Polygenic dynamics underlying the response of quantitative traits to directional selection

Hannah Götsch[a,b,*], Reinhard Bürger[a]

[a] *Faculty of Mathematics, University of Vienna, 1090 Vienna, Austria*
[b] *Vienna Graduate School of Population Genetics, Austria*

Abstract

We study the response of a quantitative trait to exponential directional selection in a finite haploid population, both at the genetic and the phenotypic level. We assume an infinite sites model, in which the number of new mutations per generation in the population follows a Poisson distribution (with mean Θ) and each mutation occurs at a new, previously monomorphic site. Mutation effects are beneficial and drawn from a distribution. Sites are unlinked and contribute additively to the trait. Assuming that selection is stronger than random genetic drift, we model the initial phase of the dynamics by a supercritical Galton-Watson process. This enables us to obtain time-dependent results. We show that the copy-number distribution of the mutant in generation n, conditioned on non-extinction until n, is described accurately by the deterministic increase from an initial distribution with mean 1. This distribution is related to the absolutely continuous part W^+ of the random variable, typically denoted W, that characterizes the stochasticity accumulating during the mutant's sweep. A suitable transformation yields the approximate dynamics of the mutant frequency distribution in a Wright-Fisher population of size N. Our expression provides a very accurate approximation except when mutant frequencies are close to 1. On this basis, we derive explicitly the (approximate) time dependence of the expected mean and variance of the trait and of the expected number of segregating sites. Unexpectedly, we obtain highly accurate approximations for all times, even for the quasi-stationary phase when the expected per-generation response and the trait variance have equilibrated. The latter refine classical results. In addition, we find that Θ is the main determinant of the pattern of adaptation at the genetic level, i.e., whether the initial allele-frequency dynamics are best described by sweep-like patterns at few loci or small allele-frequency shifts at many. The number of segregating sites is an appropriate indicator for these patterns. The selection strength determines primarily the rate of adaptation. The accuracy of our results is tested by comprehensive simulations in a Wright-Fisher framework. We argue that our results apply to more complex forms of directional selection.

*Corresponding author
Email address: hannah.goetsch@univie.ac.at (Hannah Götsch)

Keywords: polygenic adaptation, mutation, infinite-sites model, Galton-Watson branching process, Wright-Fisher model

1. Introduction

Phenotypic adaptation is ubiquitous and can occur rapidly as a response to artificial selection or to a gradual or sudden change in the environment. On the basis of accessible phenotypic measurements, the response of the mean phenotype can often be predicted or described by simple equations, such as the breeder's equation or its evolutionary counterpart, Lande's (1976) equation (Falconer and Mackay, 1996, Walsh and Lynch, 2018). During this process genotype frequencies change, and new genotypes and new alleles can be generated by recombination and mutation. Even if the response of the mean phenotype is simple to describe, it is a highly complex, often insurmountable, task to infer the response on the genetic level.

There are multiple reasons for this problem. The response of a trait may be due to small frequency changes at a large number of loci of minor effect (Hill and Kirkpatrick, 2010, Barghi et al., 2019, Boyle et al., 2017), as originally proposed by Fisher (1918) and idealized and formalized in the infinitesimal model (Bulmer, 1980, Barton et al., 2017). The selection response may also be due to selective sweeps at a few loci of large effect (Kaplan et al., 1989, Hermisson and Pennings, 2005, Messer and Petrov, 2013, Stephan, 2019). Finally, a mixture of both is a third realistic scenario (Lande, 1983, Chevin and Hospital, 2008, Chevin, 2019). In principle, different genetic architectures in combination with different initial conditions can lead to similar phenotypic responses, at least as far as the mean phenotype is concerned. Even if loci contribute additively to the trait and their number and effects are known, it is notoriously difficult to derive the allele-frequency dynamics at individual loci and the resulting dynamics of the trait's mean and variance from the selection pressure acting on the trait. Additional complications arise if linkage disequilibrium, epistasis, or pleiotropy have to be taken into account (Turelli and Barton, 1990, Bürger, 2000, Barton and Keightley, 2002, Zhang and Hill, 2005a, Walsh and Lynch, 2018).

For quite some time the focus in studying adaptation at the genetic and molecular level has been on selective sweeps, first hard sweeps and more recently soft sweeps which can occur from standing variation or immigration and are not restricted to mutation-limited evolutionary scenarios (Hermisson and Pennings, 2017). Although footprints of hard and soft selective sweeps are detectable in genetic data by various methods (Nielsen et al., 2005, Huber et al., 2016, Setter et al., 2020), they seem to be sparse in the data, especially hard sweeps in humans (Pritchard et al., 2010, Hernandez et al., 2011). The detection of polygenic responses by small allele-frequency changes at loci of minor effect is much more demanding. It became feasible at a larger scale

only when sequencing became cheaper (Berg and Coop, 2014). Genome-wide association studies (GWASs) became an important tool in identifying associations between genetic variants, typically single-nucleotide polymorphisms (SNPs), and phenotypic traits. Although GWASs are highly valuable in elucidating the genetic architecture underlying phenotypic traits, they provide no direct information on the contribution of the causal variants to adaptation of the trait, which can be deleterious, neutral or advantageous. More refined approaches, such as experimental evolution methods that yield time-series data, have been pursued successfully to demonstrate truly polygenic adaptation (reviewed in Sella and Barton, 2019, Barghi et al., 2020).

Theoreticians only recently started to explore when adaptation of polygenic traits is likely to be based on selective sweeps at a few loci vs. on small polygenic allele-frequency changes. Essential for this line of research is a good understanding of which genetic architectures and parameters (number and effects of loci, distribution of mutation effects, linkage, epistasis, pleiotropy) as well as population sizes and forms and strengths of selection can cause which adaptive patterns. By this we mean sweep patterns (hard or soft), polygenic patterns, or a mixture of both.

Some studies focused on polygenic traits under stabilizing selection, where the trait mean is close to the optimum or at most a few phenotypic standard deviations away and the selective response is due to standing genetic variation established previously by mutation-selection-(drift) balance. The findings are qualitatively similar, whether the population size is so large that random genetic drift can be ignored (De Vladar and Barton, 2014, Jain and Stephan, 2017a) or moderately large so that it plays a role but is weaker than selection (John and Stephan, 2020, Hayward and Sella, 2022). Roughly, if most standing variation is due to many polymorphic loci of small effect, then the response of the mean is mainly caused by small frequency changes. Only under exceptional circumstances do selective sweeps occur and drive the response. This result is not unexpected because (i) loci of small effect are more like to carry alleles of intermediate frequency and (ii) close to the phenotypic optimum fine tuning at the contributing loci is important.

Other studies focused on polygenic adaptation under directional selection (Höllinger et al., 2019), for instance caused by a sudden big environmental shift (Thornton, 2019, Höllinger et al., 2023). Although the assumptions on the genetic architecture and on the form of directional selection differ among these studies, they show that selective sweeps at several loci (parallel or successive) occur more often than in the above discussed scenario. The main determinant of the adaptive pattern at the genomic level is the so-called population-wide background mutation rate, but not the strength of selection or the population size (Höllinger et al., 2019).

In this study, we also explore how various evolutionary and genetic factors determine the rate and pattern of polygenic adaptation, both at the phenotypic and the genetic level. However, our assumptions depart in several respects from those imposed in previous studies. First, we

consider a trait subject to exponential directional selection, which is nonepistatic, does not induce linkage disequilibrium, and leads to a constant selection strength s. Second, in contrast to the investigations cited above, we focus on the response due to *de novo* mutations, i.e., we ignore standing variation. Third, but similar to Thornton (2019), we assume an infinite sites model. The number of mutations per generation follows a Poisson distribution and each mutation occurs on a new, previously monomorphic locus. Loci are unlinked and contribute additively to the trait. Mutation effects are beneficial and follow a general distribution. This is justified by the assumption that in our population of size N selection is stronger than random genetic drift ($Ns > 1$), so that prolonged segregation or fixation of non-advantageous mutations can be ignored. These assumptions enable us to derive analytical expressions for the time dependence of the mutant frequency distribution at each locus, of the expected number of segregating sites, and of the expected mean and variance of the trait.

The paper is structured as follows. In Section 2, we introduce our basic model and give an overview of the analytical concepts regarding the one-locus model and its extension to infinitely many sites. Also our simulation setting is described. In Section 3, we derive an explicit, approximate expression for the mutant frequency distribution in a finite Wright-Fisher population as a function of time. This becomes feasible by using a supercritical Galton-Watson process with a quite general offspring distribution to describe the spreading of a beneficial mutant. Comparison with simulation results of a corresponding Wright-Fisher model demonstrates that our approximation is highly accurate for a wide range of parameters and a large time range.

Key results are explicit and computationally efficient time-dependent expressions for the expected mean and the expected variance of an additive quantitative trait under exponential selection. They are presented in Section 4, seem to be entirely new, and provide highly accurate approximations to corresponding Wright-Fisher simulations. Interestingly, they even allow the derivation of expressions for the long-term, quasi-stationary response of the trait's mean and variance. They are not only derived from first principles by assuming the infinite sites model but also recover and refine classical results. Proofs are given in Appendix D. In Section 5.1 (and Appendix E), we derive explicit, approximate expressions for the evolution of the expected number of segregating sites. They are based on a combination of branching process methods with (in part new) approximations for the expected time to loss or to fixation of a new beneficial mutant. The latter are deduced and numerically tested in Appendix B. In Section 5.2, we use the approximation for the number of segregating sites to characterize the numerically determined initial response patterns. This allows us to examine the genomic patterns associated with the early phase of phenotypic adaptation.

Because this paper is rather technical, we summarize, explain, and discuss our findings comprehensively in Section 6. A central quantity in our theory is the expected number Θ of new

beneficial mutations occurring per generation in the total population. We provide quantitative support to previous findings in related but different contexts that sweep-like patterns will occur if Θ is sufficiently much smaller than 1, whereas polygenic shifts will drive the initial selection response if Θ is much greater than 1. Other model parameters have only a minor influence on the initial adaptive pattern, but may have a major influence on the rate of response of the trait. We propose to use the expected number of segregating sites as an indicator for distinguishing between sweep-like vs. polygenic shift-like patterns in the early phases of adaptation. In Section 6.4, we discuss the applicability of our approach to more general selection scenarios and to more general genetic architectures as well as the limitations inherent in our model assumptions.

2. The model

We assume a haploid, panmictic population which has discrete generations and a constant population size of N. Our main goal is to explore the dependence of the pattern and rate of allele-frequency change at the loci driving the adaptation of a quantitative trait on the model's parameters. To achieve this aim, we derive the dynamics from first principles. First we set up the basic model for the quantitative trait, then the underlying one-locus model, the multilocus model, which we assume to be an infinite-sites model, and the simulation model. A glossary of symbols is provided in Table 1.

2.1. The quantitative-genetic model

We consider a quantitative trait subject to nonepistatic directional selection and investigate the evolution of the distribution of allele frequencies at the loci contributing to this trait during adaptation. The fitness of an individual with phenotypic value G is given by $\exp(sG)$, where the selection coefficient $s > 0$ is assumed to be constant in time. The trait is determined additively by a (potentially infinite) number of diallelic loci at which mutation occurs. We ignore environmental contributions to the phenotype which, in the absence of genotype-environment interaction and of epistasis, would only weaken selection on the underlying loci. Therefore, genotypic and phenotypic values are equal.

Each locus i can be in the ancestral state a_i or in the derived state A_i. The effect on the phenotype of a locus in the ancestral state is zero, whereas a mutation at locus i contributes a value $\alpha_i > 0$ to the phenotype G. Thus, the phenotype of an individual is given by $G = \sum_i \alpha_i \delta_{A_i}$, where $\delta_{A_i} = 1$ if locus i is in the derived state, and $\delta_{A_i} = 0$ otherwise. The fitness of the ancestral allele is normalized to 1. Therefore, the fitness of the ancestral genotype, which has value $G = 0$, is also one. Because of our additivity assumption and the shape of the fitness function, a mutation at locus i has a fitness of $\sigma_i = e^{s\alpha_i} > 1$, where $s > 0$. Hence, all mutations are advantageous. Because we mostly assume $Ns > 1$, deleterious mutations are very unlikely to reach high frequency

or become fixed, and thus will not contribute to a permanent selection response of the mean phenotype.

Unless stated otherwise, we assume that initially all loci are in the ancestral state and adaptation occurs from *de novo* mutations. New mutations occur from the ancestral state to the derived state and back-mutations are ignored. We define the population-wide *advantageous* mutation rate by $\Theta = NU$, where U is the expected number of mutations per individual that increase the trait value. Thus, Θ is the expected number of new mutations per generation that affect the trait. We assume an *infinite sites model*, i.e., each mutation hits a new locus. This assumption is appropriate if per-locus mutation rates are small enough such that the next mutation occurs only after the previous one becomes lost or fixed.

In the infinite sites model we investigate two different scenarios. In the first, all mutation effects are equal. In the second, mutation effects α_i at any locus i are drawn from a distribution defined on $[0, \infty)$ that has a density f. Its mean is denoted by $\bar{\alpha}$ and all its higher moments are assumed to exist. Thus, the loci are exchangeable and all mutations are beneficial. For specification of analytical results and their comparison with simulation results, we use an exponential distribution (for empirical support, see Eyre-Walker and Keightley, 2007, Tataru et al., 2017) and a truncated normal distribution (see Remark 4.10).

2.2. The one-locus model

We consider a single diallelic locus in isolation and assume that allele frequencies evolve according to a haploid Wright-Fisher model with selection. The (relative) fitness of the ancestral allele a and of the derived allele A (a new mutant) are set to 1 and $\sigma = e^s$, $s > 0$. Therefore, the selective advantage of the mutant relative to the wildtype is $e^s - 1 \approx s$. We assume that in generation $n = 0$ the new beneficial mutant A occurs as a single copy and no other mutation occurs while this allele is segregating.

Because it seems unfeasible to derive analytic results for the time dependence of the allele frequency for either the Wright-Fisher model or its diffusion approximation, we approximate the stochastic dynamics in the initial phase, where stochasticity is most important, by a branching process (e.g. Athreya and Ney, 1972, Allen, 2003, Haccou et al., 2005). During this initial phase, interactions between mutant lineages can be ignored if N is large enough. The first step is to approximate the evolution of the frequency distribution of the mutant by a Galton-Watson process. In Section 3, we will approximate this discrete process by a continuous one and couple it with the deterministic allele-frequency dynamics to obtain an analytically accessible model for the long-term evolution of the mutant in a population of size N.

To be specific, we consider a Galton-Watson process $\{Z_n\}$, where $Z_n = \sum_{j=1}^{Z_{n-1}} \xi_j$, $Z_0 = 1$, and ξ_j denotes the (random) number of offspring of individual j in generation $n-1$. Thus, Z_n counts

the number of descendants of the mutant that emerged in generation 0. We assume that the ξ_i are independent, identically distributed random variables having at least three finite moments, where

$$\mathrm{E}[\xi_j] = \sigma > 1 \text{ and } \mathrm{Var}[\xi_j] = v\,. \tag{2.1}$$

Therefore, the process $\{Z_n\}$ is supercritical. The ξ_i are also independent of n. We denote the generating function of ξ_j (hence of Z_1) by φ.

Whenever we consider extinction or survival probabilities as functions of $\sigma = e^{s\alpha}$, we assume that $v = 2\sigma/\rho$ holds for the underlying family of offspring distributions, where $\rho > 0$ is a constant (cf. Remark 3.3). In this case, we also assume that the probability of ultimate extinction is continuous in $\sigma = e^{s\alpha}$ and decreases with increasing σ. When we compare our analytical results with simulation results from a Wright-Fisher model, we assume that the offspring distribution is Poisson, because this provides a close approximation to the binomial distribution if the population size N is sufficiently large, especially if $Ns \gg 1$ (cf. Lambert, 2006).

2.3. The infinite-sites model

We assume that the trait is determined by a potentially infinite number of loci at which new, beneficial mutants can occur. We assume an infinite-sites model, so that every mutation occurs at a new, previously monomorphic locus.

The initial population starts evolving in generation $\tau = 0$ in which the first mutations may occur. Prior to generation 0, the ancestral allele is fixed at every site potentially contributing to the trait. We assume that in each generation $\tau \geq 0$, the total number of new mutants emerging in the population follows a Poisson distribution with mean Θ. Each mutant occurs at a new site (locus). Because selection is multiplicative in our model and random genetic drift is a weak force relative to selection (because N is assumed to be large), we ignore linkage disequilibrium among loci. Hence, we assume that the allele-frequency distributions at the different loci evolve independently.

According to the mutation model, the ith mutation will occur at a random time τ_i. The corresponding locus will be labeled locus i, and the mean offspring number of the mutant - when the mutant is still rare in the population - is denoted by $\sigma_i > 1$. For mathematical convenience, we approximate the discrete distribution of the waiting time τ_i by the Erlang distribution with density

$$m_{i,\Theta}(t) = \frac{\Theta^i t^{i-1}}{(i-1)!} \exp(-\Theta t)\,, \quad \text{if } t > 0 \tag{2.2}$$

(and $m_{i,\Theta}(t) = 0$ if $t \leq 0$). Throughout, we use t as an integration variable and emphasize that generations τ are discrete. The mean waiting time until mutation i is i/Θ. We note that if mutations occurred according to a Poisson point process with rate Θ, the Erlang distribution would be the exact waiting-time distribution between i events. Here, we use it as an approximation. The

total number of mutation events k until generation τ is Poisson distributed with mean $\Theta\tau$

$$\mathrm{Poi}_{\Theta\tau}(k) = \frac{(\Theta\tau)^k}{k!} \exp(-\Theta\tau)\,. \tag{2.3}$$

2.4. The simulation model

The simulation algorithm written in *C++* is based on previous versions by Höllinger (2018) and Wölfl (2019). With the above assumptions, we generate Wright-Fisher simulations forward in discrete time to examine the adaptation of a trait under selection, mutation and random genetic drift in the case of an originally monomorphic ancestral population. The first mutation events can occur in generation 0. The algorithm acts on each locus because we assume that loci evolve independently. The basic idea for the algorithm is that each generation is represented by a vector, whose length is determined by the number of mutated loci. A mutation event occurs according to the mutation rate, adds a new locus and thus increases the length of the vector by one. The entries of the vector contain the number of mutants at each locus. Random genetic drift is included via binomial sampling with sampling weights due to selection (Ewens, 2004, p. 24). We use the *GNU Scientific Library* (Galassi et al., 2009) and the *MT19937* (Matsumoto and Nishimura, 1998) random number generator. Unless stated otherwise, a simulation for a set of parameters consists of $20\,000$ replicate runs. In almost all cases, $20\,000$ replicates yielded standard error bars smaller than the size of the symbols in the figures; in a few cases, more runs were performed. Therefore, error bars are not shown. Runs are stopped at a given number n (or τ) of generations. From these replicate runs, means (and variances) of the quantities of interest are computed, or histograms if distributions are the objects of interest.

3. Dynamics of allele-frequency distributions

It is well known that, conditioned on fixation, the frequency of a beneficial mutant grows faster than predicted by the corresponding deterministic model (e.g. Maynard Smith and Haigh, 1974, Otto and Barton, 1997). For models related to ours, it has been shown that conditioned on fixation and once a certain threshold has been exceeded, the evolution of the frequency of mutants can be approximated by a deterministic increase from a random variable, often labeled W, that describes the stochasticity that accumulates during the spread of the mutant (Desai and Fisher, 2007, Uecker and Hermisson, 2011). The latter authors called this the *effective initial mutant population size*. Martin and Lambert (2015) employed a variant of this approach and approximated the initial and the final phase of evolution by a Feller process conditioned on fixation. Based on this, they derived a semi-deterministic approximation for the distribution of the (beneficial) allele frequency at any given time.

Below, we develop a refined variant that provides a highly accurate, analytically tractable, and computationally efficient approximation for the frequency distribution of a beneficial mutant in any generation n in a diallelic haploid Wright-Fisher population of size N (Result 3.4). Its extension in eqs. (3.20) and (3.21) to the ith mutant occurring in the infinite sites model paves the way for an analytical treatment of our applications in Sections 4 and 5.

3.1. Approximating the random variable W

Our starting point is the supercritical Galton-Watson process $\{Z_n\}$ defined in Section 2.2. The probability of ultimate extinction, denoted by $P_{\text{ext}}^{\infty}(\sigma)$, is the unique fixed point of the offspring generating function φ in $(0,1)$, i.e., it satisfies $\varphi(P_{\text{ext}}^{\infty}) = P_{\text{ext}}^{\infty}$ and $0 < P_{\text{ext}}^{\infty} < 1$. When no confusion can occur, we suppress the dependence of P_{ext}^{∞} on the mean offspring number $\sigma > 1$.

We define $W_n = Z_n/\sigma^n$. According to a classical result (Kesten and Stigum, 1966, Athreya and Ney, 1972, Haccou et al., 2005, p. 154) there exists a random variable W such that

$$\lim_{n\to\infty} W_n = W \tag{3.1}$$

with probability 1. The limiting random variable W has expectation $\mathrm{E}[W] = 1$, satisfies $W = 0$ precisely if the process dies out, and W is positive and absolutely continuous[1] otherwise. We define the random variable W^+ corresponding to the absolutely continuous part of W by

$$\mathrm{Prob}[W \le x] = P_{\text{ext}}^{\infty} + (1 - P_{\text{ext}}^{\infty})\,\mathrm{Prob}[W^+ \le x]\,, \quad \text{where } x \ge 0\,, \tag{3.2}$$

so that W^+ has a proper distribution with density w^+. Because $\mathrm{E}[W] = 1$, eq. (3.2) informs us that $\mathrm{E}[W^+] = (1 - P_{\text{ext}}^{\infty})^{-1}$.

In general, not much is known about the shape of the distribution W^+. However, for a geometric offspring distribution with mean σ, W^+ is exponential with parameter $1 - 1/\sigma$ (for a proof, see e.g. Haccou et al., 2005, pp. 155–156). As pointed out by a reviewer, W^+ is exponential if and only if the offspring distribution is a fractional linear, or modified geometric, distribution (which has a fractional linear generating function). We treat the properties of the associated Galton-Watson processes that are relevant in our context in Appendix A. We will use them when extending results in Sections 4 and 5 to cases when the effective population size N_e differs from the actual population size N.

We denote the Laplace transform of W by $\Phi(u) = \mathrm{E}[e^{-uW}]$, where $u \ge 0$. It is well known (Athreya and Ney, 1972, p. 10) that Φ is uniquely determined as the solution of Poincaré's (or

[1] Absolute continuity of a function f is equivalent to having a derivative f' almost everywhere and f' being Lebesgue integrable.

Abel's) functional equation

$$\Phi(u) = \varphi(\Phi(u/\sigma))\,. \tag{3.3}$$

In general, this cannot be solved explicitly. However, if the offspring distribution is geometric, then (3.3) becomes

$$\Phi(u) = \frac{1}{1+\sigma(1-\Phi(u/\sigma))}\,. \tag{3.4}$$

This is (uniquely) solved by the Laplace transform of W with (cumulative) distribution

$$\mathrm{Prob}[W \le x] = P_{\mathrm{ext}}^{\infty} + (1-P_{\mathrm{ext}}^{\infty})\,(1-\exp[-(1-P_{\mathrm{ext}}^{\infty})x])\ . \tag{3.5}$$

Hence, W^+ is exponentially distributed with parameter $1-P_{\mathrm{ext}}^{\infty} = 1-\sigma^{-1}$ (Haccou et al., 2005, p. 155).

For a Poisson offspring distribution, which has the generating function $\varphi(x) = e^{-\sigma(1-x)}$, we use the approximation $\varphi(x) \approx (1+\sigma(1-x))^{-1}$ which has the same error as the first-order Taylor approximation near $x=1$, but is much more accurate on the whole interval $[0,1]$. With this approximation, we obtain from (3.3) that $\Phi(u) \approx \Big(1+\sigma(1-\Phi(u/\sigma))\Big)^{-1}$. Therefore, for a Poisson offspring distribution, we approximate W^+ by an exponential distribution with parameter $1-P_{\mathrm{ext}}^{\infty}$, where $P_{\mathrm{ext}}^{\infty}$ is the corresponding extinction probability (see Sect. 3.2).

Remark 3.1. A straightforward calculation shows that

$$\mathrm{Var}[W^+] = \frac{1}{1-P_{\mathrm{ext}}^{\infty}}\left(\mathrm{Var}[W] + 1 - \frac{1}{1-P_{\mathrm{ext}}^{\infty}}\right)\,, \tag{3.6}$$

where $\mathrm{Var}[W] = v/(\sigma(\sigma-1))$ and v is the variance of the offspring distribution (cf. Athreya and Ney, 1972, p. 4). If the offspring distribution is geometric with mean $\sigma > 1$, then $P_{\mathrm{ext}}^{\infty} = 1/\sigma$, $\mathrm{Var}[W^+] = \sigma^2/(\sigma-1)^2 = \mathrm{E}[W^+]^2$, as is required for an exponential distribution. For a Poisson offspring distribution with mean $\sigma > 1$, it follows that the coefficient of variation of W^+ has the series expansion $1-\frac{1}{3}(\sigma-1)+O((\sigma-1)^2)$. It can also be shown that the skew has the expansion $2-(\sigma-1)+O((\sigma-1)^2)$. This suggests that as σ decreases to 1, the accuracy of the exponential approximation increases. Therefore, our approximations perform best in the slightly supercritical case, i.e., if σ is only slightly larger than 1. It is straightforward to infer that for a Poisson offspring distribution, W^+ is not gamma distributed because the skew is too small.

Remark 3.2. As mentioned above and shown in Appendix A.2 for self-containedness, the distribution of W^+ is exponential if and only if the offspring distribution is fractional linear. Our argument above shows that W^+ is approximately exponential with parameter $1-P_{\mathrm{ext}}^{\infty}$ for a Poisson offspring distribution, and second and third moments converge to that of an exponential as $\sigma \to 1$. We expect that this is the case for a large class of offspring distributions, i.e., when for $\sigma \to 1$ the generating function is asymptotically equivalent to that of a fractional linear offspring distribution.

Indeed, Uecker and Hermisson (2011) derived an exponential distribution for W^+ for a birth-death branching processes with time-dependent mean offspring number (as they consider a model with temporal variation in its population size and the selection pressure). Martin and Lambert (2015) show that W^+ is approximately exponential with parameter $1 - P_{\text{ext}}^{\infty}$ for Feller diffusions. They start the deterministic phase in their Wright-Fisher model with selfing from a gamma distribution because they admit one or more initially segregating mutants.

3.2. Extinction probability by generation n for a Poisson offspring distribution

Because we use a Poisson offspring distribution for comparisons of our analytic results with simulation results from the corresponding Wright-Fisher model, we summarize the properties of the probabilities of extinction of the mutant until generation n. We assume that the mutant starts as a single copy in generation $n = 0$, and the offspring distribution is Poisson with mean $\sigma = e^s$, $s > 0$. Then the probability $P_{\text{ext}}^{\infty}(\sigma)$ of ultimate loss, or extinction, of the mutant is the unique solution of $\varphi(x) = e^{-\sigma(1-x)} = x$ in $(0, 1)$. It can be expressed in terms of the (principal branch of the) product logarithm, or the Lambert function:

$$P_{\text{ext}}^{\infty}(\sigma) = -\frac{\text{ProductLog}[-\sigma \exp(-\sigma)]}{\sigma}\,. \tag{3.7}$$

With $\sigma = e^s$, this has the series expansion

$$P_{\text{ext}}^{\infty}(\sigma) = 1 - 2s + \frac{5s^2}{3} + O(s^3)\,. \tag{3.8}$$

The extinction probability $P_{\text{ext}}^{\infty}(e^s)$ decreases from 1 to 0 as s increases from 0 to ∞. We write $P_{\text{sur}} = 1 - P_{\text{ext}}^{\infty}$ for the (long-term) survival probability. It follows directly from (3.8) that Haldane's (1927) approximation, $2s$, serves as an upper bound for P_{sur} and approaches P_{sur} as $s \to 0$ (indeed, Haldane derived an equation equivalent to $e^{-\sigma(1-x)} = x$).

We define $P_{\text{ext}}(n, \sigma)$ as the probability that the mutation gets lost before generation n, i.e., $1 - P_{\text{ext}}(n, \sigma)$ is the probability that the mutation is still segregating in generation n. If no confusion can occur, we write $P_{\text{ext}}(n)$ instead of $P_{\text{ext}}(n, \sigma)$. Then $P_{\text{ext}}(n)$ satisfies the recursion

$$P_{\text{ext}}(0) = 0, \tag{3.9a}$$

$$P_{\text{ext}}(n) = \exp[-\sigma(1 - P_{\text{ext}}(n-1))]. \tag{3.9b}$$

because $P_{\text{ext}}(n) = \varphi_{(n)}(0) = \varphi(\varphi_{(n-1)})(0)$, the nth iteration of the generating function φ of the Poisson distribution. By monotonicity of $P_{\text{ext}}(n)$ and continuity of φ it follows that $P_{\text{ext}}(n, \sigma) \to P_{\text{ext}}^{\infty}(\sigma)$ for $n \to \infty$ and for every σ. Indeed, it is well known that $P_{\text{ext}}(n, \sigma) \to P_{\text{ext}}^{\infty}(\sigma)$ holds for very general offspring distributions.

An explicit formula for $P_{\text{ext}}(n)$ seems to be unavailable. However, the recursion (3.9) is readily

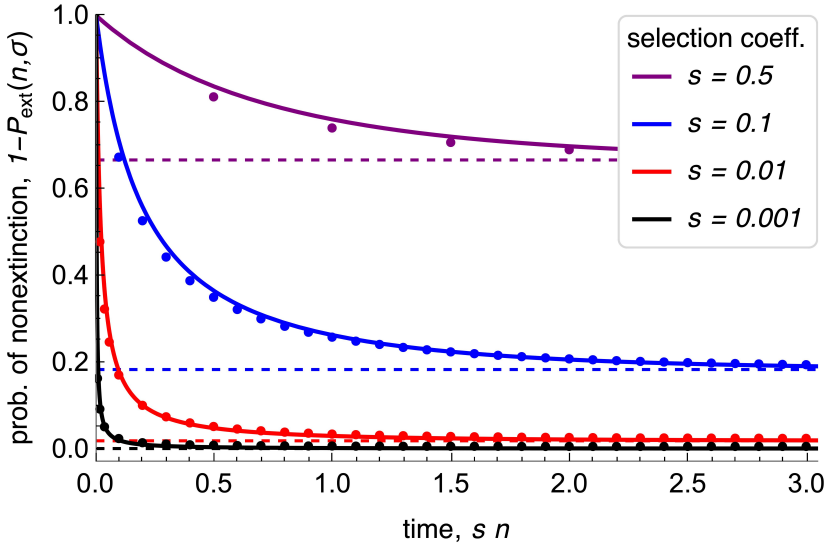

Figure 3.1: Probability of non-extinction of a mutant by generation n, $1 - P_{\text{ext}}(n, \sigma)$, for various values of the mean $\sigma = e^s$ of a Poisson offspring distribution. Time is scaled to facilitate comparison for different selection coefficients s. The dots show $1 - P_{\text{ext}}(n, \sigma)$ as obtained from numerical iteration of (3.9). The solid curves show the respective approximations (and upper bounds) obtained from (3.10) with $P^{\infty}_{\text{ext}}(\sigma)$ from (3.7). The dashed lines show the survival probabilities $P_{\text{sur}}(\sigma) = 1 - P^{\infty}_{\text{ext}}(\sigma)$, as calculated from (3.7).

evaluated numerically and convergence of $1 - P_{\text{ext}}(n, \sigma)$ to $P_{\text{sur}}(\sigma) = 1 - P^{\infty}_{\text{ext}}(\sigma)$ is fast (Fig. 3.1). As is documented by Figure 3.1 and supported by extensive numerical evalution, the inequality

$$\frac{1 - P_{\text{ext}}(n, \sigma)}{1 - P^{\infty}_{\text{ext}}(\sigma)} \leq 1 + \frac{P^{\infty}_{\text{ext}}(\sigma)}{\sigma^n - P^{\infty}_{\text{ext}}(\sigma)} \tag{3.10}$$

apparently holds always, and the upper bound serves as a highly accurate approximation unless σ is large and n is very small. For fractional linear offspring distributions, equality holds in (3.10) (Appendix A.1). Indeed, convergence of $1 - P_{\text{ext}}(n, \sigma)$ to $1 - P^{\infty}_{\text{ext}}(\sigma)$ at the geometric rate $1/\sigma$ follows immediately from Corollary 1 in Sect. 1.11 of Athreya and Ney (1972). The time T_ϵ needed for the probability of non-extinction by generation n to decline to $(1 + \epsilon)(1 - P^{\infty}_{\text{ext}}(\sigma))$ is explicitly given in (A.11) for the fractional linear case, and it provides an accurate approximation and upper bound for the Poisson case (see Appendix A.3).

Remark 3.3. For a large class of offspring distributions having bounded third moments, $P_{\text{sur}}(\sigma)$ has an approximation of the form $\frac{2(\sigma-1)}{v} + O((\sigma - 1)/v)^2)$ as σ decreases to 1 (Athreya, 1992). If $\sigma = e^{s\alpha}$ and $s\alpha$ is small, this yields $P_{\text{sur}}(e^{s\alpha}) \approx \rho s\alpha$, where $\rho = 1$ for a geometric offspring distribution, and $\rho = 2$ for a Poisson distribution or any other distribution with $v = \sigma$; cf. (3.8) and (A.20). By our monotonicity and continuity assumption on the extinction probability (Sect. 2.2), P_{sur} is monotone increasing in σ, $P_{\text{sur}}(e^{s\alpha}) \leq \min\{\rho s\alpha, 1\}$ for every $s\alpha$, and $P_{\text{sur}}(e^{s\alpha}) \to \rho s\alpha$ as

$s\alpha \to 0$.

3.3. Evolution of the distribution of mutant frequencies at a single locus

Now we start the investigation of the evolution of the frequency distribution of a mutation that occurred as a single copy in generation $n = 0$. Our first goal is to approximate the distribution of Z_n conditioned on $Z_n > 0$ by a simple continuous distribution. Therefore, we define the distribution of the (discrete) random variable W_n^+

$$\mathrm{Prob}[W_n \leq x] = P_{\mathrm{ext}}(n) + (1 - P_{\mathrm{ext}}(n))\,\mathrm{Prob}[W_n^+ \leq x]\,, \quad \text{where } x \geq 0\,. \tag{3.11}$$

Then $\mathrm{Prob}[Z_n \leq y | Z_n > 0] = \mathrm{Prob}[\sigma^n W_n^+ \leq y]$. We approximate this distribution by the exponential distribution

$$\Psi_n(y) = 1 - e^{-\lambda_n y}\,, \text{ where } y \geq 0 \text{ and } \lambda_n = \frac{1 - P_{\mathrm{ext}}(n)}{\sigma^n}\,, \tag{3.12}$$

and denote the corresponding random variable by Y_n and its density by ψ_n. For fractional linear offspring distributions, this approximation is best possible in a rigorous sense (see Appendix A). Because of the convergence of W_n^+ to W^+, the exponential distribution Ψ_n will provide a close approximation to the true distribution of W_n^+ (with the possible exception of very small values of n) if W^+ is approximately exponential. Figure 3.2 shows that for a Poisson offspring distribution, the exponential density ψ_n provides an excellent approximation for the number of mutants in generation n, conditioned on non-extinction until n, in a Wright-Fisher model. As σ decreases to 1, the accuracy of the approximation will increase.

The above approximation, $\mathrm{Prob}[Z_n \leq y | Z_n > 0] \approx \Psi_n(y)$, informs us that in an infinite population the distribution of the number of mutants after n generations, conditioned on non-extinction until generation n and originating from a single mutation, can be described by a deterministic exponential increase that starts with an exponentially distributed initial mutant population (according to ψ_0) of mean size 1. Then the mean of this population in generation n is $\sigma^n/(1 - P_{\mathrm{ext}}(n))$, which approaches $\sigma^n/P_{\mathrm{sur}}$ as n increases, and the distribution remains exponential. Thus, according to this model a beneficial mutant destined for survival does indeed grow faster than expected from the corresponding deterministic model (which has a growth rate of σ), in particular in the early generations.

As already noted, Desai and Fisher (2007), Uecker and Hermisson (2011), and Martin and Lambert (2015) used similar approaches, but conditioned on fixation, instead of non-extinction until the given generation. Therefore, they took the random variable W^+, or rather its exponentially distributed approximation with the fixation probability as parameter, as their *effective initial mutant population size* (which on average is larger than one). For large n, the distribution of $\sigma^n W^+$ has the same asymptotic behavior as Ψ_n in (3.12), but our distribution Ψ_n provides a more

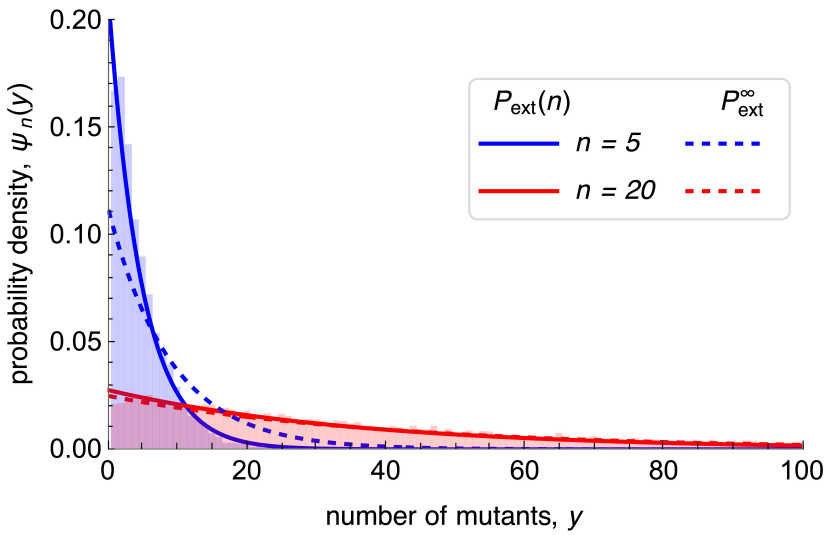

Figure 3.2: The probability density $\psi_n(y)$ of the number of descendants of a single mutation in generation $n = 5$ (blue) and $n = 20$ (red). Solid curves show the density $\psi_n(y)$ of (3.12). Dashed curves show its simplification obtained by substituting P_{ext}^{∞} for $P_{\text{ext}}(n)$ in (3.12), which is equivalent to conditioning on long-term survival. The mean of the Poisson offspring distribution is $\sigma = e^s$ with $s = 0.1$. The histograms are from Wright-Fisher simulations with a population size of $N = 10^4$, conditioned on non-extinction until generation n.

accurate approximation for the early phase of the spread of the mutant (Figure 3.2) because in each generation it has the correct mean. The distributions shown by dashed curves in Figure 3.2 are obtained by conditioning on long-term survival in the branching process.

Now we apply the above results to derive an approximation for the frequency of a mutant in a finite population of large and constant size N, conditioned on non-extinction until generation n. Our starting point is the approximate exponential growth of the number of mutants, as given by the distributions Ψ_n in (3.12) of the random variables Y_n. Because in a finite population exponential growth is impossible, we follow the population genetics tradition and use relative frequencies (probabilities). We assume that the number of resident types remains at its initial frequency of $N - 1 \approx N$ (because N is large and residents produce one offspring). We define the random variables X_n, measuring the relative frequency of the mutants in the total population in generation n, by

$$X_n = \frac{Y_n}{Y_n + N} \,. \tag{3.13}$$

We note that in the absence of stochastic variation ($\mathrm{Var}[Y_n] = 0$), we have $Y_n = \sigma^n Y_0$ and therefore $p(n) = X_n = Y_n/(Y_n + N - 1)$ solves the corresponding deterministic diallelic selection equation, (4.5), if $p(0) = Y_0/N = 1/N$. Because (3.13) is equivalent to $Y_n = NX_n/(1 - X_n)$, the density of X_n is approximated by

$$g_{a(n)}(x) = \psi_n\left(N\frac{x}{1-x}\right)\frac{d}{dx}\left(\frac{Nx}{1-x}\right) \,, \tag{3.14}$$

where ψ_n is the density of Ψ_n in (3.12). A simple calculation yields (3.15) in our key result of this section:

Result 3.4. *Conditioned on non-extinction until generation n and for given N, the discrete distribution of the mutant frequency originating from a single mutation in generation 0 with fitness σ can be approximated by the distribution of the positive continuous random variable X_n with density*

$$g_a(x) = \frac{a}{(1-x)^2} \cdot \exp\left[-a\frac{x}{1-x}\right], \tag{3.15a}$$

where

$$a = a(n, \sigma, N) = N(1 - P_{\text{ext}}(n, \sigma))\sigma^{-n}. \tag{3.15b}$$

This result does not provide a controlled approximation of the Wright-Fisher or the associated diffusion dynamics, simply because the time dependence of allele frequencies is prohibitively difficult to deal with analytically in these stochastic dynamical systems. However, its utility and accuracy will be demonstrated below.

Remark 3.5. Our expression (3.15) for the density $g_a(x)$ has the same structure as that given by Martin and Lambert (2015) in their equation (4). The difference is that their β_t decays exponentially with t with a constant parameter, whereas our $a = a(n, \sigma, N)$ decays exponentially with a parameter depending on n. This difference originates from the fact that we condition on non-extinction until n, whereas Martin and Lambert (2015) condition on fixation.

To interpret and understand the analytic results derived below, we first need to study the properties of the density $g_a(x)$ in (3.15) and of its constituent terms. The only parameter on which the mutant frequency distribution $g_a(x)$ in (3.15) depends is the compound parameter $a = a(n, \sigma, N)$. This dependence is displayed in Fig. 3.3. Obviously, $a(n, \sigma, N)$ is proportional to N. As a function of n, $a(n, \sigma, N)$ decreases approximately exponentially from its maximum value N, assumed at $n = 0$, to 0 as $n \to \infty$. As a function of σ (or s), $a(n, \sigma, N)$ decreases as well (see Fig. S.1).

We observe that $g_{a(n,\sigma,N)}(0) = a(n, \sigma, N)$. If $a > 2$, then $g_a(x)$ attains its maximum at $x = 0$ (Fig. 3.3). In this case $g_a(x)$ decays with increasing x (and approximately exponentially for small x). This will always be the case in the initial phase of the mutant's invasion. If $a < 2$, which will occur for sufficiently large n, then $g_a(x)$ has an unique maximum $g_a(\hat{x}) = \frac{4}{a}\exp(a-2)$ at $\hat{x} = 1 - \frac{a}{2} \in (0, 1]$, which shifts from $x = 0$ to $x = 1$ as a decreases (see the red curve in Fig. 3.3).

A crucial quantity in a is also the probability of nonextinction by generation n, $1 - P_{\text{ext}}(n, \sigma)$. Figure 3.1 documents the approach of $1 - P_{\text{ext}}(n, \sigma)$ to $1 - P^{\infty}_{\text{ext}}(\sigma)$ as n increases and compares it with the approximation (3.10). The convergence is quick if $s \gtrapprox 0.1$ (Section 3.2). For small s, however, $1 - P_{\text{ext}}(n)$ is much larger than $1 - P^{\infty}_{\text{ext}}$ for many more generations. Consequently,

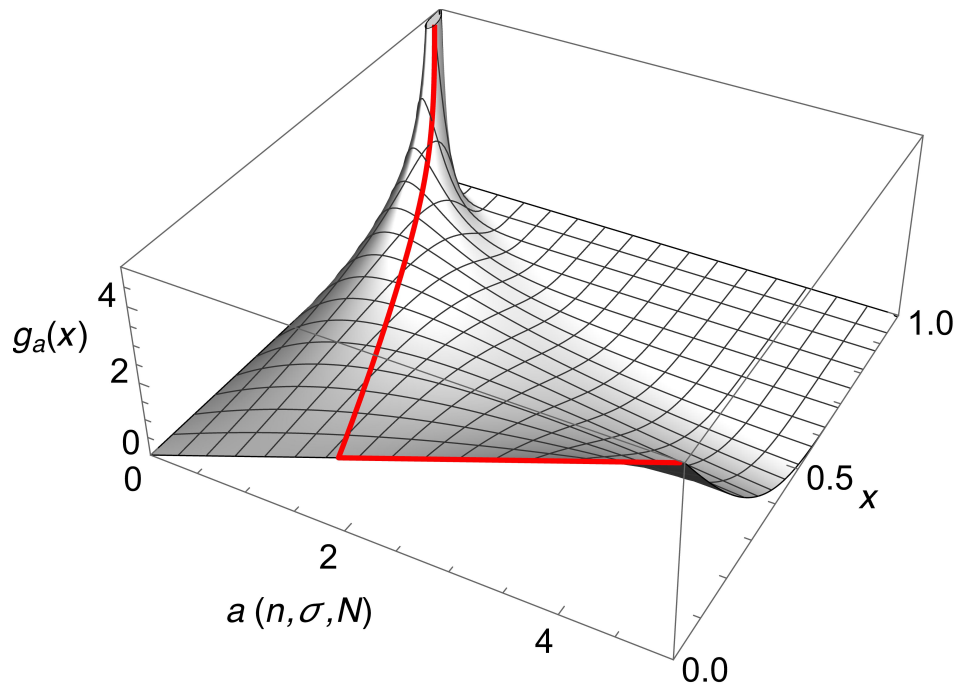

Figure 3.3: The probability density $g_a(x)$ in (3.15) is shown as a function of x and a. The red curve represents the unique maximum $g_a(\hat{x}(a))$ at $\hat{x}(a) = 1 - \frac{a}{2}$ for $a < 2$ and the unique maximum at zero for $a > 2$.

simplifying g_a by substituting P_{ext}^{∞} for $P_{\text{ext}}(n)$ will considerably decrease the accuracy of the approximation during this initial phase (cf. Fig. 3.2, which shows this effect for ψ_n).

Figures 3.4 and S.3 compare the analytic approximation (3.15) for the distribution of mutant frequencies at various generations with the corresponding histograms obtained from Wright-Fisher simulations. They show that the approximation is very accurate in the initial phase whenever $Ns \geq 1$. If $Ns \geq 100$, it remains accurate almost until fixation of the mutant. If Ns is of order 10, the approximation remains accurate for a surprisingly long period, essentially until the (mean) mutant frequency exceeds $\frac{1}{2}$, i.e., until it has become the dominating variant. That the branching-process approximation underestimates the effects of random genetic drift when the mutant reaches high frequency is an obvious consequence of its assumptions. In Section 4.1 we will see that the mean of g_a provides a highly accurate approximation to the true mean even if the mutation is close to fixation and Ns is much less than 100.

We note that Wright-Fisher simulations show that for $s > 0.5$ (i.e., very strong selection), the distribution of the mutant frequencies cannot be approximated by the density g_a. The reason is that for large σ and a Poisson offspring distribution, the density ψ_n of $Y_n = N\frac{X_n}{1-X_n}$ deviates too much from an exponential distribution. That for large σ, the variance becomes much smaller than the mean, can be inferred directly from the coefficient of variation given in Remark 3.1. In the following we exclude very strong selection and assume $s < 0.5$. In fact, we focus on parameters

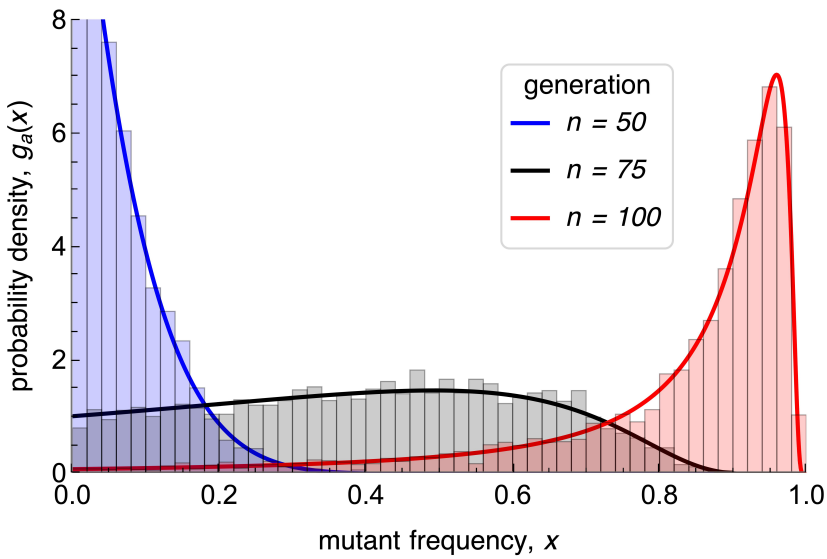

Figure 3.4: The curves show the probability density $g_{a(n,\sigma,N)}(x)$ in (3.15) of the mutant frequency in different generations n, as indicated in the legend. The histograms are obtained from simulations of the corresponding Wright-Fisher model. The population size is $N = 10^4$ and the mean of the Poisson offspring distribution is $\sigma = e^s$ with $s = 0.1$.

satisfying $s \leq 0.1$ and $Ns \geq 10$.

3.4. Evolution of the distribution of mutant frequencies in the infinite-sites model

We use the model outlined in Section 2.3. In particular, τ refers to the generation since the initial population started to evolve. Mutations $i = 1, 2, 3, \ldots$ occur at generations τ_i, as outlined above. We note that τ differs from n, as used in the above subsections on a single locus, because for locus i the generation number in the Galton-Watson process is $n = 0$ at time τ_i. From (2.2) we recall the Erlang distribution and define

$$M_{i,\Theta}(\tau) = \int_0^\tau m_{i,\Theta}(t)dt = 1 - \frac{\Gamma(i, \Theta\tau)}{\Gamma(i)}\,, \tag{3.16}$$

which approximates the probability that i mutations have occurred by generation τ. Here, $\Gamma(n, z) = \int_z^\infty x^{n-1}e^{-x}\,dx$ denotes the incomplete Gamma function (Abramowitz and Stegun, 1964, Chap. 6.5). We recall that $\lim_{z\to 0}\Gamma(n, z) = \Gamma(n) = (n-1)!$ (if n is a positive integer), and $\Gamma(0, z) = E_1(z)$, where $E_1(z) = \int_z^\infty x^{-1}e^{-x}\,dx$ is the exponential integral.

In the following, we write $[t]$ for the nearest integer to t, and we define

$$a(t, \sigma_i, N) = N(1 - P_{\text{ext}}([t], \sigma_i))\sigma_i^{-t} \quad \text{for every (real) } t \geq 0\,, \tag{3.17}$$

where σ_i is the fitness of the mutant at locus i. From Result 3.4 we infer that, conditioned on non-extinction until generation τ, the discrete distribution of the frequency of mutations at the

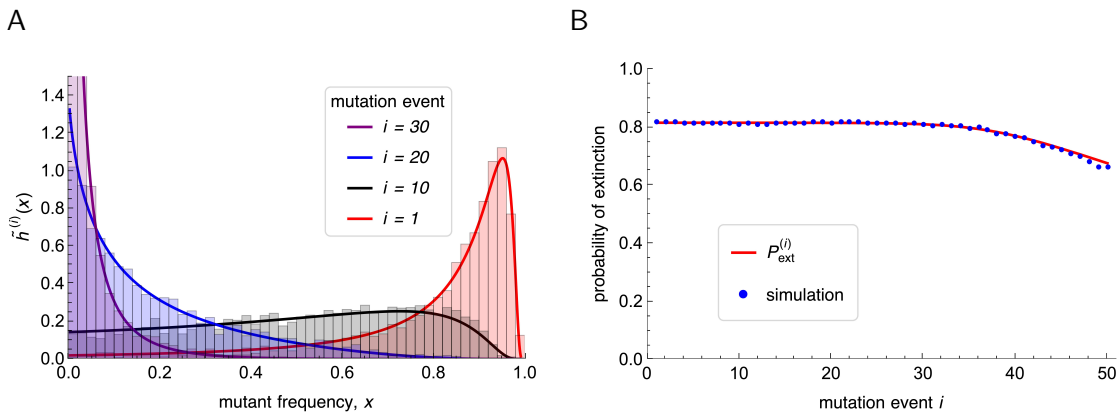

Figure 3.5: The curves in panel A show the absolutely continuous part $\tilde{h}_\tau^{(i)}(x)$ in (3.20) of the mutant frequency $\tilde{X}_\tau^{(i)}$ at the ith locus (indicated in the legend) in generation $\tau = 100$. The histograms are obtained from simulations of the corresponding Wright-Fisher model. The population-wide mutation rate is $\Theta = 0.5$, the population size is $N = 10^4$ and the mean of the Poisson offspring distribution is $\sigma = e^s$ with $s = 0.1$. The red curve in panel B displays the probability $P_{\text{ext}}^{(i)}$ in (3.19) that the mutation at locus i has been lost by generation $\tau = 100$. The blue symbols show the simulation results from the corresponding Wright-Fisher model. For larger i, the simulated values would lie below the theoretical curve because in some replicate runs considerably less than 50 mutations occur until generation 100. We note that the shape of $h_\tau^{(i)}$ is essentially identical to that of $\tilde{h}_\tau^{(i)}$, except that the vertical range has to be rescaled. For smaller τ or smaller s slight differences become visible (results not shown).

locus at which the ith mutation event occurred, locus i for short, can be approximated by the distribution of the continuous random variable with density

$$h^{(i)}(x) = h^{(i)}_{\tau,\sigma_i,N,\Theta}(x) = \frac{1}{M_{i,\Theta}(\tau)} \int_0^\tau g_{a(\tau-t,\sigma_i,N)}(x)\, m_{i,\Theta}(t)\, dt\,. \tag{3.18}$$

Here, we used that the ith mutant has occurred approximately $\tau - [t]$ generations in the past, where t is Erlang distributed with parameters i and Θ.

The probability that the mutation at locus i has been lost by generation τ, i.e., that the ancestral type is again fixed at this locus in generation τ, can be approximated by

$$P^{(i)}_{\text{ext}}(\tau,\sigma_i,\Theta) = \frac{1}{M_{i,\Theta}(\tau)} \int_0^\tau P_{\text{ext}}(\tau - [t], \sigma_i)\, m_{i,\Theta}(t)\, dt \tag{3.19}$$

(Fig. 3.5B). Below, we shall need unconditioned distributions of the mutant frequencies. Therefore, we define

$$\tilde{h}^{(i)}(x) = \tilde{h}^{(i)}_{\tau,\sigma_i,N,\Theta}(x) = \frac{1}{M_{i,\Theta}(\tau)} \int_0^\tau (1 - P_{\text{ext}}([\tau - t], \sigma_i))\, g_{a(\tau-t,\sigma_i,N)}(x)\, m_{i,\Theta}(t)\, dt \tag{3.20}$$

(see Fig. 3.5A). Then $\int_0^1 \tilde{h}^{(i)}_{\tau,\sigma_i,N,\Theta}(x)dx = 1 - P^{(i)}_{\text{ext}}(\tau,\sigma_i,\Theta)$ is the probability of nonextinction of the ith mutant until generation τ. Therefore, in the absence of conditioning on non-extinction, we obtain for the probability that in generation τ the frequency $\tilde{X}_\tau^{(i)}$ of the i-th mutant is less or

equal $p \in [0,1]$:

$$\mathrm{Prob}[\tilde{X}_\tau^{(i)} \le p] = P_{\mathrm{ext}}^{(i)}(\tau, \sigma_i, \Theta) + \int_0^p \tilde{h}_\tau^{(i)}(x)\, dx\,. \tag{3.21}$$

4. Evolution of the phenotypic mean and variance

Based on the results obtained above, we are now in the position to derive approximations for the expected response of the phenotype distribution from first principles. As usual in quantitative genetics, we concentrate on the evolution of the phenotypic mean and the phenotypic variance, i.e., the within-population variance of the trait. For this purpose, we will need the first and second moments pertaining to the density $g_a(x)$. They are given by

$$\gamma_1(a) = \int_0^1 x g_a(x)\, dx = 1 - a e^a E_1(a) \tag{4.1a}$$

and

$$\gamma_2(a) = \int_0^1 x^2 g_a(x)\, dx = 1 + a - a(2+a) e^a E_1(a)\,, \tag{4.1b}$$

respectively, where $E_1(a) = \int_a^\infty x^{-1} e^{-x}\, dx$ is the exponential integral. In addition, we will need the within-population variance of the mutant's allele frequency,

$$\gamma(a) = \int_0^1 x(1-x) g_a(x)\, dx = a(1+a) e^a E_1(a) - a\,. \tag{4.1c}$$

By the properties of the exponential integral, the mean $\gamma_1(a)$ decays from one to zero as a increases from 0 to ∞ (Abramowitz and Stegun, 1964, Chap. 5.1). Together with the properties of $a = a(t, \sigma, N)$, this implies that $\gamma_1(a(t))$ increases from approximately $1/N$ to 1 as t increases from 0 to ∞ (see Fig. 4.1A). For the variance $\gamma(a)$, we find that $\gamma(a) \to 0$ as either $a \to 0$ or $a \to \infty$, and γ assumes a unique maximum at $a \approx 0.719$, where $\gamma(0.719) \approx 0.196$. As function of t, we have $\gamma(a(0)) \approx 1/N$ and $\gamma(a(t)) \to 0$ as $t \to \infty$ (see Fig. 4.1B).

The value t_M of t at which the variance is maximized can be computed from (3.17) with $a \approx 0.719$. Because $P_{\mathrm{ext}}([t], \sigma)$ converges to $P_{\mathrm{ext}}^\infty(\sigma)$ quickly (Sects. 3.2 and A.3), we obtain $t_M \approx \ln(2.8Ns/v)/s$, where v is the variance of the offspring distribution. For a Poisson offspring distribution, $N = 10^4$ and $s = 0.001$ or $s = 0.1$, we obtain $t_M \approx 3.33$ or 7.84, which is in excellent agreement with Fig. 4.1B.

Remark 4.1. For large values of a (such as $a > 80$), we encountered numerical problems when using *Mathematica* to evaluate $\gamma_1(a)$ or $\gamma(a)$. The reason is that it multiplies the then huge value e^a with the tiny value $E_1(a)$, which is set to 0 if it falls below a prescribed value. However, the right-hand side of

$$e^a E_1(a) = \int_1^\infty \frac{1}{x} e^{a(1-x)}\, dx \tag{4.2}$$

can be integrated numerically without problems. An efficient, alternative possibility is to use the approximation $e^a E_1(a) \approx \frac{1}{a}(1-\frac{1}{a})$ (e.g., if $a > 50$), which follows from the asymptotic expansion (D.7).

Beginning with Section 4.3, we define the expected phenotypic mean, $\bar{G}(\tau)$, and the expected phenotypic variance, $V_G(\tau)$, of the trait in generation τ in terms of the distributions of the random variables $\tilde{X}_\tau^{(i)}$ defined in (3.21) and (3.20). The definitions of $\bar{G}(\tau)$ and $V_G(\tau)$ are given in eqs. (4.12) and (4.13). (In the introductory Section 4.1 a simpler random variable can be used.) Therefore, our results below are precise mathematical statements about these quantities, and not statements about the mean and variance of the trait derived from the corresponding Wright-Fisher model. The figures illustrate the accuracy at which Wright-Fisher simulation results can be approximated by our theory, which we base on the *assumption* that allele-frequency distributions evolve as stated in Result 3.4. A Poisson offspring distribution is used only for comparison with the simulation results or when stated explicitly.

4.1. A single locus

We start by investigating the evolutionary dynamics of the trait's mean $\bar{G}_i$ and variance V_{G_i} caused by the contribution of a single locus i. Let α_i be the phenotypic effect of the mutant and $\sigma_i = e^{s\alpha_i}$ its fitness effect. Further, let $X_\tau^{(i)}$ be the random variable describing the mutant's frequency τ generations after its occurrence, conditioned on non-extinction until generation τ. Thus, $X_\tau^{(i)}$ corresponds to X_n defined in (3.13) with $\tau = n$, $\sigma_i = \sigma$, and i indicating the locus. The density of $X_\tau^{(i)}$ is $g_{a(N,\sigma_i,\tau)}$, as given in (3.15). The mutant's expected contribution to the phenotypic mean is $\alpha_i \,\mathrm{E}[X_\tau^{(i)}]$ and to the phenotypic variance it is $\alpha_i^2 \,\mathrm{E}[X_\tau^{(i)}(1-X_\tau^{(i)})]$. By employing equations (4.1a) and (4.1c), we can define the contributions of locus i to the expected mean and the expected variance of the trait by

$$\bar{G}_i(\tau) = \alpha_i \int_0^1 x g_{a(\tau,\sigma_i)}(x)\,dx = \alpha_i \gamma_1(a(\tau,\sigma_i)) \tag{4.3}$$

and

$$V_{G_i}(\tau) = \alpha_i^2 \int_0^1 x(1-x) g_{a(\tau,\sigma_i)}(x)\,dx = \alpha_i^2 \gamma(a(\tau,\sigma_i))\,, \tag{4.4}$$

respectively.

In Figure 4.1 the branching process approximations $\bar{G}_i(\tau)$ and $V_{G_i}(\tau)$ (dashed curves) are compared with simulation results from the diallelic Wright-Fisher model (symbols) and with $\alpha_i p(\tau)$ and $\alpha_i^2 p(\tau)(1-p(\tau))$ (solid curves), where

$$p(\tau) = \frac{\sigma^\tau p(0)}{1+(\sigma^\tau - 1)p(0)} \tag{4.5}$$

is the solution of the discrete, deterministic selection equation $p(\tau+1) = \sigma p(\tau)/\bar{w}(\tau)$, and

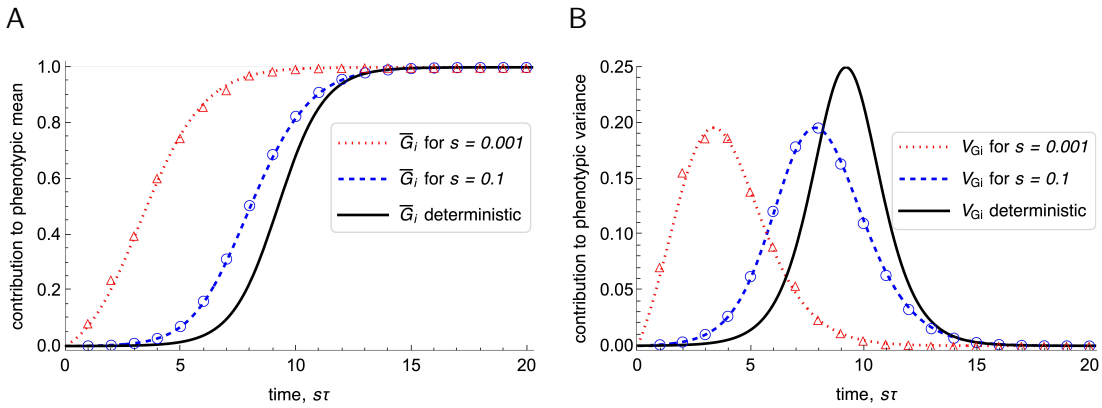

Figure 4.1: Contribution of a single locus to the expected phenotypic mean $\bar{G}_i(\tau)$ (panel A) and to the expected phenotypic variance $V_{G_i}(\tau)$ (panel B), τ generations after the mutation has occurred. The dashed and the dotted curves show $\bar{G}_i(\tau)$ in (4.3) and $V_{G_i}(\tau)$ in (4.4) for $s = 0.1$ and $s = 0.001$, respectively. The symbols show the simulation results from the corresponding Wright-Fisher models. The solid curve shows mean and variance computed from the solution (4.5) of the deterministic model with $p(0) = 1/N$. In this case, the selection coefficient affects the evolution of the mutant frequency, hence of the mean and the variance, only through a proportional change in time scale. Rescaling time by the selection coefficient s makes the deterministic solution independent of s. In a finite population, weaker selection leads to a disproportionally faster spreading of the mutant, which is conditioned on non-extinction up to generation τ. The population size is $N = 10^4$, and the mutation effect is normalized to $\alpha_i = 1$. Because we have $\alpha_i = 1$, $\bar{G}_i(\tau)$ and $V_{G_i}(\tau)$ coincide with the expected mean (4.1a) and variance (4.1c), respectively, of the mutant frequency distribution $g_a(x)$ defined in (3.15). In this and subsequent figures, analytical expressions are displayed as continuous curves.

$\bar{w}(\tau) = 1 + (\sigma - 1)p(\tau)$ is the mean fitness (e.g., Bürger, 2000, p. 30). Figure 4.1 shows excellent agreement between the branching process approximations and the Wright-Fisher simulations. It also demonstrates that in a finite population and conditioned on non-extinction until generation τ, the mutant spreads faster than in a comparable infinite population. The smaller the selective advantage is, the faster the (relative) initial increase of the mutant frequency, and hence of the mean and the variance — on the time scale $s\tau$. This is related to the fact that, to leading order in Ns (for $Ns \gg 1$), the scaled fixation time of a mutant of effect s in a Wright-Fisher model is

$$t_{\text{fix}} \approx \frac{2\ln(2Ns)}{s}\,, \tag{4.6}$$

which is a simplified version of eq. (B.2). For $N = 10^4$ and $s = 0.1$, $s = 0.001$, we obtain from (B.2) $st_{\text{fix}} \approx 16.4$, 7.0, respectively, which provide good estimates for the time to completion of the sweep (Fig. 4.1). Additional numerical support for the expected duration time of t_{fix} is provided by Fig. 4.2. Indeed, the expected duration time (4.6) of a selective sweep has been derived by Etheridge et al. (2006, Lemma 3.1).

The phenomenon described above is well-known (see also Durrett and Schweinsberg, 2004, Uecker and Hermisson, 2011). The intuitive explanation is that mutants that initially spread more rapidly than corresponds to their (small) selective advantage have a much higher probability of becoming fixed than mutants that remain at low frequency for an extended time period. The

smaller the selection coefficient s, the more important a quick early frequency increase is.

4.2. Diffusion approximations and length of the phases of adaptation

Although our results and approximations are derived on the basis of a branching process approximation, the beginnings and lengths of the phases when they are most accurate are best described using diffusion approximations for the (expected) fixation times. This is motivated and supported by the results of Etheridge et al. (2006) and our simulation results. Because in diffusion theory quantities are typically given as functions of the selection coefficient (and the population size), we denote the selection coefficient of a mutant of effect α by

$$\tilde{s}(\alpha) = e^{s\alpha} - 1\,. \tag{4.7}$$

For given N and s, we will need the expected mean fixation time of a single mutant (destined for fixation). The expectation is taken with respect to the distribution f of the mutation effects α which has mean $\bar{\alpha}$:

$$\bar{t}_{\text{fix}}(s, f, N) = \frac{\int_0^\infty t_{\text{fix}}(\tilde{s}(\alpha), N) P_{\text{fix}}(\tilde{s}(\alpha), N) f(\alpha)\, d\alpha}{\int_0^\infty P_{\text{fix}}(\tilde{s}(\alpha), N) f(\alpha)\, d\alpha}\,. \tag{4.8}$$

Here, $t_{\text{fix}}(\tilde{s}(\alpha), N)$ denotes the mean time to fixation of a single mutant with effect α and $P_{\text{fix}}(\tilde{s}(\alpha), N)$ its probability of fixation, both in a population of size N. For comparison with simulation results, we use the well-known diffusion approximation (B.1) for P_{fix} (where (4.10) yields (B.1) if $N_e = N$) and the simplified version (B.4) of the diffusion approximation for t_{fix}. Details and an efficient numerical procedure to compute $\bar{t}_{\text{fix}}(s, \bar{\alpha}, N)$ are described in Appendix B.2. If f is exponential with mean 1, an accurate approximation of $\bar{t}_{\text{fix}}(\tilde{s}(\alpha), N)$ is

$$\bar{t}_{\text{fix}}(s, \bar{\alpha}, N) \approx \frac{2}{s\bar{\alpha}} \ln(2Ns\bar{\alpha}) \tag{4.9}$$

if $2/N \leq s\bar{\alpha} \leq 0.2$ (see Appendix B.2). For $\bar{\alpha} = 1$, the approximation (4.9) corresponds to the simplified approximation (4.6) for equal mutation effects.

We note from (B.1) and (3.8) that $P_{\text{fix}}(\tilde{s}(\alpha), N) > 1 - e^{2\tilde{s}(\alpha)} > P_{\text{sur}}(e^{s\alpha})$ holds for every N, where P_{sur} is based on a Poisson offspring distribution. We emphasize that we distinguish between the fixation probability P_{fix} in the diffusion approximation of the Wright-Fisher model and the survival probability P_{sur} in the Galton-Watson process, although they are quantitatively very similar if Ns is sufficiently large and s small. We use the former only to compute fixation times, whereas the latter is used in our results and their derivation.

Remark 4.2. Our approach can be extended to cases when the effective population size N_e differs from the actual population size N. This occurs if the variance of the offspring distribution differs from its mean. We follow Ewens (2004, p. 120) and define the variance-effective population size

by $N_e = N/(v/\sigma)$ (because we assume large N, we simplified his $N - \frac{1}{2}$ to N). This yields the diffusion approximation

$$P_{\text{fix}}(s, N, N_e) = \frac{1 - e^{-2sN_e/N}}{1 - e^{-2N_e s}} \tag{4.10}$$

instead of (B.1). Numerical comparison of this fixation probability with the survival probability P_{sur} computed for fractional linear offspring distributions from (A.8) shows that these values are almost identical if $Ns > 1$ and $v \geq \sigma/2$. For instance, if $s = 0.1$ (i.e., $\sigma = 1.1$), $N = 1000$, $v = \sigma/2, \sigma, 5\sigma$, we obtain $P_{\text{fix}}(s, N, N_e) \approx 0.3297, 0.1813, 0.0392$, respectively, and $P_{\text{sur}}^{(\text{FL})}(1+s) \approx 0.3333, 0.1818, 0.0392$. For the corresponding Poisson offspring distribution, $P_{\text{sur}}^{(\text{Poi})} \approx 0.1761$ is obtained. Essentially identical numbers are obtained if $N = 100$. Interestingly, the value 0.1761 coincides almost exactly with the true fixation probability of ≈ 0.1761 in the Wright-Fisher model with $N = 1000$. Indeed, it has been proved that the diffusion approximation always overestimates the true fixation probability in the Wright-Fisher model (Bürger and Ewens, 1995). These considerations are an additional reason why we distinguish between P_{sur} and P_{fix}.

4.3. Infinitely many sites

Now we proceed with our infinite-sites model. Let α_i denote the effect of the ith mutation (at locus i) on the trait. Then its fitness effect is $\sigma_i = e^{s\alpha_i}$. The (unconditioned) distribution of the frequency $\tilde{X}_\tau^{(i)}$ of the ith mutant in generation τ was defined in (3.21) and its absolutely continuous part $\tilde{h}_\tau^{(i)}$ in (3.20). Below we shall need the k-th moment of $\tilde{X}_\tau^{(i)}$:

$$\mathrm{E}\left[(\tilde{X}_\tau^{(i)})^k\right] = \int_0^1 x^k \tilde{h}_\tau^{(i)}(x)\, dx\,. \tag{4.11}$$

Now we are in the position to introduce general expressions for the expectations of the mean phenotype $\bar{G}(\tau) = \bar{G}(\tau, f, s, N, \Theta)$ and of the phenotypic variance $V_G(\tau) = V_G(\tau, f, s, N, \Theta)$ in any generation τ, where we recall that the monomorphic population starts to evolve at $\tau = 0$ and mutations occur when $\tau > 0$. The mutation effects α_i are drawn independently from the distribution f. Because we assume linkage equilibrium in every generation τ, the random variables $\tilde{X}_\tau^{(i)}$ $(i = 1, 2, 3, \ldots)$ are mutually independent. For a given sequence of mutation events and given mutant frequency $x_\tau^{(i)}$ at locus i in generation τ, the phenotypic mean is $\sum_i \alpha_i x_\tau^{(i)}$ and the phenotypic variance is $\sum_i \alpha_i^2 x_\tau^{(i)}(1 - x_\tau^{(i)})$. Therefore, by taking expectations with respect to the Poisson distributed mutation events and the evolutionary trajectories of allele frequencies, we define the expected phenotypic mean $\bar{G}$ and variance V_G as follows:

$$\bar{G}(\tau) = \sum_{n=1}^{\infty} \mathrm{Poi}_{\Theta\tau}(n) \left(\sum_{i=1}^{n} \int_0^\infty \alpha_i \,\mathrm{E}\left[\tilde{X}_{\tau, e^{s\alpha_i}, \Theta}^{(i)}\right] f(\alpha_i)\, d\alpha_i \right) \tag{4.12}$$

and

$$V_G(\tau) = \sum_{n=1}^{\infty} \mathrm{Poi}_{\Theta\tau}(n) \left(\sum_{i=1}^{n} \int_0^{\infty} \alpha_i^2 \, \mathrm{E}\left[\tilde{X}^{(i)}_{\tau, e^{s\alpha_i}, \Theta} (1 - \tilde{X}^{(i)}_{\tau, e^{s\alpha_i}, \Theta}) \right] f(\alpha_i) \, d\alpha_i \right) . \tag{4.13}$$

The reader may note that the assumption of linkage equilibrium is not needed for the mean phenotype, and that assuming vanishing pairwise linkage disequilibria would be sufficient for the variance.

Proposition 4.3. *The expected mean and variance of the trait, as defined in* (4.12) *and* (4.13)*, respectively, can be expressed as*

$$\bar{G}(\tau) = \Theta \int_0^{\infty} \alpha f(\alpha) \int_0^{\tau} (1 - P_{\mathrm{ext}}([t], e^{s\alpha})) \gamma_1(a(t, e^{s\alpha})) \, dt \, d\alpha \tag{4.14}$$

and

$$V_G(\tau) = \Theta \int_0^{\infty} \alpha^2 f(\alpha) \int_0^{\tau} (1 - P_{\mathrm{ext}}([t], e^{s\alpha})) \gamma(a(t, e^{s\alpha})) \, dt \, d\alpha \, . \tag{4.15}$$

In particular, the population-wide mutation rate Θ affects both quantities only as a multiplicative factor.

The proof is relegated to Appendix D.

Figure 4.2 displays evolutionary trajectories of the expected mean and variance given in Proposition 4.3 and compares them with simulation results from corresponding Wright-Fisher models. On this scale of resolution the agreement is excellent, both for equal and exponentially distributed effects. After an initial phase the trajectories reach a quasi-stationary phase, in which the phenotypic variance has stabilized and the phenotypic mean increases approximately linearly in time. Similar to the one-locus case treated above, the expected mean fixation times given in the figure caption provide decent approximations for the average time required to enter the quasi-stationary phase. For refined comparisons with more explicit approximations for the quasi-stationary phase and for the initial phase we refer to Fig. 4.3 and to Fig. 4.4, respectively.

Notable in all panels of Fig. 4.2 is the faster increase of the mean and the variance for smaller selection coefficients on the time scale $s\tau$ (see the discussion in the connection with Fig. 4.1). In addition, for the same selection coefficient the evolutionary response is, especially initially, much faster for exponentially distributed mutation effects (panels C and D) than for equal mutation effects (panels A and B), as shown by different steepnesses (A vs. C) and shapes (B vs. D) of the trajectories. The reason is that mutations of larger effect tend to fix earlier and more frequently than mutations of smaller effect. Therefore, large mutation effects speed up the response in the initial phase, whereas small mutation effects delay the entry into the quasi-stationary phase, with both types occurring in an exponential distribution. As shown by the numerical values of st_{fix} and $s\bar{t}_{\mathrm{fix}}$ in the caption of Fig. 4.2, this results in similar mean fixation times for equal and

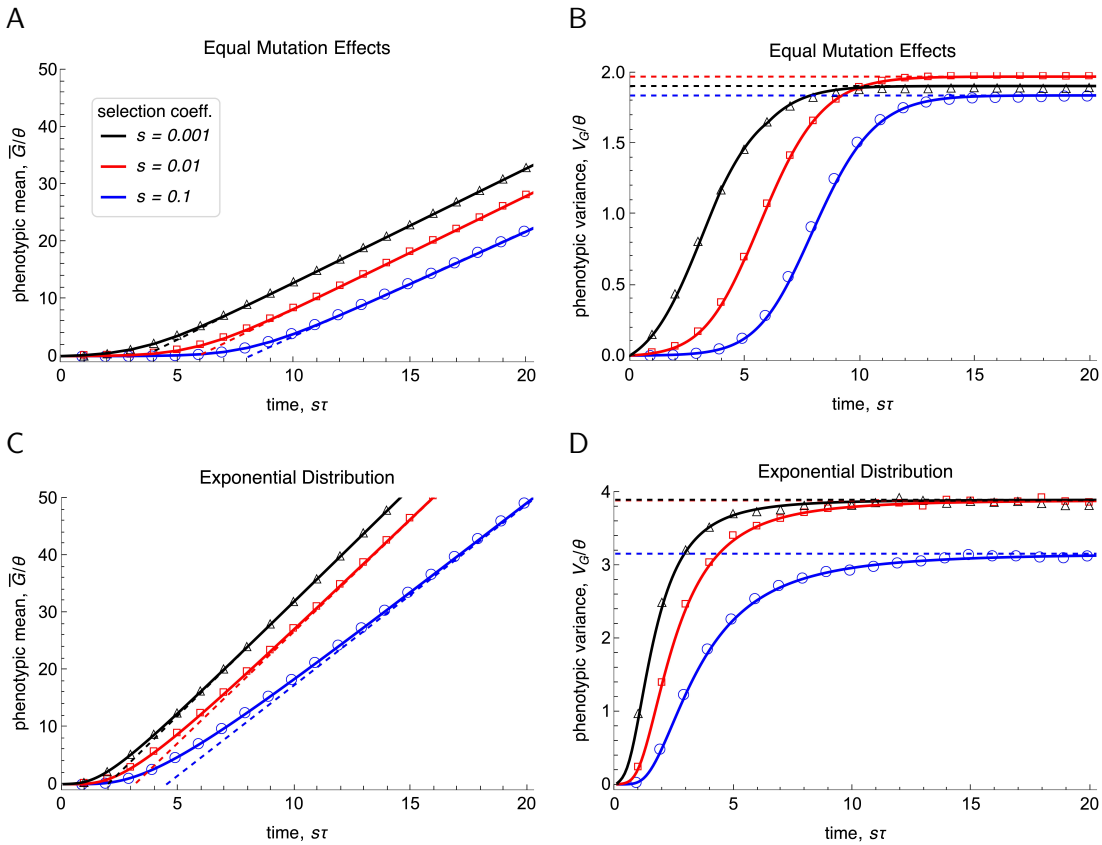

Figure 4.2: The scaled expected phenotypic mean $\bar{G}(s\tau)/\Theta$ (panels A and C) and the scaled expected phenotypic variance $V_G(s\tau)/\Theta$ (panels B and D) are shown for the selection coefficients s given in the legend. Generations τ are scaled by s to facilitate comparison of curves and data with differing s. In A and B all mutation effects are $\alpha_i = 1$, and in C and D the distribution f of mutation effects is exponential with mean $\bar{\alpha} = 1$. The solid curves show $\bar{G}(s\tau)/\Theta$ and $V_G(s\tau)/\Theta$ as computed for equal effects from (4.16) and (4.17), respectively, and for exponentially distributed effects from (4.14) and (4.15). The dashed curves in A and C show the approximation (D.40c) for the mean close to and at stationarity. The dashed lines in B and D show the stationary value of the variance, i.e., (4.22). The values of the mean fixation times multiplied by $s\,(= 0.1, 0.01, 0.001)$ are $st_{\text{fix}} \approx 15.6,\ 11.7,\ 7.0$ for equal effects $\alpha = 1$ in A and B, and $s\bar{t}_{\text{fix}} \approx 15.8,\ 10.6,\ 5.9$ for exponentially distributed effects in C and D, where t_{fix} is calculated from (B.2) and $\bar{t}_{\text{fix}}$ from (4.8). The symbols are from Wright-Fisher simulations with $\Theta = \frac{1}{2}$. The population size is $N = 10^4$. Analogous results for a truncated normal distribution ($\alpha \geq 0$ and $\bar{\alpha} = 1$) are shown in Figure S.5.

for exponentially distributed mutation effects, and thus in similar average entry times into the quasi-stationary phase. Finally, with an exponential distribution the equilibrium variance is about twice as large as for equal effects (see also Corollary 4.8).

Whereas the simulation results for $2\Theta \geq 1$ are accurately described by the analytic approximations for the phenotypic mean and the phenotypic variance (Figure 4.2), for $2\Theta \ll 1$ much more fluctuation is observed in the simulations (Fig. S.2). The reason is that the time between successful mutation events is much larger for smaller Θ, and this results in more pronounced stochastic effects. For small Θ one sweep after another occurs, whereas for large Θ parallel sweeps are common. This will be quantified in Section 5.1 using the number of segregating sites in the population.

Remark 4.4. In general, the integrals in (4.14) and (4.15) cannot be simplified. Of course, for equal effects α, $\int_0^\infty \alpha^k f(\alpha) \int_0^\tau \ldots dt\, d\alpha$ simplifies to $\alpha^k \int_0^\tau \ldots dt$. An alternative approach is to assume that the ith mutation occurs at its mean waiting time i/Θ and the number of mutation events until time τ is (approximately) $[\Theta\tau]$. For given mutation effects α_i, we obtain instead of eqs. (4.14) and (4.15) the simple approximations

$$\bar{G}(\tau) \approx \sum_{i=0}^{[\Theta\tau]-1} (1 - P_{\text{ext}}([i/\Theta], \sigma_i))\bar{G}_i\,(i/\Theta) \tag{4.16}$$

and

$$V_G(\tau) \approx \sum_{i=0}^{[\Theta\tau]-1} (1 - P_{\text{ext}}([i/\Theta], \sigma_i))V_{G_i}\,(i/\Theta)\ , \tag{4.17}$$

respectively, where we have used the one-locus results (4.3) and (4.4). The derivation is given in Appendix C. A meaningful application of these formulas requires $[\Theta\tau] > 1$. For values of Θ smaller than 1, we use linear interpolation between $\Theta = 0$ and $\Theta = 1$ (Appendix C). The approximations (4.16) and (4.17) are evaluated much faster than the integrals in Proposition 4.3 and are applied in Fig. 4.2A,B and Fig. S.2 to the case of equal mutation effects. These figures demonstrate their high accuracy.

In the following we derive simpler and more explicit expressions for the expected phenotypic mean and variance of the trait for different phases of adaptation. First, we characterize their long-term behavior, then we provide simpler approximations of the exact expressions in Proposition 4.3, and finally we derive very simple approximations for their initial increase.

4.4. Approximations for the quasi-stationary phase

Although our approach was designed to study the early phase of adaptation, we show here that it yields accurate and simple approximations for the quasi-stationary phase, when the equilibrium variance has stabilized and the response of the expected mean is approximately linear in time. This phase is reached when the flux of incoming and fixing mutations balances. This is the case after about τ_c generations (see Fig. 4.2), where

$$\tau_c = 1/\Theta + \bar{t}_{\text{fix}}\ , \tag{4.18}$$

is the expected time by which the first mutation, which appeared at (about) $\tau = 1/\Theta$, becomes fixed. If $\tau \geq \tau_c$, then from the $\Theta\tau$ mutations, which are expected to arise until generation τ, $\Theta(\tau - \tau_c) + 1 = \Theta(\tau - \bar{t}_{\text{fix}})$ mutations are expected to have had enough time to go to fixation. Fixed mutations contribute α to the $\bar{G}$ and 0 to the V_G. The mutants from the remaining $\Theta\tau_c - 1 = \Theta\bar{t}_{\text{fix}}$ mutation events either segregate in the population or have been lost before generation τ.

In the results derived below, we will need the following assumptions.

Assumption A1. Let $N \to \infty$ and assume

$$Ns^K = C^K \tag{4.19}$$

where $C > 0$ is an arbitrary constant and the constant K satisfies $K \geq 2$.

Assumption A2. Offspring distributions of mutants of effect α satisfy $v = 2\sigma/\rho$, where ρ is a positive constant, and have uniformly bounded third moments as σ decreases to 1 (Athreya, 1992). In addition, the inequality (3.10) holds.

We note that the scaling Assumption A1 for s and N reflects the fact that, compared to diffusion approximations when $K = 1$, in our model selection is stronger than random genetic drift. In Boenkost et al. (2021), Assumption A1 with $K > 2$ is called moderately strong selection and used to prove that the fixation probability of a single mutant in Cannings models is asymptotically equivalent to $2(\sigma - 1)/v$ (in our notation) as $N \to \infty$. The first requirement in Assumption A2 implies that $\lim_{s\alpha\to\infty} P_{\mathrm{sur}}(e^{s\alpha})/(\rho s\alpha) = 1$ (see Remark 3.3). It is satisfied by fractional linear and by Poisson offspring distributions. As discussed in Sect. 3.2 and shown in Appendix A, equality holds in (3.10) for fractional linear offspring distributions and (3.10) is also fulfilled for Poisson offspring distributions (see also Fig. 3.1).

We denote the values of $\bar{G}(\tau)$ and $V_G(\tau)$ in the quasi-stationary phase by $\bar{G}^*(\tau)$ and V_G^*, respectively. For notational simplicity, we set

$$\nu = \nu(s, \alpha, N) = NP_{\mathrm{sur}}(e^{s\alpha})\,. \tag{4.20}$$

We recall from Section 2.1 that the mutation distribution f has bounded moments. Here is our first main result.

Proposition 4.5. *(1) At stationarity, the expected per-generation response of the phenotypic mean, $\Delta\bar{G}^* = \lim_{\tau\to\infty}(\bar{G}(\tau+1) - \bar{G}(\tau))$, is given by*

$$\Delta\bar{G}^* = \Theta \int_0^\infty \alpha P_{\mathrm{sur}}(e^{s\alpha}) f(\alpha)\, d\alpha\,, \tag{4.21}$$

where $\Delta\bar{G}^*$ depends on Θ, s, and f.

(2) Assuming A1 and A2, the expected phenotypic variance at stationarity, $V_G^* = \lim_{\tau\to\infty} V_G(\tau)$, has the approximation

$$V_G^* = \Theta \int_0^\infty \frac{\alpha}{s} P_{\mathrm{sur}}(e^{s\alpha}) \nu e^\nu E_1(\nu) f(\alpha)\, d\alpha \; + O\Big(N^{-K_1/K}\Big)\,, \tag{4.22}$$

for every choice of $K_1 > 1$ and $K > K_1 + 1$, whence $K > 2$ needs to hold.

The proof of this Proposition is relegated to Appendix D. We discuss the choice of K_1 and K in the error term in Remark 4.9.

Remark 4.6. For mutation effects equal to α, (4.21) simplifies to

$$\Delta\bar{G}^* = \Theta\alpha P_{\mathrm{sur}}(e^{s\alpha})\,, \tag{4.23}$$

and (4.22) simplifies to

$$V_G^*(\alpha, s, N, \Theta) \approx \Theta\frac{\alpha}{s}P_{\mathrm{sur}}(e^{s\alpha})\nu e^{\nu}E_1(\nu)\,. \tag{4.24}$$

Remark 4.7. (1) The expression (4.21) for the expected per-generation response of the mean phenotype reflects the (permanent) response due to mutations that are going to fixation. Therefore, during the quasi-stationary phase, the phenotypic mean increases linearly with time as mutants reach fixation. This response is independent of the population size (in constrast to the early response; see below). In Appendix D (Remark D.1), we provide the approximation (D.40c) for $\bar{G}(\tau, f, s, N, \Theta)$ when τ is sufficiently large. It is shown in Fig. 4.2.

(2) Because only segregating mutations contribute to the phenotypic variance, the number of segregating sites and the phenotypic variance remain constant over time once the quasi-stationary phase has been reached. The term $\frac{\alpha}{s}\nu e^{\nu}E_1(\nu)$ in (4.22) and (4.24) is precisely the total variance accumulated by a mutant with effect α during its sweep to fixation. This follows immediately from (4.4) and (D.3b).

The following result provides much simpler approximations and not only recovers but refines well known results from evolutionary quantitative genetics. For $n \geq 2$ we write $\alpha_n(f) = \int_0^\infty \alpha^n f(\alpha)d\alpha$ for the nth moment about zero of the mutation distribution f.

Corollary 4.8. *We assume a Poisson offspring distribution and the assumptions in Proposition 4.5. Then the following simple approximations hold in the quasi-stationary phase.*

(1) The per-generation change of the mean phenotype at stationarity can be approximated as follows:

$$\Delta\bar{G}^* \approx 2\Theta s\left(\alpha_2(f) - \frac{5s}{6}\alpha_3(f)\right)\,. \tag{4.25}$$

(2) For mutation effects equal to α, the change in the mean is

$$\Delta\bar{G}^* \approx 2\Theta s\alpha^2(1 - \tfrac{5}{6}s\alpha)\,. \tag{4.26}$$

(3) For an exponential distribution f of mutation effects with expectation $\bar{\alpha}$, the change in the mean is

$$\Delta\bar{G}^* \approx 4\Theta s\bar{\alpha}^2(1 - \tfrac{5}{2}s\bar{\alpha})\,. \tag{4.27}$$

(4) The stationary variance has the approximation

$$V_G^*(f, s, N, \Theta) \approx \frac{1}{s}\Delta\bar{G}^* - \frac{\Theta\bar{\alpha}}{Ns}\,. \tag{4.28}$$

(5) If mutation effects are equal to α, then (4.28) *and* (4.26) *yield*

$$V_G^* \approx 2\Theta\alpha^2 \left(1 - \frac{5}{6}s\alpha - \frac{1}{2Ns\alpha}\right) . \tag{4.29}$$

(6) If f is exponential (with mean $\bar{\alpha}$), then (4.28) *and* (4.27) *yield*

$$V_G^* \approx 4\Theta\bar{\alpha}^2 \left(1 - \frac{5}{2}s\bar{\alpha} - \frac{1}{4Ns\bar{\alpha}}\right) . \tag{4.30}$$

The proof is given in Appendix D.5. It is based on series expansions of the expressions given in Proposition 4.5.

Remark 4.9. (1) Assumption A1 yields $s = O(N^{-1/K})$ and $1/(Ns) = O(N^{-1+1/K})$. Thus, s is of lower (or equal) order than $1/(Ns)$ as $N \to \infty$. The error in (4.22) is $O(N^{-K_1/K})$, where $K_1 > 1$ and $K > K_1 + 1$. By the latter assumption, we have $N^{-K_1/K} < N^{-1/K}$, but $N^{-K_1/K} > N^{-1+1/K}$. Thus, the term of order $1/(4Ns\bar{\alpha})$ in (4.30) is not secured because the error term is of larger order. With more accurate estimates in our proof of (4.22) a smaller error term might be obtainable. For illustration, we choose $K_1 = 3/2$ and $K = 3$. Then we obtain $s = O(N^{-1/3})$, $1/(Ns) = O(N^{-2/3})$, and the error term is of order $O(N^{-1/2})$.

(2) Approximations analogous to (4.25) can be obtained from (4.21) for other offspring distributions by series expansion of the survival probability $P_{\text{sur}}(e^{s\alpha})$, or simply from $P_{\text{sur}}(e^{s\alpha}) \approx \rho s\alpha$ (cf. Remark 3.3).

In Figure 4.3 the approximations derived in Proposition 4.5 and Corollary 4.8 (shown as curves) are compared with results from Wright-Fisher simulations (shown as symbols). The figure shows that the approximation (4.22) for the stationary variance (with its simplification (4.24) for equal mutation effects) is highly accurate in a parameter range containing $5 \leq Ns\bar{\alpha} \leq 0.5N$ (dotted curves in A and B), although it has been derived under the assumption $Ns\bar{\alpha} \gg 1$. The corresponding simplified approximations (4.29) and (4.30) (solid curves), derived under the additional assumption $s \ll 1$, are accurate in a range containing $5 \leq Ns\bar{\alpha} \leq 0.1N$. Whereas these approximations correctly predict a decrease in V_G^*/Θ as N decreases, the analytically derived dependence of V_G^*/Θ on s exaggerates the decrease of variance observed for small s in the Wright-Fisher simulations. Indeed, in the limit of $s \to 0$ (neutrality), the true scaled variance V_G^*/Θ should converge to α^2 if mutation effects are equal to α, and to $2\bar{\alpha}^2$ if they are exponentially distributed with mean $\bar{\alpha}$. This is indicated by the Wright-Fisher simulations and follows from the classical result (e.g., Clayton and Robertson, 1955, Lynch and Hill, 1986) that in the absence of selection, the diploid stationary variance is to a close approximation $2\Theta \int_{-\infty}^{\infty} \alpha^2 f(\alpha) d\alpha$, where f is the density of a mutation distribution with arbitrary mean. The missing factors of 2 in comparison to the classical neutral result are due to the assumption of haploidy. The reason why for small s our approximation (4.22) (dotted curves) decays below the neutral prediction is that for weak selection ($Ns\bar{\alpha} \leq 1$), the

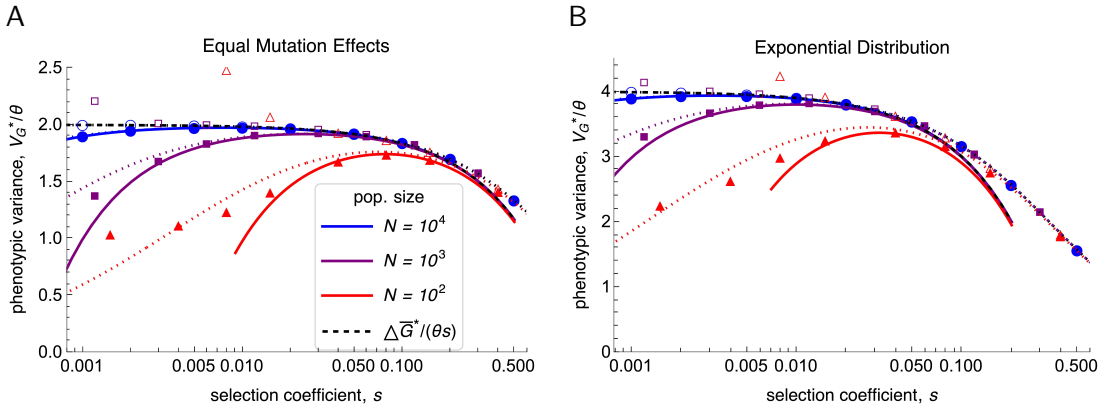

Figure 4.3: The scaled expected phenotypic variance V_G^*/Θ in the quasi-stationary phase for mutation effects equal 1 (panel A) and for exponentially distributed effects with mean 1 (panel B). The solid curves show the simple approximations (4.29) (in A) and (4.30) (in B) as functions of the selection coefficient s for three population sizes N (see legend). The most accurate available approximations are (4.24) and (4.22); they are shown as dotted curves in A and B, respectively. Simulation results from the Wright-Fisher model with $\Theta = \frac{1}{2}$ are shown as symbols (circles for $N = 10^4$, squares for $N = 10^3$, triangles for $N = 100$). In addition, the scaled expected per-generation change in the phenotypic mean, $\Delta\bar{G}^*/(\Theta s)$, is displayed (black dashed curves); it as calculated from (4.26) and (4.27) in (A) and (B), respectively. For the Wright-Fisher simulations, filled symbols are used for the scaled phenotypic variance and open symbols for the change in the scaled phenotypic mean. Note that in a Wright-Fisher model, $\Delta\bar{G}^*/(\Theta s)$ diverges to infinity if $s \to 0$ because neutral mutations can become fixed. Analogous results for a truncated normal distribution ($\alpha \geq 0$) with mean $\bar{\alpha} = 1$ are shown in Figure S.5.

allele-frequency dynamics are increasingly driven by random genetic drift which is not appropriately reflected by the branching-process approximation.

The approximations (4.26) and (4.27) for $\Delta\bar{G}^*$ (black dashed curves in panels A and B of Fig. 4.3) accurately match the response calculated from Wright-Fisher simulations (open symbols) in a range containing $2/N \leq s\bar{\alpha} \leq 0.1$. If higher-order terms in s had been included, then these approximations would be accurate for $s\bar{\alpha}$ up to 0.5. (The dotted curves for the variance are accurate up to this value because they are computed from (4.24) and (4.22), which do not use series expansion in s.) Comparison of the terms of order s^2 in the approximations (4.26) and (4.27) with the term $\frac{\Theta\alpha}{Ns}$ in (4.28) shows that the latter can be neglected in (4.28) if $cs^2\bar{\alpha}^2 \gg 1/N$, where $c = \frac{5}{3}$ for equal mutation effects and $c = 10$ for exponentially distributed effect. The other way around, random genetic drift will distort the classical relation $\Delta\bar{G}^* \approx sV_G^*$ between the response of the mean and the genetic variance if (approximately) $s\bar{\alpha} < \sqrt{10/(cN)}$. For equal mutation effects ($c = \frac{5}{3}$), this becomes $s\bar{\alpha} < 2.4/\sqrt{N}$; for exponentially distribution effects ($c = 10$), it becomes $s\bar{\alpha} < 1/\sqrt{N}$. Both relations confirm the observations in Fig. 4.3, where $\bar{\alpha} = 1$.

Remark 4.10. In quantitative genetic models often a normal distribution with mean 0 is assumed for the mutation effects. To apply our analysis, which ignores mutations with a negative effect, to this setting, we have to adapt this distribution and use a truncated normal distribution instead. If the original normal distribution has variance v_N^2, then the truncated normal (restricted to positive

effects) is defined by $f(\alpha) = \frac{2}{\sqrt{2\pi v_N^2}} \exp(-\alpha^2/(2v_N^2))$ if $\alpha \geq 0$ (and 0 otherwise). Also our mutation rate Θ will correspond to $\Theta/2$ in a quantitative genetic model in which mutations have (symmetric) positive and negative effects. This truncated normal distribution has mean $\bar{\alpha} = \sqrt{2v_N^2/\pi}$. Taking this as the parameter, we obtain $\alpha_2(f) = v_N^2 = \frac{1}{2}\bar{\alpha}^2\pi$ and $\alpha_3(f) = \bar{\alpha}^3\pi$. Therefore, (4.25) yields

$$\Delta\bar{G}^* \approx \pi\Theta s\bar{\alpha}^2(1 - \tfrac{5}{3}s\bar{\alpha}), \tag{4.31}$$

instead of (4.27), and the stationary variance is approximately

$$V_G^* \approx \pi\Theta\bar{\alpha}^2\left(1 - \frac{5}{3}s\bar{\alpha} - \frac{1}{\pi N s\bar{\alpha}}\right). \tag{4.32}$$

After this excursion to the long-term evolution of the trait, we first provide approximations for $\bar{G}(\tau)$ and $V_G(\tau)$ in Proposition 4.3 that can be evaluated expeditiously. Then we turn to the initial phase of adaptation.

4.5. Approximations for the dynamics of $\bar{G}(\tau)$ and $V_G(\tau)$

Proposition 4.11. *We posit Assumptions A1 and A2.*

(1) The phenotypic mean has the approximation

$$\begin{aligned}\bar{G}(\tau, f, s, N, \Theta) = \Theta \int_0^\infty \frac{1}{s} P_{\text{sur}}(e^{s\alpha})\Big(\exp(\nu e^{-s\alpha\tau})E_1(\nu e^{-s\alpha\tau}) - e^\nu E_1(\nu)\Big) f(\alpha)\, d\alpha \\ + O\Big(N^{-1+2/K}\Big) \text{ as } N \to \infty,\end{aligned} \tag{4.33}$$

where $K > 3$ is required.

(2) The phenotypic variance has the approximation

$$\begin{aligned}V_G(\tau, f, s, N, \Theta) = \Theta \int_0^\infty \frac{\alpha}{s} P_{\text{sur}}(e^{s\alpha})\Big(\nu e^\nu E_1(\nu) - \nu e^{-s\alpha\tau}\exp(\nu e^{-s\alpha\tau})E_1(\nu e^{-s\alpha\tau})\Big) f(\alpha)\, d\alpha \\ + O\Big(N^{-K_1/K}\Big) \text{ as } N \to \infty,\end{aligned} \tag{4.34}$$

where $\nu = NP_{\text{sur}}(e^{s\alpha})$, $K_1 > 1$ and $K > K_1 + 1$ (whence $K > 2$ is necessary).

The proof is provided in Appendix D.6. For many offspring distributions, including the Poisson, an insightful simplification is obtained by applying $P_{\text{sur}}(e^{s\alpha}) \approx \rho s\alpha$ for an appropriate constant ρ (cf. Remark 3.3).

We note that the error term is not satisfactory for the very early phase of evolution when τ is still very small. For this very early phase, we derive simple explicit approximations with time-dependent error terms in Proposition 4.13. For large τ, $V_G(\tau, f, s, N, \Theta)$ converges to V_G^* because the time-dependent term in (4.34) vanishes. Therefore, this approximation will be very accurate for large τ. Analogously, $\bar{G}(\tau, f, s, N, \Theta)$ provides an accurate approximation for large τ as shown in Remark D.1 and illustrated in Figure 4.2.C.

Remark 4.12. For equal mutation effects α, the approximations (4.33) and (4.34) simplify to

$$\bar{G}(\tau, f, s, N, \Theta) \approx \frac{\Theta}{s} P_{\mathrm{sur}}(e^{s\alpha}) \Big(\exp(\nu e^{-s\alpha\tau}) E_1(\nu e^{-s\alpha\tau}) - e^{\nu} E_1(\nu) \Big) \tag{4.35}$$

and

$$V_G(\tau, f, s, N, \Theta) \approx \frac{\Theta\alpha}{s} P_{\mathrm{sur}}(e^{s\alpha}) \Big(\nu e^{\nu} E_1(\nu) - \nu e^{-s\alpha\tau} \exp(\nu e^{-s\alpha\tau}) E_1(\nu e^{-s\alpha\tau}) \Big) , \tag{4.36}$$

respectively. Again, it is insightful to keep in mind the approximation $P_{\mathrm{sur}}(e^{s\alpha}) \approx \rho s \alpha$.

4.6. Approximations for the initial phase

For the very early phase of adaptation simple explicit approximations can be derived for $\bar{G}(\tau)$ and $V_G(\tau)$.

Proposition 4.13. *We require Assumptions A1 and A2. In addition, we assume $0 \leq \tau \leq 1/(4\bar{\alpha}s)$ and that the mutation distribution f is exponential with mean $\bar{\alpha}$.*

(1) The phenotypic mean and the phenotypic variance have the approximations

$$\bar{G}(\tau) = \frac{\Theta\tau}{N} \left[\frac{\bar{\alpha}}{1 - \bar{\alpha}s\tau} + O\Big(N^{-1+1/K}\Big) \right] , \tag{4.37}$$

and

$$V_G(\tau) = \frac{2\Theta\bar{\alpha}^2\tau}{N} \left[\frac{1 - s\bar{\alpha}\tau/2}{(1 - s\bar{\alpha}\tau)^2} + O\Big(N^{-1+1/K}\Big) \right] , \tag{4.38}$$

respectively (and the error terms are independent of τ).

(2) For sufficiently small s the leading order terms have the series expansions

$$\bar{G}(\tau) \approx \frac{\Theta\bar{\alpha}\tau}{N} \Big(1 + s\bar{\alpha}\tau + (s\bar{\alpha}\tau)^2 + (s\bar{\alpha}\tau)^3 \Big) \tag{4.39}$$

and

$$V_G(\tau) \approx \frac{2\Theta\bar{\alpha}^2\tau}{N} \Big(1 + \tfrac{3}{2} s\bar{\alpha}\tau + 2(s\bar{\alpha}\tau)^2 + \tfrac{5}{2}(s\bar{\alpha}\tau)^3 \Big) . \tag{4.40}$$

The proof is given in Appendix D.7.

Remark 4.14. If the mutation effects are equal to α, then the approximations

$$\bar{G}(\tau) \approx \frac{1}{\alpha} V_G(\tau) \approx \frac{\Theta}{Ns} \left(e^{s\alpha\tau} - 1 \right) \tag{4.41}$$

are obtained. The errors are of order $\tau\, O\Big((Ns)^{-1}\Big) = \tau\, O\Big(N^{-1+1/K}\Big)$. The validity of this approximation, in particular of the error term, requires only $e^{2s\alpha\tau} = O(1)$, thus effectively $s\alpha\tau = O(1)$ (see Appendix D.7).

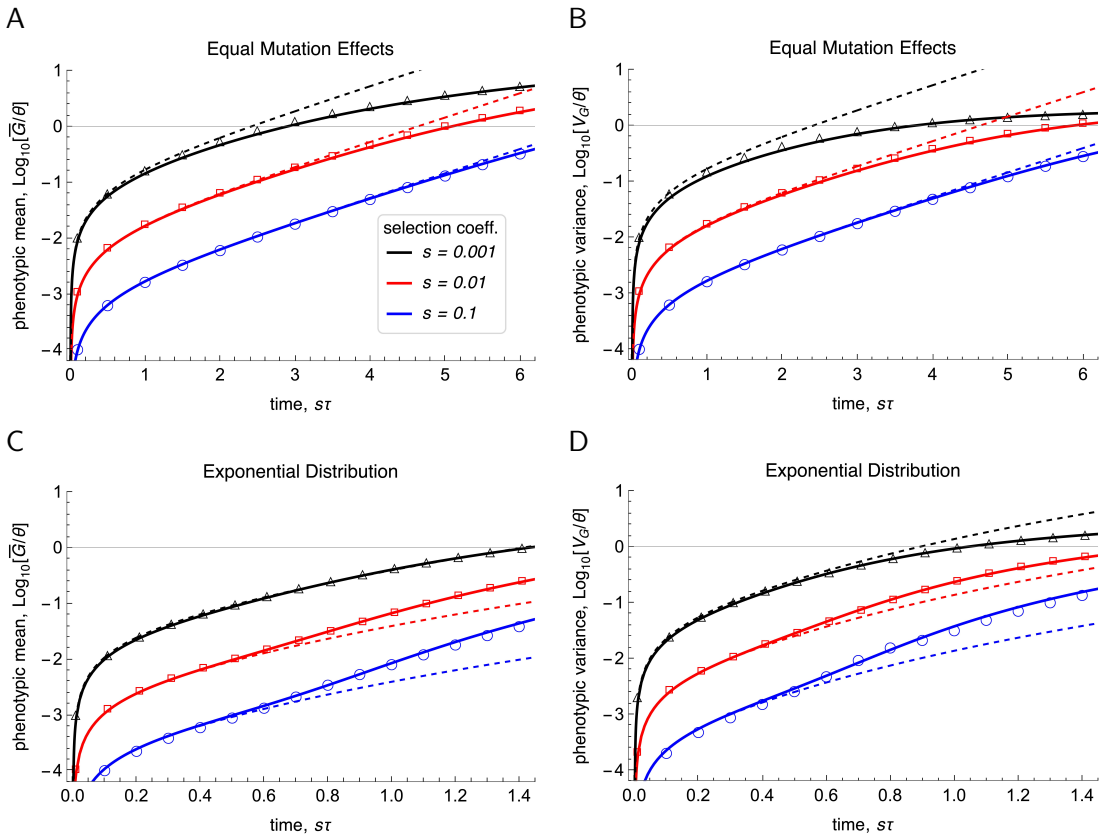

Figure 4.4: The scaled phenotypic mean $\log_{10}(\bar{G}(s\tau)/\Theta)$ (panels A, C) and the scaled phenotypic variance $\log_{10}(V_G(s\tau)/\Theta)$ (panels B, D) in the initial phase for selection coefficients s as given in the legend. For equal mutation effects $\alpha = 1$, the solid curves in A and B show the approximations (4.35) and (4.36), respectively. The dashed curves show the simplified approximation in (4.41). For exponentially distributed effects with $\bar{\alpha} = 1$ the solid curves in C and D show the approximations (4.33) and (4.34), respectively. The dashed curves show the simplified approximations (4.39) and (4.40). In all cases, symbols present results from Wright-Fisher simulations with $\Theta = \frac{1}{2}$, the population size is $N = 10^4$, and the offspring distribution is Poisson.

In Figure 4.4 the approximations for the expected phenotypic mean and variance obtained in Proposition 4.11 and Proposition 4.13 are compared with Wright-Fisher simulations for various selection coefficients. For equal mutation effects (panels A and B) the approximations are accurate in a much wider range of values $s\tau$ than for exponentially distributed effects (panels C and D). In both cases, the more complicated approximations, (4.35) and (4.36) (solid curves in A and B) and (4.33) and (4.34) (solid curves in C and D), are accurate for a much longer time span (in fact for $\tau \to \infty$ as discussed above) than the corresponding simple approximations, (4.41) (dashed curves in A, B) and (4.39), (4.40) (dashed curves in C, D). For an exponential mutation distribution, the simple approximations are accurate if, approximately, $\tau < 1/(2s\bar{\alpha})$. For weak selection, this can still be quite a long time span. The approximations (4.37) and (4.38) are not shown because on this scale of resolution they are almost indistinguishable from the corresponding simple series expansions if $\tau \lessapprox 1/(2s\bar{\alpha})$, but then diverge as $\tau s\bar{\alpha} \to 1$. For large $s\tau$, the approximations in

Proposition 4.11 and Remark 4.12 for $\bar{G}(\tau)$ and $V_G(\tau)$ are almost identical to the exact expressions in Proposition 4.3 (results not shown). Visible differences on the scale of resolution in our figures occur only for small or moderately large $s\tau$ if $s = 0.001$ and $N = 10^4$, i.e., $Ns = 10$ (see Figure S.4).

The approximations (4.37) and (4.38) inform us that initially, as long as $s\bar{\alpha}\tau \ll 1$, $\bar{G}(\tau)$ and $V_G(\tau)$ increase nearly linearly. Comparison with (4.41) shows that for an exponential mutation distribution with mean $\bar{\alpha}$ this early increase is considerably larger than for mutation effects equal to $\bar{\alpha}$. These approximations also show that initially the relation $\Delta\bar{G} \approx sV_G$, typically expected if the trait variance is high, fails; cf. (4.28). We note that (4.37) and (4.38) are always overestimates because their error terms are negative (see Appendix D.7).

5. Sweep-like vs. polygenic shift-like patterns

An important role in examining whether sweep-like patterns or polygenic shift-like patterns characterize the early phase of adaptation is played by the number of segregating sites. If sweeps occur successively, at most one segregating site is expected for most of the time. As the number of segregating sites increases, parallel sweeps occur and the sweep pattern will be transformed to a shift-like pattern. If adaptation is dominated by subtle allele frequency shifts, the number of segregating sites is expected to be large. Of course, there will be intermediate patterns, where no clear distinction between a sweep and a shift pattern can be made. Roughly, we call a pattern of adaptation 'sweep-like' if a mutant has already risen to high frequency, so that its loss is very unlikely, when the next mutant starts its rise to (likely) fixation. We call a pattern shift-like if several or many mutants increase in frequency simultaneously (see Fig. 5.2).

We emphasize that we focus on the patterns occurring during the early phases of adaptation because by our assumption that mutations are beneficial and random genetic drift is weak ($s > 0$ and $Ns \gg 1$), all mutants that become 'established' will sweep to fixation, independently of other parameters such as Θ. In addition, although exponential selection may provide a good approximation for selection occurring in a moving optimum model or by continued truncation selection of constant intensity, it can approximate selection caused by a sudden shift in the phenotypic optimum only during the early phase when the population mean is still sufficiently far away from the new optimum (see Discussion).

5.1. Number of segregating sites

To characterize the patterns of adaptation, an approximation for the expected number of segregating sites, $\mathrm{E}[S]$, is needed. This expectation is taken not only with respect to the distribution of the number of segregating sites, but also with respect to the mutation distribution f, and it assumes a Poisson offspring distribution. For a precise definition see eq. (E.3) in Appendix E.

We assume the infinite sites model, as used in Sects. 4.3 and onwards. However, we complement it by *diffusion approximations* for the fixation and the loss probabilities as well as for the corresponding expected times to fixation and loss (see Appendix B). To derive a simple approximation $\bar{S}$ for $\mathrm{E}[S]$, we distinguish between mutations that eventually become fixed, which occurs with probability $P_{\text{fix}}(\tilde{s}(\alpha), N)$ (see (4.7) for $\tilde{s}(\alpha)$), and those that get lost, which occurs with probability $1 - P_{\text{fix}}(\tilde{s}(\alpha), N)$. In addition, we assume that mutations that are destined for fixation and occurred $t \geq t_{\text{fix}}(\tilde{s}(\alpha), N)$ generations in the past are already fixed. Thus, in a given generation τ, mutations destined for fixation contribute to $\bar{S}$ if and only if they occurred less than or exactly $\min\{t_{\text{fix}}(\tilde{s}(\alpha), N) - 1, \tau\}$ generations in the past (including 0 generations in the past, i.e., now, in generation τ). Therefore, the number of generations in which mutants occur that contribute to the current variation is $\min\{t_{\text{fix}}(\tilde{s}(\alpha), N), \tau + 1\}$. This is consistent with the fact that mutations can already occur in generation $\tau = 0$. Analogously, mutations destined for loss contribute to $\bar{S}$ at time τ only if their time t of occurrence (prior to τ) satisfies $0 \leq t \leq \min\{t_{\text{loss}}(\tilde{s}(\alpha), N) - 1, \tau\}$, where $t_{\text{loss}}(s, N)$ denotes the expected time to loss of a single mutant with selective advantage s in population of size N; see (B.7). In summary, we define the approximation $\bar{S}$ as follows:

$$\begin{aligned}\bar{S}(\tau, f, s, N, \Theta) = \Theta \int_0^\infty & f(\alpha) P_{\text{fix}}(\tilde{s}(\alpha), N) \min\{t_{\text{fix}}(\tilde{s}(\alpha), N), \tau + 1\}\, d\alpha \\ & + \Theta \int_0^\infty f(\alpha)(1 - P_{\text{fix}}(\tilde{s}(\alpha), N)) \min\{t_{\text{loss}}(\tilde{s}(\alpha), N), \tau + 1\}\, d\alpha\,. \end{aligned} \tag{5.1}$$

For the quasi-stationary as well as for the initial phase, we can specify relatively simple explicit approximations by employing diffusion approximtions for P_{fix}, t_{fix}, and t_{loss}. We recall from (4.8) that $\bar{t}_{\text{fix}}(s, \bar{\alpha}, N)$ is the expected time to fixation of a single mutant with effect α on the trait drawn from the (exponential) distribution f; see also Appendix B.2. In the limit of large τ, i.e., $\tau \gg \bar{t}_{\text{fix}}$, we obtain for $\bar{S}^*(f, s, N, \Theta) = \lim_{\tau \to \infty} \bar{S}(\tau, f, s, N, \Theta)$ the following simplifications:

$$\begin{aligned}\bar{S}^*(f, s, N, \Theta) &\approx \Theta \int_0^\infty f(\alpha) P_{\text{fix}}(\tilde{s}(\alpha), N) t_{\text{fix}}(\tilde{s}(\alpha), N)\, d\alpha \\ &\qquad + \Theta \int_0^\infty f(\alpha)(1 - P_{\text{fix}}(\tilde{s}(\alpha), N)) t_{\text{loss}}(\tilde{s}(\alpha), N)\, d\alpha \\ &\approx \begin{cases} 2\Theta\big[(1 - s\bar{\alpha})\big(\ln(N) + \ln(2Ns\bar{\alpha})\big) + 1 + \gamma - (\gamma + \frac{1}{2})s\bar{\alpha}\big] & \text{if } \alpha = \bar{\alpha}\,, \\ 2\Theta\big[(1 - s\bar{\alpha})\big(\ln(N) + \ln(2Ns\bar{\alpha})\big) + 1 - \frac{3}{2}s\bar{\alpha}\big] & \text{if } \alpha \sim \mathrm{Exp}[1/\bar{\alpha}]\,. \end{cases} \end{aligned} \tag{5.2}$$

To evaluate the integrals, we used $\tilde{s}(\alpha) \approx s\alpha$ from (4.7), $P_{\text{fix}}(s, N) \approx 1 - e^{-2s}$ from (B.1), $t_{\text{fix}}(s, N) \approx \frac{2}{s}(\ln(2Ns) + \gamma)$ from (B.2), and $t_{\text{loss}}(s, N) \approx -2(1 + s)\ln(2s) + 2(1 - \gamma) + s(3 - 2\gamma)$ from (B.7) and (B.5a). Finally, a series expansion for small s was performed. The approximate formulas in (5.2) are accurate if $Ns\bar{\alpha} > 2$ and $s\bar{\alpha} < 0.1$; see Figs. 5.1B and D.

We observe that in both cases of (5.2) the leading-order term is approximately $2\Theta \ln(N)$, which

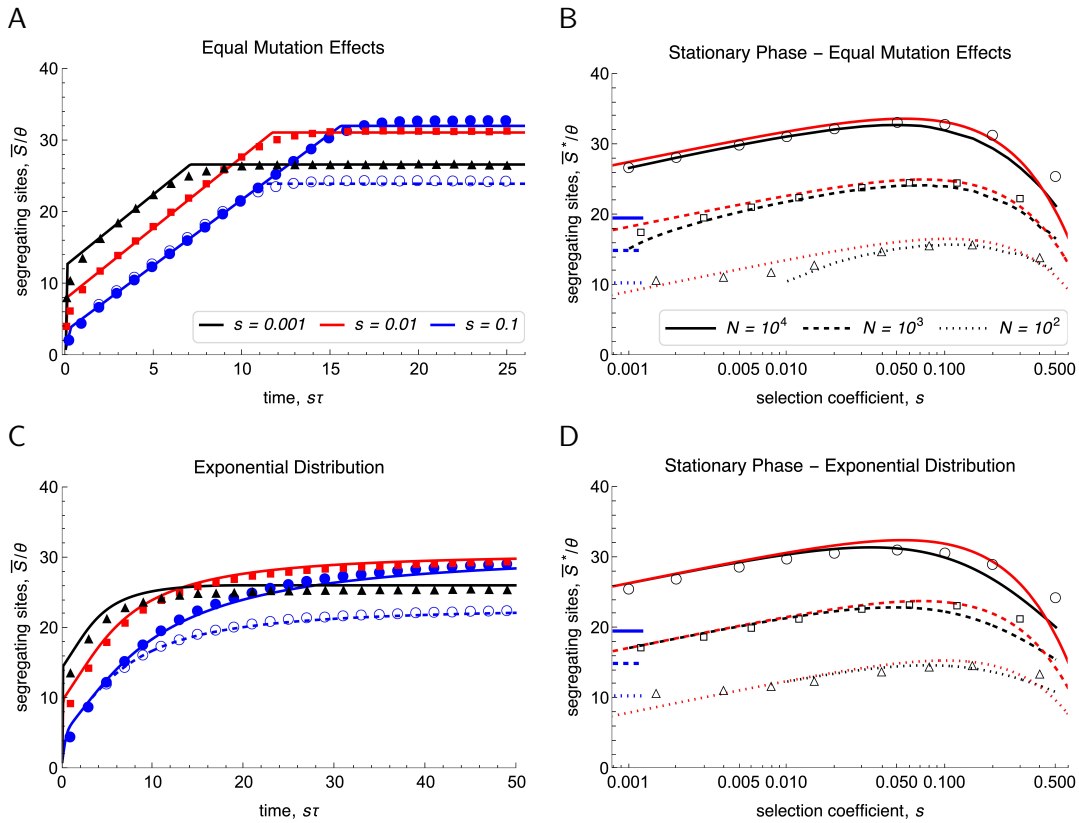

Figure 5.1: The expected number of segregating sites (rescaled by Θ) as a function of $s\tau$ (panels A, C) and at quasi-stationarity as a function of s (panels B, D). In A and B the mutation effects are equal ($\alpha = 1$), in C and D they are drawn from an exponential distribution with mean $\bar{\alpha} = 1$. Analytic approximations are computed from (5.1) and are shown as curves. Wright-Fisher simulation results for $\mathrm{E}[S]$ (with $\Theta = \frac{1}{2}$) are shown as symbols. In panels A and C, $\bar{S}/\Theta$ is shown for $N = 10^4$ and $s = 0.1, 0.01, 0.001$ (solid curves and filled symbols in blue, red, black), and for $N = 10^3$ and $s = 0.1$ (dashed curve and open symbols in blue). In panels B and D, $\bar{S}^*/\Theta$ is displayed for $N = 10^4, 10^3, 10^2$ (solid, dashed, dotted curves, and corresponding open symbols). Black curves represent (5.1), red curves the approximations in (5.2). Under neutrality, we obtain $\mathrm{E}[S] \approx 19.6, 15.0, 10.4$ for $N = 10^4, 10^3, 10^2$, respectively (indicated by the short horizontal blue lines in B and D). Analogous results for a truncated normal distribution ($\alpha \geq 0$) with mean $\bar{\alpha} = 1$ are shown in Figure S.6.

is close to the classical neutral result $\mathrm{E}[S] = \Theta \sum_{i=1}^{N-1} \frac{1}{i} \approx \Theta(\ln(N) + \gamma)$, in which the factor 2 is contained in Θ (and N is the sample size) (Ewens, 1974). Jain and Kaushik (2022) investigated the site frequency spectrum assuming periodic selection coefficients. In the absence of dominance, their approximation (15b) for strong directional selection (i.e., $N\bar{s} \gg 1$, $\bar{s}$ the average selection coefficient) becomes independent of $\bar{s}$ and yields the neutral result, thus essentially our leading order term. We note that for equal mutation effects the stationary value $\bar{S}^*$ of $\mathrm{E}[S]$ is slightly higher than that for exponentially distributed effects.

In analogy to $\bar{t}_{\mathrm{fix}}(s, \bar{\alpha}, N)$, in Appendix B.3 we introduce the expected time to loss, $\bar{t}_{\mathrm{loss}}$, of a single mutant with effect α on the trait drawn from the (exponential) distribution f. If τ is small,

but not tiny, i.e., $\bar{t}_{\text{loss}} < \tau \ll \bar{t}_{\text{fix}}$, then

$$\begin{aligned}\bar{S}(\tau, f, s, N, \Theta) &\approx \Theta(\tau+1)\int_0^\infty f(\alpha)P_{\text{fix}}(\tilde{s}(\alpha), N)\,d\alpha \\ &\quad + \Theta\int_0^\infty f(\alpha)(1-P_{\text{fix}}(\tilde{s}(\alpha), N))t_{\text{loss}}(\tilde{s}(\alpha), N)\,d\alpha \\ &\approx \begin{cases}\Theta\tau(1-e^{-2s\bar{\alpha}}) + 2\Theta\big[-(1-s\bar{\alpha})(\ln(2s\bar{\alpha})+\gamma)+1+\frac{1}{2}s\bar{\alpha}\big] & \text{if } \alpha = \bar{\alpha}\,, \\ \Theta\tau\dfrac{2s\bar{\alpha}}{1+2s\bar{\alpha}} + 2\Theta\big[-(1-s\bar{\alpha})\ln(2s\bar{\alpha})+1+\frac{3}{2}s\bar{\alpha}\big] & \text{if } \alpha \sim \mathrm{Exp}[1/\bar{\alpha}]\,,\end{cases}\end{aligned} \tag{5.3}$$

where the same approximations as for (5.2) are used. If $\tau < \bar{t}_{\text{loss}}$, then (5.1) yields $\bar{S}(\tau, f, s, N, \Theta) \approx \Theta(\tau+1)$, which is an overestimate except when $\tau = 0$ (cf. Fig. E.1).

As Figure 5.1 demonstrates, (5.1) yields a quite accurate approximation for the expected number of segregating sites, both as a function of time (panels A and C), and as function of s at quasi-stationarity (panels B and D). The piecewise linear shape of the curves in A results from the simplifying assumption of taking the minima of τ and t_{loss} or t_{fix}, i.e., the kinks occur at t_{loss} and at t_{fix}. In C, the curves are smooth because effects α are drawn from the distribution f. Panels A and C in Figure 5.1 and Figure E.1 show that as a function of $s\tau$ the rate of increase of the number of segregating sites is nearly independent of s, except for very early generations (e.g. $s\tau < 0.1$). This is also consistent with (5.3). Therefore, if measured in generations, smaller values of s lead to a slower increase of the number of segregating sites, as is supported by the rough approximations in (5.3). The intuitive reason is that invasion becomes less likely. Panels B and D (but also A and C) show that the stationary value $\bar{S}^*$ of the number of segregating sites is maximized at an intermediate selection strength. For small values of s, $\bar{S}^*$ is increasing in s because larger s increases the probability of a sweep. For sufficiently large values of s, $\bar{S}^*$ drops again because selected mutants spend less time segregating.

Based on (5.2), it is straightforward to show that, as a function of s, the maximum value of $\bar{S}^*$ is achieved close to

$$s_{\text{max}} = \frac{1}{\bar{\alpha}\,\mathrm{ProductLog}[2e^b N^2]} \approx \frac{1}{2\bar{\alpha}\ln(N)}\,, \tag{5.4}$$

where $b = \frac{3}{2} + 2\gamma \approx 2.65$ for equal effects and $b = \frac{5}{2}$ for exponentially distributed effects. In each case, the maximum values of $\bar{S}^*$ are very close to $2\Theta(2\ln(N) - \ln\ln(N))$. Thus, the maximum possible value of $\bar{S}^*$ under directional selection is about twice that under neutrality. This shows that once a quasi-stationary response to directional selection has been reached, the number of segregating sites depends only weakly on N, essentially logarithmically, and exceeds the neutral value by at most a factor of two, whatever the strength of selection. In summary, by far the most important factor determining the number of segregating sites at quasi-stationarity is Θ.

Our branching process model leads to an alternative approximation, $\tilde{S}$, for the expected number of segregating sites. If all mutation effects are equal to α, the approximation

$$\tilde{S}(\tau, \alpha, s, N, \Theta) \approx \Theta \sum_{j=0}^{[\tilde{\tau}]} (1 - P_{\text{ext}}(j, e^{s\alpha})) \tag{5.5}$$

is derived in Appendix E. Here, $\tilde{\tau} = \min\{t_{\text{fix}}(\tilde{s}(\alpha), N), \tau\}$. For values $s\tau < 2$, this approximation is considerably more accurate than the approximation $\bar{S}$ given above. This is illustrated by Fig. E.1.A, which is based on a zoomed-in versions of panel A of Fig. 5.1. A disadvantage of the approximation (5.5) is that the terms $P_{\text{ext}}(j, e^{s\alpha})$ have to be computed recursively, whereas the above approximations $\bar{S}$ are based on analytically explicit expressions.

For mutation effects drawn from a distribution f, the following approximation of $\mathrm{E}[S]$ is derived in Appendix E under the assumption $\tau \ll \bar{t}_{\text{fix}}(\tilde{s}(\alpha), N)$:

$$\tilde{S}(\tau, f, s, \Theta) \approx \Theta \sum_{j=0}^{\tau} \left(1 - \int_0^\infty f(\alpha) P_{\text{ext}}(j, e^{s\alpha}) d\alpha\right). \tag{5.6}$$

Figure E.1.B demonstrates that, in contrast to $\bar{S}$ which is based on diffusion approximations, $\tilde{S}$ is highly accurate if $s\tau < 2$. However, it is computationally more expensive than $\bar{S}$.

Equations (5.5) and (5.6) show that the expected number of segregating sites is proportional to Θ. We already know that for large τ, i.e., in the quasi-stationary phase, the number of segregating sites depends at most logarithmically on N. The approximation (5.6) of $\tilde{S}$ is independent of N, and $\tilde{S}$ in (5.5) depends on N only through $\tilde{\tau}$, which can be assumed to equal τ if $\tau \ll \bar{t}_{\text{fix}}$. According to the approximation (4.9) of $\bar{t}_{\text{fix}}$, we have $\bar{t}_{\text{fix}} > 2.77/(s\bar{\alpha})$ if $Ns\bar{\alpha} \geq 2$, because $2 \ln 4 \approx 2.77$. Therefore, we can expect independence of $\tilde{S}(\tau)$ from N if $\tau \leq 2.7/(s\bar{\alpha})$. Comparison of the dashed blue curves with the solid blue curves in panels A and C of Fig. 5.1 shows that the expected number of segregating sites is essentially independent of N for a much longer time. Also the dependence of $\tilde{S}(\tau)$ on s is much weaker than that on Θ.

5.2. Number of segregating sites as an indicator for sweep-like vs. shift-like patterns

Motivated by the work of Höllinger et al. (2019), we explored the polygenic pattern of adaptation at the loci contributing to the trait at the generation T when the mean phenotype $\bar{G}$ in a population first reaches or exceeds a given value, which for Figure 5.2 we chose to be 1 (in units of mutation effects α). For various parameter combinations and for equal and exponentially distributed mutation effects, Figure 5.2 displays the frequency distributions of the first four *successful* mutants at this generation T. By successful mutant we mean that at the stopping time T, we condition on the sites at which mutants are present (segregating or fixed), and we number them from 1 to 4, depending on the order in which these mutations occurred. Thus, the mutation at site 1 is the oldest in

the population. The sites segregating the first four successful mutations are only a small subset of sites at which mutations have occurred because their majority has been lost. The histograms are obtained from Wright-Fisher simulations. We computed the average stopping time, $\bar{T}$, the average number of segregating sites at time T, $\bar{S}_T$, and the average of mean phenotypic values in generation T, $\bar{G}_T$. These values are given in each panel (with subscripts omitted for visibility reasons). Because time is discrete, $\bar{G}_T \geq 1$. For large selection coefficients (and large mutation effects) the per-generation response tends to be large once several mutations are segregating. Therefore, $\bar{G}_T$ can be noticeably larger than 1. Indeed, the distribution of $\bar{G}$ (under the stopping condition $\bar{G} = 1$) has a sharp lower bound at 1 and may have substantial skew to the right, especially for exponentially distributed effects (results not shown).

Figure 5.2 shows clear sweep patterns in the panels with $2\Theta = 0.1$ (where $\bar{S}_T < 2.5$) and distinctive small-shift patterns at many sites if $2\Theta = 10$ (where $\bar{S}_T > 35$). The patterns occurring for $2\Theta = 1$ suggest a series of partially overlapping sweeps. The main effect of varying the selection strength s is on the time T to reach (or exceed) $\bar{G} = 1$. Successive sweeps are indeed expected if $2\Theta = 0.1$ because the average waiting time between successful mutation events is approximately $1/(2s\Theta)$. If $s = 0.001$, this waiting time is $10\,000$ and exceeds the expected fixation time of roughly $7\,000$ generations. If $s = 0.1$, the waiting time is 100 generations and falls short of the expected fixation time, which is about 156 generations by (4.9). In addition, we observe that for small Θ the first successful mutant corresponds mostly to the site with the highest mutant frequency at the time of stopping, whereas for large Θ the distributions of the first four successful mutations are hardly distinguishable.

In the panels with $2\Theta = 0.1$, the expected stopping times $\bar{T}$ exceed their corresponding expected fixation times (given in the caption of Fig. 5.2) considerably. Hence, most of these populations have already reached the quasi-stationary phase. Somewhat unexpectedly, the observed average numbers of segregating sites (from top to bottom: 2.2, 2.1, 1.9) exceed their respective (approximate) stationary values $\bar{S}^*$ (1.37, 1.65, 1.60) considerably. This observation seems to be at variance with the results in panels B and D of Figure 5.1, which show that $\bar{S}^*$ is a very accurate approximation of the numerically determined $\mathrm{E}[S]$. The reason for this discrepancy is the different stopping condition, more precisely the fact that the distribution of stopping times T is very wide and strongly skewed to the right if Θ is small (Fig. S.7). A slowly evolving population may carry many more mutations once they reach $\bar{G} = 1$ than the average or a fast evolving population because it has to wait long for the first successful mutant(s). As a result, the stopping condition $\bar{G} = 1$ inflates the number of segregating sites compared to a process that is stopped at a given generation.

For $2\Theta \geq 1$ our branching-process approximations for $\tilde{S}$ in (5.5) and (5.6) predict the observed average $\bar{S}_T$ very well if $\tilde{S}$ is evaluated at $\bar{T}$. Indeed, very accurate approximations of the observed

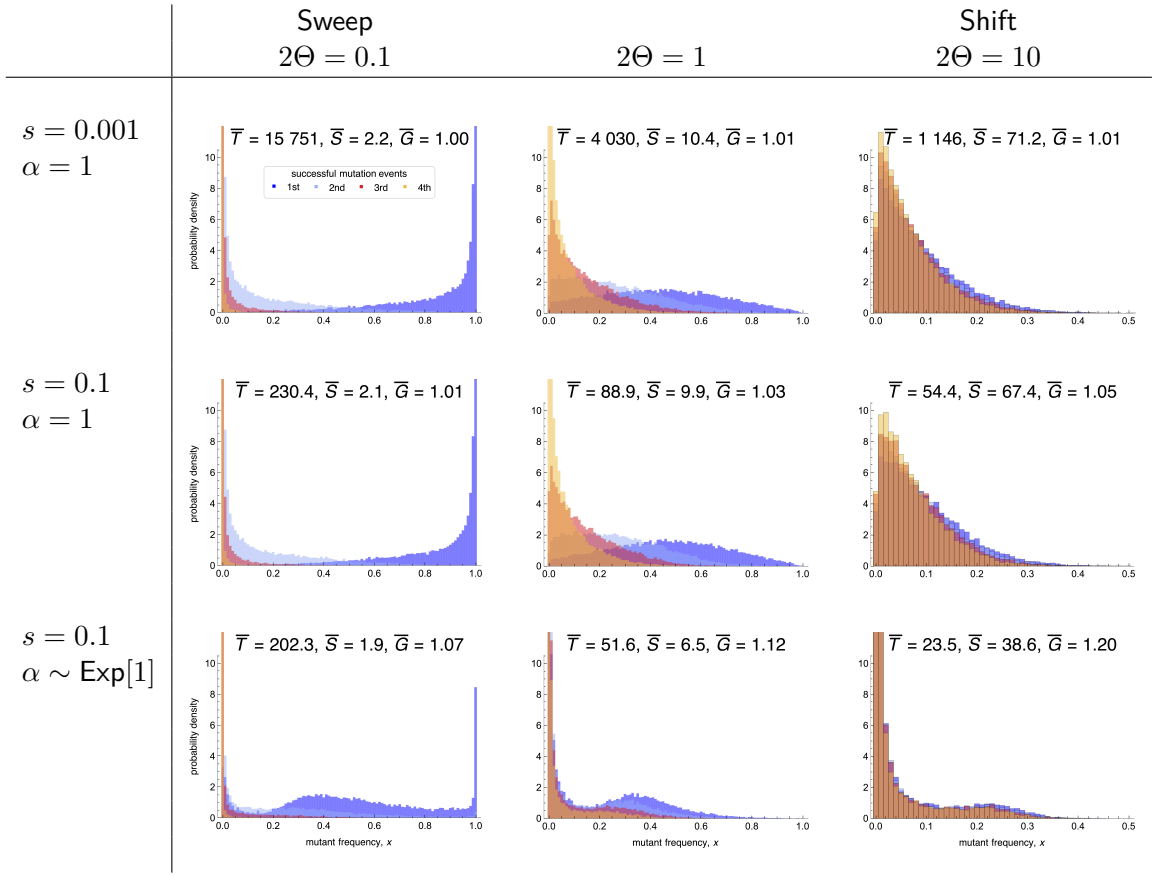

Figure 5.2: Histograms of the frequency x of the first four successful mutants at the first generation T in which the mean phenotype $\bar{G}$ in a single population reaches or exceeds the value 1. The averages of the stopping time T, of the number of segregating sites at T, and of the mean phenotypic values at T, denoted by $\bar{T}$, $\bar{S}(= \bar{S}_T)$, and $\bar{G}(= \bar{G}_T)$, are given in each panel. Histograms and averages are based on $20\,000$ replicate runs of the Wright-Fisher model. The dark blue color is for the first successful mutation, light blue for the second, red for the third, and orange for the fourth successful mutation. The selection coefficients s and the mutation effects α are indicated for each row. Columns show results for different values of Θ. The population size is $N = 10^4$ in all cases. One unit on the vertical axis corresponds to 200 runs. The expected times to fixation for rows 1, 2, and 3 are $t_{\text{fix}} \approx 7\,039$, $t_{\text{fix}} \approx 156$, and $\bar{t}_{\text{fix}} \approx 153$, respectively. The corresponding scaled stationary values of the number of segregating sites, $S^*/(2\Theta)$, are approximately 13.7, 16.5, and 16.0 (calculated from eq. (5.2)). Thus, in the first column S has reached its stationary value in rows 1 and 2, and 60% of it in row 3. The distributions of the stopping times T corresponding to the above panels are shown in Fig. S.7.

$\bar{T}$ are obtained for equal effects by solving the equation $\bar{G}(t) = 1$ using (4.35). The reason is that the distribution of T becomes increasingly symmetric and less variable as Θ increases. It is nearly perfectly symmetric if $2\Theta = 10$ (Fig. S.7). For exponentially distributed effects, when the stopping time distribution is more skewed, an accurate approximation of $\bar{T}$ is obtained from (4.33) by solving $\bar{G}(t) = 1$ only if $2\Theta \geq 10$ (results not shown). Accordingly, for such large values of Θ, the Wright-Fisher simulations stopped at the previously determined value $\bar{T}$ yield distributions very similar to those obtained by stopping when $\bar{G} = 1$ has been reached (results not shown).

Figure 5.2 confirms the above discussed finding that for an exponential distribution of mutation

effects, the response of the mean is faster than for equal effects also under the current stopping condition. As a consequence, the number of segregating sites when $\bar{G} = 1$ is reached is smaller, and considerably so if $\bar{G} = 1$ is reached well before the quasi-stationary phase is entered, as is the case of $2\Theta \geq 1$. Unsurprisingly, the distribution of stopping times T is much wider than for equal effects (Fig. S.7) because mutations of larger (beneficial) effect have a higher fixation probability and lead to faster sweeps, especially for strong selection. Mutation distributions with a heavier tail than the exponential may intensify this effect.

6. Discussion

The goal of our study has been to provide essential theory for answering the following questions: Which are the main evolutionary and genetic factors determining the rate and pattern of polygenic adaptation? How do these factors determine the response to selection, both on the phenotypic and on the genetic level? When is the selective response mainly due to sweeps at few loci and when is it due to small frequency shifts at many loci?

For this purpose, we developed and analysed a simple population-genetic model, starting from a branching process approximation to describe the spreading of a beneficial mutant in a Wright-Fisher population of size N. Our basic result is an accurate, analytically tractable, and computationally efficient approximation for the time dependence of the mutant-frequency distribution. Extension to an infinite sites model and the assumption of an additive quantitative trait under exponential directional selection enabled us to derive explicit expressions for the evolution of the expected mean phenotype and the expected phenotypic variance, as well as for the expected number of segregating sites underlying the adaptive response. As demonstrated by comparison with extensive simulations, they provide accurate approximations for the corresponding dynamics resulting from a Wright-Fisher model. Our theory focuses on the response from new mutations and assumes that selection is stronger than random genetic drift ($Ns > 1$). In the sequel, and structured according to the main sections (3, 4, and 5) of the paper, we discuss the most relevant findings and their implications for answering the last question posed above. Finally, we provide brief conclusions and discuss the limitations caused by our model assumptions (Sect. 6.4).

6.1. Branching process approximation for the time dependence of the mutant frequency distribution

The spread of a new mutation in a population as a function of time can be described by the transition matrix of an appropriate Wright-Fisher model or the transition density of the approximating diffusion. Although diffusion theory leads to quite simple expressions for many important quantities, explicit and analytically tractable time-dependent results, tracing for instance the distribution of allele frequencies, seem to be out of reach (see Steinrücken et al., 2013, for a semi-explicit approach).

We combined two methods to derive an explicit and accurate time-dependent approximation for the mutant's density in any generation n: a branching process approach capturing the stochastic effects and the deterministic logistic growth model. We developed this approach quite generally with the help of a slightly supercritical Galton-Watson process that allows for quite general offspring distributions.

The approximation $g_a(x)$ in (3.15) for the mutant frequency density depends only on the compound parameter a, which in turn depends on the generation number n, the mean $\sigma = e^s$ of the offspring distribution, the extinction probability $P_{\text{ext}}(n, \sigma)$ until generation n in the branching process, and the population size N. Its shape is displayed in Figure 3.3 and compared with Wright-Fisher simulations in Figures 3.4 and S.3. It is highly accurate in the initial phase, and requires only $Ns \geq 1$. If $Ns \geq 10$ then it is accurate until about $\frac{1}{s}\ln(2Ns)$, which is approximately one half of the expected fixation time in the Wright-Fisher model. If $Ns \geq 100$, then it remains accurate for even longer and even if x is relatively close to 1. Because, in contrast to previous related investigations (Desai and Fisher, 2007, Uecker and Hermisson, 2011, Martin and Lambert, 2015), we condition on non-extinction until generation n, our approximation is more accurate than those obtained there, especially in the early initial phase (cf. Figure 3.2, which shows the density ψ_n in the branching process, from which g_a is obtained by a proper transformation). However, by relying on a branching process to capture the inherent stochasticity, our model underestimates the effect of genetic drift near fixation, in particular if Ns is not much larger than 1. Remarkably, the approximations for the mean and variance of this distribution are accurate in a much wider parameter range. This may be explained by the fact that mutants which are rare or near fixation hardly contribute to the variance and to the response of the mean. These approximations clearly display the well known fact that, conditioned on non-extinction (in our case, until the time of observation), the mutant frequency grows faster than predicted by the corresponding deterministic model, particularly when selection is weak (see Fig. 4.1 and the discussion in Sect. 3).

Finally, these results are extended to the infinite sites model, in which the number of new (beneficial) mutants emerging in the population per generation is Poisson distributed with mean $\Theta = NU$, and every mutation occurs at a new site. Here, U is the expected number of beneficial mutations per individual per generation affecting the trait. Sites are assumed to be unlinked (and in linkage equilibrium). The monomorphic initial population started to evolve at $\tau = 0$. Equation (3.18) provides the probability density $h^{(i)}(x) = h^{(i)}_{\tau,\sigma_i,N,\Theta}(x)$ of the ith mutation (which occurred at the random time τ_i at the ith site), conditioned on non-extinction until time τ. Mutations of unequal effects are admitted, as signified by σ_i. The integration is necessary because the waiting time distribution for the ith mutation event (an Erlang distribution) needs to be taken into account. Its cousin, the unconditioned distribution $\tilde{X}^{(i)}_\tau$ with absolutely continuous part $\tilde{h}^{(i)}_\tau(x)$ is defined in

(3.20) and (3.21). It forms the basis for the applications to the quantitative genetic model.

6.2. Evolution of the mean phenotype and the genetic variance

The approximations (3.15) and (3.21) for the time dependence of the density $g_{a(\tau)}$ and of the distribution of $\tilde{X}_\tau^{(i)}$ allow us to obtain highly accurate approximations for the evolution of the expected mean ($\bar{G}$) and variance (V_G) of an additive trait subject to exponential directional selection in a finite population of size N. We assume that potentially infinitely many, stochastically independent, loci (sites) can contribute additively to the trait. This assumption of global linkage equilibrium will be a good approximation if sites are physically unlinked because multiplicative fitnesses, as caused by exponential directional selection, do not generate linkage disequilibrium (Bürger, 2000, p. 79).

The expressions (4.14) for the expected mean $\bar{G}(\tau)$ and (4.15) for the expected variance $V_G(\tau)$ of the trait are exact within our model based on the quasi-deterministic branching process approach, i.e., assuming that the density g_a in (3.15) is exact. Comparison with the results from Wright-Fisher simulations shows that they are astonishingly accurate for the whole evolutionary process, i.e., from the initial phase until the quasi-stationary response has been achieved (Figure 4.2). These expressions require integration with respect to two parameters: one over the time span until the time τ of interest, the other with respect to the distribution $f(\alpha)$ of mutation effects. The former is necessary because the extinction probability is time dependent, at least in the initial phase. The latter integration is unavoidable unless all mutation effects are equal, when accurate approximations involve only the computation of the finite sums given in Remark 4.4. We note that these approximations hold for very general offspring distributions, so that they can cover cases where the effective population size N_e differs from the actual population size N. The offspring distribution enters these equations through the extinction probabilities P_{ext}.

The general equations (4.14) and (4.15) do not provide much direct insight into the evolution of $\bar{G}$ and V_G, except that the population-wide mutation rate Θ enters exclusively as a multiplicative factor. Their numerical evaluation is straightforward but not very fast. However, they provide the basis for deriving very simple and highly accurate analytical approximations for the initial phase and for the quasi-stationary phase, when emergence and fixation of new mutations balance and the genetic variance has reached a stationary value.

6.2.1. The quasi-stationary phase

Figure 4.2 documents that the quasi-stationary phase is reached once time exceeds the expected fixation time $\bar{t}_{\text{fix}}$, which is defined in (4.8) and has the approximation (4.9). During the quasi-stationary phase the expected per-generation change in the mean, $\Delta\bar{G}^*$, and the expected variance, V_G^*, are constant. They are given by (4.21) and (4.22), respectively, and invoke the generally unavoidable integration over the distribution of mutation effects. For equal effects, the much

simpler expressions (4.23) and (4.24) are obtained. By recalling $\Theta = NU$, equation (4.28) reveals the following general and fundamental relation between the change of the mean and the variance:

$$\Delta\bar{G}^* \approx sV_G^* + U\bar{\alpha}\,. \tag{6.1}$$

If $U\bar{\alpha} \approx 0$, this is a special case of Lande's (1976) equation, which was derived under the assumption of a normal distribution of breeding values. An approximation of $\bar{G}(\tau)$, applicable already during the approach to the quasi-stationary phase, is provided in (D.40c), and shown as dashed curves in Figs. 4.2 A,C.

For a Poisson offspring distribution, an exponential distribution f of mutation effects, and weak selection, the response of the mean has the simple approximation $\Delta\bar{G}^* \approx 4\Theta s\bar{\alpha}^2(1 - \frac{5}{2}s\bar{\alpha})$. If the mutation effects are equal, the response is about one half of this value (Corollary 4.8). For large population sizes, $N \geq 10^4$, the simple approximations in Corollary 4.8, both for $\Delta\bar{G}^*$ and V_G^*, are highly accurate for selection coefficients s ranging over more than three orders of magnitude (Figure 4.3). The relation (6.1) is violated when the population size is so small that $Ns < 2$, as can be observed in Fig. 4.3 by noting the deviation of the open symbols from the black dashed curves. The reason is that for such small values of Ns random genetic drift becomes an important force and the fixation probability of very slightly beneficial mutations is higher than predicted by the branching process approximation, i.e., in this case the fixation probability under neutrality, $1/N$, exceeds Haldane's approximation, $2s$, for an advantageous mutation.

Formally, the per-generation change of the mean phenotype at stationarity in (4.21) is equivalent to Hill's (1982) approximation for the asymptotic response of a quantitative trait coming from fixation of new mutations. In our terminology and for a haploid population, his general expression [3] becomes $\Delta\bar{G} = \Theta(N_e/N)\int_{-\infty}^{\infty} \alpha P_{\text{fix}}(s\alpha)f(\alpha)\,d\alpha$, where his mutation distribution is not necessarily confined to positive values. Here, P_{fix} is the diffusion approximation (B.1) for the fixation probability in the Wright-Fisher model and thus nearly identical to P_{sur} because we assume $Ns > 1$. Thus, we recover classical results, but from a different perspective. It seems somewhat surprising that the branching process approach, which is best suited to study the initial fate of a mutation subject to selection, yields highly accurate approximations for the long-term response of a quantitative trait. For a more general discussion of the long-term response to various forms of directional selection, see Bürger (2000, pp. 331–338).

6.2.2. The transient and the initial phase

In Proposition 4.11, we derived the approximations (4.33) for $\bar{G}(\tau)$ and (4.34) for $V_G(\tau)$. They can be evaluated much more quickly than (4.14) and (4.15), and they are highly accurate for all times, except for a very early phase (e.g., Fig. S.4). They perform particularly well when $\tau > 1/\Theta + T_\epsilon$, where T_ϵ is the time needed such that the probability of non-extinction by generation

T_ϵ has declined to $(1+\epsilon)P_{\rm sur}$ (eq. (A.21) and Section 3.2). Above this lower bound the (iterative) computation of the time-dependent extinction probabilities $P_{\rm ext}(t)$ in the general equations (4.14) and (4.15), hence the integration over the time, is not needed. Thus, for equal mutation effects, no integration at all is required to evaluate $\bar{G}(\tau)$ and $V_G(\tau)$, which are then given by (4.35) and (4.36), respectively.

For the very early phase when $\tau \lessapprox 1/(4s\bar{\alpha})$, the utterly simple approximations in Proposition 4.13 are obtained for exponentially distributed mutation effects. The series expansions in (4.39) and (4.40) are accurate for a somewhat longer initial phase, even beyond $\tau \approx 1/(s\bar{\alpha})$ provided s is sufficiently small (Figure 4.4). For mutation effects equal to α, the mean phenotype grows approximately exponentially, i.e., $\bar{G}(\tau) \approx \frac{\Theta}{Ns}(e^{s\alpha\tau}-1)$ (Remark 4.14). This yields an accurate approximation if $\tau \lessapprox 1/(s\bar{\alpha})$ for $Ns = 10$ and for a longer time span as Ns increases while s decreases (Figure 4.4). During this phase, variance and mean are related through $V_G \approx \alpha\bar{G}$ and the relation $\Delta\bar{G} \approx sV_G$ is violated. For an exponential distribution of mutation effects, the initial increase of the mean and the variance is considerably faster than for equal mutation effects ($=\bar{\alpha}$), indeed we have $\bar{G}(\tau) \approx \frac{\Theta}{N}\bar{\alpha}\tau\Big(1+s\bar{\alpha}\tau+(s\bar{\alpha}\tau)^2\Big)$ (Proposition 4.13). This yields an increase of the mean that is slightly faster than that for equal effects with twice the selection intensity, $2s\bar{\alpha}$. The reason is that with an exponential distribution of effects, mutations of large effect have a much higher probability of becoming established (and eventually fixed) than mutations of small effect. In fact, the average effect of mutations drawn from an exponential distribution with mean $\bar{\alpha}$, conditioned on eventual fixation, is approximately $2\bar{\alpha}$.

6.2.3. Comparison of the response from new mutations with that predicted by the breeder's equation

We developed our model for an initially monomorphic population and adaptation from *de novo* mutations. Here, we compare our results on the response of the mean phenotype with corresponding results derived from classical quantitative genetic models (thus, essentially statistical models) and provide a rough guide for which parameter regions they conform.

In the classical additive model of quantitative genetics, environmental effects are not ignored, and it is based on a given phenotypic variance. More precisely, the trait value is $P = G + E$, where G is the (additive) genetic contribution and E the environmental contribution. The simplest version assumes that environmental effects have a distribution with mean zero and variance σ_E^2 and are independent of the genetic effects, so that $\Delta\bar{P} = \Delta\bar{G}$ and $\sigma_P^2 = \sigma_G^2 + \sigma_E^2$. The classical breeder's equation, which reads

$$\Delta\bar{P} = h^2\Delta_s\bar{P} = ih^2\sigma_P\,, \tag{6.2}$$

where $h^2 = \sigma_G^2/\sigma_P^2$ is the (narrow sense) heritability of the trait (and σ_G^2 the additive genetic

variance, which corresponds to our V_G), $\Delta_s \bar{P}$ is the selection differential, i.e., the difference in mean phenotype before and after selection, and $i = \Delta_s \bar{P}/\sigma_P$ is the so-called selection intensity, or the standardized selection differential (e.g., Bürger, 2000, pp. 317-318). Under exponential directional selection one obtains $i = s\sigma_P$ (Bürger, 2000, p. 320), so that (6.2) becomes

$$\Delta \bar{P} = h^2 s \sigma_P^2 \,. \tag{6.3}$$

Stationary phase. Our equation (4.25) informs us that, once stationarity is achieved, the response from new mutations is approximately

$$\Delta \bar{G}^* \approx \rho \Theta s \alpha_2(f) = \rho \Theta s \bar{\alpha}^2 (1 + \mathrm{CV}_\alpha^2) \,, \tag{6.4}$$

where we have used $P_{\mathrm{sur}}(e^{s\alpha}) \approx \rho s \alpha$ (Remark 3.3). If the offspring distribution is Poisson, then $\rho = 2$ and $N_e \approx N$. Here, CV_α^2 is the squared coefficient of variation of the distribution of mutation effects. For equal effects $\mathrm{CV}_\alpha^2 = 0$, for exponentially distributed effects $\mathrm{CV}_\alpha^2 = 1$, for Gamma-distributed effects with density proportional to $\alpha^{\beta-1} e^{-\alpha}$, $\mathrm{CV}_\alpha^2 = 1/\beta$, and for effects drawn from the truncated normal distribution in Remark 4.10, $\mathrm{CV}_\alpha^2 = \frac{1}{2}\pi - 1$. As already noted above, an equation equivalent to (6.4) was derived and discussed by Hill (1982a,b).

Equation (6.4) demonstrates that the response of the mean depends not only on the selection intensity, the mutation rate and the average effect of beneficial mutations, but also on the shape of the mutation distribution, in particular its coefficient of variation. Therefore, distributions with a large coefficient of variation of beneficial effects, such as Gamma-distributions with small β, lead to a faster response. This effect is already clearly visible by comparing panels A and C in Fig. 4.2, where in A we have $\mathrm{CV}_\alpha^2 = 0$ and in B, $\mathrm{CV}_\alpha^2 = 1$; see also panel A in Fig. S.5, where $\mathrm{CV}_\alpha^2 = \frac{1}{2}\pi - 1 \approx 0.507$.

Assuming a Poisson offspring distribution and exponentially distributed mutation effects with mean $\bar{\alpha}$, the selection response predicted by (6.1) equals that predicted by (6.3) if

$$4\Theta \bar{\alpha}^2 = h^2 \sigma_P^2 \,, \tag{6.5}$$

where, in the stationary phase and by (4.30), $\sigma_P^2 = V_G^* + \sigma_E^2 \approx 4\Theta\bar{\alpha}^2 + \sigma_E^2$. Inferences from empirical estimates suggest that $\bar{\alpha}$ may often be not only of the same order of magnitude as σ_P, but actually quite similar to it (Walsh and Lynch, 2018, p. 1019). In fact, they suggest $\sigma_\alpha^2 \approx \sigma_E^2$, where σ_α^2 is the variance of a symmetric distribution of mutation effects with mean zero and σ_E^2 is the environmental variance. Under our assumption of exponentially distributed (positive) effects, we obtain $\sigma_\alpha^2 = 2\bar{\alpha}^2$, which yields $\sigma_P^2 = 4\Theta\bar{\alpha}^2 + 2\bar{\alpha}^2$. Therefore, (6.5) is satisfied if $h^2 = 2\Theta/(2\Theta + 1)$. Heritabilities are, by definition, between 0 and 1, but are often in the range between 0.1 and 0.7 (e.g., Table 10.1 in Falconer and Mackay, 1996; Chap. 7 in Lynch and Walsh,

1998), for morphological traits more often in the lower half of this range. Thus a value of $h^2 \leq \frac{1}{2}$ is consistent with $\Theta \leq \frac{1}{2}$, and a value of $h^2 \leq \frac{2}{3}$ with $\Theta \leq 1$.

Initial phase. A comparison of the response of the mean phenotype from standing variation and new mutations during the initial phase of adaptation is delicate. Whereas for the response from standing variation (6.3) can still be used, the response from new mutations increases nonlinearly (see Fig. 4.4). In particular, equation (4.33) as well as its simplification (4.35) for equal effects, are too complicated to derive analytical comparisons between the two scenarios. Although numerical solution is simple and computationally fast, there are many scenarios to be compared with of how standing variation is maintained and how it is structured (e.g., Chevin and Hospital, 2008, De Vladar and Barton, 2014, Stephan and John, 2020, Hayward and Sella, 2022, and references therein). Thus, a detailed comparison has to be postponed to a future treatment.

6.3. Number of segregating sites as an indicator for sweep-like vs. shift-like patterns

Taking up the arguments in Sect. 5, we propose to use the number of segregating sites as a main indicator for the distinction between the scenarios of adaptation being mainly due to selective sweeps or mainly due to small frequency shifts at many loci. This is supported by our theory in Sect. 5.1 showing that once a stationary response to directional selection has been achieved, the expected number of segregating sites, $\bar{S}^*$, is proportional to Θ, depends logarithmically on N, and exceeds the neutral expectation of $2\Theta \ln N$, but never by more than a factor of two if s is varied between 0 and 0.5 (see eq. (5.2) and Fig. 5.1). As a function of s, the maximum value of $\bar{S}^*$ is achieved if $\bar{\alpha}s \approx 1/(2 \ln N)$.

For the initial phase of adaptation, the branching process approximations $\tilde{S}(\tau)$ in (5.5) and (5.6) are much more accurate than the simple approximations $\bar{S}(\tau)$ in (5.3) (see Fig. E.1). However, all these approximations show that the expected number of segregating sites is proportional to Θ, depends only very weakly on the product of s and the average mutation effect $\bar{\alpha}$, and is independent of the population size N.

Further support for our proposition derives from Figure 5.2. Motivated by Höllinger's (2019) work on a different model of selection, in this figure we illustrate the polygenic pattern of adaptation at the loci that contribute to the trait at the time T when the mean phenotype $\bar{G}$ first reaches or exceeds a given value. Although the expected number of segregating sites $\bar{S}_T$ deviates from our approximations in Sect. 5 for $2\Theta = 0.1$, the main conclusion remains the same. Clear sweep pattern can be observed for $2\Theta = 0.1$ (where $\bar{S}_T < 2.5$) and shift-like patterns for $2\Theta = 10$ (where $\bar{S}_T > 35$).

Because we assume exponential directional selection, beneficial mutations, and weak random genetic drift, all mutants that become 'established' will sweep to fixation in the long run. Therefore, we focus on the initial response when characterizing patterns of adaptation as sweep-like vs. shift-like.

In Section 6.4, we argue that our results are of relevance for other modes of directional selection. A popular model is one in which a population is initially in equilibrium under mutation-stabilizing selection-drift balance and then starts to evolve due to a sudden or a continuous change in the environment, modeled by a sudden shift in the optimum phenotype or by a continuously moving optimum (e.g. Bürger and Lynch, 1995, Gomulkiewicz and Holt, 1995, Matuszewski et al., 2015). For models like this, our findings may be of relevance either for the initial response (in case of a sudden shift) or even in the long term (in case of a moving optimum). Below, we briefly review recent treatments focusing on the polygenic dynamics underlying trait evolution.

If a small sudden shift in the optimum phenotype occurs, mutations of relatively small effect may be most conducive to selective sweeps because then large-effect mutations become deleterious when the population closely approaches the optimum. If a sudden shift in the optimum of appreciable magnitude occurs and the initial population has been in mutation-stabilizing selection balance before, then a kind of threshold behavior has been identified for deterministic models (De Vladar and Barton, 2014, Jain and Stephan, 2017a). Under mutation-selection balance the initial variance will be high if allelic effects are small because such alleles are maintained at intermediate frequencies. By contrast, the initial variance will be low if allelic effects are very large because such alleles will be nearly fixed or absent. Alleles of very small effect will respond slowly after the onset of selection because they are nearly neutral. If their number is large, as assumed by the infinite sites model, then the total response of the mean phenotype can still be large (proportional to the variance, as in Lande's equation discussed above). Alleles of very large effect will initially also respond very slowly because they are so rare, and this will lead to a slow response of the mean phenotype because there is little variation on which selection can act. As a result, alleles with effects close to an intermediate threshold value will be the first to respond quickly and start sweeping. On a longer time scale and for a large shift in the optimum, the response from a few loci with large effects is comparable with that from many loci with small effects if the sum of effect sizes is the same. If most loci have large effects, then the response will clearly be much faster than the response from the same number of small-effect loci; the larger the effect the faster the sweeps (De Vladar and Barton, 2014, Jain and Stephan, 2017a,b).

In finite populations the conclusions are even more subtle and may deviate from those in the deterministic case. Mutations of small effect may contribute much less to the response than in large populations because their evolutionary fate is dominated by genetic drift and their initial distribution under mutation-stabilizing selection-drift balance will be U-shaped. Again, mutations of moderate effects may contribute most to the selective response after a sudden shift in the optimum (John and Stephan, 2020, Devi and Jain, 2023, Hayward and Sella, 2022). The latter authors also conclude that mutations of large effect are unlikely to sweep, but this is a consequence

of their assumptions that the trait is highly polygenic (i.e., $\sqrt{\Theta} \gg 1$), the population is initially in mutation-stabilizing selection-drift balance (whence alleles of large effect are extremely rare), and there is only a small shift in the optimum of a few phenotypic standard deviations. The latter assumption implies that the equilibration phase is reached quickly, and then fine-tuning by alleles of small effect is essential for adaptation. If alleles of large effect contribute substantially to the initial variance and the shift is large, then they will sweep (Chevin and Hospital, 2008, Thornton, 2019, John and Stephan, 2020, Stephan and John, 2020).

For a gradually moving optimum, after a relatively short initial phase, a quasi-stationary phase is entered during which, except for stochastic fluctuations, a constant average variance is attained and the mean phenotype lags behind the optimum by an, on average, constant amount. Of course, a prolonged quasi-stationary response will depend crucially on a constant input of new mutations. During this phase, the trait experiences a mixture of directional and stabilizing selection, and the variance settles to a value that is lower than the variance under mutation-drift balance (which is nearly the same as that achieved under exponential selection), but higher than that under mutation-stabilizing selection-drift balance (Lynch and Lande, 1993, Bürger and Lynch, 1995, Jones et al., 2004). Matuszewski et al. (2015) showed that compared to adaptation from *de novo* mutations, adaptation from standing variation proceeds by the fixation of more alleles of small effect. If the optimum is stopped after some time or new mutations are ignored, then the phenotypic dynamics becomes more complex and may depend on details of the underlying genetics (De Vladar and Barton, 2014).

In our finite-population model, in which initial mutant frequencies are $1/N$, all mutants are beneficial, and no epistatic interactions occur, the response of the mean is faster the larger the allelic effects. In addition, mutations that sweep, sweep faster than predicted by deterministic growth. Nevertheless, the response of the mean phenotype depends linearly on N through $\Theta = NU$ (Corollary 4.8 and Proposition 4.11), except in the very early phase when it depends mainly on $U\bar{\alpha}$ but is nearly independent of N (Proposition 4.13).

6.4. Conclusions, limitations, and outlook

We conclude from our analyses that Θ, the expected number of beneficial mutations occurring in the population per generation (and contributing to the trait) is the main determinant of the pattern of adaptation at the genetic level. Selective sweeps at a few loci are the dominant pattern if $\Theta \leq 0.1$, and small allele frequency shifts at many loci are observed if $\Theta \geq 10$. Given the same mean (beneficial) mutation effect, $\bar{\alpha}$, higher second and third moments of the mutation distribution favour sweeps. The expected number of segregating sites is an excellent indicator for the pattern of adaptation. However, it may be challenging to estimate the number of segregating sites contributing to a trait. The selection coefficient s and the average effect $\bar{\alpha}$ primarily determine

the rate of adaptation via their product $s\bar{\alpha}$. Except for Θ, the population size has a weak effect on the rate of adaptation and on the population variance; often it enters expressions approximately logarithmically. We emphasize that our analysis is based on the assumption $Ns\bar{\alpha} \gg 1$, i.e., that selection is considerably stronger than random genetic drift. Most of our approximations, however, perform well when $Ns\bar{\alpha} > 2$.

Our conclusions concerning the influence of the parameters on the pattern of adaptation are similar to those of Höllinger et al. (2019), although their model and approach deviate from ours in important respects. Most significantly, their genetic architecture differs drastically from ours. They study a polygenic trait that assumes only two or very few values, such as resistance and non-resistance. The contributing loci are thus highly redundant, and a mutation at one or at very few loci is sufficient to complete adaptation. This is a very strong form of epistasis. In addition, mutation effects are the same at all loci. Selection on the discrete trait acts similarly as in our model: they assume a linear fitness function in continuous time, which corresponds to an exponential in discrete time. In contrast to our model, their initial populations are polymorphic and (mostly) assumed to be in mutation-stabilizing selection-drift balance. Nevertheless, they find that their population-scaled background mutation rate Θ_{bg}, which corresponds closely to our 2Θ, explains the main differences in the patterns of adaptation, i.e., few sweeps versus many slight shifts. This main conclusion is confirmed by Höllinger et al. (2023) for an additive trait subject to a sudden shift in the optimum phenotype, again for equal mutation effects.

The validity of our conclusions may be limited by a number of important simplifying assumptions, which are unlinked loci and linkage equilibrium, no pleiotropy or epistasis, and exponential directional selection. In addition, we assume that selection is stronger than random drift and we study only the selection response from new mutations, i.e., we ignore the response from standing variation. Our results are derived for haploid populations, but will apply as well to diploids without dominance by substituting $2N$ for N.

Our assumption of unlinked sites will be biologically reasonable unless the number of (simultaneously) segregating sites exceeds the number of chromosomes by far or many sites occur in regions of reduced recombination. Although exponential directional selection by itself does not induce linkage disequilibrium, deviations from this form of selection as well as random genetic drift will do so. Therefore, it would desirable to extend the above results to loosely linked loci and perform a kind of quasi-linkage equilibrium analysis.

Quite extensive simulation results for models of artificial directional selection (such as truncation selection) by Keightley and Hill (1983, 1987) suggest that linkage reduces the asymptotic variance, hence the response of the mean, compared to unlinked loci. The reduction will be small unless linkage is very tight. A similar observation was reported in Bürger (1993) for exponential directional

selection. However, these results included only population sizes of a few hundred individuals. For a moving optimum model and population sizes up to about 10^4, a qualitatively similar finding was obtained in Bürger (2000, pp. 335–338). In all these investigations, mutation distributions with positive and negative effects were used. Then, mutations of both signs appear on the same chromosome, whence low recombination will reduce fixation probabilities, especially of weakly beneficial mutants, and hamper adaptation (Barton, 1995). In concordance with these results, a simulation study by Thornton (2019) found that weak or moderate linkage has almost no effect on the response of the mean and on the variance of the trait, and also little effect on the fixation probabilities of advantageous mutations, but the resulting interference increases fixation times and affects haplotype diversity.

Analytical work on the effects of linkage is scarce, based on specific assumptions, and shows that linkage can increase or decrease the variance of a trait under directional selection, depending on the model assumptions and the magnitude of the parameters (e.g. Robertson, 1977, Hospital and Chevalet, 1996, Zhang and Hill, 2005b). The most comprehensive analysis may be that by Turelli and Barton (1994). They showed that even under strong epistatic selection, such as truncation selection, the trait distribution is significantly affected by linkage disequilibrium only if loci are tightly linked. This provides additional support for our expectation that our results derived under the assumptions of linkage disequilibrium and absence of epistasis will provide decent approximations if recombination among the contributing loci is moderate or strong. However, the explicit determination of the effects of linkage on the likelihood of complete sweeps will pose a formidable challenge (e.g. Kim and Stephan, 2003).

Moderate or strong linkage can lead to selective interference. An extreme scenario is provided by a single locus with recurrent mutation. For this case our theory remains applicable if $2N\mu \leq 1$, where μ is the per-locus mutation rate. Then it is sufficient to consider the first *successful* mutation event. If $2N\mu > 1$, competition of mutants from different mutation events can no longer be ignored and complicates the evolutionary dynamics (unpublished results by HG).

We expect that our assumption of exponential directional selection provides a good approximation for the early phases of adaptation under other forms of directional selection. A property of exponential selection that is crucial for our analysis is that the strength of selection remains constant over time. It shares this property with truncation selection (if the truncation probability remains constant), and with selection imposed by a steadily moving optimum during the phase when the population mean follows the optimum at the same rate and with a constant average lag. If a sudden shift in the optimum phenotype occurs then selection will remain approximately constant initially, but eventually gets increasingly weaker as the mean proceeds towards the optimum. Some of the above discussed results provide support for the expectation that our results may be applicable to

these forms of selection. Exponential directional selection will not be suitable as an approximation if the selection strength is fluctuating in time, if there are pleiotropic fitness effects, or if selection is epistatic, strong, and recombination is weak.

Our results concerning the time dependence of the distribution of sweeping *de novo* mutations will be applicable to very rare allelic variants already present in the population. This will be the case if their effect is large and the population has been under stabilizing selection before directional selection started. If more than one copy is initially present, the quasi-deterministic phase could be started from a gamma-distributed effective initial population size (Martin and Lambert, 2015) instead of an exponential. The extent to which our results on the pattern of adaptation from *de novo* mutations remain applicable in the presence of standing variation would be an interesting and challenging topic for future research.

Table 1: Glossary of symbols. For both the Roman and Greek alphabets, uppercase letters precede lowercase ones. For each uppercase or lowercase letter, listing is in order of appearance of the definition in the text. The second column indicates whether the symbol is used in the one-locus model or in the infinite sites (and/or quantitative-genetic) model. The references are to the equation closest to the definition of each symbol. Thus, (2.1), (2.1)+, (2.1)- refers to (2.1), the text below (2.1), the text above (2.1), respectively. Sometimes the respective subsection is given. Symbols that occur only in the Discussion (mainly Sect. 6.2.3) or only in the Appendix are not listed.

Symbol	1L	ISM	Reference	Definition
a			(3.15b), (3.17)	important parameter for g_a
f		×	Sect. 2.1	density of the distr. of mutation effects α_i
G		×	Sect. 2.1	genotype (or phenotype)
$\bar{G}_i$		×	(4.3)	contributions of locus i to $\bar{G}$
$\bar{G}$		×	(4.12)	expected phenotypic mean
$\bar{G}^*$		×	(4.20)-	expected phenotypic mean in the quasi-stationary phase
$\Delta\bar{G}^*$		×	(4.21)	expected per-generation response in the quasi-stationary phase
$\bar{G}_T$		×	Sect. 5.2	average of mean phenotypic value in gen. T
g_a	×		(3.15)	density of X_n
$h^{(i)}$		×	(3.18)	density of r.v. measuring the freq. of the i-th mut. cond. on non-ext. by gen. τ
$\tilde{h}^{(i)}$		×	(3.20)	describes the positive part of $\tilde{X}_\tau^{(i)}$
k		×	(2.3)-	total number of mutation events until time τ
$M_{i,\Theta}$		×	(3.16)	normalization constant
$m_{i,\Theta}$		×	(2.2)	density of Erlang distr. (approx. waiting time τ_i)
N			Sect. 2	population size
N_e			(4.10)-	effective population size
n	×		Sect. 2.2	generation
$\mathrm{Poi}_{\Theta\tau}$		×	(2.3)	Poisson distr.
$P_{\mathrm{ext}}^{\infty}$	×		(3.1)-, (3.7)	probability of ultimate extinction
P_{sur}	×		(3.8)+	(long-term) survival probability (Galton-Watson process)
P_{fix}	×		(B.1)	probability of ultimate fixation (diffusion theory)
$P_{\mathrm{ext}}(n)$	×		(3.9)	probability of extinction before gen. n
$P_{\mathrm{ext}}^{(i)}(\tau)$		×	(3.19)	probability that the mutation at locus i has been lost by gen. τ
$p(\cdot)$	×		(4.5)	solution of the discrete, deterministic selection equation
S		×	(5.1)-	number of segregating sites
$\bar{S}$		×	(5.1)	expected number of seg. sites
$\bar{S}^*$		×	(5.2)-	expected number of seg. sites in the stat. phase
$\tilde{S}$		×	(E.4), (5.5), (5.6)	expected number of seg. sites under branching process model
$\bar{S}_T$		×	Sect. 5.2	average number of seg. sites at time T
s			Sect. 2.1	selection coefficient
$\tilde{s}$	×		(4.7)	selection coefficient used in relation to diffusion theory
$s_{\max}$		×	(5.4)	sel. coeff. for which $\bar{S}^*$ is maximized
T_ϵ	×		(A.11)	time needed such that $1 - P_{\mathrm{ext}}(T_\epsilon) = (1+\epsilon)(1 - P_{\mathrm{ext}}^{\infty})$ for $\epsilon > 0$
T		×	Sect. 5.2	stopping time
$\bar{T}$		×	Sect. 5.2	average stopping time
t_M	×		(4.1)+	time at which the variance γ is maximized
t_{fix}	×		(B.4)	expected time to fixation
$\bar{t}_{\mathrm{fix}}$	×		(4.8)	expected mean time to fixation
t_{loss}	×		(B.7)	expected time to loss
$\bar{t}_{\mathrm{loss}}$	×		(B.15)	expected mean time to loss
U		×	Sect. 2.1	expected number of beneficial mutations per individual per generation
V_{G_i}		×	(4.4)	contributions of locus i to V_G
V_G		×	(4.13)	expected phenotypic variance
V_G^*		×	(4.20)-	expected phenotypic variance in the quasi-stationary phase
v	×		(2.1)	variance of the offspring distribution ξ_j
W_n	×		(3.1)-	rescaled (discrete) random variable obtained from Z_n
W	×		(3.1)	limiting random variable of W_n
W^+	×		(3.2)	absolutely continuous part of W
w^+	×		(3.2)+	density of W^+

Table 1: (continued)

Symbol	1L/ISM		Reference	Definition
W_n^+	×		(3.11)	positive part of W_n
$\bar{w}(\cdot)$	×		(4.5)+	mean fitness needed for the discrete, det. sel. equation
X_n	×		(3.13)	random variable measuring the (positive) mut.-frequency in gen. n
$\tilde{X}_\tau^{(i)}$		×	(3.21)	(unconditioned) r.v. for the frequency of the i-th mutant
$X_\tau^{(i)}$		×	(4.3)-	X_n for locus i (with $\tau = n$, $\sigma_i = \sigma$)
$\hat{x}$	×		(3.15)+	mut. freq. for which g_a has unique maximum for $a < 2$
Y_n	×		(3.12)+	random variable with distr. Ψ_n (approx. for positive part of Z_n)
Z_n	×		(2.1)-	Galton-Watson process
α_i		×	Sect. 2.1	mutation effect on phenotype of locus i
$\bar{\alpha}$		×	Sect. 2.1	mean of the mutation effects
$\alpha_n(f)$		×	(4.25)-	nth moment about zero of the mutation distribution f
γ_1	×		(4.1a)	first moment of g_a
γ_2	×		(4.1b)	second moment of g_a
γ	×		(4.1c)	within-population variance of the mutant's allele frequency
Θ		×	Sect. 2.1	expected total number of mutations occurring in the population per generation
λ_n	×		(3.12)	parameter of exponential distribution Ψ_n
ν		×	(4.20)	notational simplification
ξ_j	×		(2.1)-	(random) number of offspring of individual j
ρ	×		Rem. 3.3	constant, such that $P_{\mathrm{sur}} \approx \rho s \alpha$
σ_i		×	Sect. 2.1	fitness effect of a mutation at locus i
σ	×		Sect. 2.2	fitness of the derived state (mean of ξ_j, mean offspring number)
τ		×	Sect. 2.3	(discrete) generation since the initial population started to evolve
τ_i		×	(2.2)-	random time at which ith mutation event occurred
τ_c		×	(4.18)	expected time by which the first mutation becomes fixed
τ_1		×	(4.33)-	bound for time
$\tilde{\tau}$		×	(5.5)+	bound for time
Φ	×		(3.3)-	Laplace transform of W
φ	×		(2.1)+	offspring generating function of ξ_j (of Z_1)
Ψ_n	×		(3.12)	exponential distribution
ψ_n	×		(3.12)+	density of distr. Ψ_n
E				expectation
Var				variance
Prob				probability
$\mathrm{ProductLog}$			(3.7)	product logarithm (Lambert function)
Γ			(3.16)+	incomplete Gamma function
$[\cdot]$			(3.17)-	nearest integer to $\cdot$
E_1			(4.1)+	exponential integral
Exp				exponential distribution
γ				Euler–Mascheroni constant

Acknowledgments

We thank Joachim Hermisson, Emmanuel Schertzer, Ben Wölfl, and Ilse Höllinger for helpful discussions. We gratefully acknowledge extraordinarily detailed and helpful comments by three reviewers and the editor, John Wakeley. Financial support by the Austrian Science Fund (FWF) through the Vienna Graduate School of Population Genetics (GrantDK W1225-B20) to HG is gratefully acknowledged.

Additional materials

We provide a comprehensive Wolfram *Mathematica* notebook, containing additional visualizations of the analytical predictions and efficient numerical procedures. Upon publication, the notebook and the simulation code (C++) will be made available via github: `https://github.com/Ha-nn-ah/EvolutionOfQuantitativeTraits`.

Appendices

A. Fractional linear offspring distributions and associated Galton-Watson processes

A.1. Basics

The fractional linear, or modified geometric, distribution is given by

$$p_0^{(\mathrm{fl})} = r \;\text{ and }\; p_k^{(\mathrm{fl})} = (1-r)(1-p)p^{k-1} \;\text{ if }\; k \geq 1\,, \tag{A.1}$$

where $0 < r < 1$ and $0 < p < 1$; e.g. Athreya and Ney (1972, pp. 6-7) or Haccou et al. (2005, pp. 16). The name fractional linear derives from the fact that its generating function is

$$\varphi(x) = \frac{r + x(1-p-r)}{1-xp}\,, \tag{A.2}$$

hence fractional linear. With $r = 1-p$, the geometric distribution is recovered. It is straightforward to show that every fractional linear generating function generates a modified geometric distribution.

Mean and variance of $\{p_k^{(\mathrm{fl})}\}$ are

$$m = \frac{1-r}{1-p} \text{ and } v = \frac{(1-r)(p+r)}{(1-p)^2}\,. \tag{A.3}$$

In terms of m and v, we obtain

$$p = 1 - \frac{2m}{m+m^2+v} \text{ and } r = 1 - \frac{2m^2}{m+m^2+v}\,. \tag{A.4}$$

Therefore $m > 1$ if and only if $0 < r < p < 1$. In this case the associated Galton-Watson process $\{Z_n\}$ is supercritical. We note that $m > v$ if and only if $2p < 1-r$, which implies $p < \frac{1}{2}$.

If $m \neq 1$, the n-times iterated generating function $\varphi_{(n)} = \varphi_n$ is (rearranged from the parameterization in Athreya and Ney (1972, p. 7))

$$\varphi_n(x) = \frac{r(1-x) - m^{-n}(r-px)}{p(1-x) - m^{-n}(r-px)}\,. \tag{A.5}$$

It is fractional linear with parameters

$$p_n = \frac{p(1-m^{-n})}{p-rm^{-n}} \text{ and } r_n = \frac{r(1-m^{-n})}{p-rm^{-n}}\,. \tag{A.6}$$

From now on we assume $m > 1$, i.e., $r < p$. The probability of extinction by generation n, $P_{\text{ext}}^{(n)}$, is given by

$$P_{\text{ext}}^{(n)} = \varphi_n(0) = \frac{r(1-m^{-n})}{p-rm^{-n}} = r_n\,, \tag{A.7}$$

and the (ultimate) extinction probability is

$$P_{\text{ext}}^{\infty} = \frac{r}{p}\,. \tag{A.8}$$

From (A.7) and (A.8) we obtain the relation

$$\frac{1-P_{\text{ext}}^{\infty}}{1-P_{\text{ext}}^{(n)}} = 1 - m^{-n}P_{\text{ext}}^{\infty} \tag{A.9a}$$

and its equivalent

$$\frac{1-P_{\text{ext}}^{(n)}}{1-P_{\text{ext}}^{\infty}} = 1 + \frac{P_{\text{ext}}^{\infty}}{m^n - P_{\text{ext}}^{\infty}}\,. \tag{A.9b}$$

This allows to compute the time needed for the probability of non-extinction by generation n, $1-P_{\text{ext}}^{(n)}$, to decline to $(1+\epsilon)(1-P_{\text{ext}}^{\infty})$. For $\epsilon > 0$ (not necessarily small) we define T_ϵ as the (real) solution of

$$1 - P_{\text{ext}}(T_\epsilon) = (1+\epsilon)(1-P_{\text{ext}}^{\infty})\,. \tag{A.10}$$

By (A.9b), this time can be calculated explicitly:

$$T_\epsilon = \frac{\ln\Bigl(P_{\text{ext}}^{\infty}\Bigl(1+\frac{1}{\epsilon}\Bigr)\Bigr)}{\ln m}\,. \tag{A.11}$$

Therefore, $\dfrac{1}{\ln m}\ln\Bigl(1+\frac{1}{\epsilon}\Bigr)$ is always an upper bound to T_ϵ.

A.2. Distributions of W_n and W

The cumulative distribution function of $W_n = Z_n/m^n$ can be calculated explicitly (e.g., from φ_n):

$$P(W_n \le x) = 1 - \frac{(p-r)p_n^{\lfloor m^n x\rfloor}}{p-rm^{-n}}\,, \tag{A.12}$$

where p_n is given by (A.6) and $\lfloor z\rfloor$ denotes the largest integer smaller than z. From the equation

$$\text{Prob}[W_n \le x] = P_{\text{ext}}^{(n)} + (1-P_{\text{ext}}^{(n)})\,\text{Prob}[W_n^+ \le x]\,, \tag{A.13}$$

we infer

$$\text{Prob}[W_n^+ \le x] = 1 - p_n^{\lfloor m^n x\rfloor}\,. \tag{A.14}$$

This shows that the cumulative distribution of W_n^+ is imbedded into (hence approximated by) the exponential distribution with parameter

$$\lambda_n = -m^n \ln p_n = -m^n \ln \frac{p(1-m^{-n})}{p - rm^{-n}} \,. \tag{A.15}$$

We note that $\lim_{n\to\infty} p_n^{m^n x} = e^{-(1-r/p)x}$. As a consistency check, it can be confirmed directly from the CDF that $\mathrm{E}[W_n^+] = (1 - P_{\mathrm{ext}}^{(n)})^{-1}$.

If we take the exponential distribution with parameter $\lambda_n^* = 1 - P_{\mathrm{ext}}^{(n)}$, this provides an approximation to the (discrete) distribution of W_n^+ which has the property that the integral of the difference of the two distribution functions vanishes, i.e., such that

$$\int_0^\infty (1 - p_n^{\lfloor m^n x \rfloor}) - (1 - e^{\lambda_n^* x})\, dx = 0 \,. \tag{A.16}$$

Conditional on this requirement, the exponential with parameter $\lambda_n^* = 1 - P_{\mathrm{ext}}^{(n)}$ provides the best possible exponential approximation to W_n^+.

We obtain that $W^+ = \lim_{n\to\infty} W_n^+$, which exists quite generally (Athreya and Ney, 1972, e.g.,), is exponentially distributed with parameter

$$\lambda = \lim_{n\to\infty} \lambda_n^* = 1 - \frac{r}{p} = 1 - P_{\mathrm{ext}}^{\infty} \,. \tag{A.17}$$

This follows as well from Poincaré's functional equation (3.3), i.e., $\Phi(u) = \varphi(\Phi(u/m))$, which implies that the Laplace transforms of W is

$$\Phi(u) = \mathrm{E}[e^{-uW}] = \frac{p - r + ru}{p - r + pu} = 1 - \lambda + \lambda \frac{1}{1 + u/\lambda} \,, \tag{A.18}$$

with λ given by (A.17).

It is easy to show that if W^+ is exponential, then the generating function of the offspring distribution is fractional linear. This follows because the Laplace transform Φ is strictly monotone decreasing so that φ is uniquely determined by eq. (3.3). Therefore, as pointed out by a reviewer, W^+ is exponential if and only if the offspring distribution is fractional linear.

A.3. Comparison with the Galton-Watson process originating from a Poisson offspring distribution

We assume a fractional linear distribution with $v = m$, as for a Poisson distribution with mean m. A simple calculation shows that $v = m$ if and only if $p = \frac{m}{2+m}$ and $r = \frac{2-m}{2+m}$, where we assume $m < 2$. Then we obtain

$$P_{\mathrm{ext}}^{\infty} = \frac{2}{m} - 1 \text{ and } P_{\mathrm{ext}}^{(n)} = \frac{(2-m)(1-m^{-n})}{m - (2-m)m^{-n}} \,. \tag{A.19}$$

If $m = e^{s\alpha}$, as we assume in the main text, then $P_{\mathrm{ext}}^{\infty} = 2e^{-s\alpha} - 1$ and

$$P_{\mathrm{sur}} = 1 - P_{\mathrm{ext}}^{\infty} = 2(1 - e^{-s\alpha}) = 2s\alpha - (s\alpha)^2 + O((s\alpha)^3) \,. \tag{A.20}$$

We note that $P_{\mathrm{sur}} < 1$ if and only if $s\alpha < \ln 2 \approx 0.69$ (otherwise, $P_{\mathrm{sur}} = 1$). For comparison, the series expansion of P_{sur} for the corresponding Poisson offspring distribution is $2s\alpha - \frac{5}{3}(s\alpha)^2 + O((s\alpha)^3)$, and the survival probability is always < 1. In particular, the survival probability for the

Poisson distribution is always smaller than that for the corresponding linear fractional. In the limit of $s\alpha \to 0$, they are asymptotically equivalent.

If we assume in the fractional linear case that v and m are proportional as $m = e^{s\alpha}$ is varied, i.e., $v = 2m/\rho$ (cf. Remark 3.3), equation (A.11) simplifies to

$$T_\epsilon(s\alpha) = \frac{1}{s\alpha}\,\ln\Big(P_{\text{ext}}^{\infty}\Big(1+\frac{1}{\epsilon}\Big)\Big) \tag{A.21a}$$

$$= \frac{1}{s\alpha}\ln\Big(1+\frac{1}{\epsilon}\Big) - \rho + O(s\alpha)\,, \tag{A.21b}$$

where the asymptotic estimate in the second lines holds if $s\alpha \to 0$ because $P_{\text{ext}}^{\infty} \approx 1 - \rho s\alpha$. Comprehensive numerics show that (A.21a) apparently always (i.e., if $P_{\text{ext}}^{\infty}\Big(1+\frac{1}{\epsilon}\Big) \geq 1$) provides an upper bound for any Poisson offspring distribution with mean $e^{s\alpha}$ if the corresponding P_{ext}^{∞} is used. This bound is a highly accurate approximation with relative errors of order s or smaller. Finally, we note that $P_{\text{ext}}^{\infty} = 0$ if $m \geq 1 + 2/\rho$.

B. Mean times to fixation or loss of a mutant

Approximations for the expected times to absorption conditioned on either fixation or loss of a new mutation, for simplicity called mean fixation time or mean loss time, have been an important tool in theoretical population genetics ever since its inception in the 1920s. Even in large populations, the initial fate of a mutant, as well as its fate in the final stages before fixation, is subject to strong stochastic influences, and the derivation of simple and accurate approximations is still a topic of research (e.g., Charlesworth, 2020, and references therein). In this appendix, we add simple estimates to this literature that are particularly useful to compute the mean fixation or loss times of beneficial mutants drawn from a distribution of effects.

For the computation of mean times to fixation or loss in a finite population, we use diffusion approximations because the branching process model is not directly applicable. Although the approximation $g_a(p)$ in (3.15) for the distribution of the number of mutations is quite accurate for late stopping times, the derivation of useful approximations for the mean fixation and loss times requires more delicate transformations than that used for Result 3.4 and is work in progress.

B.1. Diffusion approximations

Here, we collect results on diffusion approximations of the fixation probability and of the mean times to fixation or loss of favorable mutants with a fixed selection coefficient $s > 0$. In a haploid population, the well-known diffusion approximation for the fixation probability of a single beneficial mutant is

$$P_{\text{fix}}(s,N) = \frac{1-e^{-2s}}{1-e^{-2Ns}}\,. \tag{B.1}$$

In the applications to our model, we replace s by $\tilde{s}(\alpha)$ defined in (4.7). The diffusion approximation $P_{\text{fix}}(s,N)$ is always an upper bound to the true fixation probability in the Wright-Fisher model, and is highly accurate for every $N \geq 8$ if $0 < s \leq 0.1$ (Bürger and Ewens, 1995).

The mean times to fixation or loss can be expressed only in terms of integrals of the corresponding sojourn time densities (Ewens, 2004, Sects. 4.6 and 5.3). For the mean fixation time, the simple approximation

$$t_{\text{fix}}^{(\text{HP})}(s,N) = \frac{2}{s}\left(\ln(2Ns) + \gamma - \frac{1}{2Ns}\right) \tag{B.2}$$

was provided by Hermisson and Pennings (2005), where γ is the Euler gamma. It is very accurate if $2Ns \geq 3$.

If $2Ns < 3$, we use

$$t_{\text{fix}}^{(\text{small})}(s,N) = 2N\left(1 - 2Ns\,\frac{11 - 6\gamma - 6\ln 3}{27}\right)\,. \tag{B.3}$$

This is the linear function in s which connects the mean fixation time $2N$ at $s = 0$ with the value of $t_{\text{fix}}^{(\text{HP})}(s,N)$ at $s = 3/(2N)$. As a function of s, $t_{\text{fix}}^{(\text{HP})}$ assumes its maximum very close to this value and becomes negative for very small s. As illustrated by Fig. B.1, the function

$$t_{\text{fix}}(s,N) \approx \begin{cases} t_{\text{fix}}^{(\text{HP})}(s,N) & \text{if } s \geq 3/(2N)\,, \\ t_{\text{fix}}^{(\text{small})}(s,N) & \text{if } s < 3/(2N)\,, \end{cases} \tag{B.4}$$

obtained by concatenation of $t_{\text{fix}}^{(\text{HP})}$ and $t_{\text{fix}}^{(\text{small})}$ yields a highly accurate approximation of the true diffusion approximation.

For the mean loss time of a single advantageous mutant, we obtain (see below)

$$t_{\text{loss}}^{(\text{app})}(s,N) \approx -2(1+s)\ln(2s) + 2(1-\gamma) + s(3-2\gamma) + \frac{1}{Ns} \tag{B.5a}$$

$$\approx -2\ln(2s) + 2(1-\gamma) + \frac{1}{Ns}\,, \tag{B.5b}$$

which is very accurate if $Ns \geq 2$ but breaks down if $Ns \leq 1$.

If $Ns < 2$, we use

$$t_{\text{loss}}^{(\text{small})}(s,N) = 2\ln(N) - sN\left(\ln 4 + \gamma - \tfrac{5}{4}\right)\,. \tag{B.6}$$

This is the linear function in s which connects the (approximate) mean loss time $2\ln(N)$ at $s = 0$ (Kimura and Ohta, 1969) with the value of $t_{\text{loss}}^{(\text{app})}(s,N)$ at $s = 2/N$. As illustrated by Fig. B.1, the function

$$t_{\text{loss}}(s,N) \approx \begin{cases} t_{\text{loss}}^{(\text{app})}(s,N) & \text{if } s \geq 2/N\,, \\ t_{\text{loss}}^{(\text{small})}(s,N) & \text{if } s < 2/N \end{cases} \tag{B.7}$$

yields a highly accurate approximation of the true diffusion approximation.

Proof of (B.5a). To derive this approximation, we start with the well-known diffusion approximations for the conditional sojourn time densities (eqs. (5.54) and (5.55) in Sect. 5.4 of Ewens, 2004). To account for haploidy, we set $\alpha = 2Ns$ and $p = 1/N$ in these expressions (p the initial mutant frequency). By integration with respect to x, we obtain with the help of *Mathematica* and after simple algebraic rearrangement

$$\begin{aligned} t_{\text{loss}}^{(1)}(s,N) = \frac{1}{(e^{2Ns}-1)s}\Big[&e^{2Ns}\Big(-\operatorname{Ei}(-2s) + \gamma + \ln(2Ns) - \ln(N-1)\Big) - \operatorname{Ei}(2Ns) + \operatorname{Ei}(2(N-1)s) \\ &- \operatorname{Ei}(2s) + \gamma + \ln(2Ns) - \ln(N-1) + e^{2Ns}\big(\operatorname{Ei}(-2(N-1)s) - \operatorname{Ei}(-2Ns)\big)\Big] \end{aligned} \tag{B.8}$$

for the expected time the mutant frequency spends in $(0, 1/N]$, and

$$t^{(2)}_{\text{loss}}(s,N) = \frac{e^{2s}-1}{(e^{2Ns}-1)(e^{2Ns}-e^{2s})s}\Big[-e^{4Ns}\,\text{Ei}(-2s) + e^{2Ns}\,\text{Ei}(2(N-1)s) \\ + e^{4Ns}\,\text{Ei}(-2Ns) + \text{Ei}(2Ns) + 2e^{2Ns}(\gamma + \ln(2Ns) - \ln(N-1)) \\ - \text{Ei}(2s) + e^{2Ns}\,\text{Ei}(-2(N-1)s)\Big] \tag{B.9}$$

for the time spent in $(1/N, 1)$. We note that the times $t^{(1)}_{\text{loss}}(s,N)$ and $t^{(2)}_{\text{loss}}(s,N)$ have been rescaled to generations, i.e., the diffusion time has been multiplied by N. Omitting terms of order e^{-2Ns} or smaller in (B.8) and using the approximation $\text{Ei}(x) \approx e^x/(x-1)$ if $|x| > 4$, we obtain

$$\begin{aligned} t^{(1)}_{\text{loss}}(s,N) &= \frac{1}{(e^{2Ns}-1)s}\Big[e^{2Ns}\Big(-\text{Ei}(-2s) + \gamma + \ln(2Ns) - \ln(N-1)\Big) - \text{Ei}(2Ns) + \text{Ei}(2(N-1)s)\Big] \\ &\approx \frac{1}{s}\Big[-\text{Ei}(-2s) + \gamma + \ln(2s) + \ln(N) - \ln(N-1) - \frac{1}{e^{2Ns}}\Big(\text{Ei}(2Ns) - \text{Ei}(2(N-1)s)\Big)\Big] \\ &\approx 2 - s + 1/N\,. \end{aligned}$$

The latter approximation is obtained by performing a series expansion up to order s^2 and by keeping Ns constant (as in the diffusion approximation). For $t^{(2)}_{\text{loss}}$ we obtain in an analogous way, by omitting terms of order e^{-2Ns} or smaller,

$$\begin{aligned} t^{(2)}_{\text{loss}}(s,N) &\approx \frac{e^{2s}-1}{(e^{2Ns}-1)(e^{2Ns}-e^{2s})s}\Big[-e^{4Ns}\,\text{Ei}(-2s) + e^{2Ns}\,\text{Ei}(2(N-1)s)\Big] \\ &\approx \frac{e^{2s}-1}{s}\Big[-\text{Ei}(-2s) + e^{-2Ns}\,\text{Ei}(2(N-1)s)\Big] \\ &\approx -2\gamma - 2(1+s)\ln(2s) + 2(2-\gamma)s + \frac{1-s}{Ns}\,. \end{aligned}$$

By summing up, i.e., $t_{\text{loss}} = t^{(1)}_{\text{loss}} + t^{(2)}_{\text{loss}}$, and omitting terms of order $1/N$, we arrive at (B.5a). □

B.2. Mean fixation time averaged over the mutation distribution f

Now we derive an approximation for the mean fixation time of a single mutant if its effect α on the selected trait is drawn from an exponential distribution f with mean $\bar{\alpha}$. For given N and s, we defined in (4.8) the corresponding expected mean fixation time by

$$\bar{t}_{\text{fix}}(s,f,N) = \frac{\int_0^\infty t_{\text{fix}}(\tilde{s}(\alpha),N)P_{\text{fix}}(\tilde{s}(\alpha),N)f(\alpha)\,d\alpha}{\int_0^\infty P_{\text{fix}}(\tilde{s}(\alpha),N)f(\alpha)\,d\alpha}\,. \tag{B.10}$$

We recall from (4.7) that $\tilde{s}(\alpha) = e^{s\alpha} - 1$ is the selection coefficient of a mutant of effect α. For $P_{\text{fix}}(\tilde{s}(\alpha),N)$ we use the approximation (B.1), and for $t_{\text{fix}}(\tilde{s}(\alpha),N)$ we use (B.4). With this simplified diffusion approximation for $t_{\text{fix}}(\tilde{s}(\alpha),N)$, the integrals in (B.10) are readily computed numerically. Some values are given in the legend of Fig. 4.2.

An analytically explicit, still quite accurate, approximation of $\bar{t}_{\text{fix}}$ can be derived as follows. We use the approximations $P_{\text{fix}}(s,N) \approx 1 - e^{-2s}$ and $t^{(\text{HP})}_{\text{fix}}(s,N) \approx \frac{2}{s}\left(\ln(2Ns) + \gamma\right)$, which are valid for sufficiently large Ns. In addition, we assume that the selective advantage of the mutant is $s\alpha$

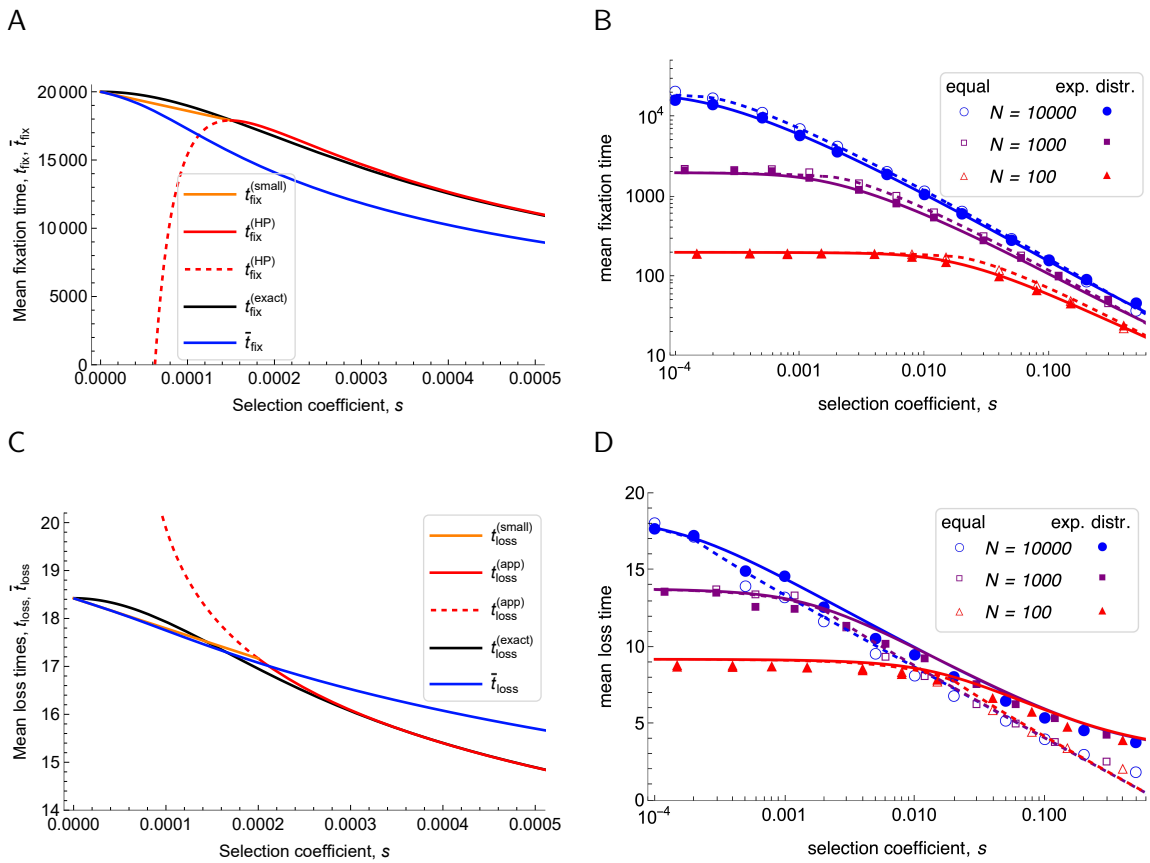

Figure B.1: Panels A and C demonstrate the accuracy of our method of approximating the true diffusion approximation for the mean time to fixation ($t_{\rm fix}$) or loss ($t_{\rm loss}$) by concatenation of a linear approximation for small s and a simplified approximation, i.e., of eqs. (B.4) and (B.7). In panel A, the mean fixation times $t_{\rm fix}^{\rm (small)}(s,N)$ for $s < 3/(2N)$ (orange, (B.3)) and $t_{\rm fix}^{\rm (HP)}(s,N)$ for $s \geq 3/(2N)$ (red, (B.2)) are compared with the exact diffusion approximation (black) for $N = 10^4$. Panel C is analogous to A and displays the mean loss times $t_{\rm loss}^{\rm (small)}(s,N)$ for $s < 2/N$ (orange, (B.6)) and $t_{\rm loss}^{\rm (app)}(s,N)$ for $s \geq 2/N$ (red, (B.5a)). In both panels, the exact diffusion approximations are computed by numerical integration of the corresponding sojourn time densities (Ewens, 2004, Sects. 4.6 and 5.3), and the dashed red curves show $t_{\rm fix}^{\rm (HP)}$ or $t_{\rm loss}^{\rm (app)}$, where they are inaccurate and substituted by $t_{\rm fix}^{\rm (small)}$ or $t_{\rm loss}^{\rm (small)}$, respectively. In addition, for mutation effects drawn from an exponential distribution with mean 1, the expected mean times to fixation ($\bar{t}_{\rm fix}$ in (B.10)) or loss ($\bar{t}_{\rm loss}$ in (B.15)) are displayed in blue. Panel B shows $t_{\rm fix}(\tilde{s}(\alpha),N)$ from (B.4) (dashed curves) and $\bar{t}_{\rm fix}(s,1,N)$ from (B.10) (solid curves) as functions of $s \geq 10^{-4}$ for different population sizes (blue for $N = 10^4$, purple for $N = 10^3$, red for $N = 10^2$). Simulation results from the Wright-Fisher model for equal mutation effects $\alpha = 1$ and exponentially distributed effects with mean $\bar{\alpha} = 1$ are shown as open and filled symbols, respectively. Panel D is analogous and depicts $t_{\rm loss}(\tilde{s}(\alpha),N)$ from (B.7) (dashed curves) and $\bar{t}_{\rm loss}(s,1,N)$ from (B.15) (solid curves). In accordance with (B.5a), panel D illustrates that $t_{\rm loss}$ and $\bar{t}_{\rm loss}$ depend only very weakly on N if $s > 2/N$.

instead of $\tilde{s}(\alpha)$. If f is exponentially distributed with mean $\bar{\alpha}$, we obtain

$$\int_0^\infty P_{\rm fix}(s\alpha,N) f(\alpha)\, d\alpha \approx \frac{2s\bar{\alpha}}{1+2s\bar{\alpha}} \tag{B.11}$$

and

$$\int_0^\infty t_{\rm fix}^{\rm (HP)}(s\alpha,N) P_{\rm fix}(s\alpha,N) f(\alpha)\, d\alpha \approx \frac{\ln(1+2s\bar{\alpha})}{s\bar{\alpha}} \left[2\ln(2Ns\bar{\alpha}) - \ln(1+2s\bar{\alpha})\right]. \tag{B.12}$$

Therefore,

$$\bar{t}_{\text{fix}}(s,\bar{\alpha},N) \approx \frac{(1+2s\bar{\alpha})\ln(1+2s\bar{\alpha})}{2(s\bar{\alpha})^2}\left[2\ln(2Ns\bar{\alpha}) - \ln(1+2s\bar{\alpha})\right]. \tag{B.13}$$

The high accuracy of this approximation is demonstrated in the supplementary *Mathematica* notebook. Assuming $s\bar{\alpha} \ll 1$ and performing a series expansion in $s\bar{\alpha}$, we arrive at

$$\bar{t}_{\text{fix}}(s,\bar{\alpha},N) \approx 2\ln(2Ns\bar{\alpha})\left(\frac{1}{s\bar{\alpha}} + 1 - \frac{2}{3}s\bar{\alpha}\right) - 2. \tag{B.14}$$

If the selective coefficient of the mutants is $\tilde{s}(\alpha)$, instead of $s\alpha$, then some of the above integrations cannot be performed analytically. However, numerics shows that the simple approximation (4.9) still works very well, especially if $2/N \leq s\bar{\alpha} \leq 0.2$ (see *Mathematica* notebook).

B.3. Mean loss time averaged over the mutation distribution f

In analogy to (4.8) and (B.10), we define the expected mean loss time of a single mutant, with effect α on the selected trait drawn from an exponential distribution f with mean $\bar{\alpha}$, by

$$\bar{t}_{\text{loss}}(s,f,N) = \frac{\int_0^\infty t_{\text{loss}}(\tilde{s}(\alpha),N)(1-P_{\text{fix}}(\tilde{s}(\alpha),N))f(\alpha)\,d\alpha}{\int_0^\infty (1-P_{\text{fix}}(\tilde{s}(\alpha),N))f(\alpha)\,d\alpha}. \tag{B.15}$$

Here, $t_{\text{loss}}(\tilde{s}(\alpha),N)$ is taken from (B.7). With this simplified diffusion approximation for $t_{\text{loss}}(\tilde{s}(\alpha),N)$, the integrals in (B.15) are readily computed numerically (see *Mathematica* notebook). For graphs of t_{loss} and $\bar{t}_{\text{loss}}$, see Fig. B.1. It shows that, unless s is very small, $\bar{t}_{\text{loss}} > t_{\text{loss}}$. The reasons are that mutations of larger effect are lost with lower probability than those of smaller effect, and small effect mutations stay longer in the population. The following is a simple and accurate approximation if (approximately) $Ns\bar{\alpha} > 1$ and $s\bar{\alpha} < 0.2$ (see *Mathematica* notebook):

$$\bar{t}_{\text{loss}}(s,\bar{\alpha},N) \approx 2(-\ln(2s\bar{\alpha}) + 1 + 2s\bar{\alpha}) - \frac{1.42 - \ln(Ns\bar{\alpha})}{Ns\bar{\alpha}}. \tag{B.16}$$

Usually, these times to loss are very small. For instance, if the mean of the exponential distribution is $\bar{\alpha} = 1$, $N = 10^4$, we obtain $\bar{t}_{\text{loss}} \approx 5.5$, 8.8, and 14.5 for $s = 0.1$, 0.01, and 0.001, respectively. For larger values of N almost the same numbers are obtained.

C. Mutations occur at their expected time - approximations for the phenotypic mean and variance

We outline a simpler approach than using the Erlang distribution for the waiting time to new mutations. We assume that the ith mutation occurs at its mean waiting time i/Θ and the number of mutation events until time τ is (approximately) $[\Theta\tau]$. This is most efficient if the mutation effects α_i are given, in particular if they are equal across sites. Then we can approximate (4.11) by

$$\mathrm{E}\left[(\tilde{X}_\tau^{(i)})^k\right] \approx (1 - P_{\text{ext}}([\tau - i/\Theta])) \int_0^1 x^k g_{a(\tau - i/\Theta)}(x)\,dx. \tag{C.1}$$

Then, instead of (4.14), we obtain the approximation

$$\begin{aligned}\bar{G}(\tau) &\approx \sum_{j=1}^{[\Theta\tau]} (1 - P_{\text{ext}}([\tau - i/\Theta], e^{s\alpha_i}))\, \alpha_i \gamma_1(a(\tau - i/\Theta, e^{s\alpha_i})) \\ &= \sum_{j=0}^{[\Theta\tau]-1} (1 - P_{\text{ext}}([j/\Theta], e^{s\alpha_j}))\, \alpha_j \gamma_1(a(j/\Theta, e^{s\alpha_j}))\,, \end{aligned} \tag{C.2}$$

where we have reversed the summation order (i.e., $j = [\Theta\tau] - i$) to obtain the second equation. This yields (4.16) by using (4.3). Analogously, we obtain instead of (4.15) the approximation (4.17).

For equal mutation effects, (4.16) and (4.17) provide highly accurate approximations of (4.14) and (4.15), respectively (results not shown, but expected because the right-hand sides of (4.16) and (4.17) can be interpreted as approximating Riemann sums).

For values of Θ much smaller than 1, few new mutations are expected until generation τ (unless τ is sufficiently large) and the rounded, expected number of mutation events, $[\Theta\tau]$, can be zero for many generations. Moreover, the Poisson distribution is highly asymmetric for small $\Theta\tau$ and gets more symmetric for larger $\Theta\tau$. This also implies that the phenotypic mean and variance may be only poorly described by their first moments when one sweep after another occurs (as with very small Θ). By Proposition 4.3, the mutation rate Θ enters the expected phenotypic mean and variance only as a multiplicative factor (see also Fig. S.2). Hence, we use $\bar{G}(\tau, \Theta) \approx \Theta \cdot \bar{G}(\tau, 1)$ and $V_G(\tau, \Theta) \approx \Theta \cdot V_G(\tau, 1)$ to obtain rough approximations for small values of Θ from equations (4.16) and (4.17).

D. Proofs of the results in Section 4

Here, we give the proofs of the main results of Section 4. We start by collecting some formulas needed in the proofs. Integrals and series expansions were mostly calculated with the help of *Mathematica*. However, integrals involving E_1 can be obtained as well by partial integration using $\int E_1(x)\, dx = -e^{-x} + xE_1(x)$.

D.1. Some integral formulas

We recall the abbreviation $\nu = NP_{\text{sur}}(e^{s\alpha})$ from (4.20). In analogy to (3.17) we define

$$a^{\infty}(t, e^{s\alpha}) = \nu e^{-s\alpha t}, \tag{D.1}$$

where we omit the dependence on N. Furthermore, we introduce the (simplified) notation

$$\gamma_1(t, e^{s\alpha}) = \gamma_1(a(t, e^{s\alpha})) \text{ and } \gamma_1^{\infty}(t, e^{s\alpha}) = \gamma_1(a^{\infty}(t, e^{s\alpha})) \tag{D.2a}$$

as well as

$$\gamma(t, e^{s\alpha}) = \gamma(a(t, e^{s\alpha})) \text{ and } \gamma^{\infty}(t, e^{s\alpha}) = \gamma(a^{\infty}(t, e^{s\alpha}))\,. \tag{D.2b}$$

We note that $a(t, e^{s\alpha}) > a^{\infty}(t, e^{s\alpha})$ and, because γ_1 is monotone decreasing in a, $\gamma_1^{\infty}(t, e^{s\alpha}) > \gamma_1(t, e^{s\alpha})$.

We will need the following integrals:

$$\int_0^{\tau} \gamma^{\infty}(t, e^{s\alpha})\, dt = \frac{1}{s\alpha} \left[e^{\nu} \nu E_1(\nu) - \exp(\nu e^{-s\alpha\tau}) \nu e^{-s\alpha\tau} E_1(\nu e^{-s\alpha\tau}) \right]\,, \tag{D.3a}$$

$$\lim_{\tau\to\infty}\int_0^\tau \gamma^\infty(t,e^{s\alpha})\,dt = \frac{\nu}{s\alpha}\,e^\nu E_1(\nu)\,. \tag{D.3b}$$

In addition, we obtain for $T_1 > 0$:

$$\int_{T_1}^\tau \gamma_1^\infty(t,e^{s\alpha})\,dt = \frac{1}{s\alpha}\left[\exp(\nu e^{-s\alpha\tau})E_1(\nu e^{-s\alpha\tau}) - \exp(\nu e^{-s\alpha T_1})E_1(\nu e^{-s\alpha T_1})\right] \tag{D.4}$$

and

$$\int_0^{T_1} \gamma_1^\infty(t,e^{s\alpha})\,dt = \frac{1}{s\alpha}\left[\exp(\nu e^{-s\alpha T_1})E_1(\nu e^{-s\alpha T_1}) - e^\nu E_1(\nu)\right]\,. \tag{D.5}$$

By setting $T_1 = \tau - 1$ in (D.4) and using $\lim_{t\to\infty}\frac{1}{s\alpha t}\exp(\nu e^{-s\alpha t})E_1(\nu e^{-s\alpha t}) = 1$, we arrive at

$$\lim_{\tau\to\infty}\int_{\tau-1}^\tau \gamma_1^\infty(t,e^{s\alpha})\,dt = 1\,. \tag{D.6}$$

From the asymptotic properties of the exponential integral (Abramowitz and Stegun, 1964, Chap. 5.1), in particular

$$e^x E_1(x) = \sum_{k=0}^\infty (-1)^k \frac{k!}{x^{k+1}} = \frac{1}{x} - \frac{1}{x^2} + \frac{2}{x^3} - \frac{6}{x^4} + \dots\ (x\to\infty)\,, \tag{D.7}$$

we derive for arbitrary $T_1 > 0$

$$\lim_{T\to\infty}\frac{1}{s\alpha T}\left[\exp(\nu e^{-s\alpha(T_1+T)})E_1(\nu e^{-s\alpha(T_1+T)}) - \exp(\nu e^{-s\alpha T_1})E_1(\nu e^{-s\alpha T_1})\right] = 1\,.$$

Combining this with (D.4), we obtain

$$\lim_{\tau\to\infty}\frac{1}{\tau - T_1}\int_{T_1}^\tau \gamma_1^\infty(t,e^{s\alpha})\,dt = 1\,. \tag{D.8}$$

D.2. A relation between the Poisson and Erlang distribution

We prove

$$\sum_{n=1}^\infty \mathrm{Poi}_{\Theta\tau}(n)\left(\sum_{i=1}^n \frac{m_{i,\Theta}(t)}{M_{i,\Theta}(\tau)}\right) = \Theta\,. \tag{D.9}$$

Indeed, we have

$$\begin{aligned}\sum_{n=1}^\infty \mathrm{Poi}_{\Theta\tau}(n)\left(\sum_{i=1}^n \frac{m_{i,\Theta}(t)}{M_{i,\Theta}(\tau)}\right) &= \sum_{n=1}^\infty e^{-\Theta\tau}\frac{(\Theta\tau)^n}{n!}\sum_{i=1}^n e^{-\Theta t}\frac{\Theta(\Theta t)^{i-1}}{\Gamma(i)M_{i,\Theta}(\tau)}\\ &= \Theta e^{-\Theta t}\sum_{i=1}^\infty \frac{(\Theta t)^{i-1}}{\Gamma(i)M_{i,\Theta}(\tau)}\sum_{n=i}^\infty e^{-\Theta\tau}\frac{(\Theta\tau)^n}{n!}\\ &= \Theta e^{-\Theta t}\sum_{i=1}^\infty \frac{(\Theta t)^{i-1}}{\Gamma(i)M_{i,\Theta}(\tau)}M_{i,\Theta}(\tau) = \Theta\,.\end{aligned} \tag{D.10}$$

The first equality is simply based on the definitions of $\mathrm{Poi}_{\Theta\tau}(n)$ in (2.3) and $m_{i,\Theta}(t)$ in (2.2). The second is obtained by reordering the sums, the third from a well-known relation between the Poisson distribution and the incomplete Gamma function (e.g. Abramowitz and Stegun, 1964, 6.5.13) and the definition of $M_{i,\Theta}(\tau)$ in (3.16), and the last from the exponential series.

D.3. Proof of Proposition 4.3

We will need the following integrals. Let $\phi(x)$ be a (continuous) function. Then we obtain directly from the definition of $\tilde{h}^{(i)}$ by a change in the order of integration:

$$\int_0^1 \phi(x)\tilde{h}^{(i)}_{\tau,\sigma_i,\Theta}(x)\,dx = \int_0^\tau \frac{m_{i,\Theta}(t)}{M_{i,\Theta}(\tau)}(1-P_{\text{ext}}([\tau-t],\sigma_i))\int_0^1 \phi(x)g_{a(\tau-t,\sigma_i)}(x)\,dx\,dt\,. \quad \text{(D.11)}$$

With $\phi(x)=x$, we get from (4.1a) and (D.11)

$$\int_0^1 x\tilde{h}^{(i)}_{\tau,\sigma_i,\Theta}(x)\,dx = \int_0^\tau \frac{m_{i,\Theta}(t)}{M_{i,\Theta}(\tau)}(1-P_{\text{ext}}([\tau-t],\sigma_i))\gamma_1(a(\tau-t,\sigma_i))\,dt\,. \quad \text{(D.12)}$$

With $\phi(x)=x(1-x)$, we get from (4.1c) and (D.11)

$$\int_0^1 x(1-x)\tilde{h}^{(i)}_{\tau,\sigma_i,\Theta}(x)\,dx = \int_0^\tau \frac{m_{i,\Theta}(t)}{M_{i,\Theta}(\tau)}(1-P_{\text{ext}}([\tau-t],\sigma_i))\gamma(a(\tau-t,\sigma_i))\,dt\,. \quad \text{(D.13)}$$

Proof of (4.14). From the definition (4.12) of $\bar{G}(\tau)$ we deduce:

$$\begin{aligned}
\bar{G}(\tau) &= \sum_{n=1}^\infty \text{Poi}_{\Theta\tau}(n)\left(\sum_{i=1}^n \int_0^\infty \alpha\,\text{E}\left[\tilde{X}^{(i)}_{\tau,e^{s\alpha},\Theta}\right] f(\alpha)\,d\alpha\right) && \text{(D.14a)}\\
&= \sum_{n=1}^\infty \text{Poi}_{\Theta\tau}(n)\int_0^\infty \alpha f(\alpha)\left(\sum_{i=1}^n \int_0^1 x\tilde{h}^{(i)}_{\tau,e^{s\alpha},\Theta}(x)\,dx\right)d\alpha && \text{(D.14b)}\\
&= \sum_{n=1}^\infty \text{Poi}_{\Theta\tau}(n)\int_0^\infty \alpha f(\alpha)\\
&\qquad \times \int_0^\tau \left(\sum_{i=1}^n \frac{m_{i,\Theta}(t)}{M_{i,\Theta}(\tau)}\right)(1-P_{\text{ext}}([\tau-t],e^{s\alpha}))\gamma_1(a(\tau-t,e^{s\alpha}))\,dt\,d\alpha && \text{(D.14c)}\\
&= \Theta\int_0^\infty \alpha f(\alpha)\int_0^\tau (1-P_{\text{ext}}([\tau-t],e^{s\alpha}))\gamma_1(a(\tau-t,e^{s\alpha}))\,dt\,d\alpha\,. && \text{(D.14d)}
\end{aligned}$$

We obtained (D.14a) by using that the mutation effects α_i are drawn independently from the same distribution f; (D.14b) by using (4.11); and (D.14c) by employing (D.12) and changing the order of integration and summation. Then (D.14c) yields (D.14d) by applying (D.9), and (D.14d) yields (4.14) by applying the time transformation $\tau - t \to T$, and returning to t (instead of T). Note also that we suppressed the dependence on N in $\tilde{X}^{(i)}$, $\tilde{h}^{(i)}$, and a. □

Proof of (4.15). Analogously, by using (D.13) instead of (D.12), we obtain for the variance from its definition (4.13),

$$\begin{aligned}
V_G(\tau) &= \sum_{n=1}^\infty \text{Poi}_{\Theta\tau}(n)\left(\sum_{i=1}^n \int_0^\infty \alpha^2\,\text{E}\left[\tilde{X}^{(i)}_{\tau,e^{s\alpha},\Theta}(1-\tilde{X}^{(i)}_{\tau,e^{s\alpha},\Theta})\right] f(\alpha)\,d\alpha\right)\\
&= \sum_{n=1}^\infty \text{Poi}_{\Theta\tau}(n)\int_0^\infty \alpha^2 f(\alpha)\left(\sum_{i=1}^n \int_0^1 x(1-x)\tilde{h}^{(i)}_{\tau,e^{s\alpha},\Theta}(x)\,dx\right)d\alpha\\
&= \sum_{n=1}^\infty \text{Poi}_{\Theta\tau}(n)\int_0^\infty \alpha^2 f(\alpha)
\end{aligned}$$

$$\times \int_0^\tau \left(\sum_{i=1}^n \frac{m_{i,\Theta}(t)}{M_{i,\Theta}(\tau)} \right) (1 - P_{\text{ext}}([\tau - t], \sigma_i)) \gamma(a(\tau - t, \sigma_i))\, dt\, d\alpha \tag{D.15a}$$

$$= \Theta \int_0^\infty \alpha^2 f(\alpha) \int_0^\tau (1 - P_{\text{ext}}([\tau - t], e^{s\alpha})) \gamma(a(\tau - t, e^{s\alpha}))\, dt\, d\alpha\,, \tag{D.15b}$$

where (D.15a) yields (D.15b) with the help of (D.9). Finally, (4.15) is obtained by the time transformation $\tau - t \to T$ (and returning from T to t). Note that for the derivation of the variance, the independence of the random variables $\tilde{X}^{(i)}$ is crucial. □

D.4. Proof of Proposition 4.5

Proof of (4.21). From (4.14) we obtain

$$\bar{G}(\tau + 1) - \bar{G}(\tau) = \Theta \int_0^\infty \alpha f(\alpha) H(e^{s\alpha}, \tau)\, d\alpha\,, \tag{D.16}$$

where

$$\begin{aligned} H(e^{s\alpha}, \tau) &= \int_0^{\tau+1} (1 - P_{\text{ext}}([t], e^{s\alpha})) \gamma_1(t, e^{s\alpha})\, dt \\ &\quad - \int_0^\tau (1 - P_{\text{ext}}([t], e^{s\alpha})) \gamma_1(t, e^{s\alpha})\, dt \\ &= \int_\tau^{\tau+1} (1 - P_{\text{ext}}([t], e^{s\alpha})) \gamma_1(t, e^{s\alpha})\, dt\,. \end{aligned} \tag{D.17}$$

Now we use $\lim_{\tau\to\infty} (1 - P_{\text{ext}}([\tau], e^{s\alpha})) = P_{\text{sur}}(e^{s\alpha})$, $\lim_{\tau\to\infty} \frac{\gamma_1^\infty(\tau, e^{s\alpha})}{\gamma_1(\tau, e^{s\alpha})} = 1$ (by continuity of γ_1 and the fact that $\lim_{\tau\to\infty} \frac{a^\infty(\tau, e^{s\alpha})}{a(\tau, e^{s\alpha})} = 1$), and

$$\lim_{\tau\to\infty} \int_\tau^{\tau+1} \gamma_1(t, e^{s\alpha})\, dt = \lim_{\tau\to\infty} \int_\tau^{\tau+1} \gamma_1^\infty(t, e^{s\alpha})\, dt = 1\,, \tag{D.18}$$

where we used (D.6). Therefore, $\lim_{\tau\to\infty} H(e^{s\alpha}, \tau) = P_{\text{sur}}(e^{s\alpha})$ and (D.16) yields

$$\Delta \bar{G}^* = \lim_{\tau\to\infty} \Big(\bar{G}(\tau + 1) - \bar{G}(\tau) \Big) = \Theta \int_0^\infty \alpha f(\alpha) P_{\text{sur}}(e^{s\alpha})\, d\alpha\,, \tag{D.19}$$

which is (4.21). □

Proof of (4.22). We start from

$$V_G(\tau) = \Theta \int_0^\infty \alpha^2 f(\alpha) \int_0^\tau (1 - P_{\text{ext}}([t], e^{s\alpha})) \gamma(t, e^{s\alpha})\, dt\, d\alpha\,. \tag{4.15}$$

In order to derive (4.22), we need to estimate

$$\begin{aligned} D_V(\tau) &= \int_0^\infty \alpha^2 f(\alpha) \int_0^\tau (1 - P_{\text{ext}}([t], e^{s\alpha})) \gamma(t, e^{s\alpha})\, dt\, d\alpha \\ &\quad - \int_0^\infty \alpha^2 f(\alpha) \int_0^\tau (1 - P_{\text{ext}}^\infty(e^{s\alpha})) \gamma^\infty(t, e^{s\alpha})\, dt\, d\alpha \end{aligned}$$

$$= \int_0^\infty \alpha^2 f(\alpha) \int_0^\tau I_V(t, e^{s\alpha})\, dt\, d\alpha \tag{D.20}$$

as $\tau \to \infty$, where

$$I_V(t, e^{s\alpha}) = (1 - P_{\text{ext}}([t], e^{s\alpha}))\gamma(t, e^{s\alpha}) - (1 - P_{\text{ext}}^\infty(e^{s\alpha}))\gamma^\infty(t, e^{s\alpha})\,. \tag{D.21}$$

With $D_V(\infty) = \lim_{\tau\to\infty} D_V(\tau)$ we obtain

$$\begin{aligned} V_G^*/\Theta = \lim_{\tau\to\infty} V_G(\tau)/\Theta &= \int_0^\infty \alpha^2 f(\alpha)(1 - P_{\text{ext}}^\infty(e^{s\alpha})) \int_0^\infty \gamma^\infty(t, e^{s\alpha})\, dt\, d\alpha + D_V(\infty) \\ &= \int_0^\infty \frac{\alpha}{s} f(\alpha) P_{\text{sur}}(e^{s\alpha}) \nu e^\nu E_1(\nu)\, d\alpha + D_V(\infty)\,, \end{aligned} \tag{D.22}$$

where we have used (D.3b) in the last step. Because V_G^*/Θ is of order 1 ($xe^x E_1(x) < 1$ and $P_{\text{sur}} \le \rho s\alpha$), proving that $|D_V(\infty)|$ becomes arbitrarily small as $N \to \infty$ (in a sense to specified), is sufficient to establish the desired result.

In view of the approximations (4.29) and (4.30), it is desirable to have an error of order $o(s) + o(1/(Ns))$ as $Ns \to \infty$ and $s \to 0$. For this reason, we will need the scaling Assumption A1, i.e., we start with $Ns^K = C^K$ as $N \to \infty$, where $C > 0$ and $K > 1$ are constants, and K will be determined at the end of the proof.

(1) In this first main step we fix the mutational effect α. Then we deal with the complications arising from drawing α from a distribution f with mean $\bar{\alpha}$ in a second main step.

Let $\epsilon > 0$ be sufficiently small, where we will quantify this at the end of step (1). From the definition of T_ϵ in (A.10) and its explicit expression in (A.21a), we infer

$$1 - P_{\text{ext}}^\infty < 1 - P_{\text{ext}}([t]) \le (1+\epsilon)(1 - P_{\text{ext}}^\infty) \text{ for every } t \ge T_\epsilon\,, \tag{D.23}$$

where here and below we suppress the dependence of $P_{\text{ext}}([t])$, P_{ext}^∞, and T_ϵ on $e^{s\alpha}$.

In the limit $\tau \to \infty$, we split the integral with respect to time in (D.20) as follows

$$\int_0^\infty I_V(t, e^{s\alpha})\, dt = \int_0^{T_\epsilon} I_V(t, e^{s\alpha})\, dt + \int_{T_\epsilon}^\infty I_V(t, e^{s\alpha})\, dt\,. \tag{D.24}$$

(i) First, we treat $\int_{T_\epsilon}^\infty I_V(t, e^{s\alpha})\, dt$ by assuming $t \ge T_\epsilon$. We recall the notation from (D.1) and (D.2). Then (D.23) implies $a^\infty(t, e^{s\alpha}) < a(t, e^{s\alpha}) \le (1+\epsilon)a^\infty(t, e^{s\alpha})$. Using

$$a(t) = (1 + u(t)\epsilon)a^\infty(t)\,, \tag{D.25}$$

where $0 < u(t) \le 1$ if $t \ge T_\epsilon$ (indeed, $u(t) \to 0$ as $t \to \infty$, and for a fractional linear offspring distribution $u(t)$ can be calculated explicitly from (A.9b)), we obtain

$$(1 - P_{\text{ext}}([t]))\gamma(a(t)) = (1 + u(t)\epsilon)(1 - P_{\text{ext}}^\infty)\, \gamma^\infty((1 + u(t)\epsilon)a^\infty(t))\,. \tag{D.26}$$

This yields

$$\begin{aligned} I_V(t) &= (1 + u(t)\epsilon)(1 - P_{\text{ext}}^\infty)\, \gamma^\infty((1 + u(t)\epsilon)a^\infty(t)) - (1 - P_{\text{ext}}^\infty)\gamma^\infty(a^\infty(t)) \\ &= \epsilon u(t)(1 - P_{\text{ext}}^\infty)\gamma_D(a^\infty(t)) + \epsilon^2 u(t)^2 (1 - P_{\text{ext}}^\infty) R_1(a^\infty(t))\,, \end{aligned} \tag{D.27}$$

where

$$\gamma_D(x) = x\Big[\Big(2+4x+x^2\Big)e^x E_1(x) - 3 - x\Big] \tag{D.28}$$

and R_1 are obtained by series expansion for $\epsilon \to 0$. The function $\gamma_D(x)$ is positive and bounded (by ≈ 0.217) and can be integrated. If we write $T_\epsilon = \ln(Z)/(s\alpha)$ (with $Z = P_{\text{ext}}^\infty(1+\epsilon)/\epsilon$) and recall $a^\infty(t) = \nu e^{-s\alpha t}$ (with $\nu = N(1-P_{\text{ext}}^\infty)$), we obtain

$$\begin{aligned}\int_{T_\epsilon}^\infty \gamma_D(a^\infty(t))\,dt &= -\frac{\nu/Z}{s\alpha}\Big(1 - e^{\nu/Z}(2+\nu/Z)E_1(\nu/Z)\Big)\\ &= \frac{1}{s\alpha}\Big(1 - 2(\nu/Z)^{-2} + O\Big((\nu/Z)^{-3}\Big)\Big)\end{aligned} \tag{D.29}$$

as $\nu/Z \to \infty$. In essentially the same way

$$\int_{T_\epsilon}^\infty R_1(a^\infty(t))\,dt \le \frac{4}{s\alpha}(\nu/Z)^{-2} \tag{D.30}$$

follows.

In order to satisfy $\nu/Z \to \infty$, we need $Ns\epsilon \to \infty$ as $N \to \infty$. In particular, it follows from the above results that $I_V(t) > 0$ if $t \ge T_\epsilon$. Therefore, we obtain

$$\begin{aligned}\int_{T_\epsilon}^\tau I_V(t)\,dt &\le \int_{T_\epsilon}^\infty I_V(t)\,dt\\ &\le \epsilon(1-P_{\text{ext}}^\infty)\int_{T_\epsilon}^\infty \gamma_D(a^\infty(t))\,dt + \epsilon^2(1-P_{\text{ext}}^\infty)\int_{T_\epsilon}^\infty R_1(a^\infty(t))\,dt\\ &\le \epsilon\rho\Big(1 + O\big((N\epsilon)^{-2}\big)\Big) + \epsilon^2 O\Big((N\epsilon)^{-2}\Big) \le 2\epsilon\rho\,.\end{aligned} \tag{D.31}$$

Here we used (D.27) together with $0 < u(t) \le 1$ in the second estimate, and (D.29) and (D.30) together with $1 - P_{\text{ext}}^\infty \le \rho s\alpha$ (Remark 3.3) in the third.

Thus, to achieve our desired result, we will need the conditions $\epsilon = o(s)$, $\epsilon = o((Ns)^{-1})$, and $(N\epsilon)^{-1} = o(1)$.

(ii) Second, we derive an estimate for $\int_0^{T_\epsilon} I_V(t, e^{s\alpha})\,dt$, where we recall $I_V(t, e^{s\alpha})$ from (D.21). For small t, $a(t)$ is large because it is decreasing from N at $t = 0$ to $a(T_\epsilon) \approx \rho\alpha N s\epsilon$ (if $s\alpha$ is small), where we recall $a(t) = N(1 - P_{\text{ext}}([t], e^{s\alpha}))e^{-s\alpha t}$. We will choose ϵ and the scaling factor K such that $Ns\epsilon \to \infty$ as $N \to \infty$. Then we use the approximation $\gamma(a(t)) = \frac{1}{a(t)} - \frac{4}{a(t)^2} + O(a(t)^{-3})$ which is readily derived from (D.7). Therefore, we obtain by using the expansion of γ:

$$\begin{aligned}I_V(t, e^{s\alpha}) &= (1 - P_{\text{ext}}([t]))\gamma(t) - (1 - P_{\text{ext}}^\infty)\gamma^\infty(t)\\ &= \Big(\frac{e^{s\alpha t}}{N} - \frac{4e^{2s\alpha t}}{N^2(1 - P_{\text{ext}}([t]))}\Big) - \Big(\frac{e^{s\alpha t}}{N} - \frac{4e^{2s\alpha t}}{N^2(1 - P_{\text{ext}}^\infty)}\Big) + O(a(t)^{-3})\\ &= \frac{4e^{2s\alpha t}}{N^2}\Big(\frac{1}{1 - P_{\text{ext}}^\infty} - \frac{1}{1 - P_{\text{ext}}([t])}\Big) + O(a(t)^{-3})\,,\end{aligned} \tag{D.32}$$

where again we have suppressed the dependence on $e^{s\alpha}$ in the terms a and P_{ext}.

From (3.10) we obtain by a brief calculation

$$\frac{1}{1 - P_{\text{ext}}^\infty} - \frac{1}{1 - P_{\text{ext}}([t])} \le \frac{P_{\text{ext}}^\infty}{1 - P_{\text{ext}}^\infty}\, e^{-s\alpha t}\,. \tag{D.33}$$

Now we integrate $I_V(t, e^{s\alpha})$ with respect to t:

$$\begin{aligned}\int_0^{T_\epsilon} I_V(t, e^{s\alpha})\, dt &\le \frac{4}{N^2}\frac{P_{\text{ext}}^\infty}{1-P_{\text{ext}}^\infty}\int_0^{T_\epsilon} e^{s\alpha t}\, dt + R \\ &= \frac{4}{N^2}\frac{P_{\text{ext}}^\infty}{1-P_{\text{ext}}^\infty}\frac{e^{s\alpha T_\epsilon}-1}{s\alpha} + R \\ &= \frac{4}{N^2}\frac{P_{\text{ext}}^\infty}{1-P_{\text{ext}}^\infty}\frac{P_{\text{ext}}^\infty(1+\epsilon)-\epsilon}{s\alpha\epsilon} + R \\ &\le \frac{4}{N^2 s\alpha\epsilon(1-P_{\text{ext}}^\infty)} + R \\ &\le \frac{8}{\rho\alpha^2}\frac{1}{(Ns)^2\epsilon} + \frac{1}{\alpha^2}\, O\Big(\frac{1}{(Ns)^3\epsilon}\Big)\,, \end{aligned} \tag{D.34}$$

where R arises from integration of terms of order $a(t)^{-3}$ or smaller, hence is of smaller order than the first term, and $e^{s\alpha T_\epsilon} = P_{\text{ext}}^\infty(1+\epsilon)/\epsilon$ (see (A.21a)). In the final step we also used $(1-P_{\text{ext}}^\infty) \ge \frac{1}{2}\rho s\alpha$ for small $s\alpha$. To keep the error in (D.34) sufficiently small, we need that $((Ns)^2\epsilon)^{-1} = o((Ns)^{-1})$ (and $((Ns)^2\epsilon)^{-1} = o(s)$, which is weaker).

In addition to $Ns^K = C^K$, choose $\epsilon = s^{K_1}$, where $K_1 > 0$. The lower bound $K > 1$ arises from the fact that in our model selection is stronger than random genetic drift. With this choice, we obtain $s = O(N^{-1/K})$, $\epsilon = O(N^{-K_1/K})$, $1/(Ns) = O(N^{-1+1/K})$, $1/((Ns)^2\epsilon) = O(N^{-2+(2+K_1)/K})$, and $Ns\epsilon = O(N^{1-(1+K_1)/K})$. To keep all error terms sufficiently small, we need $\epsilon = o(s)$, $1/((Ns)^2\epsilon) = o((Ns)^{-1})$, $N\epsilon \to \infty$ and $Ns\epsilon \to \infty$ as $N \to \infty$. These inequalities are satisfied if and only if $K_1 > 1$ and $K > K_1 + 1$. With this choice, we get error terms of smaller order than s and $a(T_\epsilon) = \rho\alpha Ns\epsilon \to \infty$ (as $N \to \infty$), which is required in (i) and (ii) above. In summary, we obtain

$$\int_0^{T_\epsilon} I_V(t, e^{s\alpha})\, dt = O\Big(N^{-2+(2+K_1)/K}\Big)\,, \tag{D.35a}$$

$$\int_{T_\epsilon}^{\infty} I_V(t, e^{s\alpha})\, dt = O\Big(N^{-K_1/K}\Big)\,, \tag{D.35b}$$

where $K_1/K < 2 - (2+K_1)/K$ because $K_1 > 1$ and $K > K_1 + 1$. A simple choice for K_1 and K is $K_1 = \frac{3}{2}$ and $K = 3$. Then $s = O(N^{-1/3})$, $\epsilon = O(N^{-1/2})$, $1/(Ns) = O(N^{-2/3})$, $1/((Ns)^2\epsilon) = 1/((Ns^3)(Ns^{1/2})) = O(N^{-5/6})$, and $Ns\epsilon = O(N^{1/6})$. Then the error term in (D.35a) is $O(N^{-5/6})$, and that in (D.35b) is $O(N^{-1/2})$.

(2) In this second main step we assume that exponentially distributed effects α with mean $\bar\alpha$. However, except for some constants occurring during integration with respect to α, the same order estimates are obtained for a f having mean $\bar\alpha$ and bounded higher moments. We adopt the scaling assumption from step (1) with $K_1 > 1$ and $K > K_1 + 1$.

(i) First we treat the case $t \le T_\epsilon$. We set $z = K_2 \ln(Ns)(\ge 1)$, where we choose s small enough such that $P_{\text{sur}} \ge \frac{1}{2}\rho s\alpha$ holds for every $\alpha \le z\bar\alpha$. This is possible by Remark 3.3 and our scaling assumption. Furthermore, we choose $\epsilon = r(N)/\min\{\alpha, z\bar\alpha\}$, where $r(N) = N^{-K_1/K}$. Assuming $\alpha \le z\bar\alpha$ and using (A.21a), we obtain

$$\begin{aligned} a(T_\epsilon) = N(1+\epsilon)P_{\text{sur}}e^{-s\alpha T_\epsilon} &= N(1+\epsilon)\frac{P_{\text{sur}}}{P_{\text{ext}}^\infty}\frac{\epsilon}{1+\epsilon} \\ &\ge \tfrac{1}{2}\rho Ns\, r(N)\frac{\alpha}{\min\{\alpha, z\bar\alpha\}} = \tfrac{1}{2}\rho\, Nsr(N)\,, \end{aligned} \tag{D.36}$$

and this inequality holds for every $t \le T_\epsilon$. Therefore, the bound (and approximation) $\gamma(a(t)) \le \frac{1}{a(t)}$ will be accurate in this parameter region since $a(t) \to \infty$ as $N \to \infty$.

Therefore, we can use (D.34) and obtain

$$\begin{aligned}\int_0^{z\bar\alpha} \alpha^2 f(\alpha) \int_0^{T_\epsilon} I_V(t, e^{s\alpha})\, dt\, d\alpha &\le \frac{2\tilde c}{(Ns)^2 r(N)} \int_0^{z\bar\alpha} \alpha^2 f(\alpha) \frac{\alpha}{\alpha^2}\, d\alpha\,, \\ &\le \frac{2\tilde c \bar\alpha}{(Ns)^2 r(N)} \\ &= O\Big(N^{-2+(2+K_1)/K}\Big)\,, \end{aligned} \tag{D.37}$$

where we have used $s\alpha T_\epsilon \le \ln\Big(1 + \frac{\min\{\alpha, z\bar\alpha\}}{r(N)}\Big) = \ln\Big(1 + \frac{\alpha}{r(N)}\Big)$ if $\alpha \le z\bar\alpha$, and $\tilde c > 1$ is a constant which also accounts for the error term R. Thus, the order of the error term is the same as in (D.35a).

If $\alpha > z\bar\alpha$, we use the brutal estimate $I_V(t) \le \gamma(t) < 1$ and observe that $T_\epsilon \le \frac{1}{s\alpha} \ln\Big(1 + \frac{z\bar\alpha}{r(N)}\Big)$. Straightforward integration shows that

$$\begin{aligned}\int_{z\bar\alpha}^\infty \alpha^2 f(\alpha) \int_0^{T_\epsilon} I_V(t)\, dt\, d\alpha &\le \int_{z\bar\alpha}^\infty \alpha^2 f(\alpha) \int_0^{T_\epsilon} 1\, dt\, d\alpha \\ &= \int_{z\bar\alpha}^\infty \alpha^2 f(\alpha) T_\epsilon\, d\alpha \\ &\le \frac{1}{s} \ln\Big(1 + \frac{z\bar\alpha}{r(N)}\Big) \int_{z\bar\alpha}^\infty \alpha f(\alpha)\, d\alpha \\ &= \frac{1}{s} \ln\Big(1 + \frac{z\bar\alpha}{r(N)}\Big) \bar\alpha e^{-z}(1+z) \\ &= \frac{\bar\alpha}{s(Ns)^{K_2}} \ln\Big(1 + \frac{\bar\alpha K_2 \ln(Ns)}{r(N)}\Big)(1 + K_2 \ln(Ns)) \end{aligned} \tag{D.38}$$

and this holds for every $\tau > 0$. With $K_2 = K$ (which is not best possible), we obtain $s(Ns)^{K_2} = O(N^{K-1-1/K})$ which, by our scaling assumption, is is of greater order than the error term $O(N^{2-(2+K_1)/K})$ arising in (D.37). Therefore, the error arising from (D.38) is of smaller order than that in (D.37). For instance, with $K_2 = K = 3$, we have $(s(Ns)^{K_2})^{-1} = O(N^{-5/3})$ so that, after taking care of the logarithmic terms, the error in (D.38) is $o(N^{-4/3})$, hence smaller than $O(N^{-5/6})$ in (D.37).

(ii) Now we assume $t > T_\epsilon$. Then we obtain from (D.31)

$$\begin{aligned}\int_0^\infty \alpha^2 f(\alpha) \int_{T_\epsilon}^\tau I_V(t)\, dt\, d\alpha &\le 2\rho \int_0^\infty \alpha^2 f(\alpha) \frac{r(N)}{\min\{\alpha, z\bar\alpha\}}\, d\alpha \\ &= 2\rho r(N) \int_0^{z\bar\alpha} \alpha f(\alpha)\, d\alpha + \frac{2\rho}{z\bar\alpha} r(N) \int_{z\bar\alpha}^\infty \alpha^2 f(\alpha)\, d\alpha \\ &< 2\rho r(N)\bar\alpha + 4\rho\bar\alpha r(N) = O\Big(N^{-K_1/K}\Big) \end{aligned} \tag{D.39}$$

(because $z \ge 1$). With our standard choice of $K_1 = \frac{3}{2}$ and $K = 3$, this is $O(N^{-1/2})$.

In summary, we obtain $0 < D_V(\infty) \le O(N^{-K_1/K})$ as $N \to \infty$. □

Because for fractional linear offspring distributions with $v = 2m/\rho$, we have $P_{\text{ext}}^\infty = 0$ if

$m \geq 1+2/\rho$ (Appendix A.3), in some of the integrals above the upper bound ∞ should be reduced accordingly because then $I_V(t) = 0$ for large α.

Remark D.1. We provide an approximation for the time dependence of $\bar{G}(\tau)$ for sufficiently large τ. We assume $\tau \gg \bar{t}_{\text{fix}}$ (in particular, $\tau > \tau_c$) and suppress the dependence of a and ν on N. Then we obtain from (4.14)

$$\begin{aligned}
\bar{G}(\tau) &= \Theta \int_0^\infty \alpha f(\alpha) \int_0^\tau (1 - P_{\text{ext}}([t], e^{s\alpha}))\gamma_1(a(t, e^{s\alpha}))\, dt\, d\alpha \\
&\approx \Theta \int_0^\infty \alpha f(\alpha)(1 - P_{\text{ext}}^\infty(e^{s\alpha})) \int_0^\tau \gamma_1(a(t, e^{s\alpha}))\, dt\, d\alpha && \text{(D.40a)} \\
&= \Theta \int_0^\infty \alpha f(\alpha) P_{\text{sur}}(e^{s\alpha}) \Big(\int_{\bar{t}_{\text{fix}}}^\tau \gamma_1(a(t))\, dt + \int_0^{\bar{t}_{\text{fix}}} \gamma_1(a(t))\, dt \Big) d\alpha && \text{(D.40b)} \\
&\approx \Theta(\tau - \bar{t}_{\text{fix}}) \int_0^\infty \alpha f(\alpha) P_{\text{sur}}(e^{s\alpha})\, d\alpha \\
&\quad + \Theta \int_0^\infty f(\alpha) \frac{1}{s} P_{\text{sur}}(e^{s\alpha}) \left[\exp(\nu e^{-s\alpha \bar{t}_{\text{fix}}}) E_1(\nu e^{-s\alpha \bar{t}_{\text{fix}}}) - e^\nu E_1(\nu) \right] d\alpha\,. && \text{(D.40c)}
\end{aligned}$$

Here, we used $P_{\text{ext}}([\tau - t], e^{s\alpha}) \approx P_{\text{ext}}^\infty(e^{s\alpha})$ for sufficiently large τ to obtain (D.40a), a simple integral splitting (and $P_{\text{sur}} = 1 - P_{\text{ext}}^\infty$) for (D.40b), and finally (D.8) and (D.5) to arrive at (D.40c).

The second term in (D.40c) accounts for the contribution of the segregating mutations to the phenotypic mean. On average, these are the mutations that appeared less than $\bar{t}_{\text{fix}}$ generations in the past and have not yet had enough time to become fixed. A permanent response of the mean is achieved by mutations that have gone to fixation. These are accounted for by the first term in (D.40c), which is consistent with (4.21).

D.5. Proof of Corollary 4.8

(1) The approximation (4.25) follows directly from (4.21) by using (3.8).

(2) For equal effects, (4.25) immediately simplifies to (4.26) by using (3.8).

(3) follows from (4.25) by the properties of the exponential distribution.

(4) Using $\nu e^\nu E_1(\nu) \approx 1 - 1/\nu$ (because $\nu = N P_{\text{sur}}(e^{s\alpha}) \gg 1$) in (4.22) we obtain

$$\begin{aligned}
V_G^*(f, s, N, \Theta) &\approx \Theta \int_0^\infty \frac{\alpha}{s} P_{\text{sur}}(e^{s\alpha}) \Big(1 - \frac{1}{\nu}\Big) f(\alpha)\, d\alpha && \text{(D.41a)} \\
&= \frac{\Theta}{s} \int_0^\infty \alpha P_{\text{sur}}(e^{s\alpha}) f(\alpha)\, d\alpha - \frac{\Theta}{sN} \int_0^\infty \alpha f(\alpha)\, d\alpha\,, && \text{(D.41b)}
\end{aligned}$$

which yields (4.28).

We prove the exponential case (6) before (5).

(6) We require Assumption A1. In (4.22), we substitute $2Ns\alpha$ for ν and $2s\alpha - \frac{5}{3}(s\alpha)^2$ for $P_{\text{sur}}(e^{s\alpha})$ to obtain

$$\begin{aligned}
V_G^*(f, s, N, \Theta) &\approx 4\Theta N s \int_0^\infty \alpha^3 \Big(1 - \frac{5}{6} s\alpha\Big) e^{2Ns\alpha} E_1(2Ns\alpha) f(\alpha)\, d\alpha && \text{(D.42a)} \\
&= \frac{4\Theta N s \bar{\alpha}^3}{(2Ns\bar{\alpha} - 1)^4} \left[(2Ns\bar{\alpha} - 1)(8(Ns\bar{\alpha})^2 - 14Ns\bar{\alpha} + 11) - 6\ln(2Ns\bar{\alpha}) \right]
\end{aligned}$$

$$-\frac{20\Theta N s^2\bar{\alpha}^4}{3(2Ns\bar{\alpha}-1)^5}\Big[(2Ns\bar{\alpha}-1)\big(24(Ns\bar{\alpha})^3-52(Ns\bar{\alpha})^2+46Ns\bar{\alpha}-25\big)+12\ln(2Ns\bar{\alpha})\Big]\,. \tag{D.42b}$$

By letting $N\to\infty$, we obtain to order $1/N$:

$$V_G^*(f,s,N,\Theta)\approx 4\Theta\bar{\alpha}^2\left(1-\frac{5}{2}s\bar{\alpha}-\frac{1}{4Ns\bar{\alpha}}+\frac{5}{12N}\right)\,, \tag{D.43}$$

where Assumption A1 yields $s=O(N^{-1/K})$ and $1/(Ns)=O(N^{-1+1/K})$, where $K>2$. This yields (4.30) by omitting the term of order N^{-1}.

(5) For equal effects, (D.42a) simplifies to $4\Theta Ns\alpha^3\Big(1-\frac{5}{6}s\alpha\Big)e^{2Ns\alpha}E_1(2Ns\alpha)$. Using $e^xE_1(x)\approx x^{-1}-x^{-2}$ if $x\to\infty$ (with $x=2Ns\alpha$ and the asymptotic assumption in (6)), we obtain

$$V_G^*(\alpha,s,N,\Theta)\approx 2\Theta\alpha^2\left(1-\frac{5}{6}s\alpha-\frac{1}{2Ns\alpha}+\frac{5}{12N}\right)\,, \tag{D.44}$$

which yields (4.29), in which the term of smallest order, $\frac{5}{12N}$, has been omitted.

D.6. Proof of Proposition 4.11

The proof of the approximation (4.34) for $V_G(\tau)$ is based on that of the approximation (4.22) for the stationary variance V_G^*.

We recall the definition of $D_V(\tau)$ from (D.20) and that of $I_V(t)$ from (D.21). Then we obtain

$$\begin{aligned}V_G(\tau)&=\Theta\int_0^\infty\alpha^2f(\alpha)(1-P_{\text{ext}}^\infty)\int_0^\tau\gamma^\infty(t,e^{s\alpha})\,dt\,d\alpha+\Theta D_V(\tau)\\&=\Theta\int_0^\infty\frac{\alpha}{s}P_{\text{sur}}\Big(\nu e^\nu E_1(\nu)-\nu e^{-s\alpha\tau}\exp(\nu e^{-s\alpha\tau})E_1(\nu e^{-s\alpha\tau})\Big)f(\alpha)\,d\alpha\\&\quad+\Theta D_V(\tau)\,,\end{aligned} \tag{D.45}$$

where the second equality follows from (D.3a).

As in step (2) of the proof of (4.34), we choose $T_\epsilon=\frac{1}{s\alpha}\ln\Big(P_{\text{ext}}^\infty\frac{1+\epsilon}{\epsilon}\Big)$, with $\epsilon=r(N)/\min\{\alpha,z\bar{\alpha}\}$, $z=K_2\ln(Ns)$ and, $r(N)=N^{-K_1/K}$.

If $\tau\le T_\epsilon$, we obtain from (D.37) and (D.38)

$$D_V(\tau)\le D_V(T_\epsilon)=O\Big(N^{-2+(2+K_1)/K}\Big) \tag{D.46}$$

because the integrands $I_V(t)$ and their bounds are nonnegative. If $\tau>T_\epsilon$, we obtain from (D.39)

$$D_V(\tau)\le D_V(\infty)=O\Big(N^{-K_1/K}\Big)\,. \tag{D.47}$$

We leave the proof of (4.33) for $\bar{G}(\tau)$ to the interested reader. It is based on analogous estimates. The main difference is that instead of γ_D in (D.28) the function

$$\tilde{\gamma}_D(x)=1+x-(2+x)xE_1(x) \tag{D.48}$$

occurs, which has a diverging integral. However, instead of $u(t) \le 1$ we can use the estimate

$$u(t) \le \frac{P_{\text{ext}}^{\infty} e^{-s\alpha t}}{\epsilon(1 - P_{\text{ext}}^{\infty} e^{-s\alpha t})} \le \frac{P_{\text{ext}}^{\infty} e^{-s\alpha t}}{\epsilon(1 - P_{\text{ext}}^{\infty})}\,, \tag{D.49}$$

which follows from (3.10). Then

$$\int_{T_\epsilon}^{\infty} e^{-s\alpha t} \tilde{\gamma}(a^{\infty}(t))\, dt = \frac{1}{s\alpha\nu} + \text{terms of lower order} \tag{D.50}$$

replaces (D.29). However, the leading-order term is now $1/(s\alpha\nu)$, from which the stronger requirement $K > 3$ is deduced.

D.7. Proof of Proposition 4.13

Proof of (4.37). We recall Assumption A1 and let $N \to \infty$. As a consequence $Ns \to \infty$ and $s \to 0$. The mean mutational effect $\bar{\alpha}$ is given and fixed. From the asymptotic expansion (D.7) and the definition of γ_1 in (4.1a), we obtain the asymptotic equivalence

$$\gamma_1(a) = 1 - ae^a E_1(a) \sim 1/a - 2/a^2 \text{ as } a \to \infty\,. \tag{D.51}$$

In addition, $\max\{0, 1/a - 2/a^2\} < \gamma_1(a) < \min\{1/a, 1\}$ holds for every $a > 0$ and $\gamma_1(a(t))$ is monotonically increasing from 0 to 1 as t increases from 0 to ∞; see text below (4.1c).

The basic idea of the proof is to use the approximation

$$\gamma_1(a(t, e^{s\alpha}, N)) \approx \frac{1}{a(t, e^{s\alpha}, N)} = \frac{e^{s\alpha t}}{N(1 - P_{\text{ext}}([t], e^{s\alpha}))} \tag{D.52}$$

for small t because $a(t, e^{s\alpha}, N)$, defined in (3.15b), is monotonically decreasing in t from $a(0, e^{s\alpha}, N) = N$ to 0 as t increases to ∞. Then one could proceed as follows by first substituting (D.52) into (4.14):

$$\bar{G}(\tau) = \Theta \int_0^{\infty} \alpha f(\alpha) \int_0^{\tau} (1 - P_{\text{ext}}([t], e^{s\alpha}))\gamma_1(a(t, e^{s\alpha}, N))\, dt\, d\alpha \tag{D.53a}$$

$$\approx \frac{\Theta}{N} \int_0^{\infty} \alpha f(\alpha) \int_0^{\tau} e^{s\alpha t}\, dt\, d\alpha \tag{D.53b}$$

$$= \frac{\Theta}{Ns} \int_0^{\infty} f(\alpha)(e^{s\alpha\tau} - 1)\, d\alpha\,. \tag{D.53c}$$

The order of the error of the approximation in (D.53b) can be obtain by integrating $2/a(t, e^{s\alpha}, N)^2$ accordingly. However, on the way several technical problems occur. In particular, if mutation effects are drawn from a distribution, in our case the exponential distribution with mean $\bar{\alpha}$, then for given τ, s, and N, the approximation (D.53) will be accurate if α is sufficiently small, but will fail if α is extremely large because then $a(\tau, e^{s\alpha}, N)$ can drop from N to close to 0 in the first generation.

Therefore, we will proceed as in (D.53) only if $\alpha \le z\bar{\alpha}$, where we choose $z = \tilde{K} \ln(Ns\bar{\alpha})$ and determine $\tilde{K}(\ge 1)$ below. Because we assume an exponential mutation distribution f with mean $\bar{\alpha}$, we have $\text{Prob}[\alpha > z\bar{\alpha}] = e^{-z}$. If $\alpha > z\bar{\alpha}$, we use $0 < \gamma_1(a) < 1$ (see above). Straightforward

integration shows that

$$
\begin{aligned}
0 &< \int_{z\bar\alpha}^{\infty} \alpha f(\alpha) \int_0^\tau (1 - P_{\text{ext}}([t], e^{s\alpha}))\gamma_1(a(t, e^{s\alpha}, N))\, dt\, d\alpha \\
&< \int_{z\bar\alpha}^{\infty} \alpha f(\alpha) \int_0^\tau 1\, dt\, d\alpha \\
&= \tau\bar\alpha e^{-z}(1+z) \\
&= \tau\bar\alpha\, \frac{1+\tilde K \ln(Ns\bar\alpha)}{(Ns\bar\alpha)^{\tilde K}} \\
&= \tau\, O\Big(\frac{\ln(Ns)}{(Ns)^{\tilde K}}\Big) \quad \text{as } N\to\infty\,,
\end{aligned}
\tag{D.54}
$$

and this holds for every $\tau > 0$.

For $\alpha \le z\bar\alpha$, we start from (4.14) and use (D.51) and (D.52) to obtain

$$
\begin{aligned}
&\int_0^{z\bar\alpha} \alpha f(\alpha) \int_0^\tau (1 - P_{\text{ext}}([t], e^{s\alpha}))\gamma_1(a(t, e^{s\alpha}, N))\, dt\, d\alpha \\
&= \int_0^{z\bar\alpha} \alpha f(\alpha) \int_0^\tau \frac{1}{N} e^{s\alpha t}\, dt\, d\alpha + R(\tau) \\
&= \frac{1}{Ns} \int_0^{z\bar\alpha} f(\alpha)(e^{s\alpha\tau} - 1)\, d\alpha + R(\tau) \\
&= \frac{1}{Ns}\Big(\frac{\bar\alpha s\tau}{1-\bar\alpha s\tau} - R_1(\tau)\Big) + R(\tau)\,.
\end{aligned}
\tag{D.55}
$$

Here,

$$
R_1(\tau) = e^{-z}\Big(\frac{e^{z\bar\alpha s\tau}}{1-\bar\alpha s\tau} - 1\Big) \tag{D.56}
$$

and $R(\tau)$ is, to leading order in $1/a^2$, the error term arising from the approximation (D.52), i.e.,

$$
R(\tau) = -2\int_0^{z\bar\alpha} \alpha f(\alpha) \int_0^\tau \frac{1 - P_{\text{ext}}([t], e^{s\alpha})}{a(t, e^{s\alpha}, N)^2}\, dt\, d\alpha\,. \tag{D.57}
$$

To obtain bounds on $R_1(\tau)$ and $R(\tau)$, we assume $s\bar\alpha\tau \le \frac{1}{4}$ (this is not best possible; any upper bound $< \frac{1}{2}$ would be sufficient to arrive at (D.62)). For $R_1(\tau)$ we obtain by observing its convexity and bounding it by the linear function assuming 0 at $\tau = 0$ and $e^{-z}(\frac{4}{3}e^{z/4} - 1)$ at $\tau = 1/(4\bar\alpha s)$:

$$
\begin{aligned}
R_1(\tau) &\le 4\bar\alpha s\tau e^{-z}\Big(\frac{4}{3}e^{z/4} - 1\Big) \le \frac{16}{3}\bar\alpha s\tau e^{-3z/4} \\
&= \frac{16}{3}\bar\alpha s\tau (Ns\bar\alpha)^{-3\tilde K/4}\,,
\end{aligned}
\tag{D.58}
$$

which is of smaller order in magnitude than the leading order term $\frac{\bar\alpha s\tau}{1-\bar\alpha s\tau}$ in (D.55) (this holds for every $\tilde K \ge 1$).

Now we derive a bound for $R(\tau)$. From the monotonic decay in t of the probabilities of non-extinction $(1 - P_{\text{ext}}([t], e^{s\alpha}))$ and the approximation $1 - P_{\text{ext}}^{\infty}(e^{s\alpha}) = P_{\text{sur}}(e^{s\alpha}) \approx \rho s\alpha$ in

Remark 3.3, which requires that $s\alpha$ is sufficiently small, we infer

$$1 - P_{\text{ext}}([t], e^{s\alpha}) > 1 - P_{\text{ext}}^{\infty}(e^{s\alpha}) > cs\alpha \tag{D.59}$$

if $c < \rho$. We can choose, for instance, $c = \rho/2$. Henceforth, we assume that s is small enough such that (D.59) holds for all $\alpha \leq z\bar{\alpha}$. (This is possible because with $Ns^K = C^K$ we obtain $sz = s\tilde{K}\ln(Ns\bar{\alpha}) = CN^{-1/K}\tilde{K}\ln(N^{(K-1)/K}C\bar{\alpha}) \to 0$ as $N \to \infty$. Hence, $czs\bar{\alpha} \to 0$ as $N \to \infty$, which guaranties that (D.59) holds for every $\alpha \leq z\bar{\alpha}$ if N is large enough). Then, we obtain for every t and $\alpha \leq z\bar{\alpha}$:

$$\frac{2\big(1 - P_{\text{ext}}([t], e^{s\alpha})\big)}{a(t, e^{s\alpha}, N)^2} = \frac{2e^{2s\alpha t}}{N^2\big(1 - P_{\text{ext}}([t], e^{s\alpha})\big)} < \frac{2e^{2s\alpha t}}{N^2\big(1 - P_{\text{ext}}^{\infty}(e^{s\alpha})\big)} < \frac{2e^{2s\alpha t}}{N^2 cs\alpha}\,. \tag{D.60}$$

Therefore, the integral with respect to t in (D.57), is bounded by

$$\int_0^{\tau} \frac{2e^{2s\alpha t}}{N^2 cs\alpha} = \frac{1}{N^2 s^2 c}\frac{e^{2s\alpha\tau} - 1}{\alpha^2}\,. \tag{D.61}$$

Substituting this into (D.57) and integrating with respect to α we obtain

$$\begin{aligned} |R(\tau)| &\leq \frac{1}{N^2 s^2 c}\int_0^{z\bar{\alpha}} f(\alpha)\frac{e^{2s\alpha\tau} - 1}{\alpha}\,d\alpha \\ &= \frac{1}{N^2 s^2 c\bar{\alpha}}\Big(-\ln(1 - 2s\bar{\alpha}\tau) - R_2(\tau)\Big) \\ &< \frac{1}{N^2 s^2 c\bar{\alpha}}\Big(3s\bar{\alpha}\tau - R_2(\tau)\Big)\,, \end{aligned} \tag{D.62}$$

where $R_2(\tau) = E_1(z(1 - 2s\bar{\alpha}\tau)) - E_1(z)$ and the linear bound on the logarithmic term applies if $s\bar{\alpha}\tau < \frac{1}{4}$. Upon two-fold differentiation, we observe that $R_2(\tau)$ is convex and increases from 0 at $\tau = 0$ to $E_1(z/2)) - E_1(z)$ at $\tau = 1/(4s\bar{\alpha})$. Therefore, we bound $R_2(\tau)$ by the corresponding linear function in τ and obtain

$$R_2(\tau) \leq 4s\bar{\alpha}\tau\Big(E_1(z/2) - E_1(z)\Big) = s\tau\, O\Big(\frac{1}{(Ns)^{\tilde{K}/2}\ln(Ns)}\Big) \tag{D.63}$$

by series expansion for $Ns \to \infty$ and using $z = \tilde{K}\ln(Ns\bar{\alpha})$. In summary, we obtain from (D.62) and (D.63), and then choosing $\tilde{K} = 4$:

$$|R(\tau)| \leq \frac{3\tau}{N^2 sc}\Big(1 + O\Big(\frac{s^{2-\tilde{K}/2}}{N^{\tilde{K}/2}\ln(Ns)}\Big)\Big) = \tau\, O\Big(\frac{1}{N^2 s}\Big)\,. \tag{D.64}$$

We note that estimates similar to those in (D.60) show that

$$a(\tau, e^{sz\bar{\alpha}}, N) \geq \text{const}\,(Ns)^{3/4}\ln(Ns\bar{\alpha})\,, \tag{D.65}$$

where we have used $s\bar{\alpha}\tau \leq \frac{1}{4}$. This shows that the approximation (D.51) is indeed justified for every $\alpha \leq z\bar{\alpha}$.

By combining (D.55) and (D.54), and substituting the bounds (D.58) for $R_1(\tau)$ and (D.64)

for $R(\tau)$, we obtain

$$\begin{aligned}
\bar{G}(\tau) &= \Theta \int_0^\infty \alpha f(\alpha) \int_0^\tau (1 - P_{\text{ext}}([t], e^{s\alpha})) \gamma_1(a(t, e^{s\alpha}, N))\, dt\, d\alpha \\
&= \frac{\Theta}{Ns}\Big(\frac{\bar{\alpha} s\tau}{1-\bar{\alpha} s\tau} - R_1(\tau)\Big) + \Theta R(\tau) + \Theta\tau\, O\Big(\frac{\ln(Ns)}{(Ns)^{\tilde{K}}}\Big) \\
&= \frac{\Theta\tau}{Ns}\Big(\frac{\bar{\alpha} s}{1-\bar{\alpha} s\tau} + O\Big(\frac{s}{(Ns)^{3\tilde{K}/4}}\Big)\Big) + \Theta\tau\, O\Big(\frac{1}{N^2 s}\Big) + \Theta\tau\, O\Big(\frac{\ln(Ns)}{(Ns)^{\tilde{K}}}\Big) \\
&= \frac{\Theta\tau}{N}\Big[\frac{\bar{\alpha}}{1-\bar{\alpha} s\tau} + O\Big((Ns)^{-3\tilde{K}/4}\Big) + O\Big((Ns)^{-1}\Big) + O\Big(\frac{\ln(Ns)}{N^{\tilde{K}-1} s^{\tilde{K}}}\Big)\Big] \\
&= \frac{\Theta\tau}{N}\Big[\frac{\bar{\alpha}}{1-\bar{\alpha} s\tau} + O\Big(N^{-1+1/K}\Big)\Big], \qquad \text{(D.66)}
\end{aligned}$$

where in the last step we chose $\tilde{K} = 4$ so that the first error term is $O\Big((Ns)^{-3}\Big)$ and the last error term is

$$O\Big(\frac{\ln(Ns)}{Ns}\frac{1}{N^2 s^3}\Big) = O\Big(\frac{\ln(Ns)}{Ns} N^{-2+3/K}\Big) = o\Big(N^{-1+1/K}\Big) = o\Big((Ns)^{-1}\Big) \qquad \text{(D.67)}$$

by using once again the scaling assumption A1. This finishes the proof of (4.37). □

Proof of (4.38). From (4.1c) and (D.7) we get $\gamma(a) \sim 1/a - 4/a^2$ as $a \to \infty$. Therefore, the proof is analogous to that of (4.37), i.e., it is based on the idea of substituting this approximation into (4.15), which yields

$$\begin{aligned}
V_G(\tau) &= \Theta \int_0^\infty \alpha^2 f(\alpha) \int_0^\tau (1 - P_{\text{ext}}([t], e^{s\alpha})) \gamma(a(t, e^{s\alpha}))\, dt\, d\alpha \\
&\approx \frac{\Theta}{N} \int_0^\infty \alpha^2 f(\alpha) \int_0^\tau e^{s\alpha(t)}\, dt\, d\alpha \\
&= \frac{\Theta}{Ns} \int_0^\infty \alpha f(\alpha)(e^{s\alpha\tau} - 1)\, d\alpha\,, \qquad \text{(D.68)}
\end{aligned}$$

which yields (4.38) if f is exponential with mean $\bar{\alpha}$.

We omit several of the details of the proof and present only the most relevant expression. Instead of (D.54) we obtain

$$\begin{aligned}
0 &< \int_{z\bar{\alpha}}^\infty \alpha^2 f(\alpha) \int_0^\tau (1 - P_{\text{ext}}([t], e^{s\alpha})) \gamma_1(a(t, e^{s\alpha}, N))\, dt\, d\alpha \\
&< \tau\, O\Big(\frac{\ln(Ns)^2}{(Ns)^{\tilde{K}}}\Big)\,. \qquad \text{(D.69)}
\end{aligned}$$

Instead of (D.55) we obtain

$$\begin{aligned}
&\int_0^{z\bar{\alpha}} \alpha^2 f(\alpha) \int_0^\tau (1 - P_{\text{ext}}([t], e^{s\alpha})) \gamma_1(a(t, e^{s\alpha}, N))\, dt\, d\alpha \\
&\quad = \frac{\bar{\alpha}}{Ns}\Big(\frac{1}{(1-s\bar{\alpha}\tau)^2} - 1 - R_1^V(\tau)\Big) + R^V(\tau)\,, \qquad \text{(D.70)}
\end{aligned}$$

where $R_1^V(\tau)$ and $R^V(\tau)$ are defined analogously to $R_1(\tau)$ and $R(\tau)$ above. For $R_1^V(\tau)$ one obtains

$$\begin{aligned} R_1^V(\tau) &= \bar{\alpha} e^{-z} \Big(\frac{e^{sz\bar{\alpha}\tau}(1 + z(1 - s\bar{\alpha}\tau)}{(1 - s\bar{\alpha}\tau)^2} - (1 + z) \Big) \\ &\leq \tau\, O\Big(\frac{s \ln(Ns)}{(Ns)^{3\tilde{K}/4}} \Big) , \end{aligned} \tag{D.71}$$

where the expression in z is monotone increasing and convex in τ, and we bound it by the linear function assuming 0 at $\tau = 0$ and the corresponding value at $\tau = 1/(4s\bar{\alpha})$. Finally, we obtain after similar calculations and series expansions

$$\begin{aligned} |R^V(\tau)| &\leq \frac{2}{N^2 s^2 c} \int_0^{z\bar{\alpha}} f(\alpha)(e^{2s\alpha\tau} - 1)\, d\alpha \\ &= \frac{2}{N^2 s^2 c\bar{\alpha}} \Big(\frac{2s\bar{\alpha}\tau}{1 - 2s\bar{\alpha}\tau} + R_2(\tau) \Big) = \tau O\Big(\frac{1}{N^2 s} \Big) . \end{aligned} \tag{D.72}$$

Putting all this together, writing the leading order term in (D.70) as

$$\frac{1}{(1 - s\bar{\alpha}\tau)^2} - 1 = 2s\bar{\alpha}\tau\, \frac{1 - s\bar{\alpha}\tau/2}{(1 - s\bar{\alpha}\tau)^2} \tag{D.73}$$

and choosing $\tilde{K} = 4$, we obtain

$$\begin{aligned} V_G(\tau) &= \Theta \int_0^\infty \alpha^2 f(\alpha) \int_0^\tau (1 - P_{\text{ext}}([t], e^{s\alpha}))\gamma_1(a(t, e^{s\alpha}, N))\, dt\, d\alpha \\ &= \frac{2\Theta\bar{\alpha}^2\tau}{N} \Big[\frac{1 - s\bar{\alpha}\tau/2}{(1 - s\bar{\alpha}\tau)^2} + O\Big(N^{-1+1/K} \Big) \Big] . \end{aligned} \tag{D.74}$$

□

The approximations (4.39) and (4.40) are then obtained by a series expansion of the leading order terms.

Proof of (4.41). For equal mutation effects the proof is much simpler because we do not have to integrate with respect to α. Indeed, we observe directly from (D.52) that, to leading order in Ns, $\bar{G}(\tau) = \frac{\Theta}{Ns}(e^{s\alpha\tau} - 1)$. The estimate of the error term, analogous to $R(\tau)$, is also much simpler and straightforwardly yields (cf. (D.61))

$$|R(\tau)| \leq \int_0^\tau \frac{2e^{2s\alpha t}}{N^2 cs\alpha} = \frac{1}{N^2 s^2 c} \frac{e^{2s\alpha\tau} - 1}{\alpha^2} , \tag{D.75}$$

which is of smaller order, by a factor of $1/(Ns)$, than the leading order term. Of course, this requires $e^{2s\alpha\tau} = O(1)$, but it does not require the stronger constraint $s\alpha\tau < \frac{1}{4}$. The approximation for $V_G(\tau)$ is derived analogously. □

E. Segregating sites: A branching process approximation

Here we derive an alternative approximation for the number of segregating sites by relying on our branching process approach, similar as in Sect. 4.3, and amending it by a diffusion approximation for the mean fixation time. At a given generation τ, $\Theta\tau$ mutations have occurred on average. A

mutant with fitness σ is still present in the population with probability $1 - P_{\text{ext}}(\tau - t, \sigma)$, where t is the number of generations since its occurrence and $P_{\text{ext}}(n, \sigma)$ is the probability of loss until generation n; see (3.9). On average, $t = i/\Theta$ for the ith mutation event. Some of those mutations may have already reached fixation and are no longer segregating in the population. Fixation occurs on average after $\bar{t}_{\text{fix}}$ generations (see eqs. 4.8 and 4.9).

With this in mind and following the definitions the expectations of the phenotypic mean and variance in (4.12) and (4.13), respectively, we define random variables $S^{(i)}_{\tau,e^{s\alpha},\Theta}$, which are 1 when locus i is segregating in the population and 0 otherwise. Using the expected fixation time t_{fix} in (B.1) from the diffusion approximation, we set

$$\tilde{\tau} = \min\{t_{\text{fix}}(\tilde{s}(\alpha), N), \tau\}\,, \tag{E.1}$$

which is the expected time a mutant present in generation τ has already been segregating. Then

$$\mathrm{E}[S^{(i)}_{\tau,e^{s\alpha},\Theta}] \approx \int_0^{\tilde{\tau}} \frac{m_{i,\Theta}(t)}{M_{i,\Theta}(\tau)}(1 - P_{\text{ext}}([\tilde{\tau} - t], e^{s\alpha}))dt\,. \tag{E.2}$$

We define the expected number of segregating sites in our model, where mutation effects are drawn from a distribution f and the offspring distribution is Poisson, by

$$\mathrm{E}[S](\tau, f, s, N, \Theta) = \sum_{n=1}^{\infty} \mathrm{Poi}_{\Theta\tau}(n) \left(\sum_{i=1}^{n} \int_0^{\infty} \mathrm{E}\left[S^{(i)}_{\tau,e^{s\alpha},\Theta}\right] f(\alpha)\, d\alpha \right)\,, \tag{E.3}$$

By assuming equality in (E.2), we obtain the following approximation for $\mathrm{E}[S]$:

$$\begin{aligned}
\tilde{S}(\tau, f, s, N, \Theta) &= \sum_{n=1}^{\infty} \mathrm{Poi}_{\Theta\tau}(n) \left(\sum_{i=1}^{n} \int_0^{\infty} \int_0^{\tilde{\tau}} \frac{m_{i,\Theta}(t)}{M_{i,\Theta}(\tau)}(1 - P_{\text{ext}}([\tilde{\tau} - t], e^{s\alpha}))\, dt\, f(\alpha)\, d\alpha \right) \\
&= \int_0^{\infty} \int_0^{\tilde{\tau}} \sum_{n=1}^{\infty} \mathrm{Poi}_{\Theta\tau}(n) \left(\sum_{i=1}^{n} \frac{m_{i,\Theta}(t)}{M_{i,\Theta}(\tau)} \right) (1 - P_{\text{ext}}([\tilde{\tau} - t], e^{s\alpha}))\, dt\, f(\alpha)\, d\alpha \\
&= \Theta \int_0^{\infty} f(\alpha) \int_0^{\tilde{\tau}} (1 - P_{\text{ext}}([\tilde{\tau} - t], e^{s\alpha}))\, dt\, d\alpha \\
&\approx \Theta \int_0^{\infty} f(\alpha) \sum_{i=0}^{[\tilde{\tau}]} (1 - P_{\text{ext}}([\tilde{\tau}] - i, e^{s\alpha}))\, d\alpha \\
&= \Theta \int_0^{\infty} f(\alpha) \sum_{j=0}^{[\tilde{\tau}]} (1 - P_{\text{ext}}(j, e^{s\alpha}))d\alpha\,.
\end{aligned} \tag{E.4}$$

The above calculations are analogous to those in the derivation of (4.14) for $\bar{G}(\tau)$ in (D.14) upon substitution of $\alpha\,\mathrm{E}[\tilde{X}^{(i)}_{\tau,e^{s\alpha},\Theta}]$ by $\mathrm{E}[S^{(i)}_{\tau,e^{s\alpha},\Theta}]$.

If all effects are equal to α, then $\tilde{\tau} = \tilde{\tau}(\alpha)$ is constant and the integration in (E.4) collapses to the computation of the sum, i.e., to (5.5) in the main text. Unless the distribution $f(\alpha)$ is a delta function, as for equal mutation effects, no further simplification of (E.4) seems possible, except when $\tau \ll \bar{t}_{\text{fix}}(\tilde{s}(\alpha), N)$. Then, for given small τ, we have $\tilde{\tau} = \tau$ except for the (very rare) very large values of α in the tail of f for which $t_{\text{fix}}(\tilde{s}(\alpha), N) < \tau$. As a consequence, we obtain from (E.4) the approximation (5.6) in the main text.

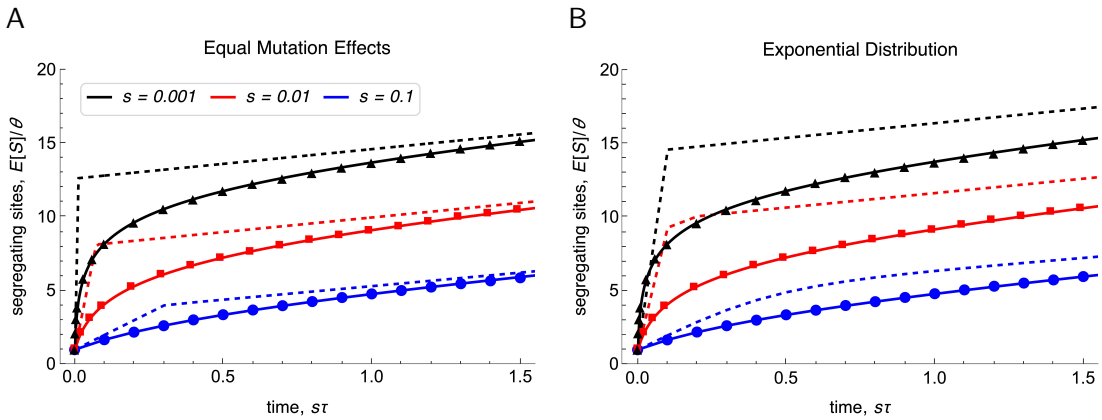

Figure E.1: Scaled expected number of segregating sites, $E[S]/\Theta$ as a function of $s\tau$ in the early phase of adaptation ($s\tau \leq 2$) for the selection intensities $s = 0.1, 0.01, 0.001$ (blue, red, black). In A the mutation effects are equal ($\alpha = 1$), in B they are drawn from an exponential distribution with mean $\bar{\alpha} = 1$. The branching process approximations $\tilde{S}$ in (5.5) and (5.6) are shown by solid curves, the diffusion approximation $\bar{S}$ in (5.1) by dashed curves. The kinks occur at st_{loss}. Results from Wright-Fisher simulations (with $\Theta = \frac{1}{2}$) are shown as symbols. The mutation effects are $\alpha = 1$ and the population size is $N = 10^4$.

S. Supplementary figures

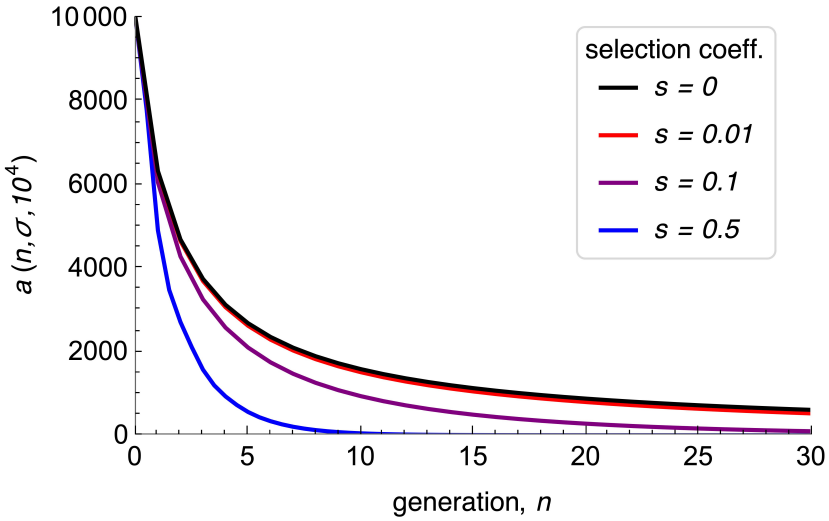

Figure S.1: Shape of $a(n, \sigma, N)$ in eq. (3.15b) as a function of the number of generations n for different selection coefficients s. The population size is $N = 10^4$ and the offspring distribution is Poisson with mean $\sigma = e^s$.

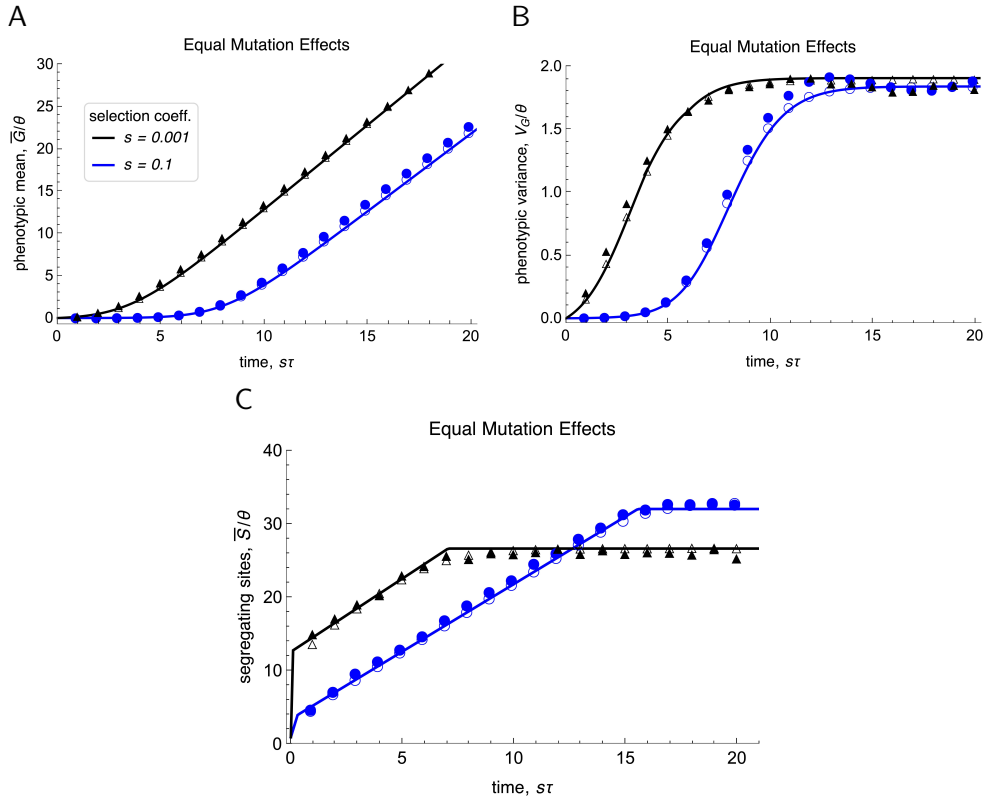

Figure S.2: The scaled expected phenotypic mean $\bar{G}(s\tau)/\Theta$ (panel A), the scaled expected phenotypic variance $V_G(s\tau)/\Theta$ (panel B), and the scaled expected number of segregating sites $\bar{S}(s\tau)/\Theta$ (panel C) are shown as functions of $s\tau$ for the selection coefficients $s = 0.1$ and $s = 0.001$ (blue and black, respectively). Equal mutation effects are assumed, i.e., $\alpha_i = 1$. Analytic approximations, with $\bar{G}(s\tau)$ computed from (4.16), $V_G(s\tau)$ from (4.17), and $\bar{S}(s\tau)$ from (5.1), are shown as solid curves. Wright-Fisher-simulation results with $2\Theta = 10$ and $2\Theta = 0.01$ (open and filled symbols, respectively) are shown as symbols. Whereas Θ is a multiplicative factor in the analytic approximations, these simulation results show fluctuations around the predicted curve if $2\Theta = 0.01$. Presumably, the reason is that with such small Θ, the waiting time for successful mutations is very long, about $1/(2s\Theta)$. On the time scale of the figure, this yields $1/(2\Theta) = 100$ units. Such rare successful mutations will entail substantial stochastic effects. We note that very slight fluctuations are also visible in Figures 4.2 and 5.1, where $2\Theta = 1$.

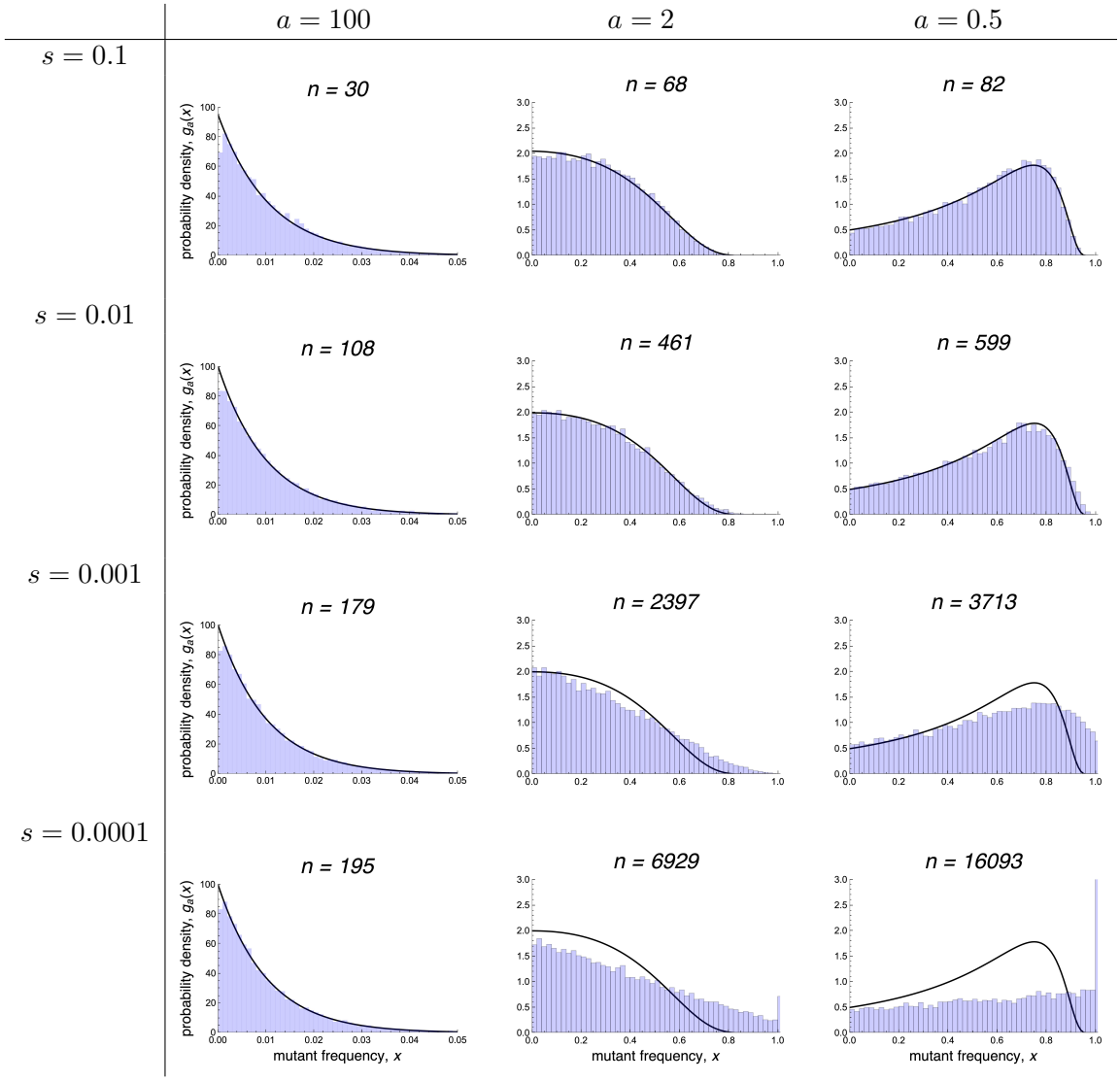

Figure S.3: The black curves show the probability density $g_a(x)$ in (3.15) of the mutant frequency in different generations n for various combinations of $a = a(n, \sigma, N)$ and $\sigma = e^s$ (conditioned on non-extinction until generation n). The number of discrete generations n (indicated above each graph) is chosen such that a (see eq. (3.15)) is closest to 100, 2 and 0.5, respectively. The population size is $N = 10^4$. The blue histograms are obtained from simulations of the corresponding Wright-Fisher model. The analytic approximation $g_a(x)$ underestimates the effect of random genetic drift near fixation (x large and $n > \frac{1}{2} t_{\text{fix}} \approx \frac{1}{s} \ln(2Ns)$), unless selection is very strong relative to random drift (top row, where $Ns = 10^3$). For $s = 0.1$, 0.01, 0.001, and 0.0001, we have $t_{\text{fix}} \approx 156$, 1060, 5991, and $13\,863$, respectively. In general, and according to intuition and assumptions, the accuracy of the approximation by $g_a(x)$ decreases with larger x and n, and smaller Ns. Nevertheless, for very early stopping times n relative to t_{fix}, the approximation remains useful even if $Ns = 1$ (left bottom). Its mean, given by (4.1a), stays accurate in a much wide parameter range than $g_a(x)$ (Fig. 4.1.A).

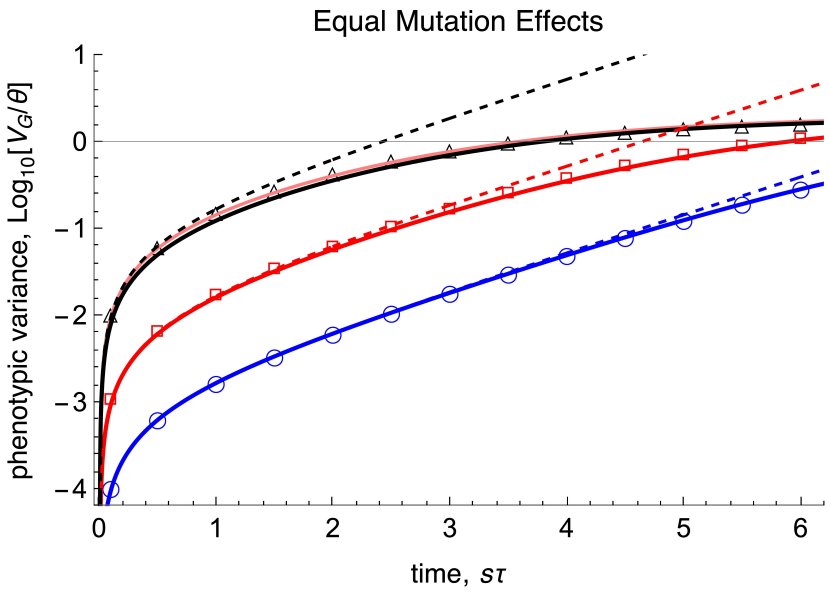

Figure S.4: This is analogous to panel B of Figure 4.4, the scaled phenotypic variance $\log_{10}(V_G(s\tau)/\Theta)$ in the initial phase for equal mutation effects $\alpha = 1$ and selection coefficients $s = 0.001, 0.01, 0.1$ (represented by black, red and blue color, respectively) is shown. The solid curves show the approximation (4.36). The dashed curves show the simplified approximation in (4.41). For $s = 0.001$, the pink solid curve shows the most accurate approximation (4.15). For all other parameter combinations in Figure 4.4 (also for the phenotypic mean) no deviation from the solid curves to the approximations from Proposition 4.3 is visible. Symbols present results from Wright-Fisher simulations with $\Theta = \frac{1}{2}$, the population size is $N = 10^4$, and the offspring distribution is Poisson.

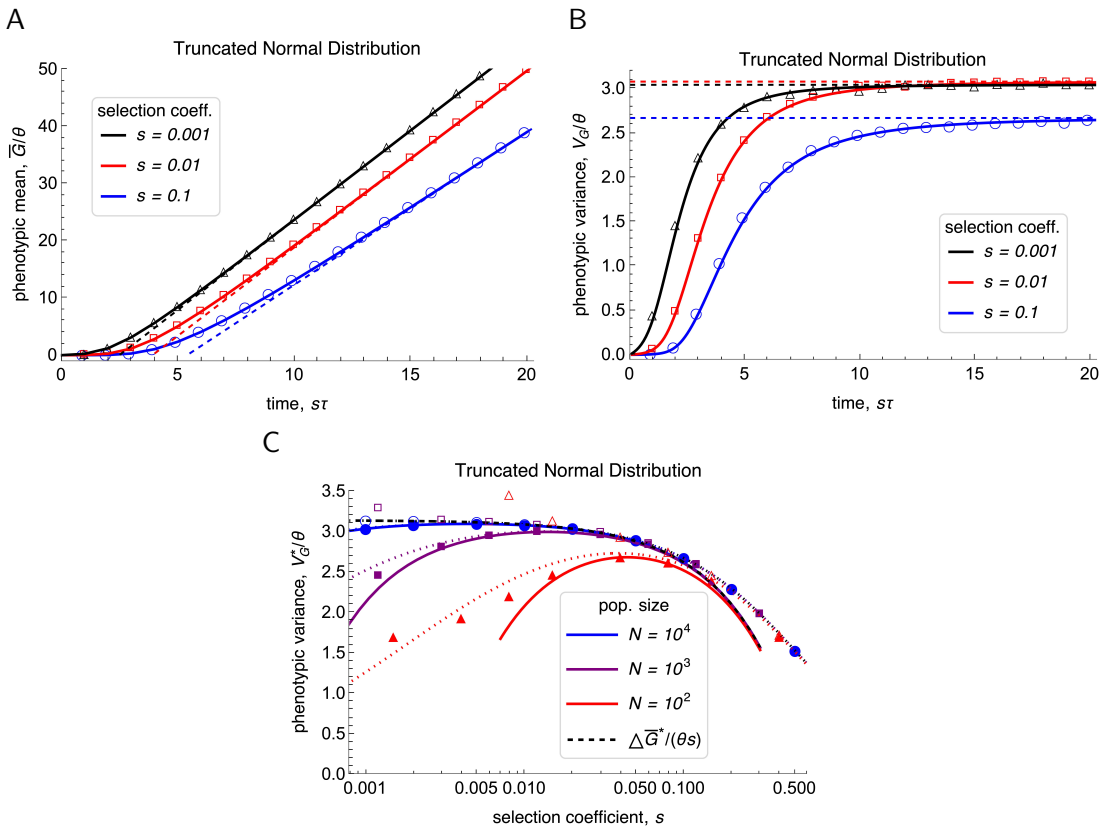

Figure S.5: The scaled expected phenotypic mean and variance, $\bar{G}(s\tau)/\Theta$ and $V_G(s\tau)/\Theta$ as function of scaled time (panels A and B), and the scaled expected phenotypic variance at stationarity, $V_G^*(s\tau)/\Theta$ (panel C) as a function of s for a truncated normal distribution with $\alpha \geq 0$ and $\bar{\alpha} = 1$ (cf. Remark 4.10). Panels A and B are analogous to Fig. 4.2 and panel C is analogous to Fig. 4.3. Analytic approximations are shown as curves, Wright-Fisher simulation results (for $\Theta = \frac{1}{2}$) as symbols.

In panels A and B, we have $N = 10^4$ and selection coefficients and color code are shown in the legends. The solid curves show $\bar{G}(s\tau)/\Theta$ and $V_G(s\tau)/\Theta$ as computed from (4.14) and (4.15). The dashed lines show the equilibrium values given by (4.21) and (4.22). The values of the mean fixation times multiplied by $s = 0.1, 0.01, 0.001$ are $s\bar{t}_{\text{fix}} \approx 16.1, 11.0, 6.2$, respectively (computed from (4.8)).

In panel C, different colors correspond to different population sizes, as shown in the legend. The most accurate, available approximation is (4.22); it is shown as dotted curves. In addition, the scaled variance V_G^*/Θ is compared with the scaled expected change in the phenotypic mean $\Delta\bar{G}^*/(\Theta s)$, as calculated from (4.31) (black dashed curve). For the Wright-Fisher simulation results (circles for $N = 10^4$, squares for $N = 10^3$, triangles for $N = 10^2$), filled symbols are used for the scaled phenotypic variance and open symbols for the change in the scaled phenotypic mean.

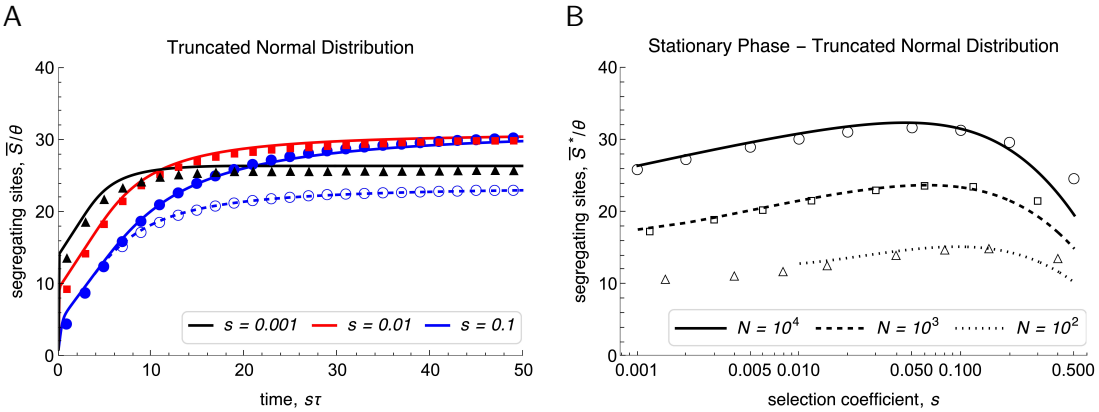

Figure S.6: The expected number of segregating sites (rescaled by Θ) as a function of $s\tau$ (panel A) and at stationarity as a function of s (panel B). This figure is analogous to the rows in Fig. 5.1 but mutation effects are drawn from a normal distribution truncated at 0 with mean $\bar{\alpha} = 1$ (cf. Remark 4.10). Analytic approximations are computed from (5.1) and are shown as curves. Wright-Fisher simulation results for $\mathrm{E}[S]$ (with $\Theta = \frac{1}{2}$) are shown as symbols. In A, $\bar{S}/\Theta$ is shown for $N = 10^4$ and $s = 0.1, 0.01, 0.001$ (solid curves and filled symbols in blue, red, black), and for $N = 10^3$ and $s = 0.1$ (dashed curve and open symbols in blue). In B, $\bar{S}^*/\Theta$ is displayed for $N = 10^4, 10^3, 10^2$ (solid, dashed, dotted curves, and corresponding open symbols).

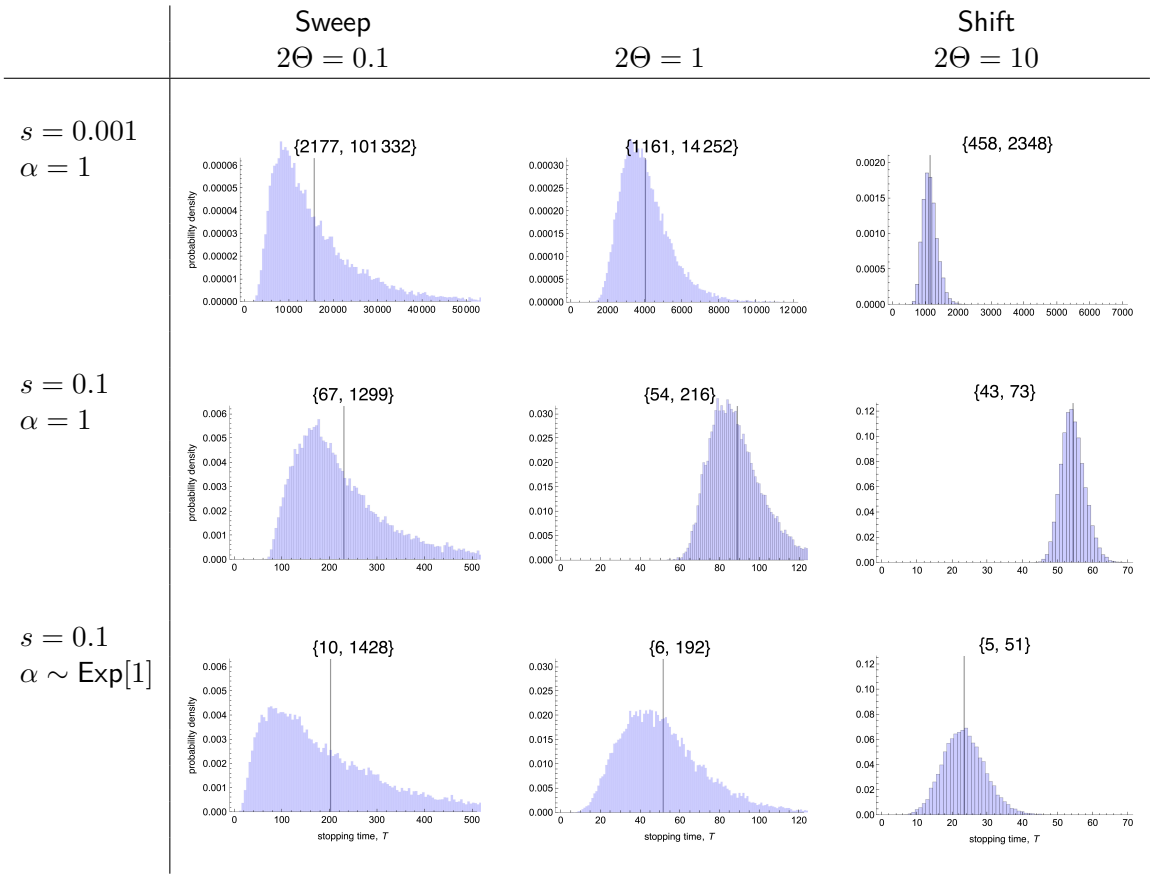

Figure S.7: Histograms of the stopping times T at which the mean phenotype $\bar{G}$ in a single population first reaches or exceeds the value 1 (this figure complements Fig. 5.2). Histograms are based on $20\,000$ replicate runs of the Wright-Fisher model. The vertical black lines represent the corresponding means $\bar{T}$, which are given in Fig. 5.2. The corresponding minimum and maximum values of T are given in curly braces. The selection coefficients s and the mutation effects α are indicated for each row. Columns show results for different values of Θ. The population size is $N = 10^4$ in all cases.

References

Abramowitz, M., Stegun, I.A., 1964. Handbook of mathematical functions with formulas, graphs, and mathematical tables. volume 55. US Government Printing Office.

Allen, L.J.S., 2003. An introduction to stochastic processes with applications to biology. Pearson/Prentice Hall, Upper Saddle River, N.J.

Athreya, K.B., 1992. Rates of decay for the survival probability of a mutant gene. Journal of mathematical biology 30, 577–581.

Athreya, K.B., Ney, P.E., 1972. Branching processes. Springer, Berlin-Heidelberg.

Barghi, N., Hermisson, J., Schlötterer, C., 2020. Polygenic adaptation: A unifying framework to understand positive selection. Nature Reviews Genetics 21, 769—-781.

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

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

Barton, N.H., Etheridge, A.M., Véber, A., 2017. The infinitesimal model: Definition, derivation, and implications. Theoretical Population Biology 118, 50–73.

Barton, N.H., Keightley, P.D., 2002. Understanding quantitative genetic variation. Nature Reviews Genetics 3, 11–21.

Berg, J.J., Coop, G., 2014. A population genetic signal of polygenic adaptation. PLoS Genetics 10, e1004412–e1004412.

Boenkost, F., González Casanova, A., Pokalyuk, C., Wakolbinger, A., 2021. Haldane's formula in cannings models: the case of moderately strong selection. Journal of Mathematical Biology 83, 70.

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

Bulmer, M.G., 1980. The mathematical theory of quantitative genetics. Clarendon Press.

Bürger, R., 1993. Predictions of the dynamics of a polygenic character under directional selection. Journal of Theoretical Biology 162, 487–513.

Bürger, R., 2000. The mathematical theory of selection, recombination, and mutation. Wiley, Chichester.

Bürger, R., Ewens, W.J., 1995. Fixation probabilities of additive alleles in diploid populations. Journal of Mathematical Biology 33, 557–575.

Bürger, R., Lynch, M., 1995. Evolution and extinction in a changing environment: A quantitative-genetic analysis. Evolution 49, 151–163.

Charlesworth, B., 2020. How long does it take to fix a favorable mutation, and why should we care? The American Naturalist 195, 753–771.

Chevin, L.M., 2019. Selective sweep at a QTL in a randomly fluctuating environment. Genetics 213, 987–1005.

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

Clayton, G., Robertson, A., 1955. Mutation and quantitative variation. The American Naturalist 89, 151–158.

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

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

Devi, A., Jain, K., 2023. Polygenic adaptation dynamics in large, finite populations. bioRxiv .

Durrett, R., Schweinsberg, J., 2004. Approximating selective sweeps. Theoretical Population Biology 66, 129–138.

Etheridge, A., Pfaffelhuber, P., Wakolbinger, A., 2006. An approximate sampling formula under genetic hitchhiking. The Annals of Applied Probability 16, 685—729.

Ewens, W.J., 1974. A note on the sampling theory for infinite alleles and infinite sites models. Theoretical Population Biology 6, 143–148.

Ewens, W.J., 2004. Mathematical population genetics: Theoretical introduction. Springer, New York.

Eyre-Walker, A., Keightley, P.D., 2007. The distribution of fitness effects of new mutations. Nature Reviews Genetics 8, 610–618.

Falconer, D.S., Mackay, T.F.C., 1996. Introduction to quantitative genetics. Longman Group, UK.

Fisher, R.A., 1918. The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh 52, 399–433.

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., 2009. GNU scientific library reference manual. 3 ed.

Gomulkiewicz, R., Holt, R.D., 1995. When does evolution by natural selection prevent extinction? Evolution , 201–207.

Haccou, P., Jagers, P., Vatutin, V.A., 2005. Branching processes: Variation, growth, and extinction of populations. volume 5 of *Cambridge studies in adaptive dynamics*. Cambridge University Press.

Haldane, J.B.S., 1927. A mathematical theory of natural and artificial selection, part V: Selection and mutation. Mathematical Proceedings of the Cambridge Philosophical Society 23, 838—-844.

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

Hermisson, J., Pennings, P.S., 2005. Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. Genetics 169, 2335–2352.

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

Hernandez, R.D., Kelley, J.L., Elyashiv, E., Melton, S.C., Auton, A., McVean, G., Project, .G., Sella, G., Przeworski, M., 2011. Classic selective sweeps were rare in recent human evolution. Science 331, 920–924.

Hill, W.G., 1982a. Rates of change in quantitative traits from fixation of new mutations. Proceedings of the National Academy of Sciences - PNAS 79, 142–145.

Hill, W.G., 1982b. Predictions of response to artificial selection from new mutations. Genetics Research 40, 255–278.

Hill, W.G., Kirkpatrick, M., 2010. What animal breeding has taught us about evolution. Annual Review of Ecology, Evolution, and Systematics 41, 1–19.

Höllinger, I., 2018. Mathematical models of speciation and polygenic adaptation. Ph.D. thesis. University of Vienna.

Höllinger, I., Pennings, P.S., Hermisson, J., 2019. Polygenic adaptation: From sweeps to subtle frequency shifts. PLoS Genetics 15, 1–26.

Höllinger, I., Wölfl, B., Hermisson, J., 2023. A theory of oligogenic adaptation of a quantitative trait. Genetics , iyad139.

Hospital, F., Chevalet, C., 1996. Interactions of selection, linkage and drift in the dynamics of polygenic characters. Genetics Research 67, 77–87.

Huber, C.D., Degiorgio, M., Hellmann, I., Nielsen, R., 2016. Detecting recent selective sweeps while controlling for mutation rate and background selection. Molecular Ecology 25, 142–156.

Jain, K., Kaushik, S., 2022. Joint effect of changing selection and demography on the site frequency spectrum. Theoretical Population Biology 146, 46–60.

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

Jain, K., Stephan, W., 2017b. Modes of rapid polygenic adaptation. Molecular Biology and Evolution 34, 3169–3175.

John, S., Stephan, W., 2020. Important role of genetic drift in rapid polygenic adaptation. Ecology and Evolution 10, 1278–1287.

Jones, A.G., Arnold, S.J., Bürger, R., 2004. Evolution and stability of the g-matrix on a landscape with a moving optimum. Evolution 58, 1639–1654.

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

Keightley, P.D., Hill, W.G., 1983. Effects of linkage on response to directional selection from new mutations. Genetics Research 42, 193–206.

Keightley, P.D., Hill, W.G., 1987. Directional selection and variation in finite populations. Genetics 117, 573–582.

Kesten, H., Stigum, B.P., 1966. A limit theorem for multidimensional galton-watson processes. The Annals of Mathematical Statistics 37, 1211–1223.

Kim, Y., Stephan, W., 2003. Selective sweeps in the presence of interference among partially linked loci. Genetics 164, 389–398.

Kimura, M., Ohta, T., 1969. The average number of generations until fixation of a mutant gene in a finite population. Genetics 61, 763–771.

Lambert, A., 2006. Probability of fixation under weak selection: A branching process unifying approach. Theoretical Population Biology 69, 419–441.

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.

Lynch, M., Hill, W.G., 1986. Phenotypic evolution by neutral mutation. Evolution 40, 915–935.

Lynch, M., Lande, R., 1993. Evolution and extinction in response to environmental change, in: Kareiva, P.M., Kingsolver, J.G., Huey, R.B. (Eds.), Biotic Interactions and Global Change, Sinauer Associates, Sunderland MA. pp. 234–250.

Lynch, M., Walsh, B., 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland MA.

Martin, G., Lambert, A., 2015. A simple, semi-deterministic approximation to the distribution of selective sweeps in large populations. Theoretical Population Biology 101, 40–46.

Matsumoto, M., Nishimura, T., 1998. Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. ACM Transactions on Modeling and Computer Simulation 8, 3–30.

Matuszewski, S., Hermisson, J., Kopp, M., 2015. Catch me if you can: Adaptation from standing genetic variation to a moving phenotypic optimum. Genetics 200, 1255–1274.

Maynard Smith, J., Haigh, J., 1974. The hitch-hiking effect of a favourable gene. Genetics Research 23, 23–35.

Messer, P.W., Petrov, D.A., 2013. Population genomics of rapid adaptation by soft selective sweeps. Trends in Ecology & Evolution 28, 659–669.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G., Bustamante, C., 2005. Genomic scans for selective sweeps using snp data. Genome Research 15, 1566–1575.

Otto, S.P., Barton, N.H., 1997. The evolution of recombination: Removing the limits to natural selection. Genetics 147, 879–906.

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

Robertson, A., 1977. Artificial selection with a large number of linked loci, in: Pollak, E.J., Kempthorne, O., Bailey, T.B., et al. (Eds.), Proceedings of the 1st International Conference on Quantitative Genetics, Iowa State University Press. pp. 307–322.

Sella, G., Barton, N.H., 2019. Thinking about the evolution of complex traits in the era of genome-wide association studies. Annual Review of Genomics and Human Genetics 20, 461–493.

Setter, D., Mousset, S., Cheng, X., Nielsen, R., DeGiorgio, M., Hermisson, J., 2020. Volcanofinder: Genomic scans for adaptive introgression. PLoS Genetics 16, e1008867.

Steinrücken, M., Wang, Y.R., Song, Y.S., 2013. An explicit transition density expansion for a multi-allelic wright–fisher diffusion with general diploid selection. Theoretical Population Biology 83, 1–14.

Stephan, W., 2019. Selective sweeps. Genetics 211, 5–13.

Stephan, W., John, S., 2020. Polygenic adaptation in a population of finite size. Entropy 22, 907.

Tataru, P., Mollion, M., Glémin, S., Bataillon, T., 2017. Inference of distribution of fitness effects and proportion of adaptive substitutions from polymorphism data. Genetics 207, 1103–1119.

Thornton, K.R., 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.

Turelli, M., Barton, N.H., 1990. Dynamics of polygenic characters under selection. Theoretical Population Biology 38, 1–57.

Turelli, M., Barton, N.H., 1994. Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? Genetics 138, 913–941.

Uecker, H., Hermisson, J., 2011. On the fixation process of a beneficial mutation in a variable environment. Genetics 188, 915–930.

Walsh, B., Lynch, M., 2018. Evolution and selection of quantitative traits. Oxford University Press, Oxford.

Wölfl, B., 2019. Footprints of polygenic adaptation of a quantitative trait under stabilizing selection. Master's thesis. Vrije Universiteit Amsterdam and University of Amsterdam.

Wolfram Research, I., 2022. Mathematica, Version 13.2.

Zhang, X.S., Hill, W.G., 2005a. Genetic variability under mutation selection balance. Trends in Ecology & Evolution 20, 468–470.

Zhang, X.S., Hill, W.G., 2005b. Predictions of patterns of response to artificial selection in lines derived from natural populations. Genetics 169, 411–425.