Fitness Distributions in Exponentially Growing Asexual Populations

Susanna C. Manrubia, Ester Lázaro, and Juan Pérez-Mercader
Centro de Astrobiología, CSIC-INTA, Carretera de Ajalvir kilómetro 4, 28850 Torrejón de Ardoz, Madrid, Spain

Cristina Escarmís and Esteban Domingo
Centro de Biología Molecular Severo Ochoa, CSIC-UAM, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain
(Received 20 December 2002; published 8 May 2003)

We explore a mean-field model for the evolution of exponentially growing populations of mutating replicators. Motivated by recent *in vitro* experiments devised to analyze phenotypic properties of bacterial and viral populations subjected to serial population transfers, we allow our *in silico* individuals to undergo unrestricted growth before applying bottleneck events. Different dynamical regimes of our model can be mapped to different experimental situations. Numerical and analytical results for fitness distributions calculated at the statistically stationary states of the dynamics compare favorably with available experimental data. Our model and results provide a common framework to better understand populations evolving under different selection pressures.

DOI: 10.1103/PhysRevLett.90.188102 PACS numbers: 87.23.–n, 87.10.+e

Mutation and selection are the main drivers in biological evolution. Rates of spontaneous mutation per genome and generation differ largely among groups, and result in differences in the adaptive capacity of organisms. RNA lytic viruses with genome length of around 10^4 nucleotides have average mutation rates of one nucleotide per genome and generation [1]. In bacteria, this rate decreases to $1/300$, while in higher eukaryotes varies from 0.1 to 100 per genome and sexual generation [2]. The selective pressure exerted upon evolving groups depends strongly on the initial population size and the time that the individuals are allowed to replicate unrestrictedly.

The effect of population size on fitness evolution has been extensively studied for RNA viruses. A common protocol is the one of serial transfer experiments [3]. First, a population of given size is allowed to multiply for a number of generations. Then, a subpopulation of the final sample is used to restart the process. Serial transfers with large founder populations usually lead to increases in the average fitness due to selection of the best adapted phenotypes [4,5], while repeated bottlenecks are associated with decreases in fitness and fixation of deleterious mutations (Muller's ratchet) [6–8]. The number of generations during which the population grows between successive bottlenecks is directly related to the time allowed for advantageous mutants to appear and become established in the system.

Models aimed at describing the effect of Muller's ratchet in asexual populations usually consider that deleterious mutations occur at a low but positive rate, and that the probability that a back mutation takes place is negligible [9–11]. As a result, a population tends to be represented only by the less fit variant, sits close to the extinction threshold, and displays very low diversity. In spite of this, serial bottleneck experiments report large fluctuations in fitness and phenotypically diverse populations [8,12]. Indeed, increases in fitness as a result of compensatory mutations have often been observed [5]. A simple model studying the increase of fitness in a growing population has been analyzed in [13].

In this Letter, we analyze a population dynamics model originally devised to explain the presence of stationary states of fitness in populations of RNA viruses [7]. The model combines a deterministic mean-field description of the growth of a mutating replicators' ensemble with population bottlenecks applied stochastically after a fixed development time. This second step introduces selection in the system. Our model is general enough that it can describe different experimental situations found in the literature, and therefore can provide a general framework to better understand the *in vitro* evolution of populations of mutating replicators. We will focus on the study of the fitness distribution in different dynamical regimes. Our analytical and numerical results will be compared with available experimental data.

The deterministic growth of the population is implemented as follows. Consider a heterogeneous ensemble where each individual is characterized by a fitness f. This quantity represents the number of offspring produced after one generation g. To introduce phenotypic variations in the progeny, we consider a simple situation where deleterious and beneficial mutations happen with probability p and q, respectively. Here $1 > p > q > 0$. With probability $(1 - p - q)$, offspring inherit the fitness of the parent individual. If a mutation takes place, the fitness of the offspring changes in one unit, plus or minus depending on the mutation being beneficial or deleterious, respectively. The whole process can be studied in two different situations corresponding to the different experimental protocols carried out in the laboratory: (i) The situation of replication plus mutation is repeated a small number g of generations. A small sample (even a single individual) of positive fitness is randomly selected and used as a founder population [14]. The process is repeated.

(ii) The population is allowed to grow without restriction for a very long time, eventually attaining a mutation-selection equilibrium.

The mean-field equations describing the dynamics of the previous model in the deterministic growing phase are as follows. Let us call $n_f(g)$ the number of individuals with fitness f at generation g. The evolution of this quantity depends on replication without error of individuals in that class (this happens with probability $1 - p - q$) and on the contributions from adjacent classes,

$$n_f(g+1) = n_f(g) + (1-p-q)fn_f(g) + p(f+1)n_{f+1}(g) + q(f-1)n_{f-1}(g). \tag{1}$$

There is an open boundary at $f = 1$, such that the previous equation applies with $n_0(g) = 0$, and a reflecting barrier at $f = F$, $n_F(g+1) = (1 + F(1-p))n_F(g) + q(F-1)n_{F-1}(g)$. The initial condition is $n_f(0) = \delta_{f,f_0}$; that is, a single individual of fitness $f_0 \geq 1$ acts as a seed to start the replication process. The total population in the system is obviously $N(g) = \sum_{i=1}^{F} n_i(g)$, and diverges in g. The following are different regimes which can be explored through the above model equations: (I) $p, q \ll 1$, *limited number of generations.* For small enough p and q, the total population grows proportionally to the fitness of the seed individual and $N(g) \simeq (1 + f_0)^g$ during the first generations. (II) *p, q arbitrary, $g \to \infty$.* In this limit, a stationary distribution sets in where the relative number of individuals in each fitness class is constant. The total population grows proportionally to the highest fitness, $N(g) \propto (1 + F)^g$. (III) *p, q arbitrary, limited number of generations.* This crossover region corresponds to the actual dynamical regime where fast-mutating organisms (such as RNA viruses) evolve. In Fig. 1, we represent the growth of the total population as a function of g for the model of Eq. (1). The three different dynamical regimes

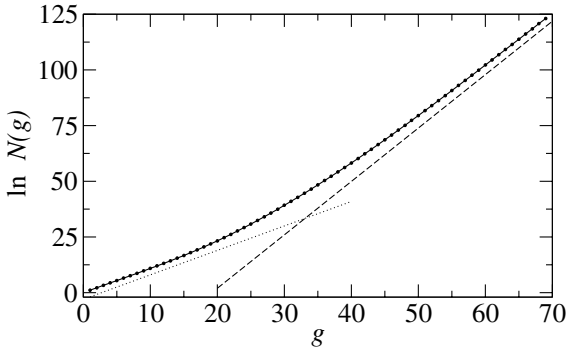

FIG. 1. Crossover between the asymptotic dynamical regimes. We show the growth of the logarithm of the total population as a function of g. Parameters are $p = 0.05$, $q = 0.005$, $f_0 = 2$, and $F = 10$. The dotted and dashed lines have slopes ln(3) and ln(11), respectively.

can be clearly identified. We now explore these three regimes in turn.

Regime (I).—For small mutation rates p and q and as long as g is limited, the total number of individuals in the system are mainly produced through replication of individuals in the fitness class f_0. This reproduces the experimental situation of serial transfers in DNA organisms (for instance, bacteria) where genetic bottlenecks are applied after a certain development time (fixed g). In [16], it was experimentally observed that the effect of Muller's ratchet is much softer on individuals with low mutation rates, where decreases in fitness were observed only after many population bottlenecks and in a relatively small fraction (5%) of the assayed populations. If mutations are sufficiently rare, we can make the approximation that only the fitness classes $f_0 + 1$ and $f_0 - 1$ are produced during the g generations and can be selected as seeds for the next passage. In this regime, the dynamical Eqs. (1) can be written as

$$\begin{aligned} n_{f_0+1}(g+1) &= (f_0+2)n_{f_0+1}(g) + qf_0n_{f_0}(g), \\ n_{f_0}(g+1) &= (f_0+1)n_{f_0}(g), \\ n_{f_0-1}(g+1) &= f_0n_{f_0-1}(g) + pf_0n_{f_0}(g), \end{aligned} \tag{2}$$

where we have assumed that the class f_0 contributes to the population of adjacent classes $f_0 \pm 1$ but that the back contribution is negligible. We are also assuming that $q \ll p$, such that $n_F(g) \simeq 0$ due to the effect of the ratchet. Those equations can be explicitly solved to yield

$$\begin{aligned} n_{f_0+1}(g) &= qf_0[(2+f_0)^g - (1+f_0)^g], \\ n_{f_0}(g) &= (1+f_0)^g, \\ n_{f_0-1}(g) &= pf_0[(1+f_0)^g - f_0^g]. \end{aligned} \tag{3}$$

At this point, it is of interest to know the statistical distribution of seed values f_0 when serial bottlenecks are applied. In many plate transfer experiments, the relevant measured quantity is the population size after a fixed time and after each bottleneck event. Though f_0 is a parameter in the initial deterministic growth of the population, it becomes a random variable when the process is coupled to stochastic selection through population bottlenecks. Using (3), the problem of calculating the distribution of initial fitness values $P_I(f_0)$ can be mapped onto a Markov chain with state-dependent transition probabilities: $p_{f_0,f_0\pm1} = n_{f_0\pm1}/n_{f_0}$, where $p_{f_0,f'}$ is the conditional probability to move to state f' if the system is at state f_0. For matrices such that $p_{f_0,f'} = 0$ if $|f_0 - f'| > 1$, the invariant distribution $P_I(f_0)$ can be calculated in a particularly simple way [17]:

$$P_I(f_0) = \frac{p_{1,2}p_{2,3}\cdots p_{f_0-1,f_0}}{p_{2,1}p_{3,2}\cdots p_{f_0,f_0-1}} P_I(1), \tag{4}$$

where the normalization factor $P_I(1)$ has to be chosen in such a way that $\sum_i P_I(i) = 1$. We finally get

$$P_{\mathrm{I}}(f_0) = P_{\mathrm{I}}(1)\left(\frac{q}{p}\right)^{f_0-1}\left(\frac{f_0+1}{2}\right)^{g}, \tag{5}$$

$$P_{\mathrm{I}}(1) = 2^{g}[\Phi(q/p, -g, 2) - (q/p)^{F}\Phi(q/p, -g, 2+F)], \tag{6}$$

where $\Phi(z, s, a) = \sum_{k=0}^{\infty} z^k/(a+k)^s$ is Lerch's transcendent function. For this regime, the expected distribution of population numbers $Q(N) \simeq \sum_{f_0} P_{\mathrm{I}}(f_0) \times \delta[N - (f_0+1)^g]$ upon serial transfers becomes a stretched exponential function,

$$Q(N) \propto N \exp\{N^{1/g}\ln(q/p)\}. \tag{7}$$

Regime (II).—In the limit $g \to \infty$, a stationary distribution of fitness values $P_{\mathrm{II}}(f)$ sets in [18]. In this case, we consider a single, exponentially growing population. In order to solve the model equations (1) in this regime, we will use a theorem by Hoppe [19] derived in the context of supercritical multitype branching processes. His result applies to our case, in particular, and states that (i) the relative proportions of the various fitness types stabilizes to a deterministic distribution $P_{\mathrm{II}}(f)$, (ii) the growth rates of all types are identical, and (iii) this asymptotic growth rate ρ corresponds to the maximal eigenvalue of the mean matrix M. The elements $m_{f_0,f}$ of the mean matrix M are the expected number of individuals with fitness f in the first generation conditional on having as a seed an individual of fitness f_0 [20]. For our case,

$$m_{f_0,f} = \begin{cases} q, & \text{if } f = f_0 + 1 \\ (1-p-q)f_0 + 1, & \text{if } f = f_0 \\ p, & \text{if } f = f_0 - 1. \end{cases} \tag{8}$$

For $p, q \ll 1$, the eigenvalues λ_i of M are just the growth rates of individuals in each fitness class, $\lambda_i = i + 1$, for $i = 1, \ldots, F$. The largest eigenvalue and the one which dominates the growth of the population in regime (II) is $\rho \simeq F + 1$ (see Fig. 1 for a numerical example). This approximation worsens for increasing F [21].

Taking into account that for $g \to \infty$ the ratios

$$\frac{n_f(g+1)}{n_f(g)} \equiv \rho_f(g) \to \rho\,, \qquad \frac{n_{f+1}(g)}{n_f(g)} \equiv \beta_f(g) \to \beta_f, \tag{9}$$

the stationary solution for the dynamic equations (1) amounts to solving the recursion

$$\beta_f = \frac{\rho - 1 - f(1-p-q)}{p(f+1)} - \frac{q(f-1)}{p(f+1)\beta_{f-1}}, \tag{10}$$

with $\beta_1 = (\rho - 2 + p + q)/(2p)$ and $\beta_{F-1} = (F-1)q/(\rho - 1 - F(1-p))$. The coefficients β_f are directly related to the asymptotic probability distribution of fitness values, $P_{\mathrm{II}}(f) = P_{\mathrm{II}}(1)\beta_1\beta_2\ldots\beta_{f-1}$, where again $P_{\mathrm{II}}(1)$ is obtained from the normalization condition. The recursion (10) cannot be explicitly solved for arbitrary parameters. In Fig. 2, we show the result of a numerical simulation with the full mean-field model and compare it with the solution for the parameters given.

There are a number of experiments with viral [22] and bacterial [23] populations which intend to optimize its fitness. Duarte and co-workers [22] compared a high fitness clone of an RNA virus with a diverse subpopulation of the same clone. The relative fitness f^E of the subclones was measured against the original clone, of fitness F^* and near maximal for the environment considered. Their results are reproduced in the inset of Fig. 2. Interestingly, two apparently puzzling empirical observations, namely, the very broad distribution of fitness obtained in a highly adapted subpopulation and the fact that the maximal fitness class is *not* the most represented class, are both expected results in a population of replicators with high mutation rates [24].

Finally, a recent model [11] has explored the properties of Muller's ratchet in the approximation where the system can attain mutation-selection equilibrium between bottlenecks—equivalent to the limit $g \to \infty$, and assuming $q = 0$. If we would carry out serial transfers in regime (II), the distribution of initial values $P_{\mathrm{II}}(f_0)$ becomes trivially $P_{\mathrm{II}}(f)$, since the information on the initial condition is lost for $g \to \infty$.

Regime (III).—This regime shares some features with regime (I), though Eqs. (1) cannot be fully solved and we resort to numerical simulations to estimate the stationary distribution of initial fitness values $P_{\mathrm{III}}(f_0)$. Populations of fast-mutating replicators evolve in this regime: The frequency of variants more than one step away from

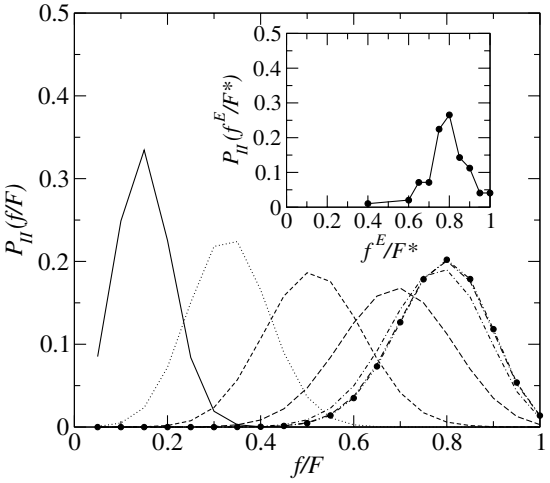

FIG. 2. Regime (II). Transient probability distributions of fitness values and asymptotic shape for $g \to \infty$. Parameters are $p = 0.25$, $q = 0.01$, and $F = 20$, for which $\rho = 16.0027$. The distribution P_{II} is shown every 20 generations. The inset shows experimental results obtained from an optimized, high-fitness, RNA viral quasispecies [22]. Just for comparison, fitness classes have been defined relative to the highest fitness variant.

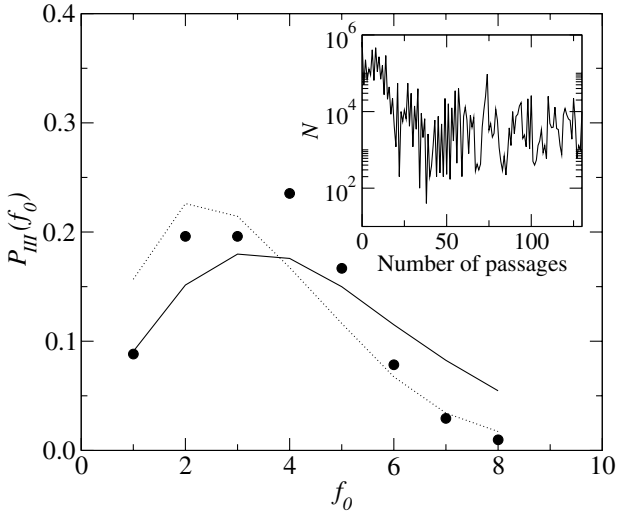

FIG. 3. Regime (III). Experimental distribution of initial fitness values estimated as $f_0 \simeq N^{1/5} - 1$ (symbols). For qualitative comparison, we show the results of our model in regime (III) with $p = 0.07$ and $q = 0.03$ (solid line) and $p = 10^{-3}$, $q = 3.5 \times 10^{-4}$ (dotted line). In both cases $F = 8$ and $g = 5$, as in the experimental case. Inset: Experimentally measured viral yield through serial transfers with foot-and-mouth disease virus [8].

the seed is high enough for f_0 to jump by more than one unit in the simulations, but g is low enough so that the selection-mutation equilibrium has not been reached. In [8], a viral population was subjected to more than 100 sequential transfers with bottlenecks of unit size. After an initial transient where the average fitness of the seed (measured in terms of the viral yield) clearly decreased, a stationary state with large fluctuations settled in (see Fig. 3). With these experimental measures, and considering that the estimated number of generations was small ($g \simeq 5$), we can roughly calculate the experimental distribution of fitness values for the seed after the bottlenecks, $P_{\rm III}(f_0^E)$: $f_0^E \simeq N^{1/g} - 1$ [25]. The experimental distribution $P_{\rm III}(f_0^E)$ is represented in Fig. 3 together with numerically obtained distributions in regime (III).

The model studied here admits a number of generalizations in order to better represent the internal complexity of evolving organisms. Including population fluctuations and the design of a more careful mapping between mutations of the genotype and phenotypic effects will be the subject of future work.

The large number of experiments studying *in vitro* evolution of fast replicating organisms is extraordinarily increasing our current understanding of evolutionary mechanisms. Simple models and statistical approaches such as the one employed here have the potential to provide general scenarios where a variety of experimental situations can be unified.

S. C. M. thanks the MCyT for support under the Ramón y Cajal program.

[1] J. W. Drake and J. J. Holland, Proc. Natl. Acad. Sci. U.S.A. **96**, 13 910 (1999).
[2] J. W. Drake, B. Charlesworth, D. Charlesworth, and J. F. Crow, Genetics **148**, 1667 (1998).
[3] D. R. Mills, R. L. Peterson, and S. Spiegelman, Proc. Natl. Acad. Sci. U.S.A. **58**, 217 (1967).
[4] I. S. Novella, E. A. Duarte, S. F. Elena, A. Moya, E. Domingo, and J. J. Holland, Proc. Natl. Acad. Sci. U.S.A. **92**, 5841 (1995).
[5] C. Escarmís, M. Dávila, and E. Domingo, J. Mol. Biol. **285**, 495 (1999).
[6] H. J. Muller, Mut. Res. **1**, 2 (1964).
[7] E. Lázaro, C. Escarmís, E. Domingo, and S. C. Manrubia, J. Virology **76**, 8675 (2002).
[8] C. Escarmís, G. Gómez-Mariano, M. Dávila, E. Lázaro, and E. Domingo, J. Mol. Biol. **315**, 647 (2002).
[9] I. Gordo and B. Charlesworth, Genetics **154**, 1379 (2000).
[10] A. S. Kondrashov, Genetics **136**, 1469 (1994).
[11] A. Colato and J. F. Fontanari, Phys. Rev. Lett. **87**, 238102 (2001).
[12] E. Lázaro, C. Escarmís, J. Pérez-Mercader, S. C. Manrubia, and E. Domingo (to be published).
[13] L. S. Tsimring, H. Levine, and D. A. Kessler, Phys. Rev. Lett. **76**, 4440 (1996).
[14] We restrict to individuals of positive fitness to mimic the protocols in [8]. Allowing for individuals of class $f = 0$ to act as seeds implies that the extinction threshold can be crossed in one time step, a situation not tackled in this work. However, large fluctuations preceding eventual extinction of the population have been observed [15].
[15] A. Grande-Pérez, S. Sierra, M. Castro, E. Domingo, and P. Lowenstein, Proc. Natl. Acad. Sci. U.S.A. **99**, 12 938 (2002).
[16] D. I. Andersson and D. Hughes, Proc. Natl. Acad. Sci. U.S.A. **93**, 906 (1996).
[17] W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1968), 3rd ed., Vol. XV.7.
[18] M. Eigen and P. Schuster, Naturwissenschaften **64**, 541 (1977).
[19] F. Hoppe, Ann. Probab. **3**, 393 (1976).
[20] K. B. Athreya and P. Ney, *Branching Processes* (Springer-Verlag, New York, 1972).
[21] It is suitable to calculate the asymptotic growth of the population when $q, p \ll 1$, but not to approximate the values of β_f, where corrections of order p and q become relevant.
[22] E. A. Duarte, I. S. Novella, S. Ledesma, D. K. Clarke, A. Moya, S. F. Elena, E. Domingo, and J. J. Holland, J. Virology **6**, 4295 (1994).
[23] R. E. Lenski and M. Travisano, Proc. Natl. Acad. Sci. U.S.A. **91**, 6808 (1994).
[24] N. Rohde, H. Daum, and C. K. Biebricher, J. Mol. Biol. **249**, 754 (1995).
[25] Note that the interval of N values mapping to the same value of f_0^E is indeed large, from $(1 + f_0)^g$ to $(2 + f_0)^g$, so the former approximation turns out to be a very good estimation of the initial fitness.