Chapter 12
Phase-Type Distribution Approximations of the Waiting Time Until Coordinated Mutations Get Fixed in a Population

Ola Hössjer, Günter Bechly and Ann Gauger

Abstract In this paper we study the waiting time until a number of coordinated mutations occur in a population that reproduces according to a continuous time Markov process of Moran type. It is assumed that any individual can have one of $m + 1$ different types, numbered as $0, 1, \ldots, m$, where initially all individuals have the same type 0. The waiting time is the time until all individuals in the population have acquired type m, under different scenarios for the rates at which forward mutations $i \to i + 1$ and backward mutations $i \to i - 1$ occur, and the selective fitness of the mutations. Although this waiting time is the time until the Markov process reaches its absorbing state, the state space of this process is huge for all but very small population sizes. The problem can be simplified though if all mutation rates are smaller than the inverse population size. The population then switches abruptly between different fixed states, where one type at a time dominates. Based on this, we show that phase-type distributions can be used to find closed form approximations for the waiting time law. Our results generalize work by Schweinsberg [60] and Durrett et al. [20], and they have numerous applications. This includes onset and growth of cancer for a cell population within a tissue, with type representing the severity of the cancer. Another application is temporal changes of gene expression among the individuals in a species, with type representing different binding sites that appear in regulatory sequences of DNA.

Keywords Coordinated mutations · Fixed state population · Moran model Phase-type distribution · Waiting time

O. Hössjer (✉)
Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden
e-mail: ola@math.su.se

G. Bechly · A. Gauger
Biologic Institute, 16310 NE 80th Street, Redmond, WA 98052, USA
e-mail: bechly@mac.com

A. Gauger
e-mail: ann.gauger@gmail.com

S. Silvestrov et al. (eds.), *Stochastic Processes and Applications*,
Springer Proceedings in Mathematics & Statistics 271,
https://doi.org/10.1007/978-3-030-02825-1_12

12.1 Introduction

A central problem of population genetics is to calculate the probability that a new germline point mutation survives and spreads from one individual to the rest of the population. This fixation probability depends not only on the selective fitness of the mutant compared to the wildtype variant, but also on the size of the population. Fisher [22, 23], Haldane [30], and Wright [67, 69] derived formulas for the fixation probability of a homogeneous one-sex or two-sex population without any subdivision. Their results were generalized by Kimura [34, 35], who formulated the fixation probability as a solution of Kolmogorov's backward equation. More recently, Lambert [41] gave a unified continuous branching process framework for calculating fixation probabilities of different population models. However, in order to know how fast genetic changes occur in a population, it is not only important to know the fixation probability, but also how long it takes for a surviving mutation to spread. This can be quantified in terms of the expected time until fixation, and for a homogeneous population this expected time was derived by Kimura and Ohta [38], Maruyama and Kimura [47, 48], and Kimura [36].

The above mentioned results have been generalized in different directions. First, a number of authors have analyzed fixation probabilities or the time to fixation of one single point mutation for models with geographic subdivision (Maruyama [46], Slatkin [61], Barton [3], Whitlock [65], Greven et al. [28]). Second, others have studied the waiting time until a more general type of DNA target gets fixed in a population, a process which involves several point mutations. This target could, for instance, be a double mutant at two loci with or without recombination (Bodmer [10], Christiansen et al. [14]). Another target is a subset of all possible DNA sequences at a number of tightly linked nucleotides. The evolutionary process then becomes a random walk on a fitness landscape of DNA strings, until the target set is reached (Gillespie [26], Chatterjee et al. [13]). This DNA string could, for instance, represent a regulatory region of a gene, and the target may consist of all sequences that contain a certain binding site of length 6–10 nucleotides, to which a transcription factor attaches and affects the expression of the gene (Stone and Wray [63], MacArthur and Brockfield [45], Yona et al. [70]). The waiting time until the new binding site arrives and gets fixed not only depends on the mutation rate, its selective advantage, and the size of the population, but also on the length of the regulatory region and the binding site, see for instance Durrett and Schmidt [18], Behrens and Vingron [7], Behrens et al. [8], Nicodéme [51], Tuğrul et el. [64], and Sanford et al. [56, 57].

Third, if several point mutations are required to reach a target that represents a complex adaption, these mutations must be coordinated in some way. For instance, it has been known for long that it is very difficult for several coordinated mutations to spread and get fixed if the intermediate states convey a selective disadvantage. In order for this to happen, the population has to be small or the mutations have to arrive fairly close in time. Wright's shifting balance theory (Wright [67, 68]) is an early attempt to explain this through geographic subdivision, where the coordinated mutations first occur and get fixed locally, before they spread to other subpopulations.

Kimura [37] considered a diploid model, and used a diffusion approach in order to find the expected waiting time until two coordinated mutations get fixed in the population, when each mutation by itself is negatively selected for, and the two loci are tightly linked or have a small recombination fraction between them. He approximated the two-dimensional process for the frequencies of the two mutant genes by a simpler, one-dimensional process. Stephan [62] generalized Kimura's model by allowing the two pathways towards the double mutation to have different mutation rates and selective disadvantages. Phillips [53] studied waiting times for two coordinated mutations to appear, using a somewhat similar model. He applied the solution to the first local phase of Wright's shifting-balance theory, and argued that this phase dominates the total waiting time until global fixation occurs.

The waiting time problem for coordinated mutations has several applications. It is widely believed, for instance, that many types of cancer occur when several somatic mutations spread in a population of cells within a tissue (Knudson [39]). This has been analyzed mathematically by Komarova et al. [40], Iwasa et al. [32, 33], Nowak [52], and Schinazi [58, 59]. A second related application is immune system response, where coordinated somatic mutations are triggered in reaction to certain antigens (Radmacher et al. [54]). A third application is to analyze the waiting time until multiple germline mutations arrive in duplicate genes in order to make them functional (Behe and Snoke [5, 6], Lynch [43]). A fourth application is coordinated germline mutations in regulatory regions, where changes at two different binding sites have to occur in a given order (Carter and Wagner [11], Durrett and Schmidt [19]). A fifth application is coordinated mutations in bacterial populations, where each surviving mutant gives rise to a daughter population that grows at an exponential rate (Axe [2]).

It is challenging to define a population genetic model that gives explicit formulas for the waiting time until several coordinated mutations occur. The reason is that such a model has to incorporate random gene frequency variation in terms of genetic drift, apart from selection and mutation. It is therefore necessary to study the time dynamics of the population's genetic composition by means of a stochastic process, and with at least two coordinated mutations, the state space of this process gets huge for all but very small populations sizes. Under certain assumptions the problem can be simplified though. For models with two coordinated mutations, this has been done in the above mentioned papers by Komarova et al. [40], Iwasa et al. [32, 33], and Durrett and Schmidt [19]. More recently, Schweinsberg [60] and Durrett et al. [20] obtained the asymptotic distribution for the waiting time until an arbitrary number m of coordinated mutations occur, when the intermediate alleles are neutral and no backward mutations are allowed. Their models have been used and extended by Lynch and Abegg [44], in order to study the waiting time until complex adaptive mutations are fixed. This work has been criticized by Axe [2], who argued that backward mutations should be included in models of complex adaptations.

The purpose of this paper is to generalize the framework of Schweinsberg [60] and Durrett et al. [20]. We derive asymptotic properties of the waiting time distribution until an arbitrary number m of coordinated mutations appear and the last one of them gets fixed, in a large population without any type of subdivision.

The mutations are allowed to have different selective fitness and mutation rates, and backward mutations are possible. The mutation probabilities are assumed to be smaller than the inverse population size, so that the genetic composition of the population changes rapidly between fixation of different genetic variants. This fixed state population model (Komarova et al. [40], Tuğrul et al. [64]) is conveniently modeled by a continuous time Markov process with a finite state space; the wildtype genetic variant and the m mutants. It is shown that asymptotically, the time until the mth mutant gets fixed in the population, has a phase-type distribution, that is, the distribution of the time the Markov process spends in non-absorbing states (or phases) before the absorbing state is reached (Neuts [50], Asmussen et al. [1]). We also give explicit approximations of the transition intensities of the Markov process. This includes transitions between non-adjacent mutations through stochastic tunneling (Carter and Wagner [11], Komarova et al. [40], Iwasa et al. [32]), where the intermediate genetic variants (the tunnel) are kept at a low frequency.

The paper is organized as follows: In Sect. 12.2 we introduce the framework for how the genetic composition of the population evolves over time by means of a Moran model (Moran [49], Section 3.4 of Ewens [21]) with deaths, births, and mutations with different selective fitness. In Sect. 12.3 we introduce the Markov process for fixed population states when the mutation rates are smaller than the inverse population size, and define the phase-type distribution for the time until this Markov process reaches its absorbing state. Then in Sect. 12.4 we give conditions under which the waiting time until the last mutant gets fixed, converges weakly towards a phase-type distribution, as the size of the population grows. After stating some results for the fixation of one single mutant in Sect. 12.5, we then provide explicit approximations, in Sect. 12.6, of the transition rates of the Markov process between different fixed states. Then we illustrate the theory for a number of asymptotic scenarios in Sect. 12.7, provide some adjustments of the asymptotic theory in Sect. 12.8, and give a summary with further extensions in Sect. 12.9. In Appendix A we provide a simulation algorithm, in Appendix B we derive an explicit approximation of the expected waiting time for one single mutant to get fixed, and in Appendix C we sketch proofs of main results.

12.2 Moran Model with Mutations and Selection

Consider a homogeneous and haploid population of constant size that consists of N individuals, all of which have the same sex. Each individual has one of $m+1$ possible types $0, 1, \ldots, m$. We can think of these types as different genetic variants or alleles, where 0 is a wildtype allele that is modified by m successive mutations. The genetic composition of the population is summarized by means of an $(m+1)$-dimensional vector

$$Z_t = (Z_{t0}, \ldots, Z_{tm}) \in \mathcal{Z}, \tag{12.1}$$

whose components represent the fraction of all alleles at time $t \geq 0$. It is assumed that t is a continuous parameter counted in units of generations. The allele frequency

configuration (12.1) is a stochastic process whose state space $\mathcal{Z}$ is the intersection of the m-simplex

$$\Delta = \left\{ z = (z_0, \ldots, z_m);\ z_i \geq 0, \sum_{i=0}^{m} z_i = 1 \right\}$$

spanned by the vectors

$$e_0 = (0, \ldots, 0), e_1 = (1, 0, \ldots, 0), \ldots, e_m = (0, 0, \ldots, 1),$$

and the set $\mathbb{N}^{(m+1)}/N$ of vectors z whose coordinates are natural numbers divided by N. More specifically, we will assume that (12.1) is a Moran model, where mutations between neighboring types $i \to i+1$ and $i \to i-1$ are possible, and where individuals with allele i have a selective fitness s_i, with $s_0 = 1$ and $s_i > 0$ for $i = 1, \ldots, m$. In our model, these numbers correspond to negative selection, neutral selection, and positive selection for allele i, depending on whether $0 < s_i < 1$, $s_i = 1$, and $s_i > 1$ respectively. The population starts with all individuals having type 0, so that $Z_0 = e_0$. It has overlapping generations, with a reproduction scheme defined as follows:

(i) Each individual dies independently according to a Poisson process with rate 1.
(ii) When an individuals dies, an offspring of some randomly chosen individual (including the one that dies) replaces it. The parent is chosen among the N individuals in the population, with probabilities proportional to their selection coefficients s_i.
(iii) If the parent has type $i < m$, the offspring in step (ii) mutates to $i+1$ with probability $u_{i+1} > 0$ and to $i-1$ with probability $v_{i-1} \geq 0$ (with $v_{-1} = 0$).

It follows from these reproduction rules that $\{Z_t;\ t \geq 0\}$ is a continuous time and time homogeneous Markov process on $\mathcal{Z}$. Our primary objects of study are the waiting time

$$T_m = \inf\{t \geq 0;\ Z_t = e_m\} \tag{12.2}$$

until allele m gets fixed in the population; and the waiting time

$$\tilde{T}_m = \inf\{t \geq 0;\ Z_{tm} > 0\} \tag{12.3}$$

until this allele first appears. Notice that T_m is the time until Z_t reaches the absorbing state e_m. On the other hand, $\tilde{T}_m$ is the hitting time of $\mathcal{Z}_m = \{z = (z_0, \ldots, z_m) \in \mathcal{Z};\ z_m > 0\}$, which is not an absorbing set of states, since the descendants of a type m individual may die out before this allele gets fixed in the population. However, if we modify the dynamics of Z_t and stop it as soon as it reaches $\mathcal{Z}_m$, we may treat this set as one single absorbing state.

It is also possible to allow for backward mutations when offspring of type m individuals are born, with probability v_{m-1}. Although this will affect the distribution of T_m, it will not impact the approximations of this distribution that we discuss in the following sections.

12.3 Phase-Type Distribution Approximation of Waiting Time

12.3.1 Asymptotic Notation

We will analyze the waiting times $T_m = T_{m,N}$ and $\tilde{T}_m = \tilde{T}_{m,N}$ asymptotically as the population size N tends to infinity. The various parameters of the model will in general depend on N as well, such as $u_i = u_{i,N}$, $v_i = v_{i,N}$, and $s_i = s_{i,N}$. We will use Bachmann–Landau asymptotic notation as $N \to \infty$, for instance $a_N \sim b_N$ if $a_N/b_N \to 1$, $b_N = O(a_N)$ if b_N/a_N stays bounded, $b_N = \Omega(a_N)$ if b_N/a_N is bounded away from zero, $b_N = \Theta(a_N)$ if a_N and b_N are of the same order (that is, $b_N = O(a_N)$ and $b_N = \Omega(a_N)$), and $b_N = o(a_N)$ if $b_N/a_N \to 0$. We will also make use of the analogous notation for sequences of random variables Y_N, with $Y_N = O_p(a_N)$ if Y_N/a_N stays bounded in probability and $Y_N = o_p(a_N)$ if Y_N/a_N converges to zero in probability. Suppose Y is a random variable with distribution F. We denote this as $Y \overset{\mathcal{L}}{\in} F$, and if Y_N is a sequence of random variables converging weakly towards Y, we often use the shorthand notation $Y_N \xrightarrow{\mathcal{L}} F$. For simplicity of notation, we will mostly omit index N for sequences of numbers or random variables that are functions of N. Sometimes, we also write $a_N \ll b_N$ or $b_N \gg a_N$ instead of $a_N = o(b_N)$.

12.3.2 Simplified Markov Process Between Fixed Population States

For all but very small N, it is not possible to get explicit and easily computable expressions for the distributions of T_m and $\tilde{T}_m$, since the state space $\mathcal{Z}$ gets huge when N grows. It is however possible to get accurate approximations of these distributions under appropriate conditions. The most crucial assumption is that the forward and backward mutation rates u_i and v_i tend to zero at a rate faster than the inverse population size, i.e.

$$u_i = o(N^{-1}), \tag{12.4}$$

and

$$v_i = o(N^{-1}) \tag{12.5}$$

for all i as $N \to \infty$. The implication of (12.4)–(12.5) is that most of the time, all individuals of the population will have the same type, and changes of this type occur rapidly when an individual with a new mutation gets many descendants that eventually take over the population. For this reason it may appear that a certain allele $i < m$ has been fixed permanently. But this is only temporary, since forward or backward mutations may later drive the population towards other fixed states.

This phenomenon was referred to as quasi-fixation in Hössjer et al. [31] for certain one-way mutation models with two possible alleles. Because of these rapid changes of the genetic decomposition Z_t of the population, it is well approximated by a continuous time Markov process defined on the finite subset

$$\mathcal{Z}_{\text{hom}} = \{e_0, \ldots, e_m\} \tag{12.6}$$

of $\mathcal{Z}$ that consists of all possible states of a type-homogeneous population. Here e_i refers to fixed state i of the population, so that all its individuals have the same type i. The simplified process has intensity matrix $\Lambda = (\lambda_{ij})_{i,j=0}^{m}$, where $\lambda_{ij} > 0$ is the rate of jumping from e_i to e_j when $j \neq i$, and $-\lambda_{ii} = \sum_{j;\, j\neq i} \lambda_{ij}$ is the rate of leaving e_i. Since e_m is an absorbing state, the intensity matrix can be decomposed as

$$\Lambda = \begin{pmatrix} \Lambda_0 & \lambda \\ 0 & 0 \end{pmatrix}, \tag{12.7}$$

where $\Lambda_0 = (\lambda_{i,j})_{i,j=0}^{m-1}$ contains the transition rates from and among the non-absorbing states, $0 = (0, \ldots, 0)$ is a row vector with m zeros, T denotes vector transposition, and $\lambda = (\lambda_{0m}, \ldots, \lambda_{m-1,m})^T$ is a column vector containing the transition rates from all non-absorbing states to e_m. A transition of Z_t from e_i to e_j corresponds to a stochastic tunneling event when $|j - i| \geq 2$. For instance, when $j \geq i + 2$, it represents a scenario where some individual who lives in a homogeneous type i population, has descendants from the same line of descent that experience mutations to $i + 1, i + 2, \ldots, j$, and then type j spreads to the whole population before any of the intermediate types do.

12.3.3 Defining Phase-Type Distribution Approximation

Since T_m is the time until the absorbing state e_m is reached, the simplified Markov process assumption with state space (12.6) and intensity matrix (12.7) implies that approximately

$$T_m \overset{\mathcal{L}}{\in} \text{PD}(\tilde{e}_0, \Lambda_0) \tag{12.8}$$

has a phase-type distribution, where $\tilde{e}_i$ is a unit vector of length m that contains the first m components of e_i. This phase-type distribution has two arguments, where the first, $\tilde{e}_0$, refers to the starting distribution of the Markov process among the non-absorbing states, and the second argument gives the intensity matrix among and from the non-absorbing states. From (12.8) we get very explicit approximate expressions for the density function

$$f_{T_m}(t) = \tilde{e}_0 \exp(\Lambda_0 t)\lambda, \quad t > 0, \tag{12.9}$$

the expected value

$$E(T_m) = -\tilde{e}_0 \Lambda_0^{-1} 1 \tag{12.10}$$

and the variance

$$\mathrm{Var}(T_m) = 2\tilde{e}_0 \Lambda_0^{-2} 1 - \left(\tilde{e}_0 \Lambda_0^{-1} 1\right)^2 \tag{12.11}$$

of T_m, with $1 = (1, \ldots, 1)^T$ a column vector of m ones.

In order to approximate the law of the waiting time $\tilde{T}_m$ in (12.3), we approximate $\{Z_t\}$ by a Markov process on

$$\tilde{\mathcal{Z}}_{\mathrm{hom}} = \{e_0, \ldots, e_{m-1}, \mathcal{Z}_m\}. \tag{12.12}$$

Then we may use (12.8) as a distributional approximation for $\tilde{T}_m$ rather than T_m, if λ_{im} is interpreted as a transition rate from e_i to $\mathcal{Z}_m$ when $i < m$, rather than from e_i to e_m.

12.4 Waiting Time Asymptotics

12.4.1 Regularity Conditions

In order to formulate precise asymptotic distributional results for T_m and $\tilde{T}_m$ when $N \to \infty$, we need some additional definitions and assumptions. We will focus on T_m, and then briefly point out the differences for $\tilde{T}_m$.

As a first step, let $\{\tau_k\}_{k=0}^M$ be the time points when a new allele gets fixed in the population. They are defined recursively as $\tau_0 = 0$ and

$$\tau_k = \inf\{t > \tau_{k-1};\ Z_t \in \{e_0, \ldots, e_m\} \setminus Z_{\tau_{k-1}}\}, \tag{12.13}$$

for $k = 1, 2, \ldots, M$, with $\tau_M = T_m$ the time point when Z_t reaches its absorbing state e_m. Clearly, $\{Z_{\tau_k};\ k = 0, 1, \ldots\}$ is a Markov chain with state space $\mathcal{Z}_{\mathrm{hom}}$ and transition probabilities

$$p_{ij} = p_{ij,N} = \begin{cases} P(Z_{\tau_{k+1}} = e_j | Z_{\tau_k} = e_i), & i = 0, \ldots, m-1, \\ & j = 0, \ldots, m, \\ 0, & i = m,\ j = 0, \ldots, m-1. \end{cases} \tag{12.14}$$

Since $Z_{\tau_k} \neq Z_{\tau_{k+1}}$ for $k < M$, the diagonal elements of the transition matrix $P = (p_{ij})_{i,j=0}^m$ vanish for all non-absorbing states, i.e. $p_{ii} = 0$ for $i < m$.

Assume that $Z_{\tau_k} = e_i$ for some $k < M$ and $i < m$. We will study what happens between the time points τ_k and τ_{k+1}, and refer to a forward mutation $i \to i + 1$ as successful if its descendants eventually take over the population. Let $f_i = f_{i,N}$ be the probability that a forward mutation that happens while all individuals of the

population have the same type i, is successful. Likewise, if $i > 0$, a successful backward mutation of type $i \to i-1$ is one whose descendants eventually take over the population. Denote by $b_i = b_{i,N}$ the probability that a backward mutation is successful, given that it happens in a homogeneous type i population. For definiteness, we also put $b_0 = 0$. A successful mutation from type i is either forward or backward, and due to (12.4)–(12.5), it will arrive when the population is homogeneous or almost homogeneous of type i. Therefore, a successful mutation from type i arrives at a rate close to

$$\mu_i = \mu_{i,N} = N v_{i-1} b_i + N u_{i+1} f_i, \quad i = 0, \ldots, m-1, \tag{12.15}$$

since new backward and forward mutations appear at rates $N v_{i-1}$ and $N u_{i+1}$ among N individuals with the same type i, but only a fraction b_i and f_i of them are successful, and cause a change in the population to another fixed state. For the absorbing state we put $\mu_m = 0$.

There will be at least one successful mutation within (τ_k, τ_{k+1}), and let τ'_{k+1} be the time point when the first of these mutations arrives. We will assume below that $\tau_{k+1} - \tau'_{k+1}$ is asymptotically negligible in comparison to the total waiting time T_m, which reflects that fact that all transitions of Z_t occur rapidly. This suggests that it is asymptotically accurate to use transition rates

$$\lambda_{ij} = \begin{cases} -\mu_i, & i = j, \\ \mu_i p_{ij}, & i \neq j, \end{cases} \tag{12.16}$$

in (12.8). In order to verify this we need to make some additional assumptions on how the rates in (12.16) behave as $N \to \infty$. We will first of all assume that the transition probabilities in (12.14) satisfy

$$p_{ij} \to \pi_{ij}, \quad i, j = 0, 1, \ldots, m, \tag{12.17}$$

as $N \to \infty$, so that $\Pi = (\pi_{ij})_{i,j=0}^{m}$ is the asymptotic transition matrix of the embedded Markov chain $\{Z_{\tau_k};\ k = 0, 1, \ldots\}$. We will then postulate that

$$(I - \Pi_0)^{-1} \text{ is invertible}, \tag{12.18}$$

with I the identity matrix of order m and Π_0 a square matrix of order m that contains the first m rows and first m columns of Π. Condition (12.18) guarantees that the asymptotic Markov chain reaches its absorbing state e_m with probability 1, since it implies $P(M < \infty) = \tilde{e}_0 (I - \Pi_0)^{-1} \pi = 1$, where $\pi = (\pi_{0m}, \ldots, \pi_{m-1,m})^T$. Let

$$I_{\text{as}} = \{i;\ 0 \leq i \leq m-1, \tilde{e}_0 (I - \Pi_0)^{-1} \tilde{e}_i^T > 0\} \tag{12.19}$$

refer to the asymptotic states. It consists of those non-absorbing states that are visited with a positive probability asymptotically as $N \to \infty$, since $i \in I_{\text{as}}$ is

equivalent to requiring that $(\Pi^k)_{0i} > 0$ for at least one $k = 0, 1, \ldots$. The remaining non-asymptotic states are denoted as

$$I_{\text{nas}} = \{1, \ldots, m-1\} \setminus I_{\text{as}}. \tag{12.20}$$

Among the asymptotic states, it is also important to know for how long time they are visited. We therefore express the expected waiting time

$$E(T_m) = E_0 + E_1 + \cdots + E_{m-1}$$

until allele m gets fixed as a sum of m terms, with $E_i = E_{i,N} = -\tilde{e}_0 \Lambda_0^{-1} \tilde{e}_i^T$ the expected time spent in state e_i before absorption into state e_m takes place. Notice that Λ_0 is invertible for each finite N, and therefore each E_i is well defined with $0 < E_i < \infty$. Indeed, since all $u_i > 0$, it follows that any fixed population state e_j with $j = i+1, \ldots, m$ can be reached from fixed population state e_i in one step. Therefore, all entries of Λ_0 above the diagonal are strictly positive, whereas the diagonal elements and row sums of Λ_0 are strictly negative. From the Gershgorin Circle Theorem we deduce that all eigenvalues of Λ_0 have a strictly negative real part, so that Λ_0 is invertible. We will assume that the limits

$$E_i / E(T_m) \to c_i, \quad 0, 1, \ldots, m-1, \tag{12.21}$$

exist as $N \to \infty$, and define

$$I_{\text{long}} = \{i;\ 0 \le i \le m-1, c_i > 0\} \tag{12.22}$$

as the set of asymptotic states e_i that are visited for such a long time that they have an asymptotic contribution to the expected waiting time (12.10). We also put

$$I_{\text{short}} = I_{\text{as}} \setminus I_{\text{long}} \tag{12.23}$$

for those non-absorbing states that are asymptotic, but visits to them are too short to have an asymptotic impact on the expected waiting time. It follows from (12.21) that the transition rates from the states in I_{long} have the same order

$$\mu_{\min} = \mu_{\min,N} = \min\{\mu_i;\ i \in I_{\text{long}}\}, \tag{12.24}$$

and it is the inverse of (12.24) that determines the asymptotic size of the waiting time (12.2). We will therefore rescale time in units of $\mu_{\min}$ and assume that

$$\frac{\mu_i}{\mu_{\min}} \to \kappa_i, \quad i = 0, \ldots, m, \tag{12.25}$$

as $N \to \infty$, where the normalized rate κ_i of leaving state e_i satisfies $1 \le \kappa_i < \infty$ for $i \in I_{\text{long}}$, $\kappa_i = \infty$ for $i \in I_{\text{short}}$, $0 \le \kappa_i \le \infty$ for $i \in I_{\text{nas}}$, and $\kappa_m = 0$. In order to

ensure that the time between the appearance of a successful mutation and fixation of a new allele is asymptotically negligible, we assume that

$$P\left(\tau_{k+1} - \tau'_{k+1} > \varepsilon \mu_{\min}^{-1} | Z_{\tau_k} = e_i\right) \to 1 \ \forall \varepsilon > 0 \text{ and } i \in I_{\text{as}}, \tag{12.26}$$

as $N \to \infty$. Notice that the probability on the left hand side of (12.26) does not depend on k, because of the Markov property of $\{Z_{\tau_k}\}$.

12.4.2 Main Results on Waiting Time Asymptotics

The following theorem specifies the asymptotic phase-type distribution of the waiting time T_m until allele m gets fixed. A proof of it is sketched in Appendix C.

Theorem 12.1 *Consider a Moran model for a population of size N with types (alleles) $0, \ldots, m$ that starts with all its individuals in allelic state 0 and then reproduces according to (i)–(iii) of Sect. 12.2, so that forward ($i \to i+1$) and backward ($i \to i-1$) mutations between nearby alleles are possible. Assume that the forward and backward mutation rates satisfy (12.4)–(12.5), that the transition probabilities between fixed population states where all individuals have the same allele, converge as in (12.17)–(12.18), that the expected times spent in various fixed states converge as in (12.21), that the rates of leaving the various fixed states satisfy (12.25), and that the time between appearance of a new successful mutation and fixation of a new allele is asymptotically negligible (12.26), as $N \to \infty$. Then the waiting time T_m until allele m gets fixed has a phase-type distribution*

$$\mu_{min} T_m \xrightarrow{\mathcal{L}} PD(\tilde{e}_0, \Sigma_0), \tag{12.27}$$

asymptotically as $N \to \infty$, when rescaled by μ_{min} in (12.24), the minimal rate of leaving a fixed state, among those that are visited for a positive fraction of time. The second argument Σ_0 on the right hand side of (12.27) contains the first m rows and first m columns of the intensity matrix $\Sigma = (\Sigma_{ij})_{i,j=0}^{m}$, with

$$\Sigma_{ij} = \begin{cases} -\kappa_i, & j = i, \\ \kappa_i \pi_{ij}, & j \neq i, \end{cases} \tag{12.28}$$

π_{ij} is the asymptotic transition probability (12.17) between fixed states i and j, and κ_i is the normalized rate (12.25) of leaving fixed state i.

We will make some comments on the asymptotic inverse size $\mu_{\min}$ of the waiting time T_m, and the matrix Σ_0 of the limit distribution in (12.27). In some applications, it convenient to generalize (12.24) and let

$$\mu_{\min} = C \min\{\mu_i;\ i \in I_{\text{long}}\} \tag{12.29}$$

for some constant $C > 0$, chosen in order to get a simple expression for $\mu_{\min}$. It is straightforward to see that Theorem 12.1 remains unchanged with this minor modification. The matrix Σ_0 contains asymptotic transition rates among and from all non-absorbing states, after the change of time scale in (12.24). It will be degenerate when either I_{short} or I_{nas} are non-empty. However, it turns out that (12.27) is still well defined, if we disregard those rows and columns of Σ_0 that correspond to I_{nas} and take the limit $\kappa_i \to \infty$ for all $i \in I_{\text{short}}$.

Consider the special case when either all $v_i = 0$, or that backward mutations have no asymptotic impact on the waiting time distribution. An important instance of (12.27) occurs if, in addition, successful forward mutations in a homogeneous type i environment always causes the same allele $F(i) > i$ to get fixed in the population, i.e.

$$\pi_{i,F(i)} = 1, \quad i \in I_{\text{as}}. \tag{12.30}$$

With this extra regularity condition, we obtain the following corollary of Theorem 12.1:

Corollary 12.1 *Consider the Moran model of Sect. 12.2. Assume that the conditions of Theorem 12.1 hold, that only forward mutations have an asymptotic impact on the population dynamics in such a way that the forward jumps between fixed population states occur according to* (12.30)*. Then the waiting time T_m until allele m gets fixed has a hypoexponential limit distribution*

$$\mu_{min} T_m \xrightarrow{\mathcal{L}} \sum_{i \in I_{long}} \kappa_i^{-1} X_i \tag{12.31}$$

as $N \to \infty$, where $X_0, \ldots, X_{m-1}$ are independent and exponentially distributed random variables with expected value 1.

Remark 12.1 The asymptotic result for the waiting time distribution of $\tilde{T}_m$ is analogous to (12.27), if we replace e_m by $\mathcal{Z}_m$ in all definitions. In particular, we interpret Σ_{im} as a normalized transition rates from e_i to $\mathcal{Z}_m$ (rather than to e_m) for $i = 0, \ldots, m-1$.

12.5 Fixation in a Two Type Moran Model Without Mutations

As a preparation for the next sections, we will state two well known result on the fixation probability and expected time to fixation, for a Moran model with two alleles ($m = 1$) and no mutations. If these two alleles start at frequencies $N - 1$ and 1, and have selection coefficients 1 and $s > 0$ respectively,

$$\beta(s) = \beta_N(s) = \begin{cases} 1/N, & s = 1, \\ (1 - s^{-1})/(1 - s^{-N}), & s \neq 1 \end{cases} \tag{12.32}$$

is the probability that the second allele gets fixed, whereas $1-\beta(s)$ is the probability that the first allele does (Komarova et al. [40], Section 6.1 of Durrett [17]). We will make frequent use of asymptotic expressions for the fixation probability of large populations. It follows from (12.32) that

$$\beta(s) \sim \begin{cases} (s^{-1}-1)\times s^N, & 1-s \gg 1/N, \\ x/\left[1-\exp(-x)\right]\times 1/N, & s=1+x/N, \\ 1-s^{-1}, & s-1 \gg 1/N, \end{cases} \tag{12.33}$$

as $N\to\infty$, where $x\neq 0$ in the second line is a constant, not depending on N.

Given that the second alleles takes over, we let $\alpha(s)$ be the expected time it takes for this to happen. Kimura and Ohta [38] derived a general diffusion approximation of $\alpha(s)$ for a large class of models with two alleles, see also Section 8.9 of Crow and Kimura [15], or Theorems 1.32 and 6.3 of Durrett [17]. In Appendix B we calculate this diffusion approximation $\alpha(s)$ for the Moran model of Sect. 12.2. In particular, we show that this diffusion approximation is of the order

$$\alpha(s)=\alpha_N(s)\sim \begin{cases} (1+s)\log(N)/(1-s), & \text{if } s<1, \\ N, & \text{if } s=1, \\ (1+s)\log(N)/(s-1), & \text{if } s>1, \end{cases} \tag{12.34}$$

asymptotically as $N\to\infty$, if s is kept fixed. The expected time to fixation in (12.34) is much different for neutral and non-neutral alleles. This is also true for the more accurate diffusion approximation of $\alpha(s)$ in Appendix B, although it has a somewhat smoother transition between $s=1$ and $s\neq 1$.

12.6 Explicit Approximate Transition Rates Between Fixed Population States

Returning to the general model with m mutations, we recall that Theorem 12.1 gives quite general conditions under which the normalized waiting times $\mu_{\min}T_m$ and $\mu_{\min}\tilde{T}_m$ have asymptotic phase-type distributions as the population size $N\to\infty$. Under these assumptions the unnormalized waiting times T_m (cf. (12.8)) and $\tilde{T}_m$ are also well approximated by phase-type distributions. But in order to apply these results we still need to find explicit approximations of the Markov transition rates λ_{ij} in (12.8) and (12.15)–(12.16) between fixed population states. As in Sect. 12.4 we focus on T_m and then pinpoint the difference when $\tilde{T}_m$ is of interest.

12.6.1 Defining Approximate Transition Rates

We introduce

$$\hat{\lambda}_{ij} = \begin{cases} N u_{i+1} r_{ij} \beta(s_j/s_i), & j > i, \\ N v_{i-1} r_{ij} \beta(s_j/s_i), & j < i \end{cases} \tag{12.35}$$

as an approximation of λ_{ij} when $i < m$ and $j \neq i$. The quantity $r_{ij} = \hat{q}_{ij}$ approximates a certain probability q_{ij}. When $|j - i| = 1$ we put $q_{i,i-1} = q_{i,i+1} = 1$. When $j \geq i + 2$, q_{ij} is a probability of tunneling from $i + 1$ to j. In more detail, q_{ij} is the probability that a forward mutation $i \to i + 1$, that occurs in a homogeneous type i population, gets a at least one descendant that mutates from $j - 1$ to j before any other allele gets fixed. Analogously when $j \leq i - 2$, q_{ij} is the probability of tunneling backwards from $i - 1$ to j. That is, q_{ij} is the probability for a backward mutation $i \to i - 1$, that occurs in a homogeneous population of type i individuals, to get at least one descendant that mutates to from $j + 1$ to j before any other allele gets fixed. For definiteness, we also put $\hat{\lambda}_{mj} = 0$ for all j.

It follows from (12.32) that the $\beta(s_j/s_i)$ term of (12.35) is the probability that descendants of one single type j individual take over a population where all the others have type i, if further mutations do not occur. In our setting, it is an approximation of the probability that the descendants of the individual that first mutated into j, take over the population before any new mutations occur. In order for this approximation to be accurate, it is required that no other allele than i attains a high frequency before j gets fixed (recall that the type j mutation itself was a descendant of a successful $i \to i \pm 1$ mutation, that appeared in homogeneous or almost homogeneous type i population).

In order to finalize the definition of $\hat{\lambda}_{ij}$ in (12.35) we must specify how r_{ij} approximates q_{ij}. When $|j - i| = 1$ we put $r_{ij} = 1$. When $|i - j| \geq 2$, we introduce explicit approximations

$$r_{ij} = \begin{cases} \prod_{l=i+1}^{j-1} R(\rho_{ilj})^{2^{-(l-i-1)}} u_{l+1}^{2^{-(l-i)}}, & j > i, \\ \prod_{l=j+1}^{i-1} R(\rho_{ilj})^{2^{-(i-l-1)}} v_{l-1}^{2^{-(i-l)}}, & j < i \end{cases} \tag{12.36}$$

of the tunneling probabilities q_{ij} that are accurate when no other allele reaches a high frequency during a transition from i to j. The parameters ρ_{ilj} in (12.36) quantify the difference between selection coefficients s_i and s_l, on a scale determined by a tunneling probability from l to j. When $i < j$, they are defined through

$$\frac{s_l}{s_i} = 1 + \rho_{ilj} \sqrt{u_{l+1} r_{ilj}} \tag{12.37}$$

for $l = j - 1, \ldots, i + 1$. This involves some other quantities r_{ilj}, which are also defined recursively, for $l = j - 1, \ldots, i$, starting with $r_{i,j-1,j} = 1$ and then using the relation

$$r_{ilj} = R(\rho_{i,l+1,j}) \sqrt{r_{i,l+1,j} u_{l+2}}. \tag{12.38}$$

When this recursion has stopped at $l = i$ we obtain the upper row of (12.36) by putting $r_{ij} = r_{iij}$. Here r_{ilj} approximates the probability q_{ilj} that a mutation $l \to l+1$, which occurs in a homogeneous type i population, gets a least one descendant that mutates into j, before any other allele gets fixed. In particular, $q_{iij} = q_{ij}$. Similarly, when $i > j$, we have that

$$\frac{s_l}{s_i} = 1 + \rho_{ilj}\sqrt{\nu_{l-1} r_{ilj}} \tag{12.39}$$

for $l = j+1, \ldots, i-1$. The probabilities r_{ilj} are defined recursively for $l = j+1, \ldots, i$, starting with $r_{i,j+1,j} = 1$, then iterating

$$r_{ilj} = R(\rho_{i,l-1,j})\sqrt{r_{i,l-1,j}\nu_{l-2}}, \tag{12.40}$$

and finally getting the lower row of (12.36) from $r_{ij} = r_{iij}$. The function

$$R(\rho) = \frac{\sqrt{\rho^2+4}+\rho}{2} \tag{12.41}$$

specifies the way in which differences between the selection coefficients $s_i, \ldots, s_{j-1}$ affect the probability r_{ij} in (12.36), see also equation (10) of Iwasa et al. [32] or equation (5) of Durrett and Schmidt [19]. Intuitively, if type l is more fit than i, then $\rho_{ilj} > 0$, and the probability in (12.36) increases (since $R(\rho) > 1$ when $\rho > 0$, in particular $R(\rho) \sim \rho$ when $\rho \to \infty$), if $s_l = s_i$ then $\rho_{ilj} = 0$ will have no impact on r_{ij} (since $R(0) = 1$), and finally, if l is less fit than i and therefore $\rho_{ilj} < 0$, this will decrease the probability in (12.36) (since $R(\rho) < 1$ when $\rho < 0$, in particular $R(\rho) \sim -1/\rho$ as $\rho \to -\infty$).

It is possible to obtain a more accurate approximation of q_{ij} than (12.36), without using the quantities ρ_{ilj} nor the function R (see the end of the proof of Lemma 12.3 in Appendix C for details). Formula (12.36) is more explicit though, and therefore it gives more insight into how the mutation rates and the selection coefficients affect the approximate tunneling probabilities r_{ij}.

12.6.2 Conditions Under Which Approximate Transition Rates Are Accurate

It turns out that Eqs. (12.35) and (12.36) are good approximations of λ_{ij} for those forward transition rates ($j > i$) and backward transition rates ($j < i$) that dominate asymptotically, provided there is exactly one forward rate and at most one backward rate from i that dominate. This can be formulated as follows. Define

$$\hat{\mu}_i = -\hat{\lambda}_{ii} = \sum_{j;j\neq i} \hat{\lambda}_{ij} \text{ for } i = 0, \ldots, m-1, \tag{12.42}$$

as an approximation of the rate μ_i in (12.15) at which a successful mutation occurs in a type i population, and suppose that

$$\hat{p}_{ij} = \frac{\hat{\lambda}_{ij}}{\hat{\mu}_i} \to \hat{\pi}_{ij} \tag{12.43}$$

as $N \to \infty$ for all $0 \le i \le m-1$ and $j \ne i$. For definiteness we also put $\hat{\pi}_{ii} = 0$ and $\hat{\pi}_{mj} = 0$ for $j = 0, \ldots, m-1$. We assume there is at most one index $0 \le B(i) < i$ for each $i = 1, \ldots, m-1$, and exactly one index $i < F(i) \le m$ for each $i = 0, \ldots, m-1$, such that fixation events from i will always be to $B(i)$ for backward mutations, and to $F(i)$ for forward mutations. This can phrased as

$$\begin{aligned} \hat{\pi}_{0,F(0)} &= 1, \\ \hat{\pi}_{i,B(i)} + \hat{\pi}_{i,F(i)} &= 1, i = 0, \ldots, m-1, \\ \hat{\pi}_{i,F(i)} &> 0, i = 1, \ldots, m-1, \end{aligned} \tag{12.44}$$

with $\hat{\pi}_{i,B(i)} = 0$ in the middle equation when $B(i) = \emptyset$, i.e. when backward mutations from a type i population have no asymptotic impact. In particular, the forward fixation from i involves stochastic tunneling if $F(i) \ge i+2$. In order for this to happen, $\hat{\lambda}_{iF(i)}$ must have a larger order asymptotically than all other $\hat{\lambda}_{ij}$ with $j > i$. It follows from (12.36) and some of the regularity conditions below, that a necessary condition for this to happen is that type j is more beneficial for reproduction than all the intermediate alleles. A similar condition applies for backward mutations, and we can summarize these necessary tunneling conditions as follows:

$$\begin{aligned} F(i) \ge i+2 &\Longrightarrow s_{F(i)} > \max\left(s_{i+1}, \ldots, s_{F(i)-1}\right), \\ B(i) \le i-2 &\Longrightarrow s_{B(i)} > \max\left(s_{B(i)+1}, \ldots, s_{i-1}\right), \end{aligned} \tag{12.45}$$

where the lower equation only applies when $B(i) \ne \emptyset$. We will need some additional regularity conditions. The first one consists of four relations

$$\begin{aligned} u_i/u_{i+1} &= O(1), && i = 0, \ldots, m-1, \\ v_{i-1} &= O(u_{i+1}), && i = 1, \ldots, m-1, \\ v_j/v_{j-1} &= O(1), && \text{if } B(i) < j < i \text{ for some } i = 2, \ldots, m-1, \\ u_{j+1} &= O(v_{j-1}), && \text{if } B(i) < j < i \text{ for some } i = 2, \ldots, m-1, \end{aligned} \tag{12.46}$$

each of which imposes some restrictions on the mutation rates. The second and fourth relations of (12.46) guarantee that backward mutations will have no asymptotic impact on forward fixations, and vice versa. The first equation of (12.46) requires that mutation rates to higher types are at least of the same order as mutation rates to lower types. Otherwise forward stochastic tunneling will be more difficult, and the formulas for some of the tunneling probabilities in (12.36) will look different. The third equation of (12.46) is the analogous requirement on backward mutations.

Notice that neither the third nor the fourth relation of (12.46) apply when back mutations do not exist or have no asymptotic impact, i.e. if $B(i) = \emptyset$ for all i.

In order to assure that condition (12.26) holds, i.e. that the time for successful mutations to get fixed are asymptotically negligible, we will assume that

$$\begin{aligned} &N \min_{i \in I_{\text{long}}} u_{F(i)}^{2-2^{-(F(i)-i-1)}} \beta(s_{F(i)}/s_i) \\ &= o\left(\min_{i \in I_{\text{as}}} \min\left[\alpha^{-1}\left(\tfrac{s_{B(i)}}{s_i}\right), \alpha^{-1}\left(\tfrac{s_{F(i)}}{s_i}\right)\right]\right), \end{aligned} \tag{12.47}$$

where $\beta(s)$ and $\alpha(s)$ are the fixation probability (12.32) and expected fixation time (12.34), respectively. If (12.47) does not hold, T_m will not only be affected by the waiting times for successful mutations to occur, but their fixation time will also have an impact.

The next regularity condition requires that the parameters of Eqs. (12.37)–(12.39) are bounded, i.e.

$$|\rho_{ilj}| = O(1), \quad i < l < j \text{ or } j < l < i \tag{12.48}$$

when $N \to \infty$. This means that the fitness $s_1, \ldots, s_{m-1}$ of the first $m-1$ mutant alleles approach 1 as N grows, so that each one of them is either slightly deleterious, neutral or slightly advantageous compared to the wildtype allele 0. The case of strong negative or positive selection ($\rho_{ilj} \to \pm\infty$ respectively) is not included in (12.48), but has been studied by Komarova et al. [40].

12.6.3 Asymptotic Distribution of Wating Time Based on Approximate Transition Rates

Equipped with the definitions and regularity conditions of Sect. 12.6.2, we are ready to formulate an asymptotic distributional result for the waiting time T_m (see Appendix C for a sketch of proof), where its limiting phase-type distribution can be derived from the explicit approximation (12.35) of the transition intensities between the fixed population states of the simplified Markov process.

Theorem 12.2 *Consider a Moran model for a population with N individuals and alleles $0, \ldots, m$ that starts with all its individuals in allelic state 0, and then reproduces according to (i)–(iii) of Sect. 12.2. Assume, as in Theorem 12.1, that (12.4)–(12.5), (12.17), (12.21), and (12.25) hold, and let λ_{ij} be the Markov transition rate (12.16) between two fixed population states i and j where all individuals have the same allele i and j respectively. Define $\hat{\lambda}_{ij}$ in (12.35) as an approximation of λ_{ij}, with $\hat{\mu}_i$ the approximate rate (12.42) of leaving state i and $\hat{\pi}_{ij}$ an approximation (12.43) of the probability π_{ij} in (12.17) of jumping from fixed state i to fixed state j, whereas $\hat{\mu}_{min} = \min_{i \in I_{long}} \hat{\mu}_i$ is an approximation of the minimal rate of leaving a fixed state, among those that are visited for a positive fraction of time. Assume further that (12.44)–(12.48) hold. Then $\pi_{ij} = \hat{\pi}_{ij}$ and $\hat{\mu}_i / \hat{\mu}_{min} \to \kappa_i$ for $i = 0, \ldots, m$*

as $N \to \infty$, where κ_i is the normalized rate (12.25) of leaving fixed state i. Moreover, the waiting time T_m until allele m gets fixed has an asymptotic phase-type distribution

$$\hat{\mu}_{min} T_m \xrightarrow{\mathcal{L}} PD(\tilde{e}_0, \Sigma_0) \tag{12.49}$$

as $N \to \infty$, where Σ_0 contains the first m rows and m columns of the intensity matrix Σ in (12.28).

Remark 12.2 The limit result for $\tilde{T}_m$ is analogous to Theorem 12.2. One simply puts $s_m = \infty$ everywhere, which corresponds to immediate fixation of a type m mutation, once it appears.

12.7 Illustrating the Theory

In this section we will illustrate Theorem 12.2. Recall that it gives the asymptotic waiting time distribution until the m:th mutant gets fixed in the population, based on the transition rates $\hat{\lambda}_{ij}$ in (12.35)–(12.36) that approximate λ_{ij} in (12.15)–(12.16). In order to determine the approximate waiting time distribution, it suffices to specify $\hat{\lambda}_{ij}$ for $i = 0, \ldots, m-1$ and $j \neq i$, $j = 0, \ldots, m$, and then look at the properties of these rates as the population size grows. We will consider different scenarios, not all of which satisfy the regularity conditions of Theorem 12.2. But in these cases we will argue why (12.49) still provides a fairly accurate asymptotic approximation of the waiting distribution T_m. On the other hand, it is implicit that the mutation rates are smaller than the inverse population size, according to (12.4)–(12.5), for all examples of this section.

12.7.1 The Case of Two Coordinated Mutations

When there are $m = 2$ coordinated mutations, formula (12.35) simplifies to

$$\begin{gathered}\hat{\lambda}_{01} = N u_1 \beta(s_1), \quad \hat{\lambda}_{02} = N R(\rho) u_1 u_2^{1/2} \beta(s_2), \quad \hat{\lambda}_{10} = N v \beta(1/s_1), \\ \hat{\lambda}_{12} = N u_2 \beta(s_2/s_1),\end{gathered} \tag{12.50}$$

where $v = v_0$, and $\rho = \rho_{012}$ is a real-valued constant (not depending on N) defined in (12.37) that is either negative, zero or positive. Here, this equation simplifies to

$$s_1 = 1 + \rho u_2^{1/2}. \tag{12.51}$$

We will investigate the limit distribution of the waiting time T_2 for different asymptotic scenarios.

12.7.1.1 No Backward Mutations, and Final Allele Has High Fitness

In this subsection we will make two favorable assumptions for the waiting time T_2; that there are no backward mutations ($v = 0$), and that allele 2 has a high fitness ($s_2 = \infty$). It turns out that the relative size of the two forward mutation rates, u_1 and u_2, is crucial for the asymptotic properties of T_2. We will look at four different cases.

Case 1: Second mutation rate very small. Assume that

$$u_2 = o(u_1 N^{-1}) \tag{12.52}$$

as $N \to \infty$. In this case, the three nonzero rates in (12.50) simplify to

$$\hat{\lambda}_{01} \sim u_1, \quad \hat{\lambda}_{02} = NR(\rho)u_1 u_2^{1/2}, \quad \hat{\lambda}_{12} = Nu_2. \tag{12.53}$$

It follows from (12.4) to (12.52) that $\hat{\lambda}_{02} \ll \hat{\lambda}_{01}$ and $\hat{\lambda}_{12} \ll \hat{\lambda}_{01}$, so that $\hat{\mu}_0 \sim \hat{\lambda}_{01} \gg \hat{\mu}_1 = Nu_2$. The asymptotic states with short and long waiting times are $I_{\text{short}} = \{0\}$ and $I_{\text{long}} = \{1\}$ respectively, the time rate to absorption is $\hat{\mu}_{\min} = Nu_2$, the mutation rates on the new time scale are $\kappa_0 = \infty$ and $\kappa_1 = 1$, and the nonzero asymptotic transition probabilities from the non-asymptotic states, are $\pi_{01} = \pi_{12} = 1$. This gives a normalized intensity matrix

$$\Sigma = \begin{pmatrix} -\infty & \infty & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

in (12.28). Because of the smallness of the second mutation rate u_2, there is asymptotically no tunneling from 0 to 2, but allele 1 gets fixed at first. After that it takes much longer time for the first allele 2 to arrive, in spite of the fact that this $1 \to 2$ mutation is successful with probability 1 (since $s_2 = \infty$, and therefore $\beta(s_2) = 1$). Consequently, the asymptotic distribution of T_2 will be dominated by the waiting time for allele 2 to appear, after allele 1 has first been fixed in the population, i.e.

$$Nu_2 \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.54}$$

as $N \to \infty$. Notice that the exponential limit distribution in (12.54) is a special case of Corollary 12.1, although (12.52) violates regularity condition (12.46) of Theorem 12.2. However, this condition is only needed in order to get a good approximation of the tunneling rate $\hat{\lambda}_{02}$. But since stochastic tunneling $0 \to 2$ has no asymptotic impact ($\hat{\lambda}_{02} \ll \hat{\lambda}_{01}$), we still believe (12.54) is accurate.

Case 2: Second mutation rate small. If

$$u_1 N^{-1} \ll u_2 \ll N^{-2} \tag{12.55}$$

as $N \to \infty$, we get slightly different asymptotics compared to Case 1. The transition rates $\hat{\lambda}_{ij}$ are the same as in (12.53), but their asymptotic ordering $\hat{\lambda}_{02} \ll \hat{\lambda}_{01} \ll \hat{\lambda}_{12}$ is different. The states with short and long waiting time are therefore switched compared to Case 1 ($I_{\text{long}} = \{0\}$, $I_{\text{short}} = \{1\}$), with a time rate $\hat{\mu}_{\min} = \hat{\mu}_1 \sim \hat{\lambda}_{01} \sim u_1$ to absorption. The rescaled mutation rates from states 0 and 1 are $\kappa_0 = 1$ and $\kappa_1 = \infty$ respectively, whereas the nonzero asymptotic transition probabilities from the non-asymptotic states are the same as for Case 1 ($\pi_{01} = \pi_{12} = 1$). This gives a normalized intensity matrix

$$\Sigma = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -\infty & \infty \\ 0 & 0 & 0 \end{pmatrix}.$$

The second mutation rate u_2 in (12.55) is too small to allow for tunneling, but large enough to make the waiting time for allele 2 much shorter than the waiting time until allele 1 gets fixed at first. Notice that (12.55) allows for any of u_1 or u_2 to dominate asymptotically. In either case, the waiting time for allele 2 to fix is shorter, because of the selective advantage of this allele ($s_2 = \infty$). Therefore, the asymptotic distribution of T_2 will be dominated by the waiting time for allele 1 to fix, i.e.

$$u_1 \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.56}$$

as $N \to \infty$. This limit result is also special case of Corollary 12.1, and it agrees with Theorem 2 of Durrett and Schmidt [19].

Case 3: Second mutation rate of intermediate size. We assume that

$$u_2 = \frac{\gamma}{N^2} \tag{12.57}$$

for some constant γ as $N \to \infty$. The transition rates in (12.50) then simplify to

$$\hat{\lambda}_{01} \sim u_1 \eta(\rho \gamma^{1/2}), \quad \hat{\lambda}_{02} = R(\rho) \gamma^{1/2} u_1, \quad \hat{\lambda}_{12} = \gamma / N, \tag{12.58}$$

where $\eta(x)$ is an asymptotic approximation of $N\beta(1 + x/N)$. From formula (12.33) we deduce that

$$\eta(x) = \begin{cases} 1, & x = 0, \\ x/\left[1 - \exp(-x)\right], & x \neq 0. \end{cases}$$

It follows from (12.4) and (12.58) that $\hat{\lambda}_{0i} \ll \hat{\lambda}_{12}$ for $i = 1, 2$, whereas $\hat{\lambda}_{01}$ and $\hat{\lambda}_{02}$ have the same asymptotic order. Therefore, $I_{\text{long}} = \{0\}$, $I_{\text{short}} = \{1\}$, $\hat{\mu}_{\min} = \hat{\mu}_0 = \hat{\lambda}_{01} + \hat{\lambda}_{02} \ll \hat{\mu}_1 = \hat{\lambda}_{12}$, $\kappa_0 = 1$, and $\kappa_1 = \infty$. This gives an asymptotic rescaled intensity matrix of the form

$$\Sigma = \begin{pmatrix} -1 & \pi_{01} & \pi_{02} \\ 0 & -\infty & \infty \\ 0 & 0 & 0 \end{pmatrix}, \tag{12.59}$$

where $\pi_{02} = \frac{R(\rho)\gamma^{1/2}}{R(\rho)\gamma^{1/2}+\eta(\rho\gamma^{1/2})}$ is the asymptotic probability for tunneling to occur, and $\pi_{01} = 1 - \pi_{02}$ is the corresponding probability of no tunneling. Since the two transition rates from allele 0 are of similar size asymptotically, allele 2 will either get fixed directly through stochastic tunneling, or in two steps where allele 1 spreads in the population at first, and then almost immediately after that, allele 2 takes over. Formula (12.49) suggests that

$$\left[\eta(\rho\gamma^{1/2}) + R(\rho)\gamma^{1/2}\right] u_1 \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.60}$$

as $N \to \infty$. However, (12.60) is not correct since (12.44) is violated, that is, there is asymptotic competition between the two forward rates from allele 0, so that $F(0)$ does not exist. In order to see that (12.60) is wrong, consider the case when the intermediate allele 1 has the same selective advantage as allele 0 ($\rho = 0$). Then (12.60) simplifies to

$$(1 + \gamma^{1/2})u_1 \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.61}$$

as $N \to \infty$, since $\eta(0) = R(0) = 1$. But this is different from Theorem 3 of Durrett et al. [20], which states that

$$\chi(\gamma)u_1 \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.62}$$

as $N \to \infty$, where

$$\chi(\gamma) = \frac{\sum_{k=1}^{\infty} \frac{\gamma^k}{(k-1)!(k-1)!}}{\sum_{k=1}^{\infty} \frac{\gamma^k}{k!(k-1)!}}. \tag{12.63}$$

In order to quantify the difference between (12.61) and (12.62), we have plotted the ratio

$$\xi(\gamma) = \frac{1 + \gamma^{1/2}}{\chi(\gamma)} \tag{12.64}$$

of the two intensities in Fig. 12.1. It can be seen that the approximate intensity is always a bit larger than the exact one, with a maximum difference of 40%, although for most values of γ the difference is less than 20%. This implies that the approximate approach will underestimate the expected waiting time by up to 40%, since competition between the two fixation rates $0 \to 1$ and $0 \to 2$ is ignored. In Sect. 12.8 we will discuss a method that to some extent corrects for this.

Case 4: Second mutation rate large. Suppose

$$u_2 \gg N^{-2}, \tag{12.65}$$

so that the transition rates in (12.50) simplify to

$$\hat{\lambda}_{01} \sim N u_1 \psi(\rho u_2^{1/2}), \quad \hat{\lambda}_{02} = N R(\rho) u_1 u_2^{1/2}, \quad \hat{\lambda}_{12} = N u_2, \tag{12.66}$$

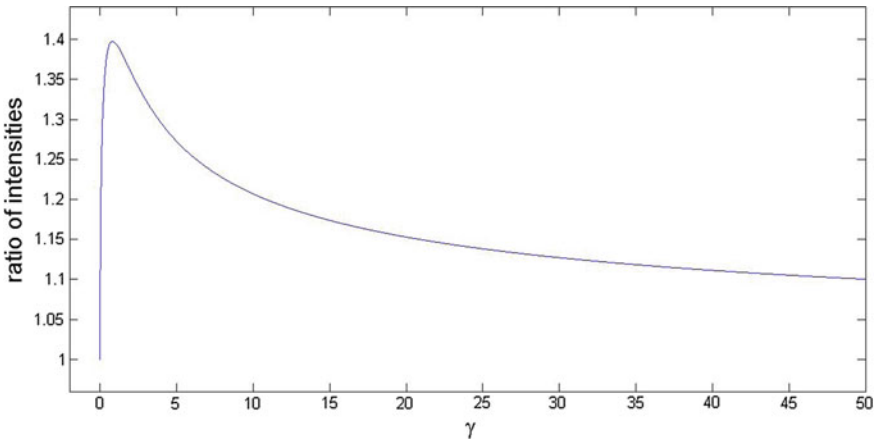

Fig. 12.1 Plot of the ratio $\xi(\gamma)$ between the approximate and exact asymptotic rates of the exponential limit distribution for the waiting time T_2 until the second mutation gets fixed in model with no backward mutations and neutral alleles ($s_1 = s_2 = 1$, i.e. $\rho = 0$ in (12.51)). The argument γ is the normalized rate (12.57) at which the second mutation occurs. It can be shown that $\xi(\gamma) > 1$ for all $\gamma > 0$, with $\xi(\gamma) \to 1$ as either $\gamma \to 0$ or $\gamma \to \infty$, and the maximum value $\xi(\gamma) = 1.40$ is attained for $\gamma = 0.82$

where

$$\psi(\rho u_2^{1/2}) = \begin{cases} 0, & \rho < 0, \\ 1/N, & \rho = 0, \\ \rho u_2^{1/2}, & \rho > 0, \end{cases}$$

relies on the asymptotic approximation of $\beta(s_1) = \beta(1 + \rho u_2^{1/2})$ defined in (12.33), when s_1, the selective fitness of allele 1, is given by (12.51). If follows from (12.4) that $\max(\hat{\lambda}_{01}, \hat{\lambda}_{02}) \ll \hat{\lambda}_{12}$ as $N \to \infty$, so that $I_{\text{long}} = \{0\}$. Regarding $\hat{\lambda}_{01}$ and $\hat{\lambda}_{02}$, their asymptotic ordering will depend on s_1. We have that $\hat{\lambda}_{01} \ll \hat{\lambda}_{02}$ if $\rho \leq 0$, whereas $\hat{\lambda}_{01}$ and $\hat{\lambda}_{02}$ are of the same order when $\rho > 0$. This means that 1 is an asymptotic state with a short waiting time when $\rho > 0$ ($I_{\text{short}} = \{1\}$), whereas it is a non-asymptotic state when $\rho \leq 0$ ($I_{\text{nas}} = \{1\}$). We follow the remark below Theorem 12.1 in (12.29), and let $\hat{\mu}_{\min} = N u_1 u_2^{1/2}$ be the asymptotic rate until allele m gets fixed in the population, which differs from μ_0 by a conveniently chosen constant. The normalized rates of leaving states 0 and 1, on the new time scale determined by $\hat{\mu}_{\min}$, are $\kappa_0 = 1(\rho > 0)\rho + R(\rho)$, $\kappa_1 = \infty$, where $1(A)$ is the indicator function for the event A (that is, it equals 1 if A occurs and 0 it if does not). The asymptotic probability of tunneling from 0 to 2, is $\pi_{02} = \frac{R(\rho)}{1(\rho>0)\rho+R(\rho)}$, and the other nonzero asymptotic transition probabilities from non-absorbing states, are $\pi_{01} = 1 - \pi_{02}$ and $\pi_{12} = 1$. This gives a rescaled intensity matrix Σ that equals (12.59). Therefore, when the mutation rate of allele 2 is large, as in (12.65), it will always become fixed in the population through tunneling $0 \to 2$ when allele 1 is selectively neutral or deleterious compared to allele 0 ($\rho \leq 0$). On the other hand, when allele 1 has higher fitness

than allele 0 ($\rho > 0$), it is possible to reach allele 2 either by tunneling, or by first having allele 1 fixed. In the latter case, the subsequent waiting time for allele 2 to spread is negligible. Formula (12.49) suggests a limit distribution

$$[1(\rho > 0)\rho + R(\rho)]\, N u_1 u_2^{1/2} \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.67}$$

as $N \to \infty$ for the total waiting time T_2 until allele 2 takes over the population. However, we only expect (12.67) to be correct when $\rho \leq 0$, since (12.44) is violated when $\rho > 0$, due to the competition between alleles 1 and 2 to take over the population at first. When $\rho \leq 0$, formula (12.67) agrees with Theorem 4 in Durrett and Schmidt [19]. In particular, when $\rho = 0$ we find that

$$N u_1 u_2^{1/2} \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.68}$$

as $N \to \infty$, since $R(0) = 1$. This agrees with a result given on pp. 231–232 of Nowak [52], and (12.68) is also a special case of Theorem 1 of Durrett et al. [20].

12.7.1.2 Backward Mutations, and Final Allele Is Neutral

In this subsection we will make three assumptions that increase the difficulty of having allele 2 fixed in the population, so that the waiting time T_2 gets longer compared to Sect. 12.7.1.1. First, we allow for backward mutations ($v > 0$), second, we assume that the fitness of the final allele 2 is the same as for allele 0 ($s_2 = 1$), and third, the intermediate allele 1 does not have a selective advantage in comparison to the other two alleles, so that $\rho \leq 0$ in (12.51). In order to avoid too many parameters of the model, we will also assume that the two forward mutation rates are identical, i.e., $u_1 = u_2 = u$. We will not consider the case when the forward mutation rate is small in comparison to the backward rate ($u = o(v)$), since (12.18) is violated then. Formally, the expected value of the limit distribution (12.27) is infinite when $u = o(v)$, since the asymptotic intensity matrix Σ_0 of fixation rates is not invertible. This is due to the fact that when backward mutations are frequent, they will effectively block the opportunities for allele 2 to spread to the whole population. In order to handle such a scenario we need to generalize Theorem 12.1 and let $\mu_{\min}^{-1}$ be determined by the asymptotic growth rate in (12.10). Here, we will therefore confine ourselves to scenarios where the backward mutation rate satisfies $v = Cu$ for some constant $C > 0$.

The above mentioned assumptions imply that the intensities (12.50) at which new alleles get fixed, simplify to

$$\begin{gathered}\hat{\lambda}_{01} = N u \beta (1 + \rho u^{1/2}),\ \ \hat{\lambda}_{02} = R(\rho) u^{3/2},\ \ \hat{\lambda}_{10} = NCu\beta(1 - \rho u^{1/2}), \\ \hat{\lambda}_{12} = N u \beta (1 - \rho u^{1/2}).\end{gathered} \tag{12.69}$$

It turns out that the asymptotic properties of T_2 depend on the size of the mutation rates, and we will look at three different scenarios.

Case 1. Small mutation rate. If

$$u = o(N^{-2}), \tag{12.70}$$

as $N \to \infty$, then (12.69) simplifies to

$$\hat{\lambda}_{01} \sim u, \quad \hat{\lambda}_{02} \sim R(\rho)u^{3/2}, \quad \hat{\lambda}_{10} \sim Cu, \quad \hat{\lambda}_{12} \sim u, \tag{12.71}$$

so that the tunneling rate $\hat{\lambda}_{02} \ll \min(\hat{\lambda}_{01}, \hat{\lambda}_{10}, \hat{\lambda}_{12})$ can be ignored, whereas the other three fixation rates $\hat{\lambda}_{01}, \hat{\lambda}_{10}$, and $\hat{\lambda}_{12}$ are of the same order. This implies that the two non-absorbing states are asymptotic, and they both contribute to the total waiting time ($I_{\text{long}} = \{0, 1\}$), with $\hat{\mu}_0 \sim \hat{\lambda}_{01} = u$ and $\hat{\mu}_1 = \hat{\lambda}_{10} + \hat{\lambda}_{12} = (C + 1)u$. Putting $\hat{\mu}_{\min} = u$, we find that the normalized rates of leaving states 0 and 1 are $\kappa_0 = 1$ and $\kappa_1 = C + 1$ respectively, whereas the nonzero asymptotic transition probabilities from the non-absorbing states are $\pi_{01} = 1, \pi_{10} = C/(C + 1)$, and $\pi_{12} = 1/(C + 1)$. This gives a matrix

$$\Sigma = \begin{pmatrix} -1 & 1 & 0 \\ C & -(C+1) & 1 \\ 0 & 0 & 0 \end{pmatrix} \tag{12.72}$$

of rescaled fixation rates. Formula (12.49) implies a limit distribution

$$u \times T_2 \xrightarrow{\mathcal{L}} \text{PD}((1, 0), \Sigma_0) \tag{12.73}$$

of the waiting time for allele 2 to take over the population as $N \to \infty$. In particular, without backward mutations ($C = 0$), we find that T_2 has an asymptotic gamma distribution

$$u \times T_2 \xrightarrow{\mathcal{L}} \Gamma(2, 1), \tag{12.74}$$

where 2 is the form parameter and 1 the intensity parameter. Since the form parameter is integer valued, the limit is also referred to as an Erlang distribution. Notice that (12.74) is a special case of Corollary 12.1, with $\kappa_0 = \kappa_1 = 1$.

Case 2. Intermediate sized mutation rate. Suppose $u = \frac{\gamma}{N^2}$ for some positive constant γ. The fixation intensities in (12.50) then simplify to

$$\begin{aligned} &\hat{\lambda}_{01} \sim u\rho\gamma^{1/2}/(1 - \exp(-\rho\gamma^{1/2})), \quad \hat{\lambda}_{02} = R(\rho)u^{3/2}, \\ &\hat{\lambda}_{10} \sim Cu\rho\gamma^{1/2}/(\exp(\rho\gamma^{1/2}) - 1), \quad \hat{\lambda}_{12} \sim u\rho\gamma^{1/2}/(\exp(\rho\gamma^{1/2}) - 1), \end{aligned} \tag{12.75}$$

as $N \to \infty$. The distribution of the waiting time T_2 turns out to be similar to Case 1, with the difference that the selection parameter ρ will have an asymptotic impact. As in Case 1, the tunneling from allele 0 to 2 can be ignored

($\hat{\lambda}_{02} \ll \min(\hat{\lambda}_{01}, \hat{\lambda}_{10}, \hat{\lambda}_{12})$), whereas the other three fixation rates $\hat{\lambda}_{01}$, $\hat{\lambda}_{10}$, and $\hat{\lambda}_{12}$ have the same order of magnitude. Therefore, both non-absorbing states are asymptotic, with a long waiting time ($I_{\text{long}} = \{0, 1\}$). Following the remark below Theorem 12.1, we standardize the time scale with an appropriately chosen constant, so that $\hat{\mu}_{\min} = u$ has a simple form. On the new time scale, the intensities to leave states 0 and 1 are $\kappa_0 = \rho\gamma^{1/2}/(1 - \exp(-\rho\gamma^{1/2}))$, $\kappa_1 = (C+1)\rho\gamma^{1/2}/(\exp(\rho\gamma^{1/2}) - 1)$ respectively. Since the nonzero transition probabilities π_{ij} of jumping between various fixation states, are the same as in Case 1, the rates of fixation between all pairs of states, after the time transformation, are

$$\Sigma = \begin{pmatrix} -\kappa_0 & \kappa_0 & 0 \\ \frac{C}{C+1}\kappa_1 & -\kappa_1 & \frac{1}{C+1}\kappa_1 \\ 0 & 0 & 0 \end{pmatrix}. \tag{12.76}$$

It follows from formula (12.49) that the asymptotic distribution for the total waiting time to reach allele 2, is given by

$$u \times T_2 \xrightarrow{\mathcal{L}} \text{PD}\left((1,0), \Sigma_0\right) \tag{12.77}$$

as $N \to \infty$. In particular, when there are no backward mutations ($C = 0$), (12.77) simplifies to

$$u \times T_2 \xrightarrow{\mathcal{L}} \kappa_0^{-1} X_0 + \kappa_1^{-1} X_1. \tag{12.78}$$

This is a special case of Corollary 12.1, with X_0 and X_1 two independent and exponentially distributed random variables with expected value 1. Notice also that Case 1 is essentially a $\rho \to 0$ limit of Case 2.

The expected value of the limit distribution of $u \times T_2$, on the right hand side of (12.77), has an explicit form. Using formula (12.10) for the expected value of a phase-type distribution, and putting $x = \rho\gamma^{1/2}$, we find that

$$\begin{aligned} u \times E(T_2) &\sim -(1,0)\Sigma_0^{-1}(1,1)^T \\ &= \begin{cases} 2 + C, & \rho = 0, \\ \left[(e^x - 1) + (1 - e^{-x})(1 + C)\right]/x, & \rho < 0, \end{cases} \end{aligned} \tag{12.79}$$

increases linearly with the backward rate C. In Fig. 12.2 we have plotted $u \times E(T_2)$ as a function of C for various values of the selection parameter ρ, and validated the accuracy of (12.79) with simulations.

Further details for the neutral case ($x = 0$) are given in Fig. 12.3, where the density function f_{T_2} of T_2 based on (12.9) is compared with simulation based histograms, for different values of C. Whereas f_{T_2} is gamma distributed for $C = 0$, it can be seen that its form approaches an exponential density as C grows.

Case 3. Large mutation rate. Assume that the mutation rate and the selection coefficient of allele 1 satisfy

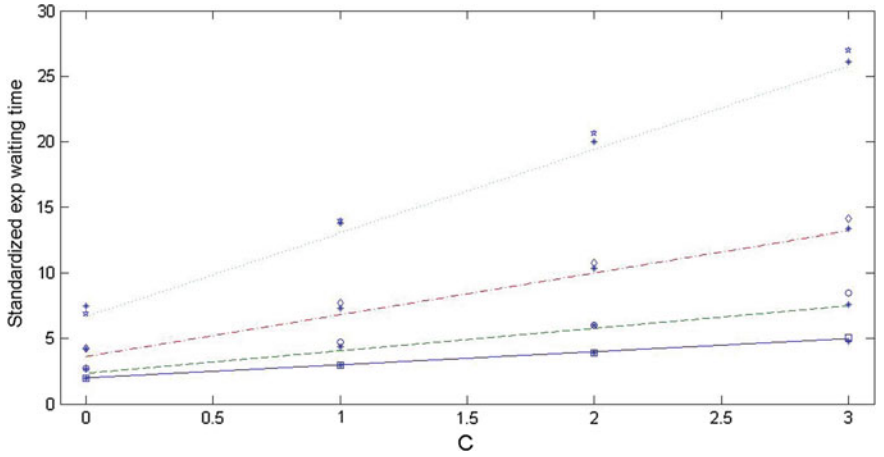

Fig. 12.2 Plot of the rescaled expected waiting time $u \times E(T_2)$, for a model with $m = 2$, forward mutation rates $u_1 = u_2 = u = \gamma/N^2$, and backward mutation rate $v_0 = Cu$. The lines are based on the approximate formula (12.79), and shown as functions of C. All lines have $s_2 = 1$, but the value of $s_1 = 1 + \rho\sqrt{u}$ varies. The intermediate allele is either neutral $\rho = 0$ (solid line), or has a selective disadvantage with $\rho = -1/\gamma^{1/2}$ (dashed line), $\rho = -2/\gamma^{1/2}$ (dash-dotted line), and $\rho = -3/\gamma^{1/2}$ (dotted line). Result from 1000 simulations, for a population of size $N = 100$ with $\gamma = 1$, are shown for $\rho = 0$ (squares), $\rho = -1/\gamma^{1/2}$ (circles), $\rho = -2/\gamma^{1/2}$ (diamonds), and $\rho = -3/\gamma^{1/2}$ (pentagrams). The parameters of the simulation algorithm are $N_c = 10$ and $\varepsilon = 0.2$ (see Appendix A). The simulation based estimates are also compared with the more accurate analytical solution (stars) based on (12.10) and (12.35)

$$\begin{aligned} &u \gg N^{-2}, \\ &\rho < 0, \end{aligned} \tag{12.80}$$

respectively. (If $\rho = 0$, it turns out that the asymptotics of T_2 is identical to Case 1.) The fixation rates in (12.50) then simplify to

$$\begin{aligned} \hat{\lambda}_{01} &= 0, \\ \hat{\lambda}_{02} &= R(\rho)u^{3/2}, \\ \hat{\lambda}_{10} &\sim -NC\rho u^{3/2}, \\ \hat{\lambda}_{12} &\sim -N\rho u^{3/2} \end{aligned} \tag{12.81}$$

as $N \to \infty$. We notice that $\hat{\lambda}_{01} \ll \hat{\lambda}_{02} \ll \min(\hat{\lambda}_{10}, \hat{\lambda}_{12})$. This implies that there is one asymptotic state $I_{\text{long}} = \{0\}$ with a long waiting time, and one non-asymptotic state $I_{\text{nas}} = \{1\}$. Time is therefore rescaled according to $\hat{\mu}_{\min} = \hat{\mu}_0 \sim \hat{\lambda}_{02} = R(\rho)u^{3/2}$, so that $\kappa_0 = 1$ and $\kappa_1 = \infty$ are the rescaled rates of leaving states 0 and 1, whereas $\pi_{02} = 1$, $\pi_{10} = C/(C+1)$, and $\pi_{12} = 1/(C+1)$ are values of the three nonzero transition probabilities from non-absorbing states. The matrix of standardized fixation rates is

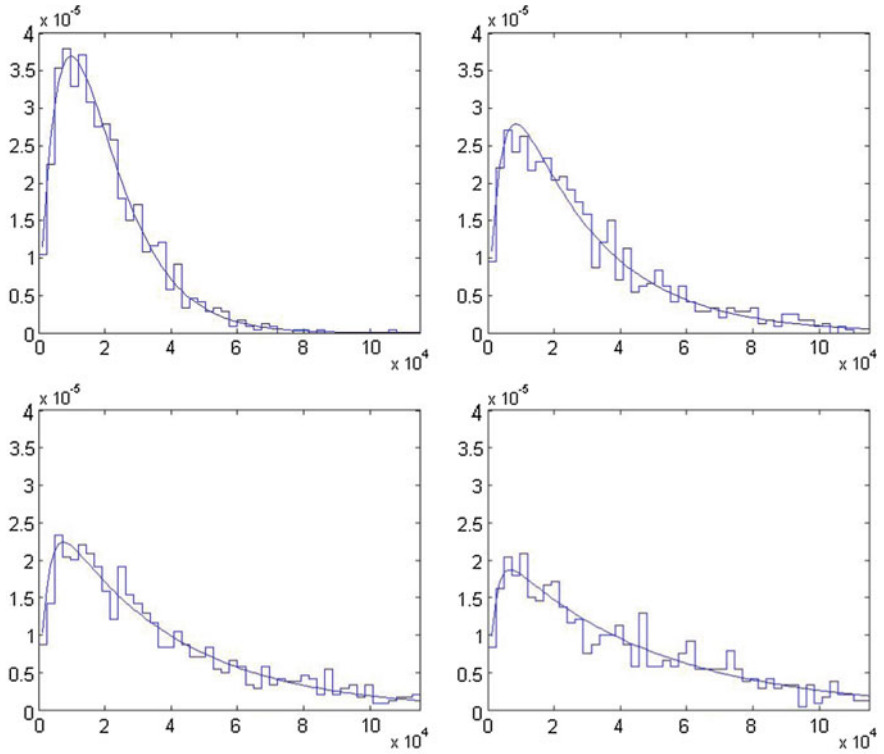

Fig. 12.3 Density functions (12.9) of the waiting time T_2 for a model with $N = 100$ individuals and $m = 2$ selectively neutral coordinated mutations ($s_1 = s_2 = 1$). The forward mutation rates are $u_1 = u_2 = 1/N^2$, whereas the backward mutation rates are $v_0 = C/N^2$ and $v_1 = 0$. The four graphs have $C = 0$ (upper left), $C = 1$ (upper right), $C = 2$ (lower left), and $C = 3$ (lower right), corresponding to the four simulations of Fig. 12.2 that are marked with squares. Shown in each plot is also a histogram from 1000 simulations, with parameters $N_c = 10$ and $\varepsilon = 0.2$ (see Appendix A). The estimated coefficients of variation $\sqrt{\text{Var}(T_2)}/E(T_2)$ from these four simulations are 0.704, 0.882, 0.941, and 0.965. This agrees well with the coefficients of variation of the density functions, which are $1/\sqrt{2} = 0.707$ for $C = 0$, and 1 in the limit as $C \to \infty$

$$\Sigma = \begin{pmatrix} -1 & 0 & 1 \\ C \times \infty & -(C+1) \times \infty & \infty \\ 0 & 0 & 0 \end{pmatrix}, \tag{12.82}$$

where ∞ of the second row should be interpreted as a limit. However, since the selective disadvantage of allele 1 is so large compared to alleles 0 and 2, allele 1 will never be fixed in a large population, and the second row of Σ will have no impact. Therefore, the only way of reaching allele 2 is through stochastic tunneling from allele 0. Formula (12.49) gives the limit distribution

$$R(\rho)u^{3/2} \times T_2 \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.83}$$

as $N \to \infty$ for the waiting time of allele 2 to get fixed. This is also a special case of Corollary 12.1.

12.7.2 Arbitrary Number of Coordinated Mutations

In this subsection we look at models with an arbitrary number m of coordinated mutations and number $m+1$ of alleles. We will consider two different kinds of models. The first one has no backward mutations, but the forward mutations have to arrive in a pre-specified order. The second model incorporates backward mutations, but the forward mutations may enter the population in any order.

12.7.2.1 Equal Forward Mutation Rates, No Backward Mutations

Assume there are no backward mutations ($v_0 = \cdots = v_{m-1} = 0$), and that forward mutations have to appear in a pre-determined order with identical mutation rates, i.e.

$$u_1 = \cdots = u_m = u. \tag{12.84}$$

We will also assume that all intermediate alleles are neutral or deleterious with the same selective fitness

$$s_1 = \cdots = s_{m-1} = s \le 1, \tag{12.85}$$

where

$$s = 1 + \rho u^{1-2^{-(m-1)}} \tag{12.86}$$

for some fixed constant $\rho \le 0$, not depending on N, and that the final allele m has a high fitness ($s_m = \infty$).

With these assumptions, formulas (12.35)–(12.36) for the fixation rates between different pairs of alleles simplify to $\hat{\lambda}_{ij} = 0$ when $j < i$, and to

$$\hat{\lambda}_{ij} \sim \begin{cases} NR(\rho)^{I(j=m)} u^{2-2^{-(j-1)}} \beta (1 + \rho u^{1-2^{-(m-1)}})^{I(j<m)}, & i = 0, \\ N u^{2-2^{-(j-i-1)}} (1/N)^{I(j<m)}, & i = 1, \ldots, m-1 \end{cases} \tag{12.87}$$

when $j > i$, where $I(\cdot)$ is the indicator function. In (12.87) we simplified the expressions for the terms $R(\rho_{ilj})$ on the right hand side of Eq. (12.36), for all i, l, j with $0 \le i < l < j \le m$. More specifically, we utilized that

$$\begin{aligned} \rho_{01m} &\sim \rho, \\ \rho_{0lj} &= o(1), (l, j) \neq (1, m), \\ \rho_{ilj} &= 0, \quad i = 1, \ldots, m-2, \end{aligned} \tag{12.88}$$

which, in view of (12.41), implies $R(\rho_{01m}) \sim R(\rho)$ and $R(\rho_{ilj}) \sim 1$ for the other terms of (12.36). In order to motivate (12.88), we use formulas (12.37) and (12.85)–(12.86) to find that

$$\begin{aligned} &1 + \rho_{ilj} u^{1/2} r_{ilj}^{1/2} = \tfrac{s_l}{s_i} \\ &= \begin{cases} 1 + \rho u^{1-2^{-(m-1)}}, & i = 0,\ l = 1, \ldots, j-1, \\ 1, & i = 1, \ldots, m-2,\ l = i+1, \ldots, j-1. \end{cases} \end{aligned} \tag{12.89}$$

When $i > 0$, (12.88) follows immediately from (12.89). When $i = 0$, we find that

$$\rho_{ilj} = \rho u^{1/2 - 2^{-(m-1)}} r_{ilj}^{-1/2}, \tag{12.90}$$

and therefore we also need to find expressions for r_{ilj}. To this end, we use formula (12.143) of Appendix C to deduce

$$\begin{aligned} r_{0lj} &= \Theta(u^{1-2^{-(j-l-1)}}), \\ r_{01m} &\sim u^{1-2^{-(m-2)}}. \end{aligned} \tag{12.91}$$

Then we insert (12.91) into (12.90) and use formula (12.4) to notice that $u \to 0$ as $N \to \infty$, in order to prove the upper two equations of (12.88).

Having established formula (12.87) for the transition rates between different fixed states, we will next investigate which jumps from state $i \leq m-2$ that are possible when N gets large. It follows from (12.87) that the transition rates from i to the intermediate states $i+1, \ldots, m-1$ are related as

$$\hat{\lambda}_{i,i+1} \gg \hat{\lambda}_{i,i+2} \gg \cdots \gg \hat{\lambda}_{i,m-1} \tag{12.92}$$

when $N \to \infty$, and therefore it is not possible to have a direct transition from state i to any of $j = i+2, \ldots, m-1$. This can also be deduced directly from formula (12.45). A transition from i to $i+2 \leq j < m$ is not possible asymptotically, since s_j is not larger than $\max(s_{i+1}, \ldots, s_{j-1})$.

Consequently, it is only possible for a population with allele i, to transfer either to a population with $i+1$ alleles, or to one in which the final allele m has been fixed. Therefore, the rate of leaving state i is of the order

$$\begin{aligned} \hat{\mu}_i &= \textstyle\sum_{j=i+1}^{m} \hat{\lambda}_{ij} \\ &= \Theta\left(\max(\hat{\lambda}_{i,i+1}, \hat{\lambda}_{im})\right) \end{aligned} \tag{12.93}$$

Table 12.1 Some possible scenarios for m coordinated mutations when all mutation rates (12.84) are identical, and the selective fitness (12.85) is the same for all alleles $i < m$. The dots indicate successive transitions $i \to i+1$ between neighboring alleles

Case	Scenario	Transitions	Mutation rate u
3	0	$0 \to m$	Large
2	$1 \le n \le m-2$	$0 \to \cdots \to n \to m$	Intermediate
1	$m-1$	$0 \to \cdots \to m$	Small

as $N \to \infty$. The asymptotic properties of the waiting time T_m until allele m gets fixed, will depend on which of the rates on the right hand side of (12.93) that dominate as $N \to \infty$ among the asymptotic states. We will consider m different scenarios, numbered as $n = 0, \ldots, m-1$, where Scenario n is characterized by a set

$$I_{\text{as}} = \{0, \ldots, n\} \tag{12.94}$$

of asymptotic states. These scenarios can be divided into three groups, depending on the size of the mutation rate u (see Table 12.1). As a general rule, the larger the mutation rate is, the earlier stochastic tunneling will kick in and drive the population towards its final state, where allele m has been fixed.

Case 1. Small mutation rate. Suppose

$$u = o(N^{-2}) \tag{12.95}$$

as $N \to \infty$, so that the rates of fixation between different pairs of alleles in (12.87) simplify to

$$\hat{\lambda}_{ij} \sim NR(\rho)^{I(i=0,j=m)} u^{2-2^{-(j-i-1)}} (1/N)^{I(j<m)}, \tag{12.96}$$

for $i = 0, \ldots, m-1$ and $i < j \le m$. We used (12.33) to simplify the fixation probability in (12.87) to

$$\beta(1 + \rho u^{1-2^{-(m-1)}}) \sim N^{-1}, \tag{12.97}$$

since (12.95) implies $u^{1-2^{-(m-1)}} = o(N^{-1})$.

Recall from (12.93) that asymptotically, we only have to consider transitions from i to $i+1$ and m. We deduce from formula (12.96) that the rates of leaving the non-absorbing states are $\hat{\mu}_i \sim \hat{\lambda}_{i,i+1} = u$ for $i = 0, 1, \ldots, m-2$, and $\hat{\mu}_{m-1} = \hat{\lambda}_{m-1,m} = Nu$. This corresponds to Scenario $m-1$ in (12.94), but only the first $m-1$ asymptotic states will contribute to the overall waiting time, so that $I_{\text{long}} = \{0, \ldots, m-2\}$ and $I_{\text{short}} = \{m-1\}$. Rescaling time by a factor $\hat{\mu}_{\min} = u$, we find that the normalized rates of leaving state i are $\kappa_i = 1$ for $i \in I_{\text{long}}$ and $\kappa_{m-1} = \infty$. Since the nonzero asymptotic transition probabilities for jumps from non-absorbing allelic states are $\pi_{i,i+1} = 1$ for $i = 0, \ldots, m-1$, the conditions of Corollary 12.1 are satisfied. It follows from formula (12.31) that

$$u \times T_m \xrightarrow{\mathcal{L}} \Gamma(m-1,1) \tag{12.98}$$

as $N \to \infty$, in agreement with Case 1 in Theorem 2 of Schweinsberg [60].

Case 2. Intermediate mutation rate. Let $n \in \{1, \ldots, m-2\}$ be a fixed number, and assume that the mutation rate has size

$$N^{-1/(1-2^{-(m-n-1)})} \ll u \ll N^{-1/(1-2^{-(m-n)})} \tag{12.99}$$

as $N \to \infty$. The transition rates $\hat{\lambda}_{ij}$ and the fixation probability satisfy (12.96) and (12.97), also for Case 2. It follows that the rate at which allele i is lost in a large population is $\hat{\mu}_i \sim \hat{\lambda}_{i,i+1} = u$ for $i = 0, \ldots, n-1$, and $\hat{\mu}_n \sim \hat{\lambda}_{nm} = Nu^{2-2^{-(m-n-1)}} \gg u$ for $i = n$. The nonzero asymptotic transition probabilities from non-absorbing states are therefore $\pi_{i,i+1} = 1$ for $i = 0, \ldots, n-1$, and $\pi_{nm} = 1$. This corresponds to Scenario n in (12.94), where the first n asymptotic states contribute to the overall waiting time ($I_{\text{long}} = \{0, \ldots, n-1\}$), the remaining asymptotic state n does not ($I_{\text{short}} = \{n\}$), and the other non-absorbing states are non-asymptotic ($I_{\text{nas}} = \{n+1, \ldots, m-1\}$). If time is standardized by $\hat{\mu}_{\min} = u$, the rescaled rates of leaving state i are $\kappa_i = 1$ for $i \in I_{\text{long}}$, and $\kappa_n = \infty$. Since the conditions of Corollary 12.1 are satisfied, we apply formula (12.31) and deduce an asymptotic distribution

$$u \times T_m \xrightarrow{\mathcal{L}} \Gamma(n,1) \tag{12.100}$$

as $N \to \infty$ for the waiting time until allele m gets fixed in the population. This corresponds to Case 2 of Theorem 2 in Schweinsberg [60].

Case 3. Large mutation rate. Assume that

$$N^{-1/(1-2^{-(m-1)})} \ll u \ll N^{-1/(1-2^{-m})} \tag{12.101}$$

as $N \to \infty$. It can be seen that the fixation probability satisfies (12.97) when $\rho = 0$, whereas $\beta(1+\rho u^{1-2^{-(m-1)}}) = o(N^{-1})$ when $\rho < 0$. In any case, it follows from (12.87) that $\hat{\lambda}_{01} \le u \ll \hat{\lambda}_{0m} = NR(\rho)u^{2-2^{-(m-1)}}$. Transitions from 0 will therefore be to state m when N is large, so that the rate of leaving state 0 is of the order $\hat{\mu}_0 \sim \hat{\lambda}_{0m}$. This corresponds to Scenario 0 in (12.94), with one single state $I_{\text{long}} = \{0\}$ that contributes to the waiting time T_m asymptotically, and since $\pi_{0m} = 1$, all other non-absorbing states are non-asymptotic ($I_{\text{nas}} = \{1, \ldots, m-1\}$). With $\hat{\mu}_{\min} = \hat{\mu}_0$, the normalized rate of leaving state 0 is $\kappa_0 = 1$. Since Corollary 12.1 is satisfied, we deduce from (12.31) that the waiting time for the m:th mutant to get fixed, has a limiting distribution

$$NR(\rho)u^{2-2^{-(m-1)}} \times T_m \xrightarrow{\mathcal{L}} \text{Exp}(1) \tag{12.102}$$

as $N \to \infty$. This results generalizes Case 3 of Theorem 2 in Schweinsberg [60] from the neutral case $\rho = 0$ to $\rho \le 0$.

Boundary scenarios. As in Schweinsberg [60], it is possible to consider $m-1$ additional asymptotic scenarios for the mutation rate, which can be interpreted as boundaries between the m scenarios of Table 12.1. For simplicity, we confine ourselves to the neutral case $\rho = 0$. Suppose $n \in \{0, \ldots, m-2\}$. As a boundary between scenarios n and $n+1$ of Table 12.1, we assume that the mutation rate satisfies

$$u = \frac{\gamma}{N^{-1/(1-2^{-(m-n-1)})}} \tag{12.103}$$

as $N \to \infty$, for some constant $\gamma > 0$. In this case the population dynamics starts with n fixation events $0 \to 1 \to \cdots \to n$. Then in the next step there is competition between fixation $n \to n+1$ and tunneling $n \to m$. If $n+1$ gets fixed, in the next step there will be a much faster transition $n+1 \to m$ that does not contribute to the overall waiting time. Therefore, among the asymptotic states $I_{\text{as}} = \{0, \ldots, n+1\}$, only those in $I_{\text{long}} = \{0, \ldots, n\}$ contribute asymptotically to T_m.

In more detail, combining the arguments for Cases 1–3 above, it can be seen that the rates of leaving state i is $\hat{\mu}_i \sim \hat{\lambda}_{i,i+1} = u$ for $i = 0, \ldots, n-1$, whereas $\hat{\mu}_n \sim \hat{\lambda}_{n,n+1} + \hat{\lambda}_{nm} = u + Nu^{2-2^{-(m-n-1)}} = (1 + \gamma^{1-2^{-(m-n-1)}})u$ for state n. We then transform the time scale by $\hat{\mu}_{\min} = u$, and find that the normalized rates are $\kappa_i = 1$ of leaving states $i = 0, \ldots, n-1$, $\kappa_n = 1 + \gamma^{1-2^{-(m-n-1)}}$ to leave state n, and it is $\kappa_{n+1} = \infty$ to leave state $n+1$. Therefore, Theorem 12.2 suggests a limit distribution

$$u \times T_m \xrightarrow{\mathcal{L}} X_0 + \cdots X_{n-1} + \frac{1}{1 + \gamma^{1-2^{-(m-n-1)}}} X_n \tag{12.104}$$

as $N \to \infty$, for the waiting time until allele m gets fixed, where $X_0, \ldots, X_n$ are independent random variables with an identical distribution that is exponential with expected value 1. However, the limit distribution in (12.104) is incorrect. The reason is that regularity condition (12.44) is violated for transitions from state n. Asymptotically it is possible to either have a transition $n \to n+1$ or stochastic tunneling $n \to m$, and therefore $\pi_{n,n+1}$ and π_{nm} are both positive. The correct limit distribution for T_m is given in Theorem 3 of Schweinsberg [60]. It states that

$$u \times T_m \xrightarrow{\mathcal{L}} X_0 + \cdots X_{n-1} + \frac{1}{\chi(\gamma^{2(1-2^{-(m-n-1)})})} X_n \tag{12.105}$$

as $N \to \infty$, with $\chi(\cdot)$ defined in (12.63). We notice that (12.104)–(12.105) generalize (12.61)–(12.62), which corresponds to the special case $n = 0$ and $m = 2$. The approximate limit distribution in (12.104) has a slightly lower expected value than the correct one in (12.105), and their ratio will depend on n and $\gamma' = \gamma^{2(1-2^{-(m-n-1)})}$. In Table 12.2 we have displayed the maximal possible ratio between expected values of the correct and approximate limit distribution, as a function of n. It can be seen that this ratio quickly approaches 1 as n grows. For most values of γ' (or γ), the ratio will be even closer to 1. See also Sect. 12.8, were we introduce a method that to some extent corrects for the different expected waiting times of (12.104) and (12.105).

Table 12.2 The table refers to a model with no backward mutations and m forward mutations with equal rate ($= u$) that satisfies (12.103) for some $\gamma > 0$ and $0 \le n \le m-2$, so that a direct transition $n \to n+1$ and tunneling $n \to m$ are both possible. Displayed is the maximal possible ratio between the expected values of the correct and approximate limit distributions of the time until allele m gets fixed, in (12.105) and (12.104) respectively, as a function of the number n of transitions $0 \to \cdots \to n$ without any tunneling. The maximum ratio in the table is attained for a value of γ that depends on n. When $n = 0$, it equals the maximum of the function that is plotted in Fig. 12.1

n	Maximal ratio
0	1.398
1	1.143
2	1.088
3	1.064
4	1.050

12.7.2.2 Forward and Backward Mutations in Any Order

When forward and backward mutations are allowed to arrive in any given order, it is reasonable to identify type i with the number of mutations that have appeared in the population so far. Suppose that u and $v = Cu$ are the rates at which each single forward and backward mutation arrive. When i mutants have been fixed in the population, there are $m - i$ additional forward mutations not present in the population, and i possible types of back mutations. Consequently, $u_{i+1} = (m-i)u$, $v_{i-1} = Ciu$, for $i = 0, \ldots, m-1$. We will also assume a neutral model, so that $s_1 = \cdots = s_m = 1$. Then formulas (12.35)–(12.36) simplify to

$$\hat{\lambda}_{ij} = \begin{cases} \prod_{l=i}^{j-1} [(m-l)u]^{2^{-(l-i)}}, & j > i, \\ \prod_{l=j+1}^{i} (Clu)^{2^{-(i-l)}}, & j < i. \end{cases} \tag{12.106}$$

Since the model is neutral, the tunneling condition in (12.45) is violated for all pairs i, j of states. We may therefore disregard the possibility of tunneling, asymptotically as $N \to \infty$, so that the rate of leaving state i is of the order

$$\hat{\mu}_i \sim \hat{\lambda}_{i,i-1} + \hat{\lambda}_{i,i+1} = [m + (C-1)i]\, u,$$

for $i = 0, \ldots, m-1$. Since all $\hat{\mu}_i$ have the same asymptotic order, it follows that all non-absorbing states are asymptotic with a long waiting time ($I_{\text{long}} = \{0, \ldots, m-1\}$). It is convenient to transform the time scale by $\hat{\mu}_{\min} = u$, so that the asymptotic rescaled intensity matrix in (12.28) has elements

$$\Sigma_{ij} = \begin{cases} m-i, & j = i+1, \\ Ci, & j = i-1, \\ 0, & |j-i| \ge 2, \\ -[m + (C-1)i], & j = i, \end{cases} \tag{12.107}$$

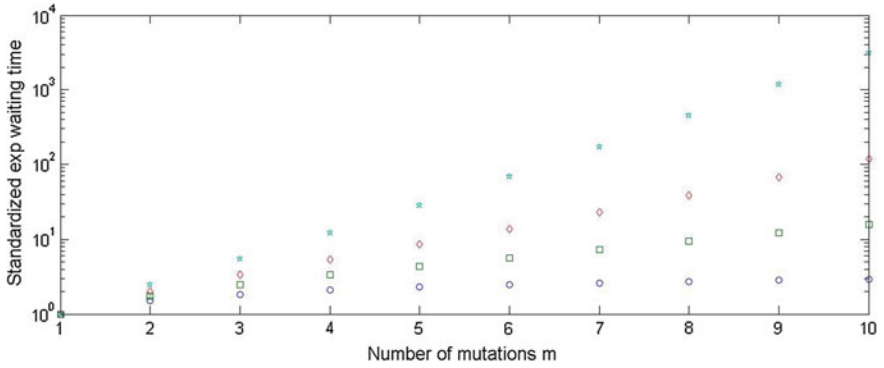

Fig. 12.4 Plot of standardized asymptotic expected waiting time $u \times E(T_m)$, according to formula (12.108), as a function of the number of required mutations m. The forward and backward mutations may appear in any order. The forward mutation rate per allele is u, and the symbols correspond to different rates $v = Cu$ of backward mutations per allele, with circles ($C = 0$), squares ($C = 0.5$), diamonds ($C = 1$), and pentagrams ($C = 2$)

for the rows that correspond to non-absorbing states ($i < m$). Combining (12.10) and (12.49), we find that the expected waiting time is given by

$$E(T_m) \sim \tilde{e}_0 \Sigma_0^{-1} 1 \times u^{-1}, \tag{12.108}$$

asymptotically as $N \to \infty$. In Fig. 12.4 we plotted the expected waiting time in (12.108) as a function of m, for different values of C. While $E(T_m)$ increases quite slowly with m in absence of backward mutations, there is a dramatic increase of $E(T_m)$ for positive C as the number m of required mutations increases. In Appendix C we derive an explicit formula for the asymptotic approximation (12.108) of $E(T_m)$. It follows from this derivation that (12.108) can be approximated by the simpler but somewhat less accurate expression

$$u \times E(T_m) \sim \begin{cases} \log(m) + 0.577, & C = 0, \\ (1 + C)^m/(Cm), & C > 0, \end{cases} \tag{12.109}$$

when C is fixed and m gets large. Formula (12.109) underscores the staircase behavior of the expected waiting time with increasing m when $C > 0$. This behavior would be even more dramatic if the intermediate states had a selective disadvantage ($s_i < 1$ for $i = 1, \ldots, m - 1$), cf. Figure 2 of Axe [2].

12.8 Some Improvements of the Asymptotic Waiting Time Theory

The practical implication of Theorem 12.2 is to approximate the distribution of the waiting time T_m until the mth mutant gets fixed. We expect this distribution to be accurate for large populations with mutation rates (12.4)–(12.5) smaller than the

inverse population size. Second, according to (12.44) there should not be competition between different alleles to get fixed. That is, for any allele i there should be at most one forward rate of fixation from i, and at most one backward rate of fixation from i, that dominate. Third, the time it takes for alleles to get fixed should be asymptotically negligible because of (12.47). In this section we will highlight some possible improvement of formula (12.10) for the expected waiting time $E(T_m)$ based on transition rates (12.35), when some of these conditions fail. Our discussion is not at all complete, but we hope it will open up for further research.

We will first revisit Case 4 of Sect. 12.7.1.1, that is, a model with $m = 2$ mutants and a large second forward mutation rate u_2. We will see what happens when the first forward mutation rate u_1 is no longer of smaller order than the inverse population size. The following result, which generalizes Theorem 1 of Durrett et al. [20], is proved in Appendix C:

Theorem 12.3 *Consider a Moran model with $m = 2$ and no backward mutations ($v_0 = 0$), where the sizes of the two forward mutation rates satisfy $Nu_1 \to a$ for some $a \geq 0$ and $N\sqrt{u_2} \to \infty$ as $N \to \infty$. Assume that the first selection coefficient $s_1 = s$ is given by (12.51) for some fixed $\rho \leq 0$, and the second one is large ($s_2 = \infty$). Let also T_2'' be the time point when the first successful mutant 2 appears in the population. Then*

$$P(NR(\rho)u_1\sqrt{u_2} \times T_2'' \geq t) \sim \exp\left(-\int_0^t h(x)dx\right) \tag{12.110}$$

as $N \to \infty$, where $R(\rho)$ is defined in (12.41), $h(x) = h(x; a, \rho)$ is a hazard function that satisfies $h(x) = 1$ when $a = 0$, and

$$h(x) = \frac{1 - \exp\left(-\frac{2\sqrt{\rho^2+4}}{\rho+\sqrt{\rho^2+4}} \times \frac{x}{a}\right)}{1 + \frac{\sqrt{\rho^2+4}+\rho}{\sqrt{\rho^2+4}-\rho} \exp\left(-\frac{2\sqrt{\rho^2+4}}{\rho+\sqrt{\rho^2+4}} \times \frac{x}{a}\right)} \tag{12.111}$$

when $a > 0$. In particular, the expected waiting time is approximated by

$$E(T_2'') \sim \left[NR(\rho)u_1\sqrt{u_2}\right]^{-1} \theta(a, \rho) = \hat{\lambda}_{02}^{-1}\theta(a, \rho), \tag{12.112}$$

where $\hat{\lambda}_{02}$ is the transition rate defined in (12.66), and

$$\theta(a, \rho) = \int_0^\infty \exp\left(-\int_0^t h(x; a, \rho)dx\right) dt.$$

In Theorems 12.1 and 12.2, we imposed conditions so that the time of tunneling and fixation were asymptotically negligible. Theorem 12.3 reveals that this is no longer the case when $u_1 = \Theta(N^{-1})$, since the waiting time T_2'' includes two parts of comparable size; the time T_2' until the first successful $0 \to 1$ mutation appears, *and* the time $T_2'' - T_2'$ of tunneling, that is, the time between the arrival of the first

successful 1 mutant and the first successful 2 mutant. It follows from the proof of Theorem 12.3 that the time T_2' until the first successful 1 mutant appears has an asymptotic exponential distribution with expected value $E(T_2') \sim \hat{\lambda}_{02}^{-1}$. Therefore, in view of (12.112), we find that tunneling *multiplies* the expected waiting time by a factor $\theta(a, \rho)$. On the other hand, we recall from Sect. 12.5 that the time it takes for allele 2 to become fixed after its first appearance, *adds* a term $\alpha(s_2) \sim E(T_2 - T_2'')$ to the expected waiting time $E(T_2) = E(T_2'') + E(T_2 - T_2'')$.

We will apply these findings as follows: Let $\hat{\lambda}_{ij}$ be the approximate fixation rates in (12.35). When $j \neq i$ we modify these rates as

$$\tilde{\lambda}_{ij} = \begin{cases} \left[\hat{\lambda}_{ij}^{-1} + \alpha(s_j/s_i)\right]^{-1}, & j \neq i, |j-i| \neq 2, \\ \left[\hat{\lambda}_{ij}^{-1}\theta(a_{i,i+1}, \rho_{i,i+1}) + \alpha(s_{i+2}/s_i)\right]^{-1}, & j = i+2, \\ \left[\hat{\lambda}_{ij}^{-1}\theta(a_{i,i-1}, \rho_{i,i-1}) + \alpha(s_{i-2}/s_i)\right]^{-1}, & j = i-2, \end{cases} \tag{12.113}$$

to take tunneling and fixation into account, where $a_{i,i+1} = Nu_{i+1}\beta(s_{i+2}/s_i)$ and $a_{i,i-1} = Nv_{i-1}\beta(s_{i-2}/s_i)$ are the size normalized rates at which new mutations appear *and* get fixed, conditionally on that tunneling is successful, whereas $s_{i+1}/s_i = 1 + \rho_{i,i+1}\sqrt{u_{i+2}}$ and $s_{i-1}/s_i = 1 + \rho_{i,i-1}\sqrt{v_{i-2}}$ are special cases of (12.37) and (12.39) for tunneling over one allele ($|j-i| = 2$). When $i = j$, we define $\tilde{\lambda}_{ii}$ so that all row sums of the matrix with elements $\tilde{\lambda}_{ij}$, are zero. The modified transition rates in (12.113) only incorporate the impact of tunneling over one allele, because it is more complicated to correct for tunneling over larger distances, and it is likely that this has less impact in many applications.

As a next step, we will correct for competition between different states to become fixed. We will confine ourselves to the case when all mutants except the last one are selectively neutral ($s_1 = \cdots = s_{m-1} = 1$), and the last mutant has a selective advantage ($s_m > 1$). It follows from this and the discussion above (12.45) that it is only possible to have competition between fixation events $i \to i+1$ and $i \to m$ for a population whose current fixed state is i. We therefore compare these two transition rates, as defined in (12.113), and denote their squared ratio by $\gamma_i = \left(\frac{\tilde{\lambda}_{im}}{\tilde{\lambda}_{i,i+1}}\right)^2$. In Appendix C we motivate that the forward transition rates in (12.113) should be modified as

$$\bar{\lambda}_{ij} = \frac{\chi\left[\gamma_i/\beta(s_m)\right]}{1+\sqrt{\gamma_i}} \times \tilde{\lambda}_{ij}, \quad j = i+1, \ldots, m, \tag{12.114}$$

where $\chi(\gamma)$ was introduced in (12.63). We put $\bar{\lambda}_{ij} = \tilde{\lambda}_{ij}$ when $j < i$, whereas the diagonal terms $\bar{\lambda}_{ii}$ are chosen so that all row sums of the matrix $\bar{\Lambda} = (\bar{\lambda}_{ij})$, are zero. When $s_m = \infty$ (so that $\beta(s_m) = 1$), we notice that the multiplicative correction factor of (12.114) is $\xi(\gamma)^{-1}$, where $\xi(\gamma)$ is the function defined in (12.64) and plotted in Fig. 12.1. Therefore, when the last mutant has high fitness, this figure tells how much the expected waiting time of a forward fixation from i will increase when competition

between fixed states $i+1$ and m is taken into account. This also agrees with formula (12.105).

Putting everything together, we define the adjusted expected waiting time as

$$E(T_m)_{\text{adj}} = \tilde{e}_0 \bar{\Lambda}_0^{-1} 1^T, \tag{12.115}$$

where $\bar{\Lambda}_0$ is a matrix containing the first m rows and the first m columns of $\bar{\Lambda}$, and adj is an acronym for adjusted. We regard (12.115) as the expected time until a semi-Markov process of allele frequencies Z_t with state space (12.6) reaches the absorbing state e_m. By this we mean that jumps between fixed states follow a Markov chain with a transition probability $-\bar{\lambda}_{ij}/\bar{\lambda}_{ii}$ from e_i to e_j. But the holding time in each state is no longer exponentially distributed, when tunneling and the time of fixation of alleles is taken into account. Although the time until a semi-Markov processes reaches an absorbing state does not have a phase-type distribution, if $-\bar{\lambda}_{ii}^{-1}$ is the expected holding time in fixed state i, formula (12.115) will still give the correct expected waiting time until the m:th mutant gets fixed.

12.8.1 One Mutation

In this subsection we consider a model with only one mutant ($m = 1$). Formulas (12.10) and (12.35) approximate the expected waiting time until fixation as

$$E(T_1) = \frac{1}{Nu_1\beta(s_1)}. \tag{12.116}$$

For a model with only two alleles, there is no tunneling and no competition between different states to become fixed. It is therefore only the expected time of a successful mutation to get fixed, that will influence the adjusted waiting time formula (12.115). It can be seen that this equation simplifies to

$$E(T_1)_{\text{adj}} = \frac{1}{Nu_1\beta(s_1)} + \alpha(s_1), \tag{12.117}$$

for a model with one single mutant. In Table 12.3, we have compared the accuracy of (12.116) and (12.117) with simulation based estimates of the expected waiting time. It can be seen that (12.117) is consistently a much more accurate approximation of the simulation based values. We also notice from this table that the smaller the mutation rate is, the smaller is the impact of the expected fixation time $\alpha(s_1)$. The general condition for asymptotic negligibility of the fixation time is (12.47). It simplifies to $\alpha(s_1) \ll [Nu_1\beta(s_1)]^{-1}$ for a model with one mutant, that is, scenarios for which the second term of (12.117) is small in comparison to the first term.

Table 12.3 Comparison between the expected waiting time formulas $E(T_1)$ and $E(T_1)_{\text{adj}}$, defined in (12.116) and (12.117) respectively, for a model with $m = 1$ mutant. The rightmost column are sample averages from 10000 simulations, with $\varepsilon = 0.04$ and $N_c = 10$ in the algorithm of Appendix A

N	u_1	s_1	$E(T_1)$	$E(T_1)_{\text{adj}}$	$\hat{E}(T_1)$
100	0.001	10/9	100.00	153.19	152.62
		2	20.00	33.95	33.36
		5	12.50	20.53	20.05
		9	11.25	18.21	17.50
		1000	10.01	15.89	15.34
1000	0.0001	10/9	100.00	198.78	197.76
		2	20.00	40.90	40.12
		5	12.50	23.99	23.52
		9	11.25	21.10	20.33
		1000	10.01	18.20	17.55
100	0.0005	10/9	200.00	253.19	250.33
		2	40.00	53.95	53.11
		5	25.00	33.03	32.72
		9	22.50	29.46	28.74
		1000	20.02	25.90	25.03
100	0.0001	10/9	1000.0	1053.2	1049.15
		2	200.0	213.95	211.63
		5	125.0	133.03	131.97
		9	112.5	119.46	117.30
		1000	100.1	105.98	105.64

12.8.2 Two Coordinated Mutations, and No Back Mutations

Here, we will revisit the model of Sect. 12.7.1.1, with two mutants ($m = 2$) and no back mutations ($v_0 = 0$). We assume that the first allele is selectively neutral ($s_1 = 1$), whereas the second one has high fitness ($s_2 = 10^5$).

In Table 12.4 we compare the accuracy of two analytical formulas for the expected waiting time until the second mutant gets fixed, with simulation based estimates. We follow the scenarios of Durrett and Schmidt [19], with different population sizes and forward mutation rates u_1 and u_2. The first expected waiting time formula is based on (12.10) and (12.35), whereas the second formula is the adjustment defined in (12.115).

It can be seen from Table 12.4 that the unadjusted expected waiting time is too low, whereas the adjusted expected waiting time is consistently much closer to the simulation based estimates. The reason for this discrepancy varies between scenarios. For those scenarios where Nu_1 is not small (Case 1–2 and Drosophila), an important feature of the adjusted formula is to incorporate the time it takes for the first successful

Table 12.4 Comparison between the expected waiting time formula $E(T_2)$ based on (12.10) and (12.35), and the adjusted waiting time formula $E(T_2)_{\text{adj}}$, based on (12.115), for a model with $m = 2$ mutants with selective fitness $s_1 = 1$ and $s_2 = 10^5$. We use the same scenarios as in Table 12.2 of Durrett and Schmidt [17], with different values of the two forward mutation rates u_1 and u_2, and no backward mutations. The quantity $\hat{E}(T_2)$ refers to a sample average from B simulations based on the algorithm of Appendix A, with the first simulation parameter ε reported in the rightmost column, and the second simulation parameter N_c set to 10

Scenario	N	Nu_1	$N\sqrt{u_2}$	$E(T_2)$	$E(T_2)_{\text{adj}}$	$\hat{E}(T_2)$	B	ε
Case 1	1000	1	10	92.6	163.2	166.9	10000	0.04
	10000			919.0	1557.8	1588	10000	0.2
Case 2	1000	1/4	10	367.7	463.8	470.5	10000	0.04
	10000			3649	4564	4644	2000	0.1
Case 3	1000	1/10	10	917.9	1051	1074	10000	0.1
	10000			9108	10434	10143	1000	0.1
Case 4	1000	1/10	4	2018	2523	2484	5000	0.1
	10000			20131	25150	26420	1000	0.2
Case 5	1000	1/10	1	5501	8053	8240	2000	0.1
	10000			55002	80471	85288	1000	0.2
Drosophila	1000	1/2	$10/\sqrt{3}$	301.0	449.2	460.5	10000	0.04
	10000			2998	4417	4440	2000	0.1

allele 1 to tunnel into allele 2. For those scenarios where $\lambda_{02}/\lambda_{01} = N\sqrt{u_2}$ is not large (Case 4–5 and Drosophila), an important fact is rather that the adjusted formula incorporates competition between alleles 1 and 2 to get fixed. On the other hand, it is not crucial, for any of the scenarios of Table 12.4 to correct for the time it takes for alleles to get fixed. This has two reasons. First, the expected fixation time of allele 2 is very short ($\alpha(10^5) \sim \log(N)$). Although the expected fixation time of allele 1 is much larger ($\alpha(1) \sim N$), for those scenarios where a transition $0 \to 1$ happens fairly often (that is, when $N\sqrt{u_2}$ is not too large, as for Case 5), the overall expected waiting time is still much larger than $\alpha(1)$.

12.9 Discussion

In this paper we analyzed the waiting time until the last of m mutations appears and gets fixed in a population of constant size without any substructure. We showed that approximately, this waiting time has a phase-type distribution whenever a fixed state model is applicable, where one genetic variant at a time dominates the population. The rationale behind this result is to approximate the dynamics of the genetic composition of the population by a continuous time Markov process with $m + 1$ states; the wildtype variant and the m mutants. We also provided a general scheme for calculating the intensity matrix of this process, and thereby obtained an explicit approximation of the waiting time distribution. Our model allows for forward and

backward mutations, with different selective fitness, to appear at different rates. Once the intensity matrix of the Markov process is known, the phase-type distribution of the waiting time automatically incorporates all the pathways towards the m:th mutant that the model allows for.

We believe the findings of this paper can be extended in several ways. First, we have provided quite a detailed sketch of proofs of main results, derived previously known results as special cases, and confirmed several others by simulations. While it is outside the scope of this paper to provide full proofs; this is an important topic for further research.

Second, we argued that our explicit approximation of the expected waiting time has the correct order of magnitude, even when some of the assumptions behind the intensity rate calculations are violated. In more detail, the transition rates between pairs of fixed states will only be correct when competition between different forward and backward transitions can be neglected asymptotically. We provided an adjustment of the expected waiting time for neutral models when such competition is present, using Theorem 3 of Schweinsberg [60] and Theorem 3 of Durrett et al. [20]. A challenging task is to generalize these results to scenarios where the mutants of the model have different selective fitness.

Third, we have assumed a homogeneous population of haploid individuals with constant size. We believe our main results can be extended to include varying population size, diploidy, and recombination, as well as geographic subdivision and other types of population structure.

Fourth, in some applications there are several possible orders in which the m mutations may arrive. This can still be handled by a fixed state population model, with a phase-type distribution for the waiting time, as in Sect. 12.7.2.2. But for some scenarios of partially ordered mutations, the state space of the Markov process has to be enlarged in order to keep track of the subset of mutations that has occurred (Gerstung and Beerenwinkel [24]).

Fifth, a challenging generalization is to derive a phase-type distribution approximation of the waiting time until m coordinated targets have been fixed in the population. For instance, each type $i \in \{1, \ldots, m\}$ could represent a sequence of DNA, which, compared to previous targets $j < i$, requires one or several additional point mutations. This would extend results in Durrett and Schmidt [18] from $m = 1$ to higher values of m.

Sixth, the results of this paper could serve as a building block in order to understand the genomewide rate of molecular evolution of m coordinated mutations. In order to obtain such a rate, a selection model has to be specified, whereby the selection coefficients of the m mutants at various loci are drawn from some multivariate distribution. This can viewed as an extension of the simulation studies in Gillespie [27] and Rupe and Sanford [55] for single mutations ($m = 1$) to larger m.

Seventh, our phase-type distribution approximation of the waiting time relies heavily on the assumption that all mutation rates are smaller than the inverse population size, in order to guarantee that successful mutations arrive so infrequently and then spread so quickly that one genetic variant at a time dominates. While this is a reasonable assumption for moderately-sized populations, it is not appropriate for

large populations where different mutations will coexist, interfere, and overlap. This includes virus, bacterial or simple eukaryotic populations, as well as large cell populations of cancer progression with diverse mutational patterns. While we adjusted for non-small mutation rates for some of these models in Sect. 12.8, it is still important to derive more general results for the waiting time of coordinated mutations in large populations. Several papers have addressed this issue, see for instance Iwasa et al. [33], Desai and Fisher [16], Beerenwinkel et al. [4], Gerstung and Beerenwinkel [24], Theorem 4 of Schweinsberg [60], and Theorem 1 of Durrett et al. [20]. It is an interesting topic of future research to generalize these results to our setting of forward and backward mutations, where the mutants have a varying selective fitness.

Finally, we have developed analytical and simulation based tools in Matlab for the waiting time of coordinated mutations, based on the results of this paper. They are freely available from the first author upon request.

Acknowledgements The authors wish to thank an anonymous reviewer for several helpful suggestions that improved the clarity and presentation of the paper.

Appendix A. A Simulation Algorithm

Recall from Sect. 12.2 that the allele frequency process Z_t of the Moran model is a continuous time and piecewise constant Markov process with exponentially distributed holding times at each state $z = (z_0, \ldots, z_m) \in \mathcal{Z}$. For all but very small population sizes, it is infeasible to simulate this process directly, since the distances between subsequent jumps are very small, of size $O_p(N^{-1})$. The τ-leaping algorithm was introduced (Gillespie [25], Li [42]) in order to speed up computations for a certain class of continuous time Markov processes. It is an approximate simulation algorithm with time increments of size τ. According to the leaping condition of Cao et al. [12], one chooses $\tau = \tau(\varepsilon)$ in such a way that

$$E\left[|Z_{t+\tau,i} - Z_{ti}||Z_{ti} = z_i\right] \leq \varepsilon z_i \tag{12.118}$$

for $i = 0, \ldots, m$ and some fixed, small number $\varepsilon > 0$, typically in a range between 0.01 and 0.1.

Zhu et al. [71] pointed out that it is not appropriate to use τ-leaping for the Moran model when very small allele frequencies are updated. For this reason they defined a hybrid algorithm that combines features of exact simulation and τ-leaping. Although most time increments are of length τ, some critical ones are shorter. Then they showed that (12.118) will be satisfied by the hybrid algorithm for a neutral model with small mutation rates, when

$$\tau \leq \varepsilon/2. \tag{12.119}$$

We will extend the method of Zhu et al. [71] to our setting, where forward and backward mutations are possible. In order to describe the simulation algorithm, we

first need to define the transition rates of the Moran model. From any state $z \in \mathcal{Z}$, there are at most $(m+1)m$ jumps $z \to z + \delta_{ij}/N$ possible, where $\delta_{ij} = e_j - e_i$, $0 \le i, j \le m$ and $i \ne j$. Each such change corresponds to an event where a type i individual dies and gets replaced by a another one of type j. Since the process remains unchanged when $i = j$, we need not include these events in the simulation algorithm. It follows from Sect. 12.2 that the transition rate from z to $z + \delta_{ij}/N$ is

$$\begin{aligned} a_{ij} &= a_{ij}(z) \\ &= z_i \times \tfrac{z_j s_j}{\sum_{k=0}^m z_k s_k}(1 - u_{j+1} - v_{j-1}) + z_i \times \tfrac{z_{j-1} s_{j-1}}{\sum_{k=0}^m z_k s_k} u_j + z_i \times \tfrac{z_{j+1} s_{j+1}}{\sum_{k=0}^m z_k s_k} v_j \\ &= \tfrac{z_i}{\sum_{k=0}^m z_k s_k}\left[z_j s_j (1 - u_{j+1} - v_{j-1}) + z_{j-1} s_{j-1} u_j + z_{j+1} s_{j+1} v_j\right], \end{aligned} \tag{12.120}$$

with $u_{m+1} = v_{-1} = z_{-1} = z_{m+1} = 0$. Let N_c be a threshold. For any given state z, define the non-critical set Ω of events as those pairs (i, j) with $i \ne j$ such that both of z_i and z_j exceed N_c/N. The remaining events (i, j) are referred to as critical, since at least one of z_i and z_j is N_c/N or smaller. The idea of the hybrid simulation method is to simulate updates of critical events exactly, whereas non-critical events are updated approximately. In more detail, the algorithm is defined as follows:

1. Set $t = 0$ and $Z_t = e_0 = z$.
2. Compute the $m(m+1)$ transition rates $a_{ij} = a_{ij}(z)$ for $0 \le i, j \le m$ and $i \ne j$.
3. Compute the set $\Omega = \Omega(z)$ of critical events for the current state z.
4. Determine the exponentially distributed waiting time $e \overset{\mathcal{L}}{\in} \mathrm{Exp}(a)$ until the next critical event occurs, where $a = \sum_{(i,j) \notin \Omega} a_{ij}$ is the rate of the exponential distribution.
5. If $e < \tau$, simulate a critical event $(I, J) \notin \Omega$ from the probability distribution $\{a_{ij}/a;\ (i, j) \notin \Omega\}$, and update the allele frequency vector as $z \leftarrow z + \delta_{IJ}/N$. Otherwise, if $e \ge \tau$, simulate no critical event and leave z intact.
6. Let $h = \min(e, \tau)$. Then simulate non-critical events over a time interval of length h, and increment the allele frequency vector as

$$z \leftarrow z + \frac{1}{N} \sum_{(i,j) \in \Omega} n_{ij} \delta_{ij},$$

   where $n_{ij} \sim \mathrm{Po}(a_{ij} h)$ are independent and Poisson distributed random variables.
7. Update date time ($t \leftarrow t + h$) and the allele frequency process ($Z_t \leftarrow z$).
8. If $z = e_m$, set $T_m = t$ and stop. Otherwise go back to step 2.

We have implemented the hybrid algorithm, with N_c and ε as input parameters and $\tau = \varepsilon/2$. When the selection coefficients s_i are highly variable, a smaller value of τ is needed though in order to guarantee that (12.118) holds.

Appendix B. The Expected Waiting Time for One Mutation

In this appendix we will motivate formula (12.34). It approximates the expected number of generations $\alpha(s)$ until a single mutant with fitness s spreads and get fixed in a population where the remaining $N-1$ individuals have fitness 1, given that such a fixation will happen and that no further mutations occur. This corresponds to a Moran model of Sect. 12.2 with $m=1$ mutant, zero mutation rates ($u_1=v_0=0$), and initial allele frequency distribution $Z_0=(1-p,p)$, where $p=1/N$. For simplicity of notation we write $Z_t=Z_{t1}$ for the frequency of the mutant allele 1.

Kimura and Ohta [38] derived a diffusion approximation of $\alpha(s)$, for a general class of models. It involves the infinitesimal mean and variance functions $M(z)$ and $V(z)$ of the allele frequency process, defined through

$$\begin{aligned} E(Z_{t+h}|Z_t=z) &= z+M(z)h+o(h),\\ \mathrm{Var}(Z_{t+h}|Z_t=z) &= V(z)h+o(h) \end{aligned}$$

as $h\to 0$. In order to apply their formula to a mutation-free Moran model, we first need to find $M(z)$ and $V(z)$. To this end, suppose $Z_t=z$. Then use formula (12.120) with $m=1$ to deduce that

$$z\to z+1/N \text{ at rate } a_{01}(z)=N(1-z)\frac{zs}{1-z+zs}, \tag{12.121}$$

whereas

$$z\to z-1/N \text{ at rate } a_{10}(z)=Nz\frac{1-z}{1-z+zs}. \tag{12.122}$$

From this it follows that

$$M(z)=\frac{1}{N}\left[a_{01}(z)-a_{10}(z)\right]=(s-1)\frac{(1-z)z}{1+z(s-1)} \tag{12.123}$$

and

$$V(z)=\frac{1}{N^2}\left[a_{01}(z)+a_{10}(z)\right]=\frac{1}{N}(1+s)\frac{(1-z)z}{1+z(s-1)}. \tag{12.124}$$

We will also need the function

$$G(z)=\exp\left(-\int_0^z\frac{2M(y)}{V(y)}dy\right)=\exp(-2Ns'z),$$

with $s'=(s-1)/(s+1)$. The formula of Kimura and Ohta [38] takes the form

$$\alpha(s)=\int_p^1\psi(z)\hat{\beta}(z)\left[1-\hat{\beta}(z)\right]dz+\frac{1-\hat{\beta}(p)}{\hat{\beta}(p)}\int_0^p\psi(z)\hat{\beta}^2(z)dz, \tag{12.125}$$

where

$$\hat{\beta}(z) = \hat{\beta}(s; z) = \frac{\int_0^z G(y)dy}{\int_0^1 G(y)dy} = \frac{1 - e^{-2Ns'z}}{1 - e^{-2Ns'}} \tag{12.126}$$

approximates the fixation probability of a mutant allele that starts at frequency $Z_0 = z$. In particular, $\hat{\beta}(1/N)$ approximates the exact probability (12.32) that one single copy of an allele with fitness s takes over a population where all other individuals have fitness 1. This diffusion approximation is increasingly accurate in the limit of weak selection ($s \to 1$).

The other function of the two integrands in (12.125), is

$$\psi(z) = \frac{2\int_0^1 G(y)dy}{V(z)G(z)} = \frac{1 - e^{-2Ns'}}{e^{-2Ns'z}} \times \frac{1 + z(s-1)}{1+s} \times \frac{1}{s'z(1-z)}. \tag{12.127}$$

In order to verify (12.34) we will approximate (12.125) separately for neutral ($s = 1$), advantageous ($s > 1$), and deleterious ($s < 1$) alleles. In the neutral case $s = 1$ we let $s' \to 0$ and find that $\hat{\beta}(z) = z$ and $\psi(z) = N/[z(1-z)]$. Inserting these functions into (12.125), we obtain an expression

$$\alpha(1) = -\frac{1}{p}\left[N(1-p)\log(1-p)\right]$$

for the expected fixation time. This is essentially the middle part of (12.34) when $p = 1/N$.

When $s > 1$, we similarly insert (12.126)–(12.127) into (12.125). After some quite long calculations, it can be shown that

$$\begin{aligned}\alpha(s) \sim{}& \frac{1+s}{s-1}\log(N) \\ &+ \frac{s}{s-1}\left[\log(2s') + \int_0^1 \frac{1-e^{-y}}{y}dy - \int_1^\infty \frac{e^{-y}}{y}dy - \frac{1}{s}\int_{2s'}^\infty \frac{e^{-y}}{y}dy\right] \\ &+ \frac{e^{-2s'}}{1-e^{-2s'}} \times \frac{1}{s-1}\int_0^{2s'} \frac{1}{y}e^y(1-e^{-y})^2 dy\end{aligned} \tag{12.128}$$

as $N \to \infty$. The first term of this expression dominates for large N, and it agrees with the lower part of (12.34).

When $s < 1$, a similar calculation yields

Table 12.5 Approximations of the expected waiting time $\alpha(s) = \alpha_N(s)$ of fixation, in units of generations, for a single mutant with selection coefficient s, in a population of size N. The columns marked Diff are based on the diffusion approximation (12.125), whereas the columns marked AsDiff are asymptotic approximations of the diffusion solution, based on the middle part of (12.34) for $s = 1$, Eq. (12.128) for $s > 1$ and Eq. (12.129) for $s < 1$. The latter two formulas only work well when $|s - 1| \gg 1/N$. They have been omitted when they depart from the diffusion solution by more than 10%

s	$N = 100$		$N = 1000$		$N = 10000$	
	Diff	AsDiff	Diff	AsDiff	Diff	AsDiff
1/5	7.38	7.39	10.84	10.85	14.30	14.30
1/2	13.62	13.67	20.57	20.58	27.48	27.48
1/1.5	20.60	20.73	32.23	32.24	43.76	43.76
1/1.1	56.42	58.92	107.06	107.27	155.61	155.63
1/1.01	98.15	–	554.51	577.34	1038.1	1040.2
1/1.001	99.48	–	985.94	–	5535.0	5710.7
1	99.50	100.00	999.50	1000.0	9999.5	10000.0
1.001	99.48	–	985.94	–	5535.0	5710.8
1.01	98.16	–	554.52	577.35	1038.1	1040.2
1.1	56.47	58.97	107.11	107.32	155.66	155.68
1.5	20.80	10.93	32.43	32.44	43.96	43.96
2	13.95	14.00	20.90	20.91	27.81	27.82
5	8.03	8.04	11.49	11.50	14.95	14.95

$$\begin{aligned}\alpha(s) \sim\ & \frac{1+s}{1-s}\log(N) \\ & + \frac{s}{1-s}\left[\log(2s'') + \int_0^1 \frac{1-e^{-y}}{y}dy - \int_1^\infty \frac{e^{-y}}{y}dy - \frac{1}{s}\int_{2s''}^\infty \frac{e^{-y}}{y}dy\right] \\ & + \frac{e^{-2s''}}{1-e^{-2s''}} \times \frac{1}{1-s}\int_0^{2s''} \frac{1}{y}e^y(1-e^{-y})^2 dy\end{aligned} \tag{12.129}$$

as $N \to \infty$, with $s'' = (1-s)/(s+1)$. The first, leading term of this formula is consistent with the upper part of (12.34). The various approximations of $\alpha(s)$ are shown in Table 12.5.

Appendix C. Sketch of Proofs of Main Results

Lemma 12.1 *Let $\{\tau_k\}_{k=0}^M$ be the fixation times of the process Z_t, defined in (12.13), and τ'_{k+1} the time points when a successful mutation first occurs between two successive fixation events ($\tau_k < \tau'_{k+1} < \tau_{k+1}$). Let also μ_i be the rate in (12.15) at which successful mutations appear in a homogeneous type i population. Then*

$$P\left(\tau'_{k+1} - \tau_k > \frac{\zeta}{\mu_i} | Z_{t_k} = e_i\right) \to \exp(-\zeta) \tag{12.130}$$

as $N \to \infty$ *for all* $\zeta > 0$ *and* $i = 0, 1, \ldots, m-1$.

Sketch of proof. Let $f_i(z) = f_{i,N}(z)$ and $b_i(z) = b_{i,N}(z)$ be the probabilities that the offspring of a type $i \in \{0, \ldots, m-1\}$ individual who mutates to $i+1$ or $i-1$ is a successful forward or backward mutation, given that the allele frequency configuration is z just before replacement occurs with the individual that dies (when $i = 0$ we put $b_0(z) = 0$). Notice in particular that $f_i = f_i(e_i)$ and $b_i = b_i(e_i)$, since these two quantities are defined as the probabilities of a successful forward or backward mutation in an environment where all individuals have type i just before the mutation, that is, when $z = e_i$.

When an individual is born in a population with allele configuration z, with probability $1 - u_{i+1} f_i(z) - v_{i-1} b_i(z)$ it is not the first successful mutation between two fixation events τ_k and τ_{k+1}, given that no other successful has occurred between these two time points. Let $0 \le t_1 < t_2 < \cdots$ be the time points when a type i individual gets an offspring, and if we choose $\{Z_t\}$ to be left-continuous, the probability of no successful mutation $i \to i \pm 1$ at time t_l, where $\tau_k < t_l < \tau_{k+1}$, is $1 - u_{i+1} f_i(Z_{t_l}) - v_{i-1} b_i(Z_{t_l})$, given that no other successful mutation has occurred so far ($\tau'_{k+1} \ge t_l$). Since the left hand side of (12.130) is the probability of no mutation $i \to i \pm 1$ being successful among those that arrive at some time point in $\mathbb{T}_i(\zeta) = \{t_l;\ \tau_k < t_l \le \tau_k + \zeta/\mu_i\}$, we find that

$$\begin{aligned} &P(\tau'_{k+1} - \tau_k > \zeta/\mu_i | Z_{\tau_k} = e_i) \\ &= E\left[\prod_{t_l \in \mathbb{T}_i(\zeta)} \left(1 - u_{i+1} f_i(Z_{t_l}) - v_{i-1} b_i(Z_{t_l})\right)\right] \\ &\approx E\left[\exp\left(-u_{i+1} \sum_{t_l \in \mathbb{T}_i(\zeta)} f_i(Z_{t_l}) - v_{i-1} \sum_{t_l \in \mathbb{T}_i(\zeta)} b_i(Z_{t_l})\right)\right], \end{aligned} \tag{12.131}$$

where expectation is with respect to variations in the allele frequency process Z_t for $t \in \mathbb{T}_i(\zeta)$.

Because of (12.4)–(12.5), with a probability tending to 1 as $N \to \infty$, Z_t will stay close to e_i most of the time in (τ_k, τ'_{k+1}), that is, all alleles $l \ne i$ will most of the time be kept at low frequencies. In order to motivate this, we notice that by definition, all mutations that arrive in (τ_k, τ'_{k+1}) are unsuccessful. It is known that the expected lifetime of an unsuccessful mutations is bounded by $C \log(N)$ for a fairly large class of Moran models with selection, where C is a constant that depends on the model parameters, but not on N (Crow and Kimura [15], Section 8.9). Since mutations arrive at rate $N(v_{i-1} + u_{i+1})$, this suggest that all alleles $l \ne i$ are expected to have low frequency before the first successful mutation arrives, if

$$C \log(N) \times N(v_{i-1} + u_{i+1}) = o(1)$$

as $N \to \infty$, i.e. if the convergence rate towards zero in (12.4)–(12.5) is faster than logarithmic. This implies that it is possible to approximate the sums on the right hand sides of (12.131) by

$$\begin{array}{l} \sum_{t_l \in \mathbb{T}_i(\zeta)} f_i(Z_{t_l}) \approx f_i |\mathbb{T}_i(\zeta)| \approx f_i N \times \zeta/\mu_i, \\ \sum_{t_l \in \mathbb{T}_i(\zeta)} b_i(Z_{t_l}) \approx b_i |\mathbb{T}_i(\zeta)| \approx b_i N \times \zeta/\mu_i, \end{array} \tag{12.132}$$

where $|\mathbb{T}_i(\zeta)|$ refers to the number of elements in $\mathbb{T}_i(\zeta)$. In the first step of (12.132), we used that $f_i(z) \to f_i$ and $b_i(z) \to b_i$ as $z \to e_i$ respectively, and therefore $f_i(Z_{t_l}) \approx f_i$ and $b_i(Z_{t_l}) \approx b_i$ for most of the terms in (12.132). In the second step of (12.132) we used that $|\mathbb{T}_i(\zeta)|$ counts the number of births of type i individuals within a time interval of length ζ/μ_i, and that each $t_{l+1} - t_l$ is approximately exponentially distributed. By the definition of the Moran model in Sect. 12.2, the intensity of this exponential distribution is approximately

$$N \times \frac{Z_{t_l i} s_i}{\sum_{j=0}^{m} Z_{t_l j} s_j} \approx N,$$

for the majority of time points t_l such that Z_{t_l} stays close to e_i. Consequently, $|\mathbb{T}_i(\zeta)|$ is approximately Poisson distributed with expected value $N\zeta/\mu_i$. We know from (12.4)–(12.5) and (12.15) that $\mu_i = o(1)$. Because this implies that $N\zeta/\mu_i \gg 1$ is large, and since the coefficient of variation of a Poisson distribution tends to zero when its expected value increases, $|\mathbb{T}_i(\zeta)|/(N\zeta/\mu_i)$ converges to 1 in probability as $N \to \infty$, and therefore we approximate $|\mathbb{T}_i(\zeta)|$ by $N\zeta/\mu_i$. To conclude; (12.130) follows from (12.15), (12.131), and (12.132). □

Proof of Theorem 12.1. Let $X_\zeta = Z_{\zeta/\mu_{\min}}$ denote the allele frequency process after changing time scale by a factor $\mu_{\min}$. Let $S_k = \mu_{\min}\tau_k$ refer to time points of fixation when $\{X_\zeta\}$ visits new fixed states in $\mathcal{Z}_{\text{hom}}$, defined in (12.6), $S'_{k+1} = \mu_{\min}\tau'_{k+1}$ the time point when a successful mutation first appears after S_k, and $S = \mu_{\min}T_m = S_M$ the time when allele m gets fixed. We need to show that

$$S \xrightarrow{\mathcal{L}} \text{PD}(\tilde{e}_0, \Sigma_0) \text{ as } N \to \infty. \tag{12.133}$$

To this end, write

$$S = \sum_{k=0}^{M-1}(S'_{k+1} - S_k) + \sum_{k=1}^{M}(S_k - S'_k) =: S_{\text{appear}} + S_{\text{tunfix}}, \tag{12.134}$$

where S_{appear} is the total waiting time for new successful mutations to appear, and S_{tunfix} is the total waiting time for tunneling and fixation, after successful mutations have appeared. We will first show that

$$S_{\text{appear}} \xrightarrow{\mathcal{L}} \text{PD}(\tilde{e}_0, \Sigma_0) \text{ as } N \to \infty. \tag{12.135}$$

It follows from (12.14) to (12.17) that $\{X_{S_k}\}$ is a Markov chain that starts at $X_{S_0} = e_0$, with transition probabilities

$$P(X_{S_{k+1}} = e_j | X_{S_k} = e_i) = p_{ij,N} \to \pi_{ij} \quad \text{for } i = 0, \ldots, m-1,\ j \neq i. \tag{12.136}$$

Because of (12.25) and Lemma 12.1, the waiting times for successful mutations $i \to i \pm 1$ have exponential or degenerate limit distributions as $N \to \infty$, since

$$P(S'_{k+1} - S_k > \zeta | X_{S_k} = e_i) \to \begin{cases} \exp(-\kappa_i \zeta), & i \in I_{\text{long}}, \\ 0, & i \in I_{\text{short}}, \end{cases} \tag{12.137}$$

where I_{long} and I_{short} refer to those asymptotic states in (12.22) and (12.23) that are visited for a long and short time, respectively. Since by definition, the non-asymptotic states $i \in I_{\text{nas}}$ in (12.20) will have no contribution to the limit distribution of S_{appear} as $N \to \infty$, it follows from (12.136) to (12.137) that asymptotically, S_{appear} is the total waiting time for a continuous time Markov chain with intensity matrix Σ, that starts at e_0, before it reaches its absorbing state e_m. This proves (12.135).

It remains to prove that S_{tunfix} is asymptotically negligible. It follows from (12.26) that

$$P(\varepsilon) = P_N(\varepsilon) = \max_{i \in I_{\text{as}}} P\left(S_k - S'_k > \varepsilon | X_{S_{k-1}} = e_i\right) = o(1) \tag{12.138}$$

as $N \to \infty$ for any $\varepsilon > 0$. Write $M = \sum_{i=0}^{m-1} M_i$, where M_i is the number of visits to e_i by the Markov chain $\{X_{S_k};\ k = 0, \ldots, M\}$, before it is stopped at time M. Let K be a large positive integer. We find that

$$\begin{aligned} P(S_{\text{tunfix}} > \varepsilon) &\leq E\left[\textstyle\sum_{k=1}^{\min(K,M)} P(S_k - S'_k > \varepsilon/K)\right] + P(M > K) \\ &\leq K P(\varepsilon/K) + \textstyle\sum_{i \in I_{\text{nas}}} P(M_i > 0) + E(M)/K \\ &\leq 2E(M)/K \end{aligned} \tag{12.139}$$

for all sufficiently large N. In the second step of (12.139) we used that

$$E(M) = \tilde{e}_0 (I - P_0)^{-1} 1^T \to \tilde{e}_0 (I - \Pi_0)^{-1} 1^T < \infty, \tag{12.140}$$

where P_0 is a square matrix of order m that contains the first m rows and m columns of the transition matrix P of the Markov chain X_{S_k}, so that its elements are the transition probabilities among and from the non-absorbing states. We used in (12.140) that M is the number of jumps until this Markov chain reaches its absorbing state, and therefore it has a discrete phase-type distribution (Bobbio et al. [9]). And because of (12.17)–(12.18), the expected value of M must be finite. In the last step of (12.139) we used (12.138) and the definition of non-asymptotic states, which implies $P(M_i > 0) = o(1)$ for all $i \in I_{\text{nas}}$.

Since (12.139) holds for all $K > 0$ and $\varepsilon > 0$, we deduce $S_{\text{tunfix}} = o(1)$ by first letting $K \to \infty$ and then $\varepsilon \to 0$. Together with (12.134)–(12.135) and Slutsky's Theorem (see for instance Gut [29]), this completes the proof of (12.133). □

In order to motivate Theorem 12.2, we first give four lemmas. It is assumed for all of them that the regularity conditions of Theorem 12.2 hold.

Lemma 12.2 *Let r_{ilj} be the probabilities defined in* (12.37)–(12.40). *Then*

$$\begin{aligned} r_{ilj} &= O(u_j^{1-2^{-(j-l-1)}}), \\ r_{ilj} &= \Omega(u_{l+2}^{1-2^{-(j-l-1)}}), \end{aligned} \quad i \le l \le j-2, \tag{12.141}$$

and

$$\begin{aligned} r_{ilj} &= O(v_j^{1-2^{-(l-j-1)}}), \\ r_{ilj} &= \Omega(v_{l-2}^{1-2^{-(l-j-1)}}), \end{aligned} \quad j+2 \le l \le i, \tag{12.142}$$

as $N \to \infty$. The corresponding formulas for $r_{ij} = r_{iij}$ in (12.36) *are obtained by putting $l = i$ in* (12.141)–(12.142).

Proof. In order to prove (12.141), assume $i \le l \le j-2$. Since $r_{i,j-1,j} = 1$, repeated application of the recursive formula $r_{i,k-1,j} = R(\rho_{ikj})\sqrt{r_{ikj}u_{k+1}}$ in (12.38), for $k = j-1, \ldots, l+1$, leads to

$$r_{ilj} = \prod_{k=l+1}^{j-1} R(\rho_{ikj})^{2^{-(k-l-1)}} u_{k+1}^{2^{-(k-l)}}. \tag{12.143}$$

We know from (12.48) that all $\rho_{ilj} = O(1)$ as $N \to \infty$. From this and the definition of the function $R(\rho)$ in (12.41), it follows that $R(\rho_{ilj}) = \Theta(1)$ as $N \to \infty$, so that

$$r_{ilj} = \Theta\left(\prod_{k=l+1}^{j-1} u_{k+1}^{2^{-(k-l)}}\right). \tag{12.144}$$

Then both parts of (12.141) follow by inserting the first equation of (12.46) into (12.144). The proof of (12.142) when $j+2 \le l \le i$ is analogous. Since $r_{i,j+1,j} = 1$, we use a recursion for $k = j+1, \ldots, l-1$ in order to arrive at the explicit formula

$$r_{ilj} = \prod_{k=j+1}^{l-1} R(\rho_{ikj})^{2^{-(l-k-1)}} v_{k-1}^{2^{-(l-k)}}.$$

Then use (12.48) and the third equation of (12.46) to verify that r_{ilj} satisfies (12.142). □

Lemma 12.3 *Let q_{ij}, q_{ilj}, r_{ij}, and r_{ilj} be the probabilities defined in connection with* (12.35)–(12.40). *Consider a fixed $i \in \{0, 1, \ldots, m-1\}$, and let $F(i)$ and $B(i)$ be the indices defined in* (12.44). *Then,*

$$\begin{aligned} &q_{ilF(i)} \sim r_{ilF(i)}, l = i, i+1, \ldots, F(i)-1, \\ &q_{ilB(i)} \sim r_{ilB(i)}, l = B(i)+1, \ldots, i, \text{ if } i > 0 \text{ and } \hat{\pi}_{iB(i)} > 0 \end{aligned} \tag{12.145}$$

as $N \to \infty$. *In particular,*

$$\begin{aligned} &q_{iF(i)} \sim r_{iF(i)}, \\ &q_{iB(i)} \sim r_{iB(i)}, \text{ if } i > 0 \text{ and } \hat{\pi}_{iB(i)} > 0. \end{aligned} \tag{12.146}$$

Sketch of proof. Notice that (12.146) is a direct consequence of (12.145), since $q_{iij} = q_{ij}$ and $r_{iij} = r_{ij}$. We will only motivate the upper part of (12.145), since the lower part is treated similarly. Consider a fixed $i \in \{0, \ldots, m-1\}$, and for simplicity of notation we write $j = F(i)$. We will argue that

$$q_{ilj} \sim r_{ilj} \tag{12.147}$$

for $l = j-1, \ldots, i$ by means of induction. Formula (12.147) clearly holds when $l = j-1$, since, by definition, $q_{i,j-1,j} = r_{i,j-1,j} = 1$. As for the induction step, let $i+1 \le l \le j-1$, and suppose (12.147) has been proved for l. Then recall the recursive formula

$$r_{i,l-1,j} = R(\rho_{ilj})\sqrt{u_{l+1} r_{ilj}} \tag{12.148}$$

from (12.38), with R defined in (12.41). If

$$q_{i,l-1,j} \sim R(\rho_{ilj})\sqrt{u_{l+1} q_{ilj}} \tag{12.149}$$

holds as well, then (12.147) has been shown for $l-1$, and the induction proof is completed. Without loss of generality we may assume that $j \ge i+2$, since otherwise the induction proof of (12.147) stops after the first trivial step $l = j-1$.

In order to motivate (12.149), we will look at what happens when the population is in fixed state i. Suppose $Z_{\tau_k} = e_i$, and recall that τ'_{k+1} is the time point when the first successful mutation $i \to i+1$ in (τ_k, τ_{k+1}) arrives. Therefore, if $Z_{\tau_{k+1}} = e_j$, there is a non-empty set $J = \{i+1, \ldots, j-1\}$ of types that must be present among some of the descendants of the successful mutation, before a mutation $j-1 \to j$ arrives at some time point $\tau''_{k+1} \in (\tau'_{k+1}, \tau_{k+1})$. Put $Z_{tJ} = \max_{l \in J} Z_{tl}$. The regularity condition

$$P\Big(\sup_{\tau'_{k+1} < t < \tau''_{k+1}} Z_{tJ} > \varepsilon | Z_{\tau_k} = e_i\Big) \to 0 \tag{12.150}$$

for all $\varepsilon > 0$ as $N \to \infty$, assures that with high probability, none of the alleles in J reaches a high frequency after the successful $i \to i+1$ mutation occurred, and before allele j first appears. We will need this condition below, for verifying the induction step (12.149).

The rationale for (12.150) is that fixation events $i \to j$ will happen much more frequently than other types of fixation events $i \to l$ with $l \in J$, because of (12.44). We will motivate that

$$P = P\Big(\sup_{\tau_k < t < \tau_k + a\mu_i^{-1}} Z_{tJ} > \varepsilon | Z_{\tau_k} = e_i\Big) \to 0 \tag{12.151}$$

for any $a > 0$ and $\varepsilon > 0$ as $N \to \infty$, with μ_i the rate of leaving fixation state i. In Lemma 12.1 we motivated that $\tau'_{k+1} - \tau_k = O_p(\mu_i^{-1})$, and in Lemma 12.5 we will argue that $\tau''_{k+1} - \tau'_{k+1} = o_p(\mu_i^{-1})$. Since this implies $\tau''_{k+1} - \tau_k = O_p(\mu_i^{-1})$, formula (12.150) will follow from (12.151).

In order to motivate (12.151), assume for simplicity there are no backward mutations (the proof is analogous but more complicated if we include back mutations as well). If allele $l \in J$ exceeds frequency ε, we refer to this as a semi-fixation event. Let $\lambda_{il}(\varepsilon)$ be the rate at which this happens after time τ_k, and before the next fixed state is reached. Then, the rate at which semi-fixation events happen among some $l \in J$, is

$$\begin{aligned}
\lambda_{iJ}(\varepsilon) &= \sum\nolimits_{l\in J} \lambda_{il}(\varepsilon) \\
&\sim N u_{i+1} \sum\nolimits_{l\in J} q_{il} \beta_{N\varepsilon}\left(\frac{s_l}{s_i}\right) \\
&\leq C(\varepsilon) \times N u_{i+1} \sum\nolimits_{l\in J} q_{il} \beta\left(\frac{s_l}{s_i}\right) \\
&\sim C(\varepsilon) \sum\nolimits_{l\in J} \lambda_{il}.
\end{aligned} \tag{12.152}$$

In the second step of (12.152) we introduced $\beta_{N\varepsilon}(s)$, the probability that a single mutant with fitness s reaches frequency ε, if all other individuals have fitness 1 and there are no mutations. We made use of

$$\lambda_{il}(\varepsilon) \sim N u_{i+1} q_{il} \beta_{N\varepsilon}\left(\frac{s_l}{s_i}\right). \tag{12.153}$$

This is motivated as in the proof of Lemma 12.4, in particular Eqs. (12.163), (12.164) and variant of (12.167) for semi-fixation rather than fixation. In the third step of (12.152) we utilized that $\beta_{N\varepsilon}(s)$ is larger than the corresponding fixation probability $\beta(s) = \beta_N(s)$ for a population of size N. In order to quantify how much larger the fixation probability of the smaller population of size $N\varepsilon$ is, we introduced $C(\varepsilon)$, an upper bound of $\beta_{N\varepsilon}(s_l/s_i)/\beta(s_l/s_i)$ that holds for all $l \in J$. An expression for $C(\varepsilon)$ can be derived from (12.32) if s_l/s_i is sufficiently close to 1. Indeed, we know from (12.48) that $s_l/s_i \to 1$ as $N \to \infty$. However, we need to sharpen this condition somewhat, to

$$s = \frac{s_l}{s_i} \geq 1 + \frac{x}{N} \tag{12.154}$$

for all $l \in J$ and some fixed $x < 0$. Then it follows from (12.32) that

$$\frac{\beta_{N\varepsilon}(s)}{\beta_N(s)} = \frac{s^{-N} - 1}{s^{-N\varepsilon} - 1} \leq \frac{(1 + x/N)^{-N} - 1}{(1 + x/N)^{-N\varepsilon} - 1} \to \frac{e^{-x} - 1}{e^{-\varepsilon x} - 1} =: C(\varepsilon)$$

is a constant not depending on N. Finally, in the last step of (12.152) we assumed

$$\lambda_{il} \sim N u_{i+1} q_{il} \beta \left(\frac{s_l}{s_i} \right), \quad l \in J. \tag{12.155}$$

This is motivated in the same way as Eq. (12.153), making use of (12.163)–(12.164) and (12.167).

Assuming that semi-fixation events arrive according to a Poisson process with intensity $\lambda_{iJ}(\varepsilon)$, formula (12.151) follows from (12.44) to (12.152), since

$$\begin{aligned} P &\sim 1 - \exp\left(-\lambda_{iJ}(\varepsilon) \times \frac{a}{\mu_i} \right) \\ &\le 1 - \exp\left(-C(\varepsilon) \sum\nolimits_{l \in J} \lambda_{il} \times \frac{a}{\mu_i} \right) \\ &= 1 - \exp(-C(\varepsilon) a \sum\nolimits_{l \in J} p_{il}) \\ &\to 1 - \exp(-C(\varepsilon) a \sum\nolimits_{l \in J} \pi_{il}) \\ &= 1 - \exp(-C(\varepsilon) a \sum\nolimits_{l \in J} \hat{\pi}_{il}) \\ &= 0 \end{aligned} \tag{12.156}$$

as $N \to \infty$. In the third step of (12.156) we used (12.16) to conclude that $p_{il} = \lambda_{il}/\mu_i$, and in the fourth step we utilized (12.17). In the fifth step of (12.156) we claimed that $\pi_{il} = \hat{\pi}_{il}$ for $l \in J$, Although we have not given a strict proof of this, it seems reasonable in view of the definitions of π_{il} and $\hat{\pi}_{il}$ in (12.17) and (12.43), together with (12.35), (12.155), and the fact that $q_{il} \sim r_{il}$ for $i < l < F(i)$ (which can be proved by induction with respect to l). Finally, in the last step of (12.156) we invoked (12.44), which implies $\hat{\pi}_{il} = 0$ for all $l \in J = \{i+1, \ldots, F(i)-1\}$.

Equation (12.150) enables us to approximate the allele frequency Z_{tl} by a branching process with mutations, in order to motivate (12.149). (A strict proof of this for a neutral model $s_0 = \cdots s_{m-1} = 1$ can be found in Theorem 2 of Durrett et al. [20].) We will look at the fate of the first $l-1 \to l$ mutation at time $\tau \in (\tau'_{k+1}, \tau''_{k+1})$, that is a descendant of the first successful $i \to i+1$ mutation at time τ'_{k+1}, and arrives before the first $j-1 \to j$ mutation at time τ''_{k+1}. Recall that $q = q_{i,l-1,j}$ is the probability that this l mutation gets an offspring that mutates into type j, and $q' = q_{ilj}$ is the corresponding probability that one of its descendants, an $l \to l+1$ mutation, gets a type j offspring. Let also $r' = r_{ilj}$ be the approximation q', and write $s = s_l/s_i$ for the ratio between the selection coefficients of alleles l and i. With this simplified notation, according to (12.149), we need to show that

$$q \sim R(\rho)\sqrt{uq'} \tag{12.157}$$

as $N \to \infty$, where $u = u_{l+1}$, and $\rho = \rho_{ilj}$ is defined in (12.37), i.e.

$$s = 1 + \rho\sqrt{ur'}. \tag{12.158}$$

We make the simplifying assumption that at time τ, the population has one single type l individual, the one that mutated from type $l-1$ at this time point, whereas all other $N-1$ individuals have type i. (Recall that we argued in Lemma 12.1 that such an assumption is asymptotically accurate.) In order to compute the probability q for the event A that this individual gets a descendant of type j, we condition on the next time point when one individual dies and is replaced by the offspring of an individual that reproduces. Let D and R be independent indicator variables for the events that the type l individual dies and reproduces respectively. Using the definition of the Moran process in Sect. 12.2, this gives an approximate recursive relation

$$\begin{aligned}
q &= P(A) \\
&= P(D=0, R=0)P(A|D=0, R=0) \\
&+ P(D=0, R=1)P(A|D=0, R=1) \\
&+ P(D=1, R=0)P(A|D=1, R=0) \\
&+ P(D=1, R=1)P(A|D=1, R=1) \\
&= \left(1-\frac{1}{N}\right)\frac{N-1}{N-1+s} \times q \\
&+ \left(1-\frac{1}{N}\right)\frac{s}{N-1+s} \\
&\times \left[u(q'+q-q'q) + vq + (1-u-v)(2q-q^2)\right] \\
&+ \frac{1}{N}\frac{N-1}{N-1+s} \times 0 \\
&+ \frac{1}{N}\frac{s}{N-1+s} \times \left[uq' + v \times 0 + (1-u-v)q\right]
\end{aligned} \tag{12.159}$$

for q, where $v = v_{l-1}$ is the probability of a back mutation $l \to l-1$. In the last step of (12.159) we retained the exact transition probabilities of the Moran process, but we used a branching process approximation for the probability q that the type l mutation at time τ gets a type j descendant. This approximation relies on (12.150), and it means that descendants of the type l mutation that are alive at the same time point, have independent lines of descent after this time point. For instance, in the second term on the right hand side of (12.159), a type i individual dies and the type l individual reproduces ($D=0$, $R=1$). Then there are three possibilities: First, the offspring of the type l individual mutates to $l+1$ with probability u. Since the type l individual and its type $l+1$ offspring have independent lines of descent, the probability is $1-(1-q')(1-q) = q'+q-q'q$ that at least one of them gets a type j descendant. Second, if the offspring mutates back to $l-1$ (with probability v), its type l parent has a probability q of getting a type j descendant. Third, if the offspring does not mutate (with probability $1-u-v$), there are two type l individuals, with a probability $1-(1-q)^2 = 2q-q^2$ that at least one of them gets a type j offspring.

Equation (12.159) is quadratic in q. Dividing both sides of it by $s/(N-1+s)$, it can be seen, after some computations, that this equation simplifies to $aq^2+bq+c=0$, with

$$\begin{aligned} a &= (1-u-v)\left(1-\frac{1}{N}\right) \sim 1, \\ b &= \frac{N-1}{N}\times\frac{1-s}{s}+u(1+q'-\frac{q'}{N})+v \\ &\sim -\frac{\rho\sqrt{ur'}}{1+\rho\sqrt{ur'}}+(1+q')u+v \\ &\sim -\rho\sqrt{uq'}, \\ c &= -uq', \end{aligned} \tag{12.160}$$

as $N\to\infty$. When simplifying the formula for b, we used (12.158) in the second step, the induction hypothesis (12.147) in the last step (since it implies $q'\sim r'$), and additionally we assumed in the last step that $(1+q')u+v=o(\sqrt{ur'})$. In order to justify this, from the second equation of (12.46) we know that $v=O(u)$, and since $q'\le 1$, it suffices to verify that $u=o(\sqrt{ur'})$, or equivalently that $r'=\Omega(u)$. But this follows from (12.46), (12.141), and the fact that $u=u_{l+1}$, since

$$r'=r_{ilj}=\Omega\left(u_{l+2}^{1-2^{-(j-l-1)}}\right)=\Omega\left(u^{1-2^{-(j-l-1)}}\right)=\Omega(u),$$

where in the last step we used that $l\le j-1$. This verifies the asymptotic approximation of b in (12.160).

To conclude, in order to prove of (12.157), we notice that the only positive solution to the quadratic equation in q, with coefficients as in (12.160), is

$$\begin{aligned} q &\sim \frac{\rho\sqrt{uq'}}{2}+\sqrt{\frac{\rho^2uq'}{4}+uq'} \\ &= \frac{\rho+\sqrt{\rho^2+4}}{2}\sqrt{uq'} \\ &= R(\rho)\sqrt{uq'}, \end{aligned}$$

where in the last step we invoked the definition of $R(\rho)$ in (12.41). This finishes the proof of the induction step (12.149) or (12.157), and thereby the proof of (12.147).

We end this proof by a remark: Recall that r_{ij} in (12.36) is an approximation q_{ij}, obtained from recursion (12.38) or (12.148) when $j>i$, and from (12.40) when $j<i$. A more accurate (but less explicit) approximation of q_{ij} is obtained, when $i<j$, by recursively solving the quadratic equation $ax^2+bx+c=0$, with respect to $x=r_{i,l-1,j}$ for $l=j-1,\dots,i+1$, and finally putting $r_{ij}=r_{iij}$. The coefficients of this equation are defined as in (12.160), with $r'=r_{ilj}$ instead of q'. When $j<i$, the improved approximation of q_{ij} is defined analogously. □

Lemma 12.4 *Let μ_i be the rate (12.15) at which a successful forward or backward mutation occurs in a homogeneous type i population, and let $\hat{\mu}_i$ in (12.42) be its approximation. Define the asymptotic transition probabilities π_{ij} between fixed population states as in (12.17), and their approximations $\hat{\pi}_{ij}$ as in (12.43). Then*

$$\mu_i \sim \hat{\mu}_i, \quad i = 0, \ldots, m-1, \tag{12.161}$$

as $N \to \infty$, and

$$\pi_{ij} = \hat{\pi}_{ij}, \quad i, j = 0, 1, \ldots, m. \tag{12.162}$$

Sketch of proof. Consider a time point τ_k when the population becomes fixed with type i, so that $Z_{\tau_k} = e_i$. Denote by f_{ij} the probability a forward mutation $i \to i+1$, which appears at a time point later than τ_k, is the first successful mutation after τ_k, that its descendants have taken over the population by time τ_{k+1}, and that all of them by that time have type j (so that $Z_{\tau_{k+1}} = e_j$). Likewise, when $j < i$ and $i \geq 1$, we let b_{ij} refer to the probability that if a backward mutation $i \to i-1$ arrives, it is successful, its descendants have taken over the population by time τ_{k+1}, and all of them have type j. For definiteness we also put $b_{0j} = 0$. We argue that

$$\lambda_{ij} \sim \begin{cases} N u_{i+1} f_{ij}, & j > i, \\ N v_{i-1} b_{ij}, & j < i, \end{cases} \tag{12.163}$$

since the event that the population at time τ_{k+1} have descended from more than one $i \to i \pm 1$ mutation that occurred in the time interval (τ_k, τ_{k+1}), is asymptotically negligible.

Let $\beta_j(z)$ be the probability that the descendants of a type j individual, who lives in a population with a type configuration z, takes over the population so that it becomes homogeneous of type j. Although $\beta_j(z)$ depends on the mutation rates $u_1, \ldots, u_m, v_0, \ldots, v_{m-1}$ as well as the selection coefficients $s_1, \ldots, s_m$, this is not made explicit in the notation. The probabilities f_{ij} and b_{ij} in (12.163) can be written as a product

$$\begin{aligned} f_{ij} &= q_{ij} E\left[\beta_j(Z_{\tau''_{k+1}}) | A_j, Z_{\tau_k} = e_i\right], \; j > i, \\ b_{ij} &= q_{ij} E\left[\beta_j(Z_{\tau''_{k+1}}) | A_j, Z_{\tau_k} = e_i\right], \; j < i \end{aligned} \tag{12.164}$$

of two terms. Recall that the first term, q_{ij}, is the probability that the first successful mutation $i \to i \pm 1$ at time $\tau'_{k+1} > \tau_k$ has a descendant that mutates into type j at some time $\tau''_{k+1} \in (\tau'_{k+1}, \tau_{k+1})$. The second term is the probability that this mutation has spread to the rest of the population by time τ_{k+1}. The conditional expectation of this second term is with respect to variations in $Z_{\tau''_{k+1}}$, and the conditioning is with respect to A_j, the event that the mutation at time τ''_{k+1} is into type j.

In order to compare the transition rates in (12.163) with the approximate ones in (12.35), we notice that the latter can be written as

$$\hat{\lambda}_{ij} = \begin{cases} Nu_{i+1}\hat{f}_{ij}, & j > i, \\ Nv_{i-1}\hat{b}_{ij}, & j < i, \end{cases} \tag{12.165}$$

where

$$\begin{aligned} \hat{f}_{ij} &= r_{ij}\beta(s_j/s_i), \quad j > i, \\ \hat{b}_{ij} &= r_{ij}\beta(s_j/s_i), \quad j < i, \end{aligned} \tag{12.166}$$

r_{ij} is the approximation of q_{ij} defined in (12.36), whereas $\beta(s_j/s_i)$ is the probability that a single type j individual gets fixed in a population without mutations, where all other individuals have type i.

We will argue that the probabilities in (12.166) are asymptotically accurate approximations of those in (12.164), for all pairs i, j of states that dominate asymptotically, that is, those pairs for which $j \in \{B(i), F(i)\}$. In Lemma 12.3 we motivated that r_{ij} is an asymptotically accurate approximation of q_{ij} for all such pairs of states. Likewise, we argue that $\beta(s_j/s_i)$ is a good approximation of the conditional expectation in (12.164). Indeed, following the reasoning of Lemma 12.3, since none of the intermediate alleles, between i and j, will reach a high frequency before the type j mutant appears at time τ''_{k+1}, it follows that most of the other $N-1$ individuals will have type i at this time point. Consequently,

$$E\left[\beta_j(Z_{\tau''_{k+1}})|A_j, Z_{\tau_k} = e_i\right] \sim \beta_j\left(\frac{N-1}{N}e_i + \frac{1}{N}e_j\right) \sim \beta\left(\frac{s_j}{s_i}\right) \tag{12.167}$$

as $N \to \infty$. In the last step of (12.167) we used that new mutations between time points τ''_{k+1} and τ_{k+1} can be ignored, because of the smallness (12.4)–(12.5) of the mutation rates. Since $\beta_j\left((N-1)e_i/N + e_j/N\right)$ is the fixation probability of a single type j mutant that has selection coefficient s_j/s_i relative to the other $N-1$ type i individuals, it is approximately equal to the corresponding fixation probability $\beta(s_j/s_i)$ of a mutation free Moran model. It therefore follows from (12.164) and (12.166) that

$$\begin{aligned} \hat{f}_{iF(i)} &\sim f_{iF(i)}, i = 0, \ldots, m-1, \\ \hat{b}_{iB(i)} &\sim b_{iB(i)}, i = 1, \ldots, m-1 \text{ and } B(i) \neq \emptyset \end{aligned} \tag{12.168}$$

as $N \to \infty$.

Next we consider pairs of types i, j such that $j \notin \{B(i), F(i)\}$. We know from (12.44), (12.165) and (12.166) that $\hat{f}_{il} = o(\hat{f}_{iF(i)})$ for all $l > i$ such that $l \neq F(i)$. It is therefore reasonable to assume that $f_{il} = o(f_{iF(i)})$ as well for all $l > i$ with $l \neq F(i)$, although $\hat{f}_{il}$ need not necessarily be a good approximation of f_{il} for all these l. The same argument also applies to backward mutations when $B(i) \neq \emptyset$ and $\hat{\pi}_{iB(i)} > 0$, that is, we should have $f_{il} = o(f_{iB(i)})$ for all $l < i$ such that $l \neq B(i)$.

Putting things together, it follows from (12.44), (12.163), (12.165), (12.168), and the last paragraph that the approximate rate (12.42) at which a homogeneous type i population is transferred into a new fixed state, satisfies

$$\begin{aligned}\hat{\mu}_i &= Nv_{i-1}\sum_{j=0}^{i-1}\hat{b}_{ij} + Nu_{i+1}\sum_{j=i+1}^{m}\hat{f}_{ij}\\ &\sim 1\left(\hat{\pi}_{iB(i)} > 0\right) Nv_{i-1}\hat{b}_{iB(i)} + Nu_{i+1}\hat{f}_{iF(i)}\\ &\sim 1\left(\hat{\pi}_{iB(i)} > 0\right) Nv_{i-1}b_{iB(i)} + Nu_{i+1}f_{iF(i)}\\ &\sim Nv_{i-1}\sum_{j=0}^{i-1}b_{ij} + Nu_{i+1}\sum_{j=i+1}^{m}f_{ij}\\ &\sim \mu_i,\end{aligned} \tag{12.169}$$

as $N \to \infty$, in agreement with (12.161). Formulas (12.16)–(12.17), (12.43)–(12.44), (12.163), (12.165), and (12.168)–(12.169) also motivate why π_{ij} should equal $\hat{\pi}_{ij}$, in accordance with (12.162). □

Lemma 12.5 *The regularity condition* (12.47) *of Theorem 12.2 implies that* (12.26) *holds.*

Sketch of proof. Suppose $Z_{\tau_k} = e_i$ and $Z_{\tau_{k+1}} = e_j$ for some $i \in I_{\text{as}}$ and $j \neq i$. Write

$$\tau_{k+1} - \tau'_{k+1} = \begin{cases} \sum_{l=i+1}^{j-1}\sigma_l + \sigma_{\text{fix}} := \sigma_{\text{tunnel}} + \sigma_{\text{fix}}, & j > i,\\ \sum_{l=j+1}^{i-1}\sigma_l + \sigma_{\text{fix}} := \sigma_{\text{tunnel}} + \sigma_{\text{fix}}, & j < i.\end{cases} \tag{12.170}$$

If $j > i$, then the successful mutation at time τ'_{k+1} is from i to $i+1$. This type $i+1$ mutation has a line of descent with individuals that mutate to types $i+2, \ldots, j$, before the descendants of the type j mutation take over the population. The first term $\sigma_{\text{tunnel}} = \tau''_{k+1} - \tau'_{k+1}$ on the right hand side of (12.170) is the time it takes for the type $i+1$ mutation to tunnel into type j. It is the sum of σ_l, the time it takes for the type $l+1$ mutation to appear after the type l mutation, for all $l = i+1, \ldots, j-1$. The second term $\sigma_{\text{fix}} = \tau_{k+1} - \tau''_{k+1}$ on the right hand side of (12.170) is the time it takes for j to get fixed after the j mutation first appears. When $j < i$, we interpret the terms of (12.170) analogously. It follows from (12.170) that in order to prove (12.26), it suffices to show that

$$\begin{aligned}\sigma_{\text{tunnel}} &= o_p(\mu_{\min}^{-1}),\\ \sigma_{\text{fix}} &= o_p(\mu_{\min}^{-1}),\end{aligned} \tag{12.171}$$

as $N \to \infty$ for all asymptotic states $i \in I_{\text{as}}$. When $j > i$, we know from (12.44) to (12.162) that with probability tending to 1, $j = F(i)$. Following the argument from the proof of Theorem 2 of Durrett et al. [20], we have that

$$\sigma_l = O_p(q_{ilj}^{-1}). \tag{12.172}$$

In the special case when $l = i+1$ and $j = i+2$, formula (12.172) can also be deduced from the proof of Theorem 12.3, by looking at $G(x)/G(\infty)$ in (12.191). Using (12.172), we obtain the upper part of (12.171), since

$$\begin{aligned}\sigma_{\text{tunnel}} &= \sum_{l=i+1}^{j-1} \sigma_l \\ &= O_p\left(\sum_{l=i+1}^{j-1} q_{ilj}^{-1}\right) \\ &= o_p(q_{iij}^{-1}) \\ &= o_p(q_{ij}^{-1}) \\ &= o_p(\mu_i^{-1}) \\ &= o_p(\mu_{\min}^{-1}).\end{aligned} \tag{12.173}$$

In the second step of (12.173) we used that $q_{iij} \leq q_{ilj}$ for $i < l$, which follows from the definition of these quantities, in the third step we invoked $q_{ij} = q_{iij}$, and in the fourth step we applied the relation

$$\mu_i = \Theta\left(N u_{i+1} q_{ij} \beta\left(\frac{s_i}{s_j}\right)\right) = o(q_{ij}). \tag{12.174}$$

The first step of (12.174) is motivated as in Lemma 12.4, since $j = F(i)$ and hence $\pi_{ij} > 0$, whereas the second step follows from (12.4) and the fact that $\beta(s_i/s_j)$ is bounded by 1. Finally, the fourth step of (12.173) follows from the definition of $\mu_{\min}$ in (12.24), since (12.174) applies to any $i \in I_{\text{as}}$. When $j < i$, the first part of (12.171) is shown analogously.

In order to verify the second part of (12.171), we know from the motivation of Lemma 12.4 that with high probability, σ_{fix} is the time it takes for descendants of the type j mutation to take over the population, ignoring the probability that descendants of other individuals first mutated into j and then some of them survived up to time τ_{k+1} as well. We further recall from Lemma 12.4 that because of the smallness (12.4)–(12.5) of the mutation rates, right after the j mutation has arrived at time τ''_{k+1}, we may assume that the remaining $N-1$ individuals have type i, and after that no other mutation occurs until the j allele gets fixed at time τ_{k+1}. With these assumptions, σ_{fix} is the time for one single individual with selection coefficient s_j/s_i to get fixed in a two-type Moran model without mutations, where all other individuals have selection coefficient 1. From Sect. 12.5 it follows that $E(\sigma_{\text{fix}}) \sim \alpha(s_j/s_i)$, and therefore the second part of (12.171) will be proved if we can verify that

$$\alpha\left(\frac{s_j}{s_i}\right) = o(\mu_{\min}^{-1})$$

holds for all $i \in I_{\text{as}}$ and $j \in \{B(i), F(i)\}$ as $N \to \infty$. This is equivalent to showing that

$$\mu_{\min} = o\left(\min_{i \in I_{\text{as}}} \min\left[\alpha^{-1}\left(\frac{s_{B(i)}}{s_i}\right), \alpha^{-1}\left(\frac{s_{F(i)}}{s_i}\right)\right]\right) \tag{12.175}$$

as $N \to \infty$, where the $\alpha^{-1}(s_{B(i)}/s_i)$-term is included only when $B(i) \neq \emptyset$ (or equivalently, when $\pi_{iB(i)} > 0$). Using (12.44), (12.46), (12.141), (12.161), (12.168), and

(12.169), we find that

$$\begin{aligned}\mu_i \sim & \hat{\mu}_i \\ =& O\left(N u_{i+1} r_{iF(i)} \beta(s_{F(i)}/s_j)\right) \\ =& O\left(N u_{i+1} u_{F(i)}^{1-2^{-(F(i)-i-1)}} \beta(s_{F(i)}/s_j)\right) \\ =& O\left(N u_{F(i)}^{2-2^{-(F(i)-i-1)}} \beta(s_{F(i)}/s_j)\right).\end{aligned} \tag{12.176}$$

Inserting (12.176) into the definition of $\mu_{\min}$ in (12.24), we obtain

$$\mu_{\min} = O\left(\min_{i \in I_{\text{long}}} N u_{F(i)}^{2-2^{-(F(i)-i-1)}} \beta(s_{F(i)}/s_j)\right),$$

and formula (12.175) follows, because of (12.47). □

Proof of Theorem 12.2. We need to establish that the limit result (12.49) of Theorem 12.2 follows from Theorem 12.1. To this end, we first need to show that all $\hat{\lambda}_{ij}$ are good approximations of λ_{ij}, in the sense specified by Theorem 12.2, i.e. $\pi_{ij} = \hat{\pi}_{ij}$ and $\hat{\mu}_i / \hat{\mu}_{\min} \to \kappa_i$ as $N \to \infty$. But this follows from Lemma 12.4, and the definitions of $\mu_{\min}$ and $\hat{\mu}_{\min}$ in (12.24) and Theorem 12.2. Then it remains to check those two regularity conditions (12.18) and (12.26) of Theorem 12.1 that are not present in Theorem 12.2. But (12.18) follows from (12.44) to (12.162), since these two equations imply $\pi_{iF(i)} > 0$ for all $i = 0, \ldots, m-1$, and (12.26) follows from Lemma 12.5. □

Proof of (12.109). Let

$$\theta_i = u \times E(T_m | Z_0 = e_i) \tag{12.177}$$

be the standardized expected waiting time until all m mutations have appeared and spread in the population, given that it starts in fixed state i. Our goal is to find an explicit formula for θ_0, and then show that (12.109) is an asymptotically accurate approximation of this explicit formula as $m \to \infty$.

Recall that Σ_{ij} in (12.107) are the elements of the intensity matrix, for the Markov process that switches between fixed population states, when time has been multiplied by $\hat{\mu}_{\min} = u$. When the population is in fixed state i, the standardized expected waiting time until the next transition is $1/(-\Sigma_{ii})$. By conditioning on what happens at this transition, it can be seen that the standardized expected waiting times in (12.177), satisfy a recursive relation

$$\theta_i = \frac{1}{-\Sigma_{ii}} + \frac{\Sigma_{i,i-1}}{-\Sigma_{ii}} \times \theta_{i-1} + \frac{\Sigma_{i,i+1}}{-\Sigma_{ii}} \times \theta_{i+1}, \tag{12.178}$$

for $i = 0, 1, \ldots, m-1$, assuming $\theta_{-1} = 0$ on the right hand side of (12.178) when $i = 0$, and similarly $\theta_m = 0$ when $i = m-1$. Inserting the values of Σ_{ij} from (12.107) into (12.178), we can rewrite the latter equation as

$$\theta_0 - \theta_1 = \frac{1}{m} =: b_0 \tag{12.179}$$

and

$$\theta_i - \theta_{i+1} = \frac{Ci}{m-i}(\theta_{i-1} - \theta_i) + \frac{1}{m-i} =: a_i(\theta_{i-1} - \theta_i) + b_i, \tag{12.180}$$

for $i = 1, \ldots, m-1$, respectively. We obtain an explicit formula for θ_0 by first solving the linear recursion for $\theta_i - \theta_{i+1}$ in (12.179)–(12.180), and then summing over i. This yields

$$\theta_0 = \sum_{i=0}^{m-1}(\theta_i - \theta_{i+1}) = \sum_{i=0}^{m-1}\sum_{k=0}^{i}\theta_{ik}, \tag{12.181}$$

where

$$\theta_{ik} = b_k \prod_{j=k+1}^{i} a_j = \frac{\binom{m-1}{k}}{(m-k)\binom{m-1}{i}} \times C^{i-k}. \tag{12.182}$$

Formulas (12.181)–(12.182) provide the desired explicit formula for θ_0. When $C = 0$, it is clear that

$$\begin{aligned}\theta_0 &= \sum_{i=0}^{m-1}\theta_{ii}\\ &= \sum_{i=0}^{m-1} 1/(m-i)\\ &\sim \log(m) + \gamma,\end{aligned}$$

where $\gamma \approx 0.5772$ is the Euler–Mascheroni constant. This proves the upper half of (12.109). For $C > 0$, we will show that when m gets large, the (standardized) expected waiting time until the last mutant gets fixed, $\theta_{m-1} - \theta_m = \theta_{m-1}$, dominates the first sum in (12.181). To this end, we first look at θ_{m-1}, and rewrite this quantity as

$$\begin{aligned}\theta_{m-1} &= \sum_{k=0}^{m-1}\theta_{m-1,k}\\ &= \tfrac{1}{\binom{m-1}{m-1}}\sum_{k=0}^{m-1}\tfrac{1}{m-k}\binom{m-1}{k}C^{m-1-k}\\ &= (1+C)^{m-1}\sum_{k=0}^{m-1}\tfrac{1}{m-k}\binom{m-1}{k}\left(\tfrac{1}{1+C}\right)^k\left(\tfrac{C}{1+C}\right)^{m-1-k}\\ &= (1+C)^{m-1}E\left(\tfrac{1}{m-X_{m-1}}\right)\\ &= (1+C)^{m-1}E\left(\tfrac{1}{1+Y_{m-1}}\right),\end{aligned} \tag{12.183}$$

where

$$\begin{aligned}X_{m-1} &\overset{\mathcal{L}}{\in} \mathrm{Bin}\left(m-1, \tfrac{1}{1+C}\right),\\ Y_{m-1} = m-1-X_{m-1} &\overset{\mathcal{L}}{\in} \mathrm{Bin}\left(m-1, \tfrac{C}{1+C}\right)\end{aligned}$$

are two binomially distributed random variables. For large m, we apply the Law of Large Numbers to Y_{m-1} and find that

$$\begin{aligned}\theta_{m-1} \approx &(1+C)^{m-1}\tfrac{1}{1+E(Y_{m-1})}\\ \approx &(1+C)^{m-1}\tfrac{1}{mC/(1+C)}\\ = &(1+C)^m/(Cm),\end{aligned} \tag{12.184}$$

in agreement with the lower half of (12.109). In view of (12.181), in order to finalize the proof of (12.109), we need to show that the sum of $\theta_{m-j}-\theta_{m-j+1}$ for $j=2,3,\ldots,m$, is of a smaller order than (12.184). A similar argument as in (12.183) leads to

$$\begin{aligned}\theta_{m-j}-\theta_{m-j+1} = &\sum\nolimits_{k=0}^{m-j}\theta_{m-j,k}\\ = &(j-1)!(1+C)^{m-j}E\left[\tfrac{1}{\prod_{n=1}^{j}(n+Y_{m-j})}\right]\\ \le &\tfrac{2}{j}(1+C)^{m-j}E\left[\tfrac{1}{(1+Y_{m-j})(2+Y_{m-j})}\right],\end{aligned} \tag{12.185}$$

where

$$Y_{m-j} \overset{\mathcal{L}}{\in} \mathrm{Bin}\left(m-j,\frac{C}{1+C}\right).$$

For large m we have, by the Law of Large Numbers, that

$$\begin{aligned}&\theta_{m-j}-\theta_{m-j+1} \le \tfrac{2}{j}(1+C)^{m-j}\tfrac{1}{[1+(m-j)C/(1+C)]^2}\\ \le &\begin{cases}4(1+C)^{m/2}/m, & j>m/2,\\ (1+C)^{m-j}/\left[m/2\times C/(1+C)\right]^2, & 2\le j\le m/2.\end{cases}\end{aligned} \tag{12.186}$$

By summing (12.186) over j, it is easy to see that

$$\sum_{j=2}^{m}(\theta_{m-j}-\theta_{m-j+1}) \ll (1+C)^m/(Cm) \sim \theta_m$$

as $m\to\infty$. Together with (12.184), this completes the derivation of the lower part of (12.109). □

Sketch of proof of Theorem 12.3. Our proof will parallel that of Theorem 1 in Durrett el al. [20], see also Wodarz and Komarova [66]. We first use formula (12.66) in order to deduce that the ratio between the two rates of fixation from a type 0 population, satisfies $\hat{\lambda}_{02}/\hat{\lambda}_{01}\to\infty$ as $N\to\infty$. When $\rho=0$ in (12.51), this is a consequence of $\hat{\lambda}_{02}/\hat{\lambda}_{01}\sim N\sqrt{u_2}$ and the assumption $N\sqrt{u_2}\to\infty$ on the second mutation rate u_2. When $\rho<0$, $\hat{\lambda}_{02}/\hat{\lambda}_{01}$ tends to infinity at an even faster rate, due to the $\psi(\rho u_2^{1/2})$-term

of $\hat{\lambda}_{01}$ in (12.66). In any case, it follows that condition (12.44) is satisfied, with $F(0) = 2$ and $\hat{\pi}_{02} = 1$. That is, tunneling from 0 to 2 will occur with probability tending to 1 as $N \to \infty$ whether $\rho = 0$ or $\rho < 0$. As in the proof of Lemma 12.3 we conclude from this that the fraction $Z_t = Z_{t1}$ of allele 1 will stay close to 0, and we may use a branching process approximation for Z_t. A consequence of this approximation is that type 1 mutations arrive according to a Poisson process with intensity Nu_1, and the descendants of different type 1 mutations evolve independently. Let $0 < \sigma \leq \infty$ be the time it takes for the first type 2 descendant of a type 1 mutation to appear. In particular, if $\sigma = \infty$, this type 1 mutation has no type 2 descendants. Letting $G(x) = P(\sigma \leq x)$ be the distribution function of σ, it follows by a Poisson process thinning argument that

$$P(T_2'' \geq t) \sim \exp(-Nu_1 \int_0^t G(x)dx). \tag{12.187}$$

We use Kolmogorov's backward equation in order to determine G. To this end, we will first compute $G(x + h)$ for a small number $h > 0$, by conditioning on what happens during the time interval $(0, h)$. As in formulas (12.121)–(12.122) of Appendix B, we let $a_{ij}(z)$ refer to the rate at which a type i individual dies and gets replaced by the offspring of a type j individual, when the number of type 1 individuals before the replacement is Nz. Since we look at the descendants of one type 1 individual, we have that $z = Z_0 = 1/N$. Using a similar argument as in Eq. (12.159), it follows from this that

$$\begin{aligned} &G(x+h) = a_{00}(1/N)h \times G(x) \\ &+a_{01}(1/N)h\left[u_2 \times 1 + (1-u_2)(2G(x) - G(x)^2)\right] \\ &+a_{10}(1/N)h \times 0 + a_{11}(1/N)h \times [u_2 \times 1 + (1-u_2)G(x)] \\ &+\left[1 - \sum\nolimits_{ij} a_{ij}(1/N)h\right] G(x) + o(h) \end{aligned} \tag{12.188}$$

for small $h > 0$. Notice that the two $a_{00}(1/N)$ terms cancel out in (12.188), whereas $a_{11}(1/N)(1 - G(x))u_2 \times h = O(N^{-2}u_2 \times h)$ is too small to have an asymptotic impact. Using formulas (12.121)–(12.122) for $a_{01}(1/N)$ and $a_{10}(1/N)$, it follows that (12.188) simplifies to

$$G(x+h) = s \times h\left[u_2 + 2G(x) - G(x)^2\right] + 1 \times h \times 0 + [1 - (s+1)h]\, G(x) + o(h),$$

when all asymptotically negligible terms are put into the remainder term. Letting $h \to 0$, we find that $G(x)$ satisfies the differential equation

$$\begin{aligned} G'(x) &= -sG(x)^2 + (s-1)G(x) + su_2 \\ &= -s(G(x) - r_1)(G(x) - r_2), \end{aligned} \tag{12.189}$$

where

$$r_1 = (s-1)/(2s) + \sqrt{[(s-1)/(2s)]^2 + u_2},$$
$$r_2 = (s-1)/(2s) - \sqrt{[(s-1)/(2s)]^2 + u_2}$$

are the two roots of the quadratic equation $-sy^2 + (s-1)y + su_2 = 0$. Recall from (12.51) that $s = 1 + \rho\sqrt{u_2}$. We may therefore express these two roots as

$$\begin{aligned} r_1 &= \sqrt{u_2}\left(\rho + \sqrt{\rho^2 + 4s^2}\right)/(2s) \sim \sqrt{u_2}\left(\rho + \sqrt{\rho^2+4}\right)/(2s) \\ &= \sqrt{u_2}R(\rho)/s, \\ r_2 &= \sqrt{u_2}\left(\rho - \sqrt{\rho^2 + 4s^2}\right)/(2s) \sim \sqrt{u_2}\left(\rho - \sqrt{\rho^2+4}\right)/(2s), \end{aligned} \tag{12.190}$$

where in the second step we used that $u_2 \to 0$ and $s \to 1$ as $N \to \infty$, and in the last step we invoked (12.41), the definition of $R(\rho)$. Since $r_2 < 0 < r_1$, and $G'(x) \to 0$ as $x \to \infty$, it follows from (12.189) that we must have $G(\infty) = r_1$. Together with the other boundary condition $G(0) = 0$, this gives as solution

$$G(x) = r_1 \frac{1 - e^{-(r_1 - r_2)sx}}{1 - \frac{r_1}{r_2}e^{-(r_1-r_2)sx}} \tag{12.191}$$

to the differential equation (12.189), with

$$r_1 - r_2 \sim \frac{\sqrt{u_2} \times \sqrt{\rho^2+4}}{s}$$

and

$$-\frac{r_1}{r_2} \sim \frac{\sqrt{\rho^2+4} + \rho}{\sqrt{\rho^2+4} - \rho}. \tag{12.192}$$

Putting things together, we find that

$$\begin{aligned} P\left(NR(\rho)u_1\sqrt{u_2} \times T_2'' \geq t\right) &\sim P\left(Nu_1r_1s \times T_2'' \geq t\right)) \\ &\sim \exp\left(-Nu_1 \int_0^{t/(Nu_1r_1s)} G(x)dx\right) \\ &\sim \exp\left(-\int_0^t h(y)dy\right), \end{aligned} \tag{12.193}$$

where formula (12.190) was used in the first step, (12.187) in the second step, in the third step we changed variables $y = Nu_1r_1s \times x$ and introduced the hazard function $h(x) = G\left(x/(Nu_1r_1s)\right)/(sr_1)$. If $Nu_1 \to a > 0$ as $N \to \infty$, it follows from (12.191) and the fact that $s \to 1$ that we can rewrite the hazard function as

$$\begin{aligned} h(x) &\sim \tfrac{1}{sr_1} G\left(\tfrac{x}{sar_1}\right) \\ &= \tfrac{1}{s} \times \frac{1-\exp\left(-\frac{r_1-r_2}{r_1}\times\frac{x}{a}\right)}{1-\frac{r_1}{r_2}\exp\left(-\frac{r_1-r_2}{r_1}\times\frac{x}{a}\right)} \sim \frac{1-\exp\left(-\frac{r_1-r_2}{r_1}\times\frac{x}{a}\right)}{1-\frac{r_1}{r_2}\exp\left(-\frac{r_1-r_2}{r_1}\times\frac{x}{a}\right)}. \end{aligned} \tag{12.194}$$

We finally obtain the limit result (12.110)–(12.111) when $a > 0$ from (12.193) to (12.194), using (12.192) and the fact that

$$\frac{r_1 - r_2}{r_1} \sim \frac{2\sqrt{\rho^2+4}}{\rho + \sqrt{\rho^2+4}}.$$

When $Nu_1 \to 0$, one similarly shows that (12.193) holds, with $h(x) = 1$. Finally, formula (12.112) follows by integrating (12.193) with respect to t. □

Motivation of formula (12.114). We will motivate formula (12.114) in terms of the transition rates $\hat{\lambda}_{ij}$ in (12.35), rather than those in (12.113) that are adjusted for tunneling and fixation of alleles.

Since we assume $s_1 = \cdots = s_{m-1} = 1 < s_m$ in (12.114), it follows from (12.35) that it is increasingly difficult to have backward and forward transitions over larger distances, except that it is possible for some models to have a direct forward transition to the target allele m. By this we mean that the backward and forward transition rates from any state i satisfy $\hat{\lambda}_{i,i-1} \gg \cdots \gg \hat{\lambda}_{i0}$, and $\hat{\lambda}_{i,i+1} \gg \cdots \gg \hat{\lambda}_{i,m-1}$ respectively, as $N \to \infty$. For this reason, from any fixed state i, it is only possible to have competition between the two forward transitions $i \to i+1$ and $i \to m$ when $0 \le i \le m-2$. Since $\gamma_i = (\hat{\lambda}_{im}/\hat{\lambda}_{i,i+1})^2$, and since the transition rates to the intermediate alleles $i+1, \ldots, m-1$ are of a smaller order than the transition rate to $i+1$, it follows that (12.35) predicts a total forward rate of fixation from fixed state i of the order

$$\begin{aligned} Nu_{i+1} f_i \sim& \hat{\lambda}_{i,i+1} + \hat{\lambda}_{i,i+m} \\ =& \hat{\lambda}_{i,i+1}(1+\sqrt{\gamma_i}) \\ =& Nu_{i+1}\beta\left(\tfrac{s_{i+1}}{s_i}\right)(1+\sqrt{\gamma_i}) \\ =& u_{i+1}(1+\sqrt{\gamma_i}), \end{aligned} \tag{12.195}$$

where in the last step we used that $s_i = s_{i+1}$ and $\beta(1) = 1/N$. We will extend the argument in the proof of Theorem 3 in Durrett et al. [20], and indicate that the total forward rate of fixation from i should rather be

$$Nu_{i+1} f_i \sim \hat{\lambda}_{i,i+1}\chi\left(\frac{\gamma_i}{\beta(s_m)}\right) = u_{i+1}\chi\left(\frac{\gamma_i}{\beta(s_m)}\right), \tag{12.196}$$

where $\chi(\cdot)$ is the function defined in (12.63). This will also motivate (12.114), since this formula serves the purpose of modifying the incorrect forward rate of fixation (12.195), so that it equals the adjusted one in (12.196), keeping the relative sizes of the different forward rates $i \to j$ of fixation intact for $j = i+1, \ldots, m$.

The rationale for (12.196) is that type $i+1$ mutations arrive according to a Poisson process at rate Nu_{i+1}, and χ/N is the probability that any such type $i+1$ mutation has descendants of type $i+1$ or m that spread to the whole population. We need to show that

$$\chi = \chi\left(\frac{\gamma_i}{\beta(s_m)}\right). \tag{12.197}$$

To this end, let X_t be the fraction of descendants of a $i \to i+1$ mutation, Nt time units after this mutation appeared. We stop this process at a time point τ when X_t reaches any of the two boundary points 0 or 1 ($X_\tau = 0$ or 1), or when a successful mutation $i+1 \to i+2$ appears before that, which is a descendant of the type $i+1$ mutation that itself will have type m descendants who spread to the whole population, before any other type gets fixed ($0 < X_\tau < 1$). We have that $x = X_0 = 1/N$, but define

$$\bar{\beta}(s_m; x) = \bar{\beta}(x) = P(X_\tau = 0 | X_0 = x)$$

for any value of x. This is a non-fixation probability, i.e. the probability that the descendants of Nx individuals of type $i+1$ at time $t = 0$ neither have a successful type $i+2$ descendant, nor take over the population before that. Since the descendants of a single type $i+1$ mutation take over the population with probability $1 - \bar{\beta}(1/N)$, it is clear that

$$\chi = N\left[1 - \bar{\beta}\left(\frac{1}{N}\right)\right] \sim \lim_{x\to 0} \frac{1-\bar{\beta}(x)}{x} = -\bar{\beta}'(0). \tag{12.198}$$

Durrett et al. [20] prove that it is possible to neglect the impact of further $i \to i+1$ mutations after time $t = 0$. It follows that X_t will be a version of the Moran process of Appendix B with $s = s_{i+1}/s_i = 1$, during the time interval $(0, \tau)$, when time speeded up by a factor of N. Using (12.123)–(12.124), we find that the infinitesimal mean and variance functions of X_t are

$$\begin{aligned} M(x) =& N \times 0 = 0, \\ V(x) =& N \times 2x(1-x)/N = 2x(1-x), \end{aligned} \tag{12.199}$$

respectively. At time t, a successful type $i+2$ mutation arrives at rate

$$\begin{aligned} N \times NX_t \times u_{i+2} q_{i+1,m} \beta\left(\tfrac{s_m}{s_i}\right) \sim& N^2 X_t \times u_{i+2} r_{i+1,m} \beta(s_m) \\ =& N^2 X_t \times r_{im}^2 \beta(s_m) \\ =& X_t \times (\hat{\lambda}_{im}/\hat{\lambda}_{i,i+1})^2 \beta(s_m)^{-1} \\ =& X_t \times \gamma_i \beta(s_m)^{-1} \\ =:& X_t \times \gamma', \end{aligned} \tag{12.200}$$

where in the second step we used $r_{im}^2 = u_{i+2} r_{i+1,m}$, which follows from (12.36), since all $R(\rho_{ilj}) = 1$ when $s_1 = \cdots = s_{m-1} = 1$. Then in the third step we used $\hat{\lambda}_{im}/\hat{\lambda}_{i,i+1} = N r_{im} \beta(s_m)$, which follows from (12.35), and in the last step we introduced the short notation $\gamma' = \gamma_i \beta(s_m)^{-1}$. (One instance of γ' is presented for the boundary scenarios of Sect. 12.7.2.1, below formula (12.105).)

We will use (12.199)–(12.200) and Kolmogorov's backward equation in order to derive a differential equation for $\bar{\beta}(x)$. Consider a fixed $0 < x < 1$, and let $h > 0$ be a small number. Then condition on what happens during time interval $(0, h)$. When h is small, it is unlikely that the process X_t will stop because it hits any of the boundaries 0 or 1, i.e.

$$\begin{aligned} P(\tau < h, 0 < X_\tau < 1) =& x\gamma' h + o(h), \\ P(\tau < h, X_\tau \in \{0, 1\}) =& o(h) \end{aligned}$$

as $h \to 0$. The non-fixation probability can therefore be expressed as

$$\begin{aligned} \bar{\beta}(x) =& x\gamma' h \times 0 + (1 - x\gamma' h) \int_0^t \bar{\beta}(y) dP(X_h = y | X_0 = x) + o(h) \\ =& (1 - x\gamma' h) \left[\bar{\beta}(x) + \tfrac{1}{2} V(x) \bar{\beta}''(x) h \right] + o(h). \end{aligned}$$

Letting $h \to 0$, we find from (12.199) that $\bar{\beta}(x)$ satisfies the differential equation

$$x(1 - x)\bar{\beta}''(x) - x\gamma' \bar{\beta}(x) = 0. \tag{12.201}$$

Durrett et al. [20] use a power series argument to prove that the solution of (12.201), with boundary conditions $\bar{\beta}(0) = 1$ and $\bar{\beta}(1) = 0$, is

$$\bar{\beta}(x) = \frac{\sum_{k=1}^{\infty} \frac{(\gamma')^k}{k!(k-1)!} (1 - x)^k}{\sum_{k=1}^{\infty} \frac{(\gamma')^k}{k!(k-1)!}}. \tag{12.202}$$

Recalling (12.63) and that $\gamma' = \gamma_i / \beta(s_m)$, we deduce formula (12.197) from (12.198) and differentiation of (12.202) with respect to x. □

References

1. Asmussen, S., Nerman, O., Olsson, M.: Fitting phase-type distributions via the EM algorithm. Scand. J. Stat. **23**, 419–441 (1996)
2. Axe, D.D.: The limits of complex adaptation: an analysis based on a simple model of structured bacterial populations. BIO-Complex. **2010**(4) (2010)
3. Barton, N.H.: The probability of fixation of a favoured allele in a subdivided population. Genet. Res. **62**, 149–158 (1993)
4. Beerenwinkel, N., Antal, T., Dingli, D., Traulsen, A., Kinzler, K.W., Velculescu, V.W., Vogelstein, B., Nowak, M.A.: Genetic progression and the waiting time to cancer. PLoS Comput. Biol. **3**(11), e225 (2007)

5. Behe, M., Snoke, D.W.: Simulating evolution by gene duplication of protein features that require multiple amino acid residues. Protein Sci. **13**, 2651–2664 (2004)
6. Behe, M., Snoke, D.W.: A response to Michael Lynch. Protein Sci. **14**, 2226–2227 (2005)
7. Behrens, S., Vingron, M.: Studying evolution of promoter sequences: a waiting time problem. J. Comput. Biol. **17**(12), 1591–1606 (2010)
8. Behrens, S., Nicaud, C., Nicodéme, P.: An automaton approach for waiting times in DNA evolution. J. Comput. Biol. **19**(5), 550–562 (2012)
9. Bobbio, A., Horvath, Á., Scarpa, M., Telek, M.: A cyclic discrete phase type distributions: properties and a parameter estimation algorithm. Perform. Eval. **54**, 1–32 (2003)
10. Bodmer, W.F.: The evolutionary significance of recombination in prokaryotes. Symp. Soc. General Microbiol. **20**, 279–294 (1970)
11. Carter, A.J.R., Wagner, G.P.: Evolution of functionally conserved enhancers can be accelerated in large populations: a population-genetic model. Proc. R. Soc. Lond. **269**, 953–960 (2002)
12. Cao, Y., et al.: Efficient step size selection for the tau-leaping simulation method. J. Chem. Phys. **124**, 44109–44119 (2006)
13. Chatterjee, K., Pavlogiannis, A., Adlam, B., Nowak, M.A.: The time scale of evolutionary innovation. PLOS Comput. Biol. **10**(9), d1003818 (2014)
14. Christiansen, F.B., Otto, S.P., Bergman, A., Feldman, M.W.: Waiting time with and without recombination: the time to production of a double mutant. Theor. Popul. Biol. **53**, 199–215 (1998)
15. Crow, J.F., Kimura, M.: An Introduction to Population Genetics Theory. The Blackburn Press, Caldwell (1970)
16. Desai, M.M., Fisher, D.S.: Beneficial mutation-selection balance and the effect of linkage on positive selection. Genetics **176**, 1759–1798 (2007)
17. Durrett, R.: Probability Models for DNA Sequence Evolution. Springer, New York (2008)
18. Durrett, R., Schmidt, D.: Waiting for regulatory sequences to appear. Ann. Appl. Probab. **17**(1), 1–32 (2007)
19. Durrett, R., Schmidt, D.: Waiting for two mutations: with applications to regulatory sequence evolution and the limits of Darwinian evolution. Genetics **180**, 1501–1509 (2008)
20. Durrett, R., Schmidt, D., Schweinsberg, J.: A waiting time problem arising from the study of multi-stage carinogenesis. Ann. Appl. Probab. **19**(2), 676–718 (2009)
21. Ewens, W.J.: Mathematical Population Genetics. I. Theoretical Introduction. Springer, New York (2004)
22. Fisher, R.A.: On the dominance ratio. Proc. R. Soc. Edinb. **42**, 321–341 (1922)
23. Fisher, R.A.: The Genetical Theory of Natural Selection. Oxford University Press, Oxford (1930)
24. Gerstung, M., Beerenwinkel, N.: Waiting time models of cancer progression. Math. Popul. Stud. **20**(3), 115–135 (2010)
25. Gillespie, D.T.: Approximate accelerated simulation of chemically reacting systems. J. Chem. Phys. **115**, 1716–1733 (2001)
26. Gillespie, J.H.: Molecular evolution over the mutational landscape. Evolution **38**(5), 1116–1129 (1984)
27. Gillespie, J.H.: The role of population size in molecular evolution. Theor. Popul. Biol. **55**, 145–156 (1999)
28. Greven, A., Pfaffelhuber, C., Pokalyuk, A., Wakolbinger, A.: The fixation time of a strongly beneficial allele in a structured population. Electron. J. Probab. **21**(61), 1–42 (2016)
29. Gut, A.: An Intermediate Course in Probability. Springer, New York (1995)
30. Haldane, J.B.S.: A mathematical theory of natural and artificial selection. Part V: selection and mutation. Math. Proc. Camb. Philos. Soc. **23**, 838–844 (1927)
31. Hössjer, O., Tyvand, P.A., Miloh, T.: Exact Markov chain and approximate diffusion solution for haploid genetic drift with one-way mutation. Math. Biosci. **272**, 100–112 (2016)
32. Iwasa, Y., Michor, F., Nowak, M.: Stochastic tunnels in evolutionary dynamics. Genetics **166**, 1571–1579 (2004)

33. Iwasa, Y., Michor, F., Komarova, N.L., Nowak, M.: Population genetics of tumor suppressor genes. J. Theor. Biol. **233**, 15–23 (2005)
34. Kimura, M.: Some problems of stochastic processes in genetics. Ann. Math. Stat. **28**, 882–901 (1957)
35. Kimura, M.: On the probability of fixation of mutant genes in a population. Genetics **47**, 713–719 (1962)
36. Kimura, M.: Average time until fixation of a mutant allele in a finite population under continued mutation pressure: studies by analytical, numerical and pseudo-sampling methods. Proc. Natl. Acad. Sci. USA **77**, 522–526 (1980)
37. Kimura, M.: The role of compensatory neutral mutations in molecular evolution. J. Genet. **64**(1), 7–19 (1985)
38. Kimura, M., Ohta, T.: The average number of generations until fixation of a mutant gene in a finite population. Genetics **61**, 763–771 (1969)
39. Knudson, A.G.: Two genetic hits (more or less) to cancer. Nat. Rev. Cancer **1**, 157–162 (2001)
40. Komarova, N.L., Sengupta, A., Nowak, M.: Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. J. Theor. Biol. **223**, 433–450 (2003)
41. Lambert, A.: Probability of fixation under weak selection: a branching process unifying approach. Theor. Popul. Biol. **69**(4), 419–441 (2006)
42. Li, T.: Analysis of explicit tau-leaping schemes for simulating chemically reacting systems. Multiscale Model. Simul. **6**, 417–436 (2007)
43. Lynch, M.: Simple evolutionary pathways to complex proteins. Protein Sci. **14**, 2217–2225 (2005)
44. Lynch, M., Abegg, A.: The rate of establishment of complex adaptations. Mol. Biol. Evol. **27**(6), 1404–1414 (2010)
45. MacArthur, S., Brockfield, J.F.Y.: Expected rates and modes of evolution of enhancer sequences. Mol. Biol. Evol. **21**(6), 1064–1073 (2004)
46. Maruyama, T.: On the fixation probability of mutant genes in a subdivided population. Genet. Res. **15**, 221–225 (1970)
47. Maruyama, T., Kimura, M.: Some methods for treating continuous stochastic processes in population genetics. Jpn. J. Genet. **46**(6), 407–410 (1971)
48. Maruyama, T., Kimura, M.: A note on the speed of gene frequency changes in reverse direction in a finite population. Evolution **28**, 161–163 (1974)
49. Moran, P.A.P.: Random processes in genetics. Proc. Camb. Philos. Soc. **54**, 60–71 (1958)
50. Neuts, M.F.: Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. John Hopkins University Press, Baltimore (1981)
51. Nicodéme, P.: Revisiting waiting times in DNA evolution (2012). arXiv:1205.6420v1
52. Nowak, M.A.: Evolutionary Dynamics: Exploring the Equations of Life. Belknap Press, Cambridge (2006)
53. Phillips, P.C.: Waiting for a compensatory mutation: phase zero of the shifting balance process. Genet. Res. **67**, 271–283 (1996)
54. Radmacher, M.D., Kelsoe, G., Kepler, T.B.: Predicted and inferred waiting times for key mutations in the germinal centre reaction: evidence for stochasticity in selection. Immunol. Cell Biol. **76**, 373–381 (1998)
55. Rupe, C.L., Sanford, J.C.: Using simulation to better understand fixation rates, and establishment of a new principle: Haldane's Ratchet. In: Horstmeyer, M. (ed.) Proceedings of the Seventh International Conference of Creationism. Creation Science Fellowship, Pittsburgh, PA (2013)
56. Sanford, J., Baumgardner, J., Brewer, W., Gibson, P., Remine, W.: Mendel's accountant: a biologically realistic forward-time population genetics program. Scalable Comput.: Pract. Exp. **8**(2), 147–165 (2007)
57. Sanford, J., Brewer, W., Smith, F., Baumgardner, J.: The waiting time problem in a model hominin population. Theor. Biol. Med. Model. **12**, 18 (2015)
58. Schinazi, R.B.: A stochastic model of cancer risk. Genetics **174**, 545–547 (2006)

59. Schinazi, R.B.: The waiting time for a second mutation: an alternative to the Moran model. Phys. A. Stat. Mech. Appl. **401**, 224–227 (2014)
60. Schweinsberg, J.: The waiting time for *m* mutations. Electron. J. Probab. **13**(52), 1442–1478 (2008)
61. Slatkin, M.: Fixation probabilities and fixation times in a subdivided population. Evolution **35**, 477–488 (1981)
62. Stephan, W.: The rate of compensatory evolution. Genetics **144**, 419–426 (1996)
63. Stone, J.R., Wray, G.A.: Rapid evolution of *cis*-regulatory sequences via local point mutations. Mol. Biol. Evol. **18**, 1764–1770 (2001)
64. Tuğrul, M., Paixão, T., Barton, N.H., Tkačik, G.: Dynamics of transcription factor analysis. PLOS Genet. **11**(11), e1005639 (2015)
65. Whitlock, M.C.: Fixation probability and time in subdivided populations. Genetics **164**, 767–779 (2003)
66. Wodarz, D., Komarova, N.L.: Computational Biology of Cancer. Lecture Notes and Mathematical Modeling. World Scientific, New Jersey (2005)
67. Wright, S.: Evolution in Mendelian populations. Genetics **16**, 97–159 (1931)
68. Wright, S.: The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: Proceedings of the 6th International Congress on Genetics, vol. 1, pp. 356–366 (1932)
69. Wright, S.: Statistical genetics and evolution. Bull. Am. Math. Soc. **48**, 223–246 (1942)
70. Yona, A.H., Alm, E.J., Gore, J.: Random sequences rapidly evolve into *de novo* promoters (2017). bioRxiv.org, https://doi.org/10.1101/111880
71. Zhu, T., Hu, Y., Ma, Z.-M., Zhang, D.-X., Li, T.: Efficient simulation under a population genetics model of carcinogenesis. Bioinformatics **6**(27), 837–843 (2011)