Fate and Fortuity: Branching Process Models for the Establishment and Fixation of Beneficial Alleles

---

Dissertation

zur Erlangung des akademischen Grades
"Doctor rerum naturalium"
an der Universität Wien

---

verfaßt von
Hildegard Uecker
aus Würzburg

Wien 2013	
Studienkennzahl lt. Studienblatt:	A 091 405
Dissertationsgebiet lt. Studienblatt:	Mathematik
Betreuer:	Prof. Joachim Hermisson
Eingereicht am:	08.07.2013

Contents

Abstract

The establishment of beneficial alleles is fundamental for the genetic adaptation of populations to environmental change. Often, a favorable allele is initially only present in few copies in a population. For adaptive evolution to proceed, it has to escape stochastic loss and finally rise to fixation. The mathematical theory of branching processes provides an elegant framework to estimate the establishment probability and the speed of early growth of a beneficial allele in a large population. In this thesis, I apply this approach to study adaptation in ecologically or genetically complex scenarios. A special emphasis is put on situations where the fitness of the allele is time dependent. The analytical results are complemented by computer simulations.

The first chapter provides concise analytical approximations for the establishment probability of a single beneficial allele in a variable environment and the time that it needs to reach a given frequency. It turns out that even slight changes in the selection pressure over time can strongly alter both quantities. In the second chapter, the results are applied to the biological problem of evolutionary rescue in structured populations. I show how various, partially antagonistic effects intertwine, leading to a complex dependence of the rescue probability on ecological characteristics. The third chapter addresses the topic of adaptive gene introgression. Linked and unlinked deleterious alleles strongly affect the introgression process of a beneficial allele. Approximations for the introgression probability and the hitchhiking probability of closely linked deleterious alleles are derived.

Zusammenfassung

Die Etablierung vorteilhafter Allele ist elementar für die genetische Anpassung einer Population an veränderte Umweltbedingungen. Ein günstiges Allel liegt anfangs oft nur in geringer Zahl in einer Population vor. Damit Adaptation stattfinden kann, darf es nicht durch stochastische Fluktuationen verlorengehen und muß schließlich fixieren. Die mathematische Theorie der Verzweigungsprozesse bildet einen eleganten Rahmen, um die Etablierungswahrscheinlichkeit und die Geschwindigkeit des anfänglichen Frequenzanstiegs des vorteilhaften Allels in einer großen Population näherungsweise zu bestimmen. In dieser Dissertation wende ich diesen Ansatz an, um Adaptation in ökologisch oder genetisch komplexen Szenarien zu untersuchen. Ein besonderer Schwerpunkt liegt auf Situationen, in denen die Fitness des Alles zeitabhängig ist. Die analytischen Ergebnisse werden durch Computersimulationen ergänzt.

Das erste Kapitel liefert analytische Approximationen für die Etablierungswahrscheinlichkeit eines einzelnen vorteilhaften Allels in einer veränderlichen Umwelt und die Zeit, die es benötigt, um eine gegebene Frequenz zu erreichen. Es zeigt sich, daß selbst geringfügige zeitliche Änderungen im Selektionsdruck beide Größen beträchtlich beeinflussen können. Im zweiten Kapitel werden die Ergebnisse auf die Frage angewandt, ob eine vom Aussterben bedrohte Population ihren Fortbestand durch genetische Anpassung an die geänderte Umwelt sichern kann ("evolutionary rescue"). Ich zeige, wie verschiedene, teils antagonistische Effekte ineinandergreifen und dadurch zu einer komplexen Abhängigkeit der Überlebenswahrscheinlichkeit der Population von ökologischen Faktoren führen. Das dritte Kapitel befaßt sich mit adaptiver Introgression von Genen (d.h. der Etablierung vorteilhafter Allele einer Art im Genom einer anderen in der Folge von Hybridisierung). Gelinkte und ungelinkte schädliche Allele haben einen starken Einfluß auf den Fixationsprozeß des positiv selektierten Allels. Näherungen für die Etablierungswahrscheinlichkeit des vorteilhaften Allels in der fremden Population und die Wahrscheinlichkeit, daß schädliche Allele mitfixieren, werden hergeleitet.

Introduction

Adaptive evolution by natural selection has brought about remarkable morphological, physiological, and behavioral traits that allow organisms to cope with their biotic and abiotic environment. Of the countless examples, think of the excellent camouflage of stick or leaf insects: their appearance resembles a twig or a leaf in form, color, and structure so much that they are almost impossible to spot among the branches and foliage around them (see Figure 1). The defining feature of an adaptation is its positive impact in a given environment on fitness, i.e., on the expected reproductive success of an organism.

Often, we can only observe the outcome but not the process of adaptive evolution because the evolutionary change took place in the past. However, evolutionary change is not necessarily slow, and it is sometimes possible to observe adaptive evolution in action: examples of immediate relevance are the evolution of drug or insecticide resistance. The analysis of genomic data moreover allows us to identify loci that have probably been under recent positive selection. Unambiguous identification of these loci is, however, difficult since different processes can lead to similar patterns in nucleotide diversity, making it hard to infer the true history of the population. Even if with high probability other hypotheses (e.g., demographic processes) can be ruled out, the function of the gene is not necessarily known so that we cannot deduce which trait has evolved as a response to selection. Laboratory experiments allow us to follow the evolutionary process of adaptation in real time. In that way, we can make quantitative measurements about the speed and mode of adaptation and investigate the genetic basis of adaptive change (for an insightful review on experimental evolution with microorganisms see Elena and Lenski, 2003).

Adaptive evolution is often triggered by environmental change. If the environment severely deteriorates, adaptive evolution might be required for persistence of the population that, without evolution, is doomed to extinction. However, an immediately precedent change is no prerequisite for adaptive evolution. Adaptation also occurs in periods of environmental stasis as fitter variants originate and become prevalent. In a long-term experiment with *E. coli*, Richard Lenski

(a) Macleay's Spectre *(Extatosoma tiaratum)* (b) Walking Leaf Insect *(Phyllium giganteum)*

Figure 1: Leaf and stick insects possess excellent camouflage.
Photo courtesy of Jeff Whitlock, www.theonlinezoo.com

and his lab observed that adaptation still takes place after tens of thousands of generations (Elena and Lenski, 2003).

Adaptation occurs through the establishment and fixation of beneficial alleles. For a long time, the prevailing belief was that many genes of small effect contribute to adaptive change (the neo-Darwinian view). We now know that the genetic architecture of adaptations can take many forms and, frequently, few genes of large effect are responsible for most of the fitness increase (Orr and Coyne, 1992; Elena *et al.*, 1996; Elena and Lenski, 2003; Orr, 2005, and references therein). For some adaptations, mutation at a single locus is sufficient. Such a simple genetic basis has been found repeatedly in the evolution of insecticide or drug resistance (Milani, 1963; McKenzie *et al.*, 1980; Roush and McKenzie, 1987; Daborn *et al.*, 2002; Gerstein *et al.*, 2012). The serial fixation of mutations of large effect leads to a step-like dynamics of the adaptive process as observed by Elena *et al.* (1996).

Mutation generates the genetic variation for selection to act on. We can classify mutations into two categories: mutations that already segregate in the population prior to environmental change ("standing genetic variation") and mutations that arise later ("de-novo" or "new mutations"). In many cases, alleles involve a trade-off: while they are beneficial in one environment, they are deleterious in another. For example, alleles that render bacteria resistant to antibiotics decrease the competitive ability so that their carriers suffer a disadvantage with respect to sensitive strains in the absence of drugs. For this reason, alle-

les that prove beneficial after an environmental change often only segregate at low frequencies in mutation-selection balance prior to the environmental switch. Initially, favorable alleles are hence normally rare in the population. This has a very important consequence: even if beneficial, they are likely to suffer stochastic loss because the few individuals in this early phase might fail to reproduce by chance. This is because a beneficial allele increases the expected number of an individual's offspring but the variance in offspring number is non-zero. The role of chance in the adaptive process is hence substantial. Only if an adaptive allele survives the chance fluctuations when rare, is it picked up by selection to rise in frequency.

In many mathematical models of the adaptive process, the establishment probability of beneficial alleles plays a key role. Two classic approaches exist for the assessment of fixation probabilities: branching process and diffusion theory. Both methods were first brought up in a seminal paper by Fisher (1922). A few years later, in 1927, Haldane estimated by means of branching process theory that an allele with selective advantage σ has a chance of about 2σ to establish in the population if it appears in a single copy and offspring numbers are Poisson distributed (Haldane, 1927). Since then, both approaches have been developed further and applied to many biological scenarios (for a review see Patwa and Wahl (2008); cf. also Wahl (2011)). Assessing the establishment probability of beneficial alleles by means of branching process theory is at the core of this thesis.

Branching processes are a special class of Markov processes that consider the amplification of individuals or particles; in classical theory as applied in this thesis, the state space is discrete, consisting of the natural numbers (including zero). The distinctive assumption underlying branching processes is that individuals reproduce and die independently of each other. This special structure of branching processes allows for the derivation of many beautiful mathematical results. In the simplest case, there is only one type of individual and the process is furthermore homogeneous in time. Here we briefly focus on time-homogeneous single-type branching processes to review some key notions. For a rigorous and comprehensive introduction to the theory, we refer to the textbooks by Athreya and Ney (1972), Jagers (1975), Harris (1963), and Sewastjanow (1974). Let $P_{ij}(t)$ be the probability that i individuals turn into j individuals in the time

interval t. The branching property is then expressed by the following relation (Sewastjanow, 1974, p. 1):

$$P_{ij}(t) = \sum_{j_1+j_2+\cdots+j_i=j} P_{1j_1}(t)P_{1j_2}(t)\cdots P_{1j_i}(t),$$

where the right-hand side sums over all possibilities for i individuals having j descendants. Time can either be discrete ($t = 0, 1, 2, \dots$) or continuous ($t \in [0, \infty)$). In a process with discrete generations, each individual leaves a random number of offspring before it dies. Such processes are called Galton-Watson branching processes. If time is continuous, one usually assumes

$$\begin{aligned} P_{11}(h) &= 1 + p_1 h + o(h), \\ P_{1k}(h) &= p_k h + o(h), \quad k \neq 1 \end{aligned}$$

with $p_k \geq 0$ for $k \neq 1$ and

$$\sum_k p_k = 0.$$

If $p_k = 0$ for $k > 2$, we encounter a special instance of a birth-death process in which the per-capita birth and death rates are independent of the number of individuals.

Branching processes display a dichotomy: either the process dies out or it grows to infinity (except for the degenerate case in which each individual has exactly one offspring; we will ignore this case in the following). For a discrete-time process, denote by m the average number of offspring of an individual. For a continuous-time process, denote by $ah + o(h)$ the expected per-capita change in the total number of individuals in the time interval h. Depending on m (discrete time) or a (continuous time), branching processes are classified as subcritical ($m < 1$, $a < 0$), critical ($m = 1$, $a = 0$), or supercritical ($m > 1$, $a > 0$). Only supercritical processes have a chance to survive indefinitely.

Generating functions are an efficient tool for the analysis of branching processes. In a Galton-Watson branching process, denote by P_k the probability that an individual produces k offspring. The offspring distribution is characterized by the probability generating function

$$F(s) = \sum_k P_k s^k.$$

For the analysis of continuous time branching processes, one relies on the related generating function

$$f(s) = \sum_k p_k s^k.$$

A very fundamental result of branching process theory states that the extinction probability of a branching process is given by the smallest non-negative root of

$$F(s) = s \quad \text{or} \quad f(s) = 0$$

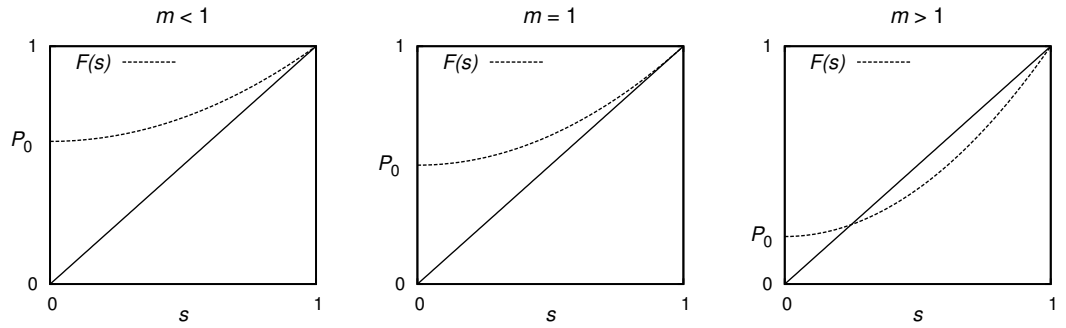
in $[0, 1]$ (Sewastjanow, 1974, p. 31). For the discrete-time process, it is intuitively clear that the extinction probability q is a fixed point of $F(s)$: assume that the initial individual leaves k offspring. Each of this k offspring founds an independent lineage. The process goes extinct when all these k lineages go extinct. Since the lineages are independent, this happens with probability q^k. The probability that the initial individual has k offspring is P_k and therefore

$$q = \sum_k P_k q^k = F(q).$$

It moreover readily follows from the above result that the extinction probability is one for subcritical and critical processes and smaller than one for supercritical processes: $F(s)$ is an increasing function with $F(0) = P_0 \geq 0$ and $F(1) = 1$ and $F'(s) < F'(1)$ for $s < 1$. For the slope at $s = 1$, we obtain $F'(1) = \sum_k kP_k = m$. For subcritical and critical processes ($m \leq 1$), it therefore holds that $F(s) > s$ for $s \in [0, 1)$. In contrast, for supercritical processes ($m > 1$), $F(s)$ and s intersect for some $s \in [0, 1)$. This reasoning is illustrated in Figure 2.

The application of branching process theory to the establishment of adaptations relies on the following idea: While rare, carriers of the beneficial allele suffer nearly independent fates, as do the individuals in a branching process. The early phase of spread of a beneficial allele in a large population (or a population whose size is far below carrying capacity) can therefore be modeled by a branching process. Once the allele is so frequent that its frequency path starts deviating from the branching process, it is highly unlikely to get lost. The survival probability of the branching process hence approximates the establishment probability of the adaptive allele.

In this thesis, I apply branching process theory to study adaptation at a single locus in ecologically or genetically complex scenarios. The branching processes that appear in the analysis are for the most part time inhomogeneous. This



(a) Subcritical process (b) Critical process (c) Supercritical process

Figure 2: For subcritical and critical branching processes, $F(s)$ and s do not intersect in $[0, 1)$. In contrast, for supercritical processes, there exists a fixpoint of $F(s)$ in $[0, 1)$.

means that the transition probabilties explicitly depend on time: for a time-inhomogeneous process, $P_{ij}(t_1, t_2)$ with $t_1 \leq t_2$ is the probability that i individuals at time t_1 leave j descendants at time t_2; in contrast to a time-homogeneous process, the probability does hence not simply depend on the interval $t_2 - t_1$. For processes in continuous time, this implies that the transition rates p_i depend on time ($p_i = p_i(t)$). The mathematical literature provides a body of relevant theory on time-inhomogeneous branching processes (e.g., Kendall, 1948; Cohn and Jagers, 1994). Yet, these results have only been little applied in population genetics so far.

From a biological perspective, time inhomogeneity in the adaptive process arises naturally when ecological changes happen at the same time scale as evolutionary change. Environmental change but also allele frequency shifts at other loci can entail changes in the strength and even the sign of selection. Likewise, the population size is often not constant. Chapter 1 of this thesis establishes key results for the fixation process of a beneficial allele in a variable environment. In particular, I derive approximations for the building blocks of the process, namely, the fixation probability of a single beneficial allele and the time that it needs to reach a given frequency in the population. In Chapter 2, I consider the biological problem of evolutionary rescue, i.e., the question of whether a population can escape extinction through adaptive evolution when faced with severe environmental deterioration. The scenario as described in the chapter is an instance

where ecological and evolutionary processes happen simultaneously. Chapter 3 is concerned with adaptive gene introgression, i.e., the transfer of adaptations across species boundaries. Here, the external environment remains constant. However, the genomic context of the allele changes so that, effectively, it again experiences a time-dependent selection pressure. For this chapter, I extend the modeling framework by multitype branching processes to describe the fate of the various haplotypes that are generated by backcrossing.

Through the work described in this thesis, we gain a better understanding of how a dynamical environment impacts evolutionary adaptation. In particular, the role of spatial and temporal ecological change for adaptive evolution (or extinction) becomes elucidated. I hope that the insights provided by this work inspire new ways of thinking about adaptation in an ecological context.

Chapter 1

On the fixation process of a beneficial mutation in a variable environment

Abstract A population that adapts to gradual environmental change will typically experience temporal variation in its population size and the selection pressure. On the basis of the mathematical theory of inhomogeneous branching processes, we present a framework to describe the fixation process of a single beneficial allele under these conditions. The approach allows for arbitrary time dependence of the selection coefficient $s(t)$ and the population size $N(t)$, as may result from an underlying ecological model. We derive compact analytical approximations for the fixation probability and the distribution of passage times for the beneficial allele to reach a given intermediate frequency. We apply the formalism to several biologically relevant scenarios, such as linear or cyclic changes in the selection coefficient, and logistic population growth. Comparison with computer simulations shows that the analytical results are accurate for a large parameter range, as long as selection is not very weak.

With minor changes, this chapter has been published as: Uecker H. & Hermisson J. (2011), *On the fixation process of a beneficial mutation in a variable environment.* Genetics 188:915–930.

1.1 Introduction

For adaptive evolution to proceed, it is not enough that new beneficial mutations enter a population. To complete an adaptive step, these mutations also need to escape stochastic loss due to genetic drift, get established, and finally rise to fixation. The fixation process of beneficial (or neutral or deleterious) alleles is one of the building blocks of population genetic theory and many of the key results on fixation probabilities and times date back to its early days. Two alternative mathematical frameworks have been developed to derive analytical expressions for these quantities: branching processes (Fisher, 1922, 1930; Haldane, 1927) and diffusion theory (Kimura, 1962; Kimura and Ohta, 1969). Today, a large body of literature exists to study fixation under various ecological scenarios and genetic conditions (reviewed in Patwa and Wahl, 2008), such as the effects of population structure (Whitlock, 2003) and spatial heterogeneity (Whitlock and Gomulkiewicz, 2005), interference due to selection on linked loci (Barton, 1995) or due to epistatic interaction (Takahasi and Tajima, 2005).

In this article, we consider the fixation process in a variable environment, leading to time-dependent selection coefficients and population sizes. Aspects of this problem have already been studied in previous work: In particular, the impact of various scenarios of demographic change (growth, decline, cycles) on the fixation probability has been treated in a series of papers (Ewens, 1967; Chia, 1968; Kimura and Ohta, 1974; Otto and Whitlock, 1997; Pollak, 2000; Parsons and Quince, 2007a; Orr and Uncкless, 2008). Studies on time-dependent selection mostly concentrate on stochastic fluctuations of the selection coefficient (Jensen, 1973; Karlin and Levikson, 1974; Takahata *et al.*, 1975; Huillet, 2011). Since the distribution of the selection coefficients is constant across generations, these models are still time homogeneous in a probabilistic sense. In contrast, surprisingly little is known when the changes of the selection coefficient $s = s(t)$ follow an explicit trend. Ohta and Kojima (1968) derive an expression for the fixation probability of a mutation with time-dependent selective advantage in the context of the evolution of chromosomal inversions. Apart from that, only particular functions have been considered: Kimura and Ohta (1969) discuss the case where selection decreases exponentially in time, and Pollak (1966) derives expressions for the fixation probability under two alternating selection pressures. No previous work seems to exist where both population size and selection strength are variable, although this is a generic

case under realistic ecological conditions. Also, there does not seem to be an investigation of fixation or passage times in variable environments.

In the following, we present a formalism to describe the fixation process under a wide range of scenarios of environmental variation. We use branching processes in continuous time to derive analytical approximations for the fixation probability and the passage time needed for the mutant allele to reach some intermediate frequency x_c. After the introduction of our model, we describe how a general approximation for the fixation probability can be obtained from known mathematical results on inhomogeneous branching processes. Afterwards, we discuss applications of this result to several biologically relevant scenarios. In the second part of the article, we introduce and apply a method to calculate the distribution of the passage time needed for a beneficial mutation to reach an intermediate frequency. The method works by combining the stochastic fluctuations from the branching approximation with the deterministic growth of the full model. This technique has been used before (for constant selection and population size) by Desai and Fisher (2007) in a model of clonal evolution. All analytical results are complemented by computer simulations, which are briefly described in a separate section. We close with a short discussion. In Appendix A, we discuss a generalized version of the model to include allele-frequency-dependent population demographies. We illustrate how the formalism can be used by applying it to an illustrative ecological scenario: the fixation probability of a "rescue mutation" in a population that is threatened by extinction (cf. Orr and Uncкless, 2008). Additional material is devoted to the Appendices B–D. We discuss in some detail the scope and limits of the approach and the accuracy of the approximation. In Appendix C, we present an alternative treatment to derive the fixation probability in a variable environment from a diffusion approach.

1.2 The model

We consider a large population of haploid individuals with time-dependent population size N_t. The population dynamics are modeled as a time-inhomogeneous birth-death process with birth and death rates $b(t, N_t)$ and $d(t, N_t)$:

$$\begin{aligned} N_t \to N_t + 1: &\quad b(t, N_t) N_t, \\ N_t \to N_t - 1: &\quad d(t, N_t) N_t. \end{aligned} \tag{1.1}$$

The impact of the changes in the external environment on the population size is reflected in the explicit time dependence of the rates on t. The dependence on N_t accounts for density dependence (e.g., logistic: $b(t, N_t) = b_1(t) - b_2(t)N_t$). We call $r(t, N_t) = b(t, N_t) - d(t, N_t)$ the growth parameter. Obviously, the expected change of N_t over a small time intervall dt reads

$$E[\Delta N|N_t] = r(t, N_t)N_t \, \mathrm{d}t. \tag{1.2}$$

Consider now two alleles, a beneficial mutant allele A and the ancestral (resident) allele a, that segregate in the population at a single locus. Recurrent mutations in both directions are ignored. In general, birth and death rates might be different for residents and mutants. These rates can depend on time and on the (absolute) frequencies of both allelic types, allowing for general frequency-dependent selection. As a consequence, also the population dynamics depends on the allelic composition and cannot be described by Eq. (1.1) anymore. We discuss this model in Appendix A. For the main part of the article, however, we assume that the rates are the same for mutants and residents and that all model parameters are independent of allele frequencies. This means, in particular that selection is soft; i.e., changes in the allelic composition due to selection or drift do not interfere with the population dynamics. Population growth and decline of the polymorphic population are then correctly described by Eq. (1.1).

In this setting, selection is modeled as competitive replacement between individuals, which does not change the population size, and is implemented as follows: At per capita rate $\xi(t, N_t) + s(t, N_t)$, a mutant additionally reproduces and succeeds in replacing a randomly chosen individual from the population by its offspring. Residents do the same at rate $\xi(t, N_t)$. Again, the selective advantage $s(t, N_t)$ of the mutant may thus depend on the external environment (modeled by the dependence of $s(t, N_t)$ on t) and the population size (modeled by the dependence on N_t). Changes in the number of mutants then occur at rates:

$$\begin{aligned} n_t \to n_t + 1: &\quad \big(\xi(t, N_t) + s(t, N_t)\big)\frac{n_t(N_t - n_t)}{N_t} + b(t, N_t)n_t, \\ n_t \to n_t - 1: &\quad \xi(t, N_t)\frac{n_t(N_t - n_t)}{N_t} + d(t, N_t)n_t. \end{aligned} \tag{1.3}$$

The model corresponds to a continuous-time Moran model, but with a population size that may change in time. Putting $b(t, N_t) = d(t, N_t) = 0$, $\xi(t, N_t) = 1$

and $s(t, N_t) = s = \text{const.}$ reproduces the standard Moran model (Moran, 1958a,b; Novozhilov *et al.*, 2006). The free parameter $\xi(t, N_t)$ has been introduced to our model to allow for easy interpolation to other models (see below) and additionally to make the analysis of density-dependent competition possible.

To further clarify the relation to other models, we calculate how the frequency of mutants $x_t := n_t/N_t$ changes over time. Let Δx be its change in an infinitesimal time interval $\mathrm{d}t$. The expectation and the variance of Δx are calculated to be:

$$\mathrm{E}[\Delta x|x_t, N_t] = s(t, N_t)x_t(1 - x_t)\mathrm{d}t, \tag{1.4a}$$

$$\mathrm{Var}[\Delta x|x_t, N_t] \approx \mathrm{E}[(\Delta x)^2|x_t, N_t] \approx \frac{x_t(1 - x_t)}{N_{e,t}}\mathrm{d}t \tag{1.4b}$$

with the time-dependent variance effective population size

$$N_{e,t} = \frac{N_t}{2\xi(t, N_t) + b(t, N_t) + d(t, N_t) + s(t, N_t)}. \tag{1.5}$$

In the last step we approximated $N_t + 1 \approx N_t$ and $N_t - 1 \approx N_t$ (see Appendix D for the derivation of Eq. (1.4a) and (1.4b)).

We see that the strength of drift, measured as $N_{e,t}^{-1}$, is proportional to the total rate of events in the model. The choice $2\xi(t, N_t) + b(t, N_t) + d(t, N_t) = 2$ coincides with the strength of drift in the standard Moran model, while $2\xi(t, N_t) + b(t, N_t) + d(t, N_t) = 1$ is consistent with the scaling in the Wright-Fisher model. In contrast to many diffusion or coalescent approaches, we do not rescale time with the effective population size (which would be impractical since $N_{e,t}$ itself depends on t). Generation time in the continuous-time Moran model is defined as the inverse of the total death rate of an individual, $(\xi(\tau, N_\tau) + d(\tau, N_\tau))^{-1}$, and may again depend on time in our model.

1.3 Fixation probability

1.3.1 Analytical theory

Following pioneering work by Haldane (1927) and Fisher (1930), there has been a long tradition in population genetics to calculate fixation probabilities by branching process methods (reviewed in Haccou *et al.* (2005) and Patwa and Wahl (2008)). For the general time-dependent case, the relevant results have long been known in the mathematical literature (e.g., Kendall, 1948; Allen,

2011). However, only specific cases (usually in the context of changing population sizes) have been discussed in the population genetics context (Ewens, 1967; Otto and Whitlock, 1997; Pollak, 2000; Wahl and Gerrish, 2001). We will therefore give a brief outline of the general theory below and show how it applies to the biological problem at hand. Previous results are recovered as special cases.

The branching process approximation is based on the following reasoning: Initially, the fate of a new beneficial mutation arising in a population will be strongly determined by genetic drift. In most cases, it will actually get lost again. Once the mutation has survived this early phase, it is, however, almost sure to get fixed given its selective advantage is large enough. To calculate the fixation probability it is therefore often sufficient to consider the stage at which the mutant population size n_t is still small relative to the total population size N_t. In this early phase, the mutant individuals suffer nearly independent fates, as do the individuals in a Galton-Watson branching process (this assumption is precisely met in an infinite population). The extinction probability of the latter can therefore be used as an approximation for the probability that the mutation gets lost. Because a mutation in a finite population is in the long term either fixed or lost, the fixation probability is the complementary probability. (We exclude the unbiological case of a population that increases without bounds, where possibly neither wildtypes nor mutants become extinct.)

Ignoring terms proportional to $x_t = n_t/N_t$ in the birth-death model (1.3), which corresponds to the limit $N_t \to \infty$, leads to transition rates that are proportional to n_t. Following standard practice in ecological modeling, we further ignore stochastic fluctuations in the population dynamics. This is done by replacing the stochastic variable N_t by its deterministic approximation denoted as $N(t)$, with dynamics

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = r(t, N(t))N(t). \tag{1.6}$$

Inserting the deterministic solution $N(t)$ into the rates for birth, death and selection, reduces the dependence of these rates on t and N_t to a dependence on t only ($s(t, N_t) \to s(t, N(t)) =: s(t)$ etc.). We arrive at a branching process with time-dependent per capita birth and death rates:

$$\begin{aligned} &\text{birth rate} : \lambda(t) = b(t) + \xi(t) + s(t), \\ &\text{death rate} : \mu(t) = d(t) + \xi(t). \end{aligned} \tag{1.7}$$

As explained above for the birth-death process, the total rate of events determines the strength of genetic drift, while $\lambda(t) - \mu(t) = s(t) + r(t)$ corresponds to the absolute expected rate of increase of a small mutant population (its Malthusian fitness parameter).

To derive the extinction probability of this process, we follow Allen (2011, p. 278ff). Let $p_i(n_0, t)$ be the probability that there are i individuals at time t when the process started with n_0 individuals at time $t = 0$. Using the Kolmogorov forward equation,

$$\frac{\mathrm{d}p_i(n_0,t)}{\mathrm{d}t} = \lambda(t)(i-1)p_{i-1}(n_0,t) + \mu(t)(i+1)p_{i+1}(n_0,t) - (\lambda(t)+\mu(t))ip_i(n_0,t), \tag{1.8}$$

a differential equation for the probability generating function $\sum_{i=0}^{\infty} p_i(n_0,t)z^i = P_{n_0}(z,t)$ of the branching process can be derived (see Allen, 2011, p. 279):

$$\frac{\partial P_{n_0}(z,t)}{\partial t} = \left[\lambda(t)(z^2-z) + \mu(t)(1-z)\right]\frac{\partial P_{n_0}(z,t)}{\partial z}, \qquad P_{n_0}(z,0) = z^{n_0}. \tag{1.9}$$

The solution is known from the mathematical literature (Kendall, 1948; Allen, 2011, p. 280) and given as

$$P_{n_0}(z,t) = \left(1 + \frac{1}{\frac{\mathrm{e}^{-\rho(t)}}{z-1} - \int_0^t \lambda(\tau)\mathrm{e}^{-\rho(\tau)}\mathrm{d}\tau}\right)^{n_0} \tag{1.10}$$

with

$$\rho(t) = \int_0^t [\lambda(\tau) - \mu(\tau)]\mathrm{d}\tau = \int_0^t [s(\tau) + r(\tau)]\mathrm{d}\tau. \tag{1.11}$$

To keep notation short, we introduce the abbreviations

$$A(t) := \mathrm{e}^{-\rho(t)}, \tag{1.12a}$$

$$B(t) := \int_0^t \lambda(\tau)\mathrm{e}^{-\rho(\tau)}\mathrm{d}\tau. \tag{1.12b}$$

The extinction probability $p_0(n_0, t)$ is immediately obtained from the generating function via

$$p_0(n_0,t) = P_{n_0}(0,t) = \left(1 - \frac{1}{(A+B)(t)}\right)^{n_0} = \left[\frac{\int_0^t \mu(\tau)\mathrm{e}^{-\rho(\tau)}\mathrm{d}\tau}{1+\int_0^t \mu(\tau)\mathrm{e}^{-\rho(\tau)}\mathrm{d}\tau}\right]^{n_0} . \quad (1.13)$$

The probability that the mutation will eventually fix in the population is therefore given by

$$p_{\mathrm{fix}}(n_0) = 1 - \lim_{t\to\infty} p_0(n_0,t) = 1 - \left(1 - \frac{1}{A+B}\right)^{n_0}, \quad (1.14)$$

where we introduced

$$A + B = \lim_{t\to\infty}(A+B)(t) = 1 + \int_0^\infty \mu(t)\mathrm{e}^{-\rho(t)}\mathrm{d}t = \int_0^\infty \lambda(t)\mathrm{e}^{-\rho(t)}\mathrm{d}t. \quad (1.15)$$

The last equality is valid for $\lim_{t\to\infty} \exp(-\rho(t)) = 0$. This condition is met in all examples considered below. For a single mutant ($n_0 = 1$), it thus holds:

$$p_{\mathrm{fix}} = \frac{1}{A+B} = \frac{2}{1+\int_0^\infty (\lambda+\mu)(t)\exp\left(-\int_0^t (s+r)(\tau)\mathrm{d}\tau\right)\mathrm{d}t} \quad (1.16a)$$

$$= \frac{2}{1+\int_0^\infty \frac{N(0)}{N_e(t)}\exp\left(-\int_0^t s(\tau)\mathrm{d}\tau\right)\mathrm{d}t}, \quad (1.16b)$$

where we have used $\int_0^t r(\tau)\mathrm{d}\tau = \int_0^t \frac{\dot{N}(\tau)}{N(\tau)}\mathrm{d}\tau = \ln(N(t)/N(0))$ for the last equality and $N_e(t)$ is defined analogously to Eq. (1.5). A similar expression was also derived by Ohta and Kojima (1968). Restricted to a constant population size and a Poisson offspring distribution, their result is, however, less general.

The result depends on two independent parameters, which are compositions of three biologically relevant factors: the strength of selection given by $s(t)$, the combined rate of birth and death events $(\lambda+\mu)(t)$, and the changes in the total population size (modeled by $r(t)$ or $N(t)$). In the above equation, we formulated the result via two different combinations of these three variables. In the first version (1.16a), it is expressed in terms of the absolute rate of increase of mutants

in the population $(s+r)(t)$ and the total rate $(\lambda+\mu)(t)$ at which events happen, which defines the time scale of the problem and also quantifies the influence of drift. In the second formulation (1.16b) of the result, we combined the birth- and death rates and the changing population size to the time-dependent variance effective population size. The second decisive parameter is the selection coefficient of the mutation. Depending on the question to be answered, one or the other version is more favorable. In the first version, the correspondence between a mutation with time-dependent selective advantage and a mutation in a population of changing size can be easiliy seen: a mutation with time-dependent selective advantage $s(t)$ in a population of (on average) constant size ($b(t) = d(t) = \text{const.}$) has the same chance to reach fixation as a mutation with constant selective advantage s_0 in a population with time-dependent death rate $d(t)$ and growth parameter $r(t) = s(t) - s_0$. (Note, however, that s_0 must be larger than 0, such that fixation is (almost) certain once the mutation has survived genetic drift.) The second version is closer to the traditional view in population genetics. It is advantageous if the variance effective population size is directly given, and allows, in particular, also for the treatment of discontinuous changes in the population size.

We see that the fixation probability is independent of many details of the individual level dynamics of the original process (Eq. (1.3)), which depends on four rates ($b(t, N_t)$, $d(t, N_t)$, $s(t, N_t)$ and $\xi(t, N_t)$). Further aspects of the stochastic model are ignored by our deterministic approximation for the population dynamics. In particular, the analytical results become independent of the particular form of density regulation. As an example, consider three scenarios: (1) a population with inherently constant size N_0 with $b(t, N_t) = d(t, N_t) = 0$ and $\xi(t, N_t) = c$; (2) a population with density regulation according to $b(t, N_t) = c + \rho(1 - N_t/N_0)$ and $d(t, N_t) = c$, with initial size N_0, and $\xi(t, N_t) = 0$; and (3) a population without density regulation with $b(t, N_t) = d(t, N_t) = c$, initial size N_0, and $\xi(t, N_t) = 0$. In all three cases, the deterministic dynamics of the population size are the same ($N(t) = N_0$) and the analytical predictions for the fixation probability coincide for arbitrary $s(t, N_t)$. Simulation results of all three scenarios indeed showed no significant difference, justifying the approximation (see Appendix B). This observation agrees with findings by Parsons and Quince (2007a) that demographic stochasticity does not significantly influence the fixation probability of advantageous alleles (Parsons and Quince (2007a) discuss this issue for a population that starts in the vicinity of the dynamic equilibrium). For concreteness, we will use the notion of "constant population size" in the following to refer to the case of a strictly constant population size ($b(t, N_t) = d(t, N_t) = 0$). We further set

$\xi(t) = \xi =$ const. in all applications. If not stated otherwise, we use $\xi = 1$ (corresponding to the Moran model scaling) for the results shown in the figures.

A necessary condition for the branching approximation to yield meaningful results is that the fate of the mutation is decided as long as the mutant frequency $x_t = n_t/N_t$ is still small. As x_t increases, mutants are no longer independent in the original birth-death process (Eq. (1.3)), and both processes differ significantly from each other. In particular, the birth-death process has a second absorbing boundary at $x_t = 1$. As a consequence, neutral or even deleterious mutations, too, can become fixed by genetic drift. In the corresponding branching model, an upper absorbing boundary does not exist. Consequently, neutral or deleterious alleles must go extinct in the long term. For an allele with a general time-dependent selection coefficient, the branching process approximation is thereby valid only if the allele is "sufficiently beneficial on average". We can formalize this condition as follows: Since divergence of the (first) integral in Eq. (1.15) leads to the wrong prediction of a zero fixation probability, we need to require that the integral converges. If we assume that $\mu(t)$ is bounded below by a constant $C_\mu > 0$, a necessary condition for the convergence of the integral is divergence of the integral in Eq. (1.11). More precisely: if we assume that $\mu(t)$ has bounds $C_\mu > 0$, C^μ such that $C_\mu < \mu(t) < C^\mu$, it must hold that $\lim_{t\to\infty} \rho(t) > C \ln[t]$ for some constant $C > 1$. A prominent example where this condition is not fulfilled is a mutation with an exponentially decreasing selection coefficient in a population of constant size (Kimura and Ohta (1969), cf. also Pollak (1966) and Ohta and Kojima (1968) for mutations that lose their advantage over time). The condition is, however, not sufficient to obtain good results: it is also necessary that the contribution of times beyond the initial phase be negligible. In particular, changes in the environment at times larger than the fixation time must not have an impact on the results. As shown in Appendix B, the approximation will usually be excellent if the estimate of p_{fix} from Eq. (1.16) fulfills $p_{\text{fix}} N \gtrsim 10$. By a similar reasoning it is clear that the requirement that the mutation be beneficial all the time is at this point unnecessarily strict. It is sufficient that the extinction probability is in both processes – the original birth-death process and the branching approximation – negligible once the number of mutants has reached a certain size.

1.3.2 Applications

Constant selection and constant population size. Setting $\lambda(t) = \xi + s$ and $\mu(t) = \xi$ yields for the fixation probability

$$p_{\text{fix}} = \frac{s}{\xi + s} \approx \frac{s}{\xi} \approx 2s\frac{N_e}{N}. \tag{1.17}$$

The well-known result found by HALDANE (1927) is therefore reproduced for $N_e = N$. In the special case of a constant environment, it is possible to calculate the exact fixation probability from the transition matrix defined via Eq. (1.3) (see EWENS, 2004, p. 90). One obtains

$$p_{\text{fix}}^{(\text{exact})} = \frac{\frac{s}{\xi+s}}{1 - \left(\frac{\xi}{\xi+s}\right)^N} \approx \frac{\frac{s}{\xi+s}}{1 - \exp\left(-\frac{s}{\xi+s}N\right)}, \tag{1.18}$$

which shows that the branching approximation is very accurate for large values of sN/ξ. Furthermore, it is possible to calculate the fixation probability of a deleterious mutation with selective disadvantage $-s$. To do so, we switch the roles of the transition rates defined in Eq. (1.3). Every path leading to fixation $n = N$, has now a chance to be realized which is by a factor of $(\xi + s)^{-N}$ lower than the corresponding path for a beneficial mutation. It immediately follows that

$$p_{\text{fix}}(-s) = \frac{p_{\text{fix}}(s)}{(\xi + s)^N} \overset{\xi=1}{\approx} \frac{p_{\text{fix}}(s)}{\exp(sN)}. \tag{1.19}$$

Constant selection and changing population size. OTTO and WHITLOCK (1997) analyzed the fixation probability for several important scenarios of demographic change. Our general result (Eq. (1.16b)) contains all those scenarios, among arbitrary others. In addition, it provides some insight into the influence of the individual-level dynamics on the fixation probability. As an example, we consider a population that follows logistic growth (or decline) until it has reached its new carrying capacity K. There are different ways to describe this global dynamics at the individual level. We discuss two possibilities, which arise naturally in a biological context: The first one assumes that a decreasing availability of resources per individual leads to a lower birth rate while the death rate stays constant (e.g., fertility is reduced). The second one assumes that the same circumstances lead to a higher death rate while the birth rate stays constant. The

selection coefficient s of the beneficial mutant is constant in both cases. In the first scenario, the birth and death rates are given by

$$\begin{aligned} b(t, N_t) &= b + r\left(1 - \frac{N_t}{K}\right), \\ d(t, N_t) &= b. \end{aligned} \tag{1.20}$$

The total population size thus changes according to

$$N(t) = \frac{KN_0}{(K - N_0)\mathrm{e}^{-rt} + N_0}, \tag{1.21}$$

and using Eq. (1.16b) yields

$$p_{\mathrm{fix}}^{(1)} = \frac{s(r+s)}{s(s+r) + (b+\xi)(s+r\gamma_0)} \tag{1.22}$$

with $\gamma_0 = \frac{N_0}{K}$.

In the second scenario, birth and death rates are then given by

$$\begin{aligned} b(t, N_t) &= b + r, \\ d(t, N_t) &= b + r\frac{N_t}{K}. \end{aligned} \tag{1.23}$$

While the deterministic dynamics at the population level are the same for both scenarios (given by Eq. (1.21)), this is not true for the fixation probability. From Eq. (1.16b), we obtain for the fixation probability in the second scenario

$$p_{\mathrm{fix}}^{(2)} = \frac{s(r+s)}{s(r+s) + (b+\xi)(s+r\gamma_0) + \gamma_0 r(r+s)} \leq p_{\mathrm{fix}}^{(1)}, \tag{1.24}$$

i.e., a reduced probability relative to the first scenario. This is explained by the fact that genetic drift is stronger in the second scenario. For small values of s and r, both fixation probabilities are approximately the same $p_{\mathrm{fix}}^{(1)} \approx p_{\mathrm{fix}}^{(2)} \approx \frac{s(r+s)}{(b+\xi)(s+r\gamma_0)}$ and reproduce the result found by Otto and Whitlock (1997) if we choose $b+\xi = 0.5$ such that the variance of the increase in the mutant number $\mathrm{Var}[\Delta n|n_t]$ is the same as in their model.

Note that for a sudden jump in population size, the result of our theory coincides with the result derived by Otto and Whitlock (1997) (where selection is effective during the change in population size), while Wahl and Gerrish (2001)

consider a slightly different situation (where selection is switched off during the bottleneck; cf. Patwa and Wahl (2008)).

Linearly increasing selection. When environmental conditions develop continuously in a given direction, the selective advantage of a mutation may gradually increase during the fixation process. Let us assume that the total population size stays constant and that the selection coefficient of the mutation linearly increases in time, thus $s(t) = s_0 + s_1 t$. We obtain the following expression for p_{fix} from Eq. (1.16a):

$$p_{\text{fix}} = \left[1 + \xi\sqrt{\frac{\pi}{2s_1}}\mathrm{e}^{\frac{s_0^2}{2s_1}}\operatorname{erfc}\left(\frac{s_0}{\sqrt{2s_1}}\right)\right]^{-1}, \tag{1.25}$$

where $\operatorname{erfc}(x) = \frac{2}{\pi}\int_x^\infty \mathrm{e}^{-t^2}\mathrm{d}t$ is the complementary error function. For the special case $s_0 = 0$, the result simplifies,

$$p_{\text{fix}}(s_0 = 0) = \left[1 + \xi\sqrt{\frac{\pi}{2s_1}}\right]^{-1} \approx \frac{\sqrt{s_1}}{\xi}\sqrt{\frac{2}{\pi}}. \tag{1.26}$$

In Figure 1.1 the analytical results are compared to simulation results, showing very good agreement. The fixation probability increases significantly with s_1. Generally:

$$\begin{aligned}\frac{p_{\text{fix}}(s(t) = s_0 + s_1 t)}{p_{\text{fix}}(s(t) = s_0)} &= \frac{\xi + s_0}{s_0 + \xi\sqrt{\frac{\pi s_0^2}{2s_1}}\mathrm{e}^{\frac{s_0^2}{2s_1}}\operatorname{erfc}\left(\frac{s_0}{\sqrt{2s_1}}\right)} \\ &\approx \left[\sqrt{\frac{\pi}{2}}y\mathrm{e}^{\frac{y^2}{2}}\operatorname{erfc}\left(\frac{y}{\sqrt{2}}\right)\right]^{-1} =: f(y)\end{aligned} \tag{1.27}$$

with $y := \frac{s_0}{\sqrt{s_1}}$. The function $f(y)$ is shown in the inset of Figure 1.1. For $y = 1$ it is evaluated to be $f(1) \approx 1.52$. This means that if $y \approx 1$, i.e., $s_0 \approx \sqrt{s_1}$, we observe an increase in the fixation probability of $\sim 50\%$ in comparison to a mutation with constant selective advantage. E.g., for initially moderately strong selection $s_0 = 0.01$ the fixation probability is still increased by $\approx 50\%$ if selection increases as slightly as $s(t) = 0.01 + 0.0001t$.

Periodically changing selection. Cyclic environmental changes, such as seasonal changes or cyclic climate fluctuations (like the El Niño phenomenon), are

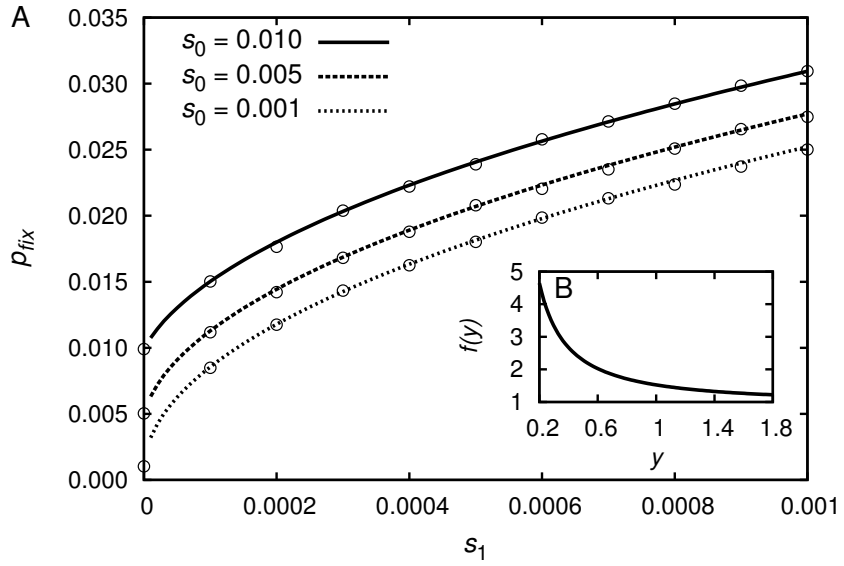


Figure 1.1: A: Fixation probability for linearly increasing selection strength $s(t) = s_0 + s_1 t$ in dependence of s_1 for various values of s_0. Simulations were performed for a population of 100 000 individuals, and each simulation point is the average over 10^6 runs. B: Function $f(y)$ (see Eq. (1.27)). The fixation probability increases significantly with increasing s_1.

frequent in nature. In the following, we consider strictly periodic changes, although our theory does not rely on this condition. Let us assume that again the total population size stays unaffected, but that the selective advantage of a mutation changes periodically with time, e.g.,

$$s(t) = s_0 + (s_{\max} - s_0)\cos(\omega t + \varphi). \tag{1.28}$$

Depending on the parameter values, it is thus possible that the mutation is disadvantageous at certain periods of time. We obtain for the fixation probability:

$$p_{\text{fix}} = \left[1 + \xi \int_0^\infty \mathrm{e}^{-s_0 t - (s_{\max}-s_0)\left(t + \frac{1}{\omega}\sin(\omega t+\varphi) - \frac{1}{\omega}\sin(\varphi)\right)} \mathrm{d}t\right]^{-1}. \tag{1.29}$$

The integral can be evaluated only numerically. Comparison to simulated data (see Figure E.5) shows that the theory provides an accurate prediction of the fixation probability also for scenarios in which the mutation temporarily gets disadvantageous. For $s_0 \leq 0$, however, it predicts a fixation probability of zero and therefore underestimates the true value.

To slightly reduce the parameter space, we concentrate now on the special case $s_{\max} = 2s_0$, i.e.,

$$s(t) = s_0(1 + \cos(\omega t + \varphi)). \tag{1.30}$$

Figure 1.2 shows how the fixation probability changes with ω for various values of φ. For small and intermediate values of ω, the value of φ has a strong impact on the result. If ω increases, the fixation probability converges to the value for a mutation with selective advantage $s = s_0$ for all values of φ. This behavior is further illustrated in Figure 1.3, which shows the fixation probability and the initial selection strength in dependence of φ for various values of ω. In the limit $\omega \to 0$, the fixation probability equals $s(0)/(1 + s(0)) \approx s(0)$. It therefore follows approximately the curve of $s(0)$ for small values of ω. For large values, it converges to s_0 for all values of φ. In an intermediate regime $s_0/10 \lesssim \omega \lesssim 10s_0$, more complex behavior is found: here, not only the initial value $s(0)$ but also the following time-development of the selection strength (i.e., not only $s(0) = s_0(1 + \cos(\varphi))$, but φ itself) becomes important. The extrema of $p_{\text{fix}}(\varphi)$ are

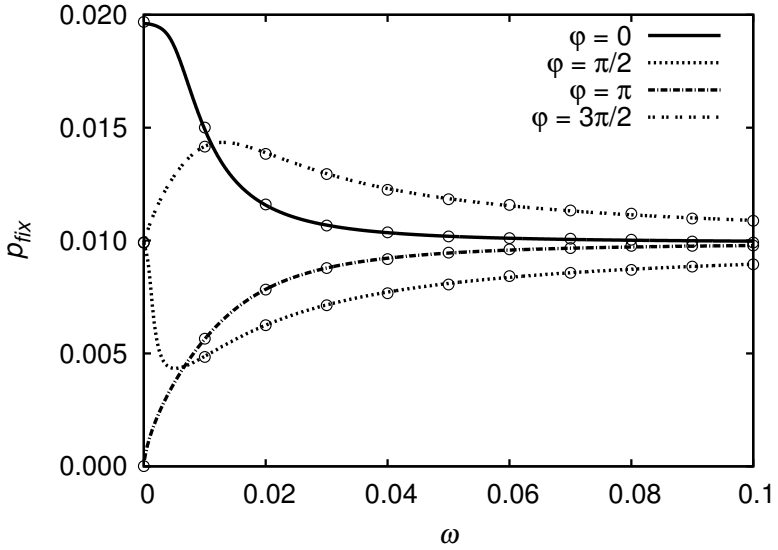


Figure 1.2: Comparison between the analytical theory and simulation results for the fixation probability in the case of periodically changing selection $s(t) = s_0(1 + \cos(\omega t + \varphi))$, $s_0 = 0.01$. Simulations were performed for a population of 100 000 individuals, and every simulation point is the average over 10^7 runs.

attained for smaller values of φ than the extrema of $s(0)$. A straightforward calculation shows furthermore that

$$p_{\text{fix}}(\varphi_e) = \frac{s_0(1 + \cos(\varphi_e))}{1 + s_0(1 + \cos(\varphi_e))}, \tag{1.31}$$

where φ_e is the value where the extremum is attained. We thus see that the fixation probability at the extrema φ_e are the same for cyclic selection and constant selection with $s = s(0) = s_0(1 + \cos(\varphi_e))$.

1.4 Time to reach an intermediate frequency x_c

1.4.1 Analytical theory

In models of the adaptive process, it is often necessary to know the time that it takes for a successful mutation to become established and to reach a certain threshold frequency (e.g., Desai and Fisher, 2007; Kopp and Hermisson,

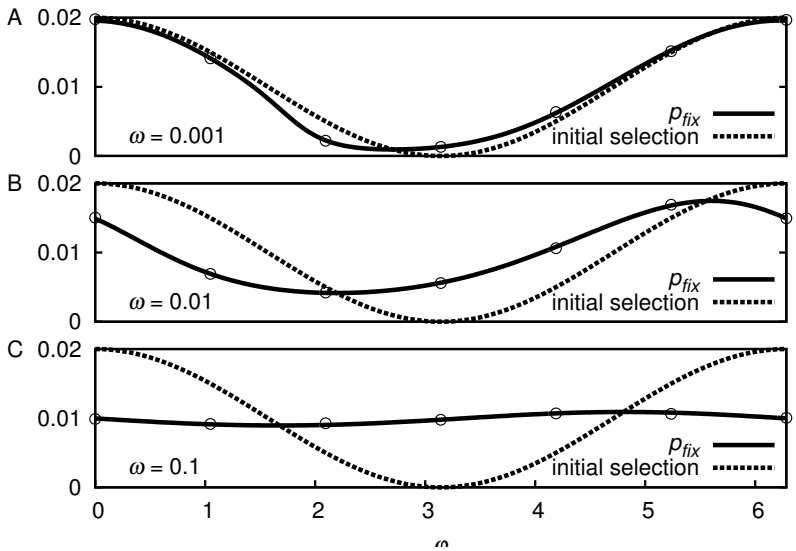


Figure 1.3: Fixation probability and initial selection strength $s(0) = s_0(1 + \cos(\varphi))$ in dependence of φ for various values of ω when the selection coefficient changes according to $s(t) = s_0(1 + \cos(\omega t + \varphi))$, $s_0 = 0.01$. Again, one observes the expected limiting behavior for small and for large values of ω. For intermediate values of ω a non-trivial behavior is found. Simulations were performed for a population of 100 000 individuals, and each simulation point is the average over 10^6 runs.

2009a). It is well-known that the deterministic model of the allele frequency increase yields a poor approximation even for the expected value of this time. The reason is that a mutation that has survived the initial phase will on average have grown faster during this phase than predicted by the deterministic model. When the frequency is finally large enough that stochasticity can be neglected and the path be modeled deterministically, the frequency will thus be larger than if it had *always* grown following the deterministic path. In that phase, it is well described by the deterministic path that is not started from a single individual, but from an (on average) larger "effective initial population size". The method presented here builds on work by Cohn and Jagers (1994) and Desai and Fisher (2007). It consists of subsuming all the stochasticity of the path under this effective initial population size as a single random variable and then modeling the path deterministically. The procedure consists of two steps: in a first step, the distribution of the effective initial population size is estimated via a branching process. In a second step, the deterministic approximation of the full birth-death model is used to describe the allele frequency path starting from the effective initial population size.

Let us again consider the phase in which the mutation is rare and in which the dynamics can be described by the time-inhomogeneous branching process (Eq. (1.7)). The key to the method is that it is possible to separate stochastic fluctuations and deterministic growth for this process. Here, deterministic growth coincides with the time-development of the expected number of mutants:

$$E[n_t] = n_0 \exp(\rho(t)). \tag{1.32}$$

In the following, we restrict to the case $n_0 = 1$. Define now a new random variable:

$$\nu_t := \frac{n_t}{E[n_t]} = n_t \exp(-\rho(t)). \tag{1.33}$$

ν_t describes all the stochasticity that has accumulated in the branching process until time t. Crucially, a theorem by Cohn and Jagers (1994) guarantees that for $t \to \infty$, ν_t converges (almost surely) to a positive random variable ν that summarizes the entire stochasticity of the process. Since $n_t = \nu \exp(\rho(t))$ in the limit $t \to \infty$, we can interprete ν as the random initial population size of an ensemble of deterministically growing paths that approximate the original process n_t for large t.

For the fixation process, in particular, we are interested in the distribution of ν conditioned on non-extinction. We proceed as follows: From the probability

generating function $P(z,t) \equiv P_1(z,t)$ (see Eq. (1.10)), we obtain for the probability $p_n(t) \equiv p_n(1,t)$ to have n individuals at time t (proof by induction, see Appendix D):

$$p_n(t) = \frac{1}{n!} \left. \frac{\mathrm{d}^n P(z,t)}{\mathrm{d}z^n} \right|_{z=0} = \frac{B^{n-1}(t)A(t)}{(A+B)^{n+1}(t)}, \quad n \geq 1. \tag{1.34}$$

Conditioning on non-extinction (which requires the biologically meaningful condition $p_0(t) \neq 1$) leads to:

$$\mathrm{Prob}[n,t|\text{not extinct}] = \frac{p_n(t)}{1-p_0(t)} = \frac{A(t)}{B(t)} \left(\frac{B(t)}{(A+B)(t)} \right). \tag{1.35}$$

By induction, we find:

$$E[n_t^k|\text{not extinct}] = \frac{1}{1-p_0(t)} E[n_t^k] \quad \text{and} \quad E[n_t^k] \sim e^{k\rho(t)}, \tag{1.36}$$

i.e., for large times the kth moment grows in the conditioned as well as in the unconditioned process as $e^{k\rho(t)}$. In particular, we obtain for the expected value:

$$E[n_t|\text{not extinct}] = \frac{1}{1-p_0(t)} e^{\rho(t)}. \tag{1.37}$$

(Since $\frac{1}{1-p_0(t)} \geq 1$, a comparison with Eq. (1.32) confirms that on average the conditioned path grows faster than the unconditioned path.) The distribution function of $\nu_t|$not extinct is immediately obtained from the distribution function of $n_t|$not extinct:

$$P(n_t \leq n_0|\text{not extinct}) = 1 - \left(\frac{B(t)}{(A+B)(t)} \right)^{n_0} \tag{1.38}$$

$$\begin{aligned} \Rightarrow & P(\nu_t \leq \nu_0|\text{not extinct}) \\ & = 1 - \exp\left[\nu_0 \exp\left(\rho(t)\right) \ln\left(\frac{B(t)}{(A+B)(t)} \right) \right]. \end{aligned} \tag{1.39}$$

As $\lim\limits_{t\to\infty} \exp\left(\rho(t)\right) \ln\left(\frac{B(t)}{(A+B)(t)} \right) = \frac{-1}{A+B} = -p_{\mathrm{fix}}$, we obtain in the limit $t \to \infty$ the stationary distribution

$$P(\nu \leq \nu_0|\text{not extinct}) = 1 - \exp\left(-p_{\mathrm{fix}}\nu_0\right). \tag{1.40}$$

This effective initial mutant population size may now be used as the starting value for the deterministic solution of the full model. If the birth and death rates are the same for mutants and residents, the latter can be obtained from Eq. (1.4a) and is given by a generalized logistic,

$$x(t) = \frac{\exp(\beta(t))}{\frac{N_0}{\nu} - 1 + \exp(\beta(t))} \tag{1.41}$$

where $x(t)$ is the mutant frequency and

$$\beta(t) = \int_0^t s(\tau)\mathrm{d}\tau. \tag{1.42}$$

We want to calculate the time needed to reach an intermediate frequency x_c. Here, intermediate frequency means a frequency at which the dynamics is well described by Eq. (1.41), i.e., not too close to either 0 or 1. Using Eq. (1.41), we can express the random variable T_{x_c} defined via $x(T_{x_c}) = x_c$ in terms of ν

$$T_{x_c} = \begin{cases} \beta^{-1}\left(\ln\left(\frac{x_c(N_0-\nu)}{(1-x_c)\nu}\right)\right), & \nu < x_c N_0, \\ 0, & \nu \geq x_c N_0. \end{cases} \tag{1.43}$$

For the definition of T_{x_c} to be unique for all x_c, the condition $s(t) > 0$ (up to single points) is necessary. It is possible that the process grows so quickly that the effective initial mutant population size is already larger than $x_c N$ (cf. the exponential tail in the distribution function for the effective initial population size ν Eq. (1.40)). In this case, we have set $T_{x_c} = 0$ so that the distribution of T_{x_c} will have a mass on $T_{x_c} = 0$. However, if x_c is not too small, this is very unlikely and the mass will be small. As the deterministic path Eq. (1.41) ignores stochastic fluctuations, there is only a single passage time in the deterministic approximation in contrast to the stochastic path in which the frequency x_c can be hit several times. T_{x_c} as defined in Eq. (1.43) is best interpreted as the average over all times at which the path crosses the frequency x_c from lower to higher values.

By a parameter transformation ($\nu \to T_{x_c}$ using Eq. (1.43)), the distribution of T_{x_c} is found from Eq. (1.40):

$$P(T_{x_c} \leq T) = \exp\left[-\frac{p_{\text{fix}} N_0}{\frac{1-x_c}{x_c}\exp\left(\beta(T)\right)+1}\right]. \tag{1.44}$$

Eq. (1.44) constitutes the main result of this section.

$\beta(t)$ is a function of $s(t)$, and p_{fix} can be expressed in terms of $s(t)$ and $N_e(t)$ (Eq. 1.16b). Like the fixation probability, the distribution of T_{x_c} therefore depends on two parameters that summarize the time-dependence of the model. For the fixation probability, we have seen in Eq. (1.16a) that there is a second way to summarize the dynamics in terms of the absolute rate of increase $(s+r)(t)$ and the total rate of birth and death events. This is not possible for the distribution of T_{x_c}, linked to the fact that variable selection and demographic change are not equivalent in this case. While the selection function $s(t)$ has a strong influence on the deterministic part of the frequency path $x(t)$, changes in the total population size (with equal effect on mutants and residents) have no influence on the latter.

To calculate moments of the distribution, we perform the parameter transformation

$$T_{x_c} \to Y := \frac{p_{\text{fix}} N_0}{\frac{1-x_c}{x_c}\exp\left(\beta(T_{x_c})\right)+1}. \tag{1.45}$$

Solving for T_{x_c} yields

$$T_{x_c} = \beta^{-1}\left(\ln\left[\frac{x_c}{1-x_c}\left(\frac{p_{\text{fix}}N_0 - Y}{Y}\right)\right]\right) \tag{1.46}$$

and we obtain for the moments

$$\langle T_{x_c}^n \rangle = \int_0^{p_{\text{fix}}N_0 x_c} \left[\beta^{-1}\left(\ln\left[\frac{x_c}{1-x_c}\left(\frac{p_{\text{fix}}N_0 - Y}{Y}\right)\right]\right)\right]^n \exp\left(-Y\right)\mathrm{d}Y. \tag{1.47}$$

For constant selection and therefore $\beta(t) = st$ we obtain for the mean of T_{x_c}:

$$\begin{aligned}\langle T_{x_c}\rangle = \frac{1}{s}\Bigg[&\ln\left[p_{\text{fix}}N_0\right] + \gamma + \ln\left[\frac{x_c}{1-x_c}\right] - \frac{1}{p_{\text{fix}}N_0} - \left(\frac{1}{p_{\text{fix}}N_0}\right)^2 \\ &+ \mathcal{O}\left(\left(\frac{1}{p_{\text{fix}}N_0}\right)^3\right)\Bigg].\end{aligned} \tag{1.48}$$

1.4.2 Applications

Constant selection and constant population size. For the distribution of the passage time T_{x_c}, we obtain with $p_{\text{fix}} = s/(\xi + s) \approx 2s\frac{N_e}{N}$ (see Eq. (1.17))

$$P(T_{x_c} \leq T) = \exp\left[-\frac{s}{\xi + s}\frac{N}{\frac{1-x_c}{x_c}\mathrm{e}^{sT} + 1}\right] \approx \exp\left[-\frac{2sN_e}{\frac{1-x_c}{x_c}\mathrm{e}^{sT} + 1}\right]. \tag{1.49}$$

Plots of the probability density for $x_c = 0.5$ and various values of s are depicted in Figure 1.4A. With increasing selection strength the distribution gets narrower and is shifted to the left.

In the particular situation of constant selection and population size, additional results can be obtained. First, we note that T_{x_c} can also be interpreted as the age of a derived allele that is currently found at frequency x_c in the population. This is a consequence of the time-homogeneity of the model. Second, the distribution of the time to reach fixation at $x = 1$ starting from a frequency $(1 - x_c)$ equals the distribution of T_{x_c} (see Appendix D for an explanation). In particular, the times needed from 0 to 0.5 and from 0.5 to 1 follow the same distribution. Since both random variables are independent, we obtain the probability density $\tilde{p}(t_{\text{fix}})$ of the whole fixation time from the density $p(T_{1/2})$ as

$$\tilde{p}(t_{\text{fix}}) = \int_0^{t_{\text{fix}}} p(T_{1/2})p(t_{\text{fix}} - T_{1/2})\mathrm{d}T_{1/2}. \tag{1.50}$$

An alternative expression for the distribution of t_{fix} has been derived before by Wang and Rannala (2004), using diffusion techniques. We note that our approximation Eq. (1.50) is simpler than the series expansion in terms of Gegenbauer polynomials using the eigenvalues of the oblate spheroidal angular function and the intermediate coefficients of the spheroidal harmonics obtained by these authors. As discussed in Appendix B, it is nevertheless highly accurate if selection is not very weak or the population size small. The cumulants of the fixation time are just two times the cumulants of the time to reach frequency 0.5. For the mean fixation time, in particular, we obtain:

$$\langle t_{\text{fix}} \rangle = \frac{2}{s}\left(\ln\left(\frac{sN}{\xi + s}\right) + \gamma - \frac{\xi + s}{sN} - \frac{(\xi + s)^2}{(sN)^2} + \mathcal{O}\left(\frac{(\xi + s)^3}{(sN)^3}\right)\right), \tag{1.51}$$

where $\gamma \approx 0.577$ is the Euler constant.

For the higher cumulants, approximations of leading order in sN can be obtained by the approximation $\exp(sT)+1 \approx \exp(sT)$ in the distribution function Eq. (1.49) and by extending the integral Eq. (1.47) to infinity. We state the results for the variance, the skewness and the kurtosis:

$$\kappa_2 = \langle t_{\text{fix}}^2 \rangle - \langle t_{\text{fix}} \rangle^2 \approx \frac{1}{3}\frac{\pi^2}{s^2}, \tag{1.52a}$$

$$\kappa_3 \approx \frac{4\zeta(3)}{s^3} > 0, \tag{1.52b}$$

$$\kappa_4 \approx \frac{2}{15}\frac{\pi^4}{s^4} > 0, \tag{1.52c}$$

where $\zeta(z)$ is the Riemann Zeta function and $\zeta(3) \approx 1.202$ is the Apéry constant. It can be shown that all cumulants of order ≥ 2 are to leading order of the form

$$\kappa_j = c_j \left(\frac{1}{s}\right)^j \tag{1.53}$$

with some numerical constant c_j.

Since $s > 0$ is required for the branching approximation, deviations from the exact values must be expected for weak selection. To estimate the precision of our approximation in this case, we compare the result for the mean fixation time to the result obtained from the diffusion approximation. For this purpose we scale time with N ($t_{\text{fix}} \to \tau_{\text{fix}} = t_{\text{fix}}/N$) and take the diffusion limit (i.e., $s \to 0$, $N \to \infty$ with $\alpha := sN/\xi$ remaining constant) of Eq. (1.51). We obtain:

$$\langle \tau_{\text{fix}} \rangle = \frac{2}{\alpha}\left(\ln(\alpha) + \gamma - \frac{1}{\alpha} - \frac{1}{\alpha^2} + \mathcal{O}\left(\frac{1}{\alpha^3}\right)\right). \tag{1.54}$$

This is in perfect agreement with the approximation for the mean fixation time found in Hermisson and Pennings (2005), using the diffusion approximation. As shown in Appendix B, the deviation of our branching process approximation from the exact diffusion result is exponentially small in $\alpha/2$.

Figure 1.4 B and C show the mean time to reach a frequency x_c in dependence of x_c. Significant deviations between the theoretical predictions of Eq. (1.49) and simulation results are found only for values of x_c that are extremely close to 1 (or 0). Generally, it holds that the approximation gets less good for $x_c \gtrsim 1 - 1/\alpha$ (or $x_c \lesssim 1/\alpha$), as in this regime the deterministic frequency path Eq. (1.41), which

we used in the derivation, is not a good approximation to the real frequency path of the mutation (see EWENS, 2004, pp. 167f).

It is interesting to note that the distribution of T_{x_c} is the same for beneficial mutations with selective advantage s and deleterious mutations with selective disadvantage $-s$ (cf. EWENS, 2004): for every point of a particular path leading from $n = 1$ to $n = n_c$, the sum of the birth and the death rate is the same for both scenarios. This implies that the distribution of the run-time of this particular path is the same. The chance to be realized differs for each such path by a constant factor $(\xi+s)^{-n_c}$. If we therefore condition on $n = n_c$ being reached, the distribution of the time to reach n_c, is the same for beneficial and deleterious mutations.

Constant selection and changing population size. Using the expression for p_{fix} calculated for a logistically growing population, we immediately get the distribution of the passage time T_{x_c}. Figure E.6 shows a comparison to simulation results for the second scenario. With Eq. (1.48) we obtain the result for the mean of T_{x_c}:

$$\langle T_{x_c} \rangle \approx \frac{1}{s}\left[\ln\left(N_0 \frac{s(r+s)}{s(r+s)+(b+\xi)(s+r\gamma_0)+\gamma_0 r(r+s)} \right) + \gamma + \ln\left(\frac{x_c}{1-x_c} \right) \right]. \tag{1.55}$$

Linearly increasing selection. Putting the expression obtained for p_{fix} Eq. (1.25) into Eq. (1.44) one obtains the distribution of the time T_{x_c} at which a fraction x_c of the population consists of mutant individuals. The distribution can be used to calculate the mean of T_{x_c} numerically. A comparison to simulation results for the distribution and the mean of T_{x_c} in dependence of s_1 can be found in Figure E.7. From Figure 1.5, it can be seen that for small values of s_0 even a very slight increase of the selection strength over time leads to a drastically smaller mean value of T_{x_c} in comparison to an environment with constant selection $s = s_0$.

Periodically changing selection. Figure 1.6 shows the mean time to reach frequency $x_c = 0.5$ for a periodic selection coefficient $s(t) = s_0(1 + \cos(\omega t + \varphi))$. As for the fixation probability, the initial selection strength $s(0)$ is decisive if the environment changes very slowly; φ itself gains importance with increasing (but still small) ω. If the environment changes fast, $\langle T_{x_c} \rangle$ becomes independent of φ

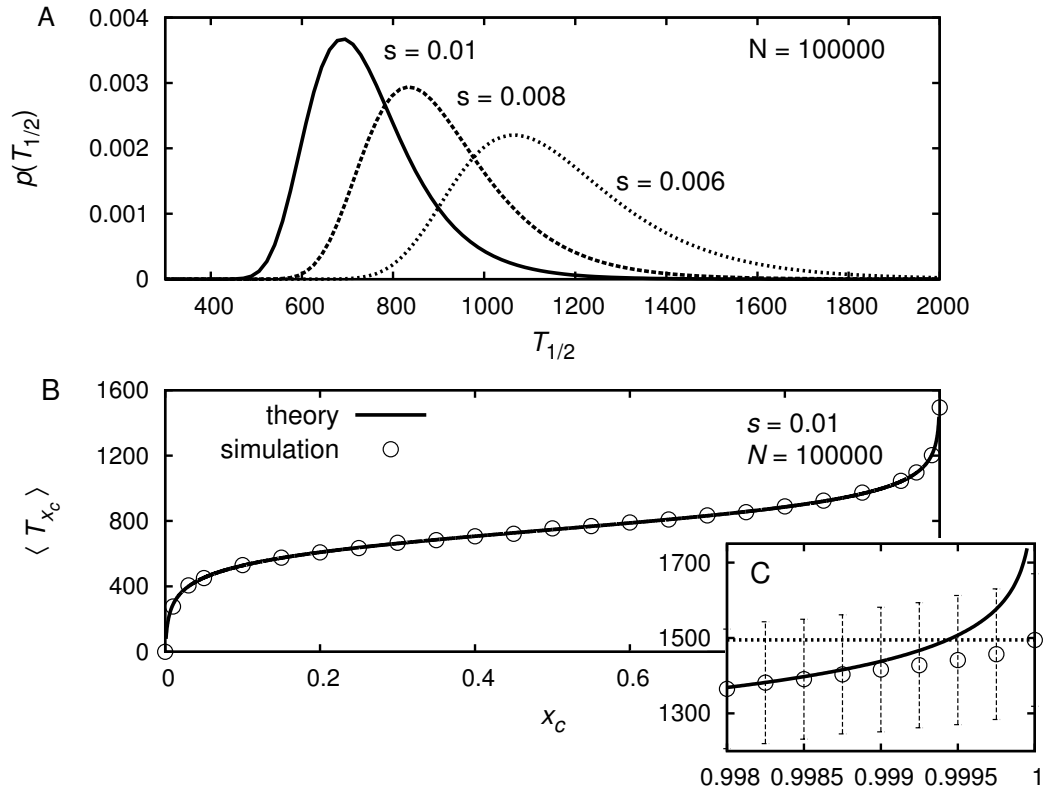


Figure 1.4: A: Probability density of the time to reach frequency $x_c = 0.5$ in the case of constant selection and constant population size. With decreasing selection strength, the density function becomes broader and is shifted to the right. B: Mean time to reach a frequency x_c in dependence of x_c. Theoretical results are compared to simulation results. Each simulation point is the average over 1000 runs. C: Amplification of the region of high frequencies. Simulations show the mean average passage time. One sees that the theoretical predictions gets less good for $x_c > 1 - 1/\alpha$. The dotted line is the result for the whole fixation time obtained by the convolution, i.e., two times the mean time needed to reach the frequency $x_c = 0.5$. The bars around the simulation points denote the standard deviation of T_{x_c} as obtained from the simulations. In contrast, the standard error of the simulations is much smaller and vanishes in the symbols.

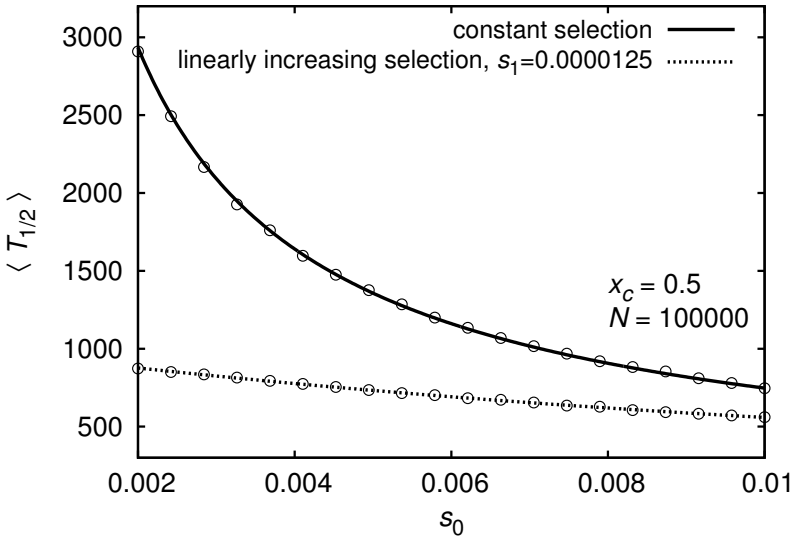


Figure 1.5: Mean time to reach frequency $x_c = 0.5$ in dependence of s_0, where a scenario with constant selection and a scenario with linearly increasing selection ($s_1 = 1/80000$) are compared. The mean fixation time is significantly smaller in the case of linearly increasing selection than in the case of constant selection. Each simulation point is the average over 1000 runs.

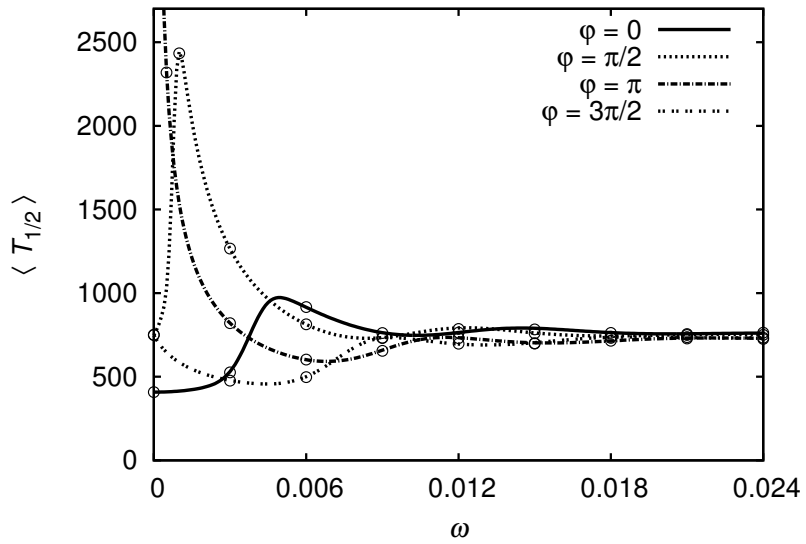


Figure 1.6: Expected value of $T_{1/2}$. Selection is periodic according to $s(t) = s_0(1 + \cos(\omega t + \varphi))$ with $s_0 = 0.01$. The total population size is $N = 100\,000$. Simulation results are averaged over 2000 runs.

and converges to its value for constant selection $s(t) = s_0$. Convergence is much faster than for the fixation probability.

1.5 Simulations

To test our analytical results, we performed individual based computer simulations for which we use a Gillespie algorithm (Gillespie, 1977). Events happen at rate $(2\xi(t, N_t) + s(t, N_t))n_t(N_t - n_t)/N_t + b(t, N_t)N_t + d(t, N_t)N_t$, where $s(t, N_t)$, $b(t, N_t)$ and $d(t, N_t)$ are assumed to be constant between events. Once the time of an event is fixed, which kind of event takes place is determined using the respective probabilities. Subsequently, the values of N_t and n_t are updated and the rates are set to their new values (additional update steps between events did not change the results).

For most of the simulation runs, we used the first passage time of a given frequency x_c to determine T_{x_c}, i.e., we neglected fluctuations around x_c, which for large value of α and not too small or large values of x_c has no significant effect. For the data shown in panel C of Figure 1.4 and in Figures B.3 and B.4, where we

pushed the boundaries of the theory, we took the average over all passage times, more precisely the average over all times at which the path crossed the frequency x_c from lower to higher values (cf. section on the analytical theory).

For all simulation results, the number of runs was chosen sufficiently large that the standard error bars vanish in the symbols. All programs were written in C, making use of the Gnu Scientific Library (Galassi *et al.*, 2009). We used Mathematica (Wolfram Research, Champaign, USA) for all numerical evaluation of integrals.

1.6 Discussion

Adaptation is the evolutionary response of a population to an environmental challenge: variation in the environmental conditions leads to altered selection pressures and changing population sizes. In nature, these changes occur on all time scales, from rapid shifts within a single generation to long-term geological trends. It therefore seems natural that models on the genetics of adaptation, too, should account for the ecological dynamics that drives the process. In contrast to this expectation, however, the vast majority of published studies with a genetic focus assumes a constant environment (reviewed, e.g., in Orr, 2010). They rely on the idea that a fast - almost instantaneous - change in the environment is followed by a period of environmental stasis. If this period is long compared to the total time it takes for one or several beneficial mutations to appear and to rise to fixation, ecology and evolution are effectively decoupled. In many cases of ecological interest, however, this separation of time scales is not appropriate. In this case, the parameters of the evolutionary model (such as selection coefficients or population sizes) are turned into time-dependent variables, which must be determined from an underlying ecological model.

In a series of recent publications by several authors, the impact of the ecological dynamics on the genetics of adaptation has been studied for the so-called moving optimum model (Bello and Waxman, 2006; Collins *et al.*, 2007; Kopp and Hermisson, 2007, 2009a,b). The consensus, at least for this model, is that the adaptive process is strongly affected by the dynamics of the selective environment. In this article, we present a more detailed treatment of the most basic aspect of the genetics of adaptation: the fixation process of a single beneficial mutation in a variable environment. All relevant parameters of the process, i.e., selection coefficient, population size, and the variance in offspring number (or, equivalently, the birth and death rates of wildtypes and mutants) may depend

explicitly on time. An example of how this can result from an explicit ecological model is given in the Appendix, where we discuss the fixation probability of a "rescue mutation" in a population that is otherwise doomed to extinction. Again, the results are strongly influenced by the ecological dynamics, leading to qualitative differences relative to previous studies that assume constant selection (Orr and Uncкless, 2008).

Even if the external environment is constant, individual alleles in a population can experience variable selection pressures if multiple selected alleles segregate in the population and if these alleles interfere due to either epistasis or linkage (cf. Hartfield and Otto, 2011). Examples include the evolution of compensatory mutations or classical problems of clonal interference, where the beneficial mutation rate is high relative to the recombination rate. Another potential application is adaptive gene-flow across a genetic barrier, where adaptations need to be purged from linked deleterious alleles by recombination.

Using branching process techniques, we obtain analytical approximations for the fixation probability and the distribution of the time for the mutant to reach a given frequency x_c in the population.

Fixation probability. The derivation of fixation probabilities of rare beneficial mutations from a supercritical branching process is a standard approach in population genetics. In particular, results have previously been obtained for the most important modes of demographic changes (e.g., Otto and Whitlock, 1997; Pollak, 2000). A look into the mathematical literature reveals, however, that a general formalism with arbitrary time-dependent birth and death rates has been available since the work of Kendall (1948). In contrast to most of the previous studies in population genetics, which use a branching process in discrete time (following Haldane, 1927), this approach is based on a continuous-time framework, which simplifies some technical aspects. Our adaptation of this formalism to the genetic context leads to a compact, yet general formula for the fixation probability p_{fix} (Eq. (1.16a) and (1.16b)) that covers previous results as special cases. In Appendix C, we show how an analogous result can also be obtained using the diffusion approach.

It turns out that the ecological dynamics affect p_{fix} through two independent variables. In two alternative formulations of the result, these can either be the time-dependent birth- and death rates of rare mutants (Eq. (1.16a)), or the selection coefficient $s(t)$ and the variance effective population size $N_e(t)$ (Eq. (1.16b)). Our applications to various scenarios show that relative to the case with constant

selection s_0, consistent changes in the selection coefficient Δs per generation of the order of $\Delta s \geq s_0^2$ have a strong effect on p_{fix}. This is to be expected since, for constant selection, the fate of a new mutation (fixation or loss) is decided once the mutant allele reaches a frequency on the order of $(Ns_0)^{-1}$. According to Eq. (1.48), this will take, on average, in the order of s_0^{-1} generations. The observation is of practical importance since it shows that predictions about fixation probabilities cannot be based on short-term fitness assays. Unmeasurably small fitness changes among generations may have a large effect. The limitations of the branching approach are the usual ones: The fate of the mutation must be decided while the mutant frequency is small and the independence assumption of the branching process applies. Deviations from the simulation results are found in the case of mutations that are almost neutral on average, with fixation probabilities $p_{\text{fix}} \lesssim 10/N$ (cf. Appendix B).

Time T_{x_c} to reach frequency x_c. The usual method in population genetics to derive fixation times or (more generally) first-passage times for mutant alleles to reach certain frequencies is diffusion theory. Since this proves difficult for the time-inhomogeneous case, we again turn to a branching process approach. We can use the fact that (almost) all stochasticity of a mutant trajectory $x(t)$ is due to its early phase while the mutant number is still small. We can therefore (approximately) describe this stochasticity in the branching framework and combine it with the deterministic growth of the full process. For constant selection and population sizes, this idea has previously been used by Desai and Fisher (2007) in the context of a model for clonal evolution.

For the general case, the crucial step of the method has again been anticipated in the mathematical literature (Cohn and Jagers, 1994). There, it has been shown that a clean separation exists for the random variable n_t of a inhomogeneous supercritical branching processes into a growth function that describes the asymptotic growth and a time-constant random variable ν that describes the stochastic fluctuations in the large-time limit. We can interpret ν as the "effective initial size" of the mutant population. It turns out that the distribution function of ν, $P(\nu \leq \nu_0|\text{not extinct}) = 1 - \exp(-p_{\text{fix}}\nu_0)$ (Eq. 1.40), is pleasingly simple even in a variable environment. In particular, the impact of the ecological dynamics is conveniently summarized in the fixation probability p_{fix}. When this initial size is combined with the deterministic growth of the full model, an approximation for the distribution function for T_{x_c} is obtained by a simple transformation of the probability density Eq. (1.44).

Computer simulations show that the results from the branching approach are usually highly accurate as long as selection is not very weak. This is also confirmed by a comparison with previous results for the expected value of the fixation time T_1 for constant s and N (Kimura and Ohta, 1969; Ewens, 2004) derived from the diffusion approximation: the order of the error term for the average fixation time is $o(\exp(-Ns/2))$ (see Appendix C). Previous results also exist for the entire probability density of T_1 in a constant environment (Wang and Rannala, 2004). Here, the branching approximation leads to a much simpler analytical result.

As a major caveat of the method, we note that for variable s or N, we do not obtain a closed expression for the time to reach full fixation $x_c = 1$. More generally, the approximation is accurate only as long as $1/(sN) \lesssim x_c \lesssim 1 - 1/(sN)$. For x_c close to 0, estimates can easily be found directly from the conditioned branching process, if needed. For mutant frequencies near 1, the fixation process enters another stochastic phase when the wildtype alleles become rare. Although this phase can be described, in principle, as a subcritical branching process, its details depend on the random time when this stochastic phase is entered, which complicates the treatment. In natural systems, an allele frequency of $x_c = 1$ is not an absorbing state and may never be reached in the face of back mutation, migration, or ongoing adaptation. For many practical applications, other thresholds (such as 5%, 50%, or 95%) are therefore more relevant and have previously been used. For example, Desai and Fisher (2007) use a threshold of $x_c = 1/(Ns)$ to characterize their establishment time τ_{est}. Kopp and Hermisson (2007, 2009a) use $x_c = 0.5$ as the critical value where the mutant comes to dominate the population in an analysis of the order and step sizes of adaptive substitutions.

Although we have formulated our results for haploids, they also apply to diploids as long as the mutant is not fully recessive (i.e., hs must be sufficiently large on average). Only the selection coefficients (or birth and death rates) of heterozygotes enter into the stochastic part of the model described by the branching process. In contrast, the deterministic frequency path used in the derivation of T_{x_c} depends on the fitness values of both heterozygous and homozygous mutants. In the purely recessive case, the dynamics of rare mutants can no longer be described by a supercritical branching process. Diffusion methods (for constant selection) show that also the scaling of the fixation time with the selection parameters is altered in this case (Ewing *et al.*, 2011).

To summarize, our results show that inhomogeneous branching processes provide a powerful framework to describe the fixation process under a wide range of

ecological scenarios and genetic conditions. We therefore hope that the methods and results can provide a step to combine ecological and genetic modelling traditions to study the genetics of adaptation under realistic ecological conditions.

Note: After acceptance of this article we became aware of a recent preprint by Waxman on the fixation process under variable selection and demography. Both articles are complementary: While our approach focuses on analytical approximations for the fixation process of a definitely beneficial mutant, Waxman (2011) presents numerical methods to derive fixation probabilities for alleles with arbitrary selection coefficients.

Acknowledgements

We thank Nick Barton, Ulf Dieckmann, Peter Pfaffelhuber and Matthew Hartfield for helpful discussions and Lindi Wahl and two anonymous referees for many useful comments. This work was made possible with financial support by the Vienna Science and Technology Fund (WWTF), and by the Deutsche Forschungsgemeinschaft (DFG), Research Unit 1078 *Natural selection in structured populations*.

A Appendix: Fixation in general ecological models

In this Appendix, we describe how the branching process approach can be applied to a general ecological scenario where the genetic composition of the population and the population dynamics are mutually dependent. Assume that at time t a population consists of w_t wild-type "residents" and n_t mutants. Both types reproduce and die at different per-capita rates, which may depend on the number of residents and mutants, w_t and n_t, and on external factors that are independent of the population, but also may change over time. A general framework, which

takes the full dynamics with all kinds of interactions between types into account, is given by the following scheme of transition rates:

$$\begin{aligned} n_t \to n_t + 1: &\quad n_t[b_1^{(m)}(t) + b_2^{(m)}(t) n_t + b_3^{(m)}(t) w_t + \dots], \\ n_t \to n_t - 1: &\quad n_t[d_1^{(m)}(t) + d_2^{(m)}(t) n_t + d_3^{(m)}(t) w_t + \dots], \\ w_t \to w_t + 1: &\quad w_t[b_1^{(w)}(t) + b_2^{(w)}(t) w_t + b_3^{(w)}(t) n_t + \dots], \\ w_t \to w_t - 1: &\quad w_t[d_1^{(w)}(t) + d_2^{(w)}(t) w_t + d_3^{(w)}(t) n_t + \dots]. \end{aligned} \tag{A.1}$$

Higher order interaction terms with transition rates proportional to any polynomial in n_t and w_t can be added as needed. The model corresponds to a general two-dimensional birth-death process with time-dependent coefficients. For such a model, the corresponding deterministic evolution of n_t and w_t is generically given by a system of two coupled differential equations. In most cases, it is impossible to solve this system analytically. Since the determinstic frequency path of the mutant is needed for the theory developed to calculate T_{x_c}, derivations for T_{x_c} must resort to numerical solutions to apply the methods outlined in the main text. In contrast, an analytical approximation for the fixation probability (and thus also the distribution of the effective initial population size ν, cf. Eq. (1.40)) is often still possible.

In the branching limit of rare mutant individuals, all interactions of either residents or mutants with (other) mutant individuals can be neglected. Mathematically, this corresponds to the approximation $n_t \approx 0$ in the birth and death rates of the residents, and negligence of all terms of order $\mathcal{O}(n_t) > 1$ in the transition rates of the mutants. As a consequence, the evolution of the residents is decoupled from the mutants. In the deterministic limit, the time-development of the resident population size $w(t) \approx N(t)$ is then given by an ordinary differential equation that can frequently be solved explicitly. We can then describe the dynamics of rare mutants by a one-dimensional branching process. Inserting the deterministic solution for $w(t)$ into all mutant transition rates of linear order in the mutant number n, the theory developed in the paper can be applied by choosing $\lambda(t)$ and $\mu(t)$ appropriately. Note, however, that the first expression for p_{fix} Eq. (1.16a) must be used, since the step leading from Eq. (1.16a) to Eq. (1.16b) is not possible if the growth rates of mutants and wildtypes are not the same. Previous results on fixation probabilities in two-dimensional birth-death chains (with constant coefficients) have been obtained by Parsons and Quince (2007a). Their approach is based on singular pertubation methods and

has the advantage that it can be applied to neutral and deleterious mutations, too (Parsons and Quince, 2007a,b), which is beyond the reach of the branching model. However, it cannot easily be extended to the time-inhomogeneous case that is our focus here.

Rescue mutation. Consider, as an example, the following scenario: a population is subject to a new and severe selection pressure (e.g., new insecticide, drug, parasite, competitor ...). As a consequence, the population size decreases, and selection will drive the population to extinction unless a "rescue mutation" can be established that confers immunity to its carriers. A similar scenario has been discussed by Orr and Uncкless (2008).

A plausible ecological model for this situation could be as follows: the original resident population evolves under a logistic population dynamics with growth parameter r and time-dependent carrying capacity $K(t)$. Due to the environmental challenge, the carrying capacity decreases exponentially like $K(t) = K_0 \exp(-at)$. At the individual level, the model can be described in the scheme of (A.1),

$$b_1^{(w)}(t) = r, \qquad d_2^{(w)}(t) = d_3^{(w)}(t) = r\frac{1}{K(t)}, \tag{A.2}$$

with all other coefficients $b_i^{(w)}(t)$ and $d_i^{(w)}(t)$ equal to zero. As long as the number of beneficial mutants is small, the wildtype population size changes according to

$$w(t) = \frac{K_0(a+r)}{a\exp(-rt) + r\exp(at)}. \tag{A.3}$$

Assume now that a beneficial mutation – if it survives stochastic loss – restores population size to its original carrying capacity $K_0 = K(0)$, i.e.,

$$b_1^{(m)}(t) = r, \qquad d_2^{(m)}(t) = d_3^{(m)}(t) = r\frac{1}{K_0} \tag{A.4}$$

(and again all other coefficients are equal to zero). Then the fixation probability of a single mutation arising τ time units after the decline in the carrying capacity started is then given by:

$$p_{\text{fix}} = 2\left[1 + \int_0^\infty r\left(1 + \frac{w(\tau+t)}{K_0}\right)\exp\left[-\int_0^t r\left(1 - \frac{w(\tau+x)}{K_0}\right)\mathrm{d}x\right]\mathrm{d}t\right]^{-1}. \tag{A.5}$$

In Figure A.1 the fixation probability of a single mutation and the ratio $\gamma_0(\tau) = w(\tau)/K_0 = N(\tau)/K_0$ are shown in dependence of τ. Note that the fixation probability increases with τ: While the mutant does not have a higher intrinsic growth rate than the resident (in a variant of the model, it could even be lower), it thrives due to relaxed competition as the resident population declines. However, for a given mutation probability per time unit u from resident to mutant, the new total rate of beneficial mutations is proportional to the resident population size and thus decreases with τ. The rate at which new successful mutations arise is given by $\theta_{\text{fix}}(\tau) = uw(\tau)p_{\text{fix}}(\tau)$, which reaches a maximum at some intermediate time τ^* (see Figure A.1). The probability that a successful mutation arises before time T

$$P(t \leq T) = 1 - \exp\left[-\int_0^T \theta_{\text{fix}}(\tau)\mathrm{d}\tau\right]. \tag{A.6}$$

The probability that a successful mutation arises at all, i.e., that the population is rescued from extinction, is therefore given by

$$P_{\text{rescue}} = 1 - \exp\left[-\int_0^\infty \theta_{\text{fix}}(\tau)\mathrm{d}\tau\right]. \tag{A.7}$$

This probability is equal to 1 if and only if the integral $\int_0^\infty \theta_{\text{fix}}(\tau)\mathrm{d}\tau$ diverges, i.e., if the rate $\theta_{\text{fix}}(\tau)$ declines sufficiently slowly. Otherwise the probability that the population is saved from extinction by a rescue mutation is smaller than 1, as was also found by Orr and Uncкless (2008). For the parameter values used in Figure A.1, the survival probability of the population is calculated to be $P_{\text{rescue}} \approx 0.39$. Given the population is rescued, the probability that the rescuing

mutation arises in the infinitesimal time interval $[\tau, \tau + \mathrm{d}\tau]$ is given by $p(\tau)\mathrm{d}\tau$ with

$$p(\tau) = \frac{\theta_{\mathrm{fix}}(\tau) \exp\left[-\int_0^\tau \theta_{\mathrm{fix}}(\hat{\tau})\mathrm{d}\hat{\tau}\right]}{P_{\mathrm{rescue}}}. \tag{A.8}$$

For not too large mutation probabilities ($u \lesssim 0.5/K_0$ for the chosen parameter values), this function has a maximum at an intermediate value of $\tau = \tau^{**}$, as also shown in Figure A.1. This result is in contrast to Orr and Uncкless (2008), who find a monotonic decrease in the (conditioned) probability that a successful mutation arises in a given generation independently of the mutation rate. This is due to the assumption in Orr and Uncкless (2008) that the selection coefficient of mutants (and thus the fixation probability p_{fix}) of a single mutant is constant over time. In our model, selection pressure on mutants is not simply included by assumption, but results from the underlying ecological dynamics. As such, it explicitly depends on time. The scenario by Orr and Uncкless (2008) is approximately met only for large $\tau > \tau^{**}$ when competition from the resident population gets very small.

B Appendix: Accuracy of the approximation

Weak selection. As pointed out in the main text, deviations from the exact solution are expected for weak selection.

For an allele with constant selective advantage and $N = const.$, a comparison to the exact solution is immediately possible (see main text).

As a further example, we consider the fixation probability of an allele with $s(t) = s_1 t$ and examine the relative error between theoretical prediction and simulation results in dependence of $p_{\mathrm{fix}}N$ (see Figure B.2). A relative error of less than 2% is found for $p_{\mathrm{fix}}N < 10$. The results for small values of $p_{\mathrm{fix}}N$ can be improved if we correct the branching approximation by a factor $1/(1 - \exp(-p_{\mathrm{fix}}N))$:

$$p^*_{\mathrm{fix}} = \frac{p_{\mathrm{fix}}}{1 - \exp(-p_{\mathrm{fix}}N)}. \tag{B.9}$$

This heuristic approximation is inspired by Eq. (1.18). The relative deviation from the simulation results is added to Figure B.2. It is seen that the approximation is considerably improved.

In Figure B.3, we compare the distribution function of $T_{1/2}$ to simulation results for various values of s and N where s and N are constant. While agreement

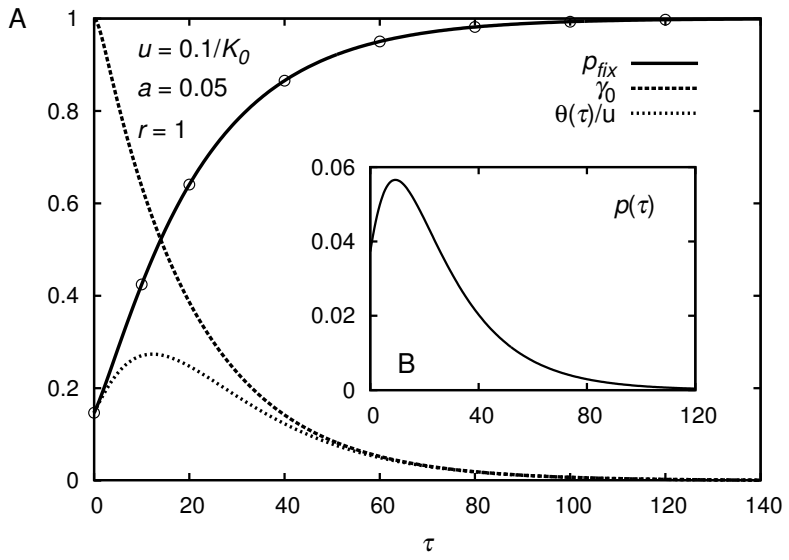


Figure A.1: A: Fixation probability $p_{\text{fix}}(\tau)$ of a single rescue mutation, the ratio $\gamma_0(\tau) = N(\tau)/K_0$ and the product of the two, which characterizes the rate of successful mutants, in dependence of the time τ since the onset of the selection pressure. B: Density function $p(\tau)$ which determines the probability that the successful rescue mutation arises in the time interval $[\tau, \tau + \mathrm{d}\tau]$ conditioned on population survival. While the fixation probability per mutant increases with τ, the latter peaks at an intermediate value for not too large mutation probabilities u. Simulations were performed for a population of 100 000 individuals, and every simulation point is the average ove 10^6 runs.

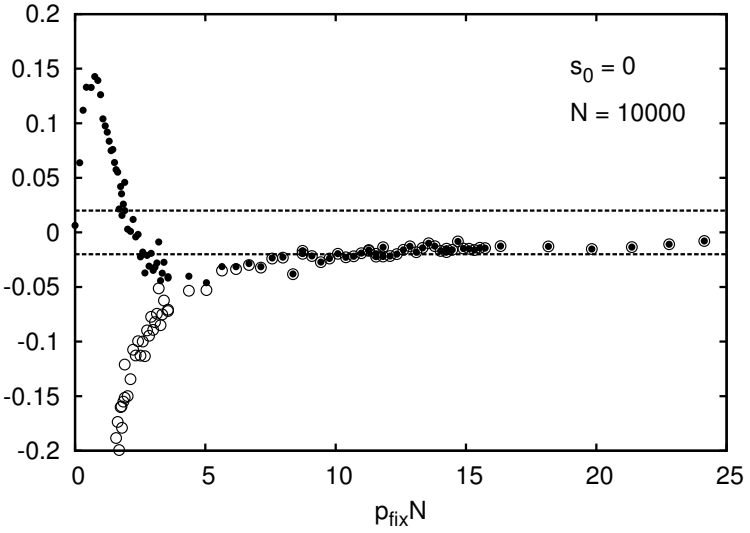


Figure B.2: Relative deviation between analytical and simulation results for an allele with selective advantage $s(t) = s_1 t$. The empty dots are obtained from branching process approximation, the small filled dots from the corrected version Eq. (B.9). One sees that the corrected version provides a much better approximation for small values of $p_{\text{fix}}N$; for large values of $p_{\text{fix}}N$, the results coincide. Simulations were performed for a population of $N = 10\,000$ individuals, and each simulation point is the average over $5 \cdot 10^7$ runs.

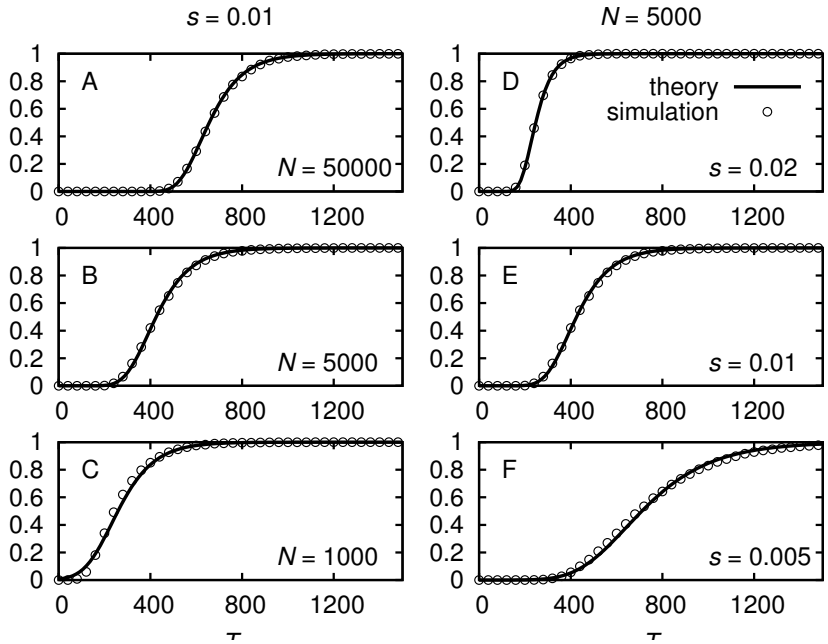


Figure B.3: Distribution of the time to reach frequency $x_c = 0.5$ for various values of s and N, where s and N are constant. Analytical and simulation results are compared. Simulation results are averaged over 5000 runs. A maximum absolute deviation of ≥ 0.05 between theory and simulations is found for $\alpha = Ns \lesssim 10$ (see panel C).

is excellent for high values of α, deviations increase for decreasing values of α. A maximum absolute deviation of ≥ 0.05 between theory and simulations is found for $\alpha = Ns \lesssim 10$.

Mode of density-regulation. To test whether the mode of density-regulation influences the outcome, some exemplary simulations for three different scenarios of density regulation were performed ($b(t, N_t) = d(t, N_t) = \text{const.}$, Lotka-Volterra dynamics, i.e., $b(t, N_t) = b + \rho(1 - N_t/K)$ and $d(t, N_t) = b$, and $b(t, N_t) = d(t, N_t) = 0$ with accordingly chosen ξ). Figure B.4 shows a comparison of the simulation results to the theoretical curve. We see that demographic stochasticity has only a slight effect (note that we used a relatively low (initial) population size of $N_0 = 2000$ for the examples to make demographic stochasticity strong).

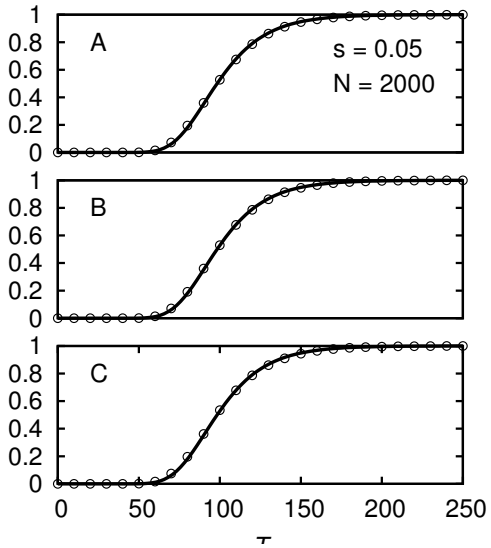


Figure B.4: Distribution function of $T_{1/2}$ for three different modes of density regulation. The selection strength is $s = 0.05$ and the initial population size $N_0 = 2000$. A: Inherently constant population size, i.e. $b(t, N_t) = d(t, N_t) = 0$, with $\xi(t, N) = 1$. B: Lotka-Volterra dynam ics, i.e. $b(t, N_t) = b + \rho(1 - N_t/K)$ and $d(t, N_t) = b$, with $b = 1.0$, $r = 1.0$ and $K = N_0 = 2000$. C: Freely fluctuating population size, i.e. $b(t, N_t) = d(t, N_t) = b$, with $b = 1.0$. Simulations are averaged over 50 000 runs. The maximum absolute deviation is of the same order of magnitude for all three scenarios.

C Appendix: Diffusion approximation

Pioneered by Kimura (1957), the diffusion approximation has been established as the second main approach for the analytical analysis of the fixation process. In this appendix, we show how an approximation for the fixation probability with variable selection and population size can be derived within the diffusion framework. We also use diffusion results for the mean fixation time for constant s and N to estimate the error of the branching process approach in this case.

As in the main text, we assume that the population inhabits a variable environment. All key model parameters may explicitly depend on time. In particular, let $s = s(t)$ be the selection coefficient of the A allele and $N = N(t)$ the total population size. Finally, the variance in offspring number is $\sigma^2 = \sigma^2(t)$. Following standard practice, we can combine N and σ^2 in the variance effective population size, $N_e(t) = N(t)/\sigma^2(t)$.

We can formulate the model as a diffusion process as follows. Let $g_t(x \to x+\epsilon)$ be the transition probability that the allele frequency changes from x at time t to $x + \epsilon$ at time $t + \mathrm{d}t$. The diffusion is characterized by the mean and the variance of the transition probability in the limit of small dt and ϵ:

$$E_{x,t}[\epsilon]dt = \int_{-\infty}^{\infty} \epsilon g_t(x \to x + \epsilon)\mathrm{d}\epsilon, \tag{C.10a}$$

$$\mathrm{Var}_{x,t}[\epsilon]\mathrm{d}t = \int_{-\infty}^{\infty} \epsilon^2 g_t(x \to x + \epsilon)\mathrm{d}\epsilon + \mathcal{O}[\mathrm{d}t^2]. \tag{C.10b}$$

To leading order in N_e^{-1},

$$E_{x,t}[\epsilon] = s(t)x(1-x), \tag{C.11a}$$

$$\mathrm{Var}_{x,t}[\epsilon] = \frac{\sigma^2(t)}{N(t)}x(1-x) = \frac{x(1-x)}{N_e(t)}. \tag{C.11b}$$

Define $f(x, \tau|p, t)$ as the probability density for an allele frequency of x at time $t + \tau$, given that the frequency at time t was p. Then

$$f(x, \tau + \mathrm{d}t|p, t) = \int_{-\infty}^{\infty} g_t(p \to p + \epsilon) f(x, \tau|p + \epsilon, t + \mathrm{d}t)\mathrm{d}\epsilon$$
$$\approx f(x, \tau|p, t + \mathrm{d}t) + \Big(\frac{\partial f(x, \tau|p, t + \mathrm{d}t)}{\partial p} E_{p,t}[\epsilon] + \frac{1}{2}\frac{\partial^2 f(x, \tau|p, t + \mathrm{d}t)}{\partial p^2}\mathrm{Var}_{p,t}[\epsilon]\Big)\mathrm{d}t. \tag{C.12}$$

In the limit $\mathrm{d}t \to 0$, we obtain the Kolmogorov backward equation for the conditional density $f(x, \tau|p, t)$ of a time-inhomogeneous diffusion:

$$\frac{\partial f}{\partial \tau} - \frac{\partial f}{\partial t} = s(t)p(1 - p)\frac{\partial f}{\partial p} + \frac{p(1 - p)}{2N_e(t)}\frac{\partial^2 f}{\partial p^2}. \tag{C.13}$$

If $s(t)$ and $N_e(t)$ are both constant, $\partial f/\partial t = 0$ and (C.13) reduces to the backward equation of the classical homogeneous case. With time dependence, the state of the population can no longer be described by the allele frequency x alone; f depends on both, x and t. Note that we measure time on a generations scale, not in $2N$ generations, as it is often done in diffusion theory.

Denote now as $P(p, t)$ the fixation probability of an allele with frequency p in the population at time t. Since fixation is the probability of eventual absorption of the diffusion at the boundary $x = 1$, we can write $P(p, t)$ in terms of the transition probability as

$$P(p, t) = \lim_{\substack{\tau \to \infty \\ y \to 1}} \int_y^1 f(x, \tau|p, t)\mathrm{d}x. \tag{C.14}$$

Integrating (C.13) over any frequency interval $[x_1, x_2]$, we see that the Kolmogorov backward equation also holds for the corresponding probabilities. For the probability of eventual fixation, in particular, the τ dependence vanishes and we obtain (Kimura and Ohta, 1974)

$$-\dot{P}(p, t) = s(t)p(1 - p)P'(p, t) + \frac{p(1 - p)}{2N_e(t)}P''(p, t), \tag{C.15}$$

where $\dot{P}$ and P' denote derivatives with respect to t and p, respectively. In contrast to the time-homogeneous case, there is no exact solution to the time-inhomogeneous equation. However, it is possible to derive an approximate solu-

tion for small p by setting $1-p \approx 1$ in (C.15). This approximation was first used by KIMURA and OHTA (1974) to derive the fixation probability for a logistically growing population. As we will see, a full solution of the general time-dependent model is possible under the same assumption. We use the following Ansatz,

$$P(p,t) = 1 - \exp\left(-Q(t)p\right). \tag{C.16}$$

Substituting this relation into the approximated PDE (ignoring $1-p$ terms in Eq. (C.15)) leads to the following ODE for $Q(t)$,

$$\frac{\mathrm{d}Q(t)}{\mathrm{d}t} = -s(t)Q(t) + \frac{Q^2(t)}{2N_e(t)}. \tag{C.17}$$

A general solution to this differential equation can be found,

$$Q(t) = \frac{2}{\int\limits_t^\infty \frac{1}{N_e(\tau)} \exp\left(-\int\limits_t^\tau s(y)dy\right) \mathrm{d}\tau}, \tag{C.18}$$

and we obtain an approximation for the fixation probability from Eq. (C.16). If the fixation process starts from a single copy of the derived allele that enters the population at time $t = 0$, in particular, we find

$$p_{\text{fix,diff}} := P(1/N(0), 0) = 1 - \exp\left(-Q(0)/N(0)\right) \approx Q(0)/N(0). \tag{C.19}$$

This expression may be directly compared to the result for the fixation probability p_{fix} (1.16b) from the branching process derivation. We find that

$$p_{\text{fix}} = \frac{p_{\text{fix,diff}}}{1 + p_{\text{fix,diff}}/2}. \tag{C.20}$$

Both expressions are thus equal to leading order in $p_{\text{fix,diff}}$ (i.e. to leading order in the selection strength).

The approximate solution from the diffusion implies the absence of competition among mutant alleles (which appears in (C.15) through the p^2 terms). This is equivalent to the independence assumption of the branching process. Note also that the approximation is constructed in a way that it fulfills the boundary condition $P(0,t) = 0$, but not the condition $P(1,t) = 1$. It is therefore only valid if the initial frequency of the allele, p, is sufficiently small.

For constant selection, the diffusion equation (with correct boundary conditions at $p = 0$ and $p = 1$) can be solved exactly. In particular, the derivation of the mean fixation time is possible (Kimura and Ohta, 1969; Ewens, 2004). This allows for a comparison of our approximate solution with the exact diffusion result in this case. From Eq. (1.47), we obtain the following expression for the mean fixation time in the diffusion limit ($s \to 0$, $\alpha = sN/\xi = \text{const.}$):

$$\langle \tau_{\text{fix}} \rangle^{(1)} = \frac{2}{\alpha} \left(\int_0^{\alpha/2} \ln \left(\frac{\alpha - Y}{Y} \right) \exp(-Y) \mathrm{d}Y \right). \tag{C.21}$$

Within the diffusion framework, the mean fixation time is given by (see Ewens, 2004, p. 140ff and p. 167ff),

$$\begin{aligned} \langle \tau_{\text{fix}} \rangle^{(2)} &= \frac{2}{\alpha[\exp(\alpha) - 1]} \int_0^1 \frac{[\exp(\alpha x) - 1][\exp(\alpha(1-x)) - 1]}{x} \mathrm{d}x \\ &= \frac{2}{\alpha[\exp(\alpha) - 1]} \int_0^\alpha \frac{\exp(\alpha) - \exp(Y) - \exp(\alpha - Y) + 1}{Y} \mathrm{d}Y \\ &\approx \frac{2}{\alpha} \left(\int_0^\alpha \frac{1 - \exp(-Y)}{Y} \mathrm{d}Y + \exp(-\alpha) \int_0^\alpha \frac{1 - \exp(Y)}{Y} \mathrm{d}Y \right). \end{aligned} \tag{C.22}$$

In the last step, we approximated $(\exp(\alpha) - 1)^{-1} \approx \exp(-\alpha)$. This approximation is of order $o(\exp(\alpha))$. By integration by parts,

$$\int_0^\alpha \ln \left(\frac{\alpha - Y}{Y} \right) \exp(-Y) \mathrm{d}Y = \int_0^\alpha \frac{1 - \exp(-Y)}{Y} \mathrm{d}Y + \exp(-\alpha) \int_0^\alpha \frac{1 - \exp(Y)}{Y} \mathrm{d}Y. \tag{C.23}$$

We thus obtain for the error term $\Delta(\alpha)$ of the branching result:

$$\Delta(\alpha) = \langle \tau_{\text{fix}} \rangle^{(2)} - \langle \tau_{\text{fix}} \rangle^{(1)} = \frac{2}{\alpha} \int_{\alpha/2}^\alpha \ln \left(\frac{\alpha - Y}{Y} \right) \exp(-Y) \mathrm{d}Y + o(\exp(-\alpha)). \tag{C.24}$$

Since

$$\begin{aligned}\exp\left(\frac{\alpha}{2}\right)\Delta(\alpha) &= \frac{2}{\alpha}\int_{\alpha/2}^{\alpha} \ln\left(\frac{\alpha - Y}{Y}\right)\exp\left[-\left(Y-\frac{\alpha}{2}\right)\right]\mathrm{d}Y \\ &= \int_0^1 \ln\left(\frac{1-Y}{1+Y}\right)\exp\left(-\frac{\alpha}{2}Y\right)\mathrm{d}Y \xrightarrow[\alpha\to\infty]{} 0,\end{aligned} \tag{C.25}$$

we conclude that $\Delta(\alpha)$ is of order $o(\exp(-\alpha/2))$.

D Appendix: Additional explanations

Derivation of Eq. (1.4a) **and** (1.4b)**.** To calculate the expected frequency change $\mathrm{E}[\Delta x|x_t, N_t]$ in an infinitesimal time interval $\mathrm{d}t$, we need to consider all events that change the frequency $x_t = n/N_t$ of mutants in the population. These events are summarized in Table 1.1. The second and third column give the probability of the event and the change Δx induced by it. Taking all events, their respective probabilities and effects in account, we obtain:

$$\begin{aligned}&\mathrm{E}[\Delta x|x_t, N_t] \\ &= \Bigg[(\xi(t,N_t)+s(t,N_t))\frac{n(N_t-n)}{N_t}\left(\frac{n+1}{N_t}-\frac{n}{N_t}\right) \\ &\quad + \xi(t,N_t)\frac{n(N_t-n)}{N_t}\left(\frac{n-1}{N_t}-\frac{n}{N_t}\right) \\ &\quad + b(t,N_t)n\left(\frac{n+1}{N_t+1}-\frac{n}{N_t}\right) + b(t,N_t)(N_t-n)\left(\frac{n}{N_t+1}-\frac{n}{N_t}\right) \\ &\quad + d(t,N_t)n\left(\frac{n-1}{N_t-1}-\frac{n}{N_t}\right) + d(t,N_t)(N_t-n)\left(\frac{n}{N_t-1}-\frac{n}{N_t}\right)\Bigg]\mathrm{d}t \\ &= s(t,N_t)x_t(1-x_t)\mathrm{d}t.\end{aligned} \tag{D.26}$$

Since $\mathrm{E}[\Delta x|x_t, N_t]^2$ is of order $\mathcal{O}((\mathrm{d}t)^2)$, it holds for the variance of Δx:

$$\mathrm{Var}[\Delta x|x_t, N_t] = \mathrm{E}[(\Delta x)^2|x_t, N_t] - \mathrm{E}[\Delta x|x_t, N_t]^2 \approx \mathrm{E}[(\Delta x)^2|x_t, N_t]. \tag{D.27}$$

We use again Table 1.1 for the calculation:

$$
\begin{aligned}
&\mathrm{E}[(\Delta x)^2|x_t, N_t] \\
&= \Bigg[(\xi(t,N_t)+s(t,N_t))\frac{n(N_t-n)}{N_t}\left(\frac{n+1}{N_t}-\frac{n}{N_t}\right)^2 \\
&\quad + \xi(t,N_t)\frac{n(N_t-n)}{N_t}\left(\frac{n-1}{N_t}-\frac{n}{N_t}\right)^2 \\
&\quad + b(t,N_t)n\left(\frac{n+1}{N_t+1}-\frac{n}{N_t}\right)^2 + b(t,N_t)(N_t-n)\left(\frac{n}{N_t+1}-\frac{n}{N_t}\right)^2 \\
&\quad + d(t,N_t)n\left(\frac{n-1}{N_t-1}-\frac{n}{N_t}\right)^2 + d(t,N_t)(N_t-n)\left(\frac{n}{N_t-1}-\frac{n}{N_t}\right)^2\Bigg]\mathrm{d}t \\
&= \left[\frac{1}{N_t}(2\xi(t,N_t)+s(t,N_t))x_t(1-x_t) + \frac{1}{N_t}\frac{n(N_t-n)}{(N_t+1)^2}b(t,N_t)\right. \\
&\quad \left. + \frac{1}{N_t}\frac{n(N_t-n)}{(N_t-1)^2}d(t,N_t)\right]\mathrm{d}t \\
&\approx \left[\frac{1}{N_t}(2\xi(t,N_t)+s(t,N_t))x_t(1-x_t) + \frac{1}{N_t}x_t(1-x_t)b(t,N_t)\right. \\
&\quad \left. + \frac{1}{N_t}x_t(1-x_t)d(t,N_t)\right]\mathrm{d}t \\
&= \frac{2\xi(t,N_t)+b(t,N_t)+d(t,N_t)+s(t,N_t)}{N_t}x_t(1-x_t)\mathrm{d}t = \frac{x_t(1-x_t)}{N_{e,t}}\mathrm{d}t,
\end{aligned}
\tag{D.28}
$$

where we approximated $N_t + 1 \approx N_t$ and $N_t - 1 \approx N_t$ in the course of the calculation.

Derivation of Eq. (1.34)**.** In order to proof Eq. (1.34), we proof the following relation from which Eq. (1.34) follows immediately by choosing $z = 0$:

$$
\frac{\mathrm{d}^n P(z,t)}{\mathrm{d}z^n} = n!\left[\frac{B^{n-1}(t)}{(A(t)-B(t)(z-1))^n} + \frac{B^n(t)(z-1)}{(A(t)-B(t)(z-1))^{n+1}}\right], \qquad n \geq 1 \tag{D.29}
$$

with the probability generating function (cf. Eq. (1.10))

$$
P(z,t) = 1 + \frac{z-1}{A(t)-B(t)(z-1)}. \tag{D.30}
$$

event	**probability**	$\mathbf{\Delta x}$
a mutant replaces a resident	$(\xi(t,N_t)+s(t,N_t))\frac{n(N_t-n)}{N_t}\mathrm{d}t$	$\left(\frac{n+1}{N_t}-\frac{n}{N_t}\right)$
a resident replaces a mutant	$\xi(t,N_t)\frac{n(N_t-n)}{N_t}\mathrm{d}t$	$\left(\frac{n-1}{N_t}-\frac{n}{N_t}\right)$
a mutant is born	$b(t,N_t)n\mathrm{d}t$	$\left(\frac{n+1}{N_t+1}-\frac{n}{N_t}\right)$
a resident is born	$b(t,N_t)(N_t-n)\mathrm{d}t$	$\left(\frac{n}{N_t+1}-\frac{n}{N_t}\right)$
a mutant dies	$d(t,N_t)n\mathrm{d}t$	$\left(\frac{n-1}{N_t-1}-\frac{n}{N_t}\right)$
a resident dies	$d(t,N_t)(N_t-n)\mathrm{d}t$	$\left(\frac{n}{N_t-1}-\frac{n}{N_t}\right)$

Table 1.1: Summary of the events that change the frequency x of mutants in the population.

We carry out a proof by induction. Eq. (D.29) builds our induction hypothesis.

- Base case $n=1$:

$$\frac{\mathrm{d}P(z,t)}{\mathrm{d}z}=\frac{1}{A(t)-B(t)(z-1)}+\frac{B(t)(z-1)}{(A(t)-B(t)(z-1))^2}.$$

- Inductive step $n\to n+1$:

$$\begin{aligned}
&\frac{\mathrm{d}^{n+1}P(z,t)}{\mathrm{d}z^{n+1}}=\frac{\mathrm{d}}{\mathrm{d}z}\frac{\mathrm{d}^{n}P(z,t)}{\mathrm{d}z^{n}}\\
&=\frac{\mathrm{d}}{\mathrm{d}z}\left[\frac{B^{n-1}(t)}{(A(t)-B(t)(z-1))^n}+\frac{B^n(t)(z-1)}{(A(t)-B(t)(z-1))^{n+1}}\right]\\
&=n!\Big[n\frac{B^{n+1}(t)B(t)}{A(t)-B(t)(z-1))^{n+1}}+\frac{B^n(t)}{(A(t)-B(t)(z-1))n+1}\\
&\qquad+(n+1)\frac{B^n(t)B(t)(z-1)}{(A(t)-B(t)(z-1))^{n+2}}\Big]\\
&=(n+1)!\left[\frac{B^n(t)}{(A(t)-B(t)(z-1))^{n+1}}+\frac{B^{n+1}(t)(z-1)}{(A(t)-B(t)(z-1))^{n+2}}\right].
\end{aligned}$$

It follows:

$$\begin{aligned} p_n(t) &= \frac{1}{n!}\left.\frac{\mathrm{d}P(z,t)}{\mathrm{d}z^n}\right|_{z=0} = \frac{B^{n-1}(t)}{(A+B)^n(t)} - \frac{B^n(t)}{(A+B)^{n+1}(t)} \\ &= \frac{B^{n-1}(t)A(t)}{(A+B)^{n+1}(t)}, \qquad n \geq 1. \end{aligned} \tag{D.31}$$

□

Symmetry of the frequency path in a constant environment. Let Ω_1 be the set of paths which lead from $n_0 = 1$ to $n = n_c - 1$, and Ω_2 be the set of paths which lead from $n = N - n_c$ to $n = N - 1$. Then there exists a bijection between these two sets $f : \Omega_1 \to \Omega_2$ which can basically be constructed by mirroring the path at $x_c = 0.5$. Concretely: Be $\omega_1 = (n_1, \ldots, n_l) \in \Omega_1$ where $n_1 = 1$ and $n_l = n_c - 1$. Define then $f(\omega_1) = \omega_2 = (m_1, \ldots, m_l)$ by $m_i = N - n_{1+l-i}$. Consider now the paths $\omega_1' = (n_1, \ldots, n_l, n_c)$ and $\omega_2' = (m_1, \ldots, m_l, N)$. It now holds: (1) The probability that path ω_1' is realized equals the probability that ω_2' is realized as both make the same number of steps in both directions. (2) The distribution of the run-time along both paths ω_1' and ω_2' is identical because the transition rates of the Moran model defined in Eq. (1.3) are symmetric. From (1) and (2) it follows that the distribution of the time to reach fixation starting from $(N - n_c + 1)/N \approx (1 - x_c)$ is the same as the distribution of the time to reach frequency $x_c = n_c/N$ starting from $1/N \approx 0$. (This is true whether one considers the first, the last or the mean passage time; the latter is approximated by our theory.)

E Appendix: Additional figures

We here compile some additional figures that show results on the fixation probability and the passage time for various scenarios.

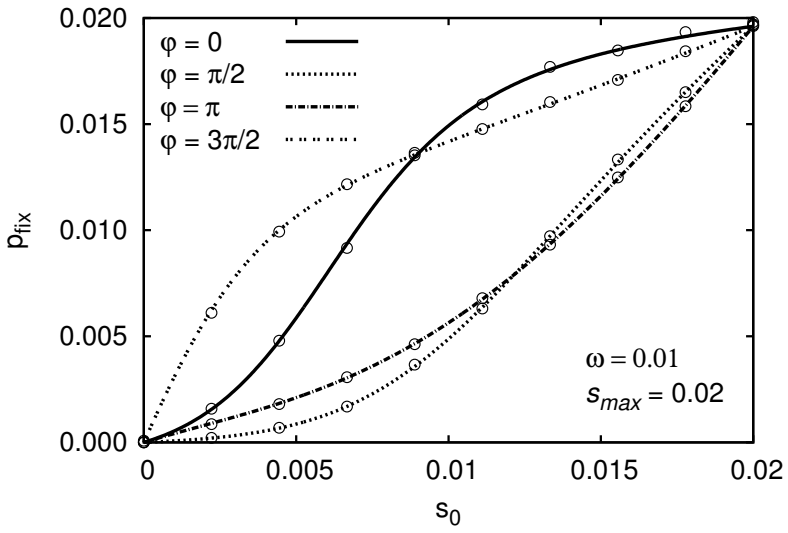


Figure E.5: Fixation probability for a mutation with periodically changing selection strength $s(t) = s_0 + (s_{\max} - s_0)\cos(\omega t + \varphi)$ in dependence of s_0 for various values of φ; $s_{max} = 0.02$, $\omega = 0.01$. Small values of s_0 correspond to relatively long and pronounced periods of selective disadvantage. As s_0 approaches $s_{\max}$ the fixation probability tends to $\approx 2s_{\max}$ as expected. Comparison to simulated data shows that the theory provides an accurate prediction of the fixation probability also for scenarios in which the mutation temporarily gets disadvantageous. For $s_0 \leq 0$, however, it predicts a fixation probability of zero and therefore underestimates the true value. Simulations were performed for a population of 100 000 individuals, and each simulation point is the average over 10^6 runs.

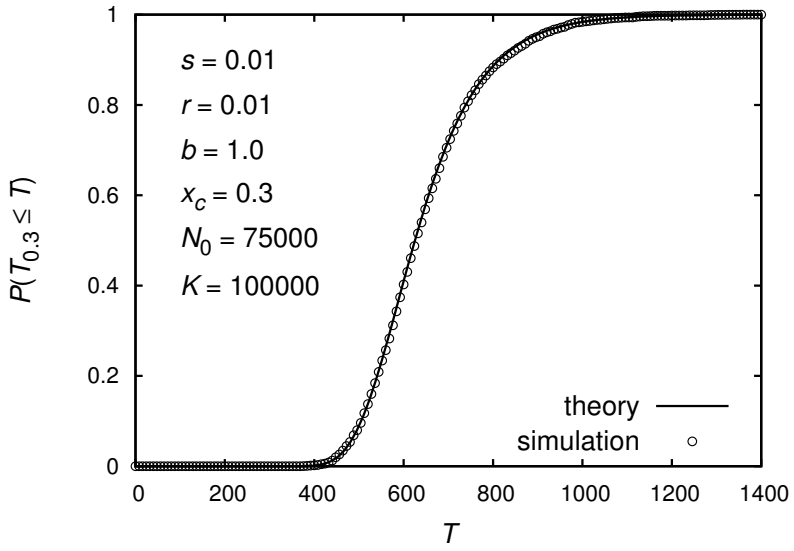


Figure E.6: Distribution of the time to reach frequency $x_c = 0.3$ in a population which grows logistically according to the second scenario with $\xi = 0$ (see text). Analytical and simulation results are compared. The simulation curve is the average over 5000 runs.

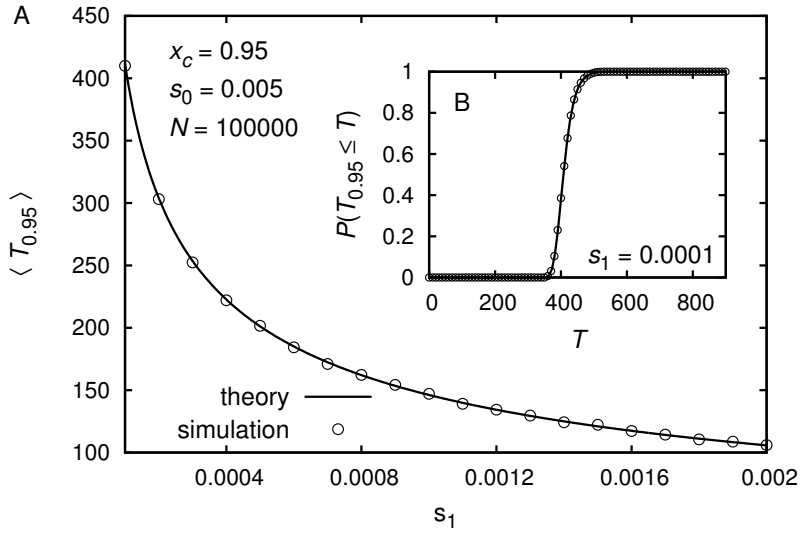


Figure E.7: A: Mean time to reach frequency $x_c = 0.95$ for linearly increasing selection $s(t) = s_0 + s_1 t$ in dependence of s_1. Every simulation point is the average over 100 runs. B: Distribution function for a fixed value of s_1. For the simulation curve 5000 runs were performed.

Chapter 2

Evolutionary rescue in structured populations

Abstract Environmental change, if severe, can drive a population extinct unless the population succeeds in adapting to the new conditions. How likely is a population to win the race between population decline and adaptive evolution? Assuming that environmental degradation progresses across a habitat, we analyze the impact of several ecological factors on the probability of evolutionary rescue. Specifically, we study the influence of population structure and density-dependent competition as well as the speed and severity of environmental change. We also determine the relative contribution of standing genetic variation and new mutations to evolutionary rescue. To describe population structure, we use a generalized island model, where islands are affected by environmental deterioration one after the other. Our analysis is based on the mathematical theory of time-inhomogeneous branching processes and complemented by computer simulations. We find that in the interplay of various, partially antagonistic effects, the probability of evolutionary rescue can show non-trivial and unexpected dependence on ecological characteristics. In particular, we generally observe a non-monotonic dependence on the migration rate between islands. Counterintuitively, under some circumstances, evolutionary rescue can occur more readily in the face of harsher environmental shifts, because of the reduced competition experienced by mutant individuals. Similarly, rescue sometimes occurs more readily when the entire habitat degrades rapidly, rather than progressively over time, particularly when migration is high and competition strong.

This chapter will, with minor changes, be published as: Uecker H. & Otto S. P. & Hermisson J. (in press), *Evolutionary rescue in structured populations.* The American Naturalist.

2.1 Introduction

Environmental change can pose severe challenges to a population. A population, previously well-adapted to the ecological conditions of its habitat, might become maladapted and risk extinction. Examples for such serious alterations are manifold and include global warming and its consequences, the invasion of a new species competing for the same ecological niche, or the onset of drug therapies, selecting for resistance. There are basically three ways in which a population might respond to the environmental deterioration: disperse to another still favorable habitat, adapt by phenotypic plasticity without any change in genotype, or evolve genetic adaptations. Under which conditions populations can escape extinction by rapid adaptive evolution is one of the key questions of evolutionary biology.

Empirical evidence for evolutionary rescue comes from various sources. The evolution of antibiotic or insecticide resistance in natural populations provides prominent examples with a considerable amount of data (e.g., Chevillon *et al.*, 1999; Normark and Normark, 2002; Karasov *et al.*, 2010). A recent survey of vertebrate studies reports a number of cases of population rescue by successful adaptation (Vander Wal *et al.*, 2013). In recent years, several lab experiments have demonstrated the ability of populations to adapt rapidly to highly stressful conditions. These studies have investigated the role of potentially important factors such as the speed of environmental deterioration, population size, genetic variation, the history of stress, modes of dispersal between sub-populations, or recombination and sexual reproduction (Bell and Gonzalez, 2009; Agashe *et al.*, 2011; Bell and Gonzalez, 2011; Lachapelle and Bell, 2012; Bell, 2013; Gonzalez and Bell, 2013).

On the theoretical side, research into evolutionary rescue has followed two directions: one type of study uses a quantitative genetic approach. Many loci with small effects contribute to fitness, and the additive genetic variance (often assumed to be constant) plays a key role. Usually, the focus is on a population's capacity to track an optimum that gradually changes over time (e.g., Lynch *et al.*, 1991; Lande and Shannon, 1996; Bürger and Lynch, 1995) or moves in space (Pease *et al.*, 1989; Polechová *et al.*, 2009; Duputié *et al.*, 2012). Adaptation after a sudden environmental change has also been analyzed in this framework (Gomulkiewicz and Holt, 1995). The simulation studies by Boulding and Hay (2001) and Schiffers *et al.* (2013), where a finite, but large number of loci contribute to fitness, are similar in tracking rescue through the spread of

alleles at multiple loci. The second class of theoretical studies starts from the other end of the scale: adaptation relying on a single mutation (or sometimes a series of mutations at a single locus). In this second class, all models published so far consider a panmictic population that is exposed to a sudden severe change in its environment (e.g., Gomulkiewicz and Holt, 1995; Iwasa *et al.*, 2003, 2004a; Bell and Collins, 2008; Orr and Uncckless, 2008; Martin *et al.*, 2013) or a series of catastrophes (Martin *et al.*, 2013). As a consequence of the environmental deterioration, the population size declines. For rescue of the population, a fitter genotype – either from standing genetic variation or newly mutated – must establish before the resident population becomes extinct.

In this study, we follow the latter approach and analyze a single-locus model with two alleles, representing the wildtype and the rescue mutant. However, we include several key ecological factors that have been left aside in the previous studies, in particular population structure. We note that this latter aspect has been shown to be of importance in a recent empirical study by Bell and Gonzalez (2011). While all previous one-locus models assume a sudden change in the environment that affects the whole population at once, we consider an environmental change that proceeds gradually across the habitat. There are thus temporal refuges, where the resident population can survive for extended periods even after environmental degradation has started. Eventually, however, the whole habitat deteriorates, and the population will go extinct unless a fitter mutant establishes. This mutant might already exist in the population before the shift in the environment (evolutionary rescue from standing genetic variation) or arise afterwards due to recurrent mutation (evolutionary rescue by de-novo mutations). In this context, we determine how ecological factors such as density-dependent competition, migration rates, and the speed and severity of the environmental change influence the probability of evolutionary rescue.

The rate of adaptive evolution – and hence the probability of evolutionary rescue – generically depends on the genetic variation (number of mutants) in the population and on the strength of selection, which determines the establishment probability of each mutation. The habitat ecology can crucially impact both factors, either directly or via ecological interactions. Importantly, an environmental variable can affect the two quantities in various, sometimes antagonistic ways. Effects on the rate of evolutionary rescue may differ, depending on whether rescue occurs mainly from the standing genetic variation or from new mutations. The intertwined influences on genetic variation and selection can lead to surprising results. For panmictic populations, it has previously been pointed out that

harsher environmental change can lead to higher rates of rescue if selection is density dependent, with important implications for the evolution of drug resistance (Gatenby, 2009; Gatenby *et al.*, 2009; Read *et al.*, 2011). In the present paper, we examine the non-trivial patterns that can result if an additional layer of ecological complexity is added, considering gradual deterioration in a structured environment.

The structure of the paper is as follows: We first introduce our model of population structure and gradual environmental degradation. We then present an overview of the phenomena observed in the full model based on simulation results. Afterwards, we consider a set of simpler sub-models that allow for an analytical treatment in order to crystallize the main effects and to reveal the basic principles behind them. All mathematical analyses are presented in the Appendix.

2.2 The model

Consider a population of haploid individuals that lives in a patchy environment. We focus on one locus with two alleles: the wildtype and the rescue mutant allele. Every generation, a fraction m of the offspring enters a migrant pool, which is thereafter distributed with equal probability over all demes (the "island model", Wright (1943)). The migration probability m, along with the number of demes D, is thus a measure of the degree of fragmentation of the habitat. Initially, selection is homogeneous across space. The population is well adapted to the ecology of its habitat, such that the population size in each deme is at its carrying capacity. The mutant allele has a disadvantage relative to the resident type and is only present at low frequencies at mutation-selection-migration-drift balance. Subsequently, the environment starts to deteriorate. Environmental deterioration does not affect the whole metapopulation at once, but proceeds gradually across the habitat. We assume that the change progresses at constant speed. That is, every ϑ generations another patch switches to the new environment (Figure 2.1), where $\vartheta = 0$ corresponds to an instantaneous degradation of the entire habitat. While the populations in demes with the original environment can maintain their size at carrying capacity, the fitness of the wildtype drops below one in the perturbed demes, and the population size starts to decline. Whereas the wildtype cannot persist in the new environment, mutants have – at least at low population densities – a fitness greater than one. They are, however, initially rare (or absent) in the population and therefore likely to suffer stochastic loss. At

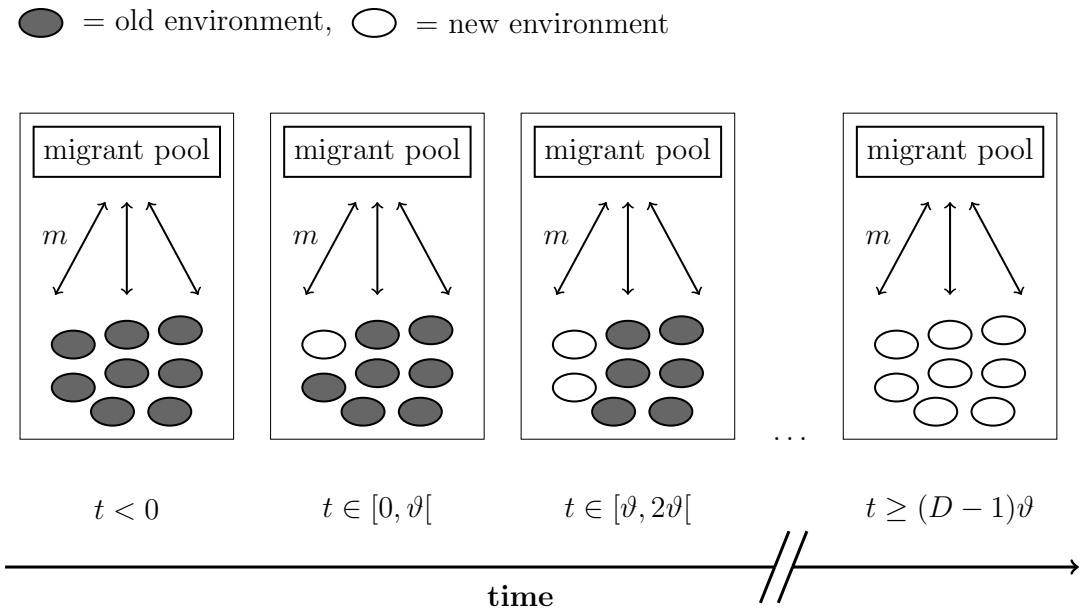


Figure 2.1: Sketch of the spatial and temporal habitat structure. The population lives in a patchy environment with D demes. Each generation a fraction m of the offspring enters a migrant pool, which is then randomly distributed over the islands. Beginning at time $t = 0$, the environment changes in intervals of ϑ generations, affecting one deme at a time, until all demes are affected. The environmental change is thus idealized as a discrete switch and may, for example, represent a shift in habitat type, the appearance of an invading species, or an alteration in climate.

this stage, the metapopulation corresponds to a source-sink system where the unperturbed demes act as sources and the perturbed demes as sinks for wildtype individuals. Because the size of the sink grows over time and the source vanishes, the metapopulation will ultimately go extinct unless it is rescued by adaptive mutation.

We denote the carrying capacity for each deme as K and define $K_{\text{total}} = DK$. $N_w^{(i)}$ and $N_m^{(i)}$ are the number of wildtypes and mutants in deme i. The life cycle is assumed to be:

1. Reproduction and mutation: Each individual of the parent generation, irrespective of type, produces a large number X of offspring. A fraction u of the wildtype offspring mutates. Back mutation is ignored.

2. Migration: A fraction $1 - m$ of the offspring remains in its home deme, a fraction m enters the migrant pool. A fraction $1/D$ of the migrant pool settles in each deme.

3. Selection & density regulation: Each deme has a hard carrying capacity K. Under the original environmental conditions, the offspring viability is sufficiently high such that the deme is fully occupied after completion of the life cycle. Following the classical island model, we determine the genetic composition of the next generation in an unperturbed deme by binomial sampling of K individuals from the local offspring pool with frequencies weighted by fitness. The number of mutants thus follows a binomial distribution with parameter

$$p = \frac{\alpha(N_m^{(i)\prime} + uN_w^{(i)\prime})}{(1-u)N_w^{(i)\prime} + \alpha(N_m^{(i)\prime} + uN_w^{(i)\prime})}, \tag{2.1}$$

where α is the relative fitness of the mutant in the old environment and

$$\begin{aligned} N_w^{(i)\prime} &= (1-m)N_w^{(i)} + \frac{m}{D}\sum_{k=1}^{D} N_w^{(k)}, \\ N_m^{(i)\prime} &= (1-m)N_m^{(i)} + \frac{m}{D}\sum_{k=1}^{D} N_m^{(k)}. \end{aligned} \tag{2.2}$$

We denote by $z_0 := 1 - \alpha$ the strength of selection against mutants in unperturbed demes. Under the deteriorated conditions in the new demes, survival probabilities are lower, and at least as long as the mutant is rare, the carrying

capacity will in general not be reached. Wildtype individuals are unable to replace themselves, and we set the probability of survival for each of their X offspring to $(1-r)/X$. In contrast, mutant individuals are – at least at low densities – able to positively grow; mutant offspring survive with probability $(1+S_i)/X$ in deme i. In the limit $X \to \infty$, the number of wildtype and mutant individuals after selection but before density-regulation follows a Poisson distribution:

$$\begin{aligned} N_w^{(i)} \to &\quad \text{Poisson with parameter} \quad (1-u)(1-r)N_w^{(i)\prime}, \\ N_m^{(i)} \to &\quad \text{Poisson with parameter} \quad (1+S_i)(N_m^{(i)\prime} + uN_w^{(i)\prime}). \end{aligned} \tag{2.3}$$

If necessary, the population size is thereafter reduced to carrying capacity K. For large r, this is very rarely needed until rescue has occurred.

Note that (unless density regulation is needed) $(1-r)$ plays the role of an absolute fitness of wildtype parents under the perturbed conditions (ignoring mutation). We use r as a measure of the severity of the environmental change. Similarly, $(1+S_i)$ is the absolute fitness of mutant parents in the perturbed deme i, where we model S_i as follows:

$$S_i = \max\left[-z, s\left(1-\beta\frac{N_w^{(i)\prime}+N_m^{(i)\prime}}{K}\right)\right] \tag{2.4}$$

with $s, z > 0, \beta \geq 0$. For $\beta = 0$, we have $S_i = s$ and fitness is density independent. For $\beta > 0$, growth of the mutant population is reduced in the presence of competing wildtype or mutant individuals. For $\beta > 1$, this entails an effective reduction of the mutant carrying capacity in the new environment to K/β. The parameter z sets a limit to the harmful effects of competition. Choosing $z \leq 1$ prevents absolute fitness from becoming negative. For $z = 1-\alpha$, the absolute fitness of mutants in a fully occupied deme is the same under both environmental conditions. The effect of competition on mutants in a fully occupied deme is weaker in the new environment for $z < 1-\alpha$ and stronger for $z > 1-\alpha$. For simplicity, we refrain from explicitly modeling density dependence of the wildtype fitness. Note that this latter assumption is more a technicality: the relevant feature of the wildtype population size in the degraded environment is a decay that is still sufficiently fast at low densities. The rate of decay (the harshness of the environmental change) is controlled by the parameter r; the specific mode of the decay is of minor importance. Note also that under certain scenarios, mutant

fitness might indeed be density dependent, while the wildtype fitness is not. E.g., it is possible that mutants are able to convert a particular resource in a manner that allows them to grow in the new environment; while all individuals use up this resource, it only affects the growth rate of mutant individuals and only in the new environment.

All model parameters and further notations used throughout the manuscript are summarized in Table 2.1.

2.3 Simulations

Algorithm The simulation program implements the discrete-generation life cycle, with offspring numbers in the old and the new demes determined by drawing from a binomial or a Poisson distribution, following Eq. (2.1) or (2.3) (plus possibly density-regulation), respectively. We start the simulations with all demes fully occupied by wildtype individuals and let the population evolve for a large number of generations to generate mutation-selection-drift equilibrium before the deterioration of the habitat sets in. After all demes have deteriorated, we track the population until the wildtype has gone extinct. If any mutants are present at this point, we let the simulations run, until the mutant has either reached a threshold density or has gone extinct. As a threshold density, we choose 90% of the total carrying capacity of the mutant if $m > 0$ and 90% of the carrying capacity of the mutant in a single deme if $m = 0$. The latter implies that the population is considered as rescued even if only a single sub-population survives. If not stated otherwise in the figure legend, simulation points represent averages over 10^7 replicates. All computer simulations were written in the C programming language, making use of the *Gnu Scientific Library* (Galassi *et al.*, 2009).

Observations Our most significant findings from the numerical analysis are summarized in Figures 2.2, 2.3, and 2.4. We focus on the dependence of the rescue probability on the strength of migration m, the severity of change (reflected by r), and the speed of environmental change (modeled by ϑ).

The dependence of the rescue probability on the migration rate is explored in Figures 2.2A and 2.2B. In all cases, we observe a (local) maximum in the probability of evolutionary rescue for low, but non-zero levels of migration ($m_{\max} \approx s$). In some cases, rescue increases again for strong migration. This secondary in-

D, K	number of demes and carrying capacity of a single deme, $K_{\text{total}} = KD$
$N_{w/m}^{(i)}$, $N_{w/m}^{(\text{old})}$, $N_{w/m}^{(\text{new})}$, $N_{w,m}^{(\text{total})}$	number of wildtype/mutant individuals, respectively, in deme i, the unperturbed (old) and perturbed (new) part of the habitat, and in the entire metapopulation
m	fraction of offspring that enter the migrant pool
u	mutation probability
ϑ	intervall between the deterioration of two demes
α, z_0	$\alpha = 1 - z_0$: relative fitness of the mutant in the old environment
$1 - r$	absolute fitness of a wildtype parent in the new environment
$1 + S_i$	absolute fitness of a mutant parent in deme i with the new environment (cf. Eq. (2.4))
β	strength of density dependence experienced by a mutant in the new environment
$1 - z$	minimum absolute fitness of a mutant parent in the new environment (at high density) if $\beta > 0$
$1 + s$	maximum absolute fitness of a mutant parent in the new environment (at low density) if $\beta > 0$; absolute fitness of a mutant in deme i if $\beta = 0$
s_{eff}	effective strength of selection experienced by a mutant
$p_{\text{est}}^{(i)}$	establishment probability of a single mutant in deme i

Table 2.1: Summary of notations.

crease for large m is observed if α (see Figure 2.2A) and r (see Figure 2.2B) are not too small. The chances of rescue become very high for strong migration, if r is large (cf. $r = 0.4$ in Figure 2.2B). All effects of strong migration are observed across a larger parameter range (i.e., even for smaller values of α and r), if the total number D of demes is larger. As we will see, the maximum for intermediate migration is generated by the interplay of two antagonistic effects: on the one hand, migration leads to an increased mutational input, on the other hand, by migration, mutants end up in unperturbed demes where they suffer a disadvantage (see section "Evolutionary rescue in an island model without standing genetic variation"). The potential increase for very large m is a consequence of relaxed competiton in the old environment (see "Evolutionary rescue in a Levene model").

Figures 2.2C and 2.2D and Figure 2.3 summarize the dependence of the rescue probability on the severity of the environmental change, $P_{\text{rescue}}(r)$. We first compare Figures 2.2C and 2.2D. As expected, the rescue probability declines with r for many parameter combinations. However, we observe a counterintuitive behavior in a relevant part of the parameter space, where $P_{\text{rescue}}(r)$ exhibits a pronounced minimum. This means that a harsher environmental change (larger r) can increase the probability of survival in some cases. Figure 2.2C shows the typical behavior for weak and intermediate migration. In this case, we observe a non-monotonic behavior for relatively high mutant fitness in the old environment and strong density dependence in the new environment (high values of both α and β). Further requirements are a sufficiently large value of s and a large, but not too large, value of z (i.e., sufficiently strong selection and density dependence in the new habitat). Figure 2.2D represents extensive migration. In this case, the parameter space yielding non-monotonic behavior is much enlarged. We observe a non-monotonic shape even for intermediate mutant fitness α and even in the absence of any additional density dependence in the perturbed environment ($\beta = 0$). For the parameters explored in Figures 2.2C and D, standing genetic variation is required in order to observe a minimum in the rescue probability ($\alpha > 0$). However, under other parameters, $P_{\text{rescue}}(r)$ can display a non-monotonic behavior even when $\alpha = 0$ and there is hence no standing variation, as illustrated in Figure 2.3. This occurs if density dependence is strong (β large), selection strong relative to migration (s large, with larger values of m requiring larger values of s), and the speed of deterioration slow (ϑ large). As we will explain in the analysis section, the advantage of a faster decay (higher r) comes about because of weakened competition either on the islands with the new environment (see

section "Evolutionary rescue in scenarios where habitat structure is absent or immaterial") or on the islands with the old environment (see section "Evolutionary rescue in a Levene model"). While a harsher change reduces the number of rescue mutations that appear by mutation, it increases the establishment probability of each single mutant due to these effects.

Figure 2.4 focusses on the influence of the speed of environmental change modeled by ϑ. In Figure 2.4A, we see that the probability of evolutionary rescue can either increase or decrease as a function of ϑ, i.e., a slowly progressing change can be better or worse for the population than an instantaneous degradation of the whole habitat. Figure 2.4A demonstrates how this depends on the strength of density dependence β. Figure 2.4B shows the probability of evolutionary rescue as a function of m for various values of ϑ. β is fixed to the value for which in Figure 2.4A, a rapid change was favoured over a slow one. We observe that for weak and intermediate migration, the survival probability significantly increases as ϑ gets larger (slower change), while in line with Figure 2.4A, it decreases with increasing ϑ for strong migration. Overall, we find that a rapid change facilitates rescue if m, α, z, s, and r are large and β not too small. The reason for this behavior is similar to that indicated above when considering the harshness of change (faster degradation can relax competition, see section "Evolutionary rescue in a Levene model"), but the parameter range yielding the counterintuitive behavior is much more restricted. For strong density dependence and migration, the fitness of mutants gets strongly suppressed in the new environment due to competition with the immigrating wildtype individuals even for large r. If standing genetic variation is large and de-novo mutations rare (large α, large r, small u), the probability of rescue is therefore larger for an instantaneous degradation of the entire habitat without temporal refugia.

In the following, we will explain within three analytically tractable sub-models how these patterns arise. In a first step, we will consider a class of models in which habitat structure is absent or immaterial ($D = 1$ or $m = 0$ or $\vartheta = 0$). The second sub-model is a generalization of the Levene (1953) model ($m = 1$). Last, we will analyze an island model with $\alpha = \beta = 0$, i.e., where the mutation is lethal in the original environment and mutant fitness in the perturbed environment is density-independent.

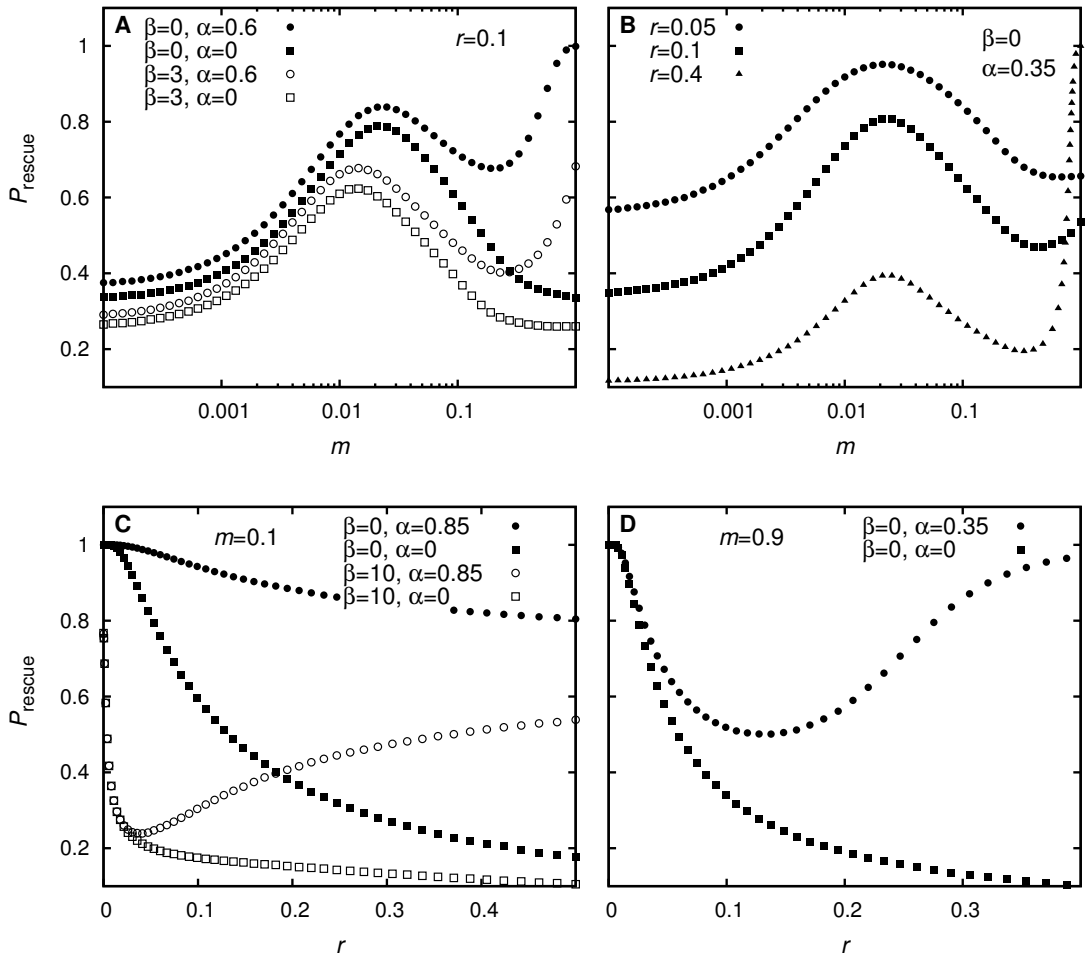


Figure 2.2: Panels A and B: The probability of evolutionary rescue as a function of m. Filled and open symbols indicate the absence or presence of density dependence (beyond the hard carrying capacity K). In Panel A, squares and circles distinguish scenarios with and without standing genetic variation. In all cases, we observe a local maximum for intermediate migration. In some cases, the rescue probability increases again for strong migration. Panels C and D: The probability of evolutionary rescue as a function of r. Filled symbols correspond to scenarios with no explicit density dependence ($\beta = 0$), while open symbols denote strong density dependence ($\beta = 10$). Squares represent scenarios where the mutation is lethal under the original conditions ($\alpha = 0$), circles denote scenarios with standing genetic variation ($\alpha > 0$). The plots show that a harsher change (larger r) is sometimes better for the survival of the population than a milder one. Each simulation point is the average over 10^6 replicates. (Parameters: $D = 8$, $z = 0.2$, $s = 0.02$, $K = 2500$, $\vartheta = 250$, $u = 0.5 \cdot 10^{-4} = 1/K_{\text{total}}$)

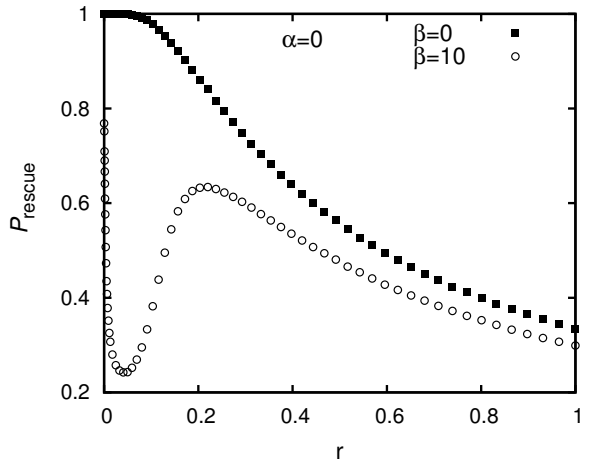


Figure 2.3: The probability of evolutionary rescue as a function of r. $P_{\text{rescue}}(r)$ can attain a minimum and a maximum for intermediate values of r. Each simulation point is the average over 10^6 replicates. (Parameters: $D = 2$, $z = 0.2$, $s = 0.02$, $m = 0.01$, $K = 10000$, $\vartheta = 5000$, $u = 0.5 \cdot 10^{-4} = 1/K_{\text{total}}$)

2.4 Analysis

General approach

The probability of evolutionary rescue is determined by the rate of successful mutants, i.e., the rate of mutants that not only arise but also establish in the population. In a panmictic population, this rate is proportional to the number of wildtype individuals, which determines – together with the mutation probability u – the number of mutants that are generated each generation, and the establishment probability of a new mutant. In a spatially structured population, both the mutational input and the establishment probability vary across demes. In a gradually deteriorating environment, we are inherently far from equilibrium, and both factors are time-dependent.

In our model, the number of new mutations that arise in generation t and survive selection and density regulation in that generation is given by the respective terms in Eq. (2.1) and (2.3). As long as the rescue allele is rare, the binomial distribution in Eq. (2.1) can be approximated by a Poisson distribution with mean $u\alpha K$. Consequently, the number of successful mutants that is generated in deme i in generation t follows a Poisson distribution with mean $u\alpha K p_{\text{est}}^{(i)}(t+1)$ if the

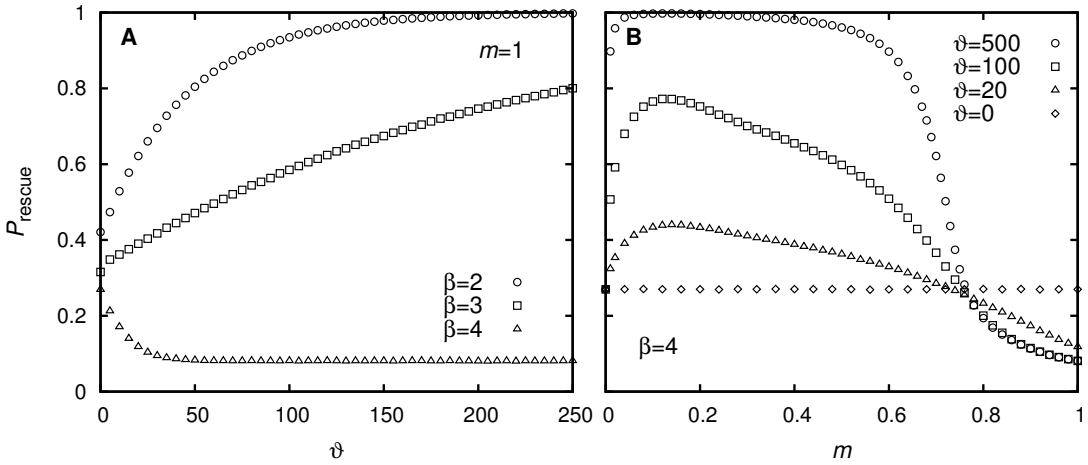


Figure 2.4: Panel A: The probability of evolutionary rescue as a function of ϑ for three levels of density dependence. We see that for strong migration and density-dependent selection, an instantaneous deterioration of the whole habitat ($\vartheta = 0$) can be better for population survival than a slowly progressing change. Panel B: The probability of evolutionary rescue as a function of m. For $m \gtrsim 0.8$, a fast change leads to a higher probability of evolutionary rescue than a slow change. Each simulation point is the average over 10^6 replicates. (Parameters: $D = 2$, $\alpha = 0.99$, $z = 0.685$, $s = 0.411$, $K = 10000$, $r = 0.5$, $u = 0.5 \cdot 10^{-5} = 0.1/K_{\text{total}}$)

deme has not yet deteriorated and $u(1+S_i(t))N_w^{(i)\prime}(t)p_{\text{est}}^{(i)}(t+1)$ in the deteriorated environments, respectively. $p_{\text{est}}^{(i)}(t)$ is the establishment probability of a mutant in deme i in generation t at the start of the life cycle (i.e., before reproduction) and $N_w^{(i)\prime}(t)$ is defined as in Eq. (2.2). For a given demographic history, the probability of evolutionary rescue is hence given by

$$P_{\text{rescue}} \approx 1 - \exp\left[-\sum_{i=1}^{D}\left(\sum_{t=-\infty}^{(i-1)\vartheta-1} u\alpha K p_{\text{est}}^{(i)}(t+1) \; + \sum_{t=(i-1)\vartheta}^{\infty} u(1+S_i(t))N_w^{(i)\prime}(t)p_{\text{est}}^{(i)}(t+1)\right)\right], \tag{2.5}$$

where the exponential is the probability that no sucessful mutant is generated in any deme in any generation. The sum $\sum_{i=1}^{D}\sum_{t=-\infty}^{(i-1)\vartheta-1}$ captures the contribution of standing genetic variation (defined as mutations arising before a deme degrades), the sum $\sum_{i=1}^{D}\sum_{t=(i-1)\vartheta}^{\infty}$ the contribution of de-novo mutations (subsequent to environmental degradation).

Throughout the paper, we model the dynamics of the wildtype population size deterministically, i.e., we neglect demographic stochasticity (see Martin *et al.* (2013) for a study that includes demographic stochasticity). The establishment phase of new mutants, however, is characterized by strong stochasticity and thus requires a probabilistic treatment. Following a long tradition in population genetics, we calculate establishment probabilities via time-inhomogeneous branching processes (Kendall, 1948; Allen, 2011; Uecker and Hermisson, 2011). This approach is based on the following reasoning: As long as the mutation is rare in the population, mutant offspring suffer (nearly) independent fates, and their early spread is therefore well described by a branching process. In a structured population, we must a priori distinguish several types of individuals, according to deme, and consequently describe the dynamics of the wildtype by a set of coupled difference equation and the spread of the mutant by a multitype branching process. In the following sections, we focus, however, on limiting cases which all allow for an effective reduction of the dimensionality of the problem. In these cases, we can model the deterministic dynamics of the wildtype by a single difference equation (see Eq. (A.1), applicable to all three sub-models) and the early phase of mutant growth by a single-type branching process with a time-dependent effective growth parameter $s_{\text{eff}}(t)$.

Details on the analysis can be found in the Appendix. All numerical evaluation of integrals is done in Mathematica (Wolfram Research, Champaign, USA).

Comparison to computer simulations shows that the analytical results are highly accurate.

2.4.1 Evolutionary rescue in panmictic populations with $D = 1$ and scenarios where habitat structure is immaterial

We first focus on scenarios where habitat structure is either absent ($D = 1$) or proves to be immaterial, i.e., does not affect the probability of evolutionary rescue. We find that habitat fragmentation is irrelevant if individuals do not migrate ($m = 0$; rescue is defined as survival of at least one sub-population) or if the environment changes simultaneously on all islands ($\vartheta = 0$). In both cases, the probability of rescue is the same as in an unstructured population of size $K_{\text{total}} = KD$ irrespective of the number D of demes. For zero migration, this initially surprising result essentially follows because in both the unstructured and the island model, the local population size declines at rate r after the local environmental shift, regardless of when this shift occurs. If the environment changes at the same time in the whole habitat ($\vartheta = 0$), the wildtype population size decays simultaneously in all demes with the same rate r as in the panmictic case, regardless of the migration rate. Hence, migration of mutants has no effect on their establishment probability, even if mutant fitness is density dependent. A formal proof of why the cases $D = 1$, $m = 0$, and $\vartheta = 0$ coincide as well as details on the analysis are given in Appendix B. For simplicity, we stick to $D = 1$ in the following. In that case, we naturally deal with only one type of wildtype and one type of mutant individual.

For density-independent fitness ($\beta = 0$), we are able to derive simple analytical approximations for the rescue probability (see Appendix B). For s and z_0 small, we obtain:

$$P_{\text{rescue}}^{\text{sgv}} \approx 1 - \exp\left[-2uK \ln\left(\frac{s+z_0}{z_0}\right)\right] = 1 - \left(\frac{s+z_0}{z_0}\right)^{-2uK}, \tag{2.6}$$

$$P_{\text{rescue}}^{\text{dnm}} \approx 1 - \exp\left[-u\frac{K}{r}2s\right], \tag{2.7}$$

where $P_{\text{rescue}}^{\text{sgv}}$ and $P_{\text{rescue}}^{\text{dnm}}$ denote the probability of evolutionary rescue by mutations from standing genetic variation and de-novo mutations, respectively. These formulae agree with the results by Orr and Uncкless (2008), who provide a

detailed discussion of evolutionary rescue in unstructured populations when fitness is density independent. (Note that the absolute fitness of a mutant in the new environment is $1 + s$ in our model, while it is $1 + s_b - r$ in Orr and Uncкless (2008); the parameter z_0 corresponds to s_d in Orr and Uncкless (2008).) In this case, the decay rate r of the wildtype population size enters the result solely via the mutational input. As the latter is larger for smaller values of r, the probability of evolutionary rescue monotonically decreases as r increases. The probability of evolutionary rescue from standing genetic variation is unaffected by the severity of change as long as mutant fitness is density independent below the hard carrying capacity K. In contrast, if fitness is density dependent ($\beta \neq 0$), the establishment probability of mutations from both standing genetic variation and de-novo mutation depends on how fast the wildtype population size decays, and more complex behavior may arise.

Figure 2.5 shows the probability of evolutionary rescue as a function of r for various combinations of β (affecting the strength of density dependence) and α (affecting the amount of standing genetic variation). If either α or β (or both) are small, the rescue probability decreases as a function of r (apart from bearly visible non-monotonic behavior, see Appendix B for details). However, when both parameters are large, we observe a pronounced minimum in the probability of evolutionary rescue. How can we understand this behavior? Both de-novo mutations and mutations from standing genetic variation contribute to evolutionary rescue, and it is helpful to consider both contributions separately. The mutational input ($\sim uN_w$) during population decline decreases as a function of r, while the establishment probability p_{est} increases because of weakened competition. The probability of evolutionary rescue by de-novo mutations depends on both factors; its overall trend is governed by the declining mutational input. By contrast, the probability of evolutionary rescue from standing genetic variation increases with r because mutant fitness is density dependent. The two contributions to population survival – rescue by de-novo mutations and rescue by mutations from standing genetic variation – thus exhibit opposite behavior as a function of r. When β is large, the population size has to be greatly reduced for the mutant growth parameter to become positive. In that case, the wildtype population size is already low, by the time that the establishment probability of de-novo mutations becomes significant. Rescue by a de-novo mutation is therefore only likely if r is very small such that the number of mutants generated before extinction of the wildtype is nevertheless high. Consequently, $P_{\text{rescue}}^{\text{dnm}}$ decays rapidly as a function of r. At the same time, mutations from standing genetic variation can

only contribute to rescue if the wildtype individuals are rapidly eliminated such that the mutants can survive up to the time when their fitness finally exceeds one. This contribution is substantial if the amount of standing genetic variation is high (large α). This implies that if β and α are both large, the contribution from de-novo mutations is high for small r and the contribution from standing genetic variation is high for large r, while for intermediate r none of the two contributions is particularly strong, leading to a minimum in the total probability of evolutionary rescue. To illustrate this point, we included the probability of evolutionary rescue from standing genetic variation and by de-novo mutations individually in Figure 2.5.

A pronounced minimum is generated if (1) density dependence is strong (sufficiently large β and z in our model) and (2) the rescue probability from standing genetic variation is high for large r. The latter condition requires large values of α and s. Furthermore, z must not be too large (in particular, $z \neq 1$), in order for mutants from the standing genetic variation to survive the first few generations after the environmental change, while the population is still large.

We point out that for extreme parameter values, a third pattern is possible: due to the antagonistic effects of r, $P_{\text{rescue}}^{\text{dnm}}$ and along with it the total probability of evolutionary rescue attains a minimum, then a maximum, and decays afterwards (Figure B.3). For details, we refer to Appendix B.

In structured populations with $m > 0$ and $\vartheta = 0$, additional genetic variation and targets of mutation are provided by immigration of individuals from the unperturbed to the perturbed demes. For slow change (large ϑ), the wildtype population in the perturbed demes approaches migration-selection balance. With density-dependent selection, mutants in these demes will be able to grow if and only if the density is sufficiently low, which is the case for large r. As a consequence of wildtype immigration from unperturbed to perturbed demes, the requirements on α are strongly relaxed, and we find a non-monotonic behavior even in parameter regions with $\alpha = 0$ (cf. Figure 2.3).

2.4.2 Evolutionary rescue in a Levene model

As a next step, we investigate the influence of gradually changing heterogeneous selection with "good" and "bad" islands, but without population structure (i.e., $m = 1$). This leads to a variant of the Levene (1953) model with environmental deterioration. To begin with, we confine the treatment to density-independent selection ($\beta = 0$).

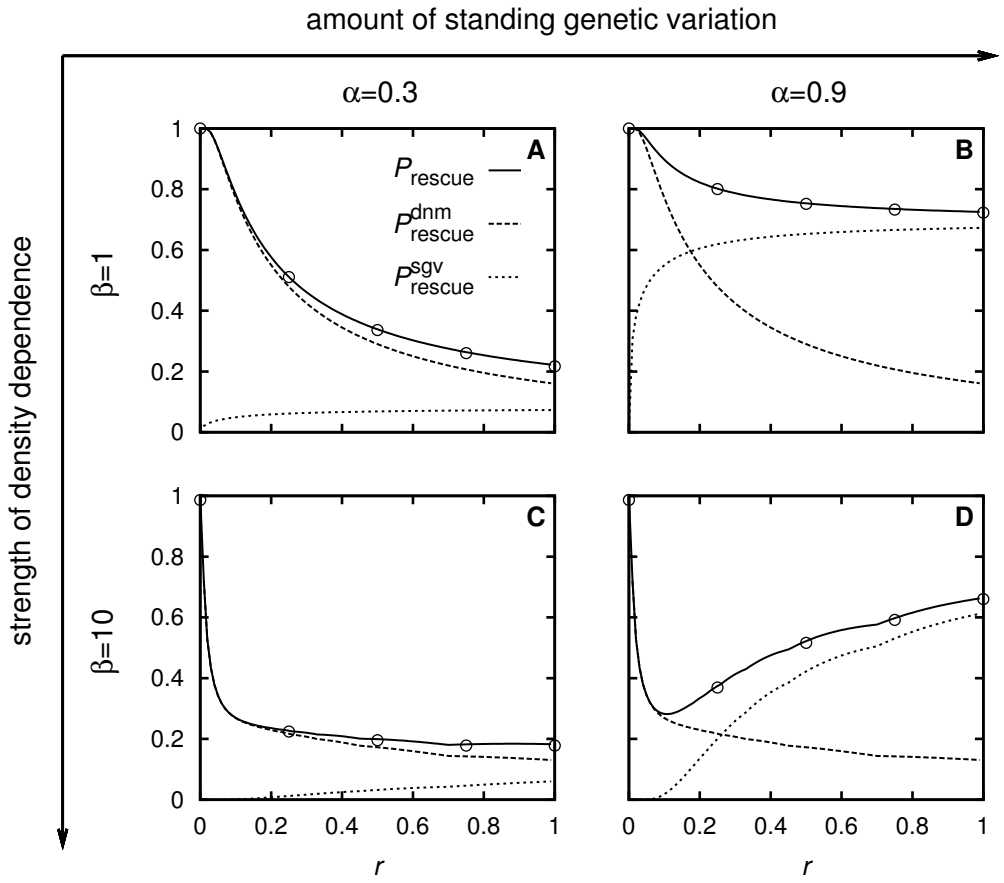


Figure 2.5: Evolutionary rescue in an unstructured population ($D = 1$). Dashed curves give the probability of rescue from de-novo mutations (i.e. those occuring after the environmental change), while dotted curves give the probability of rescue from standing genetic variation. Depending on the amount of standing genetic variation (α) and the strength of density dependence (β), the probability of evolutionary rescue decreases with r or shows a pronounced minimum for intermediate values of r. The parameter values are: $z = 0.2$, $s = 0.1$, $K = 20000$, $u = 0.5 \cdot 10^{-4} = 1/K$. The theoretical curves are based on Eq. (B.3), where the establishment probability has been calculated via Eq. (A.7). Circles denote simulation results.

Since all offspring enter the migrant pool, all wildtypes are equivalent, regardless of source population, and the same holds for mutants. Thus, there is again only one kind of each allelic type, enabling analytical treatment. The crucial quantity for the understanding of the functional behavior of P_{rescue} is the effective growth parameter of a mutant (which enters the establishment probability). To derive the growth parameter, consider first a single old patch. The genetic composition of the next generation is determined by binomial sampling of K individuals, and the number of mutants follows a binomial distribution with parameter

$$\frac{\alpha\frac{N_m^{(\text{total})}(t)}{D}}{\alpha\frac{N_m^{(\text{total})}(t)}{D}+\frac{N_w^{(\text{total})}(t)}{D}}=\frac{\alpha N_m^{(\text{total})}(t)}{\alpha N_m^{(\text{total})}(t)+N_w^{(\text{total})}(t)}, \tag{2.8}$$

where $N_m^{(\text{total})}(t)$ and $N_w^{(\text{total})}(t)$ are the number of mutants and residents in the whole meta-population. As long as mutants are rare relative to the number of wildtypes, the binomial distribution can be approximated by a Poisson distribution with parameter $\alpha(N_m^{(\text{total})}/N_w^{(\text{total})})K$. Since the offspring get distributed with equal probability over the demes, we can also (artificially) assign $N_m^{(\text{total})}/D$ mutant parents to each deme. For each mutant individual associated with a particular unperturbed deme, we thus obtain a Poisson distributed number of offspring with parameter

$$\alpha\frac{KD}{N_w^{(\text{total})}(t)}. \tag{2.9}$$

As $N_w^{(\text{total})}(t)/D$ will be smaller than K once the environmental deterioration has started, $\alpha DK/N_w^{(\text{total})}(t)$ will be larger than α. The reason is relaxed competition with the wildtype in this pre-disturbance patch: as the wildtype population size drops, fewer wildtypes contribute to the migrant pool and consequently to the local offspring pool. We now average the growth parameter of the mutant over all patches, weighting good and bad demes according to their respective numbers. For a period with d deteriorated demes, we obtain:

$$s_{\text{eff}}(t)=\frac{d}{D}(1+s)+\frac{D-d}{D}\alpha\frac{KD}{N_w^{(\text{total})}(t)}-1. \tag{2.10}$$

In Appendix C, we derive an approximation for the growth parameter of the mutant if ϑ is large enough that $N_w^{(\text{total})}(t)$ can be replaced by its stationary value, yielding:

$$s_{\text{eff}}(t) \approx \begin{cases} \alpha - 1 & \text{for} \quad t < 0, \\ \alpha - 1 + \frac{1-\alpha+s+\alpha r}{D} d = & \text{for} \quad t \in [(d-1)\vartheta, d\vartheta[, \quad d \in \{1, \ldots, D-1\}, \\ s & \text{for} \quad t \geq (D-1)\vartheta \, . \end{cases} \tag{2.11}$$

Thus, the growth parameter increases with the number of perturbed demes up to time $(D-1)\vartheta$. It might even exceed the growth parameter s in the new environment, if:

$$s_{D-1} = \alpha - 1 + \frac{1-\alpha+s+\alpha r}{D}(D-1) > s \;\Leftrightarrow\; \alpha > \frac{1+s}{1+r(D-1)}, \tag{2.12}$$

where s_{D-1} approximates the growth parameter in the period with $D-1$ deteriorated demes, following Eq. (2.11). For "infinitely" many islands, in particular, this condition is always met except for $\alpha = 0$: the strength of effective selection goes up to $s + \alpha r$ before it drops to s after the last deme has deteriorated. The temporal development of the growth parameter for small and large r is depicted in Figure 2.6A. Figure 2.6B shows the corresponding establishment probabilities. For large values of r, the effect can be quite strong. Ultimately, this is again a consequence of density-dependent fitness: density regulation in the old patches renders the absolute fitness of mutants density dependent, even if the relative fitness α is constant. When the wildtype population size decreases, competition is relaxed and the absolute fitness of mutants is increased. For the successful establishment of a mutation, absolute offspring numbers and thus the density-dependent absolute fitness matters. These considerations show that again a low fitness of the wildtype in the perturbed habitat has both a positive effect – the establishment probability increases with r – and a negative effect – the mutational input decreases. The positive effect gets stronger for larger values of α and D.

Figure 2.7 shows the probability of evolutionary rescue as a function of α and r, respectively. As can be seen from all three panels, P_{rescue} significantly increases with α; the increase is particularly strong for large D (Figure 2.7A) and r (Figures 2.7B and C). Note that $\alpha = 0$ implies that evolutionary rescue entirely relies on de-novo mutations. We also see that due to the antagonistic effects of a fast decay of the wildtype population size, the probability of evolutionary rescue

can have a pronounced minimum as a function of r (see Figures 2.7B and 2.7C). If α is small, the effect of relaxed competition is weak. In this case, the survival probability decreases as r increases because the total mutational input (and in particular the mutational input after the last deme has deteriorated) diminishes with increasing r (solid curve in Figure 2.7C).

We now include density-dependent mutant fitness into the analysis and consider the limits $\vartheta \to 0$ and $\vartheta \to \infty$. As discussed in the previous section, the case $\vartheta = 0$ compares to a scenario with no habitat structure. If ϑ is very large, it is sufficient to focus on times after all demes but one have turned bad: Two scenarios can be distinguished. Either the mutant fitness in $[(D-2)\vartheta, (D-1)\vartheta[$ is larger than one. In that case, as mutations arise recurrently, the mutant type will certainly establish during this last period. Or the mutant fitness is smaller than one. Then, mutations that were generated before time $(D-2)\vartheta$ (i.e., before the second last deme deteriorated) will not survive up to time $(D-1)\vartheta$ and we can safely ignore them. We can thus restrict our attention to a single environmental switch at time $(D-1)\vartheta$. As $m = 1$, all wildtype and all mutant individuals are equivalent with respect to selection. Therefore, the situation is once again formally equivalent to the one of an unstructured population subject to a single environmental shift (see Appendix C for details). For $\beta = 0$, the effective growth parameter of a mutant in the time interval $[(D-2)\vartheta, (D-1)\vartheta[$ is given by Eq. (2.11). The generalization to density-dependent mutant fitness is straightforward: we simply replace s by $S(t)$ and approximate $S(t)$ by its stationary value $\bar{S}$ (see Eq. (C.6)):

$$s_{D-1} = \alpha - 1 + \frac{1 - \alpha + \bar{S} + \alpha r}{D}(D-1). \tag{2.13}$$

As discussed above, if s_{D-1} is larger than zero and ϑ large, the mutant type will certainly establish before the last environmental shift. If s_{D-1} is smaller than zero, a mutation-selection equilibrium will evolve. With increasing ϑ, the total probability of evolutionary rescue hence converges to a limit value. The speed of convergence is set by the time to approach equilibrium.

Figure 2.8 compares an instantaneous shift in all demes ($\vartheta = 0$) to a very slowly progressing change ($\vartheta = \infty$). While for the parameter values chosen for Figure 2.8A, a slower change is better, we see in Figure 2.8B that the rescue probability can be higher for a very fast change than for a very slow one. How can we explain this? We give an illustrative numerical example for the case of two demes: With $r = 0.5$, we find a total equilibrium wildtype population size of

$4K/3$ before the last deme turns bad at time ϑ. For $\beta = 2$, we obtain $\bar{S} = -0.137$, and if we choose $\alpha = 0.99$, $s_{D-1} = 0.174 > 0$, i.e., mutants have a good chance to establish. In contrast, for $\beta = 4$, we obtain $\bar{S} = -z = -0.685$ and with $\alpha = 0.99$, $s_{D-1} = -0.1$. This means that selection against the mutation is even stronger in the period $[0, \vartheta[$ than before time 0 because of the strong density dependence of growth in the new environment. Additionally, the number of wildtype individuals and thus the mutational input is lower. Consequently fewer mutants are present in the population at time ϑ than at time 0, which leads to a lower probability of evolutionary rescue for a very slowly progressing change as compared to an instantaneous degradation of the whole habitat. We note that for density-independent mutant fitness $\beta = 0$, $P_{\text{rescue}}(\vartheta = 0)$ is never larger than $P_{\text{rescue}}(\vartheta \to \infty)$; for $\alpha = \beta = 0$ and s small, $P_{\text{rescue}}(\vartheta = 0) \approx P_{\text{rescue}}(\vartheta \to \infty)$ (see Appendix C for a derivation).

Note that the strength of s_{D-1} increases with r whenever either α or β is larger than zero. In that case, a harsher change leads to a higher establishment probability for the rescue mutant. Consequently, if $\beta > 0$, we may observe a non-monotonic behavior of P_{rescue} even if the rescue mutation is lethal ($\alpha = 0$) under the old environmental conditions.

2.4.3 Evolutionary rescue in an island model without standing genetic variation

We now turn to an island model with arbitrary migration in order to include population structure. We assume that there is no standing genetic variation in the population for the locus in question ($\alpha = 0$), i.e., the mutation is lethal in the old environment. We also restrict our investigation to density-independent selection ($\beta = 0$).

The analysis is based on the following reasoning: due to the particular migration pattern of the island model, we can merge all patches in the old environment into one habitat and all patches with the new environment into a second one. The number of wildtypes in the old part of the habitat is given by its current carrying capacity; the number of wildtypes in the new part is governed by a single difference equation (Eq. (A.1)). As the mutation is lethal in the old environment ($\alpha = 0$), we effectively deal with one type of mutant individuals only. As in the Levene model, it is helpful to consider the effective growth parameter of a mutant. A mutation that arises in the new environment either stays there with a probability proportional to the size of the new habitat or migrates to the old

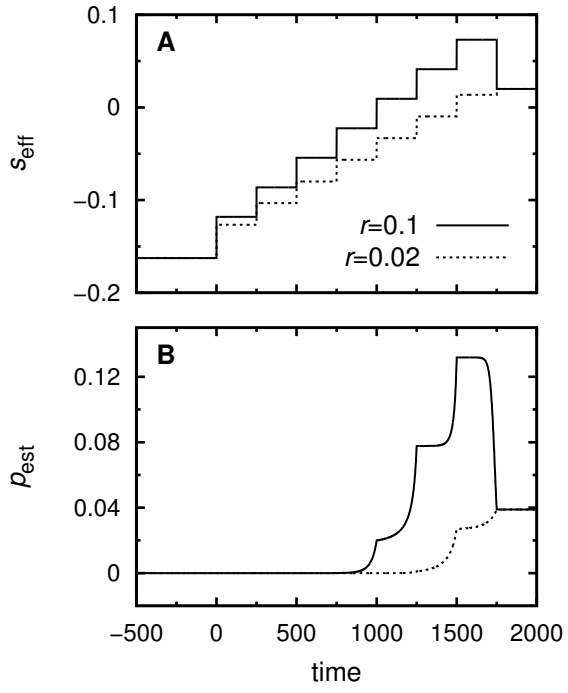


Figure 2.6: Effective mutant growth rate s_{eff} and establishment probability as a function of time for two values of r. Both quantities exceed their value for constant fitness s (far right) if r is large. The curves for s_{eff} and $p_{\text{est}}(t)$ follow Eq. (2.11) and Eq. (A.7). Parameter values: $s = 0.02$, $\beta = 0$, $\alpha = 0.85$, $D = 8$, $D\vartheta = 2000$, $K_{\text{rescue}} = 20000$, $u = 0.5 \cdot 10^{-4} = 1/K_{\text{total}}$.

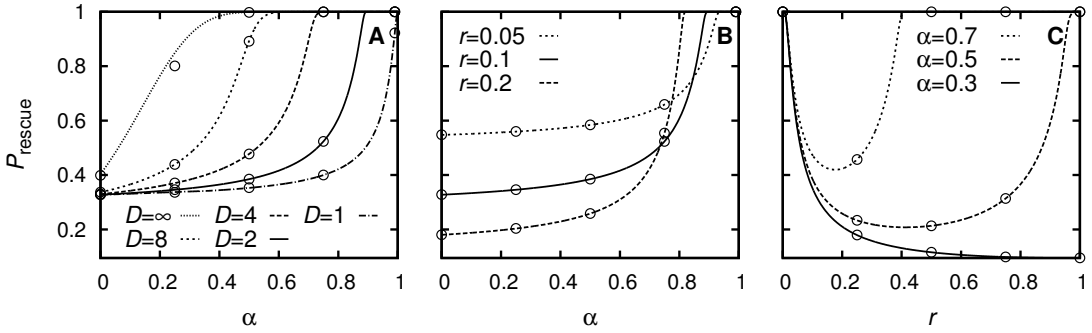


Figure 2.7: Evolutionary rescue in a Levene model. The probability of evolutionary rescue significantly increases with the fitness α of mutants in the unperturbed environment (Panel A). For sufficiently large values of α, it has a minimum for intermediate values of r (Panels B and C). If not specified otherwise, the parameter values are: $s = 0.02$, $\beta = 0$, $D = 2$, $K_{\text{total}} = 20000$, $u = 0.5 \cdot 10^{-4} = 1/K_{\text{total}}$, $\vartheta = 1000$; furthermore $r = 0.1$ for Panel A. The theoretial approximation for infinitely many islands is compared to simulations with $D = 100$. The theoretical curves are based on Eq. (C.5) (finite number of demes) and Eq. (E.12) (infinite number of demes). Circles denote simulation results.

environment with a probability proportional to the size of the latter. As it cannot survive in the old habitat, its growth rate is hence reduced by migration out of the new environment. Since the old habitat shrinks and the new habitat grows, the effective growth rate of the mutant increases in time. This implies that again, the early spread of the mutation can be described as a time-inhomogeneous branching process with a growth rate that gradually increases until the environmental conditions have changed on all islands. Following this reasoning, the effective growth parameter is given by

$$s_{\text{eff}}(t) = \begin{cases} (1+s)\left(1 - \frac{D-d}{D}m\right) - 1 & \text{for} \quad t \in [(d-1)\vartheta, d\vartheta[, \quad d \in \{1, \ldots, D-1\}, \\ s & \text{for} \quad t \geq (D-1)\vartheta\ . \end{cases} \tag{2.14}$$

Migration thus reduces the effective growth parameter of mutants and along with it their establishment probablity. At the same time, a large migration probability keeps the wildtype population size in the new environment high, which means that there is a large supply of new mutants that might possibly establish. Migration therefore has two antagonistic effects. As a result, the probability of evolutionary rescue has an intermediate maximum as a function of m (Figure 2.9A and

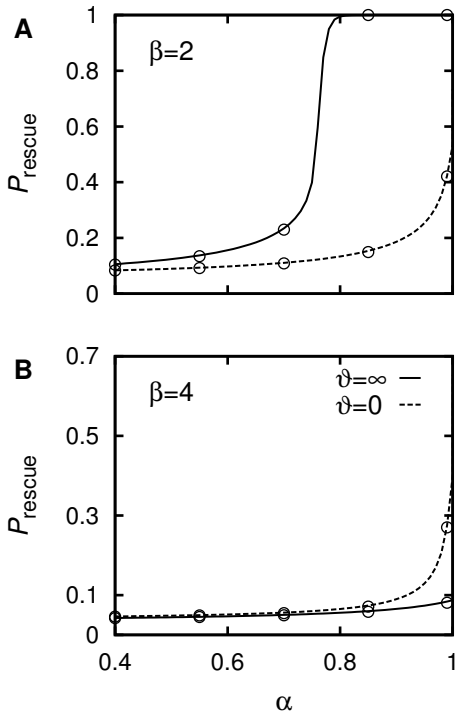


Figure 2.8: Evolutionary rescue in a Levene model. Comparing Panels A and B, we see that the survival probability of the population can be higher or lower for an instantaneous degradation of the whole habitat ($\vartheta = 0$) than for a very slowly progressing change ($\vartheta = \infty$). The parameter values are: $s = 0.411$, $z = 0.685$, $r = 0.5$, $D = 2$, $K_{\text{total}} = 20000$, $u = 0.5 \cdot 10^{-5} = 0.1/K_{\text{total}}$. The theoretical curves are based on Eq. (B.3) (with Eq. (C.8) for $\vartheta = \infty$). For the simulations, we chose $\vartheta = 5000$ to represent ∞. Circles denote simulation results.

Figure 2.10): For small m, the positive effect of migration dominates and P_{rescue} increases with m. As m gets larger, the negative effect gets stronger and finally prevails such that the probability of evolutionary rescue decreases. For details on the analysis, we point to Appendix D.

Figures 2.9B–D, illustrate the dynamics for small, intermediate and large values of m in more detail. For the sake of simplicity, we stick to a two deme model. Figure 2.9B shows the total number of residents $N_w^{(\text{new})}$ that live in the bad habitat, Figure 2.9C the establishment probability of a single mutant, and Figure 2.9D the rate of successful mutants. If individuals do not migrate at all, the number of wildtypes in the bad habitat exhibits two peaks: one when the first deme turns bad, the second when the second deme turns bad. The establishment probability is constant in time ($\approx 2s$). The rate of successful mutants therefore has two identical peaks, i.e., a rescue mutation will most likely arise briefly after the change of the environment has occurred in one of the demes. The other extreme case is $m = 1$. In that case, $N_w^{(\text{new})}$ is kept relatively high. However, the establishment probability is virtually zero up to a few generations before the second deme turns bad and then increases almost instantaneousy up to $2s$. Therefore, the rate of successful mutants has only one peak (Figure 2.9D). We find that this peak comprises approximately the same area as the two peaks for $m = 0$ together, leading to approximately the same probability of evolutionary rescue. For small r and $D = 2$, as used in Figure 2.9, this is intuitively clear: The total number of wildtype individuals stays close to K_{total} until the second deme degrades. Thus, the number of wildtypes that still exist after time ϑ (and can act as a source for rescue mutations), too, is twice as large as for $m = 0$ (see the peak in Figure 2.9B). For large r and $D > 2$, we still have $P_{\text{rescue}}(m = 0) \approx P_{\text{rescue}}(m = 1)$, although this is not obvious (see Appendix C for a derivation). For intermediate migration, both $N_w^{(\text{new})}$ and p_{est} assume non-negligible values in the period between the deterioration of the first and the second deme, leading to a significant rate of successful mutants. Consequently, the length of this period strongly influences the survival probability of the population.

The essential elements of this discussion can also be seen in the following simple approximation, which captures the characteristic behavior of $P_{\text{rescue}}(m)$: For small s, we can approximate the establishment probability of new mutants in $[0, \vartheta[$ by

$$p_{\text{est}} \approx \max\left[2s_{\text{eff}}(t), 0\right] = \max\left[2\left(s - \frac{m}{2}\right), 0\right]. \tag{2.15}$$

We furthermore approximate the mutational input during this period via the equilibrium wildtype population size in the deteriorated part of the habitat (cf. Eq. (D.16))

$$\hat{N}_w^{(\text{new})} = \frac{Km(1-r)}{m(1-r)+2r} \approx Km\frac{1-r}{2r}, \tag{2.16}$$

where the approximation is valid for $r \gg m$. This neglects the phases where the wildtype population size decays to its equlibrium or to zero, respectively, after the first and second deme deteriorates (the "peaks" in Figure 2.9D). We estimate this contribution by its value for $m = 0$ (cf. Eq. (2.7)) and obtain for the probability of evolutionary rescue:

$$P_{\text{rescue}} \approx 1 - \exp\left[-2u\vartheta \max\left[\left(s - \frac{m}{2}\right), 0\right]\frac{Km(1-r)}{2r} - \frac{2uK}{r}2s\right]. \tag{2.17}$$

This simple approximate is remarkably accurate for small m and s (see Figure D.1 in the Appendix). From its functional form, we see that the speed of deterioration set by ϑ has a strong effect on P_{rescue} if and only if m is in an intermediate range, such that both $\hat{N}_w^{(\text{new})}$ and p_{est} are non-negligible. We also see that the maximum in $P_{\text{rescue}}(m)$ is located at $m \approx s$. These results generalize to $D > 2$ (see Appendix D).

We close with some observations on how various parameters shape P_{rescue}. First, we observe that the number of demes D only has a moderate influence on population survival for $D \geq 2$ (Figure 2.10A). Further, the peak for intermediate m gets broader as s increases (Figure 2.10B): the negative effect of migration sets in later for larger s. Last, as already discussed above, the value of ϑ has a strong influence on the probability of evolutionary rescue unless migration is extremely weak or quite strong (Figure 2.10C).

2.5 Discussion

Severe environmental change can drive a population extinct unless it is able to rapidly adapt to the new conditions. Environmental change is ubiquitous and greatly enhanced by human interference. A profound understanding of how such change affects biodiversity might help to develop successful conservation strategies. On the other hand, awareness of the factors that promote rapid evolution is essential whenever we seek to inhibit it, such as in treatment plans to avoid the evolution of drug resistance. In scenarios of population extinction or evolu-

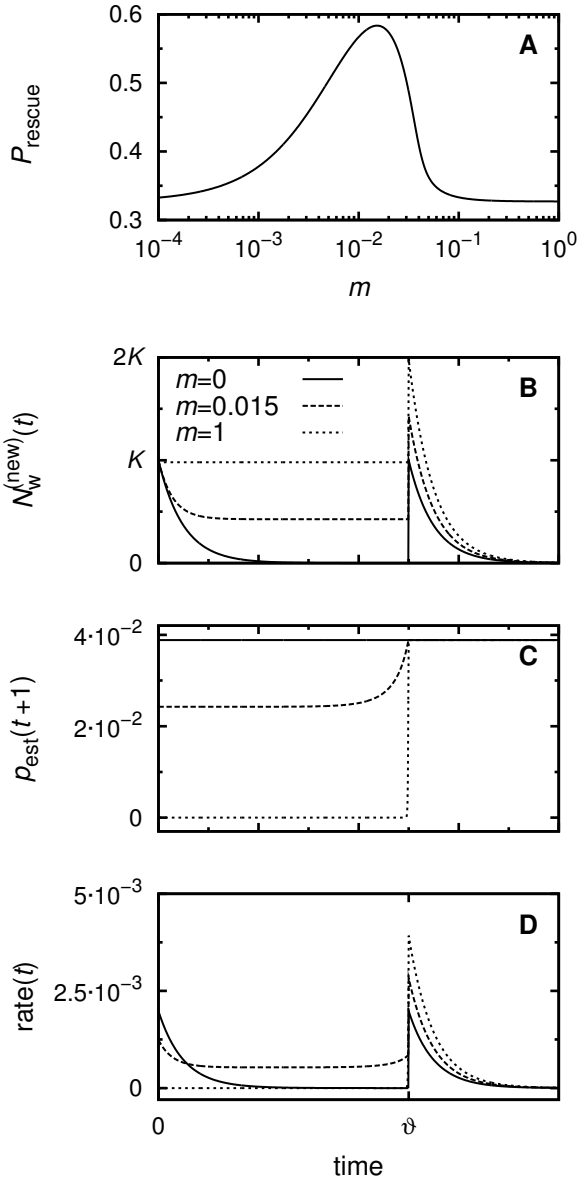


Figure 2.9: Panel A: Probability of evolutionarly rescue. The curve is based on Eq. (D.14). Panels B,C,D: Wildtype population size in the new habitat as obtained by Eq. (A.1), establishment probability (calculated following Eq. (A.7) with Eq. (2.14)), and rate of successful mutations for $m = 0$, $m = 0.015$, and $m = 1$. Panels B–D illustrate the existence of the maximum in Panel A. The parameter values are: $s = 0.02$, $\beta = 0$, $r = 0.01$, $D = 2$, $K_{\text{total}} = 20000$, $\vartheta = 1000$, $u = 0.5 \cdot 10^{-5} = 0.1/K_{\text{total}}$.

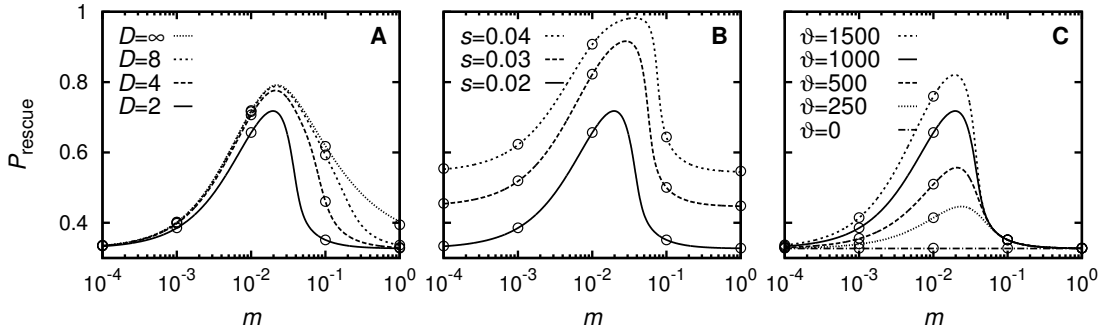


Figure 2.10: Probability of evolutionary rescue in an island model without standing genetic variation (mutant allele is lethal in the old environment). One observes a maximum for small values of m. If not specified otherwise in the figure legend the parameter values are: $s = 0.02$, $\beta = 0$, $r = 0.1$, $D\vartheta = 2000$, $K_{\text{total}} = 20000$, $u = 0.5{\cdot}10^{-4} = 1/K_{\text{total}}$. The analytical approximation for infinitely many islands is compared to simulations with $D = 40$. The theoretical curves are based on Eq. (D.14) (finite number of demes) and (E.15) (infinitely many demes). Circles denote simulation results.

tionary rescue, evolution and ecology are necessarily intertwined. For a thorough assessment of risks and chances, it is therefore indispensable to be mindful of both, evolution and ecology.

Many ecological alterations will not affect the whole habitat at once, but propagate gradually across the species range. In that case, parts of the population still experience the old environment, to which it is well-adapted, while others already face the new unfavorable conditions. The system corresponds to a source-sink-system with a shrinking – finally disappearing – source and a growing sink. Despite the obvious importance of population structure and gradual habitat deterioration for the probability of population survival, these aspects have not been considered in ecologically explicit models on evolutionary rescue so far. In this paper, we provide a baseline model of evolutionary rescue in structured populations and investigate the implications for the probability of evolutionary rescue. In addition, we allow for density-dependent mutant fitness in the new environment. On the other hand, we restrict ourselves to the most basic genetic model of one locus with two alleles. Because of its simplicity, this basic genetic model can provide insight into the fundamental mechanisms underlying evolutionary rescue in ecologically complex scenarios. In some situations, a simple

genetic basis may be appropriate. For example, mutation at a single locus can be sufficient to confer insecticide or drug resistance (Milani, 1963; McKenzie *et al.*, 1980; Daborn *et al.*, 2002; Gerstein *et al.*, 2012). Analysis of the model reveals several non-monotonic relationships and unexpected patterns. In this context, three quantitites – the speed of change, the severity of change, and the migration probability – are of particular interest.

The speed of change How does the speed at which the deterioration proceeds across the species range influence the probability of evolutionary rescue? If $m = 0$, the survival probability of the population is independent of the speed of change, as the sub-populations in the single demes suffer independent fates. For small migration rates, the probability of evolutionary rescue is drastically increased for a slow compared to a rapid change. This is because for sufficiently weak migration, wildtype individuals are rare in the altered habitats, competition is weak, and mutations have a non-negligiable establishment probability, and the slower the change, the more mutants are generated over time. For strong migration, a more complicated picture arises: While a slower change is still often favorable, a slow deterioration of the habitat can sometimes hamper adaptation of the population if mutant fitness is strongly density dependent. We can understand this unexpected behavior as follows. For a slow change, the population encounters extended periods of environmental stasis before the last demes deteriorate. If migration is strong, the number of wildtype individuals in the perturbed demes remains relatively high during this time. As a consequence, mutant fitness, if density-dependent, can be lower than it had been before the environment started deteriorating. Furthermore, the total wildtype population size and hence the number of new mutants per generation is lower than it was before time zero. The total number of mutants that are maintained in the balance of mutation, selection, and migration can therefore even be reduced relative to the number of mutants in the standing genetic variation before the environmental detoriation sets in. As a consequence, long periods of environmental stasis can lead to a reduced probability of evolutionary rescue. For density-independent mutant fitness, however, our results show that a slow change is always at least as good as a fast change (cf. Appendix C).

The severity of change The probability of successful adaptation is usually expected to decrease with increasing maladpatation of the resident population (cf. Holt and Gomulkiewicz (2004) for an overview): the slower the decay

of the wildtype population size, the more time for adaptative mutations to occur. However, it is not enough that mutations arise, they also have to survive stochastic loss and establish. Due to competition, the growth rate of new mutants will often depend negatively on the population density. In that case, a fast decay – and thus a harsh change in the environment – increases the establishment probability of mutations. As a consequence, a harsher change (larger r) is not necessarily worse for population survival than a milder one. Instead, our results show that the probability of evolutionary rescue can assume a minimum for intermediate levels of wildtype maladaptation. We find that this occurs if (1) density dependence in the new environment is strong and either the amount of standing genetic variation large or selection for the mutant strong relative to migration and the speed of change slow or if (2) migration is strong and the mutant reasonably fit in the old environment. The entire parameter space where an inverse dependence on the severity of environmental change is observed is generally larger than the parameter space where faster speed leads to higher rates of rescue. The advantage of a harsh change is that a fast decrease in the number of wildtype individuals relaxes competition, either in the new environment (scenario 1) and/or in the old environment (scenario 2). For panmictic populations without habitat structure, the first scenario has been described verbally by Read *et al.* (2011). In that case, rescue is likely for harsh environmental change because a fast decay of the wildtype population size enhances the establishment probability of mutations from standing genetic variation. In structured populations, not only mutants from standing genetic variation but also mutant descendants of immigrants from unperturbed to perturbed demes benefit from reduced competition in the new environment and experience a higher establishment probability for a more severe change. If standing genetic variation is low and rescue relies mainly on de-novo mutations, the probabiliy of evolutionary rescue may decrease again as r increases further because the reduced mutational input outweighs the benefits of a high establishment probability. The first scenario is also reminiscent of studies with stable source-sink dynamics, where constant mutational input from the source and low competition in the sink can lead to high rates of adaptation (Greulich *et al.*, 2012; Hermsen *et al.*, 2012). In the second scenario with high migration and a reasonable amount of standing variation, the positive effect of a harsh change arises for similar reasons. Fundamentally, absolute mutant fitness - and this is what decides the fate of mutations – is density dependent as a consequence of simple population regulation in the unperturbed demes. This holds true even if the relative fitness α in the old environment and the fitness

$1 + S_i$ in the new environment are density independent. When the change is harsh, wildtype individuals get depleted in the perturbed part of the habitat every generation. Hence, migration between the old and the new part of the habitat is strongly unbalanced with increasing effect for stronger migration. As a consequence, competition in the old part of the habitat is relaxed and the absolute fitness of mutants accordingly enhanced. In this context, note that in our model, demes in the original environmental state get filled up to carrying capacity every generation; the genetic composition is determined by binomial sampling. This represents a certain overidealization and requires that offspring numbers are large enough. At least for not too large deme numbers D, this requirement is, however, easily fulfilled.

Awareness that intermediate environments may represent harder challenges for adaptation could be of striking importance for the design of drug treatment strategies: our results imply that a fast eradication of the pathogen might not necessarily be the best strategy to avoid drug resistance. As discussed above, in the first scenario without habitat structure, the effect is only observed if mutant fitness is strongly density dependent, which may or may not apply to a particular species. Read *et al.* (2011) recently discussed this idea in the context of malaria. Their arguments are based on a series of data sets showing that standing genetic variation is usually high in malaria infections and that a fast eradication of the drug sensitive pathogen allows rare resistant types to quickly amplify. This discussion of experimental evidence lends empirical support to our theory, which in turn quantifies the effect. Similarly, a recent study by Peña-Miller *et al.* (2013) combining deterministic mathematical models and experimental evolution in *E. coli* shows that competitive release due to a harsh treatment with a mixture of two drugs promotes the rapid emergence of drug-resistant strains.

The migration probability The dependence of the rescue probability on migration is shaped by four effects: First, migration is advantageous because the old part of the habitat acts as a source for wildtype individuals that might possible mutate. Second, mutants migrate to the old habitat where they have a disadvantage with respect to the wildtype. Migration thus reduces the effective growth rate of mutants. With increasing migration, this effect outweighs the first one such that the rescue probability has a local maximum for intermediate migration. Third, when migration gets very strong, the effect of relaxed competition in the old demes sets in and the rescue probability can become again high. This latter effect only happens when the relative mutant fitness α in the old habitat, the

mutant fitness $1 + S_i(t)$ in the new habitat, and the decay rate r of the wildtype population size in the new habitat are sufficiently large (the parameter range increases with increasing D). Finally, if mutant fitness in the perturbed demes is density dependent, migration leads to a reduced fitness of mutants in these demes. This counteracts the effect of relaxed competition in the old part of the habitat and can entail a very low probability of evolutionary rescue for strong migration. The interplay of all four forces can lead to a surprisingly complex dependence of the rescue probability on the migration probability (see Figures 2.2 and 2.4). A maximum in the probability of successful adaptation for intermediate migration rates has been found in previous studies for partially related, partially different reasons: Gomulkiewicz *et al.* (1999) analyse the potential of local adaptation in a sink, which is coupled to a source. Unlike in our model, the source never degrades and the population hence never dies out. The focus of interest is niche evolution within the sink. Immigration is necessary to provide targets for mutation to act on (our first effect). The disadvantage of high immigration arises, because absolute mutant fitness in the sink is assumed to be density dependent and thus decreases for increasing immigration (our fourth effect). If immigration is too strong, absolute mutant fitness is depressed below one such that the mutation cannot spread at all. Note, however, that in our model the local maximum in the rescue probability exists even if selection is density independent in the new habitat due to migration of mutants out of the perturbed demes (our second effect). Emigration of mutants out of the sink is not taken into account in Gomulkiewicz *et al.* (1999). Pease *et al.* (1989) analyse a model of population persistence in a spatially continuous habitat. An optimum moves in space, and the population has both the possibility to adapt and to follow the optimum by migration. As space is continuous, the quantity of interest is not the migration probability (which is one), but the mean square distance that individuals travel per generation. Similar to our model, migration is harmful in that it brings individuals to unfavorable places (cf. also Kirkpatrick and Peischl, 2013). In contrast to our scenario, however, the advantage of migration arises because it is needed for the population to keep track of the moving optimum. The optimal amount of migration (in the sense of optimal mean square displacement) depends on the speed of the optimum and the additive genetic variance of the population among other factors. Likewise, an intermediate dispersal distance maximizes the probability of evolutionary rescue in a recent simulation study by Schiffers *et al.* (2013). The model can be seen as complementary to ours: Schiffers *et al.* (2013) consider a population in a deteriorating heterogeneous environment.

However, the factors creating the spatial heterogeneity differ from the factors that cause the temporal degradation. In contrast to our model, the heterogeneity is thus stable in time, the gradual deterioration homogeneously affects the entire habitat. Adaptation relies on a reasonably large number of loci (15 loci per trait), making the model similar to a quantitative genetics model. Within this framework, migration has two antagonistic effects: as in our model, migration brings individuals to regions in which they are maladapted. Contrary to our model, however, this effect does not act via alleles that provide adaptation to the changing conditions – these alleles are adaptive everywhere –, but on alleles that determine local adaptation to the stable heterogeneity. On the other hand, the patches are so small in SCHIFFERS *et al.* (2013) that migration is necessary for the establishment of mutations that adapt the population to the globally degraded environment. This effect does not appear in our model as deme sizes are sufficiently large that the mutation can locally establish.

Limitations and extensions of our model Our analysis has several important limitations. First, our population follows the migration scheme of an island model. In particular, this implies that dispersal is global. In order to arrive at a comprehensive picture of evolutionary rescue in structured populations, local dispersal and isolation by distance should be included into the model as a next step. Experimentally, a comparsion of the impact of local versus global dispersal for weak migration has been highlighted by BELL and GONZALEZ (2011), showing that the rate of environmental deterioration influences how the dispersal mode affects rescue. In this context, not only a model with discrete patches, but also a model in continuous space is of interest. In experimental evolution, serial transfer of individuals with small inoculum sizes represent regular catastrophes, which would need to be included in the model for a quantitative comparison of experimental data and theory (see MARTIN *et al.*, 2013). Although in some cases evolutionary rescue relies on single mutations, the simple genetic basis of adaptation is a major restriction of our study. Often, adaptation is more complex, and mutations at several loci are required to restore fitness above one. Types with only some of these mutations might potentially even perform worse (or at least not better) than the wildtype. Adaptation then includes stochastic tunneling: an inferior (or neutral) genotype is generated and has to produce a fitter mutant before it goes extinct (IWASA *et al.*, 2004b; WEISSMANN *et al.*, 2009; LYNCH and ABEGG, 2010; MARTIN *et al.*, 2013). For panmictic populations, evolutionary rescue requiring stochastic tunneling has been considered by IWASA *et al.* (2003)

and Iwasa *et al.* (2004a) with a focus on biomedical applications (evolution of drug resistance, escape of tumor cells from chemotherapy etc.). Last, many loci might contribute to adaptation. In that case, a quantitative genetics approach suggests itself, but also models with an explicit genetic basis of the trait would be valuable (cf. Boulding and Hay (2001) and Schiffers *et al.* (2013)).

In conclusion, our results confirm the importance of ecological factors, specifically habitat structure and density-dependent fitness for the probability of evolutionary rescue. They provide insight into how various mechanisms intertwine to decide the race between population decline and adaptive evolution. As we have seen, this interplay of mechanisms can lead to surprising patterns in the probability of evolutionary rescue, with rapid changes in the environment (small ϑ) sometimes being easier for evolutionary rescue and sometimes harder and with rescue showing non-monotonic relationships with both migration probability (m) and the severity of the environmental perturbation (r).

Acknowlegdments

H.U. thanks the Otto lab and the Timme lab in Göttingen for helpul discussions and hospitality. We furthermore thank Jitka Polechová, Peter Pfaffelhuber, and Reinhard Bürger for helpful discussions and Troy Day, Sean Rice, and two anonymous referees for valuable comments on the manuscript. This work was made possible with financial support by the Vienna Science and Technology Fund (WWTF), grant MA6-01, by the Deutsche Forschungsgemeinschaft (DFG), Research Unit 1078 *Natural selection in structured populations*, by a Discovery Grant from the Natural Science and Engineering Research Council of Canada, and by travel grants from the Universities of Vienna and British Columbia.

A Appendix: General notes on the analysis

The wildtype population size We model the dynamics of the wildtype population size deterministically and assume that mutants are rare enough to be ignored. For the sub-models considered in this paper, it is not necessary to determine the number of wildtype individuals in each single deme. It is sufficient to determine the total number of wildtypes in the unperturbed (or "old") and the total number of wildtypes in the deteriorated (or "new") part of the habitat. Let $N_w^{(\mathrm{new})}(t)$ be the total number of wildtype individuals that live in the new habitat

at time t before migration and selection. Using Eq. (2.3) and ignoring the effects of mutation, for $t+1>0$:

$$\begin{aligned} E[N_w^{(\text{new})}(t+1)|N_w^{(\text{new})}(t)] =& (1-r)\left(1-m+\frac{d_t}{D}m\right)N_w^{(\text{new})}(t) \\ &+ m\frac{d_t}{D}(1-r)(D-d_t)K \\ &+ K\delta\Big((t+1)\bmod\vartheta\Big), \end{aligned} \tag{A.1}$$

where d_t is the number of demes in the new environmental state at time t and $\delta(0)=1$ and $\delta(x)=0$ otherwise. The function δ takes into account that every ϑ generations a new island turns bad, bringing with it approximately K new wildtype individuals. As demes in the old environmental state get filled up tp carrying capacity every generation, the number of wildtypes in the old part of the habitat is (ignoring rare mutant individuals)

$$N_w^{(\text{old})}(t) = K(D-d_t).$$

Establishment probabilities As explained in the main text, we restrict our analytical results to scenarios where we deal with only one type of mutant individual. Across the entire life cycle (except for density regulation), each mutant produces a Poisson distributed number of offspring with mean $1+s_{\text{eff}}(t)$, where the effective growth parameter $s_{\text{eff}}(t)$ depends on the specific scenario. To make use of analytical theory, we approximate the discrete-time branching process by a continuous-time branching process. As selection can be strong in scenarios of population decline and evolutionary rescue, details matter in the transition from discrete to continuous time. For the continuous-time branching process, we use the following per capita birth and death rates:

$$\lambda(t) = 0.5 + 0.5\cdot\text{sign}(\ln(1+s_{\text{eff}}(t)))\cdot\min[|\ln(1+s_{\text{eff}}(t))|,1], \tag{A.2a}$$

$$\mu(t) = 0.5 - 0.5\cdot\text{sign}(\ln(1+s_{\text{eff}}(t)))\cdot\min[|\ln(1+s_{\text{eff}}(t))|,1], \tag{A.2b}$$

and define

$$\hat{s}_{\text{eff}}(t) := \text{sign}(\ln(1+s_{\text{eff}}(t)))\cdot\min[|\ln(1+s_{\text{eff}}(t))|,1]. \tag{A.3}$$

With the logarithm, we assure that the average long-term growth $\hat{s}_{\text{eff}}(t) = \lambda(t) - \mu(t)$, is the same as in the discrete-time process. The restriction to values between

-1 and 1 is necessary for rates to remain non-negative. Last, drift has to be scaled appropriately. In the continuous-time process, the sum $\lambda(t) + \mu(t)$ measures the strength of drift. In the diffusion limit, $\lambda(t) + \mu(t)$ must be one in order to match continuous-time and discrete-time dynamics. This leaves some freedom for the incorporation of selection (affecting the death rate or the birth rate or both). While this choice is irrelevant in the diffusion limit (and hence for weak selection), it matters if the growth parameter is large as it can be in our model. Comparison to computer simulations shows that the best agreement is obtained if we equally distribute it between the death and the birth rate, as above.

If $\exp\left[-\int_T^t \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau\right] \underset{t\to\infty}{\longrightarrow} 0$, the establishment probability of a mutation arising at time T is given by (Uecker and Hermisson, 2011)

$$p_{\text{est}}(T) = \frac{2}{1 + \int_T^\infty \exp\left[-\int_T^t \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau\right]\mathrm{d}t}. \tag{A.4}$$

In the following sections, we will encounter effective growth rates that change stepwise in time:

$$s_{\text{eff}}(t) = \begin{cases} s_0 & \text{for} \quad t < 0, \\ s_l & \text{for} \quad t \in [(l-1)\Phi, l\Phi[, \quad l \in \{1, \dots, L-1\}, \\ s_L = s & \text{for} \quad t \geq (L-1)\Phi, \end{cases} \tag{A.5}$$

where the steps occur at regular intervals Φ, depending on the model. $\hat{s}_{\text{eff}}(t)$ is defined accordingly, with:

$$\begin{aligned} &\hat{s}_i = \text{sign}(\ln(1+s_i)) \cdot \min(|\ln(1+s_i)|, 1) \quad \text{for} \quad i \in 0, \dots, L, \\ &\hat{s} = \hat{s}_L. \end{aligned} \tag{A.6}$$

Assuming that $\hat{s}_k \neq 0$ for all $k \in 0, \dots, L$, we obtain (see below for a derivation):

$$p_{\text{est}}(T) = \begin{cases} \frac{2}{1+I_0(T)} & \text{for} \quad T < 0, \\ \frac{2}{1+I_l(T)} & \text{for} \quad T \in [(l-1)\Phi, l\Phi[, \\ \frac{2\hat{s}_L}{1+\hat{s}_L} & \text{for} \quad T \geq (L-1)\Phi \end{cases} \tag{A.7}$$

with

$$I_l(T) = \frac{1}{\hat{s}_l} + \exp\left[\hat{s}_l \Delta T_l\right] \sum_{k=l}^{L-1} \frac{\hat{s}_k - \hat{s}_{k+1}}{\hat{s}_k \hat{s}_{k+1}} \exp\left[-\sum_{j=l}^{k} \hat{s}_j \Phi\right], \tag{A.8}$$

where $\Delta T_l = T - (l-1)\Phi$. For $l > 0$, ΔT_l is the time that has elapsed since the lth island deteriorated. If one or more of the $\hat{s}_k$ are 0, the result is obtained by taking the limit $\hat{s}_k \to 0$.

We turn to the derivation of Eq. (A.7). We need to evaluate the integral

$$\int_T^\infty \exp\left[-\int_T^t \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau\right]\mathrm{d}t. \tag{A.9}$$

For $T \geq (L-1)\Phi$, the calculation is straightforward. So, we focus on $T < (L-1)\Phi$. We assume throughout the derivation that $\hat{s}_k \neq 0$ for all $k \in 0, \ldots, L$. For $T < l\Phi$, if $l = 0$, or $T \in [(l-1)\Phi, l\Phi[$, if $l \in \{1, \ldots, L-1\}$, we have:

$$\begin{aligned} I_l(T) = &\int_T^\infty \exp\left[-\int_T^t \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau\right]\mathrm{d}t = \\ &\int_T^{l\Phi} \exp\left[-\int_T^t \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau\right]\mathrm{d}t + \sum_{k=l+1}^{L-1} \int_{(k-1)\Phi}^{k\Phi} \exp\left[-\int_T^t \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau\right]\mathrm{d}t \\ &+ \int_{(L-1)\Phi}^\infty \exp\left[-\int_T^t \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau\right]\mathrm{d}t. \end{aligned} \tag{A.10}$$

The first integral gives

$$\begin{aligned} \int_T^{l\Phi} \exp\left[-\int_T^t \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau\right]\mathrm{d}t = \int_T^{l\Phi} \exp\left[-\int_T^t \hat{s}_l \mathrm{d}\tau\right]\mathrm{d}t = \int_0^{l\Phi - T} \exp\left[-\hat{s}_l t\right]\mathrm{d}t \\ = \frac{1 - \exp\left[-\hat{s}_l(l\Phi - T)\right]}{\hat{s}_l}. \end{aligned} \tag{A.11}$$

The components of the sum ($k \in \{l+1, \ldots, L-1\}$) are:

$$\begin{aligned}
&\int_{(k-1)\Phi}^{k\Phi} \exp\left[-\int_{T}^{t} \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau\right]\mathrm{d}t \\
&= \int_{(k-1)\Phi}^{k\Phi} \exp\left[-\int_{T}^{l\Phi} \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau - \sum_{j=l+1}^{k-1}\int_{(j-1)\Phi}^{j\Phi} \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau - \int_{(k-1)\Phi}^{t} \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau\right]\mathrm{d}t \\
&= \exp\left[-\hat{s}_l(l\Phi - T) - \sum_{j=l+1}^{k-1} \hat{s}_j\Phi\right]\int_{0}^{\Phi} \exp\left[-\hat{s}_k t\right]\mathrm{d}t \\
&= \exp\left[-\hat{s}_l(l\Phi - T) - \sum_{j=l+1}^{k-1} \hat{s}_j\Phi\right]\frac{1-\exp\left[-\hat{s}_k\Phi\right]}{\hat{s}_k}.
\end{aligned} \tag{A.12}$$

For the last integral, we obtain:

$$\begin{aligned}
&\int_{(L-1)\Phi}^{\infty} \exp\left[-\int_{T}^{t} \hat{s}_{\text{eff}}(\tau)\mathrm{d}\tau\right]\mathrm{d}t \\
&= \int_{(L-1)\Phi}^{\infty} \exp\left[-\int_{T}^{l\Phi} \hat{s}_l\mathrm{d}\tau - \sum_{j=l+1}^{L-1}\int_{(j-1)\Phi}^{j\Phi} \hat{s}_j\mathrm{d}\tau - \int_{(L-1)\Phi}^{t} \hat{s}_L\mathrm{d}\tau\right]\mathrm{d}t \\
&= \exp\left[-\hat{s}_l(l\Phi - T)\right]\exp\left[-\sum_{j=l+1}^{L-1} \hat{s}_j\Phi\right]\frac{1}{\hat{s}_L}.
\end{aligned} \tag{A.13}$$

We now use the transformation $T \to \Delta T_l = T - (l-1)\Phi$. With this, we obtain for $l \in \{0, \ldots, L-1\}$:

$$
\begin{aligned}
I_l(T) = & \frac{1}{\hat{s}_l} - \frac{1}{\hat{s}_l} \exp\left[\hat{s}_l \Delta T_l\right] \exp\left[-\hat{s}_l \Phi\right] \\
& + \exp\left[\hat{s}_l \Delta T_l\right] \sum_{k=l+1}^{L-1} \exp\left[-\sum_{j=l}^{k-1} \hat{s}_j \Phi\right] \frac{1 - \exp\left[-\hat{s}_k \Phi\right]}{\hat{s}_k} \\
& + \exp\left[\hat{s}_l \Delta T_l\right] \exp\left[-\sum_{j=l}^{L-1} \hat{s}_j \Phi\right] \frac{1}{\hat{s}_L} \\
= & \frac{1}{\hat{s}_l} - \frac{1}{\hat{s}_l} \exp\left[\hat{s}_l \Delta T_l\right] \exp\left[-\hat{s}_l \Phi\right] \\
& + \exp\left[\hat{s}_l \Delta T_l\right] \sum_{k=l}^{L-2} \exp\left(-\sum_{j=l}^{k} \hat{s}_j \Phi\right) \frac{1 - \exp\left[-\hat{s}_{k+1} \Phi\right]}{\hat{s}_{k+1}} \\
& + \exp\left[\hat{s}_l \Delta T_l\right] \exp\left[-\sum_{j=l}^{L-1} \hat{s}_j \Phi\right] \frac{1}{\hat{s}_L} \\
= & \frac{1}{\hat{s}_l} - \frac{1}{\hat{s}_l} \exp\left[\hat{s}_l \Delta T_l\right] \exp\left[-\hat{s}_l \Phi\right] \\
& + \exp\left[\hat{s}_l \Delta T_l\right] \sum_{k=l}^{L-2} \left[\frac{\exp\left[-\sum_{j=l}^{k} \hat{s}_j \Phi\right] - \exp\left[-\sum_{j=l}^{k+1} \hat{s}_j \Phi\right]}{\hat{s}_{k+1}} \right] \\
& + \exp\left[\hat{s}_l \Delta T_l\right] \exp\left[-\sum_{j=l}^{L-1} \hat{s}_j \Phi\right] \frac{1}{\hat{s}_L} \\
= & \frac{1}{\hat{s}_l} + \exp\left[\hat{s}_l \Delta T_l\right] \sum_{k=l}^{L-1} \frac{\hat{s}_k - \hat{s}_{k+1}}{\hat{s}_k \hat{s}_{k+1}} \exp\left[-\sum_{j=l}^{k} \hat{s}_j \Phi\right],
\end{aligned}
\tag{A.14}
$$

as given by Eq. (A.8).

B Appendix: Panmictic populations with $D = 1$ and scenarios where habitat structure is immaterial

We first treat the case $D = 1$ and show at the end of the section that the results coincide with the results for $D > 1$, replacing K by K_{total}, if either $m = 0$ or $\vartheta = 0$.

After the shift in the environment, the wildtype population size decays geometrically:

$$N_w^{(\text{total})}(t) = \begin{cases} K & \text{for} \quad t < 0, \\ K(1-r)^t & \text{for} \quad t \geq 0. \end{cases} \tag{B.1}$$

As we model the population size deterministically, the selection coefficient, too, becomes a deterministic function of time:

$$s_{\text{eff}}(t) = \begin{cases} \alpha - 1 = z_0 & \text{for} \quad t < 0, \\ S(N_w^{(\text{total})}(t)) = S(t) & \text{for} \quad t \geq 0. \end{cases} \tag{B.2}$$

For the calculation of establishment probabilities, we approximate $s_{\text{eff}}(t)$ in continuous-time as a stepped function with each step lasting one generation. We can then use Eq. (A.7) with $\Phi = 1$ to calculate the establishment probability of a mutation, setting $s_{\text{eff}}(t) = s$ (accordingly $\hat{s}_{\text{eff}}(t) = \hat{s}$) when $K(1-r)^t < 1$. Calculations based on a continuous change in $s_{\text{eff}}(t)$ (see Appendix E) work well for small r, but break down as r increases, since for large r the differences between discrete and continuous time dynamics become significant.

Following Eq. (2.5), the overall probability of evolutionary rescue reads:

$$P_{\text{rescue}} \approx 1 - \exp\left[- \sum_{t=-\infty}^{-1} uK\alpha p_{\text{est}}(t+1) - \sum_{t=0}^{\infty} uN_w^{(\text{total})}(t)(1+S(t))p_{\text{est}}(t+1) \right], \tag{B.3}$$

where for numerical evaluation we again set $N_w^{(\text{total})}(t) = 0$ when $K(1-r)^t < 1$.

For $\beta = 0$, formula (A.7) for the establishment probability reduces to

$$p_{\text{est}}(T) = \begin{cases} \frac{2}{1+\exp\left[-\hat{z}_0 T\right]\left(\frac{1}{\hat{s}}+\frac{1}{\hat{z}_0}\right)-\frac{1}{\hat{z}_0}} & \text{for} \quad T < 0, \\ \frac{2\hat{s}}{1+\hat{s}} & \text{for} \quad T \geq 0 \end{cases} \tag{B.4}$$

with

$$\hat{z}_0 = \max[-\ln(1 - z_0), 1]. \tag{B.5}$$

We can give an explicit formula for the probability of evolutionary rescue:

$$\begin{aligned} P_{\text{rescue}} &\approx 1 - \exp\left[-\sum_{t=-\infty}^{-1} uK\alpha p_{\text{est}}(t+1) - \sum_{t=0}^{\infty} uK(1-r)^t(1+s)\frac{2\hat{s}}{1+\hat{s}}\right] \\ &\approx 1 - \exp\left[-\int_{t=-\infty}^{0} uK\alpha p_{\text{est}}(t) - \sum_{t=0}^{\infty} uK(1-r)^t(1+s)\frac{2\hat{s}}{1+\hat{s}}\right] \\ &= 1 - \exp\left[-2uK\frac{\alpha}{1-\hat{z}_0}\ln\left(\frac{\hat{s}+\hat{z}_0}{(1+\hat{s})\hat{z}_0}\right) - u\frac{K}{r}(1+s)\frac{2\hat{s}}{1+\hat{s}}\right] \quad \text{for} \quad \hat{z}_0 \neq 1. \end{aligned} \tag{B.6}$$

For $\hat{z}_0 = 1$ (i.e., $\alpha \leq \exp[-1]$), we obtain:

$$P_{\text{rescue}} \approx 1 - \exp\left[-2\alpha uK\frac{\hat{s}}{1+\hat{s}} - u\frac{K}{r}(1+s)\frac{2\hat{s}}{1+\hat{s}}\right]. \tag{B.7}$$

Returning to Eq. (B.6), the respective contributions of mutations from standing genetic variation and de-novo mutations are given by

$$P_{\text{rescue}}^{\text{sgv}} \approx 1 - \exp\left[-2uK\frac{\alpha}{1-\hat{z}_0}\ln\left(\frac{\hat{s}+\hat{z}_0}{(1+\hat{s})\hat{z}_0}\right)\right] \tag{B.8}$$

$$\approx 1 - \exp\left[-2uK\ln\left(\frac{s+z_0}{z_0}\right)\right] \tag{B.9}$$

$$= 1 - \left(\frac{s+z_0}{z_0}\right)^{-2uK}, \tag{B.10}$$

$$P_{\text{rescue}}^{\text{dnm}} \approx 1 - \exp\left[-u\frac{K}{r}(1+s)\frac{2\hat{s}}{1+\hat{s}}\right] \approx 1 - \exp\left[-u\frac{K}{r}2s\right], \tag{B.11}$$

where the approximation is valid for small s and z_0. For small s and z_0, our results coincide with formulas (3) and (5) in Orr and Uncкless (2008) (note that the absolute fitness of a mutant is $1+s$ in our model, while it is $1+s_b-r$ in Orr and Uncкless (2008); z_0 corresponds to s_d in Orr and Uncкless (2008)) and are similar to formula (8) for $P_{\text{rescue}}^{\text{sgv}}$ in Hermisson and Pennings (2005).

For $\beta > 0$, the formula for the probability of evolutionary rescue does not reduce to a compact expression. Evaluation of the complex formula and com-

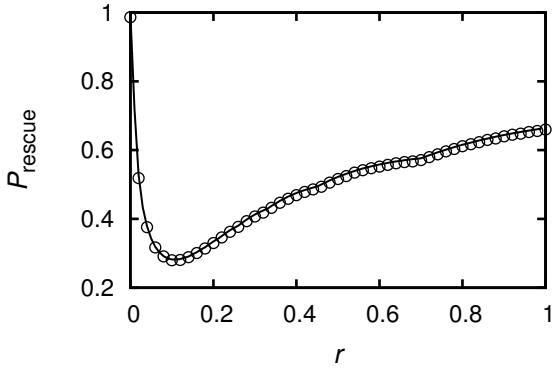


Figure B.1: Probability of evolutionary rescue as a function of r with a single deme ($D = 1$). The plot shows a detailed comparison between theory and simulation results. The parameters are chosen as in Figure 2.5D. The theoretical curve is based on Eq. (B.3) and Eq. (A.7). Simulation results are denoted by circles. Each simulation point is the average of 10^6 replicates.

parison to computer simulations shows that it yields highly accurate results. In particular, Figure B.1 demonstrates that the kinks in the graphs are not an artefact of our analytical approximation, but that the theory accurately reproduces the correct behavior. The existence of kinks can be understood if we consider the generation T_c at which for the first time $S(t) > -z$:

$$T_{\text{c}} = \max\left[0, \lfloor \frac{1}{r} \ln\left(\frac{s\beta}{s+z}\right) \rfloor + 1\right], \tag{B.12}$$

where $\lfloor \cdot \rfloor$ denotes the floor function, which maps a real number to the largest previous integer. T_c only takes discrete values and therefore jumps as a function of r. As a consequence, $P_{\text{rescue}}(r)$ is not everywhere differentiable.

We pointed out in the main text that the decay of the rescue probability as a function of r is not completely monotonic in Figure 2.5C. This can be seen in more detail in Figure B.2, which zooms in on larger r. A slight local minimum exists at $r = 0.7$. This is precisely the point where $T_c = T_c(r)$ jumps from 1 to 2 giving a little advantage to values of r larger than 0.7.

In the main text, we have discussed scenarios where the probability of evolutionary rescue either decays with r or exhibits a mimimum for intermediate values of r. In addition to these patterns, a third pattern is possible: the probability of

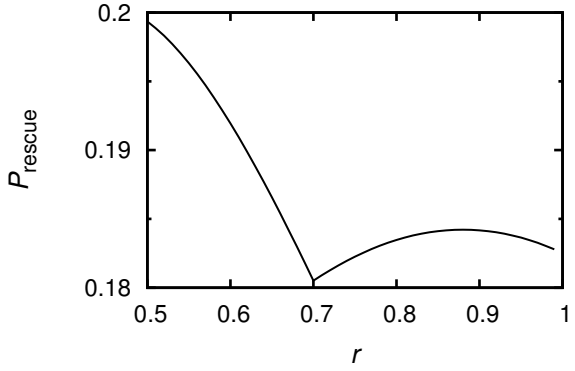


Figure B.2: Probability of evolutionary rescue as a function of r with a single deme ($D = 1$). The plot expands Figure 2.5C, for large values of r, showing a minimum for $r = 0.7$. The theoretical curve is based on Eq. (B.3) and Eq. (A.7).

evolutionary rescue attains a minimum, then a maximum, and decays afterwards (Figure B.3). This pattern can arise, because the probability that a mutation generated between time 0 and τ_0, at which $S(t)$ turns from negative to positive, rescues the population has a maximum for intermediate r: For large r, only few mutations are generated; for small r, they have a low establishment probability. If this maximum is pronounced enough, it shapes the overall curve. This is the case if β is extremely large such that the period between 0 and τ_0 is long and z small such that the establishment probability is high. The maximum gets masked if α is very large. The overall effect on the curve is generally weak, however.

It remains to prove that the results for a structured population with $m = 0$ or $\vartheta = 0$ reduce to the unstructured case with $D = 1$ (replacing K by K_{total}). To do so, we consider the general formula Eq. (2.5).

We start with $m = 0$: The dynamics in the single demes are then independent from each other. Thus, $N_w^{(i)\prime}(t) = N_w^{(1)\prime}(t-(i-1)\vartheta)$ and $S_i(t) = S_1(t-(i-1)\vartheta)$.

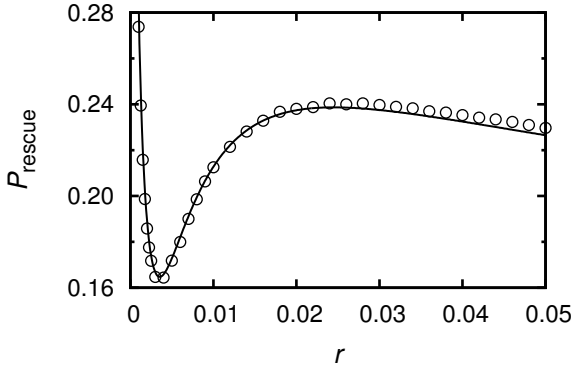


Figure B.3: Probability of evolutionary rescue in an unstructured population ($D = 1$). We see that the probability of evolutionary rescue can attain a minimum, followed by a maximum. Parameter values: $\alpha = 0.9$, $z = 0.005$, $s = 0.01$, $\beta = 40$, $K = 10^6$, $u = 1 \cdot 10^{-6} = 1/K$. The theoretical curve is based on approximation (E.1). Circles denote simulation results. Each simulation point is the average of 10^6 replicates.

The latter implies furthermore $p_{\text{est}}^{(i)}(t) = p_{\text{est}}^{(1)}(t - (i-1)\vartheta)$. Plugging this into the formula for P_{rescue} yields:

$$\begin{aligned}
P_{\text{rescue}} &\approx 1 - \exp\Bigg[-\sum_{i=1}^{D}\Bigg(\sum_{t=-\infty}^{(i-1)\vartheta-1} u\alpha K p_{\text{est}}^{(1)}(t-(i-1)\vartheta+1) \\
&\qquad + \sum_{t=(i-1)\vartheta}^{\infty} u\left(1+S_1(t-(i-1)\vartheta)\right) N_w^{(1)\prime}(t-(i-1)\vartheta) p_{\text{est}}^{(i)}(t-(i-1)\vartheta+1)\Bigg)\Bigg] \\
&= 1 - \exp\left[-\sum_{i=1}^{D}\left(\sum_{t=-\infty}^{-1} u\alpha K p_{\text{est}}^{(1)}(t+1) + \sum_{t=0}^{\infty} u(1+S_1(t)) N_w^{(1)\prime}(t) p_{\text{est}}^{(1)}(t+1)\right)\right] \\
&= 1 - \exp\left[-\sum_{t=-\infty}^{-1} u\alpha D K p_{\text{est}}^{(1)}(t+1) + \sum_{t=0}^{\infty} u(1+S_1(t)) D N_w^{(1)\prime}(t) p_{\text{est}}^{(1)}(t+1)\right].
\end{aligned} \tag{B.13}$$

For the wildtype population size, $N_w^{(1)\prime}(t) = N_w^{(1)}(t) = K(1-r)^t$. Using this, $S_1(t) = \max(-z, s(1-\beta(1-r)^t))$. Consequently, the mutant fitness and along

with it the establishment probability p_{est} of a mutation are independent of the carrying capacity. A comparison of Eq. (B.13) with Eq. (B.3) completes the proof.

We now turn to $\vartheta = 0$: In that case, the wildtype population size decays simultaneously on all demes, and we have $N_w^{(i)\prime}(t) = N_w^{(i)}(t) = K(1-r)^t$ and consequently $S_i(t) = \max(-z, s(1-\beta(1-r)^t))$ for all $i \in \{1, \ldots, D\}$. This implies in particular that the establishment probability is the same in every deme and again the same as in a population of size KD. We immediately obtain from Eq. (2.5):

$$P_{\text{rescue}} \approx 1 - \exp\left[-\sum_{t=-\infty}^{-1} u\alpha DK p_{\text{est}}^{(1)}(t+1) + \sum_{t=0}^{\infty} u(1+S_1(t))DN_w^{(1)\prime}(t)p_{\text{est}}^{(1)}(t+1)\right], \tag{B.14}$$

which again coincides with Eq. (B.3).

The formulae imply that for $m = 0$ or $\vartheta = 0$, P_{rescue} depends only on the product DK (i.e., the carrying capacity of an equivalent unstructured population) and not on D and K separately.

C Appendix: Levene model

We now consider $D \geq 1$, but focus on the limiting case $m = 1$. In a first step, we furthermore restrict the analysis to $\beta = 0$. As derived in the main text Eq. (2.10), the effective growth parameter of a mutant in the meta-population in a period with d deteriorated demes is given by

$$s_{\text{eff}}(t) = \frac{d}{D}(1+s) + \frac{D-d}{D}\alpha\frac{K}{N_w^{(\text{total})}(t)/D} - 1. \tag{C.1}$$

The total number of wildtypes $N_w^{(\text{total})}(t) = N_w^{(\text{new})}(t) + K(D-d)$ is a function of time. However, after another deme has turned bad, it will quickly decay to its new steady state (until the next deme deteriorates). For sufficiently large values of ϑ, we can therefore approximate s_d as constant by taking the equilibrium value $\hat{N}_{w,d}^{(\text{total})}$ of the total wildtype population size in the period within which d demes

are perturbed. The stationary value of the total wildtype population size $\hat{N}_{w,d}^{(\mathrm{new})}$ in the bad environment is found by solving the equation

$$0 = (1-r)\frac{d}{D}\left(\hat{N}_{w,d}^{(\mathrm{new})} + (D-d)K\right) - \hat{N}_{w,d}^{(\mathrm{new})}. \tag{C.2}$$

This yields:

$$\hat{N}_{w,d}^{(\mathrm{new})} = \frac{(1-r)\frac{d}{D}(D-d)K}{1-(1-r)\frac{d}{D}}. \tag{C.3}$$

And we obtain

$$\hat{N}_{w,d}^{(\mathrm{total})} = \hat{N}_{w,d}^{(\mathrm{new})} + K(D-d). \tag{C.4}$$

Inserting Eq. (C.4) into Eq. (C.1) yields s_d as in Eq. (2.11). With Eq. (2.11), we can now use Eq. (A.7) with $\Phi = \vartheta$ to determine the establishment probability of a new mutant.

The probability of evolutionary rescue in a Levene model can be approximated by (cf. Eq. (2.5)):

$$\begin{aligned} P_{\mathrm{rescue}} \approx 1 - \exp\left[-u\sum_{t=-\infty}^{-1} KD\alpha p_{\mathrm{est}}(t+1) - u\sum_{t=0}^{(D-1)\vartheta-1} N_w^{(\mathrm{total})}(t)(1+s_{\mathrm{eff}}(t))p_{\mathrm{est}}(t+1)\right] \\ \times \exp\left[-u\frac{N_w^{(\mathrm{total})}((D-1)\vartheta)}{r}(1+s)\frac{2\hat{s}}{1+\hat{s}}\right], \end{aligned} \tag{C.5}$$

where $N_w^{(\mathrm{total})}((D-1)\vartheta)$ is the wildtype population size immediately after the last deme has deteriorated. The first sum captures the contributions of mutations that arose before time $t = 0$. The second sum takes all mutations into account that are generated as the degradation proceeds across the demes. From time $(D-1)\vartheta$ on, the population size decays geometrically, which leads to the last term in Eq. (C.5).

We now allow for density-dependent mutant fitness ($\beta \geq 0$). We assume that the periods of environmental stasis are very long. As discussed in the main text, it is in that case sufficient to consider the two phases where either one or no deme is unperturbed. We can approximate $S(t)$ during the phase where all but one deme have deteriorated by its steady state value:

$$\bar{S} = \max\left[-z, s\left(1 - \beta\frac{\hat{N}_{w,D-1}^{(\mathrm{total})}/D}{K}\right)\right] \tag{C.6}$$

with

$$\hat{N}_{w,D-1}^{(\text{total})} = \frac{KD}{(D-1)r+1}. \tag{C.7}$$

Analogously to Eq. (2.11), we then obtain Eq. (2.13) for the effective growth parameter during that period. In the main text, we pointed out that the situation corresponds to an unstructured population with a single environmental change. We can thus use Eq. (B.3) with the following substitutions to calculate the probability of evolutionary rescue:

$$\begin{array}{lll} K & \text{is substituted by} & \hat{N}_{w,D-1}^{(\text{total})}, \\ \alpha & \text{is substituted by} & 1+s_{D-1}, \\ \beta & \text{is substituted by} & \frac{\beta \hat{N}_{w,D-1}^{(\text{total})}}{DK}. \end{array} \tag{C.8}$$

We close the section with a comparison of $P_{\text{rescue}}(\vartheta \to 0)$ and $P_{\text{rescue}}(\vartheta \to \infty)$ if $\beta = 0$. In order to do so, we approximate the amount of genetic variation which is present at the time when the last deme deteriorates by its expected value. We only consider the case $s_{D-1} < 0$. For $\vartheta = 0$ and $\vartheta \to \infty$, we obtain

$$E[\text{sgv}_0] = uDK\frac{\alpha}{1-\alpha} \quad \text{and} \quad E[\text{sgv}_\infty] = u\hat{N}_{w,D-1}^{(\text{total})}\frac{1+s_{D-1}}{-s_{D-1}}, \tag{C.9}$$

respectively. Analogously, we introduce

$$E[\text{dnm}_0] = \frac{uKD(1+s)}{r} \quad \text{and} \quad E[\text{dnm}_\infty] = \frac{u\hat{N}_{w,D-1}^{(\text{total})}(1+s)}{r} \tag{C.10}$$

for the expected number of de-novo mutations that are generated after deterioration of the last deme in both scenarios. With this notation:

$$P_{\text{rescue}}(\vartheta = 0) \approx 1 - \exp\left[-(E[\text{sgv}_0] + E[\text{dnm}_0])\frac{2\hat{s}}{1+\hat{s}}\right], \tag{C.11a}$$

$$P_{\text{rescue}}(\vartheta \to \infty) \approx 1 - \exp\left[-(E[\text{sgv}_\infty] + E[\text{dnm}_\infty])\frac{2\hat{s}}{1+\hat{s}}\right]. \tag{C.11b}$$

We now compare the exponents:

$$\begin{aligned}
\frac{E[\mathrm{sgv}_\infty]+E[\mathrm{dnm}_\infty]}{E[\mathrm{sgv}_0]+E[\mathrm{dnm}_0]} &= \frac{\hat{N}^{(\mathrm{total})}_{w,D-1}}{DK}\frac{\frac{1+s_{D-1}}{-s_{D-1}}+\frac{1}{r}}{\frac{\alpha}{1-\alpha}+\frac{1}{r}} \\
&= \frac{\hat{N}^{(\mathrm{total})}_{w,D-1}}{DK}\frac{1-\alpha}{-s_{D-1}}\frac{r-s_{D-1}(1-r)}{1-\alpha+\alpha r} \qquad \left[\begin{array}{l}\text{Note: This expression gets minimal}\\ \text{when } -s_{D-1} \text{ gets maximal.}\\ -s_{D-1} \text{ as a function of } s \text{ gets maximal}\\ \text{for } s\to 0.\end{array}\right] \\
&= \frac{1}{1+r(D-1)}\frac{1}{1+(D-1)r-\frac{s+r}{1-\alpha}(D-1)} \\
&\quad \times \frac{(r+s+(1-\alpha)(1-r))(1+(D-1)r)-sD}{1-\alpha+\alpha r} \\
&= \frac{1}{1+(D-1)r-\frac{s+r}{1-\alpha}(D-1)}\frac{1-\alpha+\alpha r+s-\frac{sD}{1+r(D-1)}}{1-\alpha+\alpha r} \\
&\overset{s\to 0}{\geq} \frac{1}{1-\frac{\alpha}{1-\alpha}(D-1)r} \geq 1.
\end{aligned} \tag{C.12}$$

This means that for $\beta = 0$, $P_{\mathrm{rescue}}(\vartheta \to \infty) \geq P_{\mathrm{rescue}}(\vartheta = 0)$. For $\alpha = 0$, we obtain equality. The approximation thus suggests that for arbitrary values of r, the probability of evolutionary rescue is approximately the same for a very fast and and a very slowly progressing change if $\alpha = 0$ and s small, which is confirmed by computer simulations (not shown). This implies in particular, that for $\alpha = 0$, s small, and ϑ large, $P_{\mathrm{rescue}}(m = 0) \approx P_{\mathrm{rescue}}(m = 1)$, as can be seen in Figure 2.10A.

D Appendix: Island model without standing genetic variation

We here restrict ourselves to $\alpha = 0$ (i.e., mutants are lethal in the old environment) and $\beta = 0$ (no additional density dependence beyond the hard carrying

capacity). In the main text Eq. (2.14), we derived the following effective growth rate of a mutant:

$$s_{\text{eff}}(t) = \begin{cases} (1+s)(1-m) + (1+s)m\frac{d}{D} - 1 & \text{for} \quad t \in [(d-1)\vartheta, d\vartheta[, \\ & \qquad d \in \{1, \dots, D-1\}, \\ s & \text{for} \quad t \geq (D-1)\vartheta\ . \end{cases} \tag{D.13}$$

Using this, we can calculate the establishment probability with Eq. (A.7), setting again $\Phi = \vartheta$. For the total rescue probability, we obtain (cf. Eq. (2.5)):

$$\begin{aligned} P_{\text{rescue}} \approx 1 - \exp &\left[-u \sum_{t=0}^{(D-1)\vartheta-1} (1+s) \left(1 - m + \frac{d_t}{D} m\right) N_w^{(\text{new})}(t) p_{\text{est}}(t+1) \right] \\ &\times \exp \left[-u \sum_{t=0}^{(D-1)\vartheta-1} (1+s) m K (D - d_t) \frac{d_t}{D} p_{\text{est}}(t+1) \right] \\ &\times \exp \left[-u(1+s) \frac{N_w^{(\text{total})}((D-1)\vartheta)}{r} \frac{2\hat{s}}{1+\hat{s}} \right], \end{aligned} \tag{D.14}$$

where $N_w^{(\text{total})}((D-1)\vartheta)$ is the wildtype population size immediately after the last deme has deteriorated. The first term takes mutants into account that originate in the new part of the habitat. The second term considers mutant offspring of individuals from old demes that migrate to the new part where they can survive. The last term is the same as in the Levene model. As $\alpha = 0$, there are no mutants in the population before time $t = 0$.

In the main text, we gave an approximation for the probability of evolutionary rescue for $D = 2$ (see Eq. (2.17)), which generalizes to more than two islands in a straightforward way. The stationary value $N_{w,d}^{(\text{new})}$ of wildtype individuals in the perturbed part of the habitat in a period with d deteriorated demes is obtained as the solution of

$$0 = (1-r)\left(1 - \frac{D-d}{D} m\right) N_{w,d}^{(\text{new})} + m\frac{d}{D}(1-r)(D-d)K - N_{w,d}^{(\text{new})}. \tag{D.15}$$

This yields:

$$\hat{N}_{w,d}^{(\text{new})} = \frac{d(D-d)Km(1-r)}{(D-d)m(1-r) + Dr}. \tag{D.16}$$

And with

$$2s_{\text{eff}}(t) \approx 2\left(s - m + m\frac{d_t}{D}\right), \tag{D.17}$$

we obtain:

$$P_{\text{rescue}} \approx$$
$$1 - \exp\left[-u\vartheta\sum_{d=1}^{D-1} 2\max\left[\left(s - m + m\frac{d}{D}\right), 0\right]\frac{d(D-d)Km(1-r)}{(D-d)m(1-r)+Dr} - u\frac{DK}{r}2s\right]. \tag{D.18}$$

Figure D.1 shows a comparison between the exact formula (D.14) and the approximation. The approximation captures the behavior for small m very well. In particular, it reproduces the maximum in the probability of evolutionary rescue. As m increases, the approximation becomes worse.

If the number of demes D is large and $r \gg m$, we can approximate Eq. (D.18) by

$$P_{\text{rescue}} \approx$$
$$1 - \exp\left[-u\vartheta\int_0^D 2\max\left[\left(s - m + m\frac{d}{D}\right), 0\right]\frac{d(D-d)Km(1-r)}{Dr}\mathrm{d}d - u\frac{DK}{r}2s\right]$$
$$= 1 - \exp\left[-u\vartheta\frac{1}{3}D^2Km\frac{1-r}{r}\max\left[\left(s - \frac{1}{2}m\right), 0\right] - u\frac{DK}{r}2s\right]. \tag{D.19}$$

From Eq. (2.17) and Eq. (D.19), we find that the maximum is at $m \approx s$.

E Appendix: Further approximations

Unstructured population, small r and s The calculation of P_{rescue} based on Eq. (B.3) and (A.7) gets computationally expensive for small r. An approximation for small r can be obtained when we assume that the selection $s_{\text{eff}}(t)$ changes continuously in time and additionally replace sums by integrals in Eq. (B.3).

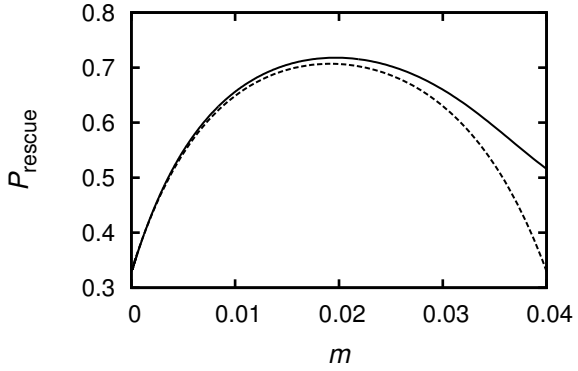


Figure D.1: Evolutionary rescue in a two island model. The plot compares the exact result (D.14) – solid line – with approximation (2.17) – dashed line. The parameter values are the same as in Figure 2.10.

We furthermore approximate $\ln(1+s(1-\beta N_w(t)/K)) \approx s(1-\beta N_w(t)/K)$ and $-\ln(1-r) \approx r$. For $\hat{z} \neq 0$, the establishment probability of a mutation then is

$$p_{\text{est}}(T) = \begin{cases} \frac{2}{1+\left(-\frac{1}{\hat{z}_0}+\mathrm{e}^{-\hat{z}_0 T}\left(\frac{1}{\hat{z}_0}-\frac{1}{\hat{z}}+\frac{1}{\hat{z}}\mathrm{e}^{\hat{z}T_\mathrm{c}}+\mathrm{e}^{\hat{z}T_\mathrm{c}}\int_0^\infty \mathrm{e}^{-st+\frac{s\beta}{r}\mathrm{e}^{rT_\mathrm{c}}(1-\mathrm{e}^{rt})}\mathrm{d}t\right)\right)}, & T \leq 0, \\ \frac{2}{1+\left(-\frac{1}{\hat{z}}+\mathrm{e}^{-\hat{z}(T-T_\mathrm{c})}\left(\frac{1}{\hat{z}}+\int_0^\infty \mathrm{e}^{-st-\frac{s\beta}{r}\mathrm{e}^{rT_\mathrm{c}}(1-\mathrm{e}^{rt})}\mathrm{d}t\right)\right)}, & 0 \leq T < T_\mathrm{c}, \\ \frac{2}{1+\int_0^\infty \mathrm{e}^{-st-\frac{s\beta}{r}\mathrm{e}^{rT}(1-\mathrm{e}^{rt})}\mathrm{d}t}, & T \geq T_\mathrm{c} \end{cases} \tag{E.1}$$

with

$$\hat{z}_0 = \max[-\ln(\alpha), 1], \tag{E.2}$$

$$\hat{z} = \max[-\ln(1-z), 1]. \tag{E.3}$$

To obtain the result for $\hat{z} = 0$, use $\lim_{\hat{z}\to 0} \frac{\exp(\hat{z}T)-1}{\hat{z}} = T$.

Infinitely many islands We here give an approximation for the probability of evolutionary rescue, when the number D of demes is large. We take the limits $D \to \infty$, $\vartheta \to 0$, $K \to 0$ with $D\vartheta = \Theta$ and $DK = K_{\text{total}}$ considered constant. We treat time as continuous and use the correspondence $\frac{d}{D} \hat{=} \frac{t}{\Theta}$.

We obtain the following differential equation for the number of wildtype individuals in the new part of the habitat (cf. the difference equation (A.1)):

$$\begin{aligned}\dot{N}_w^{(\text{new})}(t) = -rN_w^{(\text{new})} + (1-r)m\left(\frac{t}{\Theta} - 1\right)N_w^{(\text{new})} \\ + m\frac{t}{\Theta}(1-r)\left(1 - \frac{t}{\Theta}\right)K_{\text{total}} + \frac{K_{\text{total}}}{\Theta}.\end{aligned} \tag{E.4}$$

We now turn to the establishment probability of a new mutant. Both in the Levene and in the island model, s_d takes the form $s_d = \sigma_0 + \sigma_1 \frac{d}{D}$, $d \in \{1, \ldots, D-1\}$. We can approximate this as $s_{\text{eff}}(t) = \sigma_0 + \tilde{\sigma}_1 t$ with $\tilde{\sigma}_1 = \sigma_1/\Theta$ for $0 \leq t < \Theta$. We state the establishment probability for:

$$\hat{s}_{\text{eff}}(t) = \begin{cases} \hat{s}_0 & \text{for} \quad t < T_c^{(1)}, \\ \ln\left(1 + \tilde{\sigma}_0 + \tilde{\sigma}_1 t\right) & \text{for} \quad T_c^{(1)} \leq t < T_c^{(2)}, \\ 1 & \text{for} \quad T_c^{(2)} \leq t < \Theta \\ \hat{s} & \text{for} \quad t \geq \Theta \end{cases} \tag{E.5}$$

with

$$\begin{aligned} T_c^{(1)} &= \min\left[0, \frac{1}{\tilde{\sigma}_1}(\exp[-1] - 1 - \tilde{\sigma}_0)\right], \\ T_c^{(2)} &= \min\left[\Theta, \frac{1}{\tilde{\sigma}_1}(\exp[1] - 1 - \tilde{\sigma}_0)\right]. \end{aligned} \tag{E.6}$$

We obtain:

$$p_{\text{est}}(T) = \begin{cases} \frac{2}{1+\tilde{I}_0(T)} & \text{for} \quad T < T_c^{(1)}, \\ \frac{2}{1+\tilde{I}_1(T)} & \text{for} \quad T_c^{(1)} \leq T < T_c^{(2)}, \\ \frac{2}{1+\tilde{I}_2(T)} & \text{for} \quad T_c^{(2)} \leq T < \Theta, \\ \frac{2\hat{s}}{1+\hat{s}} & \text{for} \quad T \geq \Theta \end{cases} \tag{E.7}$$

with

$$\begin{aligned}
\tilde{I}_0(T) &= \frac{1}{\hat{s}_0} + \exp\left[\hat{s}_0(T-T_c)\right]\Bigg(- \frac{1}{\hat{s}_0} + \exp\left[f(T_c^{(1)})\right] \int\limits_{T_c^{(1)}}^{T_c^{(2)}} \exp\left[f(t)\right]\mathrm{d}t \\
&\qquad + \frac{1}{\hat{s}} \exp\left[-f(T_c^{(2)}) + f(T_c^{(1)}) - \Theta + T_c^{(2)}\right]\Bigg), \\
\tilde{I}_1(T) &= \exp\left[f(T)\right] \int\limits_{T}^{T_c^{(2)}} \exp\left[f(t)\right]\mathrm{d}t \\
&\qquad + \exp\left[-f(T_c^{(2)}) + f(T)\right](1 - \exp\left[\Theta - T_c^{(2)}\right]) \\
&\qquad + \exp\left[-f(T_c^{(2)}) + f(T) - \Theta + T_c^{(2)}\right]\frac{1}{\hat{s}}, \\
\tilde{I}_2(T) &= \left(\frac{1}{\hat{s}} - 1\right)\exp\left[-(\Theta - T)\right] + 1
\end{aligned} \tag{E.8}$$

and

$$f(t) = -t + \frac{(1 + \tilde{\sigma}_0 + \tilde{\sigma}_1 t)\ln\left(1 + \tilde{\sigma}_0 + \tilde{\sigma}_1 t\right)}{\tilde{\sigma}_1}. \tag{E.9}$$

In the Levene model, we find:

$$\sigma_0 = \alpha - 1, \tag{E.10}$$

$$\tilde{\sigma}_1 = (1 - \alpha + s + \alpha r)\frac{1}{\Theta}. \tag{E.11}$$

Analogously to Eq. (C.5), we obtain:

$$\begin{aligned}
P_{\text{rescue}} \approx 1 - \exp\Bigg[&- u \int\limits_{-\infty}^{0} \alpha K_{\text{total}} p_{\text{est}}(t+1)\mathrm{d}t \\
&- u \int\limits_{0}^{\Theta} (1 + s_{\text{eff}}(t)) \left(N_w^{(\text{new})} + K_{\text{total}}\left(1 - \frac{t}{\Theta}\right)\right) p_{\text{est}}(t+1)\mathrm{d}t\Bigg] \\
&\times \exp\left[-u\frac{N_w^{(\text{new})}(\Theta)}{r}(1+s)\frac{2\hat{s}}{1+\hat{s}}\right].
\end{aligned} \tag{E.12}$$

Finally, in the island model with $\alpha = \beta = 0$, we have:

$$\sigma_0 = (1+s)(1-m) - 1, \tag{E.13}$$

$$\tilde{\sigma}_1 = (1+s)m\frac{1}{\Theta}. \tag{E.14}$$

The total probability of evolutionary rescue becomes:

$$\begin{aligned} P_{\text{rescue}} \approx 1 - \exp\Bigg[- u \int_0^{\Theta} (1+s) \Bigg(& N_w^{(\text{new})}(t) \left(1 - m + m\frac{t}{\Theta}\right) \\ & + m\left(1 - \frac{t}{\Theta}K_{\text{total}}\frac{t}{\Theta}\right)\Bigg) p_{\text{est}}(t+1)\mathrm{d}t\Bigg] \\ & \times \exp\left[-u\frac{N_w^{\text{new}}(\Theta)}{r}(1+s)\frac{2\hat{s}}{1+\hat{s}}\right]. \end{aligned} \tag{E.15}$$

Chapter 3

Adaptive gene introgression after secondary contact

Abstract By hybridization and backcrossing, alleles can surmount species boundaries and get incorporated into the genome of a related species. This introgression of genes is of particular evolutionary relevance if it involves the transfer of adaptations between populations. However, any beneficial allele will typically be associated with other alien alleles, which are often deleterious for various reasons and hence hamper the introgression process. In order to describe the introgression of an adaptive allele, we set up a stochastic model with an explicit genetic makeup of linked and unlinked deleterious alleles. Based on the theory of reducible multitype branching processes, we derive a recursive expression for the establishment probability of the beneficial allele after a single hybridization event. We furthermore study the probability that slightly deleterious alleles hitchhike to fixation. The key to the analysis is a split of the process into a stochastic phase in which the process establishes itself and a deterministic phase in which the advantageous allele sweeps to fixation. We thereafter apply the theory to a set of biologically relevant scenarios such as introgression in the face of several unlinked or few closely linked deleterious alleles. A comparison to computer simulations shows that the approximations work well over a large parameter range.

A manuscript is in preparation for the Journal of Mathematical Biology.
The simulations to this chapter have been contributed by Derek Setter.

3.1 Introduction

Hybridzation between related species is a common phenomenon. Indeed, Mallet (2005) estimates that at least 25% of plant species and 10% of animal species still interbreed. Disappearance of natural habitat barriers following environmental change, the introduction of foreign species, escape of domesticated animals into the wild, or crop cultivation all create new regions of species range overlap and consequently cause high rates of hybridization. Despite reproductive barriers, hybridization between related species is often not completely prohibited and leads to the production of viable and fertile offspring. In the course of backcrossing with a parental species, not all alien genetic material might get lost; instead, part of it can become permanently incorporated into the genome of the sister species. The introgression of genes from one species into another has been shown to occur over a wide range of taxa (Rhymer and Simberloff, 1996; Lindner *et al.*, 1998; Arnold *et al.*, 1999; Arnold, 2004; Miller *et al.*, 2012). The introgression of genes from feral domestic to wild animals (Adams *et al.*, 2003; Beaumont *et al.*, 2001; Gottelli *et al.*, 1994; Rhymer and Simberloff, 1996) or from introduced to native species (Rhymer and Simberloff, 1996; Fitzpatrick *et al.*, 2010) involves potential ecological risks and can – if extensive – entail a loss of biodiversity.

In addition, evidence for the transfer of adaptations across species boundaries is growing (Arnold *et al.*, 1999; Arnold, 2004; Whitney *et al.*, 2006; Schwenk *et al.*, 2008; Arnold and Martin, 2009; The *Heliconius* Genome Consortium, 2012). Hybridization followed by intogression of genes can hence take direct influence on the evolutionary routes of a species and speed up adaptation. For example, the introduced sunflower species *Helianthus annuus* is likely to have acquired resistance genes from the native, locally adapted species *H. debilis*, which allowed it to expand its species range southwards (Heiser C. B., Jr., 1951; Whitney *et al.*, 2006). Similarly, Abi-Rached *et al.* (2011) suggest that positively selected immune system alleles from Neanderthals and Densiovians might have introgressed into modern humans. In agriculture, adaptive gene introgression can potentially constitute a major risk: adaptive herbivore, insecticide, or pathogen resistance genes from (possibly genetically modified) crops can spread to wild relatives, severely complicating weed control (Snow, 2002; Snow *et al.*, 2003). Importantly, Snow *et al.* (2003) show that a transgene can indeed reduce herbivory and increase fitness in a wild sunflower under natural conditions.

Early-generation hybrids, even if not entirely infertile or inviable, frequently suffer from a strongly reduced fitness. Often, hybrids display an intermediate phenotype which is maladapted to either parental niche. The low hybrid fitness can also result from genetic incompatibilities. By backcrossing with one of the parental species, alleles that prove to be deleterious on the foreign genetic background or cause maladaptation to the parental niches can be purged and fitness be restored (Heiser C. B., Jr., 1951; Arnold *et al.*, 1999). The probability of successful gene introgression critically depends on the strength of this fitness bottleneck.

Theoretical models on (adaptive) gene introgression that take a reduction in hybrid fitness into accound usually assume that a pre-defined number of backcrosses are required in order to lose the deleterious material and obtain a positively selected type (Demon *et al.*, 2007; Gosh and Haccou, 2010; Gosh *et al.*, 2012a,b). This basically assumes that the deleterious effects are homogeneously spread over the genome and that an appreciable amount of deleterious alleles is required to have a measurable impact on fitness. Focusing on other than genetical aspects of gene introgression such as the impact of a temporally varying environment (Gosh *et al.*, 2012a) or life history traits (Demon *et al.*, 2007), these model hence greatly simplify the underlying genetics. A step towards more realistic population genetic models has been made by Gosh *et al.* (2012a). Their analysis remains, however, restricted to the most basic scenario in which a single deleterious allele is linked to the locus under positive selection.

In this paper, we focus on a single hybridization event and examine the impact of linked and unlinked deleterious alleles on the introgression process of an adaptive allele. We first set up a Moran-like model which describes the evolution of the population by genetic drift, selection, and recombination. In the first part of the model analysis, we apply the theory of reducible multitype branching processes in order to determine by how much deleterious alleles reduce the introgression probability of a favorable allele, depending on the strength of selection and linkage. The second part considers the probability that closely linked deleterious alleles "hitchhike" to fixation. The analysis relies on a separation of the process into a strongly stochastic phase in which a haplotype carrying the beneficial allele establishes and a deterministic phase in which it sweeps through the population, possibly losing deleterious material by recombination with wildtype individuals. These recombination events and the subsequent establishment or loss of haplotypes with fewer deleterious alleles are again subject to strong stochasticity. In this analysis, we again resort to the theory of branching processes. The derived

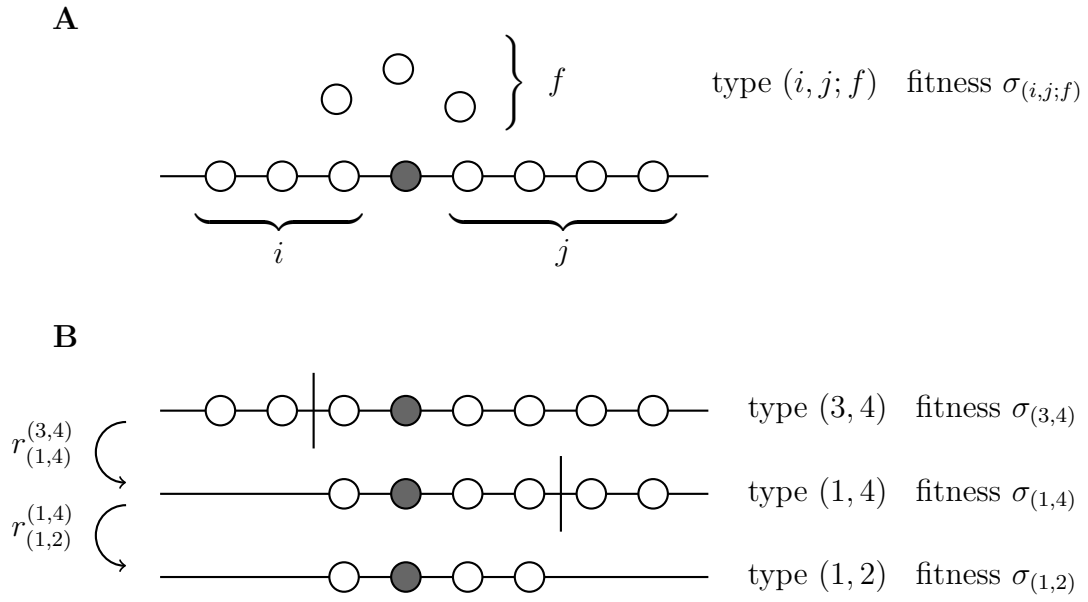


Figure 3.1: Illustration of the branching model. The dark dot represents the advantageous allele, the blank dots represent deleterious alleles. i and j give the number of deleterious allele to the left and right of the advantageous allele and f the number of unlinked alleles. Panel B illustrates how linked deleterious alleles are lost by recombination with wildtype individuals.

approximations are applied to a variety of biological scenarios, complemented by computer simulations. We close the paper by a brief discussion.

3.2 Full model and simulations

We consider a large population of N haploid individuals. The theory also applies to diploids without dominance if we can assume Hardy-Weinberg equilibrium. We use, however, the haploid formalism throughout the paper. By a single hybridization event, a hybrid individual is introduced into the population (for diploids, the alien alleles arrive in the foreign habitat at the haploid stage, i.e., for plants, by pollen dispersal). The hybrid carries an adaptive allele as well as a number of deleterious alleles. Deleterious alleles are either physically linked to the adaptive allele or unlinked. We assume that this initial introgressed haplotype carries

I and J linked deleterious to the left and the right of the beneficial allele, respectively, and F unlinked alleles. By recombination with wildtype individuals, haplotypes with fewer introgressed alleles can be generated, leading to a hybrid swarm. Selection on the deleterious alleles relies on maladaptation to the environment and is independent of the genomic context. We assume that there are no functional differences among wildtype individuals, i.e., all wildtype individuals have the same fitness and introgressed alleles interact identically with all wildtype backgrounds. The fitness of an individual is thus fully determined by the introgressed alleles that it carries.

The evolution of the population is described by the following scheme: At rate N, two individuals are chosen to reproduce and generate a single offspring. During reproduction, recombination can take place. We restrict ourselves to single crossover among the linked alleles. Double crossover is unlikely to happen over recombination distances r with $r^2 \ll r$ so that the model approximates scenarios of tight linkage. Considering larger recombination distances or gene conversion requires a straightforward extension of the formalism. Unlinked alleles are inherited with probability one half. The offspring replaces an individual that is chosen based on its fitness. For notational simplicity, we assign the numbers 1 to N to the individuals. Individual number k is then chosen with probability $\frac{1-\sigma^{(k)}}{\sum_{i=1}^{N}(1-\sigma^{(i)})}$, where $\sigma^{(k)} = 0$ for wildtype individuals. I.e., $\sigma^{(k)}$ is the Malthusian fitness of individual k in a wildtype population.

The simulation program implements the successive events without consideration of the time spans between them. As we are only interested in probabilities, this does not influence the results. The simulation program is written in the C++ programming language, making use of the *Gnu Scientific Library* (Galassi *et al.*, 2009).

The full model does not allow for an analytical treatment. In the following sections, we therefore consider approximations of the introgression process.

3.3 The early phase of spread

A single evolutionary step involves three individuals: two individuals that reproduce and one that dies (where the same individual might be chosen twice). In a large population, as long as hybrid haplotypes are rare, it is very unlikely that more than one hybrid individual is engaged in a single event (or the same individual twice). Formally, this corresponds to negligence of terms of order (n_{intro}/N)

in the transition rates, where n_{intro} denotes the number of individuals with introgressed material. In the early phase of spread, hybrids consequently suffer (nearly) independent fates, and the process is therefore well described by a multitype branching process. The branching process is strictly recovered in the limit $N \to \infty$.

The non-interaction of hybrids entails in particular that types with introgressed material only recombine with wildtype individuals. This implies that by recombination, they can only lose, not gain deleterious alleles, and we encounter a special instance of a reducible multitype branching process (cf. also Barton and Bengtsson, 1986; Demon *et al.*, 2007; Gosh *et al.*, 2012b). By recombination, types that carry only deleterious alleles, but not the beneficial allele are generated. We do not consider these types in the following (within a branching process approach, they are doomed to extinction), but focus on carriers of the advantageous allele. For the main part of the paper, we assume that all unlinked deleterious alleles have the same effect. A generalization of the main results to arbitrary effects is given in Appendix C.

We call an individual with i deleterious alleles to the left and j deleterious alleles to the right of the focal beneficical allele and f unlinked deleterious alleles an individual of type $(i, j; f)$ (cf. Figure 3.1A). Its net selection coefficient is denoted by $\sigma_{(i,j;f)}$. We set $(i, j) \equiv (i, j; 0)$. Recombinant offspring of a type (i, j) individual are either of type (i, k) with $k < j$ or of type (k, j) with $k < i$. Such recombination events may happen with probabilities $r^{(i,j)}_{(i,k)}$ and $r^{(i,j)}_{(k,j)}$, respectively. An instance of repeated recombination events is depicted in Figure 3.1B. The overall probability that a recombination event takes place is given by the sum $r_{(i,j)} = \sum_{k=0}^{j-1} r^{(i,j)}_{(i,k)} + \sum_{k=0}^{i-1} r^{(i,j)}_{(k,j)}$. The number of unlinked deleterious alleles that are inherited by an offspring individual is binomially distributed with parameter 0.5. We obtain for the per capita transition rates of the possible events in the branching process:

$$\begin{aligned}
P((i,j;f) \to 0) &= 1 - \sigma_{(i,j;f)}, \\
P((i,j;f) \to \{(i,j;f);(i,j;g)\}) &= \binom{f}{g}\left(\frac{1}{2}\right)^f (1 - r_{(i,j)}), \\
P((i,j;f) \to \{(i,j;f);(i,k;g)\}) &= \binom{f}{g}\left(\frac{1}{2}\right)^f r^{(i,j)}_{(i,k)} \quad \text{for} \quad k < j, \\
P((i,j;f) \to \{(i,j;f);(k,j;g)\}) &= \binom{f}{g}\left(\frac{1}{2}\right)^f r^{(i,j)}_{(k,j)} \quad \text{for} \quad k < i.
\end{aligned} \tag{3.1}$$

3.4 The probability of adaptive gene introgression

First, we focus on the probability that the beneficial allele establishes in the population. Once the beneficial allele is sufficiently frequent, it is very unlikely to get lost again. The extinction probability of the branching process as described in the previous section is thus a good approximation for the extinction probability of the beneficial allele in the full model. We denote by $Q_{(i,j;f)}$ the extinction probability of the process which is initiated by exactly one individual of type $(i,j;f)$.

Theorem 3.4.1. *The extinction probability $Q_{(I,J;F)}$ can be calculated by recursively solving the system of quadratic equations*

$$\begin{aligned}
&(1-r_{(i,j)})\left(\frac{1}{2}\right)^f Q_{(i,j;f)}^2 \\
&+\sum_{g=0}^{f}\binom{f}{g}\left(\frac{1}{2}\right)^f\left\{\sum_{k=0}^{i-1} r_{(k,j)}^{(i,j)} Q_{(k,j;g)}+\sum_{k=0}^{j-1} r_{(i,k)}^{(i,j)} Q_{(i,k;g)}\right\} Q_{(i,j;f)} \\
&+\sum_{g=0}^{f-1}\binom{f}{g}\left(\frac{1}{2}\right)^f (1-r_{(i,j)}) Q_{(i,j;g)} Q_{(i,j;f)} - (2-\sigma_{(i,j;f)}) Q_{(i,j;f)} + 1 - \sigma_{(i,j;f)} = 0, \\
&\qquad i\in\{0,\dots,I\},\quad j\in\{0,\dots,J\},\quad f\in\{0,\dots,F\},
\end{aligned} \tag{3.2}$$

where always the smaller root of the equation has to be used.

We only give an illustrative derivation of Eq. (3.2) here and move the full proof to Appendix A.

Consider a branching process initiated by an individual of type $(i,j;f)$. With probability $\frac{1-\sigma_{(i,j;f)}}{2-\sigma_{(i,j;f)}}$, the founding individual dies before it reproduces, in which case the lineage is immediately extinct. With probability $\binom{f}{g}\left(\frac{1}{2}\right)^f \frac{1-r_{(i,j)}}{2-\sigma_{(i,j;f)}}$, it reproduces and generates a non-recombinant offspring with g unlinked alleles, i.e., an offspring of type $(i,j;g)$. With probability $\binom{f}{g}\left(\frac{1}{2}\right)^f \frac{r_{(i,k)}^{(i,j)}}{2-\sigma_{(i,j)}}$ (or $\binom{f}{g}\left(\frac{1}{2}\right)^f \frac{r_{(k,j)}^{(i,j)}}{2-\sigma_{(i,j)}}$), it reproduces and gives birth to a type $(i,k;g)$ (or $(k,j;g)$) individual with $0\le k\le j$ (or $0\le k\le i$). Both individuals that exist after reproduction are the founding individual of a lineage. In order for the original lineage to go

extinct, both these lineages have to die out. It therefore holds for the extinction probability $Q_{(i,j;f)}$:

$$Q_{(i,j;f)} = \frac{1-\sigma_{(i,j;f)}}{2-\sigma_{(i,j;f)}} + \sum_{g=0}^{f} \binom{f}{g} \left(\frac{1}{2}\right)^f \frac{1-r_{(i,j)}}{2-\sigma_{(i,j;f)}} Q_{(i,j;g)} Q_{(i,j;f)} + \sum_{g=0}^{f} \binom{f}{g} \left(\frac{1}{2}\right)^f \left\{ \sum_{k=0}^{i-1} \frac{r^{(i,j)}_{(k,j)}}{2-\sigma_{(i,j;f)}} Q_{(k,j;g)} + \sum_{k=0}^{j-1} \frac{r^{(i,j)}_{(i,k)}}{2-\sigma_{(i,j;f)}} Q_{(i,k;g)} \right\} Q_{(i,j;f)}. \tag{3.3}$$

By rearrangement of terms, we obtain Eq. (3.2).

For the special case $F = 0$, Eq. (3.2) simplifies to

$$\begin{aligned} 0 = (1-r_{(i,j)})Q^2_{(i,j)} + \sum_{k=0}^{j-1} r^{(i,j)}_{(i,k)} Q_{(i,j)} Q_{(i,k)} + \sum_{k=0}^{i-1} r^{(i,j)}_{(k,j)} Q_{(i,j)} Q_{(k,j)} \\ + 1 - \sigma_{(i,j)} - (2-\sigma_{(i,j)}) Q_{(i,j)}, \quad i \in \{0,\dots,I\}, \quad j \in \{0,\dots,J\}. \end{aligned} \tag{3.4}$$

In contrast, if all deleterious alleles are unlinked (i.e., $I = J = 0$; abbreviate $(0,0;g) \equiv g$), Eq. (3.2) yields:

$$0 = \left(\frac{1}{2}\right)^f Q_f^2 + \left(\frac{1}{2}\right)^f \sum_{g=0}^{f-1} \binom{f}{g} Q_g Q_f + 1 - \sigma_f - (2-\sigma_f) Q_f, \quad f \in \{0,\dots,F\}. \tag{3.5}$$

Implications of this result are discussed below in section 3.6.

3.5 The hitchhiking probability

3.5.1 General idea

If the effects of closely linked deleterious alleles are not too harmful, namely if $\sigma_{(i,j)} - r_{(i,j)} > 0$, the beneficial allele can drag (some of) these deleterious alleles along to fixation. In this section, we develop a framework for determining the hitchhiking probabilities conditioned on fixation of the beneficial allele. The approach is based on a split of the process into two phases. After the original hybridization event, the beneficial allele has to establish itself. We call this the "stochastic phase". In the previous section, we have been concerned with the

establishment probability. Here, we further derive which haplotype (i, j) will escape stochastic loss in this initial phase. We assume that only one haplotype escapes. This is a very likely outcome of the stochastic phase under many circumstances because the establishment probability of each type is low. Since this establishment happens while the introgressed types are rare, we can base the derivation on the multitype branching process as before. The further increase in frequency of type (i, j) can be well described by deterministic growth. If no further recombination events happened, it would rise to fixation following the logistic equation

$$\dot{x}_{(i,j)}(t) = \sigma_{(i,j)} x_{(i,j)}(t)(1 - x_{(i,j)}(t)). \tag{3.6}$$

However, during the sweep, types with less deleterious material can still be generated by recombination. If one of these types establishes, it will outcompete type (i, j). Building on theory by Hartfield and Otto (2011), we describe the production and possible establishment of these types by a time-inhomogeneous branching process with immigration. Although the generation and establishment of new haplotypes is subject to strong stochasticity, we refer to this phase as to the "deterministic phase" because we model the frequency paths of haplotypes deterministically once they are established in the population.

We first give a derivation for the case without unliked deleterious material ($F = 0$), and subsequently generalize the approximation to $F > 0$.

3.5.2 The stochastic phase

As a first step, we determine which haplotype "rescues" the introgression process given that the process does not go extinct. For this initial phase, we again resort to the multitype branching process as defined in Eq. (3.1). As before, the process is initiated by a single individual of type (I, J). If $\sigma_{(I,J)} - r_{(I,J)} > 0$, type (I, J) has the chance to establish a permanent lineage of its own type. If $\sigma_{(I,J)} - r_{(I,J)} \leq 0$, type (I, J) itself will go extinct with probability 1. However, until extinction, recombinant offspring with fewer deleterious alleles can be generated and rescue the process. In that case, in order to determine the "rescue type", we can consider all recombination pathways that lead to establishment of the beneficial allele and determine with which (relative) probability the various paths get realized. This idea is key for the derivation of the approximation in this section.

Throughout the analysis, the total number of recombination events from type (I, J) to any other type until extinction of type (I, J) constitutes a central quantity. This follows Serra (2006) and Serra and Haccou (2007). For

$\sigma_{(I,J)} - r_{(I,J)} \leq 0$, we denote the corresponding probability generating function (p.g.f.) by $h(s)$. For $\sigma_{(I,J)} - r_{(I,J)} > 0$, we consider the number of recombination events conditioned on extinction of type (I, J) and denote the p.g.f. by $\hat{h}(s)$. $h(s)$ and $\hat{h}(s)$ can be explicitely calculated for our model and are given by Lemma D.5.

For the following lemma, we group the recombinant offspring of type (I, J) individuals into two classes: (1) individuals that found processes that survive (2) individuals that found processes that go extinct. We denote by Y_+ and Y_- the random number of recombination events from type (I, J) to type 1 and type 2 individuals, respectively. In the lemma, we rewrite the survival probability of the process in terms of the expected number of sucessful recombinant lineages and an error term. This lemma is essentially equivalent to the result in Serra and Haccou (2007, Eq. (8)). A similar result also appears already in Iwasa *et al.* (2004b).

Lemma 3.5.1. *Let $\sigma_{(I,J)} - r_{(I,J)} < 0$. The survival probability of the process can be written as*

$$1 - Q_{(I,J)} = \left(\sum_{k=0}^{I-1} r_{(k,J)}^{(I,J)} (1 - Q_{(k,J)}) + \sum_{k=0}^{J-1} r_{(I,k)}^{(I,J)} (1 - Q_{(I,k)}) \right) \frac{\frac{\mathrm{d}}{\mathrm{d}s} h(s)|_{s=1}}{r_{(I,J)}} - R_1, \tag{3.7}$$

where the error term R_1 is given by

$$R_1 = \frac{\partial}{\partial s_0} \left(\frac{h_1(s_0, s_1) - h_1(0, s_1)}{s_0} \right) \Bigg|_{s_0 = s_1 = 1} \tag{3.8}$$

with

$$h_1(s_0, s_1) = h(P_{\text{success}} s_0 + (1 - P_{\text{success}}) s_1) \tag{3.9}$$

and

$$P_{\text{success}} = \frac{\sum_{k=0}^{I-1} r_{(k,J)}^{(I,J)} (1 - Q_{(k,J)}) + \sum_{k=0}^{J-1} r_{(I,k)}^{(I,J)} (1 - Q_{(I,k)})}{r_{(I,J)}}. \tag{3.10}$$

Proof. A recombinant offspring of a type (I, J) individual founds an infinite lineage with probability P_{success} and a lineage that goes extinct with probability $1 - P_{\text{success}}$. According to Lemma D.6, the joint p.g.f of Y_+ and Y_- is given by

$$h_1(s_1, s_2) = h(P_{\text{success}} s_1 + (1 - P_{\text{success}}) s_2), \tag{3.11}$$

and we obtain for the expected number of type 1 individuals:

$$E[Y_+] = \frac{\partial}{\partial s_1} h_1(s_1, s_2)|_{s_1=s_2=1} = P_{\text{success}} \frac{\mathrm{d}}{\mathrm{d}s} h(s)|_{s=1}. \tag{3.12}$$

Now note:

$$1 - Q_{(I,J)} = P(Y_+ > 0) = E[Y_+] - R_1 \tag{3.13}$$

with

$$\begin{aligned} R_1 &= E[Y_+] - P(Y_+ > 0) = P(Y_+ = 2) + 2P(Y_+ = 3) + 3P(Y_+ = 4) + \dots \\ &= \frac{\partial}{\partial s_0} \left(\frac{h_1(s_0, s_1) - h_1(0, s_1)}{s_0} \right) \Bigg|_{s_0=s_1=1}. \end{aligned} \tag{3.14}$$

I.e., if $P(Y_+ > 1) \approx 0$, the expected number of recombination events from type (I, J) individuals to individuals with fewer mutations that found a successful lineage approximates the survival probability of the process.

□

In order to proceed, we need a formal definition of a "rescue type". Analogous to the lemma, we can then derive a recursive formula for the probability that an individual of type (i, j) rescues the process.

Definition 3.5.1. *We call an (i, j) individual an (i, j) "rescue type", denoted as $(i, j, +)$, if*
(1) it founds an infinite lineage of type (i, j) individuals
(2) there is no individual in its ancestry that founds an infinite lineage of its own type.

We denote by $X^{(k,l)}_{(i,j,+)}$ the number of rescue types $(i, j, +)$ in a process which is founded by an individual of type (k, l).

We define

$$P^{(I,J)}_{(i,j)} = \text{Prob}(X^{(I,J)}_{(i,j,+)} > 0 | \text{survival of the process}). \tag{3.15}$$

That is, $P^{(I,J)}_{(i,j)}$ gives the probability that there exists an (i, j) rescue type. A priori, that does not exclude the simultaneous existence of several rescue types. For the following theorem, we again group the recombinant offspring of a type (I, J) individual into two classes: (1) individuals that found a lineage resulting

in at least one individual of type $(i,j,+)$ (2) individuals that do not do that. We denote the number of recombinants of the first and second type with $Y_{(i,j,+)}$ and $Y_{(i,j,-)}$, respectively.

Theorem 3.5.1. *(1) Let $\sigma_{(I,J)} - r_{(I,J)} < 0$. It holds that*

$$P_{(i,j)}^{(I,J)} = \frac{\left(\sum_{k=0}^{I-1} r_{(k,J)}^{(I,J)}(1-Q_{(k,J)})P_{(i,j)}^{(k,J)} + \sum_{k=0}^{J-1} r_{(I,k)}^{(I,J)}(1-Q_{(I,k)})P_{(i,j)}^{(I,k)}\right)\frac{\frac{\mathrm{d}}{\mathrm{d}s}h(s)|_{s=1}}{r_{(I,J)}} - R_1}{\left(\sum_{k=0}^{I-1} r_{(k,J)}^{(I,J)}(1-Q_{(k,J)}) + \sum_{k=0}^{J-1} r_{(I,k)}^{(I,J)}(1-Q_{(I,k)})\right)\frac{\frac{\mathrm{d}}{\mathrm{d}s}h(s)|_{s=1}}{r_{(I,J)}} - R_2}, \tag{3.16}$$

where R_1 is defined as before and R_2 given by

$$R_2 = \frac{\partial}{\partial s_0}\left(\frac{h_2(s_0,s_1) - h_2(0,s_1)}{s_0}\right)\Big|_{s_0=s_1=1} \tag{3.17}$$

with

$$h_2(s_0,s_1) = h(P_{(i,j,+)}s_0 + (1-P_{(i,j,+)})s_1) \tag{3.18}$$

and

$$P_{(i,j,+)} = \frac{\sum_{k=0}^{I-1} r_{(k,J)}^{(I,J)}(1-Q_{(k,J)})P_{(i,j)}^{(k,J)} + \sum_{k=0}^{J-1} r_{(I,k)}^{(I,J)}(1-Q_{(I,k)})P_{(i,j)}^{(I,k)}}{r_{(I,J)}}. \tag{3.19}$$

(2) For $\sigma_{(I,J)} - r_{(I,J)} > 0$, it holds:

$$P_{(I,J)}^{(I,J)} = \frac{1-q_{(I,J)}}{1-Q_{(I,J)}} \tag{3.20}$$

with

$$1 - q_{(I,J)} = \frac{\sigma_{(I,J)} - r_{(I,J)}}{1 - r_{(I,J)}}, \tag{3.21}$$

where $q_{(I,J)}$ is the unconditioned probability that type (I,J) itself goes extinct.

For $(i,j) \neq (I,J)$, it holds:

$$P^{(I,J)}_{(i,j)} = \left(1 - \frac{1-q_{(I,J)}}{1-Q_{(I,J)}}\right) \times \frac{\left(\sum_{k=0}^{I-1} r^{(I,J)}_{(k,J)}(1-Q_{(k,J)})P^{(k,J)}_{(i,j)} + \sum_{k=0}^{J-1} r^{(I,J)}_{(I,k)}(1-Q_{(I,k)})P^{(I,k)}_{(i,j)}\right)\frac{\frac{\mathrm{d}}{\mathrm{d}s}\hat{h}(s)|_{s=1}}{r_{(I,J)}} - \hat{R}_1}{\left(\sum_{k=0}^{I-1} r^{(I,J)}_{(k,J)}(1-Q_{(k,J)}) + \sum_{k=0}^{J-1} r^{(I,J)}_{(I,k)}(1-Q_{(I,k)})\right)\frac{\frac{\mathrm{d}}{\mathrm{d}s}h(s)|_{s=1}}{r_{(I,J)}} - \hat{R}_2}, \tag{3.22}$$

where $\hat{h}_1$, $\hat{h}_2$, $\hat{R}_1$, and $\hat{R}_2$ are defined analogously to before (using $\hat{h}$ instead of h).

Proof. We first prove the first part of the theorem. With probability $P_{(i,j,+)}$, a recombinant offspring founds a lineage resulting in at least one individual of type $(i,j,+)$. Analogous to before, we obtain

$$\begin{aligned} P(Y_{(i,j,+)} > 0) &= E[Y_{(i,j,+)}] - R_2 \\ &= P_{(i,j,+)}\frac{\mathrm{d}}{\mathrm{d}s}h(s)|_{s=1} - R_2 \end{aligned} \tag{3.23}$$

with

$$R_2 = \frac{\partial}{\partial s_0}\left(\frac{h_2(s_0,s_1) - h_2(0,s_1)}{s_0}\right)\Bigg|_{s_0=s_1=1}. \tag{3.24}$$

It holds:

$$P^{(I,J)}_{(i,j)}(1-Q_{(I,J)}) = P(Y_{(i,j,+)} > 0). \tag{3.25}$$

Substituting $1-Q_{(I,J)}$ by the approximation Eq. (3.7) yields Eq. (3.16).

If $\sigma_{(I,J)} - r_{(I,J)} > 0$, type (I,J) establishes a lineage of its own type with probability

$$1 - q_{(I,J)} = \frac{\sigma_{(I,J)} - r_{(I,J)}}{1 - r_{(I,J)}} \tag{3.26}$$

(cf. Lemma D.3). It therefore holds:

$$P^{(I,J)}_{(I,J)} = \frac{1-q_{(I,J)}}{1-Q_{(I,J)}}. \tag{3.27}$$

The probability that type (I, J) goes extinct conditioned on survival of the process is accordingly given by

$$P(\text{type } (I,J) \text{ goes extinct}|\text{survival of the process}) = 1 - \frac{1-q_{(I,J)}}{1-Q_{(I,J)}}. \tag{3.28}$$

We can now repeat the proof of the first part of the theorem for the process conditioned on extinction of type (I, J). □

Remark. For $\sigma_{(I,J)} - r_{(I,J)} < 0$, we have $\frac{\mathrm{d}}{\mathrm{d}s}h(s)|_{s=1} = \frac{r_{(I,J)}}{r_{(I,J)}-\sigma_{(I,J)}}$. For $\sigma_{(I,J)} - r_{(I,J)} > 0$, we have $\frac{\mathrm{d}}{\mathrm{d}s}\hat{h}(s)|_{s=1} = \frac{r_{(I,J)}}{\sigma_{(I,J)}-r_{(I,J)}}$.

For $\sigma_{(I,J)} - r_{(I,J)} < 0$ not too close to zero, it is likely that only one of the few recombinant offspring of type (I, J) individuals founds an infinite lineage, and we can approximate $P(Y^+ \geq 2) \approx 0$ and consequently also $P(Y_{(i,j,+)} \geq 2) \approx 0$. This implies that the error terms R_1 and R_2 can be ignored. For $\sigma_{(I,J)} - r_{(I,J)} > 0$ and $r_{(I,J)}$ small, survival of the process is with high probability contingent on establishment of type (I, J) so that

$$P_{(i,j)}^{(I,J)} \approx \delta_{I,i}\delta_{J,j} \quad \text{with} \quad \delta_{k_1,k_1} = \begin{cases} 1 & \text{for} \quad k_1 = k_2, \\ 0 & \text{else.} \end{cases} \tag{3.29}$$

We can therefore formulate the following corollary:

Corollary 3.5.1. *For $\sigma_{(I,J)} - r_{(I,J)} < 0$ not too close to zero and close linkage, we can approximate*

$$P_{(i,j)}^{(I,J)} \approx \frac{P_{(i,j,+)}}{P_{\text{success}}} = \frac{\sum_{k=0}^{I-1} r_{(k,J)}^{(I,J)}(1-Q_{(k,J)})P_{(i,j)}^{(k,J)} + \sum_{k=0}^{J-1} r_{(I,k)}^{(I,J)}(1-Q_{(I,k)})P_{(i,j)}^{(I,k)}}{\sum_{k=0}^{I-1} r_{(k,J)}^{(I,J)}(1-Q_{(k,J)}) + \sum_{k=0}^{J-1} r_{(I,k)}^{(I,J)}(1-Q_{(I,k)})} \tag{3.30}$$

with

$$P_{(k,l)}^{(i,j)} \approx \delta_{i,k}\delta_{j,l} \tag{3.31}$$

for $\sigma_{(i,j)} - r_{(i,j)} > 0$. Within this approximation, the $P_{(i,j)}^{(I,J)}$, $i \leq I$, $j \leq J$, form a probability distribution with

$$\sum_{i,j} P_{(i,j)}^{(I,J)} = 1. \tag{3.32}$$

The proof for relation Eq. (3.32) is given in Appendix B.

The approximation (Eqs. (3.30) and (3.31)) implies that exactly one rescue type establishes in the population during the stochastic phase (types with fewer deleterious alleles can still arise later during the deterministic phase), i.e.,

$$P(X^{(I,J)}_{(i,j,+)} + X^{(I,J)}_{(k,l,+)} > 1) = 0 \quad \text{for any pair} \quad (i,j),(k,l). \tag{3.33}$$

This assumption appears to be justified for a large parameter region with tightly linked deleterious alleles. It is also the basis for most of our analytical anaysis of particular cases below. As discussed below, the approximation becomes less accurate if deleterious alleles are relatively loosely linked and/or haplotypes are only slightly deleterious.

We can extend the approximation to include unlinked deleterious alleles and obtain for $\sigma_{(I,J)} - r_{(I,J)} < 0$:

$$\begin{aligned}
P^{(I,J;F)}_{(i,j;0)} &\approx \frac{A+B}{C+D} \\
A &= \sum_{l=0}^{F} \binom{F}{l} \left(\frac{1}{2}\right)^F \left\{ \sum_{i=0}^{I-1} r^{(I,J)}_{(i,J)} (1 - Q_{(i,J;l)}) P^{(i,J;l)}_{(i,j;0)} \right. \\
&\qquad \left. + \sum_{j=0}^{J-1} r^{(I,J)}_{(I,j)} (1 - Q_{(I,j;l)}) P^{(I,j;l)}_{(i,j;0)} \right\} \\
B &= \sum_{l=0}^{F-1} \binom{F}{l} \left(\frac{1}{2}\right)^F (1 - r_{(I,J)})(1 - Q_{(I,J;l)}) P^{(I,J;l)}_{(i,j;0)} \\
C &= \sum_{l=0}^{F} \binom{F}{l} \left(\frac{1}{2}\right)^F \left\{ \sum_{i=0}^{I-1} r^{(I,J)}_{(i,J)} (1 - Q_{(i,J;l)}) + \sum_{j=0}^{J-1} r^{(I,J)}_{(I,j)} (1 - Q_{(I,j;l)}) \right\} \\
D &= \sum_{l=0}^{F-1} \binom{F}{l} \left(\frac{1}{2}\right)^F (1 - r_{(I,J)})(1 - Q_{(I,J;l)}).
\end{aligned} \tag{3.34}$$

For $\sigma_{(I,J)} - r_{(I,J)} > 0$, we approximate:

$$P^{(I,J;F)}_{(i,j;0)} \approx \delta_{I,i}\delta_{J,j}. \tag{3.35}$$

3.5.3 The deterministic phase

It remains to determine whether the haplotype that establishes itself in the stochastic phase rises to fixation or whether types with less deleterious material can establish during the sweep of the beneficial allele. In order to arrive at an approximation for the deterministic phase, we apply and extend an approach developed in Hartfield and Otto (2011). Hartfield and Otto (2011) determined the hitchhiking probability of a single deleterious allele which is closely linked to a beneficial one. For a single hitchhiker, their method can easily be adapted to our model, as shown below. In the Appendix, we further argue that the approach can be extended to a larger number of hitchhikers. Explicit results for two hitchhikers are derived in Appendix E.

For a single potential hitchhiker, assume that type $(0,1)$ with $\sigma_{(0,1)} - r_{(0,1)} > 0$ has been introduced and established in the population. It's further growth can be well described deterministically as given by the differential equation Eq. (3.6). However, in the initial phase, it will on average have grown faster than the deterministic path predicts. Following Uecker and Hermisson (2011), we account for the fast initial increase by the use of an "effective initial population size" ν, which we use as an initial condition for the solution of Eq. (3.6) (cf. also Desai and Fisher, 2007). ν is an exponentially distributed random variable with

$$P(\nu \leq \nu_0) = 1 - \exp\left(-p_{\text{est}}\nu_0\right), \tag{3.36}$$

where $p_{\text{est}} = 1 - q_{(0,1)}$ (cf. Eq. (3.26)) denotes the establishment probability of a single type $(0,1)$ individual in a wildtype population (Uecker and Hermisson, 2011, Eq. (40)). To leading order approximation, we can approximate the distribution by its mean $\bar{\nu} = 1/p_{\text{est}}$. For the relative frequency of type $(0,1)$, it then holds (ignoring recombination):

$$x_{(0,1)}(t) = \frac{\bar{\nu}\exp\left(\sigma_{(0,1)}t\right)}{N - \bar{\nu} + \bar{\nu}\exp\left(\sigma_{(0,1)}t\right)}. \tag{3.37}$$

This is a good approximation up to freqency $x_{(0,2)} \approx 1 - \frac{\bar{\nu}}{N}$. At higher frequencies, the frequency will again grow faster. Individuals of type $(0,0)$ are recurrently generated by recombination at rate $r_{(0,1)}Nx_{(0,1)}(1 - x_{(0,1)})$. As long as they are rare, their dynamics is strongly determined by stochasticity and can once again be approximated by a branching process. Their fitness depends on $x_{(0,1)}(t)$ and hence on time. The dynamics is thus described by a time-inhomogeneous branching

process with birth rate 1 and death rate $1 - \sigma_{(0,0)} + x_{(0,1)}(t)\sigma_{(0,1)}$. Following Eq. (16a) in Uecker and Hermisson (2011), the fixation probability of a single individual of type $(0,0)$ generated at time T is given by

$$p_{\text{fix}}^{(0,0)}(T) = \frac{\sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(0,1)})}{(\sigma_{(0,0)} - \sigma_{(0,1)})(1 - x_{(0,1)}(T)) + \sigma_{(0,0)}x_{(0,1)}(T)}. \tag{3.38}$$

"Successful" individuals of type $(0,0)$ are generated at rate

$$r_{(0,0)}^{(0,1)} N x_{(0,1)}(t)(1 - x_{(0,1)}(t)) p_{\text{fix}}^{(0,0)}(t). \tag{3.39}$$

Using this, we obtain for the probability that type $(0,1)$ fixes in the population:

$$\begin{aligned} P_{\text{det}}^{((0,1)\to(0,1))} &= \exp\left[-\int_0^\infty N r_{(0,0)}^{(0,1)} x_{(0,1)}(t)(1 - x_{(0,1)}(t)) p_{\text{fix}}^{(0,0)}(t) \mathrm{d}t \right] \\ &\approx \exp\left[-\int_{\bar\nu/N}^{1-\bar\nu/N} N r_{(0,0)}^{(0,1)} \frac{1}{\sigma_{(0,1)}} \frac{\sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(0,1)})}{(\sigma_{(0,0)} - \sigma_{(0,1)})(1-x) + \sigma_{(0,0)}x} \mathrm{d}x \right] \\ &= \left(\frac{\sigma_{(0,0)} - \frac{\bar\nu}{N}\sigma_{(0,1)}}{\sigma_{(0,0)} - \sigma_{(0,1)}(1 - \frac{\bar\nu}{N})} \right)^{-\frac{N r_{(0,0)}^{(0,1)} \sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(0,1)})}{\sigma_{(0,1)}^2}}, \end{aligned} \tag{3.40}$$

where the last equality holds for $\bar\nu < 0.5N$. For a single introgressed individual at time $t = 0$ in a large population, we can approximate $\bar\nu/N \approx 0$ and obtain

$$P_{\text{det}}^{((0,1)\to(0,1))} \approx \left(\frac{\sigma_{(0,0)}}{\sigma_{(0,0)} - \sigma_{(0,1)}} \right)^{-\frac{N r_{(0,0)}^{(0,1)} \sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(0,1)})}{\sigma_{(0,1)}^2}}. \tag{3.41}$$

Eq. (3.41) corresponds to Eq. (5) in Hartfield and Otto (2011) up to a model-specific factor of 2 in the exponent if we identify $s_a \equiv \sigma_{(0,0)}$, $s_d \equiv \sigma_{(0,0)} - \sigma_{(0,1)}$, and $r \equiv r_{(0,0)}^{(0,1)}$. If $N(1 - q_{(0,1)}) \approx N\sigma_{(0,1)}$ is small or if there are already other introgressed haplotypes sweeping in the population as in the generalization to more potential hitchhikers, it makes a quantitative difference whether one accounts for the fast initial increase or not, and we cannot approximate $\bar\nu/N \approx 0$ (see Appendix E). Alternatively, one can resort to a diffusion approach for these cases

(see HARTFIELD and OTTO (2011) and Appendix F). Note that both approaches assume that recombination is so weak that by itself, it does not influence the frequency path of type $(0,1)$ (i.e., $\sigma_{(0,1)} \gg r_{(0,0)}^{(0,1)}$).

3.5.4 Concatenation of the stochastic and the deterministic phase

In order to determine which haplotype fixes in the population, we need to concatenate the stochastic and the deterministic phase. Let $\mathcal{A}$ be the set of all types with positive fitness. Type (k,l) establishes in the stochastic phase with probability $P_{(k,l)}^{(I,J)}$ as derived and discussed in section 3.5.2 and hence enters the deterministic phase. We always assume that only one type does so (and it does so in a single copy). Given establishment of type (k,l), we denote by

$$P_{\text{det}}^{((k,l)\to(i,j))} \tag{3.42}$$

the probability that during the deterministic phase, type $(i,j) \in \mathcal{A}$ is generated and finally fixes in the population. Summing over all $(k,l) \in \mathcal{A}$ yields:

$$P(\text{“type } (i,j) \text{ fixes”}) = \sum_{(k,l)\in\mathcal{A}} P_{(k,l)}^{(I,J)} P_{\text{det}}^{((k,l)\to(i,j))}. \tag{3.43}$$

If not stated otherwise, the results presented in section 3.6 are based on Eq. (3.43) with $P_{(k,l)}^{(I,J)}$ obtained by Eq. (3.30) and Eq. (3.31). The recursions are performed by a program written in the C programming language. Approximations for $P_{\text{det}}^{(0,2)\to(0,\cdot))}$ with $\sigma_{(2,0)} - r_{(0,2)} > 0$ and $P_{\text{det}}^{(1,1)\to(\cdot,\cdot))}$ with $\sigma_{(1,1)} - r_{(1,1)} > 0$ are derived in Appendix E. All numerical evaluation of integrals that appear in these approximations is done in Mathematica (Wolfram Research, Champaign, USA).

3.6 Application to various biological scenarios

3.6.1 The impact of unlinked alleles

If $I = J = 0$, the extinction probability is given by Eq. (3.5). For Q_0, Q_1 and Q_2, we obtain:

$$Q_0 = 1 - \sigma_0, \tag{3.44a}$$

$$Q_1 = 2 - \sigma_1 - \frac{1}{2}Q_0 - \sqrt{\left(2 - \sigma_1 - \frac{1}{2}Q_0\right)^2 - 2(1 - \sigma_1)}, \tag{3.44b}$$

$$Q_2 = 4 - 2\sigma_2 - \frac{1}{2}Q_0 - Q_1 - 2\sqrt{\left(\frac{1}{4}Q_0 + \frac{1}{2} - 2 + \sigma_2\right)^2 - (1 - \sigma_2)}. \tag{3.44c}$$

How does the number of unlinked deleterious alleles impact the introgression probability if their total effect is kept constant? A comparison of Q_1 with $\sigma_1 = \sigma_0 - 2\sigma$ and Q_2 with $\sigma_1 = \sigma_0 - \sigma$ and $\sigma_2 = \sigma_0 - 2\sigma$ yields:

$$Q_2(\sigma)/Q_1(2\sigma) = 1 + \mathcal{O}(\sigma^2), \tag{3.45}$$

i.e., unless the deleterious effect is very strong, the extinction probability is approximately the same for both scenarios (either one deleterious allele of effect 2σ or two deleterious alleles each of effect σ). Figure 3.2 generalizes this result to $F > 2$. One sees that unlinked alleles significantly reduce the introgression probability. However, it is basically irrelevant whether there is one strongly deleterious allele or many slightly deleterious alleles. By how much do unlinked deleterious alleles of compound effect S_{del} reduce the introgression probability? Making use of the previous observation, it is sufficient to consider a single unlinked allele of effect S_{del}. A Taylor expansion yields:

$$\frac{1 - Q_1}{1 - Q_0} = 1 - \frac{2}{1 + \sigma_0} S_{\mathrm{del}} + \mathcal{O}(S_{\mathrm{del}}^2) \approx 1 - 2S_{\mathrm{del}} + \mathcal{O}(S_{\mathrm{del}}^2), \tag{3.46}$$

i.e., unlinked alleles approximately reduce the introgression probability by a factor that is independent of σ_0.

While unlinked alleles have a significant impact on the probability of adaptive gene introgression, they do not visibly influence the hitchhiking probability of closely linked deleterious alleles (cf. Figure 3.3).

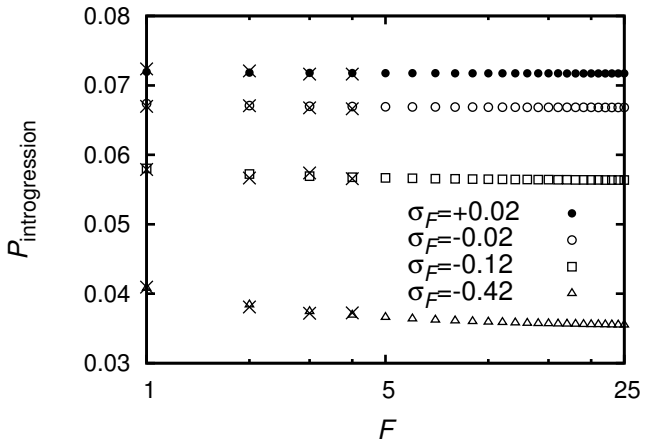


Figure 3.2: The introgression probability as a function of the number of unlinked deleterious alleles. The total effect on fitness is kept constant. The introgression probability is approximately the same whether the effect is distributed over few stronglly or many slightly deleterious alleles. The advantageous allele has Malthusian fitness $\sigma_0 = 0.08$. In the absence of deleterious alleles, it would establish with probability $1 - Q_0 = \sigma_{(0,0)} = 0.08$. Crosses denote simulation results. Each simulation point is the average over 10^6 introgression attempts.

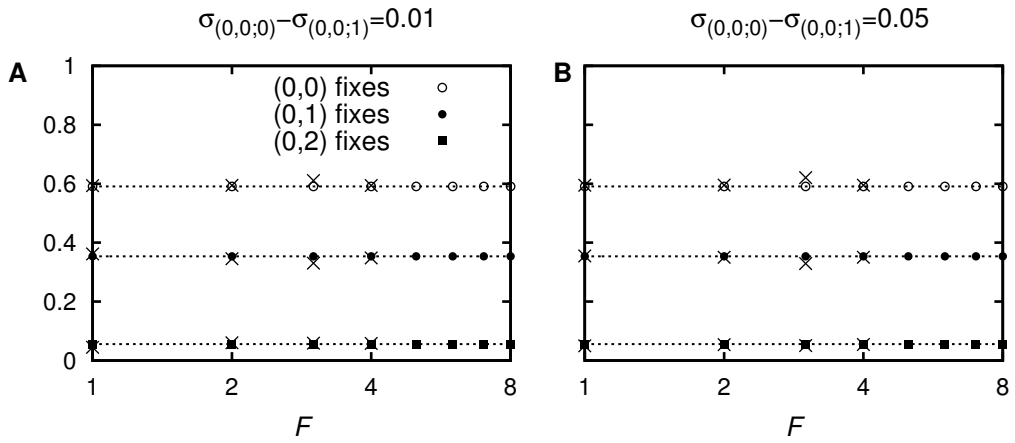


Figure 3.3: The hitchhiking probability as a function of the number of unlinked deleterious alleles. The dotted lines corrspond to the respective values for $F = 0$. Unlinked alleles do not visibly impact the hitchhiking probability of closely linked deleterious alleles. Parameter values are: $I = 0$, $J = 3$, $\sigma_0 = 0.075$, $\sigma_{(0,1)} = 0.07$, $\sigma_{(0,2)} = 0.05$, $\sigma_{(0,3)} = -0.015$, $N = 10,000$, $r^{(\cdot,\cdot)}_{(\cdot,\cdot)} = 0.0001$. Crosses denote simulation results. Each simulation point is the average over 2000 successful introgression events.

3.6.2 The impact of a single deleterious allele

In this section, we consider the impact of a single linked deleterious allele (cf. also Iwasa *et al.*, 2004a). From Eq. (3.4), we obtain:

$$Q_{(0,0)} = 1 - \sigma_{(0,0)}, \tag{3.47a}$$

$$Q_{(0,1)} = \frac{1}{2(1-r_{(0,1)})}\Bigg(2 - \sigma_{(0,1)} - r_{(0,1)}Q_{(0,0)} \tag{3.47b}$$

$$- \sqrt{\left(2 - \sigma_{(0,1)} - r_{(0,1)}Q_{(0,0)}\right)^2 - 4(1-\sigma_{(0,1)})(1-r_{(0,1)})}\Bigg) \tag{3.47c}$$

$$\approx \begin{cases} (1-\sigma_{(0,1)})\left(1 - \frac{\sigma_{(0,0)}-\sigma_{(0,1)}}{\sigma_{(0,1)}} r_{(0,1)}\right) & \text{for} \quad \sigma_{(0,1)} > 0, \\ 1 + \frac{\sigma_{(0,0)}}{\sigma_{(0,1)}} r_{(0,1)} & \text{for} \quad \sigma_{(0,1)} < 0. \end{cases} \tag{3.47d}$$

The approximation is a first order Taylor expansion in $r_{(0,1)}$, which yields accurate results for small r if $\sigma_{(0,1)}$ is not too close to zero. Due to the assumption of single crossover only, $Q_{(0,1)}$ exactly corresponds to Q_1 for $r_{(0,1)} = 0.5$ (where $\sigma_{(0,0)} \equiv \sigma_0$ and $\sigma_{(0,1)} \equiv \sigma_1$). How does a single deleterious allele impact the probability of adaptive gene introgression? We can measure the impact by the relative reduction of the introgression probability

$$\Delta P = 1 - \frac{1 - Q_{(0,1)}}{1 - Q_{(0,0)}} = 1 - \frac{1 - Q_{(0,1)}}{\sigma_{(0,0)}}. \tag{3.48}$$

If ΔP is close to zero, the deleterious allele has a weak impact; if ΔP is close to one, it has a strong impact. The influence is obviously strongest for $r_{(0,1)} = 0$. If $\sigma_{(0,1)} > 0$, the maximal relative reduction in the introgression probability is given by $s_{\text{del}}/\sigma_{(0,0)}$ with $s_{\text{del}} := \sigma_{(0,0)} - \sigma_{(0,1)}$. For $\sigma_{(0,1)} \leq 0$ and $r_{(0,1)} = 0$, the advantageous allele can not introgress at all (except through fixation by drift which is not considered here). For tight linkage, we use the foregoing Taylor expansion and obtain

$$\Delta P \approx \frac{s_{\text{del}}}{\sigma_{(0,0)}}\left(1 - \frac{r_{(0,1)}}{\sigma_{(0,0)} - s_{\text{del}}}\right) \quad \text{for} \quad \sigma_{(0,1)} > 0, \tag{3.49a}$$

$$\Delta P \approx 1 - \frac{r_{(0,1)}}{\sigma_{(0,1)}} \quad \text{for} \quad \sigma_{(0,1)} < 0. \tag{3.49b}$$

The maximal impact is strongest for a weak beneficial mutation (where introgression is easily reduced to zero for tight linkage). The impact gets weaker with increasing recombination on the scale of $\sigma_{(0,1)}$. I.e., if either $\sigma_{(0,0)} \gg s_{\text{del}}$, or if $s_{\text{del}} \gg \sigma_{(0,0)}$, the impact declines only slowly. In order to determine the behavior for strong recombination, we perform a Taylor expansion of ΔP in s_{del}:

$$\Delta P = \frac{1}{r_{(0,1)} + \sigma_{(0,0)}(1 - r_{(0,1)})} s_{\text{del}} + \mathcal{O}\left(s_{\text{del}}^2\right). \tag{3.50}$$

For $r_{(0,1)} \gg \sigma_{(0,0)}$, the relative reduction becomes independent of the strength of the beneficial allele. This is in strong contrast to the behavior for tight linkage case. The impact declines on a scale of s_{del} and becomes irrelevant for $r_{(0,1)} \gg s_{\text{del}}$. In particular, an unlinked deleterious allele leads to a relative reduction of $2s_{\text{del}}$.

If $\sigma_{(0,1)} - r_{(0,1)} > 0$, the deleterious allele can hitchhike to fixation. Type $(0,1)$ establishes with probability

$$1 - q_{(0,1)} = \frac{\sigma_{(0,1)} - r_{(0,1)}}{1 - r_{(0,1)}}, \tag{3.51}$$

i.e.,

$$P_{(0,1)}^{(0,1)} = \frac{\sigma_{(0,1)} - r_{(0,1)}}{1 - r_{(0,1)}} \frac{1}{1 - Q_{(0,1)}}, \tag{3.52}$$

and

$$P_{\text{hitchhiking}} = P_{(0,1)}^{(0,1)} P_{\text{det}}^{((0,1)\to(0,1))}. \tag{3.53}$$

In order to asses the respective relevance of the stochastic and the deterministic phase, we first perform a first-order Taylor expansion of $P_{(0,1)}^{(0,1)}$ in $r_{(0,1)}$:

$$P_{(0,1)}^{(0,1)} \approx 1 - r_{(0,1)} \sigma_{(0,0)} \frac{1 - \sigma_{(0,1)}}{\sigma_{(0,1)}^2} \approx 1 - r_{(0,1)} \frac{\sigma_{(0,0)}}{\sigma_{(0,1)}^2}. \tag{3.54}$$

I.e., changes in $P_{(0,1)}^{(0,1)}$ occur on the scale of $r_{(0,1)} \sim \sigma_{(0,1)}^2 / \sigma_{(0,0)}$. For the deterministic phase (Eq. (3.41)), the scale is set by $r_{(0,1)} \sim \sigma_{(0,1)}^2 / (N \sigma_{(0,0)} s_{\text{del}})$. This allows us to distinguish two parameter regimes: if $N s_{\text{del}} \gg 1$, the deterministic phase dominates. However, if $N s_{\text{del}} \approx 1$ or smaller, the stochastic phase cannot be ignored. Figures 3.4C and D illustrate how stochastic and deterministic phase combine to form the probability of hitchhiking for $N = 10000$ and $N = 500$, respectively. For $N = 10000$, the behavior is dominated by the deterministic phase which decays quickly as a function of $r_{(0,1)}$. In the parameter range, where $P_{\text{det}}^{((0,1)\to(0,1))}$ is

appreciable, one can ignore the influence of the stochastic phase. For $N = 500$, however, $P_{\text{det}}^{((0,1)\to(0,1))}$ decays slowly, and the stochastic phase has a non-negligible impact on hitchhiking: E.g., for $r_{(0,1)} = 0.003$, we find $P_{\text{det}}^{((0,1)\to(0,1))} \approx 0.8$ and $P_{(0,1)}^{(0,1)} P_{\text{det}}^{((0,1)\to(0,1))} \approx 0.53$. Note also that if $N(1 - q_{(0,1)})$ is small, we have to account for deviations from the deterministic path. These deviations can be accounted for via the parameter $\bar{\nu}$ as in Eq. (3.40) or via a diffusion approach as in Hartfield and Otto (2011). In Appendix F, we give the diffusion equation adjusted to our model and compare the result to Eq. (3.40).

A comparison between Panels A/B with Panels C/D of Figure 3.4 shows that the introgression probability changes only slightly over the depicted range of recombination, while the hitchhiking probability significantly decreases with increasing recombination distance; how quickly, is strongly affected by the population size.

3.6.3 The impact of a second deleterious allele

In this paragraph, we start from an individual of type $(0, 1)$ and investigate how a second deleterious allele affects the introgression and the hitchhiking probability, depending on the strength of selection and linkage. The results are summarized in Figures 3.5–3.8.

Figure 3.5 and Figure 3.6 consider the dependence on linkage. One sees that a strongly deleterious allele significantly affects the introgression probability even if it is only loosely linked. The impact on the hitchhiking probability is more subtle, and several cases have to be distinguished. We first turn to Figure 3.5, in which the second deleterious alleles is on the same side of the beneficial allele as the first one. Panel C shows the behavior for $\sigma_{(0,2)} - r_{(0,2)} < 0$. In that case, applying Eq. (3.30) with Eq. (3.31), it approximately holds:

$$P_{(0,1)}^{(0,2)} \approx \frac{r_{(0,1)}^{(0,2)}(1 - Q_{(0,1)})}{r_{(0,0)}^{(0,2)}(1 - Q_{(0,0)}) + r_{(0,1)}^{(0,2)}(1 - Q_{(0,1)})} \approx \frac{r_{(0,1)}^{(0,2)}\sigma_{(0,1)}}{r_{(0,0)}^{(0,2)}\sigma_{(0,0)} + r_{(0,1)}^{(0,2)}\sigma_{(0,1)}}, \quad (3.55)$$

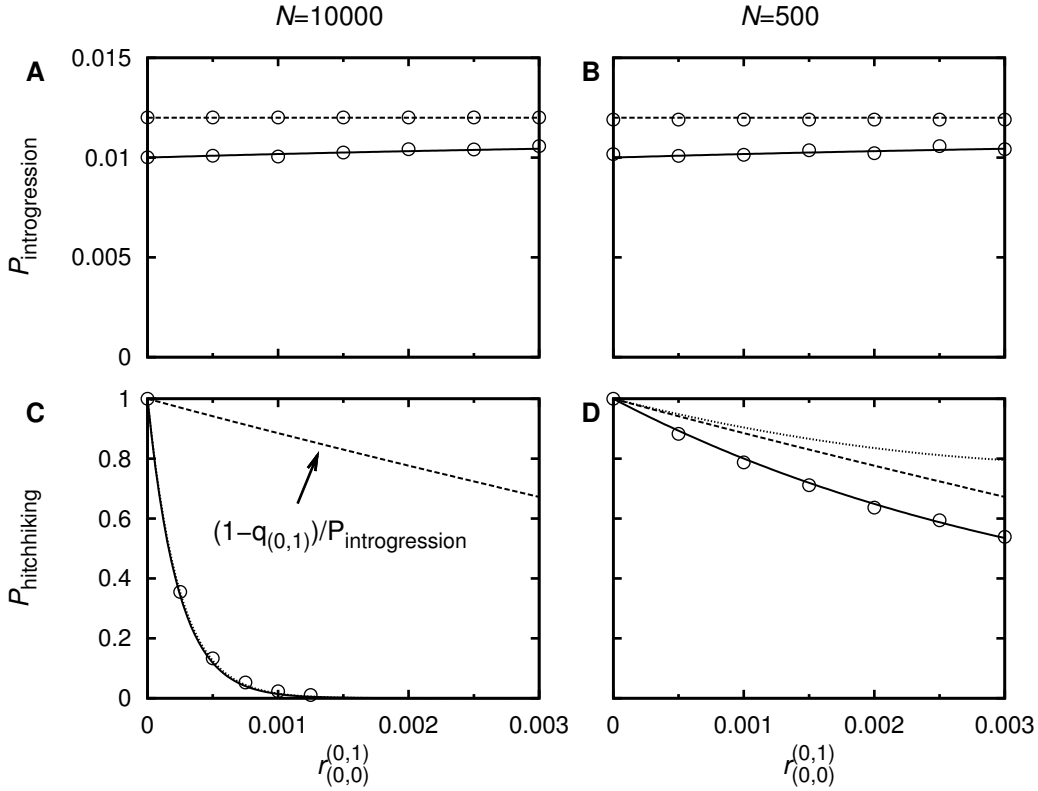


Figure 3.4: Panels A and B: Introgression probability as a function of linkage. For the solid line, it holds $I = 0, J = 1$. For the dashed line, $I = J = 0$. Panels C and D: The hitchhiking probability of a single deleterious allele ($I = 0$, $J = 1$). The solid line represents the hitchhiking probability as given by Eq. (3.53). The dashed and the dotted lines show the impact of the stochastic phase (Eq. (3.52)) and the deterministic phase (Eq. (3.40)), respectively. In Panel C, the solid and the dotted line are virtually indistinguishable; the deterministic phase dominates. In Panel D, the stochastic phase has a significant impact on the hitchhiking probability. Parameter values are: $\sigma_{(0,0)} = 0.012$, $\sigma_{(0,1)} = 0.01$. Circles denote simulation results. For Panels A and B, each simulation point is the average over 10^6 intogresstion attempts. For Panels C and D, each simulation point is the average over 10^4 successful introgression events.

where we have used $Q_{(0,1)} \approx 1 - \sigma_{(0,1)}$. Hence:

$$\begin{aligned} P_{\text{hitchhiking}}((0,2) \text{ is introduced}) &\approx \frac{r^{(0,2)}_{(0,1)}\sigma_{(0,1)}}{r^{(0,2)}_{(0,0)}\sigma_{(0,0)} + r^{(0,2)}_{(0,1)}\sigma_{(0,1)}} P^{((0,1)\to(0,1))}_{\text{det}} \\ &\approx \frac{r^{(0,2)}_{(0,1)}\sigma_{(0,1)}}{r^{(0,2)}_{(0,0)}\sigma_{(0,0)} + r^{(0,2)}_{(0,1)}\sigma_{(0,1)}} P_{\text{hitchhiking}}((0,1) \text{ is introduced}). \end{aligned} \tag{3.56}$$

I.e., for close linkage of the second deleterious allele, the hitchhiking probability gets strongly reduced. However, as linkage gets looser, it converges quickly to its value in absence of the second allele $P^{((0,1)\to(0,1))}_{\text{det}}$. Denote by c the percentage by which the second deleterious allele reduces the hitchhiking probability of the first one:

$$c = \frac{P_{\text{hitchhiking}}((0,1) \text{ is introduced}) - P_{\text{hitchhiking}}((0,2) \text{ is introduced})}{P_{\text{hitchhiking}}((0,1) \text{ is introduced})}. \tag{3.57}$$

By rearrangement of terms, we obtain

$$\frac{r^{(0,2)}_{(0,1)}}{r^{(0,2)}_{(0,0)}} = \frac{1-c}{c}\frac{\sigma_{(0,0)}}{\sigma_{(0,1)}} \quad \Leftrightarrow \quad c = \frac{1}{1 + \frac{r^{(0,2)}_{(0,1)}\sigma_{(0,1)}}{r^{(0,2)}_{(0,0)}\sigma_{(0,0)}}}. \tag{3.58}$$

Importantly, the selective strength of the second deleterious allele has no effect (as long as it is strong enough for our approximation to apply). Also the other selection coefficients only play a minor role. The second deleterious allele is relevant if

$$r^{(0,2)}_{(0,1)} < r^{(0,1)}_{(0,0)}\frac{\sigma_{(0,0)}}{\sigma_{(0,1)}} \approx r^{(0,1)}_{(0,0)} \quad \text{for} \quad \sigma_{(0,0)} \gg s_{\text{del}}, \tag{3.59}$$

which is independent of the selection coefficients if $\sigma_{(0,0)} \gg s_{\text{del}}$. The hitchhiking probability is crucially determined by the ratio of recombination distances $r^{(0,2)}_{(0,1)}/r^{(0,1)}_{(0,0)}$.

In Panel D, the second deleterious allele can hitchhike to fixation, too ($\sigma_{(0,2)} - r_{(0,2)} > 0$). The total hitchhiking probability of the closest deleterious allele is then only moderately reduced. For $r^{(0,2)}_{(0,1)}$ small, the second deleterious allele fixes, too. We see that with increasing recombination distance the analytical result underestimates the true hitchhiking probability. This has two reasons: First, the approximation $P^{(0,2)}_{(0,2)} \approx 1$ becomes worse as recombination increases. Second (and

more importantly here), it is not unlikely that type $(0,1)$ and type $(0,2)$ establish simultaneously while our approximation assumes that type $(0,1)$ establishes after type $(0,2)$. A slight shift in the time of establishment as introduced by our approach can lead to appreciable deviations of the frequency paths of both types from the true paths. This, in turn, has a non-negligible impact on the assessment of the hitchhiking probability.

In Figure 3.6, the beneficial allele is flanked by deleterious alleles. If the second allele is strongly deleterious, the hitchhiking probability is not visibly reduced (Panel B). This is because for successful introgression, the allele has to recombine away very early. The situation in Panel D looks similar to Figure 3.5C. We have

$$\begin{aligned} &P_{\text{hitchhiking}}((1,1) \text{ is introduced}) \\ &\approx \frac{r^{(1,1)}_{(0,1)}\sigma_{(0,1)}}{r^{(1,1)}_{(1,0)}\sigma_{(1,0)} + r^{(1,1)}_{(0,1)}\sigma_{(0,1)}} P_{\text{hitchhiking}}((0,1) \text{ is introduced}). \end{aligned} \tag{3.60}$$

This is formally similar to Eq. (3.56). There is, however, an important difference: Now, the selection coefficient of the second deleterious allele is crucial. Its influences ceases with increasing strength. If both deleterious alleles have approximately the same effect ($\sigma_{(0,1)} \approx \sigma_{(1,0)}$), the behavior is again determined by the ratio of the recombination distances $r^{(1,1)}_{(0,1)}/r^{(1,1)}_{(1,0)}$.

Figures 3.7 and 3.8 show how the selective disadvantage of a second closely linked deleterious allele affects the hitchhiking probability. If it is on the same side of the beneficial allele as the first one (Figure 3.7), the hitchhiking probability is greatly reduced unless the selective disadvantage is very slight. The reduction is greatest for intermediate values of the selection coefficient (see Figure 3.7A). In this parameter regime, type $(0,2)$ significantly increases in frequency before a successful lineage of type $(0,0)$ or $(0,1)$ is generated. As a consequence, the time to fixation of the beneficial allele is relatively long. Even if a sucessful lineage of type $(0,1)$ can establish, it is therefore likely that later, a sucessful lineage of type $(0,0)$ is generated (see Figure 3.7B for an illustration of this reasoning). If the beneficial allele is flanked at equal small recombination distances by two deleterious alleles as in Figure 3.8, the total hitchhiking probability of the deleterious allele to the right (fixation of type $(0,1)$ or type $(1,1)$) is barely influenced by the presence of the second deleterious allele, irrespective of the selective disadvantage of the latter. Note, however, that recombination is weak in Figure 3.8. For strong recombination and $\sigma_{(0,1)} \approx \sigma_{(1,0)}$, it is not unlikely that both a successful lineage

of type $(0,1)$ and of type $(1,0)$ establish and coexist for a long time, making the production of a successful $(0,0)$ recombinant very likely.

3.6.4 The impact of several linked deleterious alleles

To start with, assume $\sigma_{(I,J)} - r_{(I,J)} > 0$. If all recombination distances are small, we can generalize the result Eq. (3.41) and calculate the probability that all deleterious alleles hitchhike to fixation. For weak recombination, the frequency increase of type (I,J) is again well described by the deterministic path $x_{(I,J)}$. At time t, an individual of type (i,J) has Malthusian fitness

$$\sigma_{(i,J)}(t) = \sigma_{(i,J)} - x_{(I,J)}(t)\sigma_{(I,J)}. \tag{3.61}$$

Since recombination is assumed to be weak, the number of offspring of its own type is not significantly reduced by recombination. The establishment probability $p_{\text{est}}^{(i,J)}(t)$ of type (i,J) is hence obtained by suitable substitutions in Eq. (3.38). Taking all possible types into account, sucessful lineages with fewer deleterious alleles are generated at rate

$$Nx_{(I,J)}(t)(1-x_{(I,J)}(t))\left(\sum_{i=0}^{(I-1)} r_{(i,J)}^{(I,J)} p_{\text{est}}^{(i,J)}(t) + \sum_{j=0}^{J-1} r_{(I,j)}^{(I,J)} p_{\text{est}}^{(I,j)}(t)\right). \tag{3.62}$$

Analogous to the derivation of Eq. (3.41), we obtain for the probability that all deleterious alleles hitchhike to fixation

$$\begin{aligned} Q_{\text{det}}^{(I,J)} = &\prod_{i=0}^{I-1}\left(\frac{\sigma_{(i,J)}}{\sigma_{(i,J)}-\sigma_{(I,J)}}\right)^{-\frac{Nr_{(i,J)}^{(I,J)}\sigma_{(i,J)}(\sigma_{(i,J)}-\sigma_{(I,J)})}{\sigma_{(I,J)}^2}} \\ &\times \prod_{j=0}^{J-1}\left(\frac{\sigma_{(I,j)}}{\sigma_{(I,j)}-\sigma_{(I,J)}}\right)^{-\frac{Nr_{(I,j)}^{(I,J)}\sigma_{(I,j)}(\sigma_{(I,j)}-\sigma_{(I,J)})}{\sigma_{(I,J)}^2}}. \end{aligned} \tag{3.63}$$

We now turn to $\sigma_{(I,J)} - r_{(I,J)} < 0$. Be $\mathcal{A}$ the set of all types which have positive fitness and $(i,j) \in \mathcal{A}$. First note:

$$P((k,l) \in \mathcal{A}, k > i \text{ or } l > j, \text{ fixes}) < P_{(i,j)}^{(I,J)}. \tag{3.64}$$

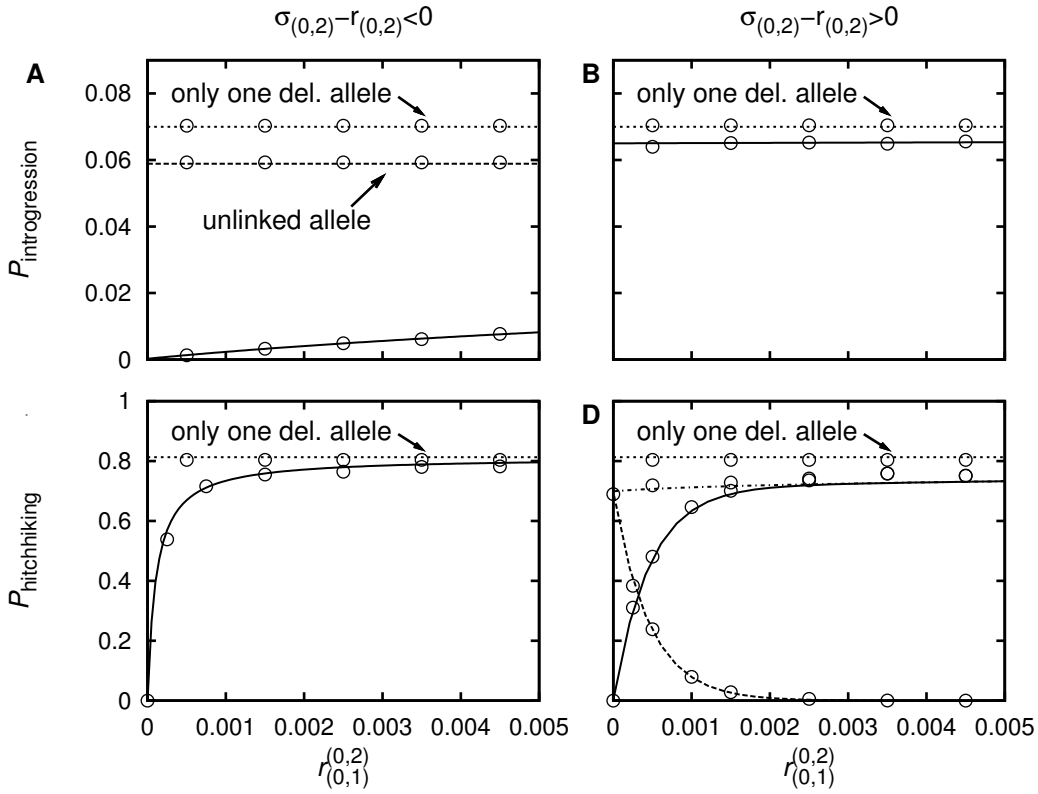


Figure 3.5: Panels A and B: Introgression probability of an adaptive allele linked to two deleterious alleles. Panels C and D: Hitchhiking probability of a closely linked deleterious allele. In each Panel, the dotted line gives the respective probability in the absence of the second deleterious allele. In Panel A, the dashed line gives the introgression probability for one closely linked and one unlinked allele. In Panel D, the solid line gives the probability that the haplotype $(0,1)$ fixes. The dashed line gives the probability that the haplotype $(0,2)$ fixes. The dash-dotted line gives the probability that either of the two fixes. For Panels A and C: $\sigma_{(0,2)} = -0.03$. For Panels B and D: $\sigma_{(0,2)} = 0.065$. The other parameter values are: $\sigma_{(0,0)} = 0.075$, $\sigma_{(0,1)} = 0.07$, $N = 10,000$, $r_{(0,0)}^{(0,1)} = 0.0001$. Circles denote simulation results. For Panels A and B, each simulation point it the average over 10^6 introgression attempts. For Panels, C and D, each simulation point is the average over 2000 successful introgression events.

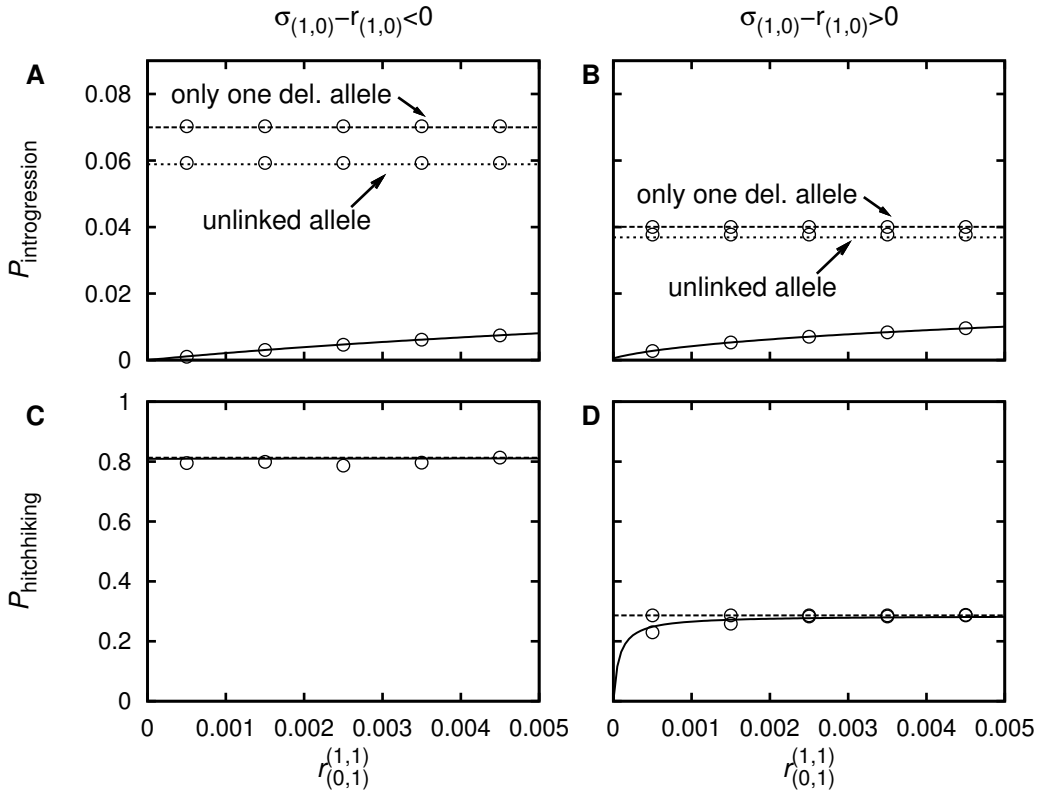


Figure 3.6: Panels A and B: Introgression probability of an adaptive allele linked to two deleterious alleles. Panels C and D: Hitchhiking probability of a closely linked deleterious allele. In each Panel, the dashed line gives the respective probability in the absence of the second deleterious allele. In Panels A and B, the dotted line gives the introgression probability for one closely linked and one unlinked allele. For Panels A and C: $\sigma_{(0,1)} = 0.07$, $\sigma_{(1,0)} = -0.025$, $\sigma_{(1,1)} = -0.03$. For Panels B and D: $\sigma_{(0,1)} = 0.04$, $\sigma_{(1,0)} = 0.03$, $\sigma_{(1,1)} = -0.005$. The other parameter values are: $\sigma_{(0,0)} = 0.075$, $N = 10,000$, $r^{(1,1)}_{(1,0)} = 0.0001$. Circles denote simulation results. For Panels A and B, each simulation point it the average over 10^6 introgression attempts. For Panels, C and D, each simulation point is the average over 2000 successful introgression events.

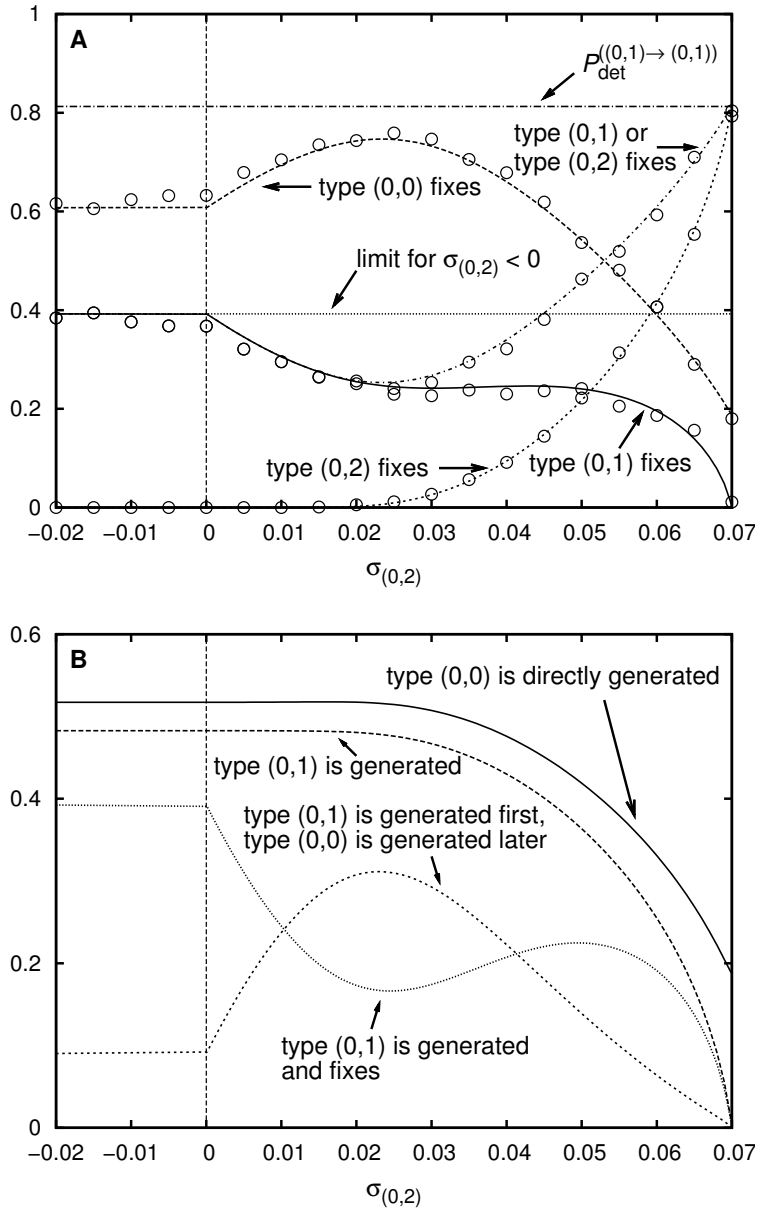


Figure 3.7: Panel A: Hitchhiking probability of closely linked deleterious alleles as a function of the Malthusian fitness parameter $\sigma_{(0,2)}$. The plot shows the probabilities at which the various haplotypes fix in the population. Panel B: Illustration of the respective pathways of recombination and establishment of types $(0,0)$ and $(0,1)$. Parameter values are: $I = 0$, $J = 2$, $\sigma_{(0,0)} = 0.075$, $\sigma_{(0,1)} = 0.07$, $N = 10,000$, $r_{(0,0)}^{(0,1)} = r_{(0,1)}^{(0,2)} = 0.0001$. Circles denote simulation results. Each simulation point is the average over 2000 successful introgression events.

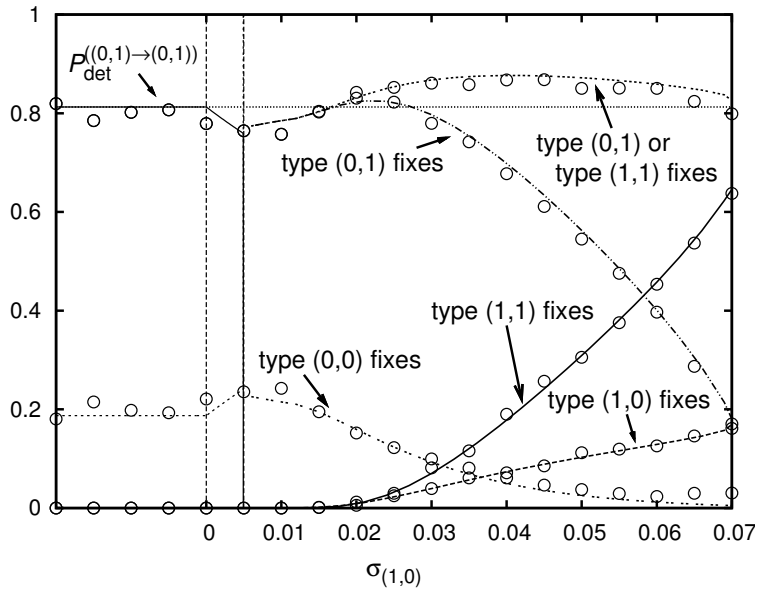


Figure 3.8: Hitchhiking probability of closely linked deleterious alleles as a function of the Malthusian fitness parameter $\sigma_{(1,0)}$. The plot shows the probabilities at which the various haplotypes fix in the population. Parameter values are: $I = 1$, $J = 1$, $\sigma_{(0,0)} = 0.075$, $\sigma_{(0,1)} = 0.07$, $N = 10,000$, $r_{(0,0)}^{(0,1)} = r_{(0,0)}^{(1,0)} = 0.0001$. Circles denote simulation results. Each simulation point is the average over 2000 successful introgression events.

Consider the special case $I = 0$, $\sigma_{(0,j)} - r_{(0,j)} > 0$, and $\sigma_{(0,l)} - r_{(0,l)} < 0$, $l > j$. For tight linkage, we can again approximate $1 - Q_{(0,k)} \approx \sigma_{(0,k)}$ for $k \leq j$ and obtain (proof by induction):

$$P_{(0,k)}^{(0,J)} \approx \frac{r_{(0,k)}^{(0,J)} \sigma_{(0,k)}}{\sum_{l=0}^{j} r_{(0,l)}^{(0,J)} \sigma_{(0,l)}}. \tag{3.65}$$

The $(j+1)$th deleterious allele has a strong effect on the hitchhiking probabilities; all further alleles have no effect though. Moreover, the strength of the $(j+1)$th allele is irrelevant as long the type $(0, j+1)$ is sufficiently deleterious that our approximation is justified.

As another example, consider the special case $J = 1$, $\sigma_{(0,1)} - r_{(0,1)} > 0$, and $\sigma_{(i,j)} - r_{(i,j)} < 0$, $i > 0$. If linkage is tight, the hitchhiking probability is barely reduced by additional deleterious mutations.

Finally, be $\sigma_{(0,1)} - r_{(0,1)} > 0$ and I and $J \geq 1$ arbitrary and $\sigma_{(i,j)} - r_{(i,j)} < 0$ for $(i,j) \notin \{(0,0),(0,1)\}$. Figure 3.9 shows how additional deleterious alleles that can themselves not hitchhike to fixation can influence the hitchhiking probability of a slightly deleterious allele. The pattern can be understood by consideration of the various paths which lead to establishment of the beneficial allele: unless $I = 0$ (or $J = 1$), at least two recombination events are necessary to generate a type with positive Malthusian fitness. The position of the first successful recombination event depends on the fitness of the types that are generated by recombination. Since $\sigma_{(1,0)}$ is only slightly deleterious, the first recombination event is likely to generate this type if $I = 1$. In this case, the hitchhiking probability is strongly reduced (cf. the dip in Figure 3.9). For $I > 1$, however, type $(1,0)$ cannot be generated via a single recombination event, and the reduction is less pronounced, getting smaller with increasing I. For large I, adding more deleterious alleles to the left or the right has only a weak effect. Generally, deleterious alleles that render haplotypes strongly disfavored have to be lost as quickly as possible, and the pathway to establishment of the beneficical allele does usually not involve tunneling via more strongly deleterious haplotypes than necessary. As a rule of thumb, the impact of these deleterious alleles can therefore be accounted for by a single allele of the compound effect. This virtual allele is located at the position of these alleles that is closest to the adaptation.

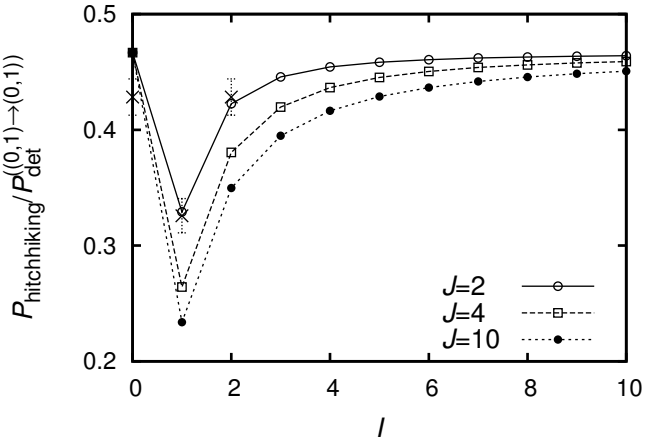


Figure 3.9: Reduction of the hitchhiking probability of a closely linked deleterious alleles by the impact of other deleterious alleles. On the x-axis, the number I of deleterious alleles to the left of the beneficical allele is varied. Only types $(0,0)$ and $(0,1)$ have positive fitness. The fixation probability of type $(0,1)$ conditioned on fixation of the adaptative allele ($P_{\text{hitchhiking}}$) displays a pronounced dip for $I = 1$. Parameter values are: $\sigma_{(0,0)} = 0.08$, $\sigma_{(0,1)} = 0.07$, $\sigma_{(0,2)} = -0.005$, $\sigma_{(1,0)} = -0.002$, all other deleterious alleles have an effect of -0.01, $r^{(k,l)}_{(i,j)} = 0.0001$ for all k, l, i, j, $N = 10,000$. Crosses denote simulation results. Each simulation point is the average over 1000 successful introgression events.

3.7 Discussion

Gene flow between related species is relatively frequent. Although many foreign alleles burden its carrier with a selective disadvantage, an exchange of genetic material between populations is often still possible. If neutral or advantageous alleles survive the fitness bottleneck caused by linked or unlinked deleterious alleles they can become permanently incorporated into the genome of the sister species. Picking up locally adaptive alleles from an indigenous species can help species to expand their range to originally inhabitable regions. Adaptive gene introgression is hence a clever evolutionary mechanism that can speed up adaptation to novel environments. Human activities create ample opportunity for hybridization between domestic animals or crop plants with their wild relatives which can cause permanent ecological damage. In this context, introgression of alleles from genetically modified organisms (e.g., insecticide resistance genes) into weed is recognized as a particular risk. A quantitative analysis of the introgression process is essential both to assess the importance of adaptive gene introgression as an evolutionary pathway to adaptation and to estimate the ecological risks associated with unwanted hybridization.

The reduction in hybrid fitness by deleterious alleles is a crucial factor in the potential introgression of adaptations. If their compound effect outweighs the benefits of the adaptation, deleterious alleles must be lost before the favorable allele can establish. Closely linked slightly deleterious alleles might be dragged along to fixation. In this paper, we developed a framework to investigate the role of linked and unlinked deleterious alleles for adaptive gene introgression. The model accounts for an explicit genetic structure and describes the genetic evolution of a haploid population under the influence of selection, recombination, and drift after a single hybridization event. The analysis is based on the theory of branching processes. In particular, the early phase of spread of the advantageous allele can be well approximated by a reducible multitype branching process with a special structure: within the branching process approximation, offspring are either of the same type as their parent or carry fewer deleterious alleles (for similar setups see Barton and Bengtsson, 1986; Demon *et al.*, 2007; Gosh and Haccou, 2010; Gosh *et al.*, 2012a,b). This distinctive structure allows for the analytical treatment of several aspects of the problem.

The introgression probability How likely is it that the advantageous allele can establish itself in the population? For large populations, the survival proba-

bility of the branching process is a very good approximation to the probability of adaptive gene introgression. The extinction probability of a multitype branching process is in general difficult to determine, and one has to resort to approximate formulas and numerical methods (e.g., BARTON, 1995; IWASA *et al.*, 2003, 2004b; SERRA and HACCOU, 2007). However, in our special case, a recursive solution can be derived and readily permits to determine the introgression probability for a specific allele configuration.

We find that deleterious alleles significantly hamper the introgression of an adaptive allele even if they are only loosely linked or unlinked. A single linked allele can lead to a strong reduction in the introgression probability. For tight linkage, its impact ceases with increasing recombination distance at a scale of the selection coefficient $\sigma_{(0,1)}$ of the initial type. For strong recombination, the rate at which its effect fades with recombination is given by the strength s_{del} of the deleterious allele and independent of the strength of the adaptation. If strongly deleterious, the effect of a single deleterious allele can still be appreciable at large distances and even if it is unlinked. We also find that the influence of several unlinked alleles is very well approximated by the influence of a single deleterious allele of the compound effect. They reduce the introgression probability by a factor which is roughly independent of the strength of the adaptive allele.

It is instructive to compare this to a situation where beneficial and deleterious alleles appear by mutation rather than introgression as discussed in BARTON (1995). In that case, the deleterious alleles segregate in mutation-selection balance in the population when the advantageous allele appears. Consider first the limiting case where a deleterious allele segregates at a single locus at frequency u. The beneficial mutation can arise on a genome carrying or not carrying the deleterious allele. Technically, the introgression probability corresponds to the fixation probability of the beneficial mutation given it arises on a genome with the deleterious allele (denoted by P_u in BARTON (1995)). In that case, its fixation probability can be significantly reduced. However, the reduction is in general much weaker than if the two alleles enter the population via a single introgression event. This is because relative fitness of the double mutant is higher if the deleterious allele segregates in the population. A numerical comparison confirms that the result derived in BARTON (1995) ($P_u/(2\sigma_{(0,0)})$ as given via Eq. (16) and (17a) in BARTON (1995)) converges to $(1 - Q_{(0,1)})/\sigma_{(0,0)}$ (cf. Eq. (3.47c)) as the mutation rate and hence the frequency of the deleterious allele tend to zero (note that the results in BARTON (1995) are based on a Poisson distribution of the offspring number such that the establishment probability of an isolated beneficial

allele is $2\sigma_{(0,0)}$ while it is $\sigma_{(0,0)}$ in our model). However, whether the beneficial allele appears on a genome with or without the deleterious allele depends on the relative frequency u of the deleterious allele in the population, and the biological relevant quantitiy is the weighted average of the fixation probabilities on the two genetic backgrounds. For $\sigma_{(0,1)} < 0$ and complete linkage, the weighted fixation probability of the beneficial mutation is reduced by u relative to its value in absence of the deleterious allele (Barton, 1995, Eq. (17b)). In contrast, the introgression probability (as well as P_u) are zero in that case. For $\sigma_{(0,1)} > 0$, the relative reduction is at most $u(s_{\text{del}}/\sigma_{(0,0)})^2$, i.e., much smaller than for introgression ($s_{\text{del}}/\sigma_{(0,0)}$). Note that the reduction in the weighted fixation probability is caused by the recurrent generation of deleterious alleles that drop on genomes carrying the adaptation. The presence of deleterious alleles itself even slightly increases the weighted fixation probability (this term is very small and neglected in Barton (1995)). Barton (1995) finds that for two loci flanking the beneficical mutation, the effects of the two deleterious alleles approximately multiply. This does not hold true for the different biological problem of adaptive gene introgression. Concerning the impact of unlinked alleles, similarities arise: in both cases, their impact on fixation of the adaptive allele can be approximated by a factor that sums the effects of the unlinked loci and is approximately independent of the strength of the adaptation. However, again, if the deleterious alleles appear by mutation, multiple genetic backgrounds exist and the adaptation can drop on any of them, while subsequently to hybridization, adaptive and deleterious alleles inherently appear together. It is hence not surprising that the factor describing the reduction in the fixation probability differs for the two scenarios.

Linked deleterious alleles can render successful introgression after a single hybridization event extremely unlikely. In order to assess whether even introgression probabilities of the order of $10^{-8} - 10^{-6}$ are still evolutionary relevant, it is helpful to compare these values to the probability of adaptation by de-novo mutations. With a point mutation probability of $\sim 10^{-8}$ and a selective advantage of 1%, the probability that a specific mutation occurs in a specific individual and thereafter rises to fixation, is $\sim 10^{-10}$. For complex adaptations, the probability is even lower. Depending on the probability of hybridization, adaptive gene introgression can hence be a relevant evolutionary process. Hybridization rates are potentially high, and even if the success probability of each single one is low, the probability that any hybridization event is followed by adaptive gene introgression is appreciable. This consideration is particularly important in an agricultural context where (genetically modified) crops grow next to wild plants

in large areas all over the world for many years. Gosh and Haccou (2010) and Gosh *et al.* (2012a,b) therefore suggest the so-called hazard rate as a measure for risk assessment as the hazard rate takes both the hybridization rate and the introgression probability into account.

The hitchhiking probability Weakly deleterious alleles that are closely linked to the adaptive allele can hitchhike to fixation. We developed a framework to estimate which haplotype finally fixes in the population, depending on the alien haplotype that was originally introduced. The approach is based on a split of the process into two phases: the establishment phase of the adaptive allele and the sweep during which further deleterious alleles can be lost. What is the respective relevance of the stochastic and the deterministic phase in this scenario? In the simplest case, there are only two loci under selection: one locus with the advantageous and one locus with a deleterious allele that can hitchhike to fixation. We can distinguish two parameter regimes: if the product of the selection coefficient s_{del} and the population size is small, the impact of the stochastic phase is significant. However, if the product of selection and population size is large, the stochastic phase can be ignored and the behavior is dominated by the deterministic phase. This is different if additional deleterious alleles render the initial haplotype itself deleterious. In that case, establishment of the adaptation is contingent on the early loss of deleterious alleles, and depending on the allelic configuration, the stochastic establishment phase will have a strong impact on hitchhiking irrespective of the population size. To good approximation, all alleles that cause serious maladaptation and are located on the same side of the beneficial allele can be summarized to a single allele of the compound effect, reducing the dimensionality of the problem. The impact of these additional alleles fades quickly with increasing recombination distance. Unlinked alleles have no visible effect on hitchhiking. These insights essentially generalize to more than one possible deleterious hitchhiker.

Hartfield and Otto (2011) analyze the hitchhiking probability of a single deleterious allele in the absence of other deleterious alleles. They present two approaches to the problem: a semi-deterministic approach based on branching process theory (which also serves as the basis for our analysis of the deterministic phase), and a diffusion approach. In both cases, however, they condition on establishment of type $(0, 1)$. Their analysis thus ignores aspects of the stochastic establishement phase and consequently applies only to the regime of large Ns_{del}.

Selection & recombination in the introgression process Summarizing, we can identify three fundamentally different genomic scales in units of the recombination rate that matter for adaptive gene introgression. First, deleterious alleles on the introgression haplotype affect the introgression probability of the beneficial allele across distances of the order of the deleterious selection coefficient ($r \sim s_{\text{del}}$). Strong deleterious alleles thus still matter even if they are unlinked. Importantly, the absolute strength of selection is crucial for the failure or success of introgression. For the hitchhiking probability of a single deleterious allele, we find two relevant scales, stemming from the stochastic and the deterministic phase of the hitchhiking process, respectively. For the stochastic phase, this scale is set by the selection coefficient of the haplotype with both the beneficial and the deleterious allele ($r \sim \sigma_{(0,1)}$). Similarly to the introgression probability it is thus the strength of selection that matters. In contrast, for the deterministic phase, the scale is primarily set by the inverse population size ($r \sim 1/N$). Usually (but not always), effects from the deterministic phase are relevant already over much shorter distances than effects from the stochastic phase and will therefore dominate. In contrast to the establishment phase, only the ratio of the selection coefficients matters. Finally, the impact of further deleterious alleles on the hitchhiking probabilty of a focal deleterious allele may depend on the genomic configuration in sublte ways. For cases where an additional deleterious allele at distance R to the hitchhiking allele suppresses the fitness of that haplotype below one, we find that the relevant scale is primarily set by the recombination distance between the beneficial allele and the deleterious hitchhiker, i.e., $R \sim r$.

Limitations and extensions The mathematical analysis of the model has two major restrictions. The first one is the common constraint of branching process approximations for establishment probabilities of beneficial alleles: for small populations, the branching process approach underestimates the true probability. In particular, our multitype branching process can contain supercritical, critial, and subcritical types. If the population size is small, fixation of deleterious haplotypes by genetic drift can be more likely than survival of the branching process. The second restriction concerns the analysis of the hitchhiking probability. The derived approximations rely on the assumption that initially, a single haplotype establishes and starts sweeping before haplotypes with less deleterious alleles possibly establish. This assumption is well justified in many scenarios. However, in particular, when linkage is not tight, deviations can arise. In that case, it seems a promising approach to analyse the multitype branching process by means of

the probability generating function for the number of the various haplotypes in dependence of time. A differential equation for this p.g.f. can be derived via infinitesimal generating functions (cf. KARLIN and TAYLOR, 1975, p. 412ff). This approch could help answering open questions on the stochastic phase. However, the results would probably take a complex form and involve numerical evaluation of terms while the approximations derived in this paper constitute intuitive expressions that can be easily evaluated. Likewise, the approximations for the deterministic phase require sufficiently tight linkage, since otherwise, recombination affects the shape of the deterministic path.

The model assumes a population of haploid individuals or diploids without dominance, i.e., it does not allow for under- or overdominance of alleles. The results for the introgression probability can be directly generalized to apply to diploids with dominance unless the beneficial allele is completely recessive: as long as introgressed alleles are rare in the population, they only appear in heterozygotes. Eq. (3.2) therefore still applies if fitness refers to heterozygote fitness. As soon as the introgressed alleles become more frequent, both copies of an individual's chromosome might carry introgressed material. The sweep of a selectively favored haplotype is therefore strongly altered if the alleles display over- or underdominance. The length and shape of this frequency path is, however, crucial for the hitchhiking probability. In principle, our approach can be extended by this element. The model furthermore assumes that selection against the deleterious alleles is independent of the genetic background. Recombination therefore increases the fitness of later generations of hybrids. However, by recombination, incompatibilities can be generated. In that case, later hybrids display a lower fitness than early generation hybrids until further recombination removes the incompatible alleles.

Beyond these genetic confinements, the second class of model extensions concerns the ecological setting. The model assumes that the alien genome arrives by long range dispersal in a panmictic population. An important extension is the incorporation of spatial structure. If the population cannot be assumed to be panmictic, the beneficial allele establishes first locally before it spreads by migration. Even if deleterious alleles can initially increase to high frequencies it is likely that they become lost later by recombination with wildtype individuals in other parts of the habitat. On the other hand, hybridization often occurs at the edge of a species range, and adaptive gene introgression can entail a range expansion. The latter reduces recombination with wildtype individuals and increases hence the hitchhiking probability of weakly deleterious alleles. A very different

situation from the one analyzed in this paper arises if dispersal is local and strong and recurrent gene flow builds up a hybrid zone. Under these circumstances, as discussed in Barton (1979), a single-locus cline does not significantly hamper the spread of a beneficial allele from one population to the other.

To summarize, we set up a minimal model in order to reveal fundamental principles that are effective in the introgression process of a favorable allele. In particular, the analysis helps to build an intuitive understanding of how deleterious alleles impact adaptive gene introgression.

Acknowlegdments

We thank Matthew Hartfield for helpful comments on the manuscript. This work was made possible with financial support by the Vienna Science and Technology Fund (WWTF), by the Deutsche Forschungsgemeinschaft (DFG), Research Unit 1078 *Natural selection in structured populations*, and by the Austrian Science Fund (FWF) via the doctoral program (DK) population genetics.

A Appendix: Proof of Theorem 3.4.1

We can reinterprete the branching process as follows: An individual of type $(i, j; f)$ "dies" at rate $2 - \sigma_{(i,j;f)}$ and at death, it produces either zero or two offspring, one of which is of its own type. We now consider the embedded discrete-time process:

- With probability $\frac{1-\sigma_{(i,j;f)}}{2-\sigma_{(i,j;f)}}$, it has 0 offspring.
- With probability $\binom{f}{g}\left(\frac{1}{2}\right)^f \frac{1-r_{(i,j)}}{2-\sigma_{(i,j;f)}}$, it produces an offspring of its own type and an offspring of type $(i, j; g)$ with $g \leq f$.
- With probability $\binom{f}{g}\left(\frac{1}{2}\right)^f \frac{r^{(i,j)}_{(i,k)}}{2-\sigma_{(i,j;f)}}$, it produces an offspring of its own type and an offspring of type $(i, k; g)$ with $g \leq f$ and $k < j$.
- With probability $\binom{f}{g}\left(\frac{1}{2}\right)^f \frac{r^{(i,j)}_{(k,j)}}{2-\sigma_{(i,j;f)}}$, it produces an offspring of its own type and an offspring of type $(k, j; g)$ with $g \leq f$ and $k < i$.

Within this scheme, the offspring generating function of an individual of type $(i, j; f)$ is given by

$$\begin{aligned} G^{(i,j;f)}(\mathbf{s}) = &\left(\frac{1}{2}\right)^f \frac{1 - r_{(i,j)}}{2 - \sigma_{(i,j;f)}} s^2_{(i,j;f)} \\ &+ \sum_{g=0}^{f} \binom{f}{g} \left(\frac{1}{2}\right)^f \left\{ \sum_{k=0}^{i-1} \frac{r^{(i,j)}_{(k,j)}}{2 - \sigma_{(i,j;f)}} s_{(k,j;g)} + \sum_{k=0}^{j-1} \frac{r^{(i,j)}_{(i,k)}}{2 - \sigma_{(i,j;f)}} s_{(i,k;g)} \right\} s_{(i,j;f)} \\ &+ \sum_{g=0}^{f-1} \binom{f}{g} \left(\frac{1}{2}\right)^f \frac{1 - r_{(i,j)}}{2 - \sigma_{(i,j;f)}} s_{(i,j;g)} s_{(i,j;f)} + \frac{1 - \sigma_{(i,j;f)}}{2 - \sigma_{(i,j;f)}}, \end{aligned} \tag{A.1}$$

where $\mathbf{s}$ is a vector with elements $\left\{ s_{(k_1,k_2;g)}, 0 \leq k_1 \leq I, 0 \leq k_2 \leq J, 0 \leq g \leq F \right\}$. Let $\mathbf{G}(\mathbf{s})$ be the vector whose components are the offspring generating functions of all possible types. According to the general theory of multitype branching processes, the extinction probability is given by the root of

$$\mathbf{G}(\mathbf{s}) = \mathbf{s} \tag{A.2}$$

in the unit cube, which is closest to the origin (Sewastjanow, 1974, p. 115).

First note that Eq. (A.2) is equivalent to Eq. (3.3) (identifying the solution of Eq. (A.2) with the vector with elements $\left\{ Q_{(k_1,k_2;g)}, 0 \leq k_1 \leq I, 0 \leq k_2 \leq J, 0 \leq g \leq F \right\}$).

Eq. (A.2) can be solved recursively starting with type $(0, 0; 0)$, for which we obtain

$$Q_{(0,0;0)} = 1 - \sigma_{(0,0;0)} < 1. \tag{A.3}$$

Following Sewastjanow (1974), in each subsequent step of the recursion, there exists a unique solution < 1. It remains to show that we obtain this solution

if we solve the quadratic equation for the smaller root. We use the following abbreviations:

$$\Sigma_g := \sum_{k=0}^{j-1} r_{(i,k)}^{(i,j)} Q_{(i,k;g)} + \sum_{k=0}^{i-1} r_{(k,j)}^{(i,j)} Q_{(k,j;g)}, \tag{A.4}$$

$$\Sigma := \left(\frac{1}{2}\right)^f \left\{ \sum_{g=0}^{f-1} \binom{f}{g} (1 - r_{(i,j)}) Q_{(i,j;g)} + \sum_{g=0}^{f} \binom{f}{g} \Sigma_g \right\}, \tag{A.5}$$

$$a := \left(\frac{1}{2}\right)^f (1 - r_{(i,j)}) > 0, \tag{A.6}$$

$$b := -(1 - \sigma_{(i,j;f)} + 1 - \Sigma), \tag{A.7}$$

$$c := 1 - \sigma_{(i,j;f)} \geq 0. \tag{A.8}$$

Note that $r \geq \Sigma_g, \forall g$.

With these abbreviations, the two roots are given by

$$q_1 = \frac{-b + \sqrt{(b^2 - 4ac)}}{2a}, \tag{A.9}$$

$$q_2 = \frac{-b - \sqrt{(b^2 - 4ac)}}{2a}. \tag{A.10}$$

We first prove that both roots are real and ≥ 0. As a preliminary step, we show that $(1 - \Sigma) \geq \left(\frac{1}{2}\right)^f (1 - r_{(i,j)})$:

$$\begin{aligned}
1 - \Sigma &= 1 - \left(\frac{1}{2}\right)^f \left\{ \sum_{g=0}^{f-1} \binom{f}{g} (1 - r_{(i,j)}) Q_{(i,j;g)} + \sum_{g=0}^{f} \binom{f}{g} \Sigma_g \right\} \\
&\geq 1 - \left(\frac{1}{2}\right)^f \left\{ \sum_{g=0}^{f-1} \binom{f}{g} (1 - r_{(i,j)}) + \sum_{g=0}^{f} \binom{f}{g} \Sigma_g \right\} \\
&= 1 - \left(\frac{1}{2}\right)^f \left\{ \sum_{g=0}^{f} \binom{f}{g} \left[(1 - r_{(i,j)}) + \Sigma_g \right] - (1 - r_{(i,j)}) \right\} \\
&\geq 1 - \left\{ \sum_{g=0}^{f} \binom{f}{g} \left(\frac{1}{2}\right)^f - (1 - r_{(i,j)}) \left(\frac{1}{2}\right)^f \right\} \\
&= (1 - r_{(i,j)}) \left(\frac{1}{2}\right)^f.
\end{aligned} \tag{A.11}$$

Notice: $b < 0$ since $\sigma_{(i,j;f)} \leq 1$ and $1 - \Sigma \geq (1 - r_{(i,j)}) \left(\frac{1}{2}\right)^f > 0$.

We now show that $b^2 - 4ac \geq 0$.

$$\begin{aligned} b^2 - 4ac &= (1 - \sigma_{(i,j)} + 1 - \Sigma)^2 - 4(1 - \sigma_{(i,j)})(1 - r_{(i,j)}) \left(\frac{1}{2}\right)^f \\ &\geq (1 - \sigma_{(i,j)} + 1 - \Sigma)^2 - 4(1 - \sigma_{(i,j)})(1 - \Sigma)^2 \\ &= (1 - \sigma_{(i,j)} - (1 - \Sigma))^2 \geq 0, \end{aligned} \tag{A.12}$$

where we have used that $(1 - r_{(i,j)}) \left(\frac{1}{2}\right)^f \leq 1 - \Sigma$. Since $a > 0$, we conclude that both roots are positive.

It remains to prove that $q_2 \leq 1$. Note that $\mathbf{1}$ is a root of the equation since $\mathbf{G}(\mathbf{1}) = \mathbf{1}$. We furthermore see that q_2 is a decreasing function of Σ and thus of all $Q_{(k,j)}$ and $Q_{(i,k)}$, if $c > 0$. For $c = 0$, it holds that $q_2 = 0$. It is hence clear that $q_2 \leq 1$. From Sewastjanow (1974), we even know that $q_2 < 1$ (because $Q_{(0,0;0)} < 1$).

□

Note that instead of considering the embedded discrete time process we could have directly resorted to the corresponding result for the extinction probability of the continuous time processes (Sewastjanow, 1974, p. 116), but it seemed more illustrative in this way.

B Appendix: Proof of Eq. (3.32)

We prove that within approximation (3.30), it holds

$$\sum_{i,j} \hat{P}^{(I,J)}_{(i,j)} = 1, \tag{B.1}$$

where $\hat{P}^{(\cdot,\cdot)}_{(\cdot,\cdot)}$ is defined by the approximative formula (3.30). We carry out a proof by induction. Eq. (B.1) builds our induction hypothesis.

- Base case $(I, J) = (0, 0)$: Since $\hat{P}^{(0,0)}_{(i,j)} = \delta_{i,0}\delta_{j,0}$, it holds $\sum_{i,j} \hat{P}^{(0,0)}_{(i,j)} = 1$.

- Inductive step: Let the hypothesis be true for all pairs (k,m) with $k<n$ and (n,k) with $k<m$. We show that it is true for (n,m):

$$\begin{aligned}
&\sum_{i,j}\hat{P}^{(n,m)}_{(i,j)}\\
=&\frac{\sum_{k=0}^{n-1} r^{(n,m)}_{(k,m)}(1-Q_{(k,m)})\sum_{i,j}P^{(k,m)}_{(i,j)} + \sum_{k=0}^{m-1} r^{(n,m)}_{(n,k)}(1-Q_{(n,k)})\sum_{i,j}P^{(n,k)}_{(i,j)}}{\sum_{k=0}^{n-1} r^{(n,m)}_{(k,m)}(1-Q_{(k,m)}) + \sum_{k=0}^{m-1} r^{(n,m)}_{(n,k)}(1-Q_{(n,k)})}\\
=&\frac{\sum_{k=0}^{n-1} r^{(n,m)}_{(k,m)}(1-Q_{(k,m)})\cdot 1 + \sum_{k=0}^{m-1} r^{(n,m)}_{(n,k)}(1-Q_{(n,k)})\cdot 1}{\sum_{k=0}^{n-1} r^{(n,m)}_{(k,m)}(1-Q_{(k,m)}) + \sum_{k=0}^{m-1} r^{(n,m)}_{(n,k)}(1-Q_{(n,k)})} = 1. \quad \square
\end{aligned} \tag{B.2}$$

  The proof works analogously if the number of unlinked alleles is larger than 0.

C Appendix: Unlinked alleles of arbitrary effect

For the main text, we assumed that each unlinked allele has the same selection coefficient. Here, we give the generalization to arbitrary effects.

Be $\mathcal{F} = \{\vec{e} = (e_1, e_2, \ldots, e_F), e_i \in \{0,1\}\}$. We define $|\vec{e}| = \sum_{i=1}^{F} e_i$. The set of unlinked deleterious alleles carried by an individual can be characterized by a vector in $\mathcal{F}$ where 1 and 0 at a given position denote the presence or absence

of a specific deleterious allele. In this notation, the transitions of the branching process read with $|\vec{e}_f| = f$:

$$\begin{aligned}
P((i,j;\vec{e}_f) \to 0) &= 1 - \sigma_{(i,j;\vec{e}_f)},\\
P((i,j;\vec{e}_f) \to \{(i,j;\vec{e}_f);(i,j;\vec{e}_g)\}) &= \left(\frac{1}{2}\right)^f (1 - r_{(i,j)}) \quad \text{for} \quad \vec{e}_f - \vec{e}_g \in \mathcal{F},\\
P((i,j;\vec{e}_f) \to \{(i,j;\vec{e}_f);(i,k;\vec{e}_g)\}) &= \left(\frac{1}{2}\right)^f r^{(i,j)}_{(i,k)} \quad \text{for} \quad k < j, \vec{e}_f - \vec{e}_g \in \mathcal{F},\\
P((i,j;\vec{e}_f) \to \{(i,j;\vec{e}_f);(k,j;\vec{e}_g)\}) &= \left(\frac{1}{2}\right)^f r^{(i,j)}_{(k,j)} \quad \text{for} \quad k < i, \vec{e}_f - \vec{e}_g \in \mathcal{F}.
\end{aligned} \tag{C.1}$$

The recursive equation for the extinction probability becomes:

$$\begin{aligned}
&\left(\frac{1}{2}\right)^f (1 - r_{(i,j)}) Q^2_{(i,j;\vec{e}_f)}\\
&\quad + \sum_{\vec{e}_g, \vec{e}_f - \vec{e}_g \in \mathcal{F}} \left(\frac{1}{2}\right)^f \left\{ \sum_{k=0}^{i-1} r^{(i,j)}_{(k,j)} Q_{(k,j;\vec{e}_g)} + \sum_{k=0}^{j-1} r^{(i,j)}_{(i,k)} Q_{(i,k;\vec{e}_g)} \right\} Q_{(i,j;\vec{e}_f)}\\
&\quad + \sum_{\substack{\vec{e}_g, \vec{e}_f - \vec{e}_g \in \mathcal{F},\\ \vec{e}_f - \vec{e}_g \neq \vec{0}}} \left(\frac{1}{2}\right)^f (1 - r_{(i,j)}) Q_{(i,j;\vec{e}_g)} Q_{(i,j;\vec{e}_f)} - (2 - \sigma_{(i,j;\vec{e}_f)}) Q_{(i,j,\vec{e}_f)}\\
&\quad + 1 - \sigma_{(i,j;\vec{e}_f)} = 0.
\end{aligned} \tag{C.2}$$

The proof is analogous to before.

Similarly, we can generalize approximation Eq. (3.34):

$$\begin{aligned}
P^{(I,J;\vec{e}_F)}_{(i,j;\vec{0})} &= \frac{\sum_{\vec{e}_l} \left(\frac{1}{2}\right)^F A_{\vec{e}_l} + \sum_{\vec{e}_l \neq \vec{e}_F} \left(\frac{1}{2}\right)^F (1 - r_{(I,J)})(1 - Q_{(I,J;\vec{e}_l)}) P^{(I,J;\vec{e}_l)}_{(i,j;\vec{0})}}{\sum_{\vec{e}_l} \left(\frac{1}{2}\right)^F B_{\vec{e}_l} + \sum_{\vec{e}_l \neq \vec{e}_F} \left(\frac{1}{2}\right)^F (1 - r_{(I,J)})(1 - Q_{(I,J;\vec{e}_l)})}.\\
A_{\vec{e}_l} &= \sum_{i=0}^{I-1} r^{(I,J)}_{(i,J)} (1 - Q_{(i,J;\vec{e}_l)}) P^{(i,J;\vec{e}_l)}_{(i,j;\vec{0})} + \sum_{j=0}^{J-1} r^{(I,J)}_{(I,j)} (1 - Q_{(I,j;\vec{e}_l)}) P^{(I,j;\vec{e}_l)}_{(i,j;\vec{0})}\\
B_{\vec{e}_l} &= \sum_{i=0}^{I-1} r^{(I,J)}_{(i,J)} (1 - Q_{(i,J;\vec{e}_l)}) + \sum_{j=0}^{J-1} r^{(I,J)}_{(I,j)} (1 - Q_{(I,j;\vec{e}_l)})
\end{aligned} \tag{C.3}$$

For $\left(\frac{1}{2}\right)^F (1 - r_{(I,J)}) - (1 - \sigma_{(I,J;F)}) > 0$, we approximate:

$$P^{(I,J;\vec{e}_F)}_{(i,j;\vec{0})} = \delta_{I,i}\delta_{J,j}. \tag{C.4}$$

D Appendix: General lemmata

We here summarize some general lemmata which we use for the derivation of hitchhiking probabilities. Although the lemmata are either not new or follow immediately from known results, we briefly sketch the proofs.

We consider a multitype branching process with $N + 1$ types. The number of individuals of type i ($i = 0, 1, \ldots, N$) in generation n is $Z_n^{(i)}$. Let $Z_0^{(N)} = 1$ and $Z_0^{(i)} = 0$ if $i \neq N$, i.e., the process is started by an individual of type N. Individuals of type i reproduce at rate 1 and die at rate $1 - \sigma_i$. The offspring of a type N individual be of type i with probability r_i. If a type N individual gives birth to an individual of type $i < N$, we call this a recombination event from type N to type i. We define $r := \sum_{k=0}^{N-1} r_k = 1 - r_N$.

Lemma D.1. *The probability generating function (p.g.f.) $f(s)$ of the offspring distribution of a type N individual is given by*

$$f(s) = \frac{1-p}{1-ps} \quad \text{with} \quad p = \frac{1}{2-\sigma_N}. \tag{D.1}$$

Proof. Individuals have a geometric offspring distribution; for the random offspring number X of a type N individual, it holds:

$$P(X = k) = \left(\frac{1}{2-\sigma_N}\right)^k \frac{1-\sigma_N}{2-\sigma_N}, \tag{D.2}$$

and therefore

$$\begin{aligned} f(s) &= \sum_{k=0}^{\infty} P(X=k)s^k = \sum_{k=0}^{\infty} \left(\frac{1}{2-\sigma_N}\right)^k \frac{1-\sigma_N}{2-\sigma_N} s^k \\ &= \frac{1-\sigma_N}{2-\sigma_N} \frac{1}{1-\frac{s}{2-\sigma_N}} = \frac{1-p}{1-ps}. \end{aligned} \tag{D.3}$$

□

Lemma D.2. *The joint p.g.f of* $(Z_1^{(0)}, Z_1^{(1)}, \ldots, Z_1^{(N)})$ *is given by*

$$F(s_0, s_1, \ldots, s_N) = f(r_0 s_0 + r_1 s_1 + \ldots r_N s_N). \tag{D.4}$$

Proof. We restrict to $N = 1$; the generalization to larger N is straightforward. Be $p(k)$ the probability that an individual of type 1 has k offspring and $p(k_0, k_1)$ the probability that it has k_0 offspring of type 0 and k_1 offspring of type 1. It holds (cf. Serra, 2006):

$$\begin{aligned} F(s_0, s_1) &= \sum_{k_0,k_1} p(k_0, k_1) s_0^{k_0} s_1^{k_1} = \sum_{k=0}^{\infty} \sum_{k_0=0}^{k} \underbrace{p(k_0, k-k_0)}_{=p(k)\left\{r_0^{k_0}(1-r_0)^{k-k_0}\binom{k}{k_0}\right\}} s_0^{k_0} s_1^{k-k_0} \\ &= \sum_{k=0}^{\infty} p(k) \underbrace{\sum_{k_0=0}^{k} \binom{k}{k_0} (r_0 s_0)^{k_0} (s_1(1-r_0))^{k-k_0}}_{=(r_0 s_0 + (1-r_0)s_1)^k} \\ &= \sum_{k=0}^{\infty} p(k)(r_0 s_0 + (1-r_0)s_1)^k = f(r_0 s_0 + (1-r_0)s_1). \end{aligned}$$

□

Lemma D.3. *The extinction probability* q_N *of type* N *is given by*

$$q_N = \begin{cases} \frac{1-\sigma_N}{1-r} & \text{if} \quad \sigma_N > r, \\ 1 & \text{if} \quad \sigma_N \leq r. \end{cases} \tag{D.5}$$

Proof. The p.g.f. for the number of offspring of its own type is given by $f(r + (1-r)s)$. From the general theory of single type branching processes, it follows that q_N is the smallest root of

$$q_N = f(r + (1-r)q_N). \tag{D.6}$$

in $[0, 1]$. □

Lemma D.4. *Offspring of type* i *be of type* $j \leq i$ *only. Consider the process on the set* $B = \left\{\omega : Z_n^{(N)} \to 0 \quad \text{as} \quad n \to \infty\right\}$. *The probability of the event* B *is denoted by* q_N. *The joint p.g.f. of* $(Z_1^{(0)}, Z_1^{(1)}, \ldots, Z_1^{(N)})$ *in the conditioned process is given by*

$$\hat{F}(s_0, s_1, \ldots, s_N) = \frac{F(s_0, s_1, \ldots, q_N s_N)}{q_N}. \tag{D.7}$$

Proof. For simplicity, we stick to $N = 1$. The generalization to $N > 1$ is straightforward. It holds (cf. Athreya and Ney, 1972, p. 52f):

$$\begin{aligned}
\hat{F}(s_0, s_1) &= \sum_{k_0=0}^{\infty} \sum_{k_1=0}^{\infty} P(Z_1^{(0)} = k_0, Z_1^{(1)} = k_1 | B) s_0^{k_0} s_1^{k_1} \\
&= \frac{\sum_{k_0=0}^{\infty} \sum_{k_1=0}^{\infty} P(Z_1^{(0)} = k_0, Z_1^{(1)} = k_1, B) s_0^{k_0} s_1^{k_1}}{q_1} \\
&= \frac{\sum_{k_0=0}^{\infty} \sum_{k_1=0}^{\infty} P(Z_1^{(0)} = k_0, Z_1^{(1)} = k_1) q_1^{k_1} s_0^{k_0} s_1^{k_1}}{q_1} \\
&= \frac{F(s_0, q_1 s_1)}{q_1}.
\end{aligned}$$

□

Lemma D.5. *Offspring of type i be of type $j \leq i$ only. Condition the process on extinction of type N. The total number of recombination events from type N individuals to other types in the whole process is determined by the p.g.f. $\hat{h}(s)$ with*

$$\hat{h}(s) = \frac{1 - psr - \sqrt{(1 - psr)^2 - 4(1 - p)p(1 - r)}}{2pq_N(1 - r)}. \tag{D.8}$$

Proof. We first proof the relation

$$\hat{h}(s) = \hat{F}(s, \hat{h}(s)). \tag{D.9}$$

(cf. Serra, 2006). Be $\tilde{p}(k)$ the probability that the total number of recombination events is k. Be $p(k_r, k_N)$ the probability that an individual of type N has k_N offspring of type N and k_r offspring of other types. First, note that:

$$\tilde{p}(k) = \sum_{k_N=0}^{\infty} \sum_{j=0}^{k} p(k - j, k_N) \tilde{p}(j | k_N). \tag{D.10}$$

Furthermore:

$$\sum_{k=0}^{\infty} \sum_{k_N}^{\infty} \sum_{j=0}^{k} p(k - j, k_N) s^k = \sum_{k_r, k_N} p(k_r, k_N) \sum_{j} \tilde{p}(j | k_N) s^{k_r + j}. \tag{D.11}$$

Using this:

$$\begin{aligned}\hat{h}(s) &= \sum_k \tilde{p}(k)s^k = \sum_{k_r,k_N} p(k_r,k_N)\sum_{j=0}^{\infty}\tilde{p}(j|k_N)s^{j+k_r} \\ &= \sum_{k_r,k_N} p(k_r,K_N)s_r^k \underbrace{\sum_{j=0}^{\infty}\tilde{p}(j|k_N)s^j}_{=\left(\hat{h}(s)\right)^{k_N}}. \quad \square\end{aligned} \tag{D.12}$$

It thus holds:

$$\begin{aligned}\hat{h}(s) = \hat{F}(s,\hat{h}(s)) &= \frac{F(s,q_N\hat{h}(s))}{q_N} = \frac{f(sr+q_N(1-r)\hat{h}(s))}{q_N} \\ &= \frac{1}{q_N}\frac{1-p}{1-p(sr+q_N(1-r)\hat{h}(s))}.\end{aligned} \tag{D.13}$$

This leads to a quadratic equation for $\hat{h}(s)$ which has two solutions. As it must hold that $\hat{h}(1) = 1$, we can exclude one of them and obtain Eq. (D.8).

Lemma D.6. *With the assumptions of the previous lemma, the joint p.g.f. of the number of recombination events from type N to type $0, 1, 2, \ldots, N-1$ is given by*

$$\tilde{h}(s_0, s_1, \ldots, s_{N-1}) = \hat{h}(\frac{r_0}{r}s_0 + \cdots + \frac{r_{N-1}}{r}s_{N-1}). \tag{D.14}$$

Proof. analogous to the proof of Lemma D.4. $\square$

E Appendix: The deterministic phase

E.1 The process is initiated by an individual of type $(0, 2)$

We now focus on three types $(0, 0)$, $(0, 1)$, and $(0, 2)$, which all have a positive selection coefficient, while all other types are assumed to be deleterious (all other types refers to all types that are taken into account in the theory). An individual of type $(0, 2)$ starts sweeping. Throughout, we assume that recombination is weak enough that it does not significantly influence how many offspring of its own type an individual has.

The deterministic frequency path is given by

$$x_{(0,2)} = \frac{\bar{\nu}_0 \exp\left(\sigma_{(0,2)} t\right)}{N - \bar{\nu}_0 + \bar{\nu}_0 \exp\left(\sigma_{(0,2)} t\right)}. \tag{E.1}$$

Recombination events occur at rate

$$\alpha_{(0,2)}(t) = \alpha_{(0,0)}(t) + \alpha_{(0,1)}(t) \tag{E.2}$$

with

$$\alpha_{(0,0)}(t) = N x_{(0,2)}(t)(1 - x_{(0,2)}(t)) r^{(0,2)}_{(0,0)} p^{(0,0)}_{\text{est}}(x_{(0,2)}(t)), \tag{E.3a}$$

$$\alpha_{(0,1)}(t) = N x_{(0,2)}(t)(1 - x_{(0,2)}(t)) r^{(0,2)}_{(0,1)} p^{(0,1)}_{\text{est}}(x_{(0,2)}(t)), \tag{E.3b}$$

where $p^{(0,0)}_{\text{est}}$ and $p^{(0,1)}_{\text{est}}$ are the establishment probabilities of an individual of type $(0,0)$ and $(0,1)$ respectively and calculated by suitable substitutions in Eq. (3.38). We assume in the following that we can set $\bar{\nu}_0/N \approx 0$ in the integration boundaries. Analogous to Eq. (3.41), the probability that no recombination event takes place is then given by

$$\begin{aligned} P^{((0,2)\to(0,2))}_{\text{det}} &= \exp\left[-\int_0^\infty \alpha_{(0,2)}(t) \mathrm{d}t \right] \\ &= \left(\frac{\sigma_{(0,0)}}{\sigma_{(0,0)} - \sigma_{(0,2)}} \right)^{-\frac{N r^{(0,2)}_{(0,0)} \sigma_{(0,0)} (\sigma_{(0,0)} - \sigma_{(0,2)})}{\sigma^2_{(0,2)}}} \times \left(\frac{\sigma_{(0,1)}}{\sigma_{(0,1)} - \sigma_{(0,2)}} \right)^{-\frac{N r^{(0,2)}_{(0,1)} \sigma_{(0,1)} (\sigma_{(0,1)} - \sigma_{(0,2)})}{\sigma^2_{(0,2)}}}. \end{aligned} \tag{E.4}$$

The probability that no successful type $(0,0)$ is generated up to t_R is given by

$$\begin{aligned}
&\exp\left[-\int_0^{t_R}\alpha_{(0,0)}(t)\mathrm{d}t\right] \\
&\approx \exp\left[-\int_0^{x_{(0,2)}(t_R)} Nr^{(0,2)}_{(0,0)}\frac{1}{\sigma_{(0,2)}}\frac{\sigma_{(0,0)}(\sigma_{(0,0)}-\sigma_{(0,2)})}{(\sigma_{(0,0)}-\sigma_{(0,2)})(1-x)+\sigma_{(0,0)}x}\mathrm{d}x\right] \\
&= \left(\frac{\sigma_{(0,0)}-(1-x_{(0,2)}(t_R))\sigma_{(0,2)}}{\sigma_{(0,0)}-\sigma_{(0,2)}}\right)^{-\frac{Nr^{(0,2)}_{(0,0)}\sigma_{(0,0)}(\sigma_{(0,0)}-\sigma_{(0,2)})}{\sigma_{(0,2)}\sigma_{(0,2)}}} \\
&\equiv Q^{(0,0)}_{\mathrm{det}}(x_{(0,2)}(t_R)).
\end{aligned} \tag{E.5}$$

The probability that no successful type $(0,1)$ is generated up to time t_R is given by

$$\begin{aligned}
&\exp\left[-\int_0^{t_R}\alpha_{(0,1)}(t)\mathrm{d}t\right] \\
&\approx \exp\left[-\int_0^{x_{(0,2)}(t_R)} Nr^{(0,2)}_{(0,1)}\frac{1}{\sigma_{(0,2)}}\frac{\sigma_{(0,1)}(\sigma_{(0,1)}-\sigma_{(0,2)})}{(\sigma_{(0,1)}-\sigma_{(0,2)})(1-x)+\sigma_{(0,1)}x}\mathrm{d}x\right] \\
&= \left(\frac{\sigma_{(0,1)}-(1-x_{(0,2)}(t_R))\sigma_{(0,2)}}{\sigma_{(0,1)}-\sigma_{(0,2)}}\right)^{-\frac{Nr^{(0,2)}_{(0,0)}\sigma_{(0,1)}(\sigma_{(0,1)}-\sigma_{(0,2)})}{\sigma_{(0,2)}\sigma_{(0,2)}}} \\
&\equiv Q^{(0,1)}_{\mathrm{det}}(x_{(0,2)}(t_R)).
\end{aligned} \tag{E.6}$$

The probability that as a first event, an individual of type $(0,0)$ is generated is given by

$$\begin{aligned}
&\int_0^\infty Q_{\text{det}}^{(0,0)}(x_{(0,2)}(t_R))Q_{\text{det}}^{(0,1)}(x_{(0,2)}(t_R))\alpha_1(t_R)\mathrm{d}t_R \\
=&\int_0^\infty Q_{\text{det}}^{(0,0)}(x_{(0,2)}(t_R))Q_{(0,1)}(x_{(0,2)}(t_R))\alpha_1(t_R)\mathrm{d}t_R \\
=&\int_0^\infty Q_{\text{det}}^{(0,0)}(x_{(0,2)}(t_R))Q_{\text{det}}^{(0,1)}(x_{(0,2)}(t_R)) \\
&\qquad\qquad \times Nx_{(0,2)}(t)(1-x_{(0,2)}(t))r_{(0,0)}^{(0,2)}p_{\text{est}}^{(0,0)}(x_{(0,2)}(t))\mathrm{d}t_R \\
=&\frac{1}{\sigma_{(0,2)}}\int_0^1 Q_{\text{det}}^{(0,0)}(x)Q_{\text{det}}^{(0,1)}(x)Nr_{(0,0)}^{(0,2)}p_{\text{est}}^{(0,0)}(x)\mathrm{d}x.
\end{aligned} \tag{E.7}$$

Equivalently the probability that as a first event, an individual of type $(0,1)$ is generated is given by

$$\frac{1}{\sigma_{(0,2)}}\int_0^1 Q_{\text{det}}^{(0,0)}(x)Q_{\text{det}}^{(0,1)}(x)Nr_{(0,1)}^{(0,2)}p_{\text{est}}^{(0,1)}(x)\mathrm{d}x. \tag{E.8}$$

In that case, the type $(0,1)$ might either rise to fixation, or a successful type $(0,0)$ individual might be generated in the following process. We now calculate the probability of these two events given a type $(0,1)$ individual has been generated at time t_R. In order to calculate the probability for the generation of a sucessful type $(0,0)$ individual, we need the frequencies of the three present types (wildtype, type $(0,2)$, type $(0,1)$) at time t after the recombination event. We have to take into account that in its early phase of growth, the type $(0,1)$ increases faster than predictic by the deterministic path. As before, we therefore replace its initial frequency by the mean of the effective initial population size ν. We furthermore assume that type $(0,1)$ individuals replace in this phase exclusively wildtype

individuals. It proves that this assumption has no visible influence on the results. We obtain:

$$\tilde{q}(t|x_{(0,2)}(t_R)) = \frac{1 - x_{(0,2)}(t_R) - \frac{\bar{\nu}(t_R)}{N}}{x_{(0,2)}(t_R)\exp(\sigma_{(0,2)}t) + 1 - x_{(0,2)}(t_R) - \frac{\bar{\nu}(t_R)}{N} + \frac{\bar{\nu}(t_R)}{N}\exp(\sigma_{(0,1)}t)}, \tag{E.9}$$

$$\tilde{x}_{01}(t|x_{(0,2)}(t_R)) = \frac{\frac{\bar{\nu}(t_R)}{N}\exp(\sigma_{(0,1)}t)}{x_{(0,2)}(t_R)\exp(\sigma_{(0,2)}t) + 1 - x_{(0,2)}(t_R) - \frac{\bar{\nu}(t_R)}{N} + \frac{\bar{\nu}(t_R)}{N}\exp(\sigma_{(0,1)}t)}, \tag{E.10}$$

$$\tilde{x}_{02}(t|x_{(0,2)}(t_R)) = \frac{x_{(0,2)}(t_R)\exp(\sigma_{(0,2)}t)}{x_{(0,2)}(t_R)\exp(\sigma_{(0,2)}t) + 1 - x_{(0,2)}(t_R) - \frac{\bar{\nu}(t_R)}{N} + \frac{\bar{\nu}(t_R)}{N}\exp(\sigma_{(0,1)}t)}. \tag{E.11}$$

The time-dependent selection coefficient of type $(0,0)$ is given by

$$\tilde{s}_{(0,0)}(t|x_{(0,2)}(t_R)) = \sigma_{(0,0)} - \sigma_{(0,1)}\tilde{x}_{01}(t|x_{(0,2)}(t_R)) - \sigma_{(0,2)}\tilde{x}_{02}(t|x_{(0,2)}(t_R)). \tag{E.12}$$

The establishment probability of a single individual of type $(0,0)$ that arises at time t after t_R can be calculated similarly to before and is given by

$$\begin{aligned} \tilde{p}_{\text{est}}^{(0,0)}(t|x_{(0,2)}(t_R)) &= \frac{1}{A}\sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(0,1)})(\sigma_{(0,0)} - \sigma_{(0,2)}), \\ A &= (\sigma_{(0,0)} - \sigma_{(0,1)})(\sigma_{(0,0)} - \sigma_{(0,2)})\tilde{q}(t|x_{(0,2)}) \\ &\quad + \sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(0,2)})\tilde{x}_{01}(t|x_{(0,2)}(t_R)) \\ &\quad + \sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(0,1)})\tilde{x}_{02}(t|x_{(0,2)}(t_R)). \end{aligned} \tag{E.13}$$

The rate of successful type $(0,0)$ individuals is

$$\tilde{\alpha}_{(0,0)}(t|x_{(0,2)}(t_R)) = N r_{(0,0)}^{(0,1)} \tilde{q}(t|x_{(0,2)}(t_R))(1 - \tilde{q}(t|x_{(0,2)}(t_R)))\tilde{p}_{\text{est}}^{(0,0)}(t|x_{(0,2)}(t_R)). \tag{E.14}$$

(Remember: $r^{(0,1)}_{(0,0)} = r^{(0,2)}_{(0,0)}$.) We assume that we can ignore deviations from the deterministic frequency path for large t close to fixation of the beneficical allele. The probability that no successful recombination event takes place is then given by

$$\text{Prob}(\text{no rec}|x_{(0,2)}(t_R)) = \exp\left(-\int_0^\infty \tilde{\alpha}_{(0,0)}(t|x_{(0,2)}(t_R))\mathrm{d}t\right). \tag{E.15}$$

The probability that type $(0,1)$ fixes is hence

$$\begin{aligned} P^{((0,2)\to(0,1))}_{\text{det}} &= \int_0^\infty Q^{(0,0)}_{\text{det}}(x_{(0,2)}(t_R))Q^{(0,1)}_{\text{det}}(x_{(0,2)}(t_R))\alpha_2(t_R)\text{Prob}(\text{no rec}|x_{(0,2)}(t_R))\mathrm{d}t_R. \\ &= \frac{1}{\sigma_{(0,2)}}\int_0^1 Q^{(0,0)}_{\text{det}}(x)Q^{(0,1)}_{\text{det}}(x)Nr^{(0,1)}_{(0,0)}p^{(0,1)}_{\text{est}}(x)\text{Prob}(\text{no rec}|x)\mathrm{d}x. \end{aligned} \tag{E.16}$$

The overall probability that type $(0,0)$ fixes is obtained as

$$\begin{aligned} P^{((0,2)\to(0,0))}_{\text{det}} &= \frac{1}{\sigma_{(0,2)}}\int_0^1 Q^{(0,0)}_{\text{det}}(x)Q^{(0,1)}_{\text{det}}(x)Nr^{(0,2)}_{(0,0)}p^{(0,0)}_{\text{est}}(x)\mathrm{d}x \\ &\quad + \frac{1}{\sigma_{(0,2)}}\int_0^1 Q^{(0,0)}_{\text{det}}(x)Q^{(0,1)}_{\text{det}}(x)Nr^{(0,1)}_{(0,0)}p^{(0,1)}_{\text{est}}(x)(1-\text{Prob}(\text{no rec}|x))\mathrm{d}x. \end{aligned} \tag{E.17}$$

E.2 The process is initiated by an individual of type $(1,1)$

We now focus on four types $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$, which all have a positive selection coefficient, while all other types are assumed to be deleterious (all other types refers again to all types that are taken into account in the theory). An individual of type $(1,1)$ starts sweeping. We derive an approximation which is valid if no more than two successful recombination events happen during the process, i.e., for weak recombination. We again assume that recombination does not significantly reduce the number of offspring of any individual's own type.

The deterministic path of type $(1,1)$ is given by

$$x_{(1,1)} = \frac{\bar{\nu}_0 \exp\left(\sigma_{(1,1)} t\right)}{N - \bar{\nu}_0 + \bar{\nu}_0 \exp\left(\sigma_{(1,1)} t\right)}. \tag{E.18}$$

Recombination events occur at rate

$$\alpha_{(1,1)}(t) = \alpha_{(1,0)}(t) + \alpha_{(0,1)}(t) \tag{E.19}$$

with

$$\alpha_{(1,0)}(t) = N x_{(1,1)}(t)(1 - x_{(1,1)}(t)) r^{(1,1)}_{(1,0)} p^{(1,0)}_{\text{est}}(x_{(1,1)}(t)), \tag{E.20a}$$

$$\alpha_{(0,1)}(t) = N x_{(1,1)}(t)(1 - x_{(1,1)}(t)) r^{(1,1)}_{(0,1)} p^{(0,1)}_{\text{est}}(x_{(1,1)}(t)), \tag{E.20b}$$

where $p^{(1,0)}_{\text{est}}$ and $p^{(0,1)}_{\text{est}}$ are the fixation probabilities of an individual of type $(1,0)$ and $(0,1)$ respectively and calculated by suitable substitutions in Eq. 3.38. Analogous to Eq. (3.41), the probability that no recombination event takes place is given by

$$\begin{aligned} P^{((1,1)\to(1,1))}_{\text{det}} &= \exp\left[- \int_0^\infty \alpha_{(1,1)}(t) \mathrm{d}t \right] \\ &= \left(\frac{\sigma_{(1,0)}}{\sigma_{(1,0)} - \sigma_{(1,1)}} \right)^{-\frac{N r^{(1,1)}_{(1,0)} \sigma_{(1,0)} (\sigma_{(1,0)} - \sigma_{(1,1)})}{\sigma^2_{(1,1)}}} \times \left(\frac{\sigma_{(0,1)}}{\sigma_{(0,1)} - \sigma_{(1,1)}} \right)^{-\frac{N r^{(1,1)}_{(0,1)} \sigma_{(0,1)} (\sigma_{(0,1)} - \sigma_{(1,1)})}{\sigma^2_{(1,1)}}}. \end{aligned} \tag{E.21}$$

The probability that no successful type $(1,0)$ is generated up to t_R is given by

$$\begin{aligned}
&\exp\left[-\int_0^{t_R}\alpha_{(1,0)}(t)\mathrm{d}t\right] \\
&\approx \exp\left[-\int_0^{x_{(1,1)}(t_R)} Nr^{(1,1)}_{(1,0)}\frac{1}{\sigma_{(1,1)}}\frac{\sigma_{(1,0)}(\sigma_{(1,0)}-\sigma_{(1,1)})}{(\sigma_{(1,0)}-\sigma_{(1,1)})(1-x)+\sigma_{(1,0)}x}\mathrm{d}x\right] \\
&= \left(\frac{\sigma_{(1,0)}-(1-x_{(1,1)}(t_R))\sigma_{(1,1)}}{\sigma_{(1,0)}-\sigma_{(1,1)}}\right)^{-\frac{Nr^{(1,1)}_{(1,0)}\sigma_{(1,0)}(\sigma_{(1,0)}-\sigma_{(1,1)})}{\sigma_{(1,1)}\sigma_{(1,1)}}} \\
&\equiv Q^{(1,0)}_{\mathrm{det}}(x_{(1,1)}(t_R)).
\end{aligned} \tag{E.22}$$

The probability that no successful type $(0,1)$ is generated up to t_R is given by

$$\begin{aligned}
&\exp\left[-\int_0^{t_R}\alpha_{(0,1)}(t)\mathrm{d}t\right] \\
&\approx \exp\left[-\int_0^{x_{(1,1)}(t_R)} Nr^{(1,1)}_{(0,1)}\frac{1}{\sigma_{(1,1)}}\frac{\sigma_{(0,1)}(\sigma_{(0,1)}-\sigma_{(1,1)})}{(\sigma_{(0,1)}-\sigma_{(1,1)})(1-x)+\sigma_{(0,1)}x}\mathrm{d}x\right] \\
&= \left(\frac{\sigma_{(0,1)}-(1-x_{(1,1)}(t_R))\sigma_{(1,1)}}{\sigma_{(0,1)}-\sigma_{(1,1)}}\right)^{-\frac{Nr^{(1,1)}_{(0,1)}\sigma_{(0,1)}(\sigma_{(0,1)}-\sigma_{(1,1)})}{\sigma_{(1,1)}\sigma_{(1,1)}}} \\
&\equiv Q^{(0,1)}_{\mathrm{det}}(x_{(1,1)}(t_R)).
\end{aligned} \tag{E.23}$$

The probability that as a first event, a type $(1,0)$ individual is generated is given by

$$\begin{aligned}
&\int_0^\infty Q_{\text{det}}^{(1,0)}(x_{(1,1)}(t_R))Q_{\text{det}}^{(0,1)}(x_{(1,1)}(t_R))\alpha_{(1,0)}(t_R)\mathrm{d}t_R \\
=&\int_0^\infty Q_{\text{det}}^{(1,0)}(x_{(1,1)}(t_R))Q_{\text{det}}^{(0,1)}(x_{(1,1)}(t_R))\alpha_{(1,0)}(t_R)\mathrm{d}t_R \\
=&\int_0^\infty Q_{\text{det}}^{(1,0)}(x_{(1,1)}(t_R))Q_{\text{det}}^{(0,1)}(x_{(1,1)}(t_R)) \\
&\qquad\qquad \times Nx_{(1,1)}(t)(1-x_{(1,1)}(t))r_{(1,0)}^{(1,1)}p_{\text{est}}^{(1,0)}(x_{(1,1)}(t))\mathrm{d}t_R \\
=&\frac{1}{\sigma_{(1,1)}}\int_0^1 Q_{\text{det}}^{(1,0)}(x)Q_{\text{det}}^{(0,1)}(x)Nr_{(1,0)}^{(1,1)}p_{\text{est}}^{(1,0)}(x)\mathrm{d}x.
\end{aligned} \tag{E.24}$$

Equivalently he probability that as a first event, a type $(0,1)$ individual is generated is given by

$$\frac{1}{\sigma_{(1,1)}}\int_0^1 Q_{\text{det}}^{(1,0)}(x)Q_{\text{det}}^{(0,1)}(x)Nr_{(0,1)}^{(1,1)}p_{\text{est}}^{(0,1)}(x)\mathrm{d}x. \tag{E.25}$$

We now choose without loss of generality $\sigma_{(1,0)} \leq \sigma_{(0,1)}$.

We first consider the case that as a first event, an individual of type $(0,1)$ is generated. In that case, the type $(0,1)$ might either rise to fixation, or a successful individual of type $(0,0)$ might be generated in the following process. We ignore the possibility that an individual of type $(1,0)$ might be generated and temporally sweep until it goes extinct. We now calculate the probability of the two events given an individual of type $(0,1)$ has been generated at time t_R. A recombination event may generate a successful type $(0,1)$ individual at time t_R. The type $(0,1)$ individual will rise to fixation unless a successful type $(0,0)$ individual is generated. In order to calculate this probability, we need the

frequencies of the three present types (wildtype, type $(0,1)$, $(1,1)$) at time t after the recombination event:

$$\tilde{q}(t|x_{(1,1)}(t_R)) = \frac{1 - x_{(1,1)}(t_R) - \frac{\bar{\nu}(t_R)}{N}}{x_{(1,1)}(t_R)\exp(\sigma_{(1,1)}t) + 1 - x_{(1,1)}(t_R) - \frac{\bar{\nu}(t_R)}{N} + \frac{\bar{\nu}(t_R)}{N}\exp(\sigma_{(0,1)}t)}, \tag{E.26a}$$

$$\tilde{x}_{01}(t|x_{(1,1)}(t_R)) = \frac{\frac{\bar{\nu}(t_R)}{N}\exp(\sigma_{(0,1)}t)}{x_{(1,1)}(t_R)\exp(\sigma_{(1,1)}t) + 1 - x_{(1,1)}(t_R) - \frac{\bar{\nu}(t_R)}{N} + \frac{\bar{\nu}(t_R)}{N}\exp(\sigma_{(0,1)}t)}, \tag{E.26b}$$

$$\tilde{x}_{11}(t|x_{(1,1)}(t_R)) = \frac{x_{(1,1)}(t_R)\exp(\sigma_{(1,1)}t)}{x_{(1,1)}(t_R)\exp(\sigma_{(1,1)}t) + 1 - x_{(1,1)}(t_R) - \frac{\bar{\nu}(t_R)}{N} + \frac{\bar{\nu}(t_R)}{N}\exp(\sigma_{(0,1)}t)}. \tag{E.26c}$$

The time-dependent selection coefficient of type $(0,0)$ is given by

$$\tilde{s}_{(0,0)}(t|x_{(1,1)}(t_R)) = \sigma_{(0,0)} - \sigma_{(0,1)}\tilde{x}_{01}(t|x_{(1,1)}(t_R)) - \sigma_{(1,1)}\tilde{x}_{11}(t|x_{(1,1)}(t_R)). \tag{E.27}$$

The establishment probability of a type $(0,1)$ individual that arises at time t after t_R can be calculated as before and is given by

$$\begin{aligned} \tilde{p}^{(0,0)}_{\text{est}}(t|x_{(1,1)}(t_R)) &= \frac{1}{A}\sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(0,1)})(\sigma_{(0,0)} - \sigma_{(1,1)}), \\ A &= (\sigma_{(0,0)} - \sigma_{(0,1)})(\sigma_{(0,0)} - \sigma_{(1,1)})\tilde{q}(t|x_{(1,1)}(t_R)) \\ &\quad + \sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(1,1)})\tilde{x}_{01}(t|x_{(1,1)}(t_R)) \\ &\quad + \sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(0,1)})\tilde{x}_{11}(t|x_{(1,1)}(t_R)). \end{aligned} \tag{E.28}$$

The rate of successful type $(0,0)$ individuals is

$$\tilde{\alpha}_{(0,0)}(t|x_{(1,1)}(t_R)) = Nr^{(0,1)}_{(0,0)}\tilde{q}(t|x_{(1,1)}(t_R))\tilde{x}_{01}(t|x_{(1,1)}(t_R))\tilde{p}^{(0,0)}_{\text{est}}(t|x_{(1,1)}(t_R)). \tag{E.29}$$

The probability that no successful recombination event takes place is therefore given by

$$\text{Prob}(\text{no rec}|x_{(1,1)}(t_R)) = \exp\left[-\int_0^\infty \tilde{\alpha}_{(0,0)}(t|x_{(1,1)}(t_R))\mathrm{d}t\right]. \tag{E.30}$$

The probability that type $(0,1)$ fixes is therefore given by:

$$\begin{aligned} P_{\text{det}}^{((1,1)\to(0,1))} &= \int_0^\infty Q_{\text{det}}^{(1,0)}(x_{(1,1)}(t_R))Q_{\text{det}}^{(0,1)}(x_{(1,1)}(t_R))\alpha_{(0,1)}(t_R)\text{Prob}(\text{no rec}|x_{(1,1)}(t_R))\mathrm{d}t_R. \\ &= \frac{1}{\sigma_{(1,1)}}\int_0^1 Q_{\text{det}}^{(1,0)}(x)Q_{\text{det}}^{(0,1)}(x)Nr_{(0,0)}^{(0,1)}p_{\text{est}}^{(0,1)}(x)\text{Prob}(\text{no rec}|x)\mathrm{d}x. \end{aligned} \tag{E.31}$$

We now turn to the case that an individual of type $(1,0)$ is generated first. A recombination event may generate a successful type $(1,0)$ individual at time t_R. The type $(1,0)$ individual will rise to fixation unless a successful type $(0,0)$ or type $(0,1)$ individual is generated. In order to calculate these probabilities, we again need the frequencies of the three present types (wildtype, type $(1,0)$, $(1,1)$) at time t after the recombination event:

$$\tilde{Q}(t|x_{(1,1)}(t_R)) = \frac{1 - x_{(1,1)}(t_R) - \frac{\bar{\mu}(t_R)}{N}}{x_{(1,1)}(t_R)\exp(\sigma_{(1,1)}t) + 1 - x_{(1,1)}(t_R) - \frac{\bar{\mu}(t_R)}{N} + \frac{\bar{\mu}(t_R)}{N}\exp(\sigma_{(1,0)}t)}, \tag{E.32a}$$

$$\tilde{X}_{10}(t|x_{(1,1)}(t_R)) = \frac{\frac{\bar{\mu}(t_R)}{N}\exp(\sigma_{(1,0)}t)}{x_{(1,1)}(t_R)\exp(\sigma_{(1,1)}t) + 1 - x_{(1,1)}(t_R) - \frac{\bar{\mu}(t_R)}{N} + \frac{\bar{\mu}(t_R)}{N}\exp(\sigma_{(1,0)}t)}, \tag{E.32b}$$

$$\tilde{X}_{11}(t|x_{(1,1)}(t_R)) = \frac{x_{(1,1)}(t_R)\exp(\sigma_{(1,1)}t)}{x_{(1,1)}(t_R)\exp(\sigma_{(1,1)}t) + 1 - x_{(1,1)}(t_R) - \frac{\bar{\mu}(t_R)}{N} + \frac{\bar{\mu}(t_R)}{N}\exp(\sigma_{(1,0)}t)}. \tag{E.32c}$$

The time-dependent selection coefficient of a type $(0,0)$ individual is given by

$$\tilde{S}_{(0,0)}(t|x_{(1,1)}(t_R)) = \sigma_{(0,0)} - \sigma_{(1,0)}\tilde{X}_{01}(t|x_{(1,1)}(t_R)) - \sigma_{(1,1)}\tilde{X}_{11}(t|x_{(1,1)}(t_R)). \quad \text{(E.33)}$$

The time-dependent selection coefficient of a type $(0,1)$ individual is given by

$$\tilde{S}_{(0,1)}(t|x_{(1,1)}(t_R)) = \sigma_{(0,1)} - \sigma_{(1,0)}\tilde{X}_{10}(t|x_{(1,1)}(t_R)) - \sigma_{(1,1)}\tilde{X}_{11}(t|x_{(1,1)}(t_R)). \quad \text{(E.34)}$$

The establishment probability of a type $(0,0)$ individual that arises at time t after t_R can be calculated as before and is given by

$$\begin{aligned} \tilde{\pi}^{(0,0)}_{\text{est}}(t|x_{(1,1)}(t_R)) &= \frac{1}{A}\sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(1,0)})(\sigma_{(0,0)} - \sigma_{(1,1)}), \\ A &= (\sigma_{(0,0)} - \sigma_{(1,0)})(\sigma_{(0,0)} - \sigma_{(1,1)})\tilde{Q}(t|x_{(1,1)}(t_R)) \\ &\quad + \sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(1,1)})\tilde{X}_{10}(t|x_{(1,1)}(t_R)) \\ &\quad + \sigma_{(0,0)}(\sigma_{(0,0)} - \sigma_{(1,0)})\tilde{X}_{11}(t|x_{(1,1)}(t_R)). \end{aligned} \quad \text{(E.35)}$$

and accordingly for the establishment probability of a type $(0,1)$ individual

$$\begin{aligned} \tilde{\pi}^{(0,1)}_{\text{est}}(t|x_{(1,1)}(t_R)) &= \frac{1}{B}\sigma_{(0,1)}(\sigma_{(0,1)} - \sigma_{(1,0)})(\sigma_{(0,1)} - \sigma_{(1,1)}), \\ B &= (\sigma_{(0,1)} - \sigma_{(1,0)})(\sigma_{(0,1)} - \sigma_{(1,1)})\tilde{Q}(t|x_{(1,1)}(t_R)) \\ &\quad + \sigma_{(0,1)}(\sigma_{(0,1)} - \sigma_{(1,1)})\tilde{X}_{10}(t|x_{(1,1)}(t_R)) \\ &\quad + \sigma_{(0,0)}(\sigma_{(0,1)} - \sigma_{(1,0)})\tilde{X}_{11}(t|x_{(1,1)}(t_R)). \end{aligned} \quad \text{(E.36)}$$

The rate of successful type $(0,0)$ individuals is

$$\tilde{\beta}_{(0,0)}(t|x_{(1,1)}(t_R)) = Nr^{(1,0)}_{(0,0)}\tilde{Q}(t|x_{(1,1)}(t_R))\tilde{X}_{10}(t|x_{(1,1)}(t_R))\tilde{\pi}^{(0,0)}_{\text{est}}(t|x_{(1,1)}(t_R)). \quad \text{(E.37)}$$

The rate of successful type $(0,1)$ individuals is

$$\tilde{\beta}_{(0,1)}(t|x_{(1,1)}(t_R)) = Nr^{(1,1)}_{(0,1)}\tilde{Q}(t|x_{(1,1)}(t_R))\tilde{X}_{11}(t|x_{(1,1)}(t_R))\tilde{\pi}^{(0,1)}_{\text{est}}(t|x_{(1,1)}(t_R)). \quad \text{(E.38)}$$

The probability that no successful recombination event takes place is therefore given by

$$\mathrm{Prob}_2(\text{no rec}|x_{(1,1)}(t_R)) = \exp\left[-\int_0^\infty \tilde{\beta}_{(0,0)}(t|x_{(1,1)}(t_R)) + \tilde{\beta}_{(0,1)}(t|x_{(1,1)}(t_R))\mathrm{d}t\right]. \tag{E.39}$$

The probability that type $(1,0)$ fixes is therefore given by:

$$\begin{aligned} P_{\mathrm{det}}^{((1,1)\to(1,0))} &= \int_0^\infty Q_{\mathrm{det}}^{(1,0)}(x_{(1,1)}(t_R))Q_{\mathrm{det}}^{(0,1)}(x_{(1,1)}(t_R))\alpha_{(1,0)}(t_R)\mathrm{Prob}_2(\text{no rec}|x_{(1,1)}(t_R))\mathrm{d}t_R. \\ &= \frac{1}{\sigma_{(1,1)}}\int_0^1 Q_{\mathrm{det}}^{(1,0)}(x)Q_{\mathrm{det}}^{(0,1)}(x)Nr_{(1,0)}^{(1,1)}p_{\mathrm{est}}^{(1,0)}(x)\mathrm{Prob}_2(\text{no rec}|x)\mathrm{d}x. \end{aligned} \tag{E.40}$$

The probability that a successful type $(0,0)$ individual is generated next at time τ_R, is given by

$$\frac{1}{\sigma_{(1,1)}}\int_0^1 Q_{\mathrm{det}}^{(1,0)}(x)Q_{\mathrm{det}}^{(0,1)}(x)Nr_{(1,0)}^{(1,1)}p_{\mathrm{est}}^{(1,0)}(x)\mathrm{e}^{-\int_0^{\tau_R}\tilde{\beta}_{(0,0)}(t|x)+\tilde{\beta}_{(0,1)}(t|x)\mathrm{d}t}\tilde{\beta}_{(0,0)}(\tau_R|x)\mathrm{d}x. \tag{E.41}$$

The probability that a successful type $(0,0)$ is generated next at any time is thus

$$\begin{aligned} &\frac{1}{\sigma_{(1,1)}}\int_0^1 Q_{\mathrm{det}}^{(1,0)}(x)Q_{\mathrm{det}}^{(0,1)}(x)Nr_{(1,0)}^{(1,1)}p_{\mathrm{est}}^{(1,0)}(x) \\ &\qquad\times\int_0^\infty \mathrm{e}^{-\int_0^{\tau_R}\tilde{\beta}_{(0,0)}(t|x)+\tilde{\beta}_{(0,1)}(t|x)\mathrm{d}t}\tilde{\beta}_{(0,0)}(\tau_R|x)\mathrm{d}\tau_R\mathrm{d}x. \end{aligned} \tag{E.42}$$

The probability that a successful type $(0,1)$ individual is generated next, is given by

$$\frac{1}{\sigma_{(1,1)}}\int_0^1 Q_{\text{det}}^{(1,0)}(x)Q_{\text{det}}^{(0,1)}(x)Nr_{(1,0)}^{(1,1)}p_{\text{est}}^{(1,0)}(x)\mathrm{e}^{-\int_0^{\tau_R}\tilde{\beta}_{(0,0)}(t|x)+\tilde{\beta}_{(0,1)}(t|x)\mathrm{d}t}\tilde{\beta}_{(0,1)}(\tau_R|x)\mathrm{d}x. \tag{E.43}$$

As numerical evaluation of these integrals is computationally expensive, we introduce the following approximation:

$$\begin{aligned}
&\int_0^\infty \exp\left[-\int_0^{\tau_R}\tilde{\beta}_{(0,0)}(t|x)+\tilde{\beta}_{(0,1)}(t|x)\mathrm{d}t\right]\tilde{\beta}_{(0,0)}(\tau_R|x)\mathrm{d}\tau_R\\
\approx&\int_0^\infty \exp\left[-\int_0^{\tau_R}\tilde{\beta}_{(0,0)}(t|x)\mathrm{d}t\right]\tilde{\beta}_{(0,0)}(\tau_R|x)\mathrm{d}\tau_R\\
&\times\frac{\int_0^\infty \exp\left[-\int_0^{\tau_R}\tilde{\beta}_{(0,0)}(t|x)+\tilde{\beta}_{(0,1)}(t|x)\mathrm{d}t\right](\tilde{\beta}_{(0,0)}(\tau_R|x)+\tilde{\beta}_{(0,1)}(\tau_R|x))\mathrm{d}\tau_R}{\int_0^\infty \mathrm{e}^{-\int_0^{\tau_R}\tilde{\beta}_{(0,0)}(t|x)\mathrm{d}t}\tilde{\beta}_{(0,0)}(\tau_R|x)\mathrm{d}\tau_R+\int_0^\infty \mathrm{e}^{-\int_0^{\tau_R}\tilde{\beta}_{(0,1)}(t|x)\mathrm{d}t}\tilde{\beta}_{(0,1)}(\tau_R|x)\mathrm{d}\tau_R}\\
=&\left(1-\exp\left[-\int_0^\infty\tilde{\beta}_{(0,0)}(t|x)\mathrm{d}t\right]\right)\\
&\times\frac{1-\exp\left[-\int_0^\infty\tilde{\beta}_{(0,0)}(t|x)+\tilde{\beta}_{(0,1)}(t|x)\mathrm{d}t\right]}{2-\exp\left[-\int_0^\infty\tilde{\beta}_{(0,0)}(t|x)\mathrm{d}t\right]-\exp\left[-\int_0^\infty\tilde{\beta}_{(0,1)}(t|x)\mathrm{d}t\right]}
\end{aligned} \tag{E.44}$$

and equivalently for the corresponding integral.

After establishment of type $(0,1)$, a successful individual of type $(0,0)$ might still be generated. We will ignore this probability. This will be a good approximation when the probability of three successful recombination events is low, i.e., in particular when recombination is low.

F Appendix: Diffusion with killing for $I = 0$, $J = 1$

Hartfield and Otto (2011) derive a diffusion equation for the probability that type $(0, 1)$ fixes in the population conditioned on fixation of the beneficial allele. We here give the equation adjusted to our model (for the derivation, we refer to Hartfield and Otto (2011)). We define:

$$S_{(0,0)} = N\sigma_{(0,0)}, \tag{F.1a}$$
$$S_{(0,1)} = N\sigma_{(0,1)}, \tag{F.1b}$$
$$\rho^{(0,1)}_{(0,0)} = Nr^{(0,1)}_{(0,0)}. \tag{F.1c}$$

The scaled version of Eq. (3.38) reads:

$$\pi(p) = \frac{S_{(0,0)}(S_{(0,0)} - S_{(0,1)})}{(S_{(0,0)} - S_{(0,1)})(1-p) + S_{(0,0)}p}, \tag{F.2}$$

where p is the relative frequency of type $(0, 1)$. The mean change in p over a time step measured in N generations is the same in both models. The variance in change of p is $V(p) = 2p(1-p)$ in our model (vs $V(p) = p(1-p)$ in the Wright-Fisher model). Following Hartfield and Otto (2011), we denote by $P^*(p)$ the conditional probability that type $(0, 1)$ fixes if its current frequency is p. We obtain:

$$\frac{\mathrm{d}^2 P^*(p)}{\mathrm{d}p^2} + \sigma_{(0,1)} \frac{1 + e^{-pS_{(0,1)}}}{1 - e^{-pS_{(0,1)}}} \frac{\mathrm{d}P^*(p)}{\mathrm{d}p} - \rho^{(0,1)}_{(0,0)} \pi(p) P^*(p) = 0. \tag{F.3}$$

The differential equation is solved with the boundary conditions

$$P^*(1) = 1, \tag{F.4a}$$
$$\left.\frac{\mathrm{d}P^*(p)}{\mathrm{d}p}\right|_{p=0} = 0. \tag{F.4b}$$

Figure F.1 compares the results for the hitchhiking probability as obtained via Eq. (F.3) and via Eq. (3.53). Note that the diffusion equation is conditioned on establishment of type $(0, 1)$, assuming an absolute rate of increase of $\sigma_{(0,1)}$, while $\bar{\nu}$ is based on an absolute rate of increase of $\sigma_{(0,1)} - r_{(0,1)}$.

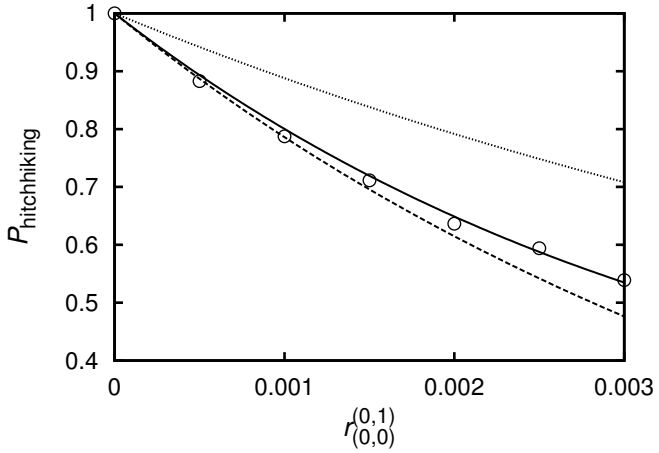


Figure F.1: The hitchhiking probability of a linked deleterious allele as a function of the recombination probability. The solid line is based on Eq. (3.53) as in Figure 3.4. The dashed line shows $P^{(0,1)}_{(0,1)} P^*(1/N)$ where $P^{(0,1)}_{(0,1)}$ and P^* are given by Eq. (3.52) and the solution of Eq. (F.3), respectively. The dotted line represents $P^*(1/N)$. The parameter values are the same as for Figure 3.4D. For $N = 10000$ as in Figure 3.4C, both results are basically indistinguishable.

References

Abi-Rached, L., *et al.*, 2011 The shaping of modern human immune systems by multiregional admixture with archaic humans. Science **334(89)**: 89–94.

Adams, J. R., J. A. Leonard, and L. P. Waits, 2003 Widespread occurence of a domestic dog mitochondrial DNA haplotype in southeastern US coyotes. Molecular Ecology **12(2)**: 541–546.

Agashe, D., J. J. Falk, and D. I. Bolnick, 2011 Effects of founding genetic variation on adaptation to a novel resource. Evolution **65**: 2481–2491.

Allen, L. J. S., 2011 *An Introduction to Stochastic Processes with Applications to Biology*. Pearson Education, Inc., New Jersey, second edition.

Arnold, M. J., 2004 Transfer and origin of adaptations through natural hybidization: Were Anderson and Stebbins right? The Plant Cell **16**: 562–570.

Arnold, M. L., M. R. Bulger, J. M. Burke, A. L. Hempel, and J. H. Williams, 1999 Natural hybridization: How low can you go and still be important? Ecology **80(2)**: 371–381.

Arnold, M. L., and N. H. Martin, 2009 Adaptation by introgression. Journal of Biology **8**: 82.

Athreya, K. B., and P. E. Ney, 1972 *Branching Processes*. Springer, Berlin.

Barton, N., and B. O. Bengtsson, 1986 The barrier to genetic exchange between hybridising populations. Heredity **56**: 357–376.

Barton, N. H., 1979 Gene flow past a cline. Heredity **43(3)**: 333–339.

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

Beaumont, M., *et al.*, 2001 Genetic diversity and introgression in the Scottish wildcat. Molecular Ecology **10**: 319–336.

Bell, G., 2013 Evolutionary rescue of a green alga kept in the dark. Biology Letters **9(1)**.

Bell, G., and S. Collins, 2008 Adaptation, extinction and global change. Evolutionary Applications **1(1)**: 3–16.

Bell, G., and A. Gonzalez, 2009 Evolutionary rescue can prevent extinction following environmental change. Ecology Letters **12**: 942–948.

Bell, G., and A. Gonzalez, 2011 Adaptation and evolutionary rescue in metapopulations experiencing environmental deterioration. Science **332**: 1327–1330.

Bello, Y., and D. Waxman, 2006 Near-periodic substitutions and the genetic variance induced by environmental change. Journal of Theoretical Biology **239**: 152–160.

Boulding, E. G., and T. Hay, 2001 Genetic and demographic parameters determining population persistence after a discrete change in the environment. Heredity **86**: 313–324.

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

Chevillon, C., M. Raymond, T. Guillemaud, T. Lenormand, and N. Pasteur, 1999 Population genetics of insecticide resistance in the mosquito *Culex pipiens*. Biological Journal of the Linnean Society **68(1–2)**: 147–157.

Chia, A. B., 1968 Random mating in a population of cyclic size. Journal of Applied Probability **5**: 21–30.

Cohn, H., and P. Jagers, 1994 General branching processes in varying environment. The Annals of Applied Probability **4(1)**: 184–193.

Collins, S. J., J. de Meaux, and C. Acquisti, 2007 Adaptive walks toward a moving optimum. Genetics **176**: 1089–1099.

Daborn, P. J., J. L. Yen, M. R. Bogwitz, G. Le Goff, E. Feli, *et al.*, 2002 A single p450 allele associated with insecticide resistance in drosophila. Science **297**: 2253–2256.

Demon, I., P. Haccou, and F. van den Bosch, 2007 Introgression of resistance genes between populations: A model study of insecticide resistance in *Bemisia tabaci*. Theoretical Population Biology **72**: 292–304.

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

Duputié, A., F. Massol, I. Chuine, M. Kirkpatrick, and O. Ronce, 2012 How do genetic correlations affect species range shifts in a changing environment. Ecology Letters **15**: 251–259.

Elena, S. F., V. S. Cooper, and R. E. Lenski, 1996 Puncutated evolution caused by selection of rare beneficial mutations. Science **272**: 1802–1804.

Elena, S. F., and R. E. Lenski, 2003 Evolution experiments with microorganisms: The dynamics and genetic basis of adaptation. Nature Reviews **4**: 457–469.

Ewens, W. J., 1967 The probability of survival of a new mutant in a fluctuating environment. Heredity **22**: 438–443.

Ewens, W. J., 2004 *Mathematical Population Genetics*. Springer, New York et al., second edition.

Ewing, G., J. Hermisson, P. Pfaffelhuber, and J. Rudolf, 2011 Selective sweeps for recessive alleles and for other modes of dominance. Journal of Mathematical Biology **63**: 399–431.

Fisher, R. A., 1922 On the dominance ratio. Proceedings of the Royal Society of Edinburgh **42**: 321–341.

Fisher, R. A., 1930 The distributions of gene ratios for rare mutations. Proceedings of the Royal Society of Edinburgh **50**: 204–219.

Fitzpatrick, B. M., *et al.*, 2010 Rapid spread of invasive genes into a threatend native species. Proceedings of the National Academy of Sciences **107(8)**: 3606–3610.

Galassi, M., J. Davies, J. Theiler, B. Gough, G. Jungman, *et al.*, 2009 *GNU Scientific Library Reference Manual*. Network Theory Ltd., Bristol, third edition.

Gatenby, R. A., 2009 A change of strategy in the war on cancer. Nature **459**: 508–509.

Gatenby, R. A., A. S. Silva, R. J. Gillies, *et al.*, 2009 Adaptive therapy. Cancer Research **69**: 4894–4903.

Gerstein, A. C., D. S. Lo, and S. P. Otto, 2012 Parallel genetic changes and non-parallel gene-environment interactions underlie nystatin resistance in yeast. Genetics **192**: 241–252.

Gillespie, D. T., 1977 Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry **81 (25)**: 2340–2361.

Gomulkiewicz, R., and R. D. Holt, 1995 When does evolution by natural selection prevent extinction? Evolution **49**: 201–207.

Gomulkiewicz, R., R. D. Holt, and M. Barfield, 1999 The effects of density-dependence and immigration on local adaptation and niche evolution in a black-hole sink environment. Theoretical Population Biology **55**: 283–296.

Gonzalez, A., and G. Bell, 2013 Evolutionary rescue and adaptation to abrupt environmental change depends upon the history of stress. Philosophical Transactions of the Royal Society B **368**: 20120079.

Gosh, A., and P. Haccou, 2010 Quantifying stochastic introgression processes with hazard rates. Theoretical Population Biology **77**: 171–180.

Gosh, A., P. G. Meirmans, and P. Haccou, 2012a Quantifying introgression risk with realistic population genetics. Proceedings of the Royal Society B **279**: 4747–4754.

Gosh, A., M. C. Serra, and P. Haccou, 2012b Quantifying time-inhomogeneous stochastic introgression processes with hazard rates. Theoretical Population Biology **81**: 253–263.

Gottelli, D., *et al.*, 1994 Molecular genetics of the most endangered canid: the ethopian wolf *Canis simensis*. Molecular Ecology **3(4)**: 301–312.

Greulich, P., B. Waclaw, and R. L. Allen, 2012 Mutational pathway determines whether drug gradients accelerate evolution of drug-resistant cells. Physical Review Letters **109**: 088101.

Haccou, P., P. Jagers, and V. A. Vatutin, 2005 *Branching Processes. Variation, Growth and Extinction of Populations*. Cambridge University Press, Cambridge et al.

Haldane, J. B. S., 1927 A mathematical theory of natural and artificial selection. v. selection and mutation. Proceedings of the Cambridge Philosophical Society **23**: 838–844.

Harris, T. E., 1963 *The Theory of Branching Processes*. Springer, Berlin, Göttingen, Heidelberg.

Hartfield, M., and S. P. Otto, 2011 Recombination and hitchhiking of deleterious alleles. Evolution **65(9)**: 2421–2434.

Heiser C. B., Jr., 1951 Hybridization in the Annual Sunflowers: Helianthus annuus × H. debilis Var. cucumerifolius. Evolution **5(1)**: 42–51.

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

Hermsen, R., J. B. Deris, and T. Hwa, 2012 On the rapidity of antibiotic resistance evolution facilitated by a concentration gradient. Proceedings of the National Academy of Sciences **109(27)**: 10775–10780.

Holt, R. D., and R. Gomulkiewicz, 2004 Conservation implications of niche conservatism and evolution in heterogeneous environments. In F. R., U. Dieckmann and D. Couvet, editors, *Evolutionary Conservation Biology*. Cambridge University Press, 244–264.

Huillet, T., 2011 On the Karlin-Kimura approaches to the Wright-Fisher diffusion with fluctuating selection. Journal of Statistical Mechanics: Theory and Experiment **2**.

Iwasa, Y., F. Michor, and M. A. Nowak, 2003 Evolutionary dynamics of escape from biomedical interention. Proceedings of the Royal Society B **270**: 2573–2578.

Iwasa, Y., F. Michor, and M. A. Nowak, 2004a Evolutionary dynamics of invasion and escape. Journal of Theoretical Biology **226**: 205–214.

Iwasa, Y., F. Michor, and M. A. Nowak, 2004b Stochastic tunnels in evolutionary dynamics. Genetics **166(3)**: 1571–1579.

Jagers, P., 1975 *Branching Processes with Applications to Biology*. John Whiley and Sons, London, New York, Sidney, Toronto.

Jensen, L., 1973 Random selective advantages of genes and their probabilities of fixation. Genetical Research **21(3)**: 215–219.

Karasov, T., P. W. Messer, and D. A. Petrov, 2010 Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. PloS Genetics **6(6)**: e1000924.

Karlin, S., and B. Levikson, 1974 Temporal fluctuations in selection intensities: case of small population size. Theoretical Population Biology **6**: 383–412.

Karlin, S., and H. M. Taylor, 1975 *A first course in stochastic processes*. Academic Press, San Diego, New York, Boston, London, Sydney, Tokyo, Toronto, second edition.

Kendall, D. G., 1948 On the generalized "birth-and-death" process. The Annals of Mathematical Statistics **19(1)**: 1–15.

Kimura, M., 1957 Some problems of stochastic processes in genetics. The Annals of Mathematical Statistics **28**: 882–901.

Kimura, M., 1962 On the probability of fixation of mutant genes in a population. Genetics **47**: 713 – 719.

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

Kimura, M., and T. Ohta, 1974 Probability of gene fixation in an expanding finite population. Proceedings of the National Academy of Sciences USA **71**: 3377 – 3379.

Kirkpatrick, M., and S. Peischl, 2013 Evolutionary rescue by beneficial mutations in environments that change in space and time. Philosophical Transactions of the Royal Society B **368**: 20120082.

Kopp, M., and J. Hermisson, 2007 Adaptation of a quantitative trait to a moving optimum. Genetics **176**: 715–719.

Kopp, M., and J. Hermisson, 2009a The genetic basis of phenotypic adaptation I: Fixation of beneficial mutations in the moving optimum model. Genetics **182**: 233–249.

Kopp, M., and J. Hermisson, 2009b The genetic basis of phenotypic adaptation II: Fixation of beneficial mutations in the moving optimum model. Genetics **183**: 1453–1476.

Lachapelle, J., and G. Bell, 2012 Evolutionary rescue of sexual and asexual populations in a deteriorating environment. Evolution **66(11)**: 3508–3518.

Lande, R., and S. Shannon, 1996 The role of genetic variation in adaptation and population persistence in a changing environment. Evolution **50(1)**: 434–437.

Levene, H., 1953 Genetic equilibrium when more than one ecological niche is available. The American Naturalist **87**: 331–333.

Lindner, C. R., I. Taha, G. J. Seiler, A. A. Snow, and L. H. Rieseberg, 1998 Long-term introgression of crop genes into wild sunflower populations. Theoretical and Applied Genetics **96**: 339–347.

Lynch, M., and A. Abegg, 2010 The rate of establishment of complex adaptations. Molecular Biology and Evolution **27(6)**: 1404–1414.

Lynch, M., W. Gabriel, and A. M. Wood, 1991 Adaptive and demographic responses of plankton populations to environmental change. Limnol. Oceanogr. **36(7)**: 1301–1312.

Mallet, J., 2005 Hybridization as an invasion of the genome. Trends in Ecology & Evolution **20(5)**: 229–237.

Martin, G., R. Aguilée, J. Ramsayer, O. Kaltz, and O. Ronce, 2013 The probability of evolutionary rescue: towards a quantitative comparison between theory and evolution experiments. Philosophical Transactions of the Royal Society B **368**: 20120088.

McKenzie, J. A., J. M. Dearn, and M. J. Whitten, 1980 Genetic basis of response to diazinon in victorian populations of the australian sheep blowfly, *Lucilia cuprina*. Australian Journal of Biological Sciences **33(1)**: 85–96.

Milani, R., 1963 Genetical aspects of insecticide resistance. Bulletin of the World Health Organization **29**: 77–87.

Miller, W., *et al.*, 2012 Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proceedings of the National Academy of Sciences : E2382–2390.

Moran, P. A. P., 1958a The effect of selection in a haploid genetic population. Mathematical Proceedings of the Cambridge Philosophical Society **54**: 463–467.

Moran, P. A. P., 1958b Random processes in genetics. Proceedings of the Cambridge Philosophical Sciences **54**: 60–71.

Normark, B., and S. Normark, 2002 Evolution and spread of antibiotic resistance. Journal of Internal Medicine **252(2)**: 91–106.

Novozhilov, A. S., G. P. Karev, and E. V. Koonin, 2006 Biological applications of the theory of birth-and-death processes. Briefings in Bioinformatics **7(1)**: 70–85.

Ohta, T., and K.-I. Kojima, 1968 Survival probabilities of new inversions in large populations. Biometrics **24(3)**: 501–516.

Orr, H. A., 2005 The genetic theory of adaptation: a brief history. Nature Reviews **6**: 119–127.

Orr, H. A., 2010 The population genetics of beneficial mutations. Philosophical Transactions of the Royal Society B **365**: 1195–1201.

Orr, H. A., and J. A. Coyne, 1992 The genetics of adaptation: a reassessment. The American Naturalist **140(5)**: 725–742.

Orr, H. A., and R. L. Unckless, 2008 Population extinction and the genetics of adaptation. The American Naturalist **172(2)**: 160 – 169.

Otto, S. P., and M. C. Whitlock, 1997 The probability of fixation in populations of changing size. Genetics **146**: 723 – 733.

Parsons, T. L., and C. Quince, 2007a Fixation in haploid populations exhibiting density dependence I: The non-neutral case. Theoretical Population Biology **72**: 121–135.

Parsons, T. L., and C. Quince, 2007b Fixation in haploid populations exhibiting density dependence II: The quasi-neutral case. Theoretical Population Biology **72**: 468–479.

Patwa, Z., and L. M. Wahl, 2008 The fixation probability of beneficial mutations. Journal of the Royal Society Interface **5**: 1279–1289.

Pease, L. M., R. Lande, and J. J. Bull, 1989 A model of population growth, dispersal, and evolution in a changing environment. Ecology **70**: 1657–1664.

Peña-Miller, R., D. Laehnemann, G. Jansen, A. Fuentes-Hernandez, P. Rosenstiel, *et al.*, 2013 When the most potent combination of antibiotics selects for the greatest bacterial load: the smile-frown transition. PLoS Biology **11(4)**: e1001540.

Polechová, J., N. H. Barton, and G. Marion, 2009 Species' range: Adaptation in space and time. The American Naturalist **174(5)**: E186–204.

Pollak, E., 1966 Some effects of fluctuating offspring distributions on the survival of genes. Biometrika **53(3/4)**: 391–396.

Pollak, E., 2000 Fixation probabilities when the population size undergoes cyclic fluctuations. Theoretical Population Biology **57**: 51–58.

Read, A. F., T. Day, and S. Huijben, 2011 The evolution of drug resistance and the curious orthodoxy of aggressive chemotherapy. Proceedings of the National Academy of Sciences **108**: 10871–10877.

Rhymer, J. M., and D. Simberloff, 1996 Extinction by hybridization and introgression. Annual Review of Ecology and Systematics **27**: 83–109.

Roush, R. T., and J. A. McKenzie, 1987 Ecological genetics of insecticide and acaricide resistance. Annual Reviews of Entomology **32**: 361–380.

Schiffers, K., E. C. Bourne, S. Lavergne, W. Thuiller, and J. M. J. Travis, 2013 Limited evolutionary rescue of locally adapted populations facing climate change. Philosophical Transactions of the Royal Society B **268**: 20120083.

Schwenk, K., N. Brede, and B. Streit, 2008 Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. Philosophical Transactions of the Royal Society B **363**: 2805–2811.

Serra, M. C., 2006 On the waiting time to escape. Journal of Applied Probability **43**: 296–302.

Serra, M. C., and P. Haccou, 2007 Dynamics of escape mutations. Theoretical Population Biology **72**: 167–178.

Sewastjanow, B. A., 1974 *Verzweigungsprozesse*. Akademie-Verlag, Berlin.

Snow, A. A., 2002 Transgenic crops – why gene flow matters. nature biotechnology **20**: 542.

Snow, A. A., *et al.*, 2003 A Bt transgene reduces herbivory and enhances fecundity in wild sunflowers. Ecological Applications **13(2)**: 279–286.

Takahasi, K. R., and F. Tajima, 2005 Evolution of coadaptation in a two-locus epistatic system. Evolution **59(11)**: 2324–2332.

Takahata, N., K. Ishii, and H. Matsuda, 1975 Effect of temporal fluctuation of selection coefficient on gene frequency in a population. Proceedings of the Natural Academy of Sciences **72**: 4541–4545.

The *Heliconius* Genome Consortium, 2012 Buttlerfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature **487**: 95–98.

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

Vander Wal, E., D. Garant, M. Festa-Bianchet, and F. Pelletier, 2013 Evolutionary rescue in vertebrates: evidence, applications and uncertainty. Philosophical Transactions of the Royal Society B **368**: 20120090.

Wahl, L. M., 2011 Fixation when N and s vary: Classic approaches give elegant new results. Genetics **188(4)**: 783–785.

Wahl, L. M., and P. J. Gerrish, 2001 The probability that beneficial mutations are lost in populations with periodic bottlenecks. Evolution **55**: 2606–2610.

Wang, Y., and B. Rannala, 2004 A novel solution for the time-dependent probability of gene fixation or loss under natural selection. Genetics **168**: 1081–1084.

Waxman, D., 2011 A unified treatment of the probability of fixation when population size and the strength of selection change over time. Genetics **188(4)**: 907–913.

Weissmann, D. B., M. M. Desai, D. S. Fisher, and M. W. Feldman, 2009 The rate at which asexual populations cross fitness valleys. Theoretical Population Biology **75(4)**: 286–300.

Whitlock, M. C., 2003 Fixation probability and time in subdivided populations. Genetics **164(2)**: 767–779.

Whitlock, M. C., and R. Gomulkiewicz, 2005 Probability of fixation in a heterogeneous environment. Genetics **171**: 1407–1417.

Whitney, K. D., R. A. Randell, and L. H. Rieseberg, 2006 Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. The American Naturalist **167(6)**: 794–807.

Wright, S., 1943 Isolation by distance. Genetics **28(2)**: 114–138.

Curriculum Vitae

Personal information:

Name:	Hildegard Uecker
Date and Place of Birth:	29th September 1982 in Würzburg, Germany
Nationality:	German

Research:

since 01/2009	PhD candidate in the Mathematics and Biosciences Group, University of Vienna & Max F. Perutz Laboratories, Austria (Advisor: Joachim Hermisson)
04/2011 – 07/2011	Short-term scholar at the University of British Columbia, Canada (Advisor: Sarah P. Otto)

Education:

10/2002 – 04/2008	Studies of Physics, Mathematics, and Philosophy at the University of Göttingen, Germany Interim exam in Mathematics (04/2004) Interim exam in Physics (07/2004) Diplom in Physics (04/2008)
09/2004 – 02/2005	Studies of Mathematics and Philosophy at the University of Caen Basse-Normandie, France
1993–2002	Wirsberg-Gymnasium (Grammar School) Würzburg Abitur (06/2002)
1989–1993	Johannes-Kepler-Grundschule (Elementary School) Würzburg

Publications

Journal Publications

- Hildegard Uecker, Sarah P. Otto, and Joachim Hermisson, Evolutionary rescue in structured populations, The American Naturalist (in press).

- Hildegard Uecker and Joachim Hermisson, On the fixation process of a beneficial mutation in a variable environment, Genetics 188: 915–930, 2011.

- Hildegard Uecker, Wolf Till Kranz, Timo Aspelmeier, and Annette Zippelius, Partitioning of energy in highly polydisperse granular gases, Physical Review E 80: 041303, 2009.

Conference Contributions

- Hildegard Uecker, Sarah P. Otto, Joachim Hermisson, On evolutionary rescue in structured populations, First Joint Congress on Evolutionary Biology, Ottawa 2012

- Hildegard Uecker and Joachim Hermisson, Fate and Fortuity: On the fixation process of a beneficial mutation in a variable environment, Workshop "Selection in Population Genetics", Paris 2011

- Hildegard Uecker, Sarah P. Otto, Joachim Hermisson, On the establishment of beneficial mutations and evolutionary rescue, Mathematical Biology Workshop and IGTC Summit, Victoria 2011

- Hildegard Uecker and Joachim Hermisson, On failure and success in a changing world, CSEE meeting, Banff 2011

- Hildegard Uecker and Joachim Hermisson, On the fixation process of a beneficial mutation in a variable environment, "Population Genetics Group", Hull 2011

- Hildegard Uecker and Joachim Hermisson, On the fixation probability and time of a beneficial mutation in a population, Symposium and Workshop "Challenges in Network Dynamics", Sestri Levante 2010

- Hildegard Uecker and Joachim Hermisson, Soft sweeps in structured populations, Workshop "Next-generation Sequencing: New Chances and Challenges for Evolutionary Genetics", Munich 2010

Other Publications

- Hildegard Uecker, Zur Temperatur und zur Geschwindigkeitsverteilung in granularen Gasen (Diploma thesis), Institut für Theoretische Physik, Universität Göttingen 2008

Acknowledgements

First, I would like to thank Professor Joachim Hermisson for the competent supervision of this thesis. Many discussions and enumerous emails helped me to solve more than one problem. Joachim gave me space and encouraged me to pursue my own ideas. His generous financial support allowed me to attend a series of workshops and conferences throughout my PhD.

I am particularly grateful to Professor Sally Otto for all her advice, her helpfulness, and mentoring. My research visit in her lab has influenced me a lot. I am also thankful to all the people in Vancouver who made me feel welcome, in particular to Thor Veen. During my time in Canada, I attended the meeting of the Canadian Society of Ecology and Evolution in Banff. Graham Bell's talk on evolutionary rescue at that meeting inspired the second chapter of this thesis.

I am most thankful to Professor Lindi Wahl and Professor Nick Barton who kindly agreed to review my thesis.

I am grateful to Professor Reinhard Bürger who was always supportive and encouraging. I have learned a lot in his lectures and seminars.

I thank the members of the DFG Forschergruppe "*Natural selection in structured populations*", in particular Prof. Peter Pfaffelhuber, for helpful discussions.

I thank Dr. Marc Timme and his group for repeated hospitality. The symposium in Sestri Levante was one of the best experiences during my PhD. I am also very thankful to many people in Göttingen who made me feel home and welcome whenever I came for a visit. I am particularly grateful to Andreas Sorge for a decade of friendship and his support with the layout of my thesis.

I would like to thank the members of the biomathematics group for good hours at the math department, especially Stephan Peischl, Christian Hilbe, Kristan Schneider, Claus Rueffler, Sebastian Matuszewski, and Derek Setter. Derek also provided me with the simulatios for Chapter 3 of this thesis. I am grateful to Claus Rueffler for his services as a bike courier. I thank Jitka Polechová for her friendship and advice.

Last, I would like to thank my family for unwavering support throughout my dissertation.