Polygenic adaptation dynamics in large, finite populations

Archana Devi[1] and Kavita Jain[2,†]

[1]Biodesign Center for Mechanisms of Evolution,

Arizona State University, Tempe, AZ 85287, USA

[2]Theoretical Sciences Unit,

Jawaharlal Nehru Centre for Advanced Scientific Research,

Bangalore 560064, India

† Corresponding author

January 26, 2023

Abstract

Although many phenotypic traits are determined by a large number of genetic variants, how a polygenic trait adapts in response to a change in the environment is not completely understood. In the framework of diffusion theory, we study the steady state and the adaptation dynamics of a large but finite population evolving under stabilizing selection and symmetric mutations when selection and mutation are moderately large. We find that in the stationary state, the allele frequency distribution at a locus is unimodal if its effect size is below a threshold effect and bimodal otherwise; these results are the stochastic analog of the deterministic ones where the stable allele frequency becomes bistable when the effect size exceeds a threshold. It is known that following a sudden shift in the phenotypic optimum, in an infinitely large population, selective sweeps at a large-effect locus are prevented and adaptation proceeds exclusively via subtle changes in the allele frequency; in contrast, we find that the chance of sweep is substantially enhanced in large, finite populations and the allele frequency at a large-effect locus can reach a high frequency at short times even for small shifts in the phenotypic optimum.

1 Introduction

Understanding the genetic basis of phenotypic variation and the phenotypic adaptation dynamics are central and challenging questions in evolution, and have attracted considerable interest, especially in the last two decades (Barton and Keightley, 2002; Rockman, 2012; Jain and Stephan, 2017a; Sella and Barton, 2019; Barghi *et al.*, 2020). The phenotypic variation depends on the genetic architecture of a trait which refers to the number of genetic variants underlying a phenotype, the effect and frequency of these variants on the phenotype, the interaction between these genetic variants and with the environment, etc. (Timpson *et al.*, 2018).

In monogenic and oligogenic traits such as insecticide resistance in Drosophila (Ffrench-Constant *et al.*, 2002), industrial melanism in peppered moth (van't Hof *et al.*, 2011) and lactase persistence in humans (Ségurel and Bon, 2017), one to few loci are associated with the phenotype, and adaptation in such traits is driven by large and rapid changes in the frequency of the selected allele (Maynard Smith and Haigh, 1974; Hermisson and Pennings, 2017). On the other hand, from genome-wide association studies (GWAS) (Visscher *et al.*, 2017), it has been established that a large number of loci are required to explain the heritability or within-population genetic variance of many traits including human height (Shi *et al.*, 2016; Yengo *et al.*, 2022) and several complex diseases (Loh *et al.*, 2015). For such polygenic traits, in response to a change in the en-

vironment, the change in the allele frequencies is correlated to each other.

Although many phenotypic traits have a polygenic basis (Boyle *et al.*, 2017; Wray *et al.*, 2018), the adaptation dynamics of such traits are not well understood (Pritchard and Di Rienzo, 2010; Pritchard *et al.*, 2010), and both empirical and theoretical routes have been taken to make progress. In recent studies, polygenic adaptation has been investigated experimentally in Drosophila (Burke *et al.*, 2010; Turner *et al.*, 2011; Barghi *et al.*, 2019), in natural populations (Therkildsen *et al.*, 2019) and in the framework of quite general theoretical models. Specifically, theoretical studies deal with models that allow one to explore various genetic architectures underlying a quantitative trait and have focused on understanding the allele frequency dynamics and their connection to phenotypic dynamics, assuming infinite population size (Lande, 1983; Chevin and Hospital, 2008; de Vladar and Barton, 2014; Jain and Stephan, 2015, 2017b), and more recently, finite populations (Höllinger *et al.*, 2019; Hayward and Sella, 2022). In an infinite population under stabilizing selection, it has been shown that the selective sweeps at large-effect loci are prevented (Chevin and Hospital, 2008; Jain and Stephan, 2017b), and in small populations where the dynamics are drift-dominated, the large-effect alleles are almost never found to sweep to fixation (Sella and Barton, 2019; Hayward and Sella, 2022).

This leaves the question of polygenic adaptation dynamics in moderately large populations; to fill this gap in our understanding, here we revisit the deterministic model studied in de Vladar and Barton (2014); Jain and Stephan (2015, 2017b) but now assuming that the population size is finite to address how random genetic drift affects the chance of sweep. We find that as in previous studies, the initial frequency of a large-effect locus plays an important role in determining if its allele frequency can reach a high frequency at short times. Naïvely, one expects that in a large population, the allele frequency follows a Gaussian-distribution centred about the deterministic result. However, using the exact stationary state distribution (Wright, 1937; Kimura, 1964, 1965; de Vladar and Barton, 2011), we find that the tail of the initial distribution is fatter than that of a Gaussian distribution and leads to an enhanced probability of a large change in the allele frequency of a large-effect locus when selection and mutation are moderately strong.

2 Model

We consider a polygenic trait controlled by ℓ diallelic loci in a panmictic, finite, diploid population of size N in linkage equilibrium. The phenotype-genotype map is assumed to be additive, and the trait value z receives contributions from $+$ and $-$ allele at the ith locus whose effect size are $\pm\gamma_i/2$ and are present in frequency x_i and $1-x_i$, respectively, in the population. Then the trait mean (averaged over

the population) in a single stochastic trajectory can be written as,

$$z = \sum_{i=1}^{\ell} \gamma_i (2x_i - 1) \tag{1}$$

On averaging over independent stochastic trajectories, we obtain $\langle z \rangle = \sum_{i=1}^{\ell} \gamma_i (2\langle x_i \rangle - 1)$. We also assume that the mutation between + and − allele occurs at an equal rate μ at each locus. The phenotypic trait evolves under stabilizing selection (Robertson, 1956; Kingsolver *et al.*, 2001; Sanjaka *et al.*, 2018; de Villemereuil *et al.*, 2020; Sodeland *et al.*, 2022) with the phenotypic fitness $w(z) \approx 1-(s/2)(z-z_{\rm opt})^2$ which falls quadratically from a (time-independent) optimal trait value $z_{\rm opt}$ with the strength of the selection being s (here, s^{-1} denotes the sum of genetic and environmental contributions to the phenotype). In much of the discussion, we assume that the effect sizes are chosen from an exponential distribution with mean $\bar{\gamma}$ (Orr, 1998; Mackay, 2004; Goddard and Hayes, 2009).

Under these assumptions, the evolution of the joint distribution of the allele frequency vector, $\vec{x} = \{x_1, ..., x_\ell\}$ can be expressed by the following Fokker-Planck equation (Kimura, 1964; Ewens, 2004),

$$\frac{\partial P(\vec{x},t)}{\partial t} = \sum_{i=1}^{\ell} \Big[-\frac{\partial}{\partial x_i}\big(M_i(\vec{x})P(\vec{x},t)\big) + \frac{1}{4N}\frac{\partial^2}{\partial x_i^2}\big(V_i(\vec{x})P(\vec{x},t)\big)\Big] \tag{2}$$

with

$$
\begin{aligned}
M_i &= \frac{x_i(1-x_i)}{2\bar{w}}\frac{\partial \bar{w}}{\partial x_i} + \mu(1-2x_i) && (3)\\
&= -\frac{s x_i(1-x_i)}{2}\left[2\gamma_i(z-z_{\rm opt}) + \gamma_i^2(1-2x_i)\right] + \mu(1-2x_i) && (4)\\
V_i &= x_i(1-x_i) && (5)
\end{aligned}
$$

on using that the population-averaged phenotypic fitness

$$\bar{w} = 1 - \frac{s}{2}(\upsilon + (z - z_{\rm opt})^2) \tag{6}$$

with genetic variance

$$\upsilon = 2\sum_{i=1}^{\ell} \gamma_i^2 x_i(1-x_i) \tag{7}$$

In the following sections, we will study the allele frequency distribution analytically.

The stochastic model described above is also studied numerically using the nonlinear Langevin equation corresponding to the Fokker-Planck equation (2). Using Itô prescription, we obtain

$$dx_i(t) = M_i(\vec{x})dt + \sqrt{\frac{V_i(x_i)}{2N}}dW_i(t) \tag{8}$$

where dW_i is the Wiener process [refer to Sec. 4.3.5, Gardiner (1997)]. Dividing the time t into $t/\delta t$ intervals of equal length δt, the allele fre-

quency at the ith locus is updated as (Sec. 4.3.1, Gardiner (1997))

$$x_i(n+1) = x_i(n) + M_i(\vec{x}(n))\delta t + \sqrt{\frac{V_i(x_i(n))}{2N}}\ \eta\ ,\ 0 \leq n \leq \frac{t}{\delta t} \quad (9)$$

where η is a random variable chosen independently at each time step from a normal distribution with mean zero and variance δt. In all the numerical data presented in this article, we have used $\delta t = 0.1$.

We are interested in the situation where the population is initially equilibrated to a phenotypic optimum at z_0 and then adapts in response to a sudden shift in the optimum when mutation is strong ($4N\mu > 1$). The dynamics of adaptation are studied by tracking the frequency of + allele for $z_{\rm f} > z_0$ (or, − allele for $z_{\rm f} < z_0$). As the initial allele frequency distribution plays a crucial role in the dynamics, we first study the stationary state in detail followed by the time-dependent properties of the allele frequencies.

In the following, we will denote the deterministic quantities by calligraphic symbols and use an asterisk for quantities in the stationary state.

3 Steady state distribution

At mutation-selection-drift balance, the exact joint distribution of the allele frequencies when the phenotypic optimum is at z_0 can be obtained by setting the left-hand side (LHS) of the Fokker-Planck equa-

tion (2) to zero. We then obtain (Sec. 9, KIMURA (1964))

$$P^*(\vec{x}) \propto \bar{w}^{2N} \prod_{i=1}^{\ell} (x_i(1-x_i))^{4N\mu-1} \tag{10}$$

$$= e^{-Ns(z-z_0)^2} \prod_{i=1}^{\ell} (x_i(1-x_i))^{4N\mu-1} e^{-2Ns\gamma_i^2 x_i(1-x_i)} \tag{11}$$

which does not factorize due to epistatic interactions in the phenotypic fitness.

To understand the allele frequency distribution at a locus, the marginal distribution $\psi^*(x_i)$ of the frequency at the ith locus can be obtained by integrating the joint distribution $P^*(\vec{x})$ over all the allele frequencies except x_i. For a large number of loci, as shown in Appendix A, the single-locus distribution can be approximated by

$$\psi^*(x_i) \propto (x_i(1-x_i))^{4N\mu-1} e^{-2Ns\gamma_i^2 x_i(1-x_i)} e^{\frac{-Ns[\gamma_i(2x_i-1)-z_0]^2}{1+2Ns\kappa_{2,i}}} \tag{12}$$

where

$$\kappa_{2,i} = \sum_{j=1, j\neq i}^{\ell} \frac{\gamma_j^2}{(1+8N\mu)} \frac{{}_1F_1(\frac{3}{2}, 4N\mu+\frac{3}{2}, \frac{Ns\gamma_j^2}{2})}{{}_1F_1(\frac{1}{2}, 4N\mu+\frac{1}{2}, \frac{Ns\gamma_j^2}{2})} \tag{13}$$

increases linearly with ℓ and depends on the effect size of all the loci except that of the ith locus, and ${}_1F_1(a, b, z)$ is the Kummer confluent hypergeometric function (OLVER *et al.*, 2022, Chapter 13)

3.1 Phenotypic optimum at zero

It is instructive to first consider the case when the phenotypic optimum, $z_0 = 0$ and the number of loci, $\ell \to \infty$. Since the last factor on the right-hand side (RHS) of (12) approaches one as $\ell \to \infty$, we obtain

$$\psi^*(x_i) \overset{\ell\to\infty}{\longrightarrow} \frac{4^{4N\mu}\Gamma(4N\mu+\frac{1}{2})}{2\sqrt{\pi}\Gamma(4N\mu)} \frac{e^{-2Ns\gamma_i^2 x_i(1-x_i)}(x_i(1-x_i))^{4N\mu-1}}{e^{-\frac{Ns\gamma_i^2}{2}}{}_1F_1(\frac{1}{2};4N\mu+\frac{1}{2};\frac{Ns\gamma_i^2}{2})} \quad (14)$$

The above expression is also obtained if the first term on the RHS of (4) can be ignored so that the + allele is underdominant and subject to symmetric mutations (Hayward and Sella, 2022). For $Ns\gamma_i^2 \ll 2$ (weak selection), the above marginal distribution reduces to the well known result, *viz.*, beta distribution for a neutral allele (Ewens, 2004).

For weak mutation ($4N\mu < 1$), the marginal distribution (14) is U-shaped (see also Fig. S1). But for strong mutation ($4N\mu > 1$), which is the parameter regime of interest here, the distribution $\psi^*(x_i)$ has the following interesting property: it is unimodal if the effect size is below the threshold effect $\hat{\gamma}_N$ and bimodal otherwise. As shown in Appendix B, the threshold effect for $z_0 = 0$ and $\ell \to \infty$ is given by

$$\hat{\gamma}_N = \sqrt{\frac{8\mu}{s} - \frac{2}{Ns}} = \sqrt{\frac{2}{Ns}(4N\mu - 1)} \quad (15)$$

The maxima of the allele frequency distribution occur at

$$x_i^* = \begin{cases} \frac{1}{2} & , \gamma_i < \hat{\gamma}_N \quad (16a) \\ \frac{1}{2}\left(1 \pm \sqrt{1 - \frac{\hat{\gamma}_N^2}{\gamma_i^2}}\right) & , \gamma_i > \hat{\gamma}_N \quad (16b) \end{cases}$$

and for $\gamma_i > \hat{\gamma}_N$, the maxima of the bimodal distribution are separated by a minimum at frequency 1/2.

For exponentially-distributed effects, the fraction of large-effect loci is equal to

$$f_L = \int_{\hat{\gamma}_N}^{\infty} d\gamma \bar{\gamma}^{-1} e^{-\gamma/\bar{\gamma}} = e^{-\frac{\hat{\gamma}_N}{\bar{\gamma}}} \tag{17}$$

Figure 1 shows our numerical results for the marginal distribution when the phenotypic optimum is at zero, the number of loci are finite and most effects are small ($\hat{\gamma}_N \gg \bar{\gamma}$), and we find them to agree well with (12). The distribution is seen to have one maximum for a *small-effect locus* ($\gamma_i < \hat{\gamma}_N(\ell)$) and two maxima for a *large-effect locus* ($\gamma_i > \hat{\gamma}_N(\ell)$) where the threshold effect $\hat{\gamma}_N(\ell)$ for finite number of loci is given by (B.7) and is larger than $\hat{\gamma}_N$ given in (15); the corrections to the mode frequencies due to finite ℓ can also be obtained (see (B.4)).

3.2 Nonzero phenotypic optimum

We now consider the case of nonzero z_0 and large ℓ.

Marginal distribution: The numerical results for the marginal distribution displayed in Fig. 2 when most effects are small (see Fig. S2 when most effects are large) are found to be in good agreement with

(12). As for $z_0 = 0$, the distribution is unimodal below a threshold effect and bimodal above it; however, the stationary marginal distribution is no longer symmetric about the allele frequency one half (as is also evident from (12)). The asymmetry in the marginal distribution results in a qualitatively different behavior of the central moments of the allele frequency compared to when the phenotypic optimum is at zero; the mean, variance and skewness of the allele frequency distribution are discussed in detail in Sec. S1.

Threshold effect and modes: As derived in Appendix B and shown in Fig. S4, when the number of loci are much larger than the phenotypic optimum, the modes of the allele frequency distribution at a locus with effect size away from the threshold effect $\hat{\gamma}_N(\ell)$ are well approximated by the result (16) for infinite loci. But close to the threshold effect, they are substantially different; in particular, for positive (negative) phenotypic optimum, the maximum in the frequency distribution of the small-effect locus increases (decreases) with the effect size and occurs at a frequency which is substantially larger (smaller) than one half (see also Fig. 2).

For small z_0 and large ℓ, (B.6) shows that the threshold effect does not differ much from the infinite-loci result (15) when selection is strong ($Ns\bar{\gamma}^2 \gg 2$); this behavior holds for large z_0 also as shown in the inset of Fig. 3. Furthermore, (B.6) predicts that the threshold effect for large but finite number of loci is always larger than $\hat{\gamma}_N$ and increases with the magnitude of the phenotypic optimum. Thus,

a locus classified as a large-effect locus for a phenotypic optimum at zero can become a small-effect locus if the phenotypic optimum is large enough as illustrated in Fig. 3. This is because for large, positive (negative) phenotypic optimum, population will be well-adapted if the $+$ $(-)$ allele frequency at most loci is close to fixation.

Trait mean: In Appendix C, using (11), we reproduce the well known result that the stationary state distribution of the trait mean is a Gaussian (Bulmer, 1972; Lande, 1976), and focus on the effect of genetic architecture on the average deviation of the trait mean. Equation (C.7) shows that in a large population, the population is better adapted, on an average, when many large-effect loci are involved than if the quantitative trait is composed of mostly small effects.

Genetic variance: The equilibrium genetic variance, $\langle v^* \rangle$ is obtained using (12) in Sec. S2. When effects are equal and the phenotypic optimum is zero, an expression for $\langle v^* \rangle$ has been obtained in Bulmer (1972); here, for nonzero z_0 and exponentially-distributed effects, we find that $\langle v^* \rangle$ is weakly affected by the location of the phenotypic optimum.

3.3 Comparison to the deterministic stationary state

In an infinitely large population, the deterministic allele frequency x_i obeys (de Vladar and Barton, 2014; Jain and Stephan, 2017b)

$$\frac{dx_i(t)}{dt} = M_i(x_i) = \frac{-sx_i(1-x_i)}{2}\left[2\gamma_i(\varkappa - z_{\text{opt}}) + \gamma_i^2(1-2x_i)\right] + \mu(1-2x_i) \quad (18)$$

with the deterministic trait mean $\varkappa(t) = \sum_{i=1}^{\ell} \gamma_i(2x_i - 1)$. Assuming that the trait mean deviation is zero in the stationary state, it has been shown that the deterministic allele frequency is in stable equilibrium below a threshold effect $\hat{\gamma} = \sqrt{8\mu/s}$ and is bistable above it (de Vladar and Barton, 2014), and given by

$$x_i^* = \begin{cases} \frac{1}{2} & , \gamma_i < \hat{\gamma} \quad (19a) \\ \frac{1}{2}\left(1 \pm \sqrt{1 - \frac{\hat{\gamma}^2}{\gamma_i^2}}\right) & , \gamma_i > \hat{\gamma} \quad (19b) \end{cases}$$

Our (15) and (16), respectively, generalize the above results for threshold effect and equilibrium frequency to large, finite populations when average trait mean deviation is zero [refer to (C.6)].

However, there are key differences in the stationary state of infinite and finite populations: The stationary state solution of the Fokker-Planck equation (2) is unique [refer to Chapter 5 of van Kampen (1997)]; that is, it is independent of the initial allele frequencies for both small- and large-effect loci. Furthermore, as Fig. S6 illustrates,

in a finite population, the allele frequency of a large-effect locus spends a long time (presumably, exponentially long in population size) near one of the maxima before shifting to the other maximum. In contrast, in an infinite population, the stationary state allele frequency of a large-effect locus depends on the initial condition and does not shift between the two solutions given by (19b).

These points are further illustrated in Fig. 4 where the stationary state frequencies in the deterministic and stochastic model, starting from the same initial condition, are found to be quite close for small effect loci but not for the large-effect ones. However, the $N \to \infty$ limit of the stochastic model can be obtained if one averages over the initial conditions in the deterministic model. Equations (B.4a) and (B.4b) for the modes of the distribution can also be written in terms of the average trait mean deviation using (C.6); a comparison between the resulting expression for large N and the corresponding results in the deterministic model shows that while (B.4a) for the mode frequency matches the deterministic result (B2) of de Vladar and Barton (2014) for small-effect loci, (B.4b) does not agree with the corresponding result for large-effect loci. But, on averaging over both modes in (B.4b) [which amounts to averaging over the allele frequency distribution] and both equilibria in (B2) of de Vladar and Barton (2014), we obtain an agreement for large N.

Due to shifts in the equilibria for large-effect loci, the numerical results for the stationary state distribution (such as those shown in

Figs. 1 and 2) were obtained by averaging over many initial conditions (ensemble-averaging) as well as long time periods in the stationary state (time-averaging) as this allowed us to sample the distribution near both the allele frequency peaks efficiently. For small N, we have checked that as expected, the marginal distribution obtained by time-averaging alone matches the result (12).

4 Dynamics after a sudden shift in optimum

We now turn to the allele frequency dynamics when the population initially equilibrated to the phenotypic optimum at z_0 adapts in response to a sudden shift in the optimum to z_f. In an infinitely large population, previous work has shown that at short times, selective sweep at a large-effect locus is unlikely if most effects are small but they can occur if most effects are large (Jain and Stephan, 2017b). Here, we wish to evaluate if genetic drift enhances the chance of sweep in the former class of genetic architectures and therefore, in this and the following section, we work in the parameter regime where the quantitative trait is determined by a large number of small-effect loci and a few large-effect loci so that $\bar{\gamma} \ll \hat{\gamma}_N(\ell)$ [see (17)]; this also means that selection is weak, that is, $Ns\gamma_i^2 \ll 2$ for most loci. However, this does not imply that the dynamics are neutral for such loci as, at least at short times, the allele frequency at any locus is subject to direc-

tional selection with a time-dependent selection coefficient [refer to the first term on the RHS of (4)].

4.1 Stochastic trajectories

The stochastic dynamics of the phenotypic trait mean, genetic variance and allele frequencies are illustrated in Fig. 5 when all loci are equilibrated to an optimum at zero and their initial frequency is kept fixed for independent stochastic runs. When the optimum is suddenly shifted to $z_{\rm f} = 1$, the average trait mean deviation, $|\langle z(t)\rangle - z_{\rm f}|$ initially decreases rapidly with time reaching a value close to zero at $t \approx 50$, and then equilibrates slowly to the stationary state at $t \sim 10^3$. In the stationary state, the trait mean deviation fluctuates about the average trait mean deviation given by (C.6) [on replacing z_0 by $z_{\rm f}$] with a width proportional to $(Ns)^{-1/2}$ [refer to (C.5)]. The genetic variance is seen to increase but only mildly as there is a substantial initial genetic variance (when averaged over initial conditions, $\mathcal{v}^* \approx \ell\bar{\gamma}^2$, see (S2.6a)).

Figure 5 also shows the allele frequency dynamics of a small- and a large-effect locus. In the stationary state, while the frequency of small-effect locus exhibits small fluctuations about its average trajectory, as discussed in the last section, peak shifts in the allele frequency occur at the large-effect locus. These figures also suggest that at short times when the trait mean deviation is large, the width of the fluctuations in the allele frequency about the average trajectory increases with

time; the dynamics of the variance of the allele frequency and the time-dependent allele frequency distribution are discussed below.

5 Allele frequency distribution: linear noise approximation

To obtain some insight into the time-dependent distribution of the allele frequencies, we employ van Kampen's system-size expansion method for a large population which assumes that the fluctuations in the number of individuals with + allele are of the order of the square root of the population size [refer to Chapter X, (VAN KAMPEN, 1997)]. Therefore, the allele frequency can be written as

$$x_i(t) = \boldsymbol{x}_i(t) + \frac{\xi_i(t)}{\sqrt{2N}} \tag{20}$$

where $\xi_i \sim \mathcal{O}(1)$ is a stochastic variable with mean zero that captures the fluctuations about the deterministic frequency $\boldsymbol{x}_i(t)$ in a large finite population. This approximation is valid when the fluctuations are small (that is, $\sqrt{\langle x_i^2 \rangle - \langle x_i \rangle^2} \ll \boldsymbol{x}_i$), and therefore, due to shift in the equilibria for large-effect loci, we do not expect (20) to work close to, or in the stationary state for such loci. However, at short times, the fluctuations are expected to be small at any locus and (20) is valid; this is indeed supported by the data in Fig. 4 where the average allele frequency is well approximated by the deterministic result for

both small- and large-effect locus.

Then, far from the stationary state, as summarized in Appendix D and discussed in detail in Sec. S3, the time-dependent marginal distribution of the allele frequency at the ith locus can be approximated by a Gaussian centred about the deterministic allele frequency $x_i(t)$ and variance $\frac{\langle \xi_i^2 \rangle}{2N}$. Given the initial frequencies $\{x_j(0)\}$, we then have

$$\psi(x_i, t|x_i(0)) \quad = \quad \sqrt{\frac{2N}{\pi \langle \xi_i^2(t) \rangle}} \exp\left[-\frac{N(x_i - x_i(t))^2}{\langle \xi_i^2(t) \rangle} \right] \tag{21}$$

where, $x_i(t)$ is obtained from (18) and, as detailed in Appendix D, $\langle \xi_i^2 \rangle$ are determined by $\ell(\ell+1)/2$ coupled ordinary differential equations given by (D.6).

Although it does not appear possible to obtain exact results for the time-dependence of the variance as (D.6) is coupled and the coefficients A_{ij} and B_i are time-dependent (see also Sec. S3), numerics suggest that at short times, the allele frequency variance increases linearly with time; in other words, the rate of change of allele frequency variance is constant in time. For a fixed set of initial frequency, the RHS of (D.6a) can be approximated by $B_i(0)$ which immediately yields

$$\langle x_i^2 \rangle - \langle x_i \rangle^2 \approx \frac{x_i(0)(1 - x_i(0))t}{2N} \tag{22}$$

The above expression suggests that one may approximate the short time dynamics of the allele frequency by that of Brownian motion with time-dependent mean, $x_i(t)$. As desired, the variance vanishes

when the population size is infinitely large.

As the variance in the allele frequency increases linearly with time and the linear noise approximation is valid when the fluctuations are much smaller compared to the deterministic frequency, the Gaussian distribution (21) describes the dynamics over the time scales that diverge with the population size. Thus the average and deterministic allele frequencies are expected to match for a longer time for a larger population size, as verified in Fig. S7, and for all times in the limit of infinite population size as the time to shift the peak will also be infinite.

6 Selective sweeps at large-effect loci

In this section, we continue to focus on quantitative traits that are determined by mostly small-effect and a few large-effect loci. For a large-effect locus, due to the bimodal nature of the stationary state marginal distribution, the allele frequency trajectory starting at a frequency below the minimum of the distribution can 'sweep' to a frequency above the minimum of the distribution with some probability.

6.1 When does a sweep occur?

More precisely, we define the probability of sweep to be the probability that a large-effect locus with initial allele frequency below one half reaches a frequency above one half on time scales over which the trait

mean deviation becomes negligible (see Fig. 6 for such representative allele trajectories). This definition of a sweep in a large population is motivated by the corresponding one in an infinite population where the + allele's trajectory at a large-effect locus reaches a high frequency if it manages to exceed a frequency one half by the time the trait mean deviation becomes close to zero as due to disruptive selection (second term on the RHS of (4)), it is then guaranteed to fix (Chevin and Hospital, 2008; Jain and Stephan, 2017b). We therefore write

$$P_{sweep} = \int_0^\infty d\tau \mathrm{Prob}(\tau) \int_0^{1/2} dx \psi^*(x = x(0)) \int_{1/2}^1 dx' \psi(x', \tau | x(0)) \tag{23}$$

where $\mathrm{Prob}(\tau)$ is the distribution of time when the trait mean deviation lies in a small interval about zero for the first time and we have dropped the subscript of the large-effect locus for brevity. To make analytical progress, below we will make a series of approximations.

In an infinitely large population and for mostly small-effects genetic architecture, the trait mean given by

$$\Delta \varkappa(t; \{x_i(0)\}) = \Delta z(0) \exp\left[-s\upsilon^*(0)t\right] \tag{24}$$

approaches the phenotypic optimum over a time determined by the initial genetic variance (Jain and Stephan, 2017b). In the following discussion, we assume that at short times, this is a good approximation in finite populations as well (see Fig. 6) and the distribution of τ is sharply peaked around $(s\upsilon^*(0))^{-1}$ where $\upsilon^*(0)$ is the initial genetic

variance in a finite population at phenotypic optimum z_0 and can be approximated by $\langle \upsilon^* \rangle$ given by (S2.4) for large ℓ.

We next approximate the allele frequency distribution $\psi(x', \tau | x(0))$ at time τ by (21) which is a Gaussian centred about the deterministic allele frequency given by (Jain and Stephan, 2017b)

$$x_\tau \approx \left[1 + \frac{1 - x(0)}{x(0)} \exp\left(\frac{\gamma(z(0) - z_{\mathrm{f}})}{\upsilon^*(0)}\right)\right]^{-1} \tag{25}$$

$$\overset{\ell \gg 1}{\approx} \left[1 + \frac{1 - x(0)}{x(0)} \exp\left(\frac{\gamma(z_0 - z_{\mathrm{f}})}{\langle \upsilon^* \rangle}\right)\right]^{-1} \tag{26}$$

(where the last expression is obtained on using (C.7)) and whose variance can be approximated by (22). On integrating over the final frequency x', we then obtain

$$P_{sweep} \propto \int_0^{1/2} dx \psi^*(x = x(0)) \mathrm{erfc}\left[\sqrt{\frac{N}{x(0)(1 - x(0))\tau}}\left(\frac{1}{2} - x_\tau\right)\right] \tag{27}$$

where $\mathrm{erfc}(x)$ is the complementary error function which decreases monotonically from 2 towards 0 as x is increased from $-\infty$ to ∞.

For consistency, like the final distribution, the initial distribution in the above equation should also be obtained in the linear-noise approximation. But, as already discussed in the preceding sections, this approximation is not valid for a large-effect locus in the stationary state. If, however, one still insists on using the Gaussian approximation about the low-frequency mode, it is found to grossly underestimate the sweep probability as the main contribution to P_{sweep} comes

from the right *tail* of the stationary state distribution which is fatter than that of the Gaussian distribution. We will therefore use the marginal distribution (12) in the above expression of sweep probability. Furthermore, for large populations, the complementary error function in the integrand of (27) can be approximated by 2 if $x_\tau > 1/2$ and zero otherwise. This implies that a large-effect locus can sweep if the initial frequency of + allele exceeds a critical value, x_c:

$$\frac{1}{2} > x(0) > x_c \quad = \quad \left[1 + \exp\left(\frac{\gamma(z_\mathrm{f} - z_0)}{\langle \upsilon^* \rangle}\right)\right]^{-1} \tag{28}$$

In an infinitely large population, the criterion (28) reduces to

$$\frac{1}{2} > x^* > x_c = \left[1 + \exp\left(\frac{\gamma(z_\mathrm{f} - z_0)}{\upsilon^*}\right)\right]^{-1} \tag{29}$$

where $x^* = \frac{1}{2}(1 - \sqrt{1 - \frac{\hat{\gamma}^2}{\bar{\gamma}^2}})$ and $\upsilon^* \approx \ell\bar{\gamma}^2$. The criterion (29) obtained in previous work (Chevin and Hospital, 2008; Jain and Stephan, 2017b) shows that sweeps are unlikely unless the shift in the phenotypic optimum is very large or the number of loci is small.

6.2 Probability of sweep

Using (28), we finally obtain a simple expression for the sweep probability in a large finite population to be

$$P_{sweep} \approx \frac{\int_{x_c}^{1/2} dx \psi^*(x)}{\int_0^{1/2} dx \psi^*(x)} \tag{30}$$

which is normalized such that P_{sweep} lies between zero and one for $x_c = 1/2$ and 0, respectively. In Fig. 6, the probability of sweep is measured numerically where if the + allele's frequency at the major locus crosses one half by the time $(s\ell\bar{\gamma}^2)^{-1}$, it is counted as the sweep event. The result (30) is compared against the sweep probability obtained numerically when there is a single large-effect locus, and we observe a good agreement.

The integrals in (30) do not appear to be exactly solvable but one can obtain some insight into the nature of P_{sweep} for large ℓ. As the stationary genetic variance is proportional to the number of loci, the critical frequency is close to one half for large ℓ. Then to leading order in ℓ^{-1}, from (30) and (12), we obtain

$$P_{sweep} \approx \frac{[(x(1-x))^{4N\mu-1}e^{-2Ns\gamma^2x(1-x)}]|_{x=\frac{1}{2}}\left(\frac{1}{2}-x_c\right)}{\int_0^{1/2}(x(1-x))^{4N\mu-1}e^{-2Ns\gamma^2x(1-x)}dx} \quad (31)$$

$$= \frac{4\Gamma(4N\mu+\frac{1}{2})}{\sqrt{\pi}\Gamma(4N\mu)}\frac{\frac{1}{2}-x_c}{{}_1F_1(\frac{1}{2};4N\mu+\frac{1}{2};\frac{Ns\gamma^2}{2})} \quad (32)$$

$$\approx \frac{\Gamma(4N\mu+\frac{1}{2})}{\sqrt{\pi}\langle\upsilon^*\rangle\Gamma(4N\mu)}\frac{\gamma(z_{\rm f}-z_0)}{{}_1F_1(\frac{1}{2};4N\mu+\frac{1}{2};\frac{Ns\gamma^2}{2})} \quad (33)$$

on using (28) to obtain the last expression.

The above expression shows that the sweep probability decreases with increasing number of loci as the initial genetic variance is large and the contribution of a single large-effect locus is negligible for adaptation. But it is higher for larger shift in the optimum $z_{\rm f}-z_0$ and same for a given optimum shift on using that the stationary genetic variance

is independent of the location of the phenotypic optimum (however, (30) and numerics show that the sweep probability can depend on z_0 and z_f for small ℓ, refer to Fig. S10). If the effect of major locus is much higher than the threshold, using Olver *et al.* (2022, Eq. 13.7.1), we find that the sweep probability decays exponentially with the scaled selection, as $e^{-Ns\gamma^2/2}$; this is because for very large effect size, the initial distribution is very narrow and therefore the contribution from the tails is negligible. As expected, the sweep probability decreases with increasing population size. Thus the chance of a large change in the allele frequency of a large-effect locus is possible if $Ns\gamma^2/2$ is not too large.

In an infinitely large population with other parameters as in Fig. 6, a sweep can not occur for the locus with effect size $\gamma = 0.8$ and new optimum $z_\mathrm{f} = 1$ but the criterion (29) is satisfied when, for example, $z_\mathrm{f} = 15$. However, in a finite population with size $N = 1000$ and $z_\mathrm{f} = 15$, (30) predicts that the chance to sweep is less than one (≈ 0.968) as the initial frequency must exceed the critical frequency $x_c \approx 0.066$. This example thus illustrates that the chance of sweep is not necessarily higher in a finite population.

To see this point more generally, we first note that as the stationary genetic variance in a finite population is smaller than the standing variation in an infinitely large population, from (28) and (29), it follows that $x_c < \boldsymbol{x}_c$. Two cases need to be considered: if $x_c < \boldsymbol{x}_c < \boldsymbol{x}^*$, the frequency at the large-effect locus sweeps for sure in an infinite

population but there is small probability that the sweep does not occur in a finite population due to the lack of contribution from the *left* tail of the distribution to (28); on the other hand, for $x_c < x^* < x_c$ or $x^* < x_c < x_c$, a sweep does not occur in an infinitely large population but there is a nonzero chance for this to happen in a finite population due to the contribution from the *right* tail of the distribution. This suggests that with increasing N, the sweep probability increases in the former case and decreases in the latter case as verified in Fig. S9. The data in Fig. S9 also show that $1 - P_{sweep}$ in the former case and P_{sweep} in the latter case decay exponentially with the population size.

When there are $n_\ell > 1$ large-effect loci, the probability that at least one of them sweeps can be written as

$$\Pi_{sweep} = 1 - \prod_{i=1}^{n_\ell}(1 - P_{sweep,i}) \tag{34}$$

The above expression assumes that the sweeping probability of each large-effect locus is independent of other large-effect loci present in the population. The inset of the Fig. 6 shows a substantial difference between the numerical results and (34) which suggests that these loci may not be independently sweeping. As the above expression overestimates the numerics, this indicates selective interference between the sweeping loci that is reducing each effect's sweeping probability and needs a more careful investigation.

7 Discussion

In this article, we studied the stationary state and dynamical properties of a quantitative trait when a large but finite population is under stabilizing selection and strong mutation ($4N\mu > 1$). The deterministic analogue of this model (de Vladar and Barton, 2014; Jain and Stephan, 2015, 2017b) and a closely related stochastic model which assumes $4N\mu < 1$ (Hayward and Sella, 2022) have already been studied. Below we present a synthesis of the results obtained in these models, and discuss the differences and similarities in adaptation dynamics when the population size is small, moderately large or infinite.

Phenotypic properties: While the phenotypic mean deviation is observed to decay rapidly towards zero or a small finite value in all the parameter regimes (rapid phase), the genetic variance has been assumed to be nearly constant in many studies (see, for example, Lande (1976); Chevin and Hospital (2008)). Although this is shown to be a good approximation for genetic architectures where most effects are small (Jain and Stephan, 2017b), the genetic variance is seen to rise substantially in course of time when the initial genetic variance is small due to weak mutation; more precisely, this occurs in an infinite population when most effects are large (Jain and Stephan, 2017b) and in a finite population when scaled mutation rate is small (Hayward and Sella, 2022). Here, we find that the average genetic variance remains roughly constant.

Allele frequency properties: For large scaled mutation rate, a threshold effect that depends on the population-genetic parameters exists relative to which an effect can be classified as large or small. The threshold effect has been obtained from the stationary state properties in de Vladar and Barton (2014) for infinitely large population and extended here to large finite populations. In contrast, for small mutation rates, such a classification is done on the basis of the dynamical properties of the allele frequency (Höllinger *et al.*, 2019; Hayward and Sella, 2022).

In an infinitely large population, the stationary state allele frequency at a large-effect locus is bistable. But, in a finite population, the allele frequency distribution at such a locus is bimodal and the frequency can switch between the two modes of the stationary state distribution (see Fig. S6 for $4N\mu > 1$) spending a time which is presumably exponentially long in N close to either mode [refer to Barton and Rouhani (1987); Barton (1989) for $4N\mu < 1$].

Linear noise approximation: Previous work has focused on the effect of genetic background presented by a large number of small-effect loci on the chance of sweep of a large-effect locus either by treating the focal locus and the background loci deterministically (Chevin and Hospital, 2008; Jain and Stephan, 2017b), or the focal locus' dynamics stochastically but neglecting stochastic fluctuations in the background (Matuszewski *et al.*, 2015). Here, we have captured the effect of random genetic drift in a large population at all loci using the

so-called linear noise approximation (van Kampen, 1997) which is often used in biophysical and chemical problems and typically involves only a few random variables (Boland *et al.*, 2008; Garai *et al.*, 2012; Schnoerr *et al.*, 2017); to our knowledge, this method has not been used in the population-genetics literature and we applied it here when the number of random variables (allele frequencies) is large. This approximation is valid so long as the fluctuations about the deterministic frequency are small. For this reason, it is not expected to work in the stationary state where peak shifts may occur and even at short times for small mutation rates due to the absorption of the allele frequency.

Probability of sweep: While some studies do not support sweeps during polygenic adaptation (Chevin and Hospital, 2008; Jain and Stephan, 2017b; Hayward and Sella, 2022), these are predicted in other work (Stetter *et al.*, 2018; Thornton, 2019; Höllinger *et al.*, 2019) and in this article. The adaptive response of a polygenic trait crucially depends on the genetic architecture of the trait and the initial allele frequency distribution.

In an infinitely large population, the rate of change of allele frequency is proportional to its effect size suggesting that a large-effect locus can sweep but, for $z_{\mathrm{f}} > z_0$, the initial frequency of the + allele at this locus is small (see (18) and (19b)). Then it is not obvious if the allele frequency can sweep but detailed analyses have shown that a large change in the allele frequency can occur provided the initial allele frequency exceeds a critical allele frequency that depends on the

effect size, optimum shift and the initial standing genetic variation (Chevin and Hospital, 2008; Jain and Stephan, 2017b). This criterion is, however, satisfied for rather extreme parameters when, for example, the effect size or optimum shift are very large.

On the other hand, in small populations, although the allele frequency of a locus with larger effect size also changes faster, all the loci contributing to adaptation start at low frequency as the stationary state distribution is U-shaped (see Fig. S1). However, the initial standing genetic variation is dominated by moderate- and large-effect loci which, importantly, is independent of the effect size (Bürger, 2000); as a result, the rate of change of allele ferquency is inversely proportional to the effect size. Thus the intermediate-effect loci by virtue of larger change in their frequency dominate by the end of rapid phase and therefore the large-effect alleles fail to sweep (Sella and Barton, 2019; Hayward and Sella, 2022).

For moderately large population considered here, the criterion for a sweep to occur is found to be essentially the same as for an infinitely large population. However, the key difference is that the deterministic stationary distribution for a large-effect locus is not merely replaced by a Gaussian centred about the deterministic frequency, and instead it has a fatter tail which leads to a wider range of initial allele frequencies that can rise to a moderately high frequency by the end of the rapid phase and continue to rise (due to disruptive selection) towards appreciable frequency. Höllinger *et al.* (2019) have considered a

binary polygenic trait to explore the role of redundancy, and found that adaptation dynamics display sweeps and subtle shifts in allele frequency when the background mutation rate is, respectively, low and high; however, the connection between their results for a binary trait and our results for a quantitative trait is not clear.

Footprints of selective sweeps?: While selective sweeps for monogenic traits leave a clear footprint on linked neutral diversity, it has been difficult to observe these for polygenic adaptation (Berg and Coop, 2014), and the current emphasis is on developing techniques that are powerful enough to resolve the signals due to polygenic adaptation or other evolutionary forces such as genetic drift. Our work shows that sweeps can occur when mutation is moderately strong ($4N\mu > 1$), most effects are small ($Ns\bar{\gamma}^2 \lesssim 2$) but the large-effect locus is under moderate selection ($Ns\gamma_i^2 > 2$) as the probability of sweep decays exponentially fast with selection strength of the large-effect locus. Although a few numerical studies of models where neutral regions are linked to selected region have been carried out (Stetter *et al.*, 2018; Thornton, 2019), these have not been done in the moderate mutation-moderate selection regime and are clearly highly desirable to judge if the sweeps observed here play a significant role in the dynamics of adaptation.

8 Data Availability

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

9 Acknowledgement

A. D. was supported by the National Institutes of Health award R35-GM122566-01, National Science Foundation award DBI-2119963, and Moore and Simons Foundations Grant 735927 during part of this research work.

References

Barghi, N., J. Hermisson, and C. Schlötterer, 2020 Polygenic adaptation: a unifying framework to understand positive selection. Nat Rev Genet **21:** 769–781.

Barghi, N., R. Tobler, V. Nolte, A. M. Jakšić, F. Mallard, K. A. Otte, M. Dolezal, T. Taus, R. Kofler, and C. Schlötterer, 2019 Genetic redundancy fuels polygenic adaptation in Drosophila. PLoS Biol. **17:** e3000128.

Barton, N. H., 1989 The divergence of a polygenic system subject to stabilizing selection, mutation and drift. Genet. Res. **54(1):** 59–77.

Barton, N. H. and P. D. Keightley, 2002 Understanding quantitative genetic variation. Nat. Rev. Genet. **3:** 11–21.

Barton, N. H. and S. Rouhani, 1987 The frequency of shifts between alternative equilibria. J. theor. Biol. **125:** 397–418.

Berg, J. J. and G. Coop, 2014 A population genetic signal of polygenic adaptation. PLoS Genet. **10:** e1004412.

Boland, R. P., T. Galla, and A. J. McKane, 2008 How limit cycles and quasi-cycles are related in systems with intrinsic noise. J. Stat. Mech. **-:** P09001.

Boyle, E. A., Y. I. Li, and J. K. Pritchard, 2017 An expanded view of complex traits: from polygenic to omnigenic. Cell **169:** 1177–1186.

Bulmer, M. G., 1972 The genetic variability of polygenic characters under optimizing selection, mutation and drift. Genet. Res., Camb. **19:** 17–25.

Bürger, R., 2000 *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, Chichester.

Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose, and A. D. Long, 2010 Genome-wide analysis of a long-term evolution experiment with Drosophila. Nature **467:** 587–90.

Chevin, L.-M. and F. Hospital, 2008 Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics **180:** 1645–1660.

de Villemereuil, P., A. Charmantier, D. Arlt, P. Bize, P. Brekke, L. Brouwer, A. Cockburn, S. D. Côté, F. S. Dobson, S. R. Evans, M. Festa-Bianchet, M. Gamelon, S. Hamel, J. Hegelbach, K. Jerstad, B. Kempenaers, L. E. B. Kruuk, J. Kumpula, T. Kvalnes, A. G. McAdam, S. E. McFarlane, M. B. Morrissey, T. Pärt, J. M. Pemberton, A. Qvarnström, O. W. Røstad, J. Schroeder, J. C.

Senar, B. C. Sheldon, M. van de Pol, M. E. Visser, N. T. Wheelwright, J. Tufto, and L.-M. Chevin, 2020 Fluctuating optimum and temporally variable selection on breeding date in birds and mammals. Proc. Natl. Acad. Sci. U.S.A. **117:** 31969–31978.

de Vladar, H. P. and N. H. Barton, 2011 The statistical mechanics of polygenic character under stabilizing selection, mutation and drift. J. R. Soc. Interface **8:** 720–739.

de Vladar, H. P. and N. H. Barton, 2014 Stability and response of polygenic traits to stabilizing selection and mutation. Genetics **197:** 749–767.

Ewens, W., 2004 *Mathematical Population Genetics*. Springer, Berlin.

Ffrench-Constant, R. H., M. Bogwitz, P. Daborne, and J. Yen, 2002 A single P450 allele associated with insecticide resistance in Drosophila. Science **27:** 2253–2256.

Garai, A., B. Waclaw, H. Nagel, and H. Meyer-Ortmanns, 2012 Stochastic description of a bistable frustrated unit. J. Stat. Mech.**:** P01009.

Gardiner, C. W., 1997 *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Springer-Verlag.

Goddard, M. E. and B. J. Hayes, 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. **10:** 381–391.

Hayward, L. K. and G. Sella, 2022 Polygenic adaptation after a sudden change in environment. eLife **11:** e66697.

Hermisson, J. and P. S. Pennings, 2017 Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. Methods in Ecology and Evolution **8:** 700–716.

Höllinger, I., P. S. Pennings, and J. Hermisson, 2019 Polygenic adaptation: From sweeps to subtle frequency shifts. PLoS Genet **15:** e1008035.

Jain, K. and W. Stephan, 2015 Response of polygenic traits under stabilising selection and mutation when loci have unequal effects. G3: Genes, Genomes, Genetics **5:** 1065–1074.

Jain, K. and W. Stephan, 2017a Modes of rapid polygenic adaptation. Mol. Biol. Evol **34:** 3169–3175.

Jain, K. and W. Stephan, 2017b Rapid adaptation of a polygenic trait after a sudden environmental shift. Genetics **206:** 389–406.

Kimura, M., 1964 Diffusion models in population genetics. J. Appl. Prob. **1:** 177–232.

Kimura, M., 1965 A stochastic model concerning the maintenance of genetic variability in quantitative characters. Proc. Natl. Acad. Sci. USA **54:** 731–736.

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli, 2001 The strength of phenotypic selection in natural populations. Am. Nat. **157:** 245–261.

Lande, R., 1976 Natural selection and random genetic drift in phenotypic evolution. Evolution **30:** 314–334.

Lande, R., 1983 The response to selection on major and minor mutations affecting a metrical trait. Heredity **50:** 47–65.

Loh, P. R., G. Bhatia, A. Gusev, H. K. Finucane, B. K. Bulik-Sullivan, S. J. Pollack, Schizophrenia Working Group of Psychiatric Genomics Consortium, T. R. de Candia, S. H. Lee, N. R. Wray, K. S. Kendler, M. C. O'Donovan, B. M. Neale, N. Patterson, and A. L. Price, 2015 Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nat Genet. **47:** 1385–92.

Mackay, T. F. C., 2004 The genetic architecture of quantitative traits: lessons from *Drosophila*. Current Opinion in Genetics and Development **14:** 253–257.

Matuszewski, S., J. Hermisson, and M. Kopp, 2015 Catch me if

you can: Adaptation from standing genetic variation to a moving phenotypic optimum. Genetics **200:** 1255–1274.

Maynard Smith, J. and J. Haigh, 1974 Hitchhiking effect of a favourable gene. Genet. Res. **23:** 23–35.

Olver, F. W. J., A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, 2022 *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.1.6 of 2022-06-30.

Orr, H. A., 1998 The population genetics of adaptation: Distribution of factors fixed during adaptive evolution. Evolution **52:** 935–949.

Pritchard, J. K. and A. Di Rienzo, 2010 Adaptation - not by sweeps alone. Nature Review Genetics **11:** 665–667.

Pritchard, J. K., J. K. Pickrell, and G. Coop, 2010 The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. Current Biology **20:** R208–R215.

Robertson, A., 1956 The effect of selection against extreme deviants based on deviation or on homozygosis. Journal of Genetics **54:** 236–248.

Rockman, M. V., 2012 The QTN program and the alleles that

matter for evolution: all that's gold does not glitter. Evolution **66-1:** 1–17.

Sanjaka, J. S., J. Sidorenkoc, M. R. Robinson, K. R. Thornton, and P. M. Visscher, 2018 Evidence of directional and stabilizing selection in contemporary humans. Proc. Natl. Acad. Sci. USA **115:** 151–156.

Schnoerr, D., G. Sanguinetti, , and R. Grima, 2017 Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. J. Phys. A: Math. Theor. **50:** 093001.

Sella, G. and N. Barton, 2019 Thinking about the evolution of complex traits in the era of Genome-Wide Association Studies. Annu Rev Genomics Hum Genet **20:** 461–493.

Shi, H., G. Kichaev, and B. Pasaniuc, 2016 Contrasting the genetic architecture of 30 complex traits from summary association data. Am J Hum Genet. **99:** 139–153.

Sodeland, M., S. Jentoft, P. E. Jorde, M. Mattingsdal, J. Albretsen, A. R. Kleiven, A.-E. Wårøy Synnes, S. H. Espeland, E. M. Olsen, C. Andrè, N. C. Stenseth, and H. Knutsen, 2022 Stabilizing selection on Atlantic cod supergenes through a millennium of extensive exploitation. Proceedings of the National Academy of Sciences **119:** e2114904119.

Stetter, M. G., K. Thornton, and J. Ross-Ibarra, 2018 Ge-

netic architecture and selective sweeps after polygenic adaptation to distant trait optima. PLoS Genetics **14:** e1007794.

Ségurel, L. and C. Bon, 2017 On the evolution of lactase persistence in humans. Annu Rev Genomics Hum Genet. **18:** 297–319.

Therkildsen, N. O., A. P. Wilder, D. O. Conover, S. B. Munch, H. Baumann, and S. R. Palumbi, 2019 Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. Science **365:** 487–490.

Thornton, K., 2019 Polygenic adaptation to an environmental shift: temporal dynamics of variation under Gaussian stabilizing selection and additive effects on a single trait. Genetics **213:** 1513–1530.

Timpson, N. J., C. M. T. Greenwood, N. Soranzo, D. J. Lawson, and J. B. Richards, 2018 Genetic architecture: the shape of the genetic contribution to human traits and disease. Nat. Rev. Genet. **19:** 110–124.

Turner, T. L., A. D. Stewart, A. T. Fields, W. R. Rice, and A. M. Tarone, 2011 Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in Drosophila melanogaster. PLoS Genet. **7:** e1001336.

van Kampen, N. G., 1997 *Stochastic processes in physics and chemistry.* North Holland Personal Library.

van't Hof, A. E., N. Edmonds, M. Daliková, F. Marec, and I. J. Saccheri, 2011 Industrial melanism in British peppered moths has a singular and recent mutational origin. Science **332:** 958–960.

Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, and J. Yang, 2017 10 years of GWAS discovery: Biology, Function, and Translation. Am J Hum Genet. **101:** 5–22.

Wray, N. R., C. Wijmenga, P. F. Sullivan, J. Yang, and P. M. Visscher, 2018 Common disease is more complex than implied by the core gene omnigenic model. Cell **173:** 1573–1580.

Wright, S., 1937 The distribution of gene frequencies in populations. Proc. Natl. Acad. Sci USA **23:** 307–320.

Yengo, L., S. Vedantam, E. Marouli, and et al., 2022 A saturated map of common genetic variants associated with human height. Nature **610:** 704–712.

Appendix A Stationary state marginal distribution

The marginal distribution $\psi^*(x_i)$ of the allele frequency in the stationary state can be found by integrating the joint distribution $P^*(\vec{x})$ given by (11) over the frequencies of all but the ith locus; this gives

$$\psi^*(x_i) \quad \propto \quad g^*(x_i) \prod_{j \neq i} \int_0^1 dx_j g^*(x_j) e^{-\frac{Ns(\Delta z^*)^2}{\ell - 1}} \tag{A.1}$$

where

$$g^*(x_i) \propto e^{-2Ns\gamma_i^2 x_i(1-x_i)} (x_i(1-x_i))^{4N\mu - 1} \tag{A.2}$$

and $\Delta z^* = z^* - z_0 = \sum_{j=1}^{\ell} \gamma_j (2x_j - 1) - z_0$. To evaluate the multiple integrals in (A.1), we rewrite its RHS as

$$\frac{\psi^*(x_i)}{g^*(x_i)} \propto \int_{-\Gamma'}^{\Gamma'} dX' e^{-Ns(X' + \gamma_i(2x_i - 1) - z_0)^2} \mathrm{Prob}(X' = \sum_{j \neq i} \gamma_j (2x_j - 1))$$

where $\Gamma' = \sum_{j \neq i} \gamma_j$, and X' is a sum of independent but non-identically distributed random variables chosen from $g^*(x_j)$. We then obtain

$$\begin{aligned} \frac{\psi^*(x_i)}{g^*(x_i)} \quad &\propto \quad \int_{-\Gamma'}^{\Gamma'} dX' e^{-Ns(X' + \gamma_i(2x_i - 1) - z_0)^2} \\ &\times \quad \int_0^1 ... \int_0^1 \delta[X' - \sum_{j \neq i} \gamma_j (2x_j - 1)] \prod_{j \neq i} dx_j g^*(x_j) \end{aligned} \tag{A.3}$$

The inner integral on the RHS of the above equation can be calculated by appealing to the central limit theorem for large ℓ, and we

get

$$\frac{\psi^*(x_i)}{g^*(x_i)} \overset{\ell\gg 1}{\propto} \int_{-\infty}^{\infty} dX' e^{-Ns(X'+\gamma_i(2x_i-1)-z_0)^2} e^{-\frac{(X'-\kappa_{1,i})^2}{2\kappa_{2,i}}} \tag{A.4}$$

$$\propto \exp\left[-\frac{Ns(\gamma_i(2x_i-1)-z_0)^2}{1+2Ns\kappa_{2,i}}\right] \tag{A.5}$$

where $\kappa_{1,i}$ and $\kappa_{2,i}$ are, respectively, the mean and variance of the sum $X' = \sum_{j\neq i}\gamma_j(2x_j-1)$ when averaged over the (normalized) distribution $\prod_{j\neq i} g^*(x_j)$. Since $g^*(x_i)$ is symmetric about $x_i = 1/2$, it follows that the mean $\kappa_{1,i} = 0$; furthermore, the variance $\kappa_{2,i}$ increases linearly with ℓ and its properties are discussed in detail below (see (A.9)).

If the phenotypic optimum is small (that is, $z_0 \sim \mathcal{O}(1)$), the exponential on the RHS of (A.5) can be expanded in powers of ℓ^{-1}; however, as explained in Sec. S4, corrections to the central limit theorem are required to obtain the correct expression for ψ^*/g^* to $\mathcal{O}(\ell^{-1})$ which finally yields

$$\frac{\psi^*(x_i)}{g^*(x_i)} \propto 1 + \frac{\kappa_4 - 4\kappa_2(\gamma_i(2x_i-1)-z_0)^2}{8\kappa_2^2} + \mathcal{O}(\ell^{-2}) \tag{A.6}$$

where $\kappa_2 = \sum_{j=1}^{\ell}\langle\gamma_j^2(2x_j-1)^2\rangle_{g^*}$ and $\kappa_4 = \sum_{j=1}^{\ell}\langle\gamma_j^4(2x_j-1)^4\rangle_{g^*} - 3\langle\gamma_j^2(2x_j-1)^2\rangle_{g^*}^2$ are, respectively, the second and fourth cumulant of the random variable $\gamma_j(2x_j-1)$ obtained using the (normalized) distribution $g^*(x_j)$. On the other hand, if the phenotypic optimum is

large (that is, $z_0 \sim \mathcal{O}(\ell)$), from (A.5), we obtain

$$\frac{\psi^*(x_i)}{g^*(x_i)} \propto e^{-\frac{\gamma_i z_0(1-2x_i)}{\kappa_2}} \tag{A.7}$$

on using that the variance $\kappa_{2,i}$ depends linearly on ℓ. As (A.5) is simpler and quite accurate for most purposes, unless specified otherwise, we work with it for much of the discussion.

In the marginal distribution $\psi^*(x_i)$, the effect of other $\ell - 1$ loci due to epistatic interactions in the phenotypic fitness appears through the variance $\kappa_{2,i}$. Since $\langle x_i \rangle_{g^*} = 1/2$ for all loci, the variance $\kappa_2 = \sum_{j=1}^{\ell} 4\gamma_j^2(\langle x_j^2 \rangle_{g^*} - \langle x_j \rangle_{g^*}^2)$ which shows that κ_2 is simply a weighted sum of the variance of the distribution $g^*(x_i)$. Performing the integrals over the allele frequency, we obtain

$$\kappa_2 = \sum_{j=1}^{\ell} 4\gamma_j^2 \left(\frac{1}{4(1+8N\mu)} \frac{{}_1F_1(\frac{3}{2}, 4N\mu + \frac{3}{2}, \frac{Ns\gamma_j^2}{2})}{{}_1F_1(\frac{1}{2}, 4N\mu + \frac{1}{2}, \frac{Ns\gamma_j^2}{2})} \right) \tag{A.8}$$

$$\approx \ell \int_0^\infty \frac{p(\gamma)\gamma^2}{1+8N\mu} \frac{{}_1F_1(\frac{3}{2}, 4N\mu + \frac{3}{2}, \frac{Ns\gamma^2}{2})}{{}_1F_1(\frac{1}{2}, 4N\mu + \frac{1}{2}, \frac{Ns\gamma^2}{2})} \tag{A.9}$$

where $p(\gamma)$ is the distribution of effects and ${}_1F_1(a, b, z)$ is the Kummer confluent hypergeometric function. For exponentially-distributed effect size with mean $\bar{\gamma}$, we have

$$\kappa_2 \approx \frac{\ell\bar{\gamma}^2}{1+8N\mu} \int_0^\infty \frac{{}_1F_1(\frac{3}{2}, 4N\mu + \frac{3}{2}, \frac{Ns\bar{\gamma}^2}{2}y^2)}{{}_1F_1(\frac{1}{2}, 4N\mu + \frac{1}{2}, \frac{Ns\bar{\gamma}^2}{2}y^2)} y^2 e^{-y} dy \tag{A.10}$$

As the above integral does not appear to be exactly solvable, we estimate it by noting that the ratio of the Kummer confluent hypergeo-

metric function in the above integrand is a monotonically increasing function of its argument, and for fixed U, this ratio may be approximated by (Olver *et al.*, 2022)

$$\frac{{}_1F_1(\frac{3}{2}, U+\frac{3}{2}, a)}{{}_1F_1(\frac{1}{2}, U+\frac{1}{2}, a)} = \begin{cases} 1 + \dfrac{4Ua}{4U^2+8U+3}\ , & a \ll 1 \quad \text{(A.11a)} \\ (2U+1)\left(1-\dfrac{U}{a}\right)\ , & a \gg 1 \quad \text{(A.11b)} \end{cases}$$

For a given $4N\mu$, we then have

$$\kappa_2 \approx \begin{cases} \dfrac{\ell\bar{\gamma}^2}{1+8N\mu}\left(2+\dfrac{1536\bar{\gamma}^2N^2\mu^2}{\hat{\gamma}^2(8N\mu+1)(8N\mu+3)}\right) & ,\ Ns\bar{\gamma}^2 \ll 2 \quad \text{(A.12a)} \\ \ell\bar{\gamma}^2\left(2-\dfrac{\hat{\gamma}^2}{\gamma^2}\right) & ,\ Ns\bar{\gamma}^2 \gg 2 \quad \text{(A.12b)} \end{cases}$$

When selection is weak ($Ns\bar{\gamma}^2 \ll 2$) or most effects are small [$\bar{\gamma}^2 \ll (4N\mu-1)\bar{\gamma}^2 \ll \hat{\gamma}_N^2$ on using (15) and (17)], (A.12a) shows that κ_2 approaches zero with increasing population size as $1/N$; in contrast, when selection is strong or most effects are large, due to (A.12b), the variance κ_2 remains nonzero in the deterministic limit. This can be understood as follows: as the width of $g^*(x_i)$ about a maximum is expected to decrease with N, for large N, one may approximate $g^*(x_i)$ by a Dirac delta function centred at 1/2 for small effect locus, and an average of two Dirac delta functions located at $x_{i,\pm} = \frac{1}{2}\left(1 \pm \sqrt{1-\frac{\hat{\gamma}^2}{\gamma_i^2}}\right)$ for large-effect locus (de Vladar and Barton, 2014). Thus, for small-effect locus, as the distribution is unimodal and sharply-peaked in large populations, the variance vanishes in the deterministic limit. But for large-effect locus, as a conse-

quence of the bimodality, the distribution remains broad resulting in a nonzero variance. In Fig. S8, it is verified that κ_2 increases with increasing selection strength in accordance with (A.12).

Appendix B Threshold effect size and mode frequency

According to (A.5), the stationary state allele frequency distribution at the ith locus is given by

$$\psi^*(x_i) \propto e^{-2Ns\gamma_i^2 x_i(1-x_i)}(x_i(1-x_i))^{4N\mu-1}e^{-\frac{Ns[\gamma_i(2x_i-1)-z_0]^2}{1+2Ns\kappa_{2,i}}} \tag{B.1}$$

The modes of the distribution, x_i^* can be found on setting the derivative $\frac{d\psi^*}{dx_i}$ equal to zero which yields the following cubic equation in x_i^*:

$$(4N\mu-1)\frac{(1-2x_i^*)}{x_i^*(1-x_i^*)}-2Ns\gamma_i^2(1-2x_i^*)+\frac{4Ns\gamma_i(z_0-\gamma_i(2x_i^*-1))}{1+2Ns\kappa_{2,i}}=0 \tag{B.2}$$

The above equation has two complex roots and one real root below the threshold effect $\hat{\gamma}_N(\ell)$ and three real roots above it. This change in the behavior of x_i^* occurs when the discriminant of the above cubic polynomial is equal to zero:

$$D = a_2^2a_1^2 - 4a_3a_1^3 + 18a_0a_2a_3a_1 - a_0\left(4a_2^3 + 27a_0a_3^2\right) = 0 \tag{B.3}$$

where $a_i, i = 0, 1, 2, 3$ are the coefficients of x^i in (B.2). The resulting equation is a 3rd order equation in $\hat{\gamma}_N^2(\ell)$ and can be solved numerically to obtain the threshold effect for finite N and ℓ.

One can, however, obtain an analytical expression for $\hat{\gamma}_N(\ell)$ when the number of loci are very large. Since κ_2 grows linearly with ℓ, for infinite number of loci, the last term on the LHS of (B.2) vanishes yielding (15) and (16) for the threshold effect $\hat{\gamma}_N$ and mode allele frequency x_i^*, respectively, which are independent of z_0. For large ℓ but small z_0, using (A.6), we find that the steady state distribution has maximum at

$$x_i^*(\ell) = \begin{cases} \dfrac{1}{2} + \dfrac{1}{2Ns\kappa_2}\dfrac{\gamma_i z_0}{\hat{\gamma}_N^2 - \gamma_i^2} & , \gamma_i < \hat{\gamma}_N(\ell) \quad \text{(B.4a)} \\ \dfrac{1}{2}\left(1 \mp \sqrt{1 - \dfrac{\hat{\gamma}_N^2}{\gamma_i^2}}\right) + \dfrac{1}{2Ns\kappa_2}\dfrac{\hat{\gamma}_N^2(z_0 \pm \sqrt{\gamma_i^2 - \hat{\gamma}_N^2})}{2\gamma_i(\gamma_i^2 - \hat{\gamma}_N^2)} & , \gamma_i > \hat{\gamma}_N(\ell) \quad \text{(B.4b)} \end{cases}$$

where $\hat{\gamma}_N(\ell)$ denotes the threshold effect for finite ℓ. For $\gamma_i > \hat{\gamma}_N(\ell)$, the minimum in the allele frequency is given by the expression on the RHS of (B.4a).

As shown in Fig. S4, a threshold effect exists below which (B.2) has only one real root and corresponds to the maximum in the unimodal distribution. But, above the threshold effect, two additional real roots of (B.2) appear which give the allele frequency at which the minimum and the second maximum of the bimodal distribution occur. Thus, at the threshold effect, for positive (negative) z_0, the minimum and the low-frequency (high-frequency) maximum of the bimodal distribution

coincide. On matching the solutions (B.4a) and (B.4b), we get

$$\pm\, 2Ns\kappa_2\Delta^3 = 3\hat{\gamma}_N^2 z_0 \pm \hat{\gamma}_N^2\Delta + 2z_0\Delta^2 \tag{B.5}$$

where $\Delta^2 = \hat{\gamma}_N^2(\ell) - \hat{\gamma}_N^2$ decreases with increasing ℓ. The above cubic equation for Δ is exactly solvable, and has two complex roots and one real root. Here, we estimate the real root by noting that the first term on the RHS which is independent of ℓ can be balanced if $\kappa_2\Delta^3$ is also independent of ℓ thus yielding

$$\hat{\gamma}_N(\ell) \approx \hat{\gamma}_N + \frac{1}{2\hat{\gamma}_N}\left(\frac{3\hat{\gamma}_N^2|z_0|}{2Ns\kappa_2}\right)^{2/3} + \mathcal{O}(\ell^{-1})\ ,\ z_0 \neq 0 \tag{B.6}$$

which shows that the deviation $\hat{\gamma}_N(\ell) - \hat{\gamma}_N$ decays rather slowly with ℓ. The above expression also shows that the threshold effect always increases with the absolute value of the phenotypic optimum. But it does not change if the phenotypic optimum shifts between z_0 and $-z_0$. The threshold effect is also larger when selection is weaker ($Ns\bar{\gamma}^2 \ll 2$) or the quantitative trait is controlled mostly by small-effect loci. For phenotypic optimum at zero, (B.5) gives

$$\hat{\gamma}_N(\ell) \approx \hat{\gamma}_N + \frac{\hat{\gamma}_N}{4Ns\kappa_2} \tag{B.7}$$

so that the deviation $\hat{\gamma}_N(\ell) - \hat{\gamma}_N$ is of order ℓ^{-1}.

Appendix C Stationary state trait mean distribution

In the stationary state, using the joint distribution (11) for ℓ loci, we find that the distribution of trait mean is given by

$$
\begin{aligned}
Pr(z^*) &= \int_0^1 dx_1 ... \int_0^1 dx_\ell P(\vec{x})\delta(z - z^*) && \text{(C.1)} \\
&\propto e^{-Ns\Delta z^{*2}} \int_0^1 ... \int_0^1 \delta(z^* - \sum_i \gamma_i(2x_i - 1)) \prod_{i=1}^{\ell} dx_i g^*(x_i) && \text{(C.2)} \\
&\propto e^{-Ns\Delta z^{*2}} e^{-\frac{z^{*2}}{2\kappa_2}} && \text{(C.3)}
\end{aligned}
$$

where $\Delta z^* = z^* - z_0$ and the last expression is obtained using the central limit theorem for large ℓ (Bulmer, 1972; Lande, 1976). The above distribution gives the average and variance of the trait mean to be (see also Sec. S4)

$$
\begin{aligned}
\langle z^* \rangle &= \frac{2Ns\kappa_2 z_0}{1 + 2Ns\kappa_2} \approx z_0 \left(1 - \frac{1}{2Ns\kappa_2} + \mathcal{O}(\ell^{-2})\right) && \text{(C.4)} \\
\langle (z^* - \langle z^* \rangle)^2 \rangle &= \frac{\kappa_2}{1 + 2Ns\kappa_2} \approx \frac{1}{2Ns} - \frac{1}{4N^2 s^2 \kappa_2} + \mathcal{O}(\ell^{-2}) && \text{(C.5)}
\end{aligned}
$$

Here, we are interested in understanding how the average deviation in the trait mean depends on the number of loci and the genetic architecture of the trait. From (C.4), we find the deviation in the

trait mean to be

$$\langle \Delta z^* \rangle \approx -\frac{z_0}{2Ns\kappa_2} = -\frac{z_0}{4\ell \times \frac{Ns\bar{\gamma}^2}{2} \times \frac{\kappa_2}{\ell\bar{\gamma}^2}} \tag{C.6}$$

which shows that if the magnitude of phenotypic optimum does not increase with the number of loci, the deviation in the mean phenotype tends to vanish and the population is perfectly adapted (on an average). But, if the phenotypic optimum varies linearly with ℓ, the average trait mean also increases but the average deviation remains unchanged.

When most loci have small effect, due to (A.12a), $2Ns\kappa_2 \to$ constant when $N \to \infty$, and therefore, the average mean deviation remains finite in the deterministic limit. But, when most effects are large, as $2Ns\kappa_2$ increases linearly with N (see (A.12b)), we expect the average mean deviation to vanish in an infinitely large population. These conclusions are verified numerically in Fig. S8, and are also consistent with the fact that for weak selection, the average trait mean deviation and the variance in trait mean are large as stabilizing selection is ineffective in keeping the population close to the phenotypic optimum while for strong selection, the population stays close to z_0 (Bulmer, 1972; Lande, 1976). Using (A.12a) and (A.12b), we may therefore write

$$\langle \Delta z^* \rangle \approx -\frac{z_0}{1 + \frac{4\ell f_S \bar{\gamma}^2}{\hat{\gamma}^2} + 4\ell f_L N s \bar{\gamma}^2} \tag{C.7}$$

where $f_L = e^{-\frac{\hat{\gamma}_N(\ell)}{\bar{\gamma}}}$ and $f_S = 1 - f_L$, respectively, denote the fraction of small- and large-effect loci in a quantitative trait. The variance in Δz^* given by (C.5) is, however, independent of the details of the genetic architecture.

In an infinitely large population, the correction to stationary allele frequency due to nonzero mean deviation has been found, assuming that this deviation is small (de Vladar and Barton, 2014). However, the number of loci and the genetic architecture for which the mean deviation will be small is not stated. Our result (C.7) show that this would be the case when the number of loci are large and selection is strong or a finite fraction of effects are large.

Appendix D Time-dependent marginal distribution

As described in Sec. S3, the joint distribution of the deviation $\{\xi_i\}, i = 1, ..., \ell$ in allele frequencies obeys the following approximate equation (see (S3.12)):

$$\frac{\partial \Pi(\vec{\xi}, t)}{\partial t} = -\sum_{i,j=1}^{\ell} A_{ij}(t) \frac{\partial}{\partial \xi_i}(\xi_j \Pi(\vec{\xi}, t)) + \frac{1}{2}\sum_{i=1}^{\ell} B_i(t) \frac{\partial^2 \Pi(\vec{\xi}, t)}{\partial \xi_i^2} \quad \text{(D.1)}$$

where

$$A_{ij}(t) = \frac{\partial M_i(\vec{x})}{\partial x_j} \tag{D.2}$$

$$B_i(t) = V_i(\vec{x}) \tag{D.3}$$

Thus, $\Pi(\vec{\xi}, t)$ is the distribution of multivariate Ornstein-Uhlenbeck process with time-dependent coefficients, and whose exact solution is a multivariate Gaussian distribution with time-dependent mean and variance; from (S3.25), it then follows that the time-dependent marginal distribution of the allele frequency at the ith locus is also a Gaussian. Thus, within linear noise approximation, the stochastic trajectory of an allele frequency is centred about the deterministic one with fluctuations of the order $N^{-1/2}$ as the variance $\langle (x_i - x_i(t))^2 \rangle = \frac{\langle \xi_i^2 \rangle}{2N}$.

The variance and covariance of the deviation ξ_i defined as

$$\langle \xi_k^2(t) \rangle = \prod_{i=1}^{\ell} \int_{-\infty}^{\infty} d\xi_i \xi_k^2 \Pi(\vec{\xi}, t) \tag{D.4}$$

$$\langle \xi_k(t)\xi_j(t) \rangle = \prod_{i=1}^{\ell} \int_{-\infty}^{\infty} d\xi_i \xi_k \xi_j \Pi(\vec{\xi}, t) \ , \ k \neq j \tag{D.5}$$

obey a set of coupled differential equations that can be derived from (D.1). On multiplying both sides of (D.1) by ξ_k^2, integrating over all

ξ_i's and carrying out integration by parts, we obtain,

$$\frac{d\langle\xi_i\xi_k\rangle}{dt} = \begin{cases} 2\sum_{j\neq i} A_{ij}(t)\langle\xi_i\xi_j\rangle + 2A_{ii}(t)\langle\xi_i^2\rangle + B_i(t)\ ,\ i=k & \text{(D.6a)} \\ \sum_{j\neq k} A_{ij}(t)\langle\xi_j\xi_k\rangle + \sum_{j\neq i} A_{kj}(t)\langle\xi_i\xi_j\rangle & \\ +A_{ik}(t)\langle\xi_k^2\rangle + A_{ki}(t)\langle\xi_i^2\rangle\ ,\ i\neq k & \text{(D.6b)} \end{cases}$$

In the above equations,

$$A_{ij} = \frac{\partial M_i}{\partial x_j} = \begin{cases} -s\gamma_i(1-2x_i)\Delta\varkappa - \frac{s\gamma_i^2}{2}(1-2x_i(1-x_i)) - 2\mu & ,\ i=j \text{ (D.7a)} \\ -2s\gamma_i\gamma_j x_i(1-x_i)(1-\delta_{\Delta\varkappa,0}) & ,\ i\neq j \text{ (D.7b)} \end{cases}$$

and

$$B_{ij} = V_i\delta_{ij} = \begin{cases} x_i(1-x_i) & ,\ i=j \quad \text{(D.8a)} \\ 0 & ,\ i\neq j \quad \text{(D.8b)} \end{cases}$$

where $\Delta\varkappa = \varkappa - z_{\mathrm{f}}$ is the trait mean deviation in an infinitely large population.

A formal solution of these coupled equations can be written (see (S3.27)) but it seems difficult to find an exact solution.

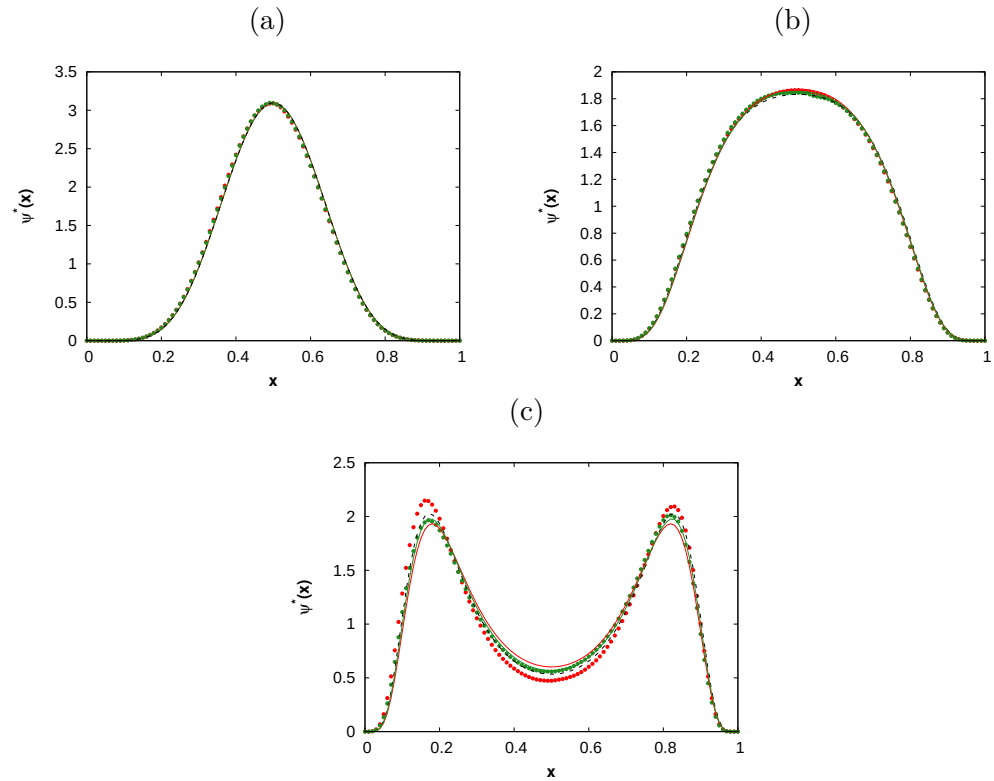

Figure 1: Stationary state marginal allele frequency distribution for $\ell = 100$ (red) and 500 (green) at a locus with effect size (a) smaller ($\gamma = 0.1$), (b) close to ($\gamma = 0.5$) and (c) larger ($\gamma = 0.7$) than the threshold size $\hat{\gamma}_N = 0.53$ for $\ell \to \infty$. The other parameters are $N = 1000$, $s = 0.05$, and $\mu = 0.002$, $\bar{\gamma} = 0.1$, and $z_0 = 0$. The points are obtained by solving the Langevin equation (9) numerically for 10^4 independently chosen initial conditions and averaged until $t = 10^5$ after a burn-in period of $t = 4 \times 10^4$. The solid line represents the analytical expression for marginal distribution (12) where $\kappa_{2,i} \approx 0.76(\ell = 100)$ and $1.4(\ell = 500)$ are calculated from the effect sizes used in this plot. The black dashed line represents the marginal distribution (14) for $\ell \to \infty$. The trait is composed of mostly small effects ($\sim 1\%$ large-effect loci). For different ℓ, different set of effects were generated keeping the effect size of the loci shown here to be the same.

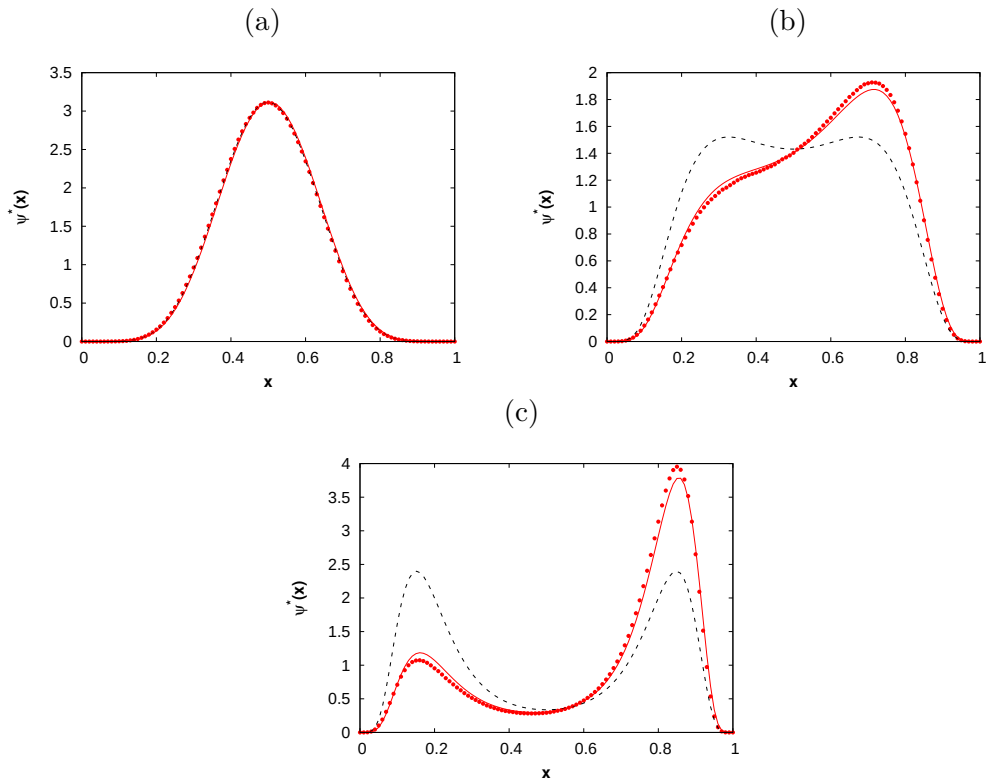

Figure 2: Stationary state marginal allele frequency distribution for a locus with effect size (a) smaller ($\gamma_i = 0.05$), (b) close to ($\gamma_i = 0.4$) and (c) larger ($\gamma_i = 0.5$) than the threshold size $\hat{\gamma}_N(\ell) \approx 0.42$. The parameters are $N = 1000$, $s = 0.1$, $\mu = 0.002, \ell = 1000$, $\bar{\gamma} = 0.08$, and $z_{\rm f} = 2$. For the set of effects used here, there were 7 large-effects loci. The points are obtained by numerically solving the Langevin equation (9) and averaging over 10^5 steady-state time average data and 10^3 ensemble avearge data after a burn-in period of $t = 4 \times 10^4$. The red solid line represents the analytical expression (12) where $\kappa_2 \approx 1.24$ is calculated from the effect sizes used in this plot, and the black dashed line represents the marginal distribution (14) for $\ell \to \infty$.

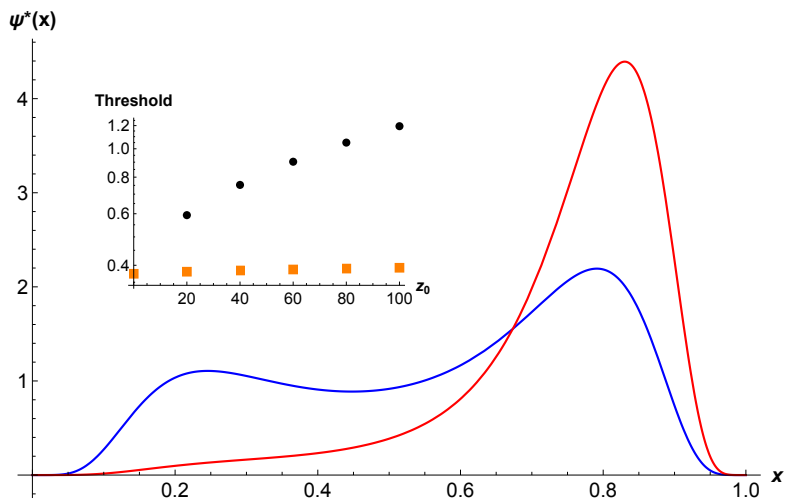

Figure 3: The inset shows the threshold effect, $\hat{\gamma}_N(\ell)$ obtained using (B.3) for weak selection ($Ns\bar{\gamma}^2 = 0.64$, black), and strong selection ($Ns\bar{\gamma}^2 = 16$, orange) where $N = 1000, \mu = 0.002, s = 0.1, \ell = 1000$. The threshold effect increases with the phenotypic optimum for weak selection but it is mildly affected when selection is strong. For weak selection, the marginal distribution at a locus with effect size $\gamma = 0.45$ is bimodal when $z_0 = 2$ (blue) but it is unimodal and heavily skewed towards high frequency when the optimum is at 10 (red).

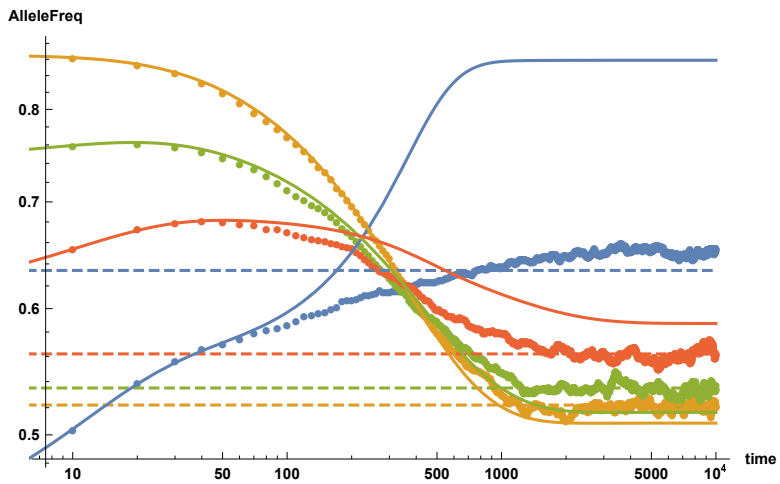

Figure 4: Comparison of deterministic allele frequency (solid lines) obtained by numerically solving (18) and average allele frequency (points) obtained by numerically solving (9) for 5000 independent stochastic runs for effect size 0.24 (yellow), 0.36 (green), 0.52 (red), 0.8 (blue), keeping the initial frequencies to be the same in both deterministic and stochastic model. The other parameters are $\ell = 200, s = 0.05, \mu = 0.002, N = 200$. The threshold frequency $\hat{\gamma}_N(\ell)$ is 0.359 and 0.493 for $z_0 = 0$ and $z_{\mathrm{f}} = 1$, respectively.

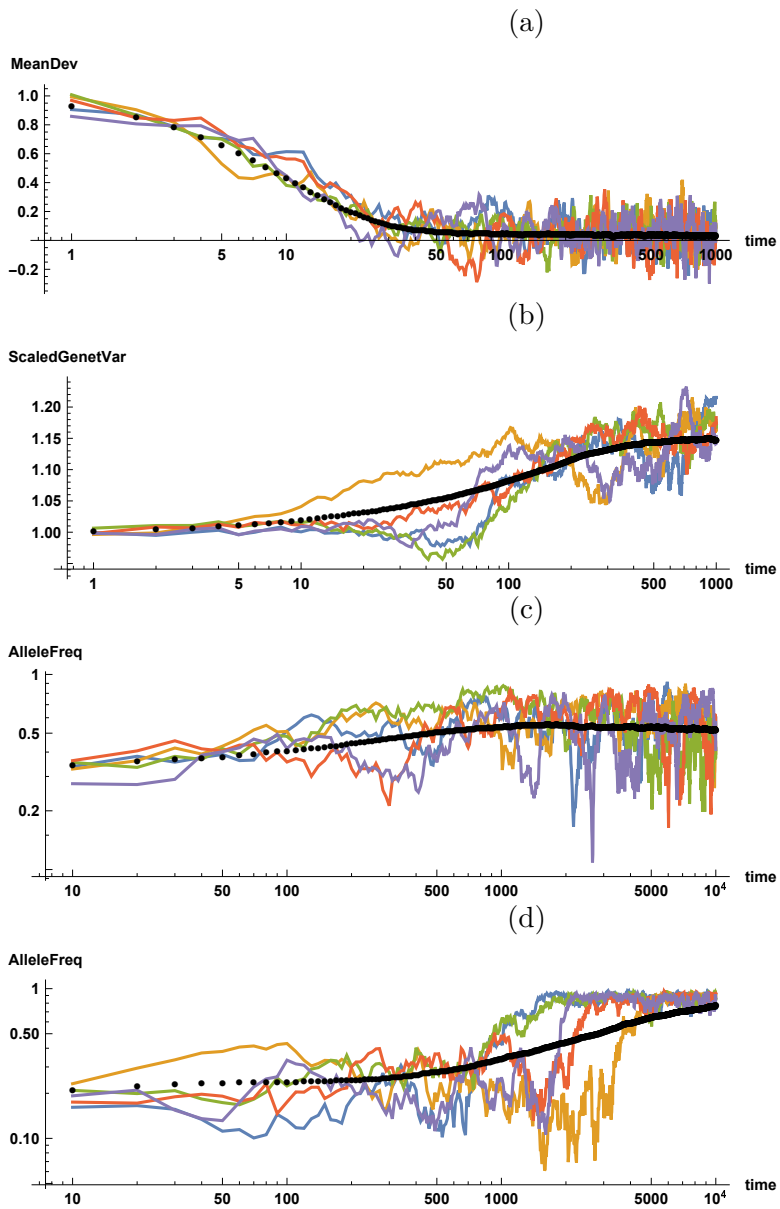

Figure 5: Stochastic trajectories of (a) magnitude of deviation in the trait mean, $|z(t) - z_{\rm f}|$, (b) scaled genetic variance $\upsilon(t)/\upsilon^*$, (c) allele frequency of a small-effect locus and (d) large-effect locus, keeping the initial frequencies of all loci fixed. The data are obtained by solving the Langevin equation (9) numerically and the solid lines in each plot show the data averaged over 10^3 independent stochastic runs. The set of effects and the parameters are the same as in Fig. 4 except that $N = 1000$ and therefore the threshold effect $\hat{\gamma}_N(\ell) \approx 0.537$. The effect size of small- and large-effect locus is, respectively, ≈ 0.36 and 0.8. For the set of effects used here, $\kappa_2 \approx 0.616$ and (C.6) predicts the stationary state average trait mean deviation to be ≈ 0.016; the initial average genetic variance is 1.75 while the stationary state genetic variance $\upsilon^* \approx 2.1$.

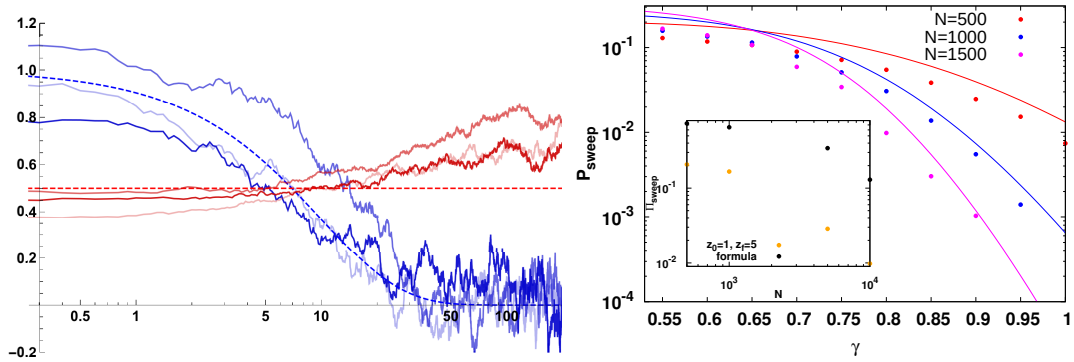

Figure 6: Left: Stochastic trajectories of the allele frequency (red) of a large-effect locus with effect size 0.7 that cross a frequency one half when the trait mean deviation $|z(t) - z_{\rm f}|$ (blue) is $(s\mathfrak{v}^*)^{-1} \approx (s\ell\bar{\gamma}^2)^{-1}$ for $N = 1000$, $s = 0.05$, $\ell = 200$, $\mu = 0.002$, $\bar{\gamma} = 0.1$, $\hat{\gamma}_N = 0.537$, $z_0 = 0$, and $z_{\rm f} = 1$. The blue dashed line shows the trait mean deviation (24) in the deterministic model and the red dashed line is at an allele frequency equal to one half. Main of Right: Probability of sweep as a function of effect size ($> \hat{\gamma}_N(\ell)$) for three population sizes; the rest of the parameters are the same as in the left figure. The points are obtained numerically from 10^4 independent stochastic runs in which the initial allele frequency was below one half and the lines show (30). In this plot, there is only one large-effect locus and the effect sizes of all $\ell - 1$ small-effect loci are kept fixed while the effect of the large-effect is varied. The large-effect locus sweeps when its initial frequency exceeds $x_c \approx 0.42, 0.4, 0.39, 0.38$ for $\gamma = 0.6, 0.7, 0.8, 0.9$, respectively. Inset of Right: Probability of sweep as a function of population size N when several loci have large effect. The parameters are $s = 0.05$, $\ell = 200$, $\mu = 0.002$, $\bar{\gamma} = 0.15$, $\hat{\gamma} = 0.56$, $z_0 = 1$, and $z_{\rm f} = 5$. The orange points represent the numerical results when at least one large effect locus out of 4 large-effect loci crosses 1/2 at $\tau \approx 1/(s\ell\bar{\gamma}^2)$ and the black points show (34). Here, the effect sizes are kept fixed for all the population sizes and the data are obtained using 10^4 independent runs.

File S1

Polygenic adaptation dynamics in large, finite populations

Supporting Information

Archana Devi and Kavita Jain

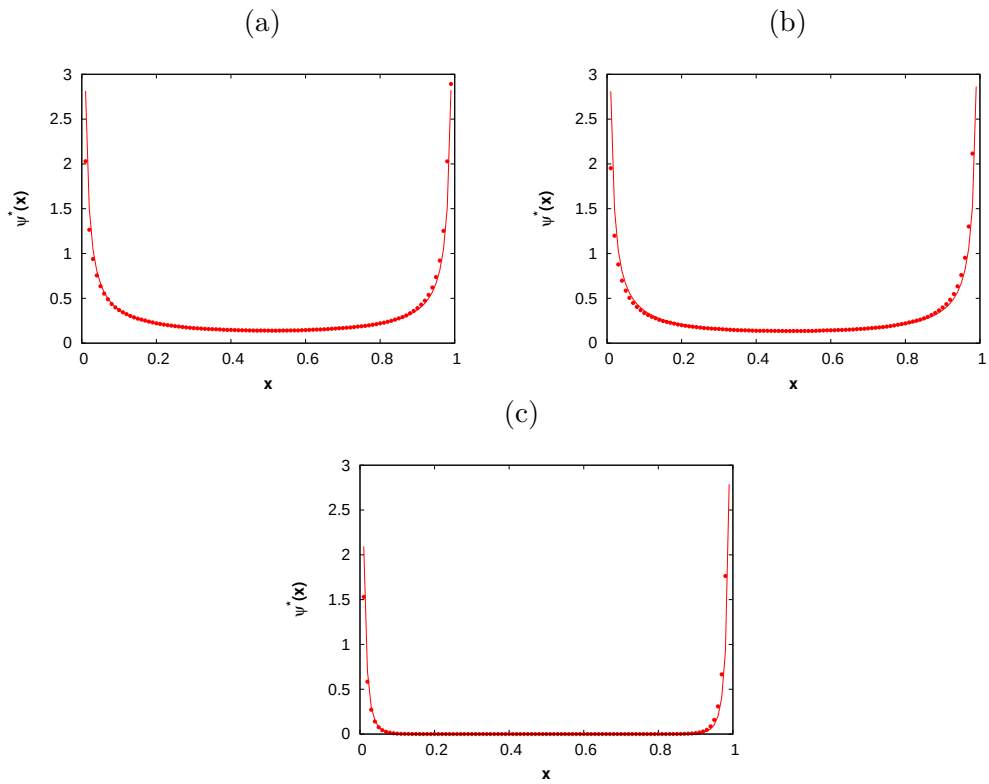

Figure S1: Stationary state marginal allele frequency distribution when the mutation rate is small for a locus with effect size (a) $\gamma_i = 0.01$, (b) $\gamma_i = 0.05$ and (c) $\gamma_i = 0.7$. The parameters are $N = 1000$, $s = 0.05$, $\mu = 0.00002, \ell = 200$, $\bar{\gamma} = 0.1$, and $z_0 = 1$. The deterministic threshold size $\hat{\gamma} = 0.056$ for these parameters whereas stochastically, there is no threshold effect. The points are obtained by numerically solving the Langevin equation (9) numerically by averaging over 10^3 independently chosen initial conditions and until $t = 10^5$ for each initial condition after a burn-in period of $t = 4 \times 10^4$. The red solid line represents the analytical expression (12) where $\kappa_2 \approx 4.8$ is calculated from the effect sizes used in this plot.

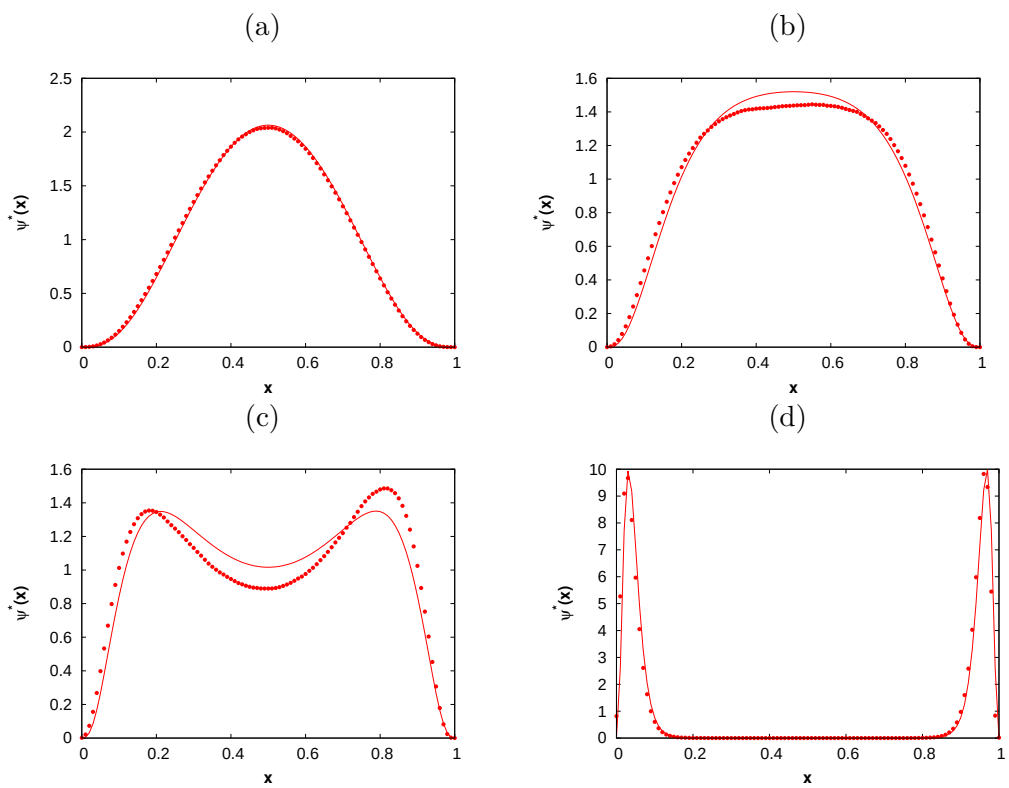

Figure S2: Stationary state marginal allele frequency distribution for a locus with effect size (a) smaller ($\gamma_i = 0.1$), (b) just below ($\gamma_i = 0.23$), (c) just above ($\gamma_i = 0.3$) and (d) larger ($\gamma_i = 0.7$) than the threshold size $\hat{\gamma}_N(\ell) = 0.24$. The parameters are $N = 1000$, $s = 0.1$, $\mu = 0.001, \ell = 200$, $\bar{\gamma} = 0.7$, and $z_0 = 1$. For the set of effects used here, there were 137 large-effects loci. The points are obtained by numerically solving the Langevin equation (9) numerically by averaging over 10^4 independently chosen initial conditions and until $t = 10^5$ for each initial condition after a burn-in period of $t = 4 \times 10^4$. The red solid line represents the analytical expression (12) where $\kappa_2 \approx 235$ is calculated from the effect sizes used in this plot.

S1 Stationary state moments

Here, we numerically study the stationary state cumulants of the allele frequency at a locus; in particular, we study the first three central moments, *viz.*, mean, variance and skewness (which is a measure of the asymmetry of the distribution). These central moments are calculated using the marginal stationary distribution $\psi^*(x_i)$ given by (12).

Figure S3a shows the mean allele frequency for various phenotypic optimum for small- and large-effects. When the phenotypic optimum is at zero, as the marginal distribution (12) is symmetric about one-half, the mean allele frequency is equal to 1/2 for all loci. But for nonzero z_0, the average allele frequency deviates from one-half to push the average trait mean towards the optimum which introduces asymmetry in the marginal distribution of the allele frequency. For $z_0 > 0 (< 0)$, the distribution for a large-effect locus is skewed with the mode at allele frequency $> 1/2$ carrying more (less) weight than the mode at allele frequency $< 1/2$.

When the phenotypic optimum is at zero and the effect size is large, as shown in Fig. S3b, the variance approaches one-quarter with increasing effect size; this is because for $\gamma_i \gg \hat{\gamma}_N(\ell)$, the mutational pressure is low and therefore the distribution is close to zero or one with equal weight, and may be approximated by $\frac{\delta(x_i)+\delta(1-x_i)}{2}$ which immediately leads to the observed result. In contrast, for nonzero z_0, as the bimodal distribution is dominated by the peak at allele frequency close to one with relatively smaller weight close to

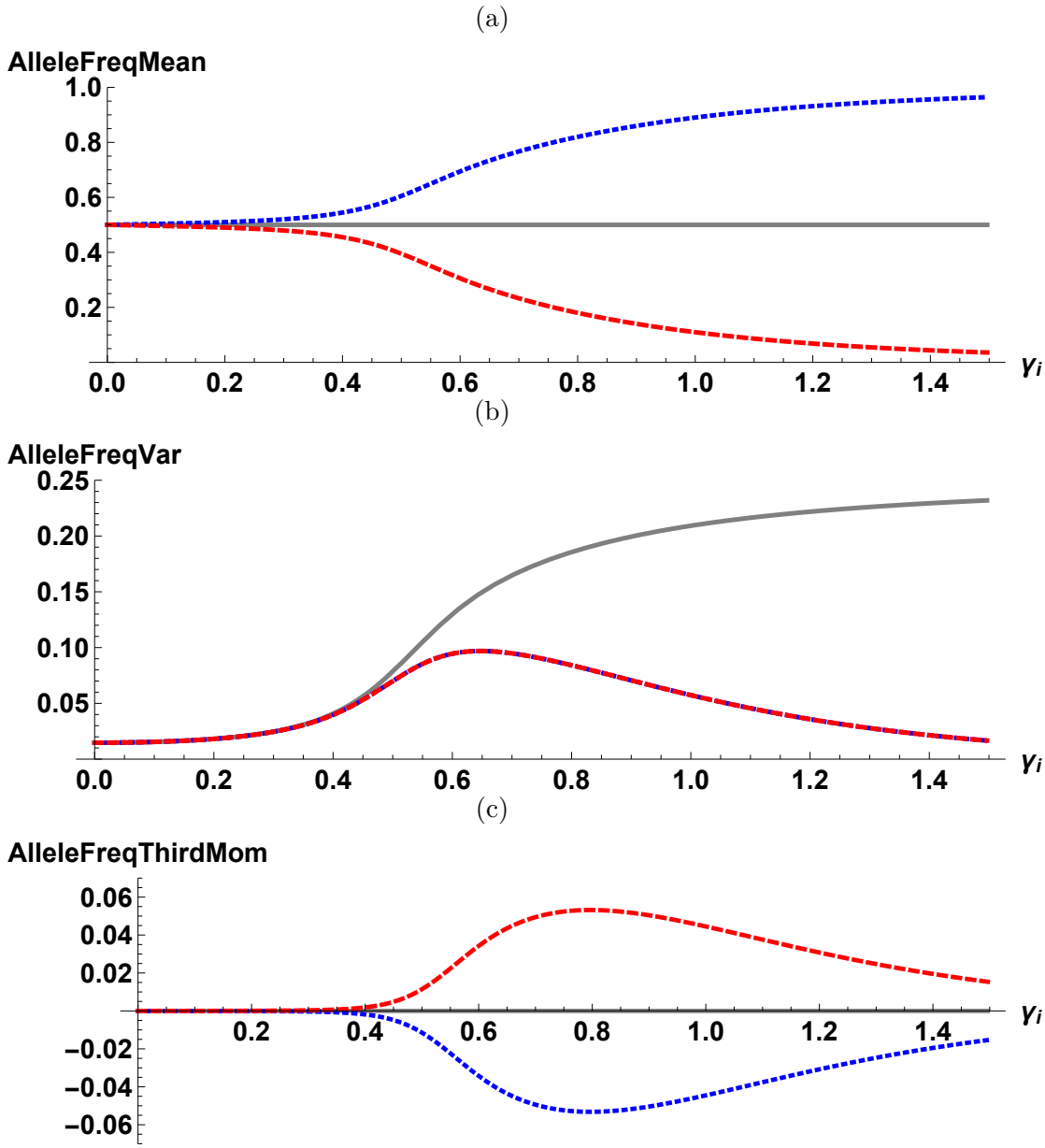

Figure S3: First three central moments of the allele frequency obtained using (12) for $z_0 = 0$ (gray), 2 (blue), -2 (red), and $N = 1000, \mu = 0.002, s = 0.1, \bar{\gamma} = 0.08, \ell = 1000$. The stochastic threshold effect is ≈ 0.416 for these parameters.

frequency zero, the allele frequency variance decreases with increasing effect size for effects larger than the threshold effect.

Similarly, for nonzero z_0 and effect sizes above the threshold effect, due to the asymmetry in the marginal distribution, the skewness (here measured by the third central moment) is nonzero. For positive z_0, as the mean allele frequency is close to the higher peak, the contribution from the lower peak results in negative third moment. The case of negative z_0 can be understood by a similar argument.

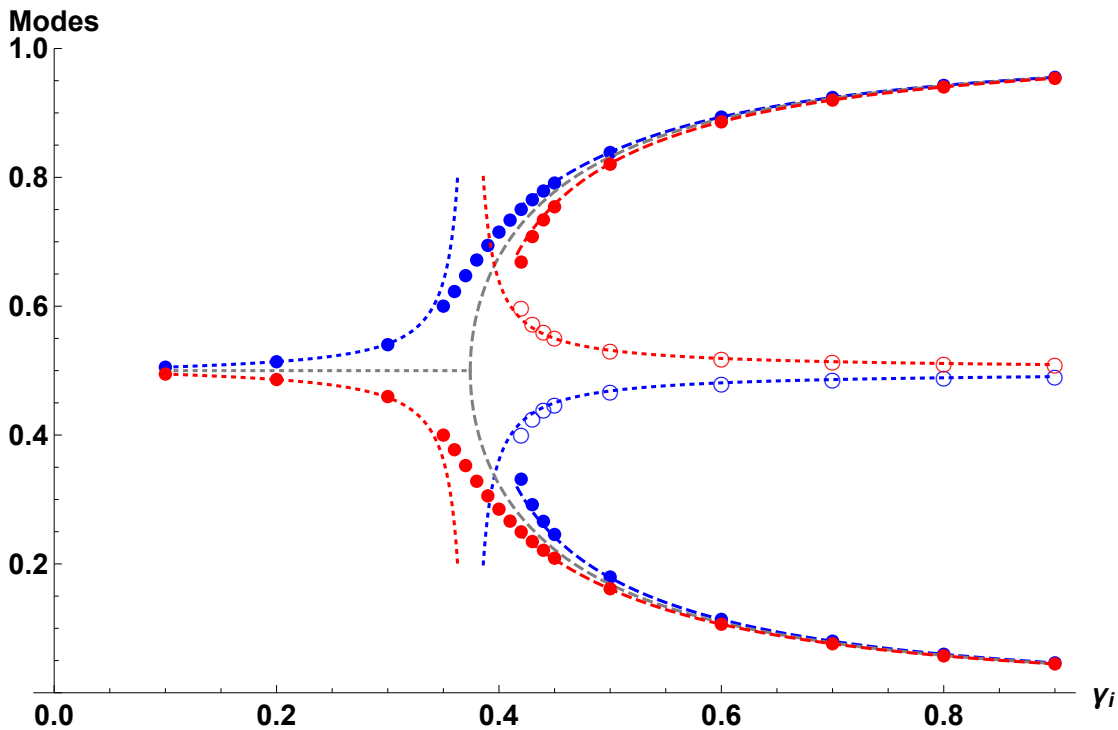

Figure S4: Allele frequency at which the modes in the stationary state marginal allele frequency distribution occur for $z_{\rm f} = 2$ (blue) and -2 (red). The points are obtained by numerically solving the cubic equation (B.2) for the modes with closed (open) symbols denoting the maximum (minimum). The dotted and dashed lines, respectively, show the approximate expressions in (B.4a) and (B.4b) for the mode frequency for finite number of loci while the gray lines show (16) for infinite number of loci. The parameters are the same as in Fig. 2, viz., $N = 1000, s = 0.1, \mu = 0.002, \bar{\gamma} = 0.08$, and $\ell = 1000$. The threshold effect $\hat{\gamma}_N(\ell)$ for finite loci obtained from the cubic equation (B.3) and the approximation (B.6) are, respectively, 0.416 and 0.401, while (15) yields $\hat{\gamma}_N(\ell \to \infty) = 0.374$.

S2 Stationary state genetic variance

We consider the genetic variance (7) when averaged over the stationary state distribution (12). For $z_0 \ll \ell$, using (A.6), we can write

$$\begin{aligned}
\langle x_i(1-x_i)\rangle &= \int_0^1 dx_i x_i(1-x_i)\psi^*(x_i) && \text{(S2.1)}\\
&\approx \frac{\int_0^1 dx_i x_i(1-x_i)g^*(x_i)\left(1-\frac{(\gamma_i(2x_i-1)-z_0)^2}{2(\kappa_2+\frac{\kappa_4}{8\kappa_2})}+...\right)}{\int_0^1 dx_i g^*(x_i)\left(1-\frac{(\gamma_i(2x_i-1)-z_0)^2}{2(\kappa_2+\frac{\kappa_4}{8\kappa_2})}+...\right)} && \text{(S2.2)}\\
&= \frac{\int_0^1 dx_i x_i(1-x_i)g^*(x_i)}{\int_0^1 dx_i g^*(x_i)}\\
&+ \frac{1}{2(\kappa_2+\frac{\kappa_4}{8\kappa_2})}\frac{\int_0^1 dx_i x_i(1-x_i)g^*(x_i)\int_0^1 dx_i g^*(x_i)(\gamma_i^2(2x_i-1)^2+z_0^2)}{[\int_0^1 dx_i g^*(x_i)]^2}\\
&- \frac{1}{2(\kappa_2+\frac{\kappa_4}{8\kappa_2})}\frac{\int_0^1 dx_i x_i(1-x_i)g^*(x_i)(\gamma_i^2(2x_i-1)^2+z_0^2)}{\int_0^1 dx_i g^*(x_i)} && \text{(S2.3)}
\end{aligned}$$

which, on summing over all loci, gives

$$\begin{aligned}
\frac{\langle \upsilon^*\rangle}{\ell} &= \frac{U}{1+2U}\int_0^\infty d\gamma\gamma^2 p(\gamma)\frac{\mathcal{F}(\frac{3}{2},\gamma)}{\mathcal{F}(\frac{1}{2},\gamma)}\\
&+ \frac{1}{(\kappa_2+\frac{\kappa_4}{8\kappa_2})}\frac{U}{1+2U}\int_0^\infty d\gamma\gamma^4 p(\gamma)\left(\frac{U+1}{2U+3}\frac{\mathcal{F}(\frac{5}{2},\gamma)}{\mathcal{F}(\frac{1}{2},\gamma)}-\frac{U}{2U+1}\frac{\mathcal{F}^2(\frac{3}{2},\gamma)}{\mathcal{F}^2(\frac{1}{2},\gamma)}\right) && \text{(S2.4)}
\end{aligned}$$

where, for brevity, we have defined $\mathcal{F}(\alpha,\gamma) = {}_1F_1\left(\frac{1}{2}, 4N\mu+\alpha, \frac{Ns\gamma^2}{2}\right)$ and $U = 4N\mu$.

The first term on the RHS of (S2.4) is independent of z_0 and is, in fact, the result obtained in Bulmer (1972) for equal effects and phenotypic optimum

at zero; the second term which is of order ℓ^{-1} is also seen to be independent of the phenotypic optimum. Thus, when $z_0 \ll \ell$, the stationary genetic variance (to $\mathcal{O}(\ell^{-1})$) does not depend on the location of the phenotypic optimum.

From (S2.4), we find that the stationary genetic variance for exponentially-distributed effects is given by

$$\langle \upsilon^* \rangle = \frac{4N\mu\ell\bar{\gamma}^2}{1+8N\mu} \int_0^\infty \frac{{}_1F_1(\frac{1}{2}; 4N\mu + \frac{3}{2}; \frac{Ns\bar{\gamma}^2}{2}x^2)}{{}_1F_1(\frac{1}{2}; 4N\mu + \frac{1}{2}; \frac{Ns\bar{\gamma}^2}{2}x^2)} x^2 e^{-x} dx \tag{S2.5}$$

where we have ignored the second term on the RHS of (S2.4). For fixed $4N\mu$, the above integral can be approximated by

$$\langle \upsilon^* \rangle \approx \begin{cases} \dfrac{8N\mu\ell\bar{\gamma}^2}{1+8N\mu}\left(1 - \dfrac{12Ns\bar{\gamma}^2}{(1+8N\mu)(3+8N\mu)}\right) & , Ns\bar{\gamma}^2 \ll 2 \quad \text{(S2.6a)} \\ \dfrac{4\ell\mu}{s} & , Ns\bar{\gamma}^2 \gg 2 \quad \text{(S2.6b)} \end{cases}$$

Thus the genetic variance is an increasing function of selection strength, and saturates to $4\ell\mu/s$ for strong selection or when most effects are large, as expected from the House-of-Cards model (Bürger, 2000). For finite ℓ, the correction to the genetic variance found above can also be approximated but are not particularly illuminating.

As mentioned above, the stationary genetic variance is independent of z_0 when the phenotypic optimum is small. Figure S5 shows that not only the total genetic variance but the contribution of small- and large-effect loci to $\langle \upsilon^* \rangle$ also depends weakly on the phenotypic optimum when z_0 is large, irrespective of selection strength.

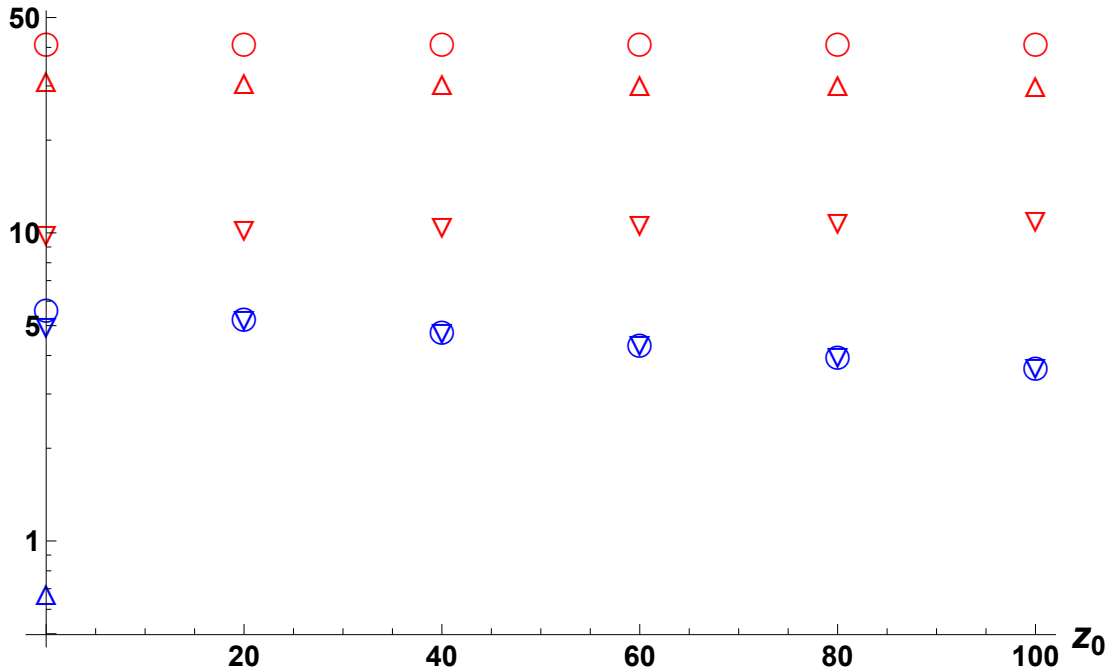

Figure S5: Stationary state genetic variance shown by ○ and obtained using (12) as a function of the phenotypic optimum for $\bar{\gamma} = 0.08$ (blue) and 0.4 (red), and $N = 1000, \mu = 0.002, s = 0.1, \ell = 1000$. The contribution of small- and large-effect loci to the total genetic variance is shown, respectively, by ▽ and △. For weak selection ($Ns\bar{\gamma}^2 = 0.64$), the contribution from large-effect loci is negligible as the threshold effect increased substantially on increasing z_0 (from 0.37 for $z_0 = 0$ to 1.2 for $z_0 = 100$) resulting in a vanishing fraction of large-effect loci contributing to a quantitative trait.

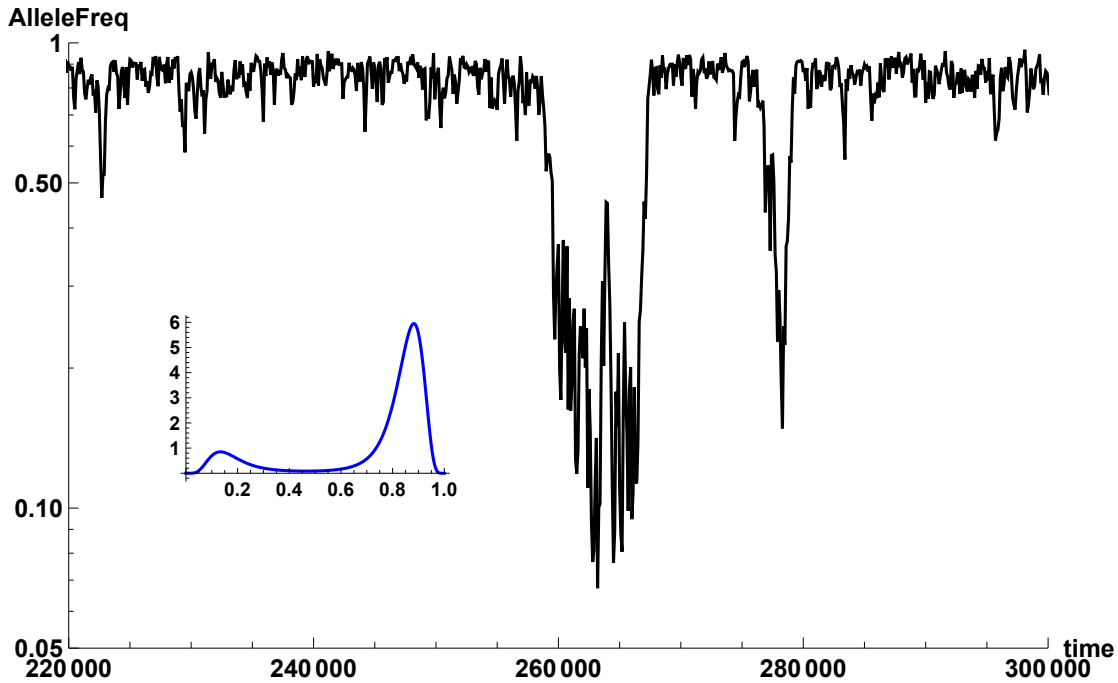

Figure S6: The main figure shows the stationary state dynamics of the allele frequency of a major effect locus while the inset shows the stationary state distribution (12) when the phenotypic optimum is at one. The allele frequency shifts between the two peaks and spends considerable time at either peak. The rest of the parameters are the same as in Fig. 5, *viz.*, $\ell = 200, N = 1000, s = 0.05, \mu = 0.002, \bar{\gamma} = 0.1$.

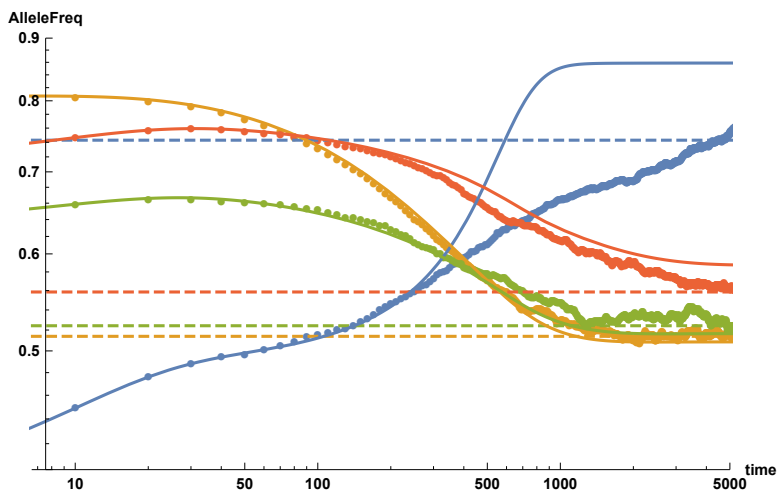

Figure S7: Comparison of deterministic allele frequency (solid lines) obtained by numerically solving (18) and average allele frequency (points) obtained by numerically solving (9) for 5000 independent stochastic runs for for effect size 0.8 (blue), 0.24 (yellow), 0.35 (green), 0.52 (red). The other parameters are $\ell = 200, \bar{\gamma} = 0.1, s = 0.05, \mu = 0.002, N = 1000$. The effect size set used is the same in Fig. 4, but due to larger N, the threshold frequency is $\hat{\gamma}_N(\ell) = 0.54$ and 0.66 for $z_0 = 0$ and $z_{\mathrm{f}} = 1$, respectively.

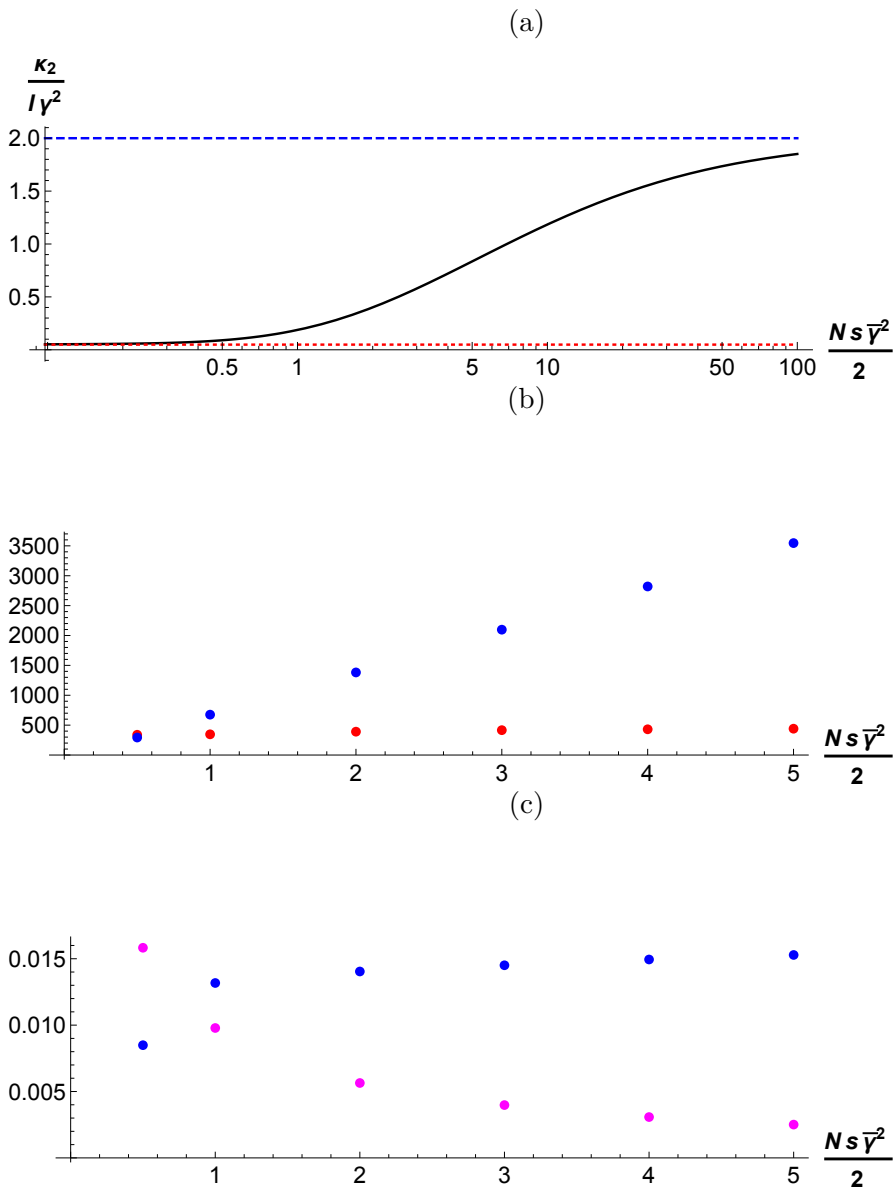

Figure S8: (a) Scaled variance κ_2 given by (A.10) as a function of scaled selection strength for $4N\mu = 20$. The dotted and dashed lines, respectively, show the result $\frac{2}{1+8N\mu}$ for weak selection and 2 for strong selection. (b) In accordance with (A.12a) and (A.12b), the contribution to $2Ns\kappa_2$ from the small-effect loci (red) remains roughly constant while that from the large-effect loci (blue) increases linearly with the population size for $z_0 = 10$, $\ell = 1000, \mu = 0.002, s = 0.1, \bar{\gamma} = 0.1$. The threshold effect $\hat{\gamma}_N(\ell)$ was calculated for each N, as discussed in Appendix B. (c) Average trait mean deviation (magenta) given by (C.6) decreases while the fraction of large-effect loci in a quantitative trait (blue) increases when the population size is increased (keeping all the other parameters same).

S3 Allele frequency distribution within linear noise approximation

Consider the joint distribution $P(\vec{x},t)$ that obeys the following Fokker-Planck equation:

$$\frac{\partial P(\vec{x},t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}(M_i(\vec{x})P(\vec{x},t)) + \frac{\epsilon^2}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}(V_{ij}(\vec{x})P(\vec{x},t)) \quad \text{(S3.7)}$$

On changing the variables from x_i to $\xi_i = \frac{x_i - x_i(t)}{\epsilon}$, we find that the probability distribution $\Pi(\vec{\xi},t) = \epsilon^\ell P(\vec{x},t)$ obeys the following partial differential equation:

$$\begin{aligned}\frac{\partial \Pi(\vec{\xi},t)}{\partial t} &= -\sum_i \frac{\partial}{\partial \xi_i}\left[\frac{M_i(\vec{x}+\epsilon\vec{\xi}) - M_i(\vec{x})}{\epsilon}\Pi(\vec{\xi},t)\right] \\ &+ \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial \xi_i \partial \xi_j}\left[V_{ij}(\vec{x}+\epsilon\vec{\xi})\Pi(\vec{\xi},t)\right] \end{aligned} \quad \text{(S3.8)}$$

where, the deterministic frequency obeys $\frac{dx_i(t)}{dt} = M_i(\vec{x})$.

The distribution and the coefficients are now expanded in a power series

in small parameter ϵ, that is,

$$\Pi(\vec{\xi}, t) = \sum_{n=0}^{\infty} \Pi^{(n)} \epsilon^n \tag{S3.9}$$

$$M_i(\vec{x} + \epsilon\vec{\xi}, t) = \sum_{n=0}^{\infty} A_i^{(n)} (\xi\epsilon)^n \tag{S3.10}$$

$$V_{ij}(\vec{x} + \epsilon\vec{\xi}, t) = \sum_{n=0}^{\infty} B_{ij}^{(n)} (\xi\epsilon)^n \tag{S3.11}$$

Using these expansions in (S3.8) and matching terms with the same order in ϵ on both sides of (S3.8), one can obtain partial differential equations for $\Pi^{(n)}(\vec{\xi}, t)$. Here, we will work only to the leading order in ϵ which results in a linear Fokker-Planck equation (*linear noise approximation*) in the sense that the first sum on the RHS of (S3.8) is linear in ξ and the second sum is constant [refer to Chapter VIII, van Kampen (1997)], and find that $\Pi^{(0)}$ is solution of the following equation (see Chapter X, van Kampen (1997), Sec. 6.3 of Gardiner (1997)),

$$\frac{\partial \Pi^{(0)}(\vec{\xi}, t)}{\partial t} = -\sum_{i,j} A_{ij}^{(1)}(t) \frac{\partial}{\partial \xi_i} (\xi_j \Pi^{(0)}(\vec{\xi}, t)) + \frac{1}{2} \sum_i B_{ij}^{(0)}(t) \frac{\partial^2 \Pi^{(0)}(\vec{\xi}, t)}{\partial \xi_i \partial \xi_j} \tag{S3.12}$$

where

$$A_{ij}^{(1)}(t) = \left. \frac{\partial M_i(\vec{z})}{\partial z_j} \right|_{\vec{z}=\vec{x}} \tag{S3.13}$$

$$B_{ij}^{(0)}(t) = V_{ij}(\vec{x}) \tag{S3.14}$$

Note that the coefficient of the first derivative term on the RHS of (S3.12) depends linearly on ξ_i and that of the second derivative term is constant in ξ_i. Thus, the joint distribution $\Pi(\xi, t)$ is the distribution for the multivariate Ornstein-Uhlenbeck process with time-dependent coefficients which, as described below, can be solved exactly using the method of characteristics (van Kampen, 1997; Gardiner, 1997).

For the diffusion equation of interest here, *viz.*, (2), as B_{ij} is diagonal, we do not consider the cross terms in the second term on the RHS of (S3.12) in the following discussion. Furthermore, we also drop the superscripts in (S3.12). Consider the Fourier transform, $G(\vec{k}, t) = \int_{-\infty}^{\infty} e^{-i\vec{k}.\vec{\xi}} \Pi(\vec{\xi}, t) \prod_i d\xi_i$ which, due to (S3.12), obeys

$$\frac{\partial G(\vec{k}, t)}{\partial t} = \sum_{i,j} A_{ij}(t) k_i \frac{\partial G(\vec{k}, t)}{\partial k_j} - \frac{1}{2} \sum_i B_i(t) k_i^2 G(\vec{k}, t) \tag{S3.15}$$

The characteristic curves obey $\frac{dk_j}{dt} = -\sum_i k_i A_{ij}(t)$ and therefore, $\vec{k}^T = \vec{c}^T \mathbf{U}^T(t, 0)$ where T denotes the transpose of the column vector $\vec{k}$, and the time evolution operator $\mathbf{U}(t, t_0)$ is given by

$$\mathbf{U}(t, t_0) = \mathbf{I} + \sum_{n=1}^{\infty} (-1)^n \int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2 ... \int_{t_0}^{t_{n-1}} dt_n \mathbf{A}^T(t_1) ... \mathbf{A}^T(t_n) \tag{S3.16}$$

$$= \mathbf{I} + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \int_{t_0}^{t} dt_1 ... \int_{t_0}^{t} dt_n \mathcal{T} (\mathbf{A}^T(t_1) ... \mathbf{A}^T(t_n)) \tag{S3.17}$$

In the above equation, $\mathcal{T}$ denotes the time-ordered product of matrices, and is required when the matrix $\mathbf{A}$ at different times do not commute: $\mathbf{A}(t_1)\mathbf{A}(t_2) \neq$

$\mathbf{A}(t_2)\mathbf{A}(t_1)$. Along the characteristic curves, the solution of (S3.15) is

$$\begin{aligned} G(\vec{k},t) &= \exp[-\frac{1}{2}\int_0^t \vec{k}^T\ \mathbf{U}^T(t',t)\mathbf{B}(t')\mathbf{U}(t',t)\vec{k}dt'] \\ &\times\ C(\mathbf{U}^{-1}(t,0)\vec{k}) \end{aligned} \tag{S3.18}$$

where the function C is determined by the initial condition $G(\vec{k},0)$.

For the initial joint distribution, $\Pi(\vec{\xi},0) = \prod_i \delta(\xi_i - \xi_i(0))$, we finally obtain

$$G(\vec{k},t) = \exp\left[-i\vec{\xi}^T(0)\mathbf{U}^{-1}(t,0)\vec{k} - \frac{1}{2}\int_0^t \vec{k}^T\ \mathbf{U}^T(t',t)\mathbf{B}(t')\mathbf{U}(t',t)\vec{k}dt'\right] \tag{S3.19}$$

Since $G(\vec{k},t)$ is a Gaussian function, using the fact that the inverse Fourier transform of a Gaussian is also a Gaussian, we conclude that for a given set of initial allele frequencies, the random variables $\{\xi_i\}$ at time t are Gaussian-distributed.

The time-dependent marginal distribution at the jth locus can be obtained from the Fourier transform:

$$\pi(\xi_j,t|\xi_j(0)) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk_j e^{ik_j\xi_j} G(0,...,0,k_j,0,...0,t) \tag{S3.20}$$

where, due to (S3.19),

$$G(0,...,0,k_j,0,...0,t) = \exp\left[-ik_jL_j(t) - \frac{k_j^2}{2}\int_0^t \mathbf{Q}_{jj}(t,t')dt'\right] \tag{S3.21}$$

and

$$\vec{L}^T(t) = \vec{\xi}^T(0)\mathbf{U}^{-1}(t,0) \tag{S3.22}$$

$$\mathbf{Q}(t,t') = \mathbf{U}^T(t',t)\mathbf{B}(t')\mathbf{U}(t',t) \tag{S3.23}$$

We thus have

$$\pi(\xi_j, t|\xi_j(0)) = \frac{1}{\sqrt{2\pi \int_0^t Q_{jj}(t,t')dt'}} \exp\left[-\frac{(\xi_j - L_j(t))^2}{2\int_0^t Q_{jj}(t,t')dt'}\right] \tag{S3.24}$$

From the above equation, it follows that the time-dependent marginal distribution of the allele frequency at the ith locus is also a Gaussian, and is given by

$$\psi(x_i, t|x_i(0)) = \sqrt{\frac{N}{\pi \int_0^t \mathbf{Q}_{ii}(t,t')dt'}} \exp\left[-\frac{N(x_i - \boldsymbol{x}_i(t) - \mathscr{L}_i(t))^2}{\int_0^t \mathbf{Q}_{ii}(t,t')dt'}\right] \tag{S3.25}$$

where

$$\vec{\mathscr{L}}^T(t) = (\vec{x}^T(0) - \vec{\boldsymbol{x}}^T(0))\mathbf{U}^{-1}(t,0) \tag{S3.26}$$

$$\mathbf{Q}(t,t') = \mathbf{U}^T(t',t)\mathbf{B}(t')\mathbf{U}(t',t) \tag{S3.27}$$

and $\mathbf{U}$ is given by (S3.17).

S4 Corrections to the central limit theorem: Edgeworth expansion

Consider the distribution of the random variable, $X = \frac{\sum_{j=1}^{\ell} \gamma_j (2x_j - 1)}{\sqrt{\kappa_2}}$ where x_i's are independently distributed according to the (normalized) distribution, $g^*(x_i)$ and $\kappa_2 = \sum_{j=1}^{\ell} \langle \gamma_j^2 (2x_j - 1)^2 \rangle_{g^*}$ where we have used that the mean $\langle \gamma_j (2x_j - 1) \rangle_{g^*} = 0$. Then, from (43) of Blinnikov and Moessner (1998), we have

$$\mathrm{Prob}(X) \propto e^{-\frac{X^2}{2}} \left[1 + \frac{\kappa_4}{\kappa_2^2} \frac{3 - 6X^2 + X^4}{24} + \mathcal{O}(\ell^{-2}) \right] \tag{S4.28}$$

where, $\kappa_4 = \sum_{j=1}^{\ell} \langle \gamma_j^4 (2x_j - 1)^4 \rangle_{g^*} - 3 \langle \gamma_j^2 (2x_j - 1)^2 \rangle_{g^*}^2$. Using the above result in (A.3) for $X' = X - \gamma_i (2x_i - 1)$ and performing the integral over X', we arrive at (A.6) in the main text. From the marginal distribution (A.6) for $z_0 \ll \ell$, we then obtain

$$\gamma_i \langle 2x_i - 1 \rangle \approx \frac{z_0 K_{2,i}}{\kappa_2} + \frac{N s z_0 (K_{2,i} \kappa_4 - \kappa_2 K_{4,i}) - z_0 \kappa_2 K_{2,i}}{2 N s \kappa_2^3} + \mathcal{O}(\ell^{-3}) \tag{S4.29}$$

where $K_{2,j} = \langle \gamma_j^2 (2x_j - 1)^2 \rangle_{g^*}, K_{4,j} = \langle \gamma_j^4 (2x_j - 1)^4 \rangle_{g^*} - 3 \langle \gamma_j^2 (2x_j - 1)^2 \rangle_{g^*}^2$. We check that, on summing over all loci, the above equation leads to (C.6).

Similarly, for the distribution of the trait mean, we have

$$
\begin{aligned}
Pr(z^*) &= \int_0^1 dx_1 ... \int_0^1 dx_\ell P(\vec{x})\delta(z - z^*) && \text{(S4.30)} \\
&\propto e^{-Ns(z^*-z_0)^2} \int_0^1 \prod_{i=1}^{\ell} dx_i g(x_i^*)\delta(z^* - \sum_i \gamma_i(2x_i - 1)) && \text{(S4.31)} \\
&\propto e^{-Ns(z^*-z_0)^2} \int_0^1 \prod_{i=1}^{\ell} dx_i g(x_i^*)\delta\left(\frac{z^*}{\sqrt{\kappa_2}} - \frac{\sum_i \gamma_i(2x_i - 1)}{\sqrt{\kappa_2}}\right) && \text{(S4.32)}
\end{aligned}
$$

On using the central limit theorem and its extension given by (S4.28), we then obtain

$$
Pr(z^*) \propto e^{-Ns(z^*-z_0)^2} e^{-\frac{z^{*2}}{2\kappa_2}} \left[1 + \frac{\kappa_4}{24\kappa_2^2}\left(3 - \frac{6z^{*2}}{\kappa_2} + \frac{z^{*4}}{\kappa_2^2}\right)\right] \quad \text{(S4.33)}
$$

where the second term in the bracket on the RHS is of order ℓ^{-1}. The above distribution gives the average and variance of the trait mean to be the same as (C.4) and (C.5) to order ℓ^{-1} and therefore the correction term does not affect the results for the trait mean.

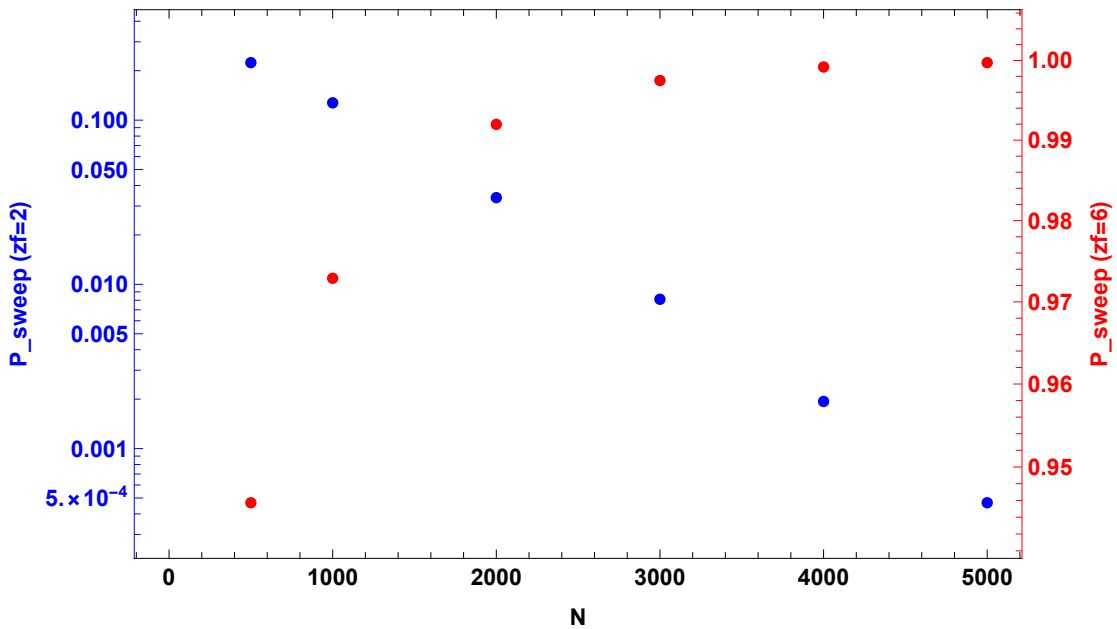

Figure S9: Sweep probability for $z_{\mathrm{f}} = 2$ (blue) and $z_{\mathrm{f}} = 6$ (red) to show the different qualitative dependence on N. The points are generated using (30) for $N = 1000$, $s = 0.05$, $\ell = 200$, $\mu = 0.002$, $\bar{\gamma} = 0.1, \gamma = 0.8, z_0 = 0$; for these parameters, the major locus frequency in an infinite population sweeps if $z_{\mathrm{f}} > 4.4$.

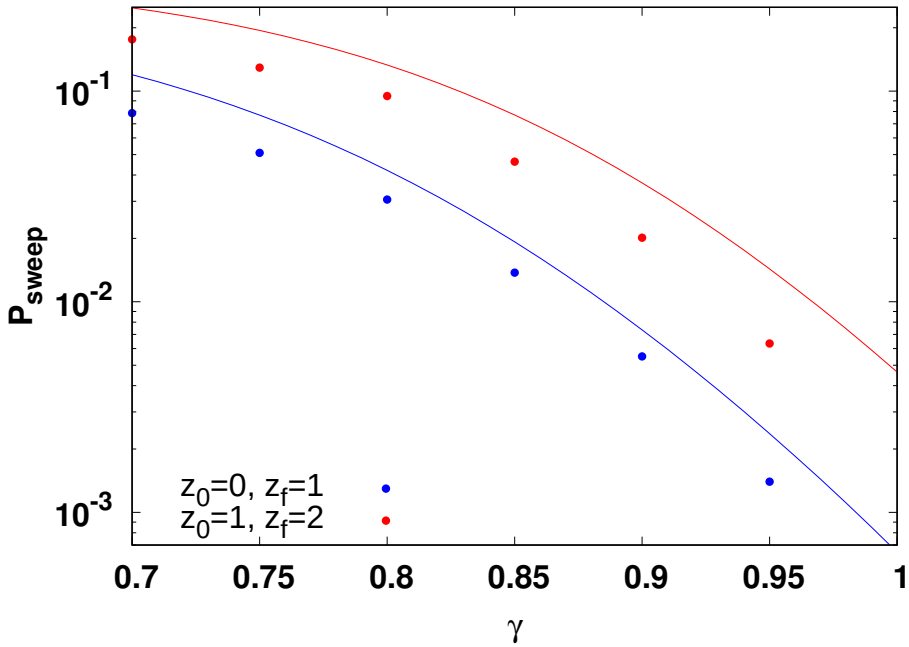

Figure S10: Probability of sweep for different initial and final optimum values keeping $z_f - z_0$ same to show that it depends on z_0 and z_f for small ℓ. This figure uses the same set of effect sizes and same parameter values as in the Fig. 6 (only the initial optimum has been changed for the red points and line). The lines and points are obtained in the similar manner as in the Fig. 6 of main text and the data are obtained using 10^4 independent runs. In case of $z_0 = 1$, the large effect locus sweeps when it exceeds a critical frequency $x_c = 0.41, 0.4, 0.39$ for $\gamma = 0.7, 0.8, 0.9$, respectively. The corresponding values for $z_0 = 0$ are given in Fig. 6 of main text.

References

Blinnikov, S. and R. Moessner, 1998 Expansions for nearly Gaussian distributions. Astron. Astrophys. Suppl. Ser. **130:** 193–205.

Bulmer, M. G., 1972 The genetic variability of polygenic characters under optimizing selection, mutation and drift. Genet. Res., Camb. **19:** 17–25.

Bürger, R., 2000 *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, Chichester.

Gardiner, C. W., 1997 *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Springer-Verlag.

van Kampen, N. G., 1997 *Stochastic processes in physics and chemistry*. North Holland Personal Library.