

# Limits to selection on standing variation in an asexual population

Nick Barton<sup>1,\*§</sup> and Himani Sachdeva<sup>2,\*</sup>

<sup>1</sup>Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg 3400, Austria.

<sup>2</sup>Department of Mathematics, University of Vienna, Vienna 1090, Austria.

\*equal contribution

§[nick.barton@ist.ac.at](mailto:nick.barton@ist.ac.at)

## Abstract

We consider how a population responds to directional selection on standing variation, with no new variation from recombination or mutation. Initially, there are  $N$  individuals with trait values  $z_1, \dots, z_N$ ; the fitness of individual  $i$  is proportional to  $e^{z_i}$ . The initial values are drawn from a distribution  $\psi$  with variance  $V_0$ ; we give examples of the Laplace and Gaussian distributions. When selection is weak relative to drift ( $N\sqrt{V_0} \ll 1$ ), variance decreases exponentially at rate  $1/N$ ; since the increase in mean in any generation equals the variance, the expected net change is just  $NV_0$ , which is the same as Robertson's (1960) prediction for a sexual population. In contrast, when selection is strong relative to drift ( $N\sqrt{V_0} \gg 1$ ), the net change can be found by approximating the establishment of alleles by a branching process in which each allele competes independently with the population mean and the fittest allele to establish is certain to fix. Then, if the probability of survival to time  $t \sim 1/\sqrt{V_0}$  of an allele with value  $z$  is  $P(z)$ , with mean  $\bar{P}$ , the winning allele is the fittest of  $N\bar{P}$  survivors drawn from a distribution  $\psi P/\bar{P}$ . When  $N$  is large, there is a scaling limit which depends on a single parameter  $N\sqrt{V_0}$ ; the expected ultimate change is  $\sim \sqrt{2 \log(1.15N\sqrt{V_0})}$  for a Gaussian distribution, and  $\sim \frac{1}{\sqrt{2}} \left( \mathcal{W} \left( \frac{0.36}{N\sqrt{V_0}} \right) - 1 \right)$  for a Laplace distribution (where  $\mathcal{W}$  is the product log function). This approach also reveals the variability of the process, and its dynamics; we show that in the strong selection regime, the expected genetic variance decreases as  $\sim t^{-3}$  at large times. We discuss how these results may be related to selection on standing variation that is spread along a linear chromosome.

We consider a simple and fundamental problem: how does a large asexual population respond to selection? We assume that selection acts steadily on standing variation, with no input from either mutation or recombination. Thus, the only parameters are the number,  $N$ , of individuals, and their fitnesses.

It is surprising that, despite its simplicity, this problem has not (to our knowledge) been addressed before. This may be because population genetics deals mainly with competition between two alleles per locus, whilst quantitative genetics has focused on genetic variance in sexual populations. Two bodies of work lie closest to the problem studied here. Robertson (1960) showed that directional selection on standing variation under the infinitesimal model causes an ultimate response, i.e., a total change in mean, equal to the genetic population size (i.e.  $2N$  for  $N$  diploid individuals) times the change in the first generation. Moreover, he showed that the result implicitly assumes that allele frequencies evolve almost neutrally. Paixão and Barton (2016) extended this result to arbitrary epistasis, and also considered the opposite regime of strong selection. However, these results all apply to a multitude of freely recombining loci in a sexual population.

More recently, there has been a mass of work on fitness waves: the steady increase in the fitness of a population as it accumulates beneficial mutations. This work, which initially sought to explain the rate of fitness gain in serial transfer experiments with RNA viruses (Novella et al 1995), showed that the long-term dynamics of the fitness distribution in an asexual population can be described as a traveling wave with a *constant* fitness variance (Tsimring et al 1996; Rouzine et al 2003). Similar models also apply to the converse problem, of the steady accumulation of deleterious mutations through Muller’s Ratchet (Haigh 1978; Rouzine et al 2008). A key challenge is to account for the stochastic loss or establishment of very fit lineages at the leading edge of the wave, which is what ultimately maintains constant variance despite continual mutational input. Most of this work thus relies on some kind of ‘semi-deterministic’ approximation of the fitness wave, patching together a deterministic description of evolution in the bulk with a stochastic treatment of the dynamics of new mutants at the leading edge; this allows for approximate predictions for the speed of the wave, its variance and the genealogies of populations described by such waves (Desai and Fisher 2007; Good et al 2012; Fisher 2013; Melissa et al 2022). Rigorous results have been obtained only in simpler cases, typically by assuming a rather extreme form of selection (Brunet et al 2007; Berard and Gouéré 2010; Berestycki et al 2013) (though see Cortines (2016); Cortines and Mallein (2017)).

However, neither of these bodies of work apply to the transient response to selection on standing variation, in the absence of mutation or recombination. For example, consider a scenario where initially, there are  $N$  alleles, with some distribution of fitness. If selection were perfectly efficient, the fittest of these would fix, and its fitness would follow an extreme value distribution. However, even beneficial alleles are likely to be lost in the first few generations just by chance, and so, even with rather strong selection, the winning allele will be the survivor from this early stochastic phase. Here we show that if the initial fitness distribution is Gaussian, the winning allele is unlikely to be more than a few standard deviations above the mean, even if the population is very large; fat-tailed distributions would allow for larger but more erratic responses to selection.

Taken literally, such a model would apply to the short-term evolution of an asexual population, but it may also be used to approximate selection on alleles of a single gene, in a sexual population. When averaged over the genome of an outcrossing eukaryote, selection is likely to be weaker than recombination. However, selection may be concentrated on variation (both regulatory and coding) at a single gene. For example, in *Drosophila*, a 10kb region would experience recombination at a rate  $10^{-4}$ ; such a region might contain hundreds of segregating sites, and hence an extremely large number of haplotypes. If the selective differences between these were greater than  $10^{-4}$ , then favourable combinations can increase intact (Sachdeva and Barton 2018). Thus, even though it neglects recombination, an asexual model may still be a useful starting point for understanding how selection acts when concentrated on small regions of a sexual genome.

In the following, we first use simulations to confirm that we can rescale time and selective value relative to the standard deviation  $\sqrt{V_0}$  of fitness; the outcome then depends on a single parameter,  $N\sqrt{V_0}$  (which corresponds to  $Ns$  in classical models). When  $N\sqrt{V_0}$  is small, variance is lost mainly by random drift, over  $\sim N$  generations. Because the increase in mean due to selection equals the genetic variance in fitness (Fisher 1930), in this quasi-neutral parameter regime, the ultimate response is  $N$  times the change in the first generation, analogous to Robertson’s (1960) result. However, when  $N\sqrt{V_0}$  is large, variance is lost much faster, through the increase of an exceptionally fit allele that eventually fixes. We show that in this regime, the outcome can be predicted by finding the value of the fittest survivor at some intermediate time  $T=t\sqrt{V_0}\sim 1$ , using a branching process approximation. This approach gives us both the distribution of the ultimate response, and the rate of loss of fitness variance.

## Model

The population consists of  $N$  haploid individuals associated with log fitness values  $\{z_i\}$ ,  $i=1, \dots, N$ , drawn from some initial distribution  $\psi(z)$  with mean zero. Individuals reproduce asexually under a Wright-Fisher model; an individual's chance of being chosen as a parent is proportional to  $e^z$ . For the most part, we assume that log fitness values are drawn from a Gaussian distribution with variance  $V_0$ ; however, we also consider heavier tailed distributions that are much more likely to generate highly fit individuals.

We neglect mutation throughout, since the aim is to understand response from standing genetic variation. Thus, in the long run, i.e., as  $t \rightarrow \infty$ , one of the alleles present at  $t=0$  must fix, resulting in a net change  $\Delta_\infty \bar{z}$  in mean log fitness, which is equal to the value of this allele. Here, we aim to understand how fixation probabilities and the dynamics of fixation depend on the initial distribution of fitness values and  $N$ .

**Scaling limit.** This model has only two parameters,  $N$  and  $\sqrt{V_0}$ , which have dimensions of time and inverse time respectively. Thus, in the scaling limit—  $N \rightarrow \infty$ ,  $\sqrt{V_0} \rightarrow 0$ , with  $N\sqrt{V_0}$  constant— evolutionary dynamics (expressed in terms of scaled time  $T=t\sqrt{V_0}$  and scaled log fitness  $Z=z/\sqrt{V_0}$ ) should depend only on a single non-dimensional parameter  $N\sqrt{V_0}$ . A scaling limit should exist for any initial distribution as long as it has finite variance. Moreover, in the scaling limit, the diffusion approximation should apply, and details of reproduction matter only via an effective population size; therefore, we expect that our results would extend more generally beyond Wright-Fisher reproduction.

Figure 1(a) shows the expected ultimate change  $\mathbb{E}[\Delta_\infty \bar{Z}]$  in mean log fitness as a function of  $N\sqrt{V_0}$  for various  $N$  (different symbols), where  $\bar{Z}=\bar{z}/\sqrt{V_0}$  is scaled relative to the initial standard deviation of fitness. Results are shown for two different initial distributions— Gaussian ( $\psi(Z)=e^{-Z^2/2}/\sqrt{2\pi}$ ; shown in shades of blue) and Laplace or bi-exponential ( $\psi(Z)=e^{-\sqrt{2}|Z|}/\sqrt{2}$ ; shades of red) both of which are symmetric about  $Z=0$ ; however, our results apply more generally also to asymmetric distributions. The expectation is obtained by averaging over  $10^3$ - $10^4$  simulation replicates for each parameter combination. Figure 1(b) quantifies the variability of net advance across replicates by plotting the the coefficient of variation (CV)— the ratio of the standard deviation to the expectation of the ultimate change— for the same parameters as in fig. 1(a). Solid black lines and dashed colored lines show analytical predictions under strong selection ( $N\sqrt{V_0} \gg 1$ ) in the scaling ( $N \rightarrow \infty$ ) limit and for finite  $N$  respectively; dotted black lines show the weak selection ( $N\sqrt{V_0} \ll 1$ ) expectations (see below). Both the expected ultimate change and its CV become independent of  $N$  for large  $N$  and follow the scaling closely. However, the convergence to the scaling limit (with increasing  $N$ ) is slower for larger values of  $N\sqrt{V_0}$  and if the initial distribution has heavier tails.

Figure 1 points towards distinct behaviours of the selection response in the  $N\sqrt{V_0} \ll 1$  (low initial variance or weak selection) and  $N\sqrt{V_0} \gg 1$  (high initial variance or strong selection) parameter regimes. We first describe these behaviours qualitatively and then outline a branching process approximation which yields surprisingly accurate predictions (solid black and dashed colored lines in fig. 1) for the distribution of the ultimate change when  $N\sqrt{V_0} \gg 1$ .

**Response under weak selection.** For  $N\sqrt{V_0} \ll 1$ , the expected ultimate (scaled) change is close to  $N\sqrt{V_0}$  (shown by the dotted black curve in fig. 1(a)), regardless of the shape of the initial distribution. This is precisely the net advance under directional selection in a sexual population under the infinitesimal model (Robertson 1960). It can be understood in the same way, by noting that under weak selection, fitness variance decays neutrally, i.e., by  $1/N$  per generation on average (see also fig. 4(b)), and that the expected change in  $\bar{z}$  per generation is equal to the

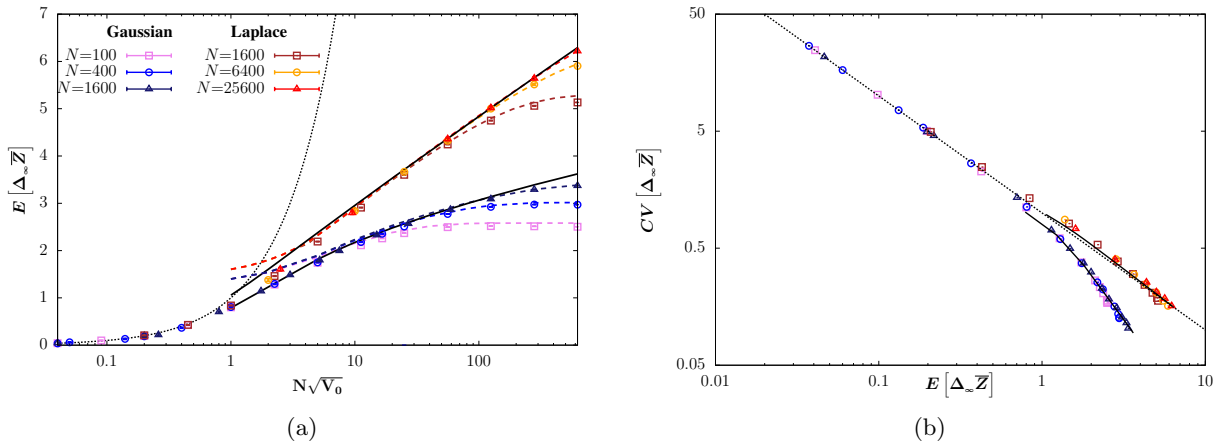


Figure 1: (a) Expected ultimate (scaled) change  $\mathbb{E}[\Delta_\infty \bar{Z}]$  in mean log fitness and (b) its coefficient of variation (CV) vs.  $N\sqrt{V_0}$ , for various  $N$  (different symbols). Results are shown for Gaussian (shades of blue) or Laplace (shades of red) initial distributions. Symbols show simulation results; dotted black lines depict  $\mathbb{E}[\Delta_\infty \bar{Z}] = N\sqrt{V_0}$  in (a) and  $CV = 1/\mathbb{E}[\Delta_\infty \bar{Z}]$  in (b), which appear to describe response under weak selection ( $N\sqrt{V_0} \ll 1$ ). Solid black lines and colored dashed lines show the scaling ( $N \rightarrow \infty$ ) and finite  $N$  predictions respectively in the strong selection ( $N\sqrt{V_0} \gg 1$ ) regime; finite  $N$  predictions for the CV are not shown since these are essentially indistinguishable from the  $N \rightarrow \infty$  predictions (solid black lines in (b)). Scaling limit predictions for  $\mathbb{E}[\Delta_\infty \bar{Z}]$  and the CV in the strong selection regime are obtained by integrating over the distribution of ultimate response in eq. (5) (and using eq. (3) to approximate the survival probability of fit alleles, with  $T_*$  set to 1 for the Gaussian and to 0.625 for the Laplace distribution); finite  $N$  predictions are obtained by integrating over eq. (12) in the Appendix.

variance of  $z$ . Therefore, the total change in mean is  $\mathbb{E}[\Delta_\infty \bar{z}] = \sum_{t=0}^{\infty} \mathbb{E}[V_t] = V_0 \sum_{t=0}^{\infty} (1 - \frac{1}{N})^t = NV_0$ ; scaling relative to the initial standard deviation gives  $N\sqrt{V_0}$ .

However, unlike in sexual populations with very many loci, here (i.e., in the asexual case), the ultimate change is determined by the fitness of the single allele that fixes; it is thus highly variable across replicates, being  $NV_0$  only *on average* (for the unscaled mean). In fact, the variance of the ultimate change is approximately equal to the initial variance  $V_0$  of log fitness values, such that the variance within populations is converted entirely to variance across replicates over time. Consequently, the CV of the ultimate change is  $1/\mathbb{E}[\Delta_\infty \bar{Z}]$  (fig. 1(b)), where  $\Delta_\infty \bar{Z} = \Delta_\infty \bar{z} / \sqrt{V_0}$  is the scaled response.

**Response under strong selection.** By contrast, for  $N\sqrt{V_0} \gg 1$ , the ultimate change is much lower than  $N\sqrt{V_0}$  and depends strongly on the initial distribution  $\psi(z)$ , being larger for longer-tailed distributions. As we show below, in this regime, selection response is governed by exceptionally fit alleles in the tail of the distribution; the fittest of these to escape initial stochastic loss is the one to fix. For instance, with a Gaussian initial distribution and for  $N\sqrt{V_0}$  in the range 10 to 1000, the allele that eventually fixes has (on average) a fitness value between 2 and 3 standard deviations above the mean, while for a Laplace initial distribution, somewhat fitter alleles (between 3 and 7 standard deviations above the mean) fix (fig. 1(a)).

With increasing  $N\sqrt{V_0}$ , response also becomes more consistent across replicates—the CV of the ultimate change is smaller than 1 in this regime, though still non-zero in the scaling limit (fig. 1(b)). The extent of variability now depends on the shape of the initial distribution, with longer-tailed distributions resulting in higher (on average) but more variable response.

There is also a marked dependence on population size in this regime, with deviations from the scaling limit noticeable even for  $N$  in the tens of thousands, e.g., in the case of the Laplace initial distribution. This can be understood by noting that the ultimate change must be limited

by the fitness of the fittest allele present in the population. This in turn follows a Gumbel distribution with a mean that scales as  $\sqrt{\log(N)}$  and  $\log(N)$  for Gaussian and Laplace initial distributions respectively. Thus, for the scaling prediction to hold,  $\log(N)$  must be sufficiently large that the fitness of the fittest allele present at  $t=0$  is not limiting. A related observation (which we discuss in detail later) is that the ultimate change does nevertheless reach a scaling limit which depends only on  $N\sqrt{V_0}$ , even though the fitness of the fittest allele present at  $t=0$  diverges (albeit weakly) with increasing  $N$ .

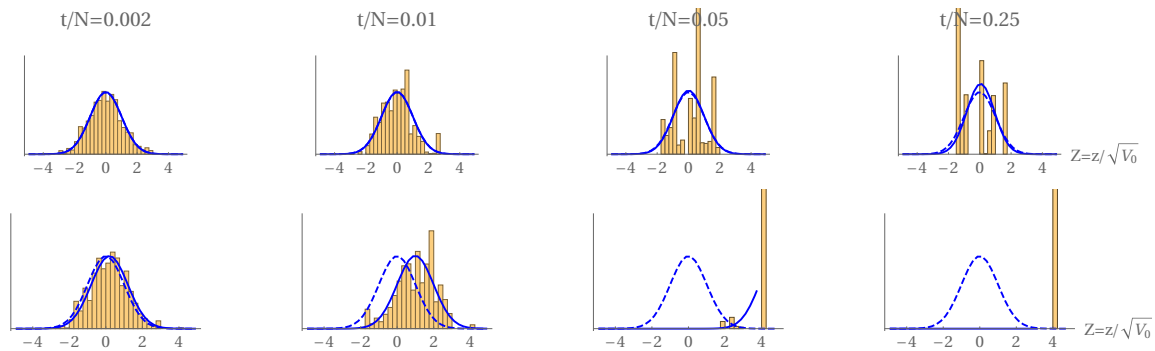


Figure 2: The evolution of a population of  $N=10^5$  individuals, starting from an initial Gaussian with  $N\sqrt{V_0}=0.25, 100$ , and hence variance  $V_0=6.25\times 10^{-12}$  or  $10^{-6}$  (top, bottom). The columns show the distributions of  $z/\sqrt{V_0}$  at  $t=200, 1000, 5000, 25000$  generations. The dashed curve shows the initial distribution at  $t=0$ ; the solid smooth curve shows a Gaussian with variance that decreases due to drift, as  $V_0e^{-t/N}$ , and mean  $\bar{z}=V_0N(1-e^{-t/N})$ .

We can understand the switch between the weak selection and strong selection regimes by noting that alleles in the tail of the distribution— with fitness values a few standard deviations above the mean— survive initial stochastic loss with a probability that is  $\mathcal{O}(\sqrt{V_0})$  and establish, i.e., attain appreciable numbers (conditional on survival), over a timescale that is  $\mathcal{O}(1/\sqrt{V_0})$ . On the other hand, drift erodes diversity in the bulk of the distribution over  $\sim N$  generations. Thus, the fittest surviving allele will dominate before drift has eroded diversity only if  $1/\sqrt{V_0}\ll N$ , i.e.,  $N\sqrt{V_0}\gg 1$ ; its subsequent trajectory as it outcompetes all other alleles is close to deterministic. Conversely, if the two timescales are comparable, i.e.,  $N\sqrt{V_0}\lesssim 1$ , all alleles, regardless of their fitness, are affected appreciably by drift over the entire time taken for one allele to fix and the dynamics do not separate neatly into a stochastic and deterministic phase.

This is illustrated in Figure 2, which shows the distribution of fitness values at various time instants for  $N\sqrt{V_0}=0.25$  (top) and  $N\sqrt{V_0}=100$  (bottom). As a reference point, the figures also show a Gaussian distribution whose variance declines over time solely due to drift (solid blue curves). In the small  $N\sqrt{V_0}$  regime (top row), the fitness distribution always lies roughly within this Gaussian envelope, though there is considerable stochastic variation across replicates, especially at long times when only a few alleles survive. By contrast, in the large  $N\sqrt{V_0}$  regime (bottom row), the (bulk of the) distribution is Gaussian only till  $t/N\sim 0.01$  (which corresponds here to  $\sqrt{V_0}t=1$ ), by which time a winning allele starts to emerge. In this example, the winning allele is approximately 3.3 standard deviations above the mean, which corresponds to deterministic growth by a factor of  $\sim e^{3.3}\approx 27$  during this period; however, growth conditional on survival is much faster and there are already 569 copies of this allele by  $t/N=0.01$ , so that it grows more or less deterministically at subsequent times and has fixed by  $t/N=0.17$ .

### A branching process approximation.

For  $N\sqrt{V_0}\gg 1$ , we can approximate the outcome by following the increase of each allele as an independent branching process in which an allele of value  $z_i$  competes with the mean  $\bar{z}(t)$  of

the rest of the population. If establishment occurs over short time scales, i.e., over  $t \ll N$  (as expected for  $N\sqrt{V_0} \gg 1$ ; see above), then the decline in variance due to drift can be neglected, and the population mean can be approximated by  $\bar{z}(t) \approx V_0 t$ . Thus, the allele experiences a time-dependent selective advantage  $s(t) \approx z_i - V_0 t$  during initial establishment.

Let  $Q(t_0)$  denote the probability that an allele with advantage  $s(t)$  is lost by some time  $t_1$ , conditional on it being present in  $n_0$  copies at time  $t_0$ . Then,  $Q(t_0) = \exp[-n_0 P(t_0)]$ , where  $P(t_0)$  is the probability of survival (until time  $t_1$ ) of the allele present as a single copy at  $t_0$ . This form follows from the branching process approximation, which assumes that each of the  $n_0$  copies is lost independently, and also assuming  $P(t_0) \ll 1$ . In the scaling limit, the change in survival probability over time can be approximated by a continuous-time diffusion process. Scaling by  $\sqrt{V_0}$ , such that  $\tilde{P}(T_0) = P(T_0)/\sqrt{V_0}$ ,  $Z = z/\sqrt{V_0}$ ,  $T = \sqrt{V_0}t$  and  $S(T) = Z - T$ , the (scaled) survival probability follows the Kolmogorov backward equation:

$$\frac{\partial \tilde{P}(T_0)}{\partial T_0} = -S(T_0) \tilde{P}(T_0) + \frac{1}{2} \tilde{P}^2(T_0) \quad \text{with} \quad \lim_{T_0 \rightarrow T_1} \tilde{P}(T_0) = \frac{1}{\sqrt{V_0}} \rightarrow \infty \text{ as } V_0 \rightarrow 0 \quad (1)$$

which has the solution:

$$\tilde{P}(T_0) = \frac{2g(T_0)}{\int_{T_0}^{T_1} g(T) dT} \quad \text{where} \quad g(T) = \exp\left(\int_T^{T_1} S(\tau) d\tau\right) \quad (2)$$

From this it follows that an allele with selective advantage  $S(T) = Z - T$ , present as a single copy at  $T=0$ , will survive until a time  $T_*$  with probability:

$$\tilde{P}(T_*) = 2\sqrt{\frac{2}{\pi}} \frac{\exp\left(-\frac{(T_* - Z)^2}{2}\right)}{\text{erf}\left(\frac{T_* - Z}{\sqrt{2}}\right) + \text{erf}\left(\frac{Z}{\sqrt{2}}\right)} \quad (3)$$

This approximation requires that the variance stay close to  $V_0$  over time  $T_*$ , i.e.,  $t_* \ll N$  (or equivalently,  $T_* \ll N\sqrt{V_0}$ ) and that  $T_* < Z$ , so that the focal allele has a positive growth rate throughout this initial phase. However,  $T_*$  must be large enough that fit alleles, having survived to this time, are very unlikely to be subsequently lost. As discussed above, both assumptions can be satisfied for some  $T_* \approx 1$  if  $N\sqrt{V_0} \gg 1$ .

We can obtain an even simpler approximation for the survival probability by neglecting the increase in the population mean, such that the allele has a *fixed* selective advantage  $Z$ . Then, the probability of survival to scaled time  $T_*$  is:

$$\tilde{P}(T_*) = \frac{2Z}{1 - \exp(-ZT_*)} \quad (4)$$

The survival probability in both eq. (3) and the simpler eq. (4) depends on the threshold  $T_*$ , which is arbitrary (though  $\mathcal{O}(1)$ , as argued above). In the following, we will explore the sensitivity of our predictions for ultimate response to the choice of  $T_*$ . Surprisingly, for  $N\sqrt{V_0} \gg 1$ , the expected ultimate change can be predicted quite accurately by using an even more drastic approximation— $\tilde{P} \approx 2Z$ , which is the asymptotic survival probability of an allele in the absence of adaptation in the bulk (see e.g., finite  $N$  predictions in fig. 1(a)).

**The distribution of ultimate selection response.** To predict the ultimate response, we assume that the fittest allele to survive until a threshold time  $T_*$  is the one to fix; the ultimate change is then the value of this fittest survivor. First, consider a given draw  $\{z\}$  of (unscaled) allelic values; we rank these such that  $z_1 > z_2 > \dots$ . If we approximate the initial dynamics of

very fit alleles in the tail of the distribution by independent branching processes (as described above), then each of these alleles can be associated with a survival probability  $P(z_1) > P(z_2) > \dots$ , which is given by eq. (3) (or more approximately by (4)). The probability that the fittest allele to survive up to time  $T_*$  has allelic value  $z_i$  or smaller, is then:  $\prod_{k=1}^{i-1} [1 - P(z_k)]$ , which is just the probability that all those with values larger than  $z_i$  are lost.

This formula gives the probability distribution of the winner from a given set of values. The fittest allele over a series of replicates will vary both because of the random initial values, and their subsequent random survival. Typically, the survival probability will be  $\sim 2z \sim 2\sqrt{V_0}$ , and so we expect that the winner will be ranked  $\sim 1/(2\sqrt{V_0})$ . For any given draw, the best that can be achieved is bounded by the fittest allele— but the actual outcome will almost always be fixation of a much lower ranked (less fit) allele.

We now find the cumulative probability density function (CDF)  $F(z)$  of the response, by taking the expectation of  $\prod_{k=1}^{i-1} [1 - P(z_k)]$  over all possible random draws of initial allelic values from a distribution  $\psi$ . For finite  $N$ , this involves first integrating over the joint distribution of  $\{z_1, z_2, \dots, z_{i-1}\}$ — the values of the top  $i-1$  ranked alleles from a set of  $N$  iid values— over the  $(i-1)$  dimensional region  $z < z_{i-1} < \dots < z_2 < z_1 < \infty$ , and then summing over all possible  $i$  (see Appendix). This gives the finite  $N$  predictions (colored dashed lines) in fig. 1(a).

However, in the scaling limit  $N \rightarrow \infty$ ,  $\sqrt{V_0} \rightarrow 0$  with  $N\sqrt{V_0}$  constant, there is a further simplification: the distribution of fitnesses of alleles, conditional on their having survived until time  $T_*$ , is  $(P\psi)/\bar{P}$  where  $\bar{P} = \int P\psi dz$  is the mean survival probability. The expected number of distinct alleles to survive the stochastic phase is thus  $N_* \approx N\bar{P}$ , and the probability that the fittest of these has value  $z$  or less is:

$$F(z) \approx \exp\left(-N_* \int_z^\infty \frac{P\psi}{\bar{P}} dw\right) \approx \exp\left(-N \int_z^\infty P\psi dw\right) \approx \exp\left(-N\sqrt{V_0} \int_{Z=z/\sqrt{V_0}}^\infty \tilde{P}\psi dw\right) \quad (5)$$

Since the probability of survival of an allele is proportional to its selection coefficient, we can write  $P = \sqrt{V_0} \tilde{P}$  (as in the last part of the above equation). Therefore, the key parameter is  $N\sqrt{V_0}$ , which under this approximation, only appears through the exponent in eq. (5). The probability that the top  $1, 2, \dots, k$  alleles to survive up till time  $T_*$  have allelic values  $z_1, z_2, \dots, z_k$ , can be recovered by differentiating  $\ln[F(z)]$ , and is given by  $[f(z_1) \dots f(z_k)]F(z_k)$  where  $f(z) = -\partial \ln[F(z)]/\partial z$ . Note that now  $z_1, z_2, \dots$  refer to the allelic values of the top survivors at time  $T_*$  rather than the top alleles present at  $T=0$ .

The theoretical predictions above depend on the unknown threshold  $T_*$ — figure 3 illustrates this dependence by comparing the CDF of the ultimate change obtained from 1000 simulation replicates of a population of size  $N=40000$  (red), with the predicted distribution of the fittest survivor at  $T_*=0.5, 1, 1.5$  (black, blue, purple) in the scaling ( $N \rightarrow \infty$ ) limit. Solid and dashed curves show predictions that account for or neglect competition between very fit survivors and the evolving bulk; these are obtained respectively by using eqs. (3) and (4), together with eq. (5). The three sets of curves in each panel correspond to  $N\sqrt{V_0}=10, 100, 1000$  (left to right); the two panels show response for Gaussian and Laplace initial distributions.

Choosing too low a value of  $T_*$  yields too high a prediction for the ultimate response since some of the alleles that have survived until a low  $T_*$  may not have established sufficient numbers and can be subsequently lost. Conversely, too high a value of  $T_*$  underestimates the aggregate selective boost enjoyed by very fit alleles during establishment (by allowing the bulk more time to catch up) and thus under-predicts ultimate response. In principle, an appropriate  $T_*$  can

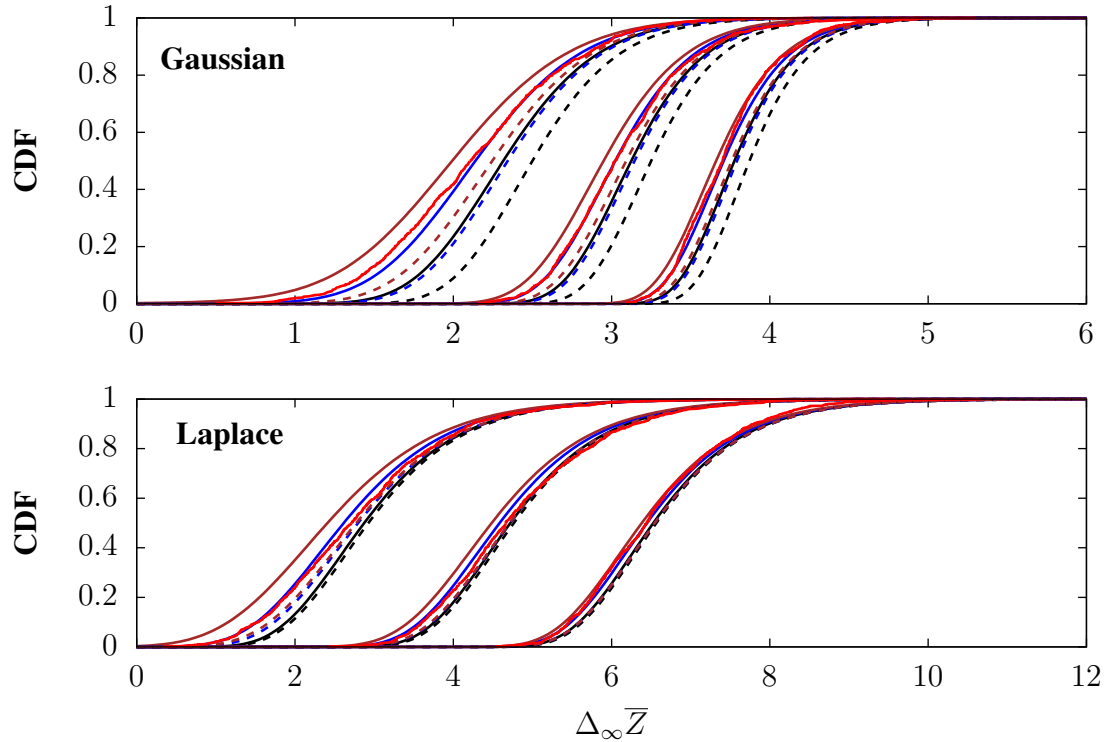


Figure 3: The cumulative distribution (CDF) of the ultimate response,  $\Delta_{\infty}\bar{Z}$ , compared with various predictions. In each panel, the three sets of curves correspond to  $N\sqrt{V_0}=10, 100, 1000$ ; the initial distribution is either Gaussian (top) or Laplace (bottom). The CDF of 1000 replicate simulations (each initialised with a random draw of allelic values), with  $N=40000$  (red) is compared with the CDF predicted by calculating the probability of survival to  $T_*=0.5, 1, 1, 5$  (black, blue, purple, solid lines right to left), from Eq. 3 (in conjunction with eq. 5). The dashed lines show the same, but using Eq. 4, which neglects the increasing mean of the bulk, and therefore predicts a somewhat higher response.

be estimated in a self-consistent way by assuming that it is the time at which the number of copies of the fittest surviving allele times its selective advantage crosses some threshold (such that the allele is almost certain to fix). The expected number of copies of an allele with fitness value  $z$ , conditional on its survival until (unscaled) time  $t_*$ , is  $\approx \exp\left[\int_{\tau=0}^{t_*} (z - \bar{z}(\tau)) d\tau\right] / P[z, t_*]$  and its selective advantage at time  $t_*$  is  $z - \bar{z}(t_*)$ , where  $\bar{z}(t_*) \approx V_0 t_*$  for  $t_* \ll N$ . Thus estimating  $T_* = \sqrt{V_0} t_*$  (given  $Z = z / \sqrt{V_0}$ ) amounts to finding the  $T_*$  at which  $\frac{Z - T_*}{P[Z, T_*]} \exp[ZT_* - T_*^2/2]$  crosses some threshold. If  $Z \gg T_*$ , we have  $\tilde{P}[Z, T_*] \approx 2Z$ , so that this criterion is approximately  $e^{ZT_*} \gg 1$ . Thus, we expect  $T_*$  to scale inversely with the value  $Z$  of the fittest surviving allele. Approximating this value by its expectation gives  $T_* \propto 1/\mathbb{E}[\Delta_{\infty}\bar{Z}]$ , where  $\mathbb{E}[\Delta_{\infty}\bar{Z}]$  depends on  $N\sqrt{V_0}$  and  $T_*$ .

This kind of argument can, in principle, allow us to estimate  $T_*$  self-consistently given  $N\sqrt{V_0}$ . However, since  $\mathbb{E}[\Delta_{\infty}\bar{Z}]$  increases only weakly with  $N\sqrt{V_0}$  (fig. 1), in practice, predictions obtained using a fixed  $T_*$  are quite accurate across a range of  $N\sqrt{V_0}$  values. For example,  $T_* \sim 1$  appears to fit well for the Gaussian initial distribution in fig. 3, while somewhat lower  $T_*$  values (closer to 0.5) provide a better match with the Laplace distribution; this is consistent with the observation that expected response is higher by a factor of about 1.5–1.7 in the latter case. Note that the choice of  $T_*$ , and more generally, accounting for adaptation in the bulk (solid vs. dashed curves) makes less difference as  $N\sqrt{V_0}$  increases.



Gaussian						
$N\sqrt{V_0}$	Mean $\Delta_\infty \bar{Z}$ (sim)	Mean $\Delta_\infty \bar{Z}$ (pred)	$\sqrt{2 \log(1.15 N\sqrt{V_0})}$	sd (sim)	sd (pred)	
10	2.140	2.1729	2.2106	0.5791	0.5448	
100	3.078	3.0671	3.0809	0.4222	0.3982	
1000	3.731	3.7501	3.7546	0.3221	0.3303	
Laplace						
$N\sqrt{V_0}$	Mean $\Delta_\infty \bar{Z}$ (sim)	Mean $\Delta_\infty \bar{Z}$ (pred)	$\frac{1}{\sqrt{2}} \left( \mathcal{W} \left( \frac{0.36}{N\sqrt{V_0}} \right) - 1 \right)$	sd (sim)	sd (pred)	
10	2.84	2.95604	2.7682	1.1530	1.0547	
100	4.84	4.83621	4.7102	1.0755	1.0212	
1000	6.55	6.6603	6.5446	0.9328	0.9977	

Table 1: Mean and standard deviation of the ultimate (scaled) change as obtained from the simulations in fig. 3, compared with predictions obtained using eqs. (5) and (3) assuming  $T_*=1$  for Gaussian and  $T_*=0.625$  for Laplace initial distributions. The simple predictions, based on setting  $N\sqrt{V_0} \int_Z^\infty \tilde{P}\psi dw = \log[2]$ , are also shown in column 4.

**A simple approximation for the expected ultimate response.** We now find simple approximations for the expected response for both Gaussian and Laplace initial distributions. This equals  $\mathbb{E}(\Delta_\infty \bar{Z}) = \int_0^\infty Z \left( \frac{\partial F}{\partial Z} \right) dZ$  (assuming that the proportion of surviving alleles with  $Z < 0$  is negligible); integrating by parts gives  $\mathbb{E}(\Delta_\infty \bar{Z}) = \int_0^\infty [1 - F(Z)] dZ$ . When  $N\sqrt{V_0}$  is large, the probability that the fittest surviving allele has value less than  $Z$  is close to zero for small  $Z$ , and increases sharply to 1 at some  $Z$  (fig. 3); thus,  $1 - F$  also changes sharply from 1 to 0 at this  $Z$ , and  $\mathbb{E}(\Delta_\infty \bar{Z})$  can be approximated as the value of  $Z$  for which  $F(Z) = \exp\left(-N\sqrt{V_0} \int_Z^\infty \tilde{P}\psi dw\right) = \frac{1}{2}$ , or  $\int_Z^\infty \tilde{P}\psi dw = \frac{\log(2)}{N\sqrt{V_0}}$ . This assumes that the distribution of ultimate response is not too skewed, which appears to be a reasonable assumption, at least for Gaussian or Laplace initial distributions (fig. 3).

If we now use the crude approximation  $P(z) = 2z$  or  $\tilde{P} = \frac{P}{\sqrt{V_0}} = 2Z$ , where  $Z = z/\sqrt{V_0}$ , then for a Gaussian initial distribution, we have:  $\int_Z^\infty \psi \tilde{P} dw = \int_Z^\infty \frac{2w e^{-\frac{w^2}{2}}}{\sqrt{2\pi}} dw = \frac{2e^{-\frac{Z^2}{2}}}{\sqrt{2\pi}}$ . Setting this equal to  $\frac{\log(2)}{N\sqrt{V_0}}$  gives:  $\mathbb{E}(\Delta_\infty \bar{Z}) \approx \sqrt{2 \log\left(\frac{2N\sqrt{V_0}}{\sqrt{2\pi} \log(2)}\right)} \approx \sqrt{2 \log(1.15 N\sqrt{V_0})}$ . For a Laplace distribution, the same argument leads to  $\frac{\log(2)}{N\sqrt{V_0}} = \left(Z + \frac{1}{\sqrt{2}}\right) e^{-\sqrt{2}Z}$ ; this gives  $\mathbb{E}(\Delta_\infty \bar{Z}) = \frac{1}{\sqrt{2}} \left( \mathcal{W} \left( \frac{\sqrt{2} \log(2)}{eN\sqrt{V_0}} \right) - 1 \right) = \frac{1}{\sqrt{2}} \left( \mathcal{W} \left( \frac{0.36}{N\sqrt{V_0}} \right) - 1 \right)$ , where  $\mathcal{W}[\alpha]$  is the product log function, which is the solution to  $\alpha = \mathcal{W} e^{-\mathcal{W}}$ . Table 1 shows that these approximations are accurate for both Gaussian and Laplace initial distributions.

Thus, essentially, the mean response to selection is given by finding the expected value of the fittest of  $N\bar{P}$  alleles, where  $\bar{P}$  is the probability of surviving to a time  $t_* \sim 1/\sqrt{V_0}$  generations. Surprisingly, this probability can be approximated by  $\approx 2\bar{z}$ , where  $\bar{z}$  is the mean selective advantage of fitter than average alleles (i.e., alleles with  $z > 0$  in case of a symmetric initial distribution) present at  $t=0$ , even though this neglects competition from the increasingly well-adapted bulk of the population.

## Dynamics of the fitness distribution.

So far, we have focused on the initial stochastic phase which determines which allele will ultimately fix. However, the dynamics of this process are more complex and differ qualitatively in

the weak vs. strong selection regimes. Figure 4 shows the variance of log fitness values as a function of scaled time, averaged over 1000 simulation replicates, for various values of  $N\sqrt{V_0}$  with  $N=10^4$ . Time is scaled either as  $T=t\sqrt{V_0}$  (fig. 4(a)) or  $\tau=t/N$  (fig. 4(b)); the two plots are drawn on a log-log and semi-log scale respectively to illustrate the different dynamics of variance decay in the two parameter regimes. Main plots and insets show variance dynamics for Gaussian and Laplace initial distributions respectively.

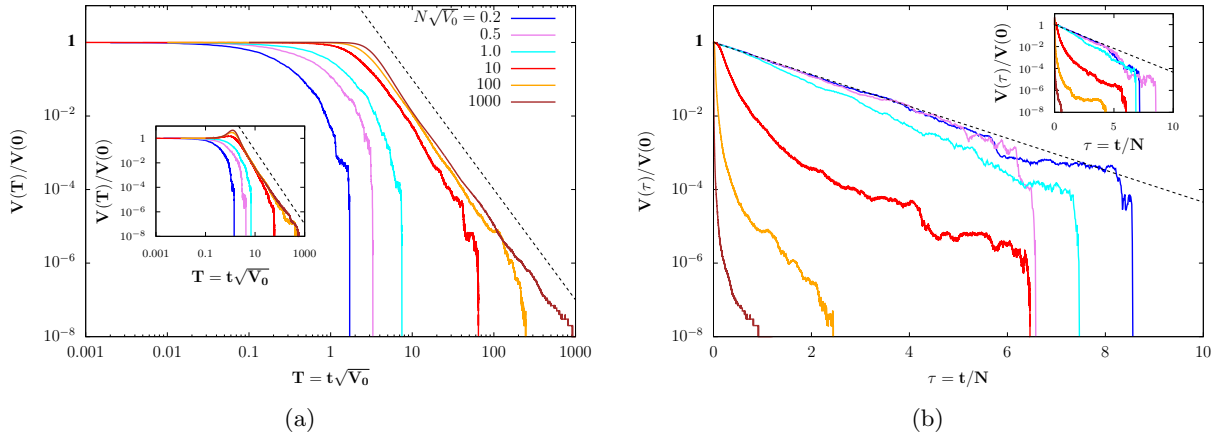


Figure 4: Expected variance (scaled by  $V_0$ ) as a function of (a)  $T=t\sqrt{V_0}$  or (b)  $\tau=t/N$  for various values of  $N\sqrt{V_0}$  (various colors) obtained by averaging over 1000 simulation replicates with  $N=10^4$ . The dashed line in (a) depicts  $V(T)/V(0) = 100/T^3$  while the dashed line in (b) depicts  $V(\tau)/V(0) = e^{-\tau}$ . For  $N\sqrt{V_0} \gg 1$ , the variance decays as  $T^{-3}$  independent of  $N\sqrt{V_0}$ , while for  $N\sqrt{V_0} \ll 1$ , it follows the neutral expectation  $e^{-\tau}$ .

In the weak selection ( $N\sqrt{V_0} \ll 1$ ) regime, the expected variance decays as  $V(t) \approx V_0 e^{-t/N}$  (depicted via a dashed black line in fig. 4(b)), more or less independently of the value of  $N\sqrt{V_0}$  (as evident in the overlapping blue and violet curves in fig. 4(b)). This is consistent with the intuition that in this regime, the increase in frequencies of fitter than average alleles due to selection approximately balances out the decrease in frequencies of less fit alleles, so that selection has no effect on the shape of the fitness distribution (on average), and the decay in fitness variance is governed essentially by drift.

By contrast, in the strong selection ( $N\sqrt{V_0} \gg 1$ ) regime, the expected variance remains constant or even slightly increases (in case of the Laplace initial distribution) over a timescale  $T = t\sqrt{V_0} \approx 1$ , but subsequently decays as a power law  $\approx T^{-3}$  (depicted by a dashed line in fig. 4(a)). The observation of power law decay appears surprising at first glance, since in this regime, we expect long-term dynamics to be governed by the competition between the fittest few alleles to have survived the initial stochastic phase. For example, if two alleles with similar fitnesses survive, then it will take a long time for the fitter of the two to fix; assuming deterministic dynamics, variance should then decay *exponentially* at a rate proportional to the fitness difference between the two alleles.

More precisely, let  $z_1 > z_2 > \dots$  denote the log fitness values and  $n_{1,0}, n_{2,0}, \dots$  the number of copies of surviving alleles at some initial time  $t_*$ . If dynamics are essentially deterministic for  $t > t_*$ , then we expect alleles to grow exponentially in *relative* abundance, so that the number of copies of allele  $i$  at time  $t$  is  $n_{i,t} = \frac{n_{i,0} e^{(t-t_*)z_i}}{\sum_j n_{0,j} e^{(t-t_*)z_j}}$ . We can use this relation to predict how the fitness distribution evolves in time, neglecting stochastic fluctuations. In particular, at very long times, when only the fittest two alleles survive, we have:  $\bar{z}(t) \approx \frac{z_1 + \alpha z_2 e^{-(z_1-z_2)(t-t_*)}}{1 + \alpha e^{-(z_1-z_2)t}}$  and  $V(t) \approx \frac{\alpha(z_1-z_2)^2 e^{-(z_1-z_2)(t-t_*)}}{[1 + \alpha e^{-(z_1-z_2)(t-t_*)}]^2}$ , where  $\alpha = \frac{n_{2,0}}{n_{1,0}}$  is the ratio of the numbers (at time  $t_*$ ) of the two

fittest surviving alleles, and is itself a random variate with expectation  $\frac{P(z_1)}{P(z_2)} e^{-(z_1-z_2)t^*}$ , which can be further approximated by  $\frac{z_1}{z_2} e^{-(z_1-z_2)t^*}$  (assuming  $P(z) \approx 2z$ ).

As one would expect based on this simple argument, fitness variance does decay nearly exponentially in individual simulation replicates at long times (and while the less fit allele is not too rare), with a rate of decay that depends on  $z_1-z_2$ . The dynamics of the expected variance (averaged over replicates) can thus be predicted from the joint distribution  $G(z, z_1)$  of the ‘gap’  $z=z_1-z_2$  between the top two surviving alleles and the value  $z_1$  of the fittest survivor as:  $\mathbb{E}[V(t)] \approx \int \int z^2 \frac{z_1}{z_1-z} \frac{e^{-zt}}{(1+\frac{z_1}{z_1-z}e^{-zt})^2} G(z, z_1) dz dz_1$ . At long enough times, regardless of the exact form of  $G(z, z_1)$  (which will depend on  $N\sqrt{V_0}$  and the initial distribution  $\psi$ ; see above), the main contribution to the expected variance comes from the  $z \rightarrow 0$  portion of the gap distribution. In other words, as we go to larger and larger times, only replicates with smaller and smaller fitness gaps between the top two surviving alleles would still have segregating variants and only these would thus contribute to the expectation over replicates as  $t \rightarrow \infty$ . Thus, we can approximate the above expectation using Laplace’s method as:

$$\begin{aligned} \mathbb{E}[V(t)] &\approx \int \int z^2 \frac{e^{-zt}}{(1+e^{-zt})^2} \left[ G(0, z_1) + z \left( \partial_z G(0, z_1) + \tanh(zt/2) \frac{G(0, z_1)}{z_1} \right) + \dots \right] dz dz_1 \\ &= \frac{1}{t^3} \left( \int_0^\infty x^2 \frac{e^{-x}}{(1+e^{-x})^2} dx \right) \left( \int_0^\infty G(0, z_1) dz_1 \right) + \mathcal{O}\left(\frac{1}{t^4}\right) \end{aligned} \quad (6)$$

## Discussion

We consider here the limits to adaptation from standing variation in a finite asexual population, allowing for an arbitrary initial distribution of fitness values across individuals. In the absence of new mutations, the maximum possible advance under selection is limited by the fittest of  $N$  alleles present at  $t=0$ , whose fitness is (on average)  $\sim \log(N)$  and  $\sim \sqrt{\log(N)}$  standard deviations respectively for the Laplace and Gaussian distributions considered above. However, the typical probability of survival is  $\sim 2\sqrt{V_0}$ , so that the actual response is governed by the fittest of  $2N\sqrt{V_0}$  surviving alleles, and is thus typically much lower than the theoretical maximum. Our derivations are an elaboration on this simple argument.

The qualitatively different—nearly neutral ( $N\sqrt{V_0} \ll 1$ ) vs. strong selection ( $N\sqrt{V_0} \gg 1$ )—regimes of response from standing variation in asexual populations that we identify in this study are analogous to those observed for small and large  $N$  in models of long-term adaptation under steady mutation, see e.g., figure 1 in Melissa et al (2022). In particular, the fact that starting with a Gaussian initial distribution, the ultimate gain in log fitness is  $\approx \sqrt{2 \log(1.15N\sqrt{V_0})}$  for  $N\sqrt{V_0} \gg 1$  under our model, is closely related to the observation that in a population described by a fitness wave with variance  $\sigma$ , the common ancestors of future individuals are located in the leading edge of the wave, approximately  $\sqrt{2 \log(N\sigma)}$  standard deviations ahead of the mean, for  $N\sigma \gg 1$  (Desai and Fisher 2007).

In this  $N\sigma \gg 1$  regime, genealogies of adapting populations are no longer described by the Kingman coalescent or its variations (e.g., with time-varying population size), but instead are strongly skewed, resulting in an excess of high-frequency derived alleles and a ‘U-shaped’ site frequency spectrum (Neher and Hallatschek 2013), as in multiple merger coalescents (Bolthausen and Sznitman 1998; Sargsyan and Wakeley 2008). One might ask how this compares with genealogies of populations adapting from standing genetic variation (as in this study) in the  $N\sqrt{V_0} \gg 1$  regime. In the absence of mutational input (of new alleles), genealogical structure at any timepoint during adaptation will be determined by the numbers of alleles surviving up

until that point, traced back through time; these numbers, in turn, depend on the fitness gaps between surviving alleles. Thus, genealogical structure itself can change considerably during the course of adaptation; characterising these changes and the associated changes in neutral diversity along the genome remains an interesting direction for future work.

What bearing do our results have on understanding selection response from standing genetic variation in *sexual* populations? Consider an extreme scenario where variation is uniformly distributed across a linear genome, with fitness variance per unit map length  $v_g$ . Under this infinitesimal model (with linkage), the fitness variance associated with a genomic region of map length  $R$  is just  $v_g R$  on average. One can now ask: what is the typical map length  $R_*$  of ‘effectively asexual’ fragments that fix (under selection) without being split up by recombination in a population of effective size  $N_e$ ? If fragments evolve nearly neutrally, i.e., if  $N_e \sqrt{v_g R_*} \ll 1$ , then the average time to fixation of any fragment is  $\approx N_e$ , implying that the typical map length  $R_*$  of fragments that fix without recombining is  $R_* \sim 1/N_e$ . Thus, a self-consistent criterion for effectively asexual blocks to evolve nearly neutrally is that  $N_e \sqrt{v_g R_*} \sim \sqrt{N_e v_g} \ll 1$ , or  $N_e v_g \ll 1$ .

One can also consider the opposite ‘strong selection’ regime— where fitness variance per unit map length  $v_g$  is sufficiently high that effectively asexual fragments can, in principle, generate sweep-like signatures as they increase under selection (Sachdeva and Barton 2018). As before, the fitness variance associated with a genomic region of map length  $R_*$  is  $v_g R_*$ ; thus, under strong selection, i.e., if  $N_e \sqrt{v_g R_*} \gg 1$ , the time scale for the establishment of a highly fit haplotype spanning a map length  $R_*$  is approximately  $\tau \sim 1/\sqrt{v_g R_*}$  (in analogy with the asexual model, where the fittest surviving allele establishes over a timescale  $\tau \sim 1/\sqrt{V_0}$ ). This argument assumes that fragments of map length  $R_*$  are effectively asexual over a timescale  $\tau$ , which requires  $R_* \tau \sim 1$  or  $R_* \sim v_g$ . Thus, this gives a complementary criterion—  $N_e \sqrt{v_g R_*} \sim N_e v_g \gg 1$  for the emergence of strongly selected effectively asexual blocks at short timescales, i.e., during *initial* response.

It seems unlikely, however, that long-term response in sexual populations (for  $N_e v_g \gg 1$ ) can be understood by extrapolating from the asexual model. In particular, even if a few fit haplotypes of map length  $\sim v_g$  increase intact early on, there may still be fine-scale recombination between these at longer timescales, generating new combinations of alleles, allowing for sustained response to selection in large populations. Such fine-scale recombination is expected to be especially efficient if the haplotypes that establish early on have very similar fitness, so that one of these does not rapidly fix. An interesting question is whether a long phase of competition between a few fit haplotypes (which can potentially generate new and even fitter combinations via recombination) might produce patterns in neutral diversity that mimic soft sweeps (Hermisson and Pennings 2005).

The above arguments all implicitly assume that a linear genome can be approximated as a collection of *independently* evolving asexual blocks, whereas in reality, blocks may interfere with one another as they fix. Such selective interference would be rather weak if the fitness variance per unit map length  $v_g$  is low, so that its effects on the dynamics of individual blocks may be captured by an appropriately defined effective population size (Robertson 1961; Santiago and Caballero 1998). However, it is less obvious whether this should hold for large  $v_g$ , i.e., if strongly selected blocks are segregating at multiple genomic locations.

Such a caricature of the recombining genome as a mosaic of quasi-independent asexual blocks is nevertheless a useful starting point, providing at least a rough criterion for when we might distinguish selection response in sexual populations— either under a steady supply of new mutations (Neher et al 2013; Weissman and Hallatschek 2014) or from standing genetic variation (as discussed above)— from neutral evolution. These arguments all furnish self-consistently derived predictions for the characteristic map length of effectively asexual blocks (or alternatively, the typical scale over which there is linkage disequilibrium or LD) in the adapting

population. This, however, obscures the fact that in reality, haplotypes that increase under selection may vary widely in length, giving rise to very heterogeneous signatures of selection along the genome. Such heterogeneity can be due to various reasons— first, variation is not uniformly spread across the genome and may at least be partly due to loci of major effect. Second, idiosyncratic patterns of LD, either in large populations or created by sampling a small number of individuals (e.g., in Evolve and Resequence experiments), may also result in a heterogeneous distribution of additive genetic variance across the genome. Importantly, this can generate considerable heterogeneity in short-term selection response along a linear genome even under the infinitesimal model (Castro et al 2019). Finally, even if the initial distribution of variance is perfectly uniform, the inherent stochasticity of evolutionary processes would still cause blocks of rather different lengths to increase during adaptation: with low initial variance ( $Nv_g \ll 1$ ), the resultant haplotype structure may be close to that under neutral evolution. A natural question then is whether heuristics based on asexual linkage blocks may be useful in characterizing the heterogeneity of response along the genome for  $Nv_g \gg 1$ , i.e., under conditions where adaptation from standing variation produces detectable signatures in sequence data.

## Appendix: Ultimate change under selection for finite $N$

Here, we calculate the distribution of the ultimate change under selection for  $N\sqrt{V_0} \gg 1$  for finite  $N$ . As in the main paper, we assume that the fittest allele to escape initial stochastic loss and survive is the one to fix. Then, given a set of ordered and scaled allelic values  $Z_1 > Z_2 > \dots$  at  $t=0$ , and under the branching process approximation, the probability that the fittest allele to survive has rank  $i$  is just:

$$g_i = g(Z_1, Z_2, \dots, Z_i) = P(Z_i) \prod_{k=1}^{i-1} [1 - P(Z_k)] \approx \sqrt{V_0} \tilde{P}(Z_i) \exp \left[ -\sqrt{V_0} \sum_{k=1}^{i-1} \tilde{P}(Z_k) \right] \quad (7)$$

where  $\tilde{P}(Z_k)$  is the scaled survival probability of an allele with scaled value  $Z_k$ ; further,  $\sqrt{V_0}$  is assumed to be sufficiently small that  $1 - \sqrt{V_0} \tilde{P} \approx e^{-\sqrt{V_0} \tilde{P}}$ . In the following, we will use the crude approximation  $\tilde{P}(Z) \approx 2Z$  in explicit computations, but our general expression for the distribution of ultimate response (eq. (10)) does not depend on this.

The distribution of the ultimate change can now be obtained by integrating  $g_i$  over the joint distribution  $\mathbb{Q}_N(Z_1, Z_2, \dots, Z_i)$  of the top  $i$  alleles present at  $t=0$ , and then summing over all possible values of the rank  $i$  of the fittest surviving allele. More precisely, the probability density  $\mathcal{M}(Z)$  of the (scaled) value  $Z$  of the fittest allele to survive the initial stochastic phase can be expressed as:

$$\mathcal{M}(Z) = \sum_i \int d\vec{Z}_{\{1, i-1\}} g_i(Z_1, Z_2, \dots, Z_i=Z) \mathbb{Q}_N(Z_1, Z_2, \dots, Z_i=Z) \quad (8)$$

where  $\int d\vec{Z}_{\{1, i-1\}}$  represents the  $(i-1)$  dimensional integral over the region  $Z < Z_{i-1} < Z_{i-2} < \dots < Z_2 < Z_1 < \infty$ . Note that here we work with the probability density  $\mathcal{M}(Z)$ , rather than the cumulative density  $F(Z)$  used in the main paper.

For an initial distribution with density  $\psi(Z)$  and cumulative density function  $\Psi(Z) = \int_{-\infty}^Z \psi(y) dy$ , the joint distribution of the top  $i$  values among a sample of  $N$  iid values can be written as:

$$\mathbb{Q}_N(Z_1, Z_2, \dots, Z_i) = \frac{N!}{(N-i)!} \left( \prod_{k=1}^i \psi(Z_k) \right) [\Psi(Z_i)]^{N-i} \quad (9)$$

Combining equations (7),(8), and (9) yields:

$$\mathcal{M}(Z) = \sqrt{V_0} \psi(Z) \tilde{P}(Z) \sum_i \frac{N!}{(N-i)!} [\Psi(Z)]^{N-i} h_{i-1}(Z)$$

where  $h_i(Z) = \int_Z^\infty dZ_i \psi(Z_i) e^{-\sqrt{V_0} \tilde{P}(Z_i)} \dots \int_{Z_3}^\infty dZ_2 \psi(Z_2) e^{-\sqrt{V_0} \tilde{P}(Z_2)} \int_{Z_2}^\infty dZ_1 \psi(Z_1) e^{-\sqrt{V_0} \tilde{P}(Z_1)}$

(10)

which is our main result for the distribution of ultimate change for finite  $N$ . The function  $h_i(Z)$  is an  $i$ -dimensional integral; we can obtain explicit expressions for specific initial distributions by approximating the survival probability as  $\tilde{P}(Z) \approx 2Z$  for  $Z > 0$  and 0 otherwise. Then for Gaussian and Laplace initial distributions, we have:

$$h_i(Z) = \frac{1}{i!} \left[ \frac{e^{2V_0}}{2} \operatorname{erfc} \left( \frac{2\sqrt{V_0} + Z}{\sqrt{2}} \right) \right]^i \quad (\text{Gaussian})$$

$$h_i(Z) = \frac{1}{i!} \left[ \frac{e^{-(\sqrt{2} + 2\sqrt{V_0})Z}}{2(1 + \sqrt{2V_0})} \right]^i \quad (\text{Laplace})$$
(11)

Now summing over  $i$ , we obtain the following approximation for (unnormalised)  $\mathcal{M}(Z)$ :

$$\mathcal{M}(Z) \approx Z e^{-Z^2/2} \left( \operatorname{erf} \left( \frac{Z}{\sqrt{2}} \right) + e^{2V_0} \operatorname{erfc} \left( \frac{2\sqrt{V_0} + Z}{\sqrt{2}} \right) + 1 \right)^{N-1} \quad (\text{Gaussian})$$

$$\mathcal{M}(Z) \approx \frac{Z e^{-\sqrt{2}Z} \left( 2 - e^{-\sqrt{2}Z} + \frac{e^{-(2\sqrt{V_0} + \sqrt{2})Z}}{\sqrt{2V_0} + 1} \right)^N}{(\sqrt{2V_0} + 1) \left( 2 - e^{-\sqrt{2}Z} + e^{-(2\sqrt{V_0} + \sqrt{2})Z} \right)}$$
(12)

if  $Z > 0$ , and 0 otherwise. In the limit  $V_0 \rightarrow 0$ ,  $N \rightarrow \infty$ , with  $N\sqrt{V_0}$  constant, the above equations simplify to the scaling limit predictions (eq. (5) in the main text).

## References

- Bérard, J., Gouéré, J.-B., 2010. Brunet-Derrida behavior of branching-selection particle systems on the line. *Comm. Math. Phys.* 298 (2), 323-342.
- Berestycki, J., Berestycki, N., Schweinsberg, J., 2013. The genealogy of branching Brownian motion with absorption. *Ann. Probab.* 41 (2), 527-618.
- Bolthausen, E., Sznitman, A.S., 1998. On ruelle's probability cascades and an abstract cavity method. *Communications in Mathematical Physics* 197: 247-276.
- Brunet, E., Derrida, B., Mueller, A.H., Munier, S., 2007. Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2007;76(4 Pt 1): 041104.
- J. P. Castro et al., An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice. *eLife* 8, e42014 (2019).
- Cortines, A., 2016. The genealogy of a solvable population model under selection with dynamics related to directed polymers. *Bernoulli* 22 (4), 2209-2236.

- Cortines, A., Mallein, B., 2017. A N-branching random walk with random selection. *ALEA : Latin American Journal of Probability and Mathematical Statistics*, 2017, 14 (1), pp.117–137.
- Desai, M.M., Fisher, D.S., 2007. Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics*. 2007;176(3): 1759-1798.
- Fisher, D.S., 2013. Asexual evolution waves: fluctuations and universality. *J Stat Mech*. 2013(1):P01011.
- Fisher, R.A., 1930. *The genetical theory of natural selection*. Oxford: Oxford University Press.
- Good, B.H., Rouzine, I.M., Balick, D.J., Hallatschek, O., Desai, M.M., 2012. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc Natl Acad Sci U S A.*; 109(13):4950-4955.
- Haigh, J., 1978. The accumulation of deleterious genes in a population *Theor. Popul. Biol.*, 14, pp. 251-267.
- Hermisson, J., Pennings, P.S., 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169, 2335-2352.
- Melissa, M.J., Good, B.H., Fisher, D.S., Desai, M.M., 2022. Population genetics of polymorphism and divergence in rapidly evolving populations. *Genetics*. 2022;221(4)
- Neher, R.A., Hallatschek, O., 2013. Genealogies in rapidly adapting populations. *Proc Nat Acad Sci* 110: 437-442.
- Neher, R.A., Kessinger, T.A., Shraiman, B.I., 2013. Coalescence and genetic diversity in sexual populations under selection. *Proc Natl Acad Sci U S A.* 110(39):15836-15841.
- Novella, I.S., Duarte, E.A., Elena, S.F., Moya, A., Domingo, E., et al., 1995. Exponential increases of RNA virus fitness during large population transmissions. *Proc. Natl. Acad. Sci. USA* 92: 5841-5844.
- Paixão, T., Barton, N. H., 2016. The effect of gene interactions on the long-term response to selection. *Proc. Natl. Acad. Sci. U.S.A.* 113, 4422-4427.
- Robertson, A., 1960. A theory of limits in artificial selection. *Proc. R. Soc. B*, 153, 234-249.
- Robertson, A., 1961. Inbreeding in artificial selection programmes. *Genet. Res.* 2: 189-194.
- Rouzine, I.M., Wakeley, J., Coffin, J.M., 2003. The solitary wave of asexual evolution. *Proc Natl Acad Sci U S A.* 100(2):587-592.
- Rouzine, I.M., Brunet, E., Wilke, C.O., 2008. The traveling-wave approach to asexual evolution: Muller’s ratchet and speed of adaptation. *Theor Popul Biol.* 73(1):24-46.
- Sachdeva, H., Barton, N.H., 2018. Introgression of a block of genome under infinitesimal selection. *Genetics* 209, 1279-1303.
- Santiago, E., Caballero, A., 1998. Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics* 149, 2105-2117.
- Sargsyan, O., Wakeley, J., 2008. A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. *Theoretical Population Biology* 74: 104-114.

Tsimring, L.S., Levine, H., Kessler, D.A., 1996. RNA virus evolution via a fitness-space model. *Phys Rev Lett.* 76(23):4440-4443.

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