Evolutionary adaptation is facilitated by the presence of lethal genotypes

Viktoria Blavatska [1,2] and Bartlomiej Waclaw [1,3]

[1] *Dioscuri Centre for Physics and Chemistry of Bacteria, Institute of Physical Chemistry PAS, Kasprzaka 44/52, 01-224 Warsaw, Poland*

[2] *Institute for Condensed Matter Physics of the National Academy of Sciences of Ukraine, 79011 Lviv, Ukraine*

[3] *School of Physics and Astronomy, The University of Edinburgh, JCMB, Peter Guthrie Tait Road, Edinburgh EH9 3FD, United Kingdom*

(Received 7 February 2023; revised 21 September 2023; accepted 29 January 2024; published 15 March 2024)

The rate of adaptation in theoretical models of biological evolution generally increases with the mutation rate. However, there is a threshold beyond which mutations into lethal states lead to extinction. It would be logical to assume that eliminating such lethal states could be advantageous for evolution. Here, we demonstrate that lethal mutations actually accelerate adaptation on rugged fitness landscapes with multiple peaks and valleys in the presence of competition for resources. We investigate a modified stochastic version of the quasispecies model, incorporating two types of genotypes—viable and lethal—and show that higher rates of lethal mutations result in shorter evolution times towards the best-fit genotype. This phenomenon can be attributed to an increased frequency of traversing fitness valleys, facilitated by reduced selection pressure against less-fit variants.

DOI: 10.1103/PhysRevResearch.6.013286

I. INTRODUCTION

Understanding what determines the rate with which biological organisms evolve and acquire new adaptive traits and how to control this process is an essential problem both in theory and practice. For example, slowing down the spread of antibiotic resistance in bacteria, which can evolve in a matter of hours [1–4], is an important objective of antimicrobial stewardship. In other situations, such as directed evolution of proteins [5,6] used, e.g., to create enzymes able to degrade plastics [7] or produce biofuels [8] and drugs [9], it is desirable to speed up evolution. Moreover, the ability to manipulate the rate of evolution holds promise in combating viral and bacterial infections, as well as cancer [10].

The rate of biological evolution depends, among others, on the mutation rate [11], the structure of the fitness landscape [12–16], its temporal dynamics [17–20], fitness fluctuations [21,22], varying environmental conditions [20,23], migration among subpopulations [24], and population's spatial structure [3,4,25].

Here, we focus on the role of the fitness landscape (FL) and the mutation rate within the framework of the quasi-species model [26], extensively studied by physicists [27–32]. It has been shown that the time it takes biological evolution to reach the best-adapted genotype decreases with increasing mutation rate [33–36]. However, excessive mutation rates can lead to population delocalization, known as "error catastrophe," which can result in population extinction if the fitness landscape contains a significant number of genotypes with null or negative fitness [37,38]. This poses a challenge in novel approaches to directed evolution, where organismal survival is used for selection, bypassing the need for individual clone screening [6]. While increasing the mutation rate is desirable to accelerate evolution, the error catastrophe sets a limit on the rate of evolution in real fitness landscapes, in which nonviable mutants represent a substantial fraction of all genotypes [39].

Lethal mutations also play an important role in a novel approach to treating infections by elevating the mutation rate of the pathogen (bacteria, viruses, or cancer cells) through drugs [10,40–42]. However, the underlying theory assumes that organismal fitness is constant, which may not be true for real diseases due to time- and density-dependent selection [43,44]. This problem is highlighted by the inability of theoretical models to reproduce experiments in which evolutionary adaptation has been shown to occur even at very high mutation rates [42].

Motivated by these examples, in the present study we use mathematical modeling to answer the following questions: (1) how the presence of lethal genotypes affects the rate of evolution, and (2) how sensitive the results are to the details of the model. Counterintuitively, we show that lethal genotypes can actually speed up evolution on some FLs, for mutation rates well below the error threshold, and that this effect is quite robust, provided that there is competition for resources in the model, which causes selection to be density dependent.

II. MODEL AND RESULTS

We consider a stochastic version of the quasispecies model with viable and nonviable genotypes. In contrast to previous works, we divide up the genotype space into a region representing genes relevant for adaptation to a new environment, and another one for housekeeping genes crucial for metabolism, DNA replication, etc., which have already been highly optimized and therefore any mutation in them is likely

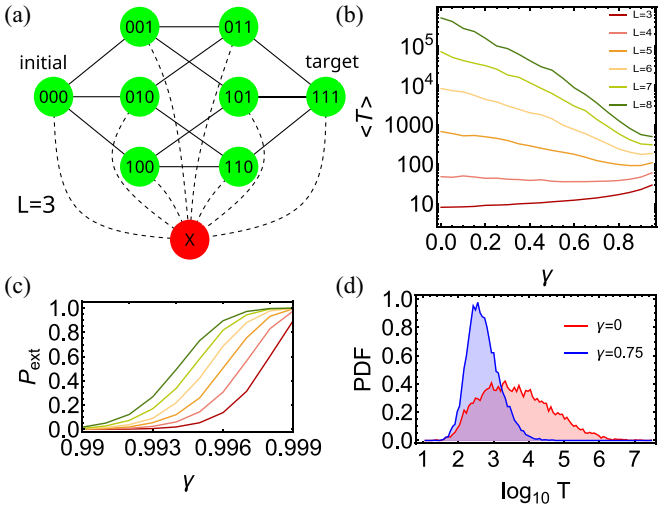

FIG. 1. (a) The genotype space of the model for $L = 3$. Green = viable states, red = lethal state. Lines represent mutations. (b) Mean adaptation time $\langle T \rangle$ decreases with increasing probability γ of lethal mutations, for $L \geqslant 5$. Colors = different L. (c) The extinction probability P_{ext} is low even for $\gamma \approx 0.99$. (d) The distribution of adaptation time $P(T)$, for $\gamma = 0$ and $\gamma = 0.75$. $L = 7$ in both cases.

to be lethal. This scenario is biologically more realistic than assigning zero fitness to random genotypes in the landscape.

To be specific, we consider a population of organisms replicating and dying stochastically. Each organism has a genotype $i = 0, \ldots, 2^L - 1$ represented by a binary sequence of length L. The state of the system is described by a vector $\{n_i\}$, where n_i is the number of organisms of type i. An organism of type i replicates with rate f_i and dies with rate N/K, where $N = \sum_i n_i$, and K is the (soft) carrying capacity of the system. Upon replication, the organism produces either a copy of itself or a mutant. The mutant can be either viable or nonviable (lethal). We assume that faithful replication occurs with probability $1 - \mu - \gamma$, a viable mutant is generated with probability μ, and a lethal mutant with probability γ. A lethal mutant is instantaneously removed from the system. The genotype of a viable mutant is obtained by inverting a randomly selected letter (0 or 1) of the binary representation of i. The genotype space is therefore an L-dimensional hypercube, with an additional node representing the lethal state connected to all other genotypes [see Fig. 1(a)].

The replication rates $\{f_i\}$ are drawn independently from the uniform distribution on [0,1), except for $f_0 = 0.5$ and $f_{2^L-1} = 1$, which we fix so that $i = 2^L - 1$ is always the fittest genotype, and the initial genotype has intermediate fitness. The system is initialized with K organisms of genotype $i = 0$ and the simulation stops when the first organism of the fittest genotype emerges.

For $\gamma = 0$, the model is essentially the stochastic quasi-species model with a maximally rugged fitness landscape [30]. For $\gamma > 0$, the fitness landscape contains a fraction $\gamma/(\mu + \gamma)$ of lethal genotypes. By construction, $\mu + \gamma$ must be smaller than one since it is the total mutation probability. We cannot therefore make either μ or γ too large.

We are interested in the large-population/large-mutation rate regime relevant for both directed evolution and lethal mutagenesis. In this limit, low-fitness variants do not fix in the population, and therefore the rate of evolution decreases monotonously with population size [45]. Since the presence of lethal genotypes reduces the population size, evolution is expected to be slower for $\gamma > 0$. We shall see that this expectation is not always correct, and that lethal genotypes, while not participating in the evolutionary process, have a significant effect on the rate of evolution.

To compare the model with and without lethal genotypes, we simulated 1000 copies of the model using a kinetic Monte Carlo algorithm [46] for $K = 1000$, $\mu = 0.04$, $L = 3, \ldots, 8$, and a range of $\gamma = 0, \ldots, 0.96$, and measured the time T it took for a single organism of best-adapted genotype to evolve. We used the same sequence of randomly generated FLs for each γ. Figure 1(b) shows that the average adaptation time increases exponentially with L. However, the increase is slower in the model with $\gamma > 0$. In fact, for $L > 4$, the presence of the lethal genotype speeds up evolution. For $L = 8$ and $\gamma = 0.75$, evolution is three orders of magnitude faster than for $\gamma = 0$. The same effect can be seen in the Moran-process version of the model, in which the population size is kept fixed and does not fluctuate (see Fig. 5 in Appendix A), which shows that it is not unique to the quasi-species model. Moreover, we additionally performed the analysis for a much larger carrying capacity and much smaller mutation rate (see Appendix B), confirming qualitatively the same tendency: the evolution process is facilitated by increasing the probability of lethal mutations.

This behavior is counterintuitive since the total number of organisms in the system is significantly reduced for $\gamma > 0$; for example, for $\gamma = 0.75$ the population reaches only about 20% of the maximum capacity. Although the per-capita birth rate does not depend on population size, the average rate with which mutants are generated is lower because there are fewer births. Moreover, the effect shows up only on large-enough fitness landscapes; for $L < 4$, the nonlethal model is actually faster. The lethal-genotype model is slower for any L also on a flat landscape (all fitnesses the same, results not shown), on which there is only genetic drift and no selection (except for nonlethality). The effect is also unrelated to the increased probability of extinction in the lethal model, since extinction becomes likely only for γ very close to 1 [Fig. 1(c)]; in this limit Fig. 1(b) shows that $\langle T \rangle$ increases slightly with γ.

In what follows we shall focus on two cases: $\gamma = 0$ (nonlethal) and $\gamma = 0.75$ (lethal), for which the extinction probability is effectively zero for $K = 1000$. The choice of $\mu = 0.04$, $\gamma = 0.75$ may be interpreted as about 1 in 20 mutations being viable, the rest being lethal (see, however, Fig. 6 in Appendix B for much lower μ and γ). Figure 1(d) shows the distribution of adaptation times for $L = 7$. While the distributions are very broad for both models (as expected due to the stochastic nature of the evolutionary process), we notice that the nonlethal model has a distribution skewed to the right, which contributes to the much longer average adaptation time.

To understand whether the two models differ only for certain fitness landscapes or all landscapes, we set $\mu = 0.04$, and $\gamma = 0$ (no lethal genotypes) or $\gamma = 0.75$ (lethal genotypes present), and run the models 20 times on 1000 random landscapes (identical for both models). We then compared evolutionary pathways in both models for 100 landscapes

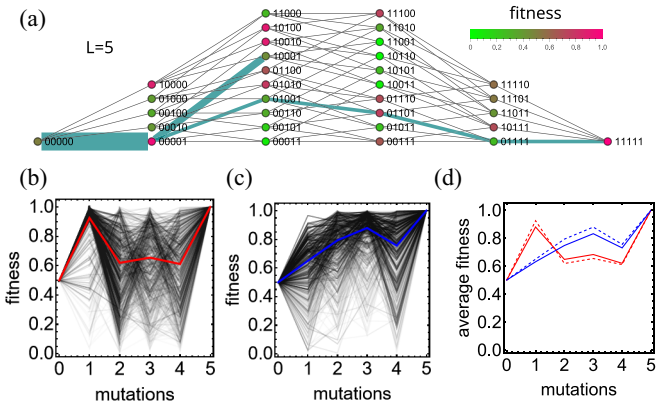

FIG. 2. Evolutionary trajectories for $L = 5$. (a) Example pathway in the nonlethal model for $\mu = 0.04$, $\gamma = 0$. (b) and (c) Fitness along evolutionary pathways, for the slowest 10% of 1000 random FLs (b) and 10% fastest FLs (c). Red and blue thick lines are averages over individual trajectories. (d) Average fitness along adaptive trajectories for slow (red) and fast (blue) FLs, for $\gamma = 0$ (solid) and $\gamma = 0.75$ (dashed).

with largest and smallest differences in the adaptation time. We shall call such landscapes "slow" and "fast," respectively. Figure 2(a) shows an example pathway in the nonlethal model on a slow landscape. The first mutation increased fitness to approximately that of the final genotype, but subsequent mutations led through a fitness valley. This is typical for slow landscapes [Fig. 2(b)]. In contrast, evolutionary pathways on fast landscapes usually involve a more gradual fitness increase, with fewer and shallower fitness valleys [Fig. 2(c)]. Interestingly, the average fitness profile along the evolutionary trajectory is similar for the lethal and nonlethal model [Fig. 2(d)]. This suggests that the evolutionary speed-up provided by lethal genotypes does not rely on following different pathways, but rather on the enhanced rate of crossing of fitness valleys.

To test this hypothesis, we investigated a simpler model with a one-dimensional (1D) fitness landscape [Fig. 3(a)]. Mutations can only change genotype i to $i \pm 1$ with probability $\mu/2$ in either direction, and the fitness values are $f_0 = 1$, $f_1 = f_2 = \cdots = f_{L-1} = 1 - \delta$, $f_L = 1$, i.e., a flat fitness valley of depth δ separates the initial and the final genotype. This represents a single evolutionary pathway from Fig. 2 and is a special case of a more generic problem considered in Ref. [13], which significantly simplifies

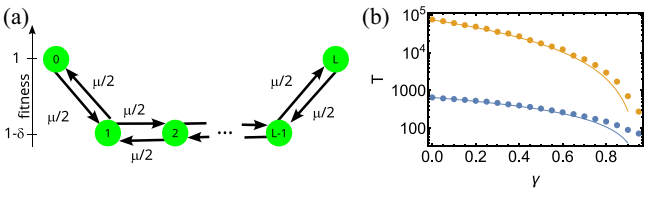

FIG. 3. Evolution in a one-dimensional model with a wide fitness valley of depth $\delta = 0.7$. (a) Schematic representation of the fitness landscape. (b) Time to reach genotype $i = L$. Points = simulation, line = theoretical predictions from Eq. (9): $L = 3$ (blue), $L = 4$ (yellow).

mathematical analysis. Numerical simulations confirm that indeed the lethal-genotype model leads to much faster adaptation time on this fitness landscape [Fig. 3(b)].

To understand this, we considered a deterministic counterpart of the stochastic model. Neglecting stochastic fluctuations, the abundances $\{n_i\}$ of organisms evolve according to the following set of equations:

$$\frac{dn_0}{dt} = n_0 f_0(1-\gamma-\mu) + \mu n_1 f_1 - n_0 N(t)/K, \quad (1)$$

$$\frac{dn_i}{dt} = n_i f_i(1-\gamma-\mu) + (\mu/2)(n_{i+1}f_{i+1} + n_{i-1}f_{i-1}) - n_i N(t)/K, \quad (2)$$

where $N(t) = \sum_i n_i(t)$. Assume for a while that the final state has fitness zero, so it acts as an absorbing boundary. The system of equations admits then a steady-state solution, with abundances determined by the nonlinear set of equations,

$$0 = n_0 f_0(1-\gamma-\mu) + \mu n_1 f_1 - n_0 N/K, \quad (3)$$

$$0 = n_i f_i(1-\gamma-\mu) + (\mu/2)(n_{i+1}f_{i+1} + n_{i-1}f_{i-1}) - n_i N/K, \quad (4)$$

with $N = \sum_i n_i$. These equations will also correctly describe the quasi-stationary distribution for the case of nonzero fitness at $i = L$, as long as the transition rate from $i = L-1$ to $i = L$ (proportional to μ) is small. For small μ, we expect $n_{i+1} \ll n_i$, which enables us to write

$$n_i \approx n_{i-1} \frac{f_{i-1}\mu/2}{N/K - f_i(1-\gamma-\mu)}. \quad (5)$$

This means that the total population size N will be dominated by n_0. Inserting $N = n_0$ into Eq. (1), we obtain that

$$n_0 = N \approx f_0 K(1-\gamma-\mu), \quad (6)$$

and hence, for $i > 0$,

$$n_i \approx n_{i-1} \frac{f_{i-1}\mu/2}{(f_0 - f_i)(1-\gamma-\mu)}, \quad (7)$$

and finally

$$n_i \approx \frac{K\mu}{2\delta}\left(\frac{(1-\delta)\mu}{2\delta(1-\gamma-\mu)}\right)^{i-1}. \quad (8)$$

The above equation shows that (1) n_i decreases exponentially with i, and (2) the rate of decrease is smaller for nonzero γ. This means that the abundance of genotype $L-1$ preceding the best-adapted genotype increases with increasing γ, even though the total population size decreases, as long as L is sufficiently large, γ is not too large, and our approximations remain valid. If the adaptation time T is now limited by the last mutational step of going from $i = L-1$ to $i = L$, we can write that

$$T = \frac{1}{\frac{\mu}{2}(1-\delta)n_{L-1}} = \frac{4\delta}{K\mu^2(1-\delta)}\left(\frac{2\delta(1-\gamma-\mu)}{\mu(1-\delta)}\right)^{L-2}. \quad (9)$$

The above formula agrees very well with computer simulations of the 1D model [Fig. 3(b)], which supports our hypothesis that long fitness valleys slow down evolution in the full model, and that the presence of lethal genotypes facilitates

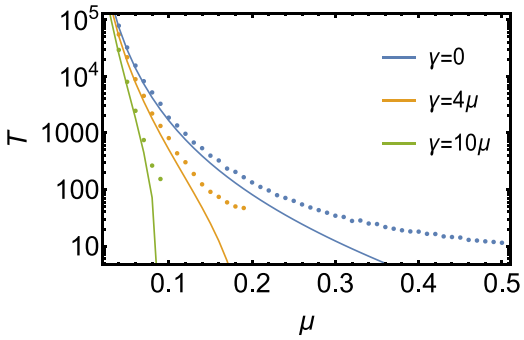

FIG. 4. The mean adaptation time T for the 1D model with $L = 4$ in which $\gamma = M\mu$, i.e., a fixed fraction $M/(M + 1)$ of mutations is lethal. Points = computer simulations averaged over 1000 replicates, lines = Eq. (9). There are no points above $\mu_{\text{crit}} = 1/(M + 1)$, which is the maximum permitted μ in our model. Parameters: $K = 10^3$, $\delta = 0.7$.

valley crossing by increasing the abundance of less-fit but viable genotypes.

It is interesting to compare this result with the observation of Ref. [47] that decreased turnover increases the fixation probability of a mutation with a fixed selective advantage, thus decreasing the time to fixation if mutants occur spontaneously and with a small rate. In our model, turnover, defined as the sum of per-capita birth and death rates, is indeed reduced for $\gamma > 0$: for strain i, per-capita turnover is $f_i + N/K$, with N decreasing with increasing γ. However, there is no fixation of valley genotypes in our model, so the result of Ref. [47] does not directly apply.

Thus far we kept μ fixed while varying γ. In reality, these two mutation rates are interdependent since increasing the viable mutation rate μ by, e.g., UV radiation also increases the lethal rate γ. Figure 4 shows that, when γ and μ are both varied such that their ratio remains constant, the presence of lethal genotypes speeds up adaptation when compared to the model without such genotypes, for the same value of μ. The figure also shows that evolution accelerates faster with μ [the slope of $T(\mu)$ gets steeper] for larger fractions of lethal genotypes.

Equation (9) can be used to make predictions for the full model with 2^L genotypes by averaging T over the distribution of fitness valley lengths (see Appendix C). Figure 8 shows that the ratio of adaptation times for $\gamma > 0$ and $\gamma = 0$ estimated in this way approximately agrees with the ratio obtained from the data presented in Fig. 1.

Motivated by our examples of directed evolution and lethal mutagenesis, we have considered high mutation rates. However, Eq. (9) shows that lethal mutations should speed up adaptation also in the case of much smaller μ, as long as γ is not too small. This is confirmed by data shown in Appendix B, Figs. 6 and 7, where we consider $\mu = 10^{-5}$ and large population sizes ($K = 10^9$) typical for microbial evolutionary experiments.

III. CONCLUSIONS

To summarize, we have observed a significant reduction of the adaptation time in the presence of lethal mutations. The effect is caused by reduced selection against less-fit genotypes; this increases the abundance of valley genotypes and hence the rate with which best-adapted mutants are generated.

Our results have important implications. Firstly, directed evolution can be substantially sped up by deliberately making some genotypes nonviable. This can be achieved, e.g., in a bioreactor (either conventional or microfluidics based) by increasing the dilution rate, which causes all variants whose replication rate is lower than the dilution rate to be washed out.

Secondly, our results urge caution regarding lethal mutagenesis as a method for treating infections. Although the presence of lethal genotypes lowers the threshold for the mutation rate that leads to the extinction of the pathogen, it also makes the time to adaptation fall off faster with μ in the presence of fitness valleys and density-dependent selection. These factors must be taken into account when modeling treatment based on lethal mutagenesis. Our results also suggest two further problems with this approach. Since the mutation rate is likely to vary in time and space as a result of nonuniform distribution of the mutagenic drug, the pathogen may evolve faster in areas of sublethal concentration of the mutagen. In addition, increasing the rate of death by, e.g., adjuvant therapy with pathogen-killing drugs will increase the rate of evolution of the surviving pathogenic organisms. In combination with drug gradients, this will create a complex, space- and time-dependent fitness landscape in which the drugs, instead of killing the pathogen, may help it evolve resistance to treatment.

The software required to reproduce the results presented here can be found in Ref. [48].

ACKNOWLEDGMENTS

We acknowledge funding under Dioscuri, a programme initiated by the Max Planck Society, jointly managed with the National Science Centre in Poland, and mutually funded by Polish Ministry of Science and Higher Education and German Federal Ministry of Education and Research (Grant No. UMO-2019/02/H/NZ6/00003). V.B. is grateful for support from the U.S. National Academy of Sciences (NAS) and the Polish Academy of Sciences (PAS) for scientists from Ukraine.

APPENDIX A: MORAN PROCESS VERSION OF THE MODEL

In order to demonstrate the validity of our main statement regarding the acceleration of adaptation by lethal genotypes in various evolutionary models, we examine the Moran process, in which the population size is kept fixed and does not fluctuate [11,49]. In the Moran-like version of our model, during each time step, one organism is randomly chosen for reproduction based on its fitness, while another organism is randomly chosen for removal, regardless of its fitness. The offspring of the selected organism replaces the removed one, ensuring that the total population size K remains unchanged.

As in the quasispecies model considered in the main text, the offspring can either inherit the genotype of its parent with probability $1 - \mu - \gamma$, or it can mutate into either another

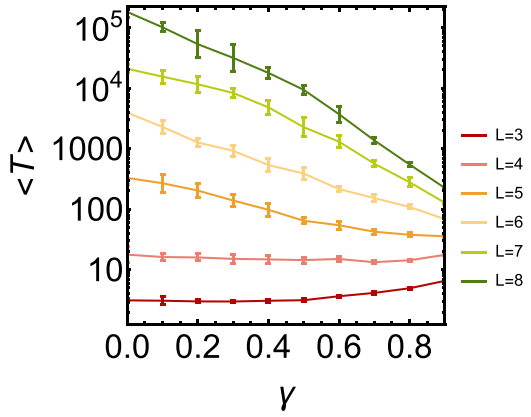

FIG. 5. The mean adaptation time $\langle T \rangle$ for the Moran model as a function of the probability γ of lethal mutations, for different L. Other parameters are $K = 1000$ and $\mu = 0.04$.

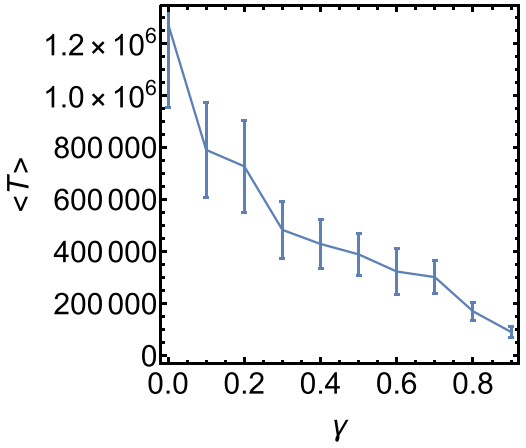

FIG. 6. The mean adaptation time $\langle T \rangle$ as a function of the lethal mutation probability γ for the full model with $L = 4, K = 10^9$, $\mu = 10^{-5}$. The simulation has been carried out using the tau-leaping algorithm.

viable genotype (with probability μ) or the lethal genotype (with probability γ). Organisms with lethal genotypes are not removed from the system but they do not reproduce (their fitness is zero). Other components of the model (fitness landscape, initial condition) are the same as for the quasi-species model.

We take N_{viable} steps of the Moran process, where N_{viable} is the number of viable organisms, to represent one unit of time, so that organisms with fitness one reproduce on average once per unit of time, similarly to the quasi-species version of the model. Therefore, the time variable is incremented in each step by $1/N_{\text{viable}}$.

Figure 5 shows the adaptation time versus the lethal mutation probability γ for this model, for different sequence lengths L and identical parameters μ, K as in Fig. 1(b) in the main text. The results are qualitatively the same as in Fig. 1(b) for the quasi-species model: the adaptation time decreases with increasing γ for $L > 4$.

APPENDIX B: LARGE POPULATION SIZES

Performing an exact kinetic Monte Carlo simulation [46] of the quasispecies model becomes impractical when dealing with large population sizes (on the order of $K \sim 10^9$ or more), which are commonly encountered in laboratory experiments on microbial evolution. This is due to the algorithm's running time, which increases linearly with the population size. To examine the applicability of our theory [Eq. (9)] to such large populations, we employ an approximate but much faster tau-leaping method [50], originally developed for stochastic simulations of chemical systems.

In this new algorithm, the state of the system is described by the same vector of genotype abundances $\{n_i(t)\}$, with $i = 0, \ldots, 2^L - 1$, as in the exact algorithm. However, instead of dealing with only a single organism in each time step, we determine by how much the state vector should change during a small but finite time interval $\Delta t > 0$, and update it accordingly. Assuming that the rates of all processes (replication, death, mutation) are approximately constant during this small time interval, Ref. [50] shows that the number of events for each process will be a binomially distributed random number with the probability of success equal to the rate of the process times Δt, and with the number of trials equal to the number of organisms that participate in said process.

A single step of the algorithm that simulates our stochastic quasispecies model consists of drawing three binomially distributed random numbers for each genotype i:

$$r_i = \text{Binomial}[\Delta t f_i(1 - \gamma - \mu), n_i],$$
$$d_i = \text{Binomial}[\Delta t (N/K), n_i],$$
$$m_i = \text{Binomial}[\Delta t f_i \mu, n_i],$$

where $\text{Binomial}(p, n)$ represents a binomial random number with n trials and the success probability p, r_i is the number of reproduction events that lead to viable, nonmutated organisms, d_i is the number of deaths, and m_i is the number of mutated viable organisms created during the time interval Δt. The number of organisms n_i is then updated as $n_i \to n_i + (r_i - d_i)$, and each of the m_i viable mutants is assigned a new genotype according to our mutation graph (hypercube) and the corresponding n_i is increased by 1. After all n_i's have been updated, the time variable t is increased: $t \to t + \Delta t$.

In the limit $\Delta t \to 0$, this algorithm reduces to the exact kinetic Monte Carlo algorithm used in the main text. For a finite Δt, a small error is introduced because the assumption of the rates not changing during the time interval is violated. In our simulations, we take $\Delta t = 1/128$ as a reasonable compromise between the speed and the accuracy of the simulation. We do not dynamically change Δt as is custom in similar algorithms because the rates in our model do not change very much; adaptive control of Δt would be an unnecessary computational burden.

We applied the algorithm to determine the time to adaptation for different lethal mutation probabilities γ for a much larger carrying capacity ($K = 10^9$) and much smaller mutation rate ($\mu = 10^{-5}$) than in Fig. 1 in the main text. Figure 6 confirms that our predictions obtained before for small population sizes remain valid also for large K: the adaptation time decreases with increasing probability of lethal mutations.

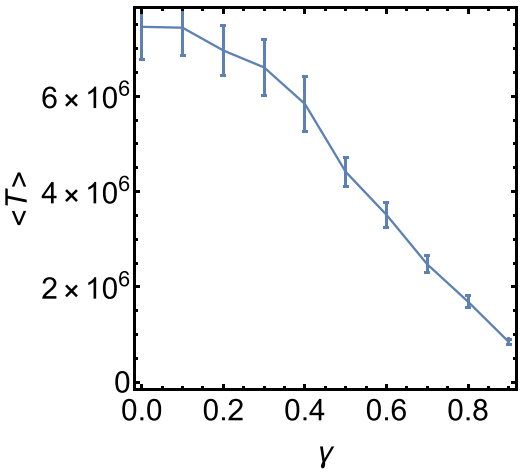

FIG. 7. The mean adaptation time $\langle T \rangle$ as a function of γ for the one-dimensional model with a wide fitness valley. Parameters: $K = 10^9$, $\mu = 10^{-5}$, $\delta = 0.6$. All results have been obtained using the tau-leaping algorithm.

We also investigated the one-dimensional fitness landscape model with a wide fitness valley as in Fig. 3(a), with the same $K = 10^9$, $\mu = 10^{-5}$, and $L = 3$. Figure 7 shows that the adaptation time again decreases with increasing γ.

We note that the mutation probability we used here ($\mu = 10^{-5}$) is still much higher than what is typical for nonmutator bacterial strains ($\mu \sim 10^{-9}$) but lower than for RNA-based viruses ($\mu \sim 10^{-4}$) [51]. Mutator strains which have a defective DNA repair system can have 100-fold higher mutation rates [52], approaching $\mu = 10^{-7}$. Simulations for such a small μ are certainly not feasible because the adaptation time (and hence the CPU time) increases quickly with decreasing μ as shown by Eq. (9).

However, in the case of directed evolution, which is our focus here, the mutation probability can be artificially increased to $\mu = 10^{-4}$–10^{-3} [6], which is within the range of μ used in this work.

APPENDIX C: ADAPTATION TIME IN THE FULL MODEL FROM THE ANALYTIC EXPRESSION FOR THE 1D MODEL

We can use Eq. (9) to make predictions for the full model with 2^L genotypes. First, we notice [Fig. 2(d)] that for a fixed fitness landscape with a sufficiently deep and wide valley, evolution quickly reaches a local fitness maximum one mutation away from the initial genotype. To use Eq. (9) we must find the distribution p_l of fitness valley length l. We can then calculate the ratio of the adaptation times for the model with and without lethal genotypes as follows:

$$\sum_{l=0}^{L-1} T_l(\gamma) p_l \Big/ \sum_{l=0}^{L-1} T_l(0) p_l. \tag{C1}$$

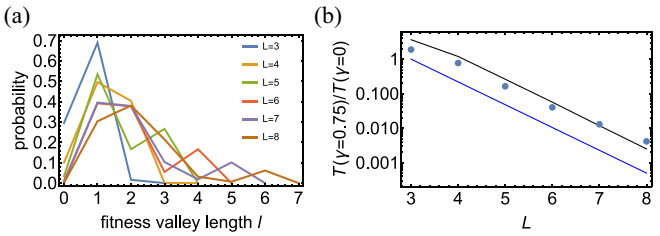

FIG. 8. The linear-FL model predicts the adaptation time in the full model. (a) Probability of finding a fitness valley of length l, for different L. See the main text for the definition of l. (b) The ratio of adaptation times for the model with $\gamma = 0.75$ and $\gamma = 0$, for different L. Points = results from Fig. 1, black line = Eq. (C1), blue line = Eq. (C2).

To find $\{p_l\}$ numerically, we generated 10 000 random FLs, and found all fitness valleys, defined as consecutive runs of genotypes with monotonously decreasing fitness. Specifically, for each FL we found all such valleys along trajectories starting at the single-mutated genotype with maximum fitness and ending at the best adapted genotype. For each trajectory, we found the length (the number of links along which fitness decreases) of the shortest valley. We then took the minimum of all maximum lengths for each trajectory, and repeated this process for all trajectories and FLs. This gave the distribution of the minimum-length valley that evolution might encounter in our model. Such valleys would form a bottleneck, limiting the rate of evolution on each specific FL.

Figure 8(a) shows the distribution of fitness valley lengths for different sizes L of the FL. Using the numerically obtained $p_l(L)$, we calculated the ratio (C2) for different L, assuming $\delta = 0.3$ (average depth of fitness valley from Fig. 2). Figure 8(b) shows that this estimate approximately agrees with the ratio obtained from the adaptation times from Fig. 1.

Interestingly, an even simpler approach amounting to replacing $L \to L - 1$ in Eq. (9) and calculating the ratio of adaptation times as

$$\frac{T_{L-1}(\gamma > 0)}{T_{L-1}(\gamma = 0)} = \left(\frac{(1 - \gamma - \mu)}{(1 - \mu)} \right)^{L-3} \tag{C2}$$

qualitatively reproduces the data [Fig. 8(b)]. This is because, as seen in Fig. 8(a), a fitness landscape for binary sequences of length L always has a nonzero probability of having a fitness valley of length $L - 2$. These longest valleys dominate the adaptation time due to its exponential dependence on the fitness valley length. Even though the probability p_{L-2} decreases exponentially with L as $\sim \exp(-0.5L)$ [Fig. 8(a)], the rate of decrease is not sufficient to overcome the exponential increase of $T_l(\gamma)$ with rate $2\delta(1 - \gamma - \mu)/[\mu(1 - \delta)]$ as long as μ is sufficiently small. However, due to the presence of shorter fitness valleys in many landscapes [Fig. 8(a)], the lethal-genotype model will perform better only on a subset of FLs which do not have such short valleys.

[1] Q. Zhang, G. Lambert, D. Liao, H. Kim, K. Robin, C.-k. Tung, N. Pourmand, and R. H. Austin, Acceleration of emergence of bacterial antibiotic resistance in connected microenvironments, Science **333**, 1764 (2011).

[2] E. Toprak, A. Veres, J.-B. Michel, R. Chait, D. L. Hartl, and R. Kishony, Evolutionary paths to antibiotic resistance under dynamically sustained drug selection, Nat. Gen. **44**, 101 (2012).

[3] P. Greulich, B. Waclaw, and R. J. Allen, Mutational pathway determines whether drug gradients accelerate evolution of drug-resistant cells, Phys. Rev. Lett. **109**, 088101 (2012).

[4] R. Hermsen, J. B. Deris, and T. Hwa, On the rapidity of antibiotic resistance evolution facilitated by a concentration gradient, Proc. Natl. Acad. Sci. USA **109**, 10775 (2012).

[5] F. H. Arnold, Design by directed evolution, Acc. Chem. Res. **31**, 125 (1998).

[6] M. S. Packer and D. R. Liu, Methods for the directed evolution of proteins, Nat. Rev. Genet. **16**, 379 (2015).

[7] E. L. Bell, R. Smithson, S. Kilbride, J. Foster, F. J. Hardy, S. Ramachandran, A. A. Tedstone, S. J. Haigh, A. A. Garforth, P. J. R. Day, C. Levy, M. P. Shaver, and A. P. Green, Directed evolution of an efficient and thermostable PET depolymerase, Nat. Catal. **5**, 673 (2022).

[8] S. Atsumi and J. C. Liao, Directed evolution of *Methanococcus jannaschii* citramalate synthase for biosynthesis of 1-propanol and 1-butanol by *Escherichia coli*, Appl. Environ. Microbiol. **74**, 7802 (2008).

[9] A. M. Davis, A. T. Plowright, and E. Valeur, Directing evolution: The next revolution in drug discovery? Nat. Rev. Drug Discov. **16**, 681 (2017).

[10] Y. Anciaux, A. Lambert, O. Ronce, L. Roques, and G. Martin, Population persistence under high mutation rate: From evolutionary rescue to lethal mutagenesis, Evolution **73**, 1517 (2019).

[11] M. A. Nowak, *Evolutionary Dynamics* (Harvard University Press, Cambridge, 2006).

[12] D. Weinreich and L. Chao, Rapid evolutionary escape by large populations from local fitness peaks is likely in nature, Evolution **59**, 1175 (2005).

[13] D. B. Weissman, M. M. Desai, D. S. Fisher, and M. W. Feldman, The rate at which asexual populations cross fitness valleys, Theor. Popul. Biol. **75**, 286 (2009).

[14] J. A. G. M. de Visser, S.-C. Park, and J. Krug, Exploring the effect of sex on empirical fitness landscapes, Am. Nat. **174**, S15 (2009).

[15] L. Tan, S. Serene, H. X. Chao, and J. Gore, Hidden randomness between fitness landscapes limits reverse evolution, Phys. Rev. Lett. **106**, 198102 (2011).

[16] I. G. Szendro, M. F. Schenk, J. Franke, J. Krug, and J. A. G. M. d. Visser, Quantitative analyses of empirical fitness landscapes, J. Stat. Mech. (2013) P01005.

[17] N. Kashtan, E. Noor, and U. Alon, Varying environments can speed up evolution, Proc. Natl. Acad. Sci. USA **104**, 13711 (2007).

[18] V. Mustonen, J. Kinney, C. G. Callan, Jr., and M. Lässig, Energy-dependent fitness: A quantitative model for the evolution of yeast transcription factor binding sites, Proc. Natl. Acad. Sci. USA **105**, 12376 (2008).

[19] P. Ashcroft, P. M. Altrock, and T. Galla, Fixation in finite populations evolving in fluctuating environments, J. R. Soc. Interface **11**, 20140663 (2014).

[20] A. Taitelbaum, R. West, M. Assaf, and M. Mobilia, Population dynamics in a changing environment: Random versus periodic switching, Phys. Rev. Lett. **125**, 048105 (2020).

[21] V. Mustonen and M. Lässig, Molecular evolution under fitness fluctuations, Phys. Rev. Lett. **100**, 108101 (2008).

[22] D. A. Charlebois, N. Abdennur, and M. Kaern, Gene expression noise facilitates adaptation and drug resistance independently of mutation, Phys. Rev. Lett. **107**, 218101 (2011).

[23] R. Hermsen and T. Hwa, Sources and sinks: A stochastic model of evolution in heterogeneous environments, Phys. Rev. Lett. **105**, 248104 (2010).

[24] P. Lombardo, A. Gambassi, and L. Dall'Asta, Nonmonotonic effects of migration in subdivided populations, Phys. Rev. Lett. **112**, 148101 (2014).

[25] E. A. Martens, R. Kostadinov, C. C. Maley, and O. Hallatschek, Spatial structure increases the waiting time for cancer, New J. Phys. **13**, 115014 (2011).

[26] M. Eigen and P. Schuster, A principle of natural self-organization, Naturwissenschaften **64**, 541 (1977).

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

[28] J. Krug and C. Karl, Punctuated evolution for the quasispecies model, Physica A **318**, 137 (2003).

[29] E. Muñoz, J.-M. Park, and M. W. Deem, Quasispecies theory for horizontal gene transfer and recombination, Phys. Rev. E **78**, 061921 (2008).

[30] J.-M. Park, E. Muñoz, and M. W. Deem, Quasispecies theory for finite populations, Phys. Rev. E **81**, 011902 (2010).

[31] M. Ancliff and J.-M. Park, Spin coherent state representation of the Crow-Kimura and Eigen models of quasispecies theory, J. Stat. Phys. **143**, 636 (2011).

[32] B. Waclaw, R. J. Allen, and M. R. Evans, Dynamical phase transition in a model for evolution with migration, Phys. Rev. Lett. **105**, 268101 (2010).

[33] P. R. A. Campos, C. Adami, and C. O. Wilke, Optimal adaptive performance and delocalization in NK fitness landscapes, Physica A **304**, 495 (2002).

[34] K. Jain and J. Krug, Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes, Genetics **175**, 1275 (2007).

[35] K. Jain and J. Krug, Adaptation in simple and complex fitness landscapes, in *Structural Approaches to Sequence Evolution* (Springer, New York, 2007), p. 299.

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

[37] H. Tejero, A. Marín, and F. Montero, Effect of lethality on the extinction and on the error threshold of quasispecies, J. Theor. Biol. **262**, 733 (2010).

[38] H.-J. Woo and J. Reifman, A quantitative quasispecies theory-based model of virus escape mutation under immune selection, Proc. Natl. Acad. Sci. USA **109**, 12980 (2012).

[39] A. Eyre-Walker and P. D. Keightley, The distribution of fitness effects of new mutations, Nat. Rev. Genet. **8**, 610 (2007).

[40] A. Arias, L. Thorne, and I. Goodfellow, Favipiravir elicits antiviral mutagenesis during virus replication in vivo, eLife **3**, e03679 (2014).

[41] J. J. Bull and C. O. Wilke, Lethal mutagenesis of bacteria, Genetics **180**, 1061 (2008).

[42] R. Springman, T. Keller, I. J. Molineux, and J. J. Bull, Evolution at a high imposed mutation rate: Adaptation obscures the load in phage T7, Genetics **184**, 221 (2010).

[43] B. R. Levin and R. Antia, Why we don't get sick: The within-host population dynamics of bacterial infections, Science **292**, 1112 (2001).

[44] Y. Fu, B. T. Ho, and J. J. Mekalanos, Tracking *Vibrio cholerae* cell-cell interactions during infection reveals bacterial population dynamics within intestinal microenvironments, Cell Host Microbe **23**, 274 (2018).

[45] K. Jain, J. Krug, and S.-C. Park, Evolutionary advantage of small populations on complex fitness landspace, Evolution **65**, 1945 (2011).

[46] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. **81**, 2340 (1977).

[47] T. Kuosmanen, S. Särkkä, and V. Mustonen, Turnover shapes evolution of birth and death rates, bioRxiv 2022.07.11.499527 (2022).

[48] https://github.com/Dioscuri-Centre/quasilethal.

[49] P. A. P. Moran, *The Statistical Processes of Evolutionary Theory* (Clarendon Press, Oxford, 2006).

[50] D. T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems, J. Chem. Phys. **115**, 1716 (2001).

[51] J. W. Drake, B. Charlesworth, D. Charlesworth, and J. F. Crow, Rates of spontaneous mutation, Genetics **148**, 1667 (1998).

[52] H. Lee, E. Popodi, H. Tang, and P. L. Foster, Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing, Proc. Natl. Acad. Sci. USA **109**, E2774 (2012).