J Stat Phys (2010) 138: 381–410
DOI 10.1007/s10955-009-9915-x

The Speed of Evolution in Large Asexual Populations

Su-Chan Park · Damien Simon · Joachim Krug

Received: 1 October 2009 / Accepted: 28 December 2009 / Published online: 14 January 2010

Abstract We consider an asexual biological population of constant size N evolving in discrete time under the influence of selection and mutation. Beneficial mutations appear at rate U and their selective effects s are drawn from a distribution $g(s)$. After introducing the required models and concepts of mathematical population genetics, we review different approaches to computing the speed of logarithmic fitness increase as a function of N, U and $g(s)$. We present an exact solution of the infinite population size limit and provide an estimate of the population size beyond which it is valid. We then discuss approximate approaches to the finite population problem, distinguishing between the case of a single selection coefficient, $g(s) = \delta(s - s_b)$, and a continuous distribution of selection coefficients. Analytic estimates for the speed are compared to numerical simulations up to population sizes of order 10^{300}.

Keywords Evolutionary dynamics · Wright-Fisher model · Clonal interference · Traveling waves

1 Introduction

The foundations of mathematical population genetics were established around 1930 in three seminal works of R.A. Fisher [19], J.B.S. Haldane [26] and S. Wright [62]. The achievement of these three pioneers is often referred to as the *modern synthesis*, because they resolved an apparent contradiction between Darwinian evolutionary theory, with its emphasis on minute

S.-C. Park · J. Krug (✉)
Institute for Theoretical Physics, University of Cologne, Köln, Germany
e-mail: krug@thp.uni-koeln.de

S.-C. Park
e-mail: psc@thp.uni-koeln.de

D. Simon
Laboratoire "Probabilités et modèles aléatoires", Université Pierre et Marie Curie, Paris, France
e-mail: damien.simon@normalesup.org

changes accumulating over long times, and the then recently rediscovered laws of Mendelian genetics, which showed that the hereditary material underlying these changes is intrinsically discrete. Like Ludwig Boltzmann faced with the problem of deriving the laws of continuum thermodynamics from atomistic models, Fisher, Haldane and Wright developed a statistical theory of evolution to explain how random mutational events occurring in single individuals result in deterministic adaptive changes on the level of populations. Not surprisingly, then, statistical physicists always have been, and are now increasingly attracted to the study of evolutionary phenomena in biology (see e.g. [4, 12, 18, 38]).

In this article we focus on a specific, rather elementary question in the mathematical theory of evolution, which was posed in the early days of the field and remains only partly understood even today: We ask how rapidly an asexually reproducing, large population adapts to a novel environment by generating and incorporating beneficial mutations. The question originates in the context of the Fisher-Muller hypothesis for the evolutionary advantage of sexual vs. asexual reproduction. Fisher [19] and H.J. Muller [42] pointed out that a disadvantage for asexual reproduction would arise in populations that are sufficiently large to simultaneously accommodate several clones of beneficial mutants. In the absence of sexual recombination, two beneficial mutations that have appeared in different individuals can be combined into a single genome only if the second mutation occurs in the offspring of the first mutant. This places a limit on the speed with which the population fitness increases in the asexual population.

The first quantitative treatment of the Fisher-Muller effect was presented by Crow and Kimura [8] for a model in which all beneficial mutations are assumed to have the same effect on the fitness of the individuals. They arrived at an expression for the speed of evolution in asexuals which saturates to a finite value in the limit of large population size $N \to \infty$, whereas for sexual populations the speed increases proportional to N. This conclusion was challenged by Maynard Smith [39], who showed (for a model with only two possible mutations) that recombination has no effect on the speed of adaptation in an infinite population. The resolution of the controversy [9, 17, 40] made it clear that the Fisher-Muller effect operates in large, but not in infinite populations; a first indication of the rather subtle role of population size, which will be a recurrent theme throughout this article.

Prompted by progress in experimental evolution studies with microbial populations [2, 14, 27, 48, 52, 57, 58], the question of the speed of evolution in the setting of Crow and Kimura has been reconsidered by several authors in recent years [3, 7, 10, 11, 50, 51, 64, 65]. Using a variety of approaches, they show that, rather than approaching a limit for large N, the speed grows as $\ln N$ in the regime of practical interest, reflecting the increasing spread of the population distribution along the fitness axis. Considerable efforts have been devoted to deriving accurate expressions for prefactors and sub-asymptotic corrections. At the same time more complex models that allow for a distribution of mutational effects have been introduced and analyzed [20, 22, 45, 61].

The purpose of this article is to review these developments on a level that is accessible to statistical physicists with no prior knowledge of population genetics. In the next section we therefore begin by introducing the basic concepts and models, primarily the discrete time Wright-Fisher model with mutations and selection. Section 3 is devoted to the dynamics of an infinitely large population. In this limit the dynamics becomes deterministic and can be solved exactly using generating function techniques. Although (as we will show) real populations operate very far from this limit, the infinite population behavior serves as a benchmark for the comparison with approximate results for finite populations, and it yields the important insight that a large population can be described as a traveling wave in fitness space [51, 55]. In Sect. 4 we review the main approaches to the finite population problem.

We provide simple derivations of reasonably accurate expressions for the speed of evolution, both for the case of a single type of beneficial mutations and for models with a distribution of mutational effects, which are compared to stochastic simulations over a wide range of population sizes. A preliminary view of the relationship between the two types of models is presented. Finally, in Sect. 5 we summarize the article and discuss some related topics which point to possible directions for future research.

2 Models

This section introduces the basic concepts and models studied in this paper. Models of evolving populations are based on three main features: reproduction with inheritance, natural selection, and mutation.[1] We describe each of these features from the point of view of stochastic processes in discrete time. For ease of explanation, our description begins with the branching process well-known in the statistical physics community.

2.1 Wright-Fisher Model

We consider here only asexual reproduction that is described by the number of offspring that each individual produces. This number is different from one individual to another, depends on many external events, and is thus described by a random variable. In the discrete time branching process without selection, an individual at time (or generation) t is replaced by n_{t+1} individuals at time $t+1$ where n_{t+1} is distributed according to a law $p(n)$ that is the same for all individuals[2] and is constant in time. The probability $p(0)$ can be seen as the death probability since the lineage of the individual disappears. The population is then completely described by its total size N_t. This stochastic process is known as the Bienaymé-Galton-Watson process and describes the growth and the death of a population without restriction on the size. Simple computations show that the average size grows as $\mathbb{E}(N_t) \propto (\bar{n})^t$ where $\bar{n} = \sum_n np(n)$ is the average number of children of one individual [16]. This simple system exhibits a transition to an absorbing (extinct) state as $\bar{n}$ varies. When $\bar{n} \leq 1$, the extinction will occur with probability 1. On the other hand, if $\bar{n} > 1$, the population grows exponentially with a finite probability. However, such a growth is not realistic because of limitations of the amount of food or resources in the environment.

In order to take this saturation effect (or environmental capacity) into account, we demand that the size of population remains constant, with a given value N. For generality, we now assume the mean number of offspring for each individual i $(1 \leq i \leq N)$ to be w_i in the unrestricted growth case described above, and we allow the w_i to be different from each other. To make the discussion concrete, we choose a Poisson distribution for the number of offspring of individual i, $p_i(n_i) = w_i^{n_i} e^{-w_i}/n_i!$. The reproduction mechanism at constant population size can then be modeled by conditioning the total number of offspring $M \equiv \sum_i n_i$ to be equal to N. The joint probability of the n_i (without restriction) is given by

$$\prod_{i=1}^{N} p_i(n_i) = e^{-N\bar{w}} \prod_{i=1}^{N} \frac{w_i^{n_i}}{n_i!}, \tag{1}$$

[1] Other important features this paper does not consider are migration and genetic recombination.

[2] The fact that all individuals have the same distribution law implies the absence of natural selection.

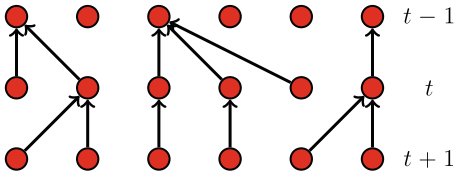

Fig. 1 A cartoon illustrating the Wright-Fisher model for a population of size $N = 6$ over three generations. The *arrows* indicate how an individual 'chooses' its parent

where $\bar{w} \equiv \sum_i w_i/N$ is the mean number of offspring per individual. The probability of observing $M = N$ is

$$\mathbb{P}(N) = \frac{(N\bar{w})^N}{N!} e^{-N\bar{w}}, \tag{2}$$

and, accordingly, the conditioned probability is given by (δ is the Kronecker delta symbol)

$$p(n_1, \ldots, n_N | N) = \frac{\delta_{MN}\, p(n_1) \cdots p(n_N)}{\mathbb{P}(N)} = \delta_{MN} \frac{N!}{n_1! \cdots n_N!} \prod_i \left(\frac{w_i}{N\bar{w}} \right)^{n_i}, \tag{3}$$

which is the widely-used Wright-Fisher (WF) model [19, 62]. It becomes then equivalent to the following process: at time $t + 1$, each individual 'chooses'[3] its parent i at time t with probability $w_i/(N\bar{w})$; see Fig. 1 as an illustration. It is obvious that this scheme is not affected by multiplying all w_i by a common factor. Inheritance is modeled by conferring to an offspring the same value of w_i as its parent.

However, the inherited genetic material may go through copying errors (or mutations), which can result in a child's having different characters from its parent. In order to take into account the effects of mutations, it is necessary to describe the characteristics of each individual. Individuals are usually characterized by a set of parameters, the *type* (either the *phenotype* that describes their biological functions and their interactions with their environment, or the *genotype* that specifies their heritable genetic material). A type is transmitted from the parent to the children up to some changes due to genetic mutations. For our purposes, the most important characteristic of an individual is its *fitness* defined as the average expected number of offspring of this individual (even if, for a given realization of the process, the effective number of offspring can be different because of the environmental capacity) in the whole population. In the reproduction scheme described above the absolute fitness of individual i is thus given by w_i, the relative fitness by $\chi_i = w_i/\bar{w}$, and the probability that individual i is chosen as a parent in the WF-model is χ_i/N. Fitness differences in the population imply selection: individuals with large fitnesses tend to generate larger fractions of the populations whereas lineages with small fitnesses tend to disappear quickly. We return to the question of how fitness is assigned to individuals below in Sect. 2.2.

In the language of statistical physics, the WF model as defined above may be seen as a mean-field model, because it does not take into account any spatial structure of the population: any individual can be the parent of any other, without any consideration of distance. This assumption is however realistic if one considers the mixing of real populations in a not-so-large environment. The role of spatial structures in evolution has also been studied for simple models, such as the island model [63] and the stepping stone model [33] which incorporate migration. The present paper focuses on mean-field reproduction models.

[3] In reality, of course, a child cannot choose its parent, but this usage of the terminology has no mathematical ambiguity and is widely used in the literature.

The WF model assumes a complete replacement of the population by children in one generation, i.e. generations do not overlap. A model with overlapping generations may be defined by splitting the replacement of the population over a longer time. A frequently used model that includes overlapping generations as well as a limited environmental capacity was introduced by Moran [41]: at each time step, one individual chosen at random is killed and replaced by the child of another individual chosen with probability χ_i/N. The time in the Moran model is still discrete, although the dynamics is evidently close to a scheme where single individuals are replaced in *continuous* time with rates proportional to the χ_i.

Both WF and Moran models have advantages and disadvantages. Unlike the WF model, the Moran model is amenable to some exact analysis, see Sect. 4.1 for an example. However, with regard to computational efficiency, the WF model is superior to the Moran model when simulating large populations. Since the conclusions relevant to biology are mostly insensitive to model details, we will base our discussion on the discrete time WF model, and comment on the corresponding continuous time or Moran model where appropriate.

2.2 Fitness Landscapes and Selection Coefficients

The main difficulty in modeling biological evolution within the framework described so far is the choice of the functional relationship $w(\mathcal{C})$ between the type $\mathcal{C}$ of an individual and its fitness, referred to as the *fitness landscape*, which encodes in a single parameter the complex interactions of a type with its environment [29]. At least two distinct approaches circumvent this difficulty: one can either try to measure the function $w(\mathcal{C})$ from experimental data if the set of types is reduced [56], or choose the fitness landscape at random from some suitable ensemble. In the last case, a widely-used further simplification consists in describing the individuals only by their fitnesses and ignoring the underlying structure of the types $\mathcal{C}$; mutations are then described only by changing the fitness of an individual by a random amount. This can be justified if the number of types is very large, so that every mutation effectively generates a new type that has never appeared before in the population. In population genetics this is known as the infinite number of sites approach [35, 46], and it will be used throughout this paper.

Each offspring has a probability U per generation of acquiring a mutation and this mutation changes the parental fitness w_i to the fitness w_i' of the offspring. In this paper, mutations are assumed to act multiplicatively on the fitness w_i and so the fitness w_i' after mutation is given by

$$w_i' = w_i(1+s) \tag{4}$$

where the *selection coefficient* s is a random variable with a distribution $g(s)$. Mutations with $s > 0$ are *beneficial* and those with $s < 0$ *deleterious*. Recall that if all the w_i are multiplied by the same quantity, then the relative fitnesses χ_i do not change, which justifies the multiplicative action of the mutations[4] [29]. One expects the relative fitnesses to reach a stationary distribution at long times such that the average fitness $\bar{w}(t)$ will increase (or decrease) exponentially with a rate referred to as the *speed of evolution*

$$v_N = \lim_{t\to\infty} \frac{\langle \ln \bar{w}(t)\rangle}{t}, \tag{5}$$

[4] An alternative scheme where the mutant fitness w_i' itself is chosen at random was investigated in [46], see Sect. 5 for further discussion.

where the angular brackets denote an average over all realizations. This speed depends on the population size N as well as on the mutation rate U and on the distribution $g(s)$ of the mutations. Two main contributions sum up to give the speed v_N: the change of mean fitness due to mutations and the selection pressure that selects individuals with larger w_i. For the WF model, these contributions are made explicit through a result obtained by Guess [23, 24]:

$$v_N = U \int \ln(1+s) g(s) ds + \left\langle \frac{1}{N} \sum_{i=1}^{N} (\chi_i - 1) \ln \chi_i \right\rangle_{\text{stat}}, \tag{6}$$

where $\langle \cdot \rangle_{\text{stat}}$ indicates an average over the stationary measure of the χ_i. The right hand side of (6) is obtained by computing the change in mean fitness over two consecutive generations, assuming that the initial fitnesses are drawn from the steady state distribution of the χ_i's. The existence of a steady state is guaranteed by the renewal property of the process, which reflects the fact that the population always returns to a genetically homogeneous state after a finite time $\sim N^N$ (see Ref. [24] for details).

The second term on the right hand side of (6) (which is always nonnegative) is related to selection. It is also the difficult part to study since the stationary distribution of the χ_i is generally unknown and hard to compute. If the distribution of relative fitness χ_i is concentrated around 1, the second term can be approximated by the variance of the distribution of relative fitness. This result is reminiscent of Fisher's fundamental theorem [19] which states that the speed of evolution is proportional to the variance of the fitness distribution.

In the present paper the dependence of the speed of evolution on the distribution $g(s)$ of selection coefficients is a central theme. As it turns out that deleterious mutations do not affect the adaptation of large populations when at least some fraction of mutations is beneficial,[5] only beneficial mutations ($s > 0$) will be considered in the following. The mutation rate (per generation) U then refers to the rate of *beneficial mutations*, which is exceedingly small in natural populations: experimental estimates for bacteria range from 10^{-7} to 10^{-4} [27, 48]. The distribution of selection coefficients of beneficial mutations is very difficult to determine experimentally, and the choice of a realistic form remains an open question [15]. Moreover, the experimental determination of evolutionary parameters such as the mutation rate U and the typical size of selection coefficients depends strongly on the assumptions made about the shape of $g(s)$ [27].

It has been argued that, because viable populations are already well adapted to their environment, fitness coefficients associated with beneficial mutations occur in the extreme high fitness tail of the underlying 'bare' fitness distribution, and therefore the shape of $g(s)$ should be given by one of the invariant distributions of extreme value statistics [30, 43]. Here we will consider two choices for this distribution. The first one (*model I*) describes the situation where all mutations have the same selective strength s_b,

$$g^{(\infty)}(s) = \delta(s - s_b). \tag{7}$$

The second class of distributions (*model II*) is supported on the whole non-negative real axis and decays as a stretched exponential [10, 20],

$$g^{(\beta)}(s) = (\beta/s_b)(s/s_b)^{\beta-1} \exp(-(s/s_b)^{\beta}), \tag{8}$$

[5] When all mutations are deleterious, the fitness decreases at constant speed and the problem is known as Muller's ratchet, see [28, 50, 51, 64] and references therein.

where the factor $(s/s_b)^{\beta-1}$ has been introduced for computational convenience. For $\beta = 1$, one recovers the widely-used exponential distribution [22, 43, 45, 61], whereas for $\beta \to \infty$ (8) reduces to (7). We note for later reference that the mean of $g^{(\beta)}$ is $\Gamma(1 + 1/\beta)s_b$. Typical values of selection coefficients obtained from evolution experiments with bacteria lie in the range $s_b \approx 0.01 - 0.05$ [27, 48]. Thus both U and s_b can be treated as small parameters, with $U \ll s_b$, in most of what follows.

3 Infinite Population Dynamics

This section studies the WF model in the infinite population limit which is described by a deterministic evolution equation. Some of the material of this section is also found in the online supporting information of [45]. Since it was shown in [45] that deleterious mutations do not contribute to the speed in the infinite population limit, all mutations are assumed to be beneficial in the following. The model will first be solved using a discrete set of fitness values, and the transition to a continuous fitness space will be performed in Sect. 3.3.2.

3.1 The Evolution Equation and Its Formal Solution

Let $f_t(n, k)$ denote the frequency of individuals with n (beneficial) mutations and with fitness e^{ks_0} at generation t; here $s_0 > 0$ and k is a non-negative integer. Note that $f_t(n, k)$ does not discern different types which have the same number of mutations and the same fitness. The restriction to fitnesses ≥ 1 is irrelevant due to the invariance of the dynamics under multiplication of absolute fitnesses by a common factor. The mean fitness of the population at generation t is

$$\bar{w}(t) = \sum_{n,k} e^{ks_0} f_t(n, k). \tag{9}$$

If there are no mutations, the frequency at the next generation is given by

$$\tilde{f}_{t+1}(n, k) = \frac{1}{\bar{w}(t)} e^{ks_0} f_t(n, k), \tag{10}$$

which is equal to the expected frequency at generation $t + 1$ for a finite population.

After reproduction, mutations can change the type of the offspring. With probability U, mutations hit an individual and with probability $1 - U$ the offspring keeps the type inherited from its parent. For simplicity, we assume that a single mutation occurs in a single mutation event (see [45] for more general cases). For each mutation a positive integer from a distribution $g_0(l)$ with strictly positive l is drawn and then the fitness of the offspring is that of its parent multiplied by e^{ls_0}. It is convenient to introduce the generating function of $g_0(l)$,

$$G(z) = \sum_{l=1}^{\infty} z^l g_0(l), \tag{11}$$

with the normalization $G(1) = 1$. Including the effect of mutations along with the selection step in (10), the frequency change becomes

$$f_{t+1}(n, k) = (1 - U)\tilde{f}_{t+1}(n, k) + U \sum_{l=1}^{k} \tilde{f}_{t+1}(n - 1, k - l) g_0(l), \tag{12}$$

which is the main equation to be analyzed in this section.

The generating function of the frequency

$$F_t(\xi, z) = \sum_{n,k} \xi^n z^k f_t(n,k) \tag{13}$$

satisfies

$$F_{t+1}(\xi, z) = \frac{F_t(\xi, ze^{s_0})}{F_t(1, e^{s_0})} \left[1 - U + U\xi G(z)\right], \tag{14}$$

where we have used the relations

$$\sum_{n,k} \xi^n z^k \tilde{f}_t(n,k) = \frac{F_t(\xi, ze^{s_0})}{F_t(1, e^{s_0})}, \tag{15}$$

$$\bar{w}(t) = F_t(1, e^{s_0}), \tag{16}$$

and the property of the convolution. Iterating (14) backwards until the initial time gives

$$F_t(\xi, z) = \frac{F_0\left(\xi, ze^{s_0 t}\right)}{F_0\left(1, e^{s_0 t}\right)} \prod_{\tau=0}^{t-1} \frac{1 + u\xi G\left(e^{s_0 \tau} z\right)}{1 + uG\left(e^{s_0 \tau}\right)}, \tag{17}$$

where $u = U/(1-U)$. One can check that (17) solves (14) by substitution.

3.2 General Asymptotic Behavior

Using (16) and (17), the mean fitness at generation t becomes

$$\ln \bar{w}(t) = \ln \frac{F_0(1, e^{s_0(t+1)})}{F_0(1, e^{s_0 t})} + \ln\left(1 - U + UG\left(e^{s_0 t}\right)\right). \tag{18}$$

If initially there is a finite K_0 such that $f_0(n,k) = 0$ for $k > K_0$, the first term arising from the initial condition saturates and does not contribute to the speed in the long time limit. On the other hand, if such a K_0 does not exit, the initial condition can affect the fitness increase indefinitely. For example, let $f_0(n,k) = \delta_{n,0} e^{-\eta} \eta^k / k!$ with the generating function $F_0(\xi, z) = e^{\eta(z-1)}$. The first term on the right hand side of (18) then becomes $\eta(e^{s_0} - 1)e^{s_0 t}$ which does not allow a finite increase rate even in the absence of mutations. This is a peculiarity of the selection dynamics in the infinite population limit and it is not difficult to understand why this happens. Since the selection confers exponential growth to all types with fitness larger than the average $\bar{w}(t)$ and there are always individuals of such types at any generation t due to the unbounded initial condition, the mean fitness can grow indefinitely without recourse to beneficial mutations. Because this is a rather artificial situation which has no biological relevance, we assume the existence of K_0 in what follows. Actually, for simplicity the initial condition

$$f_0(n,k) = \delta_{n0}\delta_{k0}, \qquad F_0(z, \xi) = 1 \tag{19}$$

will be used throughout this paper.

As $t \to \infty$, the speed is determined solely by the generating function of beneficial mutations. Let $K_{\max} = \max_k\{k : g_0(k) \neq 0\}$, then due to the exponential growth of the argument of $G(e^{s_0 t})$ the speed for the infinite size population becomes

$$v_\infty = K_{\max} s_0. \tag{20}$$

This shows that the mutation of largest effect governs the speed, which is not surprising because genetic drift (a term referring to the stochastic loss of a beneficial mutation in a finite population, see Sect. 4.1) is not operative. If we take $K_{\max} \to \infty$ with s_0 fixed, the speed diverges. Note that the speed in the infinite population limit does not depend on the mutation rate. It is only determined by the maximum value of the fitness increase by a single beneficial mutation event.

Another peculiarity of the infinite population limit is the possibility that the fitness becomes infinite at finite time. If $G(z)$ is not an entire function, the series defining the generating function has a finite radius of convergence, say $\mathcal{R}$, beyond which the series diverges. Hence when $e^{s_0 t} > \mathcal{R}$ or $t > \ln \mathcal{R}/s_0$, the mean fitness becomes infinite. For example, let $g(l) = (1-p)p^{l-1}$ which yield $G(z) = (1-p)z/(1-pz)$ for $pz < 1$ and infinite otherwise. Hence for $t > -\ln p/s_0$, the fitness becomes infinite. The radius of convergence for this example is $\mathcal{R} = 1/p$. Also note that the radius of convergence cannot be smaller than 1 because the generating function of probability is absolutely convergent for $|z| \le 1$ by definition. In the following, $G(z)$ is assumed to be an entire function, that is, $g_0(l)$ is assumed to decay faster than exponential in the asymptotic regime.

We now proceed to calculate the mean and variance of the number of accumulated mutations in the infinite population limit. First, the mean number of mutations is calculated as

$$\bar{n}(t) = \frac{\partial}{\partial \xi} \ln F_t(\xi, 1)\Big|_{\xi=1} = t - \sum_{\tau=0}^{t-1} \frac{1}{1 + uG(e^{s_0 \tau})}. \tag{21}$$

Since $G(e^{s_0\tau})$ grows at least exponentially with τ, the second term approaches a finite value. Clearly (21) gives the large population limit of the *substitution rate* k, defined here as the infinite time limit of $\bar{n}(t)/t$:

$$k = \lim_{t\to\infty} \frac{\bar{n}(t)}{t} = 1. \tag{22}$$

The variance of the number of mutations reads

$$\delta n(t)^2 = \left(\xi \frac{\partial}{\partial \xi}\right)^2 \ln F_t(\xi, 1)\Big|_{\xi=1} = \sum_{\tau=0}^{t-1} \frac{uG(e^{s_0\tau})}{[1 + uG(e^{s_0\tau})]^2}, \tag{23}$$

which has finite limit as $t \to \infty$.

3.3 Case Studies

Using the results presented above, we study the detailed evolution for two specific examples. To begin with, Sect. 3.3.1 studies the simple case that $g_0(l) = \delta_{l1}$ which corresponds to (7) with $s_b = s_0$. Then in Sect. 3.3.2 we generalize our solution to a continuous fitness distribution such as (8).

3.3.1 The Case of a Single Selection Coefficient

When $g_0(l) = \delta_{l1}$, the calculation is rather straightforward. Because the number of mutations fully specifies the fitness of a type, we replace $f_t(n, n)$ by $f_t(n)$ throughout this subsection. From (18), the mean fitness becomes

$$\ln \bar{w}(t) = \ln\left(1 - U + Ue^{s_0 t}\right) \approx s_0 \left(t - \frac{1}{s_0} \ln \frac{1}{U}\right) = s_0(t - t_0) \tag{24}$$

with

$$t_0 = \frac{1}{s_0} \ln \frac{1}{U}, \tag{25}$$

which gives $v_\infty = s_0$. The mean number of mutations in the long time limit can be calculated from (21) as

$$\bar{n}(t) \to t - \sum_{\tau=0}^{\infty} \frac{1}{1+ue^{s_0\tau}} \approx t - \int_0^\infty d\tau \frac{1}{1+ue^{s_0\tau}} = t - \frac{1}{s_0} \ln \frac{1+U}{U} \approx t - t_0, \tag{26}$$

where we approximate the summation by an integral assuming $s_0 \ll 1$, and $U \ll 1$. Not surprisingly, $s_0 \bar{n}(t) \approx \ln \bar{w}(t)$ in the long time limit. Likewise, the variance of the number of mutations is calculated as

$$\delta n(t)^2 \to \sum_{\tau=0}^{\infty} \frac{ue^{s_0\tau}}{(1+ue^{s_0\tau})^2} \approx \int_0^\infty d\tau \frac{ue^{s_0\tau}}{(1+ue^{s_0\tau})^2} = \frac{1-U}{s_0}. \tag{27}$$

Now we will show that the frequency distribution in the asymptotic limit can be approximated by a Gaussian. From (12) with $\bar{w} \approx e^{s_0(t-t_0)}$, the frequency at generation t can be approximated as

$$\begin{aligned} f_t(n) &\approx f_{t-1}(n) e^{s_0(n-(t-1)+t_0)} \\ &\approx f_n(n) \exp\left(s_0 \sum_{\tau=1}^{t-n} (n+t_0-t+\tau) \right) \approx f_n(n) e^{s_0 t_0^2/2} e^{-s_0(n-t+t_0)^2/2}, \end{aligned} \tag{28}$$

where n and t are assumed sufficiently large and we neglect the effect of mutations. Next we show that $f_n(n) \approx e^{-s_0 t_0^2/2}$ at long times, which concludes the demonstration that $f_t(n)$ becomes Gaussian. Under the assumptions of our model the largest number of mutations accumulated by an individual up to t is t, and from (17) the frequency of such individuals is

$$f_t(t) = \prod_{\tau=0}^{t-1} \frac{ue^{s_0\tau}}{1+ue^{s_0\tau}}. \tag{29}$$

Since $ue^{s_0 t}$ becomes larger than 1 at $t \approx t_0$, the term $ue^{s_0\tau}$ in the denominator of (29) makes a dominant (negligible) contribution for $t > t_0$ ($t < t_0$). Thus, we may approximate (29) in the long time limit as

$$\lim_{t\to\infty} f_t(t) \approx U^{t_0} e^{s_0 t_0 (t_0-1)/2} \approx e^{-s_0 t_0^2/2}, \tag{30}$$

which shows that $f_t(n)$ is well described by a traveling wave in the form of Gaussian.

The above consideration gives an interesting criterion for the population size beyond which the infinite population dynamics becomes valid. If the population size is larger than

$$N_c \equiv \exp(s_0 t_0^2/2) = \exp[\ln^2 U/(2s_0)], \tag{31}$$

the number of fittest individuals at a given generation is not smaller than 1 for all times t (note that $f_t(t)$ is a decreasing function of t). Since the selection coefficient of the types with t mutations compared to the mean fitness is approximately $e^{s_0 t}/\bar{w}(t) - 1 \approx 1/U \gg 1$, we can neglect the possible loss of such a type by genetic drift even if it is rare, which means

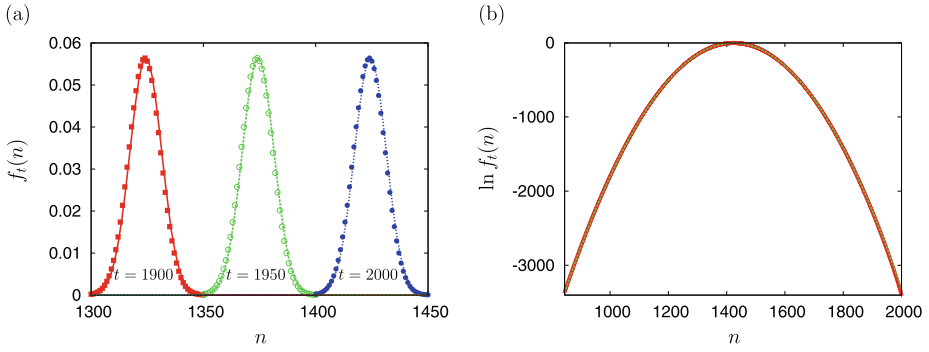

Fig. 2 (**a**) Frequency distribution of the infinite population dynamics for the case of $g_0(l) = \delta_{l1}$ with $s_0 = 0.02$ and $U = 10^{-5}$. The distributions are shown at $t = 1900$ (*left*), $t = 1950$ (*middle*), and $t = 2000$ (*right*). The peak is located at $t + (\ln U)/s_0 \approx t - 575.65$. (**b**) Plot of $\ln f_t(n)$ at $t = 2000$ as a function of n in comparison to (32). Only a tiny deviation around $n = 2000$ is visible

that the infinite population dynamics describes a finite population with $N \geq N_c$. To provide an impression of how large N_c is, we choose typical values $s_0 = 0.02$ and $U = 10^{-5}$, which gives[6] $N_c \approx 10^{1439}$.

To include the effect of mutations, we use (26) and (27) to write the frequency distribution in the form

$$f_t(n) \approx \frac{1}{\sqrt{2\pi(1-U)/s_0}} \exp\left(-\frac{(n-\bar{n}(t))^2}{2(1-U)/s_0}\right), \tag{32}$$

where the prefactor is fixed by normalization. Note that for sufficiently small U, (28) is consistent with (32). Figure 2 compares the numerically obtained frequency distribution with (32) for $U = 10^{-5}$ and $s_0 = 0.02$.

The idea that evolution can be described as a traveling wave moving at constant speed along the fitness axis was first presented by Tsimring et al. [31, 55], who considered the continuous time version of the model with multiplicative mutations and a single selection coefficient. In the continuous time case the speed of evolution diverges in the infinite population limit, because there is no bound on the number of mutations that a single individual can accumulate in a given time. A wave moving at finite speed is obtained only if the finite size of the population is introduced at least on the level of a lower cutoff on the frequency distribution.

3.3.2 Solution for a Continuous Fitness Space

In this section, we explain how the above calculation can be generalized to a continuous fitness space. We will use (8) for the distribution of selection coefficients. To connect to the results in Sect. 3.1, we perform a change of variables such that $e^x = 1 + s$, where x denotes the continuous version of ks_0. When s is drawn from $g^{(\beta)}(s)$, the probability density for x becomes

$$g_0^{(\beta)}(x) = \beta \left(\frac{e^x - 1}{s_b}\right)^{\beta-1} \exp\left(-\left(\frac{e^x - 1}{s_b}\right)^{\beta}\right) e^x \frac{1}{s_b}. \tag{33}$$

[6] The more accurate value obtained by exact numerical calculation is $N_c \approx 2 \times 10^{1477}$.

Setting $x = ls_0$, the corresponding discrete distribution is

$$g_0(l) = \mathcal{N}(s_0)\beta \left(\frac{e^{s_0 l} - 1}{s_b}\right)^{\beta-1} \exp\left(-\left(\frac{e^{ls_0} - 1}{s_b}\right)^{\beta}\right) e^{ls_0} \frac{s_0}{s_b}, \tag{34}$$

where $\mathcal{N}(s_0)$ is the normalization constant which approaches 1 as $s_0 \to 0$. We now follow Sect. 3.1, and calculate $G(e^{s_0 t})$ as

$$\begin{aligned} G(e^{s_0 t}) &= \mathcal{N}(s_0) \sum_{k=1}^{\infty} \beta \left(\frac{e^{s_0 k} - 1}{s_b}\right)^{\beta-1} \exp\left(-\left(\frac{e^{ks_0} - 1}{s_b}\right)^{\beta}\right) e^{ks_0(t+1)} \frac{s_0}{s_b} \\ &\longrightarrow \int_0^{\infty} dx \left(\frac{e^x - 1}{s_b}\right)^{\beta-1} \exp\left(-\left(\frac{e^x - 1}{s_b}\right)^{\beta}\right) e^{x(t+1)} \frac{\beta}{s_b} dx, \end{aligned} \tag{35}$$

where we take $s_0 \to 0$ with $s_0 k = x$ finite. Letting $y^{1/\beta} = (e^x - 1)/s_b$, the above integral, say W_t, becomes

$$W_t = \int_0^{\infty} (1 + s_b y^{1/\beta})^t e^{-y} dy = \int_0^{\infty} \exp\left(-y + t \ln(1 + s_b y^{1/\beta})\right) dy. \tag{36}$$

Using the steepest descent method, this can be approximated as

$$W_t \sim \exp\left(-y_c + t \ln\left(1 + s_b y_c^{1/\beta}\right)\right), \tag{37}$$

where y_c is the solution of the saddle point equation

$$y_c^{1-\frac{1}{\beta}} + s_b y_c = \frac{s_b t}{\beta} \quad \to \quad y_c \sim \frac{t}{\beta}. \tag{38}$$

Thus, the leading behavior of $\ln W_t$ becomes $t \ln t/\beta$, which along with (18) yields

$$\ln \bar{w}(t) \sim \ln W_t \sim \frac{t \ln t}{\beta} \tag{39}$$

for any finite β. On the other hand, for $\beta \to \infty$ (36) yields $W_t = (1 + s_b)^t$ and hence $\ln \bar{w}(t) \sim t$.

For $\beta = 1$, a more accurate approximation can be found in Ref. [45]. Here we observe that the speed $\ln \bar{w}(t)/t$ increases logarithmically with time irrespective of the value of β. The linear relation (21) for the mean number of beneficial mutations is still valid, because the summation on the right hand side of (21) approaches to a non-negative finite number in the continuum limit $s_0 \to 0$.

We conclude that the speed of evolution is infinite in the infinite population limit for distributions of selection coefficients like (8), which have unbounded support. Superficially this is reminiscent of the situation in the continuous time model with a single selection coefficient considered in [31, 55], but it is important to note that the reasons for the divergence of the speed are quite different in the two cases. In the continuous time setting the speed diverges because the *number* of mutations accumulated in a given time is unbounded, whereas in the discrete time model the divergence reflects that a *single* mutation can have an arbitrarily large effect. We will encounter a similar dichotomy in the discussion of the finite population dynamics in the next section.

4 Finite Populations

4.1 Genetic Drift, Fixation and Clonal Interference

Consider a single beneficial mutation with selection coefficient $s > 0$ which is introduced into an initially homogeneous population. Following the evolution of the population under WF dynamics without allowing for further mutations ($U = 0$), one can distinguish different time scales. The survival of the mutation during the first few generations is very fragile, due to the stochasticity of the reproduction process: the number of individuals carrying the mutation is small and the variance is of the same order as the mean. These fluctuations are called *genetic drift*. After this drift phase, either the mutation goes extinct (with some probability $1 - \pi_N(s)$) or the number of individuals becomes large enough (with probability $\pi_N(s)$), so that stochastic fluctuations can then be neglected and the evolution can be considered deterministic. A mutation that has reached the latter regime is called *established*[7] [17, 18, 40], and it will (in the absence of other mutations) eventually take over the entire population. This process is called *fixation*, $\pi_N(s)$ is the fixation probability, and the time needed for a mutation that survives to spread all over the population is the fixation time t_{fix}.

The fixation probability for the Moran model is given by [13]

$$\pi_N(s) = \frac{s}{1 + s - (1 + s)^{-(N-1)}}, \tag{40}$$

but for the Wright-Fisher model only approximate expressions are available [1, 54]. A widely used formula is [34]

$$\pi_N(s) = \frac{1 - e^{-2s}}{1 - e^{-2Ns}}. \tag{41}$$

For $s \to 0$ both (40) and (41) reduce to $\pi_N = 1/N$, as is obvious from a symmetry argument: When the fitness of the mutant is equal to that of the background population, the probability of fixation is the same for all N individuals. Both expressions show that the fixation of deleterious mutations ($s < 0$) is exponentially suppressed for large N, while the fixation probability for beneficial mutations becomes independent of N, reducing for (41) to

$$\pi_\infty(s) \equiv \pi(s) = 1 - e^{-2s}. \tag{42}$$

When s is small (as will often be the case) this can be further simplified to

$$\pi(s) \simeq 2s, \tag{43}$$

while $\pi(s) \simeq s$ for the Moran model.

In the limit $N \to \infty$ the restriction on the size of the growing mutant clone is irrelevant and the WF-model reduces to[8] a Bienaymé-Galton-Watson branching process with a Poisson offspring distribution of mean $1 + s$. The fixation probability is then equal to the survival probability of the branching process, which satisfies the implicit relation [1, 16, 25]

$$\pi = 1 - e^{-(1+s)\pi}. \tag{44}$$

[7] Although this terminology is mathematically ambiguous, it is widely used in the community because, we think, it is inspirational.

[8] The derivation of the WF model from the branching process in Sect. 2 easily explains this connection.

Expanding (44) to second order we recover (43) for small s, but for large s the exact fixation probability approaches unity as $1-\pi \approx e^{-(1+s)}$, in contrast to (42). In the following we nevertheless use (42) when values of the fixation probability are required for the full range of selection coefficients, and (43) when s is small.

The approximation by a branching process is also useful in deriving a heuristic estimate of the population size required for a mutant clone to become established [40]. In this approximation the average population size of the clone grows as $(1+s)^t \approx e^{st}$. However, since this average includes also instances where the clone goes extinct (with probability $1-\pi$), the population size conditioned on survival of the clone is larger by a factor $1/\pi \approx 1/2s$. Such a clone thus looks as if it started out containing already $\sim 1/2s$ individuals, which is precisely the threshold size separating stochastic from deterministic growth (see e.g. [7, 10] for a detailed treatment of this point).

In order to get some intuition about the fixation time $t_{\rm fix}$, one can look at the deterministic evolution of a mutant of type A that appears in a population consisting of the "wild type" B. The fitnesses can be taken as $w_A = (1+s)$ and $w_B = 1$. We assume that the type A has survived genetic drift and we have a frequency a_t of individuals of type A and $b_t = 1-a_t$ of individuals of type B. The deterministic evolution is thus given by

$$\begin{cases} a_{t+1} = \frac{1+s}{\bar{w}_t} a_t, \\ b_{t+1} = \frac{1}{\bar{w}_t} b_t, \\ \bar{w}_t = (1+s)a_t + b_t \end{cases} \tag{45}$$

and the solution is

$$b_t = \frac{b_0}{a_0(1+s)^t + b_0}, \qquad a_t = 1 - b_t. \tag{46}$$

For a finite population of large size N, the type B can be considered extinct when $b_t = 1/N$. With the initial condition of a single mutant, $a_0 = 1/N$, this expression gives the fixation time for large N as

$$t_{\rm fix} \simeq \frac{2\ln(N-1)}{\ln(1+s)} \simeq \frac{2\ln N}{s} \tag{47}$$

when s is small.

For later purposes, we also need the total number of individuals of type B that have existed during the fixation time of A. We note that

$$a_{t_{\rm fix}-t} = \frac{(1+s)^{t_{\rm fix}-t}}{(1+s)^{t_{\rm fix}-t} + N - 1} = \frac{N-1}{(1+s)^t + N - 1} = b_t \tag{48}$$

where we have used $(1+s)^{t_{\rm fix}} = (N-1)^2$. We can thus conclude that, during the fixation of type A, one has $\int a_t dt \simeq \int b_t dt$, so that the total number of individuals of type B is $\simeq N t_{\rm fix}/2 \simeq N \ln N / \ln(1+s)$.

This simple example shows the dependence of $t_{\rm fix}$ on N when the mutation rate U is set to 0 after the emergence of the mutant type A. If U is non-zero, the expression for $t_{\rm fix}$ is valid only as long as U is small enough, so that no new mutation emerges before the fixation of the previous one. The average time between two mutations that survive genetic drift is

$$t_{\rm mut} = \frac{1}{NU\pi(s)} \simeq \frac{1}{2NUs}. \tag{49}$$

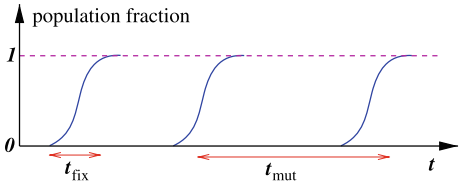

Fig. 3 In the periodic selection regime $t_{\text{mut}} \gg t_{\text{fix}}$ and beneficial mutations fix independently of each other. Each *blue line* represents a selective sweep

If $t_{\text{mut}} \gg t_{\text{fix}}$, i.e. if

$$N \ln NU \ll 1 \tag{50}$$

for s small, then no mutation interferes and we are in the *periodic selection regime* for which

$$v_N = \frac{s}{t_{\text{mut}}} \propto s^2 NU. \tag{51}$$

This situation is sketched in Fig. 3. The main feature of this regime is that the *selective sweeps* associated with different beneficial mutations are independent and well separated in time, and therefore the speed of evolution is directly proportional to the supply of beneficial mutations NU.

On the other hand, if t_{mut} and t_{fix} are of the same order, then mutations can occur during the fixation process of previous mutations [40, 61] and the distinction between t_{mut} and t_{fix} becomes unclear. In Fig. 4, we present an example showing how the population dynamics changes when the criterion (50) is violated. Following [22] we refer to the interaction among beneficial mutant clones in this regime as *clonal interference*.

In the remaining parts of this section we present the main analytic approaches that have been developed to compute the speed of evolution in the clonal interference regime. We begin by considering the case where all mutations have the same selection coefficient (*model I*) and then treat the case of a continuous distribution of selection coefficients[9] (*model II*).

4.2 Model I: Single Selection Coefficient of Beneficial Mutations

4.2.1 The Crow-Kimura-Felsenstein Approach

The first attempt to compute the speed of evolution in the presence of clonal interference is due to Crow and Kimura [8]. We present their calculation in the form given by Felsenstein [17], which takes into account that only established mutations contribute to the adaptation process. Such mutations (with selection coefficient s_b) appear in the population at rate $\pi(s_b)NU$. Assuming that a mutation was established at time $t = 0$, we now ask for the waiting time τ until a second mutation is established *in the offspring of the first*. We take s_b to be small, such that $\pi(s_b) \approx 2s_b$ and $(1 + s_b)^t \approx e^{s_b t}$. Then, according to (46), the frequency a_t of the mutant starting at $a_0 = 1/(2s_b N)$ is

$$a_t = \frac{1}{1 + (2s_b N - 1)e^{-s_b t}}. \tag{52}$$

The number of mutants at time t is Na_t, and each mutant generates an established second mutant with probability $2s_b U$ per generation. We therefore need to compute the accumulated

[9] Note that in part of the literature [7, 10, 50] the term "clonal interference" is restricted to model II.

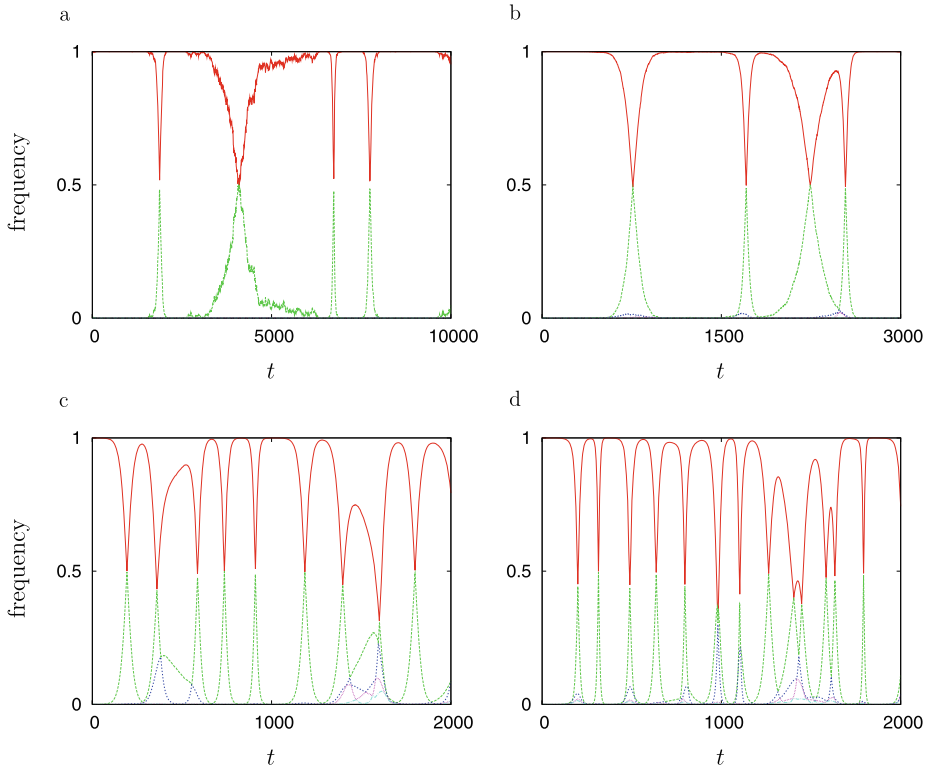

Fig. 4 The frequencies of the five most populated genotypes are shown in different colors for the WF model using the distribution (8) with $\beta = 1$, $U = 10^{-6}$, $s_b = 0.02$. The population sizes are (**a**) $N = 10^4$, (**b**) $N = 10^5$, (**c**) $N = 10^6$, and (**d**) $N = 10^7$, respectively. From $N = 10^5$ onward, where $NU \ln N \approx 1.15$, the third most populated genotype becomes visible and the distinction between t_{mut} and t_{fix} becomes blurred, which signals the onset of clonal interference

number of mutants N_{acc} that have existed up to time t, where each individual is weighted by the number of generations during which it has existed. Approximating the sum over generations by an integral, this is given by

$$N_{\text{acc}}(t) \simeq N \int_0^t dt'\, a_{t'} = \frac{N}{s_b} \ln\left[\frac{e^{s_b t}}{2Ns_b} + 1 - \frac{1}{2Ns_b}\right] \approx \frac{N}{s_b} \ln\left[\frac{e^{s_b t}}{2Ns_b} + 1\right] \tag{53}$$

for $Ns_b \gg 1$. The waiting time τ is then determined from the condition $2s_b U N_{\text{acc}}(\tau) = 1$, which yields the speed

$$v_N^{\text{CKF}} = \frac{s_b}{\tau} = \frac{s_b^2}{\ln[(2Ns_b)(e^{1/2UN} - 1)]}. \tag{54}$$

For small N ($UN \ll 1$) this reduces to the expression (51) valid in the periodic selection regime, while for large N a finite speed limit $v_\infty = s_b^2 / \ln(s_b/U)$ is reached.

In writing the relation (54) it is implicitly assumed that the situation at the appearance of the second mutation is identical to that at the appearance of the first, which is not true: the second mutation competes against a background consisting of a mixture of mutant and wild

type with mean fitness (relative to the wild type fitness of unity) $\bar{w} = (1+s_b)a_t + (1-a_t) = 1 + a_t s_b < 1 + s_b$. The selective advantage of the second mutant compared to the background population is therefore larger than s_b, and it will grow faster than the first mutant population. For this reason the expression (54) is a lower bound on the actual speed. To improve on this bound we need to take into account the coexistence of several mutant clones in the population, which will be the subject of the next subsection.

4.2.2 The Traveling Wave Approach

As the discussion in Sect. 3.3.1 shows, the deterministic evolution of an infinite population is well described as a traveling wave of approximately Gaussian shape. In order to extend this approach to large but finite populations, the deterministic dynamics of the bulk of the wave is combined with a stochastic description of the appearance of new mutants at the high-fitness edge of the frequency distribution. This idea was first proposed by Rouzine, Wakeley and Coffine [51] and has since been further elaborated [7, 10, 50]. In this section we follow the particularly simple and transparent derivation presented in [3].

Motivated by the analysis of Sect. 3.3.1, we denote by $f_t(n)$ the frequency of individuals with n mutations, and assume for this distribution the Gaussian form

$$f_t(n) \approx \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(n - v_N t/s_b)^2}{2\sigma^2}\right). \tag{55}$$

Here we have used that the mean number of mutations acquired up to time t is $\bar{n} = v_N t/s_b$. The speed v_N and the variance σ^2 of the traveling wave are related by Fisher's fundamental theorem or, more generally, by the Guess relation (6). Neglecting the direct mutation contribution $U \ln(1+s_b)$ because $U \ll 1$, and evaluating the selection term using the approximation $\chi_i \approx 1 + s_b(n_i - \bar{n})$ (where n_i is the number of mutations acquired by individual i) we see that

$$v_N \approx s_b^2 \sigma^2, \tag{56}$$

which is also true for the infinite population case when $U \ll 1$ (compare to (23)).

It is clear that at any finite time t, there is a maximal number of mutations $n_{\max}(t)$ such that $f_t(n) = 0$ for $n > n_{\max}(t)$. Let

$$L(t) \equiv n_{\max}(t) - \frac{1}{s_b} \ln \bar{w}(t) \tag{57}$$

denote the *lead* of this class of fittest individuals relative to the mean population fitness [10]. Let t_n be the generation when $n_{\max} = n$ for the first time. We assume that $\langle L(t_n)\rangle \to L_0$ as $t \to \infty$ and $\langle t_{n+1} - t_n\rangle \to \tau$, with constant L_0 and τ, which reflects the existence of a stationary traveling wave with speed

$$v_N = s_b/\tau. \tag{58}$$

For times $t_n < t < t_{n+1}$, $L(t)$ then behaves as $L_0 - \ln \bar{w}/s_b \simeq L_0 - v_N t/s_b$. We further assume that, for very large N, the lead satisfies $L(t)s_b \gg 1$, which implies that the loss by genetic drift of new mutants arising from the most fit class can be neglected. Analogous to Sect. 4.2.1, we can now compute the accumulated number of mutants in the most fit class that have existed during the time $t_n < t < t_{n+1}$ according to

$$N_{\text{acc}}(\tau) = \sum_{t=1}^{\tau} \exp\left(s \sum_{u=1}^{t} L(u)\right) = \sum_{t=1}^{\tau} e^{(L_0 s_b t - \frac{1}{2} v_N t^2)} \approx \frac{e^{L_0 s_b \tau} - 1}{1 - e^{-L_0 s_b}} \approx e^{L_0 s_b \tau}, \tag{59}$$

which is a good approximation if $v_N \tau = s_b \ll L_0 s_b$ or $L_0 \gg 1$.

The mean waiting time until the appearance of a mutant with $n_{\max} + 1$ mutations is the solution of the equation $N_{\rm acc}(\tau) U = 1$, which yields

$$\tau \approx \frac{1}{L_0 s_b} \ln\left(\frac{1}{U}\right). \tag{60}$$

Finally, we close the system of relations by noting that, as long as N is not extremely large,[10] the new fittest class will most likely appear as a single individual, which implies that

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{L_0^2}{2\sigma^2}} = \frac{1}{N} \quad \rightarrow \quad L_0 s_b = \left(2 v_N \ln \frac{N s_b}{\sqrt{2\pi v_N}}\right)^{1/2}. \tag{61}$$

From (56), (60), and (61), v_N becomes the solution of the equation

$$v_N \simeq \frac{2 s_b^2 \ln(\frac{N s_b}{\sqrt{2\pi v_N}})}{(\ln U)^2}, \tag{62}$$

which leads to the final result

$$v_N^{\rm Gauss} = \frac{2 s_b^2 \ln(N)}{(\ln U)^2} \tag{63}$$

for very large N. The logarithmic growth of the speed with N must saturate when the infinite population limit $v_N = s_b$ is approached. According to (63) this happens when $N \sim N_c \sim e^{(\ln U)^2/2 s_b}$, in agreement with the estimate (31). For population sizes exceeding N_c the relation (61) is no longer valid, because the initial frequency of the fittest genotype at t_n can be much larger than $1/N$ once $N \gg N_c$. The existence of an absolute speed limit $v_N = s_b$ is evident from (58), because τ cannot be smaller than one generation time in the discrete time model. For models with overlapping generations such a restriction does not exist, because a larger number of offspring can be generated within much less than an average generation time, and the speed increases proportional to $\ln N$ for arbitrary N.

In this context, it is instructive to compare the discrete and the continuous time models in different population size regimes. When the population size is small ($NU \ll 1$), there is a slight difference between these two models. For example, the fixation probability for small s is Cs with a model dependent constant C (compare (40) and (41)). Once the population becomes large enough so that the loss of the fittest type by genetic drift can be neglected, there is no difference between the continuous and discrete time models. However, for very large $N \geq N_c$, there is a large difference due to the restriction $\tau \geq 1$ in the discrete time model.

4.2.3 Comparison to Simulations

The above derivation of the speed of evolution involves a number of rough, uncontrolled approximations, such that the result (63) can hardly be expected to be quantitatively accurate. A much more careful analysis along similar lines was presented by Rouzine, Brunet

[10] More precisely, $N \ll N_c$.

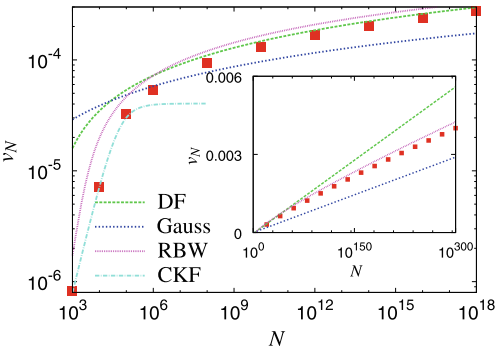

Fig. 5 Comparison of the theoretical expressions (65), (63), (64), and (54) with simulations using $s_b = 0.02$ and $U = 10^{-6}$. In the *inset* the comparison is extended up to $N = 10^{300}$ except for (54), which predicts a speed limit. The algorithm used to obtain these data is described in Appendix

and Wilke (RBW) [50], who find the implicit expression[11]

$$\ln N \approx \frac{v_N^{\mathrm{RBW}}}{2s_b^2}\left(\ln^2 \frac{v_N^{\mathrm{RBW}}}{eUs_b} + 1\right) - \ln\sqrt{\frac{s_b^3 U}{v_N^{\mathrm{RBW}} \ln(v_N^{\mathrm{RBW}}/(Us_b))}}. \tag{64}$$

A related approach, which however does not explicitly use the Gaussian shape of the deterministic part of the traveling wave, was presented by Desai and Fisher [10], who find[12]

$$v_N^{\mathrm{DF}} \approx \frac{2s_b^2 \ln N}{\ln^2(U/s_b)}. \tag{65}$$

In Fig. 5, we compare the different theories with simulation results for the WF model. For moderately large population size, (64) and (65) are of comparable quality, but for extremely large population, as shown in the inset of Fig. 5, the predictive power of (64) is superior to the other approaches.

In the asymptotic regime, (64) predicts that $v_N \sim \ln N / \ln^2 \ln N$, but (63) and (65) predict $v_N \sim \ln N$. Rigorous work [64, 65] has established that the speed in the asymptotic regime is not smaller than $\mathcal{O}(\ln^{1-\delta} N)$ for any positive δ, which does not exclude the possibility of a multiplicative $\ln^2 \ln N$-correction. Even with the dedicated algorithm used to generate the data in the inset of Fig. 5, it seems hardly possible to settle this issue using numerical simulations.

4.3 Model II: Continuous Distribution of Selection Coefficients

In Sect. 4.2, we reviewed theories aimed at calculating the speed of evolution when the selection coefficient takes a single value (model I). In this subsection, we will allow the selection coefficient to take a continuous range of values drawn from a distribution like (8) (model II). Unlike model I, two mutants arising from the same progenitor now have different selection coefficients and selection is operative between these two mutations. In contrast, in model I the competition between two mutants derived from a single progenitor is purely stochastic, and selection operates only between clones that have accumulated a different *number* of mutations.

[11] The speed V in Ref. [50] is v_N/s_b. To conform to our notation, we slightly modified (52) in Ref. [50].

[12] A detailed analysis of this approach can also be found in [7].

The qualitative picture of a wave of fixed shape traveling along the fitness axis that we developed for model I is expected to apply to model II as well, but it is more difficult to quantify, because continuous fitness cannot be reduced to the discrete number of acquired mutations. Two approaches have so far been proposed to deal with this problem. The first is related to an "equivalence principle" discovered in microbial evolution experiments [27], which suggests that a given distribution of selection coefficients can be represented by an effective single selection coefficient along with a suitably rescaled effective mutation rate. A heuristic scheme to implement this idea was given in [10] and tested against numerical simulations in [20]. As one might expect, the representation by a single "predominant" selection coefficient is quantitatively accurate only if the distribution is very narrow, such as $g^{(\beta)}$ with $\beta = 10$, and it fails completely when $\beta \leq 1$ [20].

The second approach, first proposed by Gerrish and Lenski (GL) [22], attempts to extend the periodic selection picture into the clonal interference regime by focusing on mutations of exceptionally large effect. Clonal interference is seen as a filter that eliminates mutations whose effect is small enough to be superseded by a mutation of larger effect arising later in the process. Once the size of selection coefficients of mutations that survive the competition by other clones has been identified, along with the rate at which such mutations appear, the speed of evolution is obtained from a simple relation similar to (51) used in the periodic selection regime.

In the following we outline the GL approach, derive its asymptotic predictions, and compare it to simulation results.

4.3.1 Gerrish-Lenski Theory

The GL-theory is based on two assumptions [22, 45]. First, the type of any individual at any time is either the wild type or a mutant derived directly from the wild type. The contributions from *multiple* mutations arising from an extant mutant are neglected. Since the fixation of a mutation under this assumption becomes a renewal process [21], we will refer to this assumption as the renewal assumption. Second, the loss of a beneficial mutation by stochastic sampling error when rare (genetic drift) is determined solely by its selection coefficient compared to the wild type. Other beneficial mutations do not play any role in determining the fate of the mutation at early times. We will refer to this assumption as the assumption of establishment.

The picture underlying these two assumptions is that the adaptive process can still be decomposed into separate selective sweeps in which a mutation grows in a fixed background and eventually takes over the population (compare to Fig. 3). A signature of this kind of dynamics is a step-like increase of the mean fitness. As can be seen in Fig. 4, this step-like behavior is pronounced for small populations in the periodic selection regime. However, as the population size increases, the mean fitness becomes more and more smooth, see Fig. 6, although distinct steps still occur when a mutation of exceptionally large strength appears. Thus, the GL approach is expected to be useful in a restricted range of population sizes, which goes slightly beyond the periodic selection regime. It is similar in spirit to the Crow-Kimura-Felsenstein approach reviewed in Sect. 4.2.1, which also successfully captures the slowing down of adaptation near the onset of clonal interference but fails for larger N (compare to Fig. 5).

To formulate the GL-theory quantitatively, we make use of two functions introduced previously: the probability distribution $g(s)$ of selection coefficients (like $g^{(\beta)}$ in (8)), and the probability $\pi(s)$ for the fixation (or, equivalently, the establishment) of a mutation of strength s. By the assumption of establishment, the distribution of the mutations that can

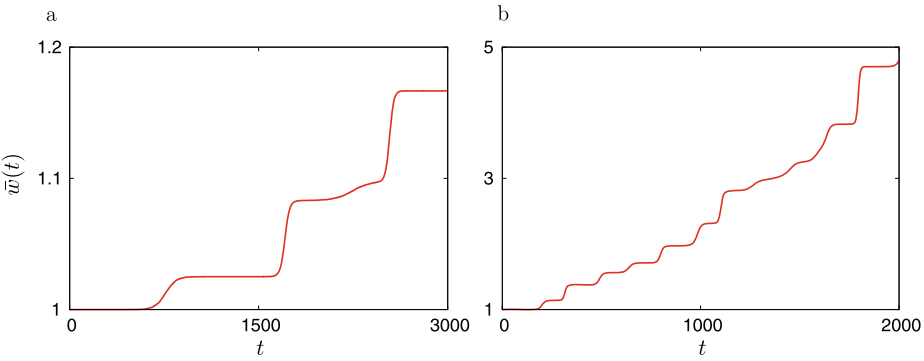

Fig. 6 Plots of mean fitness corresponding to the two panels on the right hand side of Fig. 4. The population sizes are (**a**) 10^5 and (**b**) 10^7. Although assumptions of the GL approach are not strictly applicable, one observes regions where the mean fitness behaves in a step-like fashion

spread in the population after the initial fluctuations and are really competing is then given by[13] $\pi(s)g(s)$.

For a mutation with selection coefficient s to be fixed, it is necessary that no fitter mutation is established during the time required for the first mutation to fix. The expected number of established fitter mutations that appear during this time is

$$\lambda(s) = (NUt_{\text{fix}}(s)/2)\int_s^\infty du\pi(u)g(u) \tag{66}$$

where[14] $t_{\text{fix}}(s)$ is given in (47), the factor $1/2$ comes from the renewal assumption [see also discussion below (48)], and the integral gives the probability that the selection coefficient of an established mutation is larger than s. Note that the renewal assumption prohibits a secondary mutation with selection coefficient s'' arising in the offspring of a primary mutation s' with $s' < s$ but $s' + s'' > s$, which would make (66) much more complicated. Hence, within the GL approximation the probability of not encountering any fitter mutation during fixation is $\exp(-\lambda(s))$ and, accordingly, the fixation probability of a mutation with selection coefficient s becomes

$$P_{\text{fix}}(s) = \pi(s)g(s)\exp\left(-\frac{NU\ln N}{\ln(1+s)}\int_s^\infty \pi(u)g(u)du\right). \tag{67}$$

In words, for a mutation with selection coefficient s to be fixed, it must first survive genetic drift (with probability $\pi(s)g(s)$), then should outcompete all other mutations (with probability $\exp(-\lambda(s))$). Thus, the substitution rate (the number of fixed mutations per generation) is

$$k_{\text{eff}} = NU\int_{s=0}^\infty P_{\text{fix}}(s)ds. \tag{68}$$

[13] Without the assumption of the establishment, the survival probability of a mutation should also depend on the population structure at the time when this mutation arises.

[14] Note that in previous work on the GL approach the expression $t_{\text{fix}} = 2\ln N/s$ was used irrespective of the size of s [22, 45, 61]. We will get back to the consequences of this replacement in Sect. 4.3.2.

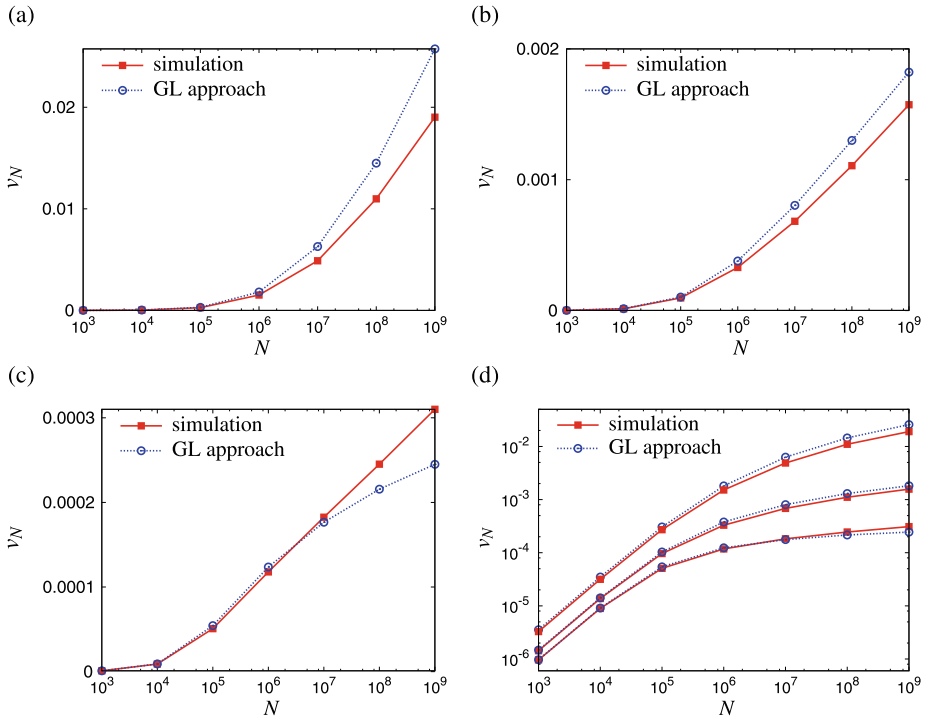

Fig. 7 Comparison of the GL theory with the simulation results of the WF model for $U = 10^{-6}$ and mean selection coefficient $\Gamma(1 + 1/\beta)s_b = 0.02$ for (**a**) $\beta = \frac{1}{2}$, (**b**) $\beta = 1$, and (**c**) $\beta = 2$. Panel (**d**) shows the data from (**a**)–(**c**) in double logarithmic scales

To calculate the speed v_N, we need the mean selection coefficient of fixed mutations which is readily obtained as

$$s_{\text{eff}} = \frac{\int s P_{\text{fix}}(s) ds}{\int P_{\text{fix}}(s) ds}. \tag{69}$$

Along with k_{eff} this determines the speed according to[15] [61]

$$v_N^{\text{GL}} = k_{\text{eff}} \ln(1 + s_{\text{eff}}). \tag{70}$$

In Fig. 7, we compare (70) to simulations using $g^{(\beta)}$ with three different values of β. The integrals in (67), (68) and (69) were evaluated numerically. We see that the GL approach is remarkably accurate also beyond the periodic selection regime, as becomes evident by comparing the double-logarithmic graph in Fig. 7(d) to the corresponding Fig. 5 for model I. However the deviations grow as N increases, in particular for $\beta = 2$, where the GL-prediction shows a negative curvature in $\ln N$ which is not present in the simulation data. We will return to this point at the end of the next subsection. As the numerical

[15] The reader may wonder why $\ln(1 + s_{\text{eff}})$ on the right hand side of (70) is not replaced by the average of $\ln(1 + s)$ with respect to $P_{\text{fix}}(s)$. For small s_{eff} the difference between the two is obviously negligible, but the same is true when $s_{\text{eff}} \gg 1$, because then P_{fix} becomes very narrow due to clonal interference. For a numerical test of (70) see [45].

scheme employed for model I relies on the discreteness of the fitness space (see Appendix), we have no information about the behavior of v_N for very large N. Since we have shown in Sect. 3.3.2 that the speed of evolution is infinite in the infinite population model, we merely know that (in contrast to model I) $\lim_{N\to\infty} v_N = \infty$.

4.3.2 Asymptotic Behavior of the GL Theory

Although we cannot analytically evaluate the expression for v_N predicted by the GL approach, an accurate asymptotic approximation can be derived, which is the topic of this section. Throughout the distribution $g^{(\beta)}$ of selection coefficients is used. The calculation follows the idea presented (for $\beta = 1$) in Ref. [45]; see also Ref. [61]. The only difference is that $t_{\text{fix}} = 2\ln N/\ln(1+s)$ is used rather than $2\ln N/s$, which will turn out to affect the conclusion significantly.

The integrations involved in the GL theory take the form

$$I[A;n] = \int_0^\infty ds s^n f(s)\exp(-Ah(s)), \tag{71}$$

where $f(s) = \pi(s)g^{(\beta)}(s)$, $h(s) = \int_s^\infty f(u)du/\ln(1+s)$, and $A = NU\ln N$. Note that $h(s)$ is a decreasing function with the range $[0,\infty]$. By the change of variable $y = h(s)$, the above integral becomes

$$I[A;n] = \int_0^\infty \Psi(y)e^{-Ay}dy, \tag{72}$$

where

$$\Psi(h(s)) = s^n\frac{f(s)}{|h'(s)|} = s^n\ln(1+s) + \frac{s^n}{1+s}\left(\frac{d}{ds}\ln h(s)\right)^{-1}. \tag{73}$$

To arrive at (72), we have used $f(s) = -(d/ds)(\ln(1+s)h(s))$ and the fact that $h(s)$ is a (monotonic) decreasing function. As $A\to\infty$, $I[A;n]$ is dominated by the contribution around $y = 0$, or equivalently around $s = \infty$. When s is very large, we can approximate $\pi(s)\approx 1$ and hence $h(s)\sim\exp(-(s/s_b)^\beta)/\ln s$. Hence for large s, we can approximate $s = s_b(-\ln y)^{1/\beta}$ $(y \ll 1)$, and

$$\Psi(y)\approx s^n\ln s - s^{n-1}\frac{s_b}{\beta}\left(\frac{s_b}{s}\right)^{\beta-1} \approx s_b^n(-\ln y)^{n/\beta}\ln(\ln(1/y))/\beta, \tag{74}$$

where we have kept only the leading order term. Hence

$$I[A;n]\approx\frac{s_b^n}{\beta}\int_0^\infty(-\ln y)^{n/\beta}\ln\ln(1/y)e^{-Ay}dy \approx \frac{s_b^n}{A\beta}(\ln A)^{n/\beta}\ln\ln A. \tag{75}$$

The substitution rate is then

$$k_{\text{eff}} = NUI[NU\ln N;0] = \frac{\ln\ln(NU\ln N)}{\beta\ln N}, \tag{76}$$

which, as $N\to\infty$, approaches 0 for any β. The asymptotic behavior of the speed is

$$v_N^{\text{GL}} = k_{\text{eff}}\ln\left(1+\frac{I[A;1]}{I[A,0]}\right)\sim\frac{(\ln\ln N)^2}{\beta^2\ln N}, \tag{77}$$

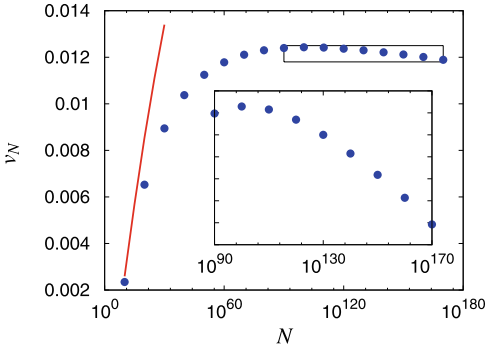

Fig. 8 Comparison of the GL theory using $t_{\rm fix} = 2\ln N/\ln(1+s)$ (*symbols*) with that using $t_{\rm fix} = 2\ln N/s$ (*line*) for $U = 10^{-6}$, $s_b = 0.02$, and $\beta = 1$. *Inset*: Close up of the boxed area. As anticipated by the asymptotic analysis, the speed decays, though slowly, with N

which also approaches 0 as $N \to \infty$.

This asymptotic behavior is easily understandable from an extremal statistics argument. The maximal mutation coefficient $s_{\rm max}$ observed over $\mathcal{M}$ mutation events is given approximatively as a solution of

$$\text{Prob}\{s > s_{\rm max}\} = \int_{s_{\rm max}}^{\infty} g(u)du = \exp\left(-(s_{\rm max}/s_b)^{\beta}\right) \simeq \frac{1}{\mathcal{M}}. \tag{78}$$

Following the GL hypothesis, the selection coefficient that gets fixed has to be the maximum of all selection coefficients that appear within its own fixation time, i.e. one has to consider a typical number

$$\mathcal{M} \sim NUt_{\rm fix} \sim NU\ln N/\ln(1+s_{\rm max}) \tag{79}$$

of mutations. Thus, the leading behavior of $s_{\rm max}$ becomes $s_{\rm max} \sim s_b \ln^{1/\beta} A \sim s_b \ln^{1/\beta} N$, the effective substitution rate is given by (up to leading order)

$$k_{\rm eff} \approx 1/t_{\rm fix}(s_{\rm max}) \sim \frac{\ln\ln(N)}{\beta \ln N} \tag{80}$$

as in (76), and the velocity is the same as in (77).

The asymptotic behavior obtained in this section is completely different from previous reports[16] [22, 45, 61]. The reason is clearly the factor $\ln(1+s)$ in the denominator of $t_{\rm fix}(s)$, which is very different from s when $s_{\rm eff} \gg 1$. However, this effect is only relevant when N is extremely large. As Fig. 8 shows, the true asymptotic behavior (77) is approached only when $N \gg 10^{100}$ for $U = 10^{-6}$ and $s_b = 0.02$ with $\beta = 1$, and the difference between using $\ln(1+s)$ and s in $t_{\rm fix}(s)$ is small when $N \leq 10^{20}$. So for realistic values of N, replacing $\ln(1+s)$ by s can provide a good approximation for the speed.

In fact, if the mutant fitness is derived from the parental fitness by multiplication with e^s rather than with $1+s$, which would correspond to a continuous time picture, the fixation time is $2\ln N/s$ for all s. The speed is then given by the expression $v_N = k_{\rm eff} s_{\rm eff}$ used in Ref. [22], rather than by (70). The leading asymptotic behavior of GL theory within this scheme can be obtained along the lines of [45] or, more directly, by adapting the extremal statistics argument given above. Since the leading behavior of $s_{\rm max}$ is the same as before, the

[16] Note that in the original paper of Gerrish and Lenski [22], it was erroneously concluded that v_N approaches a finite "speed limit" for $N \to \infty$.

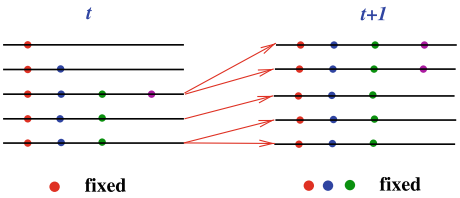

Fig. 9 Fixation of multiple mutations in a population of size $N = 5$. At time t, four types are present, and only the *red* mutation is fixed (= shared by all individuals). In the next generation, the individuals with one and two mutations leave no offspring, and consequently the *blue* and the *green* mutation go to fixation simultaneously

asymptotic behavior becomes

$$k_{\text{eff}} \sim s_{\max}/\ln N \sim s_b \ln^{1/\beta-1} N, \quad v_N^{\text{GL}} \sim s_b^2 \ln^{2/\beta-1} N. \tag{81}$$

Thus the graph of v_N versus $\ln N$ within GL theory is positively curved when $\beta < 1$ and negatively curved when $\beta > 1$, as is visible in Figs. 7(a)–(c). The simulation results for $\beta = 2$ do however not share this feature, and lie distinctly above the GL prediction for large N. In the next subsection we elaborate on this observation.

4.3.3 Importance of Multiple Mutations

It is instructive to compare (81) to the result $v_N \sim s_b^2 \ln N$ obtained for model I in Sect. 4.2. Evidently, the speed in model I should be minimal among all distributions $g(s)$ with the same mean selection coefficient, which implies that v_N should increase at least as fast as $\ln N$ also for model II.[17] However, according to (81) v_N grows more slowly than $\ln N$ when $\beta > 1$, and even decreases with N when $\beta > 2$. Moreover, the rate of substitution decreases with increasing $\ln N$ for $\beta > 1$, although we know that $k \to 1$ in the infinite population limit.

This is not really surprising, as the GL approach takes into account only the mutations of largest effect, ignoring the cumulative effect of multiple mutations of average effect which drive the dynamics in model I. On the basis of (81), one might speculate that the evolutionary process is dominated by large, extremal selection coefficients when $\beta < 1$, and by multiple mutations of typical effect when $\beta > 1$. This could also account for the breakdown of the "predominant mutation" approach for $\beta < 1$ [10, 20]. Interestingly, the exponential distribution of selection coefficients, which is most widely used in this context [22, 43, 45, 61], would then turn out to represent the marginal case separating the two regimes.

A quantitative measure of the importance of multiple mutations in the evolutionary dynamics can be obtained by asking how many mutations typically go to fixation in a single fixation event. The way in which the fixation of different mutations can become linked is illustrated in Fig. 9. It was observed numerically in [45] (for model II with $\beta = 1$) that the probability J_n of n mutations fixing in a single event is well described by a geometric distribution,

$$J_n = (1-q)^{n-1} q. \tag{82}$$

The left panel of Fig. 10 shows that the same relationship holds for model I. The parameter $1/q$ is the mean number of simultaneously fixed mutations, and it increases with N in a

[17] For the purpose of this discussion we ignore the saturation of the speed that occurs at extremely large N in model I.

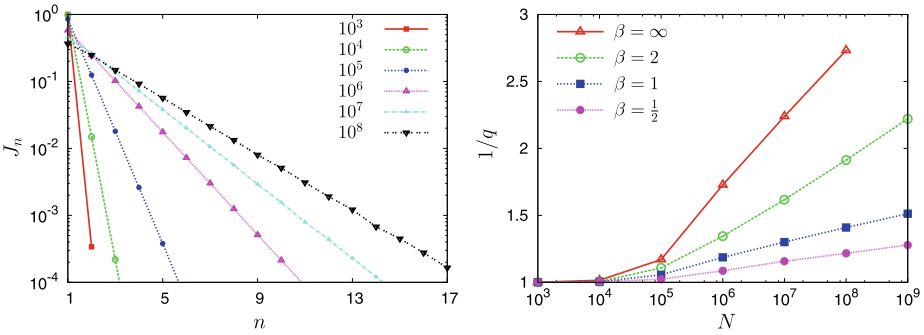

Fig. 10 *Left*: Distribution of the number of fixed mutations per fixation events for $\beta = \infty$ with $s_b = 0.02$ and $U = 10^{-6}$ in semi-logarithmic scales. From left to right, the population sizes are 10^3, 10^4, 10^5, 10^6, 10^7, and 10^8. Clean geometric distributions are observed. *Right*: Mean number of fixed mutations per fixation event $(1/q)$ as a function of population size

logarithmic fashion (right panel of Fig. 10). As expected, multiple mutations are more prevalent for larger β, but there does not seem to be any qualitative difference in the behaviors for $\beta < 1$ and $\beta > 1$. An analytic understanding of the relation (82) is so far only available for the case without selection, where $1/q$ increase linearly with population size N [59, 60]. We note, finally, that the time series of fixation events has interesting statistical properties [21, 45], which are however outside the scope of the present article.

5 Summary and Outlook

In this paper we have reviewed some aspects of evolutionary dynamics in the arguably simplest setting: A population of fixed size N evolving in a time-independent environment, supplied by independently acting beneficial mutations at a constant rate U. The quantity of primary interest is the speed v_N of logarithmic fitness increase, which is determined by the parameters N and U and by the probability distribution $g(s)$ of mutational effects with typical scale s_b.

On a qualitative level, one finds three distinct evolutionary regimes. For small populations, in the sense of (50), beneficial mutations are well separated in time and sweep through the population independently. As a consequence, the speed v_N is proportional to NU. For larger populations the clones generated by different mutations interfere and the increase of the speed is only logarithmic in N. Finally, in the limit of infinite populations, the speed saturates to a finite value (for the discrete time WF model). In the last regime the problem can be solved exactly, but, due to a conspiracy of the small parameters U and s_b, this description applies only to hyperastronomically large populations (see (31)). Real microbial populations of the kind used in evolution experiments typically operate in the intermediate clonal interference regime, which has been the main focus of the article.

Most work on the finite population problem has considered the case of a single selection coefficient (model I), where fitness is discrete. This offers considerable advantages for both approximate and rigorous analytic studies as well as for numerical simulations which are able to explore the asymptotic regime where $\ln N$ (and not just N) is large. A summary of the present state of affairs with regard to analytic approximations for the speed is given in Fig. 5. The case of a continuous distribution of selection coefficients (model II) is less well understood. Despite its conceptual shortcomings, the Gerrish-Lenski approximation

provides a quantitatively rather satisfactory description of the speed over the experimentally relevant range of population sizes (see Fig. 7), although it fails completely when the infinite population limit is approached. We have argued above that model I should provide a lower bound on the speed of evolution for general distributions of selection coefficients, which implies that the speed increases at least as fast as $\ln N$ also for model II, and possibly faster for distributions $g(s)$ that decay more slowly than exponentially.

The unifying paradigm used throughout the article is the description of the evolutionary process in terms of a traveling wave of constant shape moving along the fitness axis [51, 55]. This idea has proved to be successful also in the related but distinct context of competitive evolution, where selection is decoupled from reproduction [37, 47]. Competitive evolution models mimic a process of artificial (rather than natural) selection, where the character ("trait") of the types that is being selected is not the reproductive ability (fitness) of individuals. In one variant, individuals are assigned a scalar trait which is handed on to the offspring subject to random mutations. In one round of reproduction, each individual creates the same number of offspring, and subsequently the N with the values of the trait are selected for the next round [6]. This model falls into the large class of noisy traveling waves of Fisher-Kolmogorov type [5, 44, 53], which are much better understood than the problems described in the present article. Apart from accurate analytic approximations to the wave speed, also the genealogies of populations can be addressed, which display an interesting relation to the statistical physics of disordered systems [5]. In contrast, the genealogical properties of the WF model with selection are largely unknown.

Although the models described here are of considerable interest for the interpretation of evolutionary experiments [11, 14, 27, 48, 52, 55, 57], the reader should not be left with the impression that they provide a description that is realistic in all or even most respects. For example, the assumption of a constant supply of beneficial mutations cannot be true at arbitrarily long times, and indeed the rate of fitness increase is generally observed to slow down in experiments [2, 58]. One way to take this effect into account is by modeling the genotype as a sequence with a finite number of sites at which mutations can take place [32].

Another approach, known as Kingman's house-of-cards model [36], retains the infinite number of sites approximation but modifies the basic mutation step (4) such that the mutant fitness w_i' itself is drawn from a fixed fitness distribution $\tilde{g}(w)$. The probability of choosing a beneficial mutation with $w_i' > w_i$ then decreases as the mean fitness grows, and correspondingly the logarithmic fitness increases in a sublinear manner determined by the tail of $\tilde{g}$ [46]. In fact this problem turns out to be simpler than the one discussed in the present article, because the diminishing rate of beneficial mutations ($U \to 0$) drives the system into the periodic selection regime where selective sweeps can be treated as independent.

Kingman's assumption that the fitness of the offspring is uncorrelated with that of the parent is hardly more realistic than the assumption of independent fitness effects of different mutations which underlies (4). The few examples available so far indicate that real fitness landscapes lie between these two extremes [49, 56], which implies that the structure of the type space cannot be ignored. Like the modeling of evolutionary dynamics which we have discussed in this article, the mathematical characterization of such fitness landscapes offers a host of challenging problems that can be fruitfully explored by statistical physicists.

Acknowledgements This work was supported by DFG within SFB 680, and by the Alexander von Humboldt foundation through a fellowship to DS. We acknowledge the kind hospitality of the Lorentz Center, Leiden, where part of this paper was written.

Appendix: Simulating the Wright-Fisher model

This Appendix is devoted to explaining how we simulated model I for population sizes up to 10^{300}, as displayed in Fig. 5. The algorithm is based on that of [45], which we describe first. As in Sect. 3, we denote the frequency of individuals with fitness e^{ns_b} at generation t by $f_t(n)$. Assume that at time t there are $k+1$ distinct fitness values present in the population, i.e. $f_t(n) = 0$ if $n \leq n_0$ or $n > n_0 + k$. It is straightforward to see from (3) that the number m_i of individuals having $n_i \equiv n_0 + i$ $(i = 1, \ldots, k+1)$ mutations at generation $t+1$ is determined by the multinomial distribution

$$p(m_1, \ldots, m_{k+1}) = N! \prod_{i=1}^{k+1} \frac{p_i^{m_i}}{m_i!}, \tag{83}$$

where

$$p_i = f_t(n_i)(1-U)\frac{e^{n_i s_b}}{\bar{w}(t)} + f_t(n_i - 1)U\frac{e^{(n_i-1)s_b}}{\bar{w}(t)}. \tag{84}$$

Note that the effect of mutations is already implemented in the above algorithm, which is equivalent to the WF model in Sect. 2 (first selection then mutation). Since this multinomial distribution can be written as

$$p(m_1, \ldots, m_{k+1}) = \prod_{j=2}^{k+1} \binom{N_i}{m_i} (1-q_i)^{N_i - m_i} q_i^{m_i}, \tag{85}$$

where

$$q_i = \frac{p_i}{\sum_{j=1}^{i} p_j}, \tag{86}$$

$N_i = N_{i+1} - m_{i+1}$ and $N_{k+1} = N$, the multinomially distributed random numbers can be generated by drawing binomial random numbers k times. To be specific, we first draw m_{k+1} from the distribution

$$\binom{N}{m_{k+1}} (1-q_{k+1})^{N-m_{k+1}} q_{k+1}^{m_{k+1}}, \tag{87}$$

then the m_j are determined in the order of $j = k, k-1, \ldots, 2$ by the conditional distribution

$$\binom{N_j}{m_j} (1-q_j)^{N_j - m_j} q_j^{m_j}. \tag{88}$$

Finally, m_1 is given by $N_1 = N - \sum_{j=2}^{k+1} m_j$.

Since it is not possible to generate integers as large as 10^{100} in present day computers, in our simulations of very large populations we treat the m_j as real numbers. To be specific, we use the following algorithm. If $N_j < 10^9$, we generate binomially distributed integer random variables. If $N_j > 10^9$, we first check if the mean $\bar{m}_j \equiv N_j q_j$ is larger than prescribed number M which was set as 100 in our simulations.[18] If $\bar{m}_j < M$, we generate Poisson distributed random numbers with mean $\bar{m}_j$. Since N_j is sufficiently large and $q_j < 10^{-7}$, the Poisson distribution accurately approximates the binomial distribution in this situation. On

[18] The results do not depend on this choice.

the other hand, if $\bar{m}_j > M$, we invoked the central limit theorem to approximate the binomial distribution by a Gaussian; that is, $m_j = \bar{m}_j + \sqrt{N_j q_j (1-q_j)} N(0,1)$, where $N(0,1)$ is a normally distributed random number with mean 0 and variance 1.

Needless to say, the above algorithm is successful up to hyperastronomical population sizes because the fitness space is quantized and the number of possible fitness values at each generation, determined by the lead L_0, increases only as $\sim \ln N$. The direct application of this method to model II is not feasible, because in that case the number of different fitness values is at least NU.

References

1. Barrett, R.D.H., M'Gonigle, L.K., Otto, S.P.: The distribution of beneficial mutant effects under strong selection. Genetics **174**, 2071–2079 (2006)
2. Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., Kim, J.F.: Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature **461**, 1243–1247 (2009)
3. Beerenwinkel, N., Antal, T., Dingli, D., Traulsen, A., Kinzler, K.W., Velculescu, V.E., Vogelstein, B., Nowak, M.A.: Genetic progression and the waiting time to cancer. PLoS Comput. Biol. **3**, e225 (2007)
4. Blythe, R.A., McKane, A.J.: Stochastic models of evolution in genetics, ecology and linguistics. J. Stat. Mech.: Theory Exp. **P07018** (2007)
5. Brunet, E., Derrida, B., Mueller, A., Munier, S.: Phenomenological theory giving the full statistics of the position of fluctuating pulled fronts. Phys. Rev. E **73**, 056126 (2006)
6. Brunet, E., Derrida, B., Mueller, A., Munier, S.: Effect of selection on ancestry: An exactly soluble case and its phenomenological generalization. Phys. Rev. E **76**, 041104 (2007)
7. Brunet, E., Rouzine, I., Wilke, C.: The stochastic edge in adaptive evolution. Genetics **179**, 603–620 (2008)
8. Crow, J.F., Kimura, M.: Evolution in sexual and asexual populations. Am. Nat. **99**, 439–450 (1965)
9. Crow, J.F., Kimura, M.: Evolution in sexual and asexual populations: A reply. Am. Nat. **103**, 89–91 (1969)
10. Desai, M.M., Fisher, D.S.: Beneficial mutation-selection balance and the effect of linkage on positive selection. Genetics **176**, 1759–1798 (2007)
11. Desai, M.M., Fisher, D.S., Murray, A.W.: The speed of evolution and maintenance of variation in asexual populations. Curr. Biol. **17**, 385–394 (2007)
12. Drossel, B.: Biological evolution and statistical physics. Adv. Phys. **50**, 209–295 (2001)
13. Durrett, R.: Probability Models for DNA Sequence Evolution. Springer, Berlin (2002)
14. Elena, S.F., Lenski, R.E.: Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. Nat. Rev. Genet. **4**, 457–469 (2003)
15. Eyre-Walker, A., Keightley, P.D.: The distribution of fitness effects of new mutations. Nat. Rev. Genet. **8**, 610–618 (2007)
16. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. I. Wiley, New York (1950)
17. Felsenstein, J.: The evolutionary advantage of recombination. Genetics **78**, 737–756 (1974)
18. Fisher, D.S.: Evolutionary dynamics. In: Bouchaud, J.P., Mézard, M., Dalibard, J. (eds.) Complex Systems. Elsevier, Amsterdam (2007)
19. Fisher, R.A.: The Genetical Theory of Natural Selection. Clarendon Press, Oxford (1930)
20. Fogle, C.A., Nagle, J.L., Desai, M.M.: Clonal interference, multiple mutations and adaptation in large asexual populations. Genetics **180**, 2163–2173 (2008)
21. Gerrish, P.J.: The rhythm of microbial adaptation. Nature **413**, 299–302 (2001)
22. Gerrish, P.J., Lenski, R.E.: The fate of competing beneficial mutations in an asexual population. Genetica **102–103**, 127–144 (1998)
23. Guess, H.A.: Evolution in finite population with infinitely many types. Theor. Popul. Biol. **5**, 417–430 (1974)
24. Guess, H.A.: Limit theorems for some stochastic evolution models. Ann. Prob. **2**, 14–31 (1974)
25. Haldane, J.B.S.: A mathematical theory of natural and artificial selection, part V: Selection and mutation. Proc. Camb. Philos. Soc. **23**, 838–844 (1927)
26. Haldane, J.B.S.: The Causes of Evolution. Longmans Green, London (1932)
27. Hegreness, M., Shoresh, N., Hartl, D., Kishony, R.: An equivalence principle for the incorporation of favorable mutations in asexual populations. Science **311**, 1615–1617 (2006)

28. Jain, K.: Loss of least-loaded class in asexual populations due to drift and epistasis. Genetics **179**, 2125–2134 (2008)
29. Jain, K., Krug, J.: Adaptation in simple and complex fitness landscapes. In: Bastolla, U., Porto, M., Roman, H.E., Vendruscolo, M. (eds.) Structural Approaches to Sequence Evolution: Molecules, Networks and Populations. Springer, Berlin (2007)
30. Joyce, P., Rokyta, D.R., Beisel, C.J., Orr, H.A.: A general extreme value theory model for the adaptation of DNA sequences under strong selection and weak mutation. Genetics **180**, 1627–1643 (2008)
31. Kessler, D.A., Levine, H., Ridgway, D., Tsimring, L.: Evolution on a smooth landscape. J. Stat. Phys. **87**, 519–544 (1997)
32. Kim, Y., Orr, H.A.: Adaptation in sexuals *vs.* asexuals: clonal interference and the Fisher-Muller model. Genetics **171**, 1377–1386 (2005)
33. Kimura, M.: "Stepping stone" model of population. Ann. Rep. Natl. Inst. Genet. Jpn. **3**, 62–63 (1953)
34. Kimura, M.: On the probability of fixation of mutant genes in a population. Genetics **47**, 713–719 (1962)
35. Kimura, M.: The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics **61**, 893–903 (1969)
36. Kingman, J.F.C.: A simple model for the balance between selection and mutation. J. Appl. Prob. **15**, 1–12 (1978)
37. Kloster, M.: Analysis of evolution through competitive selection. Phys. Rev. Lett. **95**, 168701 (2005)
38. Lässig, M., Valleriani, A. (eds.): Biological Evolution and Statistical Physics. Springer, Berlin (2002)
39. Maynard Smith, J.: Evolution in sexual and asexual populations. Am. Nat. **102**, 469–473 (1968)
40. Maynard Smith, J.: What use is sex? J. Theor. Biol. **30**, 319–335 (1971)
41. Moran, P.A.P.: Random processes in genetics. Proc. Camb. Philos. Soc. **54**, 60–71 (1958)
42. Muller, H.J.: Some genetic aspects of sex. Am. Nat. **66**, 118–138 (1932)
43. Orr, H.A.: The distribution of fitness effects among beneficial mutations. Genetics **163**, 1519–1526 (2003)
44. Panja, D.: Effects of fluctuations on propagating fronts. Phys. Rep. **393**, 87 (2004)
45. Park, S.C., Krug, J.: Clonal interference in large populations. Proc. Natl. Acad. Sci. USA **104**, 18135–18140 (2007)
46. Park, S.C., Krug, J.: Evolution in random fitness landscapes: the infinite sites model. J. Stat. Mech.: Theory Exp. **P04014** (2008)
47. Peng, W., Gerland, U., Hwa, T., Levine, H.: Dynamics of competitive evolution on a smooth landscape. Phys. Rev. Lett. **90**, 088103 (2003)
48. Perfeito, L., Fernandes, L., Mota, C., Gordo, I.: Adaptive mutations in bacteria: High rate and small effects. Science **317**, 813–815 (2007)
49. Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., Tans, S.J.: Empirical fitness landscapes reveal accessible evolutionary paths. Nature **445**, 383–386 (2007)
50. Rouzine, I.M., Brunet, E., Wilke, C.O.: The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. Theor. Popul. Biol. **73**, 24–46 (2008)
51. Rouzine, I.M., Wakeley, J., Coffin, J.M.: The solitary wave of asexual evolution. Proc. Natl. Acad. Sci. USA **100**, 587–592 (2003)
52. Rozen, D.E., de Visser, J.A.G.M., Gerrish, P.J.: Fitness effects of fixed beneficial mutations in microbial populations. Curr. Biol. **12**, 1040–1045 (2002)
53. van Saarloos, W.: Front propagation into unstable states. Phys. Rep. **386**, 29 (2003)
54. Sella, G., Hirsh, A.E.: The application of statistical physics to evolutionary biology. Proc. Natl. Acad. Sci. USA **102**, 9541–9546 (2005)
55. Tsimring, L.S., Levine, H., Kessler, D.A.: RNA virus evolution via a fitness-space model. Phys. Rev. Lett. **76**, 4440–4443 (1996)
56. de Visser, J.A.G.M., Park, S.C., Krug, J.: Exploring the effect of sex on empirical fitness landscapes. Am. Nat. **174**, S15–S30 (2009)
57. de Visser, J.A.G.M., Rozen, D.E.: Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*. Genetics **172**, 2093–2100 (2006)
58. de Visser, J.A.G.M., Zeyl, C.W., Gerrish, P.J., Blanchard, J.L., Lenski, R.E.: Diminishing returns from mutation supply rate in asexual populations. Science **283**, 404–406 (1999)
59. Watterson, G.A.: Mutant substitution at linked nucleotide sites. Adv. Appl. Prob. **14**, 206–224 (1982)
60. Watterson, G.A.: Substitution times for mutant nucleotides. J. Appl. Prob. A **19**, 59–70 (1982)
61. Wilke, C.O.: The speed of adaptation in large asexual populations. Genetics **167**, 2045–2053 (2004)
62. Wright, S.: Evolution in Mendelian populations. Genetics **16**, 97–159 (1931)
63. Wright, S.: Isolation by distance. Genetics **28**, 114–156 (1943)
64. Yu, F., Etheridge, A.: Rate of adaptation of large populations. In: Pontarotti, P. (ed.) Evolutionary Biology from Concept to Application. Springer, Berlin (2008)
65. Yu, F., Etheridge, A., Cuthbertson, C.: Asymptotic behaviour of the rate of adaptation (2007). arXiv: 0708.3453v3 [math.PR]