Evolution in random fitness landscapes: the infinite sites model

Su-Chan Park and Joachim Krug

Institut für Theoretische Physik, Universität zu Köln, Zülpicher Strasse 77, D-50937 Köln, Germany
E-mail: psc@thp.uni-koeln.de and krug@thp.uni-koeln.de

Received 13 November 2007
Accepted 14 March 2008
Published 14 April 2008

Online at stacks.iop.org/JSTAT/2008/P04014
doi:10.1088/1742-5468/2008/04/P04014

Abstract. We consider the evolution of an asexually reproducing population in an uncorrelated random fitness landscape in the limit of infinite genome size, which implies that each mutation generates a new fitness value drawn from a probability distribution \( g(w) \). This is the finite population version of Kingman’s house of cards model (Kingman 1978 J. Appl. Probab. 15 1). In contrast to Kingman’s work, the focus here is on unbounded distributions \( g(w) \) which lead to an indefinite growth of the population fitness. The model is solved analytically in the limit of infinite population size \( N \rightarrow \infty \) and simulated numerically for finite \( N \). When the genome-wide mutation probability \( U \) is small, the long-time behavior of the model reduces to a point process of fixation events, which is referred to as a diluted record process (DRP). The DRP is similar to the standard record process except that a new record candidate (a number that exceeds all previous entries in the sequence) is accepted only with a certain probability that depends on the values of the current record and the candidate. We develop a systematic analytic approximation scheme for the DRP. At finite \( U \) the fitness frequency distribution of the population decomposes into a stationary part due to mutations and a traveling wave component due to selection, which is shown to imply a reduction of the mean fitness by a factor of \( 1 - U \) compared to the \( U \rightarrow 0 \) limit.

Keywords: models for evolution (theory), mutational and evolutionary processes (theory), slow relaxation and glassy dynamics
1. Introduction

A fruitful exchange of concepts and methods has taken place between evolutionary population biology and the statistical physics of disordered systems over the past several decades [1]–[3]. On the most basic level, one imagines that a biological population evolves by searching a high-dimensional landscape for fitness peaks, in much the same way as a disordered system relaxes towards its low energy configurations [4, 5]. Not surprisingly, extremal statistics arguments play a prominent role in both contexts [5]–[11]. To make the analogy more precise, we note that the inheritable characters of an individual (its genotype) are encoded in a genetic sequence (consisting of nucleotide letters or the alleles of genes), which for many purposes can be reduced to a binary sequence \( \sigma = (\sigma_1, \ldots, \sigma_L) \) of fixed length \( L \). For a statistical physicist it is very natural to assign the values \( \sigma_i = \pm 1 \) to the letters, and to treat the sequence as, for example, a row of spins in the two-dimensional Ising model [12] or a configuration of a quantum spin chain [13]. A fitness landscape is then a real-valued function \( W(\sigma) \) on the \( L \)-dimensional sequence space, analogous to the (negative) energy of the spin system.

The notion of a fitness landscape is a venerable and persistent image in evolutionary biology [14, 15], but it has always been plagued by a certain elusiveness, in the sense that very little is known about the fitness landscapes in which real organisms evolve. This situation may eventually change, as the experimental mapping of genotypic fitness
Evolution in random fitness landscapes: the infinite sites model

becomes feasible for simple microbial systems [16, 17]. Meanwhile it is reasonable to handle our ignorance of real fitness landscapes by treating $W(\sigma)$ as a realization of a suitably chosen ensemble of random functions. This approach was pioneered by Kauffman and coworkers [7, 8], who introduced the NK family of random fitness landscapes which are closely analogous to Derrida’s $p$-spin model of spin glasses developed a few years earlier [6, 18]. Two limiting cases of the model are of interest here: the random energy model (REM), in which fitness (or energy) is assigned randomly without correlations to the genotypes (or spin configurations), and the case in which the letters $\sigma_i$ contribute independently (multiplicatively for discrete time dynamics [15]) to the fitness. In the latter case there is always a single fitness maximum, which explains why this is also referred to as the Fujiyama landscape. In the evolutionary context deviations from multiplicative fitness are associated with epistasis [16]. Within the NK family, the REM landscape is maximally epistatic [19].

In previous work the evolutionary process in the REM landscape has been studied mostly in the limit of infinite population size, where fluctuations due to sampling noise (also known as genetic drift in population genetics) are ignored [15]. While this allows one to derive a rather complete picture of both stationary [20, 21] and time-dependent [22]–[25] properties of the model, the assumption of an infinite population is unrealistic, because the number of possible genotypes $2^L$ exceeds any conceivable population size $N$ already for moderate values of $L$. On the other hand, individual-based simulation studies which explicitly follow the population through sequence space are restricted to rather short sequences [26, 27].

In this contribution we therefore propose to perform the limit of infinite sequence length at finite population size $N$. This kind of limit is well known in population genetics, where it is viewed alternatively as a limit on the number of genetic loci or sites (the sequence length $L$ in our setting) or as a limit on the number of alleles (the number of possible values that the variables $\sigma_i$ can take). Although mathematically distinct, the two variants are equivalent for our purposes.

The implementation of the infinite sites limit for the REM landscape is straightforward: For $L \to \infty$ every mutation leads to a new genotype, whose fitness can be randomly generated without the need to keep track of the neighborhood relations of the sequence space. In this way large populations can be simulated efficiently for many generations. Moreover, we show that for long times and small mutation rates the model reduces to a simple point process which is partly tractable analytically. We also present an analytic solution of the infinite population limit of the model, which is useful for describing the evolution of finite populations at early times and provides important insights into the structure of the fitness distribution.

The infinite population version of our model was introduced by Kingman in 1978 [28] and is known as the house of cards model [29]. This term refers to the idea that the genetic organization of an organism is very fragile, such that it is completely disrupted by any change in the genome which therefore leads to a new fitness that is uncorrelated with the fitness of the parent genotype. Previous studies of the finite population dynamics of this model have been concerned mostly with the regime of weak selection, where a balance is established between deleterious and beneficial mutations [30]–[32]. In contrast, in the present paper we consider the strong selection regime where the population fitness increases without bound (see section 4.1 for a definition of these regimes).
We give a brief outline of this paper. In the next section we introduce the basic Wright–Fisher dynamics for asexually reproducing populations of constant size \( N \) and describe how to implement the infinite sites limit. The evolution of the fitness distribution in the deterministic limit \( N \to \infty \) is discussed in section 3. In section 4 we turn to the long-time behavior of finite populations. The key observation is that (for unbounded fitness distributions) fixation events in which a favorable mutation spreads in the population become rare and well separated as the mean fitness increases. The evolution then reduces to a simple point process closely related to the dynamics of records [9, 11, 33, 34]. We show that, to leading order, the mean population fitness achieved up to time \( t \) is given by the mutation of largest fitness that has been encountered, and we develop an analytic framework to compute sub-leading contributions to the mean fitness and the fitness variance. Finally, conclusions and some open problems are presented in section 5.

2. Wright–Fisher dynamics in the infinite sites limit

We consider a population of \( N \) individuals reproducing asexually, in discrete, non-overlapping generations. The basic Wright–Fisher (WF) dynamics [35, 36] of evolution can be described as follows. Each individual \( i \) is assigned fitness \( w_{i,t} \) \((i = 1, \ldots, N)\) at generation \( t \). Initially, all individuals have the same genotype and accordingly the same fitness \( w_{i,0} = 1 \). At every generation, all individuals are replaced. The probability that a new individual is an offspring of the parent \( i \) is proportional to the parent’s fitness \( w_{i,t} \). To be more accurate, this probability is \( w_{i,t} / \bar{w}_t N \), where \( \bar{w}_t = \sum_{i=1}^{N} w_{i,t} / N \) is the mean fitness\(^1\) at generation \( t \). In the actual simulation, we do not discern different progenitors if they have the same genotype. Instead, the number of progeny of a given genotype is determined from the multinomial distribution with a probability also proportional to the population of that genotype. The multinomial distributed numbers are chosen by sampling correlated binomial random numbers; see, e.g., [37, 38]. Since this reproduction scheme is invariant under the multiplication of a constant to the fitness of all individuals, setting the average initial fitness to unity implies no loss of generality. Once all individuals are replaced, a mutation can change the genotype of each one with probability \( U \). A cartoon illustrating the WF model is depicted in figure 1.

In the infinite sites limit, every genotype occurring by mutations can appear only once (there are no recurrent mutations [39]). In the REM landscape the fitness of the mutant is a random number \( w \) drawn from some fixed distribution \( g(w) \), independent of the number of sites affected by the mutation as well as of the parental genotype. In contrast, the multiplicative fitness landscape is represented in the infinite sites limit by drawing a random selection coefficient \( s \) and generating the new fitness \( w' \) from the parental fitness \( w \) by multiplication, \( w' = (1 + s)w \) [37], [40]–[42].

The model is completely specified by the population size \( N \), the mutation probability \( U \) and the choice of the mutation distribution \( g(w) \). To make contact with the evolutionary dynamics in a finite space of sequences of length \( L \) [27], we note that \( U \) is the probability

\(^1\) In this paper, we use three different notations for ‘mean fitness’. First, \( \tilde{w}_t \) is a random variable defined as the population average of the fitness in a specific realization of the WF model. Second, we use \( \langle \tilde{w}_t \rangle \) to denote the average of \( \tilde{w}_t \) over independent realizations; for infinite populations, this distinction is unnecessary. Third, by \( \bar{w}(\tau) \) we mean \( \langle \tilde{w}_t \rangle \) at \( t = \tau / (NU) \) for given \( N \) and \( U \) to emphasize the similarity between the simulation studies and the continuous time point process introduced in section 4.
that at least one mutation occurs in a genotype. We therefore have the relation

\[ U = 1 - (1 - \mu)^L \approx 1 - \exp(-\mu L), \]

where \( \mu \) is the mutation probability per site and the limit \( \mu \to 0, L \to \infty \) is implied. Typical values for \( U \) range from 0.0025 for \( E. \ coli \) to 0.15 for humans and 0.9985 for the bacteriophage \( Q\beta \) [15].

An important difference between the models with finite and infinite genome size is that, in the latter case, there are no local fitness maxima in which populations can get trapped when \( U \) is small [7, 8, 19, 27]. Hence we expect an indefinite increase of the population mean fitness \( \bar{w}_t \) when the fitness distribution \( g \) is unbounded. In the following sections we explore how the behavior of \( \bar{w}_t \) depends on the model parameters and the choice of \( g \). Some representative results from simulations with different values of \( N \) and \( U \) and the exponential mutation distribution

\[ g(w) = e^{-w} \]  \hspace{1cm} (2)

are shown in figure 2. The majority of our results were obtained with the choice (2).

3. Infinite populations

In this section, we consider the infinite population dynamics in the REM fitness landscape with an infinite number of sites. In this paper, we take the infinite sites \( L \to \infty \) limit first, followed by the infinite population limit \( N \to \infty \); note that these two limiting procedures do not obviously commute.
Evolution in random fitness landscapes: the infinite sites model

Figure 2. Mean fitness of the WF model for finite and infinite populations. For finite populations, the mean fitness is also averaged over independent runs. The mutation probability $U$ is set to 0.01 and the mutation distribution is $g(w) = e^{-w}$. From bottom to top, the population size $N$ increases ($N = 10^3, 10^5, 10^7$ and $10^9$). The number of independent runs is $8 \times 10^6$ ($N = 10^3$), $8 \times 10^5$ ($N = 10^5$), 20 000 ($N = 10^7$), and 4000 ($N = 10^9$). Left panel shows the short-time simulation results and compares them to the infinite population calculation in section 3. Right panel depicts log-linear plots of the mean fitness for finite populations which clearly shows a logarithmic increase of the mean fitness at long times.

3.1. Calculation of the mean fitness

Let $f_t(w) \, dw$ be the fraction of individuals in the population with a fitness between $w$ and $w + dw$. In the limit of infinite population size this evolves deterministically according to [28]

$$f_{t+1}(w) = (1 - U) \frac{w f_t(w)}{\bar{w}_t} + U g(w),$$

(3)

where $\bar{w}_t = \int_0^\infty dw \, w f_t(w)$ is the mean fitness. If $U = 1$, trivially and intuitively $f_t(w) \equiv g(w)$. The nonlinearity can be removed by introducing [15, 28] $h_t(w) = f_t(w) \prod_{k=0}^{t-1} \bar{w}_k$ with $h_0(w) = f_0(w)$ which satisfies

$$h_{t+1}(w) = (1 - U) w h_t(w) + U g(w) X_{t+1},$$

(4)

where

$$X_{t+1} = \prod_{k=0}^{t} \bar{w}_k = \int_0^\infty dw \, h_{t+1}(w).$$

(5)

The last equality is due to the fact that $f_t(w)$ remains normalized. Formally, we can solve the above equation such that

$$\frac{h_t(w)}{(1 - U)^t} = f_0(w) w^t + U g(w) \sum_{k=1}^{t} \frac{w^{t-k} X_k}{(1 - U)^k}.$$
From the self-consistency condition (5), one can find a recursion relation for $X_t$:

$$Y_t = \phi_t + U \sum_{k=1}^{t} Y_k G_{t-k},$$  \hspace{1cm} (7)

where $Y_t = X_t/(1 - U)^t$ and the quantities

$$\phi_t = \int_0^{\infty} dw f_0(w) w^t, \quad G_t = \int_0^{\infty} dw g(w) w^t$$  \hspace{1cm} (8)

are moments of the initial condition $f_0(w)$ and the fitness distribution $g(w)$, respectively. The recursion relation for $Y_t$ can be rewritten in the form

$$Y_t = \frac{\phi_t}{1 - U} + \frac{U}{1 - U} \sum_{k=1}^{t-1} Y_k G_{t-k},$$  \hspace{1cm} (9)

which is (at least numerically) solvable by iteration. Once we have calculated $Y_t$, the mean fitness follows naturally from

$$\bar{w}_t = (1 - U) \frac{Y_{t+1}}{Y_t}.$$  \hspace{1cm} (10)

Although we cannot find a complete analytic expression for $\bar{w}_t$, the leading asymptotic behavior can be easily found for some cases. If, for large $t$

$$G_{t-1} \gg \phi_t \quad \text{and} \quad G_t \gg G_k G_{t-k} \quad \text{for all} \quad 1 \leq k \leq t - 1,$$  \hspace{1cm} (11)

one can say with small error that

$$Y_t \simeq \frac{U}{1 - U} Y_{t-1} \Rightarrow \bar{w}_t \simeq (1 - U) \frac{G_t}{G_{t-1}}.$$  \hspace{1cm} (12)

The criterion (11) is actually not very restrictive. If $g(w)$ has the form of a stretched exponential multiplied by a power law

$$g(w) = \frac{\beta}{w_0 \Gamma((\nu + 1)/\beta)} \left( \frac{w}{w_0} \right)^\nu \exp \left( - \left( \frac{w}{w_0} \right)^\beta \right),$$  \hspace{1cm} (13)

where $w_0$ is a constant and $\Gamma(x)$ is the gamma function, the $t$th moment of $g(w)$ becomes

$$G_t = w_0^\nu \Gamma \left( \frac{\nu + t + 1}{\beta} \right) \Gamma \left( \frac{\nu + 1}{\beta} \right)^{-1},$$  \hspace{1cm} (14)

which satisfies (11). Using Stirling’s formula $\Gamma(z) \sim z^z e^{-z} \sqrt{2\pi/z}$, one finds that

$$\frac{\bar{w}_t}{w_0(1 - U)} \approx \left( \frac{\nu + t + 1}{\beta} \right)^{1/\beta} \left( \frac{t}{\beta} \right)^{1/\beta},$$  \hspace{1cm} (15)

a result that was also reported by Kingman [28].

When $\beta \to \infty$, one might expect that the power law (15) turns into a logarithmic increase of the fitness. As an example of an unbounded distribution that decays more
Evolution in random fitness landscapes: the infinite sites model

Figure 3. Log–log plots of mean fitness in the infinite population limit as a function of generation \( t \) for exponential \((\nu = 0\) and \(\beta = 1\) in (13); upper three datasets), Gaussian \((\nu = 0\) and \(\beta = 2\) in (13); middle three datasets) and Gumbel-type distributions (16) (lower three datasets). For all datasets, \( w_0 \) and \( f_0(w) \) are set to 1 and \( \delta(w-1) \), respectively. For each case, the mutation probability is \( U = 0.5 \) (red), \( U = 0.1 \) (purple) and \( U = 0.01 \) (blue) from bottom to top. The lines are the approximate solutions (15) and (A.5) with the respective parameters.

rapidly than (13), we consider in appendix A the Gumbel-type mutation distribution

\[
g(w) = \frac{e}{w_0} \exp \left( \frac{w}{w_0} - e^{w/w_0} \right).
\]  

(16)

Figure 3 compares the exact behavior obtained by iterating (9) to the asymptotic expression (15) and (A.5) for the exponential, Gaussian and Gumbel-type distributions. In this comparison, the initial fitness distribution is set to \( f_0(w) = \delta(w-1) \) and \( w_0 = 1 \). The analytic approximations are seen to be very accurate in all cases, with error less than 1% after 100 generations.

Up to now, we have assumed that \( G_t \) is finite for any \( t \). However, if \( g(w) \) has a power law tail such as

\[
g(w) = \alpha (1 + w)^{-(\alpha+1)},
\]

(17)

\( G_t \) becomes infinite at \( t = t_0 = \{\alpha\} \) where \( \{x\} \) means the smallest integer not smaller than \( x \). This means that \( Y_{t_0} \) is finite but \( Y_{t_0+1} \) is infinite, i.e. \( \bar{w}_t \) becomes infinite at \( t = t_0 \). This peculiarity of the deterministic selection dynamics has also been observed in the multiplicative case [37] and it is consistent with the behavior of the infinite population dynamics in a finite sequence space, where the population reaches the global fitness maximum in a single time step for large \( L \) [23].

For mutation distributions \( g(w) \) with bounded support the recursion (3) approaches a limiting distribution \( f_\infty(w) \) which has been described in detail by Kingman [28]. Remarkably, for certain choices of \( g(w) \) one finds a condensation phenomenon in which

doi:10.1088/1742-5468/2008/04/P04014
\( f_\infty(w) \) develops a \( \delta \)-function singularity at the maximally possible fitness \( w_{\text{max}} \) (set by the upper limit of the support of \( g \) or of \( f_0 \), whichever is larger). The asymptotic mean fitness is bounded from below by \((1 - U)w_{\text{max}}\). In the remainder of the paper we will restrict the discussion to unbounded mutation distributions.

### 3.2. Fitness distribution

Note that the mutation rate (provided it is nonzero) enters (15) only through the prefactor \( 1 - U \), which is known in population genetics as the \textit{mutational load} [29]. Its origin can be traced back to the fact that, in each generation, the fitness is reset randomly for a fraction \( U \) of the population, so that selection can act only on the remaining fraction \( 1 - U \) (compare to (3)).

To clarify this effect in a quantitative manner, let us find out what the frequency distribution looks like in the asymptotic regime. To this end, we calculate higher moments of \( f_t(w) \) from (3) along with (12). For convenience, let

\[ \zeta_n(t) \equiv \frac{1}{(1 - U)w_0^n} \int w^n f_t(w) \, dw. \quad (18) \]

A recursion relation for the moments can be found by multiplying by \( w^{n-1} \) both sides of (3) and integrating over \( w \), which yields

\[ \zeta_n(t) = \zeta_1(t) \left( \zeta_{n-1}(t+1) - \frac{U}{1 - U} \frac{G_{n-1}}{w_0^{n-1}} \right), \quad (19) \]

and \( \zeta_1(t) \) can be read off from (12) or (15). The formal solution for \( \zeta_n(t) \) is

\[ \zeta_n(t) = \prod_{k=0}^{n-1} \zeta_1(t + k) - \frac{U}{1 - U} \sum_{k=1}^{n-1} \frac{G_k}{w_0^n} \prod_{\ell=0}^{n-k-1} \zeta_1(t + \ell). \quad (20) \]

Since \( w_0 \zeta_1(t) \approx G_t / G_{t-1} \) for large \( t \), \( \zeta_n(t) \) becomes

\[ (1 - U)w_0^n \zeta_n(t) \approx U G_n + (1 - U) \frac{G_{t+n-1}}{G_{t-1}} - U \sum_{k=1}^{n} \frac{G_k G_{t+n-1-k}}{G_{t-1}}. \quad (21) \]

Constructing the Laplace transform (or the moment generating function) of the frequency distribution using (21) such that

\[ \int e^{-zw} f_t(w) \, dw \approx U \tilde{g}(z) + (1 - U) \hat{\varphi}_t(z) - U \hat{\psi}_t(z), \quad (22) \]

where

\[ \tilde{g}(z) = \sum_{n=0}^{\infty} \frac{(-z)^n}{n!} G_n = \int dw \, e^{-zw} g(w) \quad (23) \]

is the Laplace transform of \( g(w) \) and

\[ \hat{\varphi}_t(z) = \sum_{n=0}^{\infty} \frac{(-z)^n}{n!} \frac{G_{t+n-1}}{G_{t-1}}, \quad \hat{\psi}_t(z) = \sum_{n=1}^{\infty} \frac{(-z)^n}{n!} \sum_{k=1}^{n} \frac{G_k G_{t+n-1-k}}{G_{t-1}}, \quad (24) \]
we can in turn find the frequency distribution through inverse Laplace transformation. Since the Laplace transformation is linear, \( f_t(w) \) can be written as

\[
  f_t(w) \approx U g(w) + (1 - U) T_t(w) - U \Psi_t(w),
\]

where \( T_t(w) \) and \( \Psi_t(w) \) are the inverse Laplace transformations of \( \tilde{\varphi}_t(z) \) and \( \tilde{\psi}_t(z) \), respectively. Since \((n > 1)\)

\[
  G_n \ll \sum_{k=1}^{n} \frac{G_k G_{t+n-1-k}}{G_{t-1}} \ll \frac{G_{t+n-1}}{G_{t-1}}
\]

because of the criterion \((11)\), we expect that, when \( w \) is small (large), the dominant contribution for \( f_t(w) \) comes from \( g(w)[T_t(w)] \). Hence we neglect the contribution from \( \Psi_t(w) \), which gives

\[
  f_t(w) \approx U g(w) + (1 - U) T_t(w).
\]

For the exponential distribution \((\nu = 0, \beta = 1)\), the analytic form of \( \tilde{\varphi}(z) \) can be found. For this case, \( G_t = w_0^t t! \), \( \tilde{g}(z) = (1 + zw_0)^{-1} \), and

\[
  \tilde{\varphi}_t(z) = \sum_{n=0}^{\infty} \frac{(-z)^n G_{t+n-1}}{n!} = \sum_{n=0}^{\infty} \left( \frac{t-n-1}{n} \right)^t (-zw_0)^n = (1 + zw_0)^{-t}.
\]

Hence the frequency distribution is

\[
  f_t(w) \approx \frac{U}{w_0} e^{-w/w_0} + \frac{1 - U}{w_0(t-1)!} \left( \frac{w}{w_0} \right)^{t-1} e^{-w/w_0}.
\]

One can easily check that the mean and the variance at large \( t \) become\(^2\)

\[
  \bar{w}_t \approx (1 - U) w_0 t, \quad \delta w^2 \approx U (1 - U) w_0^2 t^2,
\]

where \( \delta w^2 = \int w^2 f_t(w) \, dw - \bar{w}_t^2 \). Standard deviation and mean are of the same order and \( \delta w \) is even larger than \( \bar{w}_t \) if \( U > \frac{1}{2} \). The origin of such a large spread is the division of \((29)\) into two widely separated distributions: a time-independent part, arising from the mutations, and a traveling wave reflecting the selection dynamics. Then the mean and the variance of \( T_t(w) \) in the asymptotic regime are found to be \( t w_0 \) and \( t w_0^2 \), respectively. Hence the spread of \( T_t \) is much smaller than the mean in the asymptotic regimes, which cannot be appreciated from the full variance in \((30)\). By the central limit theorem\(^3\), \( T_t(w) \) is approximated by the Gaussian distribution

\[
  T_t(w) \approx \frac{1}{\sqrt{2\pi t w_0^2}} \exp \left( -\frac{(w - w_0 t)^2}{2t w_0^2} \right).
\]

Although we cannot find an analytic form of \( f_t(w) \) for the general class of distributions \((13)\), the qualitative form is expected to be the same as in the exponential case, that is, a superposition of a stationary distribution due to mutations and a Gaussian traveling wave. Based on this conjecture, we can approximate the traveling wave for the general case. From the generating function of \( T_t(w) \) which is \( \tilde{\varphi}(z) \), one can get the leading

\(^2\) In \([28]\), \( \delta w^2 \) was also calculated but the factor \((1 - U)\) is missing in the result.

\(^3\) Recall that the Gamma distribution with integer \( t \) can be interpreted as the distribution of the sum of \( t \) independent and identically distributed random variables with exponential distribution \([43]\).
Figure 4. Comparison of the exact numerical solution for the frequency distribution with the traveling wave equation. Left panel: frequency distributions at generation 100, 200 and 400 (from left to right) are shown for $U = 0.5$ (square) and $U = 10^{-4}$ (circle). There is a slight mismatch due to the neglect of $\Psi_t(w)$ but in general the traveling wave solution approximates the true distribution quite well. Right panel: similar study to the left panel with the Gaussian distribution. The data are collected at generation 500, 1000 and 2000. The Gaussian approximation is almost perfect. The frequency distribution for small $w$ is $U g(w)$ for both panels as reasoned in the text (data not shown).

behavior of the mean and variance of $T_t(w)$ such as

$$-\tilde{\phi}'(0) = \frac{G_t}{G_{t-1}} \approx \left( \frac{t}{\beta} \right)^{1/\beta},$$

$$\tilde{\phi}''(0) - \tilde{\phi}'(0)^2 = \frac{G_{t+1}}{G_{t-1}} - \left( \frac{G_t}{G_{t-1}} \right)^2 \approx \frac{1}{\beta^2} \left( \frac{t}{\beta} \right)^{(2-\beta)/\beta},$$

that is, the traveling wave is Gaussian with mean $\bar{w}_t/(1-U)$ and width $\sim t^{(2-\beta)/(2\beta)}$.

Since we know the exact values of $\bar{w}_t$ from the numerical iteration of (9), we can also numerically calculate the frequency density from (3). Figure 4 numerically confirms for the exponential and Gaussian mutation distribution that the frequency distribution is divided into two parts and the traveling part takes the Gaussian form with the predicted mean and variance in this section. The decomposition (27) will play a considerable role in understanding the behavior of finite populations with large $U$, see section 4.6.

4. Finite populations

4.1. Fixation and clonal interference

An important element in the evolution of finite populations is the process of fixation, in which a mutation that is initially present in a single individual spreads in the population and eventually is shared by all individuals (see figure 1 for illustration). Consider the simple case of a single mutant of fitness $w'$ entering a genetically homogeneous population.
Evolution in random fitness landscapes: the infinite sites model

in which all individuals have the same fitness \( w \). The success of the mutant is determined by the selection coefficient

\[
s = \frac{w'}{w} - 1,
\]

which is positive (negative) for beneficial (deleterious) mutations. For the WF model the fixation probability is given approximately by [44]

\[
\pi_N(s) \approx \frac{1 - e^{-2s}}{1 - e^{-2Ns}}.
\]

When selection is strong, in the sense that

\[
N|s| \gg 1,
\]

it can be seen from (35) that fixation of deleterious mutations becomes exponentially unlikely, while a beneficial mutation is fixed with probability

\[
\pi(s) \approx 1 - e^{-2s}.
\]

Previous work on the house of cards model in finite populations has been concerned with the weak selection regime, where \( N|s| \approx 1 \) [30]–[32]. This regime can be realized by choosing a mutation distribution \( g(w) \) whose standard deviation is much smaller than the mean. However, in the context of the present paper the strong selection criterion (36) is always satisfied.

The mean time to fixation of a beneficial mutation is given by [44]

\[
t_{\text{fix}} \approx \frac{\ln N}{s}.
\]

Different evolutionary regimes arise from the comparison of \( t_{\text{fix}} \) to the expected time interval between the emergence of beneficial mutations that are destined for fixation. Denoting by \( U_b \) the beneficial mutation probability per individual, beneficial mutations arise at rate \( NU_b \). A mutation fixes with probability \( \pi(s_b) \approx 2s_b \), where \( s_b \) is the typical selection coefficient which is assumed to be small, so that (37) can be approximated by \( 2s \).

Then the waiting time between fixation events is \( t_{\text{mut}} \approx 1/(2NU_b s_b) \). When \( t_{\text{fix}} \ll t_{\text{mut}} \) or [42]

\[
2N \ln NU_b \ll 1,
\]

beneficial mutations arise rarely and fix independently, a regime that is referred to as periodic selection [45]. In the opposite case \( 2N \ln NU_b \gg 1 \) clones originating from different mutants compete for fixation, a phenomenon that is known as clonal interference [40].

In previous studies of clonal interference [37], [40]–[42], [45] it has usually been assumed that \( U_b \) is a constant parameter, in which case (39) is a condition on the population size \( N \) which is violated when \( N \) becomes large. However, in a rugged landscape the supply of beneficial mutations decreases as the mean fitness grows. If the fitness distribution of the population is well clustered around its mean \( \bar{w} \), the probability of beneficial mutations can be estimated by

\[
U_b(\bar{w}) = U \, \text{Prob}[w > \bar{w}] = U \int_{\bar{w}}^\infty dw \, g(w),
\]

\[\text{doi:10.1088/1742-5468/2008/04/P04014}\]
which vanishes for $\bar{w} \to \infty$ for any unbounded distribution $g(w)$. Thus clonal interference is a \textit{transient} phenomenon in rugged fitness landscapes. Asymptotically almost all mutations are deleterious, and hence $U$ can be identified with the probability of deleterious mutations.

4.2. Instantaneous fixation and the diluted record process

The fact that the criterion (39) is asymptotically satisfied for unbounded $g(w)$ implies that the fixation of beneficial mutations occurs as independent events that can then as well be treated as instantaneous ($t_{\text{fix}} \to 0$). Deleterious mutations do not fix, but they lower the mean population fitness through the mutational load $1 - U$ (compare to section 3). For the time being, we will neglect this effect, which amounts to taking $U \to 0$. We will show in section 4.6 how the influence of deleterious mutations can be approximately reinstated.

When the effects of deleterious mutations are ignored, the population fitness between fixation events is equal to the fitness of the last beneficial mutation that was successfully fixed, and the population is genetically homogeneous at all times (except for the instances of fixation). The model reduces to a simple point process that can be informally described as follows [31, 32]:

(i) Mutations with fitness values drawn randomly and independently from $g(w)$ are generated in discrete time according to a Poisson distribution with mean $NU$.

(ii) The fitness of the mutant is compared to the current population fitness; deleterious mutations are discarded while beneficial mutations are fixed with probability $\pi(s)$ given by (37).

(iii) The fitness of the successfully fixed mutant replaces the current population fitness, which therefore evolves according to a piecewise constant, strictly increasing jump process.

If every beneficial mutation were fixed, such that $\pi(s) = \Theta(s)$, the process described above would be identical to a variant of the well-known problem of record statistics for sequences of independent, identically distributed variables [33, 34], in which a Poisson-distributed number of new variables is created in each discrete time step. When $\pi(s) < 1$ some of the record events are lost in a way that is correlated to the corresponding record values. The process defined by rules (i)–(iii) is referred to in the following as the \textit{diluted record process} (DRP), and it will be studied extensively in the next three subsections.

For long times, when beneficial mutations become increasingly rare, the discrete unit of time is unimportant and the dependence on the system parameters $N$ and $U$ can be eliminated by using the dimensionless time variable $\tau = UNt$. Asymptotically the DRP is therefore fully specified by the functions $\pi(s)$ and $g(w)$. A comparison between the DRP and the full WF dynamics is shown in figure 5. For large $N$ an initial regime can be identified in which clonal interference reduces the fitness in the WF model compared to the DRP, and for large $U$ the fitness is reduced by the mutational load effect, but for long times and $U \ll 1$ the agreement is seen to be essentially perfect. In this and the following three subsections, we restrict ourselves to $U \ll 1$ and the analysis of the behavior for large $U$ is deferred to section 4.6.

Clearly the record process (RP) provides an upper bound on the DRP at all times [32]; for completeness, a derivation of the fitness distribution for the RP is provided in

doi:10.1088/1742-5468/2008/04/P04014
Evolution in random fitness landscapes: the infinite sites model

Figure 5. Left panel: semilogarithmic plots of the fitness versus dimensionless time $\tau = N U t$ for $U = 10^{-5}$ ($N = 10^3, 10^4, 10^5$), $U = 10^{-2}$ and $U = 0.5$ ($N = 10^3, 10^4, 10^5$) obtained from simulations of the WF model. The mutation distribution and the initial frequency distribution are the same as those in figure 2. Right panel: the comparison of the simulation of the WF model to the DRP for $N = 10^3$ and $U = 10^{-5}$ ($\tau = 10^{-2} t$). The difference in the fitness is barely observable. Inset: the same type of comparison for the variance $\kappa_2$ (see (61)) of the fitness.

appendix B. We will now argue that the RP bound is, in fact, saturated for the leading order behavior of the mean fitness, in the sense that

$$\lim_{t \to \infty} \frac{\langle w \rangle_{\text{DRP}}}{\langle w \rangle_{\text{RP}}} = 1. \tag{41}$$

To see why this is so, denote by $w_1$ the largest fitness value that has appeared up to time $\tau$, and by $w_2$ the largest fitness that has also been fixed. Then $w(\tau) = w_1$ in the record process and $w(\tau) = w_2$ in the DRP. If $w_1$ were asymptotically larger than $w_2$ in the sense that $w_1/w_2 \geq C > 1$, then the selection coefficient of $w_1$ in a background of $w_2$ is $s_{12} = w_1/w_2 - 1 \geq C - 1 > 0$ and the corresponding fixation probability $\pi(s_{12})$ is bounded away from zero. It follows that $w_1$ is fixed with finite probability, in contradiction to our assumption that $w_2$ is the largest fitness that has been fixed.

We will see later that the average fitness $w_2$ at time $\tau$ for the exponential distribution is of order $w_2 \sim \ln \tau - \ln \ln \tau$, while $w_1 \sim \ln \tau + \text{const}$. As the difference between the largest and $k$th largest value is of order $\ln k$ for exponential random variables [46], this implies that the rank of $w_2$ among the $\tau$ fitness values that have been created up to time $\tau$ is $O(\ln \tau)$.

In the following subsections we will develop some analytic tools to systematically compute the mean fitness and higher fitness moments for the DRP.

4.3. Mean field approximation

In the mean field approximation (MFA) the fitness distribution of the population is characterized only by its mean, which will be denoted by $m(\tau)$ in the following. The
probability that a new mutation with arbitrary fitness \( w' > m \) is fixed is given by

\[
p_{\text{fix}} = \int_{m}^{\infty} \pi \left( \frac{w' - m}{m} \right) g(w') \, dw' = m \int_{0}^{\infty} \pi(x) g(mx + m) \, dx,
\]

and the waiting time until this happens (in dimensionless units) is \( \Delta \tau = 1/p_{\text{fix}} \). Once fixation occurs, the population fitness increases by the amount

\[
\Delta w = \int_{m}^{\infty} \frac{w'}{m} \pi \left( \frac{w' - m}{m} \right) g(w') \, dw' - m = m \int_{0}^{\infty} \frac{x \pi(x) g(mx + m) \, dx}{\int_{0}^{\infty} \pi(x) g(mx + m) \, dx},
\]

and \( m(\tau) \) is obtained by solving the differential equation

\[
\frac{dm}{d\tau} = \frac{\Delta w}{\Delta \tau} = p_{\text{fix}} \Delta w.
\]

For the exponential distribution (2) this takes the explicit form

\[
\frac{dm}{d\tau} = 4 \frac{m + 1}{(m + 2)^2} e^{-m},
\]

with the solution

\[
e^m (m + 2) + I(m) = 4(\tau - \tau_0) + e^{w_0}(w_0 + 2) + I(w_0),
\]

where

\[
I(m) \equiv \int_{0}^{m} \frac{e^x}{x + 1} \, dx \sim \sum_{n=0}^{\infty} n! \left( \frac{e^m}{(m + 1)^{n+1}} - 1 \right).
\]

The asymptotic expansion which is a divergent series is obtained by integration by parts. Now assume that \( \tau \) is extremely large, then the approximate inversion formula of (46) will be the solution of the equation

\[
me^m \approx 4\tau \Rightarrow m + \ln m \approx \ln(4\tau).
\]

From (48), one can easily see that \( \lim_{\tau \to \infty} \ln \tau / m = 1 \). Let \( m(\tau) = (1 + f(\tau)) \ln(4\tau) \) where \( f(\tau) \to 0 \) as \( \tau \to \infty \). Then

\[
f(\tau) \ln(4\tau) + \ln(4\tau) + \ln(1 + f(\tau)) = 0 \to f(\tau) \approx -\frac{\ln \ln 4\tau}{1 + \ln 4\tau},
\]

which, as expected, approaches zero in the asymptotic regime. According to extreme value statistics, the largest fitness up to time \( \tau \) is of order \( \ln \tau + \gamma \), where \( \gamma(\approx 0.5772) \) is the Euler number. But the leading behavior of \( m(\tau) \) is \( \ln \tau + \ln 4 \), which seems contradictory. This apparent paradox can be resolved by looking at the difference \( \ln(\tau) - m(\tau) = O(\ln \ln \tau) > 0 \), which diverges for \( \tau \to \infty \). In figure 6, the MFA solution (48) with the correction (49) is compared to the WF simulation and the record problem.

For fitness distributions with a power law tail (17), we get \( (\alpha > 1) \)

\[
p_{\text{fix}} = \alpha m^{-\alpha} J(\alpha + 1, m), \quad \Delta w = m \frac{J(\alpha, m) - J(\alpha + 1, m)}{J(\alpha + 1, m)} - 1,
\]

\[\text{doi:10.1088/1742-5468/2008/04/P04014}\]
Figure 6. Comparison of the WF simulation with the record, mean field and the improved approximation scheme in section 4.4 with the stable value of $C_{DRP} = 8$. The WF simulation results are obtained with the exponential mutation distribution (2) with $N = 10^5$ and $U = 10^{-5}$ ($\tau = t$). As explained in the text, mean field theory and record dynamics give lower and upper bounds on the true asymptotics.

where

$$J(\alpha, m) = \int_{1}^{\infty} \left(1 - e^{-2(x-1)}\right) \left(\frac{1}{m} + x\right)^{-\alpha} \, dx.$$  \hspace{1cm} (51)

To have a meaningful result, we should restrict ourselves to the case $\alpha > 1$. Hence (44) for the power law case becomes

$$\frac{dm}{d\tau} = \alpha m^{-\alpha+1}(J(\alpha, m) - J(\alpha + 1, m)) - \alpha m^{-\alpha} J(\alpha + 1, m)$$

$$\approx \alpha m^{-\alpha+1}(K(\alpha) - K(\alpha + 1)), \hspace{1cm} (52)$$

where $K(\alpha) = J(\alpha, \infty)$. Evaluation of the differential equation (52) in the asymptotic regime ($m \gg 1$) yields

$$m \approx \tau^{1/\alpha}(\alpha^2(K(\alpha) - K(\alpha + 1)))^{1/\alpha} \equiv \tau^{1/\alpha} L(\alpha).$$  \hspace{1cm} (53)

To complete the analysis, let us compare $L(\alpha)$ to the prefactor obtained in (B.10) for the record process. When $\alpha - 1 \ll 1$, the two prefactors have the expansion

$$L(\alpha) = \frac{1}{\alpha - 1} + \ln(\alpha - 1) + 0.916\,014 + o(1),$$

$$\Gamma\left(\frac{\alpha - 1}{\alpha}\right) = \frac{1}{\alpha - 1} + (1 - \gamma) + o(1),$$  \hspace{1cm} (54)

which shows that the RP prefactor is larger than that of the MFA in this regime. When $\alpha \gg 1$, the asymptotic behavior of the MFA prefactor can be obtained by integration.

doi:10.1088/1742-5468/2008/04/P04014
J. Stat. Mech. (2008) P04014

Evolution in random fitness landscapes: the infinite sites model

Figure 7. Log-log plots of $\bar{w}(\tau)$ versus $\tau$ for the simulation results of the WF dynamics with the power law distribution (17) with $\alpha = \frac{3}{2}$ (upper dataset), $\alpha = 2$ (middle dataset) and $\alpha = 4$ (lower dataset). The mutation probability is set to $U = 10^{-4}$ and the population sizes are $10^4$ and $10^5$ for $\alpha = \frac{3}{2}$ and 2, and $10^5$ and $10^6$ for $\alpha = 4$, respectively. For comparison, the record solution (B.10) (straight line segments above the simulation results) and the MF prediction (53) (straight line segments below the simulation results) are also depicted.

by parts. One finds

\[ L(\alpha) = 1 + \frac{1}{\alpha} \ln(4/\alpha) + o(1/\alpha) < 1, \]

\[ \Gamma \left( 1 - \frac{1}{\alpha} \right) = 1 + \frac{\gamma}{\alpha} + o(1/\alpha) > 1, \]

which also suggests that the MFA prediction is smaller than the RP value. Actually, the MFA prefactor becomes smaller than unity when $\alpha > 3.53318$ while $\Gamma((\alpha - 1)/\alpha)$ remains larger than unity. In between, one can numerically check that the RP prefactor is always larger than that of the MFA. In fact, we will show in the next subsection that the MFA always provides a lower bound on the true mean fitness of the DRP. To summarize the mean field theory for the power law distribution, the MFA fitness is of the same order as the extremal statistics estimate (B.11), but the smaller prefactor is not consistent with the relation (41).

Figure 7 compares the WF simulation results with the record solution (B.10) and the MF result (53) for three values of $\alpha$. For $\alpha = \frac{3}{2}$, the RP result is in good agreement with the WF simulations but, as $\alpha$ increases, both bounds increasingly deviate from the simulation data. Since for $\alpha \to \infty$ the power law distribution approaches a distribution of exponential type, the results for the exponential distribution summarized in figure 6 indicate that a similar discrepancy should be expected for larger values of $\alpha$. For $\alpha \leq 2$, the variance of the distribution of record values is infinite, which implies that the value
of a new record is usually much larger than the previous record and \( \pi(s) \approx 1 \). In this case the DRP becomes identical to the RP. On the other hand, as \( \alpha \) gets bigger, the ratio of two consecutive records approaches unity in the asymptotic regime. This implies stronger corrections to the leading behavior which, according to (41), should still be given by the RP.

4.4. Master equation

To improve on the mean field approximation, we will first derive an equation for the transition probability \( p(w, \tau|w_0,0) \) of the DRP [19,31]. Since the model is a Markov process, the conditional probability \( p(w,t+dt|w',t') \) completely specifies the evolution equation. We have

\[
p(w, \tau + d\tau|w', \tau) = \delta(w-w') \left( 1 - d\tau \int_{w'}^{\infty} \pi \left( \frac{x-w'}{w'} \right) g(x) \, dx \right) \\
+ d\tau \Theta(w-w')g(w)\pi \left( \frac{w-w'}{w'} \right) .
\]

(56)

Using the Markov property, the equation for \( p(w, \tau) \equiv p(w, \tau|w_0,\tau_0) \) can be found as follows:

\[
p(w, \tau + d\tau) = \int dx \, p(w, \tau + d\tau|x, \tau)p(x, \tau) = p(w, \tau) \left( 1 - d\tau \int_{w}^{\infty} \pi \left( \frac{x-w}{w} \right) g(x) \, dx \right) \\
+ g(w) \, d\tau \int_{w_0}^{w} \pi \left( \frac{w-x}{x} \right) p(x, \tau) \, dx,
\]

(57)

and therefore\(^4\)

\[
\frac{dp(w, \tau)}{d\tau} = -p(w, \tau) \int_{w}^{\infty} \pi \left( \frac{x-w}{w} \right) g(x) \, dx + g(w) \int_{w_0}^{w} \pi \left( \frac{w-x}{x} \right) p(x, \tau) \, dx.
\]

(58)

The description by the master equation (58) is appropriate for a Markov process like the DRP which has non-continuous sample paths [47]. One can easily check that the record probability (B.3) solves (58) when \( \pi(x) = \Theta(x) \).

From (58) evolution equations for arbitrary expectation values \( \langle f(w, \tau) \rangle \) can be obtained in the form

\[
\frac{d}{d\tau} \langle f(w, \tau) \rangle \equiv \frac{d}{d\tau} \int dw \, p(w, \tau)f(w, \tau) = \int dw \left( f(w, \tau) \frac{\partial p(w, \tau)}{\partial \tau} + \frac{\partial f(w, \tau)}{\partial \tau} p(w, \tau) \right) \\
= \left\langle w \int_{0}^{\infty} [f(wx+w, \tau) - f(w, \tau)] \pi(x) g(wx+w) \, dx \right\rangle + \left\langle \frac{\partial f(w, \tau)}{\partial \tau} \right\rangle.
\]

(59)

\(^4\) When the fixation of deleterious mutations is included by using the expression (35) for the fixation probability, for suitable choices of \( g(w) \) the master equation satisfies detailed balance with respect to a stationary distribution [31]. In the present setting, where selection is assumed to be strong in the sense of (36), this stationary distribution cannot be reached on reasonable timescales.

doi:10.1088/1742-5468/2008/04/P04014
For example, the equation for the centered normalized moment $\kappa_n \equiv (w - \langle w \rangle)^n/n!$ is
\[
\frac{d\langle w \rangle}{d\tau} = \left\langle w^2 \int_0^\infty x \pi(x) g(wx + w) \, dx \right\rangle,
\]
\[
\frac{d\kappa_n}{d\tau} = \sum_{r=1}^n \left\langle \frac{w^{r+1}}{r!} (\delta w)^{n-r} \int_0^\infty dx \, x^r \pi(x) g(wx + w) \right\rangle - \kappa_{n-1} \frac{d\langle w \rangle}{d\tau},
\]
where $\delta w \equiv w - \langle w \rangle$. For the cases of the exponential distribution (2) and the power law (17), (60) becomes
\[
\frac{d\langle w \rangle}{d\tau} = \left\langle \frac{4(w + 1)}{(w + 2)^2} e^{-w} \right\rangle,
\]
\[
\frac{d\langle w \rangle}{d\tau} = \alpha \langle w^{-\alpha+1}(J(\alpha, w) - J(\alpha + 1, w)) - w^{-\alpha} J(\alpha + 1, w) \rangle,
\]
which reduces to the mean field equations (45) and (52) if we approximate $\langle \chi(w) \rangle \approx \chi(\langle w \rangle)$, where $\chi(w)$ is the function inside the brackets on the right-hand side of (62) and (63). Again note that (63) is meaningful only if $\alpha > 1$. Since $\chi(w)$ is a convex function asymptotically, that is, $\chi''(w) > 0$ for $w \gg 1$, $\chi(\langle w \rangle) \leq \langle \chi(w) \rangle$, which means that the mean field theory yields a lower bound on the true asymptotics.

### 4.5. Moment expansion

We now develop a systematic approximation scheme which extends beyond mean field theory. First we write down the differential equations for $m \equiv \langle w \rangle$, $\kappa_2, \ldots , \kappa_\ell$ which are available from equations (60) and (61) once $g(w)$ is given. Then we expand the terms on the right-hand side, of the general form $\langle \chi(w) \rangle$, up to $\kappa_\ell$ in such a way that
\[
\langle \chi(m + \delta w) \rangle = \chi(m) + \kappa_2 \frac{d^2 \chi(m)}{dm^2} + \kappa_3 \frac{d^3 \chi(m)}{dm^3} + \cdots
\]
and keep terms only up to $\kappa_\ell$ in all equations for $\kappa_n$ ($\ell \geq n$). If we only keep terms up to $\ell = 1$, then we arrive at the mean field equation. If we keep terms up to $\ell = 2$, then we have
\[
\frac{dm}{d\tau} = \frac{4(m + 1)}{(m + 2)^2 e^m} + \frac{4(m^3 + 7m^2 + 14m + 2)}{(2 + m)^4 e^m} \kappa_2,
\]
\[
\frac{d\kappa_2}{d\tau} = \frac{2(3m^2 + 6m + 4)}{(m + 2)^3 e^m} - \frac{2(m^4 + 8m^3 + 30m^2 + 68m + 16)}{(m + 2)^5 e^m} \kappa_2
\]
for exponential $g(w)$. Since we are interested in the asymptotic behavior, the above equations are approximated for large $m$ by
\[
\frac{dm}{d\tau} \simeq \frac{4}{m e^m} (1 + \kappa_2), \quad \frac{d\kappa_2}{d\tau} \simeq \frac{2}{m e^m} (3 - \kappa_2).
\]
Thus we see that $\kappa_2 \to 3$ and the mean field equation for $m$ receives a multiplicative fluctuation correction. By increasing $\ell$, the solution should become more accurate.

Clearly the above scheme cannot be applicable to all $g(w)$. For example, if $g(w)$ has a power law tail (17), $\kappa_n$ is infinite if $n \geq \{\alpha\}$. However, even if $\kappa_n$ is well defined for all...
n, it is not at all guaranteed that the solution for \( \langle w \rangle \) becomes better as we take more and more \( \kappa_n \) into account. To clarify this point, let us think about the record problem for the exponential distribution (2) which corresponds to \( \pi(x) = \Theta(x) \) and is exactly solvable. For this case (60) and (61) yield

\[
\frac{dm}{d\tau} = \langle e^{-w} \rangle = e^{-m} \sum_{k=0}^{\infty} (-1)^k \kappa_k, \quad (67)
\]

\[
\frac{d\kappa_n}{d\tau} = e^{-w} \left[ \left( \sum_{r=0}^{n-1} \frac{(\delta w)^r}{r!} \right) - \kappa_{n-1} \right] \\
\sim e^{-m} \left( \sum_{r=0}^{n-1} \sum_{k=0}^{\infty} \binom{k+r}{k} (-1)^k \kappa_{r+k} - \kappa_{n-1} \langle e^{-\delta w} \rangle \right), \quad (68)
\]

where \( \kappa_0 = 1 \) and \( \kappa_1 = 0 \) by definition and \( \sim \) means that the order of summation and integration is interchanged (without any legitimation). The approximation scheme described above implies that we keep terms only up to \( \ell \) in (67) and (68). Now, assume that this solution becomes exact as \( \ell \to \infty \). If this is true, we expect that \( \kappa_k \to M_k/k! \) as \( \tau \to \infty \) where \( M_k \) is defined in (B.5). Naively interchanging the order of summation and integration, we then get

\[
\langle e^{-\delta w} \rangle \sim \sum_{k=0}^{\infty} (-1)^k \frac{M_k}{k!} \int_{0}^{\infty} \sum_{k=0}^{\infty} \frac{(\gamma + \ln x)^k}{k!} e^{-x} \, dx = e^{\gamma}, \quad (69)
\]

which along with (67) gives the exact asymptotic behavior \( m \sim \ln \tau + \gamma \). Similarly we obtain by commuting summation and integration that

\[
\sum_{r=0}^{n-1} \sum_{k=0}^{\infty} \binom{k+r}{k} (-1)^k \kappa_{r+k} \sim e^{\gamma} \sum_{r=0}^{n-1} I_r = \kappa_{n-1} e^{\gamma}, \quad (70)
\]

where

\[
I_r = \frac{1}{r!} \int_{0}^{\infty} (-\gamma - \ln x)^r x e^{-x} \, dx = \kappa_r - \kappa_{r-1} \quad (71)
\]

for \( k \geq 1 \) and \( I_0 = 1 \). Thus it seems that our approximation scheme solves the problem accurately as \( \ell \) increases.

However, this agreement is in fact fortuitous. Equation (69) illustrates the problem. One can easily see that \( \langle e^{-\delta w} \rangle \) is indeed equal to the integral yielding \( e^{\gamma} \), but the intermediate series \( S_\ell \equiv \sum_{k=0}^{\ell} (-1)^k \kappa_k \) is actually not convergent. In fact, even if we use the exact values for \( \kappa_n \), \( S_\ell \) oscillates between 1.500 34 and 2.0618 whose average is 1.781 07 = \( e^{\gamma} \). We conclude, therefore, that the approximation scheme can at best be expected to yield an asymptotic series.

Despite this difficulty, we now demonstrate that the moment expansion yields reliable results when used with caution. We have applied the scheme to the record problem and the DRP with exponential \( g(w) \). Assuming that all \( \kappa_n \) will saturate as \( \tau \to \infty \), the mean fitness then follows an equation of the form

\[
\frac{dm}{d\tau} \approx C_{\text{RP}}(\ell)e^{-m}, \quad (72)
\]
Figure 8. The results of the approximation scheme for the mean ($C_{RP}$ and $C_{DRP}$) and variance ($\kappa_2$) including terms up to $\kappa_\ell$ for the RP (left panel) and for the DRP (right panel). Since we know the exact value of $C_{RP}$ and $\kappa_2$ for the record problem, in the left panel we compare the approximation to the exact values.

for the record process

$$\frac{dm}{d\tau} \approx \frac{C_{DRP}(\ell)}{me^m},$$

(73)

for the DRP, respectively. The argument $\ell$ means that the constants $C_{RP}$ and $C_{DRP}$ are evaluated keeping the $\kappa_n$ up to $n = \ell$. For instance, we have already shown that $C_{DRP}(1) = 4$ from (45) and $C_{DRP}(2) = 16$ from (66).

Figure 8 summarizes the results of the approximate evaluation of the $C$’s and of $\kappa_2$ with increasing $\ell$. In the range $4 \leq \ell \leq 10$, the approximation yields rather stable values. The comparison with the exact results for the record problem shows that the method is excellent in this range. However, for $\ell$ larger than 10, the error becomes uncontrollable. In fact, the solutions for $\ell > 10$ in figure 8 are meaningless because some $\kappa_n$’s are negative which should be positive. The approximation suggests that $C_{DRP} \approx 8$, which gives a highly accurate estimate of the asymptotic fitness, see figure 6. On the other hand, the estimate $\kappa_2 \approx 0.8$ appears to be somewhat smaller than the simulation results, see the inset of the right panel in figure 5.

4.6. Finite $U$

Until now, the mutation probability $U$ has been assumed to be very small. The natural extension of the previous study is to ask what will happen if the mutation rate is high, which is the topic of this subsection.

One can get some insight from the infinite population calculation in section 3, where it was shown that the frequency distribution can be approximated by a superposition (27) of two distributions which are well separated from each other. This is still expected to happen in the finite population case when the mean fitness is much larger than the average fitness due to mutations (the average of $g(w)$). Figure 9 depicts the cumulative frequency distribution obtained from simulations of the WF model at $t = 10^2$, $10^3$ and $10^4$ for
Figure 9. The cumulative frequency distribution for $N = 10^5$ and $U = 0.5$ at $t = 10^2$, $10^3$ and $10^4$ (from left to right). The data are collected from 64000 independent simulations. The approximate expressions $e(w)$ in (76) and $ec(w)$ in (77) with $\xi = 0.4725$ are also drawn for comparison. Inset: the same but the abscissa is shifted by the amount $m(\tau)$ which is the solution of (73) with $C_{DRP} = 8$. The datasets for $t = 10^3$ and $t = 10^4$ show a nice collapse.

For $N = 10^5$ and $U = 0.5$, defined as

$$F(w) \equiv \left\langle \int_w^{\infty} f_t(x; N) \, dx \right\rangle,$$

(74)

where $\langle \cdots \rangle$ means an average over independent samples and $f_t(x; N)$ is the frequency distribution for population size $N$. As expected, the cumulative distribution displays a plateau corresponding to the region of low probability between the two peaks, but in contrast to the ansatz (27) the height of the plateau is below $U$. To get a quantitative explanation of this effect, we assume that the frequency distribution is of the form $f_t(w; N) = (1 - \xi)g(w) + \xi\delta(w - m)$, where $\xi$ is the weight of the high fitness peak which is approximated by a $\delta$ function when $m \gg 1$. Then the population fraction of the genotype with fitness $m$ increases after selection to $\xi m / (\xi m + 1 - \xi)$, out of which only a fraction $1 - U$ remains after mutation. If $m$ and $\xi$ change slowly on the timescale of one generation, one obtains the stationarity condition

$$\xi = \frac{(1 - U)\xi m}{\xi m + 1 - \xi} \to \xi = (1 - U) - \frac{U}{m - 1},$$

(75)

which shows that $\xi$ approaches $1 - U$ only for $m \to \infty$. This suggests that the cumulative distribution should take the form

$$F(w) \approx e(w) \equiv (1 - \xi)g(w) + \xi,$$

(76)

for $w < m$. For the case considered in figure 9, the mean fitness at $t = 10^4$ is $\approx 19.15$ which gives $\xi \approx 0.4725$ in good agreement with the simulation data. Due to the logarithmic
Figure 10. Plots of $\bar{w}(\tau; U)/(1 - U)$ versus $\tau$ for the datasets in the left panel of figure 5. All curves now collapse in the asymptotic regime.

increase of $m$ with $t$ in the case of exponential $g(w)$, the approach to the asymptotic value $\xi \rightarrow 1 - U$ is very slow. For example, to reach $\xi = 0.49$ requires us to simulate $t \sim 10^{16}$ generations for our parameters.

For a more accurate description of the high fitness part of the frequency distribution we assume that, as in the infinite population case, the traveling wave contribution in the decomposition (27) becomes Gaussian at long times. If this is true, the cumulative frequency distribution should take the form

$$F(w) \approx ec(w) \equiv \frac{\xi}{2} \text{erfc} \left( \frac{w - m(\tau)}{2\sqrt{\kappa_2}} \right),$$

(77)

where erfc is the complementary error function and $\kappa_2(\approx 0.8)$ is obtained from the analysis of the DRP in the previous section. As shown in figure 9, $ec(w)$ approximates the numerical data quite well, and the distributions obtained at different times $10^3$ and $10^4$ collapse (inset of figure 9). Though $ec(w)$ is not excellent, we would say that it is a reasonable approximation in the asymptotic regime.

As a consequence of these considerations, the effect of the mutational load on the mean fitness is found to be remarkably simple: asymptotically we have

$$\bar{w}(\tau; U) \approx \xi m(\tau) + (1 - \xi) = (1 - U)m(\tau),$$

(78)

where $m(\tau)$ is the mean fitness of the DRP. This prediction is confirmed in figure 10.

5. Conclusions

In this paper we have explored several aspects of the evolution of an asexually reproducing population in a random fitness landscape which (in the sense of the NK family [8, 19]) is maximally epistatic. In contrast to previous work on the REM fitness landscape [20, 21, 23, 26, 27] we adopt the limit of infinite genome length, which leads to a
finite population version of Kingman’s house of cards model [28]. Although the model can hardly be expected to provide a realistic description of empirical fitness landscapes [16, 17], it serves as a useful counterpart to the much studied non-epistatic multiplicative landscape model and can help to develop some intuition for the generic features of adaptation under strong epistasis.

An example for such a generic feature is the slowing down of the speed of adaptation compared to the multiplicative model, which reflects the fact that the supply of beneficial mutations dwindles with increasing fitness. This effect is well documented in evolution experiments with microbial populations [48]–[50]; in fact, it has been argued [9] that the data obtained by Lenski and Travisano [48] for populations of E. coli can be quantitatively described as a logarithmic increase in fitness, which would be consistent with our finite population results for exponential $g(w)$.

We have extended Kingman’s analysis of the infinite population limit to include the shape of the frequency distribution, which was found to become bimodal at long times. The finite population dynamics was shown to reduce to a diluted record process (DRP) for long times and small mutation probability $U \ll 1$. This representation allows us to bound the mean population fitness from above by the standard record process and from below by a mean field approximation to the DRP, which can be improved systematically through a moment expansion. Although connections between record statistics and evolutionary processes have been suggested before [7]–[11], [32], we have here, for the first time, established a precise quantitative relation of the theory of records to one of the cornerstones of population genetics, the WF model. Finally, using insights from the infinite population case, we have shown that the fitness distribution at large $U$ can be understood as a superposition of the DRP distribution and the mutation distribution $g(w)$.

A number of important questions concerning this model are left to future work. For example, it would be interesting to consider the finite population dynamics for bounded mutation distributions $g(w)$, for which the infinite population calculation predicts the appearance of singularities in the stationary fitness distribution [28]. Furthermore, the temporal statistics of the fixation events should be investigated with regard to its relation to record dynamics, for which detailed results are available [33, 34], as well as in comparison to recent work on the corresponding issue for multiplicative fitness landscapes [37].

Acknowledgments

We are grateful to Kavita Jain for useful discussions, and acknowledge financial support from DFG within SFB 680 Molecular basis of evolutionary innovations. This paper is dedicated to Thomas Nattermann on the occasion of his 60th birthday.

Appendix A. Infinite population calculation for Gumbel-type $g(w)$

The first step is the calculation of the $t$th moment of (16):

$$G_t = \int w^t g(w) \, dw = e w_0^t \int_1^\infty (\ln(y))^t e^{-y} \, dy.$$  

(A.1)
To get an asymptotic expression, let us make a change of variables $y = x t / \ln t$:

$$G_t = \frac{e w_0^t}{\ln t} \left( \ln \left( \frac{t}{\ln t} \right) \right)^t \int_{\ln t/t}^{\infty} dx \left( 1 + \frac{\ln x}{\ln(t/\ln t)} \right)^t \exp \left( - \frac{t}{\ln t} x \right)$$

$$= \frac{e w_0^t}{\ln t} \left( \ln \left( \frac{t}{\ln t} \right) \right)^t \int_{\ln t/t}^{\infty} dx \exp \left( - \frac{t}{\ln t} \left( x - \ln t \ln \left( 1 + \frac{\ln x}{\ln(t/\ln t)} \right) \right) \right).$$

(A.2)

Since $t$ is assumed to be very large, the integral is dominated by the regime where the terms in the curly bracket attain a minimum which is easily seen to be unique. One can show that the contribution from the boundaries of the integral is exponentially small and the maximal contribution comes from the region around $x_c$ which is the solution of the equation

$$(1 - x_c) \ln t = x_c (\ln x_c - \ln \ln t) \Rightarrow x_c \simeq 1 + \frac{\ln \ln t}{\ln t} + O \left( \frac{\ln \ln t}{\ln t} \right)^2. \quad (A.3)$$

Expanding the terms in the curly bracket around $x_c$ and performing the Gaussian integral, we get

$$G_t \simeq e w_0^t \sqrt{\frac{2\pi t}{\ln t}} \left( \ln \left( \frac{t}{\ln t} \right) \right)^t \exp \left( - \frac{t}{\ln t} x_c \ln \left( 1 + \frac{\ln x_c}{\ln(t/\ln t)} \right) \right). \quad (A.4)$$

Hence, in the long run, the Gumbel-type distribution also meets the criterion (11) and the mean fitness becomes

$$\bar{w}_t \simeq w_0 (1 - U) \ln \frac{t}{\ln t} \exp \left( \frac{\ln \ln t}{(\ln t)^2} \left( 1 + \frac{(3/2) \ln \ln t - 1}{\ln t} \right) \right), \quad (A.5)$$

which increases logarithmically for long times.

**Appendix B. Record dynamics**

Here we derive an expression for the record probability distribution $p_{RP}(w, \tau)$ of the population fitness at scaled time $\tau$, in the long time limit where the generation of new random variables is equivalent to a Poisson process in continuous time [51]. During time $\tau$, the probability that there are $n$ mutations is $e^{-\tau \tau^n/n!}$. Hence the probability for the fitness at time $\tau$ to be larger than $x$ is

$$\text{Prob}[w > x, \tau] = \sum_{n=0}^{\infty} e^{-\tau \tau^n/n!} (1 - (1 - Q(x))^n) = 1 - \exp(-\tau Q(x)), \quad (B.1)$$

where $Q(x) = \int_{x}^{\infty} dw \ g(w)$ is the cumulative mutation distribution and $x$ is larger than the initial value $w_0$ of the process. In addition there is a contribution from the possibility that no new record has appeared up to time $\tau$:

$$\text{Prob}[w = w_0, \tau] = 1 - \text{Prob}[w > w_0, \tau] = \exp(-\tau Q(w_0)). \quad (B.2)$$

Together the two contributions yield

$$p_{RP}(w, \tau) = \tau g(w) e^{-\tau Q(w)} \Theta(w - w_0) + e^{-\tau Q(w_0)} \delta(w - w_0), \quad (B.3)$$

doi:10.1088/1742-5468/2008/04/P04014
from which moments and cumulants can be derived. For example, for the exponential case with \( g(w) = Q(w) = e^{-w} \) we obtain
\[
\langle w \rangle_{\text{RP}} - w_0 e^{-\tau y_{\text{eq}}_0} = \int_{w_0}^{\infty} dw \, w \tau e^{-w} = \ln \tau + \gamma + O(e^{-\tau y_{\text{eq}}_0}), \tag{B.4}
\]
\[
\langle (w - \langle w \rangle_{\text{RP}})^n \rangle_{\text{RP}} = \int_0^{\infty} (-\gamma - \ln x)^n e^{-x} \, dx + O(e^{-\tau y_{\text{eq}}_0}) \rightarrow M_n, \tag{B.5}
\]
where \( \gamma (\approx 0.5772) \) is the Euler number. Hence
\[
\langle w \rangle_{\text{RP}} \approx \ln \tau + \gamma, \quad \langle (w - \langle w \rangle_{\text{RP}})^2 \rangle_{\text{RP}} \approx M_2 = \frac{s^2}{6}, \tag{B.6}
\]
which is consistent with extreme value statistics [46]. For a Gaussian \( g(w) = 2e^{-w^2/\sqrt{\pi}} \), one finds
\[
\langle w^n \rangle_{\text{RP}} \approx \tau \int_0^{\text{erfc}(w_0)} \left( \text{erfc}^{-1}(y) \right)^n e^{-y} \, dy \approx \int_0^{\infty} \left( \ln(\gamma) - \ln x \right)^{n/2} e^{-x} \, dx \tag{B.7}
\]
\[
\approx \left( \ln(\gamma) \right)^{n/2} + \frac{n}{2} \left( \ln(\gamma) \right)^{n/2-1} \gamma + \frac{n^2 - 2n}{8} \left( \ln(\gamma) \right)^{n/2-2} \left( \frac{\gamma^2}{2} + \frac{\gamma}{6} \right), \tag{B.8}
\]
where erfc is the complementary error function \( \text{erfc}(x) = 2/\sqrt{\pi} \int_x^{\infty} e^{-y^2} \, dy \) and \( \text{erfc}^{-1}(y) \) is its inverse function whose leading asymptotic behavior is \( \left( \ln c - \ln y \right)^{1/2} / c = \sqrt{2/\pi} \) when \( y \ll 1 \). Hence, for the Gaussian case:
\[
\langle w \rangle_{\text{RP}} \approx \ln^{1/2}(\gamma), \quad \langle (w - \langle w \rangle_{\text{RP}})^2 \rangle_{\text{RP}} \approx \frac{s^2}{24 \ln(\gamma)}. \tag{B.9}
\]
For a distribution with a power law tail (17), we have (for \( n < \alpha \))
\[
\langle w^n \rangle_{\text{RP}} \approx \int_{w_0}^{\infty} dw \, \tau \alpha w^n (1 + w)^{-1} e^{-\tau(1+w)^{-\alpha}} \approx \tau^{n/\alpha} \Gamma \left( \frac{\alpha - n}{\alpha} \right), \tag{B.10}
\]
which yields (if both exist)
\[
\langle w \rangle_{\text{RP}} \sim \sqrt{\langle (w - \langle w \rangle_{\text{RP}})^2 \rangle_{\text{RP}}} \sim \tau^{1/\alpha}. \tag{B.11}
\]
If \( n \geq \alpha \) in (B.10), \( \langle w^n \rangle_{\text{RP}} \) is infinite.

References

[1] Anderson P W, Suggested model for prebiotic evolution—the use of chaos, 1983 Proc. Natl Acad. Sci. USA 80 3386
[2] Stein D L (ed), 1992 Spin Glasses and Biology (Singapore: World Scientific)
[3] Lässig M and Valveriani A (ed), 2002 Biological Evolution and Statistical Physics (Berlin: Springer)
[4] Ebeling W, Engel A, Esser B and Feistel R, Diffusion and reaction in random media and models of evolution processes, 1984 J. Stat. Phys. 37 369
[5] Nattermann T and Renz W, Diffusion in a random catalytic environment, polymers in random media, and stochastically growing interfaces, 1989 Phys. Rev. A 40 4675
[6] Bovier A, 2006 Statistical Mechanics of Disordered Systems (Cambridge: Cambridge University Press)
[7] Kauffman S and Levin S, Towards a general theory of adaptive walks on rugged landscapes, 1987 J. Theor. Biol. 128 11
[8] Kauffman S A, 1993 The Origins of Order (Oxford: Oxford University Press)

doi:10.1088/1742-5468/2008/04/P04014
Evolution in random fitness landscapes: the infinite sites model

[9] Sibani P, Brandt M and Alström P, Evolution and extinction dynamics in rugged fitness landscapes, 1998 Int. J. Mod. Phys. B 12 361

[10] Orr H A, The genetic theory of adaptation: a brief history, 2005 Nat. Rev. Genet. 6 119

[11] Krug J and Jain K, Breaking records in the evolutionary race, 2005 Physica A 358 1

[12] Leuthäuser I, Statistical mechanics of Eigen’s evolution model, 1987 J. Stat. Phys. 48 343

[13] Baake E, Baake M and Wagner H, Ising quantum chain is equivalent to a model of biological evolution, 1997 Phys. Rev. Lett. 78 559

[14] Cavalli S, 2004 Fitness Landscapes and the Origin of Species (Princeton, NJ: Princeton University Press)

[15] Jain K and Krug J, Adaptation in simple and complex fitness landscapes, 2007 Structural Approaches to Sequence Evolution: Molecules, Networks and Populations ed U Bastolla, M Porto, H E Roman and M Vendruscolo (Berlin: Springer)

[16] Weinrich D M, Watson R A and Chao L, Sign epistasis and genetic constraint on evolutionary trajectories, 2005 Evolution 59 1165

[17] Poelwijk F J, Kiviet D J, Weinreich D M and Tans S J, Empirical fitness landscapes reveal accessible evolutionary paths, 2007 Nature 445 383

[18] Derrida B, Random-energy model—an exactly solvable model of disordered systems, 1981 Phys. Rev. B 24 2613

[19] Welch J J and Waxman D, The nk model and population genetics, 2005 J. Theor. Biol. 234 329

[20] Franz S, Peliti L and Sellitto M, An evolutionary version of the random energy model, 1993 J. Phys. A: Math. Gen. 26 L1195

[21] Franz S and Peliti L, Error threshold in simple landscapes, 1997 J. Phys. A: Math. Gen. 30 4481

[22] Krug J and Karl C, Punctuated evolution for the quasispecies model, 2003 Physica A 318 137

[23] Jain K and Krug J, Evolutionary trajectories in rugged fitness landscapes, 2005 J. Stat. Mech. P04008

[24] Sire C, Majumdar S and Dean D S, Exact solution of a model of time-dependent evolutionary dynamics in a rugged fitness landscape, 2006 J. Stat. Mech. L07001

[25] Jain K, Evolutionary dynamics of the most populated genotype on rugged fitness landscapes, 2007 Phys. Rev. E 76 031922

[26] Amitrano C, Peliti L and Saber M, Population dynamics in a spin-glass model of chemical evolution, 1989 J. Mol. Evol. 29 513

[27] Jain K and Krug J, Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes, 2007 Genetics 175 1275

[28] Kingman J F C, A simple model for the balance between selection and mutation, 1978 J. Appl. Probab. 15 1

[29] Bürger R, 2000 The Mathematical Theory of Selection, Recombination, and Mutation (Chichester: Wiley)

[30] Obta T and Tachida H, Theoretical study of near neutrality, 1990 Genetics 126 219

[31] Tachida H, A study on a nearly neutral mutation model in finite populations, 1991 Genetics 128 183

[32] Gillespie J H, Substitution processes in molecular evolution. III. Deleterious alleles, 1994 Genetics 138 943

[33] Glick N, Breaking records and breaking boards, 1978 Am. Math. Mon. 85 2

[34] Arnold B C, Balakrishnan N and Nagaraja H N, 1998 Records (New York: Wiley)

[35] Wright S, Evolution in Mendelian populations, 1931 Genetics 16 97

[36] Fisher R A, 1930 The Genetical Theory of Natural Selection. (New York: Wiley)

[37] Kimura M, Substitution processes in molecular evolution. III. Deleterious alleles, 1991 Genetics 128 183

[38] Park S-C and Krug J, Clonal interference in large populations, 2006 Genetics 171 1377

[39] Kim Y and Allen Orr H, Adaptation in sexual versus asexuals: clonal interference and the Fisher–Muller model, 2005 Genetics 167 4481

[40] Gerrish P J and Lenski R E, The fate of competing beneficial mutations in an asexual population, 1998 Genetics 120 123

[41] Orr H A, The rate of adaptation in asexuels, 2000 Genetics 155 961

[42] Wilke C O, The speed of adaptation in large asexual populations, 2004 Genetics 167 2045

[43] Feller W, 1971 An Introduction to Probability Theory and its Applications vol II (New York: Wiley)

[44] Kimura M, On the probability of fixation of mutant genes in a population, 1962 Genetics 47 713

[45] de Visser J A G M and Rozen D E, Clonal interference and the periodic selection of new beneficial mutations in Escherichia coli, 2006 Genetics 172 2093

[46] David H, 1970 Order Statistics (New York: Wiley)

[47] Gardiner C W, 1990 Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences 2nd edn (Berlin: Springer)

[48] Lenski R E and Travisano M, Dynamics of adaptation and diversification: A 10000-generation experiment with bacterial populations, 1994 Proc. Natl Acad. Sci. USA 91 6808

doi:10.1088/1742-5468/2008/04/P04014
Evolution in random fitness landscapes: the infinite sites model

[49] de Visser J A G M, Zeyl C W, Gerrish P J, Blanchard J L and Lenski R E, Diminishing returns from mutation supply rate in asexual populations, 1999 Science 283 404

[50] Elena S F and Lenski R E, Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation, 2003 Nat. Rev. Genet. 4 457

[51] Davidsen J, Grassberger P and Paczuski M, Networks of recurrent events, a theory of records, and an application to finding causal signatures in seismicity, 2007 Preprint physics/0701190