

Nonequilibrium model for estimating parameters of deleterious mutations

Isabel Gordo* and Francisco Dionisio

Instituto Gulbenkian de Ciência, Rua Quinta Grande, 6, Apartado 14, P-2781-901 Oeiras, Portugal
(Received 12 November 2003; revised manuscript received 12 October 2004; published 18 March 2005)

Deleterious mutations are of extreme evolutionary importance because, even though they are eliminated by natural selection, their continuous pressure creates a pool of variability in natural populations. They are of potential relevance for the existence of several features in evolution, such as sexual reproduction, and pose a risk to small asexual populations. Despite their extreme importance, the deleterious mutation rate and the effects of each mutation on fitness are poorly known quantities. Here we analyze a simple model that can be applied to simple experiments, in microorganisms, aiming at the quantification of these values.

DOI: 10.1103/PhysRevE.71.031907

PACS number(s): 87.23.Kg, 87.10.+e

INTRODUCTION

The rate at which deleterious mutations arise in natural populations and their effects on fitness are amongst the most important parameters in evolutionary biology. For example, the value of the mutation rate to deleterious alleles is of key importance in theories of the evolution of sex and recombination, mating systems, mate choice, aging, sex chromosomes, diploidy, and in the explanation of the correlation between variability and recombination [1–4]. Here we analyze a mutation-selection model with strong bottlenecks, which might be applicable to the estimation of these parameters in microbial experimental evolution. Several experiments in a variety of model organisms have been performed in order to estimate the parameters associated with deleterious mutations and/or to understand the consequences of strong reductions in population size, which may be experienced by microbial organisms when they recurrently infect new hosts. These latter experiments typically involve the repeated random choice of a particular bacterium or virus after a period of exponential growth of the population [5,6]. Because of the experimental setup involving the bottleneck of one individual, it is normally assumed that one can safely ignore the occurrence of mutations that cause increments in fitness.

Some models of the evolution of microbial and viral populations that undergo strong bottlenecks have been studied [7–10]. Recently a simple model has been studied that provides an elegant way of estimating the deleterious mutation parameters in microbial populations [7]. However, a very strong assumption is made that is potentially not met in real experiments. The assumption was that after a bottleneck the population restores equilibrium with respect to the evolutionary forces that it is most likely subjected to. The violation of this assumption can have serious consequences in the estimation of the parameters. We analyze this classical population genetics model, both analytically and with Monte Carlo simulations, removing the equilibrium assumption. We provide simple analytical expressions that can be used for

estimating two evolutionarily relevant parameters from experiments.

THE MODEL

Consider an asexual population derived from a single individual that is reproducing until it exhausts the resources in the environment. In every generation each individual survives with a certain probability, which relates to the number of deleterious mutations that it carries. Its offspring will have the same mutations as the parent plus some mutations that may occur during reproduction. Let the number of new mutations be a Poisson random variable with mean U and let the effect on fitness caused by each mutation be s ($0 < s < 1$). Assume that the fitness of an individual with i deleterious mutations is $w_i = (1-s)^i$. It is well known that under this recurrent mutation and selection, the distribution of mutations will achieve an equilibrium given by a Poisson distribution with mean U/s [11]. However, this equilibrium takes some time to reach [12]. We are interested in the distribution of mutations at a given generation g before the population reaches equilibrium. The frequency of individuals with i deleterious mutations follows the recursion

$$f_i(g+1) = \frac{1}{\bar{w}} \sum_{k=0}^i f_k(g) w_k \frac{U^{i-k}}{(i-k)!} e^{-U} \quad (1)$$

where \bar{w} is such that $\sum_i f_i = 1$. Assume, without loss of generality, that at $g=0$, all individuals are free from mutations; then $f_0(0)=1$. At $g=1$, the number of mutations increases by new mutations from a Poisson distribution, therefore $f_i(1)$ will be Poisson with mean U . Selection will not change the distribution but will decrease its mean; namely, after selection $f_i^s(2)$ will be Poisson with mean $U(1-s)$. Because new mutations are Poisson distributed, after mutation the distribution $f_i(2)$ will be Poisson with mean $U(1-s)+U$. Therefore at any given generation g , the distribution of mutations will be Poisson with a mean $\lambda(g)$ that increases with g according to the following recursion:

$$\lambda(g+1) = \lambda(g)(1-s) + U$$

with $\lambda(1)=U$ whose solution is

*Author to whom correspondence should be addressed. Email address: igordo@gulbenkian.pt

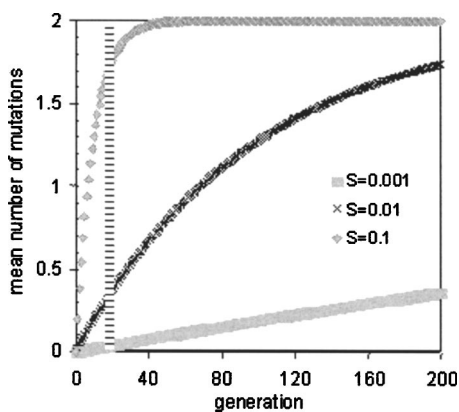


FIG. 1. Approach to equilibrium. The mean number of mutations at equilibrium is 2. The vertical dashed line shows the time corresponding to the bottleneck event.

$$\lambda(g) = \frac{U}{s} [1 - (1-s)^g].$$

Clearly for strong selection ($s \rightarrow 1$) the equilibrium (where the mean number of mutations is U/s) is approached very quickly, but for weaker selection the equilibrium takes a much longer time to approach.

Typically, in experimental plaque-to-plaque transfers, the number of generations between transfers is small. In Fig. 1 we show several possible cases and the time point that most likely corresponds to repeated bottleneck experiments. In all cases the equilibrium mean number of mutations is 2. Except for the cases where $s > 0.1$, the population, at the time of bottleneck, has a mean number of mutations much smaller than 2. This implies that if we assume the equilibrium distribution to estimate U , we will greatly underestimate the mutation rate. At the time g immediately before the first bottleneck the distribution of mutations will be Poisson with mean $\lambda(g) = \lambda$ and the probability of picking a clone with i mutations at the bottleneck is $e^{-\lambda} \lambda^i / i!$. Because $w_i = (1-s)^i$ the mean fitness at the first bottleneck will be

$$\bar{w}_{B=1} = e^{-\lambda s}.$$

Generally in microbial experiments the number of generations until the next passage is around 30 (or less) and several passages are made. We wish to get good estimates of the parameters under such circumstances. If we call $P_B(n)$ the probability of picking an individual with n deleterious mutations at bottleneck B then

$$P_B(n) = e^{-\lambda B} \frac{(\lambda B)^n}{n!}.$$

The expected mean fitness of a clone at bottleneck B will be

$$\bar{w}_B = \sum_{n=0} w_n P_B(n) = \sum_{n=0} (1-s)^n e^{-\lambda B} \frac{(\lambda B)^n}{n!} = e^{-\lambda s B}$$

and, following [7],

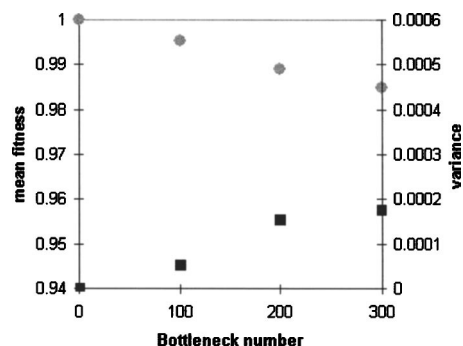


FIG. 2. Mean fitness and bottlenecks. Mean fitness is represented by circles and the variance in fitness by squares.

$$F(\bar{w}) = \frac{\bar{w}^2}{w^2} = e^{\lambda s^2 B}.$$

Therefore the regression of $\log(\text{mean fitness})$ with bottleneck number will give an estimate of λs . Note that this would give a direct estimate of U if the population had achieved mutation-selection equilibrium between bottlenecks, i.e., if the number of generations between bottlenecks were extremely large. The regression of the logarithm of $F(\bar{w})$ with bottleneck number will give an estimate of λs^2 . Therefore by comparing these estimates one can extract estimates of U and s . There is one potential problem with this analysis, which results from the fact that in the initial period after the bottleneck, the population size is so small that there can be deviations from the deterministic expectations calculated above.

STOCHASTIC MODEL

In order to evaluate the importance of these stochastic effects we built a Monte Carlo simulation model in an attempt to mimic an experimental population undergoing mutation, selection, and random genetic drift. This last evolutionary force was modeled using a multinomial distribution on class frequencies as follows.

- (1) Time is measured in generations.
- (2) Every generation, the expected values of the frequency classes of mutations are calculated according to Eq. (1).
- (3) Then the actual numbers of individuals in each class are calculated with a multinomial distribution with those frequencies as parameters.
- (4) The number of individuals doubles each generation, until the population size reaches a maximum (generally the maximum was set to 10^8 individuals). At this point in time the population size crashes to one individual and a new expansion begins.

In Fig. 2 we show the results of \bar{w}_B and the variance in mean fitness as the number of bottlenecks increases, when $U=0.0002$ and $s=0.012$. The results of this figure are very similar to those obtained in experiments using the bacteria *E. coli* [6]. In Table I we show the results of several simulations in an attempt to compare the estimates of U and s with the true values used in the simulations. We chose parameter values that are of potential relevance for the evolution of bac-

TABLE I. Accuracy of the estimated values.

Real		Estimated	
U	s	U	s
0.0002	0.012	0.00018	0.012
0.002	0.012	0.0025	0.01
0.002	0.001	0.0016	0.0012
0.001	0.001	0.00085	0.0013
0.005	0.001	0.0038	0.001
0.5	0.02	0.46	0.024
0.5	0.1	0.5	0.096

teria and viruses [13] for which several plaque-to-plaque transfers are commonly done. For the lower mutation rates, which are closest to those in bacteria, the number of clones used for estimation was 50 and the number of bottlenecks 300. For the higher mutation rates, which are closest to those in viruses, the number of clones was 50 and the number of bottlenecks 15. As can be seen the estimates are quite close to the true values.

EFFECT OF THE BOTTLENECK SIZE

Since it has been observed in experiments, mainly with viruses, that the fitness decline is dependent on the number of microorganisms at the bottleneck, we wish to understand the effect of bottleneck size in the dynamics of fitness decline. In order to get some understanding for the rate of accumulation of deleterious mutations we are going to compare a population that fluctuates in size according to the following equation:

$$N(g) = N_B 2^{g-1}$$

with a population of constant size (N_e) given by the harmonic mean of N :

$$\frac{1}{N_e} = \frac{1}{g_{\max}} \sum_{g=1}^{g_{\max}} \frac{1}{N(g)}.$$

The harmonic mean is normally considered as the effective population size [14] when populations change in size or when the numbers of males and females are different. In Table II we show that there is a pretty good agreement for the decline in $\log(\text{mean fitness})$ between the two models. In the first five cases of Table II the parameters are $U=0.02$ and $s=0.01$ and the decline in $\log(\text{mean fitness})$ was evaluated after 1000 generations, whereas for the last two cases the parameters are $U=0.01$ and $s=0.01$ and the decline in $\log(\text{mean fitness})$ was evaluated after 200 generations (standard deviations are in parentheses). We can conclude that the accumulation of deleterious mutations in a model where the population size fluctuates [$N(t)$] is equivalent to that in a model of constant population size given by the harmonic mean of $N(t)$.

When the bottleneck size is bigger than 1, a very important difference between asexual and sexual populations will

TABLE II. Comparison of the decline in $\log(\text{mean fitness})$ between fluctuating and constant size populations. See text for details.

Bottleneck size	Fitness decline	
	Fluctuating size	Constant size
1	-0.16 (0.04)	-0.19 (0.05)
2	-0.16 (0.04)	-0.17 (0.04)
7	-0.13 (0.03)	-0.12 (0.03)
30	-0.07 (0.03)	-0.07 (0.02)
100	-0.04 (0.01)	-0.04 (0.01)
1	-0.19 (0.05)	-0.19 (0.04)
7	-0.14 (0.03)	-0.15 (0.03)

arise. This is because when $N_B=1$, a particular individual with a particular mutation will be the ancestor of all the individuals of the expanding population, so that a particular mutation will be fixed in the whole population. This means that sexual reproduction will make no difference in the accumulation of deleterious mutations. However, when $N_B=2$, two individuals, possibly with mutations in different genes, can eventually recombine during the expansion of the population and restore genotypes free of mutations. Therefore in these circumstances, a highly recombining population will be at advantage. It has been noticed that HIV-1 [15] shows a very large fitness decline when subjected to bottlenecks of size 1, despite having a high recombination rate. In face of the previous argument this is precisely what is expected. It would be interesting to do the same type of experiments but with a larger bottleneck size, since a much lower fitness decline (or no decline at all) would be expected. In addition, because HIV-1 is a diploid virus and co-infection with two genetically different viruses can result in offspring very different from the parental viruses [16], a model incorporating these features would be needed for the proper interpretation of the results of such experiments.

EFFECT OF COMPENSATORY MUTATIONS

Very recently it has been observed in viruses [17,18] that after many bottlenecks of one individual, the fitness does not continually decline but reaches an approximate stable value, showing fluctuations around a certain mean value. It is clear that a model that assumes that the only types of mutations that affect fitness are deleterious (as the previous model) cannot account for this observation. It is reasonable to think that compensatory mutations become very likely when the fitness of a population of viruses becomes small. We have changed the previous model to incorporate this possibility. We assumed that the probability of compensatory mutations is initially very small, but that this probability increases as fitness decreases. There is experimental evidence for compensatory mutations in virus populations [19] and also in populations of bacteria that carry mutations conferring antibiotic resistance [20]. In Fig. 3 we show an example of the effect of incorporating this assumption. The dynamics of change in fitness across bottlenecks is similar to that observed experi-

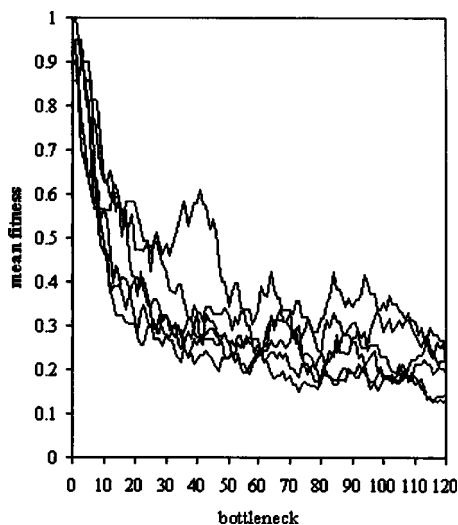


FIG. 3. Effect of compensatory mutations on fitness. Five independent simulations with a $U=0.15$, $s=0.05$, and with the probability of compensatory mutations increasing with decreasing mean fitness.

mentally [18]. Initially there is a decrease in fitness, since compensatory mutations are extremely rare. In this region one can estimate the deleterious mutation rate. After this initial decrease in fitness, compensatory mutations start fixing in the population, releasing fitness decline and avoiding the path toward extinction.

CONCLUSIONS

We have shown that when the strength of selection against deleterious alleles is weak it takes a very long time for a population that is founded by a single individual to reach equilibrium. We have proposed a way of estimating mutation

rates and fitness effects of mutations when the population is in a nonequilibrium state, which can be applied to several mutation accumulation experiments in microbes.

By comparing the simplest constant population size model with a nonconstant one, we found that the decline in mean fitness of a population subjected to repeated bottlenecks, i.e., whose size fluctuates in time according to some function $N(t)$, is similar to that of a population with constant size, whose value is given by the harmonic mean of $N(t)$. This means that approximations for the rate of accumulation of deleterious mutations in a simple model of constant population size [21,22] can be readily applied to populations that fluctuate in size. Finally we have shown how a model that allows for the occurrence of compensatory mutations can account for the dynamics in fitness observed in experimental evolution of some viruses. Recently a slightly different model to simulate viral evolution has been studied [9,10]. In that model, unlike the one here studied, it is assumed that an individual cannot survive if it acquires more than a certain number of deleterious mutations (truncation selection). It is also assumed that both deleterious and advantageous mutations occur. Such a model produces results similar to those obtained experimentally [18] and similar to those obtained in our Fig. 3. The effect of bottleneck size on the accumulation of mutations has also been addressed for that model [9].

Further work incorporating more specific features of each organism (such as recombination and diploidy in the case of the HIV virus) can help interpreting the results and consequences of bottlenecks in the evolution of microbes.

We thank M. G. Gomes for helpful suggestions. This work was supported by project POCTI/BSE/46856 through Fund. para a Ciência e Tecnologia (FCT) and FEDER. F.D. and I.G. were supported by FCT Grants No. SFRH/BPD/14820 and No. SFRH/BPD/8104.

-
- [1] N. H. Barton and B. Charlesworth, *Science* **281**, 1986 (1998).
 - [2] L. Rowe and D. Houle, *Proc. R. Soc. London, Ser. B* **263**, 1415 (1996).
 - [3] B. Charlesworth and K. A. Hughes, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6140 (1996).
 - [4] B. Charlesworth and D. Charlesworth, *Genetica (The Hague, Neth.)* **102/3**, 3 (1998).
 - [5] L. Chao, *Nature (London)* **348**, 454 (1990).
 - [6] T. T. Kibota and M. Lynch, *Nature (London)* **381**, 694 (1996).
 - [7] A. Colato and J. F. Fontanari, *Phys. Rev. Lett.* **87**, 238102 (2001).
 - [8] J. F. Fontanari *et al.*, *Phys. Rev. Lett.* **91**, 218101 (2003).
 - [9] E. Lazaro *et al.*, *Virology* **76**, 8675 (2002).
 - [10] S. C. Manrubia *et al.*, *Phys. Rev. Lett.* **90**, 188102 (2003).
 - [11] J. Haigh, *Theor. Popul. Biol.* **14**, 251 (1978).
 - [12] T. Johnson, *Proc. R. Soc. London, Ser. B* **266**, 2389 (1999).
 - [13] J. W. Drake *et al.*, *Genetics* **148**, 1667 (1998).
 - [14] J. Crow and M. Kimura, *Introduction to Population Genetics* (Harper and Row, New York, 1970).
 - [15] E. Yuste *et al.*, *Virology* **73**, 2745 (1999).
 - [16] D. S. Burke, *Emerg. Infect. Dis.* **3**, 253 (1997).
 - [17] C. Escarmis *et al.*, *J. Mol. Biol.* **315**, 647 (2002).
 - [18] E. Lazaro *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 10830 (2003).
 - [19] C. L. Burch and L. Chao, *Genetics* **151**, 921 (1999).
 - [20] S. J. Schrag and V. Perrot, *Nature (London)* **381**, 120 (1996).
 - [21] I. Gordo and B. Charlesworth, *Genetics* **154**, 1379 (2000).
 - [22] W. Stephan *et al.*, *Genet. Res.* **61**, 225 (1993).