Soluble Model for the Accumulation of Mutations in Asexual Populations

A. Colato and J. F. Fontanari
Instituto de Física de São Carlos, Universidade de São Paulo,
Caixa Postal 369, 13560-970 São Carlos SP, Brazil
(Received 19 July 2001; published 16 November 2001)

Finite asexual populations can accumulate an increasing number of deleterious mutations by a process known as Muller's ratchet, which consists of successive losses of the fittest or least-loaded classes of individuals in the population. We present here a simplified theoretical framework to describe the serial bottleneck passages setup used in experiments to demonstrate the decrease of the population mean fitness due to the operation of ratchet. In particular, we calculate the expected time between consecutive clicks of the ratchet and derive expressions relating the moments of the mean fitness distribution to the mutation and selection parameters.

DOI: 10.1103/PhysRevLett.87.238102 PACS numbers: 87.10.+e, 87.23.Kg

In an asexual population random loss of all individuals with the fewest mutations, i.e., those in the least-loaded class, is irreversible because the chance of occurrence of back mutations is negligible in very long gene sequences. As the vast majority of new mutations are probably slightly deleterious and the offspring have at least as many mutations as their parents, asexual populations are at risk of degenerating, in the sense that their mean fitness is continually decreasing. This process of irreversible accumulation of deleterious mutations, known as Muller's ratchet, has been advanced as a possible reason for the evolution of recombination since sexual reproduction can recreate individuals which have fewer mutations than their parents [1,2]. Despite recent results questioning the relevance of Muller's ratchet to the evolution of sex [3], the fitness loss in asexual populations, mainly RNA viruses, is already a well-established experimental fact [4,5]. Moreover, the understanding of the operation of the ratchet is of great importance since this process may be involved in the degeneration of the Y chromosome [6] and mitochondria [7].

From a theoretical standpoint one important issue concerns the rate with which the ratchet clicks, usually defined as the inverse of the mean time between successive losses of the least-loaded classes. Since the pioneering work of Haigh, studies of the ratchet rate have been carried out in the framework of the classical Wright-Fisher model of an asexually reproducing population of fixed size N [8]. Although many approximations to the ratchet rate, valid for different ranges of the control parameters of the model, have been proposed [8–12], a reliable general expression for this quantity remains to be obtained yet. From the practical side, criticisms have been directed to the fact that the ratchet becomes effective only for rather small population sizes, which are not usually realized in nature [13]. In addition, the experimental protocol used to verify the operation of Muller's ratchet [4,5] cannot be described by a fixed size population model. In fact, those experiments are based on the passage of a large population through a bottleneck of one or a few individuals, which provides the necessary conditions to engage the ratchet. In Fig. 1 we present the scheme of the serial transfer protocol used in those experiments. In particular, in the seminal experiment of Chao on the RNA bacteriophage $\phi 6$, only one randomly chosen individual is transferred to a fresh plaque (test tube) at each growth cycle, and during the incubation time $T = 24$ h the population increases from that single individual to about 8×10^9 viral particles per plaque [4].

Here we set out to model the bottleneck transmission experiments assuming that the incubation stage lasts long enough for the population to attain the deterministic mutation-selection equilibrium. Then the loss of a class of individuals can occur only in the passage from a test tube to another, which we model by randomly choosing N individuals from the infinite population at equilibrium. The simplicity of the proposed theoretical framework allows

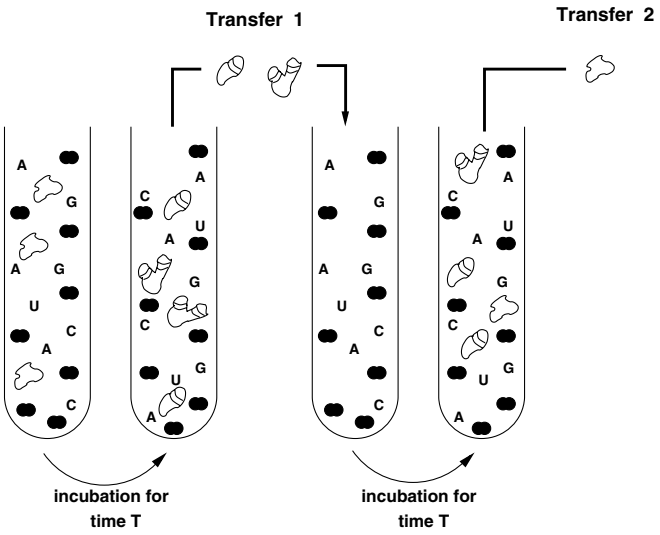

FIG. 1. Scheme of a serial transfer experiment. During the incubation period the population reaches the deterministic selection-mutation balance. Loss of the least-loaded class due to random sampling may occur only in the transfer stage.

us to address elusive issues such as the effect of epistasis among mutations. In particular, it has been suggested that synergistic epistasis, i.e., the harmful effect of a new mutation increasing with the number of mutations already present in the individual, can effectively halt the action of Muller's ratchet [14]. In our framework we show analytically that the ratchet never stops; rather, for sufficiently long times it enters a stationary regime of constant rate, akin to the case where the mutations act independently. Another situation of interest, but of less biological relevance, is that of diminishing epistasis where the disadvantageous effect of a new mutation is attenuated by the previous ones. In addition, assuming the absence of epistasis and a single transferred individual per passage we derive explicit expressions relating the moments of the population mean fitness at a given passage to the mutation rate per genome and the selective cost against mutations.

Although we are concerned with the limit of infinite genomes only, it is instructive to begin modeling each asexually reproducing haploid individual by a sequence of L sites, each one labeled 0 or 1: the bit 0 denotes the correct nucleotide type, and the bit 1 a mutant type. The probability of erroneous replication per site, u, is assumed to be the same for all sites. Taking the limit $L \to \infty$ and $u \to 0$, such that the mean number of new mutations per individual per generation $U = uL$ is finite, yields the celebrated infinite-sites model. In this limit, the probability of a new mutation occurring at a mutant site vanishes as $1/L$ and so back mutations can be safely neglected. Hence the probability that k new mutations occur in one individual is given by the Poisson distribution

$$M_k = e^{-U} \frac{U^k}{k!}. \tag{1}$$

In addition, assuming that all mutations are deleterious, the fitness of an individual with k mutations is

$$w_k = (1 - s)^{k^\alpha}, \tag{2}$$

where $s \in (0, 1)$ is the selective advantage per site of the correct nucleotide type, and $\alpha \geq 0$ is the epistasis parameter. The case $\alpha = 1$ corresponds to absence of epistasis; i.e., each new mutation reduces the fitness of the individual by the same amount, irrespective of the number of previous mutations. Synergistic and diminishing epistasis are described by $\alpha > 1$ and $\alpha < 1$, respectively.

Next we proceed with the modeling of the incubation period, when the population undergoes unlimited growth; i.e., each individual produces very many offspring in a single generation. Individuals in the kth class, i.e., with k mutations, produce a number of offspring which is proportional to w_k; each offspring carries the k mutations inherited from its parent plus a random number i of new mutations, distributed according to Eq. (1). Assuming, as usual, nonoverlapping generations, in a very large population the average number of individuals in class k at generation $t + 1$ is

$$n_k(t + 1) \propto \sum_{j=0}^{k} M_{k-j} w_j n_j(t) \tag{3}$$

so that the frequency $C_k(t)$ of individuals carring k mutations is given by

$$C_k(t + 1) = \frac{n_k(t + 1)}{\sum_{k=0}^{\infty} n_k(t + 1)} = \frac{1}{w(t)} \sum_{j=0}^{k} M_{k-j} w_j C_j(t), \tag{4}$$

where $w(t) = \sum_{j=0}^{\infty} w_j C_j(t)$ is the mean fitness of the population at generation t.

The opposing forces of mutation and selection against deleterious mutations create an equilibrium distribution of mutations across the population, given by the stationary solution of Eq. (4), $C_k(t) = \hat{C}_k$. The distinct equilibrium solutions are identified by the index $m \geq 0$ of the least-loaded class, i.e., $\hat{C}_0 = \hat{C}_1 = \ldots, \hat{C}_{m-1} = 0$, and $\hat{C}_k > 0$ for $k \geq m$. For a given m the equilibrium mean population fitness takes on a very simple form, namely,

$$\hat{w} = w_m e^{-U}, \tag{5}$$

which allows us to write $\hat{C}_k$ in terms of the fitter classes only,

$$\hat{C}_k = \frac{1}{w_m - w_k} \sum_{j=m}^{k-1} \frac{U^{k-j}}{(k - j)!} w_j \hat{C}_j \qquad k > m. \tag{6}$$

Hence starting with $k = m + 1$ we can calculate the ratios $\hat{C}_k/\hat{C}_m$ for $k > m$ recursively and then obtain $\hat{C}_m$ using the normalization condition $\sum_k \hat{C}_k = 1$.

An analytical solution to Eq. (6) is available only in the case $\alpha = 1$ [8],

$$\hat{C}_k = \frac{\theta^{k-m}}{(k - m)!} \exp(-\theta) \qquad k \geq m, \tag{7}$$

where $\theta = U/s$. Interestingly, in this case the frequency of the fittest class is $\hat{C}_m = e^{-\theta}$, regardless of m. Although for $\alpha \neq 1$ the class distribution at equilibrium can be obtained numerically only, we can easily obtain an analytical solution for $\hat{C}_m$ in the limit of large m. In particular, for $\alpha < 1$ we find

$$\hat{C}_m \approx \exp\left[\frac{U m^{1-\alpha}}{\alpha \ln(1 - s)}\right] \tag{8}$$

so that the frequency of the fittest class vanishes exponentially with increasing m, while for $1 < \alpha < 2$ we find

$$\hat{C}_m \approx e^{-U}[1 - U(1 - s)^{\alpha m^{\alpha-1}}]. \tag{9}$$

Similar results hold for $\alpha \geq 2$ as well, except that $\hat{C}_m$ approaches e^{-U} much faster with increasing α. Hence, as far as the frequency of the fittest class is concerned, after very many clicks of the ratchet, synergistic epistasis reduces to the strong selection limit of the case where mutations act independently.

The next stage is to model the random sampling of N individuals from a pool containing an infinite population

at equilibrium with the frequencies $\hat{C}_k$. In this case the probability of picking $n_k \leq N$ individuals in class $k \geq m$ is proportional to $(\hat{C}_k)^{n_k}$. Hence, the best class of individuals in the population is irreversibly lost whenever $n_m = 0$ and the probability that this event takes exactly $\delta\tau = 1, 2, \ldots$ transfers to happen is given by the geometric distribution

$$\mathcal{P}_{\delta\tau} = [1 - (1 - \hat{C}_m)^N]^{\delta\tau - 1}(1 - \hat{C}_m)^N. \qquad (10)$$

Then given that the best class is m the expected number of passages or waiting time for the loss of this class is simply

$$\langle \delta\tau \rangle_m = (1 - \hat{C}_m)^{-N} - 1. \qquad (11)$$

In Fig. 2 we show the dependence of this waiting time on m for $N = 30$ and several values of the epistasis parameter. Clearly, if any class different than the least-loaded one is lost in the sampling process, it will be recreated in the incubation period through mutations of the least-loaded class, leading eventually to the stationary distribution that satisfies Eq. (6). A difficulty arises with the interpretation of $\langle \delta\tau \rangle_m$ as the average number of passages between consecutive clicks of the ratchet because it may happen that classes $m, m + 1, \ldots$ are lost in the same passage; i.e., the ratchet yields many simultaneous clicks. However, provided that the product $N\hat{C}_m$ is not too small, the probability of such simultaneous clicks is negligible and so knowledge of the time between losses of consecutive classes suffices to describe the operation of Muller's ratchet in serial transfer experiments. Of course, simultaneous clicks are the rule rather than the exception in experiments involving the transference of a single individual ($N = 1$) as well as in the case of diminishing epistasis ($\hat{C}_m \to 0$). These caveats apply to the studies of Muller's ratchet based on the Wright-Fisher model as well, though only $\langle \delta\tau \rangle_m$ has been considered in that framework [8–12].

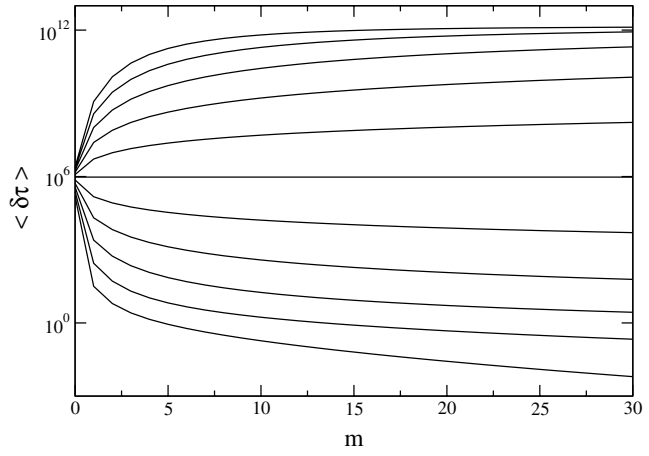

FIG. 2. Expected time $\langle \delta\tau \rangle_m$ between consecutive clicks of the ratchet as a function of the number of mutations m of the lost best class for $s = U = 0.5$, $N = 30$ and (bottom to top) $\alpha = 0.5, 0.6, \ldots, 1.5$.

Our results point out that synergistic epistasis ($\alpha > 1$) does not halt the ratchet, which would be the case if $\langle \delta\tau \rangle_m$ would increase without bounds as m increases [14]. Rather, the ratchet rate $1/\langle \delta\tau \rangle_m$ tends to the constant value $(1 - e^{-U})^N$ after a very long transient regime. The existence of such a long transient may explain the steady increasing of the ratchet rate observed in the simulations of the Wright-Fisher model under synergistic epistatic selection. In fact, those simulations run typically up to 10^4 generations with $N > 100$ [14], which is far from being sufficient to estimate the asymptotic behavior of the ratchet rate. Actually, were it not for the analytical expression Eq. (8), we would not be able to tell from inspection of Fig. 2 whether $\langle \delta\tau \rangle_m$ vanishes or not for large m in the case $\alpha < 1$.

For the parameter setting of Fig. 2, the estimate of $\langle \delta\tau \rangle_m$ in the case of absence of epistasis ($\alpha = 1$) is about 3 orders of magnitude higher than for the Wright-Fisher model. The main reason for this discrepancy is that in the latter model there is an additional mechanism to wind the ratchet on, namely, many different mutations may arise such that each individual in the fittest class acquires at least one new mutation, leading thus to the loss of that class. Since in our framework mutations act only in the incubation stage where the number of individuals in any class is infinite, this additional mechanism does not operate. In this sense, Eq. (11) can be viewed as an upper bound to the value of $\langle \delta\tau \rangle_m$ in the Wright-Fisher model, thus strengthening our claim that synergistic epistasis does not halt the ratchet. We note that in this analysis m is viewed as a fixed control parameter that specifies a particular equilibrium solution of Eq. (4), namely, the solution for which there are no individuals carrying less than m mutations. In the sequel, we show that m is a random variable whose mean $\langle m \rangle$ increases linearly with the number of passages in the absence of epistasis. We expect this result holds for $\alpha > 1$ as well in the case of large τ, while $\langle m \rangle$ should probably increase more rapidly with τ for $\alpha < 1$.

Despite the theoretical interest, the ratchet rate is not accessible experimentally; rather what is measured is the decrease of relative fitness of the transferred individuals using paired-growth experiments. Briefly, in these experiments the individual chosen at passage τ competes with the ancestral type (i.e., a mutation-free individual) in an appropriate medium, its relative fitness being defined as the ratio between the concentrations of the two types of individuals [4,5]. We note that according to Eq. (5) the fitness of the transferred individual at a given passage determines completely the mean fitness of the population in the subsequent incubation period. To describe the experimental setup, henceforth we set $N = 1$ and, since there is no compelling evidence of synergism between mutations (see, e.g., [15,16]), $\alpha = 1$ as well. The relevant quantity to describe Chao's bottleneck experiments is the probability distribution $\mathcal{P}_\tau(m)$ that m is the minimum number of mutations immediately after passage τ or, equivalently,

that one individual of class m is transferred in passage τ. Explicit calculation for $m = 0$ and 1 yields

$$\mathcal{P}_\tau(m) = e^{-\theta\tau}\,\frac{(\theta\tau)^m}{m!}\,, \tag{12}$$

where $\theta = U/s$ as before. The proof that this equation holds for $m > 1$ as well is as follows. Clearly, $\mathcal{P}_\tau(m)$ is nothing but the equilibrium frequency of class m just before passage τ which, according to Eq. (7), depends on the class that has the minimum number of mutations immediately after passage $\tau - 1$. Hence $\mathcal{P}_\tau(m)$ is given by the recursion equation

$$\mathcal{P}_\tau(m) = \sum_{n=0}^{m} \mathcal{P}_{\tau-1}(n) e^{-\theta}\,\frac{\theta^{m-n}}{(m-n)!}\,, \tag{13}$$

which we can easily verify to be satisfied by the Poisson distribution (12). Using Eq. (5) we find that the expected fitness of the population immediately after passage τ is simply

$$\langle \hat{w} \rangle = e^{-U} \sum_{m=0}^{\infty} w_m \mathcal{P}_\tau(m) = \exp[-(1+\tau)U], \tag{14}$$

which, rather surprisingly, is independent of the selection coefficient s. However, the ratio

$$\frac{\langle \hat{w}^2 \rangle}{\langle \hat{w} \rangle^2} = e^{Us\tau} \tag{15}$$

increases exponentially with s and so does the variance. Equations (14) and (15) have great potential for practical use in that they allow a direct estimate of the mutation rate per genome U and the selection coefficient s through measurements of the fitness distribution over different samples used to initiate independent passage series. In fact, we note that plots of the log fitness of the evolved wild type (least-loaded class) relative to the ancestral individual (mutation-free class) *versus* passage number for four samples are well fitted by straight lines with negative slopes [16].

As serial bottleneck passages together with paired-growth experiments are becoming standard techniques to study the effects of mutations, the formulation of a theoretical framework to interpret the experimental results is of great importance and may lead to alternative methods for estimating selection coefficients and mutations rates. In particular, while we have assumed that the virus multiplies extremely rapidly so that the population becomes effectively infinite in a single generation, a careful modeling should take into account the possibility that the ratchet clicks in the few generations immediately after the transfer, when the population is still very small. A more realistic framework to study Muller's ratchet in serial passages experiments should consider then an expanding population with overlapping generations. The present contribution represents the first step to tackle this important but largely unexplored research issue.

The work of J. F. F. is supported in part by CNPq and FAPESP, Project No. 99/09644-9. A. C. is supported by FAPESP.

[1] H. J. Muller, Mutat. Res. **1**, 1 (1964).
[2] J. Felsenstein, Genetics **78**, 737 (1974).
[3] P. D. Keightley and A. Eyre-Walker, Science **290**, 331 (2000).
[4] L. Chao, Nature (London) **348**, 454 (1990).
[5] E. Duarte, D. Clarke, A. Moya, E. Domingo, and J. Holland, Proc. Natl. Acad. Sci. U.S.A. **89**, 6015 (1992).
[6] W. R. Rice, Science **263**, 230 (1994).
[7] C. T. Bergstrom and J. Pritchard, Genetics **149**, 2135 (1998).
[8] J. Haigh, Theor. Pop. Biol. **14**, 251 (1978).
[9] W. Stephan, L. Chao, and J. G. Smale, Genet. Res. **61**, 225 (1993).
[10] D. D. G. Gessler, Genet. Res. **66**, 241 (1995).
[11] P. G. Higgs and G. Woodcock, J. Math. Biol. **33**, 677 (1995).
[12] I. Gordo and B. Charlesworth, Genetics **154**, 1379 (2000).
[13] R. E. Lenski, Science **248**, 901 (1990).
[14] A. S. Kondrashov, Genetics **136**, 1469 (1994).
[15] S. F. Elena and R. E. Lenski, Nature (London) **390**, 395 (1997).
[16] M. de La Peña, S. F. Elena, and A. Moya, Evolution **54**, 686 (2000).