau_0 = \sqrt{8\log(2/\tilde{r})}-2\tilde{r}$. Using this τ_0 and $\Theta_0 \approx \sqrt{-4\log\tilde{r}}$ (compare with Equation A4), we can match the short-time solution to the long-time solution in Equation A8.

To illustrate the validity of the approximations made in the analysis we solved Equation A2 numerically. Fig. A1 shows the square root of the logarithm of the rescaled survival probability, $\sqrt{-2\tilde{r}^2\log\Phi(\tau)}$, plotted as a function of $W(\tilde{r}^3\tau e^{-1-\tilde{r}^2/2})$. Data for different outcrossing rates collapse onto the same curve, indicating that the expression in Equation A8 captures the dependence on $\tilde{r}$.

The above analysis assumed the communal recombination model, *i.e.*, a model where the offspring fitness is independent of the parental fitness. To demonstrate that the communal recombination model captures the behavior of the infinitesimal model, we compare results obtained from stochastic simulations of the branching process for the two models in Figure A1. The infinitesimal model exhibits the same scaling collapse, indicating that the dependence on the parameters is similar. The two models do, however, show quantitative differences, which can be interpreted as less efficient recombination in the infinitesimal model. This is expected, as recombination in the infinitesimal model only halves the correlation with either parent, while this correlation is completely destroyed in the communal recombination model. Note, however, that the dependence of P_S on $\tilde{r}$ is strong, so that a rescaling of $\tilde{r}$ substantially changes the absolute numbers.

Beneficial and deleterious mutations

To analyze the behavior of $\Phi(\tau)$ in the case of $\tilde{s} \neq 0$, we track back to Equation A6, focusing on the case where $\tilde{r}\Theta_c \gg 1$:

$$\partial_\tau\Phi(\tau) = (\tilde{s} - \tilde{r}e^{-\tilde{r}\Theta_c})\Phi(\tau). \tag{A14}$$

This equation highlights the fact that as long as $\tilde{s} \ll \tilde{r}e^{-\tilde{r}\Theta_c}$, the dynamics of $\Phi(\tau)$ are dominated by fluctuations and are therefore similar to the neutral case. For a negative $\tilde{s}$, the decay of $\Phi(\tau)$ is accelerated, while a positive $\tilde{s}$ slows down and eventually halts the decay of $\Phi(\tau)$; see Figure 5. The crossover to the behavior dominated by the intrinsic fitness effect occurs when the two terms are equal. Comparing with Equation A9 yields the crossover condition

$$\tilde{s} \approx \frac{\log(\tilde{r}^3\tau)}{\tilde{r}^2\tau} \text{ or } \tau \approx \frac{\log(\tilde{r}/\tilde{s})}{\tilde{r}^2\tilde{s}}. \tag{A15}$$

Before this crossover, the effect of selection on $\tilde{s}$ is a simple attenuation or amplification of $\Phi(\tau) = e^{\tilde{s}\tau}\Psi(\tau)$, where $\Psi(\tau)$ is approximately the neutral solution, with the slight difference that the crossover point $\Theta_c \approx \sqrt{-2\log(\varepsilon\tilde{r}\Psi(\tau))-2\tilde{s}\tau}$. The correction to the crossover point is still small when selection on the intrinsic fitness effect s starts to dominate the dynamics of $\Phi(\tau)$.

Distribution of the allele copy number

The probability $p(n, T)$ of finding n copies of the allele after time T can be calculated from the generating function $\hat{p}(\lambda, T)$ by contour integration in the complex λ-plane:

$$p(n, \tau) = \frac{1}{2\pi i} \oint_C d\lambda \frac{\hat{p}(\lambda, T)}{\lambda^{n+1}}. \tag{A16}$$

The contour C has to encircle the origin and no other singularities of $\hat{p}(\lambda, T)$. For $n > 0$, we can evaluate the contour integral by deforming the contour to run alongside the singularities away from zero, which are tightly controlled by λ^{-n} for $n \gg 1$. To do so, we have to study the analytic properties of $\hat{p}(\lambda, T)$ or equivalently of $\Phi(\tau) = \sigma^{-1}\varepsilon^{-1}(1-\hat{p}(\lambda, T))$, where $\varepsilon = 1 - \lambda$. For negative $\varepsilon = 1 - \lambda$ and $s \geq 0$, $\psi(\chi, \tau)$ has a singularity at Θ_c, the value separating the regime of small θ where the linear equation is valid and the saturated regime where $\psi(\tau) = \theta/\varepsilon$; compare Equation A4. $\Phi(t)$ is the integral of $\psi(\tau, \chi)$ over χ and therefore has a branch cut for negative real ε. Otherwise, $\Phi(\tau)$ is an analytic function of ε. We can use this branch cut to

evaluate the contour integral in Equation A16. Substituting $\Phi(\tau)$ into Equation A16 in changing the integration variable to ε, we have

$$p(n,\tau) = \frac{\sigma}{2\pi i} \oint_{\mathcal{B}} d\varepsilon \, \frac{\varepsilon\Phi(\tau)}{(1-\varepsilon)^{n+1}} \approx \frac{\sigma}{2\pi i} \oint_{\mathcal{B}} d\varepsilon \, e^{\log\varepsilon + n\varepsilon + \log\Phi(\tau)}, \tag{A17}$$

where $\mathcal{B}$ is the contour encircling the branch cut. For large n the dominant contributions to the integral come from $\varepsilon \ll 1$ and we have to find a solution for $\Phi(\tau)$ for small ε.

To this end, we backtrack to Equation A11 and integrate the evolution equation for $\Phi(\tau)$ for $\varepsilon \ll 1$. We find

$$\Phi_0(\tau) = \begin{cases} e^{-\tilde{r}\tau} & \tau < \sqrt{-2\log\varepsilon} \\ \frac{1}{\sqrt{\pi\varepsilon}\tau} \exp\left[-\frac{\tau^2}{8} - \frac{\tilde{r}\tau}{2} - \frac{\tilde{r}^2}{2}\right] & \tau > \sqrt{-2\log\varepsilon}. \end{cases} \tag{A18}$$

For intermediate $\tilde{r}$, the matching between the short-term solution and the long-time solution occurs at $\tau_0 = \Theta_0 = \sqrt{-2\log\varepsilon}$, after which we have

$$\Phi(\tau) = \frac{\Theta_\varepsilon}{\varepsilon\tilde{r}} e^{-\Theta_c(\tau)^2/2} \text{ with } \Theta_c(\tau) \approx \tilde{r}^{-1}\log\left(\tilde{r}^3(\tau-\tau_0) + e^{r\Theta_0}\right). \tag{A19}$$

For large Θ_0, $\Theta_c(\tau) \approx \Theta_0 + \tilde{r}^2(\tau-\tau_0)e^{-\tilde{r}\Theta_0}$ and the branch-cut integral becomes

$$p(n,\tau) \approx \frac{\sigma}{2\pi\tilde{r}i} \oint_{\mathcal{B}} d\varepsilon\Theta_0 e^{ne^{-\Theta_0^2/2} - \Theta_0^2/2 - \Theta_0\tilde{r}^2(\tau-\tau_0)e^{-\tilde{r}\Theta_0}}. \tag{A20}$$

The exponent of the integrant peaks when

$$n\Theta_0 e^{-\Theta_0^2/2} + \Theta_0 - \tilde{r}^2\left(\tilde{r}\Theta_0 - 1\right)(\tau - \tau_0)e^{-\tilde{r}\Theta_0} = 0. \tag{A21}$$

For large n, the dominant balance is between the first term of Equation A21 and the remainder, which requires that $n \approx e^{\Theta_0^2/2}$ and hence $\varepsilon \approx n^{-1}$. Using $\Theta_0 = \tau_0 = \sqrt{2\log n}$, the asymptotic copy number distribution becomes

$$p(n,\tau) \sim \frac{1}{n\sqrt{2\log n}} e^{-\left(\Theta_0 + \tilde{r}^{-1}\log(1 + \tilde{r}^3(\tau-\tau_0)e^{-\tilde{r}\Theta_0})\right)^2/2} \tag{A22}$$

The leading dependence of $p(n,\tau)$ on n is $p(n,\tau) \sim n^{-2}$. Note, however, that the cross-term from the exponent yields another subdominant factor $e^{\alpha\sqrt{\log n}}$. By contrast, the corresponding expression in the high recombination regime is $p(n,T) \sim T^{-2}e^{-n/T}$.

The allele frequency spectrum, $f(n)$, i.e., the probability to find n copies of a derived allele irrespective of its age, is given by the time average of $p(n,\tau)$ over mutations that arose at different times τ in the past. It depends on how quickly the survival probability decays, as well as on the shape of $p(n,\tau)$. For large n one can expand the logarithm in the exponent of Equation A22 and integrate $p(n,\tau)$ over τ to obtain

$$f(n) = \int_0^\infty d\tau p(n,\tau) \sim \begin{cases} \frac{e^{\tilde{r}\sqrt{2\log n}}}{n^2\log n} & \tilde{r} \ll 1 \\ n^{-1} & \tilde{r} \gg 1. \end{cases} \tag{A23}$$

Even though $p(n,\tau)$ decays much faster for $\tilde{r} \gg 1$ (exponential) than for $\tilde{r} \ll 1$ (algebraically), the allele frequency spectrum is steeper for $\tilde{r} \ll 1$ than for $\tilde{r} \gg 1$. The long tail of the latter is due to the contribution of very old alleles with very flat exponential $p(n,\tau)$. For $\tilde{r} \ll 1$, the clonal amplification and rapid extinction of most alleles give rise to steep allele frequency spectra.

The double mutation probability: stochastic tunneling

To calculate tunneling probabilities (Iwasa *et al.* 2004; Weissman *et al.* 2009), we need to know the distribution of $w = \int_0^t dt' n(t')$ treated as a stochastic variable, i.e., the size of transient "bubbles" of mutant individuals in a large population. In analogy to Equation 1, the distribution $p(w,T|k,t,x)$ for a mutation being present on background x in k copies obeys the following equation:

$$- (\partial_t - k\partial_w)p(w, T \mid k, t, x)$$
$$= - k(B + D + r)p(w, T \mid k, t, x) + kBp(w, T \mid k + 1, t, x) + kDp(w, T \mid k - 1, t, x)$$
$$+ rk \int dw' \int dx' K_{xx'} p(w - w', T \mid k - 1, t, x)p(w', T \mid 1, t, x'). \tag{A24}$$

Instead of the generating function, we now consider the Laplace transform $\hat{p}\,(z, T \mid k, t, x) = \int dz e^{-zw} p(w, T \mid k, t, x)$. Because of the convolution property of Laplace transforms and the fact that the fates of the k alleles present at time t are independent of each other, we have $\hat{p}\,(z, T \mid k, t, x) = \hat{p}^{\,k}(z, T \mid 1, t, x)$. As before, we scale rates by σ^{-1} and times by σ. The equation for $\phi(\tau, z, x) = \sigma^{-1}(1 - \hat{p}(z, T \mid 1, t, x))$ then is

$$\partial_\tau \phi(\tau, z, \chi) = z + \tilde{r}\Phi(\tau, z) + \theta\phi(\tau, z, \chi) - \phi^2(\tau, z, \chi), \tag{A25}$$

where $\theta = \chi + \tau + \tilde{s} + z - \tilde{r}$. At $\tau = \sigma(T - t) = 0$, $w = 0$ and therefore $\phi(\tau, z, \chi) = 0$. If tunneling is rare and happens only on timescales longer than the lifetime of bubbles, we do not need to know the full time-dependent solution for $\phi(\tau, z, \chi)$ but can send τ to infinity. The above equation can again be solved by matching the asymptotic solutions at large negative and positive θ, which in the steady state read

$$\phi(z, \chi) \approx \begin{cases} e^{\theta^2/2}\left[z + \tilde{r}\Phi(z)\right] & \theta \ll \Theta_c \\ \theta & \theta \gg \Theta_c. \end{cases} \tag{A26}$$

The crossover point Θ_c is determined by $e^{\Theta_c^2/2}[z + \tilde{r}\Phi(z)] = \Theta_c$. The "solvability condition" obtained by integrating Equation A25 with regard to the Gaussian distribution of χ yields

$$0 = z + \tilde{s}\Phi(z) - \int \frac{d\chi}{\sqrt{2\pi}}\, e^{-\chi^2/2}\phi^2(z, \chi). \tag{A27}$$

For small $\tilde{r}$, the integral on the right-hand side is dominated by the vicinity of the crossover point Θ_c. Combining with the matching condition, one finds for $\Phi(z)$

$$\Phi(z) = z\frac{1 - e^{-\tilde{r}\Theta_c}}{\tilde{r}e^{-\tilde{r}\Theta_c} - s}. \tag{A28}$$

Substituting into the crossover condition we obtain an equation for Θ_c:

$$z\frac{\tilde{r} - \tilde{s}}{\tilde{r}e^{-\tilde{r}\Theta_c} - \tilde{s}} = \Theta_c e^{-\Theta_c^2/2}. \tag{A29}$$

In the neutral case ($|\tilde{s}| \ll \tilde{r}e^{-\tilde{r}\Theta_c}$), the location of the crossover is at $\Theta_c = \sqrt{-2\log(ze^{\tilde{r}\Theta_c}/\Theta_c)}$, which implies for $\Phi(z)$

$$\Phi(z) = z\frac{e^{\tilde{r}\Theta_c} - 1}{\tilde{r}} = \begin{cases} z\sqrt{-2\log z} & \tilde{r}\Theta_c \ll 1 \\ ze^{\tilde{r}\Theta_c}/\tilde{r} & \tilde{r}\Theta_c \gg 1. \end{cases} \tag{A30}$$

In the case that the intrinsic effect of the mutation dominates the denominator of Equation A29 ($|\tilde{s}| \gg \tilde{r}e^{-\tilde{r}\Theta_c}$), one finds

$$\Phi(z) = \frac{z}{|\tilde{s}|}\left(1 - e^{-\tilde{r}\Theta_c}\right) = \begin{cases} \frac{z\tilde{r}\sqrt{-2\log z}}{|\tilde{s}|} & \tilde{r}\Theta_c \ll 1 \\ \frac{z}{|\tilde{s}|} & \tilde{r}\Theta_c \gg 1. \end{cases} \tag{A31}$$

The result for $\Phi(z)$ given by Equation A28 and its different limits is directly relevant to the calculation of the probability of secondary events, *e.g.*, an additional mutation as discussed in the earlier in *Stochastic tunneling and complex adaptations*.

For completeness we also calculate the probability distribution of $w = \int_0^\infty dt\, n(t)$ that is given by the inverse Laplace transform of $\Phi(z)$,

$$P(w) = \oint_{\mathcal{C}} \frac{dz}{2\pi i}\, e^{zw}\, \bar{\Phi}_z, \tag{A32}$$

where $\mathcal{C}$ is the contour encircling the branch cut. For example, in the intermediate asymptotic $\tilde{r}\Theta_c \ll 1$ regime when $\Phi(z) = z\sqrt{\log z^{-2}}$, we have

$$P(w) = \oint_{\mathcal{C}} \frac{dz}{2\pi i} \, e^{zw} z \sqrt{\log z^{-2}} = \int_0^\infty \frac{dy}{\pi} e^{-yw} y \, \mathrm{Im}\left[\sqrt{2\log y^{-1} + i2\pi}\right]. \tag{A33}$$

$$\approx \int_0^\infty \frac{dy\, y\, e^{-yw}}{\sqrt{\log y^{-2}}} \sim \frac{w^{-2}}{\sqrt{\log w^2}}, \tag{A34}$$

where we used an approximation valid for $w \gg 1$. This can be compared to the "draftless" result, obtained also in the $\tilde{r} \gg 1$ limit: in that case $\Phi(z) = \sqrt{z}$, the inverse Laplace transform of which gives $P(w) \sim w^{-3/2}$ (Weissman *et al.* 2009).

Appendix B

Computer simulation methods

Fisher–Wright simulation

To efficiently simulate the dynamics of large populations, we keep track of classes of individuals with identical genomes, which are encoded by bitstrings. The fitness x of all individuals of a class is simply the sum of the contributions from all loci. In our discrete generation scheme, a pool of gametes is produced to which each class contributes a Poisson-distributed number of gametes with mean $c\exp(-(x-\bar{x}+\alpha))$. Here, c is the size of the class, $\bar{x}$ is the mean fitness in the population, and $\alpha = 1 - N/\bar{N}$ keeps the population size N approximately constant at $\bar{N}$. Fitness defined as growth rate in the continuous-time model naturally assumes this exponential form in a discrete generation model, simply by integration over one generation.

To implement facultative mating, a fraction r of the gametes are paired up and from each pair two offspring are produced by assigning at random the genes of the parents to either offspring. The remaining $1 - r$ fraction of gametes is copied into the next generation without recombination; *i.e.*, they have gone through an asexual reproduction cycle.

The genome of each class has a fixed length $L = 1024$ and new mutations are introduced at a locus whenever this locus is monomorphic, *i.e.*, whenever a previous mutation either went extinct or fixed. For each monomorphic locus, an individual is chosen at random and the mutation is introduced in this single individual. This scheme of keeping a fixed number of loci polymorphic has the advantage of making optimal use of the computational resources, *i.e.*, keep every locus polymorphic while maintaining an overall low mutation rate. The mutation rate, however, becomes a dependent quantity that fluctuates around a value that depends on other parameters of the population: NU_b is simply the average number of loci that become monomorphic in one generation; see Neher *et al.* (2010). For large L, the number of mutations introduced in each generation is much greater than one and does not fluctuate greatly.

To establish a steadily adapting population, 80% of the loci are kept polymorphic with beneficial mutations with effect size s_0 which is rescaled each generation such that the overall fitness variance is $\sigma^2 = 0.0025$. This rescaling is done to be able to specify r/σ and s/σ explicitly. Since fluctuations of σ in large populations with many polymorphic loci are small, this does not change the properties of the dynamics and the same results are obtained letting σ be freely determined by the population. The remaining 20% of the loci are used to study the fate of mutations with fixed effect size s. Measurements are performed after an equilibration period of 20,000 generations at intervals of 100 generations. The source code is available from the authors on request.

Simulation of the branching process

The simulation of the population dynamics is complemented by a simulation of the branching process that can be directly compared to analytic calculations. We simulate the process described by Equation 1, using an event-driven algorithm (Gillespie 1977). The simulation keeps track of all individuals that currently carry the allele in question. For each individual, the time $t + \Delta t$ of the next event is determined by drawing a Δt from an exponential distribution with parameter $B(t) + D + r$, where B and D are the birth and death rates, respectively. At time $t + \Delta t$, a birth, death, or recombination event is performed with probabilities proportional to $B(t + \Delta t)$, D, and r, respectively. In the case of death, the individual is deleted; in the case of birth, it is duplicated with the exact same fitness; and in the case of recombination, its fitness is redrawn according to the recombination kernels in Equations 2 and 3. (Note that the waiting-time distribution for the next event is not exactly exponential, since the birth rate is time dependent. This, however, amounts only to a correction of order $\sigma^2 \ll 1$.) For the recombination function of the communal model, we need not simulate Equation 1 but can solve numerically for the generating function of the $p(n, T, t)$, given in Equation 17. To this end, we discretize the fitness x and solve for the vector $\psi(\tau, x)$, using the ODE solver of SciPy (Oliphant 2007). The results obtained through stochastic simulation of the communal model agree with the numerical solution for the generating function, as it has to be.
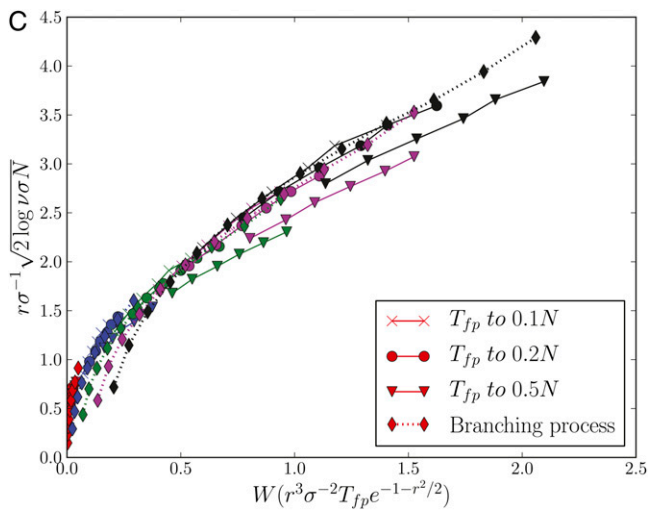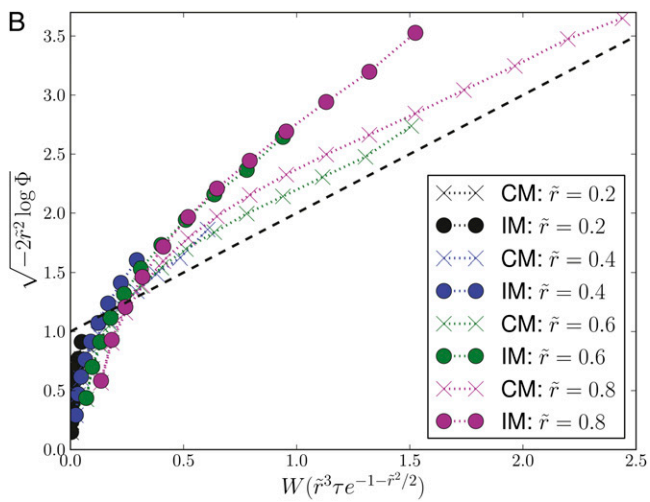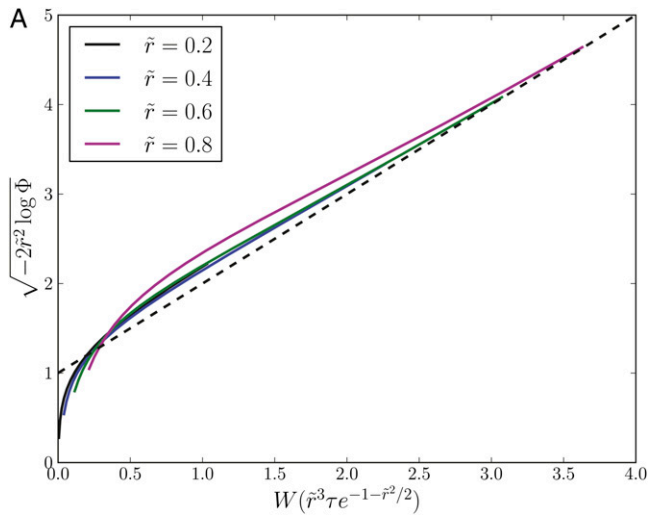
Figure A1 Asymptotic behavior of the survival probability for the communal and infinitesimal recombination models. (A) The survival probability $\Phi(\tau)$, obtained by numerical solution of Equation A1 for the communal recombination model, exhibits the predicted asymptotic behavior $\sqrt{-2\tilde{r}^2\log\Phi} \sim W(\tilde{r}^3\tau e^{-1-\tilde{r}^2/2})$ (Equation 5). (B) The survival probabilities in the infinitesimal model (IM) and the communal model (CM), determined by simulating the branching process, exhibit a similar scaling collapse, indicating that the dependence on parameters is similar. (C) The first passage time, T_{fp}, to allele copy number νN, is well described by the branching process approximation for the dynamics of alleles in a finite population as long as the frequency $\nu \ll 1$. This is confirmed here by plotting $\sqrt{2\tilde{r}^2\log\sigma\nu N}$ *vs.* $W(\tilde{r}^3\sigma T_{\text{fp}}e^{-1-\tilde{r}^2/2})$. The collapse of data for different $\tilde{r}$ and different ν confirms the predicted parameter dependence (Equation 7), while the concordance with the branching process results confirms the branching process model. Deviation from the branching process prediction is apparent for $\nu = 0.5$. Recombination rates in C follow the same color code as in A and B, while black symbols correspond to $\tilde{r} = 1$.

GENETICS

Genetic Draft and Quasi-Neutrality in Large Facultatively Sexual Populations
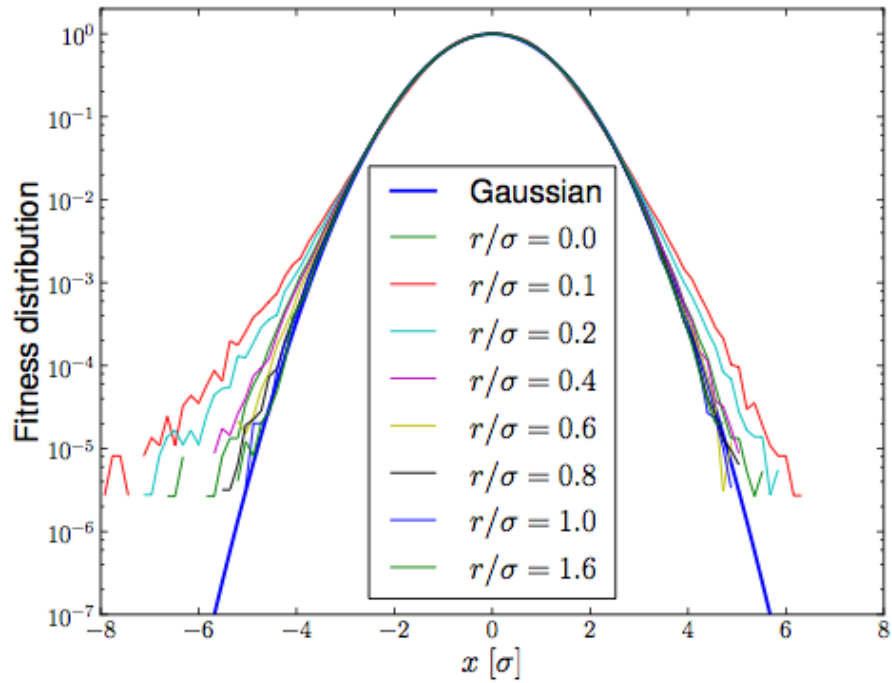
R. A. Neher and B. I. Shraiman

Figure S1 The fitness distribution of a population adapting at many loci is Gaussian. The figure shows the fitness distribution measured using our computational model for different ratios *r/σ*. Only at *r/σ* = 0.1 or 0.2, one sees (slight) deviations from the Gaussian. Our analysis applies in the range 1 > *r/σ* > 1/√log N , where the fitness distribution is Gaussian all the way into the tails.
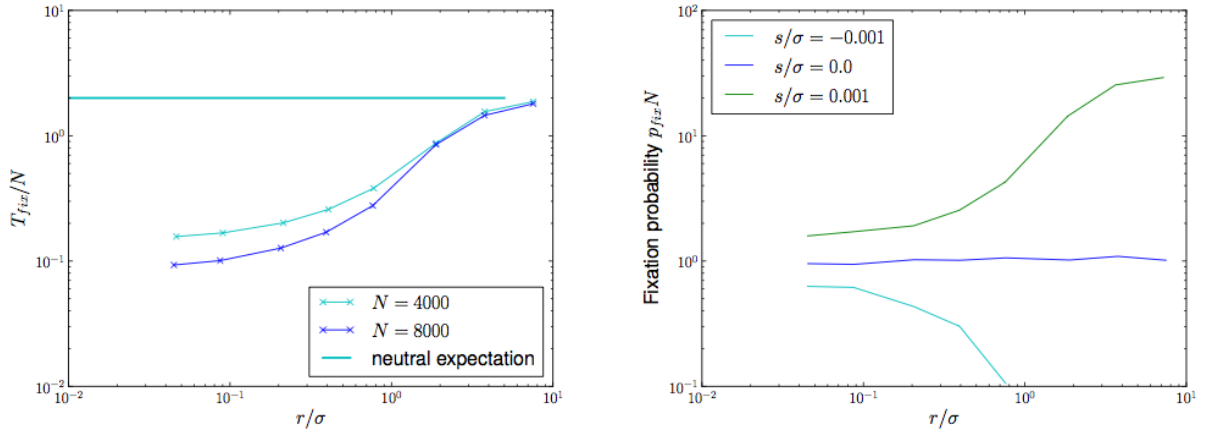
R. A. Neher and B. I. Shraiman

Figure S2 Fixation times and fixation probabilities with background selection, in analogy to Fig. 2 of the manuscript. The left panel shows the mean fixation time of neutral mutations in populations of different sizes at different ratios of r/σ. Fixation times are normalized by N. In contrast to Fig. 2 of the manuscript, the fitness variation here does not result from sweeping beneficial mutations but from many deleterious mutations. Comparing Fig. 2 to this figure, one sees that the effect of background selection on the fixation time of neutral mutations is very similar to that of multiple sweeps in a facultatively sexual population. The right panel shows the fixation probability of mutations with different effects on fitness for different ratios r/σ, normalized to the neutral expectation N^{-1}. Again, the fitness variance is due to background selection rather than sweeps, but the effect on the fixation probability is similar. These observations are consistent with the argument that from the perspective of a novel mutation, the nature of the fitness variation is irrelevant. What matters is the dynamics of the fitness of individual genotypes relative to the mean: For background selection, the fitness of genotypes declines due to accumulation of deleterious mutations, while in the case of adaptation the mean increases steadily.
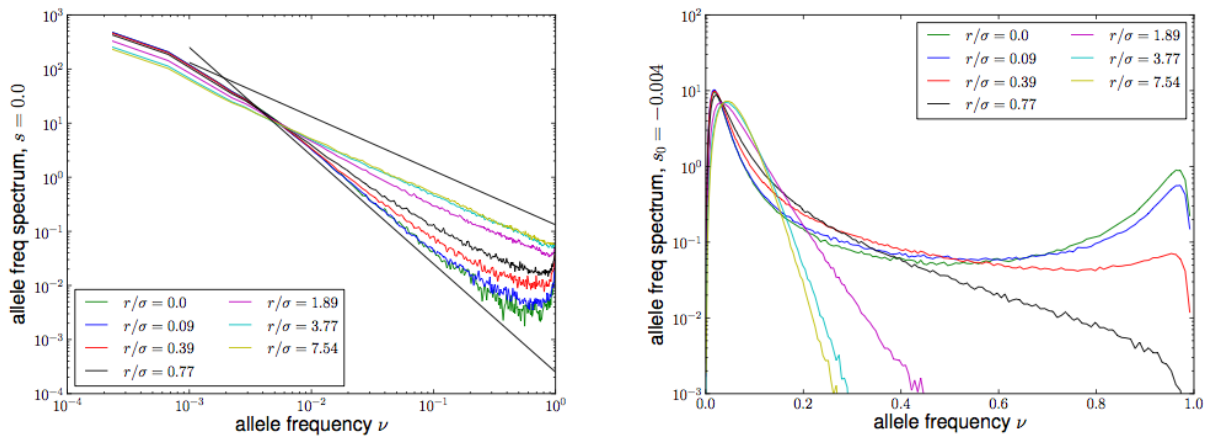
Figure S3 Allele frequency spectra of neutral (left) and deleterious mutations (right) in a background selection scenario, in analogy to Fig. 3b of the manuscript. At low recombination rates ($r/\sigma < 1$), the frequency spectrum of neutral mutations falls off much more rapidly than expected in a neutral model, very similar to what is observed for the scenario of continuous adaptation. The predicted behavior $\sim \nu^{-2}$ is indicated by the steeper black line. Only when $r/\sigma \gg 1$ does the spectrum agree with the neutral prediction ($\sim \nu^{-1}$ indicated by upper straight black line). The right panel shows the frequency spectrum of the deleterious mutations responsible for the fitness variation with effect size $s_0 = -0.004$. At low recombination rates, allele frequencies are close to fixation, either in the bad ($\nu = 1$) or the good ($\nu = 0$) state. At high recombination rates, the allele frequency spectra are distributed around their equilibrium value $\nu = \mu/s_0 = 0.0625$.

Derivation of the backward Master equation for $p(n, T)$

The fundamental quantity of the branching process is the probability distribution, $p(n, T)$, of observing n copies of the allele T generations after it originated. A backward equation for $p(n, T)$ can be derived by considering the probability $p(n, T|k, t, x)$ of having n copies at time T, given their were k copies on a background with fitness x at time t. $p(n, T|k, t - \Delta t, x)$ now can be expressed as a sum over possible intermediate states at time t:

$$
\begin{aligned}
p(n, T|k, t - \Delta t, x) = & p(n, T|k, t, x)(1 - \Delta t k(2 + x - \bar{x}(t) + s + r)) \\
& + \Delta t \left[k(1 + x - \bar{x}(t) + s)p(n, T|k + 1, t, x) + kp(n, T|k - 1, t, x) \right] \\
& + \Delta t r k \sum_{n'} \int_{x'} K_{xx'} p(n - n', T|k - 1, t, x)p(n', T|1, t, x') ,
\end{aligned}
\tag{1}
$$

where $\bar{x}(t)$ is the mean fitness of the population at time t. The different terms have straightforward interpretations: The first term is the probability that nothing happens in the small time interval Δt, the second accounts for a division of one of k individuals, which happens with rate $k(1 + x - \bar{x}(t) + s)$, the third term accounts for the death of one of the k, while the last term accounts for outcrossing of one of the k, producing a new individual with fitness x' and removing one with fitness x. The outcrossing term is then summed over all possible ways the $k - 1$ individuals with fitness x and the one with fitness x' can give rise to n individuals at time T. Sending Δt to zero and rearranging terms results in an ODE for $p(n, T|k, t, x)$

$$
\begin{aligned}
-\partial_t p(n, T|k, t, x) = & -p(n, T|k, t, x)k(2 + x - \bar{x}(t) + s + r) \\
& + k(1 + x - \bar{x}(t) + s)p(n, T|k + 1, t, x) + kp(n, T|k - 1, t, x) \\
& + rk \sum_{n'} \int_{x'} K_{xx'} p(n - n', T|k - 1, t, x)p(n', T|1, t, x')
\end{aligned}
\tag{2}
$$

This is equation 1 from the main text.

Derivation of the equation for the generating function

To remove the convolution over n it is convenient to consider the generating function $\hat{p}(\lambda, T|k, t, x) = \sum_n \lambda^n p(n, T|k, t, x)$. Multiplying the above equation by λ^n and summing over n yields

$$
\begin{aligned}
-\partial_t \hat{p}(\lambda, T|k, t, x) = & -k(2 + x - \bar{x}(t) + s + r)\hat{p}(\lambda, T|k, t, x) \\
& + k(1 + x - \bar{x}(t) + s)\hat{p}(\lambda, T|k + 1, t, x) + k\hat{p}(\lambda, T|k - 1, t, x) \\
& + rk \int_{x'} K_{xx'} \hat{p}(\lambda, T|k - 1, t, x)\hat{p}(\lambda, T|1, t, x)
\end{aligned}
\tag{3}
$$

All k initial individuals are independent, hence $\hat{p}(\lambda, T|k, t, x) = \hat{p}^k(\lambda, T|t, x)$ and the right hand side is $-\partial_t \hat{p}^k(\lambda, T|t, x) = -k\hat{p}^{k-1}(\lambda, T|t, x)\partial_t \hat{p}(\lambda, T|t, x)$. We can therefore divide the equation by $\hat{p}_x^{k-1}(\lambda, T|t, x)$ to obtain

$$
-\partial_t \hat{p}(\lambda, T, t, x) = -(2 + x - \bar{x}(t) + s + r)\hat{p}(\lambda, T, t, x) + (1 + x - \bar{x}(t) + s)\hat{p}^2(\lambda, T, t, x) + 1 + r \int_{x'} K_{xx'} \hat{p}(\lambda, T, t, x)
\tag{4}
$$

The generating function has the boundary condition $\hat{p}(\lambda, T|T, x) = \lambda$, which follows from $p(n, T|k, T, x) = \delta_{nk}$. Substituting $\hat{p}(\lambda, T|t, x) = 1 - \phi(\lambda, T, t, x)$ removes the constant term.

$$
\partial_t \phi(\lambda, T, t, x) = -r \int_{x'} K_{xx'} \phi(\lambda, T, t, x') - (x - \bar{x}(t) + s - r)\phi(\lambda, T, t, x) + (1 + x - \bar{x}(t) + s)\phi^2(\lambda, T, t, x)
\tag{5}
$$

which now has boundary condition $\phi(\lambda, T, T, x) = 1 - \lambda$. Assuming that selection is weak on the timescale of one generation ($\sigma \ll 1$), we can approximate $1 + x - \bar{x}(t) + s$ by 1 and arrive at Eq. 18 of the main text:

$$-\partial_t \phi(\lambda, T, t, x) = r \int_{x'} K_{xx'} \phi(\lambda, T, t, x') + (x - \bar{x}(t) + s - r)\phi(\lambda, T, t, x) - \phi^2(\lambda, T, t, x) \qquad (6)$$

Solution for $\Phi(\tau)$ at low $\tilde{r}$

In the main text, we presented a solution to the two equations (Eqs. (21) and (23))

$$\psi(\tau, \chi) = \begin{cases} r e^{\theta^2/2} \int_0^\tau d\tau' \Phi(\tau') e^{-\theta'^2/2} & \theta < \Theta_c \\ \theta/\epsilon & \theta > \Theta_c \end{cases} \qquad (7)$$

and

$$\partial_\tau \Phi(\tau) = \tilde{s}\Phi(\tau) - \epsilon \int d\chi P(\tau, \chi)\psi^2(\tau, \chi) \qquad (8)$$

in a regime of intermediate $\tilde{r}$ ($1/\sqrt{\log N} < \tilde{r} < 1$). An additional solution with qualitatively different properties exists at low $\tilde{r}$. While the assumption of a Gaussian fitness distribution is questionable in this range of $\tilde{r}$, we nevertheless present the solution here for completeness. Proceeding as before, we evaluate the integral in Eq. (7) by expanding exponent around its maximum. The maximum is located at $\tau' = \tau - \theta - \alpha(\tau')$ where $\alpha(\tau) = -\Phi(\tau)^{-1}\partial_\tau \Phi(\tau)$. Assuming $\alpha(\tau)$ changes slowly with time, we have

$$\psi(\tau) \approx \sqrt{2\pi}\tilde{r}\Phi(\tau)e^{\frac{(\theta+\alpha)^2}{2}} \qquad (9)$$

This solution is valid below Θ_c, where $\psi(\tau)$ crosses over the linear saturated form θ/ϵ. In addition to the matching condition at Θ_c, we use Eq. 8 for $s = 0$ to determine α:

$$\partial_\tau \Phi(\tau) = -\alpha\Phi(\tau) = -\tilde{r}\Phi(\tau)e^{\Theta_c(\alpha-\tilde{r})+\alpha^2/2-\tilde{r}^2/2} \qquad (10)$$

For $\tilde{r}\Theta_c \gg 1$ we recover the solution with small α given in the main text. For smaller $\tilde{r}$, however, we find $\alpha(\tau) \approx \frac{\log \tilde{r}}{\Theta_c}$. Solving the matching condition for Θ_c and differentiating with respect to τ yields $\partial_\tau \Theta_c = -\frac{1}{\Theta_c}\alpha(\tau)$. This is readily solved for this case of $\Theta_c\tilde{r} < 1$

$$\Theta_c(\tau) = (-3(\tau - \tau_0)\log \tilde{r} + \Theta_0^3)^{1/3} . \qquad (11)$$

Substituting this solution for $\Theta_c(\tau)$ into the expression for the rescaled generating function yields

$$\Phi(\tau) = \frac{\Theta_c}{\epsilon\tilde{r}}e^{-\frac{(\Theta_c+\alpha)^2}{2}} \sim e^{-\frac{\left[-3(\tau-\tau_0)\log \tilde{r}+\Theta_0^3\right]^{2/3}}{2}} \qquad (12)$$

Hence in this low $\tilde{r}$ regime, the decay of the survival probability is qualitatively different from the regime of intermediate $\tilde{r}$.

Effective clone-based model

To rationalize the behavior of the continuous time branching process, we considered the following simplified model discussed in the main text: Genotypes expand clonally and produce recombinant offspring. The offspring start growing simultaneously in the next "effective" generation after all clones from the previous generation have disappeared. The relevant quantity now is the number of clones or distinct genotypes, rather than the number of individuals. To understand the dynamics of the number of clones, we need to know how many clones a single clone can produce.

Consider a single genotype with background fitness χ which is carrying a mutation of effect size $\tilde{s}$. The expected number of recombinant offspring from this genotype is $\xi = \tilde{r} \int_0^\infty dt n_\chi(t)$, where $n_\chi(t)$ is the copy number trajectory. The Laplace transform $\hat{p}(\xi)$ of $p(\xi)$ obeys the equation ($\phi(z) = 1 - \hat{p}(z)$)

$$\partial_\chi \phi(\chi, z) = \tilde{r}z + (\chi + \tilde{s} - \tilde{r}(1 - z))\phi(\chi, z) - \phi(\chi, z)^2 , \qquad (13)$$

which is a simpler version of the Eq. (42) (main text) since only one single clone is considered and recombination to daugther clone is ignored. This equation can be solved asymptotically in the regimes of large and small $\chi - r$.

$$\phi(\chi, z) = \begin{cases} \tilde{r}ze^{(\chi+\tilde{s}-\tilde{r}(1-z))^2/2} \int_{-\infty}^{\chi+\tilde{s}-\tilde{r}(1-z)} dx e^{-x^2/2} & \chi - \tilde{r} \ll \Theta_c \\ (\chi - \tilde{r}(1-z)) & \chi - \tilde{r} \gg \Theta_c \end{cases} \tag{14}$$

where $\Theta_c \approx \sqrt{-2\log \tilde{r}z}$. The initial fitness of the genotype, χ, is Gaussian distributed and the Laplace transform has to be averaged over χ.

$$\phi(z) = \int_{-\infty}^{\infty} \frac{d\chi}{\sqrt{2\pi}} e^{-\chi^2/2} \phi(\chi, z) \tag{15}$$

Let's look at the mean number $\langle \xi \rangle$ of recombinant offspring first, which is given by differentiating with respect to z and setting $z = 0$, which in turn sends Θ_c to infinity. Integration by parts yields

$$\langle \xi \rangle = \partial_z \int_{-\infty}^{\infty} \frac{d\chi}{\sqrt{2\pi}} e^{-\chi^2/2} \phi(\chi) = \tilde{r} \int_{-\infty}^{\infty} \frac{d\chi}{\sqrt{2\pi}} e^{-\chi(\tilde{r}-\tilde{s})+r^2/2} \int_{-\infty}^{\chi-\tilde{r}+\tilde{s}} dx e^{-x^2/2} = \frac{\tilde{r}}{\tilde{r} - \tilde{s}} \tag{16}$$

Hence the mean number of recombinant offspring is 1 for a neutral mutation and approximately $1 + \tilde{s}/\tilde{r}$ for mutations with small effect. To evaluate the integral in Eq. (15) at finite z, we have to account for the cross-over of $\phi(\chi, z)$ at $\chi - \tilde{r} = \Theta_c \approx \sqrt{-2\log \tilde{r}z}$. For small z, the cross-over translates into a cut-off of the integral. Again, the integral can be evaluated by parts:

$$\begin{aligned} \phi(z) &= \int_{-\infty}^{\Theta_c+\tilde{r}-\tilde{s}} \frac{d\chi}{\sqrt{2\pi}} \tilde{r}z e^{-\chi(\tilde{r}(1-z)-\tilde{s})+\tilde{r}^2(1-z)^2/2} \int_{-\infty}^{\chi-\tilde{r}(1-z)+\tilde{s}} dx e^{-x^2/2} + \int_{\Theta_c+\tilde{r}-\tilde{s}}^{\infty} \frac{d\chi}{\sqrt{2\pi}} e^{-\chi^2/2} (\chi - \tilde{r}(1-z)) \\ &\approx \frac{\tilde{r}z}{\tilde{r} - \tilde{s}} \left[1 - e^{-\Theta_c(\tilde{r}-\tilde{s})-\tilde{r}^2/2} \right] \end{aligned} \tag{17}$$

Hence, the generating function of the average number of recombinant offspring generated by a random genotype is given by

$$\hat{p}(z) = 1 - \phi(z) \approx 1 - \frac{\tilde{r}z}{\tilde{r} - \tilde{s}} \left[1 - e^{-\Theta_c(\tilde{r}-\tilde{s})} \right] \,, \tag{18}$$

where the last expression is valid for $z \ll 1$ and $\tilde{r} \ll 1$.

The actual number m of recombinant offspring (novel clones) generated by one clone is Poisson distributed with mean ξ. The generating function of m is therefore

$$\sum_m \lambda^m P(m) = \sum_m \lambda^m \int_0^{\infty} d\xi \, p(\xi) e^{-\xi} \frac{\xi^m}{m!} = \int_0^{\infty} d\xi \, p(\xi) e^{-\xi(1-\lambda)} = \hat{p}(1 - \lambda) = 1 - \phi(1 - \lambda) \tag{19}$$

We will now use this result to calculate how the number of clones that descend from a particular genotype evolves over time.

The stochastic dynamics of the number genotypes

As long as the clones we are tracking constitute a small fraction of the population, different clones are independent. The probability to go from k to m clones in one effective generation has therefore the generating function $\hat{P}(\lambda, k) = \hat{P}^k(\lambda, 1) = (1 - \phi(1 - \lambda))^k$. To study the dynamics of the number of clones over many generations, we need to know how this propagator behaves when iterated.

$$\begin{aligned} \sum_m \lambda^m \sum_{m_1, \dots m_n} P(m, m_n) P(m_n, m_{n-1}) \dots P(m_1, k) &= \sum_{m_n} \hat{P}^{m_n}(1 - \lambda) \sum_{m_1, \dots m_{n-1}} P(m_n, m_{n-1}) \dots P(m_1, k) \\ &= \left[\hat{P}(1 - \hat{P}(1 - \hat{P}(1 - \dots \hat{P}(1 - \lambda)))) \right]^k = [1 - \phi \circ \phi \dots \circ \phi(1 - \lambda)]^k = [1 - \Phi_n(1 - \lambda)]^k \end{aligned} \tag{20}$$

Using the result for the Laplace transform in Eq. (18), we arrive at the difference equation

$$\Phi_{n+1} - \Phi_n \approx \left[\tilde{s}/\tilde{r} - e^{-\tilde{r}\sqrt{-2\log \tilde{r}\Phi_n}} \right] \Phi_n \tag{21}$$

This is exactly the differential equation we have derived using the continuous time mode when the effective generation time is set to $\tilde{r}^{-1}$. The latter is reasonable since the $\tilde{r}^{-1}$ is the turnover time by recombination.