Effect of mutators on adaptability in time-varying fitness landscapes

Pavel Gorodetsky* and Emmanuel Tannenbaum†
Ben-Gurion University of the Negev, Be'er-Sheva, Israel
(Received 15 January 2008; published 10 April 2008)

This Brief Report studies the quasispecies dynamics of a population capable of genetic repair evolving on a time-dependent fitness landscape. We develop a model that considers an asexual population of single-stranded, conservatively replicating genomes, whose only source of genetic variation is due to copying errors during replication. We consider a time-dependent, single-fitness-peak landscape where the master sequence changes by a single point mutation at every time τ. We are able to analytically solve for the evolutionary dynamics of the population in the point-mutation limit. In particular, our model provides an analytical expression for the fraction of mutators in the dynamic fitness landscape that agrees well with results from stochastic simulations.

DOI: 10.1103/PhysRevE.77.042901 PACS number(s): 87.23.Kg

Genetic repair plays an important role in maintaining the integrity of nearly all organismal genomes [1]. An important repair mechanism is mismatch repair (MMR), which is of particular interest to evolutionary biologists, because it is believed that mismatch-repair-deficient strains, or *mutators*, play a crucial role in the emergence of antibiotic drug resistance, and also act as gateway cells for the emergence of cancer [2] (and references therein).

In static environments, mutators are generally a relatively small fraction of the population [3], because their higher than wild-type mutation rate means that they accumulate deleterious mutations more quickly than nonmutators. However, in dynamic environments, mutators can adapt more quickly than nonmutators, and so can rise to relatively large fractions of the population. Experimentally, it is known that in static environments the mutator fraction of *Escherichia coli* is on the order of 0.1%, while pathogenic strains under pressure to respond to an adaptive immune system can have mutator fractions on the order of 10% or more [3].

Because of their importance, mutators have been the subject of considerable experimental and theoretical work [2,4–10]. Theoretical work on the subject has consisted both of numerical studies modeling the evolution of mutation rate in time-varying environments, as well as analytical work modeling how mutators influence adaptation to a new environment.

In this Brief Report, we develop an analytically solvable model describing the influence of mutators on adaptation in time-varying environments. Such a model is an extension of previous analytical work on the subject, which focused either on adaptation to a new, but static, environment [6,8], or on adaptation in environments that alternate between two states [10]. This Brief Report is also an extension of previous numerical work, in that we provide approximate analytical expressions for the fraction of mutators in dynamic environments. The work here therefore provides a starting point for understanding the role of mutators in environments where coevolutionary dynamics is important, as is the case with an adaptive immune system.

*pavelgor@bgu.ac.il
†emanuelt@bgu.ac.il

An analytically solvable model describing quasispecies evolution in a dynamic fitness landscape was first developed in [11]. Following the approach in [11], we assume a single-peak-fitness landscape in which there is one high-fitness sequence, the master sequence $\sigma_{\mathrm{via},0}$, in an otherwise flat landscape. We also assume that there exists a sequence $\sigma_{\mathrm{rep},0}$ which corresponds to a working MMR mechanism. Our organisms then have genomes that may be denoted by $\sigma = \sigma_{\mathrm{via}}\sigma_{\mathrm{rep}}$, where σ_{via} is the genome region that controls viability and σ_{rep} is a region that controls repair. Each base of the genome is chosen from an alphabet of size S (=4 for known terrestrial life).

σ_{via} consists of L_{via} bases, and σ_{rep} consists of L_{rep} bases, so that the total genome length is $L=L_{\mathrm{via}}+L_{\mathrm{rep}}$. Note that we are essentially considering a two-locus model, which is the approach taken by previous authors [6,8]. The difference here is that we are assuming master-sequence-based fitness and "repair" landscapes, so that, in contrast to previous work, various transition probabilities do not need to be considered as independent variables, but rather may be computed from per base error probabilities and sequence lengths.

A viable organism, i.e., an organism for which $\sigma_{\mathrm{via}} = \sigma_{\mathrm{via},0}$, has a first-order growth rate constant $\kappa_\sigma = k > 1$, while for an unviable organism, $\kappa_\sigma = 1$. We define a repair landscape in an analogous manner: $\epsilon_\sigma = \epsilon_r \epsilon$ for organisms with $\sigma_{\mathrm{rep}} = \sigma_{\mathrm{rep},0}$ and $\epsilon_\sigma = \epsilon$ otherwise, where ϵ is the per base replication error probability and ϵ_r is the per base repair error probability.

To create a dynamic landscape, we move the fitness peak in the genotype space to one of its nearest neighbors (chosen randomly) at regular time intervals τ. We then define the parameters n_{00}, n_{01}, n_{10}, and n_{11} as follows: n_{00} is the number of viable nonmutators, n_{01} is the number of viable mutators, n_{10} is the number of nonmutators that will be viable after the next peak shift, and n_{11} is the number of mutators that will be viable after the next peak shift.

We assume that ϵ is sufficiently small that only point mutations are important, and we also assume that n_{10} and n_{11} are zero immediately after a peak shift. That is, when the fitness peak shifts to a new master sequence, the number of organisms that will be viable after the next peak shift is assumed to be negligible. It should also be noted that we neglect back-mutations, since we assume that the sequence lengths are sufficiently long that the probability of a mutation occurring

at a base that has already undergone mutation is negligibly small.

Based on these assumptions, we have that, during the time interval $[n\tau,(n+1)\tau)$ immediately following the nth peak shift, the quasispecies equations take the form

$$\frac{d}{dt}\begin{pmatrix} n_{00} \\ n_{01} \\ n_{10} \\ n_{11} \end{pmatrix} = \begin{pmatrix} C_1 & 0 & 0 & 0 \\ C_3 & C_2 & 0 & 0 \\ C_5 & 0 & C_4 & 0 \\ 0 & C_7 & C_8 & C_6 \end{pmatrix} \begin{pmatrix} n_{00} \\ n_{01} \\ n_{10} \\ n_{11} \end{pmatrix}, \tag{1}$$

where the C_i coefficients are defined as

$$C_1 = k(1-\epsilon\epsilon_r)^L$$

$$C_2 = k(1-\epsilon)^{L_{\text{via}}},$$

$$C_3 = k(1-\epsilon\epsilon_r)^{L_{\text{via}}}[1-(1-\epsilon\epsilon_r)^{L_{\text{rep}}}],$$

$$C_4 = (1-\epsilon\epsilon_r)^L,$$

$$C_5 = \frac{k\epsilon\epsilon_r}{S-1}(1-\epsilon\epsilon_r)^{L-1},$$

$$C_6 = (1-\epsilon)^{L_{\text{via}}},$$

$$C_7 = \frac{k\epsilon}{S-1}(1-\epsilon)^{L_{\text{via}}-1},$$

$$C_8 = (1-\epsilon\epsilon_r)^{L_{\text{via}}}[1-(1-\epsilon\epsilon_r)^{L_{\text{rep}}}]. \tag{2}$$

We may solve Eq. (1) for $t\in[n\tau,(n+1)t)$ with the initial conditions $n_{10}(n\tau)=n_{11}(n\tau)=0$. Noting that at $t=(n+1)\tau$ the fitness peak shifts to the next master sequence, we have

$$n_{00}((n+1)\tau) = A_{10}(\tau)n_{00}(n\tau),$$

$$n_{01}((n+1)\tau) = A_{11}(\tau)n_{00}(n\tau) + B_{11}(\tau)n_{01}(n\tau), \tag{3}$$

where the coefficients $A_{10}(\tau)$, $A_{11}(\tau)$, and $B_{11}(\tau)$ are given by

$$A_{10}(\tau) = \frac{(e^{k\tau C_4} - e^{\tau C_4})C_5}{(k-1)C_4},$$

$$\begin{aligned} A_{11}(\tau) = {} & \frac{(e^{\tau C_4} - e^{k\tau C_4})C_5C_6C_8}{(k-1)(C_4-C_6)(kC_4-C_6)C_4} \\ & + \frac{k(e^{\tau C_6} - e^{k\tau C_6})C_4C_7C_8}{(k-1)(C_4-C_6)(kC_4-C_6)C_6} \\ & + \frac{[e^{k\tau C_4} - e^{\tau C_6} + k(e^{\tau C_6} - e^{\tau C_4})]C_5C_8}{(k-1)(C_4-C_6)(kC_4-C_6)} \\ & + \frac{[(k-1)e^{k\tau C_4} - ke^{\tau C_6} + e^{k\tau C_6}]C_7C_8}{(k-1)(C_4-C_6)(kC_4-C_6)C_4C_6}, \end{aligned}$$

$$B_{11}(\tau) = \frac{(e^{k\tau C_6} - e^{\tau C_6})C_7}{(k-1)C_6}. \tag{4}$$

Let us denote by r_n the ratio of viable mutators to viable nonmutators immediately following the nth peak shift. Then from the solution given above we have

$$r_{n+1} = \alpha(\tau) + \beta(\tau)r_n, \tag{5}$$

where $\alpha(\tau)\equiv A_{11}(\tau)/A_{10}(\tau)$ and $\beta(\tau)=B_{11}(\tau)/A_{10}(\tau)$, from which it can be shown that $r_n=\alpha(\tau)\Sigma_{i=0}^{n-1}[\beta(\tau)]^i+\beta^n(\tau)r_0$. If $0\le\beta<1$, then, as $n\to\infty$, we obtain that $r_\infty\equiv\lim_{n\to\infty} r_n = \frac{A_{11}(\tau)}{A_{10}(\tau)-B_{11}(\tau)}$, which implies a periodic solution where the fraction of viable organisms that are nonmutators at the beginning of each cycle is nonzero. When $\beta\ge1$ we obtain that r_n diverges, and so the fraction of viable organisms that are mutators goes to 1. This phenomenon is known as the *repair catastrophe* [2].

For the periodic solution that develops after a sufficient number of iterations, the fraction of viable organisms that are mutators at the beginning of a peak shift is then $r_\infty/(1+r_\infty)$. As $\tau\to\infty$, it may be readily shown that the fraction of viable organisms that are mutators just before the next peak shift is given by the static landscape expression obtained in [4].

In order for the population to form a stable quasispecies and remain viable in the time-varying landscape, the relative growth of the viable population between the peak shifts should be larger than that of the background (genome sequences some distance away from the master sequence). Since the growth constant of nonviable organisms is equal to 1, the criterion for adaptability is given by

$$\frac{n_{00}((n+1)\tau) + n_{01}((n+1)\tau)}{n_{00}(n\tau) + n_{01}(n\tau)} > e^\tau. \tag{6}$$

If this condition is not met there will be a drift from the master sequence to the "background," corresponding to the error catastrophe.

When the fraction of viable organisms that are nonmutators is nonzero, Eq. (6) may be shown to be equivalent to

$$\kappa_{nm} \equiv \frac{k}{k-1}\frac{\epsilon\epsilon_r}{1-\epsilon\epsilon_r}\frac{e^{[k(1-\epsilon\epsilon_r)^L-1]\tau} - e^{[(1-\epsilon\epsilon_r)^L-1]\tau}}{S-1} > 1, \tag{7}$$

while when the fraction of viable organisms that are nonmutators is zero, i.e., when the population has undergone the repair catastrophe, Eq. (6) may be shown to be equivalent to

$$\kappa_m \equiv \frac{k}{k-1}\frac{\epsilon}{1-\epsilon}\frac{e^{[k(1-\epsilon)^{L_{via}}-1]\tau} - e^{[(1-\epsilon)^{L_{via}}-1]\tau}}{S-1} > 1. \tag{8}$$

We may define three distinct parameter regimes, each corresponding to the viability or nonviability of the nonmutator and mutator populations, respectively.

The first region is defined by $\kappa_{nm}>1,\kappa_m$. Here, the nonmutators' effective growth rate is larger than both that of the background and that of the mutators, so that a stable quasispecies is formed with a nonzero fraction of viable nonmutators.

The second region is defined by $\kappa_m>1,\kappa_{nm}$. Here the

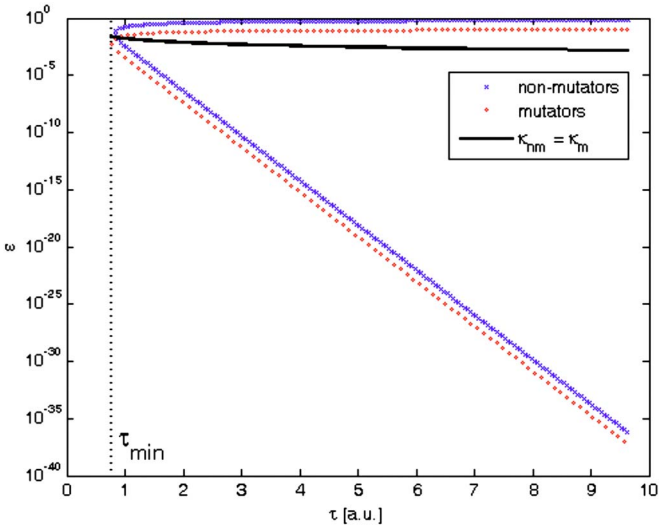

FIG. 1. (Color online) Solution of $\kappa_{nm}=1$ and $\kappa_m=1$ as a function of cycle length τ for $L_{via}=20$, $L_{rep}=10$, $S=4$, $\epsilon_r=0.1$. The area confined by the red diamonds is the survivability region of the mutators, while the area confined by the blue asterisks is the survivability region of the nonmutators. In the region above the solid curve the fraction of nonmutators is nonzero, while below the solid curve only mutators exist. τ_{min} is computed from the analytical formula given in the text.

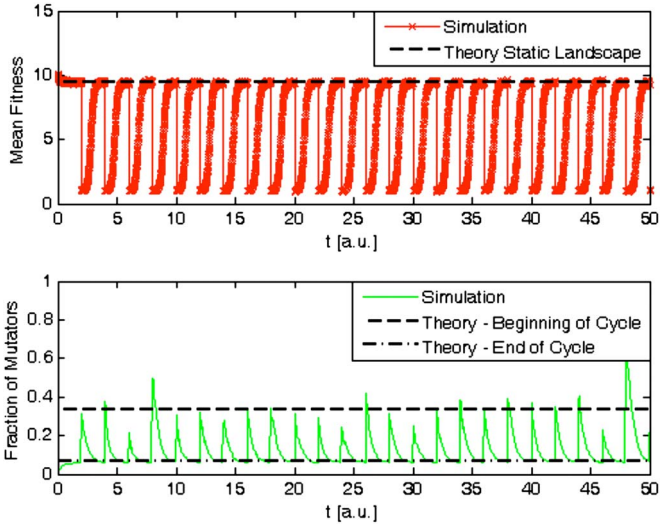

FIG. 2. (Color online) A single trajectory from a stochastic simulation for $N=1.5\times10^5$, $L_{via}=20$, $L_{rep}=10$, $S=4$, $\epsilon=0.02$, $\epsilon_r=0.1$, and $\tau=2$. The top graph shows the mean fitness of the population as a function of time, while the bottom graph shows the fraction of viable organisms that are mutators. Note that the mean fitness reaches its static landscape value. The simulated fraction of mutators at the beginning and at the end of each cycle shows good agreement with the analytical results.

mutators' effective growth is larger than both that of the background and that of the nonmutators, so that a stable quasispecies is formed, but it consists entirely of mutators. The transition between the first and second regions corresponds to a localization to delocalization transition over the repair portion of the genome that is termed the repair catastrophe [2].

Finally, the third region is defined by $\kappa_{nm},\kappa_m<1$, so that no stable quasispecies forms, and the population is unable to adapt to the changing fitness landscape.

The solutions to the equations $\kappa_{nm},\kappa_m=1$ set the upper and lower mutation thresholds for the population to survive (see Fig. 1). The upper mutation threshold is the ordinary error catastrophe for the static landscape, while the lower threshold appears only in the case of a dynamic landscape. This lower threshold arises because, in a dynamic fitness landscape, the mutation rate must have a minimal value to allow the population to adapt. Below this dynamic error threshold, the population is unable to adapt, and so no stable quasispecies is formed [11].

There is some minimal value of τ, denoted τ_{min}, below which the population cannot adapt. This corresponds to a threshold rate of change of the fitness landscape above which the population cannot adapt, irrespective of the error threshold. This minimal value for τ arises because, as τ is decreased, the mutation rate must increase to maintain a stable quasispecies. Once this minimal mutation rate exceeds the static error threshold, then no quasispecies can exist.

An analytical approximation for τ_{min} can be found by expanding Eq. (8) to second order in ϵ, solving for ϵ, and then solving for the value of τ where the two solutions are equal. Assuming that k is sufficiently larger than 1, and that τ is sufficiently large, so that $e^{k\tau}\gg e^{\tau}$, we obtain that $\tau_{min}\approx(1/k)\ln[4(k-1)(S-1)L_{via}]$.

An ϵ versus τ plot of the solution to the equation $\kappa_{nm}=\kappa_m$ yields a continuous curve, above which it can be shown that $\kappa_{nm}>\kappa_m$, and below which it can be shown that $\kappa_{nm}<\kappa_m$. An approximate expression describing this curve, one that agrees reasonably well with numerical calculations for $\tau>\tau_{min}$, is $\epsilon_{\kappa_{nm}=\kappa_m}(\tau)\approx\ln(1/\epsilon_r)/[kL(L_{via}/L-\epsilon_r)\tau-(1-\epsilon_r)]$.

The region above this curve that also lies between the lower and upper mutational thresholds of the nonmutators corresponds to a parameter regime where there is a stable quasispecies with a nonzero fraction of nonmutators. The region below this curve that also lies between the lower and upper mutational thresholds of the mutators corresponds to a parameter regime where there is a stable quasispecies consisting entirely of mutators.

We developed a stochastic code to simulate the dynamics of a population of self-replicating genomes of the form $\sigma=\sigma_{via}\sigma_{rep}$. At each time step, a given genome replicates with probability $\kappa_\sigma\Delta t$, where κ_σ is the growth constant and Δt is the size of the time step. We choose Δt to be sufficiently small so that the replication probability is much smaller than 1. Since we only allow a genome to replicate itself at most once during an iteration, making the replication probability small ensures that this restriction is a good approximation of the dynamics. After every time interval τ, a shift occurs and a new master sequence is generated to be one point mutation away from the old one. The mutation is generated at a different random location in the viable region to avoid population accumulation (to meet the initial conditions of the analytical model), i.e., the mutation position repeats itself every L_{via} peak shifts. Each realization simulated at least 10^4 time steps to make sure a periodic solution was reached. We should note that, in contrast to the analytical model, the stochastic simulations do not make any additional assumptions

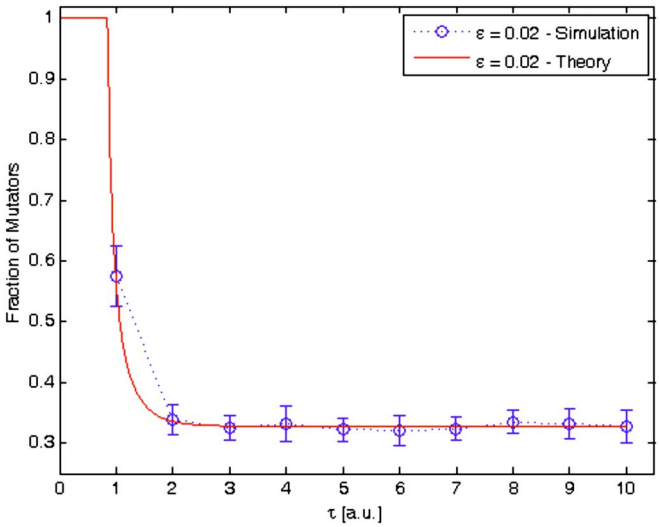

FIG. 3. (Color online) Fraction of mutators right after a peak shift for the long-time periodic solution of the quasispecies dynamics as a function of τ. $N=1.5\times10^5$, $L_{\mathrm{via}}=20$, $L_{\mathrm{rep}}=10$, $S=4$, $\epsilon=0.02$, and $\epsilon_r=0.1$. The results are averaged over ten independent runs.

beyond the ones detailed here. In particular, backmutations are allowed.

As can be seen from Figs. 2 and 3, as long as ϵ is sufficiently small, we obtain good agreement between theory and simulation, which seems to justify the various approximations made in our analytical model. However, due to stochastic effects, the fraction of mutators requires a larger population size than other evolution parameters (e.g., mean fitness, mean Hamming distance) before reasonable agreement with the theoretical results is obtained.

For sufficiently large ϵ, the agreement between our point-mutation model and the simulation results breaks down. The reason for this is that, when ϵ is small, the $n_{00}\rightarrow n_{01}$, $n_{00}\rightarrow n_{10}$, and $n_{01}\rightarrow n_{11}$ transition probabilities are all first order in ϵ, while the transition probability $n_{00}\rightarrow n_{11}$ is second order in ϵ, and is therefore neglected in a model that considers only mutation probabilities that are up to first order in ϵ.

For larger ϵ, the nonmutator to mutator transition probability, which is $1-(1-\epsilon_r\epsilon)^{L_{\mathrm{rep}}}$, goes from being first order in ϵ to zeroth order, so that the $n_{00}\rightarrow n_{11}$ transition becomes first order in ϵ as well. Therefore, if we wish to use a first-order model for larger ϵ, then this additional transition probability must be included. Including the $n_{00}\rightarrow n_{11}$ transition for larger ϵ significantly improves the theoretical prediction of the mutator fraction. For future research, we wish to develop a more systematic approach for estimating the mutator fraction in a dynamic fitness landscape that does not require the assumption of small ϵ. Furthermore, we would like to move beyond mean-field descriptions of the evolutionary dynamics and develop analytical approaches for quantifying various stochastic effects.

This research was supported by the Israel Science Foundation and by the United States–Israel Binational Science Foundation.

[1] D. Voet and J. G. Voet, *Biochemistry*, 3rd ed. (Wiley, New York, 2004).
[2] E. Tannenbaum, E. J. Deeds, and E. I. Shakhnovich, Phys. Rev. Lett. **91**, 138105 (2003).
[3] J. LeClerc, B. Li, W. Payne, and T. Cebula, Science **274**, 1208 (1996).
[4] E. Tannenbaum and E. I. Shakhnovich, Phys. Rev. E **69**, 011902 (2004).
[5] A. Sasaki and M. A. Nowak, J. Theor. Biol. **224**, 241 (2003).
[6] D. A. Kessler and H. Levine, Phys. Rev. Lett. **80**, 2012 (1998).
[7] J. M. J. Travis and E. R. Travis, Proc. R. Soc. London, Ser. B **269**, 591 (2002).
[8] P. R. Painter, Genetics **79**, 649 (1975).
[9] P. D. Sniegowski, P. J. Gerrish, and R. E. Lenski, Nature (London) **387**, 703 (1997).
[10] M. E. Palmer and M. Lipsitch, Genetics **173**, 461 (2006).
[11] M. Nilsson and N. Snoad, Phys. Rev. Lett. **84**, 191 (2000).